Report

Genetic Characterization of a CHO Cell Line with a Transgene Integrated Using Random Integration

Prepared for:	Company name Company address
Customer name:	Name Position within a company email
Internal project number:	XXX
Quote number:	XXX
Version:	1
Date:	__-___-____

Goal

In this study, 1 transgenic CHO cell line sample with the vector XXX sequence was analyzed.

The aim of this analysis was to:

1. Study the vector integrity:
    - Determine the presence of sequence variants and their allele frequency.
    - Determine the presence of vector-vector breakpoints that represent concatemerization 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. (optional)

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 transgene & integration site TLA analyses & ddPCR_v3".

Summary

Sample	Vector Integrity	Integration site(s)	Structural variants at the integration site	Copy number estimation
Sample 1	6 sequence variants, 2 structural variants	chr3:169,680,259-169,680,260	none	3-5

Conclusion

In Sample 1, 3-5 copies of the vector have integrated in chromsome 3. 6 sequence variants and 2 structural variants are found within the integrated vector sequence.

TLA, sequencing and data mapping

Viable frozen CHO-K1 cells were used and processed according to the published 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 transgene & integration site TLA analyses & ddPCR_v3".

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	132	X
		FW	256	X
2	GOI	RV	2,745	X
		FW	3,186	X

The NGS reads were aligned to the vector sequence and host genome. The Chinese Hamster CriGri-PICRH1.0 genome assembly GCF_003668045.3 was used as host reference genome sequence.

Results Sample 1

Vector integrity

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

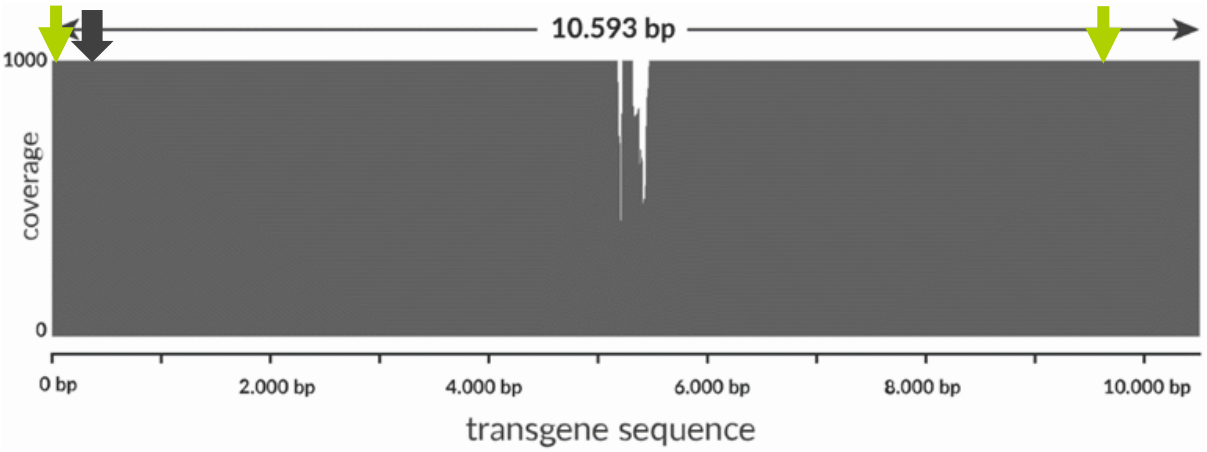

Figure 1: NGS sequencing coverage (in grey) across the vector with primer set 1. The black arrow indicates the primer location. The green arrows indicate the locations of the identified vector-genome breakpoint sequences (described below). The vector map is shown on the bottom. Y-axes are limited to 1000x. In an actual report the data of all primer sets will be presented.

High coverage is observed across the complete vector sequence Vector: 1-10,593, indicated by the grey areas in Figure 1, demonstrating that this sequence is integrated in this sample. 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. A total of 6 sequence variants were identified in the sample. Sequence variants near or at 100% mutation frequency were detected in all samples and most likely represent deviations present in the provided reference sequence of the vector before its introduction into the sample.

Table 2: Identified sequence variants

				Primer set 1		Primer set 2	
Region	**Position**	**Reference**	**Mutation**	**Coverage**	**%**	**Coverage**	**%**
Amp	141	A	C	21,254	20	788	25
GOI	1,013	T	-4AGTT	1,881	100	7,501	100
GOI	2,956	T	C	1,278	27	34,122	21
GOI	5,698	A	+1G	751	18	2,221	15
Backbone	9,487	T	G	1,358	100	1,523	99
Backbone	10,037	G	A	2,145	20	854	20

Vector concatemerization and 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. In total, 2 structural variants were identified indicating concatemerization. Breakpoint site 2 is in close proximity to the linearization site at position #. Using TLA it is not possible to determine the exact order of (partial) copies and to confirm the presence of at least one complete copy. In samples with multiple vector copies the number of vector-vector junctions may be underestimated.

Table 3: Vector-vector breakpoints

Breakpoint		Vector	Vector		Orientation of the breakpoint	Homology	Insert
1	→	6,945	7,657	→	tail to head	1	-
2	→	14	5,051	←	tail to tail	-	2

Integration sites

Whole genome coverage plot

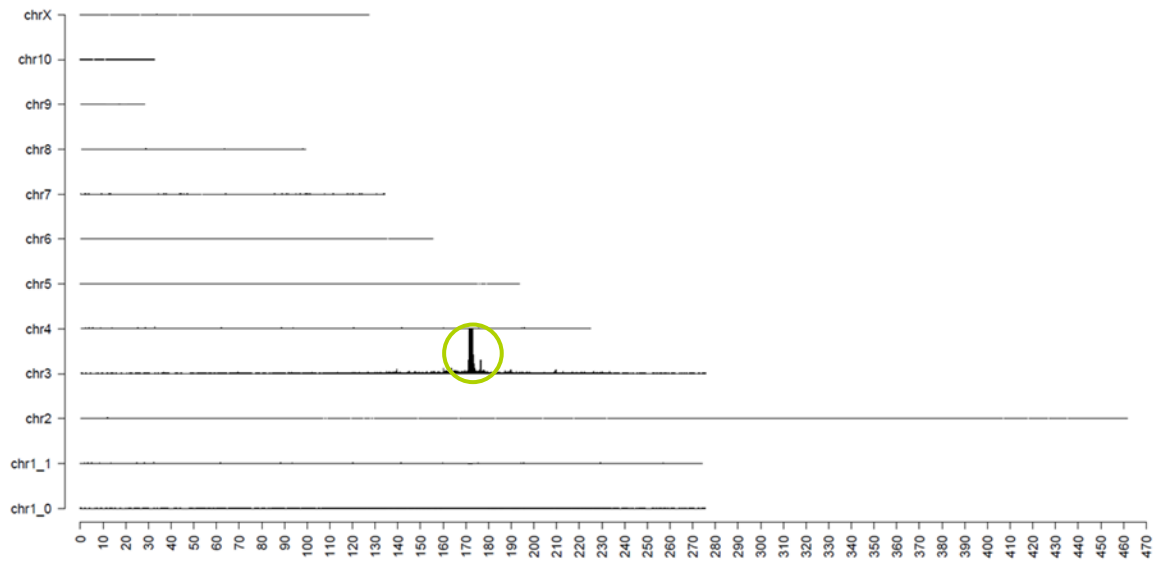

Figure 2: TLA sequence coverage across the Chinese Hamster genome using primer set 1. The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis. Identified integration site is encircled in green.

As shown in Figure 2, the vector has integrated on chromosome 3. Similar results were obtained with primer set 2.

Locus-wide coverage

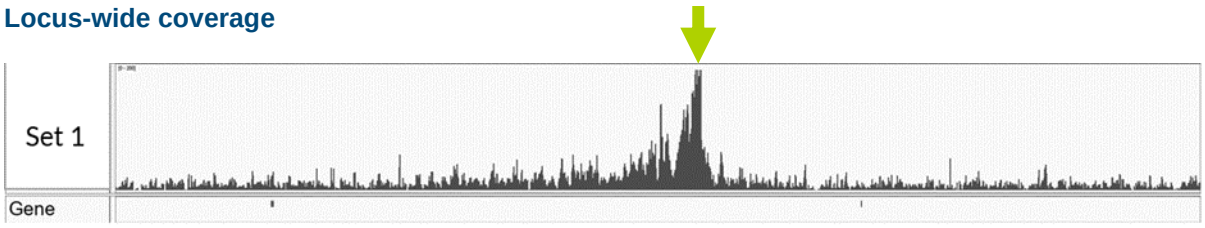

Figure 3: TLA sequence coverage (in grey) across the vector integration locus, Chinese Hamster chr3:169,530,000-169,830,000. The green arrow indicates the location of the breakpoint sequences. Y-axes are limited to 200x. In an actual report the data of all primer sets will be presented.

Coverage is observed across the vector integration site as shown in Figure 3.

Breakpoint sequences

The following breakpoint sequences were identified marking the vector integration:

5' integration site:
chr3:169,680,259 (tail) fused to Vector:4 (head) with 5 inserted bases
ATTGCACGTACGTACGTTTGGCAAACACTGTGCCTCGACTGCCGTCGGCGTAACGTCAGCTAGT
TTACCCTGTTGTACACACTGTGATAGGATGGTCGAATCGATGCTAAGCTTCGTAAATCGATATCG
ATCGTAGCTATGCTAGGGTCGCC

3' integration site:
Vector:9,527 (tail) fused to chr3:169,680,260 (head) with 3 homologous bases
CACTATGGGTACGTACGTTATATCCCTGATCGTGCTCGTAGCTGCCTGCTAAGCTAGCTGATGCT
GCCGCTTGTTGTACACTTAGGACTGTGATAGCTACGTCGTAAGCTGCTCGATGCTAGATCGCTAG
CGGCGGCTAGCTAGTGGCTGAGT

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 at chr3: 169,680,259 - 169,680,260 as shown in Figure 4. According to RefSeq, there are no genes annotated here.

Figure 4: Schematic representation of the integration site.

Copy number estimation (optional)

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 junctions are found. The copy number is estimated to be 3-5 (partial) vector copies.

QC information

Sample and Study details

Sample receipt date	X
Condition of sample at receipt	X
Start date in the lab	X
Sequencing run	X
Date data analysis	X
Deviations from the protocol	X
TLApp version:	X

Study Personnel

Lab technician	X
Data Analyst	X
QC Analysis and Report	X

Quality control

The results are independently verified and reviewed and are an accurate and complete representation of the study. The scope of accreditation for ISO/IEC 17025:2017, accredited by the Dutch Accreditation Council RvA, Registration number L671, entails all analytical services including: determination of the integrity of the transgene vector sequence; determination of the vector integration site(s) and breakpoint sequences between the vector and genome, determination of the presence of structural variants surrounding the vector integration site(s), next generation sequencing (NGS) and bio-informatic data analysis. The copy number estimation is not included in the scope of this accreditation.

Scientific approval	X
Date	X
Signature	X