

# Different sequence signatures in the upstream regions of plant and animal tRNA genes shape distinct modes of regulation

Gong Zhang<sup>1</sup>, Radoslaw Lukoszek<sup>1,2</sup>, Bernd Mueller-Roeber<sup>2,\*</sup> and Zoya Ignatova<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry and <sup>2</sup>Department of Molecular Biology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Potsdam, Germany

Received October 6, 2010; Revised November 16, 2010; Accepted November 20, 2010

## ABSTRACT

In eukaryotes, the transcription of tRNA genes is initiated by the concerted action of transcription factors IIIC (TFIIIC) and IIIB (TFIIB) which direct the recruitment of polymerase III. While TFIIIC recognizes highly conserved, intragenic promoter elements, TFIIB binds to the non-coding 5'-upstream regions of the tRNA genes. Using a systematic bioinformatic analysis of 11 multicellular eukaryotic genomes we identified a highly conserved TATA motif followed by a CAA-motif in the tRNA upstream regions of all plant genomes. Strikingly, the 5'-flanking tRNA regions of the animal genomes are highly heterogeneous and lack a common conserved sequence signature. Interestingly, in the animal genomes the tRNA species that read the same codon share conserved motifs in their upstream regions. Deep-sequencing analysis of 16 human tissues revealed multiple splicing variants of two of the TFIIB subunits, Bdp1 and Brf1, with tissue-specific expression patterns. These multiple forms most likely modulate the TFIIB–DNA interactions and explain the lack of a uniform signature motif in the tRNA upstream regions of animal genomes. The anticodon-dependent 5'-flanking motifs provide a possible mechanism for independent regulation of the tRNA transcription in various human tissues.

## INTRODUCTION

The genetic code is degenerate and the full set of the 20 naturally occurring amino acids (except for Met and Trp) are decoded by multiple codons. Theoretically, a set

of tRNA species that decode each codon individually can translate the entire genetic information from all mRNAs in a cell. Considering the wobbling effect by which one tRNA species can recognize more than one codon, the minimal tRNA set required for translation would be approximately 30 species. In bacterial genomes, the number of tRNA genes ranges from 29 to 120 genes. The low number of tRNA genes in these organisms has probably resulted from genome minimization favoring rapid replication and relatively simple regulation (1). The number of the tRNA genes increases parallel with the increase of the complexity of the organisms. The genomes of higher eukaryotes encode several distinct tRNA species (isoacceptors) to read the 61 sense codons, which also differ in their copy number among the genomes (2). A significant fraction of the tRNAs represents also species that bear the same anticodon but differ in the sequence elsewhere in the tRNA body and are defined as isodecoders (3). According to the recent version of the genomic tRNA database (November 2010) (4,5), 506 tRNAs genes in human, 605 in *Caenorhabditis elegans*, 430 in mouse and 630 in *Arabidopsis thaliana* decode the standard 20 amino acids. Systematic analysis of the tRNA composition in human tissues has revealed that the concentration of tRNA isoacceptors and isodecoders varies by as much as 10-fold among different tissues and stages of differentiations (2,3,6–8) and is therefore unlikely to be driven by neutral genetic drift. The mature tRNAs are very stable, total tRNA levels likely reflect the tRNA transcription rates (9–11). The tRNA composition differs so that it mirrors the expression needs of the cell: some tRNA species can be upregulated to meet the larger demand for some specific tRNAs by overproduction of proteins with overrepresented codons or by enhanced cell proliferation in cancer cells (6–8). Clearly, the eukaryotic genomes encode a complex tRNA pool; however, the expression of the tRNA species is not random and is subject to

\*To whom correspondence should be addressed. Tel: +49 331 977 5130; Fax: +49 331 977 5128; Email: ignatova@uni-potsdam.de  
Correspondence may also be addressed to Bernd Mueller-Roeber. Tel: +49 331 977 2810; Fax: +49 331 977 2512; Email: bmr@uni-potsdam.de

tight regulation. Transcriptional control of the tRNA genes might play a role in the development and cell differentiation of tissues of higher eukaryotes (12,13). The information embedded in the genome to maintain the differential regulation and transcription of the tRNA genes remains elusive.

In eukaryotes, the tRNA genes are transcribed by RNA polymerase III (pol III) [reviewed in (14–16)] and transcription is primarily dependent on intragenic promoter elements. The highly conserved A and B boxes of tRNA genes (coding for the D- and T $\Psi$ C-stem loops) serve as internal promoters to which the transcription factor TFIIC (TFIIIC) binds (17). TFIIC-binding is followed by a recruitment of a second transcription factor, TFIIB, which recognizes upstream region of the tRNA genes and directs the recruitment of pol III (14). TFIIB is a multisubunit complex consisting of pol III-specific B-double prime 1 (Bdp1), TFIIB-related factor 1 (Brf1) and TATA-binding protein (TBP) (18). In vertebrates, TFIIB exists in two isoforms: TFIIB $\alpha$  (composed of Bdp1, Brf2 and TBP), which is specialized to transcribe genes with promoter elements upstream of the initiation site, e.g. U6-type RNA pol III promoters (19), and TFIIB $\beta$  (comprising Bdp1, Brf1 and TBP), which is required for transcription of genes with internal promoter elements, e.g. tRNA and 5S RNA pol III promoters (20,21). TBP is a shared element with the polymerase II (pol II) and recognizes pol II and pol III-specific transcription factors through different regions of its surface (22). TBP occupies a key position in the TFIIB: analogous to its function in the pol II promoters, TBP contacts the TATA-like elements in the promoter sequences and through distortion of these segments strengthens the binding of the transcription factor complex (23). Brf1 plays an integral role for the functional stability of the TFIIB complex by interacting with both, TBP and Bdp1 [interactions are reviewed in (14)]. The third component, Bdp1, determines the physical stabilities of the TFIIB complex by modulating the DNA deformation in sequence-independent fashion (24). A cDNA library derived from human breast cancer cells reveals a presence of several active Brf1 isoforms which are generated by alternative splicing and are suggested to serve as molecular adaptors that link the TFIIB to the DNA-recognition complexes (25). Similarly, various Bdp1 forms with different molecular mass and TFIIC activity have been isolated from human embryonic kidney cells (26) which are most likely a result of an alternative splicing (27).

TATA-box-like elements and TA-rich sequences are present in the upstream regions of some tRNA genes usually between positions –34 and –29 (numbered from the 5'-end of the mature tRNA) (28,29). TFIIB can directly be recruited to a typical TATA box even in the absence of TFIIC and independently initiate transcription by pol III recruitment in yeast (30,31). In silkworms, TFIIB binds stably to the promoter sequence, but is transcriptionally incompetent in the absence of TFIIC (32). Importantly, TA-enriched regions located immediately upstream of the tRNA coding sequence (within the first 50 nt) enhance transcription, whereas multiple copies of such sequences in the far upstream regions reduce the

tRNA transcription levels; sequestration of the TFIIB complex at multiple binding sites decreases its effective concentration thus reducing the formation of the initiation complex with pol III (32). In addition, the regions upstream of the 5'-initial nucleotide of the mature tRNAs in plant and yeast (usually at positions –7 to –3) bear another conservative CAA motif which together with the TATA boxes cooperatively enhances transcription (28,33). Not all active tRNA genes, however, have this promoter structure. In general, the average proportion of AT-nucleotides in the promoter region is very high in plants (72.5%) and yeast (69.4%) compared to mammals (52.5%) (28). In a set of human (302 species) and mouse (267 species) tRNAs, the upstream region lacks TA-rich motifs (34). Similarly, the analysis of 267 species of *Drosophila melanogaster* did not reveal the typical TATA box which is most likely due to the replacement of TBP with the transcription factor TRF in this organism (35). Taken together, these data imply that sequences outside the tRNA coding regions are essential for transcription and suggest that they might play a crucial role in regulating the tRNA transcription.

Taking advantage of the large number of sequenced and annotated genomes, we have undertaken a systematic analysis of the upstream promoter regions of tRNA genes in several multicellular eukaryotic genomes. Our data revealed a clear difference in the conserved motifs located upstream of the transcription initiation site of plant versus animal tRNAs. Using deep-sequencing data of the transcriptome of 16 human tissues, we identified multiple splicing variants of Bdp1 and Brf1 with a tissue-specific expression pattern. We explain the lack of a common conserved motif in the upstream promoter region of animal tRNAs by the presence of different splice forms of Bdp1 and Brf1, which through their structural variations may differently modulate the TFIIB–DNA interactions. This finding suggests that distinct mechanisms in the regulation of the tRNA transcription exist in plant and animal genomes.

## MATERIALS AND METHODS

### tRNA information and genome sequences

The tRNA genes encoded in the nuclear genomes were extracted from the genomic tRNA database (gtRNAdb, <http://gtRNAdb.ucsc.edu/>) (4,5). Pseudo tRNA genes were excluded from the analysis. Note that the tRNA genes encoded in mitochondrial and chloroplast genomes were not included in the search. The following genomes (release version and date are indicated) were used: *Homo sapiens* (human; hg19, NCBI build 37.1, February 2009), *Mus musculus* (mouse; mm9, July 2007), *Rattus norvegicus* (rat; rn4, November 2004), *Gallus gallus* (galGal3, May 2006), *D. melanogaster* (release 5, April 2006), *Anopheles gambiae* (anoGam1, February 2003), *C. elegans* (ce4, January 2007), *A. thaliana* (February 2004), *Medicago truncatula* (March 2009), *Populus trichocarpa* (version 2.0, January 2010), *Glycine max* (soybean; glyma 1.0, December 2008).

## Statistical analysis of the upstream sequences of the tRNA genes

**Motif 'word' search.** The upstream sequences of the tRNA genes (nucleotide positions  $-100$  to  $-1$  nt; numbered from the first 5'-nt of the mature tRNA) were retrieved from the genomic tRNA data base and a simple 'word' search to extract a conserved sequence using the algorithm originally proposed by Waterman and Jones (36) and implicated for the search of consensus motifs in the upstream regions of yeast tRNAs (34). The 100 nt-long upstream sequences were screened for the frequency of occurrence of 6-nt 'word' comprising all possible combinations of 6 nt, thereby allowing none or one mismatch within a 9-nt sliding window. The 'word' length was set to 6 nt which is the common length of transcription factor binding sites in eukaryotes (37–39) and has been used in sequence analysis of transcription factor binding motifs elsewhere (34,40). The 'word' with the highest frequency is identified as the best frequency 'word' for each genome. To avoid a bias toward the highly homologous sequences of mature tRNAs, the word search was terminated at position  $-9$ . To sort out significant motifs (peaks), we set a threshold which is defined as already described (34). Briefly, the individual sequences of the upstream region for each organism were randomized and the mean, minimum and maximum values occurring within this randomized data set were calculated to determine the random noise range.

**Elicitation of conserved motifs via MEME.** To identify conserved motifs in the upstream sequences of tRNA genes, we used the MEME version 4.4.0 (available online at [http://meme.sdsc.edu/meme4\\_4\\_0/cgi-bin/meme.cgi](http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi)) (41). The parameters were set as motif minimum width of 6, maximum width of 10, max number of motifs of 3 or 5 and the option 'zero or one per sequence'.

## Deep-sequencing data sets

The deep-sequencing data sets of poly(A)<sup>+</sup>-selected mRNAs of 16 human tissues and organs were analyzed as a part of the 'Human Body Map 2.0' project and the sequencing was performed using an Illumina HiSeq 2000 platform with an error rate of 0.2–0.6%. The sequencing reads with a length of 75 nt were kindly provided by Gary Schroth and his group at Illumina Inc.

## RESULTS

### Conserved motifs are present in the upstream sequences of plant tRNA genes but are lacking in the animal genomes

TFIIIB recognizes conserved TATA-box-like elements in the upstream regions of the tRNA genes in plants and yeast through its TBP subunit (28,34,42). To address whether a conserved sequence motif is a characteristic feature of the upstream sequences of all eukaryotic tRNAs we performed a 'word' search within the 100 upstream nucleotides of the tRNA genes and compared seven animal with four plant genomes (Figure 1). A significant peak around position  $-35$  with a TATA motif

was detected in all plant genomes as reported already for *A. thaliana* and *S. cerevisiae* (28,34,42,43). In contrast, the upstream tRNA regions of the seven animal genomes did not reveal any uniform and conserved motif (Figure 1); motifs with rather individual signature for each genome were identified and the sequence and position of the motifs was dependent on the search settings (Figure 1; compare the outcome of the search with none or one nucleotide mismatch). Allowing 1 nt mismatch in the search yielded slightly more significant peaks positioned at  $\sim -40$  to  $-10$  nt upstream of the 5'-initial nucleotide of the mature tRNA in all seven animal genomes (Figure 1B). Extending the search up to 2000 nt upstream did not reveal any further conserved sequences in the animal genomes (Supplementary Figure S1). The height of the identified highest peaks in animal genomes is one-third of those specified in the plant genomes. An exception among the animal genomes is *D. melanogaster*: independent of the search parameters (none or one mismatch) a clear 6-nt motif (TTTGGC) at position  $-22$  was detected.

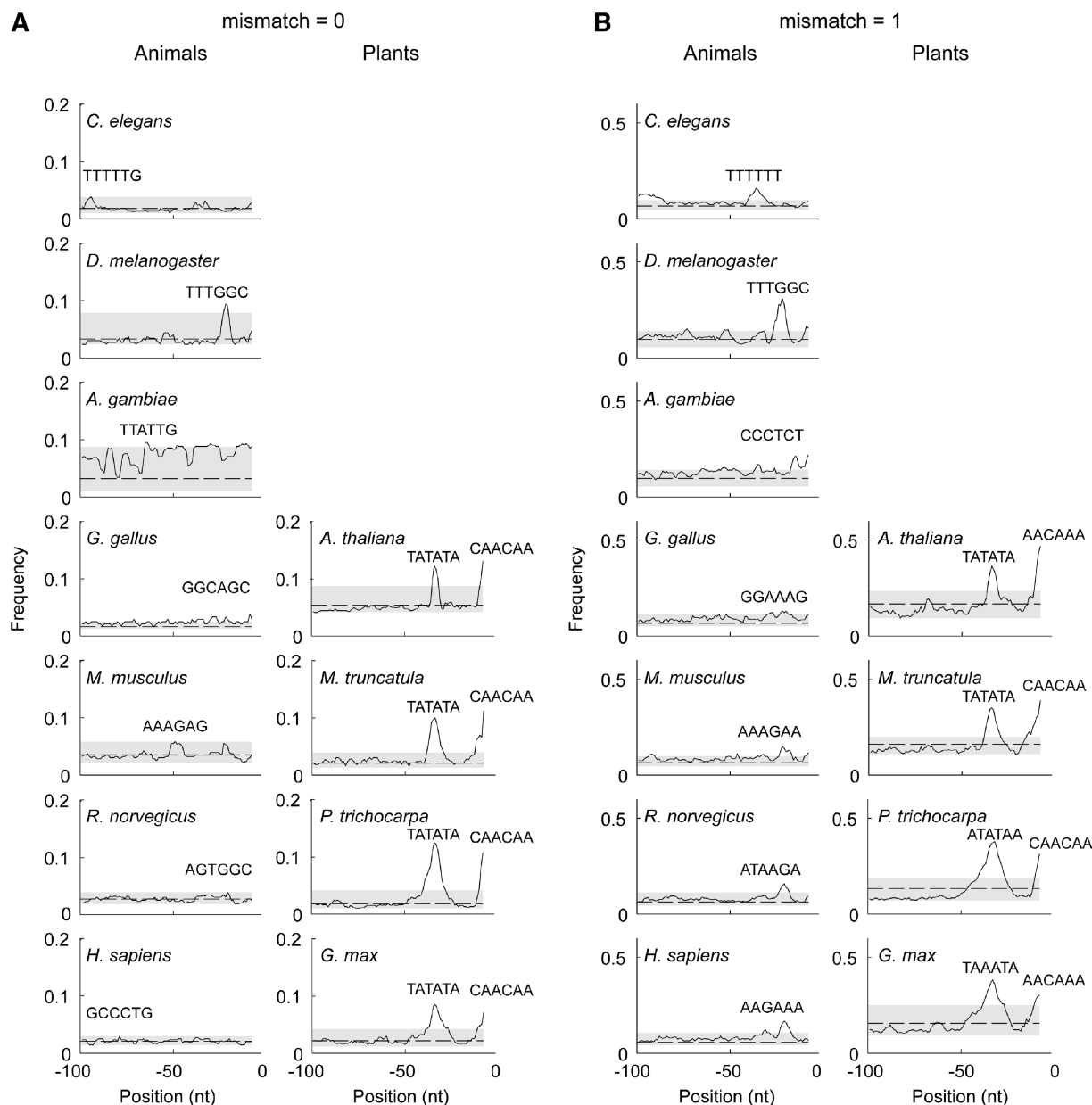
Furthermore, the 'word' search revealed an interesting feature in the plant genomes: a significant CAA motif was detected at positions  $-7$  to  $-3$  upstream of the 5'-initial nucleotide of the mature tRNAs (Figure 1). The signal is completely absent in all analyzed animal genomes. The distance between the processed 5'-end of the mature tRNA and the transcription initiation site is usually 9–12 bases [G. Kassavetis, a supplement article to (44)] and conserved motifs in these segments play a central role in transcriptional initiation (33), particularly in plants (28,43). Interestingly, the same motif was identified in the unicellular yeast *S. cerevisiae* while it was absent in *Schizosaccharomyces pombe* (Supplementary Figure S2). The best six-letter 'word' identified in *S. cerevisiae* was the TATA motif which resembled the sequence identified in the non-coding upstream regions of the plant tRNA genes (Supplementary Figure S2). Notably, in the two yeast species, the motifs with the highest frequencies centered at positions identical to those identified for the plant tRNA upstream regions (Figure 1).

To further verify the presence of a conserved motif signature within the upstream sequences of the tRNA genes, we used the algorithm introduced by Bailey and Elkan (41). Similar to the 'word' search, this algorithm validated the presence of the TA-rich motif in the upstream sequences of plant tRNA genes, whereas more variable GC-rich sequences were extracted from the animal genomes (Supplementary Figure S3). Taken together, these analyses suggest that distinct patterns of regulation of tRNA transcription in animal and plant genomes may exist.

### Animal tRNA genes share common motifs in the upstream regulatory regions in an anticodon-dependent manner

Since we could not identify a common, conserved motif in the upstream segments of animal tRNA genes, we next analyzed separately the isoacceptor and isodecoder tRNAs that read codons encoding the same amino acid.

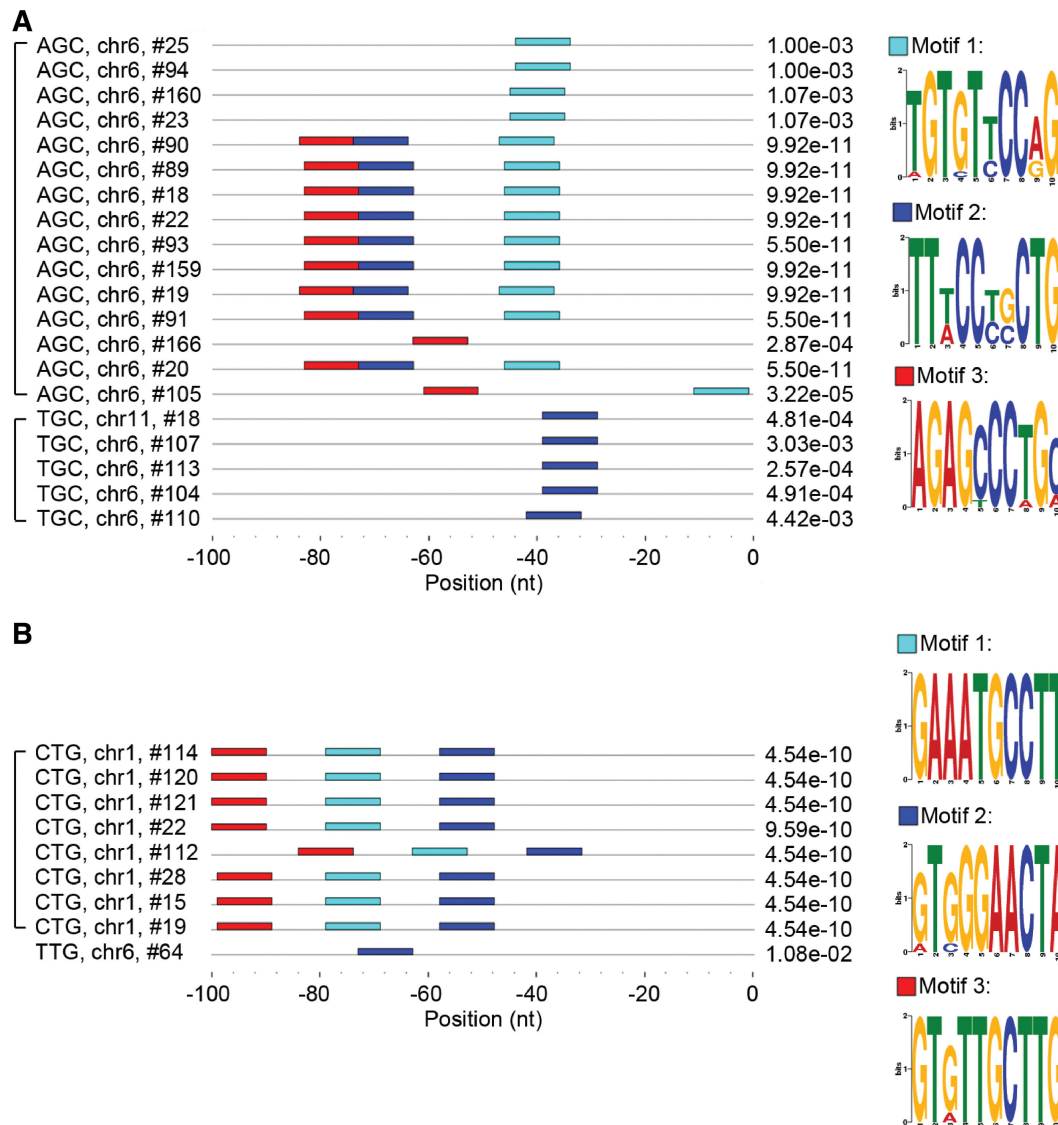




**Figure 1.** Conserved patterns are only present in the upstream sequences of plant tRNA genes. The best frequency 'word' search was performed in the 5'-region (positions -100 to -9 nt) upstream of the initial nucleotide of the mature nuclear-encoded tRNAs of different organisms allowing none (A) or one mismatch (B) (for details see 'Materials and Methods' section). The occurrence of the most frequent six-letter 'word' at each position (x-axis) is plotted on the y-axis. Position 0 corresponds by the first nucleotide of the mature tRNA. The six-letter 'word' with the highest frequency (highest peak) is indicated above the corresponding peak. The CAA-motifs immediately upstream of the 5'-initial nucleotide of the mature tRNAs are also depicted. The dashed line represents the average frequency value obtained from randomized sequences. The shadowed area indicates the area between the maximum and minimum frequency values derived from the randomized data set. nt, nucleotide.

Importantly, this search extracted anticodon-dependent motifs in the upstream segments of the majority of the tRNA families (Figure 2 and Supplementary Figure S4) in all tested animal genomes (Table 1 and Supplementary Table S1); however, no such dependence was observed for any of the plant genomes. In general, this trend is slightly more prominent in the genomes of higher vertebrates than those of lower, invertebrate animals. Comparison of the tRNA species pairing to the codons for alanine (AGC, TGC and CGC are the corresponding anticodons)

revealed a unique TGTGTtCCaG motif for Ala-tRNAs with the AGC anticodon and a characteristic TTtCCt gCTG motif for the Ala-tRNAs with the TGC anticodon (Figure 2A). The TTtCCtgCTG motif centers at position -40 to -30 upstream of the 5'-site of the mature Ala-tRNAs with the TGC anticodon; it is also present within the 5'-upstream flanking region of the Ala-tRNAs that contain the AGC anticodon albeit at a different position (Figure 2A). In contrast, the upstream regions of the third type of Ala-tRNAs bearing the CGC



**Figure 2.** Animal tRNA genes with the same anticodon bear conserved motifs in their upstream regions. Examples of typical anticodon-dependent motifs in the upstream sequences of the human tRNAs encoding Ala (**A**) and Gln (**B**) identified with MEME 4.4. Each tRNA is specified by its anticodon, chromosome location and serial chromosome number (data were retrieved from the genomic tRNA database). The square brackets on the left group the tRNAs with the same anticodon. Each horizontal line represents the 100-nt-long region upstream of the 5'-end of the mature tRNA genes; position 0 determines the first nucleotide of the mature tRNA. The positions of the three most significant motifs are colored (cyan, blue and red) and the corresponding sequence logos are presented on the right side of each group. The numbers on the right side represent the combined *P*-value and are the product of the *P*-values of all motifs detected within the upstream sequence; they are inversely proportional to the significance of the motifs. Motifs with *P* < 0.05 were considered as significant.

anticodon do not have any motif in common. Another example, i.e. the genes for Gln-tRNA, showed a more complex pattern: tRNAs bearing the CTG anticodon share three different motifs GAAATGCCTT, gTgGGA ACTA, GTgTTGCTTG, whereas Gln-tRNAs with the TTG anticodon possess only the gTgGGA ACTA motif which is also present in Gln-tRNAs with the CTG anticodon (Figure 2B). The number of the common motifs varies among the tRNA species harboring the same anticodon; common motifs are also shared with other tRNAs bearing the same amino acid but pairing to another, alternative codon (Supplementary Figure S4A).

Next to Trp and Met which are encoded by a single codon, due to the wobbling some amino acids, albeit encoded by different codons, are decoded by tRNA species bearing one and the same anticodon. Interestingly, these tRNA families also share one to three common motifs (Supplementary Figure S4B). Different combinations among the common motifs divide the tRNA families into subgroups suggesting also a possibility for differential regulation of the expression even among the tRNA families bearing the same anticodon. The variations in the significance value (*P*-value) imply that the homologous motifs observed within a tRNA family bearing the same

**Table 1.** Statistical analysis of anticodon-dependent motifs of animal genomes

Genome <sup>a</sup>	Number of amino acids with anticodon-dependent upstream tRNA motif	Number of amino acids without anticodon-dependent upstream tRNA motif
<i>Homo sapiens</i> (human)	10	4
<i>Rattus norvegicus</i> (rat)	8	5
<i>Mus musculus</i> (mouse)	8	4
<i>Gallus gallus</i> (hen)	9	5
<i>Drosophila melanogaster</i>	5	7
<i>Anopheles gambiae</i>	5	7
<i>Caenorhabditis elegans</i>	8	5

<sup>a</sup>A detailed list of the motifs within the tRNA upstream sequences dependent on the anticodon is given in Supplementary Table S1. Note that amino acids decoded by tRNAs with one and the same anticodon are not included in the table.

anticodon do not originate from simple duplications of the tRNA genes (Figure 2 and Supplementary Figure S4).

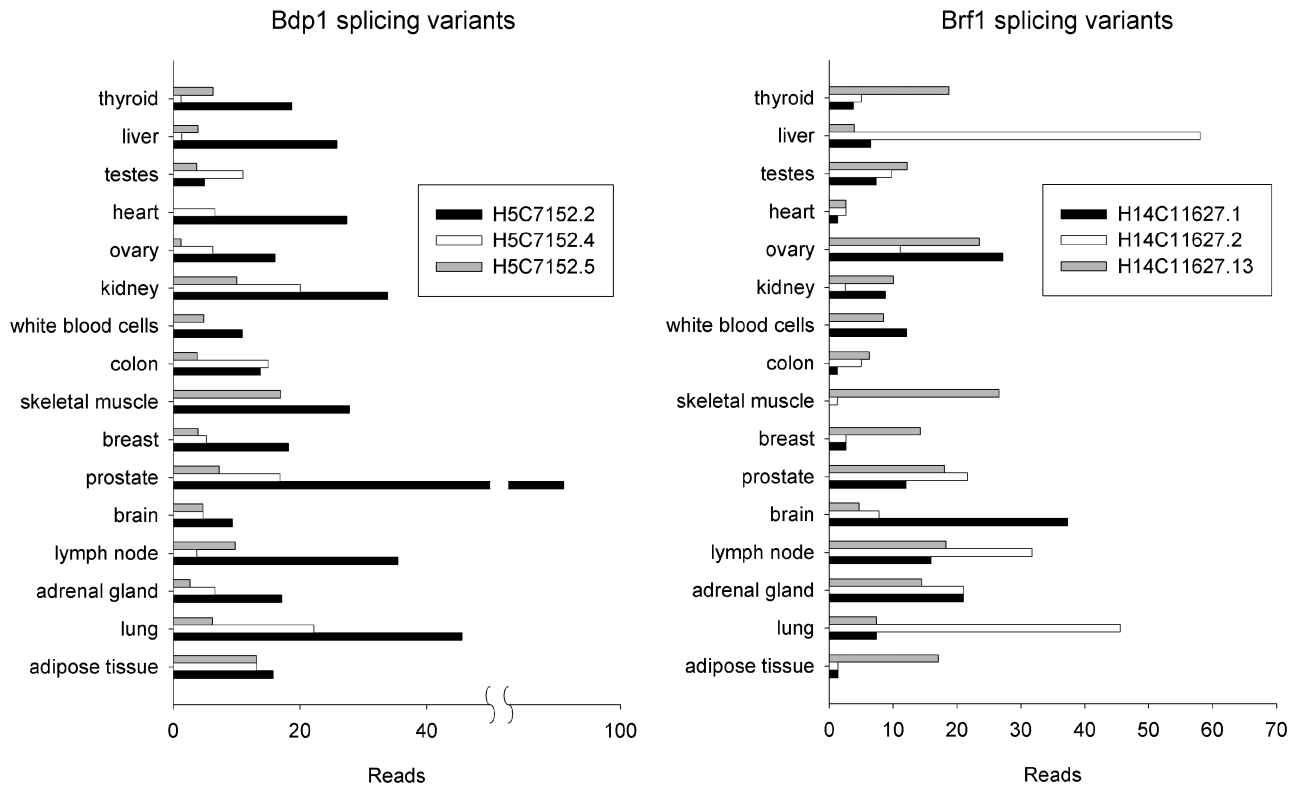
**Tissue-specific expression patterns of TFIIB isoforms may explain the anticodon-specific regulation of tRNA expression**

TFIIB is the essential factor that recognizes specific motifs in the upstream sequences of tRNA genes [preferably TA-rich regions in plants (28,43) and yeast cells (42)] and its binding stability to the cognate motif determines the expression level within one family of isodecoders (32). In human cells, expression of the genes encoding the three subunits of the TFIIB complex, Brf1, Bdp1 and TBP, is either coordinated (45) or differentially modulated (46,47). The human *BRF1* and *BDP1* genes encode various isoforms of considerably different molecular masses [UniProt database (<http://www.uniprot.org/>)] that result from alternative splicing [<http://genome.ewha.ac.kr/ECgene/> (48)] (Supplementary Table S2). We hypothesized that the anticodon-dependent motifs detected in the tRNA families of the animal genomes might bear the information for differential regulation of the expression of the tRNA sets for each codon in different tissues. This, in turn, would require a differential expression of the TFIIB isoforms in various tissues and cell types. We tested this hypothesis using high-throughput deep-sequencing data of total mRNA isolated from various human tissues. To search for specific splice variants of Brf1 and Bdp1, we aligned all deep-sequencing reads (75 nt) to the specific junction of the target splicing variants. Only reads that aligned to at least 10 nt upstream and downstream of the splice junction (allowing thereby a maximum of one mismatch) were recorded as hits. The read hits were then normalized to the total read count of the data sets to calculate the relative concentrations of the different isoforms (Figure 3 and Supplementary Figure S5). Intriguingly, various Bdp1 and Brf1 isoforms are

differentially enriched in different tissues so that a clear tissue-specific pattern emerges. For example, the H5C7152.4 isoform of Bdp1 dominates in the thyroid, whereas the isoform H5C7152.2 is the main species in the liver and prostate gland (Figure 3 and Supplementary Figure S5). Liver and lung tissues are enriched in the H14C11627.2 isoform of Brf1 and H14C11627.13 is the dominating isoform in the skeletal muscle (Figure 3 and Supplementary Figure S5). Variations in the expression of the isoforms in each tissue suggest that they may modulate the TFIIB-binding affinity to various promoter elements and thus may serve as a regulatory tool to modulate the tRNA expression levels of the isodecoders or isoacceptors in each tissue.

**DISCUSSION**

We have compared on genome-wide scale the conserved signatures within the upstream regions of nuclear tRNAs from 11 different multicellular genomes. The upstream segments of plant tRNAs showed a conserved TA-enriched motif, followed by a CAA motif immediately upstream of the 5'-initial nucleotide of the mature tRNA transcript. Similar motifs were identified in the upstream regions of yeast tRNAs [Supplementary Figure S2 and (29)]. In contrast, the 5'-flanking regions of the tRNAs from animal genomes are highly heterogeneous and vary significantly among the different species. This observation was surprising to us as TBP is present as a subunit of both, plant and animal TFIIB complexes. TBP recognizes TATA-like elements in the promoter regions; however, the entry of Brf1 and Bdp1 into the TFIIB complex alters the dynamics of TBP–DNA interactions (24). For example, the sequence preference of the assembled TFIIB complex with a mutant TBP differs significantly from the mutant TBP alone (49). The crystal structure of the yeast TFIIB complex shows that the convex surface of the TBP is extensively engaged in exceptionally extended contacts with the Brf1 subunit which are three to five times greater than the contacts observed in the pol II complexes: a total of 31 Brf1 amino acid residues interact with 29 residues of TBP through hydrogen bonds and van der Waals contacts (22). Based on sequence conservation between human and yeast Brf1, it can be predicted that 28 of the 31 hydrogen bonds observed in the yeast complex TBP–Brf1 are retained (22) suggesting also a large interaction interface between Brf1 and TBP in the human TFIIB complex. In addition, Brf1 is structurally flexible and undergoes large unfolding transitions by complex formation with TBP (22). The primary sequence variations in the human Brf1 and Bdp1 isoforms may introduce some structural variations among them, which consequently may differently modulate the TFIIB–DNA interactions and consequently the interactions with TFIIC. Consistent with this interpretation, isolated forms of Bdp1 with molecular masses of 220 and 250 kDa show higher TFIIC activity compared to the smaller Bdp1 forms with apparent masses of 150, 100 and 90 kDa (26). In the animal genomes, the presence of multiple isoforms of Bdp1 and Brf1 correlates with the lack of a uniform signature in the



**Figure 3.** Various RNA splice variants of *Bdp1* and *Brf1* genes are expressed in different human tissues. Normalized read hits for specific splicing junctions of three *Bdp1* and *Brf1* isoforms in 16 human tissues. All read hits were normalized to 100 million total sequence reads. Numbering of the splice variants is according to the ECGene alternative splicing database (Supplementary Table S2). The expression of other isoforms is presented in Supplementary Figure S5.

upstream regions of the tRNA genes suggesting that different *Brf1* and *Bdp1* splice forms may differently modulate the specificity and dynamics of TFIIB-promoter interactions. A genome-wide search in the *A. thaliana* genome revealed a single, uncharacterized gene (AT4G39160) which possesses limited overall sequence homology to human and yeast *Bdp1* (below 9% amino acid similarity). A fragment of it, comprising the phylogenetically most conserved region of *Bdp1*, the SANT domain (14), showed higher amino acid similarity to the human and yeast *Bdp1* proteins (55–63% amino acid similarity). Currently, however, there is no evidence for a contribution of this protein to the tRNA transcription initiation in plants. Two genes, which might be homologous to the *Brf1* were identified in *Arabidopsis* (Figure 4). Note, that so far no splice variants have been identified for *Brf1*-like transcripts in plants.

The genomic sequence database is updated frequently based on the higher accuracy of the new sequencing technologies and improved genome annotations. Consequently, the automated annotation of tRNA genes increases rapidly (5). For example, 446 tRNA genes were identified in the human genome in 2006 (2); this number increased to 506 species and additional 110 pseudogenes in 2009 (hg19, NCBI Build 37.1, February 2009). The differential expression of tRNA genes in a tissue- and developmental stage-specific manner (6,8) requires a sophisticated mode to control the transcription. Degenerate regulatory



**Figure 4.** Putative plant homologues of *Brf1* share some conservative elements with other eukaryotic *Brf1* proteins. Homology searches with the *S. cerevisiae* and *H. sapiens* *Brf1* splice variant H14C11627.9 (Supplementary Table S2) identified two putative homologues, AT3G09360 and AT2G45100 from *A. thaliana*. The conserved motifs derived from the sequence alignment (Supplementary Figure S6) are depicted in boxes: gray—cyclin fold boxes, white—*Brf1*-like TBP-binding domain. Both *Arabidopsis* proteins do not possess the conserved TFIIB zinc-binding domain (black box). Note that each conserved segment is scaled to the position within the corresponding nucleotide sequence (black line).

information embedded in the upstream non-coding sequences of the tRNA genes would reduce the complexity of regulation, which may result in lower metabolic costs and thus provide an evolutionary advantage. This may explain the presence of conserved signature motifs in the flanking 5'-upstream regions of tRNA genes bearing the same anticodon. The presence of more than one motif for the majority of the tRNAs with the same anticodon suggests a complex regulation mechanism of transcription. Importantly, even within these groups of tRNAs, the number of the conserved motifs varies (Figure 2A,



Supplementary Figure S4) implying a fine-tuning function of some of the motifs. The different motifs might have evolved to bind various TFIIB isoforms coexisting in some tissues; however, in different concentration (Figure 3 and Supplementary Figure S5), so that differential expression of a tRNA species bearing one and the same codon can be achieved in a tissue-specific manner. The new information, we present here on the distinct modes of regulation of tRNA transcription in plants and animals provides an excellent cross-validation with the wealth of experimental data in plants and yeast and presents new twist to understand variations in the tRNA regulation in multicellular eukaryotic organisms.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

We are grateful to Gary Schroth and his group at Illumina for providing the deep-sequencing data sets from 16 the human tissues and organs.

## FUNDING

Funding for open access charge: The Deutsche Forschungsgemeinschaft (grant IG 73/10-1 to Z.I.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F. and Sockett, R.E. (2005) Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.*, **33**, 1141–1153.
- Goodenbour, J.M. and Pan, T. (2006) Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res.*, **34**, 6137–6146.
- Geslain, R. and Pan, T. (2010) Functional analysis of human tRNA isocodons. *J. Mol. Biol.*, **396**, 821–831.
- Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.*, **37**, D93–D97.
- Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
- Dittmar, K.A., Goodenbour, J.M. and Pan, T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, e221.
- Hentzen, D., Chevallier, A. and Garel, J.P. (1981) Differential usage of iso-accepting tRNA<sup>Ser</sup> species in silk glands of *Bombyx mori*. *Nature*, **290**, 267–269.
- Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M.R. and Pan, T. (2009) tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.*, **37**, 7268–7280.
- Czech, A., Fedyunin, I., Zhang, G. and Ignatova, Z. (2010) Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations. *Mol. Biosyst.*, **6**, 1767–1772.
- Komar, A.A. (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.*, **34**, 16–24.
- Zhang, G., Fedyunin, I., Miekley, O., Valleriani, A., Moura, A. and Ignatova, Z. (2010) Global and local depletion of ternary complex limits translational elongation. *Nucleic Acids Res.*, **38**, 4778–4787.
- Ouyang, C., Martinez, M.J., Young, L.S. and Sprague, K.U. (2000) TATA-Binding protein-TATA interaction is a key determinant of differential transcription of silkworm constitutive and silk gland-specific tRNA(Ala) genes. *Mol. Cell. Biol.*, **20**, 1329–1343.
- Reynolds, W.F. (1995) Developmental stage-specific regulation of *Xenopus* tRNA genes by an upstream promoter element. *J. Biol. Chem.*, **270**, 10703–10710.
- Geiduschek, E.P. and Kassavetis, G.A. (2001) The RNA polymerase III transcription apparatus. *J. Mol. Biol.*, **310**, 1–26.
- Huang, Y. and Maraia, R.J. (2001) Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. *Nucleic Acids Res.*, **29**, 2675–2690.
- Paule, M.R. and White, R.J. (2000) Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.*, **28**, 1283–1298.
- Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature*, **294**, 626–631.
- Kassavetis, G.A., Braun, B.R., Nguyen, L.H. and Geiduschek, E.P. (1990) *S. cerevisiae* TFIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIA and TFIIC are assembly factors. *Cell*, **60**, 235–245.
- Teichmann, M. and Seifart, K.H. (1995) Physical separation of two different forms of human TFIIB active in the transcription of the U6 or the VAI gene in vitro. *EMBO J.*, **14**, 5974–5983.
- Schramm, L., Pendergrast, P.S., Sun, Y. and Hernandez, N. (2000) Different human TFIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. *Genes Dev.*, **14**, 2650–2663.
- Teichmann, M., Wang, Z. and Roeder, R.G. (2000) A stable complex of a novel transcription factor IIB-related factor, human TFIIB50, and associated proteins mediate selective transcription by RNA polymerase III of genes with upstream promoter elements. *Proc. Natl Acad. Sci. USA*, **97**, 14200–14205.
- Juo, Z.S., Kassavetis, G.A., Wang, J., Geiduschek, E.P. and Sigler, P.B. (2003) Crystal structure of a transcription factor IIB core interface ternary complex. *Nature*, **422**, 534–539.
- Patikoglou, G. and Burley, S.K. (1997) Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 289–325.
- Grove, A., Kassavetis, G.A., Johnson, T.E. and Geiduschek, E.P. (1999) The RNA polymerase III-recruiting factor TFIIB induces a DNA bend between the TATA box and the transcriptional start site. *J. Mol. Biol.*, **285**, 1429–1440.
- McCulloch, V., Hardin, P., Peng, W., Ruppert, J.M. and Lobo-Ruppert, S.M. (2000) Alternatively spliced hBRF variants function at different RNA polymerase III promoters. *EMBO J.*, **19**, 4134–4143.
- Weser, S., Gruber, C., Hafner, H.M., Teichmann, M., Roeder, R.G., Seifart, K.H. and Meissner, W. (2004) Transcription factor (TF)-like nuclear regulator, the 250-kDa form of Homo sapiens TFIIB<sup>h</sup>, is an essential component of human TFIIC1 activity. *J. Biol. Chem.*, **279**, 27022–27029.
- Kelter, A.R., Herchenbach, J. and Wirth, B. (2000) The transcription factor-like nuclear regulator (TFNR) contains a novel 55-amino-acid motif repeated nine times and maps closely to SMN1. *Genomics*, **70**, 315–326.
- Yukawa, Y., Sugita, M., Choise, N., Small, I. and Sugiura, M. (2000) The TATA motif, the CAA motif and the poly(T) transcription termination motif are all important for transcription re-initiation on plant tRNA genes. *Plant J.*, **22**, 439–447.
- Dieci, G., Yukawa, Y., Alzapiedi, M., Guffanti, E., Ferrari, R., Sugiura, M. and Ottonello, S. (2006) Distinct modes of TATA box utilization by the RNA polymerase III transcription machineries from budding yeast and higher plants. *Gene*, **379**, 12–25.
- Gerlach, V.L., Whitehall, S.K., Geiduschek, E.P. and Brow, D.A. (1995) TFIIB placement on a yeast U6 RNA gene in vivo is directed primarily by TFIIC rather than by sequence-specific DNA contacts. *Mol. Cell. Biol.*, **15**, 1455–1466.
- Kassavetis, G.A., Bardeleben, C., Kumar, A., Ramirez, E. and Geiduschek, E.P. (1997) Domains of the Brf component of RNA polymerase III transcription factor IIB (TFIIB): functions in assembly of TFIIB-DNA complexes and recruitment of RNA polymerase to the promoter. *Mol. Cell. Biol.*, **17**, 5299–5306.



32. Parthasarthy, A. and Gopinathan, K.P. (2005) Transcription of individual tRNA<sup>Ala</sup> genes from within a multigene family is regulated by transcription factor TFIIIB. *FEBS J.*, **272**, 5191–5205.
33. Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) Transcription by RNA polymerase III. *Annu. Rev. Biochem.*, **57**, 873–914.
34. Giuliadori, S., Percudani, R., Braglia, P., Ferrari, R., Guffanti, E., Ottonello, S. and Dieci, G. (2003) A composite upstream sequence motif potentiates tRNA gene transcription in yeast. *J. Mol. Biol.*, **333**, 1–20.
35. Takada, S., Lis, J.T., Zhou, S. and Tjian, R. (2000) A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell*, **101**, 459–469.
36. Waterman, M.S. and Jones, R. (1990) Consensus methods for DNA and protein sequence alignment. *Methods Enzymol.*, **183**, 221–237.
37. Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I. and Rubin, E.M. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.*, **12**, 832–839.
38. Bilu, Y. and Barkai, N. (2005) The design of transcription-factor binding sites is affected by combinatorial regulation. *Genome Biol.*, **6**, R103.
39. Husbands, A., Bell, E.M., Shuai, B., Smith, H.M. and Springer, P.S. (2007) Lateral organ boundaries defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res.*, **35**, 6663–6671.
40. Wang, G., Wang, X., Wang, Y., Yang, J.Y., Li, L., Nephew, K.P., Edenberg, H.J., Zhou, F.C. and Liu, Y. (2008) Identification of transcription factor and microRNA binding sites in responsible to fetal alcohol syndrome. *BMC Genomics*, **9**(Suppl. 1), S19.
41. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
42. Joazeiro, C.A., Kassavetis, G.A. and Geiduschek, E.P. (1996) Alternative outcomes in assembly of promoter complexes: the roles of TBP and a flexible linker in placing TFIIIB on tRNA genes. *Genes Dev.*, **10**, 725–739.
43. Yukawa, Y., Mizutani, T., Akama, K. and Sugiura, M. (2007) A survey of expressed tRNA genes in the chromosome I of Arabidopsis using an RNA polymerase III-dependent in vitro transcription system. *Gene*, **392**, 7–13.
44. Bachman, N., Eby, Y. and Boeke, J.D. (2004) Local definition of Ty1 target preference by long terminal repeats and clustered tRNA genes. *Genome Res.*, **14**, 1232–1247.
45. Zhong, S. and Johnson, D.L. (2009) The JNKs differentially regulate RNA polymerase III transcription by coordinately modulating the expression of all TFIIIB subunits. *Proc. Natl Acad. Sci. USA*, **106**, 12682–12687.
46. Cabarcas, S., Jacob, J., Veras, I. and Schramm, L. (2008) Differential expression of the TFIIIB subunits Brf1 and Brf2 in cancer cells. *BMC Mol. Biol.*, **9**, 74.
47. Athineos, D., Marshall, L. and White, R.J. (2010) Regulation of TFIIIB during F9 cell differentiation. *BMC Mol. Biol.*, **11**, 21.
48. Kim, P., Kim, N., Lee, Y., Kim, B., Shin, Y. and Lee, S. (2005) ECgene: genome annotation for alternative splicing. *Nucleic Acids Res.*, **33**, D75–D79.
49. Tsihlis, N.D. and Grove, A. (2006) The *Saccharomyces cerevisiae* RNA polymerase III recruitment factor subunits Brf1 and Bdp1 impose a strict sequence preference for the downstream half of the TATA box. *Nucleic Acids Res.*, **34**, 5585–5593.