Comprehensive genetic analysis of lentiviral integrations in genetically modified T cells

Emily Caulfield, Irina Sergeeva, Joost Swennenhuis, Ellen Stelloo, Gabrielle Dijksteel, Mario Cangiano, George Georgiou, Judith Bergboer

Cergentis B.V., a Solvias company, Utrecht, The Netherlands

Summary

Comprehensive genetic QC by TLA analysis:

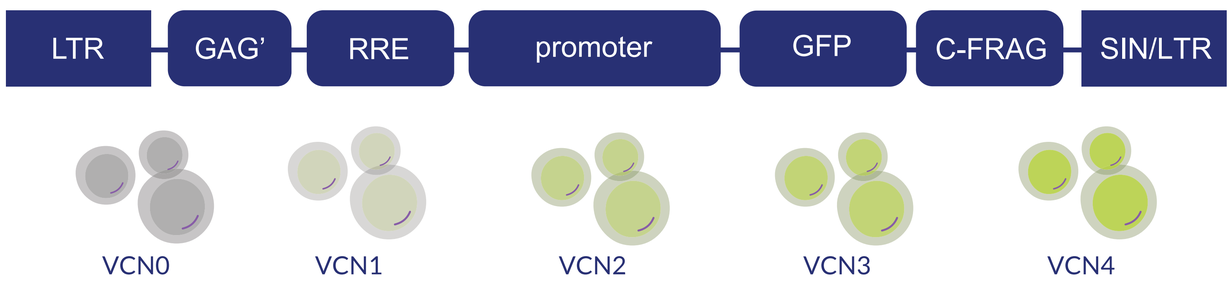
- Detects single nucleotide variants in the vector sequence
- Robustly identifies all vector integration sites & PLIER software automatically identifies these events
- Allows analysis of insertional mutagenesis
- Detects genomic rearrangements at or near the integration sites

Vector copy number (VCN) determination by ddPCR:

- Validates TLA-identified integration sites/junctions
- Precisely determines vector copy number

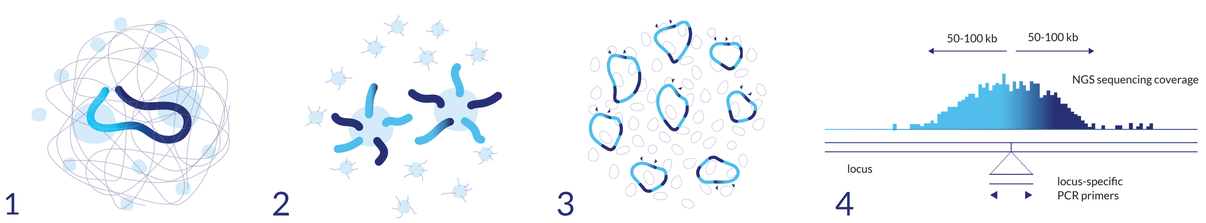
GFP+ Jurkat T cells

NIST National Institute of Standards and Technology

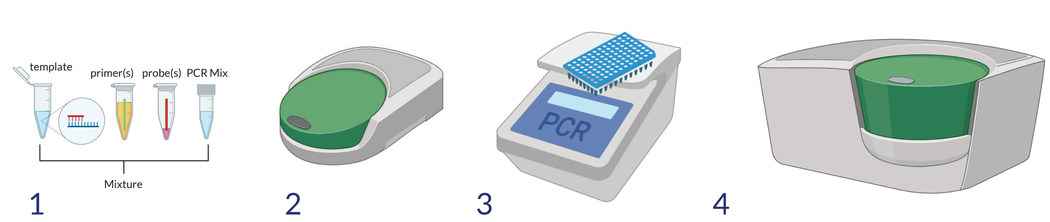


NIST transduced Jurkat (Clone E6-1) T cells with a lentiviral construct expressing GFP. [Paugh et al. *Sci Rep* (2021)]. The transduced T cells were clonally isolated to produce stable cell lines with 1, 2, 3, and 4 copies of the integrated provirus (VCN). The parental cell line (VCN=0) and the clonal transduced cell lines were evaluated as candidate reference materials for integration site analysis and vector copy number measurements within an interlab study. Cergentis joined the interlab study and performed comprehensive genetic QC by TLA and ddPCR analysis.

TLA technology & ddPCR



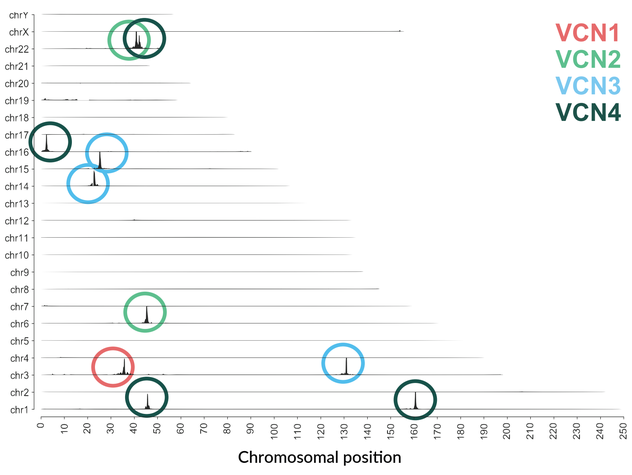
The TLA technology. 1) Genomic DNA is crosslinked. The crosslinking occurs between sequences in close proximity, coming from the same locus. 2) Crosslinked DNA is fragmented, religated and decrosslinked. 3) Digestion and ligation results in large DNA circles composed of unique combinations of DNA fragments from the same locus. Circular fragments from the locus of interest are amplified with inverse primers complementary to a short locus-specific sequence. 4) Broad NGS sequencing coverage is obtained across the locus of interest and >100 kb of neighboring DNA sequences.



VCN measurement by ddPCR. [Adapted from Lambrescu et al Int J Mol Sci (2022)] 1) ddPCR mixture (buffer, dNTPs, primers, probes, and DNA) is prepared. 2) The mixture is randomly distributed in 20 000 droplets. 3) PCR amplification is performed in all droplets simultaneously.4) The numbers of negative and positive droplets are counted based on fluorescence amplitude and analyzed with the QuantaSoft software. ddPCR assays are designed on 2 reference genes, breakpoint sequences of the integration sites, and vector element RRE.

Integration site analysis (ISA)

A



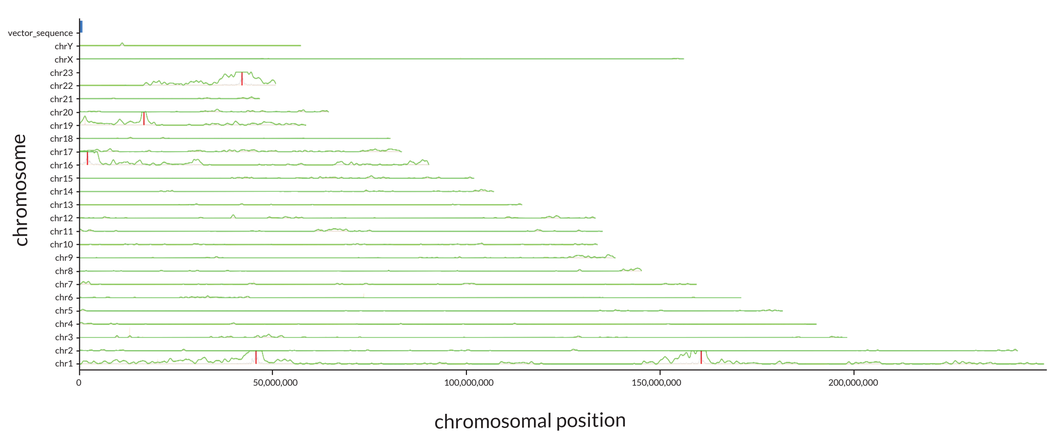
VCN1
VCN2
VCN3
VCN4

B

Sample	Integration site(s) Paugh *et al.* 2021	TLA	Gene
VCN0	none	✓	-
VCN1	chr3:35,632,784	✓	-
VCN2	chr6:45,210,754	✓	SUPT3H
	chr22:40,844,112	✓	ST13
VCN3	chr3:130,934,986	✓	ATP2C1
	chr14:22,564,447	✓	-
	chr15:25, 124,969	✓	SNHG14
VCN4	chr1:45,634,499	✓	GPBP1L1
	chr1:160,440,748	✓	-
	chr4:5,614,465	chr16:2,259,698	RNPS1
	chr22:42,082,676	✓	SMDT1

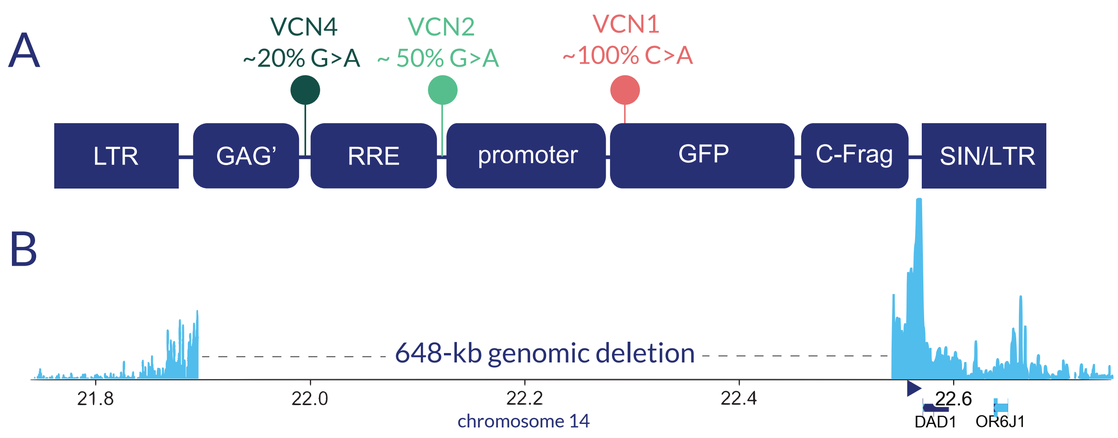
A) Whole genome coverage plot showing the integration sites of lentiviral vector GFP in Jurkat cell lines VCN1 (red), VCN2 (light green), VCN3 (blue) and VCN4 (dark green) as identified with a TLA primer set specific for vector element RRE. The primer specific peak on chr19 is masked. B) All but one of the genomic positions of the integrations were previously identified. An integration site previously estimated at chr4 in a non-unique genomic LINE1 sequence, has now been confirmed to be at chr16. 7 of 10 integrations occured intragenic.

Automated integration site analysis



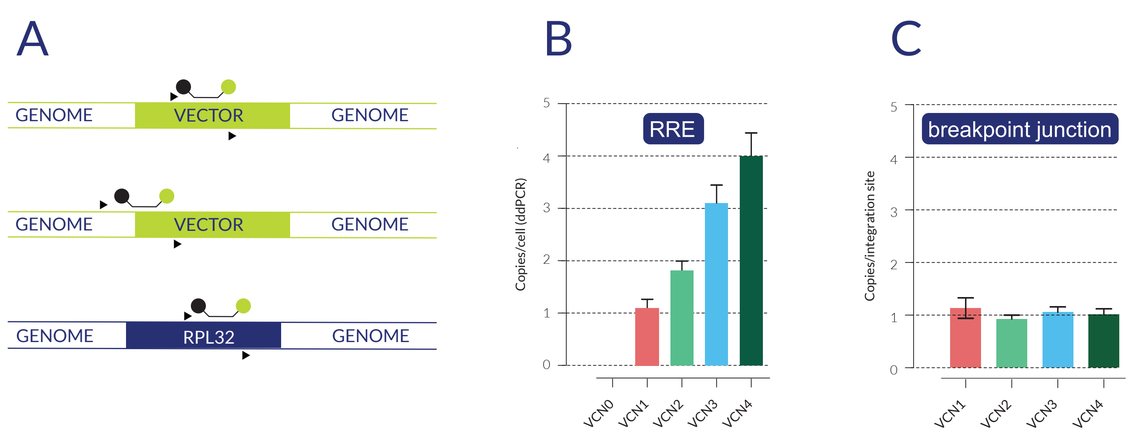
PLIER enrichment score distribution for cell line VCN4 (single TLA primer set shown, specific for vector element RRE). PLIER identifies enrichments by counting the number of proximity-ligation products, originated from the vector sequence, over genomic segments of 5 and 75 kb. A derived proximity score is compared to an expected reference value obtained by random reshuffling of all scores generated genome-wide. Regions showing both 5 and 75 kb above-threshold scores are hence identified as integration sites. The tool has been tested on a dilution series (1:10, 1:100 and 1:1000) of the Jurkat cell lines, providing a recall of 1 in all cases.

Vector & genomic integrity



A) High coverage across the entire vector sequence enabled small sequence variant analysis. A single variant was identified in the integrated vector sequence of each cell line VCN1, -2, and -4. Based on the frequency, these variants are only present in one of the integrated vector copies. B) The broad coverage surrounding the integration site shows a 648-kb genomic deletion near the integration site (green arrow) on chr14. No genes are annotated in the deleted region.

Vector Copy Number (VCN)



A) ddPCR assays are designed on the RRE vector element, breakpoint junction of the integration site and human RPL-32 gene. B-C) ddPCR assays show 0, 1, 2, 3 and 4 vector copies per cell for the VCN1-4 cell lines, and 1 vector copy per integration site which aligns with the TLA data.

CERGENTIS A SOLVIAS COMPANY