# Supplemental Material for

# PlatinumCRISPr: A structure-optimized sgRNA selection for efficient Cas9 generation of knock-outs

**IRMGARD U. HAUSSMANN, THOMAS C. DIX, DAVID W. J. MCQUARRIE, VERONICA DEZI, ABDULLAH I. HANS, ROLAND ARNOLD AND MATTHIAS SOLLER**

## Supplemental information and discussion

### A guide to generate a gene knockout

Historically, imprecise excision of a marked *P* element transposon represents a versatile approach for directed mutagenesis of a gene in *Drosophila* and other organisms to generate a knock-out as illustrated for a knock-out of *mRNA cap methyltransferase 1* (*CMTr1*) gene (Supplemental Figure 10A) (Soller et al. 2006; Haussmann et al. 2016; Bawankar et al. 2021; Haussmann et al. 2022).

A *P* element transposon (unlike *Minos* or *PiggyBac* elements) can generate deletions by imprecise excision in the presence of a transposase. To increase excision efficiency, this is generally done over a larger chromosomal deficiency (marked with w+, available from stock centres) to prevent repair from the homologous chromosome. This type of widely used mutagenesis has been facilitated by large collections of *P* element insertions in the *Drosophila* genome and availability of corresponding strains from stock centers.

Accordingly, to generate a knock-out of *mRNA cap methyltransferase 1* (*CMTr1*) gene, a *P* element transposon (marked with w+) inserted in the 5'UTR was mobilized and an imprecise

excisions were recovered. These deletion were then mapped by inverse PCR and a deletion was selected that removes parts from the 5'UTR extending into the catalytic domain rendering this allele functionally null (Supplemental Figure 10A). Like chemical mutagenesis, imprecise excision of a marked *P* element requires screening for loss of the marker and then for a deletion of the desired extension.

Now, this process is massively facilitated by the use of sgRNA/Cas9 to introduce the desired deletion to generate a gene knock-out. In principal, deletions can be recovered after injection of a plasmid expressing two sgRNA under the *U6* promoter into *Drosophila* embryos expressing Cas9 under a germline promoter. To increase efficiency, however, we advise to use a transposon marker in the gene. In addition, to prevent repair from the homologous chromosome, the transposon is best excised in the presence of a marked larger chromosomal deficiency (available from stock centres). In this configuration, however, only 25-50% of embryos will have the desired genotype, because of lethality of the deficiency chromosome and potentially the transposon insert. Hence, a transgene expressing two sgRNA under the *U6* promoter facilitates the procedure, but will take longer.

To generate a knock-out for a typical *Drosophila* gene by sgRNA/Cas9, parts of the promoter (indicated by the TATA box) and a main functional domain (e.g. a catalytic domain, an RNA Recognition Motif, a zinc-finger DNA binding domain, etc) are deleted. Accordingly, gRNAs are then selected in these parts of the gene by entering the relevant sequence into PlatinumCRISPr (dashed lines, Supplemental Figure 10B).

Potential deletions are first identified by loss of the transposon marker and then validated by PCR using primers flanking the deletion. If a deletion is present, a short PCR product will be amplified, while a PCR for the parental chromosome will be much larger and likely unproductive. To check for re-insertion of the deleted fragment, PCR with primers flanking an

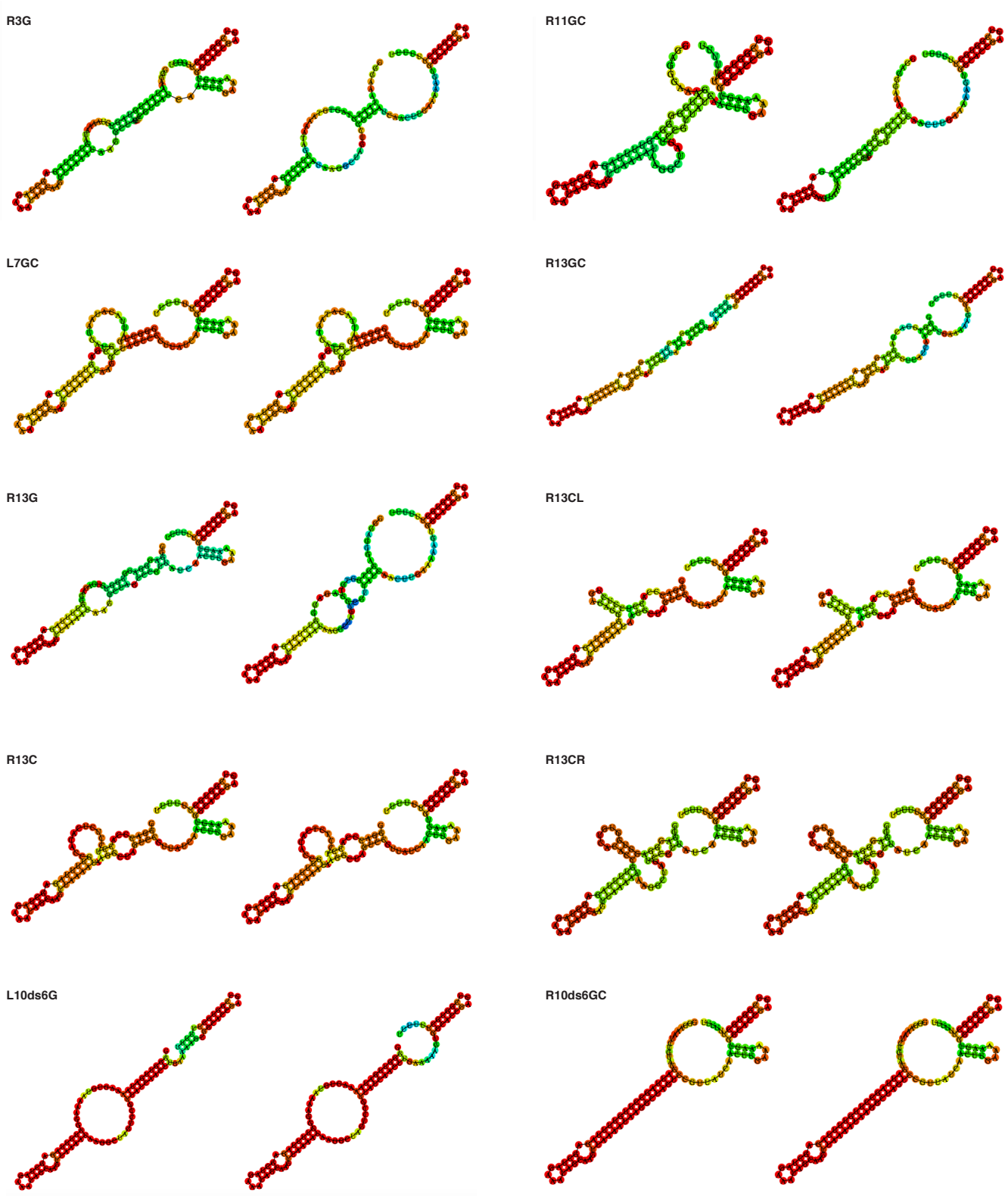
intron can be used. If the deleted fragment reinserted in the sense orientation and is expressed, a shorter fragment will be detected by RT-PCR compared to a larger fragment when the insert is expressed from the anti-sense strand.

Human genes are on average ten times larger than *Drosophila* genes due to more and larger introns (Soller 2006). Typically, human genes have short exons of around 150 nucleotides which are separated by large introns. These features require adaptation of the knock-out approach.

One possibility to overcome this challenge of increased gene size is to remove the promotor together with the first exon. However, this is often not feasible because most human genes have several transcription start-sites, physically separated by several kilobases.

An alternative approach is to delete an exon such that splicing will generate a frame-shift. To identify suitable gRNAs to delete this exon, the sequences before and after are entered into PlatinumCRISPr. To avoid splicing complications, it is best to avoid gRNAs in splice sites (select gRNAs before the branchpoint and about 30 nucleotides after the 5' splice site, Supplemental Figure 10C).

Such a deletion can conveniently be monitored by RT-PCR with primers in the flanking exons as they generate a short PCR product if the desired deletion is present, but a longer PCR for the parent. If sgRNA/Cas9 induced deletions are made in a population of cells (e.g. under puromycin selection), sgRNA/Cas9 efficiency can be monitored this way through the presence of two PCR products (with and without the deleted exon).

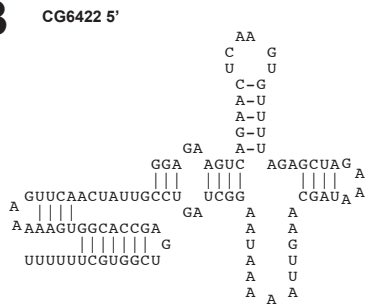


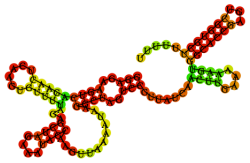

**Supplemental Figure S1.** Secondary structures of sgRNAs used in Figs 1-3 predicted by RNAfold.

**A**

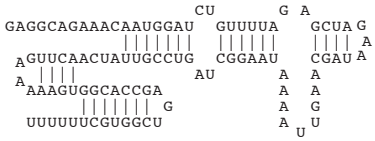
| sgRNA | length | sequence |
|---|---|---|
| **CG6422 5'** | 20 | GGAGAAGUCAGAACUCAAGU |
| **CG6422 3'** | 20 | GAGGCAGAAACAAUGGAUCU |

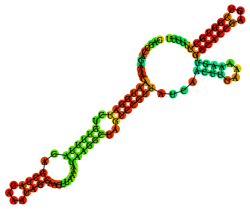
**B** **CG6422 5'**



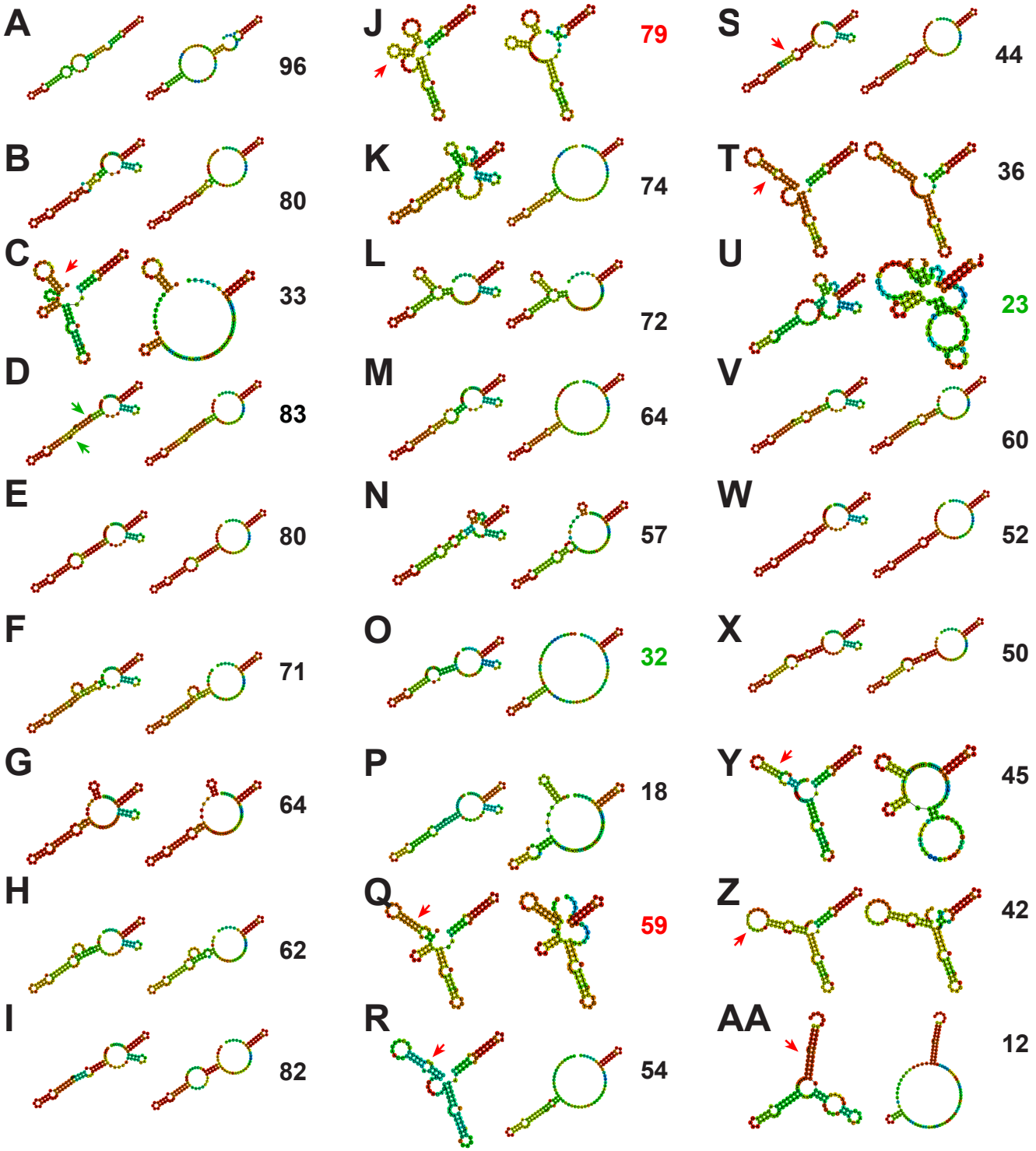


**C** **CG6422 3'**

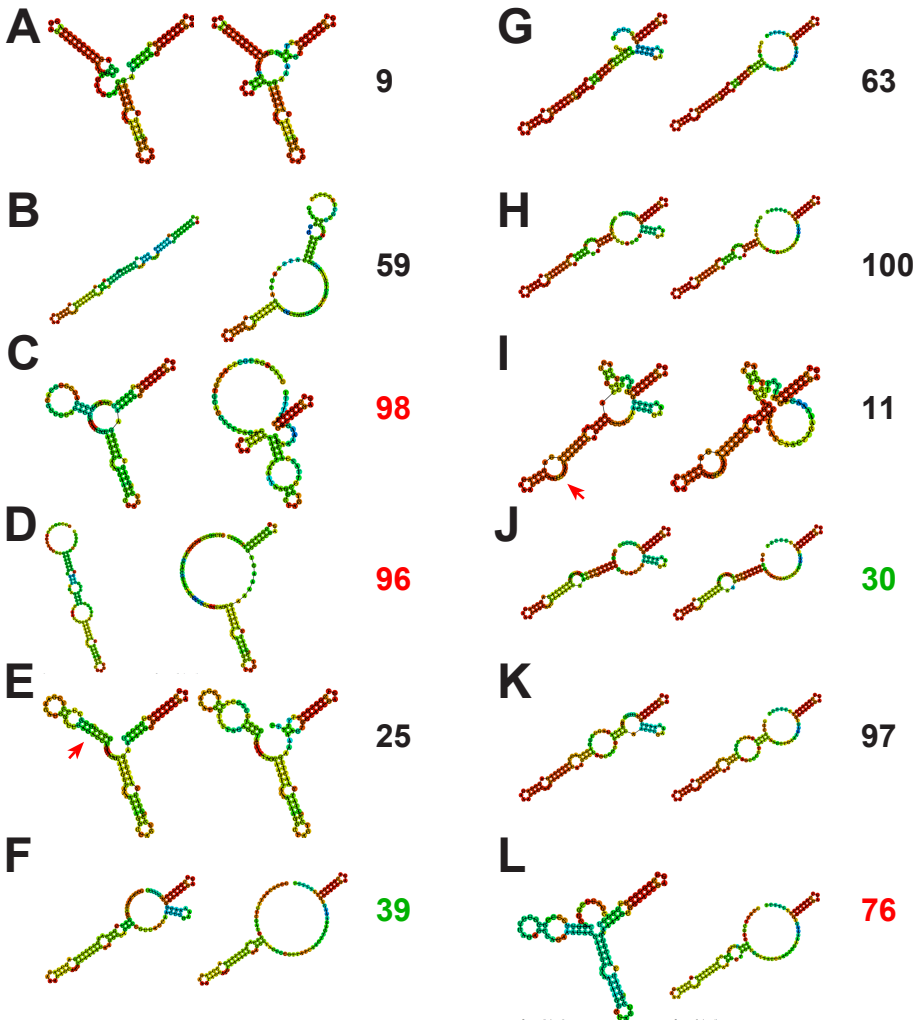




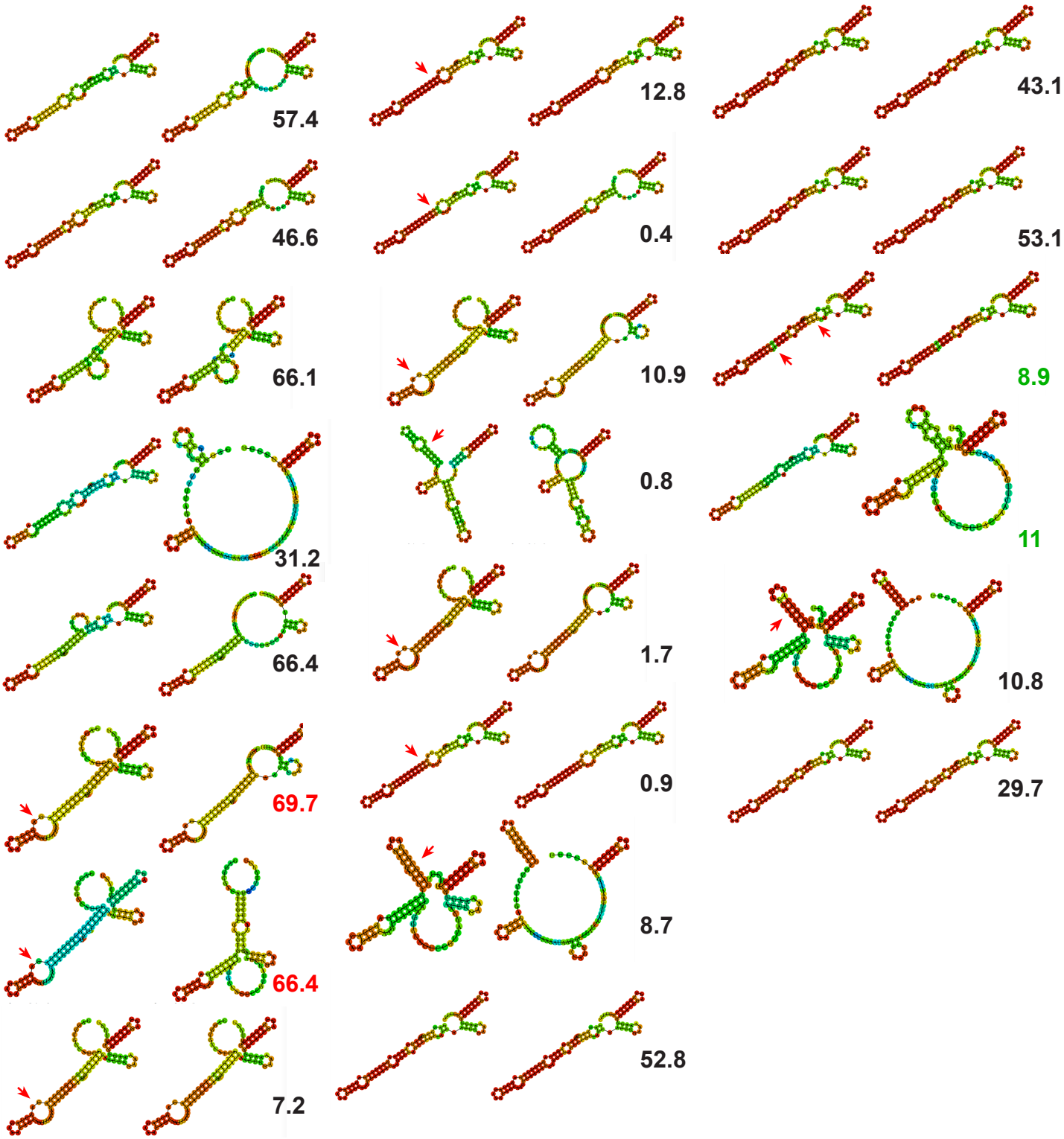
**Supplemental Figure S2.** Inactive sgRNAs targeted to the Drosophila ythdf (CG6422) gene.
**(A)** Sequences of sgRNAs targeting the CG6422 locus for generating a deletion. **(B, C)** Secondary structures of sgRNAs. The scaffold is shown on the left indicating Watson-Crick base-pairing by lines and the predicted secondary structure by RNAfold is shown on the right.



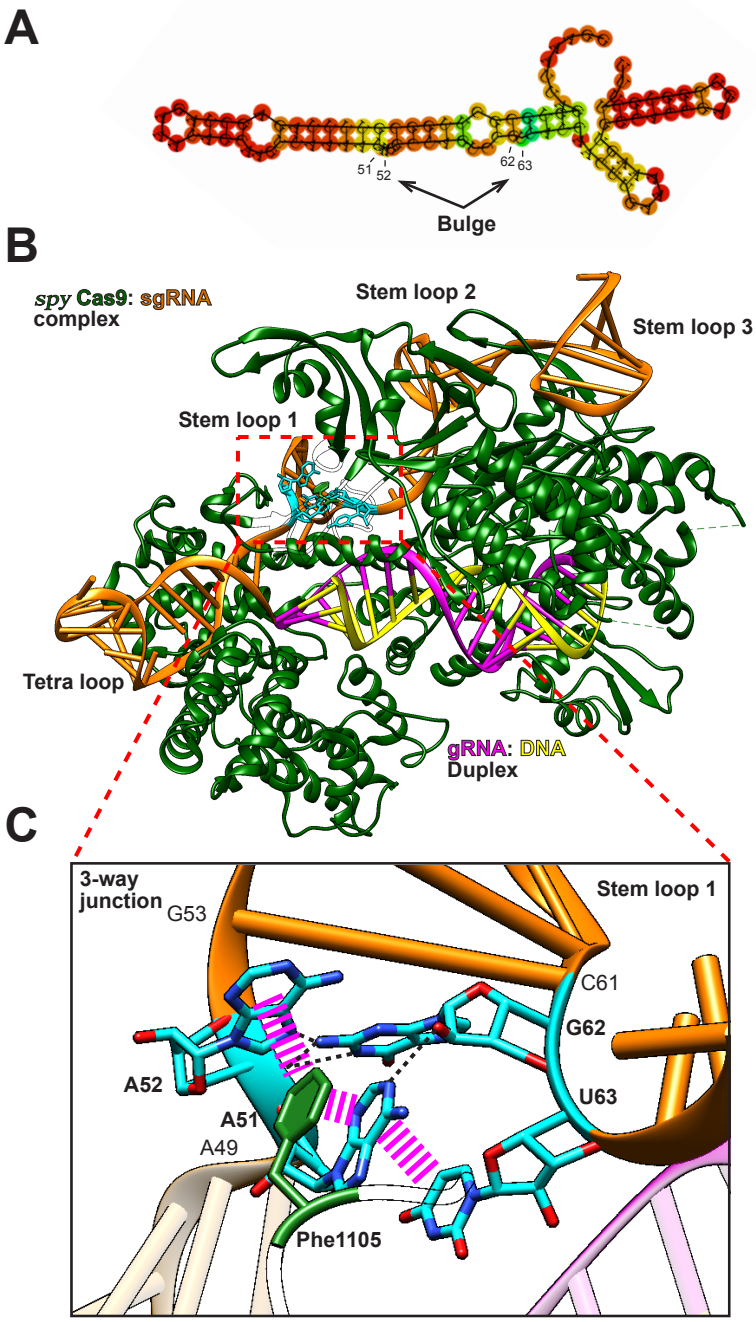

**Supplemental Figure S3.** Secondary structure of sgRNAs targeted to the Drosophila white gene from Ren et al. (2014).
**(A-AA)** Secondary structures of sgRNAs predicted by RNAfold. Minimal energy and proximity base-pair structures are shown on the left and right, respectively. Red and blue indicated high and low probabilities for the adopted structural base-pairing assignment, respectively. The number on the right indicates the effectiveness of inducing heritable mutations after injection into fly embryos as determined by Ren et al. (2014). Red and green numbers indicate an effectiveness which is too high or too low compared to predicted DNA cleavage efficiency. Red arrows point towards structural features likely limiting DNA cleavage and green arrows point towards structural features supporting DNA cleavage.



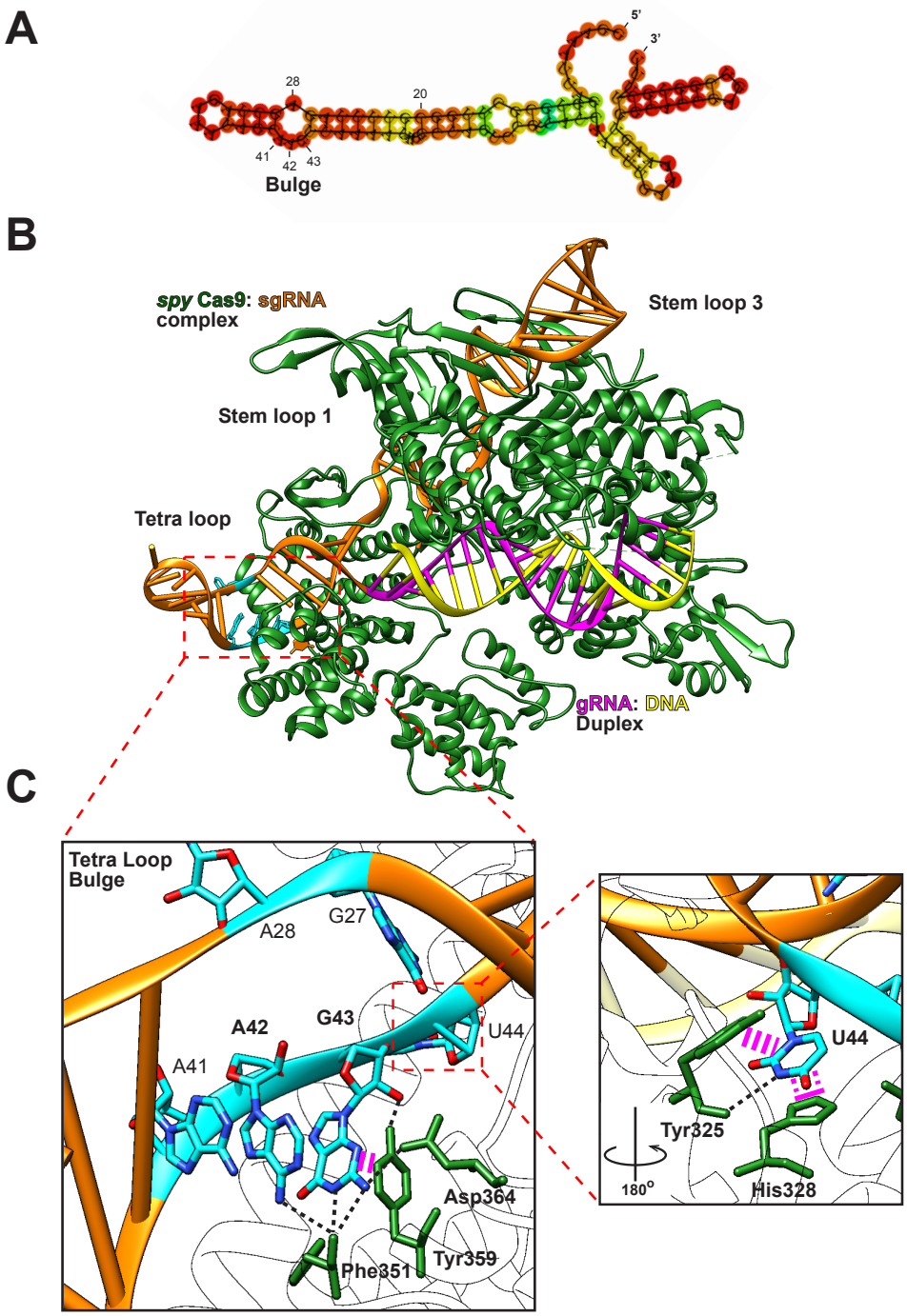

**Supplemental Figure S4.** Secondary structure of sgRNAs targeted to the Drosophila vermillion, ebony and yellow gene from Ren et al. (2014).
**(A-L)** Secondary structures of sgRNAs predicted by RNAfold targeting the vermillion (A-D), the ebony (E-H) and the yellow gene (I-L) . Minimal energy and proximity structures are shown on the left and right, respectively. Red and blue indicated high and low probabilities for the adopted structural base-pairing assignment, respectively. The number on the right indicates the effectiveness of inducing heritable mutations after injection into fly embryos as determined by Ren et al. (2014). Red and green numbers indicate an effectiveness which is too high or too low compared to predicted DNA cleavage efficiency. Red arrows point towards structural features likely limiting DNA cleavage.



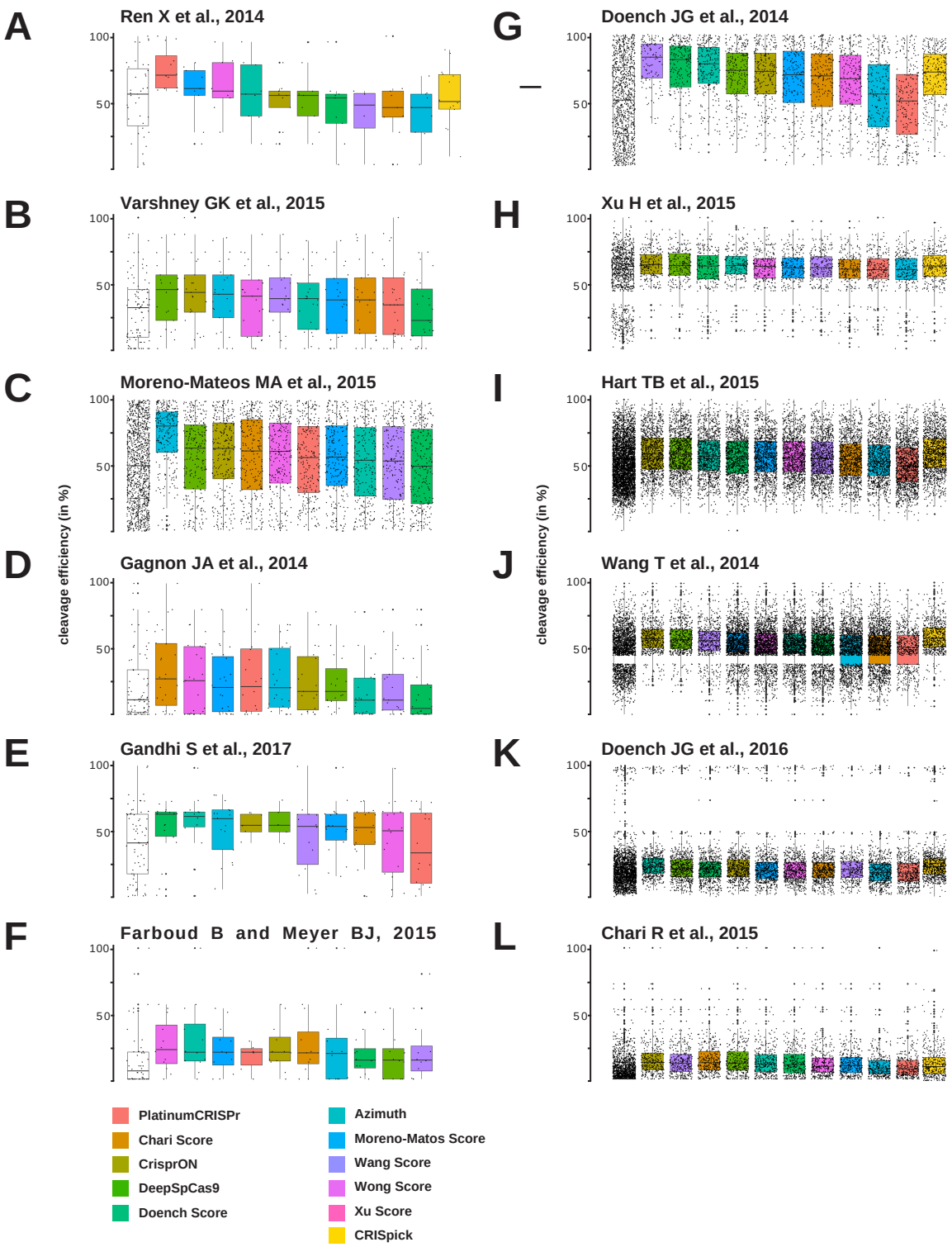

**Supplemental Figure S5.** Secondary structure of sgRNAs targeted to human CD22 from Graf et al. (2019). **(A-V)** Secondary structures of sgRNAs predicted by RNAfold. Minimal energy and proximity base-pair structures are shown on the left and right, respectively. Red and blue indicated high and low probabilities for the adopted structural base-pairing assignment, respectively. The number on the right indicates the effectiveness of inducing heritable mutations after injection into fly embryos as determined by Ren et al. (2014). Red and green numbers indicate an effectiveness which is too high or too low compared to predicted DNA cleavage efficiency. Red arrows point towards structural features likely limiting DNA cleavage.



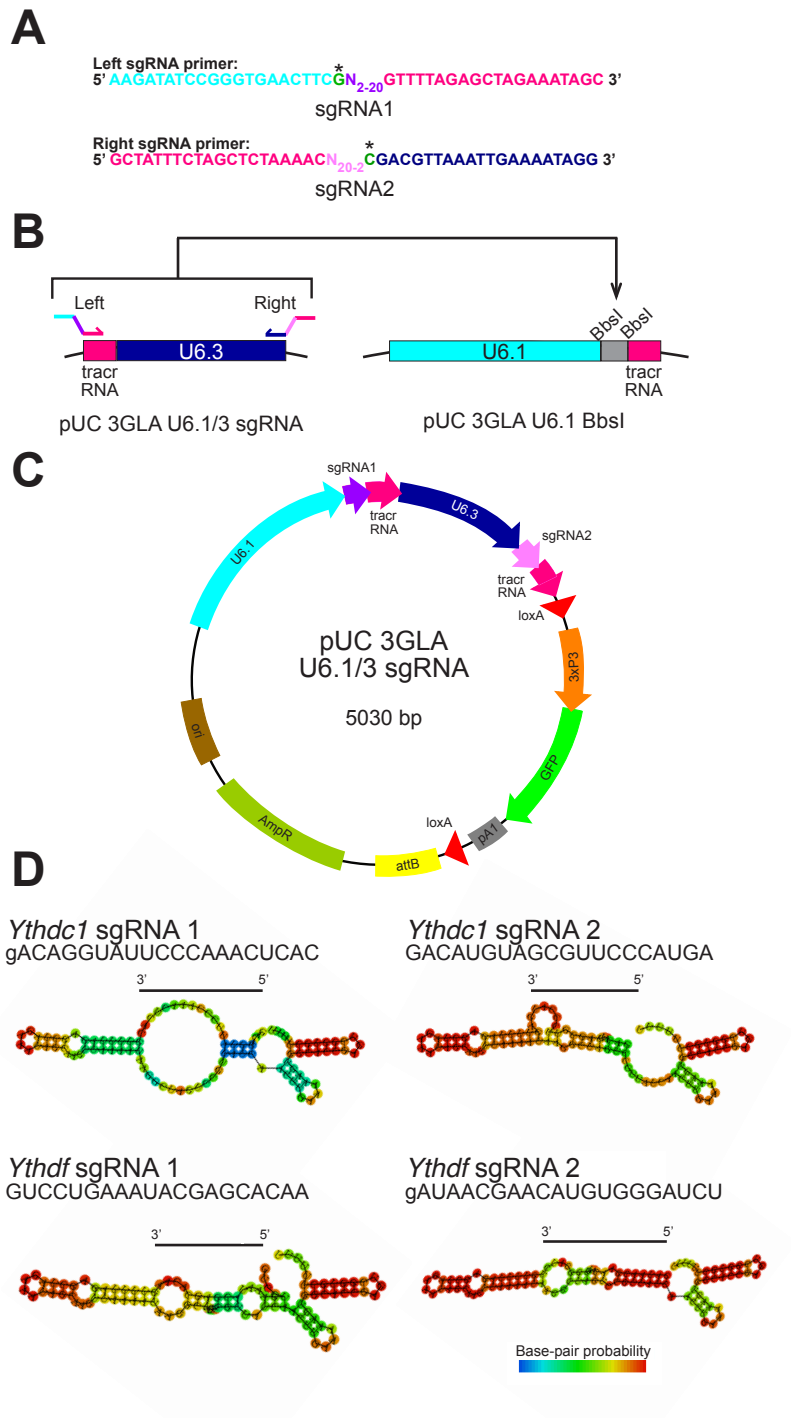

**Supplemental Figure S6.** Key sgRNA secondary structural features directly interact with the Spy Cas9 endonuclease in the three way junction.
**(A)** Minimum Free Energy (MFE) predicted secondary structure of the sgRNA used for co-crystallization of Spy Cas9 (Nishimasu et al., 2014). Nucleotides involved in the three way junction interactions are indicated ($A_{51}$, $A_{52}$, $G_{62}$ and $U_{63}$).
**(B)** X-ray crystal structure of†spyCas9 endonuclease in complex with chimeric sgRNA bound to genomic DNA target (PDB: 4OO8) (Nishimasu et al 2014). The gRNA portion is coloured in pink and the remainder of the sgRNA is coloured in orange.Genomic DNA is coloured yellow and the protein chain is coloured in green. Nucleotide residues situated at key structural features in the sgRNA are coloured cyan. Note that a small portion of the Cas9 protein chain depicting amino acid side chains to view the sgRNA three-way junction is transparent.†
**(C)** Magnified view of the sgRNA three-way junction. Hydrogen bonds are indicated by black dashed lines and aromatic stacking interactions are shown by pink dashed lines. Here, the interaction of $Phe_{1105}$ with $A_{51}$, $A_{52}$, and $U_{63}$ and the interaction of $G_{62}$ with $A_{51}$, $A_{52}$ and and the phosphate backbone can be seen.



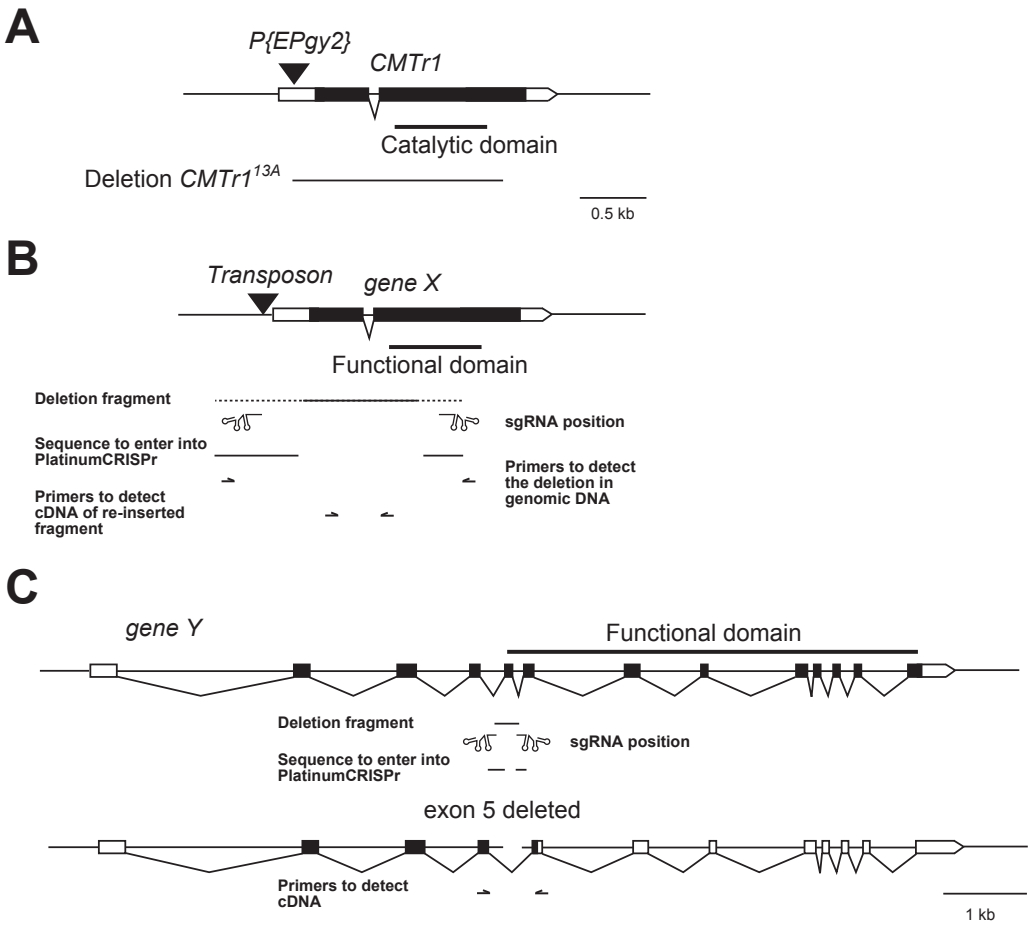

**Supplemental Figure S7.** Key sgRNA secondary structural features directly interact with the Spy Cas9 endonuclease in the tetraloop bulge.
**(A)** Minimum Free Energy (MFE) predicted secondary structure of the sgRNA used for co-crystallization of spyCas9 (Nishimasu et al., 2014). Nucleotides forming the bulge are indicated ($A_{28}$, $A_{41}$, $A_{42}$ and $G_{43}$).
**(B)** X-ray crystal structure of Spy Cas9 endonuclease in complex with chimeric sgRNA bound to genomic DNA target (PDB: 4OO8) (Nishimasu et al 2014). The gRNA portion is coloured in pink and the remainder of the sgRNA is coloured in orange. Genomic DNA is coloured yellow and the protein chain is coloured in green. Nucleotide residues situated at key structural features in the sgRNA are coloured cyan. Note that the tetra loop after the bulge does not interact with Cas9 and sticks out of the structure.
**(C)** Magnified view of the sgRNA bulge present in the tetra loop. Hydrogen bonds are indicated by black dashed lines and aromatic stacking interactions are shown by pink dashed lines. Here, the interactions of $Phe_{351}$, $Tyr_{359}$ and $Asp_{364}$ with $A_{42}$ and $G_{43}$ can be seen. Inlet on the right: 180º turn to visualize base-stacking of $U_{44}$ with $Tyr_{325}$ and $His_{328}$. Through these interactions base-pairing with $G_{27}$ is prevented.




**Supplemental Figure S8.** Comparison of sgRNA selection by different sgRNA selection tools shown as median of the cleavage efficiency for individual data sets (A-L). The distribution of cleavage efficiencies for all sgRNAs is shown on the left (white box).




**Supplemental Figure S9.** *Drosophila* GFP-marked transformation vector for U6 promoter mediated expression of two sgRNAs.
**(A)** Forward and return primer sequences to incorporate sgRNA sequences. The first nucleotide of the gRNA is indicated by an asterisk.
**(B)** Cloning scheme for incorporating two sgRNAs into the destination vector.
**(C)** Plasmid map of the fly transformation vector *pUC 3GLA U6.1/U6.3* sgRNA expressing two sgRNAs under U6.1 and U6.3 promoters, respectively.
**(D)** Secondary structure of sgRNAs targeting *Ythdc1* and *Ythdf*.




**Supplemental Figure S10.** Generation of gene knock-outs by targeted deletions. **(A)** Schematic of a the *Drosophila CMTR1* gene indicating transcribed parts in boxes with translated parts in black. Gaps between boxes indicate introns which are spliced out. The triangle indicates a positively marked transposon. The *w+* marked P element was mobilized over a *w+* marked chromosomal deficiency and a deletion starting at the P element insertion site and extending into the catalytic domain was recovered (deletion *CMTr1*$^{13A}$). **(B)** Schematic of a typical *Drosophila* gene indicating the position of sgRNAs and a deletion to be generated to obtain a knock-out below the gene model. The line indicates the desired deletion and flanking dashed lines indicate positions where gRNAs can be selected. Below are primers indicated to screen for the deletion which will result in a short PCR product if the deletion is present compared to a long PCR product in the absence of a deletion, which will likely not amplify efficiently. Then, to screen for re-insertions, primers are indicated flanking an intron in the deleted part. If re-integration occurs, a short PCR product will be amplified for integration in the sense direction. If a longer PCR product is amplified, indicating a non-spliced intron, integration will be in the anti-sense direction.
**(C)** Schematic of a typical human gene composed of short exons and large introns with the translated part marked in black. Removal of an exon can lead to a frameshift resulting in a truncated protein, that is non-functional (below). The line indicates the desired deletion induced by the indicated sgRNAs. RT-PCR with primers in exons flanking the deleted exon can be used to monitor CRISPR/Cas9 induced deletion efficiency in cultured cells based on the ration of splice products.