Cell Research (2002); 12(5-6):339-352
http://www.cell-research.com

The promoter analysis of the human *C17orf25* gene, a novel chromosome 17p13.3 gene

JIAN YING GUO*, JIAN XU*, DA QIN MAO, LI LI FU, JIAN REN GU, JING DE ZHU**

The State-Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Ln 2200/25, Xie-Tu Road, Shanghai 200032, China

ABSTRACT

The human C17orf25 gene (Accession No. AF177342) is one of thirteen genes cloned from a region displaying a high score of loss of heterozygosity within chromosome 17p13.3 in human hepatocellular carcinoma in China[1]. To unveil the underlying mechanisms for the transcription regulation of this gene and understand its implication to the hepatocellular carcinogenesis, we looked into the relevant aspects by both bioinformatic and experimental executions. We found: 1, The abundant expression of the *C17orf25* gene was evident in all the cell lines and tissue samples tested, showing little hepatoma-selectivity; 2, Its transcription starts at a single site, locating at -60 from the translation initiation codon; 3, A 58 bp fragment containing the transcription start, extending from -112 to -55, represents the minimal promoter; 4, The consensus sequence within this fragment recognized by SP1 contributes predominantly to the activity of the minimal promoter; 5, The bioinformatic analysis suggests that the *C17orf25* gene may encode a protein in the family of the glyoxalase. Our data has provided some deep insight into both function and regulation of the C17orf25 gene in the context of the normal liver and hepatocellular carcinoma.

Key words: *C17orf25 gene, SP1, transcription regulation, chromosome 17p13.3.*

INTRODUCTION

Both previous works in this lab and others' have suggested that human chromosome 17p13.3 suffers from LOH at an extremely high frequency in hepatocelluar carcinoma[1-5] as well as other tumorous lesions of various tissue origins[6], [7]. It is generally believed that a single gene in the category of the tumor suppressor would reside within the LOH region[8],[9], turning abnormality of both alleles of which is necessary for the tumor formation. After fishing out the genes from the candidate region (approximately from 0.5 to 0.67 megabases from the telomere) of 17p13.3, a better knowledge is a must as to its function and regulation of expression of each of the genes within this region. The *C17orf25* gene is one of the thirteen genes identified and cloned via both bioinformatic prediction and cDNA screening, and annotated as such that stands for Chromosome 17 open reading frame 25[1]. The gene region of the C17orf25 is approximately 23 kb in length and comprises 10 exons and 9 introns. The full-length cDNA sequence is 1814 bp long and encodes a single ORF consisting 313 amino acids.

The regulation of transcript initiation of any newly discovered gene is often the first issue to be addressed, simply because of its unique status in gene expression and the comparatively advanced technological status of the relevant methodologies than that available for the studies of post-transcriptional and translational regulation of gene expression[10]. Both availability of the sequence specific

* Both authors have contributed equally to this work

** **Corresponding author:** Prof. Jing De ZHU
E-mail: zhujingde@yahoo.com Tel/Fax: 00 86 21 64224285

Abbreviations used in this paper:? LOH, loss of heterozygosity; HCC, hepatocellular carcinoma; SP1, specificity protein 1; AP-1, activating protein 1; RT-PCR, reverse transcription-PCR; EMSA, electrophoretic mobility shift assay; IVT, the in vitro transcription/translation.

Received Sep-12-2002 Revised Oct-21-2002 Accepted Nov-5-2002

transcription factors and accessibility of their cognate sites in chromatin, the real entity of the genome, are the major determinants of whether and how fast the transcription of a given gene occurs[11]. The detailed molecular analysis of the promoter is often the first characterization to be pursued, involving mapping the transcription initiation site, delineating the minimal region of the segment with a full promoter activity, and defining the key cis-elements along with their cognate transcription factors. In this report, we described our results from such an effort, which revealed the expression profile, transcription initiation, promoter structure, the key transcription factors for the promoter function of the C17orf25 gene. Our data may pave the way for the future studies of the detail molecular mechanisms underscoring the transcription regulation of the eukaryotic genes in general and the C17orf25 gene in particular.

MATERIALS AND METHODS

Cell culture, and transient transfection /reporter gene assays

SMMC7721 (human hepatocellular carcinoma cell line), U2OS (human osteosarcoma cell line), and LO2 (normal human liver cell line) were grown at 37℃ with 5% CO_2 atmosphere in Dulbecco's modified Eagle's medium (DMEM) supplied with 10% new born bovine serum.

The transient transfection/dual-luciferase reporter assays were carried out as previously described[12]. Briefly, cultured cells were transfected with a conventional $CaCl_2$/phosphate protocol and all the experiments were carried out in duplicate. Both the tested promoter fragments used to drive the Photimus pyralis (firefly) luciferase gene and the pRL-TK vector (Promega) where Renilla reniformis luciferase is driven by the promoter of the Herpes Simplex Viral thymidine kinase gene were transfected into cells at the same time. The cell extracts from the recipient cells were assayed for luciferase activities by a single-tube assay system (Dual-luciferase reporter assay system, Promega, USA). The promoter potency of the tested fragments were presented as the mean and standard deviation of the ratio of the firefly luciferase over the renilla luciferase activity of the duplicates and plotted against the tested constructs. The strength of the promoter was described as ratio of its luciferase activity over that of the control plasmid pGL3-Control, which contains a luciferase gene driven by SV40 early region promoter/enhancer (Promega, USA). The results are the mean ±S.D. of duplicate from one of three independent experiments.

RNA preparation and expression profiling by a semi-quantitative PCR

Total RNA was prepared from cultured cells and tissue samples with Trizol reagent according to the manufacturer's instruction (Invitrogen, USA). Then the total RNA was reverse transcribed using oligo(dT) as primer with Superscript II TMReverse Transcriptase (Invitrogen, USA). Semi-quantitative PCR was carried out with a pair of primers, each of which resides at different exons of the C17orf25 gene along with the pair of primers for the human b-actin used as the total cDNA loading control (Tab 1). The PCR condition is 94℃ for 5 min, following by 25 cycles of three steps: 94℃ for 30 s, 60℃ for 30 s, and 72℃ for 30 s. The PCR products were separated on a 1.5% agarose gel and the ethidium bromide stained bands were visualized under UV illumination.

Mapping of the transcription start site(s) by primer extension analysis

Primer extension was performed as described previously[10]. Primer extension on total RNA from LO2, SMMC7721, BEL7402 and C33A cells were performed using the C17orf25 specific anti-sense oligonucleotides PE1 5'-AATACGAAGTGCAGAGCTCTGCGAGCA -3' (+6 to +32, numbering from the initiating ATG). The size of the extended products was determined by referring to sequencing ladders generated with the same labelled primer on the relevant plasmid DNA templates in a T7 DNA polymerase based manu-sequencing protocol (Amersharm, UK).

Constructs and mutagenesis by the bridging-based two round PCR method

A ~1.2 Kb fragment upstream of the translation start site was PCR-amplified and subcloned into the promoterless firefly luciferase reporter gene vector pGL3-Basic (Promega) and designated F(-1278/-55). The oligonucleotide sequences were verified by DNA sequencing using vector-specific primers. The deletion mutants of the consruct F were made by deleting the fragment with flanking restriction enzymes as indicated in the relevant figures in the result section.

Site-directed mutagenesis of the C17orf25 gene's minimal promoter construct S(-112/-55) (Fig 3B) was carried out by a bridging-based two round PCR method as described previously [12]. Briefly, in the first round of PCRs, both 5' and 3' part of segments of the designated replacement mutants, in which 12 or 6 nucleotide sequences were expected to be substituted by EcoRI (GAATTC) and Bgl II (AGATCT) or either of them, were PCR out with two internal primers, the 12 or 6 nucleotides of 3' part of which are fully complementary to each other. The plasmid S (-112/-55) was used as template in the first round of PCR reaction. Then the products were gel purified, and used in the second round in pair. The second round of PCR was carried in 50 μl reaction solution containing 5 μl of 10 ×buffer, 1 ml of 10 mMdNTPs and 2.5 Units of pfu polymerase (Promega) for three rounds (95℃ 40 s, 55℃ 30 s, 72℃ 1 min) without any primers. Then 20 pmol of flanking primers were added into the reaction and amplified for another 25 cycles at 95℃ 40 s, 55℃ 30 s, 72℃60 s. Then the PCR products were inserted into pGL3-Basic vector's multiple cloning sites Kpn I and Hind III. All the mutants were verified by DNA sequencing. (The detail protocol will be provided under request.)

In vitro transcription/ translation of the transcrip-

tion factors E2F1, DP1, and SP1

The proteins encoded by the E2F1, DP1, and SP1 cDNA sequences were made in presence of [^{35}S] methionine by the in vitro transcription/translation assay (Promega), according to the manufacturer's instruction. The quantity of the products was estimated by autoradiograph on the dried SDS-polyacrylamide gel with the 1/10 input of products. The luciferase gene construct driven by T7 promoter provided by manufacturer was used as the positive control. The comparable amount of the each IVT products was used for the DNA-protein interaction assay described in the result section.

Electrophoresis mobility shift assay[12, 13]

Nuclear extracts were prepared from cultured cells by a modi fied protocol of the Dygnan[13]. The equal molar amount of the t wo complementary strands of the oligonucleotides were 5' kinase labelled with [γ-^{32}P]ATP (~5000Ci/mmol) as previously describe d[12]. The reactions were placed at 65℃ for 5 min to inactivate t he kinase and NaCl was added into the reaction to a final concen tration of 0.1 M. Then the reannealling was carried out by keepin g the reaction vials at room temperature for 2 h. The oligonucle otide used for this study was: 5' GCGCGCCGTCGGCGGCTAG TGAC 3' '(Only the sense sequence is presented). Then 10 fmol o f ^{32}P-end-labelled reannealled oligonucleotides were incubated wi th 5 μg of the protein extract on ice for 20 min in a 20 μl solution containing 1 μg poly dI:dC (Pharmacia), 12.5 m*M* Herpes, pH 7.9, 6.25 m*M* $MgCl_2$, 50 m*M* KCl and 10% glycerol. The DNA-protein complexes were separated by electrophoresis on a non-denaturin g 4% polyacrylamide gel with 0.5×TBE as running buffer (1×TB E: 89 mM Tris, 89 mM boric acid, 8 mMEDTA, pH 8.0) in a cold r oom (~7℃). For competition analyses, 20~100 fold molar exces s of cold oligonucleotides were incubated in the binding reaction. After electrophoresis, the gels were dried onto DEAE paper (Wha tman DE81) before autoradiography was taken.

To evaluate methylated CpG effects on the protein-DNA binding, the reannealled double-stranded oligonucleotide SP1 consensus (Con-U) was methylated in vitro} by SssI methylase in presence of S-adenosyl methionine (NEB, Boston, USA). The following mutant oligonucleotides: MutT, MutC and MutG, (for sequences see Tab 1), in which a central C/G was replaced by T, C and G respectively, were used as competitor in this study.

RESULTS

Expression profile

The LOH regions identified by comparing the tumor samples with the matched non-cancerous tissues for the occurrence of the molecular genetic markers (VNTR, STR and SNP, etc.) should be the primary target of the tumor suppressor gene with a casual role in the pathogenesis of the corresponding tumorous diseases[7],[9]. Whether expression of the candidate tumor suppressor gene has any tumor-selectivity is often the first aspect to be looked into. Once the *C17orf25* gene was cloned from the LOH region of 17p13.3, we used a semi-quantitative RT-PCR assay to establish the expression pattern at the level of the steady state mRNA of the *C17orf25* gene in the hepatocellular carcinoma tissues in parallel with the corresponding non-cancerous tissues, as well as a panel of the established cell lines. With the b-actin transcript as the loading control and the PCR reactions with a limited number of cycles, the best approximation of the relative expression level of the *C17orf25* gene over the b-actin gene was achieved [14]. As shown in Fig 1, no significant difference in the density of the PCR band was observed among all the cell samples, including LO2 (a fetal liver cell line) SMMC7721(a hepatocellular carcinoma cell line), BEL7402 (a hepatocellular carcinoma cell line), U20S (an osteosarcoma cell line), and C33A (a colon cancer cell lines) (Panel A), as well as six pairs of patients samples (hepatoma tissue, K and the matched non-cancerous tissue, L) (Panel B). This observation suggests that the *C17orf25* gene ubiquitously expresses and unlikely complies with the definition of a conventional tumor suppressor gene. In addition to the lack in hepatoma-selective expression (both in positive and negative senses), the steady state mRNA level of the *C17orf25* gene exceeded drastically the level of the internal control, β-actin, implying the *C17orf25* gene mRNA is abundant in

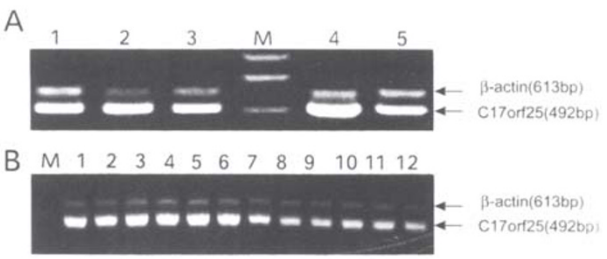

Fig 1. Expression profiling of the C17orf25 gene
Total RNA from cultured cells and tissues were isolated and semi-quantitative RT-PCR was performed using b-actin as an internal control. **A,** Lane 1-5, LO2, SMMC7721, BEL7402, U2OS, C33A cells. **B,**Lane 1-2, 1K and 1L, lane 3-4, 6K and 6L, lane 5-6, 12K and 12L, lane 7-8, 24K and 24L, lane 9-10, 28K and 28L, lane 11-12, 29K and 29L. "K" indicates HCC samples while "L" indicates adjacent non-cancerous liver tissues.

the majority of, if not all, cell types. However, the possibility for this gene to have a vital role in the

hepatocellular carcinogenesis remains if there were hepatoma specific mechanisms responsible for the differential expression at the protein level of the *C17orf25* gene.

The transcription of the human C17orf25 gene starts at a single site at -60 from the A residue of the first codon

The information crucial to the downstream boundary of the promoter of any given gene is concerning the start of transcription. Despite of the considering prediction power from the software based on the statistic survey of an increasing number of the genes, it remains necessary to draw the relevant conclusion experimentally. To this end, we performed primer extension analyses with the complementary oligonucleotide probe PE1 (+6 to +32, numbering from the initiation ATG) (Fig 2B). As shown in Fig 2, we found three distinct extended products from total RNA of LO2, SMMC7721, BEL7402 and C33A cells. To facilitate the mapping, the sequencing ladders made with the identical primer and the plasmid DNA containing the relevant insert as template were run in parallel. There are three bands representing the starts at the -60, -21 and -15 at upstream of the gene (Panel A and B, Fig 2). However, the primer-extended products may also result from the premature termination of the reverse transcription at the GC rich region of the mRNA template. Indeed, both downstream starts (-21 and -15) are embedded in a high GC rich region (Panel C, Fig 2). In this connection, the transient transfection/dual reporter assays for the promoter analysis (described in the late section, Fig 3), which showed that the promoter segments devoid of the -21 and -15 sites maintain the potent promoter activity, favors the notion that the *C7porf25* gene may initiate at a single site that was mapped at -60.

A 58 bp fragment extending from -112 to -55 is the minimal promoter of the human C17orf25 gene

To evaluate the promoter activity in cells, a common and efficient approach is to transfect the constructs in which the reporter gene, such as firefly luciferase gene is at the downstream of the tested DNA fragment, into the cells and assess the reporter

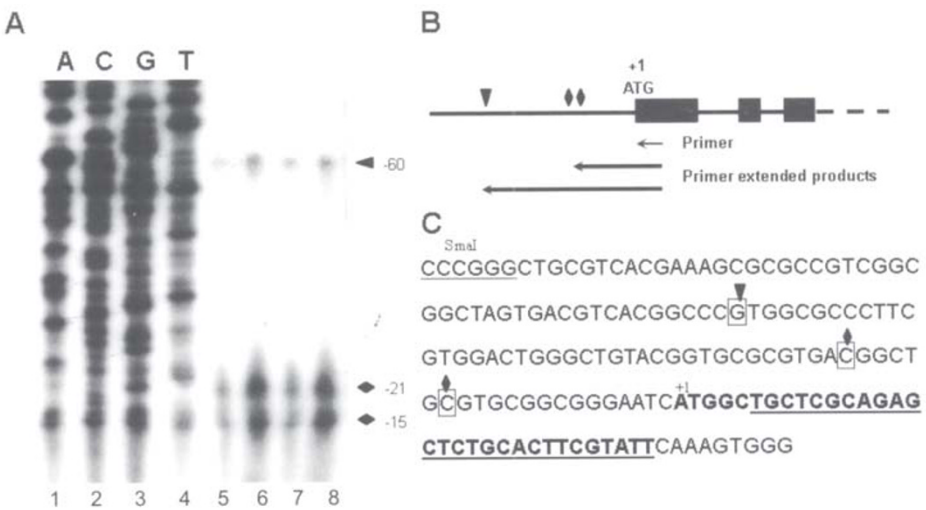

Fig 2. Determination of the start sites by primer extension **A**, Primer extension mapping of the *C17orf25* initiation sites using 32P-labelled oligonucleotide PE1. Lane 1-4, the nucleotide of the sequencing reaction using primer PE1 and a genomic fragment; lane 5-8, primer extension reaction with 10 mg total RNA from LO2, SMMC7721, BEL7402 and C33A cell lines. The primer-extended products are indicated by black arrowhead (major start, -60) and black diamond} (minor starts, -15 and -21. **B**, Schematic illustration of the relevant positions of the primer to the templates, black bars indicate coding region of *C17orf25* gene. **C**,? Nucleotide sequence of the region surrounding initiation sites. The location of the translation start site ATG is marked as position +1 and the oligonucleotides used for primer extension analysis are underlined.

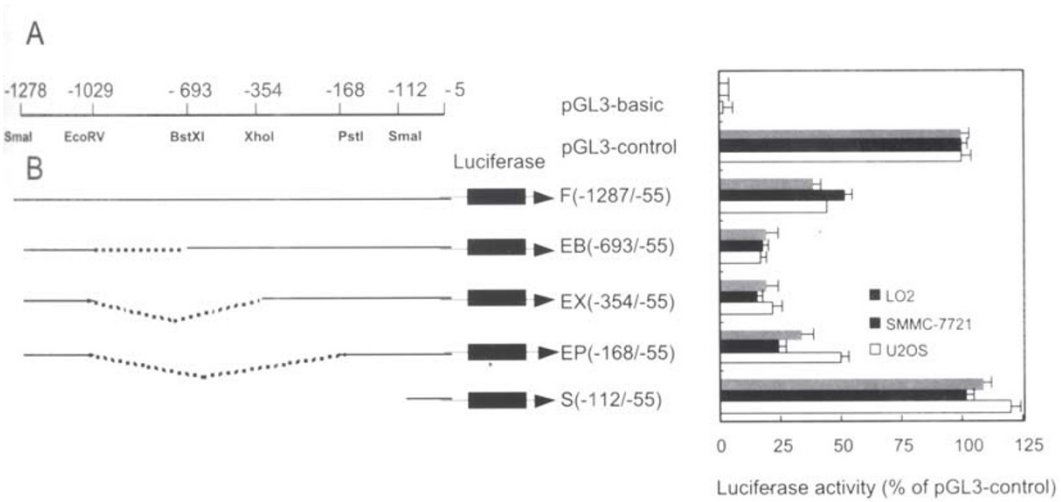

Fig 3. Deletion mapping of the *C17orf25* minimal promoter **A**, A partial restriction map of the C17orf25 promoter (-1278 to -55, numbering from the initiation ATG). **B**,? Schematic illustration of the C17orf25 reporter constructs used in transient transfection analysis of promoter activity in LO2, SMMC7721 and U2OS cell lines. Thin lines indicate vector sequences, black bars indicate luciferase reporter gene, and thick lines indicate the C17orf25 promoter fragments. Positions relative to the translation initiation site are indicated. The promoter-reporter constructs are co-transfected with a control plasmid (pRL-TK) and assayed 24 h later. The results are the mean ±S.D. of duplicate from one of three separate experiments and are shown as percentage of the activity of pGL3-Control (see the relevant section in Materials and Methods).

protein activity in the cell extracts after a couple of days? culture. To start with our functional analysis, a 1224 bp fragment covering from -1278 to -55 region of the *C17orf25* gene was PCR cloned to the 5' end of the firefly luciferase gene in the pGL3-Basic vector, and designated as F(-1278/-55).

This fragment displayed approximately 50% promoter activity of that of the combination of the SV40 early promoter/enhancer in pGL3-Control construct in all the three recipient cell lines: LO2 (an established cell line from the human embryonic liver), SMMC 7721 (a hepatoma cell line) and U2OS (an osteosarcoma cell line). The internal deletion mutants: EB(-693/-55), and EX(-354/-55) had their promoter activities reduced to the half of the level of the full length promoter. The internal mutant EP(-168/-55), in which a region from -354 to -169 were further deleted from mutant EX(-354/-55), had a similar level of promoter activity in SMMC7721 cells, and a higher level in LO2 cell and a two fold high level in U2OS cells to that of mutant EX(-354/-55). The significance of such differences among all the three cell lines remains to be investigated. However, in a simplistic term, the region between -1029 and -693 may have a positive determinant to the promoter activity of the *C17orf25* gene. The deletion mutant S (-112/-55) where the majority of the upstream region (from -1278 to -113) is deleted, exhibited a potent promoter activity, approximately two fold higher than that of the full length promoter F(-1278/-55). The observed potency of the promoter activity of this 58 bp fragment (-112 to -55) is fully compatible with the abundant level of the endogenous *C17orf25* gene expression, shown in Fig 1. Therefore, we inclined to conclude that the minimal promoter of *C17orf25* gene spans from -112 to -55, where the +1 is the A residue in the translation initiation codon, ATG and gathered our efforts in identifying the crucial cis-elements in this 58 bp fragment, followed by further detailed molecular genetic and biochemical elaboration.

The SP1 site (-92 to -87) is crucial to the activity of the minimal promoter

The availability of the databases accompanied with the robust searching engines, the disposal of the DNA motifs recognized by the known transcription factors of which are expanding rapidly have facilitated the experimental identification and verification of the key cis-elements crucial to the promoter activity of any newly discovered genes[15]. We analyzed a 500 bp segment extending from -471 to +29

with 'Human promoter mapping using CpG islands' (http://cgsigma.cshl.org/CpG_-promoter/) and the software Promoterscan (http://www.ifti.org/). As summarized in Fig 2, there are two GC rich regions falling into the category of the CpG islands. Lacking of the classic TATA and CCAAT boxes near the transcription start site put the C17orf25 gene into the category of the TATA-less gene[11],[16]. A number of consensus sequences recognized by the known transcription factors[17] have been identified, including one E2F-DP/SP1 consensus site, one AP-1 and two AP-2 sites near the initiation site (Fig 4). Within the minimal promoter (-112 to -55, a 58 bp fragment), there are several motifs identified (Fig 5A): one AP-1 site (-105 to -99) and one E2F-DP/SP1 site (-92 to -87). However, the computer prediction is largely based upon the data from the in vitro binding assay, and tends to over-estimate the significance of the sequence similarity. The question as to whether these sites do have any functional role needs to be answered experimentally. Therefore, we systematically mutated and functionally assayed four mutants, each of which in the context of the minimal promoter (-112 to -55) had respectively 12 nucleotides sequentially replaced with the following sequence: EcoRI (GAATTC) and BglII (AGATCT) restriction enzyme sites that lack any known cis-elements (named as mutAB, mutCD, mutEF, and mutGH in Fig 5B). All of the four mutants have the reduced promoter activity to various extents (Fig 5C), indicating that the sequence affected should have some contribution to the minimal promoter activity. For instance, the mutAB, where an AP-1 site is located, deliver a promoter activity no more than 40% of the minimal promoter. It is quite likely that the AP-1 based

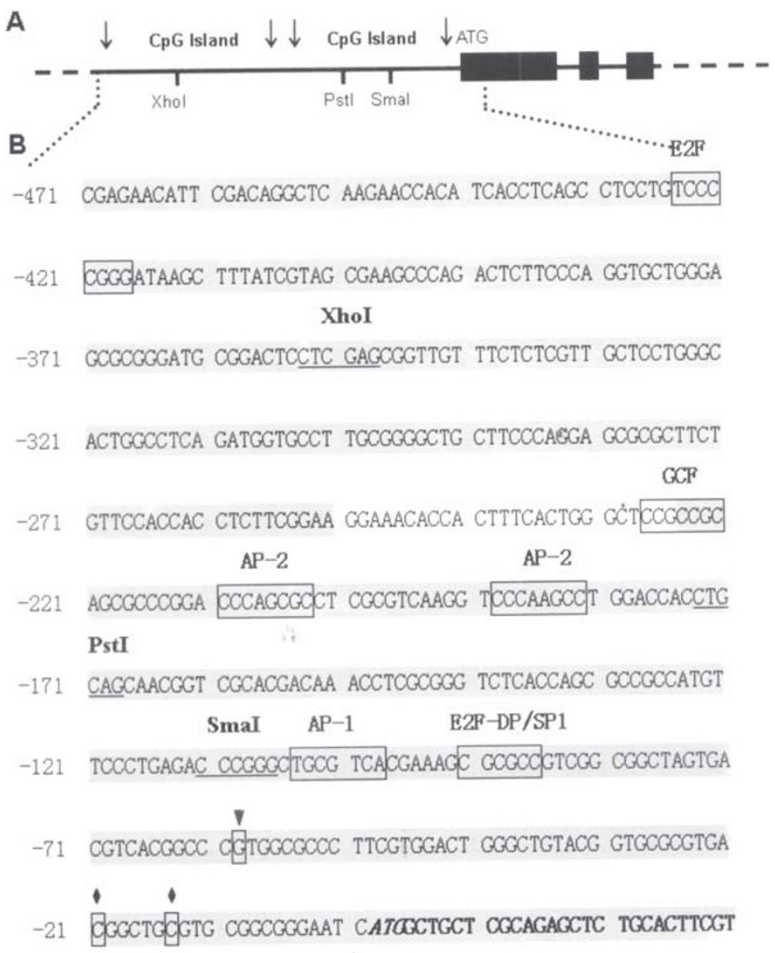

Fig 4. The primary structure of the human *C17orf25* gene promoter **A**, A partial restriction map of a 500 bp genomic fragment containing the 5'-flanking region of the *C17orf25* **B**.Nucleotide sequence (-471 to +29) of the genomic region surrounding the transcription starts (the major start is indicated by *a black arrowhead* while the other start site are indicated by *black diamond*). Two CpG islands are indicated with shaded characters. Potential cis-acting elements are indicated by box and discussed under "Result" Nucleotides indicated in boldface are coding sequences of the *C17orf25* gene.

mechanism may indeed be responsible for the effect revealed by the analysis of this mutant. mutGH, in which the sequence 5' TCACGGCCCGT3' were replaced by 5' AATTCAGATCT3', delivered no more than 60% of the promoter activity of the minimal promoter. But the changes represented by the mutants: mutCD(-94/-83) and mutEF(-82/-71), were most detrimental, no more than 15% to 25% of the wild type activity remained. The sequence changes in mutCD (-94/-83) affect the integrity of the cis-acting elements: E2F-DP/SP1 site (-92 to -87). There is no known cis-element within the sequence being altered in mutEF(-82/-71). Additional four mutants where six bp sequence was sequentially replaced within the segment (-94/-89) with EcoRI (GAATTC) or BglII (AGATCT) were made and tested in the transient transfection/dual reporter assay (Panel B, Fig 5). As shown in panel C, Fig 5, mutCD1 (-94/-89) and mutCD2(-88/-83), reduced to a fully compatible extent in the promoter activity of the mutCD(-92/-83). Therefore, the sequence representing the E2F-DP/SP1 site initially identified by bioinformatic prediction is likely vital to the promoter activity of the *C17orf25* promoter. MutEF1 (-82/-77) having the upstream six nucleotide sequence where mutEF2 (-76/-71) having the rest six that mutated in mutEF respectively exhibited 3 and 1 fold recovery in the promoter activity, caused by the mutation in mutEF. This suggests that the sequence changes shared by both mutEF and mutEF2 may represent an important novel cis-element that has not been recognized by the bioinformatic analysis in this study, characterization of this novel cis-element and its cognate transcription factors is a subject of the future investigation.

Both E2F-DP and SP1 transcription factors share a considerable degree of similarity in their cognate DNA sequences. The distinction should be made as to each's role in the promoter activity of the *C17orf25* gene. To this end, we employed the transient transfection/dual reporter assay in which the constructs enclosing a given target transcription factor gene (in this case, E2F1, DP1 and SP1, respectively) driven by a potent mammalian promoter were co-transfected with the relevant promoter constructs. As shown in panel A, Fig 6, the co-trans-

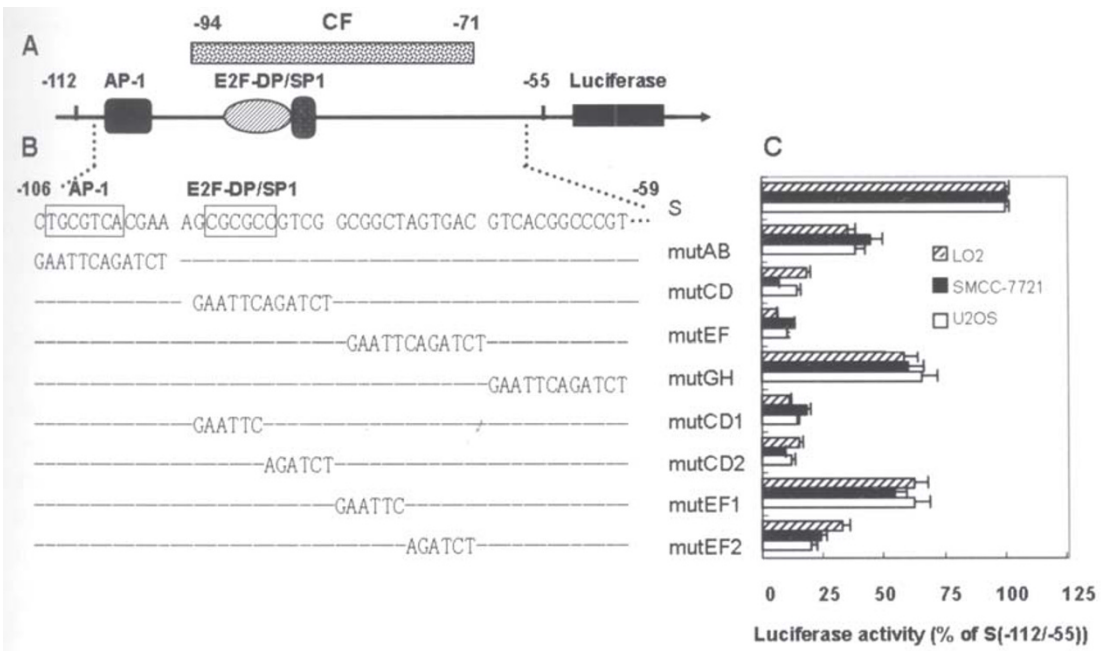

Fig 5. The promoter activity of the sequence replacement mutants of the *C17orf25* minimal promoter **A**, A partial map of the cis-elements recognized by the prediction software. The CF fragment covering a region from -94 to -71 is the double-stranded oligonucleotides used in EMSA. **B**, The alignment of the mutated sequence in each 12 or 6 nucleotide replaced mutants over the wild-type, S(-112/-55). **C**, The luciferase activity of each mutant in LO2, SMMC7721, U2OS cells is presented as the mean±S.D. of duplicate samples from one of three separate experiments using S(-112/-55) as reference.

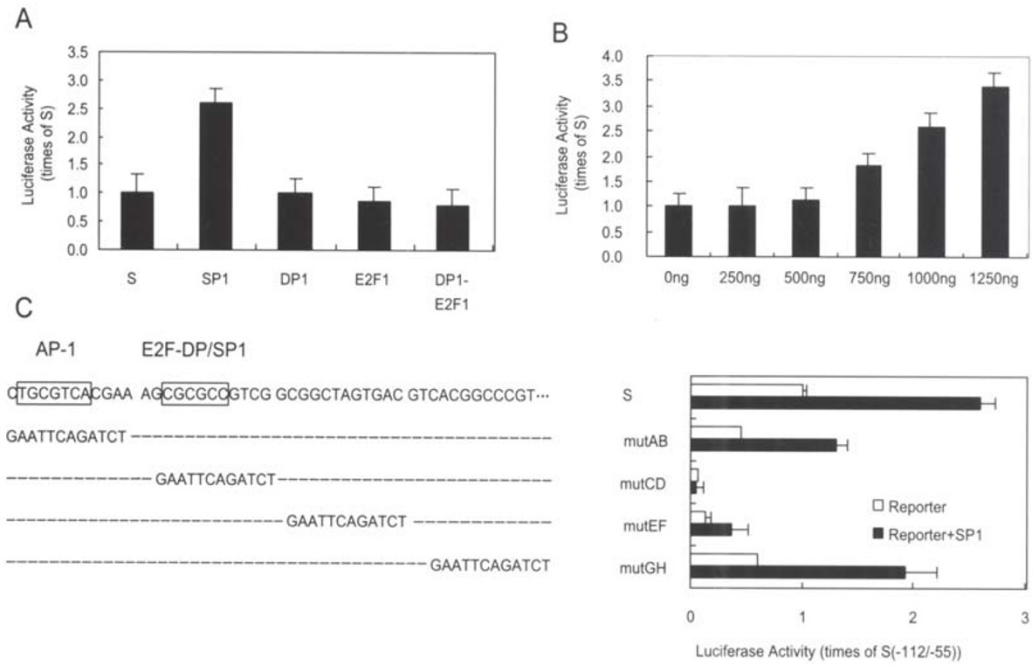

Fig 6. SP1 promotes the *C17orf25* promoter activity by co-transfection assays **A**,Co-transfection of SP1, DP1 and E2F1 expression plasmid (1 μg) with C17orf25 minimal promoter construct S(-112/-55) in SMMC7721 cells. **B**, Effect of co-transfection of a SP1 expression vector HA-SP1 on the C17orf25 promoter activity in SMMC7721 cells. 1 mg of S(-112/-55) promoter-reporter construct plus up to 1.25 μg of HA-SP1 were co-transfected with a control plasmid (pRL-TK) and assayed 24 h post-transfection. **C**,Effect of co-transfection of HA-SP1 on sequence replacement mutants.

fected SP1 construct elevated the minimal promoter activity [in the context of S(-112/-55)] as much as 2. 5 fold and it is dosage-dependent (Panel B, Fig 6). The vital role of SP1 was further strengthened from the co-transfection experiments with mutAB, mutCD, mutEF and mutGH along with the SP1 construct, showing that mutCD where the SP1 consensus were mutated, refused to respond to the input SP1 construct (Panel C, Fig 6), whereas the rest responded equally well as the wild type of the minimal promoter construct.

By a sharp contrast, E2F1 and DP1 on their own as well as in combination had no detectable effect. This implies that E2F1/DP1 may have no functional role to play in the context of the minimal promoter, its picking-up by the computer prediction simply reflects the inherent imperfection of the existing software designed for such a purpose.

The biochemical evidence supporting to any conclusion drawn from the experiments by the transient transfection/reporter assay described above, is routinely required in the study of this kind. We, hence, carried out a sensitive in vitro DNA-protein interaction assay, in which various DNA-protein complexes can be distinguished in a polyacrylamide gel electrophoresis system, namely, EMSA. In view of the demonstrated functional role of the sequence from -94 to -71, the affected sequence, jointly represented by mutCD and mutEF, we used kinase labelled double stranded oligonulcotides of this region (Tab 1) as the probe to interact with the protein extracts from the LO2 and SMMC 7721 cells, respectively. There were two predominant bands in the lanes with either LO2 or SMMC7721 nuclear extracts, both were effectively competed out by the added 20 to 100-fold molar excess of the unlabelled identical oligonucleotides or SP1 consensus Con-U (Tab 1). However, the nature of the protein components in each of these two complexes remains unclear. An important question is whether E2F1/

DP1 transcription factor may contribute to the formation of any of these two complexes, as the data from the co-transfection experiment failed to support their involvement in the function of the minimal promoter (Fig 5). To this end, we made the E2F1, DP1 and SP1 proteins by the in vitro transcription/translation method in a reticulocytic lysate based system (Promega, USA), in which the proteins

Tab 1. Oligonucleotides used in this study

Oligonucleotide	Sequences
C17orf25-L	GAGGGAAGCAAATTGTTGGA
C17orf25-R	CGACAAGGCATTATTAATGTGGA ÊÊÊÊRT-PCR
β-actin-L	AAGTACTCCGTGTGGATCGG ÊÊÊÊÊÊÊÊEMSA
β-actin-R	TCAAGTTGGGGGACAAAAAG
CF	AGCGCGCCGTCGGCGGCTAGTGAC
Con-U	CTCTACTCCCAGAAGGCCGCGGGGGGTGGAC
Con-M*	CTCTACTCCCAGAAGGC\underline{C}G\underline{C}GGGGGGTGGAC
MutT**	CTCTACTCCCAGAAGGC\underline{T}GCGGGGGGTGGAC
MutC**	CTCTACTCCCAGAAGGCC\underline{C}CGGGGGGTGGAC
MutG***	CTCTACTCCCAGAAGG\underline{G}CGCGGGGGGTGGAC

* The C residues which were specifically methylated in vitro by SssI methylase in CpG dinucleotides were underlined.
**The mutated nucleotides within the SP1 consensus sequence were underlined.

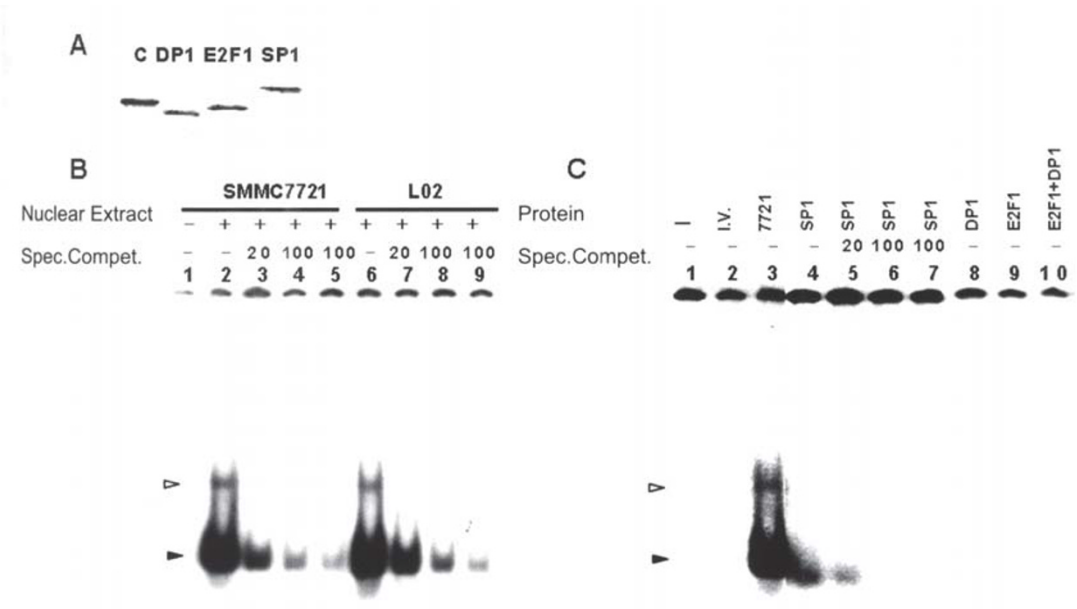

Fig 7. Physical association of SP1/SP1-like proteins with the *C17orf25* promoter region by EMSA
A, The plasmids HA-DP1, HA-E2F1 and HA-DP1 were translated in vitro using TNT®uick Coupled Transcription/Translation System (Promega) and the radioactive fusion proteins were separated by 8% SDS-PAGE. The protein (C, 61KD) was used as a positive control. **B**.^{32}P-labelled double-stranded oligonucleotides CF (position -94 to -71) were incubated with 5 μg nuclear extract from SMMC7721 (*lane 2*) and L02 cells (*lane 6*). The DNA-protein complexes were competed with 20 or 100-fold molar excess of specific competitor (*lane 3-4, lane 7-8*), or SP1 consensus oligonucleotide Con-U in the form of double-stranded oligonucleotides (*lane 5 and 9*, for sequences see Tab 1). C, ^{32}P-labelled double-stranded oligonucleotides CF were incubated with 5 mg SMMC7721 nuclear extract or 2 μl of the in vitro translated proteins (*lane 4, 8, 9 and 10*) and competed with 20 or 100-fold molar excess of specific competitor (*lane 5-6*), or SP1 consunsus Con-U (*lane 7*).

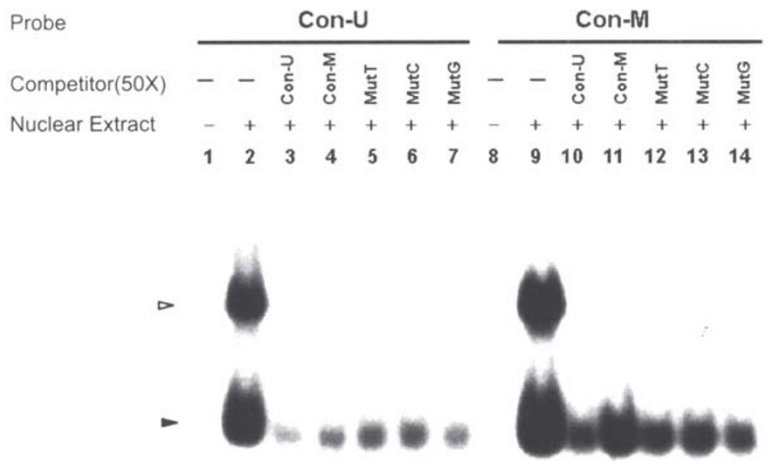

Fig 8. Methylation analysis of SP1 consensus sequence ^{32}P-labelled double-stranded oligonucleotides Con-U (SP1 consensus) and Con-M, in which C residue in CpG dinucleotides were methylated in vitro by SssI methylase, were incubated with 5 μg SMMC7721 nuclear extract (lane 2 and 9). The DNA-protein complexes were competed with various competitor Con-U, Con-M, MutT, MutC and MutG (lane 3-7, lane 10-14).

were labelled with [S^{35}] methioneine. As showing in panel A Fig 7, all four reactions resulted in a comparable level of products, namely: the firefly luciferase (lane C,) DP1, E2F1 and SP1 in panel A. Both the control (lane 2, IV, the in vitro transcription/translation system without any DNA input) and DP1, E2F1 on each own and in combination (lane 8, 9, 10) shew no sign of binding to the CF fragment (Panel C, Fig 7), while the SP1 could and its binding was amenable to competition by excess of the cold same oligonucleotides (lanes 4 and 5, panel C). This piece of evidence further supported the notion, i.e., SP1 and the relevant sequence-specific binding within the *C17orf25* promoter indeed contribute significantly to the in vivo transcription of this gene. It was also noticed thatonly the band similar to the fast moved complex of the two complexes produced by 5 μg SMMC7721 nuclear extract was present in the lanes of the IVT SP1. This may suggest that the band that was retarded more (indicated by open triangle) was associated with the additional nuclear proteins that were absent in the IVT reaction. It is also noticed that the complexes created by the IVT SP1 moved further than the complexes from the nuclear extracts (comparing lane 3, SMMC 7721 nuclear extract andlanes 4 and 5). In light of the fact that the SP1 is heavily post-translationally modified in cell, including polyglycosylation[18],[19], the above observation suggests that the IVT SP1 lacks the most, if not all, of the post-translational modification.

It has been well established that methylation of C residues in the CpG dinucleotide can significantly

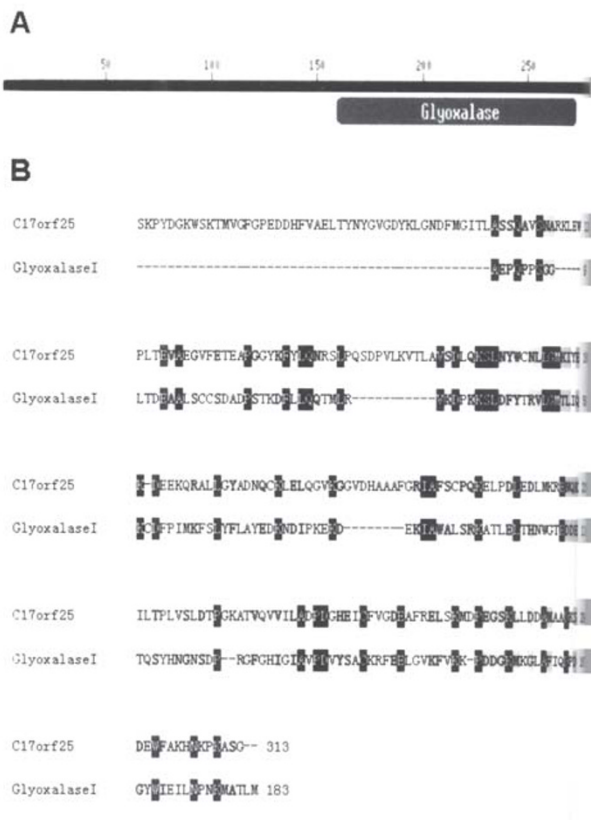

Fig 9. Homology analysis by alignment analysis of the *C17orf25* protein with Human Glyoxalase I A, Human C17orf25 protein contains a conserved domain belonging to glyoxalase family revealed by searching the conserved domain database (NCBI). **B,**Alignment of *C17orf25* protein with Human Glyoxalase I. The amino acid sequence was compared using CLUSTAL W software (http://www.ebi.ac.uk/clustalw/). The gray boxes indicate the positive amino acids in these proteins.

affect the relevant sequence specific protein-DNA interaction with a profound functional conse quence in the gene transcription[20]. In this context, SP1' s consensus, "GGCGCG" have two CpG dinucleotides. It is desirable to find out whether methylation CpG dilucleotides within SP1' consensus may have any effect on the interaction between SP1 and its cognate site. We compared the SP1 consensus (Con-U) with its methylated version (Con-M) for their capacity to be bound by nuclear proteins from SMMC7721 cells. The oligonucleotides were in vitro methylated by SssI methylase and lablled with ^{32}P. As shown in Fig 8, the nuclear extracts can bind equally well to the unmethylated and methylated version of the double-stranded oligonucleotides containing the SP1 consensus, and the binding was equally well abolished by the excess of the cold unmethylated and methylated oligonucleotides. The methylated C residue shares a great deal of the resemblance at the secondary structural level with the T residue. Therefore, the excess of each of three double stranded oligonucleotides where the C was replaced with T were used as the cold competitors and shew no reduction in their competition ability(Fig 8, lanes 5-7, 12-14). Both observations suggested that the SP1 mediated transcription regulation is not under the methylation related constraints of its cognate sites in vivo. It naturally follows that the transcription regulation of the *C17orf25* gene is less likely subjected to the control at the level of the DNA methylation.

DISCUSSION

The functional aspects of the C17orf25 protein

The vast disposal of the sequence information of an increasing number of organisms? genome in the public database has drastically enhanced our capacity of predicting almost every aspect of any novel gene before the full-scale of experimental investigation is ever initiated. The sequence similarity of the target protein with the protein(s), the function of which has been well established experimentally has been commonly used, as it would shed some lights on our understanding of the biological function of the target protein[21]. The *C17orf25* gene (No. AF177342) gene was one of thirteen genes cloned from the 116 Kb fragment within the 17p13.3 region that displays a frequent loss of heterozygosity in the human hepatocarcinoma, where the gene in the category of the tumor suppressor should reside[1], [5]. We blasted the protein sequence database (http://www.ncbi.nlm.nih.gov/) with the *C17orf25* amino acid sequence and identified a high conserved region with glyoxalase family, in the C17orf25 gene. Then we aligned the amino acid sequence of *C17orf25* with that of human glyoxalase I gene using the software at http://www.ebi.ac.uk/clustalw/. As shown in Fig 9, the alignment illustrates the homology to Glyoxalase I, a member of the metalloglutathione (GSH) transferase superfamily, that plays a critical detoxification role in cells by catalyzing the conversion of cytotoxic methylglyoxal (as the diastereomeric GSH-thiohemiacetals) to S-D-lactoylglutathione via a 1,2-hydrogen transfer[22]. The percentage of identity is 13% at level of amino acid sequence. It has been reported that the expression of the glyoxalase I gene can be tumor-specific [23], implying its potential contribution to the genesis of the human hepatocellular carcinoma. However, the roles in the hepatocellular carcinogenesis of the *C17orf25* gene remain to be experimentally determined.

Transcription regulation of the C17orf25 gene

Although the expression profile of the *C17orf25* gene does not show any hepatoma specificity at the steady state level of mRNA (Fig 1), indicating that it may not fall into the category of the conventional tumor suppressor genes or its "desirable function" may be executed beyond the level of transcription, the study of its transcription regulation of the *C17orf25* gene as a newly discovered gene, remains of importance. The feature and implication of its transcription may be unique. Therefore, we dissected the primary structure of the *C17orf25* gene promoter with a range of molecular approaches.

Using the primer extension assay, we have identified three extended products, probably representing the starts at -60, -21 and -15 upstream of the A residue of the initiation codon of translation (Fig 1). The reverse transcription reaction on the RNA template tends to stop at any regions rich in the GC nucleotides. Therefore, the information as to the 3' boundary of the promoter segment delineated from the transient transfection/dual reporter assay (Fig 3) should be taken into account before the functional

nature of each of these three extended products is concluded. Indeed, the downstream two "starts" were mapped at an extremely high GC rich region (5' CGCGTGACGGCTGCGTGCGGCGGG3' GC content is 84%) and the DNA segment devoid of these two sites maintained the potent promoter activity (Fig 2, 3). Therefore, it is likely true that the *C17orf25* gene starts at a single site for transcription, mapped at -60 upstream of the A residue of the initiation codon of translation.

By transient transfection/dual reporter assay, we initially compared the relative promoter activity of the fragments lacking various parts of the full length promoter, i.e, a 1224 bp fragment covering from -1278 to -55 region of the *C17orf25* gene. The internal deletion mutants: EB(-693/-55), EX(-354/-55) and EP(-168/-55) had their promoter activity reduced to the half of the level of the full length promoter. The mutant S(-112/-55) exhibited a potent promoter activity, approximately two fold higher than that of the full length promoter F(-1278/-55). The observed potency of the promoter activity of the latter is fully compatible with the abundant level of the endogenous *C17orf25* gene expression, shown in Fig 1. Therefore, the simplistic explanation is likely to be that the region between -1029 and -113 may have a negative determinant to the promoter activity of the *C17orf25* gene, by taking the minimal promoter S as reference, and a positive determinant in the context of the full promoter in construct F. Although the work presented in this report is exclusively concerned with the details of the minimal promoter, it remains of a great significance to study the roles of and the underlying mechanisms of the fragment (-1029 to -113) in the transcription regulation of the *C17orf25* gene in the future.

The systematic analysis was carried out of the central 48 nucleotide (-106 to -59), 12 or 6 nucleotides of which were sequentially replaced by the sequence devoid of any motifs recognized by known transcription factors in each of a panel of the mutants, for their relative promoter activity (Fig 5). mutAB(-106/-95), and mutGH(-70/-59) exhibited approximately 40% and 60%, whereas mutCD(-94/-83) and mutEF(-82/-71) delivered no more than 15% to 25% of the promoter potency of the minimal promoter, respectively. It is logically correct that the sequence between -94 to -71, the altered sequence represented by the mutCD and mutEF should have a more important role in the promoter function and be worthwhile being pursued further. Therefore, the investigation aiming at the cis-element within this segment was performed further. In addition to the motif of E2F-DP/SP1 within the region from -94 to -83 (mutCD) that was analyzed and reported in great detail in this paper, the altered sequences presented in both mutEF and mutEF2 had also contributed to the promoter activity significantly. There are no motifs recognized by any known transcription factors, presenting a tantalizing possibility that a novel cis-element may reside.

Although the transient transfection/dual report assay is a powerful approach to address the matter of transcription regulation in general, and the cis-element in particular, of any given promoter, the biochemical evidences for the specific multi-molecular interactions should also be obtained. Therefore, after the initial identification of the crucial segments CF (-94 to -71) (Fig 5), we carried out the EMSA analyses of the sequence-specific interaction between the CF fragment and the nuclear proteins from the cells that were used as recipients for transfection analysis. As shown in Fig 7, the identity of the crucial element largely responsible for the transcription activity of the C17orf25 gene promoter, i.e., SP1 and its cognate transcription factor was confirmed by the following: 1, the candidate site within the *C17orf25* gene promoter is similar to the consensus; 2, the interaction with relevant fragments in absence or presence of the specific competitors did unambiguously support the conclusion drawn from the transfection experiments; 3, the SP1 protein made from the in vitro transcription/translation possessed the desirable pattern of the binding activity. After establishing the predominant role of the SP1 in the transcription of the minimal promoter of the *C17orf25* gene, we have looked into a very important aspect of the regulation that is concerning any motifs containing CpG dinucleotide, the C residue of which is amenable to the methylation in the higher eukaryotes and alteration of its methylation state would result in changes in the expression pattern of the relevant genes, via affecting the sequence-specific DNA-protein interaction[20]. With the in vitro methylated oligonucleotides containing SP1 consensus, we had shown that SP1 bound equally

well to its cognate sequences no matter whether the CpG within is methylated or not. Therefore, although the promoter of the *C17orf25* belongs to the gene family, the promoter of which has the CpG islands nearby, its transcription may not be subjected to the DNA methylation mediated control, at least, in the context of the SP1 site.

In this content, we have also made efforts to distinguish the SP1 role from the E2F/DP' s, both can play important functional roles in transcription regulation of a large variety of genes, including those participating in cell proliferation and apoptosis. Firstly, we closely looked into the relative confidence value of each of these two motifs in the context of the sequence, the altered version of which is represented in the mutCD. The E2F-DP' s consensus is SGCGCS (where S is either C or G), while the SP1' s consensus is GGCGCG. The motif within the minimal promoter of the *C17orf25* gene is CGCGCC. The expect value for this sequence to be bound by E2F/DP proteins is 6.12e-01, which is higher than that (2.10e-01) by SP1, implying that the element within the minimal promoter is more likely to be actually recognized by SP1 rather than E2F/DP proteins (http://www.ifti.org/cgi-bin). Secondly, the E2F1 and DP1made by IVT on their own or in combination did not bind to the CF fragment as SP1 did (Panel C, Fig 7). Furthermore, the in vitro SP1 binding can accommodate the change at the level of the single point mutation within the consensus, as shown in Fig 8. Finally, the co-transfection data with the exogenously introduced E2F1 and DP1 on their own or in combination did not affect the minimal promote activity, while the SP1 did elevate the promoter activity of the minimal promoter, but not that of the mutCD and mutCD1 and mutCD2, in which the consensus were drastically changed (Fig 6). In conclusion, SP1 rather than E2F1/DP1 is likely to play a vital role in the promoter activity of the *C17orf25* gene in cell.

Although the results described in this report has considerably advanced our understanding of the transcription regulation of the *C17orf25* gene, a novel gene identified and cloned from the 17p13.3 region that suffers a high score of LOH, many questions remain. To fully comprehend the biological function, transcription regulation along with the realization of the relative value of each of the molecular approaches used in this report, much effort should be put into with the emphasis at the systematic (holistic) view regarding to both theoretic framework and experimental executions.

ACKNOWLEDGEMENTS

This work was supported by the 973 project of China (G1998051004) to Jing De ZHU.

REFERENCES

[1] Qin WX, Wan F, Sun FY. Cloning and characterication of a novel gene (*C17orf25*) from the deletion region on chromosome 17p13.3 in hepatocelular carcinoma. Cell Res 2001; **11(3)**:209-16.
[2] Fujimori M, Tokinom T, Hino O. Allelotype study of primary hepatocellular carcinoma. Cancer Res 1991; **51**:89-93.
[3] Nagai H, Pineau P, Tiollais P, Buendia MA, Dejean A. Comprehensive allelotyping of human hepatocellular carcinoma. Oncogene 1997; **14**:2927-33.
[4] Piao Z, Park C, Park JH, Kim H. Allelotype analysis of hepatocellular carcinoma. Int. J. Cancer 1998; **75**:29-33.
[5] Wang G, Huang CH, Zhao Y. Genetic aberration in primary hepatocellular carcinoma: correlation between p53 gene mutation and loss-of-heterozygosity on c hromosome 16q21-q23 and 9p21-p23. Cell Res 2000; **10**:311-23.
[6] Ittmann MM. Loss of heterozygosity on chromosomes 10 and 17 in clinically localized prostate carcinoma. Prostate 1996; **28(5)**:275-81.
[7] Lothe RA, Fossli T, Danielsen HE. Molecular genetic studies of tumor suppressor gene regions on chromosomes 13 and 17 in colorectal tumors. J Natl Cancer Inst 1992; **84 (14)**:1100-8.
[8] Stack M, Jones D, White G. Detail mapping and loss of heterozygosity analysis suggests a suppressor locus involved in sporadic breast cancer within a distal region of chromosome band 17p13.3. Hum Mol Genet 1995; **4**: 2047-55.
[9] Zhao XT, Li JJ, He YH. A Novel Growth Suppressor Gene on Chromosome 17p13.3 with a High Frequency of Mutation in Human Hepatocellular Carcinoma. Cancer Res 2001; **61**:7383-7.
[10] Zhu JD. Myeloidcell-lineage and premylocytic-stage-specific-expression of the mouse myeloperoxidase gene is controlled at initiation as well as elongation levels of transcription. Cell Res 1999; **9**:107-34.
[11] Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et Biophucica Acta 1997; **1351**:73-88.
[12] Xu J, Zhu JD, Ni M. The ATF/CREB site is the major regulatory element for transcription of the human RNA methyltransferase like 1(RNMTL1) gene, a newly discovered 17p13.3 gene. Cell Res 2002; **12**:177-97.
[13] Zhu JD. FT I, a novel positive myeloid-lineage-specific transcription regulatory element within the mouse myeloperoxidase gene enhancer, En 1. Cell Res 1995; **5**: 75-91.
[14] Ferre F. Quantitative or semi-quantitative PCR: reality

versus myth. PCR Methods Appl 1992; **2(1)**:1-9.
[15] Kashuba VI, Gizatullin RZ, Protopopov AI. Analysis of NotI linking clones isolated from human chromosome 3 specific libraries. Gene 1999; **239(2)**:259-71.
[16] Block KL, Shou Y, Thorton M, Poncz M. The regulated expression of a TATA-less, platelet-specific gene, alphaIIb. Stem Cells 1996; 14 Suppl **1**:38-47.
[17] Wingender E, Chen X, Hehl R.TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 2000; **28(1)**:316-9.
[18] Cook TB, Gebelein, R Urrutia. Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. Ann N Y Acad Sci 1999; **880**:94-102.
[19] Berg JM. Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites. Proc Natl Acad Sci USA 1992; **89(23)**:11109-10.
[20] Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-reponsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes and Dev 1989; **3**:612-9.
[21] Pellegrini M. Computational methods for protein function analysis. Curr Opin Chem Biol 2001; **5(1)**:46-50.
[22] Creighton DJ, Hamilton DS. Brief history of glyoxalase I and what we have learned about metal ion-dependent, enzyme-catalyzed isomerizations. Arch Biochem Biophys 2001; **387(1)**:1-10.
[23] Thornalley PJ. Glutathione-dependent detoxification of alpha-oxoaldehydes by the glyoxalase system: involvement in disease mechanisms and antiproliferative activity of glyoxalase I inhibitors. Chem Biol Interact 1998; **111-112**:137-51.