SCIENTIFIC SERVICES

Transgene Mapping Analysis Report

Customer: Eric Minikel

Date: March 2020

Project Number: 17898

Cergentis B.V.
Yalelaan 62
3584 CM Utrecht
The Netherlands
+31 30 760 1636
info@cergentis.com
cergentis.com

Taconic Biosciences, GmbH
Marie-Curie-Str. 10
51377 Leverkusen
Germany
+45 70 23 04 05
info@taconic.com
taconic.com

Goal

In this study, 1 transgenic mouse sample with the transgene (TG): huPRNP_mod_RP11-715K24 sequence was analysed.
The aim of this analysis was to:

1. Identify genetic alterations in the transgene:
    1) Single Nucleotide Variants (SNVs) and their allele frequency.
    2) TG-TG fusions that represent concatemers of individual copies of the transgene and/or structural rearrangements in a transgene sequence.
2. Identify transgene integration site(s) and breakpoint sequences between transgene and genome.
3. Assess the presence of structural variants surrounding the transgene integration site(s),
4. Estimate the copy number of the transgene.

Summary

Sample	Integration site(s)	Notes
59003,2	mouse chr12 (expected at 70,908,646)	- 18 SNVs and 3 TG-TG fusions were detected - 2 genomic deletions near the integration site

Conclusion

The analysis of sample 59003,2 showed an integration site on chromosome 12. The exact breakpoint between TG and genome could not be identified in the current dataset. The integration was accompanied by two genomic deletions of 413 kb and 62 kb, in which exons 1-8 of the *Tmx1* gene and exon 2-14 of the *Frmd6* gene are deleted.

Methods

TLA AND SEQUENCING

Viable frozen mouse spleen cells were used and processed according to Cergentis' TLA protocol (de Vree et al. Nat Biotechnol. Oct 2014).

2 primer sets were designed on the transgene (Table 1). The primer sets were used in individual TLA amplifications. PCR products were purified and library prepped using the Illumina Nextera flex protocol and sequenced on an Illumina sequencer.

Table 1: Primers used in TLA analysis

Primer set	Name/View point	Direction	Binding position	Sequence
1	14 KB	RV	14123	CCTTTCTTGCTTTATTAGATCCA
		FW	14279	TGCGGATCATTCACTGGTA
2	40 KB	RV	40842	TCTTTGACTGGTAGGATCCT
		FW	41495	GCAACTGTGTAGAATACTGC

ALIGNMENT OF SEQUENCING READS

Reads were mapped using BWA-SW (Li et al. Bioinformatics, 2010 [PMID: 20080505]), version 0.7.15-r1140, settings bwasw -b 7.

The NGS reads were aligned to the TG sequence and host genome. The mouse mm10 genome was used as host reference genome sequence.

SINGLE NUCLEOTIDE VARIANT DETECTION

The presence of SNVs is determined using samtools mpileup (samtools version 1.3.1) (Li et al. Bioinformatics, Jun 2009 [PMID: 19505943], Li et al. Bioinformatics, Nov 2011 [PMID: 21903627]).

SNVs are reported that meet the following criteria:

- allele frequency (relative amount of reads with the variant, compared to total coverage on the variant position) of at least 20%,
- the variant is present in the data of both primer-sets,
- for at least one of the primer-sets the coverage is >=30X,
- the variant is identified in both forward and reverse aligning sequencing reads,
- low frequency variants (between 5-20% mutant allele frequency) are not found with similar frequencies in a negative control (if included).

TG-TG FUSION DETECTION

Fusion sequences consisting of two parts of the TG, are identified using a proprietary Cergentis script. Fusions resulting from the TLA procedure itself are recognized by the restriction enzyme-specific sequence at the junction site and removed.

TG-TG fusions are reported that meet the following criteria:

- the TG-TG fusion is present in >1% of the reads at the position of the fusion,
- the TG-TG fusion is observed in data of both primer-sets, unless the data provides a clear explanation why the fusion is not found in one of the data sets,
- the TG-TG fusion is not present in negative control sample(s) (if included),
- visual inspection of the TG-TG fusions in an NGS data browser is performed to remove fusions that are sequencing artefacts, e.g. fusions found at hairpin structures or low-complexity regions.

INTEGRATION SITE DETECTION

Integration sites are detected based on a) coverage peak(s) in the genome and b) the identification of fusion-reads between the TG sequence and host genome.

Results 59003,2

TRANSGENE SEQUENCING COVERAGE

Figure 1 depicts the NGS coverage across the TG sequence. Coverage is defined as the number of NGS reads that cover a locus (in this case the transgene sequence).

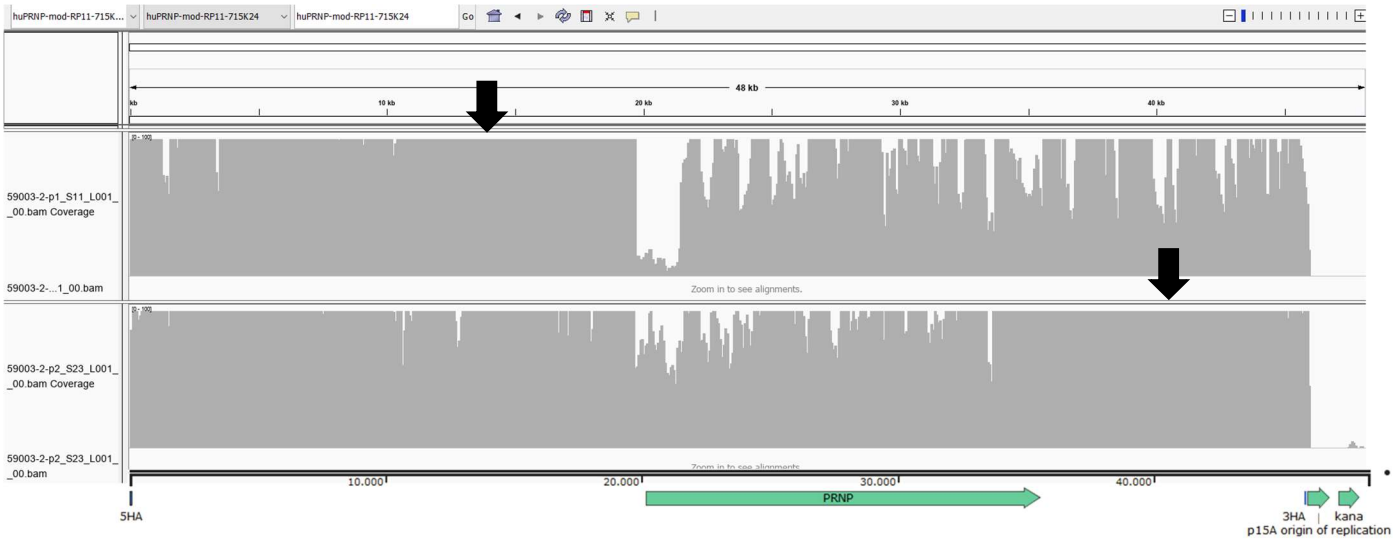


Figure 1: NGS sequencing coverage across the TG with primer set 1 (top) and primer set 2 (bottom). Black arrows indicate primer location. The vector map is shown on the bottom. Y-axis is limited to 100x.

Good coverage is observed across the TG sequence TG:2- 45,967. No coverage is observed in TG:45,968-48,422, which is the backbone between the annotated HA's.

Single Nucleotide Variants

SNVs were determined, using the criteria as described in the methods section. Detected SNVs are presented in table 2. Table 2A shows SNVs at or near 100% mutation frequency that are most likely deviations from the provided reference sequence before introduction into the sample. Table 2B shows SNVs introduced into the sample after integration.

Table 2A. Identified SNVs.

'-' indicates a deletion.

			Primer set 1		Primer set 2	
Position	**Reference**	**Mutation**	**Coverage**	**%**	**Coverage**	**%**
895	C	T	115	100	104	100
1921	A	G	190	100	215	100
2659	A	G	64	98	37	100
3574	T	C	249	100	175	99
4775	G	A	191	100	95	98
4866	A	G	179	99	106	100
6955	C	G	163	99	100	100
7968	G	A	125	100	116	100

9143	C	T	87	100	88	100
11821	A	-3TTG	195	100	114	100
18645	A	G	217	100	204	100
21111	G	T	8	100	51	98
21573	T	-1A	65	97	78	97
35890	C	T	169	96	179	97
42801	C	G	112	100	167	100

Table 2B. Identified SNVs.

			Primer set 1		Primer set 2	
Position	**Reference**	**Mutation**	**Coverage**	**%**	**Coverage**	**%**
277	C	A	166	20	87	26
17234	T	C	524	23	191	16
44766	C	-1A	79	42	152	57

Structural variation

The presence of TG-TG fusions was assessed as described in the methods section. A total of 3 TG-TG fusions were found within the TG indicating structural variations (number 1 and 2) and concatemerization (number 3) of the TG sequence:

1. TG:23,008 (tail) fused to TG:31,459 (tail) with 2 inserted bases

TCACAGGTTTTAATTTTTTAGATGAAATGGACCCACAGTTTTCTGTAAGAGAAAGGAGAGATTGTTATATTTGCTATTTGGAGCAGTTACAGCTACTTGAGATATTATGTTATTTCCTCTGTGCTTATTATTCATG

2. TG:24,251 (head) fused to TG:1,886 (head) with 1 homologous base

AAAGCCTTCTAGCCCATAAAACACATTACAACTTGAAGCTTAATTAGTTACATCGGGAAGGGGCAAATGTCAAACAACCCCCGACCTTCACCCCATAATGTCCCTTGTCTCTTCTGAGCATAAAGAAGACTGAAAGAAGAATGTAATAAAA

3. TG:45,959 (tail) fused to TG:10 (head) with 8 homologous bases

TACGGGAATTACCCCAATGGTGCTATTGCGGCCGCTTTAAGAGTCATGCAGCTGGAGACAGATTCTGACCCTCCCCAAATTGCTCCTGGGGATAATATCA

INTEGRATION SITES

Integration sites were determined as described in the methods section.

Whole genome coverage plot

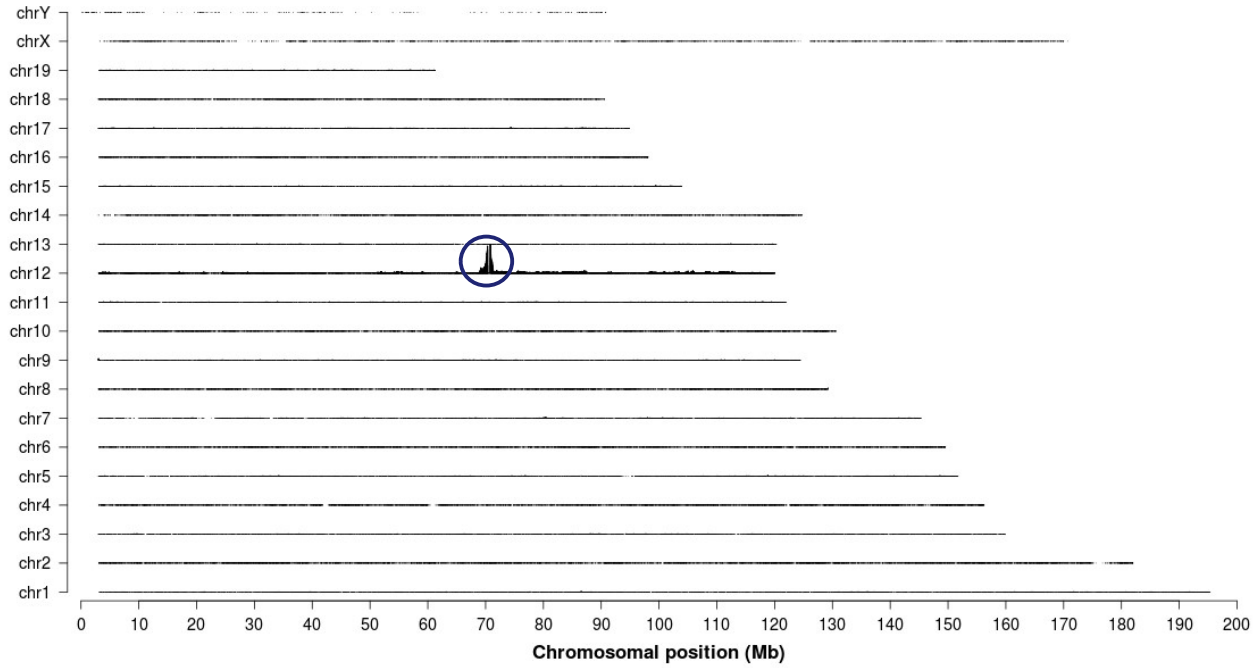


Figure 2: TLA sequence coverage across the mouse genome using primer set 1. The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis. Similar results were obtained with primer set 2.

As shown in figure 2, the TG has integrated in chromosome 12.

Locus-wide coverage

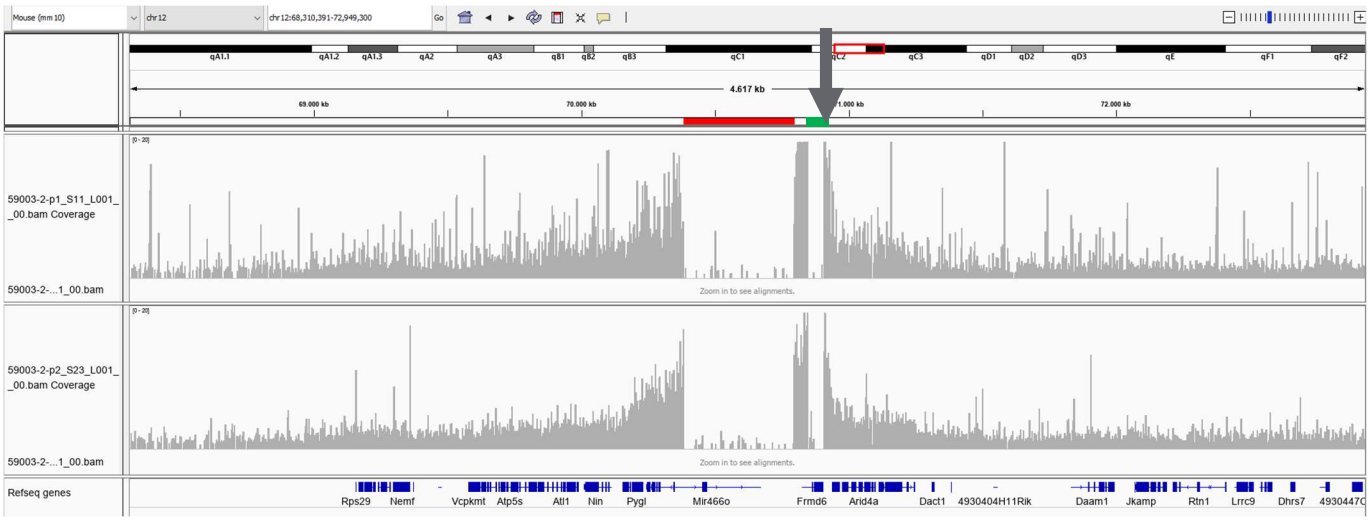


Figure 3: TLA sequence coverage across the TG integration locus; mouse chromosome chr12:68,310,391-72,949,300. Sequence coverage (in grey) generated with primer set 1 (top) and primer set 2 (bottom). The grey arrow points toward the breakpoint sequence representing the potential integration site. The red and green bars indicate genomic deletions. Y-axis is limited to 20x.

Breakpoint sequence

Both the 3' and the 5' integration site could not be identified with the current data set.

A fusion read to an unknown sequence was found which could indicate the integration site, but because of the presence of a CATG site close to the breakpoint, the origin of the fusion partner could not be identified. Further studies could be performed to gain more insight.

Unknown fused to chr12:70,908,646 (head)
CATGTTTGGGGCTATTTGGTAGGGACTCTCCTATAGGTCATTGCCATTTTTTCATTCCAGTCA

The coverage profile in figure 3 shows that two genomic deletions occurred near the integration site, which are marked by the following fusion reads:

A genomic deletion of 413 kb (red bar in Fig. 3) leading to the deletion of exons 1-8 of the *Tmx1* gene.
chr12:70,383,100 (tail) fused to chr12:70,796,163 (head)
GGCGCACACCTTCAATCCCAGCACTCGGGAGGCAGAGGCAGAGGCAGGTGGATTTCTGAGTTCGAGGCCAGCCTGGTCTACAAATGTTATTGCCTGTGTGCTGAGTAGTGCTGAAATTTGAAAATCCACTTTTTCTCATTTGTTACACTTG

A genomic deletion of 62 kb (green bar in Fig. 3) leading to the deletion of exon 2-14 of the *Frmd6* gene.
chr12:70,847,021 (tail) fused to chr12:70,908,598 (head)
TCCTAGAAATTTGTTCTGTGTGATAGTGGTGTATTCAGATAGCCTGTTTCTCAAGTGAGCTCTGGGGGTTTGTCCTTTGAGACATCTGAAGGTGAGGGCGTGGCTCCCAGGTAGGG

From this data it is concluded that the TG might have integrated in mouse chr12:70,908,646.

COPY NUMBER

An exact copy number cannot be determined using TLA. However, an estimation can be made based on the number of integration sites, number of fusion reads and the ratio of coverage on the TG and genome integration site.

Because the exact breakpoint between TG and genome could not be identified, the copy number is based on the number of TG-TG fusions identified. A copy number of at least 3 copies is estimated.

Proposed genotyping assays

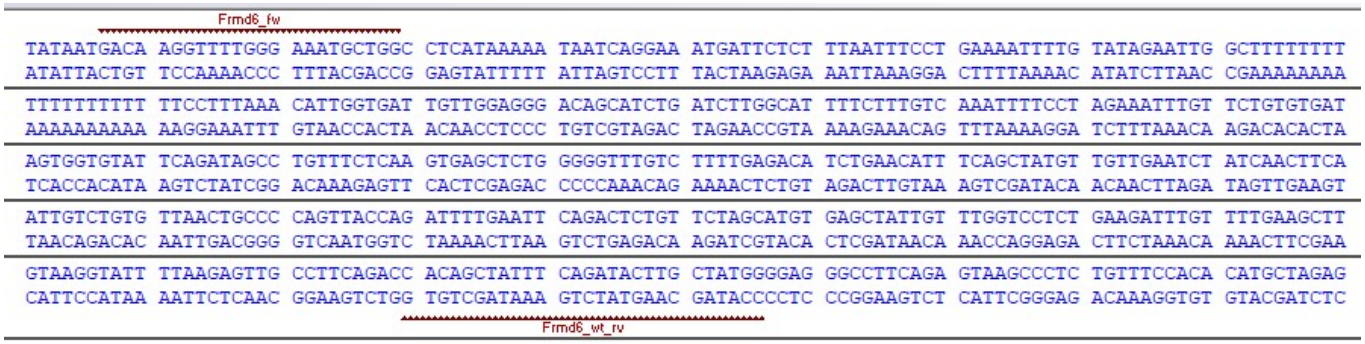
Because the exact breakpoint between TG and genome could not be identified, the genotyping assay is based on the accompanied genomic deletion within the *Frmd6* gene (exon 2-14).

***Frmd6* deletion exon 2-14:**

Frmd6_fw	GACAAGGTTTTGGGAAATGCTGG
Frmd6_tg_rv	TTTCACACCATCACTCCCTGG
Frmd6_wt_rv	CCATAGCAAGTATCTGAAATAGCTGTG

wt amplicon:	450 bp
tg amplicon:	388 bp

wt



***Frmd6* deletion /TG**