Biochemical and Biophysical Research Communications **261,** 64–70 (1999)
Article ID bbrc.1999.0963, available online at http://www.idealibrary.com on IDEAL®

A Novel Human Gene Encoding an F-box/WD40 Containing Protein Maps in the SHFM3 Critical Region on 10q24

Peter Ianakiev,* Michael W. Kilpatrick,* Caroline Dealy,† Robert Kosher,† Julie R. Korenberg,‡ Xiao-Ning Chen,† and Petros Tsipouras*,[1]

**Department of Pediatrics and †Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, Connecticut 06030; and ‡Medical Genetics Birth Defects Center, Cedar–Sinai Medical Center, UCLA, Los Angeles, California*

Received June 4, 1999

We report the cloning and characterization of a new human gene, Dactylin, encoding a novel member of the F-box/WD40 protein family. The Dactylin gene comprises nine exons distributed in more than 85 kb of genomic DNA and encoding a protein with four WD40 repeats and an F-box motif. Northern blot analysis demonstrates a single 2.8 kb transcript in brain, kidney, lung and liver. FISH hybridization localized Dactylin to 10q24.3. Using an Msc I SNP identified in the first exon of the gene, we were able to assign Dactylin within the critical region for Split Hand Split Foot malformation (SHFM3) that has been mapped to 10q24. The SHFM3 phenotype includes absence or hypoplasia of the central digital rays, a deep median cleft and syndactyly of the remaining digits. Recent studies have demonstrated the importance of F-box/WD40 proteins in the regulation of developmental processes, by a mechanism of specific ubiquitinization and subsequent proteolysis of target proteins belonging to the Wnt, Hh and NF-κB signaling pathways. The chromosomal location of Dactylin and its putative function as an F-box/WD40 repeat protein, likely to be involved in key signaling pathways crucial for normal limb development, make it a promising candidate gene for SHFM3.

The Split Hand-Split Foot malformation is a naturally occurring phenotype that appears to result from errors in the developmental patterning of the limbs (1). The malformation is characterized by a median cleft in the hand and foot associated with absence or hypoplasia of the central digits, and frequently fusion of the remaining digits. At least three different Split Hand-Split Foot malformation loci have been identified (2, 3, 4). A locus for Split Hand-Split Foot malformation (SHFM3) has been localized on human chromosome 10 (10q24) and it is the only one of the three not associated with cytogenetic abnormalities (3, 5). An analogous phenotype, *Dactylaplasia*, has been described in mouse (6) and mapped in a syntenic region of chromosome 19 (7).

Meiotic mapping of families with SHFM3 established a critical region on 10q24 (5, 8). A detailed database search for the identification of a suitable candidate gene revealed the presence of several characterized genes (FGF8, PAX2, and HOX11) and of numerous ESTs mapping in the SHFM3 critical region.

We report the cloning and characterization of a novel gene, which is localized within the SHFM3 critical region. This gene is expressed in several tissues at different stages of development. The potential involvement of the protein encoded by this gene in the etiology of SHFM3 led us to name it Dactylin (from the Greek *dactylon* for finger). The structure of dactylin comprises an F-box and WD40 motifs. Proteins containing F box/WD40 motifs have been found in the ubiquitin-dependent proteolytic pathways that are common to several species ranging from yeast to *homo sapiens* (9). Dactylin might be involved in the molecular tagging of protein molecules destined for proteolysis. However, its relationship to SHFM3 remains unknown.

MATERIALS AND METHODS

EST identification and DNA sequencing. ESTs were identified using BLAST search of dbEST database (10). Sequencing reactions were performed on an ABI 373 sequencer using the PE Big Dye terminator sequencing kit. Double stranded PCR products and plasmid DNA were purified using Qiagen purification kits. Nucleotide sequences were organized in a contig using the DNA Star package (11).

[1] Corresponding author. Department of Pediatrics, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030. Fax: 860-679-1355. E-mail: Tsipouras@nso1.uchc.edu.

0006-291X/99 $30.00

RNA isolation and generation of cDNA. Total RNA was isolated from a ten-week old human embryo using a procedure previously described (12). Five μg of mRNA was reverse-transcribed (Super-Script II, Gibco BRL) with oligo dT_{20} primer using the conditions recommended by the supplier. Ten μl of RT reaction was then subjected to PCR with primers Ex1F1: 5′-GGCAAGGCGGAAGAG-GAAGC-3′ (Nt. 297–316) and i9R1: 5′-ATTGGCAGCCTCACCT-TTCC-3′ (Nt. 2249–2268) in a total reaction volume of 100 μl. The conditions of the reaction were the following: 0.2 mM dNTP, 1× XL III PCR buffer (Boehringer), 1 μM each primer, 1 U Taq polymerase at 97°C/20 sec, 55°C/30 sec, 72°C/2 min for 30 cycles. A second nested PCR was then performed with primers Ex1F2: 5′-GAAGGA-GGAATGGAGGGTC-3′ (Nt. 511–529), and Ex9R2: 5′-GAGGTTG-TAAGACAGGGCAGC-3′ (Nt. 1811–1831) using the same conditions. The resulting product was gel purified, cloned into pGEM-T vector (Promega) and sequenced.

Determination of transcription initiation site. Total RNA from a ten-week old human embryo was isolated and reverse-transcribed into cDNA as described above. PCR was then performed using 0.2 mM dNTP, 1× XL III PCR buffer (Boehringer), 1 μM each primer, 1 U Taq polymerase at 97°C/20 sec, 55°C/30 sec, 72°C/2 min for 30 cycles. The products were separated on 1.8% agarose gel and transferred to a nylon membrane by alkali blot. Fragments were hybridized overnight at 45°C in 5× SSC, 7% SDS with 1 pmol ^{32}P-labelled oligonucleotide probe corresponding to nucleotide position 737–760. Filters were washed with 2× SSC, 1% SDS at 37°C and the fragments were detected by autoradiography.

Determination of genomic organization. Exon-intron borders were isolated using an improvement of the vectorette procedure, described previously (13). Briefly, DNA from a BAC clone (325g20), containing the human Dactylin gene was digested overnight with RsaI or AvaII restriction endonuclease. Twenty nanograms of the resulting fragments were ligated to 100 pmol of adapter comprised of strands of unequal length. Hot start PCR using one specific and one adapter primer was then employed to amplify fragments situated between the most proximal restriction site and the specific primer. These fragments were then purified and sequenced. Genomic primers were generated for each exon using the PrimerSelect program (14).

Southern and Northern blot hybridization. For Southern blot analysis PCR-amplified DNA fragments were transferred to Hybond N+ membrane (Amersham), UV cross-linked, baked, hybridized, and washed following established protocols. A human fetal multiple tissue filter (Clontech) was used for Northern blot analysis. The filter contained 2 μg per lane of poly-A+ RNA extracted from an approximately twenty-week old fetal brain, lung, liver, and kidney tissue. The filter was hybridized following the conditions recommended by the supplier.

Genotyping. Fifty nanograms of genomic DNA were amplified using primers specific for the 5′ end of the gene. The generated amplicons were digested overnight with 1 U of MscI restriction endonuclease and the resulting fragments were separated by electrophoresis on a 2% agarose gel. The gel was stained in ethidium bromide and the fragments were visualized and photographed.

In situ hybridization. In situ hybridization was performed on two sections from a 5 1/2 week human paraffin-embedded embryo. The probe used was EST clone H45642 consisting of 55 bp of coding region plus 632 bp of 3′ UTR. Hybridization to embryo sections was performed as described (15), except that hybridization was at 45°C and washing at 55°C.

Fluorescent in situ hybridization (FISH) on metaphase chromosomes. DNA from ESTs H45642 and AA397652 were labeled with biotin-14-dATP (GIBCO-BRL) using nick translation and were hybridized together to metaphase chromosomes prepared from normal male peripheral blood lymphocytes made by the bromodeoxy-uridine synchronization method (16). FISH was performed essentially according to the method described (16). Briefly, 40 ng of each cDNA probe, 2 μg Cot 1 DNA and 8 μg sonicated salmon sperm DNA in 10 μl of prehybridization mixture were pre-annealed at 37°C for 20 minutes and then applied to a denatured chromosome slide. Post-hybridization washes were at 44°C in 2× SSC/50% formamide (5 min. × 3) followed by washing in 1× SSC at 55°C (5 min. × 3). Hybridized DNAs were detected with avidin-conjugated fluorescein isothiocyanate (Vector Labs). Two amplifications were carried out using biotinylated-anti-avidin. To generate clear reverse bands, metaphase chromosomes were counterstained with chromomycin A3 followed by distamycin A (16). Color images were captured using the Photometrics Cooled-CCD camera (CH250) and the BDS image analysis system (ONCOR Imaging, Gaithersburg, MD).

RESULTS

cDNA Assembly and Cloning

Initially three human EST clones, H45642, AA397652, and AA130939 were identified by BLAST search as having a high degree of homology to the mouse BAC clone CIT282B21. This clone had been localized to the region of mouse chromosome 19 syntenic to human chromosome 10q24, which contains the SHFM3 critical region. Further database screening, using the three human EST clones identified a number of additional ESTs, from a variety of tissues and developmental stages, which were then assembled in a contig representing more than 1500 bp of cDNA. To confirm that these ESTs were derived from a single transcript, a cDNA was cloned and sequenced from a ten-week old human embryo. Additional 5′ sequence was generated from the human BAC clone 325g20 which contains the Dactylin gene. The resulting cDNA spanned 2490 bp and contained a 1782 bp ORF, 76 bp 5′UTR, and 632 bp 3′UTR.

Determination of the Transcription Initiation Site

We identified three potential in-frame initiator methionine residues (nucleotide positions 77, 155, and 620) (Figure 1a). In order to determine which of the three residues is likely to be used, we attempted to define the transcription initiation site. To determine the site of transcription initiation, total embryonic RNA was subjected to RT-PCR using primers Ex1F3: 5′-CGCTCATGGGCAGCCAG-3′; (Nt. 150–166), or Ex1F4: 5′-CAAGGCTGCAGGTGTGAG-3′; (Nt. 10–27) with 568R2: 5′-CCAGGATGAAATTAGCCTGGGA-3′; (Nt. 1019–1040). As a control we used amplification from genomic DNA using primers Ex1F3 and Ex1F4 paired with Ex1gR1: 5′-AGTGCTCGGGAAAGGGT-GTA-3′ located within intron 1. Fragments of the expected 885 bp and 1025 bp size were generated from genomic DNA using both, Ex1F3; Ex1gR1 and Ex1F4; Ex1gR1. In contrast, the expected 890 bp fragment was generated from cDNA using Ex1F3;568R2, but no fragment was generated using Ex1F4;568R2 (Figure 1b). This suggests that the transcription initiation site is

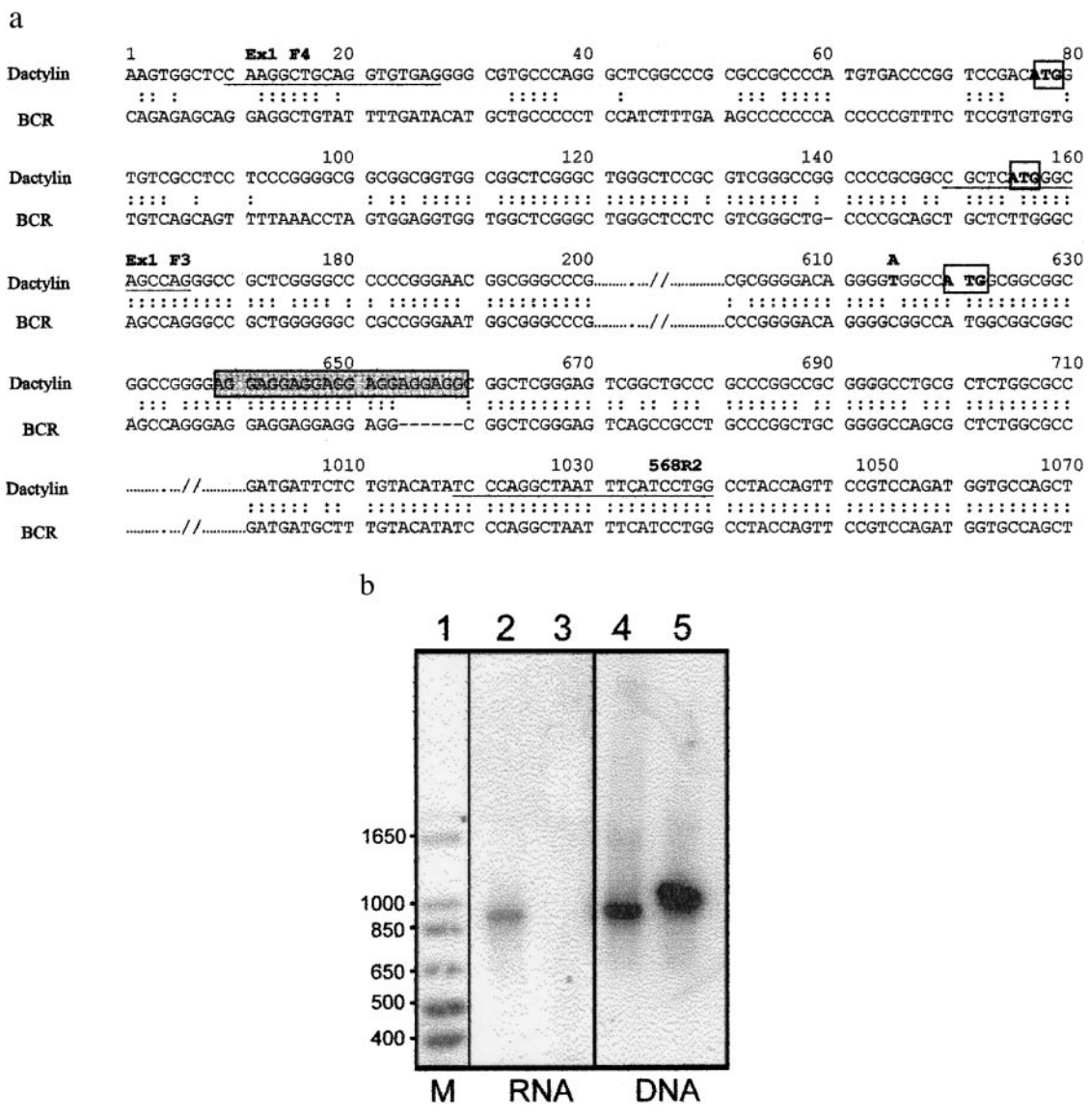

FIG. 1. (a) Sequence of the 5′-end of the Dactylin cDNA. The sequence of the homologous region of the pseudogene on 22q11.2 is also shown. Three ATG putative initiator codons are boxed. The AGG/TCC repeat is shadowed. Primers Ex1F3, Ex1F4, and 568R2 are underlined. The T/A SNP site in position 615 is in bold. (b) Determination of transcription start site by RT-PCR. Lane1: marker; lanes 2 and 3: RT-PCR of fetal RNA, using primers Ex1F3;568R2 and Ex1F4;568R2, respectively; lanes 4 and 5: amplification of genomic DNA control using primers Ex1F3;Ex1gR1 and Ex1F4;Ex1gR1.

located between Ex1F3 and Ex1F4 (nucleotide position 27 to 150) (Figure 1a).

Identification of a SNP at the 5′ End of the cDNA

Sequencing of genomic DNA from a panel of unrelated individuals identified a T/A transversion at nucleotide position 615 (Figure 1a). This nucleotide change manifests itself as the presence or absence of a MscI restriction endonuclease site. The SNP was inherited as a Mendelian trait. An estimated allele frequency of 0.84 and 0.16 was derived from the analysis of 86 chromosomes. This SNP was utilized to genotype a panel of SHFM3 families. The inheritance of the SNP was independent of the segregation of the SHFM3 phenotype.

The predicted protein shows no apparent homology to any known vertebrate protein. It comprises 412 amino-acids with an estimated molecular weight of 46.3 kDa and pI of 7.52. An apparent F-box motif is located near the amino-terminus, followed by a series of WD40 repeats (Figure 2).

Identification of a Pseudogene

BLAST search using the human Dactylin ESTs also identified a highly homologous sequence on chromo-

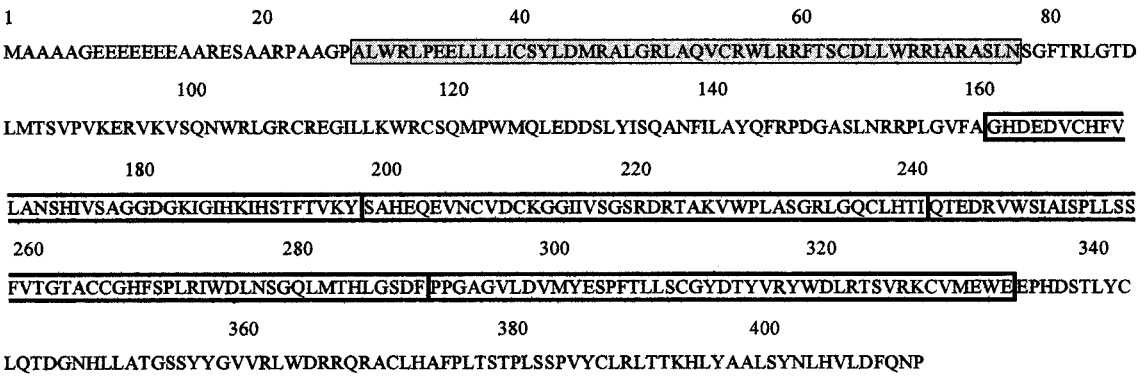

FIG. 2. Predicted amino-acid sequence of the dactylin protein. The putative F-box is shown in a shaded box and the four WD40 repeats are enclosed in clear boxes.

some 22. This sequence corresponds to the breakpoint cluster region (BCR) gene on 22q11.2. The homology between the pseudogene and human Dactylin is more than 93% at the nucleotide level. Consistent with that was the observation of a secondary signal by FISH analysis on chromosome 22q11.2 where the BCR gene has been localized (unpublished data).

Structure of the Dactylin Gene

The BAC clone (325g20) encompassing the human Dactylin gene was used for identification of exon intron structure. For each exon both 5′ and 3′ flanking boundaries were isolated and sequenced. The identified exons ranged in size from 66 bp (exon 6) to 884 bp (exon 1). Amplification across introns using exon specific primers was used to estimate the corresponding intron sizes which ranged in size from 500 bp to more than 45 kb (Figure 3). The total gene size was estimated to be approximately 85 kb.

Expression of the Dactylin Gene

Hybridization of a multiple tissue Northern blot with a probe encompassing the 3′UTR detected a single 2.8 kb band in all examined tissues (Figure 4). The strongest expression of the gene was detected in human fetal brain, followed by kidney, lung, and liver (Figure 4). In situ hybridization of Dactylin in two sections of 5 1/2 week human embryo confirmed Dactylin expression in the developing CNS (Figure 5 A, B) and liver (Figure 5 C, D). In addition Dactylin transcripts were present in spinal ganglia and differentiating myotome (Figure 5 A, B), and at low levels in mesoderm and ectoderm of the developing limb (Figure 5 E, F).

Gene Localization

Cytogenetic localization. Two overlapping ESTs, H45642, AA397652, were mapped to chromosome band 10q24.3 using FISH (Figure 6a). Over 50 metaphase cells were evaluated. Signals were clearly seen on two chromatids of at least one chromosome 10q24.3 in 80% of the cells. Secondary signals were noticed on chromosome 22q11.2, at the 13q21.3/22 border and on 15q22.3. The intensity of the signals on 13q and 15q was much lower than on 10q and 22q and may reflect regions with small degree of homology.

Genetic mapping. The SNP at position 615 was used to genotype critical meioses for previously described recombination events in members of SHFM3 families. In one family, recombination events localized Dactylin distal to D10S541 and proximal to D10S587. In addition, recombination events in a second family placed Dactylin between markers D10S185 and D10S187. Thus, Dactylin was genetically localized within the previously defined SHFM3 critical region (Figure 6b).

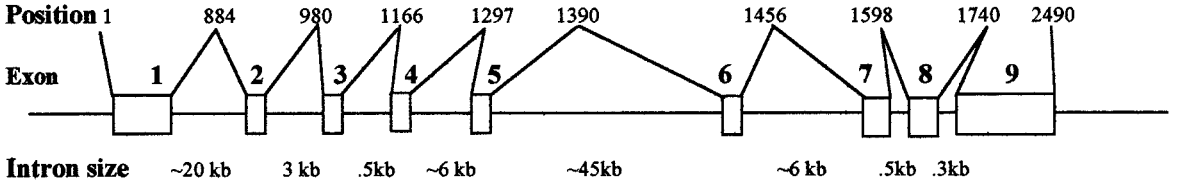

FIG. 3. Genomic organization of the Dactylin gene. The nine exons are shown as boxes, the positions of the introns shown above, and the estimated intron size below.

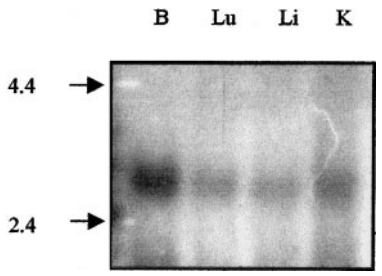

FIG. 4. Northern blot hybridization with Dactylin 3′ UTR probe: Lanes loaded with fetal RNA extracted from: B, brain; Lu, lung; Li, liver; and K, kidney. Positions of 2.4 and 4.4 kb markers are shown on the left.

DISCUSSION

We identified a novel gene, which we mapped on 10q24 within the SHFM3 critical region. This gene encodes a protein, which we have named dactylin, showing no homology to any known protein in vertebrates. The only identifiable homology is to the yeast met30 (26). The predicted structure of dactylin comprises 412 amino acids and includes an F-box and four WD40 domains, however no leader peptide sequence or transmembrane domain could be identified, suggesting that dactylin is a cytoplasmic protein.

The Dactylin gene extends over more than 85kb of DNA. The coding sequence is distributed in 9 exons varying in size. A highly homologous sequence to Dactylin was identified on chromosome 22 both by FISH analysis and database scanning. This suggests the presence of a retro-transposed pseudogene. The pseudogene appears not to be expressed as suggested by the lack of deposited pseudogene-derived EST sequences.

Our data indicated that the transcription initiation site is located between nucleotides 27 and 150. The SNP at nucleotide 615 would predict a Val to Glu amino-acid substitution. As this SNP appears to have no phenotypic effect, this suggests it falls outside the coding region. In addition, the degree of nucleotide homology between Dactylin and the chromosome 22 pseudogene, while >90% 3′ from nucleotide 100 drops to 40% 5′ from position 100. Taken together, these data suggest the most likely translation initiation site is the

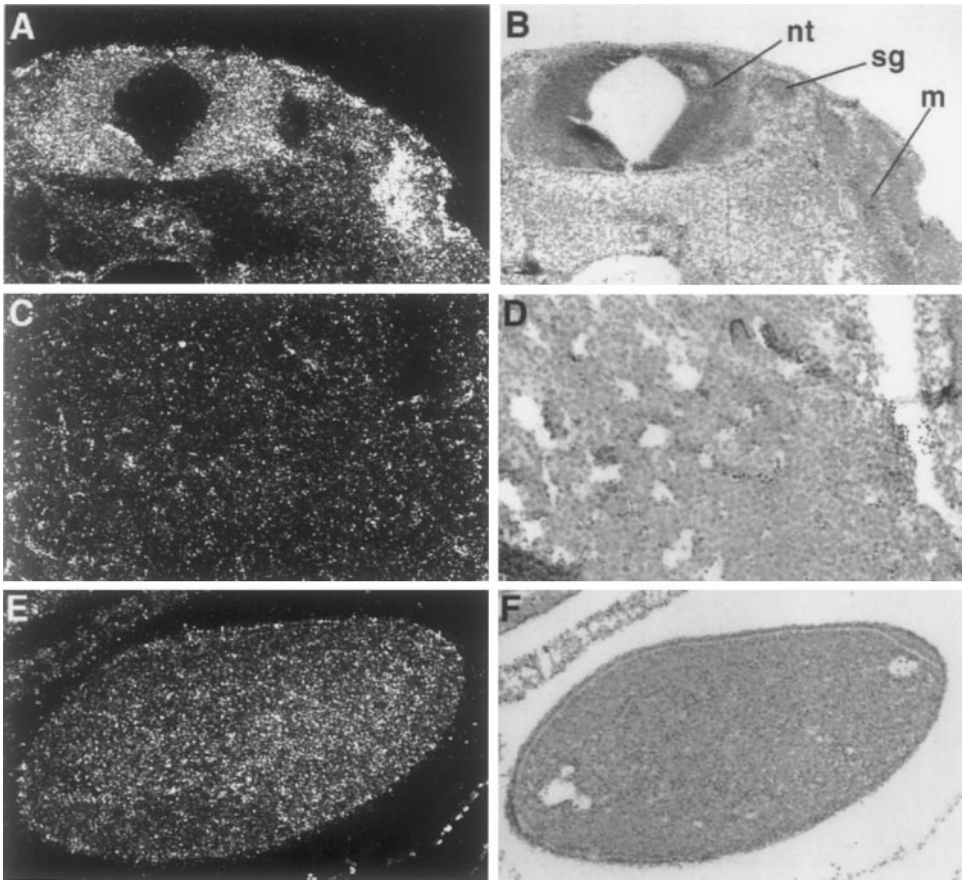

FIG. 5. In situ hybridization of Dactylin expression in a 5 1/2 week human embryo. The panels on the left and right are from dark and light microscopy respectively. (A, B): Dactylin is expressed in the neural tube (nt), spinal ganglia (sg) and myotome (m). Dactylin transcripts are also detected at low levels throughout liver parenchyma (C, D) and in mesoderm and ectoderm of the limb (E, F).

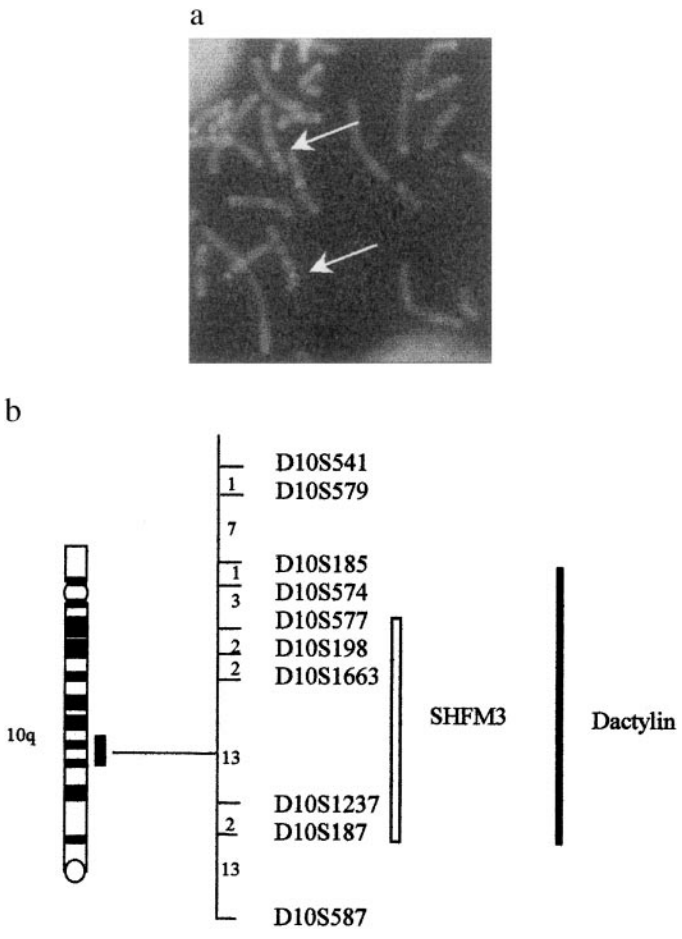

FIG. 6. (a) FISH of a metaphase chromosome spread hybridized with a Dactylin 3′ UTR probe. Arrows show the position of the chromosome 10q24 signal. (b) Genetic map of 10q24 showing, from left to right, ideogram of chromosome 10q, FISH localization of Dactylin, and genetic map with distances in centimorgans. The SHFM3 critical region is shown by the vertical open bar, Dactylin genetic localization is presented in solid vertical bar on the right.

methionine residue at nucleotide 620. This site falls within a strong Kozak consensus sequence context (17). The reading frame remains open at the 5′-end leaving the possibility of additional exons. One interesting structural characteristic of the Dactylin gene is the presence of a trinucleotide repeat $(AGG)_7$ at the 5′-end of the coding sequence. The importance of trinucleotide repeats in the etiology of genetic disorders has been well documented (18).

A single 2.8 kb gene transcript was detected and that size was in agreement with the determined cDNA sequence. Northern blot analysis and *in situ* hybridization demonstrate that Dactylin is expressed in a broad range of embryonic tissues including central and peripheral nervous system, muscle, liver and lung. Consistent with its candidacy as a gene involved in the SHFM3 phenotype, Dactylin also appears to be expressed in the developing limb.

F-box containing proteins have been suggested to play a crucial role in ubiquitin-dependent proteolysis by targeting specific protein substrates for degradation (19). These substrates include Cubitus interruptus, IκBα and β-catenin which are members of the Hedgehog, NF-κB and Wnt signaling pathways, respectively (20). In vertebrates, sonic hedgehog, NF-κB and β-catenin mediated Wnt signaling, have all been shown to be essential for normal limb development (21–25). In this paper we describe a new member of the F-box/WD40 protein family, dactylin. Based on its chromosomal location within the SHFM3 critical region and its expression in the limb, Dactylin is a promising candidate for Split Hand Split Foot malformation. Furthermore, through its F-box/WD40 domains, Dactylin has the potential to be a novel modulator of Wnt, Hedgehog or NF-κB signaling during limb morphogenesis.

ACKNOWLEDGMENTS

The authors thank Arend Sidow for suggesting the term dactylin, a generous gift of BAC clones, for sharing unpublished data, and for many useful discussions; Nada Zecevic for a gift of human fetal tissue; Ivo Kremensky for a gift of embryonic RNA; and Stylianos Antonarakis and Tessa Homphray for contributing DNA from families with SHFM3. This work was supported in part by HD-22610 (CD, RK and PT), a grant from the Coles Family Foundation (P.T.), by DOE Grants 92ER61402 and 96ER62294 (J.R.K.) and HL50025 (J.R.K.). J.R.K. holds the Brawerman Chair in Molecular Genetics at CSMC.

REFERENCES

1. Temtamy, S. A., and McKusick, V. A. (1978) *Birth Defects Orig. Artic. Ser.* **14,** 1–619.
2. Crackower, M. A., Scherer, S. W., Rommens, J. M., Hui, C., Poorkaj, P., Soder, S., Cobben, J. M., Hudgins, L., Evans, J. P., and Tsui, L. C. (1996) *Hum. Molec. Genet.* **5,** 571–579.
3. Nunes, M. E., Schutt, G., Kapur, R. P., Luthardt, F., Kukolich, M., Byers, P., and Evans J. P. (1995) *Hum. Mol. Genet.* **4,** 2165–2170.
4. Faiyaz ul Haque, M., Uhlass, S., Knapp, M., Schuler, H., Friedl, W., Ahmad, M., and Propping, P. (1993) *Hum. Genet.* **91,** 17–19.
5. Gurrieri, F., Prinos, P., Tackels, D., Kilpatrick, MW., Allanson, J., Genuardi, M., Vuckov, A., Nanni, L., Sangiorgi, E., Garofalo, G., Nunes, ME., Neri, G., Schwartz, C., and Tsipouras, P. (1996) *Am. J. Med. Genet.* **62,** 427–436.
6. Chai, C. K. (1981) *J. Hered.* **72,** 234–237.
7. Johnson, K. R., Lane, P. W., Ward-Bailey, P., and Davisson, M. T. (1995) *Genomics* **29,** 457–464.
8. Raas-Rothschild, A., Manouvrier, S., Gonzales, M., Farriaux, J. P., Lyonnet, S. X., and Munnich, A. (1996) *J. Med. Genet.* **33,** 996–1001.
9. Patton, E. E., Willems, A. R., and Tyers, M. (1998) *Trends Genet.* **14,** 236–243.
10. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
11. Clewley, J. P. (1995) *Mol. Biotechnol.* **3,** 221–224.

12. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
13. Arnold, C., and Hodgson, I. J. (1991) *PCR Methods Appl.* **1,** 39–42.
14. Plasterer, T. N. (1997) *Methods Mol. Biol.* **70,** 291–302.
15. Coelho, C. N. D., Krabbenhoft, K., Upholt, W. B., Fallon, J. F., and Kosher, R. A. (1991) *Development* **113,** 1487–1493.
16. Korenberg, J. R., Chen, X. N., Adams, M. D., and Venter, J. C. (1995) *Genomics* **29,** 364–370.
17. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
18. Ashley, C. T., Jr., and Warren, S. T. (1995) *Annu. Rev. Genet.* **29,** 703–728.
19. Patton, E. E., Willems, A. R., and Tyers, M. (1998) *Trends Genet.* **14,** 236-243
20. Maniatis, T. A. (1999) *Genes Dev.* **13,** 505–510.
21. Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A. (1996) *Nature* **383,** 407–413.
22. Kengaku, M., Capdevila, J., Rodriguez-Esteban, C., De La Pena, J., Johnson, R. L., Belmonte, J. C. I., and Tabin, C. J. (1998) *Science* **280,** 1274–1277.
23. Kanegae, Y., Tavares, A. T., Izpisua Belmonte, J. C., and Verma, I. M. (1998) *Nature* **392,** 611–614.
24. Bushdid, P. B., Brantley, D. M., Yull, F. E., Blaeuer, G. L., Hoffman, L. H., Niswander, L., and Kerr, L. D. (1998) *Nature* **392,** 615–618.
25. Galceran, J., Farinas, I., Depew, M. J., Clevers, H., and Grosschedl, R. (1999) *Genes Dev.* **13,** 709–717.
26. Kaiser, P., Sia, R. A., Bardes, E. G., Lew, D. J., and Reed, S. I. (1998) *Genes Dev.* **12,** 2587–2597.