



ELSEVIER

Gene 255 (2000) 195–203

GENE

AN INTERNATIONAL JOURNAL ON
GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Double FYVE-containing protein 1 (DFCP1): isolation, cloning and characterization of a novel FYVE finger protein from a human bone marrow cDNA library

A.R. Derubeis^a, M.F. Young^a, L. Jia^b, P.G. Robey^{a,*}, L.W. Fisher^a

^a *Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA*

^b *Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA*

Received 26 April 2000; accepted 3 July 2000

Received by D. Schlessinger

Abstract

Double FYVE-containing protein 1 (DFCP1) encodes a 777 amino acid protein that contains: (1) an N-terminal Cys–His cluster with some homology to many zinc finger domains; (2) a consensus sequence consistent with an ATP/GTP binding site; and (3) a C-terminal domain unique because it contains two zinc-binding FYVE domains. The gene, *ZNFN2A1* (GenBank accession no. AF251025) was localized to chromosome 14q22–q24 and shown to be composed of 11 exons. Northern blot analysis revealed the presence of three different mRNA transcripts (4.2, 3 and 1.2 kb). The two longer transcripts appear to be expressed in a variety of different tissues, especially in endocrine tissues, while the shorter messenger is limited to testis. Both of the larger transcripts are unusual due to the presence of a 463 bp long 5' UTR. Furthermore, the 4.2 kb transcript contains a non-standard polyadenylation consensus sequence while the 3 kb transcript contains a standard consensus sequence but within the open reading frame. Following in vitro transfection of a DFCP1-containing expression construct, confocal microscopy studies showed a vesicular distribution of DFCP1 suggesting that this protein, like other FYVE-containing proteins, might be involved in membrane trafficking. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Bone marrow stroma; Expressed sequence tag; Vesicles; Zinc finger

1. Introduction

Bone marrow stromal cells (BMSC) are non-hematopoietic cells residing in the marrow cavity. While they exhibit a number of fibroblastic features, they are distinct from fibroblastic cells found in other connective tissues. Furthermore, they lack characteristics of endothelial cells and macrophages (Castro-Malaspina et al., 1980; Wang et al., 1990). After extensive proliferation in vitro, the BMSC population retains the capacity to differenti-

ate into at least four types of connective tissue: bone, cartilage, myelosupportive stroma, and associated adipocytes (Friedenstein et al., 1974; Owen and Friedenstein, 1988).

High throughput sequencing of a human BMSC cDNA library led to the identification of many previously reported genes, as well as a number of unknown expressed sequence tags (ESTs). We report here for the first time the cloning and characterization of a new protein containing two FYVE finger domains that we have designated DFCP1 (double FYVE-containing protein 1).

The function of most FYVE-containing proteins is still unknown, however, FYVE domains have been shown to interact specifically with the membrane lipid, phosphatidyl inositol-3-phosphate (PtdIns-3-P) (Stenmark et al., 1996; Burd and Emr, 1998; Gaullier et al., 1998; Patki et al., 1998). Several of the FYVE-containing proteins have been shown to be recruited to

Abbreviations: BMSC, bone marrow stromal cells; DFCP1, double FYVE-containing protein 1; EST, expressed sequence tag; FYVE, Fab1p/YOTB/Vac1p/EEA1; PCR, polymerase chain reaction; PtdIns-3-P, phosphatidyl inositol-3-phosphate; STS, sequence tag site; UTR, untranslated region; ZNFN2A1, zinc finger protein, subfamily 2A, protein 1.

* Corresponding author. Tel.: +1-301-496-4563; fax: +1-301-402-0824.

E-mail address: probey@dir.nidcr.nih.gov (P.G. Robey)

vesicular structures and to participate in intracellular membrane trafficking (Simonsen et al., 1998; Bean et al., 1997; Weisman and Wickner, 1992; Piper et al., 1995; Yamamoto et al., 1995). We found this novel protein to be widely expressed, particularly in endocrine tissues. Upon *in vitro* transfection using DFCEP1 expression constructs, DFCEP1 did in fact localize to vesicular structures within the cytoplasm, similar to what has been reported for other FYVE-containing proteins.

2. Materials and methods

2.1. cDNA library construction

Human bone marrow stromal cells from three different donors [32 year old (yo) black female, 35 yo black male, 43 yo white male] were established according to the method published previously (Kuznetsov et al., 1997). Extracted RNA was oligo dT purified and then submitted to Stratagene (La Jolla, CA) for custom library construction into the Uni-Zap system (Dieudonne et al., 1999).

2.2. Skeletal Genome Anatomy Project (SGAP)

The SGAP is a project under development through the collaboration of the Medical Genetics Branch of the National Human Genome Research Institute (NHGRI), the Center for Information and Technology (CIT) and the Craniofacial and Skeletal Diseases Branch at the National Institute of Dental and Craniofacial Research (NIDCR) (Jia et al., 1998). SGAP's goal is to build a catalogue of genes expressed by normal and abnormal bone, cartilage, tendon, ligament and synovial tissues. To date, two cDNA libraries constructed from BMSCs, which contain osteoprogenitors (Kuznetsov et al., 1997), and more mature trabecular bone cells (TBCs) (Robey and Termine, 1985) have been sequenced.

2.3. DNA sequencing

Individual clones from the BMSC library were chosen and the 3' end was sequenced by the NIH Intramural Sequencing Center (NISC). The 3' ESTs were deposited in the National Center for Biotechnologies (NCBI) dbEST database. To date, approximately 10 000 clones have been sequenced. An EST from the BMSC library (GenBank accession no. AA545744) was selected and sequence verified and the clone was sequenced in full. Sequences were determined at least four times with universal T3, T7 or cDNA specific primers, by ABI 373 and 377XL Automated DNA sequencers. The Basic Logic Alignment Search Tool (BLAST) at the NCBI server (<http://www.ncbi.nlm.nih.gov/BLAST/>) was used

to search the GenBank database for related sequences (Altschul et al., 1990). DNA and protein sequence comparisons were carried out with the GCG software and domain searches were performed through the Prosite database (<http://www.isrec.isb-sib.ch/>). The genomic sequence was identified from GenBank using the DFCEP1 encoding cDNA sequence as query in the High Throughput Genomic Sequences Database (HTGS), and has been assigned the designation *ZNFN2A1*.

2.4. Chromosomal localization

The NCBI dbSTS (Sequence Tag Sites database) (<http://www.ncbi.nlm.nih.gov/BLAST/>) of GenBank was queried using the genomic sequence, and this led to the identification of one sequence tag site (STS). The chromosomal position of the STS marker had been localized previously by radiation hybrid mapping, and the mapping information was available in the public database.

2.5. Northern blot analysis

Total RNA was extracted using RNA STAT (Tel-Test, Inc., Friendswood, TX) as directed by the manufacturer. Poly A⁺ RNA was prepared from total RNA using a Micro Poly (A) Pure kit (AMBION Inc., Austin, TX). Poly A⁺ RNA blots (2 µg/lane) of human tissues were purchased from Clontech, Palo Alto, CA. Northern blots were hybridized with a 209 bp (α -³²P)CTP random prime labeled cDNA fragment (derived from PCR amplification of DFCEP1 encoding cDNA using 5'-cagtgccgggtctgtgacaact-3' at position 2367 bp as the forward primer and 5'-tggtcaggcaccagtagc-3' at position 2556 bp as the reverse primer). A β -actin cDNA (Clontech, Palo Alto, CA) probe was used as a control. The blots were washed free of non-specifically bound probe, and radioactivity was visualized and quantified using a PhosphoImager and ImageQuant Software from Molecular Dynamics (Eugene, OR).

2.6. Fusion protein

For expression in mammalian cells, the coding sequence (428–2761 bp) was amplified using specific primers with built in restriction sites (BamHI 5'-agaataggatcccatgagtgcccagactccc-3' forward; EcoRI 5'-ttctggaattcttaaaggtcaccgggctttt-3' reverse) and subsequently placed in frame behind the 9E10 myc sequence under the CMV promoter of pCMV-tag3b vector (Stratagene, La Jolla, CA). The integrity of the construct was then verified by DNA sequencing.

2.7. Transfections

NIH-3T3 cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) containing 10% fetal bovine serum on cover slips in six well plates. When the cells reached a confluence of 70% they were transfected with the DFCP1–c-myc fusion protein (1 µg DNA/well) with Lipofectamine Plus Reagent (Life Technologies, MD) and incubated in serum-free medium for 4 h following the manufacturer's directions. After 4 h, the serum-free media was replaced with serum-containing DMEM and the cells were allowed to recover at 37°C for 24 h.

2.8. Immunolocalization and confocal microscopy

Following 24 h of incubation, transfected NIH-3T3 cells were washed thoroughly with phosphate-buffered saline (PBS) and fixed for 20 min at room temperature (RT) in 2% phosphate-buffered formalin freshly prepared from paraformaldehyde. After extensive washing with PBS 1 ×, cells were incubated in 0.1 M glycine/PBS (pH 7.4) for 15 min at RT. Permeabilization of the cells was achieved by incubating the cells in ice-cold methanol for 5 min at –20°C. Blocking in 0.5% bovine serum albumin in PBS for 15 min at RT was followed by incubation of the cells with a 1:100 dilution of the primary antibody, monoclonal mouse anti-c-myc (Boehringer Mannheim) for 1 h at RT. After extensive washing, cells were incubated with a 1:40 dilution of the secondary antibody [donkey anti-mouse IgG F(ab')₂ fragment FITC-conjugated, Jackson Laboratories, Inc., PA] for 1 h at RT in 0.5% BSA/PBS in the dark. After washing, immunolocalization of DFCP1 fusion protein was detected by confocal microscopy. Images were captured on a Leica TCS-4D confocal laser scanning microscope system (Leica Microsystems, Heidelberg, Germany) equipped with a Kr/Ar laser. Z-series images were acquired using a 100 ×, 1.4 NA objective with conditions constant for both control and experimental samples.

3. Results and discussion

3.1. Structural characterization and chromosomal localization

At the time when the SGAP was initiated, sequencing of the BMSC cDNA library produced approximately 50 3' ESTs which appeared to be unique to the library based on BLAST search of GenBank. One of these ESTs was selected for further characterization. The cDNA clone (2761 bp) from which this EST was derived was sequenced in full and further analysis revealed that it corresponded to a novel transcript (Fig. 1A). The open reading frame (ORF) encodes a protein of 777

amino acids with a predicted molecular mass of 87 230 Da and pI of 7.46. Plots of hydrophilicity (Kyte and Doolittle, 1982), surface probability (Janin and Wodak, 1978; Emini et al., 1985), and antigenic index (Jameson and Wolf, 1988) predict a largely hydrophilic structure, suggesting that the protein is not an integral membrane protein. Submission of the deduced amino acid sequence to the SignalP database (<http://www.cbs.dtu.dk/services/SignalP/>) of the Center of Biological Sequence Analysis (CBS) showed that no known nuclear localization signal peptide for transport of the protein into the nuclei, or signal peptide for secretion from the cell, are present, assigning this protein as primarily a cytoplasmic protein.

The deduced amino acid sequence of the cDNA reveals that the protein has many interesting features. It contains a putative zinc finger domain (a Cys- and His-rich region) at the NH₂-terminus (dashed box, Fig. 1A). Comparison of this region with other zinc finger sequences in the database revealed a loose correlation with different types of zinc finger domain (B-box: aa 13–63 and aa 66–121, C₂H₂: aa 40–69). No perfect matches were found, suggesting that it represents either a novel type of zinc finger domain or is simply a Cys- and His-rich cluster. The protein has a glycine-rich region usually referred to as the P-loop or the 'A' consensus sequence (underlined with a solid line, Fig. 1A) that has been shown to be very conserved in a large number of proteins that bind ATP or GTP.

Further analysis of the deduced amino acid sequence revealed that the protein contains two putative double zinc finger domains, termed the FYVE finger, hence we have named this protein 'double FYVE-containing protein 1' (DFCP1). While zinc finger proteins are usually involved in protein–DNA interactions, a number of studies have shown that certain zinc fingers can be involved in protein–protein interactions (Freemont, 1993; Feuerstein et al., 1994) and even in the binding of lipids (Zhang et al., 1995). The FYVE domain was first characterized by Stenmark et al. (1996), the name being derived from the first letter of the first four proteins discovered to contain it: Fab1p, YOTB, Vac1p, EEA1. The amino acid sequence of the FYVE domain is highly conserved and characterizes a family of proteins with more than 30 members found in species ranging from yeast to mammals (Gauillier et al., 1999).

In DFCP1, both FYVE domains are approximately 70 amino acids long (solid lined box, Fig. 1A) and share 53% amino acid identity and 56% similarity (Fig. 2A, first two rows). The two FYVE domains are approximately 40% identical and 53% similar to FYVE domains of other proteins (Fig. 2A). The two FYVE domains of DFCP1 show the greatest degree of homology to human and yeast FYVE domains. The lack of homology to other mammals may be due to the lack of information on FYVE-containing proteins in rat and mouse. The

A

1 TGA AAT ACT GAC TTC AGG TCG AAT TAT ATT GAA AAG CTC CTG ACC ACT TTC TTT CAT TAC CAA AAC TTT GTA GCT GAT GTC CAA CCG ATG 90
91 AAC CCA CCA CCG TGA ACC CAT CAG ACC TCT CTC AGA TAG CCA TAA AAG ACC CTT CCA AGT CAA TTT TGA CCA CAT CTT TGC TTG CAC TTT 180
181 ATG GAG GAT GAA ACC ATC AAA CCA AAT CAA CGT TGC TGC TAA TAC AAG AGT CTT AGA GGC AGC AAA TTA AAA ATT TGA ACA TTT GTT TGT 270
271 GAA GAA CTA TAA CAG GAC ATG AAA GGT GTT CTT TTT TAA AGT GTT CAG AAC CCT GTG GAA GTT TCG TGC AGT CTT CAG ACT CAA ATT TTC 360
361 GTC TTC ACC CCC GGG GCA AGC TCA GTG ACT ATT ATA TGG TGG GTG TGT TTC CTT ACC AGC GTG AGT ATG AGT GCC CAG ACT TCC CCA GCA 450
Met Ser Ala Gln Thr Ser Pro Ala 8
451 GAG AAG GGC CTG AAT CCG GGG CTG ATG TGC CAG GAA AGT TAC GCT TGC AGC GGG ACT GAT GAA GCT ATC TTT GAG TGT GAT GAG TGC TGC 540
9 Glu Lys Gly Leu Asn Pro Gly Leu Met Cys Gln Glu Ser Tyr Ala Cys Ser Gly Thr Asp Glu Ala Ile Phe Leu Cys Asp Glu Cys Cys 38
541 AGT CTG CAG TGT CTC CGC TGC GAG GAG GAG CTC CAT CGG CAG GAG CGC CTG AGA AAC CAT GAG CGG ATA AGA CTC AAA CCT GGC CAT GTC 630
39 Ser Leu Gln Cys Leu Arg Cys Glu Glu Glu Leu His Arg Gln Glu Arg Leu Arg Asn His Glu Arg Ile Arg Leu Lys Pro Gly His Val 68
631 CCT TAC TGT GAC CTC TGC AAG GGT CTC AGT GGG CAT TTA CCA GGT GTT AGG CAG AGG GCA ATA GTG AGG TGC CAG ACC TGC AAA ATT AAC 720
69 Pro Tyr Cys Asp Leu Cys Lys Gly Leu Ser His Leu Asn His Thr Phe Thr Tyr Gly Val Arg Gln Arg Ala Ile Val Arg Cys Gln Thr Cys Lys Ile Asn 98
721 TTG TGC CTG GAG TGC CAG AAG AGG ACT CAT TCT GGG GGT AAC AAA AGG AGA CAC CCT GTT ACT GTG TAC AAT GTC AGT AAT CTC CAG GAG 810
99 Leu Cys Leu Glu Cys Gln Lys Arg Thr His Ser Gly Gly Asn Lys Arg Arg His Pro Val Thr Val Tyr Asn Val Ser Asn Leu Gln Glu 128
811 TCA CTG GAG GCA GAA GAG ATG GAT GAG GAG ACC AAG AGG AAG AAG ATG ACT GAG AAG GTT GTG AGT TTC CTC CTA GTA GAC GAA AAT GAA 900
129 Ser Leu Asp Glu Glu Ser His Thr Leu Ser Gly His Leu Pro Phe Tyr Val Arg Gln Arg Ala Ile Val Ser Phe Leu Val Asp Glu Ser Cys Glu 158
901 GAA ATT CAG GTA ACA AAT GAA GAA GAC TTT ATT AGA AAA TTG GAC TGC AAA CCT GAT CAG CAT CTG AAA GTG GTT TCC ATT TTT GGA AAT 990
159 Glu Ile Gln Val Thr Asn Glu Glu Asp Phe Ile Arg Lys Leu Asp Cys Lys Pro Asp Gln His Leu Lys Val Val Ser Ile Phe Gly Asn 188
991 ACT GGT GAT GGA AAG TCT CAT ACT CTC AAC CAC ACT TTC TTT TAT GGT CGT GAA GTC TTT AAA ACC TCC CCG ACC CAG GAG TCC TGC ACT 1080
189 Thr Gly Asp Gly Leu Ser His Thr Ser Phe Lys Thr Ser Pro Thr Phe Lys Thr Ser Pro Thr Phe Lys Thr Ser Pro Thr Phe Lys Thr Ser Thr 218
1081 GTG GGA GTG TGG GCA GCC TAT GAC CCA GTT CAC AAA GTA GCA GTG ATC GAT ACG GAA GGG CTC CTG GGG GCC ACC GTG AAT CTA AGC CAG 1170
219 Val Gly Val Trp Ala Ala Tyr Asp Pro Val His Lys Val Ala Val Ile Asp Thr Glu Gly Leu Leu Gly Ala Thr Val Asn Leu Ser Gln 248
1171 AGA ACA CGG CTG CTG CTT AAG GTC CTG GCC ATC TCA CAG CTC GTC ATC TAT CGA ACT CAT GCA GAC CGG CTG CAT AAC GAC CTC TTC AAA 1260
249 Arg Thr Arg Cys Leu Leu Lys Val Leu Ala Ile Ser Asp Leu Val Ile Tyr Arg Thr His Ala Asp Arg Leu His Asn Asp Arg Phe Lys 278
1261 TTC CTT GGG GAT GCC TCA GAA GCT TAT CTG AAG CAC TTC ACC AAG GAG CTC AAG GCC ACC ACT GGT CGC TGT GGC CTG GAT GTC CCT TTA 1350
279 Phe Leu Gly Asp Ala Ser Glu Ala Tyr Leu Lys His Phe Thr Lys Glu Leu Lys Ala Thr Thr Ala Arg Cys Gly Leu Asp Glu Pro Leu 308
1351 TCC ACA CTG GGC CCT GCA GTT ATC ATC Ile Phe His Glu Thr Val His Thr Gln Leu Leu Gly Ser Asp His Pro Ser Glu Val Pro Glu Lys 1440
309 Ser Thr Leu Gly Pro Ala Val Ile Ile Phe His Glu Thr Val His Thr Gln Leu Leu Gly Ser Asp His Pro Ser Glu Val Pro Glu Lys 338
1441 CTC ATC CAG GAC CGG TTC CGG AAG CTG GGC CGT TTC CCT GAA GCC TTT AGT TCC ATT CAC TAC AAG GGA ACG AGG ACT TAC ACG CCT CCC 1530
339 Leu Ile Gln Asp Arg Phe Arg Lys Leu Gly Arg Phe Pro Glu Ala Phe Ser Ser Ile His Tyr Lys Gly Thr Arg Thr Tyr Asn Pro Pro 368
1531 ACG GAC TTT TCT GGG CTT CGG CGT GCT TTG GAG CAG CTA CTA GAG AAT AAC ACC CGT TCT CCC CGG CAC CGG GGA GTC ATC TTC AAA 1620
369 Thr Asp Phe Ser Gly Leu Arg Arg Ala Leu Glu Gln Leu Leu Glu Asn Asn Thr Thr Arg Ser Pro Arg His Pro Gly Val Ile Phe Lys 398
1621 GCC CTG AAG GCA CTA AGT GAC CGC TTC AGC GGT GAG ATC CCC GAT GAC CAG ATG GCG CAC AGC TCC TTT TTT CCA GAT GAG TAT TTC ACC 1710
399 Ala Leu Lys Ala Leu Ser Asp Arg Phe Ser Gly Glu Ile Pro Asp Asp Gln Met Ala His Ser Ser Phe Pro Asp Glu Tyr Phe Thr 428
1711 TGC TCC TCC TTG TGC CTC AGT TGT GGG GTT GGA TGT AAG AAA AGC ATG AAT CAT GGG AAG GAA GGA GTG CCT CAT GAA GCC AAG AGC CGC 1800
429 Cys Ser Ser Leu Cys Leu Ser Cys Gly Val Gly Cys Lys Lys Ser Met Asn His Gly Lys Glu Gly Val Pro His Glu Ala Lys Ser Arg 458
1801 TGC AGA TAC TCC CAC CAG TAT GAC AAC CGA GTG TAT ACC TGC AAG GCC TGC TAT GAG AGA GGC GAG GAA GTC AGT GTA GTG CCC AAA ACA 1890
459 Cys Arg Tyr Ser His Gln Tyr Ser Asp Asn Arg Val Tyr Thr Cys Lys Ala Cys Tyr Glu Arg Gly Glu Glu Val Ser Val Val Pro Lys Thr 488
1891 TCT GCT TCC ACT GAC TCC CCC TGG ATG GGT CTC GCA AAA TAT GCC TGG TGT GGG TAT GTG ATC GAA TGT CCT AAC TGT GGC GTG GTC TAT 1980
489 Ser Ala Ser Thr Asp Ser Pro Trp Met Gly Leu Ala Lys Tyr Ala Trp Ser Gly Tyr Val Ile Glu Cys Pro Asn Cys Gly Val Val Tyr 518
1981 CGT AGT CGG CAG TAC TGG TTT GGA AAC CAA GAT CCT GTG GAT ACG GTG GTG CGG ACA GAG ATT GTG CAT GTG TGG CCT GGA ACT GAT GGG 2070
519 Arg Ser Arg Gln Tyr Trp Phe Gly Asn Gln Asp Pro Val Asp Thr Val Val Arg Thr Glu Ile Val His Val Trp Pro Gly Thr Asp Gly 548
2071 TTT CTG AAG GAC AAC AAT GGT GCC CAG CGC CTG TTG GAC GGG ATG AAC TTC ATG GCT CAG TCG GTG TCC GAG CTT GAG CTT GCA CCC 2160
549 Phe Leu Lys Asp Asn Asn Asn Ala Ala Gln Arg Leu Leu Asp Gly Met Asn Phe Met Ala Gln Ser Val Ser Glu Leu Ser Leu Gly Pro 578
2161 ACC AAG GGT GTG ACT TCC TGG CTG ACA GAC CAG ATC GCC CCT GCC TAC TGG AGG CCC AAC TCC CAG ATT CTG AGC TGC AAC AAG TGT GGG 2250
579 Thr Lys Ala Val Thr Ser Trp Leu Thr Asp Gln Ile Ala Pro Ala Tyr Trp Arg Pro Asn Ser Gln Ile Leu Ser Cys Asn Lys Cys Ala 608
2251 AGC TCC TTT AAA GAT AAC GAC ACT AAG CAT CAC TGC CGA GCC TGT GGG GAG GGC TTC TGT GAC AGC TGT TCA TCA AAG ACT CGC CCA GTG 2340
609 Thr Ser Phe Lys Asp Asn Asp Thr Lys His His Cys Arg Ala Cys Gly Glu Gly Phe Cys Asp Ser Cys Ser Ser Lys Thr Arg Pro Val 638
2341 CCT GAG CGG GGC TGG GGC CCT GCG CCA GTG CCG GTC TGT GAC AAC TGC TAC GAA GCC AGG AAC GTC CAG TTA GGT GTT ACC GAG GCA CAA 2430
639 Pro Glu Arg Gly Trp Glu Ala Pro Val Arg Val Cys Asp Asn Cys Tyr Glu Ala Arg Asn Val Gln Leu Ala Val Thr Glu Ala Gln 668
2431 GTG GAC GAT GAA GGT GGA ACA CTC ATT GCT CGG AAG GTG GGC GAG GCC GTG AAC ACT CTG GGA GCC GTG GTG ACA GCT ATT GAC ATA 2520
669 Val Asp Asp Glu Gly Gly Thr Leu Ile Ala Arg Lys Val Gly Glu Ala Val Gln Asn Thr Leu Gly Ala Val Val Thr Ala Ile Asp Ile 698
2521 CCA CTA GGT CTG GTA AAG GAC GCG GCC AGG CCT GCG TAC TGG GTG CCT GAC CAC GAA ATC CTC CAC TGC CAC AAC TGC CCG AAG GAG TTC 2610
699 Pro Leu Gly Leu Val Lys Asp Ala Ala Arg Pro Ala Tyr Trp Val Pro Asp His Glu Ile Leu His Cys His Asn Cys Arg Lys Glu Phe 728
2611 AGC ATC AAG CTC TCC AAG CAC CAC TGC CCG GCC TGC GGA CAG GGC TTC TGT GAT GAG TGC TCC CAT GAC CGC CGG GCT GTT CCT TCT CGT 2700
729 Ser Ile Lys Leu Ser Lys His His Cys Arg Ala Cys Gly Gln Gly Phe Cys Asp Glu Cys Ser His Asp Arg Arg Ala Val Pro Ser Arg 758
2701 GGC TGG GAC CAT CCC GTC CGA GTC TGC TTC AAC TGC AAT AAA AAG CCC GGT GAC CTT TAA CCC Cag ccc cct ctc cga gtc ctt cac aat 2790
759 Gly Trp Asp His Pro Val Arg Val Cys Phe Asn Cys Asn Lys Lys Pro Gly Asp Leu Stop 778
2791 tcc tta ggt tct cag ggt tag aaa cag tct tgc gag gta ggc cct cct ccc agt cac ctg ctg tgg tgt gtc tcc tct cct ctc cgc atc 2880
2881 cag ggc cac ttt ccc tca gtc ggg gtc agc ctg gcg gca ggc ccg aag gtc tgg acc cct cag ggc agg gga cct tgc aac tta tgc caa 2970
2971 agg gga atg aac cct aat ccg ttt cat tta ttt cag tta aaa ata atg aat ata tat gtc tat atc tct ctc tca tat ata cat atg aaa 3060
3061 ggc act cgg ggc gta tgc agc ctg ctg gct gtc aag act ctc agc atc ctc cgc aca ggg tga ggt ggc agt ggc agc agt tct 3150
3151 tcc tca tga gcc gag cca ggt cca tgg cca cca cgt ggc tgg ccc ctt cct ctg ctg ctc ttg gag cct tgg aag cct ctc ctg tcc ttg 3240
3241 gct ctt ccc tcc atg cct gtc agc tgc ctg ggg agt gac cct ccc tgg tcc ttc ctg cct gaa aca gcc tga agg gaa ttc tcc cta ggt 3330
3331 ctc ctg gga gtc gag tcc caa ttc ttg gct taa gcc tgt ttt agt cag aga cca ccc aac tta cgc tgc agg tca ccc gag tgg gtc gag 3420
3421 ggt cag agg tgc ggt ctt cgg ccc tga gaa gta gaa atg cag ggc ccg tgc tgt ccc tgg tcc ccc agg gaa cag caa gga agt aac tga 3510
3511 gcc ttc tcc agc agg gct tcc tgt ccc gat gct tgt gtc tcc act cgg cct tcc caa aag gcg gca ccc agc tcc tca atc gaa gca tct 3600
3601 gcc tcc cac ccc tgc gcc ccc tca agc cca cca tct gct tct gag tgc agc act agt att ttc att gct tat ttt aaa gct tct taa tcc 3690
3691 ttt gtt ccc aga cac aca acc cct cta gct ctc gga ggc gcg atc atg aca aac ctt cca ggg aaa ctg agc aca gga tga act gtt agt 3780
3781 tgt ttt taa aag tct ata taa ata ttt caa cag atc gta aag aaa aaa ttt atc tct ttg gtc ctt gca aga gaa gtc aaa gga act ttt 3870
3871 gtt tct cct caa gac cct gga cat ctc tgt ctg tca tga ctg gaa agg gcc cgt tgt gct gaa atc cta tca tca tgg tgg att tga tct 3960
3961 tca gtc gcc aaa cac gaa tta aag tat aat tct taa ctg aac tgg tgg gtc gct tta ggg aag tat atg gcg gtc att cct ggt 4050
4051 gcg cct atc ggt gtc aca tga gcc ctg gag tgt gtc gtc cct ctc agc cct gct cct cct gcc tcc tgg ggt cca gtc act ggg acc cta 4140
4141 ctc tag aac tgt gta cca tcc aat tgc cca tca taa agg aat ctt cct gca 4191

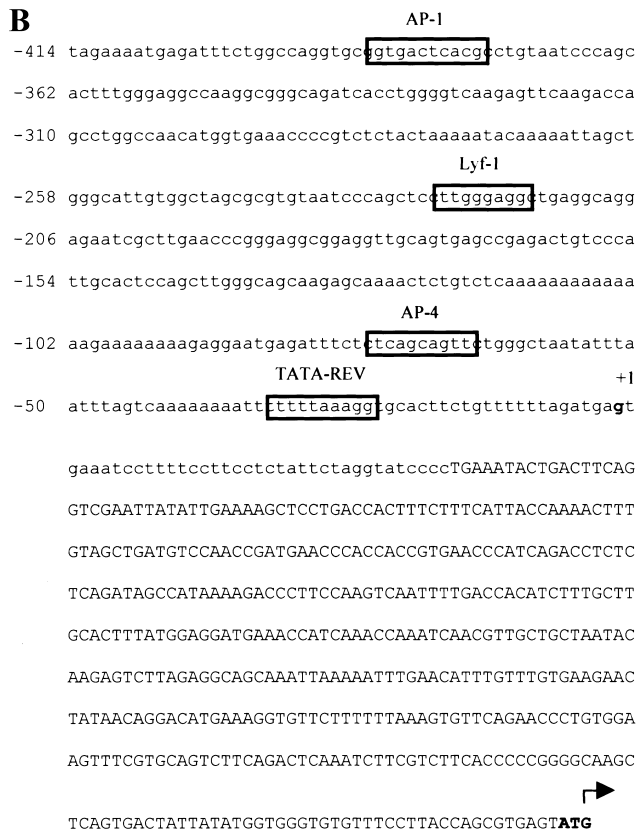


Fig. 1. (A) Nucleotide and predicted amino acid sequence of DFCP1 cDNA. A Cys–His cluster at the amino terminus (dashed line box) has similarities, but is not identical to a number of zinc finger proteins. An ATP/GTP ‘A’ consensus sequence is underlined with a solid line. The two FYVE domains at the carboxy terminus are boxed and the potential zinc-coordinating cysteine residues are in bold. The basic patches of amino acids found to be important for binding to PtdIns-3-P within the FYVE domain are underlined with a dotted line. The two Poly A⁺ consensus sequences are shaded. The intron–exon junctions indicated with arrow heads are types 0, I, 0, II, 0, II, 0, 0, I, I, respectively. Capital letters represent the sequence derived from the original cDNA clone. Lower case letters indicate the sequence derived from the genomic sequence in the database. Poly A⁺ of the ~3 kb message occurs at the junction of the upper and lower case sequences, just after the stop codon. (B) 5′ untranslated region (463 bp), predicted transcription start site and putative partial promoter region. The predicted transcription start is 37 bp upstream of the cDNA clone (shown here in upper case). Also indicated are putative transcription factor binding sequences.

most conserved amino acids within the FYVE domains are eight cysteines (bold in Fig. 1A) which are thought to coordinate Zn²⁺ binding (Stenmark et al., 1996). There is also a motif that is rich in basic residues, Arg/Lys–Arg/Lys–His–His–Cys–Arg, surrounding the third cysteine residue (dashed underline in Fig. 1A) that further characterizes the FYVE domain (Stenmark et al., 1996; Burd and Emr, 1998), and is thought to be a binding pocket for PtdIns-3-P (Misra and Hurley, 1999). It is interesting to note that the two FYVE domains in DFCP1 have a Ser and a Thr where other FYVE domains have an Arg (Fig. 2A).

By using the cDNA sequence of DFCP1 to search the HTGS database (<http://www.ncbi.nlm.nih.gov/BLAST/>), we were able to identify the complete genomic sequence (Human Genome Project, GenBank accession no. AC006345) (Fig. 2B). The gene locus name and symbol approved by the Human Genome Nomenclature Committee is *ZNFN2A1*, indicating a ‘zinc finger protein, subfamily 2A (FYVE domain-containing), 1’. Comparative analysis of cDNA and genomic sequences revealed that *ZNFN2A1* contains 11 exons and 10 introns (Fig. 2B). The Cys/His-rich cluster is entirely encoded by exon 1 and the ATP/GTP binding site is within exon 2. The two FYVE domains are encoded separately in exons 9 and 11. The genomic sequence allowed us not only to verify the 5′ untranslated region (UTR) upstream seen in the cDNA clone, but also to identify the polyadenylation consensus sequence that gives rise to the longer transcripts found in many tissues. Furthermore, PCR amplification of our own library cDNAs, as well as GenBank database searches of newly entered cDNA sequences, confirmed the length and sequence of the 3′ UTR. (Fig. 1 shows the sequence of the cDNA clone derived from the library in capital letters and the subsequently confirmed sequence first derived from the genomic sequence in lower case letters.)

The putative transcription start site was identified using a eukaryotic promoter prediction program (Neural Network Program, LBNL; <http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html>). The predicted transcription start site was located 463 bp upstream of the initiating codon and 37 bp upstream of the sequence of the cDNA clone (Fig. 1B). Analysis of the genomic sequence upstream from the putative start of transcription using the Transcription Factor Search (TFSEARCH; <http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>) revealed the presence of a number of transcription factor binding sequences including: a reverse TATA box (GGAAATTTTT at position –23); AP-4 (CTCAGCAGTT at –75); Lyf-1 (CTTGGGAGG at –225); and AP-1 (GGTGACTCAGC at –397) (Fig. 1B).

We determined the chromosomal localization of *ZNFN2A1* through a search of the dbSTS database (<http://www.ncbi.nlm.nih.gov/BLAST/>). The genomic sequence showed 100% nucleotide sequence identity to the STS marker, stSG16028, which has been localized to chromosome 14q22–q24 (NCBI GenMap '99).

3.2. Expression analysis

Because we isolated the then unique DFCP1 encoding cDNA from a cDNA library derived from cultured human bone marrow stromal cells, we wanted to compare the expression level of this mRNA in this cell population with other fibroblastic cell cultures that share various morphologic features with BMSC, but vary with

A

```

DFCP1b .....YWVDEHELLHCNCRKEFS..IKLSKHHCRACGGFCDECSHDDRRAVPSRGWDH.....PVRVCFNCKKPGDL
DFCP1a ....APAYWRNSQILSCNKCATSFK..DNDIKHHCRACGGFCDCSSSKTRPVPERGWGPA.....PVRVCDNCEARNV
Hrs2fyve AAERAPLWV.DAEE..CHRCRVQFG..VMTRKHHCRACGGIFCGKCSSKYSTIPKFGIEK.....EMRVCEPCYEQLNRKAEGK
vps27fyve SKTPALWI.DSDA..CMICSKKFS..LLMRKHHCRSCGGVFCQEHSSNSIPLFDLGIYE.....PVRVCDSCFEDYDLK
kiaa00647fyve .VTRWVEDHMASHCYNDCCEFW..LAKRRHHCRNCGNVFCAGCCHLKLPIFDQQLYD.....PVLVCNSCYEHIQVS
EEA1fyve QALNRWVAEDNEVQNCMACGKGF..VTVRRHHCRQCGNIFCAECSAKNALTESSKK.....PVRVCDACFNDLQG
fab1fyve LSKEYWMDDESSKECFSCGKTFN..TFRRKHHCRICGGIFCSCSTLLIDGDRFGCHAKM.....RVCYNCEYHADTYE
pib1fyve FWOADEEAHSCFOCKTNFS..FLVRRHHCRQCGRIFCSCSTENFVNYNKKRVHALQKNSDVESPPYRITCNECYDNLHL
fabrinfyve RAPRWIRNNEVTMCCKESFNA.LTRRRHHCRACGHVVCWCSQDYKAQLEYDGGRLN.....KVCCKDCYDIMG
    
```

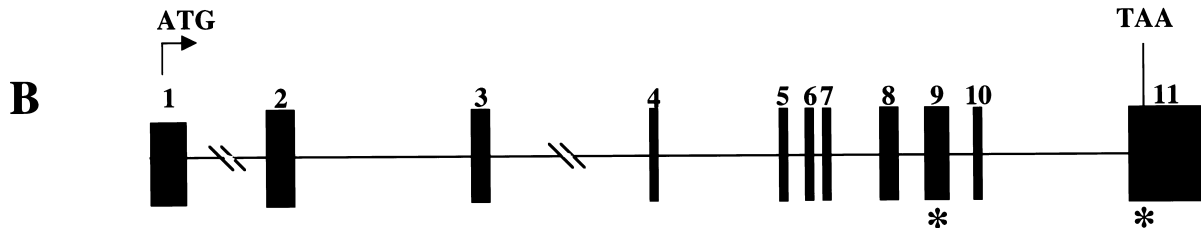
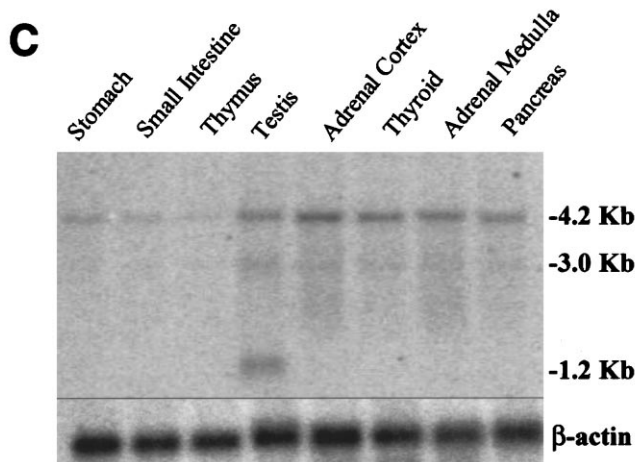
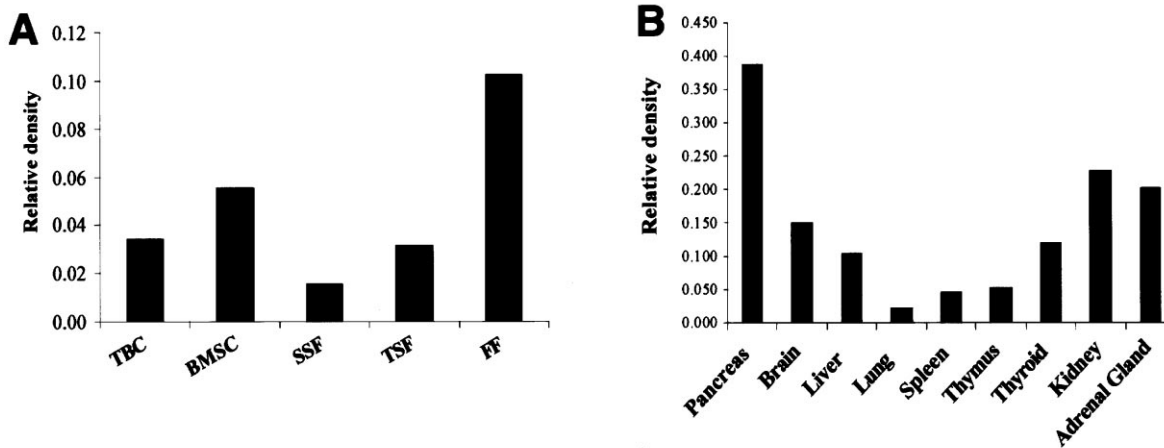


Fig. 2. The amino acid sequence alignment of the FYVE domain and the genomic structure of *ZNF2A1*. (A) DFCP1a and 1b represent the two FYVE domains in DFCP1. Sequences shown are derived from yeast: Vps27 (GenBank accession no. U24281), Pib1p (GenBank accession no. U28374), Fab1 (GenBank accession no. U01017); human: EEA1 (human early endosomal antigen, GenBank accession no. L40157), Hrs-2 (GenBank accession no. D84064) and KIAA0647 (unknown protein human, GenBank accession no. AB014547); and rat: Fabrin (GenBank accession no. AF038388). Boxes indicate amino acid identities or conservative amino acid substitutions between the two FYVE domains in DFCP1 and among the FYVE domains of other proteins. (B) The genomic structure is characterized by 11 exons represented by boxes. The putative zinc finger at the NH₂-terminus is encoded by exon 1, and the putative ATP/GTP binding site is within exon 2. The first FYVE domain is encoded by exon 9, the second FYVE domain by exon 11 (*).



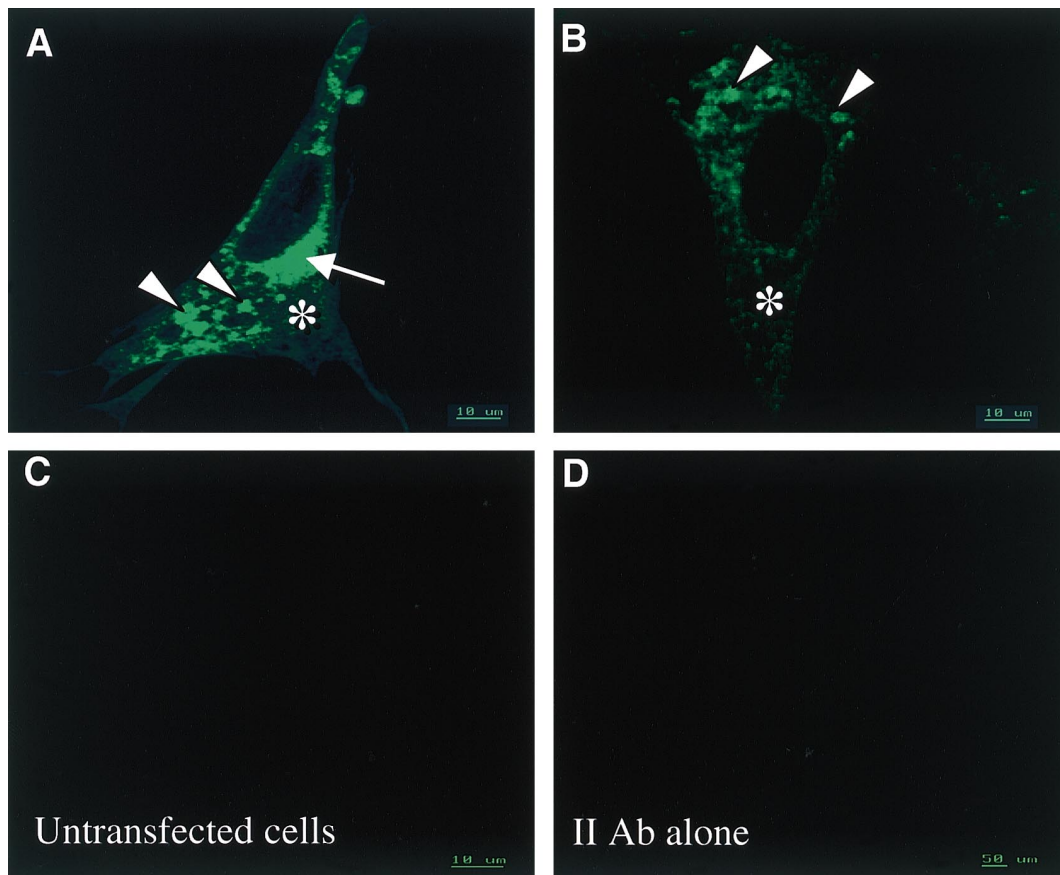


Fig. 4. Immunofluorescence staining of the c-myc–DFCP1 fusion protein in NIH-3T3 cells. NIH-3T3 cells were transfected, fixed and incubated with mouse monoclonal anti-c-myc (A and B) followed by FITC-conjugated donkey anti-mouse antibody. Staining of the Golgi (arrow), cytoplasmic vesicles (arrowheads) and diffuse staining characteristics of rough endoplasmic reticulum (asterisk) were routinely observed. (C) Untransfected cells; (D) cells transfected and incubated in the presence of only the secondary antibody.

respect to their ability to form bone. Specifically, we compared expression levels of DFCP1 in mature osteoblastic cells (TBC) with BMSCs which are osteogenic precursors, and to similar, but non-osteogenic stromal cells derived from spleen, thymus and foreskin fibroblasts (SSF, TSF, FF, respectively). DFCP1 mRNA was expressed in all cell populations examined, with the highest level of expression in FF (Fig. 3A). These results indicate that there is no obvious correlation between DFCP1 mRNA expression levels and osteogenic potential, and that DFCP1 is not unique to bone-forming cells.

Northern blot analysis of mRNA isolated from a variety of human tissues confirmed that DFCP1 is not specific to BMSCs, but is widely expressed (Fig. 3B and C). Quantitative analysis showed relatively higher expression of DFCP1 mRNA in endocrine tissues such as pancreas, adrenal gland and thyroid (Fig. 3B).

Hybridization of the DFCP1 cDNA probe to an endocrine tissue blot (Clontech, Palo Alto, CA) confirmed these findings, and furthermore revealed the presence of three different size transcripts: 4.2, 3 and 1.2 kb (Fig. 3C). The 4.2 and 3 kb appeared to be ubiquitously expressed in all tissues tested, whereas the 1.2 kb transcript was found only in testis. The presence of a 1.4 kb long 3' UTR accounts for the difference in length between the 4.2 and 3 kb mRNAs, however, the possibility of alternative splicing cannot be ruled out at this time. A search of the dbEST database with the mRNA sequences identified a group of recently deposited ESTs derived from other library sources that shared the same 3' end as the 4.2 kb transcript, as well as a separate group that shared the 3' end of the 3 kb transcript.

The 4.2 kb mRNA contains a polyadenylation con-

Fig. 3. Northern blot analysis and tissue distribution of DFCP1 mRNA. (A) Relative expression level of DFCP1 mRNA in cell cultures that differ with respect to their ability to form bone (TBC, trabecular bone cells; BMSC, bone marrow stromal cells; SSF, spleen stromal fibroblasts; TSF, thymus stromal fibroblasts; FF, foreskin fibroblasts). (B) Relative expression levels of DFCP1 mRNA in different tissues. Relative levels were determined by counting the amount of radioactivity (PhosphoImager) and then normalizing the value against β -actin. (C) The expression pattern by Northern blot of DFCP1 transcripts in a variety of tissues, demonstrating the presence of the three different transcripts (4.2, 3 and 1.2 kb).

sensus at the end of the 3' UTR (4172 bp), but with a substitution of the first nucleotide, A, with a C (CATAAA). A substitution of the first nucleotide of the AATAAA consensus sequence appears to be rare among reported mRNAs. Previous *in vitro* findings suggest that mutations in the AATAAA consensus sequence decrease the rate of cleavage and polyadenylation of the mRNA transcripts (Sheets et al., 1990). However, the 4.2 kb transcript appears in high abundance in several tissues (primarily endocrine tissues). Thus, cleavage and polyadenylation may not be rate limiting for this gene, or there may be sequences elsewhere in the precursor which compensate for the non-conventional polyadenylation consensus. Interestingly, the 3 kb mRNA does have a canonical AATAAA polyadenylation consensus sequence near the end, but still within the ORF (at 2738 bp), an unusual placement in eukaryotic cells. The fact that other ESTs recently deposited from libraries made from Poly A-containing mRNAs are identical to the 3' ends of both the 4.2 and 3.0 kb mRNAs indicates that both the non-standard sequence in the 4.2 kb form and the sequence within the ORF of the 3.0 kb form are in fact utilized. To date, the 1.2 kb message found only in testis remains unexplained. Whether this transcript is an alternative splice of *ZNFN2A1* or represents an independent gene with some homology to *ZNFN2A1* is unknown at this time.

3.3. Immunolocalization

Based on previous reports that other FYVE-containing proteins may be involved in intracellular membrane trafficking (Simonsen et al., 1998; Bean et al., 1997; Weisman and Wickner, 1992; Piper et al., 1995; Yamamoto et al., 1995), we wanted to determine where DFCP1 would localize within the cell. NIH-3T3 fibroblasts were therefore transfected with a pCMV-tag3b plasmid containing the DFCP1 ORF preceded by a sequence tag coding for the c-myc epitope (MEQKLISE). The DFCP1 was then localized using the monoclonal anti-c-myc. Confocal microscopy revealed an absence of nuclear accumulation, but a pattern typical of localization in the Golgi apparatus, endoplasmic reticulum and cytoplasmic vesicles. This pattern of localization supports the hypothesis that DFCP1 may indeed be recruited to vesicular structures due to the presence of one or both of the FYVE domains (Fig. 4A and B).

4. Conclusions

In this study we have identified and cloned a novel gene (*ZNFN2A1*) which encodes for a protein that we have named DFCP1. Double FYVE is expressed in a

wide variety of tissues, but the highest levels were seen clearly in endocrine tissues.

1. DFCP1 contains sequences encoding for: (1) a Cys- and His-rich cluster in the amino terminal portion of the protein (possibly a new type of zinc finger); (2) an ATP/GTP binding site; and (3) two FYVE domains near the carboxy terminus.
2. This is the first time that two FYVE domains have been found in one protein.
3. DFCP1 mRNA has a relatively long (463 bp) 5' UTR.
4. The polyadenylation consensus sequence for the longer transcript (4.2 kb) is non-standard, while in the shorter transcript (3 kb) it is within the coding sequence, both of which are relatively rare findings.
5. Confocal microscopy studies revealed that when transfected into NIH-3T3 cells, the fusion protein appears to localize to membranous structures in the cytoplasm.

Further analysis is required to characterize the role that this protein plays in cells, in particular, on membrane trafficking, as has been suggested for other FYVE-containing proteins.

Acknowledgements

We thank Dr. Bill Swaim for invaluable help with the confocal microscopy studies and Dr. Lynn King for a critical review of this manuscript.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Bean, A.J., Seifert, R., Chen, Y.A., Sacks, R., Scheller, R.H., 1997. Hrs-2 is an ATPase implicated in calcium-regulated secretion. *Nature* 385, 826–829.
- Burd, C.G., Emr, S.D., 1998. Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domains. *Mol. Cell* 2, 157–162.
- Castro-Malaspina, H., Gay, R.E., Resnick, G., Kapoor, N., Meyers, P., Chiarieri, D., McKenzie, S., Broxmeyer, H.E., Moore, M.A., 1980. Characterization of human bone marrow fibroblast colony-forming cells (CFU-F) and their progeny. *Blood* 56, 289–301.
- Dieudonne, S.C., Kerr, J.M., Xu, T., Sommer, B., DeRubeis, A.R., Kuznetsov, S.A., Kim, I.S., Gehron Robey, P., Young, M.F., 1999. Differential display of human marrow stromal cells reveals unique mRNA expression patterns in response to dexamethasone. *J. Cell Biochem.* 76, 231–243.
- Emini, E.A., Hughes, J.V., Perlow, D.S., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55, 836–839.
- Feuerstein, R., Wang, X., Song, D., Cooke, N.E., Liebhaber, S.A., 1994. The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proc. Natl. Acad. Sci. USA* 91, 10655–10659.
- Freemont, P.S., 1993. The RING finger. A novel protein sequence motif related to the zinc finger. *Ann. NY Acad. Sci.* 684, 174–192.

- Friedenstein, A.J., Chailakhyan, R.K., Latsinik, N.V., Panasyuk, A.F., Keiliss-Borok, I.V., 1974. Stromal cells responsible for transferring the microenvironment of the hemopoietic tissues. Cloning in vitro and retransplantation in vivo. *Transplantation* 17, 331–340.
- Gaullier, J.M., Simonsen, A., D'Arrigo, A., Bremnes, B., Stenmark, H., Aasland, R., 1998. FYVE fingers bind PtdIns(3)P. *Nature* 394, 432–433.
- Gaullier, J.M., Simonsen, A., D'Arrigo, A., Bremnes, B., Stenmark, H., 1999. FYVE finger proteins as effectors of phosphatidylinositol 3-phosphate. *Chem. Phys. Lipids* 98, 87–94.
- Jameson, B.A., Wolf, H., 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci.* 4, 181–186.
- Janin, J., Wodak, S., 1978. Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* 125, 357–386.
- Jia, L., Wilkin, D.J., King, L.M., Young, M.F., Derubeis, A., Powell, J., Yang, L., Robey, P.G., Francomano, C.A., 1998. Bone marrow stromal cells and trabecular bone cells: comparative functional genomics. *Am. J. Hum. Genet.* 63 (Suppl.), A250.
- Kuznetsov, S.A., Krebsbach, P.H., Satomura, K., Kerr, J., Riminucci, M., Benayahu, D., Robey, P.G., 1997. Single-colony derived strains of human marrow stromal fibroblasts form bone after transplantation in vivo. *J. Bone Miner. Res.* 12, 1335–1347.
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.
- Misra, S., Hurley, J.H., 1999. Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell* 97, 657–666.
- Owen, M., Friedenstein, A.J., 1988. Stromal stem cells: marrow-derived osteogenic precursors. *Ciba Found. Symp.* 136, 42–60.
- Patki, V., Lawe, D.C., Corvera, S., Virbasius, J.V., Chawla, A., 1998. A functional PtdIns(3)P-binding motif. *Nature* 394, 433–434.
- Piper, R.C., Cooper, A.A., Yang, H., Stevens, T.H., 1995. VPS27 controls vacuolar and endocytic traffic through a prevacuolar compartment in *Saccharomyces cerevisiae*. *J. Cell Biol.* 131, 603–617.
- Robey, P.G., Termine, J.D., 1985. Human bone cells in vitro. *Calcif. Tissue Int.* 37, 453–460.
- Sheets, M.D., Ogg, S.C., Wickens, M.P., 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* 18, 5799–5805.
- Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., Stenmark, H., 1998. EEA1 links PI(3)K function to Rab5 regulation of endosome fusion. *Nature* 394, 494–498.
- Stenmark, H., Aasland, R., Toh, B.H., D'Arrigo, A., 1996. Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.* 271, 24048–24054.
- Wang, Q.R., Yan, Z.J., Wolf, N.S., 1990. Dissecting the hematopoietic microenvironment. VI. The effects of several growth factors on the in vitro growth of murine bone marrow CFU-F. *Exp. Hematol.* 18, 341–347.
- Weisman, L.S., Wickner, W., 1992. Molecular characterization of VAC1, a gene required for vacuole inheritance and vacuole protein sorting. *J. Biol. Chem.* 267, 618–623.
- Yamamoto, A., DeWald, D.B., Boronenkov, I.V., Anderson, R.A., Emr, S.D., Koshland, D., 1995. Novel PI(4)P 5-kinase homologue, Fab1p, essential for normal vacuole function and morphology in yeast. *Mol. Biol. Cell* 6, 525–539.
- Zhang, G., Kazanietz, M.G., Blumberg, P.M., Hurley, J.H., 1995. Crystal structure of the cys2 activator-binding domain of protein kinase C delta in complex with phorbol ester. *Cell* 81, 917–924.