CERGENTIS
COMPLETE GENE SEQUENCING

Example report

Prepared for:

Company name:
Customer name:
Quote number:
Project number:
Version:
Date:

Goal

In this study, 1 transgenic human cell line with the vector X sequence was analysed.
The aim of this analysis was to:

1. Sequence the vector;
    1) Determine the presence of sequence variants and their allele frequency.
    2) Determine the presence of vector-vector breakpoints that represent concatemers of multiple copies of the vector and/or structural rearrangements in a single vector sequence.
2. Identify vector integration site(s) and breakpoint sequences between the vector and genome.
3. Assess the presence of structural variants surrounding the vector integration site(s).
4. Estimate the copy number of the vector.

An overview of the TLA technology and technical details of the performed analyses is provided in the manual "Introduction to the terminology and methods used in TLA analyses".

Summary

Sample	Sequence and structural variants in vector	Integration site(s)	Structural variants at the integration site	Notes
Sample 1	6 sequence variants, no deletions, 3 concatemers	chr14:24,733,900-24,733,901	no	-

Conclusion

In Sample 1, 3-5 copies of the vector have integrated in chromosome 14. 6 sequence variants are found within the vector sequence.

TLA, sequencing and data mapping

Viable frozen cells were used and processed according to Cergentis' TLA protocol (de Vree et al. Nat Biotechnol. Oct 2014). An overview of the TLA technology and technical details of the performed analyses is provided in the manual "Introduction to the terminology and methods used in TLA analyses".

TLA was performed with 2 independent primer sets specific for the vector sequence (Table 1).

Table 1: Primers used in TLA analysis

Primer set	Name/View point	Direction	Binding position	Sequence
1	Amp	Rv	1,132	X
		Fw	1,256	X
2	GOI	Rv	3,564	X
		Fw	3,842	X

The NGS reads were aligned to the vector sequence and host genome. The human hg38 genome was used as the host reference genome sequence.

Results SAMPLE 1

Vector sequencing coverage

Figure 1 depicts the NGS coverage across the vector sequence using primer set 1. Same results were obtained with primer set 2.

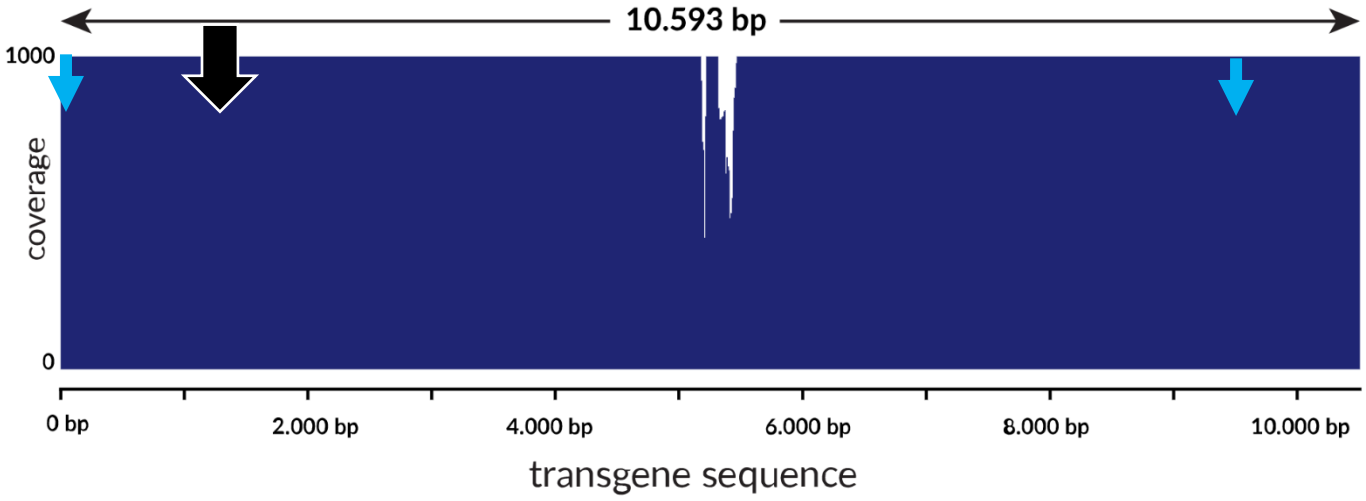

Figure 1: NGS sequencing coverage across the vector with primer set 1. The black arrow indicates the primer location. The light blue arrows indicate the locations of the identified vector-genome breakpoint sequences (described below). Y-axis is limited to 100x. Same results were obtained with primer set 2.

High coverage is observed across the complete vector sequence Vector: 1-10,250. Local dips in coverage are due to GC rich regions that are less efficiently sequenced.

Sequence variants and structural variants were called in the covered regions.

Sequence variants

Detected sequence variants are presented in table 2.

Table 2: Identified sequence variants

				Primer set 1		Primer set 2	
Region	Position	Reference	Mutation	Coverage	%	Coverage	%
Amp	141	A	C	967	30	1354	28
GOI	1013	T	-4AGTT	1181	100	1542	100
GOI	2956	T	C	1578	50	1262	56
GOI	5698	A	+1G	1631	52	1147	50
Backbone	9487	T	G	1845	100	1098	100
Backbone	10037	G	A	1455	20	897	21

Structural variants

The identified vector-vector breakpoint sites are shown in table 3. In the accompanying excel tables the sequences and frequencies of the breakpoints are presented.

Table 3: Vector-vector breakpoints

Breakpoint	Vector		Vector		Orientation of the breakpoint	Homology	Insert
1	→	8945	9657	←	tail to tail	1	-
2	←	21	5054	→	head to head	-	2

2 vector-vector breakpoints were found. Intact reads were also found at the positions of both breakpoints indicating that (partial) vector sequences have concatemerized.

Integration sites

Whole genome coverage plot

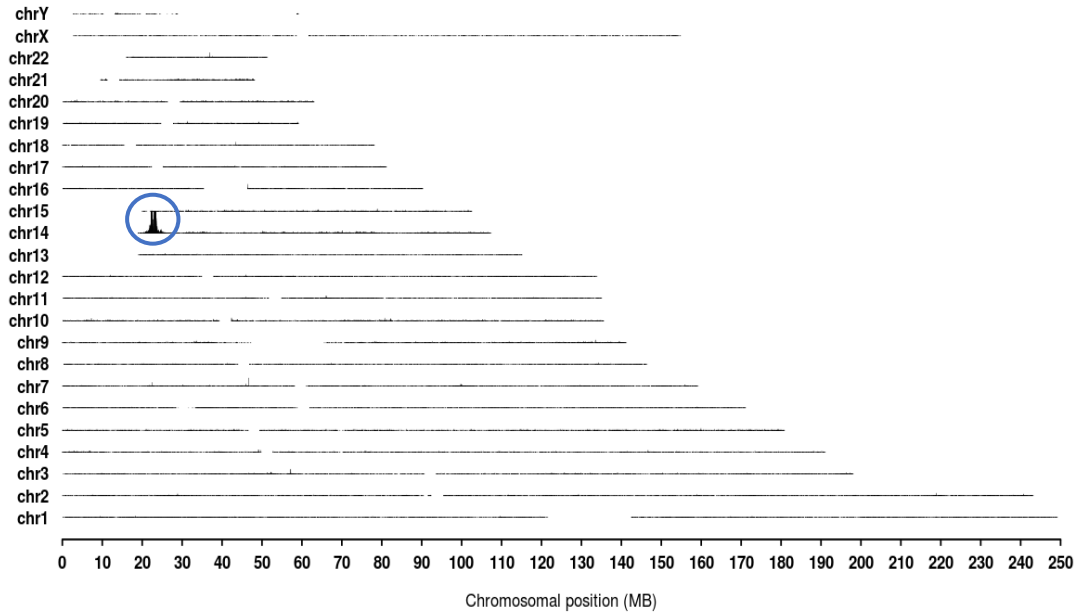

Figure 2: TLA sequence coverage across the human genome using primer set 1. The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis. The identified integration site is encircled in blue.

As shown in figure 2 and figure 3, the vector has integrated in chromosome 14. The same integration site was identified with the primer set 2.

Locus-wide coverage

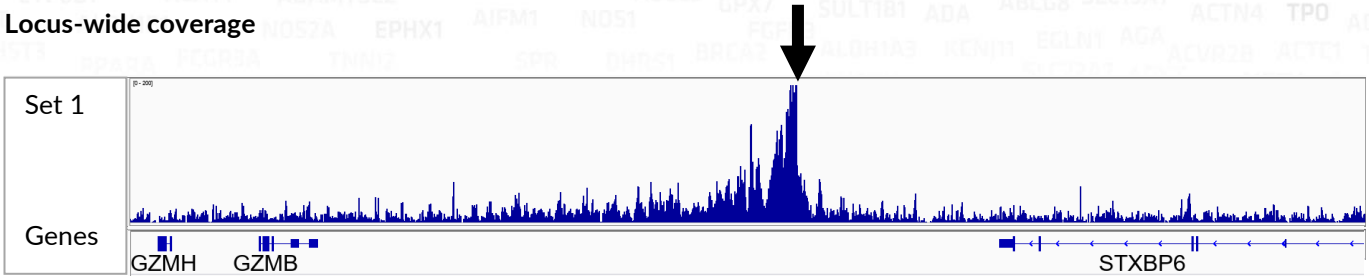

Figure 3: TLA sequence coverage (in blue) across the vector integration locus, human chr14:24,600,000-24,900,000. The black arrow indicates the location of the breakpoint sequences. Y-axis is limited to 200x. Same results were obtained with primer set 2.

Breakpoint sequences
The following breakpoint sequences were identified marking the vector integration:

5' integration site:
chr14:24,733,900 (tail) fused to Vector: 4 (head) with 5 inserted bases
AAAGAAAAAAATCACACATTGAAAATCACTCTGAGACACATTCATCTTTTTCGAAAACATACAGCCATTTCTGTACA
AAATACAGCCACATAGCAAGCTCAATTCTACCACTGAACAAGCTACTTTTTTACCCTGTTGTACACACTGTGATAGC
ATGGTCGAATCGATGCTAAGCTTCGTAAATCGATATCGATCGTAGCTATGCTAGGGTCGCC

3' integration site:
Vector: 9,527 (tail) fused to chr14:24,733,901 (head) with 3 bases homology
CACCATGGGTACGTACGTTATATCCCTGATCGTGCTCGTAGCTGCCTGCTAAGCTAGCTGATGCTGCCGCTTGATTA
TCATCTTAGCAGTCTTGATTATTAGTGTCAGTGGTAAAGCTTTGTACCTAGCTAAGAGCTTTTCTAGGAGAAGTAAG
CATGATACCTGCAGAAATTTTCCAAATGTTTGAGAGCCTAGTATGAGTTCAGGTATTGTGTC

The coverage profile in figure 3 shows that no genomic rearrangements have occurred in the region of the integration site.

From this data it is concluded that the vector has integrated in chr14:24,733,900-24,733,901. According to RefSeq, there are no genes annotated here.

Copy number estimation

In this sample, the coverage on the vector-side is 4-5 times higher than on the genome-side of the integration site. 1 integration site and 2 vector-vector breakpoints are found. The copy number is estimated to be 3-5 copies.

QC information

Internal project number:
Lab technician:
TLApp version:
Analysis:
QC-approval:
Version report:
Date:

Signed by: