Navigating Viral Integration Landscapes with CRISPR-Cas9 and Long-Read Sequencing

Anne J Hout, Nikki Claassen, Ilse H Wolters, 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 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 their use needs genetic safety testing because of concerns regarding vector genotoxicity by activation of proto-oncogenes.
- Vector integration sites (VIS) in the host genome need to be identified precisely to confirm that the therapeutic gene is safely inserted into the genome and will not give rise to therapy induced oncogenesis.

Materials & Methods:

- We designed guide RNAs (gRNA) against conserved regions in the lentiviral vectors to selectively enrich for regions of interest in a background of human DNA.
- We used synthetic lentiviral construct, single viral integration cell lines and World Health Organization (W.H.O.) reference material (10 different integration sites).
- The MinION flow cells and relevant sequencing library preparation kit (V14 kit) for the MinION Mk1D 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 designed gRNAs to target the vector sequences for enrichment of integration sites with CRISPR-Cas9 and followed by sequencing on Oxford Nanopore's Mk1D sequencer.
- We first used a synthetic DNA construct having the gRNA target sequences to verify the pipeline. The sequencing runs met all quality control criteria and no errors occurred. Basecalling was done and resulting data was analyzed using a custom-made bioinformatics pipeline, which essentially identified the vector (Figure 6).
- Next, we tested the pipeline to identify VIS in W.H.O. control material (Figure 7). The resulting data was analyzed using our bioinformatics pipeline, which by mapping reads to both the viral vector and human reference genomes, determining the alignment breakpoints of the chimeric reads thereby identifying the integration sites.
- Using a combination of the gRNAs in cis and trans, we could enrich and reliably detect all the ten reported integrated sites in the W.H.O. control sample.
- We then used a cell line, that has 1 lentiviral integration event to verify our pipeline. Our pipeline identified the integration site (data not shown) .
- Our pipeline enables detection of integrated LVVs used in gene therapies (CAR / TCR-T and other gene therapies) – both the integration sites and the vector sequence.
- We are currently establishing the sensitive range of the assay.

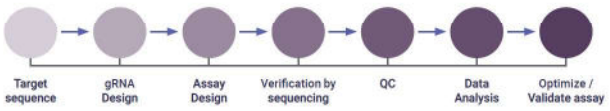


Figure 1: VIS Assay Workflow for CRISPR-Cas9 targeted ONT sequencing

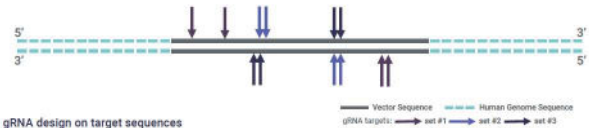


Figure 2: gRNA design on target sequences

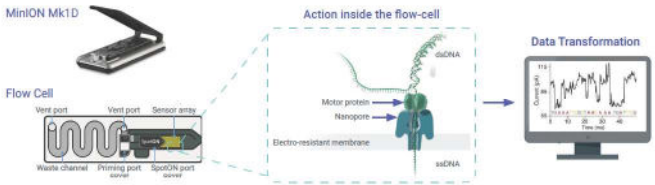


Figure 3: Snapshot of sequencing process inside a MinION flow cell and transformation of electrical signals to nucleotides.

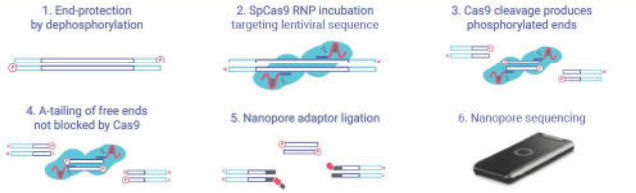


Figure 4: Assay workflow overview. The genomic DNA is processed and incubated with Ribo-Nucleo- Protein (RNP) complex to enrich insertion sites for nanopore sequencing.

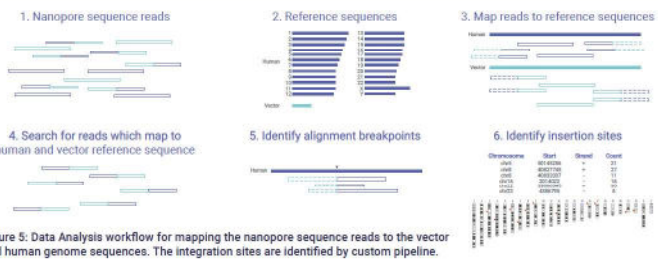
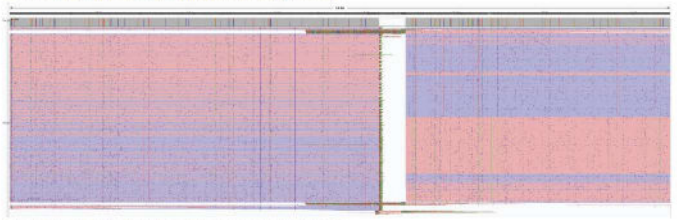


Figure 5: Data Analysis workflow for mapping the nanopore sequence reads to the vector and human genome sequences. The integration sites are identified by custom pipeline.

Key References
1. WHO/BS/2019.2373 2. https://doi.org/10.1089/hgtb.2017.078 3. https://doi.org/10.1093/nar/gkaa1152 4. https://doi.org/10.1038/s41587-020-0407-5

A. Reads for the synthetic DNA sequence targeted by gRNA set#1.



B. Reads for the synthetic DNA sequence targeted by gRNA set#3.

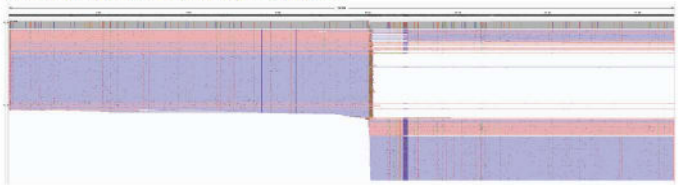
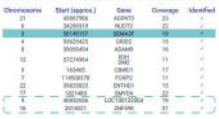
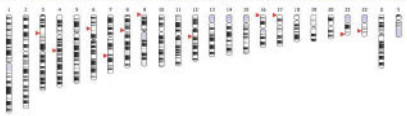


Figure 6: The above images show the reads from the synthetic DNA sequence screened for gRNAs targeted nanopore sequencing. A., shows data from gRNA set#1. B., shows the data gRNA set#3. The gRNA set#1 was better than gRNA set#3 with more reads (coverage).

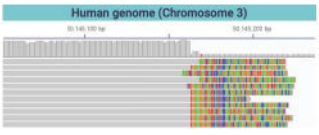
A. Table of identified Vector Integration Sites



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



C. The IGV plot shows the sequence reads that map to chromosome 3, highlighted in A.



D. The IGV plot shows the proviral vector sequence (~4Kb) that is mapped to the reference viral vector sequence (~7Kb). The gaps in the reads correspond to the Cas9 cleavage sites.

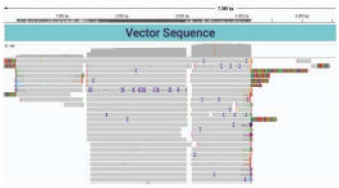


Figure 7: The above images and tables show the VIS data obtained with targeted nanopore sequencing. A., Lists the identified integration sites from the sequencing runs (n=4). The sites in blue box within the table shows the sites identified by many laboratories but not defined by W.H.O. B., visualises 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 W.H.O. have been identified in our assay.
- We also identified the additional sites documented and reported by other labs that participated in the W.H.O. study for the reference sample.
- We also established that the pipeline can detect single integration events.

Future Considerations:

- Explore approaches to improve the coverage.
- Actively establishing the sensitivity of the pipeline.
- Looking for collaborations to test in lentiviral vector therapies.