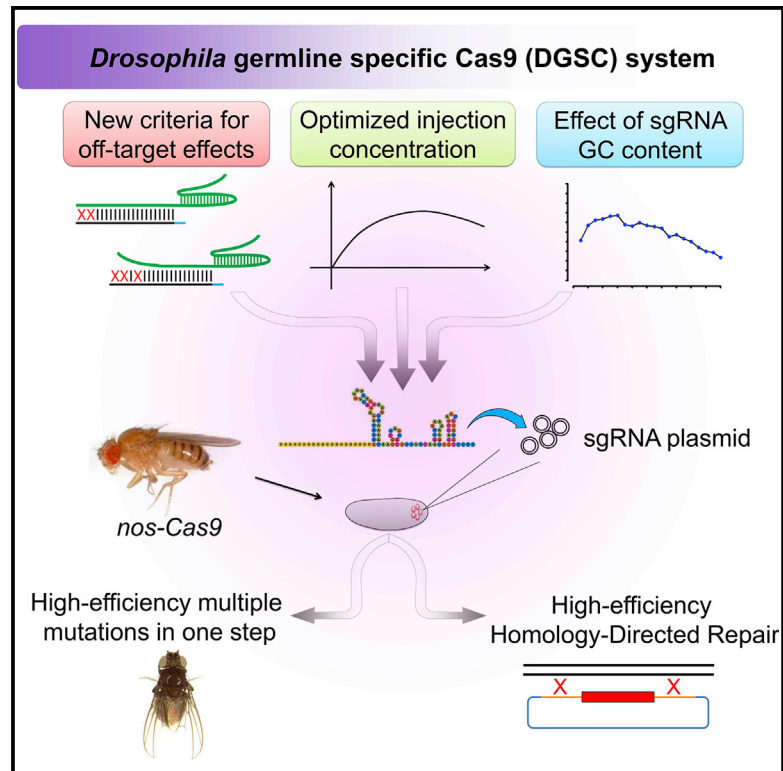


# Cell Reports

## Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in *Drosophila*

### Graphical Abstract



### Authors

Xingjie Ren, Zhihao Yang, ..., Jin Billy Li, Jian-Quan Ni

### Correspondence

jin.billy.li@stanford.edu (J.B.L.),  
nijq@mail.tsinghua.edu.cn (J.-Q.N.)

### In Brief

Ren et al. examine sgRNA properties that enhance efficiency of CRISPR/Cas9 in *Drosophila*. They demonstrate that such optimization allows more efficient mutagenesis of multiple genes and of homozygous lethal genes through homology-directed repair.

### Highlights

sgRNAs have limited off-target effects at loci with three or more mismatches

Efficiency correlates with the sgRNA GC content 6 nt adjacent to PAMs

Enhanced efficiency allows for the mutation of four genes together at high rates

Homology-directed repair rates are significantly higher than previously reported



# Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in *Drosophila*

Xingjie Ren,<sup>1,9</sup> Zhihao Yang,<sup>1,9</sup> Jiang Xu,<sup>1,2,3,4,9</sup> Jin Sun,<sup>1,9</sup> Decai Mao,<sup>1,5,9</sup> Yanhui Hu,<sup>6</sup> Su-Juan Yang,<sup>1</sup> Huan-Huan Qiao,<sup>1</sup> Xia Wang,<sup>1</sup> Qun Hu,<sup>2</sup> Patricia Deng,<sup>7</sup> Lu-Ping Liu,<sup>1,2</sup> Jun-Yuan Ji,<sup>8</sup> Jin Billy Li,<sup>7,\*</sup> and Jian-Qun Ni<sup>1,\*</sup>

<sup>1</sup>Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing 100084, China

<sup>2</sup>Tsinghua Fly Center, Tsinghua University, Beijing 100084, China

<sup>3</sup>School of Basic Medical Sciences, Wuhan University, Wuhan 430071, China

<sup>4</sup>College of Bioengineering, Hubei University of Technology, Wuhan 430068, China

<sup>5</sup>Sichuan Academy of Grassland Science, Chengdu 611731, China

<sup>6</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

<sup>7</sup>Department of Genetics, Stanford University, Stanford, CA 94305, USA

<sup>8</sup>Department of Molecular and Cellular Medicine, College of Medicine, Texas A&M Health Science Center, College Station, TX 77843, USA

<sup>9</sup>Co-first author

\*Correspondence: [jin.billy.li@stanford.edu](mailto:jin.billy.li@stanford.edu) (J.B.L.), [njq@mail.tsinghua.edu.cn](mailto:njq@mail.tsinghua.edu.cn) (J.-Q.N.)

<http://dx.doi.org/10.1016/j.celrep.2014.09.044>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

## SUMMARY

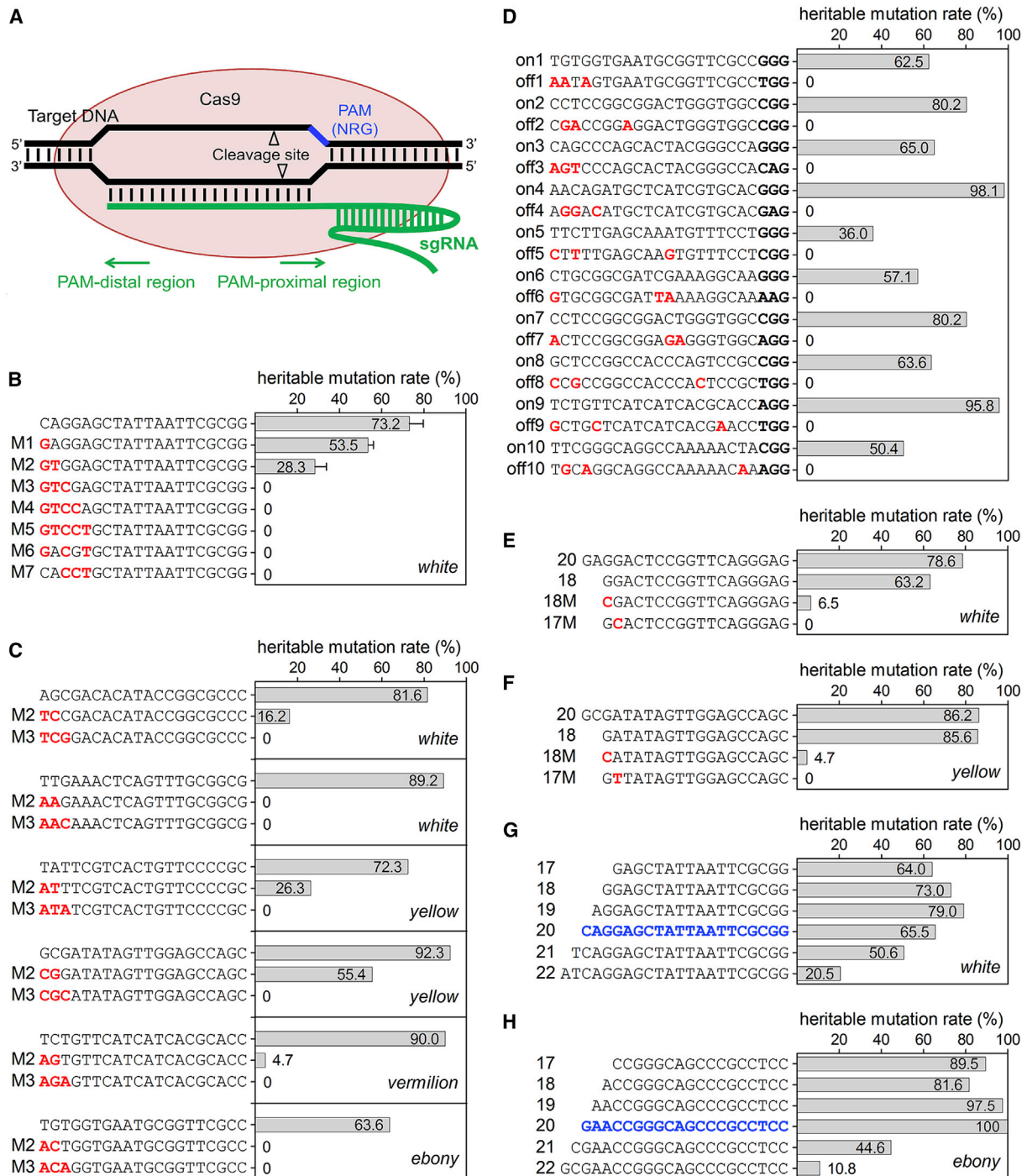
The CRISPR/Cas9 system has recently emerged as a powerful tool for functional genomic studies in *Drosophila melanogaster*. However, single-guide RNA (sgRNA) parameters affecting the specificity and efficiency of the system in flies are still not clear. Here, we found that off-target effects did not occur in regions of genomic DNA with three or more nucleotide mismatches to sgRNAs. Importantly, we document for a strong positive correlation between mutagenesis efficiency and sgRNA GC content of the six protospacer-adjacent motif-proximal nucleotides (PAMPNs). Furthermore, by injecting well-designed sgRNA plasmids at the optimal concentration we determined, we could efficiently generate mutations in four genes in one step. Finally, we generated null alleles of *HP1a* using optimized parameters through homology-directed repair and achieved an overall mutagenesis rate significantly higher than previously reported. Our work demonstrates a comprehensive optimization of sgRNA and promises to vastly simplify CRISPR/Cas9 experiments in *Drosophila*.

## INTRODUCTION

The past few years have witnessed a remarkable development of sequence-specific DNA endonuclease technologies in model organisms for precise genome editing, which holds the promise to greatly improve our understanding of developmental biology and diseases (Rong and Golic, 2000; Bibikova et al., 2002; Gong and Golic, 2003; Gao et al., 2008; Huang et al., 2009; Liu et al., 2012; Bassett et al., 2013; Chang et al., 2013; Cong et al., 2013; Friedland et al., 2013; Gaj et al., 2013; Gratz et al., 2013; Jiang et al.,

2013; Kondo and Ueda, 2013; Mali et al., 2013b; Ren et al., 2013; Wang et al., 2013; Yu et al., 2013). Considering cost, design, and efficiency, the CRISPR/Cas9 system has a number of advantages over zinc-finger or TALE protein-fused nucleases (Mali et al., 2013b). In this system, the *Streptococcus pyogenes* Cas9 contains two nuclease active sites, and the chimeric single-guide RNA (sgRNA) consists of a CRISPR RNA (crRNA) and a transactivating crRNA (tracrRNA) module (Jinek et al., 2012). The tracrRNA module is required for Cas9 nuclease activity, and the crRNA module guides the Cas9 protein to cleave double-stranded genomic DNA with sequence specificity provided by base pairing between the 20 nt targeting sequence and the targeted DNA (Figure 1A). The 3 nt (NGG for the *S. pyogenes* Cas9 system) protospacer-adjacent motif (PAM) in the genomic DNA is not included in the sgRNA sequence, but it is required for sgRNA targeting. Despite the wide utility of the powerful CRISPR/Cas9 system, its specificity and efficiency are two common concerns in the user community.

Studies in mammalian cells showed that the number, position, and distribution of nucleotide mismatches between the sgRNAs and the genomic DNA are major parameters affecting off-target effects and mutagenesis efficiency. These studies either introduced sgRNAs with mismatches to transgenes or endogenous loci, evaluated the binding patterns of nuclease-deficient Cas9, or directly tested mutagenesis rates at potential off-targets (Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013; Kuscu et al., 2014; Wu et al., 2014). To reduce off-target effects, a general design principle of avoiding sgRNAs with three or fewer nucleotide mismatches to potential off-targets has been suggested (Hsu et al., 2013; Mali et al., 2013a), but off-target mutation had been detected at genomic DNA regions with up to five mismatches to the sgRNA (Fu et al., 2013). Of the 20 nt targeting sequence of sgRNAs, mismatches in the “seed region,” the 5 to 12 nt closest to the PAM, have the largest impact on mutagenesis efficiency. Mismatches in the PAM-distal nucleotides generally have less effect on



**Figure 1. The Criteria for Off-Target Effects in the DGSC System**

(A) Schematic showing the Cas9/sgrRNA system. Cas9 is shown as the red background. The sgRNA is shown in green. The 3 nt PAM sequence (NGG or NAG) in the DNA is shown in blue. The PAM-distal and PAM-proximal ends are labeled by the green arrows.

(B) Mutagenesis efficiency of sgRNAs with mismatches (red) to a target of the *white* locus. Results are from three independent experiments, and the error bar shows SEM.

(C) Mutagenesis efficiency of sgRNAs with mismatches (red) to targets of different loci.

(D) Off-target mutagenesis efficiency of ten sgRNAs (numbered 1 to 10), compared with the on-target efficiency. The potential off-targets have 3 nt mismatches (red) to the sgRNAs. PAMs are indicated in bold text.

(legend continued on next page)

disrupting the sgRNA-target DNA hybrid than those in the PAM-proximal nucleotides (PAMPNs) (Jinek et al., 2012; Cong et al., 2013; Fu et al., 2013; Mali et al., 2013b; Pattanayak et al., 2013; Wu et al., 2014). It has been shown that the expression levels of Cas9 and sgRNA together affect mutagenesis efficiency and off-target effects (Bassett et al., 2013; Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). However, most cell-line studies did not evaluate Cas9 and sgRNA levels independently, and an in vivo study showed that sgRNA parameters are the major factors affecting specificity and efficiency within a certain Cas9 expression range (Wang et al., 2013). In addition, truncated sgRNAs with 18 nt targeting sequences have been shown to reduce off-target effects while retaining similar efficiency (Fu et al., 2014).

We and others have recently developed the CRISPR/Cas9 system in *Drosophila* (Bassett et al., 2013; Gratz et al., 2013, 2014; Kondo and Ueda, 2013; Ren et al., 2013; Yu et al., 2013, 2014; Port et al., 2014; Xue et al., 2014). However, there have been no systematic studies evaluating the efficiency and specificity in this system. Based on specificity studies in mammalian cells that reported off-target effects at genomic regions with five mismatches (Fu et al., 2013; Kuscu et al., 2014; Wu et al., 2014), 70% of the potentially usable sgRNAs could have off-target effects if five mismatches were tolerated in *Drosophila* (Ren et al., 2013), which greatly decreases the flexibility of the CRISPR/Cas9 system.

The efficiency of mutation through the nonhomologous end-joining (NHEJ) pathway after CRISPR/Cas9-induced double-strand breaks (DSBs) is high, especially with flies carrying transgenic Cas9, such as the *Drosophila* germline-specific Cas9 (DGSC) system we developed (Ren et al., 2013). However, with different sgRNAs targeting the same gene, the efficiency can vary ~10-fold by our estimation (our unpublished data). On the other hand, CRISPR/Cas9-mediated homology-directed repair (HDR) usually has an efficiency of less than 5.0% and rarely reaches 10% for homozygous viable genes (Baena-Lopez et al., 2013; Gratz et al., 2014; Xue et al., 2014; Yu et al., 2014). The HDR recovery rates through CRISPR/Cas9 systems for recessive lethal alleles have not been reported in *Drosophila*, but they are very likely to be even lower when generating loss-of-function alleles, due to the cellular lethality of homozygous mutations.

Therefore, a systematic investigation of parameters affecting the specificity and efficiency of the CRISPR/Cas9 system in flies is highly desired. In this study, we first systematically evaluated the parameters that affect the specificity and efficiency of the CRISPR/Cas9 system in *Drosophila*. We demonstrate that off-target effects do not occur in regions of genomic DNA with three or more nucleotide mismatches to the 104 sgRNAs (20 nt) that we evaluated and that the mutagenesis efficiencies of sgRNAs with a targeting sequence from 17 to 20 nt are similar, while reducing the length of sgRNA to 18 nt can further limit off-target effects. In addition, we optimize parameters for microinjection of

sgRNA in *Drosophila* embryos, based on our assays for efficiency and fertile G0 rate. Furthermore, we report a strong, positive correlation between mutagenesis efficiency and sgRNA GC content of the six PAMPNs. With these optimized parameters, we then demonstrate the potential for high-efficiency sgRNA target selection by generating mutations in up to four genes in one injection. Furthermore, we successfully carried out CRISPR/Cas9-mediated HDR to generate a null allele of *HP1a*, a homozygous lethal gene, with a high mutagenesis rate (12%).

## RESULTS

### New Criteria for Off-Target Effects

We employed two strategies to better understand the off-target effects of CRISPR-mediated mutagenesis in *Drosophila*. First, we made artificial off-targets by altering sgRNAs to contain a variety of mismatches to their targets. Second, we examined sgRNAs with natural potential off-targets. For the first strategy, we generated a series of derived sgRNAs targeting *white* (*w*) with up to five mismatched nucleotides to an intended target and evaluated the mutagenesis efficiency of these sgRNAs in the DGSC system. The original intact sgRNA efficiently triggered mutagenesis, while those with one or two mismatches in the most PAM-distal region significantly reduced the efficiency. sgRNAs with three or more mismatches, even most distal to PAMs, could not generate heritable mutations (>3,000 F1 screened) (Figure 1B). Next, we examined an sgRNA with three separated mismatches and another one with three consecutive mismatches but closer to the PAM, and we again observed no mutants (Figure 1B).

To test whether this phenomenon is common in *Drosophila*, we picked another six highly efficient sgRNAs (average mutagenesis efficiency >80%) targeting the coding region of the *yellow* (*y*), *w*, *vermillion* (*v*), and *ebony* (*e*) genes and designed corresponding sgRNAs with either two or three mismatches in the PAM-distal nucleotides. These four recessive genes have visible phenotypes when homozygous for the null alleles (Figure S1A), thus allowing us to assess germline mutation rate by crossing the G0 injected flies to the corresponding null mutants. We tested the heritable mutagenesis rate of these mismatched sgRNAs, together with the original intact sgRNAs. Consistent with the above results, we observed that four out of six sgRNAs with two mismatches generated heritable mutations at a much lower rate; the other two had mutation rates of zero. No mutants were recovered using sgRNAs with three mismatches (Figure 1C). To test if the positions of the mismatches affect efficiency, we picked two sgRNAs each with two mismatches and introduced a third mutation (transition and transversion) at each of the other 18 nucleotides. We found that these 72 sgRNAs we tested were unable to generate heritable mutants (Tables 1 and S1). These results suggest that three mismatches between sgRNAs and targets are not well tolerated in generating DSBs in *Drosophila*.

(E and F) Mutagenesis efficiency of sgRNAs with 20 nt targeting sequences, those with 18 nt, and those with 18 nt sequences and one-base mismatches (red). (G and H) Mutagenesis efficiency of sgRNAs with targeting sequences ranging from 17 to 22 nt. The regular 20 nt sgRNAs are shown in blue. Each row in (A), (B), and (E)–(H) represents an sgRNA sequence, and its mutagenesis rate. Each row in (C) and (D) represents a site of the fly genome and the mutagenesis rate at that site. See also Figure S1.

**Table 1. The Heritable Mutation Rates of 72 Derived sgRNAs with Three Mismatches to the Original sgRNAs**

sgRNA Name	sgRNA Sequence <sup>a</sup>	Heritable Mutation Rate, % (n) <sup>b</sup>	sgRNA Name	sgRNA Sequence <sup>a</sup>	Heritable Mutation Rate, % (n) <sup>b</sup>
yellow-sgRNA-1:TATTCGTCACTGTTCCCCGC			yellow-sgRNA-2:CGGATATAGTTGGAGCCAGC		
1-18TA	<u>A</u> TATCGTCACTGTTCCCCGC	0 (0/823)	2-18GC	<u>C</u> GCATATAGTTGGAGCCAGC	0 (0/663)
1-17TA	<u>A</u> TT <u>A</u> CGTCACTGTTCCCCGC	0 (0/913)	2-17AT	<u>C</u> GG <u>T</u> TATAGTTGGAGCCAGC	0 (0/594)
1-16CG	<u>A</u> TTT <u>G</u> GTCACTGTTCCCCGC	0 (0/559)	2-16TA	<u>C</u> GGAAATAGTTGGAGCCAGC	0 (0/692)
1-15GC	<u>A</u> TTT <u>C</u> CTCACTGTTCCCCGC	0 (0/472)	2-15AT	<u>C</u> GGAT <u>T</u> TAGTTGGAGCCAGC	0 (0/684)
1-14TA	<u>A</u> TTTC <u>G</u> ACTGTTCCCCGC	0 (0/584)	2-14TA	<u>C</u> GGATAAAGTTGGAGCCAGC	0 (0/634)
1-13CG	<u>A</u> TTTC <u>G</u> TACTGTTCCCCGC	0 (0/563)	2-13AT	<u>C</u> GGATAT <u>T</u> GTTGGAGCCAGC	0 (0/479)
1-12AT	<u>A</u> TTTC <u>G</u> T <u>C</u> TGTTCCCCGC	0 (0/510)	2-12GC	<u>C</u> GGATATA <u>C</u> TGGAGCCAGC	0 (0/683)
1-11CG	<u>A</u> TTTC <u>G</u> TCA <u>G</u> TGTTCCCCGC	0 (0/665)	2-11TA	<u>C</u> GGATATAGATGGAGCCAGC	0 (0/621)
1-10TA	<u>A</u> TTTC <u>G</u> TCA <u>C</u> AGTTCCCCGC	0 (0/583)	2-10TA	<u>C</u> GGATATAGTAGGAGCCAGC	0 (0/465)
1-9GC	<u>A</u> TTTC <u>G</u> TCACT <u>T</u> TCCCCGC	0 (0/654)	2-9GC	<u>C</u> GGATATAGTT <u>C</u> GAGCCAGC	0 (0/464)
1-8TA	<u>A</u> TTTC <u>G</u> TCACTGATCCCCGC	0 (0/751)	2-8GC	<u>C</u> GGATATAGTTG <u>C</u> AGCCAGC	0 (0/594)
1-7TA	<u>A</u> TTTC <u>G</u> TCACTGT <u>A</u> CCCCGC	0 (0/457)	2-7AT	<u>C</u> GGATATAGTTGG <u>T</u> GCCAGC	0 (0/820)
1-6CG	<u>A</u> TTTC <u>G</u> TCACTGTT <u>G</u> CCCCGC	0 (0/835)	2-6GC	<u>C</u> GGATATAGTTGGAG <u>C</u> CCAGC	0 (0/649)
1-5CG	<u>A</u> TTTC <u>G</u> TCACTGTT <u>C</u> GCCGC	0 (0/719)	2-5CG	<u>C</u> GGATATAGTTGGAG <u>G</u> CAGC	0 (0/788)
1-4CG	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>G</u> CGC	0 (0/521)	2-4CG	<u>C</u> GGATATAGTTGGAG <u>C</u> GAGC	0 (0/580)
1-3CG	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> GCG	0 (0/489)	2-3AT	<u>C</u> GGATATAGTTGGAG <u>C</u> CTGC	0 (0/489)
1-2GC	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> CC	0 (0/507)	2-2GC	<u>C</u> GGATATAGTTGGAG <u>C</u> CCACC	0 (0/801)
1-1CG	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> CGG	0 (0/755)	2-1CG	<u>C</u> GGATATAGTTGGAG <u>C</u> CCAGG	0 (0/523)
1-18TC	<u>A</u> TCTCGTCACTGTTCCCCGC	0 (0/710)	2-18GA	<u>C</u> GAATATAGTTGGAGCCAGC	0 (0/829)
1-17TC	<u>A</u> TT <u>C</u> CGTCACTGTTCCCCGC	0 (0/547)	2-17AG	<u>C</u> GGTATAGTTGGAGCCAGC	0 (0/660)
1-16CT	<u>A</u> TTT <u>T</u> GTCACTGTTCCCCGC	0 (0/557)	2-16TC	<u>C</u> GGACATAGTTGGAGCCAGC	0 (0/855)
1-15GA	<u>A</u> TTTCATCACTGTTCCCCGC	0 (0/693)	2-15AG	<u>C</u> GGATGATAGTTGGAGCCAGC	0 (0/773)
1-14TC	<u>A</u> TTTC <u>G</u> CACTGTTCCCCGC	0 (0/876)	2-14TC	<u>C</u> GGATACAGTTGGAGCCAGC	0 (0/839)
1-13CT	<u>A</u> TTTC <u>G</u> T <u>I</u> ACTGTTCCCCGC	0 (0/584)	2-13AG	<u>C</u> GGATATGTTGGAGCCAGC	0 (0/605)
1-12AG	<u>A</u> TTTC <u>G</u> T <u>C</u> GCTGTTCCCCGC	0 (0/620)	2-12GA	<u>C</u> GGATATA <u>A</u> TGGAGCCAGC	0 (0/509)
1-11CT	<u>A</u> TTTC <u>G</u> TCA <u>T</u> TGTTCCCCGC	0 (0/726)	2-11TC	<u>C</u> GGATATAG <u>C</u> TGGAGCCAGC	0 (0/809)
1-10TC	<u>A</u> TTTC <u>G</u> TCA <u>C</u> CGTTCCCCGC	0 (0/482)	2-10TC	<u>C</u> GGATATAGT <u>C</u> GGAGCCAGC	0 (0/558)
1-9GA	<u>A</u> TTTC <u>G</u> TCACTA <u>T</u> TCCCCGC	0 (0/799)	2-9GA	<u>C</u> GGATATAGTTA <u>G</u> AGCCAGC	0 (0/578)
1-8TC	<u>A</u> TTTC <u>G</u> TCACTG <u>C</u> TCCCCGC	0 (0/682)	2-8GA	<u>C</u> GGATATAGTTGA <u>A</u> GCCAGC	0 (0/582)
1-7TC	<u>A</u> TTTC <u>G</u> TCACTGT <u>C</u> CCCCGC	0 (0/791)	2-7AG	<u>C</u> GGATATAGTTGG <u>G</u> GCCAGC	0 (0/715)
1-6CT	<u>A</u> TTTC <u>G</u> TCACTGTT <u>T</u> CCCCGC	0 (0/734)	2-6GA	<u>C</u> GGATATAGTTGGAA <u>C</u> CCAGC	0 (0/846)
1-5CT	<u>A</u> TTTC <u>G</u> TCACTGTT <u>C</u> TCCGC	0 (0/832)	2-5CT	<u>C</u> GGATATAGTTGGAG <u>T</u> CAGC	0 (0/543)
1-4CT	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>T</u> CGC	0 (0/817)	2-4CT	<u>C</u> GGATATAGTTGGAG <u>C</u> TAGC	0 (0/528)
1-3CT	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> TGC	0 (0/779)	2-3AG	<u>C</u> GGATATAGTTGGAG <u>C</u> CGGC	0 (0/607)
1-2GA	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> AC	0 (0/581)	2-2GA	<u>C</u> GGATATAGTTGGAG <u>C</u> CAAC	0 (0/602)
1-1CT	<u>A</u> TTTC <u>G</u> TCACTGTTCC <u>C</u> CGT	0 (0/662)	2-1CT	<u>C</u> GGATATAGTTGGAG <u>C</u> CCAGT	0 (0/638)

See also Table S1.

<sup>a</sup>The mismatched nucleotides to the original sgRNA were underlined.

<sup>b</sup>The heritable mutation rate was calculated as the number of mutant F1s divided by the number of all F1s observed.

To further test this observation, we employed our second strategy for examining off-targets. We picked ten high-efficiency sgRNAs that each have a potential off-target site with three mismatches in the *Drosophila* genome. Previous reports (Hsu et al., 2013; Jiang et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013) and our results (Table S2) showed that certain sgRNAs targeting genomic sequences with NAG PAMs could generate low-level mutations, which indicated that off-target sites with NAG PAMs should not be excluded.

Therefore, when sequencing the potential off-target genomic sites of F1 mutant flies generated by these ten sgRNAs, we included sites with NAG PAMs in addition to those with NGG PAMs (Figure 1D). We did not observe off-target mutations in the fly genome for any of these ten sgRNAs (Figure 1D). Furthermore, we evaluated the mutation rates at potential off-target sites with two mismatches and one insertion/deletion to 13 sgRNAs (Lin et al., 2014) and again observed no off-target effects (Figure S1B).

Taken together, the 81 sgRNAs (5 from Figure 1B, 6 from Figure 1C, and 70 from Table 1) we tested that have three or more mismatches, consecutive or separate, do not generate mutants, even though the seven sgRNAs they are derived from are very efficient (Figures 1B and 1C). In addition, the 23 high-efficiency sgRNAs do not have off-target effects at genomic regions with three mismatches (Figures 1D and S1B). Therefore, when designing sgRNAs in *Drosophila*, one can significantly reduce or avoid off-target effects by avoiding DNA sequences with homologous genomic regions of two or fewer mismatches. Using this criterion, about 86.9% (6,575,357/7,566,724) sgRNAs targeting the fly genome have low or no off-target effects.

### sgRNAs with Targeting Sequences of 18 nt Are Superior to Those of 20 nt

A previous report shows that removing the two nucleotides most distal to the PAM from the sgRNA does not interfere with efficiency significantly (Fu et al., 2014), but mutagenesis efficiency at genomic sites with two contiguous mismatches in these nucleotides is compromised in our assays (Figures 1B and 1C). Thus, although the two PAM-distal nucleotides are the least important for an sgRNA in *Drosophila*, they seem to be important enough to have an impact on mutagenesis rates. We wondered whether this discrepancy reflected a difference of the CRISPR/Cas9 system activities between in mammalian cell lines and in *Drosophila* or showed a difference between having mismatched base pairs and having no pairings for the sgRNA and DNA. To directly test the effects of removing the two most PAM-distal nucleotides, two sgRNAs with 18 nt targeting *w* and *y* were constructed and examined. Indeed, we found that the mutagenesis efficiency of the 18 nt sgRNAs is comparable to that of the 20 nt ones (Figures 1E and 1F). In addition, one mismatch, but not three, abolished the ability of 18 nt sgRNAs to generate heritable mutations, unless the single mismatch was in the most PAM-distal pair (Figures 1E and 1F).

We further constructed a series of sgRNAs with targeting sequences ranging from 17 to 22 nt, and the results showed that the mutagenesis efficiency of sgRNAs with targeting sequences from 17 to 20 nt were at a similar level. Surprisingly, sgRNAs with 21 and 22 nt targeting sequences were far less effective than 20 nt ones (Figures 1G and 1H). In sum, the sgRNAs of 18 nt targets not only maintained a similar mutagenesis efficiency compared with those of 20 nt but also further limited off-target effects in *Drosophila*.

### Determination of the Optimal Injection Concentration

The individual effects of Cas9 and sgRNA expression levels have not been fully evaluated in *Drosophila*. To examine whether Cas9 level affects sgRNA-triggered mutagenesis, we tested the mutagenesis efficiency of heterozygous Cas9 transgenic flies, which should express Cas9 at half of the typical homozygous level. We found that heterozygous Cas9 flies had similar heritable mutation rates as homozygous flies when injected with either high- or low-efficiency sgRNAs targeting the *white* gene (Table S3). Thus, Cas9 level does not play a critical role in determining the mutagenesis efficiency within the wide range tested in the DGSC system.

We then asked whether there was an optimal amount of sgRNA for mutagenesis. Because equal volumes (1 nl) of sgRNA

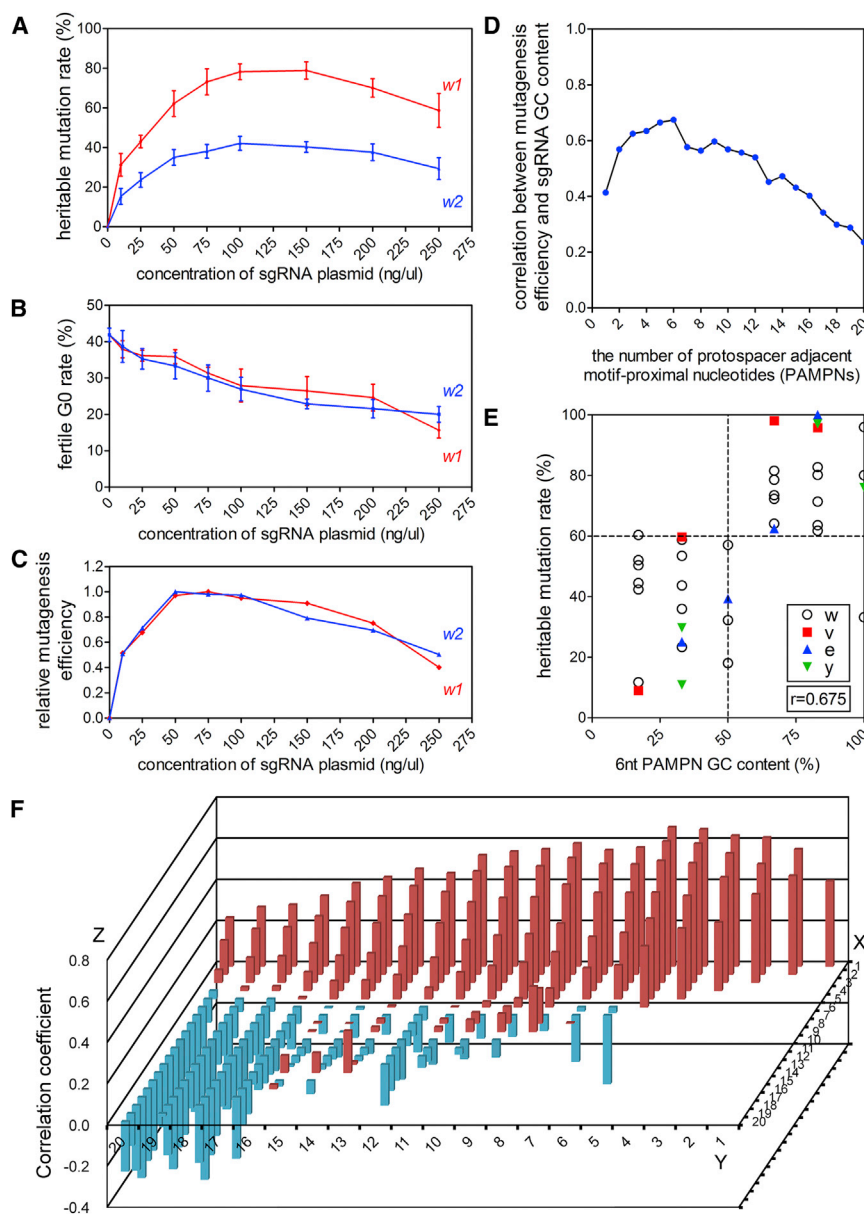
vectors are introduced into fly embryos with each injection, we can identify an optimal sgRNA concentration by assessing the mutagenesis efficiency and the fertile G0 rate. Specifically, we examined two sgRNAs targeting the *white* gene, *w1* and *w2*, which represented high and low efficiencies of mutagenesis, respectively, over a concentration range of 0–250 ng/μl. To our surprise, we found that concentrations between 75 and 150 ng/μl generated the highest mutation rate for both sgRNAs (Figure 2A). When the fertile G0 rate, which decreased with increased sgRNA concentration (Figure 2B), is taken into account, concentrations from 50 to 100 ng/μl generated the highest mutation rate for both sgRNAs (Figure 2C). Concentrations were tested using an additional sgRNA that targets *e*, and the same optimal concentrations between 50 and 100 ng/μl were observed (Figures S2A–S2C). Taken together, our data suggest that the optimal injection concentration of the sgRNA construct is approximately 75 ng/μl.

### Mutation Rate Is Highly Associated with the Sequence of the Six PAMPNs

Different sgRNAs produce different mutagenesis rates, possibly due to the epigenetic state of the target genomic locus or features of the targeting sequence of the sgRNA. Methylation of DNA does not affect Cas9 activity and, even though higher levels of Cas9 bind in euchromatic regions, Cas9 can target heterochromatin regions of the *Drosophila* chromosome (Hsu et al., 2013; Kuscu et al., 2014; Yu et al., 2013). Therefore, we focused our analysis on the targeting sequence of the sgRNAs.

Previously, when using the DGSC system to generate *w* mutants in *Drosophila* (Ren et al., 2013), we observed a correlation between mutagenesis efficiency and GC content of the most proximal nucleotides to the PAM sequence of the sgRNAs we designed. To further confirm this phenomenon, we designed a series of sgRNAs targeting the coding regions of *y*, *w*, *v*, and *e* (Table 2). We evaluated the efficiency of these sgRNAs in generating mutants in the F1 generation and, to determine the optimal number of PAMPNs to consider, we analyzed the correlation coefficient between mutagenesis efficiency and GC content of PAMPNs of different lengths (Figures 2D and 2F). The correlation was highest (0.675) for the GC content of a 6 nt PAMPN motif (Figures 2D–2F). In contrast, we did not observe such strong correlations between mutagenesis efficiency and the GC content of the 12 nt seed region (0.540) or the complete 20 nt targeting sequence (0.236) (Figures 2D, 2F, and S2D–S2U). sgRNAs with three or fewer GCs in the six nucleotides closest to the PAM rarely reach a 60% heritable mutation rate, but sgRNAs with at least four GCs in that region nearly always have a heritable mutation rate over 60% (Figure 2E). These observations suggest that effective sgRNAs can be selected according to the GC content of the six PAMPNs.

Of the ~7.6 million sgRNAs targeting the fly genome, about 2.4 million have greater than 50% GC contents at the 6 nt PAMPN and thus are more likely to have high efficiencies. A total of 892,854 of these target the coding sequence of 13,612 protein coding genes, which covers 97.5% of all protein coding genes (Table S4). If the criterion of having fewer than 3 nt mismatched potential off-targets is included, about 2.1 million sgRNAs qualify, and these cover 97.2% of all protein coding genes (Table S4).



**Figure 2. Parameters Affecting sgRNA Mutagenesis Efficiency**

(A) Heritable mutation rate of sgRNA *w1* ( $n = 3$ ) and *w2* ( $n = 3$ ).

(B) Fertile G0 rate of *w1* ( $n = 3$ ) and *w2* ( $n = 3$ ). Fertile G0 rate is calculated as the number of fertile G0 flies divided by total embryos injected.

(C) Relative mutagenesis efficiency of *w1* and *w2*. To calculate relative mutagenesis efficiency, heritable mutation rate is multiplied by fertile G0 rate and divided by the highest value.

(A–C) The x axis shows injection concentration of sgRNA vector at 0, 10, 25, 50, 75, 100, 150, 200, and 250 ng/μl. Results are from three independent experiments each for *w1* and *w2*. Error bar shows SEM.

(D) Plot figure showing the correlation between mutagenesis efficiency and sgRNA GC content, when different numbers of PAM-proximal nucleotides are considered.

(E) Scatterplot demonstrating the correlation between sgRNA heritable mutation rate and six PAMPN GC content. Data points for the sgRNA targets in *white* ( $w$ ,  $n = 27$ ), *vermillion* ( $v$ ,  $n = 4$ ), *ebony* ( $e$ ,  $n = 4$ ), and *yellow* ( $y$ ,  $n = 4$ ) are shown. Pearson's correlation coefficient ( $r$ ) = 0.675.

(F) A 3D column graph that shows the comparison of correlation coefficient when different consecutive regions of the sgRNAs are considered. Each column represents the Pearson's  $r$  between mutagenesis efficiency and GC content of a specific region of the sgRNAs. The 20 nt targeting sequence of the sgRNA is numbered with the nucleotide closest to PAM as 1 and the one farthest to PAM as 20. The x axis shows the position of the nucleotide at the PAM-distal end of the region, the y axis shows the position of the nucleotide position at the PAM-proximal end, and the z axis show the value of  $r$ . Positive values are shown in brick red and negative values in sky blue. The correlation coefficient is at the highest between mutagenesis efficiency and the GC content of the six nucleotides closest to the PAM.

See also Figure S2.

### Generating Multiple Gene Mutations in One Step

To understand how protein complexes work in vivo and how redundant genes work cooperatively, the simultaneous mutation of multiple genes using the Cas9/sgRNA system has been recently developed in mice (Wang et al., 2013). However, because only a few pole cells are present in each *Drosophila* embryo, it is a significant challenge to recover defined mutants in multiple genes in one step. We wondered whether our optimization of injection concentration and sgRNA target selection would overcome such limitations.

We injected Cas9 transgenic fly embryos with four sgRNA plasmids targeting *y*, *w*, *v*, and *e* at a combined DNA concentration of 300 ng/μl (75 ng/μl each). Two groups of sgRNAs were selected based on six PAMPN GC content to represent low and high efficiency sgRNAs for the six PAMPN (Table S5).

Since these mutations have visible recessive phenotypes, we crossed adult G0 flies to existing strains with multiple mutations and evaluated mutation rates by the visible markers in F1 (Figure 3A). However, *vermillion*-eye and *white*-eye phenotypes are not visible at the same time, since *w* is epistatic to *v*.

Therefore, for each injection, we divided adult G0 flies into two groups, and we crossed them to either existing *y w e* or *y v e* triple mutants. For the six PAMPN low-GC content group, the triple mutagenesis rates reached 0.2% and 3.0% for *y w e* and *y v e* combinations, respectively. For those with high GC content, the rates were 4.5% and 13.6% for the two different combinations (Figure 3B). We then crossed successful *y w e* triple-mutant male F1s to existing *v* females. We could therefore estimate quadruple rates by assessing *v* mutation rates in *y w e* mutants,

**Table 2. Correlation between sgRNA Mutation Rate and GC Content**

Gene Name (CG#)	sgRNA Target Sequence	6 nt GC Content (%)		Heritable Mutation Rate (%) <sup>a</sup>
<i>white</i> (CG2759)	GCACCATGGCTTGAAAATC	17	low	11.8
	CACCTATGCCTGGCACAATA	17		42.4
	CCGTTAGGGAGCCGATAAAG	17		44.5
	TTCGGGCAGGCCAAAACTA	17		50.4
	GGCACAATATGGACATCTTT	17		52.1
	CAGCAGGATGACCTCTTTAT	17		60.4
	CTGCAACGAGCGACACATAC	33		23.4
	TTCTTGAGCAAATGTTTCCT	33		36.0
	GCACAATATGGACATCTTTG	33		43.7
	TTCGAGAGCTGCATTAACC	33		53.5
	TGCGAGAGCTGCATTAACCA	33		58.8
	CGCCGGAGGACTCCGGTTCA	50		18.1
	TAGTTGGCCGCTCCCTGAAC	50		32.3
	CTGCGGCGATCGAAAGGCCA	50		57.1
	GCTGCATTAACCAGGGCTTC	67	high	64.1
	CCCAGTCCGCCGAGGACTC	67		72.3
	TTTTGGCCTGCCCGAAGCCC	67		73.5
	CCAAAACTACGGCAGCCTC	67		78.6
	CAATATGGACATCTTGGGG	67		81.6
	CAGGAGCTATTAATTCGCGG	83		61.8
GCTCCGGCCACCCAGTCCGC	83		63.6	
TTATCGGCTCCCTAACGGCC	83		71.4	
CCTCCGGCGGACTGGGTGGC	83		80.2	
CATTAACCAGGGCTTCGGGC	83		82.8	
CCGGCCACCCAGTCCGCCGG	100		33.2	
AGCGACACATACCGGCGCCC	100		80.1	
TTGAAACTCAGTTTGCGGCG	100		96.0	
<i>vermillion</i> (CG2155)	CAGAAACGATCACGATGATT	17	low	9.0
	TTCGGCGGTGCCATTAACCA	33		59.7
	AACAGATGCTCATCGTGAC	67	high	98.1
<i>ebony</i> (CG3331)	TCTGTTTCATCATCAGCACC	83		95.8
	GCTGCTGCTCCTCGAAGATG	33	low	25.1
	CCACAATTGTGCATCGTCAA	50		39.4
<i>yellow</i> (CG3757)	TGTGGTGAATGCGGTTCCGCC	67	high	62.5
	GAACCGGGCAGCCCCTCC	83		100
	GTGCACTGTTCCAGGACAAA	33	low	10.9
	TTGGGCTGCTTACAACTTC	33		29.8
	GCGATATAGTTGGAGCCAGC	83	high	97.1
	TATTCGCTACTGTTCCCGCG	100		76.1

<sup>a</sup>The heritable mutation rate was calculated as the number of mutant F1s divided by the number of all F1s observed. Six PAMPNs are underlined.

because mutant *v* allele could only originate from CRISPR/Cas9-introduced mutations in G0. Thus, we estimated the *y w v e* quadruple rate was 0% for the group of sgRNAs with low PAMPN GC content and 1.6% for the group with high GC content (Figure 3B).

To visualize quadruple-mutagenesis events directly, we chose sgRNAs that targeted *y*, *scute* (*sc*), *v*, and *e*, as the *sc* homozygous mutation has a visible phenotype that can be observed simultaneously with *y v e*. We injected Cas9 fly embryos with four sgRNAs vectors of high six PAMPN GC content at 75 ng/μl each and crossed G0 flies to existing *y sc v e* mutants. We observed a quadruple-mutagenesis rate of 7.4% (Figure 3B), which is comparable to the theoretical rate of 5.8% estimated by multiplying single-gene mutation rates. These results demonstrate that the simultaneous mutation of multiple genes is achievable in *Drosophila* with sgRNAs of high PAMPN GC content under the optimized condition.

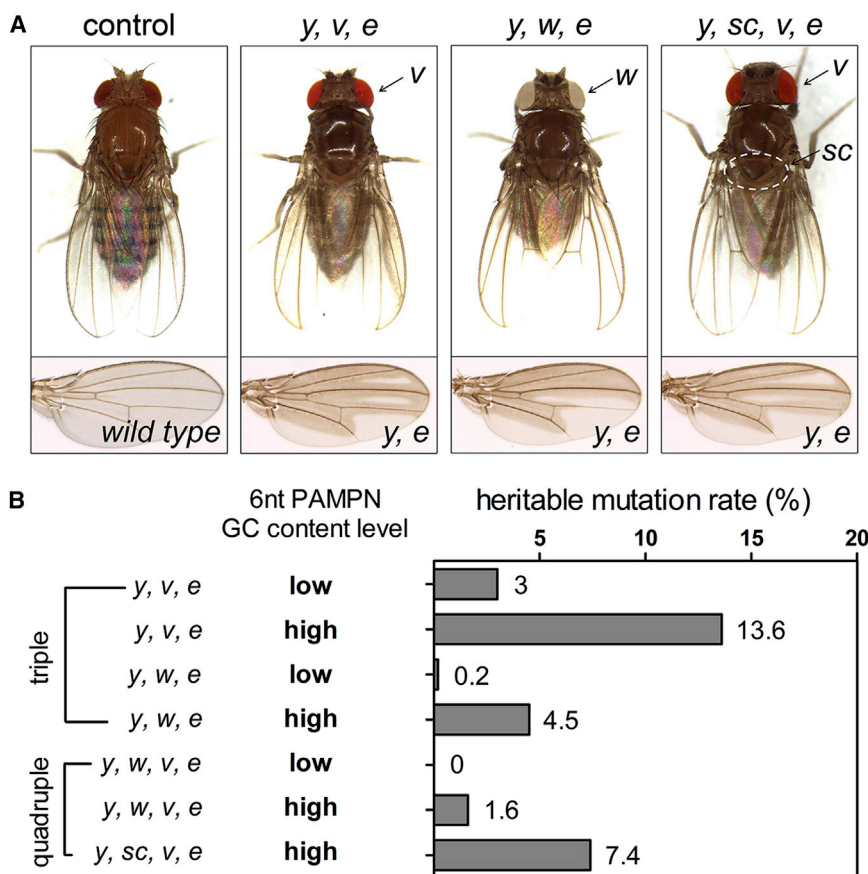
### Generating an *HP1a* Null Allele through Homology-Directed Repair

As the DGSC system was efficient in generating insertion or deletion mutants through the NHEJ pathway, we next evaluated the efficiency of mutagenesis by HDR. We chose *HP1a* for this test because known *HP1a* loss-of-function alleles are homozygous lethal before the pupal stage (James and Elgin, 1986; Eissenberg et al., 1992), allowing for evaluation of the HDR efficiency of developmentally essential genes. To replace the entire coding sequence of a gene through HDR, we usually use two sgRNAs and a plasmid DNA donor with a selective marker in combination with the DGSC system (Figure 4A). Two sgRNA plasmids with high six PAMPN GC content (Figure S3B) and a DNA donor template with 4XP3 promoter-driven mCherry flanked by homologous arms to *HP1a* (Figure 4A) were coinjected. Successful HDR mutants with mCherry eyes were counted in the F1 generation (Figure 4B) and confirmed by PCR (Figure 4C). We found the HDR efficiency occurred at a rate of 12% (98/822 F1s), with a fertile G0 rate of 11.6% (10/86 G0s) and a founder (G0s that produced any mutant F1s) rate of 50% (Table 3). We selected six successful *HP1a*<sup>HDR-mCherry</sup> mutant F1s (four from different G0s and two from the same G0) with mCherry expression in the eyes and confirmed all six lines had the same break points by sequencing (Figure S3A). We also tested for off-target effects in 24 F1s that had mCherry expression in the eyes. Since we tried to avoid off-target cuts when designing the sgRNAs, we applied two sgRNAs without close matches (i.e., with potential off-targets of at least four mismatches) in the fly genome. Thus, the two most likely off-targets had four mismatches to the sgRNAs (Figure S3B). We sequenced these sites and detected no off-target mutations in any of the 24 mutant F1s.

### DISCUSSION

Previous work, including ours, has demonstrated the feasibility of CRISPR/Cas9-mediated gene editing in *Drosophila* (Baena-Lopez et al., 2013; Bassett et al., 2013; Gratz et al., 2013, 2014; Hsu et al., 2013; Kondo and Ueda, 2013; Ren et al., 2013; Yu et al., 2013, 2014; Port et al., 2014; Xue et al., 2014). The off-target effect (i.e., specificity) and efficiency of the CRISPR/Cas9 system has been recently investigated in mammalian cells (Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013; Cho et al., 2014; Fu et al., 2014; Kuscu et al., 2014; Lin et al., 2014; Wu et al., 2014). In this work, we investigated the parameters to achieve high





**Figure 3. Triple and Quadruple Mutations in *Drosophila* with One Shot**

(A) Representative images of control flies, triple mutant F1s, and quadruple mutant F1s. The genotypes are listed on the top.

(B) Triple and quadruple mutagenesis rates shown in a bar graph. Each row represents an sgRNA combination and the corresponding mutagenesis rate. Low GC content sgRNAs have less than three GCs in the six PAM-proximal nucleotides (PAMPNs), while high GC content ones have five or six.

See also Table S5.

specificity and efficiency in *Drosophila*, which consequently enabled us to mutate up to four genes in one step and carry out HDR with higher efficiency than previously reported.

### Specificity of the CRISPR/Cas9 System

In our tests, the mutagenesis efficiency dramatically reduces to zero for the 81 sgRNAs with three or more mismatches, even at the most distal end to the PAM, when 20 nt sgRNAs are used. Since mismatches closer to the PAM are more likely to disrupt the sgRNA-DNA hybrid (Jinek et al., 2012; Cong et al., 2013; Mali et al., 2013b), genomic regions with three or more mismatches to sgRNAs have reduced or no off-target effects in *Drosophila*. This conclusion is further supported by the observation that sgRNAs targeting the *y*, *w*, *v*, and *e* genes do not induce off-target mutations in genomic DNA carrying three mismatches (Figures 1D and S1B) and is in line with recent studies in mammalian cells (Hsu et al., 2013; Mali et al., 2013a; Wang et al., 2014). In addition, although sgRNAs targeting DNA with NAG PAMs do not generate mutants efficiently, potential off-target effects with NAG PAMs cannot be completely excluded.

Consistent with a previous report (Fu et al., 2014), the sgRNAs with targeting sequences of 18 nt are as effective as those with 20 nt in *Drosophila*. This result is seemingly contradictory with our result with the mismatched sgRNAs. However, these results support a model where mismatches between the sgRNA and

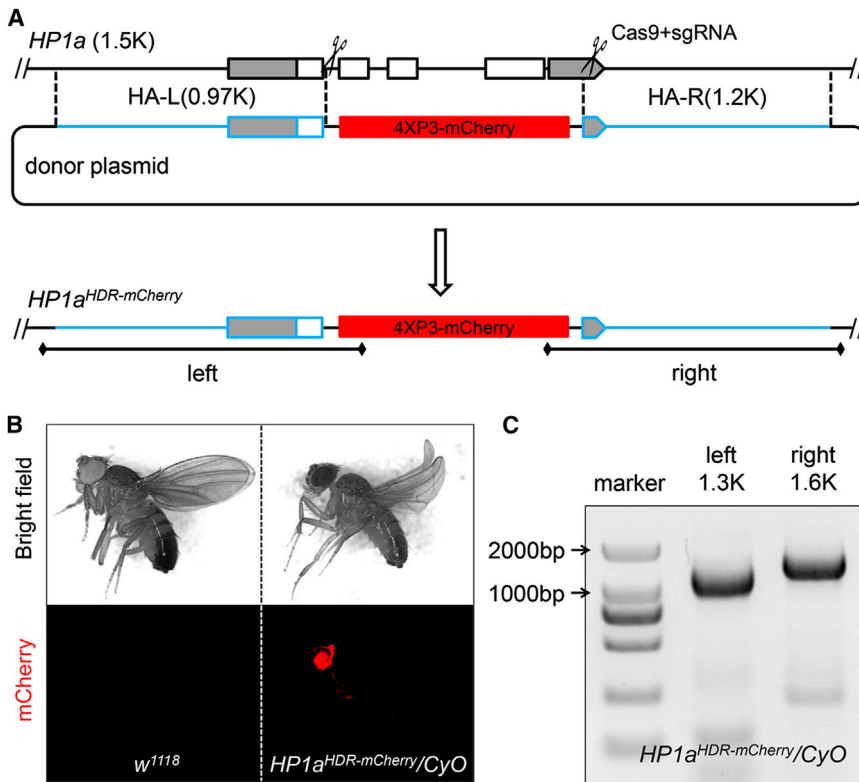
the DNA target prevent Cas9 nuclease activities but extensive matches, although fewer, allow DSBs by Cas9 (Wu et al., 2014). The finding that sgRNAs with targeting sequences shorter than 20 nt are as effective but more specific in *Drosophila* further facilitates the design of specific sgRNAs. However, it is important to note that removing 1 nt from the targeting sequence of an sgRNA increases the possibility of off-targets in the genome 4-fold, theoretically. Since 18 nt target sgRNAs nearly abolish off-target effects in regions with only a single base mismatch, decreasing sgRNA length further may not be necessary. Nevertheless, the observation that 17 to 20 nt target sgRNAs have similar muta-

genesis efficiencies in *Drosophila* demands further investigation of even shorter target lengths in vivo. It is surprising to find that 21 to 22 nt target sgRNAs have a dramatic decrease in efficiency compared with those of 20 nt. Possibly, the extra 1–2 nt have to be trimmed away in *Drosophila* just like in mammalian cell lines (Ran et al., 2013), and thus the additional processing makes 21 and 22 nt target sgRNAs less effective than 20 nt ones. Alternatively, the extra base pairs might interfere with the Cas9 enzymatic activity or the dissociation of Cas9 from genomic DNA.

The results of our study suggest the pool of specific sgRNA targets in the *Drosophila* genome is significantly larger than previously thought (Ren et al., 2013), which makes it possible to target specific locations to make fine alterations in the fly genome instead of simply knocking out genes. These results also show that the phenotypic analysis of mutant alleles generated by the Cas9/sgRNA system is more reliable than we previously assumed (Ren et al., 2013).

### Mutagenesis Efficiency of the CRISPR/Cas9 System

Our in vivo data show that the protein level of Cas9 is not critical in determining mutagenesis efficiency within the bounds of the DGSC system. We then examined a wide range of injected sgRNA concentrations, with the expectation that the mutagenesis rate would positively correlate with the sgRNA concentration. To our surprise, we found that the mutagenesis efficiency



**Figure 4. Mutating *HP1a* through HDR Using Optimal sgRNAs**

(A) Diagrams showing the repair donor plasmid and the mutation after HDR. The coding sequence of the *HP1a* locus is illustrated by the white boxes and the 5' and 3' UTRs by the shaded boxes. The *HP1a* donor *HP1a*-4XP3-mCherry contains a 4XP3-mCherry sequence (red box) to replace most of the coding sequence of *HP1a*. The two homologous arms (HA-L and HA-R; blue) of the donor template are 0.97k bp and 1.2k bp, respectively. The Cas9/sgRNA cutting sites are denoted by the scissors. Successful replacement can be detected by mCherry expression in the fly eyes or by PCR to detect the left and the right homologous arms.

(B) Images of *w<sup>1118</sup>* control and successful *HP1a<sup>HDR-mCherry</sup>* heterozygous mutant flies under bright-field (top) or epifluorescent light sources to show mCherry expression in the eyes (bottom).

(C) Agarose gel electrophoresis with PCR band sizes confirming the successful *HP1a* HDR mutation.

See also Figure S3.

peaks at an injection concentration of 75 to 150 ng/ $\mu$ l. The exact reason why increasing sgRNA concentration beyond 150 ng/ $\mu$ l decreases mutagenesis rates is not clear. However, the fact that the fertile G0 rate decreases with increased sgRNA concentration supports an idea that excessive off-target DNA DSBs may lead to cell death and thus loss of mutant germline cells. Alternatively, large amounts of sgRNA or plasmid DNA might be toxic to the cells.

Considering these observed effects of sgRNA concentration on both mutagenesis efficiency and specificity, an injection concentration of approximately 75 ng/ $\mu$ l is ideal when using the DGSC system to generate heritable mutant flies. Since most DNA miniprep kits routinely yield a final concentration of greater than 100 ng/ $\mu$ l, sgRNA vectors from minipreps can be easily applied in the DGSC system without further concentration. These results also suggest that higher levels of sgRNA do not necessarily mean higher efficiency, when applying CRISPR/Cas9 systems.

Importantly, we found the strongest positive correlation between sgRNA mutagenesis efficiency and the GC content of the six PAMPNs. We also observed that sgRNA efficiency of previous reports in *Drosophila* fit this trend (Bassett et al., 2013; Kondo and Ueda, 2013; Yu et al., 2013), even though they injected in vitro-synthesized Cas9 RNA or Cas9 DNA vector, which suggests that this strong correlation can be applied to predict sgRNA efficiency in *Drosophila*. Previous reports in mammalian cell line studies showed that medium-GC-content sgRNAs have higher mutagenesis efficiency compared with those of low or high GC content, when the whole 20 nt targeting

data from cell line studies show a trend between efficiency and GC content of sgRNAs if the seed region is considered.

Previous reports show that mismatches in the PAMPNs are more likely to disrupt complementation between sgRNA and the intended target than those in the distal nucleotides and that PAMPNs are important in maintaining the sgRNA:DNA heteroduplex in Cas9 (Jinek et al., 2014; Nishimasu et al., 2014). We therefore hypothesize that PAMPNs with higher GC content have more hydrogen bonds between the sgRNA and DNA, which stabilize the hybrid and increase the Cas9 enzyme efficiency, thus increasing overall heritable mutation rates. sgRNAs with more than 50% GCs in the six PAMPNs generally have higher mutagenesis rates than those with less than 50% GCs. However, within the group of sgRNAs that have more than 50% GCs, there is no further correlation between the actual GC content and mutation rate (Figure 2E). Nonetheless, this strong relationship suggests that the GC content of the six PAMPNs can predict the mutation efficiency of the sgRNA in *Drosophila* and thus facilitate the design of sgRNAs.

### One-Step Mutation of Multiple Genes

Generating mutations in multiple genes in one step in flies has been challenging, even with CRISPR/Cas9-based techniques. Microinjection of multiple sgRNA plasmids at the previously described concentration is toxic to the embryo and reduces the number of viable G0s (Ren et al., 2013). The determination of optimal injection parameters and new criteria for selecting an efficient sgRNA target sequence allowed us to generate flies with up to four mutant genes. Under optimized conditions, G0

**Table 3. A Comparison of CRISPR/Cas9-Mediated HDR Rates in *Drosophila* from Recent Literature**

Reference	Recessive Viable Allele	Recessive Lethal Allele	HDR Efficiency	
			% (no.) Founders	% (no.) HDR Progeny
Yu et al., 2014	CG4221	—	10.8 (3/28)	4.3 (10/230)
	CG5961	—	8.3 (1/12)	3.8 (2/52)
Gratz et al., 2014	DSH3PX1 (CG6757)	—	18.0 (9/50)	7.8 (599/7657)
		—	11.1 (2/18)	1.6 (48/3016)
		—	0 (0/28)	0 (0/5665)
		—	3.2 (1/31)	0.04 (2/4573)
Port et al., 2014	<i>wntless</i> (CG6210)	—	12.5 (2/16)	2.0 (45/2277)
		—	67 (4/6)	28 (13/46)
		—	60 (3/5)	11 (8/76)
Our method	<i>piwi</i> (CG6122)	—	53.8 (7/13)	32.8 (446/1361)
		—	<i>HP1a</i> (CG8409)	50 (5/10)

survival and fertile rates are sufficient to produce triple or quadruple mutants, in sharp contrast to using low-GC-content sgRNAs. Thus, selecting high PAMPN GC content is critical for mutating multiple genes in one step. This optimization paves the way for genetic studies of protein complexes and redundant pathways in which multiple genes may need to be mutated to manifest any phenotypes.

### Efficient Homology-Directed Repair to Target a Homozygous Lethal Gene

Previously, CRISPR/Cas9-mediated HDR was successfully applied to mutate homozygous viable genes, with the mutagenesis efficiency generally lower than 5% (Baena-Lopez et al., 2013; Gratz et al., 2014; Xue et al., 2014; Yu et al., 2014). The HDR recovery rates with homozygous lethal genes have not been reported, although they are likely to be substantially lower. We targeted *HP1a* to examine whether our optimizations enabled the targeting of a homozygous lethal gene. The G0 survival rate when generating the *HP1a*<sup>HDR-mCherry</sup> allele was at 11.6% (10/86 G0s), which was similar to that for the production of a *piwi* mutant at 16.0% (13/81 G0s) (Ren et al., 2014; our unpublished data), supporting that HDR only occurred in germ cells and thus would not affect G0 survival rate. The F1 HDR mutant rate of *HP1a* (12%) was much lower than that of *piwi* (32.8%) (Ren et al., 2014; our unpublished data), indicating that germ cells carrying homozygous *HP1a* mutations were lethal, thus reduced the efficiency. The achieved high HDR-mediated mutagenesis efficiency shows a substantial improvement over previous methods (Baena-Lopez et al., 2013; Gratz et al., 2014; Xue et al., 2014; Yu et al., 2014) and is comparable with a very recent report using high concentrations of single-stranded DNA oligonucleotide donors (Port et al., 2014) (Table 3). Therefore, the optimized sgRNA parameters we identified in this work will make HDR a lot more feasible in *Drosophila*, as long as fertile G0 flies can be recovered.

## EXPERIMENTAL PROCEDURES

### sgRNA Vector Construct

sgRNAs were designed using the online CRISPR design tool (<http://www.flyrnai.org/crispr/>) and cloned into the *U6b*-sgRNA-short vector as previously described (Ren et al., 2013). The oligonucleotides used for cloning are listed in Table S6. sgRNA sequences are also listed in Tables 2, S1, S2, S3, S5, and S9 and Figure S3.

### Donor Vector Construct for HDR at the *HP1a* Locus

The *HP1a*-4XP3-mCherry donor construct was based on the pBluescript plasmid. The left homologous arm of *HP1a* was amplified from genomic extract with primers *HP1a*-HA-left-F and *HP1a*-HA-left-R and was cloned into the HindIII and AvrII sites. The right homologous arm was amplified with primers *HP1a*-HA-right-F and *HP1a*-HA-right-R, and was cloned into the SpeI and SacI sites. The selection marker 4XP3-mCherry was constructed on a different pBluescript vector first. The gene encoding the red fluorescent protein mCherry was amplified with primers mCherry-F and mCherry-R and was cloned into the XhoI and KpnI sites. The 4XP3 promoter sequence (Horn et al., 2000) was synthesized and cloned into the HindIII and XhoI sites. The SV40 3' UTR sequence was amplified with primers SV40-F and SV40-R and cloned into the KpnI and EcoRV sites. The selection marker 4XP3-mCherry was then cut and inserted between the left and right homologous arms of *HP1a* to finish the *HP1a*-4XP3-mCherry construct. All PCR fragments were amplified with *pfu* DNA polymerase (TransGen Biotech). The donor construct was confirmed by sequencing (Invitrogen). The oligonucleotides used for cloning and PCR confirmation are listed in Table S8.

### DNA Purification and Embryo Injection

DNA plasmid solution was thoroughly mixed with 1/10 volume of 3 M sodium acetate (pH 5.2, AMRESCO) and 5 volumes of absolute ethanol and stored at  $-20^{\circ}\text{C}$  for 2 hr followed by 21,000  $\times g$  centrifugation. The DNA pellet was washed twice in 70% ethanol and twice in 100% ethanol and resuspended in injection buffer for the appropriate concentration. The *Drosophila* embryos were injected as previously described (Ren et al., 2013). The injection concentration of sgRNA plasmids was at 75 or 250 ng/ $\mu\text{l}$ , except when a range of concentrations (0–250 ng/ $\mu\text{l}$ ) was tested to determine the optimal injection concentration. The injection concentration of DNA donor for homology-directed repair was 100 ng/ $\mu\text{l}$ . The detailed injection concentrations, survival rates, and fertility rates were listed in Table S9.

### Fly Stocks and Mutation Screening

All flies were cultured on standard cornmeal food at  $25^{\circ}\text{C}$ , unless otherwise noted. To score for germline mutations, G0 adult flies that developed from injected *P{nos-Cas9}attP2* embryos were crossed to *y[1] w[67c23]*, for *white* and *yellow* mutations or to *y[1] sc[1] v[1]* for *vermillion* or to *y[1] sc[1] v[1]; Dr[1] e[1]/TM3, Sb[1]* for *ebony*. *y[1] sc[1] v[1]; Dr[1] e[1]/TM3, Sb[1]* and *y[1] w[67c23]; Dr[1] e[1]/TM6B, Tb[1]* were used in screening for triple and quadruple mutants. Embryos from *y[1] sc[1] v[1]; P{nos-Cas9}attP2* flies crossing to *y[1] sc[1] v[1]* flies were used for assessment of heterozygous Cas9 mutagenesis efficiency. *y[1] sc[1] v[1]; wg[Gla-1] Bc[1]/CyO, Cy[1]* was used as the balancer stock for HDR mutants.

The F1 progeny were screened for the first 6 days after eclosion. The heritable mutation rate was calculated as the total number of mutant F1 progeny divided by the number of progeny screened for a given sgRNA target. The correlation coefficient of the sgRNA GC content and mutation rate was calculated by the Pearson method. Mutagenesis events were confirmed by sequencing of F1 adults, and the detection primers are listed in Table S6. The representative images of mutations of relevant genes were obtained using a Leica MZ16 FA microscope with Leica Application Suite (V3.0) software. Successful *HP1a*<sup>HDR-mCherry</sup> mutants were screened by the expression of mCherry in the eyes under a Leica MZ10F fluorescent microscope.

### Genomic DNA Extraction

Fly genomic DNA was purified via phenol-chloroform extraction. Single flies were homogenized in 400  $\mu\text{l}$  of lysis buffer (1 $\times$  PBS, 0.2% SDS, 200  $\mu\text{g}/\text{ml}$

proteinase K) and incubated at 50°C for 1 hr, followed by extraction in 400  $\mu$ l of phenolchloroform. The mixture was then centrifuged at 21,000  $\times$  *g* for 20 min at 4°C. The supernatant was transferred to a new tube. An equal volume of isopropanol was added, and the tube was vortexed thoroughly. The mixture was then kept at –20°C for at least 1 hr, followed by centrifugation at 21,000  $\times$  *g* for 20 min at 4°C. The supernatant was removed, and the DNA pellet was washed with 500  $\mu$ l of 75% ethanol, followed by centrifugation at 21,000  $\times$  *g* for 5 min at 4°C. Finally, the pellet was dried for 10 min and resuspended in 30  $\mu$ l of DNase-free water.

### Off-Target Analysis

Potential off-target sites in the *Drosophila* genome were identified using CRISPR Browser (<http://www.flyrnai.org/crispr>), TagScan (<http://ccg.vital-it.ch/tagger>), and Flybase Blast (<http://flybase.org/blast>). Detailed information about off-target sites are listed in Table S10. The off-target effect was investigated by sequencing (Invitrogen), as previously described (Ren et al., 2013). Briefly, the G0 adults that developed from injected embryos were individually crossed to mutation lines of the corresponding genes or to a balancer stock, as described in the above section “Fly Stocks and Mutation Screening.” For off-target analysis on sgRNAs targeting *y*, *w*, *v*, or *e*, eight independent F1 mutant flies derived from different G0s were collected and their genomic DNA extracted. For off-target analysis on sgRNAs targeting *HP1a*, four F1s each from four G0s and eight F1s from one G0 were collected, making a total of 24 flies. Genomic DNA from single flies was used as the template, and primers flanking the potential off-target sites were used to amplify a defined DNA fragment. The PCR products were either directly sequenced by one of the primers or first cloned into the VALIUM20 vector (Ni et al., 2008) at an NheI and EcoRI site and then sequenced by specific primers. The off-target effect was identified by alignment of sequenced results to the wild-type genomic sequence. Oligonucleotides used for on/off-target analysis are listed in Tables S6–S8.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and ten tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.09.044>.

### AUTHOR CONTRIBUTIONS

J.-Q.N. designed the experiments; X.R., Z.Y., J.X., J.S., and D.M. performed the experiments; S.-J.Y., H.-H.Q., X.W., Q.H., and L.-P.L. provided technical support; J.B.L. and J.-Y.J. provided resources; J.-Q.N., X.R., Y.H., J.X., P.D., and J.B.L. analyzed the data; and J.-Q.N. and J.X. wrote the manuscript with input from P.D. and J.B.L.

### ACKNOWLEDGMENTS

We thank Dr. Norbert Perrimon (Harvard Medical School), Dr. Babak Javid (Tsinghua University), Dr. Michael Bassik (Stanford University), and Dr. Neville Sanjana (Broad Institute) for critical comments on the manuscript. This work was supported by the National Basic Research Program (973 Program; 2013CB835100), the National Natural Science Foundation of China (20131351195 and 31301008), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20121018577, 20120002110056, and 20120141120046), the Natural Science Foundation of Hubei Province (2013CFB031), the Tsinghua-Peking Center for Life Sciences, an HBUT starting grant (BSQD12142), NIH grant R01GM102484, the Ellison Medical Foundation, and the Stanford University Department of Genetics. P.D. was supported by the National Science Foundation Graduate Research Fellowship, Stanford Genome Training Program and Cell and Molecular Biology Training Program.

Received: June 3, 2014

Revised: August 22, 2014

Accepted: September 24, 2014

Published: October 23, 2014

### REFERENCES

- Baena-Lopez, L.A., Alexandre, C., Mitchell, A., Pasakarnis, L., and Vincent, J.P. (2013). Accelerated homologous recombination and subsequent genome modification in *Drosophila*. *Development* *140*, 4818–4825.
- Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Reports* *4*, 220–228.
- Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* *161*, 1169–1175.
- Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* *23*, 465–472.
- Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* *24*, 132–141.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* *339*, 819–823.
- Eissenberg, J.C., Morris, G.D., Reuter, G., and Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in *Drosophila* with dosage-dependent effects on position-effect variegation. *Genetics* *137*, 345–352.
- Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* *10*, 741–743.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* *31*, 822–826.
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* *32*, 279–284.
- Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* *31*, 397–405.
- Gao, G., McMahon, C., Chen, J., and Rong, Y.S. (2008). A powerful method combining homologous recombination and site-specific recombination for targeted mutagenesis in *Drosophila*. *Proc. Natl. Acad. Sci. USA* *105*, 13999–14004.
- Gong, W.J., and Golic, K.G. (2003). Ends-out, or replacement, gene targeting in *Drosophila*. *Proc. Natl. Acad. Sci. USA* *100*, 2556–2561.
- Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O’Connor-Giles, K.M. (2013). Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* *194*, 1029–1035.
- Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O’Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* *196*, 961–971.
- Horn, C., Jaunich, B., and Wimmer, E.A. (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.* *210*, 623–629.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* *31*, 827–832.
- Huang, J., Zhou, W., Dong, W., Watson, A.M., and Hong, Y. (2009). From the Cover: Directed, efficient, and versatile modifications of the *Drosophila* genome by genomic engineering. *Proc. Natl. Acad. Sci. USA* *106*, 8284–8289.
- James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.* *6*, 3862–3872.

- Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821.
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*.
- Kondo, S., and Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* **195**, 715–721.
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683.
- Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42**, 7473–7485.
- Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B., et al. (2012). Efficient and specific modifications of the *Drosophila* genome by means of an easy TALEN strategy. *J. Genet. Genomics* **39**, 209–215.
- Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826.
- Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., Booker, M., Perkins, L., and Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. *Nat. Methods* **5**, 49–51.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949.
- Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843.
- Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **111**, E2967–E2976.
- Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389.
- Ren, X., Sun, J., Housden, B.E., Hu, Y., Roesel, C., Lin, S., Liu, L.P., 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. USA* **110**, 19012–19017.
- Ren, X., Yang, Z., Mao, D., Chang, Z., Qiao, H.H., Wang, X., Sun, J., Hu, Q., Cui, Y., Liu, L.P., et al. (2014). Performance of the Cas9 Nickase System in *Drosophila melanogaster*. *G3 (Bethesda)* <http://dx.doi.org/10.1534/g3.114.013821>.
- Rong, Y.S., and Golic, K.G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013–2018.
- Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676.
- Xue, Z., Ren, M., Wu, M., Dai, J., Rong, Y.S., and Gao, G. (2014). Efficient gene knock-out and knock-in with transgenic Cas9 in *Drosophila*. *G3 (Bethesda)* **4**, 925–929.
- Yu, Z., Ren, M., Wang, Z., Zhang, B., Rong, Y.S., Jiao, R., and Gao, G. (2013). Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. *Genetics* **195**, 289–291.
- Yu, Z., Chen, H., Liu, J., Zhang, H., Yan, Y., Zhu, N., Guo, Y., Yang, B., Chang, Y., Dai, F., et al. (2014). Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the *Drosophila* genome. *Biol. Open* **3**, 271–280.