

The Genomic Organization and Expression of the WT1 Gene

M. GESSLER,* A. KÖNIG,* AND G. A. P. BRUNST†

*Institut für Humangenetik der Philipps Universität, Marburg, Germany; and †Genetics Division, The Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts

Received October 14, 1991

The Wilms tumor gene WT1, a proposed tumor suppressor gene, has been identified based on its location within a homozygous deletion found in tumor tissue. The gene encodes a putative transcription factor containing a Cys/His zinc finger domain. The critical homozygous deletions, however, are rarely seen, suggesting that in many cases the gene may be inactivated by more subtle alterations. To facilitate the search for smaller deletions and point mutations we have established the genomic organization of the WT1 gene and have determined the sequence of all 10 exons and flanking intron DNA. The pattern of alternative splicing in two regions has been characterized in detail. These results will form the basis for future studies of mutant alleles at this locus. © 1992 Academic Press, Inc.

INTRODUCTION

Wilms tumor is a childhood nephroblastoma that affects approximately 1 in 10,000 children and occurs in both sporadic and familial form (Matsunaga, 1981). Epidemiological data suggested a two-hit model for Wilms tumor formation (Knudson and Strong, 1972), interpreted as the inactivation of both alleles of a tumor suppressor gene. The association of Wilms tumor with cytogenetic deletions and developmental malformations in the WAGR syndrome (Wilms tumor, aniridia, genitourinary abnormalities, and mental retardation) (Francke *et al.*, 1979) formed the basis for the later identification of the Wilms tumor gene (WT1) on chromosome 11p13. Studies on loss of heterozygosity in sporadic tumors and linkage analysis in familial cases, however, provided evidence for two additional loci (WT2, WT3) that may be involved in this tumor type, either as tumor suppressor genes or alleles that predispose to Wilms tumor (Grundy *et al.*, 1988; Henry *et al.*, 1989; Huff *et al.*, 1988; Koufos *et al.*, 1989; Reeve *et al.*, 1989). The significance of each individual locus for tumor formation and the possible interactions between them has not yet been clarified.

Recently, we and others isolated a WT1 candidate gene (LK15, WT33) that is expressed in fetal kidney and encodes a zinc finger protein of the Krüppel type, suggesting a role as a transcriptional regulator (Call *et al.*, 1990; Gessler *et al.*, 1990). The gene is expressed mainly

during development of the kidney and genitourinary structures (Pritchard-Jones *et al.*, 1990). There are seven independent cases of homozygous deletions of this gene described to date (Call *et al.*, 1990; Cowell *et al.*, 1991; Davis *et al.*, 1991; Gessler *et al.*, 1990; Huang *et al.*, 1990; Huff *et al.*, 1991; Ton *et al.*, 1991). This clearly establishes that loss of WT1 function must be an important factor in the development of these tumors. The search for homozygous deletions, however, will not identify all cases in which WT1 alterations occur. To detect more subtle alterations in Wilms tumors, knowledge of the genomic organization of this gene is required to permit scanning of individual exons for small deletions or point mutations.

In this report we present the genomic structure of the WT1 gene with exon/intron boundaries established by restriction mapping and sequence analysis of genomic DNA. The pattern of alternative splicing at two exons is characterized by sequencing of independent cDNA clones and direct analysis of RNA transcripts.

MATERIALS AND METHODS

Isolation and mapping of genomic clones. Isolation of cosmids cLFT3-3 and cLFT3-6 covering the proximal part of the WT1 locus has been described (Gessler *et al.*, 1990). Additional clones encompassing the 3' end of the LK15 cDNA clone were identified by screening of a pWE15 cosmid bank from human placental DNA (Stratagene, La Jolla, CA). Clones were characterized by restriction mapping and hybridization with cDNA probes. All *EcoRI* fragments identified by cDNA probes were subcloned into pBLUESCRIPT vectors (Stratagene) and smaller exon-containing subfragments were recloned for sequence determination.

Analysis of cDNA clones. During the isolation of the LK15 cDNA clone a total of 12 cDNA clones with insert sizes >2 kb were found (Gessler *et al.*, 1990). Nine of these clones were subcloned into pBLUESCRIPT and characterized by restriction mapping. The presence of exon 5 and the occurrence of the longer version of exon 9 was tested by direct sequence analysis using cDNA primers. The same information could also be obtained using digestion with diagnostic restriction enzymes.

DNA sequencing. Plasmid DNA was prepared from 1.5-ml overnight cultures by the alkaline lysis procedure and sequenced using the Sequenase kit (United States Biochemical Corp., Bad Homburg, Germany). G/C rich regions in the 5' part of the gene were resolved by substituting 7-deaza-dGTP in all reactions. Either internal primers designed from the cDNA sequence or external universal primers for pBLUESCRIPT were employed. Sequences were determined com-

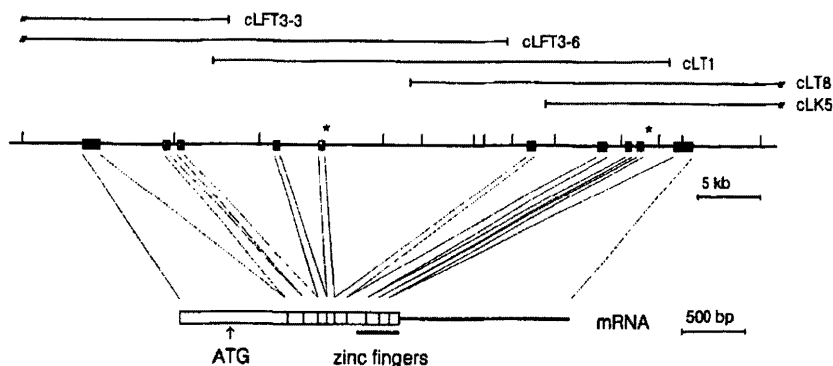


FIG. 1. Genomic organization of the WT1 locus. A part of the cosmid contig spanning the WT1 locus is shown on top with vertical lines indicating *EcoRI* cutting sites. Shaded boxes represent WT1 exons (not drawn to scale). Exons affected by alternative splicing are marked by asterisks. The assembly of exons into the 3-kb mRNA is shown in the bottom part of the figure.

pletely on both strands in most cases. Within exons 1 and 10 and downstream of exon 10, sequences were determined on one strand only. The HUSAR/UWGCG computer program package (DKFZ, Heidelberg, Germany) was used for sequence assembly and comparison.

Polymerase chain reaction. Amplification of DNA sequences was achieved by polymerase chain reaction (PCR) using 100 ng DNA template and 100 ng of each oligonucleotide primer (nucleotides 2714–2733 and 2821–2840 of the LK15 cDNA sequence for the amplification of the $(GT)_n$ repeat sequence). DNA was denatured for 30 s at 94°C, annealed with PCR primers at 55°C for 30 s, and extended with 1 unit *Taq* DNA polymerase (Amersham) for 60 s for a total of 25 cycles with the first and last step prolonged to 5 min. Polymerase buffer supplied by the vendor was used and supplemented with 100 μ g/ml bovine serum albumin and 200 μ M (dGTP, dATP, dTTP). Radioactive [α -³²P]dCTP (1 μ Ci) was used together with unlabeled dCTP at a final concentration of 2.5 μ M (adapted from Weber and May, 1989). Reaction products were resolved on a 6% polyacrylamide gel, followed by methanol/acetic acid fixation, drying, and autoradiography.

Analysis of RNA transcripts (RT-PCR). RNA was isolated from Wilms tumor tissue by the guanidinium–isothiocyanate/cesium chloride method (MacDonald *et al.*, 1987). Single-stranded cDNA was prepared from 5 μ g of total RNA (First strand kit, Stratagene) and amplification reactions were set up with 1/50 of the cDNA using primers corresponding to nucleotides 779–803 and 1581–1605 of the published LK15 cDNA sequence (Gessler *et al.*, 1990; Genbank X51630). The annealing temperature was set at 61°C with 40 cycles of amplification under standard conditions. One-half of the reaction mix was digested with *SacI* and products were visualized by ethidium bromide staining after agarose gel electrophoresis.

RESULTS

Two overlapping cosmid clones from the WT1 locus that encompass the 5' CpG island have been described previously (Gessler *et al.*, 1990). Expansion of this locus using cosmid end fragments and 3' cDNA probes resulted in a 80-kb cosmid contig that spans the entire genomic region for the WT1 gene (Fig. 1). Exon-containing fragments of individual cosmids were identified by hybridization of restriction digests with parts of the LK cDNA clones. On Southern blots of *EcoRI*-digested genomic DNA the longest cDNA clone, LK15, hybridizes to eight distinct bands. All of these fragments were found in different cosmid clones as predicted from the restriction map. Fine mapping with additional enzymes

revealed the presence of 10 different genomic restriction fragments detected by cDNA probes, suggesting the presence of 10 exons in the WT1 gene (Fig. 1).

The sequence of all exons and flanking intronic sequences were determined on appropriate genomic subclones or through the use of exon-specific oligonucleotide primers (Fig. 2). No differences were found between the cDNA sequence and genomic exon sequences. The position of mRNA splice sites was identified by comparison of sequences from genomic clones and corresponding cDNA clones. The sequences of splice donor and acceptor sites and the size of the resulting exons are listed in Table 1. Each of the four zinc fingers is encoded by a separate exon as has been noted by Haber *et al.* (1990). A stretch of 409 nucleotides upstream of the start of the LK15 cDNA sequence has been determined. As the LK15 insert size closely matches the apparent mRNA length, the sequenced upstream region likely contains part of the promoter region and transcriptional start site. There is a potential binding site for the transcription factor Sp1 (nts 232–237) and a sequence (nts 249–257) resembling the consensus binding site for the Egr-1, Krox-20, and WT1 proteins (Rauscher *et al.*, 1990). The significance of these sequences is presently not known.

The analysis of different cDNA isolates has suggested the presence of alternatively spliced transcripts for the WT1 locus. Exon 5 is absent from $\frac{1}{3}$ of the fetal kidney cDNA clones analyzed. This ratio of <50% skipping of exon 5 can also be demonstrated directly by PCR amplification of reverse transcribed RNA (RT-PCR) from fetal kidney and Wilms tumors (Fig. 3a). Using primers located in exons 1 and 9 two amplification products that differ by about 50 bp, the size of exon 5, are observed. The larger fragment containing exon 5 can be selectively cut by digestion with *SacI*, an enzyme that cuts only in exon 5 of the cDNA (Fig. 3b).

A second site of alternative splicing is located at the splice donor site of exon 9. Two different splice donor sequences can be used, resulting in the presence or absence of a nine-nucleotide sequence in the H/C link region between zinc fingers 3 and 4. Again, the presence of

```

1 gcccctccctccaaaccactcttttagattaacaacccccatctctaotcccaccgcattcgacctgcccggactcactgcttacctgaacgactctcca
101 gtgagacgaggtcccccactggcgaaggccaagaagggagggtggggggagggtgtgccacacggccagctgagagcgctgtgggttgaagagga
201 ggggtgtctcogagaggaagctocctoggaccgocctcaccacagctgcgaggggogcccccaaggagcagcgcgctgctggcgggctgggtctgc
301 tgagtgaatggagcggcagcctcctggctcctcctcttccccgcgcccggccccctctatttagctttgggaagctgagggcagccaggcagctg
401 gggtaaggagTTCAAAGGCAGCGCCACACCGGGGGCTCTCGCAACCGACCGCTGTCCGCTCCCCACTTCCCGCCTCCCTCCCACTACTCAATTC
   exon 1 SerArgGlnArgProHisProGlyAlaLeuArgAsnProThrAlaCysProLeuProHisPheProProSerLeuProProThrHisSe
501 ACCCAACCCACCCAGAGCCGGGACGGCAGCCAGGCGCCGGGCGCCCGCTCTCTCGCGCGATCTCGACTCTCTTGTCTGCAGGACCCGGCT
rProThrHisProProArgAlaGlyThrAlaAlaGlnAlaProGlyProArgArgLeuLeuAlaAlaIleLeuAspPheLeuLeuLeuGlnAspProAla
601 TCCACGTGTGTCGGGAGCCGGCTCTCAGCACAGCTCCGCTCCGGGCGTGGTGCCTACAGCAGCCAGCAGCAGGAGTCCGGGACCCGGGCGCA
SerThrCysValProGluProAlaSerGlnHisThrLeuArgSerGlyProGlyCysLeuGlnGlnProGluGlnGlnGlyValArgAspProGlyGlyI
701 TCTGGGCCAAGTTAGCGCCGAGGCCAGCGCTGAACGTCTCCAGGGCCGAGGAGCCGGGGCGTCCGGCTCAGCCTCAGCAATGGGCTCCGA
leTrpAlaLysLeuGlyAlaAlaGluAlaSerAlaGluArgLeuGlnGlyArgArgSerArgGlyAlaSerGlySerGluProGlnGlnMetGlySerAs
801 CGTGGGGACCTGAACCGCTGTGCCCCGCTCCCTCCCTGGTGGCGGGCGGGCTGTGCCCTGCCTGTGAGCGGGCGGGCAGTGGCGCGGGTG
pValArgAspLeuAsnAlaLeuLeuProAlaValProSerLeuGlyGlyGlyGlyCysAlaLeuProValSerGlyAlaAlaGlnTrpAlaProVal
901 CTGGACTTTCGCCCCCGGGCGCTTCGGCTTACGGTCTGGGGCGGCCCGCGCCACCGGCTCCGCGCCACCCCGCGCGCGGCTCACTCT
LeuAspPheAlaProProGlyAlaSerAlaTyrGlySerLeuGlyGlyProAlaProProProAlaProProProProProProHisSerP
1001 TCATCAACAGGAGCCGAGCTGGGGCGGGGAGCCGACGAGGAGCAGTGCCTGAGCGCTTCACTGTCCACTTTCCGGCCAGTTCACCTGGCACAGC
heIleLysGlnGluProSerTrpGlyAlaGluProHisGluGluGlnCysLeuSerAlaPheThrValHisPheSerGlyGlnPheThrGlyThrAl
1101 CGGAGCTGTGCTACGGCCCTTCGGTCTCTCCGCCAGCCAGGCTCATCCGGCCAGGCCAGGATGTTTCTAACGGCCCTACCTGCCAGCTGC
aGlyAlaCysArgTyrGlyProPheGlyProProProProSerGlnAlaSerSerGlyGlnAlaArgMetPheProAsnAlaProTyrLeuProSerCys
1201 CTCGAGAGCCAGCCCGCTATTCGCAATCAGGgtaagtagggcgggagcggccctacggcggggagtgccagggactctccgctctaggacacc
LeuGluSerGlnProAlaIleArgAsnGlnG
1301 cccctctctacccttttgaccgcagctcttaccagctgcttcccaagggcggtaggagtagcggagcggcggtggggaggaggccggagagtggg
1401 agtgacgcagcagcactggcccccgacatctccaaagccaggcagagctaggagcctgactgttcgcaagagcggggggcgcttggggccc.....
.....
1 agagaccgtagaagactcgggcaataataaattgggaaaaagcaggtgggggttttagtcagaaaaagtaggagaacgctgtcccaaggtcacatccagg
101 gacttgggcccgtctctctgcogaagtcctggaggcttgggggctggattctcctccgctgggggtccgaaatgcgggggttcagcgcacacccccccg
201 ccgtggctggttcagaccactgcccgtcttgcgagagcaccogctgacactgtgctctctccgocgctcccgagTTCACAGCAGCTCACCTTCGCAC
   exon 2 lyTyrSerThrValThrPheAsp
301 GGGACCCCGACTACGGTTCACCGCCCTCGCACCATGGCGGCAGTTCGCCAACCACTCATCAAGCATGAGGATCCATGGGCCAGCAGGGCTCGCTGG
GlyThrProSerTyrGlyHisThrProSerHisHisAlaAlaGlnPheProAsnHisSerPheLysHisGluAspProMetGlyGlnGlnGlySerLeuG
401 gtaagcagcgggaaccaagtgaggtcctctccctctctcctgctctctcctcctcctcccccaggaattcotaaccacagcaaataggcaaggagga
501 catctaaaaggaccttagaatacattacccaatggcactgtatttaaaggacaagactgggtctccctccatcccaagtagcttagggacatcgccatc
601 caactaagctctatgaatgctccggagggccttttggtgcaagcgcagcagctagctaggggaaatgcaaaagtgagggcttagcttctgcatcttgacg
701 gcctcgtctctgcaagcggctcctccagggcccagtgcccaggctcaggatctcgtgctcccccaaccctgcccgcagcctcgcccccgagcccttg
801 gggcgcaactcgctcagctgtctcctggtctctctcagctGTAGCAGCAGTACTCGGTCCGCCGCCCGGCTCTATGGTGGCCACCCCCACCGACAGCTGC
   exon 3 lyGluGlnGlnTyrSerValProProProValTyrGlyCysHisThrProThrAspSerCys
901 ACCGGCAGCCAGGCTTTGGTGTGAGGACGCCCTACAGCAGtaggaaagcgctgggctctgctctgggtccttgggaccccgagggggacgcatgagc
ThrGlySerGlnAlaLeuLeuLeuArgThrProTyrSerSe
1001 cggcggacttgggaacgccccagccggcagcctgactgogggcagcaggtcttggaggcagagacgctcogactccactctactcttagaggaa
1101 attaaagctgactcctcctggtctctgagttgttcttattgtgtataaaaataggacaattactcttccgagctgaagaaggagcggcgctcgag....
.....
1 tttccctcaaaggttaacttctcttagaaactcaatagtgctatacaaggttttaaaaaatgcttctgagttactctcatccctctctgctgt
101 gtagaagtgataactgtgcagagatcagtgggatgaaatgaaaataattcagatttgcatattctgtcattctggaaaatgtggaggttgcactttc
201 aagaaaagttaaaactccattgcttttgaaagaacagttgtgtatttttgggttattgttttctaactctagatgtttttcaacagTGACAAATTA
   exon 4 rAspAsnLeu
301 TACCAATGACATCCCGCTTGAATGCATGACCTGGAAATCAGATGAACTTAGGAGCCACCTTAAAGGGgtaaggtttctctattccattgcctttccac
TyrGlnMetThrSerGlnLeuGluCysMetThrTrpAsnGlnMetAsnLeuGlyAlaThrLeuLysGln
401 agtaactatacctgtttgaaccattcaaggtgottagaacttagaagaagtagcagtttagaagaaggggcaagctctctgtgacattagaacg
501 ctttctcctccccctagttggtgagttgttctcagattaaatttcaaaaacataaagtgggtgtttctctatctggacaaggaatattgaggacac
601 ttattatggtcttggtaagttttataatcttcaaggtattcaacctatttaataattgtgttttactatgatcacagctgtt.....
.....
1 agatctgatccaaagcccaagctcttgcagcctccggtggtgccttccaccccaccattggcccagtgaggactggggacttagttcagcactcttga
101 tagctagctgtatgtctctgtgtattgagggccttttcaactgattctggggtctgggggcttgccagatccatgcatgtccattccccactccccac
201 ctcttctctttcttagAGTTGGCTGGGAGCTCCAGCTCAGTGAATGGCAGAAAGGGCAGAGCAAgtagggacaatgcccagagatggcaagcaaa
   exon 5 yValAlaAlaGlySerSerSerSerValLysTrpThrGluGlyGlnSerAs
301 tggcgtaggcctctgtgactggccactgggcaatgagaggttaggactgagctgggtaaatcagggtagcatttgggtggatgggaagaatctctctccc
401 cggccccccctctcttattcatcacaacatgctggggcaaaaatggaggagtagcatttctcatcaacatgcccctggaaaacaatgtaaaaaatt

```

FIG. 2. Sequence of the WT1 gene. The nucleotide sequence of the 10 exons and flanking intronic sequences are shown. Exon sequences present in cDNA clones are shown in uppercase letters with the corresponding amino acid translation shown below. The putative translation initiation site in exon 1 (Buckler *et al.*, 1991) is marked in boldface. All intron sequences and the putative promoter region are printed in lowercase letters. The nine nucleotides that are different between the alternatively spliced forms of exon 9 are underlined. The polyadenylation signal is marked by double underline. All numbering refers to individual exons with surrounding sequences.

501 atgcagcatcctgttcatttcttctgtgaatacttcacaaaactaataactaaatttggcattacaaccataatgccacatgttccgttagatcag
 601 tcaaagctt.....
 1 ggatccaggggctcgcctcccgagaacctgcatctaaagtggcccatgaggaggcagaaaccactattttgttctgtgagctttaaacc
 101 tcattcctcctgattgacagataagcatttccaaatggcgaotgtgagccacactgagccttttccctcttctgttggcagCCACAGCACAGGGTAC
 exon 6 nHisSerThrGlyTyr
 201 GAGAGCGATAACACACAAAGCCATCCTCTGCGGAGCCCAATACAGAAATACACACGCGGTGTCTTACAGAGCATTACAGgtgacacacagacagggc
 GluSerAspAsnHisThrThrProIleLeuCysGlyAlaGlnTyrArgIleHisThrHisGlyValPheArgGlyIleGln
 301 ccagcgggaagggggccgactgctcttctacttacogcccttagtctctactgtaggactcttggccctggcagcctcttctcactgag..

 1 attgtagaatgtcctgactatcattctctgctgctgcatggggatctggagtgatggggagtggttttaggtctgacccctgccccctct
 101 ttggataactccagtgctcactctccctcaagacctacgtgaatgttccacatgtggcttaagcctccctctctactctctgctgagGATGTGC
 exon 7 AspValA
 201 GACGTGTGCTGGAGTAGCCCGACTCTTGTACGGTCGGCATCTGAGACCAGTGAGAAACGCCCTTCATGTGTGCTTACCAGGCTGCAATAAGAGATA
 rgArgValProGlyValAlaProThrLeuValArgSerAlaSerGluThrSerGluLysArgProPheMetCysAlaTyrProGlyCysAsnLysArgTy
 301 TTTAAAGCTGTCCACTTACAGATGCACAGCAGAAAGCACACTGTgaagtgtgcccgtctccagctcttgggcaaacatggttcaagagctctcttcca
 rPheLysLeuSerHisLeuGlnMetHisSerArgLysHisThrG
 401 ggctctcacactgctaaggaccagggtatttccaggatggaaagtaagcactgcttgaaggctgataccaggtgtttaaaggctgtggggccat
 501 ttgcatctgattatctatcctttattacagaggtcagtaactgtctccagaaaggcagagca.....

 1 ttttttttttaagggcagagctagatgtgagccttggggcagaatcctctggccctgagctccattcatttgaacttgaagtggggatcatcc
 101 taccctaaacagctccagcgaagtgccttaggcatttgggatctgttcagctcctagtaggagaggttgcccttaagagatccctttccagatcat
 201 tttcaaatagaatgtgtcttcccaagGTGAGAAACCATACCAGTGTGACTTCAAGGACTGTGAACGAAGGTTTCTCGTTACAGACCAGCTCAAAGA
 exon 8 lyGluLysProTyrGlnCysAspPheLysAspCysGluArgArgPheSerArgSerAspGlnLeuLysArg
 301 CACCAAAGGAGACATACAGgtttgtaggttcactctcattgtctggcagctgtgtcccttgggctagggtgatttcatgattctcttgttggat
 HisGlnArgArgHisThrG
 401 atgaatgagagagtcagccatgtgttttagttttccagttttccccccataatctctggagatgttccctctttaaataaagaaaagaaagaaag
 501 gaaacacccacatttccctgcaaacccagaaacagtagatcccactagcagtggtgatgatacactgagcatcagcaagaatgaggtgatgga
 601 gagcggaaatccaaatgtgagcgtgattcaactttctgtctctgacacagctgagatctccagctggtgagaagctggaacctctgtgtaaggaa
 701 ggtgttcccaagacagaagcaggtccttgaagattactgtgcagaatcagtgattcttgttggcctggggactgggaaatctaagggtgagggcaga
 801 tgcagacattgcaggcatggcaggaaatgctgggctcctcaagctgocggaagtcagccttgtggcctcactgtgcccacattgttagggccagggcta
 901 gaccttctctgctccatttagGTGAAACCATTCCAGTGTAAAGCTTGTGAGCGAAAGTTCTCCCGTCCGACCCTGAAGACCCACACAGGACTCAT
 exon 9 lyValLysProPheGlnCysLysAlaCysGlnArgLysPheSerArgSerAspHisLeuLysThrHisThrArgThrHis
 1001 ACAGGTAACCAAGtgcgtaaacctttctcactttatttttcaatttttttaactattgtggaatgaaattgtgatgagagggattggaagagc
 ThrGlyLysThrS
 1101 tagtgtaaaagttctaaggttagactcagttgagagaaggaatagtgctggtctatctttaaataataataaacattttta.....

 1 tttctgacctagaataataacttttacaatattcogactcctaaagcccaagaggtgctgctgttcaaggctgctgggaaatgagctcagttggtt
 101 agctcagggcagaaatgatgggaaatcagagtggtgctctgtgatcactcactcggccttgatagtgaaactgtgctctctcttctgtgcaagt
 201 gtctctgactggcaattgtgtcaacagGTGAAAGCCCTTACAGTGTGCGGTGCCAAGTTGTCAGAAAAAGTTGCCCCGGTCCAGTGAATTAGTCCGCCA
 exon 10 erGluLysProPheSerCysArgTrpProSerCysGlnLysLysPheAlaArgSerAspGluLeuValArgHi
 301 TCACAACATGCATCAGAGAAACATGACAAACTCCAGTGGCGCTTTGAGGGGTCTCCCTCGGGACCGTTCAGTGTCCAGGCAGCAGTGTGTGAAC
 sHisAsnMetHisGlnArgAsnMetThrLysLeuGlnLeuAlaLeu***
 401 TGCITTCAGTCTGACTCTCCACTCCTCCTCACTAAAAAGGAACTTCAGTTGATCTCTTCATCCAACTTCCAAGACAAGATACCGGTGCTCTCGAAA
 501 CTACCAGGTGTGCTGGAGAGTGGTCTCTGCCCCTGCTACTTTTAGTTGACTCACAGGCCCTGGAGAGCAGCTAACATGTCTGGTTAGTTAAAGC
 601 CCATTGCCATTTGCTCGGATTTTCTACTGTAAGAAGAGCCATAGCTGATCATGTCCCCCTGACCCCTTCCCTTCTTTTTTATGCTCGTTTTCTGGTGGG
 701 ATGGAATTTATGTAACATTTTCTATCATGGAATATTTATAGGCCAGGGCATGTGTATGTGTCTGTAATGTAACTTTGTCATGGTTTCCATTTACTAAC
 801 AGCAACAGCAAGAAATAAATCAGAGAGCAAGGCATCGGGGTGAATCTTGTCTAACATCCCGAGTCCAGCCAGGCTGATACCTGAAAGCAGGATGTA
 901 GTTCTGCCAGGCAACTTTTAAAGCTCATGCAATTTCAAGCAGCTGAAGAAAGAAATCAGAACTAACCCAGTACCCTGTATAGAAATCTAAAGAAATTTACC
 1001 ATTCACTTAATTCAAATGTAACACTGGCACACTGCTCTTAAAGAACTATGAAGATCTGAGATTTTTTGTGTATGTTTTGACTTTTTGAGTGTAAATC
 1101 ATATGTGTCTTTATAGATGTACATACCTCCTTGCACAAATGGAGGGAAATTCATTTTCATCACTGGGACTGTCTTGTATATAAAAACATGCTGGTAT
 1201 ATGGCTCAAGTGTAAAAATGAAAGTGAATTTAAAAAGAAATAGGGGATGGTCCAGGATCTCCACTGATAAGACTGTTTTAAAGTAACTTAAGGACCTT
 1301 TGGGCTACAAGTATATGTGAAAAAATGAGACTTACTGGGTGAGGAAATCCATGTTTAAAGATGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGT
 1401 GTTGTGTTTTTTTTAAAGGAGGGAATTTATTTACCGTTGCTTGAANTACTGTGTAATATATGTCTGATAATGATTTGCTCTTTGACAACTAA
 1501 AATTAGGACTGTATAAGTACTAGATGATCACTGGGTGTGATCTTACAAGATATTGATGATAACACTTAAATGTAACTTCACTTTTCACTTTGCTC
 1601 TCAATTAAGTCTATTCAAAGGaaagtgcaagattccatcagctgtatgtgttcaaatctcaaggccactgagcctctttaaacaggtgaagagaag
 1701 acgccaacgacatcagcggctactagtgagcctatccaggtcctgaccacaatttattctctctgtcaacttcaggagggacttatatgtcctagaag
 1801 caacagagaatattctcacttctgttctacagagtgtatgtctcatgctgtgggtgaatgggg 1863

FIG. 2—Continued

TABLE 1
Sizes of Exons and Splice Junction Sequences of the WT1 Gene

Exon	Nucleotides of cDNA	3' splice (t/c) _n /ag/GTN	Size (bp)	5' splice NAG/gtaagta
1	1-822		>822	AGG/gtaagta
2	823-945	ctctccgccgctcccgcag/GTT	123	TGG/gtaagcg
3	946-1048	tcttcggttctctctgcag/GTG	103	CAG/gtaggaa
4	1049-1126	ctagatgttttctaacag/TGA	78	GGG/gtaaggt
5	1127-1177	acctcttctctttctag/AGT	51	CAA/gtgagtg
6	1178-1274	cccttcttggtttgcag/CCA	97	CAG/gtgagca
7	1275-1425	tcttactctctgcctgcag/GAT	151	CTG/gtaagtg
8	1426-1515	aatatgtgtcttcccacag/GTG	90	CAG/gtttcta
9	1516-1599	accttctctgtccatttag/GTG	84	CAG/gtaaac
9 ^a	1516-1608		93	CAA/gtgcgta
10	1609-3007	ctggcaattgtgtcaacag/GTG	1399	

Note. Numbering of the corresponding cDNA sequence refers to the published sequence (Gessler *et al.*, 1990).
^a Alternatively used splice donor site.

the 9-bp insert can be identified in any RNA sample by RT-PCR and testing of the product for the *HphI* restriction site that is created by the sequence insert. Of the nine cDNA clones analyzed, each contained either exon 5 or the 9-bp insert at exon 9, with two clones containing both.

At the 3' end of the cDNA sequence a (GT)_n repeat that shows a different number of repeat units in other cloned DNA samples has been observed. Analysis of repeat length in 27 unrelated individuals revealed 6 different alleles with 65% heterozygosity, resulting in a calculated PIC value of 0.61 (Fig. 4, Table 2). The presence of this repeat in transcribed RNA makes it possible to identify parental origin of the allele being expressed. This can be used to assess possible parental imprinting at this locus.

DISCUSSION

The genomic sequence of the WT1 gene is an important tool for the analysis of mutations in Wilms tumor

material. In cases where no expression or only aberrantly sized transcripts can be detected, the analysis of genomic exons can reveal the nature of the underlying mutations. There are also numerous instances where only DNA is still available on tumors or tumor necrosis after chemotherapy has made extraction of intact RNA impossible.

The WT1 gene has been postulated to play a role in the genitourinary malformations seen in WAGR patients (van Heyningen *et al.*, 1990). A certain percentage of individuals with abnormalities of the genitourinary system, especially those with a family history of aniridia or Wilms tumor may carry mutations of the WT1 gene. Given the restricted expression pattern of this gene, it will be difficult to assess its regular expression in these cases. The possibility of amplifying individual exons by PCR will permit the scanning of the WT1 locus for subtle mutations that cannot be detected by Southern blot analysis. The intron sequences presented exhibit no significant homology to known human repeat sequences, so all sequences can be considered to design primers for PCR.

The presence of a highly informative intragenic (GT) repeat polymorphism will facilitate the screening for loss of heterozygosity in Wilms tumors and other tumor types for which evidence of chromosome 11 alterations exist (Atkin and Baker, 1985; Sait *et al.*, 1988). The oc-

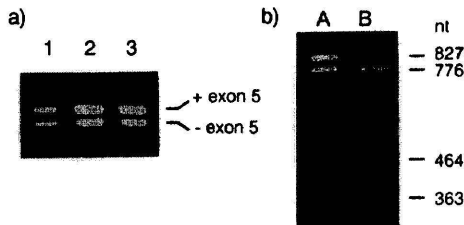


FIG. 3. Detection of alternative splicing of exon 5. (a) Two different transcripts for WT1 can be detected in unrelated Wilms tumor RNA samples (1-3) by RT-PCR using primers located in exons 1 and 9. The difference in length (827 vs 776 bp) seen upon agarose gel electrophoresis results from the presence or absence of exon 5. (b) The longer transcript detected by RT-PCR contains exon 5. There is only a unique site for the restriction enzyme *SacI* within the WT1 transcribed sequences, located in exon 5. Comparison of RT-PCR products from Wilms tumor RNA either undigested (lane A) or digested with *SacI* after amplification (lane B) shows that only the upper band containing exon 5 can be cut.

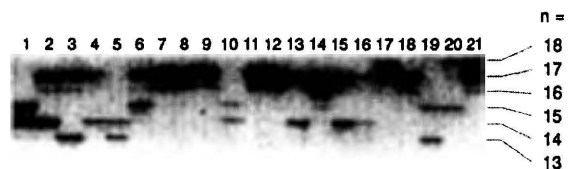


FIG. 4. A polymorphic (GT) repeat in the 3'-untranslated region of the WT1 cDNA. PCR amplification of the GT repeat sequence was carried out on DNA from 21 unrelated individuals. After gel electrophoresis and drying of the gel, six different alleles can be identified on the autoradiogram.

TABLE 2
Frequency of (GT)_n Repeat Alleles

Number of (GT) repeats	Frequency ^a (%)
13	7.4
14	53.7
15	1.8
16	11.1
17	20.4
18	5.6

^a Numbers are calculated from 54 alleles tested. Heterozygosity, 0.65; PIC = 0.61.

currence of this repeat sequence in transcribed DNA will, in addition, permit the identification of the parental origin of the transcribed allele. The preferential loss of maternal alleles in Wilms tumors has suggested a role for parental imprinting during tumor development (Wilkins, 1988). The comparison of the (GT) repeat alleles in genomic DNA and reverse transcribed mRNA can show whether there is a molecular basis for an unequal functional potential of the two parental alleles of the WT1 gene.

The WT1 gene can be alternatively spliced, thereby leaving out exon 5 or altering the length of exon 9 by nine nucleotides. In each case the two different forms occur at a similar frequency, but all of our cDNA clones contained at least one of the longer splice variants. In the analysis of the murine WT1 homolog Buckler *et al.* (1991) observed similar types of alternative splicing, suggesting that it may be of functional importance. This is further strengthened by the results of Rauscher *et al.* (1990) who showed that the addition of the three amino acid residues between the third and fourth zinc finger may alter DNA binding capacity and perhaps specificity.

The role WT1 plays in Wilms tumor development and the fraction of tumors that have mutations at this locus is still unresolved. The present report provides tools for a thorough study of the involvement of WT1 in Wilms tumor formation and in the induction of genitourinary malformations. Identification of small deletions or point mutations in these cases may identify functionally important sites in the gene product besides the DNA binding domain.

ACKNOWLEDGMENTS

We thank Professor K. H. Grzeschik for support. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Ge539/2-2) and the NIH (HG00186).

REFERENCES

- Atkin, N., and Baker, M. C. (1985). Cytogenetic study of ten carcinomas of the bladder: involvement of chromosomes 1 and 11. *Cancer Genet. Cytogenet.* **15**: 253-268.
- Buckler, A. J., Pelletier, J., Haber, D. A., Glaser, T., and Housman, D. E. (1991). Isolation, characterization, and expression of the murine Wilms' tumor gene (WT1) during kidney development. *Mol. Cell. Biol.* **11**: 1707-1712.
- Call, K. M., Glaser, T., Ito, C. Y., Buckler, A. J., Pelletier, J., Haber, D. A., Rose, E. A., Kral, A., Yeager, H., Lewis, W. H., Jones, C., and Housman, D. E. (1990). Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**: 509-520.
- Cowell, J. K., Wade, R. B., Haber, D. A., Call, K. M., Housman, D. E., and Pritchard, J. (1991). Structural rearrangement of the WT1 gene in Wilms' tumour cells. *Oncogene* **6**: 595-599.
- Davis, L. M., Zabel, B., Senger, G., Lüdecke, H.-J., Metzroth, B., Call, K., Housman, D., Claussen, U., Horsthemke, B., and Shows, T. B. (1991). A tumor chromosome rearrangement further defines the 11p13 Wilms tumor locus. *Genomics* **10**: 588-592.
- Francke, U., Holmes, L. B., Atkins, L., and Riccardi, V. M. (1979). Aniridia-Wilms' tumor association: Evidence for a specific deletion of 11p13. *Cytogenet. Cell Genet.* **24**: 185-192.
- Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H., and Bruns, G. A. P. (1990). Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343**: 774-778.
- Grundy, P., Koufos, A., Morgan, K., Li, F. P., Meadows, A. T., and Cavenee, W. K. (1988). Familial predisposition to Wilms tumour does not map to the short arm of chromosome 11. *Nature* **336**: 374-376.
- Haber, D. A., Buckler, A. J., Glaser, T., Call, K. M., Pelletier, J., Sohn, R. L., Douglass, E. C., and Housman, D. E. (1990). An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell* **61**: 1257-1269.
- Henry, I., Grandjouan, S., Couillin, P., Barichard, F., Huerre-Jeanpierre, C., Glaser, T., Philip, T., Lenoir, G., Chaussain, J. L., and Junien, C. (1989). Tumor-specific loss of 11p15.5 alleles in del11p13 Wilms tumor and in familial adrenocortical carcinoma. *Proc. Natl. Acad. Sci. USA* **86**: 3247-3251.
- Huang, A., Campbell, C. E., Bonetta, L., McAndrews-Hill, M. S., Chilton-MacNeill, S., Coppes, M. J., Law, D. J., Feinberg, A. P., Yeager, H., and Williams, B. R. G. (1990). Tissue, developmental, and tumor-specific expression of divergent transcripts in Wilms tumor. *Science* **250**: 991-994.
- Huff, V., Compton, D. A., Chao, L., Strong, L., Geiser, C., and Saunders, G. (1988). Lack of linkage of familial Wilms' tumour to chromosomal band 11p13. *Nature* **336**: 377-378.
- Huff, V., Miwa, H., Haber, D. A., Call, K. M., Housman, D., Strong, L. C., and Saunders, G. F. (1991). Evidence for WT1 as a Wilms tumor (WT) gene: Intragenic germinal deletion in bilateral WT. *Am. J. Hum. Genet.* **48**: 997-1003.
- Knudson, A. G., and Strong, L. C. (1972). Mutation and Cancer: A model for Wilms' tumor of the kidney. *J. Natl. Cancer Inst.* **48**: 313-324.
- Koufos, A., Grundy, P., Morgan, K., Aleck, K. A., Hadro, T., Lampkin, B. C., Kalbakji, A., and Cavenee, W. K. (1989). Familial Wiedemann-Beckwith syndrome and a second Wilms tumor locus both map to 11p15.5. *Am. J. Hum. Genet.* **44**: 711-719.
- MacDonald, R. J., Swift, G. H., Przypyla, A. E., and Chirgwin, J. M. (1987). Isolation of RNA using guanidinium salts. In "Methods in Enzymology" (S. L. Berger and A. R. Kimmel, Eds.), Vol. 152, pp. 219-231, Academic Press, San Diego.
- Matsunaga, E. (1981). Genetics of Wilms' tumor. *Hum. Genet.* **57**: 231-246.
- Pritchard-Jones, K., Fleming, S., Davidson, D., Bickmore, W., Porteous, D., Gosden, C., Bard, J., Buckler, A., Pelletier, J., Housman, D., van Heyningen, V., and Hastie, N. D. (1990). The candidate Wilms' tumor gene is involved in genitourinary development. *Nature* **346**: 194-197.
- Rauscher, F. J., Morris, J. F., Tournay, O. E., Cook, D. M., and Curran, T. (1990). Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science* **250**: 1259-1262.

- Reeve, A. E., Sih, S. A., Raizis, A. M., and Feinberg, A. P. (1989). Loss of allelic heterozygosity at a second locus on chromosome 11 in sporadic Wilms' tumor cells. *Mol. Cell. Biol.* **9**: 1799-1803.
- Sait, S. N. J., Dal Cin, P., and Sandberg, A. A. (1988). Consistent chromosome changes in Leiomyosarcomas. *Cancer Genet. Cytogenet.* **35**: 47-50.
- Ton, C. T., Huff, V., Call, K. M., Cohn, S., Strong, L. C., Housman, D. E., and Saunders, G. F. (1991). Smallest region of overlap in Wilms tumor deletions uniquely implicates an 11p13 zinc finger gene as the disease locus. *Genomics* **10**: 293-297.
- van Heyningen, V., Bickmore, W. A., Seawright, A., Fletcher, J. M., Maule, J., Fekete, G., Gessler, M., Bruns, G. A. P., Huerre-Jeanpierre, C., Williams, B. R. G., and Hastie, N. D. (1990). Role for the Wilms tumor gene in genital development? *Proc. Natl. Acad. Sci. USA* **87**: 5383-5386.
- Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388-396.
- Wilkins, R. J. (1988). Genomic imprinting and carcinogenesis. *Lancet* **1**: 329-331.