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

Irmgard U. Haussmann, Thomas C. Dix, David W.J. McQuarrie, et al.

Genome Res. 2024 34: 2279-2292 originally published online December 3, 2024
Access the most recent version at doi:10.1101/gr.279479.124

Method

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

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] and Matthias Soller[1,5]

[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 B15 3TN, United Kingdom;* [3] *Department of Cancer and Genomic Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom;* [4] *Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom;* [5] *Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PT, United Kingdom*

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 the 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. 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 advance the efficient use of this technique for safer applications in humans.

[Supplemental material is available for this article.]

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 nt 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% to 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, 2016; 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; 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 sequences with very low (≤35%) or very high (>80%) guanine–cytosine (GC) overall content were less effective indicating a critical aspect for binding energy in target scission. In addition, purines in the 6 nt 5′ to the PAM substantially increased Cas9 cleavage efficiency, while pyrimidines and in particular uridine resulted in a lower efficiency

Corresponding authors: matthias.soller@manchester.ac.uk, R.Arnold.2@bham.ac.uk

Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.279479.124. Freely available online through the *Genome Research* Open Access option.

(Ren et al. 2014; Wang et al. 2014; Housden et al. 2015; Graf et al. 2019). The lower efficiency of two uridines preceding the PAM site was further associated with premature termination of RNA Pol III (Graf et al. 2019), which can terminate after a stretch of four to six uridines (Gao et al. 2018). Moreover, changes in the internal structure of sgRNA have been found associated with low activity (Moreno-Mateos et al. 2015; Thyme et al. 2016; Jensen et al. 2017). Recently, also bioinformatic approaches employing machine learning have been used to improve the prediction of sgRNA cleavage efficiencies (Xiang et al. 2021). These observations have helped to improve the design of sgRNAs to yield higher efficiencies and were incorporated into sgRNA design tools, but correlations between predictions and guide activity vary considerably (Haeussler et al. 2016; Labun et al. 2016; Sledzinski et al. 2020). Accordingly, available rules to predict sgRNAs are currently not sufficient to guarantee high cleavage efficiency, and many sgRNA candidates scoring high fail to cleave efficiently (Haeussler et al. 2016; Labun et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020). In particular, the impact of sgRNA folding has not yet been analyzed in detail and incorporated in web tools for sgRNA design.

The X-ray crystal structure of Cas9 bound to sgRNA has been determined and indicates four regions of base-pairing termed tetraloop (tetraloop forms due to the fusion of crRNA and tracrRNA) and stem–loops 1–3 that might be important for its function (Anders et al. 2014; Nishimasu et al. 2014). This structure revealed many points of close interactions of the folded sgRNA with Cas9, but whether disruptions in the sgRNA structure would impact on Cas9 cleavage efficiency has not systematically been analyzed (Riesenberg et al. 2023). In particular, highly GC-rich gRNAs could disrupt the rather weak secondary structure of the sgRNA bound by Cas9.

The CRISPR–Cas9 genome editing tool is widely used to generate knockout mutants by introducing frameshifts. It has been recognized that introducing premature termination codons (PTCs) can induce the use of alternative translation initiation sites (Tuladhar et al. 2019). In addition, in CRISPR–Cas9 engineered "knockouts" of the N^6 methylated adenosines (m^6A) in mRNA of the methyltransferase *METTL3*, it has been found that a functional ORF can be restored by altered splicing leaving considerable levels of m^6A in mRNA (Poh et al. 2022). Likewise, compensatory responses have been observed involving upregulating genes and as a consequence causing stronger phenotypes than compared to removing a gene entirely (El-Brolosy et al. 2019; Ma et al. 2019). In addition, some genes have dual functions as protein and RNA (Hachet and Ephrussi 2004). Hence, introducing a frameshift will only remove the protein function. Likewise, many noncoding RNAs are present in introns suggesting that their expression is connected to the expression of the host gene (Boivin et al. 2018), but such relationships have not yet been explored comprehensively as they require more sophisticated genome editing (Deveson et al. 2017). Thus, for generating gene knockouts, removing the transcription start site or the entire gene should be considered.

The sequence complementary guided targeting of the gRNA/Cas9 complex makes this an ideal tool for genome editing, but its use is currently limited by the low predictability to cut its target in the genome (Haeussler et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020). In particular, the role of the RNA secondary structure of sgRNA has not been thoroughly evaluated and there is no sgRNA prediction tool that comprehensively takes the RNA secondary structure of sgRNAs into account. To facilitate the design of optimal sgRNAs, we developed an online tool incorporating all currently known parameters for sgRNA design including correct sgRNA folding (https://platinum-crispr.bham.ac.uk/predict.pl). Moreover, we provide guidelines to generate complete gene knockouts by generating deletions using dual sgRNAs. Taken together, these new sgRNA design and gene editing tools will help to develop CRISPR–Cas9 genome editing for safe application in humans.

Results

RNA secondary structure constraints limit sgRNA/Cas9 activity

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

It has previously been shown that extending the tetraloop in the constitutive component of sgRNAs constituted by tracrRNA and crRNA enhances cleavage efficiency in vitro using oligonucleotide substrates (Fig. 1C; Jinek et al. 2012). Introducing the extended sequence present in tracrRNA and crRNA into sgRNA (L7GCext) did not increase the efficiency of cleaving a plasmid at 37°C, but increasing the temperature to 42°C enhanced cleavage by L7GCext (Fig. 1B). Increasing the salt concentration to 200 mM also did not result in enhanced cleavage by L7GC/R3G (Fig. 1D). In contrast, both L7GC and L7GCext could cleave an oligonucleotide, while R13GC did not (Fig. 1E). Using this assay also confirms the previous observation that sgRNAs of shorter length will lead to cleavage. In addition, three guanosines introduced at the 5′ end of sgRNAs required for efficient in vitro transcription are tolerated (Fig. 1F,G; Jinek et al. 2012).

The sgRNA scaffold adopts a typical fold when bound to Cas9 consisting of the bulged tetraloop, followed by small loop 1 and the more extended loops 2 and 3, which form a protective 3′end structure (Fig. 1C; Nishimasu et al. 2014). The loop2/3 structure does not involve the uridines incorporated for termination of RNA Pol III driven expression from plasmids (Fig. 1C). When comparing the secondary structures of the four sgRNAs, we noticed that the 2 well-cutting sgRNAs L7GC or R3G maintained the secondary structure of the constitutive RNA part, while the noncutting sgRNA L11GC disrupted the structure of the tetraloop (Fig. 1H–K). The effect of R13GC seems more subtle as it could cut in the oligonucleotide assay suggesting that the repeated bulge structure is the cause for its inefficiency, which is supported by the X-ray crystal structure of the Cas9–sgRNA–DNA complex. Here, the bulge structure is recognized by Cas9 where Tyr_{359} base-stacks with G_{43}, that also forms hydrogen bonds with Asp_{364} and Phe_{351}, and Phe_{351} forms a hydrogen bond with A_{42} (Nishimasu

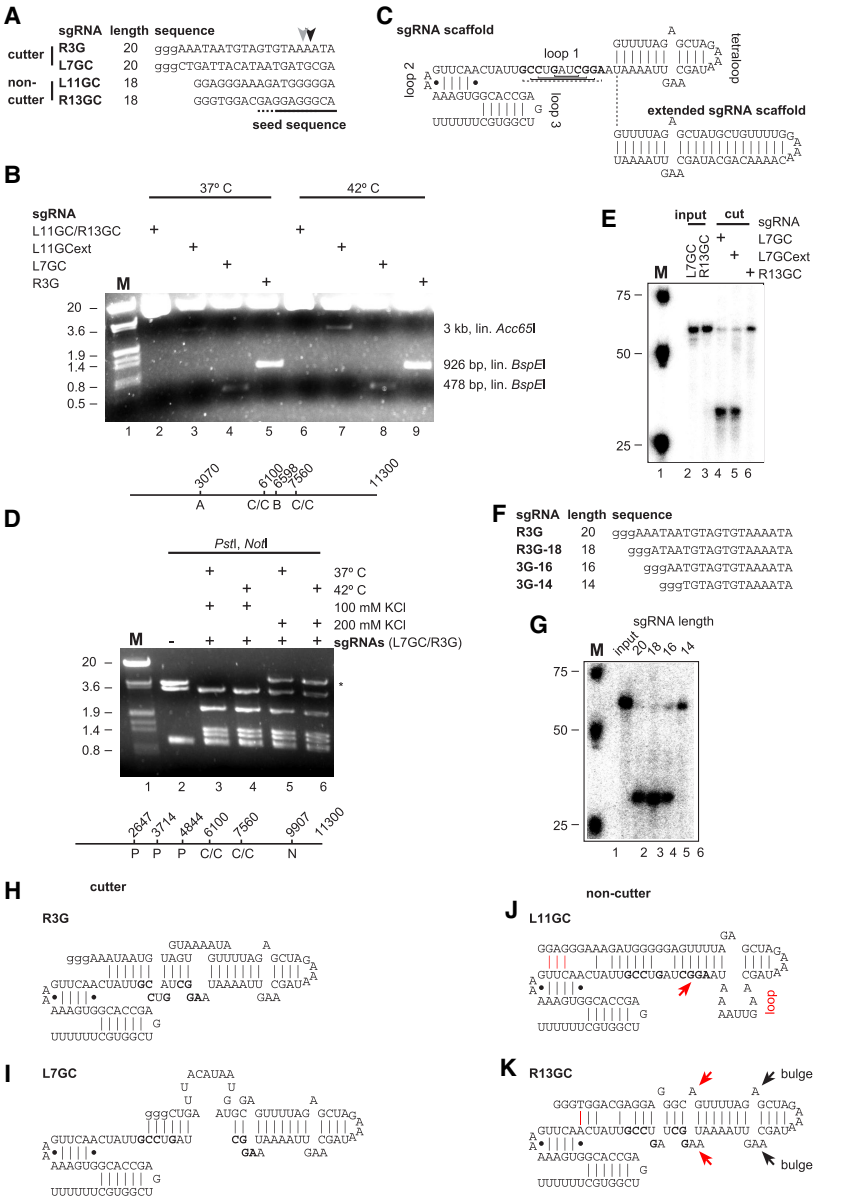

Figure 1. Sequence-dependent in vitro cleavage of oligonucleotides and plasmid DNA by the sgRNA/Cas9 complex. (*A*) Sequences of sgRNAs with observed cleavage sites indicated by arrow heads. Small letter guanosines used for in vitro transcription are not present in the target DNA sequence. The seed sequence is indicated by a line at the *bottom*. (*B*) Agarose gel showing Cas9-mediated cleavage of the 11.3 kb Dscam 3–5 plasmid for 24 h with indicated sgRNAs. Plasmids were cut with either Acc65I (lanes *2*, *3*, *6*, and *7*) or BspEI (lanes *4*, *5*, *8*, and *9*) after Cas9 cleavage. The line at the *bottom* shows a map of the plasmid with restriction sites indicated. Size markers are EcoRI/HindIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb, and 0.8 kb. (*C*) Structure of the sgRNA scaffold from cocrystallization with Cas9 (Nishimasu et al. 2014). Vertical or horizontal lines indicate Watson–Crick base-pairing, and dots or dashed lines indicate non-Watson–Crick base-pairing. Nucleotides base-pairing in loop 1 are bold. Additional base-pairing found in the tracrRNA–crRNA heterodimer is indicated in the extended scaffold (Jinek et al. 2012). (*D*) Agarose gel showing Cas9-mediated cleavage of the 11.3 kb Dscam 3–5 plasmid for 24 h with indicated sgRNAs L7GC and R3G. Plasmids were cut with PstI and NotI. The star denotes incomplete cleavage by NotI and the line at the *bottom* shows a map of the plasmid with restriction sites indicated. Size markers are EcoRI/HindIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb, and 0.8 kb. (*E*) Denaturing acrylamide gel showing Cas9-mediated cleavage of synthetic oligonucleotides with indicated sgRNAs. (*F*) Sequences of sgRNAs with variable length. Small letter guanosines used for in vitro transcription are not present in the target DNA sequence. (*G*) Denaturing acrylamide gel showing Cas9-mediated cleavage for 1 h of synthetic oligonucleotides with indicated sgRNAs of variable length. (*H–K*) Structure of sgRNAs. Nucleotides base-pairing in loop 1 are bold. Red lines in *J* and *K* indicate potential base-pairing with nucleotides in loop 2. The red arrow in *J* indicates the sequence complementarity leading to a bulge in the tetraloop. The red arrows in *K* indicate a duplication of the bulge structure present in the tetraloop.

et al. 2014). In addition, the sgRNAs initially used for deleting the *Ythdf* gene have a severely disrupted secondary structure (Supplemental Fig. S2).

When we systematically analyzed genome sequences from *Drosophila* or humans for correct folding and activity of sgRNAs using the above parameters, ~50% of sgRNAs (241 from 481 and 503 from 973, respectively) did not fold properly. Also, only ~10%–20% of randomly selected sgRNAs exert high cleavage efficiency suggesting that correct folding could essentially contribute to high cleavage efficiency (Hsu et al. 2013; Doench et al. 2014, 2016; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Moreno-Mateos et al. 2015; Liu et al. 2016; Abadi 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).

When comparing the sequences of the cutting sgRNAs L7GC or R3G with the noncutting L11GC and R13GC sgRNAs, we further noticed that L11GC and R13GC sgRNAs contained more guanosines, which in RNA can base pair with C and U. To test if guanosines in the sgRNA limit Cas9 activity, we increased their number in R3G to 13 to make sgRNA 13G (Fig. 2A). For the design of the R13 sgRNA, care was taken not to disrupt the tetraloop, but we noticed the potential to interfere with loop 2 (Fig. 2B, see below). We noticed, that sgRNA 13G is not capable of directing Cas9 cleavage if the target sequence is present in a 3 kb plasmid, but is active with a short oligonucleotide substrate (Fig. 2A–D). Likewise, adding a restriction enzyme together with sgRNA/Cas9 inhibited Cas9 in cleaving plasmid DNA suggesting that Cas9's ability to scan DNA can be impaired separately from its ability to cleave DNA.

Since the increased number of Gs in sgRNA R13G lead to enhanced base-pairing, we exchanged the Gs with Cs leading to an open structure in the seed region. This sgRNA R13C cleaved the test-plasmid efficiently (Fig. 2C,E). Introducing Cs in the left or right half of sgRNA R13G leads to short stem–loops and inefficient cleavage of the test-plasmid (Fig. 2C,F,G).

To further test to what extent base-pairing impacts on Cas9 activity, we generated sgRNAs L10ds6G and R10ds6GC, where the proximal or the distal half leads to complementary base-pairing of

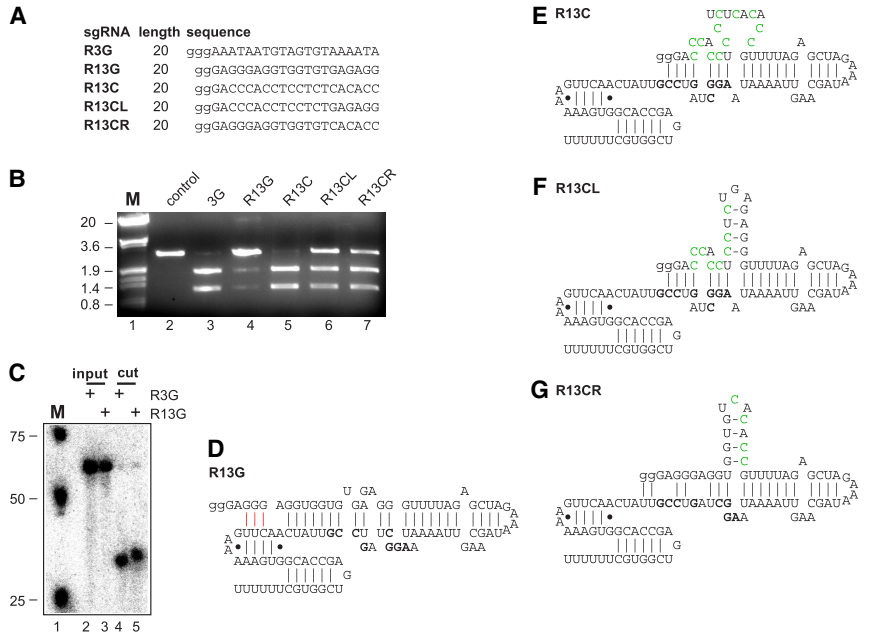

Figure 2. The secondary structure of sgRNAs affects Cas9-mediated cleavage efficiency. (*A*) Sequences of sgRNAs. Small letter guanosines used for in vitro transcription are not present in the target DNA sequence. (*B*) Agarose gel showing Cas9-mediated cleavage after 6 h of 3 kb pBS SK+ test-plasmids containing the target sequence with indicated sgRNAs. Plasmids were linearized with ScaI as in the control after Cas9 heat inactivation. Size markers are EcoRI/HindIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb, and 0.8 kb. (*C*) Denaturing acrylamide gel showing synthetic oligonucleotides before and after sgRNA/Cas9-mediated cleavage for 1 h. (*D–G*) Structure of sgRNAs. Nucleotides base-pairing in loop 1 are bold. Red lines in *D* indicate potential base-pairing with nucleotides in loop 2. Green nucleotides indicate mutations compared to sgRNA R13G.

the gRNA with the constant part, respectively (Fig. 3A,B). Although both sgRNAs supported Cas9 cleavage of oligonucleotide substrate, the R10ds6GC sgRNA base-pairing with the proximal part was mostly inactive in cleaving the plasmid indicating an impaired ability of Cas9 to scan DNA (Fig. 3C,D).

Taken together, these results demonstrate that the structure of the sgRNA is important for efficient Cas9-mediated DNA scission in vitro. Furthermore, high G content and base-pairing in the distal part of the gRNA also impair DNA scission, while base-pairing in the proximal part is tolerated.

To further substantiate these findings, we analyzed the structures of sgRNAs from previous studies in mammalian cells and *Drosophila* with regard to their cleavage efficiency of previous attempts to define rules for sgRNA cleavage efficiency in vivo (Ren et al. 2014; Graf et al. 2019). Indeed, in 39 sgRNAs designed for use in *Drosophila* reduced cleavage efficiency in nine sgRNAs is associated with disturbances of the sgRNA secondary structure resulting in a cleavage efficiency below 35% (Supplemental Figs. S3, S4; Supplemental Table S1C; Ren et al. 2014). Similarly, from 22 sgRNAs designed for use in mammalian cells, 13 had a cleavage efficiency below 35% associated with disturbances of the sgRNA secondary structure (Supplemental Fig. S5; Supplemental Table S1D; Graf et al. 2019). Similar results were also observed for the efficiency of sgRNAs in honeybees (Roth et al. 2019).

Given the requirement for correct folding of the sgRNA for efficient Cas9-mediated DNA scission, we further examined the X-ray crystal structure of the Cas9–sgRNA–DNA complex to see whether this would provide additional instructions to design sgRNAs (Nishimasu et al. 2014). Indeed, the first 2 nt, adenosine 51 and 52 (A_{51} and A_{52}) (Supplemental Figs. S6, S7) after the tetraloop form an aromatic base-stacking interaction with phenylalanine 1105 (Phe_{1105}) of Cas9. Furthermore, these interactions are stabilized by guanosine 62 (G_{62}) forming non-Watson–Crick hydrogen bonds with A_{51} A_{52} and Phe_{1105}, and uracil 63 (U_{63}) forms a base-stacking interaction with A_{52}. These interactions indicate that base-pairing of the gRNA with these nucleotides of the constant

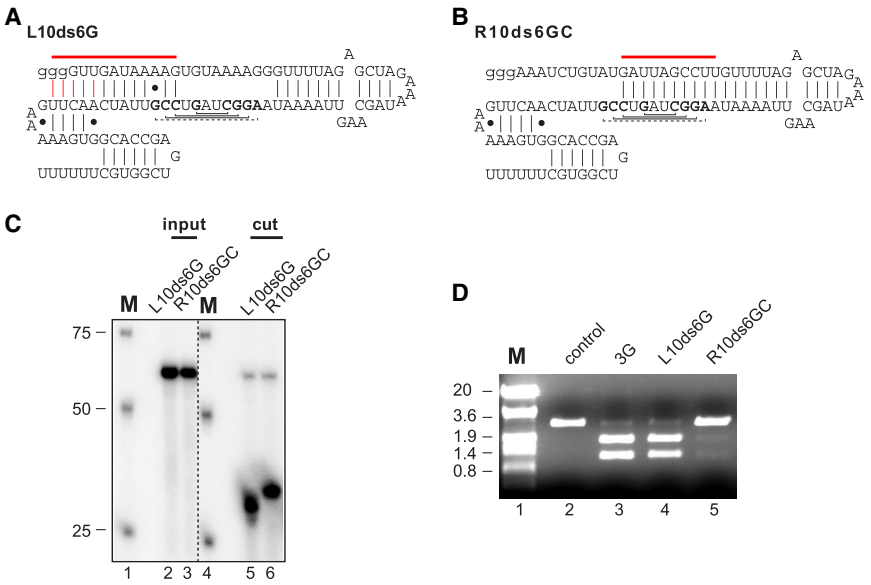

Figure 3. Base-pairing of sgRNAs in the seed-region blocks Cas9-mediated cleavage of a test-plasmid. (*A,B*) Structure of sgRNAs. Horizontal red lines in *A* and *B* indicate artificially introduced base-pairing with the sgRNA scaffold, and vertical red lines in a indicate potential base-pairing with nucleotides in loop 2. Nucleotides base-pairing in loop 1 are bold. (*C*) Denaturing acrylamide gel showing synthetic oligonucleotides before and after sgRNA/Cas9-mediated cleavage for 1 h. Note that Cas9 cleavage is heterogeneous. (*D*) Agarose gel showing Cas9-mediated cleavage after 6 h of 3 kb pBS SK+ test-plasmids containing the target sequence with indicated sgRNAs. Plasmids were linearized with ScaI as in the control after Cas9 heat inactivation. Size markers are EcoRI/HindIII digested λ DNA of 20 kb, 3.6 kb, 1.9 kb, and 0.8 kb.

part of the sgRNA reduces Cas9 activity in Cas9 cleavage assays in vitro and mutagenesis in vivo.

Large-scale evaluation of novel sgRNA design parameters

Next, we incorporated features previously published and from this study into a bioinformatic sgRNA design tool, https://platinum-crispr.bham.ac.uk/predict.pl (Supplemental Code; Hsu et al. 2013; Doench et al. 2014, 2016; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Moreno-Mateos et al. 2015; Liu et al. 2016; Abadi 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). The included features include validation of intact secondary structures (tetraloop, loops 2 and 3) (Figs. 1–3), presence of a tetraloop bulge mimic (Fig. 1K), self-complementarity of the gRNA (nucleotides 1–20) (Fig. 2F,G; Moreno-Mateos et al. 2015; Thyme et al. 2016; Jensen et al. 2017), GC content in the 6 nt seed region of the gRNA (nucleotides 15–20) (Supplemental Table S1A,C,D; Ren et al. 2014; Wang et al. 2014; Graf et al. 2019), GC content of the gRNA (nucleotides 1–20) (Supplemental Table S1A,C,D; Ren et al. 2014; Wang et al. 2014; Graf et al. 2019), the UUYY motif (nucleotides 16–20), which results in complete base-pairing (Fig. 3) and can act as a Pol III termination signal (Gao et al. 2018), the UCYG and CYGR motifs (nucleotides 16–20) associated with lower cleavage efficiency (Graf et al. 2019) and lack of base-pairing of nucleotides 40, 41, 51, and 52 that are engaged in contacts with Cas9 (Supplemental Figs. S6, S7).

PlatinumCRISPr operates at a cutoff prediction for either low or high efficiency, which is based on whether sgRNA folding is correct or whether certain detrimental motifs are present, but also on sequence composition. Although a score could be implemented for sequence composition, in practice, cleavage efficiency increases within a narrow range making a score at this point not feasible.

The genome sequence of any organism can be used with the PlatinumCRISPr server. For the most common model organisms, the PlatinumCRISPr server allows to choose the organism and will calculate off-targets (additional hits in the genome) as a complete match for the 16 nt before the tracrRNA as this is the minimal sequence for DNA scission (Fig. 1G). For other organisms off-target effects need to be determined manually by BLAST, e.g., using the NCBI webserver.

Although DNA scission is completely abrogated by a single mismatch to the target gene in *Drosophila* (Ren et al. 2014), off-target scission has been reported in tissue culture cells (Fu et al. 2013; Hsu et al. 2013). To compare off-target effects with sgRNA cleavage efficiencies predicted by either PlatinumCRISPr, DeepSpCas9, or Azimuth (standard scoring in CHOPCHOP), we used a large-scale double-strand break analysis from Cas9 expression across 59 targets using tagmentation-based tag integration site sequencing (TTISS) (Schmid-Burgk et al. 2020). In this analysis comparing off-target effects predicted by PlatinumCRISPr or CHOPCHOP, we find that off-target effects measured in tissue culture cells increase with predicted cleavage efficiency for PlatinumCRISPr (one-sided Wilcoxon rank-sum test, $P = 0.009$) and DeepSpCas9 (Spearman's rho, $P = 0.0016$), but not for CHOPCHOP (rho 0.1, $P = 0.41$).

To identify additional parameters affecting sgRNA cleavage efficiency, we performed a motif analysis among the 35% low-scoring sgRNAs for a number of different data sets (Doench et al. 2014, 2016; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Hart et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Xu et al. 2015; Gandhi et al. 2017; Xiang et al. 2021), but we did not find motifs associated with low performance in individual data sets. In particular, we found no evidence for 4 T's to be enriched, that could act as a termination signal for RNA Pol III (Gao et al. 2018; Graf et al. 2019).

We then analyzed the performance of a *Drosophila* sgRNA data set (Ren et al. 2014) according to the sgRNA design parameters described above. Our design tool PlatinumCRISPr selected 12 from 39 sgRNAs and those showed a cleavage efficiency of 58% or more (Fig. 4A). We then analyzed a number of sgRNA prediction tools for this data set including Chari score (Chari et al. 2015), CRISPRon (Xiang et al. 2021), DeepSpCas9 (Kim et al. 2019), Doench score (Doench et al. 2014), Azimuth (implemented in CHOPCHOP) (Doench et al. 2016), Moreno-Mateos score (Moreno-Mateos et al. 2015), Wang score (Wang et al. 2014), Wong score (Wong et al. 2015), and Xu score (Xu et al. 2015). PlatinumCRISPr significantly outperformed all of these prediction tools with the *Drosophila* data set (Fig. 4B; Ren et al. 2014).

Next, we analyzed 14 data sets from various organisms (*Drosophila*, zebrafish, sea squirt, worms, and cell culture cells), which determined sgRNA cleavage efficiency for their performance using the PlatinumCRISPr design tool (Doench et al. 2014, 2016; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Hart et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Xu et al. 2015; Gandhi et al. 2017; Xiang et al. 2021). For overall performance (Fig. 5), six data sets yielded significant ($P \leq 0.05$) enrichment of high-efficiency performing sgRNAs (Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Moreno-Mateos et al. 2015; Xiang et al. 2021), and five showed enrichment ($P \leq 0.25$) (Gagnon et al. 2014; Farboud

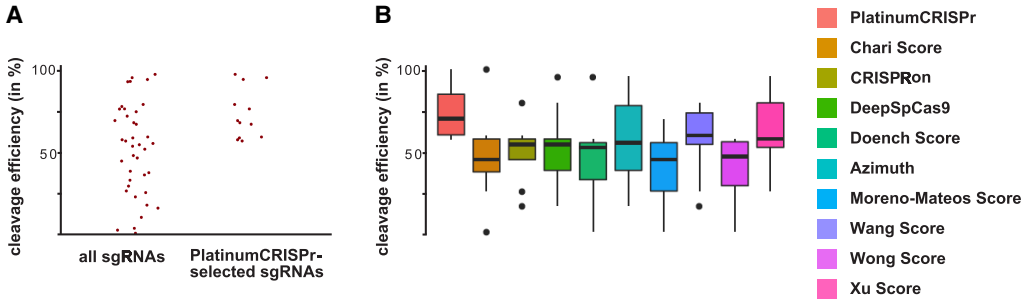

Figure 4. PlatinumCRISPr selects high-efficiency sgRNAs and outperforms other sgRNA selection tools for a *Drosophila* data set. (*A*) Comparison of the in vivo efficiency of all sgRNAs from Ren et al. (2014) data set with PlatinumCRISPr selected sgRNAs. (*B*) Comparison of the sgRNA selection performance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

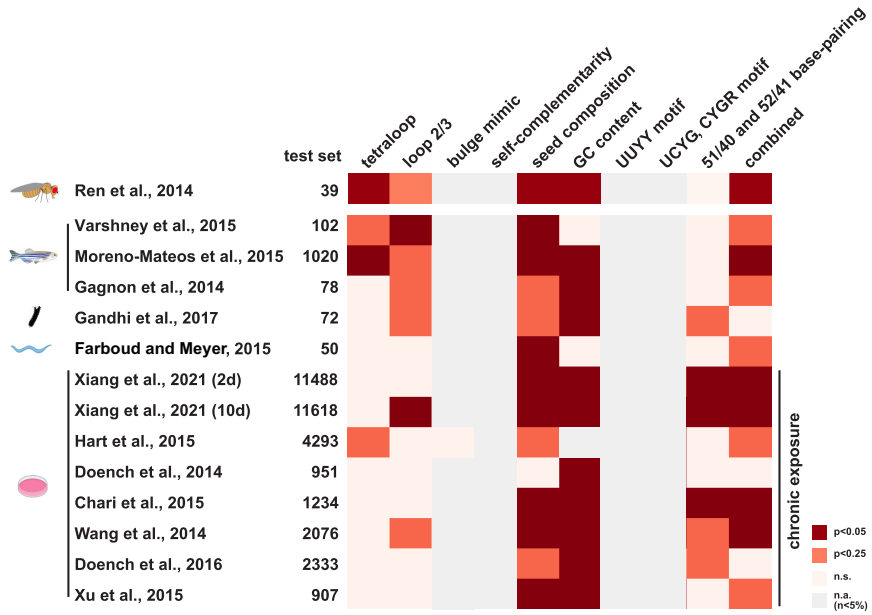

Figure 5. Comparison of individual sgRNA selection criteria for performance with different data sets. The different data sets with the number of sgRNAs tested are indicated on the *left*. The different selection criteria are shown on *top*. Significant and enriched performances are indicated in red and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

and Meyer 2015; Hart et al. 2015; Varshney et al. 2015; Xu et al. 2015), while two failed to show enrichment for most of the parameters (Doench et al. 2014, 2016). In this analysis, we noticed that structural constraints were significantly more important in cold-blooded organisms, where sgRNAs delivery is by injection in the absence of selection in contrast to cell culture cells, where delivery is by transfection and selection for chronic exposure to sgRNAs for up to 10 days before analysis.

When we analyzed the performance of PlatinumCRISPr, Chari score (Chari et al. 2015), CRISPRon (Xiang et al. 2021), DeepSpCas9 (Kim et al. 2019), Doench score (Doench et al. 2014), Azimuth (implemented in CHOPCHOP) (Doench et al. 2016), Moreno-Mateos score (Moreno-Mateos et al. 2015), Wang score (Wang et al. 2014), Wong score (Wong et al. 2015), Xu score (Xu et al. 2015), and CRISPick (Sanson et al. 2018) prediction tools (Supplemental Fig. 8A–L) with the different data sets determining sgRNA cleavage efficiency we found that Wang score performed best on the Doench data set and that PlatinumCRISPr and Moreno-Mateos performed best with *Drosophila* and zebrafish generated data sets, respectively, but we found no single prediction tool that stood out. In addition, in 5 out of 6 cell culture-generated data sets none of the prediction tools outperformed the others or substantially increased prediction efficiency (Supplemental Fig. 8A–L).

In addition, we have also analyzed data sets for the performance of sgRNA libraries in different cell lines (HeLa, HCT 116, and GBM) (Hart et al. 2015) and find very high correlations for all comparisons (Spearman's correlation between 0.71 and 0.77) indicative of minimal differences of the cleavage efficiency of sgRNAs in these cell lines. In addition, sgRNAs selected by PlatinumCRISPr or CHOPCHOP show a significantly higher cleavage efficiency than discarded sgRNAs in all 3 cell lines ($P \leq 0.03$).

The Cancer Dependency Map (DepMap) project (https://depmap.org) provides efficacy scores for 193476 gRNAs (between 0 and 1, the median efficacy in the complete data set is 0.974) (DepMap 24Q2 Public; https://doi.org/10.25452/figshare.plus.25880521.v1), of which ~25% are predicted as cutting by PlatinumCRISPr. A comparison of DepMap gRNAs resulted in a minimally better performance (0.006%, median efficacy: 0.975) of guides predicted as noncutting by PlatinumCRISPr, but the biological significance of this observation remains unclear since the overall performance substantially exceeded the gRNA performance of other cell-line based experiments with chronic exposure (Supplemental Fig. S8), possibly because gRNAs were preselected (DepMap 24Q2 Public; https://doi.org/10.25452/figshare.plus.25880521.v1).

As part of this analysis, we noticed that the average cleavage efficiency in the analyzed data sets varied substantially (from 20% to 75% average cleavage efficiency) (Supplemental Fig. 8A–L), pointing toward a bias in outcome when testing various prediction tools with different data sets. To identify common patterns among the different prediction tools applied to different data sets in the following analysis we, therefore, excluded data sets with cleavage efficiencies below 30% for further analysis (Gagnon et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Doench et al. 2016).

The default element of CHOPCHOP is Azimuth (Labun et al. 2019), but also implements elements from Doench score, Chari score, Xu score, and Moreno-Mateos score. We, therefore, reasoned that a combination of two prediction tools could result in more reliable selection of high-efficiency cleaving sgRNAs. When we combined two prediction tools the highest scoring combination was PlatinumCRISPr together with Wong score with an average cleavage prediction of 61% (Fig. 6A) and this combination also outperformed in all the remaining data sets (Fig. 6B), but this increased the stringency and only very few sgRNAs were selected.

A *Drosophila* transformation vector for expression of two sgRNAs

Next, we applied these novel sgRNA design rules to generate *Drosophila* gene deletions. For this purpose, we generated a new fly transformation vector with a GFP marker (Solomon et al. 2018), that is easier to select than previously generated *vermillion* marked vectors that require a *vermillion* mutant background for transgene identification (Port et al. 2014; Trivedi et al. 2020). This vector expresses two 20 nt sgRNAs from *U6.1* and *U6.3* promoters (Supplemental Fig. S9A–C), harboring a G as the first nucleotide as a requirement for expression from the *U6* promoter (Paule and White 2000; Ren et al. 2013). This plasmid can be generated by incorporating the two sgRNA sequences in PCR primers for single-step cloning into the plasmid, while previously published vectors require two cloning steps or plasmid recombination (Port et al. 2014; Trivedi et al. 2020). This sgRNA vector can then be injected into *Drosophila* expressing Cas9 in the germline for CRISPR to induce mutations. Alternatively, this vector can be used to generate a transgenic line via the attB site using phiC31 integrase-mediated

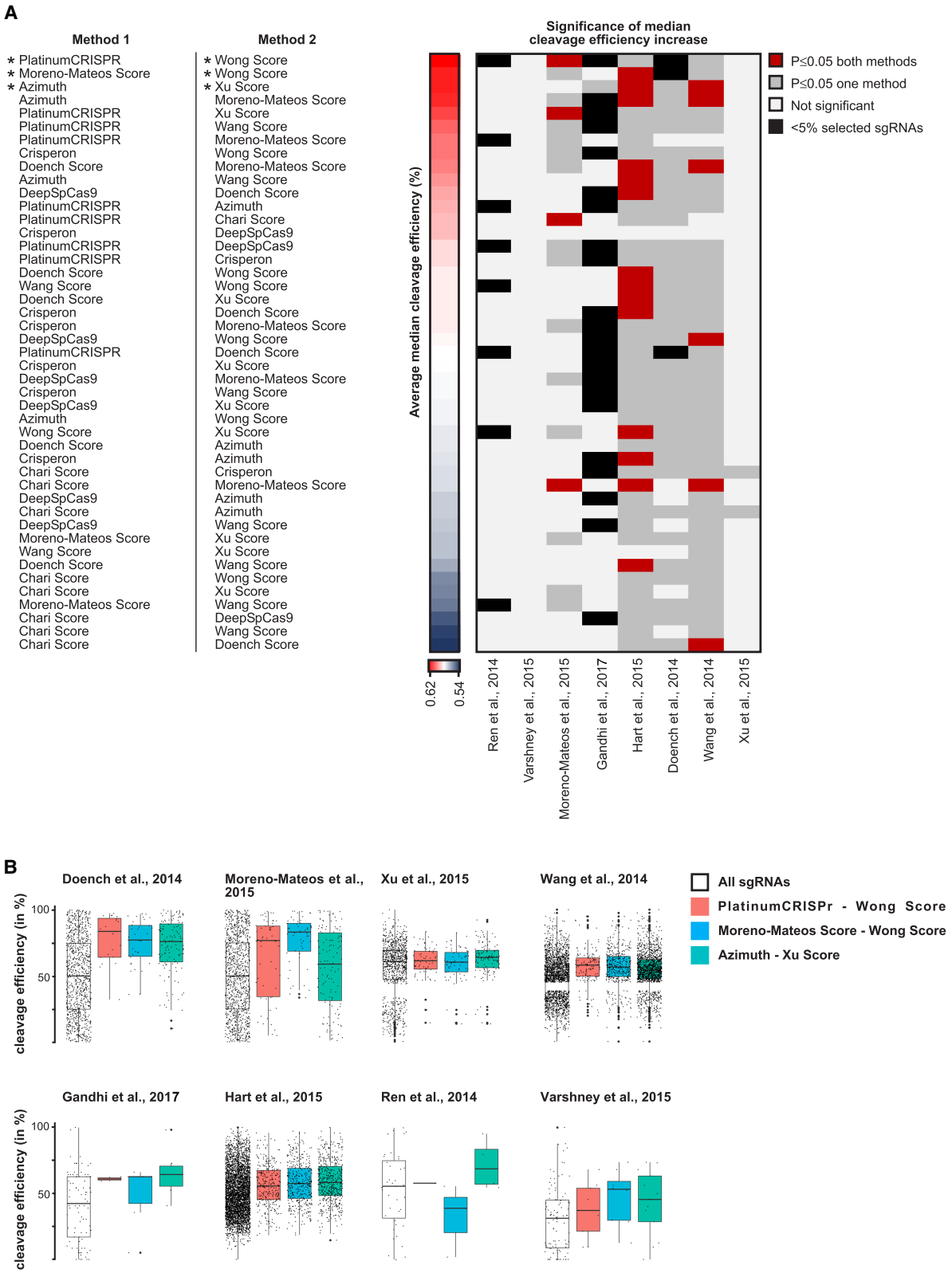

Figure 6. Combinations of two sgRNA selection tools select high-efficiency cleaving sgRNAs for several sgRNA efficiency screen data sets. Combinations of sgRNA selection tools in *A* are listed according to the overall cleavage efficiency of selected sgRNA significant ($P < 0.05$) for both methods (red) or one method (dark gray). Black indicates events with <5% of sgRNAs selected by at least one method. Comparison of sgRNA selection by different sgRNA selection tools shown as the median of the cleavage efficiency for individual data sets (*B*). The distribution of cleavage efficiencies for all sgRNAs is shown on the *left* (white box) and for PlatinumCRISPr–Wong score in red, Moreno-Mateos–Wong score in blue, and Azimuth–Xu score in green.

transformation. This fly strain is then crossed to a line expressing Cas9 in the germline for the generation of the desired genetic lesion. Transgenically provided sgRNA/Cas9 generally results in a higher efficiency, because the sgRNAs are provided maternally. Moreover, excision efficiency can be further increased by mutagenesis over a chromosomal deficiency which will remove the repair template for homologous recombination.

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

We then applied the PlatinumCRISPr tool to identify high-efficiency sgRNAs to generate gene deletions in *Drosophila Ythdc1* and *Ythdf*, that bind to m^{6}A in mRNA important in development (Dezi et al. 2016; Haussmann et al. 2016; Roignant and Soller 2017; Balacco and Soller 2019; Anreiter et al. 2021). *Ythdc1* and *Ythdf* are located on the third chromosome, and transposon inserts *Mi{MIC}YT521-B*^{MI02006} and *PBac{SAstopDsRed}*^{LL04081} marked with *GFP* or *RFP*, respectively, are available from stock centers. They were combined with an X-linked *vasCas9* or *nosCas9* for germline expression of Cas9. To allow for detection of transposon loss in the YTH protein genes the GFP and RFP markers of the *vasCas9* insert had been removed. These flies were then crossed to the GFP-marked sgRNA construct inserted on the third chromosome (Fig. 7A,F). The sgRNA insert has a weak GFP marker and can generally be distinguished from *Mi{MIC}YT521-B*^{MI02006} inserts. Females from this cross were mated with males containing *TM3 Sb/TM6 Tb* double-balancers in single crosses to recover sgRNA-induced individual deletions and avoid analysis of clonal events. The male progeny was then screened for loss of the GFP or RFP marker. Males which had lost the marker were detected in 100% ($n=9$) of the crosses for *Ythdc1* and in 88% ($n=9$) for *Ythdf*, respectively.

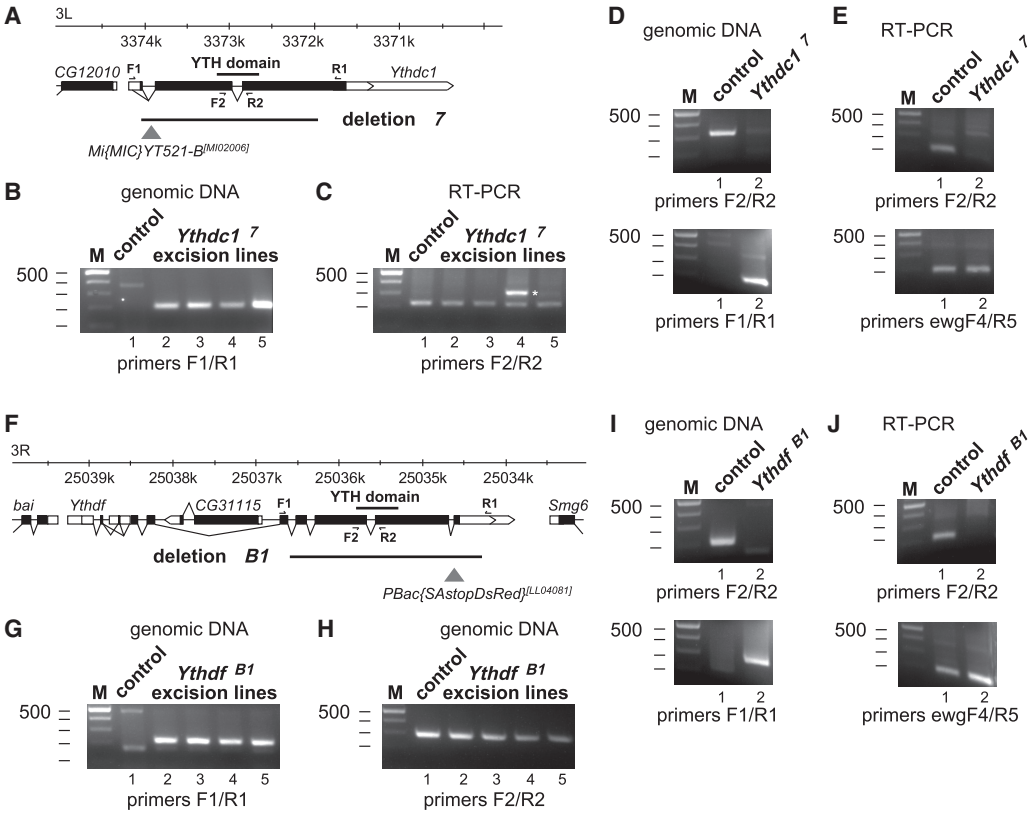

Figure 7. Generation of gene deletions in *Drosophila* YTH protein genes using two sgRNAs/Cas9 and transposon markers. (*A*) Schematic to the *Ythdc1* locus indicating transcripts (white boxes) and the ORF (black boxes) *below* the chromosome. Primers used are indicated on *top* and *below* the transcripts. The w+ marked transposon used for detecting deletions in the locus is indicated by a triangle and the deletion generated is indicated by a line. (*B*) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdc1* transposon excision lines using primers flanking the deletion. The presence of a PCR product indicates the expected gene deletion. DNA markers are indicated on the *left*. (C) Agarose gels showing RT-PCR products amplified from cDNA of control or *Ythdc1* transposon excision lines using internal primers flanking an intron. Ectopic insertion in the opposite orientation is indicated by an asterisk (lane *4*) as in this instance the transcript is not spliced. DNA markers are indicated on the *left*. (*D*) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdc1*$^{\Delta 7}$ flies using internal primers and primers flanking the deletion. DNA markers are indicated on the *left*. (*E*) Agarose gels showing RT-PCR products amplified from cDNA of control or *Ythdc1*$^{\Delta 7}$ flies using internal primers in *Ythdf* and *ewg* genes. The PCR product of the *ewg* gene was used as a loading control. DNA markers are indicated on the *left*. (*F*) Schematic to the *Ythdf* locus indicating transcripts (white boxes) and the ORF (black boxes) *below* the chromosome. Primers used are indicated on *top* and *below* the transcripts. The RFP-marked transposon used for detecting deletions in the locus is indicated by a triangle and the deletion generated is indicated by a line. (*G*) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdf* transposon excision lines using primers flanking the deletion. The presence of a PCR product indicates the expected gene deletion. DNA markers are indicated on the *left*. (*H*) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdf* transposon excision lines using primers flanking an intron. DNA markers are indicated on the *left*. (*I*) Agarose gels showing PCR products amplified from genomic DNA of control or *Ythdf*$^{\Delta B1}$ flies using internal primers and primers flanking the deletion. DNA markers are indicated on the *left*. (*J*) Agarose gels showing RT-PCR products amplified from cDNA of control or *Ythdf*$^{\Delta B1}$ flies using internal primers in *Ythdf* and *ewg* genes. The PCR product of the *ewg* gene was used as a loading control. DNA markers are indicated on the *left*.

This is a substantial increase of efficiency over imprecise P-element excision with a frequency of 0.01%–1% (Soller et al. 2006; Haussmann et al. 2016, 2022). In those crosses with loss of markers, all males for *Ythdc1* had lost the GFP marker, while for *Ythdf* the average frequency of marker loss was 42% ($n = 8$). In the reverse cross for *Ythdf*, where sgRNA/Cas9-mediated excision is induced in males, the frequency was 0% ($n = 5$) indicating that the low frequency in males is linked to the absence of recombination in males. The identified single marker-less males were then crossed to *TM3 Sb/TM6 Tb* double-balancers to establish a line and analyzed with PCR using primers next to the deletion breakpoints yielding a short PCR product (Fig. 7B,G). A PCR product had been obtained in all lines were the marker had been lost indicating that the expected deletion had indeed been generated. To generate *Ythdc1* excision lines, *nosCas9* was used for germlines, expression of Cas9, as *vasCas9* together with sgRNAs targeting *Ythdc1* resulted in female sterility.

sgRNA / Cas9 scissioned fragments insert elsewhere in the genome

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

To remove this ectopic insert(s), positive lines were crossed to *w+* marked deficiencies and out-crossed for two generations. The X and second chromosomes were then exchanged to establish null-mutant lines from single chromosomes for *Ythdc1* and *Ythdf* that were confirmed by RT-PCR to be free of any ectopic inserts (Fig. 7D,E,I,J). To avoid such complications in the future we generated a *PBac w+* containing vector which can be efficiently inserted into a locus by cloning left and right homology arms either to induce a partial deletion upon insertion or to introduce mutations. Alternatively, mutations in sgRNA cleavage sites can be introduced into the homology arms to insert point mutations. Afterward, the *PBac w+* is removed scarless by a transposase.

Discussion

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

DNA scission by the sgRNA/Cas9 complex is highly specific and requires complete base-pairing between the sgRNA and the target DNA generally not tolerating single missmatches (Ren et al. 2014; Farboud and Meyer 2015). This feature makes the sgRNA/Cas9 complex an ideal tool for genome editing, but its use is currently limited by the low predictability to cut its target in the genome (Haeussler et al. 2016; Labuhn et al. 2018; Sledzinski et al. 2020).

Here, we discovered that the structure of the sgRNA is a key determinant for the scission efficiency of the sgRNA/Cas9 complex in *Drosophila*. However, only ~50% of sequences adjacent to PAM sites constitute sgRNAs that fold properly and are not compromised by unfavorable base-pairing. In support of these two levels of interference, sgRNA R13GC correctly folds the tetraloop and loop2/3, but did not cleave a short oligonucleotide substrate. This indicates that the structure of this sgRNA blocks the catalytic activity of the sgRNA/Cas9 complex, likely by mimicking the bulge structure of the tetraloop. Second, some sgRNAs allowed cleavage of short oligonucleotide substrates (e.g., L7GC and R10ds6GC), but did not support efficient DNA scission of the target sequence in the context of a 3 kb test-plasmid. Likely, these sgRNAs interfere with the ability of the sgRNA/Cas9 complex to scan the DNA for target sites.

When we analyzed the cleavage efficiencies of sgRNAs used in *Drosophila* (Ren et al. 2014), we observed a good overlap with the ability of those sgRNAs to adopt the correct structure and having a high cleavage efficiency. Further refinement to the design of sgRNAs comes from the recognition that the GC content in the seed region is a major determinant to cleavage efficiency in addition to general GC content. Analysis of the X-ray crystal structure of the Cas9–sgRNA–DNA complex also revealed that the two As in the tracrRNA before the tetraloop engage with Cas9 through base stacking and hydrogen bonds (Nishimasu et al. 2014). Base-pairing of these two As with Us at the end of the sgRNA (N_{19} and N_{20}) before the start of the tracrRNA impacts sgRNA/Cas9 complex function and reduces cleavage efficiency (Graf et al. 2019). Likewise, if sgRNAs are made by RNA Pol III, termination occurs at the boundary of the tracrRNA if two UU precede the GUUUU of the start of the tracrRNA (Arimbasseri and Maraia 2015; Graf et al. 2019). Folding of the sgRNA is not anticipated to impact on transcription as it occurs afterward. Motif searches in various data sets to determine sgRNA cleavage efficiencies did not reveal any further motifs that impact on cleavage efficiency. If such bias exists, this would likely have been exploited by parasites of prokaryotic hosts.

Structural constraints of sgRNAs are more pronounced in cold-blooded animals

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

Evaluation of different sgRNA prediction tools

In this study, we also compared various sgRNA cleavage efficiency prediction tools with 12 data sets that have determined sgRNA cleavage efficiencies in human cells and various model organisms including *Drosophila*, zebrafish, *Caenorhabditis elegans*, honeybees,

and seasquirt (Doench et al. 2014, 2016; Gagnon et al. 2014; Ren et al. 2014; Wang et al. 2014; Chari et al. 2015; Farboud and Meyer 2015; Hart et al. 2015; Moreno-Mateos et al. 2015; Varshney et al. 2015; Wong et al. 2015; Xu et al. 2015; Gandhi et al. 2017; Kim et al. 2019; Roth et al. 2019; Xiang et al. 2021). Here, the PlatinumCRISPr tool is deemed best for *Drosophila*, Moreno-Mateos score for zebra fish, and Wang score for one human cell culture screen, but all prediction tools failed to convince for the other 5 cell culture screens. Possibly, chronic exposure over several days in cell culture systems could lead to a bias in determining sgRNA cleavage, compared to short exposure when injected into early stage embryos like in insects and zebra fish.

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

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

Initial concerns about CRISPR–Cas9 specificity were about off-target cleavage, but changing 1 nt in the spacer sequence complementary to the target efficiently abrogates cleavage activity (Ren et al. 2014). PlatinumCRISPr takes off-targets into account by screening the genome for the 16 nt distal of the PAM site, which require an exact match for cleavage, and all four possible PAM sites.

Limitations for generating gene knockouts

Generating deletions of entire genes, or essential parts of them is the preferred way to generate a null allele. This approach will avoid complications arising from introducing frameshifts at the beginning of the ORF, as translation could reinitiate from later AUG or CUG start codons (Koushika et al. 1999). In addition, in some genes, the RNA has functions on its own, as shown for *oscar* RNA that forms a large RNP particle with Oscar and other RNA-binding proteins at the posterior pole of a *Drosophila* oocyte (Hachet and Ephrussi 2004; Haussmann et al. 2011). Thus, deletion of the entire gene region will discover such additional functions harbored in gene transcripts.

When generating deletions of entire genes, we discovered that the deleted DNA fragment was inserted into the genome and transcribed, resulting in the expression of protein that rescued the flightless phenotype in *Ythdc1* deletion allele. Although the mechanism for the generation of these new inserts is not known, it seemed not to have led to complex chromosomal aberrations, because the inserts could be removed by standard recombination and/or exchange of chromosomes. Retro-transposition has been observed in the *elav* gene, which led to the loss of all introns in *Drosophila* (Samson 2008). Likewise, holometabolous insects generally have three *elav* genes, but honeybees have only one *elav* gene. The honeybee *elav* gene, however, carries features of the other two genes present in *Drosophila*. Hence, *elav* in honeybees could have collapsed in an ancestor from three to one gene by some form of recombination to include parts specific to the other two *elav* genes (Ustaoglu et al. 2021). Likewise, the *amyloid-β precursor protein* (*APP*) gene displays copy number variation in the human brain, which is increased by retro-transposition in sporadic forms of Alzheimer's disease suggesting evolutionary conserved mechanisms for reinsertion of genomic information (Lee et al. 2018). Reinsertion of fragments cut out by the CRISPR–Cas9 system seems not to be specific to *Drosophila* as it has also been observed in human cells (Geng et al. 2022). Since intron loss and gain occur during evolution, the mechanism underlying the insertion of DNA fragments from sgRNA-induced deletions might be responsible for these changes. Thus, when generating sgRNA-induced gene deletions, it is essential to test for the absence of any transcripts by RT-PCR, but also to validate a deletion at the DNA level by PCR using flanking primers or Southern blots.

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

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

For a more reliable way to generate gene knockouts in *Drosophila*, we have now developed a *PBac w+* marker that can be inserted when generating a deletion (Dix 2022). In any case, however, a marked locus will allow for rigorous cleaning of the genetic background.

In essence, we have established the rules for designing highly efficient sgRNAs and established a methodology to efficiently generate gene deletions. These findings have implications for other RNA-based methodologies including prime editing (Anzalone et al. 2019; Bosch et al. 2021).

Methods

sgRNA/Cas9 directed DNA cleavage

DNA templates for in vitro transcription were reconstituted from synthetic oligonucleotides. As only the T7 promoter needs to be double-stranded for in vitro transcription, a T7 promoter oligonucleotide (CCTGGCTAATACGACTCACTATAG) was annealed to an antisense Ultramer (IDT DNA) encoding the entire sgRNA in addition to the T7 promoter. Alternatively, a 60 nt T7 promoter oligonucleotide with a partial sgRNA was annealed to an antisense oligonucleotide encoding the tracrRNA (AAAAAAAGCACCGACT

CGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAA CTTGCTATTTCTAGCTCTAAAAC) for 15 min at 40°C (2 µM) and made double-stranded by extension with the Klenow fragment of DNA polymerase I according to the manufacturer's instructions (NEB). Klenow was then heat-inactivated 10 min at 85°C and oligonucleotides were desalted with a G-50 Autoseq Sephadex spin column (GE) before using them for in vitro transcription.

Then, sgRNAs were generated by in vitro transcription with T7 polymerase (T7 MEGAscript, Ambion) from synthetic oligonucleotides (0.2 µM) and trace-labeled with ^{32}P alpha-ATP (800 Ci/mmol, 12.5 µM, PerkinElmer) in a 20 µL reaction according to the manufacturer's instructions. After DNase I digestion, free nucleotides were removed with a G-50 Probequant Sephadex spin column (GE). Then, sgRNAs were heated for 2 min to 95°C and left at room temperature to adopt folding. Then, sgRNAs were quantified by scintillation counting and analyzed on 8%–20% denaturing polyacrylamide gels as described (Dix et al. 2022).

For synthetic substrate DNAs the sense oligonucleotide (1 µM, sgRNA flanking sequences are: TCGAGCATTATATGAAC-sgRNA-GGGTATTGGGGAATTCATTATGC) was labeled with ^{32}P gamma-ATP (6000 Ci/mmol, 25 µM, PerkinElmer) with PNK (NEB). After heat inactivation of PNK for 2 min at 95°C, sense and antisense (antisense sgRNA flanking sequences are: GGCCGCATAATGAATTCCCCAATACCC-as sgRNA-GTTCATAT AATGC) oligonucleotides were annealed by letting them cool down to room temperature and used in sgRNA–Cas9 cleavage assays. For plasmid sgRNA–Cas9 cleavage assays, these annealed oligonucleotides were cloned into a modified *pBS SK+* using an Xho I and Not I cut vector to assay sgRNA/Cas9 activity.

For sgRNA/Cas9 cleavage assays, DNA/sgRNA/Cas9 ratios of 1/10/10 were used in a 10 µL reaction using the buffer supplied (NEB) and DEPC-treated water (Haussmann et al. 2019). Typically Cas9 (100 nM final) was incubated with sgRNA (100 nM) for 10 min at 25°C before adding oligonucleotides (10 nM final) or plasmid DNA (10 nM, corresponds to ~25 ng/µL final concentration of a 3 kb plasmid). Plasmids were linearized after Cas9 digestion by first heat inactivating Cas9 for 2 min at 95°C, and then adding 10 µL of a restriction enzyme (5 U) in NEB buffer 3. Adding a restriction enzyme together with Cas9 inhibited DNA scission by Cas9. Cleavage of oligonucleotides was analyzed on 8% denaturing polyacrylamide gels and plasmid DNA was analyzed on ethidium bromide-stained agarose gels.

RNA secondary structure was analyzed with RNAfold at http://rna.tbi.univie.ac.at (Gruber et al. 2008) using the following tracrRNA sequence: GUUUUAGAGCUAGAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAG UCGGUGCUUUUUU.

Criteria for optimal sgRNA design and Cas9 structural analysis

Criteria to select sgRNAs that maintain the structure required for efficient Cas9 DNA scission were implemented in the server accessible at https://platinum-crispr.bham.ac.uk/predict.pl and are as follows. The first nucleotide of the sgRNA needs to be a G for efficient transcription initiation by RNA Pol III (Paule and White 2000). For T7-mediated in vitro transcription, three Gs need to be added to sgRNAs of 23 nt. Disruption of the tetraloop or loops 2 and 3 structures, and the sequence U A/G G C/U A/G of nucleotides 16–20, which will result in a tetraloop bulge mimic, were classified as low-efficiency sgRNAs as these parts are recognized by Cas9 (Nishimasu et al. 2014). Similarly, a hairpin loop in the gRNA consisting of four or more base-pairing nucleotides, or base-pairing of nucleotides 17–20 of the gRNA (U/C U/C U U), or base-pairing of 8 nt within the seed region with looped-out nucleotides spaced by three base-pairing nucleotides (N11–N20), or a GC content below 15% (1 or 2 nt) or above 50% (11 or more nucleotides) were also considered low efficiency. Medium efficiency was assigned for gRNAs with a GC content of 15%–25% (3–5 nt) or 40%–50% (8–10 nt), or a low CG content in the seed region (<5 nt in nucleotides 11–20). In addition, two U's at position 19/20 of the gRNA reduce efficiency because of premature transcription termination (Gao et al. 2018; Graf et al. 2019). Further, we assigned a medium impact if both of the 2 nt A_{51} A_{52} and G_{62} U_{63} in the three-way junction of loop 1 were base-paired or 7 nt within the seed region (N11–N20) base-paired with looped-out nucleotides spaced by three base-pairing nucleotides. Thus, we deemed a sgRNA optimal to allow Cas9 to cleave DNA with high efficiency, if the GC content is 30%–35% (6–7 nt) and none of the above criteria applied.

Cas9/sgRNA structural complex analysis was done using Chimera as described (Dix et al. 2022).

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

Total RNA was extracted using TRI Reagent (Sigma-Aldrich) and reverse transcribed with Superscript II (Invitrogen) according to the manufacturer's instructions using an Oligo(dT) primer. PCR was done for 40 cycles with 1 µL of cDNA, with 1 µg of genomic DNA, or from a single fly after freezing and drying in 200 µL of isopropanol. Primers to detect the sgRNA/Cas9 induced deletion in *Ythdc1* were YT F1 (GCCGCTGTGACGCAGAATTTGTGTG) and YT R1 (GGCCGTGCATGTTGCGCATGTAGTCC), and in *Ythdf* were 64F1 (GCCGAGAAAGTGCACAAGGATACGGAG) and 64R1 (CAAGGAATGGCTGAAGCAGACTCCTTG). Primers to amplify parts of the body of the RNA also flanking an intron were for *Ythdc1* YT F2 (CCACGCTGCCGCAGAACGACGCCAATC) and YT R2 (GCGGCAGATCCAGTCAAGCTCGATGAC), and in *Ythdf* were 64F2 (GAGCTGCCTGTCGATTCCCAACTCGTG) and 64R2 (CCG CCCTCTTCGTGTCGCTCCTTGAAG). Primers to amplify parts of the *ewg* gene have been described elsewhere (Koushika et al. 1999).

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

To clone two sgRNAs expressed by U6 promoters, the "tracrRNA U6.3 promoter" fragment was amplified with left (AAGATATCC GGGTGAACTTC**G**N_{19}GTTTTAGAGCTAGAAATAGC) and right (GCTATTTCTAGCTCTAAAA**C**N_{19}CGACGTTAAATTGAAAATAGG) sgRNA primers from pUC 3GLA U6.1/3 sgRNA using Pwo polymerase (Roche) with initial 30 sec denaturation at 94°C followed by two cycles 94°C/30 sec, 49°C/40 sec, 72°C/45 sec, then two cycles 94°C/30 sec, 51°C/40 sec, 72°C/45 sec, and 22 cycles 94°C/30 sec, 56°C/40 sec, 72°C/45 sec using a slow ramp (1°C/min) for increasing from 56°C to 72°C. Transcription from the U6 promoter initiates with a G (bold, underlined). Although this G does not need to be present in the targeting sequence, it needs to be included for the folding of the sgRNA. The sequences for the sgRNAs in *Ythdc1* were gACAGGTATTCCCAAACTCAC and GACATGTAGCG TTCCCATGA, and for *Ythdf* were GTCCTGAAATACGAGCACAA and gATAACGAACATGTGGGATCT. The *pUC 3GLA U6.1 BbsI* vector was cut by BbsI and the "sgRNA1 tracrRNA U6.3 promoter sgRNA2" fragment was cloned by Gibson assembly according to the manufacturer's instructions (NEB). For sequencing, primer U6.1 Fseq (GCGCGTACGTCCTTCGCATCCTTATG) was used. The sequences for *pUC 3GLA HAi Dscam 3–5, pUC 3GLA U6.1 BbsI*, and the *pUC 3GLA U6.1/3 Ythdf sgRNA* have been deposited (MK908409, MK908408, and MK908407).

Drosophila genetics and phiC31 integrase-mediated transgenesis

All *Drosophila melanogaster* strains were reared at 25°C and 40%–60% humidity on standard cornmeal-agar food in 12:12 h light:

dark cycle as described (Haussmann et al. 2013). *CantonS* was used as a wild-type control. For the *Ythdc1*, the GFP-marked *Mi{MIC}YT521-B*^{MI02006} transposon insert and the *w+* marked *Df(3L)Exel6094* deficiency were used, and for *Ythdf*, the RFP-marked *PBac{SAstopDsRed}*^{LL04081} insert and the *w+* marked deficiency *Df(3R)ED6220* were used. For phiC31 mediated transformation, constructs were injected into *y*^{1} *w** *M{vas-int.Dm}ZH-2A; PBac{y+-attP-3B}VK00013* with the landing site inserted at 76A as previously described (Haussmann et al. 2013). Before insertion of GFP marked constructs, the GFP and RFP markers had been removed from the *y1 w** *M{vas-int.Dm}ZH-2A* landing site by Cre-mediated recombination (Bischof et al. 2007; Zaharieva et al. 2015).

Implementation of PlatinumCRISPr

PlatinumCRISPr is implemented as a Perl script-based webserver iteratively evaluating the rule set described in the main text. A guide is classified as "compromised" if any of the rules is violated. For analysis of a sgRNA consisting of the target complementary sequence (spacer) and the constant crRNA fused to the tracrRNA through an artificial loop is used whereby the first nucleotide of the 20 nt spacer sequence is a G, because a G is needed for transcription (Jinek et al. 2012; Cong et al. 2013). Folding of the sgRNA is computed using RNAfold (Version 2.4.17) and further processed using bpRNA for subsequent interpretation of the dot-bracket code describing the secondary structure by a custom-made Perl script. Notably, the sequence position is calculated from the 3′ end of the tracrRNA to allow for a variable length of the spacer (between 18 and 23 nt) for custom applications using synthesized RNA consisting of the spacer fused to the crRNA and hybridized to the tracrRNA.

PlatinumCRISPr classifies guides by a binary outcome and typically reports ∼70% of sgRNAs as "compromised." Accordingly, only the top 30% were analyzed for the distribution of their reported cleavage efficiency in a given data set for each scoring application (Supplemental Fig. S8). Statistical significance was calculated using a one-sided Wilcoxon signed-rank test. We are grateful to M. Haeussler for providing published guide sequencing and cleavage efficiency scores (Haeussler et al. 2016).

CRISPRon and DeepSpCas9 guide sequencing and cleavage efficiency scores were calculated using published web-interfaces (Wong et al. 2015; Xiang et al. 2021).

For CRISpick analysis, we downloaded precalculated on-target score data (options: KO screen, CRISPR mechanism SpyCas9, Hsu tracrRNA, and On Target Scoring Rule Set was RS3seq-Chen2013 + RS3target) which are available for Mouse, Fly, and Human (Sanson et al. 2018). DepMap inferred guide efficiency data have been downloaded from DepMap (DepMap 24Q2 Public; https://doi.org/10.25452/figshare.plus.25880521.v1). For the analysis of sgRNA efficiency in different cell lines (HeLa, HCT 116, and GBM), genes essential in all cells were selected and the calculated cleavage efficiencies were compared by Spearman's correlation (Hart et al. 2015).

For the analysis of off-target effects, guides from the TTISS data set with their number of genome-wide off-target sites were used (Schmid-Burgk et al. 2020; Chen et al. 2023). For each guide, the number of expected off-targets was computed using Cas-Offinder3 (selecting targets with no bulge, and a minimum of one and a maximum of five mismatches) (Hwang et al. 2021). The ratio of measured to expected numbers of off-target sites was then compared between guides predicted to cut by PlatinumCRISPr (Wilcoxon rank-sum test) and correlated (Spearman's correlation) to DeepSpCas9 and CHOPCHOP (Azimuth 2.0 scores) using the crisprScore R-package (Hoberecht et al. 2022).

Data access

All data used for sgRNA cleavage efficiency analysis to analyze data and the source code for PlatinumCRISPr have been deposited in GitHub (https://github.com/rolandA1234/PlatinumCRISPr) and are also provided as Supplemental Code. Sequences for plasmid vectors generated in this study have been submitted to NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/about/) with the following accession numbers: MK908409 (*pUC 3GLA HAi Dscam 3–5*), MK908408 (*pUC 3GLA U6.1 BbsI*), and MK908407 (*pUC 3GLA U6.1/3 YTHDF53 sgRNA*).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Bloomington and Kyoto stock centers for fly lines, A. Handler for the PiggyBac clone, BacPac for Bac clones and plasmids, the University of Cambridge Department of Genetics Fly Facility for embryo injections for generating transgenic lines, M. Haeusler for data sets and precalculated scores, Sung-Eun Yoon from the Korea Drosophila Resource Centre and Ying Tan from GenetiVision Corporation for sgRNA efficiency data, Yuan Tian for help with art work, and Pawel Grzechnik for comments on the manuscript. For this work, we acknowledge funding from the Biotechnology and Biological Science Research Council, the Leverhulme Trust, the MRC DTP to D.W.J.M., and a Genetics Society summer studentship to A.I.H.

Author contributions: I.U.H., T.C.D., and M.S. performed biochemistry and genetic experiments. V.D. constructed vectors. A.I.H. performed genetic experiments. R.A. performed bioinformatics analysis, programming, and server implementation. D.W.J.M. analyzed data. I.U.H., R.A., and M.S. conceived the project and wrote the manuscript with help from all authors.

References

Abadi S, Yan WX, Amar D, Mayrose I. 2017. A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput Biol* **13:** e1005807. doi:10.1371/journal.pcbi.1005807

Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513:** 569–573. doi:10.1038/nature13579

Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M. 2021. New twists in detecting mRNA modification dynamics. *Trends Biotechnol* **39:** 72–89. doi:10.1016/j.tibtech.2020.06.002

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576:** 149–157. doi:10.1038/s41586-019-1711-4

Arimbasseri AG, Maraia RJ. 2015. Mechanism of transcription termination by RNA polymerase III utilizes a non-template strand sequence-specific signal element. *Mol Cell* **58:** 1124–1132. doi:10.1016/j.molcel.2015.04.002

Balacco DL, Soller M. 2019. The m^{6}A writer: rise of a machine for growing tasks. *Biochemistry* **58:** 363–378. doi:10.1021/acs.biochem.8b01166

Bawankar P, Lence T, Paolantoni C, Haussmann IU, Kazlauskiene M, Jacob D, Heidelberger JB, Richter FM, Nallasivan MP, Morin V, et al. 2021. Hakai is required for stabilization of core components of the m^{6}A mRNA methylation machinery. *Nat Commun* **12:** 3778. doi:10.1038/s41467-021-23892-5

Bischof J, Maeda RK, Hediger M, Karch F, Basler K. 2007. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci* **104:** 3312–3317. doi:10.1073/pnas.0611511104

Boivin V, Deschamps-Francoeur G, Scott MS. 2018. Protein coding genes as hosts for noncoding RNA expression. *Semin Cell Dev Biol* **75:** 3–12. doi:10.1016/j.semcdb.2017.08.016

Bosch JA, Birchak G, Perrimon N. 2021. Precise genome engineering in *Drosophila* using prime editing. *Proc Natl Acad Sci* **118:** e2021996118. doi:10.1073/pnas.2021996118

Chari R, Mali P, Moosburner M, Church GM. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* **12:** 823–826. doi:10.1038/nmeth.3473

Chen Q, Chuai G, Zhang H, Tang J, Duan L, Guan H, Li W, Li W, Wen J, Zuo E et al. 2023. Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction fingerprints. *Nat Commun* **14:** 7521. doi:10.1038/s41467-023-42695-4

Chuai G, Ma H, Yan J, Chen M, Hong N, Xue D, Zhou C, Zhu C, Chen K, Duan B, et al. 2018. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* **19:** 80. doi:10.1186/s13059-018-1459-4

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819–823. doi:10.1126/science.1231143

Deveson IW, Hardwick SA, Mercer TR, Mattick JS. 2017. The dimensions, dynamics, and relevance of the mammalian noncoding transcriptome. *Trends Genet* **33:** 464–478. doi:10.1016/j.tig.2017.04.004

Dezi V, Ivanov C, Haussmann IU, Soller M. 2016. Nucleotide modifications in messenger RNA and their role in development and disease. *Biochem Soc Trans* **44:** 1385–1393. doi:10.1042/BST20160110

Dix TC. 2022. "Characterization of ELAV multimerization in neuron development and function." PhD thesis, University of Birmingham, Birmingham, West Midlands, UK.

Dix TC, Haussmann IU, Brivio S, Nallasivan MP, HadzHiev Y, Müller F, Müller B, Pettitt J, Soller M. 2022. CMTr mediated 2′-*O*-ribose methylation status of cap-adjacent nucleotides across animals. *RNA (New York, NY)* **28:** 1377–1390. doi:10.1261/rna.079317.122

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32:** 1262–1267. doi:10.1038/nbt.3026

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34:** 184–191. doi:10.1038/nbt.3437

Doudna JA. 2020. The promise and challenge of therapeutic genome editing. *Nature* **578:** 229–236. doi:10.1038/s41586-020-1978-5

El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, Kikhi K, Boezio GLM, Takacs CM, Lai SL, et al. 2019. Genetic compensation triggered by mutant mRNA degradation. *Nature* **568:** 193–197. doi:10.1038/s41586-019-1064-z

Farboud B, Meyer BJ. 2015. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* **199:** 959–971. doi:10.1534/genetics.115.175166

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31:** 822–826. doi:10.1038/nbt.2623

Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, Montague TG, Zimmerman S, Richter C, Schier AF. 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9:** e98186. doi:10.1371/journal.pone.0098186

Gandhi S, Haeussler M, Razy-Krajka F, Christiaen L, Stolfi A. 2017. Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona. *Dev Biol* **425:** 8–20. doi:10.1016/j.ydbio.2017.03.003

Gao Z, Herrera-Carrillo E, Berkhout B. 2018. Delineation of the exact transcription termination signal for type 3 polymerase III. *Mol Ther Nucleic Acids* **10:** 36–44. doi:10.1016/j.omtn.2017.11.006

Garcia-Doval C, Jinek M. 2017. Molecular architectures and mechanisms of class 2 CRISPR-associated nucleases. *Curr Opin Struct Biol* **47:** 157–166. doi:10.1016/j.sbi.2017.10.015

Geng K, Merino LG, Wedemann L, Martens A, Sobota M, Sanchez YP, Søndergaard JN, White RJ, Kutter C. 2022. Target-enriched nanopore sequencing and de novo assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells. *Genome Res* **32:** 1876–1891. doi:10.1101/gr.276901.122

Graf R, Li X, Chu VT, Rajewsky K. 2019. sgRNA sequence motifs blocking efficient CRISPR/Cas9-mediated gene editing. *Cell Rep* **26:** 1098–1103.e3. doi:10.1016/j.celrep.2019.01.024

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Res* **36:** W70–W74. doi:10.1093/nar/gkn188

Hachet O, Ephrussi A. 2004. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428:** 959–963. doi:10.1038/nature02521

Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, et al. 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17:** 148. doi:10.1186/s13059-016-1012-2

Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, et al. 2015. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163:** 1515–1526. doi:10.1016/j.cell.2015.11.015

Haussmann IU, Li M, Soller M. 2011. ELAV-mediated 3′-end processing of *ewg* transcripts is evolutionarily conserved despite sequence degeneration of the ELAV-binding site. *Genetics* **189:** 97–107. doi:10.1534/genetics.111.131383

Haussmann IU, Hemani Y, Wijesekera T, Dauwalder B, Soller M. 2013. Multiple pathways mediate the sex-peptide-regulated switch in female *Drosophila* reproductive behaviours. *Proc Biol Sci* **280:** 20131938. doi:10.1098/rspb.2013.1938

Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, Soller M. 2016. m^{6}A potentiates *Sxl* alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature* **540:** 301–304. doi:10.1038/nature20577

Haussmann IU, Ustaoglu P, Brauer U, Hemani Y, Dix TC, Soller M. 2019. Plasmid-based gap-repair recombineered transgenes reveal a central role for introns in mutually exclusive alternative splicing in Down Syndrome Cell Adhesion Molecule exon 4. *Nucleic Acids Res* **47:** 1389–1403. doi:10.1093/nar/gky1254

Haussmann IU, Wu Y, Nallasivan MP, Archer N, Bodi Z, Hebenstreit D, Waddell S, Fray R, Soller M. 2022. CMTr cap-adjacent 2′-*O*-ribose mRNA methyltransferases are required for reward learning and mRNA localization to synapses. *Nat Commun* **13:** 1209. doi:10.1038/s41467-022-28549-5

Hemani Y, Soller M. 2012. Mechanisms of *Drosophila* Dscam mutually exclusive splicing regulation. *Biochem Soc Trans* **40:** 804–809. doi:10.1042/BST20120060

Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E. 2018. The biology of CRISPR-Cas: backward and forward. *Cell* **172:** 1239–1259. doi:10.1016/j.cell.2017.11.032

Hoberecht L, Perampalam P, Lun A, Fortin JP. 2022. A comprehensive bioconductor ecosystem for the design of CRISPR guide RNAs across nucleases and technologies. *Nat Commun* **13:** 6568. doi:10.1038/s41467-022-34320-7

Housden BE, Valvezan AJ, Kelley C, Sopko R, Hu Y, Roesel C, Lin S, Buckner M, Tao R, Yilmazel B, et al. 2015. Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci Signal* **8:** rs9. doi:10.1126/scisignal.aab3729

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31:** 827–832. doi:10.1038/nbt.2647

Hwang GH, Kim JS, Bae S. 2021. Web-based CRISPR toolkits: Cas-OFFinder, Cas-designer, and Cas-analyzer. *Methods Mol Biol* **2162:** 23–33. doi:10.1007/978-1-0716-0687-2_2

Jensen KT, Fløe L, Petersen TS, Huang J, Xu F, Bolund L, Luo Y, Lin L. 2017. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett* **591:** 1892–1901. doi:10.1002/1873-3468.12707

Jiang F, Doudna JA. 2017. CRISPR-Cas9 Structures and mechanisms. *Annu Rev Biophys* **46:** 505–529. doi:10.1146/annurev-biophys-062215-010822

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:** 816–821. doi:10.1126/science.1225829

Kim HK, Kim Y, Lee S, Min S, Bae JY, Choi JW, Park J, Jung D, Yoon S, Kim HH. 2019. SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Sci Adv* **5:** eaax9249. doi:10.1126/sciadv.aax9249

Koushika SP, Soller M, DeSimone SM, Daub DM, White K. 1999. Differential and inefficient splicing of a broadly expressed *Drosophila* erect wing transcript results in tissue-specific enrichment of the vital EWG protein isoform. *Mol Cell Biol* **19:** 3998–4007. doi:10.1128/MCB.19.6.3998

Labuhn M, Adams FF, Ng M, Knoess S, Schambach A, Charpentier EM, Schwarzer A, Mateo JL, Klusmann JH, Heckl D. 2018. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res* **46:** 1375–1385. doi:10.1093/nar/gkx1268

Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. 2016. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44:** W272–W276. doi:10.1093/nar/gkw398

Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47:** W171–W174. doi:10.1093/nar/gkz365

Lee MH, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, Liu CS, Park C, Kennedy G, Long T, et al. 2018. Somatic APP gene

recombination in Alzheimer's disease and normal neurons. *Nature* **563:** 639–645. doi:10.1038/s41586-018-0718-6
Liu X, Homma A, Sayadi J, Yang S, Ohashi J, Takumi T. 2016. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep* **6:** 19675. doi:10.1038/srep19675
Ma Z, Zhu P, Shi H, Guo L, Zhang Q, Chen Y, Chen S, Zhang Z, Peng J, Chen J. 2019. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* **568:** 259–263. doi:10.1038/s41586-019-1057-y
Michlits G, Jude J, Hinterndorfer M, de Almeida M, Vainorius G, Hubmann M, Neumann T, Schleiffer A, Burkard TR, Fellner M, et al. 2020. Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nat Methods* **17:** 708–716. doi:10.1038/s41592-020-0850-8
Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ. 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* **12:** 982–988. doi:10.1038/nmeth.3543
Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156:** 935–949. doi:10.1016/j.cell.2014.02.001
Paule MR, White RJ. 2000. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* **28:** 1283–1298. doi:10.1093/nar/28.6.1283
Poh HX, Mirza AH, Pickering BF, Jaffrey SR. 2022. Alternative splicing of METTL3 explains apparently METTL3-independent m6A modifications in mRNA. *PLoS Biol* **20:** e3001683. doi:10.1371/journal.pbio.3001683
Port F, Chen HM, Lee T, Bullock SL. 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci* **111:** E2967–E2976. doi:10.1073/pnas.1405500111
Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S, Liu LP, Yang Z, Mao D, Sun L, et al. 2013. Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. *Proc Natl Acad Sci* **110:** 19012–19017. doi:10.1073/pnas.1318481110
Ren X, Yang Z, Xu J, Sun J, Mao D, Hu Y, Yang SJ, Qiao HH, Wang X, Hu Q, et al. 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell Rep* **9:** 1151–1162. doi:10.1016/j.celrep.2014.09.044
Riesenberg S, Kanis P, Macak D, Wollny D, Düsterhöft D, Kowalewski J, Helmbrecht N, Maricic T, Pääbo S. 2023. Efficient high-precision homology-directed repair-dependent genome editing by HDRobust. *Nat Methods* **20:** 1388–1399. doi:10.1038/s41592-023-01949-1
Roignant JY, Soller M. 2017. m^6A in mRNA: an ancient mechanism for fine-tuning gene expression. *Trends Genet* **33:** 380–390. doi:10.1016/j.tig.2017.04.003
Roth A, Vleurinck C, Netschitailo O, Bauer V, Otte M, Kaftanoglu O, Page RE, Beye M. 2019. A genetic switch for worker nutrition-mediated traits in honeybees. *PLoS Biol* **17:** e3000171. doi:10.1371/journal.pbio.3000171
Samson ML. 2008. Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins. *BMC Genomics* **9:** 392. doi:10.1186/1471-2164-9-392
Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F, et al. 2018. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9:** 5416. doi:10.1038/s41467-018-07901-8
Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F. 2020. Highly parallel profiling of Cas9 variant specificity. *Mol Cell* **78:** 794–800.e8. doi:10.1016/j.molcel.2020.02.023
Singh R, Kuscu C, Quinlan A, Qi Y, Adli M. 2015. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res* **43:** e118. doi:10.1093/nar/gkv575
Sledzinski P, Nowaczyk M, Olejniczak M. 2020. Computational tools and resources supporting CRISPR-Cas experiments. *Cells* **9:** 1288. doi:10.3390/cells9051288
Soller M. 2006. Pre-messenger RNA processing and its regulation: a genomic perspective. *Cell Mol Life Sci* **63:** 796–819. doi:10.1007/s00018-005-5391-x
Soller M, Haussmann IU, Hollmann M, Choffat Y, White K, Kubli E, Schäfer MA. 2006. Sex-peptide-regulated female sexual behavior requires a subset of ascending ventral nerve cord neurons. *Curr Biol* **16:** 1771–1782. doi:10.1016/j.cub.2006.07.055
Solomon DA, Stepto A, Au WH, Adachi Y, Diaper DC, Hall R, Rekhi A, Boudi A, Tziortzouda P, Lee YB, et al. 2018. A feedback loop between dipeptide-repeat protein, TDP-43 and karyopherin-α mediates C9orf72-related neurodegeneration. *Brain* **141:** 2908–2924. doi:10.1093/brain/awy241
Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. 2016. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* **7:** 11750. doi:10.1038/ncomms11750
Trivedi D, Cm V, Bisht K, Janardan V, Pandit A, Basak B, Shwetha H, Ramesh N, Raghu P. 2020. A genome engineering resource to uncover principles of cellular organization and tissue architecture by lipid signaling. *eLife* **9:** e55793. doi:10.7554/eLife.55793.sa2
Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J, Mathews DH, Kim J, et al. 2019. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* **10:** 4056. doi:10.1038/s41467-019-12028-5
Ustaoglu P, Gill JK, Doubovetzky N, Haussmann IU, Dix TC, Arnold R, Devaud JM, Soller M. 2021. Dynamically expressed single ELAV/Hu orthologue elavl2 of bees is required for learning and memory. *Commun Biol* **4:** 1234. doi:10.1038/s42003-021-02763-1
Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, et al. 2015. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res* **25:** 1030–1042. doi:10.1101/gr.186379.114
Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343:** 80–84. doi:10.1126/science.1246981
Wierson WA, Welker JM, Almeida MP, Mann CM, Webster DA, Torrie ME, Weiss TJ, Kambakam S, Vollbrecht MK, Lan M, et al. 2020. Efficient targeted integration directed by short homology in zebrafish and mammalian cells. *eLife* **9:** e53968. doi:10.7554/eLife.53968
Wong N, Liu W, Wang X. 2015. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* **16:** 218. doi:10.1186/s13059-015-0784-0
Xiang X, Corsi GI, Anthon C, Qu K, Pan X, Liang X, Han P, Dong Z, Liu L, Zhong J, et al. 2021. Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. *Nat Commun* **12:** 3238. doi:10.1038/s41467-021-23576-0
Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS, et al. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25:** 1147–1157. doi:10.1101/gr.191452.115
Zaharieva E, Haussmann IU, Bräuer U, Soller M. 2015. Concentration and localization of co-expressed ELAV/Hu proteins control specificity of mRNA processing. *Mol Cell Biol* **35:** 3104–3115. doi:10.1128/MCB.00473-15
Zhang D, Hurst T, Duan D, Chen SJ. 2019. Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design. *Proc Natl Acad Sci* **116:** 8693–8698. doi:10.1073/pnas.1820523116

Received April 17, 2024; accepted in revised form October 7, 2024.