Original Article

Prediction of Regulatory Factor X1 Binding Sites in Promoters of RNA-Binding Proteins Genes in Mouse Brain

Xiao Jie Liu Ph.D[1], Zhi Hu Zhao PhD[2], Xuan Ming Liu PhD•[1]

Abstract

Background: There is an increasing interest in the role of regulatory factor X1 (RFX1) and RNA-binding proteins (RBPs) during neural development. However, there are few reports about their interaction.

Methods: We extracted RNA and performed reverse transcription polymerase chain reaction **(**RT-PCR) to identify RFX1 expression in imprinting control region (ICR) mouse tissues, analyzed RFX1 domains and motif consensus by comparing public databases on the Internet, tested the motif consensus with affinity-capture and western blotting experiments with mouse brain tissue, and predicted the binding sites of RFX1 in promoter regions of mouse RBPs genes.

Results: The expression of RFX1 was higher in embryonic brain compared to embryonic kidney, heart, and liver, and its expression level was relatively stable and higher in mouse embryonic brain than neonatal brain. RFX1 had several domains, including domain A as an activation domain, DBD as a DNA binding domain, domain B and C which played an important role in dimerization, and domain D as dimerization domain. RFX1 had three different profiles motif consensus RFX1M00281, RFX1M00280, and RFX1 (EF-C) M00626. There were 79 RFX1 binding sites at the promoters of 65 of 323 RBPs genes.

Conclusions: RFX1 as regulatory factor will have putative important regulating role in the expression of RBPs genes during embryonic development of mouse brain.

Keywords: Binding sites, regulatory factor X1 (RFX1), RNA-binding proteins (RBPs)

Cite this article as: Liu XJ, Zhao ZH, Liu XM. Prediction of regulatory factor X1 binding sites in promoters of RNA-binding proteins genes in mouse brain. *Arch Iran Med*. 2013; **16(7):** 390 – 396.

Introduction

The regulatory factor X (RFX), a member of transcription factors (TFs) gene family, also known as enhancer factor C (EF-C), was first discovered as a regulatory factor in mammals 20 years ago.[1] RFX binds to X-box motif that is a conserved cis-regulatory element with typically 14-mer DNA sequences,[1] which is primarily identified in the promoter regions of major histocompatibility complex (MHC) class II genes.[2,3] The RFX gene family contained members in human (RFX1-8) (http://www.uniprot.org/uniprot/Q6ZV50) and mice (RFX1-7) genomes[4–10] and has distinct functions in mammals.[11–17] The RFX1 played a more significant role in a wide variety of biologic processes, including transactivating the hepatitis B virus enhancer in the liver[18] and activating or suppressing expression of many genes.[19–27] Furthermore, it had been reported that knockout of RFX homologue in *Caenorhabditis elegans* led to severe sensory defects,[28] and the Drosophila RFX1 homologue was necessary for ciliated sensory neuron differentiation.[29]

In mammals, RFX1 was expressed higher in the brain than other tissues and organs.[20,30] Recently, it is reported that there are more than 10,000 RFX1 binding sites as conserved noncoding elements (CNEs) in human genome,[31] indicating a broad role for the RFX1 in regulating mammalian gene expression, although the mechanism remains largely unknown. In eukaryotic cells, RNA-binding proteins (RBPs) played an important role in gene expression, involving in regulating the form, abundance, and stability of both coding and noncoding RNA.[32] In the mammal brain, RBPs demonstrated a great quantity of salient features of RNA processing, such as transcription activity proteins localization and synthesis. In E13.5 (embryos at stages days post coitus (dpc) 13.5) mouse brain, there were 323 RBP genes expressed as demonstrated by in situ hybridization (ISH).[32] However, there is no report of the RFX1 binding sites in RBPs genes.

Authors' affiliations: [1]School of Biology, Hunan University, Hunan, China, [2]Beijing Institute of Biotechnology, Beijing, China.
•Corresponding author and reprints: Xuan Ming Liu PhD, School of Biology, Hunan University, 1 Denggao Read, Yuelu District, Changsha, Hunan, China, P. O. Box: 410082. E-mail: xml05@hnu.edu.cn
Accepted for publication: 19 May 2013

Materials and Methods

Animals

All studies were performed in accordance with the Guide for the Care and Use of Laboratory Animals, and all procedures were approved by the Animal Ethics Committee, at Institute of Biological Engineering, Academy of Military Medical Sciences. Strains and ages of animals examined in given experiments were provided in the text and figure legends at appropriate points.

Tissue collection

Adult imprinting control region (ICR) mice were fed with mouse food and autoclaved water and maintained on a 12-hour light: 12-hour dark in a 20 – 25°C environment. Timed mating was carried out with embryonic day 0.5 being set as midday of the day of discovery of a vaginal plug. Mouse embryonic tissues including brain, kidney, heart, and liver were collected immediately after the animals were sacrificed when the corresponding period (embryos at stages days post coitus 13.5 [E13.5], embryos at stages days

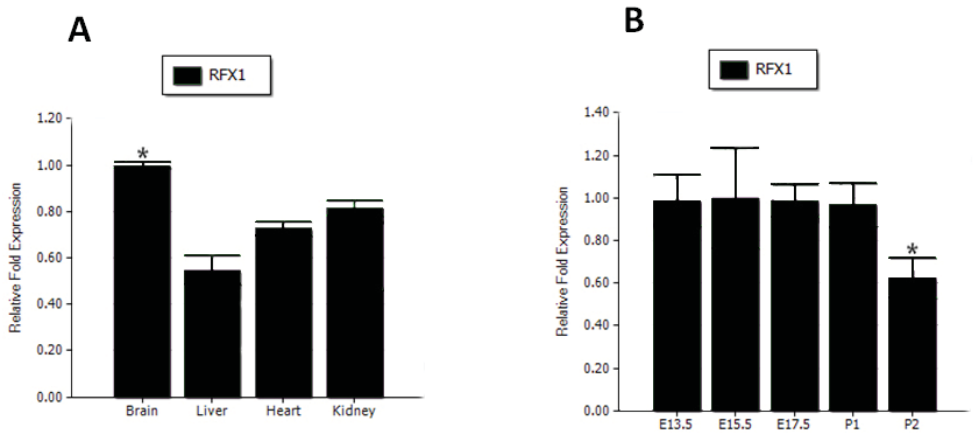

Figure 1. Expression of RFX1 **A)** in different tissues of E17.5 and **B)** in brain of mouse at different ages. GAPDH was used as an internal control for each PCR amplification. Relative expression levels in RT-PCR were obtained in each sample by normalization of absolute optical densities (ODs) of the specific target to that of the GAPDH signal.*P < 0.05.

post coitus 15.5 [E15.5], and embryos at stages days post coitus 17.5 [E17.5]; postnatal day 1 [P1] and P2) came.

RNA isolation and real-time quantitative PCR

Total RNA was extracted from embryonic brains, kidneys, hearts, and livers of E17.5 ICR mice and brains of E13.5, E15.5, E17.5, P1, and P2 ICR mice, based on our preliminary experiment and in order to reduce interference from background. The embryonic tissues were snap-frozen by using liquid nitrogen. Total RNAs were extracted with RNA extraction kit (Takara, Dalian) following treatment with DNase I. One microgram total RNA was reverse transcribed into cDNA with random hexamer primers using the RT-PCR kit (Takara, Dalian). RT-PCR was performed on qPCR system (Bio-Rad, USA) with SYBR Green (Takara, Dalian). GAPDH was set as internal control. Data were analyzed using $2^{-\Delta\Delta Ct}$ method. All experiments were repeated for at least three times. Statistical significance was evaluated using a *t*- test. *P*-values of < 0.05 were considered statistically significant.

The RFX1 primers were forward, 5-gagcacagagtggcccaggc-3, and reverse, 5-tgtcagggtccaagggattcag-3, TM 60.9°C. The primers for GAPDH were forward, 5-ggtgaaggtcggtgtgaacg-3, and reverse, 5-ctcgctcctggaagatggtg-3, TM 55-60.9°C. The PCR was performed in Bio-RadIQ5 with 40 cycles consisting of 94°C 10 sec, 94°C 10 sec, and 61.9°C 30 sec. All experiments were repeated three times using three pregnant mice, which usually breed eight to14 embryos.

Detection of RFX1 (EF-C) in E13.5 ICR mouse genome

A biotinylated 210bp mouse fibroblast growth factor 1B promoter (FGF1B) DNA[33] probe containing RFX1(EF-C) binding site[34] was incubated with nucleoprotein of ICR mouse brain tissue extracted according to the manual of nucleoprotein extraction kit (CW-Bio Beijing, China) and then captured with streptavidin beads (Invitrogen, USA). The captured protein was electrophoresed, blotted, and probed with antibody against RFX1 (Abcam, USA).

Scanning of RFX1 motifs in RBP genes

The regions spanning to 1 kb upstream of transcription start sites (TSSs) of the 323 brain expressing RBPs in E13.5 mouse[32] were identified for RFX1 recognizing motif analysis. In order to analyze RFX1 binding sites and its features in mouse genome, we preanalyzed the binding sites of RFX1 at promoters of RBPs genes excepting distal sites and other sophisticated situations.

RFX1 motif prediction and data analysis

The 1kb upstream of TSSs in 323 genes[32] from database of transcriptional start sites (DBTSS) (http://dbtss.hgc.jp) was annotated to be promoter region, and RFX1 binding sites in these promotes were searched in the database of TRANSFAC by Weight Matrix library (http://transfac.gbf-braunschweig.de), according to the literature.[35,36] A profile of the degree of conservation at each position of the matrix (e.g.,Ci-vector) was provided, ranging from zero (an equal distribution of all four nucleotides) to 100 (conservation of one nucleotide to the exclusion of all others). A "core" sequence of the matrix was also provided and defined as the (usually 4) highest conserved, consecutive positions of the matrix. The histogram was created by MATLAB7.0.

Results

Validation of RFX1 mRNA expression by real-time quantitative PCR

As shown in Figure1 B, the expression of RFX1 was relatively stable and similar in mouse embryonic brain from E13.5 to E17.5. The expression of RFX1 in brain at embryonic stage was similar to at P1 and decreased dramatically at P2 (P < 0.05). In addition, the RFX1 expression was higher in embryonic brain at E17.5 than liver of the same stage (Figure 1A) (P < 0.05).

Sequence analysis of the RFX1 domains

As shown in Figure 2, the RFX1 had five domains including A domain (residues 275 – 302), DBD domain (423 – 498), B domain (588 – 623), C domain (671 – 710), as well as D domain (727 – 893), which were analyzed by Vector_NTI10 (Invitrogen).[37,38] Of these, the DBD domain was composed of three structures, including three α- helices (H), three β-strands (S), and three connecting loops (L), in the order of H1-S1-H2-L1-H3-L2-S2-W1-S3. S2 and S3 formed the third loop, namely W1 of the winged-helix motif. H1, S1, H2, H3, L2, and S3 formed the hydrophobic core and conducted the direct or water-mediated DNA contacts (Figure 2).[38]

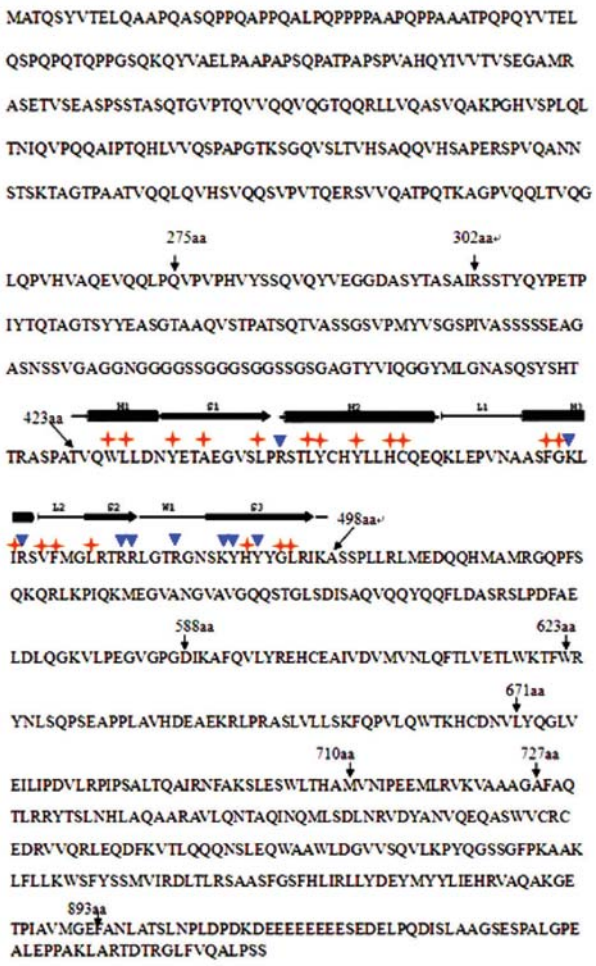

Figure 2. The sequence analysis of RFX1. The residues of 275 – 302aa is A domain. The residues of 423 – 498aa is DBD domain. The nine residues that make direct or water-mediated DNA contacts are indicated with arrow heads, residues marked with asterisks signs comprise the hydrophobic core. The sequence of 588 – 623aa is B domain. The sequence of 671 – 710aa is C domain. The sequence of 727 – 893aa is D domain.

Motif consensus of RFX1 and detection of RFX1 (EF-C)

As shown in Figure 3 (A, B, C), there were three different consensus sequences, named as RFX1M00281, RFX1M00280, and RFX1 (EF-C) M00626 (http://motifmap.ics.uci.edu/), (Table 1), respectively. The RFX1M00281 and RFX1M00280 motifs were nearly the same except an additional N at the ninth position in motif RFX1M00281 consensus (Table 1) and an "A" insertion in Logo (Figure 3A and B), resulting in very close length of RFX1M00281 and RFX1M00280 of 18 and 17, respectively. The RFX1M00626 was very different from the other two motifs in both sequence and length (Figure 3C and Table 1). Meanwhile, affinity-capture and western blotting were performed to confirm the binding capacity of RFX1 (EF-C) M00626. As shown in Figure 3D, RFX1 (EF-C) M00626 was detected in mouse genome with addition of Biotin-FGF1B and was abolished by addition of FGF1B and no DNA (Figure 3D). The fourth strip was the positive control.

Scanning RFX1 motifs in RBP genes

In the regulatory motifs in the 1kb promoter regions of 323 RBPs genes expressed in E13.5 mouse brain, there were 65 genes and 79 binding sites, and 10 of 65 genes had two motifs, three had three motifs (Table 2). Given that the core RFX1 motifs were 14- to 18-bp long, and the resolution is computed with respect to the center of the motif, the precision of identified binding sites was unprecedented.[35,36] The detailed information of the 79 binding sites, including LocusLink ID, Updated name, Position, Core match, Matrix match, Sequence, and Factor name, was listed in Table 2. From Table 2, it indicated that RFX1 (EF-C)M00626 attributed 72 of 79 binding sites. The positions of 79 binding sites at the 1kb promoter in RBPs genes were shown in Figure 4. The result indicated that these binding sites distributed at both sides of TSS (Figure 4), and in the area close to 400bp distance of TSSs there were the most binding sites. About 15% of all sites were between 0 bp and 200 bp (from -200 to 0 and from 0 to 200), 27% of sites were between 200 bp and 400 bp (from -200 to -400 and from 200 to 400), 19% of sites were between 400 bp and 600 bp, 20% of sites were between 600 bp and 800 bp, and 19% of sites were between 800 bp and 1000 bp.

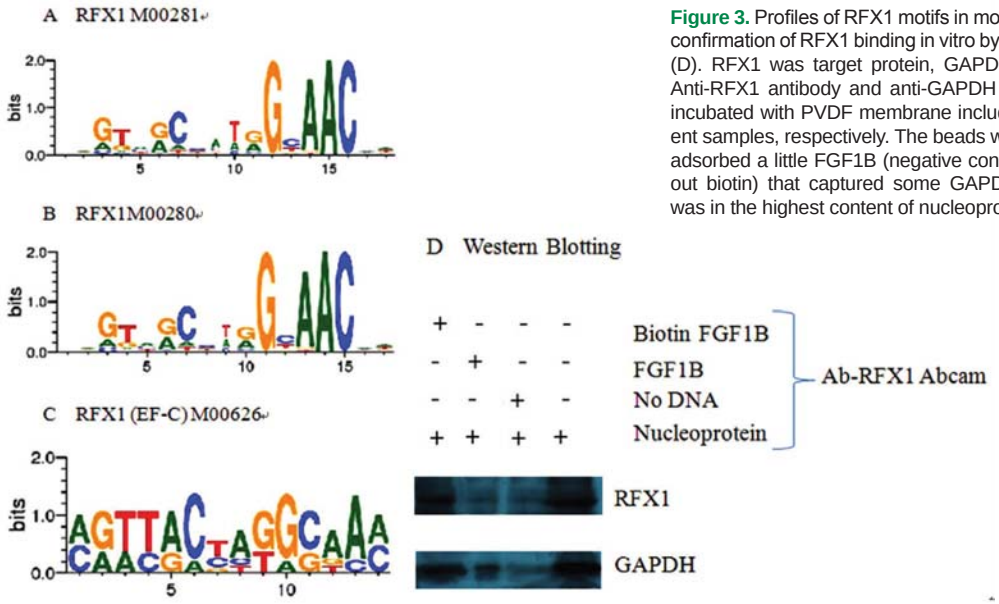

Figure 3. Profiles of RFX1 motifs in mouse (A– C) and confirmation of RFX1 binding in vitro by affinity capture (D). RFX1 was target protein, GAPDH was control. Anti-RFX1 antibody and anti-GAPDH antibody were incubated with PVDF membrane including four different samples, respectively. The beads with streptavidin adsorbed a little FGF1B (negative control probe without biotin) that captured some GAPDH protein that was in the highest content of nucleoprotein.

Table 1. There are three forms of RFX1 motif in mouse genome (http://motifmap.ics.uci.edu/).

Motif ID	TF Name	Consensus	Length
M00281	RFX1	NNRKHRCNNWRGYAACNN	18
M00280	RFX1	NNRKHRC_NWRGYAACNN	17
M00626	RFX1(EF-C)	MRWYRCYWKGSWAM	14

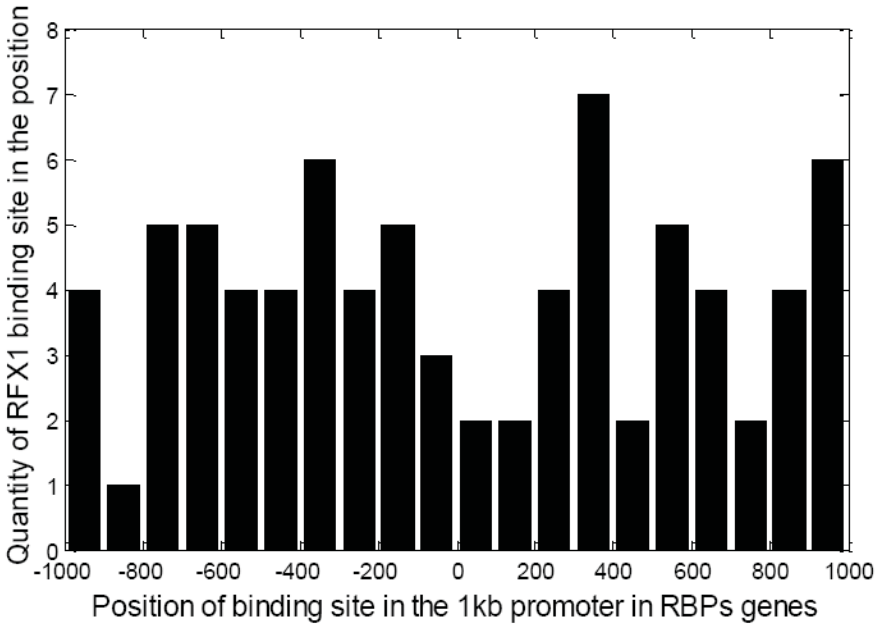

Figure 4. A histogram summarized the frequency distribution of the RFX1 binding sites in the 1kb promoters of RBPs genes. About 15% of all sites were between 0 bp and 200 bp, 27% within from 200 bp to 400 bp, 19% within from 400 bp to 600 bp, 20% within from 600 bp to 800 bp, and 19% within from 800 bp to 1000 bp.

Discussion

The present study indicated that the RFX1 was highly expressed in embryonic brain and obviously decreased in neonatal brain, and had five domains, three different motifs, and 79 binding sites in RBPs. Moreover, RFX1 expression was detected in the same stage as the expression of RBPs identified by ISH.[32] It illustrated that RFX1 had great quantity of salient features in RNA processing. These results suggest that RFX1 may play an important modulating role during embryonic development of brain in mice.

RFX1 was a transcriptional factor and had many modulation functions.[18–27] In human and mouse, RFX1 was highly expressed in the brain, neurons, microglia of hippocampus, olfactory bulb, and cerebral cortex.[20,30] It was found that knockout of RFX homologue in *Caenorhabditis elegans* led to severely sensory defects.[28] Drosophila RFX1 homologue was necessary for ciliated sensory neuron differentiation.[29] Furthermore, RFX1 was selectively enriched at neural precursor enhancers[39] and knockout of the RFX1 gene in mice led to early embryonic death.[30] These results suggest that RFX1 was important for the normal development of central nervous system. However, the detailed role of RFX1 in development of central nervous system remains to be further studied.

In the present study, we confirmed that there were five domains and three motifs in mice RFX1 protein. The domains contained domain A as an activation domain, DBD as a DNA binding domain (DBD domain), domain B and C which played an important role in dimerization and thus were called the extended dimerization domain,[40] as well as the domain D as dimerization domain.[4] The DBD domain was highly conserved throughout the eukaryotic kingdom and implicated in a diverse range of biologic systems[4] and proved to be a nuclear factor, with winged-helix conformation belonged to subfamily of helix-turn-helix (HTH) transcription factors.[41] Furthermore, the DBD domain was composed of three structures and formed the hydrophobic core of RFX1 protein. The motifs included RFX1M00281, RFX1M00280, and RFX1 (EF-C) M00626.[42] The nucleotide residues (G, C, G, A, A, C) in these three motifs were not noticeable changes. This suggests that they might be critical for DNA binding sites, and have significant information content contribute to stable DNA binding sites. Moreover, the nucleotide positions (1, 2, 5, 7, 8) of RFX1 (EF-C) M00626 were less changed compared to the other two, suggesting RFX1 (EF-C) would be more staple profile and have higher density in genome namely higher protein occupancy on the DNA.

In addition, RBPs also played an essential role at post-transcriptional gene regulation in a wide variety of events that occurred during mouse development, such regulating the form, abundance, and stability of both coding and noncoding RNA.[32] In the present study, we found that 65 genes of the 323 RBPs had 79 RFX1 binding sites at the same stage as the expression of RBPs, and 91% of these sites were RFX1 (EF-C). Recently, it was respectively reported that RFX1 and RBPs were fundamental for occurrence and development of nervous system.[29,32] These results suggest that there may be some important interaction of RBPs and RFX1 in the neuronal development.

Table 2. RFX1 binding site in the 1kb promoters in RBPs genes (there were 79 RFX1 binding sites in 323 RBPs genes in E13.5,[32] mouse brain. The value of Core match and matrix close to 1 was valid. RFX1 (EF-C) is staple profile)

Locus link ID	Updated names	Position (strand)	Core match	Matrix match	Sequence	Factor name	Seiral number
56195	Ptbp2	966(+)	1.000	0.817	cGTTACgcggctcc	RFX1(EF-C)	1
52897	D11Bwg0517e	117 (-)	0.910	0.863	tttcccgtGTAATg	RFX1(EF-C)	2
94230	Cpsf1	339 (+)	0.910	0.832	gGTCACcaggtaaa	RFX1(EF-C)	3
15384	Hnrpa/b	90 (+)	0.910	0.849	aGTCACatggcaat	RFX1(EF-C)	4
56403	Nsap1	611 (-)	1.000	0.981	caGTTGCcactgcaacag	RFX1	5
17975	Ncl	554 (+)	0.982	0.976	ctgtcgccagGGAACgc	RFX1	6
---	---	555 (+)	0.820	0.832	tGTCGCcagggaac	RFX1(EF-F-C)	7
---	---	556 (-)	0.820	0.844	gtcgccagGGAACg	RFX1(EF-C)	8
241989	Pabpc4l	901 (-)	0.982	0.979	agGTTTCctggcaacgc	RFX1	9
---	---	902 (+)	0.820	0.831	gGTTTCctggcaac	RFX1(EF-C)	10
---	---	903 (-)	0.910	0.882	gtttcctgGCAACg	RFX1(EF-C)	11
17690	Msi1h	124 (-)	1.000	0.804	ttcccttaGTAACa	RFX1(EF-C)	12
56190	Rnpc1	362 (+)	0.820	0.892	aGATGCtatgcaaa	RFX1(EF-C)	13
72459	Htatsf1	951 (+)	1.000	0.810	cGTTACagggcaga	RFX1(EF-C)	14
74111	Rbm19	297 (-)	0.910	0.800	ctggcattGTAATg	RFX1(EF-C)	15
67996	Sfrs6	374 (-)	0.820	0.805	ctttcctaGTAAAg	RFX1(EF-C)	16
20637	Snrp70	602 (+)	0.982	0.974	cggttgccggGAAACaa	RFX1	17
---	---	604 (-)	0.820	0.831	gttgccggGAAACa	RFX1(EF-C)	18
67579	Cpeb4	399 (+)	0.820	0.809	aGTAACcaggttaa	RFX1(EF-C)	19
---	---	599 (-)	0.910	0.809	tgagccttGTGACt	RFX1(EF-C)	20
233833	Tnrc6	938 (-)	0.910	0.809	tttgtcatGTGACt	RFX1(EF-C)	21
66369	2310016K04Rik	388 (-)	0.982	0.970	aaGTTTCcttggcaactc	RFX1	22
---	---	391 (-)	0.910	0.846	tttccttgGCAACt	RFX1(EF-C)	23
16201	Ilf3	811 (+)	0.820	0.864	cAATACaagggaaa	RFX1(EF-C)	24
67471	Gpatc1	466 (-)	1.000	0.855	ttttcttaGTAACt	RFX1(EF-C)	25
80912	Pum1	823 (+)	1.000	0.800	aGTTACaaggtcat	RFX1(EF-C)	26
76850	Eif2c4	192 (+)	0.820	0.837	cAATACtagacaac	RFX1(EF-C)	27
12785	Cnbp	911 (+)	0.910	0.855	aGTTGCtaaggaaa	RFX1(EF-C)	28
56463	Al033314	311 (+)	0.820	0.814	aGATGCagtgcaac	RFX1(EF-C)	29
11640	Akap1	682 (+)	0.910	0.801	tGTTGCtagtcaac	RFX1(EF-C)	30
19656	Rbmxrt	34 (-)	0.820	0.840	gtagcctaGTGATa	RFX1(EF-C)	31
74213	1700009P03Rik	514 (+)	0.820	0.802	gGCTACttggctac	RFX1(EF-C)	32
231413	Grsf1	218 (+)	0.992	0.971	tagtggcaaaGTAACat	RFX1	33
---	---	220 (-)	1.000	0.826	gtggcaaaGTAACa	RFX1(EF-C)	34
66094	Lsm7	740 (-)	0.820	0.846	tttacctgGTAAAg	RFX1(EF-C)	35
107701	Sf3b4	337 (+)	0.910	0.881	cATTACcttgcaaa	RFX1(EF-C)	36
13680	Ddx19	409 (-)	0.820	0.809	tttccaggGTGATt	RFX1(EF-C)	37
68278	Ddx39	693 (+)	0.820	0.819	cCTTACtacggaaa	RFX1(EF-C)	38
72640	Gm411	335 (-)	0.820	0.806	tatgcctgGTAAGt	RFX1(EF-C)	39
76167	6330548G22Rik	89 (-)	0.910	0.864	ttttccttGCAACg	RFX1(EF-C)	40
66599	Rad52b	785 (-)	0.910	0.823	gtgtcctcGTGACg	RFX1(EF-C)	41
12696	Cirbp	850 (+)	0.820	0.809	cGTGACccggaaac	RFX1(EF-C)	42
67920	2600016B03Rik	156 (-)	0.910	0.840	ttaccccaGTATCt	RFX1(EF-C)	43
71787	1110007F05Rik	537 (-)	1.000	0.818	ttttctgaGTAACt	RFX1(EF-C)	44
215615	Rnpep	401 (-)	0.910	0.809	ttctcatgGTGACt	RFX1(EF-C)	45
269061	5730453I16Rik	958 (-)	1.000	0.827	gttccggtGTAACg	RFX1(EF-C)	46
66877	Crnkl1	516 (-)	0.820	0.882	tttccctgGTAAAg	RFX1(EF-C)	47
75062	Sf3a3	447 (+)	0.820	0.809	aGACACctggtaac	RFX1(EF-C)	48

101214	G430041M01Rik	193 (-)	0.820	0.827	tttcccgtGTCACt	RFX1(EF-C)	49
---	---	363 (+)	1.000	0.858	cGTTACgtgggaag	RFX1(EF-C)	50
98758	Hnrpf	426 (+)	1.000	0.899	cGTTACactgcaaa	RFX1(EF-C)	51
110532	Adarb1	923 (+)	0.820	0.805	aGATGCtaggacaa	RFX1(EF-C)	52
99890	Hrmtl16	946 (+)	0.910	0.846	cGTCACgaggccag	RFX1(EF-C)	53
22668	cw17	652 (-)	0.820	0.831	gtggccagGTAAGg	RFX1(EF-C)	54
69865	1810073H04Rik	797 (+)	0.820	0.846	cGTTCCcatggaaa	RFX1(EF-C)	55
---	---	799 (-)	0.820	0.809	ttcccatgGAAACg	RFX1(EF-C)	56
23879	Fxr2	572 (+)	0.820	0.827	aAATACattgcaaa	RFX1(EF-C)	57
18459	Pabpc2	768 (+)	0.910	0.826	aGTTGCtgggctag	RFX1(EF-C)	58
16589	Kist	568 (+)	0.910	0.864	aGTTGCtaggccag	RFX1(EF-C)	59
66704	4921506I22Rik	405 (-)	0.820	0.818	ctggcctgGGAACt	RFX1(EF-C)	60
381813	LOC381813	242 (+)	1.000	0.883	aGTTACaaggcaca	RFX1(EF-C)	61
12828	Col4a3	891 (+)	0.910	0.817	cGTCACccggccag	RFX1(EF-C)	62
15547	Htf9c	96 (+)	0.820	0.827	cTTTACattgcaac	RFX1(EF-C)	63
---	---	97 (-)	0.910	0.828	tttacattGCAACg	RFX1(EF-C)	64
70044	Rbm21	629 (-)	0.910	0.816	ttcaccgaGCAACt	RFX1(EF-C)	65
93871	Wdr9	143 (+)	0.910	0.814	gGTCACtgggccaa	RFX1(EF-C)	66
---	---	749 (-)	0.910	0.832	atttcctgGTAATg	RFX1(EF-C)	67
102857	Slc6a8	243 (+)	1.000	0.878	gGTTACtatggaaa	RFX1(EF-C)	68
213236	Dnd1	627 (-)	0.910	0.832	ttgtcctgGCAACg	RFX1(EF-C)	69
218543	Sfrs12	681 (+)	0.820	0.868	aGTTCCcaggcaag	RFX1(EF-C)	70
108014	Sfrs9	242 (+)	0.820	0.828	aGTGACtgggccac	RFX1(EF-C)	71
108857	Ankhd1	335 (+)	0.820	0.810	aGTTTCggggccaa	RFX1(EF-C)	72
---	---	358 (-)	0.910	0.819	ttttcttaGTAATt	RFX1(EF-C)	73
27632	Rdbp	223 (-)	1.000	0.978	caGTTGCctagtgacat	RFX1	74
---	---	225 (-)	0.910	0.914	gttgcctaGTGACa	RFX1(EF-C)	75
77411	9530027K23Rik	737 (-)	0.820	0.816	gtcgcaaaGCATCg	RFX1(EF-C)	76
56258	Hnrph2	398 (-)	0.820	0.800	ttggccttGTGATa	RFX1(EF-C)	77
170676	Peg10	121 (-)	1.000	0.876	tttgcaaaGTAACa	RFX1(EF-C)	78
225027	Sfrs7	845 (-)	0.910	0.818	ttggcatgGCAACc	RFX1(EF-C)	79

Acknowledgments

We thank Prof. Zhao for his critical review and editing of the manuscript. This work was finished in Beijing Institute of Biotechnology, Beijing, China and supported by the National Natural Science Foundation of China (30871374, 31071119, 31030026) and the National Basic Research Program of China (2009CB825602).

References

1. Reith W, Satola S, Sanchez CH, Amaldi I, Lisowska-Grospierre B, Griscelli C, et al. Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. *Cell.* 1988; **53:** 897 – 906.
2. Dorn A, Durand B, Marfing C, Le MM, Benoist C, Mathis D. Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins. *Proc Natl Acad Sci U S A.* 1987; **84:** 6249 – 6253.
3. Sherman PA, Basta PV, Ting JP. Upstream DNA sequences required for tissue-specific expression of the HLA-DR alpha gene. *Proc Natl Acad Sci U S A.* 1987; **84:** 4254 – 4258.
4. Emery P, Durand B, Mach B, Reith W. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res.* 1996; **24:** 803 – 807.
5. Reith W, Barras E, Satola S, Kobr M, Reinhart D, Sanchez CH, et al. Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. *Proc Natl Acad Sci U S A.* 1989; **86:** 4200 – 4204.
6. Reith W, Herrero-Sanchez C, Kobr M, Silacci P, Berte C, Barras E, et al. MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. *Genes Dev.* 1990; **4:** 1528 – 1540.
7. Reith W, Ucla C, Barras E, Gaud A, Durand B, Herrero-Sanchez C, et al. RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. *Mol Cell Biol.* 1994; **14:** 1230 – 1244.
8. Steimle V, Durand B, Barras E, Zufferey M, Hadam MR, Mach B, et al. A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). *Genes Dev.* 1995; **9:** 1021 – 1032.
9. Aftab S, Semenec L, Chu JS, Chen N. Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol Biol.* 2008; **8:** 226.
10. Wu SY, McLeod M. The sak1+ gene of Schizosaccharomyces pombe encodes an RFX family DNA-binding protein that positively regulates cyclic AMP-dependent protein kinase-mediated exit from the mitotic cell cycle. *Mol Cell Biol.* 1995; **15:** 1479 – 1488.
11. Horvath GC, Kistler MK, Kistler WS. RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. *BMC Dev Biol.* 2009; **9:** 63.
12. El ZL, it-Lounis A, Morle L, Thomas J, Chhin B, Spassky N, et al. RFX3 governs growth and beating efficiency of motile cilia in mouse

and controls the expression of genes involved in human ciliopathies. *J Cell Sci.* 2009; **122:** 3180 – 3189.
13. Bonnafe E, Touka M, AitLounis A, Baas D, Barras E, Ucla C, et al. The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol Cell Biol.* 2004; **24:** 4417 – 4427.
14. it-Lounis A, Baas D, Barras E, Benadiba C, Charollais A, Nlend Nlend R, et al. Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas. *Diabetes.* 2007; **56:** 950 – 959.
15. Ashique AM, Choe Y, Karlen M, May SR, Phamluong K, Solloway MJ, et al. The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci Signal.* 2009; **2:** ra70.
16. Clausen BE, Waldburger JM, Schwenk F, Barras E, Mach B, Rajewsky K, et al. Residual MHC class II expression on mature dendritic cells and activated B cells in RFX5-deficient mice. *Immunity*. 1998; **8:** 143 – 155.
17. Smith SB, Qu HQ, Taleb N, Kishimoto NY, Scheel DW, Lu Y, et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature.* 2010; **463:** 775 – 780.
18. Siegrist CA, Durand B, Emery P, David E, Hearing P, Mach B, et al. RFX1 is identical to enhancer factor C and functions as a transactivator of the hepatitis B virus enhancer. *Mol Cell Biol.* 1993; **13:** 6375 – 6384.
19. Iwama A, Pan J, Zhang P, Reith W, Mach B, Tenen DG, et al. Dimeric RFX proteins contribute to the activity and lineage specificity of the interleukin-5 receptor alpha promoter through activation and repression domains. *Mol Cell Biol.* 1999; **19:** 3940 – 3950.
20. Ma K, Zheng S, Zuo Z. The transcription factor regulatory factor X1 increases the expression of neuronal glutamate transporter type 3. *J Biol Chem.* 2006; **281:** 21250 – 21255.
21. Zajac-Kaye M, Ben-Baruch N, Kastanos E, Kaye FJ, Allegra C. Induction of Myc-intron-binding polypeptides MIBP1 and RFX1 during retinoic acid-mediated differentiation of haemopoietic cells. *Biochem J.* 2000; **345:** 535 – 541.
22. Sengupta P, Xu Y, Wang L, Widom R, Smith BD. Collagen alpha1(I) gene (COL1A1) is repressed by RFX family. *J Biol Chem.* 2005; **280:** 21004 – 21014.
23. Xu Y, Sengupta PK, Seto E, Smith BD. Regulatory factor for X-box family proteins differentially interact with histone deacetylases to repress collagen alpha2(I) gene (COL1A2) expression. *J Biol Chem.* 2006; **281:** 9260 – 9270.
24. Liu M, Lee BH, Mathews MB. Involvement of RFX1 protein in the regulation of the human proliferating cell nuclear antigen promoter. *J Biol Chem.* 1999; **274:** 15433 – 15439.
25. Nakayama A, Murakami H, Maeyama N, Yamashiro N, Sakakibara A, Mori N, et al. Role for RFX transcription factors in nonneuronal cell-specific inactivation of the microtubule-associated protein MAP1A promoter. *J Biol Chem.* 2003; **278:** 233 – 240.
26. Wang KR, Nemoto T, Yokota Y. RFX1 mediates the serum-induced immediate early response of Id2 gene expression. *J Biol Chem.* 2007; **282:** 26167 – 26177.
27. Zhao M, Sun Y, Gao F, Wu X, Tang J, Yin H, et al. Epigenetics and SLE: RFX1 downregulation causes CD11a and CD70 overexpression by altering epigenetic modifications in lupus CD4+ T cells. *J Autoimmun.* 2010; **35:** 58 – 69.
28. Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol Cell.* 2000; **5:** 411 – 421.
29. Dubruille R, Laurencon A, Vandaele C, Shishido E, Coulon-Bublex M, Swoboda P, et al. Drosophila regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development.* 2002; **129:** 5487 – 5498.
30. Feng C, Li J, Zuo Z. Expression of the transcription factor regulatory factor X1 in the mouse brain. *Folia Histochem Cytobiol.* 2011; **49:** 344 – 351.
31. Xie X, Rigor P, Baldi P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics.* 2009; **25:** 167 – 174.
32. McKee AE, Minet E, Stern C, Riahi S, Stiles CD, Silver PA. A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev Biol.* 2005; **5:** 14.
33. Alam KY, Frostholm A, Hackshaw KV, Evans JE, Rotter A, Chiu IM. Characterization of the 1B promoter of fibroblast growth factor 1 and its expression in the adult and developing mouse brain. *J Biol Chem.* 1996; **271:** 30263 – 30271.
34. Hsu YC, Liao WC, Kao CY, Chiu IM. Regulation of FGF1 gene promoter through transcription factor RFX1. *J Biol Chem.* 2010; **285:** 13885 – 13895.
35. Fogel GB, Weekes DG, Varga G, Dow ER, Craven AM, Harlow HB, et al. A statistical analysis of the TRANSFAC database. *Biosystems.* 2005; **81:** 137 – 154.
36. Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* 1996; **24:** 238 – 241.
37. Emery P, Durand B, Reith W. RFX protein, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res*. 1996; **24 (5):** 803 – 807.
38. Gajiwala KS, Chen H, Cornille F. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature*. 2000; **403 (6772):** 916 – 921.
39. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 2010; **107:** 21931 – 21936.
40. Katan-Khaykovich Y, Shaul Y. RFX1, a single DNA-binding protein with a split dimerization domain, generates alternative complexes. *J Biol Chem.* 1998; **273:** 24504 – 24512.
41. Lai E, Clark KL, Burley SK, Darnell JE. Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proc Natl Acad Sci U S A.* 1993; **90:** 10421 – 10423.
42. Xie X, Mikelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A.* 2007; **104:** 7145 – 7150.