CHAPTER EIGHTEEN

Generation of 3′UTR knockout cell lines by CRISPR/Cas9-mediated genome editing

Sibylle Mitschka[a], Mervin M. Fansler[a,b], and Christine Mayr[a,b,*]
[a]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, United States
[b]Tri-Institutional Training Program in Computational Biology and Medicine, Weill-Cornell Graduate College, New York, NY, United States
*Corresponding author: e-mail address: mayrc@mskcc.org

Contents

Methods in Enzymology, Volume 655
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2021.03.014

Abstract

In addition to the protein code, messenger RNAs (mRNAs) also contain untranslated regions (UTRs). 3′UTRs span the region between the translational stop codon and the poly(A) tail. Sequence elements located in 3′UTRs are essential for pre-mRNA processing. 3′UTRs also contain elements that can regulate protein abundance, localization, and function. At least half of all human genes use alternative cleavage and polyadenylation (APA) to further diversify the regulatory potential of protein functions. Traditional gene editing approaches are designed to disrupt the production of functional proteins. Here, we describe a method that allows investigators to manipulate 3′UTR sequences of endogenous genes for both single- 3′UTR and multi-3′UTR genes. As 3′UTRs can regulate individual functions of proteins, techniques to manipulate 3′UTRs at endogenous gene loci will help to disentangle multi-functionality of proteins. Furthermore, the ability to directly examine the impact of gene regulatory elements in 3′UTRs will provide further insights into their functional significance.

1. Introduction

1.1 The functional significance of 3′UTRs

The 3′ untranslated region (3′UTR) is an integral part of messenger RNAs (mRNAs), encompassing the sequence between the translational stop codon and the poly(A) tail. While the coding region provides cells with the building plan for a particular protein, the 3′UTR can assist in regulating protein abundance, localization, and function (Mayr, 2019). The versatility of 3′UTRs is enabled by a wide array of *trans*-acting factors that interact with the mRNA, including RNA-binding proteins, as well as short and long non-coding RNAs. Together, these interactions support processes that modulate and specify protein function. However, technical limitations have impeded a more detailed understanding of 3′UTR-mediated functions.

Fundamentally, 3′UTRs fulfill two important functions in cells: First, 3′UTRs enable binding of the cleavage and polyadenylation machinery during co-transcriptional mRNA processing. Second, 3′UTRs allow cells to regulate the fate of an mRNA through a variety of post-transcriptional mechanisms.

1.2 Mechanism of mRNA 3′ end processing

Cleavage and polyadenylation are essential steps for the processing of primary transcripts into mature mRNAs, and therefore for protein expression. This generic mechanism of transcript processing occurs co-transcriptionally and is used by all protein-coding genes with the exception of histone genes.

3′UTRs contain the sequences required for the recruitment of the cleavage and polyadenylation machinery. This gene architecture allows the amino acid sequence to be unconstrained by the processing signal. DNA mutations in signals required for 3′ end processing cause a decrease in steady-state mRNA levels with important consequences for human health and fitness (Chang, Yeh, & Yong, 2017; Mariella, Marotta, Grassi, Gilotto, & Provero, 2019). For example, mutations in poly(A) signals in the genes encoding the transcription factor p53 or hemoglobin cause a predisposition to cancer or thalassemia, respectively (Higgs et al., 1983; Orkin, Cheng, Antonarakis, & Kazazian, 1985; Stacey et al., 2011). In addition, more than half of all genes in humans encode more than one functional poly(A) site, thus allowing alternative cleavage and polyadenylation (APA) to occur (Lianoglou, Garg, Yang, Leslie, & Mayr, 2013).

In eukaryotes, 3′ end processing is initiated upon recognition of a poly(A) signal within a favorable sequence context. In vertebrates, the poly(A) signal is a nucleotide hexamer of the sequence A(A/U)UAAA or variants thereof (Gruber et al., 2016; Ulitsky et al., 2012). The canonical hexamers AAUAAA and AUUAAA are usually more efficient at inducing cleavage and polyadenylation than other poly(A) signal variants. Therefore, they can be found at poly(A) sites of single UTR genes as well as distal poly(A) sites of multi-UTR genes. Efficient poly(A) sites also exhibit a higher degree of cross-species conservation (Wang, Nambiar, Zheng, & Tian, 2018; Wang, Zheng, Yehia, & Tian, 2018). In contrast, non-canonical hexamers and weaker sequence contexts are predominantly associated with proximal poly(A) sites of multi-UTR genes.

The polyadenylation machinery is composed of four multiprotein complexes that contact the pre-mRNA through several sequence elements (Tian & Manley, 2017) (Fig. 1A). While the poly(A) signal hexamer is an important element for 3′ end processing in vertebrates, a broader sequence context is required to recruit the cleavage and polyadenylation machinery (Martin, Gruber, Keller, & Zavolan, 2012). U-rich sequences that function as auxiliary elements are found both upstream and downstream of functional cleavage sites (Fig. 1B). The cleavage and polyadenylation specificity factor complex comprised of CPSF-160, CPSF-100, CPSF-73, CPSF-30, FIP1, and WDR33 contacts both the poly(A) signal hexamer as well as U-rich sequences in the vicinity. The auxiliary motif UGUA, which is frequently found upstream of the poly(A) site, is the preferred binding motif for cleavage factor I complex consisting of CFI_m25 plus either CFI_m59 or CFI_m68. The cleavage stimulation factor (CstF77, CstF50, CstF64, and CstF64τ)

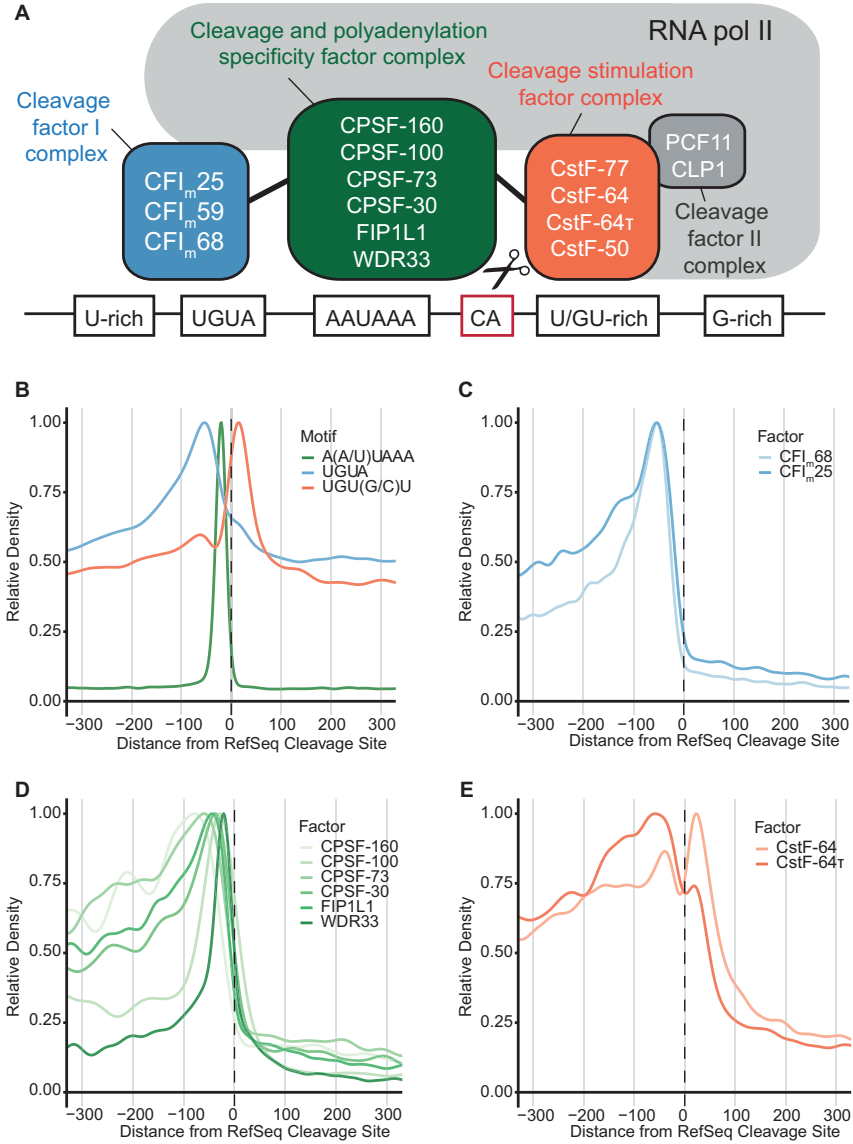

Fig. 1 mRNA cleavage and polyadenylation requires multiple sequence elements surrounding the cleavage site. (A) Schematic of the multiprotein complex responsible for mRNA cleavage and polyadenylation in humans. (B) Sequence context of functional poly(A) sites showing the poly(A) signal hexamer as well as two common auxiliary motifs. The metaplot is aligned to the transcript end in the longest isoform of RefSeq-annotated human genes. (C–E) Metagene analysis shows densities of the binding sites of protein components of the cleavage and polyadenylation machinery determined by CLIP: (C) Cleavage Factor I complex (D) Cleavage and Polyadenylation Specificity Factor complex and (E) Cleavage Stimulation Factor complex. Binding sites were retrieved from the POSTAR2 database and aligned to RefSeq-annotated transcript ends (Zhu et al., 2019).

binds to U- or GU-rich sequences downstream of the cleavage site (Fig. 1C–E). Together, the combination of upstream and downstream sequence elements determines the intrinsic strength of poly(A) sites (Cheng, Miura, & Tian, 2006). In addition, the availability of 3′ end processing factors also impacts usage rates, resulting in diverse 3′UTR isoform expression patterns across cell types and tissues (Lianoglou et al., 2013). APA-mediated differences in 3′UTR isoform expression have also been observed during proliferation, cell stress, immune cell activation, and cancer (Berkovits & Mayr, 2015; Mayr & Bartel, 2009; Sandberg, Neilson, Sarma, Sharp, & Burge, 2008; Zheng et al., 2018).

1.3 3′UTRs as regulators of mRNA stability and protein function

In addition to enabling 3′ end processing, 3′UTRs encode information that impact gene functions through post-transcriptional regulation. Since the sequence elements required for cleavage and polyadenylation are mostly found within the last 100–200 nucleotides upstream of the cleavage site (Fig. 1C–E), additional "non-essential" upstream sequences can evolve to fulfill regulatory functions. Intriguingly, the median 3′UTR length, and thus the available sequence for encoding additional information, positively correlates with morphological complexity in animal evolution (Chen, Chen, Juan, & Huang, 2012; Mayr, 2017).

Traditionally, 3′UTRs have been primarily investigated for their role in regulating protein output. *Cis*-regulatory elements located in 3′UTRs recruit RNA-binding proteins and microRNAs (miRNAs) that can modulate mRNA stability and translation efficiency. The half-life of mammalian mRNAs ranges from a few minutes to several hours. This large range is the result of finely tuned processes that guide post-transcriptional mRNA turnover and translational control. A prime example for these processes is the *Fos* mRNA, whose AU-rich elements located in the 3′UTR cause fast mRNA turnover unless stress-activated cellular pathways antagonize *Fos* mRNA degradation (Otsuka, Fukao, Funakami, Duncan, & Fujiwara, 2019). Similarly, lack of regulation by the AU-rich element in the *Tnf* 3′UTR results in chronic and fatal overproduction of the Tnf-α inflammatory cytokine in mice (Kontoyiannis, Pasparakis, Pizarro, Cominelli, & Kollias, 1999).

Additionally, 3′UTRs are known to control mRNA localization which permits spatially restricted protein synthesis. In yeast, conserved 3′UTR elements have been shown to promote translation of membrane proteins at the

endoplasmic reticulum (Chartron, Hunt, & Frydman, 2016; Loya et al., 2008) and mitochondria (Margeot et al., 2002). In eukaryotic cell models, specialized RNA granules that are intertwined with the endoplasmic reticulum enable efficient transport of proteins to the plasma membrane in a 3′UTR-dependent manner (Ma & Mayr, 2018). Similarly, 3′UTR-mediated association of the mRNA with the cytoskeleton has been shown to promote nuclear localization of the encoded protein (Levadoux, Mahon, Beattie, Wallace, & Hesketh, 1999). In neurons, cytoskeletal proteins are transported to the developing growth cone for local translation via a zipcode element located in 3′UTRs (Zhang, Singer, & Bassell, 1999).

Brain cells are known to express the highest proportion of long isoform transcripts among all cell types (Lianoglou et al., 2013). Strikingly, long and short mRNA isoforms of the same gene are often preferentially sorted into different compartments, for example the soma or the neuropil. A recent study in brain tissue found that degradation rates of long and short 3′UTR mRNA transcripts from the same gene are not correlated, suggesting that local mRNA isoform turnover is directed by independent pathways (Tushev et al., 2018). Due to the local environment during and after translation, the protein generated from a particular 3′UTR isoform can be differentially modified, as has been shown for HMGN5 (Moretti et al., 2015).

Specification of protein function without differences in mRNA localization has been demonstrated for the gene encoding the ubiquitin ligase BIRC3 (c-IAP2). The long 3′UTR of *BIRC3* was shown to facilitate cell surface expression of a transmembrane protein involved in chemokine sensing (Lee & Mayr, 2019). In contrast, both long and short 3′UTR isoforms generate BIRC3 protein involved in apoptosis regulation. Due to the unique protein functions conferred by the long *BIRC3* 3′UTR, leukemia cells benefit from upregulating expression of the long 3′UTR isoform at the expense of the short 3′UTR isoform.

Another example of essential information provided by 3′UTRs are selenoproteins. The mRNAs of these proteins rely on a pair of RNA hairpin structures, called SECIS elements, for the incorporation of the rare amino acid selenocysteine into the protein peptide chain (Berry et al., 1991; Kryukov et al., 2003).

Despite these exciting advances, we are just beginning to systematically decipher the genetic information that is encoded in 3′UTRs. The difficulty of this task is rooted in the fact that 3′UTRs rely on fundamentally different principles of encoding information than the universal triplet code found in coding regions. The most basic information layer in untranslated RNAs is

the nucleotide sequence. However, since the 4-letter nucleotide alphabet has an inherently low information density, sequence alone is often insufficient to confer specificity. RNA secondary structure incorporates additional information and enables increased specificity for recruitment of *trans*-acting factors (Dominguez et al., 2018). A large number of RNA-binding proteins interact with mRNAs on the basis of structure, or a combination of sequence and structure (Dominguez et al., 2018; Sanchez de Groot et al., 2019). However, RNA structures are believed to be more dynamic and flexible than protein folds and in silico predictions remain challenging (Bevilacqua, Ritchey, Su, & Assmann, 2016; Yu, Lu, Zhang, & Hou, 2020). Experimental data in yeast show that the RNA structure of closely related 3′ end isoforms can be dramatically different; this creates unique constraints for accessibility by *trans*-acting factors (Moqtaderi, Geisberg, & Struhl, 2018). In addition, 3′UTRs can contain a large number of *cis*-regulatory elements creating the potential for modulation and diversification of functions in a combinatorial manner (Iadevaia & Gerber, 2015). Finally, additional mRNA features including poly(A) tail length and a large repertoire of RNA modifications can further impact the fate of an mRNA (Jalkanen, Coleman, & Wilusz, 2014; Roundtree, Evans, Pan, & He, 2017). These challenges highlight the need for a more systematic evaluation of 3′UTR-mediated functions to identify and validate principles that govern mRNA-related processes.

1.4 Existing tools to study regulatory 3′UTR functions have systematic biases

Until now, most functional studies have focused on the role of 3′UTRs as regulators of protein output. Indeed, there are numerous examples showing that RNA-binding proteins and miRNAs that interact with 3′UTRs can modulate mRNA stability and translation rates (Matoulkova, Michalova, Vojtesek, & Hrstka, 2012). However, the notion that abundance regulation is the primary function of 3′UTRs is not so much a reflection of actual biology, but rather the result of the current limitations for studying alternative 3′UTR functions.

In particular, reporter assays have been commonly used as a proxy for endogenous 3′UTR behavior. While easy to perform, 3′UTR reporters are uniquely designed to measure quantitative differences in expression and cannot resolve other 3′UTR-related functions. Moreover, reporter assays analyze the impact of a 3′UTR - or even a part of it - outside of its endogenous sequence context. As part of an mRNA molecule, 3′UTRs co-evolved along

with the other mRNA components. By cropping out sequence segments, reporter assays do not consider the potential crosstalk between 3′UTR, coding region and 5′UTR. Folding and accessibility of a 3′UTR sequence can differ in the context of a synthetic reporter gene in comparison to its cognate sequence environment with consequences for post-transcriptional expression regulation (Kristjansdottir, Fogarty, & Grimson, 2015; Lautz, Stahl, & Lang, 2010; Wissink, Fogarty, & Grimson, 2016). The coding region in particular was shown to modulate post-transcriptional expression regulation. Cottrell et al. found that changes in codon optimality impact the degree of repression mediated by miRNAs binding to the 3′UTR (Cottrell, Szczesny, & Djuranovic, 2017). Furthermore, selective testing of the immediate region surrounding the putative target site, while convenient, decreases the sequence length of the 3′UTR. This can introduce an additional bias as previous research has suggested that shorter 3′UTRs exhibit stronger repression in response to overexpression of miRNAs than longer 3′UTRs (Saito & Saetrom, 2012).

Another type of experimental bias is introduced by the overexpression of 3′UTR regulators such as miRNAs and RNA-binding proteins in cells. This approach compares the response of a target gene in the presence or absence of overexpression of a *trans*-acting factor. However, regulators and their target mRNAs exist in a defined concentration equilibrium. As such, the regulatory potential is impacted by the relative concentration of regulator and target, the optimality of the binding site, as well as the abundance of alternative targets (Saito & Saetrom, 2012; Witwer & Halushka, 2016). Overexpression of miRNAs, for example, usually increases their cellular levels to several hundred-fold over their endogenous abundance. As the targeting efficiency is dose-dependent, the pool of mRNA targets may be greatly expanded at concentrations exceeding physiological levels. It is probably for these reasons that many genomic miRNA knockouts have revealed modest or no effects on previously reported target mRNAs (Baek et al., 2008; Miska et al., 2007). This suggests that the number of functionally relevant interactions could be much smaller than previously assumed.

Finally, both miRNAs and RNA-binding proteins target a large number of different mRNAs, which in turn are subject to the regulation by a number of other regulators. Factors implicated in post-transcriptional regulation usually target different mRNAs acting in a common pathway, potentially to amplify a particular biological response (Ben-Hamo & Efroni, 2015; Zanzoni, Spinelli, Ribeiro, Tartaglia, & Brun, 2019). Overexpression of an RNA-binding protein will usually cause expression changes across

hundreds of mRNAs, some of them indirectly. Separating these direct from secondary effects and evaluating their individual physiological significance remains challenging.

Given these limitations, it is perhaps not surprising that some genetic models for 3′UTR-mediated expression regulation have presented evidence contradictory to results obtained by non-endogenous studies (Mitschka & Mayr, 2020; Zhao et al., 2017).

2. Experimental design

2.1 Method overview

In this chapter, we outline strategies that will allow researchers to generate cell models to systematically investigate 3′UTR-dependent functions. Such cell models are created through defined genomic deletions of 3′UTR sequences, while preserving co-transcriptional pre-mRNA processing. We will discuss suitable approaches to target both single- and multi-UTR genes. For multi-UTR genes, we will specifically focus on genes expressing classical tandem 3′UTRs in the last exon that do not alter the amino acid sequence of the encoded protein. Importantly, alternative 3′UTR transcripts can also arise from intronic APA. In contrast to tandem APA, intronic APA events usually affect both the protein coding part as well as the 3′UTR. Usage of intronic APA sites seems to be regulated by a mutual interplay with splicing processes (Lee et al., 2018; Tian, Pan, & Lee, 2007). While genomic deletion strategies can also be applied for deleting these isoforms, we focus here on 3′UTR-dependent functions that do not alter the amino acid sequences of proteins.

CRISPR/Cas9 is now a widely used technique for genome editing. The ability to induce sequence-specific DNA double-strand breaks greatly accelerates the creation of genetically modified cell models and organisms. For traditional protein knockouts, small indel mutations arising at CRISPR/Cas9 cut sites are used to generate frame-shift mutations. However, these regional mutations are not suitable to interrogate functions of non-coding sequences. Alternatively, precision genome editing with designed DNA repair templates remains time-consuming, as the activity of the homology-directed repair (HDR) pathway is generally low in mammalian cells. Instead, our and other previous methods exploit the non-homologous end joining (NHEJ) pathway to generate defined genomic deletions (Bauer, Canver, & Orkin, 2015; Joberty et al., 2020; Mitschka & Mayr, 2020; Thomas et al., 2020; Zhao et al., 2017; Zhu et al., 2016). Specifically, a pair of CRISPR guide

RNAs (gRNA) is used to cut within the 3′UTR, thereby creating a deletion that is flanked by the two cut sites.

2.2 Useful tools for 3′ end annotations

The comprehensive annotation of mRNA 3′ ends is an ongoing project whose progress is most advanced for the human genome. Several useful online tools are available to search for gene-specific 3′ end annotations across different species. Among them are polyASite 2.0 (Herrmann et al., 2020), PolyA_DB3 (Wang, Nambiar, et al., 2018; Wang, Zheng, et al., 2018) and APASdB (You et al., 2015).

Over the last decade, increased sequencing depth has led to an increase in the number of annotated 3′ ends. Of those, very few sites have been validated using orthogonal methods that do not involve sequencing. Artifacts created by internal priming at genomic poly(A)-rich sites remain a challenge for sequencing-based annotation techniques (Gruber et al., 2016). In addition, not all 3′ end sequencing methods have been validated to deliver quantitative results. We therefore advise that any project aiming to delete a 3′UTR or to influence 3′UTR isoform usage should set out to confirm the usage of annotated poly(A) sites. Northern blot analysis can be used to verify and quantify the expression of different 3′UTR isoforms (see also Section 3.6).

2.3 Selection of cell models

While the method described in this chapter can be applied to virtually any cell line, some attention should be paid to the choice of cell model prior to starting the experiment.

2.3.1 Robust expression of the gene of interest

The deletion of the regulatory part of a 3′UTR can result in upregulation, downregulation or no change in expression of the gene of interest compared to the corresponding wild-type cells. In either case, it is the endogenous promoter that drives transcription and the baseline expression level of the gene of interest should be robust enough to enable downstream analysis. For multi-UTR genes, we recommend confirming the expression level of all relevant 3′UTR isoforms in the cell line of choice.

2.3.2 The efficiency of homozygous deletions depends on gene copy numbers

Many commonly used cell lines have aberrant karyotypes with more than two copies per gene. In order to generate a homozygous 3′UTR knockout, all gene copies need to be successfully edited. The presence of additional alleles reduces the chance of obtaining homozygous deletion clones. As is generally true for genome editing, cells with a near-diploid karyotype such as embryonic stem cells, HCT116 (modal chromosome number of 45), U87-MG (44) and WI-38 (46) are usually preferable to cell lines with complex karyotypes including HeLa (82) or MDA-MB-453 (90) cells. Due to local copy number variations, actual gene copy numbers may differ from the overall ploidy state of any given cell line. Therefore, precise karyotype information can help to inform the choice of an appropriate cell model.

In some cases, it might still be desirable to use a cell model with three or more alleles. Screening a larger number of colonies might be necessary to identify homozygous clones. As the deletion efficiency generally decreases with the distance of the two CRISPR/Cas9 cut sites (Bauer et al., 2015), generating larger deletions can further aggravate the problem. Alternatively, homozygous deletion mutants can also be created by performing a second round of transfection and selection after obtaining heterozygous clones.

2.3.3 Requirement for clonal growth

We have noticed that some cell lines are not able to grow as single cells. The addition of cell type-specific conditioned media can help to alleviate this problem. Poor cell survival can also be caused by mechanical stress during the process of cell sorting. As an alternative to growing cells in individual wells, transfected cells can be seeded sparsely in a large culture dish. In this case, individual clones need to be picked under a light microscope for subsequent screening and expansion.

2.4 Developing a gene-specific deletion strategy

CRISPR/Cas9 is now the most widely used tool for genome editing in both cell lines and organisms. Recruitment of the Cas9 nuclease by a programmable guide RNA (gRNA) makes this tool customizable to the specific needs of investigators. Stable binding of the Cas9/gRNA ribonucleoprotein (RNP) complex requires the presence of the cognate protospacer adjacent motif (PAM) downstream of the DNA region complementary to the gRNA. The canonical PAM sequence of the most commonly used Cas9 nuclease

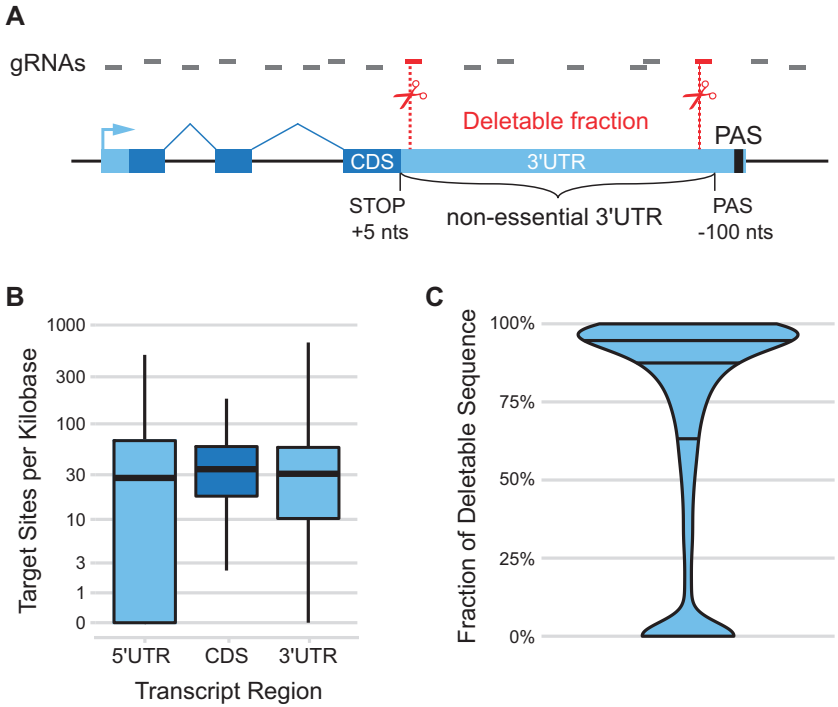

Fig. 2 Density of unique and efficient Cas9 gRNA sequences in human genes. (A) Gene model with unique and efficient gRNAs depicted above. Cleavage sites of a suitable gRNA pair for 3′UTR deletion are highlighted with dotted lines. PAS, poly(A) site; CDS, coding region; nts, nucleotides. (B) Frequency of unique and efficient *Sp*Cas9 CRISPR gRNAs in human RefSeq annotated genes separated by transcript region. Efficient gRNAs were categorized as having a Doench/Fusi score of ≥30. (C) Prediction of the deletable, non-essential 3′UTR sequence portion using gRNA criteria used in B. We defined the non-essential 3′UTR as the region between the translational stop codon +5 nucleotides and -100 nucleotides upstream of the annotated transcript end of the longest RefSeq transcript per gene, excluding 3′UTRs with a total length of less than 200 nucleotides.

variant derived from *Streptococccus pyogenes* (*Sp*Cas) is NGG. In addition, the gRNA sequence should be unique in the genome to ensure target specificity. The fact that 3′UTRs are generally more A/T-rich than coding region sequences creates some constraint on PAM site availability (Fig. 2A and B). However, human 3′UTRs still contain a median density of about 31 unique and efficient Cas9 gRNAs per kilobase of sequence (Concordet & Haeussler, 2018; Doench et al., 2016). Given the availability of gRNAs, our analysis of human 3′UTRs using the longest RefSeq-annotated transcript per gene

showed that the median percentage of removable non-essential 3′UTR sequence is still close to 90% (Fig. 2C).

2.4.1 Deletion of entire 3′UTRs in single- or multi-UTR genes to generate minimal 3′UTRs

For this type of deletion, a gRNA pair is used to delete the entire 3′UTR except for the region involved in 3′ end processing which results in mRNA transcripts with a minimal 3′UTR (Fig. 3).

Upstream gRNAs should have predicted cleavage sites close to the translational stop codon (TAA/TAG/TGA). The Cas9 cleavage site is expected to be located between nucleotide three and four upstream of the PAM site. While NHEJ can be very precise, there is the possibility that additional nucleotides will be deleted due to microhomology-mediated end joining (MMEJ). In order to ensure that this does not result in mutations of the protein coding region, the theoretical Cas9 cleavage site should not be within five nucleotides of the stop codon (Canver et al., 2014; Owens et al., 2019).

In order to avoid interference with 3′ end processing, cleavage by the downstream gRNAs should not impair cleavage and polyadenylation by deleting auxiliary sequence elements (Fig. 1). Unfortunately, the exact sequence context required for 3′ end processing is not precisely defined

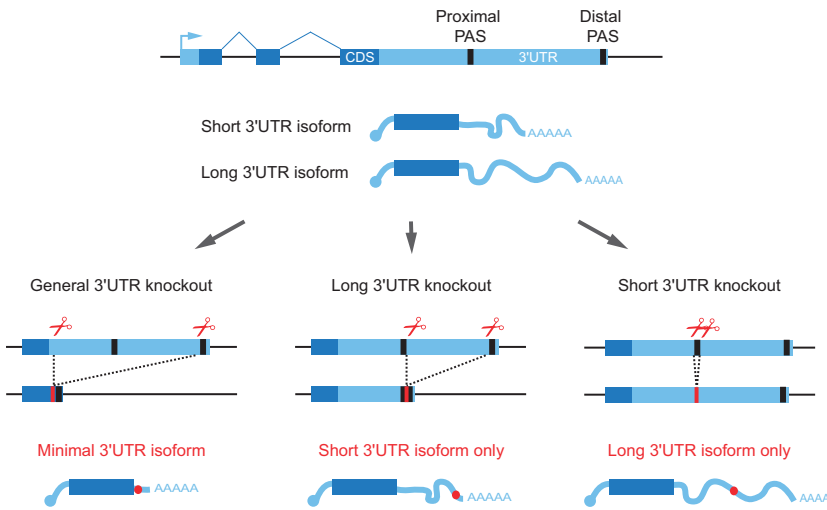

Fig. 3 Overview of deletion strategies to create 3′UTR knockouts in single and multi-UTR genes. Shown are deletion strategies at genomic gene loci as well as the results in the processed mRNAs (indicated by AAAAA, as symbol for the poly(A) tail to denote mRNAs).

for individual genes. As a general rule, we do not recommend positioning the downstream gRNA closer than 100 nucleotides upstream of the cleavage site as auxiliary elements for cleavage factor recruitment are highly enriched in these regions (Fig. 1C–E).

2.4.2 Deletion of the short 3′UTR isoform in multi-UTR genes

This type of mutation is effectively a knockout of the short 3′UTR isoform, while the overall gene-specific transcriptional output is preserved. To this end, two gRNAs with binding sites flanking the poly(A) signal used to produce the short 3′UTR isoform are selected. After the deletion, cleavage and polyadenylation can no longer occur at this site. RNA polymerase II read-through will result in exclusive usage of the downstream poly(A) site(s) and generate only mRNAs with long 3′UTRs. First, the hexamer serving as the putative poly(A) signal needs to be identified which is usually found about 25 nucleotides upstream of the cleavage site (Fig. 1B). Next, non-overlapping pairs of unique gRNAs flanking the poly(A) signal are chosen. Because the poly(A) signal is essential for successful cleavage and polyadenylation, it is not required that the deletion includes the additional sequence context surrounding the poly(A) signal. In fact, designing gRNA pairs with small distances to each other is recommended, as they are more efficient in generating deletions and minimize removal of other regulatory sequence elements.

2.4.3 Deletion of the long 3′UTR isoform in multi-UTR genes

For deleting the long 3′UTR isoform, a gRNA pair is used to delete the genomic region encoding the extended 3′UTR downstream of the proximal poly(A) site (Fig. 3). As a result, the distal, usually stronger poly(A) site will effectively move into close proximity to the proximal site. The combined effect of both poly(A) signals will likely enable highly efficient cleavage and polyadenylation at this location and is expected to preserve overall mRNA expression of the gene. After identifying both poly(A) signals, upstream gRNAs located 3′ of the proximal poly(A) signal and downstream gRNAs located 5′ of the distal poly(A) signal are selected. Because of the substantial length of some 3′UTRs, the desired deletion can be several kilobases in length. Since the deletion efficiency is inversely correlated with the distance of the Cas9 cleavage sites (Bauer et al., 2015), prior testing of individual gRNA efficiencies is particularly important for the success of this strategy (see Section 2.6). In cases with a large distance between the cleavage sites, an alternative approach is to exclusively delete the distal poly(A) site

with a pair of gRNAs flanking the site as described above. However, this approach will reduce total mRNA expression by the amount that the long 3′UTR isoform contributed to it.

2.5 Design of CRISPR gRNAs

A number of different online tools can be used to identify unique and efficient gRNAs in the target region (recently reviewed in Hanna and Doench (2020)). In addition, the UCSC genome browser also enables visualization of all possible Cas9 gRNA locations across the human (hg38) genome (Concordet & Haeussler, 2018; Haeussler et al., 2016). Whenever possible, it is recommended to choose gRNAs with two or more mismatches to other putative target sites to minimize the chances of off-target mutations and translocations. Importantly, our protocol employs vector-based expression systems in which poly(T)-stretches act as a termination signal for the endogenous Pol III. Therefore, the gRNA sense oligonucleotide, located on the strand containing the PAM sequence, is not allowed to contain four or more consecutive T's (Gao, Herrera-Carrillo, & Berkhout, 2018). Finally, gRNAs that will be transfected in pairs should not overlap in their target sequence.

Once the gRNA sequences are selected, adapter nucleotides are added to the gRNA sense and antisense sequences:

gRNA sense oligo: 5′-CACCGNNNNNNNNNNNNNNNNNNN-3′ ($N = 20$).

gRNA antisense oligo: 5′-AAACNNNNNNNNNNNNNNNNNNNC-3′ ($N = 20$).

Oligonucleotides for gRNA cloning can be ordered from any commercial vendor.

2.6 Testing of gRNAs

Ideally, two to three gRNAs should be chosen for each side flanking the desired deletion and tested for their respective cleavage efficiency. Suitable methods for determining cleavage efficiencies include the T7 endonuclease E1 mismatch detection assay, Sanger sequencing followed by tracking of indels by decomposition (TIDE) or targeted deep sequencing. A comparison of these methods and detailed protocols can be found elsewhere (Sentmanat, Peters, Florian, Connelly, & Pruett-Miller, 2018). Importantly, the deletion efficiency in cells is probably limited by the least active gRNA.

3. Protocol

3.1 Required materials

3.1.1 Reagents

- pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (available from Addgene, #42230) or a suitable alternative vector (see Section 3.3)
- gRNA oligos, designed as described in Section 2.5
- *Bbs*I-HF enzyme (NEB, R3539S)
- T4 DNA Ligase (NEB, M0202S)
- DH5α competent bacteria (homemade or commercial)
- LB liquid medium and LB agar plates
- Ampicillin sodium salt (Fisher Scientific, BP1760–5) and kanamycin sulfate (Fisher Scientific, BP906–5)
- Sequencing primer/CBh_rev: CGTCAATGGAAAGTCCCTATTGGC
- Cell line of interest at an early passage (see Section 2.3)
- Cell line-specific culture medium and additives
- *For adherent cells only:* Trypsin-EDTA (0.05%) with phenol-red (Thermo Fisher/GIBCO, 25300062)
- DAPI (Sigma-Aldrich, D9542)
- PBS
- Lipofectamine 2000 (Thermo Fisher/Invitrogen, 1166830) or reagents required for alternative transfection methods (see Section 3.3)
- Opti-MEM I Reduced Serum Medium (Thermo Fisher, 31985062)
- pmaxGFP (Lonza) or a different fluorescent marker plasmid
- FBS (e.g., Thermo Fisher/GIBCO, 26140079)
- QuickExtract DNA Extraction Solution (Lucigen, QE09050)
- *Taq* DNA polymerase with buffer (NEB, M0273)
- Deoxynucleotide (dNTP) solution mix (NEB, N0447S), diluted in water to 2 mM of each nucleotide
- Gene-specific primers for screening PCR, reconstituted and diluted to 10 μM in water (see Section 3.5)
- Gel Loading Dye, Purple 6 × (NEB, B7024S)
- UltraPure agarose (Invitrogen, 16500500)
- Ethidium bromide solution (Thermo Fisher, 17898)
- Zero Blunt TOPO PCR Cloning Kit (Thermo Fisher, 451245)
- Q5 High-fidelity DNA polymerase (NEB, M0491S)

3.1.2 Equipment

- Tabletop centrifuge
- 37 °C incubator for bacteria, equipped with a shaker

- QIAquick PCR Purification Kit (Qiagen 28104)
- QIAprep Spin Miniprep Kit (Qiagen 27104)
- PCR strips
- PCR thermal cycler
- NanoDrop spectrophotometer
- FACS sorting device capable of sorting single cells into cell culture plates under sterile conditions, e.g., BD FACSAria Cell Sorter
- *Optional:* sterile syringes, with luer-lock (e.g., BD 309653)
- *Optional:* sterile syringe filters, 0.2 μm pores (e.g., Corning, 431219)
- Light microscope
- 96-well cell culture plates (e.g., Costar 3585)
- 5 mL polystyrene round-bottom tubes with cell-strainer cap (Falcon, 352235)
- *Optional:* Multichannel pipettes and disposable reagent reservoirs
- Gel electrophoresis equipment
- GelDoc or similar instrument for gel visualization

3.2 Cloning of gRNAs into the pX330 vector

1. Make 10 × annealing buffer: 100 mM Tris pH 8, 10 mM EDTA, 500 mM NaCl.
2. Reconstitute lyophilized gRNA oligos in water at a concentration of 100 μM.
3. Digest pX330 plasmid DNA:
   - 2 μg pX330 vector DNA
   - 2 μL CutSmart buffer
   - 1 μL *Bbs*I-HF enzyme
   - Ad 20 μL sterile water
4. Incubate reaction for one hour at 37 °C.
5. Purify digested plasmid DNA using the QIAquick PCR Purification Kit according to the manufacturer's instructions. Elute silica-bound DNA in 20 μL of water.
6. Determine DNA concentration with a Nanodrop.
7. For annealing of gRNA DNA duplexes in a thermocycler, combine in a PCR strip:
   - 1 μL of sense DNA oligo
   - 1 μL antisense DNA oligo
   - 2 μL 10 × annealing buffer
   - 16 μL water
8. Incubate at 95 °C for 5 min, then decrease temperature by 0.1 °C/s until room temperature is reached.

9. Set up ligation reaction:
   - 50 μg *Bbs*I-digested pX330 plasmid DNA
   - 1 μL of 1:10 dilution of annealed oligo mix, or water as a control
   - 1 μL 10 × T4 DNA ligase reaction buffer
   - 1 μL T4 DNA ligase
   - Add water to a final volume of 10 μL
10. Incubate ligation reaction for one hour at room temperature.
11. Transform competent bacteria with 3 μL of ligation reaction using standard procedures. Plate the bacteria on ampicillin-containing LB agar plates (100 μg/mL final concentration) and grow overnight at 37 °C.
12. The next day pick 3–5 bacterial colonies per plate and inoculate 4 mL ampicillin-containing LB liquid media with bacteria. Grow overnight at 37 °C in a shaking incubator.
13. Perform plasmid DNA extraction from bacteria cultures using the QIAprep Spin Miniprep Kit according to the manufacturer's instructions. Elute plasmid DNA from spin columns in 50 μL water.
14. Measure DNA concentration using a NanoDrop. Validate correct insertion of gRNA sequences by Sanger sequencing using the CBh_rev as sequencing primer.

3.3 Transfection of target cells with gRNAs

The cell line of choice should be in an early passage and maintained in a state of exponential growth through regular splitting. For best results, cell type-specific transfection protocols should be established prior to starting the experiment. Different methods, including lipofection, calcium phosphate transfection or electroporation are suitable to deliver the CRISPR plasmids into target cells.

The relative ratios of the two gRNA plasmids to fluorescent marker plasmid are 10:10:1 (Fig. 4).

This is an example protocol for the transfection of HEK293 cells using Lipofectamine 2000:

1. The day before the transfection split 6×10^5 cells in each well of a 6-well plate containing 2.5 mL growth media per well.
2. The next day, HEK293 cells should have reached 60–90% confluency. Combine 1.2 μg of each of the two pX330-gRNA plasmids with 120 ng of pmaxGFP plasmid and add 140 μL Opti-MEM I Reduced Serum Medium.

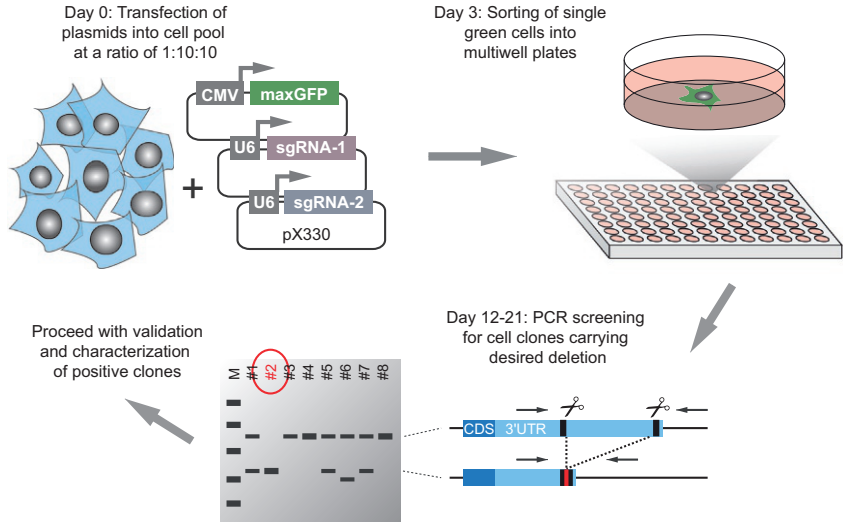

Fig. 4 Schematic overview of the major steps in the protocol to delete 3′UTRs at endogenous gene loci.

3. Add 10 µL of Lipofectamine 2000 to 140 µL Opti-MEM I Reduced Serum Medium, mix briefly, and combine with plasmid-containing solution from previous step. Mix well and incubate for 10 min at room temperature.
4. Add 250 µL of the transfection solution to a 6 well and mix gently by rocking the plate back and forth.
5. The following day change media or split cells if necessary. Continue to grow and split the cells until sorting for GFP-positive cells at day three to five after transfection.

3.4 Seeding of single cells with FACS

Three to five days after transfection, GFP-positive cells are sorted into single wells using FACS (Fig. 4). It is important not to sort the cells too early after transfection, because continued CRISPR/Cas9 activity can lead to further genome editing.

In this protocol, we recommend the use of a FACS sorter to seed transfected single cells into a 96-well culture plate. The ability to select transfected, i.e. fluorescent, cells is particularly useful for cell lines that have low transfection efficiencies. In addition, cell sorting usually yields more accurate seeding of single cells than manual seeding. However, if a cell sorter is unavailable or the cell line of choice is very sensitive to mechanical stress,

cells can instead be counted and seeded at an average concentration of 0.33 cells/well. When performing manual seeding, it is particularly important to ensure that cells are properly resuspended in medium to avoid cell clumps. For manual seeding, the fluorescent marker plasmid can be omitted in the transfection mix (Section 3.3).

1. Prepare the required number of 96-well plates by adding 200 μL of culture medium per well. Prewarm and equilibrate the plates in a cell culture incubator with 5% CO_2 (see tip 1).
2. Three days after transfection, detach transfected cells using Trypsin-EDTA (for adherent cells only).
3. Add excess volume of culture medium with 10% FBS to quench Trypsin and collect cells in a 15 mL conical tube. Centrifuge the cells for 5 min at 250 *g* and thoroughly resuspend the cell pellet in 10% FBS in PBS at 1×10^6/mL. To avoid clogging of the FACS sorter, pass cell suspension through the filter top of a 5 mL round bottom FACS tube. Prepare a sample of untransfected cells in parallel to enable accurate gating during cell sorting. Add DAPI at a final concentration of 0.1 μg/mL to the cell suspension (see tip 2). Keep cell samples on ice until cell sorting.
4. Using a FACS sorter, GFP-positive, DAPI-negative single cells are directly sorted into 96-well culture plates containing 200 μL prewarmed and CO_2-equilibrated medium. In short, intact cells are first selected in a forward versus side scatter (FSC/SSC) plot to exclude small debris and particles with high granularity. This population is further sub-gated for doublet exclusion by plotting area versus height or width (e.g., FSC-A versus FSC-H). Finally, cells with low DAPI signal and high GFP signal are selected for sorting into individual 96 wells (see tip 3).
5. After sorting, continue to culture cells at 37 °C/5% CO_2, with partial media changes every 3–5 days. To change medium, remove 100 μL of medium per well and replace with 100 μL of fresh medium.
6. Depending on the cell type, sufficiently large colonies will form within 12 to 21 days after initial cell seeding. Starting from day 10, daily inspections of the culture plates with a light microscope will help to determine the best time to harvest the cell clones. At that time, wells containing colonies will change media color and individual colonies should have reached a size of 2–5 mm in diameter. Importantly, very few wells should contain more than one independent colony. Whenever possible, wells containing multiple colonies should be excluded from the analysis.

Tips and Troubleshooting

1. To reduce cell death upon culture of single cells, the culture media can be supplemented with up to 50% of conditioned media from the same cell line. To this end, condition the medium for 24–48 h with freshly split cells from a culture of exponentially growing cells. At the time of collection, the conditioned medium should not be exhausted, i.e. the phenol-red pH indicator should not have turned orange/yellow. Clear the medium from floating cells and cell debris by centrifuging at 3000 *g* for 20 min. Afterwards, filter the supernatant through a 0.2 μm syringe filter. Store conditioned media for up to seven days at 4 °C or freeze at −20 °C for longer storage.
2. DAPI is a fluorescent DNA stain that is mostly impermeant to intact living cells. It is used to discriminate between dead (DAPI-positive) and live cells (DAPI-negative) by FACS. As not all FACS sorters are equipped with the appropriate short wavelength lasers needed to excite DAPI, propidium iodide (PI) can be used as an alternative to DAPI. However, PI and GFP have strongly overlapping emission spectra, thus requiring signal compensation during setup.
3. Cells with high GFP protein expression levels (e.g., top 25% of GFP-positive population) are more likely to also express high levels of CRISPR/Cas9 plasmids. It was shown that higher Cas9/gRNA expression levels correlate with higher cleavage efficiency (Hsu et al., 2013).

3.5 PCR screening

PCR is a fast and efficient method to screen large numbers of clones for the desired deletion. For small deletions (<500 nucleotides) a single primer pair flanking the deletion site can be used to amplify both the wild-type and edited alleles (Fig. 4). For larger deletions >1000 nucleotides, amplification of the wild-type allele is often not feasible using crude genomic DNA extracts. Therefore, an additional primer pair spanning one of the two cut sites should be used to identify wild-type alleles. For convenience, a three-primer PCR strategy can be employed to simultaneously detect both allele variants. Optimal PCR conditions should be established in advance (see also tips and troubleshooting below).

When working with suspension cells, start with protocol step 4.

1. Once the cell clones have reached a sufficiently large size, aspirate medium and wash cells with 100 μL PBS per well.

2. Add 30 µL of Trypsin-EDTA solution per well and incubate at 37 °C for 5 min or until cells start to detach.
3. Add 200 µL of culture medium to each well to quench the trypsin digest. Carefully resuspend the cells in medium by repeated pipetting.
4. Transfer 50 µL of resuspended cells into a new 96-well plate containing fresh medium to cultivate as a backup.
5. Transfer 100 µL of the cell solution into a new tube (1.5 mL tube or PCR strip/plate) and spin cells down at 1000 *g* for 3 min.
6. Remove supernatant and resuspend cell pellet in 50 µL QuickExtract DNA extraction solution. Adding more QuickExtract solution is advised if the solution appears visibly cloudy (see also tips and troubleshooting).
7. Place tubes in heat block or thermocycler at 65 °C for 6 min, followed by 98 °C for 2 min. Afterwards, place samples on ice to proceed with setup of the PCR reactions or freeze at −20 °C. For best results, use stored genomic DNA within 3 days and avoid freeze/thaw cycles.
8. Add 2 µL of the crude genomic DNA solution to 18 µL of PCR master mix containing gene-specific primers:
   - 2 µL 10 × Standard *Taq* reaction buffer
   - 1 µL screening primer forward (10 µM)
   - 1 µL screening primer reverse (10 µM)
   - 2 µL dNTP mix (10 × stock with 2 mM of each nucleotide)
   - 0.3 µL *Taq* Polymerase
   - 11.7 µL water
9. Run PCR in a thermocycler with template-specific annealing temperatures and extension times (see troubleshooting).
10. Prepare a 1–1.5% agarose gel (depending on the expected product sizes) in Tris-acetate-EDTA (TAE) buffer containing 0.5 µg/mL ethidium bromide. Add 4 µL gel loading dye to each PCR reaction. Load 12 µL on the agarose gel and run at 10 V/cm.
11. Visualize DNA bands under UV light using a GelDoc or equivalent device.

Tips and troubleshooting:

The convenience of this one-step method for DNA extraction makes it ideal for medium and high throughput screening applications. However, this procedure does not purify genomic DNA like classical genomic DNA extraction methods. Therefore, the PCR reaction needs to be optimized to work well for crude genomic DNA extracts. In general, the amplification of smaller PCR products is more efficient, while amplification of

larger DNA regions can be challenging. If the screening PCR does not work reliably the following steps can be optimized:

1. Choosing alternative primer pairs can often help to improve PCR efficiency.
2. High concentrations of denatured protein in the extraction solution can inhibit the PCR reaction. Diluting the crude genomic DNA mixture with water in ratios from 1:1 to 1:10 can solve this problem. Of note, due to residual RNA and protein contamination, quantification of genomic DNA in extracts using spectroscopy (e.g., Nanodrop) is highly inaccurate.
3. Optimization of PCR reaction conditions should be optimized if the PCR reaction continues to fail. An adjustment of PCR annealing temperatures and testing of PCR reaction additives (e.g., $MgCl_2$, DMSO) can improve yields with structured DNA templates.
4. While a standard *Taq* polymerase is usually sufficient for screening applications, the use of polymerases with higher processivity such as *Pfu* and KOD polymerase can lead to better results with difficult DNA templates.

In some clones, the interpretation of screening PCR results can be complicated when more than the expected number of allele variants are detected. These implausible genotyping results can occur due to a number of different reasons:

1. The assumed gene copy number is incorrect. Check cell line information for local allele copy numbers.
2. Multiple cell clones grew in the same well. When using FACS sorting, this should only happen in a small proportion of wells, but it is more likely to occur when manual seeding was performed. Before seeding into wells, it is essential that cell clumps are thoroughly dissociated. In addition, conditioned medium that has not been filtered can contaminate wells with cells.
3. The cell clones underwent further editing after seeding, leading to a mixed chimeric genotype. Genetic chimerism appears occasionally due to continued CRISPR/Cas9 activity. Make sure to wait at least 3 days (up to 5) after transfection before sorting single cells into individual wells.
4. DNA contamination of PCR reagents can cause additional bands to appear. If the band corresponds to DNA from the wild-type allele, the problem can be difficult to identify. Always run a PCR control without genomic DNA along with genotyping samples.

3.6 Expansion of cell clones and sequence validation

On the basis of the screening PCR, cell clones harboring the desired deletion are selected for expansion. Aliquots of edited cell clones should be frozen as backups at an early passage. In order to minimize the risk of clonal artifacts, phenotypical characterization experiments should always be performed using several cell clones. We also strongly recommend that all alleles of edited clones used for experiments are validated by Sanger sequencing. For this purpose, a target site-specific primer pair flanking the deleted region should be used to amplify the edited region by PCR using a high-fidelity polymerase (e.g., Q5 polymerase from NEB). The screening primers from Section 3.5 can be used at this step. After confirming that the PCR was successful by agarose gel electrophoresis, a small amount of the PCR product is directly sub-cloned using the Zero Blunt TOPO PCR cloning kit according to the manufacturer's instructions. Pick up to ten bacteria transformants to expand in liquid LB culture containing kanamycin. Perform a plasmid miniprep and sequence the PCR insert using one of the available primer sequences in the vector (e.g., T7, M13_for, M13_rev).

3.7 Additional validation by northern blot (recommended)

In order to confirm a change in 3′UTR isoform usage as a result of the genomic deletion, we highly recommend a validation experiment using northern blot analysis. Northern blotting is a reliable method to detect and quantify mRNAs isoform expression from a gene of interest. This is particularly important when multi-UTR genes are analyzed. Northern blot analysis can also help to detect unexpected outcomes of the 3′UTR deletion, including sequence inversions, allele heterogeneity, deficiency in poly(A) site usage due to the deletion, and activation of cryptic downstream poly(A) sites.

We have created a detailed protocol for the detection of 3′UTR isoforms using an optimized northern blot protocol (https://dx.doi.org/10.17504/protocols.io.bqqymvxw). The DNA probe should be complementary to the part of the mRNA common to both wild type and deletion cell lines, such as the coding region. For larger deletions, a second probe targeting the deleted part of the 3′UTR can be used in addition.

4. Related techniques

4.1 Alternative delivery systems for CRISPR/Cas9 gRNA pairs

The CRISPR/Cas9 toolbox continues to evolve rapidly and we anticipate further improvements in the near future. Already today, a number of

different vector designs can be used to create genomic deletions similar to the approach outlined here. For example, a single vector for simultaneous expression of a tandem pair of gRNAs can be used instead of the two-vector system described here. These systems are especially suitable for genome-wide screening approaches, as they allow expression of two gRNAs from a single lentiviral vector (Gasperini et al., 2017; Thomas et al., 2020; Vidigal & Ventura, 2015). CRISPR/Cas9 vector systems that already incorporate a fluorescent marker or an antibiotic resistance cassette can be used as well.

Notably, many labs now transfect cells with in vitro assembled Cas9-gRNA ribonucleoprotein (RNP) complexes. These RNPs can be introduced by electroporation or cationic lipid carriers and can help to circumvent the problem of low transfection efficiencies in some cell systems. In addition, RNPs have been found to be especially effective for DNA editing in a number of model organism through direct delivery by microinjection or electroporation (Chen, Lee, Lee, Modzelewski, & He, 2016; Farboud et al., 2018). We believe that these methods will provide useful variations to our protocol.

4.2 Alternative CRISPR nucleases for 3′UTR editing

A new generation of engineered CRISPR nucleases might soon be able to substitute for *Sp*Cas9 and expand the pool of targetable unique genomic sequences through different PAM-sequence requirements (Chatterjee et al., 2020; Chatterjee, Jakimo, & Jacobson, 2018; Kleinstiver et al., 2015; Legut et al., 2020). The most studied alternative to Cas9 today is called Cpf1. Unfortunately, there is only limited published information regarding the efficiency of Cpf1-mediated genomic deletions (Dumeau et al., 2019). However, we envision potential advantages of the Cpf1 nuclease over Cas9. First, the PAM sequence for Cpf1 is defined as 5′-NTTT-3′, which could be beneficial for targeting A/T-rich regions including 3′UTRs. Second, in contrast to Cas9, Cpf1 cleaves distal to its PAM site (Zetsche et al., 2015). Accordingly, small indel mutations do not automatically preclude further cleavage by Cpf1, which could potentially enhance deletion efficiencies.

4.3 Related genome editing approaches for the analysis of 3′UTR functions

Investigation of 3′UTR-dependent functions through genetic models is still rare and common standards have not been developed. Nevertheless, different approaches that are related to our procedure have been established. For example, gene "knock up" has recently been presented as a gene editing approach for bypassing 3′UTR-dependent *Gdnf* expression regulation

(Mätlik et al., 2019). Here, a CRISPR/Cas9-mediated knock-in of a strong exogenous poly(A) site is used to prematurely induce cleavage and polyadenylation. By preventing incorporation of the original 3′UTR containing repressive *cis*-regulatory elements, this method has been shown to elevate GDNF protein expression *in vivo*. Notably, the addition of Cre-inducible loxP sites creates the potential for a conditional "knock up" allele. However, the use of a strong unrelated poly(A) site can by itself increase mRNA and protein expression by making pre-mRNA cleavage and polyadenylation more efficient than in the endogenous gene independently of additional *cis*-regulatory 3′UTR elements. Nevertheless, gene "knock up" could be a useful tool to investigate the impact of elevated mRNA expression in a cell type-specific manner.

Recently, an elegant study by Bae et al. (2020) provided genetic evidence for the role of the long *Calm1* 3′UTR isoform in the development of mouse dorsal root ganglia and hippocampus. This study investigated a mouse model carrying a concise deletion of the distal *Calm1* poly(A) signal (similar to our strategy for proximal poly(A) sites). Due to the small size of the required deletion, such a strategy is expected to be very efficient in producing homozygous cell clones or embryos. However, while this approach succeeds in eliminating production of the long UTR isoform, it can reduce total mRNA and protein expression to the extent that the long 3′UTR isoform previously contributed to it.

Finally, instead of deleting the entire 3′UTR, a library containing all possible unique gRNA sequences targeting the region of interest can be introduced into target cells. The resulting pool of cells harboring small indel deletions can then be screened for a particular phenotype. Wu et al. (2017) used this type of gRNA "tiling screen" to functionally dissect 3′UTR sequences of several *Drosophila* genes. The main advantage of this method is that it can deliver precise information regarding the spatial distribution of *cis*-regulatory elements. However, in contrast to *Drosophila* genes, most human genes do not provide the necessary gRNA density to sufficiently interrogate most 3′UTR sequences (Pulido-Quetglas et al., 2017). We therefore imagine that this approach could be expanded by using tandem gRNAs that produce an array of small deletions within the 3′UTR sequence.

Acknowledgments

This work was funded by a postdoctoral fellowship from the DFG to S.M., an NIH training Grant (T32GM083937) to M.M.F. and by the NIH Director's Pioneer Award (DP1-GM123454) and the Pershing Square Sohn Cancer Research Alliance to C.M. as well as by the NCI Cancer Center Support Grant (P30 CA008748).

References

Bae, B., Gruner, H. N., Lynch, M., Feng, T., So, K., Oliver, D., et al. (2020). Elimination of Calm1 long 3'-UTR mRNA isoform by CRISPR-Cas9 gene editing impairs dorsal root ganglion development and hippocampal neuron activation in mice. *RNA*, *26*(10), 1414–1430.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64–71.

Bauer, D. E., Canver, M. C., & Orkin, S. H. (2015). Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *Journal of Visualized Experiments*, *95*, e52118.

Ben-Hamo, R., & Efroni, S. (2015). MicroRNA regulation of molecular pathways as a generic mechanism and as a core disease phenotype. *Oncotarget*, *6*(3), 1594–1604.

Berkovits, B. D., & Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, *522*(7556), 363–367.

Berry, M. J., Banu, L., Chen, Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., et al. (1991). Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region. *Nature*, *353*(6341), 273–276.

Bevilacqua, P. C., Ritchey, L. E., Su, Z., & Assmann, S. M. (2016). Genome-wide analysis of RNA secondary structure. *Annual Review of Genetics*, *50*(1), 235–266.

Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., et al. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of Biological Chemistry*, *289*(31), 21312–21324.

Chang, J. W., Yeh, H. S., & Yong, J. (2017). Alternative polyadenylation in human diseases. *Endocrinology and Metabolism (Seoul)*, *32*(4), 413–421.

Chartron, J. W., Hunt, K. C. L., & Frydman, J. (2016). Cotranslational signal-independent SRP preloading during membrane targeting. *Nature*, *536*(7615), 224–228.

Chatterjee, P., Jakimo, N., & Jacobson, J. M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. *Science Advances*, *4*(10), eaau0766.

Chatterjee, P., Lee, J., Nip, L., Koseki, S. R. T., Tysinger, E., Sontheimer, E. J., et al. (2020). A Cas9 with PAM recognition for adenine dinucleotides. *Nature Communications*, *11*(1), 2474.

Chen, C. Y., Chen, S. T., Juan, H. F., & Huang, H. C. (2012). Lengthening of 3'UTR increases with morphological complexity in animal evolution. *Bioinformatics*, *28*(24), 3178–3181.

Chen, S., Lee, B., Lee, A. Y., Modzelewski, A. J., & He, L. (2016). Highly efficient mouse genome editing by CRISPR ribonucleoprotein electroporation of zygotes. *The Journal of Biological Chemistry*, *291*(28), 14457–14467.

Cheng, Y., Miura, R. M., & Tian, B. (2006). Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics*, *22*(19), 2320–2325.

Concordet, J.-P., & Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, *46*(W1), W242–W245.

Cottrell, K. A., Szczesny, P., & Djuranovic, S. (2017). Translation efficiency is a determinant of the magnitude of miRNA-mediated repression. *Scientific Reports*, *7*(1), 14884.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191.

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., et al. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Molecular Cell*, *70*(5), 854–867 (e859).

Dumeau, C. E., Monfort, A., Kissling, L., Swarts, D. C., Jinek, M., & Wutz, A. (2019). Introducing gene deletions by mouse zygote electroporation of Cas12a/Cpf1. *Transgenic Research*, *28*(5–6), 525–535.

Farboud, B., Jarvis, E., Roth, T. L., Shin, J., Corn, J. E., Marson, A., et al. (2018). Enhanced genome editing with Cas9 ribonucleoprotein in diverse cells and organisms. *Journal of Visualized Experiments*, *135*, 57350.

Gao, Z., Herrera-Carrillo, E., & Berkhout, B. (2018). Delineation of the exact transcription termination signal for type 3 polymerase III. *Molecular Therapy. Nucleic Acids*, *10*, 36–44.

Gasperini, M., Findlay, G. M., McKenna, A., Milbank, J. H., Lee, C., Zhang, M. D., et al. (2017). CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *American Journal of Human Genetics*, *101*(2), 192–205.

Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., et al. (2016). A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Research*, *26*(8), 1145–1159.

Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J. B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), 148.

Hanna, R. E., & Doench, J. G. (2020). Design and analysis of CRISPR-Cas experiments. *Nature Biotechnology*, *38*(7), 813–823.

Herrmann, C. J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A. J., & Zavolan, M. (2020). PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Research*, *48*(D1), D174–D179.

Higgs, D. R., Goodbourn, S. E., Lamb, J., Clegg, J. B., Weatherall, D. J., & Proudfoot, N. J. (1983). Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature*, *306*(5941), 398–400.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832.

Iadevaia, V., & Gerber, A. P. (2015). Combinatorial control of mRNA fates by RNA-binding proteins and non-coding RNAs. *Biomolecules*, *5*(4), 2207–2222.

Jalkanen, A. L., Coleman, S. J., & Wilusz, J. (2014). Determinants and implications of mRNA poly(A) tail size- -does this protein make my tail look big? *Seminars in Cell & Developmental Biology*, *34*, 24–32.

Joberty, G., Falth-Savitski, M., Paulmann, M., Bosche, M., Doce, C., Cheng, A. T., et al. (2020). A tandem guide RNA-based strategy for efficient CRISPR gene editing of cell populations with low heterogeneity of edited alleles. *The CRISPR Journal*, *3*(2), 123–134.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485.

Kontoyiannis, D., Pasparakis, M., Pizarro, T. T., Cominelli, F., & Kollias, G. (1999). Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: Implications for joint and gut-associated Immunopathologies. *Immunity*, *10*(3), 387–398.

Kristjansdottir, K., Fogarty, E. A., & Grimson, A. (2015). Systematic analysis of the Hmga2 3' UTR identifies many independent regulatory sequences and a novel interaction between distal sites. *RNA*, *21*(7), 1346–1360.

Kryukov, G. V., Castellano, S., Novoselov, S. V., Lobanov, A. V., Zehtab, O., Guigó, R., et al. (2003). Characterization of mammalian Selenoproteomes. *Science*, *300*(5624), 1439.

Lautz, T., Stahl, U., & Lang, C. (2010). The human c-fos and TNFalpha AU-rich elements show different effects on mRNA abundance and protein expression depending on the reporter in the yeast Pichia pastoris. *Yeast*, *27*(1), 1–9.

Lee, S. H., & Mayr, C. (2019). Gain of additional BIRC3 protein functions through 3'-UTR-mediated protein complex formation. *Molecular Cell*, *74*(4), 701–712 (e709).

Lee, S. H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C. S., & Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature*, *561*(7721), 127–131.

Legut, M., Daniloski, Z., Xue, X., McKenzie, D., Guo, X., Wessels, H. H., et al. (2020). High-throughput screens of PAM-flexible Cas9 variants for gene knockout and transcriptional modulation. *Cell Reports*, *30*(9), 2859–2868 (e2855).

Levadoux, M., Mahon, C., Beattie, J. H., Wallace, H. M., & Hesketh, J. E. (1999). Nuclear import of Metallothionein requires its mRNA to be associated with the perinuclear cytoskeleton. *Journal of Biological Chemistry*, *274*(49), 34961–34966.

Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & Development*, *27*(21), 2380–2396.

Loya, A., Pnueli, L., Yosefzon, Y., Wexler, Y., Ziv-Ukelson, M., & Arava, Y. (2008). The 3′-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. *RNA (New York, N.Y.)*, *14*(7), 1352–1365.

Ma, W., & Mayr, C. (2018). A Membraneless organelle associated with the endoplasmic reticulum enables 3′UTR-mediated protein-protein interactions. *Cell*, *175*(6), 1492–1506 (e1419).

Margeot, A., Blugeon, C., Sylvestre, J., Vialette, S., Jacq, C., & Corral-Debrinski, M. (2002). In Saccharomyces cerevisiae, ATP2 mRNA sorting to the vicinity of mitochondria is essential for respiratory function. *The EMBO Journal*, *21*(24), 6893–6904.

Mariella, E., Marotta, F., Grassi, E., Gilotto, S., & Provero, P. (2019). The length of the expressed 3′ UTR is an intermediate molecular phenotype linking genetic variants to complex diseases. *Frontiers in Genetics*, *10*, 714.

Martin, G., Gruber, A. R., Keller, W., & Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Reports*, *1*(6), 753–763.

Mätlik, K., Olfat, S., Garton, D. R., Montaño-Rodriguez, A., Turconi, G., Porokuokka, L. L., et al. (2019). Gene Knock Up via 3′UTR editing to study gene function in vivo. *bioRxiv*, 775031.

Matoulkova, E., Michalova, E., Vojtesek, B., & Hrstka, R. (2012). The role of the 3′ untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biology*, *9*(5), 563–576.

Mayr, C. (2017). Regulation by 3′-untranslated regions. *Annual Review of Genetics*, *51*(1), 171–194.

Mayr, C. (2019). What are 3′ UTRs doing? *Cold Spring Harbor Perspectives in Biology*, *11*(10).

Mayr, C., & Bartel, D. P. (2009). Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, *138*(4), 673–684.

Miska, E. A., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., et al. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genetics*, *3*(12), e215.

Mitschka, S., & Mayr, C. (2020). Endogenous p53 expression in human and mouse is not regulated by its 3′UTR. *bioRxiv*, 2020.2011.2023.394197.

Moqtaderi, Z., Geisberg, J. V., & Struhl, K. (2018). Extensive structural differences of closely related 3′ mRNA isoforms: links to Pab1 binding and mRNA stability. *Molecular Cell*, *72*(5), 849–861.e846.

Moretti, F., Rolando, C., Winker, M., Ivanek, R., Rodriguez, J., Von Kriegsheim, A., et al. (2015). Growth cone localization of the mRNA encoding the chromatin regulator HMGN5 modulates neurite outgrowth. *Molecular and Cellular Biology*, *35*(11), 2035–2050.

Orkin, S. H., Cheng, T. C., Antonarakis, S. E., & Kazazian, H. H., Jr. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *The EMBO Journal*, *4*(2), 453–456.

Otsuka, H., Fukao, A., Funakami, Y., Duncan, K. E., & Fujiwara, T. (2019). Emerging evidence of translational control by AU-rich element-binding proteins. *Frontiers in Genetics*, *10*, 332.

Owens, D. D. G., Caulder, A., Frontera, V., Harman, J. R., Allan, A. J., Bucakci, A., et al. (2019). Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Research*, *47*(14), 7402–7417.

Pulido-Quetglas, C., Aparicio-Prat, E., Arnan, C., Polidori, T., Hermoso, T., Palumbo, E., et al. (2017). Scalable design of paired CRISPR guide RNAs for genomic deletion. *PLoS Computational Biology*, *13*(3), e1005341.

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell*, *169*(7), 1187–1200.

Saito, T., & Saetrom, P. (2012). Target gene expression levels and competition between transfected and endogenous microRNAs are strong confounding factors in microRNA high-throughput experiments. *Silence*, *3*, 3.

Sanchez de Groot, N., Armaos, A., Grana-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., et al. (2019). RNA structure drives interaction with proteins. *Nature Communications*, *10*(1), 3246.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer MicroRNA target sites. *Science*, *320*, 1643–1647.

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., & Pruett-Miller, S. M. (2018). A survey of validation strategies for CRISPR-Cas9 editing. *Scientific Reports*, *8*(1), 888.

Stacey, S. N., Sulem, P., Jonasdottir, A., Masson, G., Gudmundsson, J., Gudbjartsson, D. F., et al. (2011). A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nature Genetics*, *43*(11), 1098–1103.

Thomas, J. D., Polaski, J. T., Feng, Q., De Neef, E. J., Hoppe, E. R., McSharry, M. V., et al. (2020). RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nature Genetics*, *52*(1), 84–94.

Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature Reviews. Molecular Cell Biology*, *18*(1), 18–30.

Tian, B., Pan, Z., & Lee, J. Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Research*, *17*(2), 156–165.

Tushev, G., Glock, C., Heumuller, M., Biever, A., Jovanovic, M., & Schuman, E. M. (2018). Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron*, *98*(3), 495–511 (e496).

Ulitsky, I., Shkumatava, A., Jan, C. H., Subtelny, A. O., Koppstein, D., Bell, G. W., et al. (2012). Extensive alternative polyadenylation during zebrafish development. *Genome Research*, *22*(10), 2054–2066.

Vidigal, J. A., & Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nature Communications*, *6*, 8083.

Wang, R., Nambiar, R., Zheng, D., & Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Research*, *46*(D1), D315–D319.

Wang, R., Zheng, D., Yehia, G., & Tian, B. (2018). A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Research*, *28*(10), 1427–1441.

Wissink, E. M., Fogarty, E. A., & Grimson, A. (2016). High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics*, *17*, 177.

Witwer, K. W., & Halushka, M. K. (2016). Toward the promise of microRNAs - enhancing reproducibility and rigor in microRNA research. *RNA Biology*, *13*(11), 1103–1116.

Wu, Q., Ferry, Q. R. V., Baeumler, T. A., Michaels, Y. S., Vitsios, D. M., Habib, O., et al. (2017). In situ functional disSection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. *Nature Communications*, *8*(1), 2109.

You, L., Wu, J., Feng, Y., Fu, Y., Guo, Y., Long, L., et al. (2015). APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucleic Acids Research*, *43*(Database issue), D59–D67.

Yu, B., Lu, Y., Zhang, Q. C., & Hou, L. (2020). Prediction and differential analysis of RNA secondary structure. *Quantitative Biology*, *8*(2), 109–118.

Zanzoni, A., Spinelli, L., Ribeiro, D. M., Tartaglia, G. G., & Brun, C. (2019). Post-transcriptional regulatory patterns revealed by protein-RNA interactions. *Scientific Reports*, *9*(1), 4302.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759–771.

Zhang, H. L., Singer, R. H., & Bassell, G. J. (1999). Neurotrophin regulation of β-actin mRNA and protein localization within growth cones. *Journal of Cell Biology*, *147*(1), 59–70.

Zhao, W., Siegel, D., Biton, A., Tonqueze, O. L., Zaitlen, N., Ahituv, N., et al. (2017). CRISPR-Cas9-mediated functional disSection of 3'-UTRs. *Nucleic Acids Research*, *45*(18), 10800–10810.

Zheng, D., Wang, R., Ding, Q., Wang, T., Xie, B., Wei, L., et al. (2018). Cellular stress alters 3′UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nature Communications*, *9*(1), 2268.

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., et al. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature Biotechnology*, *34*, 1279–1286.

Zhu, Y., Xu, G., Yang, Y. T., Xu, Z., Chen, X., Shi, B., et al. (2019). POSTAR2: Deciphering the post-transcriptional regulatory logics. Nucleic Acids Research, 47(D1), D203-D211.