

# CRISPR/Cas9-based mutagenesis frequently provokes on-target mRNA misregulation

Rubina Tuladhar<sup>1</sup>, Yunku Yeu<sup>2</sup>, John Tyler Piazza<sup>1</sup>, Zhen Tan<sup>3</sup>, Jean Rene Clemenceau<sup>2</sup>, Xiaofeng Wu<sup>1</sup>, Quinn Barrett<sup>1</sup>, Jeremiah Herbert<sup>1</sup>, David H. Mathews<sup>3</sup>, James Kim<sup>4,5</sup>, Tae Hyun Hwang<sup>2\*</sup>, and Lawrence Lum<sup>1\*</sup>

<sup>1</sup>Department of Cell Biology, <sup>4</sup>Department of Internal Medicine, and <sup>5</sup>Hamon Center for Therapeutic Oncology Research Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas TX 75390 USA

<sup>2</sup>Department of Quantitative Health Sciences, Cleveland Clinic Lerner Research Institute, Cleveland, OH 44195

<sup>3</sup>Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester NY, 14642

\*corresponding authors: [hwangt@ccf.org](mailto:hwangt@ccf.org) and [lawrence.lum@utsouthwestern.edu](mailto:lawrence.lum@utsouthwestern.edu)

The introduction of insertion-deletions (INDELs) by activation of the error-prone non-homologous end-joining (NHEJ) pathway underlies the mechanistic basis of CRISPR/Cas9-directed genome editing. The ability of CRISPR/Cas9 to achieve gene elimination (knockouts) is largely attributed to the emergence of a pre-mature termination codon (PTC) from a frameshift-inducing INDEL that elicits non-sense mediated decay (NMD) of the mutant mRNA. Yet, the impact on gene expression as a consequence of CRISPR/Cas9-introduced INDELs into RNA regulatory sequences has been largely left uninvestigated. By tracking DNA-mRNA-protein relationships in a collection of CRISPR/Cas9-edited cell lines that harbor frameshift-inducing INDELs in various targeted genes, we detected the production of foreign mRNAs or proteins in ~50% of the cell lines. We demonstrate that these aberrant protein products are derived from the introduction of INDELs that promote internal ribosomal entry, convert pseudo-mRNAs into protein encoding molecules, or induce exon skipping by disruption of exon splicing enhancers (ESEs). Our results using CRISPR/Cas9-introduced INDELs reveal facets of an epigenetic genome buffering apparatus that likely evolved to mitigate the impact of such mutations introduced by pathogens and aberrant DNA damage repair, and that more recently pose challenges to manipulating gene expression outcomes using INDEL-based mutagenesis.

## INTRODUCTION

Technologies enabling the directed introduction of double-stranded DNA breaks such as CRISPR/Cas9 have transformed our ability to systematically identify DNA sequences important in biology<sup>1, 2</sup>. The repair of these double-stranded breaks by non-homologous end-joining (NHEJ) results in insertion-deletions (INDELs) of unpredictable length that upon introduction into exonic sequences could alter the coding frame and install a pre-mature termination codon (PTC). Ribosomes that encounter a PTC in nascent mRNAs, recognized by the assembly of a complex that includes proteins from the ribosome and a 3' exon-splice junction complex, induces the destruction of the mutant mRNA<sup>3, 4</sup>. On the other hand, INDELs that preserve the reading frame may yield proteins with altered sequences and thus shed light on determinants important for its function<sup>5</sup>.

Exonic sequences are laden with regulatory features that control many facets of the mRNA lifecycle including splicing and folding, two mRNA attributes that influence protein sequence composition and sites of initiation/termination, respectively<sup>6-8</sup>. Yet, the frequency with which these elements once impacted by INDELs influence gene expression outcomes remains mostly unknown. Another potential obstacle to precision gene-editing using INDEL-type mutagenesis is

the presence of pseudo-mRNAs, mRNAs harboring a PTC that can nevertheless incorporate introduced INDELs thus altering their potential to produce proteins<sup>9</sup>.

To determine the extent to which these molecular events confound our ability to predict gene expression outcomes from CRISPR/Cas9 editing, we took inventory of the post-transcriptional and -translational effects of frameshift-inducing INDELs in a panel of CRISPR-edited cell lines. We observed changes in the array of transcripts or proteins expressed from CRISPR-targeted genes in ~50% of the cell lines studied. A mechanistic account of these phenomena is presented here.

## RESULTS

### Prevalence of unanticipated gene expression outcomes following on-target CRISPR/Cas9-mediated gene editing in commercially available cell lines.

To service several ongoing research programs, we had assembled a panel of commercially available HAP1 cell lines harboring frameshift-inducing INDELs that presumably eliminate effective protein production from the targeted gene by promoting non-sense mediated decay (NMD) of the encoded mRNA (**Fig. 1A; Supp. Table 1**). HAP1 cells harbor a single copy of each chromosome thus reducing the challenges frequently associated with achieving homozygosity in diploid cells for genetic studies<sup>10</sup>. To confirm the effects of the INDEL on target gene expression, we used two antibodies each recognizing a different epitope within the targeted protein (**Fig. 1B; Supp. Table 2**). We observed in some cell lines the anticipated loss of protein presumably due to the introduced INDEL but in other instances the appearance of novel proteins detectable by Western blot analysis using a single or both antibodies (4/13 cell lines or ~30%; **Fig. 1B**). For example, in the case of the *TOP1*, *SIRT1*, *CTNNB1*, and *LRP6* knockout cell lines, we observed the substitution of the canonical protein for a faster migrating novel protein detected by Western blot analysis.

Given our inability to account for the emergence of these novel proteins based on the annotated genetic alteration introduced by CRISPR/Cas9, we next examined the effects of the INDEL on mRNA splicing given that exonic sequences harbor splicing regulatory elements<sup>8, 11, 12</sup> (**Fig. 1C; Supp. Table 3**). In the case of the *TOP1* knockout cell line where we had observed the appearance of a novel *TOP1* protein, we also witnessed the emergence of a novel mRNA species. Sequencing a cDNA-derived amplicon from the novel splice variant revealed the absence of the INDEL-containing exon suggesting the mutant protein was generated by an INDEL-induced exon exclusion event (**Supp. Table 4**). The truncated *TOP1* protein exhibited disproportionate loss of protein in the nucleus compared with in the cytoplasm, but nevertheless retained catalytic activity (**Fig. 2A-C**). The retention of catalytic activity by the truncated *TOP1* protein is consistent with the designation of *TOP1* as an essential gene in HAP1 cells from a gene trap mutagenesis screen that would preclude its elimination in viable cells<sup>10 13</sup>. In the case of the *VPS35* and *TLE3* cell lines, we observed changes in the splice variants harboring the CRISPR-targeted exons although no detectable novel proteins emerged (see **Fig. 1C**).

In contrast to the *TOP1* clones, the *CTNNB1* and *LRP6* cell lines exhibited no detectable change in mRNA splicing associated with the targeted exons suggesting the novel proteins are a consequence of alternative translation initiation (ATI) events presumably induced by the introduced INDELs (see **Fig. 1C**). Consistent with this hypothesis, the mutant *LRP6* protein is not glycosylated perhaps as a consequence of default expression in the cytoplasm in the absence of its N-terminal signal sequence (**Supp. Fig. 1A, 1C**). Similarly, the novel  $\beta$ -catenin protein co-migrates on SDS PAGE with an engineered  $\beta$ -catenin protein initiating from Met88 (**Supp. Fig. 1B**). In summary, in ~50% of CRISPR-edited cells acquired from a commercial source we observed unexpected changes in protein expression or mRNA splicing that challenge the notion that these reagents could be used to report the cellular effects of complete genetic

ablation (**Fig. 2D**). Although not investigated here, conceivably the mutant proteins could also contribute to neomorphic cellular phenotypes.

### **ATI and pseudo-mRNAs compromise a CRISPR/Cas9 based strategy to achieve a gene knockout.**

We had complemented our efforts to generate cells genetically null for various genes-of-interest with *de novo* CRISPR/Cas9-based gene targeting projects. As part of our focus on the tumor suppressor kinase LKB1, we observed the emergence of unexpected protein products – both smaller and larger proteins than the canonical protein - that were not readily explained by the presence of CRISPR-introduced INDELs (**Fig. 3A-C**). Given the INDELs created in *LKB1* are localized to the first protein coding exon (**Fig. 3D**) and the antibody recognizing the C- but not the N-terminus epitope reported the shortened LKB1 protein on SDS PAGE (see **Fig. 3B-C**), we concluded that an ATI event induced by CRISPR/Cas9-introduced INDELs likely resulted in an LKB1 protein lacking a portion of its N terminal sequence (ATI LKB1 protein).

We also noted in MIA, but not HAP1 cells, a slower migrating protein recognized by LKB1 antibodies emerged in CRISPR/Cas9-edited clones with frameshift-inducing INDELs (see **Fig. 3C**; Super LKB1 protein). The appearance of Super LKB1 protein coincided with the appearance of a new mRNA splice variant that contained a 131 bp exon not included in the transcript that encodes the canonical LKB1 protein (**Fig. 3E**). Consistent with this exon belonging to an *LKB1* pseudo-mRNA not previously annotated in MIA cells, the addition of cycloheximide (CHX) to disrupt NMD in parental MIA cells resulted in the emergence of an *LKB1* splice variant that includes this exon (**Supp. Fig. 2A**). Thus, the same INDELs that induced a frameshift in the canonical transcript now removed a PTC from an *LKB1* pseudo-mRNA and capacitated it for protein production (see **Fig. 3E**). We noted that HAP1 cells did not transcribe an mRNA containing this exon thus our introduction of INDELs into exon 1 did not result in the production of the Super LKB1 protein (**Supp. Fig. 2B**). An understanding of both the transcriptome and the pseudo-transcriptome in cells is thus critical to anticipating the net effect of frameshift-inducing INDELs introduced by CRISPR/Cas9<sup>9</sup>.

To understand how CRISPR/Cas9-introduced INDELs may have produced the ATI LKB1 protein, we generated cDNAs harboring each of the two INDELs that were found in the edited cells expressing these proteins (MIA cells, clone M2) in order to remove any potential contribution of altered mRNA splicing to the production of the mutant proteins (**Fig. 3F**). When either INDEL was introduced into the canonical *LKB1* cDNA sequence, we observed the expression of a protein that co-migrated with the ATI LKB1 protein. This unexpected protein product also co-migrated with an engineered protein that initiates at methionine 51 (see **Fig. 3F**). We noted that a cDNA harboring the 1 bp insertion that provoked the ATI LKB1 protein was not associated with a redistribution of the initiation site suggesting leaky scanning is not likely responsible for the ATI event (**Supp. Fig. 3**). Given that the PTC is located 3' to the predicted ATI site, we also assume this is not a re-initiation phenomenon in which translational termination is followed by the ribosome re-launching translation at a secondary AUG codon<sup>14</sup>. At the same time, we also evaluated the effects of these mutations on cDNAs that encode the predicted pseudo-mRNA sequence (with the 131 bp additional exon). As anticipated, we observed the emergence of proteins that co-migrated with Super LKB1 protein given that either mutation with the additional 131 bp sequence would eradicate a PTC present in the unaltered pseudo-mRNA sequence (see **Fig. 3F**).

We next considered whether alternative secondary structures of the mutant mRNAs might induce this ATI at methionine 51 (Met51; **Fig. 3G**). Hairpin structures in the 5' untranslated region (UTR) near the 5'm7G cap have been shown to inhibit translation initiation<sup>15-17</sup> although very stable hairpins also can inhibit translation initiation farther away from the 5'm7G cap<sup>16</sup>. Conversely, hairpin structures 3' to AUG start codons can facilitate translation initiation, even at non-canonical start sites<sup>18-20</sup>. We used TurboFold II<sup>21</sup>, an algorithm that predicts RNA secondary

structures conserved in a set of homologous RNA sequences and aligns the sequences, to model the possible RNA structures of several mutant clones that generated the Met51 initiated protein compared to those that did not (see **Fig. 3G, Supp. Fig. 4**). Among the sequences that produced the smaller protein, a common 3' hairpin downstream from the methionine 51 ATI site was identified at a distance that has been previously described to facilitate translation initiation<sup>18, 20</sup>. The 3' hairpin may facilitate ATI by extending the pause time of the 40S ribosomes and allowing them to engage the Met51 start site. In the clones that do not exhibit ATI, the highly stable 5' hairpin upstream of the ATI site may cause stalling of 40S ribosomes and prevent access to the Met51 start site whereas the 80S elongation ribosome that started translation at the canonical AUG start site can resolve the 5' hairpin to continue translation of the LKB1 protein as previously described<sup>16</sup>. Consistent with this model, an LKB1 cDNA that includes substitutions at the same position as the INDEL failed to produce the ATI LKB1 protein and a predicted 3' hairpin structure (**Supp. Fig. 5**). Taken together, our data suggests that ATI could be a consequence of altered mRNA structure induced by INDELs. Given targeting the gene sequences proximal to the initiating ATG is considered a robust approach for gene elimination campaigns using CRISPR/Cas9<sup>22</sup>, our observation may be applicable to other CRISPR-editing projects including those that yielded the *CTNNB1* and *LRP6* HAP1 clones (see **Fig. 1**).

### **ATI suppresses NMD and facilitates expression of a C-terminally truncated protein from an mRNA harboring a PTC.**

Despite the introduction of a frameshift-promoting INDEL in *LKB1*, we presumed that an ATI event, which restores codon usage to its native phase, would fail to elicit NMD during the pioneer round of translation. At the same time, having avoided destruction, the mutant mRNA is now able to support repeated rounds of translation including presumably short polypeptides initiating at the canonical start site and ending at the PTC. Given our initial Western blot analysis of the LKB1 CRISPR-edited clones did not capture low molecular proteins (see **Fig. 3B,C**), re-examination of LKB1 proteins in our CRISPR-edited clones indeed revealed the presence of a small LKB1 polypeptide. This protein (short LKB1) co-migrates with an engineered protein that initiates at the canonical start site but terminates at the presumed PTC introduced by the INDEL (**Supp. Fig. 6A**).

We compared the effects on mRNA stability of an INDEL associated with ATI with an INDEL that yielded no detectable LKB1 polypeptides (**Supp. Fig. 6B-C**) in order to determine if ATI suppressed NMD as a potential mechanism for promoting C-terminally truncated proteins. Comparing the levels of the two LKB1 mRNAs, we observed greater loss of the mRNA in the CRISPR-edited clone lacking any detectable ATI events (**Supp. Fig. 6D**). We observed little difference induced by CHX exposure in LKB1 mRNA abundance in the ATI-associated cells when compared to parental cells suggesting that NMD is not acting on the mRNA with an ATI-provoking mutation (see **Supp. Fig. 6D**). On the other hand, in the case of the CRISPR-edited cell line that expresses no LKB1 polypeptides, we observed a 10-fold change in LKB1 mRNA in the presence of CHX suggesting the mutant mRNA in this case is subject to robust NMD action (**Supp. Fig. 6D**). In total, we observed the production of three polypeptides in lieu of the canonical LKB1 protein following the introduction of a frameshift-inducing INDEL: Super LKB1, ATI LKB1, and Short LKB1 (**Supp. Fig. 6E**). More generally, our observations also suggest that introducing INDELs early in the transcript increases the potential for an ATI event that is able to clear off all of the splice junction complexes during the pioneer round thus enabling the synthesis of polypeptides with truncations in the C-terminal sequence.

### **Exon symmetry influences CRISPR/Cas9 outcomes following aberrant exon skipping.**

In the analysis of our assembled HAP1 cell line panel, we also observed ~30% of the clones exhibited exclusion of the targeted exon in the mRNA. Exons are replete with splicing regulatory motifs including exon splicing enhancers and suppressors (ESEs and ESSs,



respectively). These degenerate hexameric sequences dictate the extent to which exons are included within a transcript<sup>12, 23</sup>. We suspected that exon exclusion was at least in part due to the disruption of ESEs by an INDEL event. As part of our efforts focused on studying the *SUFU* tumor suppressor protein, we had generated a collection of cells that presumably were null for *SUFU* based on Western blot analysis (**Fig. 4A, B**). Yet, we noted that many of these clones exhibited exclusion of the targeted exon (**Fig. 4C**). The extent of exon exclusion notably differs suggesting other factors, perhaps RNA structure changes that contribute to exon splicing regulation, also may be compromised by the introduction of an INDEL at this position within the *SUFU* mRNA. We identified a cluster of potential ESEs in the targeted *SUFU* exon that was likely impacted by the INDEL in these clones (**Fig. 4D**). No ESSs were identified in this case. To determine how reliably we can induce exon exclusion by impacting a predicted ESE, we introduced INDELs at putative ESEs found in other *SUFU* exons and performed similar analysis of the protein and mRNA in RMS13 cell line (**Fig. 4E-L**). In every instance, we observed exon exclusion by targeted disruption of a putative ESE.

When all the clones presented so far from both commercial and *de novo* engineered were considered with respect to predicted impact on an ESE and exon exclusion, we observed a strong correlation between these two events (**Fig. 4M; Supp. Fig. 7**). A subset of the clones exhibiting alternative splicing also expressed novel polypeptides (see *TOP1* and *SIRT1*; see **Fig. 1B**). We noted in both these cases that the exons were symmetric – meaning that exon exclusion would result in a transcript lacking that sequence but nevertheless retain the original reading frame. In the case of the *SUFU* clones, the majority of exons skipped were asymmetric thus likely resulting in the lack of protein expression. However, we noted one targeted and skipped exon (exon 2) was symmetric yet the resulting transcript failed to generate a detectable protein perhaps due to misfolding of the mutant protein (see **Fig. 4E, F**). Indeed, the skipped exon encodes part of an intrinsically disordered region of the protein that is essential for interaction with members of the pro-survival BCL2 family members<sup>24</sup>. From these *SUFU* clones, we expect that decreased *SUFU* mRNA seen in CRISPR-edited cells was due to NMD provoked by the introduction of a frameshift inducing INDEL, or exclusion of the targeted asymmetric exon and the introduction of a PTC in an NMD-enabling position within the gene.

### **CRISPinatoR: a web-based guide RNA design tool that exploits ESE disruption for achieving gene elimination.**

Purposeful disruption of ESEs in asymmetric exons could improve gene knockout efficiency given that even INDELs that fail to alter the coding frame would have a second opportunity for introducing a PTC by skipping the exon altogether. To systematize this strategy, we developed the CRISPinatoR, a website that identifies asymmetric exons found in a given gene and CRISPR/Cas9 guide sequences that help to deliver double-stranded breaks within proximity of a putative ESE (**Fig. 5A, Supp. Fig. 8**). At the same time, the portal could be used to induce the skipping of an exon harboring a deleterious mutation in order to generate a novel protein that may retain function. We note that when analyzing genome-wide CRISPR libraries, that ratio of guides targeting symmetric and asymmetric exons was fairly consistent, suggesting that these algorithms do not factor in potential gene elimination efficiency based on exon symmetry (**Supp. Fig. 9A-B**). Similarly, the CRISPinatoR could be used to re-evaluate previously reported phenotypes using CRISPR/Cas9 based on the potential for the sgRNA for inducing exon skipping.

### **Targeting RNA regulatory elements for gene knockout agendas.**

We pressure-tested the ability of CRISPinatoR to design guides that induce exon skipping for either degradation of mRNA or production of novel protein-encoding mRNAs by targeting asymmetric or symmetric exons, respectively. Using the WNT receptor LRP5 as a case study, we asked the CRISPinatoR to identify sgRNAs that presumably would be able to induce exon

skipping in each exon class (**Fig. 5B**). We identified 6 clones that harbored INDELs at the anticipated LRP5 exonic sequence by targeted sequencing of isolated genomic DNA (**Fig. 5C**). Using RT-PCR analysis coupled with targeted sequencing, we observed exon skipping in clones associated with both guides (**Fig. 5D**; **Supp. Fig. 10**). We observed an absence of LRP5 protein in the clone exhibiting exclusion of an asymmetric exon (**Fig. 5E**). However, in the clone exhibiting exclusion of a symmetric exon, we observed the appearance of a faster migrating protein (see **Fig. 5E**). We confirmed that this new protein retains glycosyl moieties, suggesting that its signal sequence localized to the N-terminus is intact unlike in the case of the *LRP6* edited HAP1 clone (**Fig. 5F**; see **Supp. Fig. 1A,C**). The presence of a secreted protein and evidence for skipping of the CRISPR targeted exon suggest that the novel LRP5 protein formed would harbor a compromised  $\beta$ -propeller domain – one of two that contributes to WNT3A binding (**Fig. 5G**). Indeed, we observed response of a clone expressing the truncated LRP5 protein to exogenously supplied WNT conditioned medium using a WNT pathway reporter (**Fig. 5H**). The weakened response compared to WT HAP1 cells likely reflects reduced total LRP5 protein levels and/or reduced WNT binding affinity with deletion of exon 16 sequence. On the other hand, the cell expressing the LRP5 mRNA excluding the CRISPR-edited asymmetric exon showed a loss of WNT pathway response consistent with the absence of LRP5 protein production from an mRNA lacking an asymmetric exon (see **Fig. 5H**).

## DISCUSSION

The microRNA-like behavior of short interference RNAs (siRNAs) has long posed a challenge to using RNA interference (RNAi) for selective gene product ablation in both early discovery and therapeutic settings<sup>25, 26</sup>. Whereas this issue is not inherent in DNA-editing systems such as CRISPR/Cas9, we show here that this technological advantage is offset by the unanticipated effects stemming from the on-target changes that impact the regulation of the RNA product and the translation of the protein it encodes. Exon skipping events, for example, associated with CRISPR have been previously observed although the mechanistic basis for these phenomena was not well-understood<sup>27</sup>. Our incomplete understanding of RNA splicing regulatory mechanisms, inability to accurately predict RNA structural changes introduced by INDELs, and limited accounting of the pseudo-transcriptome challenge our ability to anticipate transcriptional and translation outcomes as a consequence of introducing INDELs in exonic sequences (**Fig. 6**). Indeed, we assume these trials extend to other INDEL-producing gene editing systems that have been applied in human cells such as the CRISPR endonuclease Cpf1<sup>28</sup>.

A number of considerations in guide design could be installed in our design workflow to increase the fidelity of DNA sequencing information for predicting protein translational outcomes. A map of RNA-regulatory motifs (such as ESEs) that might be impacted by a CRISPR/Cas9-delivered INDEL such as that generated by the CRISPinatoR for the human genome could help in improving gene elimination or protein engineering campaigns. We acknowledge that the impact of RNA structure and possibly other determinants that can influence the function of regulatory sequences involved in RNA splicing, for example, are not accounted for by our database. At the same time, an understanding of lineage-associated pseudo-transcripts that would be edited alongside the intended target transcripts, would also help to anticipate the emergence of novel protein products such as Super LKB1 from conversion of a pseudo-mRNA to a protein-encoding mRNA.

Perhaps the most daunting challenge that we encountered from our analysis of CRISPR-edited cell lines is the emergence of IRESs likely due to INDEL-induced changes in RNA structure. We anticipate that the number of ATI events associated with INDELs will be higher than what is reported here given the shortage of antibodies useful for detecting native as well as potentially truncated proteins that emerge from ATI. In this regard, the use of translation inhibitors such as CHX combined with RT-PCR could be a simple method to flag mRNAs that

harbor CRISPR/Cas9-introduced frame-shift inducing INDELs yet for reasons including ATI subversion are not substrates for NMD.

Our observations also have implications for the use of INDEL-based genome editing tools for gene rescue efforts where induced exon skipping can excise sequences that harbor a mutation thus producing a viable gene<sup>29</sup>. These outcomes are currently achieved by using two CRISPR guides that flank a mutated exon<sup>30-32</sup>, or target an exon-specific splice junction using a single guide<sup>33</sup>. However, the ability to use a single guide targeting an ESE to achieve a similar outcome should reduce the dangers of using two CRISPR guides and expand the number of single guide options with acceptable off-target risks. In this regard, guides identified by the CRISPinator targeting ESEs found in symmetric exons could be used systematically identify such opportunities in genes involved in disease.

## REFERENCES

1. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
2. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
3. Garneau, N.L., Wilusz, J. & Wilusz, C.J. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**, 113-126 (2007).
4. Lykke-Andersen, S. & Jensen, T.H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**, 665-677 (2015).
5. Maier, K. et al. Single Amino Acid Deletion in Kindlin-1 Results in Partial Protein Degradation Which Can Be Rescued by Chaperone Treatment. *J Invest Dermatol* **136**, 920-929 (2016).
6. Dominguez, D. et al. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**, 854-867 e859 (2018).
7. Julien, P., Minana, B., Baeza-Centurion, P., Valcarcel, J. & Lehner, B. The complete local genotype-phenotype landscape for the alternative splicing of a human exon. *Nat Commun* **7**, 11558 (2016).
8. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-177 (2013).
9. Frith, M.C. et al. Pseudo-messenger RNA: phantoms of the transcriptome. *PLoS Genet* **2**, e23 (2006).
10. Blomen, V.A. et al. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092-1096 (2015).
11. Ke, S. & Chasin, L.A. Context-dependent splicing regulation: exon definition, co-occurring motif pairs and tissue specificity. *RNA Biol* **8**, 384-388 (2011).
12. Sterne-Weiler, T. & Sanford, J.R. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol* **15**, 201 (2014).
13. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096-1101 (2015).
14. Sriram, A., Bohlen, J. & Teleman, A.A. Translation acrobatics: how cancer cells exploit alternate modes of translational initiation. *EMBO reports* **19** (2018).
15. Jenkins, R.H. et al. A conserved stem loop motif in the 5'untranslated region regulates transforming growth factor-beta(1) translation. *PLoS One* **5**, e12283 (2010).
16. Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.* **9**, 5134-5142 (1989).

17. Svitkin, Y.V. et al. The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382-394 (2001).
18. Jagodnik, J., Chiaruttini, C. & Guillier, M. Stem-Loop Structures within mRNA Coding Sequences Activate Translation Initiation and Mediate Control by Small Regulatory RNAs. *Mol. Cell* **68**, 158-170 e153 (2017).
19. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8301-8305 (1990).
20. Paulus, M., Haslbeck, M. & Watzele, M. RNA stem-loop enhanced expression of previously non-expressible genes. *Nucleic Acids Res* **32**, e78 (2004).
21. Tan, Z., Fu, Y., Sharma, G. & Mathews, D.H. TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic Acids Res* **45**, 11570-11581 (2017).
22. Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
23. Cartegni, L., Chew, S.L. & Krainer, A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298 (2002).
24. Wu, X. et al. Extra-mitochondrial prosurvival BCL-2 proteins regulate gene transcription by inhibiting the SUFU tumour suppressor. *Nat Cell Biol* **19**, 1226-1236 (2017).
25. Echeverri, C.J. et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Methods* **3**, 777-779 (2006).
26. Ma, Y., Creanga, A., Lum, L. & Beachy, P.A. Prevalence of off-target effects in Drosophila RNA interference screens. *Nature* **443**, 359-363 (2006).
27. Mou, H. et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol* **18**, 108 (2017).
28. Zetsche, B. et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759-771 (2015).
29. Ousterout, D.G. et al. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun* **6**, 6244 (2015).
30. Nelson, C.E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403-407 (2016).
31. Tabebordbar, M. et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407-411 (2016).
32. Long, C. et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* **351**, 400-403 (2016).
33. Long, C. et al. Correction of diverse muscular dystrophy mutations in human engineered heart muscle by single-site genome editing. *Sci Adv* **4**, eaap9004 (2018).
34. Ke, S. et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res* **21**, 1360-1374 (2011).
35. Fairbrother, W.G. et al. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res* **32**, W187-190 (2004).
36. Zerbino, D.R. et al. Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761 (2018).
37. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
38. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).



# **MATERIALS AND METHODS**

## **Cell Culture**

HELA, MIA PaCa-2 and RMS13 cell lines were purchased from ATCC. WT and CRISPR-edited HAP1 knockout commercial cell lines were purchased from Horizon Discovery (**Supp. Table 1**). Puromycin was purchased from Fisher Scientific (ICN10055225). Cycloheximide was purchased from EMD Millipore (239765). NE-PER Nuclear and Cytoplasmic extraction reagent (78833) was purchased from ThermoFisher.

## **Western blot analysis**

Cell lysates were generated with PBS/1% NP40 buffer supplemented with protease inhibitor cocktail (Sigma Cat. No. S8820). Protein sample loading buffer was added to cell lysates and proteins were separated on SDS-PAGE (BioRad Criterion TGX Precast Gel). Antibodies used for immunoblotting are listed in **Supp. Table 2**.

## **Transfection of sgRNAs**

1x10<sup>6</sup> MIA PaCa-2 or HAP1 cells were seeded per 6 well plate and co-transfected with 0.5 µg pCas-Guide plasmid using Effectene transfection reagent (Qiagen). 24 hrs after transfection, cells were trypsinized and plated in 150 mm<sup>2</sup> culture dishes in various dilutions for clonal selection.

## **Clonal isolation of CRISPR-edited cells**

Cells in 150 mm<sup>2</sup> culture plates were treated with 0.5 µg/ml of puromycin in order to enrich for cells expressing Cas9. Puromycin selection was maintained for 10 days after which single colonies were isolated and grown in a 96 well plate. Cells from single colonies were passaged multiple times until sufficient cells were available for analyzing genomic DNA, RNA and protein.

## **Genomic DNA extraction and genomic sequencing**

Genomic DNA was extracted from the CRISPR-edited cells using Genomic DNA Minikit (Bioland Scientific) according to manufacturer's instructions and used as template for PCR amplification. PCR primers encompassing the CRISPR-targeted region were designed. PCR was performed with GoTaq Green Master Mix (Promega M7122) with following conditions: 98°C for 2 mins (initial denaturation), 25 cycles of 98°C for 30 secs, 56°C for 30 secs, 72°C for 30 secs (denaturation, annealing, extension) and final 70°C for 5 minutes (final extension). Gel electrophoresis in a 1.5% agarose gel was performed and the PCR products were purified from the gel using QIA Quick PCR Purification Kit (Qiagen) and cloned into pCR-TOPO plasmid using TOPO TA cloning kit for Subcloning (ThermoFisher Scientific). pCR-TOPO plasmids containing genomic DNA sequences were transformed into TOP10 competent cells and individual colonies were selected and sequenced.

## **RNA extraction and analysis**

RNA extraction was performed using RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. cDNA synthesis was performed on 1 µg of RNA using ProtoScript First Strand cDNA Synthesis Kit (Promega). Primers recognizing exons flanking the CRISPR-targeted exon (**Supp. Table 3**) were used to amplify the cDNA sequences isolated from the CRISPR-edited cells. PCR products were electrophoresed in a 1% agarose gel and the gel bands were isolated using QIA Quick PCR Purification Kit (Qiagen). Isolated DNA was cloned into pCR-TOPO plasmids using the TOPO TA cloning kit (ThermoFisher Scientific), and clones were sequenced at the UTSW Sequencing Core.

# **RNA Secondary structure modeling**

Conserved secondary structures were modeled using TurboFold II<sup>21</sup>. The full-length sequences of the five clones without ATI (HAP1 clones H1, H6, H7, H8 and 3bp substitution) and eight clones with ATI (HAP1 clones H2, H3, H4, H5 and MIA clones M2.1, M2.2, M3.1, M4.1) were modeled separately. Default parameters were used with TurboFold II. The resulting secondary structures of each CRISPR clone were mapped to the manual alignment of all clones in dot-bracket format.

# **Nuclear and cytoplasmic fractionation**

WT and TOP1  $\Delta$ E6 cells were washed with PBS. Nuclear and cytoplasmic extract were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents according to the manufacturer's protocol (ThermoFisher).

# **ESE selection and design of CRISPinator**

ESE sequences were collected from Ke et al<sup>34</sup> (top 200 hexamers based on ESEseq score) and Rescue-ESE<sup>35</sup> (238 hexamers). 43 sequence motifs for RNA-binding proteins associated with splice regulation were also included<sup>8</sup>. Redundant hexamers and motifs were removed and finally 440 motifs were used to generate a reference collection of ESE sequences. We scored the frequency of ESE sequences that were potentially impacted by guides found in commonly used CRISPR libraries (GeckoV2, Avana, TKO V3 and Sanger). The target exon for each sgRNA and the symmetry of the target exon was identified using genomic annotation from Ensembl<sup>36</sup>. The number of ESEs within a given 20bp sgRNA sequence was annotated. To score the off-target potential for each sgRNA candidate, we modified the bwa source code<sup>37</sup>. The total 23bp sequence (20bp sgRNA + 3bp PAM) was aligned to the hg19 reference genome allowing up to 3 mismatches. The variable bp in the 5' most position of the PAM sequence was not considered for mismatch scoring. An off-target score (0-100) for the sgRNA was calculated by a method used in Hsu et al<sup>38</sup> and a cutoff of 80 was considered acceptable.

# **Acknowledgements:**

This work was supported in part by Welch Foundation (I-1665 to LL), CPRIT (RP130212 to LL), the National Cancer Institute (1R01 CA168761 to JK), the American Cancer Society (RSG-16-090-01-TBG to JK), and the National Institutes of Health (R01GM076485 to DHM).

# **Competing interests:**

LL, RT, YY, THH, JTP and QB are co-inventors on U.S. Provisional Patent Application No. 16/003,683.

## FIGURE LEGENDS

**Figure 1. Prevalence of unanticipated gene expression outcomes following on-target CRISPR/Cas9-mediated gene editing.** (A) The effect of CRISPR-introduced frameshift alterations on mRNA and protein expression was analyzed using a panel of CRISPR/Cas9-edited HAP1 cells that were commercially accessible. Red=deletion, Green=insertion. The targeted exon, anticipated PTC following insertion/deletion mutation and the recognition sites of antibodies used in “B” are indicated. (B) Appearance of novel proteins in cells edited with CRISPR/Cas9. HAP1 cells were subjected to Western blot analysis using two distinct antibodies. \* indicate novel proteins. (C) CRISPR/Cas9 gene editing induces expression of novel mRNA species. RT-PCR analysis of edited cells was performed using primers recognizing flanking exons and the amplicons generated were sequenced. \* indicate novel mRNA species.

**Figure 2. A *TOP1* gene harboring a frameshift-inducing deletion introduced by CRISPR/Cas9 retains catalytic activity.** (A) Exclusion of symmetric exon 6 produces in-frame TOP1 protein with altered subcellular redistribution. (B) Quantification of TOP1 subcellular distribution in “A”. The values for the cytoplasmic and the nuclear TOP1 protein were normalized to the total TOP1 protein in HAP1 WT or TOP1  $\Delta$ E6 cells. (C) The internally truncated TOP1  $\Delta$ E6 protein can induce relaxation of supercoiled DNA. Camptothecin (TOP1 inhibitor) prevents DNA relaxation. (D) Summary of novel mRNAs or proteins observed in 13 CRISPR-edited commercial HAP1 cell lines (Horizon Discovery).

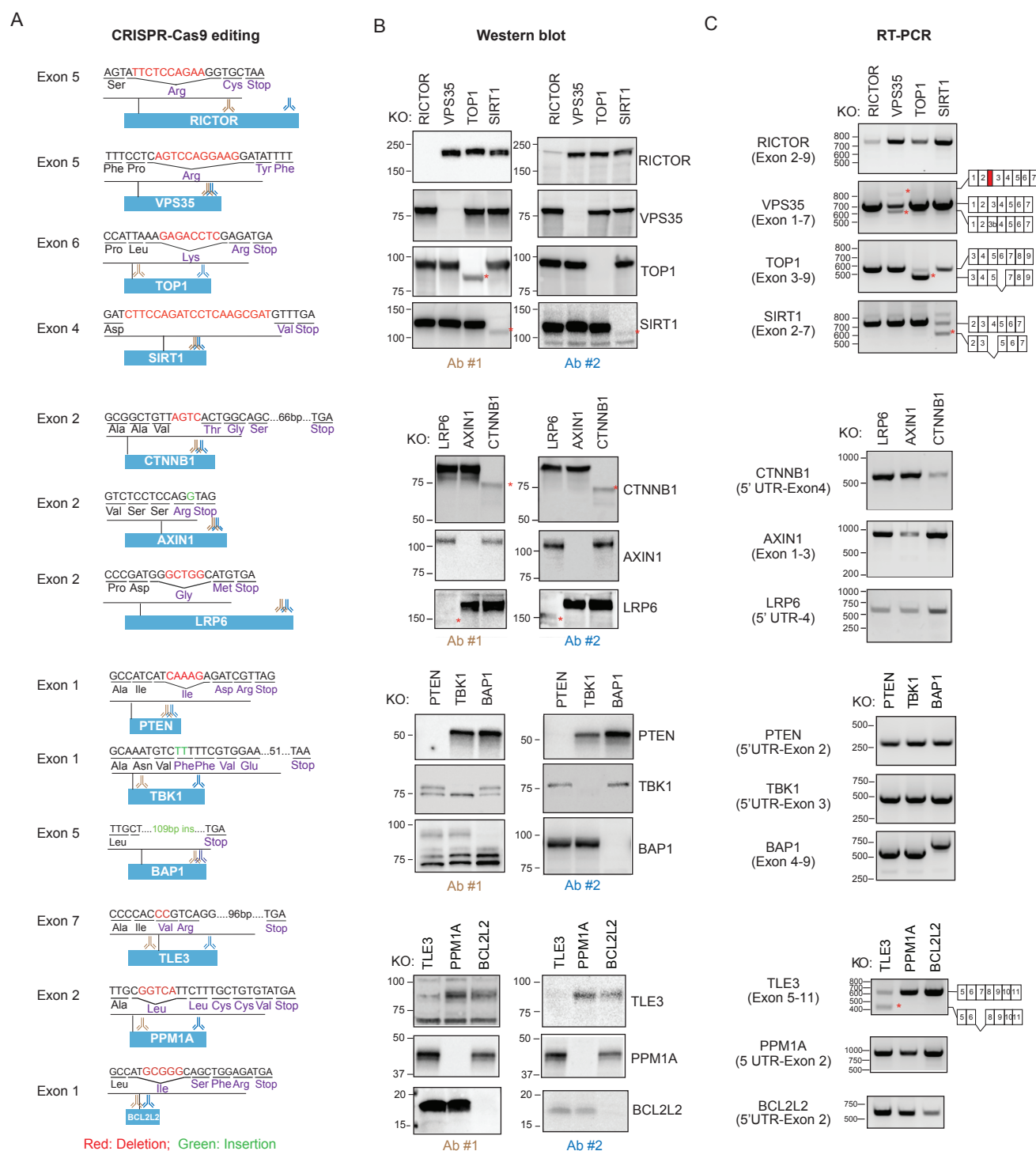
**Figure 3. Alternative translation initiation (ATI) and pseudo-mRNA editing produce foreign proteins in a CRISPR/Cas9 gene knockout campaign.** (A) Genomic structure of the *LKB1* gene and the exonic sequence targeted by the *LKB1* exon 1 sgRNA. (B) Emergence of a faster migrating novel LKB1 protein (ATI LKB1) as a consequence of CRISPR/Cas9 gene editing. Lysates generated from CRISPR-edited HAP1 clones were subjected to Western blot analysis using two distinct LKB1 antibodies recognizing either N- and C-terminus localized epitopes. (C) Western blot analysis of CRISPR/Cas9-edited MIA clones reveals the appearance of a large LKB1 protein (Super LKB1) in addition to the ATI LKB1 protein. (D) Genomic sequences of CRISPR/Cas9-edited HAP1 and MIA clones reveal on-target insertion/deletion mutations in the *LKB1* gene. Predicted gene alternation for each clone is indicated. (E) CRISPR/Cas9-introduced INDELs induce expression of an LKB1 pseudo-mRNA transcript. RT-PCR analysis was performed using primers mapping to 5' UTR and exon 4 in *LKB1* to generate amplicons from the cDNA of CRISPR/Cas9-edited clones. The amplicons generated were then sequenced. MIA clones M2 and M3, which express Super LKB1 protein, harbor an mRNA species that includes an additional exon. The 131 bp additional exon contains canonical splice acceptor and donor sequences. (F) A cDNA expression strategy for understanding allele-specific CRISPR-introduced INDELs on protein expression provides evidence for ATI. *LKB1* and Super *LKB1* cDNA expression constructs harboring genomic alterations found in *LKB1* of MIA Clone M2 were introduced into HELA cells that lack endogenous *LKB1* expression. The 1bp and 2bp insertion and deletion, respectively, in the Super *LKB1* cDNA resulted in proteins that co-migrate with the Super LKB1 protein observed in MIA Clone M2. On the other hand, the same mutations in *LKB1* cDNA gave rise to proteins that co-migrated with the ATI LKB1 protein found in Clone M2, and with the protein that initiates at Met51. (G) RNA secondary structure rearrangement triggered by CRISPR/Cas9-generated INDELs likely results in ATI. TurboFold II was used to predict common RNA secondary structures of *LKB1* clones with or without ATI LKB1 protein. RNA structures of clone H1 (WT) and clone H4 (6 bp del) are representative of the clones without the ATI LKB1 protein or with the ATI LKB1 protein, respectively. The nucleotides in yellow marks the 5' hairpin structure in WT clone H1 and the nucleotides in orange marks the 3' hairpin structure in ATI LKB1 clone H4.

**Figure 4. Compromised exon splicing enhancers (ESEs) account for INDEL-induced exon skipping.** (A) Genomic structure of the *SUFU* gene and the exonic sequence targeted by the *SUFU* exon 8 sgRNA. The recognition sites of antibodies used in “B” are indicated. (B) Western blot analysis of HAP1 cells edited with *SUFU* exon 8 sgRNA shows no detectable expression of *SUFU* protein. (C) Exon skipping is prevalent in CRISPR/Cas9-edited *SUFU* clones. RT-PCR analysis using primers flanking exon 6 and exon 10 of *SUFU* was performed for the CRISPR/Cas9-edited *SUFU* clones. Sequencing of the amplicons revealed existence of exon skipping in all of the clones except clones H9 and H10. (D) Disruption of exon splicing enhancers (ESEs) by CRISPR-introduced INDELs triggers skipping of the edited exons. Genetic mutation and the presence or the absence of exon skipping event for each clone are indicated. Putative ESEs were identified using the RESCUE-ESE web server. (E) Multiple sgRNA sequences located in symmetric or asymmetric exons of the *SUFU* gene used for targeted disruption of ESEs. (F) sgRNAs described in “E” were used to edit the *SUFU* gene in RMS13 cells. Western blot analysis of lysates derived from the CRISPR/Cas9-edited RMS13 clones showed no detectable *SUFU* protein. (G) Genomic sequences of the RMS13 clones edited with *SUFU* exon 3 sgRNAs. CRISPR-introduced mutations and putative ESE and ESS sequences are indicated. (H) RT-PCR analysis and cDNA sequencing result of clones R2 and R3 using primers flanking exon 1 and exon 5. (I) Genomic sequences of the RMS13 clones edited *SUFU* exon 2 sgRNA. CRISPR-introduced mutations and putative ESE and ESS sequences are indicated. (J) RT-PCR analysis and cDNA sequencing result of clones R1 and R4 using primers flanking 5’ UTR and exon 4. (K) Genomic sequences of the RMS13 clones edited with *SUFU* exon 8 sgRNA. CRISPR-introduced mutations and putative ESE and ESS sequences are indicated. (L) RT-PCR analysis and cDNA sequencing result of the clones R1 and R4 using primers flanking exon 6 and exon 10. (M) Disruption of ESE code is highly reliable in anticipating CRISPR/Cas9-induced exon skipping. Twenty-four CRISPR/Cas9-edited cell lines with different mutations were analyzed for the presence or the absence of exon skipping events and for the changes in the ESE code due to CRISPR-introduced INDELs.

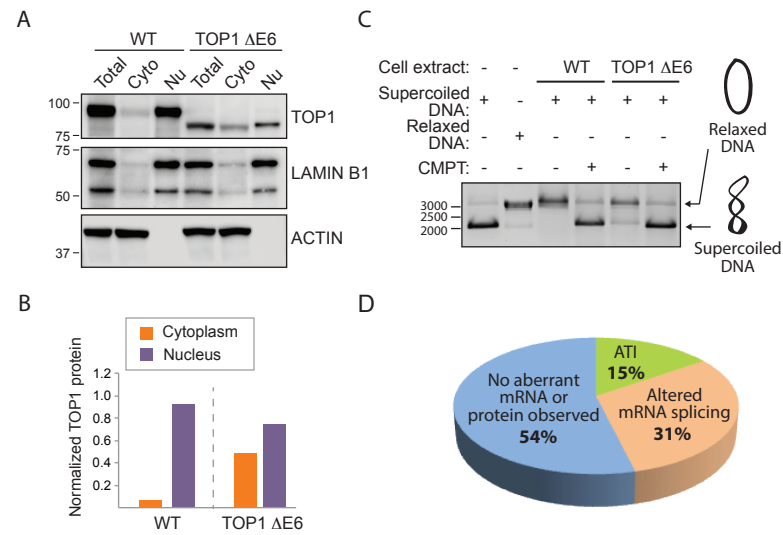
**Figure 5. Targeting RNA regulatory elements for gene knockout agendas.** (A) CRISPinatoR: a web-based guide RNA design tool that utilizes targeted ESE disruption for achieving gene elimination. CRISPinatoR identifies sgRNA sequences that target ESEs in asymmetric exons and calculates off-targeting potential and the number of splice variants impacted by the sgRNAs. A scoring system that integrates all three parameters is used to provide sgRNAs with high gene knockout potential. (B) Genome structure of the *LRP5* gene and sgRNA sequences targeting the asymmetric exon 2 and the symmetric exon 16. (C) Genomic sequencing results of HAP1 clones edited using *LRP5* exon 2 and exon 16 sgRNAs. CRISPR-introduced mutations and the putative ESE sequences are indicated. (D) Exclusion of an asymmetric or a symmetric exon with INDEL-induced changes to the putative ESE sequences. RT-PCR analysis and cDNA sequencing result of HAP1 cells edited with *LRP5* exon 2 and exon 16 sgRNAs. (E) Targeted ESE disruption in asymmetric exon increases gene knock-out potential. Western blot analysis of HAP1 clones edited with *LRP5* exon 2 sgRNA (Clone 21) and exon 16 sgRNA (Clone 3) were probed with 2 distinct antibodies indicated in “B”. ESE disruption in symmetric exon 2 produces internally truncated in-frame *LRP5* protein. (F) The internally truncated *LRP5* protein undergoes glycosylation. Lysates derived from WT or *LRP5* ΔE16 HAP1 cells were incubated with the deglycosidase PNGase F then subjected to Western blot analysis. (G) Exclusion of *LRP5* exon16 would delete a sequence adjacent to the WNT3A binding domains. (H) The *LRP5* ΔE16 protein formed post skipping of a symmetric exon is functionally active. WNT/β-catenin pathway activity in response to WNT3A conditioned medium (WNT3A CM) was measured for HAP1 WT, *LRP5* ΔE2 and *LRP5* ΔE16 cells. WNT pathway inhibitors WNT974 (PORCNI) and IWR1 (TNKSi) serve as negative and positive control respectively.



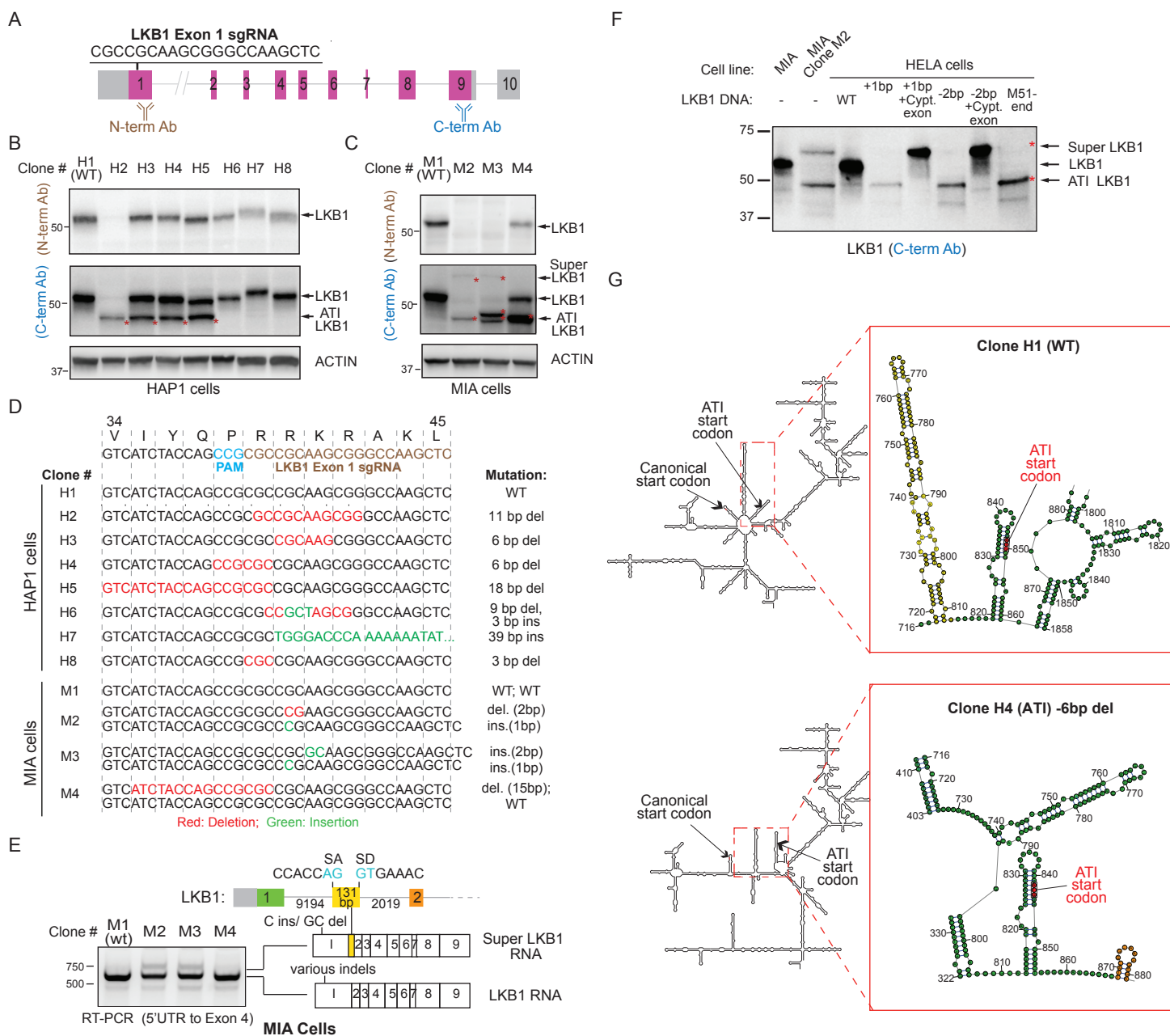
**Figure 6. Cell intrinsic response mechanisms for countering the effects of INDELs revealed by CRISPR/Cas9 failures.** INDELs introduced by Cas9 and other enzymes used for gene editing elicit transcriptional and translational responses that may have evolved to buffer the transcriptome and proteome against common environmental insults.



**Figure 1**

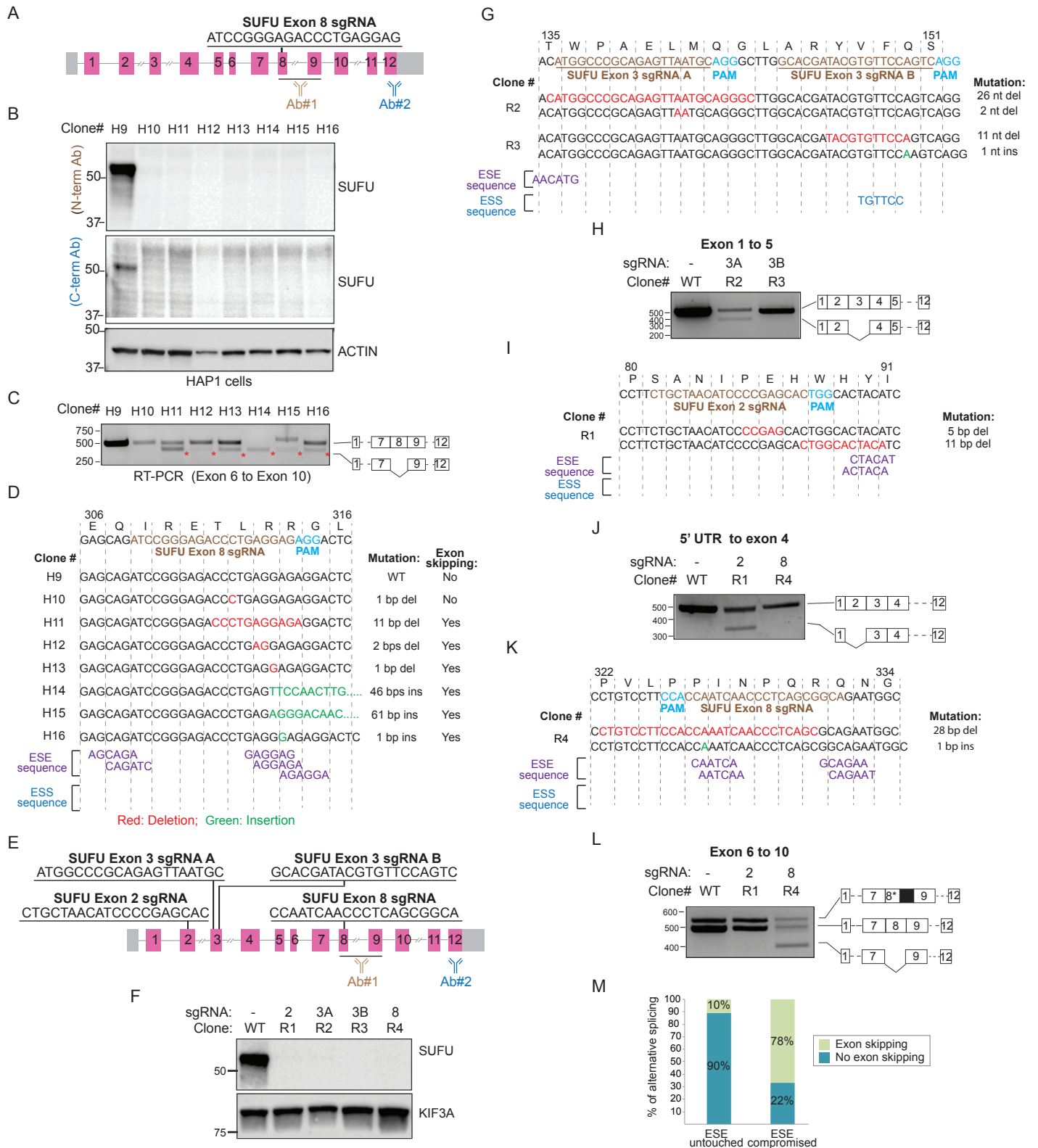


**Figure 2**

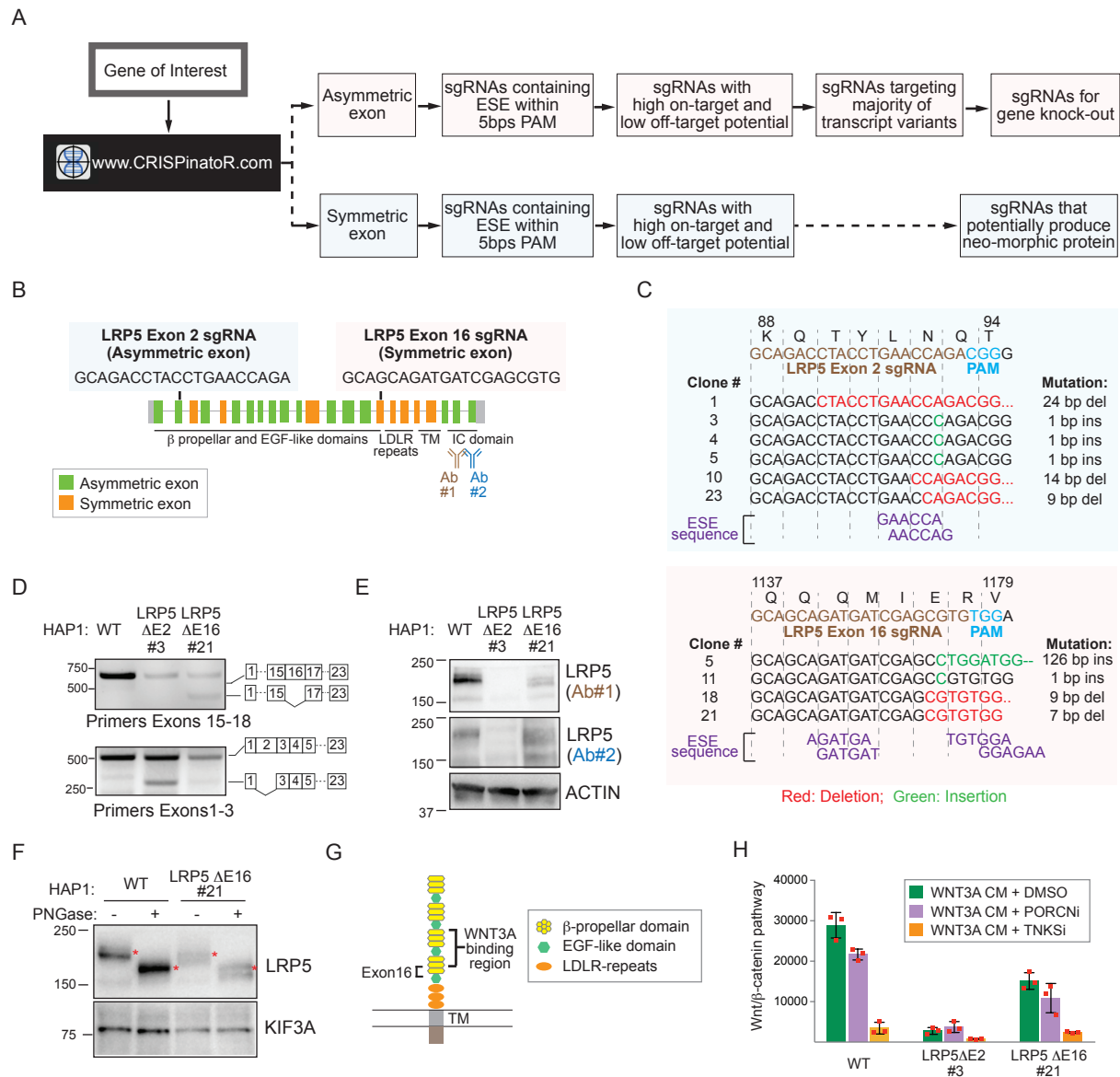


**Figure 3**

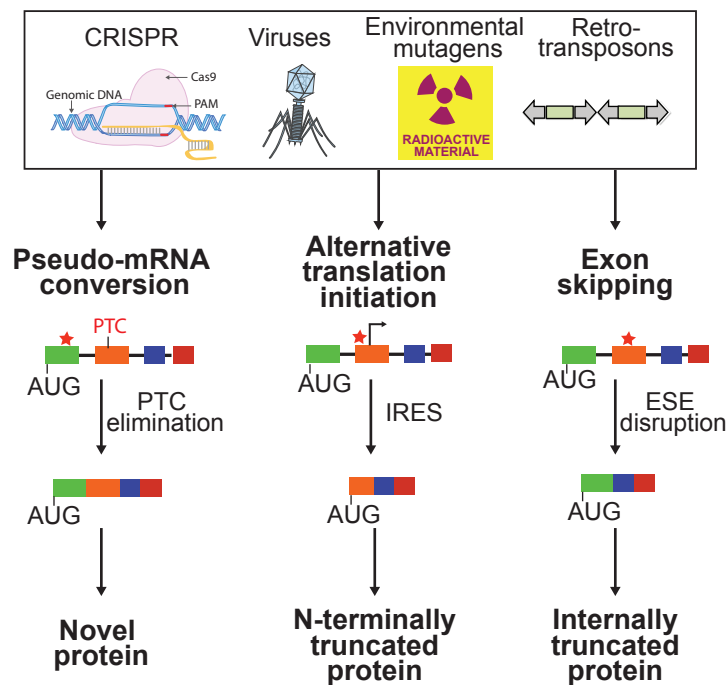




### Figure 4



**Figure 5**



**Figure 6**