Precision Mapping Viral Integration Landscapes with CRISPR-Cas9 and Long-Read Sequencing

Anne J Hout, Vera G Buyle, Johanna FB Pagano, Karthikeyan Devaraju*
Research & Development, Cerba Research (Netherlands), Rijswijk 2288ER, The Netherlands
(presenting & corresponding author)*

Cerba Research

cerbaresearch.com

Background:

- Gene therapies with integrating vectors such as transposons, lentiviral (LVV) and retroviral (RVV) vectors, are used for stable expression by integrating a therapeutic gene into the host's genome.
- The LVV or RVV are mostly used in cell and gene therapies, especially for CAR / TCR-T (chimeric antigen receptor / T-cell receptor) therapies. These vectors follow random integration patterns and concerns regarding vector genotoxicity by activation of proto-oncogenes, warrant safety testing.
- Vector integration sites (VIS) in the host genome needs to be precisely identified and reported, to confirm that the therapeutic gene is safely inserted into the genome and do not give rise to therapy induced oncogenesis.

Materials & Methods:

- We designed guide RNAs that target conserved regions in the lentiviral vectors to selectively enrich for regions of interest in a background of human DNA.
- World Health Organization (W.H.O.) reference material that has VIS reported at 10 different genomic locations, was used.
- The MinION flow cells and relevant sequencing library preparation kit (V14 kit) for the minION Mk1C were ordered from Oxford Nanopore Technologies (ONT), UK.
- Using standard pipelines for sequence alignment and enrichment, insertion sites were annotated using a custom script.

Results:

- We developed an unbiased targeted approach using CRISPR-Cas9 for enrichment followed by sequencing on Oxford Nanopore's Mk1C sequencer to identify VIS in W.H.O. control material.
- The sequencing runs met all quality control criteria and no errors occurred. Basecalling was done using the latest software tool developed by Oxford Nanopore Technologies. The resulting data were analyzed using a custom-made bioinformatics pipeline, which essentially identifies integration sites by mapping reads to both the viral vector and human reference genomes, determining the alignment breakpoints of the chimeric reads.
- Using a combination of the gRNAs and long read sequencing, we could enrich and reliably detect all the ten reported integrated sites in the W.H.O. control sample. Our pipeline enables sensitive and reliable detection of integrated LVVs used in gene therapies (CAR / TCR-T and other gene therapies).
- Interestingly, our results indicated that by altering the gRNA binding in cis or trans, we could vary the depth of sequencing for the viral vector.
- In addition to VIS, since the proviral sequence is available, we propose that it can also identify:

1. Potential recombination, if any, of the vector; or
2. Mutations in transgene that hinder the therapeutic gene expression.

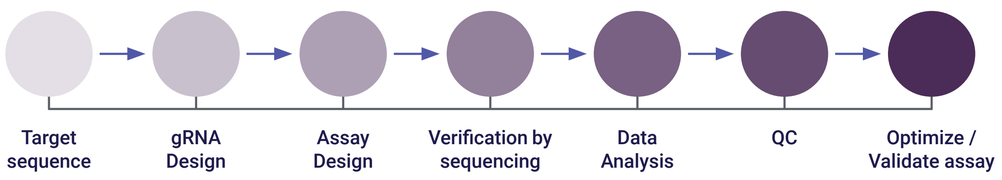


Figure 1: VIS Assay Workflow for CRISPR-Cas9 targeted ONT sequencing

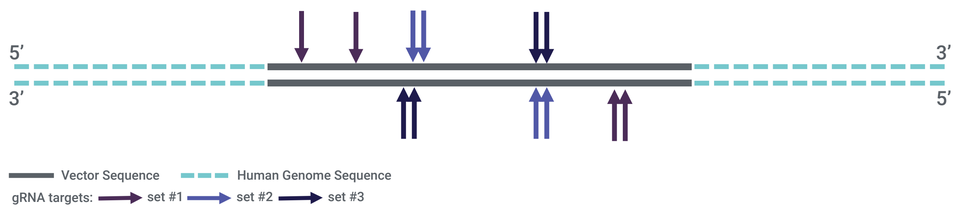


Figure 2: gRNA design on target sequences

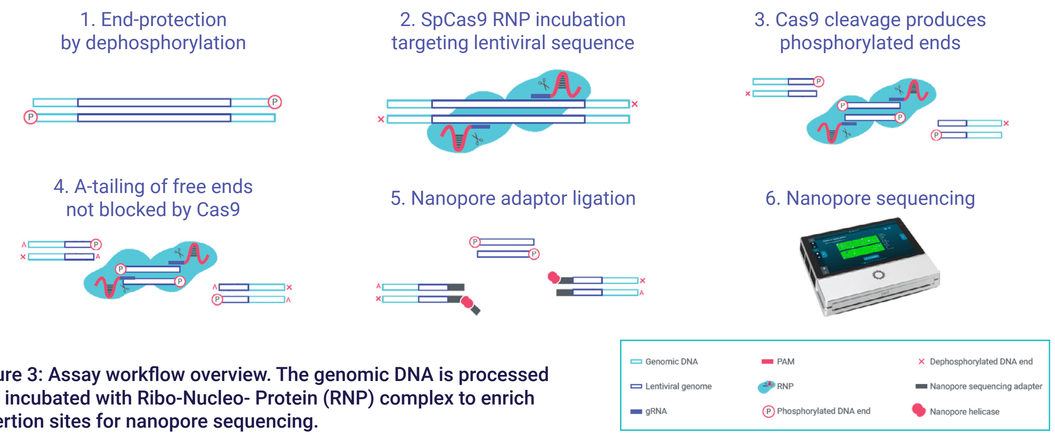


Figure 3: Assay workflow overview. The genomic DNA is processed and incubated with Ribo-Nucleo- Protein (RNP) complex to enrich insertion sites for nanopore sequencing.

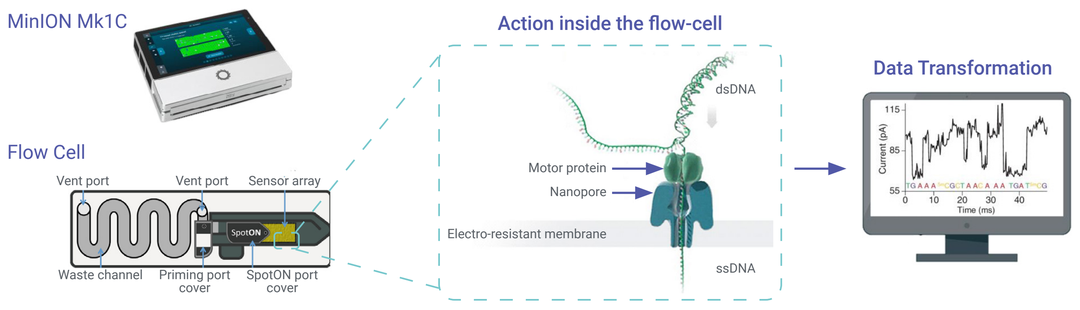


Figure 4: Snapshot of Nanopore Mk1C and sequencing process inside a MinION flow cell and transformation of electrical signals to nucleotides.

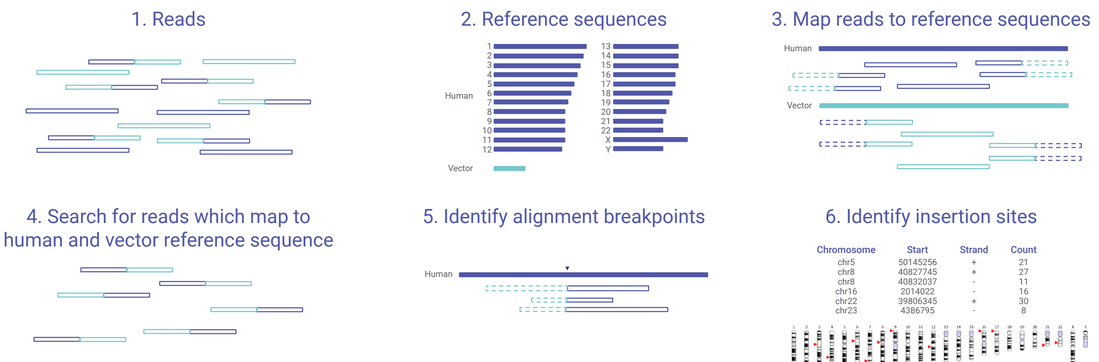
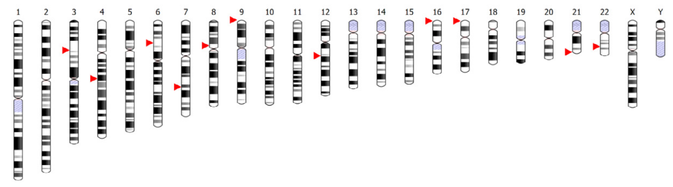


Figure 5: Data Analysis workflow for mapping the nanopore sequence reads to the vector and human genome sequences. The integration sites are identified by custom pipeline.

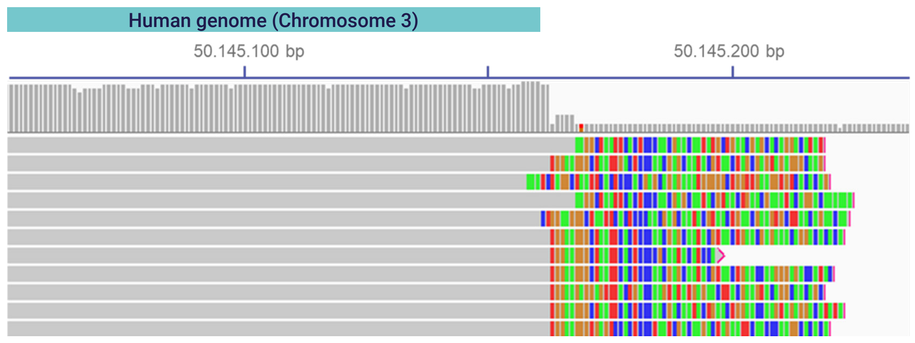
A. Table of identified Vector Integration Sites

Chromosome	Start (approx.)	Gene	Coverage	Identified
21	43867986	AGPAT3	23	✓
6	34299319	NUDT3	22	✓
3	50145157	SEMA3F	19	✓
4	93625425	GRID2	15	✓
8	39055454	ADAM9	16	✓
12	57274964	R3H DM2	11	✓
9	163465	CBWD1	17	✓
7	114508378	FOXP2	11	✓
22	39825823	ENTHD1	15	✓
17	1821489	SMYD4	22	✓
9	40832036	LOC100132004	19	✓
16	2014021	ZNF598	31	✓

B. Visualization of all the vector integration sites in the different human chromosomes



C. The IGV plot shows the sequence reads that map to 3, highlighted in A.
The coloured bars represent non-matching sequences that are the viral vector sequences.



D. The IGV plot shows the proviral vector sequence (~4Kb) that is mapped to the reference viral vector sequence (~7Kb). The reference sequence includes the integrating sequences within the LTRs (~ 4Kb) and flanking non-integrating transfer plasmid sequences (~3Kb). The gaps in the vector sequence show the Cas9 cleavage sites.

Figure 6: The above images and tables show the VIS data obtained with targeted nanopore sequencing. A., Lists the identified integration sites from the sequencing runs (n=4). The sites in blue box within the table shows the sites identified by many laboratories but not defined by W.H.O. B visualises the summary VIS data in depicting the integration sites within the chromosomes. C., shows the human genome sequence from the site highlighted in table A. D., shows the vector sequence identified from multiple genomic locations.

Conclusion:

- All integration sites documented & defined by WHO have been identified in our assay.
- We also identified the additional sites documented and reported by other labs that participated in the WHO study for the reference sample.
- Our VIS pipeline is reliable and reproducible.
- We have developed an unbiased targeted sequencing approach for vector integration sites, using CRISPR-Cas9 for enrichment followed by sequencing on Oxford Nanopore's MinION Mk1C sequencer to identify VIS in W.H.O. control material.

Future Considerations:

- Implement our VIS assay and data analysis pipelines to target integration sites from Retroviral vectors.
- Explore feasibility and implement our pipeline for detecting integration sites of transposons

Key References
1. WHO/BS/2019.2373 2. https://doi.org/10.1089/hgtb.2017.078 3. https://doi.org/10.1093/nar/gkaa1152