Available online at www.sciencedirect.com
ScienceDirect
Journal of Genetics and Genomics 38 (2011) 157−163
JOURNAL OF GENETICS AND GENOMICS
www.jgenetgenomics.org

Human transcription factor genes involved in neuronal development tend to have high GC content and CpG elements in the proximal promoter region

Yue-Sheng Long*, Jia-Ming Qin, Tao Su, Qi-Hua Zhao, Yong-Hong Yi, Wei-Ping Liao

Key Laboratory of Neurogenetics and Channelopathies of Guangdong Province and the Ministry of Education of China, Institute of Neuroscience and The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China

Received 14 September 2010; revised 1 March 2011; accepted 3 March 2011

Abstract

Transcription factors (TFs) play critical roles in the development of the nervous system, but the transcriptional regulatory mechanisms of these genes are poorly understood. Here we analyzed 5-kb of the 5′ flanking genomic DNA sequences of 41 TF genes involved in neuronal development. The results showed that the TF genes tend to have higher GC contents in the proximal region and most of the TF genes have at least one proximal GC-rich (GC content > 60%) promoter with a CpG island. The promoter distribution analysis showed that the GC-poor promoters were sporadically distributed within the 5-kb flanking genomic sequence (FGS); however, more than half (37 of 70) of the GC-rich promoters were located in the proximal region between nucleotides −1 and −500. Luciferase assays showed that partial GC-rich promoters increased gene expression in SH-SY5Y cells and that CpG methylation repressed the promoter activity. This study suggests a potential general mechanism for regulation of TF expression.

Keywords: Transcription factor; Promoter; GC content; CpG island; DNA methylation

1. Introduction

Guanine and cytosine (GC) are unevenly distributed in the human genome. High GC content is associated with gene density and composition of repeat sequences (Lander et al., 2001). Gene density is greater in regions of high GC content than in regions of low GC content (Venter et al., 2001). Variations in GC content can determine gene expression in a synergistic interplay with transcription factor-binding sites (Vinogradov, 2005). In the GC-rich regions, there are some transcription factor-binding sites, such as the CpG element (usually for Sp1 binding), which are believed to preferentially appear at the transcriptional start sites of genes to activate gene expression (Holler et al., 1988; Venter et al., 2001; Vinogradov, 2005).

Transcription factors (TFs) have been demonstrated to play critical roles in neural development (Groves et al., 1995; Ross et al., 2003), mediating various brain functions (Alberini, 2009) and the pathogenesis of neurological disorders (Chahrour et al., 2008; Lai et al., 2001). In the nervous system, TFs determine different gene expression patterns that contribute to distinct neuronal and glial cell subtypes. On the other hand, the temporal and spatial expression of TF genes is fine-tuned by other transcription factors. However, the transcriptional regulatory mechanisms of TF genes are still poorly understood.

Here, we measured the GC contents of TF gene 5′ flanking genomic sequences (FGSs) and found that GC-rich regions are located in the proximal FGSs of most TF genes. The predicted promoters in the 5-kb FGSs tended to have high GC contents and multiple CpG islands. Luciferase assays showed that these proximal GC-rich promoters drove gene expression *in vitro* and that CpG methylation repressed promoter activity, which suggests a potential mechanism for the transcriptional regulation of TF genes.

* Corresponding author. Tel: +86 20 3415 2640, fax: +86 20 3415 3378.
E-mail address: longys1972@yahoo.com.cn (Y.-S. Long).

doi:10.1016/j.jgg.2011.03.003

2. Materials and methods

2.1. Sequence data and GC content calculation

We selected 41 human TF genes that have been demonstrated experimentally to play important roles in neural development and in mediating various brain functions (Supplementary Table 1). The 5′ FGS analyzed in this study, includes the 5′ untranslated region (UTR) and the 5-kb genomic sequence upstream of the 5′ UTR. The 5′ UTRs were defined according to data published on the NCBI website (http://www.ncbi.nlm.nih.gov/). NCBI accession numbers for these genes are listed in Supplementary Table 1. The 5-kb genomic sequences upstream of the 5′ UTR were downloaded from the Ensembl website (http://www.ensembl.org/index.html). If there were untranslated exons on the 5′ UTR, we defined the 5′ FGS as the region ranging from the first 5′ UTR exon up to the 5-kb genomic sequence upstream of the first 5′ UTR exon. To analyze the GC content of the 5′ FGS, we divided the 5-kb genomic sequence into 25 segments of 200 bp each, and then calculated the GC content of each segment with DNA Baser (http://www.dnabaser.com).

2.2. Promoter prediction

The promoter regions and transcription start sites (TSSs) of the 5-kb genomic fragments were predicted using the TSSG program (Baylor College of Medicine, Houston) (Solovyev and Salamov, 1997). The predicted promoter sequences are listed in the Supplementary Material.

2.3. Luciferase constructs

Human genomic DNA was isolated and purified from peripheral blood with QuickGene DNA Whole Blood Kit L (Fuji Photo Film Co., LTD, Japan) using the Automatic Nucleic Acid Isolation System (QuikGene-610L, FUJIFILM). The potential TF promoter sequences (from −200 to +150) were amplified from the genomic DNA with primers listed in Table 1 (P1 and P2 for *ATF2*-P1; P3 and P4 for *FOXG1*-P1; P5 and P6 for *EGR2*-P1; P7 and P8 for *PHOX2A*-P1; P9 and P10 for *GATA3*-P1). PCR amplification was initiated by denaturation at 94 °C for 4 min followed by 30 cycles of 30 s at 94 °C, 30 s at 65 °C, and 3 min at 72 °C, with a final extension step of 5 min at 72 °C. The PCR products were visualized by ethidium bromide staining of agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen, USA). The purified products were cloned into the pGL4.10 vector (Promega, USA) at the *Xho* I and *Hin*d III sites to produce a series of luciferase recombinant pGL4-X constructs in which the promoter sequences were located directly upstream of the luciferase translation start site. In the luciferase experiments, the vector pGL4.10 lacking a promoter was used as a negative control.

Table 1
Primers used in this study.

Name	Sequence (5′→3′)[a]
P1	GCCTCGAGCTCCCGGGGAGCCAAGG
P2	CGGAAGCTTGGGTAGAGGGGACGGGCATC
P3	TGTCTCGAGGCGGGCTTCTCGCGGTCC
P4	GACAAGCTTCGCCCGGGATAGCCGCCAG
P5	GTCCTCGAGAACCGGAATTCCTCCCCG
P6	TCAAAGCTTCCTCGCTCAGTTAGACGGAAAGTG
P7	CGGCTCGAGGGGGAGAGCTCGCGTGTAG
P8	CCCAAGCTTCCGGAACGCCTCGCCAG
P9	CACCTCGAGGCGCTCTCCCAAACACCCTG
P10	TGGAAGCTTGGGGACAGGATCCCCGGC

[a] Restriction sites are underlined.

2.4. CpG methylation

To generate the promoters with methylated CpG, bacterial CpG methylase *M.SssI* (New England Biolabs, USA) was used according to the manufacturer's instruction. Evaluation of the completeness of the reaction was carried out as described (Robertson et al., 1995). In brief, the reaction products were digested with an appropriate methylation-sensitive restriction enzyme (*Hpa* II or *Hha* I, New England Biolabs). Complete protection from digestion indicated that the methylation reaction went to completion. After the methylation reaction, the recombinant constructs were purified using the QIAquick Gel Extraction Kit and were then used for transfections.

2.5. Cell culture and transfection procedure

Human neuroblastoma SH-SY5Y cells were grown in 45% minimum essential medium alpha medium (MEM, GIBCO, USA) with 45% F12 (GIBCO), 10% fetal bovine serum (GIBCO), and 10 μg/mL streptomycin. The cells were incubated at 37 °C in a 5% CO_2 atmosphere. The cells were plated at 4×10^5 cells/well in 24-well plates 24 h before transfection. The medium was changed to MEM without fetal bovine serum before transfection. In each well, cells were co-transfected with 100 ng of a plasmid encoding Firefly luciferase driven by the TF promoter and 2.5 ng of a plasmid encoding *Renilla* luciferase (pRL-TK; Promega) using a DNA (μg) to Lipofectamine 2000 (μL) ratio of 1:3. Cell lysates were collected 24 h after transfection.

2.6. Luciferase assays

After treatment, cells were harvested in 20 μL of lysis buffer per well (Promega). The activities of Firefly luciferase and *Renilla* luciferase were assayed using the Dual-Luciferase Reporter Assay System (Promega). Luminescence was detected by a GloMax 20/20 Luminometer (Promega) for 10 s with a 2 s delay according to the manufacturer's instructions. For each construct, at least three experiments with 20 transfections in each experiment were carried out. The transfection efficiency and cell number were based on co-transfected *Renilla* luciferase values and the total cellular protein for each plate. The reported values are expressed as the mean ± standard deviation (SD). All luciferase measurements were normalized for transfection efficiency and cellular protein.