DIS3 ribonuclease prevents the cytoplasmic accumulation of lncRNAs carrying non-canonical ORFs.

Dominika Foretek, Marc Gabriel, Isabelle Hatin, Julien Jarroux, Marina Pinskaya, Elise Pepermans, Kurt Boonen, Rachel Topno, Vera Slaninova, Marina Serna, et al.

Accumulation of lncRNAs in cytoplasm upon DIS3 depletion leads to production of cryptic peptides detected in Multiple Myeloma.

Dominika Foretek
Institut Curie
Marc Gabriel
Institut Curie
Isabelle Hatin
Institute for Integrative Biology of the Cell
Julien Jarroux
Weill Cornell Medicine https://orcid.org/0000-0002-7185-6960
Marina Pinskaya
Institut Curie-CNRS
Elise Pepermans
Centre for Proteomics (CFP), University of Antwerp
Kurt Boonen
Centre for Proteomics (CFP), University of Antwerp
Rachel Topno
Institut Genetique Humaine
Vera Slaninova
Institut Genetique Humaine
Marina Serna
Structural Biology Programme, Spanish National Cancer Research Center (CNIO)
Oscar Llorca
Structural Biology Programme, Spanish National Cancer Research Center (CNIO)
Geert Baggerman
Centre for Proteomics (CFP), University of Antwerp
Edouard Bertrand
Institut de Génétique Humaine https://orcid.org/0000-0002-9642-7994
Steven West
Living Systems Institute, University of Exeter
Olivier Namy
Institute for Integrative Biology of the Cell
Antonin Morillon (antonin.morillon@curie.fr)

ncRNA, epigenetic and genome fluidity, CNRS UMR 3244, Sorbonne Université, Université PSL, Institut Curie, Centre de Recherche, 26 rue d'Ulm, 75248 Paris https://orcid.org/0000-0002-0575-5264

Research Article

Keywords: RNA decay, DIS3, lncRNA, multiple myeloma, novel ORF, tumour-specific antigen

Posted Date: January 17th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3006132/v3

Additional Declarations: The authors declare no competing interests.

Abstract

Some long noncoding (lnc)RNAs harbor the potential to produce functional micropeptides. Despite the increasing recognition of their significance, the regulatory dynamics of cytoplasmic lncRNA expression, decay, and translation remain poorly understood. Here, we investigate the role of ribonucleases in controlling cytoplasmic levels of lncRNAs. By transcriptomic analysis we identified DIS3 but not XRN1 as a major enzyme preventing accumulation of lncRNAs in cytoplasm. Single-molecule experiments illustrate an example of DIS3-sensitive transcript (DIST) accumulation in the nucleus preceding the one in the cytoplasm, suggesting a sequential series of events. Approximately 14.5% of the DISTs contain at least one actively translated open reading frame (ORF). This finding is highly relevant to Multiple Myeloma bone marrow cancer patients' cases with mutations impairing the DIS3 enzymatic activity and revealing a subgroup of overexpressed translatable DISTs. Immunopeptidomic approach identified the association of DIST-derived peptides with the major histocompatibility complex class I (MHCI). Notably, the low expression of DISTs in healthy tissues emphasizes their potential as targets for cancer-specific immunotherapies. Our findings shed light on the intricate regulatory mechanisms governing cytoplasmic lncRNA dynamics and highlight their clinical relevance in the context of bone marrow cancers, providing a foundation for future investigations into novel therapeutic strategies.

Introduction

Long noncoding RNAs (lncRNAs), a heterogeneous class of transcripts exceeding 200 nucleotides without coding potential, have emerged as critical players in the intricate landscape of gene regulation[1]. While the majority of these molecules predominantly function within the nucleus, orchestrating a myriad of regulatory processes at both transcriptional and post-transcriptional levels[2], subset of lncRNAs has been identified with ubiquitous or cytosolic residence[3–5], underlining their importance in inferred pathways [reviewed in[6]].

Despite significant progress in understanding nuclear lncRNA functions, the regulatory dynamics of their cytoplasmic counterparts remain underexplored. A crucial determinant of the cellular abundance of these transcripts lies in their degradation, with emerging evidence suggesting that the average half-lives of lncRNAs are notably shorter than those of messenger RNAs (mRNAs)[7–9], at least for those presenting similar transcriptional rates as mRNA[10,11]. There are two main exonucleolytic pathways that degrade RNA, either in 5' - 3' or 3' - 5' direction. In the human system, the 3' - 5 pathway is governed by a plethora of catalytic enzymes, with EXOSC10 localized in the nucleolus, DIS3L and DIS3L2 in the cytoplasm, and DIS3 predominantly in the nucleus, with minor fractions in the cytosol[12–14]. Notably, EXOSC10, DIS3L and DIS3 enzymes are catalytic subunits of the exosome complex, responsible for degrading unspliced, incorrectly polyadenylated transcripts, or products of endonucleolytic cleavage [reviewed in[15]]. Distinct from the exosome complex, DIS3L2 operates independently and degrades RNAs via an oligouridylation-dependent pathway[16,17]. Meanwhile, the 5' - 3' exonucleases XRN2 and XRN1 function in the nucleus and cytoplasm, respectively [reviewed in[18]]. XRN2 is mostly responsible for co-transcriptional digestion of 3'

flanking region RNA promoting transcription termination[19]. XRN1 is one of the main contributors of the cytoplasmic mRNA surveillance, digesting the 5'end of uncapped and 3' deadenylated RNAs [reviewed in[20]]. Together with 5'-3' nucleases, XRN1 is involved in quality control mechanisms of translated RNAs, via Nonsense-Mediated Decay (NMD) of transcripts with premature stop codons (PTCs), No-Go Decay (NGD) for those with stalled ribosomes, and Non-Stop Decay (NSD) if they lack a stop codon [reviewed in[21]]. Yet it remains to be defined which of those pathways are main actors regulating the levels of the human cytoplasmic lncRNAs.

In mammalian cells XRN1 has been shown to target some lncRNAs in osteosarcoma[22] and a subset of lncRNAs that are hosts of snoRNA genes[23], but this has been observed for a limited number of targets so far. Also 3'-5' nucleases might play a role in limiting cytoplasmic lncRNA accumulation, since RNA-uridylation, triggering the DIS3L2-dependent degradation, was described for some intergenic lncRNAs (lincRNAs)[17]. In addition, DIS3, even though mostly nuclear, was implicated in degradation of several examples of cytoplasmic, poly-adenylated (polyA) transcripts in human cells[24]. Interestingly, no genome-wide study has yet addressed the identity of the exonucleases involved in regulating global levels of human lncRNAs in cytoplasm nor the functional significance of their misregulation.

In budding and fission yeasts, Xrn1 deletion resulted in the discovery of a class of cytoplasmic unstable lncRNAs called Xrn1-sensitive unstable transcripts (XUTs)[25–27], for which the translation and NMD are pivotal for their stability[28]. Multiple evidence show that also in mammals some lncRNAs can contain Open Reading Frames (ORFs) of variable length, together with other mRNA-like features and can be bound by ribosomes and potentially translated[4,29–31]. Indeed, lncRNA-derived peptides have been detected in different cellular contexts[32–35] of which some are reported to have functional significance [reviewed in[36]]. Moreover, due to lncRNAs context-specific expression[37], lncRNA-derived peptides could have specific significance in cancer as tumor specific antigens (TSAs) that can be bound and presented by the MHC and thus used targets for immunotherapy. As novel methods have increased the confidence in identifying true peptides originating from sORF[33,38], the number of novel TSAs has expanded[39,40] including lncRNA-encoded TSAs[41,42]. Furthermore, peptides derived from unstable mRNAs, sensitive to NMD have been shown to bind to the MHCI[43], further enlarging the potential source of novel antigens; it is tempting to hypothesize whether peptides originating from NMD-sensitive lncRNAs could follow the same path. In all, misregulation of RNA decay pathways could generate a large number of potential TSAs originating from cryptic lncRNAs.

In this work, we aimed to identify the enzyme(s) regulating lncRNAs levels in the cytoplasm as well as to explore the potential of cytoplasmic lncRNAs as translation templates. The large number of RNA decay co-factors and exonucleases present in mammalian cells in comparison to unicellular eukaryotes, complicates the task of assessing the importance of each of the nucleases in the decay of cytoplasmic lncRNAs. Thus, we used an auxin-inducible degron system that enables a conditional rapid depletion of a protein of interest[44]. Using cellular fractionation and RNA sequencing from cells depleted of XRN1 or DIS3, we show that XRN1 has a limited role and degrades only a small subset of lncRNAs within the

cytoplasm, while DIS3 degrades a larger proportion of lncRNAs, that we named DIS3 sensitive Transcripts (DISTs). Single-molecule RNA fluorescent *in situ* hybridization (smiFISH) on a time-course experiment, following DIS3 inactivation, indicates that, for at least some of the DISTs, the nuclear accumulation can precede the cytoplasmic one. Further, ribosome profiling and immunopeptidomic analyses show that some DISTs are bound by ribosomes, are translated, and some of these are presented by the MHCI. Moreover, some DISTs identified in our *in vitro* model are also overexpressed in multiple myeloma (MM) cases, which present mutations in the *DIS3* gene. We propose that these DISTs represent a potential novel source of TSAs, as they are specifically expressed in malignant tissues and are identified as bound by MHCI.

Results

DIS3 is the major exonuclease regulating lncRNA levels in the cytoplasm

To prioritize potential enzymes and co-factors that could regulate lncRNA decay in the cytoplasm, we first performed differential expression (DE) analyses using public total RNA-seq datasets, including a XRN1 knock-out dataset from HEK293 cells (XNR1 KO)[45], DIS3 mutants with inactivated RNB and PIN domains[13] (DIS3 PIN RNB mut), and knock-downs (KD) of cytoplasmic RNA decay cofactors HBS1LV1, HBS1LV3 and SKIV2L[46] (supplemental Table S1). Among these, the DIS3 PIN RNB mutation showed the most striking upregulation of 72.5% of all expressed lncRNAs (Fig. 1A and Supplemental Table S2). While inactivation of the cytoplasmic factors XRN1 or HBS1LV3, showed moderate percentages of lncRNAs differentially expressed (17.2% and 19.1% respectively), XRN1 KO had the higher absolute number of overexpressed lncRNAs (527 from 3059). Thus, we focused our comparative study on DIS3 and XRN1 enzymes.

To identify the direct effect of DIS3 and XRN1 in lncRNAs stability, we opted for the existing DIS3 auxin-degron cell model (DIS3-AID) and we constructed the XRN1-AID HCT116 cell line with TIR1 controlled by a Tet-On promoter (see materials and methods), which enabled timely removal of the enzymes[47] (Fig. S1A and S1B). Total RNAs from the cytoplasmic fraction of cells induced or not for degron-mediated DIS3 or XRN1 depletion (Fig. S1C- S1F) were sequenced (CYTO-seq) and DE analysis was performed.

Since DIS3 and XRN1 can degrade cryptic and *a priori* unannotated transcripts[13,25], we quantified RNA expression by mapping the read counts on both standard GENCODE v26 annotation, and the *de novo* annotation generated by Scallop transcript assembler[48]. We defined different RNA biotypes that show statistically significant increase in expression in the cytoplasm of the XRN1- or DIS3-depleted cells [see materials and methods]. LncRNAs represented 77% of all biotypes of annotated transcripts upregulated after DIS3 depletion and only 12% following XRN1 depletion (Fig 1B and Table S2). This result reinforced the previous finding that XRN1 is mostly dedicated to the degradation of protein coding transcripts in human cells, prompting us to further explore the role of DIS3 in cytoplasmic lncRNA decay.

To compare the cytoplasmic mRNA and lncRNAs sensitive to DIS3, we plot the coverage of reads within the exonic and intronic regions within the GENCODE annotated upregulated lncRNAs and mRNAs, both groups selected as containing at least 2 introns and not classified as premature transcripts (PT)[13] (Fig. 1C and Fig. S1G-I and material and methods). The resulting metagene showed that, in contrast to mRNA, the lncRNAs present reads on both exons and introns at similar levels, suggesting that the DIS3-sensitive lncRNAs in the cytoplasm either remain broadly unprocessed or display unknown isoforms. This brought our attention to another group of DIS3 targets, that Scallop profiling will miss- the intragenic or sense overlapping transcripts, such as enhancer RNAs (eRNAs), a well-known class of nuclear DIS3 substrate[49]. Using the recently published annotation of eRNAs[50], we further identified 3025 upregulated eRNAs (Fig 1D and 1E), raising a number of DIS3-sensitive transcripts (DISTs) to 5166 out of 8779 (59%) expressed lncRNAs and eRNAs.

In conclusion, our analysis showed that in contrast to the minor role of XRN1 in finetuning lncRNA steady-state levels (5.8%), DIS3 prevents the accumulation of a large number of lncRNAs and eRNAs within the cytoplasm (59%), that were previously considered as restricted mostly to the nuclear environment[13]. This raises the question of DISTs features in comparison with the nuclear DIS3-sensitive lncRNAs.

DIS3-sensitive transcripts found in cytoplasm share many features with nuclear DIS3-targets

To determine how many DIS3-sensitive transcripts were shared by both nuclear and cytoplasmic cellular compartments, we reanalysed data for DIS3 nuclei-seq[47]. Using our annotation, we identified that 77% of upregulated lncRNAs from nuclear-seq were also upregulated in CYTO-seq (Fig 2A) (with p-value< 0.0001 of the overlap, SuperExactTest[51]). In the nucleus, RNAs targeted by the exosome are often products of bidirectional transcription[13,47,49,52]. To test our DISTs for this feature, we refined transcriptional precise start and stop extremities by coverage and length thresholds (see materials and methods), we defined the bidirectionality of remaining 5162 DISTs, and classified them into upstream antisense (ua) RNAs, convergent (con) RNAs and bidirectional eRNAs as defined in a previous study[50]. Our results showed 797 bidirectional eRNA and 1157 ua- and con-RNAs in promoter regions of protein-coding genes (PCGs). In total, only ~38% of the DIS3-sensitive lncRNAs followed RNA Pol II-dependent bidirectionality (Fig. 2B and S2A), suggesting that bidirectionality of transcription is not the main trigger of DIS3-mediated degradation.

To validate and follow in more detail the dynamic of cytoplasmic DISTs accumulation, we performed single-molecule RNA fluorescent *in situ* hybridization (smiFISH)[53] on ZNF674-AS1 (Fig. 2C), after 1h and 4h DIS3-depletion. Our result showed a gradual increase of RNA molecules in the cells with significant increase from a median of 7 to 39-40 molecules in non-treated cells and after 4h of treatment, respectively (Fig. 2D and 2E). Interestingly, when RNA molecules were quantified in nucleus and cytoplasm separately, already after 1h we observed a statistically significant increase of nuclear RNA molecules in both biological replicates. Contrary, in the cytoplasm this was only true in one of two replicates (Fig. 2D). However, after 4h of DIS3 depletion, in both compartments, the number of molecules

increased to a similar level of 19-20 molecules. This suggests that DISTs accumulation in the cytoplasm could be related to their export from nucleus following their stabilization. Similar phenomenon was observed in mRNAs when their stability was modulated[54]. It is important to note that the dynamic and timing differs for different DISTs. For example, CTD-2371O3.3 dramatically accumulated in the nucleus already after 1h of DIS3 depletion (from 5-7 to 45-49 median number of molecules per cell), while more modestly increased in the cytoplasm from (3-5 to 11-9 median number of molecules per cell, Fig. S2B and S2C).

Overall, our global and single-molecule studies confirmed that DISTs are not necessarily transcribed from bidirectional promoters, rapidly accumulate in the nucleus and cytoplasm as soon as the DIS3 enzyme is depleted and raise the question of their cytoplasmic fate**.**

DISTs can be bound by ribosomes with a characteristic pattern of active translation

One of fates of many cytoplasmic RNAs is translation. To identify whether DISTs are bound to ribosomes, we performed ribosome profiling experiments as depicted in (Fig. 3A) including all the 9 possible non-AUG start codons[55], regarding recent data on lncRNAs[32,38,56,57]. In total, 431844 ORFs were predicted to be actively translated in at least one sample across all conditions (supplemental Table S4). Among these we identified 750 DISTs containing at least one ORF predicted to be actively translated in DIS3-depleted cells. Of them, 12% had a canonical AUG start (Fig. S3A) and all were classified as novel ORFs by ribotricer tool (Table S4). The metagene of the highest expressed ORF of each lncRNA showed similar characteristic peaks of ribosomes stalling around start and stop codons of the ORF, as well as periodicity, as is similarly detected in mRNAs (Fig. 3B and 3C). Interestingly, only 4% of DIST-ORFs were also present in cells in which the active exosome was present (Fig. 3D), suggesting that some but very few of these lncRNAs might escape exosome-dependent degradation and be translated in normal conditions. Moreover, not more than 5% of DIST ORFs were present in different databases based on sequence or gene name analysis[58–60], further reinforcing the notion that most of the DISTs with translatable ORFs were only found bound by ribosomes upon DIS3 depletion due to their upregulation in cytoplasm and thus, were not characterized in previous studies.

The obvious determinant potentially involved in translation is the polyA tail [reviewed in[61]]. Since eRNAs and lncRNAs are a class of transcripts with a variable status of polyadenylation[62,63], we performed polyA-enriched RNA-seq on cytoplasmic RNAs followed by a DE analysis, to define which of the translated DISTs can be polyadenylated. Our result showed that 80% of translatable DISTs were expressed in DIS3-depleted cells (Fig. 3F, Table S5), indicating that most of translatable DISTs possibly undergo a classic polyA-dependent mechanism of translation as exemplified in Fig. 3F.

Altogether, we have identified a set of translatable lncRNAs and 80% of them can be detected in the polyA enriched RNA-seq, suggesting that polyA could be an important feature for the DISTs translation capacity. Such a property makes them resemble a canonical mRNA translation template, rather than circular RNAs or histone mRNAs, both devoid of polyA tails but translated [reviewed in [64,65]].

DISTs are mostly monoexonic lncRNAs and in majority contain short ORFs

To further characterize DISTs features, we compared their lengths against DIS3-insensitive lncRNAs and PCGs. Interestingly, DISTs full transcripts including introns are shorter, in contrast to the sum of exons, which were significantly longer than the other lncRNAs and PCGs (Fig. 4A). Both results point to the fact that the majority of DISTs (89%) are monoexonic (Table S3). Interestingly, Mestdagh's group also has confirmed the existence of a large, novel, monoexonic intergenic class of lncRNAs (4,877), that has been overlooked in previous studies[62]. Notably, 12% (575) of them were also identified also as DISTs in this work.

Next, we assessed the ORF size distribution in the highest expressed ORF of each of the 750 potentially translated DISTs to compare them with those detected in DIS3-insensitive PCGs, annotated lncRNAs. As expected, the median ORFs sizes were significantly shorter for DISTs (138 nucleotides (nt) for lncRNAs and 132nt for eRNAs) than for PCGs (666nt) (Fig. 4B). However, both DIST-ORFs of lncRNA and eRNA origin were significantly longer than median length of 105 nt for sORFs in DIS3-insensitive translatable lncRNAs. Analyses of 5'UTR and 3'UTR features did not provide further information that could clearly discriminate DISTs from other transcripts and explain their sensitivity to DIS3 and translation (Fig 4B).

To conclude, DISTs and DIST-ORFs do not show any obvious features that could explain either their DIS3 targeting or translation, other than having significantly longer ORFs than the other DIS3-insensitive lncRNA-ORFs. This property might indicate some functional significance that could be further studied in physiological contexts of DIS3 inactivation.

Cytoplasmic DISTs are detected in multiple myeloma cases

DIS3 is an essential gene and it was shown to be mutated in several cancer types[66]. To understand whether DISTs could be found in a disease context, we focused on Multiple Myeloma (MM), presenting various *DIS3* mutations in up to 15% of cases [67–72]. Interestingly, some of the recurrent mutations in the *DIS3* gene have been shown *in vitro* to impact the activity of its encoded exoribonuclease[73]. Thus, we quantified DIST expression in the total RNA-seq dataset of bone marrow samples from a cohort of 211 MM patients and 64 healthy controls[74] using *kallisto*[75] and our custom-made annotation, including GENCODE, eRNAs and Scallop-identified transcripts.

To identify cases that could accumulate DISTs because of impaired DIS3 activity, we analysed a repertoire of the mutations and single nucleotide polymorphisms present in the MM patients with identified mutations in *DIS3* gene (33 patients). Based on the published structural data, we selected mutations that might affect conformation and structure of the RNB domain (see materials and methods), which is a critical domain for lncRNA degradation[13]. As a result 15 cases were classified as DIS3-mutated and one as being a protein truncation (depicted in red in the structure model in Fig. 5A, and all listed in Table S7). We performed pair-wise DE analyses comparing the counts between the *DIS3*-mutated group and the other patients, or between the *DIS3*-mutated group and the healthy controls. Intersection of MM-

DISTs and DISTs obtained from our *in vitro* model identified 849 upregulated lncRNAs in all three conditions (p-value<0.001, SuperExactTest) (Fig. 5B), of which 203 had also a translatable ORF.

To further validate selected DIST-ORFs in patients data, we performed additional DE analysis, of RNA-seq data, quantifying reads using ORF coordinates, to determine if the DIST-ORFs regions found in the MM patients are upregulated. 66 DIST-ORFs, within 61 transcripts, present a significant upregulation, as illustrated by their expression heatmap in patients (Fig. 5C). Interestingly, the levels of expression of these 61 DISTs were inversely correlated with the levels of *DIS3* expression (Fig. 5D), indicating that DISTs are sensitive not only to the catalytic efficiency of DIS3 but possibly also to the expression of DIS3 itself.

In conclusion, this shows that the DIS3-AID *in vitro* model could be used for transcript predictions in cancer patients not only presenting *DIS3* mutations but also impaired DIS3 expression.

Cytoplasmic DISTs are translated and presented by MHC class I

Encouraged by our transcriptomic results, we further aimed to evaluate whether the products of DIST-ORFs translation could be TSA in MM. After determining the HLA types of the MM patients using seq2HLA[76], we applied the netMHCpan-4.1[77,78] tool for the prediction of peptides binding to MHCI. All 66 ORFs contain at least one peptide sequence predicted to be a strong binder to at least one HLA allele present among MM patients. The median size of the DIST-ORFs does not exceed 46 amino acids (Fig. 4B), putting the detection of DIST-derived peptides at the limit of detection by classical proteomic analysis. For this reason, we opted for an immunopeptidomic approach that gives a higher detection level of non-canonical peptides than standard proteomic and mass spectrometry methods[79] and gives direct evidence of peptide presentation on the surface of the cell.

Our rationale was to use the DIS3-AID cell line, DIS3-depleted for 4h, and two MM cell lines KMS27 and KMM-1, to determine the presentation capacity of DIST-peptides. MHCI purification and analysis of eluted peptides by mass spectrometry (MS) (as depicted in Fig. 6A) revealed a total up to 9360 uniquely assigned peptides between the 3 experiments. The size distribution of detected peptides showed the most prevalent class of peptides bound by MHCI containing 9 amino acids (Fig. S4 A-C). Next, we queried these peptides using *MSFragger*[80] and our database of DIST-ORFs (see materials and methods), filtering out sequences of the UniProt catalogue. In total, 9, 10, and 4 unique DIST-derived peptides were identified in DIS3-AID, KMM-1 and KMS27 cell lines, respectively. To assess the robustness of the identification of these peptides, we compared the observed retention time (RT) to the RT calculated by the DeepLC algorithm[81] based on peptide sequences. For both proteins and DIST-peptides databases, the Pearson correlation coefficient was between 0.96-0.99 with no significant difference (F-test) between standard proteome and lncRNA peptides, confirming a high quality of the identification (Fig. 6B, 6C and S5A). Interestingly, for DIS3-AID and KMM-1, 2 peptides were assigned to multiple DIST-ORFs, from which one was in common in two cell lines. Among the 22 unique DIST-ORF-derived peptides, 3 were discarded based on manual spectra curation and 15 clustered with HLA alleles specific to each cell line by Gibbs clustering[82] (Fig. S5B-D and Table S7), further confirming their specificity. Comparing the

immunopeptidomic results to Ribo-seq data, 7 peptides from DIST-ORFs were detected to be translated by the Ribo-seq experiments in DIS3-depleted cells or other conditions. An example of spectra of a DIST-peptide, differential in the CYTO-seq and in the MM ORF transcriptomic analysis, is depicted in (Fig. 6D). The ions predicted by Prosit[83] corresponded well with the ions identified in MS.

Finally, we assessed the specificity of expression at RNA level of peptide-encoding DISTs in bone marrow samples of MM patients in comparison with healthy controls of the same cohort and 26 healthy tissues from GTEx database for 66 Ribo-seq selected DIST-ORFs and the DIST-ORFs with peptides initially detected in immunopeptidomic analysis (Fig. S5E and S5F. Expression was quantified based on transcripts per million (TPM), only for DISTs with counts above 1 of 90th percentile values and detected as polyadenylated in our study. This stringent filtering resulted in final list of 16 DISTs with translatable ORFs and specific expression only in MM (Fig. 6E). These lncRNAs encode potential TSAs in MM with *DIS3* gene mutations.

In all, these results confirm that DIST-ORF-derived peptides can be detected by immunopeptidomic approaches and, being specifically expressed in cancer, can show the potential as resource for future MM therapies.

Discussion

In this work, we focused on the regulation of lncRNA degradation by two major exonucleases which act at each extremity of the RNA template. By cell fractionation and the application of auxin-degron system for timely depletion of enzyme of interest, we detected different RNA biotypes that are affected by each pathway.

XRN1 is targeting a limited number of cytoplasmic lncRNAs in human in contrast to yeast.

Our results indicate that approximately 77% of human transcripts targeted by XRN1 are mRNA, leaving the possibility that the lncRNAs degradation in mammals, in contrast to yeast, would be targeted by other RNA decay pathway(s). This observation aligns with the CRAC data for mouse XRN1 which suggests its predominant association with mRNAs[84]. One hypothesis for such differences with yeast Xrn1 is that human RNA degradation might be more specialized with multiple associated cofactors. Certainly, yeast transcriptomes are smaller, with less intronic sequences and not all have RNAi, for instance. In line with this hypothesis, the *S. cerevisiae XRN1* deletion strain accumulates around 1600, mostly antisense, lncRNAs-XUTs[25–27], while the budding yeast- *Naumovozyma castellii,* expressing the cytoplasmic RNAi pathway (Ago and Dicer), shows much less lncRNA-XUTs and more nuclear exosome-sensitive lncRNAs[26], suggesting that in presence of an active RNAi system, XRN1 activity would favor mRNA as a primary substrate. One can speculate that the complexity of the human transcriptome together with the diversity of decay pathways could have largely constrained lncRNA decay to the 3'end RNA degradation pathways.

DIS3 is key factor for preventing lncRNAs to accumulating into the cytoplasm.

The presence of over 5000 DISTs in the cytoplasm may be surprising, as the DIS3 exonuclease is shown to be mainly nuclear (Fig. S6 and[14]) and a large number of well-known DIS3 degraded transcripts, such as eRNAs, conRNAs and uaRNAs/PROMPTs are assumed to be restricted to the nucleus[13]. Pioneer evidence already pointed that the targeted RNAs might not be limited to nuclear environment. Indeed, Ogami and colleagues have shown, that knock-down of the MTR4 RNA helicase, an exosome component, resulted in the accumulation of several RNAs in the cytoplasm[85], similar to what we observed at the genome-wide level upon DIS3 depletion. DISTs are mostly monoexonic or poorly spliced (Fig. 1C and 4A), suggesting a limited efficiency in nuclear export. We propose that the DIS3 depletion affecting many nuclear lncRNAs[13] can have an impact on the RNA flux between the nucleus and the cytoplasm. In fact, around 80% of transcripts accumulated in CYTO-seq are also detected by nuclear-seq[47]. We could hypothesize that in absence of an efficient nuclear RNA decay, such massive nuclear RNA accumulation would promote shuttling of every polyadenylated RNAs associated with PABPN1 (Poly-A Binding Protein), including the DISTs. Alternatively, it could also push the transfer of RNAs associated with ALYREF export factor. Interestingly, ALYREF has been shown to bind polyA$^+$ exosome targets[86] and to associate with Cap Binding Complex (CBC)[87]. The latter can bind also capped PROMPTs[88], another target of exosome complex. In brief, increasing the quantity of polyA+ exosome targets in the nucleus would induce the ALYREF-dependent export and would result in large amount of exosome-dependent lncRNAs in the cytoplasm. In this respect, the presence of polyA$^-$ DISTs in the cytoplasm could be the consequence of either a rapid deadenylation in the cytoplasm[61], or less selective export of polyA$^-$ RNA due to saturation of quality control machinery in the nucleus.

Is the decay of DIS3-sensitive lncRNAs controlled or triggered by translation?

It was suggested that yeast cytoplasmic lncRNAs are degraded in a translation-dependent manner[29]. Thus in theory this mechanism could be conserved in mammals. The work on the human helicase MTR4, associated with the nuclear exosome, has shown that both prematurely terminated transcripts and uaRNAs were detected in the polysome fraction[85] upon MTR4 knock-down. Similar to our results, the uaRNAs were suggested to contain putative ORFs. This raises the question of whether DIS3 and exosome-dependent degradation could also be linked to translation control. In this scenario, one could predict that blocking translation would stabilize the translated lncRNAs. However, this hypothesis remains unlikely since DISTs are not sensitive to XRN1 depletion that acts in concert with the exosome for the classical translation-dependent NMD targets[89]. Nevertheless, a recently described translation-mediated RNA decay pathway was proposed by Ibrahim et al., showing that mRNA are cleaved while bound by ribosomes, in a process called ribothrypsis[90]. Similarly, a putative endonucleolytic cut in the translated lncRNAs could produce an accessible end for further exosomal digestion by DIS3. However, this could be a mechanism explaining degradation of only14.5% of DISTs that associate with ribosomes, while for the remaining DISTs their potential degradation co-factors remain to be identified.

DISTs are templates for the immune peptidome

Among the potentially translated DISTs, among 3 tested cell lines we detected 18 peptides with high confidence (Fig. S5F and Table S8) specifically derived from DIST-ORFs in an immunopeptidomic approach. This proves not only the translation of DIST-derived peptides but also their capacity to be presented by MHCI. However, none of the peptides detected by immunopeptidomic method are among the TSAs prioritized by DIST-ORFs region differential expression (Fig 5B and 5C). The reasons could be that the putative peptides would not be good substrates for the MHC class1 presentation, and/or that the mass spectrometry is not sensitive enough to capture these specific peptides. Mass spectrometry data analysis is based on the prediction of the peptides from the spectra information and can be highly dependent on the type of equipment, sample preparation and dosage. Additionally, the same amino acid sequence could sometimes be the result of translation of different nucleotide sequences, thus it could be harder to unambiguously attribute such peptides to a given RNA/ORF. Therefore, it is becoming increasingly accepted that Ribo-seq and mass spectrometry should be considered as complementary for identifying potentially important antigens for further immunogenicity validation.

DIST expression is specifically observed in independent MM cohort

Currently, some chimeric antigen receptor (CAR)-T cell therapies and vaccines using peptides are being tested in clinical trials and being used in MM treatments, especially for patients with relapse or refractory cancer[91]. However, there is still need for further development of combinatorial antigen approaches for more efficient and specific therapies. Both CAR-T and peptide vaccines currently in clinical trials are based on target peptides derived from proteins. In this work, we focused on identification of lncRNAs that can be specifically expressed and translated in MM cells carrying *DIS3* mutations, which are associated with worse progression-free survival (PFS)[92]. The study from Antonino Neris' group[92] identified more than 2000 lncRNAs as upregulated among MM cases with *DIS3* mutations from the CoMMpas cohort of 640 bone marrow samples[93–96]. We identified 252 lncRNAs in common with our work. Moreover, 23 of the 35 lncRNAs reported to predict poorer PFS[92] were also expressed in our data, suggesting that our analysis is highly relevant to *DIS3* mutations in MM. Interestingly, among those 23 shared lncRNAs, 4 contain translatable DIST-ORFs among the predicted strong MHCI binders. So far, none of those 4 were detected in the immunopeptidomic analysis performed in the DIS3-AID cell line, indicating that their TSA potential needs to be further explored.

DIST-ORFs confers a selective advantage for tumour development?

An important question concerns the cellular function(s) of the (poly)peptides originating from sORFs embedded in unstable lncRNAs. These peptides might be selectively expressed in cancer cells carrying mutations in the RNA decay machinery to confer advantages to the tumour growth. We hypothesize that DIST-derived peptides could be found not only in multiple myeloma (MM) but also in other cancer types with DIS3 mutations such as Acute Myeloid Leukaemia[97] or DIS3 mis-regulation as Superficial spreading melanoma[98]. Thus, studying the cellular functions and immunogenicity of DIST-derived peptides will be

crucial for gaining valuable insights into the treatment of patients, particularly those with recurrent diseases.

To conclude, the main objective of this study was to identify specific enzymes regulating levels of lncRNAs in cytoplasm and we propose that DIS3 play a key role in preventing lncRNA accumulation in the cytoplasm. While we have identified a significant regulator of lncRNA levels, the exact cofactors involved in the surveillance process remain to be discovered.

Declarations

Acknowledgments

We would like to thank to Munich Leukemia Laboratory especially (MLL) especially Claudia Haferlach and Wencke Walter for access to MM patients data, imaging platform at IGH, ICGex – NGS platform Institut Curie especially Sonia Lameiras and Virginie Raynal, The Cell and Tissue Imaging Platform – PICT-IbiSA (member of France–Bioimaging – ANR-10-INBS-04) of the [UMR144 or U934/UMR3215 or UMR3664 or UMR168 or Translational Department or NIMCE] of Institut Curie for help with light microscopy, especially Olivier Leroy. Rocco Cipolla, Ugo Szachnowski, Matthieu Lejars, Setareh Aflaki, Perrine Verdys, Yago Arribas De Sandoval and Arthur Imbert for technical assistance and discussions. Michael Schertzer for critical reading of the manuscript. Nolwenn Jouvenet for support and discussions. A.M. and M.P., were supported by the European Research Council (ERC-consolidator DARK-616180-ERC-2014). This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 795109. This funding was attributed to DF. SW is funded by a Wellcome Trust Investigator Award (223106/Z/21/Z). This project was supported by a VLIR interuniversitaire onderzoeksprojecten (MIMICRY – Modulating Immunity and the Microbiome for effective CRC). Marina Serna and Oscar Llorca acknowledge the support of the National Institute of Health Carlos III to CNIO.

Authors contribution

D.F. and A. M. designed the project. D.F and M.P. performed experiments, M.G. performed most of bioinformatic data analysis, I.H. performed polysome profiling and RPF purification, J.J optimized the cell fractionation protocol, E.P. prepared MS samples for immunopeptidomic analysis, K. B. analyzed the immunopeptidomic data, R.T. prepared scripts for smiFISH image analysis, V.S. participated in smiFISH samples preparation, M.S.G and O.L. prepared structural analysis of DIS3 protein, G.B. supervised immunopeptidomic experiments, E.B. provided resources for smiFISH, S.W. provided the initial cell lines used in the project, O.N. provided resources for olysome profiling, A.M. managed the project. D.F. and A.M. wrote the manuscript with input from all authors.

Conflict of interest

The authors declare no competing interests.

Methods

Cell lines and cell culture

HCT116 cells were cultured in McCoy's 5A (*Gibco, 16600-082*) medium supplemented with 10% fetal bovine serum (FBS) (*Eurobio, CVFSVF0001*) at 37°C in a humidified 5% CO_2 atmosphere. HCT116:TIR1 cell line constitutively expressing TIR1 was described in[29] and DIS3-AID cell line was described in[55]. For construction of XRN1-AID cell line a HCT116 tet-on Tir1 cell line was created using protocol and plasmid from (Natsume et al. 2016), following protocol described in (Eaton et al. 2018). Briefly, 250.000 HCT116 cells were seeded/ well in 6 well plate a day before transfection. 1ug of all in one plasmid containing Cas9 and sgRNA targeting stop codon of XRN1 plasmid and Homology Directed Repair template containing three tandem FLAG and miniAID epitopes and neomycin selection marker was used for transfection with Jetprime (Polyplus), following the manufacturers' guidelines. The selection marker was separated from the tag by P2A site for cleaving during translation (Kim et al. 2011). Media was changed after 24 hours and, after 48 hours, cells were re-plated into 100mm dishes in media containing 700ug/ml Neomycin. Resistant colonies were picked after 7 days. Correct genomic insertion of tags was assayed by PCR and by Western blot. For DIS3- AID set of experiments, auxin was used at a concentration of 500uM for 1 hour unless stated otherwise, for XRN1-AID due to need of inducible TIR1 2 µg/ml doxycycline (dox) and 500 µM auxin was added for 24 h. The KMM-1 and KMS27 (JCRB) cell lines were grown in RPMI 1640 media (Gibco, 21875-034) supplemented with 10% FBS the suspension part of the culture was passaged by dilution while the adherent cells of KMM-1 were trypsinized and combined with suspension cells in each passage.

CRISPR clone construction

gRNA was cloned into pX459 plasmid (Addgene 62988) using two rounds of PCR amplification of gRNA insert with flanking region containing AflII and XbaI cloning sites (primers listed in Table S9). Next, both PCR product and plasmid were digested and ligated using T4 ligase (NEB) overnight. Proper insert was assayed by sanger sequencing. For homology repair template a plasmid with XRN1 homology arms (500 nt flanking stop codon form both side, synthesized by GeneScript) and 3xFLAG 3xminiAID Neomycin containing vector constructed in (Eaton et al. 2018) was used for amplification of respectful fragments using primers listed in Table S6 for cloning with NEBuilder® HiFi DNA Assembly kit. Obtained vector was assayed by Sanger sequencing and linearized before introducing into cells.

Cell Fractionation and RNA seq libraries preparation

After thawing, cells were passaged at least 2 times before seeding 2.5x 10^6 cells into T150 flasks and grown to 80-90% of confluency. For XRN1-AID, the auxin and doxycycline were added 24h before reaching this confluency while for DIS3-AID auxin was added for 1h the last day. For each condition cells were grown in 4 flasks from which 1 was used for protein extraction and fractionation control and 3 others for the replicates of RNA. Cells were washed with cold PBS and scrapped in PBS on ice. After gentle centrifugation cell pellet was resuspended in buffer containing a-amanitin, RNase (SUPERasin) and

proteinase (complete) inhibitors and low level of detergent to lyse cell membrane and keep the nuclei intact, followed by centrifugation collection of cytoplasmic fraction and lysis of nuclei and separation of chromatin. Described before in Gagnon et al. 2014[103] (with change to Tris pH 7.4 for HLB and NLB buffer). The quality of separation of the fractions was tested on western blot and by qPCR (Fig. S1 C-D). 400 ng of cytoplasmic RNA was used for preparation of libraries using Truseq Stranded Total RNA kit (Illumina) after depletion of ribosomal RNAs or Stranded mRNA Prep (Illumina). Libraries were sequenced on NovaSeq 6000 System. Between 54-112 millions of uniquely mapped reads to human genome version 38 were used to further analyse the changes in cytoplasmic RNA levels (Table S5).

After 1h or 24h depletion triplicates of cell cultures for each condition were subjected to fractionation and RNA extractions. Following quality control of the purity of fractions at the RNA and protein levels, total RNAs of the cytoplasmic fractions from the control cell lines and DIS3-AID and XRN1-AID, with or without auxin, (and doxocycline for XRN1-AID) were sequenced (CYTO-seq).

RNA extraction and qPCR

RNA was extracted using the miRNeasy Mini Kit (Quiagen, 217004) following the manufacturers instruction. An additional DNase (Qiagen) treatment was performed following the manufacturer's instructions for samples send for sequencing. For testing the cell fractionation same volume of RNA was taken of reverse transcription reaction. For other experiments 500ng of RNA was used for first-strand cDNA synthesis using SuperScript II RT (Invitrogen) and following the manufacturer's instructions with 100ng of random primers (Invitrogen) used for total cDNA synthesis with primers listed in (Supplementary Table S6). Real time qPCR was performed using SyberGreen (Roche) and Light Cycler 480.

Protein extraction

Beside cell fractionation experiment, protein extracts were prepared from cells grown in p100 plates, washed once with ice cold PBS, and lysed in 500µl RIPA lysis buffer (ThermoScientific, 89900) supplemented with 10µl/ml of a protease and phosphatase inhibitor cocktail (ThermoScientific). The lysed cells were scrapped and collected into Eppendorf tube followed by 8min sonication in 30 sec ON / 30 sec OFF cycles (Diagenode) and centrifuged 10 min at 13 000 rpm at 4°C. Protein concentration was analyzed using the Pierce BCA Protein assay kit (ThermoScientific). Antibodies used for Western blot are listed in Table S7.

Ribo-seq

For each of the cell lines, two vials were thawed and followed to obtain two independent biological replicates. Cells were passaged at least two times before seeding 1x10^6 cells into p100 dishes. Two days later, on the day of experiment, media were aspirated from the plates and fresh media with or without auxin were added for 1h. Afterwards, media were aspirated, and dishes were flash frozen in liquid nitrogen and stored in -80℃ for further use. Cells were gently thawed on ice and 250µl of lysis buffer (10mM Tris pH7.5; 10mM $(CH_3COO)_2Mg$; 100mM KCl; 1% Triton; 2mM DTT) containing complete protease inhibitor cocktail (Roche) and 2U of RNase Murine Inhibitor (NEB) was added to each dish. Cells were collected with scrapper and then centrifuged for 3 minutes at maximum speed at 4℃ an aliquot of each lysate was collected for RNA extraction and the rest was flash frozen and stored at -80℃. Cell lysates were digested with RNase I (Ambion) (1U/ per 15 unit of absorbance) and Turbo DNase (Invitrogen) (1U/50ug of material) for 1h at 25°C loaded onto a 24% sucrose cushion centrifugated 90 min at 110 krpm on a TLA110 rotor at 4°C. Next monosomes were rinsed 2 times with 500ml of lysis buffer then resuspended with 500ml of lysis buffer and digested using 5U/UA_{260nm} RNase I (Ambion) for 1h at 25℃ followed by addition of 500U of SUPErasin RNase inhibitor (Invitrogen). RNA was extracted by acid phenol at 65°C, chloroform and precipitated by ethanol with 0.3 M sodium acetate pH5.2. Resuspended RNA was loaded on 17% polyacrylamide (19:1) gel with 7M urea and run in 1xTAE buffer for 6h at 100V. RNA fragments corresponding to 28-34nt were retrieved from gel and precipitated in ethanol with 0.3 M sodium acetate pH5.2 in presence of 100mg glycogen. rRNA was depleted using riboPool (siTools), adjusting initial volume to 40µl reaction and adding 1µl of Murine RNase inhibitor (NEB) to each sample before and after hybridization step. RNA was precipitated and concentration was measured by Qubit (Thermo Fisher Scientific). 10 ng of each sample was engaged in library preparation using D-plex Small RNA-seq Kit (Diagenode) following manufacturer's instructions with 10 cycles of PCR amplification in the last step followed by DNA purification with Monarch PCR DNA cleanup kit (NEB) and AMPpure XP beads final cleanup. Library molarity was analyzed using Tape Station and an equimolar pool of libraries was sequenced on a NovaSeq 6000 system (Illumina) with 10% PhiX.

MS IP

For immunopeptidomic experiments, DIS3-AID cell line was amplified in T300 flask to 80-90% confluency and auxin was added for 4h. Next, cells were washed 3 times in ice cold 1xPBS and collected by scrapping, followed by centrifugation at 4°C. The cell pellets were flash frozen in liquid nitrogen and stored at -80°C for further use. Two biological replicates were pooled together giving 600mln of cells that were further used for immunopeptidomic analysis. Cells were resuspended in lysis buffer at 4°C. After homogenization, the lysates were cleared by centrifugation and underwent a preabsorption on empty protein A-agarose columns. After this preclearance step, a pull down of the MHCI complexes were carried out, using clone W6/32 anti-MHCI antibody coupled to the protein A-agarose columns. After multiple washing steps, the MHC complexes are eluted and dissociated in mild acid elution buffer. Immune peptides were separated from other eluted components by reverse phase separation prior to loading on evotips for chromatographic separation by an evosep one HPLC system. An Evosep One LC system was coupled with the Bruker timsTOF Pro mass spectrometer equipped with the Bruker Captive Spray source.

The 30 SPD method was used. The Endurance Column 15 cm x 150 µm ID, 1.9 µm beads (EV1106, Evosep) was connected to a Captive Spray emitter (ZDV) with a diameter 20 µm (1865710, Bruker) (both from Bruker Daltonik GmbH, Bremen). The timsTOF Pro was calibrated according to the manufacturer's guidelines. The source parameters were: Capillary voltage 1500 V, Dry Gas 3.0 l/min and Dry Temp 180 °C. The temperature of the ion transfer capillary was set at 180°C. Column was kept at 40°C. The Parallel Accumulation–Serial Fragmentation DDA method was used to select precursor ions for fragmentation with 1 TIMS-MS scan and 10 PASEF MS/MS scans, as described by Meier et al. (2018). The TIMS-MS survey scan was acquired between 0.70 and 1.45 Vs/cm2 and 100–1,700 m/z with a ramp time of 100 ms. The 10 PASEF scans contained on average 12 MS/MS scans per PASEF scan with a collision energy of 10 eV. Precursors with 1–5 charges were selected with the target value set to 20,000 a.u and intensity threshold to 2,500 a.u. Precursors were dynamically excluded for 0.4 s. The timsTOF Pro was controlled by the OtofControl 6.0 software (Bruker Daltonik GmbH). Ten PASEF scans can contain up to 12 MS/MS scans per PASEF scan. Data were analyzed MS Fragger, integrated in the FragPipe pipeline[80]. The build-in "Nonspecific-HLA" workflow was used. Precursor and fragment mass tolerance was set at 20ppm. Oxidation (M), N-terminal acetylation, pyroglutamine, pyroglutamate and cysteinylation were set as variable modifications. The standard database search validation tools from the Nonspecific-HLA workflow were also used, including MSBooster, which is a rescoring tool that uses deep learning predictions of RT fragment intensities[104]. The FDR of peptide and PSM matches was set 1% and probability above 0.9. Uniprot human reference proteome database was concatenated with the custom lncRNA ORF database. The querry list included all the DIST-ORFs predicted to be actively translated in DIS3-depleted cells in Ribo-seq experiment, combined with the ribotricer prediction of ORF based only on the lncRNA specific sequences for the remaining 638 lncRNAs that were upregulated in cell line and MM data and not detected by Ribo-seq experiment.

smiFISH

For smiFISH experiments, cells were passaged at least 2 times after thawing. Two days before experiment, the cells were seeded on coverslips coated with 0.01% poly-L lysin (Sigma) to reach 40-60% of confluency on the day of experiment. At day of the experiment, cells were incubated with fresh media with or without auxin for 1h followed by 2x wash in 1xPBS and fixation with 4% PFA (Electron Microscopy Sciences) for 20 min. Next, the slides were washed 2x with PBS followed by addition of cold 70% EtOH and stored at 4°C before hybridization with smiFISH probes. The smiFISH protocol was performed as described in[53]. The smiFISH probes were designed using Oligostan[53]. 24 specific probes passing the filters of Oligostan design tool were used for hybridization with fixed cells (listed in Table S6). The images were taken using Zeiss Axioimager Z1/Apotome) equipped with a 63x objective and a CCD camera (Axiocam MRm). Obtained images were processed and analyzed using FISH-quant v2 python packages[105] for cell segmentation and spot detection. For each image set threshold (for ZNF674-AS1) or automatic thresholding (for CTD-2371O3.3) was used to adapt for variable background levels. Two independent replicates of the experiment were performed with DIS3-AID cell line incubated with or without auxin for 1h. For each replicate at least 90 cells were quantified.

Impact of DIS3 mutations on protein structure/activity

Initially, the atomic structures available for the DIS3 protein were analyzed, with the position altered residues, with respect to the defined catalytic active sites[73]. Then a systematic analysis of the structural alterations observed for DIS3 in patients was performed, using a combination of the missense3D software together with the structural analysis of the available atomic structures of the protein[100,106]. The atomic model (PDB 6D6Q)[100] was visualized using USCF Chimera[101]. Only mutations impacting the activity but not the truncation of the protein were retained. The list of mutations with results of prediction are listed in Table S6.

Data analysis

Quality control and alignment of RNA-seq and Ribo-seq data:

The Curie bioinformatics platform performed the quality control of the RNA-seq and Ribo-seq data using MultiQC. The alignment of the reads to the human genome hg38 was done using STAR 2.6.1a, with the following parameters: –outSAMstrandField “intronMotif” –outSAMattributes “All” –outSAMtype “BAM” “SortedByCoordinate” –alignIntronMax 1000000 –outFilterMismatchNmax 999 –seedPerReadNmax 100000 –outFilterMultimapNmax 20. Multi-mapped reads were filtered-out, and for Ribo-seq alignments, only reads of length 25 to 34 nucleotides not falling into rRNA were kept for downstream analysis.

Transcriptomic profiling

Transcriptome assembly was performed on each RNA-seq alignment using Scallop v0.10.5, with the following parameters: –min_transcript_length_base 200 –library_type “first” –min_splice_bundary_hits 5 –min_transcript_coverage 10 –min_single_exon_coverage 20 scheme depicted in Fig. S6A. The obtained GTF annotations of all samples were merged with cuffmerge from cufflinks v2.2.1[107], then genes for which the exons overlap GENCODE v26 annotation were filtered out with BEDTools v2.29[108]. The result was finally combined with GENCODE v26 (https://www.gencodegenes.org/human/release_26.html) and eRNAs annotation retrieved from[50] for downstream analysis.

Enhancer RNAs

For eRNA published annotation in HCT116 of conservative eRNA was used[50]. Beside the eRNA placed outside of annotated genes, we analysed how many of these eRNA fall into regions of the upregulated genes defined in CYTO-seq. In total, 362 GENCODE annotated RNAs overlap with the upregulated eRNAs (with at least 1 nt of overlap), including 298 lncRNAs and 27 PCG, and 342 Scallop annotated lncRNA. For further analysis those RNAs were classified as eRNAs.

Premature RNAs analysis

We performed a separate quantification on exons and intron annotation of the GENCODE annotated RNAs. The genes that would have only differentially expressed first exon and intron[13,47] were selected as

PT. Among the upregulated transcripts in our DIS3-AID dependent list, 380 lncRNAs and 240 mRNAs fall into PT category (Fig. S1D and S1E).

Recursive analysis

For the gene to be retained in the final list, it had to be differentially expressed in the same direction in both control + versus tagged cells + but also between tagged cells – versus tagged cells +, if in the latter the same gene was not differentially expressed between control cells – versus control cells +. For XRN1-AID dataset first comparison: control + versus clone + and clone – versus clone + was performed for each clone independently and intersected with addition of genes differentially expressed for both clones in only comparison of clone – versus clone + but not in control cells – versus control + was retained in the list (Fig S6B), where + indicates addition of auxin.

Refining TSS and TES

For better definition of the start and end of our candidates that do not follow fully the standard annotation, we used a sliding window approach of 200nt.For a given gene, the sliding window is run on the metatranscript of exons only (successively from the start to determine the new start, and from the end, to determine the new end) ; if the first 200nt have non-null read coverage values across Dis3-AID+ samples, the original start or end is kept, otherwise, the window is slid until a minimum coverage of 5 reads inside the window is found, then the most upstream new start or the most downstream new end is reported. If no new start or new end can be reported, we use the same method, but on the gene level (the new start or new end can be in the introns in this case). Refined candidates with a length less than 200nt are re-processed with the sliding window approach at the gene level, but to find at the same time the new start and new end (as the steps above can fail for candidates with short exons and read coverage spanning the introns). After start and end refining, candidates with a length below 200nt are removed.

Transcript bidirectionality analysis

The division of DISTs to upstream antisense (ua) RNAs, convergent (con) RNAs and bidirectional eRNAs was performed as defined in[50]. For eRNA all the pairs where at least one eRNA would be upregulated were taken resulting in 777 bidirectional eRNA.

Metagenes

In order to select the wanted parts (exons and introns), we have created firstly metatranscripts as follows: the exons from all the transcripts of each gene were merged with BEDTools v2.29[108], in order to have non-redundant segments of exons, and these exons were numbered; introns were inferred from these exons. The first exon, first intron, second exon, last intron and last exon were selected for each gene. The alignment files were converted in BigWig files with RPM normalization (reads per million of mapped reads), using UCSC tools (http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/), and for each part and each gene, the read coverage was extracted. The obtained signal was scaled on 100 positions using

the R base approx function (allows to compute interpolations), with the following parameters: method="linear", ties="ordered". The average value was computed at each of the 100 positions, and the result was plotted using the R packages: rtracklayer[109] and ggplot2[110].

Active ORFs detection

Ribotricer 1.3.1[111] was used to detect long and short actively translating ORFs. The first step consisted in the generation of all candidate ORFs from the combined GTF annotation (GENCODE v26, Scallop and eRNAs) and hg38 genome, by using the command “prepare-orfs” with the following parameters: –start_codons “ATG, CTG, TTG, GTG, AGG, ACG, AAG, ATC, ATA, ATT” –min_orf_length 30 –longest. The resulting table was converted into GFF by custom command lines. The second step consisted in the use of the command “detect-orfs” to determine translating ORFs from the candidate ORFs for each Ribo-seq alignment. The GFF across all the samples were concatenated and then the redundancies were filtered out based on the following columns of the GFF: chromosome, start, end, strand, attributes. For downstream analysis, ORFs with at least 9 non-zero coverage leading codons were considered strongly active. As the ORF prediction is only based on the annotation, we further tested if the lncRNA ORFs that would span an exon-exon junction would also have such junction present in RNA-seq data in the DIS3 depleted conditions. To be sure that the ORF is translated from DIS3-sensitive isoforms of transcripts we have performed quantification of RNA-seq data on the ORF annotation followed by recursive analysis as for RNA-seq and only the ORFs that would be differentially expressed in the RNA-seq data for both gene and ORF annotation were retained as lncRNAs predicted to be translated upon DIS3-depletion. We have observed a peak from Ribo-seq on the snoRNA location in the snoRNA host genes, thus, to remove bias that could come from the snoRNA we eliminated the genes that the name would contain SNHG, SNO, SNU or RNU and one eRNA that also showed high peak from snoRNA. For Ribo-seq in total we obtained between 3 and 79 million of reads per sample (Table S8). Due to variable number of reads between the replicates we have selected ORFs that would be assigned as actively translated in at least one replicate and at least detected in the second replicate of each condition. To identify ORFs in lncRNAs specifically associated with ribosomes upon DIS3 depletion, we crossed the list of all predicted ORFs in lncRNAs for all 4 conditions, eliminating any ORFs overlapping snoRNAs and PCGs, to remove any ambiguity. P-site offset and periodicity were assessed based on 5'-end read coverage in RPM normalization, computed from the alignment file of each sample, and stored in BigWig format using BEDTools v2.29 and UCSC tools. The mean 5'-end read coverage for selected ORFs was then displayed on 50 nucleotides around their start or stop codon, using the R packages rtracklayer and ggplot2.

Read counting and Differential expression

Read counting on gene, exon, and ORF levels were performed on the combination of the human gene annotation GENCODE v26, eRNAs, and new transcripts from Scallop, using featureCounts v2.0.0[112] with the parameters -s 2 -O 9 (raw counts for Ribo-seq for each of gene features are presented in Fig. S3B and table S5). Analysis for Fig1 were done using GENCODE annotation for normalization of DESeq2 data for unreferenced genes. For further quantifications, modified annotation including eRNA and Scallop

transcripts was used. The conditions were compared using DESeq2, with the following parameters: betaPrior=FALSE, independentFiltering=F, cooksCutoff=F. Only the features with adjusted p-value <= 0.05, read counts >=20 and abs(log2FoldChange)>= 0.585 were retained as differentially expressed. Heatmaps of expression were obtained using the R package ComplexHeatmap from bioconductor[113].

HLA binding prediction

HLA alleles from MM patients were determined using the tool seq2HLA[76]. ORFs candidates were converted into peptides using the SeqinR R package, and only the ones with a length >= 8 amino acids were kept. The lists of peptides and HLA alleles were given to the tool netMHCpan-4.1 in order to determine the binding status. Only the peptides with an elution ligand rank <= 0.5 were kept for downstream analysis.

Patients Data and GTEx analysis:

The data concerning patients samples were previously published [74]. New counting on the raw reads was performed using kallisto[75]. All further analysis followed the standard DESeq2 protocol. The extended ORF annotation for kallisto indexing for counting of expression in both our MM dataset and GTEx tissue samples. We performed querry on both the Ribo-seq identified DIST-ORFs candidates in addition to DIST-ORFs with detected peptides in immunopeptidomic analysis. We used the 90^{th} percentile values and TPM≤1 to define as threshold for defining the ORFs specifically expressed in MM, excluding testis.

Code availability:

Specific code used for analysis performed in the paper can be found at: https://github.com/MorillonLab/DIS3_analysis

Data availability:

Sequencing datasets generated in this study, including Cyto-seq and Ribo-seq, were deposited into the GEO database under the accession numbers: GSE188282, GSE188195 and GSE233699.

References

1. Jarroux, J., Morillon, A., and Pinskaya, M. (2017). History, Discovery, and Classification of lncRNAs. Adv. Exp. Med. Biol. 1008, 1–46. 10.1007/978-981-10-5203-3_1. 2. Gourvest, M., Brousset, P., and Bousquet, M. (2019). Long Noncoding RNAs in Acute Myeloid Leukemia: Functional Characterization and Clinical Relevance. Cancers 11, 1638. 10.3390/cancers11111638. 3. Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. 16, 20. 10.1186/s13059-015-0586-4. 4. Heesch, S. van, Iterson, M. van, Jacobi, J., Boymans, S., Essers, P.B., Bruijn, E. de, Hao, W., MacInnes, A.W., Cuppen, E., and Simonis, M. (2014). Extensive localization of

long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol. 15, 1–12. 10.1186/gb-2014-15-1-r6. 5. Szcześniak, M.W., Wanowska, E., Mukherjee, N., Ohler, U., and Makałowska, I. (2020). Towards a deeper annotation of human lncRNAs. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1863, 194385. 10.1016/j.bbagrm.2019.05.003. 6. Noh, J.H., Kim, K.M., McClusky, W.G., Abdelmohsen, K., and Gorospe, M. (2018). Cytoplasmic functions of long noncoding RNAs. Wiley Interdiscip. Rev. RNA 9, e1471. 10.1002/wrna.1471. 7. Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res. 22, 885–898. 10.1101/gr.131037.111. 8. Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. 24, 86–96. 10.1038/nsmb.3325. 9. Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. 22, 947–956. 10.1101/gr.130559.111. 10. Maekawa, S., Imamachi, N., Irie, T., Tani, H., Matsumoto, K., Mizutani, R., Imamura, K., Kakeda, M., Yada, T., Sugano, S., et al. (2015). Analysis of RNA decay factor mediated RNA stability contributions on RNA abundance. BMC Genomics 16, 154. 10.1186/s12864-015-1358-y. 11. Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol. Cell 65, 25–38. 10.1016/j.molcel.2016.11.029. 12. 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. 10.1038/emboj.2013.135. 13. Szczepińska, T., Kalisiak, K., Tomecki, R., Labno, A., Borowski, L.S., Kulinski, T.M., Adamska, D., Kosinska, J., and Dziembowski, A. (2015). DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. Genome Res. 25, 1622–1633. 10.1101/gr.189597.115. 14. 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., et al. (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J. 29, 2342–2357. 10.1038/emboj.2010.121. 15. Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochim. Biophys. Acta 1829, 552–560. 10.1016/j.bbagrm.2013.01.006. 16. Łabno, A., Warkocki, Z., Kuliński, T., Krawczyk, P.S., Bijata, K., Tomecki, R., and Dziembowski, A. (2016). Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. Nucleic Acids Res. 44, 10437–10453. 10.1093/nar/gkw649. 17. Ustianenko, D., Pasulka, J., Feketova, Z., Bednarik, L., Zigackova, D., Fortova, A., Zavolan, M., and Vanacova, S. (2016). TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. EMBO J. 35, 2179–2191. 10.15252/embj.201694857. 18. Nagarajan, V.K., Jones, C.I., Newbury, S.F., and Green, P.J. XRN 5' → 3' exoribonucleases: Structure, mechanisms and functions. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 10.1016/j.bbagrm.2013.03.005. 19. Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. 32, 127–139. 10.1101/gad.308528.117. 20. Łabno, A., Tomecki, R., and Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - Enzymes and mechanisms. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1863, 3125–

3147. 10.1016/j.bbamcr.2016.09.023. 21. Monaghan, L., Longman, D., and Cáceres, J.F. (2023). Translation-coupled mRNA quality control mechanisms. EMBO J. 42, e114378. 10.15252/embj.2023114378. 22. Pashler, A.L., Towler, B.P., Jones, C.I., Haime, H.J., Burgess, T., and Newbury, S.F. (2021). Genome-wide analyses of XRN1-sensitive targets in osteosarcoma cells identify disease-relevant transcripts containing G-rich motifs. RNA N. Y. N 27, 1265–1280. 10.1261/rna.078872.121. 23. Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev. 28, 2498–2517. 10.1101/gad.246538.114. 24. Slomovic, S., Fremder, E., Staals, R.H.G., Pruijn, G.J.M., and Schuster, G. (2010). Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. Proc. Natl. Acad. Sci. 107, 7407–7412. 10.1073/pnas.0910621107. 25. van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475, 114–117. 10.1038/nature10118. 26. Szachnowski, U., Andus, S., Foretek, D., Morillon, A., and Wery, M. (2019). Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. Life Sci. Alliance 2, e201900407. 10.26508/lsa.201900407. 27. Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.-S., Gautheret, D., and Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Mol. Cell 61, 379–392. 10.1016/j.molcel.2015.12.020. 28. Andjus, S., Szachnowski, U., Vogt, N., Hatin, I., Papadopoulos, C., Lopes, A., Namy, O., Wery, M., and Morillon, A. (2022). Translation is a key determinant controlling the fate of cytoplasmic long non-coding RNAs. Preprint at bioRxiv, 10.1101/2022.05.25.493276 10.1101/2022.05.25.493276. 29. Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L.A., and Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA N. Y. N 22, 867–882. 10.1261/rna.053561.115. 30. Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Rep. 8, 1365–1379. 10.1016/j.celrep.2014.07.045. 31. Zeng, C., Fukunaga, T., and Hamada, M. (2018). Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. BMC Genomics 19. 10.1186/s12864-018-4765-z. 32. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. Science 367, 1140–1146. 10.1126/science.aay0262. 33. Othoum, G., Coonrod, E., Zhao, S., Dang, H.X., and Maher, C.A. (2020). Pan-cancer proteogenomic analysis reveals long and circular noncoding RNAs encoding peptides. NAR Cancer 2, zcaa015. 10.1093/narcan/zcaa015. 34. Wu, P., Mo, Y., Peng, M., Tang, T., Zhong, Y., Deng, X., Xiong, F., Guo, C., Wu, X., Li, Y., et al. (2020). Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA. Mol. Cancer 19, 22. 10.1186/s12943-020-1147-3. 35. Heesch, S. van, Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., et al. (2019). The Translational Landscape of the Human Heart. Cell 178, 242-260.e29. 10.1016/j.cell.2019.05.010. 36. Wright, B.W., Yi, Z., Weissman, J.S., and Chen, J. (2022). The dark proteome: translation from noncanonical open reading frames. Trends Cell Biol. 32, 243–258. 10.1016/j.tcb.2021.10.010. 37. Carlevaro-Fita, J., Lanzós, A., Feuerbach, L., Hong, C., Mas-Ponte, D.,

Pedersen, J.S., and Johnson, R. (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. Commun. Biol. 3, 1–16. 10.1038/s42003-019-0741-7. 38. Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat. Chem. Biol. 9, 59–64. 10.1038/nchembio.1120. 39. Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., et al. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat. Commun. 11, 1293. 10.1038/s41467-020-14968-9. 40. Ehx, G., Larouche, J.-D., Durette, C., Laverdure, J.-P., Hesnard, L., Vincent, K., Hardy, M.-P., Thériault, C., Rulleau, C., Lanoix, J., et al. (2021). Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. Immunity 54, 737-752.e10. 10.1016/j.immuni.2021.03.001. 41. Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., et al. (2022). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat. Biotechnol. 40, 209–217. 10.1038/s41587-021-01021-3. 42. Barczak, W., Carr, S.M., Liu, G., Munro, S., Nicastri, A., Lee, L.N., Hutchings, C., Ternette, N., Klenerman, P., Kanapin, A., et al. (2023). Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. Nat. Commun. 14, 1078. 10.1038/s41467-023-36826-0. 43. Apcher, S., Daskalogianni, C., Lejeune, F., Manoury, B., Imhoos, G., Heslop, L., and Fåhraeus, R. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc. Natl. Acad. Sci. U. S. A. 108, 11572–11577. 10.1073/pnas.1104104108. 44. Natsume, T., Kiyomitsu, T., Saga, Y., and Kanemaki, M.T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Rep. 15, 210–218. 10.1016/j.celrep.2016.03.001. 45. Chang, C.-T., Muthukumar, S., Weber, R., Levdansky, Y., Chen, Y., Bhandari, D., Igreja, C., Wohlbold, L., Valkov, E., and Izaurralde, E. (2019). A low-complexity region in human XRN1 directly recruits deadenylation and decapping factors in 5’-3’ messenger RNA decay. Nucleic Acids Res. 47, 9282–9295. 10.1093/nar/gkz633. 46. Kalisiak, K., Kuliński, T.M., Tomecki, R., Cysewski, D., Pietras, Z., Chlebowski, A., Kowalska, K., and Dziembowski, A. (2017). A short splicing isoform of HBS1L links the cytoplasmic exosome and SKI complexes in humans. Nucleic Acids Res. 45, 2068–2080. 10.1093/nar/gkw862. 47. Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. Cell Rep. 26, 2779-2791.e5. 10.1016/j.celrep.2019.02.012. 48. Shao, M., and Kingsford, C. (2017). Accurate assembly of transcripts through phase-preserving graph decomposition. Nat. Biotechnol. 35, 1167–1169. 10.1038/nbt.4020. 49. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461. 10.1038/nature12787. 50. Lidschreiber, K., Jung, L.A., von der Emde, H., Dave, K., Taipale, J., Cramer, P., and Lidschreiber, M. (2021). Transcriptionally active enhancers in human cancer cells. Mol. Syst. Biol. 17, e9873. 10.15252/msb.20209873. 51. Wang, M., Zhao, Y., and Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. Sci. Rep. 5, 16923. 10.1038/srep16923. 52. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA Exosome Depletion Reveals Transcription

Upstream of Active Human Promoters. Science 322, 1851–1854. 10.1126/science.1164096. 53. Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.-M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., et al. (2016). smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. 44, e165. 10.1093/nar/gkw784. 54. Loedige, I., Baranovskii, A., Mendonsa, S., Dantsuji, S., Popitsch, N., Breimann, L., Zerna, N., Cherepanov, V., Milek, M., Ameres, S., et al. (2023). mRNA stability and m6A are major determinants of subcellular mRNA localization in neurons. Mol. Cell 83, 2709-2725.e10. 10.1016/j.molcel.2023.06.021. 55. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147. 10.1016/j.cell.2011.10.002. 56. Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. U. S. A. 109, E2424-2432. 10.1073/pnas.1207846109. 57. Prensner, J.R., Abelin, J.G., Kok, L.W., Clauser, K.R., Mudge, J.M., Ruiz-Orera, J., Bassani-Sternberg, M., Moritz, R.L., Deutsch, E.W., and Heesch, S. van (2023). What Can Ribo-Seq, Immunopeptidomics, and Proteomics Tell Us About the Noncanonical Proteome? Mol. Cell. Proteomics 22. 10.1016/j.mcpro.2023.100631. 58. Lv, D., Chang, Z., Cai, Y., Li, J., Wang, L., Jiang, Q., Xu, K., Ding, N., Li, X., Xu, J., et al. (2022). TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. Nucleic Acids Res. 50, D413–D420. 10.1093/nar/gkab847. 59. Luo, X., Huang, Y., Li, H., Luo, Y., Zuo, Z., Ren, J., and Xie, Y. (2022). SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. Nucleic Acids Res. 50, D1373–D1381. 10.1093/nar/gkab822. 60. Mudge, J.M., Ruiz-Orera, J., Prensner, J.R., Brunet, M.A., Calvet, F., Jungreis, I., Gonzalez, J.M., Magrane, M., Martinez, T.F., Schulz, J.F., et al. (2022). Standardized annotation of translated open reading frames. Nat. Biotechnol. 40, 994–999. 10.1038/s41587-022-01369-0. 61. Passmore, L.A., and Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nat. Rev. Mol. Cell Biol. 23, 93–106. 10.1038/s41580-021-00417-y. 62. Lorenzi, L., Chiu, H.-S., Avila Cobos, F., Gross, S., Volders, P.-J., Cannoodt, R., Nuytens, J., Vanderheyden, K., Anckaert, J., Lefever, S., et al. (2021). The RNA Atlas expands the catalog of human non-coding RNAs. Nat. Biotechnol. 39, 1453–1465. 10.1038/s41587-021-00936-1. 63. Wu, G., Schmid, M., Rib, L., Polak, P., Meola, N., Sandelin, A., and Jensen, T.H. (2020). A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome. Cell Rep. 30, 2387-2401.e5. 10.1016/j.celrep.2020.01.068. 64. Sinha, T., Panigrahi, C., Das, D., and Chandra Panda, A. (2022). Circular RNA translation, a path to hidden proteome. WIREs RNA 13, e1685. 10.1002/wrna.1685. 65. Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 9, 843–854. 10.1038/nrg2438. 66. AACR Project GENIE Consortium (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov. 7, 818–831. 10.1158/2159-8290.CD-17-0151. 67. Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., et al. (2018). Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. Blood 132, 587–597. 10.1182/blood-2018-03-840132. 68. Robinson, S.R., Viegas, S.C., Matos, R.G., Domingues, S., Bedir, M., Stewart, H.J.S., Chevassut, T.J., Oliver, A.W., Arraiano, C.M., and Newbury, S.F. (2018). DIS3 isoforms vary in their endoribonuclease activity and are differentially expressed within haematological cancers. Biochem. J. 475, 2091–2105.

10.1042/BCJ20170962. 69. Pertesi, M., Vallée, M., Wei, X., Revuelta, M.V., Galia, P., Demangel, D., Oliver, J., Foll, M., Chen, S., Perrial, E., et al. (2019). Exome sequencing identifies germline variants in DIS3 in familial multiple myeloma. Leukemia 33, 2324–2330. 10.1038/s41375-019-0452-6. 70. Boyle, E.M., Ashby, C., Tytarenko, R.G., Deshpande, S., Wang, H., Wang, Y., Rosenthal, A., Sawyer, J., Tian, E., Flynt, E., et al. (2020). BRAF and DIS3 Mutations Associate with Adverse Outcome in a Long-term Follow-up of Patients with Multiple Myeloma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 26, 2422–2432. 10.1158/1078-0432.CCR-19-1507. 71. White, B.S., Lanc, I., O'Neal, J., Gupta, H., Fulton, R.S., Schmidt, H., Fronick, C., Belter, E.A., Fiala, M., King, J., et al. (2018). A multiple myeloma-specific capture sequencing platform discovers novel translocations and frequent, risk-associated point mutations in IGLL5. Blood Cancer J. 8, 35. 10.1038/s41408-018-0062-y. 72. Weißbach, S., Langer, C., Puppe, B., Nedeva, T., Bach, E., Kull, M., Bargou, R., Einsele, H., Rosenwald, A., Knop, S., et al. (2015). The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. Br. J. Haematol. 169, 57–70. 10.1111/bjh.13256. 73. Tomecki, R., Drazkowska, K., Kucinski, I., Stodus, K., Szczesny, R.J., Gruchota, J., Owczarek, E.P., Kalisiak, K., and 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, 1270–1290. 10.1093/nar/gkt930. 74. Höllein, A., Twardziok, S.O., Walter, W., Hutter, S., Baer, C., Hernandez-Sanchez, J.M., Meggendorfer, M., Haferlach, T., Kern, W., and Haferlach, C. (2020). The combination of WGS and RNA-Seq is superior to conventional diagnostic tests in multiple myeloma: Ready for prime time? Cancer Genet. 242, 15–24. 10.1016/j.cancergen.2020.01.001. 75. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. 10.1038/nbt.3519. 76. Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boisguérin, V., Türeci, Ö., Diken, M., Castle, J.C., and Sahin, U. (2012). HLA typing from RNA-Seq sequence reads. Genome Med. 4, 102. 10.1186/gm403. 77. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 48, W449–W454. 10.1093/nar/gkaa379. 78. Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., and Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61, 1–13. 10.1007/s00251-008-0341-z. 79. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. Science 367, 1140–1146. 10.1126/science.aay0262. 80. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat. Methods 14, 513–520. 10.1038/nmeth.4256. 81. Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L., and Degroeve, S. (2021). DeepLC can predict retention times for peptides that carry as-yet unseen modifications. Nat. Methods 18, 1363–1369. 10.1038/s41592-021-01301-5. 82. Andreatta, M., Alvarez, B., and Nielsen, M. (2017). GibbsCluster: unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res. 45, W458–W463. 10.1093/nar/gkx248. 83. Schmidt, T., Samaras, P., Dorfer, V., Panse, C., Kockmann, T., Bichmann, L., van Puyvelde, B., Perez-Riverol, Y., Deutsch, E.W., Kuster, B., et al. (2021). Universal Spectrum Explorer: A Standalone (Web-)Application for Cross-Resource Spectrum Comparison. J. Proteome Res. 20, 3388–3394.

10.1021/acs.jproteome.1c00096. 84. Tuck, A.C., Rankova, A., Arpat, A.B., Liechti, L.A., Hess, D., Iesmantavicius, V., Castelo-Szekely, V., Gatfield, D., and Bühler, M. (2020). Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. Mol. Cell 77, 1222-1236.e13. 10.1016/j.molcel.2020.01.007. 85. Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. Genes Dev. 31, 1257–1271. 10.1101/gad.302604.117. 86. Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. Cell Rep. 23, 2199–2210. 10.1016/j.celrep.2018.04.061. 87. Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., et al. (2017). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. EMBO J. 36, 2870–2886. 10.15252/embj.201696139. 88. Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., et al. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat. Struct. Mol. Biol. 20, 1367–1376. 10.1038/nsmb.2703. 89. Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol. Cell 12, 675–687. 10.1016/s1097-2765(03)00349-6. 90. Ibrahim, F., Maragkakis, M., Alexiou, P., and Mourelatos, Z. (2018). Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. Nat. Struct. Mol. Biol. 25, 302–310. 10.1038/s41594-018-0042-8. 91. Boussi, L.S., Avigan, Z.M., and Rosenblatt, J. (2022). Immunotherapy for the treatment of multiple myeloma. Front. Immunol. 13. 10.3389/fimmu.2022.1027385. 92. Todoerti, K., Ronchetti, D., Favasuli, V., Maura, F., Morabito, F., Bolli, N., Taiana, E., and Neri, A. (2022). DIS3 mutations in multiple myeloma impact the transcriptional signature and clinical outcome. Haematologica 107, 921–932. 10.3324/haematol.2021.278342. 93. Keats, J.J., Speyer, G., Christophe, L., Austin, C., Stephenson, K., Kurdoglu, A., Russell, M., Jessica, A., Lori, C., Adkins, J., et al. (2015). Identification of Initiating Trunk Mutations and Distinct Molecular Subtypes: An Interim Analysis of the Mmrf Commpass Study. Blood 126, 722. 10.1182/blood.V126.23.722.722. 94. Keats, J.J., Speyer, G., Christofferson, A., Legendre, C., Aldrich, J., Russell, M., Cuyugan, L., Adkins, J., Blanski, A., Hodges, M., et al. (2016). Molecular Predictors of Outcome and Drug Response in Multiple Myeloma: An Interim Analysis of the Mmrf CoMMpass Study. Blood 128, 194. 10.1182/blood.V128.22.194.194. 95. Keats, J.J., Craig, D.W., Liang, W., Venkata, Y., Kurdoglu, A., Aldrich, J., Auclair, D., Allen, K., Harrison, B., Jewell, S., et al. (2013). Interim Analysis Of The Mmrf Commpass Trial, a Longitudinal Study In Multiple Myeloma Relating Clinical Outcomes To Genomic and Immunophenotypic Profiles. Blood 122, 532. 10.1182/blood.V122.21.532.532. 96. Lonial, S., Yellapantula, V.D., Liang, W., Kurdoglu, A., Aldrich, J., Legendre, C.M., Stephenson, K., Adkins, J., McDonald, J., Helland, A., et al. (2014). Interim Analysis of the Mmrf Commpass Trial: Identification of Novel Rearrangements Potentially Associated with Disease Initiation and Progression. Blood 124, 722. 10.1182/blood.V124.21.722.722. 97. 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., et al. (2012). The Origin and Evolution of Mutations in Acute Myeloid Leukemia. Cell 150, 264–278. 10.1016/j.cell.2012.06.023. 98. 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., et al. (2011). Integrative Genomics Identifies Molecular Alterations that Challenge the Linear Model of Melanoma Progression. Cancer Res. 71, 2561–2571. 10.1158/0008-5472.CAN-10-2958. 99. Scientific Image and Illustration Software | BioRender https://www.biorender.com/. 100. Weick, E.-M., Puno, M.R., Januszyk, K., Zinder, J.C., DiMattia, M.A., and Lima, C.D. (2018). Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. Cell 173, 1663-1677.e21. 10.1016/j.cell.2018.05.041. 101. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. 10.1002/jcc.20084. 102. Ittisoponpisan, S., Islam, S.A., Khanna, T., Alhuzimi, E., David, A., and Sternberg, M.J.E. (2019). Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J. Mol. Biol. 431, 2197–2212. 10.1016/j.jmb.2019.04.009. 103. Gagnon, K.T., Li, L., Janowski, B.A., and Corey, D.R. (2014). Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. Nat. Protoc. 9, 2045–2060. 10.1038/nprot.2014.135. 104. Yang, K.L., Yu, F., Teo, G.C., Demichev, V., Ralser, M., and Nesvizhskii, A.I. (2022). MSBooster: Improving Peptide Identification Rates using Deep Learning-Based Features. Preprint at bioRxiv, 10.1101/2022.10.19.512904 10.1101/2022.10.19.512904. 105. Imbert, A., Ouyang, W., Safieddine, A., Coleno, E., Zimmer, C., Bertrand, E., Walter, T., and Mueller, F. (2022). FISH-quant v2: a scalable and modular tool for smFISH image analysis. RNA N. Y. N 28, 786–795. 10.1261/rna.079073.121. 106. Gerlach, P., Schuller, J.M., Bonneau, F., Basquin, J., Reichelt, P., Falk, S., and Conti, E. (2018). Distinct and evolutionary conserved structural features of the human nuclear exosome complex. eLife 7, e38686. 10.7554/eLife.38686. 107. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578. 10.1038/nprot.2012.016. 108. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. 26, 841–842. 10.1093/bioinformatics/btq033. 109. Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841–1842. 10.1093/bioinformatics/btp328. 110. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer) 10.1007/978-0-387-98141-3. 111. Choudhary, S., Li, W., and D Smith, A. (2020). Accurate detection of short and long active ORFs using Ribo-seq data. Bioinforma. Oxf. Engl. 36, 2053–2059. 10.1093/bioinformatics/btz878. 112. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. 30, 923–930. 10.1093/bioinformatics/btt656. 113. Gu, Z. (2022). Complex heatmap visualization. iMeta 1, e43. 10.1002/imt2.43.

Figures

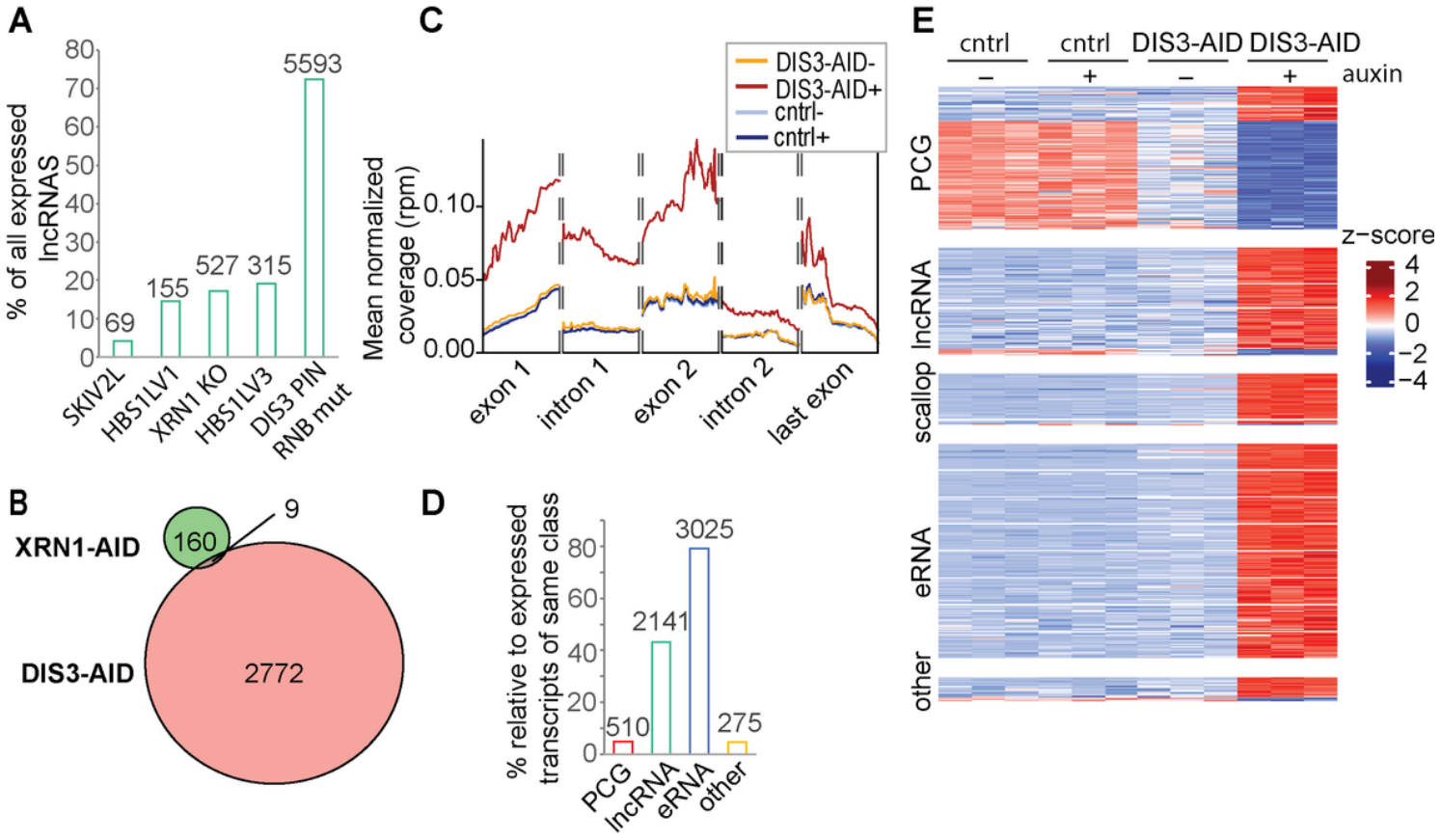

Figure 1

DIS3 prevents accumulation of lncRNAs. A) Bar plot presenting percentage of upregulated against all expressed GENCODE-annotated lncRNA, with a fold change above 1.5 and adjusted p-value <=0.05 in comparison to control conditions based on DESeq2 analysis of published dataset[13,45,46] and expression filter set to first 10% percentile (see also Table S1 and S2), the absolute number of upregulated lncRNAs is presented above the bars. B) Euler diagram of overlap between upregulated lncRNAs from GENCODE and scallop DE analysis for XRN1 and DIS3 depleted cells. C) Metagene profiling of the GENCODE annotated lncRNAs containing at least 2 introns (n=380), upregulated upon DIS3 depletion. D) Bar plot of the percentage of DIS3-sensitive RNAs (Fold change > 1.5, *p*-value < = 0.05) belonging to protein-coding gene (PCG) (n=510), lncRNA (n=2141) and eRNA (n=3025) or other (n=275) subtypes of transcripts in relation to all expressed genes from specific gene class (see also Table S1 and S2). E) Supervised clustering heatmaps of a z-score of differentially expressed genes (n=7644) in DIS3-AID in comparison to a control (cntrl), treated (+) or not (-) with auxin.

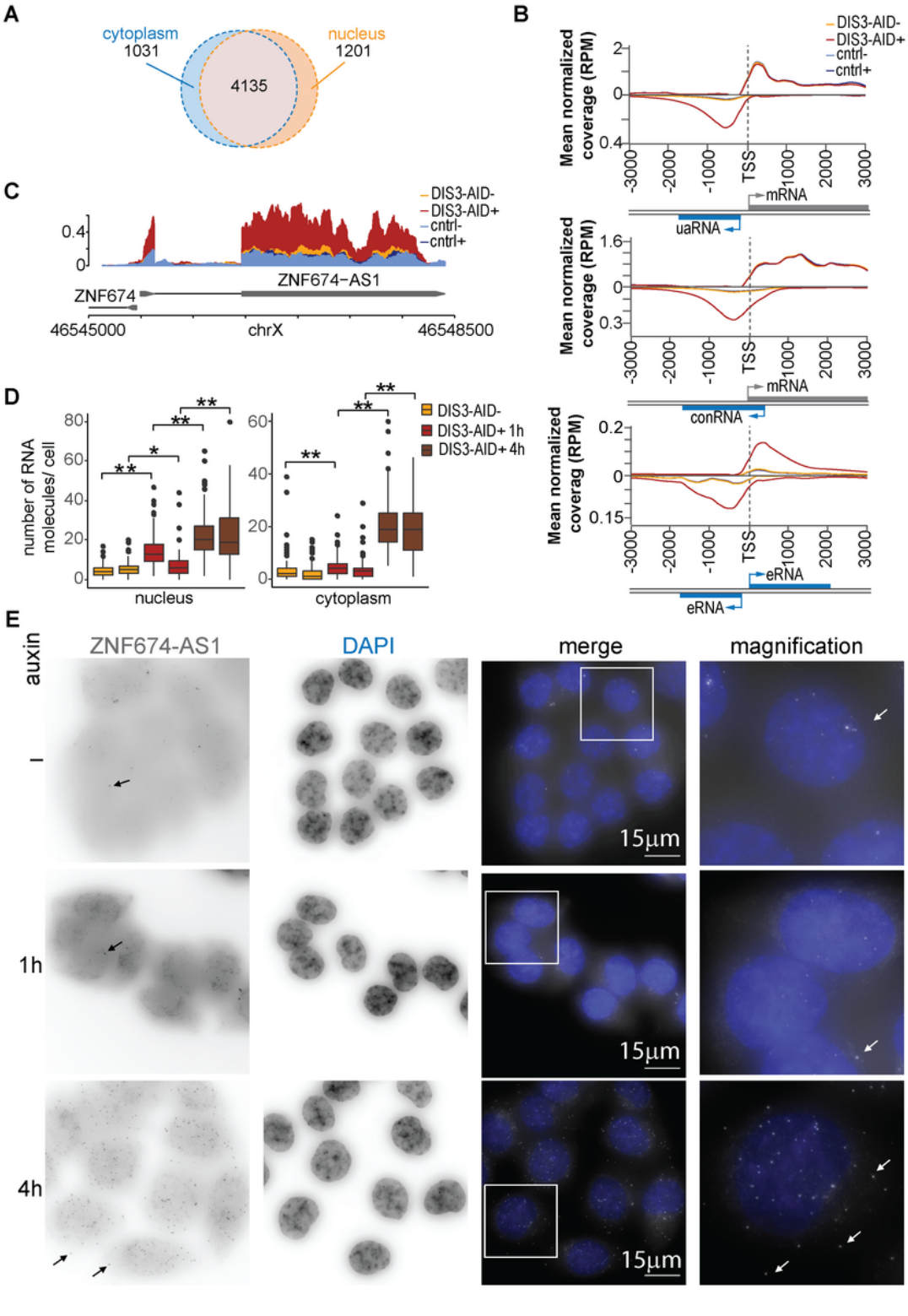

Figure 2

Some of DISTs specific features are shared between cytoplasmic and nuclear DIS3-targets. A) Venn diagram of an overlap between DISTs found in CYTO-seq and nuclei-seq47 (SuperExactTest p-value <0.0001). B) Metagene profiling of bidirectional transcripts: upstream antisense (ua) RNAs (n=667), convergent (con) RNAs (n=497) and bidirectional enhancer © RNAs (n=517) with corresponding sense transcript, identified by CYTO-seq. C) Snapshot of reads coverage of the DIST regions CYTO- seq of DIS3-

AID and control (cntrl) cells treated (+) or not (-) with auxin. D) Box plot of the number of RNA molecules per cell for at least 90 cells for two biological replicates of the smiFISH experiment. *P-value<0.05, **P-value < 0.001 (two-tailed t-test). Median number of molecules for nucleus: (-) 4 and 5, (1h) 13 and 6, (4h) 20 and 19; for cytoplasm (-) 2 and 1, (1h) 4 and 3, (4h) 19 and 19 for each of replicates respectively. E)Representative maximum projection images of smiFISH z-stacks of ZNF674_AS1 lncRNA in DIS3-AID cells treated (+) or not (-) with auxin (white points in merge images) and DNA stained by DAPI (blue in merge image). Arrows indicate examples of the lncRNA molecules present in cytoplasm.

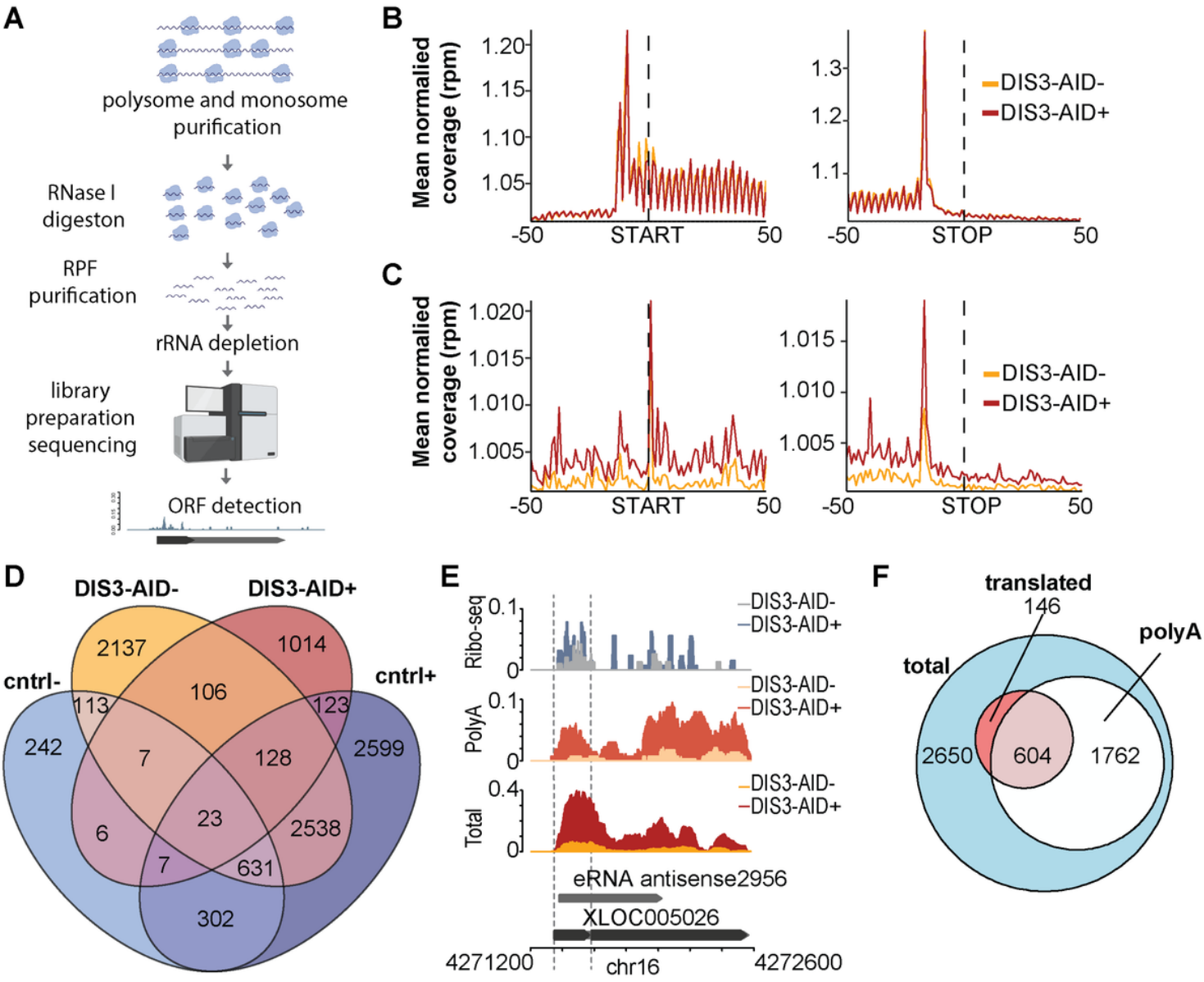

Figure 3

Cytoplasmic DISTs show ribosome binding pattern of active translation and are partially polyadenylated and conserved. A) Experimental workflow of Ribo-seq experiment (prepared with BioRender99. B) and C) Metagene profiling of Ribo-seq reads spanning across the start and stop codons respectively for highest expressed ORF per gene detected in DIS3-AID cells treated (+) or not with auxin (-) for PCG (n=13711) (B)

and DISTs identified ORFs (n=750) (C). D) Venn diagram of actively translated ORFs identified by *ribotricer* in lncRNAs. E) Snapshot of reads coverage of the DIST regions in Ribo-, polyA+ RNA- and total RNA- seq of DIS3-AID cells treated (+) or not (-) with auxin. F) Pie chart presenting a percentage of start codon identified in DIST-ORFs as actively translated in DIS3-AID+ (n=1414).

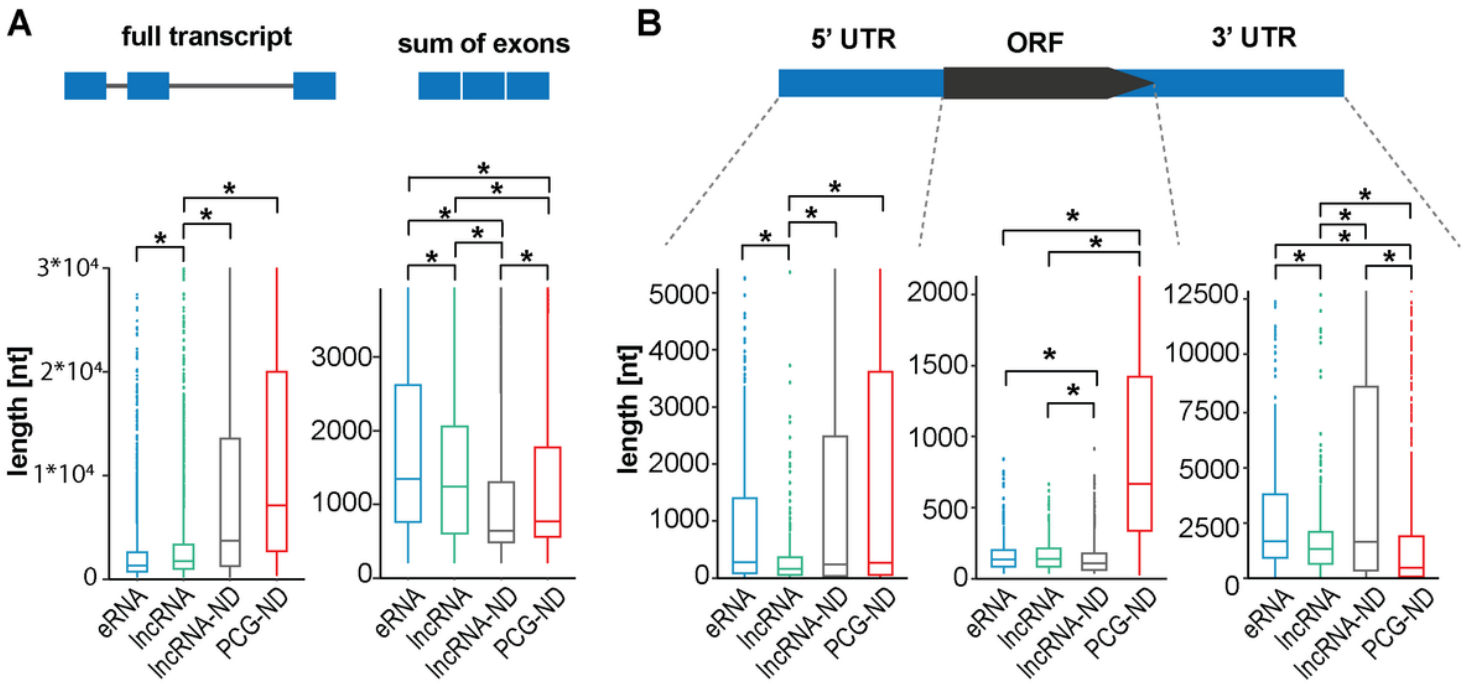

Figure 4

Features of DISTs correlate with majority monoexonic and short ORFs for translated ones. Boxplots presenting the size of full transcripts and sum of exons for upregulated DISTs (n= 3025 eRNA, n=2137 lncRNA) and expressed non-differential lncRNAs (n=7748) and protein coding genes (PCG) (n=2850) (two tailed t-test *p<0.001) (see also Table S2). B) The 3'UTR, ORFs and 5'UTR lengths of the highest differentially expressed eRNAs (n=381) and lncRNAs (n=369) and non-differential (ND) lncRNAs (n=982) and PCGs (n=1599) detected to be bound to ribosomes by Ribo-seq. **P*-value < 0.001 (two-tailed t-test). Median values of differential lncRNA and eRNA 5'UTR 159 and 281 nt respectively and 1317-1671 nt for 3' UTR.

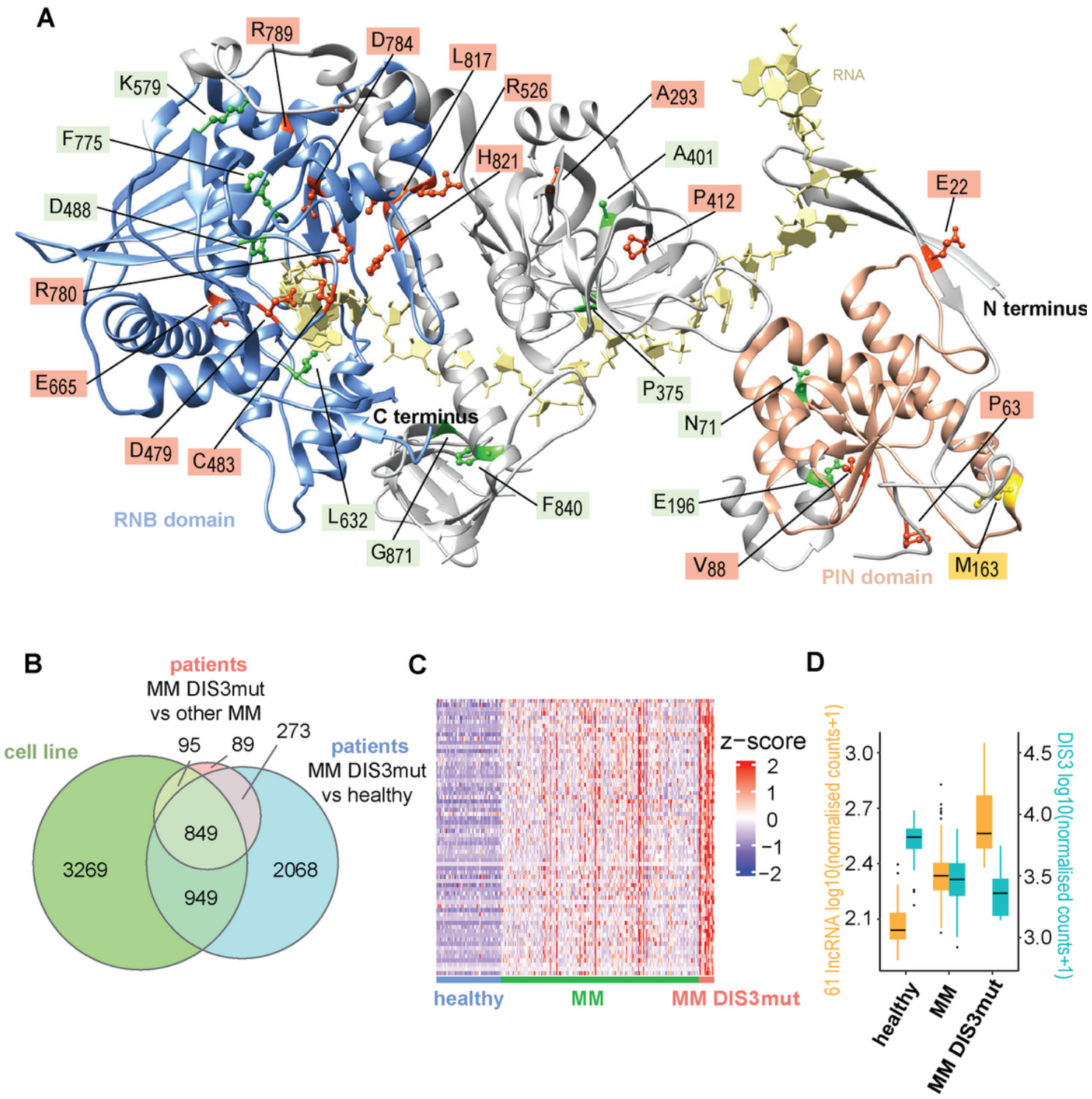

Figure 5

DISTs are detected in MM cases with *DIS3* mutations. A) Atomic model of the DIS3 structure (PDB 6D6Q)[100] displayed using USCF Chimera[101]. Mutations found in patient's specimens are indicated in red for mutations predicted to alter the structure and/or activity based on the visual inspection of the structure and using the missense3D software[102]. Mutations predicted as not affecting the structure/activity are labelled in green. Mutation creating truncated version of the protein is indicated by

yellow (see also Table S4). B) Venn diagram of an overlap between upregulated lncRNAs from the DIS3-AID cell line treated with auxin and MM samples with mutations in the *DIS3* gene comparing to other MM cases or healthy control. C) Supervised clustering heatmap of a z-score representing expression of 66 DIST-ORFs upregulated in patients with *DIS3* mutations. D) Boxplot of mean expression of 61 lncRNAs containing the ORFs from panel C, with anticorrelation of *DIS3* gene expression with (R= -0.88).

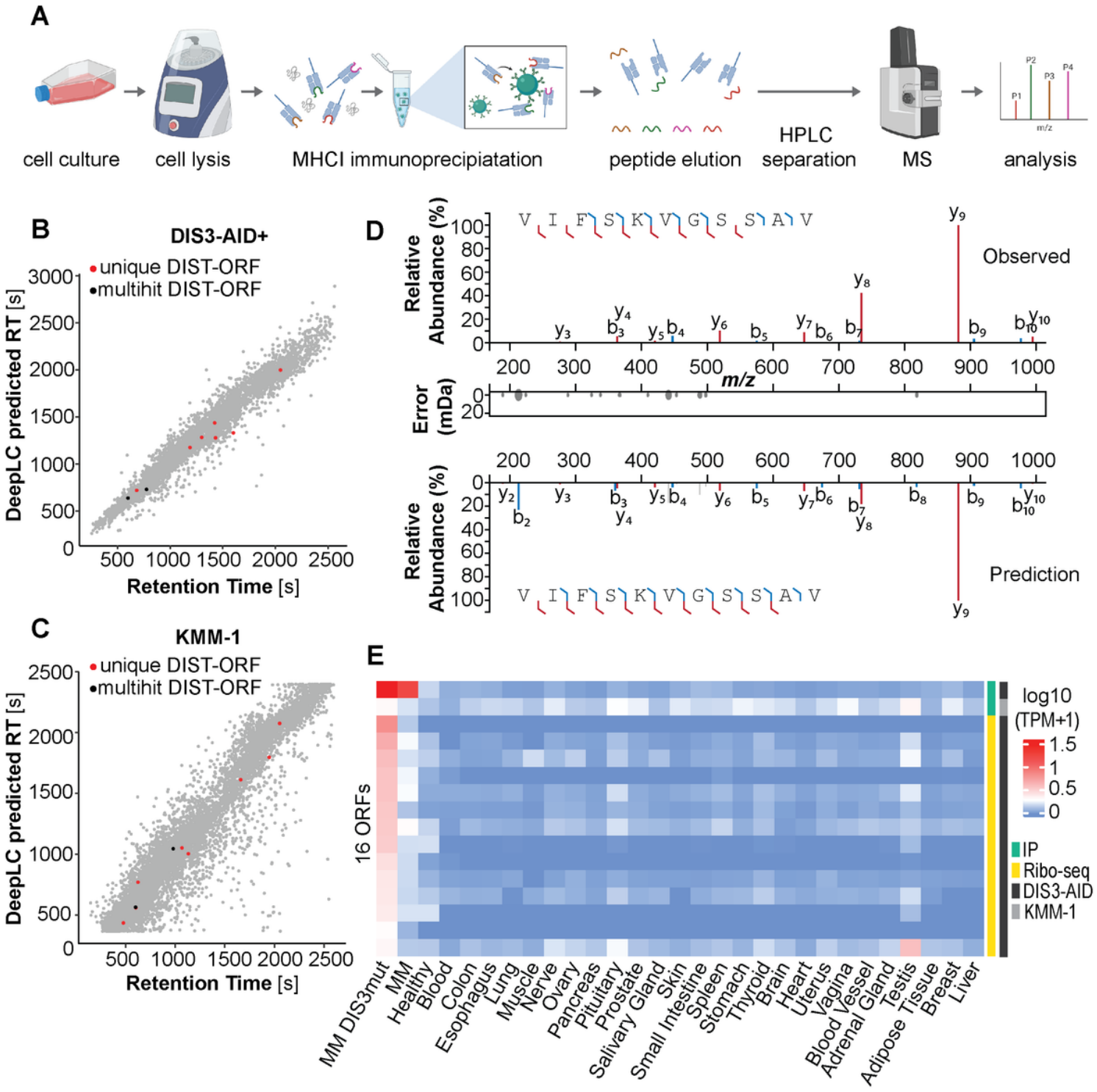

Figure 6

Some of the products of DIST-ORFs translation are presented by MHCI. A) Experimental workflow of immunopeptidomic experiments for detection of lncRNA-derived peptides (prepared with BioRender[99]). B) and C) Dot plot correlation between retention times (RT) predicted by DeelLC[81] (y-axis) and observed RT by MS (x-axis), canonical ORFs (grey), unique (red) and multiple (black) DIST-ORFs, in DIS3-AID after auxin treatment (n=5308) and KMM-1 (n=8596) cell line respectively. D) Spectra of the immunopeptide detected (top panel) and predicted by the Prosit HLA CID 2020 model (bottom panel) based on the sequence and fragment annotation tolerance set to 0.1 Da. An error between the two spectra in the middle. Plotted using Universal Spectrum Explorer[83]. E) Heatmap of 90th percentile expression (log10 of TPM+1) of 16 TSA candidate DIST-ORFs selected from Ribo-seq and Immunopeptidomic analysis in GTEx samples from heathy tissue (at least 45 samples per tissue) together with samples from the MM cohort and healthy control. The color code on the right side of the heatmap indicate the experiment type in which DIST-ORF was detected (immunpeptidomic analysis (IP) in yellow and Ribo-seq in green) and the cell line of the origin.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterialsversion3.pdf