

**Table S1 The Localization Factor (LF) for all continuous ( $X_4-N_0-X_4$ ) and split 8-mers ( $X_4-N_{1-30}-X_4$ ) was determined.** Next, the probability of a non-random distribution was determined and the top 20 for different insert length was determined. For column  $X_4-N_{5-30}-X_4$ , the insert length that produced the highest LF is presented. The most localized split 8-mers are further extended to determine the TFBS or overlapping TFBS and are presented in the column next to the sequence motifs. Two sequences representing the ETS $\Leftrightarrow$ ETS motif and the ETS $\Leftrightarrow$ CRE motif are highlighted in red.

$X_4-N_0-X_4$			$X_4-N_2-X_4$			$X_4-N_4-X_4$			$X_4-N_{5-30}-X_4$		
Sequence	Predicted		Sequence	Predicted		Sequence	Predicted		Sequence	Predicted	
	TFBS	P-value		TFBS	P-value		TFBS	P-value		TFBS	P-value
CCGGAAGT	ETS	97	GTGA--TCAC	CRE	56	TATA----CGGC	CpG-TATA	92	ACGT-N <sub>9</sub> -GCGC	CpG-CRE	56
TATAAAGG	TATA	85	CGGA--TGAC	ETS-CRE	47	TATA----CGGG	CpG-TATA	36	GCGC-N <sub>7</sub> -TGAC	CpG-CRE	44
CGGAAGTG	ETS	70	GCCC--CCCC	SP1	47	CAAT----GCGC	CpG-CAAT	28	CGCC-N <sub>8</sub> -TATA	CpG-TATA	38
GCCAATCG	CAAT	61	TATA--CGGC	CpG-TATA	46	TATA----GGGC	CpG-TATA	25	CAAT-N <sub>5</sub> -GCCC	CpG-CAAT	37
CAGCCAAT	CAAT	58	TATA--AGGG	CpG-TATA	44	TATA----GCGC	CpG-TATA	24	ACGT-N <sub>8</sub> -CGCG	CpG-CRE	36
CCCCGCC	SP1	54	TGAC--CACA	CRE	36	TATA----CCGG	CpG-TATA	21	CGCG-N <sub>6</sub> -TGAC	CpG-CRE	35
GCCATCTT		53	TATA--CGCG	CpG-TATA	35	TATA----GCCG	CpG-TATA	19	CCCT-N <sub>28</sub> -TCCC		33
TATAAAG	TATA	53	TGAC--CACG	CRE	33	AGGC----CCCC		19	TATA-N <sub>6</sub> -GCGG	CpG-TATA	33
GGCCAATC	CAAT	47	CAAT--GCGC	CpG-CAAT	28	CGGA----ACGT	ETS-CRE	18	TGAC-N <sub>11</sub> -GCGC	CpG-CRE	32
CCCGAAG	ETS	47	GGAA--GGAA	ETS-ETS	26	CAAT----AGCC	CpG-CAAT	18	TTAT-N <sub>18</sub> -CGCC	CpG-TATA	29
ACCAATCA	CAAT	46	GGGA--TGTA		21	GCGC----ACGT	CpG-CRE	17	CAAT-N <sub>24</sub> -CCGC	CpG-CAAT	28
GGAAGTGA	ETS-CRE	46	CCCC--CCCT		20	CCAA----CGCC	CpG-CAAT	15	CGCG-N <sub>5</sub> -TATA	CpG-TATA	28
GCCCCGCC	SP1	44	GCGG--CAAT	CpG-CAAT	19	TGCG----CGCG		15	GGCG-N <sub>8</sub> -TGAC	CpG-CRE	27
GCCAATCA	CAAT	44	ACCG--AGTG	ETS	18	CAAT----GCCC	CpG-CAAT	15	GGCG-N <sub>19</sub> -CAAT	CpG-CAAT	26
GCGGAAGT	ETS	43	AGCG--AGTG	ETS	18	CAAT----GGCG	CpG-CAAT	14	TTAT-N <sub>15</sub> -CCCG	CpG-TATA	25
CCAATGGG	CAAT	41	CGCC--AAGT	ETS	18	AGCA----AATG		14	TATA-N <sub>6</sub> -GCCG	CpG-TATA	24
CGGCAAT	CAAT	40	TATA--CCCG	CpG-TATA	18	TATA----GGGA	CpG-TATA	14	TATA-N <sub>11</sub> -CGGC	CpG-TATA	24
CCAATCAG	CAAT	39	TATA--AGCG	CpG-TATA	17	CCGC----ATAA	CpG-TATA	13	ACCA-N <sub>24</sub> -ATGG		23
AAGTGACG	CRE	37	CCCC--CCAC		17	TATA----GGCC	CpG-TATA	13	CGCC-N <sub>7</sub> -TTAT	CpG-TATA	23
CCATCTTG		37	CGCG--TTAT	CpG-TATA	17	CAAT----GGGC	CpG-CAAT	13	TCAT-N <sub>7</sub> -CGCC	CpG-CRE	23