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. A transcriptome-wide analysis reveals that hundreds of mRNA isoforms targeted by nonsense-mediated decay (NMD) are upregulated. 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.

47 *Introduction*

48 Messenger RNA-binding proteins (mRBPs) determine the stability, location, efficiency of
49 translation and fate of bound mRNAs and are therefore important regulators of post-transcriptional
50 gene expression (1). A central component of spliced mRNPs in metazoans is the exon-junction-complex
51 (EJC), which is deposited during splicing upstream of exon-exon boundaries (2-4). The
52 heterotetrameric core of the EJC is composed of the proteins EIF4A3, MAGOH, RBM8A (Y14) and CASC3
53 (BTZ, MLN51) (5,6). Generally, EJCs serve on spliced mRNAs as a mark that act as a binding platform
54 for peripheral EJC-interacting factors (7). The core and peripheral EJC components contribute to
55 different steps of post-transcriptional gene expression including splicing regulation, mRNA localization,
56 translation and nonsense-mediated mRNA decay (NMD) (2,4).

57 The EJC does not form spontaneously, but instead undergoes stepwise assembly in association
58 with the spliceosome, while it proceeds through different spliceosomal complexes (8). As a first step,
59 the splicing factor CWC22 recruits EIF4A3, to which the MAGOH/RBM8A heterodimer binds later on
60 (9-13). Unlike the three spliceosome-associated EJC components EIF4A3, MAGOH and RBM8A, the
61 fourth protein CASC3 is not detected in the purified spliceosomal C complex (14,15). Furthermore, it
62 cannot be detected in mRNPs formed on splicing intermediates (12). Therefore, it was suggested that
63 CASC3 binds to the initially formed trimeric pre-EJC (consisting of EIF4A3, MAGOH and RBM8A) at a
64 later stage. Interestingly, CASC3 has also been shown to be a shuttling protein that is mainly located in
65 the cytoplasm, whereas the other EJC components are predominantly detected in the nucleus (16-20).
66 It has therefore been suggested that CASC3 binds to the EJC in the nucleus and is transported with it
67 into the cytoplasm (21). A recent study demonstrated that EJCs undergo a compositional switch and
68 that the ASAP/PSAP component RNPS1 and the protein CASC3 bind to functionally different mRNPs
69 and exist in mutually exclusive EJCs (22). While EJCs of nuclear enriched transcripts were found to
70 interact with RNPS1, the EJCs of cytoplasmic enriched transcripts rather contained CASC3. This
71 observation is in line with the predominantly cytoplasmic localization of CASC3, but would argue
72 against a nuclear function.

73 Another aspect under debate is the involvement of CASC3 in the NMD pathway. According to the
74 EJC-dependent model of NMD, an EJC present more than 50-55 nucleotides downstream of a
75 premature termination codon (PTC) triggers degradation by the NMD machinery (23). This quality
76 control mechanism rids the cells of aberrant transcripts that contain PTCs due to mutations or mis-
77 splicing. Additionally, it serves as a post-transcriptional mechanism of gene expression, especially
78 when coupled to alternative splicing (24). This is a common feature of many genes coding for mRBPs,
79 e.g. most SR proteins (25,26). The EJC triggers NMD by interacting with members of the SURF complex
80 resulting in phosphorylation of the central NMD factor UPF1 (27). Phosphorylated UPF1 then
81 stimulates two distinct degradation pathways of NMD: The SMG5/7-dependent pathway results in
82 deadenylation and decapping of the transcript followed by exonucleolytic decay from the 5' end by
83 XRN1 and the 3' end by the exosome (28). Alternatively, the transcript can be cleaved in the vicinity of
84 the PTC by the endonuclease SMG6 which results in two mRNA fragments that can be
85 exonucleolytically degraded by XRN1 and the exosome (29,30). While these pathways can act
86 redundantly and in principle compensate for each other, SMG6-dependent endonucleolytic cleavage
87 (endocleavage) has been shown to be the dominant pathways for NMD in human cells (31-33). In cells
88 depleted of CASC3 a stabilizing effect on PTC-containing reporter mRNAs and selected endogenous
89 targets was reported (12,17,34). Furthermore, tethering CASC3 to an mRNA results in UPF1-dependent
90 degradation of the transcript (12). However, a recent report has challenged these observations and
91 showed that CASC3 plays a minor role in NMD and only for certain endogenous targets in contrast to
92 EIF4A3 or RNPS1 (22).

93 We were intrigued by the contrasting reports about the enigmatic role of CASC3 and decided to
94 investigate the function of CASC3 and its distinction to the other EJC core components in more detail.
95 For this purpose, we established HEK 293 CASC3 knockout (KO) cells using CRISPR-Cas9-mediated gene
96 editing. The CASC3 KO cell lines are largely unchanged in their composition of the EJC core and
97 peripheral interacting proteins. However, RNA-sequencing reveals an upregulation of transcript
98 variants containing premature-termination codons (PTC) as well as the differential expression of many

99 known NMD targets, indicating a perturbation of this decay pathway. Mechanistically, CASC3
100 stimulates SMG6-dependent turnover of NMD targets and likely acts as a link from the EJC to the NMD
101 machinery. On the basis of these results we propose a revised model of EJC-dependent NMD in human
102 cells.

103 # *Materials and Methods*

104 ## *Cell culture*

105 Flp-In 293 T-REx cells (Thermo Fisher Scientific) were maintained at 37°C, 5% CO_2 and
106 90% humidity in Dublecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific) supplemented
107 with 9% fetal bovine serum (FBS) and Penicillin-Streptomycin (both Thermo Fisher Scientific).
108 Tethering experiments were performed in HeLa Tet-Off cells (Clontech) cultured in the same
109 conditions.

110 ## *siRNA-mediated knockdowns*

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

120 ## *Generation of knockout cells using CRISPR-Cas9*

121 The knockouts were performed using the Alt-R CRISPR-Cas9 system (IDT) and reverse transfection
122 of a Cas9:guideRNA ribonucleoprotein complex using Lipofactamine RNAiMAX (Thermo Fisher
123 Scientific) according to the manufacturer's protocol. The crRNA sequences to target CASC3 were
124 /AlTR1/rGrCrGrCrGrCrUrUrCrGrCrArArGrArCrArCrCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/
125 (clone H) and
126 /AlTR1/rGrUrUrCrGrGrCrCrUrCrCrGrCrGrCrUrGrUrGrArGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/

127 (clones F and T). Reverse transfection was performed on $1.5x10^5$ cells per crRNA in 12-well dishes.
128 48 hours after transfection the cells were trypsinized, counted and seeded at a density of a single cell
129 per well in 96-well plates. Cell colonies originating from a single clone were then validated by Sanger
130 sequencing of the targeted genomic DNA locus and western blotting.

131 *Plasmid transfection*

132 All used plasmids are listed in Supplementary Table S1. To express FLAG-tagged protein constructs
133 and the reporter mRNAs detected by northern blotting, the cells were stably transfected using the Flp-
134 In T-REx system and the tetracycline inducible pcDNA5/FRT/TO vector (Thermo Fisher Scientific). The
135 constructs TPI-WT, TPI-PTC, β-globin WT and β-globin PTC are available on Addgene (IDs 108375-
136 108378). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells. Per well, 1 µg of reporter
137 construct was transfected together with 1 µg of the Flp recombinase expressing plasmid pOG44 using
138 the calcium phosphate method. 48 h after transfection, the cells were transferred into 10 cm dishes
139 and selected with 100 µg/ml hygromycin. After 10 days, the colonies were pooled. Expression of the
140 reporter mRNA was induced with 1 µg/ml doxycycline for 24 h.

141 Constructs that express V5-tagged and MS2V5-tagged proteins were stably integrated into the
142 cells using the PiggyBac (PB) Transposon system and the cumate-inducible PB-CuO-MCS-IRES-GFP-EF1-
143 CymR-Puro vector (System Biosciences). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells.
144 2.5 µg of the PB Transposon vector and 0.8 µg of PB Transposase were transfected per well using the
145 calcium phosphate method. After 48 h, the cells were pooled in 10 cm dishes and positive clones
146 selected with 2 µg/ml puromycin for a week. Expression of proteins was induced using 30 µg/ml
147 cumate for 72 h.

148 The tethering construct pSBtet-Hyg-TPI-4MS2-SMG5-4H was stably integrated into HeLa Tet-Off
149 cells using the Sleeping Beauty (SB) transposon system (35,36). pSBtet-Hyg was a gift from Eric Kowarz
150 (Addgene plasmid #60508; http://n2t.net/addgene:60508; RRID:Addgene_60508). pCMV(CAT)T7-
151 SB100 was a gift from Zsuzsanna Izsvak (Addgene plasmid #34879; http://n2t.net/addgene:34879;

152 RRID:Addgene_34879). 2.5x10^5 cells were seeded 24 h before transfection in 6-wells. Per well, 1 µg of
153 the reporter construct was transfected together with 1.5 µg of the SB Transposase using the calcium
154 phosphate method. 48 h after transfection, the cells were transferred into 10 cm dishes and selected
155 with 100 µg/ml hygromycin. After 10 days, the colonies were pooled. In absence of tetracycline the
156 reporter was constitutively expressed.

157 *RNA-Sequencing and computational analyses*

158 RNA-Seq analysis was carried out with 293 wild type (WT) cells transfected with Luciferase siRNA
159 and the CASC3 KO clones H and T transfected with either Luciferase or CASC3 siRNAs. Three biological
160 replicates were analyzed for each sample. RNA was isolated with the kit NucleoSpin RNA Plus
161 (Macherey-Nagel). The Lexogen SIRV Set1 Spike-In Control Mix (SKU: 025.03) that provides a set of
162 external RNA controls was added to the total RNA to enable performance assessment. Mix E0 was
163 added to replicate 1, mix E1 was added to replicate 2 and mix E2 to replicate 3. The Spike-Ins were not
164 used for analysis. The library preparation was performed with the TrueSeq Stranded Total RNA kit
165 (Illumina). First steps of the library preparation involve the removal of ribosomal RNA using
166 biotinylated target-specific oligos combined with Ribo-Zero Gold rRNA removal beads from 1 µg total
167 RNA input. The Ribo-Zero Gold Human/Mouse/Rat kit depletes samples of cytoplasmic and
168 mitochondrial rRNA. Following purification, the RNA is fragmented and cleaved. RNA fragments are
169 copied into first strand cDNA using reverse transcriptase and random primers, followed by second
170 strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments then have the
171 addition of a single'A' base and subsequent ligation of the adapter. The products are purified and
172 enriched with PCR to create the final cDNA library. After library validation and quantification (Agilent
173 tape station), equimolar amounts of library were pooled. The pool was quantified by using the Peqlab
174 KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and
175 sequenced on an Illumina NovaSeq6000 sequencing instrument and a PE100 protocol.

176 Read processing and alignment was performed as described previously (37). In short, adaptor
177 sequences and low quality bases were removed with Flexbar 3.0 (38). Short reads from the rRNA locus
178 were subtracted by mapping against the 45S precursor (Homo sapiens, NR_046235.1) using Bowtie2
179 (39). The remaining reads were aligned against the human genome (version 38, EnsEMBL 90 transcript
180 annotations) using the STAR read aligner (version 2.5.3a) (40).

181 To compute gene differential expression analysis, reads covering exons were counted with
182 FeatureCounts (version 1.5.1) (41) using the '—primary' and '—ignoreDup' parameters. Differential
183 gene expression analysis was performed with DESeq2 (42,43) and IWH R packages. Significance
184 thresholds were |log2FoldChange|> 1 and adjusted p-value (padj) < 0.05. Genes were designated as
185 small RNA (sRNA) host gene, if they contained other Ensembl-annotated genes of biotypes snoRNA or
186 miRNA within their genomic coordinates (44).

187 Differential splicing was detected with LeafCutter (version 0.2.7) (45) with the parameters
188 min_samples_per_intron = 2 and min_samples_per_group = 2. Significance thresholds were |deltapsi|
189 > 0.1 and adjusted p-value (p.adjust) < 0.05.

190 Transcript abundance estimates were computed with Salmon (version 0.13.1) (46) using the the -
191 validateMappings --gcBias parameters. Differential transcript usage was computed with
192 IsoformSwitchAnalyzeR (version 1.7.1) and the DEXSeq method (47-52). Significance thresholds were
193 |dIF| > 0.1 and adjusted p-value (isoform_switch_q_value) < 0.05. For the Boxplot and Kolmogorov-
194 Smirnoff test, the data were filtered only for the adjusted p-value. PTC status of transcript isoforms
195 with annotated open reading frame was determined by IsoformSwitchAnalyzeR using the 50 nt rule of
196 NMD (49,53-55). Isoforms with no annotated open reading frame in Ensembl were designated "NA" in
197 the PTC analysis.

198 The UPF1 and SMG6/7 (56) and RNPS1 (57) knockdown datasets were processed and analyzed
199 with the same programs, program versions, and scripts as the CASC3 dataset. All packages used are
200 listed in the respective analysis table (Supplementary Tables S4-6). Overlaps of data sets were

201 represented via nVenn (58), eulerr (59) and Upset plots (60). Heatmaps were generated using
202 ComplexHeatmap (61). Barcode plots were produced with barcodeplot function from the limma
203 package version 3.38.3 (50) using transcript isoform dIF as ranking statistic. To test the significance of
204 this enrichment, we used the function cameraPR, from the same package, with the use.rank parameter
205 set to TRUE (62).

206 *SILAC, co-immunoprecipitation and mass spectrometry*

207 293 WT and 293 CASC3 KO clone H cells expressing either FLAG or FLAG-EIF4A3 were labeled by
208 maintaining them for 5 passages in DMEM for SILAC medium (Thermo Fisher Scientific) supplemented
209 with FBS (Silantes), Penicillin-Streptomycin (Thermo Fisher Scientific) and the respective amino acids
210 at a final concentration of 0.798 mmol/L (Lysine) and 0.398 (Arginine). Unlabeled proline was added
211 to prevent enzymatic Arginine-to-Proline conversion. The conditions were "light" (unlabeled
212 Lysine/Arginine), "medium" (Lysine 4/Arginine 6) and "heavy" (Lysine 8/Arginine 10). A label switch
213 was performed between the three replicates according to the experimental setup listed in
214 Supplementary Table S2. 24 h before expression of the FLAG-tagged construct, the CASC3 KO clone H
215 cells were treated with siRNA against CASC3. The expression of FLAG or FLAG-EIF4A3 was induced for
216 72 h with 1 µg/ml doxycycline. The cells were lysed in buffer E with RNAse (20 mM HEPES-KOH (pH
217 7.9), 100 mM KCl, 10% glycerol, 1 mM DTT, Protease Inhibitor, 1 µg/ml RNAse A) and sonicated using
218 the Bandelin Sonopuls mini20 with 15 pulses (2.5 mm tip in 600 µl volume, 1s, 50% amplitude). 600 µl
219 of a 1.6 mg/ml total protein lysate were incubated with 30 µl Anti-FLAG M2 magnetic beads (Sigma) at
220 4°C while rotating for 2 h. The beads were washed three times for 5 min with EJC-buffer (20 mM HEPES-
221 KOH (pH 7.9), 137 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-100, 0.1% NP-40) and eluted in 43 µl of a 200
222 mg/ml dilution of FLAG peptides (Sigma) in 1x TBS. The samples were merged according to
223 Supplementary Table S2. 1 volume of 10% SDS was added and the samples were reduced with DTT and
224 alkylated with CAA (final concentrations 5 mM and 40 mM, respectively). Tryptic protein digestion was
225 performed using a modified version of the single pot solid phase-enhanced sample preparation (SP3)
226 (63). In brief, reduced and alkylated proteins were supplemented with paramagnetic Sera-Mag speed

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

235 Proteomics analysis was performed by data-dependent acquisition using an Easy nLC1200 ultra
236 high-performance liquid chromatography (UHPLC) system coupled via nanoelectrospray ionization to
237 a Q Exactive Plus instrument (all Thermo Scientific). Tryptic peptides were separated based on their
238 hydrophobicity using a chromatographic gradient of 60 min with a binary system of buffer A (0.1%
239 formic acid) and buffer B (80% ACN, 0.1% formic acid). In-house made analytical columns (length: 50
240 cm, inner diameter: 75 µm) filled with 1.9 µm C18-AQ Reprosil Pur beads (Dr. Maisch) were used for
241 separation. Buffer B was linearly increased from 3% to 27% over 41 min followed by a steeper increase
242 to 50% within 8 min. Finally, buffer B was increased to 95% within 1 min and stayed at 95% for 10 min
243 in order to wash the analytical column. Full MS spectra (300 – 1,750 m/z) were acquired with a
244 resolution of 70,000, a maximum injection time of 20 ms, and an AGC target of 3e6. The top 10 most
245 abundant peptide ions of each full MS spectrum were selected for HCD fragmentation (NCE: 27) with
246 an isolation width of 1.8 m/z and a dynamic exclusion of 10 seconds. MS/MS spectra were measured
247 with a resolution of 35,000, a maximum injection time of 110 ms and an AGC target of 5e5.

248 MS RAW files were analysed using the standard settings of the MaxQuant suite (version 1.5.3.8)
249 with the before mentioned SILAC labels (64). Peptides were identified by matching against the human
250 UniProt database using the Andromeda scoring algorithm (65). Carbamidomethylation of cysteine was
251 set as a fixed modification, methionine oxidation and N-terminal acetylation as variable modification.

252 Trypsin/P was selected as the digestion protein. A false discovery Rate (FDR) < 0.01 was used for
253 identification of peptide-spectrum matches and protein quantification. Data processing and statistical
254 analysis was done in the Perseus software (version 1.5.5.3) (66). Significantly changed proteins were
255 identified by One-sample t-testing (H0 = 0, fudge factor S0 = 0.1). The results are listed in
256 Supplementary Table S2. Visualization was performed with the Instant Clue software (version 0.5.3)
257 and the R package ggplot2 (version 3.1.0) (67,68).

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

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

262 RNA was extracted using peqGOLD TriFast reagent (VWR) according to the manufacturer's
263 instructions. Reverse transcription was performed with GoScript Reverse Transcriptase (Promega)
264 using 2 µg total RNA and oligo dT primers. Semi-quantitative PCR was carried out with MyTaq Red Mix
265 (Bioline). Quantitative real time PCR was performed with 16 ng of cDNA per reaction with GoTaq qPCR
266 Master Mix (Promega) and the CFX96 Touch Real-Time PCR Detection System (Biorad). The average cT
267 values were calculated from three technical replicates. The mean fold changes from three biological
268 replicates were calculated according to the ΔΔCt method (69). When measuring isoform switches, the
269 fold change of the PTC-containing transcript was normalized to the canonical transcript. When
270 measuring differential expression, the fold change was normalized to GAPDH. For each primer pair
271 amplification efficiencies were measured by a 2-fold dilution curve and ranged between 87 and
272 100.1%. The primer sequences are listed in Supplementary Table S3.

273 *Western blotting*

274 Protein extraction was performed with peqGOLD TriFast reagent (VWR), and proteins were
275 separated by SDS-PAGE gel electrophoresis and transferred to a PVDF membrane (GE Healthcare Life
276 Sciences). The following antibodies were used: anti-CASC3 amino acid residues 653-703 (Bethyl

277 Laboratories, #A302-472A-M), anti-CASC3 amino acid residues 367-470 (Atlas Antibodies,
278 #HPA024592), anti-EIF4A3 (Genscript), anti-FLAG (Cell Signaling Technology, #14793), anti-RBM8A
279 (Atlas Antibodies, #HPA018403), anti-SMG6 (Abcam, #ab87539), anti-SMG7 (Elabscience, #E-AB-
280 32926), anti-Tubulin (Sigma-Aldrich, #T6074), anti-V5 (QED Bioscience, #18870), anti-XRN1 (Bethyl
281 Laboratories, #A300-443A), anti-rabbit-HRP (Jackson ImmunoResearch, #111-035-006), anti-mouse-
282 HRP (Jackson ImmunoResearch, #115-035-003). Detection was performed with Western Lightning
283 Plus-ECL (PerkinElmer) or Amersham ECL prime (Ge Healthcare Life Sciences) and the
284 chemiluminescence imager Fusion FX6 EDGE (Vilber-Lourmat).

285 *Northern blotting*

286 The cells were harvested in peqGOLD TriFast reagent (VWR) and total RNA extraction was
287 performed as recommended by the manufacturer's protocol. 2.5 µg of total RNA were resolved on a
288 1% agarose/0.4 M formaldehyde gel using the tricine/triethanolamine buffer system (70) followed by
289 a transfer on a nylon membrane (Roth) in 10x SSC. The blots were incubated overnight at 65°C in
290 Church buffer containing [α-32P]-GTP body-labeled RNA probes for detection of the reporter mRNA.
291 Endogenous 7SL RNA was detected by a 5'-32P-labeled oligonucleotide (5'-
292 TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT-3'). The blots were visualized and quantified using a
293 Typhoon FLA 7000 (GE Healthcare) and ImageQuant TL 1D software.

294 # *Results*

295 ## *CASC3 is dispensable for the correct splicing of many EJC-dependent transcripts*

296 Previously, we and others have shown that the depletion of EJC core components in human
297 cells leads to pervasive re-splicing of cryptic splice sites, resulting in aberrant splice variants lacking
298 exonic sequences (37,71). Mechanistically, the EJC prevents the use of cryptic splice sites either by
299 interaction with the ASAP/PSAP component RNPS1 or by sterically masking splice sites
300 (schematically depicted in Supplementary Figure S1A) (37). In keeping with previous observations
301 from HeLa cells, the knockdowns of the EJC core components EIF4A3 or RBM8A in HEK293 cells
302 resulted in exon-skipping of the mRNAs for RER1, OCIAD1 and MRPL3 (Figure 1A-C). Surprisingly,
303 when we performed a knockdown of the EJC core factor CASC3 this did not result in mis-splicing
304 of these three selected transcripts (Figure 1A-C). This observation stands in contrast to a previous
305 study that has reported transcriptome-wide alternative splicing upon CASC3 depletion, including
306 MRPL3 (71,72). Interestingly, the knockdown of CASC3 increased the abundance of an alternative
307 transcript isoform of the SR protein SRSF2 (Figure 1D), which was previously shown to be regulated
308 by NMD. Although this altered abundance of the NMD-targeted SRSF2 transcript isoform could be
309 due to alternative splicing, we consider it very likely that reduced NMD is responsible for this
310 effect.

311 Although the knockdown efficiency of CASC3 was substantial (Supplementary Figure S1B),
312 we wished to exclude that residual amounts of CASC3 prevented a reliable assessment of the
313 protein's function. Since CASC3 was found to be non-essential in multiple genome-wide screens of
314 human immortalized cell lines, we reasoned that knockout (KO) of CASC3 should be feasible
315 (73,74). Accordingly, we obtained three cell lines by CRISPR-Cas9-mediated gene editing,
316 designated H, F, and T lacking the CASC3-specific 130 kDa band on a western blot (Figure 1E). In all
317 cell lines we detected genomic insertions of different length and sequence at the beginning of the
318 coding region of CASC3, which resulted in frame shifts of the downstream coding region or, in the

319 case of cell line T, contained in-frame termination codons (Figure 1F, Supplementary Figure S1C
320 and D). For the cell lines H and F, we observed an additional band of 100 kDa on western blots with
321 antibodies recognizing the C-terminal or central region of CASC3 (Figure 1E, red arrow,
322 Supplementary Figure S1C and E). This cross-reactive protein interacted with FLAG-tagged EIF4A3
323 and disappeared upon treatment with siRNAs against CASC3 (Figure 1G, Supplementary Figure S1E
324 and F). This suggests that the cell lines H and F produce an N-terminally truncated form of CASC3,
325 presumably by initiating protein translation at a non-canonical initiation codon. The production of
326 aberrant protein forms by alternative translation initiation has been recently described in a
327 systematic analysis to commonly occur in KO cell generated by CRISPR-Cas9 genome editing
328 (75,76). The cell line T (without further treatment) and cell line H in combination with CASC3 siRNA
329 treatment completely lack detectable CASC3 protein. This set of cell lines therefore enables a
330 hitherto unfeasible analysis of CASC3's cellular function.

331 In agreement with the data obtained from the knockdown of CASC3, transcripts containing
332 EJC-dependent splice sites were correctly spliced in CASC3 KO cells (Figure 1H and I, Supplementary
333 Figure S1G and H). Remarkably, the increased abundance of the NMD-regulated SRSF2 transcript
334 isoform was much more prominent in the CASC3 KO than in the CASC3 knockdown (Figure 1J).

335 *CASC3 regulates NMD-sensitive isoforms*

336 So far, our results have indicated that CASC3 shapes the transcriptome differently
337 compared to EIF4A3, MAGOH and RBM8A. To investigate the global effects of CASC3 depletion on
338 the transcriptome, we performed RNA-sequencing (RNA-seq) of the cell lines H and T either
339 treated with CASC3 or control siRNAs (Figure 2A) and identified differentially expressed genes
340 (Supplementary Figure S2A and B, Supplementary Table S4). To exclude clone-specific and siRNA
341 treatment-related effects, we compared the identified targets between the four conditions.
342 Overall, the high number and the substantial overlap of upregulated genes suggests that CASC3
343 KO mainly results in the accumulation of certain transcripts (Figure 2B and C, Supplementary Figure

344 S2C). Interestingly, several upregulated genes belong to the class of small RNA (e.g. snoRNA) host
345 genes, which are frequently NMD targets (Figure 2C and D) (31,33). We validated the upregulation
346 of the snoRNA host gene ZFAS1 by qPCR, which was even more pronounced in CASC3 KO than
347 UPF1 knockdown cells (Supplementary Figure S2D and E). Across the top 100 significantly
348 upregulated genes in CASC3 KO, 15 small RNA host genes were identified (Figure 2D). Comparing
349 the differentially expressed genes to recent transcriptome-wide NMD screens, many of the top
350 100 significantly upregulated genes were also differentially expressed in UPF1 and SMG6/7 KD
351 (Figure 2D, 23% and 59%, respectively) (31,56). However, none of the identified targets were
352 present in an RNPS1 knockdown (Figure 2D) (37,57). Collectively, our differential gene expression
353 analysis strengthens the proposed link between CASC3 and the NMD-machinery.

354 Next, we analyzed alternative splicing changes in CASC3 KO cells (Figure 2E). Since our earlier
355 assays showed that CASC3 was not involved in the EJC-regulated splicing of many targets (Figure
356 1), we were surprised to detect many altered splicing events in CASC3 depleted cells
357 (Supplementary Figure S2F, Supplementary Table S5). It is remarkable that hardly any alternative
358 splicing events were shared between RNPS1 knockdown and CASC3 KO cells (Figure 2E,
359 Supplementary Figure S2F). Either CASC3 regulates an RNPS1-independent set of alternative splice
360 sites or the splicing changes are due to impaired NMD, which fails to remove NMD-sensitive
361 isoforms. To test these possibilities, we investigated the functional consequence of CASC3-
362 dependent alternative splicing on the transcript isoform level (Figure 3A, Supplementary Figure
363 S3A and B, and Supplementary Table S6). Strikingly, in all CASC3 KO conditions many upregulated
364 mRNA isoforms contained a premature termination codon (PTC), rendering the transcripts
365 susceptible to NMD (Figure 3A). On the other hand, downregulated isoforms rarely contained a
366 PTC. Among the identified isoform switches was the target SRSF2, which we confirmed earlier to
367 be CASC3-dependent (Figure 1D and J). While overall SRSF2 gene expression varied only slightly
368 between wild-type and CASC3 KO cells, the isoform usage changed dramatically towards the
369 accumulation of NMD-sensitive transcripts in the CASC3 KO conditions (Figure 3B). A similar

370 accumulation of NMD-sensitive transcripts was also observed for other transcripts (Supplementary
371 Figure S3C-E).

372 We next validated a set of transcript isoform switches by qPCR (Figure 3C-E, Supplementary Figure
373 S3F and G). In the transcript isoforms stabilized by the CASC3 KO and by a UPF1 KD, the inclusion
374 of intronic regions resulted in the inclusion of PTCs. The shift of isoform usage from NMD-
375 insensitive to PTC-containing transcripts was also observed transcriptome-wide in CASC3 KO cells
376 and was comparable to NMD-compromised SMG6/7 or UPF1 depleted cells (Figure 3F). When
377 compared to transcript isoform changes in a SMG6/7 knockdown, the events in the CASC3 KO are
378 enriched in a similar fashion (Figure 3G). While the enrichment is stronger in a comparison of UPF1
379 and SMG6/7, the isoform changes occurring in the RNPS1 KD do not correlate with the ranked
380 SMG6/7 events (Figure 3G, Supplementary Figure S3H). These findings indicate that many
381 transcript isoforms upregulated upon depletion of CASC3 represent genuine endogenous NMD
382 targets.

383 *The EJC core is undisturbed if CASC3 is not present*

384 The CASC3 KO could potentially influence the composition of exon junction complexes and
385 their peripheral interacting proteins, which could underlie the observed effects on the
386 transcriptome. Therefore, we analyzed the FLAG-tagged EIF4A3 interactome in the cell line H
387 treated with CASC3 siRNAs and wild type cells using mass spectrometry. EIF4A3 was successfully
388 enriched, together with other known EJC complex members (Figure 4A and B, Supplementary
389 Figure S4). Co-precipitated CASC3 was reduced to background levels in the knockout condition,
390 further validating the absence of CASC3 in the EJC (Figure 4B and C, log2 fold change = -4.34). We
391 were interested to identify factors that significantly changed between the KO and WT condition.
392 These belong to three distinct groups: significant in both WT/CTL and KO/CTL conditions (blue
393 dots), significant in KO/CTL (orange dots), and significant in WT/CTL (magenta dots). In the first
394 group, no factor was changed between the WT and KO condition more than 1.6 fold

395 (Supplementary Figure S4). In the second group three factors were negatively enriched in the
396 KO/CTL condition or marginally altered between KO and WT conditions (Figure 4B and C, WARS,
397 CMBL and USP15).

398 The only other protein besides CASC3 that was significantly changed between KO/WT and only
399 enriched in the WT/CTL condition was the NMD factor UPF3B (Figure 4C, log2 fold change = -1.52).
400 UPF3B links the EJC to the NMD machinery via direct interactions (77) and was recently found to
401 be enriched in cytoplasmic CASC3-loaded EJCs (22). The reduction of NMD-competent EJCs could
402 contribute to the NMD impairment that we observed upon loss of CASC3. Strikingly, no other EJC
403 core factor or splicing regulatory EJC component (e.g. ASAP/PSAP) was considerably altered in the
404 CASC3-depleted condition (Figure 4D).

405 *CASC3 stimulates SMG6-dependent endonucleolytic cleavage*

406 To deepen the understanding of how a lack of CASC3 results in reduced NMD efficiency, we
407 stably integrated the well-established globin NMD reporter PTC39 in WT and CASC3 KO cell lines
408 (Figure 5A and B, Supplementary Figure S5A) (78). The analysis of a reporter mRNA enables a read-
409 out of multiple aspects of mRNA degradation: firstly, the total levels of the full-length reporter,
410 secondly the contribution of 5'->3' exonucleolytic decay by XRN1 (detection of xrFrag due to an
411 XRN1-resistant element) (79,80); and thirdly the amount of endonucleolytic cleavage by SMG6
412 (detection of 3' fragment stabilized by XRN1 knockdown). In both WT and CASC3 KO cell lines the
413 reporter was efficiently degraded, showing that the NMD pathway is still functional in CASC3
414 depleted cells. However, full-length reporter levels in CASC3 KO were slightly higher when
415 compared to wild-type cells (lane 2 vs. lane 5). Notably, the accumulation of 3' fragments following
416 XRN1 knockdown was clearly reduced in the CASC3 KO condition (lane 3 vs. lane 6). This difference
417 in endonucleolytic cleavage efficiency was also observed when expressing a minigene reporter of
418 the endogenous CASC3 target TOE1. While there was a substantial upregulation of full-length
419 reporter mRNA in CASC3 KO cells, the amount of the 3' fragment was strongly reduced, suggesting

420 that SMG6-dependent endonucleolytic cleavage is inefficient in CASC3 KO cells (Figure 5C-E,
421 Supplementary Figure S5B and C). To further address this, a TPI reporter was expressed in
422 combination with knockdowns of XRN1, SMG6 and/or SMG7 (Figure 5F and G, Supplementary
423 Figure S5D). The degree of reporter and 3' fragment stabilization of the TPI reporter following XRN1
424 knockdown was comparable to the observations made for the globin and TOE1 reporters (lanes 1-
425 3 vs. lanes 6-8). In both WT and CASC3 KO cells, the SMG6 knockdown resulted in a drastic
426 reduction of 3' fragments, as expected (lanes 4 and 9). Notably, a knockdown of SMG7 together
427 with XRN1 revealed a major difference between the cell lines. In WT cells the PTC-containing
428 reporter was only minimally stabilized by the SMG7/XRN1 knockdown and 3' fragments were
429 unaffected. Performing a SMG7/XRN1 knockdown in CASC3 KO cells lead to a more dramatic
430 stabilization of the full-length reporter and a decrease of 3' fragments compared to the XRN1
431 knockdown condition (lanes 3 and 5 vs. lanes 8 and 10). Collectively, our results indicate that in
432 CASC3 KO cells SMG6-mediated endonucleolytic cleavage is impaired. This could explain why the
433 CASC3 KO cells are more sensitive to a knockdown of SMG7 when compared to wild type cells.

434 To identify, which part of CASC3 promotes NMD, we employed a tethering reporter that was
435 designed to monitor mRNA turnover as well as endonucleolytic cleavage at the termination codon
436 (Figure 6A). Tethering the full-length CASC3 protein to the MS2 stem loops downstream of the stop
437 codon resulted in degradation of the reporter compared to tethering of the negative control GST,
438 as was previously reported for similar tethering reporters (Figure 6B and C, Supplementary Figure
439 S6) (12,32). This degradation was accompanied by the production of 3' fragments in XRN1
440 knockdown conditions, indicating that the mechanism of decay is comparable to the PTC-
441 containing reporter mRNAs (Figure 6B, lane 6). Surprisingly, C-terminally truncated deletion
442 mutants of CASC3 that contain the first 480 or even 137 amino acid residues were able to induce
443 degradation of the tethering reporter to a comparable extent as full-length CASC3 (Figure 6B, lanes
444 3 and 4, 7 and 8).

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

456 ## *Discussion*

457 The role of CASC3 within the EJC has been the subject of scientific controversy for many years.
458 CASC3 has been initially described as an EJC core protein, because it was required for the assembly of
459 the EJC from recombinant protein components *in vitro* (21). However, it has been demonstrated that
460 the mechanism of EJC assembly using recombinant proteins is mechanistically different from EJC
461 assembly in splicing extracts or in living cell (11). Furthermore, several recent publications challenged
462 the view of CASC3 being an EJC core component. For instance, CASC3 was reported to be present in
463 substoichiometric amounts compared to the other three EJC core proteins EIF4A3, RBM8A, and
464 MAGOH in HEK293 (7) and U2OS cells (81). Also, during mouse embryonic brain development CASC3
465 deficiency results in a different phenotype than the other EJC core components (82). By using CASC3
466 CRISPR-Cas9 knockout cells, we unambiguously establish that CASC3 is not required for EJC assembly
467 or EJC-regulated splicing in the nucleus (Figure 7). Even previously reported CASC3-dependent
468 alternative splice events (71) were not detectable in any of our KO or KD conditions. Therefore, our
469 molecular analyses fully support the recently emerging view of defining CASC3 as a peripheral EJC
470 component. As a mainly cytoplasmic component of the EJC we propose that the principal role of CASC3
471 is to alter the efficiency by which NMD-sensitive transcript isoforms are degraded.

472 Although NMD has been extensively studied in the past decades and many NMD factors have been
473 identified and characterized, no universal model exists that describes how they work together to elicit
474 NMD. While a function of CASC3 in NMD has been reported before, previous analyses did not show
475 consistent results, ranging from a substantial contribution of CASC3 to only a minor role in NMD
476 (12,17,22,34). In addition, none of the previous publications performed transcriptome-wide analyses
477 but concentrated on reporter mRNAs of only a few selected endogenous NMD targets. We reasoned
478 that the inconsistent results in the literature may be influenced by variable CASC3 knockdown
479 efficiency. By generating CASC3 knockout cell lines, we can for the first time analyze the global effects
480 of a complete depletion of CASC3 on the transcriptome and can exclude that residual CASC3 protein
481 masks these effects. Interestingly, the CASC3 knockout cells appear phenotypically normal, unlike

482 when the EJC components EIF4A3, RBM8A, or MAGOH are depleted. Nonetheless, since CASC3 is
483 required for mouse embryogenesis and involved in the transport of mRNAs in *D. melanogaster*, it is
484 likely that CASC3 downregulation in highly specialized cell types such as neurons or in developing
485 tissues would result in a more severe phenotype compared to HEK293 cells.

486 In recent years, high-throughput RNA-sequencing became an increasingly important method for
487 the analysis of NMD. Several RNA-Seq datasets of cells with NMD-factor knockdowns have been
488 generated and analyzed (31,33,83-85). However, these datasets were obtained in different cell lines,
489 with different amounts of replicates and due to the rapid developments of next-generation
490 sequencing, not using the same technologies. Furthermore, batch effects and divergent approaches of
491 data analyses may contribute to the fact that only a minor overlap of NMD targets could be established
492 so far (31). We compared the results of our CASC3 KO RNA-sequencings to the most recent and
493 comprehensive NMD factor analysis performed by Colombo *et al.* (31). The differential expression
494 analysis revealed that many of the top upregulated genes in the CASC3 KO datasets are also
495 significantly affected by UPF1 or SMG6/7 knockdowns (31) and/or encode for small RNA (sRNA) host
496 genes, a previously described class of NMD targets (33).

497 We detected many alternative splicing events in the CASC3 KO data, which was unexpected given
498 that CASC3 was apparently dispensable for the nuclear EJC-related functions. However, we could
499 attribute these splicing patterns to dysfunctional NMD, since isoform-specific algorithms revealed that
500 predominantly PTC-containing transcripts accumulated. A comprehensive bioinformatics analysis
501 workflow and a systematic approach to detect affected transcripts under NMD factor
502 knockdown/knockout conditions could therefore be a crucial step to paint a complete picture of the
503 regulation of transcripts by NMD in the future. Our initial screen of the KO cells showed that compared
504 to the siRNA-mediated CASC3 KD in WT cells, the genomic CASC3 KO cells demonstrated a more
505 pronounced NMD inhibition. It is therefore important to reduce the amount of residual CASC3 protein
506 as much as possible to obtain consistent and robust effects.

507 How exactly CASC3 activates NMD when bound to an EJC, is not yet fully understood. Previously,
508 we reported that the presence of EJCs in the 3' UTR enhances endonucleolytic cleavage (32). In line
509 with the proposed role as a peripheral NMD-activating EJC component, we observed that CASC3
510 stimulates SMG6-dependent endonucleolytic cleavage, thereby promoting the degradation of NMD-
511 targeted transcripts. This effect can be recapitulated by tethering full-length CASC3, its N-terminal two
512 thirds (1-480) or just its N-terminal 137 amino acids to a reporter mRNA. How the small N-terminal
513 region, which cannot assemble into the EJC or contains any known protein domains or sequence motifs
514 can elicit NMD remains to be determined. It is also unclear, if the N-terminus activates translation-
515 dependent degradation, as it was previously shown for the full length CASC3 (79). Since the N-terminal
516 segment of CASC3 is a region of low-complexity it could hypothetically undergo liquid-liquid phase
517 separation (LLPS) and be present in condensates with mRNA decay factors, such as processing bodies
518 (P-bodies). In agreement with this idea CASC3 was shown to localize to cytoplasmic granules when
519 overexpressed (86).

520 Our data suggest that CASC3 activates NMD by potentially redundant mechanisms. Binding of
521 CASC3 to the EJC could have an indirect effect on NMD stimulation by increasing the stability of the
522 bound EJC and thus maintaining the possibility of efficient endonucleolytic cleavage of the transcript.
523 An indication for this role comes from the initial *in vitro* observation that CASC3 stabilizes recombinant
524 EJCs (21). Additionally, the moderately reduced pull-down of UPF3B with EIF4A3 could indicate that
525 cytoplasmic NMD-competent EJCs are less stable in CASC3 KO cells. Alternatively, CASC3 may directly
526 contribute to the recruitment of NMD factors. We therefore propose that CASC3 potentially in
527 conjunction with UPF3B links the EJC with the NMD machinery. In particular, CASC3 influences the
528 contribution of SMG6-mediated endonucleolytic and SMG7-dependent exonucleolytic decay
529 pathways to the overall degradation efficiency of NMD. Accordingly, in wild type cells a knockdown of
530 SMG7 only had a marginal effect on the abundance of the analyzed NMD reporter mRNA, whereas it
531 clearly impaired NMD in CASC3 KO cells. Also, the amount of endonucleolytic cleavage-derived 3'
532 fragments was reduced when CASC3 is depleted, mirroring the SMG6-knockdown condition.

533 By integrating CASC3 as a specific NMD-activating factor we can now postulate a modified model
534 of EJC-dependent NMD, which is also compatible with several molecular properties of the EJC (Figure
535 7). Since CASC3 is only present in modest amounts in the cytoplasm, it will probably not immediately
536 associate with all EJCs on recently exported mRNPs. This would also not be necessary, since most EJCs
537 are located in the coding sequence and will therefore be removed by the first translating ribosome.
538 However, mRNAs containing PTCs will carry one or more EJCs in their 3' UTR, which are available for
539 binding of CASC3. The first few translating ribosomes may terminate upstream of CASC3-free EJCs,
540 which could preferentially trigger SMG7-dependent exonucleolytic degradation. Previously, NMD has
541 been proposed to occur primarily in the pioneering round of translation when newly synthesized
542 transcripts are bound to the cap-binding complex (87,88). This model has been challenged and there
543 is evidence that NMD can occur on already translating mRNAs and possibly with a constant probability
544 during every round of translation (89-91). Thus, CASC3 could bind to the EJC at a later time point and
545 then increase the probability to activate SMG6-mediated endonucleolytic degradation after each
546 round of termination. Important molecular targets of CASC3 may be mRNAs that escape initial NMD
547 activation, despite containing a PTC (89,92). CASC3 may help to reduce the amount of these mRNAs
548 by maintaining the NMD-activating function of the EJC, either by increasing its stability on the mRNA
549 or via direct interactions with the NMD machinery. This concept would be consistent with the recent
550 observation that NMD targets undergo several rounds of translation before endonucleolytic cleavage
551 occurs (89).

552 In summary, our data paint a picture, in which CASC3 has no essential EJC-related function in the
553 nucleus, but helps to sustain the EJC's ability to induce NMD. We do not exclude the possibility that
554 CASC3 is already associated with the EJC in the nucleus. However, our model of delayed binding of
555 CASC3 to the EJC in the cytoplasm would explain why only a small amount of CASC3 is sufficient to
556 activate EJC-dependent NMD. In this model, CASC3 is an indispensable cytoplasmic component of the
557 EJC that helps to degrade mRNAs that failed to unload all their bound EJCs during the initial rounds of

558 translation. Thus, the binding of CASC3 to the EJC could signal the final round(s) of translation of an
559 mRNA.

560 *Data Availability*

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

563 • RNA-seq data have been deposited in the ArrayExpress database (93) at EMBL-EBI under
564 accession number E-MTAB-8461 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-
565 8461).
566 • The mass spectrometry proteomics data have been deposited to the ProteomeXchange
567 Consortium via the PRIDE (94) partner repository with the dataset identifier PXD015754
568 (https://www.ebi.ac.uk/pride/archive/projects/PXD015754).

569 *Author Contributions*

570 Conceptualization, N.H.G., J.V.G. and V.B.;
571 Methodology, N.H.G., V.B., J.V.G., and C.K.F.;
572 Software, T.B.B., V.B., J.L.W., S.K. and C.D.;
573 Investigation, J.V.G, V.B., J.L.W., S.K., D.U.A. and S.C.;
574 Resources and Data Curation, T.B.B., J.L.W., S.K., J.A. and C.D.;
575 Writing – Original Draft, Review & Editing, N.H.G., J.V.G. and V.B.;
576 Visualization, J.V.G., V.B. and T.B.B.;
577 Supervision, N.H.G., C.D. and M.K.;
578 Funding Acquisition, N.H.G. and C.D.

579 *Funding*

580 This work was supported by grants from the Deutsche Forschungsgemeinschaft to C.D. (DI 1501/8-
581 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
582 Institutional Strategy of the University of Cologne within the German Excellence Initiative. N.H.G.
583 acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1) from the Deutsche
584 Forschungsgemeinschaft. C.D. and T.B.B. were kindly supported by the Klaus Tschira Stiftung gGmbH
585 (00.219.2013). This work was supported by the DFG Research Infrastructure as part of the Next
586 Generation Sequencing Competence Network (project 423957469). NGS analyses were carried out at
587 the production site WGGC Cologne.

588 *Acknowledgements*

589 We thank members of the Gehring lab for discussions and reading of the manuscript. We also
590 thank Marek Franitza and Christian Becker (Cologne Center for Genomics, CCG) for preparing the
591 sequencing libraries and operating the sequencer. We acknowledge Tobias Jakobi for helping with
592 infrastructure support.

593 *Conflict of Interest*

594 The authors declare no competing interests.

595 *References*

596 1. Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding
597 proteins. *Nat Rev Mol Cell Biol*, **19**, 327-341.

598 2. Boehm, V. and Gehring, N.H. (2016) Exon Junction Complexes: Supervising the Gene
599 Expression Assembly Line. *Trends Genet*, **32**, 724-735.

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

602 4. Woodward, L.A., Mabin, J.W., Gangras, P. and Singh, G. (2017) The exon junction complex: a
603 lifelong guardian of mRNA fate. *Wiley Interdiscip Rev RNA*, **8**.

604 5. Bono, F., Ebert, J., Lorentzen, E. and Conti, E. (2006) The crystal structure of the exon junction
605 complex reveals how it maintains a stable grip on mRNA. *Cell*, **126**, 713-725.

606 6. Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen,
607 J.S., Seraphin, B., Le Hir, H. and Andersen, G.R. (2006) Structure of the exon junction core
608 complex with a trapped DEAD-box ATPase bound to RNA. *Science*, **313**, 1968-1972.

609 7. Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z. and Moore, M.J. (2012)
610 The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus.
611 *Cell*, **151**, 750-764.

612 8. Gerbracht, J.V. and Gehring, N.H. (2018) The exon junction complex: structural insights into a
613 faithful companion of mammalian mRNPs. *Biochem Soc Trans*, **46**, 153-161.

614 9. Alexandrov, A., Colognori, D., Shu, M.D. and Steitz, J.A. (2012) Human spliceosomal protein
615 CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-
616 mediated decay. *Proc Natl Acad Sci U S A*, **109**, 21313-21318.

617 10. Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M. and Le Hir, H.
618 (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon
619 junction complex assembly. *Nat Struct Mol Biol*, **19**, 983-990.

620 11. Steckelberg, A.L., Boehm, V., Gromadzka, A.M. and Gehring, N.H. (2012) CWC22 connects pre-
621 mRNA splicing and exon junction complex assembly. *Cell Rep*, **2**, 454-461.

622 12. Gehring, N.H., Lamprinaki, S., Hentze, M.W. and Kulozik, A.E. (2009) The hierarchy of exon-
623 junction complex assembly by the spliceosome explains key features of mammalian nonsense-
624 mediated mRNA decay. *PLoS Biol*, **7**, e1000120.

625 13. Steckelberg, A.L., Altmueller, J., Dieterich, C. and Gehring, N.H. (2015) CWC22-dependent pre-
626 mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes.
627 *Nucleic Acids Res*, **43**, 4687-4700.

628 14. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. and Luhrmann, R. (2008) Isolation of an active
629 step I spliceosome and composition of its RNP core. *Nature*, **452**, 846-850.

630 15. Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H. and
631 Luhrmann, R. (2011) Semiquantitative proteomic analysis of the human spliceosome via a
632 novel two-dimensional gel electrophoresis method. *Mol Cell Biol*, **31**, 2667-2682.

633 16. Degot, S., Regnier, C.H., Wendling, C., Chenard, M.P., Rio, M.C. and Tomasetto, C. (2002)
634 Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast
635 cancer. *Oncogene*, **21**, 4422-4434.

636 17. Palacios, I.M., Gatfield, D., St Johnston, D. and Izaurralde, E. (2004) An eIF4AIII-containing
637 complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*, **427**,
638 753-757.

639 18. Shibuya, T., Tange, T.O., Sonenberg, N. and Moore, M.J. (2004) eIF4AIII binds spliced mRNA in
640 the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol*,
641 **11**, 346-351.

642 19. Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F. and Dreyfuss, G. (2000)
643 Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that
644 persists in the cytoplasm. *Mol Cell*, **6**, 673-682.

645 20. Le Hir, H., Gatfield, D., Braun, I.C., Forler, D. and Izaurralde, E. (2001) The protein Mago
646 provides a link between splicing and mRNA localization. *EMBO Rep*, **2**, 1119-1124.

647 21. Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B. and Le Hir, H. (2005) The exon
648 junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct*
649 *Mol Biol*, **12**, 861-869.

650 22. Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R. and
651 Singh, G. (2018) The Exon Junction Complex Undergoes a Compositional Switch that Alters
652 mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep*, **25**, 2431-2446 e2437.

653 23. Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-
654 containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*, **23**, 198-199.

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

657 25. Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A.,
658 Blume, J.E. and Ares, M., Jr. (2007) Ultraconserved elements are associated with homeostatic
659 control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev*,
660 **21**, 708-718.

661 26. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) Unproductive
662 splicing of SR genes associated with highly conserved and ultraconserved DNA elements.
663 *Nature*, **446**, 926-929.

664 27. Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M.,
665 Dreyfuss, G. and Ohno, S. (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to
666 the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA
667 decay. *Genes Dev*, **20**, 355-367.

668 28. Loh, B., Jonas, S. and Izaurralde, E. (2013) The SMG5-SMG7 heterodimer directly recruits the
669 CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with
670 POP2. *Genes Dev*, **27**, 2125-2138.

671 29. Eberle, A.B., Lykke-Andersen, S., Muhlemann, O. and Jensen, T.H. (2009) SMG6 promotes
672 endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol*, **16**, 49-55.

673 30. Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J. and Izaurralde, E. (2008) SMG6 is the
674 catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*,
675 **14**, 2609-2617.

676 31. Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R. and Muhlemann, O. (2017)
677 Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive
678 redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA*, **23**, 189-201.

679 32. Boehm, V., Haberman, N., Ottens, F., Ule, J. and Gehring, N.H. (2014) 3' UTR length and
680 messenger ribonucleoprotein composition determine endocleavage efficiencies at
681 termination codons. *Cell Rep*, **9**, 555-568.

682 33. Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A. and Jensen, T.H. (2014)
683 Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA
684 host genes. *Genes Dev*, **28**, 2498-2517.

685 34. Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W. and Kulozik, A.E.
686 (2005) Exon-junction complex components specify distinct routes of nonsense-mediated
687 mRNA decay with differential cofactor requirements. *Mol Cell*, **20**, 65-75.

688 35. Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P.,
689 Schmitt, A., Becker, K., Matrai, J. *et al.* (2009) Molecular evolution of a novel hyperactive
690 Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet*,
691 **41**, 753-761.

692 36. Kowarz, E., Loscher, D. and Marschalek, R. (2015) Optimized Sleeping Beauty transposons
693 rapidly generate stable transgenic cell lines. *Biotechnol J*, **10**, 647-653.

694 37. Boehm, V., Britto-Borges, T., Steckelberg, A.L., Singh, K.K., Gerbracht, J.V., Gueney, E.,
695 Blazquez, L., Altmuller, J., Dieterich, C. and Gehring, N.H. (2018) Exon Junction Complexes
696 Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol Cell*, **72**, 482-495
697 e487.

698 38. Dodt, M., Roehr, J.T., Ahmed, R. and Dieterich, C. (2012) FLEXBAR-Flexible Barcode and
699 Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)*, **1**, 895-905.

700 39. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat*
701 *Methods*, **9**, 357-359.

702 40. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.
703 and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.

704 41. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program
705 for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.

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

708 43. Ignatiadis, N., Klaus, B., Zaugg, J.B. and Huber, W. (2016) Data-driven hypothesis weighting
709 increases detection power in genome-scale multiple testing. *Nat Methods*, **13**, 577-580.

710 44. Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins,
711 C., Gall, A., Giron, C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res*, **46**, D754-D761.

712 45. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K. and Pritchard, J.K.
713 (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet*, **50**, 151-
714 158.

715 46. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and
716 bias-aware quantification of transcript expression. *Nat Methods*, **14**, 417-419.

717 47. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq
718 data. *Genome Res*, **22**, 2008-2017.

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

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

723 50. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma
724 powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic*
725 *Acids Res*, **43**, e47.

726 51. Soneson, C., Love, M.I. and Robinson, M.D. (2015) Differential analyses for RNA-seq: transcript-
727 level estimates improve gene-level inferences. *F1000Res*, **4**, 1521.

728 52. Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential
729 expression analysis of RNA-seq data. *Genome Biol*, **11**, R25.

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

733 54. Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K.,
734 Krogh, A., Wang, J. and Porse, B.T. (2012) Mammalian tissues defective in nonsense-mediated
735 mRNA decay display highly aberrant splicing patterns. *Genome Biol*, **13**, R35.

736 55. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis,
737 S., Gatto, L., Girke, T. *et al.* (2015) Orchestrating high-throughput genomic analysis with
738 Bioconductor. *Nat Methods*, **12**, 115-121.

739 56. Colombo, M., Karousis, E.D., Bourquin, J., Bruggmann, R. and Muhlemann, O. (2017) Gene
740 Expression Omnibus GSE86148
741 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86148). [DATASET]

742 57. Boehm, V., Britto-Borges, T., Steckelberg, A.L., Singh, K.K., Gerbracht, J.V., Gueney, E.,
743 Blazquez, L., Altmuller, J., Dieterich, C. and Gehring, N.H. (2018) ArrayExpress E-MTAB-6564
744 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6564) [DATASET]

745 58. Perez-Silva, J.G., Araujo-Voces, M. and Quesada, V. (2018) nVenn: generalized, quasi-
746 proportional Venn and Euler diagrams. *Bioinformatics*, **34**, 2322-2324.

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

749 60. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H. (2014) UpSet: Visualization of
750 Intersecting Sets. *IEEE Trans Vis Comput Graph*, **20**, 1983-1992.

751 61. Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in
752 multidimensional genomic data. *Bioinformatics*, **32**, 2847-2849.

753 62. Wu, D. and Smyth, G.K. (2012) Camera: a competitive gene set test accounting for inter-gene
754 correlation. *Nucleic Acids Res*, **40**, e133.

755 63. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M. and Krijgsveld, J. (2014)
756 Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol*, **10**, 757.

757 64. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized
758 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, **26**,
759 1367-1372.

760 65. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. and Mann, M. (2011)
761 Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome*
762 *Res*, **10**, 1794-1805.

763 66. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M. and Cox, J.
764 (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data.
765 *Nat Methods*, **13**, 731-740.

766 67. Nolte, H., MacVicar, T.D., Tellkamp, F. and Kruger, M. (2018) Instant Clue: A Software Suite for
767 Interactive Data Visualization and Analysis. *Sci Rep*, **8**, 12648.

768 68. Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

769 69. Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C(T)
770 method. *Nat Protoc*, **3**, 1101-1108.

771 70. Mansour, F.H. and Pestov, D.G. (2013) Separation of long RNA by agarose-formaldehyde gel
772 electrophoresis. *Anal Biochem*, **441**, 18-20.

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

775 72. Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley,
776 C.R., Bradley, R.K., Irimia, M. *et al.* (2018) Exon Junction Complex Shapes the Transcriptome by
777 Repressing Recursive Splicing. *Mol Cell*, **72**, 496-509 e499.

778 73. Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M.,
779 Zimmermann, M., Fradet-Turcotte, A., Sun, S. *et al.* (2015) High-Resolution CRISPR Screens
780 Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, **163**, 1515-1526.

781 74. Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M.,
782 Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A. *et al.* (2017) Evaluation and Design of
783 Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 (Bethesda)*, **7**, 2719-2727.

784 75. Tuladhar, R., Yeu, Y., Tyler Piazza, J., Tan, Z., Rene Clemenceau, J., Wu, X., Barrett, Q., Herbert,
785 J., Mathews, D.H., Kim, J. *et al.* (2019) CRISPR-Cas9-based mutagenesis frequently provokes
786 on-target mRNA misregulation. *Nat Commun*, **10**, 4056.

787 76. Smits, A.H., Ziebell, F., Joberty, G., Zinn, N., Mueller, W.F., Clauder-Munster, S., Eberhard, D.,
788 Falth Savitski, M., Grandi, P., Jakob, P. *et al.* (2019) Biological plasticity rescues target activity
789 in CRISPR knock outs. *Nat Methods*, **16**, 1087-1093.

790 77. Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H. and Conti,
791 E. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a
792 core EJC-UPF3b complex. *Proc Natl Acad Sci U S A*, **107**, 10050-10055.

793 78. Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W. and
794 Kulozik, A.E. (1998) Binary specification of nonsense codons by splicing and cytoplasmic
795 translation. *EMBO J*, **17**, 3484-3494.

796 79. Boehm, V., Gerbracht, J.V., Marx, M.C. and Gehring, N.H. (2016) Interrogating the degradation
797 pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun*, **7**, 13691.

80. Voigt, F., Gerbracht, J.V., Boehm, V., Horvathova, I., Eglinger, J., Chao, J.A. and Gehring, N.H. (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. *Nat Protoc*, **14**, 1603-1633.

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

82. Mao, H., Brown, H.E. and Silver, D.L. (2017) Mouse models of Casc3 reveal developmental functions distinct from other components of the exon junction complex. *RNA*, **23**, 23-31.

83. Hurt, J.A., Robertson, A.D. and Burge, C.B. (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res*, **23**, 1636-1650.

84. Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y. and Akimitsu, N. (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol*, **9**, 1370-1379.

85. Schmidt, S.A., Foley, P.L., Jeong, D.H., Rymarquis, L.A., Doyle, F., Tenenbaum, S.A., Belasco, J.G. and Green, P.J. (2015) Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Res*, **43**, 309-323.

86. Cougot, N., Daguenet, E., Baguet, A., Cavalier, A., Thomas, D., Bellaud, P., Fautrel, A., Godey, F., Bertrand, E., Tomasetto, C. *et al.* (2014) Overexpression of MLN51 triggers P-body disassembly and formation of a new type of RNA granules. *J Cell Sci*, **127**, 4692-4701.

87. Ishigaki, Y., Li, X., Serin, G. and Maquat, L.E. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell*, **106**, 607-617.

88. Maquat, L.E., Tarn, W.Y. and Isken, O. (2010) The pioneer round of translation: features and functions. *Cell*, **142**, 368-374.

89. Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M. and Tanenbaum, M.E. (2019) Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. *Mol Cell*, **75**, 324-339 e311.

90. Rufener, S.C. and Muhlemann, O. (2013) eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat Struct Mol Biol*, **20**, 710-717.

91. Durand, S. and Lykke-Andersen, J. (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat Struct Mol Biol*, **20**, 702-709.

92. Trcek, T., Sato, H., Singer, R.H. and Maquat, L.E. (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev*, **27**, 541-551.

93. Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T. *et al.* (2015) ArrayExpress update--simplifying data submissions. *Nucleic Acids Res*, **43**, D1113-1116.

94. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M. *et al.* (2019) The PRIDE database and related

837 tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*, **47**,
838 D442-D450.

839

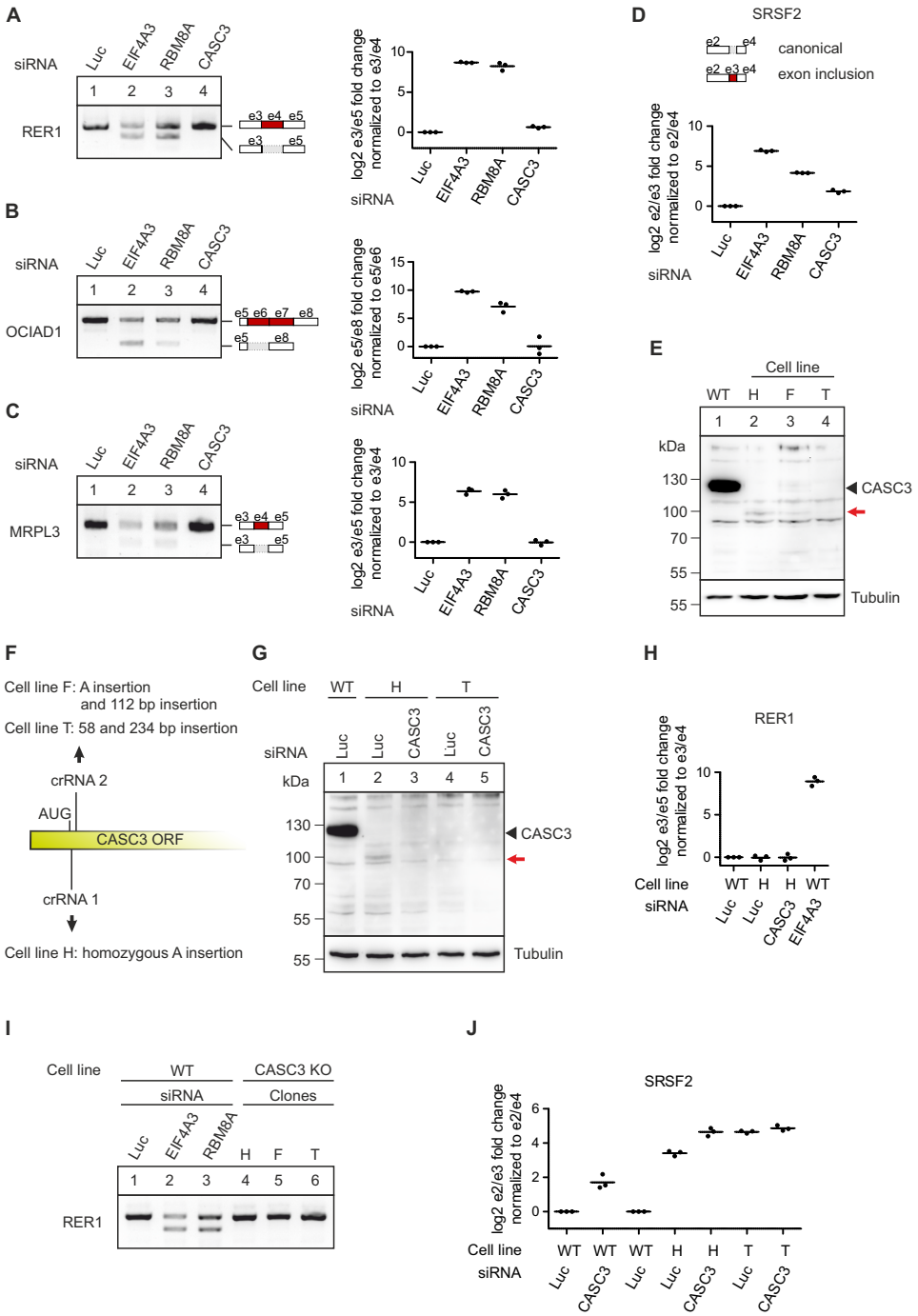


Figure 1 - CASC3 is not involved in the splicing regulation of known EJC-dependent targets.

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: 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.

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

G: 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.

H: 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.

I: 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).

J: Relative quantification of the SRSF2 transcript isoforms by qPCR in WT cells treated with Luc siRNA as a negative control, WT cells treated with CASC3 siRNA, CASC3 KO cell line H treated with Luc siRNA, CASC3 KO cell line H treated with CASC3 siRNAs and CASC3 KO cell line T treated with Luc siRNA as well as CASC3 KO cell line T treated with CASC3 siRNAs. Data points and means are plotted (n=3).

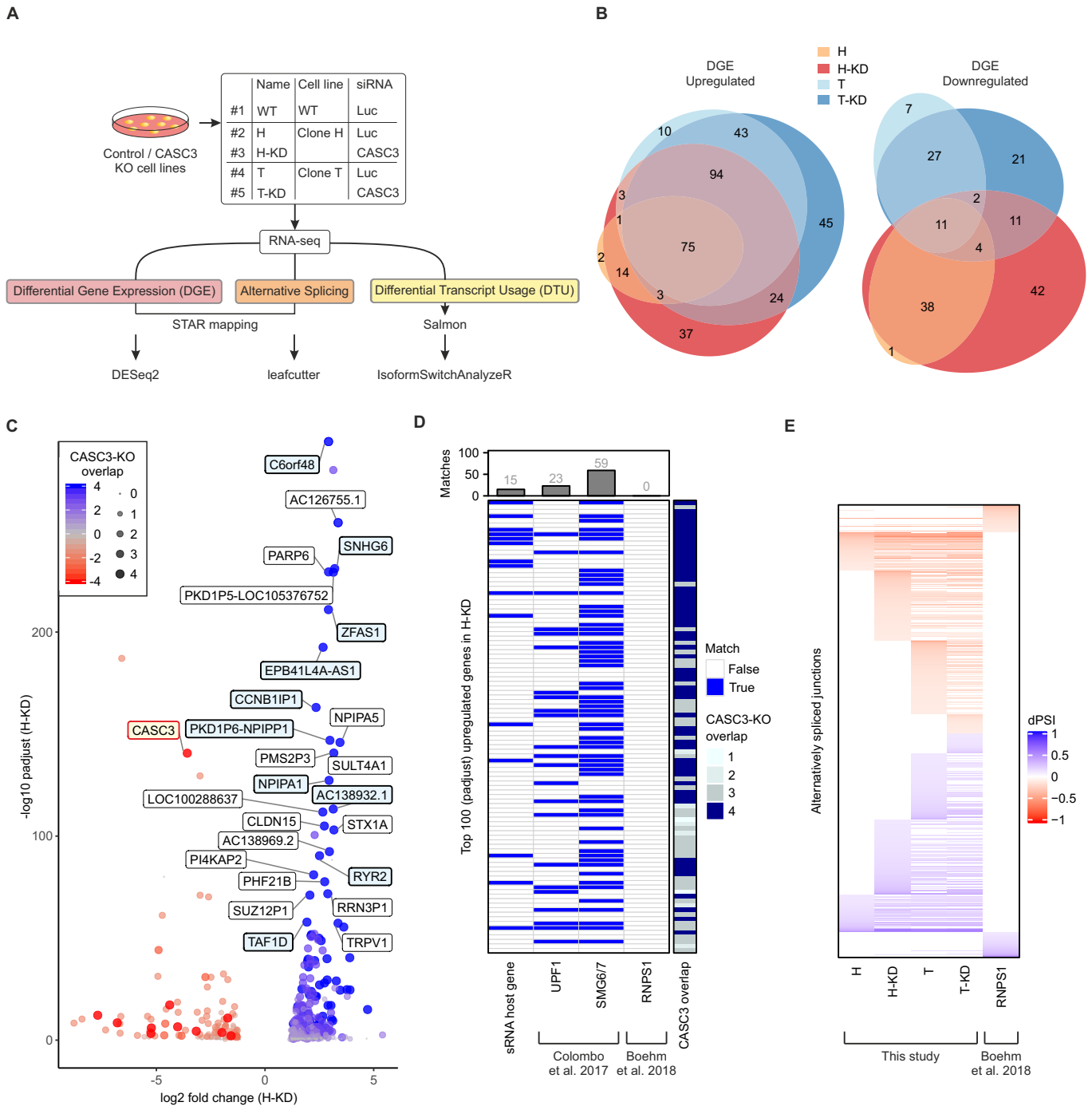


Figure 2 - Transcriptome-wide effects of CASC3 depletion.

A: Workflow 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 an Euler plot, some intersections cannot be plotted. All intersections are shown in Supplementary Figure S2C.

C: Volcano plot of differential gene expression analysis of the condition H-KD using overlap from Figure 2B 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).

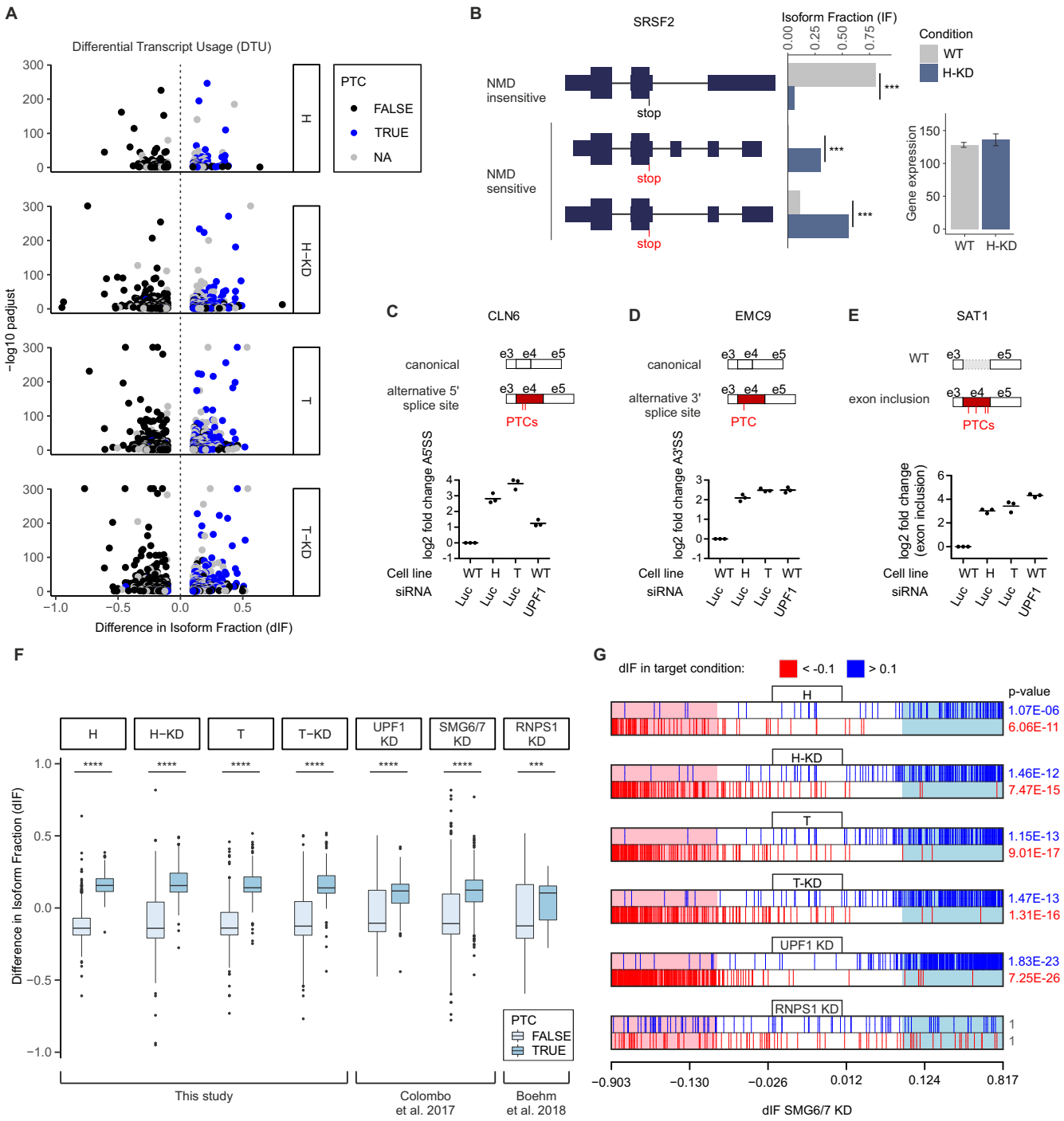


Figure 3 - Knockout of CASC3 leads to a global upregulation of NMD-sensitive transcript isoforms.

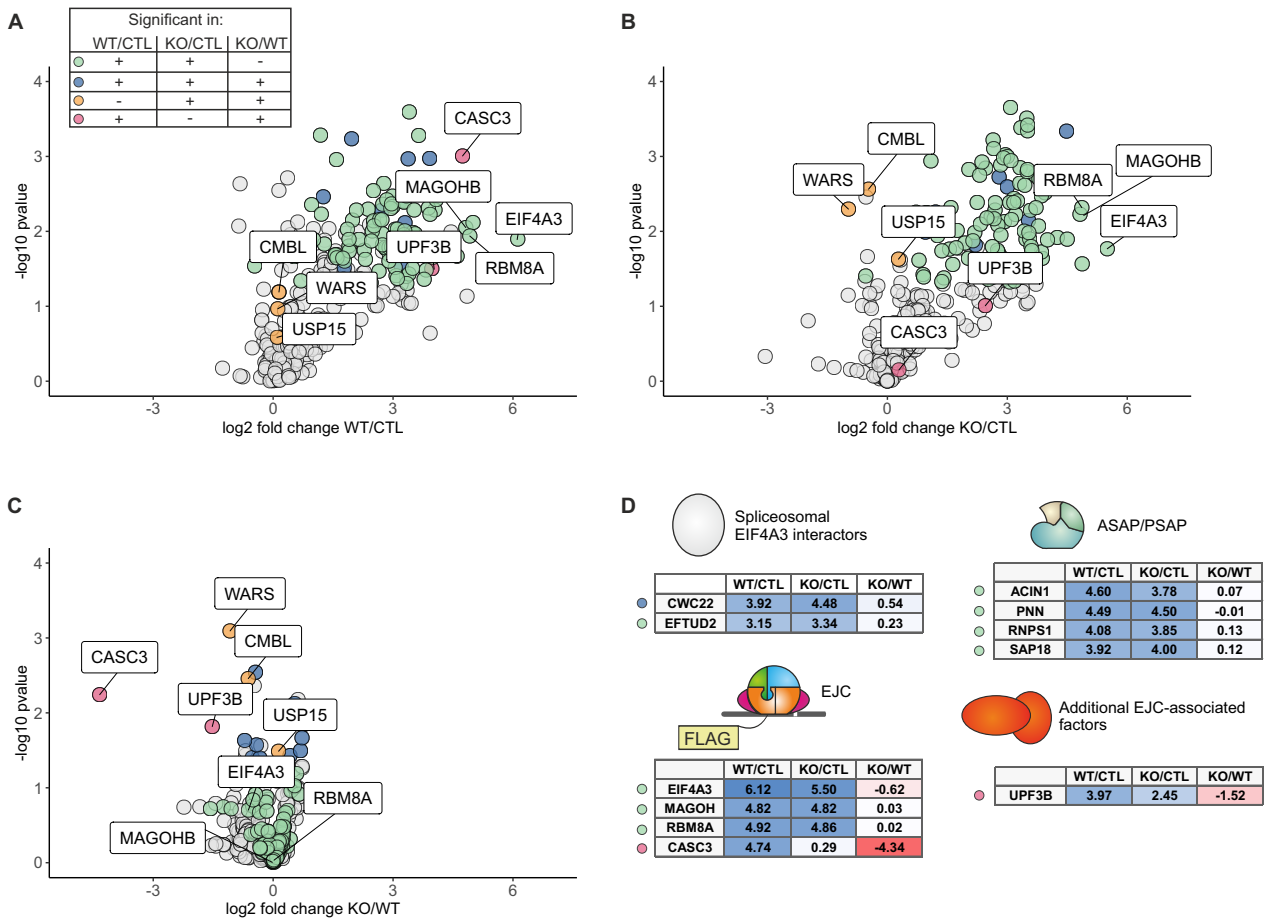
A: Results from IsoformSwitch analysis plotted as volcano plots. Transcript isoforms identified as NMD-sensitive are shown as blue dots. Isoforms with no annotated open reading frame are designated as "NA". Difference in Isoform Fraction (dIF) is plotted against –log10 padjust (adjusted p-value).

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: Barcode plots showing the enrichment of transcript isoforms that undergo isoform switching (padj < 0.05) of a target condition compared to the SMG6/7 KD dataset. On the x-axis the dIF of transcripts that undergo isoform switching in SMG6/7 are ranked according to their dIF. The regions in the barcode plot with |dIF| > 0.1 of SMG6/7 KD are shaded light red if dIF < -0.1 and light blue if dIF > 0.1. Similarly, individual transcript isoforms with a |dIF| > 0.1 in the target conditions are marked with red lines if dIF < -0.1 and blue lines if dIF > 0.1. To test whether the up- and downregulated sets of transcripts in the target conditions are highly ranked in terms of dIF relative to transcripts that are not in the respective set, the camera test from the limma R package was performed and significant p-values (< 0.05) are shown in blue (up) or red (down).



	WT/CTL	KO/CTL	KO/WT
CWC22	3.92	4.48	0.54
EFTUD2	3.15	3.34	0.23

	WT/CTL	KO/CTL	KO/WT
ACIN1	4.60	3.78	0.07
PNN	4.49	4.50	-0.01
RNPS1	4.08	3.85	0.13
SAP18	3.92	4.00	0.12

	WT/CTL	KO/CTL	KO/WT
EIF4A3	6.12	5.50	-0.62
MAGOH	4.82	4.82	0.03
RBM8A	4.92	4.86	0.02
CASC3	4.74	0.29	-4.34

	WT/CTL	KO/CTL	KO/WT
UPF3B	3.97	2.45	-1.52

Figure 4 - Cells that lack CASC3 have intact EJCs.

A-C: 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. A: EIF4A3 against FLAG control in WT cells, B: EIF4A3 against FLAG control in KO cells, C: 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.

D: Overview of the enrichment of EJC- and EJC-associated proteins.

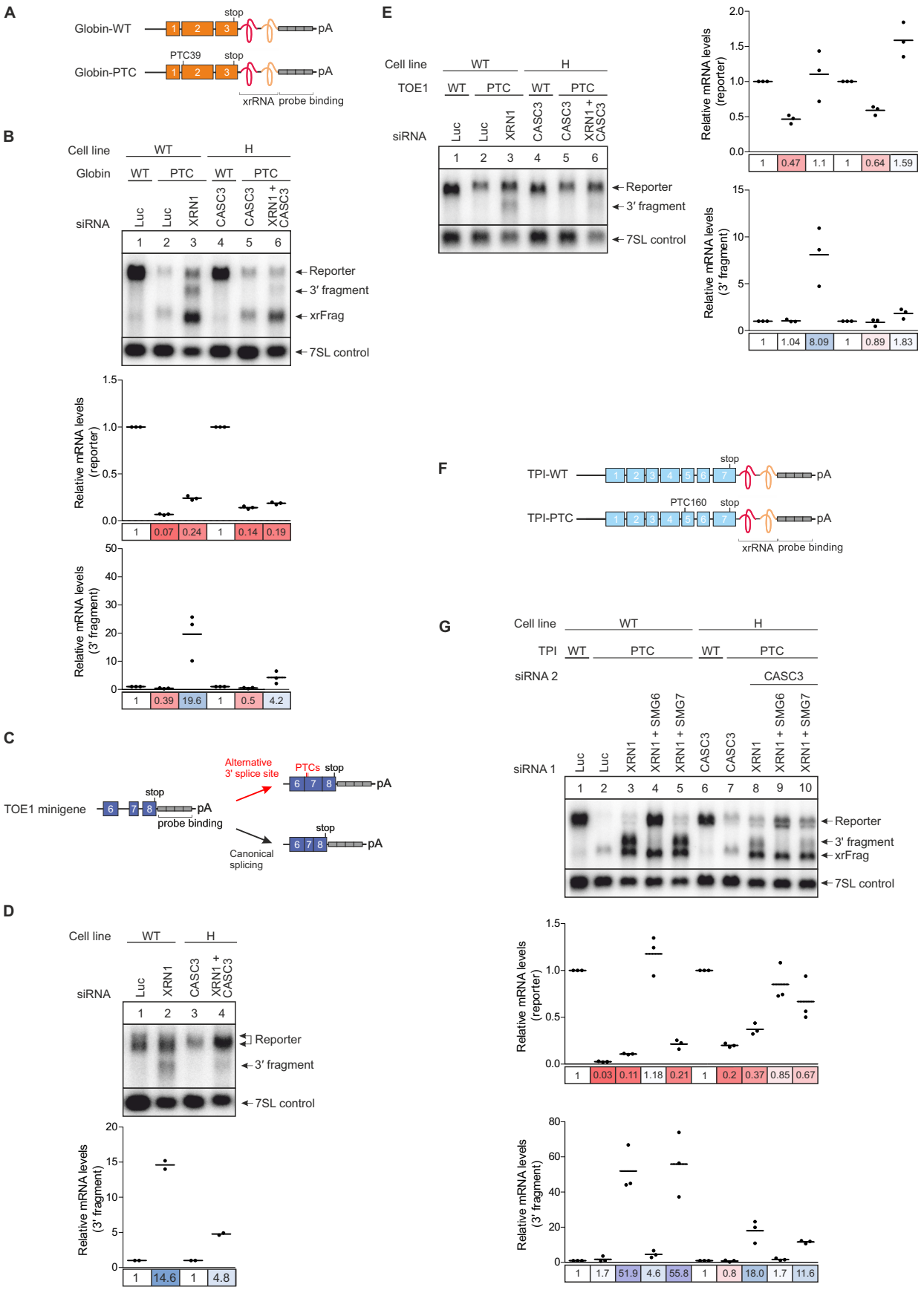

A
Globin-WT
stop
pA
Globin-PTC
PTC39
stop
pA
xrRNA
probe binding
B
Cell line
WT
H
Globin
WT
PTC
WT
PTC
siRNA
Luc
Luc
XRN1
CASC3
CASC3
XRN1 + CASC3
1 2 3 4 5 6
Reporter
3′ fragment
xrFrag
7SL control
Relative mRNA levels (reporter)
1 0.07 0.24 1 0.14 0.19
Relative mRNA levels (3′ fragment)
1 0.39 19.6 1 0.5 4.2
C
TOE1 minigene
stop
pA
probe binding
Alternative 3′ splice site
PTCs
stop
pA
Canonical splicing
stop
pA
D
Cell line
WT
H
siRNA
Luc
XRN1
CASC3
XRN1 + CASC3
1 2 3 4
Reporter
3′ fragment
7SL control
Relative mRNA levels (3′ fragment)
1 14.6 1 4.8
E
Cell line
WT
H
TOE1
WT
PTC
WT
PTC
siRNA
Luc
Luc
XRN1
CASC3
CASC3
XRN1 + CASC3
1 2 3 4 5 6
Reporter
3′ fragment
7SL control
Relative mRNA levels (reporter)
1 0.47 1.1 1 0.64 1.59
Relative mRNA levels (3′ fragment)
1 1.04 8.09 1 0.89 1.83
F
TPI-WT
stop
pA
TPI-PTC
PTC160
stop
pA
xrRNA
probe binding
G
Cell line
WT
H
TPI
WT
PTC
WT
PTC
siRNA 2
CASC3
siRNA 1
Luc
Luc
XRN1
XRN1 + SMG6
XRN1 + SMG7
CASC3
CASC3
XRN1
XRN1 + SMG6
XRN1 + SMG7
1 2 3 4 5 6 7 8 9 10
Reporter
3′ fragment
xrFrag
7SL control
Relative mRNA levels (reporter)
1 0.03 0.11 1.18 0.21 1 0.2 0.37 0.85 0.67
Relative mRNA levels (3′ fragment)
1 1.7 51.9 4.6 55.8 1 0.8 18.0 1.7 11.6

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: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing a prespliced variant of the TOE1 minigene reporter depicted in Figure 5C. The intron between exon 6 and 7 is deleted so that the reporter is constitutively spliced to contain either a normal stop codon (TOE1-WT) or a PTC (TOE1-PTC). The mRNA levels were normalized to 7SL RNA. For the relative mRNA quantification, in each condition (WT vs. CASC3 KO with KD) the reporter and 3′ fragment levels were normalized to the TOE-WT reporter (lanes 1 and 4).

F: 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.

G: 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.

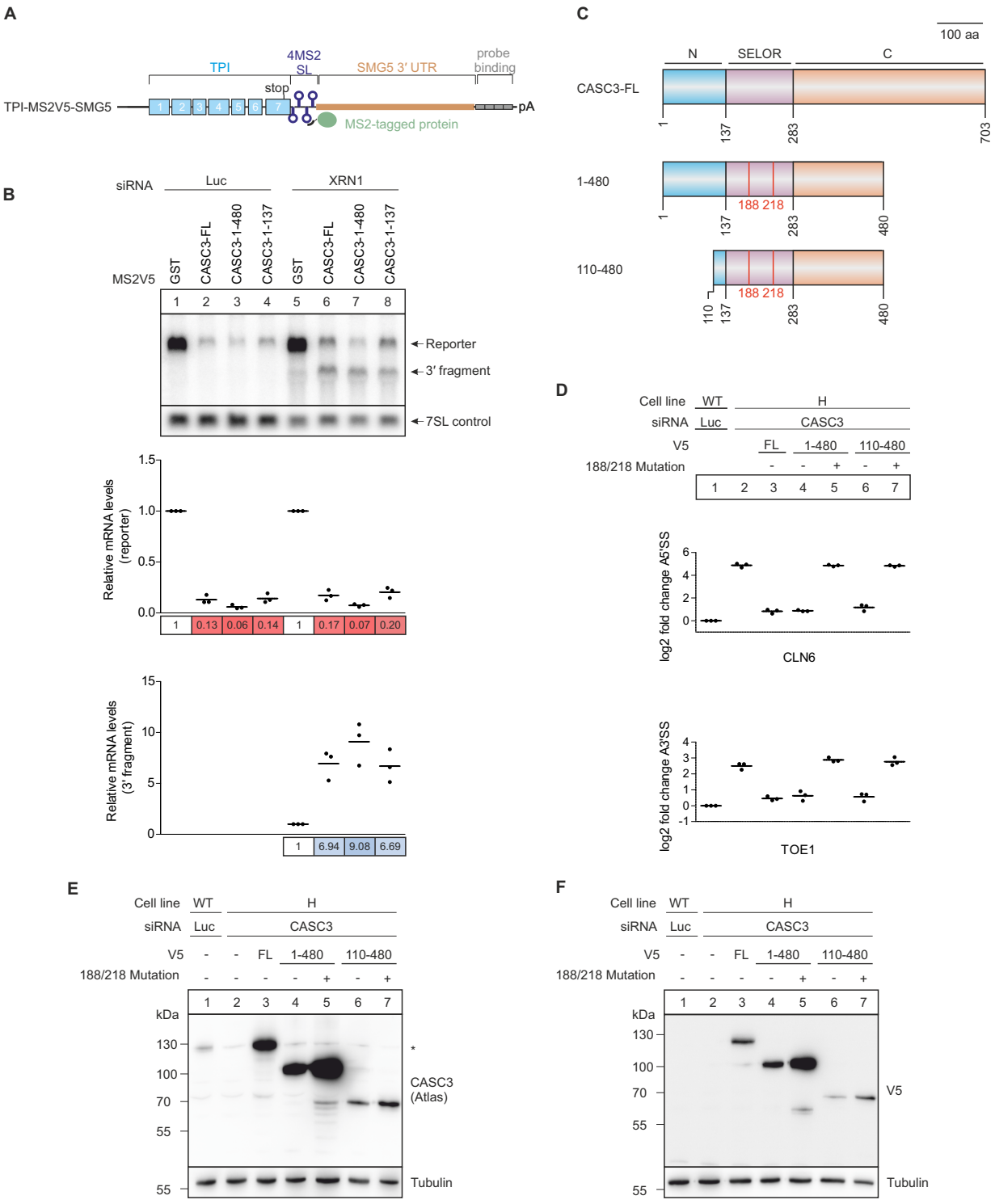


Figure 6 - The CASC3 N-terminus promotes but is not necessary to elicit NMD.

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 Figure 6A 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 schematically in Figure 6C. Rescue protein expression is confirmed in Figure 6E and F. Individual data points and means are plotted from n=3 experiments.

E and F: Western blot of samples shown in Figure 6D. The expression of rescue proteins was confirmed by an antibody against CASC3 (E) and an antibody recognizing the V5 tag (F).

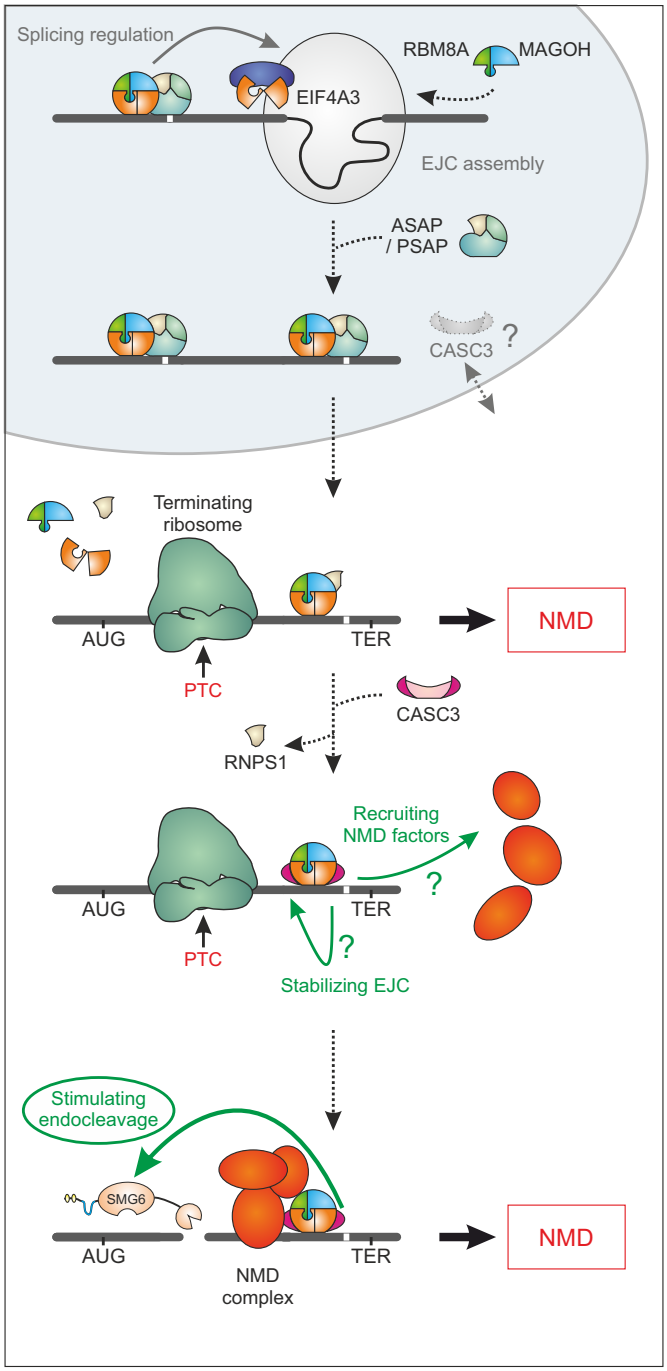


Figure 7 - Model of CASC3's cellular function.

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.