6. E. J. Beer, *Rec. Geol. Surv. India* **50**, 139 (1919).
7. B. Runnegar, unpublished observations.
8. T. M. Han, *U.S. Geol. Surv. Prof. Pap.* **769**, 76 (1975).
9. P. Cloud, S. M. Awramik, K. Morrison, D. G. Hadley, *Precamb. Res.* **10**, 73 (1979); F. Wang, *Nature* **294**, 74 (1981); X. Song, *Geol. Mag.* **121**, 179 (1984); R. Riding and L. Voronova, *ibid.*, p. 205; J. S. Peel, *Rapp. Grønlands Geol. Unders.* **137**, 5 (1988); A. H. Knoll and Y. Ohta, *Polar Res.* **6**, 59 (1988).
10. G. Vidal, *Lethaia* **22**, 375 (1989).
11. V. A. Gallardo, *Nature* **268**, 331 (1977); C. R. Smith, H. Kukert, R. A. Wheatcroft, P. A. Jumars, J. W. Deming, *ibid.* **341**, 27 (1989); H. W. Jannasch, D. C. Nelson, C. O. Wirsen, *ibid.* **342**, 834 (1989); M. P. Starr, H. Stolp, H. G. Trüper, A. Balows, H. G. Schlegel, *The Prokaryotes* (Springer-Verlag, Berlin, 1981).
12. N. H. Mendelson, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1740 (1976); _____ and J. J. Thwaites, *J. Bacteriol.* **171**, 1055 (1989).
13. B. Runnegar, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 97 (1991).
14. J. P. Gogarten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661 (1989); N. Iwabe, K. I. Kuma, M. Hasegawa, S. Osawa, T. Miyata, *ibid.*, p. 9355.
15. T. Cavalier-Smith, *Ann. N.Y. Acad. Sci.* **503**, 17 (1987).
16. M. L. Sogin, J. H. Gunderson, H. J. Elwood, R. A. Alonso, D. A. Peattie, *Science* **243**, 75 (1989); K. S. Kabnick and D. A. Peattie, *Am. Sci.* **79**, 34 (1991).
17. J. W. Schopf, in *Major Events in the History of Life*, J. W. Schopf, Ed. (Jones & Bartlett, Boston, MA, 1992), pp. 29–63.
18. H. J. Hofmann and J. Chen, *Can. J. Earth Sci.* **18**, 443 (1981).
19. R. E. Summons, T. G. Powell, C. J. Boreham, *Geochim. Cosmochim. Acta* **52**, 1747 (1988); R. E. Summons and M. R. Walter, *Am. J. Sci.* **290-A**, 212 (1990).
20. R. F. Doolittle, K. L. Anderson, D. F. Feng, in *The Hierarchy of Life*, B. Fernholm, K. Bremer, H. Jörnvall, Eds. (Elsevier, Amsterdam, 1989), pp. 73–143.
21. T. Cavalier-Smith, *Ann. N.Y. Acad. Sci.* **503**, 55 (1987).
22. C. R. Woese and G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977); C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987); _____, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
23. E. Henderson, M. Oakes, M. W. Clark, J. A. Lake, A. T. Matheson, W. Zillig, *Science* **225**, 510 (1984); J. A. Lake *Nature* **331**, 184 (1988); *Mol. Biol. Evol.* **8**, 378 (1991).
24. M. C. Rivera and J. A. Lake, *Science* **257**, 74 (1992).
25. N. Iwabe, K. I. Kuma, H. Kishino, M. Hasegawa, T. Miyata, *J. Mol. Evol.* **32**, 70 (1991).
26. M. Schidlowski, *Nature* **333**, 313 (1988).
27. J. M. Hayes, in *Earth's Earliest Biosphere. Its Origin and Evolution*, J. W. Schopf, Ed. (Princeton Univ. Press, Princeton, NJ, 1983), pp. 291–301.

28. We thank J. W. Villar and G. L. LaBerge for their enthusiastic support of this work and the Management of Cleveland-Cliffs, Inc. for permission to publish this paper.

31 January 1992; accepted 18 May 1992

Modulation of DNA Binding Specificity by Alternative Splicing of the Wilms Tumor *wt1* Gene Transcript

W. A. Bickmore, K. Oghene, M. H. Little, A. Seawright, V. van Heyningen, N. D. Hastie*

The technique of whole-genome polymerase chain reaction was used to study the DNA binding properties of the product of the *wt1* gene. The zinc finger region of this gene is alternatively spliced such that the major transcript encodes a protein with three extra amino acids between the third and fourth fingers. The minor form of the protein binds specifically to DNA. It is now shown that the major form of wt1 messenger RNA encodes a protein that binds to DNA with a specificity that differs from that of the minor form. Therefore, alternative splicing within the DNA binding domain of a transcription factor can generate proteins with distinct DNA binding specificities and probably different physiological targets.

The *wt1* gene was isolated from the region of human chromosome 11p13 implicated in predisposition to the development of Wilms tumor (*1*, *2*). The expression pattern of WT1, detection of intragenic deletions, and point mutations suggest that this gene functions in the regulation of kidney and gonadal development and in the genesis of Wilms tumor (*3*). The zinc fingers in the WT1 protein are related to those of the early growth response (EGR) family of proteins, and WT1 binds to a consensus DNA binding site for these proteins (*4*).

Studies of EGR1 and EGR2 (*5*, *6*) showed that each finger contacts a 3-bp subsite of DNA antiparallel to the guanine (G)-rich strand of the double helix, so that the most COOH-terminal finger contacts the 5′ end of the binding site. Amino acids in the NH_2-terminal portion of the α helix of each finger contact G residues in the DNA. The amino acid preceding the α helix of each finger (Fig. 1A, box I) contacts the third base of the subsite (− − G). In EGR1 finger 2, the third residue of the α helix (Fig. 1A, box II) contacts the second base (− G −), and in fingers 1 and 3 the sixth residue (Fig. 1A, box III) contacts the first base (G − −). It appears that the binding site for an EGR-type zinc finger protein can be deduced from this simple recognition code. Other nucleotide residues seem to be of little importance in determining binding specificity, although it seems unlikely that proteins with identical contact residues (for example, EGR1, EGR2, and EGR4) fulfill their respective cellular functions by binding identical DNA targets in vivo (*7*). Fingers 2 through 4 of WT1 bind to the EGR1 consensus site (*4*). However, WT1 contains an extra finger not found in other EGR-like proteins (Fig. 1A) whose function in determining binding specificity is not understood.

The *wt1* gene is atypical in another respect. In other EGR-like proteins, seven amino acids separate adjacent fingers, and these may be important in juxtaposing adjacent fingers correctly relative to the target (*6*). In WT1 use of an alternative 5′ splice junction introduces three extra amino acids (KTS) between fingers 3 and 4 (Fig. 1, A and B). This is in fact the predominant form of human and murine WT1 mRNA in all cells that express WT1 (*8*), and we refer to it here as the +KTS form of WT1. These extra amino acids may displace the fourth zinc finger relative to the EGR binding site, and indeed the +KTS form of WT1 is unable to bind this sequence (*4*, *9*). Because its conservation and cellular abundance suggest that the +KTS form of WT1 has an in vivo function, we have tested whether zinc fingers of the +KTS form bind to DNA sequences other than the EGR binding site.

We identified binding sites, from the human genome, for both forms of the WT1 zinc fingers, with whole-genome polymerase chain reaction (PCR) (*10*). The zinc fingers from both the +KTS and −KTS forms of WT1 were expressed as β-galactosidase (β-gal) fusion proteins (Fig. 1B) (*11*) and bound in vitro to human DNA, converted to a form suitable for amplification by PCR by ligation to catch-linkers (*12*). After multiple rounds of binding, the amplified DNA was cloned and sequenced. Clones derived from this procedure fell into two distinct classes depending on the splice form of WT1 used (Fig. 2A). Clones isolated with the −KTS form of WT1 contained runs of GT dinucleotides, as found in the $(CA)_n$ class of repeated DNA sequence. Such sequences can present G residues in many of the correct positions for contact with EGR-type zinc fingers. Electrophoretic mobility-shift assays were performed on the whole-genome PCR clones (Fig. 2B) (*13*). The $(GT)_n$-containing clones showed reduced mobility with the form of WT1 (−KTS) used in their isolation (Fig. 2B). As judged by competition experiments, the binding affinity of −KTS WT1 for these sequences is at least four-

Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh, EH4 2XU, Scotland.

*To whom correspondence should be addressed.

fold lower than for the EGR site because addition of excess unlabeled −P17 (or +P5) did not compete with the binding of −KTS WT1 to oligo A. Isolation of these sequences probably reflects their relative abundance (5×10^4 copies) in the genome (*14*). We expect that such sequences are not available for binding to WT1 in vivo. Although many of the clones isolated with −KTS consisted mostly of $(GT)_n$ sequences, an oligonucleotide (oligo C) that contained a similar sequence (Fig. 2A) did not appear to bind −KTS in vitro. Thus whole-genome PCR can identify DNA binding sites not recognized in the context of simple oligonucleotides, and the retardation we observed with the whole-genome PCR clones suggests that multiple protein molecules may bind to each DNA target (see multiple shifted bands in Fig. 2B).

In contrast, we did not isolate any $(GT)_n$-containing clones from genomic DNA using the +KTS form of WT1 (Fig. 2A), and the +KTS form failed to retard the mobility of the −KTS genomic clones (Fig. 2B). However, we did recover genomic clones with +KTS (Fig. 2A), and two of these (+P2 and +P5) were retarded with the +KTS zinc fingers (Fig. 2B). Therefore, the major spliced form of the WT1 gene encodes a DNA binding protein. Both +P2 and +P5 also bound −KTS.

To investigate the nature of the binding of +KTS genomic clones to WT1, we used synthetic oligonucleotides (oligo A, B, and C) in binding and competition experiments (*13*). Both +P2 and +P5 contained EGR1-like target motifs (Fig. 2A), which should at least bind −KTS. The mobility of oligo B, derived from +P5, was retarded by both forms of the protein (Fig. 3A); however, a molar excess of this oligonucleotide did not impede the binding of +KTS or −KTS to +P2 or +P5. On the other hand, an EGR consensus oligonucleotide (oligo A) did compete with the binding of +P2 and +P5 to −KTS WT1 but was unable to compete with +P2's and +P5's binding to the +KTS form of the protein (Fig. 3B). Thus we have identified conditions (under competition) where +KTS binds to a DNA sequence not recognized by −KTS.

There is another situation where +KTS-binding clones failed to bind to −KTS WT1. In some isolates of WT1, the serine (residue 365) that precedes the first contact arginine residue of finger 2 (Fig. 1A) is replaced by a phenylalanine (*1*). This substitution had little effect on binding to the EGR site (Fig. 3A) and −KTS genomic clones (Fig. 4), but it slightly decreased binding of both +KTS and −KTS to oligo B (Fig. 3A). The binding of −KTS to clones +P2 and +P5 was severely impaired by this substitution, possibly reflecting the increased importance of finger 2 in the binding of −KTS to these clones and the disruption caused to this binding by the presence of a Phe side chain adjacent to the contact Arg residue. By contrast, +KTS with Phe at position 365 still bound the +KTS genomic clones (Fig. 4). This Ser to Phe substitution was first detected in a cDNA clone from a leukemia cell line (*1*) and may represent a naturally occurring polymorphism or a mutation. We believe, however, that this may be a deleterious mutation for several reasons. First, the Ser at position 365 is conserved throughout vertebrate evolution (*15*). Second, this substitution so far has not been detected in ~100 normal chromosomes (*16*). Last, our binding studies indicate that this substitution can affect WT1 binding to selected targets. If we are correct in arguing that the Ser to Phe substitution is deleterious, the DNA targets isolated with +KTS WT1 assume physiological significance as the binding of these and not the EGR1 site is affected by the mutation.

It has been proposed that WT1 may function as an EGR1 antagonist, binding to

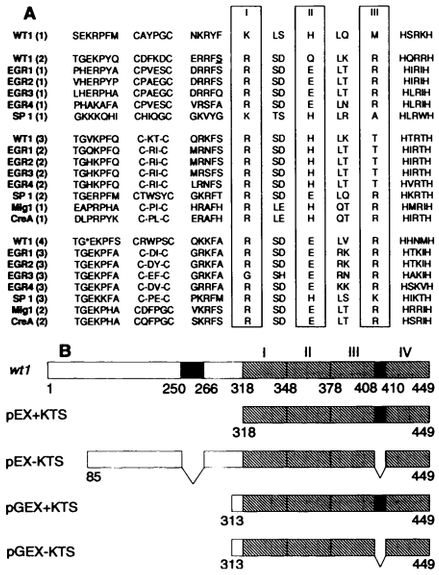

Fig. 1. Organization of WT1 proteins. (**A**) Zinc fingers of WT1 and other members of the EGR family. The numbers in parentheses indicate the number of the zinc finger, with zinc finger 1 appearing first in the primary sequence of the protein. The amino acid residues that contact G residues in the DNA are indicated by the boxed areas I, II, and III. The position at which alternative splicing introduces KTS into WT1 between fingers 3 and 4 is indicated by an asterisk. The site of the Ser-Phe variation in finger 2 of human WT1 is underlined and in bold (*1*). The amino acid sequences of the human proteins EGR 1, 2, 3, and 4, Sp I, the *Saccharomyces cerevisiae* protein Mig1, and the *Aspergillus nidulans* protein CreA are also shown (*19*). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**B**) The structure of the *wt1* gene and constructs described here (*11, 13*). The numbers below each diagram indicate nucleotide position. Zinc fingers 1 to 4 (I to IV) are at the COOH-terminal end of the protein (nucleotides 318 to 449). There are two alternatively spliced regions (filled boxes), the first (nucleotides 250 to 266) introduces 17 amino acids NH_2-terminal to the DNA binding domain. The second (nucleotides 408 to 410) introduces three amino acids (KTS) between the third and fourth zinc fingers.

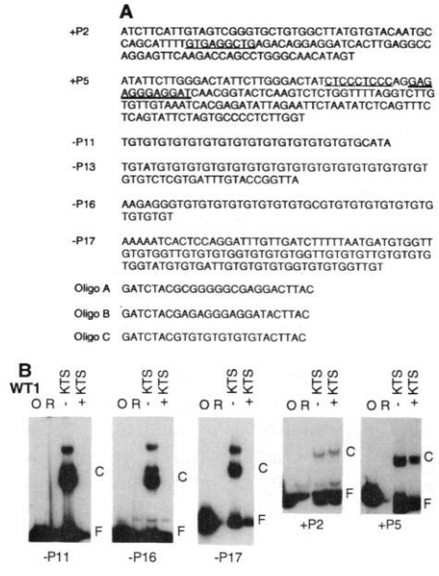

Fig. 2. Identification of genomic sequences that bind WT1. (**A**) Sequences of clones generated by whole-genome PCR with WT1. The electrophoretic mobility of these DNA fragments was retarded in the presence of WT1. The +P2 and +P5 sequences were identified with the +KTS form of WT1. EGR1-like binding motifs within these clones are underlined. Clones −P11, −P13, −P16, and −P17 were obtained with the −KTS form. Oligonucleotides A, B, and C (oligo A, B, and C) were also used in electrophoretic mobility-shift experiments. Oligo A contains the consensus binding site for EGR1; oligo B is derived from a region underlined in bold in clone +P5, and oligo C is derived from −KTS clones. (**B**) Electrophoretic mobility-shift assays of cloned genomic fragments in the presence and absence of various forms of WT1. The genomic fragment used is indicated below each panel. The protein used in each assay is indicated above each respective lane. 0, no extract added; R, extract that contained a pGEX-WT1 construct with the zinc fingers cloned in the reverse orientation so that no fusion protein is induced; −KTS, the −KTS form of WT1; +KTS, the +KTS form of WT1. The positions of the free (F) and the complexed (C) DNA are indicated.

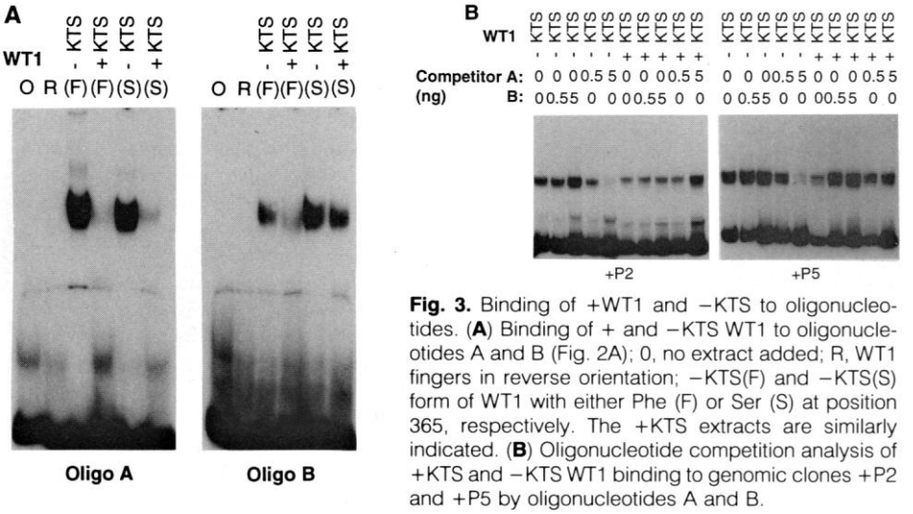

Fig. 3. Binding of +WT1 and −KTS to oligonucleotides. (**A**) Binding of + and −KTS WT1 to oligonucleotides A and B (Fig. 2A); 0, no extract added; R, WT1 fingers in reverse orientation; −KTS(F) and −KTS(S) form of WT1 with either Phe (F) or Ser (S) at position 365, respectively. The +KTS extracts are similarly indicated. (**B**) Oligonucleotide competition analysis of +KTS and −KTS WT1 binding to genomic clones +P2 and +P5 by oligonucleotides A and B.

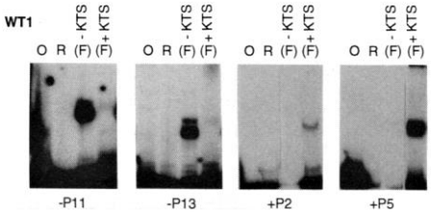

Fig. 4. Electrophoretic mobility-shift assays of +KTS and −KTS genomic targets by WT1 with a Phe at position 365. O, no extract; R, extract containing reverse orientation WT1; −KTS(F) and +KTS(F), extracts that contain both forms of WT1 fingers with Phe at position 365.

EGR targets and repressing transcription of genes activated by EGR1 (*17*). The +KTS form of WT1 is ineffectual in this mode of function because it does not bind the EGR site (*4*, *9*), yet it represents 70 to 80% of WT1 mRNA. Because no antibody is now available that is able to discriminate between the +KTS and the −KTS forms of WT1, the respective protein concentrations are not known. We propose that target DNA binding sites recognized by +KTS WT1 may be important in the understanding of WT1 function and that the different forms of WT1 may fulfill diverse cellular functions by binding different DNA targets and regulating distinct target genes. It remains to be seen what sequences confer optimal binding to the +KTS form of WT1. There is no evidence so far for tissue-specific regulation of this alternative splicing, as the ratio of these two forms of WT1 appears to be both temporally and spatially constant (*8*).

Changing the DNA binding specificity of a transcription factor by alternative splicing may be a general mechanism of gene regulation. Another example of alternative splicing in zinc fingers has been described. The mammalian *Evi-1* gene encodes a protein with ten zinc fingers, and a splice variant results in the omission of two zinc fingers (*18*). This probably affects the binding specificity of this protein, but this has not been shown directly. The alternative splicing observed in WT1 is more subtle than the elimination of entire fingers as in *Evi-1*. The link between adjacent zinc fingers is thought to be important for positioning of neighboring fingers relative to the DNA (*6*). In our assays both +KTS and −KTS bound the genomic clones isolated with +KTS. Therefore, introduction of three extra amino acids between the third and fourth zinc fingers of WT1 did not affect the ability of the protein to bind these targets. It may be that fingers 1 to 3 are most important in the binding of WT1 to such targets, so that the displacement of the fourth finger by the introduction of KTS has little effect. By contrast WT1 fingers 2 to 4 (those most similar to EGR1) are probably the most important for binding to the EGR1 site and to genomic clones isolated with −KTS.

REFERENCES AND NOTES

1. K. M. Call *et al.*, *Cell* **60**, 509 (1990).
2. M. Gessler *et al.*, *Nature* **343**, 774 (1990).
3. K. P. Pritchard-Jones *et al.*, *ibid.* **346**, 194 (1990); V. Huff *et al.*, *Am. J. Hum. Genet.* **47(S)**, 28 (1990); J. Pelletier *et al.*, *Cell* **67**, 437 (1991).
4. F. J. Rauscher III, J. F. Morris, O. E. Tournay, D. M. Cook, T. Curran, *Science* **250**, 1259 (1990).
5. J. Nardelli, T. Gibson, C. Vesque, P. Charnay, *Nature* **349**, 175 (1991).
6. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
7. B. Christy and D. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8737 (1989).
8. D. A. Haber *et al.*, *ibid.* **88**, 9618 (1991).
9. J. Pelletier *et al.*, *Cell* **67**, 437 (1991).
10. K. W. Kinzler and B. Vogelstein, *Mol. Cell. Biol.* **10**, 634 (1990).
11. The pEX+KTS and pEXWT−KTS were constructed in the pEX β-gal expression vectors. A Bam HI–Sal I fragment derived from pB2.1 was cloned into pEX2 to create pEX+KTS. We derived pEX−KTS by cloning a Sal I–Pst I fragment of pSKII-WT33 into pEX3.
12. Whole-genome PCR was carried out as described [K. W. Kinzler and B. Vogelstein, *Nucleic Acids Res.* **17**, 3645 (1989)]. Catch-linked human DNA (300 ng) and poly(dI-dC) (4.5 μg) were added to DNA binding buffer (30 μl) (50 mM Hepes (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 1 mM dithiothreitol (DTT), 20% glycerol). Insoluble nonfusion protein (30 μg) was added and incubated on ice for 20 min. Insoluble protein and bound DNA were pelleted to remove sequences binding to nonfusion proteins. This achieved a higher enrichment of sequences binding to the fusion protein. Poly(dI-dC) (4.5 μg) and insoluble protein containing either +KTS or −KTS fusion proteins (30 μg) were added to the supernatant. After 20 min on ice, insoluble protein and bound DNA were recovered. Bound DNA was released by digestion with proteinase K and amplified by PCR to act as substrate for further rounds of binding. The PCR reactions were carried out with primers A (5′-GAGTAGAATTCTAATATCTC-3′) and B (5′-GAGATATTAGAATTCTACTC-3′) and cycled (30 cycles) at 90°C for 1 min, 45°C for 2 min, and 70°C for 2 min. Selected amplified DNA (300 ng) was rebound to fusion protein. After four rounds of binding and amplification, the DNA was cut with Eco RI and was cloned. We screened transformants by hybridization to bound versus unbound DNA. Cloned sequences that hybridized strongly and specifically to the bound DNA were sequenced.
13. The pGEX+KTS and pGEX−KTS were constructed in the Bam HI site of pGEX3X. We obtained the Bam HI fragments (residues 313 to 449) by reverse transcriptase-PCR of WT1 mRNA, with primers B374 (5′-CCGTGGATCCTGGGACACTGAACGGTCC-3′) and B375 (5′-GCCTGGATCCTTGTACGGTCGGCATCTGAG-3′) located 3′ and 5′ to the WT1 fingers, respectively. Reactions were cycled (30 cycles) at 95°C for 1 min, 54°C for 2 min, and 72°C for 2 min. The sequences of the resulting products were verified, and the orientation of the fusion constructs was ascertained with a monoclonal antibody (48.83) to the COOH-terminal peptides of WT1. This antibody was also used to quantitate WT1 protein in extracts. The fusion proteins (Fig. 1B) were induced and cells were resuspended in 50 mM Hepes (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 1 mM DTT, 1% Triton X-100, phenylmethylsulfonyl fluoride (PMSF) (20 μg/ml), and lysozyme (10 μg/ml). Cells were lysed by sonication and were dialyzed against 50 mM Hepes (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 1 mM DTT, PMSF (20 μg/ml), and 20% glycerol. For electrophoretic mobility-shift analysis ^{32}P-end–labeled cloned insert DNA (1 ng) or oligonucleotide prepared as described [P. Chavrier *et al.*, *EMBO J.* **9**, 1209 (1990)] was mixed with poly(dI-dC) (1 μg) and where appropriate, with cold competitor DNA in DNA binding buffer, all in a total volume of 9 μl. Protein extract was added that contained 50 to 100 ng of WT1 fusion protein, and the mixture was incubated on ice for 20 min before electrophoresis on 6% acrylamide and 0.5× TBE (0.045 M tris-borate, 0.001 M EDTA) gels.
14. M. Litt and J. A. Luty, *Am. J. Hum. Genet.* **44**, 397 (1989).
15. J. Kent, unpublished observations.
16. D. A. Haber *et al.*, *Cell* **61**, 1257 (1990); M. H. Little *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press; K. Miyagawa, unpublished data.
17. S. L. Madden *et al.*, *Science* **253**, 1550 (1991).
18. D. Bordereaux, S. Fichelson, P. Tambourin, S. Gisselbrecht, *Oncogene* **5**, 925 (1990).
19. S. V. Suggs, J. L. Katzowitz, C. Tsai-Morris, V. P. Sukhatme, *Nucleic Acids Res.* **18**, 4283 (1990); S. Patwardhan *et al.*, *Oncogene* **6**, 917 (1991); J. O. Nehlin and H. Ronne, *EMBO J.* **9**, 2891 (1990); C. E. A. Dowzer and J. M. Kelly, *Mol. Cell. Biol.* **11**, 5701 (1991).

20. W.A.B. is a Lister Institute Research Fellow, K.O. is supported by a Medical Research Council Human Genome Mapping Project grant, and M.H.L. is supported by the Royal Society. We thank D. Housman for providing B2.1 and WT33, B. Vogelstein for catch-linked human DNA, A. Gallagher for oligonucleotide synthesis, and the Medical Research Council Human Genetics Unit photography department for preparation of figures.

24 February 1992; accepted 15 May 1992