ZNF232: structure and expression analysis of a novel human C2H2 zinc finger gene, member of the SCAN/LeR domain subfamily

Lampros A. Mavrogiannis a,2, Alexandros Argyrokastritis a, Nicholas Tzitzikas a, Emmanuel Dermitzakis a,3, Theologia Sarafidou a, Philippos C. Patsalis b, Nicholas K. Moschonas a,c,*

a Department of Biology, University of Crete, P.O. Box 2208, 71409 Heraklion, Greece
b Cyprus Institute of Neurology and Genetics, P.O. Box 3462, 1683 Nicosia, Cyprus
c Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, P.O. Box 1527, 71110 Heraklion, Greece

Received 26 October 2000; received in revised form 18 January 2001; accepted 18 January 2001

Abstract

We have identified a novel zinc finger gene, ZNF232, mapped to human chromosome 17p12. The coding region of the gene is organized in three exons corresponding to a 417 amino acid long polypeptide containing a SCAN/LeR domain and five C2H2-type zinc fingers. ZNF232 is possibly a nuclear protein, as suggested by expression analysis of GFP/ZNF232 chimeric constructs.

Keywords: Zinc finger protein; SCAN/LeR domain; Human erythroleukemia K562 cell; Normalized cDNA library; Chromosome 17; Nucleus

During the past decade, partial cDNA sequences (expressed sequence tags, ESTs) corresponding to novel human genes, obtained through systematic library screening, have been rapidly accumulated and exploited for the construction of an integrated framework genome map [1,2]. In addition to ordinary cDNA libraries, screening of normalized libraries greatly facilitated the identification of relatively rare transcripts and significantly contributed to the enrichment of the EST database [3]. Currently, determination of novel full-length cDNAs and sequence comparisons with already characterized genes frequently reveal well-known protein motifs, thus enabling essential gene annotation, i.e. classification of novel genes to various gene/protein families [4,5] and incorporation of analyzed genes to the emerging high resolution human genomic map.

Human erythroleukemia K562 cells, considered as pluripotent hematopoietic progenitors, have been proven to be useful in studying regulated gene expression. Numerous studies on important biological functions such as developmental stage-specific gene switching, signal transduction and transcriptional regulation have been based on the analysis of genes active in K562 cells [6–9]. In order to systematically search for novel genes expressed in these cells, we generated a normalized cDNA library [10] and isolated a number of partial cDNA clones. Among several ESTs of unknown nature, EST HSIMBB246 (GenBank accession No. X93860), displaying similarities to genes coding for C2H2-type zinc finger proteins, was considered for further analysis. In order to obtain a full-length cDNA and avoid possible sequence ambiguities due to normalization, we screened the corresponding native K562 library [10]. A 114 bp 32P-labeled PCR product, synthesized by using a pair of EST-deduced primers (5'-AGATGGGGTGCTCATCTTG-3' and 5'-CCGATGCTGACTTAGATGTATG-3', nucleotide (nt) positions 1149–1167 and 1243–1262, respectively), was applied as a probe. PCR conditions included 35 cycles with denaturation at 94°C (1 min), annealing at 57°C (1 min) and extension at 72°C (2 min).

* Corresponding author, at address c. Fax: +30-81-394-408; E-mail: moschon@imbb.forth.gr

1 The cDNA sequence of ZNF232 (name approved by the HUGO/GDB Nomenclature Committee) has been submitted to GenBank and assigned accession No. Y15067; the corresponding genomic DNA sequences are available under accession Nos. AF080169–171.

2 Present address: Clinical Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK.

3 Present address: Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, PA 16802, USA.
Five positive clones with insert sizes ranging from 1.1 to 1.35 kb were isolated, subcloned into pBluescript II KS+ (Stratagene), sequenced (Sequenase V.2 kit, Amersham) and evaluated in silico. A 1473 bp long cDNA corresponding to a novel human zinc finger gene, designated \textit{ZNF232}, was assembled (Fig. 1). An internal 27 bp long coding sequence, nt positions 670–696, corresponding to nine amino acid (aa) residues, was absent in two of the analyzed clones, possibly reflecting an alternative splicing event (Figs. 1 and 2).

\textit{ZNF232} cDNA and genomic DNA (a \textit{ZNF232}-specific PAC No. 202L17, see below) sequencing revealed the presence of three potential translation start codons at the 5' end of the gene. AUG at nt position 126 (Fig. 1) was considered the most suitable, as being in optimum accordance to the Kozak rule \cite{11}, whereas the others, located 81 (Fig. 1) and 390 bp upstream from it, were considered less possible. In addition, an in frame translation stop codon was determined 504 bp upstream from the putative translation start codon (not shown). Moreover, our sequence data combined with recent data from GenBank (No. AC074339) and EST database (EST No. AI344299; EST No. AW952625), suggested a \textit{ZNF232} transcription unit of at least 2.0 kb with a proposed 5' untranslated region of about 0.5 kb. The actual size of the transcript was confirmed by Northern analysis (see below). The putative 417 aa long polypeptide has a predicted molecular mass of 47.6 kDa and may be divided into two distinct regions (Figs. 1 and 2A): the amino terminal, where a finger associated motif, the \(\alpha\)-helical leucine-rich SCAN/LeR domain \cite{12,13} is recognized (aa 49–130), and the carboxy terminal zinc finger region, characterized by five consecutive repeats of the C2H2 type (aa 278–410). A putative nucleus localization signal (NLS; aa sequence: -RHRR-) \cite{14} is embedded in the fifth repeat (Fig. 1).

The C2H2 zinc finger protein family exemplified by the \textit{Drosophila melanogaster} Krüppel gene product \cite{15}, the \textit{Xenopus laevis} TFIIIA \cite{16,17} and the human transcription factor Sp1 \cite{18}, constitutes one of the largest protein families in the mammalian genomes \cite{19}. The \textit{ZNF232} zinc finger motifs are closely related; they fit the Kruppel-type consensus \cite{20} and are joined by the conserved S/TGEKPY/FX linker peptide. Application of the proposed DNA-zinc finger recognition rules \cite{21} suggests that all \textit{ZNF232} fingers, and especially the second, the fourth and the fifth, can, in principle, participate in optimal DNA binding. The SCAN/LeR domain placed about 50 aa downstream from the amino terminus, lies in a region which is conserved in the majority of the members belonging to the respective subfamily. This domain, initially identified in human p18 \cite{12} and \textit{ZNF174} \cite{13} proteins, is shared by a rapidly growing number of mammalian C2H2 zinc finger proteins. Although \textit{ZNF232} appears to be overall neutral (deduced pI = 6.73), the non-finger portion is markedly acidic (pI = 4.85; aspartate and glutamate content: 17.3%), reminiscent of several transcription factors.
However, transient expression experiments in COS-7 cells, designed to test the possible influence of a GAL4-ZNF232 chimeric protein on the transcriptional activity of a chloramphenicol acetyltransferase (CAT) reporter gene driven by a thymidine kinase (tk) promoter containing a GAL4 binding site [23], did not result in any significant change of CAT activity (data not shown). Since the presence of the zinc fingers may interfere with chimeric protein folding, thus making a potentially ZNF232 activation domain inaccessible to the transcriptional machinery, we also designed a zinc finger deletion construct, by omitting 614 bp from the 3' end of ZNF232 cDNA. Similarly, no significant change in CAT activity was observed. Moreover, transfection in mouse NIH-3T3 cells, under optimized conditions [13], also demonstrated no significant effects. Our results are in agreement with previous observations [12,13] and may indicate that ZNF232 is a weak transcriptional regulator or that the SCAN/LeR domain cannot act on a heterologous DNA binding site [12,13], suggesting that it is not an independent transcriptional regulatory element. Interestingly, a human gene encoding an isolated SCAN/LeR domain was reported recently [24]. Although the function of the domain is still unknown in depth, it was demonstrated that SCAN/LeR mediates hetero- and/or homotypic protein associations and thus modulates the function of transcriptional factors [24–26].

In order to investigate whether ZNF232 exists in a single or multiple copies in the human genome, we performed genomic DNA blot hybridization analysis. The simple hybridization pattern detected in several restriction digests (data not shown) is consistent with the existence of a unique ZNF232 gene. To elucidate the exon/intron organization of ZNF232, we screened the gridded RPCI-1 human genomic PAC library (a generous gift of UK HGMP-Resource Centre, Hinxton, UK). This was performed by probing with a 32P-labeled 730 bp EcoRI fragment derived from the 5' end of a ZNF232 cDNA clone spanning nt positions 370–1100. A positive clone, No. 202L17, was isolated. Structural integrity of the clone was confirmed by DNA blot hybridization analysis in parallel with a human genomic DNA sample (data not shown). Subcloning of appropriate PAC clone fragments, sequencing, and PCR using a series of ZNF232 cDNA-specific primers, enabled us to elucidate the genomic organization of the gene (Fig. 2). Our analysis suggested that ZNF232 encompasses a genomic region of at least 6 kb and the translated portion of the gene spans three exons, encoding 139, 42 and 236 amino acids, respectively (Figs. 1 and 2). The putative 5' UTR of the gene is interrupted by an intron

---

**Fig. 2.** Exon/intron organization and genomic sequence of ZNF232. (A) Schematic illustration of ZNF232 exon/intron organization. Boxes represent ZNF232 exons. Relative exon sizes are on scale, while the approximate intron sizes are indicated in kb. The SCAN/LeR and the zinc finger domains as well as the alternatively spliced part of exon 4 are depicted by filled boxes. The translation start and the stop codon are marked on exon 2 and 4, respectively. (B) Genomic sequences flanking ZNF232 exons. Exons are indicated by uppercase letters; lowercase letters denote intron sequences. The splice sites, all complying with the GT/AG consensus, and the alternative acceptor site in exon 4, are in bold. Intron sizes were estimated by either PCR using primer pairs corresponding to exon sequences (introns 2 and 3), or by Southern analysis exploiting appropriate cDNA fragments as probes (intron 1). Intron 2 was sequenced to its entity.
of about 2.0 kb. The sequence encoding for the SCAN/LeR domain covers the second half of exon 2, while all five zinc fingers are clustered in exon 4. An AG dinucleotide located 27 bp downstream from the 5' end of exon 4 (nt position 696, Figs. 1 and 2B), may represent an appropriate alternative acceptor site, able to be utilized interchangeably by the splicing machinery in vivo. This observation correlates perfectly with the absence of a 27 bp long coding sequence (nt positions 670–696) determined in two of the analyzed ZNF232 cDNA clones.

RNA blot hybridization analysis using 2–3 μg of poly(A)+ RNA from K562 cells, detected a single ZNF232-specific transcript of about 2.1 kb (data not shown). This is comparable with the size of the suggested ZNF232 overall cDNA sequence. Furthermore, in order to explore the ZNF232 expression profile in various human tissues and simultaneously investigate the possibility of ZNF232 in vivo alternative splicing, we employed a multiple tissue panel of normalized first-strand cDNAs (MTC, Clontech) in PCR experiments using a pair of primers (5'-ACTGTGCTGGAGGATTTAGAG-3' and 5'-ATGAACTCTCTGGTGGACAAC-3') flanking cDNA nt positions 670–696. PCR performed under non-saturating conditions (Clontech MTC manual) permitted semiquantitative estimation of relative ZNF232 transcript levels. The gene is expressed in all tissues tested (Fig. 3); however, slightly higher expression was observed in liver, testis and ovary. Furthermore, two transcript variants with a size difference comparable to the PCR products obtained by two ZNF232 cDNA templates differing internally by 27 bp at nucleotide positions 670–696, were detected in all samples (Fig. 3). This suggests that ZNF232 is subjected to alternative splicing in vivo, a feature shared by other members of the SCAN/LeR subfamily [27–29].

Fig. 3. Expression pattern of ZNF232 in human tissues. (A) Each of the indicated MTC panel cDNA was used as a PCR template with a ZNF232-specific primer pair (see text). After 35 amplification cycles (annealing at 55°C), samples were subjected to 2.2% agarose/EtBr electrophoresis. Lanes: 1–8, heart; brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, respectively; 9–16, spleen, thymus, prostate, ovary, small intestine, colon, leukocytes, respectively; a, PCR product from a ZNF232 cDNA containing the 27 bp region at nt positions 670–696; b, PCR product from a ZNF232 cDNA lacking the 27 bp region at nt positions 670–696; a/b, co-electrophoresis of a mixture of the PCR products of lanes a and b. (B) The same as above cDNA samples (lanes 1–8 and 9–16) used as PCR templates with a control glyceraldehyde-3-phosphate dehydrogenase gene (G3PDH) primer pair (Clontech) after 27 amplification cycles (annealing at 55°C). Lanes M (both A and B panels), pBluescript/HindIII size marker.

Fig. 4. Fluorescence in situ hybridization on human male metaphase spreads using PAC clone RPCI-1 No. 202L17 as a probe. The unique signal at 17p12 is denoted by an arrow.
Interestingly, the relative abundance of the two splice variants differed among tested samples (Fig. 3), indicating a tissue-specific post-transcriptional control for ZNF232. It remains to be seen whether these alternatively spliced transcripts suggesting two ZNF232 isoforms, differing internally by nine amino acids, are of functional significance.

Chromosomal assignment of ZNF232 was initially performed using a monochromosomal human-rodent somatic cell hybrid panel (Coriell Institute of Medical Research). DNA samples (50–250 ng) from each hybrid and comparable amounts from the respective human and rodent parental cell lines were used as templates in PCR amplifications with a ZNF232-specific primer pair. The experiment suggested that ZNF232 maps to chromosome 17 (data not shown). Subsequent FISH analysis with fluorescence labeled PAC No. 202L17 as a probe, performed to normal metaphase chromosome spreads according to standard high resolution procedures [30,31], resulted in a strong hybridization signal at 17p12 (Fig. 4). Database searches revealed that a partially sequenced BAC clone (GenBank accession No. AC012146) roughly mapped to chromosome 17, encompasses ZNF232. Our FISH data place this BAC to 17p12. The chromosomal distribution of human genes encoding for zinc finger proteins is clearly non-random [32,33]. A disproportionate density characterizes chromosomes 1, 3, 6, 7, 10, 11, 22 and an excessive abundance marks chromosome 19 (Genome Database; www.gdb.org). Moreover, available mapping data for members of the human SCAN/LeR group reveal an apparently clustered genomic organization. The majority of the reported members are mapped to certain chromosomal regions, i.e. 3p21-22, 6p21-22 and 16p13 [34–37]. Apart from ZNF232, at least four other zinc finger genes of the C2H2 type map to 17p12-p13, i.e. ZNF18/KOX11, ZNF29/KOX26, ZNF62 and ZFP3, thus pinpointing another zinc finger gene-rich region in the human genome. Structural abnormalities of C2H2 zinc finger genes have been found in a variety of genetic diseases. In particular, deletions at 17p12-13 have been associated with gliomas, astrocytomas, colon, and lung tumors [32,38].

Finally, in order to investigate experimentally the subcellular localization of the ZNF232 polypeptide and the functional role of the predicted NLS, we proceeded in transient expression of a green fluorescent protein-ZNF232 chimera, in African green monkey COS-7 cells. The construct, pEGFP/ZNF232, contained the enhanced green fluorescent protein (GFP) open reading frame combined with the ZNF232 coding region. This was prepared by co-ligation of the linearized pEGFP-C1 vector (Clontech) with a PCR product (primers: 5′-AGGATGGCCTG-TATCATAAC-3′ and 5′-GATCCAGTCTCAAAGTAGA-GATTAGAC-3′, nt positions 123–142 and 1418–1444, respectively) representing the entire ZNF232 translated region; the amplified DNA was inserted into the blunt-ended HindIII site of the vector polylinker. A deletion derivative, pEGFP/delZNF232, lacking the NLS region was prepared by excision of a KpnI fragment defined by the unique internal (nt position 830, Fig. 1) and the vector KpnI sites (expressed ZNF232 polypeptide: 1–235 aa). About 10⁶ COS-7 cells were transfected with 2 μg of either pEGFP/ZNF232 or pEGFP/delZNF232 DNA using the calcium phosphate co-precipitation method [39]. Twenty-four to 36 h after transfection, the cells were examined with fluorescence microscopy. The chimeric GFP/ZNF232 polypeptide was detected specifically in the nucleus; on the contrary, expression of pEGFP/delZNF232 resulted in the exclusion of the truncated polypeptide from the nucleus (Fig. 5). These findings suggest that ZNF232 is a nuclear protein and demonstrate the role of the predicted NLS.

To summarize, this study reports the detailed analysis of a novel chromosome 17p12 zinc finger gene, ZNF232, member of the SCAN/LeR subfamily. ZNF232 is widely expressed in man, encodes a nuclear protein and is subject to tissue-specific post-transcriptional regulation. The latter may lead to two structural variants differing in relative abundance among tissues. Like other members of the group, ZNF232 may be associated with the transcriptional machinery in a wide variety of human tissues. It would be interesting to investigate the potential of ZNF232 as a transcriptional modulator as well as to explore its interaction repertoire.

---

Fig. 5. Fluorescence photomicrographs showing the subcellular localization of GFP-ZNF232 chimeras in transiently transfected COS-7 cells. Plasmid DNA used: (A) pEGFP-C1 vector (control); the GFP product exhibited passive nucleocyttoplasmic diffusion, as expected [40]; (B) pEGFP/ZNF232; (C) pEGFP/delZNF232.
The authors wish to thank S. Vrontou for her contribution in the initial screening of the cDNA library, A. Parasparaki for technical support and Dr. G. Chalepakis for critical reading of the manuscript. This work was supported by grants of the University of Crete and IMBB-FORTH.

References