

Research Article

# DIS3 isoforms vary in their endoribonuclease activity and are differentially expressed within haematological cancers

Sophie R. Robinson<sup>1</sup>, Sandra C. Viegas<sup>2</sup>, Rute G. Matos<sup>2</sup>, Susana Domingues<sup>2</sup>, Marisa Bedir<sup>3</sup>, Helen J.S. Stewart<sup>1</sup>, Timothy J. Chevassut<sup>1</sup>,  Antony W. Oliver<sup>3</sup>, Cecilia M. Arraiano<sup>2</sup> and  Sarah F. Newbury<sup>1</sup>

<sup>1</sup>Medical Research Building, Brighton and Sussex Medical School, University of Sussex, Falmer, Brighton BN1 9PS, U.K.; <sup>2</sup>Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal; <sup>3</sup>School of Life Sciences, Genome Damage and Stability Centre, University of Sussex, Falmer, Brighton BN1 9RQ, U.K.

Correspondence: Sarah F. Newbury (s.newbury@bsms.ac.uk)



DIS3 (defective in sister chromatid joining) is the catalytic subunit of the exosome, a protein complex involved in the 3′–5′ degradation of RNAs. DIS3 is a highly conserved exoribonuclease, also known as Rrp44. Global sequencing studies have identified DIS3 as being mutated in a range of cancers, with a considerable incidence in multiple myeloma. In this work, we have identified two protein-coding isoforms of DIS3. Both isoforms are functionally relevant and result from alternative splicing. They differ from each other in the size of their N-terminal PIN (PiIT N-terminal) domain, which has been shown to have endoribonuclease activity and tether DIS3 to the exosome. Isoform 1 encodes a full-length PIN domain, whereas the PIN domain of isoform 2 is shorter and is missing a segment with conserved amino acids. We have carried out biochemical activity assays on both isoforms of full-length DIS3 and the isolated PIN domains. We find that isoform 2, despite missing part of the PIN domain, has greater endonuclease activity compared with isoform 1. Examination of the available structural information allows us to provide a hypothesis to explain this altered behaviour. Our results also show that multiple myeloma patient cells and all cancer cell lines tested have higher levels of isoform 1 compared with isoform 2, whereas acute myeloid leukaemia and chronic myelomonocytic leukaemia patient cells and samples from healthy donors have similar levels of isoforms 1 and 2. Taken together, our data indicate that significant changes in the ratios of the two isoforms could be symptomatic of haematological cancers.

## Introduction

DIS3 (defective in sister chromatid joining) is a highly conserved RNA exoribonuclease and a catalytic subunit of the exosome, a protein complex involved in the 3′–5′ degradation and processing of RNAs. The crucial role that DIS3 plays within RNA processing and decay is highlighted by its association with many forms of human cancer [1,2]. High-throughput studies have identified that DIS3 is recurrently mutated in different types of cancer such as multiple myeloma (10%) and in acute myeloid leukaemia (AML, 4%) [3–7]. Additionally, DIS3 is differentially expressed in superficial spreading melanoma and nodular melanoma and has also been identified as a candidate oncogene in colorectal cancer [8–11]. Therefore, the levels of DIS3 protein appear to be important and relevant to the progression of many commonly occurring cancers.

The precise role of DIS3 in cancer progression is not at all clear. Indeed, knockdown or mutation of DIS3 in model organisms or human tissue culture cells results in lethality or inhibition of proliferation. For example, in both the yeast *Saccharomyces cerevisiae* and in *Drosophila*, knockdown of DIS3 has been shown to result in lethality, due to failure in mitosis [12,13]. Similarly, in human HEK293

Received: 19 December 2017  
Revised: 16 May 2018  
Accepted: 24 May 2018

Accepted Manuscript online:  
25 May 2018  
Version of Record published:  
29 June 2018

cells expressing catalytically dead mutants in DIS3, where the endogenous activity of DIS3 has been knocked down, the cells proliferate more slowly and have reduced metabolic activity [14]. DIS3 is found in the nucleus of human cells and has been shown to be involved in the degradation of a vast range of RNAs including protein-coding RNAs, snoRNA precursors, introns, long non-coding RNAs, microRNAs and tRNAs as well as Promoter Upstream Transcripts (PROMPTs) and Cryptic Unstable Transcripts (CUTs) [15]. Therefore, inhibition of its activity is likely to have wide-ranging effects on cellular metabolism.

Human DIS3 is a multi-domain protein containing two different catalytic activities: a 3′–5′ exonucleolytic activity conferred by the RNaseII/R (RNB) domain [16] and an endonucleolytic activity via the PilT N-terminal (PIN) domain [17–19]. Other domains include a CR3 motif [20], two cold shock domains (CSDs) and an S1 domain which non-specifically binds RNA [21] (Figure 1). The human genome contains two further homologues: DIS3L1 (DIS3L), which contains an inactive PIN domain, and DIS3L2, where the PIN domain is absent [22–25]. In mammals, DIS3 functions as one of the three catalytic subunits of the exosome, along with DIS3L1 and Rrp6, a distributive exoribonuclease which belongs to the RNase D family [26,27]. The RNA substrate is usually guided through the central channel within the 9-subunit exosome barrel and then degraded by the exonuclease domain of DIS3 located at the base of the ring [26,28–30]. Alternatively, some substrates, such as highly structured RNA molecules, may access the DIS3 catalytic centre by direct entry rather than threading through the exosome barrel [28,29,31].

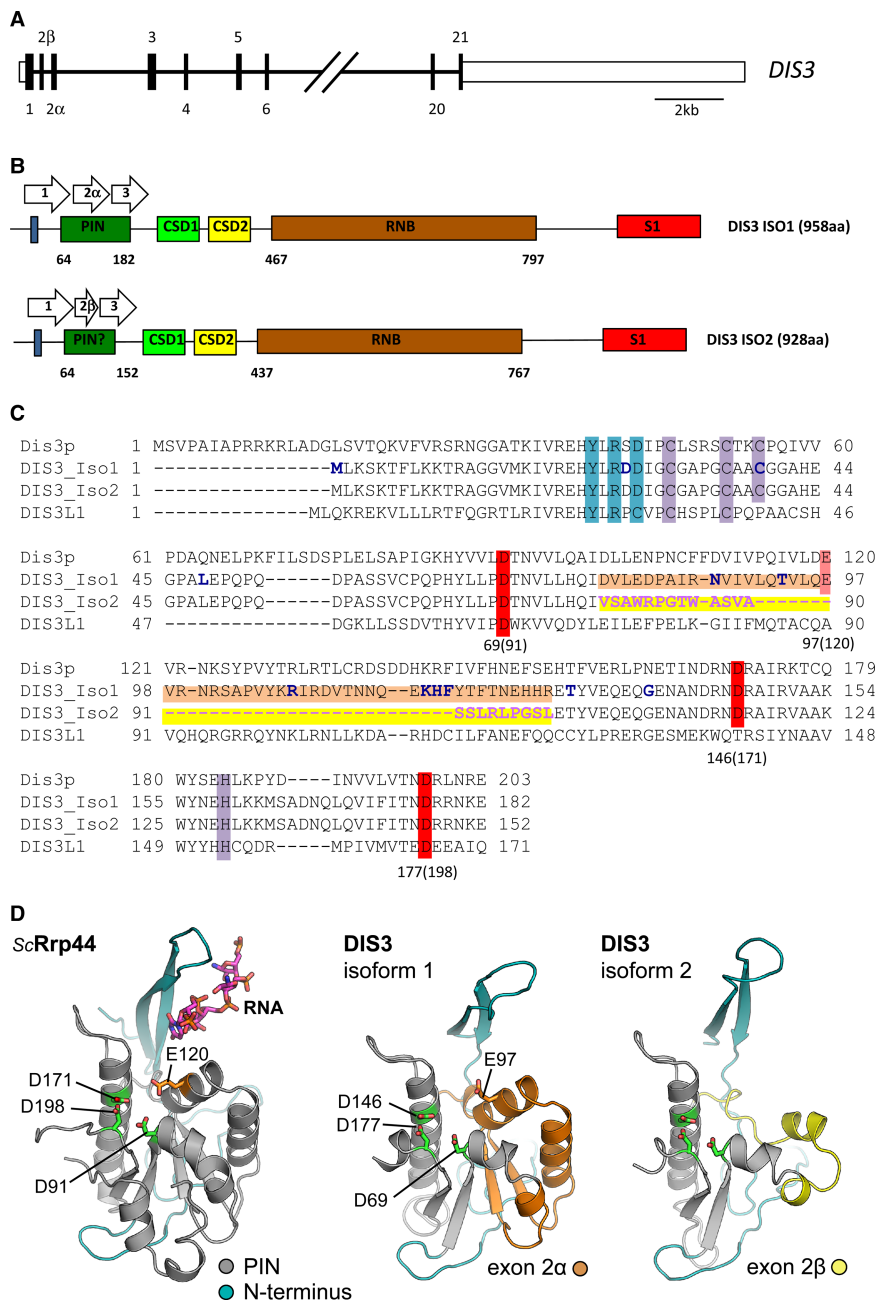
The importance of the PIN domain for DIS3 activity is demonstrated by evidence, showing that mutations which abrogate the endonuclease activity of the PIN domain together with mutations in the exonuclease domain have a synergistic effect on proliferation and metabolic activity in both human and yeast cells [14,18,19]. The PIN domain is a compact domain of ~100 amino acids which is present in many diverse proteins from all kingdoms of life. It is commonly found in RNA processing enzymes; for example, those involved in non-sense-mediated decay, such as SMG5 and SMG6 [32]. PIN domains contain three to four conserved acidic amino acid residues which serve to co-ordinate a divalent metal ion catalyzing cleavage of the RNA and release of nucleotide 5′ monophosphate products [17–19]. Crystal structure studies on the *S. cerevisiae* exosome bound to Rrp44 show that residues within the PIN domain contact the RNA together with regions of Rrp41 and Rrp43 (core exosome proteins) [26]. Furthermore, the N-terminal CR3 motif together with the YxRxD motif at the N-terminal of the PIN domain provides further anchoring of Rrp44 to the exosome barrel [20]. Taken together, these interactions produce a channel for RNA binding, directing it to the catalytic centres.

In the present paper, we identify, for the first time, the presence of two alternatively spliced protein-coding isoforms of human DIS3 that differ in the region encoding the crucial PIN domain. The shorter isoform (isoform 2) is missing conserved amino acid residues in the centre of the PIN domain. Biochemical assays using purified full-length DIS3 isoforms, as well as isolated PIN domains, demonstrate that isoform 1 has reduced endonuclease activity compared with isoform 2. Examination of the available structural information allows us to provide a hypothesis to explain this altered behaviour. *In vivo*, both of the isoforms are ubiquitously expressed but display differential expression levels in primary patient cells and many cell lines. Our results indicate that multiple myeloma patient cells and all cancer cell lines tested have higher levels of isoform 1 compared with isoform 2, whereas AML and chronic myelomonocytic leukaemia (CMML) patient samples and samples from healthy donors have similar levels of isoforms 1 and 2. Taken together, our data indicate that aberrant expression of these two isoforms can contribute to the progression of haematological cancers.

## Methods

### Overexpression, purification and *in vitro* activity assays of human Dis3 isoforms

The pReceiver-B03 vectors containing recombinant DIS3 isoform 1 or 2 were purchased from GeneCopoeia. This vector contains a T7 promoter for high-level expression, GST-tag and an ampicillin resistance gene for bacterial selection. The presence of each isoform was confirmed by DNA sequencing. To construct the truncated versions of DIS3 isoforms (ISO1 and ISO2), the expression plasmid (vector H0869 and vector H3667, respectively) was amplified by PCR with primers that cover the entire vector sequence, but create a premature termination codon after the amino acid N219 (N189 on ISO2) downstream from the PIN domain. In this way, the vectors created (pRSV-1 and pRSV-2) enable the expression of a truncated version of each protein composed only by the respective PIN domain fused to the protein purification tag. The PCR products were



**Figure 1. Schematic of the DIS3 locus together with the sequence and structure of the N-terminus of human DIS3.**

Part 1 of 2

(A) The DIS3 gene encodes two alternate exons, exon 2 $\alpha$  and 2 $\beta$  that are incorporated into isoforms 1 and 2 transcripts, respectively. (B) Exon 2 encodes a large portion of the PIN domain. Exon 2 $\beta$  is smaller than exon 2 $\alpha$ , resulting in a PIN domain 30 amino acids shorter in the isoform 2 protein. Although the rest of the protein remains in frame, this results in isoform 2 being 30 amino acids shorter in total than isoform 1. (C) Diagram showing the amino acid residues corresponding to the N-terminus and PIN domain of *S. cerevisiae* Dis3p, human DIS3 isoform 1, human DIS3 isoform 2, and human DIS3L1, together with residues thought to be important in their activity. The amino acid residues corresponding to isoform 1 are highlighted in orange, whereas the residues corresponding to isoform 2 are highlighted in yellow. The residues important for catalysis are highlighted in red with numbers below referring to human Dis3 isoform 1 amino acid sequence with corresponding yeast Dis3p sequence numbers in brackets. Acidic residues which are not important for catalysis in the SMG6 PIN domain are represented in pale red. The N-terminal Y-R-D exosome-binding site is marked in blue and the conserved CCCH motif is marked in purple [20]. Residues that are frequently mutated in multiple myeloma are marked in bold blue letters.

**Figure 1. Schematic of the DIS3 locus together with the sequence and structure of the N-terminus of human DIS3.**

Part 2 of 2

(D) Comparisons of the *S. cerevisiae* PIN domain with the human isoforms 1 and 2. In isoform 2, the loss of two alpha helices (labelled orange in isoform 1) is predicted to be highly destabilizing and affect the activity of this isoform. The three aspartic acid residues known to be important in co-ordinating the metal ion required for cleavage activity (D69, 146, and 177 in Dis3<sup>iso1</sup>) are marked in green. Co-ordinates corresponding to the PIN domain of ScRrp44 were extracted from PDB entry 4IFD [26]. Sequence-threaded homology models for both isoforms of human DIS3 were generated using the Phyre2 web-server [33].

circularized with T4 DNA ligase and used to transform competent DH5 $\alpha$  strains. The sequence of the selected clones was confirmed by DNA sequencing.

DIS3 protein isoforms ISO1 and ISO2, and the truncated versions containing only the respective PIN domain were overproduced with a glutathione *S*-transferase (GST) tag in *Escherichia coli* BL21-CodonPlus (DE3)-RIL strain containing the recombinant plasmids of interest. As a control, we have overexpressed the GST protein alone in the same strain containing the commercial plasmid pGEX-4T1 (GE Healthcare). The five proteins were purified by affinity chromatography. Briefly, cells were grown at 30°C in LB medium supplemented with ampicillin and chloramphenicol to an optical density (600 nm) near 1. Protein expression was induced by the addition of 0.5 mM of IPTG for 8 h at 20°C and cells were harvested by centrifugation. The culture pellets were resuspended in 1/20 volumes of Buffer A [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 2 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 of Benzonase (Sigma) to degrade nucleic acids and clarified by a 60 min centrifugation at 18 000 rpm, 4°C. The GST-tagged recombinant proteins were purified by affinity chromatography, using the ÄKTA FPLC™ System (GE Healthcare). The clarified extracts were loaded onto a GST-Trap 1 ml column (GE Healthcare) previously equilibrated in Buffer A. Protein elution was achieved with Buffer A with 20 mM reduced glutathione. The fractions containing the protein of interest, free of contaminants, were pooled and the buffer was exchanged for 20 mM Tris-HCl (pH 8), 100 mM KCl and 50% glycerol using Sephadex G-25 PD-10 desalting columns (GE Healthcare). Proteins were quantified using the Bradford Method and stored at –20°C. The purity of the enzymes was analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and by western blot using anti-GST antibodies, revealing >90% homogeneity.

*In vitro* activity assays of the proteins were performed using the synthetic 30-mer oligoribonucleotide ss16-A<sub>14</sub> as a substrate, labelled at its 5'-end with [ $\gamma$ -<sup>32</sup>ATP] and T4 Polynucleotide Kinase (Ambion), and circularized with T4 RNA ligase (Thermo). Protein and RNA concentrations were 50 and 25 nM, respectively, in all the cases. The experiment was performed in a buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, 3 mM MnCl<sub>2</sub>, and 1 mM DTT.

The reactions, in a total volume of 40  $\mu$ l, were started by the addition of the enzyme and further incubated at 37°C. Aliquots of 5  $\mu$ l were withdrawn at different time points, and the reactions were stopped by the addition of formamide containing dye supplemented with 10 mM EDTA. As a control, an aliquot of each reaction, without the enzyme, was incubated until the end of the assay. Reaction products were resolved in a 7 M urea/20% polyacrylamide gel and visualized by phosphorImaging (FLA-2000, Fuji, Stamford, CT, U.S.A.). Each activity assay was performed at least in triplicate. Quantification of the disappearance of the substrate was performed using the ImageQuant software (GE Healthcare).

For methods relating to Supplementary Figure S5, the reader is referred to the Supplementary Methods section. Plasmids and primers used in the present study are listed in Supplementary Tables S1–S3.

## Structural analyses

Sequence-threaded homology models of human DIS3 were generated using the structure of *S. cerevisiae* Rrp44, in the context of an 11-subunit exosome complex (PDB: 5K36), as a template [33].

## Cell culture

Cells were cultured in Dulbecco's modified Eagle's medium (HeLa, HEK-293), RPMI-1640 medium (OCI-AML3, U-266, RPMI-8226, THP-1, DG-75, GM12878, KG-1, KMS-12-BM, MOLP-8) 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 and 100  $\mu$ g/ml streptomycin, Sigma–Aldrich), at 37°C in a 5% CO<sub>2</sub> humidified

atmosphere. A total of 13 human cell lines were used in this work: the sources of these cell lines are given in Supplementary Table S4.

All patient samples were primary bone marrow aspirates with the exception of the healthy individuals which were peripheral blood samples. 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. Twelve of the CMML samples were obtained from Cambridge Blood and Stem Cell Biobank. Further information on patient samples is given in Supplementary Table S5. Mononuclear cells were isolated by Histopaque 1077 density gradient purification as per the manufacturer's instructions.

## Polymerase chain reaction and sequencing

Genomic DNA was extracted using the DNeasy Kit (Qiagen) according to the manufacturer's instructions. Semi-quantitative RT-PCR was used to investigate the presence of alternatively spliced isoforms of DIS3 as well as for amplification of cDNA from each isoform for sequencing. PCR products were purified using the DNA purification kit (Qiagen) according to manufacturer's instructions, before being sent to Eurofins MWG for sequencing.

## RNA extraction and qRT-PCR

Cells ( $2 \times 10^6$ ) were lysed and total RNA was extracted using the RNeasy Mini Kit (Qiagen) with an on-column DNase digestion, as per the manufacturer's instructions. RNA concentrations were measured on a NanoDrop 1000 spectrophotometer (Thermo Scientific). RNA was reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) as per the manufacturer's instructions. qRT-PCR was performed on each cDNA sample in triplicate using TaqMan Universal PCR Master Mix, No AmpErase UNG (Life Technologies). Each of the DIS3 isoform 1- or isoform 2-specific custom-designed Taqman assays (Supplementary Table S3) (Life Technologies) were run on a ViiA7 qRT-PCR machine (Applied Biosystems). Negative controls lacking reverse transcriptase showed negligible background. All data were normalized to GAPDH and relative expression levels were calculated using the  $2^{-\Delta\Delta C_t}$  method.

The amplification efficiencies of the DIS3 isoform-specific primers (DIS31 and DIS32) were determined using a series of cDNA dilutions (6.25, 12.5, 25, and 50 ng per 10  $\mu$ l qPCR, based on RNA concentrations) using a standard curve set-up (Supplementary Figure S1).

## Statistical analysis

Unless otherwise stated, error bars represent the SEM obtained from three or more independent experiments. All statistics were carried out using the 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.

## Results

### Characterization of two DIS3 protein-coding isoforms

Examination of the DIS3 gene in the Genome Browser ENSEMBL reveals five different DIS3 transcript annotations. 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, which encodes a large region of the endonucleolytic PIN domain and the length of their 3'-UTR (untranslated region). The PIN domain provides DIS3 with endonucleolytic activity and is thought to function in releasing natural exosome substrates that are stalled at sites of strong secondary structure [2,17].

Figure 1 shows the structure of the DIS3 gene. In humans, the length of the two annotated protein coding transcripts are 10 604 and 5232 nt, which are predicted to encode proteins of 958 and 928 amino acids (109 and 105 kDa, respectively) (www.genecards.org). Notably, exon 2 $\alpha$  (DIS3 isoform 1) is longer than exon 2 $\beta$  (DIS3 isoform 2). If isoform 2 is translated, the inclusion of exon 2 $\beta$  would result in a PIN domain 30 amino acids shorter than isoform 1 while leaving the rest of the protein in frame (Figure 1C). 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%.

Sequence-threaded homology models, using the published structure of *S. cerevisiae* Rrp44 as a template [26,33], enabled us to make predictions about the effect of exon 2 $\beta$  on the structure and function of the human DIS3 PIN domain. Amino acids 78–129 (51 amino acids; encoded by exon 2 $\alpha$ ) form approximately half of the PIN domain fold (coloured orange in Figure 1D). This is replaced by the much shorter sequence (22 amino acids) encoded by exon 2 $\beta$  (yellow, Figure 1D). The loss of two  $\beta$ -strands, which serve to create the central  $\beta$ -sheet of the PIN domain, is likely to be destabilizing, at least in the context of the PIN domain alone. Interestingly, three highly conserved aspartic acid residues (D69, D177, and D146) essential for PIN endonuclease activity are present in a similar conformation in both isoforms (Figure 1D). These data therefore suggest that the DIS3 protein encoded by isoform 2 may have a different biochemical activity than that encoded by isoform 1 because of the altered size of the PIN domain.

To obtain an indication of whether both DIS3 isoform transcripts are translated into protein and thus functionally relevant, the online tool GWIPS (Genome Wide Information on Protein Synthesis) was used to analyze and visualize Ribo-seq data obtained using the ribosome profiling technique. Data from all ribosome profiling studies indicate that both isoforms are translated and although coverage is much lower on the shorter exon 2 of isoform 2 (exon 2 $\beta$ ), ribosome binding is above the background level (Supplementary Figure S2). Therefore, isoform 2 can be translated into protein, at least under certain cellular conditions.

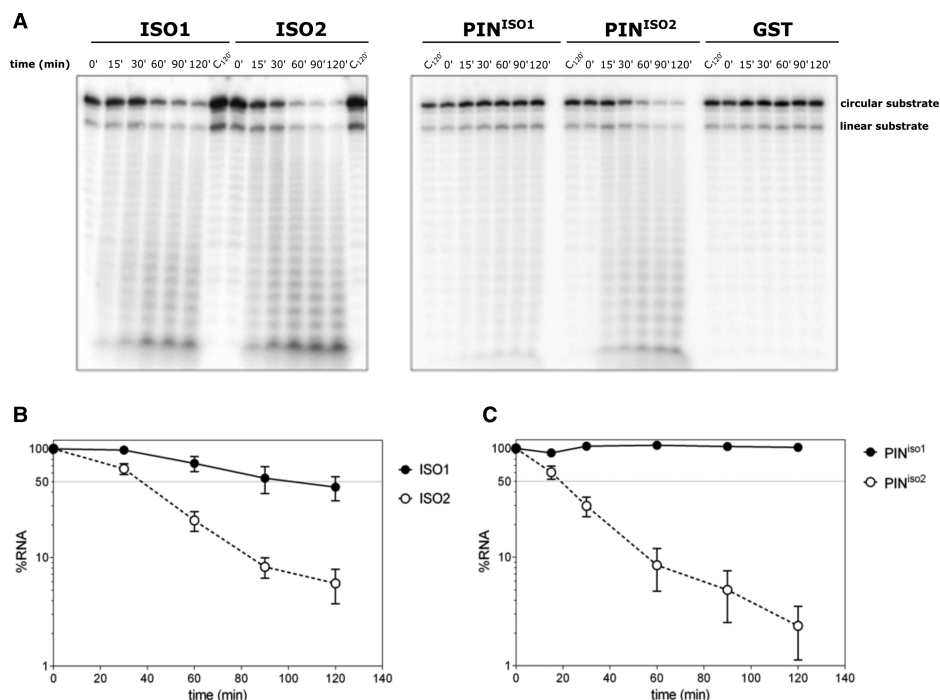
### Biochemical analyses of the endonuclease activities of Dis3 isoforms

Truncation of the PIN domain in DIS3 isoform 2 would suggest altered or absent endoribonuclease activity. To test this, we expressed and purified both isoforms of full-length DIS3 (DIS3<sup>iso1</sup> (ISO1) and DIS3<sup>iso2</sup> (ISO2)) as well as the isolated PIN domains (PIN<sup>iso1</sup> and PIN<sup>iso2</sup>). All four DIS3 variants were assayed for endoribonuclease activity using a circularized ss16-A<sub>14</sub> RNA substrate. Reactions were carried out in a high manganese concentration buffer, which favours DIS3 endonucleolytic activity [17–19]. If endoribonuclease activity occurs, the levels of the circular RNA substrate would be expected to decrease due to endonucleolytic cleavage. As can be seen from Figure 2, under our experimental conditions, ISO1, ISO2, and PIN<sup>iso2</sup> have endoribonuclease activity, as the levels of circular RNA substrate decrease over time. Surprisingly, ISO2, with the truncated PIN domain, had a higher endoribonuclease activity than ISO1. A graphical comparison of their activity shows that 50% of the circular substrate is cleaved in 40 min for ISO2, whereas ISO1 is only capable of cleaving 10% of the substrate in the same time (Figure 2). In confirmation of this result, the shorter PIN domain PIN<sup>iso2</sup> also cleaved the circular substrate more rapidly than PIN<sup>iso1</sup> with 50% being cleaved in 20 min for PIN<sup>iso2</sup>, whereas PIN<sup>iso1</sup> shows highly impaired activity. Concomitantly, there is also a higher accumulation of reaction products for PIN<sup>iso2</sup> compared with PIN<sup>iso1</sup>. Therefore, truncation of the PIN domain in isoform 2 actually appears to result in a greater endoribonuclease activity, rather than the expected reduction. No degradation of substrate was seen for a GST-only control, indicating no evidence of contamination by bacterial ribonucleases (Figure 2A).

In PIN domains, three to four conserved acidic amino acid residues co-ordinate the metal ion in the active site [32,34,35] (D69, E97, D146, and D177 in the human DIS3 protein) [17–19]. The third of these residues, an aspartic acid (D146 in hDIS3), is the most conserved in the PIN domains (except in the PIN domain of DIS3L1 homologue) and its single mutation is reported to abolish its activity *in vivo* and *in vitro* [19,32,34,35]. The second one (E97 in hDIS3) is not strictly conserved across the PIN-domain family [32]. The truncated PIN domain of hDIS3 isoform 2, however, contains only three of these acidic amino acids (D69, D146, and D177) that are conserved in identity and spatial arrangement with the PIN domains of yeast Rrp44, human SMG6, and isoform 1 of human DIS3 (Figure 1 and Supplementary Figure S3).

We confirmed the role of the highly conserved D146 residue, by mutating it to an asparagine, in the context of the PIN<sup>iso1</sup> expression construct (PIN<sup>D146N</sup>), as this substitution has previously been reported to compromise the endoribonuclease activity of the PIN domain [17,19,25]. In concordance, the D146N mutation practically abolishes the activity of the PIN protein, thereby confirming its importance in Mg<sup>2+</sup>/Mn<sup>2+</sup> co-ordination (Supplementary Figure S4). It also demonstrates that contaminating ribonucleases are not co-purified from the *E. coli* heterologous host.

To the best of our knowledge, the effect of mutating E97 in isolation has not been previously examined. Therefore, we changed the glutamic acid residue to an alanine in the context of the PIN<sup>iso1</sup> expression construct (PIN<sup>E97A</sup>) and measured its endonucleolytic activity against a circular RNA substrate. Data presented in Supplementary Figure S5 show that the activity of PIN<sup>E97A</sup> is reduced compared with the wild-type PIN<sup>iso1</sup>



**Figure 2. The shorter human DIS3 isoform 2 exhibits more endoribonuclease activity than isoform 1 on circular RNA substrates.**

(A) *In vitro* reactivity of purified wild-type full-length human DIS3 (ISO1) and the shorter variant ISO2 on a 5'-labelled 30-mer ssRNA circularized oligonucleotide. The same circularized substrate was incubated with a purified truncated version of human DIS3 isoforms 1 and 2 containing only the PIN domain (PIN<sup>iso1</sup> and PIN<sup>iso2</sup>, respectively). Incubation times are indicated on top of the panels. In each case, the shorter variant isoform 2 (ISO2 and PIN<sup>iso2</sup>) shows a higher level of cleavage of the circular substrate than the longer isoform 1 (ISO1 and PIN<sup>iso1</sup>). The control GST-tag protein (empty vector purification containing only the GST-tag) shows no endoribonuclease cleavage (right-hand panel), showing that no contaminating ribonuclease activity is carried over during the purification process. (B and C) Graphical representation of the *in vitro* endoribonuclease activity of the full-length DIS3 isoform 1 (ISO1) compared with DIS3 isoform 2 (ISO2) (B) and the isoform 1 PIN domain (PIN<sup>iso1</sup>) compared with the isoform 2 PIN domain (PIN<sup>iso2</sup>) (C). The substrate consumption was quantified over time and represented as the percentage of RNA disappearance.

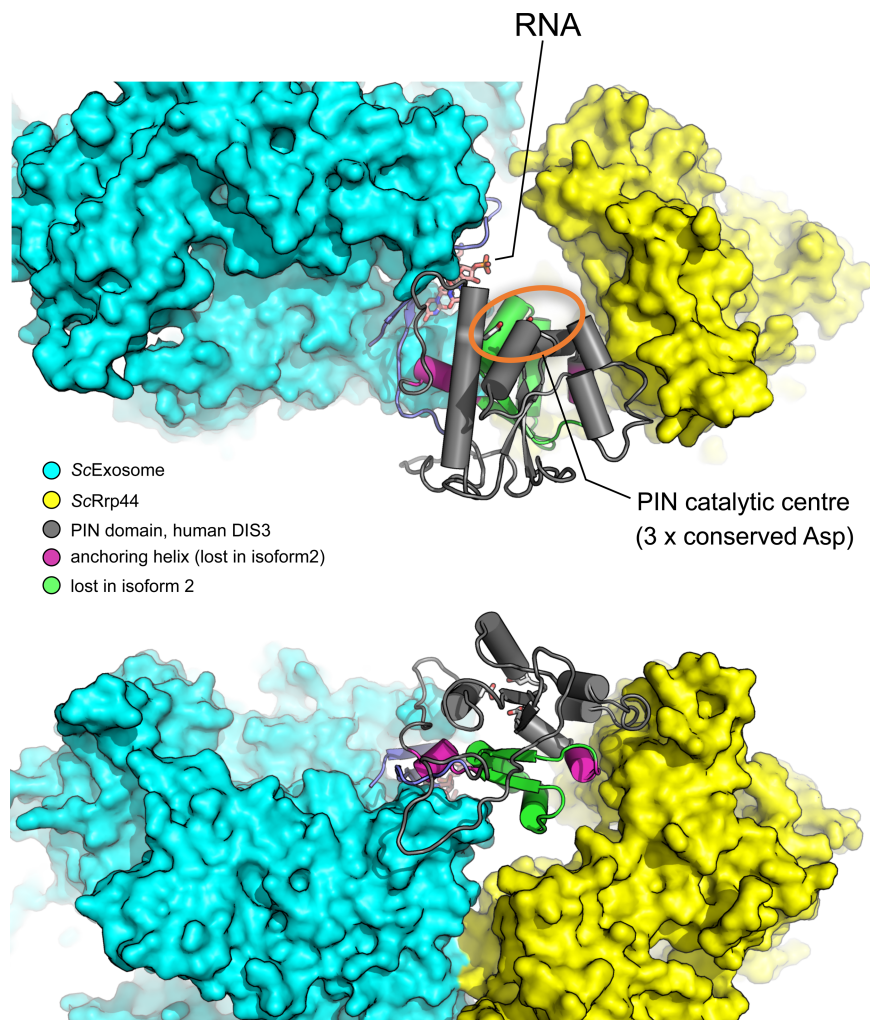
protein. Therefore, this residue also appears to be important for the catalytic activity of the PIN domain in the context of the longer isoform.

### Modelling the effect of the shorter isoform on exosome function

Sequence-threaded models generated for both isoforms of DIS3 reveal a potential explanation for the increased catalytic activity of isoform 2. Isoform 2 deletes amino acid residues 123-FTNEHHR-129 (coloured green in Figure 3). In isoform 1, these residues form a small loop that facilitates the interaction of the PIN domain of DIS3 with the rest of the exosome complex, while also forming an integral part of an RNA-interacting region. Two additional motifs, CR3 and YxRxD, also contribute to the PIN/exosome interface [20,26]. The loss of the FTNEHHR 'anchoring' motif in isoform 2 is likely to result in a higher degree of conformational plasticity and flexibility for the PIN domain, generating a wider, less obstructed channel for RNA to access the endoribonuclease active centre. This could explain the greater endonucleolytic activity of isoform 2 observed in our biochemical assays.

### Isoform 1 is the principal DIS3 isoform expressed in immortalized cell lines

To experimentally validate the presence of the two annotated DIS3 isoforms, semi-quantitative reverse-transcription PCR (RT-PCR) was employed. Primers were designed to flank the variable exon 2 by annealing to exons 1 and 3 common to both isoforms (Supplementary Figure S5A), thus amplifying two fragments, each



**Figure 3. Three-dimensional model of DIS3 illustrating the possible effects of the shorter PIN domain (isoform 2) on binding of DIS3 to the exosome.**

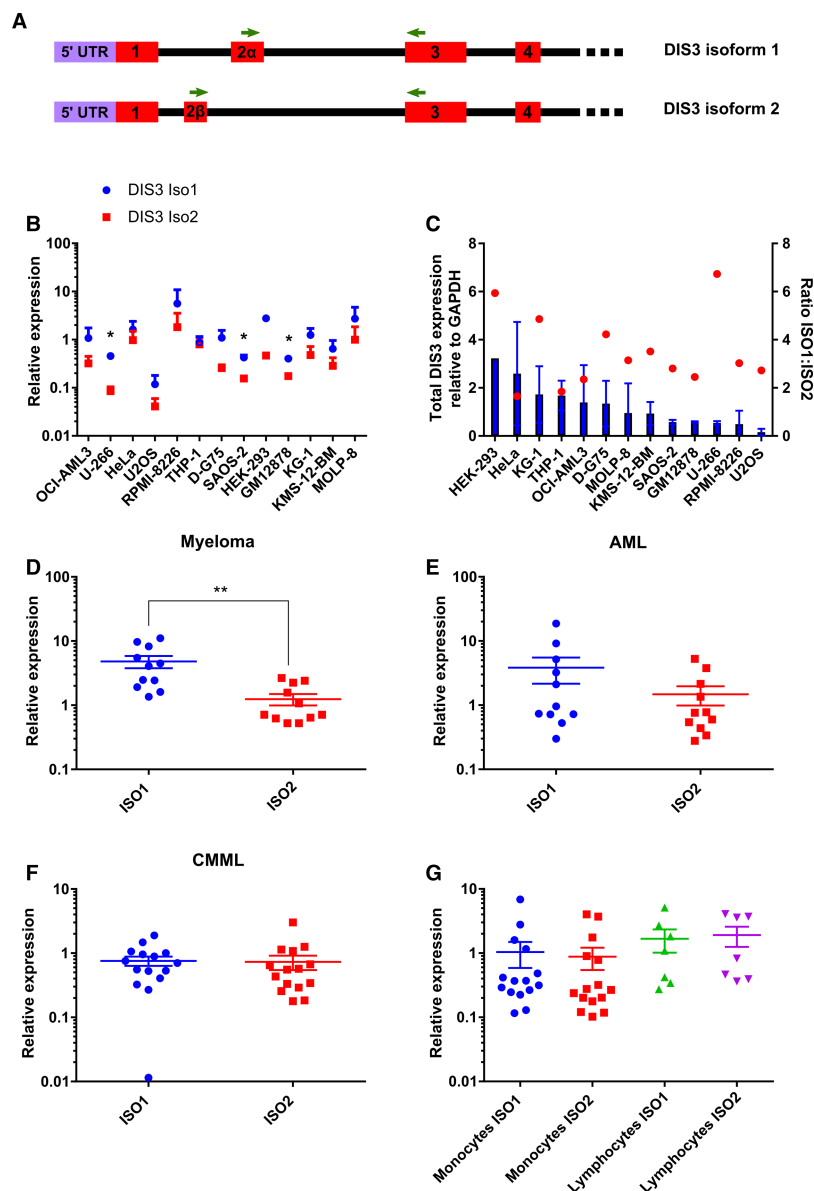
The PIN domain (in grey) links the remaining part of DIS3/Rrp44 to the exosome ring complex. The catalytic site of the PIN endoribonuclease domain (three conserved aspartic acids) is marked in orange. In isoform 2, the truncated PIN domain results in loss of a loop that serves to link the PIN domain back to the extreme N-terminal part of the protein (marked in green). These key residues normally form a little loop which serves to anchor the PIN domain back to the extreme N-terminal part of the protein, facilitating the interaction with rest of the complex, as well as forming part of an RNA-interacting region. In isoform 2, the anchoring helices (purple), seen best in the lower diagram, are also missing. Therefore, isoform 2 would most probably lose the tightly constrained conformation, allowing more access to the catalytic centre of the PIN domain than would be possible in isoform 1.

corresponding to the individual isoform transcripts. RT-PCR was performed on six cell lines of differing cancer types (Supplementary Figure S5B), predominantly AML and myeloma as well as mononuclear cells from haematological cancer patients which included six myeloma patients (Supplementary Figure S5C), four AML patients (Supplementary Figure S5D), and three CMML patients (Supplementary Figure S5E). Supplementary Figure S5 shows the amplification of two bands as expected, differing in size by exactly 100 bp. The larger fragment corresponds to isoform 1 with the full-length 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 than that of isoform 2 across all the cell lines and myeloma samples. In the AML samples, the intensities are more equal, while in two of the CMML patients the opposite is observed



(Supplementary Figure S5E). This raises the question of whether isoform expression is tissue- or disease-specific.

To address this initial observation, qRT-PCR was carried out using isoform-specific TaqMan primer-probe assays (Figure 4A). As the levels of the two targets are being directly compared with each other, the amplification efficiency of the primer probes was first tested by generating a standard curve (Supplementary Figure S1).



**Figure 4. Relative expression of the two DIS3 isoform transcripts in cell lines and patient samples.**

(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 13 cell lines relative to GAPDH ( $n = 3$ ). Asterisks (in black) indicate where the relative expression is significantly different ( $P < 0.05$ ). (C) Total DIS3 expression relative to GAPDH (blue bars) in 13 cell lines does not correlate with isoform ratio (red spots) ( $r = 0.14$ ,  $P = 0.64$ ). (D) In myeloma patient samples, expression levels of isoform 1 is consistently significantly higher than isoform 2 in myeloma patients ( $P = 0.0012$ ). (E) In AML patient samples, isoform 1 is expressed at slightly higher levels than isoform 2. (F) In CMML patient samples, levels of isoforms 1 and 2 are similar. (G) qRT-PCR shows that monocytes and lymphocytes isolated from healthy individuals show similar levels of isoforms 1 and 2. For further details on proportions of DIS3 isoforms 1 and 2 in cell lines and patient samples, see Supplementary Figure S6.

Isoform 1 is consistently more highly expressed than isoform 2 in all of the 13 cell lines tested, contributing an average of ~70% to total DIS3 levels (Figure 4B and Supplementary Figure S6A). 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 protein expressed. However, Figure 4C shows this not to be the case ( $r = 0.143$ ,  $P = 0.640$ ), which is perhaps not surprising given all cell lines show a similar isoform ratio.

### Higher levels of expression of isoform 1 are prevalent in myeloma cells but not in cells from AML and CMML patients

To extend the above findings on the isoform expression levels in cell lines to patient samples, qRT-PCR was employed using the same isoform-specific TaqMan primer probes as above (Figure 4A). In agreement with the semi-quantitative RT-PCR findings, all bone marrow samples from myeloma patients had a higher expression level of isoform 1 (Figure 4D and Supplementary Figure S6B). In contrast, for AML patient samples, the relative expression of isoform 1 was not significantly higher than isoform 2 with 3 out of 11 patient samples having a higher proportion of isoform 2 compared with isoform 1 (Figure 4E and Supplementary Figure S6C). For CMML patient samples, the relative expression of isoforms 1 and 2 was very similar and more of isoform 2 was expressed than isoform 1 in 5 of the 15 patient samples tested (Figure 4F and Supplementary Figure S6D).

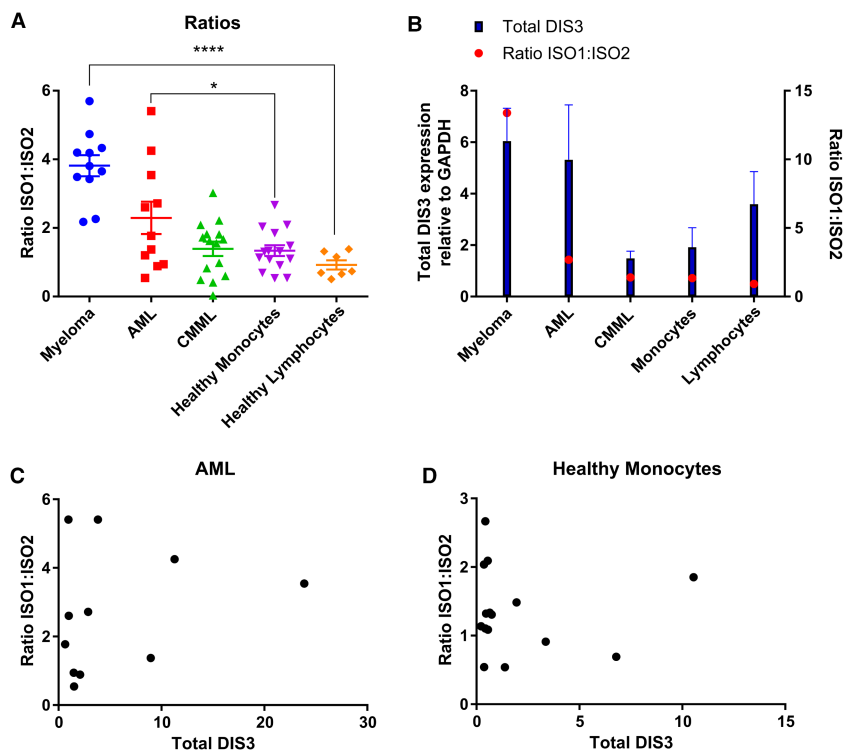
As myeloma is a lymphoid malignancy and AML and CMML are myeloid-derived, this suggests that this expression pattern is lymphocyte-specific. To test this, peripheral blood mononuclear cells were isolated from healthy individuals before being separated into monocyte and lymphocyte fractions. Monocytes were isolated using the adherence method, and anti-CD14 staining was used to test the purity, as well as to confirm IL-6 up-regulation upon lipopolysaccharide stimulation of monocytes (data not shown). Lymphocytes remained in suspension providing an easy and effective method to isolate the two cell fractions.

The qRT-PCR experiments demonstrate that, in healthy monocytes, the levels of isoforms 1 and 2 are similar, with 4 out of 15 individuals having higher levels of isoform 2 than isoform 1 (Figure 4G and Supplementary Figure S6E). Nevertheless, this is not myeloid-specific as healthy lymphocytes also show the same pattern (Figure 4G and Supplementary Figure S6F). This suggests that CMML patients are more similar to healthy controls in their expression ratio of the two DIS3 isoforms compared with myeloma patients. When the ratio of isoform 1 to isoform 2 within the three diseases are compared with healthy cells, myeloma is indeed significantly different ( $P < 0.0001$ ), whereas CMML patients are not significantly different (Figure 5). Interestingly, AML patients display a greater level of variation in isoform expression levels, but do appear to be significantly different from healthy controls ( $P = 0.04$ ). In summary, myeloma patient samples and all cancer cell lines tested have higher levels of isoform 2, whereas CMML patient samples and samples from healthy donors have similar levels of isoforms 1 and 2.

Unlike the cell line data, there is a significant correlation between the relative level of the two isoforms and the total level of DIS3 protein expressed in these primary cells (Figure 5B,  $r = 0.941$ ,  $P = 0.017$ ). However, no relationship is seen when the data within individual disease types are examined (Figure 5C,D). Taken together, these data suggest that the isoform ratios observed are a result of the cell specifically regulating the expression of isoform 1 to change the stoichiometric ratio of the two isoforms, rather than as a general means of controlling total DIS3 protein levels.

### Isoform expression ratios correlate with disease severity in CMML patients

Given the lack of evidence for the isoform ratio being a way of modulating total DIS3 protein levels, it was of interest to find out whether there was any correlation with disease severity. This was examined using the clinical parameters of plasma cell, blast and monocyte count for myeloma, AML 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 shows that the percentage of plasma cells/blast cells is not correlated with an isoform ratio in either myeloma ( $r = -0.18$ ,  $P = 0.67$ ) or AML ( $r = 0.29$ ,  $P = 0.41$ ). However, in CMML, there was a significant negative correlation between the percentage of monocytosis and isoform expression (Figure 6C). That is, CMML patients with a higher level of isoform 2 compared with isoform 1, or a smaller ISO1 : ISO2 ratio, have a higher number of monocytes in their blood ( $r = -0.62$ ,  $P = 0.01$ ).



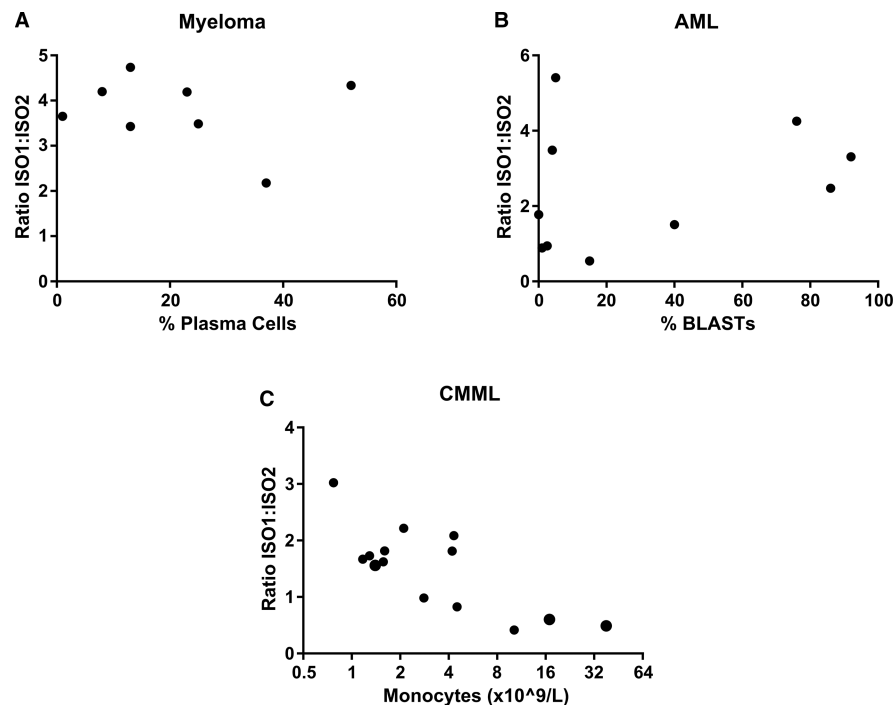
**Figure 5. Ratio of isoform expression levels versus total DIS3 levels.**

(A) The ratio of isoform 1 to isoform 2 is significantly higher between myeloma and healthy lymphocytes ( $P < 0.0001$ ) as well as AML and healthy monocytes ( $P = 0.04$ ) but not between CMML and healthy monocytes ( $P = 0.84$ ). (B) There is a positive correlation between the iso1 : iso2 ratio and total DIS3 levels across the disease types ( $r = 0.941$ ,  $P = 0.017$ ). (C and D) Within AML patients (C) and healthy monocytes (D) here is no correlation between the iso1 : iso2 ratio and total DIS3 levels.

## Discussion

The human protein DIS3 is of interest due to its frequent mutation in multiple myeloma patient cells. Here, we describe the existence of two protein-coding isoforms of human DIS3 that differ in the size of their PIN domain, as a result of alternative splicing using a mutually exclusive second exon. The shorter isoform (isoform 2) deletes several amino acids from the PIN domain, which are highly conserved with yeast Dis3/Rrp44 (Figure 1). As a result, we expected it to be highly compromised in endoribonuclease activity with respect to isoform 1. However, contrary to expectation, our biochemical analyses revealed the converse to be true. We also found that DIS3 isoform 1 was expressed at higher levels in all the immortalized cell lines tested, as well as in cells from patients affected by multiple myeloma.

Our detailed biochemical and structural analyses of the two DIS3 isoforms has provided us with insights into the function of these proteins. Our modelling shows that isoform 2 lacks the 123-FTNEHHR-129 anchoring motif (coloured green in Figure 3), which in isoform 1 facilitates the interaction of the PIN domain with the rest of the exosome complex. The loss of this motif is likely to generate a wider, flexible and less obstructed channel for RNA to access the endoribonuclease active site. It is feasible that the loss of these amino acids may also allow increased direct access of RNA to the PIN domain active site, rather than threading through the centre of the exosome. This increased direct access for RNA to the PIN domain active site may be particularly important for degradation of structured nuclear RNAs such as tRNAs and ribosomal RNAs [36]. In isoform 2, the acidic residue E97 that is conserved in yeast Dis3p is missing, suggesting that it is not required for catalytic activity in this particular context. However, biochemical analyses indicate that it is required in the context of isoform 1, as the cleavage activity of PIN<sup>iso1</sup> is reduced when this residue is mutated to alanine (PIN<sup>E97A</sup>; Supplementary Figure S4).



**Figure 6. 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 ( $r = -0.18$ ,  $P = 0.67$ ) or (B) AML, respectively ( $r = 0.29$ ,  $P = 0.412$ ). (C) In CMML, however, a significant negative correlation exists between the iso1 : iso2 ratio and monocyte count ( $r = -0.62$ ,  $P = 0.018$ ).

Our experimental data show that both of the *DIS3* transcripts are expressed in all the cancer cell lines we tested, indicating that both types of resultant protein may be required for efficient degradation of RNA. However, in all of the 13 immortalized cell lines analyzed, isoform 1 is expressed at a higher level than isoform 2 at the mRNA level, with some cell lines expressing nine times more isoform 1 than isoform 2 (HEK293). This is consistent with our results for cells from myeloma patients, which also express higher levels of isoform 1. Therefore, the isoform with the full-length PIN domain, although less active in terms of endoribonuclease activity, is often selected in highly proliferative, immortalized cells.

The importance of isoform 2, in relation to myeloma, is supported by recent data showing that the region of the PIN domain that is missing in isoform 2 is a ‘hot-spot’ for somatic mutations in humans [37]. These mutations include N87S, T93A, R108C, and a cluster of mutations at K118E, H119D, F120L and Y121S (Figures 1C and 3). Interestingly, these mutations directly abut the FTNEHHR anchoring motif. T93 is noted as a recurrent mutation and the variant Y121S is mutation in the myeloma cancer cell line OPM2. These mutations would be predicted to affect the function of isoform 1 protein and could also hinder the correct splicing of isoform 2.

The association of isoform 1 with the cancer phenotype is borne out by our results showing that it is the principal transcript in myeloma patient cells, making up ~80% of total *DIS3* protein levels in the 11 patients tested. In AML cells, overall, more patients expressed higher levels of isoform 1 than isoform 2. In contrast, monocytes and lymphocytes from healthy individuals expressed similar levels of isoforms 1 and 2. Therefore, the increase in isoform 1 in comparison with isoform 2 in myeloma cells is supportive of its role in cancer progression.

The reason for the prevalence of isoform 1 in cancer cell lines and in myeloma tumour cells is not clear. Our data suggest that reduced expression of isoform 2 may result in less degradation of target transcripts overall, particularly those that require endonucleolytic cleavage. Transcriptomic profiling of cells from myeloma patients has shown that a significant fraction of up-regulated transcripts encode proteins that are involved in RNA processing and degradation, e.g. *PATL2* (protein associated with topoisomerase II homologue 2; RNA binding), *DHX58* (DEXH-Box Helicase 58; RNA helicase), *PAN2* (PolyA-Specific Ribonuclease Subunit) and *POP1* (Processing of Precursor 1; deadenylases), *APOBEC3F* (Apolipoprotein B mRNA Editing Enzyme

Catalytic Subunit 3F; RNA editing) and RNU11 (RNA, U11 Small Nuclear; spliceosomal subunit) [37]. Therefore, mutations in DIS3 appear to be affecting post-transcriptional control pathways in the cell which presumably lead to up-regulation of specific RNAs involved in proliferation. In the case of CMML, where increased expression of isoform 2 is correlated with disease progression, it is expected that a different set of target RNAs will be up-regulated, together with mutations in other genes, resulting in alternative pathways leading to cancer phenotypes.

In conclusion, our data are consistent with human DIS3 encoding two variant forms of this protein, each of which has differences in the mode of action of their endoribonuclease activities. Aberrant expression of each of these two isoforms may contribute to the progression of haematological cancers. Further work to assess the differing effects of these two isoforms on cellular and cancer phenotypes may shed light on the progression of these cancers.

### Abbreviations

AML, acute myeloid leukaemia; CMML, chronic myelomonocytic leukaemia; DIS3, defective in sister chromatid joining; GST, glutathione *S*-transferase; PIN, PiIT N-terminus; RT-PCR, reverse-transcription PCR.

### Author Contribution

S.R.R. designed and performed most of the experiments. S.C.V., R.G.M., and S.D. designed, performed, and analyzed the biochemical experiments as well contributed to the writing of the manuscript. M.B. carried out some of the initial experiments. A.W.O. performed the structural modelling, advised on experimental interpretation, and contributed to writing the manuscript. H.J.S.S. advised on the cell biology methods and contributed to design of some experiments. T.J.C. contributed to the design of the experiments as well as provided the patient material. C.M.A. advised on the design and interpretation of the experiments. S.F.N. co-ordinated the study, contributed to the design and interpretation of the experiments, and wrote the manuscript.

### Funding

This work was funded by a Brighton and Sussex Medical School studentship [WC003-14] and a Genetics Society Training Grant [Montes24-8-2014] to S.R.R. Work at ITQB NOVA was financially supported by: Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by the European Regional Development Fund (FEDER) through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI) and by national funds through FCT — Fundação para a Ciência e a Tecnologia; project PTDC/BIA-MIC/1399/2014. R.G.M. is a recipient of a post-doctoral grant from FCT [SFRH/BPD/75887/2011]. S.C.V. was financed by an FCT grant [SFRH/BPD/103040/2014] and by program IF of 'Fundação para a Ciência e a Tecnologia' [ref. IF/00217/2015]. S.D. is the recipient of a post-doctoral fellowship from the EmPower Putida project ([grant no. 635536] from the European Union's Horizon 2020 research and innovation programme). The work was also supported by Cancer Research UK [C/302/A24386; A.W.O.] and a Biotechnology and Biological Sciences Research Council grant [BB/P021042/1] to S.F.N.

### Acknowledgements

CMML samples were provided by the Cambridge Blood and Stem Cell Biobank, which is supported by the Cambridge NIHR Biomedical Research Centre, Wellcome Trust — MRC Stem Cell Institute and the Cambridge Experimental Cancer Medicine Centre, U.K. The authors thank Rachel Sworn for her contribution to the RT-PCR and qRT-PCR work, Oliver Rogoyski for help with the figures, Ben Towler for helpful discussions plus critical reading of the manuscript, and Clare Rizzo-Singh for technical help.

### Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

### References

- 1 Reis, F.P., Pobre, V., Silva, I.J., Malecki, M. and Arraiano, C.M. (2013) The RNase II/RNB family of exoribonucleases: putting the 'Dis' in disease. *Wiley Interdiscip. Rev. RNA* **4**, 607–615 <https://doi.org/10.1002/wrna.1180>
- 2 Robinson, S.R., Oliver, A.W., Chevassut, T.J. and Newbury, S.F. (2015) The 3' to 5' exoribonuclease DIS3: from structure and mechanisms to biological functions and role in human disease. *Biomolecules* **5**, 1515–1539 <https://doi.org/10.3390/biom5031515>

- 3 Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C. et al. (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 <https://doi.org/10.1038/nature09837>
- 4 Lohr, J.G., Stojanov, P., Carter, S.L., Cruz-Gordillo, P., Lawrence, M.S., Auclair, D. et al. (2014) Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* **25**, 91–101 <https://doi.org/10.1016/j.ccr.2013.12.015>
- 5 Walker, B.A., Wardell, C.P., Melchor, L., Hulkki, S., Potter, N.E., Johnson, D.C. et al. (2012) Intracloonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. *Blood* **120**, 1077–1086 <https://doi.org/10.1182/blood-2012-03-412981>
- 6 Weißbach, S., Langer, C., Puppe, B., Nedeva, T., Bach, E., Kull, M. et al. (2015) The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Br. J. Haematol.* **169**, 57–70 <https://doi.org/10.1111/bjh.13256>
- 7 Welch, J.S., Ley, T.J., Link, D.C., Miller, C.A., Larson, D.E., Koboldt, D.C. et al. (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 <https://doi.org/10.1016/j.cell.2012.06.023>
- 8 Camps, J., Pitt, J.J., Emons, G., Hummon, A.B., Case, C.M., Grade, M. et al. (2013) Genetic amplification of the NOTCH modulator LNX2 up-regulates the WNT/β-catenin pathway in colorectal cancer. *Cancer Res.* **73**, 2003–2013 <https://doi.org/10.1158/0008-5472.CAN-12-3159>
- 9 de Groen, F.L., Krijgsman, O., Tijssen, M., Vriend, L.E., Ylstra, B., Hooijberg, E. et al. (2014) Gene-dosage dependent overexpression at the 13q amplicon identifies DIS3 as candidate oncogene in colorectal cancer progression. *Genes Chromosomes Cancer* **53**, 339–348 <https://doi.org/10.1002/gcc.22144>
- 10 Lim, J., Kuroki, T., Ozaki, K., Kohsaki, H., Yamori, T., Tsuruo, T. et al. (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**, 921–925
- 11 Rose, A.E., Poliseno, L., Wang, J., Clark, M., Pearlman, A., Wang, G. et al. (2011) Integrative genomics identifies molecular alterations that challenge the linear model of melanoma progression. *Cancer Res.* **71**, 2561–2571 <https://doi.org/10.1158/0008-5472.CAN-10-2958>
- 12 Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. and Tollervey, D. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410 <https://doi.org/10.1093/emboj/18.19.5399>
- 13 Towler, B.P., Jones, C.I., Viegas, S.C., Apura, P., Waldron, J.A., Smalley, S.K. et al. (2015) The 3′-5′ exoribonuclease Dis3 regulates the expression of specific microRNAs in *Drosophila* wing imaginal discs. *RNA Biol.* **12**, 728–741 <https://doi.org/10.1080/15476286.2015.1040978>
- 14 Tomecki, R., Drazkowska, K., Kucinski, I., Stodus, K., Szczesny, R.J., Gruchota, J. et al. (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**, 1270–1290 <https://doi.org/10.1093/nar/gkt930>
- 15 Szczepińska, T., Kalisiak, K., Tomecki, R., Labno, A., Borowski, L.S., Kulinski, T.M. et al. (2015) DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res.* **25**, 1622–1633 <https://doi.org/10.1101/gr.189597.115>
- 16 Frazão, C., McVey, C.E., Amblar, M., Barbas, A., Vonnrhein, C., Arraiano, C.M. et al. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* **443**, 110–114 <https://doi.org/10.1038/nature05080>
- 17 Lebreton, A., Tomecki, R., Dziembowski, A. and Séraphin, B. (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* **456**, 993–996 <https://doi.org/10.1038/nature07480>
- 18 Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G. Sanchez-Rotunno, M. et al. (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat. Struct. Mol. Biol.* **16**, 56–62 <https://doi.org/10.1038/nsmb.1528>
- 19 Schneider, C., Leung, E., Brown, J. and 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**, 1127–1140 <https://doi.org/10.1093/nar/gkn1020>
- 20 Schaeffer, D., Reis, F.P., Johnson, S.J., Arraiano, C.M. and van Hoof, A. (2012) The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. *Nucleic Acids Res.* **40**, 9298–9307 <https://doi.org/10.1093/nar/gks693>
- 21 Cairão, F., Arraiano, C. and Newbury, S. (2005) *Drosophila* gene tazman, an orthologue of the yeast exosome component Rrp44p/Dis3, is differentially expressed during development. *Dev. Dyn.* **232**, 733–737 <https://doi.org/10.1002/dvdy.20269>
- 22 Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H. and Dziembowski, A. (2013) Exonuclease hDIS3L2 specifies an exosome-independent 3′-5′ degradation pathway of human cytoplasmic mRNA. *EMBO J.* **32**, 1855–1868 <https://doi.org/10.1038/emboj.2013.135>
- 23 Malecki, M., Viegas, S.C., Carneiro, T., Golik, P., Dressaire, C., Ferreira, M.G. et al. (2013) The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. *EMBO J.* **32**, 1842–1854 <https://doi.org/10.1038/emboj.2013.63>
- 24 Staals, R.H., Bronkhorst, A.W., Schilders, G., Slomovic, S., Schuster, G., Heck, A.J. et al. (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome. *EMBO J.* **29**, 2358–2367 <https://doi.org/10.1038/emboj.2010.122>
- 25 Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J. et al. (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J.* **29**, 2342–2357 <https://doi.org/10.1038/emboj.2010.121>
- 26 Makino, D.L., Baumgärtner, M. and Conti, E. (2013) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495**, 70–75 <https://doi.org/10.1038/nature11870>
- 27 Wasmuth, E.V. and 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**, 133–144 <https://doi.org/10.1016/j.molcel.2012.07.012>
- 28 Han, J. and van Hoof, A. (2016) The RNA exosome channeling and direct access conformations have distinct in vivo functions. *Cell Rep.* **16**, 3348–3358 <https://doi.org/10.1016/j.celrep.2016.08.059>
- 29 Schneider, C. and Tollervey, D. (2013) Threading the barrel of the RNA exosome. *Trends Biochem. Sci.* **38**, 485–493 <https://doi.org/10.1016/j.tibs.2013.06.013>
- 30 Januszyc, K. and Lima, C.D. (2014) The eukaryotic RNA exosome. *Curr. Opin. Struct. Biol.* **24**, 132–140 <https://doi.org/10.1016/j.sbi.2014.01.011>
- 31 Wang, H.W., Wang, J., Ding, F., Callahan, K., Bratkowski, M.A., Butler, J.S. et al. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3′ end processing. *Proc. Natl Acad. Sci. U.S.A.* **104**, 16844–16849 <https://doi.org/10.1073/pnas.0705526104>
- 32 Glavan, F., Behm-Ansmant, I., Izaurralde, E. and Conti, E. (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J.* **25**, 5117–5125 <https://doi.org/10.1038/sj.emboj.7601377>
- 33 Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 <https://doi.org/10.1038/nprot.2009.2>

- 34 Bleichert, F., Granneman, S., Osheim, Y.N., Beyer, A.L. and Baserga, S.J. (2006) The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proc. Natl Acad. Sci. U.S.A.* **103**, 9464–9469 <https://doi.org/10.1073/pnas.0603673103>
- 35 Fatica, A. and Bozzoni, I. (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7–21 <https://doi.org/10.1038/nrg3606>
- 36 Delan-Forino, C., Schneider, C. and Tollervey, D. (2017) Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. *PLoS Genet.* **13**, e1006699 <https://doi.org/10.1371/journal.pgen.1006699>
- 37 Lionetti, M., Barbieri, M., Todoerti, K., Agnelli, L., Fabris, S., Tonon, G. et al. (2015) A compendium of DIS3 mutations and associated transcriptional signatures in plasma cell dyscrasias. *Oncotarget* **6**, 26129–26141

## Supplemental methods

### *Overexpression, purification and in vitro activity assays of the PIN truncated version of human Dis3 isoform1 and respective mutants*

The vector pRSV-1, expressing the truncated version of isoform 1, composed only of the PIN domain fused to the protein purification tag, was amplified by PCR with primers that cover the entire sequence but insert alterations in their sequences to create the respective mutant versions, PIN<sup>E97A</sup> and PIN<sup>D146N</sup>. In each case, a silent mutation (not creating any other amino-acid change) was inserted also through each primer, to create an extra restriction site to help with the selection of the positive clones. In this case, an extra *ScaI* restriction site was inserted with E97A mutation and a *BsmI* restriction site inserted with D146N mutation. The PCR products were circularized with T4 DNA ligase and used to transform competent DH5 $\alpha$  strains (vectors pRSV-3 and pRSV-4 for PIN<sup>E97A</sup> and PIN<sup>D146N</sup>, respectively). The sequence of the selected clones was confirmed by DNA sequencing.

The PIN truncated versions containing only the respective PIN domain (PIN<sup>iso1</sup>, PIN<sup>E97A</sup> and PIN<sup>D146N</sup>) were overproduced with a Glutathione S-transferase (GST) tag in *E. coli* BL21-CodonPlus(DE3)-RIL strain containing the recombinant plasmids of interest. The proteins were overexpressed and purified by affinity chromatography (as described in the main manuscript). *In vitro* activity assays of the proteins were performed using the synthetic 30-mer oligoribonucleotide ss16-A14 as a substrate, labelled at its 5' end with [ $\gamma$ -<sup>32</sup>ATP] and T4 Polynucleotide Kinase (Ambion), and circularized with T4 RNA ligase (Thermo). Protein and RNA concentrations were 250 nM and 25 nM, respectively. The experiment was performed in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 3mM MnCl<sub>2</sub> and 1 mM DTT and following the procedure described in the main manuscript.



**Table S1. List of plasmids used in this work.**

<b>Plasmid</b>	<b>Comments</b>	<b>Reference</b>
H0869	pReceiver-B03 vector expressing DIS3 ISO1 fused to a GST-tag; Amp <sup>R</sup>	Genecopoeia
H3667	pReceiver-B03 vector expressing DIS3 ISO2 fused to a GST-tag; Amp <sup>R</sup>	Genecopoeia
pRSV-1	pReceiver-B03 vector expressing PIN domain of DIS3 ISO1 fused to a GST-tag; Amp <sup>R</sup>	This study
pRSV-2	pReceiver-B03 vector expressing PIN domain of DIS3 ISO2 fused to a GST-tag; Amp <sup>R</sup>	This study
pRSV-3	pReceiver-B03 vector expressing PIN domain of DIS3 ISO1 <sup>E97A</sup> fused to a GST-tag; Amp <sup>R</sup>	This study
pRSV-4	pReceiver-B03 vector expressing PIN domain of DIS3 ISO1 <sup>D146N</sup> fused to a GST-tag; Amp <sup>R</sup>	This study
pGEX-4T-1	Protein expression vector with GST-tag; Amp <sup>R</sup>	GE Healthcare

**Table S2. Oligonucleotides used in this work**

<b>Name</b>	<b>Sequence 5'-3'</b>	<b>Comments</b>
30-mer	5'-CCCGACACCAACCACU-A <sub>14</sub>	RNA oligo used as substrate (circularized)
PIN FRW	5'-CTCGAGTGCGGCCGCAACCCAGCTTTC	oligo used in the construction of pRSV-1/2
PIN REV Mut	5'-CCACTTTCTATTTAATTCCCTTCTTCAG	oligo used in the construction of pRSV-1/2
E97A FRW ScaI	5'-CTACAAACAGTACTTCAAGCAGTGAGAAATCGC	insertion of E97A mutation (vector pRSV-3)
E97 REV	5'- CACAATTACATTCTGATGGC	construction of pRSV-3 vector
Primer 15 (D146N) BsmI	5'-AGGGAGAGGAATGCTAATGACAGGAATAATAGAG	insertion of D146N mutation (vector pRSV-4)
Primer 15 Rev	5'-GTTCTTGTTCTACATAGGTTTCTCTATGGTGC	construction of pRSV-4 vector

\* all the DNA primers are phosphorylated at the 5' end to enable the circularization of the vector

**Table S3. 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 accordance to MIQE guidelines.

Oligonucleotide	Accession number	Sequence (5' to 3')	Anchor nucleotide	Amplicon length
DIS31 (Forward/Reverse)	NM_014953.4	N/A	N/A	114
DIS32 (Forward/Reverse)	NM_001128226.2	N/A	N/A	84
GAPDH (Forward/Reverse)	NM_002046	N/A	1087	110

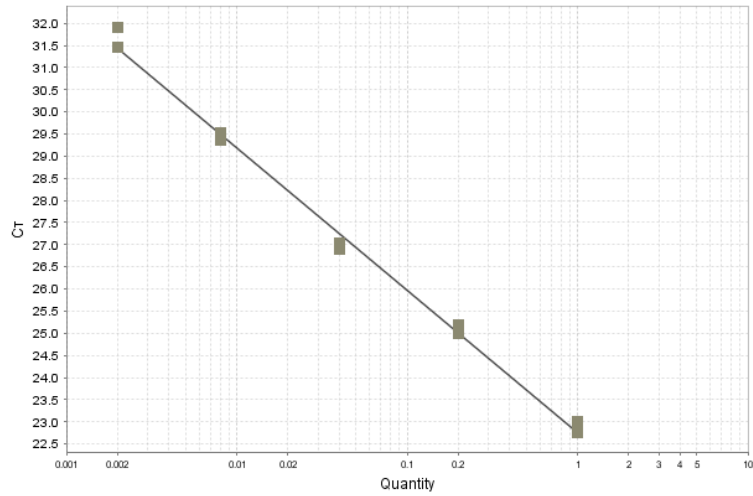
**Table S4. List of cell lines used in this study. The cell origins and tissue type are given in brackets).**

<b>Cell Line</b>	<b>Cell Type &amp; Origin</b>
RPMI-8226	B-lymphoblast from plasmacytoma in peripheral blood of 61 year old male with myeloma (B-lymphocyte; peripheral blood)
U-266	B-lymphoblast from plasmacytoma in peripheral blood of 53 year old male with myeloma (B-lymphocyte; peripheral blood)
KMS-12-BM	Bone marrow of a 64 year old female with myeloma (B-lymphocyte; peripheral blood)
MOLP-8	Peripheral blood of 52 year old man with myeloma (B-lymphocyte; peripheral blood)
HeLa	Cervical epithelial cell of a 31 year old female with cervical adenocarcinoma (epithelial; cervix)
SAOS-2	Bone cells from an 11 year old female with osteosarcoma (partially differentiated mesenchymal stem cells; bone)
THP-1	Monocytes from peripheral blood of a 1 year old male with acute monocytic leukaemia (monocyte: peripheral blood)
GM12878	B-lymphoblast from peripheral blood of female and transformed with Epstein bar virus
U-2OS	Bone epithelial cell from a 15 year old female with osteosarcoma (partially differentiated mesenchymal stem cells; bone)
D-G75	B-lymphoblast from pleural effusion of 10 year-old male with Burkitt's Lymphoma (B-lymphocyte; pleural effusion)
KG-1	Macrophage from bone marrow of 59 year old male with acute myelogenous leukaemia (macrophage;bone marrow)
OCI-AML3	Peripheral blood of a 57 year old male with acute myeloid leukaemia (myeloid; peripheral blood)
HEK-293	Epithelial cell from human embryonic kidney (epithelial-like; embryonic kidney)

**Table S5.** Summary of clinical details of human participants used in this study.

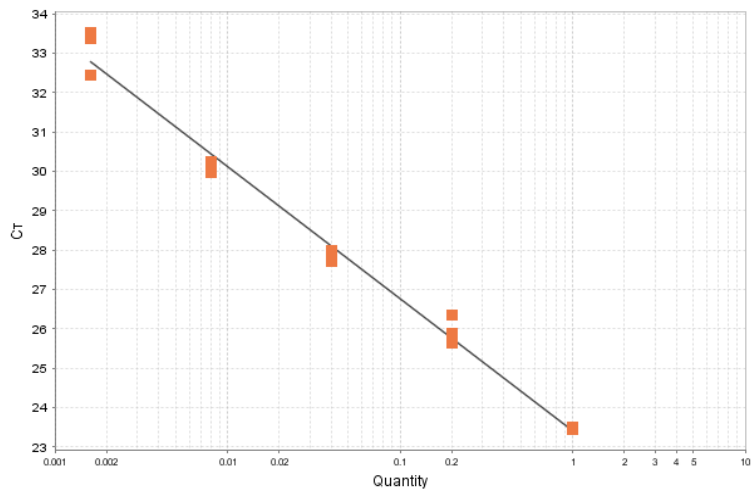
Disease	Patient	Age	Sex	Disease	Patient	Age	Sex
Myeloma	1	78	M	AML	1		
	2	69	F		2	83	M
	3	86	M		3	68	M
	4	72	M		4	74	M
	5	73	F		5	76	F
	6	71	F		6	81	M
	7	78	F		7	60	M
	8	51	F		8	46	F
	9	47	M		9	76	M
	10	65	M		10	58	M
	11	54	M		11	74	F
CMML	1	94	M				
	2	62	F				
	3	67	M				
	4	60	F				
	5	53	M				
	6	63	M				
	7	74	F				
	8	66	M				
	9	75	F				
	10	71	M				
	11	72	M				
	12	72	F				
	13	78	F				
	14	85	F				
	15	79	F				

### DIS3 Isoform 1



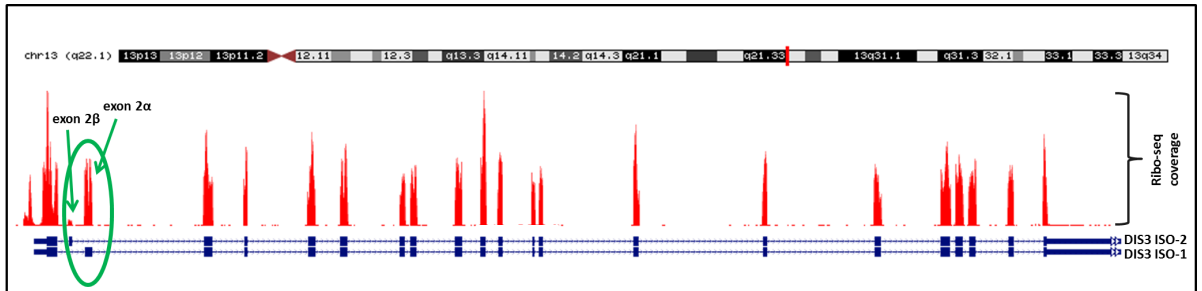
Target: dis31 Slope: -3.213 Y-Inter: 22.757  $R^2$ : 0.996 Eff%: 104.742 Error: 0.059

### DIS3 Isoform 2

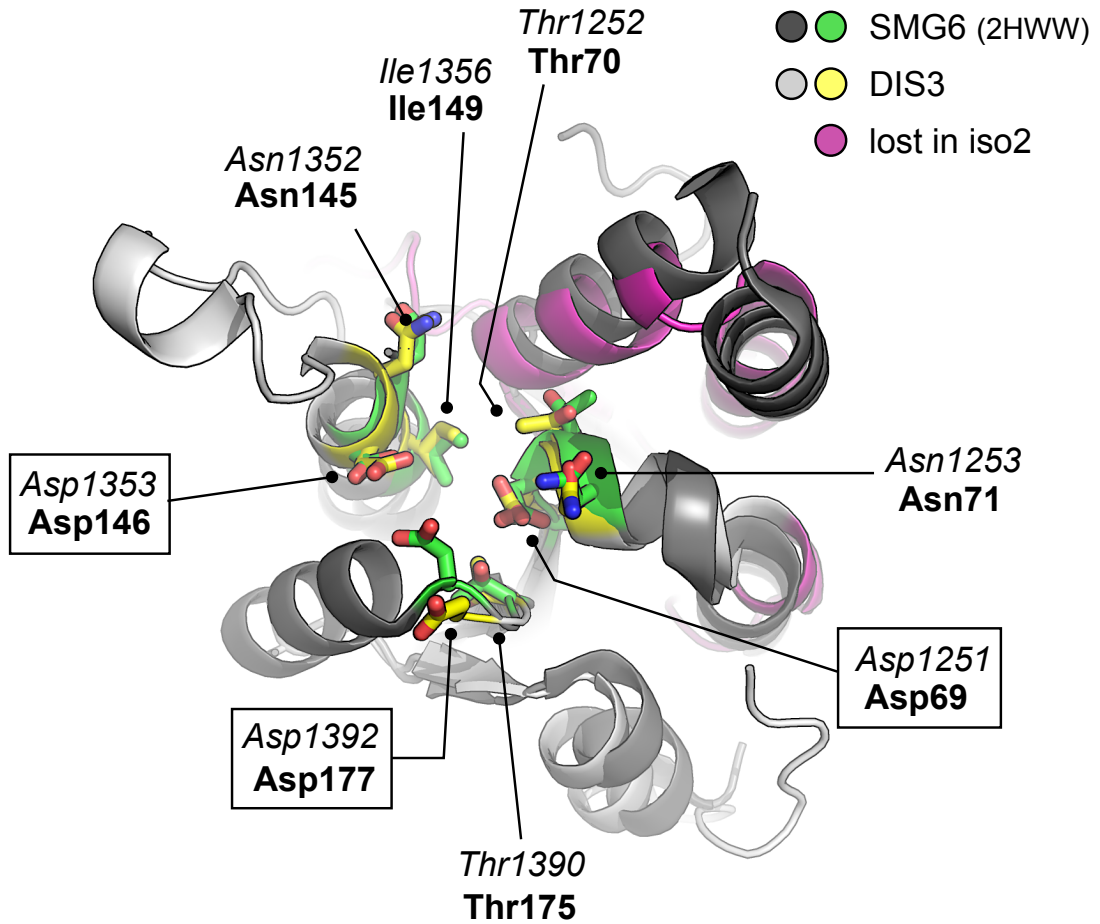


Target: dis32 Slope: -3.349 Y-Inter: 23.418  $R^2$ : 0.987 Eff%: 98.866 Error: 0.105

**Figure S1 . Standard curves and amplification efficiency of the DIS3 isoform 1 and DIS3 isoform 2 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.

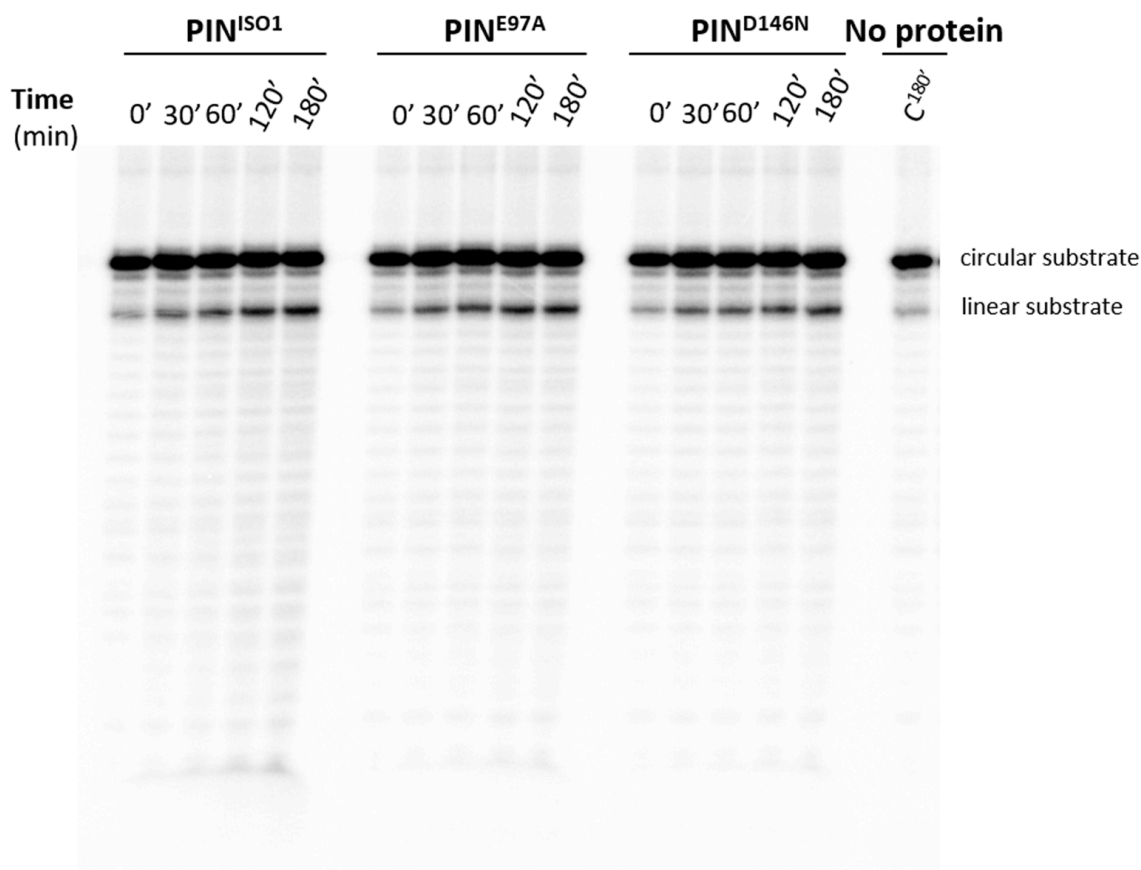


**Figure S2. Visualisation of ribo-seq data on the DIS3 transcripts obtained by ribosome profiling.** Although coverage is much lower on the shorter exon 2 of isoform 2 (exon 2 $\beta$ ), ribosome binding is above background level indicating this isoform is translated. Alternative exons are labelled. Data obtained using the online tool GWIPS.

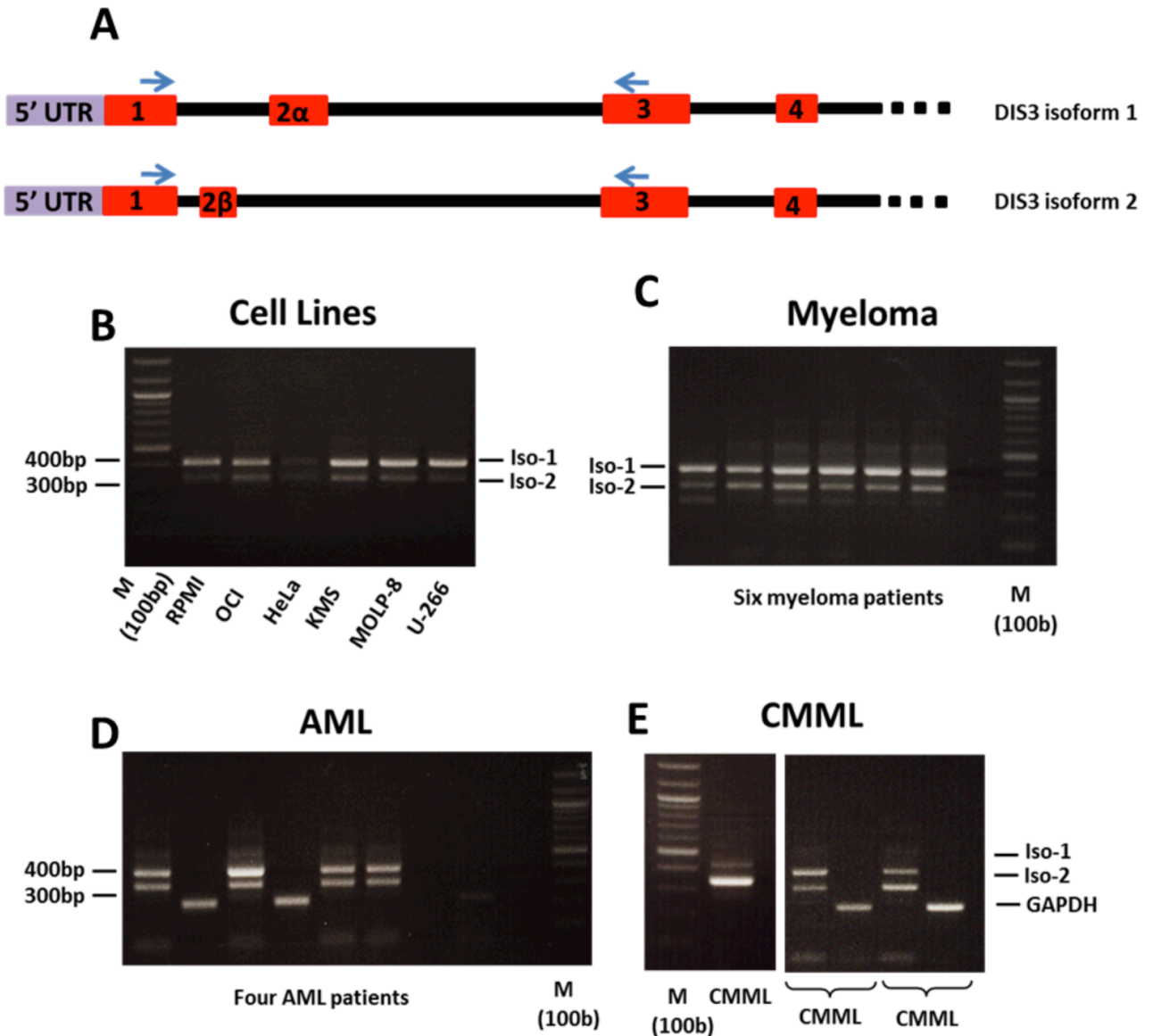


**Figure S3. Diagram comparing the active sites of human DIS3 PIN domain with human SMG6.** The three core catalytic residues (Asp 69, Asp 146 and Asp177) are conserved in both identity and spatial arrangements between the two proteins. The region lost in isoform 2 (purple) is unlikely to affect the catalytic activity of the PIN domain of isoform 2.

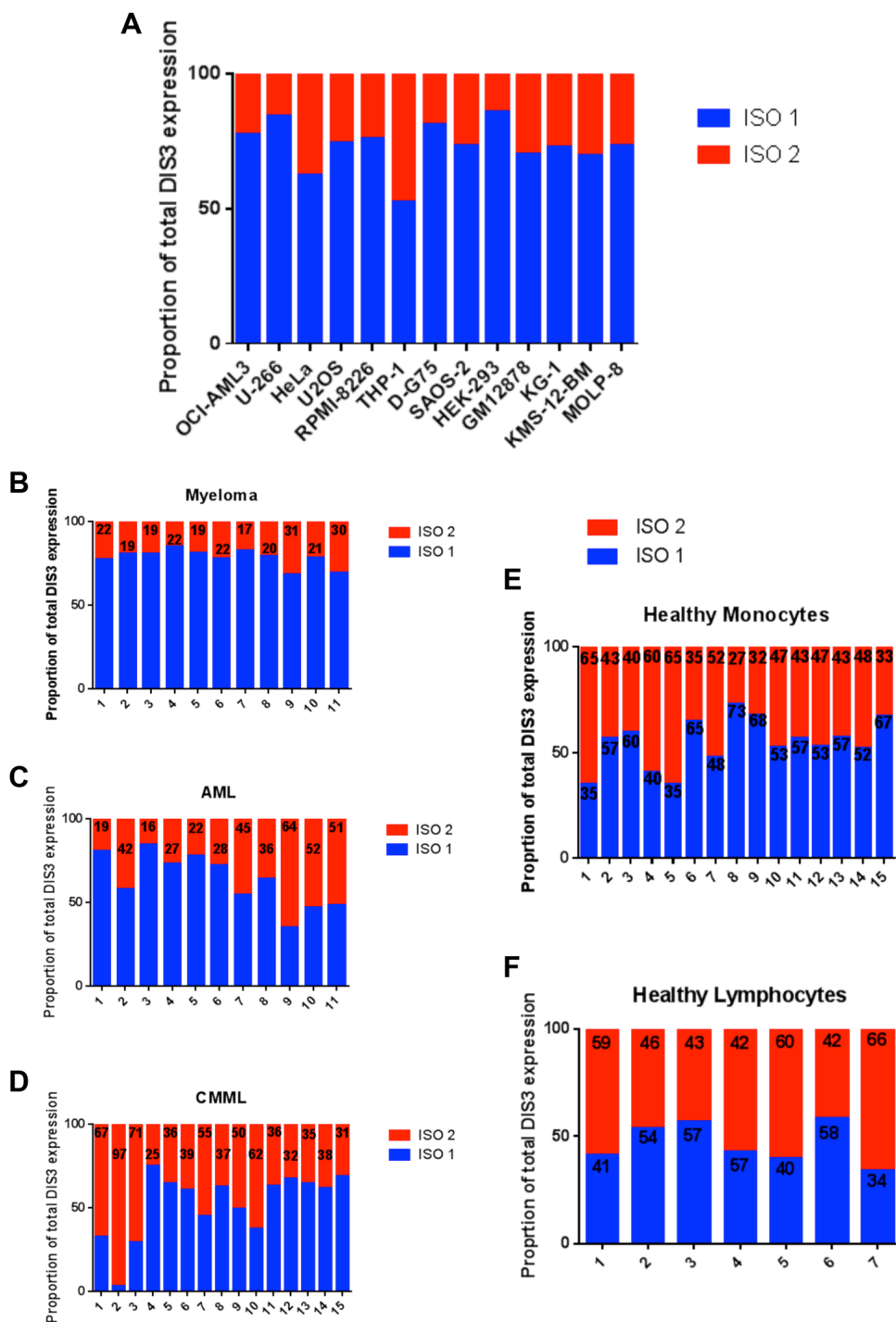




**Figure S4. Effect off two single point mutations on the endoribonucleolytic activity of PIN domain on a circular RNA substrate.** A 5'-labeled 30mer ssRNA circularized oligonucleotide (upper band) was incubated with a purified truncated version of human DIS3 isoforms 1 containing only the PIN domain (PIN<sup>iso1</sup>) and the same protein either with a mutation in the glutamic acid residue in position 97 to an alanine (PIN<sup>E97A</sup>) or a mutation in aspartic acid residue 146 to an asparagine (PIN<sup>D146N</sup>). Incubation times are indicated on top of the panels. The lower band in the substrate is a linear molecule resulting from the suboptimal yield of the circularization reaction. The E97A mutation, in one of the conserved residues of PIN domain necessary for the metal ion coordination, causes a considerable reduction in activity. The activity of the protein is practically abolished by the mutation in the catalytic residue 146 (D146N), as previously reported [17,19]. For comparison of the activity of PIN<sup>iso1</sup> with the respective mutants, the concentrations of protein and RNA were 250 nM and 25 nM, respectively. The experiments were performed in a buffer containing 3mM MnCl<sub>2</sub> (endoribonucleolytic buffer conditions). See Supplemental methods for further details.



**Figure S5. 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 in four of the samples.



**Figure S6.** Expression of the two DIS3 isoforms as a proportion of total DIS3 expression (A) Expression of the two isoforms as a proportion of total DIS3 expression in a range of cell lines. (B) In myeloma patient samples the proportion of isoform 1:isoform 2 is, on average, 78:22. (C) In AML patient samples Isoform 1 is expressed at higher levels than isoform 2; the proportion of isoform 1:isoform 2 is 64:36. (D) In CMML patient samples, levels of isoform 1 and 2 are similar with an average proportion of isoform 1:isoform 2 being 53:47. (E and F) Isoform 1 and 2 contribute approximately equally to total DIS3 levels within individual patients. In healthy monocytes, the average proportion of isoform 1;isoform 2 is 65:45 whereas in healthy lymphocytes is 49:51. Proportions represent average of three biological replicates.