Transgenic Res
https://doi.org/10.1007/s11248-019-00122-9

BRIEF COMMUNICATION

Transcription termination sequences support the expression of transgene product secreted with milk

Alexey Deykin · Maxim Tikhonov · Vladislav Kalmykov · Igor Korobko · Pavel Georgiev · Oksana Maksimenko

Received: 3 August 2018 / Accepted: 12 March 2019

Abstract Expression of the reporter gene in transgenic animals depends on the surrounding chromatin environment. Recent genome-wide studies have shown that, in mammals, the entire genome is transcribed. Transcription through a transgene often has a negative effect on the expression of a reporter gene. Here, we compared the ability of well-studied chicken chromatin insulator HS4 and bidirectional transcription terminators from the human genome to support high-level expression of the firefly luciferase gene (*Fluc*) under control of the previously characterized goat β-casein gene promoter. The insertion of HS4 or either of the two transcription terminators upstream of the promoter resulted in tenfold enhancement of *Fluc* expression in the mammary glands of transgenic mice. These results suggest that transcriptional terminators, similar to the HS4 insulator, can be used to improve the reporter gene expression in transgenic animals.

Keywords Bidirectional transcription termination · Transgenic animals · HS4 insulator · β-Casein gene promoter · Milk protein production

Introduction

A substantial problem in generating transgenic animals is that the transgene expression is dependent on integration site (Bishop and Smith 1989; Burkov et al. 2013; Choo et al. 1989; Dobie et al. 1996; Eszterhas et al. 2002; Suzuki et al. 2006). Several strategies have been developed to ensure position-independent transgene expression and, in particular, to protect the transgene from inactive surrounding chromatin and/or endogenous transcription at the locus of integration (for review, see (Houdebine 2018; Maksimenko et al. 2013), which may interfere with transgene transcription (Eszterhas et al. 2002). One of the best characterized and frequently used approaches to improve the transgene expression upon its random integration is relied on transgene flanking with the HS4 insulator derived from the chicken β-*globin* locus (Emery et al. 2002; Guglielmi et al. 2003). For example, inclusion of two tandemly repeated HS4 copies at the 5′ end of the construct resulted in improved transgene expression (Guglielmi et al. 2003; Truffinet et al. 2005). The CTCF protein is responsible for the ability of the HS4 insulator to block enhancers (Yannaki et al. 2002;

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11248-019-00122-9) contains supplementary material, which is available to authorized users.

A. Deykin · M. Tikhonov · V. Kalmykov · I. Korobko · P. Georgiev (✉) · O. Maksimenko (✉)
Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilova St., Moscow, Russia 119334
e-mail: georgiev_p@mail.ru

O. Maksimenko
e-mail: maksog@mail.ru

Published online: 28 March 2019

Yusufzai and Felsenfeld 2004). The HS4 insulator is also bound by the USF1, USF2, and BGP1/Vezf1 proteins which form the boundary between active chromatin and heterochromatin (Dickson et al. 2010; West et al. 2004).

Recent developments in targeted genome editing have made it possible to insert the transgene into a predefined genomic site (so-called safe harbor) where its expression is ensured (DeKelver et al. 2010). However, a number of experiments performed to date indicate that a major portion of the mammalian genome is transcribed and that a large percentage of the transcripts are accounted for by long non-protein-coding sequences (lncRNAs) (Bertone et al. 2004; Cheng et al. 2005; Kapranov et al. 2007). Transcription through the transgene can affect its expression by many different mechanisms. Transcripts can act as molecular scaffolds in recruiting different enzymatic complexes that repress transcription (Long et al. 2017). The RNA Pol II complex moving through the transgene regulatory region can disturb protein complexes associated with DNA (Erokhin et al. 2013; Martianov et al. 2007; Mazo et al. 2007; Palmer et al. 2011). Finally, transcription through the transgene often induces RNA interference, which leads to transcription repression (Eszterhas et al. 2002).

Based on the facts indicating that transcription through regulatory elements usually induces transgene repression, we have tested whether transcription termination at the transgene borders has a positive effect on transgene expression. The firefly luciferase gene (*Fluc*) under control of the previously characterized goat *β-casein* gene promoter was used as a reporter. The goat *β-casein* gene promoter, known for its high efficiency, is used in commercial vector pBC1 (Invitrogen) for the production of target proteins in the mammary glands of mice, goats, and cows (Edmunds et al. 1998; Goldman et al. 2012; Maksimenko et al. 2013; Tong et al. 2011; Zhang et al. 2008). The 6.2-kb regulatory region of the *β-casein* gene consists of a promoter and a hormone-dependent enhancer that stimulates the promoter only in the mammary gland cells (Kabotyanski et al. 2009). The pBC1 vector also contains two chicken HS4 insulators located upstream of the goat *β-casein* promoter.

To exclude the effect of strong virus transcriptional terminators such as SV40, we selected two bidirectional transcription terminator sequences (TTSs) from the human genome and inserted them upstream of the goat *β-casein* gene promoter. The control transgenic mice (without regulatory elements upstream of the promoter) showed low expression of the firefly luciferase reporter in the mammary glands. In contrast, insertion of HS4 or TTS upstream of the promoter resulted in approximately tenfold enhancement of the reporter gene expression. Thus, TTSs can be used for improving the reporter gene expression in transgenic lines.

Materials and methods

Plasmids

The bicistronic expression construct pRFP-IRES-EGFP (Fig. 1a) was prepared on the basis of pEGFP-N1 vector (Clontech, USA) containing cDNA for the enhanced green fluorescent protein gene (*EGFP*) under control of the constitutive early/immediate cytomegalovirus (*CMV*) promoter, in which cDNA for the red fluorescent protein gene (*RFP*) and encephalomyocarditis virus internal ribosomal entry sequence (IRES) derived from pIRES vector (Clontech) were sequentially cloned into multiple sites located between the *CMV* promoter and *EGFP* cDNA so that the *EGFP* cDNA was not in frame with the *RFP* cDNA. The bicistronic expression construct pRluc-IRES-Fluc was prepared by the same scheme, but the *Renilla* and firefly luciferase genes *Rluc* and *Fluc* were inserted instead of *RFP* and *EGFP*.

Candidate TTSs identified in silico (Table 1) were amplified with the corresponding primer pairs and cloned into the *Eco*RV restriction site of the pBluescript II SK(+) vector (Stratagene, USA) for subsequent subcloning into the pRFP-IRES-EGFP or pRluc-IRES-Fluc plasmid between the first ORF (*RFP* or *Rluc*) and IRES in direct or reverse orientation (Fig. 1a).

Plasmids for producing transgenic animals were constructed on the backbone of the pBC1 vector (Invitrogene, USA). First, the entire 3′-region of the goat *β-casein* gene located between *Xho*I and *Not*I restriction sites was replaced by the bovine growth hormone gene transcription termination sequence. Firefly luciferase cDNA was derived from pGL3-control vector (Promega, USA) and engineered so as to contain at its 5′-end, immediately after translation initiation AUG codon, the sequence from goat β-

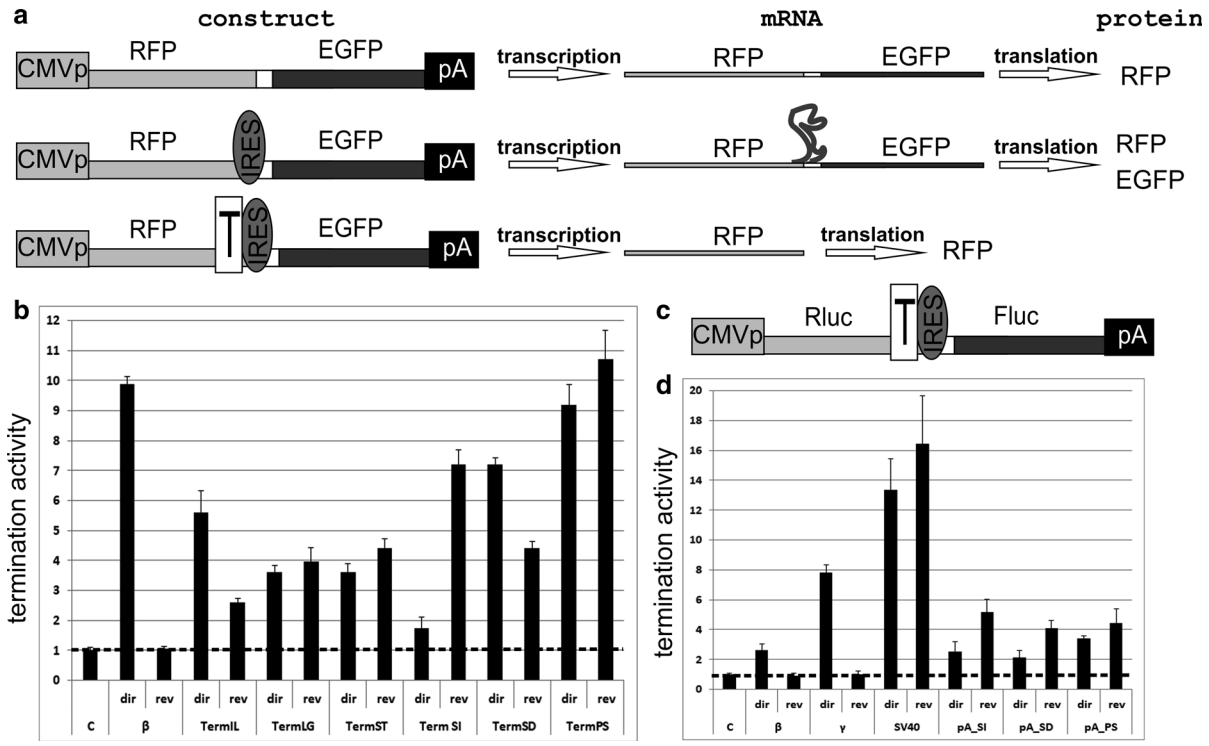

Fig. 1 **a** Scheme of bicistronic RFP/EGFP reporter constructs containing the CMV promoter (CMVp), polyadenylation signal from SV40 virus (pA), internal ribosome entry site (IRES), and test termination sequence (T). **b** Histogram of IRES-directed EGFP expression in CHO cells transfected with control pRFP-IRES-EGFP construct without TTSs ('C') and constructs with selected TTSs inserted in direct (dir) or reverse (rev) orientation. 'β' is the control termination sequence from the β-*globin* locus that work as terminator only in one direction ('dir'). Data are presented as 'termination activity' estimated from the ratio of the number of CHO cells expressing EGFP after transfection with the control construct to the number of such cells after transfection with the test constructs. Dot line shows the level for negative control. Here and in Fig. S1, TTSs are named as in Table 1. Error bars show standard deviations (n = 3). **c** Scheme of bicistronic Rluc/Fluc reporter construct. **d** Histogram of IRES-directed Fluc expression in CHO cells transfected with control pRluc-IRES-Fluc construct without TTSs (C) and constructs with polyA sequences from the test TTSs (pA_SI, pA_SD, pA_PS) inserted in direct (dir) or reverse (rev) orientation. Data are presented as 'termination activity' estimated from the ratio of Rluc expression to Fluc expression normalized with respect to that in the control (C). 'β', 'γ', are the control termination sequences from the β-*globin* and γ-*globin* loci, respectively, that work as terminator only in one direction ('dir'); 'SV40' is the control termination sequence from the SV40 virus that works as terminator in both orientations. Error bars show standard deviations (n = 3)

casein cDNA encoding a signal peptide that enables luciferase secretion into milk when expressed in the mammary gland (secreted luciferase, SLuc). SLuc cDNA was then cloned into the *Xho*I restriction site of the modified pBC1 vector (plasmid pHS4-SLuc). For microinjection to generate transgenic animals, the plasmids were digested with *Pac*I, DNA fragments were separated in agarose gel, and the plasmid backbone-less DNA fragment encompassing the transgene was purified using Zymoclean Large Fragment DNA Recovery Kit (Zymo Research).

Cell cultivation and transfection

CHO-K1, MCF7 and HeLa cells were cultured in DMEM with 10% of inactivated fetal bovine serum, 2 mM L-glutamine, 35 mg/L L-proline, and Mycokill-AB antibiotic mixture (PAA Laboratories) at a working concentration at 37 °C in a humidified atmosphere with 5% CO_2. The cells were subcultured every 5 days at a concentration of 10^5 cells/cm^2. Transient cell transfection with recombinant plasmids was performed as follows. To prepare the transfection mixture, 3–4 μg of plasmid DNA was added to 375 μL of

Table 1 Characteristics of TTSs used in experiments

TTS	5′ gene	3′ gene	Full TTS element		Minimal poly(A) element	
			Primers for amplification	Length (bp)	Primers for amplification	Length (bp)
TermIL	*ITPKA*	*LTK*	5′-ACACCCTGGAGGTATCCGAGTT-3′ 5′-ATCTTGGAGCGTCTGCAGTACTG-3′	1550		
TermLG	*LYPLA2*	*GALE*	5′-GGCTGAGAAGCTCCGGTCTGT-3′ 5′-AAGAGGAGCTGGGGTGGACAG-3′	1680		
TermST	*SELO*	*TUBGCP6*	5′-CTGGACAAGGACCTGGAAGGC-3′ 5′-TCTCTTCAAGGTGAGGCCCA-3′	950		
TermSI	*SRPK3*	*IDH3G*	5′-TTCCATGGCACAGACCACATCGCTCAC-3′ 5′-TAAGGCTGTCCTGGCATCCATG-3′	1178	5′-CGCCAGTTCTCCACAACCA-3′ 5′-TGGATTCCCCTTCCCACTC-3′	246
TermSD	*SKIV2L*	*DOM3Z*	5′-CTGAATTTTGGGCTGGTTG-3′ 5′-GCCCAGAGCACGGTTGTC-3′	768	5′-ATCGTATTTGCGGCCAGC-3′ 5′-GCCCATATGGTATGTGGAA-3′	264
TermPS	*PSMC5*	*SMARCD2*	5′-GCATGTATGCCCTGCGAGAAC-3′ 5′-GGAGAGACGAGCTGCTTTCTAC-3′	1396	5′-GGCTAGGGCATGTGTGTGT-3′ 5′-ACATCTGGCAGGTTTGGA-3′	267

serum-free cell culture medium; in a separate tube, Lipofectamine 2000 (3 μL per microgram plasmid DNA) was mixed with the same volume of serum-free cell culture medium; the solutions were then pooled and incubated at room temperature for 30–40 min. The cells that reached 70–80% confluence (8×10^4 - cells/cm^2) were washed with serum-free medium, incubated in the transfection mixture for 4–6 h, and then cultured in serum-containing DMEM. The level of reporter gene expression was assessed on day 2 after transfection (36–48 h) by measuring fluorescence or chemiluminescence intensity, using non-transfected cells as negative control.

Cytofluorometric analysis

The cells were washed with phosphate buffered saline (PBS), treated with trypsin to detach them from the substrate, washed twice to remove trypsin, and resuspended in a fresh portion of PBS. The resulting suspension (10^6 cells/mL) was transferred to 5-mL round-bottom test tubes.

Measurements were made in a MACSQuant Analyzer VYB flow cytofluorometer (Miltenyi Biotec, Germany). The voltage in the measurement channels was adjusted so as not to take into account the autofluorescence of nontransfected cells and then was maintained constant. The cells with fluorescence intensity above 10 arbitrary units on the logarithmic scale for a proper channel were counted and used to quantify the level of reporter gene expression.

Dual luciferase assay

A Firefly & Renilla Luciferase Assay Kit (Biotium) was used according to the manufacturer's protocol. Measurements were made in a Synergy 4 microplate reader (BioTek) with a sensitivity of 100 and 1 s exposure time.

Generation of transgenic mice

The purified DNA fragment containing the transgene (5 ng/μL in 10 mM tris–HCl, pH 7.4, with 0.1 mM EDTA) was injected into the male pronuclei of fertilized F1 (CBA × C57BL/6) mouse oocytes obtained by inducing superovulation (Zvezdova et al. 2010), and the surviving zygotes were implanted into pseudopregnant females.

Identification of transgenic animals

Transgene presence was determined by PCR on the template of genomic DNA (100 ng) isolated from tail biopsy samples using proteinase K digestion and phenol/chloroform extraction. PCR with primers to detect luciferase cDNA (5′-TGGAAGACGCCAAAA ACATAAAG-3′ and 5′-GCGAAATGCCCATACT GTT-3′) and HS *Taq* polymerase (Sileks) was performed under the following conditions: 94 °C, 15 s; 94 °C, 20 s, 58 °C, 30 s, 72 °C, 40 s (35 cycles); 72 °C, 5 min. The presence of PCR product amplified from the luciferase cDNA template (300 bp) was detected by agarose gel electrophoresis. As a control, PCR for the mouse actin gene with primer pair 5′-AG CTGAGAGGGAAATTGTGCG-3′ and 5′-GCAACG GAACCGCTCGTT-3′ was performed under the same conditions.

Transgene copy number determination

Transgene copy number was determined in real-time PCR using a calibration curve, which was plotted as follows. A fragment of transgene was amplified with primers 5′-GCTATGAAGAGATACGCCCTGGTT CC-3′ and 5′-CGACGATTCTGTGATTTGTATTCAG CC-3′, cloned into the *Eco*RV restriction site of pBluescript II SK(+) vector (Stratagene, USA), and reamplified with M13dir (5′-GTAAAACGACGGCCAGT-3′) and M13rev (5′-CAGGAAACAGCTATGAC-3′) primers flanking the multiple cloning site of the vector (amplification fragment length 378 bp, corresponding to 2.5×10^{11} molecules per 100 ng DNA). The amplification product was resolved by agarose gel electrophoresis, purified from the gel with ZymoClean Gel-recovery kit (Zymo Research), and its concentration was determined with a Qubit fluorometer (Invitrogene, USA). Calibration curve was plotted using different dilutions of the fragment in wild-type mouse genomic DNA solution (1650 diploid genomes per 100 ng) so as to mimic the presence of 1, 2, 4, 8, 12, or 24 transgene copies per diploid mouse genome and reflect the dependence between the PCR cycle number at which the amplification signal reached the threshold level and the number of mimicked transgene copies. The transgene copy number in experimental samples was determined based on the threshold cycle number, with reference to the calibration curve.

The PCR reaction mixture contained 2 μL of primer mix (10 pmol/μL each), 2 μL of dNTP (2 mM each), 2 μL of 10 × PCR buffer, 0.5 μL of EvaGreen dye mix (Biotium), 1 μL of genomic DNA, and 0.2 μL of HotStart *Taq* polymerase at 5 u/μL (Sileks, Russia) in a total volume of 20 μL. Amplification conditions were as follows: 94 °C, 15 min (1 cycle); 94 °C, 20 s; 62 °C, 20 s; 72 °C, 20 s (45 cycles).

To normalize results with respect to the genomic DNA content in the reaction, real-time PCR with the mouse actin gene primers, 5′-AGCTGAGAGGGAA ATTGTGCG-3′ and 5′-GCAACGGAACCGCTCG TT-3′, was performed using wild-type genomic DNA as a template.

Mouse milking and milk processing

Lactating females were removed from the nest 2 h before milking. Mice were immobilized during surgery with Avertin (a 2.5% aqueous solution of the preparation containing 1 g of 2,2,2-tribromoethanol in 1 ml of 2-methyl-2-butanol). Avertin was injected intraperitoneally at 15 μl per gram body weight, intraperitoneally injected with oxytocin (2 IU), and milking was started after 2 min (DePeters and Hovey 2009). Milk was expressed using a vacuum milking machine, collected in a 1.5 ml centrifuge tube and placed on ice.

A milk sample was diluted with two volumes of PBS, centrifuged for 10 min at 12,000*g,* and the medium aqueous fraction was collected and stored at − 70 °C.

Luciferase activity in 20-μL aliquots of processed milk samples was determined with a Firefly Luciferase Assay Kit (Biotium). For each set of samples, a calibration curve was plotted using standardized dilutions of recombinant firefly luciferase in the processed milk of non-transgenic mice.

RNA purification and quantitative analysis

Total RNA was isolated from female mammary glands using the TRI reagent (MRC) according to the manufacturer's instructions. For each replicate, 100 mg of tissue were collected and frozen in liquid nitrogen. RNA was treated with TURBO DNA-free™ Kit (Thermo Scientific) to eliminate residual genomic DNA. The synthesis of cDNA was performed using

50 U of ArrayScript reverse transcriptase (Thermo Scientific) in a reaction mixture containing 5 μg of RNA and 25 μM random hexamer or 1 μM specific primers.

The amounts of specific cDNA fragments were quantified by real-time PCR. At least three independent technical replicates were made for each RNA sample. Relative levels of mRNA expression were calculated in the linear amplification range by calibration to a standard curve. Significance in level changes was calculated by *t* test without equal variance (MS Excel). The sequences of oligonucleotides used for quantification of mRNA were: luciferase 5′-GCTATGAAGAGATACGCCCT GGTTCC-3′ and 5′-CGACGATTCTGTGATTTG-TATTCAGCC-3′; β-actin for normalization 5′-AGCTG AGAGGGAAATTGTGCG-3′ and 5′-GCAACGGAA CCGCTCGTT-3′.

Results

Identification of candidate transcription termination sequences

To test whether transcription termination sequences can positively affect transgene expression, we made an attempt to identify new bidirectional TTSs in the human genome. A combination of the most effective transcriptional terminators, introns, and regulatory elements from different organisms in the same construct is a well-established approach to improve transgene expression (Petitclerc et al. 1995). The best characterized and strong bidirectional transcription terminator is that derived from the SV40 virus which is used for many applications in a wide range of higher eukaryotes. Here, we tried to identify new bidirectional human TTS that could be used in the future to generate different transgenic mammals. For this purpose, we examined sequences between closely spaced (100–200 bp), convergently oriented pairs of housekeeping genes that are highly expressed in all tissues (Fig.S1a).

A database of annotated human genes was compiled based on analysis of the reference human genome assembly (Build 37.1) with our in-house developed software (Fig.S1). The database contained 40 283 entries, each including taxon ID, chromosome number, nucleotide positions of gene start and end, information on gene location on the plus or minus strand, contig name and its position on a chromosome, contig orientation, gene name, gene ID, record type ("gene"), project ID relevant to an annotation, transcript type synthesized from a gene, and code for annotation reliability (Table S1).

Using our in-house software, the database was analyzed for the presence on the same chromosome of pairs of genes that are transcribed in opposite, convergent directions and whose 3′-ends are spaced by a DNA segment not exceeding 100 nt in length (Fig.S1). As a result, 141 sites satisfying this requirement were identified (Table S2), but 61 of these sites were excluded from further analysis because the existence of transcripts for the adjacent gene(s) was only predicted but not demonstrated experimentally. Finally, six out of the remaining 80 sites (Table S2) were selected for in-depth analysis based on high mRNA expression level (Table 1). These regions (ca. 200 bp each) had two sets of all the key *cis* elements characterized for the mammalian mRNA 3′ processing sites (Shi and Manley 2015), including the AAUAAA, GU-rich downstream element, and U-rich upstream auxiliary element (Fig. 1a; Fig. S2). The selected candidate TTS were amplified and cloned for further experimental testing of their capacity to provide efficient termination of transcription.

Transcription termination capacity of the selected candidate sequences

To test the selected TTSs for transcription termination efficiency, we used previously described approach based on the bicistronic model system (Tikhonov et al. 2013). The bicistronic expression construct, pRFP-Term-IRES-EGFP, (Fig. 1a) contained *RFP* cDNA under control of *CMV* promoter and *EGFP* cDNA cloned at 3′ to *RFP* cDNA, with mismatched open reading frames of RFP and EGFP. The IRES from encephalomyocarditis virus (Ngoi et al. 2004) was inserted between *RFP* and *EGFP* cDNAs to support cap-independent EGFP translation from long *RFP-IRES-EGFP* transcripts. The candidate TTSs were inserted either in direct or reverse orientation between *RFP* cDNA and IRES, and their efficiency in transcription termination was estimated from the level of EGFP fluorescence: the more effective the transcription termination by the inserted TTS, the less the amounts of *RFP* transcripts containing the *EGFP* cDNA sequence and of the translated EGFP protein.

Plasmids pRFP-Term-IRES-EGFP with each of the six selected candidate TTSs and the control termination sequence from the β-*globin* locus either in direct or reverse orientation were constructed and used for transient transfection of CHO cells. The percentage of cells with EGFP fluorescence above the selected threshold (characterizing the efficiency of transcription termination) was estimated by FACS analysis 48 h after transfection. The results of a representative experiment are summarized in Fig. 1b and Table S3. They show that, among the six candidate sequences, the highest transcription termination capacity in either orientation was observed for two TTSs, TermSD, and TermPS. The termination of transcription by the selected sequences was confirmed in experiments on transfection of MCF7 and HeLa cells (Fig. S3, Table S3). To verify these results, we developed a similar bicistronic model, pRluc-Term-IRES-Fluc, based on *Renilla* and firefly luciferases (Fig. 1c). In this system, we used the minimal TTS with poly(A)signals (apr 200 bp) (Table 1) and determined the *Renilla*- to-firefly luciferase luminescence ratio as a quantitative characteristic of termination efficiency. Plasmids pRluc-Term-IRES-Fluc with three selected candidate sequences (pA_SI, pA_SD, and pA_PS) and control termination sequences from SV40 virus, β-globin locus, and γ-globin locus in direct or reverse orientation were constructed and used to transiently transfect CHO and HeLa cells. The results confirmed effective transcription termination by the pA_SD and pA_PS sequences in either orientation (Fig. 1d, S4). However, the effect of SV40 termination sequences was stronger than that of the test TTSs.

Transcription termination sequences improve transgene expression in mouse mammary glands

Next, we tested whether the TTSs could improve transgene expression in the mammary gland as was previously shown for the two tandemly arranged HS4 insulators (Giraldo et al. 2003; Goldman et al. 2012; Potts et al. 2000; Rival-Gervier et al. 2003). The plasmid for transgenesis was constructed on the basis of pBC1 vector (Invitrogen) commonly used for transgenic protein production in the mammary gland. The control construct (Luc) (Fig. 2a) contained the promoter and 5′ UTR from the goat β-casein gene followed by the secreted firefly luciferase cDNA (LUC) fused to the 3′ UTR that included the

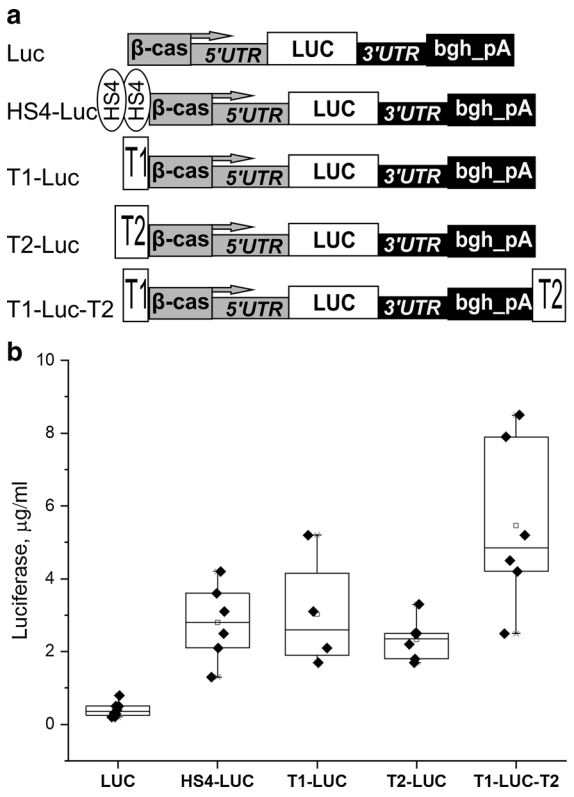

Fig. 2 (a) Schematic of constructs for mouse transgenesis that contained β-casein promoter (β-cas), polyadenylation signal from bovine growth hormone (bgh_pA), insulator from chicken β-globin locus (HS4), and test termination sequences (T1 and T2). The firefly luciferase reporter gene (LUC) fused with the sequence encoding N-terminal β-casein signal peptide was inserted between 5′-UTR from the β-casein gene and 3′-UTR from the bovine growth factor gene. **(b)** Box plots of luciferase expression in the mammary glands of F1 and F2 female progeny of the founders transgenic for the indicated constructs. Luciferase content in the milk (μg/ml) is shown only for the females with a confirmed single copy of the construct. Data are presented as dots for each single-copy transgenic line, boxes demonstrate the spread of values for each construct, the horizontal line in the box is the mean

polyadenylation signal of the bovine growth hormone gene (bdh_pA) (Goodwin and Rottman 1992). The test constructs (HS4-Luc, T1-Luc, and T2-Luc) contained HS4, T1 (TermSD), and T2 (TermPS) sequences inserted upstream of the goat β-casein promoter. In addition, we made the T1-Luc-T2 construct in which the T1 and T2 sequences flanked the reporter gene.

These linear constructs were injected into the male pronuclei of mouse zygotes, and the injected single-cell embryos were implanted into recipient females. Transgenic mice in F0 (founders) were identified by

PCR and crossed with non-transgenic mice. The construct copy numbers in the F1 lines derived from the founders were estimated using a qPCR technique. As a result, we identified females with different numbers of construct copies in the progeny of all the founders. For several transgenic males and females we obtained the F2 progeny by crossing transgenic mice with new non-transgenic mice. In many cases, parents carrying multiple copies of a construct gave birth to offspring of females with different numbers of copies of the construct. Thus, during initial injection of DNA, constructs simultaneously were inserted at different genomic regions in founders.

To compare the luciferase expression in different constructs, we have summarized the results for females carrying only a single copy of a construct, single-copy female (Fig. 2b, Fig. S5). In general, we have obtained at least one female with an independent single-copy insertion from each founder. Different females with single-copy insertions derived from a single founder, may have a common origin from a single integration event for the construct. As a result, there were at least three independent integration events with the Luc construct, four with HS4-Luc, two with T1-Luc, two with T2-Luc, and three with T1-Luc-T2 (Fig. S5, Table S4). Firefly luciferase content in the milk, as a measure of the transgene expression, was determined by a luminometric assay in heterozygous F1 and F2 females (Fig. 2b). Milk samples were taken at the end of the first and second week of lactation (Table S4). Some females had two pregnancies and milk samples were taken twice. Also, we found correlation between luciferase activity in the milk and the level of *luciferase* mRNA expression in the mammary glands of transgenic females (Fig. S6).

The level of luciferase expression in the control transgenic lines (Luc) proved to be approximately 10 times lower than in the lines carrying tandem HS4 insulators, which confirmed the role of this insulator in stimulation of the goat *β-casein* promoter. Substitution of HS4 insulators with either T1 or T2 sequence at the 5′-end of the transgenic expression unit provided for a similar level of luciferase expression, attesting to the capacity of these TTSs to insulate the transgene. Importantly, when the transgenic expression unit was flanked with TTSs on both 5′ and 3′ ends, the level of luciferase expression was further increased, suggesting a cumulative effect of complete transgene insulation with TTSs.

Discussion

Transcription termination is triggered by multipartite signals on the nascent RNA, which are recognized by the mRNA 3′ processing complex (Shi and Manley 2015). Thus, termination activity of a TTS depends on the efficiency of processing complex recruitment to RNA synthesized by RNA Pol II moving along the TTS. Our attempt to identify the strongest human TTSs was based on the premise that transcription should be completely terminated between the closely spaced, convergently oriented, and actively transcribed genes. As a result, we identified human bidirectional transcription termination sequences located between two gene pairs, *PSMC5*/*SMARCD2* and *SKIV2L*/*DOM3Z*. Similar to the SV40 terminator, these TTSs contain all motifs that are essential for transcription termination. Despite this similarity, however, they proved to function less effectively than the SV40 terminator in our assays. In our opinion, the difference in activity between different terminator sequences may be explained by as yet unknown variations in the motifs that are important for their recognition by the mRNA 3′ processing complex.

As follows from analysis of transgene product secretion with milk, the two selected TTSs indeed supported higher transgene expression when inserted upstream of the goat *β-casein* promoter. These results agree with data on the negative role of transcription through the transgene. The expression level in the presence of either TTS was enhanced approximately tenfold, as in the presence of two tandemly arranged HS4 insulators from the chicken β-*globin* locus that are traditionally used to improve the reporter gene expression in transgenic animals. Therefore, the identified TTSs can functionally substitute for the HS4 insulators in transgenic constructs, with their smaller size being an advantage in such experiments. A noteworthy fact is that transgene flanking with two different TTSs resulted in further enhancement of transgene expression. A likely explanation is that the TTSs on both 5′ and 3′ ends of the transgene prevent the spread of transcription from neighboring genomic regions. This is in line with data on the improved expression of two expression units in a single construct when one of them is flanked by insulators on both sides (Hasegawa and Nakatsuji 2002).

There are several possible mechanisms how proteins bound to the HS4 insulator can improve the

expression of a transgene. The CTCF protein can block the spreading of repressor complexes (Ghirlando and Felsenfeld 2016); the USF1/2 proteins form multiprotein complexes with NURF and hSET1 that have nucleosome-remodeling and histone H3K4 methyltransferase activities (Li et al. 2011); finally, the BGP1/Vezf1 protein protects the GC-rich sequences of an insulator against methylation (Dickson et al. 2010). Interestingly, it has also been shown that BGP1/Vezf1 is also able to terminate weak transcription in the HS4 sequences that is initiated in the neighboring heterochromatic region (Gowher et al. 2012). Thus, the HS4 insulator can function as a weak TTS. It seems likely that HS4 and strong TTSs have partially complementary activities. Therefore, a combination of these elements may have a synergetic effect in improving transgene expression. Further research is needed to elucidate this question.

Acknowledgements We are grateful to N. A. Gorgolyuk and V. A. Mogila for their help in preparing the manuscript. Experiments were performed using the equipment of the IGB RAS facilities supported by the Ministry of Science and Education of the Russian Federation. This study was supported by the Russian Science Foundation, Project No. 16-14-00150 (to I.K.).

References

Bertone P et al (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246. https://doi.org/10.1126/science.1103388

Bishop JO, Smith P (1989) Mechanism of chromosomal integration of microinjected DNA. Mol Biol Med 6:283–298

Burkov IA et al (2013) Expression of the human granulocyte-macrophage colony stimulating factor (hGM-CSF) gene under control of the 5′-regulatory sequence of the goat alpha-S1-casein gene with and without a MAR element in transgenic mice. Transgenic Res 22:949–964. https://doi.org/10.1007/s11248-013-9697-4

Cheng J et al (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308:1149–1154. https://doi.org/10.1126/science.1108625

Choo KB, Liew LN, Wang HC (1989) Production of transgenic mice carrying the human hepatitis B virus or the human papillomavirus DNA sequences in Taiwan: analysis of physical structure and hereditary mode. Proc Natl Sci Counc Repub China B 13:314–318

DeKelver RC et al (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res 20:1133–1142. https://doi.org/10.1101/gr.106773.110

DePeters EJ, Hovey RC (2009) Methods for collecting milk from mice. J Mammary Gland Biol Neoplasia 14:397–400. https://doi.org/10.1007/s10911-009-9158-0

Dickson J, Gowher H, Strogantsev R, Gaszner M, Hair A, Felsenfeld G, West AG (2010) VEZF1 elements mediate protection from DNA methylation. PLoS Genet 6:e1000804. https://doi.org/10.1371/journal.pgen.1000804

Dobie KW et al (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. Proc Natl Acad Sci USA 93:6659–6664

Edmunds T et al (1998) Transgenically produced human antithrombin: structural and functional comparison to human plasma-derived antithrombin. Blood 91:4561–4571

Emery DW, Yannaki E, Tubb J, Nishino T, Li Q, Stamatoyannopoulos G (2002) Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. Blood 100:2012–2019. https://doi.org/10.1182/blood-2002-01-0219

Erokhin M, Davydova A, Parshikov A, Studitsky VM, Georgiev P, Chetverina D (2013) Transcription through enhancers suppresses their activity in Drosophila. Epigenetics Chromatin 6:31. https://doi.org/10.1186/1756-8935-6-31

Eszterhas SK, Bouhassira EE, Martin DI, Fiering S (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol Cell Biol 22:469–479

Ghirlando R, Felsenfeld G (2016) CTCF: making the right connections. Genes Dev 30:881–891. https://doi.org/10.1101/gad.277863.116

Giraldo P, Rival-Gervier S, Houdebine LM, Montoliu L (2003) The potential benefits of insulators on heterologous constructs in transgenic animals. Transgenic Res 12:751–755

Goldman IL et al (2012) Production of human lactoferrin in animal milk. Biochem Cell Biol 90:513–519. https://doi.org/10.1139/o11-088

Goodwin EC, Rottman FM (1992) The 3′-flanking sequence of the bovine growth hormone gene contains novel elements required for efficient and accurate polyadenylation. J Biol Chem 267:16330–16334

Gowher H, Brick K, Camerini-Otero RD, Felsenfeld G (2012) Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. Proc Natl Acad Sci USA 109:2370–2375. https://doi.org/10.1073/pnas.1121538109

Guglielmi L, Le Bert M, Truffinet V, Cogne M, Denizot Y (2003) Insulators to improve expression of a 3(′)IgH LCR-driven reporter gene in transgenic mouse models. Biochem Biophys Res Commun 307:466–471

Hasegawa K, Nakatsuji N (2002) Insulators prevent transcriptional interference between two promoters in a double gene construct for transgenesis. FEBS Lett 520:47–52

Houdebine LM (2018) Production of pharmaceutical proteins by transgenic animals. Rev Sci Tech 37:131–139. https://doi.org/10.20506/rst.37.1.2746

Kabotyanski EB, Rijnkels M, Freeman-Zadrowski C, Buser AC, Edwards DP, Rosen JM (2009) Lactogenic hormonal induction of long distance interactions between beta-casein gene regulatory elements. J Biol Chem 284:22815–22824. https://doi.org/10.1074/jbc.M109.032490

Kapranov P et al (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316:1484–1488. https://doi.org/10.1126/science.1138341

Li X et al (2011) Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. Blood 118:1386–1394. https://doi.org/10.1182/blood-2010-11-319111

Long Y, Wang X, Youmans DT, Cech TR (2017) How do lncRNAs regulate transcription? Sci Adv 3:eaao2110. https://doi.org/10.1126/sciadv.aao2110

Maksimenko OG, Deykin AV, Khodarovich YM, Georgiev PG (2013) Use of transgenic animals in biotechnology: prospects and problems. Acta Nat 5:33–46

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666–670. https://doi.org/10.1038/nature05519

Mazo A, Hodgson JW, Petruk S, Sedkov Y, Brock HW (2007) Transcriptional interference: an unexpected layer of complexity in gene regulation. J Cell Sci 120:2755–2761. https://doi.org/10.1242/jcs.007633

Ngoi SM, Chien AC, Lee CG (2004) Exploiting internal ribosome entry sites in gene therapy vector design. Curr Gene Ther 4:15–31

Palmer AC, Egan JB, Shearwin KE (2011) Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription 2:9–14. https://doi.org/10.4161/trns.2.1.13511

Petitclerc D et al (1995) The effect of various introns and transcription terminators on the efficiency of expression vectors in various cultured cell lines and in the mammary gland of transgenic mice. J Biotechnol 40:169–178

Potts W, Tucker D, Wood H, Martin C (2000) Chicken beta-globin 5′HS4 insulators function to reduce variability in transgenic founder mice. Biochem Biophys Res Commun 273:1015–1018. https://doi.org/10.1006/bbrc.2000.3013

Rival-Gervier S et al (2003) The insulator effect of the 5′HS4 region from the beta-globin chicken locus on the rabbit WAP gene promoter activity in transgenic mice. Transgenic Res 12:723–730

Shi Y, Manley JL (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev 29:889–897. https://doi.org/10.1101/gad.261974.115

Suzuki O et al (2006) Transgene insertion pattern analysis using genomic walking in a transgenic mouse line. Exp Anim 55:65–69

Tikhonov M, Georgiev P, Maksimenko O (2013) Competition within introns: splicing wins over polyadenylation via a general mechanism. Acta Nat 5:52

Tong J et al (2011) Production of recombinant human lysozyme in the milk of transgenic pigs. Transgenic Res 20:417–419. https://doi.org/10.1007/s11248-010-9409-2

Truffinet V, Guglielmi L, Cogne M, Denizot Y (2005) The chicken beta-globin HS4 insulator is not a silver bullet to obtain copy-number dependent expression of transgenes in stable B cell transfectants. Immunol Lett 96:303–304. https://doi.org/10.1016/j.imlet.2004.09.002

West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G (2004) Recruitment of histone modifications by USF proteins at a vertebrate barrier element. Mol Cell 16:453–463. https://doi.org/10.1016/j.molcel.2004.10.005

Yannaki E, Tubb J, Aker M, Stamatoyannopoulos G, Emery DW (2002) Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors. Mol Ther 5:589–598. https://doi.org/10.1006/mthe.2002.0582

Yusufzai TM, Felsenfeld G (2004) The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci USA 101:8620–8624. https://doi.org/10.1073/pnas.0402938101

Zhang J et al (2008) Expression of active recombinant human lactoferrin in the milk of transgenic goats. Protein Expr Purif 57:127–135. https://doi.org/10.1016/j.pep.2007.10.015

Zvezdova ES et al (2010) Generation of transgenic animals, expressing alpha- and beta-chains of autoreactive TCR. Mol Biol (Mosk) 44:311–322