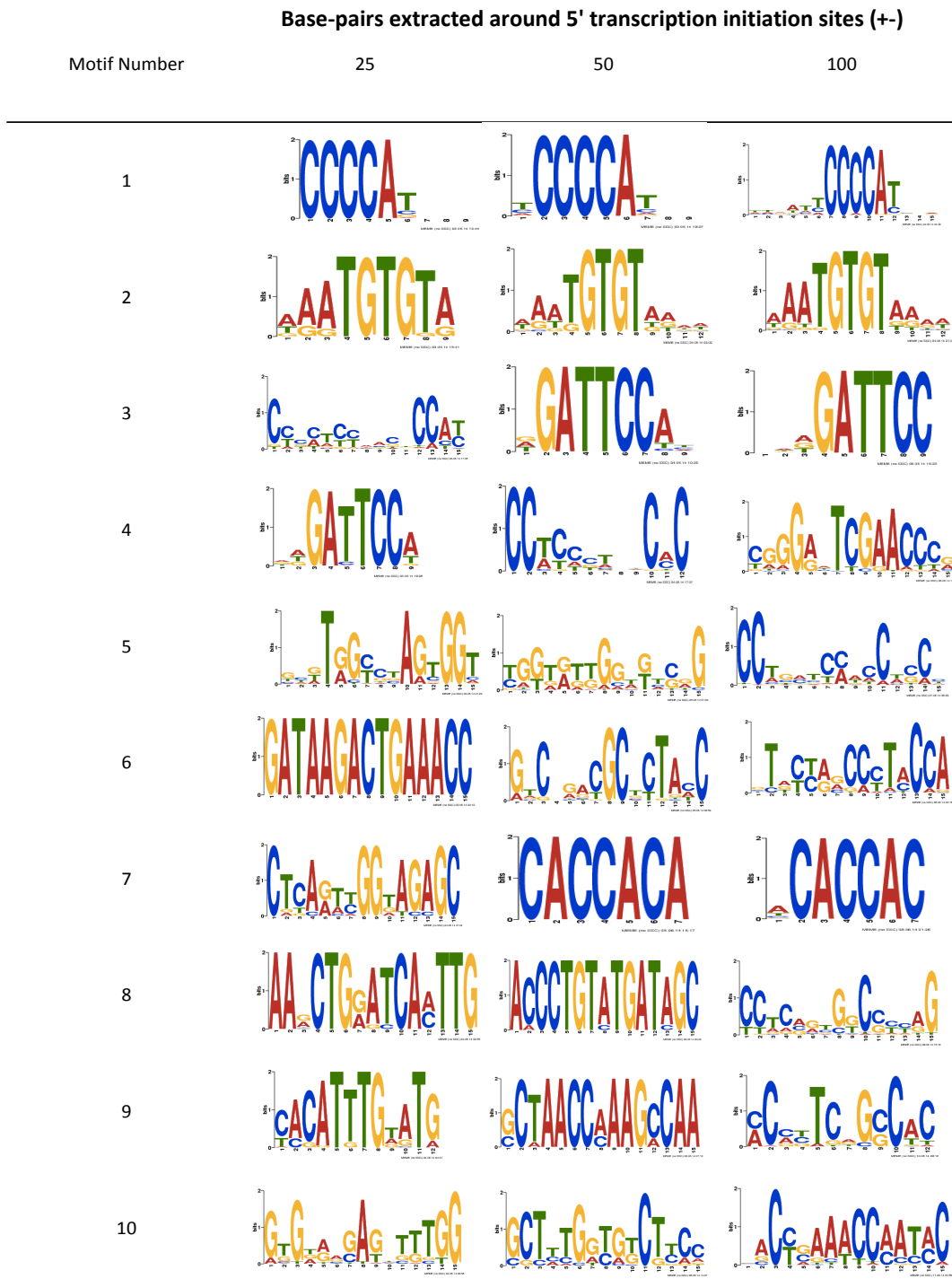


Additional File 1. Supplemental Figures.

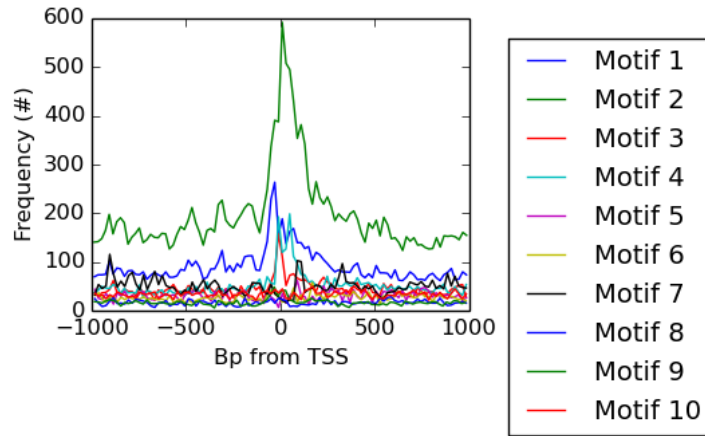
Tretina, Pelle and Silva.



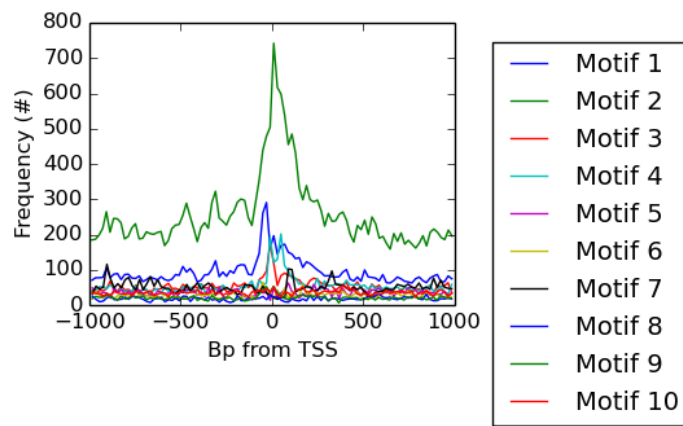
Supplemental Figure 1a. All motifs discovered around transcription initiation sites, at varying distances, and their frequency distributions.

Bp searched
around TSS

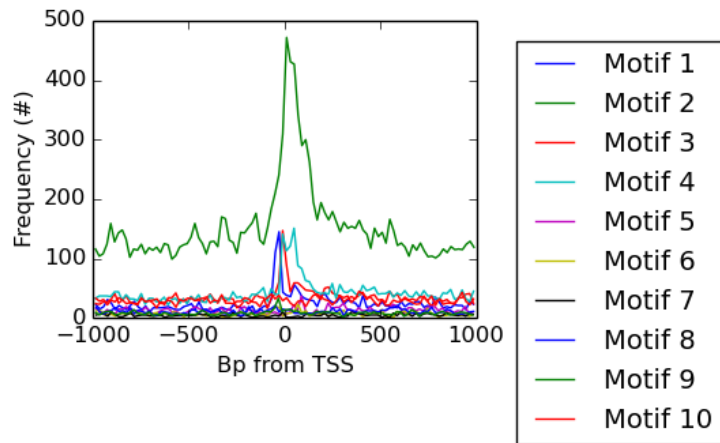
25



50

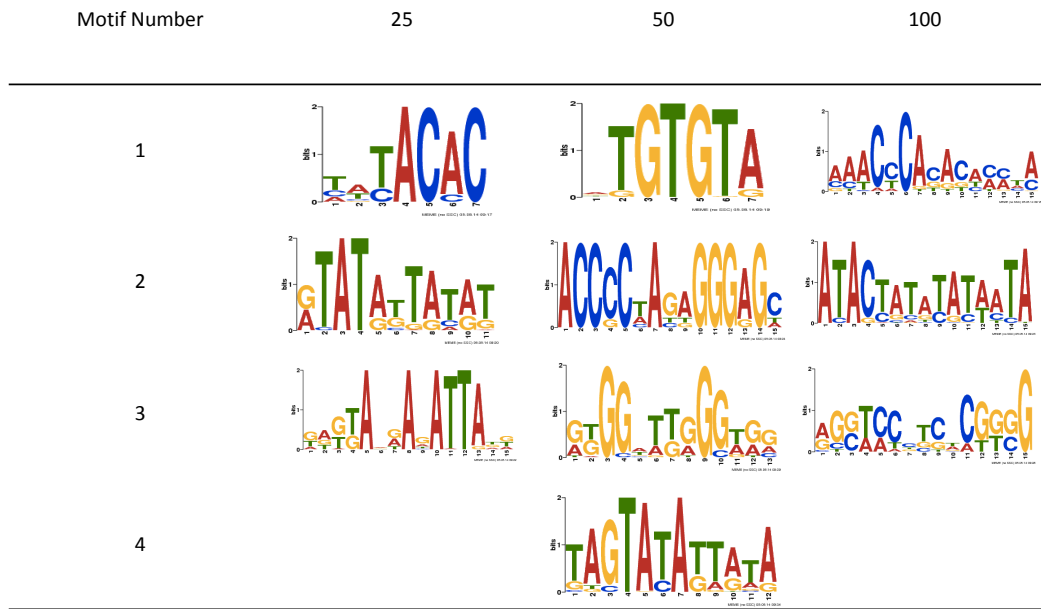


100



Supplemental Figure 1b. All motifs discovered around transcription start sites, at varying distances, and their frequency distributions.

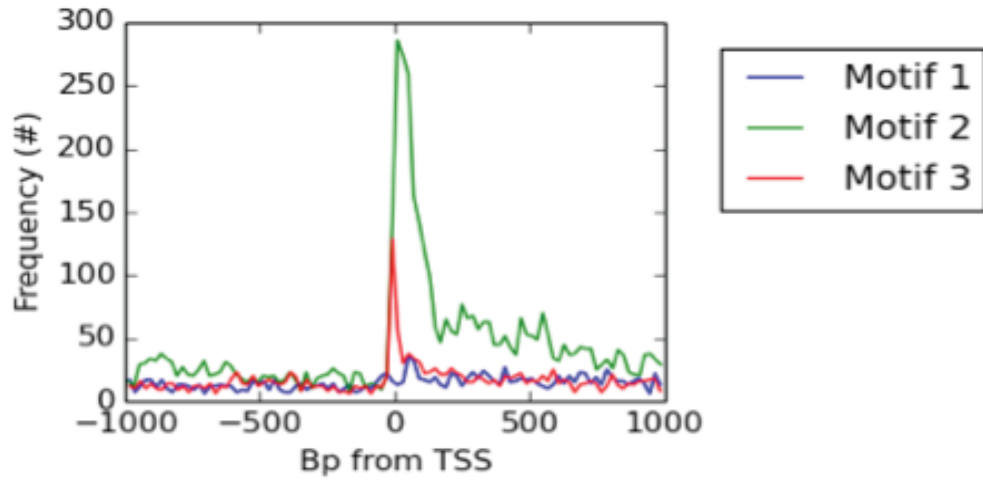
Base-pairs extracted around 3' transcription termination sites (+-)



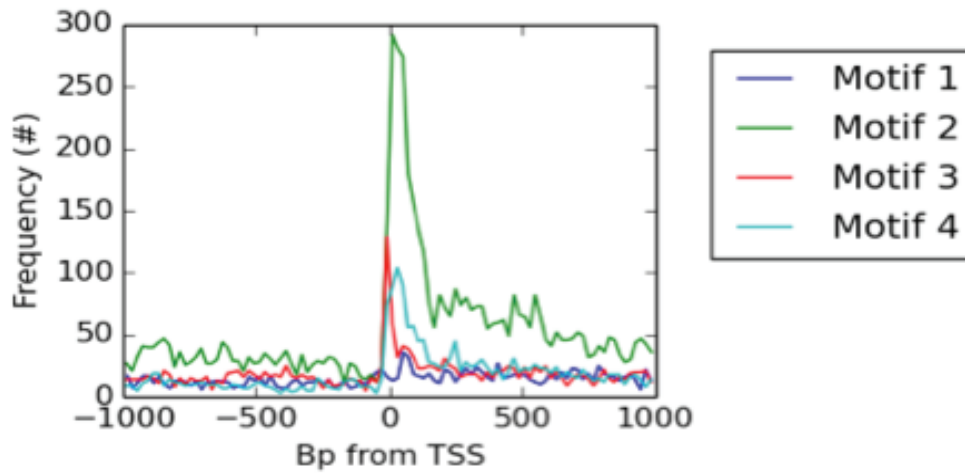
Supplemental Figure 2a. All motifs discovered around transcription termination sites at varying distances, and their distributions.

Bp searched
around TTS

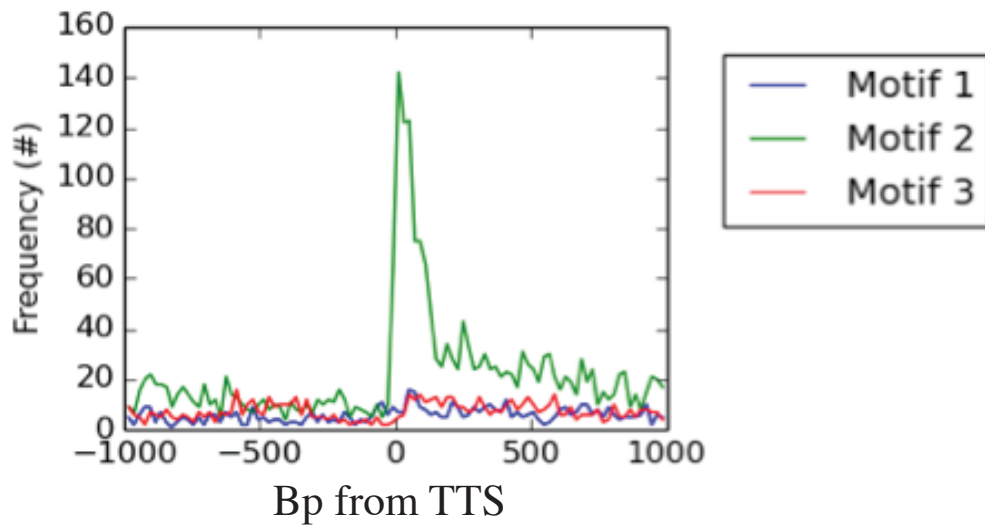
50



100








200











Supplemental Figure 2b. All motifs discovered around transcription termination sites at varying distances, and their distributions.

Motif

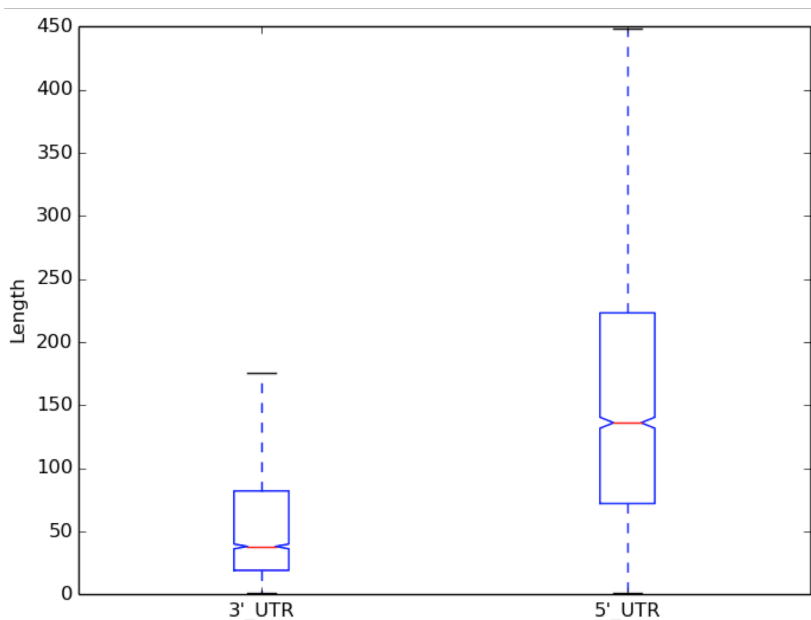




Structure Name	Image	$\Delta G(\text{kcal.mole}^{-1})$	T_m ($^{\circ}\text{C}$)	$\Delta H(\text{kcal.mole}^{-1})$	$\Delta S(\text{cal.K}^{-1}\text{mole}^{-1})$
1		-2.99	62.5	-26.8	-79.85
2		-2.64	55.4	-28.5	-86.75

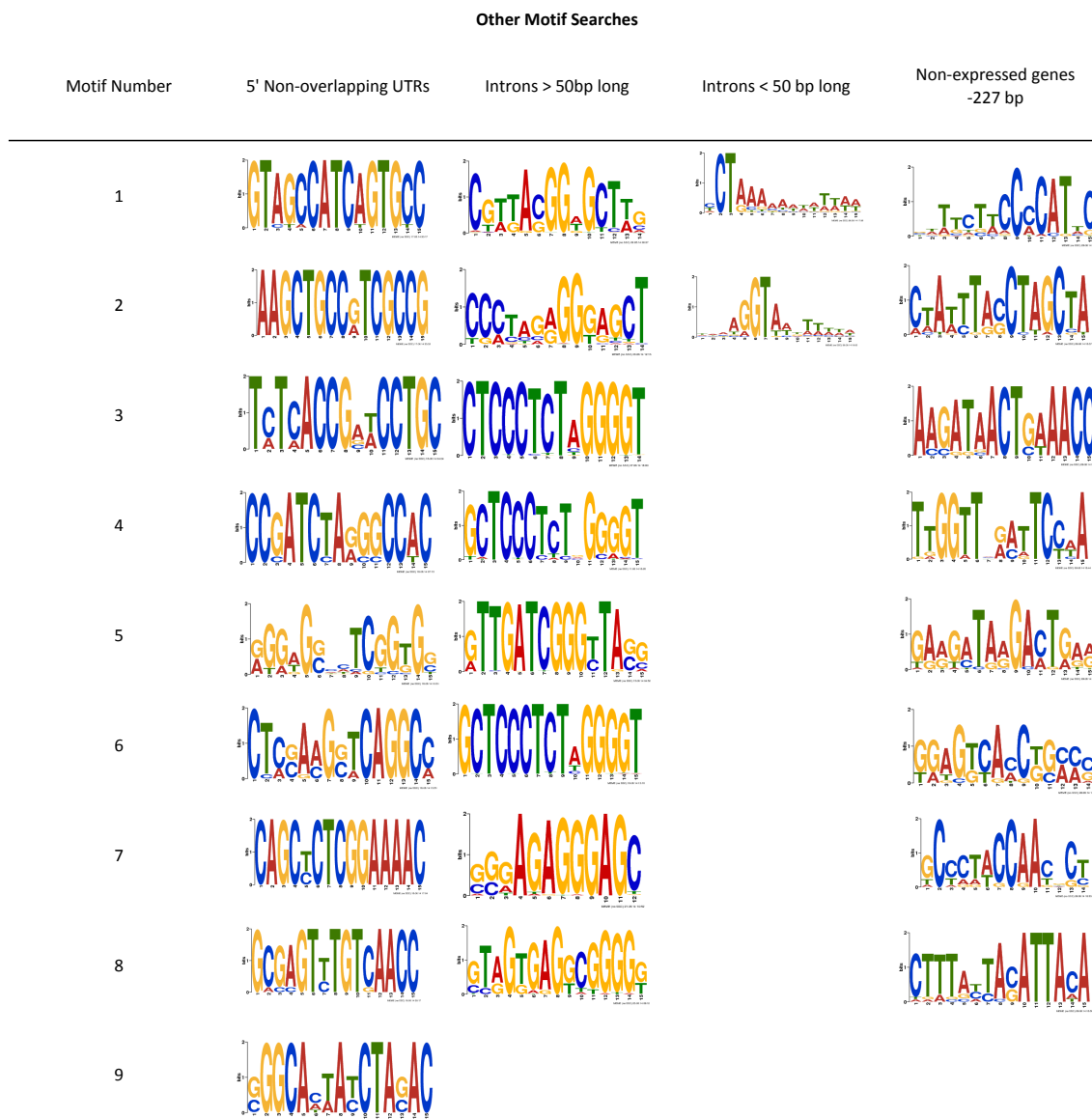
Structure Name	Image	$\Delta G(\text{kcal.mole}^{-1})$	T_m ($^{\circ}\text{C}$)	$\Delta H(\text{kcal.mole}^{-1})$	$\Delta S(\text{cal.K}^{-1}\text{mole}^{-1})$
1		1.84	-68.8	-4	-19.58
2		2.71	-273.1	0	-9.09

Structure Name	Image	$\Delta G(\text{kcal.mole}^{-1})$	T_m ($^{\circ}\text{C}$)	$\Delta H(\text{kcal.mole}^{-1})$	$\Delta S(\text{cal.K}^{-1}\text{mole}^{-1})$
1		1.25	-7	-10.4	-39.08
2		1.52	-54.3	-4.2	-19.19
3		1.73	-17.4	-10.4	-40.67
4		1.85	-21.6	-10	-39.76
5		2.11	-55.5	-5.7	-26.18
6		2.12	-81.9	-3.8	-19.87

Supplemental Figure 2c. Each motif from the MEME search using $\pm 50\text{bp}$ sequences around the transcription termination site, and the consensus sequence submitted to Integrated DNA Technology's OligoAnalyzer 3.1 (<https://www.idtdna.com/calc/analyzer>) and the output displayed above.



Supplemental Figure 3. Length of annotated 3' and 5' untranslated regions (UTRs) in *T. parva*. Box and whisker plots showing that 5' UTRs are significantly longer than 3' UTRs. Outliers are not shown, to simplify the visual display.



Supplemental Figure 4a. Other MEME motif searches found novel motifs. Regions that were searched include all 5' untranslated regions that did not overlap any adjacent gene, introns longer than 50bp long, introns less than 50bp long, and non-expressed genes (FPKM < 1).

Region	Consensus Motif	Region	Consensus Motif
25bp TSS	CCCCATNNN	25bp TTS	TATACAC
	AAATGTGTA		GTATATTATAT
	CCCCTCCAACNCCAT		GAGTANGAGATTA
	AAGATTCCA	50BP TTS	ATGTGTA
	GTGTGGCCTAGTGGT		ACCCCTAGAGGGGAGC
	GATAAGACTGAAACC		GTGGATTGGGTGG
	CTCAGTTGGTAGAGC		TAGTATATTATA
	AAGCTGGATCAATTG	100bp TTS	AAACCCACACACCTA
	CACATTTGTATG		ATACTATATATAATA
GTGTAAGAGATTGG	AGGTCCCTCTCGGGG		
50bp TSS	TCCCCATNN	5' Non-overlapping UTRs	GTAGCCATCAGTGCC
	AAATGTGTAATA		AAGCTGCCGTCGCCG
	CGATTCCAT		TCTACCCGATCCTGC
	CCTCCCTNNCAC		CCGATCTAGGGCCAC
	TGGTGTGGTGTCGG		GGGAGGCCTCGGTGG
	GTCNGACGCTCTACCCACCACA		CTCGAAGGTCAGGCC
	ACCCTGTATGATAGC		CAGCTGTCGGAAAAC
	GCTAACCAAAGCCAA		GCGAGTTTGTCAACC
	GCTTTGGTGTCTTCC		GGGCACTATCTAGAC
100bp TSS	TTCATTCCCCATCNA	Introns > 50bp long	CGTTACGGAGCTTG
	AAATGTGTAATA		CCCTAGAGGGGAGCT
	NAAGATTCC		CTCCCTCTAGGGGT
	CGGGAATCGAACCCG		GCTCCCTCTGGGGGT
	CCTGATCCACCTCCC		GTTGATCGGGTTAGG
	GTA TAGCCCTACCA		GCTCCCTCTAGGGGT
	ACACCAC		GGGAGAGGGGAGC
	CCTCAGTGGCCCCAG		GTAGTGAGGCGGGGG
	CCCTCAGCCAC		Introns < 50bp long
	NACCGAAACCAATAC	TAAAGGTAATTTTAA	
		Non-expressed genes (-227bp)	GTTCTTCCCCATTC
			CTATTTACCTAGCTA
			AAGATAACTGAAACC
			TTGGTTNGATTCCAA
			GAAGATAAGACTGAA
			GGAGTCACCTGCCC
			GCCCTACCAACNCT
		CTTTATTACATTACA	

Supplemental Figure 4b. This table summarizes all motifs found in the present study.