MOLECULAR AND CELLULAR BIOLOGY, Feb. 1991, p. 906–912
0270-7306/91/020906-07$02.00/0

Vol. 11, No. 2

The Gene Encoding the Oligodendrocyte-Myelin Glycoprotein Is Embedded within the Neurofibromatosis Type 1 Gene

DAVID VISKOCHIL,[1]* RICHARD CAWTHON,[2] PETER O'CONNELL,[2] GANGFENG XU,[2] JEFF STEVENS,[2] MELANIE CULVER,[3] JOHN CAREY,[1] AND RAY WHITE[2]

Department of Pediatrics[1] and Howard Hughes Medical Institute and Department of Human Genetics,[2] University of Utah School of Medicine, Salt Lake City, Utah 84132

Received 25 September 1990/Accepted 30 November 1990

In the course of efforts to identify the neurofibromatosis type 1 gene (*NF1*), three genes were found embedded within an intron of *NF1*. The cDNA sequence of one of these genes (*OMGP*) encodes oligodendrocyte-myelin glycoprotein. *OMGP* spans at least 2.7 kb of genomic DNA, and it maps within 4 kb of the breakpoint of a balanced chromosomal translocation carried by an individual with NF1. *OMGP* is similar in genomic structure to two other expressed genes, *EVI2A* and *EVI2B*, which lie approximately 20 and 5 kb telomeric of the *OMGP* locus, respectively. All three genes have the same transcriptional orientation and are contained within one intron of *NF1*, which is transcribed off the opposite strand. Whether altered expression of *OMGP* might play a role in the clinical heterogeneity of NF1 is as yet unclear.

Neurofibromatosis type 1 (NF1) is a heritable condition that affects 1 in approximately 4,000 people (9, 35). Although the dominant mutant allele seems fully penetrant, the clinical expressivity is highly variable, ranging from medically insignificant café-au-lait skin spots to malignant neurofibrosarcomas. The gene for NF1 has been cloned and partially characterized (7, 38, 39). Two previously characterized genes, *EVI2A* and *EVI2B*, are embedded in an intron of the NF1 gene (7, 27, 40), and both are transcribed from the complementary strand of *NF1* (6). *EVI2A* and *EVI2B* encode putative transmembrane peptides, and the mouse homologs (*Evi-2a* and *Evi-2b*) are associated with ecotropic viral insertions (5, 6a) which have been implicated in the expression of murine myeloid leukemias (4). *EVI2A* and *EVI2B* are similar in that both have single introns separating 5′ noncoding exons from single 3′ coding exons (6), and both genes map between two *NF1* translocation breakpoints (27). The functions of *EVI2A* and *EVI2B* are unknown; however, tissue-specific mRNA expression suggests distinct functions: *EVI2A* is expressed in brain, peripheral blood, and bone marrow, and *EVI2B* is expressed only in peripheral blood and bone marrow (6). We describe here a third gene, *OMGP*, lying within the same intron of *NF1*.

MATERIALS AND METHODS

Cell lines and tissue. Epstein-Barr virus-transformed lymphoblastoid cell lines from normal and NF1 individuals were suspension cultured. The NF13 (17), DCR1 (22), SP3, and L(17n)C (36) somatic cell hybrid cell lines were cultured as previously reported (27).

cDNA library screening. An oligo(dT)-primed human cDNA library derived from frontal cortex of a 2-year-old female was purchased from Stratagene, San Diego, Calif. (lambda-ZAP; catalog no. 935203). Phage from the libraries were incubated with *Escherichia coli* LE392 and plated according to the manufacturer's instructions. Duplicate plaque lifts were made with Biotrans filters (0.2-μm pore size; Pall Biodyne) according to published methods (2, 12).

DNA probes were radiolabeled by the technique of Feinberg and Vogelstein (11). Cosmid and cDNA inserts were separated from the vector by restriction enzyme digestion followed by gel electrophoresis in low-melting-point agarose (NuSieve GTG agarose; FMC Bioproducts, Rockland, Maine). DNA was labeled directly in gel slices. The filters from the cDNA library were prehybridized in a solution of 50% formamide, 5× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 50 mM sodium phosphate (pH 6.5), 2× Denhardt solution, and 500 μg of sheared, denatured human DNA per ml for 2 h at 42°C. Separate prehybridization of the radiolabeled probes was simultaneously carried out in a solution of the same composition except that DNA (20 μg/ml) from the vector used to clone the probe was included. Hybridization was carried out for 16 h at 42°C. Washes were in 0.1× SSC–0.1% sodium dodecyl sulfate (SDS), three times at room temperature for 30 min and twice at 65°C for 30 min. After washing, filters were wrapped in plastic and exposed with an intensifying screen on X-Omat AR film at −70°C.

Southern analysis. Five micrograms of *Eco*RI-digested DNA from each somatic cell hybrid and human lymphoblastoid cell line was fractionated on 1% agarose gel in Tris-acetate buffer. Gels were denatured in 0.4 N NaOH for 30 min and transferred to Hybond-N+ filters (Amersham, Arlington Heights, Ill.) overnight in 0.4 N NaOH. Membranes were washed twice in 2× SSC for 15 min and air dried.

Radiolabeling of DNA probes was carried out according to Feinberg and Vogelstein (11). Blots were prehybridized in a solution of 10% polyethylene glycol, 8% SDS, and 500 μg of total human DNA per ml for 2 to 12 h prior to hybridization with probes that had also been prehybridized for 30 to 180 min with total human DNA and with vector DNA to decrease background. Hybridization was carried out for 12 to 16 h. Before air drying and autoradiography, filters were washed at room temperature twice with 2× SSC–0.1% SDS and twice with 0.1× SSC–0.1% SDS for 20 min each time and then with 0.1× SSC–0.1% SDS for 5 min at 65°C.

DNA sequencing. All DNA sequencing was based on the dideoxy termination method of Sanger et al. (31). Double-stranded plasmid preparations of cDNA clones in Bluescript SK(-) and of Bluescript subclone pT336, excised from lambda-ZAP by in vivo excision as instructed by the manufac-

* Corresponding author.

turer (Stratagene), were made by the cesium chloride-ethidium bromide centrifugation method (21). The preparations were denatured with 0.2 N NaOH, neutralized, and sequenced from M13 and cDNA sequence-specific primers, using T7 polymerase sequencing kit 27-1682-01 (Pharmacia, Uppsala, Sweden). The ends of the cDNAs were sequenced with fluorescently tagged M13 primers on an Applied Biosystems model 370A DNA sequencer (Applied Biosystems, Inc., Foster City, Calif.). Internal cDNA was sequenced manually with cDNA sequence-specific oligonucleotide primers that had been constructed on an Applied Biosystems 380B DNA synthesizer using 0.2-μmol-scale synthesis.

Subclone pT336 originated from cosmid EVI36 (27). The original *Sau*3A site from the T3 end of the cosmid clone marks the proximal end of the human genomic sequence represented in the subclone.

RESULTS

OMGP was identified with an evolutionarily conserved cloned DNA sequence, an approach that has been effective in searching for other human genes (26, 30). Cross-hybridization between mouse and human DNAs on Southern blots revealed conservation within an 11.5-kb *Eco*RI fragment that contains the translocation breakpoint from an NF1 patient with a balanced t(1;17) chromosomal rearrangement (6). A subcloned 10-kb fragment from an *Eco*RI-partial *Sau*3A digest, pT336, identified eight cDNA clones from a human frontal cortex cDNA library. The longest clone, HB36.2, was 1.65 kb and contained a 16-base poly(A) tract and polyadenylation signal sequence (AATAAA) located 18 bases 5′ from the poly(A) tract. The nucleotide sequence of HB36.2, consisting of bases −24 through 1609 in the genomic sequence shown in Fig. 1, is identical to the cDNA sequence encoding the oligodendrocytemyelin glycoprotein (OMgp), a putative cell adhesion molecule (24, 25).

Southern blot analysis of genomic DNA probed with HB36.2 confirmed that the cDNA mapped uniquely to the expected 11.5-kb *Eco*RI fragment in total human DNA and in a somatic cell hybrid containing chromosome 17q as its only human chromosome (Fig. 2). Furthermore, Southern blot hybridization of *Eco*RI-digested DNA from a rodent × human somatic cell hybrid line (DCR1) containing the 1;17 translocation derivative chromosome (22) with HB36.2 as a probe identified a 7.8-kb derivative *Eco*RI fragment that encompasses the breakpoint (Fig. 2, lane 5). Examination of 120 NF1 individuals with this probe failed to identify a 7.8-kb *Eco*RI fragment; therefore, the 7.8-kb fragment identified by HB36.2 in the derivative chromosome is unlikely to be a polymorphism and must represent the t(1;17) derivative fragment. This result places the HB36.2 cDNA clone within 7.8 kb of the 1;17 translocation breakpoint. Surprisingly, the region of conserved rodent-human homology spanning the *OMGP* locus was contained within a rodent *Eco*RI fragment that by coincidence is similar in size to the 11.5-kb *Eco*RI fragment encompassing the t(1;17) breakpoint in human genomic DNA.

The map position of HB36.2 was further defined by Southern blot analysis of the genomic insert from subclone pT336 (Fig. 3a). The cDNA clone HB36.2 hybridized to two of four restriction fragments: a 1.9-kb *Hin*dIII-*Hin*dIII fragment (HH1.9) and the 3.2-kb *Hin*dIII-*Eco*RI fragment (HE3.2). Fragments HH1.9 and HE3.2 were shown to be adjacent because the cDNA sequence from *OMGP* was identical to the genomic sequence spanning the shared *Hin*dIII site (base 753 in Fig. 1). HE3.2 represents the distal terminus because it contains the *Eco*RI cloning site; fragment SH1.8 lies at the proximal end of the insert because it contains the *Sau*3A cloning site. Fragment HH2.9 therefore lies between SH1.8 and HH1.9.

To better define the t(1;17) translocation breakpoint with respect to *OMGP*, each of the four insert fragments from pT336 was subcloned and hybridized to Southern blots containing *Eco*RI-digested somatic cell DNA from the DCR1 hybrid cell line. Fragments HH2.9 (Fig. 3c), HH1.9 (Fig. 3d), and HE3.2 (data not shown) identified the 7.8-kb derivative genomic *Eco*RI band from DCR1 and the 11.5-kb genomic *Eco*RI fragment from control DNA, while fragment SH1.8 revealed only the 11.5-kb *Eco*RI fragment in normal DNA and failed to detect the derivative band from DCR1 (Fig. 3b). These results placed the t(1;17) breakpoint within the HH2.9 fragment. By sequence analysis, the 3′ end of *OMGP* is approximately 1 kb distal of the shared *Hin*dIII site between fragments HH2.9 and HH1.9; therefore, *OMGP* lies less than 4 kb but more than 1 kb from the t(1;17) breakpoint. The translocation breakpoint region defined by these experiments is denoted by the stippled box in Fig. 3a.

We sequenced genomic DNA encompassing *OMGP* by dideoxy-terminated primer extensions directed by oligonucleotides derived from the HB36.2 cDNA sequence. *OMGP* spans at least 2.5 kb of genomic DNA and contains one intron separating a 5′ noncoding exon from the coding exon. The genomic sequence extends 314 bases upstream from the 5′-end nucleotide (base −971) of the OMgp cDNA sequence reported by Mikol et al. (24) (Fig. 1). There are 77 bases of 5′ noncoding exon in *OMGP*, from bases −971 to −893. The 5′ splice site shows a typical primate consensus sequence, NAG|GT$^{G}_{A}$AGT, and the 814-base intervening sequence has a consensus 3′ splice acceptor site, TNTTTGTTTCTAG|G, at base −79 (33). The initiation codon reported by Mikol and Stefansson (25) is at base −51. However, an alternative initiation codon exists 21 bases upstream from the proposed start of the translated peptide. This alternative in-frame initiation codon (base −72) adds seven amino acids to the signal peptide, including two basic residues (Gln and Lys) and one acidic residue (Glu). The cDNA sequence encoding the mature polypeptide starts at base 1 and displays an uninterrupted open reading frame, terminating with a valine at base 1263. The 3′ noncoding region is 346 bases long and contains two instances of the ATTTA pentamer, a motif common to the 3′ untranslated regions of many oncogene and cytokine mRNAs (3). No obvious TATAAA- or CAATC-like sequences are evident in the genomic sequence upstream from the most 5′ end of *OMGP* cDNA sequence. However, an octamer sequence in the intervening sequence at base −119, ATTTGCAT, 40 bases upstream from the 3′ splice site, suggests the presence of an enhancer element (10). The 5′ end of the gene has not been established.

Because *OMGP* was a candidate for the NF1 gene, we screened our panel of NF1 individuals with DNA sequences from the *OMGP* locus in an effort to detect mutations. Southern blot analysis of DNAs from 87 unrelated NF1 patients, digested with *Eco*RI and *Bam*HI and probed with HB36.2 cDNA, failed to identify derivative bands or quantitative reductions in band signal intensities relative to control probes. The open reading frame of *OMGP* was screened for NF1-specific mutations by means of the polymerase chain reaction–single-strand conformation polymorphism method of Orita et al. (28), and the entire *OMGP* open reading frame from 14 non-NF controls and 12 unrelated NF1 patients was examined. We detected one single-strand conformation polymorphism that lies in the segment be-

```
                                                                        5'  tgaggcaaatgttaatgaggcaatg  -1261
ttaaatatggacccaatgtcagacaaatacatagaaaggagtaagggccaactctcatgcataaggtatcccatcctatagcaaatcagatatataggtacgcttga  -1154
tgccacaaatttttaaaaaattgtccattttgttgcgtgtgcacctcttgccataaatttgagtcagcaccagcgacagctctgcagtcctcctatgtggtactga  -1047
tcaggtggttgcagagcttcagctcacagcaacacaatgcagctgagcaggcaagcacagcccacagccagaaacaGTTCCGACTCTACAGAACAAGACGACCTTTA  -940
AGTTTCCCAGAGAAAATGAGATGCTGATGTTGAAGACGACACCACGGgtaagatgttatttaaatcagtaaaaggctgactttggaatctttttcctttttcttta  -833
agaaaaagtcaacgttaggattaaatatatattcaatagcaagtgcatgcaccagaaatttgctgcagtgtcagttgagggatatttttatacattcagtcactct  -726
gtaaatatacatattgttttcctttaaaatgggcactgaaatatacagaaaaaaatcactttataaaatgtgaggtttataggtactgtgttggtctggatttttca  -619
agtgcttttacaaagatatatttatcctaaaaacatacagataaaaatttcgaagactgctttaatatctaaataaaatctaccctatatacacacattgaattac  -512
attacctgcagagattaaaaaaaaaagacacgacagccatttttctcatctgagtaagaaagcatatcatcaaaaatagtaatagcctacaactgcaactatttatt  -405
tgcaaagaatgctattttatcatattaaggctctagaaagataaataagaaagaatatggttagaaaagggggagggagagagaaaataaaggagaaaatgcagga  -298
gagagtagggagagagtctctctctaccacatagcccaatgaaggattaagcattgactataaatgaagggagctttgttagtttaatcactggaacaattataaaa  -191
ggactcgacaacaacgaggtttattgaaaattttgcctaatgctaactgacccatgcagatgcctaaactgtatttgcatattaaaagaagggtgtatctgtttgtt  -84

tctagGCTTTG ATG GAA TAT CAG ATA TTG AAA ATG TCT CTC TGC CTG TTC ATC CTT CTG TTT CTC ACA CCT GNT ATT TTA TGC  -1
             M   E   Y   Q   I   L   K   M   S   L   C   L   F   I   L   L   F   L   T   P   X   I   L   C

ATT TGT CCT CTC CAA TGT ATA TGC ACA GAG AGG CAC AGG CAT GTG GAC TGT TCA GGC AGA AAC TTG TCT ACA TTA CCA TCT  81
 I   C   P   L   Q   C   I   C   T   E   R   H   R   H   V   D   C   S   G   R   N   L   S   T   L   P   S

GGA CTG CAA GAG AAT ATT ATA CAT TTA AAC CTG TCT TAT AAC CAC TTT ACT GAT CTG CAT AAC CAG TTA ACC CAA TAT ACC  162
 G   L   Q   E   N   I   I   H   L   N   L   S   Y   N   H   F   T   D   L   H   N   Q   L   T   Q   Y   T

AAT CTG AGG ACC CTG GAC ATT TCA AAC AAC AGG CTT GAA AGC CTG CCT GCT CAC TTA CCT CGG TCT CTG TGG AAC ATG TCT  243
 N   L   R   T   L   D   I   S   N   N   R   L   E   S   L   P   A   H   L   P   R   S   L   W   N   M   S

GCT GCT AAC AAC AAC ATT AAA CTT CTT GAC AAA TCT GAT ACT GCT TAT CAG TGG AAT CTT AAA TAT CTG GAT GTT TCT AAG  324
 A   A   N   N   N   I   K   L   L   D   K   S   D   T   A   Y   Q   W   N   L   K   Y   L   D   V   S   K

AAC ATG CTG GAA AAG GTT GTC CTC ATT AAA AAT ACA CTA AGA AGT CTC GAG GTT CTC AAC CTC AGT AGT AAC AAA CTT TGG  405
 N   M   L   E   K   V   V   L   I   K   N   T   L   R   S   L   E   V   L   N   L   S   S   N   K   L   W

ACA GTT CCA ACC AAC ATG CCC TCC AAA CTA CAT ATC GTG GAC CTG TCT AAT AAT TCT TTG ACA CAA ATT CTT CCA GGT ACA  486
 T   V   P   T   N   M   P   S   K   L   H   I   V   D   L   S   N   N   S   L   T   Q   I   L   P   G   T

TTA ATA AAC CTG ACA AAT CTC ACA CAT CTT TAC CTG CAC AAC AAT AAG TTC ACA TTC ATT CCA GAC CAA TCT TTT GAC CAA  567
 L   I   N   L   T   N   L   T   H   L   Y   L   H   N   N   K   F   T   F   I   P   D   Q   S   F   D   Q

CTC TTT CAG TTG CAA GAG ATA ACC CTT TAC AAT AAC AGG TGG TCA TGT GAC CAC AAA CAA AAC ATT ACT TAC TTA CTG AAG  648
 L   F   Q   L   Q   E   I   T   L   Y   N   N   R   W   S   C   D   H   K   Q   N   I   T   Y   L   L   K

TGG ATG ATG GAA ACA AAA GCC CAT GTG ATA GGG ACT CCA TGT TCT ACC CAA ATA TCA TCT TTA AAG GAA CAT AAC ATG TAT  729
 W   M   M   E   T   K   A   H   V   I   G   T   P   C   S   T   Q   I   S   S   L   K   E   H   N   M   Y

CCC ACA CCT TCT GGA TTT ACC TCA AGC TTA TTC ACT GTA AGT GGG ATG CAG ACA GTG GAC ACC ATT AAC TCT CTG AGT GTG  810
 P   T   P   S   G   F   T   S   S   L   F   T   V   S   G   M   Q   T   V   D   T   I   N   S   L   S   V

GTA ACT CAA CCC AAA GTG ACC AAA ATA CCC AAA CAA TAT CGA ACA AAG GAA ACA ACG TTT GGT GCC ACT CTA AGC AAA GAC  891
 V  .T   Q   P   K   V   T   K   I   P   K   Q   Y   R   T   K   E   T   T   F   G   A   T   L   S   K   D

ACC ACC TTT ACT AGC ACT GAT AAG GCT TTT GTG CCC TAT CCA GAA GAT ACA TCC ACA GAG ACT ATC AAT TCA CAT GAA GCA  972
 T   T   F   T   S   T   D   K   A   F   V   P   Y   P   E   D   T   S   T   E   T   I   N   S   H   E   A

GCA GCT GCA ACT CTA ACT ATT CAT CTC CAA GAT GGA ATG GTC ACA AAC ACA AGC CTC ACT AGC TCA ACA AAA TCA TCC CCA  1053
 A   A   A   T   L   T   I   H   L   Q   D   G   M   V   T   N   T   S   L   T   S   S   T   K   S   S   P

ACA CCC ATG ACC CTA AGT ATC ACT AGT GGC ATG CCA AAT AAT TTC TCT GAA ATG CCT CAA CAA AGC ACA ACC CTT AAC TTA  1134
 T   P   M   T   L   S   I   T   S   G   M   P   N   N   F   S   E   M   P   Q   Q   S   T   T   L   N   L

TGG AGG GAA GAG ACA ACC ACA AAT GTA AAG ACT CCA TTA CCT TCT GTG GCA AAT GCT TGG AAA GTA AAT GCT TCA TTT CTC  1215
 W   R   E   E   T   T   T   N   V   K   T   P   L   P   S   V   A   N   A   W   K   V   N   A   S   F   L

TTA TTG CTC AAT GTT GTG GTC ATG CTG GCT GTC TGA GGGTCTGCATTTTCTGAAACTAATGAAAGCACTCCTCCCTGATGTACAGTTGGGAAAAT  1310
 L   L   L   N   V   V   V   M   L   A   V

ATGTCCATATCTAACCAGTGATTCGAGCTAT[ATTTA]AGTATTCAAGAAAGCCAGTCTTAACATTTCTAACTCTGATGTAAATGAAGTAACTTGTCTTAAATAAAA  1415
GAAATGCACAATGTCTTGGTACTTGCTGCTATTTTACTGTCTTAATTAAGTAAACTAATGAGTTTCTTTTATAAAAAAAATGAAATGTTTTAAGGCTTCA[ATTTA]  1520
TTGCACAAAATATAAAGCATCTAAACTTTAATATGTATTTTATGTATGTTTACACTGTCAAACATCTGGAAAATAAAAGGTCTATGCTCataactgtgtcatttggc  1627
tttccagtcataccaactttnagcagaatcaaaatgacctcaccatttttgttctagggat 3'                                               1688
```

FIG. 1. Nucleotide sequence of a portion of the genomic clone, pT336 (see Fig. 3a), that encompasses *OMGP*. The HB36.2 cDNA sequence agrees with the cDNA sequence and the deduced amino acid sequence reported by Mikol et al. (24). Pending the establishment of the 5′ terminus of the *OMGP* transcript, we selected an arbitrary reference point for nucleotide numbering: the first base of the triplet encoding the N-terminal amino acid (25) in the mature peptide in position 1. Lowercase letters represent genomic sequence, and uppercase letters denote cDNA sequence. The 5′ noncoding exon sequence is from the fetal spinal cord cDNA, c4, reported by Mikol et al. (24). The first potential ATG initiation codon, followed by six amino acids, is underlined; the pentamer associated with decay of mRNA appears twice in the 3′ noncoding exon sequence (boxes). The polyadenylation signal sequence in the 3′ noncoding portion of the cDNA is underlined in boldface, and the arrowhead indicates the poly(A) addition site. Double underlining at base −119 indicates an octamer sequence associated with enhancer elements (10).

tween bases −114 and +167 (Fig. 1). This segment spans the base position −11 that is polymorphic for either an A or a G in several cDNA sequences (24; unpublished data); an A is in this position in the sequenced allele of pT336. No NF1-specific mutation was detected. After eliminating this gene as a candidate for the NF1 gene, we applied identical screening methods to a sequence then known as *TBR* (translocation breakpoint region), a putative coding region identified by sequence conservation that had been mapped to a 3.8-kb *Eco*RI fragment 50 kb telomeric to the *OMGP* locus (38). Probes from this region identified a 40-kb deletion in one individual with NF1 and an 11-kb deletion in another. On the basis of these subsequent studies, *TBR* was identified as the NF1 gene (7, 38).

DISCUSSION

In the course of our efforts to identify the NF1 gene, we isolated cDNAs from a human frontal cortex library with a genomic DNA probe from the t(1;17) translocation breakpoint region; this probe had already been shown to have rodent-human conservation. DNA sequence analysis revealed that one of our cDNA clones, HB36.2, was homologous to the cDNA sequence encoding OMgp. OMgp is an extracellular peptide linked to the outer cell membrane through a glycosylphosphatidylinositol lipid (25), as are Thy-1 (20) and the cell surface form of the neural cell adhesion molecule, NCAM (15). Tandemly repeated blocks of leucine-rich residues, characteristic of a family of peptides having cell-adhesive properties (13, 19, 29, 37), are also present. Mikol et al. (24) reported that antibodies directed against the L2/HNK1-carbohydrate epitope, which is involved in cell-cell and cell-substrate interactions, bind to a subpopulation of OMgp molecules. An epidermal growth factor-like motif at the N terminus of OMgp may impart specificity to the cell adhesion properties of this protein. These structural and functional analogies strongly suggest that OMgp may function as a cell adhesion molecule in the

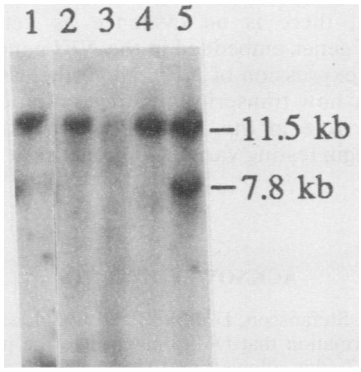

FIG. 2. Southern analysis of *Eco*RI-digested DNA from somatic cell hybrids probed with cDNA HB36.2. Lanes: 1, total human DNA; 2, L(17n)C, long arm of chromosome 17; 3, SP3, derivative t(15;17), which is degraded but hybridization is still evident at 11.5 kb; 4, NF13, derivative t(17;22) containing the chromosome 22 centromere; 5, DCR1, derivative t(1;17) containing the chromosome 1 centromere. cDNA HB36.2 was labeled with Klenow fragment, and filters were hybridized and washed under normal conditions. The shadow at 7.8 kb in lane 1 is from an adjacent lane on the original autoradiograph.

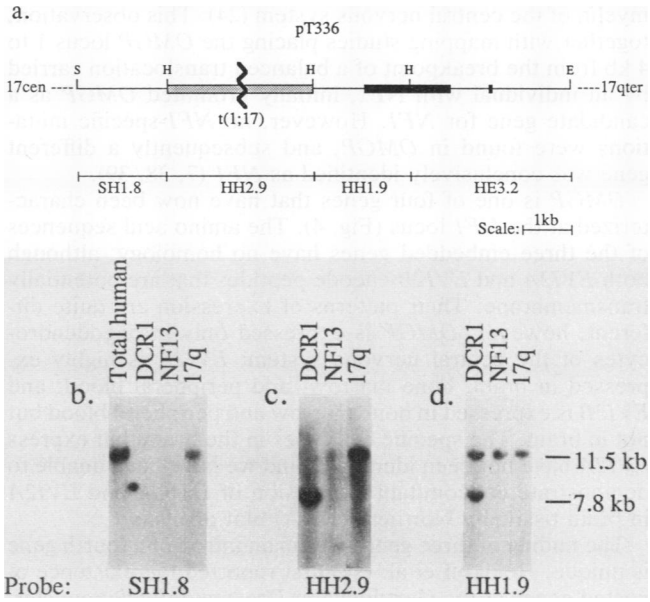

FIG. 3. (a) Restriction map of the 10-kb subcloned *Eco*RI-partial *Sau*3A fragment containing the t(1;17) breakpoint. Sites: E, *Eco*RI; H, *Hin*dIII; S, *Sau*3A. The heavy lines within parts of fragments HH1.9 and HE3.2 designate *OMGP*; this gene is divided into two exons, one 5′ noncoding exon and one 3′ sequence that encodes the entire peptide. The t(1;17) translocation breakpoint maps to the stippled box within fragment HH2.9. (b to d) Southern analysis of *Eco*RI-digested DNA from somatic cell hybrids DCR1, derivative t(1;17); NF13, derivative t(22;17); and L(17n)C, long arm of chromosome 17, probed sequentially with the subcloned restriction fragments shown in panel a. (b) Because probe SH1.8 does not hybridize to the normal 11.5-kb *Eco*RI fragment in either DCR1 or NF13, it lies proximal to both translocation breakpoints. (c) Probe HH2.9 hybridizes to a 7.8-kb derivative band in DCR1 instead of the normal 11.5-kb fragment; thus, it spans the t(1;17) translocation breakpoint. (d) Probe HH1.9 hybridizes to the 7.8-kb derivative band in DCR1. It also hybridizes to an 11.5-kb *Eco*RI fragment in each of the somatic cell hybrids; in DCR1 and NF13, this represents a rodent genomic fragment that contains conserved rodent-human sequence associated with *OMGP*.

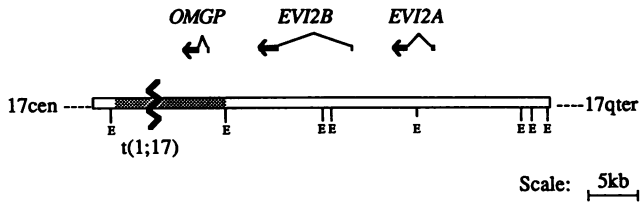

FIG. 4. Map positions of three genes, *OMGP*, *EVI2B*, and *EVI2A*, shown on a genomic map of the t(1;17) breakpoint region. Arrowheads indicate the orientation of transcription, 5′ to 3′. E, *Eco*RI sites. Bold lines denote cDNAs, connected in each case by peaked lines indicating the presence of introns. The 5′ end of *EVI2B* lies 4 kb centromeric of the 3′ end of *EVI2A*; the 5′ end of the longest cDNA clone for *OMGP* maps 5 kb centromeric of the 3′ end of *EVI2B*. Each transcript is approximately 1.6 to 2.1 kb long, and each has two exons, a 5′ noncoding exon and a 3′ coding exon. The intron for *EVI2A* is 2.3 kb; the introns for *EVI2B* and *OMGP* are 8.0 and 0.85 kb, respectively. The three genes together span approximately 28 kb of genomic sequence within one intron of *NF1* (40); the 5′ end of *EVI2A* is 4.1 kb centromeric of an exon of *NF1*. The position of the subclone pT336 is indicated by shading on the genomic map, and the t(1;17) breakpoint is approximate.

myelin of the central nervous system (24). This observation, together with mapping studies placing the *OMGP* locus 1 to 4 kb from the breakpoint of a balanced translocation carried by an individual with NF1, initially promoted *OMGP* as a candidate gene for *NF1*. However, no *NF1*-specific mutations were found in *OMGP*, and subsequently a different gene was conclusively identified as *NF1* (7, 38, 39).

OMGP is one of four genes that have now been characterized at the *NF1* locus (Fig. 4). The amino acid sequences of the three embedded genes have no homology, although both *EVI2A* and *EVI2B* encode peptides that are potentially transmembrane. Their patterns of expression are quite different, however. *OMGP* is expressed only in oligodendrocytes of the central nervous system; *EVI2A* is highly expressed in brain, bone marrow, and peripheral blood; and *EVI2B* is expressed in bone marrow and peripheral blood but not in brain. The specific cell types in the brain that express *EVI2A* have not been identified, and we have been unable to demonstrate concomitant expression of *OMGP* and *EVI2A* in brain tissue by Northern (RNA) blot analysis.

The finding of three genes within an intron of a fourth gene is unique. Henikoff et al. (14) first reported the existence of nested genes at the Gart locus in *Drosophila melanogaster*, which encoded unrelated proteins transcribed from opposite strands; Chen et al. (8) subsequently identified two genes within a large intron of the dunce gene of *D. melanogaster*. Adelman et al. (1) were the first to report transcription of mammalian genes on opposite strands, at the gonadotropin-releasing hormone locus in the rat genome. Miyajima et al. (23) subsequently found that two homologs are transcribed from opposite strands of an exon in human *erbA*. Recently, Levinson et al. (18) offered the first report of transcription of a human gene from an intron of another expressed gene. This single gene, in intron 22 of the factor VIII gene, exists in multiple copies; it contains no introns, and its function is not known. The discoveries with respect to the *NF1* locus extend these observations of human genomic organization even further.

The complexity of the *NF1* region suggests unusual modes of transcriptional regulation within this locus. The chromatin may be in an open conformation, allowing intimate control of individual genes by tissue-specific DNA-binding proteins. We find it interesting that *OMGP* has an octamer sequence in the intron at base −119, ATTTGCAT, that is also found within the enhancer of immunoglobulin heavy-chain genes and within all promoters of genes coding immunoglobulin light-chain variable regions (10). It may be a binding site for human B-cell nuclear factor (IfNF-A), a transcriptional control protein found in lymphoid cells (34). Similar proteins in oligodendrocytes may enhance transcription of *OMGP* from an open chromatin state, while other nuclear factors may enhance expression of the neighboring genes in a tissue-specific manner. The genes embedded in *NF1* may also be regulated at the transcript level, as multiple ATTTA pentamers are present in the 3′ noncoding sequences of all three. This motif is postulated to promote rapid cytoplasmic degradation of mRNA (3). Such a function in the *NF1* region would be highly beneficial if the intron genes continue to be expressed to some degree in tissues in which *NF1* is expressed. In such a case, rapid degradation of mRNA would provide another level of gene regulation. Further tissue-specific regulation may involve RNA-RNA duplex formation to stabilize specific transcripts. The first intron of the N-*myc* locus, for example, is transcribed as an antisense RNA that is sometimes polyadenylated, and it forms RNA-RNA duplexes with the sense RNA from the first exon of n-*myc* (16). The stability of this duplex suggests that modulation of RNA processing is possible via antisense transcription within a genetic locus.

It is logical to suppose that altered expression of the intron genes as a consequence of certain mutations in the *NF1* gene might cause a specific NF1 phenotype. However, we have not been able to identify a distinguishing phenotype among individuals bearing either of two mutations that have been found in the *NF1* intron containing the embedded genes. One of these mutations, a 40-kb deletion, removes a number of *NF1* exons, the 5′ exon of *EVI2B*, and all of *EVI2A*. The other mutation is the t(1;17) balanced translocation identified in a proband and two of her offspring with NF1 (32); the breakpoint lies within 4 kb of the 3′ end of *OMGP*. Although we have not evaluated the expression of *OMGP* in any of the individuals with the t(1;17) translocation, the lack of a distinguishing NF1 phenotype implies that even if *OMGP* expression is altered, it has no clinical relevance in this family. Thus, there is no evidence as yet that altered expression of genes embedded in the *NF1* gene accounts for variability in expression of NF1. Nevertheless, it is intriguing to wonder how transcriptional regulation of these genes might play a role in the NF1 phenotype, and it is now possible to begin testing various hypotheses at the molecular level.

ACKNOWLEDGMENTS

We thank K. Stefansson, D. Mikol, and J. Gulcher for communicating the information that OMgp mapped to the proximal arm of chromosome 17; this allowed us to compare directly our cDNA sequence with the OMgp amino acid sequence without an exhaustive search of data banks. We also thank G. Dewald, J. Gusella, A. Menon, V. Michaels, M. Schmidt, B. Seizinger, and D. Ledbetter for the t(1;17) cell lines; M. Skolnick and the National Neurofibromatosis Foundation for assistance in collection of NF1 patients and blood sampling; M. Robertson for the fluorescent sequencing of portions of the cDNAs on the Applied Biosystems Inc. model 370A sequencer; R. Gesteland for reviewing the manuscript, and R. Wolff and R. Weiss for helpful discussions. R. Foltz edited the manuscript and prepared the figures.

This work was supported in part by a Young Investigator award to D.V. from the National Neurofibromatosis Foundation. R.W. is an Investigator at the Howard Hughes Medical Institute.

REFERENCES

1. **Adelman, J. P., C. Bond, J. Douglass, and E. Herbert.** 1987. Two mammalian genes transcribed from opposite strands of the same DNA locus. Science **235**:1514–1517.
2. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda-gt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
3. **Brawerman, G.** 1989. mRNA decay: finding the right targets. Cell **57**:9–10.
4. **Buchberg, A., H. Bedigan, N. Jenkins, and N. Copeland.** 1990. *Evi-2*, a common integration site involved in murine myeloid leukemogenesis. Mol. Cell. Biol. **10**:4658–4666.
5. **Buchberg, A., H. Bedigan, B. Taylor, E. Brownell, J. Ihle, S. Nagata, N. Jenkins, and N. Copeland.** 1988. Localization of *evi-2* to chromosome 11: linkage to other proto-oncogene and growth factor loci using interspecific backcross mice. Oncogene Res. **2**:149–165.
6. **Cawthon, R., P. O'Connell, A. Buchberg, D. Viskochil, R. Weiss, M. Culver, J. Stevens, N. Jenkins, N. Copeland, and R. White.** 1990. Identification and characterization of transcripts from the von Recklinghausen neurofibromatosis region: the sequence and genomic structure of *EVI2* and mapping of other transcripts. Genomics **7**:555–565.

6a. **Cawthon, K., et al.** Genomics, in press.

7. **Cawthon, R., R. Weiss, G. Xu, D. Viskochil, M. Culver, J. Stevens, M. Robertson, D. Dunn, R. Gesteland, P. O'Connell, and R. White.** 1990. A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. Cell **62**:193–201.
8. **Chen, C., T. Malone, S. Beckendorf, and R. Davis.** 1987. At least two genes reside within a large intron of the *dunce* gene of *Drosophila*. Nature (London) **329**:721–724.
9. **Crowe, F. W., W. T. Schull, and J. F. Neel.** 1956. A clinical, pathological, and genetic study of multiple neurofibromatosis. Charles C Thomas, Publisher, Springfield, Ill.
10. **Falkner, F. G., and H. G. Zachau.** 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature (London) **310**:71–74.
11. **Feinberg, A., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
12. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolating of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
13. **Hashimoto, C., K. Hudson, and K. Anderson.** 1988. The *Toll* gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell **52**:269–279.
14. **Henikoff, S., M. Keene, K. Fechtel, and J. Fristrom.** 1986. Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell **44**:33–42.
15. **Jessel, T. M.** 1988. Adhesion molecules and the hierarchy of neural development. Neuron **1**:3–13.
16. **Krystal, G., B. Armstrong, and J. Battey.** 1990. N-*myc* mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. Mol. Cell. Biol. **10**:4180–4191.
17. **Ledbetter, D. H., D. C. Rich, P. O'Connell, M. Leppert, and J. Carey.** 1989. Precise localization of NF1 to 17q11.2 by balanced translocation. Am. J. Hum. Genet. **44**:20–24.
18. **Levinson, B., S. Kenwrick, D. Lakich, G. Hammonds, Jr., and J. Gitschier.** 1990. A transcribed gene in an intron of the human factor VIII gene. Genomics **7**:1–11.
19. **Lopez, J., D. W. Chung, K. Fujikawa, F. Hagen, E. Davie, and G. Roth.** 1988. The α and β chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino acid sequence. Proc. Natl. Acad. Sci. USA **85**:2135–2139.
20. **Low, M. G., and P. W. Kinkade.** 1985. Phosphatidylinositol is the membrane-anchoring domain of the Thy-1 glycoprotein. Nature (London) **318**:62–64.
21. **Maniatis, T., E. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 93–94. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Menon, A. G., D. Ledbetter, D. C. Rich, B. Seizinger, G. Rouleau, V. Michaels, M. Schmidt, G. Dewald, C. DellaTorre, J. Haines, and J. Gusella.** 1989. Characterization of a translocation within the von Recklinghausen neurofibromatosis region of chromosome 17. Genomics **5**:245–249.
23. **Miyajima, N., R. Horiuchi, Y. Shibuya, S. Fukushige, K. Matsubara, K. Toyoshima, and T. Yamamoto.** 1989. Two *erbA* homologs encoding proteins with different T_3 binding capacities are transcribed from opposite strands of the same genetic locus. Cell **57**:31–39.
24. **Mikol, D., J. Gulcher, and K. Stefansson.** 1990. The oligodendrocyte-myelin glycoprotein belongs to a distinct family of proteins and contains the HNK-1 carbohydrate. J. Cell Biol. **110**:471–480.
25. **Mikol, D., and K. Stefansson.** 1988. A phosphatidylinositol-linked peanut agglutinin-binding glycoprotein in central nervous system myelin and on oligodendrocytes. J. Cell Biol. **106**:1273–1279.
26. **Monaco, A., R. Neve, C. Colletti-Feener, C. Bertelson, D. Kurnit, and L. Kunkel.** 1986. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature (London) **323**:646–650.
27. **O'Connell, P., D. Viskochil, A. Buchberg, J. Fountain, R. Cawthon, M. Culver, J. Stevens, D. C. Rich, D. H. Ledbetter, M. Wallace, J. C. Carey, N. Jenkins, N. Copeland, F. Collins, and R. White.** 1990. The human homologue of murine *evi-2* lies between two translocation breakpoints associated with von Recklinghausen neurofibromatosis. Genomics **7**:547–554.
28. **Orita, M., Y. Suzuki, T. Sekiya, and K. Hayashi.** 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics **5**:874–879.
29. **Reinke, R., D. Krantz, D. Yen, and S. Zipursky.** 1988. Chaoptin, a cell surface glycoprotein required for Drosophila photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. Cell **52**:291–301.
30. **Rommens, J., M. Iannuzzi, B. Kerem, M. Drumm, G. Melmer, M. Dean, R. Rozmahel, J. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J. Riordan, L.-C. Tsui, and F. Collins.** 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science **245**:1059–1065.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Schmidt, M. A., V. Michels, and G. Dewald.** 1987. Cases of neurofibromatosis with rearrangements of chromosome 17 involving band 17q11.2. Am. J. Med. Genet. **28**:771–777.
33. **Shapiro, M., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications. Nucleic Acids Res. **15**:7155–7174.
34. **Singh, H., R. Sen, D. Baltimore, and P. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
35. **Stumpf, D. A., J. Alksne, J. Annegers, S. Brown, P. Conneally, D. Housman, M. Leppert, J. Miller, M. Moss, A. Pileggi, I. Rapin, R. Strohman, L. Swanson, and A. Zimmerman.** 1987. Neurofibromatosis. Arch. Neurol. **45**:575–578.
36. **van Tuinen, P., D. C. Rich, K. Summers, and D. H. Ledbetter.** 1987. Regional mapping panel for human chromosome 17: application to neurofibromatosis type 1. Genomics **1**:374–381.
37. **van Vactor, D., D. Krantz, R. Reinke, and S. L. Zipursky.** 1988. Analysis of mutants in chaoptin, a photoreceptor cell-specific glycoprotein in Drosophila, reveals its role in cellular morphogenesis. Cell **52**:281–290.
38. **Viskochil, D., A. Buchberg, G. Xu, R. Cawthon, J. Stevens, R. Wolff, M. Culver, J. Carey, N. Copeland, N. Jenkins, R. White,**

and P. O'Connell. 1990. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell **62**:187–192.

39. **Wallace, M., D. Marchuk, L. Andersen, R. Letcher, H. Odeh, A. Saulino, J. Fountain, A. Brereton, J. Nicholson, A. Mitchell, B. Brownstein, and F. Collins.** 1990. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science **249**:181–186.
40. **Xu, G., P. O'Connell, D. Viskochil, R. Cawthon, M. Robertson, M. Culver, D. Dunn, J. Stevens, R. Gesteland, R. White, and R. Weiss.** 1990. The neurofibromatosis type 1 gene encodes a protein related to GAP. Cell **62**:599–608.