Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 815–819, February 1991
Genetics

Splice-mediated insertion of an *Alu* sequence inactivates ornithine δ-aminotransferase: A role for *Alu* elements in human mutation

(splicing/inborn error/gyrate atrophy/retroposon)

GRANT A. MITCHELL*, DAMIAN LABUDA*, GISÈLE FONTAINE*, JEAN MARIE SAUDUBRAY†, JEAN PAUL BONNEFONT†, STANISLAS LYONNET†, LAWRENCE C. BRODY‡, GARY STEEL‡, CASSANDRA OBIE‡, AND DAVID VALLE‡§

*Service de Génétique Médicale, Hôpital Ste-Justine, Montreal, ON H3T 1C5, Canada; †Institut National de la Santé et de la Recherche Médicale Unité 12, Hôpital Necker-Enfants Malades, 75743 Paris, Cedex 15 France; and ‡Laboratory of Genetics, Howard Hughes Medical Institute and Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by Daniel Nathans, November 9, 1990 (received for review August 31, 1990)

ABSTRACT **In studies of mutations causing deficiency of ornithine δ-aminotransferase (EC 2.6.1.13), we found an allele whose mature mRNA has a 142-nucleotide insertion at the junction of sequences from exons 3 and 4. The insert derives from an *Alu* element in ornithine δ-aminotransferase intron 3 oriented in the direction opposite to transcription (an "antisense *Alu*"). A guanine → cytosine transversion creates a donor splice site in this *Alu*, activating a cryptic acceptor splice site at its 5′ end and causing splice-mediated insertion of an *Alu* fragment into the mature ornithine-δ-aminotransferase mRNA. We note that the complement of the *Alu* consensus sequence has at least two cryptic acceptor sites and several potential donor sequences and predict that similar mutations will be found in other genes.**

The human genome contains about 700,000 *Alu* sequences comprising about 6% of genomic DNA (1–3). These reiterated elements have been shown to play a role in human genetic disease in two ways: (*i*) by providing sites for homologous recombination that produce deletions and chromosomal rearrangements (4–8) and (*ii*) by *de novo* retroposition of an *Alu* sequence into genes (9). We document a third mechanism of mutation involving *Alu* elements that entails splicing of a portion of an intronic *Alu* element oriented in the direction opposite that of the transcription of the resident gene (an "antisense *Alu* element") into the mature transcript. We discovered this mechanism of mutation during our molecular studies of gyrate atrophy of the choroid and retina (GA), an autosomal recessive chorioretinal degeneration caused by deficiency of ornithine δ-aminotransferase (OAT; EC 2.6.1.13) (10).

MATERIALS AND METHODS

Designation of *Alu* Residues. The numbering convention of residues in *Alu* elements is as described (11). Thus, for an antisense *Alu*, the 3′-most residue with respect to the direction of transcription of the resident gene is position 1 and the most 5′, not including the poly(A) tract, is position 282.

Plasmid Vectors and DNA Probes. pGEM-4 vector (Promega Biotec) was used for cloning. As a cDNA template for radiolabeled probe synthesis and polymerase chain reaction (PCR) synthesis, we used phuOAT6 (12), which contains a human OAT cDNA insert including the entire coding and 3′ nontranslated sequence of human OAT as well as 60 base pairs (bp) of 5′ nontranslated sequence inserted into the *Eco*RI site of pGEM-4.

Cell Lines. Skin fibroblasts from GA patients and normal individuals (controls) were cultured and harvested as described (13).

Enzyme Assay. OAT activity was measured in fibroblast extracts with a radioisotopic assay as described (14).

Western Blot Analysis. Determination of OAT protein in fibroblast extracts was performed with a monospecific polyclonal rabbit anti-OAT antibody that was raised to a synthetic peptide corresponding to amino acid residues 7–25 of mature human OAT.

PCR Amplifications. For OAT cDNA synthesis and PCR amplification 20 μg of whole-cell patient RNA was pretreated with 20 units of DNase I (Pharmacia) in 40 mM Tris·HCl, pH 7.5/6 mM Mg Cl_2/0.75 unit of Inhibit-ace, (5 Prime→3 Prime, Inc.) in a final volume of 15 μl for 20 min at 37°C. The DNase was then inactivated by incubation at 92°C for 5 min and cDNA synthesis performed as follows: RNA (2 μl of the above reaction mixture) was incubated in 20 mM Tris·HCl, pH 8.4 (at 21°C)/50 mM KCl/2.5 mM $MgCl_2$/1 mM dATP/1 mM dCTP/1 mM dGTP/1 mM dTTP (Pharmacia)/bovine serum albumin (0.1 mg/ml)/Moloney murine leukemia virus reverse transcriptase (Pharmacia; 13 units)/Inhibit-ace (0.75 unit) in a final volume of 15 μl. The priming oligonucleotide was 5′-GCTCTAGATGTTAAGGTCAATTTG-3′ at 600 ng. Incubation was at 42°C for 30 min and then at 95°C for 5 min. For PCR amplification, 10 μl of the cDNA synthesis mixture was added to a 40-μl solution containing 20 mM Tris·HCl (pH 8.3 at 21°C), 1.5 mM MgCl, and 600 ng of the 5′ oligonucleotide 5′-TCAGATCTGTGGTTTTTCTA-3′. Thirty cycles of amplification were performed (94°C, 15 sec; 46°C, 15 sec; 72°C, 20 sec) followed by a final elongation of 5 min at 72°C. For the gel shown in Fig. 1 *B* and *C*, 2 μl of the reaction mixture was reamplified using the same primers and amplification conditions, and 10 μl of the reaction mixture was electrophoresed in a 2% agarose gel and transferred to Hybond (Amersham).

For the genomic PCR amplifications used to obtain the DNA whose sequence is shown in Figs. 3 and 4, we amplified 100 ng of genomic DNA using standard reagents (GeneAmp, Perkin–Elmer/Cetus). The priming oligonucleotides for the amplification reaction were DV23 (5′-GGTGCACACAACTACCAT-3′), which corresponds to the sense strand of OAT exon 3, and DV208 (5′-GCCTCCATAGCACCTAACACAG-3′), which corresponds to intronic sequence 63 bp 3′ of the OAT intron 3 antisense *Alu* element. Thirty cycles were

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

Abbreviations: GA, gyrate atrophy of the choroid and retina; OAT, ornithine δ-aminotransferase; PCR, polymerase chain reaction.

§To whom reprint requests should be addressed at: PCTB 802, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205.

performed (94°C, 15 sec; 58°C, 15 sec) followed by a final elongation of 5 min at 72°C.

DNA Sequencing. Dideoxy-nucleotide sequencing of cloned DNA was performed after alkaline denaturation of double-stranded plasmids. For direct sequencing of PCR-amplified fragments, the amplified products were passed twice through Centricon-30 filters (Amicon) and sequenced directly using ^{32}P-end-labeled DV208 as primer as described (15).

RESULTS AND DISCUSSION

The patient (GA022) is a 12-year-old boy born to consanguineous Algerian parents. He presented with typical GA, including chorioretinal degeneration and hyperornithinemia (plasma ornithine, 780 μM; normal, 60 ± 20 μM). OAT activity and immunoreactive material were undetectable in extracts of his cultured skin fibroblasts (data not shown). Northern blot analysis of fibroblast RNA showed an OAT mRNA that was slightly larger than that of normal controls and present at about 5% of the normal amount (Fig. 1*A*).

To determine the molecular basis of this mutation, we performed reverse transcription and cDNA amplification of his fibroblast OAT mRNA. We found that the region of his OAT cDNA corresponding to exons 3 and 4 was increased in length (Fig. 1*B*). On sequence analysis we found a 142-bp insertion precisely at the junction of the sequence encoded by these two exons. An in-frame stop codon (TAG) was present 108 bp into the insert that truncates the OAT peptide by 90%, explaining the lack of OAT activity and antigen.

Because this result suggested aberrant splicing, we wished to determine if GA022 produced any normally sized OAT mRNA. As shown in Fig. 1*A*, none was detectable in Northern blots of his fibroblast poly(A)$^+$ RNA; however, when we amplified a segment of his OAT cDNA spanning exons 3 and 4, we detected a trace amount of normally sized product (<3% of the mutant transcript) (Fig. 1*C*).

Interestingly, the sequence of the insertion corresponded to the complement of the right half of an *Alu* element (*Alu* consensus positions 279–138) (11). We determined by analysis of clones of the OAT structural gene (12) that a single *Alu* element normally is present in the third intron of OAT, 150 bp from its 5′ extremity (Fig. 2). Amplification and sequencing of this region in the patient and a control revealed identical sequences except at the position immediately 3′ to the block of sequence inserted in the GA022 transcript (position 137 of the *Alu* consensus) where the patient was homozygous for a C → G transversion (Fig. 3*A*). His parents and one of two clinically unaffected siblings were heterozygous for this mutation (Fig. 4). This nucleotide substitution creates a donor splice site (AG/GTAATT) (Fig. 3*B*) that activates an upstream cryptic acceptor site. The resulting "exon" flanked by these sites accounts for the 142-bp insertion and corresponds to bases 279–138 of the *Alu* consensus. We designate this allele as OAT G67ins to indicate an insertion at the glycine-67 codon. We know of three similar donor site-creating mutations in humans, all of which occur in intron 2 of β-globin (18–20).

To determine if the involvement of an *Alu* element in our patient's mutation was coincidental or if the sequence of *Alu* elements actually predisposed this event, we examined the *Alu* consensus sequence for similarity to the sequences necessary for splicing. As predicted by the consequences of the GA022 mutation, the complement of the poly(A) tail and the next three internal bases in an antisense *Alu* provides two essential components of a splice acceptor site, a pyrimidine tract followed by an AG dinucleotide (Y_nGAG, where Y is a pyrimidine) (Fig. 3*B*). This sequence, which results in splicing 5′ of base 278 in the *Alu* consensus, differs from the consensus acceptor splice site sequence ($Y_nNCAG/$) only by the presence of a guanine in the −3 position. In splicing assays, however, $Y_nGAG/$ has been shown to be a functional, albeit inefficient, acceptor splice site (21, 22). The length of the polypyrimidine tract is also important and must be >10 bases to function effectively in splicing (21–23). A recent survey of the sequence of 193 *Alu* elements found that 80% had poly(A) tracts of this size (24). Thus, the majority of antisense *Alu* elements will have a polypyrimidine tract of sufficient length for an acceptor splice site. The final component of an acceptor splice sequence is the lariat branch point that, in this case, is provided by upstream non-*Alu* genomic sequence. The requirements for the position and sequence (YNYTRAY, where N is any base and R is a purine) of branch point sites are lax in higher eukaryotes (21, 22). Thus, we expect that the 5′ end of a majority of antisense *Alu* sequences will contain all the components of an acceptor splice site.

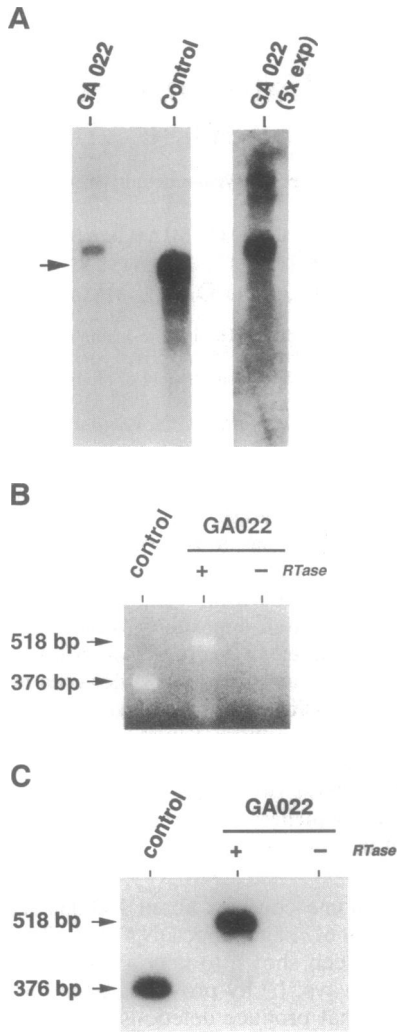

FIG. 1. Degree of abnormal splicing of OAT mRNA in GA022. (*A*) Northern blot analysis of fibroblast poly(A)$^+$ RNA from patient GA022 (3 μg) and control (1 μg) probed with human OAT cDNA. A 5-fold longer exposure of the GA022 lane is shown on the right. The arrow indicates the position of the normal 2.2-kilobase OAT mRNA. OAT cDNA amplification products visualized by staining with ethidium bromide (*B*) or hybridization to radiolabeled human OAT cDNA (12) (*C*). Both *B* and *C* show the products of amplification of a control OAT cDNA and of OAT cDNA synthesized from GA022 fibroblast RNA. No reverse transcriptase (−RTase) was added at the cDNA synthesis step. The PCR primers were complementary to the 5′ end of OAT exon 3 and the middle of exon 4 to yield a 376-bp amplified fragment from a normal OAT cDNA and a 518-bp fragment from cDNA with the 142-bp insert between exon 3 and 4.

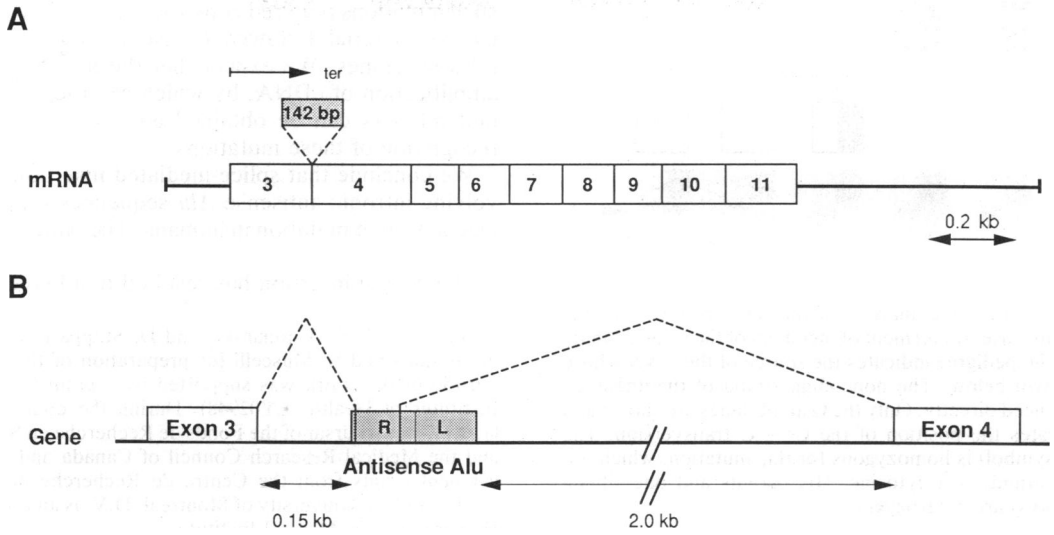

FIG. 2. Diagram of OAT cDNA and gene fragments in patient GA022. (*A*) The OAT cDNA from the patient contains a 142-bp insertion precisely at the junction of the sequence encoded by exons 3 and 4. The sequence of the insertion corresponds to the complement of residues 279–138 of the *Alu* consensus (11). (*B*) OAT intron 3 contains a single antisense *Alu* element. The dashed lines indicate the abnormal splicing pattern observed in GA022.

There are two additional potential acceptor sites in the reverse complement of the *Alu* consensus, at positions 206 and 118. In the former, the polypyrimidine tract is short, interrupted by a guanine and has a guanine in the −2 position (Y_5GY_5GG/GT). Thus, this site is unlikely to be functional unless variant from the consensus. Interestingly, in the *Alu* of OAT intron 3, this sequence is more favorable ($Y_5GY_5\underline{A}G/G$) but there was no amplified fragment of the length expected for splicing at this position in GA022 (Fig. 1 *B* and *C*). Fearon *et al.* (25) suggested that a somatic mutation converting the −2 G → A in an antisense intronic *Alu* in the DCC (deleted in colorectal carcinomas) gene may have activated this gene in a colorectal cancer of one patient. The potential acceptor splice site at position 118 (Y_5GTAY_5AG/TA) also has a relatively short polypyrimidine tract interrupted by purines but was utilized for an *Alu*-containing transcript of the c-*rel* oncogene (see below) (26).

The experience with the β-globin mutants with additional donor sites, all of which exploit the same cryptic acceptor (18–20), shows that the presence of a cryptic acceptor site can be revealed by a variety of downstream donors usually within 50–300 bp of the acceptor. There are at least seven sites in the reverse complement of the *Alu* consensus sequence that differ from the donor consensus (AG/GTRAG) by three or less nucleotides (positions 169, 157, 137, 64, 37, 22, and 20). Given that the reverse complement of the consensus *Alu* sequence contains at least two cryptic acceptor sites and several potential donor splice sites, we anticipate that similar

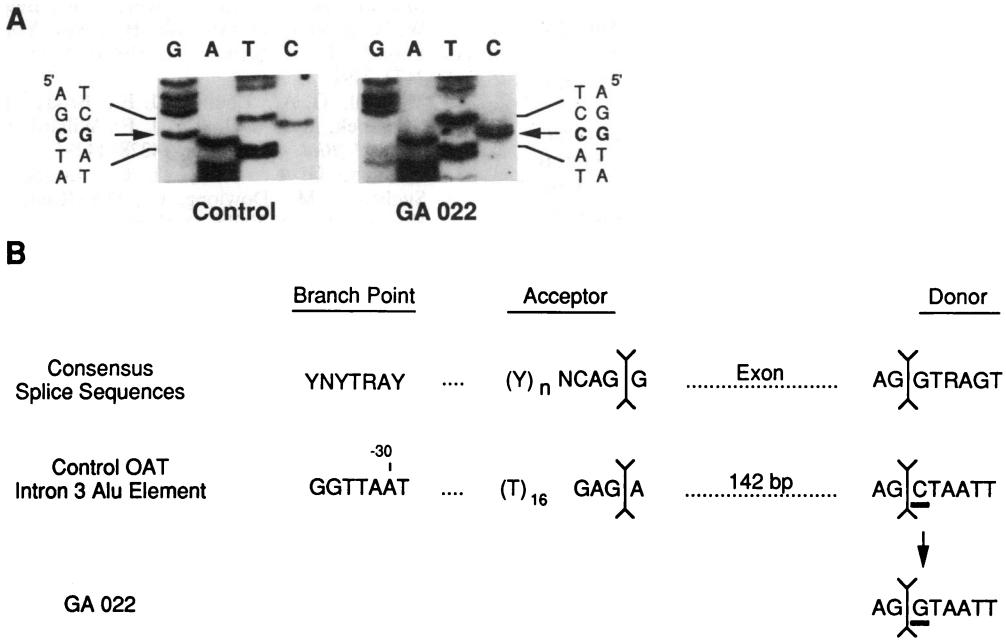

FIG. 3. Donor site-creating mutation in the intronic antisense *Alu* of GA022. (*A*) Direct sequence analysis of amplified genomic DNA from the OAT intron 3 antisense *Alu* of a normal control and of patient GA022. The sequence is of the noncoding strand. The nucleotide sequences of both strands are printed flanking the sequencing lanes and the mutated base is indicated in bold type. (*B*) Sequence comparison between the mammalian consensus for lariat branch points (16), acceptor (17) and donor sites (17), the corresponding regions of the normal OAT intron 3 antisense *Alu* sequence, and the mutation found in GA022.

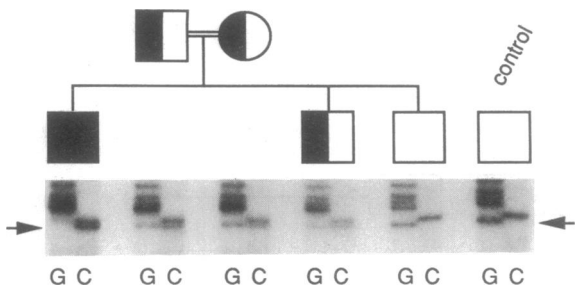

FIG. 4. Direct sequence analysis of the relevant region of the OAT intron 3 antisense *Alu* element of members of the GA022 family and a control. The pedigree indicates the source of the DNA whose sequence is shown below. The noncoding strand of the amplified DNA was sequenced directly. Only the G and C lanes are shown and the arrow indicates the position of the C → G transversion. The proband (solid symbol) is homozygous for this mutation, which, on the noncoding strand, is a cytosine. His parents and one sibling (half-solid symbols) are heterozygous.

splice-mediated insertions involving *Alu* will disrupt the transcripts of other genes. This possibility is magnified by the large number of *Alu* elements in the genome. In the 40-kilobase transcribed portion of the *HPRT* locus, for example, there are 34 *Alu* elements, 23 (68%) of which are in the antisense orientation (27). Moreover, *Alu* elements may be preferentially located in R bands, the region of the genome enriched for transcribed sequences (3, 28).

Despite these considerations, we have found only three other examples in which a portion of an *Alu* element appears to be spliced into the coding region of a mature mRNA. The clearest example is the normal human decay-accelerating factor (*DAF*) gene from which two transcripts are produced, differing only by the presence in one of a 118-nucleotide antisense *Alu* cassette inserted precisely between sequence encoded by two exons (29, 30). The 5′ boundary of this insert begins at *Alu* consensus position 275, indicating that the splicing apparatus uses the 5′ polypyrimidine tract of an antisense intronic *Alu* as in the OAT gene of GA022. The donor site used by the *DAF* antisense *Alu* exon is at position 157. The *DAF* gene contains an intronic antisense *Alu* that provides the insert (30). In contrast to the OAT G67ins allele, the *DAF* antisense *Alu* insert does not terminate translation and generates an alternative form of the protein (29).

The second example is the human protooncogene, c-*rel*. Brownell *et al.* (26) characterized two c-*rel* clones isolated from a Doudi Burkett lymphoma cell line cDNA library. One contained a 96-nucleotide antisense *Alu* cassette (*Alu* consensus positions 117–23) inserted between sequence provided by two exons (26). Although the insert maintains the open reading frame, the correspondingly altered protein was not detected in immunoprecipitation experiments. Thus, the biologic significance of this alternative form of c-*rel* is uncertain. These investigators recognized the potential for intronic antisense *Alu* elements to undergo productive splicing (26).

The third example comes from analysis of a single human cDNA clone for the fifth component of complement (C5) isolated from a normal liver cDNA library (31). In this clone an antisense *Alu* cassette from virtually the identical segment of the *Alu* consensus as that in the OAT G67ins allele replaces the expected 3′ region of C5 as predicted by comparison with the murine C5 cDNA. The relative frequency of this apparently abnormal mRNA species and documentation of an antisense *Alu* element in the appropriate intron have not been reported.

The paucity of other examples of splice-mediated *Alu* insertions may be due, in part, to the technical requirements required for their identification. Until recently, detection of such mutations required construction of cDNA libraries from patient material followed by isolation and sequencing of relevant clones. We expect that the increased use of PCR amplification of cDNA, by which even low-abundance mutant mRNAs can be obtained and analyzed, will enhance recognition of these mutations.

We conclude that splice-mediated insertion mutations involving intronic antisense *Alu* sequences may be a general mechanism of mutation in humans. The patient reported here is, to our knowledge, the first example in which splice-mediated *Alu* insertion has resulted in a hereditary disease.

We thank S. E. Antonarakis and D. Stoppa-Lyonnet for helpful discussions and S. Muscelli for preparation of this manuscript. A portion of this work was supported by a grant from the National Institutes of Health (EY02948). During the course of this work, G.A.M. was a bursar of the Fonds de Recherche en Sante du Quebec and the Medical Research Council of Canada and received establishment grants from the Centre de Recherche de l'Hôpital Ste-Justine and the University of Montreal. D.V. is an investigator in the Howard Hughes Medical Institute.

1. Rinehart, F. P., Ritch, T. G., Deininger, P. L. & Schmid, C. W. (1981) *Biochemistry* **20**, 3003–3010.
2. Hwu, H. R., Roberts, J. W., Davidson, E. H. & Britten, R. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3875–3879.
3. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M. & Goad, W. B. (1989) *Genomics* **4**, 273–289.
4. Lehrman, M. A., Schneider, W. J., Südhof, T. C., Brown, M. S., Goldstein, J. L. & Russell, D. W. (1985) *Science* **227**, 140–146.
5. Myerowitz, R. & Hogikyan, N. D. (1987) *J. Biol. Chem.* **262**, 15396–15399.
6. Nicholls, R. D., Nischel-Ghodsian, N. & Higgs, D. R. (1987) *Cell* **49**, 369–378.
7. Markert, M. L., Hutton, J. J., Wiginton, D. A., States, J. C. & Kaufman, R. E. (1988) *J. Clin. Invest.* **81**, 1323–1327.
8. Rouyer, F., Simmler, M.-C., Page, D. C. & Weissenbach, J. (1987) *Cell* **51**, 417–425.
9. Vidaud, M., Vidaud, D., Siguret, V., Lavergne, J. M. & Goossens, M. (1989) *Am. J. Hum. Genet.* **44**, A226.
10. Valle, D. & Simell, O. (1989) in *The Metabolic Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. W. & Valle, D. (McGraw-Hill, New York), pp. 599–627.
11. Labuda, D. & Striker, G. (1989) *Nucleic Acids Res.* **17**, 2477–2491.
12. Mitchell, G. A., Looney, J. E., Brody, L. C., Steel, G., Suchanek, M., Engelhardt, J. F., Willard, H. F. & Valle, D. (1988) *J. Biol. Chem.* **263**, 14228–14295.
13. Mitchell, G. A., Brody, L. C., Looney, J., Steel, G., Suchanek, M., Dowlong, C., Der Kaloustian, V., Kaiser-Kupfer, M. I. & Valle, D. (1988) *J. Clin. Invest.* **81**, 630–633.
14. Valle, D., Kaiser-Kupfer, M. I. & del Valle, L. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5159–5161.
15. Higuchi, R., von Beroldingern, C. H., Sensabaugh, G. F. & Erlich, H. A. (1988) *Nature (London)* **332**, 543–546.
16. Harris, N. L. & Senapathy, P. (1990) *Nucleic Acids Res.* **18**, 3015–3019.
17. Ohshima, Y. & Gotoh, Y. (1987) *J. Mol. Biol.* **195**, 247–259.
18. Treisman, R., Orkin, S. H. & Maniatis, T. (1983) *Nature (London)* **302**, 591–596.
19. Cheng, T.-C., Orkin, S. H., Antonarakis, S. E., Potter, M. J., Sexton, J. P., Markham, A. F., Giardina, P. J. V., Li, A. & Kazazian, H. H., Jr. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2821–2825.
20. Dobkin, C. & Bank, A. (1985) *J. Biol. Chem.* **260**, 16332–16337.
21. Reed, R. (1989) *Genes Dev.* **3**, 2113–2123.
22. Smith, C. W. J., Porro, E. B., Patton, J. G. & Nadal-Ginard, B. (1989) *Nature (London)* **342**, 243–247.
23. Wieringa, B., Hofer, E. & Weissman, C. (1984) *Cell* **37**, 915–925.
24. Economou, E. P., Bergen, A. W., Warren, A. C. & Antonarakis, S. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2951–2954.
25. 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. & Vogelstein, B. (1990) *Science* **247**, 49–56.
26. Brownell, E., Mittereder, N. & Rice, N. R. (1989) *Oncogenes* **4**, 935–942.
27. Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C. T. & Ansorge, W. (1990) *Genomics* **6**, 593–608.
28. Korenberg, J. R. & Rykowski, M. C. (1988) *Cell* **53**, 391–400.
29. Caras, I. W., Davitz, M. A., Rhee, L., Weddell, G., Martin, D. W., Jr., & Nussenzweig, V. (1987) *Nature (London)* **325**, 545–549.
30. Post, T.W., Arce, M. A., Liszewski, M. K., Thompson, E. S., Atkinson, J. P. & Lublin, D. M. (1990) *J. Immunol.* **144**, 740–744.
31. Lundwall, A. B., Wetsel, R. A., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R. & Tack, B. F. (1985) *J. Biol. Chem.* **260**, 2108–2112.