Making designer mutants in model organisms

Ying Peng1,*, Karl J. Clark1,2,3, Jarryd M. Campbell1,2,3, Magdalena R. Panetta4, Yi Guo1,5 and Stephen C. Eckert1,2,3,4,*

ABSTRACT

Recent advances in the targeted modification of complex eukaryotic genomes have unlocked a new era of genome engineering. From the pioneering work using zinc-finger nucleases (ZFNs), to the advent of the versatile and specific TALEN systems, and most recently the highly accessible CRISPR/Cas9 systems, we now possess an unprecedented ability to analyze developmental processes using sophisticated designer genetic tools. In this Review, we summarize the common approaches and applications of these still-evolving tools as they are being used in the most popular model developmental systems. Excitingly, these robust and simple genomic engineering tools also promise to revolutionize developmental studies using less well established experimental organisms.

KEY WORDS: Genome engineering, Transcription activator-like effector nuclease (TALEN), Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated systems (Cas9), Zinc finger nuclease (ZFN), Model organisms

Introduction

Modern developmental biology was born out of the fruitful marriage between traditional embryology and genetics. Genetic tools, together with advanced microscopy techniques, serve as the most fundamental means for developmental biologists to elucidate the logistics and the molecular control of growth, differentiation and morphogenesis. For this reason, model organisms with sophisticated and comprehensive genetic tools have been highly favored for developmental studies. Advances made in developmental biology using these genetically amenable models have been well recognized. The Nobel prize in Physiology or Medicine was awarded in 1995 to Edward B. Lewis, Christiane Nüsslein-Volhard and Eric F. Wieschaus for their discoveries on the ‘Genetic control of early structural development’ using Drosophila melanogaster; and again in 2002 to John Sulston, Robert Horvitz and Sydney Brenner for their discoveries of ‘Genetic regulation of development and programmed cell death’ using the nematode worm Caenorhabditis elegans. These fly and worm systems remain powerful and popular models for invertebrate development studies, while zebrafish (Danio rerio), the dual frog species Xenopus laevis and Xenopus tropicalis, rat (Rattus norvegicus), and particularly mouse (Musculus) represent the most commonly used vertebrate model systems.

Box 1. Common DNA repair pathways

NHEJ (non-homologous end joining) specifies the DNA repair process where double-strand break (DSB) ends are directly ligated without the need for a homologous template. DNA ends processed via the canonical NHEJ pathway are usually protected from significant 5′-end resection by heterodimeric Ku proteins. Many, but not all, NHEJ-based DNA changes result in small insertions or deletions near the DSB cut site.

HDR (homology-directed repair) is the general category of DNA repair pathways using sequence homology to repair DNA lesions. One common feature of different HDR pathways is the initial 5′-end processing of one or both DNA DSBs. Single-stranded ends generated in this manner are used to search for homologous sequences either from another site in the genome or from a foreign (donor) DNA. HDR may require multiple cellular steps, including DNA replication and other processes. Unlike classical homologous recombination (HR), HDR can use short DNA templates as donors, including single-stranded oligonucleotides. HDR is widely used to replace an endogenous locus with foreign DNA in genome engineering applications. HR is the most well studied mode of homology-directed repair, whereas HDR also includes HR-independent pathways. Canonical HR involves nucleotide sequence exchanges between two similar or identical molecules of DNA. The initiation of efficient homologous recombination requires long stretches of sequence homology between recombining DNAs. HR-independent HDR mechanisms include single-strand annealing (SSA), a process whereby two 5′-processed single-stranded ends are joined through base-pair complementation, and break-induced replication (BIR), where one 5′-processed DSB uses its homology sequence as template to initiate DNA replication.

restricted to systems where cell culture systems can be used to generate whole organisms.

Traditional recombinant DNA technology enables molecular biologists to ‘cut and paste’ simpler prokaryotic DNA plasmids with great precision and superb efficiency. However, equivalent applications for studying development in multicellular eukaryotes require novel tools with more stringent sequence specificity and increased versatility, due to higher genomic complexity, as well as a critical efficiency threshold to enable routine applications. Through decades of innovation, promising designer target-specific endonucleases fulfill this important need. In this Review, we will discuss the properties and limitations of different designer endonuclease platforms for the developmental biologist, each of which can be adapted to introduce molecularly distinct site-specific modifications in eukaryotic genomes. We then explore several successful examples of their use. The principles and applications of these designer targeted endonucleases are largely applicable to nearly all genomes tested so far, although with varying efficiencies and limitations. Thus, we will end with a perspective on how these tools can bring a new era of developmental biology: greatly expanding both the depth and scope of potential research avenues, and making this a very stimulating time for the design and execution of diverse experiments to test both long-standing and new hypotheses in the field.

**Designer endonucleases: principle and design**

Designer endonuclease-based genome engineering approaches involve introducing a lesion at an intended site of the genome, resulting in a deletion, insertion or replacement of genomic sequence using cellular repair pathways. The key cellular property for genome engineering is the robust recruitment of DNA repair machinery at a desired locus after a double-strand break. The goal of genome engineering is thus to produce reagents, specifically designer endonucleases, that achieve predictable, high-specificity sequence recognition with excellent efficacy. The main inspiration to overcome this hurdle came from understanding how naturally occurring sequence-specific DNA-binding proteins achieve their specificity in reading double-stranded DNA and then using these principles to generate custom enzymes. At present, three major families of designer endonucleases are commonly used: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided endonucleases.

**Zinc-finger nucleases**

The Cys2-His2 zinc-finger domain, consisting of ~30 amino acids in a stereotypical ββα configuration, is one of the most frequently used DNA-binding motifs found in eukaryotic sequence-specific transcriptional factors. Upon binding of a zinc-finger to its target site, three base pairs in the major groove of the DNA are in close contact with a few amino acids on the surface of the α-helix. As these contacts mediate the sequence recognition specificity of zinc fingers, modifying the key amino acids on the ‘fingers’ can render a degree of selectivity toward a given three base-pair DNA sequence. Proteins constructed with tandem repeats of engineered, sequence-specific DNA-binding zinc fingers constituted the first successful, custom-designed platform to target DNA sequences (Bibikova et al., 2003; Kim et al., 1996; Porteus and Baltimore, 2003). Fusing such a DNA sequence recognition module with a sequence-independent FokI endonuclease domain produced designer endonucleases that were used to introduce site-specific modifications in the human genome (Umov et al., 2005). Two individual zinc-finger nucleases (ZFNs) are required to induce a lesion at a single site due to the dimeric requirement for FokI activity (Fig. 1A). To improve sequence specificity, modifications of the dimer interface on the FokI cleavage domain were made to form obligate heterodimers (Doyon et al., 2011). However, a significant drawback for the widespread application of ZFN technology is the limited binding selectivity conferred by the zinc-finger modules, as well as complex context-dependent interactions between adjacent zinc fingers that can alter binding affinity to the DNA (Cornu et al., 2008; Händel et al., 2009). Designing an efficient ZFN is usually not a trivial task, typically involving multiple rounds of tests and modifications. In addition, a lack of consistency, with different design and assembly platforms, makes this technology less accessible to most laboratories (Carroll, 2011; Kim et al., 2010; Ramirez et al., 2008; Sander et al., 2011). Although most genome engineers have switched their strategies to use newer tools (see below), nearly all of this more recent work is based on foundational experiments by pioneering ZFN scientists.

**Transcription activator-like effector nucleases**

Transcription activator-like effector (TALE) proteins represent a molecularly unique class of transcription factors from the plant pathogenic bacteria *Xanthomonas* that contain a DNA-binding domain consisting of 33-35 amino-acid modular repeats that each recognize a single DNA base pair (Bogdanove and Voytas, 2011). A simple but stringent correspondence code of DNA recognition was discovered such that two hypervariable amino acids in each TAL domain (repeat-variable di-residues, RVD) can distinguish the four different DNA nucleotides at the recognition site (Boch et al., 2009; Moscou and Bogdanove, 2009). Building on the ZFN work, this new DNA-binding platform facilitated the rapid expansion of DNA-binding proteins using novel TAL DNA-binding domains. TALE nucleases (TALENs) are made by fusing consecutive DNA recognition TAL repeats with the type IIS FokI endonuclease domain to achieve a site-specific DNA lesion (Fig. 1B; Bedell et al., 2012; Christian et al., 2010; Li et al., 2011a; Mahfouz et al., 2011). By the end of 2013, more than 200 reports of independent successful TALEN applications had emerged. From this collected dataset, some conclusions about TALEN applications can be drawn. Current TALEN designs and assembly methodology are extremely effective and can be easily adopted in any lab. Moreover, TALENs have in practice very few restrictions with regard to potential sequence targeting applications, and the use of two 14+ base TALE recognition sites plus a spacer ranging from 13 to 20 bp provides single site recognition potential in even the most complex genomes. Importantly, new synthesis platforms assemble TALENs for as little as US$5 in supply costs per TALEN arm (Liang et al., 2014).

**RNA-guided endonucleases – clustered regularly interspaced short palindromic repeats and their associated systems**

Applications using ZFN or TALENs require making target-specific engineering endonuclease constructs involving modular assembly of protein-DNA recognition motifs. Although significant progress has been made to facilitate rapid assembly and cloning (Cermak et al., 2011; Li et al., 2011b; Reyon et al., 2012; Schmid-Burgk et al., 2013), generating a custom engineered protein requires investment and maintenance of plasmid libraries (from 40 to >400 clones), and troubleshooting complex ligations of 6-11 plasmids can be difficult when the reactions are not working well. Recently, an RNA-guided genome engineering system employing components of bacterial adaptive immune response pathway has emerged that does not require custom protein synthesis and instead
uses a unique guide RNA (gRNA) along with a single endonuclease protein (Cas9) (Barrangou et al., 2007; Horvath and Barrangou, 2010). Type II clustered regularly interspaced short palindromic repeats and their associated systems (CRISPR/Cas9 systems) from Streptococcus pyogenes are the most widely used CRISPR/Cas9 genomic engineering platform to date. Target site recognition relies on the Cas9-mediated Watson-Crick base pairing between a short stretch of CRISPR repeat RNA (originally named as ‘spacer’) and one strand of target DNA (known as the ‘protospacer’ sequence) (Jinek et al., 2012; Makarova et al., 2006). This protospacer sequence must be immediately followed by a ‘NGG’ (or ‘NAG’ with less efficiency) tri-nucleotide protospacer adjacent motif (PAM) on the opposite strand (Mojica et al., 2009); the presence of PAM is crucial for Cas9 target recognition.
To reduce the complexity of the endogenous CRISPR/Cas9 system, a simplified two-component system using a single hybrid hairpin gRNA with a generic tetraloop secondary structure to load Cas9 for sequence-specific DNA cleavage has been developed (Cong et al., 2013; Jinek et al., 2012, 2013; Mali et al., 2013c). Distinct from ZFN and TALEN systems, the separation of the target recognition and nuclease functions on orthogonal components offers extensive flexibility and simplicity for targeted genome manipulations. As Cas9 endonuclease is a universal component for any gRNA-mediated DNA cleavage, transgenic strains expressing Cas9 nuclei gene endogenously have been made in *Drosophila* (Kondo and Ueda, 2013; Ren et al., 2013; Sebo et al., 2013). Thus, in this simplest CRISPR/Cas9 application method, users need only to design, build and provide one gRNA per targeted lesion in order to introduce genetic changes in those animals.

The CRISPR/Cas9 custom endonucleases are the most accessible custom nuclease system for end users to conduct many site-specific genome manipulations. The gRNA can be in *vivo* transcribed using a designed DNA template (made from synthesized DNA oligonucleotides) or can be supplied in the form of a DNA expression vector, transcribed under the control of a RNA polymerase III promoter (Mali et al., 2013b). Multiplex targeting is also routinely achievable: by providing multiple gRNAs together, lesions can be induced simultaneously at multiple loci (Cong et al., 2013; Wang et al., 2013). Importantly, standard Cas9-mediated custom nuclease applications use a single target sequence for specificity, an approach that is comparable with the specificity found within a single TAL arm and that leads to increased rates of off-target cutting (Mali et al., 2013a). Reducing and/or minimizing the off-target effects of the lower specificity Cas9 system is an active area of genome engineering research (Fig. 1D,E; Ran et al., 2013a; Fu et al., 2014; Guilinger et al., 2014; Tsai et al., 2014).

In contrast to the ZFN and TALEN systems that employ the dimeric type IIS FokI endonuclease and thus function as pairs, the unmodified Cas9 endonuclease monomer has full double-strand DNA cleavage activity once properly guided (Jinek et al., 2012), so only a single gRNA is required. The Cas9 protein contains two independent endonuclease domains homologous to either HNH or RuvC endonucleases (Jinek et al., 2012). Each of these domains cleaves one strand of dsDNA at the target recognition site: the HNH domain cleaves the complementary DNA strand (the strand forming the duplex with gRNA), whereas the RuvC-like domain cleaves the non-complementary DNA strand (Jinek et al., 2012). Recent structural analyses (Nishimasu et al., 2014; Jinek et al., 2014) of CRISPR/Cas9 complexes have revealed a two-lobe structure for Cas9—a recognition (REC) lobe and a nuclease (NUC) lobe. Cas9 interacts with the RNA-DNA duplex via the REC lobe in a largely sequence-independent manner, implying that the Cas9 protein itself does not confer significant target sequence preference. However, one caveat with the CRISPR/Cas9 system is that gRNA-loaded Cas9 endonuclease cleavage is not completely dependent on linear guide sequence, with some off-target sequences being shown to be cut with similar or even higher efficiency than the designed target site (Cong et al., 2013; Esvelt et al., 2013; Pattanayak et al., 2013; Ran et al., 2013b). In general, mismatch(es) between the first 12 nucleotides (nt) of the gRNA and the DNA target are not well tolerated, suggesting high sequence specificity in the PAM-proximal region. However, mismatches beyond the first 12 nt can be compatible with efficient cleavage (Cong et al., 2013). A recent biophysical study (Sternberg et al., 2014) on the thermodynamic properties of Cas9 binding has provided a likely explanation for the features of specificity outlined above. Although this general rule regarding specificity holds true, it is an over-simplification, and the sequence recognition specificity of the CRISPR system is a topic of active investigation (Cho et al., 2014; Esvelt et al., 2013; Pattanayak et al., 2013; Ran et al., 2013b; Ren et al., 2013). Notably, shorter gRNAs with up to 5000-fold reduction in off-target effects have been recently described (Fu et al., 2014). Adding two additional G nucleotides on the 5′ end of gRNA in some instances improves the specificity of the CRISPR/Cas9 system (Cho et al., 2014), possibly by altering gRNA stability, concentration or secondary structure. The relaxation of sequence specificity of the RNA-guided endonuclease system remains the biggest challenge so far for its use in genome engineering (Jinek et al., 2014; Sternberg et al., 2014).

To reduce this problem of potential DNA lesions at off-target sites, two promising practical solutions have been developed to impose a requirement for two gRNA target recognition sites. The first uses a Cas9 mutant with only a single-strand endonuclease activity such as the D10A mutant that cleaves only the strand complementary with the gRNA (Cho et al., 2014; Esvelt et al., 2013; Ran et al., 2013a). Such a mutant Cas9 endonuclease can generate only a single-stranded DNA lesion and is thus named a Cas9 ‘nickase’ (Jinek et al., 2012). When two single-stranded lesions are introduced simultaneously on opposite strands, via a Cas9 nickase guided by independent gRNAs recognizing adjacent target sequences, a combined double-stranded break will be made at the targeted site (Fig. 1D). The most recent addition is the use of two nuclease-deficient Cas9 proteins fused to the FokI nuclease domain, deployed in a way that is almost analogous to ZFNs and TALENs (Fig. 1E; Guilinger et al., 2014; Tsai et al., 2014). In both of these systems, the specificity is inherently higher than with one gRNA (Cho et al., 2014; Esvelt et al., 2013; Ran et al., 2013a). The increase of targeting fidelity with both approaches comes with some sacrifice in efficiency: both Cas9 nickase and FokI-dCAS9 demonstrate lower endonuclease activity than wild-type Cas9. In addition, a key limitation of this approach is the requirement for two closely spaced PAM sequences in the same target region of the genome (Fig. 1D,E).

Beyond the type II CRISPR/Cas9 system from *Streptococcus pyogenes* that is currently in popular use today, other similar type II CRISPR/Cas9 systems are also being developed (Chylinski et al., 2013; Esvelt et al., 2013; Fonfara et al., 2013). The major differences between these diverse CRISPR systems lie in the PAM sequence recognized by the endonuclease. This provides additional flexibility in identifying suitable sites for targeted modifications, with the goal of relieving the requirement for an NGG sequence.

**Practical considerations using designer endonucleases: specificity, efficiency and target site flexibility**

Achieving high targeting specificity is important for targetable nuclease applications. One reason for the relatively low frequency of off-target cleavage by ZFNs and TALENs is the dimeric requirement for FokI DNA cleavage. The extent of potential off-target lesions using dimeric ZFN and TALEN systems has been investigated and supports the broad conclusion of limited background effects (Gabriel et al., 2011; Gupta et al., 2011; Hockemeyer et al., 2011; Mussolino et al., 2011; Osborn et al., 2013; Pattanayk et al., 2013; Tesson et al., 2011). In the standard CRISPR-Cas9 system, the reduced sequence recognition stringency that is observed might come from the evolutionary advantage of an adaptive acquired immune system: a less stringent Cas9 endonuclease that tolerates some mismatches between the crRNA and an evolving invading genome may have been selectively
preserved through evolution (Carroll, 2013). The Cas9/FokI systems (Fig. 1E) are currently the best alternative to the standard Cas9 nuclease when stringent target selection is desired (Cho et al., 2014; Esvelt et al., 2013; Ran et al., 2013a). Regardless of the choice of designer endonuclease, under situations where the delivery dose reaches saturation, a clear nuclease-associated cellular toxicity can be observed; keeping the nuclease activity at the lowest practical level would be beneficial to restrict DNA lesions on the targeted locus and reduce off-target effects (Cho et al., 2014; Ran et al., 2013b).

For model organisms, the effects of off-target lesions can be reduced through breeding to untargeted animals: unlinked mutations tend to be diluted quickly within the genetic pool through generational passing. For example, the zebrafish genome is encoded by 25 pairs of chromosomes; chromosomal segregation through out-crossing eliminates unintended changes on 49/50 linkage groups. In model organisms with much fewer chromosomes (such as Drosophila with four chromosome pairs), meiotic recombination in the germline also helps to dilute off-target lesions, thus only loci near the target sequence are likely to be refractory to genetic dilution. Modern designer nucleases have proven particularly useful in model systems in which homozygosity cannot be rapidly achieved through breeding. In these systems, efficient biallelic chromosome conversion is desirable, and both CRISPRs and TALENs have been shown to accelerate the production of homozygosity in human pluripotent stem cells (González et al., 2014), as well as in other cell culture systems, with biallelic conversion rates up to 39% (Tan et al., 2013). Given the potential for off-target lesions, it is in general good practice to include proper controls to ensure the generated molecular allele is responsible for the documented phenotype. Importantly, and similar to traditional means of mutagenesis, genetic mapping of a mutant phenotype to a specific lesion locus does not prove causality. Obtaining additional independent data, such as another mutant allele mapped to the lesion locus does not prove causality. Obtaining additional independent data, such as another mutant allele mapped to the same gene, knockdown phenocopy (morpholinos and/or RNAi) and ultimately phenotypic rescue using the wild-type gene or gene product will be necessary to consolidate the causal relationship. The ability to make specific DNA lesions in the genome at ease with these tools should not overshadow their intrinsic limitations, as it is virtually impossible to exclude the possibility of additional lesions introduced and retained in the genome.

The proliferation of these new designer nucleases has occurred over a very short period of time, and they have yet to be developed into truly foolproof tools that offer consistently high targeting efficiency with tight specificity in every system. Thus, it is still possible that a designer nuclease made by following the current ‘best’ guidelines may work poorly (Cong et al., 2013; Mali et al., 2013b; Ran et al., 2013b). It is particularly puzzling for some CRISPR/Cas9 experiments that targeting efficiency achieved by different gRNAs may vary by a few orders of magnitude, even if the Cas9 endonuclease is supplied endogenously at a constant level (Kondo and Ueda, 2013; Sebo et al., 2013). Systematic efforts are needed to pinpoint the major factors that contribute to the variations in targeting efficiencies and to provide further guidelines for improvement. One such factor that is likely to be especially confounding for in vitro and somatic cell work is the epigenetic state, particularly the DNA methylation status, of the target genomic sequence. Some TALENs work less efficiently on methylated DNA (Dupuy et al., 2013), while limited reports suggest the Cas9 system is less sensitive to DNA methylation status (Ran et al., 2013b). Such differences in DNA methylation sensitivity are likely due to their different modes of target recognition: the 5-methyl group on C does not affect proper base pairing with gRNA, but might abolish the protein-DNA interaction in the major groove for ZFNs and TALENs. It is, however, noteworthy that CpG methylation seems to be negatively correlated with Cas9 binding to DNA in vivo (Wu et al., 2014). Such seemingly contradictory observations could be resolved by the strong indication that chromatin accessibility is a major determinant of Cas9 binding to DNA in vivo (Kucsko et al., 2014; Wu et al., 2014), but that, once bound, the endonuclease activity of Cas9 is unaffected by DNA methylation status.

Among the more popular tools, target site flexibility is a major practical difference between TALENs and CRISPRs. TALENs provide high precision and the greatest options for target site selection with current tools, with no absolute sequence requirements, though the presence of a 5' T base in each TALEN arm often provides the highest activity (reviewed by Campbell et al., 2013; Lamb et al., 2013; Tsuji et al., 2013). Standard CRISPR systems require only a single PAM sequence in the targeted genomic sequence (Fig. 1C), but some CRISPR systems also have sequence requirements due to the RNA expression system used to make the gRNAs (Hwang et al., 2013); sequence constraints are therefore greater than for TALENs. The use of dual gRNA-based CRISPR/Cas9 systems (Fig. 1D,E) will increase the targeting specificity while reducing options for target site selection.

**Applications: designer genome modifications**

The designer nucleases discussed above achieve only the first step of genome engineering applications: the introduction of a double-strand break at a specified genomic site. These lesions are corrected by cellular DNA repair machineries, during which genomic DNA might be modified at the targeted site. In general, DSBs are commonly fixed either through error-prone non-homologous end-joining (NHEJ) (joining two distal DNA breakpoints; see Box 1 and Fig. 2A) or homology-directed repair (HDR) (which requires the presence of a donor DNA with homology to the sequences distal to DNA break point; Fig. 2B,C). NHEJ involves only the two DNA sequences flanking the DSB and is a very efficient repair outcome in many cell types. For this reason, NHEJ-based local sequence modification has been the most widely used genome engineering application that has been successfully demonstrated in almost all organisms tested so far (Fig. 3; and see below). HDR is necessarily a more-complex reaction as it involves the DNA sequences adjacent to the DSB in addition to the added complexity of a third DNA molecule that serves as the donor template (see Box 1). Consequently, HDR is much less efficient than NHEJ in most systems.

The outcome of these repair processes can be unpredictable, but they can also lead to a range of useful mutagenic lesions for genome engineering applications, including deletion or inversion of endogenous sequences, or insertion of exogenous sequences (discussed further below). As we have little manipulative capability to guide the cellular DNA repair process to achieve a given outcome, current best practice is to screen through the pool of targeted cells or organisms for a specific modification at the targeted locus. Indeed, the diversity of outcomes can be exploited in model systems, as a range of results can be clonally amplified by subsequent germline propagation. Given the high degree of conservation of DSB repair pathways in eukaryotes, the diverse genomic engineering applications are translatable between different model organisms, although it should be noted that different DNA repair pathways can be differentially favored between different cell types, even within the same organism. In the following sections, we present an overview of the main types of genome editing that are
commonly being achieved via different mechanisms of DSB repair induced by custom endonucleases (see also Table 1 for a summary of key examples). Although we provide examples of a few model organisms in which a particular application has been demonstrated so far, we encourage readers to go beyond the established examples and explore the most effective approach with their favorite system.

Mutagenesis by NHEJ and detection methodology

The most straightforward application of designer nucleases is to induce a small mutation at a particular site within the genome (Fig. 2A). If such a site is within a protein open reading frame, DSB repair by the error-prone NHEJ pathway can induce a deleterious missense or nonsense mutation in the protein of interest. Missense mutants generated in this manner can be used as a quick means to study structure-function relationships of a few key amino acid residues in vivo. Nonsense or frame-shift mutations will likely make truncated proteins or no protein at all, which can be valuable tools for genetic analysis. The efficiency of DSB generation by the latest generation of engineered nucleases can be sufficiently high that both alleles of the same site are modified within an individual organism in vivo (see, for example, Bedell et al., 2012; Beumer et al., 2013a; Tan et al., 2013).

Molecular assays are typically used to detect insertion or deletion (indel) mutations generated by designer nucleases. The local DNA sequence change may be reflected by RFLP (restriction fragment length polymorphism) if a restriction enzyme site is included around the target sequence. A more general approach takes advantage of local DNA sequence changes that lead to alterations in melting temperature of a PCR-based DNA fragment. Thus, HRMA (high resolution melting analysis) (Wittwer, 2009) or WAVE (based on high resolution temperature-modulated high-pressure liquid chromatography) (Yu et al., 2006) can be used to detect mutations after PCR amplification of the targeted sequence. Alternatively, a mixture of near-identical DNA with local sequence variations (created via indel generation) will make annealed duplexes during PCR with mismatched single-stranded bases at the mutated site. Such single-stranded ‘bubbles’ in the PCR product can be recognized and cut by Cel1 or T7E1 nucleases, generating two shorter DNA fragments. A sensitive nuclease assay is thus used widely for mutation detection (Qiu et al., 2004). The newest addition to the toolkit is the recent development of single-molecule real-time (SMRT) DNA-sequencing technique, which can measure in parallel the frequency of different editing events at any given targeted locus within a large DNA population (Hendel et al., 2014).
Improving homologous recombination efficiency by DSBs introduced on the recipient DNA

Homologous recombination (HR) was the preferred methodology for generating site-specific modifications before the widespread use of designer nucleases. A problem with traditional HR has been low efficiency, thus limiting the model systems where it is feasible. For example, classical HR-mediated gene targeting in *Drosophila* is so inefficient that the donor DNA cannot be directly provided as plasmids for embryonic injection due to the limitation in numbers of injections one can practically handle (Rong and Golic, 2000). However, introducing a targeted DSB results in a much-improved HR efficiency (Fig. 2B). Pioneering work with ZFNs demonstrated that the HR rate following a DSB improved up to 100-fold (Beumer et al., 2008; Bibikova et al., 2001; Porteus and Baltimore, 2003). Recent trials with TALENs and CRISPR systems have also shown a significant improvement of HR efficiency (Baena-Lopez et al., 2013; Beumer et al., 2013a). Thus, in flies, introducing DSBs onto the target site by various nuclease systems has enabled the direct injection of DNA donors into fly embryos as a viable and faster alternative to the traditional transgenic method used to generate modified flies by homologous recombination (Baena-Lopez et al., 2013; Gratz et al., 2014). The functional advantage of designer nucleases to enhance the rate of HR-mediated gene replacement has been successfully demonstrated in all major model organisms (including yeast, worm, fly, fish and mammals; see Table 1).

Making small indels or modifications by homology-directed repair using single-stranded oligonucleotides

Alternatively, homology-directed repair can be used to introduce local modifications at DNA lesion sites (Fig. 2C). A short homology arm of only 10-30 nt is usually sufficient to induce HDR-mediated DNA repair of lesions introduced by targeted endonucleases *in vivo* (Bedell et al., 2012; see Campbell et al., 2013 for a review). Such minimal homology requirement enables synthesized single-stranded DNA (ssDNA) oligonucleotides to serve as a suitable template to repair DSBs via a still poorly defined HDR pathway.
This approach is particularly applicable for engineering small changes, including as little as a single nucleotide. Interestingly, one molecular signature that distinguishes this HDR pathway from traditional homologous recombination is the asymmetry of many modified loci around the DSB. Using ssDNA as donor and reference, the 3′ side tends to be repaired accurately, whereas small indels that are more characteristic of NHEJ-based repair are often found at the 5′ side; this has been seen in mammalian tissue culture cells, and zebrafish and mouse embryos (Orlando et al., 2010; Bedell et al., 2012; Wefers et al., 2013). This molecular asymmetry around an otherwise symmetrical DSB suggests that mechanisms involving DNA synthesis may play an important role in this process.

Besides the homology arms on its 5′ and 3′ ends matching the sequences of boundary sequences at the DNA DSB, the donor ssDNA oligonucleotide can incorporate virtually any sequence in the middle as long as the overall length of oligonucleotide is within the reasonable range of synthesis (currently, up to ~200 bases for routine DNA oligonucleotide synthesis). Thus, small DNA elements can be introduced at a particular position in the selected locus. This could result in, for example, the insertion of a short defined stretch of amino acids into a particular site of a protein, an epitope tag fusing in frame with the targeted protein for endogenous tagging, or a recombination site enabling further site-specific engineering. Owing to the ease of making donor DNA through chemical synthesis of short oligonucleotides, such methods for creating local sequence modifications have been successfully demonstrated in many organisms (including yeast, worm, fly and mammals; see Table 1).

In human cell lines, precise deletions have been achieved using an appropriately designed repair ssDNA (Chen et al., 2011). In some mammalian cells, the efficiency of resection is low, thus limiting the application of this strategy to generating small deletions close to the lesion site (Yang et al., 2013). Alternatively, larger deletions can be generated by inducing two distinct DSBs, which can be repaired using an ssDNA donor harboring homology ends to DNA sequences flanking the distal sides of each cut (Cong et al., 2013; Ran et al., 2013a). Overall, this strategy can be used to generate deletions, to manipulate the sequence of a particular stretch of the targeted protein or to introduce a DNA element at a particular site for further site-specific manipulations. The successful use of ssDNA donor-mediated HDR repair also applies in cases using double Cas9 nickases, specifically when a 30-70 nt 5′ overhang is generated (Ran et al., 2013a).

The ability to make site-specific DNA lesions creates new avenues for site-specific transgenesis. Traditionally, site-specific transgenesis either relies on inefficient HR-mediated knock-in or previously deposited recombination sites (such as atP or loxP). A highly efficient site-specific transgenesis method without such limitations would be greatly beneficial to developmental biologists.

Two recent reports on high-efficiency site-specific gene insertions indicate that such an approach can be routinely adopted with the assistance of designer endonucleases (Maresca et al., 2013; Auer et al., 2014). Both reports exploit the property of NHEJ to fuse DSBs regardless of their sequence (Fig. 2D): chromosomal lesions were previously noted to frequently ‘absorb’ external DNA fragments (Gabriel et al., 2011; Lin and Waldman, 2001a,b; Miller et al., 2004). A strategy named ‘ObLiGaRe’ (obligate ligation-gated recombination) has been developed in mammalian cells (Maresca et al., 2013), whereby a ZFN or TALEN pair would cut simultaneously at one site in the genomic DNA and a similar site in the donor plasmid. The NHEJ-mediated repair could then ‘ligate’ the plasmid into the genomic site through the DSBs. Upon insertion of the ligated product, the transgene insertion will abolish the TALEN recognition site, thus enabling stable site-specific transgenesis. Plasmids up to 15 kb have been integrated through this method, whereas no off-targeted insertion was observed.

Similarly, an in vivo study using zebrafish confirmed the specificity and effectiveness of designer nuclease mediated site-specific transgenesis (Auer et al., 2014). When a ‘bait’ sequence was incorporated in a donor plasmid, simultaneous cutting of the genomic DNA and the plasmid using CRISPR/Cas9 system enabled high efficiency site-specific transgenesis. Importantly, multiple gRNAs can be used simultaneously, so the targets on the genomic DNA and donor plasmid can be distinct (Jinek et al., 2012). This finding also supports the notion that site-specific insertion of foreign DNA is indeed through homology-independent end joining, which can happen in either orientation with error-prone junctions. In theory, any linearized dsDNA could be incorporated into the lesion site, although initial tests using linearized plasmids were inefficient, perhaps due to toxicity in fish embryos (Auer et al., 2014). To date, these novel approaches have come from vertebrate models (mouse and human cells, as well as zebrafish); however, the prevailing mode of NHEJ DSB repair suggests similar or related approaches will also prove effective for invertebrate systems.

### Table 1. Described modes of DNA editing based on repair pathway and organism

<table>
<thead>
<tr>
<th>Application</th>
<th>Mode of DNA repair pathway</th>
<th>Model organisms</th>
<th>Key reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local mutagenesis by NHEJ</td>
<td>NHEJ</td>
<td>Many</td>
<td>See Fig. 3. Beumer et al., 2006; Bibikova et al., 2001; Bobis-Wozowicz et al., 2011; Carroll, 1996; Carroll et al., 2008; Cui et al., 2011; Vasquez et al., 2001; Zu et al., 2013</td>
</tr>
<tr>
<td>Gene replacement by homologous recombination</td>
<td>HR</td>
<td>Human, mouse, rat, fish, fly, worm, frog and yeast</td>
<td>Bedell et al., 2012; Beumer et al., 2013; DiCarlo et al., 2013; Radecke et al., 2006; Wefers et al., 2013; Zhao et al., 2014</td>
</tr>
<tr>
<td>Local modifications using single-stranded oligonucleotides</td>
<td>Non-HR modes of HDR</td>
<td>Human, mouse, fish, fly, worm and yeast</td>
<td>Auer et al., 2014; Maresca et al., 2013</td>
</tr>
<tr>
<td>Targeted NHEJ-mediated knock-ins</td>
<td>NHEJ</td>
<td>Human and fish</td>
<td>Bibikova et al., 2002; Brunet et al., 2009; Cradick et al., 2013; Gupta et al., 2013; Liu et al., 2013; Piganeau et al., 2013; Solli et al., 2010; Xiao et al., 2013</td>
</tr>
<tr>
<td>Generating chromosomal abnormalities by joining two distal DSBs</td>
<td>NHEJ</td>
<td>Human and fish</td>
<td>Auer et al., 2014; Maresca et al., 2013</td>
</tr>
<tr>
<td>DNA replacement with double-strand oligo after ssDNA nicking</td>
<td>Uncertain – transcription-dependent non-canonic HDR pathway?</td>
<td>Human</td>
<td>Ran et al., 2013b</td>
</tr>
</tbody>
</table>

DSB, double-stranded break; HR, homologous recombination; HDR, homology-directed repair; NHEJ, non-homologous end-joining; ssDNA, single-stranded DNA.
Generating deficiencies by joining two distal DSBs
Chromosomal deficiencies, duplication and inversions are traditionally induced with random mutagenesis tools such as DNA-damaging gamma rays. Chromosomal abnormalities with precise boundaries can be important tools to map developmental regulatory landscapes, and can be achieved with high efficiency and versatility using designer nucleases (Liu et al., 2013; Piganeau et al., 2013; Xiao et al., 2013). Successful examples of this application have been at least demonstrated for mammals, fish and flies. By design, genomic segments from tens of kilobases up to 15 megabases can be deleted by the combination of two pairs of custom nucleases, including TALENs, Cas9 nucleases or the derived nickases (Cong et al., 2013; Fujii et al., 2013; Ran et al., 2013a). Although chromosomal translocations have been observed as an unfavorable by-product of CRISPR-induced lesions due to off-target lesions on different chromosomes (Cradick et al., 2013), these observations also suggest two distal DNA fragments without homology can be efficiently joined via the NHEJ pathway, and was recently used to model genomic rearrangements observed in human cancer (Choi and Meyerson, 2014). Although the deletions are generally very reproducible in nature, these chromosomal abnormalities are ultimately repaired by the error-prone NHEJ repair pathway. As a result, the local sequence around the junctions of the two distal DNA segments is highly variable at the individual nucleotide level between independent deletion mutants. Such variations need to be taken into consideration for downstream biological studies using such deficiencies.

DNA replacement with double-strand (ds) oligo after ssDNA nicking
The ability of a DNA nickase (such as Cas9 nickase or equivalent ZFN or TALEN nickases) (Kim et al., 2012; Wu et al., 2014) to generate a lesion on a particular strand of DNA at the target site enables the generation of 5′ or 3′ single-stranded DNA overhangs when two nickases make lesions at adjacent sites (Fig. 2F). These sticky ends with defined sequences can be used to incorporate foreign dsDNA fragments with compatible sticky ends. This is conceptually very similar to sticky end-mediated molecular cloning after restriction enzyme digestion in recombinant DNA technology. Even though this application has not been extensively tested beyond one pilot study in human cells, where a 148 bp dsDNA oligo was successfully inserted via its 5′ overhangs (Ran et al., 2013a), such an approach could in theory circumvent some practical restrictions of HDR-mediated insertion of ssDNA (size limitation of insertion and the dependence of homology overhang of significant size).

Conclusions
The availability of a wide range of designer nuclease tools with successful applications in model organisms is revolutionizing the way basic developmental processes can be studied. For example, making static knockout alleles using NHEJ is now standard practice in most developmental biology systems, something either difficult or impossible to do just a few years ago. Moreover, the arrival of versatile, robust and simple genomic engineering tools means that genetic analysis of developmental processes should no longer be confined to traditional model organisms, but can be extended to less well-established organisms, where genetic tools have been lagging behind. Fig. 3 shows a snapshot of the published systems to date. Examples of organisms in which these tools have been successfully applied range from commercially important livestock (e.g. cows and pigs; Carlson et al., 2012), to species important for disease research (e.g. mosquitio; Aryan et al., 2013) and from systems that have

Table 2. Other, non-nuclease based, locus-specific DNA tools based on designer nuclease backbones

<table>
<thead>
<tr>
<th>Designer regulator</th>
<th>DNA recognition module</th>
<th>Fusion partner</th>
<th>Mode/mechanism of regulation</th>
<th>Key reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zinc-finger transcription factors</td>
<td>Tandem zinc fingers</td>
<td>VP64 (tandem copies of VP16 activation domain from herpes simplex virus)</td>
<td>Transcriptional activation</td>
<td>Beerli et al., 1998</td>
</tr>
<tr>
<td>Zinc-finger repressor</td>
<td>Tandem zinc fingers</td>
<td>KRAB (Krüppel-associated box) repression domain</td>
<td>Forming facultative heterochromatin</td>
<td>Beerli et al., 1998</td>
</tr>
<tr>
<td>TALE activators (TALEAs)</td>
<td>TALE DNA-binding domain</td>
<td>VP64</td>
<td>Transcriptional activation</td>
<td>Crocker and Stern, 2013</td>
</tr>
<tr>
<td>TALE repressors (TALERs)</td>
<td>TALE DNA-binding domain</td>
<td>Hairy repression domain</td>
<td>Histone deacetylation</td>
<td>Cong et al., 2012; Crocker and Stern, 2013</td>
</tr>
<tr>
<td>TALE epigenetic modifier</td>
<td>TALE DNA-binding domain</td>
<td>LSD1 (lysine-specific demethylase)</td>
<td>Histone demethylation</td>
<td>Maeder et al., 2013a; Mendenhall et al., 2013</td>
</tr>
<tr>
<td>TALE DNA imaging (dTALE)</td>
<td>TALE DNA-binding domain</td>
<td>GFP</td>
<td>Fluorescent detection</td>
<td>Thanisch et al., 2014</td>
</tr>
<tr>
<td>CRISPRi</td>
<td>RNA-guided endonuclease-inactivated Cas9</td>
<td>None</td>
<td>Competitive DNA binding to block transcriptional initiation or elongation</td>
<td>Qi et al., 2013</td>
</tr>
<tr>
<td>Cas9 transcriptional activator</td>
<td>RNA-guided endonuclease-inactivated Cas9</td>
<td>VP64 or VP48 (tandem copies of VP16 activation domain, from herpes simplex virus) and P65 activation domain (from human NF-κB RelA transcription factor)</td>
<td>Transcriptional activation</td>
<td>Cheng et al., 2013; Farzadfard et al., 2013; Gilbert et al., 2013; Hu et al., 2014; Maeder et al., 2013b; Mali et al., 2013a</td>
</tr>
<tr>
<td>Cas9 transcriptional repressor</td>
<td>RNA-guided endonuclease-inactivated Cas9</td>
<td>KRAB repression domain</td>
<td>Forming facultative heterochromatin</td>
<td>Farzadfard et al., 2013; Gilbert et al., 2013</td>
</tr>
</tbody>
</table>

CRISPRi, CRISPR interference; TALE, transcription activator-like effector.
previously been difficult to target genetically (e.g. *Xenopus*; Blitz et al., 2013; Guo et al., 2014), to newer models for evo-devo studies (e.g. *Ciona*; Treen et al., 2014). These tools are broadly applicable for NHEJ-induced mutations and more sophisticated approaches are likely to succeed in many cases. One caveat is that the design of the targeted nucleases relies on high-quality genomic sequence, so ongoing genomic sequencing and assembly efforts for such models will need to continue. This will not only provide a clear genetic blueprint of their development and physiology, but will offer immediate access to many of these new genetic manipulation tools. Another bottleneck in some organisms will be the development of methods for introducing nucleic acids encoding designer nucleases and their repair substrates into developing embryos and germline cells. With the advent of these new technologies in diverse species, comparative and functional studies between developmentally divergent and evolutionarily related species need no longer be simply tested in the species with the accessible genetic tools, but will likely be cross-examined in any pair-wise or even multiplexed phylogenetic combinations. Bringing models that were previously inaccessible to genetic manipulation into functional developmental studies will thus equip every developmental biologist with a novel evolutionary prospective.

Beyond the ready generation of static alleles, these tools offer new options for precise and sophisticated molecular genetic manipulations. Engineering the genomic locus of interest at its endogenous site has many advantages. For example, with single base resolution of genomic editing, it is feasible to mutate specifically a crucial amino acid or to influence the choice of multiple alternative splicing events in order to study the function of a particular protein isoform, all conducted within the context of a complete genomic complex with regulatory and other control elements intact. Introducing an epitope tag at a chosen position in a protein should also be straightforward, thus circumventing the need to generate a protein-specific antibody for immunolocalization or immunoprecipitation. It would even be conceivable to epitope tag every transcription factor within the genomes of key model organisms for systematic chromatin immunoprecipitation approaches. In addition to manipulations of protein-coding regions in the genome, developmental biologists are now equipped with a feasible means with which to probe the vast regulatory genomic landscape (which remains largely unexplored by traditional random mutagenesis approach) and to study the gene regulation network within an intact genomic context. Many developmentally regulated genes possess multiple regulatory elements, which work together to produce complex spatial and temporal expression patterns during development. Approaches using genomic engineering to examine their roles within their normal genomic context will be crucial to understand the interplay between distinct regulatory elements in controlling gene expression.

Furthermore, the ease of targeted manipulation of any genomic region enables the next era beyond genetics in developmental studies, by offering avenues to monitor and manipulate the transcription or epigenetic status of individual loci (Maeder et al., 2013a; Mendenhall et al., 2013) with single cell precision in relation to pattern formation and cell fate determination. Such designer transcriptional and epigenetic regulators (Table 2) build on similar platforms as various designer nucleases that we discussed previously. A significant current drawback shared among such approaches is that multiple targeting modules are usually required to act synergistically on a single locus to bring about any biological meaningful perturbation. Yet further refinement of this class of targeting regulators will definitely hold even greater promise to the new generation of developmental biologists.

Acknowledgements
We apologize for the editorial constraints that limit the citation of all the excellent papers that pioneered the way for applications of genomic engineering tools in various organisms. We thank Jose Agosto Rivera and anonymous reviewers for critical comments, and Mark Curry for excellent artistic renditions.

Competing interests
The authors declare no competing financial interests.

Funding
This Review was supported by the state of Minnesota [partnership grant H001275806], by National Institutes of Health grants [GM63904, P30DK84567 and P30DK90728] and by the Mayo Foundation to S.C.E.; by a Mayo Clinic Center for Cell Signaling in Gastroenterology Pilot and Feasibility Award [NIHDK P30DK084567] to K.J.C.; and by a Mayo Clinic New Investigator Startup Fund, a Richard F. Emslander Career Development Award and a Mayo Clinic Center for Cell Signaling in Gastroenterology Pilot and Feasibility Award [NIHDK P30DK084567] to Y.G. J.M.C. was supported in part by the Mayo Clinic Center for Clinical and Translational Sciences [UL1TR000135 from the National Center for Advancing Translational Sciences]. Deposited in PMC for release after 12 months.

References


