HOMOLOGY Medicines, Inc.

Method for Identification and Characterization of Sites of Homology Directed Strand Cross-Over Using rAAV Integration Vectors

Potts S, Prout J, Palladino A, Gall K, Rubin M, Selby D, Bourgeois-Tchir N, Zheng A, Thompson J and Wright J

CODE

Abstract

Homology-directed repair (HDR) is a natural DNA repair process that maintains genome integrity through precise correction of DNA damage. Genome editing tools that utilize HDR have enormous therapeutic potential for the correction of loss-of-function disease-causing mutations. Variability in homologous recombination activity is known to exist between genome locations and DNA sequence contexts; thus identifying genomic targets with the highest likelihood of HDR is an important consideration during the development of HDR-mediated therapies.

We and others have previously shown that recombinant AAVs (rAAVs) can integrate into a targeted locus by HDR without the use of a nuclease. Here, we report the application of high-throughput rAAV production and locus-wide HDR screens to identify and characterize recombinant integration at specific genomic sites that displayed increased strand cross-over. We created rAAV libraries with 0.1% – 1% base variation across the respective homology arms (HAs) of >20-donor HDR rAAV plasmids. Next, we employed high-throughput rAAV packaging and *in vitro* transduction of these libraries to create an arrayed rAAV HDR screen. Long-read sequence genotyping analysis of edited alleles allowed for mapping of strand cross-over sites. The accumulated results of strand cross-over exposed potential genomic hotspots of rAAV recombination. In total, these methods represent an approach with the potential to accelerate the screening for HDR sites with improved integration efficiency and expand existing knowledge of genome sequence characteristics that favor strand cross-over.

Goal:

Characterize crossover position along homology arms (HA) to identify regions that demonstrate higher efficiency of recombination. This information can then be used to correlate genomic landscape to predict other locations of high recombination efficiency.

1. Homology Arm (HA) Position

HA location's impact on crossover

- Reciprocal versus point insertion
- Complex SNP profiles to map crossover location

2. High Throughput Transduction

Vector creation and media transduction

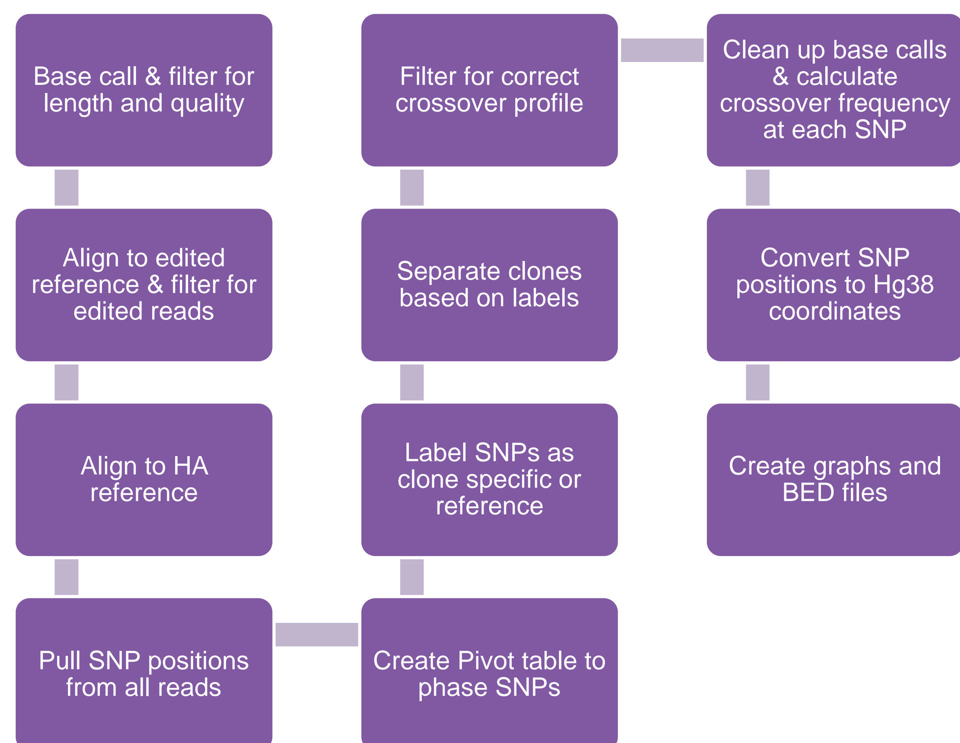
- Vector released to the media from packaging cells demonstrated to transduce and edit recipient cells effectively
- Bypassed traditional vector purification by media transduction to greatly reduce experimental duration and complexity

3. PCR, Sequencing, and Crossover Analysis

Evaluate integration and crossover

- Nested targeted-integration PCR spanning payload to genome
- Long-read NGS (ONT) sequencing to map crossover locations

Crossover Analysis



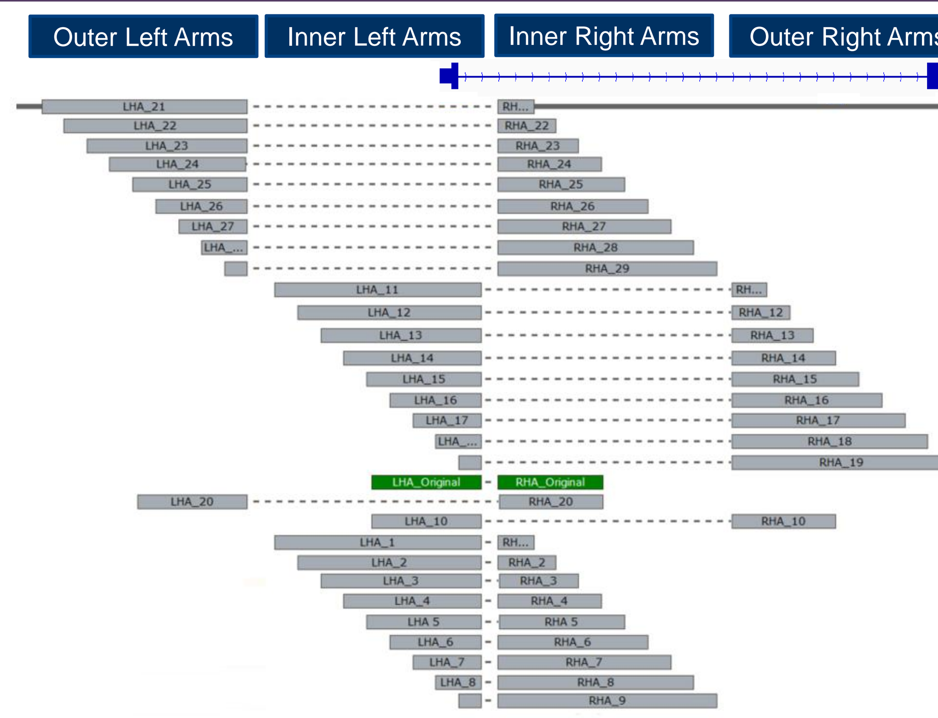
Conclusions and Next Steps

- Recombination occurred most frequently where there was higher probability of an endogenous double-strand break
- Perform ChIP-Seq and DSB assays to further characterize this relationship between crossover pattern and genomic landscape.

References

1. Raphaël M, Krzysztof G, Gaëlle L and Olivier C. (2018) Predicting double-strand DNA breaks using epigenome marks or DNA at kilobase resolution. Genome Biology 19:34.
2. Wilson D S and Keefe A D. (2001). Random Mutagenesis by PCR. Current protocols in molecular biology 8:Unit 8.3.
3. Jenjaroenpun P, Wongsurawat T, Sutheeworapong S and Kuznetsov V. (2017). R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. Nucleic Acids Research 45: D119-127.

1. HA Integration Screen Design



Study Design Highlights

- Arms ranged from 200 to 1900bp
- Point insertion versus reciprocal insertion
- Error-prone PCR [2] to create 3 clones with distinct SNP profile for each arm design
  - <1% difference from target region
- Same payload for all vectors

SNP Profiles

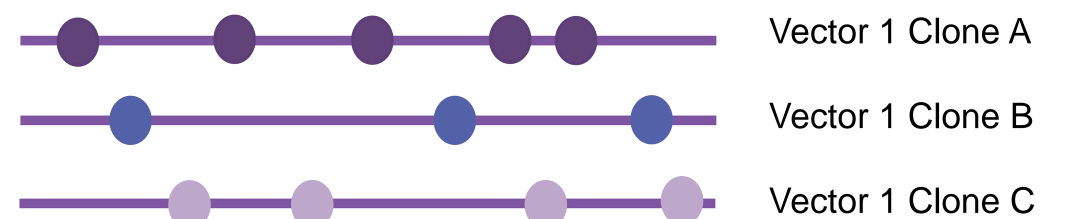


Figure 1: HAs were designed to vary in length and location. Insertion was either reciprocal to the size of the payload or a point insertion. Error-prone PCR was used to create 3 unique profiles per vector design to maintain <1% variation from the target but retain mapping potential across the whole arm.

2. High-Throughput Transduction

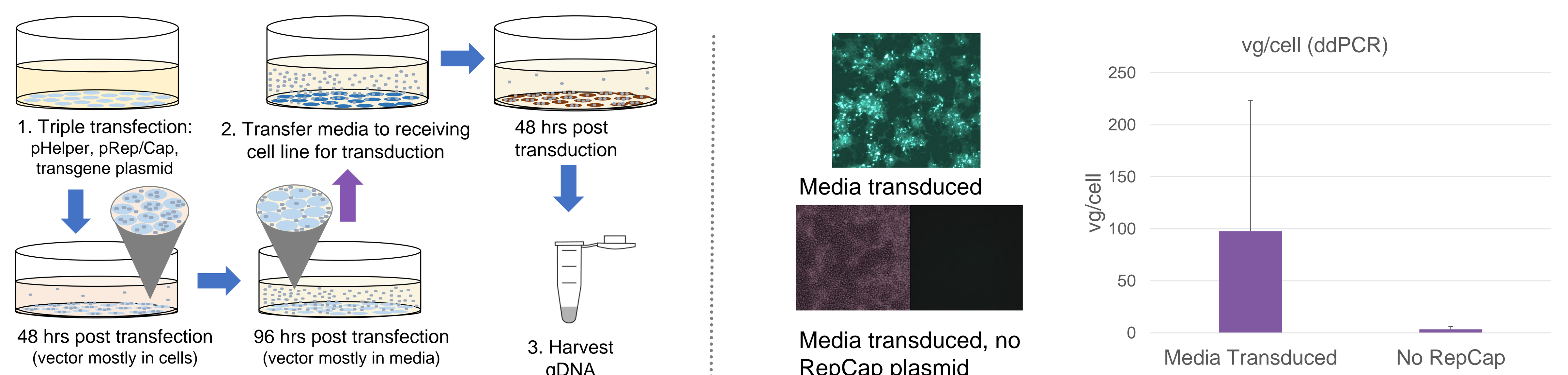


Figure 2: Individual transfection of plasmids followed by incubation of 96hrs to allow for proper packaging and release of vector into media. Media for each arm design was pooled and transferred to recipient cell line for transduction. Transduction in human lung and hepatocytes confirmed by GFP expression and ddPCR for vector genome (vg) copy number confirmed presence of vector. Negative control was transduced with media from cells transfected without the RepCap plasmid.

3. Crossover Analysis and Genomic Landscape Relationship

Figure 3. Overall Crossover Locations

Figure 3: Locus plot showing the full results of the crossover screen.

The top panel shows the genomic location of the screen, which spans the upstream region of the gene of interest, through Exon 2

The next panel predicts where double-strand breaks occur in lung tissue, generated with public data downloaded from ENCODE and using PredictDSB software[1].

The middle panel is the results for each vector. The top bars indicate where the HAs are located, and the following bars show, for each clone and replicate, the most common crossover interval. The interval is defined as the region between last reference SNP and first integrated SNP. An average of 50,000 reads per crossover interval were used (min 250 reads). Vector 15 (V15) left HA (LHA) is boxed in as more details are presented on this arm in Figure 5.

Figure 4. Total Crossover By Region

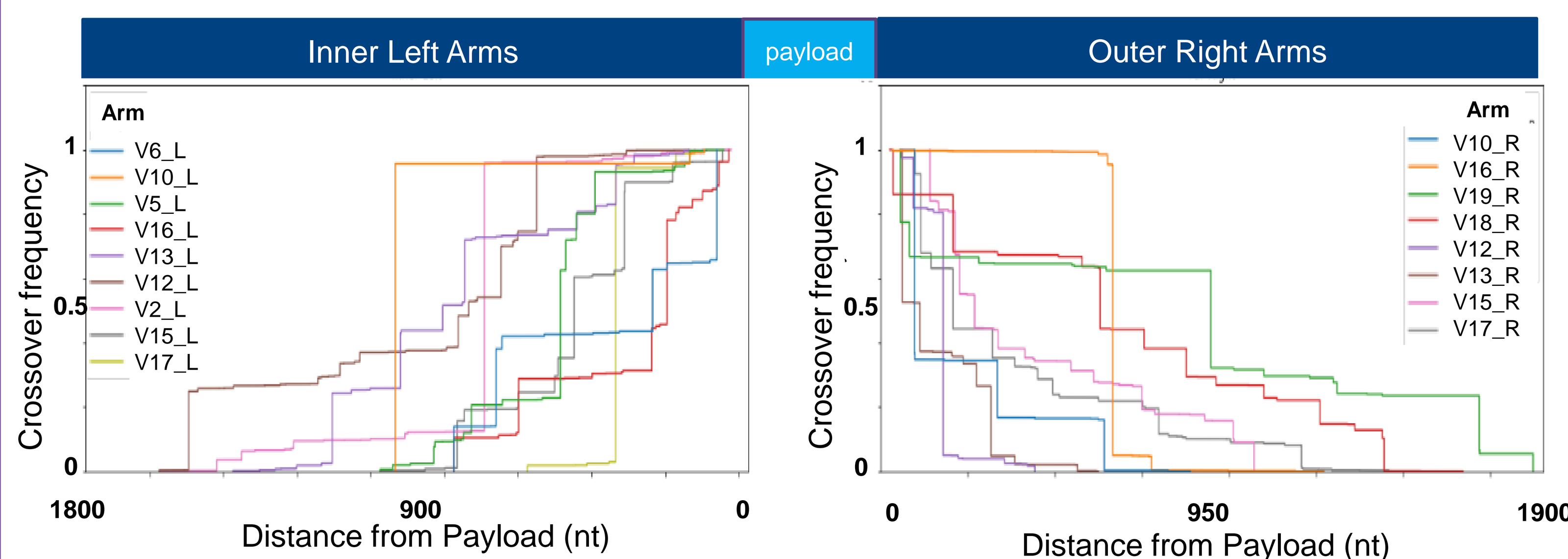


Figure 4: In the above figure, the screen design is divided into four sections as noted at the bottom.

Figure 4 shows the cumulative frequency of SNP conversions from reference to integration along each arm in the noted sections.

These traces are a composite of all clones and replicates for each vector (Left = 4.3M, Right = 1.6M reads).

Figure 5. Example: Vector 15 LHA Genomic Landscape

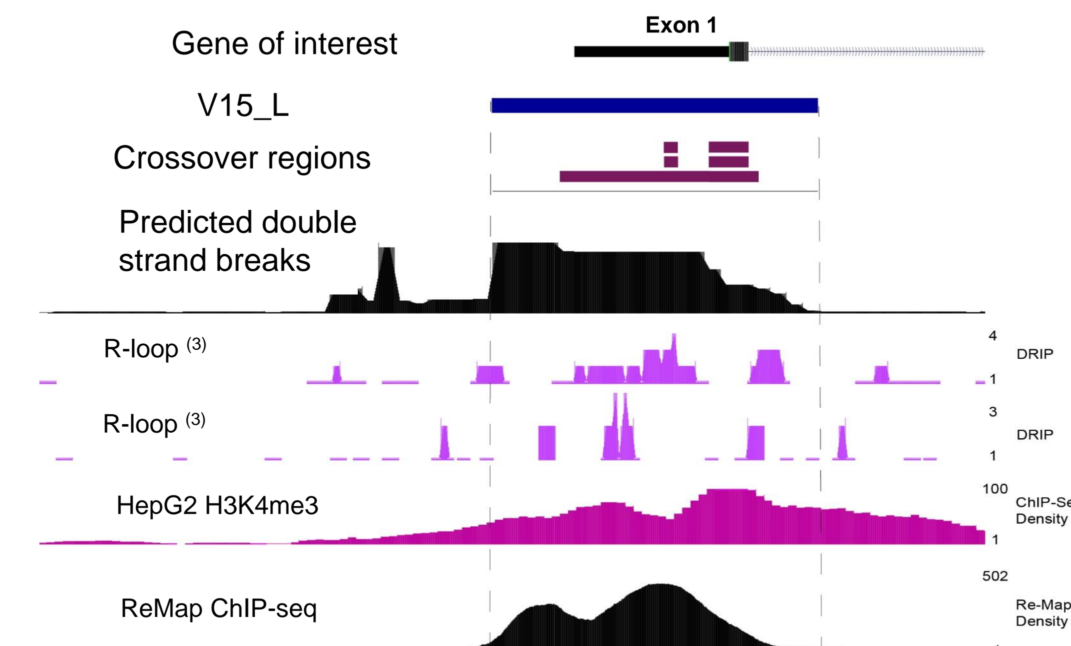


Figure 5: The genomic landscape and its relationship to Vector 15 LHA is presented here.

The genomic placement of the HA and the specific sites of crossover are shown aligned to the DSB prediction from Figure 3, as well as two R-loop tracks, an H3K4me3 track, and the Transcription Factor density from ReMap.

This alignment of genomic landscape elements hints at a possible relationship between recombination frequency and sites of endogenous double-strand breaks.