Cell, Vol. 62, 193–201, July 13, 1990, Copyright © 1990 by Cell Press

A Major Segment of the Neurofibromatosis Type 1 Gene: cDNA Sequence, Genomic Structure, and Point Mutations

Richard M. Cawthon,* Robert Weiss,* Gangfeng Xu,* David Viskochil,† Melanie Culver,* Jeff Stevens,* Margaret Robertson,* Diane Dunn,* Ray Gesteland,* Peter O'Connell,* and Ray White*
* Department of Human Genetics
and Howard Hughes Medical Institute
† Department of Pediatrics
University of Utah School of Medicine
Salt Lake City, Utah 84103

Summary

Overlapping cDNA clones from the translocation breakpoint region (TBR) gene, recently discovered at the neurofibromatosis type 1 locus and found to be interrupted by deletions and a t(17;22) translocation, have been sequenced. A 4 kb sequence of the transcript of the TBR gene has been compared with sequences of genomic DNA, identifying a number of small exons. Identification of splice junctions and a large open reading frame indicates that the gene is oriented with its 5′ end toward the centromere, in opposition to the three known active genes in the region. PCR amplification of a subset of the exons, followed by electrophoresis of denatured product on native gels, identified six variant conformers specific to NF1 patients, indicating base pair changes in the gene. Sequencing revealed that one mutant allele contains a T→C transition changing a leucine to a proline; another NF1 allele harbors a C→T transition changing an arginine to a stop codon. These results establish the TBR gene as the NF1 gene and provide a description of a major segment of the gene.

Introduction

Neurofibromatosis type 1 (NF1) is one of the most frequent and important of the inherited disorders, affecting 1/3500 individuals (Stumpf et al., 1987) with peripheral neurofibromas and café-au-lait spots and an increased risk of learning disabilities, renal hypertension, and malignancies (see Riccardi and Eichner, 1986). Recent work has identified a new gene at the NF1 locus, the translocation breakpoint region (TBR) gene, and shown it to be interrupted by deletions and a chromosome translocation breakpoint (Viskochil et al., 1990 [accompanying paper]). These findings strongly implicate the TBR gene as the NF1 gene. However, it remained possible that these rearrangements could be compromising the relationship between some other gene and its regulatory elements and that this other gene is the real NF1 gene. This hypothesis is of specific concern, as we have already characterized several other genes in the immediate vicinity that have structural or functional properties consistent with an NF1 gene (O'Connell et al., 1990; Buchberg et al., 1990; Cawthon et al., 1990; Mikol et al., 1990; Viskochil et al., unpublished data).

If, however, the TBR gene is indeed the NF1 gene, and the rearrangements are having their effect through its inactivation, then at least some NF1 mutations would be expected to be base pair changes that significantly alter the amino acid sequence of the gene product. To look for base pair change mutations specific to NF1 patients, we have developed DNA sequence from our previously identified cDNA clones. Comparison of this cDNA sequence with genomic DNA sequence from the region indicates a number of small exons distributed over a long stretch of genomic DNA.

Identification of intron–exon boundaries provided a basis for synthesis of oligonucleotide primers that will amplify exons of the TBR gene from genomic DNA of NF1 patients and controls. Because single-strand conformation polymorphism (SSCP) has been shown to be a sensitive indicator of sequence variation at the level of even a single base pair (Orita et al., 1989a, 1989b), we have used it to scan several exons of a large number of NF1 patients and controls. Sequencing of fragments of altered mobility taken directly from the SSCP gels provided confirmation of the underlying genetic change.

Results

cDNA Sequencing

Thirty cDNA clones reflecting sequences from the RNA transcript of the TBR gene had been previously isolated from human fetal brain libraries (Viskochil et al., 1990). Figure 1 shows the cDNA sequence obtained from the overlapping fetal brain clones FB4, FB6, FB8, FB9, FB13, FB16, FB18, FB20, FB35, FB37, and FB38. With one exception, DNA sequencing of overlapping fetal brain cDNAs has provided the continuous transcript sequence shown in Figure 1. The exception is in one region of exon 3 where 118 bp of genomic sequence (see below) was used to join two cDNA sequences. No intron–exon boundaries were seen within the 118 bp of genomic sequence. There are no stop codons in the open reading frame of the sequence from position 3 to position 3309. It should be noted that a second open reading frame, phase-shifted by one base, starts approximately 100 bp upstream of this stop codon and continues for ~600 bp past the stop codon. In addition, in one of our cDNA clones there is a 54 bp insertion between positions 3563 and 3564, downstream of the stop codon but suggesting inclusion of an additional exon through alternate splicing of the message. None of the cDNA clones sequenced have a poly(A) tract at the cloning site, and no two clones end with identical sequences. We conclude that our 4 kb cDNA meld is likely internal to the transcript, which may be as large as 11 kb (Viskochil et al., 1990).

The predicted amino acid sequence of the large open reading frame found in the TBR cDNA was used to search

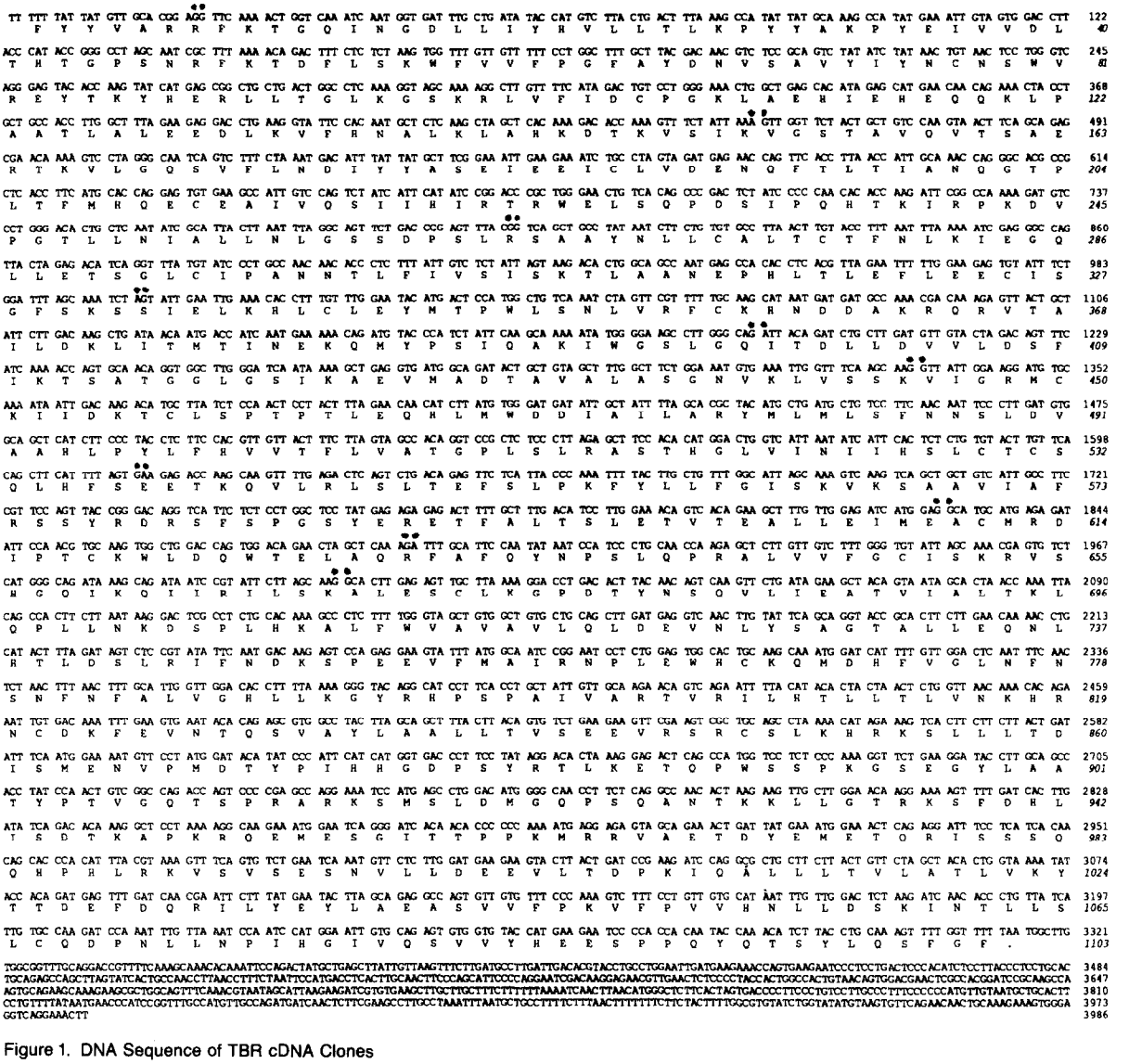

Figure 1. DNA Sequence of TBR cDNA Clones

The 3986 nucleotides of DNA sequence generated from the overlapping set of cDNA clones defining the TBR gene (Viskochil et al., 1990) is shown with translation of the 1103 amino acid open reading frame. The known exon splice junctions for exons 1 through 9 are indicated by double dots at the exon boundaries. Nucleotides 1–780 are from cDNA FB37, 942–1190 from FB9, 1120–1890 from FB20, 1190–1610 from FB16, 1792–1968 from FB6, 1886–3238 from FB18, 1969–3373 from FB4, 2877–3020 from FB9, 2886–3163 from FB35, 3020–3310 from FB38, 2554–3986 from FB13, and 1930–3986 from FB8. A gap in contiguous sequence between nucleotides 824 and 942 is spanned by genomic sequence from cEV120. FB13 contains an additional 54 nucleotide sequence (AGCTTCTCTGCCTTGCTCTAACTCAGCAGTTTTCATGCAGCTGTTCCCTCATCA) inserted at nucleotide 3563 of the sequence.

for homology to known DNA and protein sequences. No significant homologies have been detected thus far.

Genomic DNA Sequences

Genomic DNA sequences come primarily from a new method of multiplex sequencing (Weiss et al., unpublished) that was applied to the cosmid cEVI20 (O'Connell et al., 1990) and to the 3.8 kb EcoRI fragment harboring the t(17;22) breakpoint. Gaps between sequence contigs were filled in by primer walking or, in a few cases, by asymmetric polymerase chain reaction (PCR) amplification and sequencing of M13 subclones. The entire sequence of the 3.8 kb EcoRI fragment has been obtained, as has the entire sequence of the 9 kb EcoRI fragment containing the EVI2 gene. The open reading frame of the cDNA sequence and the intron–exon boundary sequences agree with the orientation of transcription as 5′ to 3′ in the centromeric to telomeric direction.

Structure of the NF1 Gene

Detailed mapping of exons from the cDNAs was performed through sequence comparison between cDNAs and genomic DNA. Figure 2 shows the sequences of the exons and their flanking intronic sequences for 50 bp on either side of the intron–exon boundaries. Figure 2 also shows the locations of the exons on the restriction frag-

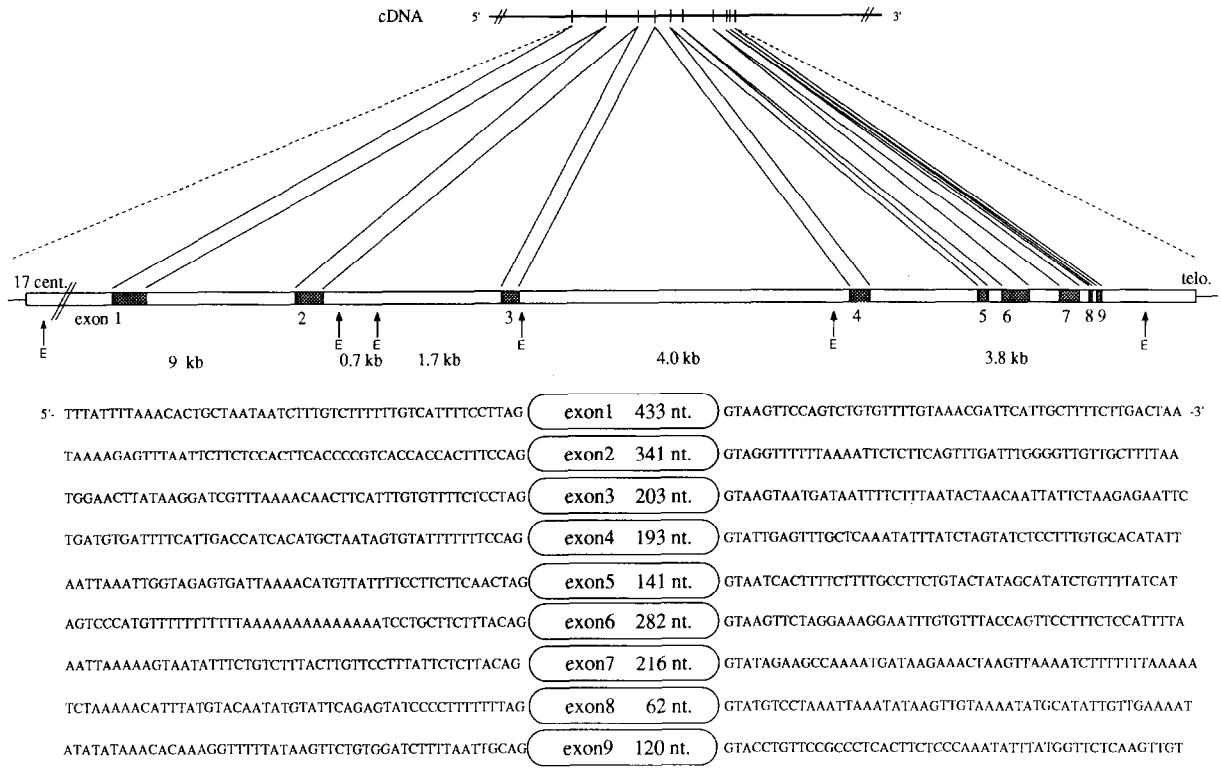

Figure 2. Intron-Exon Boundaries of TBR Exons 1 to 9

Intronic sequences (50 nucleotides in each case) flanking each of the nine exons are shown in the centromeric to telomeric orientation. These sequences were derived from cEVI20 and the 3.8 kb EcoRI fragment from cT311. The length of each exon is shown, as well as its schematic location within the TBR composite cDNA and the EcoRI contig mapped from cEVI20 and cT311. The actual sizes of the 9 and 3.8 kb EcoRI fragments are 8840 and 3889 nucleotides, respectively, by sequence determination.

ment map of the region surrounding the translocations. The most proximal (5′) exon mapping to cloned genomic sequence, provisionally denoted as exon 1, lies in the 9 kb EcoRI fragment containing the 5′ noncoding exon of EVI2 (Cawthon et al., 1990). The EVI2 5′ exon lies 4120 bp centromeric of exon 1, with an Alu repeat located midway between. Exon 1 is 433 bp long, and the intron between exons 1 and 2 also contains an Alu repeat. Exon 2 is also completely within the 9 kb EcoRI fragment. Exon 3 maps to the 1.7 kb EcoRI fragment shown previously to be deleted in an 11 kb NF1 deletion mutation (Viskochil et al., 1990). Exons 4–9 map to the 3.8 kb EcoRI fragment spanning the translocation breakpoint. The intron between exons 4 and 5 contains the t(17;22) breakpoint (R. K. Wolff, personal communication). The 22 bp of cDNA sequence extending 5′ from exon 1 is not found in the 9 kb fragment, indicating that the TBR gene extends centromeric of this fragment.

Table 1. Sequences of Primers Used for PCR–SSCP Analyses

Primer Pair	Exon(s) Flanked	Sequence
A	4	5′-ATAATTGTTGATGTGATTTTCATTG-3′
		5′-AATTTTGAACCAGATGAAGAG-3′
B	5	5′-ATCTAGTATTTTTGAGGCCTCAG-3′
		5′-CAGATATGCTATAGTACAGAAGG-3′
C	6	5′-CATATCTGTTTTATCATCAGGAGG-3′
		5′-AAGTAAAATGGAGAAAGGAACTGG-3′
D	7	5′-CAAAATGAAACATGGAACTTTAGA-3′
		5′-TAAGCATTAAGTACAAATAGCACA-3′
E	7–9	5′-TTTATGTTTGTGCTCTAACACCAAGT-3′
		5′-ATAAATGCTAGAATGATTTCTCATGCT-3′

The first primer in each pair lies 5′ of the exon or set of exons that it amplifies. PCR was performed as described in Experimental Procedures.

Identification of Nucleotide Alterations in NF1 Patients

SSCP is a rapid and sensitive assay for nucleotide alterations, including point mutations (Orita et al., 1989a, 1989b). DNA segments 100–400 bp in length are amplified by PCR, heat denatured, and electrophoresed on high resolution, nondenaturing acrylamide gels. Under these conditions each single-stranded DNA fragment assumes a secondary structure determined in part by its nucleotide sequence. Several examples of single base changes significantly affecting the electrophoretic mobility of the PCR product have been reported (Orita et al., 1989b; Dean et al., 1990).

Five pairs of oligonucleotide primers were used to screen exons 4–9 of the TBR gene. The sequences of the primers are given in Table 1. Primer pairs A–D specifically

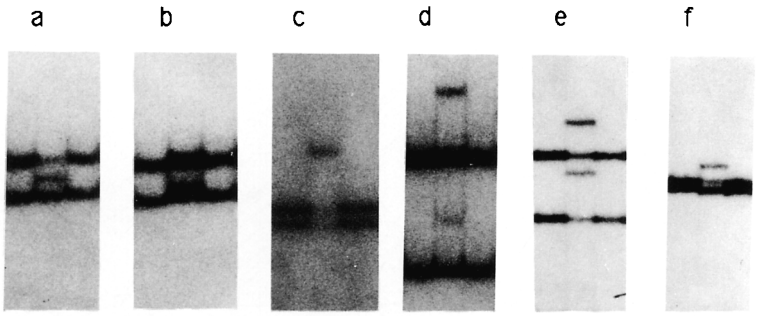

Figure 3. SSCP Variants Specific to NF1 Patients

The middle lanes of the six panels show unusual SSCP alleles that were found in NF1 individuals. The two flanking lanes in each panel show alleles with mobilities that were shared by all the non-NF1 samples as well as most of the NF1 samples. The PCR product analyzed in (a), (b), and (c) spans exon 4 of the new NF1 candidate gene. The PCR product in (d) contains exon 5; the product in (e) contains both exon 8 and exon 9; and the product in (f) contains exon 7. The patients with aberrant alleles are as follows: patient 11423 (a), patient 11476 (b), patient 11528 (c), patient 11389 (d), patient 11578 (e), and patient 11572 (f).

amplify each of exons 4 through 7, along with some flanking intron sequences. Primer pair E amplifies a single 1.3 kb genomic segment containing exons 7, 8, and 9. After amplification, the products of E-primed PCR were digested with a combination of the restriction enzymes RsaI and XbaI to yield fragments in a size range amenable to SSCP analysis. One of the resulting fragments contains exon 7 and another fragment contains exons 8 and 9.

Each primer pair was used to amplify specific DNA segments from 72 NF1 patients and 60–65 non-NF1 controls. After PCR, the products of the reaction were initially examined by agarose gel electrophoresis and ethidium bromide staining. In all samples with visible product, a single band of appropriate size was seen, and no differences were detected between NF1 individuals and non-NF1 controls (data not shown).

However, upon subjecting these PCR products to SSCP analysis (Figure 3), six alleles specific to NF1 individuals were found. Several patterns can occur when heterozygotes are detected by SSCP analysis. The simplest pattern is that of Figure 3e. The homozygous samples in the first and third lanes have only two bands, representing the two complementary single strands of DNA. The heterozygote in the middle lane has two additional bands corresponding to the two strands of the new allele. In Figure 3f, only three distinct bands are seen in the heterozygote; one strand of the new allele is able to assume a novel conformation with a different electrophoretic mobility, but the other strand, although it has experienced the complementary sequence change, had no new conformations available to it, and so continues to migrate with the corresponding strand of the normal allele. As expected, the aberrant band and one of the bands from the normal allele are of equal intensity, and each of these is approximately half the intensity of the third band. In Figure 3c we believe that both strands of the new allele migrate together; apparently under these gel conditions they have not been resolved. Our interpretation is supported by the greater intensity of the new band in comparison with the bands of the normal allele. We know that this band does not represent reannealed double-stranded DNA, because the nondenatured sample has a different mobility (data not shown).

Three of the variants were found in the segment containing exon 4, one was in the segment containing exon 5, one was in the segment spanning exon 7, and one was in an RsaI restriction fragment containing exons 8 and 9. No two individuals shared any one of the six variant alleles, and no variant allele was observed in the control samples. The six alleles were found in four familial cases and in two sporadic cases of NF1. The entire PCR–SSCP procedure was repeated on the DNA samples in which the variant alleles were detected and on several control samples. All six variant SSCP bands were reproducible.

The most likely explanation for these results is that at least some of these NF1 individuals carry a DNA sequence change within the amplified DNA segment that causes NF1. We should also consider, however, the null hypothesis that all six alleles actually represent DNA polymorphisms not associated with NF1. Assuming the null hypothesis, we can calculate the probability that none of the six alleles would be found upon screening the non-NF1 individuals. When a one-sided test of heterogeneity is performed using Fisher's exact test, the frequency of observed variants is significantly greater in NF1 patients than in controls at a nominal level of 0.05.

Genomic DNA samples from the unaffected spouse and affected child of patient 11423 were available. As shown in Figure 4, SSCP analysis revealed that the unusual allele originally detected in the father is inherited by his NF1-affected son. THE NF1 father, lane a, and NF1 child, lane c, show identical SSCP patterns consisting of four bands appearing as two doublets. The unaffacted mother has just the lower band of the upper doublet and the upper band of the lower doublet. The differences in intensity of bands from one lane to another are due to differences in amount of PCR product amplified and loaded on the gel. This experiment confirms that the sequence variant that gives rise to the variant SSCP allele is carried by the same chromosome 17 that carries the NF1 mutation.

Nucleotide Sequences of Variant Alleles Detected by SSCP Analysis

To determine the nucleotide changes responsible for the altered electrophoretic mobilities of the variant bands, we sequenced several of the associated PCR products. The band of altered mobility was cut from the dried SSCP

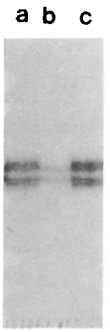

Figure 4. Coinheritance of an SSCP Variant and NF1 in the Family of Patient 11423

Lane a contains the sample from patient 11423; lane b is the sample from the unaffected spouse of patient 11423; and lane c is the sample from the NF1-affected child of this couple. (The patterns observed are different from those seen in Figure 3a because the gels were run under different conditions. See Experimental Procedures.)

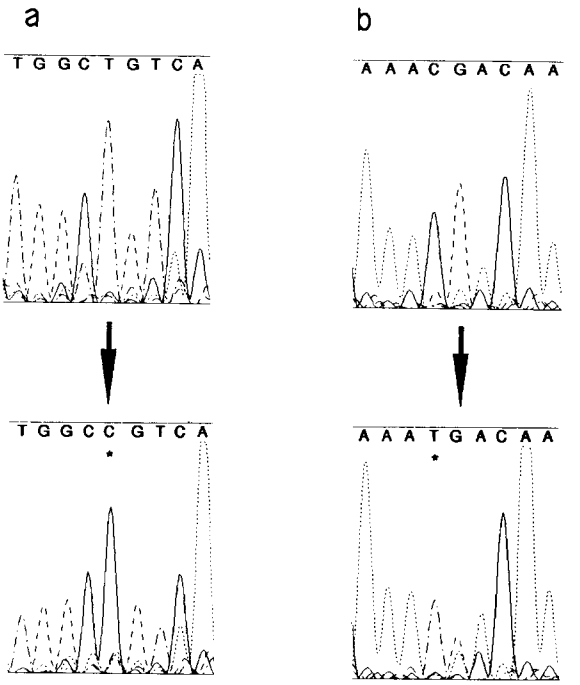

Figure 5. Sequences of Two NF1 Alleles

(a) Sequence of the variant SSCP allele from patient 11423. The normal sequence of nucleotides 1041–1049 (see Figure 1) is shown at the top; the bottom panel shows the T→C point mutation that has occurred at position 1045 in this patient.

(b) Sequence of the variant allele of patient 11528. The normal sequence of positions 1084–1092 is shown at the top. A C→T transition has introduced a stop codon in the coding sequence in this patient (bottom panel).

acrylamide gel, and DNA was eluted from the gel slice, reamplified, and sequenced on an Applied Biosystems model 373A automated sequencer (Hata et al., personal communication). Sequencing of the altered SSCP allele of exon 4 from patient 11423 (the allele shown to be inherited along with the NF1 mutation in the family) revealed that it contains a T→C transition (Figure 5a) at base 1045 of the cDNA sequence (Figure 1), changing a leucine residue to a proline residue in the predicted polypeptide. Sequencing of the altered SSCP allele derived from the same exon from patient 11528 revealed a C→T transition at base 1087 of the cDNA sequence, changing an arginine residue to a stop codon (Figure 5b). For each of these mutations, sequencing was performed on both strands of the PCR products, with complete agreement.

Discussion

The findings reported here confirm that the TBR gene is the gene for neurofibromatosis type 1. Six base pair variants were revealed among 72 NF1 patient DNAs. A set of 60–65 control, unaffected individuals showed no variant bands. The likelihood of this observation occurring due to chance if the six variants were population polymorphisms unrelated to NF1 is less than 0.05. This provides good support for the hypothesis that at least some of these base pair change variants are, in fact, NF1 mutations in the TBR gene.

Furthermore, at least some of the base pair change variants should have a major impact on the amino acid sequence of the gene product. Sequence data from two of the variant bands indicate a nucleotide substitution within exon 4. The predicted amino acid change in one case results in a leucine to proline substitution. Demonstration of the transmission of this variant band from the affected parent to the affected child confirmed that it reflected the NF1 mutant allele. The other case was even more dramatic; the nucleotide substitution creates an in-frame stop codon, truncating the protein at this point. These NF1 point mutations, unlike the deletions and translocation previously shown to interrupt the TBR gene, are unlikely to have a regulatory impact on other genes in the region and thus uniquely identify the TBR gene as the NF1 gene.

Each of the mutant alleles described here and in the accompanying paper (Viskochil et al., 1990) is expected to inactivate the NF1 gene. The dominant inheritance of NF1 raises the intriguing question of how a mutation that inactivates the gene product results in a dominant allele. One possibility is simply that there is a gene dosage effect and that cells with only 50% activity of the NF1 gene become vulnerable to secondary muations at other loci resulting in growth deregulation.

A more interesting hypothesis, by analogy with retinoblastoma, is that the mutant NF1 allele might be a recessive at the level of the cell (Cavenee et al., 1983). The phenotype, the emergence of a neoplastic clone of cells, would result from the loss or inactivation of the normal allele in the progenitor of the cell clone. In the case of retinoblastoma, loss of function of the normal allele occurs most frequently through a loss of the chromosome, or a major segment of the chromosome, carrying the normal allele, presumably as a nondisjunctional event at mitosis. Loss of chromosome 17 is seen in malignant Schwannomas from NF1 patients (neurofibromas are not amenable to such studies as they consist of cell mixtures); however, the

interpretation is complicated by the fact that p53, a known recessive oncogene, is located on 17p, the short arm of chromosome 17. The loss of the entire chromosome, therefore, can be interpreted as due to the presence of a mutated p53 gene. The only informative losses are those where only one arm of the chromosome is lost. Although most often partial chromosome loss is only of 17p (Menon et al., 1990; Ponder, personal communication), loss of chromosome 17q (the NF1 gene is on 17q) has been reported in only two cases (Skuse et al., 1989). There seem to be relatively few NF1-specific chromosome losses associated with the malignancies.

The high frequency of new mutation seen in the NF1 gene, however, suggests that loss of the normal allele through somatic mutation might occur much more frequently than mitotic nondisjuctions. If that were the case, then we would not expect to see an appreciable increase in loss of 17q in NF1 tumors and could maintain the cell recessive hypothesis. This model makes the strong prediction, which is now testable, that two mutant NF1 alleles, a new somatic mutation in addition to the inherited mutation, will often be found in NF1 malignant Schwannomas.

Interestingly, 22 bp at the 5′ end of the merged cDNA sequence shown in Figure 1 is not found in exon 1 nor anywhere within the genomic sequence of the 9 kb EcoRI fragment, suggesting an exon located 5′ of this EcoRI fragment. Furthermore, preliminary PCR amplifications of the 5′ ends of two of our cDNA clones, FB24 and FB17, using a facing primer from exon 4 and a vector primer (data not shown), suggest transcript sequences 850 bp and 1050 bp greater in length than can be accounted for by the lengths of exons 1–4. Hybridization of these two cDNAs to Southern transfers of restriction digests of cosmids from our contig spanning the region indicated no homology with restriction fragments proximal (5′) to the 9 kb EcoRI fragment. Combined with the large size of the NF1 transcript, approximately 11 kb, it is likely that the NF1 transcript spans the t(1;17) translocation breakpoint in addition to the t(17;22) breakpoint.

The suggestion that the NF1 gene spans, in opposing orientation, at least three active genes within a major intron raises the intriguing possibility that some mutations in this region may have several effects, impacting two or more genes jointly. Is it possible simultaneously to transcribe nested genes in opposite orientation? If not, then activation of any of the three genes within the intron, EVI2, RC1, or OMGP, might inhibit transcription of the NF1 gene. Careful examination of patient phenotypes in NF1 deletion cases might be revealing. EVI2, OMGP, and the NF1 gene are each transcribed in brain, and EVI2, RC1, and the NF1 gene are known to be present in peripheral blood (Cawthon et al.. 1990).

Could the learning disabilities of some NF1 patients be explained by removal or inactivation of either the EVI2 or OMGP gene? Alteration of the OMGP gene is an especially intriguing possibility, owing to its known role in oligodendrocytes as an externally located, cell adhesion molecule potentially active in mediating proper cell motility and differentiation during brain development (Mikol et al., 1990). Similarly, the higher frequency of juvenile chronic myelogenous leukemias that has been found among NF1 patients (Bader and Miller, 1978; Mays et al., 1980; Clark and Hutter, 1982) might be related to NF1 mutations that jointly disrupt EVI2 or RC1. The embedded genes provide even more motivation than usual for clinical correlations of specific mutations with phenotypic characteristics.

The large size of the gene transcript may provide a basis for understanding the high frequency of new NF1 mutations. Our preliminary Northern blot analysis (Viskochil et al., 1990) suggests a message with an approximate length of 11 kb. Interestingly, our finding in a screen of only 1014 bp of exonic DNA that 6 out of approximately 70 patients yielded variant SSCP bands, of which two were confirmed as single base substitutions, suggests that the NF1 gene may have a large proportion of point mutations.

We are, however, scanning only a specific section of the gene and may find that deletions are more abundant in other regions. For example, mapping of the cDNA sequence against genomic DNA sequences has revealed that the NF1 gene structure, for the 4 kb we have examined, consists of a series of relatively small exons distributed over more than 110 kb of genomic DNA. Intriguingly, this distribution of small exons seems characteristic of very large genes as, for example, the Duchenne muscular dystrophy gene (Koenig et al., 1987) and the chromosome 18q gene found to be deleted in colon carcinomas (Fearon et al., 1990). This raises the possibility that NF1 may also be distributed over a very large genomic region, which (if true) may also be a large target for deletions and rearrangements not yet characterized.

SSCP analysis of two additional individuals, not reported above, gave notable preliminary findings (data not shown). The first of these was the finding that the mother of patient 11578 showed the same variant band as the patient, unexpected because both parents and her sibling have been examined and diagnosed as being unaffected by NF1. One possibility is that she may actually carry a mutant NF1 allele but not be so severely affected as to be readily detected clinically. Detailed examination might reveal mild NF1 signs or symptoms. A second possibility is that the variant in this case may be a polymorphism unrelated to NF1. Although our statistical analysis indicates that it is highly unlikely that all of the variants observed among the NF1 patients would be due to chance polymorphism, a finding of one or two polymorphisms in the NF1 population but not seen in the controls would not be unexpected.

The second finding of note was the occurrence of an SSCP variant in exon 5 (data not shown) of a patient with a history of bilateral acoustic neuromas and multiple paraspinal neurofibromas, diagnosed as NF2. This variant is not one of the six reported among the NF1 patients. NF2 is believed to be distinct from NF1, and is characterized by central, as opposed to peripheral, neurofibromas and acoustic neuromas that frequently lead to deafness. Mapping studies have indicated that the mutated NF2 gene in at least some families is located on chromosome

22 (Rouleau et al., 1987). One possibility is that the variant we detected is a polymorphism unrelated to the pathogenesis of NF2. The other possibility, that the apparent NF2 presentation in this case is actually due to a mutation in the NF1 gene, is more intriguing, and would suggest that NF2 is genetically heterogeneous.

This description of the NF1 gene constitutes a beginning. As the gene becomes more fully characterized, the molecular biology and molecular pathophysiology associated with the many mutations should emerge. Molecular probes for RNA transcripts will define the developmental and tissue-specific patterns of expression. As the amino acid sequence becomes more completely predicted, clues to the functional characteristics of the gene product should suggest new and specific experiments. Antibodies to specific domains of the gene product will reveal the subcellular localization of the protein and suggest possible functions. The many mutations known for the NF1 gene will also play an important role in understanding its normal functions. Since it has now been shown that point mutations in the amino acid sequence can cause NF1, we expect to find some mutations of reduced expressivity or penetrance, increasing the sophistication of clinical diagnosis and prognosis.

Most importantly, it can now be hoped that the basic molecular pathophysiologic description of this disease will be obtained, and that it will lead to effective means of intervention.

Experimental Procedures

cDNA Sequencing

All DNA sequencing was based on the dideoxy termination method of Sanger et al. (1977). Double-stranded plasmid preparations of cDNA clones FB4, FB8, FB13, FB18, and FB20 in Bluescript SK(−) were made by the cesium chloride–ethidium bromide centrifugation method (Maniatis et al., 1982), denatured by treatment with 0.2 N NaOH, neutralized, and sequenced from M13 and cDNA sequence-specific primers using the T7 polymerase sequencing kit 27-1682-01 (Pharmacia). DNA sequence was also generated on these cDNA clones from minilysate supercoiled preparations (Birnboim, 1983). cDNA clones FB35, FB37, and FB38 (in λgt10) and FB20 and FB4 (in λ Zap) were symmetrically amplified from plaques using specific vector-based forward and reverse primers with exon-specific primers generated from previous DNA sequence. The amplified products were asymmetrically amplified, and the single strands were sequenced either manually or with fluorescently tagged M13 primers on an Applied Biosystems model 370A DNA sequencer (Applied Biosystems, Inc., Foster City, CA), using T7 DNA polymerase or Taq DNA polymerase. cDNA FB9 was symmetrically amplified with M13 primers and cDNA sequence-specific primers for asymmetric PCR sequencing. DNA sequencing using the primer walk stategy gave overlapping clone cDNA sequence that was merged using the IntelliGenetics suite of programs (IntelliGenetics Inc., Mountain View, CA) running on a VAX computer. DNA sequences were aligned, examined for open reading frames, and compared with DNA sequences in the EMBL and GenBank data bases and the PIR and SWISS protein data bases.

Oligonucleotide primers for sequencing and enzymatic amplification were synthesized on an Applied Biosystems 380B DNA synthesizer using 0.2 μmol scale synthesis.

Sequencing of Cosmid Clones

DNA sequence from cosmid clone cEVI20 was obtained by multiplexed dideoxy sequencing of random subclones (Church and Kieffer-Higgins, 1988; R. Weiss et al., unpublished data). cEVI20 (O'Connell et al., 1990) DNA was sonicated, size fractionated (500–1000 bp) on agarose gels, and blunt-end ligated to a set of ten vectors carrying multiplex identifier tag sequences (pKZ vectors). A total of 720 primary clones were isolated (72 per vector) and distributed into 72 mixtures, each containing ten clones (one clone for each of the ten vectors). Plasmid DNA pools were prepared and then sequenced as mixtures. Chain extension and termination with T7 DNA polymerase (Sequenase, United States Biochemical Corp.) were primed simultaneously from two sites directly flanking the unique identifier tags and inserts, resulting in a mixture of 20 sequence ladders per reaction. These mixed sequences were electroblotted from standard sequencing gels onto GeneScreen nylon membranes, and probed sequentially with oligonucleotides complementary to the 20 nucleotide identifier tags. A total of 690 readable sequence ladders were recovered from 20 probings; these 125,172 nucleotides melded into 61 contigs (approximate total melded sequence = 38,000 nucleotides) covering ~90% of cEVI20. The sequence of the 9 kb EcoRI fragment containing the 5′ noncoding exon of EVI2 (Cawthon et al., 1990) and exons 1 and 2 of the TBR cDNA (Viskochil et al., 1990) was completed on both strands (actual size = 8,840 nucleotides) by augmenting the sequence data recovered from the sonicated clones with 27,335 nucleotides of sequence generated from 96 isolates of the 9 kb EcoRI fragment, each containing a random insert of a Tn3-based multiplexed transposable sequencing vector. The sequence of the 3.8 kb EcoRI fragment from cT311 (Viskochil et al., 1990) was generated by a combination of random clones from the cEVI20 multiplexed pool and Tn3-based multiplexed sequencing vector inserts into a 3.8 kb EcoRI subclone. Gaps were closed in both of these cases by selected primer walking.

Contigs were melded using the GEL program in the IntelliGenetics package implemented on a VAX 8600. FASTA searches (Pearson and Lipman, 1988) of the National Biomedical Research Foundation (NBRF) protein sequence data base (release 19.0, 6/89) and translated GenBank (release 61.0) and EMBL (release 19.0) data bases were performed with the conceptual translation of the TBR open reading frame. The PAM250 matrix was used in the second step of the comparison, and the search was run using the UWGCG software package (Devereux et al., 1984). FASTDB searches were also performed with the IntelliGenetics package on GenBank (release 62.0) and EMBL (release 20.0) data bases.

Patient Tissues

NF1 patients participating in this study met criteria agreed upon in a consensus conference at the National Institutes of Health (Stumpf et al., 1987). Lymphoblastoid cell lines from normal and NF1 individuals were established in our laboratory by transformation of peripheral blood mononuclear cells with Epstein–Barr virus. All human tissue samples used for research were obtained with the approval of the Institutional Review Board at the University of Utah Health Sciences Center.

Identification of SSCPs

PCRs were performed with the primer pairs listed in Table 1, with 100 ng of genomic DNA, 70 μM each dNTP, 0.5 μM each primer, 1.5 mM $MgCl_2$, 10 mM Tris (pH 8.3), 50 mM KCl, 0.25 U of Taq polymerase, and 0.1 μl of [α-^{32}P]dCTP (3000 Ci/mmol, 10 mCi/ml) in a volume of 10 μl. The product was then diluted 100-fold in 0.1% SDS, 10 mM EDTA, followed by a 1:1 dilution in 95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol. Once in this loading solution, the samples were heated at 90°C for 3 min to denature the DNA, then loaded onto 4.5% nondenaturing acrylamide gels (49:1 ratio of acrylamide to methylene-bis-acrylamide) measuring 38 cm × 20 cm × 0.4 mm, containing 90 mM Tris–borate (pH 8.3), 2 mM EDTA. Electrophoresis was carried out at 4°C at 40 W, constant power. Samples were also screened under a second set of conditions, which consisted of the addition of 10% glycerol to the gel solution and performing the electrophoresis at room temperature.

After electrophoresis, the gel was transferred to Whatman 3MM paper and dried on a vacuum slab gel dryer. Autoradiography with Kodak X-Omat AR film at −70°C with an intensifying screen overnight was usually sufficient to allow detection of the PCR products.

Direct Sequencing of SSCP Alleles

Individual SSCP bands were cut directly from the dried gel, placed in 100 μl of distilled water, incubated at 37°C with shaking for 1 hr, and centrifuged briefly to pellet debris. An aliquot (1–10 μl) of the supernatant was then used directly in a 100 μl PCR reaction. The primers used in the original amplification were used in this amplification as well, except that the 5′ end of one primer contained additional sequence consisting of the universal sequencing primer from M13, and the 5′ end of the other primer included the sequence of the reverse sequencing primer.

The double-stranded PCR product resulting from this amplification was purified by two centrifugation-washes with a Centricon 100 column (Amicon), then sequenced following a test-site protocol suggested by Applied Biosystems, Inc. (Foster City, CA). This involved performing the dideoxy sequencing reactions with Taq polymerase in a thermal cycler, using fluorescently tagged M13 universal or reverse sequencing primers, followed by gel electrophoresis and data collection and analysis on an Applied Biosystems model 373A automated sequencer.

PCR Amplification

cDNA clone inserts were amplified by PCR under standard conditions. The λ Zap clones were amplified with the −21M13 universal primer (GTAAAACGACGGCCAGT) and the reverse primer from Bluescript polylinker sequence (AACAGCTATGACCATG). The λgt10 clones were amplified with vector-based primers synthesized by E. Meenen, HHMI: GT10FP (AGCAAGTTCAGCTGGTTAAGT) and GT10RP (TTATGAGTATTTCTTCCAGGGT). Purified plaque suspensions were diluted 1:70 and heated to 70°C for 10 min, and 10 μl was used in a 50 μl reaction volume. The amplification conditions were as follows: 50°C annealing for 1 min, 72°C extension for 3.5 min, and 94°C for 1 min for 30 cycles. Products were either digested with EcoRI and electrophoresed through low melting agarose or subjected to ultrafiltration with Centricon 100 prior to radiolabeling.

Acknowledgments

We are grateful to Jean-Marc Lalouel, Steve Hunt, and Lynne Jorde for their advice on the statistical issues arising in the comparison of the frequencies of variant SSCP bands in NF1 cases vs. controls. We are also indebted to R. A. Hepler for assistance with the preparation of the manuscript, to B. Ottrud for assistance with the figures, and to Ed Meenen for synthesis of the oligonucleotides used for sequencing and PCR. This work was supported in part by the National Neurofibromatosis Foundation (D. V.). R. W. is an Investigator, P. O'C. a Senior Associate, and R. M. C. an Associate of the Howard Hughes Medical Institute.

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.

Received June 26, 1990.

References

Bader, J. L., and Miller, R. W. (1978). Neurofibromatosis and childhood leukemia. J. Pediatr. *92*, 925–929.

Birnboim, H. C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. Meth. Enzymol. *100*, 243–255.

Buchberg, A. M., Bedigian, H. G., Jenkins, N. A., and Copeland, N. G. (1990). Identification of a putative protooncogene, *Evi-2*, in murine myeloid leukemias that is closely linked to the von Recklinghausen neurofibromatosis locus in humans. Mol. Cell. Biol., in press.

Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C., and White, R. L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature *305*, 779–784.

Cawthon, R. M., O'Connell, P., Buchberg, A. M., Viskochil, D., Weiss, R. B., Culver, M., Stevens, J., Jenkins, N. A., Copeland, N. G., and White, R. (1990). Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of *EVI2* and mapping of other transcripts. Genomics, in press.

Church, G., and Kieffer-Higgins, S. (1988). Multiplex DNA sequencing. Science *240*, 185–188.

Clark, R. D., and Hutter, J. J., Jr. (1982). Familial neurofibromatosis and juvenile chronic myelogenous leukemia. Hum. Genet. *60*, 230–232.

Dean, M., White, M. B., Amos, J., Gerrard, B., Stewart, C., Khaw, K.-T., and Leppert, M. (1990). Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. Cell *61*, 863–870.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. *12*, 387–395.

Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W., and Vogelstein, B. (1990). Identification of a chromosome 18q gene which is altered in colorectal cancers. Science *247*, 49–56.

Feinberg, A., and Vogelstein, B. (1984). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *137*, 266–267.

Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., and Kunkel, L. M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509–517.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 93–94.

Mays, J. A., Neerhout, R. C., Bagby, G. C., and Koler, R. D. (1980). Juvenile chronic granulocytic leukemia: emphasis on cutaneous manifestations and underlying neurofibromatosis. Am. J. Dis. Child. *13*, 654–658.

Menon, A. G., Ledbetter, D. H., Rich, D. C., Seizinger, B. R., Rouleau, G. A., Michaels, V. F., Schmidt, M. S., Dewald, G., DellaTorre, C. M., Haines, J. L., and Gusella, J. F. (1989). Characterization of a translocation within the von Recklinghausen neurofibromatosis region of chromosome 17. Genomics *5*, 245–249.

Menon, A. G., Anderson, K. M., Riccardi, V. M., Chung, R. Y., Whaley, J. M., Yandell, D. W., Farmer, G. E., Freiman, R. N., Lee, J. K., Li, F. P., Barker, D. F., Ledbetter, D. H., Kleider, A., Martuza, R. L., Gusella, J. F., and Seizinger, B. R. (1990). Chromosome 17p deletions and p53 gene mutations associated with the formation of malignant neurofibromas in von Recklinghausen neurofibromatosis. Proc. Natl. Acad. Sci. USA *87*, in press.

Mikol, D. D., Gulcher, J. R., and Stefansson, K. (1990). The oligodendrocyte-myelin glycoprotein belongs to a distinct family of proteins and contains the HNK-1 carbohydrate. J. Cell Biol. *110*, 471–480.

O'Connell, P., Viskochil, D., Buchberg, A. M., Fountain, J., Cawthon, R. M., Culver, M., Stevens, J., Rich, D. C., Ledbetter, D. H., Wallace, M., Carey, J. C., Jenkins, N. A., Copeland, N. G., Collins, F. S., and White, R. (1990). The human homolog of murine *Evi-2* lies between two von Recklinghausen neurofibromatosis translocations. Genomics, in press.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA *86*, 2766–2770.

Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. (1989b). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics *5*, 874–879.

Pearson, W. R., and Lipman, D. J. (1987). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444–2448.

Reed, K. C., and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. *13*, 7207–7221.

Riccardi, V. M., and Eichner, J. E. (1986). Neurofibromatosis: Phenotype, Natural History, and Pathogenesis (Baltimore: Johns Hopkins University Press).

Rouleau, G. A., Wertelecki, W., Haines, J. L., Hobbs, W. J., Trofatter, J. A., Seizinger, B. R., Martuza, R. L., Superneau, D. W., Connealy, P. M., and Gusella, J. F. (1987). Genetic linkage of bilateral acoustic neurofibromatosis to a DNA marker on chromosome 22. Nature *329*, 246–248.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Skuse, G. R., Kosciolek, B. A., and Rowley, P. T. (1989). Molecular genetic analysis of tumors in von Recklinghausen neurofibromatosis: loss of heterozygosity for chromosome 17. Genes, Chromosomes and Cancer *1*, 36–41.

Stumpf, D. A., Alksne, J. F., Annegers, J. F., Brown, S. S., Conneally, P. M., Housman, D., Leppert, M., Miller, J. P., Moss, M. L., Pileggi, A. J., Rapin, I., Strohman, R. C., Swanson, L. W., and Zimmerman, A. (1987). Neurofibromatosis. Arch. Neurol. *45*, 575–578.

Viskochil, D., Buchberg, A. M., Xu, G., Cawthon, R. M., Stevens, J., Wolff, R. K., Culver, M., Carey, J. C., Copeland, N. G., Jenkins, N. A., White, R., and O'Connell, P. (1990). Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell *62*, this issue.