Investigating

the role of the ribonuclease DIS3 in

haematological cancers

Sophie Rebecca Robinson

brighton and sussex
medical school

A thesis submitted in partial fulfilment of the requirements of the University of Brighton and the University of Sussex for the degree of Doctor of Philosophy

July 2016

Abstract

Whole genome sequencing has recently identified DIS3 as a novel tumour suppressor gene in multiple myeloma. DIS3 is a conserved RNA exonuclease and catalytic subunit of the exosome, a protein complex involved in the 3' to 5' degradation and processing of messenger RNA and small RNAs. Messenger RNA processing and degradation is important in controlling gene expression and therefore cellular function, however the role DIS3 plays in the pathogenesis of haematological cancer remains unclear.

Using RNAi as a means to knock-down DIS3, I have performed various functional assays to investigate the consequences of DIS3 loss-of function on myeloma cells. I have investigated cell viability, drug-sensitivity, mitotic errors, apoptosis and the generation of double-strand breaks in both transiently transfected myeloma cells and stable transfected adherent cells. I have also performed transcript profiling experiments in the form of RNA-sequencing to identify possible targets of DIS3 as well as synthetic lethality screens to identify proteins that may be cooperating with DIS3 mutations in myeloma pathogenesis. Overall, DIS3 knock-down did not appear to affect cellular phenotype in these assays, possibly indicating that DIS3 may be conferring a competitive advantage to cancer cells through a mechanism that only occurs in vivo. Alternatively, DIS3 mutations may not be driving tumourigenesis on their own but may either require another cellular pathway to be disrupted, or, may only be required to maintain the tumour rather than initiate it.

In addition to investigating the role of DIS3 in oncogenesis, I have also studied the normal physiological role of DIS3 within the cell. I have confirmed the presence of two alternatively spliced, protein-coding transcripts of DIS3 that differ in the size of their endoribonucleolytic PIN domain. My work has characterised the levels of these two isoforms in cell lines and in tissues from humans with various haematological cancers. Isoform 1 appeared to be the principal transcript in cell lines as well as myeloma and AML patient cells. However, in CMML and healthy controls, the ratios of each isoform are more equal and often isoform 2 is more highly expressed than isoform 1. Activity assays indicated a difference in the ability

of the shorter isoform to degrade circular RNAs, suggesting isoform 2 may have reduced endonucleolytic function. Initial work has also identified a link between the higher expression of isoform 2 in CMML patients and common mutations in the splicing gene SRSF2. This suggests the expression of the endonucleolytically-reduced DIS3 isoform 2, may contribute towards a CMML phenotype.

Although this project was unable to identify the role of DIS3 in myeloma development, there is strong evidence that mutations in this gene are being positively selected and confer an advantage to cancer cells. More sophisticated experiments may need to be conducted whereby the *in vivo* environment is mimicked more effectively, through the generation of a mutant mouse model. Only once we understand the picture more fully, can we begin to design targeted molecular therapies for affected patients.

Contents

Acknowledgements

I would firstly like to thank my supervisor, Dr. Sarah Newbury, for her genuine interest and guidance, both academically and personally, as well as her unequivocal support on this journey, from beginning to end. I would like to thank my other supervisor, Dr. Tim Chevassut, for giving me the opportunity to pursue this PhD and for his friendship and support along the way.

I would like to thank all members of the Newbury and Chevassut labs, both past and present, for all their help and support and also for being great friends and colleagues. Dr. Helen Stewart has been a great source of knowledge and discussion, both academically and socially. My fellow PhD student Ben Towler also deserves a special thanks for his keen mutual interest in my work and insightful discussions. I would also like to extend my appreciation to Clare Rizzo-Singh and Karen Scruby for the smooth running of the lab, I would have been far less productive without them. Finally, to Dr. Stefano Casserta for his wisdom and willingness to help, particularly for his contribution to the flow cytometry work.

I would like to thank my family for all their love and encouragement and my brilliant bunch of friends, all of which I am very lucky to have in my life. To my parents for their ongoing enthusiasm and for preventing me from dropping out of university all those years ago. Life could have been very different indeed! And to my sister Hannah for her love and support and for always being there when I need her.

Finally, I cannot begin to express how grateful I am to my husband Josh, without him, none of this would have been possible. His support, encouragement, understanding and belief in me throughout this rocky journey has been immeasurable. This has not only been one of the biggest challenges of my life but his also, and I'm sure the terms 'Western blot' and 'RNA-seq' will always conjure up unique memories!

Candidate's declaration

I declare that the research in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for degree.

Signed:

Dated:

List of Abbreviations

AML   Acute Myeloid Leukaemia

CLL   Chronic Lymphocytic Leukaemia

LOH   Loss of Heterozygosity

miRNA   MicroRNA

MM   Multiple Myeloma

NGS   Next-Generation Sequencing

NM   Nodular Melanoma

RNAi   RNA interference

Rrp44   Name given to the yeast DIS3 homologue

siRNA   Short-interfering RNA

SNV   Single Nucleotide Variation

SSM   Superficial Spreading Melanoma

VAF   Variant Allele Frequency

WGS/WES   Whole Genome/ Whole Exome Sequencing

mRNA   Messenger RNA

ncRNA   Non-coding RNA

UTR   Untranslated region

tRNA   Transfer RNA

rRNA   Ribosomal RNA

hnRNAs	Heterogenous nuclear RNAs
snRNAs	Small nuclear RNAs
snoRNAs	Small nucleolar RNAs
RISC	RNA-induced silencing complex
PIN	PilT N terminus domain
RNB	Ribonuclease II domain
PROMPTs	Promoter Upstream Transcripts
TSS	Transcription start site
ChIP	Chromatin Immunoprecipitation
SHM	Somatic Hypermutation
CSR	Class switch recombination
ASCT	Autologous stem cell transplant
qPCR	Quantitative real-time PCR
gDNA	Genomic DNA
cDNA	Complementary DNA
AID	Activation-induced cytidine deaminase
LTR	Long terminal repeat
shRNA	Short hairpin RNA
snRNP	Small nuclear ribonucleoprotein

Table of Figures

Chapter 1: Introduction

1.1 The expansive role of RNA within the cell

The central dogma of molecular biology describes how information encoded in the DNA within living cells is converted into functional products that form essential parts of the cellular machinery (Crick, 1970). DNA serves as a set of instructions which are carried by RNA, as a messenger, to the ribosomes where the information is translated into proteins. Proteins serve as effector molecules, carrying out the essential functions of the cell. Due to its important role within the central dogma, messenger RNA (mRNA), as it is known, is the most recognised class of RNA in cellular organisms. Nevertheless, with modern research it has become apparent that the central dogma does not always apply. Often the DNA codes for RNA molecules which are not translated into protein but are themselves functional end-products (Eddy, 2001; Eddy, 1999). These non-coding RNAs can have diverse functions within the cell and broadly fit into three categories: those involved in protein translation, RNA processing or gene regulation. This section aims to explore the structure, function and biogenesis of different classes of RNA within the cell in order to provide an overview of the critical role they play in cellular development and survival.

1.1.1 Messenger RNA

Within the DNA sequence of any organism are thousands of genes, each encoding a single functional molecule. Historically, most genes were thought to encode a single protein and although this is now known not to be the case, proteins are still the most accepted biologically-relevant effector molecules within the cell. The first step in which a gene codes for a protein is through transcription, whereby in the nucleus, the DNA sequence is transcribed into a complementary sequence of RNA, termed messenger RNA (mRNA). The mRNA is then transported into the cytoplasm where it recruits the translational machinery and is converted into a sequence of amino acids which folds into a protein.

Mature messenger RNAs consist of coding regions and untranslated regions (UTRs) as well as a poly(A) tail and a 5' 7-methylguanosine cap. The coding region is composed of codons, each of which codes for a single amino acid on the ribosome. Coding regions usually begin with a start codon and end with a stop codon which signal the start and end of translation to the ribosome respectively. The UTRs lie either side of the coding region, termed the 5' UTR and 3' UTR. The UTR sequences can differ between mRNAs and perform several roles including mRNA stabilisation, localisation and translation efficiency (Barrett et al., 2012). The poly(A) tail is a long sequence of adenine nucleotides added to the 3' end of the mRNA during transcription which promotes nuclear export and translation as well as protecting the mRNA from degradation (Guhaniyogi and Brewer, 2001). The 5' 7-methylguanosine cap is also added during transcription and performs similar roles to the poly(A) tail. As well as addition of the poly(A) tail and 5' cap, the mRNA is also subject to splicing which removes the non-coding sections, named introns, and sometimes alternative splicing to produce alternative transcripts from the same gene.

In the cytoplasm, interactions between proteins bound to the poly(A) tail and the 5' cap cause the mRNA to circularise (Wells et al., 1998). EIF-4E binds directly to the cap and poly(A)-binding protein (PABP) binds to the poly(A) tail. EIF-4G binds to both PABP and EIF-4E, bringing the cap and tail into close proximity (Figure 1.1). This circularisation of the transcript allows for efficient translation by the ribosomes and confers protection from degradation.

1.1.2 Non-protein coding RNAs

The completion of the Human Genome Project in 2001 revealed a surprising finding for the time: that most of the human genome does not code for proteins. Humans were discovered to possess approximately 22000 protein-coding genes, coded for by just 2% of the genome (Consortium, 2004), a number substantially lower than was initially predicted. This finding led to the novel concept of the G-value paradox (Hahn and Wray, 2002), whereby organism complexity is not reflected by the
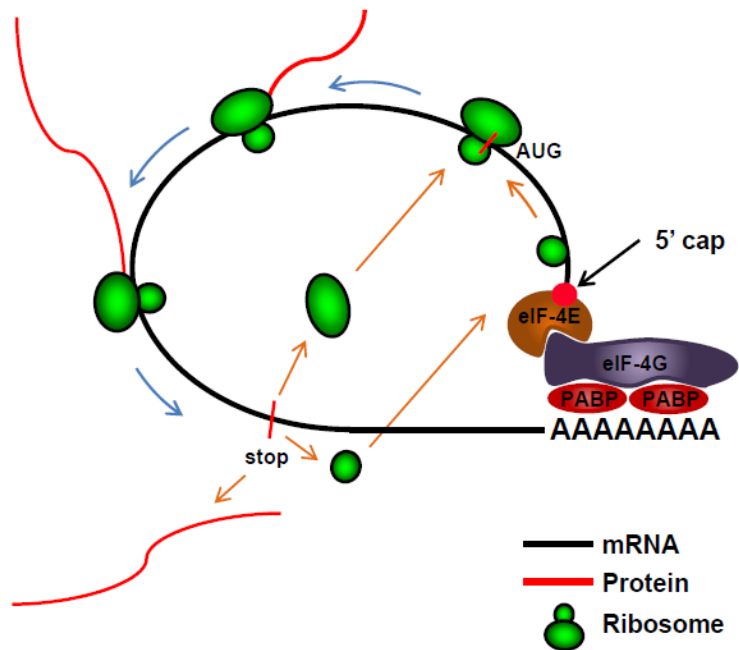
Figure 1.1. Cartoon showing the conformation of actively translated mRNAs. eIF-4G brings together eiF-4E and PABP, which are respectively bound to the 5' cap and poly(A) tail of the mRNA. This brings the 5' and 3' ends of the mRNA into close proximity and allows for efficient recycling of ribosomal subunits. Many other factors are involved in initiation, translation and release, but these are not shown.

number of protein-coding genes in its genome. This paradox has since been explained in part by the discovery that large portions of the genome encode RNAs as an end-product and these non-coding RNAs contribute vastly to cellular function (Eddy, 1999; Eddy, 2001). Indeed, evidence is emerging to suggest that non-coding content may correlate with organism complexity (Taft et al., 2007).

Non-coding RNAs can be broadly arranged into three categories: those involved in protein translation, RNA processing or gene regulation.

1.1.2.1 Non-coding RNAs in translation

Transfer RNAs

The first non-coding RNA to be characterised was an alanine transfer RNA (tRNA). tRNAs are approximately 80 nucleotide-long adapter RNA molecules whose function is to carry a specific amino acid to the ribosome during translation (Sharp et al., 1985). tRNAs have a folded structure with three hairpin loops that form a clover shape. Each tRNA has an attachment site for an amino acid at the top and a three-base pair anti-codon region for complementary base pairing to the mRNA codon at the base (Figure 1.2).

tRNAs are transcribed as pre-tRNAs in the nucleus where they undergo extensive modifications and often splicing out of introns as well as trimming of the 5' and 3' ends to form the mature tRNA molecule. The mature tRNAs are aminoacylated before being exported to the cytoplasm ready to take place in protein synthesis.

Ribosomal RNA

Ribosomal RNA (rRNA) is the catalytic component of the ribosomes, acting as a ribozyme to catalyse peptide bond formation. Eukaryotes contain four different rRNA molecules, three, 5S, 5.8S and 28S (40S), make up the large ribosomal subunit and one, 18S, makes up the small subunit (Brimacombe and Stiege, 1985). These rRNAs compose 60% of the ribosome whilst the remaining 40% is made up of
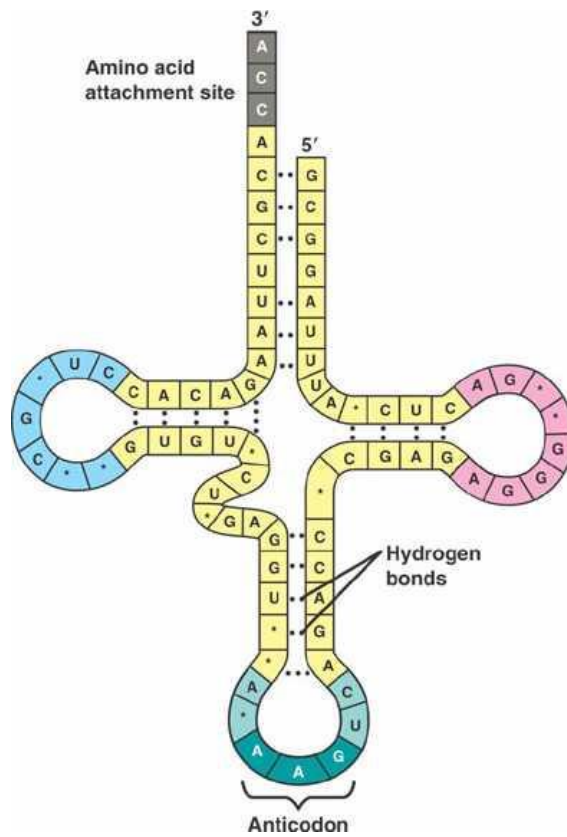
Figure 1.2. The structure of a transfer RNA. See text for details (College of DuPage). http://bio1151.nicerweb.com/Locked/media/ch17/tRNA.html

protein. The 18S, 5.8S and 25S subunits (40S) of rRNA are transcribed as a polycistronic transcript in the nucleolus and 5S is transcribed outside the nucleolus (Figure 1.3). During transcription, small RNAs that guide the addition of chemical groups to other nucleolar RNAs, called snoRNAs, facilitate covalent modifications of rRNA residues including 2'-O-ribose methylation and the production of pseudouridines. Ribosomal proteins are imported into the nucleus and associate with the rRNAs after transcription. The 40S precursor then undergoes a number of cleavage and exonucleolytic steps to produce the three types of rRNA, before being exported into the cytoplasm where they assemble along with the 5S rRNA, to make a functional ribosome.

1.1.2.2 Non-coding RNAs in gene regulation

MicroRNAs

MicroRNAs are short RNA molecules, approximately 20-23 nucleotides in length which generally act to downregulate gene expression at the post-transcriptional level (Bartel, 2004; Bartel, 2009). First discovered in *C.elegans* (Lee et al., 1993; Wightman et al., 1993), miRNAs are evolutionarily conserved and play a role in almost every cellular process. MiRNAs either repress translation or cause degradation of their mRNA targets, typically by base-pairing within the 3'UTR. In plant cells, miRNAs typically base-pair with their targets with almost perfect complementarity which results in cleavage of the target mRNA by RISC (RNA-induced silencing complex) and subsequent degradation of the 3' section. In animal cells, the lower degree of complementarity between miRNA and their targets suggests that translational repression is more prevalent

miRNAs are transcribed as long primary-miRNA (pri-miRNA) transcripts which are cleaved by Drosha and its partner DGCR8 to produce a precursor miRNA hairpin
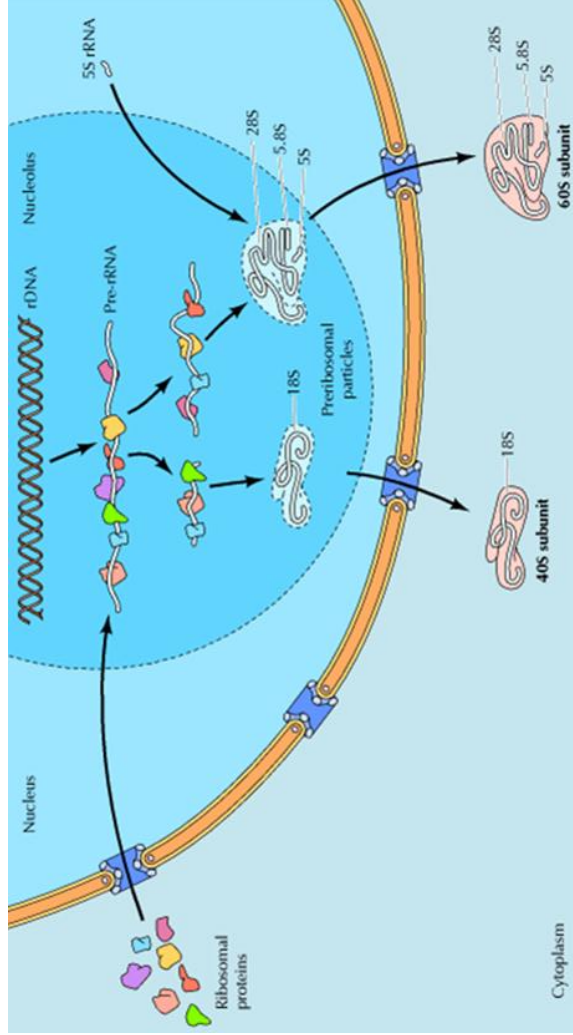
Figure 1.3. Ribosome biogenesis in eukaryotes. The 18S, 5.8S and 25S subunits (40S) of rRNA are transcribed as a polycistronic transcript in the nucleolus and 5S is transcribed outside the nucleolus. Ribosomal proteins are imported into the nucleus and associate with the rRNAs after transcription. The 40S precursor then undergoes a number of cleavage and exonucleolytic steps to produce the three types of rRNA before being exported into the cytoplasm where they assemble along with the 5S rRNA, to make a functional ribosome (Bio-Tech helpline by Rajani Sharma).

27

(pre-miRNAs) (Figure 1.4) (Bartel, 2004). Some miRNAs (mirtrons) are encoded in introns and are processed into a pre-miRNA by the spliceosome. Pre-miRNAs are exported into the cytoplasm before being cleaved by Dicer to produce a short double-stranded duplex. One strand of the miRNA-duplex is selected as the guide strand and is incorporated into an argonaute protein to form the RNA-induced silencing complex (RISC). The other strand (passenger or miRNA*) is usually degraded.

The unique combination of miRNAs contributes to a cell's specific array of protein expression and their misexpression is associated with many types of human cancer (Hata and Lieberman, 2015) (Calin and Croce, 2006). For this reason, miRNA production is itself subject to several levels of regulation (Krol et al., 2010b). As well as transcriptional regulation, post-transcriptional regulation through RNA degradation is also important (Bail et al., 2010; Heo et al., 2008).

Short-interfering RNAs

Short-interfering RNAs (siRNAs) are very similar to miRNAs in that they function in RNA interference (RNAi). SiRNA are derived from longer double-stranded RNAs that can be exogenously introduced but are also endogenously expressed (Hamilton and Baulcombe, 1999; Okamura and Lai, 2008). SiRNAs regulate the same genes that express them, whereas miRNAs regulate other genes. SiRNAs are processed in the same way as miRNAs, however, unlike miRNAs, siRNAs bind with perfect complementarity, resulting in cleavage and complete degradation of their targets.

Piwi-interacting RNAs

Piwi-interacting RNAs are a novel class of small RNAs of between 26 and 31 nucleotides in length that have a role in post-transcriptional gene silencing of retrotransposons in order to maintain genome integrity (Siomi et al., 2011). They
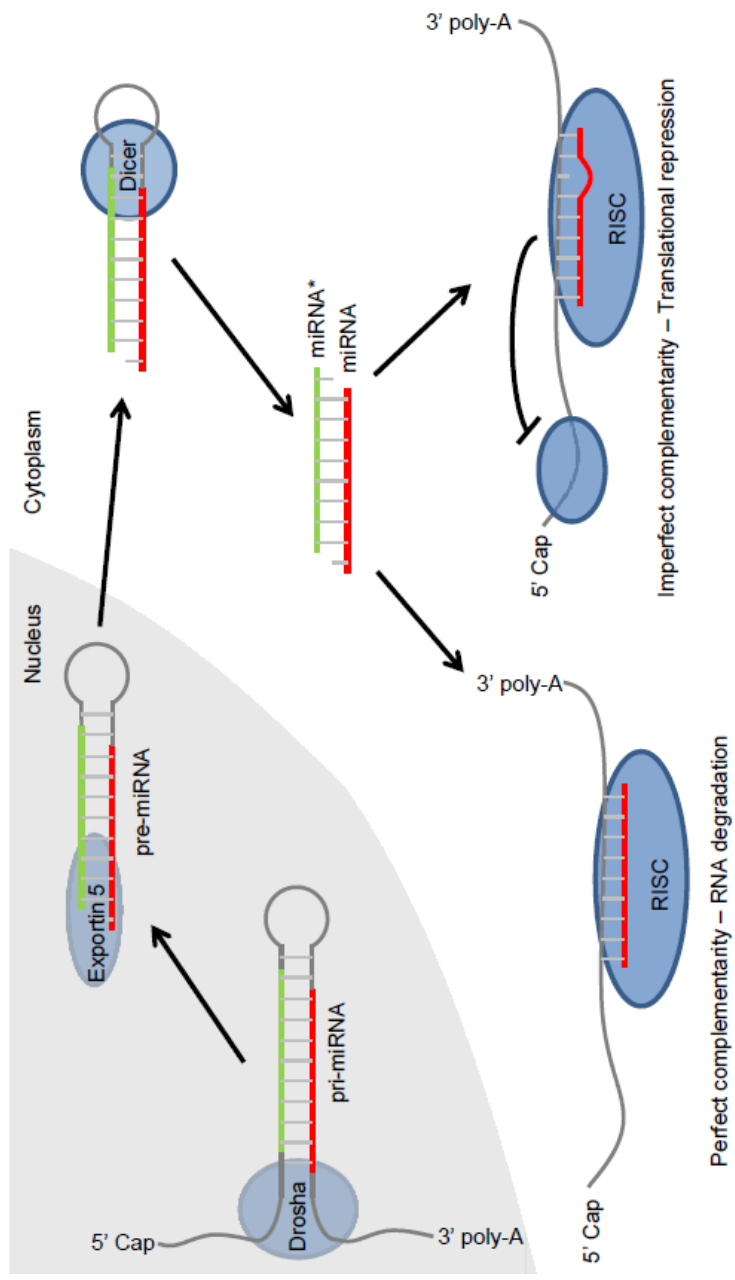
Figure 1.4. The miRNA processing pathway. The pri-miRNA is typically transcribed from intergenic or intronic regions of the genome by RNA Polymerase II and is cleaved in the nucleus by Drosha and its partner DGCR8 (Pasha) to form the pre-miRNA. The 2nt overhang at the 3' end is recognised by Exportin 5 and the pre-miRNA is exported to the cytoplasm, where it undergoes hairpin removal by Dicer-1 to produce the miRNA·miRNA* complex. The mature miRNA is separated from the passenger strand (the miRNA*) and associates with Argonaute, part of the RNA-induced silencing complex (RISC). In some instances, the miRNA* can also associate with Argonaute and itself target mRNAs. The 'seed region' (nucleotides 2-8) of the miRNA play the largest part in determining the target mRNAs. Either translational repression or degradation can occur to mRNAs targeted by the RISC (Bartel, 2004, Valencia-Sanchez et al., 2006, Gu and Kay, 2010).

29

predominantly function in mammalian germ cells and have found to be essential for murine spermatogenesis (Deng and Lin, 2002). PiRNAs interact with piwi proteins that are part of the argonaute family and direct them to their transposon targets, similar to miRNA and siRNA. The biogenesis of piRNAs is not yet fully understood but evidence suggests they are products of long single stranded precursors often encoded by repetitive intergenic sequences in the genome (Thomson and Lin, 2009).

Long non-coding RNAs

Historically, transcription was believed to generate mRNA from non-overlapping regions of the genome with defined gene boundaries. In the 1970s and 1980s the transcription of small heterogenous nuclear RNAs (hnRNAs) such as rRNA, tRNA, snRNA and snoRNA was discovered. However, until the arrival of whole-genome sequencing technologies in the 1990s and 2000s, the extent to which the genome is transcribed, was not fully appreciated. It is now estimated that transcription occurs from 70 -90% of our genome, generating a plethora of RNAs that are not derived from annotated protein coding genes (Carninci et al., 2005). Although still controversial, the concept of 'pervasive transcription' is now widely accepted (Kapranov et al., 2007; Wilusz et al., 2009; Kung et al., 2013) and long non-coding RNAs (lncRNAs) are no longer dismissed as artefacts but may have important functions in gene regulation.

LncRNAs are more than 200 nucleotides in length and are transcribed from within many genomic contexts including intergenic regions, anti-sense DNA, pseudogenes, long introns, promoters and enhancers, as well as overlapping protein-coding genes (Kung et al., 2013). Although the majority of lncRNAs need verification of biological relevance, functional studies indicate important roles for several lncRNAs and many are implicated in a range of developmental processes and diseases (Brown et al., 1992; Ng et al., 2012; Hawkins and Morris, 2010; Wang et al., 2011; Gutschner and Diederichs, 2012).

Long non-coding RNAs have been found to function in both transcriptional and post-transcriptional regulation. For instance, certain lncRNAs can modulate the function of transcription factors by regulating the association and activity of co-regulators or by functioning themselves as co-regulators (Feng et al., 2006). Alternatively, lncRNAs can form RNA-DNA triplexes within promoters to prevent binding of the basal transcription machinery (Martianov et al., 2007). Similar to small regulatory RNAs, lncRNAs function in post-transcriptional regulation often through complementary base pairing where they mask key elements within the mRNA that are usually required to bind trans-acting factors. In splicing, one such anti-sense lncRNA complements an intronic splice site within the 5'UTR of the sense transcript, repressing splicing and thus causing retention of the intron (Beltran et al., 2008). Another lncRNA named BC1, expressed in neurons, interacts with the initiation factors EIF-4A and PABP, preventing their interaction and thus causing repression of translation of certain neuron-specific mRNAs (Wang et al., 2005).

Some of the most famous lncRNAs function in epigenetic regulation. The expression of human Hox genes is controlled by differential chromatin modifications brought about spatially and temporally in the embryo by the interaction of lncRNAs, such as HOTAIR, with chromatin remodelling complexes (Rinn et al., 2007). This lncRNA-directed chromatin modification is also apparent during imprinting and X-chromosome inactivation. The best characterised lncRNA, Xist, is expressed on the inactive X-chromosome and brings about the silencing of X-linked genes in female embryos by coating its length, causing irreversible repressive chromatin modifications (Brown et al., 1992).

1.1.2.3 Non-coding RNAs in RNA processing

Small nuclear RNAs

Before an mRNA can be exported from the nucleus and used in protein synthesis, it must undergo a number of post-transcriptional processing steps (see above). One of these steps, called splicing, occurs concurrently with transcription and involves

the removal of introns from the pre-mRNA and joining of the exons, to form the mature mRNA (Tilgner et al., 2012). For the majority of eukaryotic introns, this process is catalysed by a large complex of RNA-protein complexes, called small nuclear ribonucleoproteins (snRNPs) which together make up the spliceosome. Five snRNPs exist, U1, U2, U4, U5 and U6, each consisting of an approximately 150 nucleotide small nuclear RNA (snRNA) complexed with associated proteins. Splicing occurs as a series of reactions, each stage catalysed by different sets of snRNPs.

The majority of snRNAs are transcribed as pre-snRNA and receive a 5' 7-methylguanosine cap before being exported from the nucleus. In the cytoplasm pre-snRNAs undergo 3' trimming to form a 3' stem-loop as well as hypermethylation of the 5' cap into trimethylguanosine (Hamm et al., 1990). The 3' stem loop signals for the snRNAs to be assembled into snRNPs and the modified 5' cap is required for import back into the nucleus. The role of snRNAs is to recognise the sequences of splicing signals on the pre-mRNA substrate in order to guide the snRNP to the correct site for catalysis.

Self-splicing introns

Another, albeit rare, method of pre-mRNA splicing occurs without the requirement for any proteins. Self-splicing introns act as ribozymes to catalyse their own removal from transcripts without the need for the spliceosome (Cech, 1990). These introns perform splicing using the same biochemical reactions as the spliceosome, using transesterification reactions to join together exons by releasing an intron lariat.

Small nucleolar RNAs

Some classes of RNA require specific chemical modifications in order to generate a mature and functional RNA molecule. Small nucleolar RNAs (snoRNAs) guide the addition of these chemical modifications, namely 2'-O-methylation and pseudouridylation to rRNAs, tRNAs, and snRNAs in the nucleolus (Bachellerie et al., 2002; Eddy, 2001). There are two main classes of snoRNA, C/D box which generally

guide methylation and H/ACA box which guide pseudouridylation (Figure 1.5). SnRNAs are associated with proteins in complexes called small nucleolar ribonucleproteins (snoRNP). The snoRNA contains an antisense region which is complementary to the base on the substrate targeted for modification, enabling the snoRNP to bind to the site and catalyse the reaction. The effect of the modifications on the mature RNAs is not fully understood, however they are thought to enhance folding and RNA-protein interactions.

Most vertebrate guide snoRNAs are encoded in the introns of housekeeping genes that are essential for ribosome biogenesis, suggesting coordinated accumulation of snoRNAs and their targets. Similar to other small RNAs, processing steps such as excision, cleavage and exonucleolytic trimming are all required to produce the mature molecules before RNP assembly.

1.2 RNA Degradation and Maturation

A fine balance must be achieved between the synthesis and degradation of the many classes of RNA molecules within the cell. Mutations that cause defects in RNA turnover can have significant consequences on cellular function (Cairrão et al., 2005; Waldron et al., 2015). Messenger RNAs can be regulated transcriptionally as a means of controlling gene expression but post-transcriptional gene regulation through RNA degradation is also critical. In addition, quality control pathways are in place to remove aberrant mRNAs in order to prevent the accumulation of toxic protein products (Reznik and Lykke-Andersen, 2010). Moreover, as we have seen, the majority of primary transcripts are subject to processing events to produce mature RNAs that have diverse functions within the cell. The focus of this thesis is on mRNA degradation as well as small non-coding RNA processing and decay.

1.2.1 mRNA decay

mRNA degradation in eukaryotes involves a number of complex and interconnected pathways that all converge on three common mechanisms (Figure 1.6). mRNAs
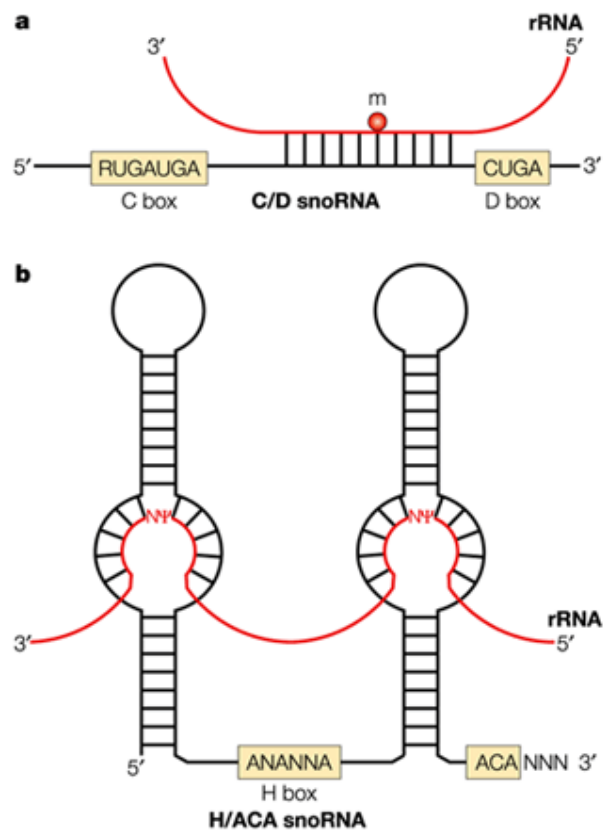
Figure 1.5. The structure and function of snoRNAs. (A) C/D Box small nucleolar RNAs (snoRNAs) use antisense complementarity to target RNA for 2'-O-ribose methylation (site marked with 'm' and red dot). R stands for A or G (purine). (B) H/ACA Box snoRNAs use antisense complementarity in an interior loop to target RNA for pseudouridylation (site marked 'NPsi'). (Eddy, 2001)
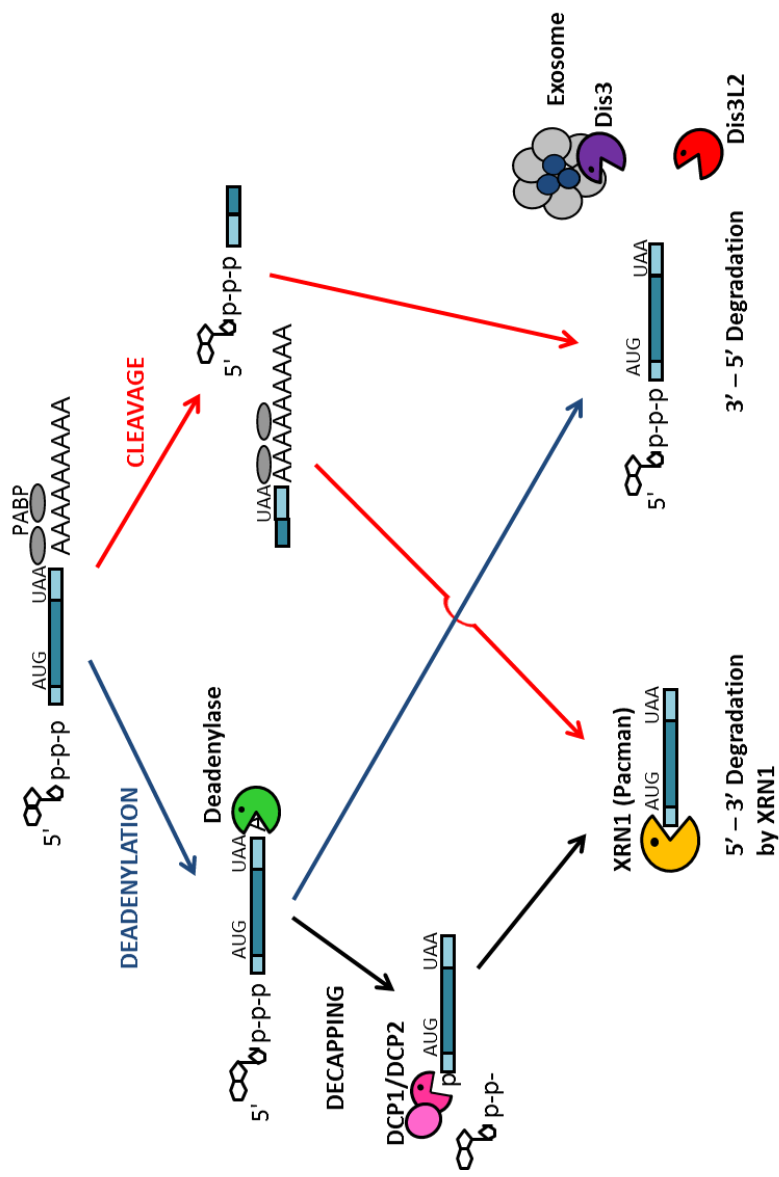
Figure 1.6. Overview of the mRNA degradatio,n pathways in eukaryotes. Messenger RNAs first undergo removal of the 3' poly-A tail (deadenylation), allowing access for 3' → 5' degradation by the exosome complex and DIS3. Following deadenylation, the mRNA can undergo removal of the 5' cap (decapping) exposing the mRNA to degradation by the 5' → 3' exoribonuclease XRN1. Alternatively mRNAs can undergo endonucleolytic cleavage (e.g., due to RNAi, or nonsense-mediated mRNA decay in some organisms) creating two fragments, each of which is susceptible to either XRN1 or the exosome and the DIS3 paralogues.

must first be either deadenylated and de-capped or internally cleaved to allow access for either the exosome and paralogues of DIS3 or the 5'-3' exoribonuclease, XRN1 (Garneau et al., 2007). Deadenylation removes the poly-A tail from transcripts to allow access for 3' to 5' degradation by DIS3 and the exosome (Doidge et al., 2012; Wahle and Winkler, 2013). The 5'-cap may then be removed by decapping enzymes, leaving the transcript vulnerable to 5' to 3' degradation by the XRN1 exoribonuclease (Newbury, 2006). Finally, transcripts can be endonucleolytically cleaved to create two fragments which are susceptible to degradation by either DIS3 and the exosome or XRN1. Many different pathways exist upstream of these processes to target particular substrates for degradation. These pathways can be divided into quality-control and regulated-decay pathways and are described in the following two sub-sections.

1.2.1.1 mRNA quality control pathways

Aberrant and faulty transcripts must be detected by the cell to prevent the production of potentially toxic protein products. Surveillance mechanisms exist both in the nucleus and in the cytoplasm to detect errors at all stages of mRNA production and maturation. In the nucleus, mRNAs that are faulty due to errors in transcription, export or processing are degraded. This includes transcripts in which transcription fails to terminate at the polyadenylation site, pre-mRNAs with retained introns, as well as mRNAs with defective packaging into mRNPs for nuclear export.

Surveillance pathways that take place in the cytoplasm are translation dependent and include nonsense mediated decay (Lejeune et al., 2003), non-stop decay (Frischmeyer et al., 2002; van Hoof et al., 2002) and no-go decay (Doma and Parker, 2006). Nonsense mediated decay (NMD) is the best studied quality control mechanism and is triggered by transcripts that contain a premature termination codon (PTC). Although the mechanism by which this occurs is still controversial and seems to vary between species, a sequence of protein binding events is triggered

which subsequently leads to the decay of the transcript by either the 5' to 3' or 3' to 5' pathway (Kashima et al., 2006).

Non-stop decay targets mRNAs that lack a stop-codon and as a consequence translation continues along the poly-A tail. In yeast and mammalian cells, a stalled ribosome at the 3' end of a transcript is detected and bound by Ski7 which recruits the Ski complex and the exosome to deadenylate and degrade the transcript (Frischmeyer et al., 2002). It is also thought that translation through the poly(A) tail would displace PABP, leading to deadenylation and subsequent decay.

The most recently discovered RNA surveillance pathway, no-go decay, prevents translation of transcripts with a strong secondary structure. Ribosomes stalled by the secondary structure are detected and the mRNA is endonucleolytically cleaved. The endonuclease responsible has not been identified but seems to require Dom34 and HSB1, proteins which are related to the eukaryotic translation release factors, eRF1 and eRF3 (Doma and Parker, 2006). Once the transcript has been cleaved into two fragments, it is degraded by either the exosome or XRN1.

1.2.1.2 Regulated mRNA decay pathways

It was initially believed that mRNA degradation performed more of a quality control role and that gene regulation occurred only at the level of transcription. However, we now know that alternative to the degradation of aberrant mRNAs, mRNAs may be subjected to regulated degradation as a means of controlling gene expression (Garneau et al., 2007). For example, there are instances where transcription does not occur such as in early development, sperm maturation or red blood cell maturation which rely solely on post-transcriptional processes. This involves the regulation of mRNA half-life in order to alter the amount of mRNA which can be translated into protein. Previously, it was assumed that regulated mRNA decay occurs primarily in the cytoplasm through cis-encoded destabilising elements in the 3' UTRs or by the RNA-induced silencing complex (RISC). However, it has become clear that mRNA decay also occurs in the nucleus, facilitated by mRNA processing.

AU-rich elements (ARE) and GU-rich elements (GRE) are two examples of cis-encoded destabilising elements, both found in the 3' UTRs of a number of mRNAs. AREs are characterised by the presence of the AUUUA pentamer, surrounded by a U rich context and are located within the 3'UTR of short-lived mRNAs coding for proteins that mediate rapid regulatory responses in the cell, such as inflammatory or stress responses (e.g., GM-CSF, c-fos, and cmyc) (Bakheet et al., 2006). AREs exert their effect on post-transcriptional gene expression by recruiting trans factors. These ARE-binding proteins (AUBPs) can promote transcript degradation by recruitment of the CCR4-NOT complex resulting in deadenylation of the mRNA and subsequent degradation by either the 5' to 3' or 3' to 5' pathway (Doidge et al., 2012; Wahle and Winkler, 2013; Mukherjee et al., 2002; Kadowaki et al., 1994). The function and abundance of GREs is less understood but they have been shown to regulate a different repertoire of genes and have a more modest effect on mRNA stability (Halees et al., 2011).

MiRNA-mediated degradation of mRNAs provides another means of modulating gene expression. As discussed previously, miRNAs are short RNA molecules which generally act to downregulate target expression by either repressing translation or causing degradation of their target mRNA by the RNA-decay machinery (Figure 1.4). In plant cells, miRNAs typically base-pair with their targets with almost perfect complementarity which results in cleavage of the target mRNA by RISC (RNA-induced silencing complex) and subsequent degradation of the 3' section by AtXRN4. The 5' section is thought to be degraded by the exosome (Huntzinger and Izaurralde, 2011). In animal cells, the molecular mechanisms of miRNA-mediated gene silencing are still not clear, probably due to the existence of a diversity of mechanisms. However, in most cases miRNAs are usually only partially complementary to their targets and direct endonucleolytic cleavage of targets rarely occurs. Although it was previously thought that the levels of miRNA-targeted mRNAs remained unchanged however, there is currently no consensus view on whether mRNA degradation or translational repression is the main mechanism of silencing (Braun et al., 2011; Guo et al., 2010; Fukaya and Tomari, 2012). A number of mechanisms have been proposed through which translational repression occurs.

For instance, GW182, a component of RISC, has been shown to recruit the deadenylase complex CCR4-NOT and PAN2-PAN3 which stimulates the release of the protective Poly-A binding protein (PABP) (Wahle and Winkler, 2013; Doidge et al., 2012). This disrupts the PABP-eif4E interaction resulting in de-circularisation of the transcript and loss of the confirmation required for efficient translation (Braun et al., 2011; Zekri et al., 2013; Chekulaeva et al., 2011; Fabian et al., 2011) Additionally, the recruitment of deadenylases has been shown to result in the association of the helicase eIF4A2, which can act as a clamp in the 5'UTR preventing ribosome progression (Meijer et al., 2013). For targets that are destined for degradation, deadenylation results in decapping and subsequent 5' to 3' degradation by XRN1 (Rehwinkel et al., 2005).

Interestingly, RNA processing can act as a negative regulator of gene expression as opposed to its classical role in generating functional mRNAs. This process targets for example, mRNAs that contain cryptic introns - introns that are usually spliced at a very low frequency. The recruitment of the spliceosome to a cryptic intron causes rapid degradation of the transcript. Therefore, the unprocessed transcripts are functional whereas RNA processing induces decay. This type of response is regulated as splicing is activated in response to external cues such as DNA damage.

1.2.2 Small-RNA processing and decay

Not all RNA substrates that are targeted by the RNA decay machinery are destined for complete degradation. As discussed previously, rRNA, small nucleolar RNA (snoRNA), small nuclear RNA (snRNA) and tRNA species are all transcribed as pre-RNAs, which must then be cleaved and/or trimmed to produce functional small RNA products (Butler, 2002).

As mentioned previously, miRNA-mediated degradation of mRNAs is an important means of modulating gene expression. The unique combination of miRNAs contributes to a cell's specific array of protein expression and their misexpression is associated with many types of human cancer (Hata and Lieberman, 2015). For this

reason, miRNA production is itself subject to several levels of regulation (Krol et al., 2010b). As well as transcriptional regulation, post-transcriptional regulation through RNA degradation is also important. This can occur indirectly through the regulation of RNA-binding proteins such as the cleavage of DCGR8 mRNA by Drosha, leading to its degradation (Triboulet et al., 2009; Han et al., 2009) , or directly by targeting either the pri, pre- or mature miRNA (Heo et al., 2008; Buck et al., 2010; Krol et al., 2010a; Katoh et al., 2009).

1.3 An overview of the ribonuclease DIS3

The focus of this thesis is specifically on an enzyme involved in the 3' to 5' pathway of RNA degradation called DIS3. DIS3 is a highly conserved 3' to 5' exoribonuclease that provides catalytic activity to a multi-subunit complex, the exosome and is involved in pathways of both mRNA decay and small-RNA processing and decay.

1.3.1 Conservation, structure, mechanistic functions and sub-cellular localisation of DIS3

DIS3 was first discovered in *S.pombe* mutants that were _defective in sister chromatid disjoining_ (Ohkura et al., 1988). DIS3 belongs to the RNase II/RNR superfamily and orthologues exist in most organisms from bacteria to humans (Zuo and Deutscher, 2001). Members of this family contain the exonucleolytic RNB domain and show very high sequence conservation as well as functional conservation, demonstrated by the genetic complementation of mutant yeast Dis3 (Rrp44) with the human homologue (Mitchell et al., 1997; Allmang et al., 1999). Some eukaryotes have paralogues of DIS3 and The domain architecture differs slightly between homologues (Figure 1.7, see (Tomecki et al., 2010) for more detail on domain composition).

Human DIS3 has an exonucleolytic RNB domain, two cold shock domains (CSDS) and an S1 domain which non-specifically bind RNA, an endonucleolytic PIN (PilT N terminus) domain (Frazão et al., 2006; Lebreton et al., 2008; Tomecki et al., 2010)
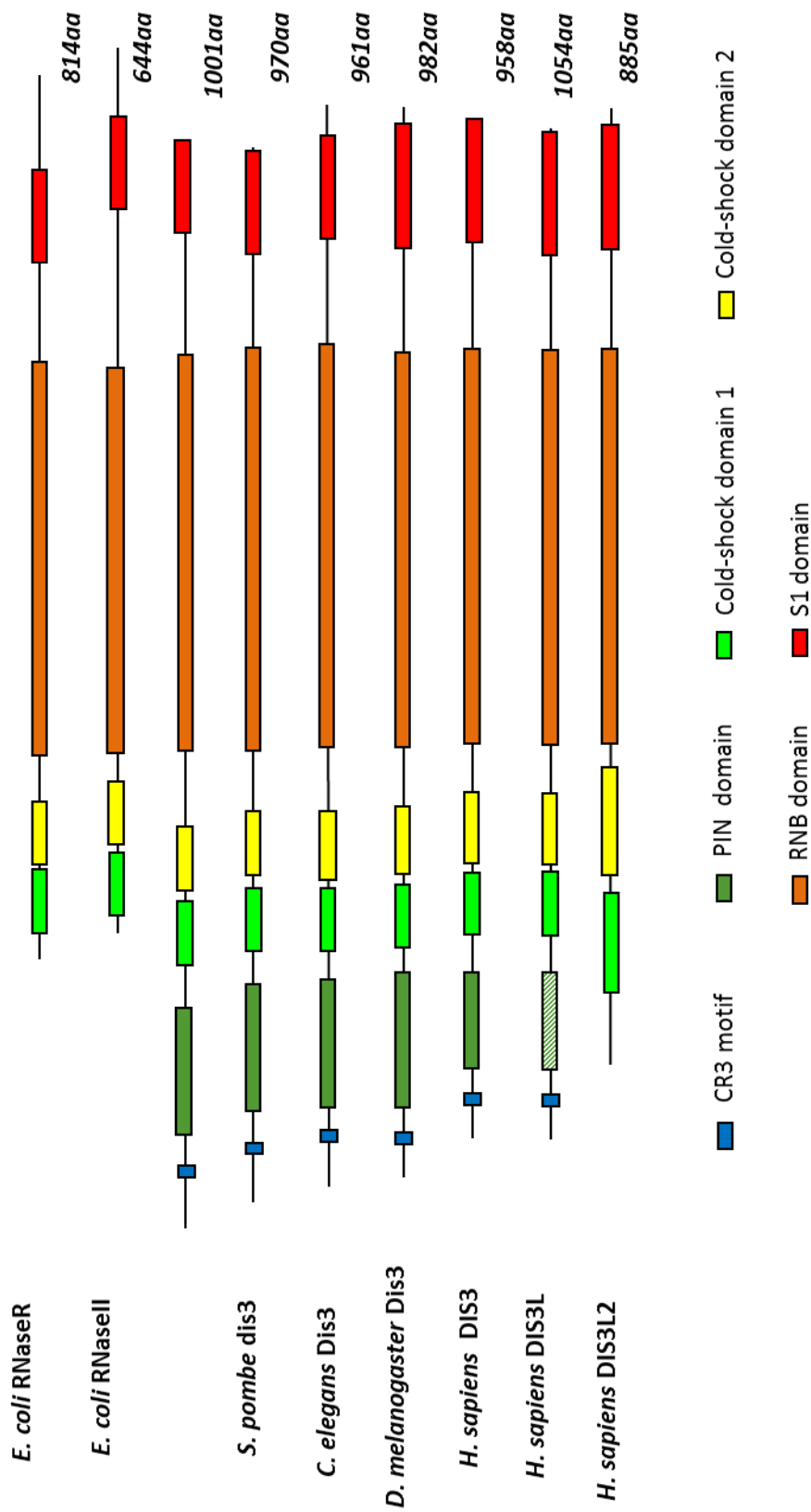
Figure 1.7. Domain organisation of members of the RNR/RNase II superfamily. Members of this family have a similar modular domain organisation. The N-terminal region is variable but cold-shock domain 1 and cold-shock domain 2 are present in all members, followed by an RNB domain and an S1 domain. At the N-terminus, *S. cerevisiae* Rrp44/Dis3 and some other members, also contain a conserved CR3 motif and a catalytic PIN domain. However, mutations in human DIS3L render the PIN domain inactive (hatched). Dis3L2 has lost the N-terminal extension but contains extended cold shock domains. See text for details of domain functions.

and at the N-terminus, a CR3 motif consisting of three cysteine residues that has an important structural role (Schaeffer et al., 2012). Humans also contain two further homologues, DIS3L and DIS3L2, which differ in the presence or absence of the PIN domain (Staals et al., 2010; Tomecki et al., 2010). DIS3L does possess a PIN domain but it is rendered inactive by mutations in two important acidic residues; whereas the PIN domain is completely absent from DIS3L2. Recent evidence suggests that DIS3L2 is a paralogue of DIS3 which functions in different pathways, independent from the exosome (Zhang et al., 2010; Kumakura et al., 2013; Chang et al., 2013). The exonucleolytic RNB domain of DIS3 is exclusive to the RNase R/RNase II family of proteins and is therefore only present in the DIS3 proteins in humans. In contrast, the endonucleolytic PIN domain, first identified in the N-termius of the PilT protein, is also found in proteins involved in nonsense mediated decay, such as SMG5 and SMG6.

Structural analyses and RNase protection experiments have revealed a common model for the mechanism of action of DIS3. DIS3 is a highly processive, hydrolytic enzyme (Schaeffer et al., 2009) that hydrolyses single-stranded RNA in a 3' to 5' direction, releasing one nucleotide at a time and leaving a product a few nucleotides long (Lorentzen et al., 2008). Exonuclease activity is dependent on four conserved aspartic acid residues in the RNB domain that coordinate two magnesium ions in the catalytic centre (Dziembowski et al., 2007; Liu et al., 2006). The RNB active site is buried at the bottom of narrow channel and can only be reached by single-stranded RNA at least 7nt long (Lorentzen et al., 2008).DIS3 can unwind substrates with intra- or intermolecular secondary structures as long as there is an unstructured region of at least 4-5 nt at the 3' end of the RNA. The force of the active site pulling on 3'end of dsRNA accumulates as elastic tension so that about every 4 nt the tension reaches a threshold value and is released in a 'burst' to unwind 4 nt of the duplex at a time (Lee et al., 2012). The endonucleolytic PIN active site consists of four acidic amino acids that coordinate two divalent metal cations and is thought to function in releasing natural exosome substrates that are stalled at sites of strong secondary structure (Lebreton et al., 2008; Schneider et al., 2012). The PIN domain cannot cleave double-stranded RNA but circular and linear

single-stranded RNA are both substrates (Lebreton et al., 2008) and activity is stimulated by a 5' phosphate.

In mammals, DIS3 functions as one of the three catalytic subunits of the exosome, along with DIS3L and Rrp6, a distributive exoribonuclease which belongs to the RNase D family (Graham et al., 2009; Wasmuth and Lima, 2012; Makino et al., 2013). The exosome is a multi-protein complex composed of nine catalytically inert subunits that make up a two-layered barrel-like structure (Figure 1.8). The upper layer is composed of a 'cap' of three S1 or KH domain RNA binding proteins, Rrp40, Csl4 and Rrp4 which rests on a 'core' ring of six proteins, Rrp41-46, all with homology to RNase PH (Symmons and Luisi, 2009). RNase PH enzymes are usually phosphorolytic meaning they use inorganic phosphate as a reactant to cleave nucleotide-nucleotide bonds, however, in the mammalian exosome, the RNase PH homologues are catalytically inactive. The recently solved crystal structure of the *S. cerevisiae* exosome complex shows Rrp44 (Dis3) to be anchored at the bottom of the exosome core through interactions with the PIN domain and CR3 motif (Schneider et al., 2009; Makino et al., 2013). Rrp6 can associate with the exosome cap, forming an 11-subunit complex. *In vitro* evidence suggests that DIS3 can act independently of the exosome although this has not been shown *in vivo* (Wasmuth and Lima, 2012; Graham et al., 2009). *In vitro*, when bound to the exosome, the activities of DIS3 and Rrp6 are suppressed through allosteric effects that diminish their RNA binding ability (Wasmuth and Lima, 2012; Drazkowska et al., 2013). This may suggest that the exosome core modulates the RNase activities as part of a regulatory process that controls RNA decay.

The central channel of the exosome is only wide enough to accommodate single-stranded RNA, so secondary structures must be unwound from the cap by either the nuclear TRAMP complex or cytoplasmic Ski complex (Halbach et al., 2013; LaCava et al., 2005; Symmons and Luisi, 2009). Substrates targeted to DIS3 can either enter the catalytic domain directly or be threaded through the central channel of the exosome to the exo- or endoribonuclease sites at the bottom.
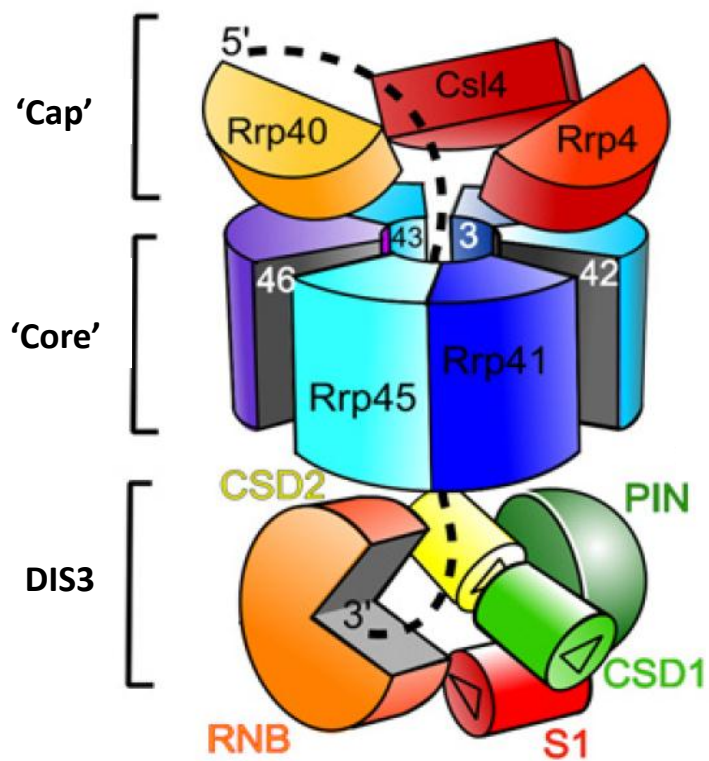
Figure 1.8. The exosome complex in association with DIS3. The inactive 'core' exosome consists of nine subunits. Three subunits form an RNA binding 'cap' structure (shades of red/orange). The remaining six subunits form a ring structure through which the RNA substrate is channelled (shades of blue). The exosome gains its activity by association with DIS3 at base of the ring structure. (Adapted from Luisi, B, 2009).

Recent data suggests that substrates for processing are targeted directly to DIS3, whereas some substrates for degradation must first be threaded through the exosome core (Schneider et al., 2012; Mitchell, 2014). Substrates targeted to Rrp6 are threaded through the central channel and divert off laterally beneath the S1/KH cap to access the Rrp6 active site. It is unknown how RNAs are differentially targeted to DIS3 or Rrp6; it appears stochastic but could be determined by additional factors *in vivo* (Wasmuth and Lima, 2012). Interestingly, Rrp6 appears to enhance the activity of DIS3 in the 11-subunit exosome complex (Wasmuth and Lima, 2012) but the mechanism behind this is unknown. How the activities of these two enzymes cooperate *in vivo* is also unknown, however they are known to work sequentially in the maturation of 5.8s rRNA (Briggs et al., 1998).

Subcellular compartmentalisation of ribonucleases is an important control mechanism in the temporal and spatial regulation of RNA processing and decay. The subcellular localisation of DIS3 homologues and the different exosome subunits has not been investigated in great depth, besides two studies in *Drosophila* (Graham et al., 2006; Graham et al., 2009) and one study in human-derived HeLa and HEK-293 cells (Tomecki et al., 2010). It is generally agreed that DIS3 is nuclear, excluded from the nucleolus with minor pools being found in the cytoplasm; whereas Rrp6 is found in both the nucleolus and nucleus and DIS3L is solely cytoplasmic (Figure 1.9). However, in some *Drosophila* S2 cells, Dis3 has shown restricted localisation to the cytoplasm and the localisation pattern differed from cell to cell (Graham et al., 2006). Furthermore, a flag-tagged version of DIS3 expressed in a stable HEK-293 cell line showed only a nuclear localisation with no signal being detected in the cytoplasm (Tomecki et al., 2010). The functional significance of these localisation patterns remains to be determined. DIS3, along with other exosome components, may localise to different regions of the cell depending on cell-cycle stage or changes in growth conditions. Interestingly, flag-tagged Dis3 co-localises with the nuclear lamina in *Drosophila* cells (Graham et al., 2006). The importance of this nucleo-peripheral localisation is unknown; however, DIS3 could be critical for surveillance during mRNA export. This hypothesis is supported by previous studies that show both *S.pombe* and human DIS3 interact
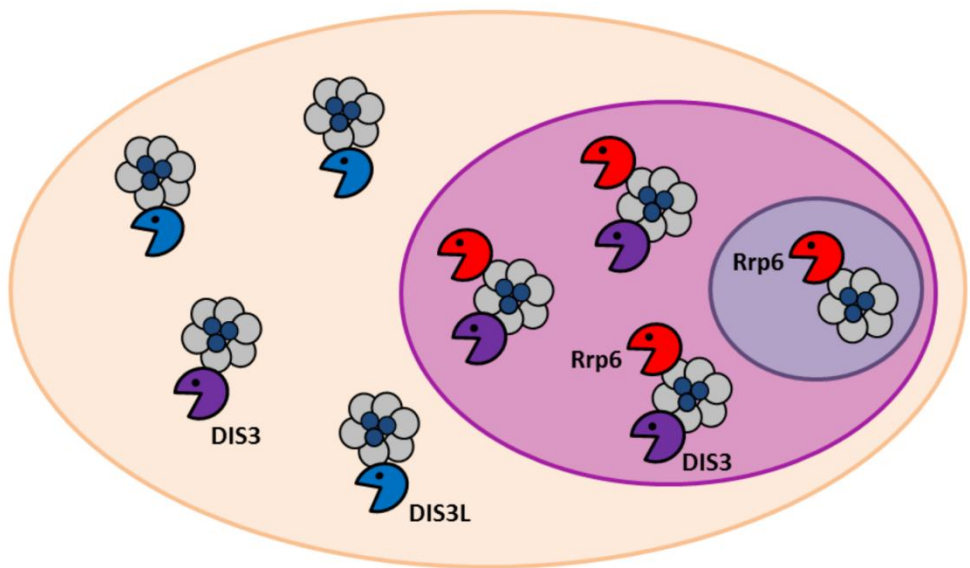
Figure 1.9. Sub-cellular localisation of the different human exosome complexes.
The non-catalytic exosome core (grey) is present in the nucleus, cytoplasm and
nucleolus but associates with different 3' to 5' catalytic subunits depending on the
compartment. In the nucleus the exosome associates simultaneously with DIS3
(purple) and Rrp6 (red). In the cytoplasm the core associates with the cytoplasmic-
restricted DIS3L (blue) and separately with DIS3 but in lower amounts. In the
nucleolus the exosome binds only to Rrp6. It remains to be determined whether
the exosome or the catalytic subunits exist on their own.

with Ran, which is essential for nucleocytoplasmic transport of proteins and ribonucleoproteins (Noguchi et al., 1996; Shiomi et al., 1998).

The predominantly observed nuclear localisation of DIS3 is thought to be controlled by two nuclear localisation signals at the C-terminus of the protein (Graham et al., 2009). DIS3 is known to target both nuclear and cytoplasmic RNAs but it is not known whether a distinct pool of DIS3 proteins exist in each compartment, or if a single, shuttling pool is responsible for the processing and/or turnover of targets in both the nucleus and cytoplasm. N-terminal domains also appear to contribute to DIS3 subcellular localisation but they do not contain nuclear localisation sequences (Mamolen et al., 2010). N-terminal domains may contain an additional regulatory sequence or they may act by maintaining the proper structure of the enzyme, such that the C-terminal nuclear localisation signal is kept in a functional conformation.

1.3.2 Molecular functions of DIS3

The continuous synthesis and degradation of RNAs allow the metabolic changes required for proper cellular function. In association with the exosome, DIS3 is the common effector of a vast range of RNA metabolic pathways functioning in mRNA quality control, regulation of gene expression and small RNA processing. Although not discussed here, the ability of the exosome to handle such a diversity of substrates is down to a network of auxiliary factors which interact with exosome to recruit it to particular substrates (Chlebowski et al., 2013). The following section aims to discuss the role of DIS3 and the exosome in the pathways of mRNA decay and small RNA processing and decay. The relative contributions of Rrp6 and Dis3 to the degradation of the many exosome substrates are still not fully understood, however where specificities are known this has been indicated.

1.3.2.1 Role of DIS3 in mRNA decay

As previously discussed, surveillance mechanisms exist both in the nucleus and in the cytoplasm to detect errors at all stages of mRNA production and maturation.

Both the 3' to 5' and 5' to 3' pathways are involved in nuclear mRNA turnover but which is used depends on the substrate. The exosome is known to specifically degrade un-spliced pre-mRNAs (Bousquet-Antonelli et al., 2000) and mRNAs with defective polyadenylation (Milligan et al., 2005). Interestingly, there is evidence that pre-mRNA surveillance by the exosome takes place during transcription. The interaction and co-localisation with the elongation factor, Spt6, and RNA polymerase II, in *Drosophila*, suggests the exosome may function co-transcriptionally *in vivo* as part of a checkpoint that monitors proper pre-mRNA processing (Andrulis et al., 2002).

Various surveillance pathways take place in the cytoplasm and the 3' to 5' pathway is thought to have some involvement in degrading targeted transcripts. Although not yet clear, evidence suggests that Rrp6 is predominant over DIS3 in targeting NMD substrates (Kiss and Andrulis, 2010). In non-stop decay, in yeast and mammalian cells at least, a stalled ribosome at the 3' end of a transcript is detected and bound by Ski7 which recruits the Ski complex and the exosome to deadenylate and degrade the transcript (Frischmeyer et al., 2002).

DIS3 is also involved in regulated degradation of ARE-containing mRNAs as a means of controlling gene expression (Chen et al., 2001). However, there is currently no evidence to suggest DIS3 or the exosome are involved in the degradation of miRNA-repressed mRNAs in animal cells.

1.3.2.2 Role of DIS3 in small non-coding RNA processing and decay

DIS3 and the exosome were originally discovered in yeast to be involved in the processing of ribosomal RNAs (rRNAs) (Mitchell et al., 1997) and were only later discovered to have a function in mRNA surveillance. rRNA, small nucleolar RNA (snoRNA), small nuclear RNA (snRNA) and tRNA species are all transcribed as pre-RNAs, which must then be cleaved and/or trimmed to produce functional small RNA products (Butler, 2002). The exosome is generally responsible for processing

these stable nuclear RNAs by trimming the extended 3' ends of primary transcripts down to their mature length.

For example, rRNA synthesis in yeast begins with the synthesis of a 35S precursor-rRNA in the nucleolus. The pre-rRNA gets internally cleaved in a series of steps to produce a number of smaller fragments including a 7S intermediate. In the nucleus, the Dis3-exosome complex is required for the 3' end processing of the 7S intermediate into the mature 5.8S rRNA and for the degradation of the 5' external transcribed spacer removed from the full length pre-rRNA transcript (Mitchell et al., 1996). Final maturation of 5.8S rRNA takes place in the cytoplasm where it undergoes exonucleolytic processing at the 3' end, also by the exosome. Additionally, snoRNAs and snRNAs which participate in rRNA processing and modification and pre-mRNA splicing respectively, are excised from polycistronic precursors or from mRNA introns and undergo multi-step 3' processing that involves the nuclear exosome (Allmang et al., 1999).

Dis3 has been found to specifically degrade tRNA species in yeast. This function was first discovered in *S. cerevisiae* tRNA methyltransferase mutants that produce hypomodified tRNAs (Kadaba et al., 2004; Schneider et al., 2007). These tRNAs lack a single modification which may subtly affect their folding but otherwise are mature and functional. The intact exosome lacking only the catalytic activity of Dis3 fails to degrade the hypomodified tRNA, showing this to be a specific Dis3 substrate. Additionally, in yeast Dis3 mutants, both mature and precursor tRNAs are markedly increased (Gudipati et al., 2012). This phenotype is intensified in Dis3 exo-endo– double mutants suggesting PIN activity contributes significantly to tRNA degradation, as expected from highly structured substrates. Interestingly, this study revealed that more than 50% of tRNAs that are transcribed are degraded by Dis3 and never reach the functional pool of mature tRNAs in wild-type cells.

As mentioned previously, miRNA expression is also subject to post-transcriptional regulation. Many miRNAs that are known to be degraded in different organisms have as yet undefined nucleases. Nevertheless, the exosome and sometimes DIS3

specifically have been found to be involved in the turnover of several miRNAs. In *Drosophila* wing imaginal discs, Dis3 knock-down has been found to increase the expression of the mature form of *miR-252-5p* but not the precursor, suggesting Dis3 may be functioning to specifically degrade the mature miRNA as a means of regulation. Another miRNA, *miR-982-5p*, decreases in expression in Dis3 knock-down discs, suggesting Dis3 may be involved in processing the precursor miRNA into its mature form (Towler et al., 2015). Also in *Drosophila*, a family of miRNAs have been discovered that are encoded in introns, which are processed in an exosome mediated biogenesis pathway. These mirtrons bypass normal Drosha cleavage and are processed into pre-miRNAs by the spliceosome. After splicing the 3' tail is trimmed by the exosome (Flynt et al., 2010). Additionally, DIS3 and Rrp6 have been found to be involved in the degradation of pre-miRNAs in mammalian cells. Unlike Rrp6, knockdown of DIS3 does not seem to affect the level of mature miRNAs but does cause an increase in several truncated pre-miRNAs, suggesting DIS3 is involved in the quality control of pre-miRNAs (Liu et al., 2014). Interestingly, this study found the activity of DIS3 on pre-miRNAs to be stimulated by uridine tails, which stimulate the uridyl transferases TUT4 and TUT7, providing a positive feedback-loop in the degradation of Ago-bound pre-miRNAs.

One of the major classes of nuclear exosome substrates in humans is PROMoter uPstream Transcripts (PROMPTs). Similar to cryptic unstable transcripts (CUTs) in yeast, PROMPTS are short-lived ncRNA species, between 200 and 600 nucleotides in length, transcribed upstream of the promoters of active protein-coding genes. PROMPTs are transcribed by any of the three RNA polymerases and have 3' poly-A tails as well as 5' cap structures. Evidence suggests that most, if not all, actively transcribed RNA pol II genes have associated PROMPTs, but they seem to be especially prominent at TATA-less, CpG-rich promoters with broad transcription start site (TSS) regions (Preker et al., 2008). PROMPTs tend to be generated between 500 and 2500nt upstream of the TSS and although not linked with TSS-associated RNAs (formed by RNA Pol II backtracking and stalling), their transcription is positively correlated with downstream gene activity. PROMPTs are currently poorly understood but could function in regulating the expression of downstream

genes by providing reservoirs of RNAPII which facilitates rapid activation of the downstream gene (Preker et al., 2011). PROMPTs are only detected when exosome subunits are depleted. Single knock-downs of DIS3 or Rrp6 yield a much lower stabilisation than double depletion of both catalytic subunits (Preker et al., 2008; Tomecki et al., 2010), however human DIS3 mutant cells show a significant stabilisation of PROMPTs without simultaneous mutation of Rrp6 (Tomecki et al., 2014).

1.3.3 Biological functions of DIS3

Although its role in RNA metabolism is well-documented, the biological functions of DIS3 responsible for the observed phenotypes in mutants are less well known. There are a number of studies in yeast and *Drosophila* which pertain to the biological activity of DIS3 (Table 1.1) however sufficient functional studies of DIS3 in human cells do not exist. Nevertheless, this protein is strikingly conserved across eukaryotes meaning studies in lower organisms may yield useful insight into its function in human cells.

1.3.3.1 Role of DIS3 in cell-cycle regulation

There are a number of studies which present evidence for a role of DIS3 in regulation of the cell-cycle. Dis3 was first discovered in a mutant fission yeast strain to cause non-separation of sister chromatids during anaphase (Ohkura et al., 1988; Kinoshita et al., 1991). Subsequently, the *S.pombe* Dis3 homologue was found to bind to the human GTPase Ran, a member of the RAS superfamily (Noguchi et al., 1996) which functions in spindle assembly and the regulation of cell cycle progression as well as in nucleocytoplasmic transport (Desai and Hyman, 1999; Sazer and Dasso, 2000). RanGTP specifically functions to activate spindle assembly factors by releasing them from complexes with importins (Carazo-Salas et al., 1999; Clarke and Zhang, 2008). At kinetochores, increased Ran-GTP levels displace some spindle assembly checkpoint (SAC) components to allow activation of the anaphase-

Phenotype/ Process Affected	Organism/ Cell Line	Knock-down or mutant (AA position)	Corresponding human DIS3 AA position	Conserved in humans?	Domain	DIS3/ exosome subunit referred to as
Non-separation of sister chromatids	*S. pombe*	P509L	P509 – conserved (Based on predicted *S.pombe* sequence)	Yes	RNB	dis3
Mitotic control and interaction with Ran GTPase	*S. cerevisiae*	G562D, E565K, V566G	G562 E565 V566	Yes, all three	RNB	Dis3sc
Aneuploidy, spindle assembly, metaphase to anaphase transition and kinetochore function	*S. pombe*	P509L	P509 – conserved (Based on predicted S.pombe sequence)	Yes	RNB	dis3
Cell-cycle regulation and microtubule production	*S. cerevisiae* *D. melanogaster*	G562D, E565K, V566G Knock-down	G562 E565 V566	Yes N/A	RNB N/A	Dis3
Larval lethality, no wings	*D. melanogaster*	Knock-down	N/A	N/A	N/A	Dis3
Centromeric transcript turnover and heterochromatin silencing	*S. pombe*, *S. Cerevisiae* (Rrp4)	P509L	P509	Yes	RNB	Dis3/Rrp44
Antibody diversification	*M. musculus* CH12F3 lymphoma cells, Human Ramos B lymphoma cells, HEK-293 cells	Knock-down	N/A	N/A	N/A	Dis3/ Rrp44/ Rrp40, exosome subunit

Table 1.1. List of phenotypes observed in DIS3/exosome mutants and knock-downs in various organisms. Where applicable the corresponding amino acid (AA) in human DIS3 has been given, along with the affected domain.

promoting complex (APC) (Arnaoutov and Dasso, 2003), facilitating cell-cycle progression. Interestingly, the same RNA processing phenotype has been observed in both Dis3 and Ran yeast mutants, suggesting that Ran may regulate the assembly or disassembly of Dis3 and the nuclear exosome (Suzuki et al., 2001).

More recently, *S.pombe* Dis3 mutants have been shown to have elongated metaphase spindles and a block in metaphase to anaphase transition (Murakami et al., 2007). Like Ran, Dis3 appears to be required for correct kinetochore formation and function. The kinetochore consists of many proteins whose functions include anchoring of chromosomes to the mitotic spindle, verification of anchoring, activation of the spindle checkpoint and participation in force generation to propel chromosome movement during cell division (Maiato et al., 2004). Kinetochore formation is monitored by the spindle checkpoint protein Mad2. In single Dis3 mutants, Mad2 restrains mitotic progression but in Dis3 Mad2 double mutants, cells proceed to anaphase without proper chromosome segregation, generating aneuploid cells.

Andrulis et al provide further evidence that DIS3 is involved in mitotic progression as perturbation of Dis3 in *S. cerevisiae* affects microtubule localization and structure (Smith et al., 2011). RNA-seq analysis showed broad changes in the levels of cell cycle and microtubule related transcripts in Dis3 mutant strains. Similar work in *Drosophila* S2 cells showed that the knock-down of Dis3 also predominantly affects the expression of cell cycle-related transcripts (Kiss and Andrulis, 2010). Another study using transgenic flies showed ubiquitous loss of Dis3 to cause larval lethality. In the same study, a spatial knock-down of Dis3 only in the wing pouch region of the imaginal disc was performed, yielding flies with a severe 'no wing' phenotype (Towler et al., 2015) (Figure 1.10), revealing the essential role of this protein in development.

 Although previous siRNA-based experiments in human cells did not show an effect of DIS3 knock-down on growth rate, Tomecki *et al* subsequently showed a mutation-specific effect on the growth of HEK-293 cells. Cell lines were created that

Figure 1.10. Knockdown of Dis3 in *D.melanogaster* wing imaginal disc results in a severe 'no wing' developmental phenotype (right) compared to wild-type (left) (Towler et al, 2015)**.**

expressed inducible exogenous variants of DIS3 with multiple myeloma associated mutations. Cells expressing DIS3 variants with D487N or R780K substitutions proliferated at a slower rate compared with the wild-type cell line. Additionally, when the endonucleolytic PIN domain is mutated alongside mutations in the RNB domain, synthetic lethality and a higher accumulation of PROMPTs are observed, suggesting the two catalytic domains cooperate to degrade substrates. The same group have shown homozygous conditional knock-outs of human DIS3 from the DT40 Cre1 cell line is lethal (Tomecki et al., 2014).

What is the mechanism by which DIS3 is affecting mitosis? One suggestion is that DIS3 could be processing a gene needed for kinetochore formation (Murakami et al., 2007). Another, supported by recent data, is that DIS3 has a link with heterochromatin silencing at the centromere (Murakami et al., 2007; Vasiljeva et al., 2008; Wang et al., 2008). Previously, dense chromatin packaging in heterochromatic regions was thought to inhibit transcription leading to low level gene activity (Grewal and Elgin, 2002). However, recent evidence from budding and fission yeasts suggests that rapid nuclear turnover of heterochromatic transcripts, reinforces transcriptional silencing (Buhler et al., 2007). The deletion of Dis3 considerably increases levels of transcripts from silent centromeric and telomeric loci (Murakami et al., 2007; Vasiljeva et al., 2008; Wang et al., 2008). As the centromere is essential for proper segregation of chromosomes, which is disrupted in Dis3 mutants, centromeric heterochromatin silencing represents a plausible role for DIS3 *in vivo*.

1.3.3.2 Role of DIS3 in generating antibody diversity

Interestingly, the exosome has also been implicated in recruiting activation-induced cytidine deaminase (AID) to chromatin in mammalian B-cells where DIS3 may be functioning specifically in degrading nascent RNA during the DNA repair process (Basu et al., 2011). Activated B-lymphocytes undergo two distinct immunoglobulin (Ig) gene diversification processes: somatic hypermutation (SHM) and IgH class switch recombination (CSR). Both these processes require transcription through

target regions, to open up the DNA duplex. On the resulting ssDNA, AID converts methylated and 5-hydroxymethylated cytidine residues into uracil and thymine respectively. This leads to the creation of double strand breaks (DSBs) which are mostly repaired locally between IgH regions but can also be joined to DSBs occurring elsewhere in the genome resulting in chromosomal translocations.

AID is only functional on ssDNA and so targets the non-template strand during transcription. However, the mechanism by which it accesses the template which is hybridised to nascent transcript is unknown as AID has no known activity on RNA/DNA hybrids. *In vivo* knockdown, ChIP, and physical association studies by Basu *et al*, provide evidence that the RNA exosome functions in targeting AID to both strands of transcribed duplex DNA substrates (Basu et al., 2011). To do this, the exosome along with one of its catalytic subunits, must remove the template RNA. As discussed previously, the exosome has been shown to interact with elongating RNA polymerase II (Andrulis et al., 2002). However, it does not engage RNA substrates that lack a free single-stranded 3' end and RNA still attached to RNA polymerase only has a free 5' end. Therefore, Basu *et al* propose a model whereby the exosome functions on stalled Pol II units that backtrack to reveal a free 3' end (Basu et al., 2011). DIS3 or Rrp6 could degrade the nascent RNA in the 3' to 5' direction, leaving the template strand as ssDNA substrate for AID.

1.3.4 DIS3 and Disease

1.3.4.1 DIS3 and cancer

The earliest study linking DIS3 to cancer identified it as a metastasis-related gene. Two independent gene expression profiling studies of colorectal cell lines and human tissues identified overexpression of DIS3 as high as 38-fold in primary and metastatic tumours relative to normal colonic mucosa (Lim et al., 1997; Liang et al., 2007). Another study has shown a significant overexpression of DIS3 in colorectal carcinomas compared to adenomas (de Groen et al., 2014; Camps et al., 2013). This observed overexpression could be due to an amplification of the DIS3 locus, 13q22,

frequently observed in colorectal cancer. Conversely, the DIS3 locus is often deleted in chronic lymphocytic leukaemia (CLL) and patients have been found to display loss-of-heterozygosity (LOH). Sequenced germline DNA from five families with CLL showed five amino acid changes within DIS3. DIS3 was also shown to be under-expressed 2.8-fold in a CLL leukaemic clone compared with normal B-cells (Liang et al., 2007), suggesting decreased expression is a consequence of the decrease in copy number. This difference in copy number between colorectal cancer and CLL may suggest a tissue-context dependent role of DIS3 in promoting cancer development.

DIS3 has also appeared in linkage studies of breast cancer patients (Rozenblum et al., 2002). However, the significance of this is unclear as it involves polymorphisms rather than deleterious mutations. DIS3 may also be biologically relevant in melanoma. In superficial spreading melanoma (SSM) cells, DIS3 has reduced expression compared to normal melanocytes because one chromosomal copy is deleted. In contrast, DIS3 is overexpressed in nodular melanoma (NM) cells (Rose et al., 2011). Furthermore, SSM cells display sensitivity to mebendazole, a microtubule-destabilizing drug, whereas NM cells are resistant. This is consistent with the function of DIS3 in the regulation of chromosomal segregation during mitosis (see section 4.1) (Murakami et al., 2007). SSM and NM are believed to represent sequential phases of linear progression from radial to vertical growth, yet recurrent differential deletions such as that of DIS3 suggest SSM and NM might be the result of an independent pathway. However, a recent meta-analysis with combined experimental validation of five microarray-based melanoma datasets did not identify DIS3 to be part of a biomarker signature for melanoma (Liu et al., 2013).

Whole genome sequencing studies have identified missense mutations in DIS3 to occur in ~ 4% (4/106) of Acute Myeloid Leukaemia (AML) patients (Ding et al., 2012). In all affected patients, mutations mapped to the exoribonucleolytic RNB domain. AML develops as a clonal evolution of haematopoietic progenitor cells (HPSC/blasts). A HPSC acquires an initiating event which increases its proliferation

and genetic instability, causing the clone to expand. Many subclones evolve from the founding clone leading to an oligoclonal malignant tumour (Welch et al., 2012). Alleles found to have a variant allele frequency (VAF) of 50% usually represent heterozygous somatic mutations that are present in all cells within the sample. DIS3 is mutated in both primary tumour and relapse samples at a VAF between 37% and 47%, suggesting a heterozygous event in these cases. However, whether DIS3 mutations initiate clonal expansion of the HSPC or cooperate to give the clone an additional advantage is still unclear (Ding et al., 2012).

1.3.4.2 DIS3 and multiple myeloma

The most striking association between DIS3 and cancer is probably the finding that DIS3 is recurrently mutated in multiple myeloma (MM). Multiple myeloma is defined by a malignant proliferation of monoclonal antibody (also called M protein)-secreting plasma cells and counts for 20% of deaths related to haematological malignancies (Kyle and Rajkumar, 2004; Laubach et al., 2011).

To date 34 of 306 (11%) myeloma patients analysed by whole genome or exome sequencing studies have been found to have missense mutations in DIS3 that may be functionally relevant (Figure 1.11) (Lohr et al., 2014; Chapman et al., 2011; Walker et al., 2012; Weißbach et al., 2015). In most patients DIS3 mutations correlate with deletions of 13q and a few patients were also found to be associated with copy neutral loss of heterozygosity (cnLOH) that results in the presence of homozygous DIS3 mutations. Six patients have been found to have the same mutation R780K, an occurrence which often indicates activation rather than suppression of a protein. However, most evidence suggests DIS3 is a tumour suppressor gene. A recent amplicon sequencing study identified three hotspot mutations (R780, D488 and E665) within the RNB domain of DIS3 and investigations in HEK-293 cells indicate that the R780K mutation leads to a lower proliferation rate compared to the WT cell line, suggesting a loss-of-function phenotype which would classify DIS3 as a tumour-suppressor gene. Moreover, biochemical assays performed using recombinant versions of DIS3 bearing MM-associated mutations
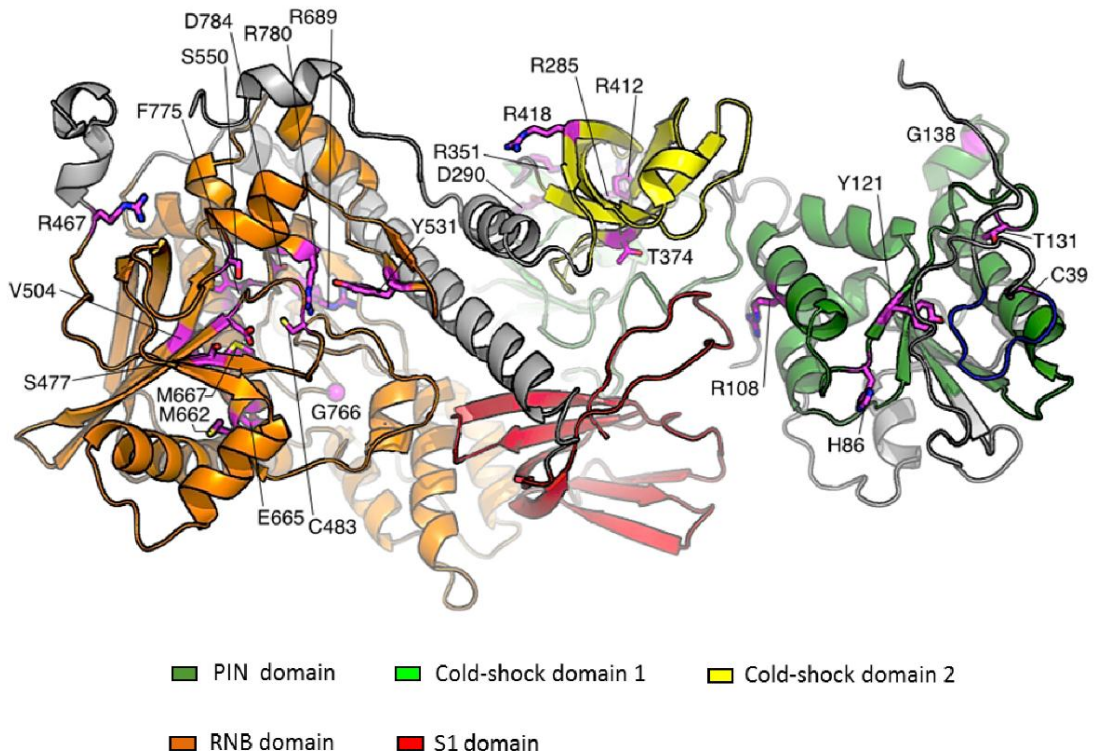
Figure 1.11. Three-dimensional model of DIS3 illustrating the position of amino acids substituted by myeloma-associated mutations. Mutations in pink. DIS3 domain functions – RNB domain: exoribonucleolytic; PIN domain: endoribonucleolytic; CSD1 and CSD2: cold-shock; S1: RNA binding. Modelled according to the recently solved *S.cerevisiae* Rrp44 structure (Makino et al, 2013) using Phyre2 and the webserver 'Site Directed Mutator'. CSD1 is at the back of the structure and is therefore 'shadowed'.

indicate that in the majority of cases, these mutations abolish DIS3 exoribonucleolytic activity (Tomecki et al., 2014). Analyses using available structural information and predictive tools also suggest that most myeloma mutations have a destabilising effect on the enzyme. For example, R780K, found in six multiple myeloma patients, involves an amino acid which is directly involved in binding to the phosphate backbone of the incoming substrate so is highly likely to affect catalysis (for a description of the structure of DIS3 see section 1.3.1). Also, S477R, found in another patient, is a mutation from a small to very large amino acid, with gain of a positive charge. It is next to a loop that contains residues involved in magnesium ion coordination and is therefore also predicted to have an impact on the catalytic activity of DIS3.

Similar to many other cancers, multiple myeloma has been found to develop as a consequence of a clonal evolution of cells (Shlush and Hershkovitz, 2015; Greaves and Maley, 2012; Melchor et al., 2014). In multiple myeloma specifically, the initial immortalisation of the cell usually occurs by the acquisition of a chromosomal abnormality (Morgan et al., 2012). Chromosomal abnormalities can be classified into hyperdiploid (trisomies) or non-hyperdiploid subtypes. Curiously, DIS3 mutations are most commonly seen in non-hyperdiploid subtypes (Lohr et al., 2014; Chapman et al., 2011; Walker et al., 2012). Non-hyperdiploidy involves translocations of the IgH locus with other chromosomes and is caused by aberrant class-switch recombination (CSR) in B-cells.

Genome wide sequencing studies have provided semi-quantitative analysis of the size of the clonal populations carrying a particular mutation within an individual tumour. It was anticipated that mutations arising in all the clones would take part in initiating myeloma whereas mutations present only in some subclones would be potentiators of the disease. However, it appears that the situation is not quite that simple. Mutations in DIS3 were found to be both clonal in some patients and sub-clonal in others meaning they are functioning sometimes as the former and sometimes as latter (Lohr et al., 2014; Weißbach et al., 2015). This observation is not restricted to DIS3 but rather applies to the other ten significantly mutated

genes in myeloma patients, KRAS, NRAS, TP53, FAM46C, BRAF, TRAF3, PRDM1, CYLD, RB1 and ACTG1.

Within patient samples, some patterns of cooperation and exclusion can be identified between mutations in DIS3 and other genes. DIS3 mutations seem to be mutually exclusive with mutations in FAM46C. Collectively, DIS3 and FAM46C mutations are observed in 21% of patients (Lohr et al., 2014). The precise function of FAM46C is yet unknown, however there is evidence it belongs to a family of nucleotidyltransferases (Kuchta et al., 2009) and it was recently shown to function as an mRNA stability factor that interacts with poly- A-binding protein cytoplasmic 1(PABPC1) and binds to CU rich motifs within the 3'UTRs of some mRNAs (Meng, 2010). In support of this function, Chapman *et al* have found its expression to be highly correlated to the expression of a set of ribosomal proteins and translation initiation factors (Chapman et al., 2011). DIS3 is known to function in the maturation of rRNA, suggesting these two genes could be involved in the same cellular pathway.

Conversely, DIS3 mutations always seem to occur in parallel with a hemizygous (monoallelic) deletion of the RB1 region (13q14), either as del (13q) or as an interstitial deletion of the RB1 locus. The gene of interest at 13q14 may be RB1 (retinoblastoma tumour suppressor protein), or one of the miRNAs at this locus which are under-expressed in CLL and MM (*miR-15a/16*). This raises the possibility that mutation and selection of DIS3 as a driver mutation in myeloma is dependent on deletion of 13q14. However, more NGS studies are needed to increase the sample size in order determine whether this correlation is significant.

Although there is very little data on the clinical impact of DIS3 mutations, one very recent study has identified a trend towards a shorter median overall survival for patients with DIS3 mutations. Patients carrying DIS3 mutations in minor subclones of their tumours also showed a significantly worse response to therapy compared to patients with DIS3 mutations in the major subclone (Weißbach et al., 2015).

Although, minor subclones also tend to accumulate 17p deletions which could also explain this trend.

Although the cohort sizes in these genome-wide sequencing studies are large, the actual number of patients with DIS3 mutations is still quite small; therefore, determining the clinical implications of DIS3 mutations will require yet more samples to be tested.

1.4   Multiple Myeloma

1.4.1 General overview and pathophysiology

Multiple myeloma is a malignancy of antibody-producing plasma cells that accumulate in the bone marrow (Figure 1.12). The malignant plasma cells undergo a massive clonal expansion resulting in the production of high levels of monoclonal immunoglobulin, or paraprotein, in the blood (Raab et al., 2009). Symptoms of the disease result from the malignant cells accumulating in the bone marrow and stimulating bone resorption, interfering with production of normal blood cells, as well as organ damage caused by the build-up of paraprotein.

Myeloma originates when genetic damage occurs in developing B cells. B cells undergo two processes in the germinal centre, somatic hypermutation (SHM) and class switch recombination (CSR), which are both mediated by the generation of double-strand breaks. Normally, these DSBs are repaired locally to generate healthy mature B-cells. However sometimes they can be joined to other locations in the genome, resulting in chromosomal translocations and aberrant mutations which lead to the generation of malignant plasma cells.

Multiple myeloma can be seen as progressing through distinct clinical phases which represent a multi-step transformation process from normal to malignant plasma cells (Figure 1.13). Myeloma usually develops from the asymptomatic pre-malignant syndrome, monoclonal gammopathy of undetermined significance (MGUS), defined

Multiple Myeloma

Red marrow where plasma cells are made

Bone

Normal plasma cells

Antibodies

Multiple myeloma cells (abnormal plasma cells)
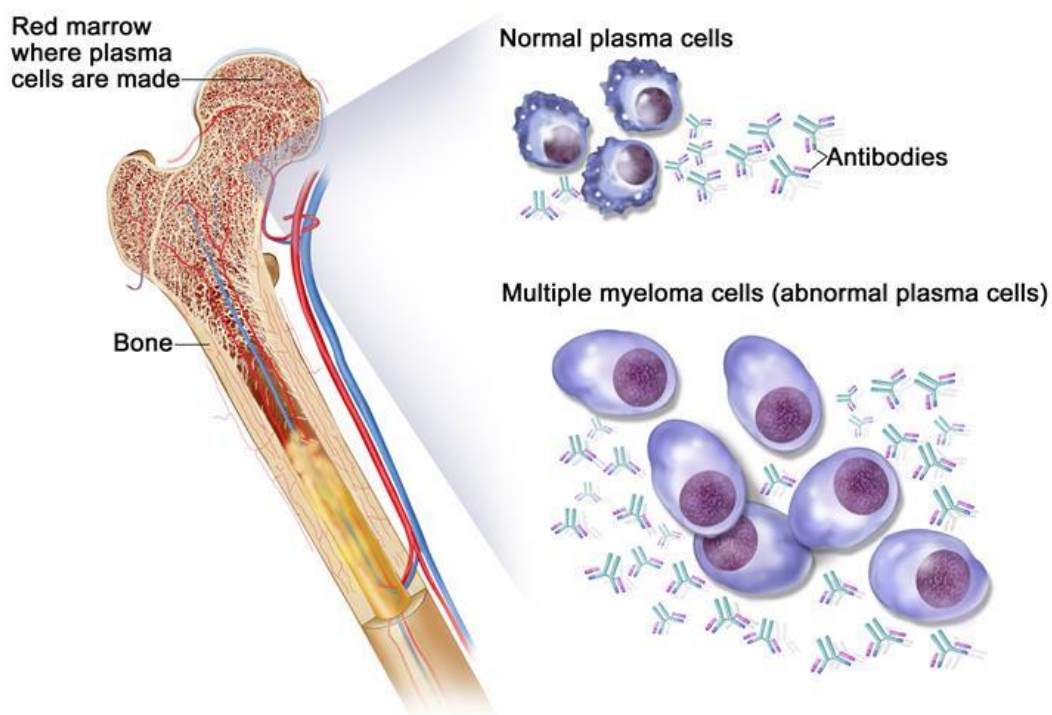
© 2014 Terese Winslow LLC
U.S. Govt. has certain rights

Figure 1.12. Multiple myeloma is a malignancy of antibody-producing plasma cells that accumulate in the bone marrow. (International Cancer Institute)
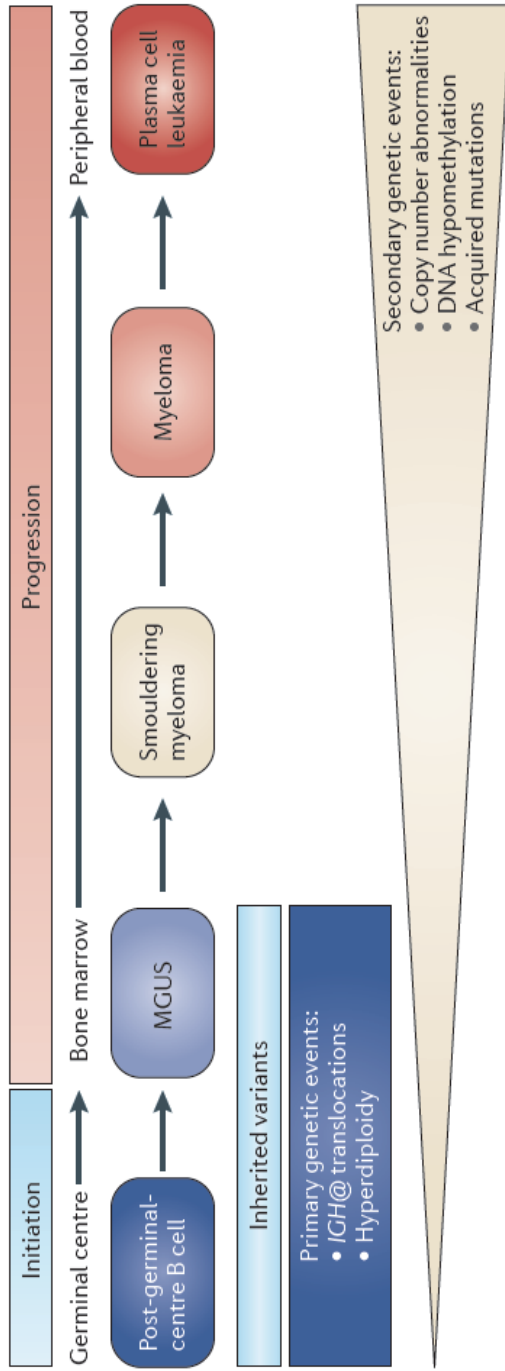
Figure 1.13. The initiation and progression of multiple myeloma is caused by distinct genetic events. Multiple myeloma progresses through distinct clinical phases from normal to malignant plasma cells due to the initial acquisition of primary genetic events and the gradual accumulation of secondary events. A post-germinal centre B cell receives a genetic "hit" which immortalises the cell and initiates transition to the indolent phase of monoclonal gammopathy of undetermined significance (MGUS). MGUS clones may then transition through the other disease phases of smouldering multiple myeloma (SMM), myeloma, and plasma cell leukaemia (PCL) as genetic "hits", which confer a survival advantage and are acquired over time. (*Morgan et al 2012*)

by a clonal expansion of plasma cells and the presence of monoclonal antibody in the blood (Morgan et al., 2012) . Although benign, approximately 1% of MGUS patients progress to myeloma per year (Kyle et al., 2002; Kyle et al., 2011). Smouldering myeloma (SMM) is a variant of MGUS which presents no symptoms but is characterised by higher levels of paraprotein and plasma cells and the annual risk of progression to malignancy is 10% to 20%. Active myeloma is diagnosed when, as well as paraprotein and clonal plasma cell expansion, there is also organ and tissue damage. The disease can eventually progress to extramedullary myeloma (EMM) whereby plasma cells develop the ability to survive outside of the bone marrow environment and can form tumours in other tissues of the body. The final stage is plasma cell leukaemia (PCL), characterised by very high levels of plasma cells in the peripheral blood and very aggressive clinical features.

Symptoms of myeloma include bone disease, renal dysfunction, anaemia and infection (Kyle and Rajkumar, 2004; Kyle and Rajkumar, 2008). Bone disease is caused by malignant plasma cells stimulating osteoclasts that cause bone resorption and inhibiting osteoblasts that form new bone, causing lytic bone lesions. Osteoclast activity can cause hypercalcaemia which in combination with high levels of paraprotein, contributes to renal dysfunction. Anaemia and increased infection are a result of myeloma cells overcrowding the bone marrow, reducing the production of red blood cells and white blood cells respectively.

In the earlier stages of the disease, myeloma cells require a specialised bone marrow microenvironment to maintain their survival. The bone marrow consists of extracellular matrix proteins as well as bone marrow stromal cells (BMSCs), immune cells, stem cells endothelial cells, osteoclasts and osteoblasts. It is well established that a bidirectional signalling loop exists between myeloma cells and bone marrow cells to increase tumour growth and survival. Initially, the interdependence between cells of the bone marrow microenvironment and the tumour cells was thought to be specifically with osteoclasts, responsible for bone resorption. Myeloma cells have a reciprocal relationship with these cells whereby they release signalling molecules which in turn stimulate the osteoclasts to release tumour-

growth factors (Mundy et al., 1974) . Now we know that myeloma cells also interact with numerous other cell types in the bone marrow to promote tumour growth and survival, for example, T-cells, dendritic cells, bone marrow stromal cells (BMSCs), endothelial cells and natural killer cells (Fowler et al., 2011). For example, myeloma cells have been shown to express DKK1 which inhibits osteoblast formation and increases bone resorption (Figure 1.14). RANKL, the ligand for receptor activator of NFkB, is also expressed by myeloma cells and is known to induce osteoclast formation (Sezer et al., 2002). Osteoclasts in turn produce high levels of the multi-functional cytokine, IL-6, which both directly promotes myeloma cell growth (Abe et al., 2004), as well as increasing the number of IL-17-secreting T cells (Korn et al., 2008). IL-17 both induces myeloma tumour growth and inhibits immune function (Prabhala et al., 2010).

As well as their effect on osteoclasts and T-cells, myeloma cells are also known to express certain adhesion molecules which mediate their attachment to extracellular matrix proteins and BMSCs, allowing them to home to the bone marrow (Teoh and Anderson, 1997). This binding also stimulates IL-6 production from BMSCs, further enhancing tumour growth. Myeloma cells can also promote bone marrow angiogenesis by the release pro-angiogenic cytokines. The function of dendritic cells and natural killer cells are also known to be impaired in patients with myeloma. One study has remarkably shown how these DCs were incapable of presenting the patient-specific tumour idiotype to autologous T cells, allowing immune evasion (Ratta et al., 2002).

These examples demonstrate the ability of myeloma cells to create a bone marrow niche which facilitates the growth and survival of tumour cells, not only by enhancing the microenvironment but also by regulating immune cells in a way that allows evasion of immune recognition. The BM niche, therefore, acquires primary interest as a pathogenic factor in MM.
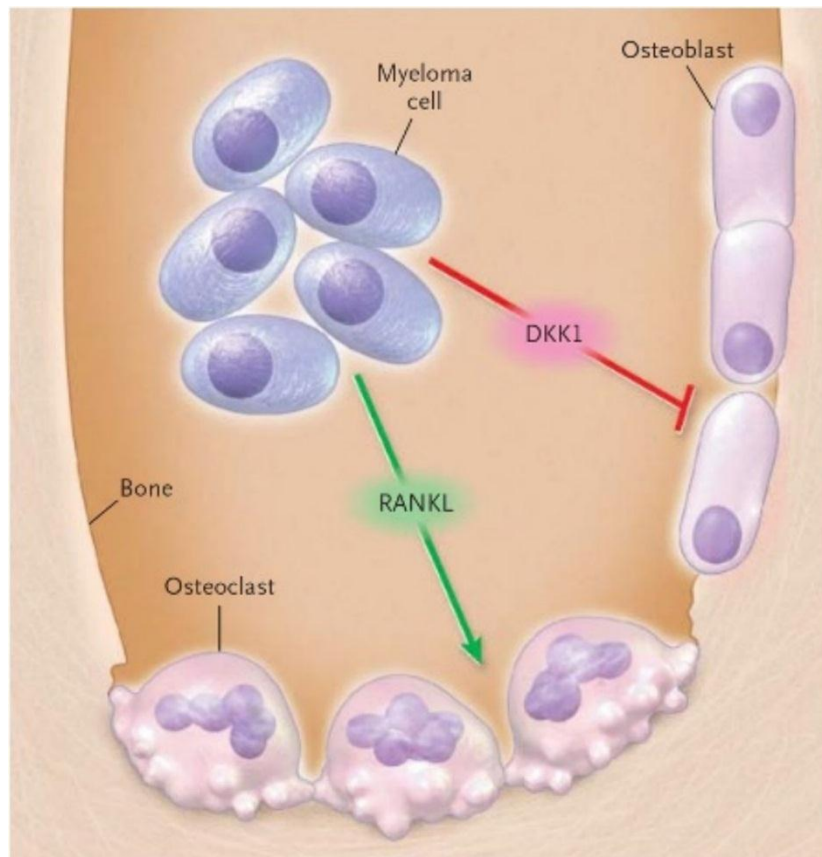
Figure 1.14. The myeloma bone marrow microenvironment. Myeloma cells release RANK ligand (RANKL), stimulating osteoclasts, and dickkopf 1 (DKK1), a protein that inhibits osteoblastic function and leads to loss of osteoblastic differentiation and destruction of osteoblast-lineage stem cells, resulting in lytic bone lesions that cannot heal. (*Glass et al, 2003*).

1.4.2 Genetics

The basic premise underlying the initiation and progression of myeloma is that multiple mutations in different pathways drive malignant change through deregulation of the intrinsic biology of the plasma cell. Genetic aberrations can be classified as primary events, contributing to plasma cell immortalisation, or secondary events, contributing to disease progression (Figure 1.13). The initial immortalisation usually occurs by the acquisition of a chromosomal abnormality which can be classified into hyperdiploid (trisomies) or non-hyperdiploid subtypes (Morgan et al., 2012). Non-hyperdiploidy involves translocations of the IgH locus with other chromosomes and is caused by aberrant class-switch recombination (CSR) in B-cells. Aberrant CSR brings oncogenes under the influence of the IGH enhancer region leading to their upregulation. The most common translocations include t(4;14) which results in the overexpression of FGFR3 and MMSET, a tyrosine kinase receptor and histone methyltansferase transcription repressor respectively, t(14;16)/t(14;20) both resulting in increased expression of the transcription factor MAF and t(11;14) which causes upregulation of CCDN1 (cyclin D1). Hyperdiploidy involves trisomies of the odd numbered chromosomes 3, 5, 7, 9, 11, 15, 19, and 21 and is thought to cause over-expression of protein biosynthesis genes, specifically those representing end points in MYC, NF-κB, and MAPK signalling pathways. A key early molecular abnormality in myeloma caused by these primary events is the deregulation of the G1/S cell cycle transition point via the overexpression of cyclin D genes.

Primary genetic events co-operate with secondary lesions such as secondary translocations, copy number variations (CNV), loss of heterozygosity (LOH), acquired mutations, and epigenetic modifications to produce the malignant phenotype of myeloma. Unlike other cancers, very few genes are recurrently mutated in myeloma but those that are include KRAS, NRAS, FAM46C, DIS3, TP53, BRAF, TRAF3, CYLD, RB1 and PRDM1 (Walker et al., 2015). These secondary lesions can be classified as generally belonging to cell signalling, epigenetic or RNA processing pathways. KRAS, NRAS and BRAF are oncogenes belonging to the MAPK

pathway that when over expressed lead to cell proliferation. The precise function of FAM46C is yet unknown, however there is evidence it belongs to a family of nucleotidyltransferases (Kuchta et al., 2009) and it was recently shown to function as an mRNA stability factor that binds to CU rich motifs within the 3'UTRs of some mRNAs (Meng, 2010). In support of this function, Chapman et al have found its expression to be highly correlated to the expression of a set of ribosomal proteins and translation initiation factors (Chapman et al., 2011). DIS3 is a ribonuclease involved in degrading mRNA for both quality control purposes and as a means of gene regulation, as well as small RNA processing. TP53 is the well-known tumour suppressor that codes for the pro-apoptotic p53 protein. TRAF3 is a TNF receptor associated factor involved in regulation of the NF-kB pathway, which when aberrantly activated, contributes to cell proliferation and survival. CYLD is a deubiquinating enzyme which also plays a role in regulating NF-kB. RB1 (Retinoblastoma 1) is a tumour suppressor gene that acts as a negative regulator of the cell cycle and PRDM1, also known as BLIMP-1 acts as a repressor of beta-interferon expression involved in the differentiation of plasma cells. With the exception of NRAS (24%) and KRAS (27%), all genes are mutated at a low percentage, indicating the deregulation of key pathways is important, rather than mutations of single genes.

The role of epigenetics in myeloma is a relatively under-studied field, however there is increasing evidence that the three types of epigenetic modification, namely DNA methylation, histone modifications and RNA interference influence pathogenesis (Dimopoulos et al., 2014). During the transformation from MGUS to myeloma, a well-recognised pattern of global DNA hypomethylation and gene-specific hypermethylation occurs (Walker et al., 2011). During the progression from myeloma to PCL, gene-specific hypermethylation is observed in genes involved in cell adhesion and signalling, suggesting an important role in promoting transition into the circulation. Histone modifications are also important as demonstrated by the HOXA9 gene which is overexpressed in myeloma (Chapman et al., 2011). HOXA9 is a transcription factor primarily regulated by HMTs. Histone methyltransferases become deregulated in myeloma through mutations in genes such as MMSET, a

histone methyltransferase repressor, whose expression correlates with that of HOXA9. In addition, miRNAs are known to be differentially expressed in myeloma. Several miRNAs have been identified as oncomirs (miRNAs that are amplified or overexpressed in cancer and were shown to have a promoting role in the development of primary tumours) or tumour suppressors (miRNAs that are deleted or reduced in cancer cells and their loss is associated with tumour development (Calin and Croce, 2006). miR-15a and miR-16-1 located at 13q14.3, a region that is commonly deleted in MM, have been shown to regulate tumour proliferation in MM (Roccaro et al., 2009).

Coupled to these secondary events, malignant cells acquire genetic abnormalities that facilitate their interaction with the bone marrow microenvironment (discussed above). Direct evidence supporting the idea that genetic events mediate stromal interactions comes from the study of cases overexpressing MAF (Morgan et al., 2012). *MAF* is overexpressed in 30% of myeloma cases and alters the homing and adhesive properties of the cell through integrin-β7 (Neri et al., 2011), as well as by the upregulation of the chemokine receptor CCR1, thus promoting the survival of plasma cells in their bone marrow niches.

Together with the characterisation of genes and pathways disrupted in myeloma, studies are also revealing a further level of genetic complexity in the form of intraclonal heterogeneity. It is becoming increasingly clear that the molecular events acquired during myeloma progression are not acquired in a linear fashion but instead through branching, nonlinear pathways that can be described as a clonal evolution of plasma cells (Figure 1.15). Each clone is defined by different molecular events and can either be described as dominant or minor depending on how much of the tumour population it makes up (Morgan et al., 2012). Clones are constantly evolving so that minor subclones can acquire driver mutations that confer a survival advantage and over time become major subclones. Intraclonal heterogeneity has therapeutic relevance as treating the dominant clone alone is likely to lead to drug resistance and relapse through the clonal evolution of minor clones.
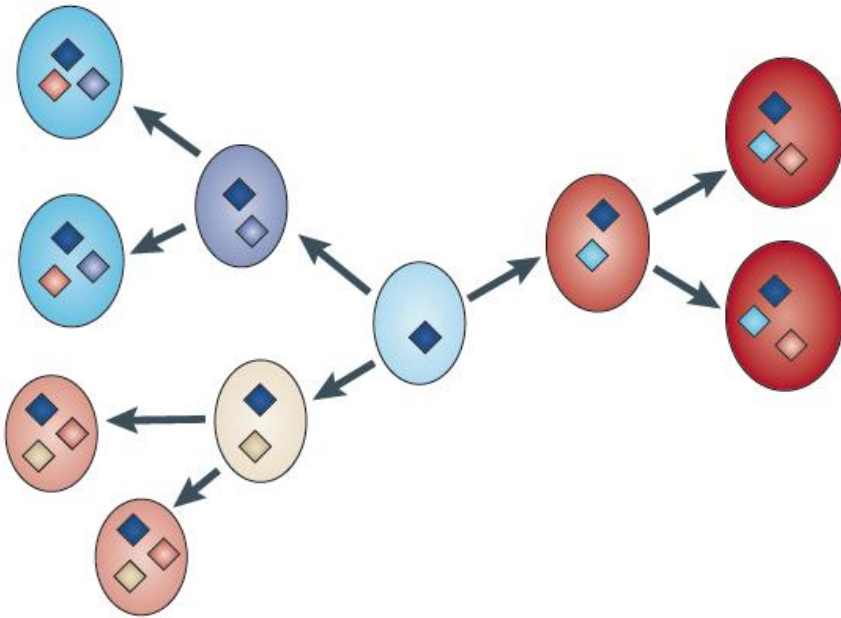
Figure 1.15. The clonal evolution of plasma cells in multiple myeloma.
Molecular events are not acquired in a linear fashion but instead through
branching pathways. The key molecular events leading to disease evolution are
represented as diamonds. Each clone is defined by a different set of molecular
events. (*Morgan et al 2012)*

As almost all cases of myeloma progress through the pre-malignant syndrome MGUS, it is important to consider factors that may lead to MGUS as well as those events promoting the transition from MGUS to myeloma. Familial studies have shown that in families with myeloma cases, there is a two-fold increased risk of developing MGUS demonstrating that inherited genetic variation can predispose to the pre-malignant condition (Altieri et al., 2006). Molecular epidemiological studies have identified three gene pairs which are associated with an increased risk of myeloma DNMT3A and DTNB (on 2p), ULK4 and TRAK1 (on 3p), and DNAH11 and CDCA7L (on 7p) (Broderick et al., 2012). The functional role of each gene pair is yet unknown but it appears that the deregulation of MYC is important.

1.4.3 Epidemiology

Multiple myeloma accounts for approximately 1% of cancers and 13% of haematological cancers (Kyle and Rajkumar, 2004). In 2012 there were 4934 new cases of myeloma in the UK and 2742 deaths (Cancer Research UK). The incidence rate is 9 per 100000 males and 7 per 1000000 females with myeloma occurring twice as frequently in African Americans compared to Caucasians, however the reasons underpinning these differences have yet to be explained. Myeloma incidence is strongly related to age with an average of 43% of cases being diagnosed in people aged 75 and over. Five-year survival stands at 47% and ten-year survival at 37% but outcome depends on the stage of myeloma at diagnosis as well as age and fitness. The decrease in survival with age may be partly attributable to fewer treatment options being available for older patients who are often unable to tolerate strong chemotherapy. Survival rates have increased over time however, which is probably due to earlier diagnosis, better detection methods as well as high-dose chemotherapy and autologous stem cell transplantation.

1.4.4 Aetiology

Due to myeloma being relatively rare, the risk factors for developing this disease are not well understood. It is estimated that in the UK less than 1% of cases are

preventable and less than 1% are linked to occupational exposures (Parkin et al., 2011). The strongest risk factor for myeloma is age where incidence and mortality increase as people get older with incidence rates rising sharply from around age 55. Another risk factor is sex with a male to female ratio of around 13:10 and ethnicity, with African Americans being twice as likely to develop myeloma (NCIN). As mentioned previously, people with a family member with myeloma or MGUS are 2 to 3 times as likely to develop the disease. People with a lowered immunity through taking medicines after organ transplantation (Engels et al., 2013) or HIV infection have an increased risk (Shiels et al., 2009). Obesity is also associated with increased risk due to a higher secretion of IL-6 (Larsson and Wolk, 2007) as well as people who have been exposed to ionising radiation, benzene or ethyl oxide (International Agency for Research on Cancer).

Some medical conditions are believed to increase the risk of myeloma for example meta-analyses have shown risk is 50% higher in patients with the autoimmune disease pernicious anaemia (McShane et al., 2014). In addition, people with a rare genetic condition called Gaucher disease which is an autosomal recessive lysosomal storage disease have a higher risk of developing myeloma (Shiran et al., 1993). People with this condition have a deficiency in an enzyme within the lysosomes which is required for the metabolism of fats, leading to the build-up of fat in certain organs. Elevated levels of IL-6 in Gaucher disease may aid the growth and survival of myeloma cells (Allen et al., 1997). In addition, fatty substances can accumulate in the bone creating skeletal lesions which change the bone microenvironment, also aiding myeloma cell survival.

1.4.5 Treatment

Traditionally, the international staging system (ISS) used clinical features such as serum albumin and β2-microglobulin to direct treatment of individual patients (Greipp et al., 2005); however, this has since been revised to include the presence or absence of cytogenetic abnormalities as well as lactate dehydrogenase level. Treatment of myeloma can be seen as stages of initial therapy, stem cell transplant

(if eligible), consolidation/maintenance therapy, and treatment of relapse (Palumbo and Anderson, 2011). Current treatment consists of combinations of chemotherapy, biological therapies and steroids as well as radiotherapy and bisphosphonates to help prevent bone damage and control pain. The drugs chosen depend on a patient's health and whether a transplant is planned as well as the risk group of disease. Certain cytogenetic abnormalities present a high risk such as t(14:16), t(14;20) and del(17p) whereas others present an intermediate or standard risk which is the case for most trisomies, t(11;14) and t(6;14).

In standard and intermediate risk patients, a combination of bortezomib, lenalidomide and dexamethasone are used, followed by autologous stem cell transplant (ASCT) and lenalidomide plus dexamethasone for standard risk and bortezomib, cyclophosphamide and dexamethasone for intermediate risk patients (Mahindra et al., 2012). Bortezomib is a non-specific proteasome inhibitor thought to induce bcl-2 phosphorylation and cleavage causing G2-M arrest and apoptosis. Dexamethasone is a glucocorticoid steroid that also induces apoptosis of myeloma cells. Lenalidomide is a biological therapy which induces tumour apoptosis through anti-angiogenic and anti-osteoclastogenic effects as well as immunomodulation. Following initial treatment, standard risk patients receive lenalidomide maintenance and intermediate risk patients received bortezomib for 2 years. High risk patients initially receive a combination of carfilozomib, lenalidomide and dexamethasone, followed by ASCT and the above combination, followed by carfilozomib or bortezomib maintenance for 2 years. If the patient is frail or over age 75 they may not tolerate a triplet regime so treatments are modified slightly and the patient may not be eligible for transplantation.

Many emerging therapies are showing promise in clinical trials including monoclonal antibodies. One such antibody called Siltuximab blocks IL-6 which may reduce myeloma cell growth in combination with bortezomib (Cancer Research UK). Another called BHQ880 works by blocking DKK1, a protein that stops osteoblasts from forming new bone. In addition, a new type of biological therapy called panobinstat is being investigated as a drug that is helpful for relapsed myeloma or

myeloma that has stopped responding to treatment. Panobinsat is a deacetylase inhibitor. Although still only experimental, vaccines are being investigated that are designed to encourage the immune system to attach myeloma cells.

Myeloma is not a single disease entity but rather a collection of related disorders that are manifested clinically as clonal proliferations of plasma cells. The eventual goal of myeloma therapy is personalised medicine, whereby biomarkers are used to define a patient subgroup that will benefit from a specific drug. The risk-stratified approaches to treatment outlined above are suboptimal to direct treatment for individual patients. Although this system has been improved by including molecular cytogenetic data such as translocations and copy number abnormalities, there is a great need to take into account molecular lesions that are predictive for response to treatment, thus achieving personalised cancer care for patients with myeloma.

1.5 Aims of this project

The principle aim of this project is to elucidate the mechanism by which DIS3 mutations are contributing to myeloma pathogenesis whilst also investigating the normal physiological role of DIS3 within human cells. The approaches employed are described in the following four results chapters and outlined briefly below.

1.5.1 Chapter 3

The aims of the experiments in this chapter were to predict the effect of myeloma-associated mutations on the stability of DIS3 and to establish a DIS3 knock-down model to investigate whether loss of DIS3 causes any phenotypic changes within cells. Techniques used include mutational bioinformatics analysis, transfection of mammalian cells, qPCR, western blot, growth and survival assays, drug sensitivity assays, immunocytochemistry and flow cytometry.

1.5.2 Chapter 4

The aim of this chapter was to perform transcriptomic profiling using RNA-sequencing to compare the transcriptomes of DIS3 knock-down and control cells in order to identify possible targets of DIS3. Techniques used include RNA-sequencing and qPCR.

1.5.3 Chapter 5

In this chapter an RNAi loss-of-function screen was used to identify targets whose knock-down acts synergistically with loss of DIS3 function. In this way, the aim was to identify new potentially therapeutic targets for myeloma patients affected by DIS3 mutations as well as potentially reveal unexpected pathway interactions that can advance our understanding of myeloma biology. Techniques used include RNAi synthetic lethality screen, transfection and qPCR.

1.5.4 Chapter 6

The aims of this chapter were to identify and characterise the protein-coding isoforms of DIS3 as well as assess their expression level in different cell lines and patient cells and investigate a possible differential function. Experiments carried out include qPCR, western blot, cloning, biochemical RNase assays, end-point PCR and DNA sequencing.

Chapter 2: Materials and Methods

2.1 Cell Culture

2.1.1 Cell Lines

A total of seven human MM cell lines were used in this work: H929, SKMM1, MM.1S, RPMI-8226, U-266, KMS-12-BM, and MOLP-8. These cells were a kind gift from Dr. E. Ocio (University of Salamanca), Dr. S. Sahota and B. Guinn (University of Southampton), and Dr. Y. Calle (King's College Hospital, London). The identity of the myeloma cell lines was previously assessed by protein electrophoresis of cell supernatants to confirm production of the expected type of monoclonal immunoglobulin (Stewart et al., 2011). The acute myeloid leukaemia cell line, OCI-AML3, was obtained from Dr. Y. Calle (King's College Hospital, London) and validated by sequencing for specific DNMT3A mutations. The GM12878 lymphoblastoid cell line, E14 (mouse embryonic stem cell), HeLa (cervical cancer) and HEK-293 (human embryonic kidney) cells were obtained from Dr.M.West, Dr. S. Wheatley and Prof. G. Richardson respectively (University of Sussex).The osteosarcoma cell lines, U-2OS and SAOS-2 were obtained from the European Collection of Cell Cultures (ECACC). See Table 2.1 for a list of cell lines used in this study.

2.1.2 Primary Cells

All patient samples were primary bone marrow aspirates with the exception of patient 238 and the healthy control samples, both of which were peripheral blood. Primary bone marrow aspirates and peripheral blood samples were taken from routine diagnostic specimens after informed consent of the patients. The project received approval from the local ethics committee (The Brighton Blood Disorder Study, references: 09/025/CHE and 09/H1107/1) and was conducted in accordance with the Declaration of Helsinki. Seven of the CMML samples were obtained from

Cell Line	Cell Type & Origin
H929	B-lymphoblast from plasmacytoma in bone marrow of 62 year old female with myeloma
SKMM1	Bone marrow of 51 year old male with plasma cell leukaemia
MM1.S	B-lymphoblast from peripheral blood of 42 year old female with myeloma
RPMI-8226	B-lymphoblast from plasmacytoma in peripheral blood of 61 year old male with myeloma
U-266	B-lymphoblast from plasmacytoma in peripheral blood of 53 year old male with myeloma
KMS-12-BM	Bone marrow of 64 year old female with myeloma
MOLP-8	Peripheral blood of 52 year old man with myeloma
HeLa	Cervical epithelial cell of 31 year old female with cervical adenocarcinoma
SAOS-2	Bone epithelial cell of 11 year old female with osteosarcoma
THP-1	Monocytes from peripheral blood of 1 year old make with acute monocytic leukaemia
GM12878	B-lymphoblast from peripheral blood of female and transformed with Epstein bar virus
U-2OS	Bone epithelial cell from 15 year old female with osteosarcoma
D-G75	B-lymphoblast from pleural effusion of 10 year-old male with Burkitt's Lymphoma
KG-1	Macrophage from bone marrow of 59 year old male with acute myelogenous leukaemia
OCI-AML3	Peripheral blood of 57 year old male with acute myeloid leukaemia
E14	Mouse embryonic stem cell
HEK-293	Epithelial cell from human embryonic kidney

Table 2.1 List of cell lines and origins used in this study.

Cambridge Blood and Stem Cell Biobank. Mononuclear cells were isolated by Histopaque 1077 density gradient purification as per manufacturer's instructions.

2.1.3 Cell Passage

Cells were cultured in Dulbecco's modified eagle's medium (HeLa, HEK-293, E14), RPMI-1640 medium (RPMI-8226, SKMM1, MM.1S, OCI-AML3, H929, GM12878, MOLP-8, KMS-12-BM, U-266, DG-75) or DMEM-F12 (U-2OS, SAOS-2) with 10% foetal calf serum (FCS), supplemented with 2 mM L-glutamine (Gibco) and antibiotics (100 IU/mL penicillin, 100 μg/ml streptomycin, Sigma-Aldrich), at 37°C in a 5% CO_2 humidified atmosphere. Adherent cell passage was performed by removing growth media, rinsing the monolayer with sterile PBS and incubating with 2 mM trypsin-EDTA for 5 minutes at 37°c. Complete growth media was added to neutralise the effect of trypsin-EDTA and detached cells were collected by centrifugation at 400 x g for 5 minutes. The supernatant was discarded before pelleted cells were re-suspended in complete growth media and plated in tissue culture flasks. Suspension cells were directly collected by centrifugation at 400 x g for 5 minutes and passaged similarly.

2.1.4 Cryopreservation of cells

Cells were pelleted by centrifugation before being re-suspended in growth media with 10% dimethyl sulphoxide (DMSO). 1 ml aliquots were transferred to cryovials and placed at -80°c in an isopropanol-containing freezing container for 24 hours before being transferred into liquid nitrogen for long term storage. When needed, frozen stocks were rapidly thawed at 37°c in a water bath and DMSO removed by centrifugation at 400 x g for 5 minutes, before being re-suspended in growth media and plated in tissue culture flasks.

2.2 Cell Transfection

2.2.1 Transient transfection of siRNA

Initially Lipofectamine 2000 and HiPerfect transfection reagents were tested according to the manufacturer's instructions. 100 µM of Smart Pool ON-TARGETplus siRNA against DIS3 or Non-Targeting Control siRNA (GE, Dharmacon) were initially used (see Figure 2.1 and Table 2.2 for details of siRNA) and gene silencing was analysed 48 hrs to 96 hrs post-transfection. The Neon™ electroporation system (Life Technologies) was subsequently used for suspension cells according to the manufacturer's instructions. Different cell lines were subjected to different transfection parameters (Table 2.3). After transfection was optimised, 25 nM each of two Silencer® Select siRNAs against DIS3 and 50 nM of a negative control siRNA (Life Technologies) were used for subsequent experiments in RPMI-8226 cells (Figure 2.1). Transfections were sometimes performed with a no-siRNA electroporated control.

2.2.2 Stable transfection of shRNA vectors

1×10^6 U-2OS cells were seeded into each of four T25 flasks in 5ml of DMEM-F12 growth media supplemented with 10% foetal calf serum (FCS) and with 2 mM L-glutamine (Gibco) but without antibiotics. Cells were left to proliferate for 2 days before being transfected using Lipofectamine LTX PLUS (Thermo Scientific). 150 µl of Opti-MEM (Gibco) and 9 µl of LTX reagent were incubated together for 5 mins at room temperature before being added to 2.5 µg of each shRNA vector (Three DIS3 shRNAs and one non-targeting control) and 3 µl of PLUS reagent. The mix was added to the cells in a dropwise fashion and after 48 hours puromycin (Thermo Scientific) was added at 3 µg/ml. Cells were left to proliferate for 3 weeks before being observed for green or red fluorescence and tested for DIS3 knock-down by qPCR.
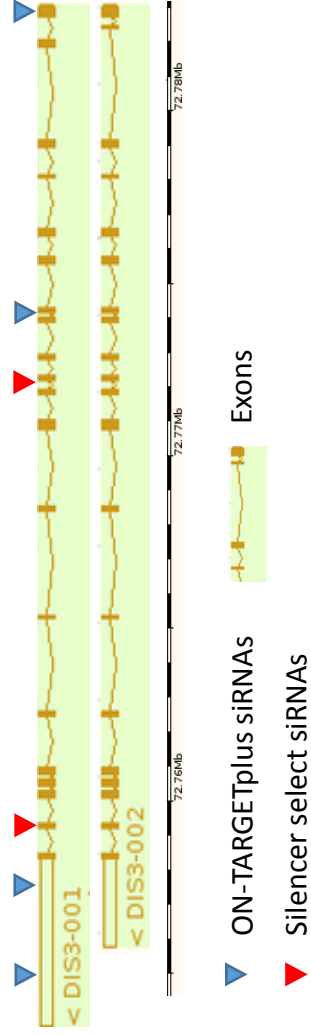
Figure 2.1. Map of the two DIS3 isoform transcripts showing the position of the two different sets of siRNAs used in this study. Transcript diagrams taken from ENSEMBL, siRNAs mapped using VectorNTI (Life Technologies).

siRNA	Sense	Anti-sense
DIS3 Smart Pool ON-TARGETplus 1	UGGAAAUGCCAGACGUUA	UAACGUCTGGCAUUUCCA
DIS3 Smart Pool ON-TARGETplus 2	GCAGAUUAAUUUUGCGUAA	UUACGCAAAAUUAAUCUGC
DIS3 Smart Pool ON-TARGETplus 3	UGAUGAAGAUCGUGCGCGA	UCGCGCACGAUCUUCAUCA
DIS3 Smart Pool ON-TARGETplus 4	AGGUAGAGUUGUAGGAAUA	UAUUCCUACAACUCUACCU
DIS3 Silencer Select 1	GGAGCAUUACUGAAAAGGAtt	UCCUUUUCAGUAAUGCUCCag
DIS3 Silencer Select 2	CAAAAGCAAAGGAAUAGUAtt	UACUAUUCCUUUGUUUUGaa
NABP1/ FLJ22833 siGenome 1	GUCAACGACCCACUUAUUU	AAAUAAGUGGGUCGUUGAC
NABP1/ FLJ22833 siGenome 2	AUGAAGUGAGAUCGUGCAA	UUGCACGAUCUCACUUCAU
NABP1/ FLJ22833 siGenome 3	GAUAUUAAGCCCGGACUGA	UCAGUCCGGGCUUAATAUC

Table 2.2. List of siRNA sequences used in this study. Lower case letters represent non-complementary 3' additions to the siRNA sequence.

Cell Line	Pulse voltage (V)	Pulse width (ms)	Pulse number	Cell density (cells per 100ul tip)	For use with plasmid or siRNA?
KG-1	1650	20	1	5.6×10^5	plasmid
KG-1	1800	15	1	5.6×10^5	plasmid
KG-1	1700	20	1	5×10^6	plasmid
OCI-AML3	1000	50	1	1.51×10^6	siRNA
U-266	1650	20	1	5.6×10^5	siRNA
MOLP-8	1650	20	1	5.6×10^5	siRNA
KMS-12-BM	1150	30	2	5.6×10^5	siRNA
RPMI-8226	1150	30	2	5.6×10^5	siRNA
H929	1150	30	2	5.6×10^5	siRNA
MM.1S	1150	30	2	5.6×10^5	siRNA

Table 2.3 Cell-type specific parameters used for transfection with the Neon™ electroporation system (Life Technologies) .

2. 3 Synthetic lethality screens

Before performing the screen, transfection efficiency was optimised. DIS3 shRNA cells (3 x10^3) as well as the non-targeting control cells were seeded into a 96 well plate in triplicate in DMEM-12 supplemented with 10% foetal calf serum (FCS), 2 mM L-glutamine (Gibco) and antibiotics (100 IU/mL penicillin, 100 μg/ml streptomycin, Sigma-Aldrich). The following day the media was removed, cells washed with PBS and 188 μl of fresh media was added. 144 μl of Opti-MEM was incubated with 9 μl of HiPerfect (Qiagen) and incubated at room temperature for 15 mins before different concentrations of either siGFP (0.625 nM, 1.25 nM, 2.5 nM or 5 nM) or a non-targeting control siRNA (Dharmacon) were added and incubated for a further 10 mins. 12 μl of the mix was added to each well and the GFP intensity was assessed after 24 hours using an Olympus Scan R system.

After achieving good knock-down efficiency with the siGFP, the DNA Damage Response (DDR) synthetic lethality screen was set up. The DDR siRNA library (Dharmacon) consists of an array of 240 siRNA pools across three 96-well plates, each pool targeting a different gene. 3.5 x10^3 of each of DIS3 shRNA 1-transfected cells and control shRNA-transfected cells were seeded together as a 50:50 ratio into three 96-well black walled, clear bottom plates (Corning) in 200 μl of DMEM-12 supplemented with 10% foetal calf serum (FCS) and with 2 mM L-glutamine (Gibco) and antibiotics (100 IU/mL penicillin, 100 μg/ml streptomycin, Sigma-Aldrich), without puromycin. The following day the growth media was changed and replaced with 188 μl of fresh media. A master mix of 3.15 ml of Opti-MEM and 202.5 μl of HiPerfect were mixed together before 12 μl was aliquoted to new 96 well plates. 1 μl of each siRNA from the DDR library was added to each well of HiPerfect mix to make a final concentration of 1.25 nM. The non-targeting, siGFP, siGAPDH and transfection reagent controls were made up next. 144 μl of Opti-MEM and 8 μl of HiPerfect were mixed together along with 1.5 μl of scrambled siRNA. For the siGFP, siGAPDH and non-transfected controls were made up by adding 72 μl of Opti-MEM, 4.5 μl of HiPerfect and 0.75 μl of siRNA before 12 μl of each control was aliquoted to the 96 well plates. 50 μl of media was then taken from the cell culture plates and

mixed with the siRNA mix before being added back to the cells. The cells were left for 72 hours before the media was removed; the cells were washed with PBS and fixed using 4% paraformaldehyde for 10 minutes. Cells were then washed and permeabilised with 0.1% Triton-X for 5 minutes before DAPI was added. Cells were washed with PBS before being screened using an Olympus Scan R system.

2.4 Growth assays

Cells were plated ($2x10^5$) in triplicate in 24-well plates with 500 µl of medium. At various time points following transfection (0, 48, 96, 144 hrs) cells were stained with trypan blue and counted on a Modified-Fuchs haemocytometer (Hawskley Medical). Cell pellets were re-suspended in complete growth media and the haemocytometer was filled with 10 µl of single cell suspension. The number of cells overlying the ruled grid was counted using low power magnification (×10) on an inverted microscope (Leica). The number of cells per millilitre was calculated (cells in 25 squares of the grid multiplied by 10^4) and the cell suspension was diluted appropriately to seed the correct number of cells for each experiment.

2.5 Cell Viability assays

To assess viability using a WST1 assay (4-[3-(4-Iodophenyl)-2-(4-nitrophenyl)-2H-5-tetrazolio]-1,3-benzene disulfonate, Roche), cells were plated ($2.5x10^3$) in triplicate in 96-well plates with 90 µl of medium. At 24 hour intervals following transfection (0 to 168 hrs), 10 µl of WST1 was added to the cultures. Cultures were incubated at 37^oC and the absorbance was recorded after 90 mins at 450 nm using a plate reader (Bio TEK) and KC4 software. The viability was plotted as percentage OD change whereby untreated cells or cells at time-point zero are set at 0% and viability in other groups was calculated by comparing the optical density readings with the controls. Similar to MTT, WST1 is a tetrazolium dye but is reduced outside of living cells on the plasma membrane, rather than inside. WST1 is water-soluble unlike MTT and therefore does not require a solubulisation step to form a coloured solution.

2.6 Caspase assays

To assess apoptosis, the Caspase-Glo 3/7 (Promega) was used according to manufacturer's instructions. Briefly, cells were plated (2.5×10^3) in triplicate in 96-well plates with 50 μl of medium. At 24 hour intervals following transfection (0 to 96 hrs), 50 μl of Caspase-Glo reagent was added to each well and incubated for 1 hr. Luminescence was measured in a luminometer. Results were plotted as the mean luminescence, relative to the luminescence at time point 0.

2.6.1 Drug sensitivity assays

RPMI-8226 cells were plated (2.5×10^3) in triplicate in 96-well plates and bortezomib (Santa-Cruz) was added to the medium at 0, 0.01, 0.1 and 1 nM concentrations. After 24 hrs, WST1 was added to the cultures and analysed as above. In the same manner, 0.1 nM bortezomib was added to transfected cells 72 hrs post-transfection and WST1 absorbance was measured over various time points (0, 24, 48 hrs).

2.7 RNA extraction, reverse transcription and quantitative real-time PCR

2×10^6 cells were lysed and total RNA was extracted using the RNAeasy mini-kit (Qiagen) as per manufacturer's instructions. The resultant total RNA was re-suspended in 30 μl of nuclease-free water. 1 μl of the sample was measured using an ND-1000 spectrophotometer (Thermo Scientific). RNA integrity was measured on a 2100 bioanalyser (Agilent). Potential contaminating DNA was removed by DNase treatment of RNA as part of the gDNA removal step of the Quantitect RT kit (Qiagen). 500 ng of RNA was reverse transcribed as per manufacturer's instructions, into a final volume of 20 μl and was diluted 1:5 in nuclease-free water. For qPCR analysis 2.5 μl of cDNA was mixed with 5 μl Precision 2 x real-time PCR MasterMix, 0.5 μl of each primer (Table 2.4 for list of all qPCR primer/probe assays) and 2 μl water and analysed in a Life Technologies ViiA™ 7 System using an annealing temperature of 50 °C. Negative controls lacking reverse transcriptase showed

Oligonucleotide	Company	Accession number	Sequence (5' to 3')	Anchor nucleotide	Context length sequence	Amplicon length
DIS3 (Forward/Reverse)	PrimerDesign	NM_014953	ACCCCCGCCAACCATTTAA / CGCCTCGTCCTCCCAAAA	N/A	N/A	84
DIS31 (Forward/Reverse)	Life Technologies	NM_014953.4	N/A	N/A	N/A	114
DIS32 (Forward/Reverse)	Life Technologies	NM_001128226.2	N/A	N/A	N/A	84
GAPDH (Forward/Reverse)	PrimerDesign	NM_002046	N/A	1087	142	110
RPL13A (Forward/Reverse)	PrimerDesign	NM_012423	N/A	727	223	153
EIF4A2 (Forward/Reverse)	PrimerDesign	NM_001967	N/A	900	146	113
B2M (Forward/Reverse)	PrimerDesign	NM_004048	N/A	362	141	114
ACTB (Forward/Reverse)	PrimerDesign	NM_001101	N/A	1195	106	92
CYC1 (Forward/Reverse)	PrimerDesign	NM_001916	N/A	929	207	145
TOP1 (Forward/Reverse)	PrimerDesign	NM_003286	N/A	2361	195	141

Table 2.4. List of qPCR primer/probes used within this study. N/A = no sequence information available. Anchor nucleotide = a nucleotide contained anywhere within the probe sequence. Accession numbers based on the most updated version. Information presented in according to MIQE guidelines.

negligible background. Analyses were performed in duplicate or triplicate. All data were normalised to the appropriate reference genes as determined by the GeNorm assay (Primer design) (see below). Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

The amplification efficiencies of the DIS3 isoform-specific primers (DIS31 and DIS32) were determined with a series of cDNA dilutions (6.25, 12.5, 25, 50ng per 10 µl qPCR reaction, based on RNA concentrations) using a standard curve set up.

2.7.1 GeNorm Analysis

GeNorm reference gene selection assay (PrimerDesign) was performed for HeLa, RPMI-8226, OCI-AML3, KMS-12-BM and MOLP-8 cells. Six samples for each cell line and 12 reference genes were tested as per manufacturer's instructions and a post-PCR melt curve was performed. Qbase + software (Biogazelle) was used to determine the optimal number of reference targets and to rank the reference genes in order of stability of expression.

2.8 Immunocytochemistry

Adherent cells were grown on sterile coverslips in 24-well plates and suspension cells were dropped (1×10^5 cells/µl) onto poly-lysine coated microscope slides. Cells were fixed with 4% paraformaldehyde for 20 mins at room temperature, permeabilised using 0.1% Triton x100 and incubated overnight at 4 ^{o}C with rabbit polyclonal anti-DIS3 (1:100; Sigma), rabbit anti-53BP1 (1:100; Bethyl labs) or anti-CD14 (Biolegend) diluted in antibody diluting solution (0.1 M lysine, 10% FCS, sodium azide in PBS). Cells were then incubated with anti-rabbit Alexa Fluor 488 (1:150) or anti-mouse Alexa Flour 546 (1:250), counterstained with DAPI (4',6-diamidino-2-phenylindole) (1 µg/ml), mounted with Citifluor anti fade mounting medium (agar scientific) and observed by fluorescence microscopy using a Leica DM 5000B microscope fitted with a Leica DPC300FX digital camera.

2.8.1 Nuclear phenotype scoring

To score for nuclear phenotype, cells were fixed and permeabilised as above then stained with DAPI, mounted and observed. For each treatment, 800 cells were scored for normal, poly-lobed or multiple nuclei.

2.9 Metaphase spreads

72 hours post-transfection, 5×10^6 of DIS3 knock-down and non-targeting control HeLa cells were seeded into a T75 flask and incubated with DMEM media. After 24 hours, 1 µg/ml of colcemid (Sigma-Aldrich) was added to the medium and the cells were incubated for 2 hours. Cells were next trypsinised before being centrifuged at 200 x g for 5 mins. Supernatant was discarded and cells were re-suspended in remaining supernatant before 5ml of 0.56% KCl was added slowly from the side of the tube. Cells were incubated for 15 mins before being centrifuged at 200 x g for 10 mins. The supernatant was aspirated leaving 1ml and the cell pellet was resuspended. 1 ml of methanol: glacial acetic acid (3:1) was added. This process was repeated three more times before the pellet was resuspended in a small volume of fixative. The cells were then dropped from 10 cm height onto a microscope slide, were air dried before being stained with 5% Giemsa dye.

2.10 Western blotting

5×10^6 cells were lysed in 1 ml of protein extraction buffer supplemented with protease inhibitor cocktail (Roche) and 5% 2-mercaptoethanol. Using a cell scraper and micropipette, the lysates were transferred to a 1.5 ml Eppendorf tube, vortexed then centrifuged at 13,000 x g for 10 mins. Protein concentrations were determined using DC Protein Assay (BioRad) according to the manufacturers protocol. One volume of Laemmli loading dye (Biorad) was added to each protein sample before 50 µg was run on each lane of a Criterion XT Tris-Acetate 7% resolving gel (Biorad) with 1 x SDS running buffer (Fisher Scientific). Proteins were transferred to a PVDF Hybond-P membrane (Amersham) using a wet transfer

blotting system (Biorad), run at 150 V for 1 hour. The membrane was then blocked with 5% milk in 1 x TBS-T (Tris buffered saline (TBS) solution with the detergent Tween® 20) for 1 hour before being incubated with either anti-DIS3 (1:600, Sigma) or anti-actin (1:20000, MP Biomedicals,) overnight rolling at 4°C. The membrane was then washed with 1 x TBS-T buffer for 3 x 10 mins before being incubated with anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (1:80000) for 1 hour at room temperature. After washing again with 1 x TBS-T buffer for 3 x 10 mins, the membrane was developed using Luminata Forte HRP substrate (Merck Millipore) for 5 mins before being incubated with RX X-ray film (Fuji) and developed. Where necessary, DIS3 protein was quantified relative to controls using the Image J software.

2.11 Polymerase chain Reaction and Sequencing

Genomic DNA was extracted using the DNeasy kit (Qiagen) according to manufacturer's instructions. Conventional PCR was used to investigate the presence of alternatively spliced isoforms of DIS3 as well as for amplification of DNA for sequencing. PCR reactions consisted of 1 μl gDNA or cDNA template, 1 μl of each primer (forward and reverse, 0.5 μg/μl), 12.5 μl Amplitaq Gold PCR Master Mix (Thermo Scientific) and 10 μl DNase-free water. PCR reaction conditions were as follows: initial denaturation step of 95°C for 5 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 60 seconds. This was followed by a final extension step of 72°C for 7 minutes and cooling to 4°C for storage. Loading Dye was added to the PCR products and 10 μl 1kb or 100bp DNA ladder (New England Biolabs) before loading on to a 1% agarose gel (Sigma-Aldrich) made in Tris-Borate-EDTA (TBE) buffer (Sigma-Aldrich) containing 1 μl Gel Red. See Table 2.5 for a list of oligonucleotides used in this study.

PCR products were purified using the DNA purification kit (Qiagen) according to manufacturer's instructions, before being sent to Eurofins MWG for sequencing. 15 μl of 15 ng/μl DNA was sent with 2 μl of either forward or reverse primer.

Primer Name	Forward	Reverse
DIS3	CGACACTAATGTGTTACTGCACC	GAGCCTAACTGCTAACCCCG
SRSF2	GTGGACAACCTGACCTACCG	CGACCGAGATCGAGAACGAG
GAPDH	AGGGGTCTACATGGCAACTG	CGACCACTTTGTCAAGCTCA

Table 2.5 List of PCR primers used in this study.

2.12 Vector Cloning

Two vectors were used in this study: pGIPZ (Thermo Scientific, Figure 2.2A) containing the DIS3 shRNA clones and pReceiver-B03 (Genecopoeia, Figure 2.2B) containing the DIS3 isoform clones.

2.12.1 Plasmid preparation

Each plasmid was added to 50 µl of Dh5α competent cells and incubated on ice for 1 hour. The cells were subjected to heat shock for 45 sec at 42°C, and then placed immediately on ice. 200 µl of LB medium was added to the cells and incubated at 37°C, agitating at 250 rpm for 30 minutes. Agar plates supplemented with ampicillin were pre-warmed at 37°C, and the cells were spread evenly onto the plate and incubated at 37°C overnight. Sterile pipette tips were used to pick up five colonies from the plate which were inoculated in a universal tube with 5 ml of LB medium containing 100 µg/ml ampicillin. These were incubated at 37°C with constant shaking overnight. The bacteria were then harvested by transferring to a 15 ml falcon tube and spinning at 3,500 rpm at 4°C for 10 minutes. The plasmid was extracted using QIAprep Miniprep as per manufacturer's instructions (Qiagen).

For large scale plasmid purification, inoculated DH5 α were poured into a conical flask with 250 ml of LB media containing ampicillin were left to shake overnight at 37 ^{o}C. Cells were harvested by centrifuging at 3500 rpm at 4 ^{o}C for 30 mins and plasmids isolated using Midi/ Maxiprep (Qiagen) as per manufacturer's instructions. Concentration of plasmid was measured by nanodrop spectrophotometer and stored at -20 ^{o}C.

2.12.2 Restriction Digests

Digests were performed by mixing 1 µg of plasmid with 0.5 µl of each of the restriction endonucleases (New England Biolabs), 2 µl of the appropriate buffer,
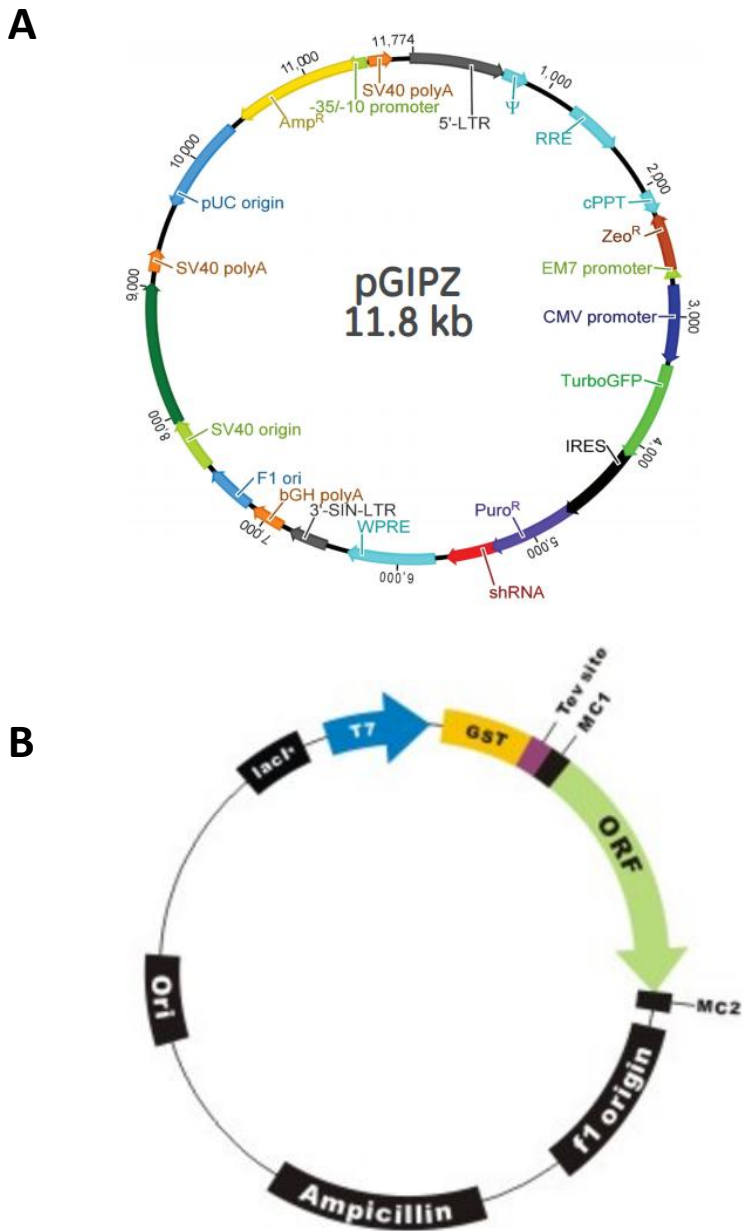
Figure 2.2. Maps of the vectors used in this study. (A) The pGIPZ vector containing DIS3-targeting shRNAs was used to create stable DIS3 knock-down cell lines. Important elements include the hCMV promoter to drive transgene expression, tGFP reporter for visual tracking of shRNA expression, 5' and 3' LTRs for genome integration and PuroR to permit selection and propagation of stable integrants. The shRNA hairpin consists of a 22nt dsRNA stem, complementary to the DIS3 transcript and a 19nt loop from miR-30 as well as 125nt of miR-30 flanking sequence on either side of the hairpin, in order to increase Drosha and Dicer processing efficiency. (B) The pReceiver-B03 vector containing recombinant DIS3 isoform cDNA clones was used for expression and purification of DIS3 isoforms for use in RNase assays. The vector contains a T7 promoter for high level expression, GST-tag and Tev cleavage site, ampicillin resistance gene for bacterial selection and multiple cloning sites.

0.2 µl of BSA (New England Biolabs) and made up to 20 µl with water. The mixture was incubated at 37 ^{o}C for 2 hours and run on a 1% TBE, 1% agarose gel. For the diagnostic digest of pReceiver-B03, XmaI and EcoRV were used. For excision of the CMV:TurboGFP region of pGIPZ, XhoI and NotI were used.

2.12.3 pGIPZ cloning

XhoI and NotI were used to excise the CMV:AcGFP-NLS or CMV:mCherry-NLS region from the relevant vectors. The same endonucleases were used to excise the CMV:TurboGFP region of pGIPZ. 3 µl of AcGFP or mCherry was mixed with 1 µl of the linearised pGIPZ vector, 5 µl of 2 x rapid ligation buffer and 1 µl of T4 DNA ligase (Promega). The ligation reaction was incubated at room temperature for 3 hrs before being propagated as above (plasmid preparation).

2.13 Biochemical assays

2.13.1 Overexpression and purification of DIS3 isoforms

Dis3 protein isoforms one (ISO1) and two (ISO2) were overproduced with a Glutathione S-transferase (GST) tag in *E. coli* Rosetta (DE3) strains containing the recombinant plasmids of interest (pReceiver-B03) and purified by affinity chromatography. Briefly, cells were grown at 37ºC in Terrific Broth medium (TB) supplemented with ampicillin and chloramphenicol to an optical density (600 nm) near 1. Protein expression was induced by addition of 0.5 mM of IPTG for 4 hrs at 30°C and cells harvested by centrifugation at 4°C. The culture pellets were resuspended in 30 mls of Buffer A (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM DTT). Suspensions were lysed using a French Press at 1000 Psi in the presence of protease inhibitors. After lysis, the crude extracts were treated with 125 U (0.5 µl) of Benzonase (Sigma) to degrade the nucleic acids and clarified by a 30 min centrifugation at 10 000 x g, 4°C. The GST-tagged recombinant proteins were purified by affinity chromatography, using the ÄKTA primer plus FPLC System (GE Healthcare). The purified extracts were loaded into a GST-Trap 1 ml column (GE

Healthcare) equilibrated in Buffer A. Protein elution was achieved with Buffer A with 20 mM reduced glutathione. The fractions containing the protein of interest were pooled and desalted using a PD10 column equilibrated with desalting buffer (10 mM Tris–HCl pH8, 10 mM KCl, 1 mM DTT). Half of the protein sample was digested with TEV protease overnight at 4°C to remove the GST-tag. An equal volume of glycerol was added, proteins were quantified using the Bradford Method and stored at −20°C. The purity of the enzymes was analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE).

2.13.2 Substrate preparation

ss16-A$_{14}$ oligoribonucleotides were labelled at the 3' end with [γ-32P]-ATP (GE Healthcare) using T4 polynucleotide kinase (NEB) according to the manufacturer's instructions. Labelled substrates were purified using a G-50 column (GE Healthcare). Circularised substrate was prepared by incubating 750 pmol ss16-A$_{14}$ oligoribonucleotide with 10 mM of unlabelled ATP in a buffer containing 50 mM Tris-HCL pH 8.0, 10 mM MgCl$_2$, 10 mM DTT in the presence of T4 ligase overnight on ice.

2.13.3 Ribonuclease assays

In vitro enzymatic assays were performed in 20 µl reaction volumes containing 10 mM Tris-HCL pH 8.0, 75 mM NaCl and 1 mM DTT supplemented with 3 mM MnCl$_2$. 125 nM protein and 50 nM substrate were used. Reactions were performed at 37°C for the indicated times and terminated by adding 20 µl of formamide loading dye (90% formamide, 20 mM EDTA, 0.03% bromophenol blue, 0.03% xylene cyanol in 1 x TBE). Reaction products were resolved on denaturing 20% polyacrylamide, 8 M urea, 1 xTBE gels and visualised using an Imagem FUJI imager.

2.14 Flow cytometry

RPMI-8226 cells were plated (2x10^5) in triplicate in 24-well plates in 500 µl of medium and subject to either starvation by omission of FCS and L-glutamine from

medium or the addition of 1 µg/ml of colchicine (Sigma). After 24 hrs cells were fixed in 70% ice-cold ethanol for 30 mins at 4°C. Cells were washed twice in PBS and centrifuged at 850 x g for 10 mins before being incubated with 0.6 mg/ml RNase A (Qiagen) for 1hr and then 10 µg/ml propidium iodide for 30 mins at room-temperature. The cells were then analysed on a flow cytometer (LSRII, Becton & Dickinson) using excitation at 488 nm and emission at 617 nm. Control electroporated, scrambled siRNA and DIS3 siRNA transfected cells were subject to the same method without starvation or colchicine treatment, 72-hours post-transfection. A no RNase A control was also set-up to confirm RNase activity.

2.15 RNA-seq sample preparation and analysis

5×10^5 RPMI-8226 cells were either transfected with DIS3 siRNA, scrambled siRNA or electroporated only and after 72 hours, RNA was extracted using the RNeasy kit (Qiagen) and treated with the RNase-free DNase kit (Qiagen). RNA purity and integrity was checked on a Nanodrop 1000 and Bioanalyser 2100 (Agilent) respectively. A total of 3 µg of RNA per sample was sent to Oxford Gene Technology for conversion to cDNA using oligo-dt primers and sequencing on a HiSeq2000 lane using TruSeq v3 chemistry (Illumina) generating an average of 28 million reads per sample. Initial quality control of samples was performed using FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by adapter removal using Scythe v0.993b (https://github.com/vsbuffalo/scythe) and quality trimming using Sickle v1.29 (https://github.com/najoshi/sickle). Reads were aligned to version CRChr37 of the human genome using TopHat v2.0.12 (Kim et al., 2013)and Bowtie v2.2.3 (Langmead and Salzberg, 2012). FPKM values and differential expression comparisons were performed using Cufflinks (Trapnell et al., 2012).

2.14 Statistical analysis

Unless otherwise stated, error bars represent the S.E.M (standard error) obtained from three independent experiments. All statistics were carried out using GraphPad

PRISM software (V.6.01) and *, **, ***, represent statistical significance at the levels of $p < 0.05$, $p < 0.01$, and $p < 0.001$ respectively.

Chapter 3: Generating a DIS3 knock-down model to investigate a functional phenotype

3.1 Introduction

In depth sequencing studies have identified DIS3 as a novel tumour suppressor gene in multiple myeloma (Walker et al., 2012; Walker et al., 2015; Chapman et al., 2011; Lohr et al., 2014; Weißbach et al., 2015; Leich et al., 2013; Lionetti et al., 2015). Mutations in DIS3 occur in ~11% of myeloma patients but the mechanism by which DIS3 mutations contribute to pathogenesis is so far unknown. Myeloma is a very genetically heterogenous disease, both between patients and within individual cases, however it is currently treated as a single entity with the result that individual therapeutic success is varied. As the characterisation of the PML-RARA fusion showed, a lot of progress can be made by the identification of a single molecular event regarding disease definition. Therefore, identifying the role of DIS3 may help to develop new targeted molecular therapies for affected patients.

As discussed in section 1, although the molecular functions of DIS3 surrounding its role in RNA metabolism are well known, fewer studies exist that investigate the biological functions of DIS3 and whether it is involved in regulating specific cellular processes. Additionally, the majority of these studies take place in yeast and *Drosophila*, which whilst yielding useful insight, are no replacement for studies in human cells. Nevertheless, as this protein is strikingly conserved across eukaryotes, findings from lower organisms may yield useful precedents through which to investigate the function in human cells.

There are a number of studies which present evidence for a role of DIS3 in regulation of the cell-cycle. Dis3 was first discovered in a mutant fission yeast strain to cause non-separation of sister chromatids during anaphase (Ohkura et al., 1988) and subsequently *S.pombe* mutants have been shown to have elongated metaphase spindles and a block in metaphase to anaphase transition (Murakami et

al., 2007). Perturbation of Dis3 *in S. cerevisiae* affects microtubule localisation and structure (Smith et al., 2011) and RNA-seq has shown broad changes in the levels of cell cycle and microtubule related transcripts in mutant strains (Kiss and Andrulis, 2010). These findings may be attributable to a link between DIS3 and heterochromatin silencing as deletion of Dis3 considerably increases levels of transcripts from silent centromeric and telomeric loci (Vasiljeva et al., 2008).

The exosome has also been implicated in recruiting activation-induced cytidine deaminase (AID) to chromatin in mammalian B-cells during immunoglobulin gene diversification (Basu et al., 2011). AID is only functional on ssDNA and so targets the non-template strand during transcription. However, the mechanism by which it accesses the template which is hybridised to nascent RNA transcript is unknown. The exosome along with one of its catalytic subunits must remove the RNA to expose the ssDNA template. However, whether DIS3 is the ribonuclease responsible for degradation of the nascent transcript is unknown.

The first step in attempting to elucidate the role of DIS3 mutations in myeloma pathogenesis is to use a reverse genetics approach to study its function within the cell. When first embarking on this project, functional studies of DIS3 in human or even mammalian cells did not exist, with only a few studies having been published since (Segalla et al., 2015; Tomecki et al., 2014; de Groen et al., 2014). In addition, there was limited evidence to show whether the mutations that had been identified in patients with multiple myeloma were causing activation or loss of function of the protein, or indeed either. Therefore, the central focus of this chapter is to predict the effect of myeloma-associated mutations on the activity of DIS3 to determine whether it is either a tumour-suppressor or oncogene and to subsequently establish a DIS3 knock-down model to investigate whether loss of DIS3 causes any phenotypic changes within cells, which may yield insight into the role of DIS3 in disease.

3.2 Aims

The aims of this chapter were to:

1. Use bioinformatics to predict the effect of myeloma-associated mutations on DIS3 protein stability
2. Generate a physiologically relevant DIS3 knock-down model using siRNA oligonucleotides
3. Investigate the effect siRNA-mediated DIS3 knock-down may have on cell phenotype and anti-myeloma drug sensitivity
4. Generate a stable knock-down model of DIS3 using shRNA-plasmids
5. Investigate the effect of stable DIS3 knock-down on cell phenotypes

3.3 Bioinformatic analysis of myeloma-associated mutations in DIS3

In collaboration with Dr Antony Oliver, an analysis of the original seven myeloma-associated mutations was performed. Through a combination of techniques, including manual inspection of structures, homology modelling according to the recently published structure of *S.cerevisiae* Rrp44 (Makino et al., 2013), using Phyre2 (Kelley and Sternberg, 2009) and the webserver 'Site Directed Mutator', we were able to analyse and make predictions about the potential consequences of each DIS3 mutation on the stability of the protein. Of the seven mutations analysed, five were predicted to have either a 'destabilising' or 'highly destabilising' effect on the domain fold which is consistent with the idea that DIS3 mutations are functionally inactivating (Table 3.1). The remaining two that are not predicted to perturb DIS3 stability may nevertheless impair specific functions of the enzyme. For example, the G766R mutation is predicted to be neutral in terms of protein fold but forms part of a turn motif and sits at the bottom of the substrate binding channel. This mutation is therefore likely to have a major effect on catalysis. A potentially important function of glycine in this particular position is further underscored by the fact that the G833D mutation is known to suppress degradation of hypomodified tRNAMet in *S. cerevisiae* (Schneider et al., 2007; Kadaba et al., 2004).

DIS3 mutation			Structure/Model	SDM analysis		Comments
COSMIC ID	Description of mutation	Domain affected	PDB code	Pseudo ΔΔG	Predicted effect of mutation	
329308	Y531D	RNB	4IFD	-3.07	Highly destabilising	In the active site and probably contributes to the binding/positioning of an incoming RNA substrate
329310	T374P	CSD2	4IFD	-1.57	Destabilising	Only slightly destabilising, does not appear to be in a protein-protein interface
329311	R780K	RNB	4IFD	-2.77	Highly destabilising	Directly involved in binding to the phosphate-backbone of incoming ssRNA. Overall charge of the residue is maintained, however it is close in proximity to a loop that gets re-modelled upon substrate binding and thus could affect catalysis
N/A	S477R	RNB	4IFD	0.14	Neutral	Mutation from a small to very large amino acid with gain of a positive charge. Next to a loop that contains residues important for activity and therefore may affect catalysis
N/A	V504G	RNB	4IFD	-5.09	Highly destabilising	Alteration to glycine will be disruptive as forms part of the hydrophobic core of the protein
N/A	G766R	RNB	4IFD	1.76	Neutral	Forms part of a turn motif and sits at the bottom of the substrate binding channel, therefore will have a major effect on catalysis
N/A	D488N	RNB	4IFD	-1.76	Destabilising	Sits at bottom of the substrate binding channel and potentially makes contact with the RNA, however the mutation is very conservative so unlikely to have a major effect

Table 3.1. Analysis of myeloma-linked mutations found in DIS3. Analysis is based on structures deposited in the protein data bank (PDB), amino acid sequence threading using Phyre2, and the use of webserver 'Site Directed Mutator' (SDM) as well as manual inspection of structures by Antony Oliver. Only 7 of the 24 currently known myeloma-linked mutations have been analysed.

Likewise, although predicted to have a neutral effect, S477R is a major mutation, changing a small amino acid to very large one, with gain of a positive charge. It is also next to a loop that contains residues important for RNB activity so may affect catalysis. Our analysis coupled with the high frequency of DIS3 mutations and loss-of heterozygosity in patients, provides good evidence for them being functionally inactivating and therefore important drivers of pathogenesis. Furthermore, three of the mutations have been created in Flp-In T-REx HEK-293 cells where they result in loss of enzymatic activity (Drazkowska et al., 2013); reinforcing the idea that DIS3 is a tumour suppressor gene.

3.4 Generating a physiologically relevant knock-down model of DIS3

3.4.1 DIS3 is ubiquitously expressed across different cell types

As DIS3 mutations had been observed in myeloma and to a lesser extent in AML patients, a disease-relevant cell line is preferred for use in generating a knock-down to recapitulate the disease scenario. However, as no information was known on the expression levels of DIS3 in different human tissues, this first needed to be established. In order to determine the expression of DIS3 in cell lines derived from various tissues, Western blot analysis was performed on a number of myeloma, acute myeloid leukaemia (AML) and adherent cell lines. Figure 3.1 shows the DIS3 protein is expressed at a similar level across these cell lines, confirming it is a ubiquitous protein with a non-tissue specific function.

3.4.2 DIS3 can only be knocked-down in adherent cell lines using lipid-based transfection reagents

For my investigation to be physiologically relevant, silencing would ideally take place in a myeloma cell line; however, suspension cells are notoriously difficult to transfect. Various lipid-based transfection reagents were initially tested to insert a GFP-plasmid into a number myeloma and AML cell lines but this method was unsuccessful (data not shown). siRNAs are often more easily transfected into cells
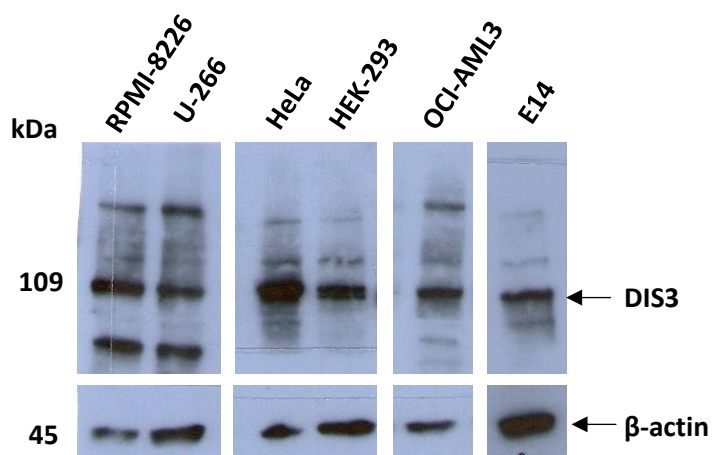
Figure 3.1 DIS3 is ubiquitously expressed across cell lines.
Total protein was extracted and an equal amount (50µg) was
loaded per lane. β-actin acts a positive control. RPMI-8226 and
U-266 are myeloma cell lines; HeLa is cervical cancer, HEK-293
human embryonic kidney, OCI-AML3 is an AML cell line and
E14 is a mouse embryonic stem cell line.

than plasmids due to their smaller size, therefore the transfection was repeated with 50nM of a set of 4 pooled siRNAs (Smart Pool ON-TARGETplus siRNA, Thermo Scientific). These siRNAs were designed to target both splice variants of DIS3 and as a control a non-targeting scrambled siRNA was used. Nevertheless, this proved unsuccessful in these cell lines; qPCR showed no difference in the relative levels of DIS3 between the scrambled and DIS3 knock-down cells using the HiPerfect transfection reagent (data not shown).

As adherent cells are well-documented to be more easily transfectable, the next step was to test the siRNAs in HeLa cells. Use of the HiPerfect reagent resulted in an efficient level of silencing of the mRNA and protein at 48 hrs and 72 hrs post-transfection respectively (Figure 3.2, paired t-test, p-value=0.0135). For the purpose of downstream applications, it was necessary to test the optimal time post-transfection in which to analyse gene silencing by quantifying the level of DIS3 at various time points. This was performed in HeLa cells at 24-hour intervals and showed the greatest level of mRNA knock-down to occur after 48-hrs (Figure 3.3).

3.4.3 DIS3 can be knocked-down in suspension cell lines using siRNA electroporation

Although a good knock-down had been achieved in the HeLa cell line, this cell line is not physiologically normal and has little relevance to myeloma. Therefore, I explored an electroporation method of transfecting suspension cells. Electroporation uses an electric field in order to increase the permeability of the cell membrane and is often used for difficult-to-transfect cells. Three sets of parameters with a different voltage, pulse width and number were initially tested by transfecting an EGFP plasmid into the AML line, KG-1. Cells showed a heterogenous level of GFP expression; therefore 100 cells were scored as having either a high, mid or low GFP intensity. Based on these results a 70% transfection efficiency was achieved with parameter set 1 (Figure 3.4, see Table 2.3 for parameter details).

A

B
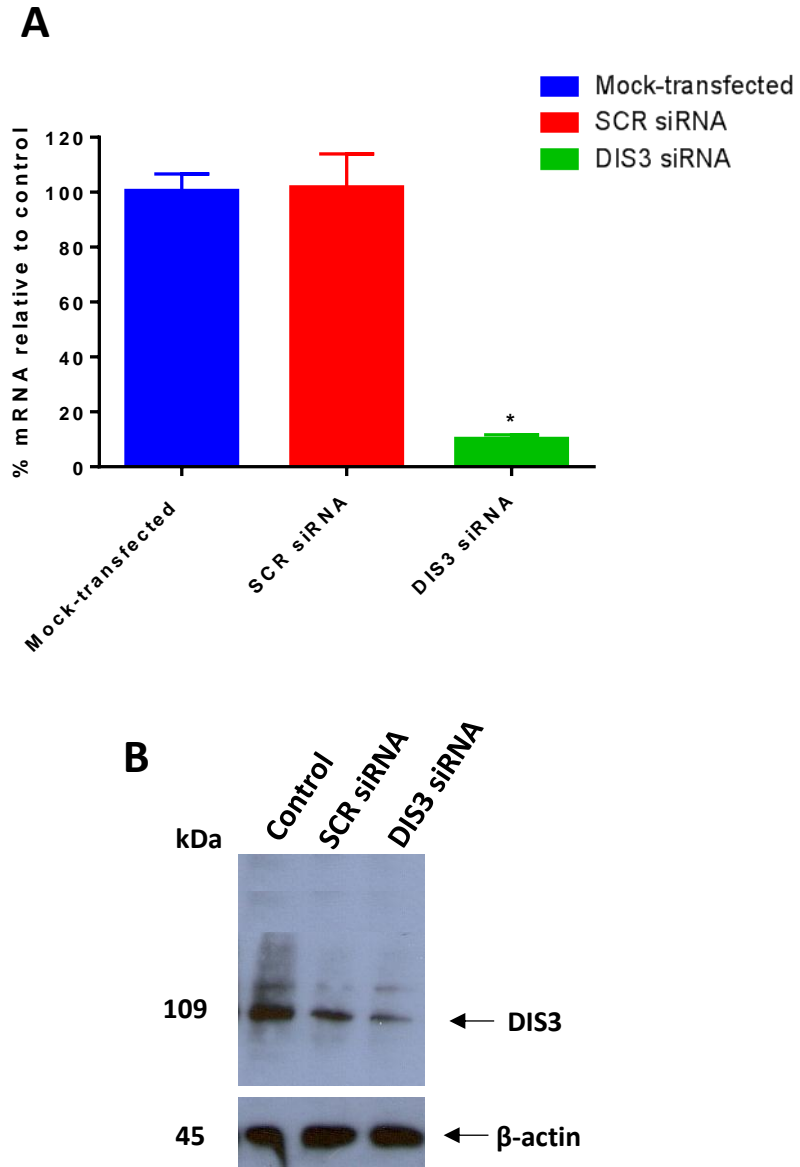
Figure 3.2. siRNA knock-down of DIS3 in HeLa cells transfected using the HiPerfect reagent . (A) Quantitative PCR showing DIS3 mRNA knock-down, (paired t-test, p-value=0.02) performed with suitable reference genes according to GeNorm analysis (GAPDH and B2M). mRNA levels are expressed as a % of the scrambled control. (B) Western blot showing DIS3 protein knock-down. Total protein was extracted and an equal amount (50µg) was loaded per lane, confirmed with an actin control. Control = vehicle control; SCR = scrambled siRNA. Error bars represent the SEM obtained from three technical experiments.
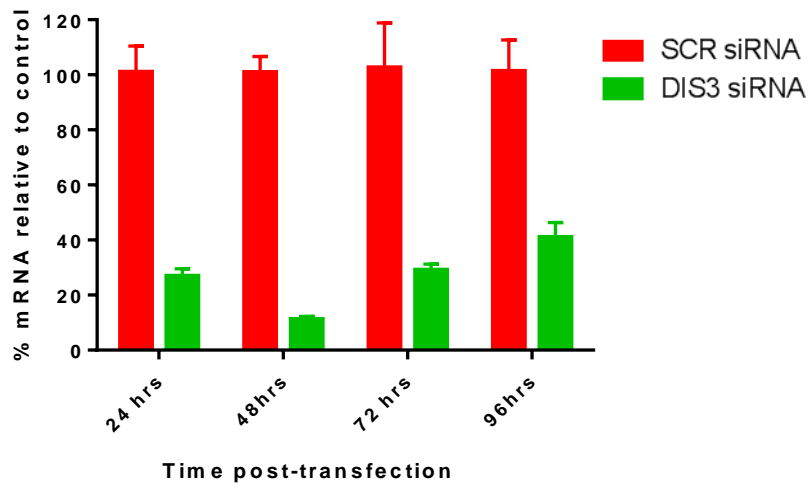
Figure 3.3. A time course experiment in HeLa cells shows the optimal time to analyse gene silencing of mRNA is 48hrs post-transfection. mRNA levels are expressed as a % of the scrambled control. Error bars represent the SEM obtained from three (24, 72, 96hrs) or six (48hrs) technical replicates.
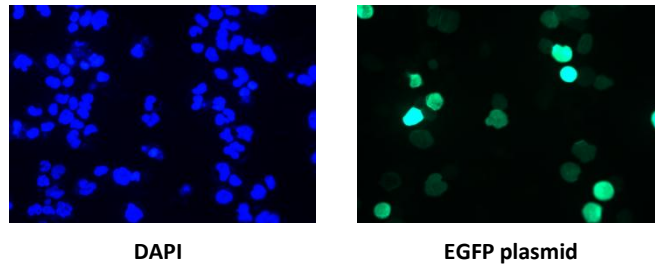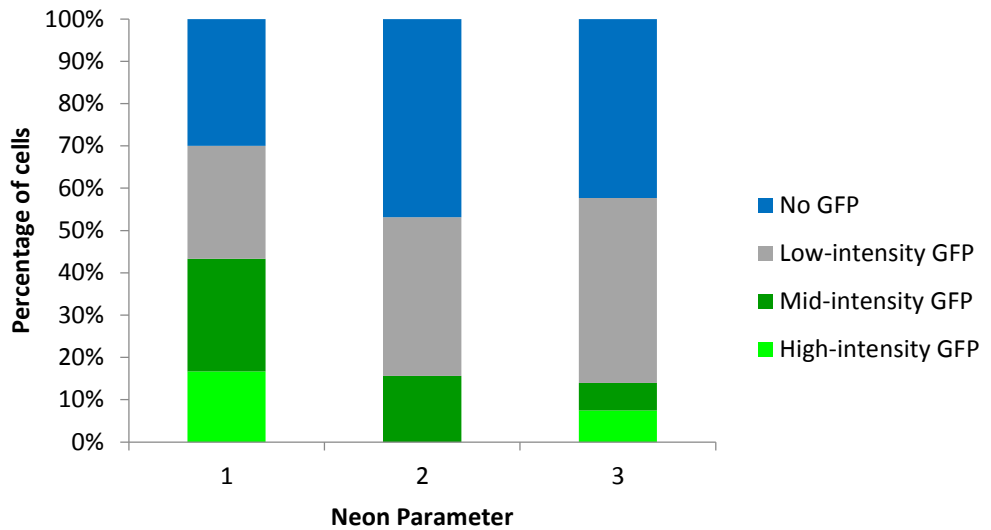
Figure 3.4. Transfection efficiency of the EGFP plasmid into KG-1 cells by electroporation. (A) EGFP transfected KG-1 cells show a heterogenous level of GFP expression, 24 hrs post-transfection. Nuclei were stained with DAPI dye. (B) Parameter set 1 (see Table 2.3) produced the most efficient transfection based on scoring for high, mid and low-intensity GFP.

After some optimisation of pulse width, frequency and voltage, DIS3 was successfully knocked-down with 100nM siRNA by electroporation in various myeloma cell lines (see Table 2.3). Knock-down was assessed by qPCR (Figure 3.5, three representative cell lines shown). As RPMI-8226 is a commonly used myeloma cell line which also showed the highest knock-down, these cells were chosen as the ex vivo model for further functional experiments.

As 100nM of this set of siRNAs (Smart Pool ON-TARGETplus siRNA, Thermo Scientific) was needed for >70% knock-down (Figure 3.6 A), lower concentrations of another two siRNAs were tested (Silencer® Select siRNAs, Life Technologies) in an attempt to reduce any off target effects. At 25nM each, these siRNAs produced a knock-down of 70% when compared to the scrambled off-target control siRNA (Figure 3.6 B); therefore, these siRNAs were chosen for use in further experiments. Figure 3.7 shows the DIS3 mRNA and protein to be knocked-down effectively in RPMI-8226 cells at 48hrs (paired t-test, p=0.001) and 72hrs post-transfection respectively, using these siRNAs.

3.4.4 GeNorm analysis reveals a differing stability of reference genes in different cell lines

There is increasing evidence that when quantifying gene expression, normalising to a singlereference gene can generate errors and lead to misinterpretation of the expression level of the target gene. Often, a commonly used reference gene e.g. GAPDH, β-actin or tubulin may be variably expressed within the same cell type, producing artefactual changes in the target gene. geNorm is a popular algorithm to determine the most stable reference (housekeeping) genes from a set of tested candidate reference genes in a given sample panel. In order to correctly quantify the level of DIS3 after knock-down compared to controls and comply with the MIQE guidelines, a geNorm experiment was performed on five cell lines that were used for DIS3 knock-down. After using the qbase software for analysis, the most stable reference genes for normalisation were found to be GAPDH and B2M for HeLa,
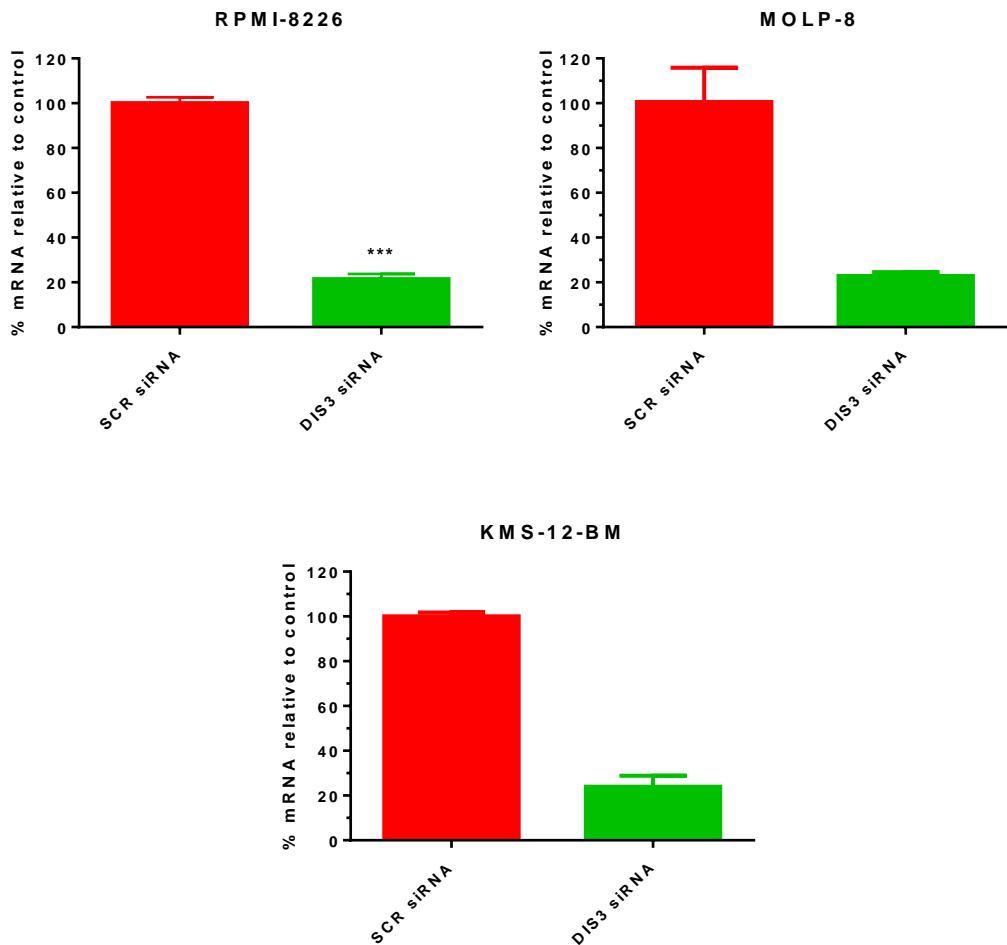
Figure 3.5. siRNA knock-down of DIS3 in three myeloma cell lines transfected using the Neon™ electroporation system. Gene silencing was analysed by qPCR 48-hours post-transfection with suitable reference genes according to GeNorm analysis (RPMI: RPL13A and EIF4A2; KMS-12BM: EIF4A2 and TOP1; MOLP-8: ACTB and CYC1). mRNA levels are expressed as a % of the scrambled control. Error bars represent the SEM obtained from two or three (RPMI-8226) independent experiments (paired t-test, p=0.002).

A
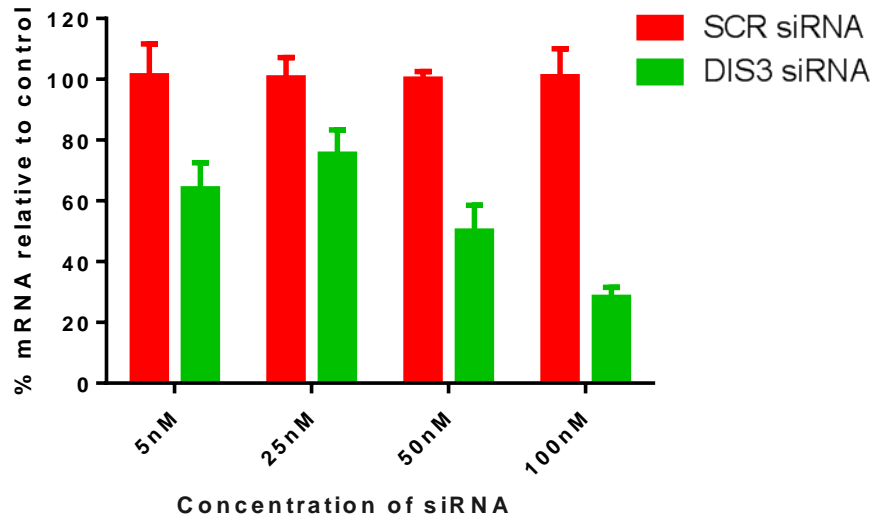
B
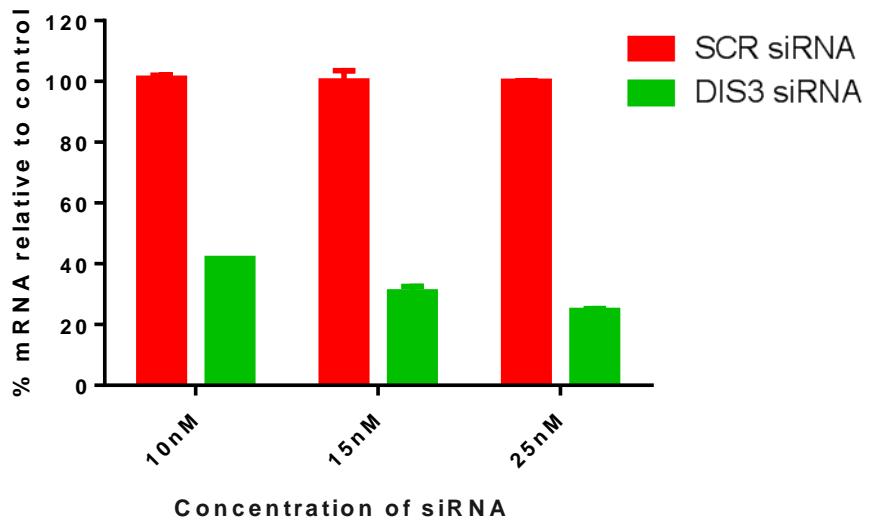
Figure 3.6. Comparison of the knock-down efficiency of ON-TARGET plus and Silencer® Select siRNAs against DIS3 in RPMI-8226 cells. Gene silencing was analysed by qPCR 48-hours post-transfection. (A) 100nM of ON-TARGETplus siRNA is required for > 70% knock-down of DIS3. (B) 25nM each of two Silencer® Select siRNAs is required for > 70% knock-down of DIS3. mRNA levels are expressed as a % of the scrambled control. Error bars represent the SEM obtained from three technical experiments. SCR = scrambled siRNA.
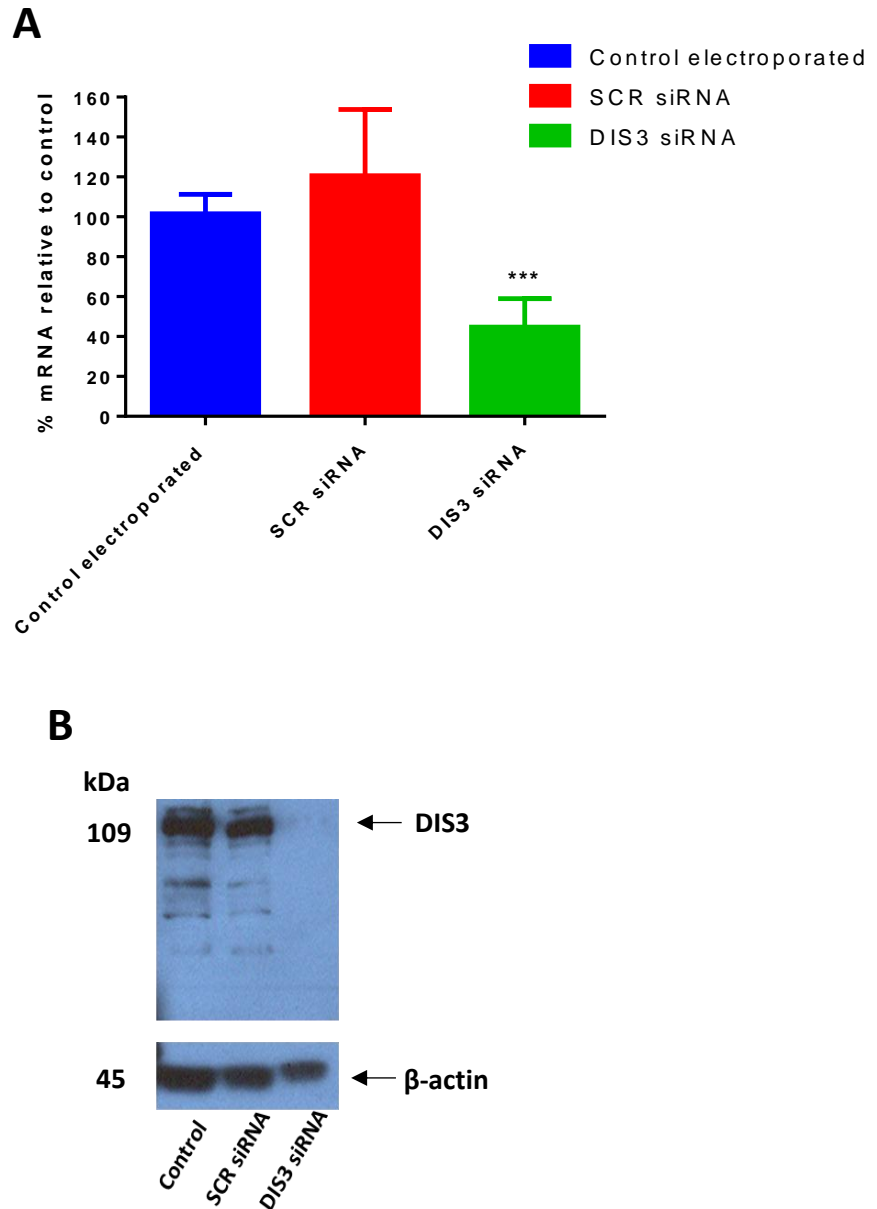
Figure 3.7. siRNA knock-down of DIS3 mRNA and protein, 48 and 72hrs post-transfection, respectively. (A) qPCR performed on five independent experiments (paired t-test, p=0.001). mRNA levels are expressed as a % of the scrambled control. Error bars represent the SEM. SCR = scrambled siRNA. (B) Western blot of one representative experiment (n=2). Total protein was extracted and an equal amount (50µg) was loaded per lane, confirmed with an actin control.

RPL13A and EIF4A2 for RPMI-8226 cells; EIF4A2 and TOP1 for KMS-12-BM cells; ACTB and CYC1 for MOLP-8 cells and B2M and GAPDH for OCI-AML3 cells (Figure 3.8). These genes were used as normalisers for these cell lines in future qPCR experiments.

3.5 Investigating the effect of siRNA-mediated DIS3 knock-down on cell phenotype

3.5.1 DIS3 knock-down does not appear to affect cell growth rate

The first step when beginning to characterise whether knocking-down DIS3 has any effect on cells is assess growth rate. As DIS3 appears to be acting as a tumour-suppressor gene based on data from biochemical assays and mutational analysis (Table 3.1), we may expect to see an increase in proliferation, consistent with a cancer phenotype. In contrast, previous studies in *S.cerevisiae* and *Drosophila* cells have demonstrated DIS3 depletion causes decreased cell proliferation (Schaeffer et al., 2012; Kiss and Andrulis, 2010). To find out whether DIS3 knock-down has an effect on the proliferation rate of RPMI-8226 cells, a WST-1 cell viability assay was performed at designated time-points post-transfection. No difference in absorbance was observed between the controls and the knock-down cells (Figure 3.9A, paired t-test, p value =0.98). Although WST1 is an indicator of cell viability, it measures the cleavage of a tetrazolium salt by a mitochondrial reductase and is therefore dependent on appropriate mitochondrial function within the cell. As I could not exclude the possibility of DIS3 knock-down affecting mitochondrial function, I repeated the experiment by directly counting the cells that excluded trypan blue. However, no difference in growth rate was observed (Figure 3.9B, paired t-test, p value =0.21). To eliminate the possibility of this being a cell type-specific effect, I repeated the experiment in three other myeloma cell lines and in HeLa cells, but again no significant difference was observed (Figure 3.9 C, D, E, F). Therefore, we can conclude that DIS3 knock-down by siRNA electroporation does not affect cell viability and thus growth rate in these cell lines. What can be
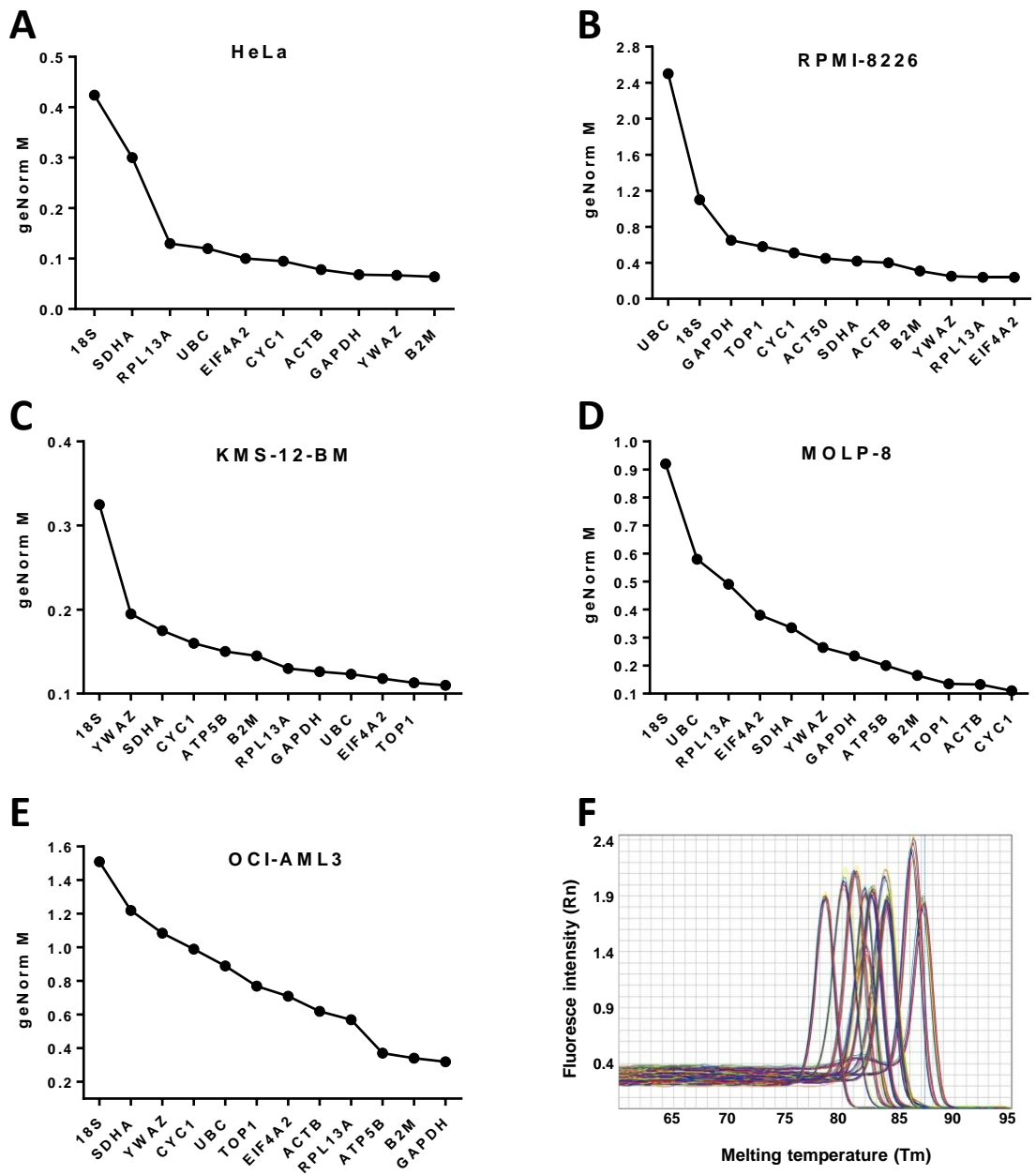
Figure 3.8. Average expression stability of a selection of reference genes within different cell lines as determined by GeNorm analysis. Genes with a geNorm M value of ≤0.2 are considered very stable. The two most stable reference genes were (A) HeLa: GAPDH and B2M; (B) RPMI-8226: RPL13A and EIF4A2; (C) KMS-12-BM: EIF4A2 and TOP1; (D) MOLP-8: ACTB and CYC1 and (E) OCI-AML3: GAPDH and B2M. (F) Representative melt curve plot demonstrates specific primer binding (KMS-12-BM).
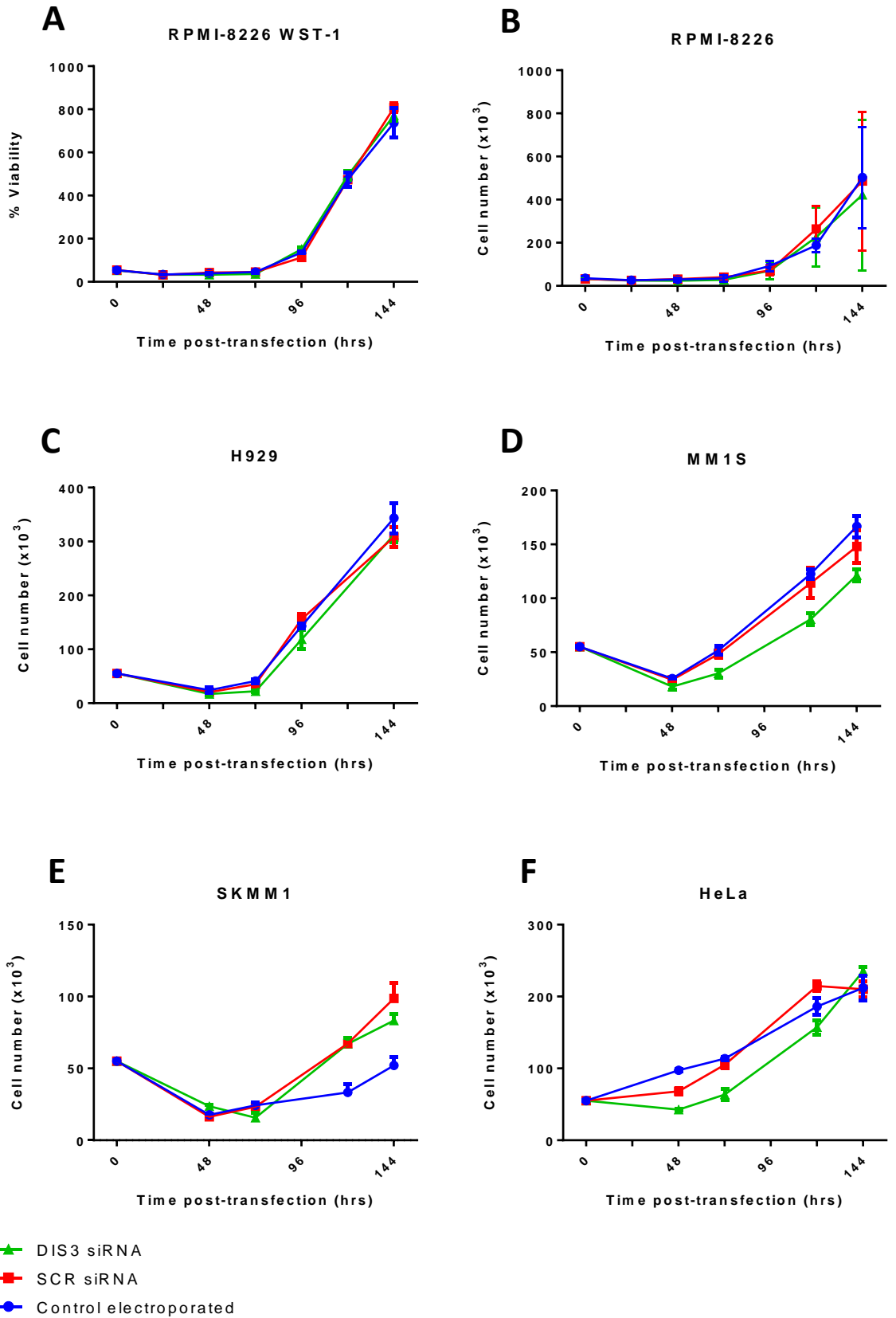
Figure 3.9. Effect of DIS3 knock-down on the growth rate of different cell lines. No significant difference was observed in (A) RPMI-8226 cells as measured by WST1 or by cell counting in (B) RPMI-8226, (C) H929, (D) MM1S, (E) SKMM1 and (F) HeLa cells. Error bars represent the SEM obtained from three (A,B) independent experiments; or three technical replicates (C,D,E,F).

observed from these experiments is the long recovery period post-electroporation, demonstrating the harsh nature of this method of transfection.

3.5.2 DIS3 knock-down does not appear to affect cell sensitivity to the anti-myeloma drug, bortezomib

The goal of myeloma therapy is personalized medicine, whereby biomarkers are used to define a patient subgroup that will benefit from a specific drug. At present, data on the prognosis of patients with DIS3 mutations and their response to particular therapies is unavailable. Traditionally, the international staging system (ISS) used clinical features such as serum albumin and β2-microglobulin as measures of prognosis (Greipp et al., 2005) but these are insufficient to direct treatment for individual patients. Although this system has been improved by including molecular cytogenetic data such as translocations and copy number abnormalities, there is a great need to take into account genetic abnormalities when selecting treatment strategies for individual patients. Dexamethasone, bortezomib and thalidomide feature frequently in treatment regimens for myeloma. For this reason, these three drugs were selected to test whether there was any differential drug sensitivity between control and DIS3 knock-down RPMI-8226 cells.

In order to determine whether DIS3 knock-down affected RPMI-8226 sensitivity to these anti-myeloma drugs, it was first necessary to obtain a dose-response curve on untransfected cells. The literature presents varying sensitive concentrations of the drugs, possibly due to differences in cell culture conditions, infection or cell-line authentication. Therefore, cells were treated with varying doses of bortezomib (0.01 to 1nM, Figure 3.10A), dexamethasone (0.1 – 1000nM, Figure 3.10B) and thalidomide (0.1-1000nM, Figure 3.10c), and) and assessed after 48 hours using the WST-1 viability assay. No change in viability was observed with dexamethasone and thalidomide at the concentrations used, however a dose-dependent response was observed with bortezomib. The lowest effective dose of bortezomib was0.1nM, therefore this concentration was used to conduct a time-course experiment.
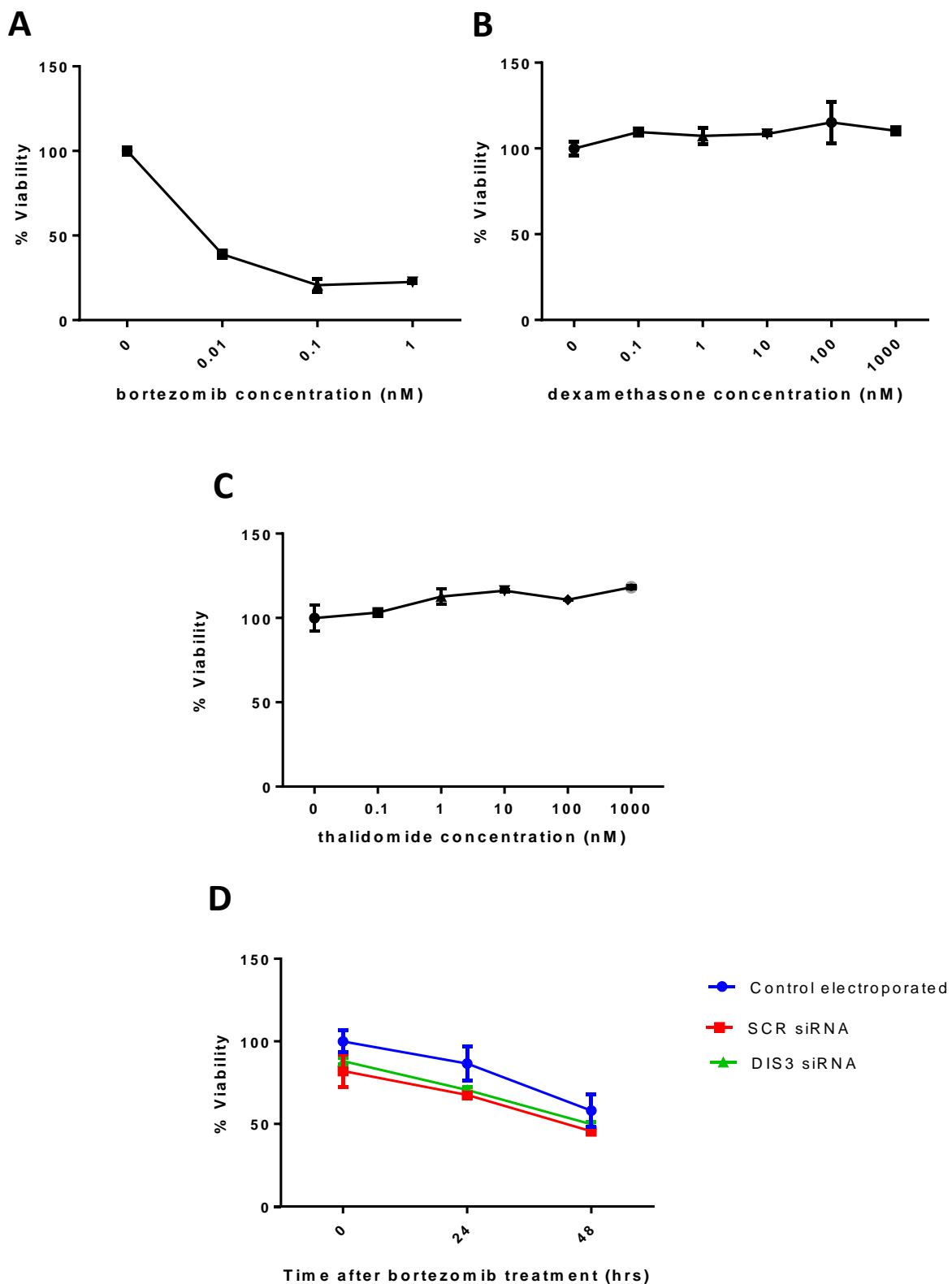
Figure 3.10. Effect of three anti-myeloma drugs on RPMI-8226 cell survival. (A) A dose-response curve shows RPMI-8226 cells are sensitive to 0.1nM bortezomib. (B) RPMI-8226 cells appear resistant to concentrations up to 1000nM of both dexamethasone and thalidomide (C). Cell survival was measured using WST1, 48-hours after drug treatment. (D) Knock-down of DIS3 does not appear to affect cell sensitivity to bortezomib. Error bars represent SEM obtained from two biological replicates. SCR = scrambled siRNA.

116

Bortezomib is a proteasome inhibitor, used frequently used for relapsed-myeloma. The mechanism by which bortezomib kills myeloma cells remains elusive, however there is evidence that it induces bcl-2 phosphorylation and cleavage causing G2-M arrest and apoptosis induction. Bortezomib was added to transfected cells at 72-hrs post-transfection and their survival monitored by WST1 over the next 48 hrs. No difference was observed in the survival of DIS3 knock-down cells compared to controls (Figure 3.10D, paired t-test, p value = 0.238).

3.5.3 DIS3 knock-down does not appear to cause mitotic errors or apoptosis

Previous studies in yeast have shown DIS3 to be required for the proper segregation of chromosomes during mitosis, potentially by regulating kinetochore function (Murakami et al., 2007; Ohkura et al., 1988). *S.cerevisiae* DIS3 mutants arrest in pre-anaphase mitosis due to abnormal chromosome segregation, subsequently causing the activation of the Mad2 spindle-assembly checkpoint (SAC) protein. In DIS3 Mad2 double mutants, cells proceed to anaphase without proper chromosome segregation, generating aneuploidic cells. This phenotype may have relevance to the pathogenesis of multiple myeloma, of which 90% of patients are aneuploid (Drach et al., 1995).

In order to investigate whether DIS3 knock-down has an effect on mitosis and ploidy, metaphase spreads were prepared from HeLa cells. Chromosomes were counted in control electroporated, scrambled and DIS3 knock-down cells but no significant difference in chromosome number was observed (*Mann-Whitney t-test, p=0.92*, Figure 3.11). In addition, RPMI-8226 cells were stained with DAPI and scored for the presence of abnormal nuclei. For each of the three treatments: control electroporated, scrambled siRNA and DIS3 siRNA, 800 cells were scored for nuclear phenotypes. Although non-significant, a small increase in the number of cells displaying poly-lobed nuclei and multiple micronuclei was observed in the DIS3 knock-down samples (Figure 3.12, X^2, p value =0.122). These phenotypes can be an indication of mitotic abnormalities or apoptotic morphological changes. In order to
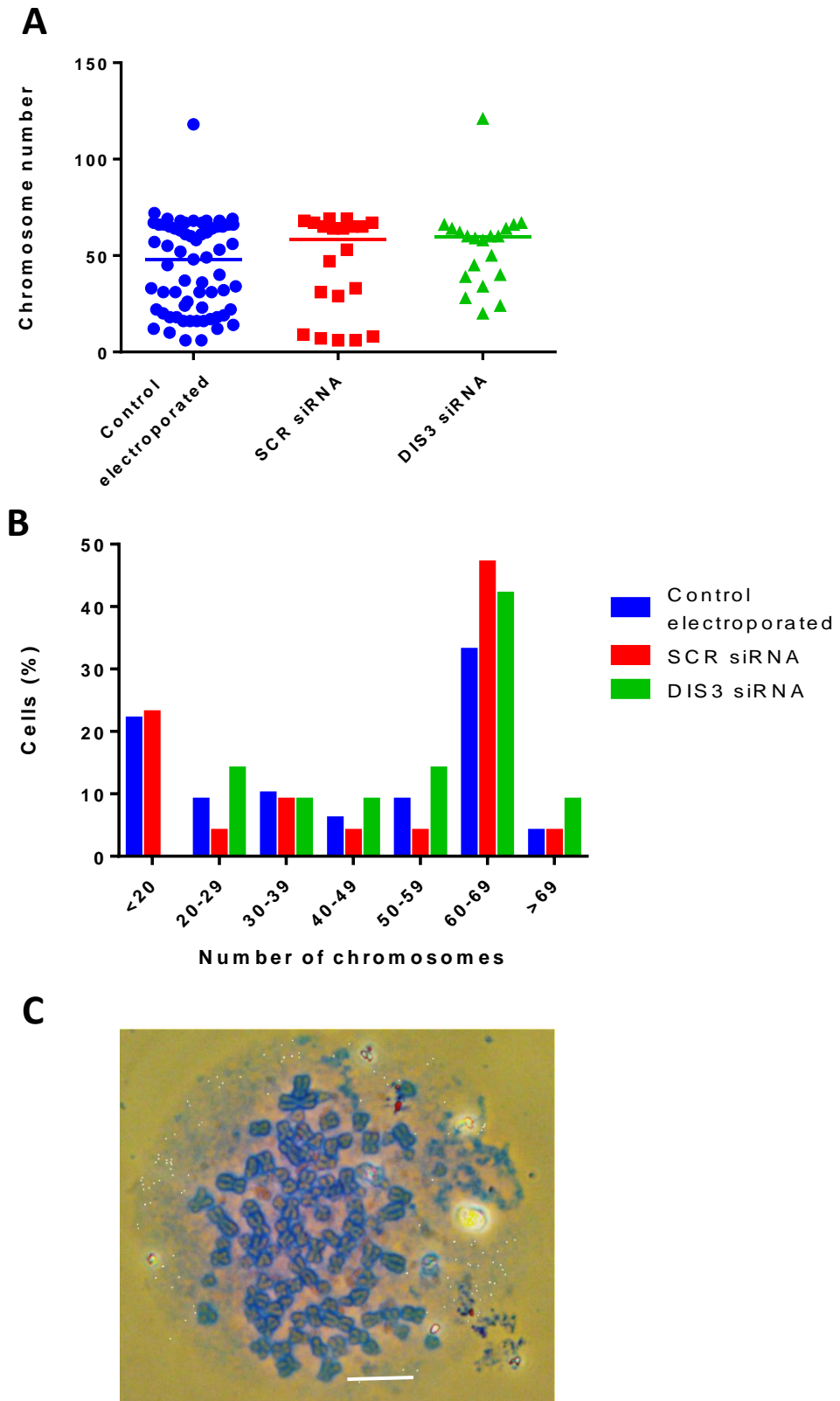
Figure 3.11. Effect of DIS3 knock-down on the chromosome number of HeLa cells.
(A) Dot plot showing the median chromosome number in control electroporated (n=65), scrambled (n=20) and DIS3 knock-down cells (n=20). Knock-down of DIS3 does not appear to affect chromosome number *(Mann-Whitney t-test, p=0.92)*. (B) Number of chromosomes are displayed as categories. (C) Representative image of HeLa cell chromosomes in metaphase, stained with giemsa dye. Scale bar, 10 μm.
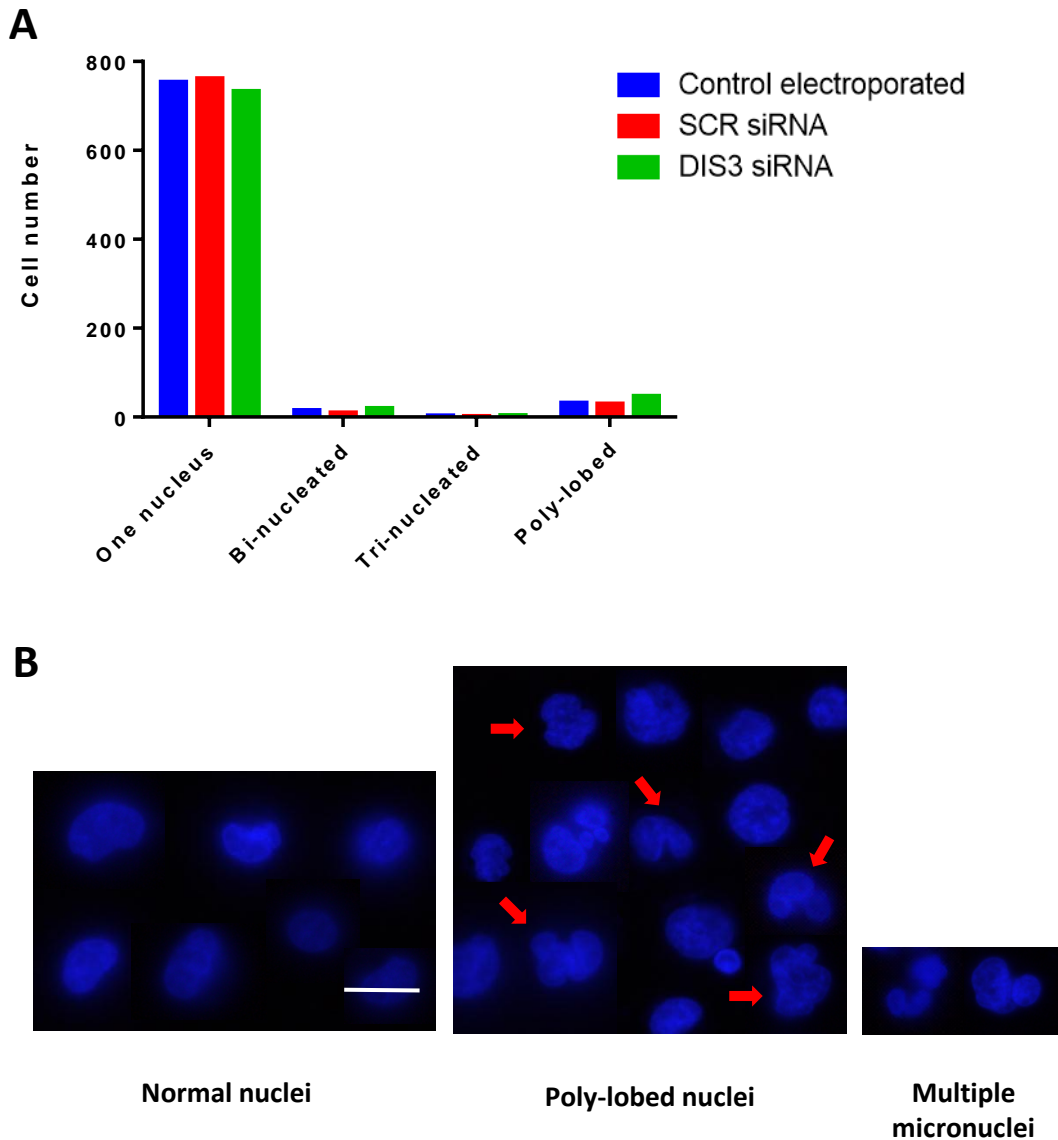
Figure 3.12. Effect of DIS3 knock-down on mitosis in RPMI-8226 cells. (A) A non-significant increase in the number of cells displaying poly-lobed nuclei and multiple micronuclei was observed in the DIS3 knock-down samples (*X², p value =0.122)*. (B) Cells with representative normal nuclei, poly-lobed nuclei and multiple micronuclei. 800 cells were scored for each treatment. Scale bar, 25 μm.

investigate this further with a more sensitive method, cells were stained with propidium iodide and their DNA content was measured by flow cytometry.

Preliminary experiments were first carried out to gate out the doublets, confirm the activity of the RNase (Figure S1) and determine the normal position of the 2n and 4n peak in this cell line (Figure 3.13). Untransfected RPMI-8226 cells were initially starved for 24 hrs (Figure 3.13 A), cultured under normal conditions (Figure 3.13 B) or treated with colchicine for 24 hrs (Figure 3.13 C). Starvation shifts cells into G1/G0 phase of the cell cycle allowing me to define the 2n peak (Figure 3.13A). Colchicine is a spindle fibre inhibitor which arrests cells in metaphase, allowing me to define the 4n peak (Figure 3.13C). When these three histograms are overlaid (Figure 3.13D) the shifts in peaks, representing shifts in position of the cell cycle, can be clearly identified.

After defining the positions of the peaks in this cell line, transfected cells were cultured under normal conditions (10% FCS) analysed at 72 hrs post-transfection and stained with PI. No difference was observed in the size of the aneuploidic peak between DIS3 knock-down cells and controls; however, there was an increase in the sub G1 peak which is characteristic of cells with fragmented DNA, as a result of cell death (Figure 3.13F). Despite there being a high level of cell death in all three treatments as a result of the electroporation method of transfection (Figure S1), knock-down samples had a significant 19.7% increase in cells within this sub-G1 fraction compared to controls (Figure 3.13F, t-test, p value =0.025).

Due to the high background level of cell death caused by the electroporation and in order to determine the mechanism of cell death represented by the sub-G1 peak, I carried out experiments to determine the level of caspase activation in DIS3 knock-down cells compared to controls. Due to a lab-wide mycoplasma infection, these assays were performed on a new batch of non-infected RPMI-8226 cells. Caspase enzymes become activated during apoptosis leading to the destruction of intracellular DNA repair elements, structural polypeptides and signalling kinases. Detection of the activity of the executioner caspases 3 and 7 is a reliable indicator
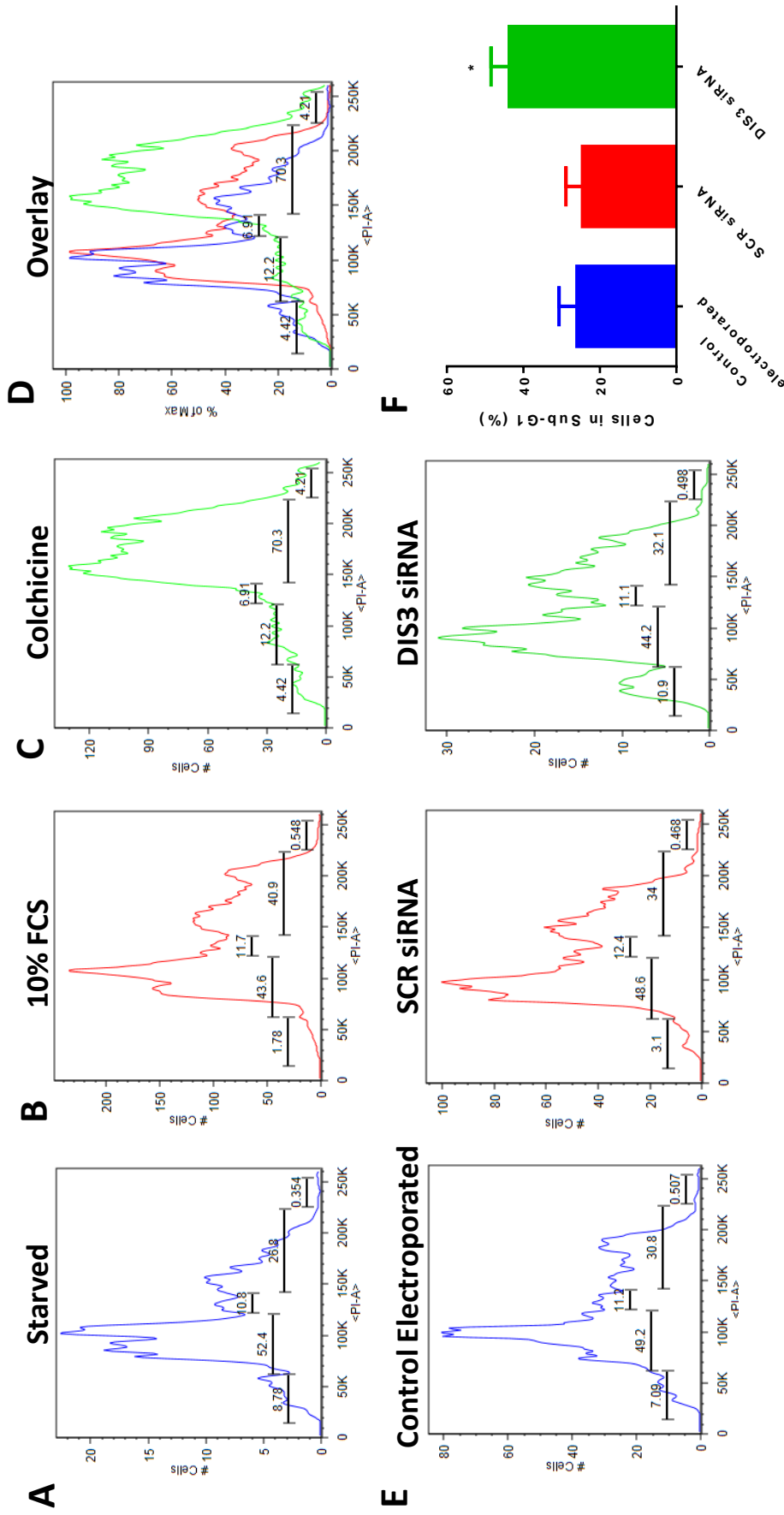
Figure 3.13. DNA content analysis of RPMI-8226 cells by flow cytometry. (A) Cells were either starved (A), cultured in 10% FCS (B), or treated with colchicine (C) for 48 hours. Starved cells show a shift towards G1/G0 whereas colchicine treated cells shift into G2/M. (D) represents an overlay of A,B and C. (E) Knock-down of DIS3 causes an increase of cells within the sub-G1 peak, which represents dying cells. Histograms are representative of one experiment of four, in the remaining three experiments % of cellz within sub-G1 was substantially higher across all three samples, explaining the higher overall % in (F). (F) DIS3 knock-down cells show a significant 19.7% increase in cells in the sub-G1 peak compared to controls (t-test, p = 0.025). Each peak represents a different phase of the cell cycle; the percentage of cells in each phase is labelled. Error bars represent the SEM obtained from four independent experiments. SCR = scrambled siRNA. PI-A = propidium iodide intensity.

of apoptosis in cells. Therefore, the caspase 3/7 Glo kit (Promega) was used and luminescence was measured at 24 hour intervals up to 96 hours post-transfection. No difference in the level of luminescence, was observed between DIS3 knock-down cells and controls (Figure 3.14). This suggests that either the increased sub-G1 peak is not a result of increased apoptosis in knock-down cells, but may represent increased necrosis, or that the mycoplasma-free cell lines are responding differently to DIS3 knock-down compared to the previously infected cells.

3.5.4 DIS3 knock-down does not appear to affect the number of DNA double-strand breaks

A previous study has implicated the exosome in recruiting activation-induced cytidine deaminase (AID) to chromatin in B-cells (Basu et al., 2011). AID creates mutations in the DNA that can lead to double-strand break (DSB) formation, an important process in antibody class switching. Myeloma often results from translocations formed from these double strand breaks. In order to investigate whether DIS3 might have a role in this process, knock-down and control cells were stained with an antibody against 53BP1, a protein which localises to DSBs and is involved in the double-strand break repair pathway (Figure 3.15A). Eighty cells were observed for each treatment and the foci counted. No difference in the number of 53BP1 foci was detected when knock-down cells are compared to both controls, (Figure 3.15B, Kruskal-Wallis test, p-value 0.5406), however when compared to the scrambled control alone, there is a significant difference (p=0.04).

3.6 Generating a stable DIS3 knock-down model

Small interfering RNA-mediated DIS3 knock-down presents limitations on the window of time in which functional assays can be performed as the protein is only knocked-down transiently. This is caused by a dilution of the siRNA as the cells divide. As Figure 3.3 illustrates, the optimal time post-transfection at which to assay changes to the mRNA is 48 hours, 24 hours after which DIS3 mRNA levels have already doubled. In order to overcome these limitations, stable-knockdown models
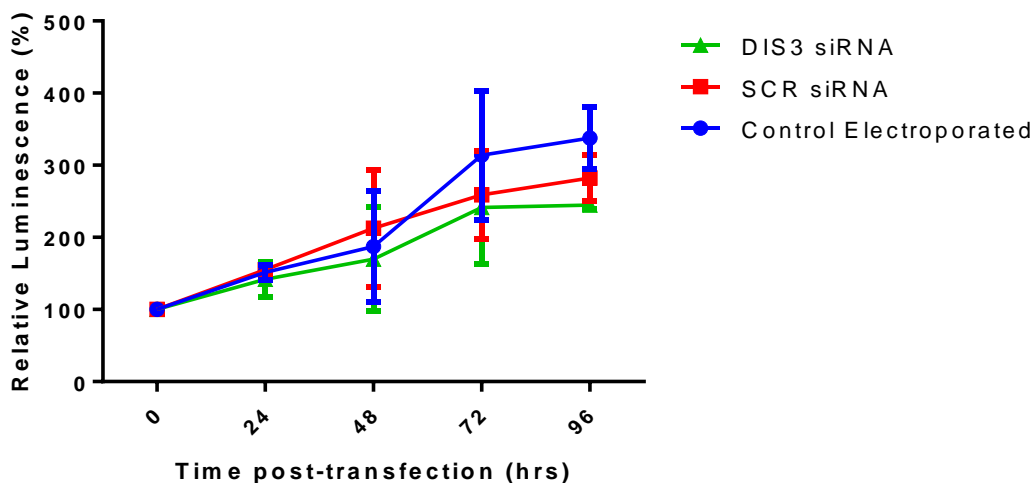
3.14 There is no difference in the level of caspase activation in DIS3 knock-down RPMI-8226 cells compared to controls. Caspase induction was measured in DIS3 knock-down, scrambled and control electroporated cell cultures using Caspase-Glo 3/7 assay kit (Promega), and the results are shown as the mean luminescence, relative to the luminescence at time point 0 ± the SEM for 3 independent experiments.

A

α-53BP1

B
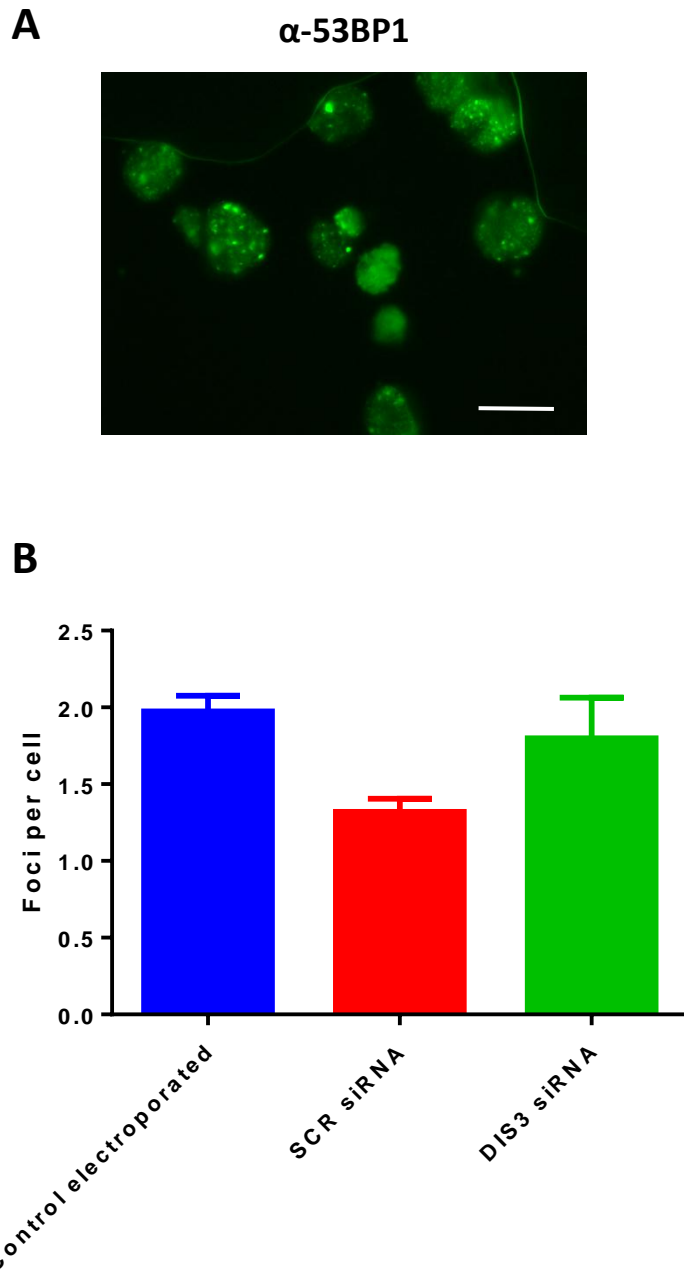
Figure 3.15. Analysis of the number of double-strand break foci within RPMI-8226 cells. (A) 53BP1 staining of a representative population of knock-down cells. (B) Knock-down of DIS3 has no significant effect on the number of double-strand breaks within cells when compared to both controls (Kruskal-Wallis test, p value 0.5406). Error bars represent the SEM obtained from three independent experiments. SCR = scrambled siRNA. Scale bar, 25 μm.

can be created in which an shRNA can be stably integrated into the cell's genome, allowing for persistent expression. For this reason, I created stable knock-down models in order to repeat the time course assays without the limitations of the transient method, as well as abrogating the problem of high cell death resulting from the electroporation method of transfection.

Three shRNA clone vectors were purchased pre-cloned into a pGIPZ vector, with at least one guaranteed to induce silencing, each targeting a different region of the DIS3 transcript (Figure 2.2). The pGIPZ vectors contain all the elements necessary for stable introduction into the genome as well as visualisation and selection of positive clones (Figure 3.16A). The shRNA hairpin consists of a 22nt dsRNA stem, complementary to the DIS3 transcript and a 19nt loop from miR-30 as well as 125nt of miR-30 flanking sequence on either side of the hairpin, in order to increase Drosha and Dicer processing efficiency. Turbo GFP allows for visual marking of positive cells, puromycin resistance gene allows for the selection of stable cell lines and the long-terminal repeats (LTRs) facilitate insertion of the vector into the genome.

The first step after receiving the vectors was to prepare the plasmid from the glycerol stock by culturing the bacteria, before isolating the DNA (Figure 3.16B) and performing a diagnostic restriction digest to confirm presence of the correct vector (Figure 3.16C). The same procedure was performed with an RFP-containing scrambled shRNA vector for use as a negative control in future knock-down experiments (data not shown). U-2OS cells were selected as an easily-transfectable cell line in which to create a stable DIS3 knock-down model using these vectors and being adherent, are more easily visualised by microscopy. Each of the three GFP-DIS3 shRNA clones was transfected into U-2OS cells in addition to the RFP-scrambled control. Puromycin was added to the cells after 48 hours and cells were left to grow to confluency for at least 4 weeks before knock-down was assessed by qPCR (Figure 3.16D) and Western blotting (Figure 3.16E). Efficient knock-down was achieved of both the mRNA and protein with all three shRNA clones.
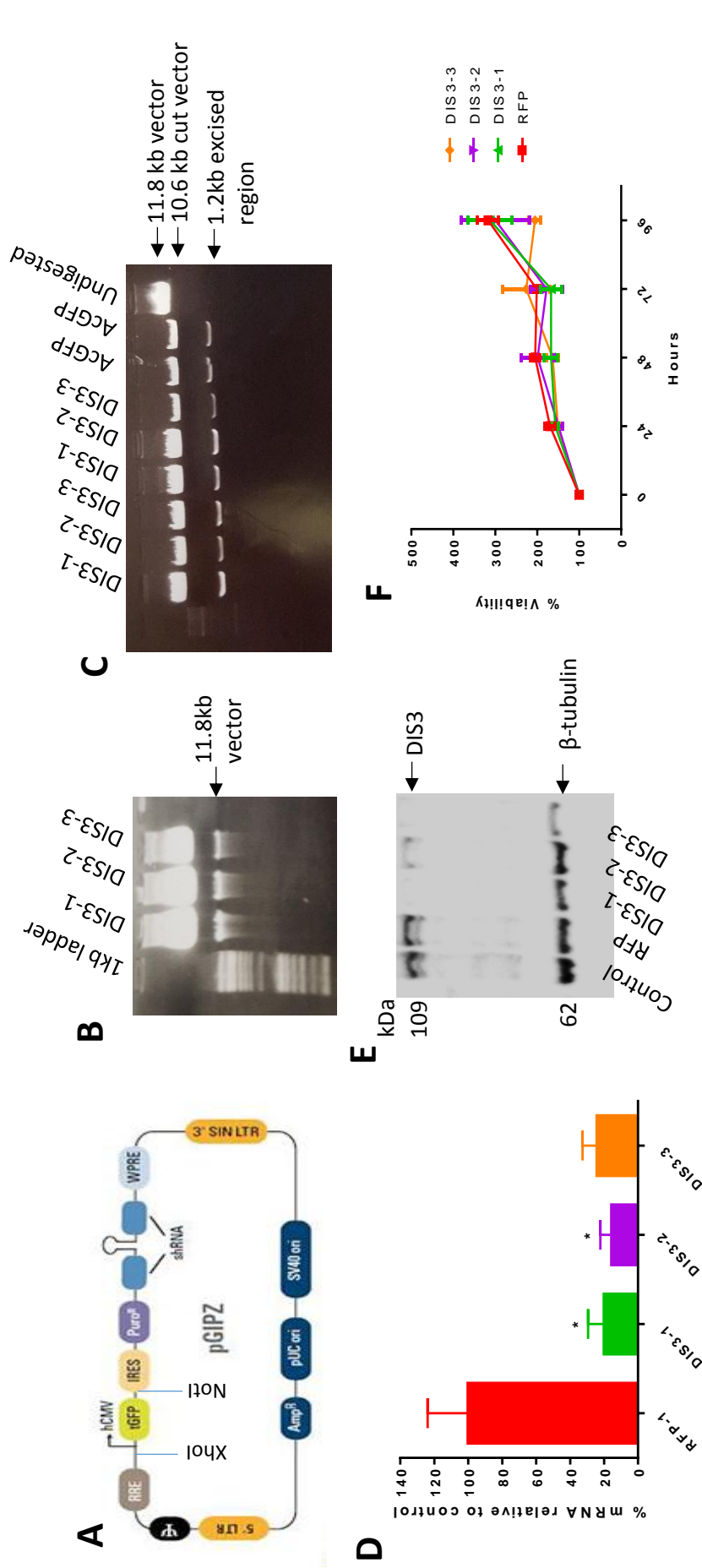
Figure 3.16. Generating a stable DIS3 knock-down model using three DIS3 targeting shRNA vectors in U-2OS cells. (A) Features of the pGIPZ vector (Dharmacon). Important elements include the hCMV promoter to drive transgene expression, tGFP reporter for visual tracking of shRNA expression, 5' and 3' LTRs for genome integration and Puro[R] to permit selection and propagation of stable integrants. The shRNA hairpin consists of a 22nt dsRNA stem, complementary to the DIS3 transcript and a 19nt loop from miR-30 as well as 125nt of miR-30 flanking sequence on either side of the hairpin, in order to increase Drosha and Dicer processing efficiency. XhoI and NotI cut sites are indicated. (B) Agarose gel showing the three DIS3 shRNA vectors after bacterial culture. (C) Diagnostic restriction digest of the three vectors with XhoI/NotI which excise the CMV/tGFP region in (A), along with two AcGFP positive controls and an undigested negative control vector. (D) qPCR showing efficient knock-down of DIS3 with all three shRNA clones when compared to a scrambled control (paired t-tests, p <0.05). Errors bars represent the SEM from four independent experiments. mRNA levels are expressed as a % of the scrambled control. (E) Western blot confirming knock-down of DIS3 at the protein level in the three clones, when compared to an untransfected and scrambled controls. (F) No difference in viability between the three knock-down cell lines and control is observed. Error bars represent the SEM from at least three independent experiments.

126

3.7 Investigating the effect of stable DIS3 knock-down on cell phenotype

To find out whether there was a growth difference in this stable knock-down model, WST-1 viability assays were performed over a 96-hour period, but similarly to the transient model, no significant difference was observed between the control and knock-downs (Figure 3.17F). Due to a lab-wide mycoplasma infection, these cell stocks had to be discarded and clean U-2OS knock-down cell lines re-generated. These cell lines underwent the same procedure as before, however a lower knock-down of DIS3 was achieved with all three shRNA clones (Figure 3.17A, paired t-tests, p >0.05). WST-1 assays were repeated on the clean cell lines to assess for any growth difference between knock-downs and controls. Two biological replicates were performed across different weeks (Figure 3.17 B, C) and appear to show an increased viability of cell lines DIS3-1 and DIS3-3, the two cell lines with the highest knock-down according to the qPCR. To assess for apoptotic differences, caspase assays were performed as before, on two individual replicates, showing what appears to be an increase in caspase activity in the knock-downs (Figure 3.17 D, E). Due to conflicting results within this model and previous knock-down models, this line of enquiry was not pursued any further.

3.8 Discussion

How loss-of-function DIS3 mutations are tumorigenic and how they contribute to multiple myeloma pathogenesis remains largely unknown. This chapter aimed to shed light on this by analysing the effect of the mutations on DIS3 activity, and subsequently develop a genetic model to mimic the disease scenario and investigate a possible phenotype.

3.8.1 DIS3 appears to be a tumour suppressor gene in multiple myeloma

When this study began, two groups had performed whole-genome/exome sequencing studies on a total of 60 multiple myeloma patients to reveal 8 distinct missense mutations within DIS3 (Chapman et al., 2011; Walker et al., 2012).
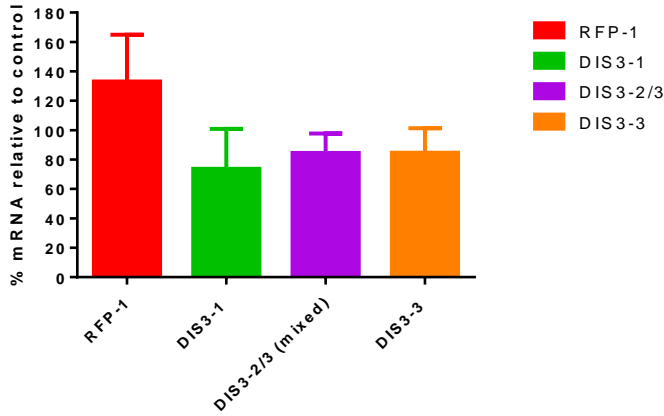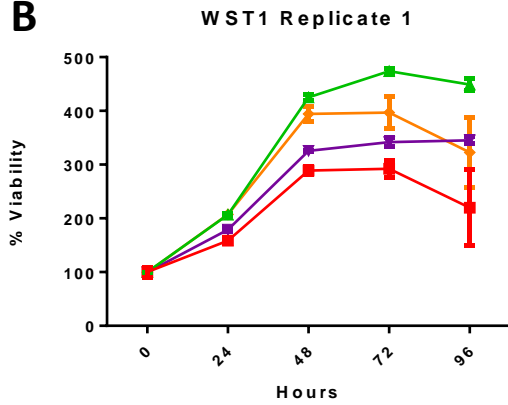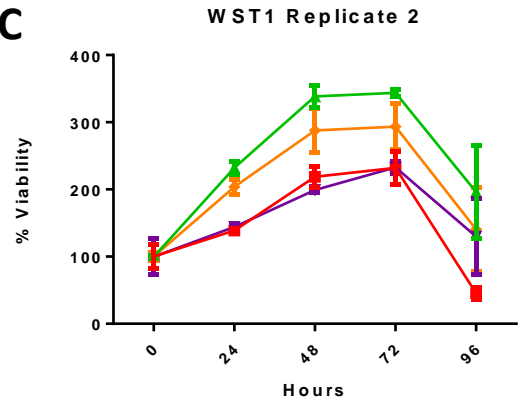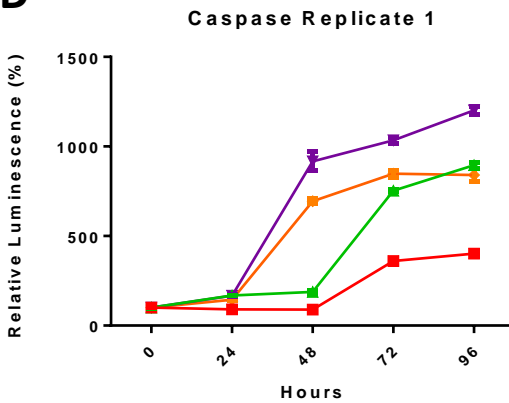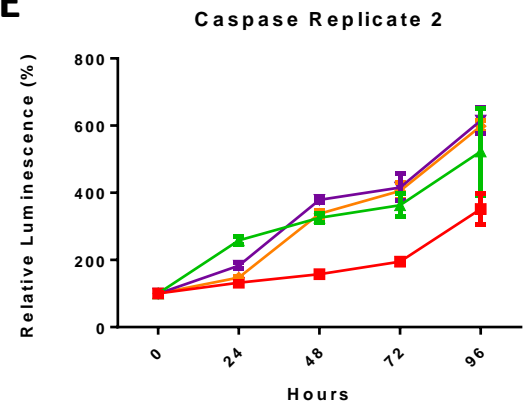
Figure 3.17. Effect of stable DIS3 knock-down on mycoplasma-free U-2OS cells. (A) qPCR showing knock-down of DIS3 with all three shRNA clones when compared to a scrambled control for five independent experiments. mRNA levels are expressed as a % of the scrambled control. (B,C) In two individual biological replicates, DIS3-1 and DIS3-3 appear to show increased viability. (D,E) In two individual biological replicates, caspase activity appears increased in the DIS3 knock-down cell lines. Results are shown as the mean optical density/luminescence for the WST-1 assays and caspase assays respectively, relative to the luminescence/OD at time point 0 ± the SEM for 3 technical replicates.

At this stage, it was unknown whether these mutations were activating or inactivating, and therefore whether DIS3 was an oncogene or tumour suppressor. Evidence for the former came from a study showing mutation of a residue within the RNB domain of the conserved bacterial homologue, RNase II, lead to a 100-fold increase in activity, turning it into a 'super-enzyme' (Barbas et al., 2009). In addition, two of the eight patients had the same single nucleotide variation, suggestive of an activating mutation. Nevertheless, 6 out of 8 of the patients exhibited loss of heterozygosity via deletion of the remaining allele and two of the mutations had been functionally characterised in yeast and bacteria where they result in loss of enzymatic activity (Barbas et al., 2009; Schneider et al., 2007), both providing evidence for DIS3 as a tumour suppressor.

To shed some light on this, we carried out a computational analysis on 7 of the 8 identified mutations at the time (Table 3.1), demonstrating that all have either a destabilising effect on protein folding or are likely to impair enzyme function. This finding strongly suggests DIS3 mutations are likely to be functionally inactivating and has since been supported by data from a number of other studies. Eight NGS studies (Chapman et al., 2011; Walker et al., 2012; Lohr et al., 2014; Leich et al., 2013; Weißbach et al., 2015; Walker et al., 2015; Lionetti et al., 2015; Bolli et al., 2014) have so far identified a total of 69 non-synonymous mutations or indels in myeloma patients. In most patients, mutations correlated with deletions of 13q (LOH), strongly pointing to DIS3 as a tumour suppressor. Moreover, biochemical assays recently performed using recombinant versions of DIS3 bearing MM-associated mutations indicate that in the majority of cases, these mutations abolish DIS3 exoribonucleolytic activity (Tomecki et al., 2014), corroborating our predictions.

Our analysis in combination with other studies provided us with enough evidence to believe the mutations are inactivating and DIS3 is a tumour suppressor gene. Thus directing the generation of a DIS3 knock-down model to investigate its function in the pathogenesis of multiple myeloma.

3.8.2 SiRNA-mediated DIS3 knock-down does not appear to affect cell phenotype

To our knowledge, when this study began, no publications existed which investigated the biological function of DIS3 in human-derived cells. In an attempt to shed light on the role of DIS3 mutations in the development of myeloma, a method of DIS3 knock-down in myeloma cells was first optimised. As was confirmed, suspension cells are notoriously difficult to transfect using conventional lipid-based methods. Therefore, a method of electroporation was adopted to facilitate siRNA-mediated knock-down. After optimising the parameters to achieve an efficient knock-down, various assays were performed based on findings from mutant/knock-downs in other model systems, to investigate a possible phenotype.

Previous studies using yeast and *Drosophila* DIS3 knock-downs and mutants have highlighted a role of DIS3 in cell-cycle regulation (Ohkura et al., 1988; Kinoshita et al., 1991; Smith et al., 2011; Kiss and Andrulis, 2010; Hou et al., 2012). Nevertheless, no effect on growth rate was observed in any of the myeloma cell lines or HeLa upon DIS3 knock-down. As well as growth rate, drug sensitivity, aneuploidy, mitosis and DNA double-strand breaks were also investigated in RPMI-8226 cells, based on findings from other studies; however, no significant difference was observed between knock-downs and controls in any of these assays. A significant difference was observed in the size of the sub-G1 peak by propidium iodide flow-cytometry, which represents dying cells. This finding was followed up with caspase assays in mycoplasma-free cells, to measure the level of apoptosis in knock-downs compared to controls, but no such difference was observed. Therefore, the increased sub-G1 peak can only either be representing a higher number of necrotic cells in the knock-down population or an enhanced response of mycoplasma-infected cells to DIS3 depletion compared to non-infected cells. Although electroporation can cause high cytotoxicity due to the perturbation of the cell membrane, leading to escape of the intracellular contents, it is not clear how an DIS3-tareting siRNA could cause more cellular damage than a non-targeting scrambled siRNA, therefore this difference may well be due to the mycoplasma

infection. The flow-cytometry experiment could be repeated in these clean cells to confirm this. However, due to the very high level of background cell death caused by the electroporation, as shown by the long recovery period on the cell growth curves and high sub-G1 peak, this model was abandoned and a stable knock-down model generated (discussed in section 3.8.5).

3.8.3 Technical reasons for a lack of phenotype

There are many technical reasons that may explain the lack of phenotype upon DIS3 knock-down. Firstly, the cells used for these assays are immortalised cell lines which already possess malignant characteristics through various mutations in oncogenes and tumour suppressors. DIS3 mutations may be drivers of the disease but they may not necessarily be needed for maintenance of MM once it has been established. Indeed, other studies that have yielded phenotypes have been performed on whole organisms, with non-malignant/transformed backgrounds. Secondly, siRNA-mediated knock-down does not completely remove the protein; in my study about 30% of the levels of the control still remained in the knock-down cells. The incomplete removal of functional protein, in combination with a potential redundancy with another ribonuclease, may be sufficient for normal functioning of the cell. Indeed, Tomecki et al were unable to observe any effect on cell growth rate using siRNA-based depletion of DIS3 in HeLa cells (Tomecki et al., 2010). Moreover, there is evidence that DIS3 and the other exosome-associated nuclear exoribonuclease, Rrp6, have partially overlapping functions (Dziembowski et al., 2007). Tomecki et al show how the depletion of DIS3 leads to slightly elevated levels of Rrp6 and vice versa, suggesting the cell is trying to compensate for the loss of one by increasing levels of another (Tomecki et al., 2010). Although, no evidence of this phenomenon could be observed in my RNA-seq data (chapter 4), upregulation of Rrp6 may well have occurred at the protein level in the DIS3 knock-down cells. Other studies have corroborated this by showing how individual knock-downs of Rrp6 and DIS3 have marginal effects on molecular phenotypes, whereas co-depletion causes complete stabilisation of substrates (Bresson and Conrad, 2013; Preker et al., 2008).

This poses the question of why Rrp6 does not compensate for the loss of DIS3 activity *in vivo*, to prevent the development of myeloma. This may be explained by a study in yeast which shows that when DIS3 is mutated (not depleted), Rrp6 activity is inhibited (Wasmuth and Lima, 2012). In the nucleus, the 10-subunit exosome complex also binds Rrp6 at the cap, forming an 11-subunit complex (Makino et al., 2013). When DIS3 is mutated, RNA trapped in an inactive RNB domain blocks the central channel, preventing RNA access to the Rrp6 active site at the top of the structure. This inhibition may be occurring in DIS3-mutated myeloma cells so that the activity of not only DIS3, but also Rrp6 is inhibited, in a dominant negative fashion. Therefore, siRNA-mediated depletion of DIS3 from the cell will not inhibit Rrp6 and may allow this enzyme to compensate for loss of DIS3 activity. This may be an important difference between the situations *in vivo* and *in vitro*. Indeed, recently when the same group created DIS3-mutant human cell lines, as opposed to using RNAi-mediated knock-down, a growth phenotype was observed in three of five of the mutants (Tomecki et al., 2014).

A further reason for a lack of phenotype may be explained by the limitations of siRNA knock-down technology. Firstly, siRNA knock-down is a transient process. The duration of the knock-down is determined by the rate of cell growth and the dilution of the siRNAs below a crucial threshold level necessary for maintaining inhibition of gene expression; as well as the half-life of the protein. The half-life of the protein will affect how long it takes post-transfection for the protein to be depleted to a sufficient level, whereas the rate of cell growth will affect how long the knock-down is maintained after this time point. Although Figure 3.7 showed a good knock-down of the protein 72-hours post-transfection in RPMI-8226 cells, the time-course assays may be directly limited by varying levels of DIS3 silencing over this period. Even the assays performed at one-time point, 72-hours post-transfection, may be limited by a delay in the cell's physiological response to this depletion. Additionally, a replenishment of the DIS3 protein after cell division, may abrogate any potential physiological effects before they are fully manifested.

3.8.4 Biological reasons for a lack of phenotype

As well as technical reasons caused by limitations of the experimental set-up, there is the possibility that perturbing DIS3 function genuinely doesn't confer a tumour-related phenotype in these cell lines, under *in vitro* conditions, or at least not one that I have examined. It may be that DIS3 is conferring a competitive advantage to the cancer cells through a mechanism that occurs only *in vivo*. Multiple myeloma is a disease of plasma cells that reside in a unique microenvironment within the bone marrow. Myeloma progression is mediated not only by direct alterations to the plasma cells themselves, but also by interactions with the stromal cells and immune cells of the bone marrow, which promote tumour growth and survival (Fowler et al., 2011). DIS3 mutations might therefore be affecting the interaction of myeloma cells with the bone marrow microenvironment which indirectly promotes pathogenesis and would not be observable in my *in vitro* experiments. Alternatively, despite the high recurrence in patients, DIS3 mutations may not be directly driving tumourigenesis, in which case we would not expect to observe a phenotype by knocking-down DIS3 alone.

3.8.4.1 DIS3 may affect the bone marrow microenvironment

Initially, the interdependence between cells of the bone marrow microenvironment and the tumour cells was thought to be specifically with osteoclasts, responsible for bone resorption. Myeloma cells have a reciprocal relationship with these cells whereby they release signalling molecules which in turn stimulate the osteoclasts to release tumour-growth factors (Mundy et al., 1974) . More recently, it has been shown that that myeloma cells also interact with numerous other cell types in the bone marrow to promote tumour growth and survival, for example, T-cells, dendritic cells, bone marrow stromal cells (BMSCs), endothelial cells and natural killer cells (Fowler et al., 2011). For example, myeloma cells have been shown to express RANKL (Sezer et al., 2002), the ligand for receptor activator of NFkB, which is known to induce osteoclast formation. Osteoclasts in turn produce high levels of the multi-functional cytokine, IL-6 which both directly promotes myeloma cell

growth (Abe et al., 2004), as well as increasing the number of IL-17-secreting T cells (Korn et al., 2008). IL-17 both induces myeloma tumour growth and inhibits immune function (Prabhala et al., 2010).

As well as their effect on osteoclasts and T-cells, myeloma cells are also known to express certain adhesion molecules which mediate their attachment to extracellular matrix proteins and BMSCs, allowing them to home to the bone marrow (Teoh and Anderson, 1997). This binding also stimulates IL-6 production from BMSCs, further enhancing tumour growth. Myeloma cells can also promote bone marrow angiogenesis by the release pro-angiogenic cytokines (Manier et al., 2012). Indeed, these pro-angiogenic factors have been shown in some cases to be produced constitutively as a result of genetic mutations (Rajkumar and Witzig, 2000). The function of dendritic cells and natural killer cells are also known to be impaired in patients with myeloma. One study has remarkably shown how these DCs were incapable of presenting the patient-specific tumour idiotype to autologous T cells, allowing immune evasion (Ratta et al., 2002).

These examples demonstrate the ability of myeloma cells to create a bone marrow niche which facilitates the growth and survival of tumour cells, not only by enhancing the microenvironment but also by regulating immune cells in a way that allows evasion of immune recognition. Although it can only be speculated at this stage, DIS3 may normally function to target these signalling molecules, which become up-regulated when the enzyme is mutated, promoting myeloma pathogenesis. Due to the nature of my experiments, I was not able to assay a possible effect of DIS3 knock-down cells on tumour microenvironment. However, one place to start may be to co-culture the transfected cells on a bone marrow stromal cell line such as HS-5 and measure cell proliferation, drug sensitivity and secretion of specific growth factors by ELISA.

3.8.4.2 DIS3 may not be a driver of tumourigenesis

The overall aim of this chapter was to try to understand what the role of DIS3 mutations are in the development of multiple myeloma and to try to validate DIS3 mutations as drivers of disease. Ultimately, there has to be functional data showing that a particular mutation can transform cells, create tumours in transgenic models or at least give a selective advantage to cells, before it can be classed as a driver. Nevertheless, no such studies have emerged in the literature since the initial discovery of DIS3 as a putative tumour-suppressor. In line with previous experiments in other organisms, two functional studies exist which show that DIS3 mutation/ knock-down reduces rather than increases tumour characteristics in human cell lines. A study using mutant HEK-293 cells observed a slowed growth phenotype (Tomecki et al., 2014) and a knock-down study in the colorectal cancer cell line, HCT116, showed reduced viability, migration and invasion (de Groen et al., 2014). This is somewhat contradictory to the phenotypes we might expect to see from depleting a tumour suppressor gene. This could be due to DIS3 possessing tissue-specific roles which implicate it as an oncogene in some tissues and a tumour-suppressor in others. Indeed, DIS3 seems to be over-expressed in some cancers and under-expressed in others. Alternatively, DIS3 mutations may act as cooperating events whereby they cannot drive tumourigenesis on their own but require another cellular pathway to be disrupted. When we consider that cancers are a result of a clonal evolution of cells that have acquired not just one, but many cooperating mutations, it is not surprising that depletion of just one of these genes *in vitro* produces a different phenotype. Interestingly, DIS3 mutations always seem to occur in parallel with hemizygous (monallelic) deletion of the RB1 region (13q14), either as del(13q) or as an interstitial deletion of the RB1 locus. The gene of interest at 13q14 may be RB1 (retinoblastoma tumour suppressor protein), or one of the miRNAs at this locus which are down-expressed in MM (miR-15a/16) (Roccaro et al., 2009), raising the possibility that mutation and selection of DIS3 as a driver mutation in myeloma is dependent on deletion of 13q14. This could be tested by creating double knock-downs or mutants to observe a potential synthetic effect.

Despite the high recurrence of DIS3 mutations in myeloma patients, there is also a chance that DIS3 mutations are not driving myelomagenesis. One of the reasons for this may be because they are required only to maintain the tumour, rather than initiate it. High throughput studies have provided semi-quantitative analysis of the size of the clonal populations carrying a particular mutation within an individual tumour. It was anticipated that mutations arising in all the clones would take part in initiating myeloma whereas mutations present only in some sub-clones would be potentiators of the disease, perhaps helping the cancer to 'adapt' to its microenvironment. However, it appears that the situation is not quite that simple. Mutations in DIS3, as well as other genes significantly mutated in myeloma, were found to be both clonal in some patients and sub-clonal in others suggesting they may be functioning sometimes as the former and sometimes as latter (Lohr et al., 2014; Weißbach et al., 2015).

Another reason why DIS3 mutations are occurring so recurrently in patients but might not be driving disease is that they have no role in cancer development but have not undergone negative selection. In cancer cells, often the DNA repair machinery is compromised meaning there is less repair of mutated DNA. Mutations which are deleterious to the cancer will undergo negative selection and be lost from the population; those that are beneficial or neutral will be retained. In WGS data, intergenic regions and genes that are lowly expressed in a particular cell type will often have fewer mutations per megabase than genes that are vital for the survival of the cancer. It is therefore possible that DIS3 is not critical to the survival of myeloma cells and therefore the high mutation rate seen in this gene is just an artefact of this.

3.8.5 Stable DIS3 knock-down in U-2OS cells produces conflicting effects on cell viability and apoptosis

In order to abrogate the issue of high cell death using the electroporation method and the limitations of a transient knock-down, stable DIS3 knock-down cell lines were created in which to investigate a functional phenotype. A high knock-down of

DIS3 was observed in the three initial DIS3 siRNA transfected cell lines; however, no effect on cell viability was observed, consistent with the result achieved in the transiently-transfected myeloma cells. After re-transfection of clean U-2OS cells due to a lab-wide mycoplasma infection, under the same selective dose of puromycin, knock-down of DIS3 was lower in the three lines. As the location in which the shRNA inserts into the genome cannot be controlled, these could be due to an effect of this or it could be a result of the mycoplasma infection. This time, two repeats of the viability assay showed an apparent increase in viability of DIS3-1 and DIS3-3, the two lines with the higher knock-down, compared to the control. Caspase assays were performed on these cell lines showing what appears to be an increase in apoptosis in the knock-downs. Although we cannot make any definitive conclusions on two replicates, this apparent contradiction of data may be explained in a number of ways. Firstly, in any one population of cells, a certain proportion will be undergoing apoptosis. Therefore, an increase in cell number will lead to a proportional increase in the number of cells undergoing apoptosis. Alternatively, as the WST-1 measures viability indirectly via mitochondrial activity, there may be more apoptosis in the knock-down population but the remaining live cells may be more metabolically active, giving increased signal in both assays. Another explanation is that there may be more live cells but also a few cells with very high caspase activity in the knock-downs compared to the control, giving a net increase in caspase signal. Further experiments would therefore need to be performed to tease out the exact situation.

The apparent discrepancy between the phenotypes seen in RPMI-8226 and U-2OS knock-down cells may a cell-type specific effect or may be as a result of off target effects of either or both the siRNA and shRNA (Kok et al., 2015). Ideally, a DIS3 mutant mouse model should be created to investigate the function of DIS3 mutations an *in vivo* set-up. Nevertheless, if DIS3 is a cooperating or passenger mutation, as discussed above, an effect on phenotype will not be observed until we understand the picture more fully. Only then can we begin to design therapeutic targets for affected patients.

Chapter 4: Using RNA-sequencing to identify potential targets of DIS3

4.1 Introduction

DIS3 a 3' to 5' exoribonuclease involved in the processing of stable RNA species and the degradation of aberrant RNAs. Known DIS3 targets include pervasive transcription products such as PROMoter uPstream Transcripts (PROMPTs) (Preker et al., 2008) and enhancer RNAs (eRNAs) (Andersson et al., 2014); small non-coding RNAs such as rRNA (Mitchell et al., 1997; Mitchell et al., 1996), tRNA (Gudipati et al., 2012), snRNAs and snoRNA (Allmang et al., 1999), as well as faulty mRNAs such as un-spliced pre-mRNAs (Bousquet-Antonelli et al., 2000), mRNAs with defective polyadenylation (Milligan et al., 2005) and mRNAs lacking a stop-codon (van Hoof et al., 2002; Frischmeyer et al., 2002). Although DIS3 and the exosome are also responsible for the degradation of ARE-containing mRNAs as a means of controlling gene expression (Chen et al., 2001), specific targets have not yet been identified. Therefore, although DIS3 loss-of-function mutants have defects in cellular processes such as the cell-cycle, the mechanism behind these phenotypes, and thus how DIS3 mutations are contributing to multiple myeloma, is not understood. In order to identify specific mRNA targets of DIS3, transcriptome profiling can be performed.

Various technologies have been developed to deduce and quantify the transcriptome under different conditions. Traditionally, hybridisation-based approaches were used that involved incubating fluorescently labelled cDNA with custom-made microarrays. A series of oligonucleotides which correspond to fragments of the genome are spotted onto a nylon membrane and their sequence determines their specificity for their target molecule. A sample of fluorescently labelled target molecules (e.g. cDNA) is applied to the slide, which allows for complementary base pairing between probes and target molecules. The amount of fluorescence is quantified at each individual spot, which is relative to the

abundance of target molecule in the original sample. Several samples can be compared to identify differentially expressed targets allowing for gene expression analysis under different conditions.

Alternative to microarrays, sequence-based approaches can be used which directly determine the cDNA sequence. Initially, Sanger sequencing was used but this approach is relatively low throughput, expensive and generally not quantitative. Recently, the development of novel high-throughput DNA sequencing methods has provided a new method for both mapping and quantifying transcriptomes. In this method, termed RNA-Seq (RNA-sequencing), complementary DNAs (cDNAs) generated from the RNA of interest are directly sequenced using next-generation sequencing technologies. The reads obtained from this can then be aligned to a reference genome in order to construct a whole genome transcriptome map.

RNA-Seq has clear advantages over existing approaches. Whereas microarrays have a reliance upon existing knowledge about transcript sequences, RNA-seq is not limited by prior knowledge and so captures both known and novel features meaning the data suffer from much lower biases (Wang et al., 2009). Additionally, due to the ability of probes to hybridise to more than one transcript, microarrays often suffer from cross-hybridisation artefacts, leading to false positives and high background levels. In contrast, RNA-seq has very low, if any, background signal because RNA sequences can be unambiguously mapped to unique regions of the genome. Microarrays also lack sensitivity for genes expressed either at low or very high levels and therefore have a much smaller dynamic range (one-hundredfold to a few-hundredfold). In contrast, RNA-seq provides digital data in the form of aligned read-counts, allowing for a very wide dynamic range (9,000-fold) (Nagalakshmi et al., 2008), improving the sensitivity of detection for rare transcripts. Moreover, comparing expression levels across different microarray experiments is often difficult and can require complicated normalisation methods.

In addition to the above mentioned advantages of RNA-seq also has very high levels of reproducibility, for both technical and biological replicates (Cloonan et al., 2008; Nagalakshmi et al., 2008) and has been shown to be highly accurate for quantifying

expression levels of exons and splicing variants (Mortazavi et al., 2008). For this reason, RNA-seq was the method of choice in attempting to identify mRNA targets of DIS3 whose misregulation may underlie the positive selection of DIS3 mutations in myeloma.

4.2 Aims

In order to gain an insight into which genes may be targeted by DIS3 this chapter aims to:

1. Perform RNA-seq on DIS3 knock-down and control cells
2. Validate any DIS3 potential targets using Taqman qPCR
3. Perform gene ontology (GO) analysis on differentially regulated genes to determine whether DIS3 targets mRNAs belonging to a particular biological pathway

4.3 Sample Preparation

In order to identify potential targets, a reverse genetics approach is needed to deplete the function of DIS3 from the cell. As discussed in the previous chapter, DIS3 can be knocked-down in suspension cells using siRNA-based electroporation. In order for this study to be relevant to multiple myeloma, the myeloma cell line RPMI-8226 was selected as the cell line of choice for transcriptome profiling. As in chapter 3, DIS3 was knocked-down using 25nM of Silencer select siRNA resulting in a 70% knock-down when compared to controls (Figure 3.7). Two sets of controls were used: scrambled siRNA transfected (SCR) and electroporation only (CE). The scrambled siRNA acts as a negative control to distinguish sequence-specific silencing of the DIS3 siRNA from non-specific effects. By grouping both controls in the later analysis, the real effects of DIS3 can be confidently established. Cells were transfected and snap frozen after 72 hours. Snap freezing the cells prevents any alterations to the transcript profile and according to the Western blot in Figure 3.7, at 72-hours post-transfection the DIS3 protein is knocked-down optimally. The RNAeasy kit (Qiagen) was used to extract RNA from two biological replicates each

of DIS3 knock-down, scrambled siRNA and transfection-only control cells before DNase treatment.

The quality of the initial RNA samples is important to ensure accurate and reproducible results are generated from RNA-seq; therefore, quality was checked using two methods. Firstly, the 260/280 and 260/230 ratios were assessed on a Nanodrop spectrophotometer (Table 4.1). The 260/280 ratio represents the RNA to protein ratio and will identify any protein contamination. The 260/230 ratio identifies any contamination with organic solvents. For RNA sample to be of high enough quality for sequencing both these ratios should be greater than 1.8. In addition to the purity of RNA, the RNA integrity is also important to ensure minimal degradation that may otherwise be mistaken for differential expression. A Bioanalyser (Agilent) was used to assess the integrity of all six samples (Figure 4.1). Using electrophoretic separation on microfabricated chips, RNA samples are separated and subsequently detected via laser induced fluorescence detection. An electropherogram is generated and RNA of high quality shows two clear peaks representing the 18S and 28S ribosomal RNAs. An RNA integrity number (RIN) is generated to give a clear measure of RNA quality. All six samples had RIN values above 8 and are therefore considered high enough quality for RNA-seq.

Once the quality of the RNA was confirmed, 3µg of each sample was sent to Oxford Gene Technology for library preparation and RNA sequencing across two lanes of an Illumina HiSeq2000 machine.

4.4 Library preparation

The preparation of cDNA libraries and the RNA sequencing was outsourced to Oxford Gene Technology. Library preparation was performed by OGT using the Illumina TruSeq kit according to the general workflow shown in Figure 4.2 (Labome). As the principal interest of this work was to identify mRNA targets of DIS3, the first step involved purifying poly-A containing RNA using oligo-dT attached

Sample	260/280	260/230	Yield (ng/µl)
CE_1	1.98	1.86	410.7
CE_2	2.08	2.04	162
SCR_1	2.09	2.00	592.5
SCR_2	2.05	1.86	197.1
KD_1	2.06	2.01	385.2
KD_2	2.10	2.02	178.5

Table 4.1. **RNA purity values for the samples sent for RNA-sequencing after DNase treatment.** The 260/280 and 260/230 ratios are above 1.8 indicating minimal contamination with organic solvents and protein respectively.

Figure 4.1 Bioanalyser report of the RNA integrity of the 6 samples used for RNA-sequencing. (A) Gel showing two clear peaks representing 18S and 28S ribosomal RNA in each of the six samples. (B) Electropherograms showing the fluorescence intensity of RNA at different lengths show two distinct peaks representing the 18S and 28S rRNA. All RIN values were above 8.5 except for sample 2 where the RIN value is absent due to a technical fault. Sample 1= KD1, sample 2=KD2, sample 3= SCR1, sample 4= SCR2, sample 5= CE1, sample 6=CE2.

Figure 4.2. Illumina TruSeq RNA library preparation work-flow. PolyA+ RNA is enriched using oligo(dT) beads followed by fragmentation and reverse transcription. The 5' and 3' ends of cDNA fragments are next prepared to allow efficient ligation of adapters containing a unique barcode and primer binding sites. Finally, ligated cDNAs are PCR-amplified and ready for cluster generation and sequencing. Rd1 and Rd2 = Read 1 and Read 2 sequencing primers for paired-end sequencing. (Labome)

magnetic beads. This method also allows for the removal of rRNA, accounting for 95% of cellular RNA, which can otherwise hinder sufficient coverage of mRNA. Oligo-dT selected mRNA was then fragmented using divalent cations under elevated temperature to produce 200bp fragments. Following fragmentation, samples were primed for first strand cDNA synthesis using random hexamers, followed by synthesis of the second cDNA strand which removes the template RNA, generating double stranded cDNA.

The fragmentation of the initial mRNA molecules generates overhangs which are undesirable for the following adaptor ligation steps. The overhangs in the resulting cDNA were therefore converted to blunt ends using a 3'-5' exonuclease to remove 3' overhangs and a polymerase to repair the 5' ends. Prior to adapter ligation, the 3' ends were adenylated to prevent self-ligation and provide an overhang to use for adaptor ligation. Following end repair the 3' and 5' adaptors were ligated. The adapters have two functions: to allow hybridisation to the flow cell during the cluster generation as well as to provide a priming site for sequencing. The ligated adaptors also contain the specific index sequences which allow for sample identification following pooling of samples prior to sequencing. Using the adaptors as priming sites, a few subsequent rounds of PCR were performed to enrich the cDNA fragments that have successfully ligated both the 3' and 5' adaptors. Minimal rounds of PCR were used to avoid generating errors which would be represented in the sequencing.

4.5 RNA-sequencing

Prior to cluster generation and sequencing all the RNA samples were pooled together. After pooling the cDNA was denatured into single stranded cDNA and passed across the flow cell. The principle of the sequencing technology used is shown in Figure 4.3A (Eurofins Genomics). Preparation of the sequencing library is done by bridge PCR and the sequencing is done by cyclic reversible termination technology. In bridge PCR, the two adaptors ligated at either end of the cDNA
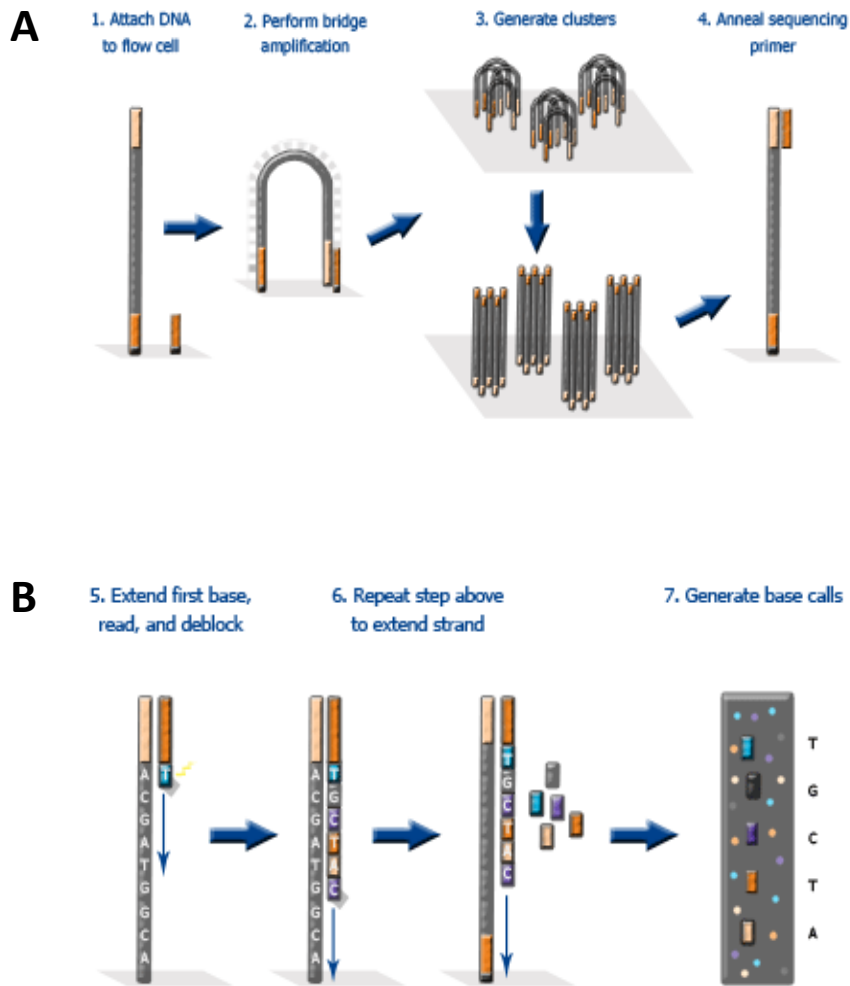
Figure 4.3. Principle of the Illumina sequencing technology. (A) Preparation of the sequencing library by bridge PCR. Single stranded DNA library fragments are attached to the flow cell at one end via complementary primers before free-ends also attach, creating bridge structures. Enzymes and nucleotides are added and the hybridised DNA strand is used as a template to create double-stranded bridges. Denaturation and repetition of this process leads to clonal clusters of localised identical strands. Finally, sequencing primers are added to begin the sequencing stage. (B) Sequencing of the clusters is performed using cyclic reversible termination. A single, fluorescently labelled terminator nucleotide is incorporated into the complementary strand during each cycle and each base is identified using base calling software. (Eurofins Genomics).

molecule hybridise with complementary oligonucleotides bound to the flow cell, creating a bridge. Unlabelled nucleotides and enzyme are added to initiate amplification and double-strand bridges are created using the hybridised cDNA strands as templates. The double-stranded molecules are then denatured before the process is repeated many times to generate clusters containing multiple, identical sequences.

Sequencing of the clusters is performed using cyclic reversible termination (Figure 4.3B). A single, fluorescently labelled terminator nucleotide is incorporated into the complementary strand during each cycle. Each nucleotide has a different dye attached. Just one labelled nucleotide is incorporated during each cycle and unincorporated nucleotides are washed away. This process occurs in both directions along the cDNA molecule to generate paired-end reads. As the cycle number increases the complementary strand is built for each hybridised cDNA molecule and each successfully incorporated base is identified or 'called' using base calling software. The number of cycles is therefore directly proportional to read length. In these experiments 100 cycles were used generating reads of 100 nucleotides in length. Additional, post-sequencing 'index' reads are then carried out to identify the index sequence to allow sample separation.

4.6 Differential expression analysis

Bioinformatics analysis of the RNA-seq data was initially performed by Oxford Gene Technology prior to additional analysis by Stewart Stevens. Below is a discussion of the analysis performed by Stewart Stevens. Both analyses used the Tuxedo protocol; however, the exact parameters used in the algorithms differed. Figure 4.4 gives an overview of the analysis pipeline.
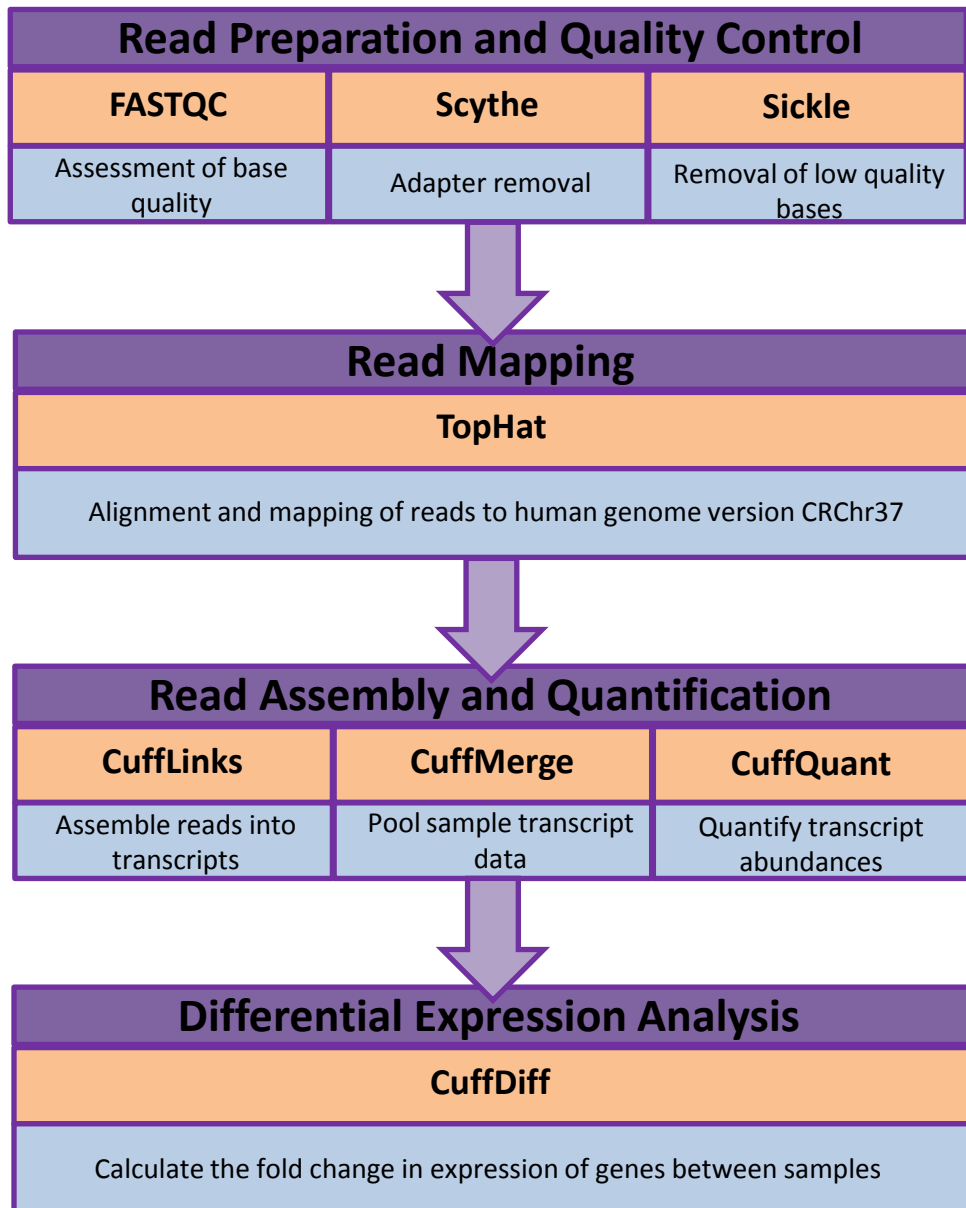
Figure 4.4. **Overview of the RNA-seq analysis pipeline using the Tuxedo protocol.** Orange boxes represent the programs used for each stage; blue boxes describe the function of each program.

4.6.1 Read Quality Control

The sequencing run produces FASTQ files which contain the information on the base calls and quality of each base. These are the initial files used in an extensive mapping and analysis pipeline to identify differentially expressed mRNAs between the control and knockdown samples. The first step in the analysis is a quality control step. Figure 4.5 gives a visual representation of the base quality of sample CE_1 produced by the FastQC program. In general, all bases with a quality score of >30 are regarded as high quality. In addition to the quality of the sequencing, FastQC was used to identify the total number of reads each sequence received together with overrepresented sequences. Only in the forward samples was an overrepresented sequence identified and each time this corresponded to the Illumina adapter, occurring on average in 0.74% of the samples.

The overrepresented adapters and any reads or bases that were of insufficient quality required removal. An algorithm called Scythe was used to remove the adapter sequences. Scythe uses base quality, a user-input sequence and probability of finding an adapter to identify the adapter derived bases for removal. Base quality aids Scythe in the identification of the 3' adapters as the quality tends to get poorer towards the 3' end where the adapters are located.

After Scythe-mediated removal of the adapter sequences, the quality of the remaining bases was assessed and those that were substandard were removed. The bases at the 5' and 3' end of reads tended to be of lower quality (Figure 4.5). An algorithm called Sickle was used to trim the read sequences so that only high quality bases remain for the mapping stages. Sickle was run in paired-end mode and reads that lost their pair (about 1% of reads) remained in a singles file for inclusion in the alignment and differential analysis stages. Table 4.2 shows the number of paired and single records kept and discarded by Sickle.

Quality scores across all bases (Sanger / Illumina 1.9 encoding)
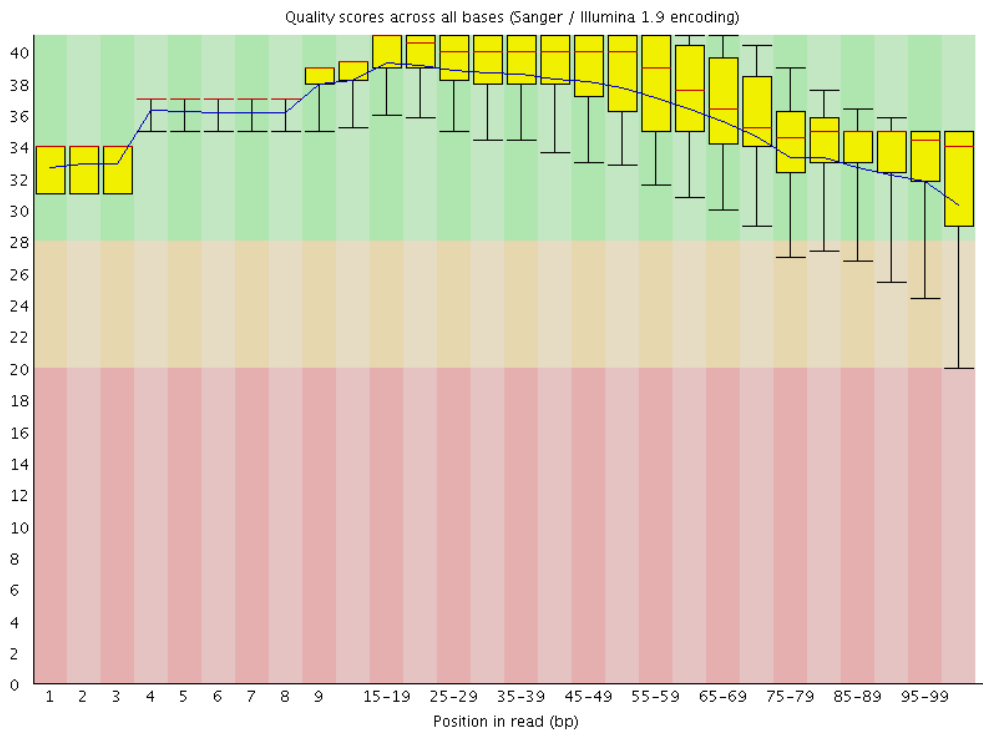
Position in read (bp)

Figure 4.5. **Graphical presentation of base quality of reads within the CE_1 sample produced by FASTQC**. Y-axis shows the base score with >30 representing high quality bases. X-axis shows the base position in the 100nt read. As expected the lowest base qualities occur at the start and end of the read.

Sample	Paired kept	Single kept	Paired discarded	Single discarded
CE_1	28131612	388719	339840	388719
CE_2	29891296	412574	243938	412574
SCR_1	33994646	496960	260958	496960
SCR_2	31607108	374511	361186	374511
KD_1	28006726	315661	247080	315661
KD_2	27886882	281790	155416	281790

Table 4.2. Summary of the number of discarded and remaining raw read sequences following Sickle and Scythe processing. Over-represented sequences were also successful removed.

4.6.2 Read alignment and mapping

Following adapter removal and quality trimming, the remaining high quality reads were mapped to the genome using TopHat. The Illumina iGenome sequences were used as a reference, prepared from ENSEMBL data, version CRChr37. TopHat2 works by breaking the reads down into smaller fragments and mapping them to the sequence supplied in the reference genome. When several of a read's segments align to the genome far apart from one another, TopHat infers that the read spans a splice junction and estimates where that junction's splice sites are. The total number of reads that were successfully mapped by TopHat for each sample are summarised in Table 4.3. The major output from TopHat2 was a file containing the genomic location of each read, allowing identification of the mRNA from which it was derived.

4.6.3 Assembling the reads

In order to quantify the total number of reads each transcript receives, the aligned reads must first be assembled into individual transcripts using an algorithm called Cufflinks (Trapnell et al., 2012). A sample may contain reads from multiple splice variants for a given gene; therefore, Cufflinks must infer the splicing structure of each gene. As genes often have multiple alternative splicing events, there are many possible reconstructions that explain the sequencing data. Cufflinks reports a parsimonious transcriptome assembly of the data, reporting the smallest possible set of transcripts that are needed to explain the input reads.

Although Cufflinks was used to assemble the samples individually, it is necessary to pool the data before proceeding to differential analysis. Samples were merged using CuffMerge before the transcript abundances were quantified by an algorithm called CuffQuant. To improve the accuracy of transcript abundance, a number of optional parameters were run such as 'frag bias correct' which corrects for any read bias. Reads which mapped to less than 20 locations within the genome were

Sample	Input Reads	Mapped reads (% input)	Multiple alignments (% input)	Aligned pairs (concordant alignment rate)
CE_1	28131612	27435495 (97.5)	854203 (3.1)	13546786 (95.2)
CE_2	29891296	29197401 (97.7)	935139 (3.2)	14423486 (95.4)
SCR_1	33994646	32006919 (94.2)	981907 (3.1)	15804779 (91.9)
SCR_2	31607108	30911634 (97.8)	938290 (3.0)	15269163 (95.6)
KD_1	28006726	27262293 (97.3)	864818 (3.2)	13467149 (95.2)
KD_2	27886882	27176120 (97.5)	874780 (3.2)	13424578 (95.2)

Table 4.3. Summary of the TopHat2 read alignment procedure. The number of input reads is presented for each sample with an average of 96% successfully mapped to the genome. Only a small fraction of reads map to multiple places within the genome; those that mapped to >20 positions were discarded. On average 94% of the read pairs mapped to the same location in the genome as would be expected (concordant alignment rate) which shows a highly successful TopHat2 run.

allowed. To improve accuracy of the weighting of these reads the 'multi read correct' function was used.

Due to the initial RNA fragmentation step, longer transcripts will contribute to more fragments and are more likely to be sequenced. Therefore, CuffQuant normalises the read counts to transcript length and sequence depth when quantifying the transcripts. The expression level metric that is calculated is called the FPKM - fragments per kilobase and million mappable reads. At 20 million reads, the resulting gene expression levels are within 5% accuracy.

4.6.4 Calculating differential expression

The next stage in the analysis was to assess the differential expression between the control and DIS3 knockdown samples by using a program called CuffDiff. CuffDiff calculates expression in two or more samples and calculates the significance of observed changes in expression. CuffDiff can also identify genes that are differentially spliced. Changes in abundance of isoforms relative to one another reflect differential splicing of their common pre-mRNA.

Due to only having two biological replicates rather than triplicates, the 'min reps for js test' function was set to 2, meaning the confidence in the statistical output will be very low. In addition, due to the variable efficiency of mRNA enrichment methods, rRNA reads were masked to improve the overall robustness of transcript abundance estimates.

Differential expression was assessed using two comparisons: scrambled (SCR) versus knock-down (KD) and control electroporated (CE) versus knock-down. Both comparisons grouped together the two control and knock-down replicates. CuffDiff provided an output showing the fold change of each transcript between the conditions in addition to the raw p-value and the p-value corrected for multiple comparisons (q-value). Finally, a program called CummeRbund was used for further quality control and to visualise the expression data. Figure 4.6 shows the cross-
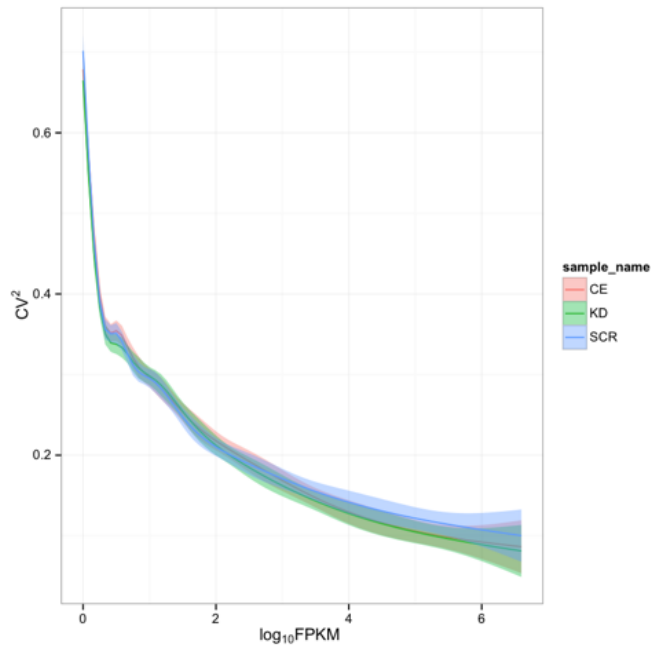
Figure 4.6. **Cross-replicate variability between replicate FPKM values as visualised by CummeRbund.** The y-axis is the squared coefficient of variation – a normalised measure of variance between replicate FPKM values. The x-axis is the mean FPKM value. The mean (line) and range of variabilities (shaded area) are displayed at different FPKM levels. CE = electroporated only control, SCR = scrambled transfected control, KD = DIS3 knock-down.

replicate variability at different FPKM levels. As expected, variation is higher at low FPKM values. The different conditions are consistent with each other in respect to noise at different expression levels.

4.7 Global overview of the RNA-seq data

Completion of the above analysis pipeline generated differential expression data that can be used to evaluate the effect of DIS3 knock-down on myeloma cells. Two sets of data were generated: those produced from the analysis by OGT and those produced from the analysis by Stewart Stevens. Unless otherwise stated, results presented represent those obtained by the latter.

Prior to assessing expression changes of individual genes between the two conditions, it is useful to explore global expression patterns across the samples. A total of 18906 genes including protein-coding and non-protein coding were expressed in RPMI-8226 cells. The levels of expression ranged from 0.0001 FPKM (APOC2) to 8512 FPKM (HBE-1) indicating a large range of gene expression levels within this cell line. In order to visualise global expression data across the six samples, a dendrogram was generated and gene expression levels were compared using kernel density plots (Figure 4.7 A). Encouragingly, the two control groups were more similar to each other than to the knock-down group. Figure 4.7 B shows the gene expression distributions of the controls and knock-down. In these control plots there were three major 'peaks' of gene expression; the first, furthest left, represents genes at a low level of expression (FPKM <1), the middle peak represents genes which are moderately expressed (FPKM ~ 100) and finally the small peak shows genes expressed at very high levels (FPKM >5000). Interestingly, the genes expressed at low levels appear to shift right in the knock-down samples indicating an increase in expression of lowly expressed genes. Additionally, there is also an increase in the number of genes that are moderately expressed in the knock-down samples. This pattern is perhaps not surprising given DIS3 is a ribonuclease that normally degrades RNAs.
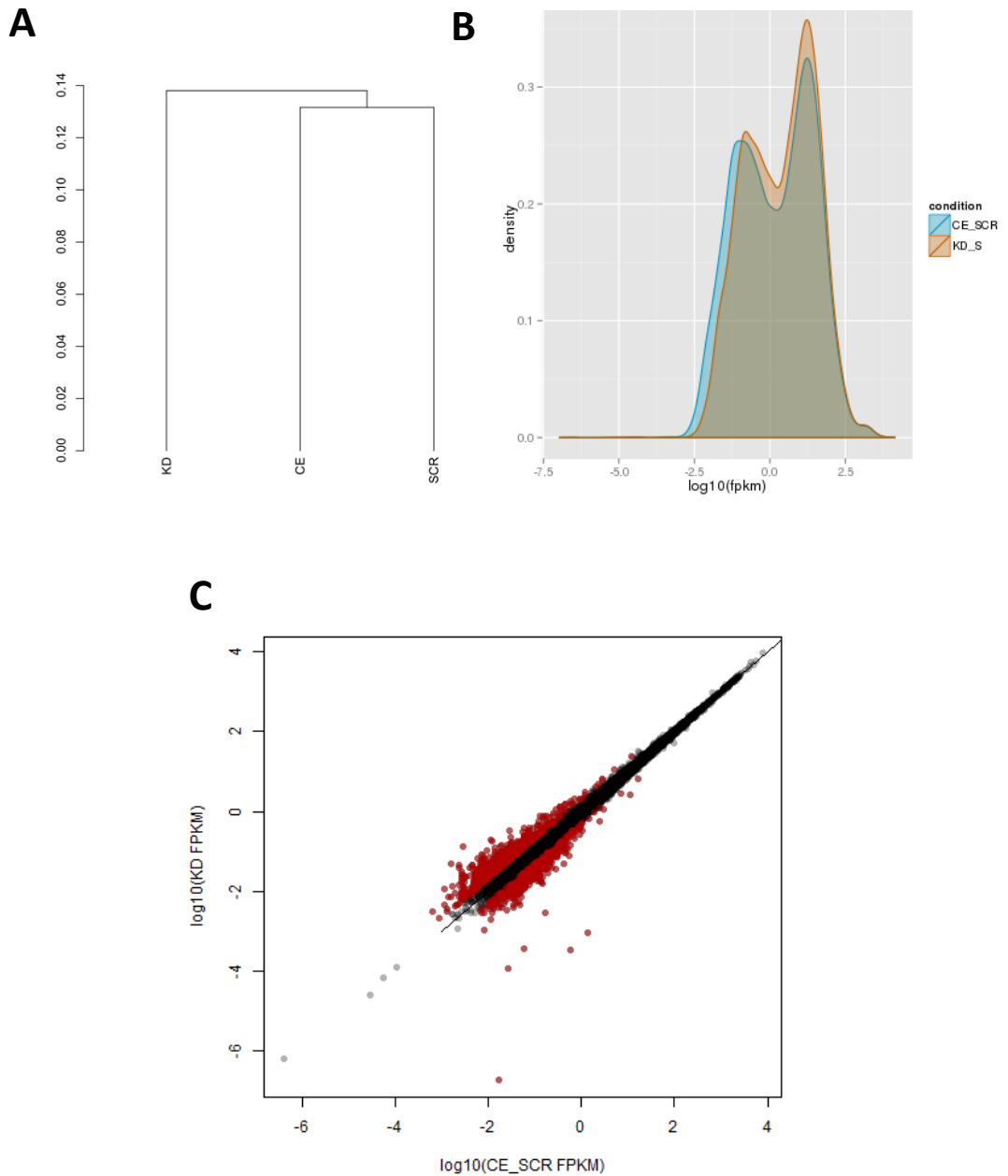
Figure 4.7. Comparison of global gene expression patterns in the DIS3 knock-down and two control groups. (A) Dendrogram showing hierarchal clustering of Jensen-Shannon distances between FPKM values. The conditions are grouped according to how similar they are. (B) Kernel density plot comparing the gene expression profiles of the two groups. The density score represents the frequency of genes falling within log10 (FPKM) levels. CE_SCR = grouped electroporated only and scrambled control, KD = DIS3 knock-down. (C) Scatter plot showing a comparison of gene FPKM values between control and DIS3 knock-down samples. Red dots represent genes with a fold change of ≥ 2 or ≤ -2.

157

In addition to the kernel density plots as a means of viewing overall changes in gene expression, scatter plots were also used. Kernal density plots give an idea of a general trend in expression changes; however, the scatter plots allowed a visual overview of transcripts that both increase and decrease in expression. Figure 4.7 C shows that a number of transcripts appear to be differentially expressed between the control and DIS3 knockdown samples, with a similar number lying above and below the line. Of 18906 genes expressed, 2113 showed a change of more than two-fold (red dots). The majority of the genes showing differential expression were within the lower region of expression (FPKMs <1) which was consistent with the expression distribution plots. Due to the low read coverage, genes with FPKMs of <1 tend to show greater technical variation as they can represent errors in mapping rather than being real reads.

4.8 Identifying differentially expressed genes

Interpretation of the analysis of the RNA-seq data was initially outsourced to Oxford Gene Technology, however due to issues with this, a second analysis was performed by Stewart Stevens. Reassuringly, DIS3 showed an average fold change of -4 fold across the two replicates. Curiously, despite the knock-down levels being consistent when tested by qPCR, in the RNA-seq data, knock-down replicate 1 showed twice the level of DIS3 knock-down compared to replicate 2 (Figure 4.8). In the following section I will discuss the two individual analyses separately.

4.8.1 Oxford Gene Technology Analysis

OGT identified 36 significantly differentially expressed genes between the knock-down and control samples. Upon further inspection however, each gene has multiple transcripts and the fold change given for the entire gene was actually a product of only one of these transcripts changing in expression. Furthermore, with the exception of DIS3, the transcripts that produced this change were all small products labelled by ENSEMBL as either 'retained intron' or 'processed transcript' rather than full length transcripts and therefore may or may not be real. This was

A

ENSG00000083520

B

ENSG00000083520
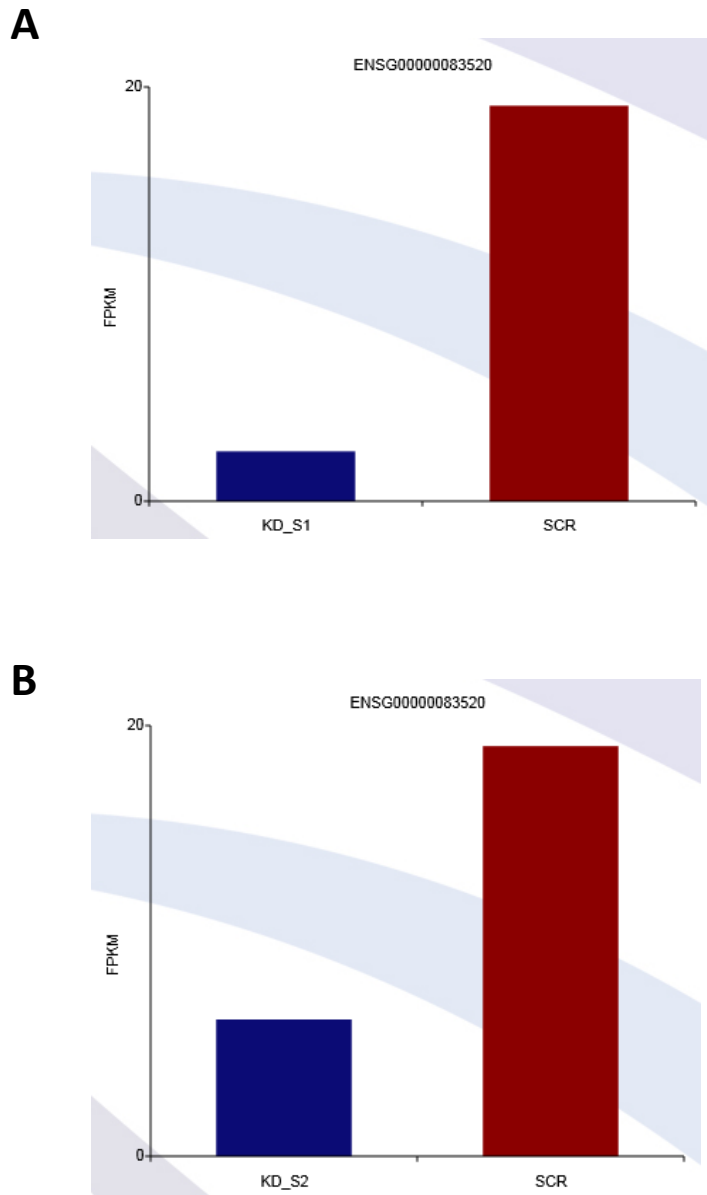
Figure 4.8. **Bar charts showing the FPKM values for DIS3 in the individual two knock-down samples compared to the grouped scrambled controls.** Knock-down replicate 1 showed more reduced levels of DIS3 compared to replicate 2. KD= DIS3 knock-down sample, SCR = scrambled control sample.

due to OGT aligning the reads to an annotation file from ENSEMBL that appears to include artefacts. To get around this problem, the transcripts were assessed individually for differential expression. OGT listed 0 transcripts as being significantly differentially expressed, however due to only using duplicates, this significance testing is relatively meaningless. Therefore, all 212368 transcripts that changed in expression were taken forward in a separate analysis.

A series of filters were applied to generate two lists: transcripts that have a confident positive fold change and transcripts that have a confident negative fold change (Figure 4.9). The filtering process included a number of parameters for the selection of genes that were able to be confidently assigned as differentially expressed. One of these parameters was that transcripts must have an FPKM of ≥ 0.5 in each of the replicates, to reduce the chance of incorporating technical errors. The second was that transcripts must show a fold change of more than 2 or less than -2. That is the levels have at least either doubled or halved in the knock-down compared to control. It was also critical to identify genes that change consistently across all the replicates and were therefore 'real' reproducible changes. Within the control group, transcripts were excluded if the variation between replicates was greater than the average of the two knock-down replicates standardised to the relative fold change. The same was applied to the knock-down samples. In order to filter out those very small, possible artefacts, labelled as 'retained intron' or 'processed transcript', the biotype for each transcript was imported from ENSEMBL and only those classed as 'protein coding' were considered. After the filtering process, 207 upregulated and 190 down-regulated transcripts remained (red dots in Figure 4.10). The 30 most upregulated and downregulated genes can be visualised in the strip plots in Figure 4.11 and Figure 4.12 respectively. The efficiency of the consistency filtering is evident in the clustering of the knockdown (red) and control (blue) replicates with the clear distinction between them.

Figure 4.9. Flow chart displaying the stages of filtering of the RNA-seq data generated by Oxford Gene Technology. 212398 transcripts were initially recorded as differentially expressed. The biotype of each transcript was imported from ENSEMBL. A series of filters was applied to generate two lists: transcripts that have a positive fold change (green sequence) and transcripts that have a negative fold change (red sequence).
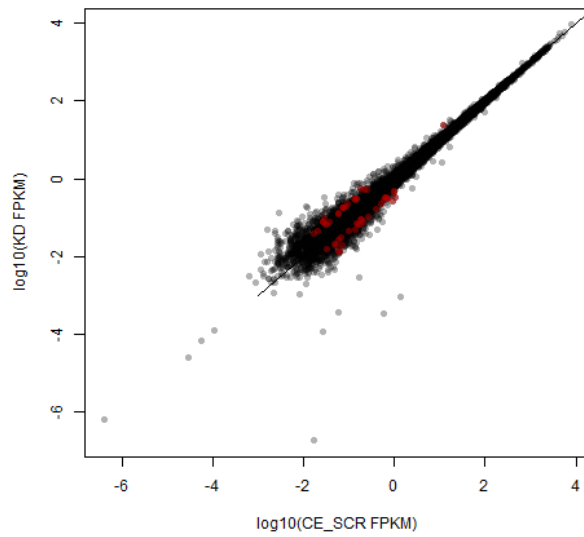
Figure 4.10. Scatter plot showing genes confidently identified as being differentially expressed between control and knock-down samples. Red dots represent genes with a consistent FPKM between replicates and a fold change of ≥ 2 or ≤ -2. CE_SCR = grouped electroporated only and scrambled control, KD = DIS3 knock-down.
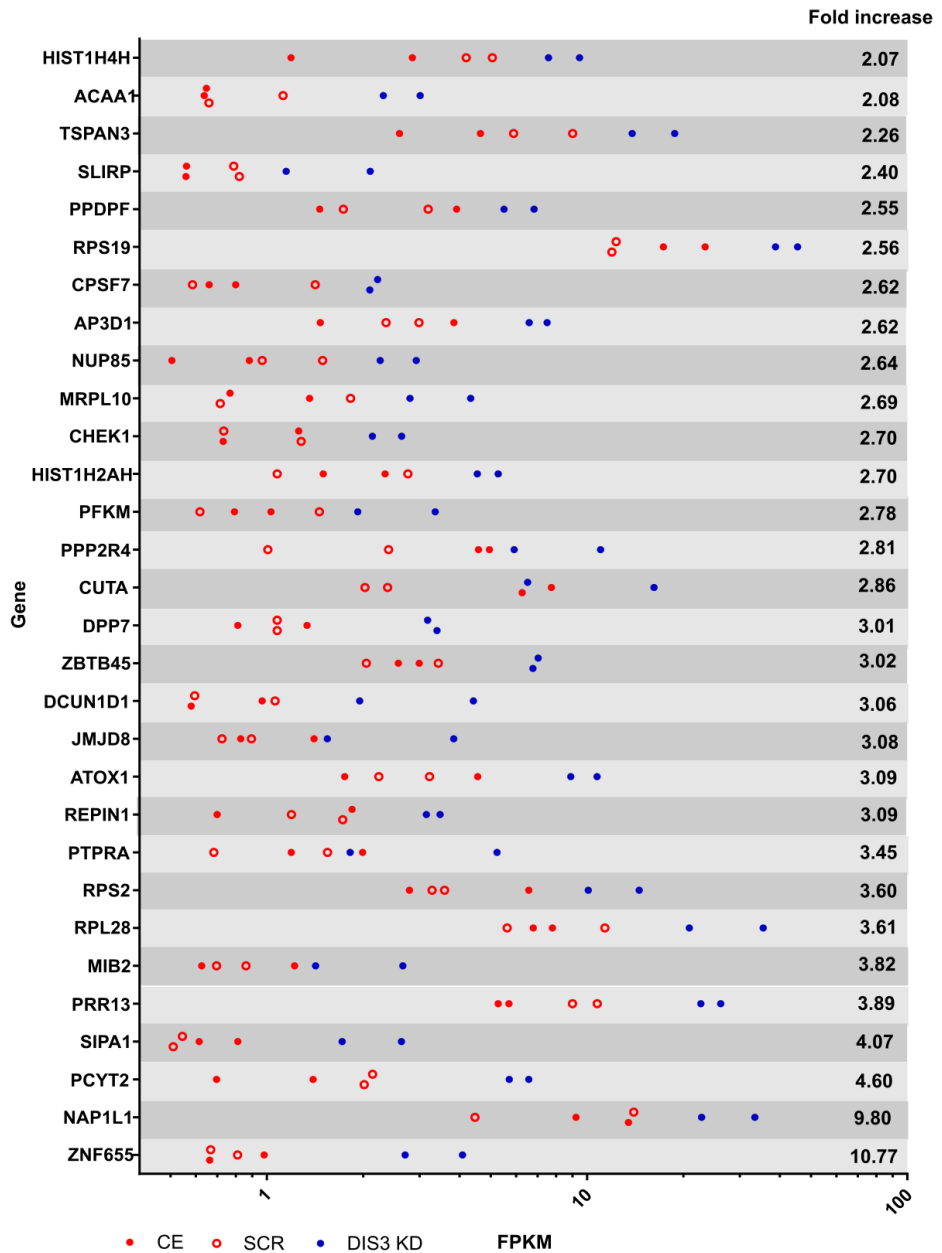
Figure 4.11. Strip plots of the thirty most upregulated genes from OGT's analysis in descending order. FPKMs of the individual replicates are plotted. Red circles represent control electroporated replicates, open circles represent scrambled replicates and blue dots DIS3 KD replicates. The fold change is shown for each gene.
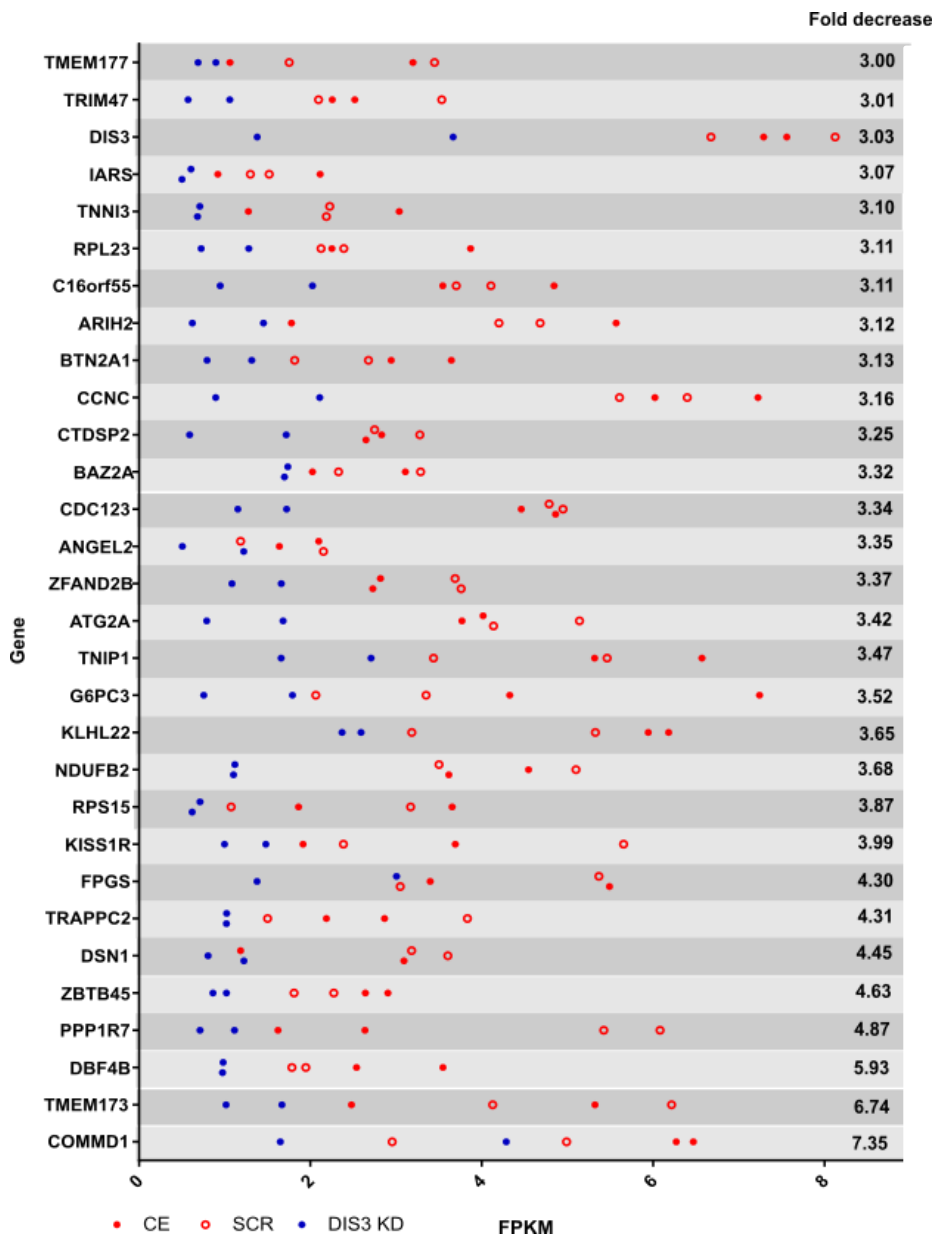
Figure 4.12. Strip plots of the thirty most down-regulated genes from OGT's analysis in descending order. FPKMs of the individual replicates are plotted. Red circles represent control electroporated replicates, open circles represent scrambled replicates and blue dots DIS3 KD replicates. The fold change is shown for each gene.

4.8.2 Stewart's Analysis

Due to the issues with the analysis by OGT, an additional analysis was performed using the same pipeline but modifying a few parameters such as 'frag bias correct' and 'multi read correct (discussed in section 4.6.3) whilst also restricting the analysis to transcripts annotated by the Consensus CDS Project. Restricting the analysis in this way avoided the small transcript annotations that complicated the initial analysis. Before correction for multiple hypothesis testing, 205 genes were identified as having differential expression between control and knock-down samples ($p < 0.01$). As with the OGT analysis, none of the genes showed a significant differential expression after correction for multiple hypothesis testing ($p < 0.05$). Figure 4.13 shows the individual replicate FPKM values for the remaining 21 genes after filtering for FPKM >0.5. There is no overlap between these genes and those identified in the previous analysis as the latter analysis focused on single transcripts whereas the former used whole genes.

4.9 Gene Ontology Analysis

The interpretation of gene expression data is based on the function of individual genes as well as their role in pathways since genes work connectively in all biological processes. In addition, for some genes, a small expression change may not be significant at a single gene level, but minor changes of several genes may be relevant in a pathway and may have dramatic biological consequences. Therefore, after having identified differentially regulated genes from the two analyses, it was of interest to identify cellular pathways that were enriched and thus potentially regulated by DIS3. The 428 differentially regulated transcripts from both analyses that were up- and downregulated ±>2-fold were submitted to the DAVID Gene Functional Classification tool (Huang et al., 2009a). DAVID searches against the background genome for the number of genes which, relative to the size of the input list, would be expected to be identified within each category. An enrichment score
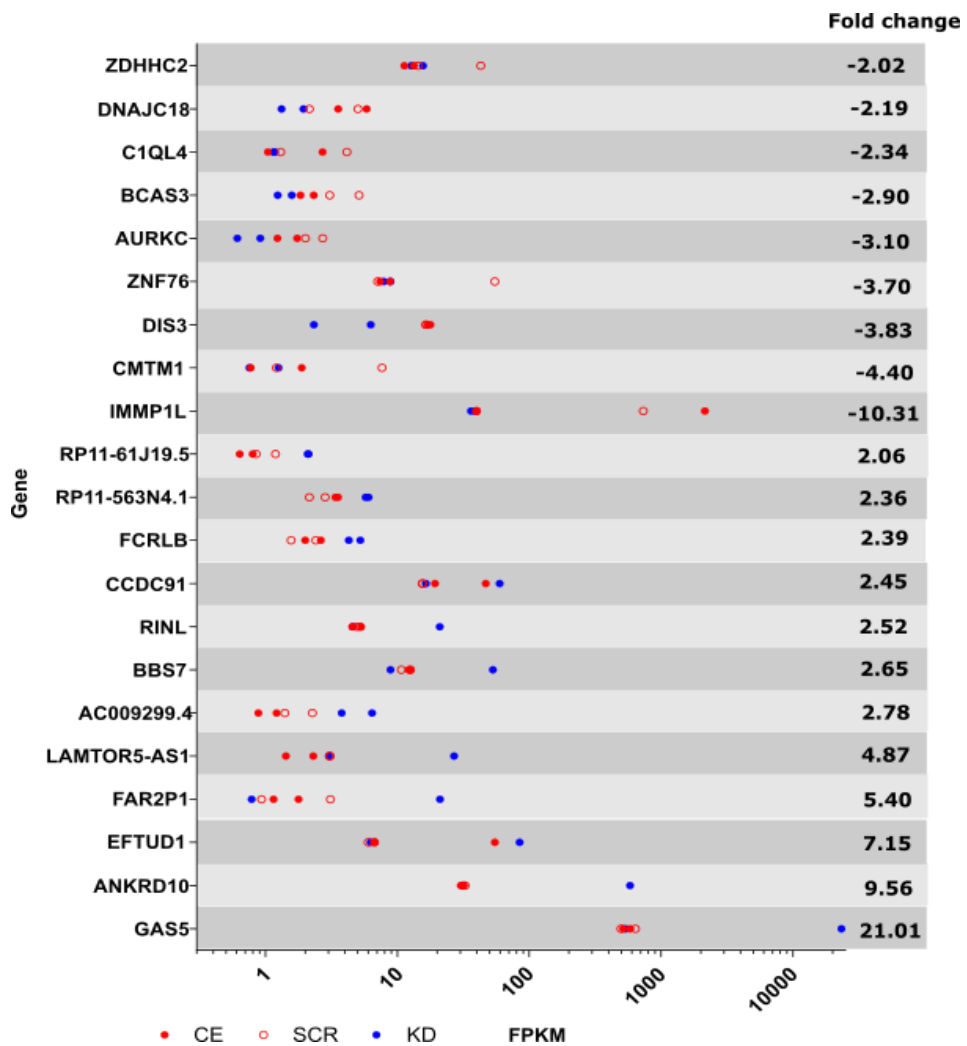
Figure 4.13. Strip plots of the twenty one differentially expressed genes from Stewart's analysis. FPKMs of the individual replicates are plotted. Red circles represent control electroporated replicates, open circles represent scrambled replicates and blue dots DIS3 KD replicates. The fold change is shown for each gene.

is calculated as –log10 (p-value) and any score >1.3 represents a significant enrichment at p < 0.05.

A limitation of DAVID is that it relies on previous annotation of a genes biological function, which for a high proportion of the differentially regulated transcripts was unknown. Additionally, the low number of input transcripts made it more difficult to identify large scale enrichment. However, five significant categories were identified for the upregulated genes and two for the down regulated genes (Tables 4.4 and 4.5). The five groups of upregulated genes were: translation, small ncRNA processing, RNA splicing, chromatin assembly and post-transcriptional gene regulation. The first category, translation, included a few multiple ribosomal proteins and translation elongation factors, however many of the genes within this category were ribosomal protein pseudogenes that are non-functional. Interestingly, in the small ncRNA processing category, two subunits of RNase P, a 3' to 5' exoribonuclease responsible for generating mature tRNA molecules as well as the RNA-binding protein Lsm6 which is thought to be involved in mRNA degradation by activating the decapping step in the 5'-to-3' mRNA decay pathway as well as mRNA splicing, were present. Within the RNA splicing category were splicing factors, hnRNPS that bind to AU-rich elements and a cleavage and polyadenylation factor involved in pre-mRNA 3' processing. The chromatin assembly category included four histone proteins and within the final category was Pumilio, a protein that regulates translation and mRNA stability by binding the 3'-UTR of mRNA targets. Therefore, as many of these genes are involved in RNA processing, the upregulation of these groups of proteins, if not targets of DIS3, may indicate a compensatory mechanism for loss of DIS3.

4.10 Validation of differentially expressed transcripts

In order to validate the differential expression changes observed in the RNA-seq data, Taqman assays needed to be available to confirm the changes by qPCR. A problem that was identified at this stage was that many of the available Taqman assays were not specific to the differentially expressed transcripts, but also to other

alternatively spliced transcripts that did not change in my RNA-seq data. Therefore, if these were used for validation, the fold change would represent that across all alternatively spliced transcripts. For this reason, only genes with Taqman assays available for the specific transcript that changed in the RNA-seq data were included. As DIS3 is a ribonuclease, direct targets are most likely to be those genes that are upregulated rather than down-regulated following DIS3 depletion. Therefore, as the principal interest was on direct targets of DIS3, only upregulated genes were considered for validation.

In order to take an unbiased approach, the five genes with the highest fold increase from the OGT analysis for which there were Taqman assays available, were selected for further validation: ZBTB45, zinc finger and BTB domain-containing protein 45 of unknown function; PPP2R4, protein phosphatase 2A activator, a Ser/Thr phosphatase implicated in the negative control of cell growth and division; as well as three histone proteins that were flagged up in the GO analysis, HIST1H2AH, HIST1H4C and HIST1H4H. From Stewart's analysis the three genes with the largest fold change for which there were Taqman assays available, GAS5, a lncRNA containing multiple C/D box snoRNA genes in its introns; ANKRD10, ankyrin-repeat domain 10 and EFTUD1, an elongation factor involved in the biogenesis of the 60S ribosomal subunit, were also selected for validation. Although there was variability between the two knock-down FPKMs for these genes, all three displayed a large upregulation in one knock-down sample compared to controls and this sample was the one with the greatest reduction in DIS3 according to the RNA-seq data (Figure 4.8), providing rationale for validation. Three further genes were selected, BBS7, a protein believed to promote ciliogenesis; RINL, a guanine nucleotide exchange factor involved in Rab activation and FCRLB, an Fc receptor involved in the immune response, as these were the next mostly highly upregulated genes which showed the least variability between replicates (Figure 4.13).

In order to identify biologically relevant and reproducible hits, RNA was extracted from new samples for the validation. Real changes would be expected to change in a similar manner across independent DIS3 knockdown samples, whereas any non-specific changes caused for example by sample handling, would not be expected to

Term category	Enrichment score (-log10(pvalue))	Gene count (% of input)
Translation	5.51	15 (7.2)
Small ncRNA processing	2.54	11 (5.3)
RNA splicing	2.45	20 (9.6)
Chromatin assembly	2.41	15 (7.2)
Post-transcriptional gene regulation	1.35	7 (3.4)

Table 4.4. Gene ontology analysis of genes upregulated by > 2-fold in DIS3 knock-down cells compared to controls.

Term category	Enrichment score (-log10(pvalue))	Gene count (% of input)
Translation	2.41	11 (5.3)
Protein metabolism	1.31	6 (2.9)

Table 4.5. Gene ontology analysis of genes downregulated by > 2-fold in DIS3 knock-down cells compared to controls.

be consistent in the fresh samples. To mimic the conditions used for the library preparation, oligo-dT primers were used to prime the reverse transcription reaction. Successful knock-down was confirmed in these fresh samples by qPCR and levels were consistent with those in the initial samples sent for sequencing. Eleven genes, five from the initial OGT analysis and six from Stewart's analysis were selected for validation (see previous paragraph). Unfortunately, although all but one of the eleven genes showed a positive fold change, none increased significantly by qPCR in RPMI-8226 myeloma cells (Figure 4.14 A). Figure 4.14 B corroborates this by showing there to be no relationship between RNA-seq fold change and qPCR fold-change. However, although the changes were not statistically significant due to small sample size, two of the histone genes showed an average fold-change of above two and all three technical replicates showed an increase in expression in the knock-down compared to control. qPCR was also performed on RNA derived from the three osteosarcoma-derived stable DIS3 knock-down cells lines (described in chapter 3), however none of the genes significantly changed in these cell lines either (Figure 4.15).
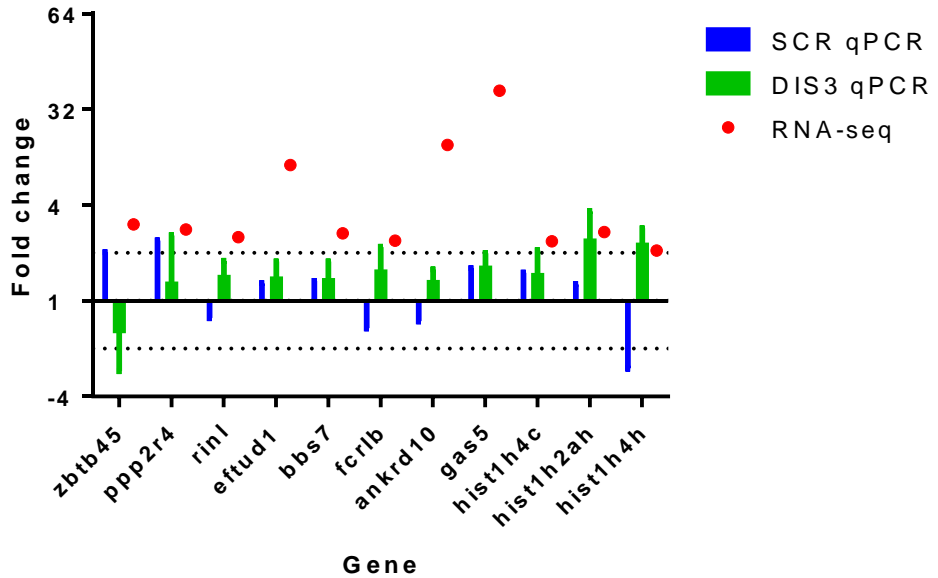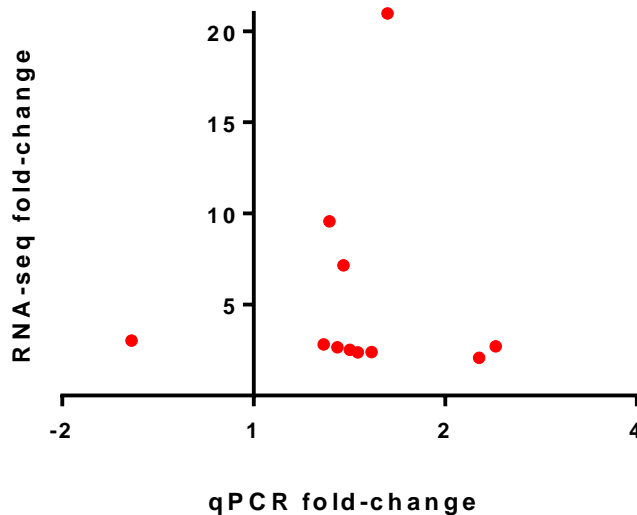
Figure 4.14. qPCR validation of the RNA-seq results in RPMI-8226 myeloma cells.
(A) Graph shows the comparison between the fold changes of transcripts by RNA-seq (red dots) and qPCR (coloured bars). Error bars represent the S.E.M obtained from three independent experiments each made up of three technical replicates. Dotted lines represent a two-fold change. (B) Scatter plot showing no relationship between the RNA-seq and qPCR fold-changes for the 11 tested genes.
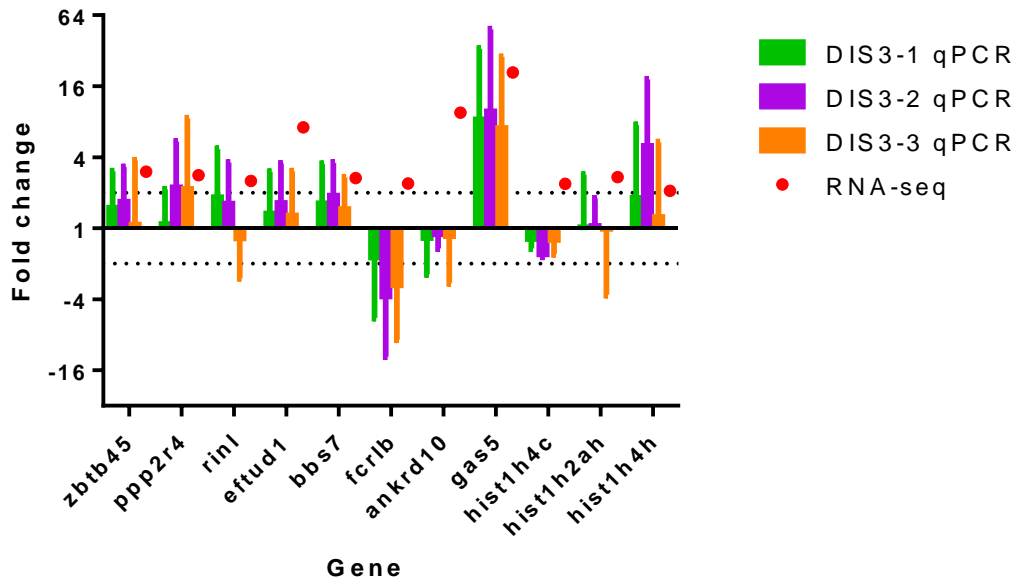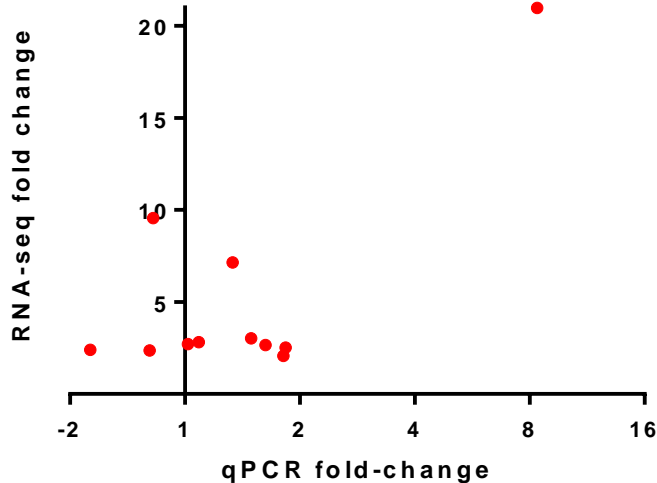
Figure 4.15. qPCR validation of the RNA-seq results in three osteosarcoma-derived stable DIS3 knock-down cells lines. Graph shows the comparison between the fold changes of transcripts by RNA-seq (red dots) and qPCR (coloured bars). Error bars represent the S.E.M obtained from three independent experiments each made up of three technical replicates. The three coloured bars each represent a different U-2OS clone. Dotted lines represent a two-fold change. (B) Scatter plot showing no relationship between the RNA-seq and qPCR fold-changes for the 11 tested genes.

4.11 Discussion

In this chapter I have used RNA-seq as a global, unbiased approach in an attempt to identify genes that are sensitive to DIS3 expression. Two analyses were performed using the Tuxedo protocol of read alignment, quantification and differential expression analysis. Each analysis differed slightly in the exact parameters used, producing two different lists of differentially expressed genes. Eleven of these genes were selected for validation by Taqman qPCR based on fold-change and consistency between replicates. Unfortunately, none of these genes significantly changed as detected by qPCR.

4.11.1 DIS3 may regulate histone proteins

Although the changes were not statistically significant by qPCR due to a small sample size, two of the histone genes showed an average fold-change of above two and all three technical replicates showed an increase in expression in the knock-down compared to control. Therefore, with a larger sample size, the observed upregulation of the histone genes may be revealed as real changes. The enrichment of four histone genes in the gene ontology analysis also supports this. This finding would be interesting as histones are already known to be targeted by exosome (Slevin et al., 2014; Mullen and Marzluff, 2008), however no knowledge exists on whether DIS3 is the nuclease responsible for 3' to 5' degradation. Furthermore, there is a well-established link between histone expression and degradation and cell-cycle progression (Zhao, 2004; Reis and Campbell, 2007). Upon DNA damage, histones are normally downregulated and entry into S-phase is stalled. Disruption of the mechanisms that maintain precise histone levels leads to chromosomal instability. As DIS3 mutants are well known to display cell-cycle defects and as a malignancy characterised by chromosomal instability, it is feasible to speculate that disruption of DIS3 may cause misregulation of histone proteins which may contribute to multiple myeloma development.

4.11.2 Limitations of the experimental design and analysis

RNA-seq is a revolutionary tool for high-throughput, high-sensitivity, and high-speed transcriptome analysis. As a fairly new technology however, the large scale data analyses associated with RNA-seq still harbour some challenges. In this section I will discuss some of the issues and limitations encountered in this particular project that may be able to be overcome with future advancements.

4.11.2.1 Limitations of the analysis by Oxford Gene Technology

In the initial analysis performed by OGT, only 36 genes were identified as significantly differentially expressed between control and knock-down samples. The statistical analysis supplied was based on that performed by CuffDiff, the algorithm used to determine transcript differential expression. Differential expression is determined as statistically significant by CuffDiff if the p-value is lower that the false discovery rate of 0.05. This would normally be a useful tool to drive the selection of consistently differentially expressed genes, however, despite the control samples being grouped to give an n of 4 there were only duplicates of the knock-down samples. A low n number minimises the statistical power and renders it almost useless. Of the 36 genes identified as differentially expressed, the vast majority showed large variations between the replicates with many being heavily dominated by a single replicate. In addition, due to the reference genome used for mapping, it later emerged that all these changes were due to probable artefacts labelled as small 'processed transcripts' so did not represent the gene as a whole. This therefore removed any confidence in the analysis performed by OGT and was thus not used in the downstream analysis. Due to cost limitations, often the experimental design is a trade-off between sequencing depth and biological replicates which is why in this particular experiment, duplicates were chosen over triplicates. However, any future experiments of a similar nature should use a larger sample size such that the CuffDiff statistical analysis can be made full use of.

After abandoning the analysis by OGT, I performed my own analysis of the data. Due to the small sample size, stringent criteria had to be set in order to have the upmost confidence in the determination of differentially expressed genes. The filtering criteria included FPKM, consistency between replicates, fold-change and biotype. Following the filtering process, the number of differentially expressed transcripts (those showing a >2-fold change) reduced from 3625 to 397. A reduction this large shows the high level of potential false positives that are often present in global data sets and emphasises the importance of stringent filtering. Nevertheless, a trade-off is made between discarding false-positives and discarding true changes and it is likely that a number of transcripts were excluded due to over stringent cut offs. For instance, excluding genes that showed a difference between the replicates will likely have resulted in a number of true changes being missed. If triplicates were used, it would have been easier to identify inconsistencies and fewer genes would have been mistaken for false-positives. Additionally, differentially expressed transcripts with an FPKM of <0.5 will also have been missed due to the filtering criteria, especially as according to Figure 4.7 B, these transcripts contributed to most of the expression changes. However, low FPKMs introduce reliability issues as genes are known to show increased variability and are therefore less likely to validate by qPCR. Additionally, those genes showing very low expression are difficult to detect at a reliable cycle threshold when using qPCR.

One question that may arise may concern the reason for such differences between replicates. Although measures were taken to ensure the conditions under which the samples were collected were as consistent as possible, the variation may be due to sample treatment during the transfection or RNA extraction. For example, one sample may have been subjected to more environmental stress than the other. Alternatively, it could be caused by natural variation and therefore represent real biological differences that we are correct in excluding from the analysis. Furthermore, variation between the knock-down replicates may represent the different extent to which DIS3 is knocked-down in each sample. Although the qPCR

and Western blot that were carried out before the samples were sent away showed a consistent knock-down, the RNA-seq data suggested otherwise, with one knock-down sample showing half the levels of DIS3 than the other (Figure 4.8).

4.11.3 Reasons for a lack of validation of differentially expressed genes

Although most of the transcripts selected for validation by qPCR showed a positive fold change, none showed a statistically significant change in expression between the control and knock-down samples. The potential reasons for this are discussed below.

RNA-seq has been praised as a method that without the dependency on specifically designed microarray probes, can quantify the levels of splice variants. However, the field of alternative splicing analysis using RNA-seq data is still in its infancy and detection of alternative splicing does not reach the same accuracy as the detection of differential expression (Rehrauer et al., 2013). Estimating the relative abundance of transcript variants is a challenging task. Many of the reads that map to a gene will map to shared exons, complicating the process of counting reads for each transcript. A read from a shared exon could have come from one or several isoforms. In order to assign reads to a particular isoform, CuffLinks implements a statistical model to estimate an assignment of abundance to each transcript, that explains the observed reads with maximum likelihood (Trapnell et al., 2012). Results are thus an estimate of isoform abundance and maximum likelihood estimates, especially for low-abundance transcripts, can be inaccurate (Bryant et al., 2012). This creates a dilemma, as in my RNA-seq data, the fold-change listed for a gene was calculated by the addition of all its transcripts, however only certain transcripts within a gene were differentially expressed. Other transcripts belonging to the same gene showed no apparent change. Therefore, as the transcripts belonging to an individual gene often changed in different directions, I decided to base the analysis on transcript changes rather than gene changes as the latter was unlikely to validate by qPCR. Thus, if the transcript abundances were not accurately calculated by CuffDiff, this could provide an explanation for why the differentially

expressed transcripts according to RNA-seq, did not change by qPCR. In future, to make for a simpler analysis, it may be worth focusing on differentially expressed genes whose transcripts all change in the same direction.

An alternative explanation for why the transcripts did not validate may be that the observed expression levels in the RNA-seq are due to natural variation in this cell line. With only duplicates to work with, there is a higher chance that the observed levels could be due to chance rather than a real effect of DIS3 knock-down. Furthermore, genes that appear differentially regulated either due to biological or technical variation will likely produce the highest fold-changes. The lack of validation may reflect the fact that the selection of genes was based on those with the highest fold-change. Therefore, in an attempt to select genes in an unbiased manner, on the contrary, a higher fold-change probably biases towards false-positives.

An alternative explanation is that the lack of validation may be attributable to the insufficient precision of qPCR to detect low fold-changes using a small sample size. This is feasible as the majority of the selected transcripts only showed a fold-change of around 2. The fact that all but one of the genes changed in the correct direction also supports this explanation. Using a larger sample size for the qPCR may get around this issue, however there is a cost limitation attached to performing multiple transfections by electroporation.

Finally, as the number of transcripts that passed the filtering process was low and none of the selected genes validated by qPCR this may suggest that DIS3 is responsible only for the regulation of a discrete set of mRNAs. In a similar study, only a small number of transcripts were identified as targets for XRN1, a 5' to 3' exoribonuclease, in *Drosophila (*Jones et al., 2016), possibly reflecting a general mechanism of transcript specific regulation by exoribonucleases. In addition, DIS3 is predominantly nuclear localised and known to target faulty mRNAs such as un-spliced pre-mRNAs (Bousquet-Antonelli et al., 2000) and mRNAs with defective polyadenylation (Milligan et al., 2005). However, due to using poly-A selection during library preparation this experiment focused on polyadenylated mRNAs only.

Therefore, changes to the aforementioned classes of RNA would not be detected by this experimental set-up, however their misexpression may well explain the contribution of DIS3 mutations to the development of multiple myeloma.

4.11.4 Final conclusions

This chapter has highlighted how RNA-sequencing allows for unbiased, high throughput, sensitive quantification of gene expression whilst at the same time, showing the importance of experimental design and interpretation of expression changes. The many small 'processed transcripts' that accounted for the supposed gene expression changes in the initial analysis did not reflect the real change in gene expression levels. Furthermore, upon performing my own analysis, many transcripts changed but other transcripts belonging to the same gene did not. Combined with the variability between biological replicates, this demonstrates how RNA-seq data needs to be interpreted with caution to ensure false-positives are not mistaken for real expression changes. The observed upregulation of histone transcripts may provide an insight into the mechanism by which DIS3 contributes to myeloma and this would be interesting to follow up in future work.

Chapter 5: Characterising DIS3 synthetic lethal interactions

5.1 Introduction

Although cancer survival rates have continued to improve over recent years, current chemotherapies are far from ideal. Standard chemotherapies were initially discovered on the basis of their ability to kill rapidly dividing cells, and thus some of their common side effects such as hair loss, nausea and immunosuppression, are due to the toxicity to rapidly dividing normal tissues. Consequently, the bottleneck to the development of safe and effective anticancer drugs does not lie in an inability to identify chemicals that will kill cancer cells, but in our inability to identify chemicals that will kill cancer cells but will not harm normal cells. With the aim of identifying therapies that have greater effectiveness and fewer side effects, cancer research is now largely focused on discovering tumour-specific traits that might be exploited for selective targeting.

An anti-cancer drug discovery strategy that is becoming popular to target cancer cells, is the use of synthetic lethality. This approach takes its name from classical genetic studies that aimed to identify genes that cooperate in the same essential process. Initially performed in model organisms such as yeast, when either gene is mutated alone, the cell is viable, however if both genes are mutated, lethality results (Figure 5.1). Synthetic lethal genetic interactions exist because of the mechanisms employed by cells to maintain homeostasis in the face of mutations and environmental challenges (Hartman et al., 2001; Masel and Siegal, 2009). Cells establish a robust buffering system to ensure that processes do not depend on any single component, predominantly by establishing a level of functional redundancy between genes (Nijman, 2011). Redundancy can be seen between paralogous genes that can still partially perform the same task but is more commonly observed between non-homologous genes operating in the same cellular process. This redundancy can be exploited for therapeutic advantage using
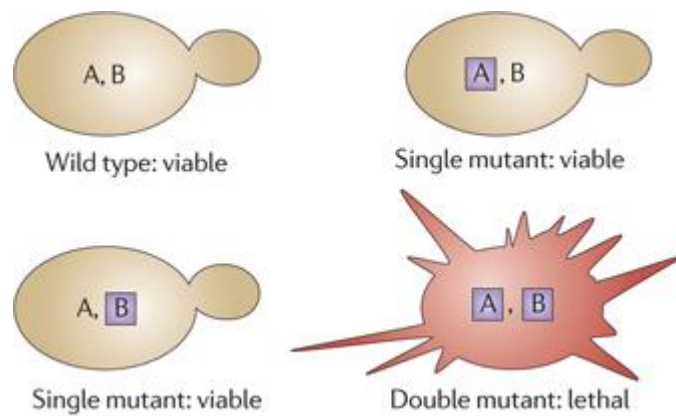
Figure 5.1 Synthetic Lethality. In model organisms, synthetic lethality describes the genetic interaction between two genes. If either gene is mutated by itself, the organism remains viable. The combination of a mutation in both genes is incompatible with viability and results in lethality. Taken from *Chan et al, 2011*.

the concept of synthetic lethality. Genes that demonstrate a redundancy with tumour-mutated genes, would constitute putative 'secondary drug targets' whose inactivation would kill cancer cells, whilst leaving healthy cells unaffected.

The best known example of a cancer therapy based on synthetic lethality, currently being tested in clinical trials, is the PARP1 and BRCA1/2 pair (Farmer et al., 2005; Bryant et al., 2005; Fong et al., 2010; Fong et al., 2009). Mutations in the breast cancer early onset (BRCA1 and BRCA2) genes, normally involved in homologous recombination (HR), occur frequently in breast and ovarian cancers. Inhibition of Poly (ADP-ribose) Polymerase (PARP), responsible for single-strand break repair, leads to synthetic lethality with loss of BRCA1/2 as breaks cannot be repaired by HR. In this way, only cells with BRCA1/2 mutations are susceptible to treatment with a PARP inhibitor, resulting in remarkably mild side-effects. Exploiting synthetic lethality therefore increases selectivity towards killing tumour cells, as well as increasing the dose of the drug required for toxic effects. Moreover, as tumour progression is a multi-step process and the driving mutations change in the different stages of tumour growth, synthetic lethality could target a range of temporal mutations that occur along the pathway of tumorigenesis.

As a result of the increased availability of chemical and genetic tools for perturbing gene function in somatic cells, screens for synthetic lethality can be carried out using libraries of chemical compounds and chemically synthesised small interfering RNA (siRNA). Isogenic cell lines that differ by only one essential cancer gene are labelled and mixed together before being treated with a library of RNAi or small-molecule compounds (Chan and Giaccia, 2011). Some siRNAs or compounds would not be toxic to either cell type; some would be selectively toxic to normal cells or to tumour cells; and some would be toxic to both genotypes (Figure 5.2). In small-molecule screens the result would be the obtainment of candidate compounds for treatment of a given cancer genotype. In RNAi screens the result would be two-fold: revealing unexpected connections that can advance drug development efforts as well as enhancing our understanding of the fundamental biology behind interactions within cancer cells.
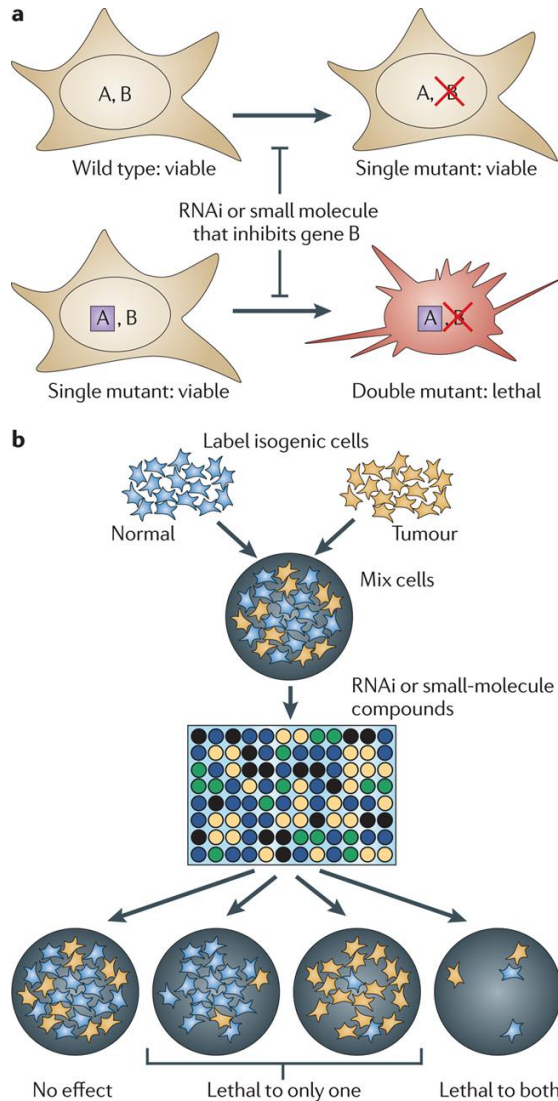
Figure 5.2. Mammalian synthetic lethality screens for anticancer efficacy.
(A) Synthetic lethal screens can be used to identify genes or small-molecule compounds to specifically target tumour cells while sparing the normal tissue. A mutation in the first gene is essential to the development of cancer (for example, a loss-of-function mutation in a tumour suppressor gene or a gain-of-function mutation in an oncogene). The second gene would be identified either through an RNA interference (RNAi) library or it would directly be inhibited by a small-molecule compound. Inhibition of this second gene through RNAi or a small molecule alone would not interfere with tumour growth. Nonetheless, inhibiting the second gene in a tumour of a given genotype would result in selective cytotoxicity of the tumour. (B) Isogenic cells that differ by only one essential cancer gene could be fluorescently tagged and mixed together in equal numbers. The cells would then be added to a 96-well plate and treated with a compound/siRNA library. Fluorescence would be read over several days. Some compounds/siRNA would not be toxic to either cell type; some would be selectively toxic to normal cells or selectively toxic to tumour cells; and some would be toxic to both genotypes. Taken from *Chan et al, 2011*.

As multiple myeloma is currently for the most part, treated as a single entity, identifying synthetic lethal interactions with DIS3 has the potential to identify genes that could be used as tumour-specific targets in DIS3-mutated patients. In addition, as the role of DIS3 beyond its function in RNA degradation and processing is unknown, synthetic lethality screens may reveal unexpected interactions that may advance our understanding of how DIS3 loss-of-function contributes to multiple myeloma pathogenesis. For this reason, this chapter aims to use siRNA screens to identify potentially synthetic lethal interactions within DIS3 knock-down cells, with the aim of identifying genes that could be used as tumour-specific targets for therapeutic treatments of myeloma patients.

5.2 Aims

The central focus of this chapter is to use siRNA libraries to identify synthetic lethal interactions with DIS3, leading to the following aims:

1. Generate nuclear-fluorescently labelled DIS3-knock down and control cell lines
2. Use a commercially available DNA-Damage Response siRNA library to perform a synthetic lethality screen on DIS3 knock-down cells
3. Validate any potential hits by performing individual RNAi experiments and survival assays

5.3 Generating nuclear-labelled isogenic cell lines

Before performing the synthetic lethality screen, isogenic cell lines that differ only by knock-down of DIS3 needed to be generated. To do this, a stable DIS3 knock-down and control cell line were generated, similar to that in chapter 3. Three shRNA clone vectors were purchased pre-cloned into a pGIPZ vector, with at least one guaranteed to induce silencing, each targeting a different region of the DIS3 transcript. The pGIPZ vectors contain all the elements necessary for stable introduction and into the genome as well as visualisation and selection of positive

clones (Figure 5.3). The shRNA hairpin consists of a 22nt dsRNA stem, complementary to the DIS3 transcript and a 19nt loop from miR-30 as well as 125nt of miR-30 flanking sequence on either side of the hairpin, in order to increase Drosha and Dicer processing efficiency. Puromycin resistance gene allows for the selection of stable cell lines and the long-terminal repeats (LTRs) facilitate insertion of the vector into the genome.

In order for control and knock-down cells to be distinguished at the end stage of the screening process, each needed to be labelled with a different fluorescent tag. The original pGIPZ vector contains TurboGFP for the visual marking of positive cells and without a nuclear localisation signal (NLS), labels the cytoplasm of cells. In order for the imaging program to distinguish individual cells during image acquisition, cells needed to be nuclear-labelled. Therefore, the CMV:turboGFP region from the pGIPZ vector was replaced with either a CMV:AcGFP:NLS for the DIS3 shRNA-containing vectors or CMV:mCherry:NLS for the non-targeting shRNA-containing vectors, both containing nuclear localisation signals. The endonucleases XhoI and NotI were used to excise the CMV:turboGFP region from pGIPZ (Figure 5.3B). The AcGFP-NLS and mCherry-NLS fragments (generated by Grant McGregor) which contain a SpeI cut site between the promoter and fluorophore, were then ligated with the linearised pGIPZ vectors to create fluorescent tagged vectors with nuclear localisation signals. A further diagnostic restriction digest was performed with SpeI and NotI to test for successful ligation (data not shown). This process resulted in the generation of nuclear RFP-tagged non-targeting shRNA vectors and nuclear GFP-tagged DIS3 shRNA vectors for use in the screen.

U-2OS, an osteosarcoma cell line, was selected as an easily-transfectable, adherent cell line in which to create a stable DIS3 knock-down model using these vectors. Once the turboGFP had been replaced with a nuclear fluorescent tag, each of the three GFP-DIS3 shRNA and the RFP-non-targeting control clones were transfected into U-2OS cells. Puromycin was added to the cells after 48 hours and cells were left to grow to confluency for at least 4 weeks before knock-down was assessed by qPCR (Figure 5.4A) and Western blotting (Figure 5.4B). Efficient knock-down was
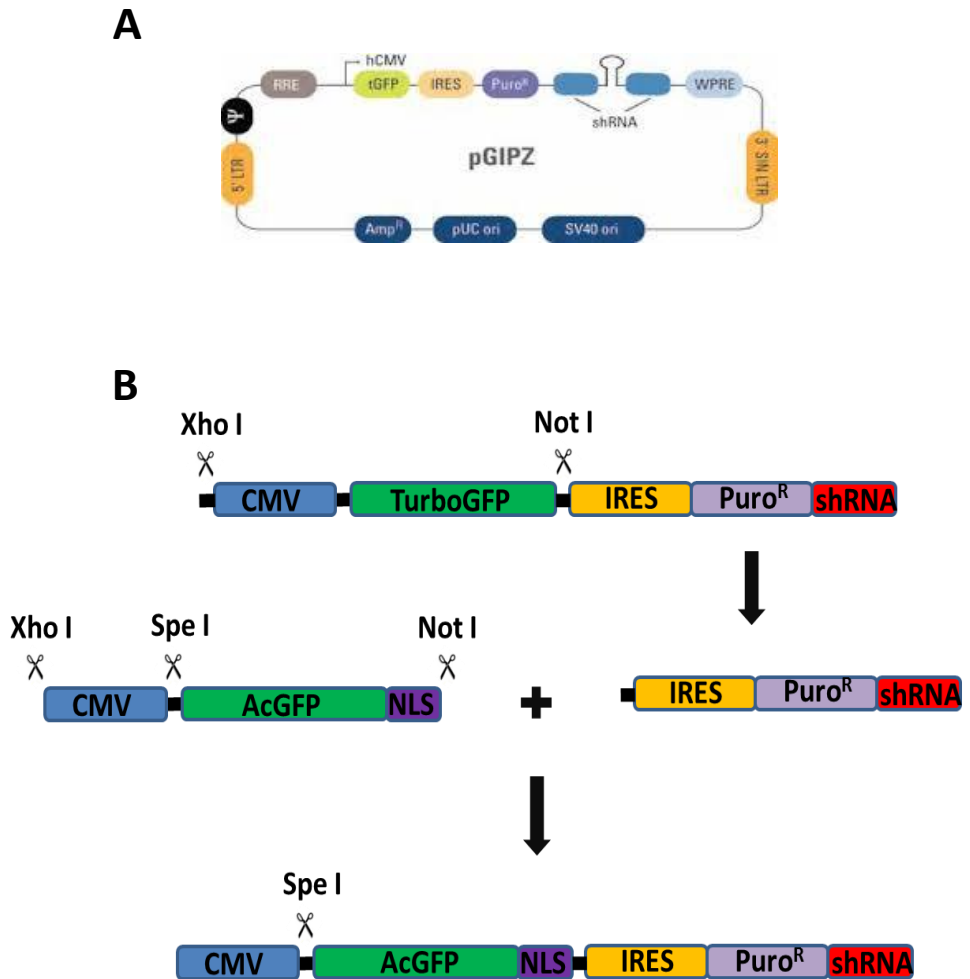
Figure 5.3 Generating a vector for the stable knock-down of DIS3 for use in synthetic lethality screens. (A) Features of the pGIPZ vector (Dharmacon). Important elements include the hCMV promoter to drive transgene expression, tGFP reporter for visual tracking of shRNA expression, 5' and 3' LTRs for genome integration and Puro^R to permit selection and propagation of stable integrants. Other elements include: IRES allowing expression of TurboGFP and puromycin resistance genes in a single transcript; Ψ - Psi packaging sequence allows viral genome packaging (not applicable); RRE - ev response element enhances titer by increasing packaging efficiency of full-length viral genomes (not applicable); WPRE -Woodchuck hepatitis post-transcriptional regulatory element enhances transgene expression in the target cells; SV40 ori - allows propagation of plasmids within mammalian cells expressing the SV40 Large T antigen and pUC ori – to allow replication initiation in bacteria. (B) Generation of a nuclear localised fluorescent signal. The CMV:turboGFP region was excised from the pGIPZ vector using XhoI and NotI and replaced with a CMV:AcGFP fragment with a nuclear localisation signal (NLS) and SpeI cut site. The same procedure was performed to create a red fluorescent non-targeting control vector by replacing CMV:turboGFP with mCherry-NLS.
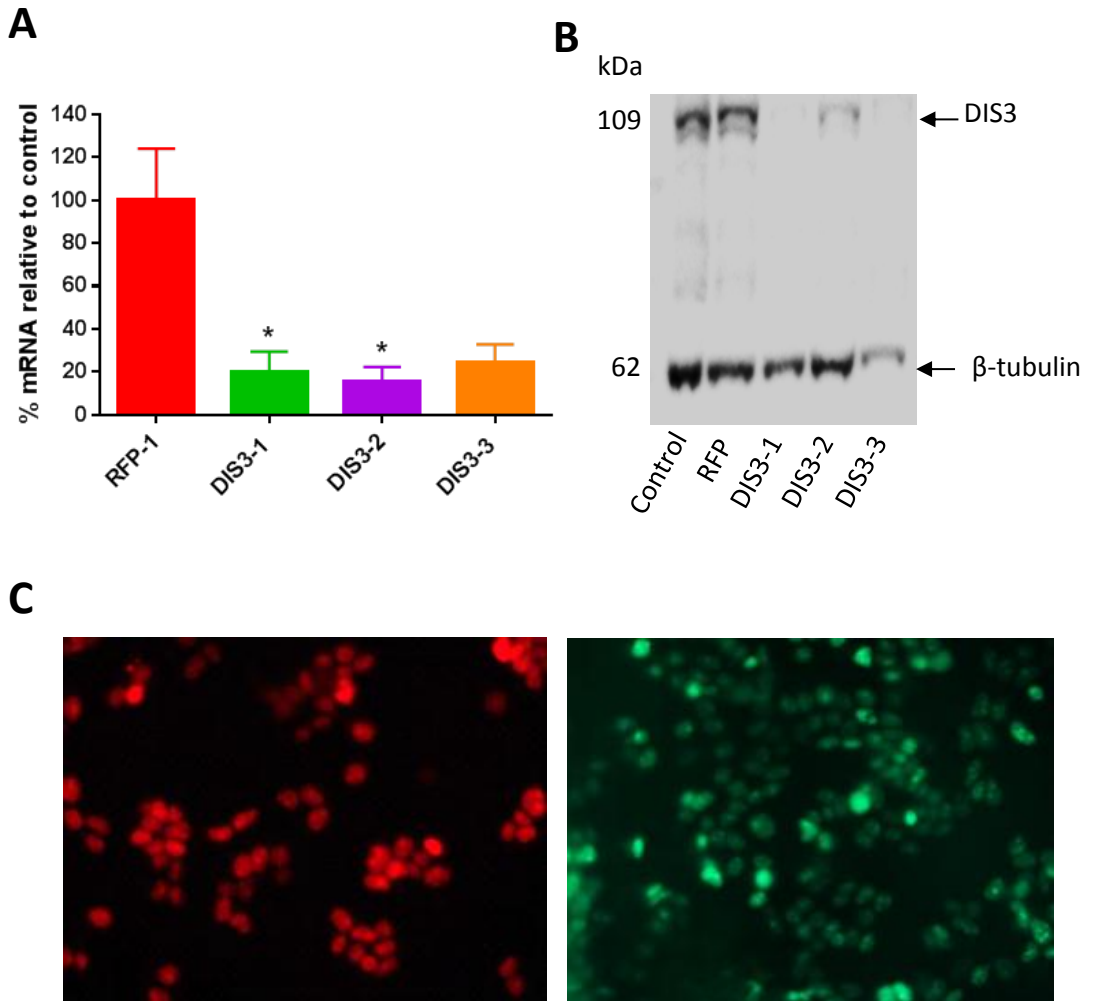
Figure 5.4. Generating a stable DIS3 knock-down model using three DIS3 targeting shRNA vectors and a non-targeting control in U-2OS cells. (A) qPCR showing efficient knock-down of DIS3 with all three shRNA clones when compared to a scrambled control (paired t-tests, $p < 0.05$). Errors bars represent the SEM from four independent experiments. (B) Western blot confirming knock-down of DIS3 at the protein level in the three clones, when compared to untransfected and scrambled controls. (C) Epifluorescence images showing the nuclear localisation of mCherry in non-targeting control U-2OS cells and AcGFP in DIS3 knock-down cells.

achieved of both the mRNA and protein with all three shRNA clones. Figure 5.4C shows the nuclear localisation of fluorescence in the control (RFP) and DIS3 knock-down (GFP) cell lines. As DIS3 clone 1 showed the largest fold change, this cell line was taken forward for use in the synthetic lethality screen.

5.4 Generating preliminary data for the synthetic lethality screen

Before the screen was carried out, some preliminary testing was performed. Firstly, the presence of two populations of cells, RFP-labelled DIS3 wild-type and GFP-labelled DIS3 knock-down cells, was confirmed by immunolabeling with anti-DIS3. As there were differences in the GFP intensity in the immunolabeled cells, FACS was used to assess each cell individually. The plots in Figure 5.5A demonstrate two populations, one of RFP-labelled cells which have a normal level of DIS3 (Cy5 labelled) and one of GFP-labelled cells which have a lower level of DIS3, representing control and knock-down cells respectively. This data corroborates the western blot in Figure 5.4. Interestingly, no relationship exists between the level of DIS3 and GFP fluorescence intensity ($r=-0.136$, $p<0.001$, Figure 5.5B). Given that the vector can insert into both heterochromatin and euchromatin regions of the genome and that shRNA and GFP are expressed from the same transcript, we may expect more of an inverse relationship. However, this data suggests that the DIS3 shRNA is at maximum function at all observable doses. This finding is reassuring as it means that in all GFP-expressing cells, DIS3 is consistently knocked-down to a high level and the fold changes seen in Figure 5.4 are representative of all cells, rather than representing an average. Therefore, a synthetic lethal interaction will be more detectable than if the GFP cell population consisted of a range of DIS3 expression levels.

As the two groups of cells will be mixed together and co-cultured for the screen, it is important to confirm that one does not have a drastically different growth rate to the other. If this was the case, after the 3-day post-transfection incubation period, there may be very few cells belonging to one of the groups, leaving a very small sample size and making the analysis difficult. In order to test this, GFP and RFP cells
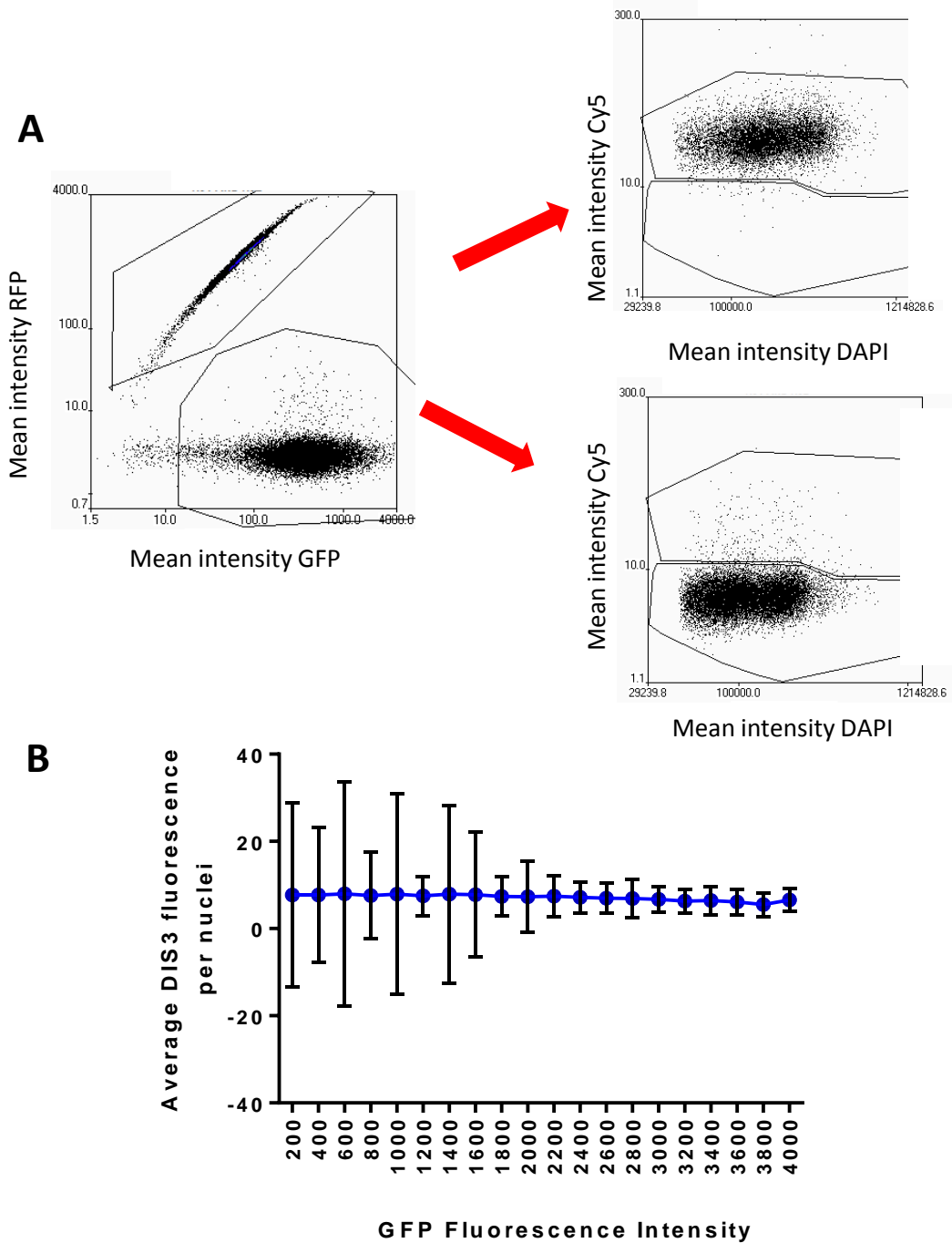
Figure 5.5. Preliminary data showing the correlation between fluorescence and DIS3 expression. (A) Dot plots illustrating two populations of cells corresponding to mCherry-labelled wild-type cells and GFP-labelled DIS3 knock-down cells. Individual cells were imaged, analysed and plotted according to their signal intensities in the fluorescence channels. Cy5 = DIS3. Gates are placed around positive fluorescently labelled cells. (B) GFP levels show a linear relationship with DIS3 levels in cells. Immunofluorescence imaging of 160 000 GFP cells. DIS3 signals were averaged according to bins of GFP intensity. Each data point represents at least 600 cells and error bars represent standard deviation.

187

were mixed according to a range of ratios and the GFP intensity measured after 3 days (Figure 5.6A). Consistent with the findings in Chapter 3, the two cell lines grew at approximately equal rates (Figure 5.7) meaning they can be seeded at a 50:50 ratio for the screen.

In order to test the transfection efficiency, differing concentrations of siGFP were transfected into knock-down and control cells. The bar graph in Figure 5.6B shows efficient reduction in GFP intensity with increasing concentrations of siGFP up to 5nM. Figure 5.6C shows the distribution of cells at different GFP intensities, showing two populations – one with low GFP intensity and one with higher GFP intensity representing siGFP-transfected and control cells respectively. As a negative control, Figure 5.6D demonstrates that siGFP transfection has no effect on RFP intensity.

5.5 A DNA-damage response screen indicates NABP1 may confer synthetic lethality when knocked-down alongside DIS3

As discussed in chapter 1, the exosome has been implicated in recruiting activation-induced cytidine deaminase (AID) to chromatin in B-cells (Basu et al., 2011), where DIS3 may be functioning specifically in degrading nascent RNA. AID creates mutations in the DNA that can lead to double-strand break formation, an important process in antibody class switching. As myeloma often results from translocations formed from these double strand breaks, if it was discovered that DIS3 had an involvement in this process, this may help uncover the role of DIS3 in myelomagenesis. Given that genomic instability and defective DNA repair are hallmarks of cancer and perturbing these processes further is likely to lead to toxicity, this provides rationale for using a DNA damage response (DDR) siRNA library for the first synthetic lethality screen. The DDR siRNA library consists of 240 siRNAs that target genes involved in damage surveillance, damage recognition and signalling, signal effector genes, repair genes, and genes involved in dissociation and resolution of DNA repair complexes. Each well of a 96-well plate consists of a
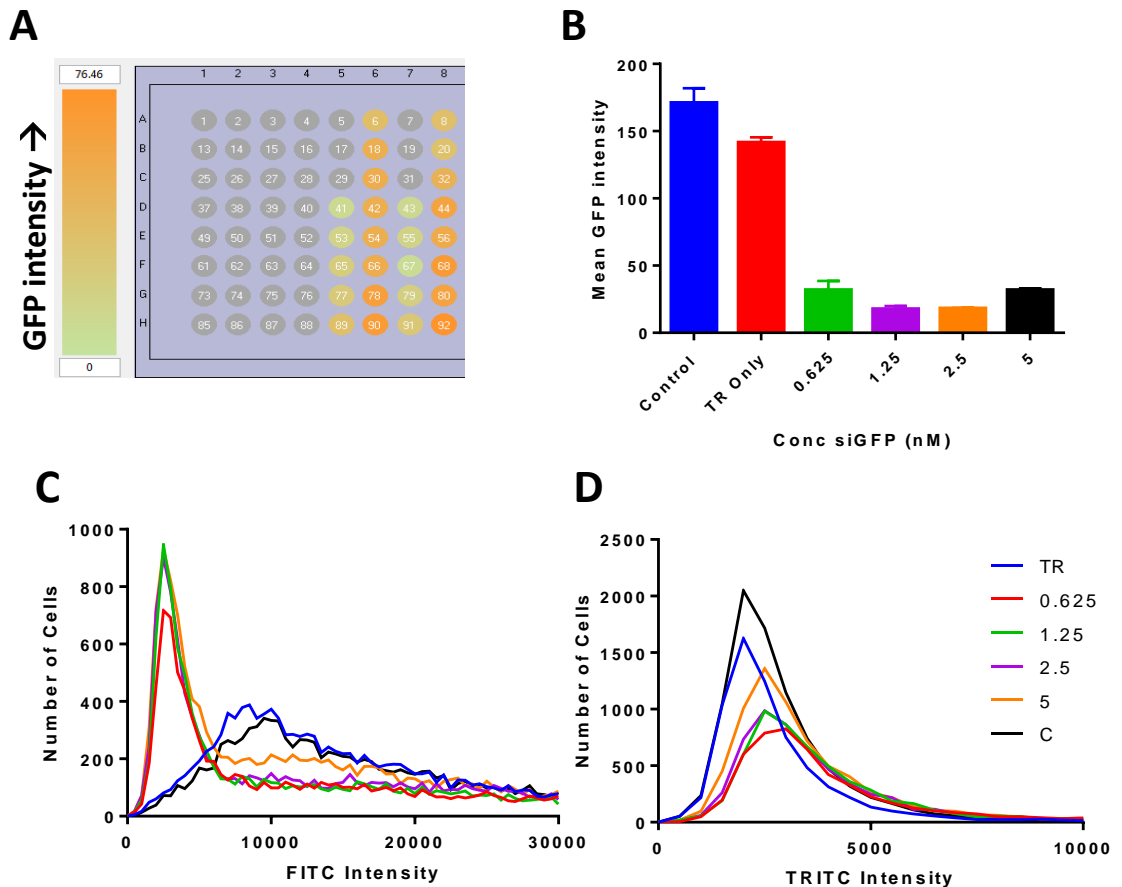
Figure 5.6. Preliminary growth rate and transfection efficiency experiments performed before the DNA-damage response synthetic lethality screen. (A) Schematic of an experiment set up in a 96-well plate of cells at a range of RFP:GFP ratios to determine a seeding ratio for the screen. (B) Bar chart illustrating GFP knock-down using a range of siGFP concentrations, to test transfection efficiency. Mean GFP intensity is a measure of the GFP intensity per cell. (C) Graph showing the distribution of cells at different GFP (FITC) intensities after transfection with a range of siGFP concentrations (see legend). (D) Distribution of cells at different RFP (TRITC) intensities after transfection with a range of siGFP concentrations (see legend).
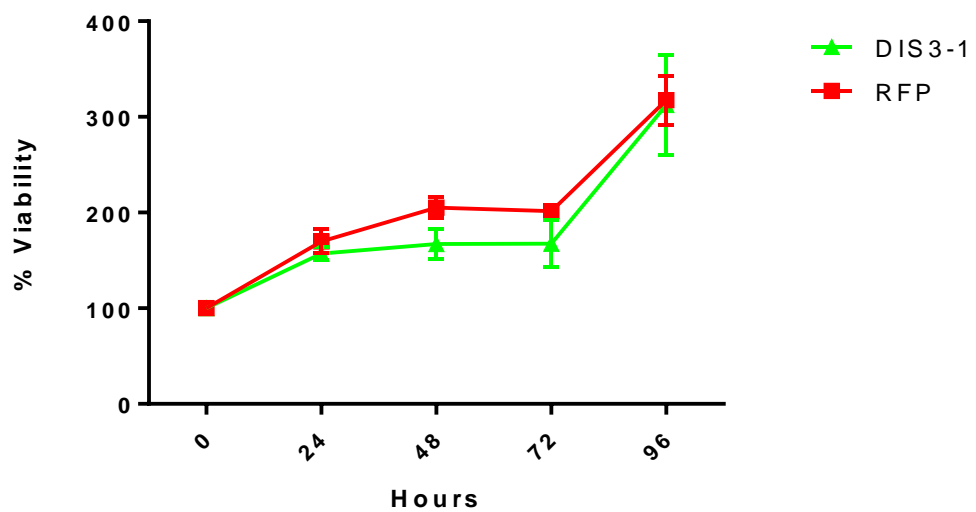
Figure 5.7. No difference in viability between the DIS3 knock-down cell line 1 and control is observed. Viability measured using the WST-1 assay. Results are shown as the mean optical density (OD), relative to the OD at time point 0 ± the SEM for 3 technical replicates.

pool of 4 individual siRNAs which are guaranteed to silence target gene expression by at least 75% at the mRNA level. Figure 5.8 shows the DDR plate layouts.

Once the control experiments had been carried out, the DDR screen was set up. The two cell lines were co-cultured at a 50:50 ratio in each well of three 96 well plates and incubated with the siRNAs according to the plate maps in Figure 5.8 for 72 hours. After 72 hours the plates were imaged using the Olympus Scan R system which automatically scans each well and acquires data on the number and intensity of GFP and RFP cells. Figure 5.9 demonstrates there to be no significant difference in the proportion of knock-down and control cells in the negative controls across the plates, indicating the method alone does not differentially impede the survival of one cell line more than the other. The data for each gene was analysed by calculating Z-scores which represent how many standard deviations the score is away from the plate mean. An example analysis is presented in Figure 5.10. Genes with a Z-score of below -2 confer synthetic lethality with DIS3; those with a Z-score of above 2 confer a growth advantage to DIS3 knock-down cells compared to controls. Three repeats of the screen were performed and an average Z-score plotted for each gene, presented in Figure 5.11 as a waterfall plot. As can be seen, the only gene with an average Z-score of less than -2 and thus possibly conferring synthetic lethality is FLJ22833, otherwise known as NABP1. Two genes, PRKCG and GTF2H3 had average Z-scores of above 2, possibly indicating a synthetic growth advantage. However, in order to take account of variability between screens, genes were sorted by covariance (Figure 5.12A). Figure 5.12A plots all three Z-scores for each gene and shows increasing variability from left to right. It is clear to see that those genes to the left-most of the axis which have the lowest variability also appear to show the lowest Z-scores. If those genes with an absolute covariance of less than 1 are selected (Figure 5.12B), the gene with the largest Z-scores is FLJ22833, therefore this gene was selected for further validation.

FLJ22833, otherwise known as NABP1 (nucleic acid binding protein 1) is a component of the heterodimeric sensor of ssDNA (SOSS) complex, a multiprotein complex that functions downstream of the MRN (Mre11, Rad50 and Nbs1) complex

DDR Plate 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	Scramble	RAD50	POLE2	RUVBL2	PRKCG	FANCC	FEN1	TCEA1	RTEL1	GCN5L2	APTX	X
B	T. Reagent	RAD18	TTRAP	GTF2H5	POLE	UBE2B	MDC1	IHPK3	SIRT1	TREX1	WRN	Scramble
C	siGFP	DDX11	APEX1	TDG	TOPBP1	RAD54L	RPA1	ATF2	VCP	ALKBH2	GTF2H4	T. Reagent
D	GAPDH	PMS1	BRCA1	POLM	REV3L	HMGB2	GADD45A	IGHMBP2	PMS2	CSNK1E	BRIP1	Scramble
E	Scramble	CSPG6	RAD52	FANCL	FANCD2	TRIP13	TYMS	XPC	HUS1	RPS27L	DNA2L	siGFP
F	X	MAD2L2	KIAA1596	SETMAR	PRKDC	C11ORF13	PARP2	POLI	RAD17	TOP2A	PER1	GAPDH
G	X	ADPRTL3	NEIL2	REV1L	SOD1	CSNK1D	MSH3	MSH4	XAB2	FANCG	ATR	X
H	X	HEL308	RAD51L3	UNG2	GTF2H2	YBX1	XRCC1	GTF2H1	ERCC5	MUS81	RAP80	X

DDR Plate 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	Scramble	FLJ13614	TRIM28	POLS	CXORF53	POLG2	DCLRE1A	UVRAG	TREX2	HTATIP	RECQL	X
B	T. Reagent	CHEK2	NUDT1	MBD4	RNF168	PRPF19	FRAP1	RAD23B	MJD	DCLRE1C	FANCB	Scramble
C	siGFP	CETN2	KUB3	TP73	OGG1	LIG3	MEN1	MLH1	MRE11A	RRM2B	FLJ40869	T. Reagent
D	GAPDH	DCLRE1B	ERCC3	GIYD1	MUTYH	TDP1	POLH	GADD45G	EYA3	XPA	RAD23A	Scramble
E	Scramble	POLK	SHFM1	NEIL3	UBE2A	HRMT1L6	RNF8	TP53	RPA2	MMS19L	MGC2731	siGFP
F	X	POLN	MIZF	MSH6	FANCE	EME2	C2ORF13	TP53BP1	MNAT1	PMS2L5	SMC6L1	GAPDH
G	X	CCNH	RBBP8	XRCC2	RECQL5	NEIL1	FLJ12610	XRCC4	DLG7	EXO1	ABL1	X
H	X	C7ORF11	HMGB1	RAD54B	ERCC6	LIG1	RPA3	CHAF1A	SPO11	DNMT1	USP1	X

DDR Plate 3

	1	2	3	4	5	6	7	8	9	10	11	12
A	Scramble	EYA1	RECQL4	RAD52B	MLH3	CIB1	BTG2	MPG	TNP1	MSH2	RAD51	X
B	T. Reagent	RAD1	FLJ21816	KIAA1018	CNOT7	CDKN2D	DDB1	CKN1	PARP1	MGC32020	MGC4189	Scramble
C	siGFP	POLA	RAD9A	RENT1	NBS1	DMC1	PCNA	BAZ1B	ALKBH	POLB	NTHL1	T. Reagent
D	GAPDH	DDB2	POLD1	MGMT	FANCF	PARG	ERCC2	TADA3L	ATRX	UBE2V1	POLL	Scramble
E	Scramble	GTF2H3	EME1	POLQ	RAD51C	MSH5	DEPC-1	LIG4	ATM	SMC1L1	CDK7	siGFP
F	X	FANCA	KIAA0625	FLJ10719	UBE2V2	SMUG1	RAD21	RAD51L1	UNG	CHEK1	ATRIP	GAPDH
G	X	DUT	PNKP	BLM	APEX2	BRCA2	G22P1	CLK2	POLG	BRE	XRCC5	X
H	X	HSU24186	XRCC3	NPM1	ASF1A	H2AFX	FLJ22833	UBE2N	ERCC4	ERCC1	RRM2	X

Figure 5.8. Plate layouts from the DNA-damage response siRNA library (Dharmacon). The library consists of 240 siRNAs spread across three 96-well plates. Wells in lanes 2 to 11 of the plate contains a pool of 4 individual siRNAs which target a particular gene involved in the DNA-damage response. Lanes 1 and 12 are reserved for the positive and negative controls. The scrambled siRNA provides the baseline for the analysis and the transfection reagent ensures treatment does not affect cells. GAPDH acts as a negative control as its knock-down should not favour the survival of either cell line and siGFP acts as a positive control.
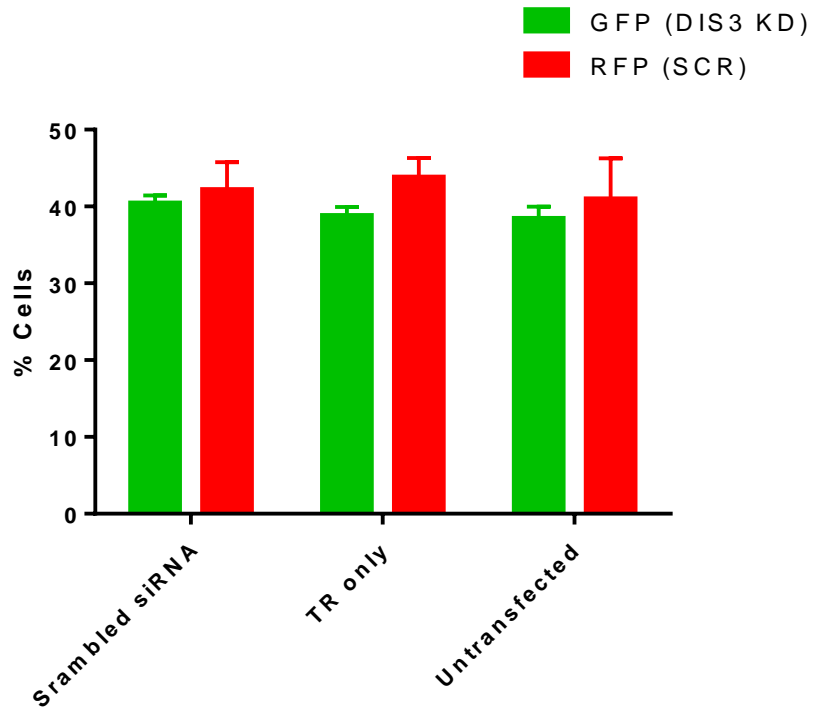
Figure 5.9. The proportion of DIS3 knock-down and control cells does not differ in the negative controls of the DNA-damage response siRNA screen. Negative controls included the use of a scrambled siRNA, a transfection-reagent (TR) only and untransfected control. The lack of a difference in the number of cells of each cell line in these controls indicates the method alone does not impede the survival of one more than the other.

GENE	Mean GFP %	Mean RFP %	Ratio Green/Red	Numerical Ratio	Ratio/ Scramble	Average scramble
RAD50	47.6	40.7	47.6:40.7	1.169	1.214	0.963

MEAN ACROSS PLATE	SD	Z - Score
0.99	0.184	1.216

Figure 5.10. Example analysis of the ratio of RFP to GFP cells for one of the genes in the DDR screen. The Olympus Scan R system detects the proportion of red and green cells in each well. A ratio is then calculated for each gene by dividing the percentage of red cells by the percentage of green. This ratio is then divided by the average ratio of red to green cells in the scrambled wells. These figures are used to generate a mean across the plate which should be close to 1 and a standard deviation. A Z-score is calculated for each gene by subtracting the mean across the plate from the ratio/scrambled, divided by the standard deviation across the plate. This score represents how many standard deviations the score for each gene is away from the plate mean. Z-scores of 0 suggest the siRNA has had no differential effect on the knock-down and control cells. Z-scores of below 0 indicate there are more control cells than knock-down cells, indicating a possible synthetic lethality. Z-scores above 0 indicate there are more knock-down than control cells, indicating a possible growth advantage to the former. In this example knock-down of RAD50 has generated a Z-score of 1.216, suggesting the DIS3 knock-down cells have a slight growth advantage. Scores below -2 or above 2 are considered potential hits.
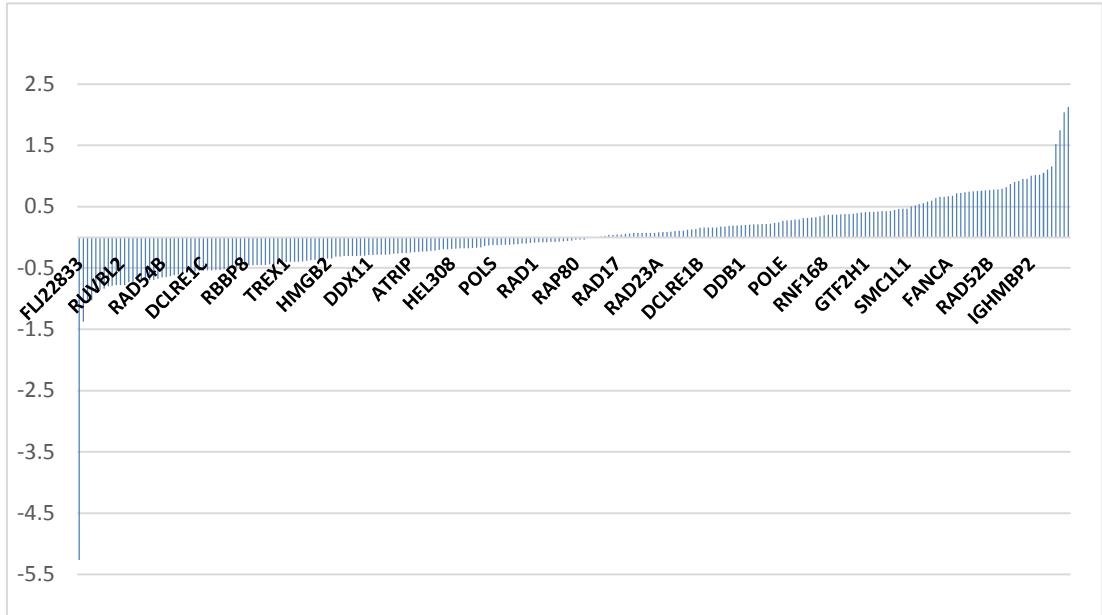
Figure 5.11. Waterfall plot presenting the average Z-score for each of 240 genes involved in the DNA-damage response from three individual siRNA screens. Z-scores represent how many standard deviations the score is away from the plate mean. Genes with a Z-score of below -2 may confer synthetic lethality with DIS3; those with a Z-score of above 2 may confer a growth advantage to DIS3 knock-down cells compared to controls.
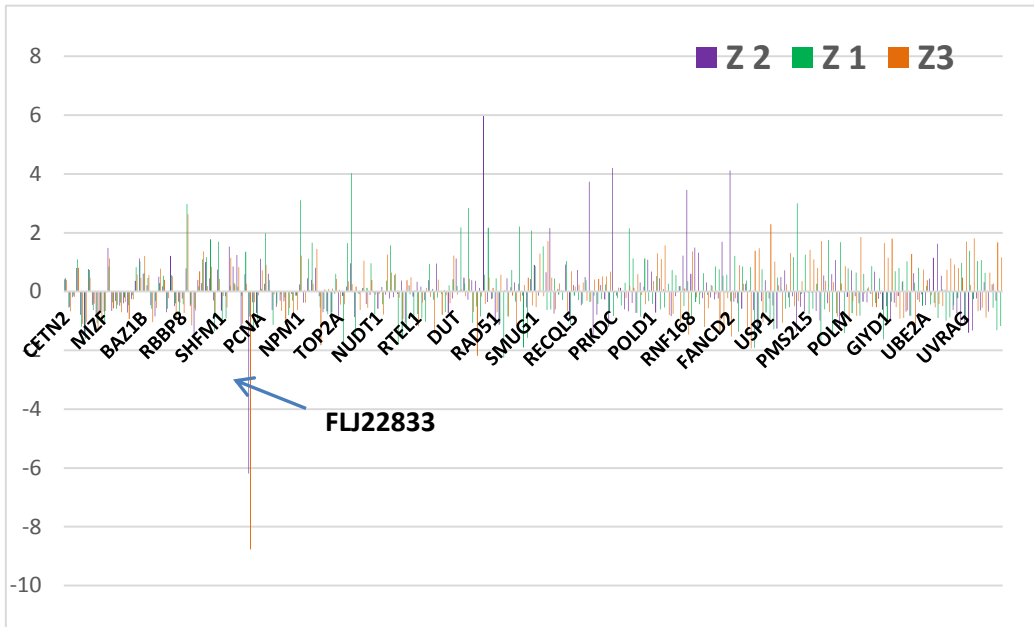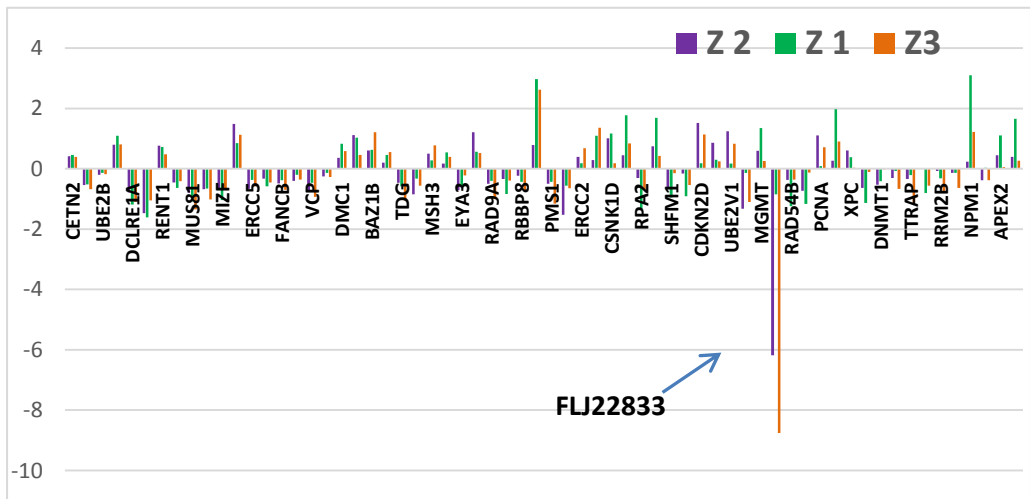
Figure 5.12. Graphs showing the Z-scores for genes involved in the DNA-damage response from three individual screens, sorted by covariance. (A) The three Z-scores are presented for each of the 240 genes, sorted by absolute covariance. (B) Genes with a covariance of less than 1 are presented as in (A). Variability between the Z-scores increases from left to right on the x-axis. Z-scores represent how many standard deviations the score is away from the plate mean. Genes with a Z-score of below -2 may confer synthetic lethality with DIS3; those with a Z-score of above 2 may confer a growth advantage to DIS3 knock-down cells compared to controls.

to promote DNA repair and G2/M checkpoint activation (Huang et al., 2009b). The MRN complex plays an important role in the initial processing of double-strand DNA breaks (Paull and Lee, 2005), potentially tethering the broken ends, prior to repair by homologous recombination or non-homologous end joining. In the SOSS complex, NABP1 binds to and acts as a sensor of single-stranded DNA, in particular to polypyrimidines (Li et al., 2009). The SOSS complex associates with DNA lesions and influences diverse endpoints in the cellular DNA damage response including cell-cycle checkpoint activation, recombinational repair and maintenance of genomic stability. NABP1 is known to be essential for efficient homologous recombination-dependent repair of double-strand breaks (DSBs) and ATM-dependent signalling pathways. Although only speculative at this stage, given this role and its potential synthetic lethal interaction with DIS3, NABP1 may also be involved in recruiting the AID enzyme to the open DNA duplex through sensing the presence of single-stranded DNA.

5.6 Validation

In order to test this potential hit is a true positive and not a result of off-target effects, an individual transfection was set up with three new siRNAs. These three siRNAs were selected based on their targeting of different regions of the FLJ22833 (NABP1) transcript to those in pool in the DDR plate. Unfortunately however, this individual transfection did not cause any difference in survival rate of the knock-down and control cells (Figure 5.13), this could possibly mean that the initial difference was caused by off-target effects of the siRNAs used in the DDR plate.

5.7 Discussion

In an effort to provide more effective therapies with fewer side effects, cancer treatments are now focusing on a selective targeting of tumour cells. Nevertheless, due to a pharmacological difficulty in restoring the function of tumour suppressors the majority of targeted anticancer drugs inhibit mutated oncogenes that display increased activity (Kaelin, 2005). DIS3 is a recurrently mutated tumour suppressor
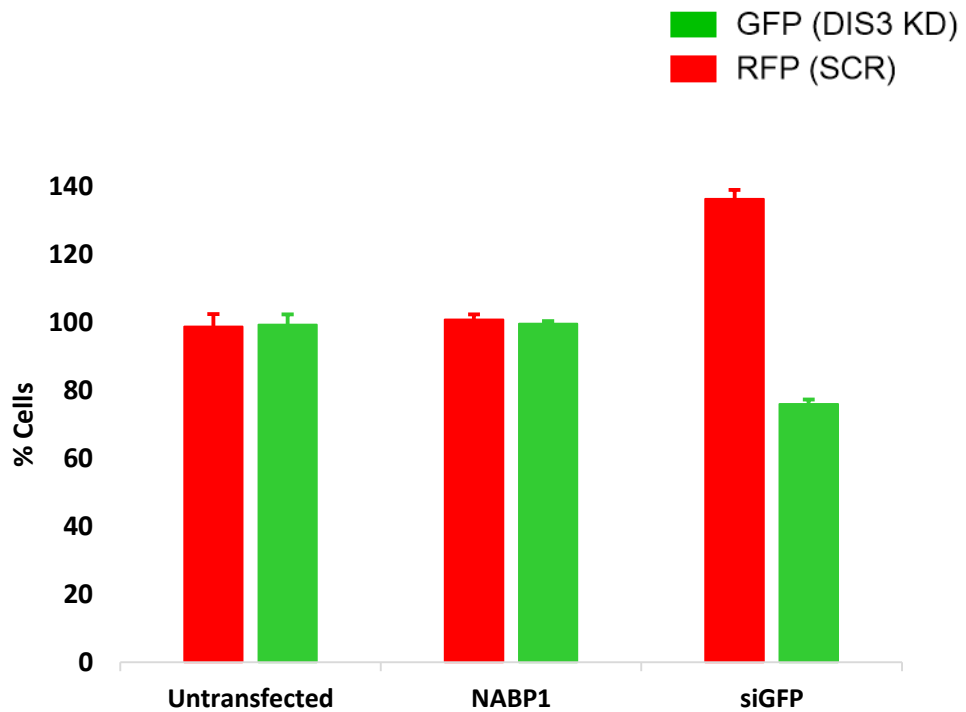
Figure 5.13. Validation of the synthetic lethal interaction observed between DIS3 and NABP1 in the DDR screen. No difference in survival was seen between the control and knock-down cells upon transfection with siNABP1. Untransfected cells were used as a negative control and siGFP as a positive control. % cells represents the percentage of cells compared to a scrambled siRNA control. Error bars represent S.E.M obtained from at least 1.6×10^5 cells imaged for each treatment.

in multiple myeloma and although its role in RNA degradation is well understood, little knowledge exists of the process through which mutations contribute to pathogenesis. Thus, the loss-of-function mutations in DIS3 coupled with the lack of knowledge on its role in myeloma make devising targeted treatments a formidable task.

In recent years, utilising the concept of synthetic lethality as a method to selectively target cancer cells with specific mutations is becoming a popular strategy. While it is difficult to predict synthetic lethal interactions if little knowledge of a process exists, with the discovery of RNAi, unbiased large-scale functional genomic screens for the identification of such targets have become possible. RNAi screens provide not only a means to discover new drug targets but also a means to enhance our understanding of the fundamental biology of interactions within cancer cells (Chan and Giaccia, 2011). Therefore, the main aim of this chapter was to use siRNA libraries to identify synthetic lethal interactors of DIS3 as a means to understand its role in myelomagenesis as well as potentially yielding drug targets for DIS3-mutated patients.

5.7.1 Summary of results

Due to a previously indicated role of the exosome in recruiting activation-induced cytidine deaminase (AID) to chromatin in B-cells which creates double-strand breaks (Basu et al., 2011), combined with the knowledge that cancer cells already exhibit genomic instability and defective DNA repair, a DNA-damage response library was selected for the synthetic lethality screen. Using an arbitrary cut off of two standard deviations above or below the mean, one siRNA, NABP1 was found to confer synthetic lethality with loss of DIS3 and two, PRKCG and GTF2H3, conferred a synthetic growth advantage.

With a predominant aim of identifying potential drug targets, NABP1 was selected to follow up and validate to ensure the observed result was not due chance or off-target effects. However, no such difference in survival was observed between

control and knock-down populations. Despite using a low concentration of siRNA in the original screen, the initial difference in survival may have been due to off-target effects of one or more of the siRNA from the pool. Alternatively, as the cells were a different passage to those used originally, the lack of reproducibility could be due to an inherent change in the cells making them less susceptible to NABP1 knock-down.

5.7.2 Reasons for lack of synthetic interactors with DIS3

Curiously, the remaining 237 genes tested showed no major impact on the survival of either DIS3 knock-down or control cells. There may be a number of reasons for this. Firstly, although there is a precedent for a role of the exosome in double-strand break formation which can lead to the translocations observed in multiple myeloma, DIS3 itself may not be the candidate exonuclease involved. Instead Rrp6 which also provides exonucleolytic function to the exosome may be functioning in this process. Therefore, it is feasible that no interactions exist between DIS3, involved in RNA degradation and genes involved in the DNA-damage response. Alternatively, there may be genes in the library that interact with DIS3, however the siRNAs may not effectively reduce their levels at the concentration used in this set-up. Another explanation originates from the fact that shRNA-mediated knock-down of DIS3 does not completely remove the protein; in my study about 30% of the levels of the control still remained in the knock-down cells. The incomplete removal of DIS3, in combination with a potential redundancy with another ribonuclease, may be sufficient for normal functioning and therefore reducing the likelihood of observing synthetic lethal interactions.

As well as aberrations affecting intracellular mechanisms, multiple myeloma is characterised by alterations to the bone marrow microenvironment. Reciprocal interactions exist between the different components of the BM microenvironment and malignant plasma cells to promote tumour growth and survival as well as immune cells to allow evasion of immune recognition (Fowler et al., 2011). Given the lack of an observable phenotype upon DIS3 knock-down as discussed in chapter

3, it is feasible that DIS3 may normally function to target bone marrow signalling molecules, which become up-regulated when the enzyme is mutated, promoting myeloma pathogenesis. Consequently, without culturing DIS3 knock-down cells in a bone-marrow microenvironment, synthetic lethal interactions may not be revealed. Further work could investigate the interaction of DIS3 with the tumour microenvironment using so-called "contextual synthetic lethality" (Chan et al., 2010). It is also worth pointing out that in the analysis, only Z-scores of less than -2 or above 2 were considered likely hits. However, this does not mean that other genes with lower Z scores should be dismissed as it may be that with further investigation, for instance more efficient knock-down, these genes do in fact yield a synthetic lethality.

5.7.3 Advantages of the screening method

Although this work is in the early stages, it demonstrates how synthetic lethality is a valuable concept to understand the functions of genes and has immediate relevance for cancer therapy. The strategy described here employs two basic components. The first involves the use of paired human cancer cell lines that differ only in a single mutant gene that is often altered in the tumour type under study. The second involves the introduction of genes encoding different fluorescent proteins into each of these lines, thus individually marking them. The two lines are then co-cultured, and their relative growth followed using fluorescence spectroscopy. SiRNAs were tested against clones that did not express fluorescent proteins to ensure that the differential toxicity was not the result of clonal variability or confounding effects of the fluorescent proteins. The design of the screen has several major advantages of others in common use. The two lines to be compared are co-cultured and assayed simultaneously, eliminating a variety of errors encountered when screening cell pairs that are maintained in separate plates. Whereas, our screen contains internal controls for each well normalising for variability of cell numbers, most previously published screens experience potentially confounding effects such as differences in proliferation rate and cell-cycle distribution between cell lines. In co-culturing parental and gene-targeted

cells, we have therefore devised a screening strategy allowing precise internal calibration of each assay and rapid throughput. Furthermore, assays based on engineered fluorescence proteins are highly cost-effective, because no additional reagents or pipetting steps are required for analysis of growth.

5.7.4 Limitations of synthetic lethality screens

Despite the potential of synthetic lethality screens to aid in cancer therapeutics, currently there are some limitations to their success. As can be observed from my study, interactions are more accurately described as synthetic sick rather than synthetic lethal so the cancer cells would not necessarily be completely eliminated. Of previously published screens, there appears to be little overlap in results (Nijman, 2011). Some of this may be due to the use of different cell lines as well as technical approach; the use of different shRNA libraries that target only partially overlapping genes and inhibit gene expression with different efficiencies. A weakness of RNAi screens stems from the fact that RNAi typically only results in 70–90% inhibition of gene expression, potentially explaining why interactions are mostly synthetic sick rather than synthetic lethal. Additionally, RNAi screens are plagued with off target effects providing an explanation for the common observation of false positives. Furthermore, an interaction identified from an RNAi screen may not necessarily lead to a therapeutic; that is, a compound to inhibit or activate the identified interaction target may not exist (Chan and Giaccia, 2011). Moreover, proteins bound to drugs might have effects that are very different from those predicted by RNA interference (RNAi) that cause quantitative reductions in protein abundance. For example, a drug might interfere with one function of a multifunctional protein, or cause a protein to act in a dominant-negative or dominant-positive manner (Kaelin, 2005). For this reason, screens for synthetic lethality that are carried out using libraries of chemical compounds are likely to be complementary to screens that are carried out using genetic tools. Unlike siRNA screens, small molecule screens directly provide potential candidates for optimisation into lead compounds.

Although synthetic lethal interactions may provide handles for novel therapies, most of the described genetic interactions in cancer cells do not strictly adhere to the definition of synthetic lethality which demands that only the combination of two perturbations results in lethality. In an *in vitro* set-up, isogenic cell lines are used that differ only in the mutation of interest, yet *in vivo*, cancer cells have acquired not just one but an array of mutations. Therefore, there is no guarantee that *in vivo*, inhibiting the interactor will have the desired effect of killing the cancer cells without causing unwanted side-effects. In addition, there are now many examples where different phenotypes have been observed following inactivation of a particular tumour suppressor gene in both mice and humans (Kaelin, 2005). These observations indicate that synthetic lethal relationships ultimately need to be discovered or validated in relevant human cells, and that caution needs to be exercised when extrapolating cell-culture results to intact organisms.

Analogous to the therapeutic window for drugs, i.e., the safe and effective dose range, "the synthetic lethality window" of the genetic interaction will be important for its potential as a therapy (Nijman, 2011). For example, with the BRCA-PARP synthetic lethal interaction, BRCA mutant cells are some two orders of magnitude more sensitive to PARP inhibitors than non-mutant cells and this strong responsiveness is probably key to its clinical success. It is currently unclear how common this type of strong genetic interaction is in human cells. At the same time, random mutations and genome plasticity, viewed at the level of a tumour, markedly increase the likelihood that rare therapy-resistant subclones will emerge. A 1-cm3 tumour already contains >10^9 cells. So, the likelihood of clinical success will increase with early diagnosis (to minimise the number of cells in the pool from which resistant cells might arise).

5.7.5 New approaches

As well as single-well screens as employed in this study, several groups have developed a second screening approach that has some distinct advantages. To circumvent the practical limitations of screening individual siRNAs in a single-well,

pooled methods are employed to genetically barcode knockout clones so they can still be accurately identified even within a mixed population of cells (Silva et al., 2008; Luo et al., 2008; Schlabach et al., 2008). Cell lines (A and B) are infected with shRNA libraries targeting thousands of gene products. Cells are cultured to allow the depletion of those containing shRNAs that target essential genes. Genomic DNA is isolated and the vectors are quantified using so-called barcode sequences (short stretches of DNA) that are unique for each shRNA vector. By comparing the genes that are required in one cell line but not the other by custom micro-array or deep sequencing, potential synthetic interactions can be identified. Such screens are much easier to handle than large-scale single well screens and thus have a higher throughput.

Furthermore, an alternative method to using isogenic lines has been established. A confounding factor of using isogenic cell lines is that the single genetic variation being studied may not actually be the only difference in 'isogenic cells'. Rather, genetic drift between pairs of isogenic cell lines may result in multiple differences that can alter responses to RNAi or drug treatment. In this strategy, a single human cancer cell line that is deficient in a gene of interest is used. Complementation of this gene is provided by a low-copy unstable episome expressing this gene (Simons et al., 2001b; Simons et al., 2001a). In the context of a drug or RNAi screen, retention of this episome is selected under synthetic lethal conditions, thus revealing novel interactions. Although isogenic lines often provide a valuable avenue for synthetic lethality screens, this work demonstrates that other approaches also have distinct advantages.

5.7.6 Final conclusions

This chapter has explored how synthetic lethality is likely to be a valuable concept to understand the functions of genes, the mechanism of action of drugs, and has immediate relevance for cancer therapy. Because such genetic approaches allow an alignment of particular molecular defects with specific drugs, there is a high probability that the serious side effects associated with many currently used

chemotherapeutics will be less problematic. We are clearly poised to move away from empirically discovered cytotoxics towards new agents that are based on a knowledge of cancer genetics and a more sophisticated view of gene–gene interactions.

Chapter 6: Characterising the two protein-coding isoforms of DIS3

6.1 Introduction

Historically the relationship between genes and proteins was described as the 'one gene-one polypeptide' hypothesis, whereby each gene is responsible for the synthesis of a single protein. Clearly an oversimplification, this theory is now known to be incorrect. This is attributable to both the presence of genes that code for RNAs as end-products and also for genes whose mRNAs undergo alternative splicing. Alternative splicing is a regulated post-transcriptional process in which particular exons may be included or excluded from the final mRNA to produce protein isoforms that contain differences in their amino acid sequence and thus often in their biological function (Black, 2003). This process permits a more varied proteome from a genome of limited size, helping to partly explain the G-value paradox in which organism complexity is not reflected by the number of protein-coding genes in its genome (Hahn and Wray, 2002). In humans at least 70% of genes are known to be alternatively spliced with each gene giving rise to an average of 4 alternatively spliced variants. This phenomenon allows the genome to direct the synthesis of a much larger proteome than would be expected from its 20,000 protein-coding genes, greatly increasing structural and functional diversity.

The general mechanism of RNA splicing occurs via an RNA-protein complex called the spliceosome composed of multiple small nuclear ribonucleoproteins (snRNPs), namely U1, U2, U4, U5 and U6 as well as a large network of auxiliary proteins. A typical eukaryotic intron contains consensus sequences such as the 5' splice site, branch point and 3' acceptor site that are recognised by the spliceosome (Dredge et al., 2001). Splicing occurs in a step-wise manner by the association and dissociation of different snRNPs within the complex at these consensus sequences (Figure 6.1). Different modes of alternative splicing are known to occur, namely exon skipping, intron retention, the use of mutually exclusive exons and the use of alternative splice donor or acceptor sites (Sammeth et al., 2008) (Figure 6.2). Alternative splicing is regulated by different combinations of trans- acting proteins and

Pre-mRNA
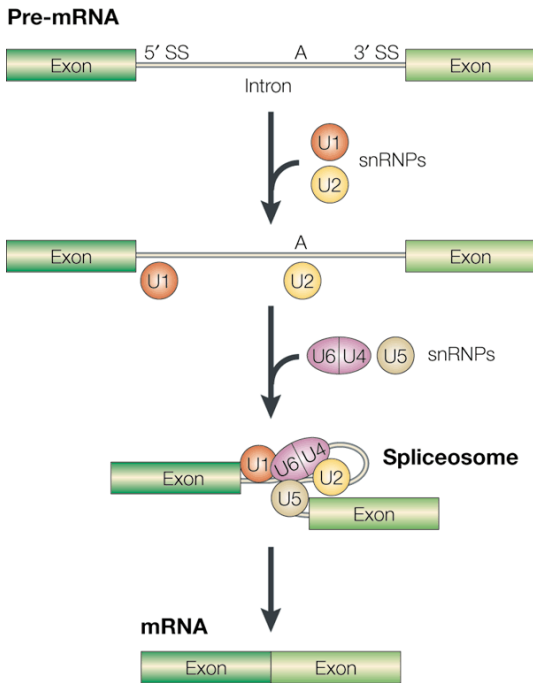
Figure 6.1. The basic mechanism of RNA splicing. Introns contain conserved motifs such as a 5' splice site, 3' splice site and adenosine branch point. These motifs are initially recognised by the snRNPS U1 and U2 leading to the recruitment of other components of the spliceosome, (U4/6 and U5 snRNPs), followed by rearrangement in the base pairing with RNA and excision of the intron. Taken *from Dredge et al, 2001*.
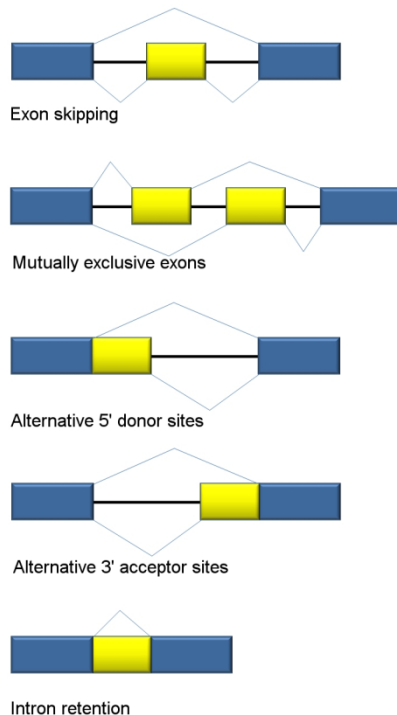
Figure 6.2. The different modes of alternative splicing. Exon skipping: an exon may be spliced out of the primary transcript or retained. Mutually exclusive exons: one of the two exons is retained in mRNAs after splicing but not both. Alternative donor site: an alternative 5' SS is used changing the 3' boundary of the upstream exon. Alternative acceptor site: an alternative 3' SS is used, changing the 5' boundary of the downstream exon. Intron retention: an intron may be spliced out or retained. *(Sammeth, 2008)*

cis-acting regulatory sites on the pre-mRNA that form a 'splicing code' governing how splicing will occur under different cellular conditions(Wang and Burge, 2008) (Barash et al., 2010). Temporal and tissue specific alternative splicing is thought to be controlled by the differential expression, concentration and/or activity of these splicing factors.

A number of studies have indicated that many diseases have a splicing component, either due to changes in the splicing machinery resulting in mis-splicing of multiple transcripts, or due to mutations within the splicing sequences of individual genes. Cancer is a prominent example of a disease in which abnormally spliced mRNAs are found in high proportions, however it is not clear whether aberrant splicing patterns are a cause or consequence of oncogenesis (Skotheim and Nees, 2007). Notably, in colorectal and prostate cancer, the number of splicing errors has been shown to vary greatly between individuals (Sveen et al., 2011), possibly indicating the former. On the other hand, in patients with myelodysplastic syndromes such as chronic myelomonocytic leukaemia, the most common mutations are those affecting spliceosomal proteins (Yoshida et al., 2011). Recently, mutations affecting the splicing factor SRSF2 which is critical for recruiting U1 and U2 snRNPs to the pre-mRNA, have been found to alter its recognition of specific splicing motifs, causing mis-splicing of key haematopoetic regulators and impairing haematopoetic differentiation (Kim et al., 2015). This undoubtedly reflects a causative role of splicing mutations in the development of myelodysplastic disorders.

Due to its importance in the pathogenesis of multiple myeloma, DIS3 is currently a gene of great interest to the scientific community, yet no information on whether it undergoes alternative splicing is present in the literature. Therefore, this chapter seeks to investigate whether protein-coding isoforms of DIS3 exist and whether differential expression of these could contribute to disease phenotypes.

6.2 Aims

The central focus of this chapter is to characterise the different alternatively spliced isoforms of DIS3 leading to the following aims:

1. Use publicly available bioinformatics data to identify alternatively spliced isoforms of DIS3
2. Determine whether DIS3 isoforms are translated into protein
3. Determine whether DIS3 isoforms are differentially expressed in different tissue-types and diseases
4. Determine whether there is a functional difference between the DIS3 isoforms

6.3 Characterisation of two DIS3 protein-coding isoforms

Examination of the DIS3 gene in the Genome Browser ENSEMBL reveals five different DIS3 transcript annotations (Figure 6.3A). Of the three transcripts listed as protein coding, two are annotated by the Consensus CDS project (CCDS) indicating consistent, high quality annotation across the different annotation platforms. These two transcripts appear to differ in the use of a mutually exclusive exon 2 and the length of their 3' UTR (Figure 6.3B). Figure 6.4A shows the structure of the DIS3 gene. Notably exon 2α used by DIS3 isoform 1 is longer than exon 2β, used by isoform 2. Exon 2 encodes a large region of the endonucleolytic PIN domain (Figure 6.4B). If isoform 2 is translated, the inclusion of exon 2β would result in a PIN domain 30 amino acids shorter than isoform 1 whilst leaving the rest of the protein in frame. As the total length of the PIN domain in isoform 1 is 118 amino acids, translation of isoform 2 would reduce its size by over 25%.

Through a combination of techniques, including manual inspection of structures, homology modelling according to the recently published structure of *S.cerevisiae* Rrp44 (Makino et al., 2013), using Phyre2 (Kelley and Sternberg, 2009) and the webserver 'Site Directed Mutator' (SDM), we were able to analyse and make predictions about the potential consequences of using exon 2β on the stability of

Figure 6.3. Annotation of the DIS3 gene from the publicly available database, ENSEMBL. (A) Five transcripts are annotated and the biotype is given. Of the three protein-coding transcripts, two are annotated by the Consensus CDS project (CCDS). (B) Diagrammatic representation of the structure of the annotated transcripts. The two yellow transcripts are those annotated by the CCDS. The two transcripts use a mutually exclusive exon 2 and appear to have different length 3′ UTRs (empty boxes). Note: DIS3 is orientated in the antisense direction therefore 5′ to 3′ is right to left.
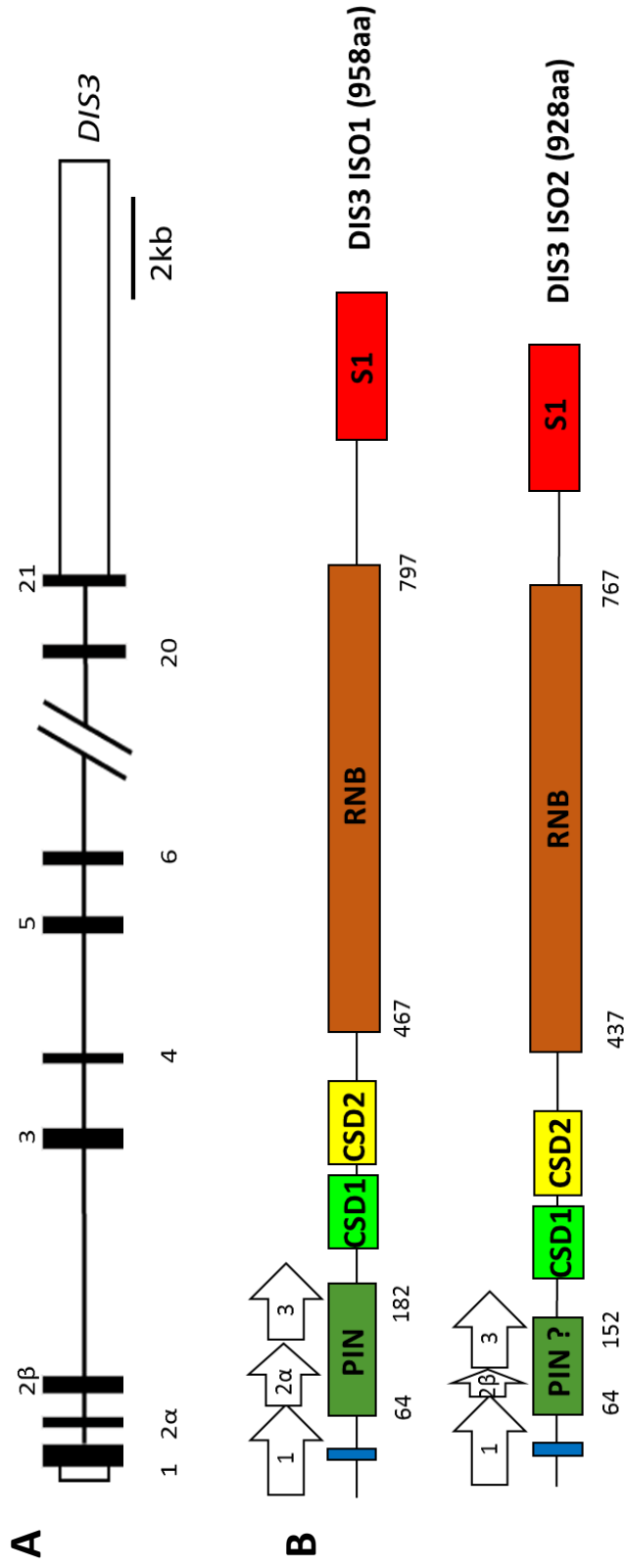
Figure 6.4. Schematic of the DIS3 locus and protein isoform domain structures. (A) The DIS3 gene encodes two alternate exons, exon 2α and 2β that are incorporated into isoform 1 and isoform 2 transcripts respectively. (B) Exon 2 encodes a large portion of the PIN domain. Exon 2β is smaller than exon 2α resulting in a PIN domain 30 amino acids shorter in the isoform 2 protein. Although the rest of the protein remains in frame, this renders isoform 2 30 amino acids shorter in total than isoform 1.

the protein. SDM uses parameters of the local structural environment such as conformation, solvent accessibility and hydrogen bonding class, to calculate a stability score (pseudo ΔΔG). Figure 6.5 shows the inclusion of exon 2β over exon 2α to result in the loss of two beta-strands and gain of an alpha-helix, likely to be highly destabilising to the PIN domain. This raises the question of whether isoform 2 contains the essential residues for catalysis as identified in *S.cerevisiae*. Figure 6.6 shows the sequence alignment of the PIN domains of four DIS3 homologues, illustrating the absence of the critical residue E120 (Lebreton et al., 2008; Schaeffer et al., 2009; Schneider et al., 2009), also missing from DIS3L which is known to have no PIN function. This finding provides the rationale to investigate these isoforms further to determine whether they are both expressed and whether they have a differential function.

6.4 Ribosome profiling data suggests both DIS3 isoforms are translated

Before pursuing an investigation into the biological significance of the two DIS3 isoforms, we needed some indication that both transcripts are translated into protein and thus functionally relevant. As an initial step in understanding whether both DIS3 isoforms are translated, the online tool GWIPS (Genome Wide Information on Protein Synthesis) was used to analyse and visualise ribo-seq data obtained using the ribosome profiling technique. Using data from all ribosome profiling studies, Figure 6.7 demonstrates that although coverage is much lower on the shorter exon 2 of isoform 2 (exon 2β), ribosome binding is above background level indicating this isoform is translated, but to a lower level than isoform 1. As such, it appears that either constitutively or under certain conditions, the cell positively selects isoform 2 for expression and translation into protein, thus providing rational for further study.

6.5 Both isoform transcripts are ubiquitously expressed in a range of cell types

To experimentally validate the presence of the two annotated DIS3 isoforms, reverse-transcription PCR was employed. Primers were designed to flank the variable exon 2 by annealing to exons 1 and 3 common to both isoforms (Figure

Figure 6.5. Three-dimensional model of the PIN region of the two DIS3 protein isoforms. The 3D-model demonstrates the loss of two beta-strands from isoform 2 and the gain of an alpha-helix, predicted to be highly destabilising. Modelled according to the recently solved *S. cerevisiae* Rrp44 structure (Makino et al, 2013) using Phyre2 (Kelley and Sternberg 2009) and the webserver "Site Directed Mutator.

```
                96          106         116         125
                |           |           |           |
86 HYVVLDTNVV LQAIDLLENP NCFFDVIVPQ IVLDEVRN-K SYPVYTRLRT   134 dis3p
64 HYLLPDTNVL LHQIDVLEDP A-IRNVIVLQ TVLQEVRN-R SAPVYKRIRD   111 DIS3 ISO1
64 HYLLPDTNVL LHQIVSAWRP G------TWA SVASSLRL-P GS--------    98 DIS3 ISO2
57 HYVIPDWKVV QDYLEILEFP E-LKGIIFMQ TACQAVQHQR GRRQYNKLRN   105 DIS3L


                145         155         165         175
                |           |           |           |
135 LCRDSDDHKR FIVFHNEFSE HTFVERLPNE TINDRNDRAI RKTCQWYSEH  184 dis3p
112 VTNNQEKH-- FYTFTNEHHR ETYVEQEQGE NANDRNDRAI RVAAKWYNEH  159 DIS3 ISO1
 99 ---------- --------L ETYVEQEQGE NANDRNDRAI RVAAKWYNEH   129 DIS3 ISO2
106 LLKDARHD-- CILFANEFQQ CCYLPRERGE SMEKWQTRSI YNAAVWYYHH  153 DIS3L


                191         201
                |           |
185 LKPY----DI NVVLVTNDRL NRE   203 dis3p
160 LKKMSADNQL QVIFITNDRR NKE   182 DIS3 ISO1
130 LKKMSADNQL QVIFITNDRR NKE   152 DIS3 ISO2
154 CQ-----DRM PIVMVTEDEE AIQ   171 DIS3L
```

Figure 6.6. Sequence alignment of the PIN domains of four DIS3 homologues.
Orange region depicts the amino acids coded for by exon 2 of isoform 2. The four
acidic residues essential for catalysis of *S.cerevisiae* DIS3 are marked in green.
S.cerevisiae E120 is absent from DIS3 ISO2. E120 is also absent from DIS3L as well
as D171. Dis3p = *S.cerevisiae*. DIS3L2 not shown as does not contain a PIN domain.
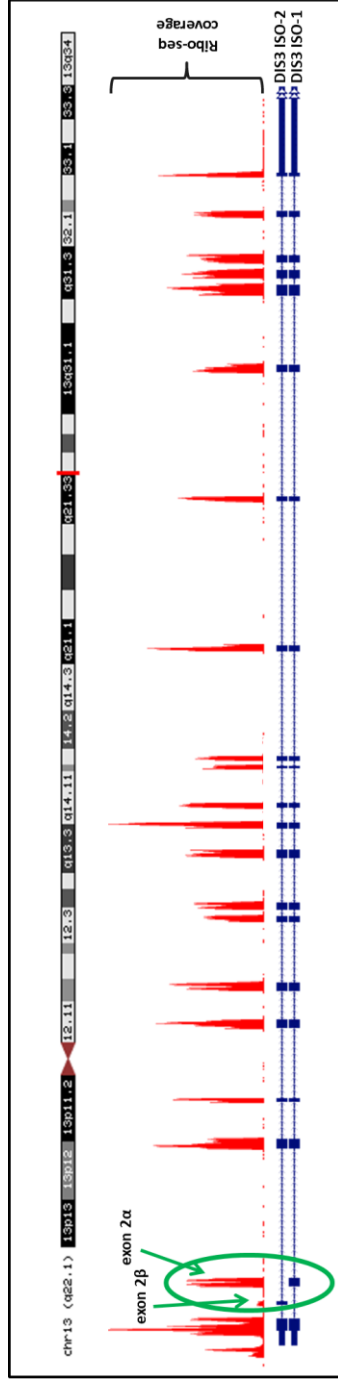
Figure 6.7. Visualisation of ribo-seq data on the DIS3 transcripts obtained by ribosome profiling. Although coverage is much lower on the shorter exon 2 of isoform 2 (exon 2β), ribosome binding is above background level indicating this isoform is translated. Alternative exons are labelled. Data obtained using the online tool GWIPS.

6.8A), thus amplifying two fragments, each corresponding to the individual isoform transcripts. RT-PCR was performed on six cell lines of differing cancer types, predominantly AML and myeloma as well as mononuclear cells from haematological cancer patients which included six myeloma patients, four AML patients and three CMML patients. Figure 6.8 shows the amplification of two bands as expected, differing in size by exactly 100bp. The larger fragment corresponds to isoform 1 with the intact PIN domain and the smaller fragment to isoform 2 containing the short PIN domain. These findings illustrate that both isoform transcripts are expressed, thus corroborating the bioinformatics data. Upon initial inspection, the intensity of the isoform 1 band appears stronger across most of the samples than isoform 2, with the exception of the CMML patients. In at least two of the three CMML samples isoform 2 appears more highly expressed than isoform 1 (Figure 6.8E). This begs the question of whether isoform expression is tissue or disease specific, requiring a more quantitative method to test this.

6.6 Isoform 1 is the principal isoform expressed in cell lines

In order to address the initial observation by RT-PCR of a differential expression level of the isoform transcripts across cell types, qPCR was carried out using isoform-specific Taqman primer-probes (Figure 6.9A). Before this however, the amplification efficiency of the primer probes was first tested using a series of dilutions to create a standard curve. This is important as the levels of the two targets are being directly compared to each other, therefore if one of the probe sets was found to amplify its target less efficiently than the other, i.e. the DNA less than doubled for each cycle, then the relative levels of this target would be underestimated and the data would be invalid. Figure 6.10 confirms a similar amplification efficiency of the two Taqman primer-probe sets.

Once the amplification efficiency of the probes had been confirmed, isoform-specific qPCR was carried out. Figure 6.9B shows the levels of the two isoforms across thirteen cell lines of different disease origin, relative to GAPDH. Isoform 1 is
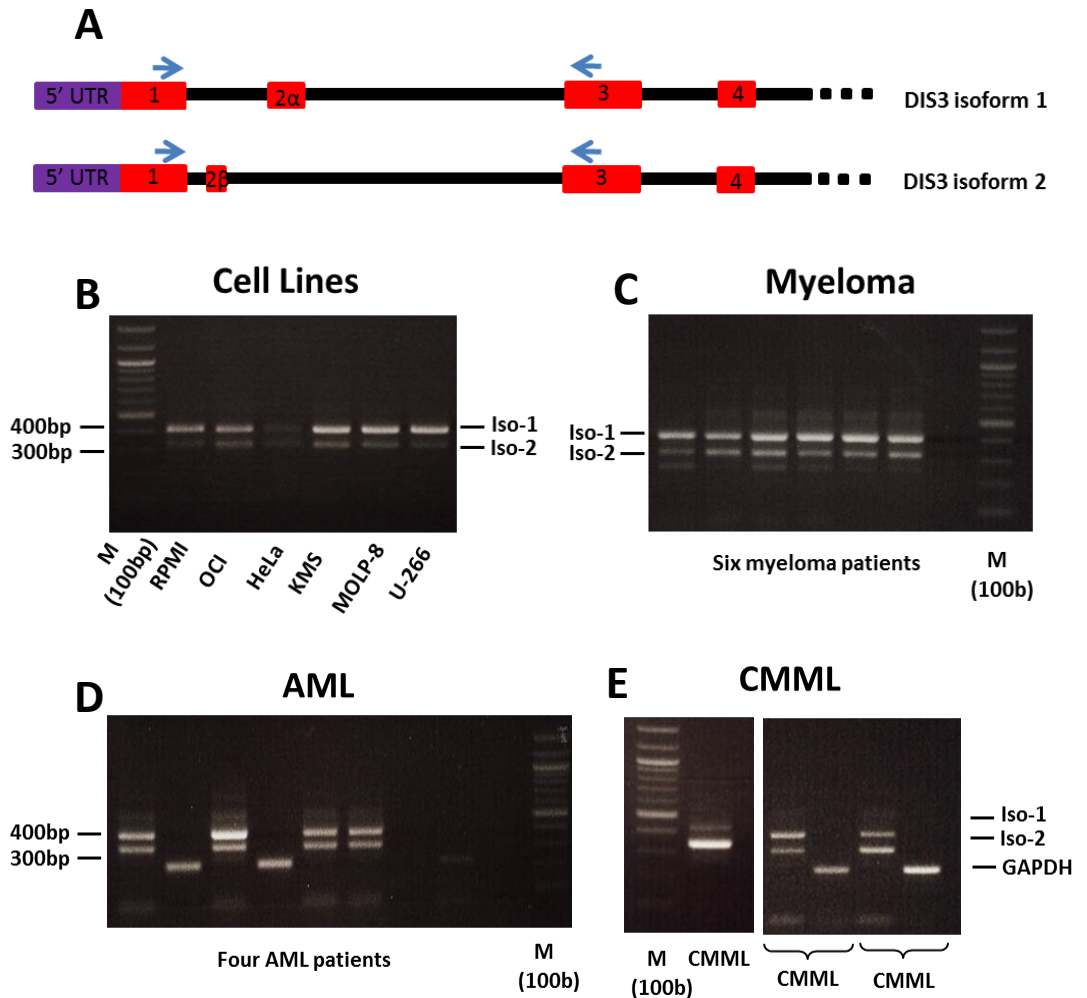
Figure 6.8. RT-PCR demonstrating the ubiquitous expression of the two DIS3 isoforms in different cell types. (A) Schematic of the two DIS3 isoforms with blue arrows showing the position of the primers flanking the variable exon 2 by annealing to exons 1 and 3 common to both isoforms. Two bands can be seen corresponding to isoform 1 (400bp) and isoform 2 (300bp) in cell lines (B), myeloma patients (C), AML patients (D) and CMML patients (E). GAPDH was used as a control.
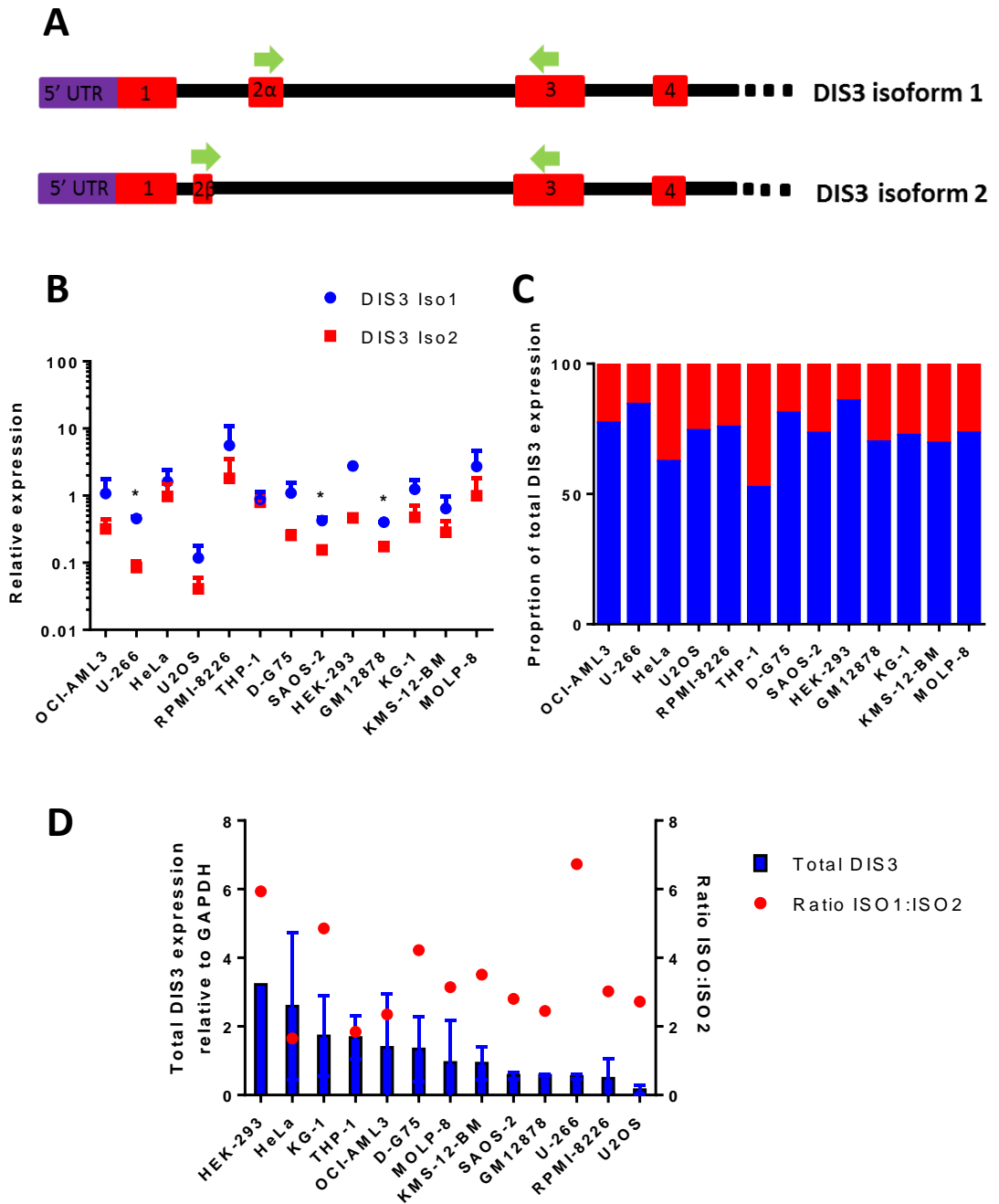
Figure 6.9. qPCR showing the expression levels of the DIS3 isoforms at the mRNA level in a range of cell lines. (A) Schematic of the two DIS3 isoform transcripts with green arrows showing the position of the isoform-specific Taqman primers. (B) Relative expression of the two isoforms in thirteen cell lines relative to GAPDH (n=3). Asterisks indicate significance at $p < 0.05$. (C) Expression of the two isoforms as a proportion of total DIS3 expression in the cell lines. Proportions represent average of three biological replicates. (D) Total DIS3 expression relative to GAPDH in thirteen cell lines does not correlate with isoform ratio.

DIS3 Isoform 1
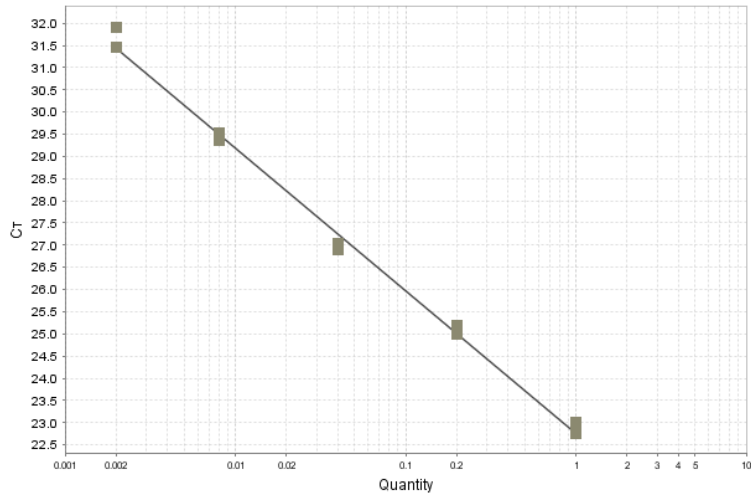
Target: dis31 Slope: -3.213 Y-Inter: 22.757 R²: 0.996 Eff%: 104.742 Error: 0.059

DIS3 Isoform 2
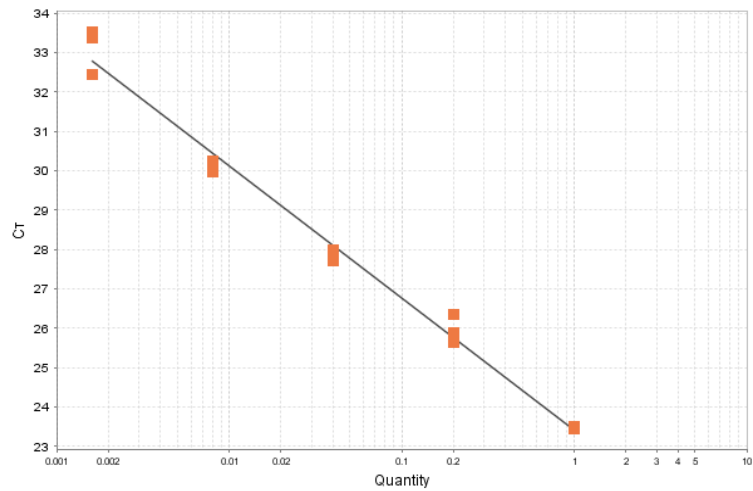
Target: dis32 Slope: -3.349 Y-Inter: 23.418 R²: 0.987 Eff%: 98.866 Error: 0.105

Figure 6.10 . Standard curves and amplification efficiency of the DIS31 and DIS32 primers. (A) DIS3 isoform 1, (B) DIS3 isoform 2. Standard curves were generated using a series of cDNA dilutions (6.25, 12.5, 25, 50ng per 10ul qPCR reaction, based on RNA concentrations) and using the standard curve set up on the Life Technologies ViiA™ 7 System.

consistently more highly expressed than isoform 2 in all cell lines. Figure 6.9C illustrates the difference in isoform expression as a proportion of total DIS3 expression. This reveals a fairly consistent pattern across cell lines with isoform 1 contributing an average of approximately 70% to total DIS3 levels. As isoform 1 appears to contribute to the majority of the total DIS3 transcripts expressed, an intriguing question is whether there is a correlation between the relative level of the two isoforms and the total level of DIS3 expressed in these cell lines. Nevertheless, Figure 6.9D shows this not to be the case (r=0.143, p = 0.640). This may not be surprising however, considering all cell lines show the same higher expression of isoform 1.

To address whether the higher level of the isoform 1 transcript in cell lines translates to the protein level, western blotting was carried out using an antibody raised to the C-terminal S1 domain of DIS3 (Figure 6.11A), a region common to both isoforms. Figure 6.11B and C demonstrate a consistent higher level of the larger isoform 1 protein across all cell lines tested, corroborating the qPCR data. However, the ratio of isoform 1 to isoform 2 at the mRNA level does not necessarily correlate with the ratio at the protein level (Figure 6.11D, r=-0.067, p=0.88).

6.7 AML and myeloma patients have a higher expression of isoform 1 whereas CMML patients show a pattern more similar to healthy controls

To address the preliminary findings from the RT-PCR that CMML patients appeared to have a higher expression level of isoform 2, rather than isoform 1 like the other disease types, qPCR was employed using isoform-specific Taqman primer probes as above (Figure 6.9A). In agreement with the RT-PCR findings, all myeloma and AML patients were shown to have a higher expression level of isoform 1 using this quantitative method (Figure 6.12, A, B). In contrast five of the ten CMML patients tested had higher isoform 2 levels and two were borderline equal (Figure 6.12 C). As CMML is a malignancy of monocytes, this begs the question of whether this expression pattern is monocyte specific. To test this, PBMCs were isolated from
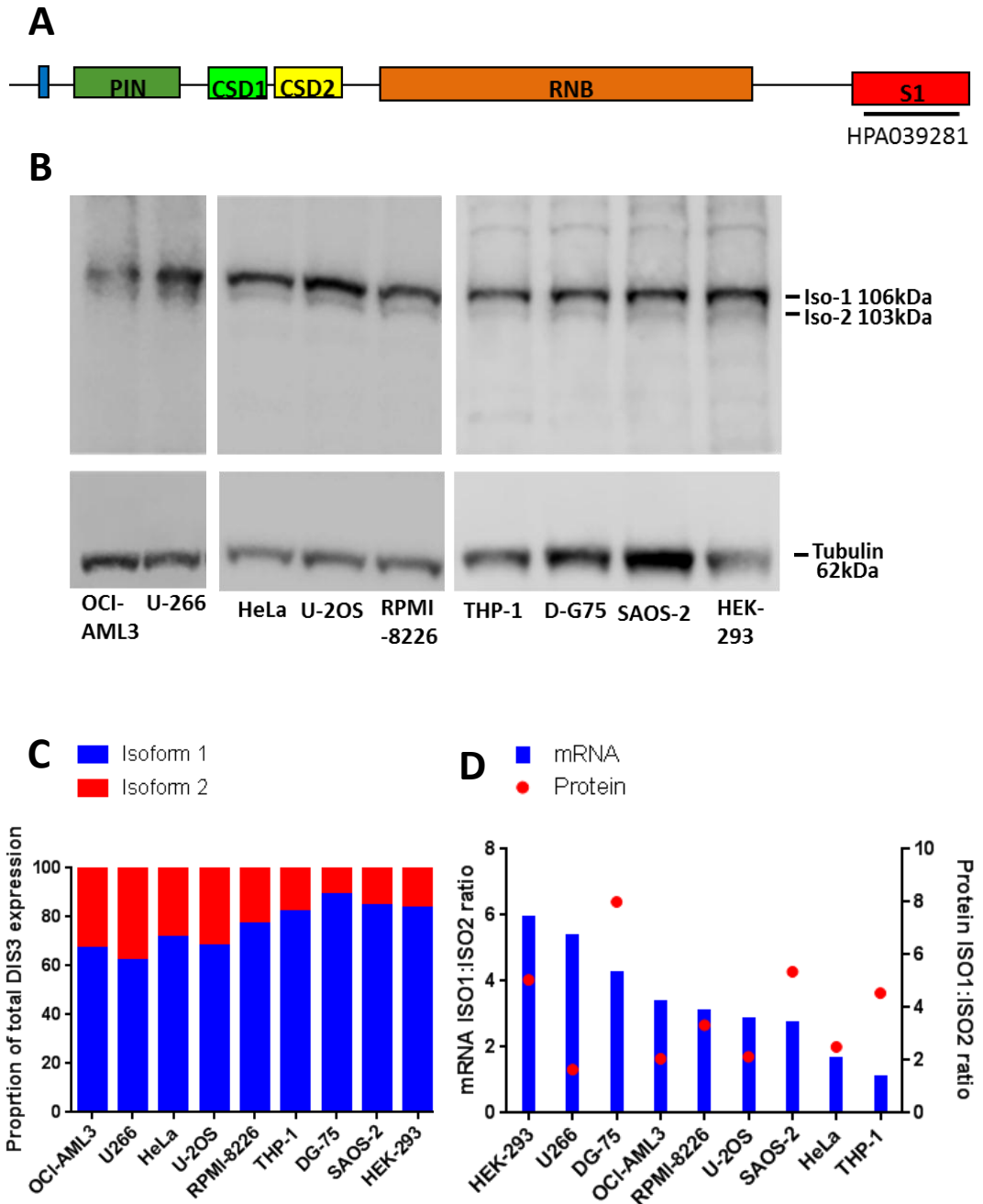
Figure 6.11. Western blot showing the expression levels of the DIS3 isoforms at the protein level in a range of cell lines. (A) Schematic of the DIS3 protein showing the position of antibody binding in the S1 domain. (B) Western blot in ten cell lines showing two bands corresponding to the two isoforms. Tubulin used as a loading control. (C) Quantitation of isoform protein levels expressed as a proportion of total DIS3 levels. (D) The levels of the two isoforms at the mRNA level in ten cell lines does not correlate with the levels at the protein level (r=-0.067, p=0.88).
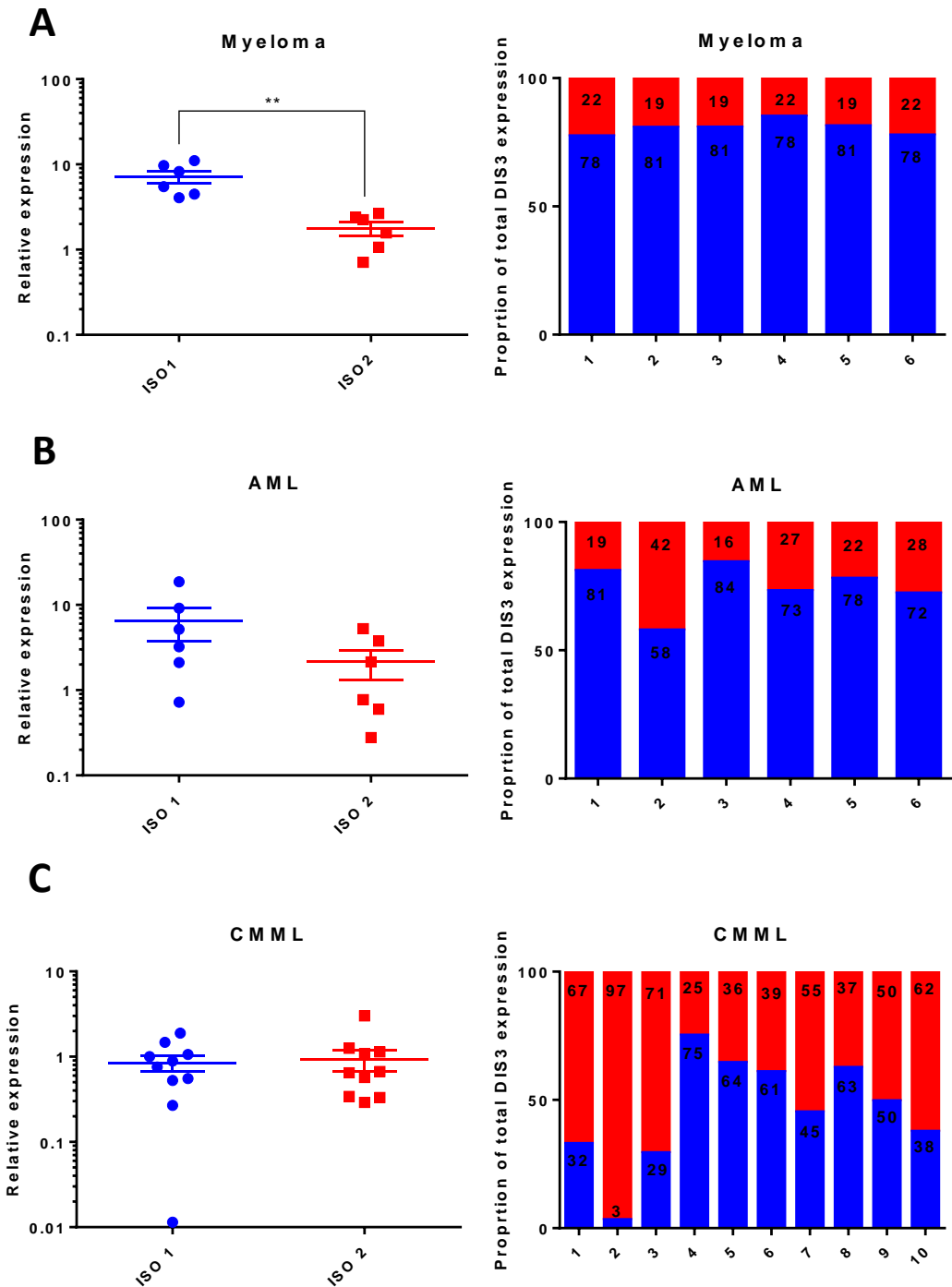
Figure 6.12. Relative expression of the two DIS3 isoform transcripts in primary cells from three different haematological malignancies. (A) Expression levels of isoform 1 is consistently significantly higher than isoform 2 in myeloma patients (p = 0.0018). Isoform 1 and isoform 2 account for approximately 80% and 20% of the total DIS3 mRNA levels respectively. (B) Isoform 1 is higher than isoform 2 in AML patients, contributing approximately 70% of the total DIS3 levels. (C) In CMML patients, levels of isoform 1 and 2 are approximately equal when patients are viewed as a whole. Within individual patients isoform 2 levels are often higher than isoform 1.

healthy individuals before being separated into monocyte and lymphocyte fractions. Monocytes were isolated using the adherence method and anti-CD14 staining was used to test the purity (Figure 6.13A), as well as confirmation of IL-6 upregulation upon LPS activation (Figure 6.13B). IL-6 is a marker of monocyte activation and therefore provides confirmation of a monocyte population. Lymphocytes remained in suspension providing an easy and effective method to isolate the two cell fractions.

The qPCR demonstrates that in healthy monocytes, the levels of isoform 1 and 2 are more equal, with some individuals having higher levels of isoform 2, and many borderline equal levels (Figure 6.13 C, D), paralleling the expression pattern seen in CMML patients. Nevertheless, this is not monocyte specific as healthy lymphocytes also show the same pattern (Figure 6.13 C, E). This suggests that CMML patients are more similar to healthy controls in their expression ratio of the two DIS3 isoforms than AML and myeloma patients. Extending these findings, when the ratio of isoform 1 to isoform 2 within the three disease are compared to healthy cells, myeloma and AML are significantly different (p = 0.0012 and 0.0002 respectively) whereas CMML patients are not (p = 0.703) (Figure 6.14A). Therefore, in contrast to all cell lines, AML and myeloma patients, CMML patients and healthy controls appear to have a higher expression of isoform 2 than isoform 1. The significance of this to cancer phenotypes can only be speculated at this stage.

Interestingly unlike the cell lines, there is a significant positive correlation between the relative level of the two isoforms and the total level of DIS3 expressed in these primary cells (Figure 6.14B, r = 0.941, p = 0.017). Examination of Figure 6.12 shows this appears to be due to a reduced relative expression of isoform 1, rather than an increase in the level of isoform 2. This raises the question of whether the isoform ratios observed are as a result of the cell regulating the expression of isoform 1 specifically to change the stoichiometry of the two isoforms or as a general means of controlling total DIS3 levels. It may be that the higher expression of isoform 2 than isoform 1 within CMML and healthy individuals is an artefact of reduced total DIS3 levels and vice versa in AML and myeloma. However, although this may be the case between disease types, this relationship is not present when the data from

A

Cell fraction	CD14+
Lymphocytes	1/100
Monocytes	52/100

B

IL-6

C

D

ISO 2
ISO 1

Healthy Monocytes

E

Healthy Lymphocytes
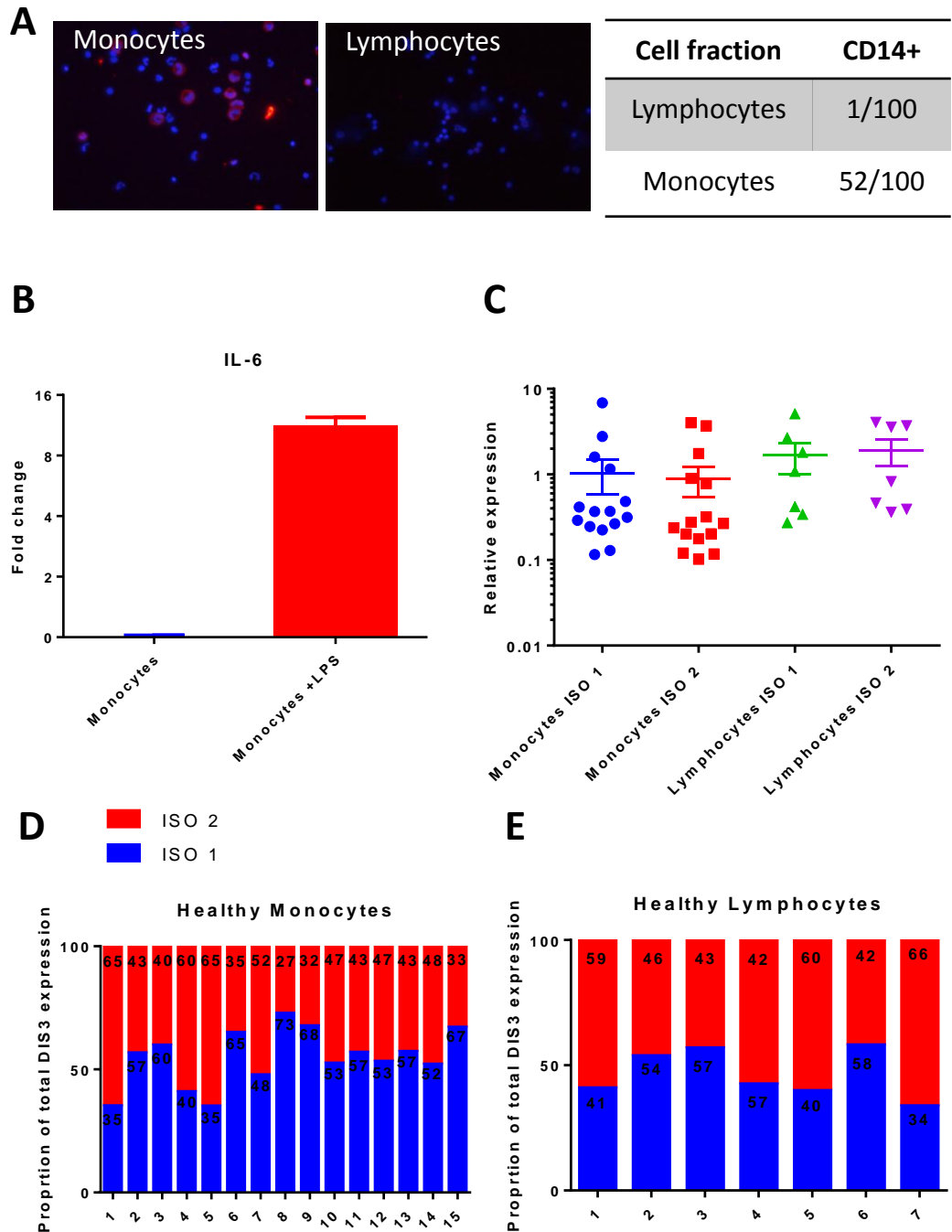
Figure 6.13. Monocytes and lymphocytes isolated from healthy individuals show approximately equal levels of isoform 1 and 2. (A) Anti-CD14 staining of monocytes and lymphocytes showing a 52-fold enrichment of monocytes in the monocyte fraction and depletion of monocytes from the lymphocyte fraction. (B) Cells stimulated with LPS showed upregulation of IL-6, confirming the presence of monocytes. (C) qPCR shows that monocytes and lymphocytes isolated from healthy individuals show approximately equal levels of isoform 1 and 2. (D,E) Isoform 1 and 2 contribute approximately equally to total DIS3 levels within individual patients.

224
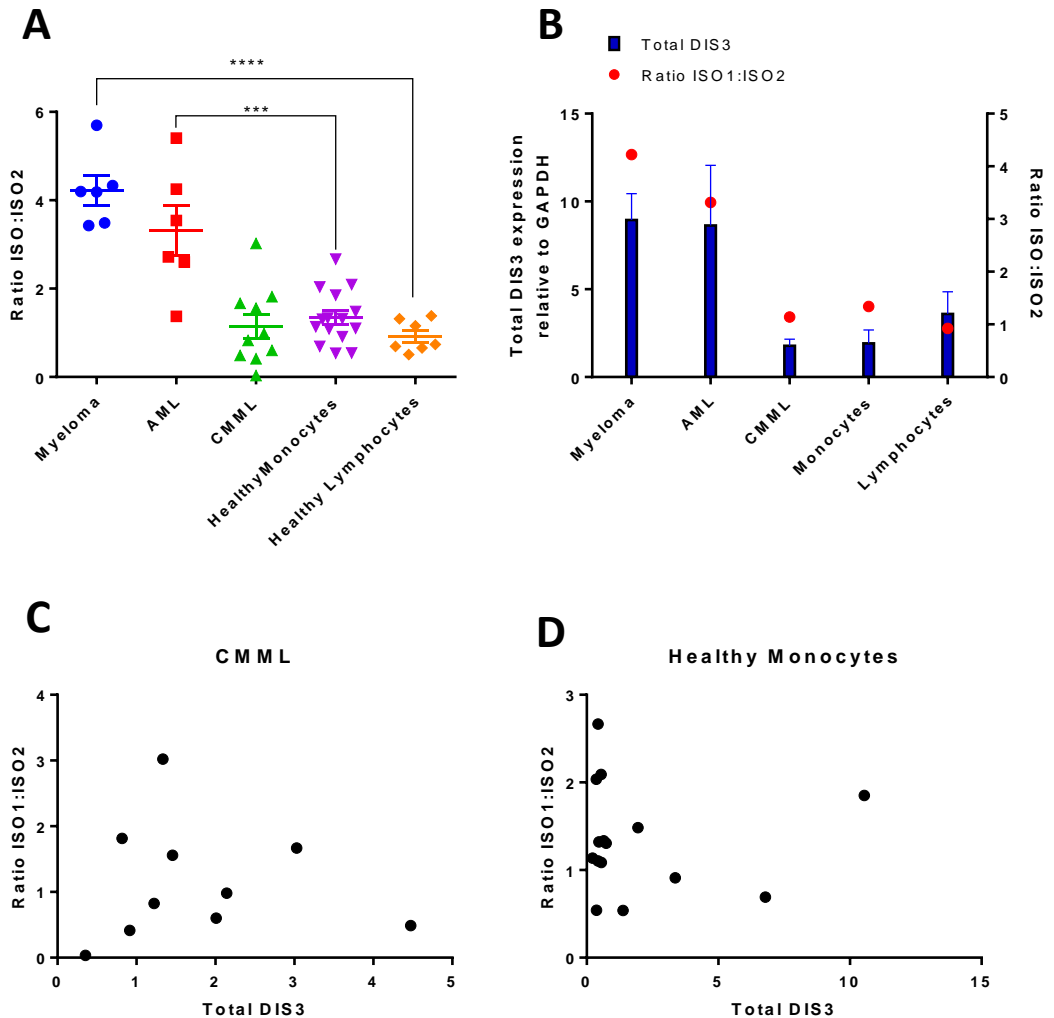
Figure 6.14. Ratio of isoform expression levels versus total DIS3 levels. (A) The ratio of isoform 1 to isoform 2 is significantly higher between myeloma and healthy lymphocytes (p = 0.0012) as well as AML and healthy monocytes (p = 0.0002) but not between CMML and healthy monocytes (p = 0.703). (B) There is a positive correlation between ratio iso1:iso2 and total DIS3 levels across the disease types (r = 0.941, p = 0.017). (C) Within CMML patients and healthy monocytes (D) there is no correlation between ratio iso1:iso2 and total DIS3 levels.

within individual disease types is examined (Figures 6.14 C, E), suggesting the former is more likely within disease types.

6.8 Isoform expression ratios correlate with disease severity in CMML patients

Given that we can be fairly confident that within disease types, the ratio of expression of the two DIS3 isoforms is unlikely to be an artefact of reduced total DIS3 levels, it was of interest to find out whether isoform expression ratios correlate with disease severity. This was measured using the clinical parameters of blast, plasma cell and monocyte count for AML, myeloma and CMML respectively. Although a crude measure of disease severity alone, these cell counts are one of the parameters used to diagnose disease. Figure 6.15 shows that although the sample sizes are very small, initial data does not demonstrate this to be the case for myeloma (r=0.402, p=0.502) and AML (r = 0.122, p= 0.844). CMML however shows a significant negative correlation between percentage monocytosis and isoform expression (Figure 6.15C). That is, CMML patients with a higher level of isoform 2 compared to isoform 1, or a smaller iso1:iso2 ratio, have a higher number of monocytes in their blood (r= -0.9, p=0.002).

Given this finding, it was of interest to test which of these CMML patients had the common oncogenic mutation in the splicing factor SRSF2. SRSF2 is a member of the serine/arginine rich (SR) protein family and contributes to splicing through binding to exonic splicing enhancers (ESE) within the pre-mRNA. SRSF2 is mutated in approximately 50% of CMML patients and is associated with adverse clinical outcome (Kim et al., 2015). The hotspot mutation at P59 has recently been found to alter SRSF2 recognition of specific splicing motifs (Figure 6.16). SRSF2 has a consensus motif of SSNG where S represents either C or G. In SRSF2 mutant cells, exons that were promoted exhibited enrichment for CCNG and depletion for GGNG versus exons that were repressed. In other words, mutant SRSF2 appears to preferentially bind CCNG and not GGNG. Interestingly, despite being a third of the length, exon 2β of isoform 2 contains more of these SRSF2 mutant-enriched motifs than exon 2α of isoform 1 (Figure 6.17). This may suggest the SRSF2 mutation
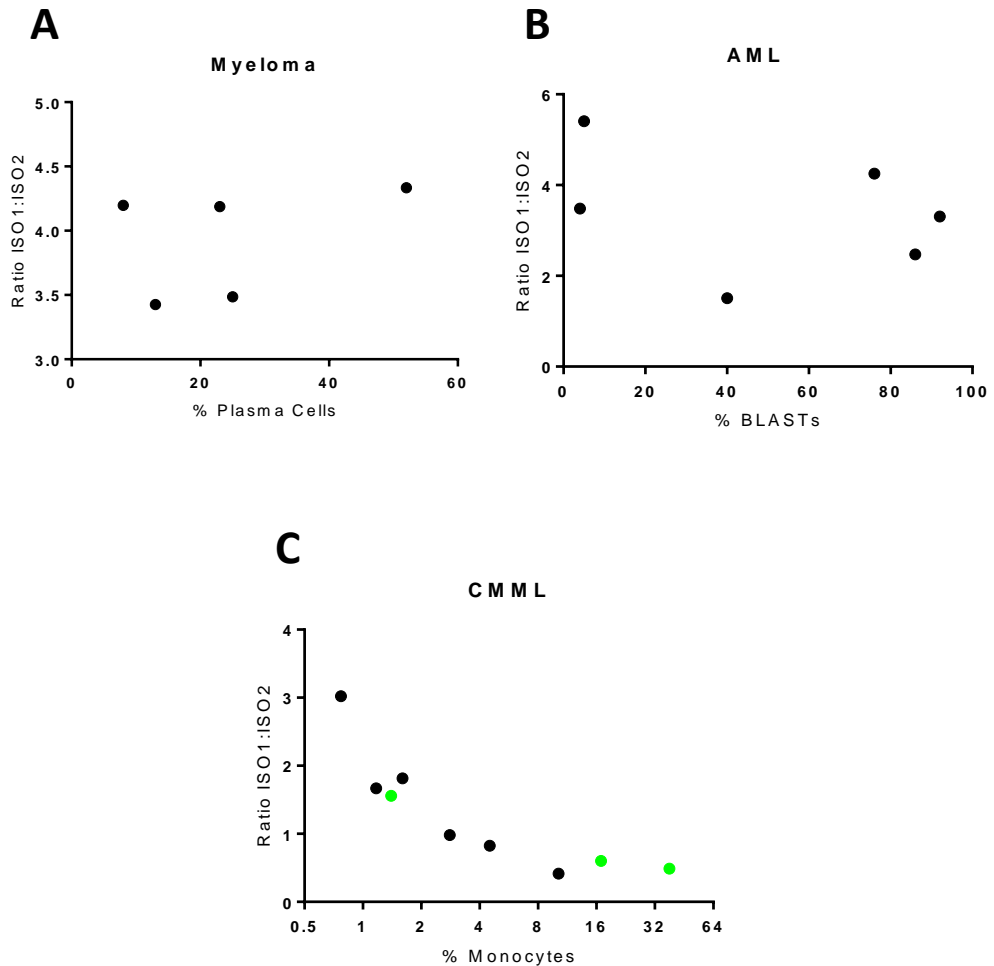
Figure 6.15. Correlation between cell count and isoform expression ratio in three haematological malignancies. The ratio of isoform 1 to isoform 2 does not correlate with plasma cell or blast count in (A) myeloma (r = 0.402, p = 0.502) or (B) AML respectively (r = 0.122, p = 0.844). (C) In CMML however, a significant negative correlation exists between ratio iso1:iso2 and monocyte count (r = -0.9, p = 0.002). Green dots represent patients with SRSF2 mutations.

227
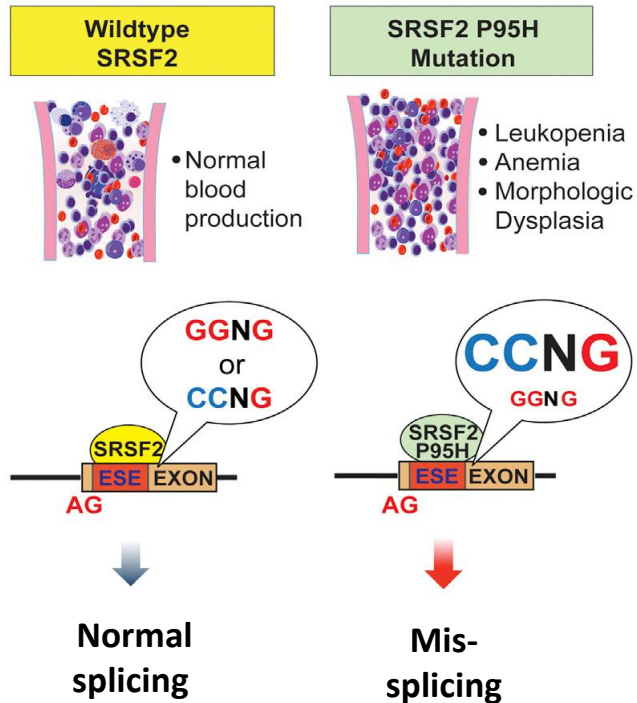
Figure 6.16 SRSF2 mutations in CMML alter SRSF2's sequence-specific RNA binding activity, leading to mis-splicing of key haematopoeitc regulators. Wild-type SRSF2 recognises the consensus motifs GGNG and CCNG within exonic splicing enhancers (ESE) leading to normal splicing and normal blood production. The P59 mutation, observed frequently in CMML patients, leads to altered recognition of splicing motifs by SRSF2. Mutant-SRSF2 preferentially binds CCNG over GGNG causing mis-splicing of targets and leading to clinical features such as leukopenia, anaemia and morphologic dysplasia. Adapted from *Kim et al, 2015.*

Isoform 1 Exon 2α

1   atgttcttga ggac**cctg**cc atcaggaatg taattgtgct acaaacagtt   **50**
51 cttcaagaag tgagaaatc**g cag**tgcc**ccc g**tatataaac gcatccgaga **100**
101 tgtgactaat aaccaagaga agcatttcta tactttcact aatgagcacc **150**
151 atag   **154**

Isoform 2 Exon 2β

1   taagtg**cctg** gagg**ccg**gg a**cctg**ggctt ctgtggcct**c cagcctg**cga **50**
51 ctc**ccaggca g**ctt   **64**

	SRSF2 mutant enriched motifs	SRSF2 mutant depleted motifs
Isoform 1	3	1
Isoform 2	7	2

Figure 6.17. Identification of mutant SRSF2 binding-motifs present in the alternatively spliced exon 2 of the two DIS3 isoforms. As discovered by *Kim et al 2012*, mutant SRSF2 preferentially binds CCNG and GCNG (enriched motifs) rather than GGNG and CGNG (depleted motifs). Exon 2β of isoform 2 possesses 4 more of these preferential motifs than isoform 1, despite its shorter length. Mutant SRSF2 binding motifs (enriched motifs) highlighted in purple.

biases for isoform 2 expression. To test this, PCR was performed to amplify this region of SRSF2 in all CMML patients (Figure 6.18A) before sending the amplicon off for sequencing. Figure 6.18B shows that three patients appear to display the hotspot mutation, represented as green dots on the graph in Figure 6.15C. Interestingly, two of these patients do have higher levels of isoform 2 and concomitantly, the highest level of monocytes in their blood. This begs the question of whether SRSF2 mutations select for expression of DIS3 isoform 2 over isoform 1, through altered recognition of splicing motifs, which contributes to CMML pathogenesis. However, at this stage in the study, with such a small sample size, this can only be speculated.

6.9 Biochemical assays indicate reduced endonucleolytic activity of DIS3 isoform 2

Bioinformatic data predicts the shortened PIN domain of isoform 2 to have reduced activity due to a shortened PIN domain and loss of a critical residue needed for catalysis in *S. cerevisiae*, however this needed to be experimentally validated. Expression clones were purchased containing cDNA for each isoform as well as a T7 promoter, GST tag with Tev cleavage site and ampicillin resistance selection marker (Figure 6.19A). The clones were propagated in *E.coli* and five individual colonies were selected before undergoing a diagnostic restriction digest. The online resource Restriction Mapper was used to identify restriction enzymes that cut either side of the isoform variable exon 2 (Figure 6.19B). The restriction digest generated the expected band sizes (Figure 6.19C), therefore PCR was performed to generate products for sequencing. The same primer pair that flank exon 2 was used to amplify both clones of isoform 1 and 2 generating bands that differed in size by 100bp (Figure 6.19D), representing the size difference of exon 2α and exon 2β from isoform 1 and 2 respectively. One clone of each isoform was sent for sequencing for validation and to ensure no missense mutations were present. The sequencing results showed that both isoform clones sequenced show an exact match with their annotated sequences in ENSEMBL (data not shown). In the next step, the isoform
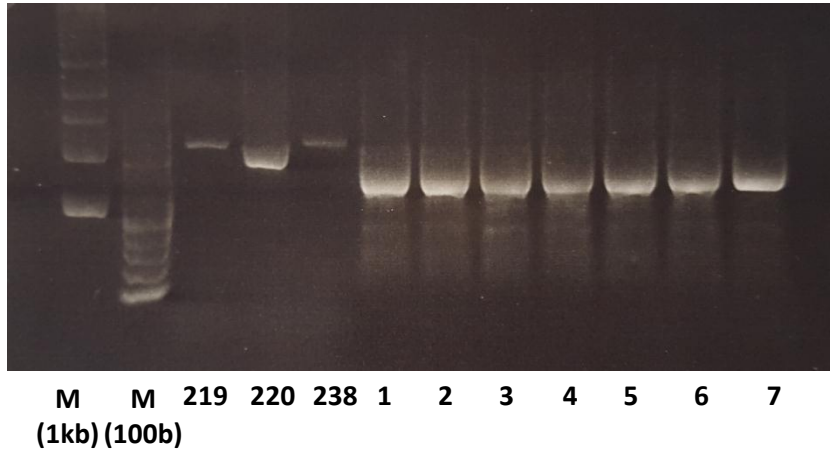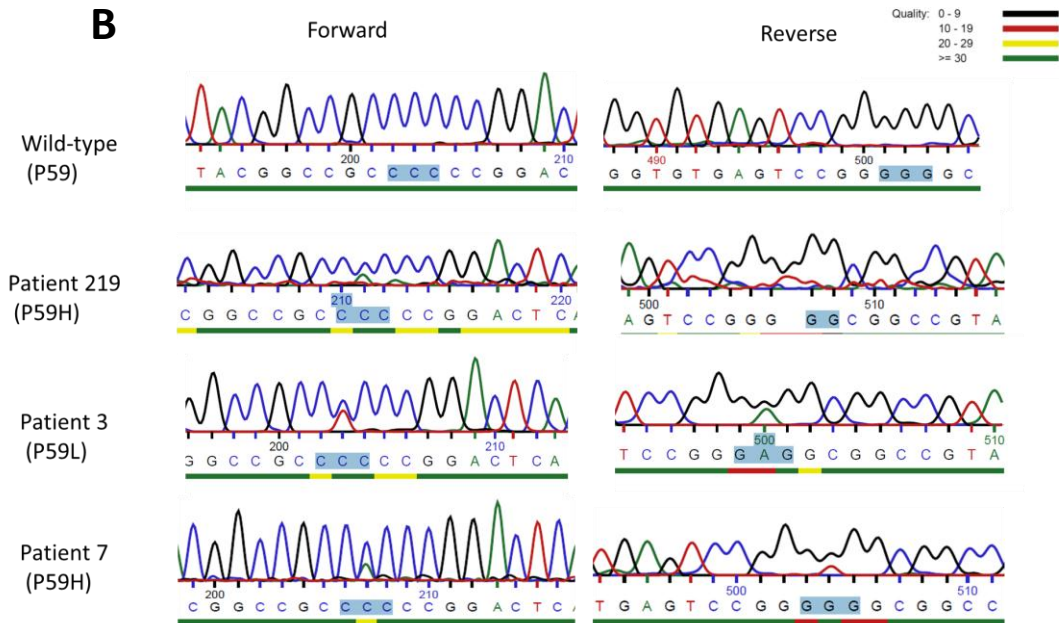
Figure 6.18. Sequencing the SRSF2 gene in CMML patients. (A) PCR amplification of the region for sequencing in the 10 CMML patients. (B) Base calls for the forward and reverse sequencing of the SRSF2 region for the three patients with SRSF2 mutations. The triplet code highlighted in blue encodes proline at position 95, the known hotspot mutation in CMML. The wild-type triplet code is CCC in the forward direction and GGG in the reverse. In three patients, in the forward direction a second peak can be seen above base 2 of the triplet code, indicating a heterozygous mutation. In the reverse direction, the base either cannot be read (patient 219) or is again read as heterozygous (patients 3 and 7).

A

B
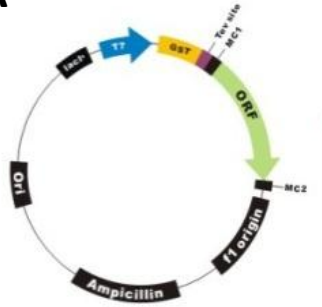
GACGTTCTTAAAAAAGA`CCCGGG`CGGGCGGCGTGATGAAGATCGTGCGCGAGCACTACCT
GCGAGACGACATCGGCTGCGGTGCG`CCCGGG`TGCGCAGCGTGTGGAGGGGCGCACGAG
GGGCCGGCCCTGGAGCCGCAGCCCCAGGACCCGGCGAGCAGCGTCTGCCCGCAACCGCA
CTACTTGCTGCCCGACACTAATGTGTTACTGCACCAGATTGATGTTCTTGAGGACCCTGCCA
TCAGGAATGTAATTGTGCTACAAACAGTTCTTCAAGAAGTGAGAAATCGCAGTGCCCCCGT
ATATAAACGCATCCGAGATGTGACTAATAACCAAGAGAAGCATTTCTATACTTTCACTAATG
AGCACCATAGAGAAACCT//TTGCTATTGATGGTTGGCCCAGAAATTCCA`GATATC`CAAATG

`CCCGGG` = XmaI cut site   `GATATC` = EcoRV cut site

ACCGCACTACTTG = Variable Exon 2

ISO 1 (clone 0869) Expected band sizes = 1203bp and 1135bp

ISO 2 (clone 3667) Expected band sizes = 1103bp and 1035bp (100bp smaller than ISO1 bands)
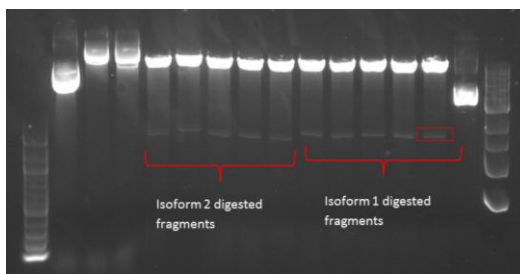
C   **D**
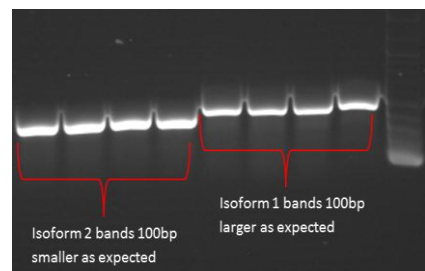
Figure 6.19. Generation of recombinant DIS3 isoform 1 and 2. (A) Diagrammatic representation of the pReceiver-B03 vector containing a T7 promoter for high level expression, GST-tag and Tev cleavage site, ampicillin resistance gene for bacterial selection and multiple cloning sites. (B) Restriction enzyme sites were mapped onto the DIS3 sequence in order to perform a diagnostic digest after proliferation of the vector. (C) Two bands of expected sizes were observed on an agarose gel after restriction digest confirming presence of the correct products. (D) PCR amplification of the two isoform clones for sequencing.

cDNA clones were expressed in the E. coli Rosetta (DE3) strain and protein expression induced using IPTG. The GST-tagged recombinant proteins were purified by affinity chromatography and half were digested with TEV protease to remove the GST-tag (Figure 6.20). After optimisation of the buffer and substrate concentration, RNase activity assays were performed with a Y32P-labelled 30-mer (16-mer + A14) circularised substrate. A circularised substrate was used as the PIN domain has endonucleolytic activity which is required for the cleavage of circularised substrates, after which the linearised substrate can be degraded exonucleolytically by the RNB domain. Four proteins were used: cleaved isoform 1 and isoform 2 as well as GST-tagged isoform 1 and isoform 2. Figure 6.21 shows the results of the assays. After 90 minutes' incubation, more substrate remained in the isoform 2 sample than the isoform 1 sample, indicating a reduced activity of the former. In both GST-tagged protein samples there is less disappearance of substrate than in their cleaved counterparts, suggesting the large GST tag affects enzyme activity. Nevertheless, as with the cleaved samples, after 90 minutes more substrate remained in isoform 2 than isoform 1. Thus, whilst degradation products can be seen in the isoform 2 sample suggesting the protein has not lost endoribonucleolytic activity completely, it does appear to have reduced activity compared to isoform 1.

6.10 The two DIS3 isoforms may undergo differential miRNA-targeting

According to bioinformatic data, isoform 2 also has a shorter 3'UTR than isoform 1 (1380bp vs 7626bp) (Figure 6.3B). The 3'UTR regulates many important processes involved with the expression, stability and targeting of the mRNA within the cell. One of the important functions of the 3'UTR is to regulate the expression level of the transcript via miRNA targeting. The extended 3' UTR of isoform 1 may contain additional binding sites for miRNAs which can decrease the expression of the mRNA. To investigate this, an online miRNA prediction tool devised by the Segal lab of the Weizmann Institute, was used to run the PITA algorithm. PITA scans the 3'UTR for potential miRNA targets using seed matching tools and scores each site. A score of less than -10 is considered a likely miRNA-mRNA interaction site based on a target accessibility algorithm in Kertesz et al, 2007 (Kertesz et al., 2007). Using this

Isoform 1   **Isoform 2**

Iso 1-GST 132kDa
Iso 1 106kDa

Iso 2-GST 129kDa
Iso 2 103kDa

-Tev   **+Tev**   **-Tev**   **+Tev**
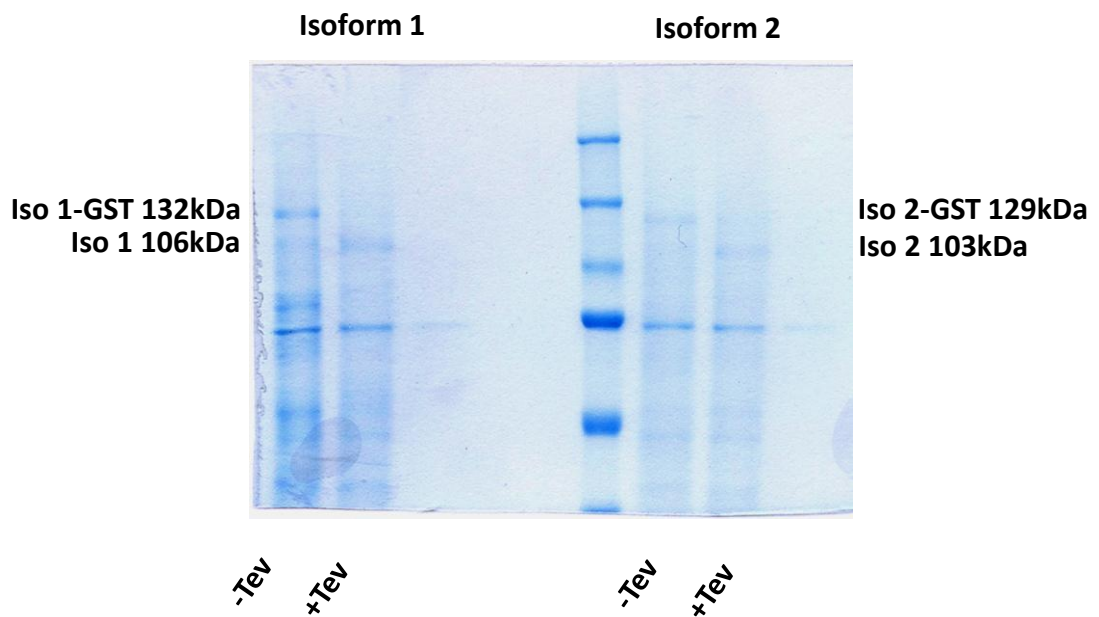
6.20 Coomassie blue-stained PAGE gel of purified DIS3 isoform proteins.
Lanes 1 and 5 represent purified isoform 1 and 2 respectively containing
GST tags. Lanes 2 and 6 represent Tev-cleaved isolates to remove the GST-
tag from isoform 1 and 2 respectively.
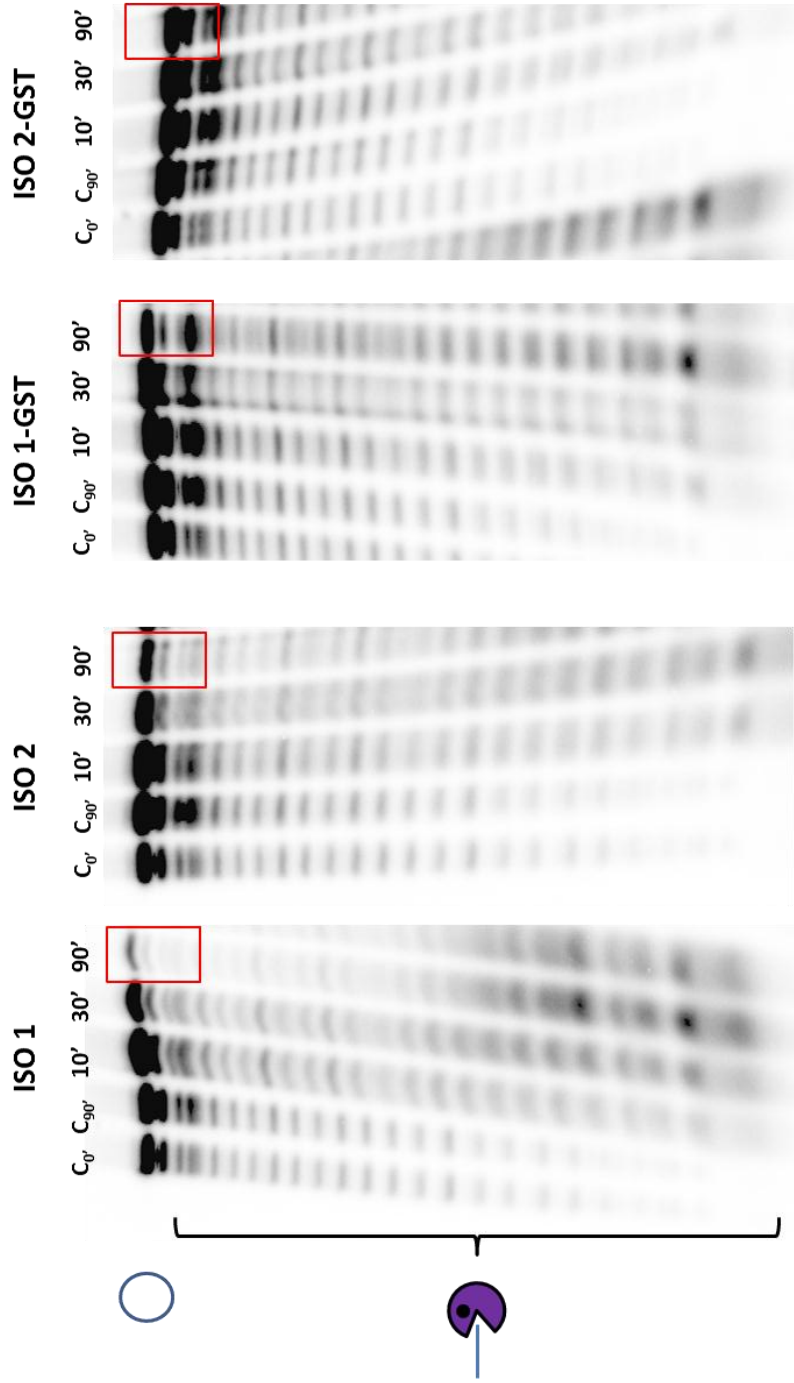
Figure 6.21. RNase activity assays of DIS3 isoform 1 and isoform 2. A Y32P-labelled 30-mer (16-mer + A14) circularised substrate was incubated for 10, 30 or 90 minutes with each of four proteins: Tev-cleaved isoform 1, Tev-cleaved isoform 2, GST-tagged isoform 1 and GST-tagged isoform 2. Samples were run on a polyacrylamide gel for 4 hours. No-enzyme controls were used and left for 0 and 90 minutes. Circular substrates are visualised at the top of the gel and degradation products as a ladder down the length of the gel.

235

algorithm, isoform 1 was predicted to contain 19 additional likely miRNA target sites compared to isoform 2 (Table S1). Considering our findings that isoform 1 is usually more highly expressed than isoform 2, these additional miRNA binding sites do not appear to be simply acting to decrease the expression level of isoform 1 but instead may provide specific temporal and spatial control of translation.

6.11 Discussion

DIS3 is currently a protein of interest due to its apparent role in the development of multiple myeloma, yet no studies exist on the alternative splicing of DIS3. This chapter presents novel information on the existence of two protein-coding isoforms of DIS3. As we have seen, the two isoforms differ in the size of their PIN domain through the use of a mutually exclusive second exon. Structural analysis predicts this to be highly destabilising to the catalytic activity of isoform 2. Bioinformatics also predicts isoform 2 to have a shorter 3'UTR.

I have experimentally validated the presence of the two DIS3 isoforms and the ubiquitous expression of both in a range of cell types suggests neither are solely tissue specific. This does not necessarily mean that the levels of either isoform does not change in specific tissues, simply that they are both at least present in the cell types I have tested. In all cell lines tested, isoform 1 is expressed at a higher level than isoform 2 at both the mRNA and protein level. As isoform 1 contains the intact and functional PIN domain, this suggests that for the most part, most cells require the endonucleolytic activity of DIS3. Interestingly the level of the isoforms at the mRNA level does not necessarily translate to the protein level according to quantitation of the western blot. This could be caused by post-transcriptional regulation or it could be due to a technical reason such as the difficulty quantifying western blots. Alternatively, although the correct size, it may be that the smaller band visible on the western blot does not correspond to isoform 2. To confirm this antibody needs to be validated using purified proteins.

Isoform 1 is also the principal transcript in AML and myeloma patients, making up approximately 80% of total DIS3 levels. However, in CMML, a malignancy of

monocytes, the opposite is observed with the average relative expression of isoform 2 higher than that of isoform 1. This pattern was found not to be CMML or monocyte specific however, as both healthy monocytes and healthy lymphocytes show relatively equal expression levels of the two isoforms.

The expression level of the two isoforms was found to correlate with total DIS3 levels across the five disease types, suggesting higher levels of isoform 1 may be an artefact of increased total DIS3 levels rather than the cell changing the stoichiometry of the two isoforms. However, the same correlation was not seen in patients within one disease type, suggesting any difference in the level of the isoforms observed within each disease type is a result of the cell modulating their stoichiometry. Given this finding, it was observed that the ratio of the two isoforms correlates with disease severity in CMML patients. That is, as the level of isoform 2 increases compared to isoform 1, the level of monocytosis increases. Given that CMML is commonly associated with mutations in the splicing factor SRSF2, it was interesting to find that the number of SRSF2-mutant recognised motifs was higher in isoform 2. In support of this, although the sample size is small, two of three of the patients found to have SRSF2 mutations had the highest level of isoform 2 of the group. This suggests the mutation of SRSF2 may bias for isoform 2 expression, potentially contributing to CMML oncogenesis.

In the final part of this study, I showed using RNase activity assays that DIS3 isoform 2 has reduced but not abolished endonucleolytic function. Given that the PIN domain is also involved in tethering DIS3 to the exosome scaffold complex, it is interesting to speculate why the cell expresses both these isoforms and to different levels in different disease types. As tethering to the exosome has been shown to dampen DIS3 activity, it may be that the specific expression of isoform 2 provides a mechanism by which to enhance enzyme activity. Reduction in PIN activity may be important for processing of particular RNA substrates.

According to bioinformatic data, isoform 2 also has a shorter 3'UTR. The 3'UTR regulates many important processes involved with the expression, stability and targeting of the mRNA within the cell. The 3'UTR contains binding sites for miRNAs which can decrease expression of the mRNA as well as binding regions for specific

proteins that can also affect the stability and the localisation of the transcript. Using the PITA algorithm, 19 additional predicted miRNA binding sites are present in the 3'UTR of isoform 1 compared to isoform 2, indicating that isoform 1 is under more regulatory control than isoform 2. Nevertheless, this does not appear to simply reduce the expression of the transcript as our data shows the opposite in that isoform 1 is usually more abundant than isoform 2. Instead, it is possible that isoform 1 undergoes more spatial and temporal regulation through interaction with a range of cis and trans factors. Interestingly, a number of reports have provided evidence for miRNA-enhanced translation of their targets in certain conditions rather than miRNA-mediated repression. Furthermore, some AU-binding proteins have also been discovered to enhance mRNA half- life and protein translation. In all, a complex interplay between a range of 3'UTR factors on the longer 3'UTR of isoform 1 may result in differential stability or localisation of the two isoforms under certain conditions. As isoform 2 has a reduced endonucleolytic function and potentially tethering to the exosome, this may provide a means for the cell to regulate the activity of DIS3 in particular contexts.

Chapter 7: Summary and Future Work

7.1 Summary of main findings

At the onset of this study, it was unknown how DIS3 mutations, observed recurrently in multiple myeloma, contribute to tumourigenesis nor how to therapeutically target the cells harbouring these mutations. Furthermore, there was no knowledge on the isoforms of DIS3, nor their differential function or relevance to disease. The work presented in this thesis has shed light on the possible mechanisms by which DIS3 mutations confer an advantage to cancer cells, as well as possible therapeutic targets, in addition to characterising the two DIS3 protein-coding isoforms.

7.1.1 DIS3 knock-down does not affect the phenotype of human cells

In chapter 3, prior to generating a DIS3 knock-down model, computational analysis was used to characterise DIS3 as a tumour suppressor rather than oncogene in multiple myeloma. Subsequently, transient siRNA and stable shRNA knock-down of DIS3 were used in myeloma and osteosarcoma cells respectively, in an attempt to identify a phenotype which may shed light on the role of DIS3 in myeloma. Based on previous findings from DIS3 mutant organisms, many aspects of cell phenotype were investigated, however no difference was observed between knock-down and controls. This lack of phenotype led to a number of speculations, both biological and technical. Hypotheses based on technical limitations included insufficient depletion of DIS3, functional compensation by another ribonuclease, as well as the limitations of using immortalised cell lines which already possess malignant characteristics. Biological hypotheses for a lack of phenotype included the possibility that DIS3 promotes pathogenesis by affecting the bone marrow microenvironment, something that would not be observed in *in vitro* experiments. Similarly, DIS3 mutations may not drive tumorigenesis on their own but may require another cellular pathway to be disrupted, or are only required to maintain

the tumour and not initiate it, both of which would not yield a phenotype *in vitro*. Finally, there is a possibility that DIS3 mutations have no role in cancer development but have not undergone negative selection because they are not deleterious to cancer cells. Therefore, although this chapter was unable to pinpoint the exact mechanism of how DIS3 contributes to myeloma, it was successful in shedding light on a number of potential possibilities that may not have previously been considered.

7.1.2 RNA-seq identifies histone proteins as potential DIS3 targets

In chapter 4, high-throughput transcriptome profiling was performed using RNA-seq in order to identify potential targets of DIS3 that may help reveal its role as a tumour suppressor in myeloma. After a stringent filtering process, 207 transcripts were identified as confidently upregulated and 190 as confidently downregulated by more than 2-fold upon DIS3 knock-down. Gene ontology analysis using the online tool DAVID, identified five significantly enriched pathways that may be regulated by DIS3 including: translation, small ncRNA processing, RNA splicing, chromatin assembly and post-transcriptional gene regulation. As many of these genes are involved in RNA processing, their upregulation may indicate a compensatory mechanism for loss of DIS3. Eleven of the most upregulated genes were selected for validation by qPCR based on fold-change and Taqman assay availability. Unfortunately, although the majority changed in the correct direction, none of these genes showed a significant differential expression by qPCR. The possible reasons for this included: limitations of the transcript quantification calculation by CuffDiff; natural or technical variation within the cell line; as well as insufficient precision of qPCR to detect low fold-changes using a small sample size. Nevertheless, although the changes were not statistically significant due to a small sample size, two of the histone genes showed an average fold-change of above 2, indicating that with a larger sample size, these genes may be discovered to be real targets of DIS3. The enrichment of histone proteins by the GO analysis also supported this. This could be an interesting finding as histone expression is tightly linked to cell cycle progression and disruption of the mechanisms that maintain

histone levels leads to chromosomal instability. In addition to the fact that most DIS3 mutants display cell-cycle defects, myeloma is a malignancy characterised by chromosomal instability. Therefore, it is feasible to speculate that misregulation of histones by an aberrant DIS3 may be a mechanism underlying the recurrence of DIS3 mutations in myeloma.

7.1.3 DNA Damage Response genes do not appear to confer synthetic lethality alongside DIS3

Chapter 5 sought to identify synthetic lethal interactors of DIS3 with the dual aim of shedding light on the role of mutations in myeloma as well as yielding potential tumour-specific therapeutic targets. A DNA damage response screen was performed using a library of 240 siRNAs. Using an arbitrary cut off of two standard deviations above or below the mean, one siRNA, NABP1 was found to confer synthetic lethality with loss of DIS3 and two, PRKCG and GTF2H3, conferred a synthetic growth advantage. However, upon validation of the NABP1 hit using a different set of siRNAs, no such synthetic lethality was observed. This may indicate off-target effects of one or more of the siRNAs in the initial pool or may have been a result of using a different passage of cells. Nevertheless, this chapter discussed a novel synthetic lethality screening strategy whereby control and treatment cells labelled with different fluorescent tags and co-cultured and their relative growth monitored relative using fluorescence spectroscopy.

7.1.4 DIS3 has two protein-coding isoforms that show differential expression and function

The focus of chapter 6 was to characterise the two DIS3 protein-coding isoforms and investigate their expression, function and relevance to disease. Both isoforms were found to be translated and ubiquitously expressed across a range of cell types

demonstrating their biological relevance. Isoform 1 was identified as the principal transcript in all cell lines tested, as well as in primary cells from AML and myeloma patients. However, isoform 2 was more highly expressed in CMML patients whilst healthy controls showed equal levels of the two. The ratio of isoform expression was found to correlate with the percentage monocytosis in CMML patients, an indicator of disease severity. Given that CMML is commonly associated with mutations in the splicing factor SRSF2, it was interesting to find that the number of SRSF2-mutant recognised motifs was higher in isoform 2. In support of this, although the sample size was small, two of three of the patients found to have SRSF2 mutations had the highest level of isoform 2 of the group. This suggested that the mutation of SRSF2 may bias for isoform 2 expression, potentially contributing to CMML oncogenesis. RNase assays showed that DIS3 isoform 2 has reduced endonucleolytic PIN activity and bioinformatics analysis identified additional miRNA binding sites within the longer 3'UTR of isoform 1, potentially revealing a mechanism of differential regulation of the two isoforms.

7.2 Future work

The results presented in this thesis indicate a number of leads that can be followed to further investigate the function of DIS3 in myeloma as well as the significance of the two DIS3 isoforms to disease.

7.2.1 Investigating the biological function of DIS3 in mammalian cells

In order to further investigate the biological function of DIS3 in mammalian cells, DIS3 could be knocked-down alongside Rrp6, the other exosome-associated nuclear exoribonuclease, that may otherwise compensate for the loss of DIS3 and be responsible for the lack of phenotype observed in this study. Alternatively, in order to recapitulate the situation *in vivo,* a better option would be to generate a DIS3 mouse mutant which would answer the question of whether DIS3 acts through altering the bone marrow microenvironment. If DIS3 knock-out was found to be lethal, a conditional-knock-out could be generated to deplete DIS3 only from the

haematopoietic lineage. Nevertheless, if mutation of DIS3 is not driving tumourigenesis but is instead cooperating with another disrupted pathway, then a functional phenotype will be more difficult to observe. More sequencing data will be required to identify aberrations that co-occur with mutations in DIS3, in order to identify interacting partners. A reverse genetic approach can then be employed where both genes are depleted simultaneously, allowing for a more accurate emulation of the situation *in vivo.*

7.2.2 Identifying DIS3 targets

Chapter 4 identified the importance of experimental design when performing transcriptome profiling experiments such as RNA-seq. Due to the difficulties that were encountered with performing the RNA-seq experiment in duplicates, future work would ideally use triplicates in order for a statistical analysis to be carried out and significantly differentially expressed genes identified. Additionally, due to the presence of multiple transcripts encoded by single genes in humans, validation of RNA-seq data should be restricted to genes whose fold-change is calculated from the sum of all the transcript changes rather than individual transcripts. This is because CuffLinks generates an estimate of isoform abundance and thus calculations of isoform expression can be inaccurate, especially for lowly expressed transcripts. This can therefore generate false-positive results which will not validate by qPCR. Further biological repeats of the qPCR should be performed in order to verify whether the histone proteins identified in this study are true targets of DIS3. If this is the case, this may reveal a mechanism by which DIS3 mutations contribute to multiple myeloma.

7.2.3 Identifying synthetic lethal interactors of DIS3

Chapter 5 discussed the results of a preliminary screen to identify genes that confer synthetic lethality when knocked-down alongside DIS3 in order to identify potential therapeutic targets. According to the cut-off limit, only one gene, NABP1, conferred synthetic lethality, however two, PRKCG and GTF2H3, appeared to confer a survival

advantage. Although these genes will not lead to a therapeutic target for DIS3 mutated cells, with further investigation they may reveal useful insights into the pathways affected by DIS3 in myeloma. Furthermore, only genes with Z-scores of less than -2 or above 2 were considered for validation; however it would also be useful to test other genes with lower Z-scores as it may be that with for instance, more efficient knock-down, these genes do in fact yield a synthetic lethality. Given the lack of an observable phenotype upon DIS3 knock-down as discussed in chapter 3, it is feasible that DIS3 may normally function to target bone marrow signalling molecules, which become up-regulated when the enzyme is mutated, promoting myeloma pathogenesis. In order to test whether DIS3 affects the bone marrow microenvironment, future work could investigate the interaction of DIS3 with the tumour microenvironment using so-called "contextual synthetic lethality" (Chan, Pires et al. 2010). Finally, further synthetic lethality screens could be performed that target other pathways, in addition to DNA damage, such as RNA processing or chromatin regulation as these pathways were identified as upregulated in the GO analysis in chapter 4.

7.2.4 Characterising the two isoforms of DIS3

To yield further insight into the role of the two DIS3 isoforms in the cells, it would be interesting to test, using the co-immunoprecipitation technique, whether DIS3 isoform 2 associates with the exosome core. As the exosome is known to reduce the activity of DIS3, this may suggest that expression of isoform 2 allows the cell to enhance exoribonucleolytic activity. The reduced PIN function may be part of a trade-off of dissociation from the exosome, rather than a functional feature in itself. Furthermore, if an antibody could be designed that can distinguish the two protein isoforms, it would be interesting to perform immunocytochemistry to assess whether any differential cellular localisation exists. It may be that in specific locations within the cell, PIN activity is detrimental to the processing of particular substrates. Additionally, it may also be of interest to uncover whether either

isoform is upregulated in response to stress conditions or different stages of the cell cycle. Additionally, the individual recombinant isoforms could be over-expressed after depleting endogenous DIS3 and transcriptional profiling could be performed to determine whether either isoform has specific targets.

Further RNase assays should be performed using only the PIN domain to ensure it is this domain which has the differential activity, as well as with linear substrates. Linear substrates are targeted by exonucleolytic activity and should be degraded equally well by both isoforms, providing a positive control. Finally, it would be interesting to obtain more CMML patients to genotype for SRSF2 mutations and find out whether these correlate with isoform expression. If this is found to be the case, it may be that one of the functions of these common mutations within this splicing factor is to increase the expression of isoform 2.

7.3 Concluding Remarks

Through a combination of phenotypic investigations and identification of potential DIS3 targets, this thesis has helped shed light on the potential mechanisms by which DIS3 mutations may contribute to multiple myeloma, as well as identifying synthetic lethal interactions that may lead to tumour-specific therapies. Aside from investigating the function of DIS3 in myeloma, this work has also characterised two novel protein-coding isoforms of DIS3, whose expression may have a relevance to CMML pathogenesis. Therefore, the ribonuclease DIS3 may be important in multiple types of haematological cancer and only by understanding its fundamental biology can we expect to gain insight into its role in human disease.

References

Abe, M., Hiura, K., Wilde, J., Shioyasono, A., Moriyama, K., Hashimoto, T., Kido, S., Oshima, T., Shibata, H., Ozaki, S., Inoue, D. & Matsumoto, T. 2004. Osteoclasts enhance myeloma cell growth and survival via cell-cell contact: a vicious cycle between bone destruction and myeloma expansion. *Blood,* 104(8)**,** pp 2484-91.

Allen, M. J., Myer, B. J., Khokher, A. M., Rushton, N. & Cox, T. M. 1997. Pro-inflammatory cytokines and the pathogenesis of Gaucher's disease: increased release of interleukin-6 and interleukin-10. *QJM,* 90(1)**,** pp 19-25.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. & Tollervey, D. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *Embo J,* 18(19)**,** pp 5399-410.

Altieri, A., Chen, B., Bermejo, J. L., Castro, F. & Hemminki, K. 2006. Familial risks and temporal incidence trends of multiple myeloma. *Eur J Cancer,* 42(11)**,** pp 1661-70.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F. O., Jørgensen, M., Andersen, P. R., Bertin, N., Rackham, O., Burroughs, A. M., Baillie, J. K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C. J., Meehan, T. F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C. O., Heutink, P., Hume, D. A., Jensen, T. H., Suzuki, H., Hayashizaki, Y., Müller, F., Forrest, A. R., Carninci, P., Rehli, M., Sandelin, A. & Consortium, F. 2014. An atlas of active enhancers across human cell types and tissues. *Nature,* 507(7493)**,** pp 455-61.

Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P. & Lis, J. T. 2002. The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature,* 420(6917)**,** pp 837-41.

Arnaoutov, A. & Dasso, M. 2003. The Ran GTPase regulates kinetochore function. *Dev Cell,* 5(1)**,** pp 99-111.

Bachellerie, J. P., Cavaillé, J. & Hüttenhofer, A. 2002. The expanding snoRNA world. *Biochimie,* 84(8)**,** pp 775-90.

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L. A., Hart, R. P. & Kiledjian, M. 2010. Differential regulation of microRNA stability. *RNA,* 16(5)**,** pp 1032-9.

Bakheet, T., Williams, B. R. & Khabar, K. S. 2006. ARED 3.0: the large and diverse AU-rich transcriptome. *Nucleic Acids Res,* 34(Database issue)**,** pp D111-4.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J. & Frey, B. J. 2010. Deciphering the splicing code. *Nature,* 465(7294)**,** pp 53-9.

Barbas, A., Matos, R. G., Amblar, M., López-Viñas, E., Gomez-Puertas, P. & Arraiano, C. M. 2009. Determination of key residues for catalysis and RNA cleavage specificity: one mutation turns RNase II into a "SUPER-ENZYME". *J Biol Chem,* 284(31)**,** pp 20486-98.

Barrett, L. W., Fletcher, S. & Wilton, S. D. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci,* 69(21)**,** pp 3613-34.

Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* 116(2)**,** pp 281-97.

Bartel, D. P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell,* 136(2)**,** pp 215-33.

Basu, U., Meng, F. L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C. R., Wesemann, D. R., Januszyk, K., Gregory, R. I., Deng, H., Lima, C. D. & Alt, F. W. 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell,* 144(3)**,** pp 353-63.

Beltran, M., Puig, I., Peña, C., García, J. M., Alvarez, A. B., Peña, R., Bonilla, F. & de Herreros, A. G. 2008. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev,* 22(6)**,** pp 756-69.

Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem,* 72(291-336.

Bolli, N., Avet-Loiseau, H., Wedge, D. C., Van Loo, P., Alexandrov, L. B., Martincorena, I., Dawson, K. J., Iorio, F., Nik-Zainal, S., Bignell, G. R., Hinton, J. W., Li, Y., Tubio, J. M., McLaren, S., O' Meara, S., Butler, A. P., Teague, J. W., Mudie, L., Anderson, E., Rashid, N., Tai, Y. T., Shammas, M. A., Sperling, A. S., Fulciniti, M., Richardson, P. G., Parmigiani, G., Magrangeas, F., Minvielle, S., Moreau, P., Attal, M., Facon, T., Futreal, P. A., Anderson, K. C., Campbell, P. J. & Munshi, N. C. 2014. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun,* 5(2997.

Bousquet-Antonelli, C., Presutti, C. & Tollervey, D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell,* 102(6)**,** pp 765-75.

Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. 2011. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell,* 44(1)**,** pp 120-33.

Bresson, S. M. & Conrad, N. K. 2013. The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet,* 9(10)**,** pp e1003893.

Briggs, M. W., Burkard, K. T. & Butler, J. S. 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem,* 273(21)**,** pp 13255-63.

Brimacombe, R. & Stiege, W. 1985. Structure and function of ribosomal RNA. *Biochem J,* 229(1)**,** pp 1-17.

Broderick, P., Chubb, D., Johnson, D. C., Weinhold, N., Försti, A., Lloyd, A., Olver, B., Ma, Y. P., Dobbins, S. E., Walker, B. A., Davies, F. E., Gregory, W. A., Child, J. A., Ross, F. M., Jackson, G. H., Neben, K., Jauch, A., Hoffmann, P., Mühleisen, T. W., Nöthen, M. M., Moebus, S., Tomlinson, I. P., Goldschmidt, H., Hemminki, K., Morgan, G. J. & Houlston, R. S. 2012. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet,* 44(1)**,** pp 58-61.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J. & Willard, H. F. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell,* 71(3)**,** pp 527-42.

Bryant, D. W., Priest, H. D. & Mockler, T. C. 2012. Detection and quantification of alternative splicing variants using RNA-seq. *Methods Mol Biol,* 883(97-110.

Bryant, H. E., Schultz, N., Thomas, H. D., Parker, K. M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N. J. & Helleday, T. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature,* 434(7035)**,** pp 913-7.

Buck, A. H., Perot, J., Chisholm, M. A., Kumar, D. S., Tuddenham, L., Cognat, V., Marcinowski, L., Dölken, L. & Pfeffer, S. 2010. Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection. *RNA,* 16(2)**,** pp 307-15.

Buhler, M., Haas, W., Gygi, S. P. & Moazed, D. 2007. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell,* 129(4)**,** pp 707-21.

Butler, J. S. 2002. The yin and yang of the exosome. *Trends Cell Biol,* 12(2)**,** pp 90-6.

Cairrão, F., Arraiano, C. & Newbury, S. 2005. Drosophila gene tazman, an orthologue of the yeast exosome component Rrp44p/Dis3, is differentially expressed during development. *Dev Dyn,* 232(3)**,** pp 733-7.

Calin, G. A. & Croce, C. M. 2006. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res,* 66(15)**,** pp 7390-4.

Camps, J., Pitt, J. J., Emons, G., Hummon, A. B., Case, C. M., Grade, M., Jones, T. L., Nguyen, Q. T., Ghadimi, B. M., Beissbarth, T., Difilippantonio, M. J., Caplen, N. J. & Ried, T. 2013. Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/β-catenin pathway in colorectal cancer. *Cancer Res,* 73(6)**,** pp 2003-13.

Cancer Research UK. *Myeloma mortality statistics* [Online]. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/mortality [Accessed 26th Feb 2015 2016].

Cancer Research UK. *Myeloma Research* [Online]. Available: http://www.cancerresearchuk.org/about-cancer/type/myeloma/treatment/whats-new-in-myeloma-research [Accessed 1st March 2016.

Carazo-Salas, R. E., Guarguaglini, G., Gruss, O. J., Segref, A., Karsenti, E. & Mattaj, I. W. 1999. Generation of GTP-bound Ran by RCC1 is required for chromatin-induced mitotic spindle formation. *Nature,* 400(6740)**,** pp 178-81.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., et al. 2005. The transcriptional landscape of the mammalian genome. *Science,* 309(5740)**,** pp 1559-63.

Cech, T. R. 1990. Self-splicing of group I introns. *Annu Rev Biochem,* 59(543-68.

Chan, D. A. & Giaccia, A. J. 2011. Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat Rev Drug Discov,* 10(5)**,** pp 351-64.

Chan, N., Pires, I. M., Bencokova, Z., Coackley, C., Luoto, K. R., Bhogal, N., Lakshman, M., Gottipati, P., Oliver, F. J., Helleday, T., Hammond, E. M. & Bristow, R. G. 2010. Contextual synthetic lethality of cancer cell kill based on the tumor microenvironment. *Cancer Res,* 70(20)**,** pp 8045-54.

Chang, H. M., Triboulet, R., Thornton, J. E. & Gregory, R. I. 2013. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature,* 497(7448)**,** pp 244-8.

Chapman, M. A., Lawrence, M. S., Keats, J. J., Cibulskis, K., Sougnez, C., Schinzel, A. C., Harview, C. L., Brunet, J. P., Ahmann, G. J., Adli, M., Anderson, K. C., Ardlie, K. G., Auclair, D., Baker, A., Bergsagel, P. L., Bernstein, B. E., Drier, Y., Fonseca, R., Gabriel, S. B., Hofmeister, C. C., Jagannath, S., Jakubowiak, A. J., Krishnan, A., Levy, J., Liefeld, T., Lonial, S., Mahan, S., Mfuko, B., Monti, S., Perkins, L. M., Onofrio, R., Pugh, T. J., Rajkumar, S. V., Ramos, A. H., Siegel, D. S., Sivachenko, A., Stewart, A. K., Trudel, S., Vij, R., Voet, D., Winckler, W., Zimmerman, T., Carpten, J., Trent, J., Hahn, W. C., Garraway, L. A., Meyerson, M., Lander, E. S., Getz, G. & Golub, T. R. 2011. Initial genome sequencing and analysis of multiple myeloma. *Nature,* 471(7339)**,** pp 467-72.

Chekulaeva, M., Mathys, H., Zipprich, J. T., Attig, J., Colic, M., Parker, R. & Filipowicz, W. 2011. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol,* 18(11)**,** pp 1218-26.

Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., Stoecklin, G., Moroni, C., Mann, M. & Karin, M. 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell,* 107(4)**,** pp 451-64.

Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. 2013. RNA decay machines: The exosome. *Biochim Biophys Acta,* 1829(6-7)**,** pp 552-60.

Clarke, P. R. & Zhang, C. 2008. Spatial and temporal coordination of mitosis by Ran GTPase. *Nat Rev Mol Cell Biol,* 9(6)**,** pp 464-77.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J. & Grimmond, S. M. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods,* 5(7)**,** pp 613-9.

Consortium, I. H. G. S. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431(7011)**,** pp 931-45.

Crick, F. 1970. Central dogma of molecular biology. *Nature,* 227(5258)**,** pp 561-3.

de Groen, F. L., Krijgsman, O., Tijssen, M., Vriend, L. E., Ylstra, B., Hooijberg, E., Meijer, G. A., Steenbergen, R. D. & Carvalho, B. 2014. Gene-dosage dependent overexpression at the 13q amplicon identifies DIS3 as candidate oncogene in colorectal cancer progression. *Genes Chromosomes Cancer,* 53(4)**,** pp 339-48.

Deng, W. & Lin, H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell,* 2(6)**,** pp 819-30.

Desai, A. & Hyman, A. 1999. Microtubule cytoskeleton: No longer an also Ran. *Curr Biol,* 9(18)**,** pp R704-7.

Dimopoulos, K., Gimsing, P. & Grønbæk, K. 2014. The role of epigenetics in the biology of multiple myeloma. *Blood Cancer J,* 4(e207.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K. & DiPersio, J. F. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature,* 481(7382)**,** pp 506-10.

Doidge, R., Mittal, S., Aslam, A. & Winkler, G. S. 2012. Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. *Biochem Soc Trans,* 40(4)**,** pp 896-901.

Doma, M. K. & Parker, R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature,* 440(7083)**,** pp 561-4.

Drach, J., Schuster, J., Nowotny, H., Angerler, J., Rosenthal, F., Fiegl, M., Rothermundt, C., Gsur, A., Jäger, U. & Heinz, R. 1995. Multiple myeloma: high incidence of chromosomal aneuploidy as detected by interphase fluorescence in situ hybridization. *Cancer Res,* 55(17)**,** pp 3854-9.

Drazkowska, K., Tomecki, R., Stodus, K., Kowalska, K., Czarnocki-Cieciura, M. & Dziembowski, A. 2013. The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro. *Nucleic Acids Res,* 41(6)**,** pp 3845-58.

Dredge, B. K., Polydorides, A. D. & Darnell, R. B. 2001. The splice of life: alternative splicing and neurological disease. *Nat Rev Neurosci,* 2(1)**,** pp 43-50.

Dziembowski, A., Lorentzen, E., Conti, E. & Seraphin, B. 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol,* 14(1)**,** pp 15-22.

Eddy, S. R. 1999. Noncoding RNA genes. *Curr Opin Genet Dev,* 9(6)**,** pp 695-9.

Eddy, S. R. 2001. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet,* 2(12)**,** pp 919-29.

Engels, E. A., Clarke, C. A., Pfeiffer, R. M., Lynch, C. F., Weisenburger, D. D., Gibson, T. M., Landgren, O. & Morton, L. M. 2013. Plasma cell neoplasms in US solid organ transplant recipients. *Am J Transplant,* 13(6)**,** pp 1523-32.

Eurofins Genomics. *The principle of the Illumina sequencingtechnology* [Online]. Available: https://www.eurofinsgenomics.eu/en/eurofins-genomics/product-faqs/next-generation-sequencing/general-technical-questions/what-is-the-principal-of-the-illumina-sequencing-technology.aspx [Accessed 5th April 2016 2016].

Fabian, M. R., Cieplak, M. K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T. F. & Sonenberg, N. 2011. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol,* 18(11)**,** pp 1211-7.

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M., Jackson, S. P., Smith, G. C. & Ashworth, A. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature,* 434(7035)**,** pp 917-21.

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P. & Kohtz, J. D. 2006. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev,* 20(11)**,** pp 1470-84.

Flynt, A. S., Greimann, J. C., Chung, W. J., Lima, C. D. & Lai, E. C. 2010. MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. *Mol Cell,* 38(6)**,** pp 900-7.

Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J., Ashworth, A., Carmichael, J., Kaye, S. B., Schellens, J. H. & de Bono, J. S. 2009. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med,* 361(2)**,** pp 123-34.

Fong, P. C., Yap, T. A., Boss, D. S., Carden, C. P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., A'Hern, R., Tutt, A., Ashworth, A., Stone, J.,

Carmichael, J., Schellens, J. H., de Bono, J. S. & Kaye, S. B. 2010. Poly(ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol,* 28(15)**,** pp 2512-9.

Fowler, J. A., Edwards, C. M. & Croucher, P. I. 2011. Tumor-host cell interactions in the bone disease of myeloma. *Bone,* 48(1)**,** pp 121-8.

Frazão, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M. & Carrondo, M. A. 2006. Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature,* 443(7107)**,** pp 110-4.

Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R. & Dietz, H. C. 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science,* 295(5563)**,** pp 2258-61.

Fukaya, T. & Tomari, Y. 2012. MicroRNAs mediate gene silencing via multiple different pathways in drosophila. *Mol Cell,* 48(6)**,** pp 825-36.

Garneau, N. L., Wilusz, J. & Wilusz, C. J. 2007. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol,* 8(2)**,** pp 113-26.

Graham, A. C., Davis, S. M. & Andrulis, E. D. 2009. Interdependent nucleocytoplasmic trafficking and interactions of Dis3 with Rrp6, the core exosome and importin-alpha3. *Traffic,* 10(5)**,** pp 499-513.

Graham, A. C., Kiss, D. L. & Andrulis, E. D. 2006. Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci. *Mol Biol Cell,* 17(3)**,** pp 1399-409.

Greaves, M. & Maley, C. C. 2012. Clonal evolution in cancer. *Nature,* 481(7381)**,** pp 306-13.

Greipp, P. R., San Miguel, J., Durie, B. G., Crowley, J. J., Barlogie, B., Bladé, J., Boccadoro, M., Child, J. A., Avet-Loiseau, H., Harousseau, J. L., Kyle, R. A., Lahuerta, J. J., Ludwig, H., Morgan, G., Powles, R., Shimizu, K., Shustik, C., Sonneveld, P., Tosi, P., Turesson, I. & Westin, J. 2005. International staging system for multiple myeloma. *J Clin Oncol,* 23(15)**,** pp 3412-20.

Grewal, S. I. & Elgin, S. C. 2002. Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev,* 12(2)**,** pp 178-87.

Gudipati, R. K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L. M., Jacquier, A. & Libri, D. 2012. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol Cell,* 48(3)**,** pp 409-21.

Guhaniyogi, J. & Brewer, G. 2001. Regulation of mRNA stability in mammalian cells. *Gene,* 265(1-2)**,** pp 11-23.

Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature,* 466(7308)**,** pp 835-40.

Gutschner, T. & Diederichs, S. 2012. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol,* 9(6)**,** pp 703-19.

Hahn, M. W. & Wray, G. A. 2002. The g-value paradox. *Evol Dev,* 4(2)**,** pp 73-5.

Halbach, F., Reichelt, P., Rode, M. & Conti, E. 2013. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell,* 154(4)**,** pp 814-26.

Halees, A. S., Hitti, E., Al-Saif, M., Mahmoud, L., Vlasova-St Louis, I. A., Beisang, D. J., Bohjanen, P. R. & Khabar, K. 2011. Global assessment of GU-rich regulatory content and function in the human transcriptome. *RNA Biol,* 8(4)**,** pp 681-91.

Hamilton, A. J. & Baulcombe, D. C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science,* 286(5441)**,** pp 950-2.

Hamm, J., Darzynkiewicz, E., Tahara, S. M. & Mattaj, I. W. 1990. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell,* 62(3)**,** pp 569-77.

Han, J., Pedersen, J. S., Kwon, S. C., Belair, C. D., Kim, Y. K., Yeom, K. H., Yang, W. Y., Haussler, D., Blelloch, R. & Kim, V. N. 2009. Posttranscriptional crossregulation between Drosha and DGCR8. *Cell,* 136(1)**,** pp 75-84.

Hartman, J. L., Garvik, B. & Hartwell, L. 2001. Principles for the buffering of genetic variation. *Science,* 291(5506)**,** pp 1001-4.

Hata, A. & Lieberman, J. 2015. Dysregulation of microRNA biogenesis and gene silencing in cancer. *Sci Signal,* 8(368)**,** pp re3.

Hawkins, P. G. & Morris, K. V. 2010. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription,* 1(3)**,** pp 165-175.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J. & Kim, V. N. 2008. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell,* 32(2)**,** pp 276-84.

Hou, D., Ruiz, M. & Andrulis, E. D. 2012. The ribonuclease Dis3 is an essential regulator of the developmental transcriptome. *BMC Genomics,* 13(359.

Huang, d. W., Sherman, B. T. & Lempicki, R. A. 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4(1)**,** pp 44-57.

Huang, J., Gong, Z., Ghosal, G. & Chen, J. 2009b. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell,* 35(3)**,** pp 384-93.

Huntzinger, E. & Izaurralde, E. 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet,* 12(2)**,** pp 99-110.

International Agency for Research on Cancer. *List of Classifications by cancer sites with sufficient or limited evidence in humans, Volumes 1 to 105** [Online]. Available: http://monographs.iarc.fr/ENG/Classification/index.php [Accessed 26th Feb 2016.

Jones, C. I., Pashler, A. L., Towler, B. P., Robinson, S. R. & Newbury, S. F. 2016. RNA-seq reveals post-transcriptional regulation of Drosophila insulin-like peptide dilp8 and the neuropeptide-like precursor Nplp2 by the exoribonuclease Pacman/XRN1. *Nucleic Acids Res,* 44(1)**,** pp 267-80.

Kadaba, S., Krueger, A., Trice, T., Krecic, A. M., Hinnebusch, A. G. & Anderson, J. 2004. Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. *Genes Dev,* 18(11)**,** pp 1227-40.

Kadowaki, T., Chen, S., Hitomi, M., Jacobs, E., Kumagai, C., Liang, S., Schneiter, R., Singleton, D., Wisniewska, J. & Tartakoff, A. M. 1994. Isolation and characterization of Saccharomyces cerevisiae mRNA transport-defective (mtr) mutants. *J Cell Biol,* 126(3)**,** pp 649-59.

Kaelin, W. G. 2005. The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer,* 5(9)**,** pp 689-98.

Kapranov, P., Willingham, A. T. & Gingeras, T. R. 2007. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet,* 8(6)**,** pp 413-23.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G. & Ohno, S. 2006. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev,* 20(3)**,** pp 355-67.

Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S. & Baba, T. 2009. Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev,* 23(4)**,** pp 433-8.

Kelley, L. A. & Sternberg, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4(3)**,** pp 363-71.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet,* 39(10)**,** pp 1278-84.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol,* 14(4)**,** pp R36.

Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., Ramakrishnan, A., Li, Y., Chung, Y. R., Micol, J. B., Murphy, M. E., Cho, H., Kim, M. K., Zebari, A. S., Aumann, S., Park, C. Y., Buonamici, S., Smith, P. G., Deeg, H. J., Lobry, C., Aifantis, I., Modis, Y., Allain, F. H., Halene, S., Bradley, R. K. & Abdel-Wahab, O. 2015. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell,* 27(5)**,** pp 617-30.

Kinoshita, N., Goebl, M. & Yanagida, M. 1991. The fission yeast dis3+ gene encodes a 110-kDa essential protein implicated in mitotic control. *Mol Cell Biol,* 11(12)**,** pp 5839-47.

Kiss, D. L. & Andrulis, E. D. 2010. Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. *RNA,* 16(4)**,** pp 781-91.

Kok, F. O., Shin, M., Ni, C. W., Gupta, A., Grosse, A. S., van Impel, A., Kirchmaier, B. C., Peterson-Maduro, J., Kourkoulis, G., Male, I., DeSantis, D. F., Sheppard-Tindell, S., Ebarasi, L., Betsholtz, C., Schulte-Merker, S., Wolfe, S. A. & Lawson, N. D. 2015. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell,* 32(1)**,** pp 97-108.

Korn, T., Mitsdoerffer, M., Croxford, A. L., Awasthi, A., Dardalhon, V. A., Galileos, G., Vollmar, P., Stritesky, G. L., Kaplan, M. H., Waisman, A., Kuchroo, V. K. & Oukka, M. 2008. IL-6 controls Th17 immunity in vivo by inhibiting the conversion of conventional T cells into Foxp3+ regulatory T cells. *Proc Natl Acad Sci U S A,* 105(47)**,** pp 18460-5.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M. B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H. J., Schübeler, D., Oertner, T. G., Schratt, G., Bibel, M., Roska, B. & Filipowicz, W. 2010a. Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell,* 141(4)**,** pp 618-31.

Krol, J., Loedige, I. & Filipowicz, W. 2010b. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet,* 11(9)**,** pp 597-610.

Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L. & Ginalski, K. 2009. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res,* 37(22)**,** pp 7701-14.

Kumakura, N., Otsuki, H., Tsuzuki, M., Takeda, A. & Watanabe, Y. 2013. Arabidopsis AtRRP44A is the functional homolog of Rrp44/Dis3, an exosome component, is essential for viability and is required for RNA processing and degradation. *PLoS One,* 8(11)**,** pp e79219.

Kung, J. T., Colognori, D. & Lee, J. T. 2013. Long noncoding RNAs: past, present, and future. *Genetics,* 193(3)**,** pp 651-69.

Kyle, R., Therneau, T., Rajkumar, S., Offord, J., Larson, D., Plevak, M. & Melton, L. 2002. A long-term study of prognosis in monoclonal gammopathy of undetermined significance. *New England Journal of Medicine,* 346(8)**,** pp 564-569.

Kyle, R. A., Buadi, F. & Rajkumar, S. V. 2011. Management of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Oncology (Williston Park),* 25(7)**,** pp 578-86.

Kyle, R. A. & Rajkumar, S. V. 2004. Multiple myeloma. *N Engl J Med,* 351(18)**,** pp 1860-73.

Kyle, R. A. & Rajkumar, S. V. 2008. Multiple myeloma. *Blood,* 111(6)**,** pp 2962-72.

Labome. *Illumina RNA library preparation* [Online]. Available: http://www.labome.com/method/RNA-seq-Using-Next-Generation-Sequencing.html [Accessed 5th April 2016 2016].

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. & Tollervey, D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell,* 121(5)**,** pp 713-24.

Langmead, B. & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9(4)**,** pp 357-9.

Larsson, S. C. & Wolk, A. 2007. Body mass index and risk of multiple myeloma: a meta-analysis. *Int J Cancer,* 121(11)**,** pp 2512-6.

Laubach, J., Richardson, P. & Anderson, K. 2011. Multiple myeloma. *Annu Rev Med,* 62(249-64.

Lebreton, A., Tomecki, R., Dziembowski, A. & Séraphin, B. 2008. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature,* 456(7224)**,** pp 993-6.

Lee, G., Bratkowski, M. A., Ding, F., Ke, A. & Ha, T. 2012. Elastic coupling between RNA degradation and unwinding by an exoribonuclease. *Science,* 336(6089)**,** pp 1726-9.

Lee, R. C., Feinbaum, R. L. & Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75(5)**,** pp 843-54.

Leich, E., Weißbach, S., Klein, H. U., Grieb, T., Pischimarov, J., Stühmer, T., Chatterjee, M., Steinbrunn, T., Langer, C., Eilers, M., Knop, S., Einsele, H., Bargou, R. & Rosenwald, A. 2013. Multiple myeloma is affected by multiple and heterogeneous somatic mutations in adhesion- and receptor tyrosine kinase signaling molecules. *Blood Cancer J,* 3(e102.

Lejeune, F., Li, X. & Maquat, L. E. 2003. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol Cell,* 12(3)**,** pp 675-87.

Li, Y., Bolderson, E., Kumar, R., Muniandy, P. A., Xue, Y., Richard, D. J., Seidman, M., Pandita, T. K., Khanna, K. K. & Wang, W. 2009. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem,* 284(35)**,** pp 23525-31.

Liang, L., Qu, L. & Ding, Y. 2007. Protein and mRNA characterization in human colorectal carcinoma cell lines with different metastatic potentials. *Cancer Invest,* 25(6)**,** pp 427-34.

Lim, J., Kuroki, T., Ozaki, K., Kohsaki, H., Yamori, T., Tsuruo, T., Nakamori, S., Imaoka, S., Endo, M. & Nakamura, Y. 1997. Isolation of murine and human homologues of the fission-yeast dis3+ gene encoding a mitotic-control protein and its overexpression in cancer cells with progressive phenotype. *Cancer Res,* 57(5)**,** pp 921-5.

Lionetti, M., Barbieri, M., Todoerti, K., Agnelli, L., Fabris, S., Tonon, G., Segalla, S., Cifola, I., Pinatel, E., Tassone, P., Musto, P., Baldini, L. & Neri, A. 2015. A compendium of DIS3 mutations and associated transcriptional signatures in plasma cell dyscrasias. *Oncotarget,* 6(28)**,** pp 26129-41.

Liu, Q., Greimann, J. C. & Lima, C. D. 2006. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell,* 127(6)**,** pp 1223-37.

Liu, W., Peng, Y. & Tobin, D. J. 2013. A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis. *PeerJ,* 1(e49.

Liu, X., Zheng, Q., Vrettos, N., Maragkakis, M., Alexiou, P., Gregory, B. D. & Mourelatos, Z. 2014. A MicroRNA precursor surveillance system in quality control of MicroRNA synthesis. *Mol Cell,* 55(6)**,** pp 868-79.

Lohr, J. G., Stojanov, P., Carter, S. L., Cruz-Gordillo, P., Lawrence, M. S., Auclair, D., Sougnez, C., Knoechel, B., Gould, J., Saksena, G., Cibulskis, K., McKenna, A., Chapman, M. A., Straussman, R., Levy, J., Perkins, L. M., Keats, J. J., Schumacher, S. E., Rosenberg, M., Getz, G., Golub, T. R. & Consortium, M. M. R. 2014. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell,* 25(1)**,** pp 91-101.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A. & Conti, E. 2008. Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol Cell,* 29(6)**,** pp 717-28.

Luo, B., Cheung, H. W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J. S., Beroukhim, R., Weir, B. A., Mermel, C., Barbie, D. A., Awad, T., Zhou, X., Nguyen, T., Piqani, B., Li, C., Golub, T. R., Meyerson, M., Hacohen, N., Hahn, W. C., Lander, E. S., Sabatini, D. M. & Root, D. E. 2008. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A,* 105(51)**,** pp 20380-5.

Mahindra, A., Laubach, J., Raje, N., Munshi, N., Richardson, P. G. & Anderson, K. 2012. Latest advances and current challenges in the treatment of multiple myeloma. *Nat Rev Clin Oncol,* 9(3)**,** pp 135-43.

Maiato, H., DeLuca, J., Salmon, E. D. & Earnshaw, W. C. 2004. The dynamic kinetochore-microtubule interface. *J Cell Sci,* 117(Pt 23)**,** pp 5461-77.

Makino, D. L., Baumgartner, M. & Conti, E. 2013. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature,* 495(7439)**,** pp 70-5.

Mamolen, M., Smith, A. & Andrulis, E. D. 2010. Drosophila melanogaster Dis3 N-terminal domains are required for ribonuclease activities, nuclear localization and exosome interactions. *Nucleic Acids Res,* 38(16)**,** pp 5507-17.

Manier, S., Sacco, A., Leleu, X., Ghobrial, I. M. & Roccaro, A. M. 2012. Bone marrow microenvironment in multiple myeloma progression. *J Biomed Biotechnol,* 2012(157496.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. 2007. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature,* 445(7128)**,** pp 666-70.

Masel, J. & Siegal, M. L. 2009. Robustness: mechanisms and consequences. *Trends Genet,* 25(9)**,** pp 395-403.

McShane, C. M., Murray, L. J., Landgren, O., O'Rorke, M. A., Korde, N., Kunzmann, A. T., Ismail, M. R. & Anderson, L. A. 2014. Prior autoimmune disease and risk of monoclonal gammopathy of undetermined significance and multiple myeloma: a systematic review. *Cancer Epidemiol Biomarkers Prev,* 23(2)**,** pp 332-42.

Meijer, H. A., Kong, Y. W., Lu, W. T., Wilczynska, A., Spriggs, R. V., Robinson, S. W., Godfrey, J. D., Willis, A. E. & Bushell, M. 2013. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science,* 340(6128)**,** pp 82-5.

Melchor, L., Brioli, A., Wardell, C. P., Murison, A., Potter, N. E., Kaiser, M. F., Fryer, R. A., Johnson, D. C., Begum, D. B., Hulkki Wilson, S., Vijayaraghavan, G., Titley, I., Cavo, M., Davies, F. E., Walker, B. A. & Morgan, G. J. 2014. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia,* 28(8)**,** pp 1705-15.

Meng, T. 2010. *The Molecular Cloning and Characterization of Fam46c RNA Stability Factor.* PhD, Harvard University.

Milligan, L., Torchet, C., Allmang, C., Shipman, T. & Tollervey, D. 2005. A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol Cell Biol,* 25(22)**,** pp 9996-10004.

Mitchell, P. 2014. Exosome substrate targeting: the long and short of it. *Biochem Soc Trans,* 42(4)**,** pp 1129-34.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell,* 91(4)**,** pp 457-66.

Mitchell, P., Petfalski, E. & Tollervey, D. 1996. The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev,* 10(4)**,** pp 502-13.

Morgan, G. J., Walker, B. A. & Davies, F. E. 2012. The genetic architecture of multiple myeloma. *Nat Rev Cancer,* 12(5)**,** pp 335-48.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5(7)**,** pp 621-8.

Mukherjee, D., Gao, M., O'Connor, J. P., Raijmakers, R., Pruijn, G., Lutz, C. S. & Wilusz, J. 2002. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J,* 21(1-2)**,** pp 165-74.

Mullen, T. E. & Marzluff, W. F. 2008. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev,* 22(1)**,** pp 50-65.

Mundy, G. R., Luben, R. A., Raisz, L. G., Oppenheim, J. J. & Buell, D. N. 1974. Bone-resorbing activity in supernatants from lymphoid cell lines. *N Engl J Med,* 290(16)**,** pp 867-71.

Murakami, H., Goto, D. B., Toda, T., Chen, E. S., Grewal, S. I., Martienssen, R. A. & Yanagida, M. 2007. Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS One,* 2(3)**,** pp e317.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. & Snyder, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320(5881)**,** pp 1344-9.

NCIN. *Cancer Incidence and Survival by Major Ethnic Group, England, 2002-2006* [Online]. Available: http://www.ncin.org.uk/publications/reports/ [Accessed 26th Feb 2016.

Neri, P., Ren, L., Azab, A. K., Brentnall, M., Gratton, K., Klimowicz, A. C., Lin, C., Duggan, P., Tassone, P., Mansoor, A., Stewart, D. A., Boise, L. H., Ghobrial, I. M. & Bahlis, N. J. 2011. Integrin β7-mediated regulation of multiple myeloma cell adhesion, migration, and invasion. *Blood,* 117(23)**,** pp 6202-13.

Newbury, S. F. 2006. Control of mRNA stability in eukaryotes. *Biochem Soc Trans.* England.

Ng, S. Y., Johnson, R. & Stanton, L. W. 2012. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J,* 31(3)**,** pp 522-33.

Nijman, S. M. 2011. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. *FEBS Lett,* 585(1)**,** pp 1-6.

Noguchi, E., Hayashi, N., Azuma, Y., Seki, T., Nakamura, M., Nakashima, N., Yanagida, M., He, X., Mueller, U., Sazer, S. & Nishimoto, T. 1996. Dis3, implicated in mitotic

control, binds directly to Ran and enhances the GEF activity of RCC1. *EMBO J,* 15(20)**,** pp 5595-605.

Ohkura, H., Adachi, Y., Kinoshita, N., Niwa, O., Toda, T. & Yanagida, M. 1988. Cold-sensitive and caffeine-supersensitive mutants of the Schizosaccharomyces pombe dis genes implicated in sister chromatid separation during mitosis. *Embo j,* 7(5)**,** pp 1465-73.

Okamura, K. & Lai, E. C. 2008. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol,* 9(9)**,** pp 673-8.

Palumbo, A. & Anderson, K. 2011. Multiple myeloma. *N Engl J Med,* 364(11)**,** pp 1046-60.

Parkin, D. M., Boyd, L. & Walker, L. C. 2011. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer,* 105 Suppl 2(S77-81.

Paull, T. T. & Lee, J. H. 2005. The Mre11/Rad50/Nbs1 complex and its role as a DNA double-strand break sensor for ATM. *Cell Cycle,* 4(6)**,** pp 737-40.

Prabhala, R. H., Pelluru, D., Fulciniti, M., Prabhala, H. K., Nanjappa, P., Song, W., Pai, C., Amin, S., Tai, Y. T., Richardson, P. G., Ghobrial, I. M., Treon, S. P., Daley, J. F., Anderson, K. C., Kutok, J. L. & Munshi, N. C. 2010. Elevated IL-17 produced by TH17 cells promotes myeloma cell growth and inhibits immune function in multiple myeloma. *Blood,* 115(26)**,** pp 5385-92.

Preker, P., Almvig, K., Christensen, M. S., Valen, E., Mapendano, C. K., Sandelin, A. & Jensen, T. H. 2011. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res,* 39(16)**,** pp 7179-93.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H. & Jensen, T. H. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science,* 322(5909)**,** pp 1851-4.

Raab, M. S., Podar, K., Breitkreutz, I., Richardson, P. G. & Anderson, K. C. 2009. Multiple myeloma. *Lancet,* 374(9686)**,** pp 324-39.

Rajkumar, S. V. & Witzig, T. E. 2000. A review of angiogenesis and antiangiogenic therapy with thalidomide in multiple myeloma. *Cancer Treat Rev,* 26(5)**,** pp 351-62.

Ratta, M., Fagnoni, F., Curti, A., Vescovini, R., Sansoni, P., Oliviero, B., Fogli, M., Ferri, E., Della Cuna, G. R., Tura, S., Baccarani, M. & Lemoli, R. M. 2002. Dendritic cells are

functionally defective in multiple myeloma: the role of interleukin-6. *Blood,* 100(1)**,** pp 230-7.

Rehrauer, H., Opitz, L., Tan, G., Sieverling, L. & Schlapbach, R. 2013. Blind spots of quantitative RNA-seq: the limits for assessing abundance, differential expression, and isoform switching. *BMC Bioinformatics,* 14(370.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA,* 11(11)**,** pp 1640-7.

Reis, C. C. & Campbell, J. L. 2007. Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in Saccharomyces cerevisiae. *Genetics,* 175(3)**,** pp 993-1010.

Reznik, B. & Lykke-Andersen, J. 2010. Regulated and quality-control mRNA turnover pathways in eukaryotes. *Biochem Soc Trans,* 38(6)**,** pp 1506-10.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. & Chang, H. Y. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell,* 129(7)**,** pp 1311-23.

Roccaro, A. M., Sacco, A., Thompson, B., Leleu, X., Azab, A. K., Azab, F., Runnels, J., Jia, X., Ngo, H. T., Melhem, M. R., Lin, C. P., Ribatti, D., Rollins, B. J., Witzig, T. E., Anderson, K. C. & Ghobrial, I. M. 2009. MicroRNAs 15a and 16 regulate tumor proliferation in multiple myeloma. *Blood,* 113(26)**,** pp 6669-80.

Rose, A. E., Poliseno, L., Wang, J., Clark, M., Pearlman, A., Wang, G., Vega Y Saenz de Miera, E. C., Medicherla, R., Christos, P. J., Shapiro, R., Pavlick, A., Darvishian, F., Zavadil, J., Polsky, D., Hernando, E., Ostrer, H. & Osman, I. 2011. Integrative genomics identifies molecular alterations that challenge the linear model of melanoma progression. *Cancer Res,* 71(7)**,** pp 2561-71.

Rozenblum, E., Vahteristo, P., Sandberg, T., Bergthorsson, J. T., Syrjakoski, K., Weaver, D., Haraldsson, K., Johannsdottir, H. K., Vehmanen, P., Nigam, S., Golberger, N., Robbins, C., Pak, E., Dutra, A., Gillander, E., Stephan, D. A., Bailey-Wilson, J., Juo, S. H., Kainu, T., Arason, A., Barkardottir, R. B., Nevanlinna, H., Borg, A. & Kallioniemi, O. P. 2002. A genomic map of a 6-Mb region at 13q21-q22 implicated in cancer development: identification and characterization of candidate genes. *Hum Genet,* 110(2)**,** pp 111-21.

Sammeth, M., Foissac, S. & Guigó, R. 2008. A general definition and nomenclature for alternative splicing events. *PLoS Comput Biol,* 4(8)**,** pp e1000147.

Sazer, S. & Dasso, M. 2000. The ran decathlon: multiple roles of Ran. *J Cell Sci,* 113 (Pt 7)(1111-8.

Schaeffer, D., Reis, F. P., Johnson, S. J., Arraiano, C. M. & van Hoof, A. 2012. The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. *Nucleic Acids Res,* 40(18)**,** pp 9298-307.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F. P., Dastidar, E. G., Sanchez-Rotunno, M., Arraiano, C. M. & van Hoof, A. 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol,* 16(1)**,** pp 56-62.

Schlabach, M. R., Luo, J., Solimini, N. L., Hu, G., Xu, Q., Li, M. Z., Zhao, Z., Smogorzewska, A., Sowa, M. E., Ang, X. L., Westbrook, T. F., Liang, A. C., Chang, K., Hackett, J. A., Harper, J. W., Hannon, G. J. & Elledge, S. J. 2008. Cancer proliferation gene discovery through functional genomics. *Science,* 319(5863)**,** pp 620-4.

Schneider, C., Anderson, J. T. & Tollervey, D. 2007. The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. *Mol Cell,* 27(2)**,** pp 324-31.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A. & Tollervey, D. 2012. Transcriptome-wide analysis of exosome targets. *Mol Cell,* 48(3)**,** pp 422-33.

Schneider, C., Leung, E., Brown, J. & Tollervey, D. 2009. The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res,* 37(4)**,** pp 1127-40.

Segalla, S., Pivetti, S., Todoerti, K., Chudzik, M. A., Giuliani, E. C., Lazzaro, F., Volta, V., Lazarevic, D., Musco, G., Muzi-Falconi, M., Neri, A., Biffo, S. & Tonon, G. 2015. The ribonuclease DIS3 promotes let-7 miRNA maturation by degrading the pluripotency factor LIN28B mRNA. *Nucleic Acids Res,* 43(10)**,** pp 5182-93.

Sezer, O., Heider, U., Jakob, C., Eucker, J. & Possinger, K. 2002. Human bone marrow myeloma cells express RANKL. *J Clin Oncol,* 20(1)**,** pp 353-4.

Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J. & Söll, D. 1985. Structure and transcription of eukaryotic tRNA genes. *CRC Crit Rev Biochem,* 19(2)**,** pp 107-44.

Shiels, M. S., Cole, S. R., Kirk, G. D. & Poole, C. 2009. A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals. *J Acquir Immune Defic Syndr,* 52(5)**,** pp 611-22.

Shiomi, T., Fukushima, K., Suzuki, N., Nakashima, N., Noguchi, E. & Nishimoto, T. 1998. Human dis3p, which binds to either GTP- or GDP-Ran, complements Saccharomyces cerevisiae dis3. *J Biochem,* 123(5)**,** pp 883-90.

Shiran, A., Brenner, B., Laor, A. & Tatarsky, I. 1993. Increased risk of cancer in patients with Gaucher disease. *Cancer,* 72(1)**,** pp 219-24.

Shlush, L. I. & Hershkovitz, D. 2015. Clonal evolution models of tumor heterogeneity. *Am Soc Clin Oncol Educ Book,* 35(e662-5.

Silva, J. M., Marran, K., Parker, J. S., Silva, J., Golding, M., Schlabach, M. R., Elledge, S. J., Hannon, G. J. & Chang, K. 2008. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science,* 319(5863)**,** pp 617-20.

Simons, A., Dafni, N., Dotan, I., Oron, Y. & Canaani, D. 2001a. Establishment of a chemical synthetic lethality screen in cultured human cells. *Genome Res,* 11(2)**,** pp 266-73.

Simons, A. H., Dafni, N., Dotan, I., Oron, Y. & Canaani, D. 2001b. Genetic synthetic lethality screen at the single gene level in cultured human cells. *Nucleic Acids Res,* 29(20)**,** pp E100.

Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol,* 12(4)**,** pp 246-58.

Skotheim, R. I. & Nees, M. 2007. Alternative splicing in cancer: noise, functional, or systematic? *Int J Biochem Cell Biol,* 39(7-8)**,** pp 1432-49.

Slevin, M. K., Meaux, S., Welch, J. D., Bigler, R., Miliani de Marval, P. L., Su, W., Rhoads, R. E., Prins, J. F. & Marzluff, W. F. 2014. Deep sequencing shows multiple oligouridylations are required for 3' to 5' degradation of histone mRNAs on polyribosomes. *Mol Cell,* 53(6)**,** pp 1020-30.

Smith, S. B., Kiss, D. L., Turk, E., Tartakoff, A. M. & Andrulis, E. D. 2011. Pronounced and extensive microtubule defects in a Saccharomyces cerevisiae DIS3 mutant. *Yeast,* 28(11)**,** pp 755-69.

Staals, R. H., Bronkhorst, A. W., Schilders, G., Slomovic, S., Schuster, G., Heck, A. J., Raijmakers, R. & Pruijn, G. J. 2010. Dis3-like 1: a novel exoribonuclease associated with the human exosome. *EMBO J,* 29(14)**,** pp 2358-67.

Stewart, H. J., Kishikova, L., Powell, F. L., Wheatley, S. P. & Chevassut, T. J. 2011. The polo-like kinase inhibitor BI 2536 exhibits potent activity against malignant plasma cells and represents a novel therapy in multiple myeloma. *Exp Hematol,* 39(3)**,** pp 330-8.

Suzuki, N., Noguchi, E., Nakashima, N., Oki, M., Ohba, T., Tartakoff, A., Ohishi, M. & Nishimoto, T. 2001. The Saccharomyces cerevisiae small GTPase, Gsp1p/Ran, is involved in 3' processing of 7S-to-5.8S rRNA and in degradation of the excised 5'-A0 fragment of 35S pre-rRNA, both of which are carried out by the exosome. *Genetics,* 158(2)**,** pp 613-25.

Sveen, A., Agesen, T. H., Nesbakken, A., Rognum, T. O., Lothe, R. A. & Skotheim, R. I. 2011. Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. *Genome Med,* 3(5)**,** pp 32.

Symmons, M. F. & Luisi, B. F. 2009. Through ancient rings thread programming strings. *Structure,* 17(11)**,** pp 1429-31.

Taft, R. J., Pheasant, M. & Mattick, J. S. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays,* 29(3)**,** pp 288-99.

Teoh, G. & Anderson, K. C. 1997. Interaction of tumor and host cells with adhesion and extracellular matrix molecules in the development of multiple myeloma. *Hematol Oncol Clin North Am,* 11(1)**,** pp 27-42.

Thomson, T. & Lin, H. 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol,* 25(355-76.

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T. R. & Guigó, R. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res,* 22(9)**,** pp 1616-25.

Tomecki, R., Drazkowska, K., Kucinski, I., Stodus, K., Szczesny, R. J., Gruchota, J., Owczarek, E. P., Kalisiak, K. & Dziembowski, A. 2014. Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res,* 42(2)**,** pp 1270-90.

Tomecki, R., Kristiansen, M. S., Lykke-Andersen, S., Chlebowski, A., Larsen, K. M., Szczesny, R. J., Drazkowska, K., Pastula, A., Andersen, J. S., Stepien, P. P., Dziembowski, A. & Jensen, T. H. 2010. The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J,* 29(14)**,** pp 2342-57.

Towler, B. P., Jones, C. I., Viegas, S. C., Apura, P., Waldron, J. A., Smalley, S. K., Arraiano, C. M. & Newbury, S. F. 2015. The 3'-5' exoribonuclease Dis3 regulates the expression of specific microRNAs in Drosophila wing imaginal discs. *RNA Biol,* 0.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. & Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7(3)**,** pp 562-78.

Triboulet, R., Chang, H. M., Lapierre, R. J. & Gregory, R. I. 2009. Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA,* 15(6)**,** pp 1005-11.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science,* 295(5563)**,** pp 2262-4.

Vasiljeva, L., Kim, M., Terzi, N., Soares, L. M. & Buratowski, S. 2008. Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell,* 29(3)**,** pp 313-23.

Wahle, E. & Winkler, G. S. 2013. RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. *Biochim Biophys Acta,* 1829(6-7)**,** pp 561-70.

Waldron, J. A., Jones, C. I., Towler, B. P., Pashler, A. L., Grima, D. P., Hebbes, S., Crossman, S. H., Zabolotskaya, M. V. & Newbury, S. F. 2015. Xrn1/Pacman affects apoptosis and regulates expression of hid and reaper. *Biol Open*.

Walker, B. A., Boyle, E. M., Wardell, C. P., Murison, A., Begum, D. B., Dahir, N. M., Proszek, P. Z., Johnson, D. C., Kaiser, M. F., Melchor, L., Aronson, L. I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J. R., Brioli, A., Mikulasova, A., Cairns, D. A., Gregory, W. M., Quartilho, A., Drayson, M. T., Russell, N., Cook, G., Jackson, G. H., Leleu, X., Davies, F. E. & Morgan, G. J. 2015. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol,* 33(33)**,** pp 3911-20.

Walker, B. A., Wardell, C. P., Chiecchio, L., Smith, E. M., Boyd, K. D., Neri, A., Davies, F. E., Ross, F. M. & Morgan, G. J. 2011. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood,* 117(2)**,** pp 553-62.

Walker, B. A., Wardell, C. P., Melchor, L., Hulkki, S., Potter, N. E., Johnson, D. C., Fenwick, K., Kozarewa, I., Gonzalez, D., Lord, C. J., Ashworth, A., Davies, F. E. & Morgan, G. J. 2012. Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. *Blood,* 120(5)**,** pp 1077-86.

Wang, H., Iacoangeli, A., Lin, D., Williams, K., Denman, R. B., Hellen, C. U. & Tiedge, H. 2005. Dendritic BC1 RNA in translational control mechanisms. *J Cell Biol,* 171(5)**,** pp 811-21.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A. & Chang, H. Y. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature,* 472(7341)**,** pp 120-4.

Wang, S. W., Stevenson, A. L., Kearsey, S. E., Watt, S. & Bähler, J. 2008. Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Mol Cell Biol,* 28(2)**,** pp 656-65.

Wang, Z. & Burge, C. B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA,* 14(5)**,** pp 802-13.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10(1)**,** pp 57-63.

Wasmuth, E. V. & Lima, C. D. 2012. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol Cell,* 48(1)**,** pp 133-44.

Weißbach, S., Langer, C., Puppe, B., Nedeva, T., Bach, E., Kull, M., Bargou, R., Einsele, H., Rosenwald, A., Knop, S. & Leich, E. 2015. The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Br J Haematol,* 169(1)**,** pp 57-70.

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klco, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., DiPersio, J. F., Ding, L., Mardis, E. R. & Wilson, R. K. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell,* 150(2)**,** pp 264-78.

Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell,* 2(1)**,** pp 135-40.

Wightman, B., Ha, I. & Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell,* 75(5)**,** pp 855-62.

Wilusz, J. E., Sunwoo, H. & Spector, D. L. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev,* 23(13)**,** pp 1494-504.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S. & Ogawa, S. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature,* 478(7367)**,** pp 64-9.

Zekri, L., Kuzuoğlu-Öztürk, D. & Izaurralde, E. 2013. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *EMBO J,* 32(7)**,** pp 1052-65.

Zhang, W., Murphy, C. & Sieburth, L. E. 2010. Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. *Proc Natl Acad Sci U S A,* 107(36)**,** pp 15981-5.

Zhao, J. 2004. Coordination of DNA synthesis and histone gene expression during normal cell cycle progression and after DNA damage. *Cell Cycle,* 3(6)**,** pp 695-7.

Zuo, Y. & Deutscher, M. P. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res,* 29(5)**,** pp 1017-26.