[CANCER RESEARCH 61, 4130–4135, May 15, 2001]

Fusion of a Novel Gene, *RCC17*, to the *TFE3* gene in t(X;17)(p11.2;q25.3)-bearing Papillary Renal Cell Carcinomas[1]

Pierre Heimann,[2] Hakim El Housni, Gönül Ogur, Marian A. J. Weterman, Elizabeth M. Petty, and Gilbert Vassart

Department of Medical Genetics and Institute of Interdisciplinary Research, Faculty of Medicine, Free University of Brussels, 1070 Brussels, Belgium [P. H., H. E. H., G. O., G. V.]; Department of Human Genetics 417, University Hospital Nijmegen, 6500 HB Nijmegen, the Netherlands [M. A. J. W.]; and Departments of Human Genetics and Internal Medicine, University of Michigan Health System, Ann Arbor, Michigan 48109-0638 [E. M. P.]

ABSTRACT

A subset of childhood and young adult renal cell carcinomas displays a recurrent translocation t(X;17)(p11;q25) as the sole cytogenetic abnormality. In two young girls, we demonstrate that this translocation results in the fusion of a novel gene, designated *RCC17*, at chromosome 17q25, to the transcription factor *TFE3* located on the Xp11 chromosomal region. In both cases, the t(X;17) fuses the NH_2-terminal region of *RCC17* to the COOH-terminal part of *TFE3* including the basic helix-loop-helix DNA-binding domain and the leucine zipper dimerization domain. The reciprocal fusion transcript *TFE3/RCC17* is also expressed. *RCC17* encodes a putative protein of 553 amino acids. It is ubiquitously expressed in normal adult tissues. No significant similarity was found with other fusion partners of *TFE3* or with any relevant functional protein domains, precluding informed speculation about the normal function of this gene.

INTRODUCTION

Sporadic RCCs[3] in children are rare, accounting for less than 6% of all childhood kidney tumors (1). They differ morphologically and genetically from those arising in adults, in which two main types of RCC are distinguished: clear cell carcinoma, or nontubulopapillary RCC, and its variants, which account for almost 80% of RCC and the tubulopapillary type, which represents 15–20% of adult RCC. These two histological types are associated with well-defined genetic abnormalities. In clear cell carcinomas, loss of heterozygosity (LOH) and mutation of the Von Hippel-Lindau (*VHL*) tumor suppressor gene located on chromosome 3p25 is often observed (2). Papillary RCCs generally exhibit recurrent numerical abnormalities of other chromosomes including tetrasomy 7, trisomy 10, 12, 16, 17, and/or loss of the Y chromosome (3). Overexpression of a candidate gene associated with the polysomy of chromosome 7 could be the *MET* proto-oncogene (4), which encodes the receptor for hepatocyte growth factor/scatter factor (*HGF/SF*; Ref. 5).

In childhood, sporadic RCCs are mainly (but not exclusively) of tubulopapillary type (6) and none of the genetic aberrations cited above have been thus far observed. Instead, anomalies in Xp11.2 region, such as the translocation t(X;1)(p11.2;q21) or its variants are recurrently reported (7). This t(X;1)translocation accounts for ~50% of the described cases (8) and results in the fusion of a novel gene on chromosome 1, called *PRCC*, to *TFE3* gene located on the X chromosome (9, 10). *TFE3* encodes a member of the basic helix-loop-helix family of transcription factors, which was originally demonstrated to bind to the μE3 element of the immunoglobulin heavy chain enhancer (11). *PRCC* is involved in the pre-mRNA splicing process (12).

Among the variants, translocations t(X;1)(p11.2;p34) and t(X;17)(p11.2;q25) have been the most frequently reported (8), with the former resulting in the fusion of the *TFE3* gene to the splicing factor gene *PSF* at chromosome 1p34 (13). In the present study, we have identified the gene involved in the chromosome 17q25 translocation and characterized the fusion transcript. This gene on 17q25 has been coined *RCC17*. *RCC17* contains a conserved UBX domain that is possibly involved in the ubiquitylation pathway but no relevant functional protein motif suggestive of its function.

MATERIALS AND METHODS

Case Reports. Case 1 was a five-year-old-female of African origin who presented with hematuria. An ultrasound scan revealed a solid mass of 14 mm at the lower pole of the right kidney. Partial nephrectomy with locoregional lymphadenectomy was performed. Histological examination demonstrated a grade III papillary RCC. One regional lymph node appeared to be invaded by the tumor. This mass was classified as $pT_{1a}N_1M_0$ according to the TNM staging system.

Case 2 was a five-year-old-female of African origin admitted to the hospital because of fever and pain in the right hypochondria. Ultrasound scan revealed a retroperitoneal mass associated with mild hydronephrosis as well as a small nodule 10 mm in diameter in the upper pole of the left kidney. Histological examination of the kidney nodule showed a grade II papillary RCC (Fig. 1), and the retroperitoneal mass corresponded to an infiltration of regional kidney lymph nodes by this papillary RCC. Further analysis confirmed that this tumor did not extend beyond the regional lymph nodes. This tumor was classified as $pT_{1a}N_1M_0$ according to the TNM staging system.

Cytogenetic Analysis. For both cases, short-term cultures were initiated after collagenase treatment of the tumoral tissue and used for chromosomal analysis. Culture conditions, harvesting, slide preparation, and G-banding were carried out as described previously (14). Karyotypes were expressed according to the ISCN 1995 (15). Thereafter, a long-term cell line containing the t(X;17)was obtained from case 1.

FISH. To delineate the breakpoint on chromosome 17q25, several P1 and BAC clones spanning the 17q25.2–25.3 chromosomal regions at (proximal to distal) D17S937, D17S939, D17S802, *TK1*, D17S836, *ERBA2L*, and *ILF1* were used (16). A 3.2-kb *TFE3* clone (9), a 1.8-kb *RCC17* clone (corresponding to the cDNA identified in the present study), and α satellite probes for chromosomes X and 17 (DXZ1 and D17Z1; Oncor, Gaithersburg, MD) were also used. All of the FISH analyses were performed as described previously (14) on metaphases from the long-term cell line containing the t(X;17) (case 1) and on metaphases from a short-term cell culture (case 2). Chromosomal preparations from normal peripheral leukocytes were used with the *RCC17* cDNA probe.

5′ RACE PCR. 5′ RACE PCR was performed on poly(A)+RNA extracted from case 1 using the 5′ RACE system for rapid amplification of cDNA ends kit (Life Technologies, Inc., Paisley, PA4 9RF, United Kingdom). Poly(A)+ RNA was isolated from the long-term cell line t(X;17) metaphases and using the Micro-FastTrack kit (Invitrogen, Carlsbad, CA). In brief, 500 ng of poly(A)+RNA were reverse transcribed with a *TFE3*-specific primer 5′-CAGTTCCTTGATCCTGTCG-3′. The first and second round PCRs were performed according to the manufacturer's instructions with the specific *TFE3* reverse primers 5′-CGCCTGCGACGCTCAATTAG-3′ and 5′-CGTTTGATGTTGGGCAGCTC-3′, respectively.

Received 12/13/00; accepted 3/20/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] P. H. is a recipient of the "Fondation Erasme."

[2] To whom requests for reprints should be addressed, at the Department of Medical Genetics and Institute of Interdisciplinary Research, Faculty of Medicine, Campus Erasme, 808 route de Lennik, 1070 Brussels, Belgium. Phone: 32-2-555-41-45; Fax: 32-2-555-42-12; E-mail: pheimann@ulb.ac.be.

[3] The abbreviations used are: RCC, renal cell carcinoma; FISH, fluorescence *in situ* hybridization; 5′ RACE PCR, 5′ rapid amplification of cDNA ends PCR; RT-PCR, reverse transcription-PCR; ASPS, alveolar soft part sarcoma; UBA and UBX, ubiquitin regulatory domain protein; TGF-β, transforming growth factor β; EST, expressed sequence tags; ISCN, international system for human cytogenetic nomenclature; MMLV, moloney murine leukemia virus; GAPDH, glyceraldehyde phosphate deshydrogenase.

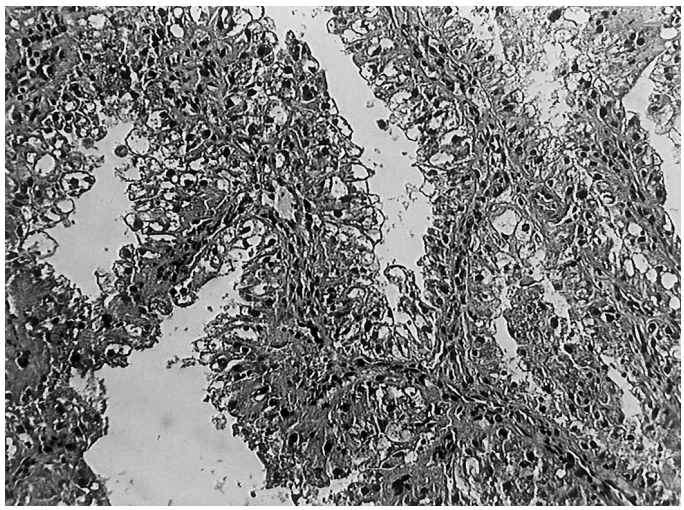

Fig. 1. Histology of case 2 showing the typical papillary formations observed in papillary RCC. Case 1 displayed the same morphological features.

RT-PCR. Total RNA was isolated from the long-term cell line containing the t(X;17) (case 1) and from a tumor sample (case 2) with the Trizol Reagent kit (Life Technologies, Inc.) and subsequently treated with DNase using a DNA-free kit (Ambion, Austin, TX). Reverse transcription was performed with the MMLV reverse transcriptase and random hexamers (Life Technologies, Inc.) according to the manufacturer's protocol. The *RCC17-TFE3* hybrid transcripts were detected by amplification with the *RCC17* primer 5′-GTTGCCGAAGTCCCTCTCCA-3′ (forward primer) and the *TFE3* primer 5′-CGCCTGCGACGCTCAATTAG-3′ (reverse primer). The reciprocal *TFE3-RCC17* hybrid transcripts were detected using the *TFE3* primer 5′-CGCAAGTGCCCAGCCACT-3′ (forward primer) and the *RCC17* primer 5′-GCCCCCTCCTCAGCAGCT-3′ (reverse primer). The presence of normal *RCC17* transcripts was assessed with primers 5′-GCCATTTTCCACAGATCAGG-3′ (forward primer) and 5′- GCTGGAGAGGGACTTCGG-3′ (reverse primer). Normal *TFE3* transcripts were detected using the same forward and reverse *TFE3* primers described above. For both *RCC17-TFE3* and *TFE3-RCC17* fusion transcripts, amplification conditions were 94°C for 60 s, 60°C for 30 s, and 72°C for 60 s (30 cycles) using the standard Taq DNA polymerase (Applied Biosystems, Foster City, CA) in a final volume of 50 μl. For normal *TFE3* transcripts, the PCR conditions were similar except for the fact that 35 cycles were performed. PCR amplifications for the normal *RCC17* transcript were done with the PCR Qiagen kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Amplification conditions were 94°C for 60 s, 55°C for 30 s, and 72°C for 60 s (35 cycles) in a final volume of 50 μl. The products were separated by electrophoresis in a 2% agarose gel followed by staining with ethidium bromide.

Northern Blot Analysis. An-839 bp fragment, representing the 5′ part of the cDNA identified in this study, was used as a probe to screen various adult human tissue poly(A)+RNAs including heart, brain, liver, skeletal muscle, kidney, pancreas, and placenta.

DNA Sequencing. For sequence analysis, PCR products were either subcloned with the TA Cloning kit (Invitrogen) following instructions of the manufacturer or were sequenced directly after having been isolated by electrophoresis in 1% agarose gels and purified with the Geneclean II kit (BIO101, Vista, CA) when necessary. The 5′ Race PCR product from case 1, the *RCC17-TFE3* RT-PCR product from case 2, and both strands of the *RCC17* gene cDNA were sequenced with the dideoxymethod using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). The different primers used for sequencing were the M13 forward and reverse primers (for subcloned PCR products), the *TFE3* reverse primer 5′-CGCCTGCGACGCTCAATTAG-3′, and several *RCC17* primers given in Table 1.

RESULTS

Karyotypic Analysis and Rearrangement of *TFE3*. Cytogenetic analysis demonstrated a karyotype of 46, t(X;17)(p11;q25)[20] for case 1 and a karyotype of 46, t(X;17)(p11;q25)[30]/46,XX[15] for case 2 (data not shown). In both cases, dual color FISH showed that the breakpoint on chromosome 17q25 lies between the *ERBA2L* gene located proximally and the *ILF1* gene located distally, hence demonstrating that the breakpoint is located in the 17q25.3 subband (data not shown). FISH analysis on t(X;17) metaphases from both cases with the *TFE3* cDNA probe, showed splitting of this probe, hence confirming the involvement of *TFE3* gene in the chromosomal rearrangement (data not shown).

Identification of the *RCC17* Gene. Sequence of the 5′ Race product from case 1 diverged from the normal *TFE3* sequence at position corresponding to the *TFE3* exon3-exon4 junction. Searches of the EMBL databases showed that the unknown 5′sequence had 339 bp of overlap with a 183-kb homo sapiens chromosome clone RP11-498C9 consisting of 16 unordered contigs (GenBank accession no. ACO69004), from which 71 bp of additional sequence at the 5′ end of the unknown sequence were also identified (Fig. 2). These additional nucleotides completed the 5′ end of the coding sequence because they contained an ATG codon in agreement with the Kozak consensus sequence (17) and displayed an in-frame TGA stop codon located 36 bp upstream of the initiation triplet. The 3′ end of the novel sequence had 113 bp of overlap with a 227-bp EST of human colon (Stratagene, GenBank accession no. AA056378) which, in turn, matched with a 409-bp human placental EST (GenBank accession no. R62723). This last one yielded 382 bp of additional sequence on the 3′end of our unknown sequence and showed 255 bp of overlap with a 387 bp Soares breast homo sapiens cDNA clone (GenBank accession no. H51880). A total of 17 additional overlapping ESTs were identified corresponding to the 3′ end of the cDNA (GenBank accession no. AW302150, AW302160, W63713, AA303110, AW194381, AI923728, AA485942, AA972192, AI024586, AI825028, AI985056, AI031531, AI625114, AI826958, AA766699, AA782421, and AW452860). Altogether, they gave an additional sequence of 642 bp and completed the 3′ end of the unknown cDNA including the poly(A) tail. To test the colinearity of the unknown 5′ Race product and the 17 additional ESTs, we performed two independent PCRs on human fetal brain cDNA (Clontech, Palo Alto, CA) and human placental cDNA, using a forward and a reverse primer located at position 804–823 (corresponding to the 3′ end of the unknown

Table 1 *List of forward and reverse primers used to sequence both of the strands of the RCC17 cDNA*

All of the primer locations are given according to the RCC17 coding sequence (starting from the initiation codon). All of the primers are within the open reading frame (ORF) except CH17S1, CH17S2, and CH17R8.

Primers name	Primers sequence	Primers location
CH17S1	5′-CGTCCTGGCTGTTGCCGA-3′	[a]
CH17S2	5′-GTCACGTGAGCGGAAAATGG-3′	[a]
CH17S3	5′-AGAGGAGCGTGCTCGACC-3′	nt 164–181
CH17R1	5′-ATGCGAACCATGTTCTCAGG-3′	nt 212–231
CH17S4	5′-GCCATTTTCCACAGATCAGG-3′	nt 356–375
CH17R2	5′-ACAAACCTGATGGTGGCG-3′	nt 495–512
CH17S5	5′-AGTGCTACGACCCCGTGG-3′	nt 518–535
CH17S6	5′-TGCTGCGAGCACACTCAG-3′	nt 670–687
CH17S7	5′-GTTGCCGAAGTCCCTCTCCA-3′	nt 804–823
CH17R3	5′-TGGAGAGGGACTTCGGCAAC-3′	nt 804–823
CH17S8	5′-CAGCGGAGCTGCCTGATG-3′	nt 1010–1027
CH17R4	5′-CAAGGGGGCTTCTTCCAG-3′	nt 1096–1113
CH17S8	5′-GGAGCCACCTGGGGAACC-3′	nt 1262–1279
CH17R5	5′-GGAGATGGCATGCTCCAG-3′	nt 1429–1446
CH17R6	5′-GCCCCCTCCTCAGCAGCT-3′	nt 1545–1562
CH17R7	5′-GCCCAGCTCCTCTTCACG-3′	nt 1602–1620
CH17R8	5′-ATTCCTGCTGGCAGAGGAGG-3′	[b]

[a] Primers CH17S1 and CH17S2 have been chosen within the human clone RP11-498C9 upstream of the RCC17 ORF. CH17S1 and CH17S2 start 64 and 16 nucleotides (nt) upstream of the initiation codon, respectively. CH17S2 overlaps the initiation codon at its 3′ extremity.

[b] Primer CH17R8 starts 43 nt downstream of the STOP codon.

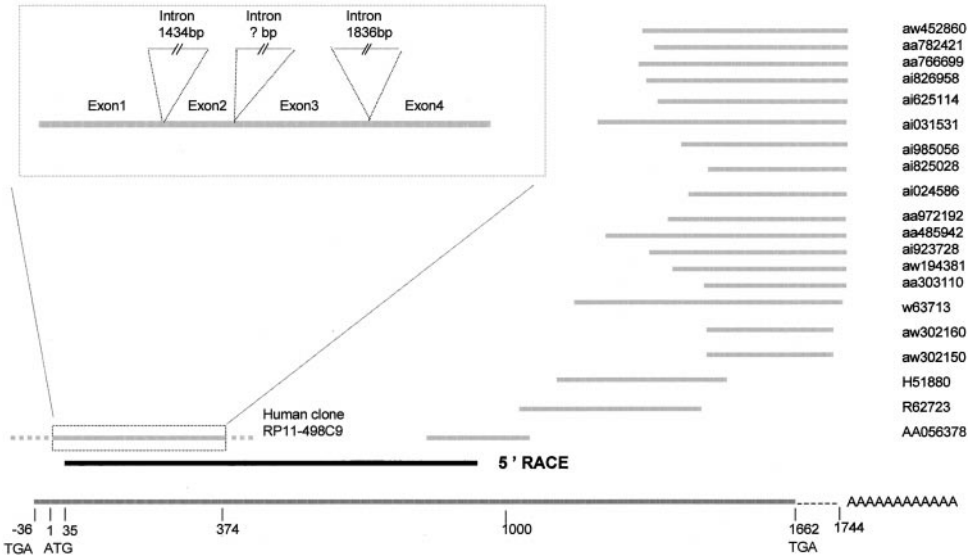

Fig. 2. Schematic representation of the RCC17 cDNA contig. The 5′ RACE product (*bold line*) measured 898 bp and extended up to 35 bp downstream of the ATG codon. The 5′ portion of the cDNA contig including the start codon was deduced from the assembly of the unknown clone from the 5′ RACE product with 71 bp from the 18-kb human chromosomal clone RP11-498C9. Twenty overlapping ESTs allowed identification of the 3′ portion of the unknown sequence. This composite sequence contains the RCC17 open reading frame extending from nucleotide 1 (*ATG* codon) to 1662 (*TGA* codon). Sequence comparison between the human chromosomal clone RP11-498C9 and the first 374 bp of the unknown cDNA sequence allowed to determine the location of the first four exons as well as the exon/intron boundaries between exons 1 and 2, and exons 3 and 4 ("See Results").

cDNA) and position 1545–1562 of the composite sequence, respectively. For both tissues, we successfully amplified a fragment of ~760 bp that was cloned and sequenced. The nucleic acid sequences obtained were similar to the 3′ part of the composite sequence. The 71 additional bp yielded by the human chromosomal clone RP11-498C9, the unknown clone from the 5′ Race PCR product and the different overlapping ESTs, were then assembled into one single contig resulting in an open reading frame of 1662 nucleotides (Fig. 2). Several PCRs were performed from human placental cDNA with primers scattered along the whole length of this sequence and giving overlap amplicons of 300–400 bp. Direct sequencing of each amplicon was performed, yielding a contig with continuous open reading frame. The corresponding gene encoding a 553-amino acid putative protein was named *RCC17* (for *RCC* chromosome *17*; Fig. 3).

A full-length *RCC17* cDNA clone was isolated from human placental cDNA and used as a probe for FISH analysis on metaphase preparations from normal peripheral leukocytes. This probe allowed the localization of the *RCC17* gene to the very distal part of chromosome 17q, within the 17q25.3 subband (Fig. 4). Sequence comparisons of the predicted *RCC17* protein using the BLAST software and the PROSITE database did not show significant similarity to any known proteins, including other fusion partners of *TFE3*, namely *PRCC*, *PSF*, and *NonO*. However, searches with the SMART website tool (18) revealed within the 3′ portion of *RCC17* (residues 382–464) a significant similarity with a UBX domain (Fig. 3), a poorly characterized motif present in some ubiquitin-regulatory proteins. Moreover, three P-X-X-P motifs, which are known to bind proteins with SH3 domains (19), were observed at residues 241–244, 258–261, and 276–279 (Fig. 3).

Partial Genomic Structures of *RCC17*. A comparison of the first 374 bp of RCC17 cDNA with the human chromosomal clone RP11-498C9 allowed determination of the first four exonic sequences of RCC17 (see Fig. 2). The first and second exons were localized within the contig nr 8 (assembly name) of RP11-498C9 and separated by an intronic sequence of 1434 bp. Starting from the initiation codon, exon 1 extended down to nucleotide 102, whereas exon 2 had a length of 56 bp. The third and fourth exons were localized within the contig nr 12 (assembly name) of RP11-498C9 and were separated by an intron of 1836 bp. Exons 3 and 4 were 115 and 101 bp long, respectively.

Expression of *RCC17-TFE3* Transcripts. The presence of a RCC17-TFE3 hybrid transcript was demonstrated by RT-PCR using a 5′ *RCC17* primer and a 3′ *TFE3* primer. A product of the predicted size (321 bp) was obtained in both PRCC cases but not in several cancer cell lines including K562 (chronic myelogenous leukemia cell line), HL60 (acute myeloid leukemia cell line), and MZ2mel (melanoma cell line; Fig. 5). Sequencing of the RT-PCR product of case 2 demonstrated an identical *RCC17-TFE3* junction as for the 5′ Race PCR product obtained in case 1. Thus, in both cases, the 311-amino acids NH_2-terminal region of *RCC17* was fused to the COOH-terminal DNA binding domain (bHLH and LZ regions) of *TFE3*. Reciprocal *TFE3-RCC17* hybrid transcripts were also detected in both cases (data not shown).

Expression of Normal *TFE3* and *RCC17* Transcripts. Detection of both normal *TFE3* and *RCC17* transcripts was performed by RT-PCR on both RCC cases and on the three cancer cell lines mentioned above. For *TFE3* and *RCC17*, we used primers flanking the respective breakpoint. For both genes, amplicons of the predicted size (411 bp for *TFE3* and 468 bp for *RCC17*) were observed in samples from all of the cell types (Fig. 5). The distribution of *RCC17* gene expression was explored by Northern blotting and RT-PCR. Hybridization of a *RCC17* probe to a Northern blot with a collection of poly(A)+RNA from various tissues detected a ~2-kb transcript weakly expressed in heart, placenta, liver, skeletal muscle, kidney, and pancreas but not in brain (Fig. 6). The *RCC17* expression study was completed by RT-PCR on various human adult tissue cDNA (Clontech) including whole brain with the same *RCC17* primers as mentioned above. An amplicon of the expected size (468 bp) was observed in all of the samples, demonstrating ubiquitous expression of the *RCC17* gene (Fig. 7). Moreover, a Northern blot analysis performed on fetal brain demonstrated a ~2-kb transcript also (data not shown).

DISCUSSION

In the present study, we have demonstrated that the t(X;17)(p11.2; q25.3) translocation observed in a subset of papillary RCCs results in the fusion of a novel gene on chromosome 17, which we have called *RCC17*, to the *TFE3* transcription factor gene located on chromosome X. In both cases studied, translocation results in the fusion of the 311 first amino acids of the NH_2-terminal part of *RCC17* to the COOH-terminal part of *TFE3*, including the basic helix-loop-helix DNA binding domain and the leucine zipper dimerization domain. Surprisingly, there was no loss of normal *TFE3* transcripts. In case 2, the persistence of *TFE3* expression could be attributable to contamination

```
GGAAAATGGCGGCCCCGGCAGGCGGCGGAGGCTCCGCGGTGTCGGTGCTGGCCCCGAACG 60
     M  A  A  P  A  G  G  G  G  S  A  V  S  V  L  A  P  N   18

GCCGGCGCCACACGGTGAAGGTGACGCCGAGCACCGTGCTGCTTCAGGTTCTGGAGGACA 120
G  R  R  H  T  V  K  V  T  P  S  T  V  L  L  Q  V  L  E  D   38

CGTGCCGGCGGCAGGACTTCAACCCCTGTGAATATGATCTGAAGTTTCAGAGGAGCGTGC 180
T  C  R  R  Q  D  F  N  P  C  E  Y  D  L  K  F  Q  R  S  V   58

TCGACCTTTCTCTCCAGTGGAGATTTGCCAACCTGCCCAACAATGCCAAGCTGGAGATGG 240
L  D  L  S  L  Q  W  R  F  A  N  L  P  N  N  A  K  L  E  M   78

TGCCCGCTTCCCGGAGCCGTGAGGGGCCTGAGAACATGGTTCGCATCGCTTTGCAGCTGG 300
V  P  A  S  R  S  R  E  G  P  E  N  M  V  R  I  A  L  Q  L   98

ACGATGGCTCGAGGTTGCAGGACTCTTTCTGTTCAGGCCAGACCCTCTGGGAGCTTCTCA 360
D  D  G  S  R  L  Q  D  S  F  C  S  G  Q  T  L  W  E  L  L   118

GCCATTTTCCACAGATCAGGGAGTGCCTGCAGCACCCCGGCGGGGCCACCCCAGTCTGCG 420
S  H  F  P  Q  I  R  E  C  L  Q  H  P  G  G  A  T  P  V  C   138

TGTACACGAGGGATGAGGTGACGGGTGAAGCTGCCCTGCGGGGCACGACGCTGCAGTCGC 480
V  Y  T  R  D  E  V  T  G  E  A  A  L  R  G  T  T  L  Q  S   158

TGGGCCTGACCGGGGGCAGCGCCACCATCAGGTTTGTCATGAAGTGCTACGACCCCGTGG 540
L  G  L  T  G  G  S  A  T  I  R  F  V  M  K  C  Y  D  P  V   178

GCAAGACCCCAGGAAGCCTGGGCTCGTCAGCGTCGGCTGGCCAGGCAGCCGCCAGCGCTC 600
G  K  T  P  G  S  L  G  S  S  A  S  A  G  Q  A  A  A  S  A   198

CACTTCCCTTGGAATCTGGGGAGCTCAGCCGCGGCGACTTGAGCCGTCCGGAGGACGCGG 660
P  L  P  L  E  S  G  E  L  S  R  G  D  L  S  R  P  E  D  A   218

ACACCTCAGGGCCCTGCTGCGAGCACACTCAGGAGAAGCAGAGCACAAGGGCACCCGCAG 720
D  T  S  G  P  C  C  E  H  T  Q  E  K  Q  S  T  R  A  P  A   238

CTGCCCCCTTTGTTCCTTTCTCGGGTGGGGGACAGAGACTGGGGGGCCCTCCTGGGCCCA 780
A  A  P  F  V  P  F  S  G  G  G  Q  R  L  G  G  P  P  G  P   258

CGAGGCCTCTGACATCATCTTCAGCTAAGTTGCCGAAGTCCCTCTCCAGCCCTGGAGGCC 840
T  R  P  L  T  S  S  S  A  K  L  P  K  S  L  S  S  P  G  G   278

CCTCCAAGCCAAAGAAGTCCAAGTCGGGCCAGGATCCCCAGCAGGAGCAGGAGCAGGAGC 900
P  S  K  P  K  K  S  K  S  G  Q  D  P  Q  Q  E  Q  E  Q  E   298

GGGAGCGGGATCCCCAGCAGGAGCAGGAGCGGGAGCGGCCCGTGGACCGGGAGCCCGTGG 960
R  E  R  D  P  Q  Q  E  Q  E  R  E  R  P  V  D  R  E  P  V   318

ACCGGGAGCCGGTGGTGTGCCACCCCGACCTGGAGGAGCGGCTGCAGGCCTGGCCAGCGG 1020
D  R  E  P  V  V  C  H  P  D  L  E  E  R  L  Q  A  W  P  A   338

AGCTGCCTGATGAGTTCTTTGAGCTGACGGTGGACGACGTGAGAAGACGCTTGGCCCAGC 1080
E  L  P  D  E  F  F  E  L  T  V  D  D  V  R  R  R  L  A  Q   358

TCAAGAGTGAGCGGAAGCGCCTGGAAGAAGCCCCCTTGGTGACCAAGGCCTTCAGGGAGG 1140
L  K  S  E  R  K  R  L  E  E  A  P  L  V  T  K  A  F  R  E   378

CGCAGATAAAGGAGAAGCTGGAGCGCTACCCAAAGGTGGCTCTGAGGGTCCTGTTCCCCG 1200
A  Q  I  K  E  K  L  E  R  Y  P  K  V  A  L  R  V  L  F  P   398

ACCGCTACGTCCTACAGGGCTTCTTCCGCCCCAGCGAGACAGTGGGGGACTTGCGAGACT 1260
D  R  Y  V  L  Q  G  F  F  R  P  S  E  T  V  G  D  L  R  D   418

TCGTGAGGAGCCACCTGGGGAACCCCGAGCTGTCATTTTACCTGTTCATCACCCCTCCAA 1320
F  V  R  S  H  L  G  N  P  E  L  S  F  Y  L  F  I  T  P  P   438

AAACAGTCCTGGACGACCACACGCAGACCCTCTTTCAGGCGAACCTCTTCCCGGCCGCTC 1380
K  T  V  L  D  D  H  T  Q  T  L  F  Q  A  N  L  F  P  A  A   458

TGGTGCACTTGGGAGCCGAGGAGCCGGCAGGTGTCTACCTGGAGCCTGGCCTGCTGGAGC 1440
L  V  H  L  G  A  E  E  P  A  G  V  Y  L  E  P  G  L  L  E   478

ATGCCATCTCCCCATCTGCGGCCGACGTGCTGGTGGCCAGGTACATGTCCAGGGCCGCCG 1500
H  A  I  S  P  S  A  A  D  V  L  V  A  R  Y  M  S  R  A  A   498

GGTCCCCTTCCCCATTGCCAGCCCCTGACCCTGCACCTAAGTCTGAGCCAGCTGCTGAGG 1560
G  S  P  S  P  L  P  A  P  D  P  A  P  K  S  E  P  A  A  E   518

AGGGGGCGCTGGTCCCCCCTGAGCCCATCCCAGGGACGGCCCAGCCCGTGAAGAGGAGCC 1620
E  G  A  L  V  P  P  E  P  I  P  G  T  A  Q  P  V  K  R  S   538

TGGGCAAGGTGCCCAAGTGGCTGAAGCTGCCGGCCAGCAAGAGGTGAGAGCTGCCAGCCT 1680
L  G  K  V  P  K  W  L  K  L  P  A  S  K  R  *               553

GAGGTGCCCACTCCG 1695
```

Fig. 3. The *RCC17* cDNA sequence and predicted amino acid sequence. *Vertical arrow*, the breakpoint location, which was identical for both of the t(X;17) cases. *Italic boldface*, the P-X-X-P motifs. *Boldface*, the segment of the putative protein showing a UBX domain profile.

by normal cell because the RT-PCRs were performed on a tumor fragment containing, most likely, clusters of nontumor cells. In case 1, this observation remains intriguing because a pure t(X;17) cell line was used. If we assume that the translocation involved the active X chromosome, the remaining normal *TFE3* allele should be silent according to lyonization and, one of the explanations for the persistent *TFE3* expression observed here could be loss of inactivation of the normal copy. However, it is also known that some genes escape X-chromosome inactivation. *TFE3* could be one of them. Alternatively, it is possible that the Xp11 translocation would involve randomly the active or inactive *TFE3* allele with no disruption of the inactivation mechanism. If the rearrangement would occur on the inactive chromosome, reexpression of the *TFE3* COOH-terminal part could be attributable to the promoter of the partner gene located on 5′ of the fusion transcript. Finally, because more than 50% of patients carrying a Xp11 rearrangement are males, one may speculate that the functional integrity of the wild type *TFE3* gene is not essential to the viability of the tumoral cells.

The *TFE3* gene belongs to the bHLH-Zip family of transcription factors and has been initially identified as a protein binding to the μE3 element of the immunoglobulin heavy-chain enhancer (11). However, its ubiquitous expression indicates that this gene must play a much broader role in transcriptional control. For example, it has recently been suggested that the *TFE3* gene may act on phosphate transport in the kidney by regulating transcription of the type II sodium-dependent phosphate (Na/P_i) cotransporters (NPT2; Ref. 20). These are major transporters associated with renal P_i reabsorption and are located at the apical membrane of renal proximal tubular cells. *TFE3* would also have a function in the regulation of melanogenesis through the control of tyrosinase and *Tyrp1* gene expression (21). Moreover, *TFE3*, in

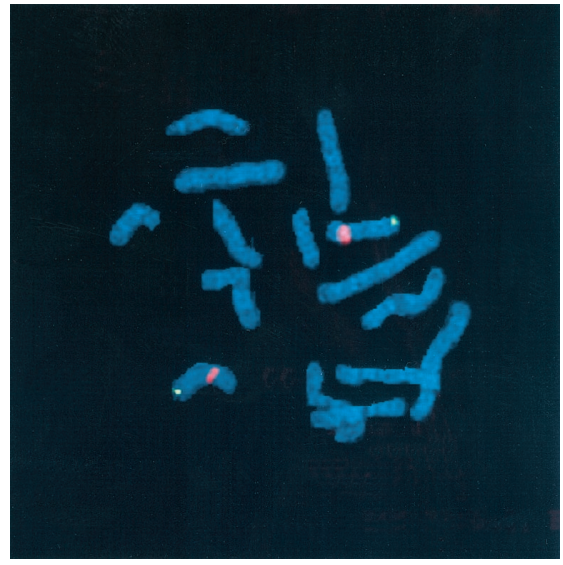

Fig. 4. FISH on a metaphase preparation from normal human peripheral leukocytes. The *RCC17* gene was localized on the 17q25.3 subband using a full-length *RCC17* cDNA clone as a probe (*green*). Both of the chromosomes 17 were identified with an α satellite chromosome 17 probe (D17Z1-Oncor; *red*).

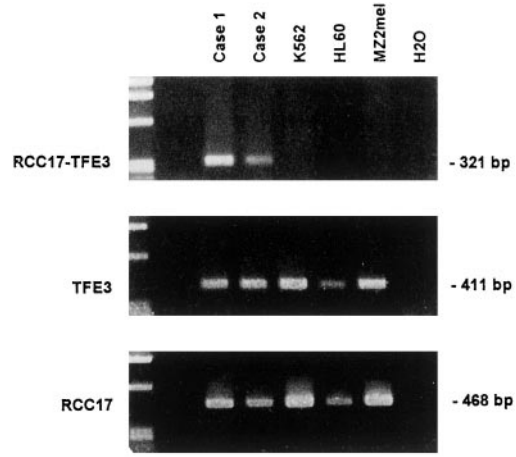

Fig. 5. Detection of *RCC17-TFE3* fusion transcripts by RT-PCR (*upper panel*). An amplicon of the expected size was obtained for both t(X;17) cases but not in three cancer cell lines including K562 (chronic myelogenous leukemia cell line), HL60 (acute myeloid leukemia cell line), or MZ2mel (melanoma cell line). Normal TFE3 and RCC17 transcripts were detected in all of the samples (*middle* and *lower panels*, respectively).

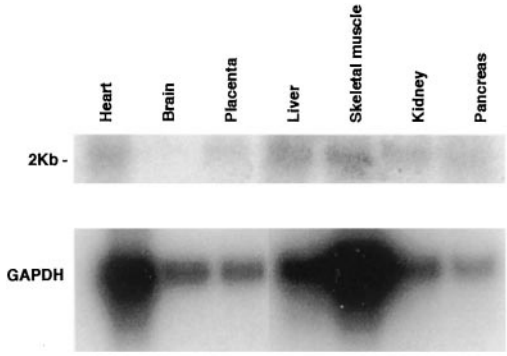

Fig. 6. Multiple tissue Northern blot analysis. A 837-bp fragment, representing the 5′ portion of RCC17, was used as a probe. *Upper panel*, A ~2-kb transcript was weakly expressed in the heart, placenta, liver, skeletal muscle, kidney, and pancreas but not in whole brain. *Lower panel*, GAPDH control probe.

cooperation with the SMAD3 protein, and through the *TGF-β*-activated signal transduction pathway, activates the plasminogen activator inhibitor-1 gene (*PAI1*; Ref. 22); a factor involved in cancer cell invasion and angiogenesis (23).

The presence of some P-X-X-P motifs in *RCC17* suggests that the encoded protein could bind proteins with SH3 modules; domains that are known to mediate protein-protein interactions. (19). Because these P-X-X-P motifs are included in the 5′ *RCC17* fragment that fuses to *TFE3*, the *RCC17/TFE3* chimeric transcript would enact novel regulatory interactions of *TFE3* with SH3-containing proteins. *RCC17* contains also an UBX domain located within its COOH-terminal portion. The UBX domain is observed in some human proteins such as hFAF1 (24), REP8 (25), and the predicted p47 protein (GenBank accession no. NM 018839.1). The functional properties of this motif remains unclear, but its cooccurrence with the ubiquitin-regulatory UBA domain (26) in some proteins [human predicted protein LOC51035 (GenBank accession no. XM 012093.1) and HUMORF (GenBank accession no. M68864)] suggests that it is involved in the ubiquitylation pathway.

Finally, *RCC17* does not possess any peculiar domains suggestive of its function. Nevertheless, its role could tentatively be predicted from the functional data available from the other *TFE3* fusion genes. Indeed, *PRCC*, *PSF*, and *NonO* are all involved in the pre-mRNA splicing process (12, 13), and, alteration of their function as a possible consequence of fusion to *TFE3* could result in inappropriate splicing of specific transcripts and in tumorigenesis (12). Although *RCC17* does not show any homology with other *TFE3* fusion partners, it may be hypothesized that *RCC17* could also represent a splicing factor.

Because both the *RCC17/TFE3* and *TFE3/RCC17* fusion transcripts were expressed in the tumoral tissue, their respective role in carcinogenesis remains to elucidate. However, it has been demonstrated by means of Zebra-luciferase reporter assays that the *PRCC/TFE3* fusion protein in t(X;1)(p11;q21)papillary RCC, has a transactivating capability that is 3-fold higher than for the wild-type *TFE3*, whereas the activation of the reciprocal product *TFE3/PRCC* is not significantly elevated, compared with that caused by the normal PRCC protein (27). This suggests that generation of the *PRCC/TFE3* fusion protein (and not of the reciprocal product) is the most critical factor for tumor development.

An interesting point would be to determine whether the various partner genes fused with *TFE3* have a different functional importance and to determine their possible influence on tumor morphology and clinical outcome. In human cancer exhibiting variant translocations (*i.e.*, one common gene fused to different possible partner genes), there are several examples demonstrating a significant relation between the type of fusion and the morphological findings as well as the prognosis. In synovial sarcoma a para-articular soft tissue tumor of unknown histogenesis, the morphological aspects vary according to the *SYT-SSX1* or *SYT-SSX2* fusion transcript (28). Moreover, the metastasis-free survival is significantly longer among patients with the *SYT-SSX2* fusion transcript. Clinical differences attributable to genetic heterogeneity are also observed in alveolar rhabdomyosarcoma, an aggressive pediatric soft-tissue tumor with striated muscle differentiation. Indeed, tumors with the *PAX7-FKHR* fusion more often present as a localized disease and are associated with longer event-free survival as compared with the *PAX3-FKHR* tumors (29).

Of note, the *TFE3* gene displays some variability of genomic breakpoints among the different types of Xp11 fusion transcripts. In our two cases, the breakpoint was similar to that observed in *NonO-TFE3* and *PSF-TFE3* hybrid transcripts [corresponding to the t(X;1)(p11.2;p34) and inv(X)(p11.2;q12) bearing RCC, respectively], and localized between exons 3 and 4 of the *TFE3* gene. The fusion protein generated lacks the *TFE3* acidic activation domain, unlike the *PRCC-TFE3* fusion type [corresponding to the translocation t(X;1)(p11;q21)], which retains the *TFE3* acidic activation domain, because the breakpoint is localized at the exon 1–2 junction. It may be that this structural heterogeneity has no functional significance. However, it has been demonstrated in Ewing sarcoma, a primitive neuroectodermal tumor of children and young adults, that structural variations in the *EWS-FLI1* fusion gene correlated closely with the various clinical behaviors exhibited by this neoplasm (30).

Molecular studies on large series of childhood RCC are needed to determine whether fusion of the *TFE3* gene to different partners as well as various *TFE3* breakpoints correlate with particular morphological and clinical phenotypes.

Surprisingly, after initial submission of the present article, it has been reported that the same t(X;17)fusion transcript was present in the ASPS (31), a soft tissue tumor of unclear histogenesis that shows histological features and clinical behaviors different from the RCCs. The *TFE3* breakpoint in ASPS was in the majority of cases similar to the one observed in t(X;17) RCCs, *i.e.*, located between *TFE3* exons 3 and 4. In few cases, this breakpoint occurred within intron 2. The novel gene on chromosome 17q25 was designated *ASPL* (for ASPS locus) and showed the same breakpoint as the *RCC17* gene. *ASPL* matched perfectly the *RCC17* sequence except for the absence in *RCC17* of a 90-nucleotide segment corresponding to the 3′ portion of *ASPL* exon 2. When we had compared the 5′ end of *RCC17* with the human chromosome clone RP11-498C9 (see "Results" and Fig. 2), this 90-nucleotide fragment was considered as part of the RCC17 intron 2 in accordance with appropriate exon/intron junctions. This difference, which may correspond to alternative splicing, led Ladanyi *et al.* (31) to assign the initiating codon of *ASPL* 321 nucleotides downstream as compared with the initiating ATG of *RCC17*. The existence of the *RCC17* sequence as determined here is confirmed by numerous additional ESTs [including, *e.g.*, a pediatric acute myelog-

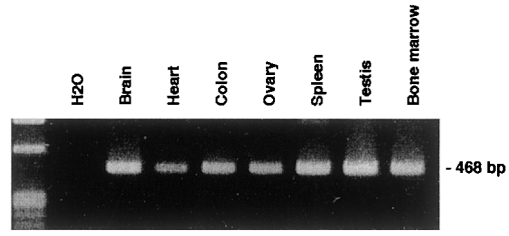

Fig. 7. Detection of normal RCC17 transcript by RT-PCR on various human adult tissue cDNA samples. An amplicon of the expected size (468 bp) was observed in all of the samples.

enous leukemia cell cDNA clone (GenBank accession no. BE 241763.1) and a Soares breast 3Nb HBst homo-sapiens cDNA clone (GenBank accession no. H45871.1)] showing perfect alignment with the 5′ segment of *RCC17*.

The implication of identical cytogenetic and molecular anomalies in tumors of different sites is not unprecedented. In congenital mesoblastic nephroma (CMN) and congenital fibrosarcoma (CFS), the same *ETV6-NTRK3* fusion transcript has been identified (32). Both tumors share a very similar histological pattern, and the discovery of an identical fusion transcript gives strong evidence that they represent a single neoplastic entity occurring in different sites (32). In PRCC, the t(X;17)(p11;q25) translocation is reciprocal, whereas it seems to be unbalanced in ASPS (14, 33). It is conceivable that in ASPS, the occurrence of a chimerical fusion gene would be associated with a still-undefined loss and/or gain of function, thus accounting for the different phenotypes of ASPS and PRCC.

REFERENCES

1. Pratt, C. P., Douglass, E. C. Management of the less common cancers of childhood. *In:* P. A Pizzo, and D. G. Poplack (eds.), Principles and Practice of Pediatric Oncology, pp.765–767. Philadelphia: J. B. Lippincott Company, 1989.
2. Gnarra, J. R., Tory, K., Weng, Y., Schmidt, L., Wei, M. H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F., Lubensky, I. A., Duan, R., Florence, C., Pozatti, R., Walther, M. M., Bander, N. H., Grossman, H. B., Brauch, H., Pomer, S., Brooks, J. D., Issacs, W. B., Lerman, M. I., Zbar, B., and Linehan, W. M. Mutation of the *VHL* tumour suppressor gene in renal carcinoma. Nat. Genet., *7:* 85–90, 1994.
3. van den Berg, E., Dijkhuizen, T., Oosterhuis, J. W., Geurts van Kessel, A., de Jonk, B., and Störkel, S. Cytogenetic classification of renal cell cancer. Cancer Genet. Cytogenet., *95:* 103–107, 1997.
4. Glukhova, L., Goguel, A. F., Chudoba, I., Angevin, E., Pavon, C., Terrier-Lacombe, M. J., Meddeb, M., Escudier, B., and Bernheim, A. Overrepresentation of 7q31 and 17q in renal cell carcinoma. Genes Chromosomes Cancer, *5:* 348–356, 1998.
5. Naldini, L., Vigna, E., Narsimhan, R. P., Gaudino, G., Zarnegar, R., Michalopoulos, G. K., and Comoglio, P. M. Hepatocyte growth factor (HGF) stimulates the tyrosine kinase activity of the receptor encoded by the proto-oncogene c-*MET*. Oncogene, *6:* 501–504, 1991.
6. Renshaw, A. A., Granter, S., Fletcher, J. A., Kozakewitch, H., Corless, C., and Perez-Atayde, A. Renal cell carcinoma in children and young adults: increased incidence of papillary architecture and unique subtypes. Am. J. Surg. Pathol., *23:* 795–802, 1999.
7. Dijkhuizen, T., van den Berg, E., Störkel, S., Terpe, H-J., Bürger, H., and de Jong, B. Distinct features for chromophilic renal cell cancer with Xp11.2 breakpoints. Cancer Genet. Cytogenet., *104:* 74–76, 1998.
8. Désangles, F., Camparo, P., Fouet, C., Houlgatte, A., and Arborio, M. Translocation (X;1) associated with a nonpapillary carcinoma in a young woman: a new definition for an Xp11.2 RCC subtype. Cancer Genet. Cytogenet., *113:* 141–144, 1999.
9. Weterman, M. A. J., Wilbrink, M., and Geurts van Kessel, A. Fusion of the transcription factor *TFE3* gene to a novel gene, *PRCC*, in t(X;1)(p11;q21)-positive papillary renal cell carcinomas. Proc. Natl. Acad. Sci, USA, *93:* 15294–15298, 1996.
10. Sidhar, S., Clark, J., Gill, S., Hamoudi, R., Crew, A. J., Gwilliam, R., Ross, M., Linehan, W. M., Birdsall, S., Shipley, J., and Cooper, C. S. The t(X;1)(p11.2;q21.1) translocation in papillary renal cell carcinoma fuses a novel gene *PRCC* to the *TFE3* transcription factor gene. Hum. Mol. Genet., *5:* 1333–1338, 1996.
11. Beckman, H., Su, L., and Kadesch, T. A helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Gene Dev., *4:* 167–179, 1990.
12. Skalsky, Y. M., Ajuh, P. M., Parker, C., Lamond, A. I., Goodwin, G., and Cooper, C. S. *PRCC*, the commonest *TFE3* fusion partner in papillary renal carcinoma is associated with pre-mRNA splicing factors. Oncogene, *20:* 178–187, 2001.
13. Clark, J., Lu, Y. J., Sidhar, S. K., Parker, C., Gill, S., Smedley, D., Hamoudi, R., Linehan, W., Shipley, J., and Cooper, C. Fusion of splicing factor genes *PSF* and *NonO* (*p54nrb*) to the *TFE3* gene in papillary renal cell carcinoma. Oncogene, *15:* 2233–2239, 1997.
14. Heimann, P., Devalck, C., Debusscher, C., Sariban, E., and Vamos, E. Alveolar soft-part sarcoma: further evidence by FISH for the involvement of chromosome 17q25. Genes Chromosomes Cancer, *23:* 194–197, 1998.
15. ISCN. F. Mitelman (ed.), An International System for Human Cytogenetic Nomenclature. Basel: Karger, 1995.
16. Kalikin, L. M., George, R. A. V., Keller, M. P., Bort, S., Bowler, N. S., Law, D. J., Chance, P. F., and Petty, E. M. An integrated physical and gene map of human distal chromosome 17q24-proximal 17q25 encompassing multiple disease loci. Genomics, *57:* 36–42, 1999.
17. Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell, *15:* 1109–1123, 1978.
18. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA, *95:* 5857–5864, 1998.
19. Sudol, M. From Src homology domains to other signaling modules: proposal of the "protein recognition code." Oncogene, *17*: 1469–1474, 1998.
20. Kido, S., Miyamoto, K., Mizobush, H., Taketani, Y., Ohkido, I., Ogawa, N., Kaneko, Y., Harashima, S., and Takeda, E. Identification of regulatory sequences and binding proteins in the Type II Sodium/Phosphate cotransporter NPT2 gene responsive to dietary phosphate. J. Biol. Chem., *274:* 28256–28263, 1999.
21. Verastegui, C., Bertolotto, C., Bille, K., Abbe, P., Ortonne, J. P., and Balotti, R. TFE3, a transcription factor homologous to microphthalmia, is a potential transcriptional activator of tyrosinase and TyrpI genes. Mol. Endocrinol., *14:* 449–456, 2000.
22. Hua, X., Miller, Z. A., Wu, G., Shi, Y., and Lodish, H. F. Specificity in transforming growth factor beta-induced transcription of the plasminogen activator inhibitor-1 gene: Interaction of promoter DNA, transcription factor μE3, and Smad protein. Proc. Natl. Acad. Sci. USA, *96:* 13130–13135, 1999.
23. Bajou, K., Noel, A., Gerard, R. D., Masson, V., Brunner, N., Holst-Hansen, C., Skobe, M., Fusenig, N. E., Carmeliet, P., Collen, D., and Foidart, J. M. Absence of host plasminogen activator inhibitor 1 prevents cancer invasion and vascularization. Nat. Med., *4:* 923–928.
24. Ryu, S-W., Chae, S-K., Lee, K-J., and Kim, E. Identification and characterization of human Fas associated factor 1, hFAF1. Biochem. Biophys. Res. Commun., *262:* 388–394, 1999.
25. Yamabe, Y., Ichikawa, K., Sugawara, K., Imamura, O., Shimamoto, A., Suzuki, N., Tokutake, Y., Goto, M., Sugawara, M., and Furuichi, Y. Cloning and characterization of Rep8 (D8S2298E) in the human chromosome 8p11.2-p12 Genomics, *39:* 198–204 1997.
26. Hofmann, K., and Bucher, P. The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. Trends Biochem. Sci., *21:* 172–173, 1996.
27. Weterman, M., van Groningen, J. J. M., Jansen, A., and Geurts van Kessel, A. Nuclear localization and transactivating capacities of the papillary renal cell carcinoma-associated TFE3 and PRCC (fusion) proteins Oncogene, *19*: 69–74, 2000.
28. Kawai, A., Woodruff, J., Healey, J. H., Brennan, M. F., Antonescu, C. R., and Ladanyi, M. *SYT-SSX* gene fusion as a determinant of morphology and prognosis in synovial sarcoma. N. Engl. J. Med., *338:* 153–160, 1998.
29. Barr, F. G. The role of chimeric paired box transcription factors in the pathogenesis of pediatric rhabdomyosarcoma. Cancer Res., *59* (Suppl.)*:* 1711s–1715s, 1999.
30. Lin, P. P., Brody, R. I., Hamelin, A. C., Bradner, J. E., Healey, J. H., and Ladanyi, M. Differential transactivation by alternative EWS-FLI1 fusion proteins correlates with clinical heterogeneity in Ewing's sarcoma. Cancer Res., *59:* 1428–1432, 1999.
31. Ladanyi, M., Lui, M. Y., Antonescu, C. R., Krause-Boehm, A., Meindl, A., Argani, P., Healey, J. H., Ueda, T., Yoshikawa, H., Meloni-Ehrig, A., Sorensen, P. H. B., Mertens, F., Mandahl, N., van den Berghe, H., Sciot, R., Dal Cin, P., and Bridge, J. The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the *TFE3* transcription factor gene to *ASPL*, a novel gene at 17q25. Oncogene, *20:* 48–57, 2001.
32. Rubin, B. P., Chen, C-J., Morgan, T. W., Xiao, S., Grier, H. E., Kozakewich, H. P., Perez-Atayde, A. R., and Fletcher, J. A. Congenital mesoblastic nephroma t(12;15) is associated with *ETV6-NTRK3* gene fusion: cytogenetic and molecular relationship to congenital (infantile) fibrosarcoma. Am. J. Pathol., *153:* 1451–1458, 1998.
33. Kiuru-Kuhlefelt, S., El-Rifai, W., Sarlomo-Rikala, M., Knuutila, S., and Miettinen, M. DNA copy number changes in alveolar soft part sarcoma: a comparative genomic hybridization study. Mod. Pathol., *11:* 227–231, 1998.