Efficient creation and characterization of multi-kilobase knockins by CRISPR/Cas9 and multiple rAAVs

Monica F. Sentmanat[1], Zi Teng Wang[1], Evguenia Kouranova[1], Samuel Peters[1], Wan Ching Chan[1], Jed Lin[1], Yong Miao[1], Julie Ritchey[2], Julie O'Neal[2], John Dipersio[2], J. Michael White[3], Mia Wallace[4], Xiaoxia Cui[1#]

[1]Genome Engineering & Stem Cell Center, McDonnell Genome Institute, Dept. of Genetics, Washington University in St. Louis School of Medicine, 660 S. Euclid Ave, St. Louis, MO, 63110

[2]Department of Medicine, Washington University in St. Louis School of Medicine, 660 S. Euclid Ave, St. Louis, MO, 63110

[3]Transgenic, Knockout and Microinjection Core, Department of Pathology & Immunology, Washington University in St. Louis School of Medicine, 660 S. Euclid Ave, St. Louis, MO, 63110

[4]Mouse Genetics Core, Department of Pediatrics, Washington University in St. Louis School of Medicine, 660 S. Euclid Ave, St. Louis, MO, 63110

[#]Corresponding author: x.cui@wustl.edu

Abstract

We report the successful creation of multi-kilobase knockin mice and cell lines using three rAAV donors and CRISPR/Cas9 via consecutive homology-directed repair and reliable generation of over 120 KI mice with rAAV donors. For comprehensive analysis of the post-edit genome, we combined target capture and long-read sequencing for a multiplexable and high-throughput, all-in-one method to detect on-target insertion and as importantly, various undesired editing outcomes in the genome. Together, the two novel methods greatly improve the efficiency and accuracy of challenging large KI mouse and cell models.

Main Text

Site-specific large insertions, or knockins (KIs), remain the more challenging type of gene edits, compared to simple knockouts, deletions, and point mutations or tagging. This is partially because of the difficulty in nuclear delivery of long DNA templates, commonly as plasmid. CRISPR RNPs and single-stranded oligodeoxyribonucleotides can be efficiently electroporated into single-cell mouse embryos, whereas plasmid donors must be microinjected, limiting throughput, survival and founder rate. Recently, rAAV was shown to deliver donor templates by simple incubation with mouse embryos prior to RNP electroporation[1,2]. Using this method, we report here the successful creation of 107 KI mice with insert size ranging from 400 bp to 3.5 kb (**Fig.1a**, **Supplementary Table 1**). We observed a general trend that the longer the insert, the lower efficiency. However, the stronger correlation is between KI efficiency and length of insert over homology, each doubling of homology length relative to the insert results in a 10% or greater increase in efficiency R=-0.29, p-value =0.001 (**Fig.1b**). Fifteen out of the 107 models were gene replacements, generated using one donor and two gRNAs. Gene replacement was as efficient as direct insertion among these models (**Fig.1c, Supplementary Table 1)**.

To overcome the 4.7 kb size limitation of an rAAV genome, we applied a successive insertion strategy, where the first donor delivers a portion of the insert along with a gRNA target site, cleavage of which mediating subsequent insertion from the second donor[3]. The gRNA target site in the donor can be identical to the genomic target site, or unrelated, such as a previously validated target site from a different species or a new target site created at the insertion junction. The gRNA target is not cleavable in the single-stranded rAAV genome. For even larger inserts, three rAAVs are used, and the first two rAAVs each brings in a unique gRNA target site to mediate second and third insertions (**Fig.1d)**. To date, we have completed thirteen models with 2-rAAVs and three with 3-rAAVs (**Fig.1e & Supplementary Table 2**), based on junction PCR screen. Examples of junction PCR screen for 1-, 2- and 3- AAV models are shown in **Supplementary Figs.1-3,** respectively.

All but two of the multi-rAAV models target the mouse ROSA26 locus, nine of which are for conditional expression of a cDNA at the ROSA26 locus, using the configuration CAG-LSL-cDNA-pA. Given the CAG-LSL-pA cassette is shared, we created a common rAAV donor to first insert the cDNA-less cassette (XCD427d, **Supplementary Fig.4**), and a second (and a third if needed) rAAV will deliver the cDNA. Among the nine projects using the common XCD427d donor, the six with one cDNA donor (total 2 rAAVs) had an average founder rate at 16%, whereas the three with two cDNA donors (total 3 rAAVs) had an average founder rate at 8%.

Encouraged by the data successful in mice, we used the same strategy on cancer and stem cell lines. In iPSCs, insertion of a GFP-fused cDNA into the AAVS1 site, project XCB910b, was more efficient with rAAV donor than plasmid donor (**Supplementary Fig.5**). Project XCD47c, using 2-rAAV donors also for insertion into the AAVS1 site, produced junction-positive clones in both HEK293T cells and iPSCs, with or without the second gRNA, even though it was much

more efficient with both gRNAs (**Supplementary Fig.6**). This result implies the possibility of recombination between the two rAAV donors prior to HDR but further confirmation is needed. Finally, using rAAV donors, we were also able to generate 7-kb KIs using 3 rAAVs in mouse BV2 cells, where long KIs were previously not feasible because BV2 cells do not survive plasmid transfection (**Supplementary Fig.7**).

Five 1-rAAV and three 2-rAAV mouse projects failed to produce a founder with two mouse sessions each (**Supplementary Tables 1&2**), even though gRNAs cut efficiently in cultured cells during validation (**Supplementary Table 3**). Some projects had animals with one or more positive junctions but at least one junction was not detectable in any (**Supplementary Fig.8**), implying either incomplete insertion or simply failed PCRs. To exclude the latter, several rounds of PCR optimization were often performed.

As insert size and number of rAAV donors increase, genotyping becomes increasingly challenging. Multiple junctions are necessary and can often be inconclusive. PCRs do not always work across the insertion junctions, depending on the locus and sequence context. For a better genotyping method, we first tried Cas9-targeted sequencing (nCATS)[4] (**Supplementary Fig.9**), but it was inefficient, low throughput and unable to detect random insertion of rAAV donors we frequently observed in founder genomes (**Supplementary Fig.10**) andas previously reported for viral donors[5–7]. To overcome these limitations, we implemented multiplexable target captured long-read sequencing (TCLR, **Fig.2a**). DNA is first fragmented with Covaris g-Tubes or tagmented with Tn5 and barcoded, then amplified post-hybridization using adapter-specific primers to enrich for donor and flanking genomic DNA regions and sequenced, enabling up to 96 samples from the same or different projects to be screened in one flow cell. After demultiplexing and alignment to the donor and reference genome, KI and wild-type alleles are detected with

high read depth (**Fig.2b, Supplementary Table 4**), and deviations from expected coverage are easily discernable in some single- (**Fig.2c**) and multi-rAAV samples (**Fig.2d**). Additional alignments of raw reads from **Fig.2c** are shown in **Supplementary Fig.11**, highlighting contiguous coverage across the full-length insert. Instead of struggling with potentially troublesome junction PCRs, animals positive for a simple internal PCR can then be screened using TCLR to conclusively identify founders.

Deviations from expected read depth per position indicate undesired copy number variation, resulting from donor concatenation, random integration, deletion, or partial insertions (schematics shown in **Supplementary Fig.12**). Junctions of random insertions and concatenations are confirmed by PCRs and sequencing (**Supplementary Fig.13&14**, respectively). Using TCLR, we found that F1s for one of the 3-rAAV projects that missed a junction by PCR, XCE54a, contain deletions up to 3.7 kb within homology arms (**Supplementary Fig.15**), which was not observed in other projects, including the three XCC113 projects. We hypothesize that because all three rAAV donors are at or bigger than 4.7 kb to accommodate the 9.7 kb total insert, possibly leading to incomplete packaging of rAAVs. Undesired editing events, including large deletions, imperfect insertions, random integrations as well as rearrangements, are not always detected by PCR-based genotyping methods, but their presence can seriously skew phenotype and reduces the precision of the research models. Previous work to document unwanted editing events used ddPCR and Sourthern blots to estimate copy number or low-pass WGS and ITR-specific amplification[5,7]. While these methods are useful for flagging the presence of some undesired integration events, they are not scalable and can miss errors not attributable to CNVs and are not as accurate as the single-base pair resolution by TCLR. TCLR provides the much needed, comprehensive analysis of the post-edit genome.

114 Both Tn5 and Covaris g-Tubes are valid DNA shearing methods for TCLR. Whereas g-Tubes
115 produce longer average read lengths, Tn5 tagmentation allows higher throughput, lower cost and
116 the use of crude lysate, instead of purified gDNA. Shown in **Supplementary Fig.16** is a
117 successful TCLR screen of multiple clones for XCB910b rAAV donor-mediated KI in iPSC cells
118 using Tn5 on lysates. Like founder animal screens, clones positive for an insert-specific PCR can
119 then be reliably screened by TCLR.

120 TCLR is applicable to genotyping cells and animals created using plasmid donors as well as cell
121 therapy products involving genomic insertions. As an example, we sampled three different
122 production batches of CAR-T cells using TCLR and readily identified three different sets of
123 random integration sites of the CAR cassette (**Supplementary Fig.17a**), as the cassette was
124 introduced into the genome via lentiviral infection. No clonal expansion was observed, and
125 insertion sites are mostly intragenic, occasionally in coding sequences (**Supplementary**
126 **Fig.17b**). TCLR can potentially be a sensitive QC method for CAR-T cells with further
127 optimization and deeper coverage to ensure therapeutic batches are without deleterious,
128 tumorigenic insertions, and similarly, for post-treatment patient monitoring.

129 In summary, we report two novel methodologies addressing the unmet needs in the gene editing
130 field: co-delivery of CRISPR/Cas9 with multiple rAAV donors for consecutive HDR events to
131 overcome low efficiency of large KIs and TCLR, a much needed, all-in-one genotyping method
132 for thorough characterization of the edited genome, with potential applications in cell therapy
133 products.

134

135 **References**

1. Chen, S. *et al.* CRISPR-READI: Efficient generation of knock-in mice by CRISPR RNP Electroporation and AAV Donor Infection. *Cell Rep.* **27**, 3780-3789.e4 (2019).
2. Yoon, Y. *et al.* Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses. *Nat. Commun.* **9**, 412 (2018).
3. Bak, R. O. & Porteus, M. H. CRISPR-Mediated Integration of Large Gene Cassettes Using AAV Donor Vectors. *Cell Rep.* **20**, 750–756 (2017).
4. Gilpatrick, T. *et al.* Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
5. Suchy, F. P. *et al.* Genome engineering with Cas9 and AAV repair templates generates frequent concatemeric insertions of viral vectors. *Nat. Biotechnol.* 1–10 (2024) doi:10.1038/s41587-024-02171-w.
6. Nelson, C. E. *et al.* Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat. Med.* **25**, 427–432 (2019).
7. Hanlon, K. S. *et al.* High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* **10**, 4439 (2019).

Acknowledgements

We thank Ben Humphreys, Gaya Amarasinghe, Daisy Leung and Ting Wang for comments and suggestions, Robert Fulton, Catrina Fronick, Paul Cliften for technical assistance. This work was partially supported by the Shared Resource Investment Program from Siteman Cancer Center.

Figure 1

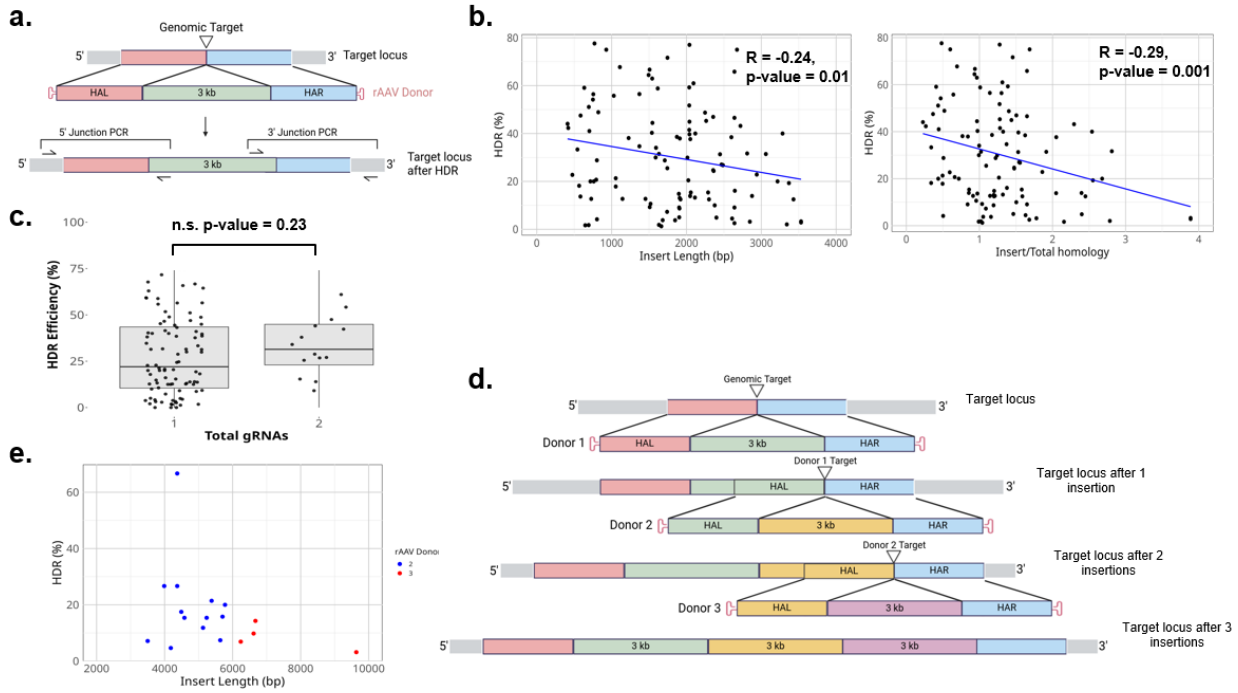

Figure 1. rAAVs are an efficient donor format for editing mouse embryos and sequential insertions from rAAV templates enable multi-kilobase KIs. **a.** Schematic of rAAV donor and junction PCR designs. Grey bars indicate genomic sequence flanking homology arms. **b.** Correlations of HDR efficiency by junction PCRs vs. insert length or vs. insert length/total homology length in 107 KI models. **c.** Boxplots comparing HDR efficiency of direct insertion vs. gene replacement KI models. **d.** Schematic of sequential insertions using 2- or 3-rAAV donors. **e.** KI efficiency by junction PCRs vs. insert length in 17 models with two (below 6000 bp) and four (above 6000 bp) rAAVs.

Figure 2

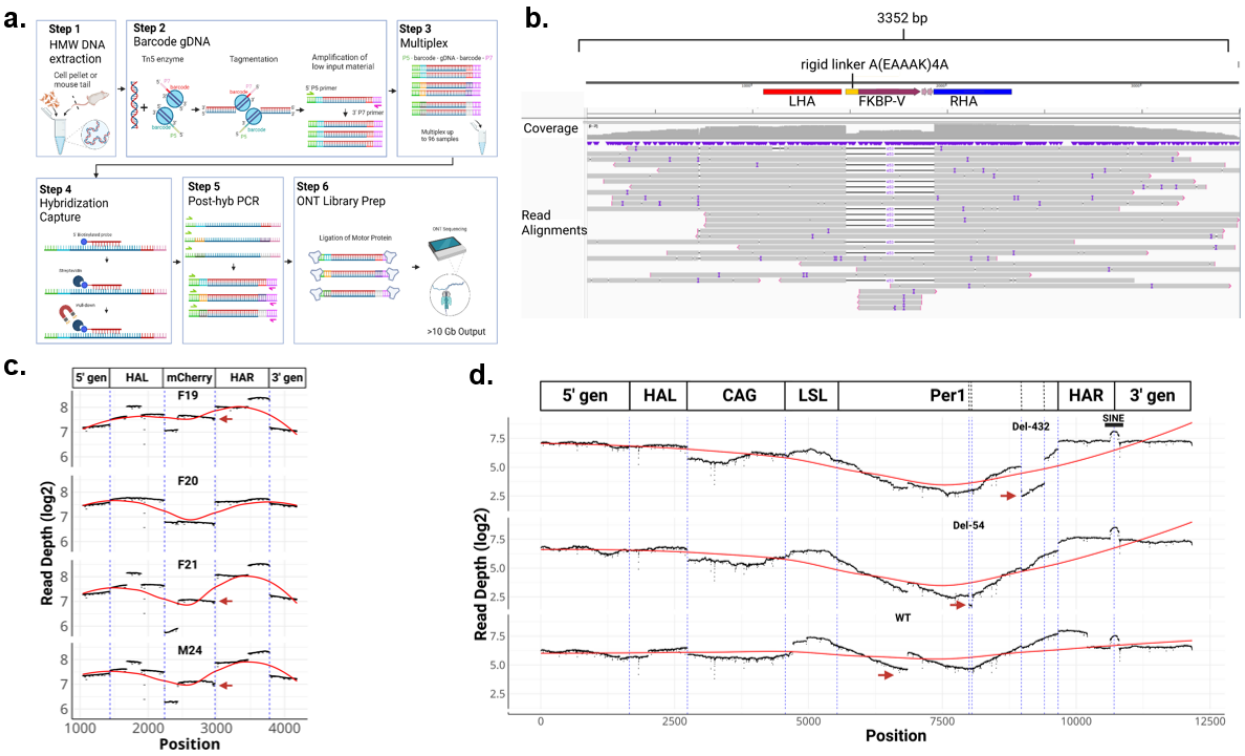

Figure 2. Target Captured Nanopore Long Read Sequencing (TCLR) as a thorough, high throughput genotyping method. **a.** Schematic of TCLR. **b.** IGV screenshot of aligned bam file for MS2757 illustrating read lengths cover genomic to genomic regions across donor cassette for both KI and wild-type alleles. **c.** Dotplot of read depth (log2 normalized) per base across XCB528 mCherry insertion (schematic on top). Dashed vertical lines (blue) mark boundaries of insert and homology arms. Red arrows point to copy number variation. **d.** Dotplot of read depth (log2 normalized) per base across the longer insert mediated by XCC113 3-rAAV donors. Dashed vertical lines (blue) mark boundaries of genetic regions: gen5, 5' genomic; L and R, left and right homology arms, CAG, CAG promoter; LSL: floxed polyA signals; Per1, cDNA variants (hashed boxes); pA, polyA signal. Red arrows mark copy number variation.