CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex

Jennifer V. Gerbracht[1], Volker Boehm[1], Thiago Britto-Borges[2,3], Sebastian Kallabis[4], Janica L. Wiederstein[4], Simona Ciriello[1,5], Dominik U. Aschemeier[1], Marcus Krüger[4], Christian K. Frese[4,6], Janine Altmüller[7,8], Christoph Dieterich[2,3], Niels H. Gehring[1]

[1] Institute for Genetics, University of Cologne, 50674 Cologne, Germany
[2] Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III and Klaus Tschira Institute for Integrative Computational Cardiology, University of Heidelberg, 69120 Heidelberg, Germany
[3] DZHK (German Centre for Cardiovascular Research), Partner site Heidelberg/Mannheim, 69120 Heidelberg, Germany
[4] CECAD Research Center, University of Cologne, Joseph-Stelzmann-Str. 26, 50931 Cologne, Germany
[5] present address: AO Research Institute Davos, Clavadelerstrasse 8, CH-7270 Davos Platz, Switzerland
[6] present address: Max Planck Unit for the Science of Pathogens, 10117 Berlin, Germany
[7] Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany
[8] Center for Molecular Medicine Cologne, University of Cologne, 50937 Cologne, Germany

Contact
Niels H. Gehring, University of Cologne, Institute for Genetics, Zuelpicher Str. 47a, 50674 Cologne, Germany; email: ngehring@uni-koeln.de

Running Title (40 Characters)
CASC3 promotes EJC-dependent NMD

Keywords (5, alphabetical order, separated by slash)
CRISPR-Cas9/gene expression/mRNA quality control/NMD/RNA degradation

Abstract

The exon junction complex (EJC) is an essential constituent and regulator of spliced messenger ribonucleoprotein particles (mRNPs) in metazoans. As a core component of the EJC, CASC3 was described to be pivotal for EJC-dependent nuclear and cytoplasmic processes. However, recent evidence suggests that CASC3 functions differently from other EJC core proteins. Here, we have established human CASC3 knockout cell lines to elucidate the cellular role of CASC3. In the knockout cells, overall EJC composition and EJC-dependent splicing are unchanged, whereas mRNA isoforms targeted by nonsense-mediated decay (NMD) are upregulated on a transcriptome-wide scale. Mechanistically, recruiting CASC3 to reporter mRNAs by direct tethering or via binding to the EJC stimulates mRNA decay and endonucleolytic cleavage at the termination codon. Building on existing EJC-NMD models, we propose that CASC3 equips the EJC with the ability to communicate with the NMD machinery in the cytoplasm. Collectively, our results characterize CASC3 as a peripheral EJC protein that tailors the transcriptome by promoting the degradation of EJC-dependent NMD substrates.

Introduction

Messenger RNA-binding proteins (mRBPs) determine the stability, location, and fate of bound mRNAs and are therefore important regulators of post-transcriptional gene expression (Hentze *et al.*, 2018). A central component of spliced mRNPs in metazoans is the exon-junction-complex (EJC), which is deposited during splicing upstream of exon-exon boundaries (Boehm & Gehring, 2016, Le Hir *et al.*, 2016, Woodward *et al.*, 2017). The heterotetrameric core of the EJC is composed of the proteins EIF4A3, MAGOH, RBM8A (Y14) and CASC3 (BTZ, MLN51) (Andersen *et al.*, 2006, Bono *et al.*, 2006). Generally, EJCs serve on spliced mRNAs as a mark that act as a binding platform for peripheral EJC-interacting factors (Singh *et al.*, 2012). The core and peripheral EJC components contribute to different steps of post-transcriptional gene expression including splicing regulation, mRNA localization, translation and nonsense-mediated mRNA decay (NMD) (Boehm & Gehring, 2016, Woodward *et al.*, 2017).

The EJC does not form spontaneously, but instead undergoes stepwise assembly in association with the spliceosome, while it proceeds through different spliceosomal complexes (Gerbracht & Gehring, 2018). As a first step, the splicing factor CWC22 recruits EIF4A3, to which the MAGOH/RBM8A heterodimer binds later on (Gehring *et al.*, 2009, Alexandrov *et al.*, 2012, Barbosa *et al.*, 2012, Steckelberg *et al.*, 2012, Steckelberg *et al.*, 2015). Unlike the three spliceosome-associated EJC components EIF4A3, MAGOH and RBM8A, the fourth protein CASC3 is absent from the purified spliceosomal C complex (Bessonov *et al.*, 2008). Furthermore, it cannot be detected in mRNPs formed on splicing intermediates (Gehring *et al.*, 2009). Therefore, it was suggested that CASC3 binds to the initially formed trimeric pre-EJC (consisting of EIF4A3, MAGOH and RBM8A) at a later stage. Interestingly, CASC3 has also been shown to be a shuttling protein that is mainly located in the cytoplasm, whereas the other EJC components are predominantly detected in the nucleus (Kataoka *et al.*, 2000, Le Hir *et al.*, 2001, Degot *et al.*, 2002, Palacios *et al.*, 2004, Shibuya *et al.*, 2004). It has therefore been suggested that CASC3 binds to the EJC in the nucleus and is transported with it into the cytoplasm (Ballut *et al.*, 2005). A recent study demonstrated that EJCs undergo a compositional switch

and that the ASAP/PSAP component RNPS1 and the protein CASC3 bind to functionally different mRNPs and exist in mutually exclusive EJCs (Mabin *et al.*, 2018). While EJCs of nuclear enriched transcripts were found to interact with RNPS1, the EJCs of cytoplasmic enriched transcripts rather contained CASC3. This observation is in line with the predominantly cytoplasmic localization of CASC3, but would argue against a nuclear function.

Another aspect under debate is the involvement of CASC3 in the NMD pathway. According to the EJC-dependent model of NMD, an EJC present more than 50-55 nucleotides downstream of a premature termination codon (PTC) triggers degradation by the NMD machinery (Nagy & Maquat, 1998). This quality control mechanism rids the cells from aberrant transcripts that contain PTCs due to mutations or mis-splicing. Additionally, it serves as a post-transcriptional mechanism of gene expression, especially when coupled to alternative splicing (Nasif *et al.*, 2018). This is a common feature of many genes coding for mRBPs, e.g. most SR proteins (Lareau *et al.*, 2007, Ni *et al.*, 2007). The EJC triggers NMD by interacting with members of the SURF complex resulting in phosphorylation of the central NMD factor UPF1 (Kashima *et al.*, 2006). Phosphorylated UPF1 then stimulates two distinct degradation pathways of NMD: The SMG5/7-dependent pathway results in deadenylation and decapping of the transcript followed by exonucleolytic decay from the 5’ end by XRN1 and the 3’ end by the exosome (Loh *et al.*, 2013). Alternatively, the transcript can be cleaved in the vicinity of the PTC by the endonuclease SMG6 which results in two mRNA fragments that can be exonucleolytically degraded by XRN1 and the exosome (Huntzinger *et al.*, 2008, Eberle *et al.*, 2009). While these pathways can act redundantly and in principle compensate for each other, SMG6-dependent endonucleolytic cleavage (endocleavage) has been shown to be the dominant pathways for NMD in human cells (Boehm *et al.*, 2014, Lykke-Andersen *et al.*, 2014, Colombo *et al.*, 2017). In cells depleted of CASC3 a stabilizing effect on PTC-containing reporter mRNAs and selected endogenous targets was reported (Palacios *et al.*, 2004, Gehring *et al.*, 2005, Gehring *et al.*, 2009). Furthermore, tethering CASC3 to an mRNA results in UPF1-dependent degradation of the transcript (Gehring *et al.*, 2009). However, a

recent report has challenged these observations and showed that CASC3 plays a minor role in NMD and only for certain endogenous targets in contrast to EIF4A3 or RNPS1 (Mabin *et al.*, 2018).

We were intrigued by the contrasting reports about the enigmatic role of CASC3 and decided to investigate the function of CASC3 and its distinction to the other EJC core components in more detail. For this purpose, we established HEK 293 CASC3 knockout (KO) cells using CRISPR-Cas9-mediated gene editing. The CASC3 KO cell lines are unchanged in their composition of the EJC core and peripheral interacting proteins. However, RNA-sequencing reveals an upregulation of transcript variants containing premature-termination codons (PTC) as well as the differential expression of many known NMD targets, indicating a perturbation of this decay pathway. Mechanistically, CASC3 stimulates SMG6-dependent turnover of NMD targets and likely acts as a link from the EJC to the NMD machinery. On the basis of these results we propose a revised model of EJC-dependent NMD in human cells.

Results

CASC3 is dispensable for nuclear EJC functions

Previously, we and others have shown that depletion of EJC core components in human cells leads to pervasive re-splicing of cryptic splice sites, resulting in aberrant splice variants lacking exonic sequences (schematically depicted in Supplementary Figure 1) (Wang *et al.*, 2014, Boehm *et al.*, 2018). Mechanistically, the EJC prevents the use of cryptic splice sites either by interaction with the ASAP/PSAP component RNPS1 or by sterically masking splice sites (Boehm *et al.*, 2018). In keeping with previous observations from HeLa cells, the knockdowns of the EJC core components EIF4A3 or RBM8A in HEK293 cells resulted in exon-skipping of the mRNAs for RER1, OCIAD1 and MRPL3 (Figure 1 A-C). Surprisingly, a knockdown of the EJC core factor CASC3 did not result in mis-splicing of these three selected transcripts. (Figure 1A-C). However, the knockdown of CASC3 increased the abundance of an alternative transcript isoform of the SR protein SRSF2, which is known to be targeted by NMD (Figure 1D) (Sureau *et al.*, 2001, Wollerton *et al.*, 2004). Together, these results indicate that CASC3 contributes to NMD activation by the EJC, but is not required for EJC-dependent splicing regulation. Given that the generation of EJC-specific splice isoforms is a readout for EJC assembly, the initial formation of EJCs does not seem to be affected by the knockdown of CASC3.

EJC assembly is unaffected in CASC3 knockout cells

Although the knockdown efficiency of CASC3 was substantial (Figure 1E), we wished to exclude that residual amounts of CASC3 prevented a reliable assessment of the protein's function. Since CASC3 was found to be non-essential in multiple genome-wide screens of human immortalized cell lines, we reasoned that knockout (KO) of CASC3 should be feasible (Hart *et al.*, 2015, Hart *et al.*, 2017). Accordingly, we obtained three cell lines by CRISPR-Cas9-mediated gene editing, designated H, F, and T lacking the CASC3-specific 130 kDa band on a western blot (Figure 2A). In all cell lines we detected genomic insertions of different length and sequence at the

beginning of the coding region of CASC3, which resulted in frame shifts of the downstream coding region or, in the case of cell line T, contained in-frame termination codons (Figure 2B, Supplemental Figure 2A and B). For the cell lines H and F, we observed an additional band of 100 kDa on western blots with antibodies recognizing the C-terminal or central region of CASC3 (Figure 2A, red arrow, Supplementary Figure 2A and C). This cross-reactive protein interacted with FLAG-tagged EIF4A3 and disappeared upon treatment with siRNAs against CASC3 (Figure 2C, Supplementary Figure 2C and D). This suggests that the cell lines H and F produce an N-terminally truncated form of CASC3, presumably representing a novel, non-canonical protein product. Such aberrant proteins have been recently described in a systematic analysis to be common by-products of CRISPR-Cas9 genome editing (Tuladhar *et al.*, 2019).

To assess the composition of exon junction complexes and their peripheral interacting proteins in the absence of CASC3 KO, we analyzed the FLAG-tagged EIF4A3 interactome in the cell line H and wild type cells using mass spectrometry. EIF4A3 was successfully enriched, together with other known EJC complex members (Supplemental Figure 2E). Although co-precipitated CASC3 was strongly reduced in the knockout cell line (Supplemental Figure 2G, log2 fold change = -3.75), it was enriched relative to the control (Supplemental Figure 2F, log2 fold change = 2.43). This confirmed that residual amounts of CASC3 protein are still present in the knockout cell line H. When the cells were treated additionally with CASC3 siRNAs, the amount of CASC3 pulled down by EIF4A3 was reduced to background levels (Figure 2D and E). Therefore, the cell lines H and F are comparable to a constitutive knockdown of CASC3, whereas the cell line T (without further treatment) and cell line H in combination with CASC3 siRNA treatment completely lack detectable CASC3 protein. This set of cell lines enables a hitherto unfeasible analysis of CASC3's cellular function as part of the EJC.

Strikingly, no other EJC core factor or splicing regulatory EJC component (e.g. ASAP/PSAP) was significantly altered in the CASC3-depleted condition (Figure 2F). This is in line with our initial conclusion that splicing-regulatory EJCs can assembly even in the absence of CASC3. Further support comes from co-immunoprecipitation analysis of EIF4A3, which demonstrated unchanged interaction with the EJC component MAGOH in the CASC3 knockout cell line H (Supplementary Figure 2H). In agreement with the data obtained from the knockdown of CASC3 in Figure 1, transcripts containing EJC-dependent splice sites were correctly spliced in CASC3 KO cells (Figure 2G and H, Supplementary Figure 2I and J), indicating that CASC3 is not necessary for correct and stable deposition of the EJC during splicing and its subsequent nuclear splice function. The role of CASC3 during EJC assembly was also characterized by performing *in vitro* splicing of a MINX intron in cell lysates obtained from the cell line H (Figure 2I and Supplementary Figure 2K). Notably, the overall splicing efficiency of the MINX transcript was slightly reduced in the input of CASC3 KO lysates when compared to wild-type lysates. However, spliced RNA was pulled down to the same extent by EIF4A3 (2.97% in KO vs. 3.01% in WT), suggesting that EIF4A3 is deposited on spliced mRNA even when CASC3 is strongly depleted. In the absence of CASC3, the only other protein that was significantly, although mildly reduced in EIF4A3-IPs was the NMD factor UPF3B (Figure 2F, log2 fold change = -1.52). UPF3B links the EJC to the NMD machinery via direct interactions (Buchwald *et al.*, 2010) and was recently found to be enriched in cytoplasmic CASC3-loaded EJCs (Mabin *et al.*, 2018). The reduction of NMD-competent EJCs could contribute to the NMD impairment that we observed upon loss of CASC3.

CASC3 regulates NMD-sensitive isoforms

To investigate the global effects of CASC3 depletion on the transcriptome, we performed RNA-sequencing (RNA-seq) of the cell lines H and T either treated with CASC3 or control siRNAs (Figure 3A) and identified differentially expressed genes (Supplementary Figure 3A and 3B, Supplementary Table 4). To exclude clone-specific and siRNA treatment-related effects, we compared the identified targets between the four conditions. Overall, the high number and the substantial

184 overlap of upregulated genes suggests that CASC3 KO mainly results in the accumulation of certain
185 transcripts (Figure 3B and C, Supplementary Figure 3C). This observation fits with the hypothesis
186 that CASC3 is required for efficient execution of NMD. Accordingly, several upregulated genes
187 belong to the class of small RNA (e.g. snoRNA) host genes, which are frequently NMD targets
188 (Figure 3C and D) (Lykke-Andersen *et al.*, 2014). We validated the upregulation of the snoRNA host
189 gene ZFAS1 by qPCR, which was even more pronounced in CASC3 KO than UPF1 knockdown cells
190 (Supplementary Figure 3D and E). Across the top 100 significantly upregulated genes in CASC3 KO,
191 15 small RNA host genes were identified (Figure 3D). Comparing the differentially expressed genes
192 to recent transcriptome-wide NMD screens, many of the top 100 significantly upregulated genes
193 were also differentially expressed in UPF1 and SMG6/7 KD (Figure 3D, 23% and 59%, respectively)
194 (Colombo *et al.*, 2017, Data ref: Colombo *et al.*, 2017). However, none of the identified targets
195 were present in an RNPS1 knockdown (Figure 3D) (Boehm *et al.*, 2018, Data ref: Boehm *et al.*,
196 2018). Collectively, our differential gene expression analysis strengthens the proposed link
197 between CASC3 and the NMD-machinery.

198 Next, we analyzed alternative splicing changes in CASC3 KO cells (Figure 3A). Since our earlier
199 assays showed that CASC3 was not involved in EJC-regulated splicing (Figures 1 and 2), we were
200 surprised to detect many altered splicing events in CASC3 depleted cells (Supplementary Figure 3F,
201 Supplementary Table 5). It is remarkable that hardly any alternative splicing events were shared
202 between RNPS1 knockdown and CASC3 KO cells (Figure 3E, Supplementary Figure 3F). Either
203 CASC3 regulates an RNPS1-independent set of alternative splice sites or the splicing changes are
204 due to impaired NMD, which fails to remove NMD-sensitive isoforms. To test these possibilities,
205 we investigated the functional consequence of CASC3-dependent alternatively splicing on the
206 transcript isoform level (Figure 3A, Supplementary Figure 4A and B, and Supplementary Table 6).

207 Strikingly, in all CASC3 KO conditions the majority of upregulated mRNA isoforms contained a
208 premature termination codon (PTC), rendering the transcripts susceptible to NMD (Figure 4A). On

the other hand, downregulated isoforms rarely contained a PTC. Among the identified isoform switches was the NMD target SRSF2, which we confirmed earlier to be CASC3-dependent (Figure 1D). While overall SRSF2 gene expression varied only slightly between wild-type and CASC3 KO cells, the isoform usage changed dramatically towards the accumulation of NMD-sensitive transcripts in the CASC3 KO conditions (Figure 4B).

We next validated a set of transcript isoform switches by qPCR (Figure 4C-E, Supplementary Figure C and D). In the transcript isoforms stabilized by the CASC3 KO and by a UPF1 KD, the inclusion of intronic regions resulted in the inclusion of PTCs. For example, an NMD-sensitive transcript variant of the gene TOE1 is produced alongside the canonical isoform by usage of an alternative 3' splice site (Supplementary Figure 4D). This event was also prominently detected in the SMG6/7 and UPF1 dataset (Supplementary Figure 4E).

The shift of isoform usage from NMD-insensitive to PTC-containing transcripts was also observed transcriptome-wide in CASC3 KO cells and was comparable to NMD-compromised SMG6/7 or UPF1 depleted cells (Figure 4F). These findings indicate that many transcript isoforms upregulated upon depletion of CASC3 represent genuine endogenous NMD targets. On the gene level, nearly 50% of CASC3-dependent alternatively used transcript isoforms were also found in the SMG6/7 or UPF1 conditions (Figure 4G and Supplementary Figure 4F). The incomplete overlap between CASC3 and SMG6/7 or UPF1 could be due to technical differences of the data sets or because not all NMD sensitive transcripts in the cell are affected by a knockout of CASC3. This could include transcripts that are degraded by EJC-independent NMD (Buhler *et al.*, 2006, Metze *et al.*, 2013).

CASC3 stimulates SMG6-dependent endocleavage

To deepen the understanding of how a lack of CASC3 results in reduced NMD efficiency, we stably integrated the well-established globin NMD reporter PTC39 in WT and CASC3 KO cell lines (Figure 5A and B, Supplementary Figure 5A) (Thermann *et al.*, 1998). The analysis of a reporter mRNA enables a read-out of multiple aspects of mRNA degradation: firstly, the total levels of the

full-length reporter, secondly the contribution of 5'->3' exonucleolytic decay by XRN1 (detection of xrFrag due to an XRN1-resistant element (Boehm *et al.*, 2016, Voigt *et al.*, 2019); and thirdly the amount of endonucleolytic cleavage by SMG6 (detection of 3' fragment stabilized by XRN1 knockdown). In both WT and CASC3 KO cell lines the reporter was efficiently degraded, showing that the NMD pathway is still functional in CASC3 depleted cells. However, full-length reporter levels in CASC3 KO were slightly higher when compared to wild-type cells (lane 2 vs. lane 5). Notably, the accumulation of 3' fragments following XRN1 knockdown was clearly reduced in the CASC3 KO condition (lane 3 vs. lane 6). This difference in endocleavage efficiency was also observed when expressing a minigene reporter of the endogenous CASC3 target TOE1. While there was a substantial upregulation of full-length reporter mRNA in CASC3 KO cells, the amount of the 3' fragment was strongly reduced, suggesting that SMG6-dependent endocleavage is inefficient in CASC3 KO cells (Figure 5C and D, Supplementary Figure 5B-D). To further address this, a TPI reporter was expressed in combination with knockdowns of XRN1, SMG6 and/or SMG7 (Figure 5E-G). The degree of reporter and 3' fragment stabilization of the TPI reporter following XRN1 knockdown was comparable to the observations made for the globin reporter (lanes 1-3 vs. lanes 6-8). In both WT and CASC3 KO cells, SMG6 knockdown resulted in a drastic reduction of 3' fragments, as expected (lanes 4 and 9). Notably, a knockdown of SMG7 together with XRN1 revealed a major difference between the cell lines. In WT cells the PTC-containing reporter was only minimally stabilized by the SMG7/XRN1 knockdown and 3' fragments were unaffected. Performing a SMG6/XRN1 knockdown in CASC3 KO cells lead to a more dramatic stabilization of the full-length reporter and a decrease of 3' fragments compared to the XRN1 knockdown condition (lanes 3 and 5 vs. lanes 8 and 10). Collectively, our results indicate that in CASC3 KO cells SMG6-mediated endocleavage is impaired. This could explain why the CASC3 KO cells are more sensitive to a knockdown of SMG7 when compared to wild type cells.

To identify, which part of CASC3 promotes NMD, we employed a tethering reporter that was designed to monitor mRNA turnover as well as endocleavage at the termination codon (Figure 6A).

Tethering the full-length CASC3 protein to the MS2 stem loops downstream of the stop codon resulted in degradation of the reporter compared to tethering of the negative control GST (Figure 6B and C, Supplementary Figure 6A). This degradation was accompanied by the production of 3' fragments in XRN1 knockdown conditions, indicating that the mechanism of decay is comparable to the PTC-containing reporter mRNAs (Figure 6B, lane 6). Surprisingly, C-terminally truncated deletion mutants of CASC3 that contain the first 480 or even 137 amino acid residues were able to induce degradation of the tethering reporter to a comparable extent as full-length CASC3 (Figure 6B, lanes 3 and 4, 7 and 8).

Finally, CASC3 deletion mutants were expressed in the CASC3 KO cells to identify the minimal part necessary to rescue the effects on endogenous NMD targets (Figure 6D, Supplementary Figure 6B and C). As in the tethering experiment, expression of full-length CASC3 and the C-terminal truncated variant 1-480 resulted in transcript isoform levels comparable to wild-type cells for the targets CLN6 and TOE1 (Figure 6D lanes 1-4). An EJC binding-deficient mutant of CASC3 (188/218 double point mutation) was unable to rescue, supporting the notion that CASC3 is recruited to the mRNA by binding to the EJC (Figure 6D, lane 5). Deleting the N-terminal 109 amino acids of CASC3 (110-480) did not alter the rescue ability (Figure 6D, lane 6). While in the tethering assay it was sufficient to place the N-terminus downstream of a termination codon, this part of CASC3 was not necessary to rescue NMD activity in the KO cells. This suggests that different domains of CASC3 act in a redundant manner during the activation of NMD by the EJC.

Discussion

The role of CASC3 within the EJC has been the subject of scientific controversy for many years. CASC3 has been initially described as an EJC core protein, because it was required for the assembly of the EJC from recombinant protein components *in vitro* (Ballut *et al.*, 2005). However, it has been demonstrated that the mechanism of EJC assembly using recombinant proteins is mechanistically different from EJC assembly in splicing extracts or in living cell (Steckelberg *et al.*, 2012). Furthermore, several recent publications challenged the view of CASC3 being an EJC core component. For instance, CASC3 was reported to be present in substoichiometric amounts compared to the other three EJC core proteins EIF4A3, RBM8A, and MAGOH in HEK293 (Singh *et al.*, 2012) and U2OS cells (Beck *et al.*, 2011). Also, during mouse embryonic brain development CASC3 deficiency results in a different phenotype than the other EJC core components (Mao *et al.*, 2017). By using CASC3 CRISPR-Cas9 knockout cells, we unambiguously establish that CASC3 is not required for EJC assembly or EJC-regulated splicing in the nucleus (Figure 7). Therefore, our molecular analyses fully support the recently emerging view of defining CASC3 as a peripheral EJC component. As a mainly cytoplasmic component of the EJC we propose that the principal role of CASC3 is to alter the efficiency by which NMD-sensitive transcript isoforms are degraded.

Although NMD has been extensively studied in the past decades and many NMD factors have been identified and characterized, no universal model exists that describes how they work together to elicit NMD. While a function of CASC3 in NMD has been reported before, previous analyses did not show consistent results, ranging from a substantial contribution of CASC3 to only a minor role in NMD (Palacios *et al.*, 2004, Gehring *et al.*, 2005, Gehring *et al.*, 2009, Mabin *et al.*, 2018). In addition, none of the previous publications performed transcriptome-wide analyses but concentrated on reporter mRNAs of only a few selected endogenous NMD targets. We reasoned that the inconsistent results in the literature may be influenced by variable CASC3 knockdown efficiency. By generating CASC3 knockout cell lines, we can for the first time analyze the global effects of a complete depletion of CASC3 on the transcriptome and can exclude that residual CASC3 protein masks these effects. Interestingly,

the CASC3 knockout had no deleterious effect on the viability of HEK293 cells, unlike the depletion of the EJC components EIF4A3, RBM8A, or MAGOH. Nonetheless, since CASC3 is required for mouse embryogenesis and involved in the transport of mRNAs in *D. melanogaster*, it is likely that CASC3 downregulation in highly specialized cell types such as neurons or in developing tissues would result in a more severe phenotype compared to HEK293 cells.

In recent years, high-throughput RNA-sequencing became an increasingly important method for the analysis of NMD. Several RNA-Seq datasets of cells with NMD-factor knockdowns have been generated and analyzed (Tani *et al.*, 2012, Hurt *et al.*, 2013, Lykke-Andersen *et al.*, 2014, Schmidt *et al.*, 2015, Colombo *et al.*, 2017). However, these datasets were obtained in different cell lines, with different amounts of replicates and due to the rapid developments of next-generation sequencing, not using the same technologies. Furthermore, batch effects and divergent approaches of data analyses may contribute to the fact that only a minor overlap of NMD targets could be established so far (Colombo *et al.*, 2017). We compared the results of our CASC3 KO RNA-sequencings to the most recent and comprehensive NMD factor analysis performed by Colombo *et al.* (2017). The differential expression analysis revealed that many of the top upregulated genes in the CASC3 KO datasets are also significantly affected by UPF1 or SMG6/7 knockdowns (Colombo *et al.*, 2017) and/or encode for small RNA (sRNA) host genes, a previously described class of NMD targets (Lykke-Andersen *et al.*, 2014).

We detected many alternative splicing events in the CASC3 KO data, which was unexpected given that CASC3 was apparently dispensable for the nuclear EJC-related functions. However, we could attribute these splicing patterns to dysfunctional NMD, since isoform-specific algorithms revealed that predominantly PTC-containing transcripts accumulated. Using all the available bioinformatics analysis pipelines and a systematic approach to detect affected transcripts under NMD factor knockdown/knockout conditions could therefore be a crucial step to paint a complete picture of the regulation of transcripts by NMD in the future.

How exactly CASC3 activates NMD when bound to an EJC, is not yet fully understood. Previously, we reported that the presence of EJCs in the 3' UTR enhances endocleavage (Boehm *et al.*, 2014). In line with the proposed role as a peripheral NMD-activating EJC component, we observed that CASC3 stimulates SMG6-dependent endocleavage, thereby promoting the degradation of NMD-targeted transcripts. This effect can be recapitulated by tethering full-length CASC3, its N-terminal two thirds (1-480) or just its N-terminal 137 amino acids to a reporter mRNA. How the small N-terminal region, which cannot assemble into the EJC or contains any known protein domains or sequence motifs can elicit NMD remains to be determined. It is also unclear, if the N-terminus activates translation-dependent degradation, as it was previously shown for the full length CASC3 (Boehm *et al.*, 2016). Since the N-terminal segment of CASC3 is a region of low-complexity it could hypothetically undergo liquid-liquid phase separation (LLPS) and be present in condensates with mRNA decay factors, such as processing bodies (P-bodies). In agreement with this idea CASC3 was shown to localize to cytoplasmic granules when overexpressed (Cougot *et al.*, 2014).

Our data suggest that CASC3 activates NMD by potentially redundant mechanisms. This fits very well into the general picture of NMD, which uses multiplexed degradation pathways to efficiently and robustly remove mRNAs. Mechanistically, binding of CASC3 to the EJC could have an indirect effect on NMD stimulation by increasing the stability of the bound EJC and thus maintaining the possibility of efficient endonucleolytic cleavage of the transcript. An indication for this role comes from the initial *in vitro* observation that CASC3 stabilizes recombinant EJCs (Ballut *et al.*, 2005). Additionally, the moderately reduced pull-down of UPF3B with EIF4A3 could indicate that cytoplasmic NMD-competent EJCs are less stable in CASC3 KO cells. Alternatively, CASC3 may directly contribute to the recruitment of NMD factors. We therefore propose that CASC3 either alone or in conjunction with UPF3B links the EJC with the NMD machinery. In particular, CASC3 influences the contribution of SMG6-mediated endonucleolytic and SMG7-dependent exonucleolytic decay pathways to the overall degradation efficiency of NMD. Accordingly, in wild type cells a knockdown of SMG7 only had a marginal effect on the abundance of the analyzed NMD reporter mRNA, whereas it clearly impaired NMD in CASC3 KO

355 cells. Also, the amount of endocleavage-derived 3' fragments was reduced when CASC3 is depleted,
356 mirroring the SMG6-knockdown condition.

357 By integrating CASC3 as a specific NMD-activating factor we can now postulate a modified model
358 of EJC-dependent NMD, which is also compatible with several molecular properties of the EJC (Figure
359 7). Since CASC3 is only present in modest amounts in the cytoplasm, it will probably not immediately
360 associate with all EJCs on recently exported mRNPs. This would also not be necessary, since most EJCs
361 are located in the coding sequence and will therefore be removed by the first translating ribosome.
362 However, mRNAs containing PTCs will carry one or more EJCs in their 3' UTR, which are available for
363 binding of CASC3. The first few translating ribosomes may terminate upstream of CASC3-free EJCs,
364 which could preferentially trigger SMG7-dependent exonucleolytic degradation. Previously, NMD has
365 been proposed to occur primarily in the pioneering round of translation when newly synthesized
366 transcripts are bound to the cap-binding complex (Ishigaki *et al.*, 2001, Maquat *et al.*, 2010). This model
367 has been challenged and there is evidence that NMD can occur on already translating mRNAs and
368 possibly with a constant probability during every round of translation (Durand & Lykke-Andersen,
369 2013, Rufener & Muhlemann, 2013, Hoek *et al.*, 2019). Thus, CASC3 could bind to the EJC at a later
370 time point and then increase the probability to activate SMG6-mediated endonucleolytic degradation
371 after each round of termination. Important molecular targets of CASC3 may be mRNAs that escape
372 initial NMD activation, despite containing a PTC (Trcek *et al.*, 2013, Hoek *et al.*, 2019). CASC3 may help
373 to reduce the amount of these mRNAs by maintaining the NMD-activating function of the EJC, either
374 by increasing its stability on the mRNA or via direct interactions with the NMD machinery. This concept
375 would be consistent with the recent observation that NMD targets undergo several rounds of
376 translation before endocleavage occurs (Hoek *et al.*, 2019).

377 In summary, our data paint a picture, in which CASC3 has no essential EJC-related function in the
378 nucleus, but helps to turn EJCs into a persistent RNA degradation mode. We do not exclude the
379 possibility that CASC3 is already associated with the EJC in the nucleus. However, our model of delayed

380 binding of CASC3 to the EJC in the cytoplasm would explain why only a small amount of CASC3 is
381 sufficient to activate EJC-dependent NMD. In this model, CASC3 is an indispensable cytoplasmic
382 component of the EJC that helps to degrade mRNAs that failed to unload all their bound EJCs during
383 the initial rounds of translation. Thus, the binding of CASC3 to the EJC could signal the final round(s) of
384 translation of an mRNA.

Materials and Methods

Cell culture

Flp-In 293 T-REx cells (Thermo Fisher Scientific) were maintained at 37° C, 5% CO_2 and 90% humidity in Dublecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific) supplemented with 9% fetal bovine serum (FBS) and Penicillin-Streptomycin (both Thermo Fisher Scientific). Tethering experiments were performed in HeLa Tet-Off cells (Clontech) cultured in the same conditions.

siRNA-mediated knockdowns

The cells were seeded in 6-well plates at a density of $2x10^5$ cells per well and reverse transfected using 2.5 µl Lipofectamine RNAiMAX and 60 pmol of the respective siRNA(s) according to the manufacturer's instructions. In preparation for mass spectrometry, the cells were reverse transfected in 10 cm dishes using 10 µl Lipofectamine RNAiMAX and 300 pmol siRNA. siRNAs were targeted against Luciferase (5'-CGTACGCGGAATACTTCGA-3'), EIF4A3 (5'-AGACATGACTAAAGTGGAA-3'), RBM8A (5'-TTCGCAGAATATGGGGAAA-3'), CASC3 (5'-CTGATGACATCAAACCTCGAAGAAT-3', 5'-CGTCATGAACTTTGGTAATCCCAGT-3'), UPF1 (5'-GATGCAGTTCCGCTCCATT-3'), XRN1 (5'-AGATGAACTTACCGTAGAA-3'), SMG6 (5'-GGGTCACAGTGCTGAAGTA-3') or SMG7 (5'-CGATTTGGAATACGCTTTA-3').

Generation of knockout cells using CRISPR-Cas9

The knockouts were performed using the Alt-R CRISPR-Cas9 system (IDT) and reverse transfection of a Cas9:guideRNA ribonucleoprotein complex using Lipofactamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's protocol. The crRNA sequences to target CASC3 were /AlTR1/rGrCrGrCrGrCrUrUrCrGrCrArArGrArCrArCrCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/ (clone H) and /AlTR1/rGrUrUrCrGrGrCrCrUrCrCrGrCrGrCrUrGrUrGrArGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/ (clones F and T). Reverse transfection was performed on $1.5x10^5$ cells per crRNA in 12-well dishes.

48 hours after transfection the cells were trypsinized, counted and seeded at a density of a single cell per well in 96-well plates. Cell colonies originating from a single clone were then validated by Sanger sequencing of the targeted genomic DNA locus and western blotting.

Plasmid transfection

All used plasmids are listed in Supplementary Table 1. To express FLAG-tagged protein constructs and the reporter mRNAs detected by northern blotting, the cells were stably transfected using the Flp-In T-REx system and the tetracycline inducible pcDNA5/FRT/TO vector (Thermo Fisher Scientific). The constructs TPI-WT, TPI-PTC, β-globin WT and β-globin PTC are available on Addgene (IDs 108375-108378). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells. Per well, 1 µg of reporter construct was transfected together with 1 µg of the Flp recombinase expressing plasmid pOG44 using the calcium phosphate method. 48 h after transfection, the cells were transferred into 10 cm dishes and selected with 100 µg/ml hygromycin. After 10 days, the colonies were pooled. Expression of the reporter mRNA was induced with 1 µg/ml doxycycline for 24 h.

Constructs that express V5-tagged and MS2V5-tagged proteins were stably integrated into the cells using the PiggyBac (PB) Transposon system and the cumate-inducible PB-CuO-MCS-IRES-GFP-EF1-CymR-Puro vector (System Biosciences). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells. 2.5 µg of the PB Transposon vector and 0.8 µg of PB Transposase were transfected per well using the calcium phosphate method. After 48 h, the cells were pooled in 10 cm dishes and positive clones selected with 2 µg/ml puromycin for a week. Expression of proteins was induced using 30 µg/ml cumate for 72 h.

The tethering construct pSBtet-Hyg-TPI-4MS2-SMG5-4H was stably integrated into HeLa Tet-Off cells using the Sleeping Beauty (SB) transposon system (Mates *et al.*, 2009, Kowarz *et al.*, 2015). pSBtet-Hyg was a gift from Eric Kowarz (Addgene plasmid #60508; http://n2t.net/addgene:60508; RRID:Addgene_60508). pCMV(CAT)T7-SB100 was a gift from Zsuzsanna Izsvak (Addgene plasmid #34879; http://n2t.net/addgene:34879; RRID:Addgene_34879). $2.5x10^5$ cells were seeded 24 h before

435 transfection in 6-wells. Per well, 1 µg of the reporter construct was transfected together with 1.5 µg
436 of the SB Transposase using the calcium phosphate method. 48 h after transfection, the cells were
437 transferred into 10 cm dishes and selected with 100 µg/ml hygromycin. After 10 days, the colonies
438 were pooled. In absence of tetracycline the reporter was constitutively expressed.

439 *RNA-Sequencing and computational analyses*

440 RNA-Seq analysis was carried out with 293 WT cells transfected with Luciferase siRNA and the
441 CASC3 KO clones H and T transfected with either Luciferase or CASC3 siRNAs. Three biological
442 replicates were analyzed for each sample. RNA was isolated with the kit NucleoSpin RNA Plus
443 (Macherey-Nagel). The Lexogen SIRV Set1 Spike-In Control Mix (SKU: 025.03) that provides a set of
444 external RNA controls was added to the total RNA to enable performance assessment. Mix 0 was added
445 to replicate 1, mix 1 was added to replicate 2 and mix 3 to replicate 3. The Spike-Ins were not used for
446 analysis. The library preparation was performed with the TrueSeq Stranded Total RNA kit (Illumina).
447 First steps of the library preparation involve the removal of ribosomal RNA using biotinylated target-
448 specific oligos combined with Ribo-Zero gold rRNA removal beads from 1ug total RNA input. The Ribo-
449 Zero Human/Mouse/Rat kit depletes samples of cytoplasmic and mitochondrial rRNA. Following
450 purification, the RNA is fragmented and cleaved. RNA fragments are copied into first strand cDNA using
451 reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA
452 Polymerase I and RNase H. These cDNA fragments then have the addition of a single'A' base and
453 subsequent ligation of the adapter. The products are purified and enriched with PCR to create the final
454 cDNA library. After library validation and quantification (Agilent tape station), equimolar amounts of
455 library were pooled. The pool was quantified by using the Peqlab KAPA Library Quantification Kit and
456 the Applied Biosystems 7900HT Sequence Detection System and sequenced on an Illumina
457 NovaSeq6000 sequencing instrument and an PE100 protocol.

458 Read processing and alignment was performed as described previously (Boehm *et al.*, 2018). In
459 short, adaptor sequences and low quality bases were removed with Flexbar 3.0 (Dodt *et al.*, 2012).

460 Short reads from the rRNA locus were subtracted by mapping against the 45S precursor (Homo
461 sapiens, NR_046235.1) using Bowtie2 (Langmead & Salzberg, 2012). The remaining reads were aligned
462 against the human genome (version 38, EnsEMBL 90 transcript annotations) using the STAR read
463 aligner (version 2.5.3a) (Dobin *et al.*, 2013).

464 To compute gene differential expression analysis, reads covering exons were counted with
465 FeatureCounts (version 1.5.1) (Liao *et al.*, 2014) using the '—primary' and '—ignoreDup' parameters.
466 Differential gene expression analysis was performed with DESeq2 (Love *et al.*, 2014, Ignatiadis *et al.*,
467 2016) and IWH R packages. Significance thresholds were |log2FoldChange|> 1 and adjusted p-value
468 (padj) < 0.05. Genes were designated as small RNA (sRNA) host gene, if they contained other Ensembl-
469 annotated genes of biotypes snoRNA or miRNA within their genomic coordinates (Zerbino *et al.*, 2018).

470 Differential splicing was detected with LeafCutter (version 0.2.7) (Li *et al.*, 2018) with the
471 parameters min_samples_per_intron = 2 and min_samples_per_group = 2. Significance thresholds
472 were |deltapsi| > 0.1 and adjusted p-value (p.adjust) < 0.05.

473 Transcript abundance estimates were computed with Salmon (version 0.13.1) (Patro *et al.*, 2017)
474 using the the -validateMappings --gcBias parameters. Differential transcript usage was computed with
475 IsoformSwitchAnalyzeR (version 1.7.1) and the DEXSeq method (Robinson & Oshlack, 2010, Anders *et*
476 *al.*, 2012, Ritchie *et al.*, 2015, Soneson *et al.*, 2015, Vitting-Seerup & Sandelin, 2017, Vitting-Seerup &
477 Sandelin, 2019). Significance thresholds were |dIF| > 0.1 and adjusted p-value
478 (isoform_switch_q_value) < 0.05. For the Boxplot and Kolmogorov-Smirnoff test, the data were filtered
479 only for the adjusted p-value. PTC status of transcript isoforms with annotated open reading frame
480 was determined by IsoformSwitchAnalyzeR using the 50 nt rule of NMD (Weischenfeldt *et al.*, 2012,
481 Vitting-Seerup *et al.*, 2014, Huber *et al.*, 2015, Vitting-Seerup & Sandelin, 2017). Isoforms with no
482 annotated open reading frame in Ensembl were designated "NA" in the PTC analysis.

483 The UPF1 and SMG6/7 (Data ref: Colombo *et al.*, 2017) and RNPS1 (Data ref: Boehm *et al.*, 2018)
484 knockdown datasets were processed and analyzed with the same programs, program versions, and

485 scripts as the CASC3 dataset. All packages used are listed in the respective analysis table
486 (Supplementary Tables 4-6). Sashimi plots were generated using ggsashimi (Garrido-Martin *et al.*,
487 2018). Overlaps of data sets were represented via nVenn (Perez-Silva *et al.*, 2018), eulerr (Larsson,
488 2019) and Upset plots (Lex *et al.*, 2014). Heatmaps were generated using ComplexHeatmap (Gu *et al.*,
489 2016).

490 *SILAC, co-immunoprecipitation and mass spectrometry*

491 293 WT and 293 CASC3 KO clone H cells expressing either FLAG or FLAG-EIF4A3 were labeled by
492 maintaining them for 5 passages in DMEM for SILAC medium (Thermo Fisher Scientific) supplemented
493 with FBS (Silantes), Penicillin-Streptomycin (Thermo Fisher Scientific) and the respective amino acids
494 at a final concentration of 0.798 mmol/L (Lysine) and 0.398 (Arginine). Unlabeled proline was added
495 to prevent enzymatic Arginine-to-Proline conversion. The conditions were “light” (unlabeled
496 Lysine/Arginine), “medium” (Lysine 4/Arginine 6) and “heavy” (Lysine 8/Arginine 10). A label switch
497 was performed between the three replicates according to the experimental setup listed in
498 Supplementary Table 2. The expression of FLAG or FLAG-EIF4A3 was induced for 72 h with 1 µg/ml
499 doxycycline. The cells were lysed in buffer E with RNAse (20 mM HEPES-KOH (pH 7.9), 100 mM KCl,
500 10% glycerol, 1 mM DTT, Protease Inhibitor, 1 µg/ml RNAse A) and sonicated using the Bandelin
501 Sonopuls mini20 with 15 pulses (2.5 mm tip in 600 µl volume, 1s, 50% amplitude). 600 µl of a 1.6 mg/ml
502 total protein lysate were incubated with 30 µl Anti-FLAG M2 magnetic beads (Sigma) at 4° C while
503 rotating for 2 h. The beads were washed three times for 5 min with EJC-buffer (20 mM HEPES-KOH (pH
504 7.9), 200 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-100, 0.1% NP-40, 0.05% Sodium deoxycholate) and
505 eluted in 43 µl of a 200 mg/ml dilution of FLAG peptides (Sigma) in 1x TBS. The samples were merged
506 according to Supplementary Table 2. 1 volume of 10% SDS was added and the samples were reduced
507 with DTT and alkylated with CAA (final concentrations 5 mM and 40 mM, respectively). Tryptic protein
508 digestion was performed using a modified version of the single pot solid phase-enhanced sample
509 preparation (SP3) (Hughes *et al.*, 2014). In brief, reduced and alkylated proteins were supplemented
510 with paramagnetic Sera-Mag speed beads (Thermo Fisher Scientific) and mixed in a 1:1-ratio with 100%

511 acetonitrile (ACN). After 8 min incubation protein-beads-complexes were captured using an in-house
512 build magnetic rack and two times washed with 70% EtOH. Afterwards, samples were washed once
513 with 100% ACN, air-dried and reconstituted in 5 µl 50 mM Triethylamonium bicarbonate
514 supplemented with 0.5 µg trypsin and 0.5 µg LysC and incubated overnight at 37°C. On the next day
515 the beads were resuspended and mixed with 200 µl ACN, incubated for 8 min and again placed on the
516 magnetic rack. Tryptic peptides were washed once with 100% ACN, airdried, dissolved in 4% DMSO
517 and transferred into 96-well PCR tubes. After acidification with 1 µl of 10% formic acid, samples were
518 ready for LC-MS/MS analysis

519 For the MS experiment combined with CASC3 knockdown, the 293 CASC3 KO clone H cells were
520 treated with siRNA against CASC3 24 h before expression of the FLAG-tagged construct. The cells were
521 lysed and samples prepared for MS as described above. However, the washing steps were performed
522 in a less stringent EJC buffer (20 mM HEPES-KOH (pH 7.9), 137 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-
523 100, 0.1% NP-40).

524 Proteomics analysis was performed by data-dependent acquisition using an Easy nLC1200 ultra
525 high-performance liquid chromatography (UHPLC) system coupled via nanoelectrospray ionization to
526 a Q Exactive Plus instrument (all Thermo scientific). Tryptic peptides were separated based on their
527 hydrophobicity using a chromatographic gradient of 60 min with a binary system of buffer A (0.1%
528 formic acid) and buffer B (80% ACN, 0.1% formic acid). In-house made analytical columns (length: 50
529 cm, inner diameter: 75 µm) filled with 1.9 µm C18-AQ Reprosil Pur beads (Dr. Maisch) were used for
530 separation. Buffer B was linearly increased from 3% to 27% over 41 min followed by a steeper increase
531 to 50% within 8 min. Finally, buffer B was increased to 95% within 1 min and stayed at 95% for 10 min
532 in order to wash the analytical column. Full MS spectra (300 – 1,750 m/z) were acquired with a
533 resolution of 70,000, a maximum injection time of 20 ms and an AGC target of 3e6. The top 10 most
534 abundant peptide ions of each full MS spectrum were selected for HCD fragmentation (NCE: 27) with

an isolation width of 1.8 m/z and a dynamic exclusion of 10 seconds. MS/MS spectra were measured with a resolution of 35,000, a maximum injection time of 110 ms and an AGC target of 5e5.

MS RAW files were analysed using the standard settings of the MaxQuant suite (version 1.5.3.8) with the before mentioned SILAC labels (Cox & Mann, 2008). Peptides were identified by matching against the human UniProt database using the Andromeda scoring algorithm (Cox *et al.*, 2011). Carbamidomethylation of cysteine was set as a fixed modification, methionine oxidation and N-terminal acetylation as variable modification. Trypsin/P was selected as digestion protein. A false discovery Rate (FDR) < 0.01 was used for identification of peptide-spectrum matches and protein quantification. Data processing and statistical analysis was done in the Perseus software (version 1.5.5.3) (Tyanova *et al.*, 2016). Significantly changed proteins were identified by One-sample t-testing (H0 = 0, fudge factor S0 = 0.1). The results are listed in Supplementary Table 2 as "Results without knockdown" (MAP conditions) and "Results with knockdown". Visualization was performed with the Instant Clue software (version 0.5.3) (Nolte *et al.*, 2018).

Co-immunoprecipitation experiments followed by western blotting were performed as described above except that a 15 min incubation step in SDS buffer (600 mM Tris pH 6.8, 100 mM DTT, 10% Glycerol, 2% SDS, 0.002% Bromophenolblue) was used for elution from the beads.

In vitro splicing assay

FLAG-emGFP and FLAG-EIF4A3 was expressed in 293 WT and 293 CASC3 KO clone H cells for 72 h. Whole cell lysate preparation and *in vitro* splicing was performed as previously described (Steckelberg & Gehring, 2014), however the nuclear extract was substituted with whole cell extract from the respective cell line. Pulldown efficiency of spliced mRNA was calculated by dividing the amount of spliced mRNA in the pulldown by the amount in the respective input and dividing the results by 10 since 10% of the input was loaded.

Semi-quantitative and quantitative reverse transcriptase (RT)-PCR

559 RNA was extracted using peqGOLD TriFast reagent (VWR) according to the manufacturer's
560 instructions. Reverse transcription was performed with GoScript Reverse Transcriptase (Promega)
561 using 2 µg total RNA and oligo dT primers. Semi-quantitative PCR was carried out with MyTaq Red Mix
562 (Bioline). Quantitative real time PCR was performed with 16 ng of cDNA per reaction with GoTaq qPCR
563 Master Mix (Promega) and the CFX96 Touch Real-Time PCR Detection System (Biorad). The average cT
564 values were calculated from three technical replicates. The mean fold changes from three biological
565 replicates were calculated according to the ΔΔCt method (Schmittgen & Livak, 2008). When measuring
566 isoform switches, the fold change of the PTC-containing transcript was normalized to the canonical
567 transcript. When measuring differential expression, the fold change was normalized to GAPDH. For
568 each primer pair amplification efficiencies were measured by a 2-fold dilution curve and ranged
569 between 87 and 100.1%. The primer sequences are listed in Supplementary Table 3.

570 *Western blotting*

571 Protein extraction was performed with peqGOLD TriFast reagent (VWR), separated by SDS-PAGE
572 gel electrophoresis and transferred to a PVDF membrane (GE Healthcare Life Sciences). The following
573 antibodies were used: Anti-β-actin (Sigma-Aldrich, #A5441), anti-CASC3 amino acid residues 653-703
574 (Bethyl Laboratories, #A302-472A-M), anti-CASC3 amino acid residues 367-470 (Atlas Antibodies,
575 #HPA024592), anti-EIF4A3 (Genscript), anti-FLAG (Cell Signaling Technology, #14793), anti-MAGOH
576 (Santa Cruz Biotechnology, #sc-271365), anti-RBM8A (Atlas Antibodies, #HPA018403), anti-SMG6
577 (Abcam, #ab87539), anti-SMG7 (Elabscience, #E-AB-32926), anti-Tubulin (Sigma-Aldrich, #T6074), anti-
578 V5 (QED Bioscience, #18870), anti-XRN1 (Bethyl Laboratories, #A300-443A), anti-rabbit-HRP (Jackson
579 ImmunoResearch, #111-035-006), anti-mouse-HRP (Jackson ImmunoResearch, #115-035-003).
580 Detection was performed with Western Lightning Plus-ECL (PerkinElmer) or Amersham ECL prime (Ge
581 Healthcare Life Sciences) and the chemiluminescence imager Fusion FX6 EDGE (Vilber-Lourmat).

582 *Northern blotting*

583 The cells were harvested in peqGOLD TriFast reagent (VWR) and total RNA extraction was
584 performed as recommended by the manufacturer's protocol. 2.5 µg of total RNA were resolved on a
585 1% agarose/0.4 M formaldehyde gel using the tricine/triethanolamine buffer system (Mansour &
586 Pestov, 2013) followed by transfer on a nylon membrane (Roth) in 10x SSC. The blots were incubated
587 overnight at 65 °C in Church buffer containing [α-32P]-GTP body-labeled RNA probes for detection of
588 the reporter mRNA. Endogenous 7SL RNA was detected by a 5'-32P-labeled oligonucleotide (5'-
589 TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT-3'). The blots were visualized and quantified using a
590 Typhoon FLA 7000 (GE Healthcare) and ImageQuant TL 1D software.

591 *Data Availability*

592 The datasets produced in this study are available in the following databases. These data will be
593 made publicly accessible upon publication.

- 594 RNA-seq data have been deposited in the ArrayExpress database (Kolesnikov *et al.*, 2015) at
595 EMBL-EBI under accession number E-MTAB-8461
596 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8461).
- 597 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
598 Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) partner repository with the dataset
599 identifier PXD015754 (https://www.ebi.ac.uk/pride/archive/projects/PXD015754).

Author Contributions

Conceptualization, N.H.G., J.V.G. and V.B.;

Methodology, N.H.G., V.B., J.V.G., and C.K.F.;

Software, T.B.B., V.B., J.L.W., S.K. and C.D.;

Investigation, J.V.G, V.B., J.L.W., S.K., D.U.A. and S.C.;

Resources and Data Curation, T.B.B., J.L.W., S.K., J.A. and C.D.;

Writing – Original Draft, Review & Editing, N.H.G., J.V.G. and V.B.;

Visualization, J.V.G., V.B. and T.B.B.;

Supervision, N.H.G., C.D. and M.K.;

Funding Acquisition, N.H.G. and C.D.

Acknowledgements

We thank members of the Gehring lab for discussions and reading of the manuscript. We also thank Marek Franitza and Christian Becker (Cologne Center for Genomics, CCG) for preparing the sequencing libraries and operating the sequencer. We acknowledge Tobias Jakobi for helping with infrastructure support. This work was supported by grants from the Deutsche Forschungsgemeinschaft to C.D. (DI 1501/8-1, DI1501/8-2) and N.H.G (GE 2014/6-1, GE 2014/6-2 and GE 2014/10-1). V.B. was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative. N.H.G. acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1) from the Deutsche Forschungsgemeinschaft. C.D. and T.B.B. were kindly supported by the Klaus Tschira Stiftung gGmbH (00.219.2013)

Conflict of Interest

The authors declare no competing interests.

References

Alexandrov A, Colognori D, Shu MD, Steitz JA (2012) Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc Natl Acad Sci U S A* 109: 21313-8

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22: 2008-17

Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, Pedersen JS, Seraphin B, Le Hir H, Andersen GR (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313: 1968-72

Ballut L, Marchadier B, Baguet A, Tomasetto C, Seraphin B, Le Hir H (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* 12: 861-9

Barbosa I, Haque N, Fiorini F, Barrandon C, Tomasetto C, Blanchette M, Le Hir H (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* 19: 983-90

Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (2011) The quantitative proteome of a human cell line. *Mol Syst Biol* 7: 549

Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452: 846-50

Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) ArrayExpress E-MTAB-6564 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6564)

Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol Cell* 72: 482-495 e7

Boehm V, Gehring NH (2016) Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends Genet* 32: 724-735

Boehm V, Gerbracht JV, Marx MC, Gehring NH (2016) Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* 7: 13691

Boehm V, Haberman N, Ottens F, Ule J, Gehring NH (2014) 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep* 9: 555-68

Bono F, Ebert J, Lorentzen E, Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126: 713-25

Buchwald G, Ebert J, Basquin C, Sauliere J, Jayachandran U, Bono F, Le Hir H, Conti E (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc Natl Acad Sci U S A* 107: 10050-5

Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol* 13: 462-4

661 Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O (2017) Gene Expression Omnibus
662 GSE86148 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86148). [DATASET]

663 Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O (2017) Transcriptome-wide
664 identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and
665 SMG7-mediated degradation pathways. *RNA* 23: 189-201

666 Cougot N, Daguenet E, Baguet A, Cavalier A, Thomas D, Bellaud P, Fautrel A, Godey F, Bertrand E,
667 Tomasetto C, Gillet R (2014) Overexpression of MLN51 triggers P-body disassembly and formation of
668 a new type of RNA granules. *J Cell Sci* 127: 4692-701

669 Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range
670 mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-72

671 Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search
672 engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794-805

673 Degot S, Regnier CH, Wendling C, Chenard MP, Rio MC, Tomasetto C (2002) Metastatic Lymph Node
674 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene* 21: 4422-34

675 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013)
676 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21

677 Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-Flexible Barcode and Adapter Processing for
678 Next-Generation Sequencing Platforms. *Biology (Basel)* 1: 895-905

679 Durand S, Lykke-Andersen J (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent
680 translation in human cells. *Nat Struct Mol Biol* 20: 702-9

681 Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH (2009) SMG6 promotes endonucleolytic
682 cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16: 49-55

683 Garrido-Martin D, Palumbo E, Guigo R, Breschi A (2018) ggsashimi: Sashimi plot revised for browser-
684 and annotation-independent splicing visualization. *PLoS Comput Biol* 14: e1006360

685 Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE (2005) Exon-junction
686 complex components specify distinct routes of nonsense-mediated mRNA decay with differential
687 cofactor requirements. *Mol Cell* 20: 65-75

688 Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE (2009) The hierarchy of exon-junction complex
689 assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay.
690 *PLoS Biol* 7: e1000120

691 Gerbracht JV, Gehring NH (2018) The exon junction complex: structural insights into a faithful
692 companion of mammalian mRNPs. *Biochem Soc Trans* 46: 153-161

693 Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in
694 multidimensional genomic data. *Bioinformatics* 32: 2847-9

695 Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M,
696 Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat
697 J (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities.
698 *Cell* 163: 1515-26

699 Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, Chandrashekhar M, Hustedt N,
700 Seth S, Noonan A, Habsid A, Sizova O, Nedyalkova L, Climie R, Tworzyanski L, Lawson K, Sartori MA,
701 Alibeh S, Tieu D, Masud S et al. (2017) Evaluation and Design of Genome-Wide CRISPR/SpCas9
702 Knockout Screens. *G3 (Bethesda)* 7: 2719-2727

703 Hentze MW, Castello A, Schwarzl T, Preiss T (2018) A brave new world of RNA-binding proteins. *Nat*
704 *Rev Mol Cell Biol* 19: 327-341

705 Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M,
706 Tanenbaum ME (2019) Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA
707 Decay. *Mol Cell* 75: 324-339 e11

708 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke
709 T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles
710 AK, Pages H et al. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat*
711 *Methods* 12: 115-21

712 Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J (2014) Ultrasensitive proteome
713 analysis using paramagnetic bead technology. *Mol Syst Biol* 10: 757

714 Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E (2008) SMG6 is the catalytic endonuclease
715 that cleaves mRNAs containing nonsense codons in metazoan. *RNA* 14: 2609-17

716 Hurt JA, Robertson AD, Burge CB (2013) Global analyses of UPF1 binding and function reveal expanded
717 scope of nonsense-mediated mRNA decay. *Genome Res* 23: 1636-50

718 Ignatiadis N, Klaus B, Zaugg JB, Huber W (2016) Data-driven hypothesis weighting increases detection
719 power in genome-scale multiple testing. *Nat Methods* 13: 577-80

720 Ishigaki Y, Li X, Serin G, Maquat LE (2001) Evidence for a pioneer round of mRNA translation: mRNAs
721 subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106:
722 607-17

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

726 Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G (2000) Pre-mRNA splicing
727 imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell*
728 6: 673-82

729 Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi
730 M, Burdett T, Megy K, Pilicheva E, Rustici G, Tikhonov A, Parkinson H, Petryszak R, Sarkans U, Brazma
731 A (2015) ArrayExpress update--simplifying data submissions. *Nucleic Acids Res* 43: D1113-6

732 Kowarz E, Loscher D, Marschalek R (2015) Optimized Sleeping Beauty transposons rapidly generate
733 stable transgenic cell lines. *Biotechnol J* 10: 647-53

734 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357-9

735 Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes
736 associated with highly conserved and ultraconserved DNA elements. *Nature* 446: 926-9

Larsson J (2019) eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.0.0, https://cran.r-project.org/package=eulerr

Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E (2001) The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep* 2: 1119-24

Le Hir H, Sauliere J, Wang Z (2016) The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol* 17: 41-54

Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H (2014) UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* 20: 1983-92

Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* 50: 151-158

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923-30

Loh B, Jonas S, Izaurralde E (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* 27: 2125-38

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen TH (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* 28: 2498-517

Mabin JW, Woodward LA, Patton RD, Yi Z, Jia M, Wysocki VH, Bundschuh R, Singh G (2018) The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep* 25: 2431-2446 e7

Mansour FH, Pestov DG (2013) Separation of long RNA by agarose-formaldehyde gel electrophoresis. *Analytical biochemistry* 441: 18-20

Mao H, Brown HE, Silver DL (2017) Mouse models of Casc3 reveal developmental functions distinct from other components of the exon junction complex. *RNA* 23: 23-31

Maquat LE, Tarn WY, Isken O (2010) The pioneer round of translation: features and functions. *Cell* 142: 368-74

Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J, Ma L, Samara-Kuko E, Gysemans C, Pryputniewicz D, Miskey C, Fletcher B, VandenDriessche T, Ivics Z, Izsvak Z (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet* 41: 753-61

Metze S, Herzog VA, Ruepp MD, Muhlemann O (2013) Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA* 19: 1432-48

Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23: 198-9

775 Nasif S, Contu L, Muhlemann O (2018) Beyond quality control: The role of nonsense-mediated mRNA
776 decay (NMD) in regulating gene expression. *Semin Cell Dev Biol* 75: 78-87

777 Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M, Jr.
778 (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by
779 alternative splicing and nonsense-mediated decay. *Genes Dev* 21: 708-18

780 Nolte H, MacVicar TD, Tellkamp F, Kruger M (2018) Instant Clue: A Software Suite for Interactive Data
781 Visualization and Analysis. *Sci Rep* 8: 12648

782 Palacios IM, Gatfield D, St Johnston D, Izaurralde E (2004) An eIF4AIII-containing complex required for
783 mRNA localization and nonsense-mediated mRNA decay. *Nature* 427: 753-7

784 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware
785 quantification of transcript expression. *Nat Methods* 14: 417-419

786 Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J,
787 Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain
788 E, Walzer M, Jarnuczak AF et al. (2019) The PRIDE database and related tools and resources in 2019:
789 improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

790 Perez-Silva JG, Araujo-Voces M, Quesada V (2018) nVenn: generalized, quasi-proportional Venn and
791 Euler diagrams. *Bioinformatics* 34: 2322-2324

792 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential
793 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47

794 Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of
795 RNA-seq data. *Genome Biol* 11: R25

796 Rufener SC, Muhlemann O (2013) eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA
797 decay in mammalian cells. *Nat Struct Mol Biol* 20: 710-7

798 Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ (2015)
799 Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of
800 endogenous NMD targets in human cells. *Nucleic Acids Res* 43: 309-23

801 Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nature*
802 *protocols* 3: 1101-8

803 Shibuya T, Tange TO, Sonenberg N, Moore MJ (2004) eIF4AIII binds spliced mRNA in the exon junction
804 complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* 11: 346-51

805 Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ (2012) The cellular EJC
806 interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151: 750-764

807 Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates
808 improve gene-level inferences. *F1000Res* 4: 1521

809 Steckelberg AL, Altmueller J, Dieterich C, Gehring NH (2015) CWC22-dependent pre-mRNA splicing and
810 eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res* 43: 4687-700

811 Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH (2012) CWC22 connects pre-mRNA splicing and
812 exon junction complex assembly. *Cell Rep* 2: 454-61

Steckelberg AL, Gehring NH (2014) Studying the composition of mRNPs in vitro using splicing-competent cell extracts. *Methods* 65: 342-9

Sureau A, Gattoni R, Dooghe Y, Stevenin J, Soret J (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* 20: 1785-96

Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol* 9: 1370-9

Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeier C, Hentze MW, Kulozik AE (1998) Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* 17: 3484-94

Trcek T, Sato H, Singer RH, Maquat LE (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev* 27: 541-51

Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J, Mathews DH, Kim J, Hyun Hwang T, Lum L (2019) CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* 10: 4056

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13: 731-40

Vitting-Seerup K, Porse BT, Sandelin A, Waage J (2014) spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* 15: 81

Vitting-Seerup K, Sandelin A (2017) The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res* 15: 1206-1220

Vitting-Seerup K, Sandelin A (2019) IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*

Voigt F, Gerbracht JV, Boehm V, Horvathova I, Eglinger J, Chao JA, Gehring NH (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. *Nature protocols* 14: 1603-1633

Wang Z, Murigneux V, Le Hir H (2014) Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol* 15: 551

Weischenfeldt J, Waage J, Tian G, Zhao J, Damgaard I, Jakobsen JS, Kristiansen K, Krogh A, Wang J, Porse BT (2012) Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol* 13: R35

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 13: 91-100

Woodward LA, Mabin JW, Gangras P, Singh G (2017) The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdiscip Rev RNA* 8

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR et al. (2018) Ensembl 2018. *Nucleic Acids Res* 46: D754-D761

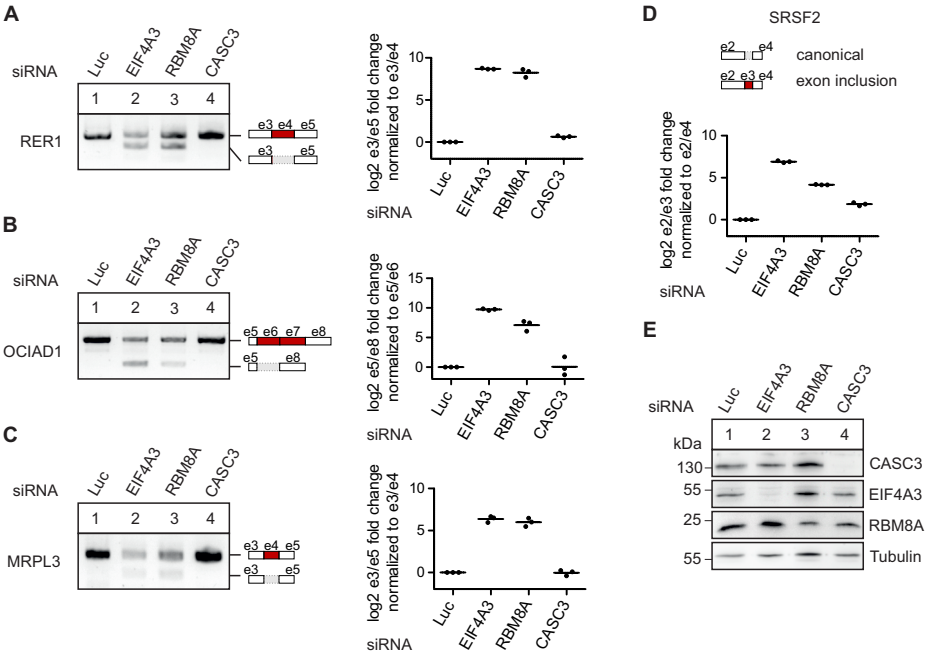

Figure 1 - CASC3 is not involved in the splicing regulation of known EJC-dependent targets.

A-C: RT-PCR- and quantitative RT-PCR-based detection (qPCR) of transcript isoforms of the genes RER1 (A), OCIAD1 (B), and MRPL3 (C) after siRNA-mediated knockdown of the indicated EJC components or Luciferase (Luc) as a negative control. Skipped exons are depicted schematically (e: exon). Data points and means from the qPCRs are plotted (n=3).

D: Relative quantification of the SRSF2 transcript isoforms by qPCR following knockdown of the indicated EJC components or Luciferase (Luc) as a negative control. The transcript variants at the position of the included exon are depicted schematically. Data points and means are plotted (n=3).

E: Confirmation of the knockdowns shown in A-D by western blotting.

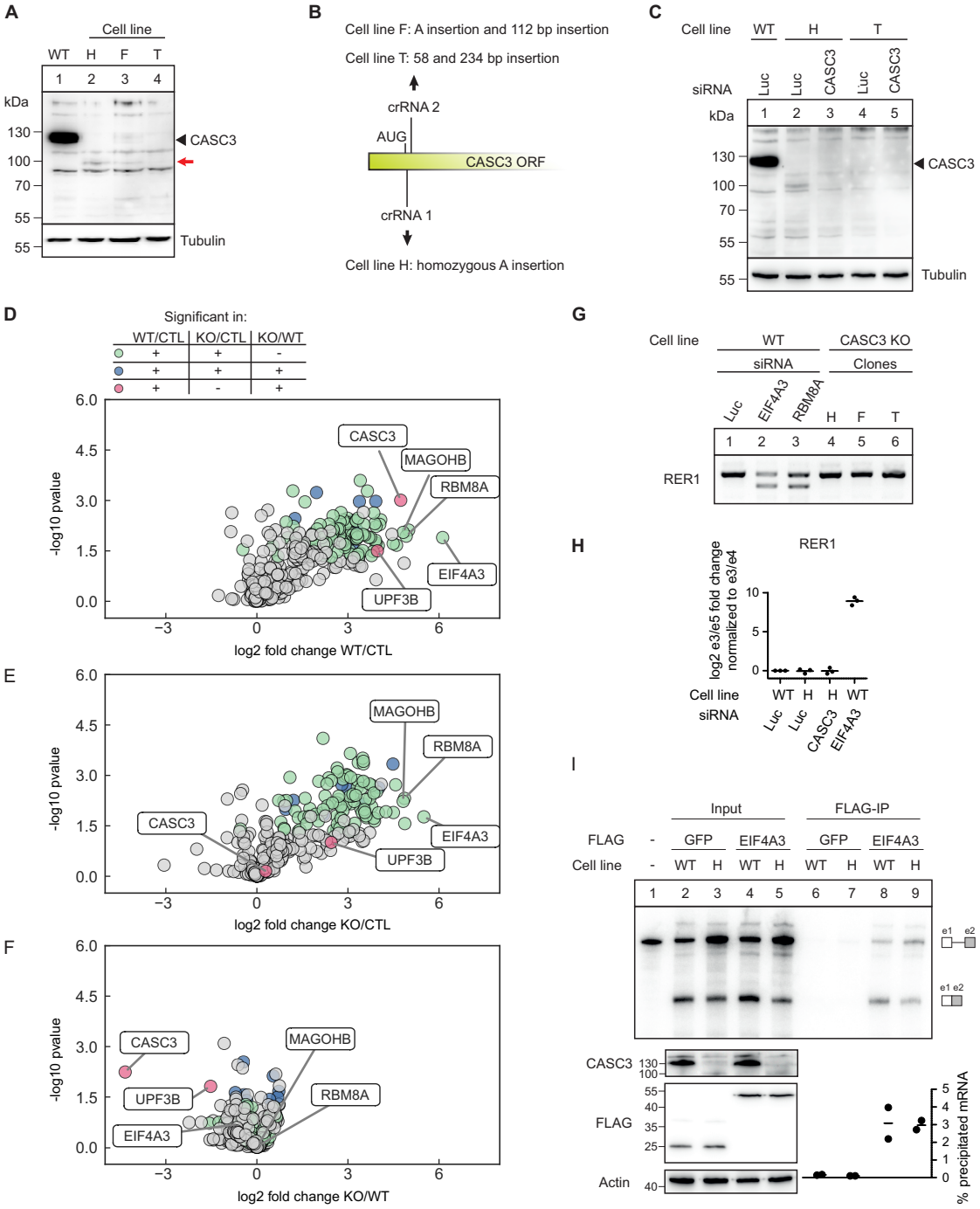

Figure 2 - Cells that lack CASC3 have intact EJCs and are splicing competent.

A: Total protein lysates from wild-type cells (WT) and CASC3 knockout (KO) cell lines H, F and T were separated by SDS-PAGE and CASC3 was detected by western blotting. The red arrow indicates an additional band detected in the cell lines H and F and which is not visible in the WT condition or in the cell line T.

B: Schematic depiction of the insertions resulting in a CASC3 KO or constitutive knockdown in the indicated clones.

C: Total protein lysates from WT and CASC3 KO cell lines H and T were separated by SDS-PAGE and CASC3 was detected by western blotting. In lanes 3 and 5 the cells have additionally been treated with siRNAs targeting CASC3.

D-F: Volcano plots of mass spectrometry-based analysis of the interaction partners of EIF4A3 in WT cells and in the CASC3 KO cell line H treated with siRNAs targeting CASC3. D: EIF4A3 against FLAG control in WT cells, E: EIF4A3 against FLAG control in KO cells, F: EIF4A3 in KO cells against EIF4A3 in WT cells. The color labeling indicates targets that are significant in the respective comparisons after one-sample t-testing.

G: RT-PCR of transcript isoforms of the gene RER1 after siRNA-mediated knockdown of the indicated EJC components or Luciferase (Luc) as a negative controle, compared to CASC3 KO cell lines H, F and T.

H: Relative quantification of the RER1 transcript isoforms by qPCR in WT cells treated with Luc siRNA as a negative control, CASC3 KO cell line H treated with Luc siRNA, CASC3 KO cell line H treated with CASC3 siRNAs and WT cells treated with EIF4A3 siRNA. Data points and means are plotted (n=3).

I: *In vitro* splicing assay of a MINX transcript in cell lysates obtained from WT cells or the CASC3 KO cell line H which stably expressed FLAG-GFP or FLAG-EIF4A3. The protein levels of CASC3 and the recombinant proteins in the cell lysates were analyzed by western blotting. The pulldown efficiencies of spliced mRNA were calculated and are plotted as individual data points and means (n=2).

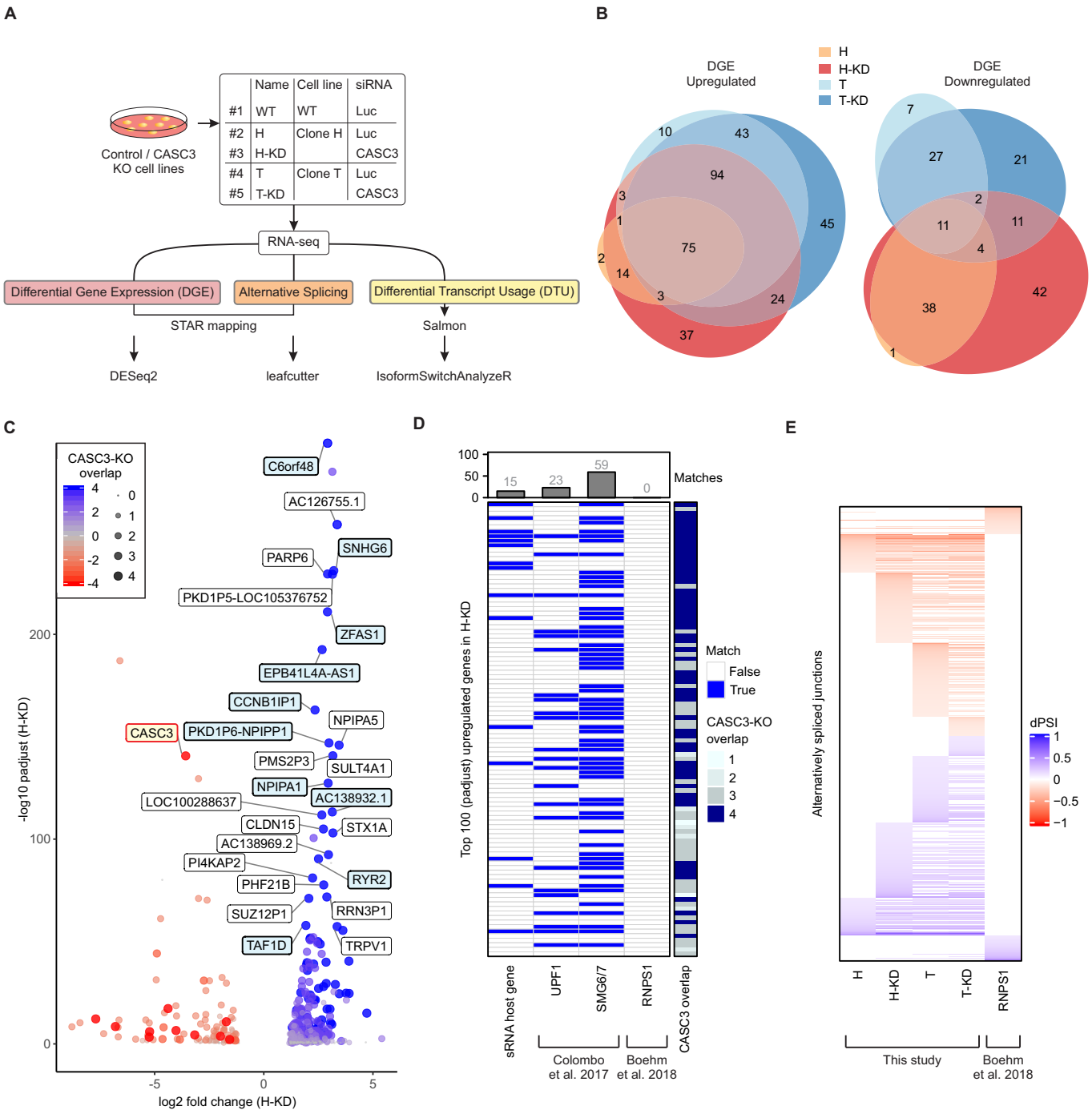

Figure 3 - CASC3-dependent targets are also affected by a perturbation of the NMD machinery.

A: Pipeline for RNA-sequencing analysis.

B: Overlap of up- and downregulated genes in the CASC3 KO cell lines H and T, +/- CASC3 siRNAs. DGE: Differential gene expression. Due to the visualization as Euler plot, some intersections cannot be plotted. All intersections are shown in Supplementary Figure 3C.

C: Volcano plot of differential gene expression analysis of the condition H-KD using overlap from B as color and point size definition. Gene symbols are indicated for the top 25 upregulated genes detected in all four conditions and for CASC3 (colored in light red). Labels of small RNA host genes are colored in light blue. Log2 fold change is plotted against –log10 padjust (adjusted p-value).

D: Matching of the top 100 upregulated genes sorted by padjust (adjusted p-value) in condition H-KD with small RNA (sRNA) host genes and comparison to knockdowns of UPF1, SMG6/7 and RNPS1.

E: Heatmap of all identified alternatively spliced junctions in the respective condition, measured in delta percent spliced in (dPSI).

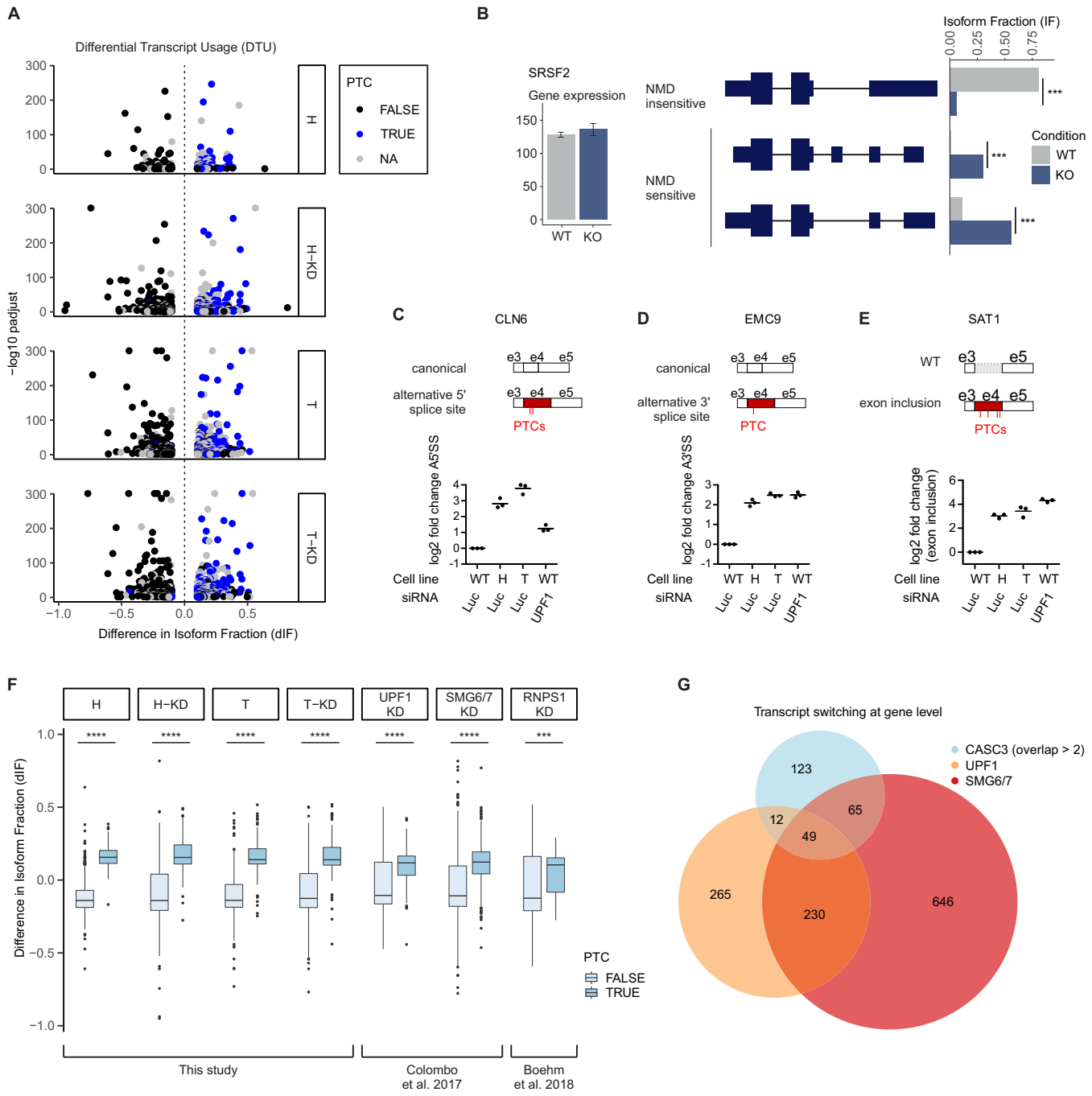

Figure 4 - Knockout of CASC3 leads to a global upregulation of NMD-sensitive transcript isoforms.

A: Results from IsoformSwitch analysis plotted as volcano plots. Transcript isoforms identified as NMD-sensitive are shown as blue dots. Isoforms with no annotated open reading frame are designated as “NA”. Difference in Isoform Fraction (dIF) is plotted against –log10 padjust (adjusted p-value).

B: Quantification of transcript isoforms from SRSF2 by IsoformSwitchAnalyzeR.

C-E: Relative quantification of the schematically depicted transcript isoforms of the genes CLN6 (C), EMC9 (D), and SAT1 (E) by qPCR in WT cells, CASC3 KO cell lines H and T and WT cells treated with siRNA targeting UPF1. PTC: premature termination codon. Individual data points and means are plotted (n=3).

F: Boxplot of PTC-containing vs. non-PTC-containing transcript isoforms after IsoformSwitch analysis for all CASC3 KO conditions compared to UPF1, SMG6/7 and RNPS1. A Kolmogorov-Smirnoff test was applied (p-value < 0.001 ***, p-value < 10-16 ****).

G: Euler plot showing overlaps of DTU-affected genes in the CASC3 (> 2 overlaps), UPF1-KD and SMG6/7-KD condition.

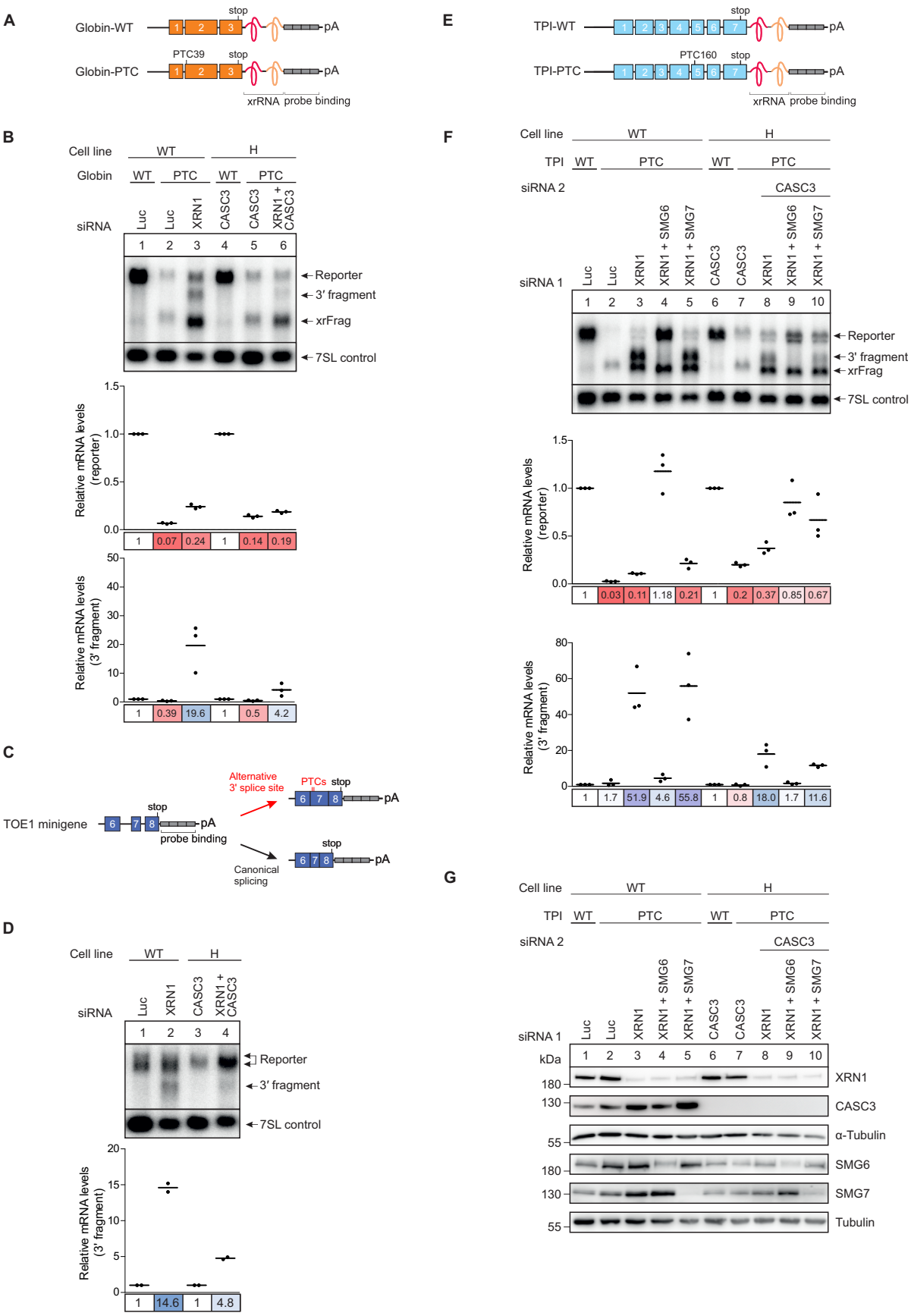

Figure 5 - SMG6-mediated endocleavage is impaired when CASC3 is not present.

A: Schematic depiction of the globin mRNA reporter. The reporter consists of three exons (orange boxes) followed by an XRN1-resistant element (xrRNA) and a probe binding cassette (gray boxes). The PTC reporter contains a premature termination codon (PTC) in the second exon.

B: Northern blot of RNA extracted from the indicated cell lines that stably express the globin reporter. The xrFrag corresponds to the 3′ part of the reporter that is resistant to degradation by XRN1 due to the xrRNA. The cell lines in lane 3 and 6 were additionally treated with XRN1 siRNA which results in the appearance of a 3′ degradation fragment below the full-length reporter. Reporter and 3′ fragment mRNA levels were normalized to 7SL RNA which is shown as a loading control. For the relative mRNA quantification, in each condition (WT vs. CASC3 KO with KD) the reporter and 3′ fragment levels were normalized to the globin WT reporter (lanes 1 and 4). Individual data points and means are plotted from n=3 experiments.

C: Schematic depiction of the TOE1 minigene reporter consisting of exons 6-8 (purple boxes) followed by a probe binding cassette (gray boxes). The reporter can be spliced to either contain the canonical stop codon (bottom right) or, by usage of an alternative 3′ splice site, a PTC in exon 7 (top right).

D: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing the TOE1 minigene reporter. The 3′ fragment levels were first normalized to the 7SL RNA loading control and for every cell line the XRN1 knockdown condition to the condition without XRN1 knockdown (n=2).

E: Schematic depiction of the triose phosphate isomerase (TPI) mRNA reporter. The reporter consists of seven exons (blue boxes) followed by an XRN1-resistant element (xrRNA) and a probe binding cassette (gray boxes). The PTC reporter contains a premature termination codon (PTC) in the fifth exon.

F: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing the either the TPI WT or TPI PTC mRNA reporter. The reporter and 3′ fragment mRNA levels were normalized to the 7SL control. For each cell line, the mRNA levels were then normalized to the respective TPI WT reporter or 3′ fragment levels. Individual data points and means are plotted from n=3 experiments.

G: Western blot of samples shown in Figure 5F.

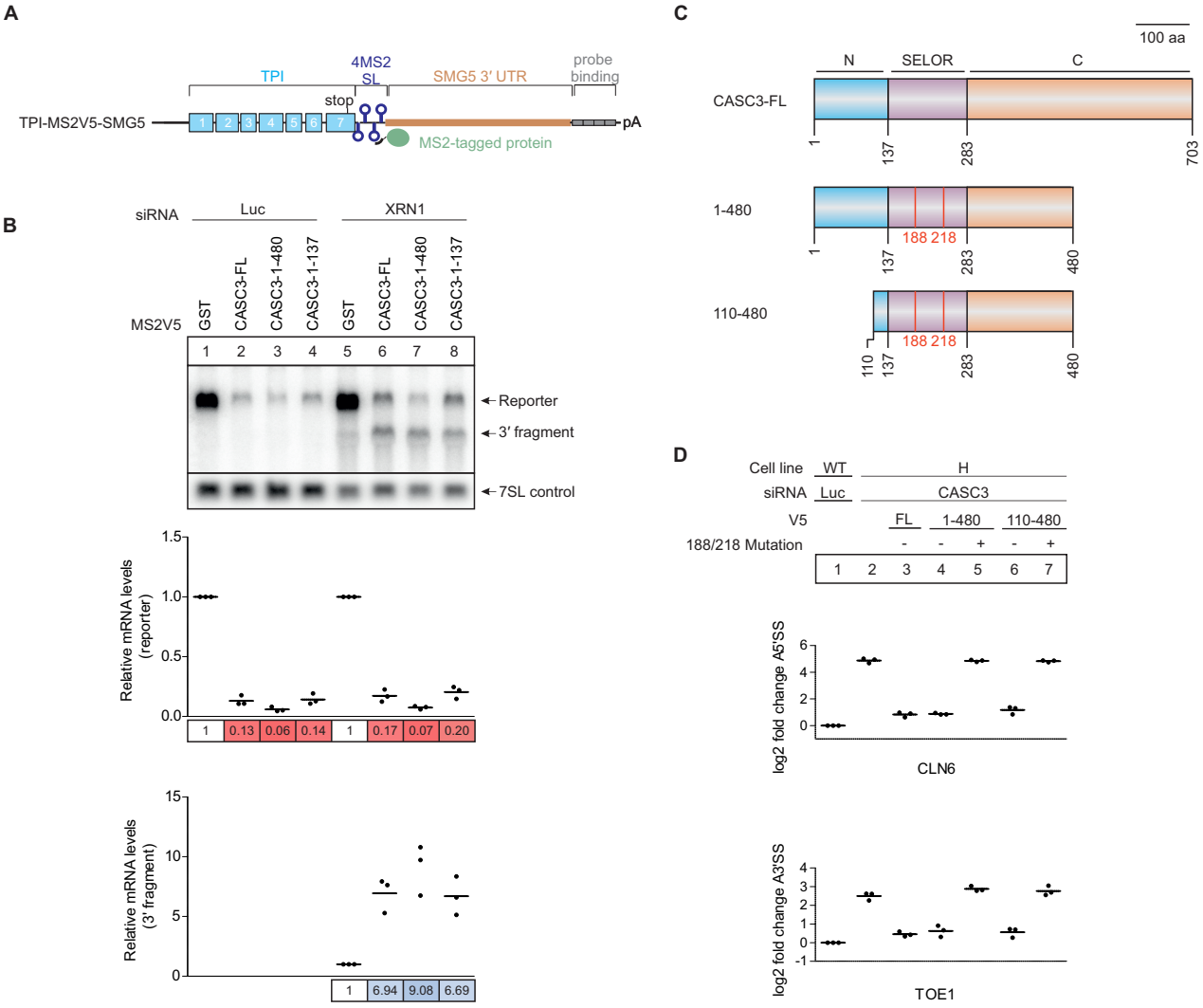

Figure 6 - The CASC3 N-terminus promotes but is not necessary to elicit NMD.

A: Schematic depiction of the TPI-MS2V5-SMG5 tethering reporter. The reporter consists of the TPI ORF (blue boxes) followed by 4 MS2 stem loops (SL). Downstream the SMG5 3′ untranslated region (UTR) is inserted to increase the size of 3′ fragments that result from cleavage at the termination codon. Reporter and 3′ fragment mRNAs can be detected via the probe binding cassette (gray boxes).

B: Northern blot of a tethering assay performed in HeLa Tet-Off cells. The cells stably express the tethering reporter shown in A together with the indicated MS2V5-tagged proteins. When the cells are additionally treated with XRN1 siRNA, a 3′ degradation fragment can be detected below the full-length reporter. The reporter and 3′ fragment mRNA levels are normalized to the 7SL RNA. For the calculation of the relative mRNA levels in each condition (Luc vs. XRN1) the levels were normalized to the MS2V5-GST control (lanes 1 and 5).

C: Schematic depiction of CASC3 rescue protein constructs. The full-length (FL) protein consists of an N-terminal (blue), C-terminal (orange) and central SELOR domain (purple). The construct 1-480 has a C-terminal deletion, whereas in the construct 110-480 both the N- and C-terminus are truncated. Both deletion constructs were also rendered EJC-binding deficient by mutating the amino acid residues 188 and 218 (F188D, W218D).

D: Relative quantification of the CLN6 (top) and TOE1 (bottom) transcript isoforms by qPCR in the indicated cell lines. The V5-tagged rescue proteins expressed in the KO condition are shown in C. Individual data points and means are plotted from n=3 experiments.

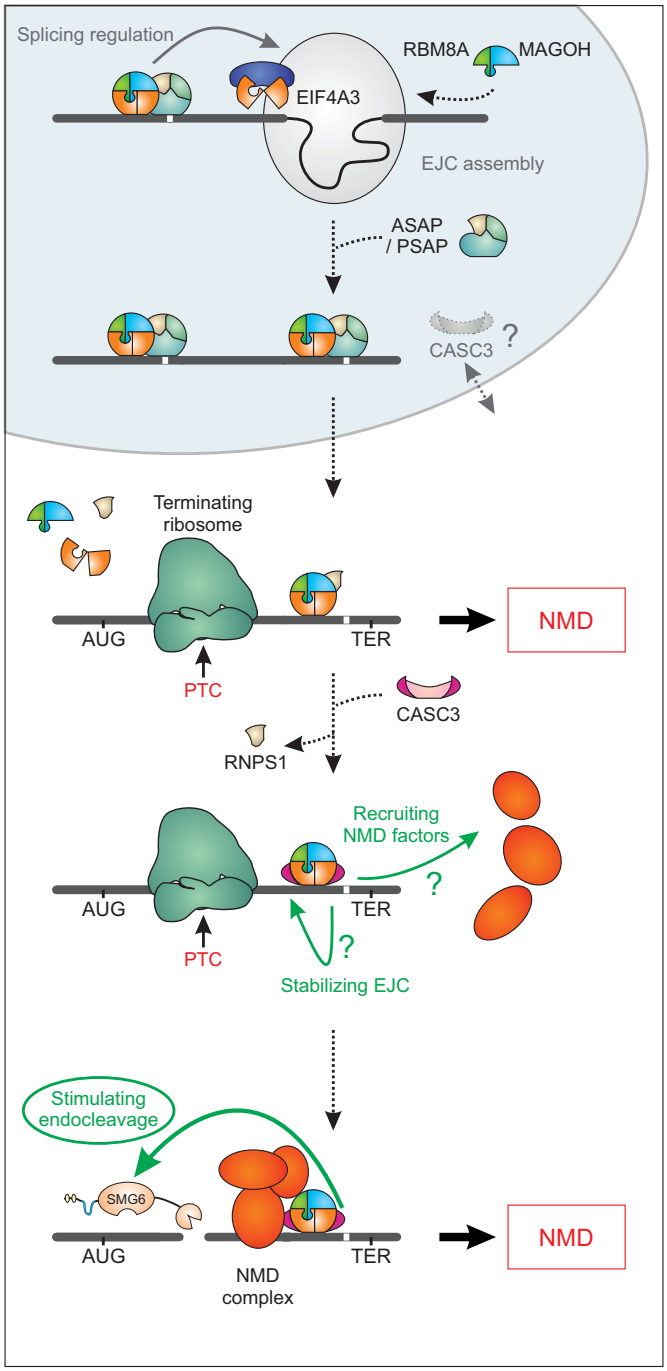

Figure 7 - Model of CASC3's cellular function.

Schematic depiction of the proposed function of CASC3 in the cell. We have found no evidence that CASC3 is necessary for EJC assembly in the nucleus, although CASC3 shuttles between cytoplasm and nucleus. Transcripts that require the deposition of the EJC to be correctly spliced were not affected by a lack of CASC3. In the cytoplasm, premature termination codon (PTC)-containing transcripts may still be degraded by NMD during the initial round(s) of translation in a CASC3-independent manner. CASC3 association with the EJC maintains and/or promotes the NMD-stimulating effect of the EJC, resulting in the degradation of transcripts that evaded initial NMD activation.