CHAPTER EIGHTEEN

Cas9-Based Genome Editing in Zebrafish

Andrew P.W. Gonzales*,†, Jing-Ruey Joanna Yeh*,†,1
*Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, USA
†Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
1Corresponding author: e-mail address: jyeh1@mgh.harvard.edu

Contents

1. Introduction 378
   1.1 CRISPR/Cas adaptive immunity 378
   1.2 The Type II CRISPR/Cas system 379
   1.3 The development of CRISPR/Cas genome-editing technology 380
   1.4 The zebrafish animal model and CRISPR/Cas 383
2. Targeted Generation of Indel Mutations 385
   2.1 Cas9 modification and delivery platforms 385
   2.2 Single-guide RNA design considerations 388
   2.3 Introduction and identification of Cas9–sgRNA-induced indels 395
3. Other Targeted Genome-Editing Strategies 396
   3.1 Precise sequence modifications mediated by single-stranded oligonucleotides 396
   3.2 Targeted integration of long DNA fragments 397
   3.3 Chromosomal deletions and other rearrangements 400
4. Future Directions 401

Acknowledgments 403
References 403

Abstract

Genome editing using the Cas9 endonuclease of Streptococcus pyogenes has demonstrated unprecedented efficacy and facility in a wide variety of biological systems. In zebrafish, specifically, studies have shown that Cas9 can be directed to user-defined genomic target sites via synthetic guide RNAs, enabling random or homology-directed sequence alterations, long-range chromosomal deletions, simultaneous disruption of multiple genes, and targeted integration of several kilobases of DNA. Altogether, these methods are opening new doors for the engineering of knock-outs, conditional alleles, tagged proteins, reporter lines, and disease models. In addition, the ease and high efficiency of generating Cas9-mediated gene knock-outs provides great promise for
high-throughput functional genomics studies in zebrafish. In this chapter, we briefly review the origin of CRISPR/Cas technology and discuss current Cas9-based genome-editing applications in zebrafish with particular emphasis on their designs and implementations.

1. INTRODUCTION

1.1. CRISPR/Cas adaptive immunity

In order to persist and thrive within threatening virus-rich environments, prokaryotes over the course of evolutionary history have developed various kinds of defense mechanisms for fending off invading viral genetic elements (Labrie, Samson, & Moineau, 2010), one of which being an immune mechanism mediated by clustered regularly interspaced short palindromic repeat (CRISPR) loci (Barrangou et al., 2007; Ishino, Shinagawa, Makino, Amemura, & Nakata, 1987; Jansen, Embden, Gaastra, & Schouls, 2002). CRISPR loci are common among prokaryotes and have been estimated to be in ~40% and ~90% of all genomically sequenced bacteria and archaea, respectively (Grissa, Vergnaud, & Pourcel, 2007a; Kunin, Sorek, & Hugenholtz, 2007; Sorek, Kunin, & Hugenholtz, 2008). These loci, together with their neighboring CRISPR-associated (Cas) genetic elements, form an unique adaptive immune system called CRISPR/Cas, which utilizes short RNA-guided endonucleases to target, cleave, and degrade specific viral sequences during a recurring infection (Bhaya, Davison, & Barrangou, 2011; Bolotin, Quinquis, Sorokin, & Ehrlich, 2005; Horvath & Barrangou, 2010; Marraffini & Sontheimer, 2010a).

CRISPR loci are characterized by arrays of conserved ~20–50 base-pair (bp) repeats with distinct “spacer” sequences of comparable length interspaced between them (Grissa, Vergnaud, & Pourcel, 2007b; Rousseau, Gonnet, Le Romancer, & Nicolas, 2009). These loci are flanked by a cluster of cas genes which encode some of the enzymatic machinery utilized for normal CRISPR/Cas function (Makarova, Grishin, Shabalina, Wolf, & Koonin, 2006). Within a given CRISPR locus, each spacer sequence is unique and derived from fragments of invading viral nucleic acids acquired from a previous pathogenic exposure, thus allowing the prokaryote to generate a genetically stored immunological memory of past infections (Bolotin et al., 2005; Mojica, Diez-Villasenor, Garcia-Martinez, & Soria, 2005). All CRISPR/Cas systems generate this immunological memory by following a common three-step process (Wiedenheft, Sternberg, & Doudna, 2012).
First, during preliminary exposure to a pathogen, invading foreign nucleic acids must be cleaved into fragments called protospacers, which then become integrated as spacers in the CRISPR locus (Barrangou et al., 2007; Garneau et al., 2010). Second, during any subsequent infection, the CRISPR locus is transcribed to produce a single long pre-CRISPR RNA, which is then processed into an active genetic library of many spacer-derived short CRISPR RNAs (crRNAs) (Brouns et al., 2008). Third, crRNAs combined with one or more Cas proteins form RNA-guided endoribonuclease surveillance complexes, which by base-pair interactions between the crRNA spacer region and complementary viral protospacer sequences, allow the complexes to target and cleave invading foreign nucleic acids (Brouns et al., 2008).

1.2. The Type II CRISPR/Cas system

Three types of CRISPR/Cas systems are known to exist in nature, each of which uses distinct mechanisms to carry out the aforementioned three-step process to produce CRISPR-mediated adaptive immunity (Makarova, Aravind, Wolf, & Koonin, 2011; Makarova, Haft, et al., 2011; Wiedenheft et al., 2012). Among these three systems, the Type II CRISPR/Cas system is the best characterized and the simplest in certain key aspects. One important difference is that its surveillance complex requires a single Cas9 endonuclease (Chylinski, Makarova, Charpentier, & Koonin, 2014; Sapranauskas et al., 2011), while Type I and Type III systems require several proteins (Makarova, Aravind, Wolf, & Koonin, 2011; Makarova, Haft, et al., 2011).

In addition to the Cas9 protein, the Type II surveillance complex also consists of two RNA components, a crRNA and a transactivating crRNA (tracrRNA) (Deltcheva et al., 2011). The tracrRNA is required for normal crRNA processing and Type II surveillance complex formation, and the crRNAs contain 20-nucleotide (nt) spacer regions derived from the original CRISPR locus (Deltcheva et al., 2011). By complementary base-pair interactions, these crRNAs guide the surveillance complexes to target, bind, and degrade foreign genetic elements that contain protospacer sequences complementary to the spacer, as well as a Cas9-specific protospacer adjacent motif (PAM) directly 3’ to the target protospacer (Gasiunas, Barrangou, Horvath, & Siksnys, 2012).

Having the correct PAM sequence directly adjacent to the protospacer is necessary for DNA interrogation by the Type II surveillance complex and for the triggering of Cas9 cleavage activity. Indeed, mismatches within or
nearby the first few nucleotides of the PAM have been shown to inhibit heteroduplex formation and unwinding of the dsDNA target (Sternberg, Redding, Jinek, Greene, & Doudna, 2014). In this manner, the 3′ PAM sequences allow the Type II system to distinguish between sequences belonging to “self” and those that are foreign in order to prevent the destruction of its own CRISPR loci (Horvath et al., 2008; Marraffini & Sontheimer, 2010b). Among Type II Cas9 endonucleases found in various prokaryotic species, PAM sequences vary in complexity, one of the simplest being the 5′-NGG PAM of *Streptococcus pyogenes* Cas9 (Jinek et al., 2012). The natural RNA-guided Cas9 endonuclease from *S. pyogenes* (SpCas9) possesses the ability, in principle, to target any invading protospacer sequence in the form of 5′-N_{20}-NGG-3′. Thus, it is both the relatively small number of components required by the Type II system, combined with the flexibility of its required target sequence, which have allowed the recent adaptation of the Type II CRISPR/Cas system as a novel, powerful, and amendable genome-editing platform.

1.3. The development of CRISPR/Cas genome-editing technology

Contemporary genome editing relies on the usage of programmable nucleases to artificially produce gene disruptions, DNA insertions, targeted mutations, or chromosomal rearrangements in a predictable and controlled manner (Segal & Meckler, 2013). These engineered nucleases “edit” the genome by introducing targeted double-strand DNA breaks (DSBs), which in turn allow the cell’s natural repair mechanisms—e.g., nonhomologous end-joining (NHEJ) mediated repair and homology-directed repair (HDR)—to be co-opted for the purpose of site-specific DNA manipulation (Bibikova, Beumer, Trautman, & Carroll, 2003; Bibikova et al., 2001; Bibikova, Golic, Golic, & Carroll, 2002). The potential applications of such genome-editing technologies are far-reaching, including the bioengineering of disease-resistant, nutrient-rich crops and livestock (Carlson et al., 2012; Li, Liu, Spalding, Weeks, & Yang, 2012), the generation of various animal models and human pluripotent stem cell models that can be used for preclinical drug studies (Brunet et al., 2009; Carbery et al., 2010; Ding et al., 2013; Yang et al., 2013), and even the development of therapies involving the direct delivery of genetically corrected, patient-derived pluripotent stem cells or somatic cells (Schwank et al., 2013; Sebastiano et al., 2011). In light of the various potential benefits programmable nucleases present, the engineering of more facile, precise, and efficient genome-editing platforms is
highly sought after and have made the recent development of CRISPR/Cas genome-editing systems all the more valuable.

The first published instance of an engineered CRISPR/Cas system for the purpose of genome editing was in 2012, when researchers adapted the Type II CRISPR/Cas system of *S. pyogenes* and demonstrated that SpCas9 could be guided by a programmable chimeric dual-RNA to target and cleave various DNA sites *in vitro* (Jinek et al., 2012). In this study, the authors simplified the system even further to create a genome-editing platform that required only two components—SpCas9 and a synthetic single guide RNA (sgRNA) consisting of a fusion of the essential features of Type II crRNA and tracrRNA (Fig. 18.1). A few months after this CRISPR/Cas platform’s initial debut, its utility quickly expanded to a variety of cellular systems, exhibiting its efficacy in introducing targeted mutations within various species of bacteria (Jiang, Bikard, Cox, Zhang, & Marraffini, 2013), as well as, cultured human cancer cell lines and human pluripotent stem cells (Cho, Kim, Kim, & Kim, 2013; Cong et al., 2013; Jinek et al., 2013; Mali, Yang, et al., 2013). Around the same time, our group reported efficient genome editing in zebrafish using CRISPR/Cas, demonstrating its potential in a whole multicellular organism (Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). Since then, the platform has proven its effectiveness and versatility in editing the genes of various flora and fauna, only a sample of these being

![Figure 18.1](image_url)

**Figure 18.1** A graphic representation of DNA targeting by sgRNA-guided Cas9. An sgRNA, consisting of a 20-nt crRNA spacer and a tracrRNA tail, guides the *S. pyogenes*-derived Cas9 endonuclease to bind to and unwind a specific 20-nt genomic target site. The target site should be in the form of 5'-N20-NGG, where NGG is the PAM sequence (highlighted in yellow (light gray in the print version)). The top and bottom strands of the genomic DNA are then cleaved by the RuvC-like nuclease domain and the HNH nuclease domain of Cas9 (indicated by the “scissors”) to produce a DNA double-strand break (DSB) approximately three base pairs proximal to the PAM.
yeast, rice, wheat, C. elegans, silk worms, fruit flies, frogs, mice, and non-human primates (DiCarlo et al., 2013; Friedland et al., 2013; Nakayama et al., 2013; Niu et al., 2014; Shan et al., 2013; Wang, Yang, et al., 2013; Wang, Li, et al., 2013; Yu et al., 2013). It is very rare in biology for a single biotechnology to have the degree of versatility as CRISPR/Cas to work with such effectiveness in the wide scope of organisms that it does, giving this new technology the potential to fulfill many of the research, engineering, and therapeutic goals of the genetic engineering field.

Beyond the extraordinary applicability of CRISPR/Cas, this novel genome-editing platform has also exhibited several key advantages for laboratory use compared to other programmable nuclease systems, such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). The first advantage is the greater ease by which CRISPR/Cas can be designed and implemented. Unlike ZFNs and TALENs which require the complex design of zinc-finger and TALE DNA-binding arrays for every new genomic target site, CRISPR/Cas simply requires changing the 20-nt sgRNA spacer sequence so that it matches the target site (Sander & Joung, 2014). The second advantage of CRISPR/Cas is its comparable or greater genome-editing efficiencies than that of ZFNs or TALENs. In general, CRISPR/Cas functions with greater consistency, efficacy, and less toxicity than lab-produced ZFNs (Cornu et al., 2008; Maeder et al., 2008; Ramirez et al., 2008), and they are likely to be more effective at targeting methylated genomic sites compared to TALENs (Hsu et al., 2013). Although the success rate and mutation efficiency of CRISPR/Cas in human cells and in zebrafish appear to be comparable to those of TALENs, CRISPR/Cas is far superior than TALENs in its capability for multiplex genome editing. It has been shown that high-efficiency multiplex genome editing can be achieved using CRISPR/Cas by simply combining Cas9 with multiple sgRNAs (Cong et al., 2013; Guo et al., 2014; Jao, Wente, & Chen, 2013; Ma, Chang, et al., 2014; Ma, Shen, et al., 2014; Mali, Yang, et al., 2013). However, multiplex genome editing using several ZFN or TALEN pairs carries the risk of exacerbating off-target effects by the cross reaction between nuclease pairs (Sollu et al., 2010). In light of these various advantages of the CRISPR/Cas system over previous programmable nuclease platforms, CRISPR/Cas, also known as RNA-guided nucleases, have rapidly risen to become the flagship of contemporary genome-editing technologies.
1.4. The zebrafish animal model and CRISPR/Cas

The zebrafish is a powerful and tractable animal model for functional genomics analysis, the study of human disease pathogenesis, as well as, for the discovery and development of new drugs (Campbell, Hartjes, Nelson, Xu, & Ekker, 2013; Helenius & Yeh, 2012; Lieschke & Currie, 2007). The key strength of the zebrafish model lies in its intermediate evolutionary relationship to humans, between mammalian model systems, such as mice, on one hand, and invertebrate model systems, such as *Drosophila* and *C. elegans*, on the other. The zebrafish has an upper hand over invertebrate models due to its common vertebrate ancestry with humans. This closer ancestry gives the zebrafish greater genetic and anatomical similarity to humans than invertebrates, meaning that orthologous genes carry similar functions as in humans, and most of the organ systems and structures between zebrafish and humans are homologous (Kettleborough et al., 2013; Lieschke & Currie, 2007; Santoriello & Zon, 2012). Due to these genetic and anatomical similarities, various zebrafish models have been developed to study the pathogenesis of human diseases, ranging from genetic disorders such as Duchenne muscular dystrophy and forms of cardiomyopathy (Bassett et al., 2003; Kawahara et al., 2011; Xu et al., 2002), to acquired diseases, such as melanoma and tuberculosis (Cambier et al., 2014; Ceol et al., 2011; Patton et al., 2005; Swaim et al., 2006; White et al., 2011).

Conversely, though the mouse model exhibits greater molecular and anatomical similarity with humans due to their shared mammalian ancestry, the zebrafish carries many key advantages over mouse models due to its non-mammalian features. Because zebrafish reproduce by external fertilization, all stages of zebrafish embryogenesis are accessible to the researcher for study, unlike mammals wherein embryogenesis occurs within the body. This benefit combined with the natural optical transparency of the zebrafish allows for real-time observation of studied processes by fluorescent reporters with great ease (Ignatius & Langenau, 2011; Moro et al., 2013; Pantazis & Supatto, 2014; Weber & Koster, 2013). These observational qualities, in conjunction with the relative size, rapid development, and fecundity of zebrafish compared to mice, enables low animal maintenance and husbandry infrastructure expenses that allow the affordability of high-throughput, whole-animal zebrafish drug screens and reverse-genetic experimentation at a scale simply unfeasible with mouse models (Kari, Rodeck, & Dicker, 2007; Peal, Peterson, & Milan, 2010). Therefore, the zebrafish model serves...
as an especially prime candidate for the application of state-of-the-art genome-editing technologies such as CRISPR/Cas.

As mentioned above, Hwang et al. was first to demonstrate that the CRISPR/Cas genome-editing platform could be adapted in vivo in zebrafish by using it to introduce site-specific insertion/deletion (indel) mutations with mutation frequencies between 24% and 59% at 8 out of 10 tested genes (Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). Interestingly, some of the successfully mutated target sites were within genomic regions previously untargetable by TALENs. Thus, this pioneer study displayed the robustness and power of the CRISPR/Cas platform in zebrafish. Furthermore, CRISPR/Cas-induced indel mutations were later shown to be heritable with transmission rates up to ~100%, opening the possibility for using CRISPR/Cas to create genetic knock-out lines for specific genes (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013). The goal of Section 2 of this chapter will be to discuss the methodologies that have been developed since these initial studies for the generation of targeted indel mutations in zebrafish using CRISPR/Cas.

It has recently been shown that the co-injection of several sgRNAs targeting multiple genomic loci can result in the simultaneous generation of DSBs at multiple genomic target sites. This mechanism may cause random-length insertion/deletion (indel) mutations (red asterisk) at the target site (A). Additional approaches can be used to create predetermined sequence modifications. Linear donor DNA fragments containing a desired functional cassette without any sequence homology to the target locus may be inserted into the target site during DNA repair (B). Moreover, donor DNA containing homologous sequences of the target locus, in the form of small single-stranded oligonucleotides (C) or plasmid DNA (D), may be recombined with the genomic DNA and replace the target site sequence.

Figure 18.2 Cas9-mediated genome editing. The RNA-guided Cas9 endonuclease can induce DSBs at its genomic target site. Subsequently, a DSB may be repaired by a non-homologous end-joining (NHEJ) repair mechanism. This mechanism may cause random-length insertion/deletion (indel) mutations (red asterisk) at the target site (A). Additional approaches can be used to create predetermined sequence modifications. Linear donor DNA fragments containing a desired functional cassette without any sequence homology to the target locus may be inserted into the target site during DNA repair (B). Moreover, donor DNA containing homologous sequences of the target locus, in the form of small single-stranded oligonucleotides (C) or plasmid DNA (D), may be recombined with the genomic DNA and replace the target site sequence.
of multigene mutations in zebrafish embryos, demonstrating even further the power by which CRISPR/Cas can generate targeted indel mutations (Jao et al., 2013; Ota, Hisano, Ikawa, & Kawahara, 2014). Nevertheless, CRISPR/Cas can also be used in zebrafish for purposes beyond the generation of indels (Fig. 18.2). CRISPR/Cas has been used to create small but precise sequence modifications such as point mutations, to integrate long DNA fragments at target sites, and to facilitate long-range chromosomal deletions and inversions (Fig. 18.2). As the genome-editing repertoire of the CRISPR/Cas system in zebrafish continues to rapidly grow, it will be the goal of Section 3 of this chapter to discuss the other available genome-editing strategies beyond the introduction of indels.

2. TARGETED GENERATION OF INDEL MUTATIONS

2.1. Cas9 modification and delivery platforms

The most studied and commonly implemented version of Cas9 endonuclease used for CRISPR/Cas genome editing is SpCas9. This is in part due to its short PAM 5'-NGG which is simpler than that of many other Type II Cas9 nucleases (Westra et al., 2012). However, because of the innate differences in cellular contexts between prokaryotic and eukaryotic systems, this bacterial Cas9 must be modified for in vivo eukaryotic experimentation.

In the preliminary studies implementing CRISPR/Cas in vivo in cultured human cells, SpCas9 was human-codon optimized, allowing the primary structure of SpCas9 to be encoded by codons preferentially used by human cells in order to boost SpCas9 translation efficiency (Cho et al., 2013; Cong et al., 2013; Mali, Yang, et al., 2013). Also in these experiments, one or more commonly used nuclear localization signals (NLS), such as the SV40 NLS, were added to one or both sequence termini of the Cas9 protein to facilitate endonuclease entrance into the eukaryotic nucleus. In our published studies applying CRISPR/Cas to zebrafish, we used an SpCas9 vector (pMLM3613) containing the natural noncodon optimized SpCas9 sequence and an NLS attachment to our construct’s C-terminus (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013; Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). Though we did not use a codon-optimized version of SpCas9 in our initial studies, our studies demonstrated that a simple NLS attachment to natural SpCas9 suffices for CRISPR/Cas to efficiently generate indel mutation rates up to ~60% in zebrafish (Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013).

Since the time of our initial publications, we have started using a version of SpCas9 that has been optimized for human codon usage. Based upon our
experiences, we have consistently seen higher somatic mutation frequencies in zebrafish with this codon-optimized SpCas9 version (pJDS246), and therefore recommend using this version over natural SpCas9. In addition, we have found that the activity of pJDS246 is comparable to a zebrafish codon-optimized SpCas9 (pCS2-nCas9n) (Gonzales, unpublished results), which was reported to produce indel mutagenesis rates up to \textasciitilde 75–99\% at various loci (Jao et al., 2013). Last, it is also worth considering the possibility that other modified Type II Cas9 orthologs besides SpCas9 may provide more potent Cas9 options in the future (Esvelt et al., 2013).

All of the Cas9 plasmids mentioned above can be ordered from Addgene (http://www.addgene.org/CRISPR/). After receiving the Cas9-containing plasmid, it should be linearized by the appropriate restriction enzyme and then \textit{in vitro} transcribed to produce Cas9 mRNA. Most Cas9 plasmids contain either a T7 or SP6 promoter upstream of the Cas9 sequence to allow for standard \textit{in vitro} transcription. In order for \textit{in vitro}-transcribed Cas9 mRNA to be translated efficiently in zebrafish embryos, the mRNAs should have a 5'-cap and a 3'-poly(A) tail. Though most published papers to-date implement CRISPR/Cas in zebrafish by co-injecting Cas9 mRNA and sgRNA together into zebrafish embryos, it has recently been argued that the direct injection of preformed Cas9 protein–sgRNA complexes may be more advantageous because it eliminates the need for Cas9 translation before CRISPR/Cas can start functioning. However, results from these experiments are conflicting as to whether this method can more consistently produce efficient site-directed mutagenesis than direct RNA injections (Gagnon et al., 2014; Sung et al., 2014). Nonetheless, the study by Gagnon et al. strongly suggests that the injection of such complexes can raise the indel mutation rates of normally weak sgRNAs up to approximately sixfold.

\textbf{Protocol for preparation of SpCas9 mRNA for microinjection}

1. Linearize the human-codon optimized SpCas9 vector, pJDS246 (Addgene, Cambridge, MA), with the \textit{PmeI} restriction enzyme (New England Biolabs, Ipswich, MA) by setting up the following reaction: 5 μg of pJDS246 vector DNA, 10 μL of 10× CutSmart™ Buffer (New England Biolabs), 1 μL of \textit{PmeI} (10 units/μL), and sterile deionized water to a total volume of 100 μL. Add \textit{PmeI} last to the reaction mixture. Incubate the reaction at 37 °C for 3 h to overnight to ensure complete linearization.

2. Purify the \textit{PmeI}-cut vector using Qiagen’s QIAquick PCR Purification kit and elute with 25 μL of EB Buffer. Measure the vector DNA
concentration with a spectrometer. Run 100 ng of uncut and cut vector DNA on a 1% wt/vol agarose gel to confirm complete digestion of the vector sample. The purified vector sample can be stored at $-20\, ^\circ C$.

3. **In vitro** transcribe capped and poly(A)-tailed SpCas9 mRNA using a mMESSAGE mMACHINE® T7 Ultra kit (Life Technologies, Beverly, MA). First, thaw the $2 \times$ NTP/ARCA and $10 \times$ T7 Reaction Buffer solutions at room temperature while keeping the T7 Enzyme Mix on ice at all times. Put the $2 \times$ NTP/ARCA solution on ice as soon as it has thawed. Once the $10 \times$ T7 Reaction Buffer has thawed, vortex it to redissolve any precipitate. Next, set up the transcription reaction by mixing the listed reagents in a nuclease-free microfuge tube in the following sequence: 5 µL of $2 \times$ NTP/ARCA, 1 µL of $10 \times$ T7 Reaction Buffer, 1 µL of T7 Enzyme Mix, and then 3 µL of linearized SpCas9 vector (from Step 2). Gently flick and briefly microfuge the tube to collect the reaction mixture at the bottom of the tube. Incubate the tube at $37\, ^\circ C$ for 3 h to overnight for **in vitro** transcription to proceed. After the transcription step, add 1 µL of TURBO DNase to the reaction mixture. Gently flick and briefly microfuge the tube to mix. Incubate the tube at $37\, ^\circ C$ for 30 min for DNA removal.

4. Prepare the poly(A) tailing reaction master mix by combining the following reagents supplied in the same kit in a nuclease-free microfuge tube: 10 µL of $5 \times$ EPAP Buffer, 2.5 µL of 25 mM MnCl$_2$, 5 µL of ATP Solution, and 21.5 µL of nuclease-free water. After the TURBO DNase incubation step is complete (end of Step 3), add the poly(A) tailing reaction master mix to the reaction mixture. Gently flick and briefly microfuge the tube to mix. Aliquot 2 µL of this new mixture to a clean nuclease-free tube labeled “−poly(A)” and store this tube at $-20\, ^\circ C$ for later gel analysis. Next, add 2 µL of E-PAP Enzyme to the reaction mixture. Gently flick and briefly microfuge the tube to mix. Incubate the reaction mixture at $37\, ^\circ C$ for 1–2 h for poly(A) tailing reaction to proceed.

5. After poly(A) tailing is complete, aliquot 2 µL of the reaction mixture to another clean nuclease-free tube labeled “+poly(A)” and store this tube at $-20\, ^\circ C$ for later gel analysis. Then, add 25 µL of Lithium Chloride Precipitation Solution to the remaining reaction mixture. The volume of Lithium Chloride Precipitation Solution added should be half the volume of the reaction mixture. Thoroughly mix the solution and incubate it at either 0.5–1 h on dry ice, or preferably overnight at $-20\, ^\circ C$ for a greater overall mRNA yield.
6. During the mRNA precipitation step, add 5 μL of the Formaldehyde Loading Dye into the 2 μL “−poly(A)” and “+poly(A)” aliquots from before and after the poly(A) tailing reaction (from Steps 4 and 5). Run the samples on a 1% wt/vol agarose gel to check for successful poly(A) tailing by looking for an upshift in the “+poly(A)” sample relative to the “−poly(A)” sample.

7. After the mRNA precipitation step, spin the sample in a microcentrifuge at >10,000 × g, 4 °C for 30 min. After the spin, check for an opaque white mRNA pellet at the bottom of the tube. Carefully aspirate the supernatant without dislodging the pellet. Next, add 1 mL of RNase-free 70% ethanol to the tube and wash the pellet by inverting the tube several times. Centrifuge the tube at >10,000 × g, 4 °C for 15 min. Again, check for the pellet at the bottom of the tube. Aspirate as much of the supernatant as possible without dislodging the pellet so that the pellet will air-dry quickly. Leave the tube with the lid open in a fume hood until all of the supernatant has evaporated and the pellet becomes dry and translucent.

8. Dissolve the SpCas9 mRNA pellet with 15 μL of non-diethylpyrocarbonate (DEPC)-treated, nuclease-free water. As soon as the mRNA pellet completely dissolves, put the tube on ice. Measure the dissolved SpCas9 mRNA concentration using a spectrometer (the yield is typically 1000–2000 ng/μL). Aliquot the SpCas9 mRNA into multiple nuclease-free microfuge tubes (~1500 ng/tube) to prevent freeze–thaw cycles. Store these tubes at −80 °C.

2.2. Single-guide RNA design considerations

The sgRNA design that we use is a ~100-nt sequence in which the first 20 nucleotides interact with the complementary strand of the target site, while the remaining portion interacts with SpCas9 (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013; Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). This ~100-nt sgRNA has a longer tracrRNA region compared to the sgRNA first described in the in vitro study done by Jinek et al. (2012), and it appears to be more effective in vivo compared to sgRNAs that have shorter tracrRNA regions (Jinek et al., 2012, 2013). This sgRNA design is the most common sgRNA design in use (Sander & Joung, 2014), with the same or similar sgRNA design being used in other published zebrafish studies (Auer, Duroure, De Cian, Concordet, & Del Bene, 2014; Chang et al., 2013; Hruscha et al., 2013; Jao et al., 2013).
To express sgRNAs in early stage zebrafish embryos, the sgRNA is usually in vitro transcribed and then microinjected. The sgRNA should not have a 5′-cap or a 3′-poly(A) tail, and the sgRNA vectors used for sgRNA production should have a T7 or SP6 promoter. The transcribed sgRNA can recognize any DNA target in a 5′-GG-N_{18}-NGG-3′ format, with the 5′-GG required for T7-driven transcription, and the “NGG” being the *S. pyogenes* PAM. Thus, the theoretical targeting range is 1 site for every 128 bp of DNA (Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). However, we have shown that sgRNA targeting can often tolerate 2-nt mismatches at its 5′-end, extending the theoretical targeting range to 1 site for every 32 bps (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013). In addition, we had initially proposed to loosen the constraint of the T7 promoter, which would enable targeting of sequences of the form 5′-(G/A)(G/A)-N_{18}-NGG-3′. Nonetheless, a recent report by Gagnon et al. suggests that any change to the 5′-GG can reduce sgRNA efficiency. This result is likely due to T7-driven transcription errors because in all three cases tested, the 5′-GA sgRNAs transcribed by the SP6 promoter showed similar activities to their 5′-GG sgRNA counterparts (Gagnon et al., 2014). Thus, the SP6 promoter may be more flexible than the T7 promoter, tolerating a G→A change at the second position from the transcription start site (Helm, Brule, Giege, & Florentz, 1999; Imburgio, Rong, Ma, & McAllister, 2000; Kuzmine, Gottlieb, & Martin, 2003).

At least two studies have proposed guidelines for choosing sgRNA sequences that are more likely to be effective (Gagnon et al., 2014; Wang, Wei, Sabatini, & Lander, 2014). Together, these studies suggest that the last 1–4 nucleotides of the spacer region should preferably be purines. In addition, the GC content of the spacer region near the PAM should be >50%, but not too high. Although these guidelines may be taken into consideration when there are many sites to choose from, they do not necessarily guarantee success or failure of a sgRNA as shown by the authors and our data (Gagnon et al., 2014; Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013; Wang et al., 2014).

In addition to the above sgRNA design considerations, the location of the sgRNA target site within a gene must also be considered. When designing a sgRNA to produce a gene knock-out, for example, a suitable genome browser (e.g., http://useast.ensembl.org) in combination with free sgRNA design programs (e.g., http://zifit.partners.org/ZiFiT/) should be used to examine the regions of the gene that one wishes to target. Ideally, the sgRNA target site should be as far upstream as possible within the gene’s open reading frame to ensure that an introduced indel can disrupt almost
the entire protein. However, in addition to this targeting principle, one must also search for alternative splice variants for a given gene, as well as annotated alternative translational start sites. If there are any of these additional confounding factors, the same targeting principle can be applied by designing a sgRNA target site after the most downstream annotated alternative start site within the most upstream exon shared by all alternative transcripts. Upon choosing a sgRNA target site, complementary oligonucleotides containing the designed sgRNA spacer region can be ordered, annealed, and inserted into plasmids for cloning, restriction linearization, and subsequent T7 or SP6 in vitro transcription. The sgRNA expression vector used for our previous publications (pDR274) was originally developed by the Joung lab (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013; Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013). The pDR274 vector and those from other labs are available at the non-profit Addgene (http://www.addgene.org/CRISPR/). In addition, there are now published cloning-free methods for sgRNA synthesis that may make sgRNA production quicker (Cho et al., 2013; Gagnon et al., 2014; Hruscha et al., 2013).

Despite the remarkable success of CRISPR/Cas, a potential drawback of the current platform is the possibility of off-target effects due to sgRNA mistargeting. In published studies within human cells, sgRNA mistargeting has been demonstrated to cause indel frequencies at off-target sites at rates up to ~125% compared to its on-target sites (Fu et al., 2013) and has even been shown to inadvertently induce large chromosomal deletions at several loci (Cradick, Fine, Antico, & Bao, 2013). Although earlier CRISPR/Cas studies had proposed the theory of a seed sequence (Cong et al., 2013; Jiang et al., 2013; Jinek et al., 2012), a ~10–12-bp mismatch intolerant region directly adjacent to the target loci’s PAM motif, more recent studies that intentionally investigated sgRNA off-target effects have shown the reality to be much more complicated. Though sequence homology in the first ~10–12 bps directly adjacent to the PAM are relatively more important, all base pairs at a target site, including those in the seed region, confer varying degrees of sgRNA targeting specificity in an sgRNA-specific manner (Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). Clearly, increasing the number of mismatches reduces the efficiency of target cleavage at a given site, especially as the mismatches become more proximal to the PAM (Fu et al., 2013; Hsu et al., 2013). One study suggests that rC:dC base-pairing causes the greatest destruction to sgRNA–Cas9 targeting activity (Hsu et al., 2013). Another study suggests that shorter sgRNA designs may be less active but more specific than longer sgRNAs (Pattanayak et al., 2013). A couple of
studies have also shown that in addition to the 5′-NGG PAM, Cas9 can sometimes target sequences that have a 5′-NAG or 5′-NNGGN PAM (Hsu et al., 2013; Jiang et al., 2013).

Nevertheless, in spite of this progress, exceptions have been demonstrated to these principles. For example, recent zebrafish publications have reported off-target cleavage by sgRNAs at multiple loci, with one study exhibiting detectable cleavage by T7 endonuclease I (T7EI) mismatch assays at an off-target site containing only two mismatches outside the seed region (Jao et al., 2013), while in another study, next-generation sequencing identified indel rates ~1.0–2.5% at off-target sites containing up to six base-pair mismatches at their 5′ most distal ends (Hruscha et al., 2013). On the other hand, not all sgRNAs are promiscuous, and not all potential off-target sites are mistargeted by sgRNAs (Fu et al., 2013; Hruscha et al., 2013; Hsu et al., 2013; Jao et al., 2013; Pattanayak et al., 2013). In light of the complexity regarding sgRNA off-target effects and the lack of in vivo genome-wide bioinformatics studies done on this topic, sgRNA targeting zebrafish genomic loci should be designed in conjunction with available software to minimize potential specificity issues (Bae, Park, & Kim, 2014; Hsu et al., 2013; O’Brien & Bailey, 2014; Xiao et al., 2014).

Other recently developed strategies to optimize or go beyond the standard Cas9–sgRNA platform should be considered. One strategy involves simply lowering the concentrations of Cas9 and sgRNA injected into the zebrafish embryo, since it has been shown in human cells that reduced concentrations lower off-target indel rates, although on-target indel rates will also diminish to varying degrees (Fu et al., 2013; Hsu et al., 2013). Another strategy uses truncated sgRNAs (tru-gRNAs) that contain shortened ~17 to 18-nt spacer regions, which have been shown to maintain on-target efficiencies while reducing indel frequencies at off-target sites up to ~5000-fold compared to standard sgRNAs (Fu, Sander, Reyon, Cascio, & Joung, 2014). The implementation of D10A Cas9 nickases guided by paired sgRNAs targeting adjacent genomic sites have likewise been shown to reduce the off-target effects of CRISPR/Cas up to ~1500-fold in human cells without sacrificing on-target activity. These work by doubling the number of base pairs required for double-strand DNA cleavage and by relying on less erroneous DNA repair for any off-target DNA single-strand nicks (Cho et al., 2014; Mali, Aach, et al., 2013; Ran et al., 2013). Recently, two independent groups developed a crRNA-guided FokI nuclease platform, fusing catalytically inactive Cas9 with a FokI nuclease domain. Taking advantage of the RNA-guided capability of Cas9 and the dimerization
requirement of the FokI nucleases, this platform recognizes extended target sequences and has been shown to be more specific compared to wild-type Cas9 and paired nickases (Guilinger, Thompson, & Liu, 2014; Tsai et al., 2014). Due to differences in the availability of sgRNA target sites, regardless of which of the options are chosen for generating targeted indels, the varied strategies mentioned should be considered on a case-by-case basis.

**Protocol for preparation of sgRNAs for microinjection:**

1. Linearize the sgRNA vector, pDR274 (Addgene, Cambridge, MA), with the BsaI restriction enzyme (New England Biolabs) by setting up the following reaction: 5 μg of pDR274 vector DNA, 10 μL of 10× CutSmart™ Buffer (New England Biolabs), 1 μL of BsaI (10 units/μL), and sterile deionized water to a total volume of 100 μL. Add BsaI last to the reaction mixture. Incubate the reaction at 37 °C for at least 3 h to overnight to ensure complete digestion.

2. Purify the BsaI-cut vector using Qiagen’s QIAquick PCR Purification kit and elute with 25 μL of EB Buffer. Measure the vector DNA concentration with a spectrometer. Run 100 ng of both uncut and cut vector DNA on a 1% wt/vol agarose gel to confirm complete digestion of the vector DNA sample. Dilute the vector sample with EB Buffer to a final concentration of 5–10 ng/μL. Keep a stock of this purified cut vector sample at −20 °C for future use.

3. Design a pair of 22-nt DNA oligos that contain within them complementary sequences that correspond to the 18-bp sequence adjacent to the PAM in a sgRNA target site. These oligos when annealed together by these sequences can be inserted into the pDR274 vector for T7 promoter-driven transcription of a given sgRNA. For insertion into the pDR274 vector, Oligo #1 also contains at its 5’-end a TAGG, which is a part of the T7 promoter sequence, and Oligo #2 contains at its 5’-end an AAAC. These short sequences are necessary to provide the annealed oligo the sticky ends necessary for unidirectional insertion into BsaI-cut pDR274. Investigators can also use free online software ZiFiT Targeter (http://zifit.partners.org/ZiFiT/) to generate the oligo sequences needed for any specified target site.

4. Obtain these DNA oligos from a reliable source. Anneal these DNA oligos together by combining 45 μL of 100 μM Oligo #1, 45 μL of 100 μM Oligo #2, and 10 μL of 10× NEBuffer 2.1 (New England Biolabs) in a microfuge tube. Place the tube in a removable heat block at 95 °C for 5 min. Then, remove the heat block and allow it to
gradually cool until it reaches below 37 °C. The annealed oligos can be stored at −20 °C.

5. **Ligate the annealed oligos (from Step 4) into the linearized pDR274 vector (from Step 2) by setting up the following reaction and letting it incubate at room temperature for 1 h or at 4 °C overnight:**

   - 1 µL of the annealed oligos, 1 µL of purified BsaI-cut pDR274 vector,
   - 2.5 µL of 2 × Rapid Ligation Buffer (Promega, Madison, WI), and
   - 0.5 µL of T4 DNA Ligase (Promega). If you use a different ligase and ligation buffer, follow the ligation condition recommended by the manufacturer. Transform chemically competent bacterial cells with 5 µL of the ligation product. After transformation is complete, spread the cells on a LB/kanamycin plate and incubate it at 37 °C overnight.

6. **Pick at least three colonies per transformation and inoculate each colony in a culture tube containing 1.5 mL of LB/kanamycin. Incubate the culture tubes in a shaker at 37 °C overnight. The next day, extract the plasmid DNA using a plasmid DNA miniprep kit.** Send the extracted plasmid DNA samples for sequencing using a M13F primer. Confirm whether the sequenced plasmid samples have the correct elements in the appropriate order: (from 5′ to 3′) a T7 promoter, the customized target sequence, and the appropriate sgRNA backbone sequence (Hwang, Fu, Reyon, Maeder, Tsai, et al., 2013).

7. **Re-inoculate a bacteria sample (from Step 6) that contains the correct sgRNA vector.** Miniprep the sgRNA vector using a gravity-flow column-based QIAGEN Plasmid Mini kit, and measure the vector DNA concentration using a spectrometer. Linearize the sgRNA vector with the *DraI* restriction enzyme (New England Biolabs) by setting up the following reaction: 5 µg of customized sgRNA vector DNA, 10 µL of 10 × CutSmart™ Buffer (New England Biolabs), 1 µL of *DraI* (10 units/µL), and sterile deionized water to a total volume of 100 µL. Add *DraI* last to the reaction mixture. Incubate the tube at 37 °C for at least 3 h to overnight to ensure complete digestion.

8. **Purify the *DraI*-cut vector using Qiagen’s QIAquick PCR Purification kit and elute with 25 µL of EB Buffer.** Measure the vector DNA concentration with a spectrometer. Run 100 ng of both uncut and cut vector DNA on a 3% wt/vol agarose gel to confirm complete digestion of the vector DNA sample. The digested sgRNA vector should exhibit only two fragments at 1.9 kb and 282 bp. The smaller DNA fragment contains the T7 promoter and the customized sgRNA sequence. Gel purification is not necessary. The purified DNA can be stored at −20 °C.
9. *In vitro* transcribe the sgRNA using the MAXIscript\textsuperscript{©} T7 kit (Life Technologies) by setting up the following reaction in a nuclease-free microfuge tube: 1 μg of purified *Dra*I-cut sgRNA vector DNA (from Step 8), 2 μL of 10× Transcription Buffer, 1 μL of each of the ATP, UTP, GTP, CTP 10 mM solutions, 2 μL of T7 Enzyme Mix, and nuclease-free water to a total volume of 20 μL. Incubate the reaction mixture at 37 °C for 2 h to overnight for sgRNA transcription. After sgRNA transcription, add 2 μL of TURBO DNase to the reaction mixture and incubate at 37 °C for 30 min to remove DNA from the sample. Next, add 1 μL of 0.5 M EDTA to the reaction mixture. Gently flick and briefly microfuge the tube to stop the reaction. Subsequently, add 30 μL of nuclease-free water and 5 μL of 5 M ammonium acetate to the reaction mixture and thoroughly mix. Last, add 150 μL of 100% RNase-free ethanol and thoroughly mix. Incubate the sample on dry ice for 0.5–1 h, or preferably at −20 °C overnight for a greater overall sgRNA yield.

10. After the sgRNA precipitation step, spin the sample in a microcentrifuge at >10,000 × g, 4 °C for 30 min. After the spin, check for an opaque white sgRNA pellet at the bottom of the tube. Carefully aspirate the supernatant without dislodging the pellet. Next, add 1 mL of RNase-free 70% ethanol to the tube and wash the pellet by inverting the tube several times. Centrifuge the tube at >10,000 × g, 4 °C for 15 min. Again, check for the pellet at the bottom of the tube. Aspirate as much of the supernatant as possible without dislodging the pellet so that the pellet will air-dry quickly. Leave the tube with the lid open in a fume hood until all of the supernatant has evaporated and the pellet becomes dry and translucent.

11. Dissolve the sgRNA pellet with 11 μL of non-DEPC-treated, nuclease-free water. As soon as the sgRNA pellet completely dissolves, put the tube on ice. Measure the dissolved sgRNA concentration using a spectrometer (the yield is typically 100–200 ng/μL). Aliquot the sgRNA into multiple nuclease-free microfuge tubes (~100 ng/tube) to prevent freeze–thaw cycles. Store the tubes at −80 °C. Check the sgRNA integrity by mixing 1 μL of the sgRNA with 5 μL of Formaldehyde Loading Dye and running it on a 3% wt/vol agarose gel containing 0.2–0.5 μg/mL of ethidium bromide. There should be a distinct sgRNA band without smearing.
2.3. Introduction and identification of Cas9–sgRNA-induced indels

After selecting and preparing the appropriate Cas9 and sgRNA(s) for a given experiment, CRISPR/Cas genome editing can be implemented by zebrafish embryo microinjections. Zebrafish embryos must be collected and immediately injected at the one-cell stage before the first round of cellular mitosis to facilitate homogenous distribution of CRISPR/Cas components among all future daughter cells. As previously discussed, the most common practice uses in vitro-prepared Cas9 mRNAs for injections, though studies using Cas9 protein–sgRNA complexes have also been reported (Gagnon et al., 2014). Various concentrations of Cas9 mRNA and sgRNAs have been used in different studies. For sgRNAs that show a high on-target mutation rate, reducing the concentration of Cas9 may reduce potential off-target activities. On the other hand, the quality of in vitro-transcribed Cas9 mRNA and sgRNA have a direct influence on the observed on-target activity.

Depending on the particular purpose of the study, indel mutation rates may be determined by the T7EI mismatch assay, PCR subcloning and sequencing, high resolution melt analysis (Dahlem et al., 2012), and/or next-generation sequencing. Due to its speed and independence of any sophisticated instrumentation, the T7EI assay is probably the most widely used method. T7EI analysis relies upon the ability of T7EI to recognize and cleave nonperfectly annealed DNA. In this method, PCR amplicons of targeted genomic loci are denatured and gradually cooled to allow for partial hybridization between differentially sized indel-containing single-stranded DNA fragments. These partially hybridized PCR strands will contain mismatched regions at the designed target sites that may be cleaved by T7EI to produce two fragments of predictable lengths. Upon electrophoresis, the percentage of cleaved PCR products can be measured, and the estimated NHEJ percentage rates can be calculated using the calculations given by Guschin et al.: % target site indel rate = 100 × (1 − (1 − fraction cleaved)\(^{1/2}\)) (Guschin et al., 2010). Nonetheless, the estimated detection limit of this assay is above ~3% (Hwang, Peterson, & Yeh, 2014). In addition, genetic polymorphisms surrounding the target site may cause false positives and should be carefully controlled. Moreover, if a low amount of or no cleaved product is detected, sequencing of the PCR products should be considered.
3. OTHER TARGETED GENOME-EDITING STRATEGIES

3.1. Precise sequence modifications mediated by single-stranded oligonucleotides

Previously, it was found that in the presence of DSBs, single-stranded oligonucleotides (ssODNs) flanked by two arms with tens of base pairs of homology to the sequences surrounding the breaks may be co-opted to introduce predesigned sequence modifications in human cells (Chen et al., 2011). Later, a similar strategy was used to introduce small yet precise targeted insertions in conjunction with TALENs (Bedell et al., 2012). We and others have recently shown that small predetermined sequences may be targetedly inserted in zebrafish using CRISPR/Cas and ssODNs containing \( \sim 20–50 \)-nt homology arms (Chang et al., 2013; Hruscha et al., 2013; Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013).

Thus far, this ssODN-mediated CRISPR/Cas method has been used to knock-in \( \sim 30 \)-nt mloxp sites (Chang et al., 2013), as well as, \( \sim 30 \)-nt HA tags at two zebrafish loci (Hruscha et al., 2013). We have used ssODNs to precisely insert up to \( \sim 40 \) nucleotides at different zebrafish loci (Gonzales, unpublished results) and to successfully generate precise point mutations in one zebrafish gene (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013). It should be noted that even though insertion rates can be as high as \( \sim 20\% \), it has been found in all of these studies, including the aforementioned TALEN study, that only a small portion of the sequence changes are free of additional mutations. These results are very different from the results of studies in human cells and mice, wherein sequence modifications introduced by ssODNs are almost always precise (Chen et al., 2011; Kayali, Bury, Ballard, & Bertoni, 2010; Shen et al., 2013). The exact mechanism of DNA repair in zebrafish that mediates the insertion under these conditions is still unclear.

To implement this ssODN-mediated method, we have noticed that, in some cases, sense or antisense ssODNs containing the sequences either homologous or complementary to the sgRNA target site show varied efficiencies (Hwang, Fu, Reyon, Maeder, Kaini, et al., 2013). Nonetheless, the differences in their efficiencies are locus-dependent, and there is no consensus as to which strand is better to use. Interestingly, a recent study shows that ssODNs containing a stop cassette, which if inserted into a gene provides stop codons at all possible reading frames, has allowed this ssODN-mediated technique to serve as an alternative to producing genetic knock-outs by
indel frameshift mutations (Gagnon et al., 2014). Moreover, this study demonstrated the insertion and heritability of these stop cassettes in three zebrafish genes. Therefore, the demonstrated ssODN-mediated insertion strategy provides great versatility to CRISPR/Cas genome editing in zebrafish.

3.2. Targeted integration of long DNA fragments

Targeted genome modifications involving insertions of long DNA fragments can be achieved via homologous recombination (HR) or a homology-independent mechanism such as NHEJ (Fig. 18.3). In zebrafish embryos, NHEJ appears to be responsible for most DNA repair activities, whereas HR is estimated to be at least 10-fold less active (Dai, Cui, Zhu, & Hu, 2010; Hagmann et al., 1998; Liu et al., 2012). Zu et al. were the first to successfully demonstrate HR-mediated insertion in zebrafish by

Figure 18.3 Engineered DNA nucleases can facilitate targeted integration of long DNA fragments via homology-dependent and homology-independent mechanisms. (A) Targeted integration mediated by homologous recombination. In this approach, a plasmid donor DNA containing the DNA cassette (red (dark gray in the print version) box) to be inserted, flanked by several hundred base pairs to a couple kilobase pairs of sequences (gray boxes) upstream and downstream of the genomic target site, is co-injected with the engineered DNA nucleases into zebrafish embryos. The DNA cassette can be inserted into the nuclease target site via homologous recombination. The sequence surrounding the cassette and the joining ends should be precise. (B) Targeted integration mediated by a homology-independent mechanism. In this approach, the donor DNA should be linearized in vitro or in vivo, and need not have sequence homology to the genomic target locus. The linearized donor DNA can be inserted into the nuclease target site, but the sequences at the joining ends will not be precise.
the use of highly efficient TALEN pairs. In their study, the authors reported targeted insertion of EGFP in three zebrafish genes; however, germline transmission was found for only one gene at a low frequency of 1.5% (Zu et al., 2013). Therefore, although HR-mediated gene targeting affords the ability to generate precise sequence changes, this method may not become a widely used approach unless its efficiency can be improved (Beumer et al., 2008; Liu et al., 2012). To date, an example of Cas9-assisted HR-mediated DNA integration has not been reported in zebrafish.

On the other hand, Auer et al. demonstrated efficient homology-independent targeted integration of >5.7-kilobase (kb) DNA fragments in zebrafish using CRISPR/Cas (Auer et al., 2014). In their study, a donor DNA construct containing a modified Gal4 gene was co-injected with Cas9

![Figure 18.4](image)

Figure 18.4 Designs of donor DNAs. A plasmid donor DNA may be linearized in vivo by adding a CRISPR/Cas9 target site (blue (dark gray in the print version) vertical line) to it. This target sequence can be the same as or different from the genomic target sequence (yellow (light gray in the print version) vertical line). (A) In this design, Gal4 and a self-processing peptide E2A will be expressed only if the integrated DNA cassette is in the correct orientation and in the right coding frame. Thus, theoretically one-sixth of the integration events will result in the expression of E2A–Gal4. This is the approach utilized by Auer et al. (B) An alternative approach will be to insert a DNA cassette including a splice acceptor site (SA) into an intronic region. In this approach, the reading frame can be predicted because no indel mutation is introduced in the coding region. Thus, approximately one-half of the integration events will result in the expression of E2A–Gal4.
mRNA and a GFP-targeting sgRNA into zebrafish embryos carrying a tissue-specific GFP reporter and the UAS:RFP transgene (Fig. 18.4). In theory, one-sixth of the targeted integration events would be in-frame, resulting in Gal4 and subsequent RFP expression. Thus, this experimental setup enabled visual detection of in-frame knock-in events by a green-to-red switch. Detection of RFP+ cells in the expected expression domains, which are the same as the GFP+ domains in the noninjected transgenic embryos, would indicate successful integration at the target site. In contrast, detection of RFP+ cells outside the expected expression domains would suggest off-target or random integration. Moreover, the incorporation of the Gal4–UAS system likely enhanced the sensitivity of this assay.

By using this approach, Auer et al. made a number of important findings (Auer et al., 2014). First, they found that the knock-in efficiency by homology-independent repair can be significantly improved if the donor DNA is linearized in the injected embryos rather than in vitro (Fig. 18.4). In vivo cleavage of the plasmid donor DNA was accomplished by adding to the donor construct a sgRNA “bait” sequence that could be cleaved by co-injection of a corresponding sgRNA. By co-injecting a sgRNA with a 66% mutation efficiency and a plasmid donor containing the corresponding bait sequence, they observed a near sevenfold increase of knock-in efficiency compared to the use of an in vitro linearized donor DNA.

Second, the authors demonstrated that the bait sequence could be made either the same or different from the genomic target sequence. While the former strategy simplifies experimental design due to the need for only one sgRNA, the latter allows the flexibility to use the same donor construct for any given target site (Fig. 18.4). Indeed, this latter strategy may be useful because the Cas9-mediated knock-in efficiency is directly related to the targeting efficiency of a given sgRNA. For example, in the study by Auer et al., there were two cases wherein a single sgRNA was used for both the bait and the target gene, and sgRNAs with 66% and 20% mutation rates were able to induce the green-to-red switch in 75% and 15% of the injected embryos, respectively. Therefore, as shown also by the authors, it may be beneficial when using this knock-in strategy to reuse a sgRNA-bait sequence pair that is known to have a high in vivo linearization efficiency in order to increase the knock-in rates at a difficult target site.

Third, Auer et al. found that the germline transmission of the RFP knock-in allele was quite high and that the observed frequencies of identified founder fish could be increased by screening only RFP+ F0 fish. For
example, in a case wherein a single sgRNA with a 66% mutation rate was used, Auer et al. observed founder fish frequencies of 10% and 33–40% by screening all of the fish or only RFP + F0 fish, respectively. Further examination of the F1 fish showed that both single-copy and multiple-copy integrations of the RFP allele could be found. Moreover, despite some low level off-target activities for a single sgRNA they used (probably under 3% at the sites screened), off-target integration was not found in the F0 fish by PCR, Southern blots, or fluorescence.

Overall, this pioneer study significantly extended the power of CRISPR/Cas genome editing and opens the possibility of unprecedented strategies for creating novel functional alleles in zebrafish. Conceivably, one potential means of improving the current system will be to insert DNA cassettes with splice acceptor sites into the intronic regions of target genes (Fig. 18.4). This approach is likely to better guarantee in-frame integrations irrespective of sequence alterations at the joining ends.

3.3. Chromosomal deletions and other rearrangements

Functional analysis of noncoding RNAs, transcriptional enhancers, regulatory elements in promoters, gene clusters, and tandem-duplicated genes require methods that, in a targeted fashion, can delete genomic segments ranging from hundreds of base pairs to several hundred kilobase pairs. Consequently, genome-editing tools that can induce large segmental deletions are of particular interest for zebrafish research due to the prevalence of gene duplications in this model organism (Lu, Peatman, Tang, Lewis, & Liu, 2012).

Using two customized pairs of ZFNs or TALENs to target two distant sequences on the same chromosome, deletions of large DNA segments ranging from several kilobase pairs to megabase pairs have been reported in

![Figure 18.5](image.png)

**Figure 18.5** Cas9-mediated chromosomal rearrangements. (A) Chromosomal deletions, inversions, and duplications may be induced by two sgRNA–Cas9 complexes that target two distant sites on the same chromosome. (B) Chromosomal translocations may be induced by two sgRNA–Cas9 complexes that target two different chromosomes.
cultured mammalian cells (Carlson et al., 2012; Lee, Kim, & Kim, 2010), silkworms (Ma et al., 2012), plants (Qi et al., 2013), and zebrafish (Gupta et al., 2013; Xiao et al., 2013) (Fig. 18.5). Interestingly, re-insertions of the DNA segments cleaved off by these programmable nucleases have also been observed as inversions and duplications (Gupta et al., 2013; Lee, Kweon, Kim, Kim, & Kim, 2012; Qi et al., 2013; Xiao et al., 2013) (Fig. 18.5). Furthermore, targeted chromosomal translocations have also been successfully created between two ZFN or TALEN sites located on two different chromosomes (Brunet et al., 2009; Piganeau et al., 2013; Simsek et al., 2011) (Fig. 18.5). In light of these studies, ZFN and TALEN-mediated chromosomal rearrangements have been applied to cultured cells for the development of human disease models involving genome rearrangements (Piganeau et al., 2013). Consequently, it has become a very attractive prospect to apply the more facile CRISPR/Cas platform for the generation of such disease models (Choi & Meyerson, 2014).

To date, zebrafish chromosomal deletions and inversions mediated by CRISPR/Cas have been shown by only two groups (Ota et al., 2014; Xiao et al., 2013). In one study, Xiao et al. was able to perform chromosomal deletions of two loci with CRISPR/Cas; however, lower deletion rates were reported by CRISPR/Cas compared to those by TALENs for the tested loci, though the reasons for this discrepancy are unclear (Xiao et al., 2013). More recently, Ota et al. demonstrated germline transmission of a 7.1-kb CRISPR/Cas deletion that had been successfully identified in one of eleven screened potential founder fish (Ota et al., 2014). Because of the ease of use and general robustness of the CRISPR/Cas system, it is likely that Cas9-based approaches will play a more prominent role in future studies in zebrafish involving chromosomal deletions and other rearrangements.

### 4. FUTURE DIRECTIONS

Over the past year-and-a-half, the CRISPR/Cas system has proven itself to be a powerful yet facile and efficient genome-editing platform in zebrafish, demonstrating its broad adaptability by its ability to generate targeted indels, exact point mutations, site-specific insertions of varying lengths, and chromosomal rearrangements in a rapid and low-cost manner. In spite of its quick rise to fame, the CRISPR/Cas platform still has several hurdles to overcome, if it is to advance the field of genetic engineering even further. The two most crucial obstacles that the CRISPR/Cas system faces
relate to its targeting range and its specificity. Though the current CRISPR/Cas platform has a relatively widespread targeting range of 1 site in every 8 bps, this is still lower than that of previous programmable nuclease-ases such as TALENs, which has been engineered to theoretically target any sequence in the genome (Joung & Sander, 2013; Lamb, Mercer, & Barbas, 2013). The investigation of other Type II Cas9 orthologs besides that of S. pyogenes, coupled with the application of continuous directed evolutionary efforts (Esvelt, Carlson, & Liu, 2011), will likely provide an avenue to expand the CRISPR/Cas targeting range. With regard to the variable targeting specificity of CRISPR/Cas, global genome-wide off-target studies will be needed to accurately test the effectiveness of current tru-sgRNA, paired Cas9 nickase, and other strategies. Off-target studies will also be needed for the potential development of more reliable methods to increase CRISPR/Cas specificity.

Finally, CRISPR/Cas possesses untapped potential for future zebrafish research. From a genetic engineering standpoint, CRISPR/Cas shows great promise to be used for conditional, tissue-specific genome editing. The development of such a spatially and temporally controlled genome-editing application has yet to be demonstrated by CRISPR/Cas, but its development is certainly underway and will be of tremendous benefit to future disease and functional genomics research when applied to zebrafish. Beyond genome editing, CRISPR/Cas has already been adapted in human cells as a system for transcriptional regulation (Gilbert et al., 2013) and as a means for dynamic imaging of genetic structures (Chen et al., 2013). These adaptations are made possible by deactivating the nuclease activity of Cas9 without removing its sgRNA-mediated targeting function, and in principle, these methods can be extended to zebrafish. It is logical that such engineering could allow CRISPR/Cas to be adapted for purposes of epigenetic editing (Maeder et al., 2013) and modulation of genomic architectures (Deng et al., 2012). Also, a recent publication has shown that orthogonal Cas9 enzymes could be used in bacteria to allow for simultaneous indel targeting and transcriptional repression at two different targets, thus opening up the possibility of combining several CRISPR/Cas capabilities, in parallel, within a single organism (Esvelt et al., 2013). Such a report points to the enormous potential CRISPR/Cas has to one day become a universal, all-purpose biomolecular engineering platform that can revolutionize not only zebrafish research, but could influence all future biomedical pursuits and medical treatments.
ACKNOWLEDGMENTS

This work is supported by the National Institutes of Health (R01 GM088840 and R01 CA140188) and by the Hassenfeld Scholar’s Award.

REFERENCES


Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology, 43(6), 1565–1575 (Research Support, Non-U.S. Gov’t).


Ma, S., Chang, J., Wang, X., Liu, Y., Zhang, J., Lu, W., et al. (2014). CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in bombyx Mori. *Scientific Reports, 4*, 4489. [http://dx.doi.org/10.1038/srep04489](http://dx.doi.org/10.1038/srep04489) (Research Support, Non-U.S. Gov’t).


Simsek, D., Brunet, E., Wong, S. Y., Katyal, S., Gao, Y., McKinnon, P. J., et al. (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal...


