Transcriptional regulation of human-specific SVA_{F1} retrotransposons by cis-regulatory *MAST2* sequences

Anastasia A. Zabolotneva, Olga Bantysh, Maria V. Suntsova, Nadezhda Efimova, Galina V. Malakhova, Gerald G. Schumann, Nurshat M. Gayfullin, Anton A. Buzdin

PII: S0378-1119(12)00591-4
DOI: doi: 10.1016/j.gene.2012.05.016
Reference: GENE 37546

To appear in: *Gene*

Accepted date: 4 May 2012

Please cite this article as: Zabolotneva, Anastasia A., Bantysh, Olga, Suntsova, Maria V., Efimova, Nadezhda, Malakhova, Galina V., Schumann, Gerald G., Gayfullin, Nurshat M., Buzdin, Anton A., Transcriptional regulation of human-specific SVA_{F1} retrotransposons by cis-regulatory *MAST2* sequences, *Gene* (2012), doi: 10.1016/j.gene.2012.05.016

Transcriptional regulation of human-specific SVA_{F1} retrotransposons by cis-regulatory *MAST2* sequences

Anastasia A. Zabolotneva[1], Olga Bantysh[1], Maria V. Suntsova[1], Nadezhda Efimova[1], Galina V. Malakhova[1], Gerald G. Schumann[2], Nurshat M. Gayfullin[3], and Anton A. Buzdin[1]*

[1]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya, Moscow 117997, Russia, [2]Division of Medical Biotechnology, Paul-Ehrlich-Institut, Paul-Ehrlich-Straße 51-59, D-63225 Langen, Germany, [3]Lomonosov Moscow State University, Lomonosovsky pr., 31-5, Moscow 119192, Russia

*To whom correspondence should be addressed. Tel/Fax: +7 495 727 38 63; Email: bu3din@mail.ru

ABSTRACT

SVA elements represent the youngest family of hominid non-LTR retrotransposons. Recently, a human-specific subfamily (termed SVA_{F1}, CpG-SVA, or MAST2-SVA) was discovered representing fusion of the CpG island-containing exon 1 of the *MAST2* gene and a 5'-truncated SVA. SVA_{F1} includes at least 84 members, which suggests exceptionally high retrotransposition level. We investigated if the acquirement of the *MAST2* CpG-island might play a role in the success of the SVA_{F1} subfamily. We observed that in 16 samples representing seven human tissues, *MAST2* was cotranscribed with the members of the SVA_{F1} subfamily, but not with other retrotransposons. We found that the methylation status of the *MAST2*-derived sequences of SVA_{F1} elements reversely correlates with transcriptional activity of *MAST2*. The *MAST2* sequence at the 5' end of SVA_{F1} acts as positive transcriptional regulator in human germ cells. Finally, in various testicular tissue samples we uncovered transcriptional correlation of *MAST2* with the human L1, *Alu* and SVA retrotransposons.

Key words: transposable element, transcriptional regulation, human genome, germ cells, SVA retrotransposon.

1. Introduction

Retroelements (REs) are mobile genetic elements that spread throughout the host genome by reverse transcription of their respective RNA intermediate (Gogvadze and Buzdin, 2009, Goodier and Kazazian, 2008, Schumann et al., 2010). They constitute ~42% of the human genomic DNA and represent the only class of mobile elements naturally transposing in primates. In modern humans, only the non-LTR retrotransposon families L1, *Alu* and SVA are thought to include transpositionally active elements (Mills et al., 2007). Among the non-LTR retrotransposons, only L1 elements are termed autonomous, because they encode the protein machinery required for their retrotransposition in *cis* (Wei et al., 2001). The L1-encoded protein machinery was also demonstrated to be responsible for the *trans*-mobilization of the non-autonomous retrotransposon families *Alu* and SVA (Dewannieux et al., 2003, Hancks et al., 2011, Hancks and Kazazian, 2010, Raiz et al., 2012) and for processed pseudogene formation (Esnaul et al., 2000). SVA elements are hominid-specific non-LTR retrotransposons, distinguished by their organization as composite mobile elements. They are present in ~2700 copies in the human genome (Wang et al., 2005), and represent the evolutionarily youngest, currently active family of human retrotransposons. SVAs may impact the host genome through a variety of mechanisms and sporadically generate disease-causing insertions in human DNA.

The origin of SVA elements can be traced back to the beginnings of hominid primate evolution, about 18-25 Mya (Xing et al., 2006). Starting at the 5′-end, a full-length SVA element is composed of a $(CCCTCT)_n$ hexamer repeat region; an *Alu*-like region consisting of three antisense *Alu* fragments adjacent to an additional sequence of unknown origin; a *variable number of tandem repeats* (VNTR) region, which is made up of copies of a 36- to 42-bp sequence or of a 49- to 51-bp sequence (Ostertag et al., 2003), presumably derived from the SVA2 element found in Rhesus macaques and humans (Damert et al., 2009, Han et al., 2007, Jurka and Gentles, 2006); and a *short interspersed element of retroviral origin* (SINE-R)

region. The latter is derived from the 3′-end of the *env* gene and the 3′-LTR of the endogenous retrovirus HERV-K-10 (Ono et al., 1987). A poly(A) tail is positioned downstream of the predicted conserved polyadenylation signal AATAAA (Ostertag et al., 2003). SVAs are currently evolving in humans as indicated by the acquisition of the MAST2 sequence via an aberrant splicing event which resulted in the formation of the SVA subfamily termed SVA_{F1} (Damert et al., 2009), MAST2-SVA (Hancks and Kazazian, 2010), or CpG-SVA (Bantysh and Buzdin, 2009) (Fig.1A, B; sequences are shown in Supplemental dataset S1). To date, 84 members of the human-specific SVA_{F1} subfamily have been identified. Extensions of the MAST2 sequences located upstream of the 5'-truncated *Alu*-like region range from 14 to 382 bp. In this subfamily, SVA_F sequences range from 662 to 4255 bp. These 5'-transduced retrotransposons, in turn, may harbor different additional 5'-, 3'- transductions, or both (Bantysh and Buzdin, 2009, Damert et al., 2009, Hancks et al., 2011, Hancks and Kazazian, 2010). To date it is unknown, if the 5' transduced sequence has any effect on SVA retrotransposition efficiency. However, we hypothesize that CpG islands-encompassing *MAST2* sequences at the 5' end of SVA_{F1} source elements were and still are beneficial to the transcription of those SVA_{F1} elements and might function as positive transcriptional regulator.

We set out to investigate if the acquirement of the *MAST2* CpG-island played a role in the success of the SVA_{F1} subfamily, and if transcriptional activity of the *MAST2* gene correlates with the expression of functional mammalian non-LTR retrotransposons. The gene *MAST2* was reported to be highly active in mammalian testicular tissue (Walden and Cowan, 1993, Walden and Millette, 1996), although no thorough characterization of its promoter sequence has been performed yet. *MAST2* codes for a microtubule-associated serine/threonine-protein kinase 2 that functions in a multi-protein complex in spermatid maturation (Walden and Millette, 1996). Testis-specific *MAST2* RNA increases in abundance during prepuberal testis development, peaking at the spermatid stage. According to publically available tissue array data, *MAST2* is also expressed in the heart and the central nervous system (UCSC Browser data on MAST2

transcription in human tissues, http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc001cov.2&hgg_prot=Q6P0Q8&hgg_chrom=chr1&hgg_start=46269284&hgg_end=46501797&hgg_type=knownGene&db=hg19&hgsid=184239277).

In order to investigate any potential correlation between the transcriptional regulation of *MAST2* and the expression of functional mammalian non-LTR retrotransposons, we performed quantitative real-time reverse transcription PCR (qRT-PCR) assays on total RNA preparations from various human, mouse and rat tissues. We could confirm that human *MAST2* transcription peaks in testicular and heart tissues. *MAST2* was cotranscribed with the SVA_{F1} subfamily in human tissues. We found that in human testicles, *MAST2* transcription correlates also with the transcription pattern of non-LTR retrotransposon families L1, *Alu* and SVA, and with the expression of L1 ORF1 protein. We show that the 324 bp-long *MAST2* sequence that is part of SVA_{F1} subfamily members act as positive transcriptional regulators in the human testicular germ cell tumor cell line Tera-1. Therefore, we conclude that the acquisition of the *MAST2*-derived CpG islet might be beneficial to the expression of members of the SVA_{F1} subfamily that is characterized by an exceptionally efficient amplification process during the recent evolution of the human genome.

2. Material and Methods

2.1. *In silico* sequence analysis.

The consensus sequences of the mammalian REs were obtained from the Repbase Update database (http://www.girinst.org/repbase/update/index.html). Oligonucleotide primers were designed using GeneRunner and Primer 3 software. Homology searches against GenBank were performed using the BLAST web server at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For multiple alignments, BLAST pairwise search, Vector NTI and Clustal W programs (Thompson et al., 1994) were used. Putative RNA secondary structure features were predicted using RNAfold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

2.2. Oligonucleotides.

Oligonucleotides were purchased from Evrogen (Russia) and their sequences are listed in supplementary table 4.

2.3. Tissue samples.

Samples of human tumor (seminoma and bladder cancer) were provided by P.A.Hertzen Moscow Oncological Reserach Institute. Tumor tissues were sampled from surgical specimens with testicular germ cell tumors (defined as testicular samples T3 and T4 in the text) and bladder cancer under non-neoplastic conditions. Representative samples were divided into two parts, one of them being immediately frozen in liquid nitrogen, and the other formalin-fixed and paraffin-embedded for histological analysis. Normal human tissue samples were taken from the adult (19-43 years old) donors killed in road accidents up to 24 hours after death. The biosampling manipulations were done according to European Union ethical guidelines and approved by the local institutional ethical committees.

2.4. RNA isolation and cDNA synthesis.

Total RNA was isolated from frozen tissues using an RNeasy Mini RNA purification kit (Qiagen). A special effort was made to completely degrade residual DNA from the samples using RNase-free DNase (Sigma). cDNA synthesis was performed with 2μg of total RNA using random hexamer primer set (SibEnzyme, Russia) with the ThermoScript RT-PCR System (Invitrogen) or KAPA SYBR One-Step qRT-PCR kit (KAPA Biosystems) in 5% DMSO, according to the manufacturer recommendations. All RT-PCR experiments were performed against the negative RT "-" controls (no reverse transcriptase added at the stage of the first strand cDNA synthesis) to control for DNA contamination. Only those samples displaying negative results on the RT "-" control experiments were further analyzed.

2.5. RT-PCR and qRT-PCR.

All reverse transcription-PCR (RT-PCR) and quantitative real-time reverse transcription-PCR (qRT-PCR) experiments were reproduced at least three times, using independent cDNA

preparations. Prior to the assay, the priming capacities of the oligonucleotides were examined by genomic PCRs at various temperatures, depending on the primer combination used, with 40 ng of human, mouse or rat genomic DNA. Primer sequences for qRT-PCR (supplemental dataset S1) were chosen using Primer Express software (Applied Biosystems) and are presented in supplemental material (supplementary table 4).

qRT-PCR was performed with cDNA samples from normal and cancerous human testicular parenchyma, heart, brain, lung, ovaries and bladder with the equivalent of 20 ng total RNA being used as a template in each PCR, performed in a final volume of 40 µl. Five-microliter aliquots of the reaction mixture after 24, 27, 30, 33, 36, and 39 cycles of amplification were analyzed by electrophoresis on 1.2% agarose gels. In all cases, the transcriptional levels were normalized to the transcriptional level of the housekeeping gene *ACTB* encoding β-actin, using Beta-Actin RT-PCR Primer Set (Agilent). RT "-" and "no-template" controls were performed for each RT-PCR amplification. qRT-PCR assays were performed in quadruplicate using MxPro3000 thermocycler (Stratagene) and Eva-Green real-time PCR kit (SibEnzyme). Expression levels were normalized to *ACTB* expression using the following formulae: $2^{-\Delta Ct} = 2^{-(Ct\ sample\ -\ Ct\ refrence)}$ (Livak and Schmittgen, 2001).

2.6. Luciferase reporter gene assay.

For transient transfection experiments, the human testicular embryonal germ cell tumor cell line Tera1 (Jewett, 1978) was used. Cells were grown in DMEM/F12 (1:1) medium containing 10% fetal calf serum (Invitrogen) at 37°C and 5% CO_2.

Transfections were carried out in 25-cm^2 tissue culture flasks using Lipofectamine 2000 (Invitrogen) according to the manufacturer's recommendations. For each transfection, we used 0.5 µg of a 5:1 mixture of the analytical plasmid (including the expression plasmid in which the firefly luciferase gene is under transcriptional control of the upstream regulatory sequence of interest) and normalization plasmid pCMV_LacZ (carrying the β-galactosidase gene under transcriptional control of the CMV promoter). The normalization plasmid was used in all

experiments as internal control to minimize errors caused by differences in transfection efficiencies between the independent replicates. Luciferase transcription was scored only for the experiments with transfection efficiencies of 65-70%. Transfection efficiencies were measured by GFP fluorescence after cotransfection of the cells with the plasmid pTurbo-GFP (Evrogen).

The following luciferase reporter constructs were used:

-"pGL3-No promoter", identical to the promoterless pGL3 basic vector (Promega);

-"pGL3-L1 5'UTR", including the full-length 5' untranslated region of human L1Hs retrotransposon upstream the luciferase gene (Olovnikov, *et al.*, 2007), kindly provided by S. E. Dmitriev (Belozersky Institute of Physico-Chemical Biology, Moscow State University);

-"pGL3-CpG" in which the 324 bp-long fragment (supplemental dataset S1) covering the CpG islet that the *MAST2* exon 1 and by the SVA_{F1} family members have in common, was cloned into the promoterless pGL3 basic vector upstream the luciferase reporter gene using *MluI* and *HindIII* restriction sites. Prior cloning, the fragment was PCR-amplified using human genomic DNA as the template, with the primer pairs: CSPfor, 5'-*GGCTGGGGTCGGTGGCTTAG*-3', and CSPrev, 5'-*AGGCAGGCGGCTGCTCCTTG*-3';

-"pGL3-5'SVA" carrying the 5' terminal fragment that corresponds to the consensus sequence of the SVA_F family members (supplemental dataset S2) which is located upstream of the luciferase reporter gene which was cloned using *MluI* and *HindIII* restriction sites. Prior to cloning, the fragment was PCR-amplified from human genomic DNA using the primer pair: ASVAfor, 5'-*TTCCTGAGGCATTGAGTGCTAATAAATAATAAATGC*-3', and ASVArev, 5'-*GCTGGGGATGTGTAGGTTGTAGTGAGC*-3'. The PCR fragment was cloned and sequenced.

The sequences placed upstream of the luciferase reporter gene are identical to the respective genomic queries, as confirmed by Sanger-sequencing. 48-hours after transfection, cells were lysed and total RNA was isolated using the RNeasy kit (QIAGEN). After DNase treatment, the cDNA was synthesized as described above. Luciferase transcription was

quantified by qRT-PCR with the luciferase gene-specific primers LucF (5'-*AGAATCGTCGTATGCAGTGAAAAC*-3') and LucR (5'-*CTTTAGGCAGACCAGTAGATCCAGA*-3'). *LacZ* transcription was measured by qRT-PCR with the primers LZF (5'-*CCAGCTGGCGTAATAGCGAAGAG*-3') and LZR (5'-*GTCAAATTCAGACGGCAAACGAC*-3'). Cycling conditions were as follows: initial denaturation for 10 min at 95C, followed by a three-step profile: denaturation for 30 sec at 95˚C, annealing for 30 sec at 60˚C, and extension for 1 min at 72˚C. qRT-PCR reactions were performed in a 25 μl reaction mixture containing 10 pM of each of the two oligonucleotides, cDNA first strands and 1x qPCRmix-HS SYBR mix (Evrogen). The obtained values for the firefly luciferase transcriptional products were normalized for *LacZ* transcription. All measurements were done at least in quadruplicate. Normalized transcriptional levels were calculated according to the formulae: $2^{-\Delta Ct} = 2^{-(Ct\ sample\ -\ Ct\ refrence)}$ (Livak and Schmittgen, 2001).

2.7. Bisulfite sequencing.

Bisulfite treatment was carried out using EpiTect kit (Qiagen) following the manufacturers' recommendations. Prior to bisulfite conversion, DNA isolated from human tissues was digested with *EcoRI* endonuclease. In all the instances, two or more independent, duplicate bisulfite experiments were performed. Bisulfite-treated DNA was then PCR-amplified with the following primer sets:

MS, specific to the unique fragment of the CpG island present only in gene *MAST2*, but absent from SVA_{F1} elements: UMF1 (5'-*TTTGGTTTAGTAGTTTTATGAATG*-3'), UMF2 (5'-*GATGAATGAAAGATTTAATG*-3'), UMR1 (5'-*CCTTCTCCCACCCAACAC*-3'), UMR2 (5'-*CCTACCTACAACCAACCT*-3').

For the first PCR performed with the primers UMF1 and UMR1, the thermocycling conditions were as follows: initial denaturation for 10 min at 95C, followed by a three-step profile: denaturation for 30 sec at 95˚C, annealing for 30 sec at 50˚C, and extension for 1 min

at 72˚C, for 20 cycles. The nested PCR with the primers UMF2 and UMR2 was carried out under the same conditions, for 30 cycles.

CS, specific to the common fragment of the *MAST2* CpG island present also in the SVA_{F1} elements: CMF1 (*5'-GGTTTTAGTTGTAGATATGAAG-3'*), CMF2 (*5'-GGAGTTGTTTTAGTTTTTG-3'*), CMR1 (*5'-CACTTCCCAAATAAAATAAC-3'*), CMR2 (*5'-CTCACTTCCTAAATATAATAAC-3'*). For the first PCR, done with the primers CMF1 and CMR1, the thermocycling conditions were as follows: initial denaturation for 10 min at 95˚C, followed by a three-step profile: denaturation for 30 sec at 95˚C, annealing for 30 sec at 50˚C, and extension for 1 min at 72˚C, for 20 cycles. The nested PCR with the primers CMF2 and CMR2, was carried out under the same conditions, for 30 cycles.

The nested PCR products were agarose gel-purified using Wizard gel and PCR clean-up system (Promega) and ligated into the pGEM-T easy vector (Promega) according to the manufacturer protocol, followed by the cloning in *E. coli* and sequencing of plasmid preparations from the individual clones. In order to determine the methylation statuses of the individual CG dinucleotides, the sequence data were analyzed using the BiqAnalyzer software (Which Company????).

2.8. Immunoblot analysis.

Testicular tissue samples were homogenized in PBS-8M Urea-Proteinase Inhibitor Cocktail. Protein concentrations were measured in cleared lysates according to the Bradford protocol. The protein lysates were boiled in Laemmli buffer, loaded on 12 % polyacrylamide gels (8,3 mcg of total protein/lane), subjected to SDS-PAGE, and electroblotted onto nitrocellulose membranes. After protein transfer, membranes were blocked for two hours at room temperature in a 10% solution of non-fat milk powder in 1x PBS-T (137 mM NaCl, 3mM KCl, 16.5 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.05 % Tween 20 (Sigma)), washed in 1xPBS-T, and incubated overnight with the respective primary antibody at 4°C. To detect L1 ORF1p, the polyclonal rabbit-anti-L1ORF1p antibody (Raiz et al., 2012) was used in a 1:2000 dilution in

1xPBS-T containing 5% milk powder. Membranes were washed thrice in 1xPBS-T and incubated with an HRP-conjugated, secondary anti-rabbit IgG antibody (Amersham Biosciences) at a dilution of 1:30,000 in 1xPBS-T/5% milk powder for 2 hours. Subsequently, the membrane was washed 3 times for 10 min in 1xPBS-T. ß-actin expression was detected using a monoclonal anti-ß-actin antibody (clone AC-74, Sigma-Aldrich) as primary antibody at dilution of 1:30,000. Anti-mouse HRP-linked species-specific antibody (from sheep) at a dilution of 1:10,000 served as secondary antibody specific for anti-ß-actin antibody. Immunocomplexes were visualized using lumino-based ECL immunoblot reagent (Amersham Biosciences).

2.9. Statistical analysis.

Statistical tests were done using Statistica 7.0 software. Correlation analyses were performed using linear regression model. Graphs and diagrams were built using Microsoft Excel program.

3. Results and discussion

3.1. Transcriptional activities of the *MAST2* gene and the SVA RE family in human tissues.

To address the question if transcriptional activities of the *MAST2* gene and of those members of the SVA_{F1} subfamily that include *MAST2* sequences are correlated, we firstly quantified the levels of RNA encoded by the *MAST2* gene in normal and cancerous human testicular and bladder tissues as well as in brain, heart, lung, ovary by performing qRT-PCR assays. We detected increased *MAST2* transcript levels in normal testicles, brain and heart (Fig.2), which is in agreement with the publicly available microarray gene expression data showing that *MAST2* transcription reaches maximum levels in testicles, brain and heart in both human and mouse (UCSC Browser data on MAST2 transcription in human tissues, UCSC Browser data on MAST2 transcription in mouse tissues, http://genome.ucsc.edu/cgi-

bin/hgGene?hgg_gene=uc008ugn.1&hgg_prot=NP_001036208&hgg_chrom=chr4&hgg_start=115979366&hgg_end=116136788&hgg_type=knownGene&db=mm9&hgsid=184826333).

At least 84 members of the SVA_{F1} subfamily harbor a sequence at their 5'-end that is derived from the CpG islet of the *MAST2* gene (Fig.1). To amplify the junction between 5'-transduced MAST2 sequences and the SVA_F sequence of the retrotransposon, we designed a qRT-PCR based assay using primers that are specific for the 5'-transduced MAST2 sequence and the SVA_F part of the SVA_{F1} retrotransposons (Fig.1C; primers Cf + Cr). We found that on the panel of 16 RNA samples that were isolated from seven different human tissues, transcription of the *MAST2* gene correlated with the transcription of those SVA_{F1} subfamily members that include MAST2 sequences with the correlation coefficient 0.706, p=0.00224 (Fig.2). The data obtained using EST database analysis supported the finding that the SVA_{F1} family is transcriptionally active. Eleven human ESTs covered the junctions between 5'-transduced MAST2 sequences and the adjacent 5'-truncated Alu-like region that is specific for SVA_{F1} subfamily members (Damert et al., 2009). Among these ESTs, three represented transcripts from testis, three minus-RNAs from prostate, and the others were unique RNAs from embryonic stem cells, spleen, lung, thymus and T-lymphocytes. This data does not contradict the hypothesis that the SVA_{F1} family is transcriptionally upregulated in testis, similarly to *MAST2* gene.

In order to investigate overall expression of the entire SVA retrotransposon family in human tissues, we designed primers specific for the sequences of the SINE-R region that all SVA subfamilies have in common (Fig.1B,C; primers Sf + Sr). qRT-PCR assays uncovered that there is no correlation between the overall transcriptional activity of SVA elements and the transcription of the *MAST2* gene or SVA_{F1} subfamily members carrying MAST2 sequences in the majority of tested human tissues (Fig.2), correlation coefficient; p-value for *MAST2*-SVA and SVA_{F1}-SVA comparison, respectively: 0.401; 0.12 and 0.048; 0.86. Taken together, transcriptional activities of the *MAST2* gene and the SVA_{F1} subfamily members correlate,

while the overall transcription of all SVA elements does not. This suggests that the *MAST2* gene and those members of the SVA_{F1}-subfamily that include MAST2 sequences are transcriptionally controlled by the same MAST2-derived regulatory sequences.

3.2. Functional tests of *MAST2*-derived sequences of SVA_{F1} subfamily members.

In the case of SVA_{F1} subfamily members, 5' terminal SVA sequences were replaced by exon 1 of the *MAST2* gene as a consequence of an ancient aberrant splicing event. To compare promoter activities of MAST2-derived sequences and of the "authentic" SVA_F 5'-terminal regions, we performed luciferase reporter assays in Tera-1 cells (Fig.3). In pGL3-CpG (Fig.3B), the luciferase reporter gene was set under transcriptional control of the 324-bp fragment covering the CpG islet that *MAST2* exon 1 and a subset of SVA_{F1} family members have in common (supplemental dataset S2). In parallel, we generated pGL3-5'SVA, in which the luciferase gene was set under transcriptional control of the same 5'-terminal SVA_{F1} sequence that was replaced by the *MAST2* exon 1 sequence as aconsequence of an ancient splicing event. The 328-bp fragment was amplified from the genomic SVA_F locus 3q21.3 (supplemental dataset S4, Fig.3A). As negative control, we used the promoterless, empty luciferase reporter vector pGL3-Basic (Promega). A reporter construct in which the luciferase gene was controlled by the 900-bp L1Hs 5'UTR served as positive control (Olovnikov et al., 2007).

Since both *MAST2*- and SVA_F- derived sequences that were inserted upstream of the luciferase reporter gene contain multiple ATG start codons which can interfere with the luciferase protein production, we used qRT-PCR to quantify luciferase mRNA production. We used the *LacZ* gene controlled by the human CMV promoter on plasmid pCMV_lacZ as reference gene for normalization of luciferase expression in Tera-1 cells that were transiently cotransfected with pCMV_lacZ and pGL3-CpG, pGL3-L1 5'UTR or pGL3-Basic, respectively. Total RNA was isolated from the differently cotransfected Tera-1 cells to

quantify luciferase- and *LacZ* transcript levels. We found that the 324-bp *MAST2* sequence had a promoter activity, which was about two-fold higher than the activity of the 5' terminal 328-bp fragment of the canonical SVA_F element (Figure 3C). This indicates that the acquisition of the *MAST2*-derived CpG islet increased the SVA_{F1} transcriptional activity in germ cells and, therefore, might be beneficial for the propagation of the SVA_{F1} family in the human genome. The *MAST2*-derived fragments of SVA_{F1} subfamily members contain numerous putative transcription factor binding sites (Supplemental figure 1) which might play a role in the transcriptional upregulation of SVA_{F1} subfamily members.

After we had quantified MAST2 transcripts in human germ line tissues (testicles and ovaries, Fig. 4A), we next tested the methylation status of the CpG island encoded by exon 1 of the *MAST2* gene and by the *MAST2* sequences acquired by members of SVA_{F1} subfamily (Fig. 4B). To this end, we separately mixed genomic DNA preparations isolated from three human ovarian samples, four human testicular samples expressing *MAST2* at relatively high levels (>8% of beta actin gene *ACTB* transcription), and four human testicular samples expressing *MAST2* at low levels (<0,5% of *ACTB* transcription), respectively (Figure 4A). We found that the CpG island located in the *MAST2* gene was almost completely unmethylated in all of the examined tissue samples, including the mixed sample of human ovaries, where the *MAST2* gene itself was transcribed at an especially poor level (Fig.4C). In contrast, the *MAST2*-derived CpG-islet in SVA_{F1} elements was heavily methylated in samples where *MAST2* gene expression was low (ovaries, mixed testicular sample 1), and slightly methylated when the *MAST2* transcription was high (mixed testicular parenchyma sample 2). These data suggest that the *MAST2*-derived CpG islets captured by the source element of the SVA_{F1} subfamily may play an important functional role in the transcriptional activation of many members of this subfamily. Hypomethylation of the *MAST2* CpG-islet in the tissues that do not express the *MAST2* gene itself may represent a rather typical situation for most of the human CpG islands

which are normally not methylated independently on the transcriptional status of the neighboring genes.

3.3. Transcriptional correlation between active retrotransposon families and the *MAST2* gene in human testicular tissue.

Testicular tissue samples were obtained from eight different adult human donors (Fig. 4A). *MAST2* transcription levels varied heavily among the tested samples, which might reflect interindividual variability in proportions of germ cells in the different tissue specimens. In contrast to the data obtained using a panel of seven different human tissues, examination of transcription levels on a panel of human testicular parenchyma tissue samples revealed correlating expression patterns for the *MAST2* gene and the total set of genomic SVA elements, as indicated by a correlation coefficient of 0.9701 (Fig.5A). We also compared transcriptional activities of *MAST2* and the remaining transpositionally active human retrotransposon families *Alu* and L1. Primers were designed to specifically bind to the conserved regions of the consensus sequences of these retrotransposon families. This way, we facilitated the amplification of most human *Alu* and L1 subfamilies. Consistent with the previous findings, we detected a strong positive correlation between the transcriptional activities of *MAST2* and both *Alu* and L1 elements in human testicular samples reflected by the correlation coefficients greater than 0.94 (Fig.5B-D; Supplementary Tab.1). The method that was used to quantify human *Alu* sequences does not distinguish between RNA polymerase II and III transcripts. This is important, because authentic SINE RNAs are transcribed by RNA polymerase III, although SINE sequences can also be cotranscribed by cellular RNA polymerase II, are therefore abundant in the mRNA fraction and can easily be amplified by RT-PCR (Shaikh et al., 1997). Interestingly, the strongest transcriptional correlation between *MAST2* and other retrotransposons was detected in testicular tissue samples where *MAST2* was transcribed at higher levels. The removal of the two “outlier” samples exhibiting the highest *MAST2*

transcription levels resulted in a decrease of the correlation coefficients for the pairs *MAST2* – SVA, *MAST2* –Alu, *MAST2* –L1 ORF1 and *MAST2* –L1 3'UTR to the values of 0.81, 0.31, 0.20 and 0.42, respectively.

3.4. Biological significance of coexpression of the *MAST2* gene and human non-LTR retrotransposons.

The relatively young SVA_F subfamily has expanded about 3.18 Myrs ago, after the human and chimpanzee divergence (~4-6 Mya) (Wang, H. et al., 2005). At the time of human-chimpanzee radiation, the ancestral genome comprised more than 2.000 SVA copies, which are shared now by the human and chimpanzee genomes. In terms of genomic proliferation, the evolutionary young SVA_{F1} subfamily should be considered a very successful one: the offspring of only one among approximately 2000 SVA copies that resided in human DNA at that time (i.e. < 0.05%) has generated at least 84 new fixed insertions (~10% of all 860 human specific SVA elements) (Bantysh and Buzdin, 2009) which indicates exceptionally high genomic proliferation and fixation rates. Moreover, there may be numerous additional genomic SVA copies derived from the SVA_{F1} source elements that have been 5' truncated and, therefore, lost any distinguishing 5′-transduced *MAST2* sequence. Interestingly, it has been discovered recently that the human lineage displays significantly lower methylation of SVA elements in sperm compared to the chimpanzee lineage (Molaro et al., 2011), which might be at least one reason for the expansion of the SVA_{F1} subfamily in humans. Finally, individual SVA_{F1} subfamily members are characterized by specific 5'- and 3' transductions which were suggested to alter SVA transposition efficiencies by mechanisms other than transcriptional regulation (Bantysh and Buzdin, 2009, Damert et al., 2009, Hancks et al., 2011). For example, the acquisition of an *Alu*Sp element by 3' transduction was recently demonstrated to increase the *trans*-mobilization frequency of the respective SVA_{F1} elements by up to ~25 fold (Damert

et al., 2009, Hancks et al., 2011), most likely by making the 3'-transduced SVA_{F1} element a better target for the L1 protein machinery.

Gaining *MAST2*-derived regulatory sequences might be beneficial for the propagation of the SVA_{F1} subfamily as it guarantees SVA_{F1} expression and subsequent retrotransposition in the germ line which facilitates fixation of the new SVA insertions in the genome. We demonstrate that in human Tera-1 cells, the *MAST2*-derived sequences of SVA_{F1} elements display a promoter activity which is similar to that of the human L1 5-UTR. We show that *MAST2* gene expression in human testicular tissues correlates with transcriptional levels of human retrotransposon families SVA, L1 and *Alu*. Transcriptional coactivation of L1, *Alu* and SVA with the *MAST2* gene cannot simply be explained by any structural similarities shared between these elements and the *MAST2* gene (Schumann et al., 2010, Kramerov and Vassetzky, 2005). However, there may exist common patterns in binding of some specific transcriptional factors, which will be a matter of future studies. An alternative explanation could be that the transcriptional activation of the *MAST2* gene might be directly or indirectly associated with derepression of mammalian retrotransposons, *e.g.* by the mechanism including demethylation of retrotransposon copies within the host genomic DNA.

MAST2 transcription culminates during the first meiotic division, and its intracellular RNA concentrations remain highly elevated on the following stages/cell types: pachytene spermatocytes, round spermatids, and residual bodies (Walden and Cowan, 1993, Walden and Millette, 1996). Consequently, male germ cells undergoing meiosis may exhibit the highest upregulation of both *MAST2* and SVA_{F1} transcription in testicles *in vivo.* The acquisition of the *MAST2* regulatory sequence by the SVA_{F1} retrotransposon was probably an advantageous modification, which provided the opportunity for the SVA_{F1} element to become activated in the adequate cell type at the same time, when also expression of the *trans*-mobilizing L1 protein machinery is guaranteed. Immunoblot analysis with an anti-L1 ORF1p antibody showed that the *MAST2* transcription is correlated with the production of the L1-encoded protein machinery

in testicular tissue *in vivo* (Fig.6). Since non-autonomous non-LTR retrotransposons, like SVA, require the presence of the L1-encoded protein machinery for their own proliferation, spatio-temporal coexpression of SVA elements with functional L1s in the germ line is obviously essential for their propagation.

4. Conclusions

We conclude that the acquirement of the *MAST2* CpG-island might play a positive role in the success of the SVA_{F1} subfamily. *MAST2* sequences at the 5' end of SVA_{F1} elements act as positive transcriptional regulator in human germ cells. We observed that in 16 tissue samples representing seven different human tissues, *MAST2* was cotranscribed with the members of the SVA_{F1} subfamily. Methylation status of the *MAST2*-derived sequences of SVA_{F1} elements reversely correlates with transcriptional activity of *MAST2*. Finally, in various testicular tissue samples we observed transcriptional correlation of *MAST2* with human retrotransposon families L1, *Alu* and SVA.

5. Acknowledgements

This work was supported by the Molecular and Cellular Biology Program of the Presidium of the Russian Academy of Sciences and by the grant 10-04-00593-a and 11-04-01752-a from the Russian Foundation for Basic Research and by the *Deutsche Forschungsgemeinschaft* [DA 545/2-1 to G.G.S.]. We thank Sergey Dmitriev for providing plasmid pGL3-L1 5'UTR. We are grateful to Peter Shegay and Dmitry Sokov who provided us with the human cancerous tissue samples.

References

Bantysh, O. B., Buzdin, A. A., 2009. Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. Biochemistry (Moscow). 74, 1393-1399

BLAST web server at NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi

Damert, A., Raiz, J., Horn, A. V. et al., 2009. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res. 19, 1992-2008

Dewannieux, M., Esnault, C., Heidmann, T., 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35, 41-48

Esnault, C., Maestre, J., Heidmann, T., 2000. Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24, 363-367

Gogvadze, E., Buzdin, A., 2009. Retroelements and their impact on genome evolution and functioning. Cell. Mol. Life Sci. 66, 3727-3742

Goodier, J. L., Kazazian, H. H., 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. 135, 23-35

Han, K., Konkel, M. K., Xing, J. et al., 2007. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science. 316, 238-240

Hancks, D. C., Goodier, J. L., Mandal, P. K. et al., 2011. Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum. Mol. Genet. 20, 3386-3400

Hancks, D. C., Kazazian, H. H., 2010. SVA retrotransposons: Evolution and genetic instability. Semin. Cancer Biol. 20, 234-245

Jewett, M. A., 1978. Testis carcinoma: transplantation into nude mice. Natl. Cancer Inst. Monogr. 49, 65-66

Jurka, J., Gentles, A. J., 2006. Origin and diversification of minisatellites derived from human Alu sequences. Gene. 365, 21-26

Kramerov, D. A., Vassetzky, N. S., 2005. Short retroposons in eukaryotic genomes. Int. Rev. Cytol. 247, 165-221

Livak, K. J., Schmittgen, T. D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25, 402-408

Mills, R. E., Bennett, E. A., Iskow, R. C. et al., 2007. Which transposable elements are active in the human genome? Trends Genet. 23, 183-191

Molaro, A., Hodges, E., Fang, F. et al., 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell. 146, 1029-1041

Olovnikov, I. A., Adyanova, Z. V., Galimov, E. P. et al., 2007. [A key role of the internal region of the 5'-untranslated region in the human L1 retrotransposon transcription activity]. Mol. Biol. (Mosk). 41, 508-514

Ono, M., Kawakami, M., Takezawa, T., 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res. 15, 8725-8737

Ostertag, E. M., Goodier, J. L., Zhang, Y. et al., 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am. J. Hum. Genet. 73, 1444-1451

Raiz, J., Damert, A., Chira, S. et al., 2012. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 40, 1666-1683

Repbase Update database, http://www.girinst.org/repbase/update/index.html

RNAfold software, http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Schumann, G. G., Gogvadze, E. V., Osanai-Futahashi, M. et al., 2010. Unique functions of repetitive transcriptomes. Int. Rev. Cell. Mol. Biol. 285, 115-188

Shaikh T.H., Roy A.M., Kim J. et al., 1997. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. J. Mol. Biol. 271, 222-234.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680

UCSC Browser data on MAST2 transcription in human tissues, http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc001cov.2&hgg_prot=Q6P0Q8&hgg_chrom=chr1&hgg_start=46269284&hgg_end=46501797&hgg_type=knownGene&db=hg19&hgsid=184239277

UCSC Browser data on MAST2 transcription in mouse tissues, http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc008ugn.1&hgg_prot=NP_001036208&hgg_chrom=chr4&hgg_start=115979366&hgg_end=116136788&hgg_type=knownGene&db=mm9&hgsid=184826333

Walden, P. D., Cowan, N. J., 1993. A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with microtubules of the spermatid manchette. Mol. Cell. Biol. 13, 7625-7635

Walden, P. D., Millette, C. F., 1996. Increased activity associated with the MAST205 protein kinase complex during mammalian spermiogenesis. Biol. Reprod. 55, 1039-1044

Wang, H., Xing, J., Grover, D. et al., 2005. SVA elements: a hominid-specific retroposon family. J. Mol. Biol. 354, 994-1007

Wei, W., Gilbert, N., Ooi, S. L. et al., 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429-1439

Xing, J., Wang, H., Belancio, V. P. et al., 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc. Natl. Acad. Sci. U S A. 103, 17608-17613

LEGENDS TO FIGURES

Figure 1. Schematic representation of the structures of the human *MAST2* gene, canonical full-length SVA retrotransposons and the precursor source element of the SVA_{F1} subfamily. (A) The *MAST2* gene harbours a CpG island at its 5' end that overlaps with its exon 1: ex, exon; arrows (Mr, Mf) represent PCR primers used to amplify *MAST2* transcripts by qRT-PCR. (B) A canonical full-length SVA retrotransposon consists of CCCTCT repeats, *Alu*-like-, VNTR and SINE-R regions and a polyA tail at its 3' end. SVA elements are flanked by TSDs ranging from 4-20-bp. Arrows (Sf, Sr) indicate SINE-R-specific primers. (C) Organization of the precursor source element of the SVA_{F1} subfamily. P2 and P3 indicate potential regulatory sequences that could be responsible for the generation of chimeric MAST2-SVA pre mRNAs (Damert et al., 2009). Arrows indicate primers (Cr, Cf) that bind upstream and downstream of the MAST2-SVA junction, respectively.

Figure 2. Relative transcription levels of the *MAST2* gene (A), SVA_{F1} subfamily members (B) and of the entire SVA retrotransposon family (C) in human tissues. Transcriptional activities were quantified relative to the endogenous *ACTB* (beta-actin) gene expression by qRT-PCR for the tissue samples taken from testicular parenchyma, heart, brain, lung, ovaries, normal and cancerous urinary bladder.

Figure 3. Luciferase reporter gene assay to assess the promoter activity of the *MAST2*-derived 324-bp fragment of SVA_{F1} subfamily members. In reporter plasmids pGL3-5'SVA (A) and pGL3-CpG (B), the firefly luciferase gene is under transcriptional control of the 328-bp SVA_F-derived sequence (Supplemental Dataset S4) and the 324-bp MAST2-derived sequence (Supplemental Dataset S3), respectively. (C) Luciferase transcriptional activity was quantified relative to lacZ gene expression by qRT-PCR on total RNA isolated from Tera-1

cells that were cotransfected with pCMV_lacZ and pGL3-5'SVA, pGL3-CpG, pGL3-Basic or pGL3-L1 5'UTR, respectively. In each case, qRT-PCR-measured luciferase transcription was normalized against the qRT-PCR-measured LacZ transcription level.

Figure 4. Comparison of the *MAST2* transcription with the methylation pattern of the MAST2 CpG-islet. (A) Transcription of the *MAST2* gene was quantified relative to *ACTB* gene expression by qRT-PCR in mixed human ovarian samples and testicular parenchyma samples, expressing the *MAST2* gene at low levels (mixed testicular sample 1), and at high levels (mixed testicular sample 2) (B) Schematic representation of SVA_{F1} family members and of the *MAST2* gene. (C) Methylation patterns were determined by bisulfite sequencing of the *MAST2*-derived fragments of SVA_{F1} elements (using primers Sf and Sr), and of the CpG-rich region specific for the promoter of the *MAST2* gene (primers Mf and Mr). For bisulfite sequencing of the CpG island of the *MAST2* gene, we analyzed a region of the MAST2 CpG island that is located exclusively on exon 1 of the MAST2 gene (Mf, Mr). To determine the methylation status of the CpG islet of SVA_{F1} elements, we probed a region that multiple copies of SVA_{F1} elements and the MAST2 gene have in common (Sf, Sr). Empty circles, unmethylated CG dinucleotides; Full circeles, methylated CG dinucleotides.

Figure 5. Correlation between mRNA levels of the *MAST2* gene and the RNA levels of functional retrotransposon families in human testicular parenchyma samples. Transcriptional levels were determined relative to ACTB gene expression by qRT-PCR. X-axis, transcription of the MAST2 gene. Y-axis, overall transcription of all genomic SVA subfamilies (A), Alu subfamilies (B) and L1 elements using primers specific for the L1 ORF1-coding region (C) and for the L1 3'UTR (D), respectively. Each panel represents experiments with the same testicular tissue samples T28, T30, T31, T33, T38, T39, T40, T41, T44 and T46. Only normal (non-tumor) testicular tissue samples were used for these analyses.

Figure 6. Endogenous L1 protein expression correlates with overall SVA and *MAST2* gene transcription levels. (A) Immunoblot analysis of cell lysates from testicles of seven different individuals (T81 – T87) using an anti-L1 ORF1p antibody. As loading control, an immunoblot anlysis was performed using an anti-β-actin antibody. Cell extracts from NTera2D1 cells served as positive control for L1ORF1p expression. (B) *MAST2* gene transcription levels were quantified relative to β-actin mRNA levels by qRT-PCR.

FIGURES

Figure 1

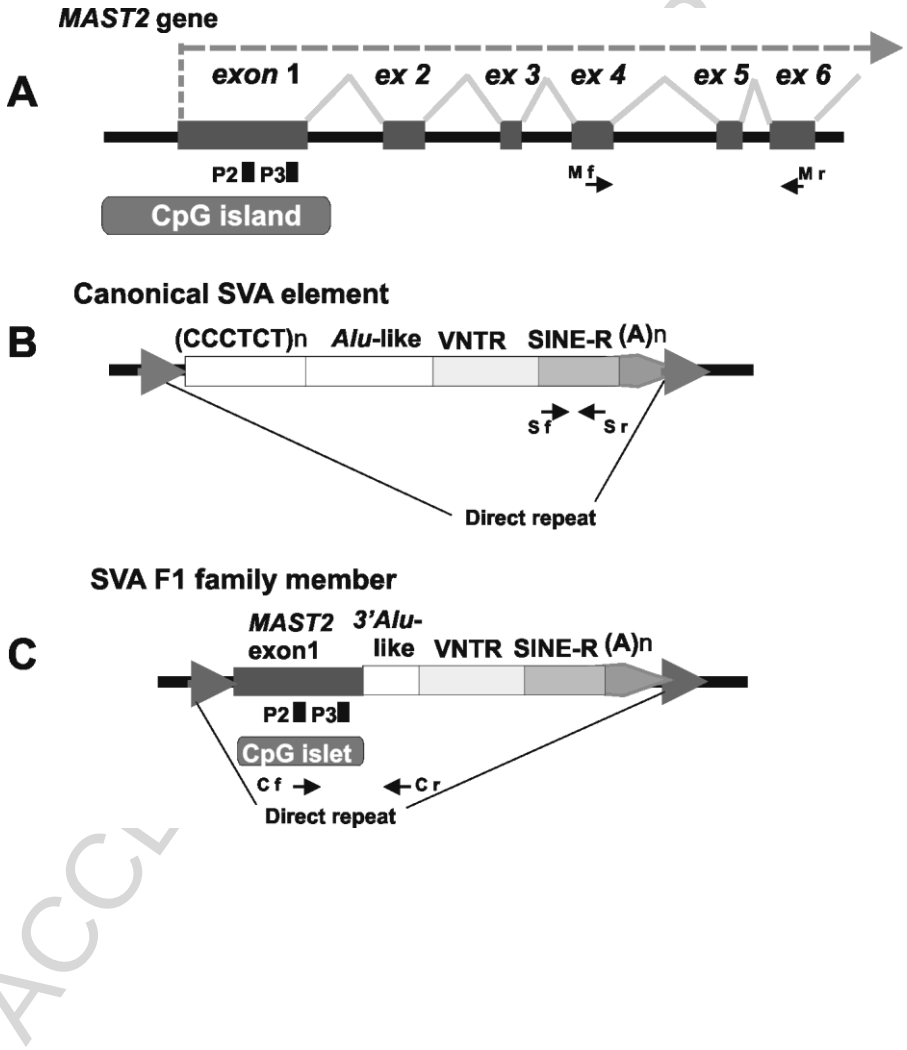


Figure 2

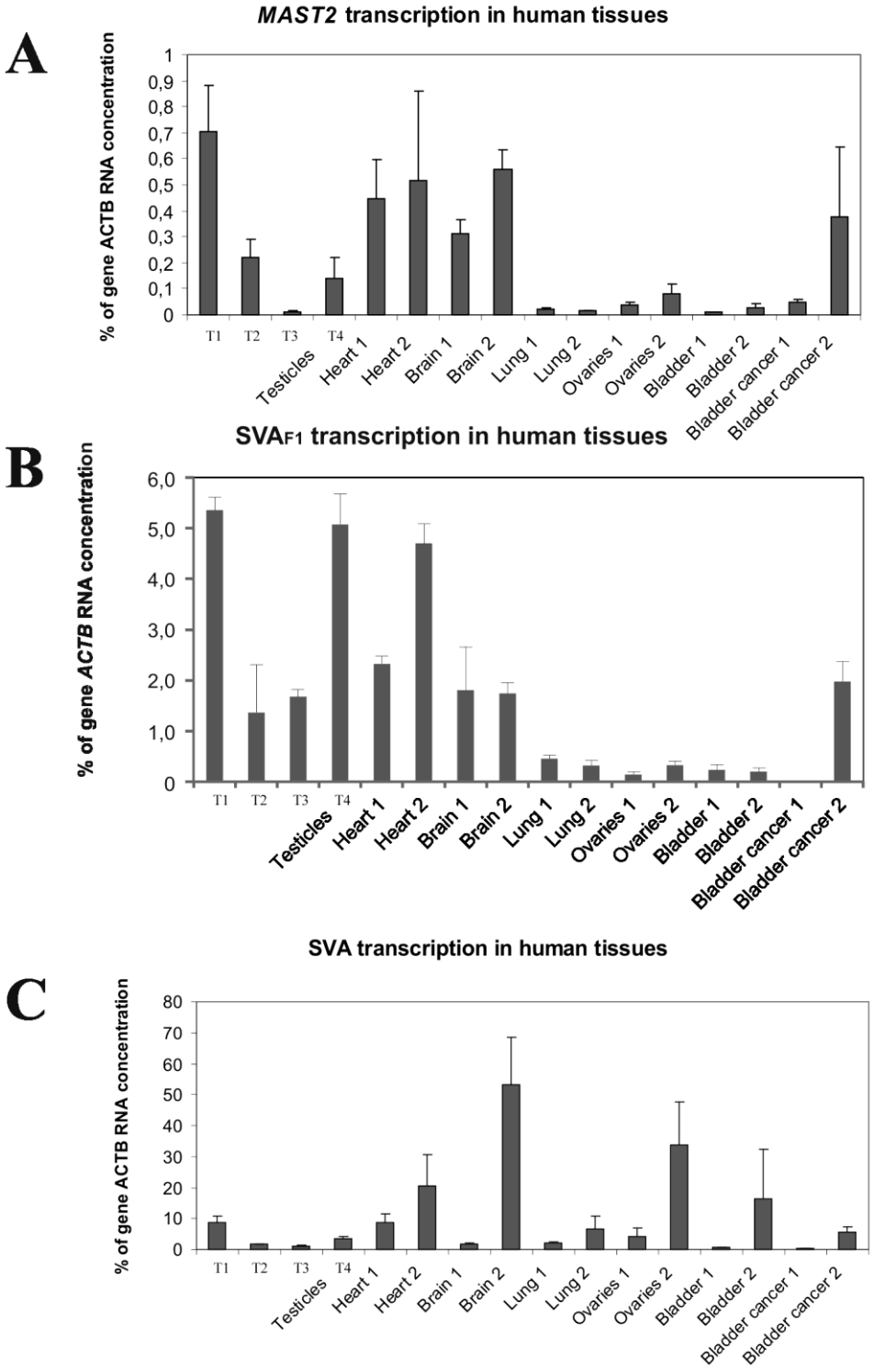


Figure 3

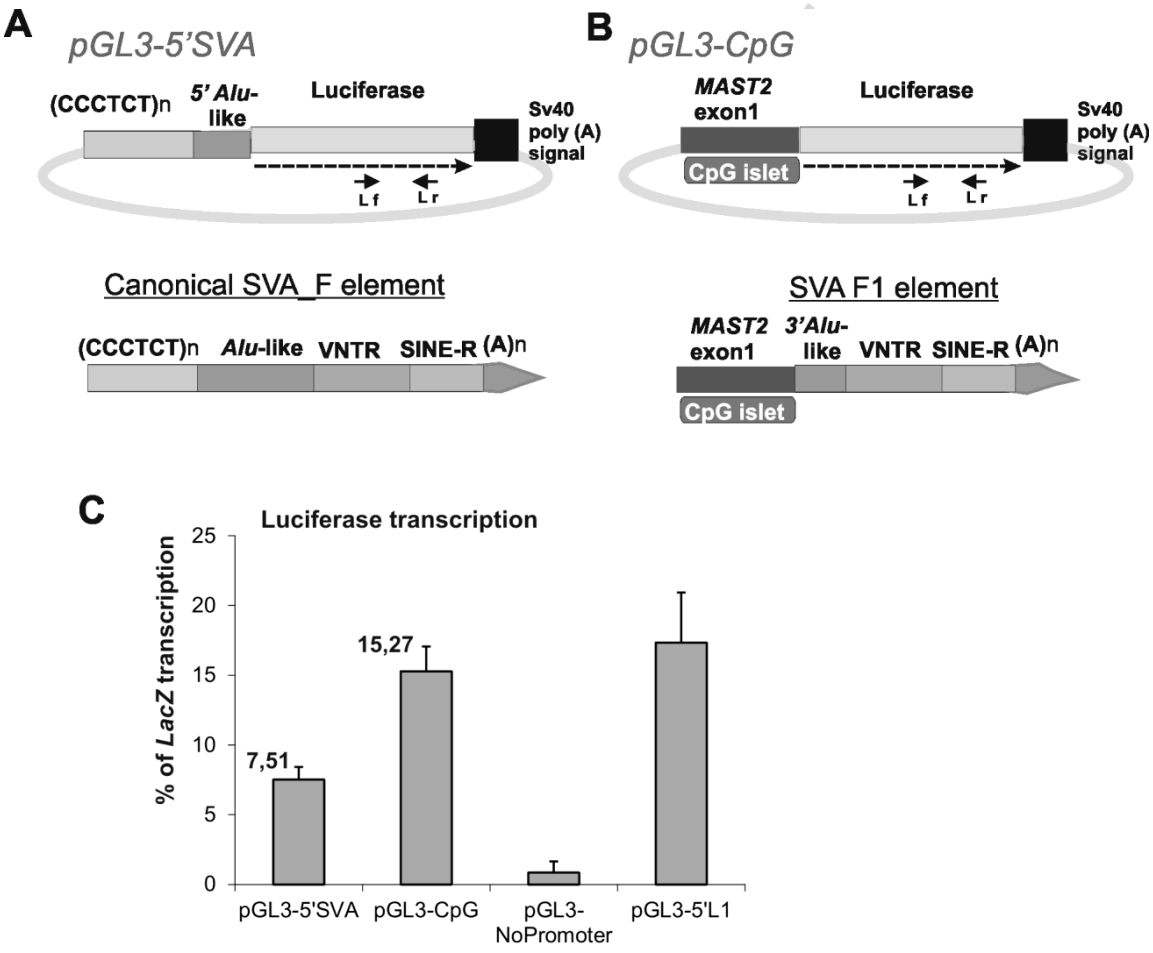


Figure 4

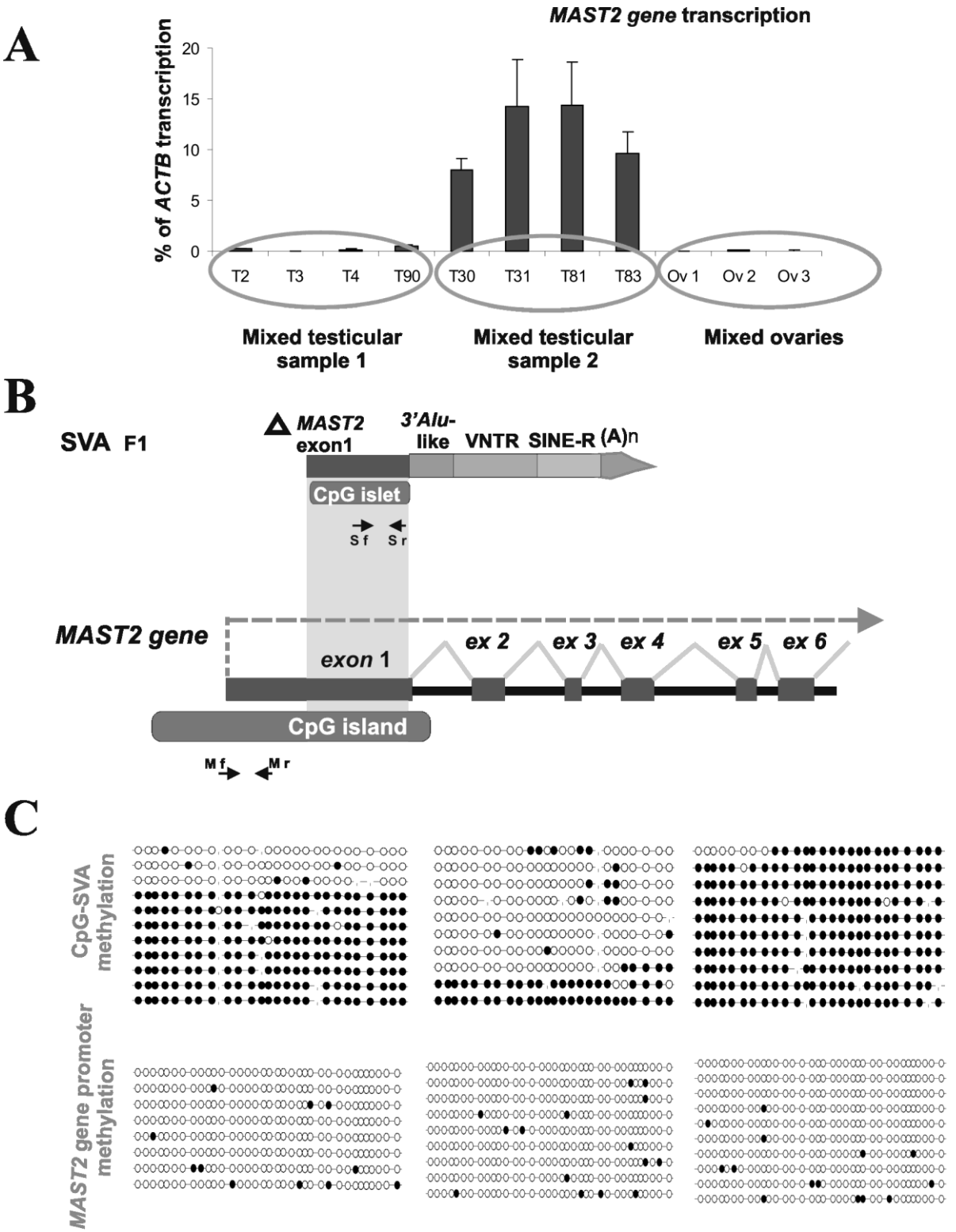

A
MAST2 gene transcription
% of ACTB transcription
20
15
10
5
0
T2
T3
T4
T90
T30
T31
T81
T83
Ov 1
Ov 2
Ov 3
Mixed testicular sample 1
Mixed testicular sample 2
Mixed ovaries
B
SVA F1
MAST2 exon1
3'Alu-like
VNTR
SINE-R
(A)n
CpG islet
Sf
Sr
MAST2 gene
exon 1
ex 2
ex 3
ex 4
ex 5
ex 6
CpG island
Mf
Mr
C
CpG-SVA methylation
MAST2 gene promoter methylation

Figure 5

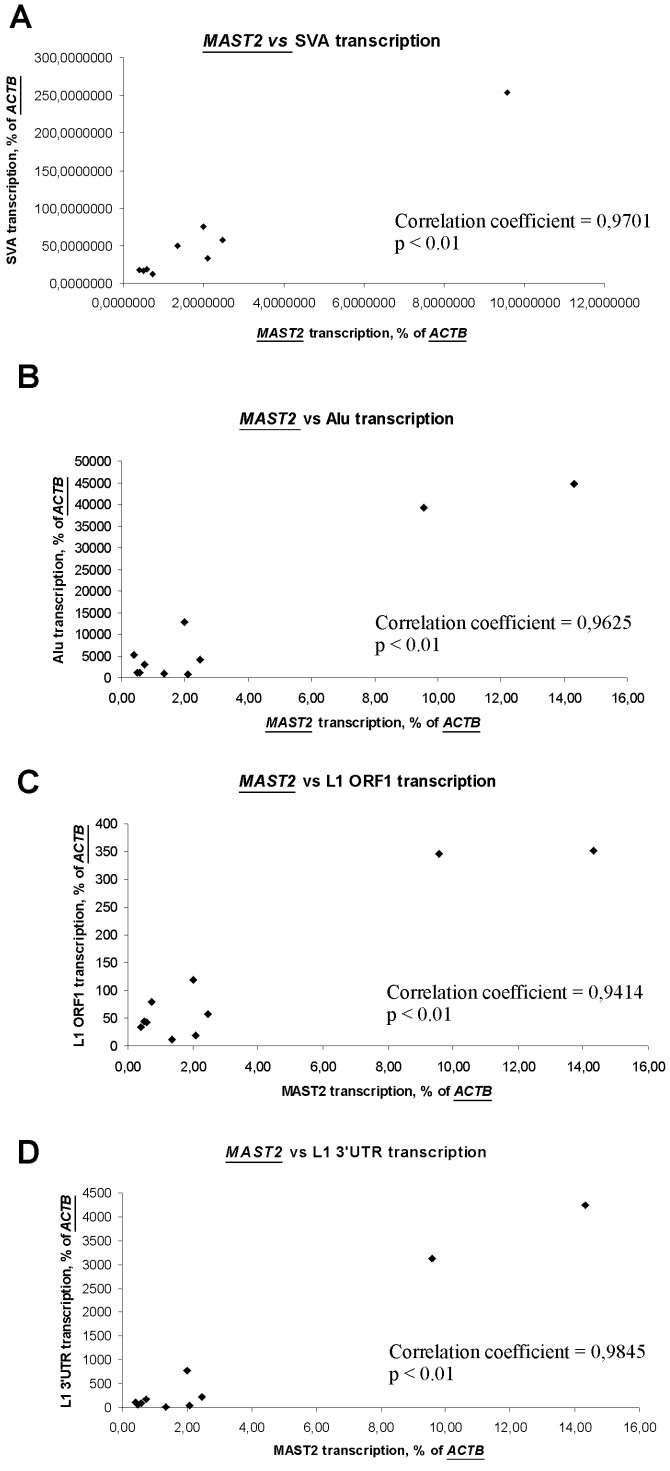
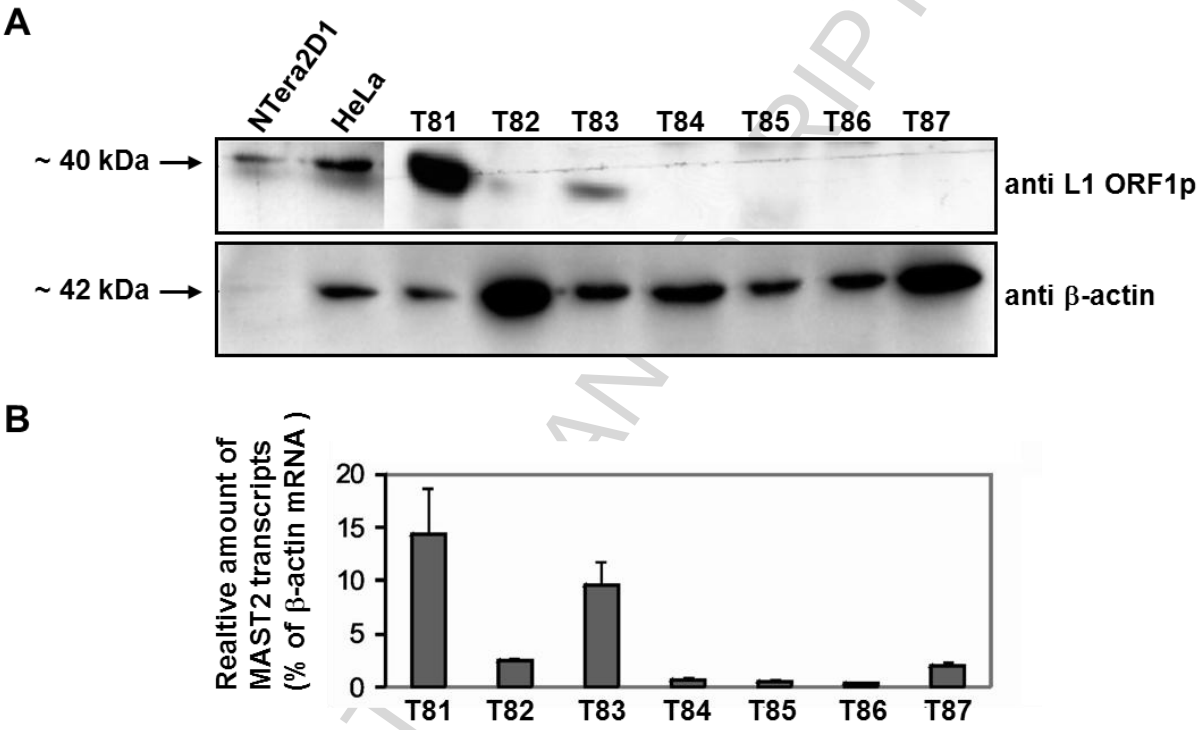


Figure 6



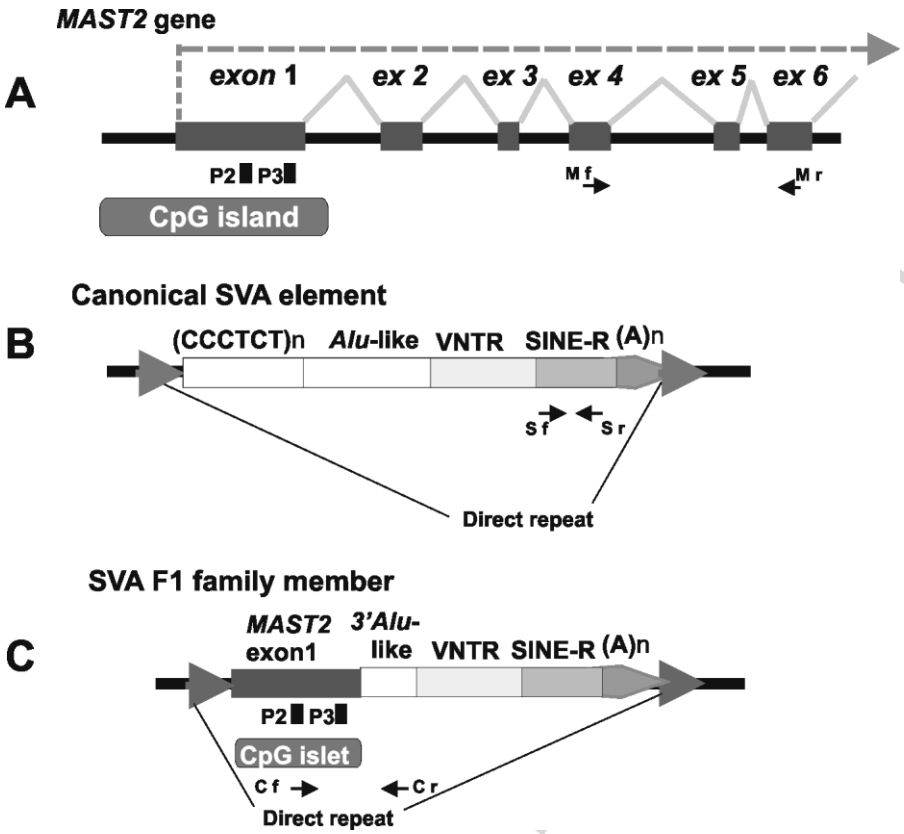


Fig. 1

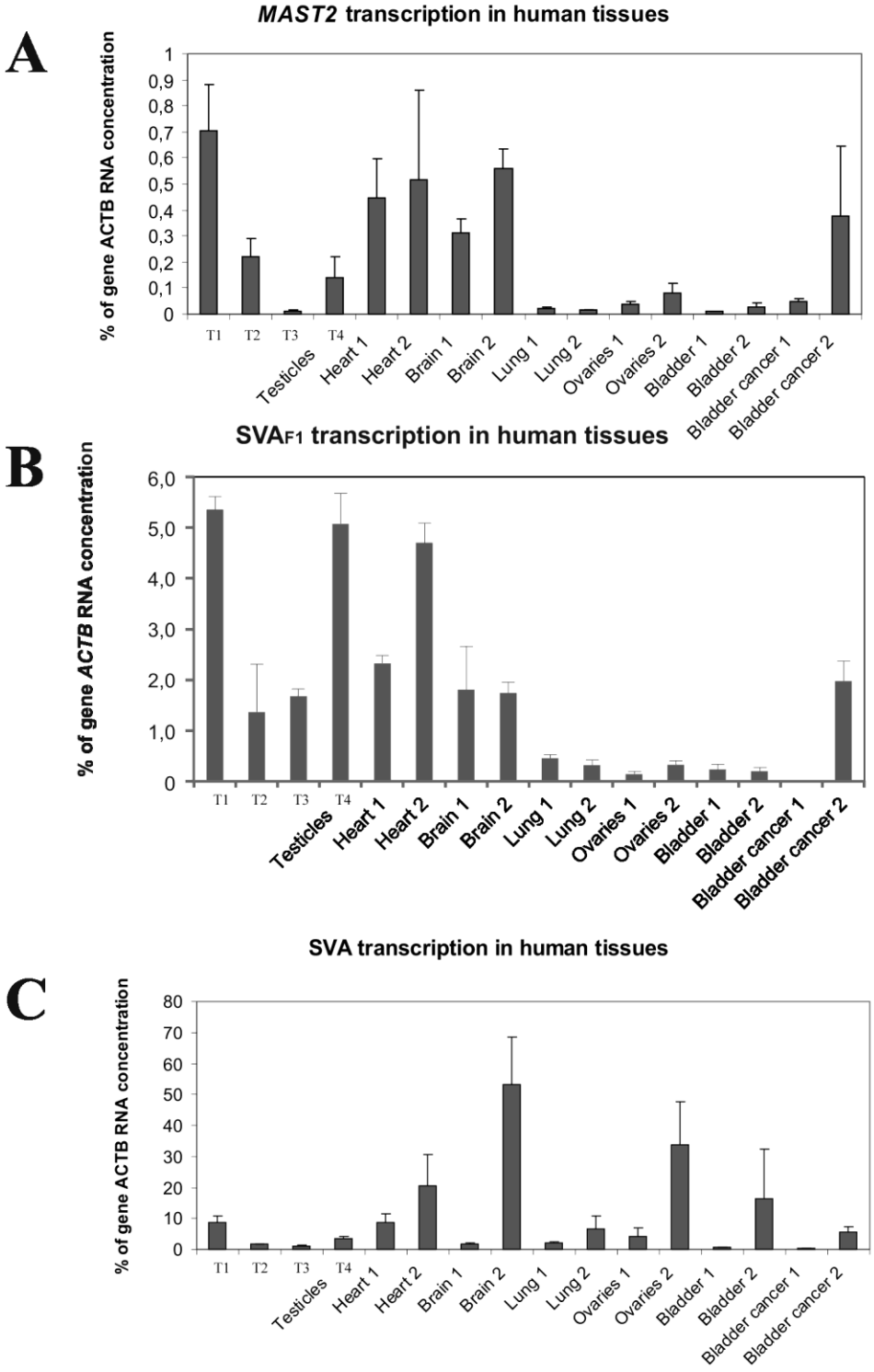


Fig. 2

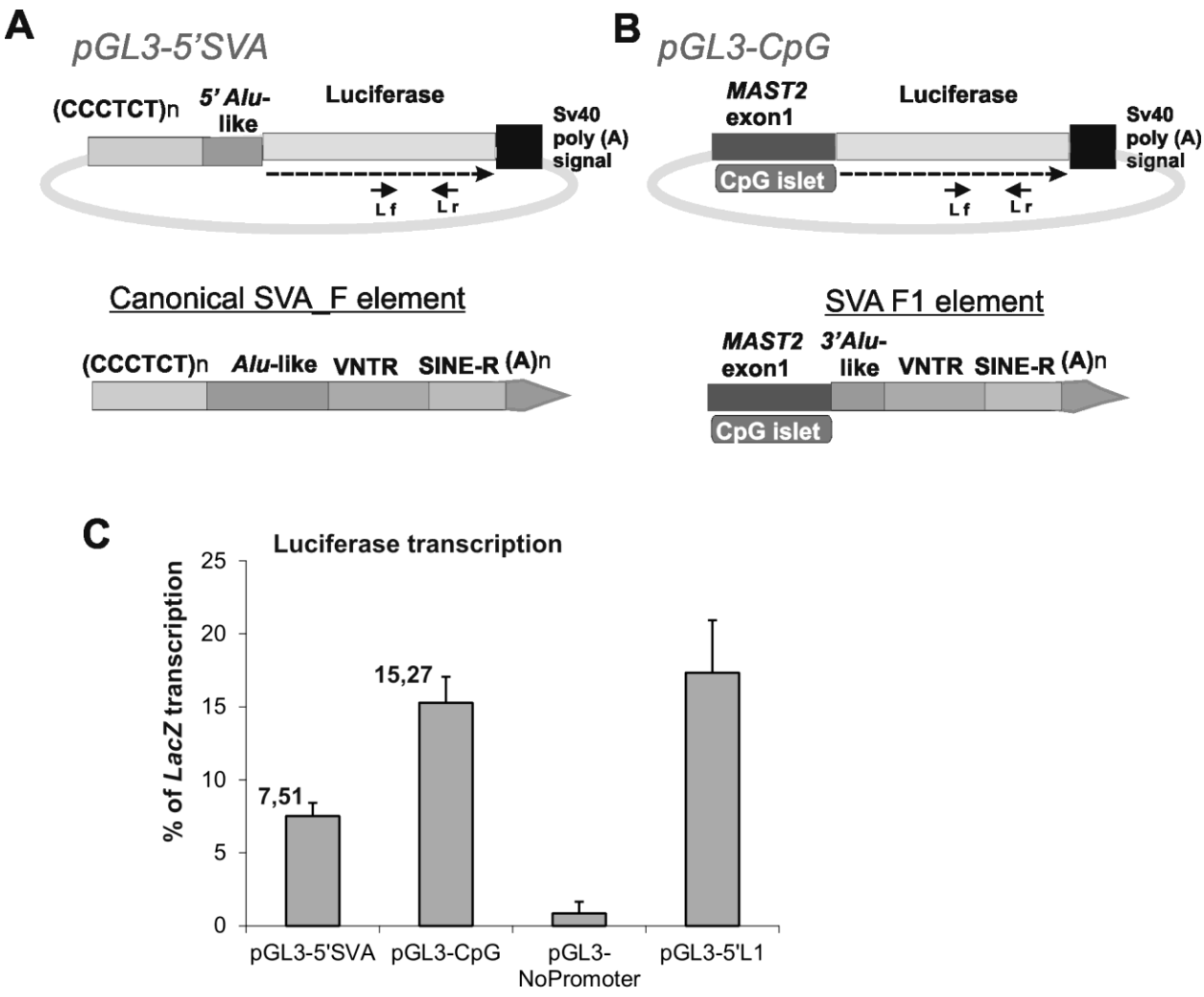


Fig. 3

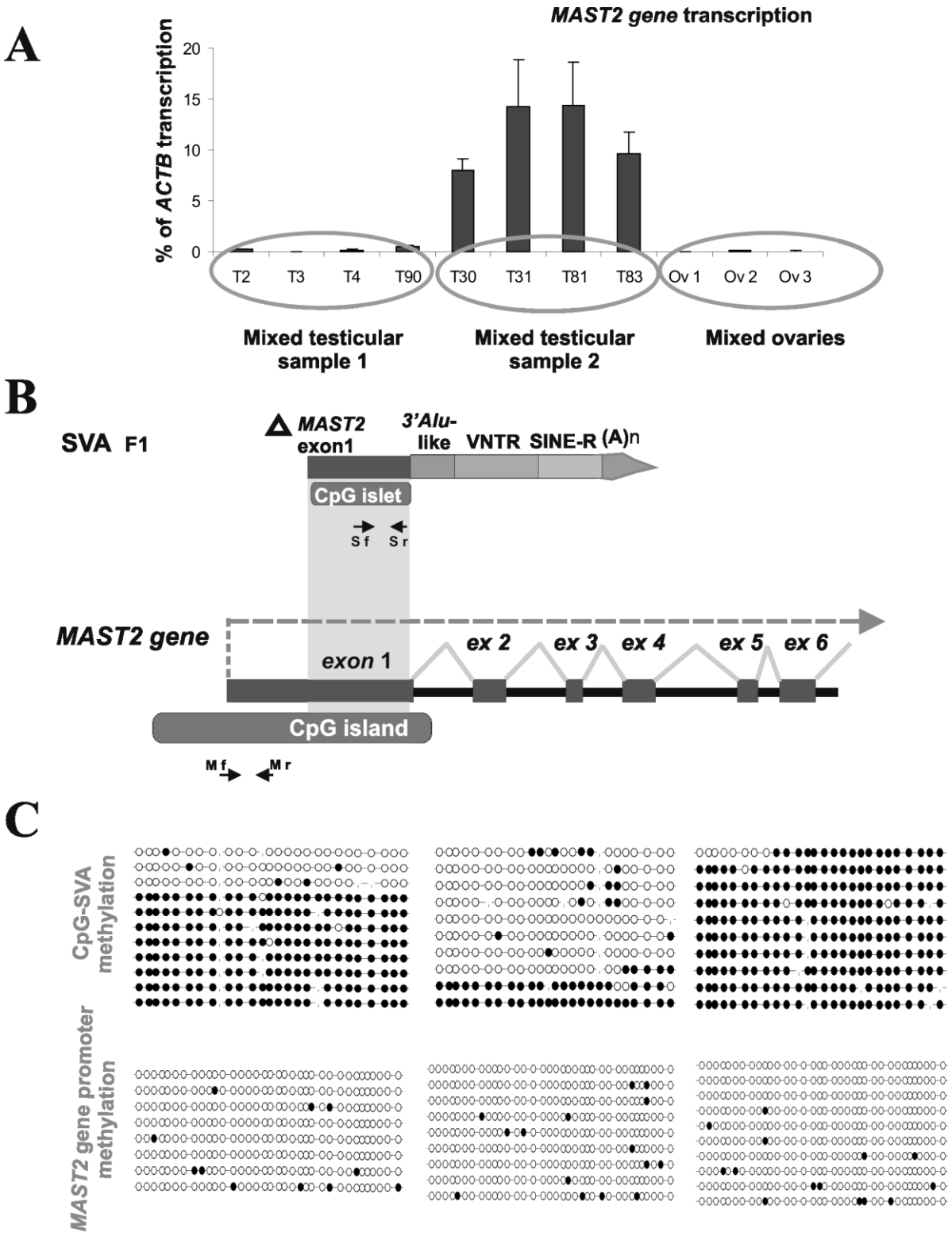


Fig. 4

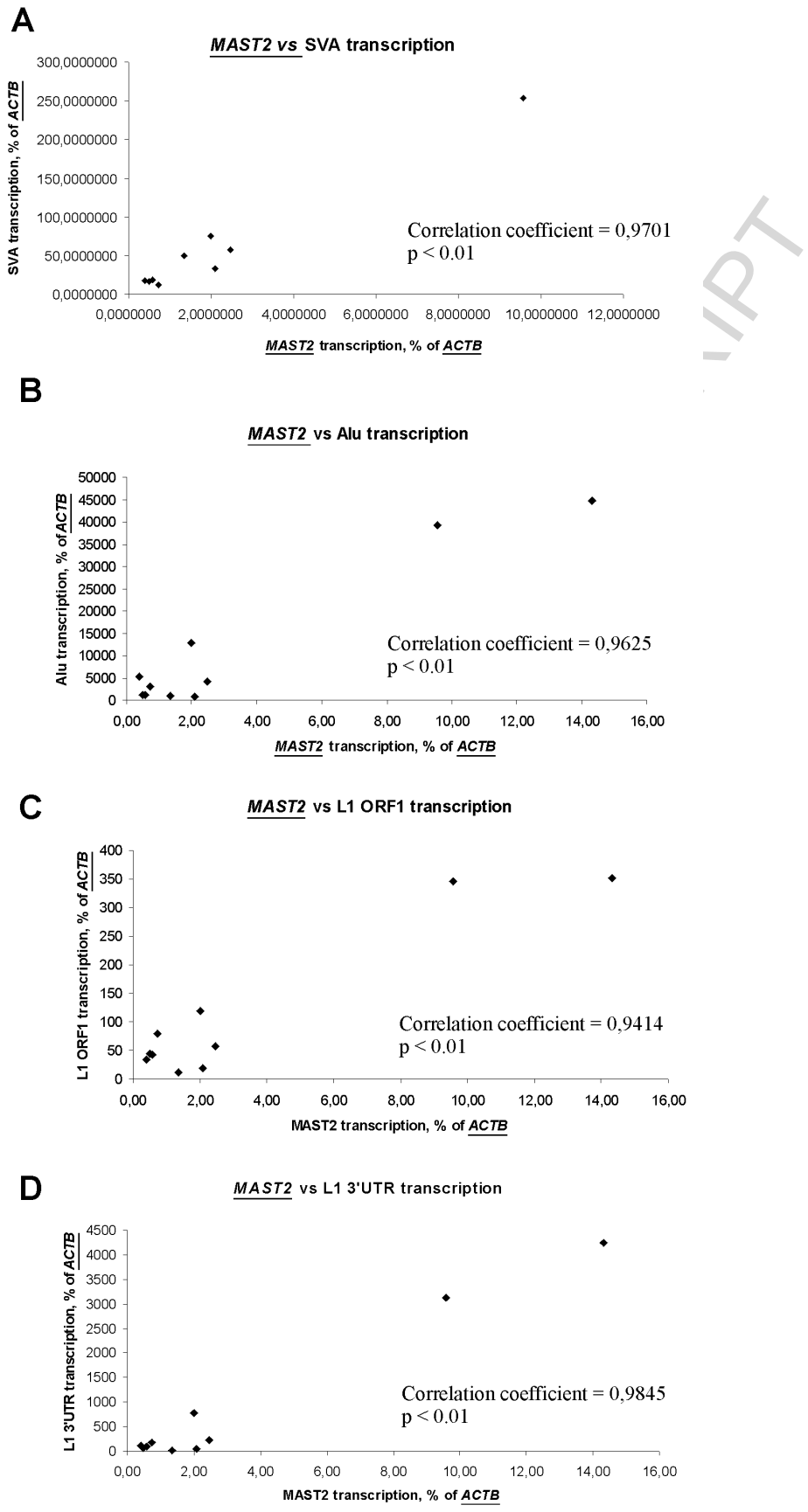


Fig. 5

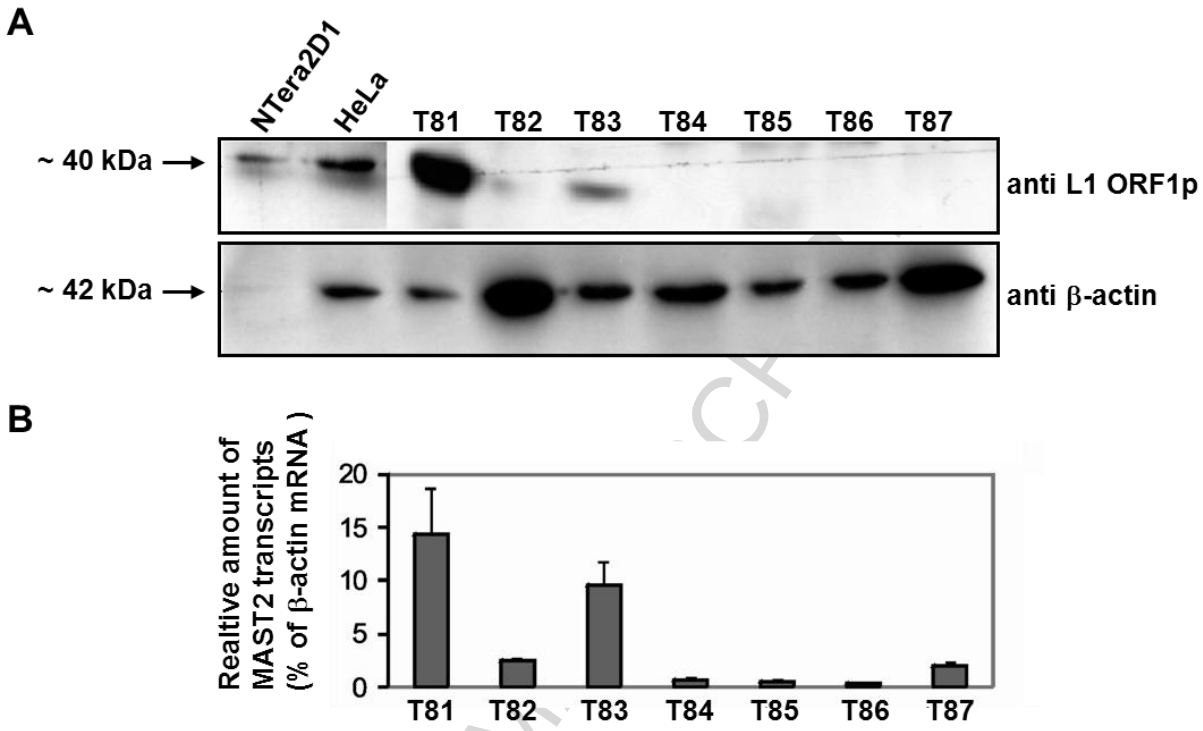


Fig. 6

Highlights

- SVA_{F1} (CpG-SVA) family consists of at least 84 human specific hybrid retrotransposons
- 5' terminal sequence of SVA_{F1} evolved from the exon 1 of the gene *MAST2*
- We show that SVA_{F1} and *MAST2* display similar transcriptional pattern in human tissues
- *MAST2*–derived sequence serves as the positive transcriptional regulator of SVA_{F1} elements
- *MAST2* is cotranscribed with many mammalian retrotransposons in testicular tissue