Efficient Ends-Out Gene Targeting In *Drosophila*

Juan Huang¹ *, Wenke Zhou¹ *, Annie M. Watson¹, Yuh-Nung Jan², Yang Hong¹

¹ Department of Cell Biology and Physiology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261

² Howard Hughes Medical Institute, Departments of Physiology and Biochemistry, University of California San Francisco, San Francisco, CA 94143

* These authors contributed equally to this paper.
Running Title:
Efficient Gene Targeting in Drosophila

Keywords:
ends-out targeting, homologous recombination, Drosophila, Reaper, Crumbs, Arf6

Corresponding Author:
Yang Hong
Department of Cell Biology and Physiology
University of Pittsburgh School of Medicine
S325 BST, 3500 Terrace Street
Pittsburgh, PA 15261

Phone: 412-648-2845
Fax: 412-648-8330
Email: yhong@pitt.edu
In this report, we describe several approaches to improve the scalability and throughput of major genetic crosses in ends-out gene targeting. We generated new sets of targeting vectors and fly stocks, and introduced a novel negative selection marker that drastically reduced the frequency of false positive targeting candidates.

The development of homologous recombination-based gene targeting is a landmark breakthrough in *Drosophila* genetics (Gong and Golic 2003; Rong and Golic 2000). In particular, the so-called “ends-out” or replacement-type gene targeting offers a straightforward approach for generating either knock-out or knock-in alleles. To date, there are already more than twenty genes that have been modified by ends-out targeting (Supplementary Table 1). Nonetheless, the frequency of target-specific homologous recombination in *Drosophila* varies tremendously, ranging from higher than 1/200 gametes (Manoli et al. 2005) to less than 1/350,000 (Jones et al. 2007) (also Y. H, unpublished data) -- i.e., a more than 1800-fold difference. In cases of low targeting efficiency (<1/100,000 gametes), ends-out targeting can be exceedingly time- and labor-intensive. Here, we optimized the current ends-out targeting scheme by focusing on improving the scalability and throughput of its major genetic crosses. As illustrated in Figure 1a, there are three major genetic crosses in a typical ends-out targeting. In the **targeting cross**, virgin females of a transgenic line bearing the donor DNA (“P(donor)”) are crossed with *hs-FLP, hs-I-SceI* males, and their larval progeny are heat shocked to induce generation of linear donor DNA fragments by FLPase and I-SceI enzymes. In the **screening cross**, virgin females from the targeting cross that are of the correct genotype (*P(donor)/hs-FLP, hs-I-SceI*) are crossed with proper chromosome balancer males, and preliminary targeting candidates are recovered based on their w* marker. However, many of these candidates might be false-positives due to the failure of
excision or non-targeting integration of the donor DNA. In the mapping cross, only preliminary candidates whose \( w^+ \) marker is mapped to the target gene chromosome are selected for further analysis.

For the targeting cross, the number of \( P\{\text{donor}\}^{hs-FLP, \ hs-I-SceI} \) virgin females directly determines the scale of the whole targeting experiment. Genes that are resistant to homologous recombination may require collecting and sorting >15,000 virgins from the targeting cross (LARSSON et al. 2004), which is extremely labor-intense due to the time-sensitive nature of virgin collection and the genotyping process. To eliminate this major bottleneck in scaling up the targeting cross, we modified the original \( hs-FLP, \ hs-I-SceI \) stocks by replacing their Y chromosomes and balancer chromosomes with ones that contain \( hs-hid \) transgenes (GRETHER et al. 1995). We named these modified stocks "6934-hid" and "6935-hid" (Fig. 1b) after the original stock numbers. Ubiquitous expression of the cell death gene \( hid \) induced by heat-shock causes strong lethality. As illustrated in Figure 1c, in a targeting cross using 6934-hid, all male progeny and those female progeny carrying \( hs-hid \) balancer chromosome are eliminated. Since \( P\{\text{donor}\}/hs-FLP, \ hs-I-SceI \) females are the only genotype that survives, 6934-hid and 6935-hid completely eliminate the time-sensitive virgin collection and genotyping process.

For screening and mapping crosses, we found their throughput was often severely limited by the high background of false positives, which may represent more than 95% - 99.9% of preliminary candidates (J.H, W.Z. and Y.H., unpublished results, and see below). Therefore, we introduced a negative selection marker into the current ends-out targeting scheme, so the majority of non-targeted integrations may be directly eliminated before they are subject to any further screening and mapping efforts. Ectopic
expression of another cell death gene *reaper* (*rpr*), similar to *hid*, also causes strong lethality (WHITE et al. 1996). As illustrated in Figure 2b, a *UAS-Rpr* module can be tagged to the 3’ end of a transgenic donor DNA fragment (e.g., *P{crb::mEosFP*KO}). Once the donor DNA fragment is recombined into the target gene locus, *UAS-Rpr* will be lost due to homologous recombination. In contrast, non-targeted integrations will likely retain the donor DNA fragment with an intact *UAS-Rpr* module (“Rpr+”). By using proper *Gal4* driver stocks to set up the screening cross, Rpr+/Gal4 false-positive candidates will be directly eliminated due to the ectopic expression of Rpr.

To implement the *UAS-Rpr* selection, we made a new set of ends-out targeting vectors, pRK1 and pRK2 (Fig. 2a) that were based on the integration and modification of pEndsOut2 and pBS70W (available from http://dgrc.cgb.indiana.edu/). Both pRK1 and pRK2 contain a *UAS-Rpr* module, while pRK2 also has a GMR (HAY et al. 1994)-enhanced w’ marker to further facilitate the recovery of targeting candidates. We made two targeting constructs, a *crb::mEosFP*KO knock-in construct (Fig. 2b) and a *dArf6*KO knock-out construct (Fig. 2c), based on pRK1 and pRK2, respectively. Crb is a transmembrane protein essential for developing cell polarity (TEPASS et al. 1990). We plan to study the trafficking and dynamics of Crb by tagging it with a photo-convertible fluorescent protein mEosFP (WIEDENMANN et al. 2004). *dArf6* (*Arf51F*) is a small GTPase that may play key roles in *Drosophila* muscle and nervous system development, although no *dArf6* mutants are currently available. *dArf6*KO targeting aims to delete 2.158kb of the *dArf6* locus that includes all the coding exons plus the 3’UTR (Fig. 2c). We obtained multiple transgenic lines from both targeting constructs at normal frequency (Supplementary Table 2), indicating that *UAS-Rpr* in pRK1 and pRK2 was sufficiently silent in the absence of the Gal4 driver and did not adversely affect the routine P-element-based transgenic process. All the transgenic donor lines were larval or pupal
lethal when crossed with neuronal-specific drivers Gal4\textsuperscript{477} and Gal4\textsuperscript{221} (Supplementary Table 2)\cite{GRUEBER et al. 2003}. Gal4\textsuperscript{221}/UAS-Rpr also consistently produced very few adult escapers of a fully penetrated wing inflation phenotype (Supplementary Fig. 1d).

To evaluate the effectiveness of UAS-Rpr without any bias, we first carried out crb\textsubscript{::}mEosFP\textsuperscript{KI} knock-in experiments without UAS-Rpr selection (similar to Fig. 1a). From approximately 5x10\textsuperscript{4} screening cross progenies, we recovered 270 male candidates, of which fourteen were mapped to the 3\textsuperscript{rd} chromosome where crb is located. We then screened 125 non-3\textsuperscript{rd} chromosome candidates and all fourteen 3\textsuperscript{rd} chromosome candidates for the presence of UAS-Rpr by crossing them into Gal4\textsuperscript{221}. As summarized in Table 1, only 3/125 of the non-3\textsuperscript{rd} chromosome candidates (\textit{i.e.}, false positives) are Rpr\textsuperscript{f1}, while 11/14 of the 3\textsuperscript{rd} chromosome candidates are Rpr\textsuperscript{f1} of which seven were confirmed by PCR to have the correct targeting events (Supplementary Fig. 1a). Extrapolating from these data, selecting against UAS-Rpr in the screening cross of crb\textsubscript{::}mEosFP\textsuperscript{KI} would eliminate more than 96\% (253/263) of false positives (Table 1). In addition, UAS-Rpr selection also eliminates tandem-insertion mutants (GONG and GOLIC 2003), which can be difficult to distinguish from true targeting candidates by simple PCR assays (Supplementary Fig. 1b,c). The homologous recombination frequency of crb\textsubscript{::}mEosFP\textsuperscript{KI} is approximately 1/7,000 if only considering the male candidates.

We then decided to carry out a large-scale dArf6\textsuperscript{KO} targeting experiment by taking full advantage of the new reagents and methods described here, as we failed at dArf6\textsuperscript{KO} targeting based on the original pEndsOut2 vector by screening \(\sim 1.6\times 10^5\) screening cross progenies (W.Z. and Y.H., unpublished results). We re-cloned the same 5’ and 3’ homologous arms into the pRK2-based vector. By using 6935-hid to set up the targeting cross, we easily collected \(>2 \times 10^4\) virgin females. 12,000 of them were mated
with $w/Y; Pin/CyO; Gal4^{221}[w^+]$ males to set-up the screening cross (Fig. 1d). From $>7 \times 10^5$ screening cross progenies (Table 2) we recovered 315 $w^+$ males, of which five were verified as specific targeting candidates by PCR (Table 2, Supplementary Fig. 2a,b). As a control, 200 virgin females from the same targeting cross were crossed with regular $w/Y, Pin/CyO$ males. Of 124 $w^+$ male candidates recovered, 1.6% (2/124) were $Rpr^{-}$, but none harbored the true targeting event (Table 1&2). Thus, $UAS-Rpr$ selection achieved an impressive ~60-fold reduction of false positives. Effectively, $dArf6^{KO}$ targeting was accomplished at a scale equivalent to screening/mapping $>18,000$ (315x60) preliminary candidates in the absence of $UAS-Rpr$ selection. In addition, the dramatically reduced number of preliminary candidates, combined with their dark-red eye color due to GMR-enhanced $w^+$ expression in pRK2, made the screening process much easier and faster. The homologous recombination frequency of $dArf6^{KO}$ can be estimated as approximately $1/140,000$ if only considering the male candidates. Homozygotes of $dArf6^{KO}$ are viable but are male and female sterile, so it is possible that $dArf6$ only plays a specific and indispensable role in germline development. When we were preparing this manuscript, a P-element-induced deletion allele of $dArf6$ was published and Dyer et al. observed same male and female sterile phenotypes in their homozygous $dArf6$ mutant flies (DYER et al. 2007).

Compared with the “rapid scheme” in which preliminary candidates were screened for the loss of FRT sites (GONG and GOLIC 2003; RONG et al. 2002), $UAS-Rpr$ selection is more efficient since it directly eliminates false positives. In addition, we found that majority of the false positives (57%-87%) had damaged FRT sites (Table 1); therefore, they could only be eliminated by $UAS-Rpr$ selection but not by the FRT test. Since the $I-SceI$ sites are positioned rather close to the FRT sites in ends-out targeting constructs, we speculate that the frequent FRT damages seen here were most likely due
to the double-strand DNA repair process triggered by the premature cut of \textit{I-SceI} sites (BELLAICHE \textit{et al.} 1999; GONG and GOLIC 2003). Separating FRT and \textit{I-SceI} sites further away in future pRK-based targeting vectors should further reduce the frequency of false positives. Consistently, the Golic lab reported that the frequency of false positives was low using the pW25 targeting vector in which FRT and \textit{I-SceI} were separated by 100-150bp (GONG and GOLIC 2004). Since pRK-based vectors may not be suitable for making Gal4 knock-in alleles, pW25 series vectors should be excellent alternatives.

In summary, for a targeting experiment with an expected homologous recombination frequency of around $\sim 1/100,000$ gametes, we estimate that our 6934-hid/6935-hid stocks and UAS-Rpr selection reduced the work load of genetic crosses to a level comparable to a routine P-excision experiment. In addition, our new targeting vectors, such as pRK2, should significantly facilitate the molecular cloning and transgenesis of targeting constructs due to the enhanced multiple cloning sites and $w^+$ expression. Overall, these new reagents and methods should significantly increase the success rate of targeting experiments on genes that are resistant to homologous recombination.

**ACKNOWLEDGEMENTS**

We are sincerely grateful to Dr. Jeff Sekelsky for pBS70W and pEndsOut2 plasmids, Leon Perniciaro for help in screening in $d{\text{Arf}}^6{\text{KO}}$ targeting candidates, Dr. Ulrich Nienhaus for EosFP constructs, Dr. Fabrice Rogers for hs-hid stocks, Drs. Fen-Biao Gao, Sige Zou, Peizhang Xu, Koen Venken for comments on the manuscript. Y.N.J is an investigator of the Howard Hughes Medical Institute. pRK1 and pRK2 will be donated to the Drosophila Genomic Resource Center (DGRC) and 6934-hid and 6935-hid stocks
will be donated to the Bloomington Stock Center. This work is supported by start-up funds from the University of Pittsburgh School of Medicine (Y.H.).
FIGURE LEGENDS

Figure 1. Genetic crosses in targeting experiments.

a. Genetic crosses of a typical ends-out targeting experiment. The transgenic donor DNA ("P{donor}") is on the 2nd chromosome, while the target gene is on the 3rd chromosome. P{donor}*: linearized extrachromosomal donor DNA fragment that only exists transiently – it will either be lost permanently or inserted into a chromosome by targeted or non-targeted integration events. Target*: potential targeting events. Note that the majority of potential targeting events may be non-specific and not located on the 3rd chromosome (see text). ?: this copy of the 2nd chromosome is inherited from the female in the screening cross. It could be the donor chromosome or hs-FLP, hs-I-SceI, or the recombinant between the two. Nonetheless, this copy of the chromosome is irrelevant in the mapping cross.

b. Genotypes of 6934-hid and 6935-hid.

c. 6934-hid stock eliminates the virgin collection and genotype sorting in targeting cross.

d. The genetic cross scheme of the dArf6KO targeting experiment. Because dArf6 is on the 2nd chromosome, a transgenic line carrying dArf6KO donor DNA ("P{dArf6}Rpr+") on the 3rd chromosome was used. w; Pin/CyO; Gal4221[w-] stock was used to set up the screening cross in lieu of a regular Pin/CyO balancer stock. This allowed simultaneous selection against non-specific targeting candidates while balancing the potential specific targeting candidates from the screening cross. P{dArf6}Rpr+: linearized extrachromosomal dArf6KO donor DNA fragment. dArf6KO*: potential targeting events.

Figure 2. pRK1 vector and targeting of crb::mEosFPki and dArf6KO.

a. Only pRK1 is diagramed here. The hsp70::white (w+) transformation marker is flanked by two loxP sites so that w+ can be removed in the final targeted alleles by Cre
recombinase. **MCS**: multiple cloning sites. **Amp^R**: ampicillin-resistant gene. **3'P** and **5'P**: 3’ and 5’ P element sequences for transgenic insertion.

**b.** *crb::mEosFP^{Kd}*_ knock-in targeting. mEosFP (“FP”) is fused in frame at the position right outside the transmembrane domain of Crb.

**c.** *dArf6^{Kd}*_ knock-out targeting. Dotted bar indicates the targeted deletion (2.158kb).

In both **b** and **c**, gray or black boxes are exons and white boxes are introns (introns and exons are not shown for CG8155, CG8157 and CG8160 in **c**). Fine dotted lines indicate the homologous recombination event. Solid bars indicate the positions and sizes of diagnostic and verification PCRs (for PCR results please see Supplementary Figure 1&2). Diagrams are not precisely to scale.
Table 1. Genetic and PCR analyses of targeting candidates of *crb::mEosFP<sup>Ki</sup>* and *dArf6<sup>KO</sup>*

<table>
<thead>
<tr>
<th>Targeted Allele</th>
<th>Targeting Candidates</th>
<th>Rpr test</th>
<th>FRT&lt;sup&gt;+&lt;/sup&gt;</th>
<th>loxP&lt;sup&gt;+&lt;/sup&gt;</th>
<th>X-Chr.</th>
<th>PCR Verified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 122</td>
<td>17/122</td>
<td>122/122</td>
<td>0/122</td>
<td>n/d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 3</td>
<td>0/3</td>
<td>3/3</td>
<td>0/3</td>
<td>n/d</td>
</tr>
<tr>
<td>crb::mEosFP&lt;sup&gt;Ki&lt;/sup&gt;</td>
<td>Non-3&lt;sup&gt;rd&lt;/sup&gt; chr. candidates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 3</td>
<td>0/3</td>
<td>3/3</td>
<td>--</td>
<td>2&lt;sup&gt;+&lt;/sup&gt;/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 11</td>
<td>0/11</td>
<td>11/11</td>
<td>--</td>
<td>7/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 110</td>
<td>54/120</td>
<td>n/d</td>
<td>0/120</td>
<td>n/d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 0</td>
<td>2</td>
<td>n/d</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 2</td>
<td>0/2</td>
<td>n/d</td>
<td>0/2</td>
<td>n/d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 2</td>
<td>0/2</td>
<td>n/d</td>
<td>--</td>
<td>n/d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 0</td>
<td>2</td>
<td>n/d</td>
<td>--</td>
<td>n/d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 0</td>
<td>2</td>
<td>n/d</td>
<td>--</td>
<td>n/d</td>
</tr>
<tr>
<td>dArf6&lt;sup&gt;KO&lt;/sup&gt;</td>
<td>Non-2&lt;sup&gt;nd&lt;/sup&gt; chr. candidates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;-&lt;/sup&gt;: 0</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rpr&lt;sup&gt;+&lt;/sup&gt;: 0</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*Rpr<sup>-</sup>:* scored by lethality or strong wing phenotypes in the presence of *Gal4<sup>221[w-]</sup>* or *Gal4<sup>177[w-]</sup>*. The total number of *Rpr<sup>-</sup>* false positives of *crb::mEosFP<sup>Ki</sup>* can be estimated as ten (six from non-3<sup>rd</sup> chromosome candidates (3 x (256/125)), plus four from 3<sup>rd</sup> chromosome candidates). *FRT<sup>+</sup>* or *loxP<sup>+</sup>*: scored by eye color variegation in the presence of constitutively expressed FLPase or Cre recombinase. ~87% ((263-(17x2))/263) of *crb::mEosFP<sup>Ki</sup>* false positives and ~57% ((124-54)/124) of *dArf6<sup>KO</sup>* false positives showed damaged FRT sites. *X-Chr.*: candidates that were mapped to the X Chromosome. Here, none of the non-target chromosome false positives were mapped to the X chromosome. Since the 4<sup>th</sup> chromosome is extremely small, therefore unlikely to harbor any non-specific targeting events, “X-Chr” data indicates that virtually all of the non-target chromosome false positives retained their donor DNA on the original chromosome, either due to damaged FRT sites or insufficient excision of donor DNA. *:
tandem insertion mutants. **: Only \( d\text{Arf6}^{KO} \) candidates recovered from screening crosses with regular \( \text{Pin/CyO} \) balancer stock are listed here (see Table 2). \( n/d \): not done. --: not applicable.
Table 2. *dArf6*KO targeting with and without using UAS-Rpr as a negative selection marker.

<table>
<thead>
<tr>
<th>Screening Cross Set Up</th>
<th>Progenies Screened</th>
<th>Male Candidates</th>
<th>2nd Chr. Candidates</th>
<th>Rpr+</th>
<th>FRT+ PCR Verified</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X) w/Y; Pin/CyO; Gal422[w]</td>
<td>&gt;7 x 10^5</td>
<td>315</td>
<td>30/315</td>
<td>--</td>
<td>n/d</td>
</tr>
<tr>
<td>(X) w/Y; Pin/CyO;</td>
<td>~7,300</td>
<td>124</td>
<td>2 * /124</td>
<td>122/124</td>
<td>54/124</td>
</tr>
</tbody>
</table>

2nd Chr.: 2nd chromosome. n/d: not done. --: not applicable. *: These two candidates harbored non-targeted integration of donor DNA on the 2nd chromosome and were Rpr+.
REFERENCES:


DYER, N., E. REBOLLO, P. DOMINGUEZ, N. ELKHATIB, P. CHAVRIER et al., 2007 Spermatocyte cytokinesis requires rapid membrane addition mediated by ARF6 on central spindle recycling endosomes. Development 134: 4437-4447.


Targeting Cross

\( y^w \hspace{1em} \text{hs-FLP, hs-I-SceI} + \)
\( \text{y}^w \hspace{1em} \text{y}^w \rightarrow \text{CyO, hs-hid} \)
(6934-hid)

Heat shock @ 38ºC

Screening Cross

\( y^w + \hspace{1em} \text{TM3 or TM6b} \)
\( \text{y}^w \rightarrow \hspace{1em} \text{y}^w \rightarrow \)

Mapping Cross

\( y^w \hspace{1em} \text{y}^w \rightarrow \text{y}^w \rightarrow \text{y}^w \rightarrow \text{y}^w \)
\( y^w \rightarrow \text{y}^w \rightarrow \text{y}^w \rightarrow \text{y}^w \)

“6934-hid”

“6935-hid”

Targeting Cross

\( y^w \hspace{1em} \text{hs-FLP, hs-I-SceI} + \)
\( \text{y}^w \hspace{1em} \text{y}^w \rightarrow \text{CyO, hs-hid} \)
(6935-hid)

Heat shock @ 38ºC

Screening Cross

\( y^w \hspace{1em} \text{Pin} \)
\( \text{y}^w \hspace{1em} \text{y}^w \rightarrow \text{Gal4221[w-]} \)

Mapping Cross

\( y^w \hspace{1em} \text{dArf6}^{00} \)
\( \text{y}^w \hspace{1em} \text{y}^w \rightarrow \text{CyO} \)
\( \text{y}^w \hspace{1em} \text{y}^w \rightarrow \text{y}^w \rightarrow \text{y}^w \rightarrow \text{y}^w \)

chromosome mapping

J. Huang & Y. Hong  Figure 1