# ONT Lenti Detection

## Methods

### Cas9 enrichment for lentivirus insertion mapping

To prepare enriched Oxford Nanopore Technologies (ONT) sequencing libraries, we used PoreChop to design 2 gRNAs (5’-agatccgttcactaatcgaatgg-3’ and 5’-ggaacagtacgaacgcgccgagg-3’) for Cas9-mediated cleavage approximately 1kb within each end of integrated lentiviral sequences. These gRNAs were designed to not match elsewhere in the hg38 human reference genome. We confirmed their on-target efficiency by Cas9 (IDT: Alt-R® S.p. Cas9 Nuclease V3; Cat no. 1081058) cleavage of the lentiviral DNA, visualised on gel, in a separate experiment. DNA dephosphorylation (NEB: Quick CIP; Cat No. M0525S), single guide (IDT: Alt-R® CRISPR-Cas9 crRNA and Alt-R® CRISPR-Cas9 tracrRNA; Cat no. 1072532) and RNP formation, Cas9 cleavage and subsequent library preparation (ONT: SQK-CS9109) were largely performed according to the ONT Cas9 enrichment guidelines. We increased the starting amount of DNA to 5µg, and the dephosphorylation and cleavage incubation times to 2 hours and 24 hours, respectively. For two replicates of each reprogramming method, we then loaded 350ng of the enriched DNA library on to a MinION R9.4 flowcell, as per the manufacturer’s recommendations, and sequenced for 48 hours. Additionally, for the hmi32 fibroblast sample, 3µg of unenriched DNA was sequenced on a PromethION R9.4 flowcell (library prep kit SQK-LSK110) by the Kinghorn Centre for Clinical Genomics (KCCG).

### Bioinformatic analyses

Reads with a Phred score ≥ 10 were basecalled with Guppy (version 5.0.11). These reads were mapped with minimap2 (version 2.17) to both the human reference genome (hg38), and the sequence of the expected lentiviral insert (Li, 2018). Alignment maps were filtered with samtools (version 1.13) to only keep primary alignments with a length ≥ 800 bp, and a mapping quality of 60 (Li et at., 2009). Reads that mapped to both hg38 and the lentivirus sequence were retained and then subjected to another round of filtering. Here, reads were discarded when the base pair interval between the alignments to the lentiviral sequence and hg38 on the read was ≥ 51 bp. Reads that originated from the unenriched library and comprised a complete (≥ 4500 bp) putative lentiviral insert, spanned by a genomic alignment, as identified by TLDR (version 1.2.2) were kept (Ewing et al., 2020). Exact insert sites per read were identified based on the coordinates of both alignment maps (hg38 and lentiviral) to the original read. Exact insert sites were clustered together with bedtools (version 2.30.0) cluster within a 50 bp interval (Quinlan & Hall, 2010). For each cluster, the coverage was calculated and the smallest start- and largest end coordinates were selected as the exact insert site. Bedtools intersect was used to compare the occurrence of insert locations between samples.

To estimate the total diversity of the lentiviral insert population, an estimated Poisson distribution, based on the number of clusters and their respective coverages, was calculated per sample. This distribution was used to estimate the zero-counts. These zero-counts were used to improve the estimated Poisson distribution, and subsequently the zero-counts, until convergence. This generated a Poisson estimated total insert population size. Jaccard similarity scores were calculated based on the coverage of overlapping and non-overlapping insert sites between samples.