WHAT'S DIFFERENT ABOUT CHROMOSOME 16?

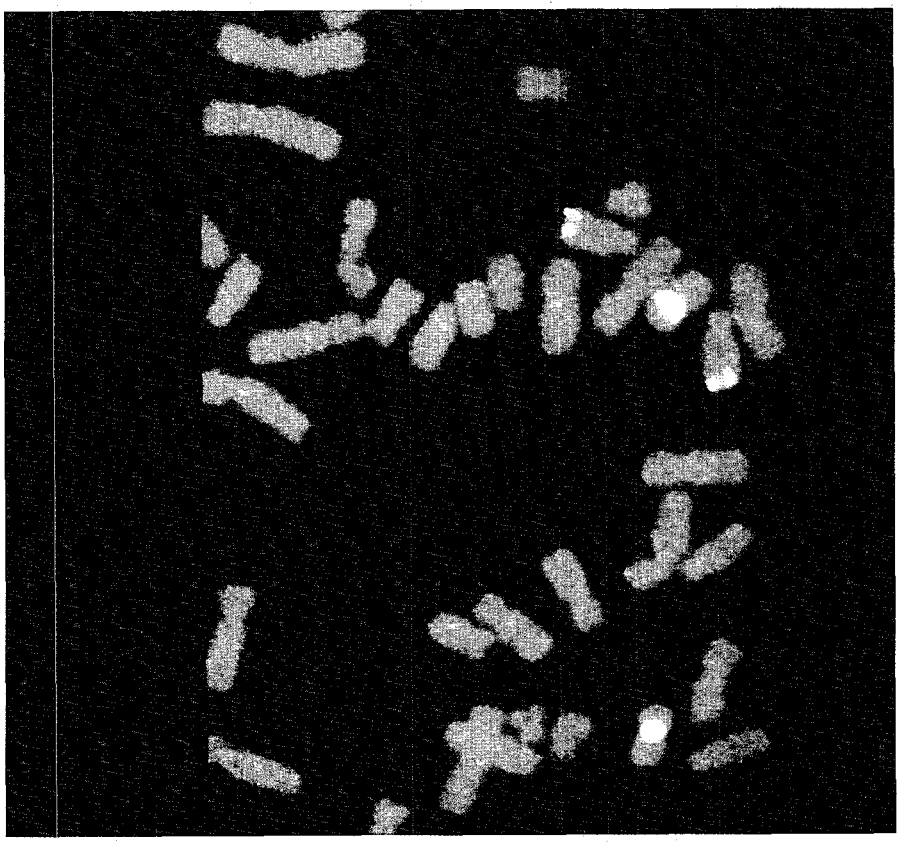
Raymond L. Stallings and Norman A. Doggett

Human chromosome 16 is different from most other human chromosomes in that it contains a larger-than-average fraction of repetitive sequences. As we will describe below, during the course of constructing a contig map for chromosome 16, we discovered several new low-abundance repetitive sequences that are present only on chromosome 16 and that may be implicated in the etiology of certain genetic diseases.

Repetitive sequences are frequently referred to as junk DNA because it has been difficult to determine whether these sequences have any role in the organization and functioning of eukaryotic genomes. Repetitive sequences are also referred to as selfish DNA because they represent such a large fraction of these genomes. For example, the fraction of repetitive DNA in the human genome is estimated to be between 25 and 35 percent. The fact that some classes of repetitive sequences, such as the alpha satellite DNA found in primates, have mutated rapidly over evolutionary time scales lends credence to the notion that at least some repetitive sequences represent mere clutter and play no functional role.

In contrast, work led by Bob Moyzis here at the Laboratory has shown that the repeat sequences that make up the functional centromeres and telomeres of human chromosomes have been highly conserved throughout evolution and serve very important functions. The centromeric repeat sequences are essential to the proper replication and parceling out of chromosomes to daughter cells during cell division. The telomeric tandem repeats maintain the ends of the chromosomes during replication. Some simple microsatellite repeat sequences, such as $(GT)_n$, are so widely distributed throughout all eukaryotic genomes that it is difficult to believe they don't have some functional significance. (See "Various Classes of Human Repetitive DNA Sequences.")

Regardless of whether different classes of repetitive sequences have specific functions or, as Orgel and Crick suggest, are "the ultimate parasite," many of these sequences are of medical interest. Recent findings demonstrate that some human repetitive sequences undergo rapid mutations or facilitate chromosomal rearrangements and that both types of changes can lead to human genetic diseases. The fragile site on the human X chromosome is an example. Like other fragile sites, the fragile X site is so named because the X chromosome at that site appears to have a nonstaining gap or break under certain experimental conditions. The fragile X site is located on the X chromosome within the region Xq27.3. Fragile X is inherited in a Mendelian fashion. Recent cloning of the fragile X region and subsequent analysis showed, first, that it contains the trinucleotide tandem repeat sequence $(CCG)_n$, and second, that the tandem repeat can undergo significant amplification (that is, n can increase significantly) between one generation and the next. Moreover, amplification of $(CCG)_n$ seems to be the cause of a very common form of mental retardation that has long been associated with the presence of the fragile X site.



Photograph courtesy of David Ward, Yale University School of Medicine

Shortly after the dramatic discovery of the fragile X site came reports that amplification of another trinucleotide repeat on chromosome X, $(CTG)_n$, is responsible for spinal and bulbar muscular atrophy and that amplification of the $(CTG)_n$ repeat on chromosome 19 is responsible for myotonic dystrophy. Evidently, when those tandem repeats undergo spontaneous amplification within germ-line cells, they disrupt the functioning of a gene or of the regulatory region for a gene in an offspring derived from a gamete containing the amplified sequence. The increasing level of amplification from one generation to the next is accompanied by an increase in the symptoms of the disease, a genetic process that has been termed anticipation. For example, amplification of $(CTG)_n$ that occurs in one generation may cause cataracts, and its further amplification in a subsequent generation will cause full-blown myotonic dystrophy.

Repetitive sequences other than trinucleotide tandem repeats have also been implicated in genetic disease. For example, it was recently discovered that the insertion of a truncated L1 sequence in the gene for blood-clotting factor VIII was responsible for a spontaneous case of hemophilia A. Similarly, de novo insertion of Alu repeats into the cholinesterase gene led to inactivation of the gene, and a comparable insertion in the NF1 gene caused the common dominant disorder known as neurofibromatosis type 1.

Our group and a group at Leiden University have recently determined that there is extensive sequence homology between two widely separated regions of chromosome 16, band 16p13 on its short arm and band 16q22 on its long arm. The homology could explain why rearrangements occur between those chromosomal regions in acute nonlymphocytic leukemia (ANLL). The sequence homology between the two bands is due to the presence of low-abundance repetitive sequences at multiple loci in bands 16p13, 16p12, 16p11, and 16q22.

We discovered those repetitive sequences on chromosome 16 in the course of developing the contig map of chromosome 16. As we grouped pairs of overlapping clones into contigs, we encountered an anomaly—a set of 78 clones, all of which seemed to overlap other clones in the set. Thus the clones appeared to form a single contig, or island of overlapping clones, much larger than the average contig, which contained only four or five clones. However, when we tried to position the clones to form a

single contig, we found that they could not be placed in a linear order, but rather the contig branched in many directions and included many clones that seemed to be piled on top of one another. Our inability to construct a linear contig indicated that many false overlaps had been deduced from the fingerprint data because of the presence of some unknown repetitive sequence in the clones.

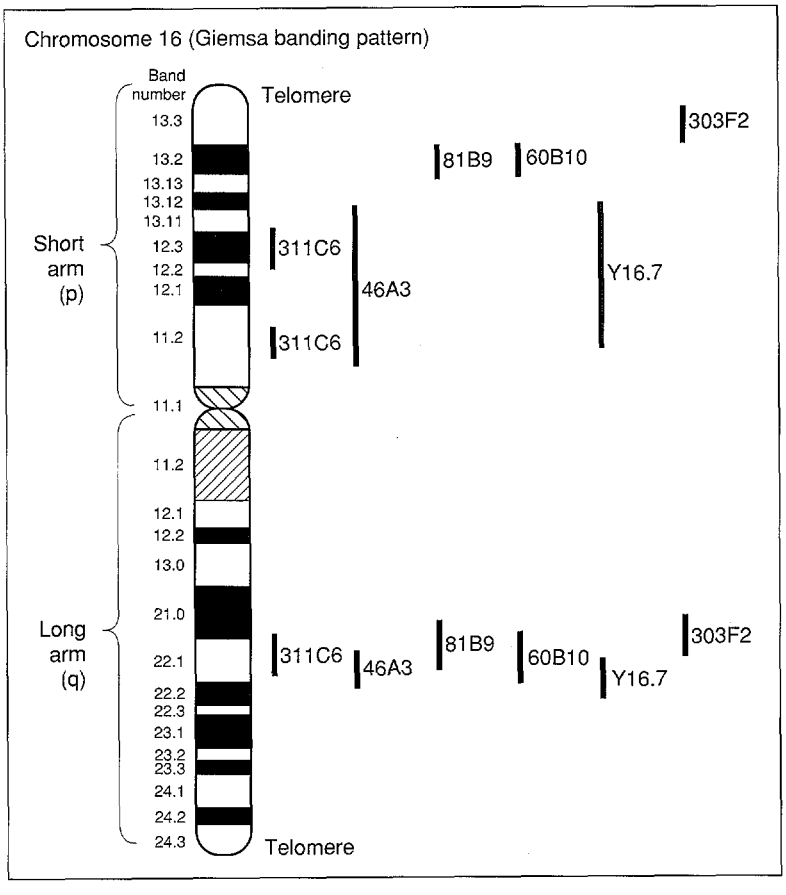
We went on to analyze the 78 clones using a variety of techniques. Fluorescence in-situ hybridization of five of the clones revealed that each one hybridized to as many as three locations on chromosome 16, and those locations occurred in four bands of chromosome 16: 16p13, 16p12, 16p11, and 16q22 (see Figure 1). The hybridization results and further analysis indicated that the four bands contain low-abundance repetitive sequences that are found only on chromosome 16. Characterization of one of those sequences revealed that it was a minisatellite-type sequence that did not possess homology to any of the known minisatellites. The consensus repeat unit of the sequence is

TCCT X TCCT CTTCCACCCT CAGTGGATGA TAATCTGAAG GA,

where X is any sequence containing between 2 and 9 nucleotides. The results of in-situ hybridization of this consensus repeat to chromosome 16 is shown in the opening pages of "The Mapping of Chromosome 16." High-stringency hybridization of the consensus sequence to Southern blots containing DNA from humans, the rhesus monkey, rat, mouse, dog, cow, rabbit, chicken, and yeast produced positive hybridization signals only from human and monkey DNA. Apparently, the sequence is present only in primates and therefore could be relatively recent in origin.

We estimate that the low-abundance repetitive sequences specific to chromosome 16 together occupy between 2 million and 6 million base pairs of the chromosome. Moreover, those sequences appear to overlap the breakpoint regions involved in the rearrangements of chromosome 16 commonly observed to accompany the particular subtype of acute nonlymphocytic leukemia referred to as ANLL subtype M4. Those chromosomal rearrangements include an inversion around the centromere between breakpoints in bands 16p13 and 16q22, a translocation between the homologs of chromosome 16 involving bands 16p13 and 16q22, and deletions in 16q22. Recombination between the low-abundance repetitive sequences in bands 16p13 and 16q22 could lead to the observed inversions and translocations. Therefore it is not unreasonable to consider that the repetitive sequences may be causally related to the inversions and translocations that occur in the chromosomes of leukemia cells. The isolation of repetitive sequences common to bands 16p13 and 16q22 is facilitating the isolation of the breakpoint regions and any gene(s) that may reside at those breakpoints.

We have discovered not only low-abundance repetitive sequences in the euchromatic arms of chromosome 16 but also novel repetitive sequences at the pericentromeric regions (regions near the centromere) of human chromosome 16 and at locations on other human chromosomes. The latter repetitive sequences are distinct from



any of the five satellite sequences (α, β, I, II, III) that are commonly found in the centromeric region of all human chromosomes. Previous work at the Laboratory had revealed that a large block of chromosome-specific, satellite-II-variant DNA occurs at the pericentromeric region of the long arm of chromosome 16 (at 16q11.1) and that a chromosome-specific α-satellite variant occurs in the centromeric region of chromosome 16. We have identified a new repetitive sequence that appears as a large block on the pericentromeric region of the short arm of chromosome 16 (at 16p11.1) and is also found in the telomeric regions of chromosome 14 (Figure 2). This block of repetitive sequence at 16p11.1 composes almost 2 percent (or 2 million base pairs) of chromosome 16. In addition, we have found another repetitive sequence that maps to 16p11.1 and 15q11.1.

The region 16p11.1 appears to be quite rich in novel repetitive DNA sequences that map to a few other human chromosomes. Another minisatellite, MS29, maps to 16p11.1 and to chromosome 6. The MS29 locus at 16p11.1 is polymorphic in that it is absent from some human chromosomes 16. Several other unusual chromosome-16 variants have also been reported that appear to have extra material added in band 16p11.1 The extra material is C-band negative; that is, it does not darken when stained by the special techniques that usually darken only the centromeric regions. Also, the extra material is not composed of α-satellite DNA.

With the extensive amount of repetitive DNA found at 16p11.1, one might expect to find occasional amplification of this region. The amplification of this DNA does not appear to have any phenotypic effect, although the possibility of increased risk of aneuploidy cannot be ruled out. Also, the possibility that further amplification in successive generations could have detrimental effects cannot be ruled out. ■

Further Reading

C. R. Bryke, W. R. Breg, R. P. Venkateswara, and T. L. Yang-Feng. 1990. Duplication of euchromatin without phenotypic effects: A variant of chromosome 16. *American Journal of Medical Genetics* 36:43–44.

J. Buxton, P. Shelbourne, J. Davies, C. Jones, T. Van Tongeren, C. Aslanidis, P. de Jong, G. Jansen, M. Anvret, B. Riley, R. Williamson, and K. Johnson. 1992. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 355:547–548.

J. G. Dauwerse, E. A. Jumelet, J. W. Wessels, J. J. Saris, A. Hagemeijer, G. C. Beverstock, G. J. B. Van Ommen, and M. H. Breuning. 1992. Extensive cross-homology between the long and short arm of chromosome 16 may explain leukemic inversions and translocations. *Blood* 79:1299–1304.

B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, and H. H. Kazazian. 1991. Isolation of an active human transposable element. *Science* 254:1805–1810.

D. L. Grady, R. L. Ratliff, D. L. Robinson, E. C. McCanlies, J. Meyne, and R. K. Moyzis. 1992. Highly conserved repetitive DNA sequences are present at human centromeres. *Proceedings of the National Academy of Sciences of the United States of America* 89:1695–1699.

G. M. Greig, S. B. England, H. M. Bedford, and H. F. Willard. 1989. Chromosome-specific alpha satellite DNA from the centromere of human chromosome 16. *American Journal of Human Genetics* 45:862–872.

E. J. Kremer, M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland, and R. I. Richards. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$. *Science* 252:1711–1714.

A. R. La Spada, E. M. Wilson, D. B. Lubahn, A. E. Harding, and K. H. Fischbeck 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77–79.

M. M. LeBeau, R. A. Larson, M. A. Bitter, J. W. Vardiman, H. M. Golomb, and J. D. Rowley. 1983. Association of an inversion of chromosome 16 with abnormal marrow eosinophils in acute myelomonocytic leukemia: A unique cytogenic-clinicopathologic association. *New England Journal of Medicine* 309:630–636.

R. K. Moyzis, K. L. Albright, M. F. Bartholdi, L. S. Cram, L. L. Deaven, C. E. Hildebrand, N. E. Joste, J. L. Longmire, J. Meyne, and T. Schwarzacher-Robinson. 1987. Human chromosome-specific repetitive DNA sequences: Novel markers for genetic analysis. *Chromosoma* 95:375–386.

K. Muratani, T. Hada, Y. Yamamoto, T. Kaneko, Y. Shigeto, T. Ohue, J. Furuyama, and K. Higashino. 1991. Inactivation of the cholinesterase gene by Alu insertion: Possible mechanism for human gene transposition. *Proceedings of the National Academy of Sciences of the United States of America* 88:11315–11319.

S. Ohno. 1972. So much "junk" DNA in our genomes. In *Evolution of Genetic Systems*, edited by H. H. Smith. New York: Gordon and Breach.

L. E. Orgel and F. H. C. Crick. 1980. Selfish DNA: The ultimate parasite. *Nature* 284:604–607.

R. L. Stallings, A. F. Ford, D. Nelson, D. C. Torney, C. E. Hildebrand, R. K. Moyzis. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* 10:807–815.

R. L. Stallings, N. A. Doggett, K. Okumura, and D. C. Ward. 1992. Chromosome 16 specific repetitive DNA sequences that map to chromosomal regions known to undergo breakage/rearrangement in leukemia cells. *Genomics* 13:332–338.

M. R. Wallace, L. B. Anderson, A. M. Saulino, P. E. Gregory, T. W. Glover, and F. S. Collins. 1991. A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353:864–866.

Z. Wong, N. J. Royle, and A. J. Jeffreys. 1990. A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics* 7:222–234.