Leading Edge
Minireview

Cell

Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine

Maximilian W. Popp[1,2] and Lynne E. Maquat[1,2,*]
[1]Department of Biochemistry and Biophysics, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA
[2]Center for RNA Biology, University of Rochester, Rochester, NY 14642, USA
*Correspondence: lynne_maquat@urmc.rochester.edu
http://dx.doi.org/10.1016/j.cell.2016.05.053

Nonsense-mediated mRNA decay (NMD) is a eukaryotic mRNA quality control and regulatory process that plays direct roles in human health and disease. In this Minireview, we discuss how understanding the molecular events that trigger NMD can facilitate strategic targeting of genes via CRISPR/Cas9 technologies and also inform disease diagnostics and treatments.

Setting the Stage: mRNA Biogenesis and NMD

Within the nucleus, pre-mRNAs are synthesized and immediately undergo *cis*-modifications and binding by various proteins prior to and during the formation of mRNA. At the 5′ end, a 7-methylguanosine residue is added to protect the transcript from 5′-to-3′ exoribonucleases and provide a binding platform for the cap-binding protein (CBP) complex (CBC), composed of CBP80–CBP20. Mammalian pre-mRNAs often contain intervening regions, or introns, that are excised by complexes of RNAs and proteins called spliceosomes. The act of pre-mRNA splicing contributes to the accurate production of mRNAs and thus proteins, in part by depositing macromolecular landmarks of proteins, named exon-junction complexes (EJCs), upstream of spliced exon-exon junctions. At the 3′ end, a series of non-templated adenosine residues generally form a platform for poly(A)-binding proteins that protect the transcript from 3′-to-5′ exoribonucleases. With an assortment of bound proteins, the mRNA is now termed an mRNP complex that—providing that pre-mRNA processing has been successful—is ready to be exported from the nucleus into the cytoplasm to engage with functional ribosomes and other factors that concomitantly direct protein synthesis and mRNA quality inspection via NMD, a process by which a premature stop codon that would lead to a truncated protein product triggers mRNA degradation.

NMD is tightly intertwined with mRNA biogenesis even at its earliest stages, since NMD is promoted by both the CBC and EJCs (Lykke-Andersen and Jensen, 2015). EJCs, which usually consist of four core proteins—eukaryotic translation initiation factor 4A3 (eIF4A3), cancer susceptibility candidate 3 (CASC3), RNA-binding motif protein 8A (RBM8A or Y14), and mago-nashi homolog (MAGOH)—are deposited ~20–24 nucleotides (nt) upstream of ~80% of all exon-exon junctions (Le Hir et al., 2016). This splicing-dependent "mark" serves to orient the NMD machinery once the mRNP is exported to the cytoplasm for translation-dependent inspection by NMD factors.

Generally, faulty transcripts are triaged for destruction by NMD immediately (i.e., < 1 min) after entry into the cytoplasm, largely before the CBC is replaced by eIF4E, which is the cap-binding protein of the bulk of cellular mRNA (Trcek et al., 2013). NMD uses an initial CBC-promoted "pioneer round" of translation to identify the termination codon within the reading frame. Concomitantly, pioneer ribosomes remove EJCs, which record mRNA splicing history, from within but not downstream of the reading frame, thereby establishing whether the termination codon falls ≥ 50–55 nt upstream of an exon-exon junction—a general indication that the termination codon is premature, since normal termination codons are largely located in the last exon.

The molecular gymnastics of proteins that enable EJC-mediated NMD are partially understood. Like the situation for a normal termination codon, in the case of a premature termination codon (PTC) with a downstream EJC that was not removed by pioneer ribosomes, eukaryotic release factor 1 (eRF1) and eRF3 associate with the terminating ribosome. In the case of a PTC, however, additional factors form a transient so-called SURF complex with eRF1-eRF3 (Yamashita, 2013) (Figure 1A). These include suppressor with morphogenic effect on genitalia 1 (SMG1), which together with SMG8 and SMG9 forms the complex SMG1C, and up-frameshift 1 (UPF1). UPF1 is the central NMD factor: it is an ATP-dependent RNA helicase containing an abundance of serine and threonine residues located within its N and C termini that function as phosphate-acceptor sites. SMG1 is a phosphatidylinositol 3-kinase-related protein kinase that phosphorylates many of these residues, and this activity is initially held in check by the SMG8-SMG9 complex (Yamashita, 2013). If a terminating ribosome-SURF complex forms ≥ 50–55 nt upstream of an exon-exon junction (Nagy and Maquat, 1998) that has a splicing-generated EJC, then NMD ensues. This is a consequence of an interaction between minimally SMG1 and UPF1 of SURF and the EJC, whose core constituents may be further outfitted with the UPF3X (also called UPF3b) and UPF2 proteins (Shum et al., 2016). When UPF2, anchored to the EJC by UPF3X, interacts with the cysteine- and histidine-rich domain of UPF1, it triggers a conformational change that activates UPF1 helicase activity (Chamieh et al., 2008). This promotes mRNA unwinding and protein removal by UPF1, which are essential for substrate decay. The resulting complex is the "decay-inducing complex" (DECID) (Figure 1B), and it is at this stage that SMG1 phosphorylates the UPF1 termini. Tissue-specific expression of the NMD antagonist UPF3 disrupts this complex and confers fine-tuning NMD activity (Shum et al., 2016).

UPF1 phosphorylation has two consequences: first, phosphorylated UPF1 represses further translation initiation through

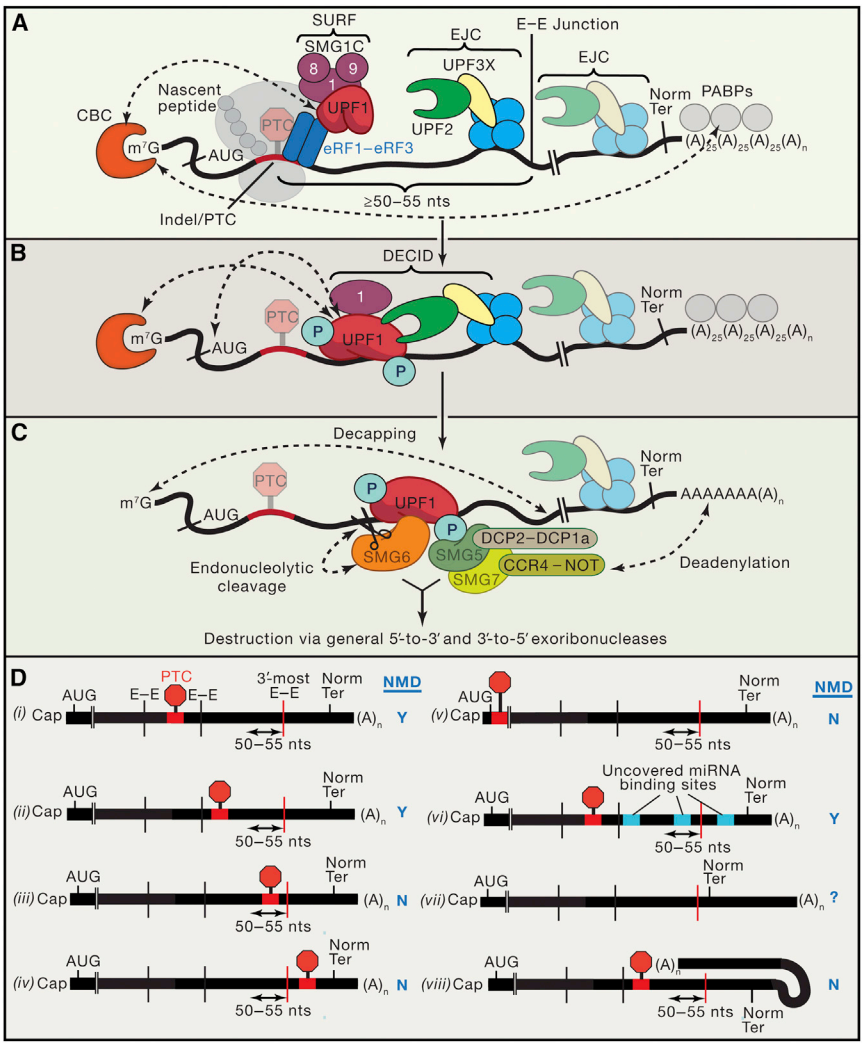

Figure 1. Understanding NMD Can Improve Generating CRISPR/Cas9-Mediated Gene Knockouts

(A) Cytoplasmic mRNA bound by the cap (m^7G)-binding protein complex (CBC) and poly(A)-binding proteins (PABPs) undergoes a pioneer round of translation coupled to inspection by the NMD machinery (Maquat et al., 2010). When the translating ribosome undergoes an altered termination event at the premature termination codon (PTC) (for example, as a result of indel formation), the SMG1 complex (SMG1C) joins factors that also typify normal termination events, namely eukaryotic release factor 1 (eRF1) and eRF3, to form the SURF complex. SMG1C is composed of the UPF1 kinase SMG1 and its repressors SMG8 and SMG9. EJCs typically display UPF2 anchored by UPF3X. Dashed lines signify that CBC promotes SURF formation and that CBC and PABPs may "circularize" the mRNA via bridging proteins. Notably, PABPs are known antagonists of NMD. AUG, initiation codon; Norm Ter, normal termination codon; nts, nucleotides.

(B) As a rule, if an EJC-associated exon-exon junction (E-E) resides ≥50–55 nt downstream of the termination event, then NMD initiates. UPF2 interacts with UPF1, causing a large conformational change that activates UPF1 helicase activity. Within the resulting decay-inducing complex (DECID), SMG1 phosphorylates UPF1. CBC also promotes SMG1-UPF1 binding to the EJC (dashed arrow), and phosphorylated UPF1 inhibits further translation initiation at the AUG codon (dashed arrow).

(C) Phosphorylated UPF1 recruits the SMG6 endonuclease, which cleaves between the PTC and EJC, and SMG5-SMG7 recruits the multisubunit CCR4-NOT complex, causing deadenylation, and/or DCP2-DCP1a, causing decapping. In all cases, destruction of mRNA fragments ensues via general exoribonucleases.

(D) Examples of indel/PTC placement within a typical transcript and its result on NMD as well as additional scenarios. Y, yes NMD occurs; N, no NMD. See text for details.

an interaction with eIF3 that is critical for mRNA decay; second, phosphorylated UPF1 residues form platforms and expose new regions of UPF1 to which RNA-degradative enzymes are either directly or indirectly recruited, resulting in targeted, rather than indiscriminate, mRNA destruction (Figure 1C). Recruitment of the endonuclease SMG6 leads to NMD substrate cleavage in the vicinity of the PTC. This generates unstable decay intermediates that lack either the 5′ cap or 3′ poly(A) and, thus, are susceptible to cellular 5′-to-3′ or 3′-to-5′ exonucleases, respectively. Recruitment of the SMG5-SMG7 complex, which is devoid of intrinsic nuclease activity but has the capacity to bring in the CCR4-NOT deadenylase complex via SMG7 and the DCP2/DCP1a mRNA-decapping proteins via SMG5, also results in target degradation. SMG5 and SMG6 additionally recruit protein phosphatase 2A to dephosphorylate UPF1, recycling it for further rounds of NMD (Lykke-Andersen and Jensen, 2015).

NMD in Action: CRISPR/Cas9-Generated Knockouts

A promising means for creating loss-of-function alleles involves using the highly publicized CRISPR/Cas9 approach (Shalem et al., 2015). We provide only a thumbnail sketch of how this tool is used to knock out protein-encoding genes so as to set the context for how NMD contributes to this process.

Owing to its portability, the microbial type II clustered regularly interspaced short palindromic repeat (CRISPR) adaptive-immune system is now widely used to modify genomic DNA. Requirements for gene manipulation are minimal: the *Streptococcus pyogenes* Cas9 nuclease (SpCas9) need only be co-expressed with a single-guide RNA (sgRNA) composed of a fusion of CRISPR RNA (crRNA) and short *trans*-activating RNA (tracrRNA). Unlike other gene-editing nucleases that rely on amino acids to bring the nuclease to its correct location, SpCas9 uses the sgRNA to locate complementary regions in the genome. The only limitation is the requirement for a protospacer-adjacent motif (PAM), which for SpCas9 is the trinucleotide 5′-NGG-3′. Thus, by obeying the PAM requirement and modifying the 20-nt targeting sequence of the sgRNA, nuclease activity can be directed virtually anywhere within the genome. Many variations of this are emerging, such as using nicking enzymes to separately target Watson and Crick DNA strands to improve