1 **Homology arms of targeting vectors for gene insertions and CRISPR/Cas9 technology:**
2 **size does not matter, quality control of targeted clones does**

3

4 Silvia Petrezsélyová[b], Slavomír Kinský[b], Dominika Fričová[b], Radislav Sedláček[b], Ingo
5 Burtscher[a] and Heiko Lickert[a]

6

7 [a]Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, 85764
8 Neuherberg, Germany

9 [b]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, v.v.i.,
10 Vídeňská 1083, 142 20 Prague 4, Czech Republic

11

12 # Address correspondence to Heiko Lickert (heiko.lickert@helmholtz-muenchen.de) and
13 Silvia Petrezsélyová (silvia.petrezselyova@img.cas.cz)

14

15 Heiko Lickert

16 Institute of Diabetes and Regeneration Research

17 Helmholtz Zentrum München

18 85764 Neuherberg

19 Germany

20 Tel.: +49-89-3187-3867

21 Fax.: +49-89-3187-2060

22

23 Silvia Petrezsélyová

24 Transgenic Models of Diseases & Czech centre for Phenogenomics, BIOCEV

25 Institute of Molecular Genetics of the ASCR, v. v. i.

26 Videnska 1083, Prague 4, CZ-142 20

27 Czech Republic

28 Tel.: +420-241-063-205

29 Fax.: +420-224 310-955

30

31 **ABSTRACT**

32

33 While clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-
34 associated 9 (Cas9) technology has rapidly progressed genome editing, gene knock-in
35 engineering still remains difficult due to construction of large targeting vectors that allow
36 homologous recombination. Here, we demonstrate that the combination of CRISPR/Cas9
37 approach and targeting vectors with short homology arms mediated insertion of large reporter
38 cassettes (~ 2.3 kb) into endogenous genes with similar or greater efficiency than with
39 conventional vectors. On the other hand, we emphasize the necessity of the quality control of
40 recombinant clones either by Southern hybridization assay or long-range PCR. In conclusion,
41 our approach will facilitate generation of sophisticated genetic modifications as endogenous
42 reporters or conditional alleles.

43

44 Keywords: CRISPR/Cas9, genome editing, reporter, targeting vector, homology arms

45

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

54

55 **HIGHLIGHTS** (3-5 bullet points, max. 85 characters per point including spaces)

56

57 • Generation of gene knock-in using the CRISPR/Cas9 approach was simplified

58 • Targeting efficiency with conventional and mini-targeting vectors is comparable

59 • Mini-targeting vectors with Cas9 nickases allow to introduce large DNA fragments

60 • Quality control of targeted clones (PCR and Southern blot-based) is recommended

61 ## 1. Introduction

62 Mario R. Capecchi, Martin J. Evans and Oliver Smithies were awarded the Nobel
63 Prize in Medicine in 2007 for their work in the development of a technique to target and
64 introduce specific gene modifications in mice by the use of embryonic stem cells. Genetic
65 changes are introduced via a process called homologous recombination that occurs between
66 an engineered exogenous DNA template and the genome of the mouse embryonic stem cells
67 (mESCs). Transgenic mESCs are selected and then injected into mouse blastocyst-stage
68 embryos or are aggregated with morula-stage embryos, which are then transferred into
69 pseudopregnant foster mothers. Offspring produced in this fashion are chimeras composed of
70 mES cell-derived and blastocyst-derived cells. Chimeras are then backcrossed to test for
71 germline transmission of the targeted gene. Finally, heterozygous mice are intercrossed to
72 derive homozygous gene-targeted mice. This method has been extensively used for decades
73 and helped to decipher functions of many mammalian genes *in vivo*. However, the use of this
74 technique has been hampered because of several limitations: low frequency of homologous
75 recombination rate in mammalian cells, labour-intensive and time-consuming
76 selection/screening strategies and construction of large targeting vectors with isogenic
77 homology regions up to 10 kb to increase the targeting efficiency Capecchi [1].

78 In recent years, discovery of new genome-editing technologies, such as ZFN (zinc
79 finger nucleases), TALENs (transcription activator-like effector nucleases) and CRISPR (the
80 Clustered Regularly Interspaced Short Palindromic Repeats) associated with Cas9 (CRISPR-
81 associated 9) protein, has revolutionized generation of targeted mutations in many model
82 organisms (see [2] for a review). All these technologies work on the same principle, *i.e.*
83 sequence-guided DNA endonucleases induce DNA double-strand breaks (DSBs) that
84 subsequently stimulate either non-homologous end joining (NHEJ) recombination and/or
85 homology-directed repair (HDR) at targeted loci. The NHEJ recombination is an error-prone

86 event forming small insertions or deletions (indels) and therefore results in a frame-shift
87 mutation. Because the HDR system requires a DNA template to repair a DSB, it is used to
88 introduce a desired mutation or reporter gene. The template molecule is supplied either in the
89 form of single-stranded oligonucleotides or double-stranded vectors. These vectors are
90 typically designed to contain a selection marker that is flanked by genomic DNA fragments
91 called 5'- and 3'- homology arms. Since it has been generally accepted that long homology
92 arms are beneficial for increased gene-targeting efficiency [3-5], targeting vectors have been
93 routinely constructed with long homology arms up to 10 kb.

94 Nowadays, with novel genome-editing technologies in hand, the homologous
95 integration of large DNA fragments has greatly enhanced the ability to generate specific gene
96 modifications. Nevertheless, ZFN and TALEN technologies still require a considerable effort
97 to assembly customized new proteins for each targeting. The RNA-guided genome-editing
98 tool, CRISPR/Cas9, is much simpler to implement. It requires only the Cas9 enzyme and an
99 easily-edited guide RNA (gRNA) comprised of a genome target sequence of 20 nucleotides
100 (nt) followed by a 3-nt photospacer adjacent motif (PAM) sequence: NGG (where "N" is any
101 base). When these components are assembled in the nucleus at a specific locus guided by the
102 23 nt target sequence, Cas9 induces a DSB in the genome 3-4 nt upstream of the NGG site.

103 In this study, we investigated the possibility to facilitate CRISPR/Cas9-based genome
104 editing, particularly at the level of preparation of targeting vectors. We asked whether mESCs
105 will repair a CRISPR/Cas9-induced DSB by homologous integration of a small or large DNA
106 fragment with similar efficiencies. For this purpose, we prepared two types of targeting
107 vectors that served as templates for homologous recombination: i) classical targeting vectors
108 with homology arms of 2 to 5 kb in length and ii) mini-targeting vectors carrying homology
109 arms long only ~ 0.3 kb. Both types of targeting vectors carrying reporter cassettes were
110 combined with a pair of gRNAs and pair of CRISPR nickases (Cas9D10A mutant version)

111 that generate single strand breaks. A DSB and DNA insertion will occur only if two nicks are
112 in close proximity (less than 200 bp) creating 5' overhangs. The 5' single-strand DNA
113 overhangs invade the double-stranded molecule of exogenous targeting construct and thereby
114 it facilitates gene knock-in at a specific locus with minimal off-target activity [6].

115 Here we demonstrate on three different examples that combination of CRISPR/Cas9
116 nickases together with mini-targeting vectors leads to an overall insertion frequency as high
117 as if vectors with large homology arms are used. Moreover, our approach using "mini-
118 homology arms" donor molecules enables integration even larger inserts (in our case,
119 transgene cassettes long ~ 2.3 kb). By Southern blot analysis we demonstrated that Cas9
120 nickases in combination with targeting vectors with short homology arms led to insertion of
121 foreign vector DNA also by non-homologous integrations, thus resulting in additional
122 genomic alternations. These phenomena were not observed at similar extents when using
123 targeting vectors with long homology arms.

124

125 ## 2. Materials and methods

126 ### *2.1. Construction of Hnf4α-Venus, TagRFP-Tcf7L2 and Ngn3-Venus targeting*
127 ### *vectors, and gRNA expressing vectors*

128 Knock-in targeting constructs for *Hnf4α*, *Tcf7L2* and *Ngn3* genes were designed as
129 shown in Figure 1. Primers used in this study to generate large targeting and mini-targeting
130 vectors are listed in Supplemental Table 1.

131 Hnf4α-Venus knock-in cassette was designed to target Exon 10. For mini-targeting
132 vector, 5' homology region (HR) and 3' HR were PCR amplified using C57B16 BAC (RP23-
133 156K9) as template and using primers as follows: EP_1170 and EP_1171 primers for 5' HR

134 and EP_1172 and EP_1173 primers for 3´ HR. 5' HR was subcloned via *Not*I and *Xba*I and
135 3' HR was subcloned via *Hind*III and *Xho*I, into the pBluescript KS (pBKS), generating the
136 pBKS-Hnf4α Ex10-HR. Using primers EP_1126 and EP_1127 on a Venus containing DNA
137 template a Venus-RGS-HisTag fragment (753 bp) was amplified and gel purified after *Xba*I
138 and *Spe*I digestion and subcloned between 5' and 3' HRs, resulting in pBKS-Hnf4α Ex10-
139 HR-Venus. PGK promoter-driven neomycin resistance gene flanked by FRT sites (FRT-Neo-
140 FRT) was released by *Bam*HI and *Eco*RI digestion from the PL451-loxP [7] and cloned into
141 these sites in 3' upstream sequence of the Venus resulting in the pBKS-Hnf4α Ex10HR-
142 Venus-Neo (Hnf4α-Venus mini-targeting vector). In order to generate Hnf4α-Venus knock-in
143 targeting vector with large (more than 5 kbp) homology region, we prepared retrieval plasmid
144 containing 5' HR (EP_1178 and EP_1179) and 3' HR (EP_1180 and EP_1181) by cloning
145 into *Hind*III/*Asc*I and *Hind*III/*Spe*I sites, respectively, into PL254 [8]. Resulting vector was
146 linearized (5949 bp) with *Hind*III and *NdeI*, and electroporated into EL350 bacteria
147 containing Hnf4α BAC retrieving the wild type sequence between PCRs via bacterial
148 homologous recombination (Hnf4α retrieval vector). Subsequently, the mini-targeting
149 cassette (3165 bp) was released by digest with *Not*I/*Xho*I and introduced into the pL254-
150 Hnf4α via bacterial homologous recombination in EL350 bacteria containing retrieval vector.
151 Final targeting construct (pL254-Hnf4α-Venus targeting vector; Fig. 1A) was confirmed by
152 sequencing.

153 For TagRFP-Tcf7L2 reporter, we cloned TagRFP into the *Tcf7L2* start codon in order
154 to create an N-terminally TagRFP-tagged Tcf7L2 fusion protein. 5' HR (EP-1210 and EP-
155 1211) and 3' HR (EP_1212 and EP_1213) of Exon1 were amplified by PCR using BAC
156 (RP23-289C7) as template and subcloned via *Not*I/*Bam*HI and *Hind*III/*Sal*I into pBKS,
157 respectively, generating the pBKS-Ex1Tcf7L2-HR. The TagRFP sequence without stop
158 codon was amplified from plasmid DNA (Evrogen) using EP_1214 and EP_1215 primers

159 and cloned between 5' and 3' HR into *Eco*RI and *Hind*III sites, resulting in pBKS-Ex1
160 Tcf7L2-HR-TagRFP. PGK promoter-driven FRT-Neo-FRT was gel purified as 1889 bp
161 fragment after *Bam*HI and *Eco*RI digestion of the PL451-loxP [7] and ligated into 5'
162 upstream sequence of TagRFP, resulting in the pBKS-Ex1 Tcf7L2-HR-TagRFP-Neo.
163 Retrieval vector for TagRFP-Tcf7L2 was prepared as follows: retrieval 5' HR (EP_1220 and
164 EP_1221) and 3' HR (EP_1222 and EP_1223) were amplified using Tcf7L2 BAC as
165 template. Products were subsequently cloned into *Asc*I/*Spe*I and *Bam*HI/*Spe*I sites of PL254
166 vector [8]. Resulting vector was linearized with *Nde*I and *Spe*I (5895 bp) and electroporated
167 into EL350 bacteria containing Tcf7L2 BAC retrieving the wild type sequence between PCRs
168 via bacterial homologous recombination (TcfL2 retrieval vector). Subsequently, Tcf7L2
169 mini-targeting cassette was released by digestion with *Not*I, *Kpn*I and 3256 bp fragment was
170 introduced into the pL254-Tcf7L2 via bacterial homologous recombination in EL350 bacteria
171 containing retrieval vector, resulting in the final targeting construct (pL254-TagRFP-Tcf7L2
172 targeting vector; Fig. 1B), which was confirmed by sequencing.

173 For cloning of the Ngn3-Venus knock-in cassette into Exon 2 of *Ngn3*, 5' (EP_ 1121
174 and EP_1122) and 3' HR (EP_1123 and EP_ 1124) were amplified by PCR using BAC
175 (RP23-121F10) as template. PCR fragments were subcloned via *Not*I/*Xba*I and *Hind*III/*Xho*I,
176 respectively into pBKS, generating the pBKS-Ngn3 Ex2-HR. The Venus sequence was
177 amplified from a Venus containing DNA template (EP_1126 and EP_1127) and cloned
178 between homology region of pBKS-Ex2 Ngn3-HR via ligation into *Xba*I and *Spe*I sites
179 resulting in pBKS-Ex2 Ngn3-HR-Venus. PGK promoter-driven FRT-Neo-FRT was cloned as
180 the 1851 bp fragment after *Sma*I and *Hind*III digestion of PL451-loxP vector [7] into the 3'
181 upstream sequence of the Venus resulting in the pBKS-Ex2 Ngn3-HR-Venus-Neo (Fig.1C).

182 Specific guide RNA sequences for target regions were selected according
183 http://crispr.mit.edu/ [9]. To generate CRISPR expression vectors targeting specific genomic

184 loci, 20 bp of sequence located 5' of the PAM sequence was cloned into pBS-U6-
185 chimericRNA (a generous gift from O. Ortiz, Institute of Developmental Genetics, Helmholtz
186 Zentrum München). A pair of oligos was annealed by heating on 95 °C and then cooled down
187 at room temperature. Self-annealing oligo duplex generating *Bbs*I overhangs were cloned into
188 the *Bbs*I-digested vector. Successfull integration of CRISPRs into pbs-U6-chimericRNA
189 vector was confirmed by sequencing.

190

191 ### *2.2. Cell culture and homologous recombination in mES cells*

192 Mouse ES cells were cultured as previously described [10]. IDG3.2 ES cells [11]
193 were electroporated either with mini-targeting vectors linearized with *Not*I or targeting
194 vectors linearized with *Asc*I together with pbs-U6-chimericRNA and vector E-235 pCAG-
195 Cas9v2D10A-bpA (a generous gift from O. Ortiz, Institute of Developmental Genetics,
196 Helmholtz Zentrum München) encoding for nickase Cas9. The final concentration in the
197 mixture for electroporation was 9 µg/µl Cas9n vector, 3 µg/µl of each gRNA and 20 µg/µl of
198 the single DNA template. Neo-resistant clones were selected at a final concentration of 300
199 µg/ml geneticin (G418 sulfate; Invitrogen/Gibco). Targeting efficiency was shown by
200 Southern blot.

201

202 ### *2.3. PCR genotyping*

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

206

207 ### *2.4. Southern blot analysis*

208 3' Southern probe for *Hnf4α* was PCR amplified (501 bp) using EP_1176 and
209 EP_1177 primers from described BAC. In Southern hybridization assay of *Eco*RI-digested
210 genomic DNA, the Hnf4α 3' probe recognized a shift from 6881 bp (wt) to 6103 bp for the
211 targeted allele. *Tcf7L2* 5' Southern probe (387 bp) was PCR amplified using EP_1216 and
212 EP_1217 primers from described BAC and in Southern hybridization assay of *Hind*III-
213 digested genomic DNA it recognized a shift from 9040 bp (wt) to 8219 bp for the targeted
214 allele. *Ngn3* 3' Southern probe (603 bp) was PCR amplified using EP_1107 and EP_1108
215 primers from described BAC and in Southern hybridization assay of *Hind*III-digested
216 genomic DNA it recognized a shift from 6937 (wt) to 4496 bp for the targeted allele.

217

218 ### *2.5. On-target and off-target analysis*

219 For on-target analysis, the genomic sequences around the gRNA target site of several
220 heterozygous clones for each targeting experiment were PCR amplified and served for
221 sequencing. Both, wild-type and mutant (carrying either Venus or TagRFP) alleles were
222 checked.

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

228

229 ## 3. Results and discussion

230 In past, much effort has been devoted to generating robust targeting vectors carrying
231 long homology arms of 2 to 5 kb in length using bacterial artificial chromosome (BAC)
232 recombineering method. Because of the manipulation of large DNA fragments, target
233 construct generation is often time-consuming due to several cloning steps. Now with the
234 implementation of the CRISPR/Cas9 system, targeting efficiency (the percentage of desired
235 mutation achieved) is boosted due to induction of a double strand break by Cas9 and cellular
236 repair mechanisms that use a targeting vector as template for repair of the DSB. This fact
237 encouraged us to investigate the possibility of shortening of homology arms of targeting
238 vectors to ~ 0.3 kb in order to test for their gene-targeting efficiency using CRISPR/Cas9
239 technology.

240 For this purpose, we generated both: i) mini-targeting constructs containing homology
241 arms around 0.3 kb and ii) targeting constructs with homology arms of 5-7 kb in length using
242 conventional cloning techniques and bacterial recombination in order to generate Hnf4α-
243 Venus and TagRFP-Tcf7L2 fusion constructs (Fig. 1A and B). The targeting strategy (see
244 Materials and methods) was designed in a way to generate a C-terminal Hnf4α-Venus fusion
245 protein or an N-terminal TagRFP-Tcf7L2 fusion protein. TagRFP sequence encodes for
246 improved monomeric bright red (orange) fluorescent protein of the conventional RFP [12].
247 For selection in mES cells, both targeting vectors contained a phospho-glycerate kinase
248 (PGK) promoter-driven Neomycin (Neo) resistance gene cassette flanked by two flippase
249 recognition target (FRT) and *lox*P sites, respectively. We decided to use the Cas9/double
250 nicking approach with paired gRNAs. Thus the mini-targeting or targeting constructs were
251 co-electroporated with E-235 plasmid carrying Cas9 nickase and plasmids with two
252 appropriate gRNAs for each locus into F1 hybrid (129Sv/C57Bl/6) mouse ES cells (IGD 3.2;
253 [11]) and 248 (149 with the mini-targeting and 99 with the targeting constructs) for Hnf4α-
254 Venus and 177 (58 with the mini-targeting and 119 with the targeting constructs) for

255 TagRFP-Tcf7L2 for neomycin-resistant clones were isolated (Table 1). Homologous
256 recombinants obtained by using mini-targeting vectors were confirmed by both PCR and
257 Southern blotting (3’ probe for Hnf4α-Venus and 5’ probe for TagRFP-Tcf7L2; Fig. 1A and
258 B and data not shown). The primary mESCs screen was performed by PCR using a primer
259 outside the arm of homology in combination with a primer specific for the cassette to amplify
260 the mutant allele. In case of Hnf4α-Venus, 26% of analysed clones were recombinants using
261 EP_1357 and EP_1358 primers and in case of TagRFP-Tcf7L2, 50% of analysed clones were
262 positive when detected by EP_1359 and EP_1360 primers. Strikingly, Southern blotting
263 revealed that, of the 149 clones, 41 were either homozygous or heterozygous Hnf4α-Venus
264 recombinants (27.5%) and for TagRFP-Tcf7L2 construct, of the 58 clones, only 11 showed to
265 have the cassette integrated in the *Tcf7L2* locus (18.9%) (Table 1). Interestingly, Southern
266 hybridization revealed that some clones showed multiple bands (Fig. 2; an example for
267 Hnf4α clones is shown). This was probably due to multiple knock-in allele insertions or
268 insertion of the cassettes into Cas9-mediated DSB via non-homologous mechanism at the
269 desired locus that resulted in genomic alternations, such as duplications or inversions (not
270 further analysed). In table 1 and figure 2 these clones represent the category called “other”.
271 These clones could only be detected by Southern blot. It is of note that this phenomenon was
272 observed only with mini-targeting vectors and not with large targeting vectors.

273 Because PCR genotyping results correlated better with the Southern blot analysis for
274 Hnf4α-Venus obtained with the mini-targeting vector, we asked whether the PCR strategy
275 applied for this case, *i.e.* DNA amplification with EP_1357 (recognises a region outside of
276 the homology arm) and EP_1358 (derived from the cassette specific sequence) primers can
277 be also used to distinguish recombinants with the cassette integrated in homozygous and
278 heterozygous manner simply by including a third primer derived from the wild type sequence
279 (EP_1374) (Fig. 1A). Our three-primer PCR strategy revealed that out of 22 analysed clones

280 (previously identified by two-primer PCR as recombinant clones) only 4 (18%) were
281 identified as heterozygous mutants and 18 (82%) clones were identified as homozygous
282 mutants (data not shown). These data were in contrast to our Southern blot analysis where
283 only 12.3% homozygous clones (7 out of 57) were identified. Sequencing analysis revealed
284 that this discrepancy was caused by a Cas9-mediated indel in the binding site of the wt-
285 specific primer in those clones that were identified as heterozygous by Southern blot but
286 homozygous by PCR (Fig. 3A). This result indicates that design of a PCR primer overlying
287 with gRNA target sequences is an efficient way to screen for unwanted indels on the wild
288 type allele. To verify this hypothesis, we designed a new primer (EP_1424) that is derived
289 from a site more downstream from the gRNA #24 site and indeed, we were able to detect the
290 presence of wild-type allele in Southern blot-identified heterozygous clones (data not shown).
291 On the other hand, small deletions were not detected by Southern hybridization assay, but
292 were found by direct sequencing. This emphasizes the necessity to verify the targeting in
293 combination by Southern blot and PCR analysis.

294 Unlike clones electroporated by mini-targeting vectors, clones obtained using large
295 targeting vectors were verified only by Southern blot analysis and their efficiency of targeting
296 was as follows: while 21.8% HDR for TagRFP-Tcf7L2 was comparable with mini-targeting
297 approach, electroporation of mES cells with Hnf4α-Venus targeting vector led to generation
298 of only one heterozygous clone (Table 1).

299 Based on these two targeting strategies it seems that vectors with short homology
300 arms together with two guide RNAs provide sufficient specificity for gene targeting in
301 mESCs. To seek further evidence for this conclusion, we used the same knock-in strategy
302 with mini-targeting vectors for another target locus. We designed two gRNAs targeting the
303 translational stop codon of Exon 2 of *Ngn3* locus to integrate Venus fluorescent marker at its

304 C-terminus (Fig. 1C). The strategy yielded 28.8% Ngn3-Venus mESC clones based on
305 Southern blot analysis (Table 1).

306 We next analysed the wild-type allele of selected heterozygous clones for each
307 targeting experiment to detect NHEJ-induced indel mutations via direct sequencing. 11 out of
308 22 Hnf4α-Venus clones showed small deletions (Fig. 3A) and in 4 clones large deletions
309 ranging from 104 to 211 bp at the target site were detected (not shown). In case of TagRFP-
310 Tcf7L2, in 2 out of 6 clones indels were observed (Fig. 3B) and 2 out of 11 Ngn3-Venus
311 clones showed small indels (Fig. 3C) and another clone exhibited a deletion of 271 bp long
312 (not shown) at the target site. The ends of the DSBs in different deletion alleles vary in length
313 because they are repaired via NHEJ pathway. The sequence of reporter (Venus or TagRFP)
314 allele of the selected clones was also verified and no DNA mutations were detected (not
315 shown).

316 Finally, we assessed the specificity of all gRNAs used in this work. Two putative off-
317 target (OT) sites per each gRNA were computationally predicted (see Materials and
318 methods), PCR amplified and sequenced using genomic DNA of 3 – 5 individual clones
319 (Table 2). No indels were detected at any locus. Our results are in line with previous findings
320 of the Zhang laboratory and suggest that the Cas9 nickases combined with two gRNAs
321 minimizes off-target effects [6].

322 In summary, our targeting strategy using mini-targeting vectors possessing ~ 0.3 kb
323 long arms combined with CRISPR/Cas9 technology on three different loci proves that it is as
324 efficient (or even more efficient depending on locus) as the strategy with traditional targeting
325 vectors, and importantly, allows insertion of large DNA fragments (2.3 kb). A recent study of
326 Li *et al*. [13] reached similar conclusions, *i.e.* they recommend the use of targeting vectors
327 possessing homology arms of 0.2-0.4 kb in length for larger DNA insertions and for insertion

328 of a small DNA fragment (~ 100 bp) even shorter homology arms (0.05 kb) are sufficient. On
329 contrary to this work, we demonstrate that performing PCR genotyping only is not
330 recommended because “mini-arms” have a higher tendency to be integrated also via non-
331 homologous mechanisms causing inversions, duplications, *etc*. that could be only detected by
332 Southern hybridization assay or long-range PCR. On the other hand, a proper 3-primer PCR
333 set-up might give valuable information about homozygosity/heterozygosity of analysed
334 clones. Our positive experience with insertion of large DNA constructs into endogenous
335 genes using mini-targeting vectors and the CRISPR/Cas9 gene-editing approach will help
336 investigators to design a strategy that will greatly facilitate more sophisticated genome
337 engineering such as insertion of an epitope, a fluorescent reporter marker or generation of a
338 conditional allele.

339

340 **Acknowledgements**

341 The excellent technical assistance of Wenke Barkey and Bianca Vogel is acknowledged for
342 cloning and Southern blot analysis, respectively. pBS-U6-chimericRNA plasmid and E-235
343 plasmid, here named also as pCAG-Cas9v2D10A-bpA, are kind gifts of Oskar Ortiz (Institute
344 of Developmental Genetics, Helmholtz Zentrum München). The abroad research training of
345 D.F., S.K. and S.P. was founded by OPVK CZ.1.07/2.3.00/30.0027 “Founding the Centre of
346 Transgenic technologies” education program. This work was funded (in part) by the
347 Helmholtz Alliance ICEMED – Imaging and Curing Environmental Metabolic Diseases.
348 H.L. thanks the Helmholtz Society, German Research Foundation and German Center for
349 Diabetes Research (DZD e.V.) for financial support.

350

351

352 **References**

353 1. Capecchi, M.R., *Gene targeting in mice: functional analysis of the mammalian*
354 *genome for the twenty-first century.* Nat Rev Genet, 2005. **6**(6): p. 507-12.
355 2. Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, *ZFN, TALEN, and CRISPR/Cas-based*
356 *methods for genome engineering.* Trends Biotechnol, 2013. **31**(7): p. 397-405.
357 3. Hasty, P., J. Rivera-Perez, and A. Bradley, *The length of homology required for gene*
358 *targeting in embryonic stem cells.* Mol Cell Biol, 1991. **11**(11): p. 5586-91.
359 4. Deng, C. and M.R. Capecchi, *Reexamination of gene targeting frequency as a*
360 *function of the extent of homology between the targeting vector and the target locus.*
361 Mol Cell Biol, 1992. **12**(8): p. 3365-71.
362 5. Shulman, M.J., L. Nissen, and C. Collins, *Homologous recombination in hybridoma*
363 *cells: dependence on time and fragment length.* Mol Cell Biol, 1990. **10**(9): p. 4466-
364 72.
365 6. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome*
366 *editing specificity.* Cell, 2013. **154**(6): p. 1380-9.
367 7. Copeland, N.G., N.A. Jenkins, and D.L. Court, *Recombineering: a powerful new tool*
368 *for mouse functional genomics.* Nat Rev Genet, 2001. **2**(10): p. 769-79.
369 8. Engert, S., et al., *Sox17-2A-iCre: a knock-in mouse line expressing Cre recombinase*
370 *in endoderm and vascular endothelial cells.* Genesis, 2009. **47**(9): p. 603-10.
371 9. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat
372 Biotechnol, 2013. **31**(9): p. 827-32.
373 10. Uetzmann, L., I. Burtscher, and H. Lickert, *A mouse line expressing Foxa2-driven*
374 *Cre recombinase in node, notochord, floorplate, and endoderm.* Genesis, 2008.
375 **46**(10): p. 515-22.

376 11. Hitz, C., W. Wurst, and R. Kuhn, *Conditional brain-specific knockdown of MAPK*
377 *using Cre/loxP regulated RNA interference.* Nucleic Acids Res, 2007. **35**(12): p. e90.
378 12. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins*
379 *derived from Discosoma sp. red fluorescent protein.* Nat Biotechnol, 2004. **22**(12): p.
380 1567-72.
381 13. Li, K., et al., *Optimization of genome engineering approaches with the CRISPR/Cas9*
382 *system.* PLoS One, 2014. **9**(8): p. e105779.

383

384 **FIGURE LEGENDS**

385 **Figure 1. Targeting and mini-targeting vectors design and detection of targeting.**
386 Targeting vector and/or mini-targeting vector design for Hnf4α-Venus (A), TagRFP-Tcf7L2
387 (B) and Ngn3-Venus (C) reporter alleles are illustrated. For each experiment, wild-type allele
388 with all exons is shown at top, the part of wild-type allele influenced by targeting on the next
389 line, the large targeting vector on the next line and the mini-targeting vector on the next line.
390 DNA probes for Southern blot analysis are selected (either 5' or 3'; in purple) which are in
391 the gene to be targeted, but external to the targeting vector and also diagnostic restriction sites
392 are depicted (in red). The detection strategy is based on size distinction of probe hybridized
393 fragments of the shorter wild-type allele and the larger targeted allele upon digestion by
394 corresponding restriction enzyme (see Materials and Methods). Also shown are primers for
395 diagnostic PCR analysis (in grey). Translated exons (black boxes), UTRs (blue boxes), Venus
396 (yellow boxes), TagRFP (red boxes), promoter regions (grey boxes) and Neomycin (Neo)
397 resistance cassette (orange boxes) flanked by FRT or loxP sites are indicated. Homology
398 regions to generate the targeting construct are indicated as 5' and 3' retrieval (green boxes).
399 Abbreviations used: base pair, (bp); flippase recognition target, (FRT); exon, (Ex); phospho-

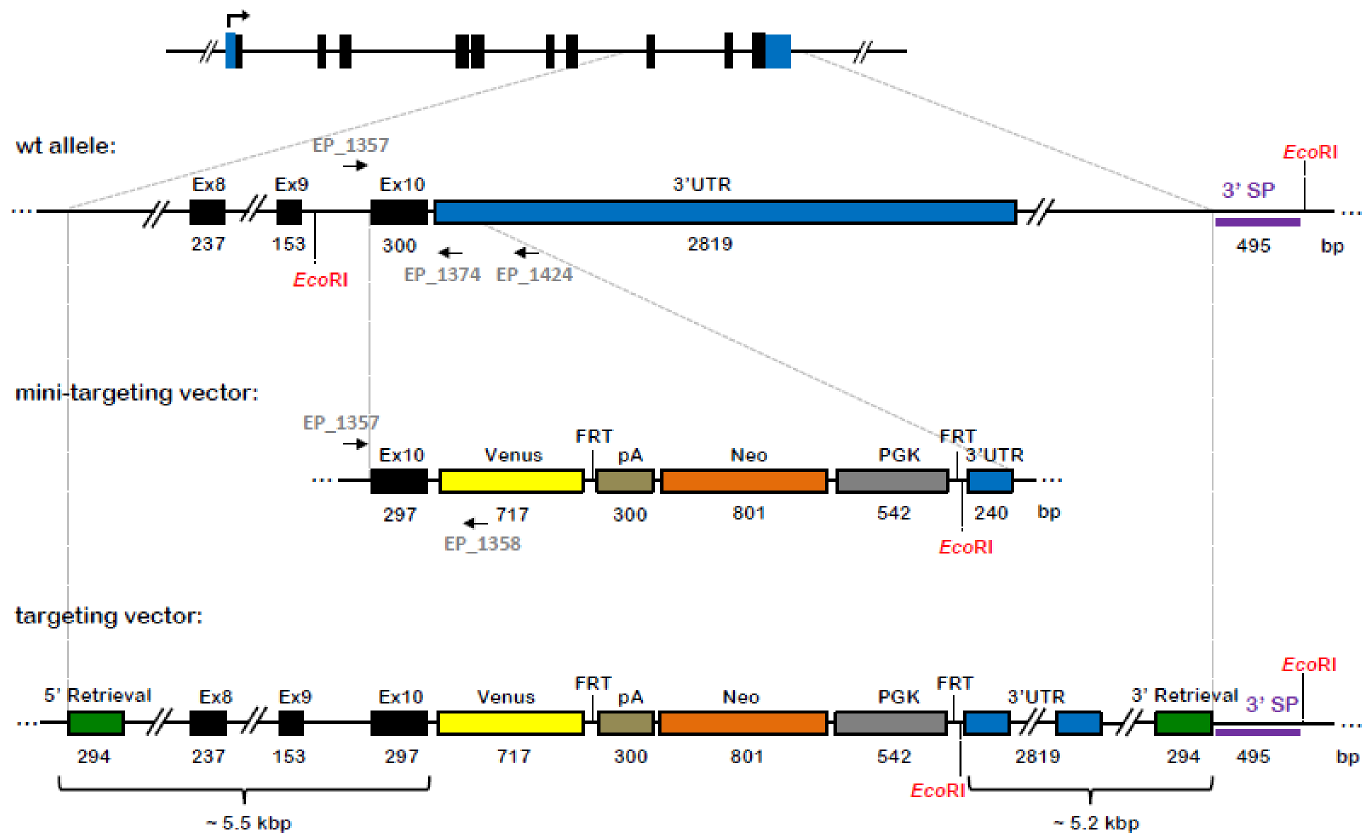
400 glycerate kinase, (PGK); red fluorescent protein, (RFP); SV40 polyadenylation sequence,
401 (pA); untranslated region, (UTR).

402

403 **Figure 2. Southern blot analysis of the mESC clones electroporated by mini-targeting**
404 **Hnf4α-Venus construct.** Genomic DNA was digested with *Eco*RI and probed with the
405 Hnf4α 3' radiolabeled probe (see Material and Methods). The probe was used to recognize wt
406 allele (6881 bp) and targeted allele (6103 bp). mESC clones were categorized into four
407 groups: wild-type (1), clones with homozygously (2) or heterozygously (3) integrated cassette
408 and clones with the cassette integrated in undesired manner (4), in the text referred also as
409 „others".

410

411 **Figure 3. On-target analysis.** A schematic illustration showing the location of the gRNAs
412 and their PAM sequences along the mouse *Hnf4α* (A), *Tcf7L2* (B) and *Ngn3* (C) loci.
413 Sequences of mutant alleles of selected clones showing representative indels for each locus
414 are listed below. gRNA and PAM sequences are indicated by blue and violet bars,
415 respectively. Cleavage site is depicted by red arrow.



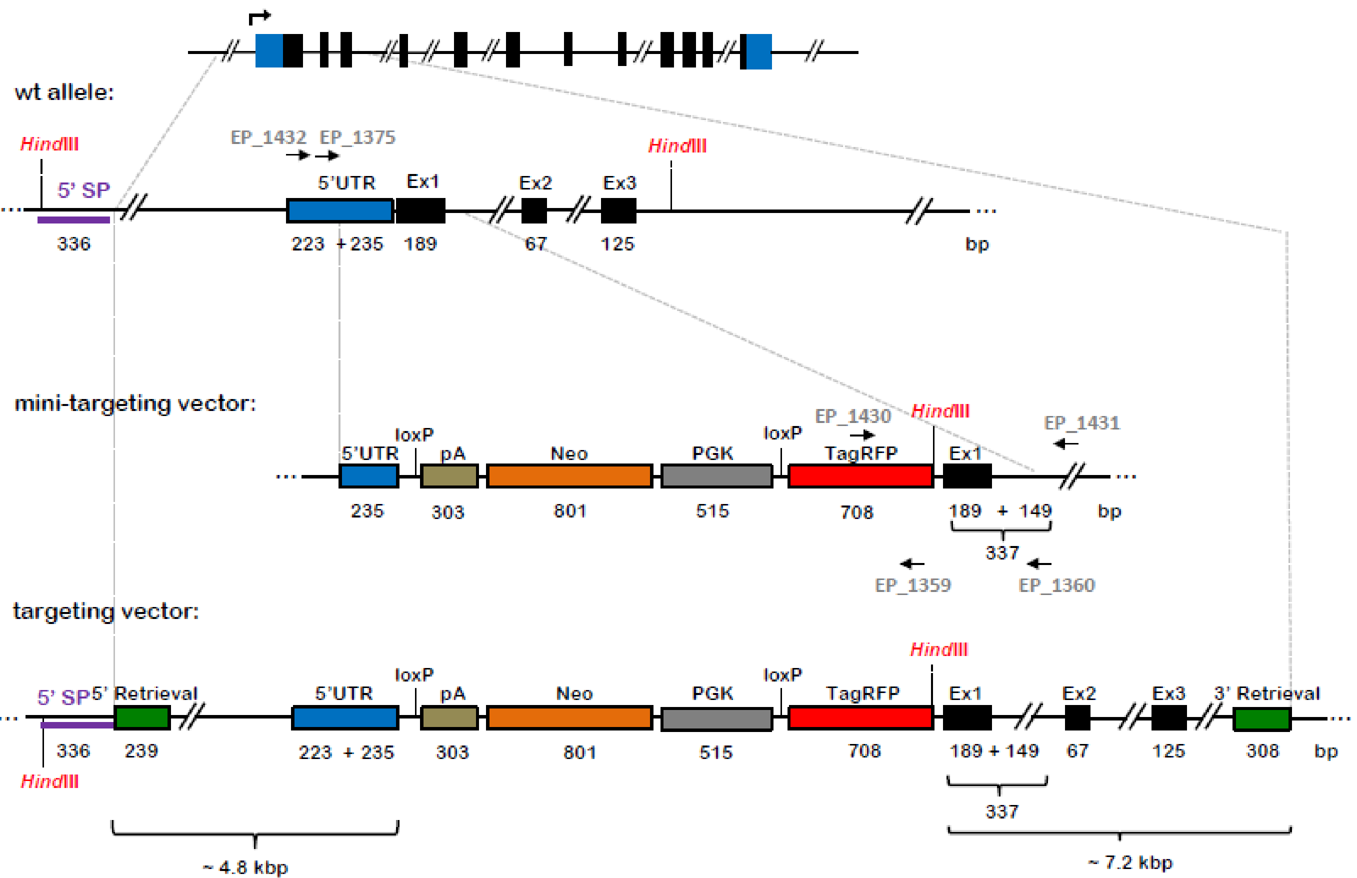


Figure 1

C

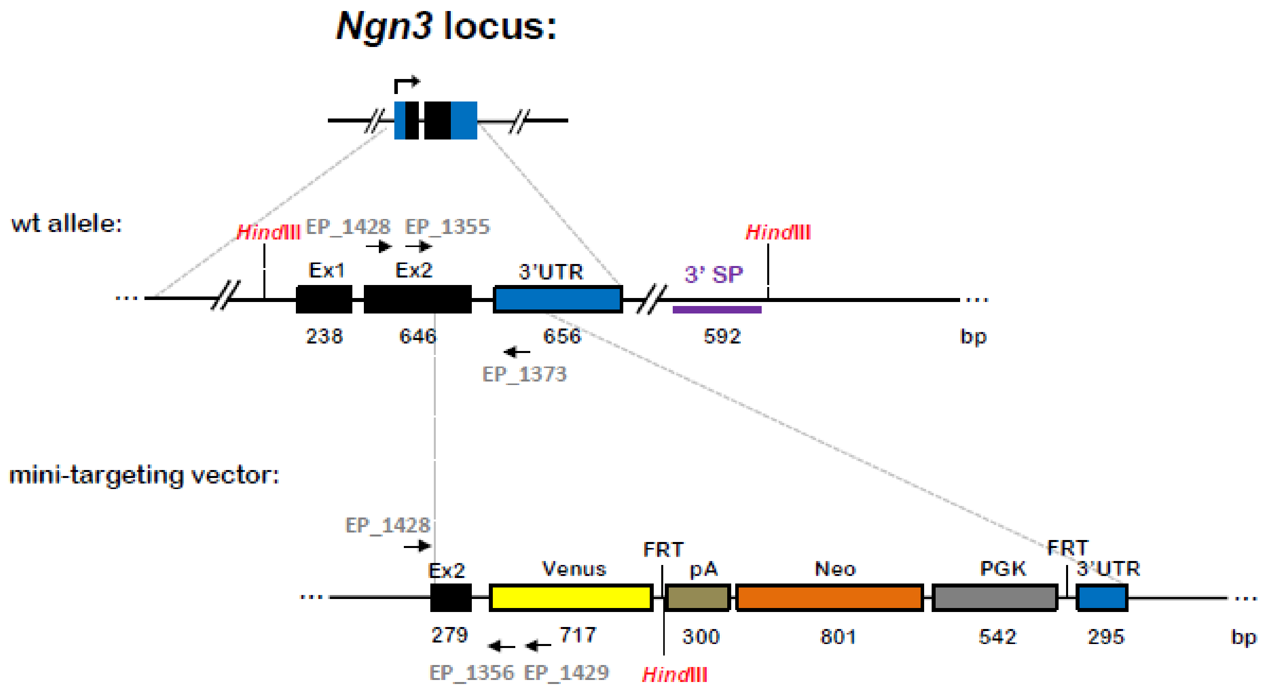


Figure 1 (continued)

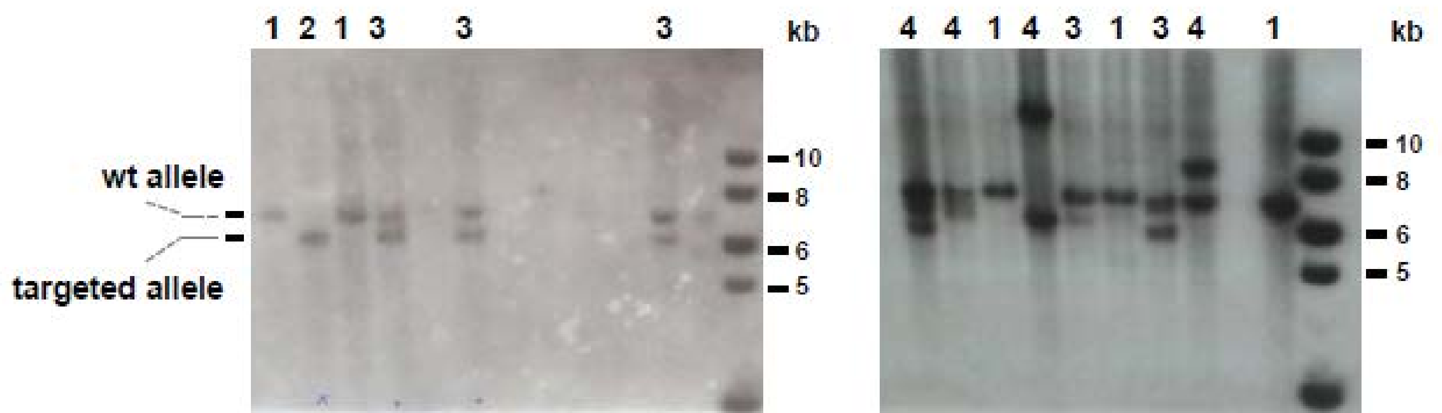


Figure 2

A

gRNA #1 — 34 bp 5′ overhang — + 3 bp gRNA offset

```
WT: 5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAGAAGCCATCTAGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
    3'…AGACGGTAAGGGGTCGGTTGCTAGTGGTTCGTTCTTCGGTAGATCGTTCGACGACCCCCCCAGTCCCCAAGACGACCGAGTATGGGAGTC…5'
```

PAM — PAM — gRNA #24

Ex10 — 3′ UTR

Indels in *Hnf4α* locus:

```
           WT:  5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAGAAGCCATCTAGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#2    (Δ9):   5'…TCTGCCATTCCCCAGCCAACGATCACCAA_________CATCTAGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#3    (Δ88):  5'…______________________________________________AGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#10   (Δ24):  5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAG________________________GTGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#15   (Δ15):  5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAGAA_______________GCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#17   (Δ41):  5'…TCTGCCATTCCCCAGCCAACGATCACC_________________________________________TTCTGCTGGCTCATACCCTCAG…3'
#18   (Δ40):  5'…TCTGCCATTCCCCAGCCAAC________________________________________GTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#28   (Δ15):  5'…TCTGCCATTCCCCAGCCAACGATCAC_____CA__________AGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#37   (Δ3):   5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAGAAGCCATCTAGCAAGCT___GGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#38   (Δ57):  5'…TCTGC_________________________________________________________CAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#39   (Δ15):  5'…TCTGCCATTCCCCAGCCAACGATCACCA_______________AGCAAGCTGCTGGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
#47   (Δ6):   5'…TCTGCCATTCCCCAGCCAACGATCACCAAGCAAGAAGCCATCTAGCAA______GGGGGGGTCAGGGGTTCTGCTGGCTCATACCCTCAG…3'
```

B

gRNA #7 — 40 bp 5′ overhang — + 6 bp gRNA offset

```
WT: 5'…GCTTTTTTTTCCTCCTTCATTTTTCTTCCAAAATTGCTGCTGCTGGTGGGTGAAAAATGCCGCAGCTGAACGGCGGTGGAGGAGATGACCT…3'
    3'…CGAAAAAAAAGGAGGAAGTAAAAAGAAGGTTTTAACGACGACGACCACCCACTTTTTACGGCGTCGACTTGCCGCCACCTCCTCTACTGGA…3'
```

PAM — PAM — gRNA #2

5′ UTR — Ex1

Indels in *Tcf7L2* locus:

```
           WT:  5'…GCTTTTTTTTCCTCCTTCATTTTTCTTCCAAAATTGCTGCTGCTGGTGGGTGAAAAATGCCGCAGCTGAACGGCGGTGGAGGAGATGACCT…3'
#2    (Δ3):   5'…GCTTTTTTTTCCTCCTTCATTTTTCTTCCAAAATTGCTGCTG___GTGGGTGAAAAATGCCGCAGCTGAACGGCGGTGGAGGAGATGACCT…3'
```

C

gRNA #2 — 51 bp 5′ overhang — + 17 bp gRNA offset

```
WT: 5'…GGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGT…3'
    3'…CCACGGGTCGAGGGGTAGGATAGACGAGGGCCCTCGTGACCACAAGAGTCTGAAGAACACTTCTCTGGACAGACCGAGACCCACCACCCA…5'
```

PAM — PAM — gRNA #6

Ex2 — 3′ UTR

Indels in *Ngn3* locus:

```
           WT:  5'…GGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGT…3'
#F1   (Δ10):  5'…GGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCAC__________AGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGT…3'
#G9   (Δ8):   5'…GGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCA________GTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGT…3'
```

Figure 3

Table 1

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

Locus	Homology arms (kbp)		Number of clones					
	5'	3'	homozygous	heterozygous	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	0.01
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

*a percentage of correctly targeted clones (homozygous and heterozygous) out of the total mESC clones.

Table2
Off-target analysis.

Targeted locus	gRNA	Sequence	PAM	on/off-target locus	# mismatches	Indel frequency
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

The CRISPR target sequence is provided for each gRNA. Potential off-target (OT) sequences are listed with mismatches to the target sequence (lowercase red letters). The localisation and number of mismatches are also indicated. The frequency of off-target indel formation is indicated as the number of independent events out of the number of individual clones analysed.

Supplemental Table 1
List of primers used in this study.

Locus	Name (type)	Sequence*	Restriction site
		Primers for cloning strategies	
Hnf4α	EP_1170 (5’ HR; mini-targ.)	NNNGCGGCCGCGTTTGAGCCATAACCACCTTAGGG	*Not*I
	EP_1171 (5’ HR; mini-targ.)	NNNTCTAGAGATGGCTTCTTGCTTGGTGATCG	*Xba*I
	EP_1172 (3’ HR; mini-targ.)	NNNAAGCTTCAAGCTGCTGGGGGGGTCAGG	*Hind*III
	EP_1173 (3’ HR; mini-targ.)	NNNCTCGAGCAGGCAGTGGAAGACAGTCCTCCC	*Xho*I
	EP_1178 (5’ HR; retrieval)	NNNGGCGCGCCGGCACCATTATTGTGGCTCTTCCATCTC	*Asc*I
	EP_1179 (5’ HR; retrieval)	NNNAAGCTTCATATG CAAGGACAGCGCCTCGCGTCCAAC	*Hind*III
	EP_1180 (3’ HR; retrieval)	NNNAAGCTTGTGGAGTCACGGTGTCAAGTACAATG	*Hind*III
	EP_1181 (3’HR; retrieval)	NNNACTAGTGACCTTTGTCACTGTCCAGCATTC	*Spe*I
	EP_1176 (3´ Southern probe)	NNNGAATTCGCGTGAGGCATTCTGGGATTACAGAG	*Eco*RI
	EP_1177 (3´ Southern probe)	NNNCTCGAGACGAAAGCCAAGCTTTCCCTAGCC	*Xho*I
Ngn3	EP_1121 (5’ HR; mini-targ.)	NNNGCGGCCGCGATCGAGACCCTGCGCTTCGCC	*Not*I
	EP_1122 (5’ HR; mini-targ.)	NNNTCTAGACAAGAAGTCTGAGAACACCAGTGCTCCC	*Xba*I
	EP_1123 (3’ HR; mini-targ.)	NNNAAGCTTAGAGACCTGTCTGGCTCTGGGTG	*Hind*III
	EP_1124 (3’ HR; mini-targ.)	NNNCTCGAGAAGTATGGGAGTCTCCAACTAGG	*Xho*I
	EP_1107 (3´ Southern probe)	NNNCTGCAGTAAATGAATTGAGGCTAGGCATCCATC	*Pst*I
	EP_1108 (3´ Southern probe)	NNNGTCGACGTTACCCAGTTAAGCAGCAATAAAGAC	*Sal*I
	EP_1210 (5’ HR; mini-targ.)	NNNGCGGCCGCGAAACCCAAACTCACGCGTGCAGAAG	*Not*I
	EP_1211 (5’ HR; mini-targ.)	NNNGGATCCCACCAGCAGCAGCAATTTTGGAAG	*Bam*HI
	EP_1212 (3’ HR; mini-targ.)	NNNAAGCTTATGCCGCAGCTGAACGGCGGTG	*Hind*III
	EP_1213 (3’ HR; mini-targ.)	NNNGTCGACGGTTACAAGTTTTCGCAGCCAGCAG	*Sal*I
Tcf7L2	EP_1220 (5’ HR; retrieval)	NNNGGCGCGCCAGGTGTCTATTCTTGTGTTGGCTAAGG	*Asc*I
	EP_1221 (5’ HR; retrieval)	NNNACTAGTGATCTGTCGACCACTGAGTTAG	*Spe*I
	EP_1222 (3’ HR; retrieval)	NNNACTAGTCATATGAGAAGTGAGAAGACTGTTGCTCTAC	*Spe*I
	EP_1223 (3’ HR; retrieval)	NNNGGATCCTGAACTGTCTGGGGCCACTTCATC	*Bam*HI
	EP_1216 (5’ Southern probe)	NNNGTCGACGGCGTCTCTTCTCAGAGCACTGG	*Sal*I
	EP_1217 (5’ Southern probe)	NNNGGATCCACAAATGTTTAAAGAAACTTCAAAGAC	*Bam*HI

		Primers used for genotyping
Hnf4α	EP_1357	GGCAATGGGAGTTTCCATGTTAC
	EP_1358	GCTGAACTTGTGGCCGTTTAC
	EP_1374	AGCAGCTTGCTAGATGGCTTC
	EP_1424	GGTGAAGAAGTTGAGGGAAGAAG
Ngn3	EP_1355	CTTCCCGGATGACGCCAAACTTAC
	EP_1356	TCGCCCTTGCTCACCATTCTAGAC
	EP_1373	AGATGCTTGAGAGCCTCCACTACC
	EP_1428	CCAATGATCGGGAGCGCAATCG
	EP_1429	TTCAGGGTCAGCTTGCCGTAGG
Tcf7L2	EP_1359	CGAGGTGGCTGTGGCCAGATAC
	EP_1360	AAGGGAGGGTCGGCGGAGAAAG
	EP_1375	ACGCGTGCAGAAGATCTCCC
	EP_1430	AGATCCAAGAAACCCGCTAAGAAC
	EP_1431	CGGGATTTATCTCGGAAACTTTCG
	EP_1432	AATCAGCGCCGCCTTTGAAC

		Primers for off-target analysis
Hnf4α	EP_1400	ACACTAATCCAGACCCCACTGT
	EP_1401	TCTTTCAAGTTAGCTGGGCAGT
	EP_1402	TTCAACTCCAAATCATCTCTGG
	EP_1403	CAGTAAAAGCCCTTTCTTGCAC
	EP_1404	TACAGCCTAGCAGGTGGAAAAT
	EP_1405	TGAGAGCAAACAAAAGTCGTGT
	EP_1406	CAAGGGCTCTTCTCTGTGTCTT
	EP_1407	TCCCAGGAAGATACTGCTCATT
Ngn3	EP_1408	AGCCCTGATGGCACATATTAGT
	EP_1409	GGATTTTCCAGGACATTCACAT
	EP_1410	GGCTAGCCTCAAACTCATAGACA

	EP_1411	CTCGGAATGTGTTTAGCCCTAC	
	EP_1412	ATCCTTGAAGTTCGTACCCAGA	
	EP_1413	GGTCCCCAGAAAGTTTACAGTG	
	EP_1414	CCCATCCAGGACTAAGATCAAC	
	EP_1415	AAATGGAAATTGGGACATTTTG	
Tcf7L2	EP_1416	TCTCCACTTTGTGAACAGGATG	
	EP_1417	TCAGAACTTTTGTGCCTTGCTA	
	EP_1418	AAGCATCAGATCTGGGGAGTAG	
	EP_1419	GCAGTTGACACTCTGTGGAGTC	
	EP_1420	AGTGGGTGCTCTGTTTGCTATT	
	EP_1421	TCCCTGTTCAAAGACAACTTCC	
	EP_1422	TTCGCTTCTCTAAGGTTTACGG	
	EP_1423	CCCTCACGATGATGAAGAAGA	
		Primers for amplification of cassettes	
Venus	EP_1126	NNNGCGGCCGCAGCCACCATGTCTAGAATGGTGAGCAAGGGCGAGGAGCTGTTC	*Not*I/*Xba*I
	EP_1127	NNNACTAGTTCAGTGATGGTGATGGTGATGCGATCCTCTCTTGTACAGCTCGTCC ATGCCGAGAGTG	*Spe*I
TagRFP	EP_1214 (AU1_tagRFP)	NNNGAATTCGCCACCATGGACACATACCGCTACATCGACACCTACAGATACATC GTGTCTAAGGGCGAAGAGCTGATTAAGGAG	*Eco*RI
	EP_1215 (TagRFP, no STOP)	NNNAAGCTTATTAAGTTTGTGCCCCAGTTTGCTAGG	*Hind*III

*NNN denotes an equimolar mixture of all four bases. Underlined sequences represent restriction sites inserted.