Structure-optimized sgRNA selection with PlatinumCRISPr for efficient Cas9 generation of knock-outs

IRMGARD U. HAUSSMANN[1,2], THOMAS C. DIX[1], DAVID W. J. MCQUARRIE[1], VERONICA DEZI[1], ABDULLAH I. HANS[1], ROLAND ARNOLD[3,4,5] AND MATTHIAS SOLLER[1,5,6]

[1]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[2]College of Life Science, Birmingham City University, Birmingham, B5 3TN, United Kingdom

[3]Institute of Cancer and Genomics Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[4]Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[6]Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, United Kingdom

Running title: Structure-optimized sgRNAs for efficient Cas9

Key Words: sgRNA/Cas9 optimization, secondary structure, retro-transposition

[5] Corresponding authors

matthias.soller@manchester.ac.uk R.Arnold.2@bham.ac.uk

Abstract

A single guide RNA (sgRNA) directs Cas9 nuclease for gene-specific scission of double-stranded DNA. High Cas9 activity is essential for efficient gene editing to generate gene deletions and gene replacements by homologous recombination. However, cleavage efficiency is below 50% for more than half of randomly selected sgRNA sequences in human cell culture screens or model organisms. We used in vitro assays to determine intrinsic molecular parameters for maximal sgRNA activity including correct folding of sgRNAs and Cas9 structural information. From comparison of over 10 data sets, we find major constraints in sgRNA design originating from defective secondary structure of the sgRNA, sequence context of the seed region, GC context and detrimental motifs, but we also find considerable variation among different prediction tools when applied to different data sets. To aid selection of efficient sgRNAs, we developed web-based PlatinumCRISPr, an sgRNA design tool to evaluate base-pairing and sequence composition parameters for optimal design of highly efficient sgRNAs for Cas9 genome editing named PlatinumCRISPr. We applied this tool to select sgRNAs to efficiently generate gene deletions in *Drosophila Ythdc1* and *Ythdf*, that bind to N^6 methylated adenosines (m^6A) in mRNA. However, we discovered, that generating small deletions with sgRNAs and Cas9 leads to ectopic reinsertion of the deleted DNA fragment elsewhere in the genome. These insertions can be removed by standard genetic recombination and chromosome exchange. These new insights into sgRNA design and the mechanisms of CRISPR-Cas9 genome editing advances efficient use of this technique for safer applications in humans.

Introduction

Bacterially derived Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated protein 9 (Cas9) from *Streptococcus pyogenes* provides a powerful tool for precise genome editing (Garcia-Doval and Jinek 2017; Jiang and Doudna 2017; Hille et al. 2018; Doudna 2020). To induce double stand-breaks in DNA at desired locations, the DNA scission enzyme Cas9 uses a guide RNA (gRNA) containing a 20 nucleotide complementary sequence to the genomic target site (protospacer), which also requires the protospacer adjacent motif (PAM) at the 3'end, comprised of an NGG sequence (whereby N is any nucleotide and G is guanine). In addition to the target complementary sequence (spacer) the gRNA also contains a constant crispr RNA (crRNA) sequence that base-pairs with *trans*-activating crRNA (tracrRNA). Alternatively, a single gRNA (sgRNA) can be used whereby the crRNA is fused to the tracrRNA through an artificial loop (Jinek et al. 2012; Cong et al. 2013).

High efficiency CRISPR-Cas9 mediated DNA scission is essential to generate mutants at high frequency in genetic screens and to provide the resource for efficient homologous recombination directed gene replacements. Large scale analysis of sgRNA efficiencies revealed the whole spectrum of on-target cleavage activities ranging from 0-100% arguing for a number of parameters that need to be correct for high efficiency cleavage leading to models incorporating weighing of features and/or thermodynamics of secondary structures (Hsu et al. 2013; Doench et al. 2014; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Hart et al. 2015; Housden et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Xu et al. 2015; Doench et al. 2016; Liu et al. 2016; Abadi et al. 2017; Gandhi et al. 2017; Chuai et al. 2018; Labuhn et al. 2018; Graf et al. 2019; Zhang et al. 2019; Michlits et al. 2020; Sledzinski et al. 2020; Trivedi et al. 2020; Xiang et al. 2021; Riesenberg et al. 2023). These studies identified that

70 sequences with very low (≤35%) or very high (>80%) guanine-cytosine (GC) overall content were
71 less effective indicating a critical aspect for binding energy in target scission. In addition, purines in
72 the six nucleotides 5` to the PAM substantially increased Cas9 cleavage efficiency, while
73 pyrimidines and in particular uridine resulted in a lower efficiency (Ren et al. 2014; Wang et al.
74 2014; Housden et al. 2015; Graf et al. 2019). The lower efficiency of two uridine preceding the
75 PAM site was further associated with premature termination of RNA Pol III (Graf et al. 2019),
76 which can terminate after a stretch of four to six uridines (Gao et al. 2018). Moreover, changes in
77 internal structure of sgRNA has been found associated with low activity (Moreno-Mateos et al.
78 2015; Thyme et al. 2016; Jensen et al. 2017). Recently, also bioinformatic approaches employing
79 machine-learning have been used to improve prediction of sgRNA cleavage efficiencies (Xiang et
80 al. 2021). These observations have helped to improve the design of sgRNAs to yield higher
81 efficiencies and were incorporated into sgRNA design tools, but correlations between predictions
82 and guide activity vary considerably (Haeussler et al. 2016; Labun et al. 2016; Sledzinski et al.
83 2020). Accordingly, available rules to predict sgRNAs are currently not sufficient to guarantee high
84 cleavage efficiency and many sgRNA candidates scoring high fail to cleave efficiently (Haeussler et
85 al. 2016; Labun et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020). In particular, the impact of
86 sgRNA folding has not yet been analysed in detail and incorporated in web tools for sgRNA design.
87 The X-ray crystal structure of Cas9 bound to sgRNA has been determined and indicates four
88 regions of base-pairing termed tetraloop (tetraloop forms due to fusion of crRNA and tracrRNA)
89 and stem loops 1-3 that might be important for its function (Anders et al. 2014; Nishimasu et al.
90 2014). This structure revealed many points of close interactions of the folded sgRNA with Cas9, but
91 whether disruptions in the sgRNA structure would impact on Cas9 cleavage efficiency has not

92 systematically been analyzed (Riesenberg et al. 2023). In particular, highly GC-rich guide RNAs
93 could disrupt the rather weak secondary structure of the sgRNA bound by Cas9.
94 The CRISPR-Cas9 genome editing tool is widely used to generate knock-out mutants by
95 introducing frameshifts. It has been recognized that introducing premature termination codons
96 (PTCs) can induce use of alternative translation initiation sites (Tuladhar et al. 2019). In addition, in
97 CRISPR-Cas9 engineered "knock-outs" of the m^6A mRNA methyltransferase METTL3, it has been
98 found that a functional ORF can be restored by altered splicing leaving considerable levels of m^6A
99 in mRNA (Poh et al. 2022). Likewise, compensatory responses have been observed involving
100 upregulating genes and as a consequence causing stronger phenotypes than compared to removing a
101 gene entirely (El-Brolosy et al. 2019; Ma et al. 2019). In addition, some genes have dual functions
102 as protein and RNA (Hachet and Ephrussi 2004). Hence, introducing a frameshift will only remove
103 the protein function. Likewise, many non-coding RNAs are present in introns suggesting that their
104 expression is connected to the expression of the host gene (Boivin et al. 2018), but such
105 relationships have not yet been explored comprehensively as they require more sophisticated
106 genome editing (Deveson et al. 2017). Thus, for generating gene knock-outs, removing the
107 transcription start site or the entire gene should be considered.
108 The sequence complementary guided targeting of the gRNA/Cas9 complex makes this an ideal tool
109 for genome editing, but its use is currently limited by the low predictability to cut its target in the
110 genome (Haeussler et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020). In particular, the role of
111 RNA secondary structure of sgRNA has not been thoroughly evaluated and there is no sgRNA
112 prediction tool that comprehensively takes RNA secondary structure of sgRNAs into account. To
113 facilitate design of optimal sgRNAs, we developed an online tool incorporating all currently known
114 parameters for sgRNA design including correct sgRNA folding (https://platinum-

115 crispr.bham.ac.uk/predict.pl). Moreover, we provide guidelines to generate complete gene
116 knockouts by generating deletions using dual sgRNAs. Taken together, these new sgRNA design
117 and gene editing tools will help to develop CRISPR-Cas9 genome editing for safe application in
118 humans.
119

120 ## Results

121 **RNA secondary structure constraints limit sgRNA/Cas9 activity**

122 Although sgRNA/Cas9 can cleave DNA efficiently, the first sgRNAs (L11GC and R13GC) we
123 designed according to previously published guidelines did not cut the *pUC 3GLA Dscam 3-5*
124 reporter we designed to study *Dscam* alternative splicing from introducing mutations by gap repair
125 recombineering (**Fig. 1A,B, Supplemental Fig. S1A, Supplemental Table S1A**)(Hemani and
126 Soller 2012; Haussmann et al. 2019). Similarly, the first sgRNAs flanking the *Drosophila Ythdf*
127 gene, a reader for m^6A mRNA methylation did not result in a deletion of the locus based on
128 screening for loss of an RFP-marked transposon even though we had validated that the target
129 sequence in this strain does not contain sequence polymorphisms (n=103, **Supplemental Fig. S2, ,**
130 **Supplemental Table S1B**)(Balacco and Soller 2019). To determine the intrinsic molecular
131 parameters for maximal sgRNA activity, we devised an in vitro assay to test the DNA scission
132 efficiency of these two sgRNA based on in vitro transcribed sgRNAs and commercially available
133 Cas9 using oligonucleotide and plasmid substrates containing matching protospacer sequences
134 followed by a PAM.

135 It has previously been shown that extending the tetraloop in the constitutive component of sgRNAs
136 constituted by tracrRNA and crRNA enhances cleavage efficiency in vitro using oligonucleotide
137 substrates (**Fig. 1C**)(Jinek et al. 2012). Introducing the extended sequence present in tracrRNA and

138 crRNA into sgRNA (L7GCext) did not increase efficiency of cleaving a plasmid at 37ºC, but
139 increasing the temperature to 42ºC enhanced cleavage by L7GCext (**Fig. 1B**). Increasing the salt
140 concentration to 200 mM also did not result in enhanced cleavage by L7GC/R3G (**Fig. 1D**). In
141 contrast, both L7GC and L7GCext could cleave an oligonucleotide, while R13GC did not (**Fig. 1E**).
142 Using this assay also confirms the previous observation that sgRNAs of shorter length will lead to
143 cleavage. In addition, three guanosines introduced at the 5'end of sgRNAs required for efficient in
144 vitro transcription are tolerated (**Fig. 1F,G**)(Jinek et al. 2012).
145 The sgRNA scaffold adopts a typical fold when bound to Cas9 consisting of the bulged tetraloop,
146 followed by small loop 1 and the more extended loops 2 and 3, which form a protective 3'end
147 structure (**Fig. 1C**)(Nishimasu et al. 2014). The loop2/3 structure does not involve the uridines
148 incorporated for termination of RNA Pol III driven expression from plasmids (**Fig. 1C**). When
149 comparing the secondary structures of the four sgRNAs, we noticed that the two well-cutting
150 sgRNAs L7GC or R3G maintained the secondary structure of the constitutive RNA part, while the
151 non-cutting sgRNA L11GC disrupted the structure of the tetraloop (**Figs. 1H-K**). The effect of
152 R13GC seems more subtle as it could cut in the oligonucleotide assay suggesting that the repeated
153 bulge structure is the cause for its inefficiency, which is supported by X-ray crystal structure of the
154 Cas9-sgRNA-DNA complex. Here, the bulge structure is recognized by Cas9 where Tyr_{359} base-
155 stacks with G_{43}, that also forms hydrogen bonds with Asp_{364} and Phe_{351} , and Phe_{351} forms a
156 hydrogen bond with A_{42} (Nishimasu et al. 2014). In addition, the sgRNAs initially used for deleting
157 the *Ythdf* gene have a severely disrupted secondary structure (**Supplemental Fig. S2**).
158 When we systematically analysed genome sequences from *Drosophila* or humans for correct
159 folding and activity of sgRNAs using the above parameters, about 50% of sgRNAs (241 from 481
160 and 503 from 973, respectively) did not fold properly. Also, only about 10-20 % of randomly

161 selected sgRNAs exert high cleavage efficiency suggesting that correct folding could essentially
162 contribute to high cleavage efficiency (Hsu et al. 2013; Doench et al. 2014; Gagnon et al. 2014; Ren
163 et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Moreno-Mateos et al.
164 2015; Doench et al. 2016; Liu et al. 2016; Abadi et al. 2017; Chuai et al. 2018; Labuhn et al. 2018;
165 Graf et al. 2019; Zhang et al. 2019; Michlits et al. 2020; Sledzinski et al. 2020).
166 When comparing the sequences of the cutting sgRNAs L7GC or R3G with the non-cutting L11GC
167 and R13GC sgRNAs, we further noticed that L11GC and R13GC sgRNAs contained more
168 guanosines, which in RNA can base-pair with C and U. To test if guanosines in the sgRNA limit
169 Cas9 activity we increased their number in R3G to 13 to make sgRNA 13G (**Fig. 2A**). For the
170 design of the R13 sgRNA, care was taken not to disrupt the tetraloop, but we noticed the potential
171 to interfere with loop 2 (**Fig. 2B**, see below). We noticed, that sgRNA 13G is not capable of
172 directing Cas9 cleavage if the target sequence is present in a 3 kb plasmid, but is active with a short
173 oligonucleotide substrate (**Figs. 2A-D**). Likewise, adding a restriction enzyme together with
174 sgRNA/Cas9 inhibited Cas9 in cleaving plasmid DNA suggesting that Cas9's ability to scan DNA
175 can be impaired separately from its ability to cleave DNA.
176 Since the increased number of Gs in sgRNA R13G lead to enhanced base-pairing, we exchanged
177 the Gs with Cs leading to an open structure in the seed region. This sgRNA R13C cleaved the test-
178 plasmid efficiently (**Figs. 2C and 2E**). Introducing Cs in the left or right half of sgRNA R13G lead
179 to short stem loops and inefficient cleavage of the test-plasmid (**Figs. 2C, 2F and 2G**).
180 To further test to what extent base-pairing impacts on Cas9 activity, we generated sgRNAs
181 L10ds6G and R10ds6GC, where the proximal or the distal half leads to complementary base-pairing
182 of the gRNA with the constant part, respectively (**Figs. 3A,B**). Although both sgRNAs supported
183 Cas9 cleavage of oligonucleotide substrate, the R10ds6GC sgRNA base-pairing with the proximal

184 part was mostly inactive in cleaving the plasmid indicating an impaired ability of Cas9 to scan DNA
185 (**Figs. 3C,D**).
186 Taken together, these results demonstrate that the structure of the sgRNA is important for efficient
187 Cas9 mediated DNA scission in vitro. Furthermore, high G content and base-pairing in the distal
188 part of the gRNA also impairs DNA scission, while base-pairing in the proximal part is tolerated.
189 To further substantiate these findings, we analyzed the structures of sgRNAs from previous studies
190 in mammalian cells and *Drosophila* with regard to their cleavage efficiency of previous attempts to
191 define rules for sgRNA cleavage efficiency in vivo(Ren et al. 2014; Graf et al. 2019). Indeed, in 39
192 sgRNAs designed for use in *Drosophila* reduced cleavage efficiency in nine sgRNAs is associated
193 with disturbances of the sgRNA secondary structure resulting in a cleavage efficiency below 35%
194 (**Supplemental Figs. S3 and S4, Supplemental Table S1C**)(Ren et al. 2014). Similarly, from 22
195 sgRNAs designed for use in mammalian cells, 13 had a cleavage efficiency below 35% associated
196 with disturbances of the sgRNA secondary structure (**Supplemental Fig S5, Supplemental Table**
197 **S1D**)(Graf et al. 2019). Similar result were also observed for the efficiency of sgRNAs in honey
198 bees(Roth et al. 2019).
199 Given the requirement for correct folding of the sgRNA for efficient Cas9 mediated DNA scission,
200 we further examined the X-ray crystal structure of the Cas9-sgRNA-DNA complex to see whether
201 this would provide additional instructions to design sgRNAs (Nishimasu et al. 2014). Indeed, the
202 first two nucleotides, adenosine 51 and 52 (A_{51} and A_{52}, **Supplemental Figs. S6 and S7**) after the
203 tetraloop form an aromatic base-stacking interaction with phenylalanine 1105 (Phe_{1105}) of Cas9.
204 Furthermore, these interactions are stabilized by guanosine 62 (G_{62}) forming non-Watson Crick
205 hydrogen bonds with A_{51} A_{52} and Phe_{1105}, and uracil 63 (U_{63}) forms a base-stacking interaction with
206 A_{52}. These interactions indicate that base-pairing of the gRNA with these nucleotides of the

207 constant part of the sgRNA reduce Cas9 activity in Cas9 cleavage assays in vitro and mutagenesis
208 in vivo.
209
210 **Large scale evaluation of novel sgRNA design parameters**
211 Next, we incorporated all the features previously published and from this study into a bioinformatic
212 sgRNA design tool, https://platinum-crispr.bham.ac.uk/predict.pl (Supplemental Code)(Hsu et al.
213 2013; Doench et al. 2014; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015;
214 Farboud and Meyer 2015; Moreno-Mateos et al. 2015; Doench et al. 2016; Liu et al. 2016; Abadi et
215 al. 2017; Chuai et al. 2018; Labuhn et al. 2018; Graf et al. 2019; Zhang et al. 2019; Michlits et al.
216 2020; Sledzinski et al. 2020). The included features include validation of intact secondary structures
217 (tetraloop, loop 2 and 3, **Figs. 1-3**), presence of a tetraloop bulge mimic (**Fig. 1K**), self-
218 complementarity of the gRNA (nts 1-20, **Figs. 2F,G**)(Moreno-Mateos et al. 2015; Thyme et al.
219 2016; Jensen et al. 2017), GC content in the six nucleotide seed region of the gRNA (nts 15-20,
220 **Supplemental Table S1A, S1C and S1D**, (Ren et al. 2014; Wang et al. 2014; Graf et al. 2019), GC
221 content of the gRNA (nts 1-20, **Supplemental Table S1A, S1C and S1D**, (Ren et al. 2014; Wang
222 et al. 2014; Graf et al. 2019), the UUYY motif (nts 16-20), which results in complete base-pairing
223 (**Fig. 3**) and can act as a Pol III termination signal(Gao et al. 2018), the UCYG and CYGR motifs
224 (nts 16-20) associated with lower cleavage efficiency (Graf et al. 2019) and lack of base-pairing of
225 nucleotides 40, 41, 51 and 52 that are engaged in contacts with Cas9 (**Supplemental Figs. S6 and**
226 **S7**).
227 PlatinumCRISPr operates at a cut-off prediction for either low or high efficiency, which is based on
228 whether sgRNA folding is correct or whether certain detrimental motifs are present, but also on
229 sequence composition. Although a score could be implemented for sequence composition, in

230 practice, cleavage efficiency increases within an narrow range making a score at this point not
231 feasible.
232 The sequence of any organism can be used with the PlatinumCRISPr server as long as the genome
233 is sequenced (e.g. accessible via the Ensembl Genome browser). For the most common model
234 organisms, the PlatinumCRISPr server allows to choose the organism and will calculate off targets
235 (additional hits in the genome) as complete match for the 16 nucleotides before the tracrRNA as this
236 is the minimal sequence for DNA scission (Fig. 1G). For other organisms off target effects need to
237 be determined manually by BLAST, e.g. using the NCBI webserver.
238 Although DNA scission is completely abrogated by single mismatch to the target gene in
239 *Drosophila* (Ren et al. 2014), off target scission has been reported in tissue culture cells (Fu et al.
240 2013; Hsu et al. 2013). To compare off target effects with sgRNA cleavage efficiencies predicted
241 by either PlatinumCRISPr, DeepSpCas9 or Azimuth (standard scoring in CHOPCHOP) we used a
242 large scale double-strand break analysis from Cas9 expression across 59 targets using tagmentation-
243 based tag integration site sequencing (TTISS)(Schmid-Burgk et al. 2020). In this analysis
244 comparing off target effects predicted by Platinum CRISPr or CHOPCHOP, we find that off target
245 effects measured in tissue culture cells increase with predicted cleavage efficiency for
246 PlatinumCRISPr (one-sided Wilcoxon Rank-Sum test p=0.009) and DeepSpCas9 (Spearman's rho,
247 p=0.0016), but not for CHOPCHOP (rho 0.1, p=0.41).
248 To identify additional parameters affecting sgRNA cleavage efficiency, we performed a motif
249 analysis among the 35% low scoring sgRNAs for a number of different data sets (Doench et al.
250 2014; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer
251 2015; Hart et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Xu et al. 2015; Doench et
252 al. 2016; Gandhi et al. 2017; Xiang et al. 2021), but we did not find motifs associated with low

253 performance in individual data sets. In particular, we found no evidence for 4 T's to be enriched,
254 that could act as a termination signal for RNA Pol III (Gao et al. 2018; Graf et al. 2019).
255 We then analysed the performance of a *Drosophila* sgRNA data set (Ren et al. 2014) according to
256 sgRNA design parameters described above. Our design tool PlatinumCRISPr selected 12 from 39
257 sgRNAs and those showed a cleavage efficiency of 58 % or more (**Fig. 4A**). We then analysed a
258 number of sgRNA prediction tools for this data set including Chariscore (Chari et al. 2015),
259 Crispron (Xiang et al. 2021), DeepSpCas9 (Kim et al. 2019), DoenchScore (Doench et al. 2014),
260 Azimuth (implemented in ChopChop)(Doench et al. 2016), Moreno-Mateos Score (Moreno-Mateos
261 et al. 2015), Wang Score (Wang et al. 2014), Wong Score (Wong et al. 2015) and Xu Score (Xu et
262 al. 2015). PlatinumCRISPr significantly outperformed all of these prediction tools with the
263 *Drosophila* data set (Ren et al. 2014)(**Fig. 4B**).
264 Next, we analysed 14 data sets from various organisms (*Drosophila*, zebrafish, sea squirt, worms
265 and cell culture cells,), which determined sgRNA cleavage efficiency for their performance using
266 the PlatinumCRISPr design tool (Doench et al. 2014; Gagnon et al. 2014; Ren et al. 2014; Wang et
267 al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Hart et al. 2015; Moreno-Mateos et al. 2015;
268 Varshney et al. 2015; Xu et al. 2015; Doench et al. 2016; Gandhi et al. 2017; Xiang et al. 2021). For
269 overall performance (**Fig. 5**), six data sets yielded significant ($p \leq 0.05$) enrichment of high
270 efficiency performing sgRNAs (Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Moreno-
271 Mateos et al. 2015; Xiang et al. 2021), and five showed enrichment ($p \leq 0.25$)(Gagnon et al. 2014;
272 Farboud and Meyer 2015; Hart et al. 2015; Varshney et al. 2015; Xu et al. 2015), while two failed
273 to show enrichment for most of the parameters (Doench et al. 2014; Doench et al. 2016). In this
274 analysis, we noticed that structural constraints were significantly more important in cold-blooded
275 organism, where sgRNAs delivery is by injection in the absence of selection in contrast to cell

276 culture cells, where delivery is by transfection and selection for chronic exposure to sgRNAs for up
277 to 10 days before analysis.
278 When we analysed the performance of PlatinumCRISPr, Chariscore (Chari et al. 2015), Crispron
279 (Xiang et al. 2021), DeepSpCas9(Kim et al. 2019), Doench Score (Doench et al. 2014), Azimuth
280 (implemented in ChopChop)(Doench et al. 2016), Moreno-Mateos Score (Moreno-Mateos et al.
281 2015), Wang Score (Wang et al. 2014), Wong Score (Wong et al. 2015), Xu Score (Xu et al. 2015)
282 and CRISPick (Sanson et al. 2018) prediction tools (**Supplemental Figs. 8A-L**) with the different
283 data sets determining sgRNA cleavage efficiency we found that Wang Score performed best on the
284 Doench data set and that PlatinumCRISPr and Moreno-Mateos performed best with *Drosophila* and
285 zebrafish generated data sets, respectively, but we found no single prediction tool that stood out. In
286 addition, in five out of six cell culture generated data sets none of the prediction tools outperformed
287 the others or substantially increased prediction efficiency (**Supplemental Figs. 8A-L**).
288 In addition, we have also analyzed datasets for performance of sgRNA libraries in different cell
289 lines (HeLa, HCT 116 and GBM) (Hart et al. 2015) and find very high correlations for all
290 comparisons (Spearman's correlation between 0.71 and 0.77) indicative of minimal differences of
291 the cleavage efficiency of sgRNAs in these cell lines. In addition, sgRNAs selected by
292 PlatinumCRISPr or CHOPCHOP show a significantly higher cleavage efficiency than discarded
293 sgRNAs in all three cell lines ($p \leq 0.03$).
294 The DepMap project provides efficacy scores for 193476 gRNAs (between 0 and 1, the median
295 efficacy in the complete data-set is 0.974) (DepMap 2024), of which ~25% are predicted as cutting
296 by PlatinumCRISPr. A comparison of DepMap gRNAs resulted in a minimally better performance
297 (0.006 %, median efficacy: 0.975) of guides predicted as non-cutting by PlatinumCRISPr, but the
298 biological significance of this observation remains unclear since the overall performance

299 substantially exceeded gRNA performance of other cell-line based experiments with chronic
300 exposure (Supplemental Fig. S8), possibly because gRNAs were preselected (DepMap 2024).
301 As part of this analysis, we noticed that the average cleavage efficiency in the analyzed data sets
302 varied substantially (from 20-75% average cleavage efficiency, **Supplemental Figs. 8A-L**).
303 pointing towards a bias in outcome when testing various prediction tools with different data sets. To
304 identify common patterns among the different prediction tools applied to different data sets in the
305 following analysis we therefor excluded data sets with cleavage efficiencies below 30% for further
306 analysis (Gagnon et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Doench et al. 2016).
307 The default element of ChopChop is Azimuth (Labun et al. 2019), but also implements elements
308 from Doench Score, Chari Score, Xu Score and Moreno-Mateos Score. We therefore reasoned that
309 a combination of two prediction tools could result in more reliable selection of high efficiency
310 cleaving sgRNAs. When we combined two prediction tools the highest scoring combination was
311 PlatinumCRISPr together with Wong score with an average cleavage prediction of 61% (**Fig. 6A**)
312 and this combination also outperformed in all the remaining data sets (**Fig. 6B**), but this increased
313 the stringency and only very few sgRNAs were selected.
314

315 **A *Drosophila* transformation vector for expression of two sgRNAs**

316 Next, we applied these novel sgRNA design rules to generate *Drosophila* gene deletions. For this
317 purpose we generated a new fly transformation vector with a GFP marker (Solomon et al. 2018),
318 that is easier to select than previously generated *vermillion* marked vectors that require a *vermillion*
319 mutant background for transgene identification(Port et al. 2014; Trivedi et al. 2020). This vector
320 expresses two 20 nt sgRNAs from *U6.1* and *U6.3* promoters (**Supplemental Figs. S9A-C**),
321 harboring a G as first nucleotide as a requirement for expression from the *U6* promoter (Paule and

322 White 2000; Ren et al. 2013). This plasmid can be generated by incorporating the two sgRNA
323 sequences in PCR primers for single step cloning into the plasmid, while previously published
324 vectors require two cloning steps or plasmid recombination (Port et al. 2014; Trivedi et al. 2020).
325 This sgRNA vector can then be injected into *Drosophila* expressing Cas9 in the germline for
326 CRISPR to induce mutations. Alternatively, this vector can be used to generate a transgenic line via
327 the attB site using phiC31 integrase mediated transformation. This fly strain is then crossed to a line
328 expressing Cas9 in the germline for generation of the desired genetic lesion. Transgenically
329 provided sgRNA/Cas9 generally results in a higher efficiency, because the sgRNAs are been
330 provided maternally. Moreover, excision efficiency can be further increased by mutagenesis over a
331 chromosomal deficiency which will removing the repair template for homologous recombination.
332

333 **Efficient generation of gene deletions by sgRNA/Cas9 using transposon markers**

334 We then applied the PlatinumCRISPr tool to identify high efficiency sgRNAs to generate gene
335 deletions in *Drosophila Ythdc1* and *Ythdf*, that bind to N^6 methylated adenosines (m^6A) in mRNA
336 important in development (Dezi et al. 2016; Haussmann et al. 2016; Roignant and Soller 2017;
337 Balacco and Soller 2019; Anreiter et al. 2021). *Ythdc1* and *Ythdf* are located on the third
338 chromosome and transposon inserts *Mi{MIC}YT521-B*^{MI02006} and *PBac{SAstopDsRed}*^{LL04081}
339 marked with *GFP* or *RFP*, respectively, are available from stock centers. They were combined with
340 an X-linked *vasCas9* or *nosCas9* for germline expression of Cas9. To allow for detection of
341 transposon loss in the YTH protein genes the GFP and RFP markers of the *vasCas9* insert had been
342 removed. These flies were then crossed to the GFP marked sgRNA construct inserted on the 3rd
343 chromosome (**Figs. 7A and F**). The sgRNA insert has a weak GFP marker and can generally be
344 distinguished from *Mi{MIC}YT521-B*^{MI02006} inserts. Females from this cross were mated with males

345 containing *TM3 Sb/TM6 Tb* double-balancers in single crosses to recover the sgRNA induced
346 individual deletions and avoid analysis of clonal events. The male progeny was then screened for
347 loss of the GFP or RFP marker. Males which had lost the marker were detected in 100 % (n=9) of
348 the crosses for *Ythdc1* and in 88 % (n=9) for *Ythdf*, respectively. This is a substantial increase of
349 efficiency over imprecise P-element excision with a frequency of 0.01 to 1% (Soller et al. 2006;
350 Haussmann et al. 2016; Haussmann et al. 2022). In those crosses with loss of markers, all males for
351 *Ythdc1* had lost the GFP marker, while for *Ythdf* the average frequency of marker loss was 42 %
352 (n=8). In the reverse cross for *Ythdf*, where sgRNA/Cas9 mediated excision is induced in males, the
353 frequency was 0% (n=5) indicating that the low frequency in males is linked to the absence of
354 recombination in males. The identified single marker-less males were then crossed to *TM3 Sb/TM6*
355 *Tb* double-balancers to establish a line and analysed with PCR using primers next to the deletion
356 breakpoints yielding a short PCR product (**Figs. 7B and G**). A PCR product had been obtained in
357 all lines were the marker had been lost indicating that the expected deletion had indeed been
358 generated. To generate *Ythdc1* excision lines *nosCas9* was used for germline expression of Cas9 as
359 *vasCas9* together with sgRNAs targeting *Ythdc1* resulted in female sterility.
360

361 **sgRNA/Cas9 scissioned fragments insert elsewhere in the genome**

362 After establishing the lines, we noticed that all lines (n=10) established for the *Ythdc1* deletion did
363 not show the flightless phenotype previously reported(Haussmann et al. 2016). Therefore, we
364 selected four lines for further analysis by RT-PCR from RNA (*Ythdc1* excision lines, **Fig. 7C**) or of
365 genomic DNA (*Ythdf* excision lines, **Fig. 7H**) of homozygous flies with primers that were within
366 the deletion and also flanked an intron. Unexpectedly, a copy of the gene was still present in all

367 *Ythdc1* and *Ythdf* excision lines analyzed suggesting that the deleted fragment had been inserted
368 elsewhere in the genome.
369 To remove this ectopic insert(s), positive lines were crossed to *w+* marked deficiencies and out-
370 crossed for two generations. The X and 2nd chromosomes were then exchanged to establish null-
371 mutant lines from single chromosomes for *Ythdc1* and *Ythdf* that were confirmed by RT-PCR to be
372 free of any ectopic inserts (**Figs. 7D,E and Figs. 7I,J**). To avoid such complications in the future
373 we generated a *PBac w+* containing vector which can be efficiently inserted into a locus by cloning
374 left and right homology arms either to induce a partial deletion upon insertion. Alternatively,
375 mutations in sgRNA cleavage sites can be introduced into the homology arms to insert point
376 mutations. Afterwards, the *PBac w+* is removed scar less by a transposase.
377

378 ## Discussion

379 **sgRNA structure and sequence composition contribute to sgRNA/Cas9 DNA scission**
380 **efficiency**

381 DNA scission by the sgRNA/Cas9 complex is highly specific and requires complete base-pairing
382 between the sgRNA and the target DNA generally not tolerating single miss-matches(Ren et al.
383 2014; Farboud and Meyer 2015). This feature makes the sgRNA/Cas9 complex an ideal tool for
384 genome editing, but its use is currently limited by the low predictability to cut its target in the
385 genome(Haeussler et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020).
386 Here, we discovered that the structure of the sgRNA is a key determinant for the scission efficiency
387 of the sgRNA/Cas9 complex in *Drosophila*. However, only about 50% of sequences adjacent to
388 PAM sites constitute sgRNAs that fold properly or are not compromised by unfavorable base-
389 pairing. In support of these two levels of interference, sgRNA R13GC correctly folds the tetraloop

390 and loop2/3, but did not cleave a short oligonucleotide substrate. This indicates that the structure of
391 this sgRNA blocks the catalytic activity of the sgRNA/Cas9 complex, likely by mimicking the
392 bulge structure of the tetraloop. Second, some sgRNAs allowed cleavage of short oligonucleotide
393 substrates (e.g. L7GC and R10ds6GC), but did not support efficient DNA scission of the target
394 sequence in the context of a 3 kb test-plasmid. Likely, these sgRNAs interfere with the ability of the
395 sgRNA/Cas9 complex to scan the DNA for target sites.
396 When we analyzed cleavage efficiencies of sgRNAs used in *Drosophila* (Ren et al. 2014), we
397 observed a good overlap with the ability of those sgRNAs to adopt the correct structure and having
398 a high cleavage efficiency. Further refinement to the design of sgRNAs comes from the recognition
399 that the GC content in the seed region is a major determinant to cleavage efficiency in addition to
400 general GC content. Analysis of the X-ray crystal structure of the Cas9-sgRNA-DNA complex also
401 revealed that the two As in the tracrRNA before the tetraloop engage with Cas9 through base
402 stacking and hydrogen bonds (Nishimasu et al. 2014). Base-pairing of these two As with Us at the
403 end of the sgRNA (N_{19} and N_{20}) before the start of the tracrRNA impact sgRNA/CAS9 complex
404 function and reduce cleavage efficiency (Graf et al. 2019). Likewise, if sgRNAs are made by RNA
405 Pol III, termination occurs at the boundary of the tracrRNA if two UU precede the GUUUU of the
406 start of the tracrRNA (Arimbasseri and Maraia 2015; Graf et al. 2019). Folding of the sgRNA is not
407 anticipated to impact on transcription as it occurs afterwards. Motif searches in various dataset to
408 determine sgRNA cleavage efficiencies did not reveal any further motifs that impact on cleavage
409 efficiency. If such bias exists, this would likely have been exploited by parasites of prokaryotic
410 hosts.
411
412 **Structural constraints of sgRNAs are more pronounced in cold-blooded animals**

413 The bacterial CRISPR-Cas9 system consists of two RNAs (crRNA and trRNA) that assemble with
414 Cas9 to form the active complex. crRNA and trRNA base-pair through sequence complementarity
415 to form the tetra loop in the active Cas9 complex (Garcia-Doval and Jinek 2017; Jiang and Doudna
416 2017; Hille et al. 2018). In type II systems the tracrRNA is required for crRNA maturation
417 suggesting that base-pairing takes place while being assembled with Cas9. After the crRNA is
418 trimmed, the entire CRISPR-Cas9 complex can scan genomic DNA for DNA scission sites.
419 Alternatively, the crRNA could hybridize first to genomic DNA and recruit tracrRNA and Cas9 to
420 form a complex for DNA scission on site. In this scenario, the crRNA would not be able to interfere
421 with tracrRNA/Cas9 complex activity by forming an aberrant RNA secondary structure, but
422 whether this second scenario could be applied to more efficient genome editing with reduced off-
423 target cleavage needs to be tested. Of note, the effect of structural constraints are stronger in cold-
424 blooded animals likely reflecting that the optimal temperature for *E. coli* is 37ºC. In this context, it
425 would be worth exploring how much longer the sgRNA can be to still support Cas9 DNA scission
426 as longer RNAs would more stably hybridize to DNA. In either case, however, understanding
427 sgRNA/CAS9 complex assembly will inform how to prevent off-target DNA scission.
428

429 **Evaluation of different sgRNA prediction tools**

430 In this study, we also compared various sgRNA cleavage efficiency prediction tools with 12
431 datasets that have determined sgRNA cleavage efficiencies in human cells and various model
432 organisms including *Drosophila*, zebrafish, *C. elegans*, honey bees and seasquirt (Doench et al.
433 2014; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer
434 2015; Hart et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Wong et al. 2015; Xu et
435 al. 2015; Doench et al. 2016; Gandhi et al. 2017; Kim et al. 2019; Roth et al. 2019; Xiang et al.

436 2021). Here, the PlatinumCRISPr tool deemed best for *Drosophila*, Moreno-Mateos Score for zebra
437 fish and Wang Score for one human cell culture screen, but all prediction tools failed to convince
438 for the other five cell culture screens. Possibly, chronic exposure over several days in cell culture
439 systems could lead to a bias in determining sgRNA cleavage, compared to short exposure when
440 injected into early stage embryos like in insects and zebra fish.

441 Taken together, optimizing sequence composition and structural constraints in sgRNA design
442 essentially contributes to high DNA cleavage efficiency. Accordingly, optimized sgRNAs show
443 very little base-pairing with sequences adjacent to the loop 1 region, or are not complementary to
444 the tetraloop structure and/or the loop2/3 structure. In addition, avoiding base-pairing in the 10 nt
445 seed region prior to the PAM site predicts high efficiency of sgRNAs for DNA scission.
446 Furthermore, avoiding two Us before the PAM site prevents interference with the sgRNA/Cas9
447 structure.

448 In any case, introducing the target sequence into a plasmid allows to reliably determine the cleavage
449 efficiency of a particular sgRNA in an vitro assay. However, cellular features such as chromatin
450 state can also impact on Cas9 mediated DNA scission (Singh et al. 2015). These limitations of
451 CRISPR-Cas9 genome editing from biological parameters, including the impact of cellular
452 environment by RNA-binding proteins or chromatin structure, remain to be adequately
453 quantified. In this context, whether genes are expressed at the time of sgRNA cleavage likely also
454 plays a role in the observed cleavage efficiency as expression is associated with less compacted
455 DNA. Accordingly, expressed genes seem to be more efficiently mutagenized than
456 developmentally silenced genes (Port et al. 2014; Trivedi et al. 2020).

457 Initial concerns about CRISPR-Cas9 specificity were about off target cleavage, but changing one
458 nucleotide in the spacer sequence complementary to the target efficiently abrogates cleavage
459 activity (Ren et al. 2014). PlatinumCRISPr takes off-targets into account by screening the genome

460 for the 16 nucleotides distal of the PAM site, which require an exact match for cleavage, and all
461 four possible PAM sites.
462
463 **Limitations for generating gene knock-outs**
464 Generating deletions of entire genes, or essential parts of them is the preferred way to generate a
465 null allele. This approach will avoid complications arising from introducing frameshifts at the
466 beginning of the ORF, as translation could reinitiate from later AUG or CUG start codons
467 (Koushika et al. 1999). In addition, in some genes the RNA has functions on its own, as shown for
468 *oscar* RNA that forms a large RNP particle with Oscar and other RNA-binding proteins at the
469 posterior pole of a *Drosophila* oocyte (Hachet and Ephrussi 2004; Haussmann et al. 2011). Thus,
470 deletion of the entire gene region will discover such additional functions harbored in gene
471 transcripts.
472 When generating deletions of entire genes, we discovered that the deleted DNA fragment was
473 inserted into the genome and transcribed, resulting in the expression of protein that rescued the
474 flightless phenotype in *Ythdc1* deletion allele. Although the mechanism for the generation of these
475 new inserts is not known, it seemed not to have led to complex chromosomal aberrations, because
476 the inserts could be removed by standard recombination and/or exchange of chromosomes. Retro-
477 transposition has been observed in the *elav* gene, which led to the loss of all introns in *Drosophila*
478 (Samson 2008). Likewise, holometabolous insects generally have three *elav* genes, but honey bees
479 have only one *elav* gene. The honey bee *elav* gene, however, carries features of the other two genes
480 present in *Drosophila*. Hence, *elav* in honey bees could have collapsed in an ancestor from three to
481 one gene by some form of recombination to include parts specific to the other two *elav* genes
482 (Ustaoglu et al. 2021). Likewise, the *amyloid-β precursor protein* (*APP*) gene displays copy

483 number variation in the human brain, which are increased by retro-transposition in sporadic forms
484 of Alzheimer's disease suggesting evolutionary conserved mechanisms for reinsertion of genomic
485 information (Lee et al. 2018). Re-insertion of fragments cut out by the CRISPR-Cas9 system seems
486 not to be specific to *Drosophila* as it has also been observed in human cells (Geng et al. 2022).
487 Since intron loss and gain occurs during evolution, the mechanism underlying insertion of DNA
488 fragments from sgRNA induced deletions might be responsible for these changes. Thus, when
489 generating sgRNA induced gene deletions, it is essential to test for the absence of any transcripts by
490 RT-PCR, but also to validate a deletion at the DNA level by PCR using flanking primers or
491 Southern blots.

492 For a more reliable way to generate gene knock-outs, CRISPR-Cas9 induced deletion of only a part
493 of a gene seems more feasible and aligns with previous mutagenesis approaches using marked *P*
494 element transposons (Supplemental information and discussion, **Supplemental Fig. 10A**)(Soller et
495 al. 2006; Haussmann et al. 2016; Bawankar et al. 2021; Haussmann et al. 2022). Such targeted
496 deletions to 5'part of the gene can remove the core promoter consisting of the TATA box and
497 extend into a functional domain (e.g. catalytic domain, DNA-binding domain, RNA recognition
498 motif, etc, **Supplemental Fig. 10B**).

499 Since human genes are much larger and often have alternative transcription start sites (Soller 2006),
500 deletion of an exon such that splicing of the remaining exons generate a frameshift is a valuable
501 option (**Supplemental Fig. 10C**). Alternatively, a GFP cassette with a poly(A) site can be inserted
502 to terminate the ORF in the beginning, but it needs to be evaluated whether the poly(A) site in the
503 beginning of the gene is used or whether the exon containing the GFP cassette is skipped (Soller
504 2006; Wierson et al. 2020).

505 For a more reliable way to generate gene knock-outs in *Drosophila*, we have now developed a *PBac*
506 *w+* marker that can be inserted when generating a deletion (Dix 2022). In any case, however, a
507 marked locus will allow for rigorous cleaning of the genetic background.
508 In essence, we have established the rules for designing highly efficient sgRNAs and established
509 methodology to efficiently generate gene deletions. These findings have implications for other RNA
510 based methodologies including prime editing (Anzalone et al. 2019; Bosch et al. 2021).
511

512 ## Methods

513 **sgRNA/Cas9 directed DNA cleavage**

514 DNA templates for in vitro transcription were reconstituted from synthetic oligonucleotides. As
515 only the T7 promoter needs to be double-stranded for in vitro transcription, a T7 promoter
516 oligonucleotide (CCTGGCTAATACGACTCACTATAG) was annealed to an anti-sense Ultramer
517 (IDT DNA) encoding the entire sgRNA in addition to the T7 promoter. Alternatively, a 60 nt T7
518 promoter oligonucleotide with a partial sgRNA was annealed to an anti-sense oligonucleotide
519 encoding the tracrRNA
520 (AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTT
521 AACTTGCTATTTCTAGCTCTAAAAC) for 15 min at 40º C (2 μM) and made double-stranded by
522 extension with Klenow fragment of DNA polymerase I according to the manufacturer's instructions
523 (NEB). Klenow was then heat-inactivated 10 min at 85º C and oligonucleotides were desalted with
524 a G-50 Autoseq Sephadex spin column (GE) before using for in vitro transcription.
525 Then, sgRNAs were generated by in vitro transcription with T7 polymerase (T7 MEGAscript,
526 Ambion) from synthetic oligonucleotides (0.2 μM) and trace-labeled with ^{32}P alpha-ATP (800
527 Ci/mmol, 12.5 μM, Perkin Elmer) in a 20 μl reaction according to the manufacturer's instructions.

528 After DNase I digestion, free nucleotides were removed with a G-50 Probequant Sephadex spin
529 column (GE). Then, sgRNAs were heated for 2 min to 95° C and left at room temperature to adopt
530 folding. Then, sgRNAs were quantified by scintillation counting and analysed on 8-20 %
531 denaturing polyacrylamide gels as described (Dix et al. 2022).
532 For synthetic substrate DNAs the sense oligonucleotide (1 μM, sgRNA flanking sequences are:
533 TCGAGCATTATATGAAC-sgRNA-GGGTATTGGGGAATTCATTATGC) was labeled with ^{32}P
534 gamma-ATP (6000 Ci/mmol, 25 μM, Perkin Elmer) with PNK (NEB). After heat-inactivation of
535 PNK for 2 min at 95° C, sense and anti-sense (anti-sense sgRNA flanking sequences are:
536 GGCCGCATAATGAATTCCCCAATACCC-as sgRNA-GTTCATATAATGC) oligonucleotides
537 were annealed by letting cool down to room temperature and used in sgRNA-Cas9 cleavage assays.
538 For plasmid sgRNA-Cas9 cleavage assays, these annealed oligonucleotides were cloned into a
539 modified *pBS SK+* using a Xho I and Not I cut vector to assay sgRNA/Cas9 activity.
540 For sgRNA/Cas9 cleavage assays, DNA/sgRNA/Cas9 ratios of 1/10/10 were used in a 10 μl
541 reaction using the buffer supplied (NEB) and DEPC-treated water (Haussmann et al. 2019).
542 Typically Cas9 (100 nM final) was incubated with sgRNA (100 nM) for 10 min at 25° C before
543 adding oligonucleotides (10 nM final) or plasmid DNA (10 nM, corresponds to ~25 ng/μl final
544 concentration of a 3 kb plasmid). Plasmids were linearized after Cas9 digestion by first heat
545 inactivating Cas9 for 2 min at 95ºC, and then adding 10 μl of a restriction enzyme (5 U) in NEB
546 buffer 3. Adding a restriction enzyme together with Cas9 inhibited DNA scission by Cas9.
547 Cleavage of oligonucleotides was analysed on 8 % denaturing polyacrylamide gels and plasmid
548 DNA was analysed on ethidium bromide stained agarose gels.
549 RNA secondary structure was analyzed with RNAfold at http://rna.tbi.univie.ac.at (Gruber et al.
550 2008) using the following tracrRNA sequence:

551 GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAA
552 AAGUGGCACCGAGUCGGUGCUUUUUU.
553

554 **Criteria for optimal sgRNA design and Cas9 structural analysis**

555 Criteria to select sgRNAs that maintain the structure required for efficient Cas9 DNA scission were
556 implemented in the server accessible at https://platinum-crispr.bham.ac.uk/predict.pl and are as
557 follows. The first nucleotide of the sgRNA needs to be a G for efficient transcription initiation by
558 RNA Pol III (Paule and White 2000). For T7 mediated in vitro transcription, three Gs need to be
559 added to sgRNAs of 23 nucleotides. Disruption of the tetraloop or loop 2 and 3 structures, and the
560 sequence U A/G G C/U A/G of nucleotides 16-20, which will result in a tetraloop bulge mimic,
561 were classified as low efficiency sgRNAs as these parts are recognized by Cas9 (Nishimasu et al.
562 2014). Similarly, a hairpin loop in the gRNA consisting of 4 or more base pairing nucleotides, or
563 base pairing of nucleotides 17-20 of the gRNA (U/C U/C U U), or base-pairing of eight nucleotides
564 within the seed region with looped-out nucleotides spaced by three base-pairing nucleotides (N11-
565 N20), or a GC content below 15% (1 or 2 nts) or above 50% (11 or more nts) were also considered
566 low efficiency. Medium efficiency was assigned for gRNAs with a GC content of 15-25 % (3-5 nts)
567 or 40-50% (8-10 nts), or a low CG content in the seed region (less than 5 nts in nucleotides 11-20).
568 In addition, two U's at position 19/20 of the gRNA reduce efficiency because of premature
569 transcription termination (Gao et al. 2018; Graf et al. 2019). Further, we assigned a medium impact
570 if both of the two nucleotides A_{51} A_{52} and G_{62} U_{63} in the three way junction of loop 1 were base
571 paired or seven nucleotides within the seed region (N11-N20) base-paired with looped-out
572 nucleotides spaced by three base-pairing nucleotides. Thus, we deemed an sgRNA optimal to allow

573 Cas9 to cleave DNA with high efficiency, if the GC content is 30-35 % (6-7 nts) and none of the
574 above criteria applied.
575 Cas9/sgRNA structural complex analysis was done using Chimera as described (Dix et al. 2022).
576

577 **RNA extraction, RT-PCR and PCR on genomic DNA**

578 Total RNA was extracted using Tri-reagent (SIGMA) and reverse transcribed with Superscript II
579 (Invitrogen) according to the manufacturer's instructions using an Oligo(dT) primer. PCR was done
580 for 40 cycles with 1 μl of cDNA, with 1 μg of genomic DNA or from a single fly after freezing and
581 drying in 200 μl of isopropanol. Primers to detect the sgRNA/Cas9 induced deletion in *Ythdc1* were
582 YT F1 (GCCGCTGTGACGCAGAATTTGTGTG) and YT R1
583 (GGCCGTGCATGTTGCGCATGTAGTCC), and in *Ythdf* were 64F1
584 (GCCGAGAAAGTGCACAAGGATACGGAG) and 64R1
585 (CAAGGAATGGCTGAAGCAGACTCCTTG). Primers to amplify parts of the body of the RNA
586 also flanking an intron were for *Ythdc1* YT F2 (CCACGCTGCCGCAGAACGACGCCAATC) and
587 YT R2 (GCGGCAGATCCAGTCAAGCTCGATGAC), and in *Ythdf* were 64F2
588 (GAGCTGCCTGTCGATTCCCAACTCGTG) and 64R2
589 (CCGCCCTCTTCGTGTCGCTCCTTGAAG). Primers to amplify parts of the *ewg* gene have been
590 described elsewhere (Koushika et al. 1999).
591

592 **Cloning of sgRNAs into *pUC 3GLA U6.1 BbsI***

593 To clone two sgRNAs expressed by U6 promoters, the "tracrRNA U6.3 promoter" fragment was
594 amplified with left (AAGATATCCGGGTGAACTTC**G**N_{19}GTTTTAGAGCTAGAAATAGC) and
595 right (GCTATTTCTAGCTCTAAAA**C**N_{19}CGACGTTAAATTGAAAATAGG) sgRNA primers

596 from pUC 3GLA U6.1/3 sgRNA using Pwo polymerase (Roche) with initial 30 sec denaturation at
597 94ºC followed by two cycles 94ºC/30 sec, 49ºC/40 sec, 72ºC/45 sec, then two cycles 94ºC/30 sec,
598 51ºC/40 sec, 72ºC/45 sec and 22 cycles two cycles 94ºC/30 sec, 56ºC/40 sec, 72ºC/45 sec.
599 Transcription from the U6 promoter initiates with a G (bold, underlined). Although this G does not
600 need to be present in the targeting sequence, it needs be included for folding of the sgRNA. The
601 sequences for the sgRNAs in *Ythdc1* were gACAGGTATTCCCAAACTCAC and
602 GACATGTAGCGTTCCCATGA, and for *Ythdf* were GTCCTGAAATACGAGCACAA and
603 gATAACGAACATGTGGGATCT. The pUC 3GLA U6.1 BbsI vector was cut by BbsI and the
604 "sgRNA1 tracrRNA U6.3 promoter sgRNA2" fragment was cloned by Gibson assembly according
605 to the manufacturer's instructions (NEB). For sequencing, primer U6.1 Fseq
606 (GCGCGTACGTCCTTCGCATCCTTATG) was used. The sequences for *pUC 3GLA HAi Dscam*
607 *3-5, pUC 3GLA U6.1 BbsI* and the *pUC 3GLA U6.1/3 Ythdf sgRNA* have been deposited
608 (MK908409, MK908408 and MK908407).
609

610 ***Drosophila* genetics and phiC31 integrase-mediated transgenesis**

611 All *Drosophila melanogaster* strains were reared at 25ºC and 40%–60% humidity on standard
612 cornmeal-agar food in 12:12 h light:dark cycle as described (Haussmann et al. 2013). *CantonS* was
613 used as a wild type control. For the *Ythdc1*, the GFP marked *Mi{MIC}YT521-B*^{MI02006} transposon
614 insert and the *w+* marked *Df(3L)Exel6094* deficiency were used, and for *Ythdf*, the RFP marked
615 *PBac{SAstopDsRed}*^{LL04081} insert and the *w+* marked deficiency *Df(3R)ED6220* were used. For
616 phiC31 mediated transformation, constructs were injected into *y*^{1} *w* M{vas-int.Dm}ZH-2A;*
617 *PBac{y+-attP-3B}VK00013* with the landing site inserted at 76A as previously
618 described(Haussmann et al. 2013). Prior to insertion of GFP marked constructs, the GFP and RFP

619 markers had been removed from the *y1 w* M{vas-int.Dm}ZH-2A* landing site by Cre mediated
620 recombination (Bischof et al. 2007; Zaharieva et al. 2015).
621
622 **Implementation of PlatinumCRISPr**
623 PlatinumCRISPr is implemented as a Perl script based web-server iteratively evaluating the rule set
624 described in the main text. A guide is classified as "compromised" if any of the rules is violated.
625 For analysis of an sgRNA consisting the target complementary sequence (spacer) and the constant
626 crispr RNA (crRNA) fused to the tracrRNA through an artificial loop is used whereby the first
627 nucleotide of the 20 nt spacer sequence is a G, because a G is needed for transcription (Jinek et al.
628 2012; Cong et al. 2013). Folding of the sgRNA is computed using RNAFold (Version 2.4.17) and
629 further processed using bpRNA for subsequent interpretation of the dot-bracket code describing the
630 secondary structure by a custom-made Perl script. Notably, the sequence position is calculated from
631 the 3'end of the tracerRNA to allow for variable length of the spacer (between 18 and 23 nt) for
632 custom applications using synthesized RNA consisting of the spacer fused to the crRNA and
633 hybridized to the tracrRNA.
634 PlatinumCRISPr classifies guides by a binary outcome and typically reports around 70% of
635 sgRNAs as "compromised". Accordingly, only the top 30% were analysed for the distribution of
636 their reported cleavage efficiency in a given data set for each scoring application (**Supplemental**
637 **Fig S8).** Statistical significance was calculated using a one-sided Wilcoxon signed-rank test. We are
638 grateful to M. Haeussler for providing published guide sequencing and cleavage efficiency scores
639 (Haeussler et al. 2016).
640 CrisperON and DeepSpCas9 guide sequencing and cleavage efficiency scores were calculated using
641 published web-interfaces (Wong et al. 2015; Xiang et al. 2021).

642 For CRISpick analysis, we downloaded pre-calculated on-target score data (options: KO screen,
643 CRISPR mechanism SpyCas9, Hsu TracerRNA and On Target Scoring Rule Set was RS3seq-
644 Chen2013+RS3target) which are available for Mouse, Fly, and Human (Sanson et al. 2018).
645 DepMap inferred guide efficiency data has been downloaded from DepMap (DepMap 2024). For
646 the analysis of sgRNA efficiency in different cell lines (HeLa, HCT 116 and GBM), genes essential
647 in all cells were selected and the calculated cleavage efficiencies were compared by Spearman's
648 correlation (Hart et al. 2015).
649 For the analysis of off-target effects, guides from the TTISS data-set with their number of genome
650 wide off-target sites were used (Schmid-Burgk et al. 2020; Chen et al. 2023). For each guide, the
651 number of expected off-targets was computed using Cas-Offinder3 (selecting targets with no bulge,
652 and a minimum of one and a maximum five mismatches) (Hwang et al. 2021). The ratio of
653 measured to expected numbers of off-target sites was then compared between guides predicted to
654 cut by PlatinumCRISPr (Wilcoxon-Rank-sum test) and correlated (Spearman's correlation) to
655 DeepSpCas9 and CHOPCHOP (Azimuth 2.0 scores) using the crisprScore R-package (Hoberecht et
656 al. 2022).
657

658 **Data access**

659 All data used for sgRNA cleavage efficiency analysis to analyze data and the source code for
660 PlatinumCRISPr have been deposited in GitHub (https://github.com/rolandA1234/PlatinumCRISPr)
661 and are also provided as Supplemental Code.tar.gz file.
662 Sequences for plasmid vectors generated in this study have been submitted to NCBI GenBank
663 (https://www.ncbi.nlm.nih.gov/genbank/about/) with the following accession numbers: MK908409

664 (*pUC 3GLA HAi Dscam 3-5*), MK908408 (*pUC 3GLA U6.1 BbsI*) and MK908407 (*pUC 3GLA*
665 *U6.1/3 YTHDF53 sgRNA*).
666

667 **Competing interest statement**

668 The authors declare no competing interests
669

670 **Acknowledgments**

671 We thank Bloomington and Kyoto stock centers for fly lines, A. Handler for the PiggyBac clone,
672 BacPac for Bac clones and plasmids, the University of Cambridge Department of Genetics Fly
673 Facility for embryo injections for generating transgenic lines, M. Haeusler for datasets and pre-
674 calculated scores, Sung-Eun Yoon from the Korea Drosophila Resource Centre and Ying Tan from
675 Genetivision Corporation for sgRNA efficiency data, Yuan Tian for help with art work and Pawel
676 Grzechnik for comments on the manuscript. For this work we acknowledge funding from the
677 Biotechnology and Biological Science Research Council, the Leverhulme Trust, the MRC DTP to
678 DWJM and a Genetics Society summer studentship to AIH.
679 *Author contributions:* IUH, TCD and MS performed biochemistry and genetic experiments, VD
680 constructed vectors and AH performed genetic experiments. RA performed bioinformatics analysis,
681 programming and server implementation. DWJM analysed data. IUH, RA and MS conceived the
682 project and wrote the manuscript with help from all authors. .
683

684 **References**

685 Abadi S, Yan WX, Amar D, Mayrose I. 2017. A machine learning approach for predicting
686 CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS*
687 *Comput Biol* **13**: e1005807.
688 Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA
689 recognition by the Cas9 endonuclease. *Nature* **513**: 569-573.
690 Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M. 2021. New Twists in Detecting mRNA
691 Modification Dynamics. *Trends Biotechnol* **39**: 72-89.
692 Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C,
693 Newby GA, Raguram A et al. 2019. Search-and-replace genome editing without double-
694 strand breaks or donor DNA. *Nature* **576**: 149-157.
695 Arimbasseri AG, Maraia RJ. 2015. Mechanism of Transcription Termination by RNA Polymerase
696 III Utilizes a Non-template Strand Sequence-Specific Signal Element. *Mol Cell* **58**: 1124-
697 1132.
698 Balacco DL, Soller M. 2019. The m(6)A Writer: Rise of a Machine for Growing Tasks.
699 *Biochemistry* **58**: 363-378.
700 Bawankar P, Lence T, Paolantoni C, Haussmann IU, Kazlauskiene M, Jacob D, Heidelberger JB,
701 Richter FM, Nallasivan MP, Morin V et al. 2021. Hakai is required for stabilization of core
702 components of the m(6)A mRNA methylation machinery. *Nat Commun* **12**: 3778.
703 Bischof J, Maeda RK, Hediger M, Karch F, Basler K. 2007. An optimized transgenesis system for
704 Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**:
705 3312-3317.
706 Boivin V, Deschamps-Francoeur G, Scott MS. 2018. Protein coding genes as hosts for noncoding
707 RNA expression. *Semin Cell Dev Biol* **75**: 3-12.

708 Bosch JA, Birchak G, Perrimon N. 2021. Precise genome engineering in Drosophila using prime
709 editing. *Proc Natl Acad Sci U S A* **118**.
710 Chari R, Mali P, Moosburner M, Church GM. 2015. Unraveling CRISPR-Cas9 genome engineering
711 parameters via a library-on-library approach. *Nat Methods* **12**: 823-826.
712 Chen Q, Chuai G, Zhang H, Tang J, Duan L, Guan H, Li W, Li W, Wen J, Zuo E et al. 2023.
713 Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction
714 fingerprints. *Nat Commun* **14:** 7521.
715 Chuai G, Ma H, Yan J, Chen M, Hong N, Xue D, Zhou C, Zhu C, Chen K, Duan B et al. 2018.
716 DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* **19**: 80.
717 Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al.
718 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.
719 DepMap Broad Institute. 2024. DepMap 24Q2 Public. Figshare Dataset. https://depmap.org/portal.
720 https://doiorg/1025452/figshareplus25880521v1.
721 Deveson IW, Hardwick SA, Mercer TR, Mattick JS. 2017. The Dimensions, Dynamics, and
722 Relevance of the Mammalian Noncoding Transcriptome. *Trends Genet* **33**: 464-478.
723 Dezi V, Ivanov C, Haussmann IU, Soller M. 2016. Nucleotide modifications in messenger RNA
724 and their role in development and disease. *Biochem Soc Trans* **44**: 1385-1393.
725 Dix TC 2022. Characterization of ELAV multimerization in neuron development and function. PhD
726 Thesis, University of Birmingham.
727 Dix TC, Haussmann IU, Brivio S, Nallasivan MP, HadzHiev Y, Müller F, Müller B, Pettitt J, Soller
728 M. 2022. CMTr mediated 2'-O-ribose methylation status of cap-adjacent nucleotides across
729 animals. *RNA (New York, NY)* **28**: 1377-1390.

730 Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z,
731 Wilen C, Orchard R et al. 2016. Optimized sgRNA design to maximize activity and
732 minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**: 184-191.
733 Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL,
734 Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-
735 mediated gene inactivation. *Nat Biotechnol* **32**: 1262-1267.
736 Doudna JA. 2020. The promise and challenge of therapeutic genome editing. *Nature* **578**: 229-236.
737 El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Gunther S, Fukuda N, Kikhi K, Boezio GLM,
738 Takacs CM, Lai SL et al. 2019. Genetic compensation triggered by mutant mRNA
739 degradation. *Nature* **568**: 193-197.
740 Farboud B, Meyer BJ. 2015. Dramatic enhancement of genome editing by CRISPR/Cas9 through
741 improved guide RNA design. *Genetics* **199**: 959-971.
742 Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency
743 off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*
744 **31:** 822-826.
745 Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, Montague TG, Zimmerman S,
746 Richter C, Schier AF. 2014. Efficient mutagenesis by Cas9 protein-mediated
747 oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**:
748 e98186.
749 Gandhi S, Haeussler M, Razy-Krajka F, Christiaen L, Stolfi A. 2017. Evaluation and rational design
750 of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona. *Developmental*
751 *biology* **425**: 8-20.

752 Gao Z, Herrera-Carrillo E, Berkhout B. 2018. Delineation of the Exact Transcription Termination
753 Signal for Type 3 Polymerase III. *Molecular therapy Nucleic acids* **10**: 36-44.
754 Garcia-Doval C, Jinek M. 2017. Molecular architectures and mechanisms of Class 2 CRISPR-
755 associated nucleases. *Curr Opin Struct Biol* **47**: 157-166.
756 Geng K, Merino LG, Wedemann L, Martens A, Sobota M, Sanchez YP, Søndergaard JN, White RJ,
757 Kutter C. 2022. Target-enriched nanopore sequencing and de novo assembly reveals co-
758 occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in
759 human cells. *Genome research* **32**: 1876-1891.
760 Graf R, Li X, Chu VT, Rajewsky K. 2019. sgRNA Sequence Motifs Blocking Efficient
761 CRISPR/Cas9-Mediated Gene Editing. *Cell Rep* **26**: 1098-1103 e1093.
762 Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite.
763 *Nucleic Acids Res* **36**: W70-74.
764 Hachet O, Ephrussi A. 2004. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic
765 localization. *Nature* **428**: 959-963.
766 Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S,
767 Shkumatava A, Teboul L, Kent J et al. 2016. Evaluation of off-target and on-target scoring
768 algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**:
769 148.
770 Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann
771 M, Fradet-Turcotte A, Sun S et al. 2015. High-Resolution CRISPR Screens Reveal Fitness
772 Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**: 1515-1526.

773 Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, Soller M. 2016.
774 m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex
775 determination. *Nature* **540**: 301-304.
776 Haussmann IU, Hemani Y, Wijesekera T, Dauwalder B, Soller M. 2013. Multiple pathways
777 mediate the sex-peptide-regulated switch in female Drosophila reproductive behaviours.
778 *Proc Biol Sci* **280**: 20131938.
779 Haussmann IU, Li M, Soller M. 2011. ELAV-mediated 3'-end processing of ewg transcripts is
780 evolutionarily conserved despite sequence degeneration of the ELAV-binding site. *Genetics*
781 **189**: 97-107.
782 Haussmann IU, Ustaoglu P, Brauer U, Hemani Y, Dix TC, Soller M. 2019. Plasmid-based gap-
783 repair recombineered transgenes reveal a central role for introns in mutually exclusive
784 alternative splicing in Down Syndrome Cell Adhesion Molecule exon 4. *Nucleic Acids Res*
785 **47**: 1389-1403.
786 Haussmann IU, Wu Y, Nallasivan MP, Archer N, Bodi Z, Hebenstreit D, Waddell S, Fray R, Soller
787 M. 2022. CMTr cap-adjacent 2'-O-ribose mRNA methyltransferases are required for reward
788 learning and mRNA localization to synapses. *Nat Commun* **13**: 1209.
789 Hemani Y, Soller M. 2012. Mechanisms of Drosophila Dscam mutually exclusive splicing
790 regulation. *Biochem Soc Trans* **40**: 804-809.
791 Hille F, Richter H, Wong SP, Bratovic M, Ressel S, Charpentier E. 2018. The Biology of CRISPR-
792 Cas: Backward and Forward. *Cell* **172**: 1239-1259.
793 Hoberecht L, Perampalam P, Lun A, Fortin JP. 2022. A comprehensive Bioconductor ecosystem for
794 the design of CRISPR guide RNAs across nucleases and technologies. *Nat Commun* **13:**
795 6568.

796 Housden BE, Valvezan AJ, Kelley C, Sopko R, Hu Y, Roesel C, Lin S, Buckner M, Tao R,
797 Yilmazel B et al. 2015. Identification of potential drug targets for tuberous sclerosis
798 complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Science*
799 *signaling* **8**: rs9.
800 Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X,
801 Shalem O et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat*
802 *Biotechnol* **31**: 827-832.
803 Hwang GH, Kim JS, Bae S. 2021. Web-Based CRISPR Toolkits: Cas-OFFinder, Cas-Designer, and
804 Cas-Analyzer. *Methods Mol Biol* **2162:** 23-33.
805 Jensen KT, Floe L, Petersen TS, Huang J, Xu F, Bolund L, Luo Y, Lin L. 2017. Chromatin
806 accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing
807 efficiency. *FEBS Lett* **591**: 1892-1901.
808 Jiang F, Doudna JA. 2017. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46**: 505-
809 529.
810 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-
811 RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.
812 Kim HK, Kim Y, Lee S, Min S, Bae JY, Choi JW, Park J, Jung D, Yoon S, Kim HH. 2019. SpCas9
813 activity prediction by DeepSpCas9, a deep learning-based model with high generalization
814 performance. *Science advances* **5**: eaax9249.
815 Koushika SP, Soller M, DeSimone SM, Daub DM, White K. 1999. Differential and inefficient
816 splicing of a broadly expressed Drosophila erect wing transcript results in tissue-specific
817 enrichment of the vital EWG protein isoform. *Mol Cell Biol* **19**: 3998-4007.

818 Labuhn M, Adams FF, Ng M, Knoess S, Schambach A, Charpentier EM, Schwarzer A, Mateo JL,
819 Klusmann JH, Heckl D. 2018. Refined sgRNA efficacy prediction improves large- and
820 small-scale CRISPR-Cas9 applications. *Nucleic Acids Res* **46**: 1375-1385.
821 Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. 2016. CHOPCHOP v2: a web tool for
822 the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**: W272-276.
823 Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. 2019. CHOPCHOP
824 v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**:
825 W171-w174.
826 Lee MH, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, Liu CS, Park C, Kennedy G,
827 Long T et al. 2018. Somatic APP gene recombination in Alzheimer's disease and normal
828 neurons. *Nature* **563**: 639-645.
829 Liu X, Homma A, Sayadi J, Yang S, Ohashi J, Takumi T. 2016. Sequence features associated with
830 the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep* **6**: 19675.
831 Ma Z, Zhu P, Shi H, Guo L, Zhang Q, Chen Y, Chen S, Zhang Z, Peng J, Chen J. 2019. PTC-
832 bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS
833 components. *Nature* **568**: 259-263.
834 Michlits G, Jude J, Hinterndorfer M, de Almeida M, Vainorius G, Hubmann M, Neumann T,
835 Schleiffer A, Burkard TR, Fellner M et al. 2020. Multilayered VBC score predicts sgRNAs
836 that efficiently generate loss-of-function alleles. *Nat Methods* **17**: 708-716.
837 Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ.
838 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo.
839 *Nat Methods* **12**: 982-988.

840 Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki
841 O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**:
842 935-949.
843 Paule MR, White RJ. 2000. Survey and summary: transcription by RNA polymerases I and III.
844 *Nucleic Acids Res* **28**: 1283-1298.
845 Poh HX, Mirza AH, Pickering BF, Jaffrey SR. 2022. Alternative splicing of METTL3 explains
846 apparently METTL3-independent m6A modifications in mRNA. *PLoS biology* **20**:
847 e3001683.
848 Port F, Chen HM, Lee T, Bullock SL. 2014. Optimized CRISPR/Cas tools for efficient germline
849 and somatic genome engineering in Drosophila. *Proc Natl Acad Sci U S A* **111**: E2967-2976.
850 Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S, Liu LP, Yang Z, Mao D, Sun L et al. 2013.
851 Optimized gene editing technology for Drosophila melanogaster using germ line-specific
852 Cas9. *Proc Natl Acad Sci U S A* **110**: 19012-19017.
853 Ren X, Yang Z, Xu J, Sun J, Mao D, Hu Y, Yang SJ, Qiao HH, Wang X, Hu Q et al. 2014.
854 Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA
855 parameters in Drosophila. *Cell Rep* **9**: 1151-1162.
856 Riesenberg S, Kanis P, Macak D, Wollny D, Düsterhöft D, Kowalewski J, Helmbrecht N, Maricic
857 T, Pääbo S. 2023. Efficient high-precision homology-directed repair-dependent genome
858 editing by HDRobust. *Nat Methods* **20**: 1388-1399.
859 Roignant JY, Soller M. 2017. m6A in mRNA: An Ancient Mechanism for Fine-Tuning Gene
860 Expression. *Trends Genet* **33**: 380-390.

861 Roth A, Vleurinck C, Netschitailo O, Bauer V, Otte M, Kaftanoglu O, Page RE, Beye M. 2019. A
862 genetic switch for worker nutrition-mediated traits in honeybees. *PLoS biology* **17**:
863 e3000171.
864 Samson ML. 2008. Rapid functional diversification in the structurally conserved ELAV family of
865 neuronal RNA binding proteins. *BMC Genomics* **9**: 392.
866 Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale
867 A, Root DE, Piccioni F et al. 2018. Optimized libraries for CRISPR-Cas9 genetic screens
868 with multiple modalities. *Nat Commun* **9**: 5416.
869 Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F. 2020. Highly Parallel
870 Profiling of Cas9 Variant Specificity. *Mol Cell* **78**: 794-800.e798.
871 Singh R, Kuscu C, Quinlan A, Qi Y, Adli M. 2015. Cas9-chromatin binding information enables
872 more accurate CRISPR off-target prediction. *Nucleic Acids Res* **43**: e118.
873 Sledzinski P, Nowaczyk M, Olejniczak M. 2020. Computational Tools and Resources Supporting
874 CRISPR-Cas Experiments. *Cells* **9**.
875 Soller M. 2006. Pre-messenger RNA processing and its regulation: a genomic perspective. *Cell Mol*
876 *Life Sci* **63**: 796-819.
877 Soller M, Haussmann IU, Hollmann M, Choffat Y, White K, Kubli E, Schäfer MA. 2006. Sex-
878 peptide-regulated female sexual behavior requires a subset of ascending ventral nerve cord
879 neurons. *Current biology : CB* **16**: 1771-1782.
880 Solomon DA, Stepto A, Au WH, Adachi Y, Diaper DC, Hall R, Rekhi A, Boudi A, Tziortzouda P,
881 Lee YB et al. 2018. A feedback loop between dipeptide-repeat protein, TDP-43 and
882 karyopherin-α mediates C9orf72-related neurodegeneration. *Brain : a journal of neurology*
883 **141**: 2908-2924.

884 Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. 2016. Internal guide RNA
885 interactions interfere with Cas9-mediated cleavage. *Nat Commun* **7**: 11750.
886 Trivedi D, Cm V, Bisht K, Janardan V, Pandit A, Basak B, H S, Ramesh N, Raghu P. 2020. A
887 genome engineering resource to uncover principles of cellular organization and tissue
888 architecture by lipid signaling. *Elife* **9**.
889 Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J,
890 Mathews DH, Kim J et al. 2019. CRISPR-Cas9-based mutagenesis frequently provokes on-
891 target mRNA misregulation. *Nat Commun* **10**: 4056.
892 Ustaoglu P, Gill JK, Doubovetzky N, Haussmann IU, Dix TC, Arnold R, Devaud JM, Soller M.
893 2021. Dynamically expressed single ELAV/Hu orthologue elavl2 of bees is required for
894 learning and memory. *Communications biology* **4**: 1234.
895 Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li
896 M et al. 2015. High-throughput gene targeting and phenotyping in zebrafish using
897 CRISPR/Cas9. *Genome research* **25**: 1030-1042.
898 Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-
899 Cas9 system. *Science* **343**: 80-84.
900 Wierson WA, Welker JM, Almeida MP, Mann CM, Webster DA, Torrie ME, Weiss TJ, Kambakam
901 S, Vollbrecht MK, Lan M et al. 2020. Efficient targeted integration directed by short
902 homology in zebrafish and mammalian cells. *Elife* **9**.
903 Wong N, Liu W, Wang X. 2015. WU-CRISPR: characteristics of functional guide RNAs for the
904 CRISPR/Cas9 system. *Genome Biol* **16**: 218.

905 Xiang X, Corsi GI, Anthon C, Qu K, Pan X, Liang X, Han P, Dong Z, Liu L, Zhong J et al. 2021.
906 Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning.
907 *Nat Commun* **12**: 3238.
908 Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS et al. 2015.
909 Sequence determinants of improved CRISPR sgRNA design. *Genome research* **25**: 1147-
910 1157.
911 Zaharieva E, Haussmann IU, Brauer U, Soller M. 2015. Concentration and localization of co-
912 expressed ELAV/Hu proteins control specificity of mRNA processing. *Mol Cell Biol* **35**.
913 Zhang D, Hurst T, Duan D, Chen SJ. 2019. Unified energetics analysis unravels SpCas9 cleavage
914 activity for optimal gRNA design. *Proc Natl Acad Sci U S A* **116**: 8693-8698.
915

916 ## Figure legends

917 **Figure 1:** Sequence dependent in vitro cleavage of oligonucleotides and plasmid DNA by the
918 sgRNA/Cas9 complex.
919 (A) Sequences of sgRNAs with observed cleavage sites indicated by arrow heads. Small letter
920 guanosines used for in vitro transcription are not present in the target DNA sequence. The seed
921 sequence is indicated by a line at the bottom.
922 (B) Agarose gel showing Cas9 mediated cleavage of the 11.3 kb Dscam 3-5 plasmid for 24 h with
923 indicated sgRNAs. Plasmids were cut with either Acc65I (lanes 2, 3, 6 and 7) or BspEI (lanes 4, 5,
924 8 and 9) after Cas9 cleavage. The line at the bottom shows a map of the plasmid with restriction
925 sites indicated. Size markers are EcoRI/HinDIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb and 0.8
926 kb.

927 (C) Structure of the sgRNA scaffold from co-crystallization with Cas9 (Nishimasu et al., 2014).
928 Vertical or horizontal lines indicate Watson-Crick base-pairing, and dots or dashed lines indicate
929 non-Watson-Crick base-pairing. Nucleotides base-pairing in loop 1 are bold. Additional base-
930 pairing found in the tracrRNA-crRNA heterodimer is indicated in the extended scaffold (Jinek et
931 al., 2012).
932 (D) Agarose gel showing Cas9 mediated cleavage of the 11.3 kb Dscam 3-5 plasmid for 24 h with
933 indicated sgRNAs L7GC and R3G. Plasmids were cut with PstI and NotI. The star denotes
934 incomplete cleavage by NotI and the line at the bottom shows a map of the plasmid with restriction
935 sites indicated. Size markers are EcoRI/HinDIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb and 0.8
936 kb.
937 (E) Denaturing acrylamide gel showing Cas9 mediated cleavage of synthetic oligonucleotides with
938 indicated sgRNAs.
939 (F) Sequences of sgRNAs with variable length. Small letter guanosines used for in vitro
940 transcription are not present in the target DNA sequence.
941 (G) Denaturing acrylamide gel showing Cas9 mediated cleavage for 1 h of synthetic
942 oligonucleotides with indicated sgRNAs of variable length.
943 (H-K) Structure of sgRNAs. Nucleotides base-pairing in loop 1 are bold. Red lines in J and K
944 indicate potential base-pairing with nucleotides in loop 2. The red arrow in J indicates the sequence
945 complementarity leading to a bulge in the tetraloop. The red arrows in K indicate a duplication of
946 the bulge structure present in the tetraloop.
947
948 **Figure 2:** Secondary structure of sgRNAs affects Cas9 mediated cleavage efficiency.

949 (A) Sequences of sgRNAs. Small letter guanosines used for in vitro transcription are not present in
950 the target DNA sequence.
951 (B) Agarose gel showing Cas9 mediated cleavage after 6 h of 3 kb pBS SK+ test-plasmids
952 containing the target sequence with indicated sgRNAs. Plasmids were linearized with ScaI as in the
953 control after Cas9 heat inactivation. Size markers are EcoRI/HinDIII digested λ DNA of 20 kb, 3.6
954 kb, 1.9 kb and 0.8 kb.
955 (C) Denaturing acrylamide gel showing synthetic oligonucleotides before and after sgRNA/Cas9
956 mediated cleavage for 1h.
957 (D-G) Structure of sgRNAs. Nucleotides base-pairing in loop 1 are bold. Red lines in D indicate
958 potential base-pairing with nucleotides in loop 2. Green nucleotides indicate mutations compared to
959 sgRNA R13G.
960
961 **Figure 3:** Base-pairing of sgRNAs in the seed-region blocks Cas9 mediated cleavage of a test
962 plasmid.
963 (A, B) Structure of sgRNAs. Horizontal red lines in A and B indicate artificially introduced base-
964 pairing with the sgRNA scaffold, and vertical red lines in a indicate potential base-pairing with
965 nucleotides in loop 2. Nucleotides base-pairing in loop 1 are bold.
966 (C) Denaturing acrylamide gel showing synthetic oligonucleotides before and after sgRNA/Cas9
967 mediated cleavage for 1h. Note that Cas9 cleavage is heterogeneous.
968 (D) Agarose gel showing Cas9 mediated cleavage after 6 h of 3 kb pBS SK+ test-plasmids
969 containing the target sequence with indicated sgRNAs. Plasmids were linearized with ScaI as in the
970 control after Cas9 heat inactivation. Size markers are EcoRI/HinDIII digested λ DNA of 20 kb, 3.6
971 kb, 1.9 kb and 0.8 kb.

972

973 **Figure 4:** PlatinumCRISPr selects high efficiency sgRNAs and outperforms other sgRNA selection
974 tools for a *Drosophila* data set.
975 (A) Comparison of the in vivo efficiency of all sgRNAs from the Ren et al. (2014) data set with
976 PlatinumCRISPr selected sgRNAs.
977 B) Comparison of the sgRNA selection performance of PlatinumCRISPr with other sgRNA
978 selection tools.
979

980 **Figure 5:** Comparison of individual sgRNA selection criteria for performance with different data
981 sets. The different data sets with the number of sgRNAs tested are indicated on the left. The
982 different selection criteria are shown on top. Significant and enriched performances are indicated in
983 red and orange, respectively ($p<0.05$ and $p<0.25$). Criteria with numbers below 5% are indicated in
984 beige, and criteria already applied to the data set are shown in grey.
985

986 **Figure 6:** Combinations of two sgRNA selection tools select high efficiency cleaving sgRNAs for
987 several sgRNA efficiency screen data sets. Combinations of sgRNA selection tools in A are listed
988 according to overall cleavage efficiency of selected sgRNA significant ($p<0.05$) for both methods
989 (red) or one method (dark grey). Black indicates events with less than 5% of sgRNAs selected by at
990 least one method. Comparison of sgRNA selection by different sgRNA selection tools shown as
991 median of the cleavage efficiency for individual data sets (B). The distribution of cleavage
992 efficiencies for all sgRNAs is shown on the left (white box) and for PlatinumCRISPr-Wong Score
993 in red, Wond Score-Xu Score in purple and PlatinumCRISPr-Moreno-Mateos Score in blue.
994

995

996 **Figure 7:** Generation of gene deletions in *Drosophila* YTH protein genes using two sgRNAs/Cas9
997 and transposon markers.
998 (A) Schematic to the *Ythdc1* locus indicating transcripts (white boxes) and the ORF (black boxes)
999 below the chromosome. Primers used are indicated on top and below the transcripts. The w+
1000 marked transposon used for detecting deletions in the locus is indicated by a triangle and the
1001 deletion generated is indicated by a line.
1002 (B) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdc1*
1003 transposon excision lines using primers flanking the deletion. Presence of a PCR product indicates
1004 the expected gene deletion. DNA markers are indicated on the left.
1005 (C) Agarose gels showing RT-PCR products amplified from cDNA of control or *Ythdc1*
1006 transposon excision lines using internal primers flanking an intron. Ectopic insertion in the opposite
1007 orientation is indicate by an asterisk (lane 4) as in this instance the transcript is not spliced. DNA
1008 markers are indicated on the left.
1009 (D) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdc1*$^{\Delta 7}$ flies
1010 using internal primers and primers flanking the deletion. DNA markers are indicated on the left.
1011 (E) Agarose gels showing RT-PCR products amplified from cDNA of control or *Ythdc1*$^{\Delta 7}$ flies
1012 using internal primers in *Ythdf* and *ewg* genes. The PCR product of the *ewg* gene was used as
1013 loading control. DNA markers are indicated on the left.
1014 (F) Schematic to the *Ythdf* locus indicating transcripts (white boxes) and the ORF (black boxes)
1015 below the chromosome. Primers used are indicated on top and below the transcripts. The RFP
1016 marked transposon used for detecting deletions in the locus is indicated by a triangle and the
1017 deletion generated is indicated by a line.

1018 (G) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdf*
1019 transposon excision lines using primers flanking the deletion. Presence of a PCR product indicates
1020 the expected gene deletion. DNA markers are indicated on the left.
1021 (H) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdf*
1022 transposon excision lines using primers flanking an intron. DNA markers are indicated on the left.
1023 (I) Agarose gels showing PCR products amplified from genomic DNA of control or $Ythdf^{\Delta B1}$ flies
1024 using internal primers and primers flanking the deletion. DNA markers are indicated on the left.
1025 (J) Agarose gels showing RT-PCR products amplified from cDNA of control or $Ythdf^{\Delta B1}$ flies
1026 using internal primers in *Ythdf* and *ewg* genes. The PCR product of the *ewg* gene was used as
1027 loading control. DNA markers are indicated on the left.

Figure 1

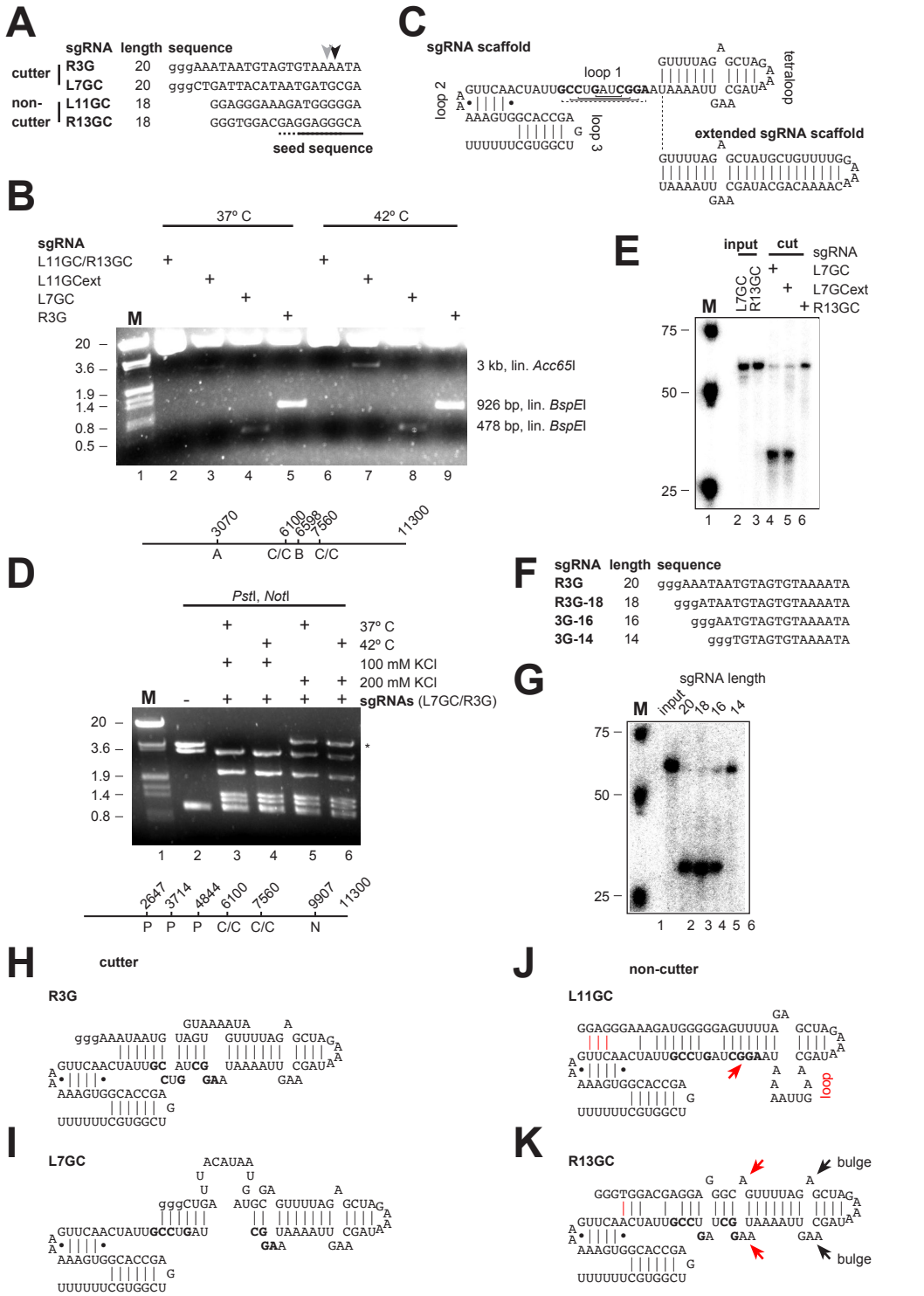
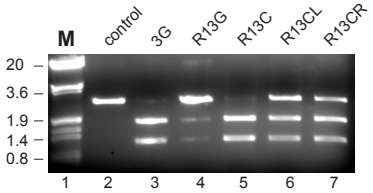

A
sgRNA length sequence
cutter R3G 20 gggAAATAATGTAGTGTAAAATA
L7GC 20 gggCTGATTACATAATGATGCGA
non-cutter L11GC 18 GGAGGGAAAGATGGGGGA
R13GC 18 GGGTGGACGAGGAGGGCA
seed sequence
B
37° C
42° C
sgRNA
L11GC/R13GC
L11GCext
L7GC
R3G
M
20
3.6
1.9
1.4
0.8
0.5
3 kb, lin. Acc65I
926 bp, lin. BspEI
478 bp, lin. BspEI
1 2 3 4 5 6 7 8 9
3070 6100 6398 7560 11300
A C/C B C/C
C
sgRNA scaffold
loop 1
loop 2
loop 3
tetraloop
extended sgRNA scaffold
D
PstI, NotI
37° C
42° C
100 mM KCl
200 mM KCl
sgRNAs (L7GC/R3G)
M
20
3.6
1.9
1.4
0.8
1 2 3 4 5 6
2647 3714 4844 6100 7560 9907 11300
P P P C/C C/C N
E
input cut
sgRNA
L7GC
L7GCext
R13GC
M
75
50
25
1 2 3 4 5 6
F
sgRNA length sequence
R3G 20 gggAAATAATGTAGTGTAAAATA
R3G-18 18 gggATAATGTAGTGTAAAATA
3G-16 16 gggAATGTAGTGTAAAATA
3G-14 14 gggTGTAGTGTAAAATA
G
sgRNA length
M input 20 18 16 14
75
50
25
1 2 3 4 5 6
H
cutter
R3G
I
L7GC
J
non-cutter
L11GC
loop
K
R13GC
bulge
bulge

Figure 2

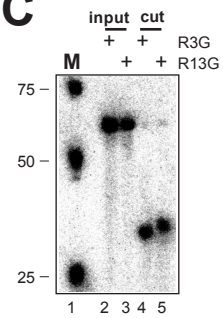
A

sgRNA	length	sequence
R3G	20	gggAAATAATGTAGTGTAAAATA
R13G	20	ggGAGGGAGGTGGTGTGAGAGG
R13C	20	ggGACCCACCTCCTCTCACACC
R13CL	20	ggGACCCACCTCCTCTGAGAGG
R13CR	20	ggGAGGGAGGTGGTGTCACACC

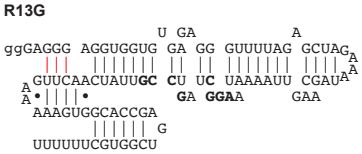
B



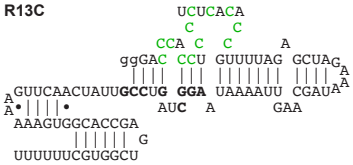
C



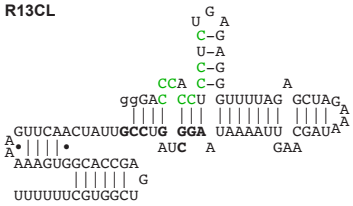
D



E



F



G

Figure 3

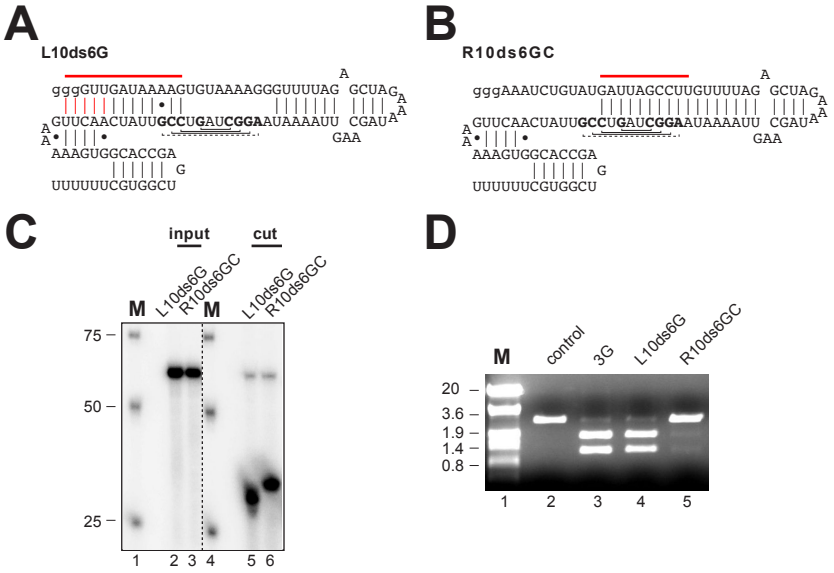

A
L10ds6G
B
R10ds6GC
C
input
cut
M
L10ds6G
R10ds6GC
75
50
25
1 2 3 4 5 6
D
M
control
3G
L10ds6G
R10ds6GC
20
3.6
1.9
1.4
0.8
1 2 3 4 5

Figure 4

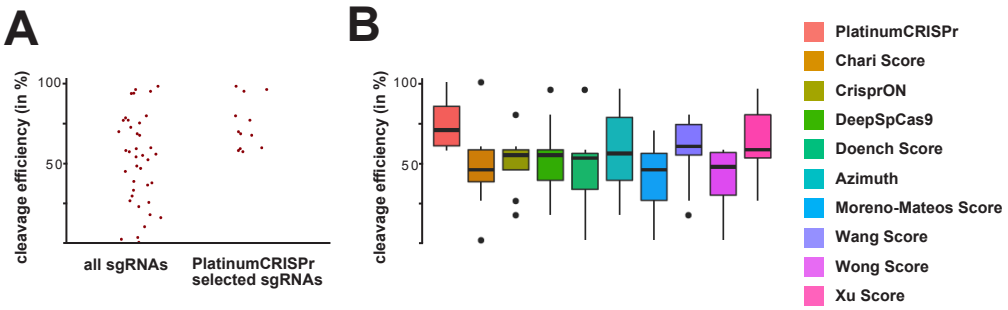

A
cleavage efficiency (in %)
100
50
all sgRNAs
PlatinumCRISPr selected sgRNAs
B
cleavage efficiency (in %)
100
50
PlatinumCRISPr
Chari Score
CrisprON
DeepSpCas9
Doench Score
Azimuth
Moreno-Mateos Score
Wang Score
Wong Score
Xu Score

Figure 5

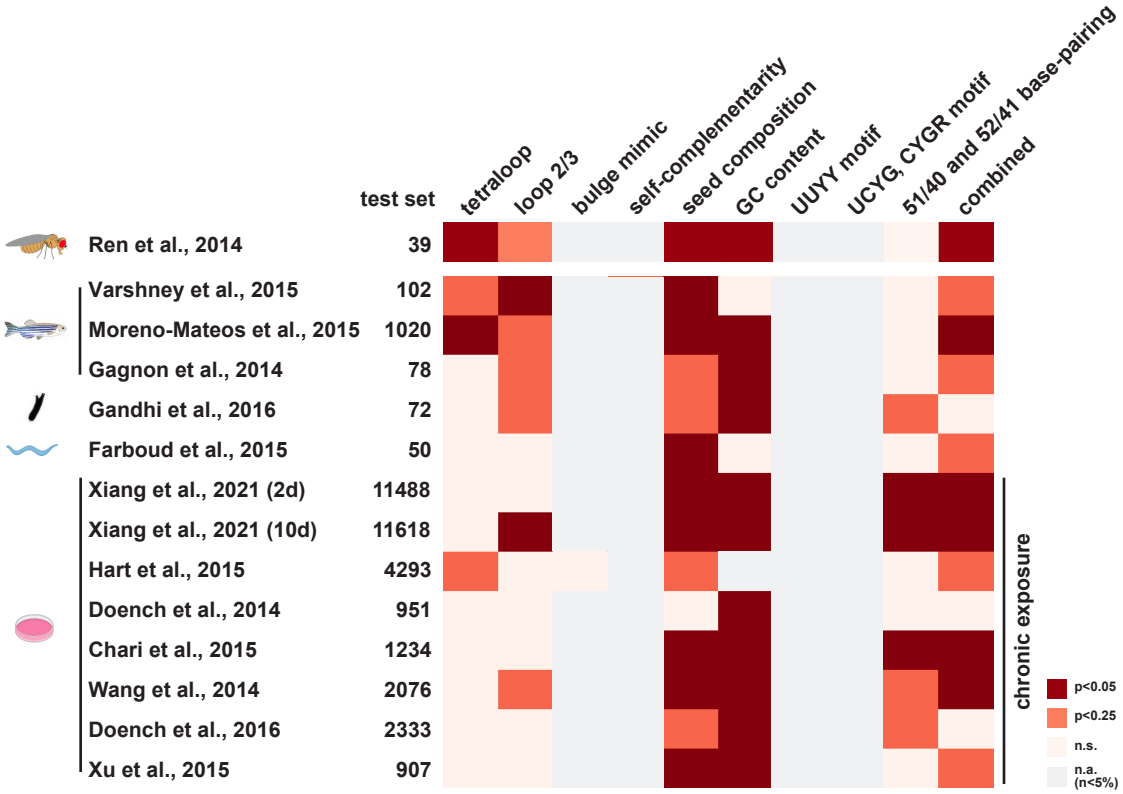

test set
tetraloop
loop 2/3
bulge mimic
self-complementarity
seed composition
GC content
UUYY motif
UCYG, CYGR motif
51/40 and 52/41 base-pairing
combined
Ren et al., 2014 39
Varshney et al., 2015 102
Moreno-Mateos et al., 2015 1020
Gagnon et al., 2014 78
Gandhi et al., 2016 72
Farboud et al., 2015 50
Xiang et al., 2021 (2d) 11488
Xiang et al., 2021 (10d) 11618
Hart et al., 2015 4293
Doench et al., 2014 951
Chari et al., 2015 1234
Wang et al., 2014 2076
Doench et al., 2016 2333
Xu et al., 2015 907
chronic exposure
p<0.05
p<0.25
n.s.
n.a. (n<5%)

Figure 6

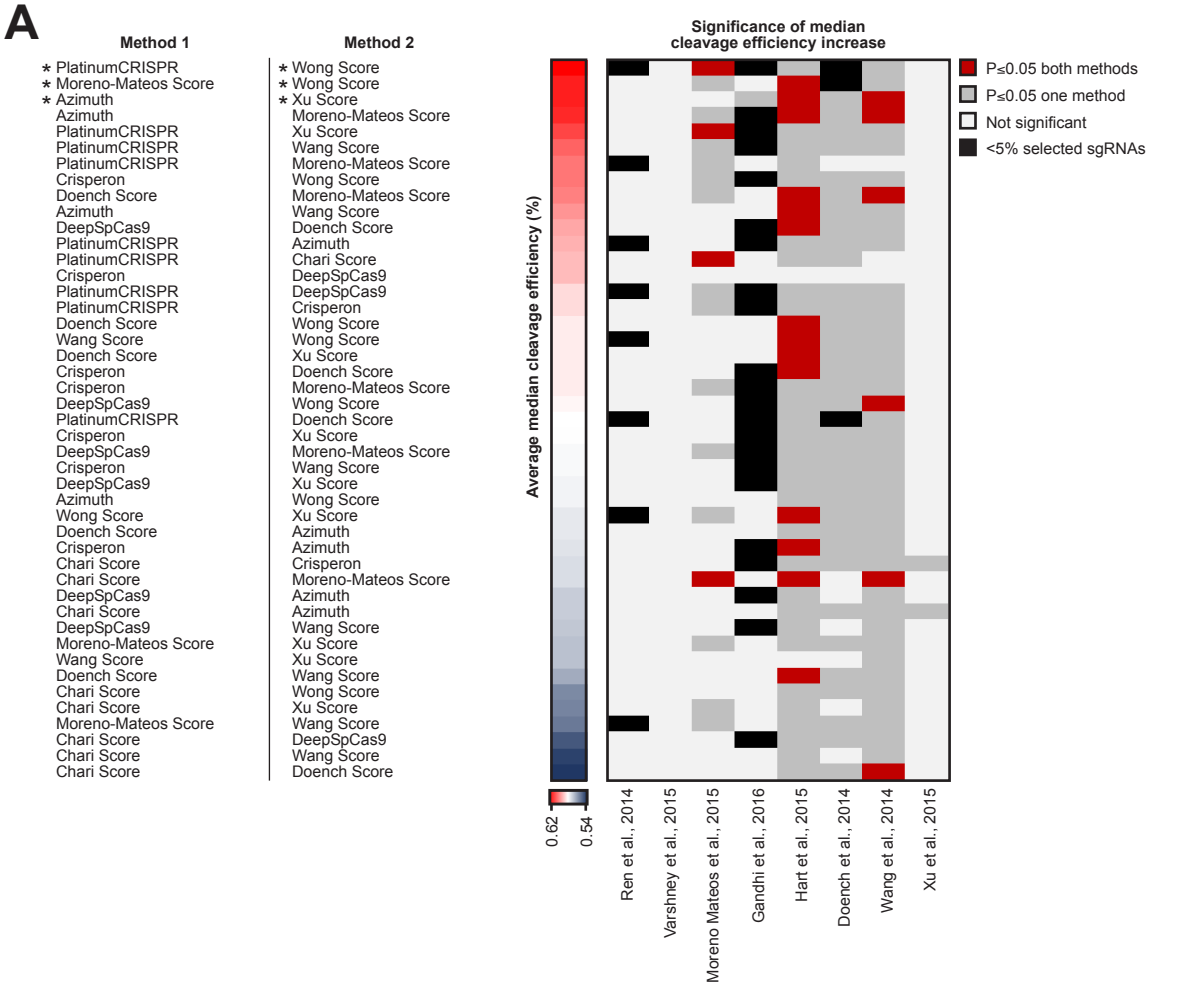
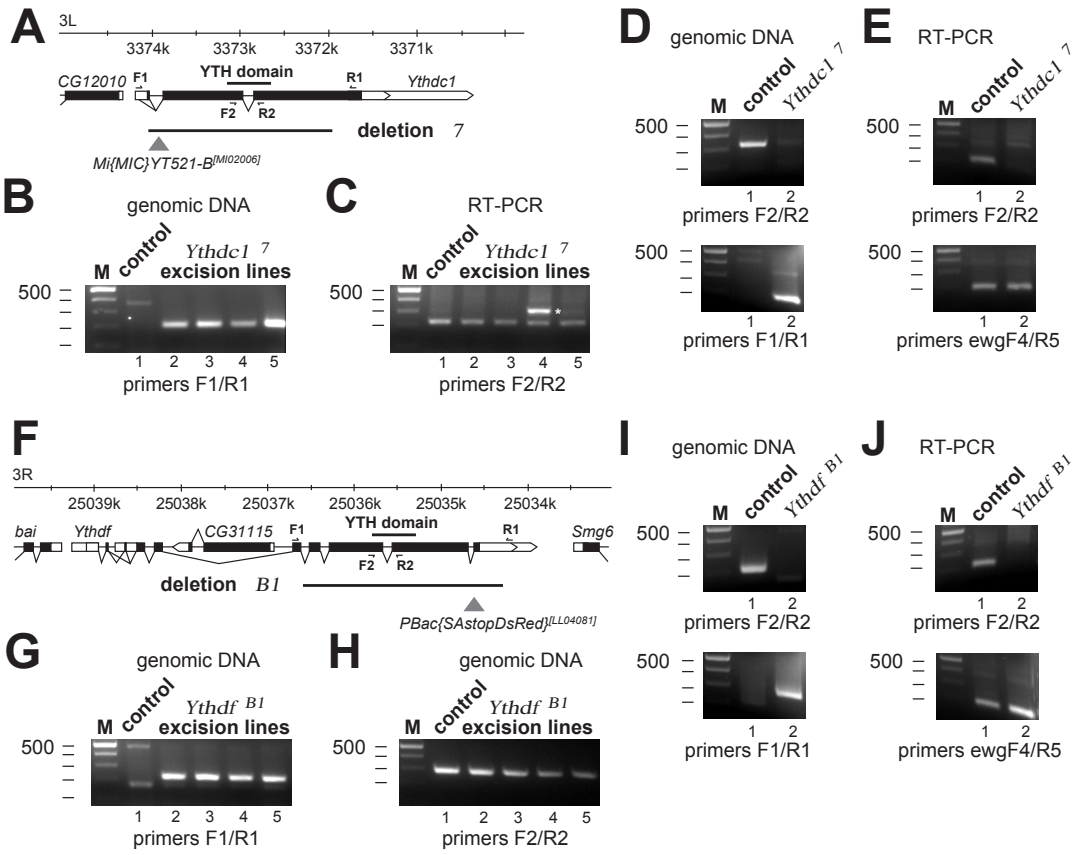

A
Method 1
Method 2
* PlatinumCRISPR
* Moreno-Mateos Score
* Azimuth
Azimuth
PlatinumCRISPR
PlatinumCRISPR
PlatinumCRISPR
Crisperon
Doench Score
Azimuth
DeepSpCas9
PlatinumCRISPR
PlatinumCRISPR
Crisperon
PlatinumCRISPR
PlatinumCRISPR
Doench Score
Wang Score
Doench Score
Crisperon
Crisperon
DeepSpCas9
PlatinumCRISPR
Crisperon
DeepSpCas9
Crisperon
DeepSpCas9
Azimuth
Wong Score
Doench Score
Crisperon
Chari Score
Chari Score
DeepSpCas9
Chari Score
DeepSpCas9
Moreno-Mateos Score
Wang Score
Doench Score
Chari Score
Chari Score
Moreno-Mateos Score
Chari Score
Chari Score
Chari Score
* Wong Score
* Wong Score
* Xu Score
Moreno-Mateos Score
Xu Score
Wang Score
Moreno-Mateos Score
Wong Score
Moreno-Mateos Score
Wang Score
Doench Score
Azimuth
Chari Score
DeepSpCas9
DeepSpCas9
Crisperon
Wong Score
Wong Score
Xu Score
Doench Score
Moreno-Mateos Score
Wong Score
Doench Score
Xu Score
Moreno-Mateos Score
Wang Score
Xu Score
Wong Score
Xu Score
Azimuth
Azimuth
Crisperon
Moreno-Mateos Score
Azimuth
Azimuth
Wang Score
Xu Score
Xu Score
Wang Score
Wong Score
Xu Score
Wang Score
DeepSpCas9
Wang Score
Doench Score
Significance of median cleavage efficiency increase
Average median cleavage efficiency (%)
0.62
0.54
Ren et al., 2014
Varshney et al., 2015
Moreno Mateos et al., 2015
Gandhi et al., 2016
Hart et al., 2015
Doench et al., 2014
Wang et al., 2014
Xu et al., 2015
P≤0.05 both methods
P≤0.05 one method
Not significant
<5% selected sgRNAs
B
Doench et al., 2014
Moreno Mateos et al., 2015
Xu et al., 2015
Wang et al., 2014
Gandhi et al., 2016
Hart et al., 2015
Ren et al., 2014
Varshney et al., 2015
cleavage efficiency (in %)
100
50
All sgRNAs
PlatinumCRISPr - Wong Score
Moreno-Mateos Score - Wong Score
Azimuth - Xu Score

Figure 7


A
3L
3374k
3373k
3372k
3371k
CG12010
F1
YTH domain
R1
Ythdc1
F2
R2
deletion 7
Mi{MIC}YT521-B[MI02006]
B
genomic DNA
M
control
Ythdc1 7
excision lines
500
1 2 3 4 5
primers F1/R1
C
RT-PCR
M
control
Ythdc1 7
excision lines
500
1 2 3 4 5
primers F2/R2
D
genomic DNA
M
control
Ythdc1 7
500
1 2
primers F2/R2
500
1 2
primers F1/R1
E
RT-PCR
M
control
Ythdc1 7
500
1 2
primers F2/R2
500
1 2
primers ewgF4/R5
F
3R
25039k
25038k
25037k
25036k
25035k
25034k
bai
Ythdf
CG31115
F1
YTH domain
R1
Smg6
F2
R2
deletion B1
PBac{SAstopDsRed}[LL04081]
G
genomic DNA
M
control
Ythdf B1
excision lines
500
1 2 3 4 5
primers F1/R1
H
genomic DNA
M
control
Ythdf B1
excision lines
500
1 2 3 4 5
primers F2/R2
I
genomic DNA
M
control
Ythdf B1
500
1 2
primers F2/R2
500
1 2
primers F1/R1
J
RT-PCR
M
control
Ythdf B1
500
1 2
primers F2/R2
500
1 2
primers ewgF4/R5

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

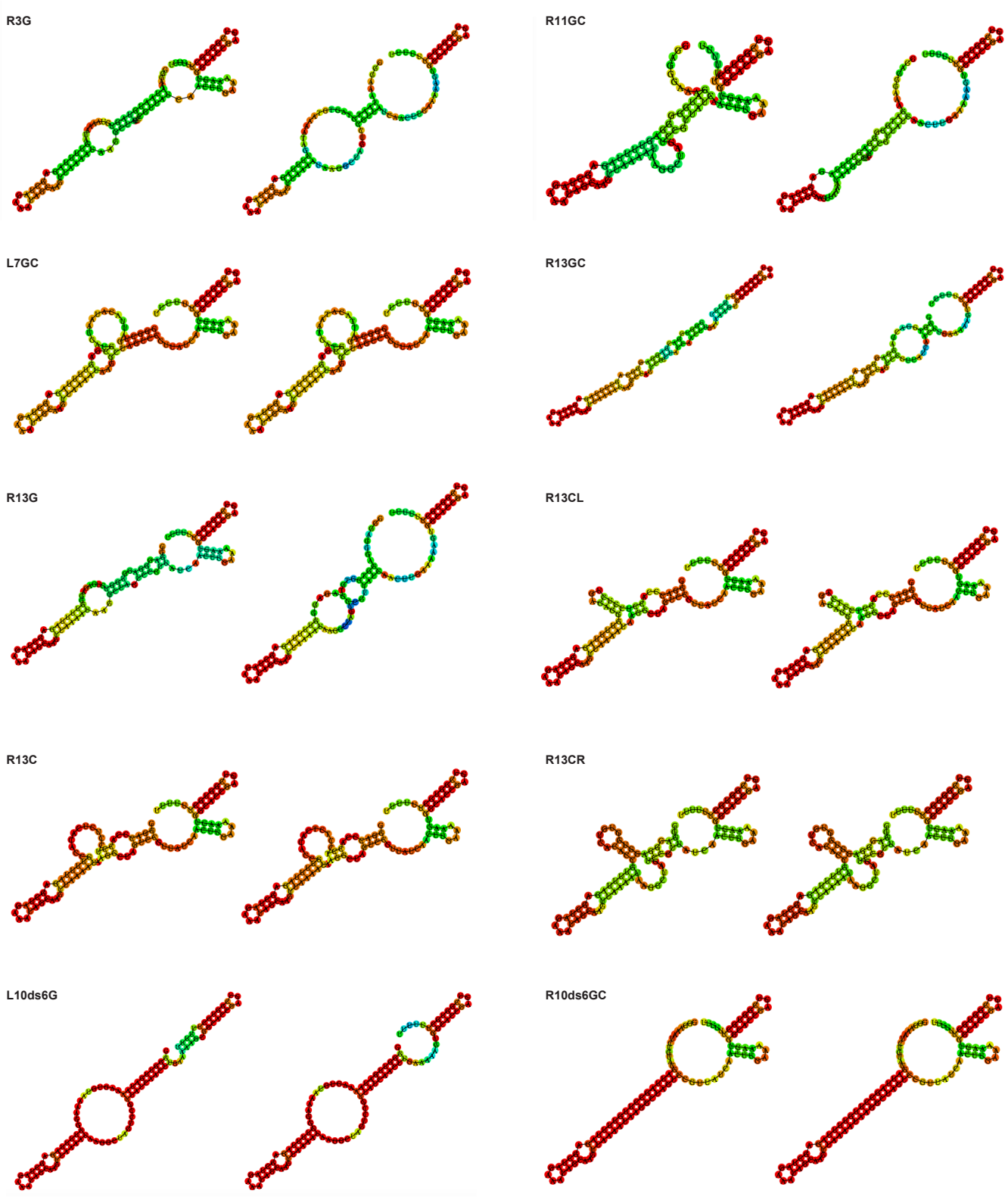
intron can be used. If the deleted fragment reinserted in the sense orientation and is expressed, a shorter fragment will be detected by RT-PCR compared to a larger fragment when the insert is expressed from the anti-sense strand.

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).



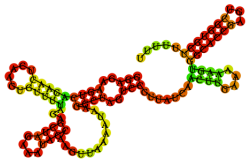
Supplemental Figure S1. Secondary structures of sgRNAs used in Figs 1-3 predicted by RNAfold.

A

sgRNA	length	sequence
CG6422 5'	20	GGAGAAGUCAGAACUCAAGU
CG6422 3'	20	GAGGCAGAAACAAUGGAUCU

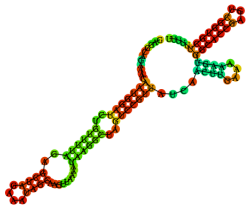
B **CG6422 5'**

```
                        AA
                       C   G
                       U   U
                        C-G
                        A-U
                        A-U
                        G-U
                   GA   A-U
               GGA  AGUC  AGAGCUAG A
               |||  ||||     ||||  A
 GUUCAACUAUUGCCU  UCGG     CGAUA
A ||||          GA    A   A
A AAAGUGGCACCGA       A   A
       ||||||| G      U   G
 UUUUUUCGUGGCU        A   U
                      A   U
                      A   A
                        A
```

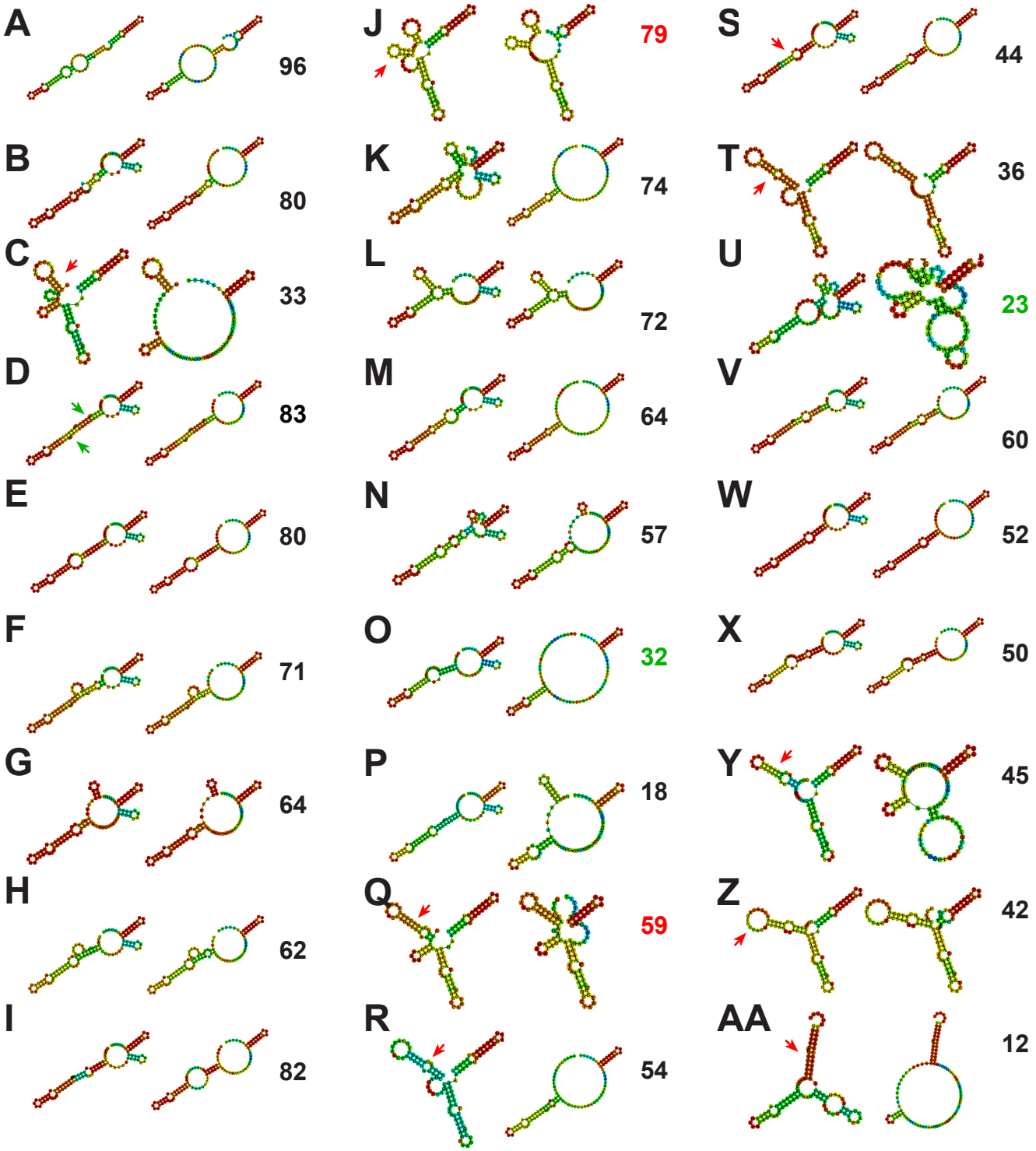


C **CG6422 3'**

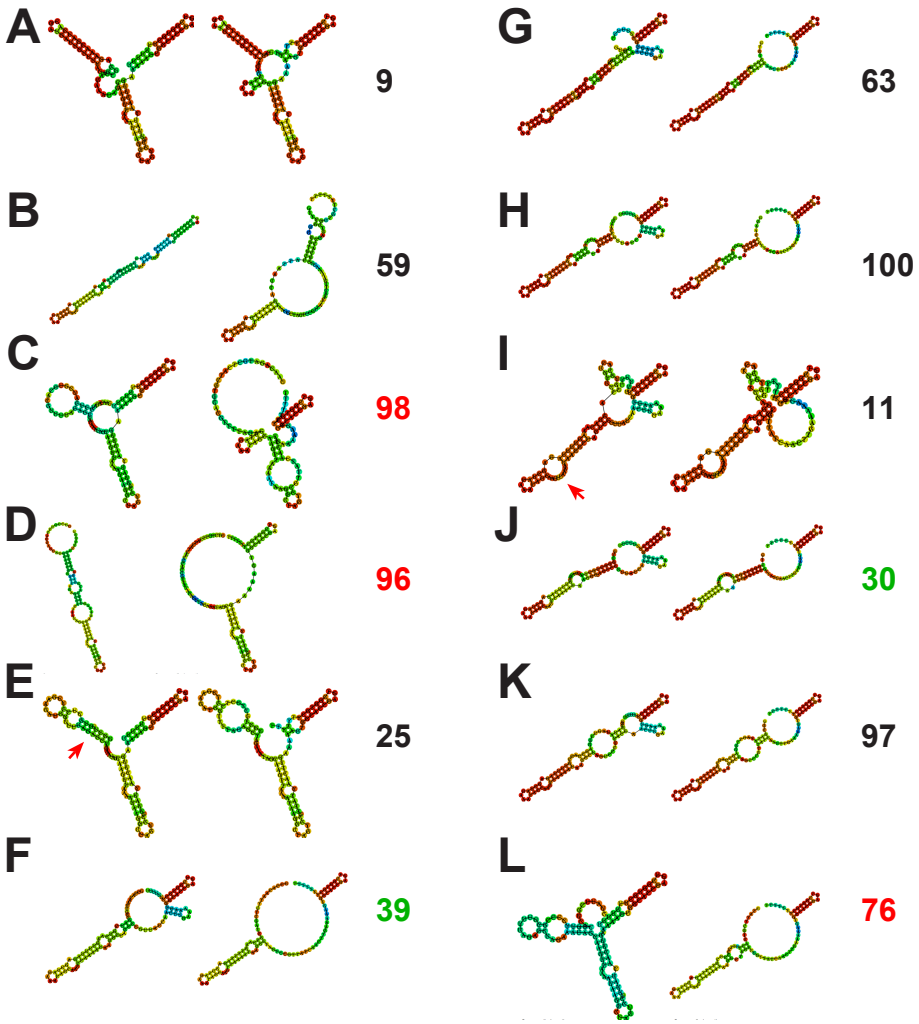
```
                        CU      G A
GAGGCAGAAACAAUGGAU  GUUUUA   GCUA G
            |||||||  ||||||   |||| A
 AGUUCAACUAUUGCCUG  CGGAAU   CGAU A
 A ||||           AU       A  A
  AAAGUGGCACCGA            A  A
        ||||||| G          A  G
 UUUUUUCGUGGCU             A  U
                             U
```



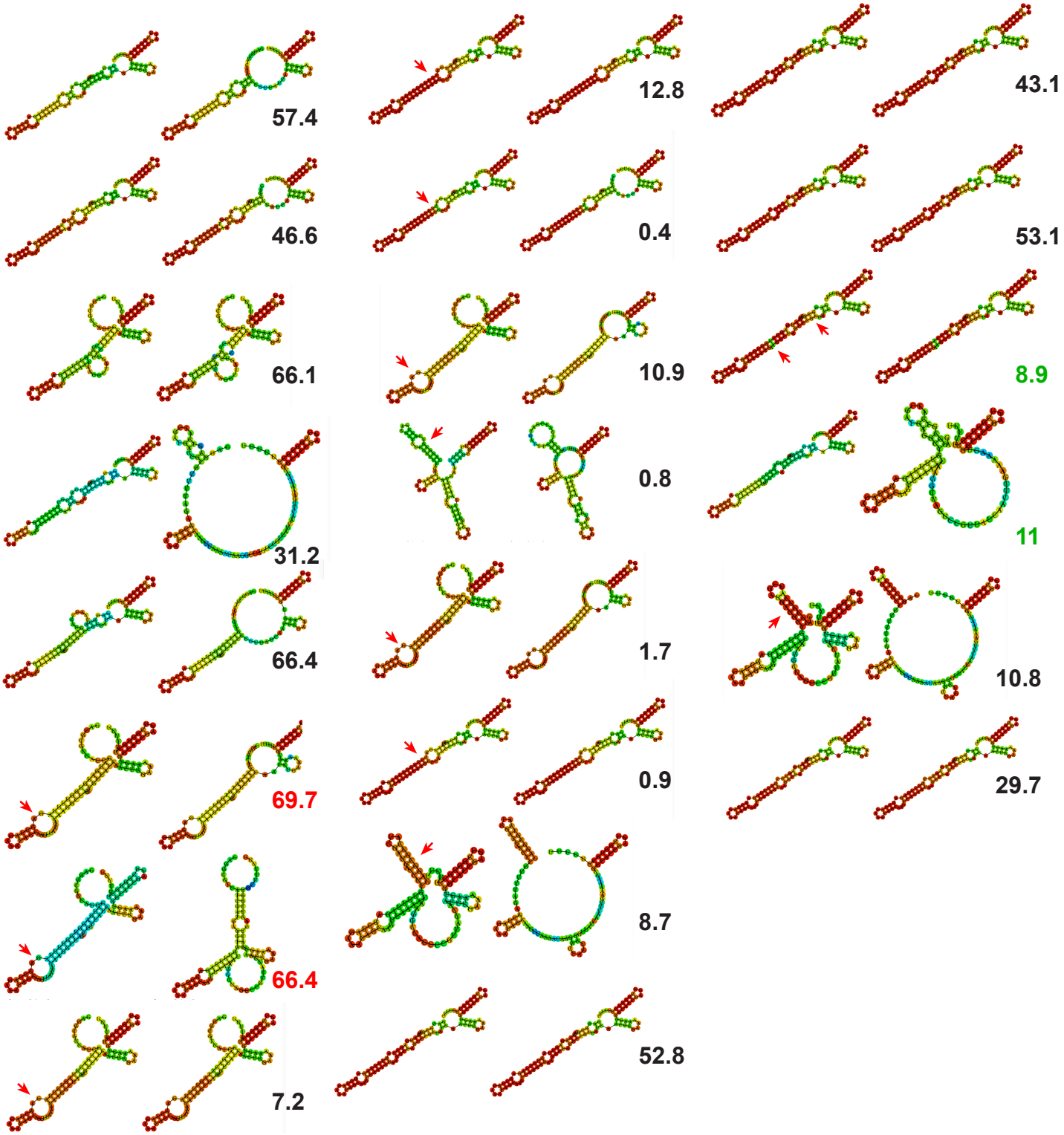
Supplemental Figure S2. Inactive sgRNAs targeted to the Drosophila ythdf (CG6422) gene.
(A) Sequences of sgRNAs targeting the CG6422 locus for generating a deletion. **(B, C)** Secondary structures of sgRNAs. The scaffold is shown on the left indicating Watson-Crick base-pairing by lines and the predicted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



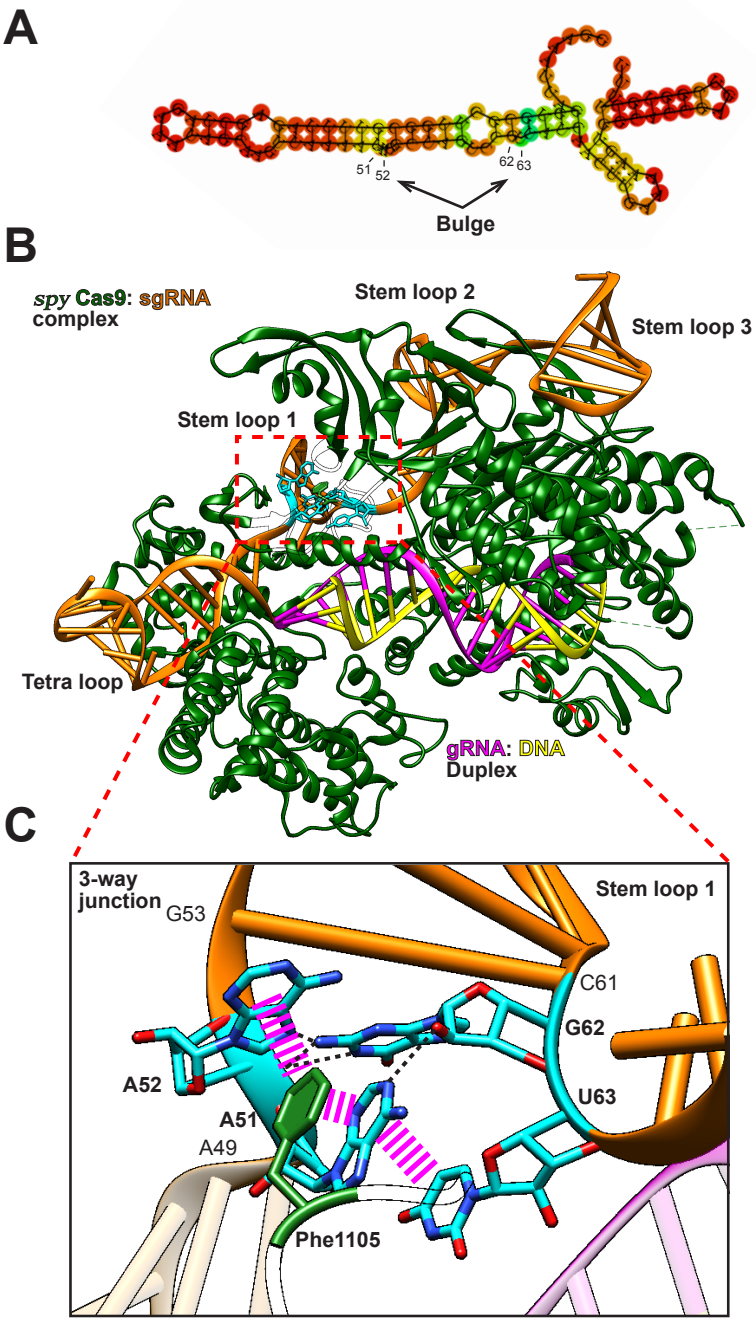
Supplemental Figure S3. Secondary structure of sgRNAs targeted to the Drosophila white gene from Ren et al. (2014).
(A-AA) Secondary structures of sgRNAs predicted by RNAfold. Minimal energy and proximity base-pair structures are shown on the left and right, respectively. Red and blue indicated high 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.



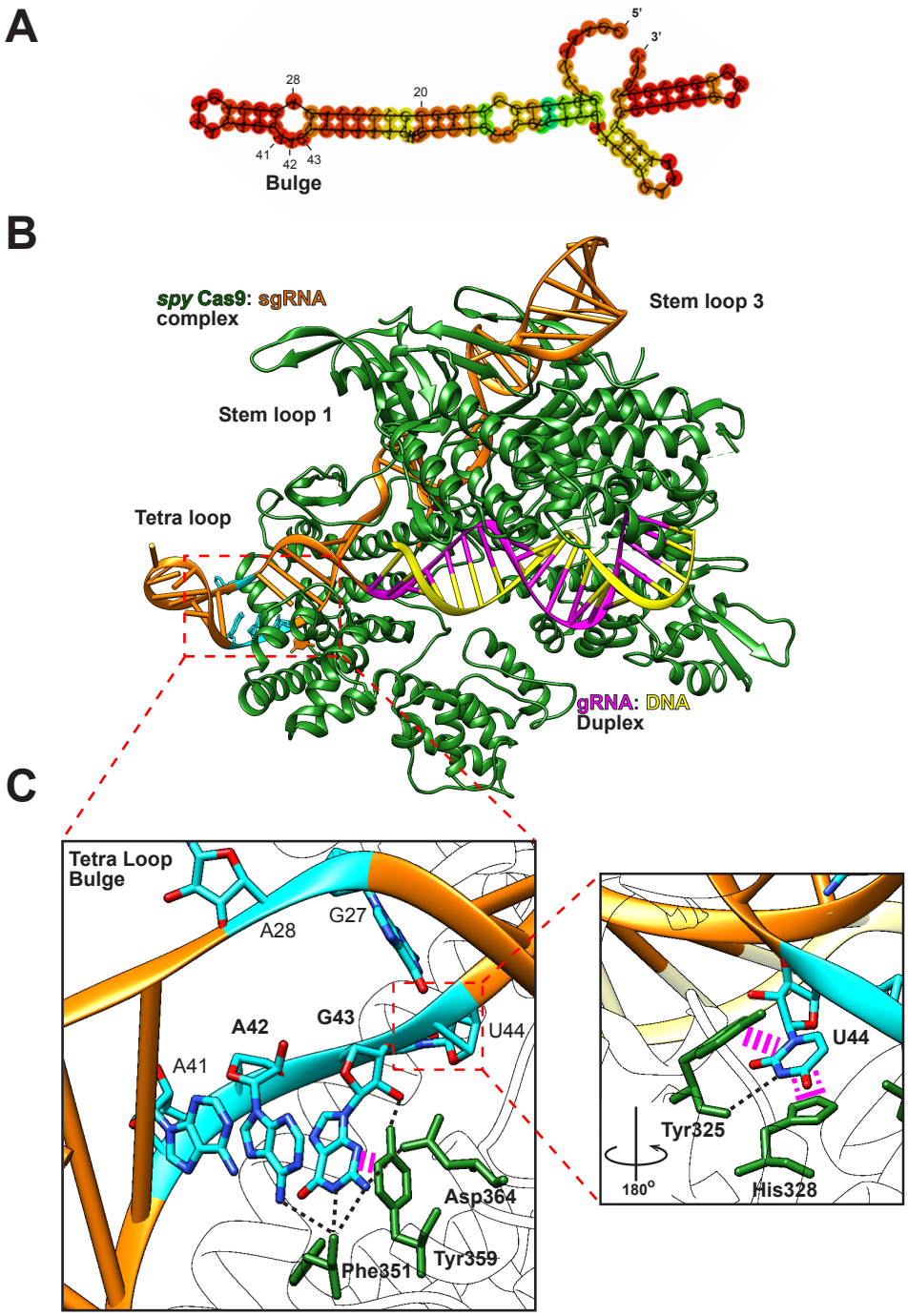
Supplemental Figure S4. Secondary structure of sgRNAs targeted to the Drosophila vermillion, ebony and yellow gene from Ren et al. (2014).
(A-L) Secondary structures of sgRNAs predicted by RNAfold targeting the vermillion (A-D), the ebony (E-H) and the yellow gene (I-L) . Minimal energy and proximity 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.



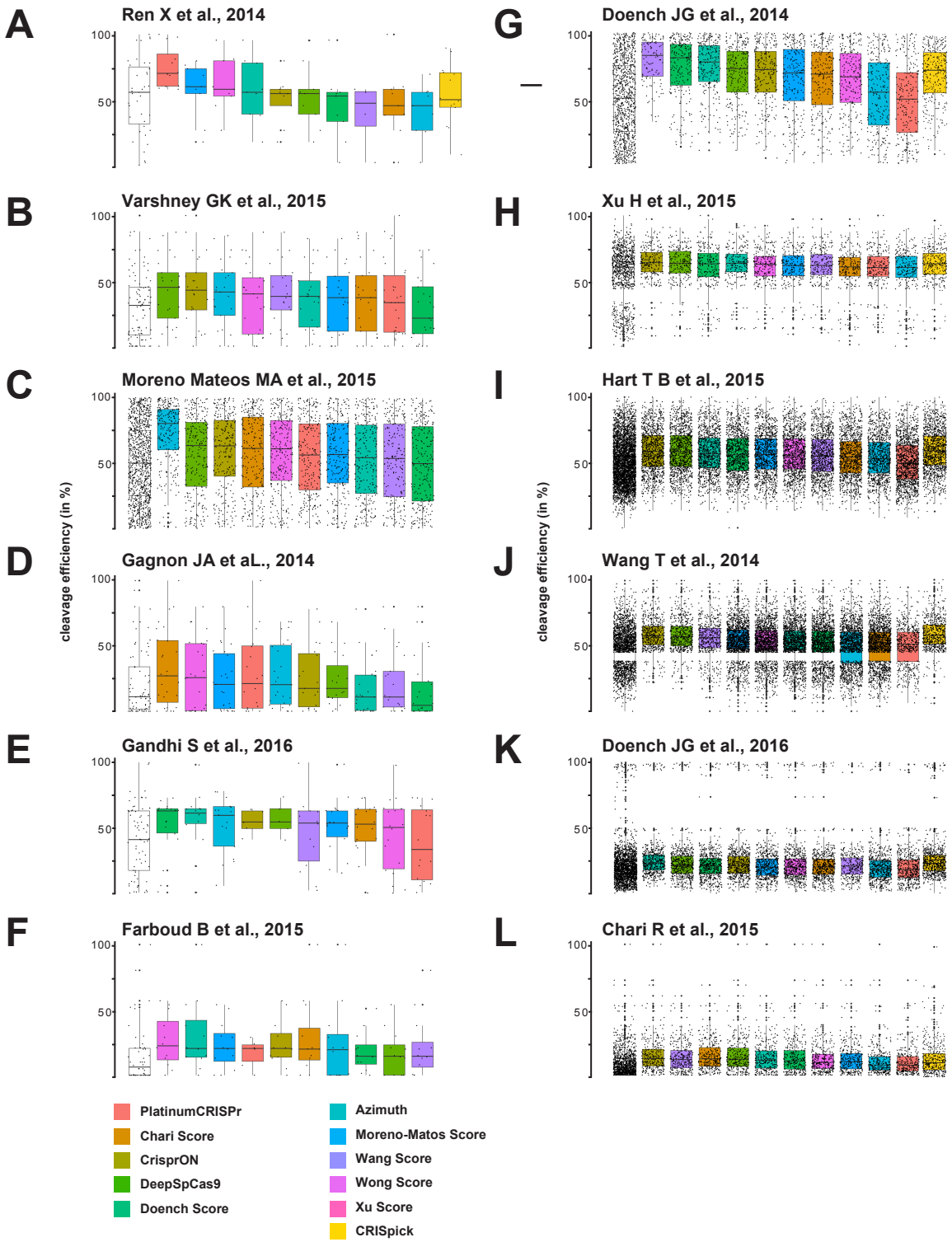
Supplemental Figure S5. Secondary structure of sgRNAs targeted to human CD22 from Graf et al. (2019). **(A-V)** Secondary structures of sgRNAs predicted by RNAfold. Minimal energy and proximity base-pair structures are shown on the left and right, respectively. Red and blue indicated high and low probabilities 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.



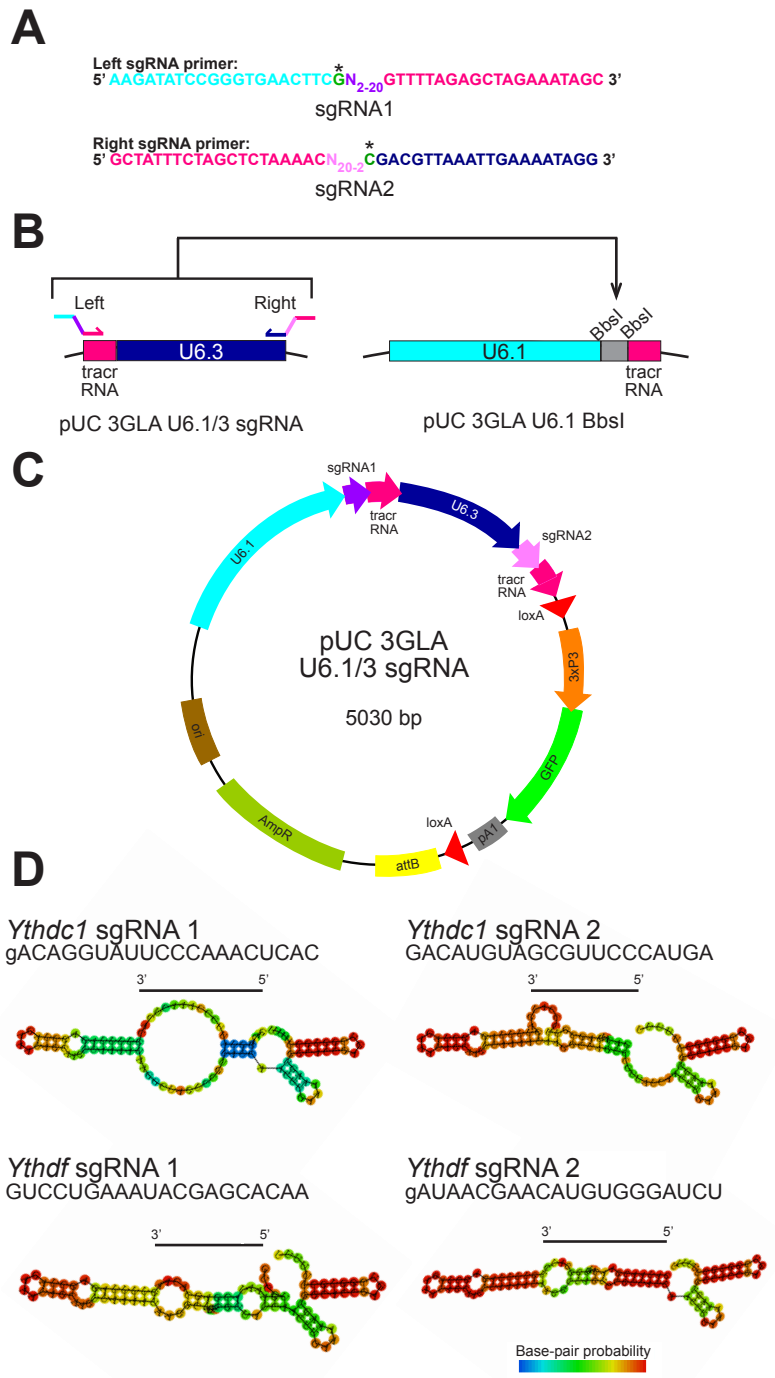
Supplemental Figure S6. Key sgRNA secondary structural features directly interact with the Spy Cas9 endonuclease in the three way junction.
(A) Minimum Free Energy (MFE) predicted secondary structure of the sgRNA used for co-crystallization of Spy Cas9 (Nishimasu et al., 2014). Nucleotides involved in 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.



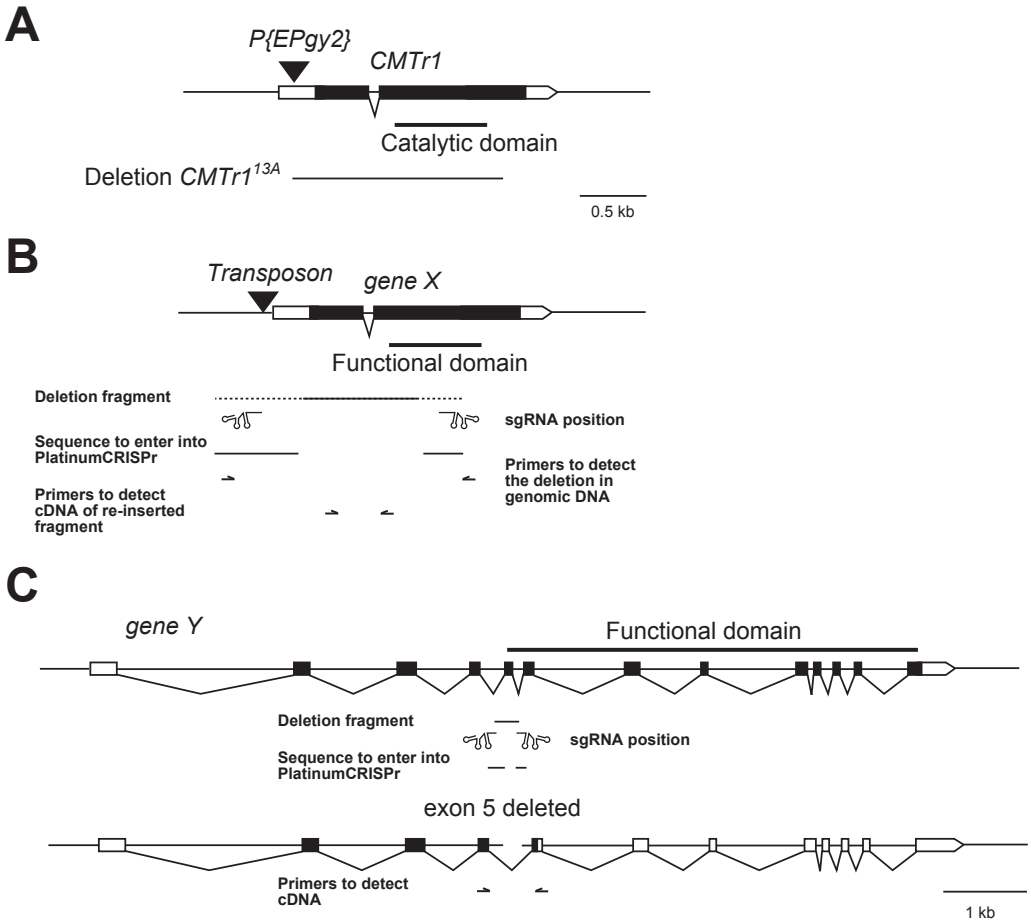
Supplemental Figure S7. Key sgRNA secondary structural features directly interact with the Spy Cas9 endonuclease in the tetraloop bulge.
(A) Minimum Free Energy (MFE) predicted secondary structure of the sgRNA used for co-crystallization of spyCas9 (Nishimasu et al., 2014). Nucleotides forming the bulge 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.