Supplemental Materials

Molecular Biology of the Cell

Reber et al.

Supplemental Information

CRISPR-Trap: A clean approach for the generation of gene knockouts and gene replacements in human cells.

Stefan Reber, Jonas Mechtersheimer, Sofia Nasif, Julio Aguila Benitez, Martino Colombo, Michal Domanski, Daniel Jutzi, Eva Hedlund and Marc-David Ruepp

Supplemental Material and Methods

Oligonucleotides

List of oligonucleotides used for this study. All oligonucleotides and TaqMan probes were purchased from Microsynth AG, Switzerland.

Primer	5'-sequence-3'	aim
5.8S rRNA fwd (SYBR)	GGTGGATCACTCGGCTCGT	qPCR
5.8S rRNA rev (SYBR)	GCAAGTGCGTTCGAAGTGTC	qPCR
5.8S rRNA fwd (TaqMan)	GCGCTAGCTGCGAGAATTA	qPCR
5.8S rRNA rev (TaqMan)	ACGCTCAGACAGGCGTA	qPCR
5.8S rRNA TaqMan probe	TGATCATCGACACTTCGAACGCACT	qPCR
human actin-β fwd (SYBR)	GACTTCGAGCAAGAGATGGC	qPCR
human actin-β rev (SYBR)	AGGAAGGAAGGCTGGAAGAG	qPCR
human actin-β fwd	CTGGCACCCAGCACAATG	qPCR
human actin-β rev	GCCGATCCACACGGAGTACT	qPCR
human actin-β TaqMan probe	ATCAAGATCATTGCTCCTCCTGAGCGC	qPCR
FUS mRNA fwd	AGCGGTGTTGGAACTTCG	qPCR
FUS mRNA rev	GACTGCTCTGCTGGGAATAG	qPCR
TDP-43 CDS fwd	AACCGAACAGGACCTGAAAGAG	qPCR
TDP-43 CDS rev	CAGTCACACCATCGTCCATCTATC	qPCR
TDP-43 3' UTR fwd	GAACGCAAGGCTGTGATATG	qPCR
TDP-43 3'UTR rev	TATGCACCACCAAGTCTCTG	qPCR
SMG7 fwd	TTGGCCCTGGTTGATTTCTC	qPCR
SMG7 rev	CTGGAAGTGGTGTCGCACTAA	qPCR
human β-globin fwd	GCTGCACTGTGACAAGCTGC	qPCR
human β-globin rev	AAAGTGATGGGCCAGCACAC	qPCR
human β-globin TaqMan probe	TCCTGAGAACTTCAGGCTCCTGGGCAAC	qPCR
mini-µ fwd	GTCTCACCTTCTTGAAGAACGTGTC	qPCR
mini-µ rev	GGGATGGTGAAGGTTAGGATGTC	qPCR
mini-µ TaqMan probe	CACATGTGCTGCCAGTCCCTCCA	qPCR
TDP-43 intron 1 fwd	CATGCCTCAGCCAGTTAG	PCR of genomic locus
TDP-43 intron 1 rev	CGAGGTGAAGGTAAGGGTAG	PCR of genomic locus
DD-TDP43 rev	CTTGCCCAGCATGAACTTG	PCR of genomic locus
FUS intron 4 fwd	TAAAGGGACCAGCAGTAG	PCR of genomic locus
FUS intron 5 rev	TGACAAGTGCCAGGCAATGAAG	PCR of genomic locus
SMG7 intron 1 fwd	AAAGATCTGTTGTAGGTGCTATAGAGAGA	PCR of genomic locus
SMG7 intron 1 rev	CTCCCATACAGGTGATCAACAAAAC	PCR of genomic locus
SMG7 exon 3 fwd	CAGAAGCATTGTAAACCAGACTTAG	PCR of genomic locus
SMG7 exon 3 rev	ATAAGACTGACCACAGACAATC	PCR of genomic locus
SMG7 exon 15 fwd	GGAGTAGGGTCTTCTTTC	PCR of genomic locus
SMG7 exon 15 rev	TAGGTAGAGGGAGTCTTC	PCR of genomic locus
sgRNA TDP-43 oligo fwd	CACCGCTGAAGGCTTATGATGCGTC	cloning
sgRNA TDP-43 oligo rev	AAACGACGCATCATAAGCCTTCAGC	cloning
sgRNA SMG7 oligo fwd	CACCGACCAAACGTTGCCGGGCACG	cloning
sgRNA SMG7 oligo rev	AAACCGTGCCCGGCAACGTTTGGTC	cloning
sgRNA SMG7 ex3 fwd	CACCGAAAATGCTAGTTACCGATT	cloning
sgRNA SMG7 ex3 rev	AAACAATCGGTAACTAGCATTTTC	cloning
sgRNA SMG7 ex15 fwd	CACCGGAGTTACTTGCTTGAGTTG	cloning

sgRNA SMG7 ex15 rev	AAACCAACTCAAGCAAGTAACTCC	cloning
HBB HindIII fw	AATCTGAAGCTTATGGTGCATCTGACTCCTGAGG	cloning
HBB NotI rev	AATCTCGCGGCCGCTTAGTGATACTTGTGGGCCAGG	cloning
2xHA mini-µ TER30	GGTCCAGACTGCTGCTACTGGACCTGGCAATG	site-directed mutagenesis
2xHA mini-µ TER100	GCTGTGCTGGAGGATTAGTCTGCAGTCAGTGTG	site-directed mutagenesis
2xHA mini-µ TER200	AAGTCATCCAGGGTATCTGAACCTTCCCAACACTG	site-directed mutagenesis
2xHA mini-µ TER300	CTGGCTAAAGGATGGGTAGCTCGTGGAATCTGG	site-directed mutagenesis
2xHA mini-µ TER400	CTGGTCTCAAACCTGGCAACCTATTAAACCCTGAATATC	site-directed mutagenesis
2xHA mini-µ TER500	AGGAGAGAAGCCCTTCTACAGGCAGGTGACTGTG	site-directed mutagenesis
FUS exon 5 fwd	CCAGAGTGGGAGCTACAG	Sanger sequencing (PCR product)
M13 fwd	GTAAAACGACGGCCAG	Sanger sequencing (TOPO vector)
SMG7 5HA fwd	TGTTTCATACCTTGCGGAGGAGGGAG	PCR for cloning SMG7KO plasmid
SMG7 5HA rev	TTCTACCAATAACGTTTGGTATCTTTCATG	PCR for cloning SMG7KO plasmid
SMG7 3HA fwd	TCATGTCTGAGCCGGGCACGGGGGCTCACG	PCR for cloning SMG7KO plasmid
SMG7 3HA rev	TGTAAGCTCTGTAAAAGGAGGGTG	PCR for cloning SMG7KO plasmid
ZeoR cassette fwd	ACCAAACGTTATTGGTAGAAACAACTACAT	PCR for cloning SMG7KO plasmid
ZeoR cassette rev	CGTGCCCGGCTCAGACATGATAAGATACAT	PCR for cloning SMG7KO plasmid

Plasmids

pTAL EF1a 023977, pTAL EF1a 021443 and pCLS22315 were purchased from Cellectis. The FUS exon 5 reporter plasmid was generated by annealing two oligonucleotides (5'-TGGACAGCAGCAAAGCTATAATCCCCCTCAGGGCTATGGACAGCAGAAACGT - 3' and 5'-TTCTGCTGTCCATAGCCCTGAGGGGGATTATAGCTTTGCTGCTGTCCAAGCT - 3') and subsequent cloning into the *AatII* and *SacI* sites of pRR-Puro described in (Flemr & Buhler, 2015). The two pU6gDNA-Cas9-GFP vectors targeting FUS intron 1 were purchased from Sigma Aldrich. The pCRISPR-EF1a-eSpCas9(1.1) was custom synthesized [General Biosystems]. The eSpCas9(1.1) is described in (Slaymaker et al, 2016). To clone the sequence targeting the locus of interest into pCRISPR-EF1a-eSpCas9, two complementary oligonucleotides (see list of oligonucleotides) were phosphorylated with T4 PNK [New England Biolabs], annealed by heating to 95 °C and subsequently cooled down to room temperature creating 5' overhangs on both ends. The phosphorylated and annealed oligonucleotides where then ligated into the *BbsI* sites of pCRISPR-EF1a-eSpCas9(1.1). The pcDNA-2xHA mini-µ WT plasmid was constructed as follows: the *NotI-SpeI* mini-µ fragment of pβ mini-µ wt uA1 (previously described by (Metze et al, 2013)) was inserted into the *NotI-XbaI* sites of pcDNA 3.1, followed by insertion of a synthetic gene fragment [General Biosystems] encoding 2xHA tag instead of the variable leader sequence into the *EcoRI-XbaI* sites, thereby creating WT mini-µ with a 2xHA tag instead of the variable leader sequence. In order to obtain a series of mini-µ constructs harboring the termination codons at positions 30, 100, 200, 300, 400 and 500, respectively, site-directed mutagenesis with primers described in list of oligonucleotides was performed. All constructs were verified by sequencing. The donor plasmid for generating the SMG7 KO (SMG7 KO ZeoR) was created by cloning the two homology arms flanking the cassette containing the beta globin intron, the zeocin resistance coding sequence (in frame with the first exon of SMG7) and the SV40 polyadenylation signal into pCRII-TOPO vector [Invitrogen]. The two homology arms were amplified by PCR using genomic DNA from HT1080 cells as a template and fused to the zeocin coding cassette by overlap extension PCR (primers in the oligonucleotide list). All plasmids were verified by sequencing.

Antibodies

The mouse Y12 antibody (anti-SmB/B'), the rabbit polyclonal anti CPSF-73 antibody, and the polyclonal rabbit anti-FUS antibody are described elsewhere (Jenny et al, 1996; Lerner et al, 1981; Raczynska et al, 2015). The rabbit polyclonal anti-CPSF-100 antibody was raised against the N-terminal peptide of CPSF-100 (Uniprot Q9P210). The DNA fragment encoding for *E.coli* codon-optimized, hexahistidine (His6) – tagged CPSF-100 (residues 593-782) was custom synthesized [GeneArt, Invitrogen] and cloned into pET28a plasmid using *SalI* and *XbaI* sites. The recombinant protein was expressed in BL21(DE3)

Codon Plus RIPL [Agilent Technolgies] and purified under denaturing conditions over Ni-NTA Superflow [IBA Lifesciences]. Rabbits were immunized with the purified protein in combination with GERBU Adjuvant LQ. Antibodies were affinity-purified from rabbit serum using a Maltose Binding Protein-CPSF-100 fragment fusion protein coupled to NHS-activated Agarose Slurry [Thermo Scientific]. Additional antibodies that were used: mouse anti-tyrosine tubulin [T9028, Sigma-Aldrich], mouse anti-TDP-43 [sc-100871, Santa Cruz], mouse anti-HA [sc-7392, Santa Cruz], mouse anti-FLAG [F1804, Sigma Aldrich], rabbit anti-SMG7 [A302-170A, Bethyl] (used in Figure 4C), rabbit anti-SMG7 [MBS820862, MyBioSource] (used in Supplementary Figure S4C), donkey anti-rabbit IRDye800CW [926-32213, LI-COR Biosciences], donkey anti-mouse IRDye800CW [926-32212, LI-COR Biosciences], donkey anti-mouse IRDye680LT [925-68022, LI-COR Biosciences], goat anti-mouse IRDye800 [925-32210, LI-COR Biosciences], donkey anti-rabbit IRDye680LT [926-68023, LI-COR Biosciences].

Computational analysis of publicly available datasets

The transcriptomics data of each study in Supplementary Table S1 was produced by different methods and was available in various formats, either as raw data or intermediate processed files. This variability required custom analyses to be carried out in every case. Richter et al. employed U133A and U133B Affymetrix GeneChips, CEL files were provided, which we analyzed with the "Affy" R package (Buszczak et al, 2014), performing both a quantile and background normalization in combination with the RMA expression method (Kristensen et al, 2013). The used probe ID for ADPGK is 7990231. Sadic et al. made use of an Illumina NextSeq 500 RNA-seq run and published a counts table matrix, produced with a BowTie (Gough et al, 2012) mapping on a custom annotation set. We performed a differential expression analysis with DESeq2 (Zhang et al, 2014). McCleland et al. sequenced their KO clones with an Illumina HiSeq 2000, reporting a table with reads per kilobase of transcript per Million mapped reads (RPKM) values, instead of counts table. In this case, we simply computed the ratio of such values. Kabir et al. interrogated the transcriptome by means of an Illumina HumanHT-12 V4.0 expression beadchip. Encrypted proprietary idat files were compared with the package "illuminaio" R package (Stanford et al, 1998). The probe ID corresponding to RAP1 is ILMN_171812. Lebedeva et al. studied their gene expression through a Illumina HiSeq 2500 sequencing, making their whole final results available online. In this case we just picked the information about FUS directly from the differential expression analysis table. An Affymetrix Mouse Genome 430 2.0 Array was chosen by Huang and colleagues, which we analyzed as described before for in (Richter et al, 2013). The chosen probe ID is 7990231. Finally, MyoD expression levels were reported directly in the manuscript by (Kim et al, 2016).

To determine the number of amino acids encoded by all known human spliced protein-coding genes (in Figure 3B) we retrieved the gene annotation from RefSeq release 78. The gene list was filtered for transcripts that code for a protein product and that have at least one intron in their structure. The first exon of these genes was taken into consideration. The corresponding figure depicts the number of amino acids that are coded in these exons. When zero, the first exon is non-coding.

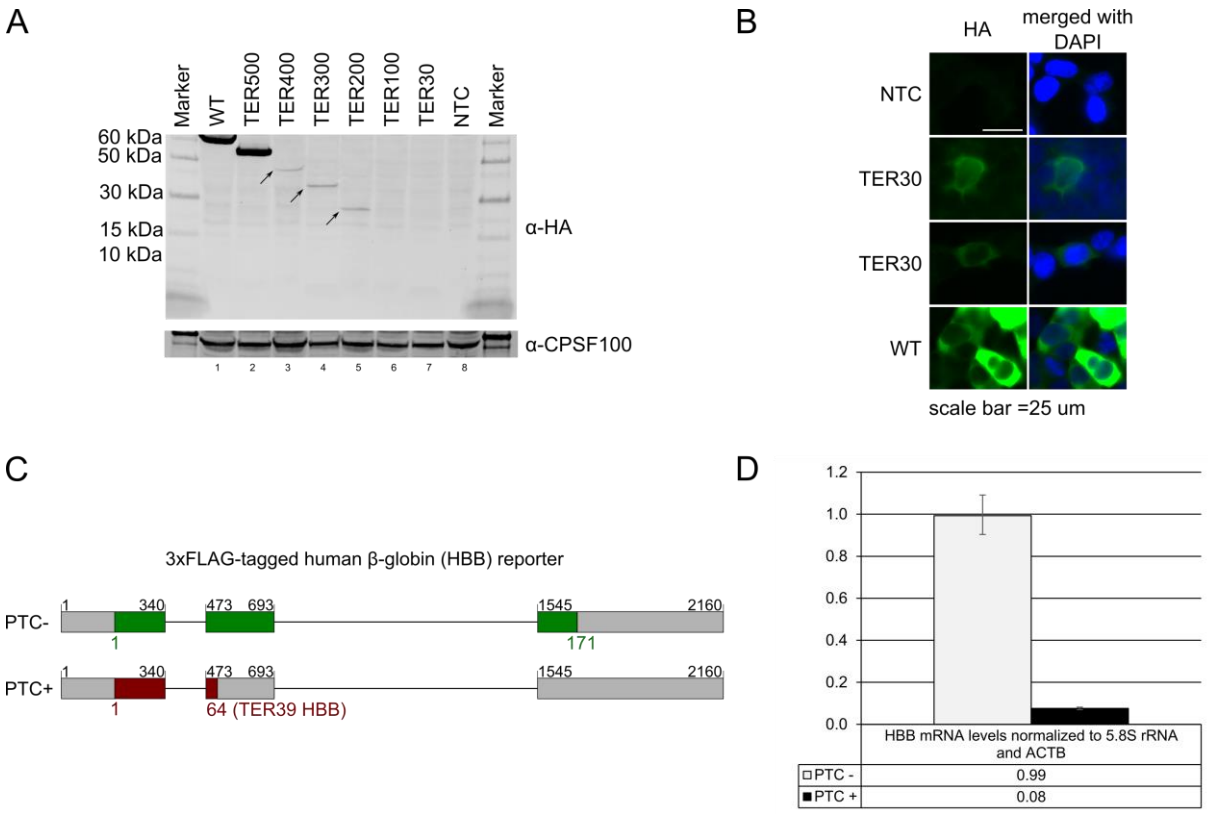
Supplementary Tables and Figures

Reference	target gene	method	target exon	target species	cell line specification	mRNA level relative to WT
10.1016/j.ajpath.2013.05.032	FXN	Cre/loxP-mediated deletion	4	mouse		33.94%
10.15252/embr.201439937	ATRX	CRISPR/Cas9	9	mouse	mESC	35.16%
10.15252/embr.201439937	DAXX	CRISPR/Cas9	2	mouse	mESC	40.91%
10.1002/path.4596	CDK8	CRISPR/Cas9	2	mouse	IMCE_clone1	73.46%
10.1002/path.4596	CDK8	CRISPR/Cas9	2	mouse	IMCE_clone19	62.42%
10.1080/15476286.2016.1256532	FUS	CRISPR/Cas9	3	zebrafish		58.89%
10.5713/ajas.16.0749	MYOD	CRISPR/Cas9	1	chicken	QM7	50.80%
10.1016/j.celrep.2014.10.014	RAP1	TALEN (whole exon deletion)	2	human	HCT116	12.45%
10.1016/j.celrep.2014.10.014	RAP1	TALEN (whole exon deletion)	2	human	HT1080	39.52%
10.1016/j.celrep.2014.10.014	RAP1	TALEN (whole exon deletion)	2	human	ARPE19	46.05%
10.1371/journal.pone.0065267	ADPGK	ZF	1	human	HCT116	59.57%
10.1371/journal.pone.0065267	ADPGK	ZF	1	human	H460	33.33%

Supplementary Table S1 related to Figure 1: Summary of seven studies where knockouts were produced by introducing a frameshift into the CDS of the gene of interest and subsequently whole transcriptome analysis was performed. The target gene, knockout method, targeted exons and the target species of each experiment are shown. If available, the cell line is specified. The last column shows the mRNA level of the knockout target relative to the mRNA level in wild type cells.

```
  wild type_NG_012889.2_pos. 4180 AGCAGCAAAGCTATGGACAGCAGCAAAGCTATAATCCCCCTCA
frameshift clone 1 (fs1)_allele A AGCAGCAAAGCTATG--------------------CCCCCTCA
frameshift clone 1 (fs1)_allele B AGCAGCAAAGCTATGGACAGCAGCAAAGCTAT--TCCCCCTCA
frameshift clone 2 (fs2)_allele A AGCAGCAAAGCTATGGACAGCAGCAAAGCTATAATC--CCTCA
frameshift clone 2 (fs2)_allele B AGCAGCAAAGCTATGGACAGCAGCAAAGCTATAATC--CCTCA
          in-frame clone_allele A AGCAG------------------------------CCCCCTCA
          in-frame clone_allele B AGCAG------------------------------CCCCCTCA
```

Supplementary Figure S1 related to Figure 1: Sequence of the FUS exon 5 genomic locus of frameshift clone 1, frameshift clone 2 and the in-frame clone aligned to the wild type sequence (RefSeq: NG_012889.2, starting from position 4180). Both alleles of each clone are shown. Deletion of respective base positions are indicated with gaps with a line for each base position. Whereas all deletions in clones fs1 and fs2 lead to a frameshift, the deletion in the in-frame clone leads to the deletion of amino acids Q141 to P152 of FUS protein.

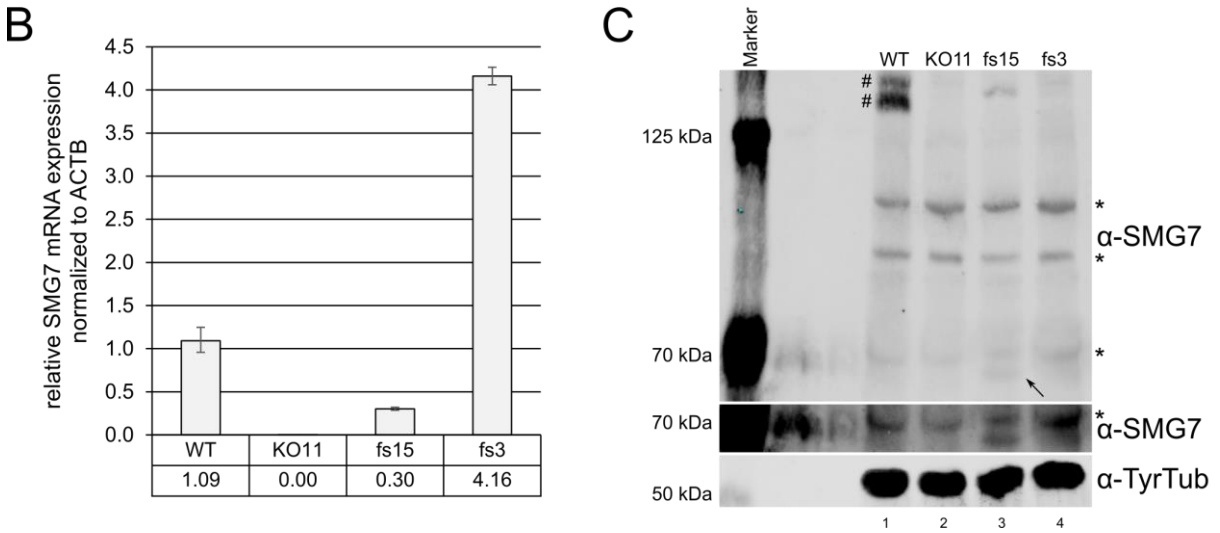


Supplementary Figure S2 related to Figure 2: A) Western blot of HeLa total cell extracts transiently transfected with different mini-µ reporter constructs. Reporter constructs were detected by anti-HA antibodies. CPSF-100 served as loading control. A non-transfected control (NTC) was included. The expected bands of TER400, TER300 and TER200 are indicated with black arrows. Note that the TER100 construct, detectable by slot blotting, escapes detection by western blotting. B) Immunostaining of HEK293T cells transfected with wild type (WT) or TER30 mini-µ reporter plasmids. The HA-tag of the reporter constructs were visualized with anti-HA antibodies (green) and the cells were counterstained with DAPI. A non-transfected control (NTC) was included to estimate background staining. Scale bar = 25 µm. C) Scheme of HBB reporter constructs. Boxes and thin lines indicating exonic and intronic sequences respectively. Base pair positions are indicated in black above the respective reporter construct. The coding sequence of the PTC- construct with the stop codon at amino acid position 171 is indicated in green. The coding sequence of the PTC+ construct with the stop codon at amino acid position 64 (corresponding to amino acid position 39 of the β-globin open reading frame) is indicated in red. D) RT-qPCR results depicting the relative human β-globin (HBB) reporter levels. HEK-293 Flp-In T-Rex cells were induced 24 h before harvest with 250 ng/ml doxycycline hyclate to express either 3xFLAG-tagged wild type (PTC-) HBB or 3xFLAG-tagged HBB containing a premature stop codon at position 39 of the HBB ORF (PTC+). HBB levels are normalized to the average of two house-keeping genes (5.8S rRNA and actin beta). Average values and standard deviations of three biological replicates are shown.

A

```
wild type_NG_029808.1_pos. 72031 AAAACACCTGTAACTCAAACCCCAACTCAAGCAAGTAACTCCCAGTTCATCCCCATT
 SMG7_exon15_frameshift_allele A AAAACACCTGTAACTCAAACCCCAA-TCAAGCAAGTAACTCCCAGTTCATCCCCATT
 SMG7_exon15_frameshift_allele B AAAACACCTGTAACTCA----------AAGCAAGTAACTCCCAGTTCATCCCCATT

wild type_NG_029808.1_pos. 43541 GCAGGCTCTGCAGGACCTGTACCAGAAAATGCTAGTTACCGATTTGGAATACGCTTT
  SMG7_exon3_frameshift_allele A GCAGGCTCTGCAGGACCTGTACCAGAAAATGCTAGTTACC-ATTTGGAATACGCTTT
  SMG7_exon3_frameshift_allele B GCAGGCTCTGCAGGACCTGTACCAGAAAATGCTAGTTACC-ATTTGGAATACGCTTT
```



Supplementary Figure S3 related to Figure 4: A) Genomic sequence of the HT1080 SMG7 locus after targeting exon 15 or exon 3, respectively. Exon 15 and exon 3 genomic loci of frameshift clone 15 (fs15) and frameshift clone 3 (fs3) are shown. Sequences are aligned to the wild type sequence (RefSeq: NG_029808.1). Both alleles of each clone are shown. Deletion of respective base positions are indicated with gaps with a line for each base position. B) SMG7 mRNA levels of KO clone 11 (compare Figure 4C and D) and the two frameshift clones described in A are displayed relative to wild type SMG7 mRNA expression as determined by RT-qPCR. Average values and standard deviations of three biological replicates are shown. C) Western blot showing the absence of SMG7 protein in clone KO11 and the presence of C-terminally truncated SMG7 protein in clone fs15 (indicated with a black arrow). For better visualization of the band indicated with the black arrow, this section of the western blot is displayed with a higher exposure (lower membrane part labeled with α-SMG7). The SMG7 antibody does not recognize the very N-terminus of SMG7, hence, the presence of a potential peptide in clone fs3 could not be assessed and can therefore not be excluded. The specific bands for SMG7 in the wild type lane are indicated with hashtags. Asterisk indicate unspecific bands recognized by the α-SMG7 antibodies. TyrTub served as a loading control.

Supplemental References

M Buszczak, Signer RA, Morrison SJ (2014) Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* **159:** 242-251

M Flemr, Buhler M (2015) Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. *Cell reports* **12:** 709-716

CA Gough et al (2012) Prediction of protein-destabilizing polymorphisms by manual curation with protein structure. *PloS one* **7:** e50445

A Jenny, Minvielle-Sebastia L, Preker PJ, Keller W (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* **274:** 1514-1517

SW Kim, Lee JH, Park BC, Park TS (2016) Myotube differentiation in CRISPR/Cas9-mediated MyoD knockout quail myoblast cells. *Asian-Australasian journal of animal sciences* **30:** 1029-1036

AR Kristensen, Gsponer J, Foster LJ (2013) Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Molecular systems biology* **9:** 689

EA Lerner, Lerner MR, Janeway CA, Jr., Steitz JA (1981) Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease. *Proceedings of the National Academy of Sciences of the United States of America* **78:** 2737-2741

S Metze, 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:** 1432-1448

KD Raczynska et al (2015) FUS/TLS contributes to replication-dependent histone gene expression by interaction with U7 snRNPs and histone-specific transcription factors. *Nucleic acids research* **43:** 9711-9728

S Richter et al (2013) Zinc finger nuclease mediated knockout of ADP-dependent glucokinase in cancer cell lines: effects on cell survival and mitochondrial oxidative metabolism. *PloS one* **8:** e65267

IM Slaymaker et al (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* **351:** 84-88

WL Stanford et al (1998) Expression trapping: identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. *Blood* **92:** 4622-4631

X Zhang et al (2014) Impact of human pathogenic micro-insertions and micro-deletions on post-transcriptional regulation. *Human molecular genetics* **23:** 3024-3034