4166–4179 *Nucleic Acids Research, 2011, Vol. 39, No. 10* *Published online 25 January 2011*
doi:10.1093/nar/gkq1363

Use of RecA fusion proteins to induce genomic modifications in zebrafish

Hsin-Kai Liao[1,2] and Jeffrey J. Essner[1,*]

[1]Department of Genetics, Development and Cell Biology and [2]Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA

Received September 2, 2010; Revised December 30, 2010; Accepted December 31, 2010

ABSTRACT

The bacterial recombinase RecA forms a nucleic acid-protein filament on single-stranded (ss) DNA during the repair of double-strand breaks (DSBs) that efficiently undergoes a homology search and engages in pairing with the complementary DNA sequence. We utilized the pairing activity of RecA–DNA filaments to tether biochemical activities to specific chromosomal sites. Different filaments with chimeric RecA proteins were tested for the ability to induce loss of heterozygosity at the *golden* locus in zebrafish after injection at the one-cell stage. A fusion protein between RecA containing a nuclear localization signal (NLS) and the DNA-binding domain of Gal4 (NLS-RecA-Gal4) displayed the most activity. Our results demonstrate that complementary ssDNA filaments as short as 60 nucleotides coated with NLS-RecA-Gal4 protein are able to cause loss of heterozygosity in ~3% of the injected embryos. We demonstrate that lesions in ~9% of the F0 zebrafish are transmitted to subsequent generations as large chromosomal deletions. Co-injection of linear DNA with the NLS-RecA-Gal4 DNA filaments promotes the insertion of the DNA into targeted genomic locations. Our data support a model whereby NLS-RecA-Gal4 DNA filaments bind to complementary target sites on chromatin and stall DNA replication forks, resulting in a DNA DSB.

INTRODUCTION

The diverse activities of bacterial RecA make it an attractive protein for use in genome manipulation and gene targeting. RecA is widely known for its recombinase activity that catalyzes strand exchange during the repair of double-strand breaks (DSBs) by homologous recombination (1–3). After a DSB occurs from ionizing radiation or some other insult, exonucleases chew back the DNA ends 5′–3′, thereby exposing one strand of the DNA (1,4). The single-stranded (ss) DNA becomes stabilized by single-strand binding protein (SSB). After binding of SSB, RecA binds the ssDNA and forms a helical nucleoprotein filament (referred to as a filament or a presynaptic filament). During DNA repair, the homology-searching functions of RecA direct the filament to homologous DNA and catalyze homologous base pairing and strand exchange. This results in the formation of a DNA heteroduplex. Following strand invasion, DNA polymerase elongates the ssDNA, using the homologous DNA as template to repair the DNA break, and crossover structures or Holliday junctions are formed. Interestingly, RecA also shows a motor function that participates in the migration of the crossover structures (5).

RecA and RecA-like proteins (called Rad51 in non-bacterial species) have been examined for activity that stimulates gene targeting and homologous recombination in a variety of eukaryotic systems. In tobacco cells, expression of bacterial RecA containing a nuclear localization signal (NLS) increases the repair of mitomycin C-induced DNA damage by homologous recombination and somatic intrachromosomal recombination from 3- to 10-fold (6). Expression of NLS-RecA in tobacco can also stimulate sister chromatid exchange 2.4-fold over wild-type levels (7). In somatic mammalian cells, overexpression of NLS-RecA stimulates gene targeting by homologous recombination 10-fold (8). However, in human cells, overexpression of a human homologue of RecA, hRAD51, stimulates recombination only 2- to 3-fold over wild-type levels under antibiotic selection (9). In zebrafish, a mutant form of the enhanced green fluorescent protein (EGFP) was corrected at low frequency after injection of ssDNA-RecA filaments directly into embryos (10). Rad52, a member of the Rad51 epistasis group, also promotes single-strand annealing and low level gene disruption in zebrafish using mutated oligonucleotides (11). Taken together, these studies indicate that ectopic expression of RecA or Rad51 alone results in a modest stimulation of homologous recombination, but the frequency is not high enough to be useful for gene targeting.

*To whom correspondence should be addressed. Tel: +1 515 294 7133; Fax: +1 515 294 8457; Email: jessner@iastate.edu

Here, we have taken a different approach and used the homology-searching activity of bacterial RecA to target specific chromosomal regions and promote site-specific mutation in zebrafish. Different RecA chimeric proteins were generated and used to coat ssDNA complementary to the *golden* (*gol*) locus. After injection of the chimeric RecA-*gol* DNA filaments at the one-cell stage, loss of heterozygosity (LOH) was monitored at the *gol* locus by examining embryos for loss of pigmentation. A chimeric RecA protein that is fused at the amino-terminus to a NLS and at the carboxyl terminus to the Gal4 DNA-binding/dimerization domain, named NLS-RecA-Gal4, promotes LOH at the *gol* locus. LOH is not observed at the *gol* locus when NLS-RecA-Gal4-DNA filaments are injected with DNA that is complementary to other genes. These results indicate that the homology-searching activity of RecA can be used to tether different biochemical activities to specific chromosomal sites.

MATERIALS AND METHODS

Zebrafish culture and strains

Zebrafish embryos were raised at 28.5°C for staging (12). The wild-type WIK and *gol*^{b1} mutant zebrafish strains were obtained from Zebrafish International Resource Center (http://zfin.org).

Cloning, expression and purification of fusion proteins

Native RecA was purchased from Sigma. Different RecA fusion proteins were cloned for this study (Figure 1A). For the different fusion proteins, the full-length sequence encoding bacterial RecA and the yeast Gal4 DNA-binding domain were amplified with the proofreading polymerase Pfx50 (Invitrogen) from pBEU14 RecA (13) and pGBT9 Gal4 plasmids (Clontech), respectively. For all proteins produced, the N-terminus of RecA was fused with a NLS from SV40 Large T Antigen to promote nuclear targeting (14). The NLS at the 5′-terminus of each fusion protein was added using the following primer: 5′-CATATGCCACCTAAAAAGAAGAGAAA GGTAGAAGACCCCAAGATGGCTATCGACGAAA ACAA-3′. NLS-RecA was amplified using the 5′-primer and the following 3′-primer: 5′-GATCGCGGCCGCAAA ATCTTCGTTAGTTTCTG-3′. The primers used to prepare the 3′-terminal of different forms of RecA fused to Gal4 were 5′-GCGGCCGCCGATACAGTCAACTGT CTTT-3′ for NLSRecAGal4 and 5′-CAGCGGAGACCT TTTGGTTTTG-3′ for NLS-RecA-Gal4ΔDD. The NLS-RecA sequence was fused to the Gal4 DNA-binding domain sequence via a unique NheI restriction enzyme site using the following oligonucleotides: 5′-GCTAGCAA AATCTTCGTTAGTTTCTG-3′ and 5′-GCTAGCATGA AGCTACTGTCTTCTATCG-3′. The purified polymerase chain reaction (PCR) products were ligated to NdeI and NotI restriction sites in the pET41a expression vector (Novagen). This vector contains a sequence coding for an 8xHis-affinity tag after the NotI site at the C-terminus.

Recombinant forms of RecA were expressed in BL21(DE3)pLysS_Cam^{+} *Escherichia coli*. After induction with 1–2 mM isopropyl β-D-1-thiogalactopyranoside, the cells were cultured at 37°C for 2 h, harvested by centrifugation and lysed using a French press/cell disrupter (Thermo Electron Corporation). Lysates were centrifuged at 12 000*g* at 4°C for 20 min, and recombinant proteins in the supernatant were purified using Ni^{2+}-chelate affinity chromatography (Qiagen). Proteins were eluted using Imidazole in a step gradient (from 50 to 500 mM of Imidazole). RecA proteins usually eluted near the 250 μM step. Protein-rich fractions were examined by 12% SDS–PAGE and concentrated using a 5 kDa cut-off Centriprep Amicon system (Millipore). The elution buffer was exchanged with phosphate buffered saline buffer. The purified proteins were stored in 50% glycerol at −20°C.

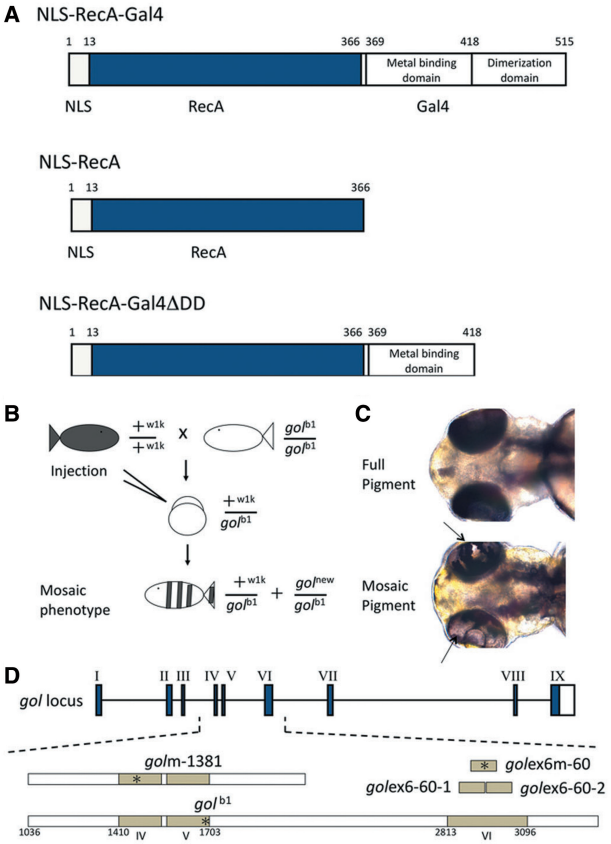

Figure 1. (**A**) RecA fusion proteins used in this study. (**B**) Injection of complementary ssDNA-NLS-RecA-Gal4 filaments leads to LOH at the *gol* locus, resulting in mosaic eye pigmentation. Genotype of the embryos used for injection is shown. (**C**) An example of LOH at the *gol* locus. Dorsal views of eyes at 3 dpf showing wild-type pigmentation patterns in a non-injected embryo (upper panel) and mosaic pigmentation patterns (bottom panel) after injection of complementary ss*gol*-NLS-RecA-Gal4 targeting filaments. (**D**) DNA complementary to either exons4/5 or exon6 of *gol* was used for making *gol* RecA filaments. *Gol*m1381 (m, mutant; 1381, 1381 bp) and *gol*ex6m-60 (ex6, exon 6; m, mutant; 60, 60 bp) DNAs carry mutations resulting in a stop codon (asterisk). *Gol*ex6-60-1(ex6, exon 6; 60, 60 bp) and *gol*ex6-60-2 (ex6, exon 6; 60, 60 bp) oligonucleotides do not have mutations.

Preparation of RecA filaments

The 60-mer oligonucleotides used for targeting were ordered from Integrated DNA Technologies (Supplementary Table S1). To produce 60-bp RecA

filaments that are complementary to the targeted gene, 45–160 ng of ssDNA was used per RecA coating reaction. For production of longer, 200–300 bp and 1381 bp filaments (Supplementary Tables S1 and S2), dsDNA was PCR amplified from genomic DNA extracted from 2 days post fertilization (dpf) zebrafish embryos (Qiagen). To create the mutant *gol* 1381 bp (*gol*m-1381) DNA fragment, the primers 5′-GACGAAGGGCGACAT CTGAGTCAGCACCATCATGG-3′ and 5′-CCATGAT GGTGCTGACTCAGATGTCGCCCTTCGTC-3′ were used to replace the endogenous HgaI site with a stop codon in the middle of the fragment. After column purification (Qiagen, PCR purification kit), 10–20 ng of the PCR product was diluted in water into 4.0 µl and was not exposed to ethidium bromide (EtBr). dsDNA was denatured by heating to 95°C in a temperature cycler (Mastercycler epgradient S, Eppendorf) for 12 min and chilled on ice for 2 min; 0.8 µl of buffer [100 mM TrisOAc, pH 7.5; 500 mM NaOAc; 10 mM DTT, 10 mM $Mg(OAc)_2$], 0.6 µl of 16.2 mM ATPγS (Sigma) and 100–200 ng of RecA, NLS-RecA or NLS-RecA-Gal4 was added to 4.0 µl ssDNA to a final reaction volume of 7.0 µl. Reactions for producing DNA–RecA filaments were then incubated at 37°C for 30 min. RecA filaments were placed on ice until injection into the embryo. The efficiency of coating was determined by mobility shift on a 2% agarose gel. Following electrophoresis, the gel was stained with EtBr to visualize the DNA.

Tracking RecA filament activity with a gene trap

The gene trap construct, pDB759 [provided by Stephen Ekker at the Mayo Clinic, MN, USA (15)], was modified to follow the effects of injection of RecA filaments in some experiments. The gene trap was modified to contain an EGFP reporter gene rather than a RFP reporter gene. The gene trap contains a carp β-actin splice acceptor (SA) in front of EGFP that is missing the first AUG, followed by the zebrafish β-actin 3′-UTR/poly adenylation signal (pA). Three different reading frames following the SA were constructed by adding one or two cytosines (C) at the XmaI site between the slice acceptor and the EGFP coding sequence (Supplementary Table S3). The forward primers used to add one or two Cs are 5′-GCAACGCCC GGG**C**TACCGGAAGGTGTGAGCAAGG-3′ and 5′-G CAACGCCCGGG**CC**TACCGGAAGGTGTGAGCAA GG-3′, and the reverse primer is 5′-GCTAGGATCGATT TCTTACTTGTACAGCTCGTC-3′. The restriction enzymes XmaI and ClaI were used to cut and replace the original GFP DNA fragment. The SA-EGFP-pA plasmids with all three frames in equal molar amounts were linearized with NotI and NdeI and directly purified with a Qiagen DNA purification column without either electrophoresis or exposure to EtBr. The 150 ng (50 ng of each frame) of the SA-EGFP-pA linear fragments was added to the 7 µl coating reaction, and water was added to a 10 µl final volume.

Insertion site analysis

Genomic DNA was extracted from zebrafish embryos at 4 dpf with the Qiagen Genomic DNA isolation kit (Qiagen). The insertion site of the reporter gene was determined using a nested PCR approach. The primers used to amplify from sequences 5′ to *gol*-270- or *gol*-300-targeting sites are Primer I 5′-GTGGGCGGAGTTACAGATCA-3′ and Primer II 5′-TGTGATGTTTTCTGGAACTATTGA ATCTT-3′, respectively. The nested PCR-used primers to amplify out from the EGFP reporter construct were 5′-GGCCAGGGCACGGGCAGC-3′ and Primer III 5′-GAGCATGTGACCATGTGGAGTCAGC-3′. Nested PCR was performed with an annealing temperature of 56°C for 35 cycles using Go Taq DNA polymerase (Promega). Similarly, insertion of the EGFP gene trap into the *prominin1a* (*prom1a*) gene was tested by amplification out from the EGFP reporter construct, using the same primers as above, and a primer 5′ to the targeted region in the *prom1a* gene: 5′-TTGCATATCCTGCGAT GTGA-3′. The PCR products were separated by electrophoresis through a 1.2% agarose gel. Amplified DNA bands were cut from the gel, purified using a Qiagen gel extraction kit and analyzed directly by DNA sequencing (Supplementary Table S4).

Real-time PCR analysis

The frequency of insertion of the EGFP gene trap into the *gol* locus was estimated using real-time PCR. The primer used to amplify from the *gol* gene was 5′ to the *gol*-270-targeting site: gol-F; 5′-AAGGGCGACATCGGCG TCA-3′. The primer used to amplify from the EGFP gene trap was GFP-R, 5′-GTCACTTACTTAACCTAG TCTCAAG-3′. The primers used to amplify a 322-bp DNA fragment from the *floating head* (*flh*) locus, a single copy gene used as a reference, were *flh*-F 5′-GAA GAGCTTACGAAACTTATGGAC-3′ and *flh*-R 5′-TAC ACAGCTCCACGACATGTTGT-3′. The genomic DNA samples were those used in the insertion site analyses. Real-time PCR was performed with an annealing temperature of 58°C for 40–60 cycles using the iQ SYBR Green Supermix reagent with the MyiQ real-time PCR detection system (Bio-Rad). The real-time PCR products were separated by electrophoresis through a 1.5% agarose gel to confirm their sizes.

Mapping

For molecular mapping of the chromosomal break points, simple sequence length polymorphism (SSLP) markers (http://www.ncbi.nlm.nih.gov/unists) on zebrafish linkage group 18 were identified that were polymorphic between wild-type or WIK and gol^{b1} zebrafish strains. A sequence length polymorphism was also observed in intron 8 of the *gol* gene, which we named *gol*in8. This region was amplified with the following primers: 5′-GAGACTCTGGGAATCC CGG-3′ and 5′-ATAAGGGGGCGCTGTAGAGC-3′. To identify new *gol* alleles, *gol*-NLS-RecA-Gal4-injected embryos were raised to adulthood and crossed with gol^{b1} zebrafish to identify F1 offspring that failed to complement the gol^{b1} allele. Offspring that failed to complement were raised to adulthood if possible. Genomic DNA was isolated from either the F1 larva or fin clips from F1 adults that had failed to complement the gol^{b1} allele and their wild-type siblings as a control.

Imaging

Live and fixed embryos were mounted in 1.5% methylcellulose and imaged on a Zeiss Axioskop 2 or Zeiss Discovery V12 microscope. Digital images were taken with a ProgRes C10plus digital camera.

RESULTS

Design of RecA fusion proteins

It has been previously shown that ssDNA RecA-coated filaments bind homologous genomic sequences and promote strand invasion and strand exchange (4). In this study, we have produced chimeric RecA proteins to test whether the homology-searching activity of RecA could be used to tether specific biochemical activities to distinct chromosomal regions in the zebrafish genome. Our goal was to create a modified RecA fusion protein that would stimulate gene disruption at sites homologous to the ssDNA RecA-coated filaments. The chimeric RecA proteins used in this study are shown in Figure 1A. To assay the gene targeting activity of the chimeric RecA proteins, each protein was tested for its ability to induce LOH at the *golden* (*gol*) locus in the zebrafish genome, as shown in Figure 1B.

The NLS-RecA-Gal4 fusion protein was originally designed to tether UAS-containing plasmids to chromosomal sites. In the course of testing the ability of NLS-RecA-Gal4 fusion protein to tether plasmids, we observed that injection of *gol* DNA filaments made with the NLS-RecA-Gal4 protein alone resulted in LOH at the *gol* locus (Figure 1C). Therefore, we decided to investigate this activity further and test if it could promote targeted gene disruption. The NLS-RecA-Gal4 fusion protein was made by addition of the NLS from SV40 Large T Antigen to the N-terminus of RecA and addition of the DNA-binding domain of Gal4 at the C-terminus (Figure 1A). The NLS was added to facilitate nuclear targeting of the ssDNA-RecA filaments after injection into the zebrafish embryo. The Gal4 DNA-binding domain can be divided into two subdomains, a dimerization and a metal-binding/DNA recognition domain. Our initial studies, presented below, indicated that the dimerization domain of Gal4 was necessary for the full gene targeting activity of NLS-RecA-Gal4-coated filaments. For control experiments we made two proteins, NLS-RecA, which is missing the entire Gal4 DNA binding domain, and NLS-RecA-Gal4ΔDD, which lacks the Gal4 dimerization subdomain, leaving the metal-binding domain intact (Figure 1A).

The NLS-RecA-Gal4 protein was tested for its ability to coat ssDNA. An *in vitro* reaction was assembled with purified protein, ssDNA and a non-hydrolysable form of ATP, ATPγS. A 250-nt region from the *flh* locus was PCR amplified and denatured to make complementary single-stranded DNA (cssDNA). The coating reaction was analyzed on a standard agarose gel to examine if there was a mobility shift in the DNA. Commercially available RecA protein efficiently coated cssDNA, resulting in a predicted mobility shift of the ssDNA after electrophoresis (Supplementary Figure S1). In contrast, incubation with an equivalent amount of BSA did not result in a mobility shift of the DNA (data not shown). Similar to RecA, the NLS-RecA-Gal4 protein also caused a mobility shift of the cssDNA, indicating that the NLS-RecA-Gal4 fusion protein retained the ability to bind cssDNA. However, much of the coated cssDNA failed to migrate into the agarose gel, suggesting the NLS-RecA-Gal4 filaments form higher order structures, possibly via the dimerization domain in the Gal4 DNA-binding region.

Injection of cssDNA-NLS-RecA-Gal4 filaments promotes LOH

To test the ability of the cssDNA-RecA filaments to promote targeted gene disruption, we used an *in vivo* assay that allowed us to visualize LOH in zebrafish embryos at a pigment gene, *gol*. We chose to target the *gol* locus for several reasons. First, the *gol* locus is required in a cell-autonomous manner for pigmentation (16), providing a direct relationship between phenotype and genotype and allowing individual cells to be examined for pigment production. Second, homozygous recessive *gol* mutants are viable but lack dark pigmentation in the retinal epithelial cells and the melanocytes of the trunk. *gol* heterozygotes display normal or wild-type levels of pigmentation (Figure 1C). LOH in *gol* heterozygous embryos results in clear patches in the eye that lack pigmentation (Figure 1C) (17), which allows for rapid screening of a visible phenotype in a mutated gene that is not essential for embryogenesis. Third, the gene corresponding to the *gol* mutation, named *slc24a5*, is cloned and the genomic region containing the *gol* locus is well characterized (18).

Previous studies have shown that ssEGFP DNA-RecA filaments can correct a mutant form of EGFP in zebrafish at a low frequency (10). Our initial strategy to target the *gol* locus was to test the ability of RecA proteins to recombine exogenously supplied homologous DNA into the endogenous *gol* locus. A 1381-bp piece of DNA complementary to a region spanning from intron 3 to intron 5 of the *gol* gene, which contained a mutated (m) exon 4, was used (called *gol*m-1381 in Figure 1D). The mutation in exon 4 results in a premature stop codon and was designed to replace an endogenous HgaI restriction enzyme recognition site in exon 4 to create a restriction fragment length polymorphism (RFLP) if it was recombined into the endogenous *gol* locus (Supplementary Table S1). The 1381-bp DNA was denatured and coated with NLS-RecA-Gal4 and ATPγS to produce css*gol*m-1381-NLS-RecA-Gal4 filaments. Following injection, embryos were examined for LOH at the *gol* locus at 3 dpf. At 3 dpf, 2.9% of the injected embryos displayed LOH at the *gol* locus as displayed by loss of pigmentation in patches of the retinal epithelium (Figure 1C and Table 1). As a control, the NLS-RecA-Gal4 protein was mixed with ds *gol*m-1381 DNA that was not denatured and injected into gol^{b1} heterozygous embryos. This condition did not show detectable levels of LOH at the *gol* locus. These results indicated that the css *gol*m-1381-NLS-RecA-Gal4 filaments were able to mutate the wild-type copy of *gol*. However, analysis of DNA isolated from embryos containing mosaic patterns of pigmentation did not show the

Table 1. Injection of complementary ssDNA-NLS-RecA-Gal4 results in LOH at the *gol*^{b1} locus

RecA type	Filaments	Dosage (pg)	Total	Total living	Normal	*gol* eye clones (%)	
NLS-RecA-Gal4	**css *gol*m-1381**	45	292	241	ND	7/241	**2.9**
NLS-RecA-Gal4	ds *gol*m-1381	45	284	237	ND	0/237	0
NLS-RecA-Gal4	ss *gol*ex6m-60 s	80	155	136	ND	0/136	0
NLS-RecA-Gal4	ss *gol*ex6m-60 a	80	258	224	ND	0/224	0
NLS-RecA-Gal4	**css *gol*ex6m-60**	80	278	201	ND	5/201	**2.5**
No	css *gol*ex6-60-1	160	74	65	58	0/58	0
No	css *gol*ex6-60-2	160	227	208	150	0/150	0
No	ss *gol*ex6-60-1 s + *gol*ex6-60-2 a	160	75	68	68	0/68	0
No	ss *gol*ex6-60-2 s + *gol*ex6-60-1 a	160	82	71	70	0/70	0
No	css *gol*ex6-60-1 + *gol*ex6-60-2	160	597	379	342	0/342	0
NLS-RecA-Gal4	**css *gol*ex6-60-1**	160	260	188	144	6/144	**4.2**
NLS-RecA-Gal4	**css *gol*ex6-60-2**	160	139	135	107	3/107	**2.8**
NLS-RecA-Gal4	ss *gol*ex6-60-1 s + *gol*ex6-60-2 a	160	222	197	190	0/190	0
NLS-RecA-Gal4	ss *gol*ex6-60-2 s + *gol*ex6-60-1 a	160	147	129	129	0/129	0
NLS-RecA-Gal4	**css *gol*ex6-60-1 + *gol*ex6-60-2**	160	979	745	591	20/591	**3.4**
RecA	css *gol*ex6-60-1 + *gol*ex6-60-2	160	317	233	206	0/206	0
NLS-RecA	css *gol*ex6-60-1 + *gol*ex6-60-2	160	301	253	220	0/220	0
NLS-RecA-Gal4△DD	**css *gol*ex6-60-1 + *gol*ex6-60-2**	160	396	341	296	3/296	**1.0**
NLS-RecA-Gal4	css *flh*-60	160	298	268	255	0/255	0
NLS-RecA-Gal4	css *prom1a*-60	160	411	314	300	0/300	0
NLS-RecA-Gal4	css *vegfa*-60	160	542	337	307	0/307	0
NLS-RecA-Gal4	**css 5′AmC6-*gol*ex6-60-2**	160	152	109	67	2/67	**3.0**
NLS-RecA-Gal4	**css 3′InvdT-*gol*ex6-60-2**	160	393	318	236	9/236	**3.8**
No-injection			471	438	ND	0/438	0

Normal indicates normal development through 4 dpf. Rows in bold are filaments that resulted in LOH or *gol* eye clones. ss, single-strand DNA; ds, double-strand DNA; css, complementary single-strand DNA probe; *gol*, *golden* gene; *flh*, *floating head* gene; *prom1a*, *prominin 1a*; *vegfa*, *vascular endothelial growth factor a* gene; ex6, exon6; 60 and 1300 refer to the number of nucleotides; s, sense; a, antisense; m, mutated to create a stop codon; 5′ AmC6, 5′ Amino modifier C6 block; 3′ InvdT, 3′ Inverted dT block; ND, non-determined.

presence of the modified RFLP (data not shown). This indicates that the LOH observed at the *gol* locus was not induced by homologous DNA replacement.

We next wanted to determine if smaller css*gol*-NLS-RecA-Gal4 filaments were able to induce LOH at the *gol* locus. The 60-bp oligos were designed that are complementary to exon 6, named ss*gol*ex6m-60 sense (s) and ss*gol*ex6m-60 antisense (a) (Figure 1D and Supplementary Table S1). The *gol* oligos were designed to contain stop codons in the center of the 60-bp sequence to create an RFLP and mutate the locus if the oligos replaced the endogenous *gol* gene by recombination. The RFLP would be detected as a change from an endogenous AfeI to an artificial HindIII restriction site. When the 60-bp oligos were coated with NLS-RecA-Gal4 protein, mixed together and injected into *gol*^{b1} heterozygous embryos, LOH was observed in 2.5% of the injected embryos as seen by patches of missing pigmentation in the eyes (Figure 1B and Table 1). However, analysis of DNA isolated from embryos displaying LOH at the *gol* locus following injection of css*gol*ex6m-60-NLS-RecA-Gal4 filaments did not reveal the expected RFLP (data not shown). These results indicated that css*gol*-NLS-RecA-Gal4 filaments as short at 60 bp are able to mutate the targeted gene, but the mutations are not likely induced by recombination.

To rule out that the induced LOH observed was a result of recombination of the cssDNA-NLS-RecA-Gal4 filaments, we tested the ability of css oligos without stop codons to promote LOH at the *gol* locus. Two adjacent pairs of short css *gol* oligos, named *gol*ex6-60-1 and -2, were designed with wild-type sequences complementary to exon 6 of the *gol* gene (Figure 1D and Supplementary Table S1). Injection of either complementary pair of *gol*ex6-60-1 or -2-NLS-RecA-Gal4 filaments into *gol*^{b1} heterozygous embryos resulted in LOH at a frequency similar to the css*gol*ex6m-60-NLS-RecA-Gal4 filaments (Figure 1D and Table 1). This frequency was not enhanced by co-injection of both css*gol*ex6-60-1 and -2-NLS-RecA-Gal4 filaments. In control experiments we tested whether injection of naked complementary ss oligos without NLS-RecA-Gal4 protein, in various combinations, could drive LOH (Table 1). Patches lacking pigment were never observed in these experiments, and demonstrate that the activity of the cssDNA-RecA filaments to promote LOH is dependent upon the NLS-RecA-Gal4 protein (Table 1). The ability of css-*gol* NLS-RecA-Gal4 filaments containing wild-type sequence to promote LOH provide further support that the mechanism is independent from recombination-induced gene replacement.

We next tested if both complementary ssDNA-NLS-RecA-Gal4 filaments were required to induce LOH. Injection of either sense or antisense ss*gol*ex6m-60-NLS-RecA-Gal4 filaments alone into *gol*^{b1} heterozygous embryos does not induce detectable LOH at the *gol* locus (Table 1). Furthermore, when non-paired but adjacent filaments produced with NLS-RecA-Gal4 (either ss*gol*ex6-60-1 sense with ss*gol*ex6-60-2 antisense or ss*gol*ex6-60-1 antisense with ss*gol*ex6-60-2 sense) were injected into *gol* heterozygous embryos, the LOH was not detected (Table 1). These results show that complementary ssDNA NLS-RecA-Gal4 filaments are required for detectable induction of LOH at the *gol* locus.

Control injections of ss or complementary *gol*ex6-60-1 or -2 oligos alone were unable to induce LOH when injected into *gol*^{b1} heterozygotes (Table 1), showing a dependence of the induction of LOH upon the NLS-RecA-Gal4 protein. Likewise, native RecA or NLS-RecA filaments produced with css-*gol*ex6-60-1 and -2 oligos were unable to produce detectable levels of LOH at the *gol* locus (Table 1). To test if the effect of the NLS-RecA-Gal4 protein was due to the dimerization domain in the Gal4 domain, the dimerization domain was removed and the metal-binding domain was left intact (Figure 1A). Filaments produced with css-*gol*ex6-60-1 and -2 oligos and the dimerization-deleted form of RecA, called NLS-RecA-Gal4ΔDD, were still able to induce LOH at a somewhat reduced efficiency (1%) when injected in to *gol*^{b1} heterozygotes (Table 1). This result suggested that while the dimerization domain contributes to the activity of the NLS-RecA-Gal4 protein, other features in the chimeric RecA protein are also important.

To test whether 5′-phosphate group on the targeting filament is required to induce LOH, css*gol*ex6-60-2 oligos were synthesized with a 5′-amino modification at the C6 group (5′ AmC6) that blocks the phosphate group. Previous work has shown that the 5′ AmC6 does not affect the ability of RecA protein to bind to the oligo and form a filament (19). However, 5′ AmC6 does impair concatemer formation (often called the train effect) by RecA filaments, which can link up >50 filaments end to end. We also tested whether the 3′-hydroxyl group was required to induce LOH by using an inverted deoxythymidine 3′-end modifier (3′ InvdT). This modification inhibits DNA polymerase-driven primer extension from the 3′ end of the filament (20). As shown in Table 1, both 5′ AmC6 or 3′ InvdT filaments produced with NLS-RecA-Gal4 retain the activity to induce LOH at the *gol* locus, indicating either 5′-phosphate group or the free 3′-hydroxyl group of DNA filaments is not required to induce LOH activity.

Finally, to determine whether the LOH observed at the *gol* locus is dependent upon the complementary sequence of the oligo in the css-NLS-RecA-Gal4 filaments to the *gol* locus, 60-bp css oligos were designed complementary to the *flh* (21), *prom1a* (22) and *vascular endothelial growth factor a* (*vegfa*) (23) genes (Supplementary Table S1). The oligos were coated with NLS-RecA-Gal4 protein to produce filaments *css flh-60*, *css prom1a-60* and *css vegfa-60* and injected into 298, 411 and 542 embryos, respectively, which were heterozygous for the *gol*^{b1} mutation. LOH was not detected at the *gol* locus in the 862 embryos examined (Table 1), indicating a requirement for complementarity of the filaments to their target. These results suggest that the complementary sequences within the filament direct the specific targeting of LOH at the *gol* locus.

Site-specific insertion of a reporter gene by cssDNA-NLS-RecA-Gal4 filaments

The data presented above indicated that the css-*gol*-NLS-RecA-Gal4 filaments are able to promote LOH at the *gol* locus. Our analysis of DNA from the mutated embryos indicated that the disrupted *gol* allele was not caused by gene replacement, resulting in mutation of the locus. However, LOH could be a result of DNA DSB, including chromosomal loss, targeted deletions or rearrangements of the *gol* locus. To test whether DSBs might contribute to the LOH, we co-injected a gene trap that contains an EGFP reporter gene and the css-*gol*-NLS-RecA-Gal4 filaments. We reasoned if the css-*gol*-NLS-RecA-Gal4 filaments were promoting site-specific DSBs at the *gol* locus, at some frequency we would observe integration of the gene trap during the repair of the locus by the non-homologous end-joining (NHEJ) pathway. The NHEJ repair pathway is the predominant pathway in the early zebrafish embryo (24–26). The reporter/gene trap used in this assay is a linear dsDNA construct that contains a SA from the carp β-actin gene followed by EGFP coding sequence that lacks the first AUG (27). These sequences are followed by the zebrafish β-actin 3′-UTR/PolyA signal (pA). The EGFP gene trap was placed in all three reading frames following the SA (Supplementary Table S3). Integration of the EGFP gene trap in an intron is expected to produce a fluorescent fusion protein from the targeted gene in one of six integrations (three reading frames and two possible orientations). The detected fluorescence signal is expected to correlate with the endogenous expression of the targeted gene if correct targeting is accomplished. A css-*gol*-270-NLS-RecA-Gal4 filament complementary to 270 bp of the *gol* locus spanning exons 4 and 5 was co-injected with linear dsDNA for all three frames of the EGFP gene trap into wild-type embryos. Following injection expression of EGFP was observed in the retinal epithelium of the eye where the *gol* gene is normally expressed in 5.1% of the embryos (Figure 2A and Table 2). Similar results were obtained with a second filament, css-*gol*-300-NLS-RecA-Gal4 that has a sequence complementary to 300 bp spanning introns 5 to 6 of the *gol* gene. In three experiments, expression of EGFP was observed in the eyes in 8.5% of the embryos (Table 2). For each filament EGFP expression in regions of the embryo outside the eye was only observed in a few embryos, suggesting effective targeting by the NLS-RecA-Gal4 filaments to the *gol* locus. The results suggest that filaments produced from two different regions of the *gol* gene are able to direct integration into the *gol* locus.

When the *flh* gene is targeted in a similar manner, expression of EGFP is observed in the notochord where the *flh* gene is normally expressed during embryogenesis (21) in ~8.5% of the injected embryos (Figure 2B and Table 2). The *prom1a* gene was also targeted and embryos examined for EGFP expression in the eye and central nervous system (22). Following injection of css-*prom1a*-NLS-RecA-Gal4 filaments with the EGFP gene trap, retinal and brain expression of EGFP was observed in 18.8% of the injected embryos (Figure 2C and Table 2). Taken together, these observations are consistent with the integration of the reporter gene construct following repair of a DSB at the targeted locus.

We examined whether the EGFP gene trap could be detected in the *gol* gene in genomic DNA isolated from

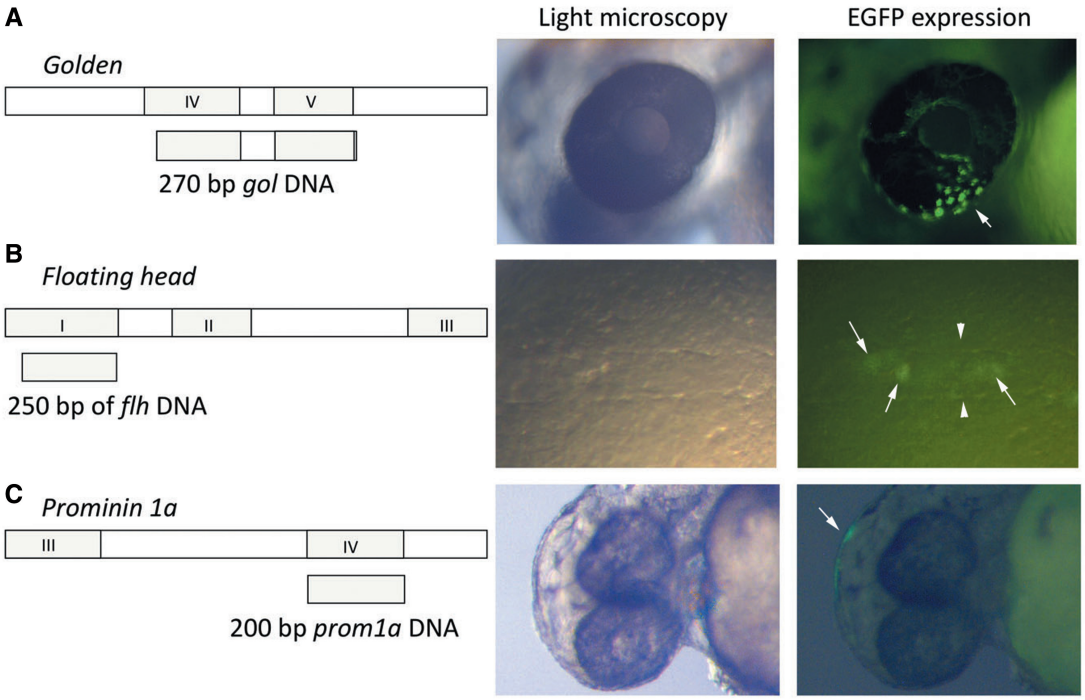

Figure 2. Expression of an EGFP gene trap is consistent with site-specific integration into the *gol*, *flh* and *prom1a* loci. Single-stranded NLS-RecA-Gal4 filaments complementary to regions of the *gol* (**A**), *flh* (**B**) and *prom1a* (**C**) genes were co-injected with the EGFP gene trap cassettes. EGFP expression consistent with targeting gene expression was observed in 5–19% of the injected embryos (Table 2). For the *gol* gene, expression was observed in the retina at 3 dpf (white arrows, A), for the *flh* gene expression was detected in the notochord at 14 hpf (arrows, B; arrowheads mark the notochord boundary) and for the *prom1a* gene expression was detected in the dorsal diencephalon at 30 hpf (arrows, C).

Table 2. Injection of a gene trap with NLS-RecA-Gal4 filaments results in EGFP expression that reflects the expression of the targeted gene

Targeted gene	Somatic reporter gene expression			
	css filaments	Total embryos	Surviving	Fluorescence (%)
golden	*gol*-270	791[a]	649	33 (5.1)
	gol-300	207	165	14 (8.5)
floating head	*flh*-250	216	212	18 (8.5)
prominin 1a	*prom1a*-200	162	154	29 (18.8)

Numbers associated with the filaments refer to the number of nucleotides.
[a]Data were collected from three individual experiments.

embryos following co-injection of either css-*gol*-270-NLS-RecA-Gal4 or css-*gol*-300-NLS-RecA-Gal4 filaments. Because these filaments target different regions in the same gene, we could test whether insertion of the EGFP reporter gene clustered around the genomic sequence in the co-injected filament. Following co-injection of either filament with the EGFP gene trap, embryos displaying EGFP fluorescence were selected for DNA isolation and PCR analysis. To analyze potential junction fragments between the reporter gene trap and the *gol* gene, primers were designed to PCR amplify between the reporter gene trap (Primer III, Figure 3A) and regions of the *gol* gene proximal to the position corresponding to the filaments (Primer I or Primer II, Figure 3A). PCR on genomic DNA isolated from embryos co-injected with the reporter gene trap and css-*gol*-270-NLS-RecA-Gal4 (primer Pairs I and III) led to amplification of several PCR products (Figure 3B, left panel). Similar PCR results with multiple products were observed after amplification using primer Pair II and III on DNA isolated from embryos co-injected with the reporter gene trap and css-*gol*-300-NLS-RecA-Gal4 filaments. Sequencing of these products revealed that they represented junction fragments between the reporter gene trap and the *gol* gene. Insertion sites were observed both 5′ and 3′ to the region complementary to the css-*gol*-270-NLS-RecA-Gal4 filament (Figure 3C black triangles, and Supplementary Table S4). Remarkably, the same junction fragment that corresponds to the 3′-end of the css-*gol*-270-NLS-RecA-Gal4 filament was recovered independently in four different embryos, and an additional embryo produced a junction fragment that differed by only one nucleotide. Regions 5′ to the css-*gol*-270-NLS-RecA-Gal4 filament also had common insertion site. Amplification with primer Pairs I and III was specific to the injection from the css-*gol*-270-NLS-RecA-Gal4 filament and did not amplify products from DNA isolated from css-*gol*-300-NLS-RecA-Gal4 filament-injected embryos, with the exception of one embryo that produced false amplification products as revealed by sequencing. The same was true for primer Pairs II and III, which only amplified PCR products from DNA isolated from embryos co-injected

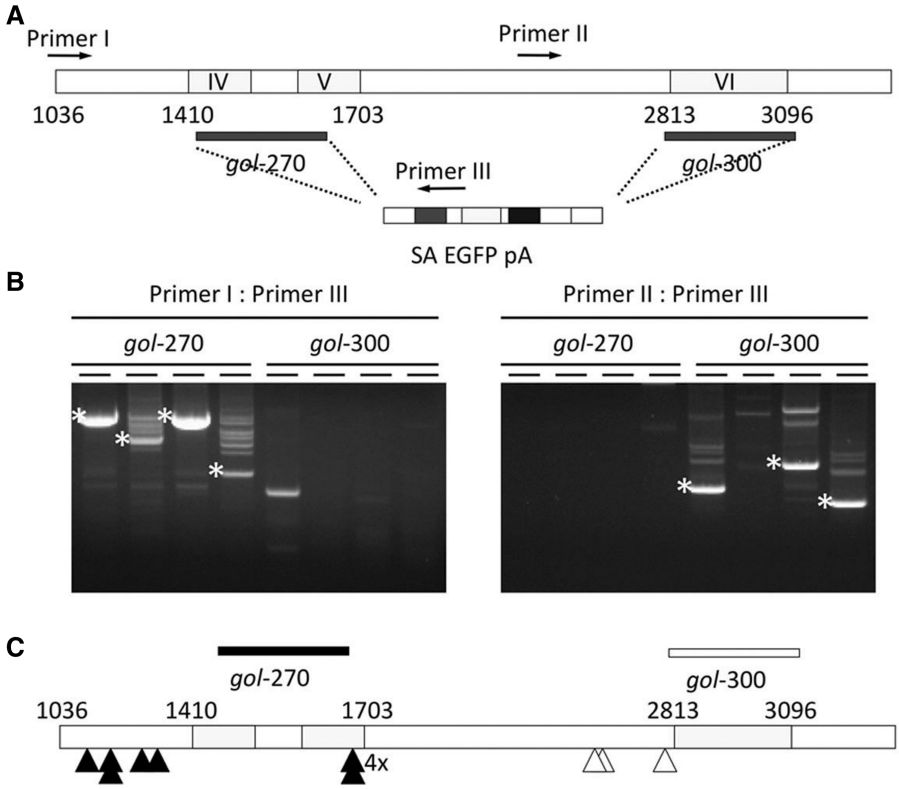

Figure 3. Complementary ss *gol*-NLS-RecA-Gal4 filaments direct site-specific insertion of a gene trap into the *gol* locus. (**A**) Two regions of the *gol* gene corresponding to exons 4 through 5 (*gol*-270, 270 bp) and exon 6 (*gol*-300) were amplified, denatured and coated with NLS-RecA-Gal4 protein to make complementary ss-*gol*-NLS-RecA-Gal4 filaments. Filaments were injected with the EGFP gene trap that contains a SA followed by EGFP in three reading frames and a poly adenylation signal (pA). (**B**) PCR amplification of junction fragments between the EGFP gene trap and the endogenous *gol* locus was observed with DNA isolated from individual embryos. Only primers proximal to the region complementary to the filament yielded amplification products that could be verified by sequencing (asterisks). (**C**) Junction fragments allow insertions to be mapped to the *gol* locus. The *gol*-270 filaments (black bar) resulted in insertion of the EGFP gene trap 5′ and 3′ to the region complementary to the filaments (black triangles). In some cases, identical junction fragments were observed from different embryos (triangles marked 4×). The triangle marked 4x corresponds to the 3′-end of the *gol*-270 filament. The *gol*-300 filament (clear bar) also promoted insertion near regions 5′ to the complementary sequence in the *gol* gene (clear triangles).

with the reporter gene trap and the css-*gol*-300-NLS-RecA-Gal4 filaments. Analysis of these PCR products revealed junction fragments that sit 5′ to the css-*gol*-300-NLS-RecA-Gal4 filaments sequence (Figure 3C, white arrows). While the degree of off-target integration was not assessed in these experiments, our results indicate that the NLS-RecA-Gal4 filaments direct integration of the EGFP reporter gene to regions near the complementary genomic location of the filament.

We used similar molecular analyses to examine targeting of the EGFP reporter to the *prom1a* gene. Five junction fragments between *prom1a* and EGFP were identified from eight embryos co-injected with the EGFP gene trap and the css-*prom1a*-200-NLS-RecA-Gal4 filaments (Supplementary Figure S2 and Table S4). All insertion sites detected were in a region proximal to the css-*prom1a*-200 filament-targeting site. These data are consistent with the pattern of EGFP expression that was observed in *prom1a*-targeted embryos. These results indicate that the targeting mediated by NLS-RecA-Gal4 filaments was not unique to the *gol* locus.

To estimate the frequency of targeted integration of the EGFP gene trap, the junction fragments between the reporter gene trap and the *gol* gene were analyzed further using real-time PCR. Primers were designed for real-time PCR to amplify between the EGFP gene trap (primer GFP-R) and regions proximal to the *gol*1715 position (primer *gol*-F), a position for which we isolated several integration events from multiple embryos (Figure 3B, left panel and Supplementary Table S4). A similar size 322 bp PCR fragment from the *flh* gene was amplified with primers *flh*-F and *flh*-R and used as a reference for a single-copy gene to determine the ratio of the frequency of integration events to a single copy locus. Cycle of threshold (C_t) values of 34.9 and 35.2 were observed with the *gol*-F and GFP-R primers from two embryos that were co-injected with the EGFP reporter gene trap and css-*gol*-270-NLS-RecA-Gal4 filaments (Supplementary Figure S3). The ΔC_t values corresponding to 2.5×10^{-4} and 1.7×10^{-4} per embryonic cell were determined using the reference (*flh*) genomic DNA amplicon from each embryo and represent the frequency of a single integration event. The relative frequency of a non-specific integration in the *flh* locus by css-*gol*-270-NLS-RecA-Gal4 mediated targeting was tested by the same real-time PCR conditions except the *flh*-F and GFP-R primers were used for amplification. As expected

based on the low frequency of 'off-target' expression of the EGFP gene trap, none of the *flh*-EGFP amplifications yielded products near the threshold value within 40 cycles. Instead, the values were similar to negative control reactions that lack DNA template (Supplementary Figure S3A). This indicates that targeted integration of the EGFP gene trap occurs at least at a 100-fold greater frequency than random integration. Comparison of the amplification of *gol*-EGFP junction fragments to a single-copy genomic locus suggests that the css-DNA-NLS-RecA-Gal4 filaments promote site-specific integration in ~1 out of every 10 000 cells.

Injection of css-gol-DNA-NLS-RecA-Gal4 filaments results in mutations that transmit to the next generation

Embryos displaying EGFP expression in the eye after co-injection of css-*gol*-DNA-NLS-RecA-Gal4 filaments with the reporter gene trap into wild-type embryos were grown to adulthood and screened for transmission of targeted alleles in the *gol* gene to the next generation. The adult F0 zebrafish were tested for non-complementation with the gol^{b1} allele by crossing to homozygous gol^{b1} fish. Out of 21 F0 adults screened, two F0 fish produced offspring that failed to complement the gol^{b1} allele (Figure 4A and Table 3), which corresponds to 9.5% of the injected F0 zebrafish. The germlines of these two founder fish are mosaic, with 0.7 and 3.7% of the offspring showing failure to complement the gol^{b1} allele (Table 3). One of the F0 founders was injected with css-*gol*-270-NLS-RecA-Gal4 filaments and the other with css-*gol*-300-NLS-RecA-Gal4 filaments (Figure 3). None of the offspring from either founder displayed fluorescence from the EGFP gene trap. This suggested that the new mutant alleles of *gol* were not insertion alleles of the EGFP gene trap, but could represent deletions or rearrangements.

Wild-type embryos (WIK) that were injected with css*gol*m-1381-NLS-RecA-Gal4 filaments were also grown to adulthood and screened for non-complementation with the gol^{b1} allele. In total, 54 F0 zebrafish were screened and 5 zebrafish were identified that produced offspring that failed to complement the gol^{b1} allele (Table 3). This corresponds to a frequency of 9.3% of the injected F0 zebrafish harboring a new mutant allele of *gol*.

The F1 offspring that failed to complement the gol^{b1} allele were examined by PCR of sequences surrounding the complementary filament to identify the molecular lesion in the new mutated *gol* alleles. We isolated DNA for sequence analysis from the F1 offspring of Founders 1, 3, 4 and 5. Sequencing of the amplified products from *gol* locus only revealed the presence of sequences from the gol^{b1} allele (data not shown), suggesting that the newly recovered alleles in the *gol* gene might correspond to deletions of the gene. Consistent with this, embryos from F0 founders that failed to complement the gol^{b1} allele often displayed cardiac edema (Founders 3–5, Table 3) and were difficult to maintain. Because this is a phenotype not observed in other alleles of *gol*, the phenotypic

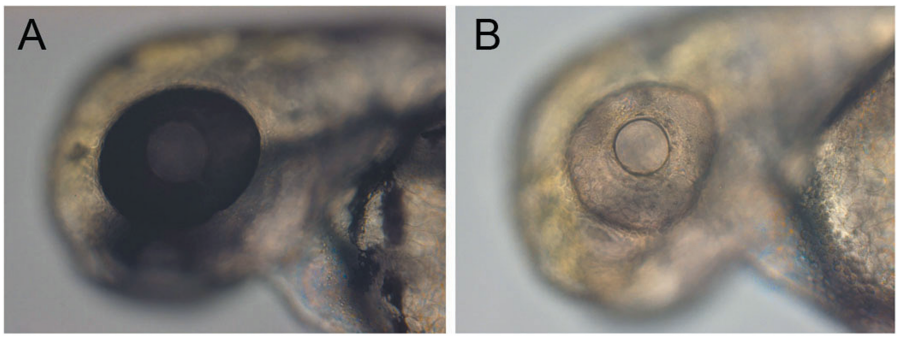

Figure 4. Targeted mutations of the *gol* locus using *gol*-NLS-RecA-Gal4 filaments were transmitted through the zebrafish germline. (**A** and **B**). A testcross between a F0 zebrafish injected with *gol*-NLS-RecA-Gal4 filaments and a gol^{b1} homozygote produces offspring that fail to complement the gol^{b1} allele (B) and siblings with normal pigment (A).

Table 3. Injection of css DNA NLS-RecA-Gal4 filaments results non-complementation of the gol^{b1} mutation in the offspring

Targeting Filament	Founder (%)	Founder ID	*gol* progeny (%)	Edema	gol^{b1} Only	z46013 2.0 CM	*gol*in8 4.3 CM	z9404 7.0 CM	z8488 32.5 CM	z21330 38.5 CM
css *gol*-270	1/13 (7.7)	1	34/919 (3.7)	None	Yes	Absent	Absent	Absent	Absent	Present
css *gol*-300	1/8 (12.5)	2	1/139 (0.7)	None	ND	ND	ND	ND	ND	ND
css *gol*m-1381	5/54 (9.3)	3	1/126 (0.8)	Edema	Yes	ND	ND	ND	ND	ND
		4	5/60 (8.3)	Edema	Yes	Absent	Absent	ND	ND	Absent
		5	4/210 (1.9)	Edema	Yes	ND	ND	ND	ND	ND
		6	1/79 (1.3)	None	ND	Absent	Absent	Absent	NI	Present
		7	1/103 (1.0)	None	ND	Absent	Absent	Absent	NI	NI

Founder percentage is the percent of F0 fish whose offspring failed to complement the gol^{b1} mutation. *gol* progeny % is the percent of offspring displaying the *gol* phenotype. Edema refers to whether the progeny displayed pericardial edema. gol^{b1} only indicates that only the gol^{b1} allele was identified by PCR. ND - not determined. NI—not informative due to a lack of polymorphic loci.

observation suggests additional genetic alterations are present in the embryos that fail to complement the *gol*^{b1} allele.

To determine if the new *gol* alleles correspond to deletions, we used polymorphic z-markers or PCR-amplified simple SSLPs to examine the genotypes of the F1 embryos around the *gol* locus on chromosome 18. In total, 22 SSLPs or z-markers were tested to identify polymorphic loci for Founders 1, 4, 6 and 7 for which we could identify the parental and test cross genotypes. Polymorphic loci were not identified for Founders 2, 3 and 5. For Founder 1, we were able to identify polymorphic amplification products between the wild-type strain used for injection experiments and the AB strain that harbors the *gol*^{b1} allele using z-markers z46013, z9404, z8488 and z21330 (Figure 5A and Table 3). Marker z46013 is close to the telomere and the *gol* locus, while z9404, z8488 and z21330 are ~7, 32.5 and 38.5 CM away from the telomere, respectively (http:/zfin.org). Founder 1 produced offspring from the test cross that lacked the z46013, z9404 and z8488 markers and retained one of the z21330 markers from the wild-type-injected parent. This suggests the deletion proximal breakpoint maps between z8488 and z21330 and is 32.5 CM to 38.5 CM in size. We also identified a polymorphic marker, called *gol*in8, within intron 8 of the *gol* gene that was also lost from the wild-type background (Figure 5A), further confirming the deletion of the *gol* locus in embryos derived from Founder 1.

For Founders 4 and 6, we were able to identify SSLP markers z46013, *gol*in8, z9404 and z21330 that displayed polymorphic amplification products in the wild-type WIK strain used for injection experiments and the AB strain that harbors the *gol*^{b1} allele (Figure 5B and C and Table 3). DNA from offspring of Founder 4 crossed with *gol*^{b1} only amplified the SSLPs associated with the *gol*^{b1} parent, suggesting a large deletion >38.5 CM is present on chromosome 18 (Figure 5B). In contrast, Founder 6 produced offspring from the test cross that lacked the z46013, *gol*in8 and z9404 products from the WIK-injected parent but retained z21330 (Figure 5C). This suggests a smaller deletion with a proximal breakpoint between 7 and 38.5 CM is associated with the mutant *gol* allele in this founder. Founder 7 produced offspring that failed to complement the *gol*^{b1} allele and lacked the z46013, *gol*in8, z9404 SSLPs associated with Founder 7, indicating deletion of the *gol* locus and at least 7 CM of distal chromosome 18 (Figure 5C). None of the other SSLPs was informative for Founder 7 (Table 3). Together, these results suggest that the targeted gene disruptions recovered at the *gol* locus were passed to the next generation and that the disruptions predominantly comprise large chromosomal deletions.

DISCUSSION

We have demonstrated that complementary ssDNA-NLS-RecA-Gal4 filaments targeted to the *gol* locus are able to induce LOH at this locus after injection into zebrafish embryos. We show that the NLS-RecA-Gal4 protein is able to coat cssDNA (Supplementary Figure S1). When cssDNA–protein complex is injected into the zebrafish embryos, the cssDNA-NLS-RecA-Gal4 filaments likely undergo a homology search to find homologous chromosomal DNA, similar to the behavior of native RecA. In support of this, we did not detect LOH at the *gol* locus when cssDNA-NLS-RecA-Gal4 filaments complementary to *vegfa*, *floating head* and *prom1a* were used (Table 1), suggesting that the targeted disruption of the *gol* locus was dependent upon the DNA sequence in the css-*gol* DNA-NLS-RecA-Gal4 filaments. We propose a model for this activity where the css-*gol* DNA-NLS-RecA-Gal4 filaments target the *gol* locus for disruption by creating arrested replication forks that result in DSBs (Figure 6). Alternatively, the cssDNA-NLS-RecA-Gal4 filaments could promote recruitment of components of the DNA repair pathway and endogenous nucleases to specific chromosomal sites, resulting in deletions or site-specific insertion of exogenous DNA.

NLS-RecA-Gal4 protein may form a dimer and stabilize the complex of single-strand sense (ss-s) and antisense (ss-a)-NLS-RecA-Gal4 filaments on the targeted genomic region through interactions of the Gal4 domain (Figure 6). This activity requires the Gal4 DNA binding/dimerization domains in NLS-RecA-Gal4 filaments (Table 1). RecA proteins that lack the Gal4 domain, such as native RecA or NLS-RecA, produce *gol* filaments that do not result in detectable levels of LOH at the *gol* locus. The dimerization domain of Gal4 is likely partially required for this activity too, as reduced LOH at the *gol* locus was observed after injection of *gol* filaments produced with the NLS-RecA-Gal4ΔDD protein (Table 1). This suggests that the interaction of NLS-RecA-Gal4 proteins via the dimerization domain may occur between individual css-*gol* DNA-NLS-RecA-Gal4 filaments and is at least partially required for the LOH activity. The filaments could also invade one another and form higher order structures on the DNA that could block a replication fork. This possibility would explain the ability of complementary filaments made with the NLS-RecA-Gal4ΔDD protein to still promote LOH, albeit at a reduced frequency compared to filaments with the NLS-RecA-Gal4 protein. Alternatively, the Gal4 domain could create a dominant negative form of RecA that is able to stabilize the interaction of complementary filaments. Moreover, interaction between complementary filaments may also be required for detectable LOH at the *gol* locus because injection of either the sense or antisense filaments alone has no detectable effect. The interaction of the complementary NLS-RecA-Gal4 filaments could occur following the formation of D-loop structures on both strands of DNA (Figure 6). Once the chromosomal target is located by homologous pairing, the cssDNA-NLS-RecA-Gal4 filaments initiate DNA strand invasion. We suggest that during the stand invasion and strand exchange steps, the ssDNA filament will invade and unwind its homologous ds genomic DNA, resulting in the formation of D-loop structures. This could allow the complementary filaments to directly interact with one another via the Gal4 domain in NLS-RecA-Gal4 protein.

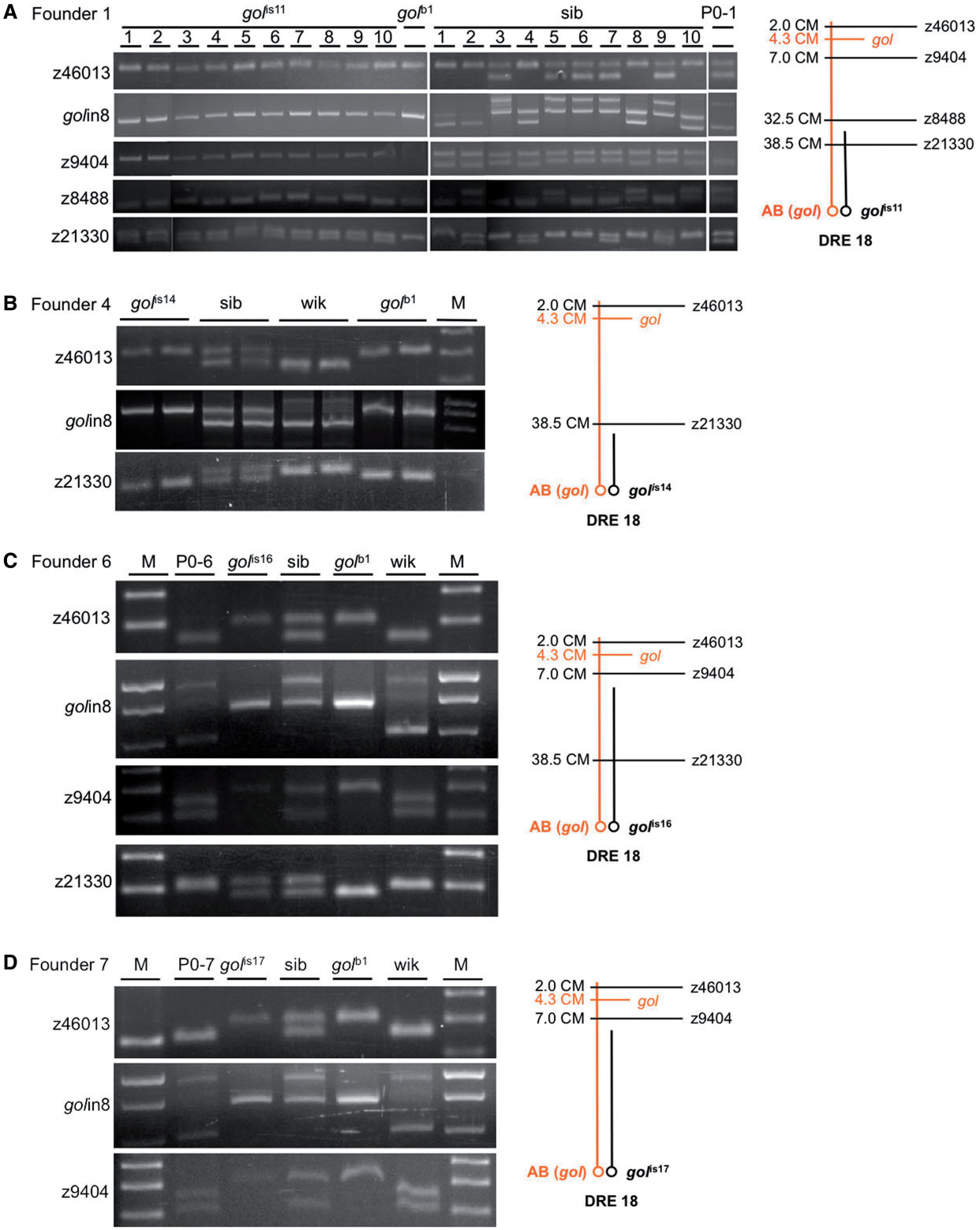

Figure 5. Mapping of new *gol* alleles from Founders 1, 4, 6 and 7 with z-markers reveals deletions at the *gol* locus. Founders were crossed with gol^{b1} homozygotes and DNA isolated from offspring displaying the gol^{b1} phenotype and from siblings (sib) with wild-type pigmentation. Parental lines (P0 and gol^{b1}) or the wild-type strain used for injection (WIK) is shown for comparison. DNA was amplified with primer pairs to identify SSLPs that were polymorphic. (**A**) Offspring from Founder 1 that fail to complement the gol^{b1} mutation have one allele for markers SSLPs z46013, *gol*in8, z9404 and z8488, but are heterozygous for marker z21330. This indicates a deletion of 32.5–38.5 CM, with a proximal breakpoint boundary that maps between z8488 and z21330. (**B**) In offspring from Founder 4 the only SSLP alleles present were from the gol^{b1} parent, indicating the deletion is >38.5 CM. (**C**) In the offspring from Founder 6 that failed to complement the gol^{b1} mutation, a single allele was present for marker z9404 but heterozygous for marker z21330, indicating a deletion of at least 7 CM. (**D**) Founder 7 produced offspring with a deletion that contained only marker z9404, indicating a deletion of at least 7 CM.

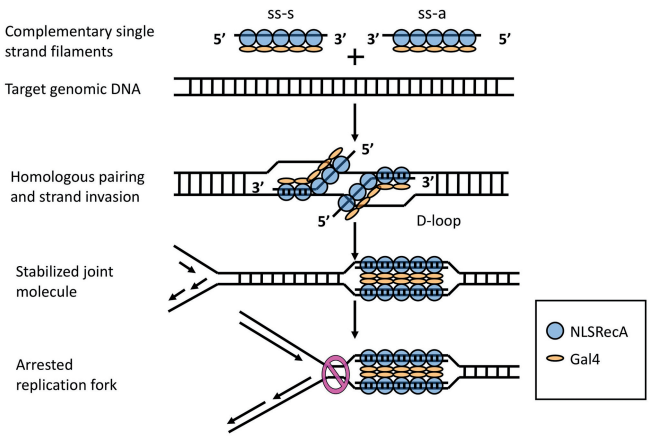

Figure 6. A model for gene targeting using complementary ss DNA-NLS-RecA-Gal4. Both sense and antisense single-strand (ss-s and ss-a)-NLS-RecA-Gal4 filaments are co-injected into zebrafish embryos. The RecA homology search activity can guide the filaments to the targeted region. The cssDNA-NLS-RecA-Gal4 filaments can undergo homologous pairing and strand invasion. This causes the formation of D-loops that are stabilized by the Gal4 dimerization domains between the complementary filaments on the target chromosome. This compact DNA joint molecule is assumed to block replication fork progression, leading to a DNA DSB.

Because the ssDNA-RecA filament disassembles upon ATP hydrolysis, we used a non-hydrolytic form of ATP, ATPγS. We reasoned that, like RecA, ATPγS would make more stable cssDNA-NLS-RecA-Gal4 filaments. Some proteins that tightly bind DNA, such as mutated transcriptional machinery, can impede replication fork progression on chromosomes, causing stalled or collapsed replication forks (28,29). Arrested replication forks can increase the stress on the chromosome, causing the formation of DNA DSB and replication fork collapse (30). We suggest that the css-*gol* DNA-NLS-RecA-Gal4 filaments create a steric block to DNA replication during the rapid (15-min) cell cycles during early zebrafish development, resulting in arrested replication forks.

The detection of targeted chromosomal deletions and integration of the EGFP gene trap near the *gol* locus is consistent with the ability to create site-specific stalled replication forks (Figures 2 and 3 and Supplementary Figure S3). This is different from randomly induced stalled replication forks from chemical inhibition of replication (31) or mutant DNA-binding proteins (28). Arrested replication forks can be processed in repair-independent and repair-dependent manners. If the replication fork is not repaired, the free DNA fragment can be lost, causing a large deletion. This is consistent with our observations that at least four of the new alleles of *gol* that we recovered are the result of deletions (Figure 4). Alternatively, the free DNA fragment could also move to a different genomic region, resulting in a chromosomal translocation (28). A large deletion or translocation can cause a deficient haploid allele during meiosis in which the resulting phenotype can be detected by a complementation test, similar to our results, or by single-generation haploid screening (32).

Arrested replication forks are also a target for nuclease digestion as part of repair-dependent processing, resulting in the formation of DNA DSBs (28). For example, Endonuclease VII from T4 bacteriophage cleaves arrested replication forks induced from antitumor drugs that complex with type II topoisomerases (33). In yeast, branched DNA structures at stalled replication forks are digested by an endonuclease formed by the association of Mus81 and Mms4/Eme1 (34). This kind of DSB creates a shorter cleft than ones caused by replication fork collapse. Short DSBs are efficiently repaired by either homologous recombination or the NHEJ. Because the NHEJ repair pathway predominates during early zebrafish embryogenesis (24), we expect that repair by this pathway would either be precise or result in small deletions or insertions (25,26).

Arrested replication forks can be repaired by a variety of mechanisms that would also be consistent with the observation that the EGFP gene trap was inserted into the *gol* locus proximal to the targeted region (Figure 3). Stalled replication forks can often trigger cell cycle arrest and stimulate a DSB repair mechanism by either the homologous recombination or the NHEJ pathway (35). In early zebrafish embryos the NHEJ pathway overshadows the homologous recombination pathway as the predominant repair pathway (24). This repair pathway could also ligate the EGFP gene trap into a stalled replication fork that was cleaved by an endonuclease. During repair of a DNA DSB, the NHEJ pathway may insert the EGFP gene trap, resulting in detection of EGFP consistent with the pattern of expression of the *gol* gene in the pigmented epithelium of the eye. Our detection of integration of the EGFP reporter gene proximal to the targeted region with the css-*gol* or DNA-NLS-RecA-Gal4 filaments is consistent with this interpretation (Figures 2 and 3). This mechanism appears to be operational at other chromosomal sites, as targeting the *flh* and *prom1a* loci with cssDNA-NLS-RecA-Gal4 filaments and the EGFP gene trap results in EGFP expression that is similar to the known expression patterns of these loci (Figure 2 and Supplementary Figure S2).

The directed integration of the EGFP gene trap likely occurs at a lower frequency than the creation of large deletions. For at least four of seven new *gol* alleles recovered, we were only able to detect genomic sequences from the gol^{b1} allele in embryos from the non-complementation tests, suggesting most of the newly recovered alleles represent deletions. In addition, for the alleles created using css-*gol* DNA-NLS-RecA-Gal4 filaments and the EGFP gene trap (Founders 1 and 2), EGFP expression was not detected in the offspring nor was DNA corresponding the EGFP gene trap found in the *gol* locus by PCR (data not shown). The real-time PCR analysis revealed a single integration event of EGFP gene trap mediated by css-*gol* DNA-NLS-RecA-Gal4 filaments reaches a frequency of 1 in 10 000 embryonic cells. Other integration events could also be present in these embryos that we were unable to detect with this method. While integration of the EGFP gene trap could be relatively infrequent compared to deletion at the *gol* locus, the targeted integration events did occur at a greatly increased frequency

compared to integration at random chromosomal sites. This is supported by the very low number of injected embryos with off-target expression of EGFP (Figure 2) and the estimation of random integration at a specific locus with real-time PCR (Supplementary Figure S3). Our data suggest that the co-injected EGFP gene trap can be used as an indicator of gene targeting by screening embryos for fluorescent gene expression patterns that are consistent with the expression pattern of the targeted gene.

The average frequency of recovery of a founder fish with at least one new *gol* allele was 9.3% of the injected embryos (Table 3). The recovery rate of new alleles could possibly be higher if a single founder has more than one new allele; however, this remains to be rigorously tested. From each founder, on average, 2.5% of the progeny failed to complement the gol^{b1} allele. Together, these results correspond to an overall frequency of 1 in 400 (0.093×0.025) recovered gametes that fail to complement the gol^{b1} allele. Interestingly, we did not recover any small deletions at the *gol* locus. This could be due to the rapid cell divisions that occur during early zebrafish embryogenesis that favor repair-independent processing of stalled replication forks. In the absence of repair, stalled replication forks are likely to resolve with chromosome breaks in mitosis. This would lead to the large chromosomal deletions we detected on the *gol*-targeted chromosomes.

Our results demonstrate that the cssDNA-NLS-RecA-Gal4 filaments target large DNA deletions to chromosomes and promote integration of linear DNA into specific chromosomal regions. The frequency of recovering chromosomal deletions suggests this method could be used to generate a deletion series of zebrafish chromosomes. Deletions are useful tools for genetic analyses, for example, to uncover tumor suppressor genes (36) or to evaluate the null phenotype of loss of function alleles. The method we developed for genome modification in zebrafish could also potentially be applied to other species.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Perry Hackett, Scott Fahrenkrug and Daniel Voytas for thoughtful comments during the development of this article. We are grateful to Drs Darius Balciunas and Stephen Ekker for providing the elements used in the gene trap construct. We also thank Drs H. Joseph Yost, David Zarling and Chris Lehman for thoughtful discussions. We are grateful to Dr Maura McGrail for comments on the article and Dr Tom Bobik for assistance with protein purification.

FUNDING

Center for Integrated Animal Genomics at Iowa State University; National Center for Research Resources at the National Institutes of Health (1R21RR25915-01). Funding for open access charge: National Center for Research Resources at the National Institutes of Health (1R21RR25915-01).

Conflict of interest. J.J.E. has an ownership interest in Recombinetics, Inc., which has proprietary rights in technology described in this article.

REFERENCES

1. McGrew,D.A. and Knight,K.L. (2003) Molecular design and functional organization of the RecA protein. *Crit. Rev. Biochem. Mol. Biol.*, **38**, 385–432.
2. Radding,C.M., Shibata,T., DasGupta,C., Cunningham,R.P. and Osber,L. (1981) Kinetics and topology of homologous pairing promoted by *Escherichia coli* recA-gene protein. *Cold Spring Harb. Symp. Quant. Biol.*, **45(Pt 1)**, 385–390.
3. Seitz,E.M., Brockman,J.P., Sandler,S.J., Clark,A.J. and Kowalczykowski,S.C. (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.*, **12**, 1248–1253.
4. Cox,M.M. (1999) Recombinational DNA repair in bacteria and the RecA protein. *Prog. Nucleic Acid Res. Mol. Biol.*, **63**, 311–366.
5. Campbell,M.J. and Davis,R.W. (1999) On the in vivo function of the RecA ATPase. *J. Mol. Biol.*, **286**, 437–445.
6. Reiss,B., Klemm,M., Kosak,H. and Schell,J. (1996) RecA protein stimulates homologous recombination in plants. *Proc. Natl Acad. Sci. USA*, **93**, 3094–3098.
7. Reiss,B., Schubert,I., Kopchen,K., Wendeler,E., Schell,J. and Puchta,H. (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by Agrobacterium. *Proc. Natl Acad. Sci. USA*, **97**, 3358–3363.
8. Shcherbakova,O.G., Lanzov,V.A., Ogawa,H. and Filatov,M.V. (2000) Overexpression of bacterial RecA protein stimulates homologous recombination in somatic mammalian cells. *Mutat. Res.*, **459**, 65–71.
9. Yanez,R.J. and Porter,A.C. (1999) Gene targeting is enhanced in human cells overexpressing hRAD51. *Gene Ther.*, **6**, 1282–1290.
10. Cui,Z., Yang,Y., Kaufman,C.D., Agalliu,D. and Hackett,P.B. (2003) RecA-mediated, targeted mutagenesis in zebrafish. *Mar. Biotechnol.*, **5**, 174–184.
11. Takahashi,N. and Dawid,I.B. (2005) Characterization of zebrafish Rad52 and replication protein A for oligonucleotide-mediated mutagenesis. *Nucleic Acids Res.*, **33**, e120.
12. Kimmel,C.B., Ballard,W.W., Kimmel,S.R., Ullmann,B. and Schilling,T.F. (1995) Stages of embryonic development of the zebrafish. *Dev. Dyn.*, **203**, 253–310.
13. Uhlin,B.E. and Clark,A.J. (1981) Overproduction of the *Escherichia coli* recA protein without stimulation of its proteolytic activity. *J. Bacteriol.*, **148**, 386–390.
14. Keller,M., Harbottle,R.P., Perouzel,E., Colin,M., Shah,I., Rahim,A., Vaysse,L., Bergau,A., Moritz,S., Brahimi-Horn,C. *et al.* (2003) Nuclear localisation sequence templated nonviral gene delivery vectors: investigation of intracellular trafficking events of LMD and LD vector systems. *Chembiochem*, **4**, 286–298.
15. Petzold,A.M., Balciunas,D., Sivasubbu,S., Clark,K.J., Bedell,V.M., Westcot,S.E., Myers,S.R., Moulder,G.L., Thomas,M.J. and Ekker,S.C. (2009) Nicotine response genetics in the zebrafish. *Proc. Natl Acad. Sci. USA*, **106**, 18662–18667.
16. Streisinger,G., Coale,F., Taggart,C., Walker,C. and Grunwald,D.J. (1989) Clonal origins of cells in the pigmented retina of the zebrafish eye. *Dev. Biol.*, **131**, 60–69.
17. Moore,J.L., Rush,L.M., Breneman,C., Mohideen,M.A. and Cheng,K.C. (2006) Zebrafish genomic instability mutants and cancer susceptibility. *Genetics*, **174**, 585–600.
18. Lamason,R.L., Mohideen,M.A., Mest,J.R., Wong,A.C., Norton,H.L., Aros,M.C., Jurynec,M.J., Mao,X., Humphreville,V.R., Humbert,J.E. *et al.* (2005) SLC24A5, a

putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, **310**, 1782–1786.
19. Simonson,T., Kubista,M., Sjoback,R., Ryberg,H. and Takahashi,M. (1994) Properties of RecA-oligonucleotide complexes. *J. Mol. Recognit.*, **7**, 199–206.
20. Dames,S., Margraf,R.L., Pattison,D.C., Wittwer,C.T. and Voelkerding,K.V. (2007) Characterization of aberrant melting peaks in unlabeled probe assays. *J. Mol. Diagn.*, **9**, 290–296.
21. Talbot,W.S., Trevarrow,B., Halpern,M.E., Melby,A.E., Farr,G., Postlethwait,J.H., Jowett,T., Kimmel,C.B. and Kimelman,D. (1995) A homeobox gene essential for zebrafish notochord development. *Nature*, **378**, 150–157.
22. McGrail,M., Batz,L., Noack,K., Pandey,S., Huang,Y., Gu,X. and Essner,J.J. (2010) Expression of the zebrafish CD133/prominin1 genes in cellular proliferation zones in the embryonic central nervous system and sensory organs. *Dev. Dyn.*, **239**, 1849–1857.
23. Liang,D., Xu,X., Chin,A.J., Balasubramaniyan,N.V., Teo,M.A., Lam,T.J., Weinberg,E.S. and Ge,R. (1998) Cloning and characterization of vascular endothelial growth factor (VEGF) from zebrafish, *Danio rerio*. *Biochim. Biophys. Acta*, **1397**, 14–20.
24. Hagmann,M., Bruggmann,R., Xue,L., Georgiev,O., Schaffner,W., Rungger,D., Spaniol,P. and Gerster,T. (1998) Homologous recombination and DNA-end joining reactions in zygotes and early embryos of zebrafish (*Danio rerio*) and *Drosophila melanogaster*. *Biol. Chem.*, **379**, 673–681.
25. Meng,X., Noyes,M.B., Zhu,L.J., Lawson,N.D. and Wolfe,S.A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 695–701.
26. Doyon,Y., McCammon,J.M., Miller,J.C., Faraji,F., Ngo,C., Katibah,G.E., Amora,R., Hocking,T.D., Zhang,L., Rebar,E.J. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 702–708.
27. Liu,Z.J., Moav,B., Faras,A.J., Guise,K.S., Kapuscinski,A.R. and Hackett,P.B. (1990) Development of expression vectors for transgenic fish. *Biotechnology*, **8**, 1268–1272.
28. Michel,B., Flores,M.J., Viguera,E., Grompone,G., Seigneur,M. and Bidnenko,V. (2001) Rescue of arrested replication forks by homologous recombination. *Proc. Natl Acad. Sci. USA*, **98**, 8181–8188.
29. Aguilera,A. and Gomez-Gonzalez,B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.*, **9**, 204–217.
30. Michel,B., Ehrlich,S.D. and Uzest,M. (1997) DNA double-strand breaks caused by replication arrest. *EMBO J.*, **16**, 430–438.
31. Feitsma,H., Akay,A. and Cuppen,E. (2008) Alkylation damage causes MMR-dependent chromosomal instability in vertebrate embryos. *Nucleic Acids Res.*, **36**, 4047–4056.
32. Imai,Y., Feldman,B., Schier,A.F. and Talbot,W.S. (2000) Analysis of chromosomal rearrangements induced by postmeiotic mutagenesis with ethylnitrosourea in zebrafish. *Genetics*, **155**, 261–272.
33. Hong,G. and Kreuzer,K.N. (2003) Endonuclease cleavage of blocked replication forks: an indirect pathway of DNA damage from antitumor drug-topoisomerase complexes. *Proc. Natl Acad. Sci. USA*, **100**, 5046–5051.
34. Boddy,M.N., Gaillard,P.H., McDonald,W.H., Shanahan,P., Yates,J.R. III and Russell,P. (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell*, **107**, 537–548.
35. Michel,B. (2000) Replication fork arrest and DNA recombination. *Trends Biochem. Sci.*, **25**, 173–178.
36. Soares,B.S., Eguchi,K. and Frohman,L.A. (2005) Tumor deletion mapping on chromosome 11q13 in eight families with isolated familial somatotropinoma and in 15 sporadic somatotropinomas. *J. Clin. Endocrinol. Metab.*, **90**, 6580–6587.