Journal of Alzheimer's Disease 34 (2013) 1–4
IOS Press

Supplementary Data

Haplotype of Single Nucleotide Polymorphisms in Exon 6 of the MZF-1 Gene and Alzheimer's Disease

Elisa Porcellini[a,*], Ilaria Carbone[a], Pier Luigi Martelli[b], Manuela Ianni[a], Rita Casadio[b], Annalisa Pession[a] and Federico Licastro[a]
[a]*Department of Experimental Pathology, School of Medicine, University of Bologna, Bologna, Italy*
[b]*Laboratory of Biocomputing, Giorgio Prodi Center/Department of Biology, University of Bologna, Bologna, Italy*

Handling Associate Editor: Calogero Caruso

Accepted 9 November 2012

In order to determine the putative location of binding sites for transcription factor, we scanned with TESS [1] the 31 bp long sequence centered on each of the four analyzed SNPs, for both the reference and the mutated alleles. The models of transcription factors binding sites we adopted are the position weighted matrices collected in TRANSFAC v6.0 [2] and the JASPAR [3]. The quality of the matching between a matrix M and a sequence S is measured with a log-likelihood (LL) score, computed as the logarithm of the ratio between the probability to observe the sequence S given the motif model M, and the probability to observe the sequence S given a background null model. High positive scores indicate that the sequence is considered to favor the binding of the transcription factor. The value of LL-score strongly depends on the adopted matrix and it is therefore specific for each transcription factor. The LL-score significance is assessed by computing a p-value. The match strength can be also evaluated as the ratio between the actual LL-score and the maximum LL-score that can be obtained with a particular matrix. When this value is above 75%, we marked the match as strong.

Supplementary Table 1 lists the transcription factor binding sites reporting a p-value for the LL-score ≤ 0.05. For each site discovered in the reference and/or the mutated sequence, Supplementary Table 1 reports the name of the transcription factor, the location of the site (considering the SNP in the position numbered with 0), the name of the matrix model, the match quality, the LL-score, the LL-score per site (obtained as the ratio between the LL-score and the site length), and the corresponding p-value. Strong matches are also highlighted. Differences between the transcription binding sites found in reference and mutated sequences are highlighted in grey.

*Correspondence to: Elisa Porcellini, PhD, Laboratory of Immunopathology and Immunogenetics, Department of Experimental Pathology, University of Bologna, Via S. Giacomo 14, 40126 Bologna, Italy. Tel.: +39 051 2094700; Fax: +39 051 2094746; E-mail: elisa.porcellini3@unibo.it.

ISSN 1387-2877/13/$27.50

Supplementary Table 1
Analysis of transcription factor binding sites in regions surrounding the SNPs associated with Alzheimer's disease risk

rs1884082					REF: GCATTTCCCAAGCAGGGGGAGGAGTTCTCTG				MUT: GCATTTCCCAAGCAGTGGGAGGAGTTCTCTG			
Transcription factor	Begin	End	Strand	Matrix	LL-score	LL-score per site	LL-score *p*-value	Match strength	LL-score	LL-score per site	LL-score *p*-value	Match strength
Ik-1	–15	–3	–	M00086	10.38	0.80	6.3e–03		10.38	0.80	6.3e–03	
C/EBPbeta	–15	–2	–	M00109	9.77	0.70	1.1e–02		9.77	0.70	1.1e–02	
C/EBPalpha	–15	–2	+	M00116	7.49	0.54	4.0e–02		7.49	0.54	4.0e–02	
C/EBPbeta	–15	–2	–	M00117	8.43	0.60	2.0e–02		8.43	0.60	2.2e–02	
C/EBPalpha	–15	–2	–	M00190	7.78	0.56	3.3e–02		7.78	0.56	3.3e–02	
Ik-2	–14	–2	–	M00087	9.30	0.78	1.8e–02		9.30	0.78	1.8e–02	
LyF-1	–12	–3		M00141	10.15	1.13	9.3e–03		10.15	1.13	9.3e–03	
GKLF*	–7	6	+	M00286	9.33	0.67	1.1e–02		–	–	–	–
p300*	–8	5	+	M00033	–	–	–	–	7.62	0.54	3.6e–02	
F$STRE_01*	–4	3	+	M00154	9.34	1.17	2.2e–02		–	–	–	–
MZF-1*	–3	4	+	M00083	8.54	1.07	3.2e–02		–	–	–	–
Sp1	–2	10	+	M00196	12.65	0.97	1.2e–03		10.88	0.84	4.0e–03	
V$GC_01	–2	11	+	M00255	12.12	0.87	2.1e–03		10.86	0.78	4.8e–03	
P300	1	14	+	M00033	7.47	0.53	3.9e–02		7.47	0.53	3.9e–02	
rs3761740					REF: AAATGTGTGGTGGGGCCATATTAGTGGTGAC				MUT: AAATGTGTGGTGGGGACATATTAGTGGTGAC			
Transcription factor	Begin	End	Strand	Matrix	LL-score	LL-score per site	LL-score *p*-value	Match strength	LL-score	LL-score per site	LL-score *p*-value	Match strength
AML1a	–10	–5	+	M00271	10.87	1.81	1.3e–02	High	10.87	1.81	1.3e–02	High
MZF-1*	–11	1	+	M00084	–	–	–	–	10.11	0.78	7.0e–03	
MZF-1*	–7	0	+	M00083	7.70	0.96	4.7e–02		11.84	1.48	1.5e–03	High
Sp1	–9	0	+	M00008	–	–	–	–	7.68	0.77	4.8e–02	
HOXA3	1	9	+	M00395	7.02	0.78	2.3e–02	High	7.02	0.78	2.3e–02	High
CBF_(2)	4	15	–	M00185	8.28	0.75	1.9e–02		8.28	0.75	1.9e–02	
AML1a	7	12	+	M00271	7.58	1.26	5.0e–02		7.58	1.26	5.0e–02	
MZF-1	7	14	+	M00083	8.04	1.00	4.2e–02		8.04	1.00	4.2e–02	
rs1800629					REF: GTTTTGAGGGGCATGGGGACGGGGTTCAGCC				MUT: GTTTTGAGGGGCATGAGGACGGGGTTCAGCC			
Transcription factor	Begin	End	Strand	Matrix	LL-score	LL-score per site	LL-score *p*-value	Match strength	LL-score	LL-score per site	LL-score *p*-value	Match strength
MZF-1	–14	–2	+	M00084	10.22	0.79	6.4e–03		8.61	1.43	3.7e–02	
LyF-1	–13	–5	+	M00141	8.63	0.96	2.3e–02		8.63	0.96	2.3e–02	
F$STRE_01	–11	–4	+	M00154	9.37	1.17	2.2e–02		9.37	1.17	2.2e–02	
CF1*	–9	–1	+	M00111	7.30	0.81	4.6e–02		7.30	0.81	4.6e–02	
AP-2alpha*	–6	+5	–	M00189	8.28	0.69	2.6e–02		–	–	–	–
USF-1*	–7	+1	–	M00217	8.66	1.08	3.3e–02		–	–	–	–
MZF-1*	–4	+3	+	M00083	9.39	1.17	2.2e–02	High	–	–	–	–
ADR1	+4	+9	+	M00048	8.61	1.43	3.7e–02	High	8.61	1.43	3.7e–02	High

Supplementary Table 1
Continued

rs1800896					REF: AAGGCTTCTTTGGGAGGGGGAAGTAGGGATA				MUT: AAGGCTTCTTTGGGAAGGGGAAGTAGGGATA			
Transcription Factor	Begin	End	Strand	Matrix	LL-score	LL-score per site	LL-score *p*-value	Match strength	LL-score	LL-score per site	LL-score *p*-value	Match strength
C/EBPalpha	–9	+3	+	M00159	–	–	–	–	9.33	0.72	8.2e–03	
Ik-2*	–8	+3	+	M00087	–	–	–	–	10.04	0.84	9.0e–03	
LyF-1	–7	+1	+	M00141	14.07	1.56	3.5e–04	High	12.86	1.43	1.1e–03	High
MZF-1*	–6	+6	+	M00084	15.14	1.16	3.4e–05	High	11.58	0.89	2.2e–03	
MZF-1*	–2	+5	+	M00083	9.54	1.19	1.7e–02	High	8.80	1.10	2.7e–02	
F$STRE_01	–2	+5	+	M00154	6.56	0.82	4.2e–02		8.32	1.04	3.2e–02	
GKLF*	–2	+11	+	M00286	–	–	–	–	9.19	0.66	1.2e–02	
GKLF	–1	+12	+	M00286	8.18	0.58	2.3e–02		9.82	0.70	7.4e–03	

For each considered SNP, the 31 residue long sequence of the reference and the mutated allele are reported. The SNP is in the central position and it is underlined. Putative binding sites are predicted by scanning the sequences with TESS [1] against the models represented by the matrices stored in the TRANSFAC [2] and JASPAR [3] databases. Only sites matching with a $p \leq 0.05$ are reported. Begin and End positions are given with reference to the SNP position, numbered with 0. For each match we list the total log-likelihood (LL) score, the LL-score normalized to the site length (LL-score per site), and the corresponding *p*-value. If the LL-score is more than 75% of the maximum score achievable with a given matrix, the match is labeled as strong (Match strength column). Differences between the transcription factor binding sites in the reference and the mutated sequences are indicated with a star (*).

REFERENCES

[1] Schug J (2008) Using TESS to predict transcription factor binding sites in DNA sequence. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2.6.

[2] Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31**, 374-378.

[3] Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**, 276-287.