SCIENTIFIC SERVICES

Transgene Mapping Analysis Report

The Broad Institute - Eric Minikel

B6;129-Tg(Prnp)a20Cwe

11-May-2023

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 fragment (TG fragment) Tga20 was analyzed. The aim of this analysis was to:

1. Identify genetic alterations in the TG fragment:
    1) Sequence variants and their allele frequency.
    2) TG fragment-TG fragment breakpoints that represent concatemerization of multiple copies of the TG fragment and/or structural rearrangements in a TG sequence.
2. Identify TG integration site(s) and breakpoint sequences between TG and genome.
3. Assess the presence of structural variants surrounding the TG integration site(s).
4. Estimate the copy number of the TG.
5. Assess the co-integration of *E. coli* genomic sequences at the integration site.

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_v2".

Summary

Sample	Sequence and structural variants in TG fragment	Integration site(s)	Structural variants at the integration site	Copy number	Notes
TBD	6 sequence variants, 8 structural variants	chr17:46,761,775-46,762,856	1 kb deletion	8-70	No co-integration of *E. coli* genomic sequences

Conclusion

In sample TBD, the TG has integrated on chromosome 17, in the intron 1 of *Ptcra*. The integration event led to a 1 kb deletion of mouse genome sequence. Additionally, 6 sequence variants and 8 structural variants were detected in the integrated TG fragment sequence. Finally, no co-integration of *E. coli* genomic sequences was detected at the TG integration site.

TLA, sequencing and data mapping

Viable frozen mouse splenocytes 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_v2".

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

Table 1: Primers used in TLA analysis

Primer set	Name/View point	Direction	Binding position (chr2, mm10)	Sequence
1	exon 1	RV	131,909,564	AGAAAGTATATTGGCCATTGC
		FW	131,910,364	CAGTTTCATTCTCAACGTCG
2	exon 2	RV	131,911,782	CCTGTACTTTGACTCTCAGT
		FW	131,911,964	GATTGCTGGGATCGAACC

The NGS reads were aligned to the TG sequence and host genome. The mouse GRCm38/mm10 genome was used as host reference genome sequence.

Results TBD

TRANSGENE INTEGRITY

Figure 1 depicts the NGS coverage across the TG sequence using primer sets 1 and 2.

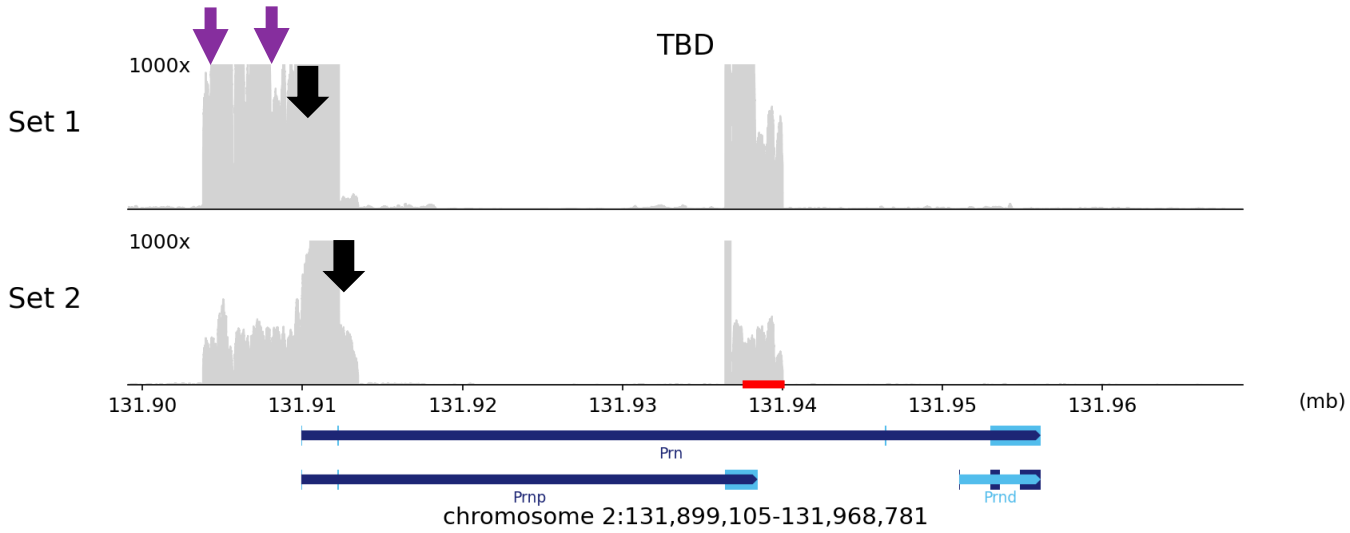


Figure 1: NGS sequencing coverage (in grey) across the mouse *Prnp* locus. Black arrows indicate the primer location. The purple arrows indicate the locations of the identified TG-genome breakpoint sequences (described below). The red bar indicates the provided reference sequence. The *Prnp* transcripts are shown on the bottom. Y-axes are limited to 1000x.

High coverage is observed across the *PrnP* locus sequence chr2:131,903,786-131,912,292 and chr2:131,936,420-131,940,019, indicating that the integrated TG is longer than the provided TG fragment (red bar in Figure 1). The following breakpoints have been identified at the borders of the coverage which represent the connection between the upstream and downstream regions with coverage (breakpoint 1) and concatemerization of the TG copies (breakpoints 2-7).

1) chr2:131,912,292 (tail) fused to chr2:131,936,420 (head)
CTGAAGCATTCTGCCTTCCTAGTGGTACCAGTCCAATTTAGGAGAGCCAAGCAGACTATCAGTCATCATGGCG
AACCTTGGCTACTGGCTGCTGGCCCTCTTTGTGACTATGTGGACTGATGTCGGCCTCTGCAAAAAGCGGCCAA
AGC

2) chr2:131,939,950 (tail) fused to chr2:131,903,786 (head) with 12 inserted bases
GTTCACCTGCTCCGTTGGCGCGGCCGCGACGGATCCAAAGGCAGCAAAAAGGCAGAGAGGGTGATACTGGG
CCTGGCTTAAGCATTTGAAACTTCAAAGCTCACCCCCAATTACACACTTCTTCCAACAAGTCCACACCTCCTAA
TTAG

3) chr2:131,939,493 (tail) fused to chr2:131,903,786 (head) with 4 inserted bases
AGGCTCAACCTTTTGGCCCAAAAGGCCACACTTGCAATTCACTTTGCATACCTGTGTCCATTGTAAGGGAAGG
CACGGGCTCATGGTGCAGGTGTCCATTGTGCGACGGATCCAAAGGCAGCAAAAAGGCAGAGAGGGTGATACT
GGGC

4) chr2:131,940,015 (tail) fused to chr2:131,904,511 (head) with 3 bp homology
ATGCCCACCTCCTGCAGGACTCAGTCAGTCAGTCAACCTATCTACCATG

5) chr2:131,940,016 (tail) fused to chr2:131,903,786 (head) with 7 inserted bases
GGGTACAACGGTAGAGGATGCTGAGGCATTGATGCCCACCTCCTGCAGGACTCAGTCAGTCAGCCGCGACG
GATCCAAAGGCAGCAAAAAGGCAGAGAGGGTGATACTGGGCCTGGCTTAAGCATTTGAAACTTCAAAGCTCAC
CCCCA

6) chr2:131,940,016 (tail) fused to chr2:131,903,797 (head) with 17 inserted bases
ACCTGCTCCGTTGGACAGGGTACAACGGTAGAGGATGCTGAGGCATTGATGCCCACCTCCTGCAGGACTCAG
TCAGTCAGATCCTGCAGGACTCAGTCAGCAAAAAGGCAGAGAGGGTGATACTGGGCCTGGCTTAAGC

7) chr2:131,940,019 (tail) fused to chr2:131,903,786 (head) with 6 inserted bases
TGATGCCCACCTCCTGCAGGACTCAGTCAGTCAGTCGCGACGGATCCAAAGGCAGCAAAAAGGCAGAGAGG
GTGATACTGGGCCTGGCTTAAGCATTTGAAACTTCAAAGCTCACCCCCAATTA

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

Sequence variants
Sequence variants were called only in the provided reference sequence. Detected sequence variants are presented in table 2. All sequence variants are near or at 100% mutation frequency (present in all reads at the specific location) and most likely represent deviations present in the provided reference sequence of the TG before its introduction into the sample. The identified sequence variants have been identified in the provided TG fragment sequence. However, they all correspond to the WT sequence of the mouse genome in the *Prnp* locus.

Table 2. Identified sequence variants

				Primer set 1		Primer set 2	
Region	Position	Reference	Mutation	Coverage	%	Coverage	%
Not annotated	897	G	C	349	100	294	100
Not annotated	1,014	T	A	332	100	230	100
Not annotated	1,117	T	-1C	409	95	303	97
Not annotated	1,172	T	+1C	320	94	263	94
Not annotated	1,195	A	G	322	100	266	99
Not annotated	1,823	G	C	375	100	254	100

TG concatemerization and structural variants
In addition to 7 structural variants described above, 1 structural variant was identified within the provided reference sequence. Intact reads were also found at the positions of this breakpoint and breakpoints 2-7 described above, indicating that (partial) TG sequences have concatemerized. A concatemer is formed prior to integration from the full length copies, and/or randomly fragmented vector copies. 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.

TG:1,931 (tail) fused to TG:2,177 (head)
AGGCCACACTTGCAATTCACTTTGCATACCTGTGTCCATTGTAAGGGAAGGCACGGGCTCATGGTGCAGGTGT
CCATTGTGAGAAGGCACAGGCTCATGGTGCAGGTGTCCAACAAGCCTAAAGACAAAGATCTGGGCAAGTGCT
TTTTCTTAAAGTCAGGGAGGAGTACACAGAAGGTAGTGGGGGATGGGGGTACTTGGGGGTGTCTTTC

Please note that due to the lack of a complete TG sequence, the number of structural variants could be underestimated.

INTEGRATION SITES

Whole genome coverage plot

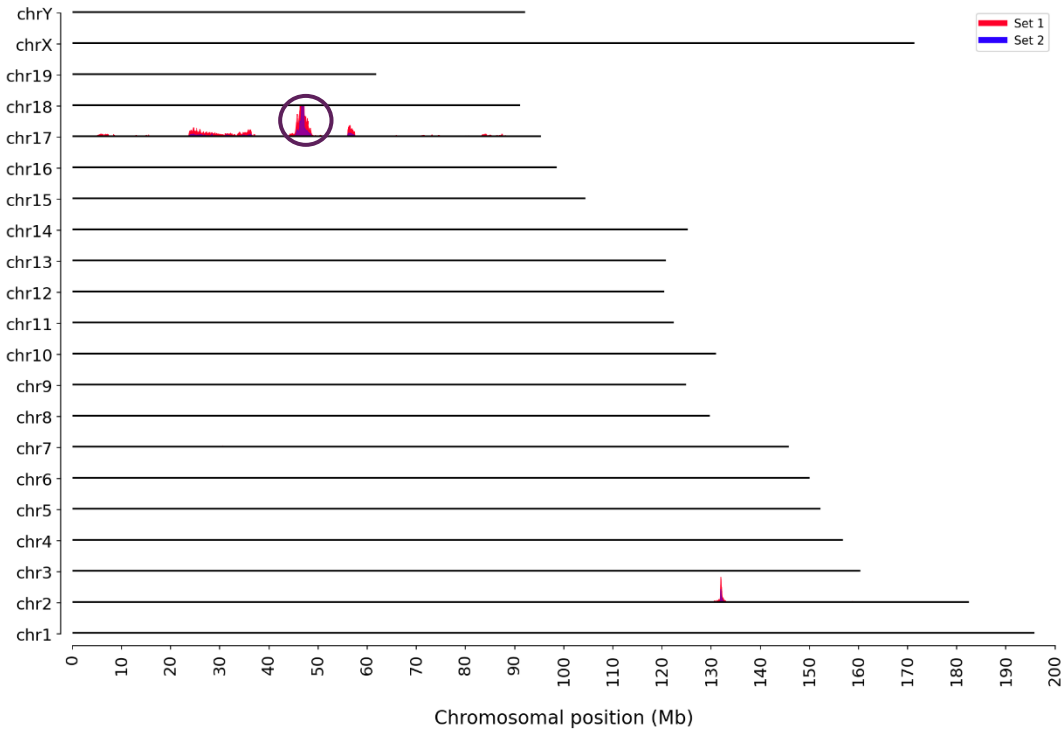


Figure 2: TLA sequence coverage across the mouse genome using primer set 1 (red) and set 2 (blue). The chromosomes are indicated on the y-axis, the chromosomal position on the x-axis. Identified integration site is encircled in purple.

As shown in figure 2, the TG has integrated in chromosome 17. The peak on chromosome 2 is due to homology of the TG with the genome (*Prnp* promoter).

Locus-wide coverage

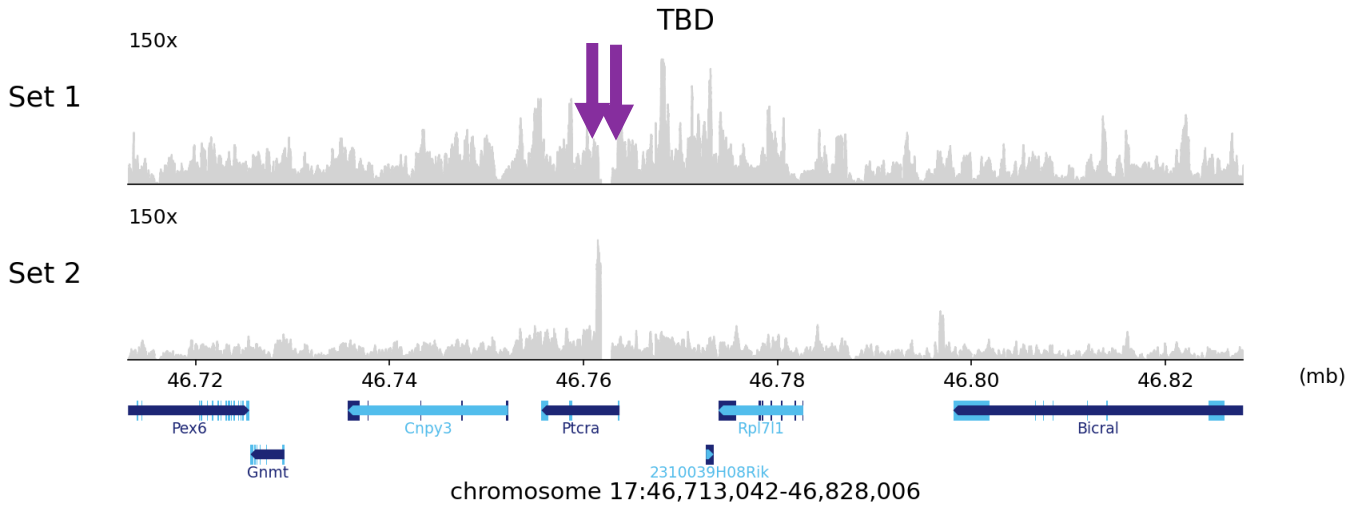


Figure 3: TLA sequence coverage (in grey) across the TG integration locus, mouse chr17:46,713,042-46,828,006. The purple arrows indicate the location of the breakpoint sequences. Y-axes are limited to 150x.

Coverage is observed across the TG integration site as shown in figure 3.

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

5' integration site:
chr17:46,761,775 (tail) fused to TG (homologous to chr2:131,909,936 (head))
ATCCCAGCGCCTACACACCCAACACTTCAATCTGTAATGAAATCCTATGCCCTCGTCTAGTGTGTCTGAAGACA
CACTCCCGGCTCCCCCGCGTTGTCGGATCAGCAGACCGATTCTGGGCGCTGCGTCGCATCGGTGGCAGGTA
AGCG

3' integration site:
TG (homologous to chr2:131,903,786 (head)) fused to chr17:46,762,856 (head) with 3 inserted bases
CTTGTTGGAAGAAGTGTGTAATTGGGGGTGAGCTTTGAAGTTTCAAATGCTTAAGCCAGGCCCAGTATCACCC
TCTCTGCCTTTTTGCTGCCTTTGGATCCGTCGCTTTGGAAATAATCTTTCTTTTTTTTAAGATTTATTTTATGTAT

The coverage profile in figure 3 shows that a genomic deletion has occurred in the region of the integration site. The 1 kb genomic sequence in between the two identified breakpoints is deleted.

From this data it is concluded that the vector has integrated at mouse chr17:46,761,775-46,762,856 as shown in Figure 4. According to the RefSeq, this is in intron 1 of *Ptcra*.

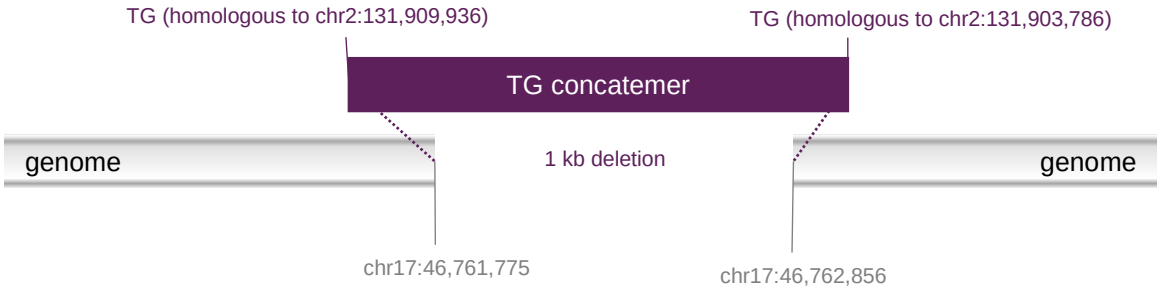


Figure 4: Schematic representation of the integration site.

COPY NUMBER

In this sample, the coverage on the TG-side (calculated based on the homology with chr2) is much higher than on the genome-side of the integration site (roughly 5-70 times). 1 integration sites and at least 7 TG-TG junctions indicating TG concatemerization are found. The copy number is estimated to be 8-70 (partial) TG copies.

Integration of *E.coli* genomic sequences

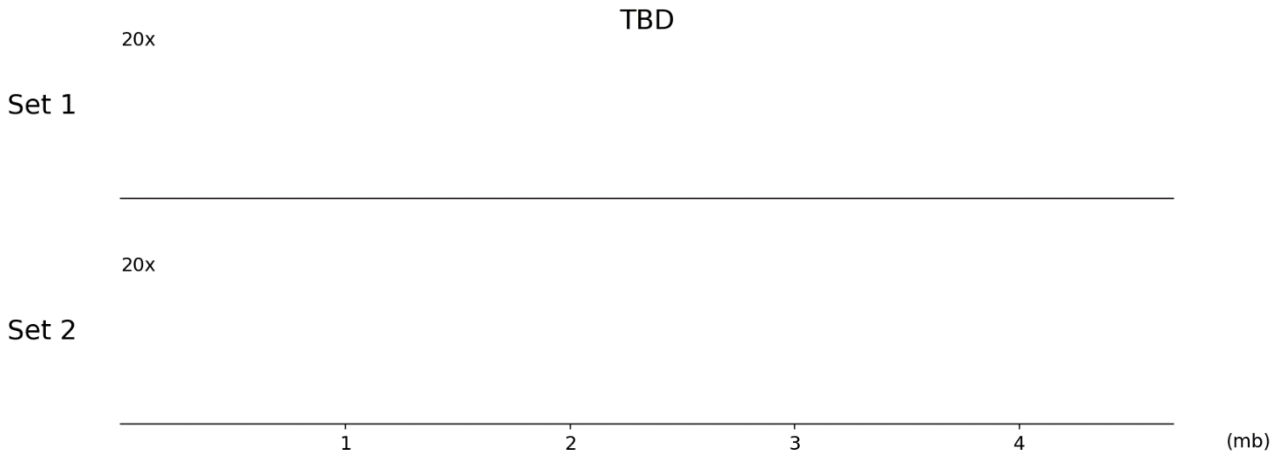


Figure 5: TLA sequence coverage (in grey) across the *E.coli* genome. Y-axes are limited to 20x.

As shown in figure 5, no coverage was observed across the *E.coli* chromosome indicating that no co-integration of the bacterial genomic sequences has occurred at the integration site in this sample.

QC INFORMATION

SAMPLE STUDY AND DETAILS

Sample receipt date	24-Mar-2023
Condition of sample at receipt	frozen
Start date in the lab	27-Mar-2023
Sequencing run	23-014
Date data analysis	24-Apr-2023
Deviations from the protocol	None
TLApp version:	1.5.1

STUDY PERSONNEL

Lab technician	Melinda Aprelia, BSc
Data Analyst	Andrea Conidi, PhD
QC Analysis and Report	Clara Esteban, PhD

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. Section “Proposed Genotyping Assay” of the report is provided by Taconic Biosciences, Inc. and is not covered by the abovementioned accreditation.

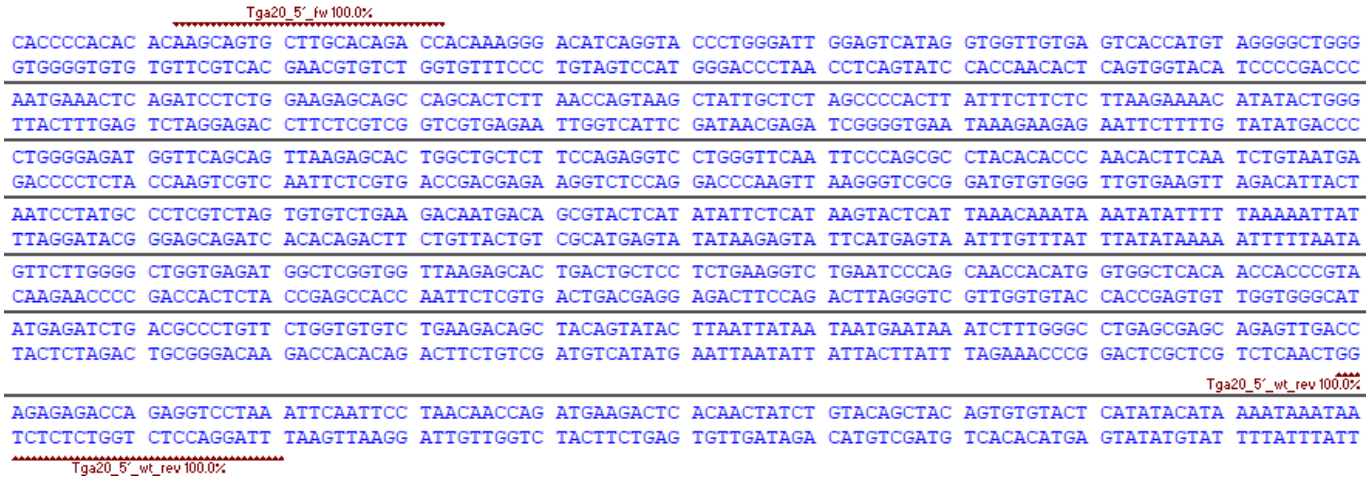
Scientific approval	Irina Sergeeva, PhD - Scientific Account Manager
Date	11-May-2023
Signature	

PROPOSED GENOTYPING ASSAYS

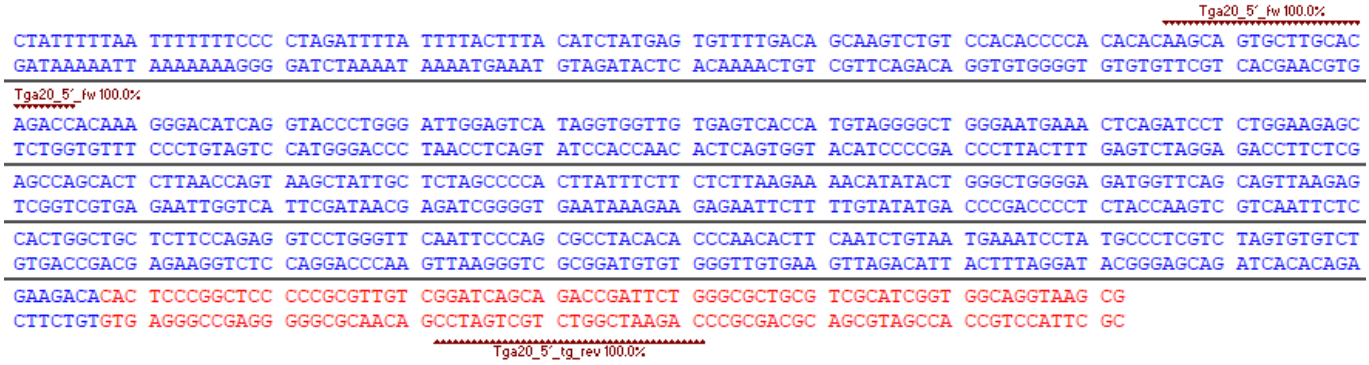
Tga20_5'_fw	AAGCAGTGCTTGCACAGACC
Tga20_5'_wt_rev	TTAGGACCTCTGGTCTCTCTGG
Tga20_5'_tg_rev	CAGAATCGGTCTGCTGATCC

wt amplicon: 608 bp
tg amplicon: 366 bp

wt



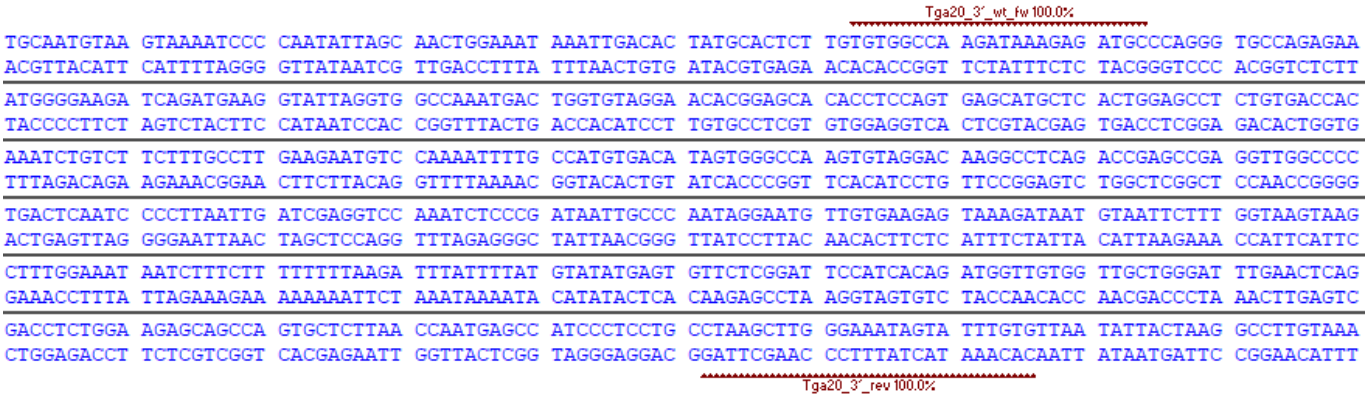
TG



Tga20_3'_wt_fw	TGTGGCCAAGATAAAGAGATGC
Tga20_3'_tg_fw	TTGGAAGAAGTGTGTAATTGGG
Tga20_3'_rev	CACAAATACTATTTCCCAAGCTTAG

wt amplicon: 514 bp
tg amplicon: 277 bp

wt



TG