Homology Arms of Targeting Vectors for Gene Insertions and Crispr/Cas9 Technology: Size Does Not Matter; Quality Control of Targeted Clones Does

Article *in* Cellular & Molecular Biology Letters · November 2015
DOI: 10.1515/cmble-2015-0047

CITATIONS
2

READS
394

6 authors, including:

Silvia Petrezselyova
Institute of Molecular Genetics AS CR, v.v.i., BIOCEV
21 PUBLICATIONS **226** CITATIONS
SEE PROFILE

Slavomir Kinsky
The Czech Academy of Sciences
7 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Dominika Truban
Mayo Clinic
24 PUBLICATIONS **151** CITATIONS
SEE PROFILE

Radislav Sedlacek
Institute of Molecular Genetics AS CR
157 PUBLICATIONS **1,151** CITATIONS
SEE PROFILE

DE GRUYTER

CELLULAR & MOLECULAR BIOLOGY LETTERS
http://www.cmbl.org.pl

Received: 17 June 2015
Final form accepted: 16 October 2015

DOI: 10.1515/cmble-2015-0047

Short communication

HOMOLOGY ARMS OF TARGETING VECTORS FOR GENE INSERTIONS AND CRISPR/CAS9 TECHNOLOGY: SIZE DOES NOT MATTER; QUALITY CONTROL OF TARGETED CLONES DOES

SILVIA PETREZSELYOVA[1], SLAVOMIR KINSKY[1], DOMINIKA TRUBAN[1], RADISLAV SEDLACEK[1], INGO BURTSCHER[2] and HEIKO LICKERT[2,*]

[1]Institute of Molecular Genetics of the ASCR, v.v.i., Videnska 1083, 142 20 Prague 4, Czech Republic, [2]Institute of Stem Cell Research, Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

Abstract: Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) technology has brought rapid progress in mammalian genome editing (adding, disrupting or changing the sequence of specific sites) by increasing the frequency of targeted events. However, gene knock-in of DNA cassettes by homologous recombination still remains difficult due to the construction of targeting vectors possessing large homology arms (from 2 up to 5 kb). Here, we demonstrate that in mouse embryonic stem cells the combination of CRISPR/Cas9 technology and targeting vectors with short homology arms (~ 0.3 kb) provides sufficient specificity for insertion of fluorescent reporter cassettes into endogenous genes with similar efficiency as those with large conventional vectors. Importantly, we emphasize the necessity of thorough quality control of recombinant clones by combination of the PCR method, Southern hybridization assay and sequencing to exclude undesired mutations. In conclusion, our approach facilitates programmed integration of

* Author for correspondence. Email: heiko.lickert@helmholtz-muenchen.de, phone: +49-89-3187-3867, fax: +49-89-3187-2060

Abbreviations used: CRISPR – clustered regularly interspaced short palindromic repeats; Cas9 – CRISPR-associated 9; DSB – double-strand break; FRT – flippase recognition target; gRNA – guide RNA; HDR – homology-directed repair; HR – homology recombination; mESCs – mouse embryonic stem cells; Neo – neomycin; NHEJ – non-homologous end joining; pBKS – pBluescript KS; OT – off-target; ORF – open reading frame; PGK – phospho-glycerate kinase; PAM – photospacer adjacent motif; RFP – red fluorescent protein; pA – SV40 polyadenylation sequence; TALEN – transcription activator-like effector nuclease; UTR – untranslated region; ZFN – zinc finger nucleases

exogenous DNA sequences at a target locus and thus could serve as a basis for more sophisticated genome engineering approaches, such as generation of reporters and conditional knock-out alleles.

Keywords: CRISPR/Cas9, Genome editing, Reporter, Targeting vector, Homology arms, Embryonic stem cell

INTRODUCTION

In 2007, Mario R. Capecchi, Martin J. Evans and Oliver Smithies were awarded the Nobel Prize in Medicine for introducing gene-specific modifications in embryonic stem cells, thereby generating knock-out and knock-in mouse models. Genetic changes are introduced by homologous recombination that occurs between an engineered exogenous DNA template and the genome of the mouse embryonic stem (mES) cells. Transgenic mES cells are then selected and used for chimera production by blastocyst injection or morula aggregation. Chimeras are backcrossed to achieve germline transmission and to generate transgenic mouse models. This method has been extensively used for decades to decipher the function of many mammalian genes *in vivo*. However, the use of this technique has been hampered due to several limitations: the low frequency of homologous recombination in mammalian cells, labor-intensive and time-consuming selection and screening strategies, as well as construction of large targeting vectors with isogenic homology regions up to 10 kb to increase the targeting efficiency [1].

In recent years, discovery of new genome-editing technologies, such as zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs) and the clustered regularly interspaced short palindromic repeats (CRISPR) associated with CRISPR-associated 9 (Cas9) protein, has revolutionized generation of targeted mutations in many model organisms (see [2] for a review). All these technologies work on the same principle, i.e. sequence-guided DNA endonucleases induce DNA double-strand breaks (DSBs) that subsequently stimulate either non-homologous end joining (NHEJ) recombination and/or homology-directed repair (HDR) at targeted loci. The NHEJ recombination is an error-prone event forming small insertions or deletions (indels) and thus often results in frame-shift mutations. Because the HDR system requires a DNA template to repair a DSB, it is used to introduce a desired mutation or reporter gene into the genome. The template molecule is supplied in the form of either single-stranded oligonucleotides or double-stranded vectors. These vectors are typically designed to contain a selection marker that is flanked by genomic DNA fragments called 5′- and 3′-homology arms. Since it has been generally accepted that long homology arms are beneficial for increased gene-targeting efficiency [3-5], targeting vectors have been routinely constructed with long homology arms up to 10 kb.

Nowadays, with novel genome-editing technologies at hand, the homologous integration of large DNA fragments has greatly enhanced the ability to generate

specific gene modifications. Nevertheless, ZFN and TALEN technologies still require a considerable cloning effort as multiple protein motifs need to be aligned in order to obtain site specificity. The CRISPR/Cas9 RNA-guided genome-editing tool is much simpler to implement. It only requires the Cas9 enzyme and an easily-edited guide RNA (gRNA) comprising a genome target sequence of 20 nucleotides (nt) followed by a 3-nt protospacer adjacent motif (PAM) nucleotide sequence: NGG. When these components are assembled in the nucleus at a specific locus guided by the 23 nt target sequence, Cas9 induces a DSB in the genome 3-4 nt upstream of the NGG site.

In this study, we investigated the possibility to facilitate CRISPR/Cas9-based genome editing, particularly at the level of targeting vector design. We asked whether mES cells will repair a CRISPR/Cas9-induced DSB by homologous integration of a small or large DNA fragment with similar efficiencies. For this purpose, we prepared two types of targeting vectors that served as templates for homologous recombination: i) classical targeting vectors with long homology arms of 2 to 5 kb and ii) mini-targeting vectors carrying homology arms of only ~ 0.3 kb in length. Both types of targeting vectors carrying reporter cassettes were combined with a pair of gRNAs and pair of CRISPR nickases (Cas9D10A mutant version) that generate single strand breaks. A DSB and DNA insertion will occur only if two nicks are in close proximity (less than 200 bp), creating 5′ overhangs. The 5′ single-strand DNA overhangs invade the double-stranded molecule of exogenous targeting construct, thereby facilitating gene knock-in at a specific locus with minimal off-target activity [6].

We demonstrate with three different examples that combination of CRISPR/Cas9 nickases together with mini-targeting vectors leads to an overall insertion frequency comparable to conventional targeting vectors. Moreover, our approach using mini-targeting vectors (~ 0.3 kb homology arms) enables integration of even larger inserts (~ 2.3 kb). By Southern blot analysis we demonstrated that Cas9 nickases in combination with mini-targeting vectors led to insertion of foreign vector DNA also by non-homologous integrations, thus resulting in additional genomic alternations. This phenomenon was not observed when targeting vectors with long homology arms were used. Therefore we strongly recommend that Cas9-mediated integration of exogenous DNA with short homology arms always be accompanied by Southern blot analysis to eliminate clones with undesired genomic alterations at the targeting locus.

MATERIALS AND METHODS

Construction of *Hnf4α*-Venus, TagRFP-*Tcf7L2* and *Ngn3*-Venus targeting vectors, and gRNA expressing vectors

Knock-in targeting constructs for *Hnf4α*, *Tcf7L2* and *Ngn3* genes were designed as shown in Figure 1. Primers used in this study to generate large targeting and mini-targeting vectors are listed in Suppl. Table 1 in Supplementary material at: 10.1515/cmble-2015-0047_sm.

Hnf4α-Venus knock-in cassette was designed to target Exon 10. For the mini-targeting vector, the 5′ homology region (HR) and 3′ HR were PCR amplified using C57B16 BAC (RP23-156K9) as a template and using primers as follows: EP_1170 and EP_1171 primers for 5′ HR and EP_1172 and EP_1173 primers for 3′ HR. 5′ HR was subcloned via *Not*I and *Xba*I and 3′ HR was subcloned via *Hind*III and *Xho*I, into the pBluescript KS (pBKS), generating the pBKS-Hnf4α Ex10-HR. Using primers EP_1126 and EP_1127 on a Venus containing DNA template, a Venus-RGS-HisTag fragment (753 bp) was amplified and gel purified after *Xba*I and *Spe*I digestion and subcloned between 5′ and 3′ HRs, resulting in pBKS-Hnf4α Ex10-HR-Venus. The PGK promoter-driven neomycin resistance gene flanked by FRT sites (FRT-Neo-FRT) was released by *Bam*HI and *Eco*RI digestion from the PL451-loxP [7] and cloned into these sites in the 3′ upstream sequence of Venus, resulting in pBKS-Hnf4α Ex10HR-Venus-Neo (Hnf4α-Venus mini-targeting vector). In order to generate a Hnf4α-Venus knock-in targeting vector with a large (more than 5 kbp) homology region, we prepared a retrieval plasmid containing a 5′ HR (EP_1178 and EP_1179) and a 3′ HR (EP_1180 and EP_1181) by cloning into *Hind*III/*Asc*I and *Hind*III/*Spe*I sites, respectively, then into PL254 [8]. The resulting vector was linearized (5949 bp) with *Hind*III and *NdeI*, and electroporated into EL350 bacteria containing Hnf4α BAC, retrieving the wild type sequence between PCRs via bacterial homologous recombination (Hnf4α retrieval vector). Subsequently, the mini-targeting cassette (3165 bp) was released by digestion with *Not*I/*Xho*I and introduced into pL254-Hnf4α via bacterial homologous recombination in EL350 bacteria containing the retrieval vector. The final targeting construct (pL254-Hnf4α-Venus targeting vector; Fig. 1A) was confirmed by sequencing.

For the TagRFP-Tcf7L2 reporter, we cloned TagRFP into the *Tcf7L2* start codon in order to create an N-terminally TagRFP-tagged Tcf7L2 fusion protein. The 5′ HR (EP_1210 and EP_1211) and 3′ HR (EP_1212 and EP_1213) of Exon1 were amplified by PCR using BAC (RP23-289C7) as a template and subcloned via *Not*I/*Bam*HI and *Hind*III/*Sal*I, respectively into pBKS, generating pBKS-Ex1Tcf7L2-HR. The TagRFP sequence without a stop codon was amplified from plasmid DNA (Evrogen) using EP_1214 and EP_1215 primers and cloned between 5′ and 3′ HR into *Eco*RI and *Hind*III sites, resulting in pBKS-Ex1 Tcf7L2-HR-TagRFP. PGK promoter-driven FRT-Neo-FRT was gel purified as a 1889 bp fragment after *Bam*HI and *Eco*RI digestion of the PL451-loxP and ligated into the 5′ upstream sequence of TagRFP, resulting in pBKS-Ex1 Tcf7L2-HR-TagRFP-Neo. The retrieval vector for TagRFP-Tcf7L2 was prepared as follows: the retrieval 5′ HR (EP_1220 and EP_1221) and 3′ HR (EP_1222 and EP_1223) were amplified using Tcf7L2 BAC as a template. Products were subsequently cloned into *Asc*I/*Spe*I and *Bam*HI/*Spe*I sites of the PL254 vector [8]. The resulting vector was linearized with *Nde*I and *Spe*I (5895 bp) and electroporated into EL350 bacteria containing Tcf7L2 BAC, retrieving the wild type sequence between PCRs via bacterial homologous recombination (TcfL2 retrieval vector). Subsequently, the Tcf7L2 mini-targeting cassette was

released by digestion with *Not*I and *Kpn*I, and a 3256 bp fragment was introduced into pL254-Tcf7L2 via bacterial homologous recombination in EL350 bacteria containing the retrieval vector, resulting in the final targeting construct (pL254-TagRFP-Tcf7L2 targeting vector; Fig. 1B), which was confirmed by sequencing.
For cloning of the Ngn3-Venus knock-in cassette into Exon 2 of *Ngn3*, the 5′ (EP_ 1121 and EP_1122) and 3′ HR (EP_1123 and EP_ 1124) were amplified by PCR using BAC (RP23-121F10) as a template. PCR fragments were subcloned via *Not*I/*Xba*I and *Hind*III/*Xho*I, respectively, into pBKS, generating the pBKS-Ngn3 Ex2-HR. The Venus sequence was amplified from a Venus-containing DNA template (EP_1126 and EP_1127) and cloned between the homology region of pBKS-Ex2 Ngn3-HR via ligation into *Xba*I and *Spe*I sites, resulting in pBKS-Ex2 Ngn3-HR-Venus. PGK promoter-driven FRT-Neo-FRT was cloned as the 1851 bp fragment after *Sma*I and *Hind*III digestion of the PL451-loxP vector [7] into the 3′ upstream sequence of Venus, resulting in pBKS-Ex2 Ngn3-HR-Venus-Neo (Fig. 1C).
Specific guide RNA sequences for target regions were selected according to http://crispr.mit.edu/ [9]. While the single strand breaks induced on opposite genomic strands were less then 100 bp apart, nicks on the targeting vectors were separated by about a 3 kb DNA sequence (by cassette-containing elements, e.g. pA, Neo, PGK) and therefore did not lead to DNA DSBs. To generate CRISPR expression vectors targeting specific genomic loci, a 20 bp sequence located 5′ of the PAM sequence was cloned into pBS-U6-chimericRNA. A pair of oligos was annealed by heating to 95°C and then cooled down at room temperature. A self-annealing oligo duplex generating *Bbs*I overhangs was cloned into the *Bbs*I-digested vector. Successful integration of CRISPRs into the pbs-U6-chimericRNA vector was confirmed by sequencing.

Cell culture and homologous recombination in mES cells

Mouse ES cells were cultured on a murine embryonic feeder (MEF) layer in Dulbecco′s Modified Eagle Medium (DMEM, Invitrogen) with high glucose (4500 mg/l) containing 15% fetal calf serum (FCS, PAN), 2 mM L-glutamine (from 200 mM stock; Invitrogen), 1 x MEM nonessential amino acids (from 100 x stock; Invitrogen), 0.1 mM β-mercaptoethanol (from 50 mM stock; Invitrogen) and 1500 U/ml LIF (leukemia inhibitory factor, Chemicon, 107 U/ml). Cells were split every two days using trypsin (0.05% Trypsin, 0.53 mM EDTA; Invitrogen) as described previously [10]. F1 hybrid (129Sv/C57Bl/6) mES cells (IDG3.2; [11]) were electroporated using a Bio Rad Gene Pulser Xcell in 4 mm cuvettes at 250 V, 500 µF capacitance and ∞ ohms resistance. Prior electroporation mini-targeting vectors were linearized with *Not*I and targeting vectors were linearized with *Asc*I and were mixed with pbs-U6-chimericRNA and vector E-235 pCAG-Cas9v2D10A-bpA encoding nickase Cas9. For electroporation, 9 µg of Cas9n vector, 3 µg of each gRNA and 20 µg of targeting vector were used. Neo-resistant clones were selected at a final concentration of 300 µg/ml Geneticin (G418 sulfate; Invitrogen/Gibco). Targeting efficiency was assessed by Southern blot.

PCR genotyping

Correct integration of mini-targeting constructs was verified by PCR using genomic DNA of expanded mES cell clones as a template to amplify sequences around the gRNA target sites using the primers (Suppl. Table 1; primers for genotyping).

Southern blot analysis

Genomic DNA was separated on 0.7% (v/w) agarose gel after restriction digests with the appropriate enzymes, transferred to a positively charged nitrocellulose membrane (Roche) and hybridized with ^{32}P-labeled probes. The 3′ Southern probe for *Hnf4α* was PCR amplified (501 bp) using EP_1176 and EP_1177 primers from the described BAC. In the Southern hybridization assay of *Eco*RI-digested genomic DNA, the Hnf4α 3′ probe recognized a shift from 6881 bp (wt) to 6103 bp for the targeted allele. The *Tcf7L2* 5′ Southern probe (387 bp) was PCR amplified using EP_1216 and EP_1217 primers from the described BAC and in the Southern hybridization assay of *Hind*III-digested genomic DNA it recognized a shift from 9040 bp (wt) to 8219 bp for the targeted allele. The *Ngn3* 3′ Southern probe (603 bp) was PCR amplified using EP_1107 and EP_1108 primers from the described BAC and in the Southern hybridization assay of *Hind*III-digested genomic DNA it recognized a shift from 6937 (wt) to 4496 bp for the targeted allele.

On-target and off-target analysis

Mouse ES cells were digested overnight in lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% (w/v) sarkosyl and 1 mg/ml proteinase K). Genomic DNA was extracted by NaCl/ethanol precipitation. For on-target analysis, the genomic sequences of several heterozygous clones for each targeting experiment were PCR amplified using primers flanking the target regions (Suppl. Table 1; primers for genotyping) and served for sequencing. Both wild-type and mutant (carrying either Venus or TagRFP) alleles were checked using primers as follows: primers used for sequencing of Hnf4α-Venus clones: wt allele – EP_1424, Venus allele – EP_1358; primers used for sequencing of TagRFP-Tcf7L2 clones: wt allele – EP_1431, TagRFP allele – EP_1430; primers used for sequencing of Ngn3-Venus clones: wt allele – EP_1373, Venus allele: EP_1429.

Potential off-target sites for selected *Hnf4α*, *Tcf7L2* and *Ngn3* gRNAs were predicted by an online tool (available at http://crispr.mit.edu/guide/). Two putative off-target sites per each gRNA were PCR amplified from the genomic DNA of 3–5 individual gRNA-injected mES clones and analyzed by direct sequencing. The sequences were aligned with the wild-type sequence to detect if any indels were created.

RESULTS AND DISCUSSION

In the past, much effort has been devoted to generating targeting vectors carrying long homology arms of up to 10 kb in length using bacterial artificial chromosome (BAC) recombineering (see [7] for a review). Because manipulations of large DNA fragments are tedious and involve several cloning steps, target construct generation is cumbersome and time-consuming. With the implementation of the CRISPR/Cas9 system, targeting efficiency is theoretically boosted due to the Cas9-induced DSB at the targeted locus and cellular repair mechanisms that use a targeting vector as a template for repair. This encouraged us to test gene-targeting efficiencies using vectors with short homology arms (~ 0.3 kb) in combination with CRISPR/Cas9 technology.

For this purpose, we generated both: i) mini-targeting constructs containing homology arms around 0.3 kb and ii) targeting constructs with homology arms of 5-7 kb in length using standard cloning techniques and bacterial recombination in order to generate Hnf4α-Venus and TagRFP-Tcf7L2 fusion constructs (Fig. 1A and B). The targeting strategy (see Materials and methods) was designed in such a way as to generate a C-terminal Hnf4α-Venus fusion protein or an N-terminal TagRFP-Tcf7L2 fusion protein. The TagRFP fluorescent reporter is an improved monomeric bright red fluorescent protein (RFP) [12]. For selection in mES cells, both targeting vectors contained a phospho-glycerate kinase (PGK) promoter-driven Neomycin (Neo) resistance gene cassette flanked by two flippase recognition target (FRT) and *lox*P sites, respectively. We decided to use the Cas9/double nicking approach with paired gRNAs that minimizes off-target cleavage [6]. The mini-targeting or targeting constructs for each locus were co-electroporated with a plasmid driving Cas9 nickase expression under the control of the CAG promoter and plasmids with two appropriate gRNAs driven by the human PolIII promoter U6 into F1 hybrid (129Sv/C57Bl/6) mES cells (IGD 3.2; [11]). We picked and expanded 248 Neo-resistant mES clones (149 with the mini-targeting and 99 with the targeting constructs) for the Hnf4α-Venus targeting and 177 Neo-resistant mES clones (58 with the mini-targeting and 119 with the targeting constructs) for the TagRFP-Tcf7L2 strategy (Table 1). Homologous recombinants obtained by using mini-targeting vectors were confirmed by both PCR genotyping and Southern blotting (Fig. 1A and B and data not shown). The primary screen was performed by PCR using a primer outside the arm of homology in combination with a primer specific for the cassette to amplify the mutant allele (PCR A; Fig. 1A and B).

Using this PCR strategy specific for the knock-in allele, 26% of Hnf4α-Venus clones and 50% of TagRFP-Tcf7L2 clones were identified as recombinants at the target locus. Strikingly, Southern blotting revealed that out of 149 clones, 7 were homozygous and 34 heterozygous Hnf4α-Venus recombinants (27.5%), whereas for the TagRFP-Tcf7L2 construct, out of 58 clones, 11 were heterozygous (18.9%) (Table 1). However, Southern hybridization revealed that

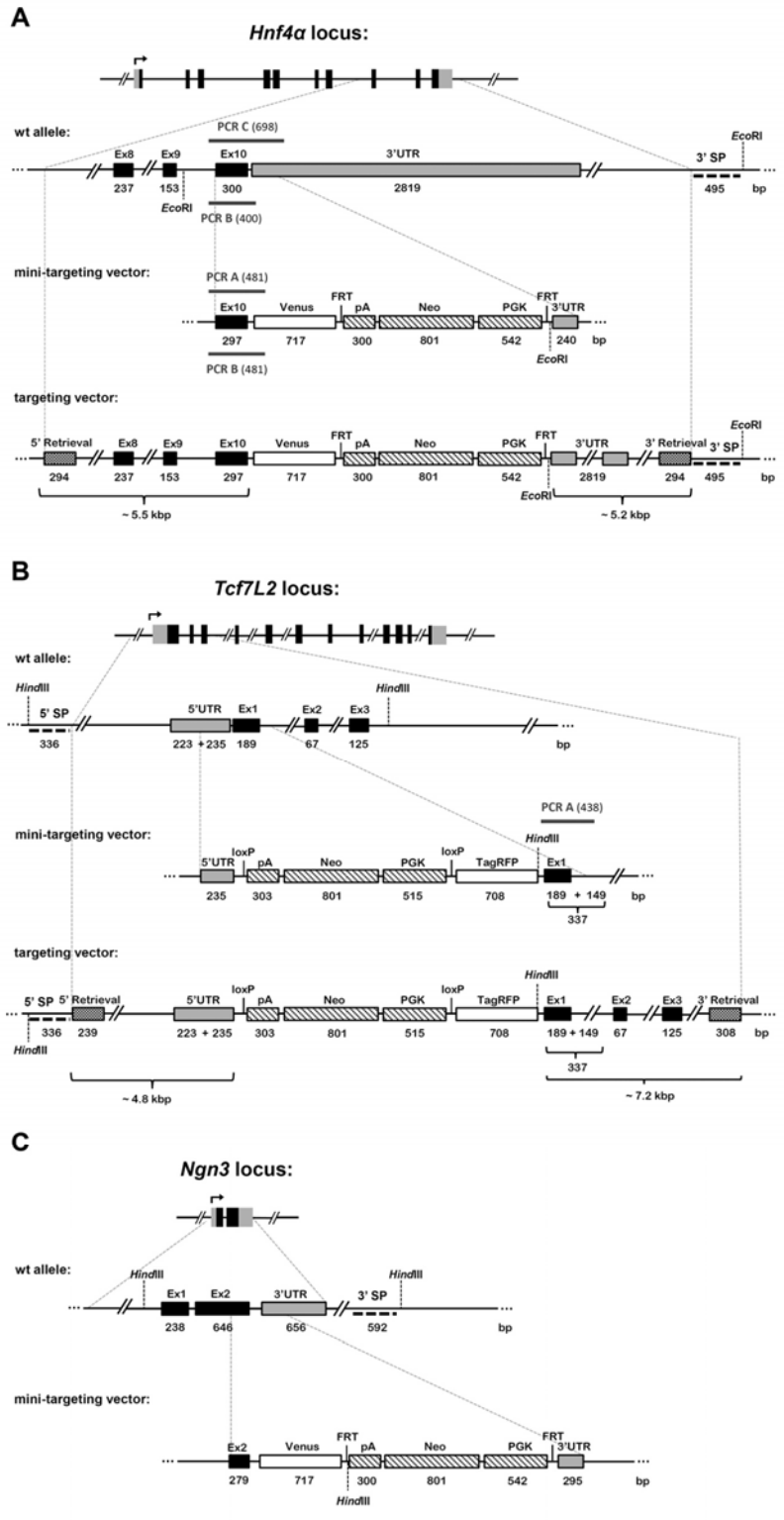
A
Hnf4α locus:
wt allele:
PCR C (698)
Ex8
Ex9
Ex10
3'UTR
3' SP
EcoRI
237
153
300
2819
495
bp
EcoRI
PCR B (400)
mini-targeting vector:
PCR A (481)
Ex10
Venus
FRT
pA
Neo
PGK
FRT
3'UTR
297
717
300
801
542
240
bp
EcoRI
PCR B (481)
targeting vector:
5' Retrieval
Ex8
Ex9
Ex10
Venus
FRT
pA
Neo
PGK
FRT
3'UTR
3' Retrieval
3' SP
EcoRI
294
237
153
297
717
300
801
542
2819
294
495
bp
~ 5.5 kbp
EcoRI
~ 5.2 kbp
B
Tcf7L2 locus:
wt allele:
HindIII
5' SP
5'UTR
Ex1
Ex2
Ex3
HindIII
336
223 + 235
189
67
125
bp
PCR A (438)
mini-targeting vector:
HindIII
loxP
5'UTR
pA
Neo
PGK
loxP
TagRFP
Ex1
235
303
801
515
708
189 + 149
bp
337
targeting vector:
HindIII
5' SP
5' Retrieval
loxP
5'UTR
pA
Neo
PGK
loxP
TagRFP
Ex1
Ex2
Ex3
3' Retrieval
336
239
223 + 235
303
801
515
708
189 + 149
67
125
308
bp
HindIII
337
~ 4.8 kbp
~ 7.2 kbp
C
Ngn3 locus:
wt allele:
HindIII
Ex1
Ex2
3'UTR
3' SP
HindIII
238
646
656
592
bp
mini-targeting vector:
Ex2
Venus
FRT
pA
Neo
PGK
FRT
3'UTR
279
717
300
801
542
295
bp
HindIII

Fig. 1. Targeting and mini-targeting vectors design and detection of targeting. A – Targeting vector and mini-targeting vector design for Hnf4α-Venus. B – Targeting vector and mini-targeting vector design for TagRFP-Tcf7L2. C – Mini-targeting vector design for Ngn3-Venus. For each experiment, wild-type allele with all exons is shown at top, the targeted allele is below, followed by the large targeting vector and the mini-targeting vector. DNA probes for Southern blot analysis (5′ probe for TagRFP-Tcf7L2 and 3′ probe for Hnf4α-Venus and Ngn3-Venus; dashed line in black) which are in the gene to be targeted, but external to the targeting vector, are selected, and also diagnostic restriction sites are depicted. The detection strategy is based on size distinction of probe hybridized fragments of the shorter wild-type allele and the larger targeted allele upon digestion by the corresponding restriction enzyme (see Materials and methods). PCR products and their size in bp using primers to verify the genotype at the *Hnf4α* and *Tcf7L2* loci are shown in dark grey. Translated exons (black boxes), UTRs (light grey boxes), Venus (white boxes), TagRFP (white boxes), and in downward diagonal boxes, the following elements are depicted: promoter regions and Neomycin (Neo) resistance cassette flanked by FRT or loxP sites and SV40 polyadenylation sequence. Homology regions to generate the targeting construct are indicated as 5′ and 3′ retrieval (black dotted boxes). Abbreviations used: base pair, (bp); exon (Ex); flippase recognition target (FRT); phospho-glycerate kinase (PGK); red fluorescent protein (RFP); Southern probe (SP); SV40 polyadenylation sequence (pA); untranslated region (UTR).

Table 1. Summary of targeting efficiency of Hnf4α, Tcf7L2 and Ngn3 loci based on Southern blot analysis.

Locus	Homology arms (kb)		Number of clones					
	5′	3′	hom	het	other	wt	total	% HDR*
Hnf4α	0.3	0.24	7	34	16	92	149	27.5
	5.5	5.2	0	1	0	98	99	1.0
Tcf7L2	0.24	0.34	0	11	5	42	58	18.9
	4.8	7.2	0	26	0	93	119	21.8
Ngn3	0.28	0.3	4	34	35	59	132	28.8

* Percentage of correctly targeted clones (hom – homozygous; het – heterozygous) out of the total mES cell clones.

some clones showed multiple bands (Fig. 2; an example for Hnf4α clones is shown). This was most likely due to insertion of the targeting vector via a non-homologous mechanism at the targeting locus that resulted in genomic alterations, such as concatemer formation, duplications or inversions. Whether linearization of targeting vectors could have caused those alternations was not further analyzed. In Table 1 and Figure 2 these clones represent the category called “other”. These clones could only be detected by Southern blot analysis and were only observed with the mini-targeting vectors, but not with the conventional targeting vectors, illustrating that larger homology arms increase the fidelity of homologous recombination. The number of recombinants identified by PCR and Southern blot analysis correlated better for mES clones obtained with the Hnf4α-Venus mini-targeting vector, though not perfectly.

This difference was likely caused by inability to amplify the mutant allele by PCR in several clones due to the presence of indels. Therefore, we wondered if we can distinguish wild type clones and recombinants with the cassette integrated in a homozygous and heterozygous manner by multiple PCR with three primers recognizing either the 5´ region of the transgene or the wild-type sequence (PCR B; Fig. 1A). This PCR strategy based on the different size of transgenic (481 bp) and non-transgenic (400 bp) products revealed that out of 22 analyzed clones (previously identified by a PCR specific for the knock-in allele) only 4 (18%) were identified as heterozygous mutants and 18 (82%) clones were identified as homozygous mutants (data not shown). This was in contrast to our Southern blot analysis where only 12.3% of homozygous clones (7 out of 57) were identified. To clarify this discrepancy we sequenced all 22 clones that were identified by PCR as homozygous or heterozygous targeted.

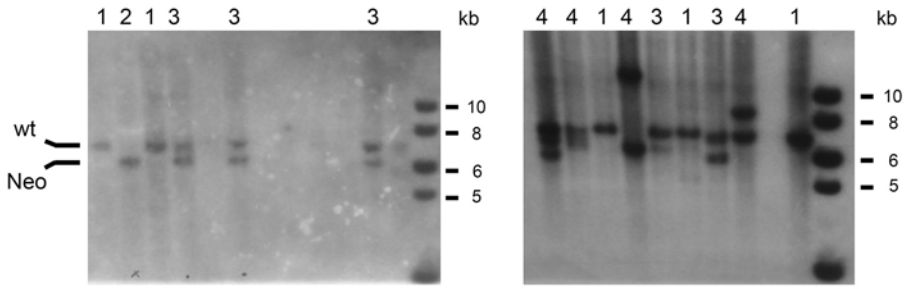

Fig. 2. Southern blot analysis of the mES cell clones electroporated by mini-targeting Hnf4α-Venus construct. Genomic DNA was digested with *Eco*RI and probed with the Hnf4α 3´ radiolabeled probe (see Material and methods). The probe was used to recognize the wild-type allele (wt; 6881 bp) and targeted allele (Neo; 6103 bp). mES cell clones were categorized into four groups: wild type (1), clones with a homozygously (2) or heterozygously (3) integrated cassette, and clones with the cassette integrated in an undesired manner (4), in the text also referred to as "others". A representative part of two membranes of the same experiment is shown to demonstrate all types of clones.

This revealed that indels caused by Cas9-mediated DSBs precluded binding of primers to the wt sequence in those clones that were identified as heterozygous by Southern blot but were homozygous by PCR (Fig. 3A). Thus, the design of a PCR primer overlying the gRNA target sequence is an efficient way to screen for indels on the wild-type allele. We further tested this hypothesis and designed a new primer that is derived from a site more downstream from the gRNA #24 binding site (Fig. 3A). Using this strategy (PCR C; Fig. 1A) we were able to detect the presence of the wild-type allele in Southern blot-identified heterozygous clones (data not shown). On the other hand, small deletions were not detected by Southern hybridization assay, but were found by direct sequencing. This emphasizes the necessity to verify the gene targeting by Southern blot, sequencing and PCR analysis. The targeting efficiencies using

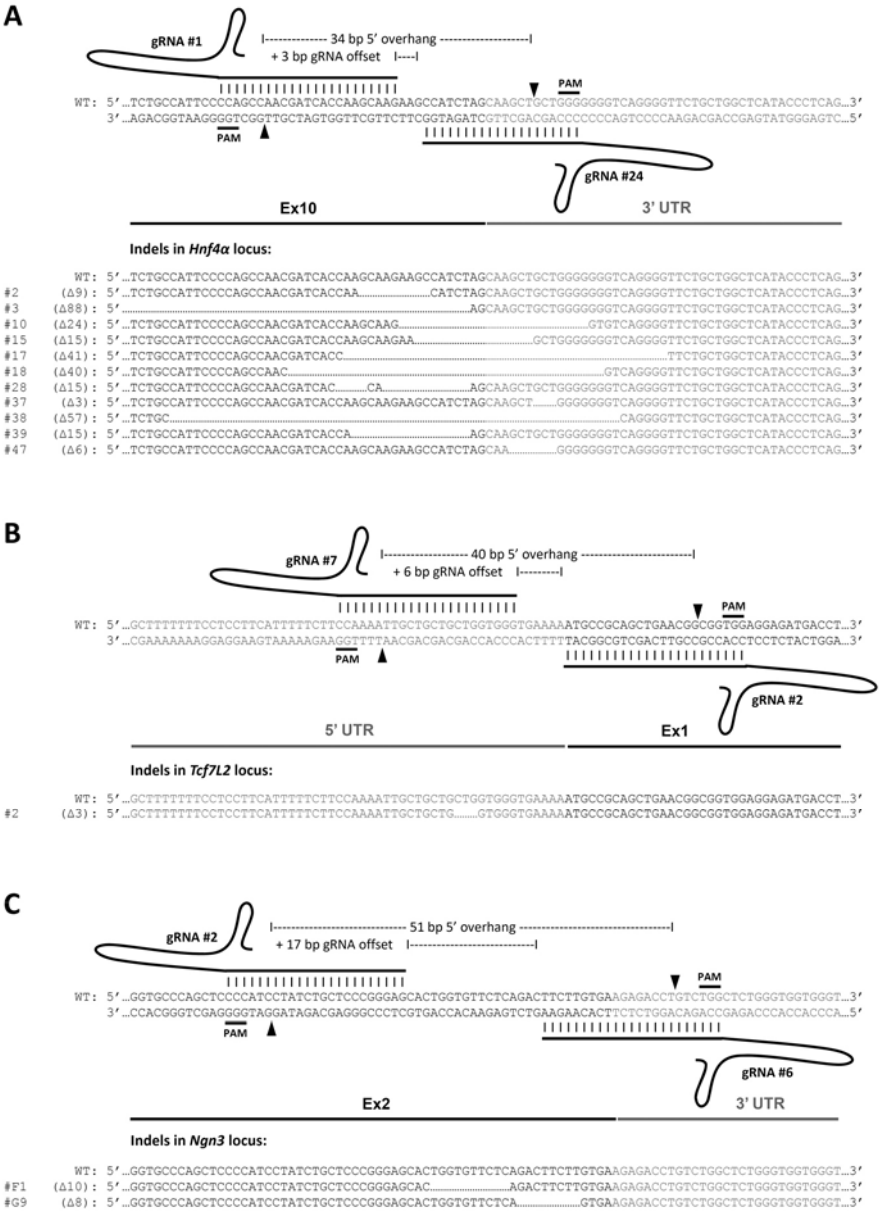

Fig. 3. On-target analysis. A–C: A schematic illustration showing the location of the gRNAs and their PAM sequences along the mouse *Hnf4α*, *Tcf7L2* and *Ngn3*, respectively, locus. Sequences of mutant alleles of selected clones showing representative indels for each locus are listed below. gRNA (black loop) and PAM (black bar) sequences are depicted. Cleavage site is depicted by black arrow.

mini-targeting (M) versus targeting vectors with long homology arms (L) were comparable at the *Tcf7L2* locus (M: 18.9%; L: 21.8%) or even better at the *Hnf4α* locus (M: 27.8%; L: 1%), as revealed by Southern blot analysis (Table 1). Remarkably, homozygous clones were only obtained with the mini-targeting constructs, likely due to the higher transfection efficiency of the smaller vectors. Based on these two targeting strategies, it appeared that vectors with short homology arms combined with two guide RNAs provide sufficient specificity for gene targeting in mES cells. To seek further evidence for this idea, we used the same knock-in strategy with mini-targeting vectors for a third target locus. We designed two gRNAs targeting the translational stop codon of Exon 2 of the *Ngn3* locus to generate a C-terminal Ngn3-Venus fusion reporter (Fig. 1C). The strategy yielded 28.8% Ngn3-Venus mES cell clones based on Southern blot analysis (Table 1), confirming that our targeting strategy reveals a high number of recombined clones at independent loci.

Table 2. Off-target analysis.

Locus	gRNA	Sequence [a,b]	PAM	On/off-target locus	# ismatch.	Indel freq.[c]
Hnf4α	#1	CTTGCTTGGTGATCGTTGGC	TGG	chr2:- 169513431	-	-
	OT-1	tgTcCTTGtTGATCGTTGGC	GAG	chr11:-101349427	4	0/4
	OT-2	CTTGCTaGGgcATCGTTGGC	TAG	chr4:-123683819	3	0/4
	#24	CCATCTAGCAAGCTGCTGGG	GGG	chr2:+ 169513454	-	-
	OT-3	CCAcCTAGaAAGCTGCTGGG	GGG	chr9:-45364291	2	0/4
	OT-4	CCATtTAGCAAGCTGCTtGG	TAG	chr9:-95115497	2	0/4
Tcf7L2	#2	ATGCCGCAGCTGAACGGCGG	TGG	chr19:+55816831	-	-
	OT-1	AcGgCGCAGCTGaACGGCGG	AGG	chr10:+86478957	3	0/3
	OT-2	ATGCgGCtGCTGgACGGCGG	CAG	chr3:-108830119	3	0/3
	#7	CCCACCAGCAGCAGCAATTT	TGG	chr19:-55816825	-	-
	OT-3	tCCACCAGCAGCAGCAtTTT	AGG	chr7:+80750910	2	0/3
	OT-4	CCCACtAGCAGCAGCAATcT	GGG	chr11:+107648519	2	0/3
Ngn3	#2	CTCCCGGGAGCAGATAGGAT	GGG	chr10:-61596829	-	-
	OT-1	CTCttGGaAGCAGATAGGAT	GGG	chr10:+37291122	3	0/5
	OT-2	CaCaCaGGAGCAGATAGGAT	CAG	chr11:+58590446	3	0/5
	#6	TTCTTGTGAAGAGACCTGTC	TGG	chr10:+61596846	-	-
	OT-3	TTCTTGacAAGAGACCTGTC	AGG	chr9:-97907955	2	0/5
	OT-4	aTCTTGTcAAGAGACCTGTa	AGG	chr18:-61923975	3	0/5

[a] The CRISPR target sequence is provided for each gRNA. [b] Potential off-target (OT) sequences are listed with mismatches to the target sequence (lowercase grey letters). The localization and number of mismatches are also indicated. [c] The frequency of off-target indel formation is indicated as the number of independent events out of the number of individual clones analyzed.

To detect NHEJ-induced indel mutations, we analyzed the wild-type allele of selected heterozygous clones for each targeting experiment by direct sequencing. Eleven out of 22 Hnf4α-Venus clones showed small deletions (Fig. 3A), and in four clones large deletions ranging from 104 to 211 bp at the target site were detected (not shown). In the case of TagRFP-Tcf7L2, in two out of six clones indels were observed (Fig. 3B) and two out of eleven Ngn3-Venus clones showed small indels (Fig. 3C). Another clone exhibited a deletion of 271 bp at the target site (data not shown). The ends of the DSBs in different deletion alleles vary in length because they are repaired via the NHEJ pathway. The sequence of the reporter (Venus or TagRFP) allele of the selected clones was also verified and no mutations were detected (data not shown). Finally, we assessed the specificity of all gRNAs used in this work. Two putative off-target (OT) sites per gRNA were computationally predicted (see Materials and methods), PCR amplified and sequenced using genomic DNA of three to five individual clones (Table 2). No indels were detected at any OT site. Our results are in line with previous findings of the Zhang laboratory and suggest that the Cas9 nickases combined with two gRNAs minimizes off-target effects [6].

In summary, our targeting strategy using mini-targeting vectors possessing ~ 0.3 kb homology arms combined with CRISPR/Cas9 technology at three different loci proved feasible and efficient. Through improvement, the strategy allowed the integration of reporter transgenes also in a homozygous manner that was not achieved with conventional targeting vectors. Additionally, generating heterozygous reporter alleles in which the other allele is either wild type or carrying an indel frameshift mutation will be extremely useful for cell lineage and gene functional analysis. Now several studies on the successful use of the CRISPR/Cas9 system and targeting vectors with very short homology arms have been published in different organisms, e.g. in zebrafish [13] and in nematode [14]. However, to our knowledge, only one study has explored the influence of the length of the homology arms on Cas9-mediated integration of donor DNA into the genome of higher eukaryotes. The study of Li and his co-workers [15] reached similar conclusions, i.e. it was recommended to use homology arms of 0.2-0.4 kb for large DNA insertions and homology arms of 0.05 kb for insertion of small DNA fragments. However, in contrast to our study, the authors investigated only the recombination borders at the targeting site by PCR and sequencing. Based on our results, it is evident that targeting vectors with small homology arms tend to integrate via NHEJ mechanisms causing various rearrangements that can only be detected by Southern blot analysis, and thus PCR genotyping alone is not sufficient. We further demonstrated that a three-primer PCR set-up can provide valuable information about homo- or heterozygosity of the targeted clones. To detect indels on the wild-type allele, further validation by sequencing is necessary but can also be detected by PCR using primers spanning the gRNA target site. Targeting three different loci with similar high efficiencies illustrates the great advantage of combining mini-targeting vectors with the CRISPR/Cas9 gene-editing approach. Our study will

help to design and validate targeting strategies for sophisticated genome engineering, such as the insertion of an epitope, a fluorescent reporter gene or generation of a conditional knock-out allele.

Acknowledgements. The excellent technical assistance of Wenke Barkey and Bianca Vogel is acknowledged for cloning and Southern blot analysis, respectively. pBS-U6-chimeric RNA and pCas9 (Cas9 D10A mutant version) plasmids are kind gifts of Oskar Ortiz (Institute of Developmental Genetics, Helmholtz Zentrum München). The main support was provided by the Helmholtz Alliance ICEMED – Imaging and Curing Environmental Metabolic Diseases and the Helmholtz Society, German Research Foundation and German Centre for Diabetes Research (DZD e.V.) to H.L. and by the Academy of Sciences of the Czech Republic (RVO 68378050), BIOCEV – Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund, and OPVK CZ.1.07/2.3.00/30.0027 to R.S.

REFERENCES

1. Capecchi, M.R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. **Nat. Rev. Genet.** 6 (2005) 507–512.
2. Gaj, T., Gersbach, C.A. and Barbas C.F. 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. **Trends Biotechnol.** 31 (2013) 397–405.
3. Hasty, P., Rivera-Perez, J. and Bradley A. The length of homology required for gene targeting in embryonic stem cells. **Mol. Cell Biol.** 11 (1991) 5586–5591.
4. Deng, C. and Capecchi, M.R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. **Mol. Cell Biol.** 12 (1992) 3365–3371.
5. Shulman, M.J., Nissen, L. and Collins C. Homologous recombination in hybridoma cells: dependence on time and fragment length. **Mol. Cell Biol.** 10 (1990) 4466–4472.
6. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. **Cell** 154 (2013) 1380–1309.
7. Copeland, N.G., Jenkins, N.A. and Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. **Nat. Rev. Genet.** 2 (2001) 769–779.
8. Engert, S., Liao, W.P., Burtscher, I. and Lickert, H. Sox17-2A-iCre: a knock-in mouse line expressing Cre recombinase in endoderm and vascular endothelial cells. **Genesis** 47 (2009) 603–610.
9. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J.,

Marraffini, L.A., Bao, G. and Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. **Nat. Biotechnol.** 31 (2013) 827–832.
10. Uetzmann, L., Burtscher, I. and Lickert, H. A mouse line expressing Foxa2-driven Cre recombinase in node, notochord, floorplate, and endoderm. **Genesis** 10 (2008) 515–522.
11. Hitz, C., Wurst, W. and Kuhn, R. Conditional brain-specific knockdown of MAPK using Cre/loxP regulated RNA interference. **Nucleic Acids Res.** 35 (2007) e90.
12. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. **Nat. Biotechnol.** 22 (2004) 1567–1572.
13. Li, K., Wang, G., Andersen, T., Zhou, P. and Pu, W.T. Optimization of genome engineering approaches with the CRISPR/Cas9 system. **PLoS One** 9 (2014) e105779.
14. Hisano**,** Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T. and Kawahara, A. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. **Sci. Rep.** 5 (2015) 8841.
15. Paix**,** A., Wang, Y., Smith, H.E., Lee, C.Y., Calidas, D., Lu, T., Smith, J., Schmidt, H., Krause, M.W. and Seydoux, G. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. **Genetics** 198 (2014) 1347-1356.