SUPPLEMENTARY MATERIAL FOR THE ARTICLE:

Corcoran *et al.* "*FOOTER**: a quantitative comparative genomics method for efficient recognition of cis-regulatory elements"*

Supplementary materials include:

- *Table Suppl1 with predictions of FOOTER on the 72 sites test set.*
- *Detailed description of the comparison of FOOTER with ConSite and rVista (includes Table Suppl2 with the results).*

1. PREDICTIONS OF *FOOTER* ON A 72 SITES TEST SET.

Factor	Promoter	Position	*Site (experimentally defined)*	FOOTER Prediction	FOOTER score	Ref
NFAT	IL2	-45	TA**TTTTTCC**A	TTTTTCC	8.8	(Rooney et al. 1995)
	IL2	-90	T**TGAAAAT**ATGTGTAATA	TGAAAAT	8.1	(Rooney et al. 1995)
	IL2	-135	A**GGAAAAA**CAAAGGTAAT	GGAAAAA	13.3	(Rooney et al. 1995)
	IL2	-160	AGAAATTCCAGAGAGTCA	*not found*		(Rooney et al. 1995)
	IL2	-280	A**GGAAAAA**CTGTTTCATA	GGAAAAA	13.3	(Rooney et al. 1995)
	IL2	*-2240*		**GGAAAAT**	8.0	
	IL4	-106	GTAAACTC**ATTTTCC**CTTG GTTTC	ATTTTCC	13.3	(Szabo et al. 1993)
	IL4	-121	GTAATAAA**ATTTTCC**AATG TAAAC	ATTTTCC	13.3	(Szabo et al. 1993)
	IL4	-238	GGTGTTTC**ATTTTCC**AATT GGTCTGATTTCACA**GGAAA** **AT**TTACC	ATTTTCC *And/Or* GGAAAAT	12.7 12.7	(Szabo et al. 1993)
	IL4	-287	TATGGTGTAATTTCCTATGC TTGA	*not found*		(Szabo et al. 1993)
	IL4	-406	GCAGTCCTCCTGGGGAAAG ATAGAGTAATATCA	*not found*		(Burke et al. 2000)
	IL4	*-1226*		**GGAAAAA**	8.8	
HNF-1α	PEPCK	-200	CAAC**ATTCATTAAC**AACCA CAAG**TTCAATCATT**ATCTC CCTGGAGTTTAT	ATTCATTAAC *And/Or* TTCAATCATT	8.6 8.4	(Patel et al. 1994)
	G6Pase	-271	CGGGGACCAGGAGGGCAG ACCCTTGCACTGCCAAGAA GCATGCCAAA**GTTAATCAT** **T**GGCCCTGCTGAGTAC	GTTAATCATT	9.0	(Lin et al. 1997)
	G6Pase	*-462*		**AATTAATAAC**	13.1	
	Pdx-1	-2114	AGCCTCTTTTCTTTCTGCAG GGCCG AGCAA**AAATATTAAA**TGG AAGCAAATGAAGCATCGA AATGGAGACCAA	AAATATTAAA	13.0	(Melloul et al. 2002)
	Pdx-1	-2980	GGTTTTCTCAACTCAGGGC ATAATTTTATTTAATTTTTA ATAGCAAAGTA ATTTTTGGGATGAATATGG TTTTAAAAATTAAAGTTCG TGTAATCCTATC	*not found*		(Melloul et al. 2002)
HNF3β	PEPCK	-455	AGCAGGTACAGACATTATC TAGAAGTCTCATGGCTCAG AGCTGAATTTCCTTCTCATG ACCTTTGGCCGTGGGAGTG ACACTCACAGCTGTG**GTGT** **TTTGACA**ACCAGCAGCCAC CGGCACACAAAATGTGCAG CC	GTGTTTTGACA	8.5	(Croniger et al. 1998)
	Pdx-1	-2037	GAATAAATGAAGCGTCGAG ATGGAAGCCAATTTACCAA AATGCATGCAATTAGACCA GAAGTGC**TAAGCAAACAT** CCTGGGGTGTGGGTTAGGC AGGC	TAAGCAAACAT	8.7	(Melloul et al. 2002)
	Pdx-1	-2657	ACACTTTAATTGGTTTACA	TTATTTATCCA	8.8	(Melloul et al. 2002)

			GCCTTTTTTGT**TTATTTATC CA**TAAGAGCTGC TGTTAAATGGCTCGGGAAG GTGCTC			
	Pdx-1	-3025	GGTTTTCTCAACTCAGGGC ATAATT**TTATTTAATTT**TT AATAGCAAAGTA ATTTTTGGGATGAATATGG TTTTAAAAATTAAAGTTCG TGTAATCCTATC	TTATTTAATTT	8.1	(Melloul et al. 2002))
	Pdx-1	*-3065*		**TGGTTTGCTTT**	12.8	
	C7AH	-175	TCTGTTTGTTCTGGAGC	*not found*		(Crestani et al. 1998)
	C7AH	*-16*		**GTGTTTGCTTT**	8.4	
	C7AH	*-225*		**CTGTTTACTTC**	8.8	
HNF-3γ	G6PASE	-100	AGACAAACGTGGTTTTTGA GTCCAAAGATCAGGG	*not found*		(Lin et al. 1997)
	G6PASE	-146	CTGAACATGTTTGCATCAA CCTACTGGTGAT	*not found*		(Lin et al. 1997)
	G6PASE	-198	GGCCGATCAGGCTG**TTTTT GTGTGCCT**GTTTTTC	TTTTTGTGTGCC T	8.5	(Lin et al. 1997)
	G6PASE	*-47*		**GGGCATATAAA AC**	13.7	
	G6PASE	*-1920*		**GGGAAATTCAG GC**	12.6	
HNF-4	C7AH	-149	T**GGACTTAGTTCA**AGGCC GGGTAAT	GGACTTAGTTC A	7.6	(Crestani et al. 1998)
C/EBP-α	ACDC	-117	CCCACTCATTGGCTATTGG CCTTGACTGGGT**TGGCCAA T**GGTAAG	TGGCCAAT	7.9	(Park et al. 2004)
	ACDC	*-2089*		**TTTCACAAT**	12.6	
	ACDC	*-2017*		**TTGTGCAAT**	8.6	
C/EBP-β	PEPCK	-91	CCTGCCCCTTACGTCAGAG GCGAGCCT	*not found*		(Croniger et al. 1998)
	PEPCK	-248	AAATG**TTGTGTAA**GGACTC ACTAT	TTGTGTAA	8.8	(Croniger et al. 1998)
	PEPCK	-332	TGCCCTTGACCCCCACCTG ACAATTAAGGCAAGAGCCT GCAGT**TTGCATCA**GCA	TTGCATCA	8.3	(Croniger et al. 1998)
	Leptin	-58	G**TTGCGCAA**G	TTGCGCAA	8.9	(Mason et al. 1998)
	IL-6	-155	TAAAGGACGTCACA**TTGCA CAA**TCTT	TTGCACAA	8.2	(Xiao et al. 2004)
CREB	PEPCK	-91	CCTGCCCCT**TACGTCA**GAG GCGAGCCT	TACGTCA	7.9	(Patel et al. 1994)
	PEPCK	-455	AGCAGGTACAGACATTATC TAGAAGTCTCATGGCTCAG AGCTGAATTTCCTTCTCATG ACCTTTGGCCGTGGGAG**TG ACACC**TCACAGCTGTGGTG TTTTGACAACCAGCAGCCA CCGGCACACAAAATGTGCA GCC	TGACACC	8.4	(Croniger et al. 1998)
	CG-α	-44	AAACTGATCTGAGGGTTGC AATGTG ATATGATCAATT**GATGTCA** TGGTAATTATACCAAGTGC CATCCAATCACT	GATGTCA	12.9	(Fowkes, RC et al. 2002)
	CG-α	-132	TCTTCATAAGCTGTCCTT**G AGGTCA**CCACTACCTCAAA	GAGGTCA	12.2	(Fowkes, RC et al. 2002)

			ATGTCTAAAAAC			
	CDC2l2	-13	TCATCATTA**GGCGTCA**ACA CAGG	GGCGTCA	8.4	(Feng et al. 2004)
	hCGα	-146	AAAT**TGACGTC**ATGGTAA AAATTGACGTCATGGTAA	TGACGTC	12.6	(Ghosh, D et al. 2005, *in print*)
	hCGα	*-240*		**TGTCGTC**	7.7	
	BDKRB2	-94	GATCTAGGCTGGAAGTGGA GGGGGGAGGTGCCCAGGA GAGTGA**TGACATCA**	TGACATCA	7.6	(Saifudeen et al. 2005)
	IL-6	-155	TAAAG**GACGTCA**CATTGC ACAATCTT	GACGTCA	8.3	(Xiao et al. 2004)
	IL-6	*-1830*		**TGATGTC**	12.3	
	CART	-153	CGGCGGGCAT**TGACGTCA** AACGGCAGC	TGACGTCA	13.1	(Lakatos et al. 2002)
GR-α	PEPCK	-455	AGCAGGTACAGACATTATC TAGAAGTCTCATGGCTCAG AGCTGAATTTCCTTCTCATG ACCTTTGGCCGTGGGAGTG ACACTCACAGCTGTGGTGT TTTGACAACCAGCAGCCAC CGGCACACAAAATGTGCAG CC	*not found*		(Croniger et al. 1998)
	PEPCK	*-750*		**TCAGTTTCCT**	7.9	
T3R-α	PEPCK	-332	**TGCCCTTGACCC**CCACCT GACAATTAAGGCAAGAGCC TGCAGTTTGCATCAGCA	TGCCCTTGACCC	13.1	(Croniger et al. 1998)
Sp1	Leptin	-100	**GGGCGG**	GGGCGG	13.1	(Mason et al. 1998)
	NES	-171	CTTTT**CCGCCC**GGCCGG	CCGCCC	13.1	(Cheng et al. 2004)
	NES	-183	TAGGGA**CCGCCC**CTTTT	CCGCCC	13.1	(Cheng et al. 2004)
	NES	*-1173*		**CCTCCC**	12.5	
	MMP9	-560	ATTCCTTCCGCCCCCAGAT G	*not found*		(Takahra et al. 2004)
	MMP9	-520		**GGGAGG**	8.6	
SRF	EGR1	-88	TGCTTC**CCATATATGG**CCA TGT	CCATATATGG	9.8	(Christy and Nathans 1989)
	EGR1	-128	GTCCTT**CCATATTAGG**GCT TCC	CCATATTAGG	10.2	(Christy and Nathans 1989)
	EGR1	-358	CCAGCG**CCTTATATGG**AG TGGC	CCTTATATGG	13.2	(Christy and Nathans 1989)
	EGR1	-412	GAAACG**CCATATAAGG**AG CAGG	CCATATAAGG	14.4	(Christy and Nathans 1989)
	ACTA1	-100	AC**CCAAATATGG**CT	CCAAATATGG	13.4	(Wasserman and Fickett 1998)
	ACTA1	-181	CT**CCTTCTTTGG**TC	CCTTCTTTGG	7.8	(Wasserman and Fickett 1998)
	ACTA1	-227	CT**CCATATACGG**CC	CCATATACGG	13.1	(Wasserman and Fickett 1998)
	CaMh	*-62*	CTCCAAATTTAGGC	*not found*		(Molkentin et al. 1996)
	CaMh	-184	**CCTTTCATGG**	CCTTTCATGG	7.9	(Molkentin et al. 1996)
	CKMM	-1236	**CCATGTAAGG**	CCATGTAAGG	13.5	(Amacher et al. 1993)
	CKMM	*-178*		**CCATACAAGG**	8.9	
MEF-2	CaMh	-328	**ATTAAAAATAACT**GA	ATTAAAAATAA CT	7.6	(Molkentin and Markham 1993)
	CaMh	*-898*		**GTGTAAATTGC CC**	12.2	
	CaMh	*-1544*		**AGCTATATTGA GA**	8.1	

	CKMM	-1078	**TCTAAAAATAACT**	TCTAAAAATAACT	8.4	(Amacher et al. 1993)
	CKMM	-1194	T**GGTTATAATTAAC**C	GGTTATAATTAAC	8.7	(Amacher et al. 1993)
NF-Y	LPL	-65	**AGCCAATAGG**	AGCCAATAGG	7.7	(Previato et al. 1991)
	LPL	*-1795*		**AACCAATCAT**	10.2	
	Cyclin B2	-281	GTGTCTAAGAAAATTC**AGCCAATGAG**AGTGCGAGAGT GCATCTTGTGTT**GGCCAATGAG**AACAGCGACCCGTGC GCAGGGCCGCCCAATGGGG CGCAAGCGACGCGGTAT	AGCCAATGAG *And/Or* GGCCAATGAG	13.7 13.0	(Wasner et al. 2003)
	ACDC	-117	CCCA**CTCATTGGCT**ATTGG CCTTGACTGGGTTGGCCAA TGGTAAG	CTCATTGGCT	7.6	(Park et al. 2004)
	ACDC	*-2229*		**AACCAAACCG**	8.5	
NF-κB	IL-6	-62	GTG**GGATTTTCCC**A	GGATTTTCCC	8.3	
	MMP9	-600	CCAG**TGGAATTCCC**CAGC CT	TGGAATTCCC	13.0	(Takahra et al. 2004)
	MMP9	*-2112*		**GGCAAATTCC**	13.2	
	Vcam-1	-90	GAAGGTCA**GGAAAAGCCA** GAGATTTATA	GGAAAAGCCA	8.0	(Tu et al. 2001)
	iNOS	-114	G**GGGACTCTCC**	GGGACTCTCC	9.1	(Wei et al. 2004)
	iNOS	-1044	GGGGATTTTCC	*not found*		(Wei et al. 2004)
	iNOS	*-2760*		**GGCATTTCTC**	7.6	
NF-1	PEPCK	-116	TCA**GTTCCAA**ACCTGACCA TGGCTAT	GTTCCAA	7.7	(Croniger et al. 1998)
GATA-1	Vcam-1	-117	CAGTAA**AGATAG**CCTTTTG GAGTCGAAGATGAGGAAA AGCCTGTATTTTA TAGTCTTGGAAGTGTCTTCT TTTGCCAGGACAGAGAGAG GAGCTTCAGCA	AGATAG	12.4	(Tu et al. 2001)
GATA-3	CG-α	-346	**TTTCTG**TTTCCTGTTGAAA TAATGTAATCCTGAAAATG TTTTTTTTTATCCTGCTTTA TGAAA	TTTCTG	12.1	(Fowkes, RC et al. 2002)
	CG-α	*-394*		**CAGATG**	8.5	
AP-1	PEPCK	-91	CCTGCCCCTTACGTCAGAG GCGAGCCT	*not found*		(Croniger et al. 1998)
	PEPCK	-285	TTTGCATCAGCAACAGGCA GGGTCAAAGT**TTAGTCA**AT C	TTAGTCA	8.5	(Croniger et al. 1998)
	Vcam-1	-346	**TGACTCA**TCAAAAGAAAT AACTTTTTCCTTTCTCTTGT AAGAGA	TGACTCA	13.0	(Tu et al. 2001)
	MMP9	-79	GGAAGC**TGAGTCA**AAGAA GGCT	TGAGTCA	9.1	(Takahra et al. 2004)
	MMP9	-533	TATAAAGCA**TGAGTCA**GA CACCTC	TGAGTCA	13.0	(Takahra et al. 2004)

Table Suppl1. **Results of FOOTER predictions of known binding sites of various transcription factors.** The analysis of twenty-four promoter regions is presented. The Table contains the names of the TFs and the name of the gene whose promoter region was analyzed, the position that the site has been identified, the reported sequence in the literature, the FOOTER predicted sequence and score. Predictions in **bold letters** are

unconfirmed. Unconfirmed predictions in **underlined blue** letters are outside the promoter regions examined in the corresponding publications. Overall, FOOTER exhibited 83% sensitivity and 72% specificity over the 3 kb region. Note that if two sites are found within a verified binding region it is still considered as only 1 true positive.

2. COMPARISON OF *FOOTER* WITH EXISTING METHODS.

We compared FOOTER with two algorithms (on-line versions) that use comparative genomics and PSSM models to identify phylogenetically conserved signals in mammals: *ConSite* (Lenhard et al. 2003) and *rVista* (Loots et al. 2002). The results are presented in **Table Suppl1**.

Comparison with *ConSite*. Program *ConSite* (http://www.phylofoot.org; (Lenhard et al. 2003)) uses a set of well-curated PSSM models to scan the human and mouse promoter regions and identify binding sites. Unlike FOOTER, however, it does not use species-specific matrices for the factors; it uses overall vertebrate, plant, or insect PSSM models instead. Also, *ConSite* does not employ a combined scoring scheme for the pairs of binding sites. Instead, it uses a filtering mechanism in which patterns are filtered out if their PSSM score does not meet a specified threshold (default: 10 bits). Another aspect of the *ConSite* program is that the sites that it considers are located within conserved regions (default percent identity: 80% in a window of 50 bases). Although, Wasserman *et al.* have reported that 98% of confirmed muscle-specific sites had been found to reside in conserved regions (Wasserman et al. 2000), this cannot be taken as a general rule. For example, in another study, Levy *et al.* (Levy and Hannenhalli 2002) reported that only 50% of the sites were found in conserved regions. Unlike *ConSite*, FOOTER does not apply any conservation threshold, allowing patterns to be compared irrespective of the region in which they are residing.

ConSite does not provide PSSM models for eight TFs in the test set. Instead, it gives the option for uploading a user-specific PSSM model. However, for comparison purposes, we used its on-line version (http://www.phylofoot.org) with the available matrices. The available matrices allow for predictions to be made for fourty nine of the seventy two verified sites (**Table Suppl2**). When it was ran with its default parameters (identity percentage: 80%, or value returned with promoter region; score threshold: 80%), *ConSite* identified twenty three of these sites (S_N= 47%) and produced a total of fifteen additional predictions (S_P= 61%; **Suppl2**). We also ran *ConSite* with a lowered score threshold (70%) and identity percentage (70% if it was defaulted at > 70%) in order to increase its sensitivity and better mimic FOOTER parameters. These parameters were used for scanning the promoters for which *ConSite* did not perform well with the defaults, namely the promoters of the genes G6Pase, CaMh, C7AH, CG-α, V-cam, NF-κB and PEPCK. This second analysis of the PEPCK and MMP9 promoters resulted in many additional predictions (false positives), so we decided to use the results of the analysis with the default parameters for these promoters. In the other promoters, *ConSite* correctly predicted eleven more sites, which increased its sensitivity to S_N= 69.4%. Furthermore, it made thirteen additional predictions, thus decreasing its specificity to S_P= 54.8%. Hence, in the best case, *ConSite*, exhibited a sensitivity of 69.4%, a value that is similar to their reported value (Lenhard et al. 2003), and a specificity of 61%. Note that searching for the same fourty nine sites, FOOTER was able to find 44 (S_N= 89.8%) making fifteen additional predictions (S_P= 74.6%) (**Table 2**).

Comparison with *rVista*. *rVista* (Loots et al. 2002) uses the general TRANSFAC matrices (Wingender 2004) (i.e., not species- or mammalian-specific) to scan the promoters of the genes and a filtering mechanism (options: “conserved” and “aligned sequences”) to exclude non-conserved sites. When tested on the NFAT sites of various interleukin promoters, *rVista* reported a sensitivity of 88% (Loots et al. 2002). However, we believe

that this high performance is due to the fact that *rVista* uses a low PSSM threshold for the initial scanning, so that it detects even subtle "signals." This results in high sensitivity scores, but at the cost of lower specificity. *rVista* relies on the genomic conservation to decrease the high false positive rate and it succeeds in that, but only to a certain degree.

We tested *rVista* on the same TFBS set as *ConSite* and FOOTER. The only factor that *rVista* doesn't provide PSSM model for is HNF-3γ. Running with the option "conserved" (option "aligned sequences" gave slightly worse results), it performed well in finding fifty four of the sixty nine binding sites, but it produced 189 additional predictions (false positives; **Table Suppl2**). So, its overall performance was S_N= 78% and S_P= 22%. Note that searching for the same sixty nine sites, FOOTER was able to find fifty nine (S_N= 85.5%) making twenty one additional predictions (S_P= 73.8%) (**Table Suppl2**).

Factor	**Promoter**	**No . of sites**	***Footer***	***ConSite (def)***	***ConSite (70%, 70%)***	***rVista***
NFAT	IL2	5	4TP, 1FP	**N/A**		5 TP 24 FP
	IL4	5	3TP, 1FP	**N/A**		5 TP, 13 FP
HNF-1α	PEPCK	1	1TP, 0FP	1 TP, 0 FP	*too many FP*	1 TP, 1 FP
	G6Pase	1	1TP, 1FP	0 TP, 0 FP	1 TP, 0 FP	1 TP, 4 FP
	Pdx-1	2	1TP, 0FP	0 TP, 0 FP	2 TP, 1 FP	1 TP, 1 FP
HNF-3β	C7AH	1	0TP, 2FP	0 TP, 2 FP	1 TP, 3 FP	1 TP, 3 FP
	PEPCK	1	1TP, 0FP	0 TP, 0 FP	*too many FP*	1 TP, 2 FP
	Pdx-1	3	3TP, 1FP	2 TP, 0 FP		3 TP, 8 FP
HNF-3 □	G6Pase	3	1TP, 2FP	**N/A**		**N/A**
HNF-4	C7AH	1	1TP, 0FP	**N/A**		1 TP, 0 FP
C/EBP-α	ACDC	1	1TP, 2FP	**N/A**	**N/A**	1TP, 12FP
C/EBP-β	PEPCK	3	2TP, 0FP	1 TP, 1 FP	*too many FP*	1 TP, 6 FP
	IL-6	1	1TP, 0FP	1TP, 1FP		1TP, 9FP
	Leptin	1	1TP, 0FP	1 TP, 0 FP		1 TP, 7 FP
CREB	PEPCK	2	2TP, 0FP	1 TP, 0 FP		2 TP, 6 FP
	CG-a	2	2TP, 0FP	0 TP, 0 FP	0 TP, 0 FP	2 TP, 3 FP
	hCG□	1	1TP, 1FP	0TP, 1FP	1TP, 3FP	1TP, 3FP
	IL-6	1	1TP, 1FP	1TP, 0FP		1TP, 2FP
	CDC2L2	1	1TP, 0FP	0TP, 0FP	0TP, 1FP	0TP, 0FP
	BDKRB2	1	1TP, 0FP	1TP, 0FP		1TP, 0FP
	CART	1	1TP, 0FP	1TP, 0FP		1TP, 0FP
GR-α	PEPCK	1	0TP, 1FP	**N/A**		0 TP, 0 FP
T3R	PEPCK	1	1TP, 0FP	**N/A**		0 TP, 3 FP
Sp1	Leptin	1	1TP, 0FP	1 TP, 1 FP		1 TP, 16 FP
	NES	2	2TP, 1FP	0TP, 0FP	2TP, 2FP	2TP, 5FP
	MMP9	1	0TP, 1FP	0TP, 2FP	*too many FP*	1TP, 8FP
SRF	EGR1	4	4TP, 0FP	3 TP, 0 FP		3 TP, 3 FP

	ACTA1	3	3TP, 0FP	3 TP, 0 FP		3 TP, 0 FP
	CaMh	2	1TP, 0FP	0 TP, 0 FP	1 TP, 0 FP	0 TP, 0 FP
	CKMM	1	1TP, 1FP	1 TP, 1 FP		0 TP, 0 FP
MEF-2	CaMh	1	1TP, 2FP	0 TP, 2 FP	1 TP, 2 FP	1 TP, 0 FP
	CKMM	2	2TP, 0FP	1 TP, 0 FP		0 TP, 0 FP
NF-Y	LPL	1	1TP, 1FP	1TP, 0FP		1TP, 5FP
	ACDC	1	1TP, 1FP	0TP, 0FP	0TP, 1FP	0TP, 2FP
	Cyclin B2	1	1TP, 0FP	0TP, 0FP	1TP, 0FP	1TP, 1FP
GATA-1	Vcam-1	1	1TP, 0FP	1 TP, 4 FP		1 TP, 6 FP
GATA-3	CG-a	1	1TP, 1FP	0 TP, 0 FP	0 TP, 0 FP	1 TP, 3 FP
NF-κB	Vcam-1	1	1TP, 0FP	0 TP, 0 FP	1 TP, 2 FP	0 TP, 0 FP
	IL-6	1	1TP, 0FP	1TP, 0FP		1TP, 2FP
	MMP9	1	1TP, 1FP	1TP, 0FP	1 TP, 1 FP	1TP, 2FP
	iNOS	2	1TP, 1FP	0TP, 0FP	0TP, 2FP	1TP, 2FP
NF-1	PEPCK	1	1TP, 0FP	**N/A**		0TP, 12FP
AP-1	PEPCK	2	1TP, 0FP	**N/A**		2TP, 8 FP
	Vcam-1	1	1TP, 0FP	**N/A**		1TP, 5FP
	MMP9	2	2TP, 0FP	**N/A**		2TP, 2FP

Table Suppl2. **Results of predictions of programs FOOTER,** *ConSite* (Lenhard et al. 2003) **and** *rVista* (Loots et al. 2002)**.** *ConSite* ran with its default parameters and also with 70% for score threshold and the minimum between 70% or the default for identity percentage. *rVista* ran with the option "conserved". Based on the results of the optimal runs of these programs on the same promoter set, their sensitivity and specificity values were measured to be 69% and 55% for *ConSite* (on 49 sites) and 78% and 22% for the *rVista* (on 69 sites), respectively. By comparison, FOOTER achieved a sensitivity of 83% and specificity of 72% (on 72 sites).

REFERENCES

Amacher, S.L., J.N. Buskin, and S.D. Hauschka. 1993. Multiple regulatory elements contribute differentially to muscle creatine kinase enhancer activity in skeletal and cardiac muscle. *Mol Cell Biol* **13:** 2753-2764.

Burke, T.F., V. Casolaro, and S.N. Georas. 2000. Characterization of P5, a novel NFAT/AP-1 site in the human IL-4 promoter. *Biochem Biophys Res Commun* **270:** 1016-1023.

Cheng, L., Z. Jin, L. Liu, Y. Yan, T. Li, X. Zhu, and N. Jing. 2004. Characterization and promoter analysis of the mouse nestin gene. *FEBS Lett* **565:** 195-202.

Christy, B. and D. Nathans. 1989. Functional serum response elements upstream of the growth factor-inducible gene zif268. *Mol Cell Biol* **9:** 4889-4895.

Crestani, M., A. Sadeghpour, D. Stroup, G. Galli, and J.Y. Chiang. 1998. Transcriptional activation of the cholesterol 7alpha-hydroxylase gene (CYP7A) by nuclear hormone receptors. *J Lipid Res* **39:** 2192-2200.

Croniger, C., P. Leahy, L. Reshef, and R.W. Hanson. 1998. C/EBP and the control of phosphoenolpyruvate carboxykinase gene transcription in the liver. *J Biol Chem* **273:** 31629-31632.

Feng, Y., A.C. Goulet, and M.A. Nelson. 2004. Identification and characterization of the human Cdc2l2 gene promoter. *Gene* **330:** 75-84.

Lakatos, A., G. Dominguez, and M.J. Kuhar. 2002. CART promoter CRE site binds phosphorylated CREB. *Brain Res Mol Brain Res* **104:** 81-85.

Lenhard, B., A. Sandelin, L. Mendoza, P. Engstrom, N. Jareborg, and W.W. Wasserman. 2003. Identification of conserved regulatory elements by comparative genome analysis. *J Biol* **2:** 13.

Levy, S. and S. Hannenhalli. 2002. Identification of transcription factor binding sites in the human genome sequence. *Mamm Genome* **13:** 510-514.

Lin, B., D.W. Morris, and J.Y. Chou. 1997. The role of HNF1alpha, HNF3gamma, and cyclic AMP in glucose-6-phosphatase gene activation. *Biochemistry* **36:** 14096-14106.

Loots, G.G., I. Ovcharenko, L. Pachter, I. Dubchak, and E.M. Rubin. 2002. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* **12:** 832-839.

Mason, M.M., Y. He, H. Chen, M.J. Quon, and M. Reitman. 1998. Regulation of leptin promoter function by Sp1, C/EBP, and a novel factor. *Endocrinology* **139:** 1013-1022.

Melloul, D., S. Marshak, and E. Cerasi. 2002. Regulation of pdx-1 gene expression. *Diabetes* **51 Suppl 3:** S320-325.

Molkentin, J.D., S.M. Jobe, and B.E. Markham. 1996. Alpha-myosin heavy chain gene regulation: delineation and characterization of the cardiac muscle-specific enhancer and muscle-specific promoter. *J Mol Cell Cardiol* **28:** 1211-1225.

Molkentin, J.D. and B.E. Markham. 1993. Myocyte-specific enhancer-binding factor (MEF-2) regulates alpha-cardiac myosin heavy chain gene expression in vitro and in vivo. *J Biol Chem* **268:** 19512-19520.

Park, S.K., S.Y. Oh, M.Y. Lee, S. Yoon, K.S. Kim, and J.W. Kim. 2004. CCAAT/enhancer binding protein and nuclear factor-Y regulate adiponectin gene expression in adipose tissue. *Diabetes* **53:** 2757-2766.

Patel, Y.M., J.S. Yun, J. Liu, M.M. McGrane, and R.W. Hanson. 1994. An analysis of regulatory elements in the phosphoenolpyruvate carboxykinase (GTP) gene which are responsible for its tissue-specific expression and metabolic control in transgenic mice. *J Biol Chem* **269:** 5619-5628.

Previato, L., C.L. Parrott, S. Santamarina-Fojo, and H.B. Brewer, Jr. 1991. Transcriptional regulation of the human lipoprotein lipase gene in 3T3-L1 adipocytes. *J Biol Chem* **266:** 18958-18963.

Rooney, J.W., Y.L. Sun, L.H. Glimcher, and T. Hoey. 1995. Novel NFAT sites that mediate activation of the interleukin-2 promoter in response to T-cell receptor stimulation. *Mol Cell Biol* **15:** 6299-6310.

Saifudeen, Z., S. Dipp, H. Fan, and S.S. El-Dahr. 2005. Combinatorial Control of the Bradykinin B2 Receptor Promoter by p53, CREB, KLF-4 and CBP: Implications for Terminal Nephron Differentiation. *Am J Physiol Renal Physiol*.

Szabo, S.J., J.S. Gold, T.L. Murphy, and K.M. Murphy. 1993. Identification of cis-acting regulatory elements controlling interleukin-4 gene expression in T cells: roles for NF-Y and NF-ATc. *Mol Cell Biol* **13:** 4793-4805.

Takahra, T., D.E. Smart, F. Oakley, and D.A. Mann. 2004. Induction of myofibroblast MMP-9 transcription in three-dimensional collagen I gel cultures: regulation by NF-kappaB, AP-1 and Sp1. *Int J Biochem Cell Biol* **36:** 353-363.

Tu, Z., V.R. Kelley, T. Collins, and F.S. Lee. 2001. I kappa B kinase is critical for TNF-alpha-induced VCAM1 gene expression in renal tubular epithelial cells. *J Immunol* **166:** 6839-6846.

Wasner, M., U. Haugwitz, W. Reinhard, K. Tschop, K. Spiesbach, J. Lorenz, J. Mossner, and K. Engeland. 2003. Three CCAAT-boxes and a single cell cycle genes homology region (CHR) are the major regulating sites for transcription from the human cyclin B2 promoter. *Gene* **312:** 225-237.

Wasserman, W.W. and J.W. Fickett. 1998. Identification of regulatory regions which confer muscle-specific gene expression. *J Mol Biol* **278:** 167-181.

Wasserman, W.W., M. Palumbo, W. Thompson, J.W. Fickett, and C.E. Lawrence. 2000. Human-mouse genome comparisons to locate regulatory sites. *Nat Genet* **26:** 225-228.

Wei, J., H. Guo, C. Gao, and P.C. Kuo. 2004. Peroxide-mediated chromatin remodelling of a nuclear factor kappa B site in the mouse inducible nitric oxide synthase promoter. *Biochem J* **377:** 809-818.

Wingender, E. 2004. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. *In Silico Biol* **4:** 55-61.

Xiao, W., D.R. Hodge, L. Wang, X. Yang, X. Zhang, and W.L. Farrar. 2004. Co-operative functions between nuclear factors NFkappaB and CCAT/enhancer-binding protein-beta (C/EBP-beta) regulate the IL-6 promoter in autocrine human prostate cancer cells. *Prostate* **61:** 354-370.