Targeted long-read sequencing approaches for fine-mapping genome editing events

Gabriel E. Rech[1], Chrystal Snow[2] and Chia-Lin Wei[1,2]

[1]The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06032, USA.
[2]The Jackson Laboratory, Bar Harbor, ME, 04609, USA.

The Jackson Laboratory
Leading the search for tomorrow's cures

ABSTRACT

The ability to modify genomes of model organisms using a wide range of genetic manipulation methods has provided us tools to study disease mechanisms and devise therapeutic strategies. Yet, the approaches to determine transgene integration or examine the modified loci are currently limited. The Jackson Laboratory is the world's largest source of genetically defined mice strains, sourcing more than 8,000 strains worldwide. The traditional and gold standard method for validating mouse models has been PCR based assays and Sanger sequencing. Despite its robustness, Sanger sequencing method is not highly sensitive on identifying novel low-frequency genomic alterations and can struggle to accurately characterize large and complex regions of the genomes. We have developed robust targeted long-read sequencing platforms for the identification and characterization of genome editing events in host genomes. Our established platform applies a CRISPR-Cas9 targeted approach[1], combined with long-read nanopore sequencing, to enhance sequencing depth at specific loci of interest. This platform provides an end-to-end workflow to fine map vector/transgene integration sites and their associated genomic aberration. The entire workflow has been optimized with a wide range of sample types. Because the robustness of the capture reactions and the simplicity of the nanopore sequencing operation, the process can be easily automated and is scalable to achieve high throughput and consistency. We expect that this platform will be valuable for vector integration assessment and reveal functional insight in the frequency, complexity, impacts and clonality of the integration events.

METHODS

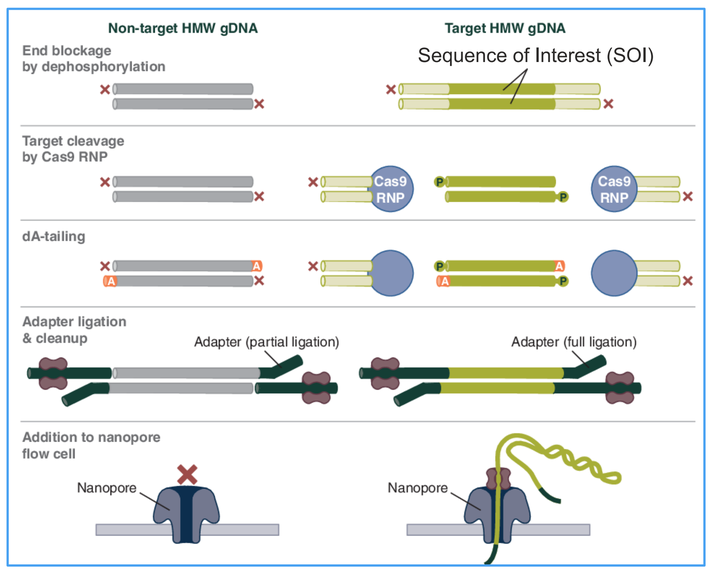
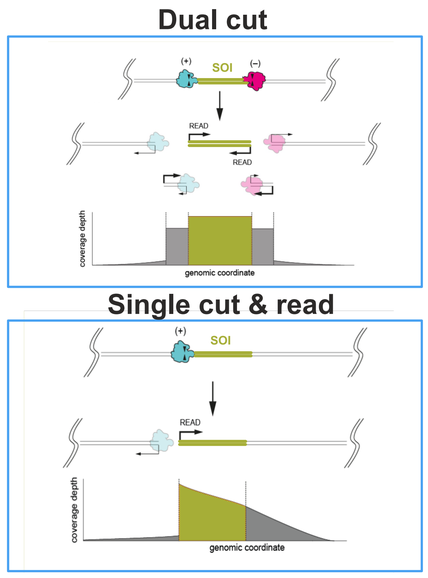


Figure 1: Overview of the Cas9 no-amplification enrichment library prep workflow[2].



- **SOI is <20 kb.**
- **Know the region on either side of the SOI.**
- **Useful for characterizing whole locus (structural variations, SNPs, methylation)**

- **Any size SOI.**
- **Only one side of the SOI is known.**
- **Useful for identifying location of random transgene / vector insertions.**

Figure 2: Two different flavors of the CRISPR-Cas9 targeted approach.

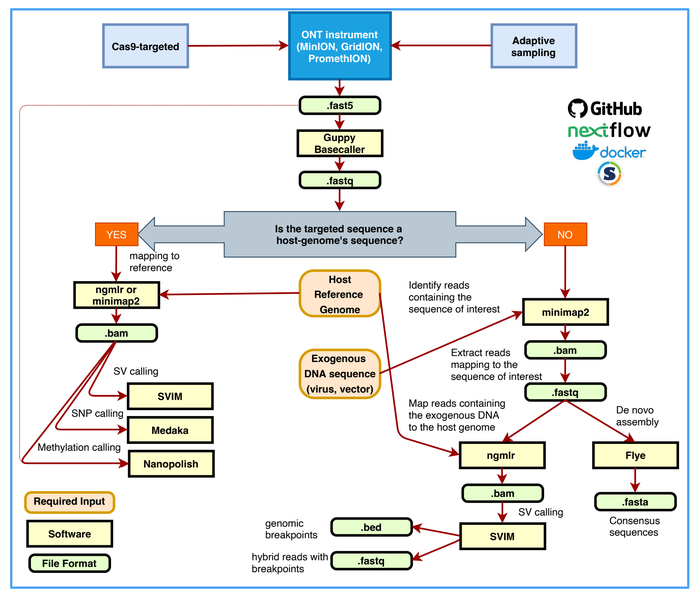


Figure 3: Customized ONT CRISPR-Cas9 targeted sequencing computational analysis.

RESULTS

Reproducibility study

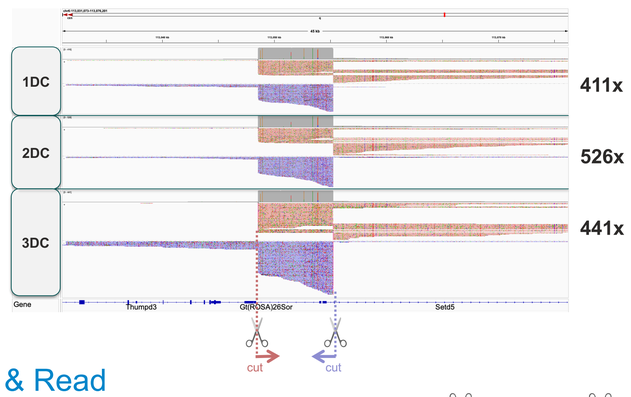
Sample ID	DNA input (µg)	Library yield (µg)	Library loaded onto FC (total µg)	Number of pores at start	Max Pore Occupancy (%)	Total Data Produced (GB)	Number of reads	Coverage over SOI	Reads with host genome junctions (insertions)	Assay Type
1DC	3.6	1.35	0.969	1356	8	11.46	73,393	411	NA	Dual Cut
2DC	3.6	1.40	0.969	1627	15	19.22	140,344	526	NA	Dual Cut
3DC	3.6	1.29	0.969	1417	10	13.96	97,355	441	NA	Dual Cut
1CR	3.6*	2.71*	2.03*	1454	10	14.34	108,001	616	73	Single Cut & Read
2CR	3.6*	3.33*	2.03*	1465	12	14.32	127,983	494	45	Single Cut & Read
3CR	3.6*	3.15*	2.03*	1347	15	20.94	169,543	762	77	Single Cut & Read

*Cut & Read assay requires 2 libraries pooled at final clean up.

Table 1: Reproducibility experiment for Dual Cut (DC) and Single Cut & Read (CR) assays. A 6.7Kb mouse genomic locus was targeted in the DC assay. A known exogenous transgene insertion of 14Kb was targeted in the CR assay.

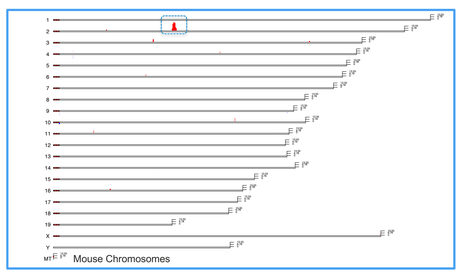
Dual Cut

Aim: To characterize mouse ROSA26 locus.

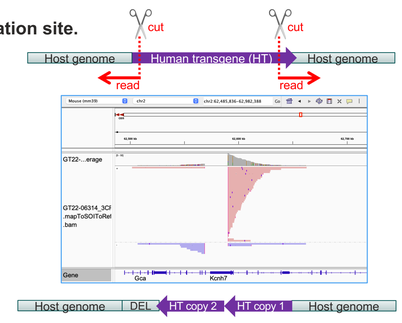


Single Cut & Read

Aim: Mouse model validation. Detect transgene integration site.



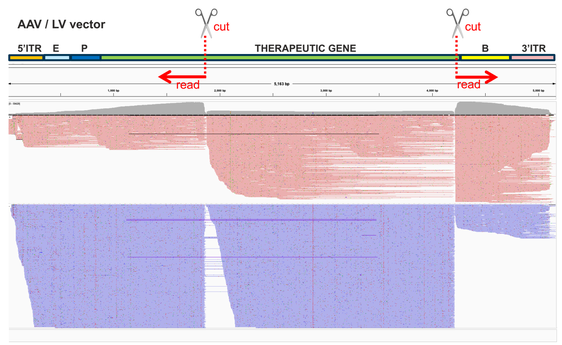
Mouse genome showing high read coverage at chromosome 2 when mapping reads containing the SOI (human transgene) in sample 2CR. Similar pattern observed in samples 1CR and 3CR.



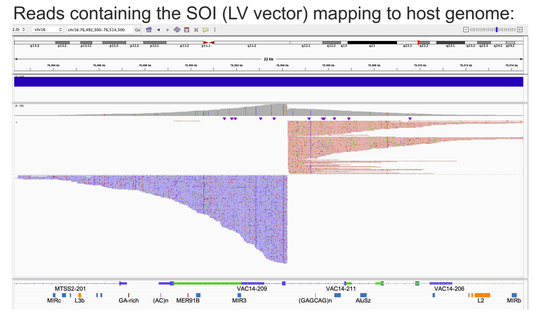
Consensus sequence shows insertion site at chromosome 2 and structure of the insertion (transgene tandemly duplicated and deletion at the host genome in the junction site).

Use case I

Aim: Assessment of gene-therapy vector genome integration.



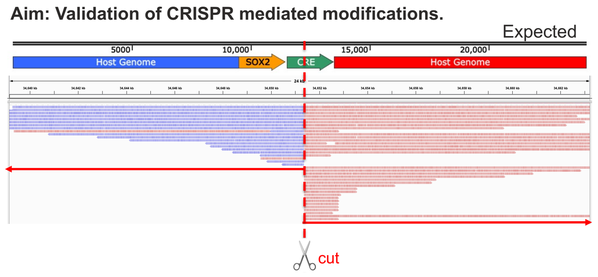
Reads containing the SOI (LV vector) mapping to host genome:



LV putative integration site detected.

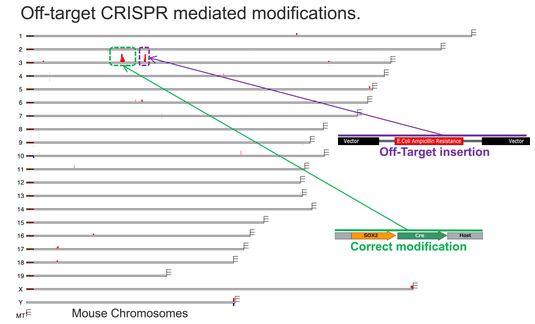
Use case II

Aim: Validation of CRISPR mediated modifications.



- Host gene SOX2 modified with the addition of CRE using CRISPR.
- On-target and unwanted off-target integrations identified in the same experiment.
- Integration of ***E.coli*** related gene.
- Identified insertion of CRISPR vector.

Off-target CRISPR mediated modifications.



Simultaneous detection of on and off target insertions.

Targeted long-read sequencing:

- Rapid identification of genome modifications, including complex structural variations.
- Amplification free.
- Insertion sites.
- Insert composition.
- Off target insertions.
- Greater variety of input compared to other methods.

www.jax.org

1. Gilpatrick, et al. "Targeted nanopore sequencing with Cas9-guided adapter ligation". Nature Biotechnology (2020).
2. Lesbirel, Simon et al. "A faster workflow for the assessment of genomic loci in mice using a novel HMW DNA extraction technology upstream of Cas9 targeted sequencing." (2021).