Transposon PiggyBac Example Report

Prepared for:
Company name
Company address

Customer representative:
Name
Job title
Email

Report: XXX

Quote number: XXX

Version: X

Date: 29-Apr-24

1) Goal

In this study, 1 transgenic CHO sample with the vector xxxxx sequence was analyzed.
The aim of this analysis was to:

1. Study the vector integrity;
    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 transgene & integration site TLA analyses & ddPCR.

2) Summary

Sample	Vector Integrity	Integration Site(s)	Structural variants at the integration site	Copy number estimation
Sample 1	3 sequence variants, 2 structural variants	28	At least 30 (partial) copies	Backbone integration

3) 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.

4) Methods

TLA, sequencing and data mapping

Viable frozen CHO-K1 cells were used and processed according to Solvias' 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.

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/Viewpoint	Direction	Binding position	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.

5) 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.

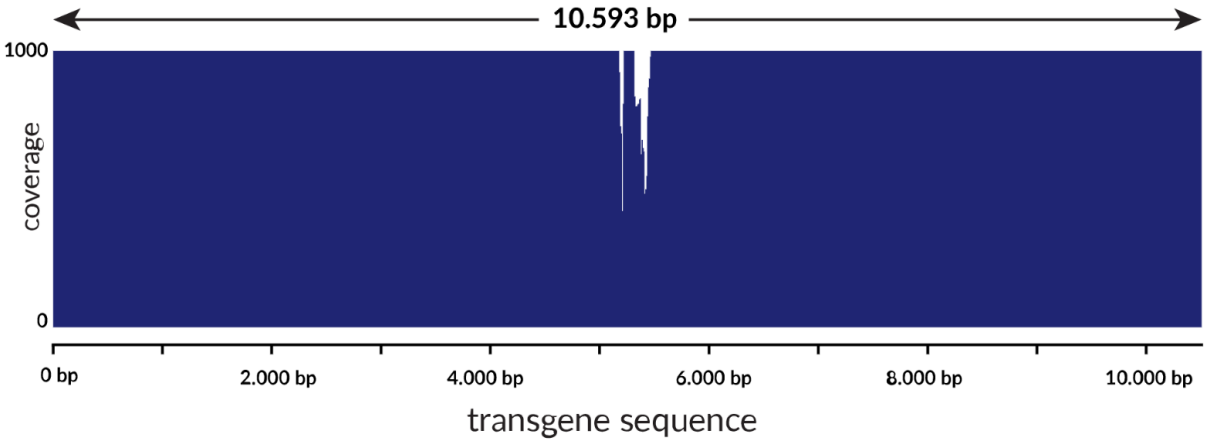

Figure 1: NGS sequencing coverage across the vector with primer set 1. The black arrow indicates the primer location. Y-axis is limited to 1000x. In an actual report the data of all primer sets will be presented.

High coverage is observed across the complete vector sequence Vector: 1-10,593. Coverage observed between 1-583 and 9,256-10,593 indicates the backbone has integrated in at least one location 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 3 sequence variants were identified in the sample.

Table 2: Identified sequence variants

Region	Position	Ref	Mut	Primer set 1		Primer set 2	
				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

'+' indicates an insertion

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. Note, the breakpoint 2 indicates that the vector has recombined with itself using ITRs and the TTAA sequence present at position 9,256.

Table 3: Vector-vector breakpoints

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

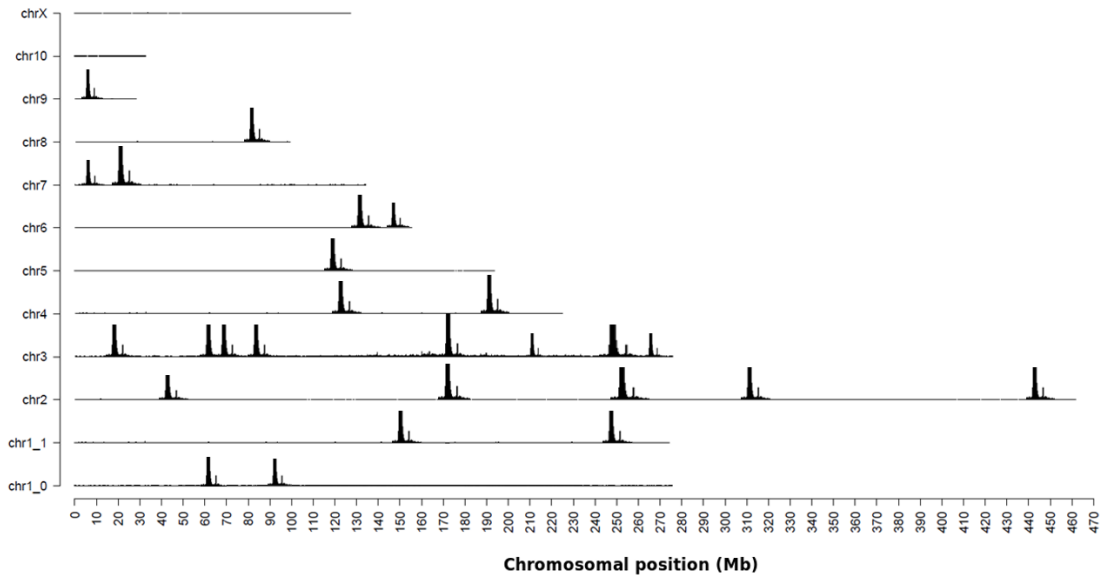

Figure 2: TLA sequence coverage across the Chinese Hamster genome using primer set 1. The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis.

As shown in figure 2, the vector has integrated on multiple chromosomes. Similar results were obtained with primer set 2. The identified integration sites are listed in Table 4. 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

at this random integration site. In the accompanying excel tables the sequences and frequencies of the breakpoints are presented.

Break-point	Vector		Chromosome			Orientation of the breakpoint	Hom	Insert	Structural variants at the integration site	Gene at integration Site
1	→	9,256	chr1_0	65,102,039	→	tail to head	4	-		*Pbxip1*
	←	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	-		*Fam184a*
	←	583	chr2	255,303,431	→	head to head	4	-		
8	←	583	chr2	315,900,623	←	head to tail	4	-		*XM_027401667.2*
	→	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	-		*XM_027409106.2*
	→	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	-	-		*XM_027410136.2*
	→	9,256	chr3	85,891,055	←	tail to tail	3	-		
14	→	9,256	chr3	170,208,201	←	tail to tail	4	-		*Zswim4*
	←	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	-		*Rpp40*
	←	583	chr3	255,641,056	←	head to tail	-	-		
18	→	9,256	chr3	275,012,875	→	tail to head	1	-	4.8kb deletion	-
	←	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	-		*Rasa2*
	←	9,256	chr4	205,159,446	→	tail to head	4	-		
21	→	579	chr5	120,097,581	→	head to tail	1	-		*Rbm25*
	←	9,256	chr5	120,097,583	←	tail to head	-	-		
22	→	9,256	chr6	136,164,486	→	tail to tail	4	-		*Ehmt1*
	→	583	chr6	136,164,491	←	head to head	4	-		
23	←	9,256	chr6	155,241,475	→	tail to tail	4	-		*Otog*
	→	584	chr6	155,241,480	←	head to head	5	-		
24	←	9,256	chr7	2,230,999	→	tail to tail	4	-		*Ephx4*
	→	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	-	-		*Ube3c*
	→	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

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.

3) QC Information

Sample and study details

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

Study personnel

Lab technician
Lab technician qPCR
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. TLA processing of cells, NGS sequencing, and data analysis (except for copy number) are ISO/IEC 17025:2017 accredited by the Dutch Accreditation Council RvA, Registration number L671.

Scientific approval

Date

Signature