Report

Transgene analysis and integration site sequencing of 1 CHO cells samples containing Transposon PiggyBac vector

Prepared for:

Customer name:

Internal project number:
Quote number:

Version: 1
Date: 26-Jun-2025

Goal

In this study, 1 transgenic CHO sample with the vector XXXX 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).

Summary

Sample	Vector integrity	Number of integration site(s)	Copy number estimation (optional)	Notes
Sample 1	3 sequence variants, 2 structural variants	28	At least 30 (partial) copies	Backbone integration

Conclusion

In samples 1, 28 integration sites were observed. Backbone coverage was also observed. 3 sequence variants and 2 structural variants were seen in this sample.

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/VP	Direction	Binding position *2025-4988*	Sequence
1	GS	RV	725	X
		FW	650	X
2	CMV	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 set 1. Similar results were obtained with primer set 2.

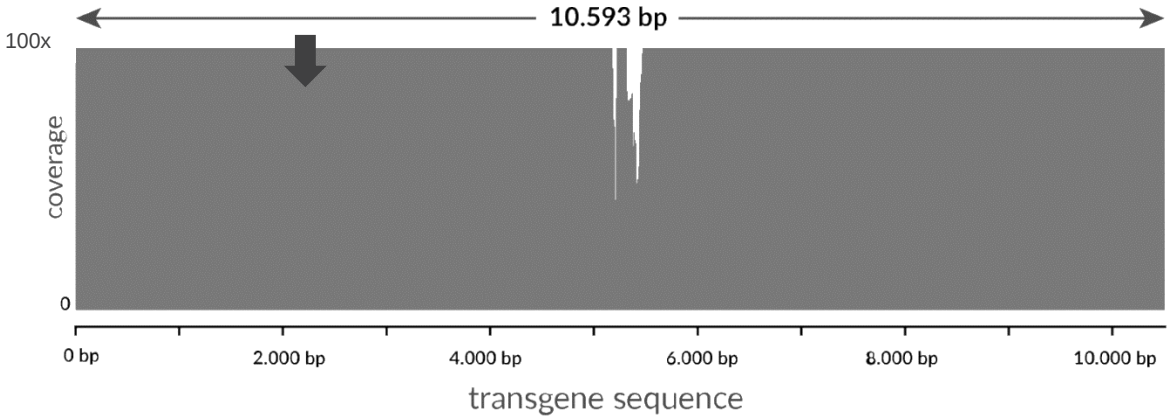

Figure 1: NGS sequencing coverage (in grey) across the vector. Black arrows indicate the primer locations.. Y-axes are limited to 100x. In an actual report the data of all primer sets will be presented.

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. Coverage outside the ITRs indicates the backbone has integrated in at least one location in the genome. Local dips in coverage are due to regions which 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. Please note that using the 5% filtering criteria, the reported allele frequencies for an individual sequence variant represent the fraction of all the occurrences of that variant among all vector copies integrated in all loci in the entire cell population.

Table 2: Identified sequence variants in the vector.

				Primer set 1		Primer set 2	
Region	**Position**	**Reference**	**Mutation**	**Coverage**	**%**	**Coverage**	**%**
CMV	141	A	C	21,254	20	788	25
Not annotated	1,013	A	+1G	751	18	2,221	15
KAN	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. 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 in the vector.

Breakpoint		**Vector**	**Vector**		**Orientation of the breakpoint**	**Homology**	**Insert**
1	→	1,457	7,657	→	tail to head	-	1
2	→	2,500	9,256	←	tail to tail	4	-

Integration sites

Whole genome coverage plot

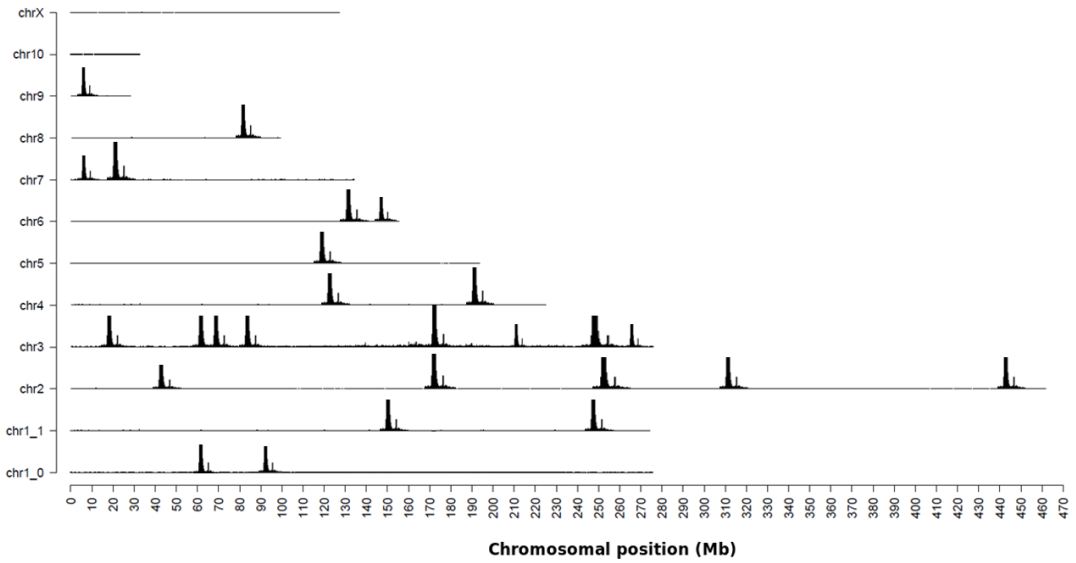

Figure 2: TLA sequence coverage across the Chinese hamster genome using primer set 1. The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis.

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

Breakpoint sequences

The identified integration sites are listed in Table 4. In the accompanying Excel tables, the sequences, frequencies and gene annotation at the breakpoints are presented. Most integration site breakpoints are identified at the expected ITR locations. A random integration event has occurred on chromosome 2 (breakpoint 6). Vector concatemerization most likely has occurred.

Table 4: Integration sites

Break point	Vector		Chromosome			Orientation of the breakpoint	Homology	Insert	Genomic structural variants at the integration site
1	→	9,256	chr1_0	65,102,039	→	tail to head	4	-	
	←	583	chr1_0	65,102,042	←	head to tail	4		
2	→	9,256	chr1_0	98,996,200	←	tail to tail	5	-	
	←	583	chr1_0	98,996,207	→	head to head	5	-	
3	→	9,256	chr1_1	150,700,514	→	tail to head	5	-	
	←	583	chr1_1	150,700,508	←	head to tail	4	-	
4	→	9,256	chr1_1	250,408,074	→	tail to head	-	-	
	←	579	chr1_1	250,408,073	←	head to tail	-	-	
5	←	583	chr2	44,323,798	←	head to tail	4	-	
	→	9,256	chr2	44,323,803	→	tail to head	4	-	
6[#]	←	1,250	chr2	179,287,078	←	head to tail	4	-	
	→	9,256	chr2	179,287,083	→	tail to head	4	-	
7	→	9,256	chr2	255,303,426	←	tail to tail	4	-	
	←	583	chr2	255,303,431	→	head to head	4	-	
8	←	583	chr2	315,900,623	←	head to tail	4	-	

	→	9,256	chr2	315,900,629	→	tail to head	5	-	
9	→	9,256	chr2	444,590,878	→	tail to head	4	-	
	←	583	chr2	444,590,871	←	head to tail	6	-	
10	→	9,256	chr3	20,921,238	→	tail to head	4	-	
	←	583	chr3	20,921,233	←	head to tail	4	-	
11	←	583	chr3	65,816,955	←	head to tail	4	-	
	→	9,256	chr3	65,816,960	→	tail to head	4	-	
12	←	583	chr3	75,428,314	←	head to tail	4	-	
	→	9,256	chr3	75,428,319	→	tail to head	4	-	
13	←	583	chr3	85,891,055	→	head to head	-	-	
	→	9,256	chr3	85,891,055	←	tail to tail	3	-	
14	→	9,256	chr3	170,208,201	←	tail to tail	4	-	
	←	585	chr3	170,208,206	→	head to head	6	-	
15	←	583	chr3	213,761,150	←	head to tail	4	-	
	→	9,256	chr3	213,761,155	→	tail to head	4	-	
16	→	9,256	chr3	255,204,534	←	tail to tail	4	-	
	←	583	chr3	255,204,540	→	head to head	5	-	
17	←	9,263	chr3	255,641,056	→	head to head	5	-	
	←	583	chr3	255,641,056	←	head to tail	-	-	
18	→	9,256	chr3	275,012,875	→	tail to head	1	-	
	←	583	chr3	275,012,875	←	head to tail	2	-	
19	→	9,256	chr4	125,860,215	←	tail to tail	6	-	
	←	583	chr4	125,860,222	→	head to head	4	-	
20	←	584	chr4	205,159,441	←	head to tail	5	-	
	←	9,256	chr4	205,159,446	→	tail to head	4	-	
21	→	579	chr5	120,097,581	→	head to tail	1	-	
	←	9,256	chr5	120,097,583	←	tail to head	-	-	
22	→	9,256	chr6	136,164,486	→	tail to tail	4	-	
	→	583	chr6	136,164,491	←	head to head	4	-	
23	←	9,256	chr6	155,241,475	→	tail to tail	4	-	
	→	584	chr6	155,241,480	←	head to head	5	-	
24	←	9,256	chr7	2,230,999	→	tail to tail	4	-	
	→	583	chr7	2,230,004	←	head to head	4	-	
25	←	9,256	chr7	25,809,621	→	tail to tail	4	-	
	→	579	chr7	25,812,758	←	head to tail	-	-	
26	←	9,256	chr8	93,163,509	←	tail to head	-	-	
	→	583	chr8	93,163,509	→	head to tail	3	-	
27	←	583	chr9	15,622,618	←	head to tail	4	-	
	←	9,256	chr9	15,622,623	←	tail to head	4	-	
28	→	9,256	chr9	14,787,373	→	tail to tail	3	-	
	→	583	chr9	14,622,618	←	head to tail	4		

Copy number estimation (optional)

In this sample, the copy number is estimated based on the number of integration sites and the number of structural variants identified for the specific vector. 28 integration sites and 2 vector-vector junctions as well as backbone integration are found in this sample. Backbone integration indicates that multiple (partial) vector copies are found at some of the identified integration sites. The copy number is estimated to be at least 30 (partial) vector copies. The numbers provided here are the minimum expected copy numbers.

QC information

Sample and Study details

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

Study Personnel

Lab technician
Data Analyst
QC Analysis and Report

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
Date
Signature