Table 2. Results of Hybridization of Inter-Alu PCR Products to YACs

Number of cosmid contigs overlapped by a single YAC	0	1	2	3	4	5	6	>6
Number of YACs	105	133	84	41	24	12	6	6

Each YAC clone has an average insert size of 215,000 base pairs, so we expect many of them will bridge the gaps between two or more cosmid contigs. To find those contigs, we adopted a hybridization strategy which is less time-consuming than the STS approach. YAC clones are not good hybridization probes for detecting overlaps between human DNA inserts because the yeast DNA in those clones contains sequences that are homologous to human DNA and can produce false-positive hybridization signals. We need, instead, to generate DNA probes from each YAC clone that we know are derived from the human DNA insert in that clone. An efficient procedure, known as inter-Alu PCR, is outlined in Figure 8(a). The procedure uses the polymerase chain reaction to amplify DNA that lies between Alu sequences within the human DNA insert of the YAC. Alu sequences are found in human DNA but not in yeast DNA. Therefore, if primers from the ends of the Alu sequence are used in the polymerase chain reaction, the reaction will amplify regions of the human DNA insert only. Figure 8(b) shows a gel containing the amplified products derived by applying inter-Alu PCR to each of eight YAC clones. Each lane of the gel contains PCR products from one YAC clone. The average number of PCR products was about six.

The inter-Alu PCR products from each YAC clone were then radiolabeled with ^{32}P nucleotides and annealed with $Cot1$ DNA, a process that covers any $Cot1$ repetitive sequences that might be present. The PCR products were then

ready to be used as a hybridization probe to screen the 4000 fingerprinted cosmid clones. To facilitate screening, the 4000 fingerprinted clones were fixed on membranes in a high-density, gridded array. Each membrane accommodates 1536

clones, so the entire set of fingerprinted cosmid clones was arrayed on three membranes (see Figure 9). Cosmids that yield positive hybridization signals must contain a DNA sequence present in the YAC clone from which the hybridization
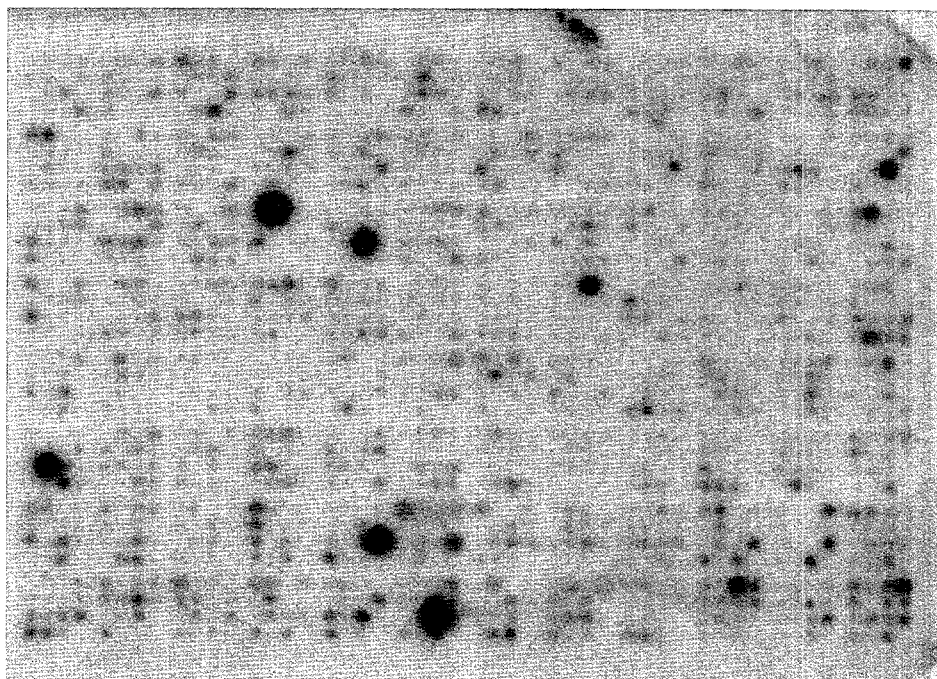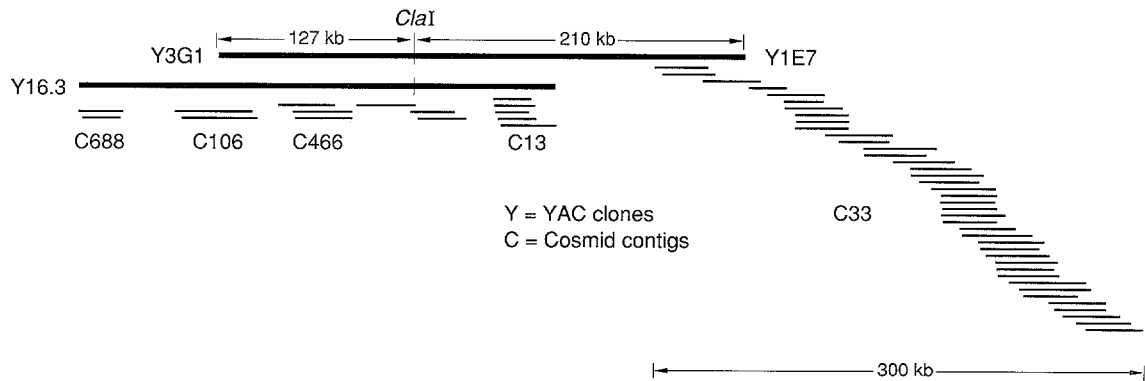

Figure 9. Hybridization of Inter-Alu PCR Products to Cosmid Clones
The figure shows an autoradiogram of a membrane containing 1536 cosmid clones. The clones from the wells of 16 different microtiter plates (8 rows and 12 columns for 96 clones per plate) were stamped onto a membrane the size of the microtiter plate by a high-precision robotic device. The resulting gridded array of clones provides a convenient tool for hybridization experiments. The darker and larger dots are the result of hybridization of YAC inter-Alu PCR products to specific cosmid clones. Here the PCR products from YAC clone Y3A12 hybridized to cosmid clones from 2 different contigs. The results suggest that the YAC clone overlaps those cosmid contigs. The automated robotic gridding device that makes the hybridization grids was designed and built by Pat Medvick, Tony Beugelsdijk, and Bob Hollen in the MEE-3 group. A photograph of the device appears on the opening pages of "DNA Libraries" and is discussed in "Libraries from Flow-sorted Chromosomes."

(a) Subcloning of YAC 16.3
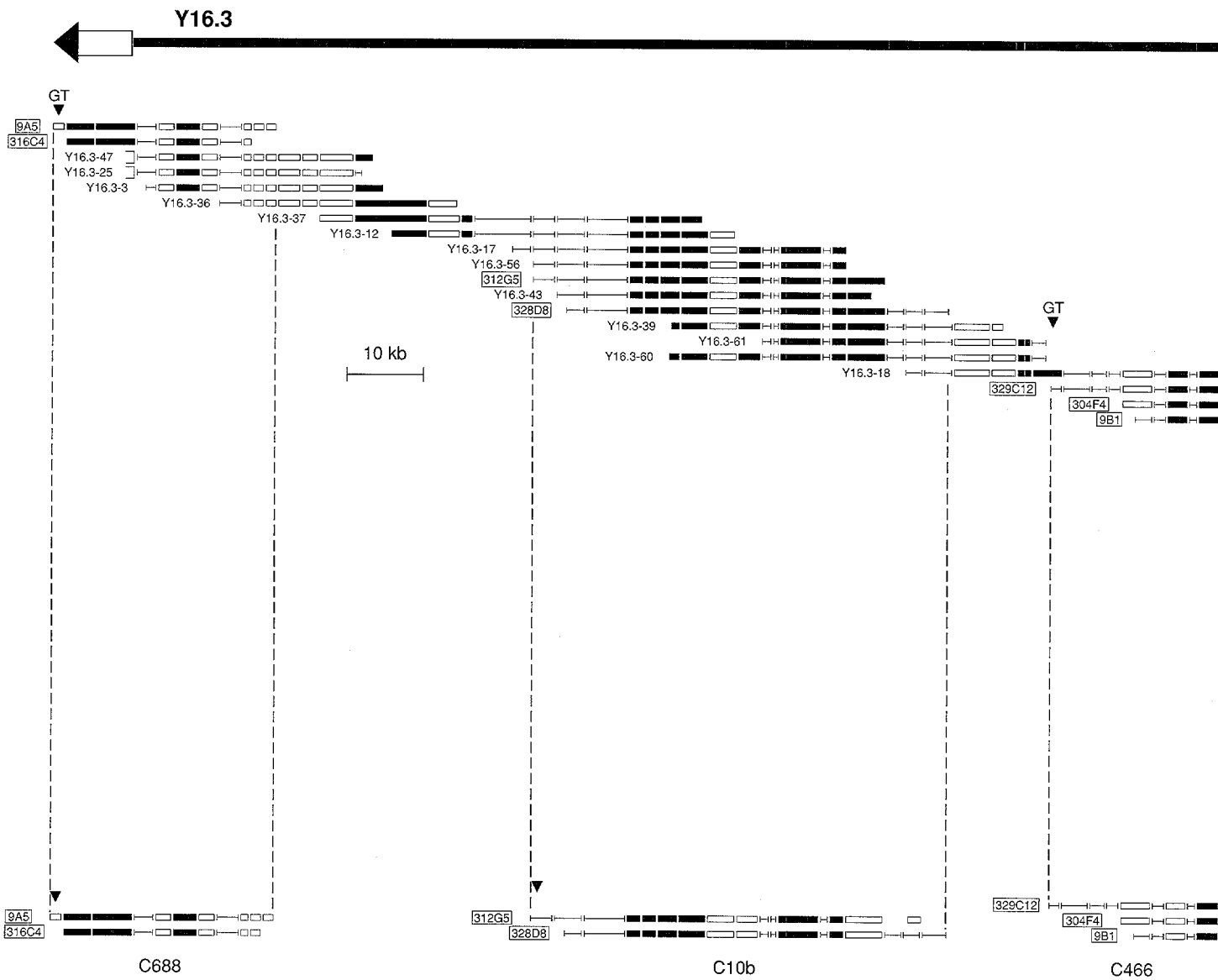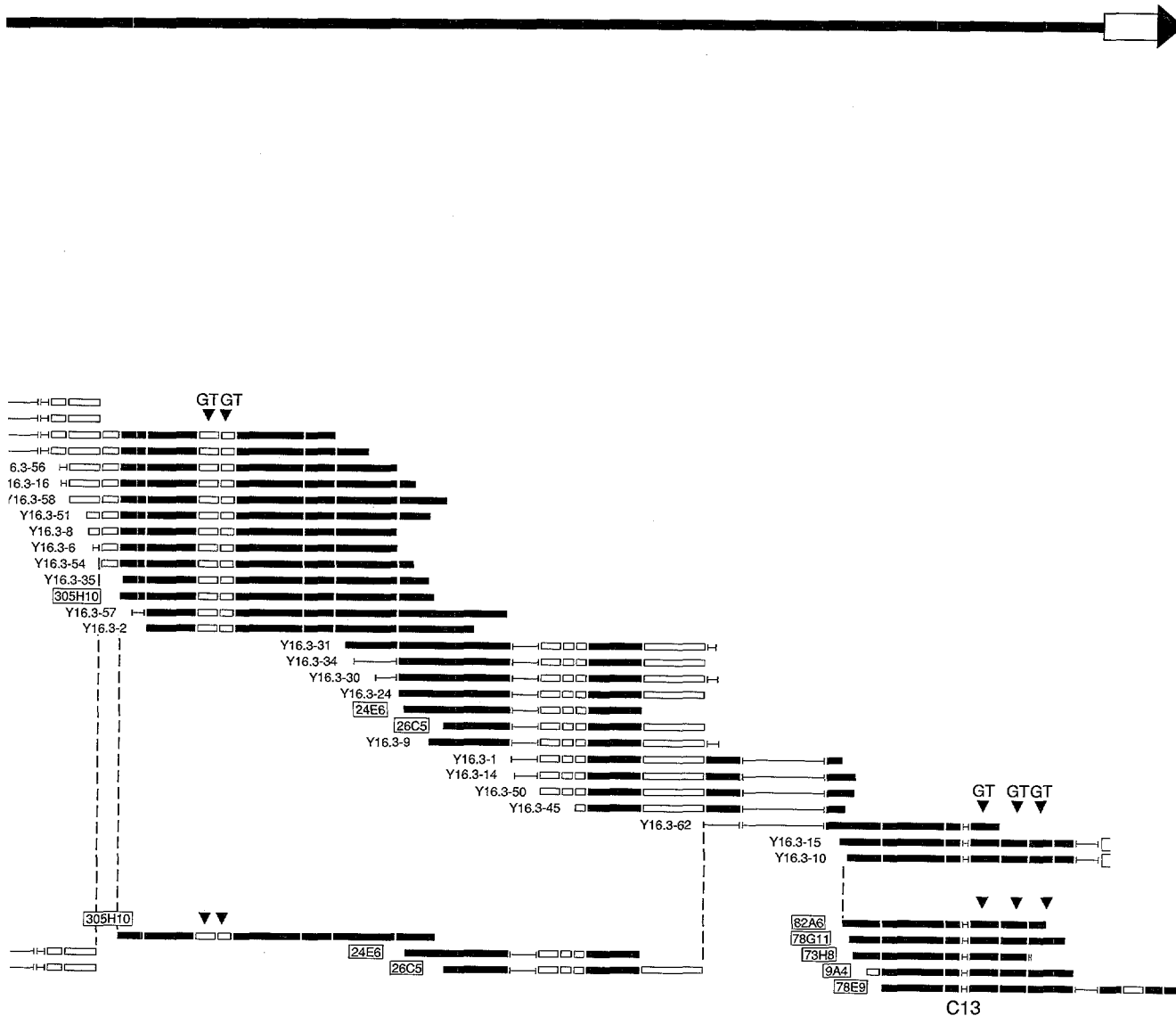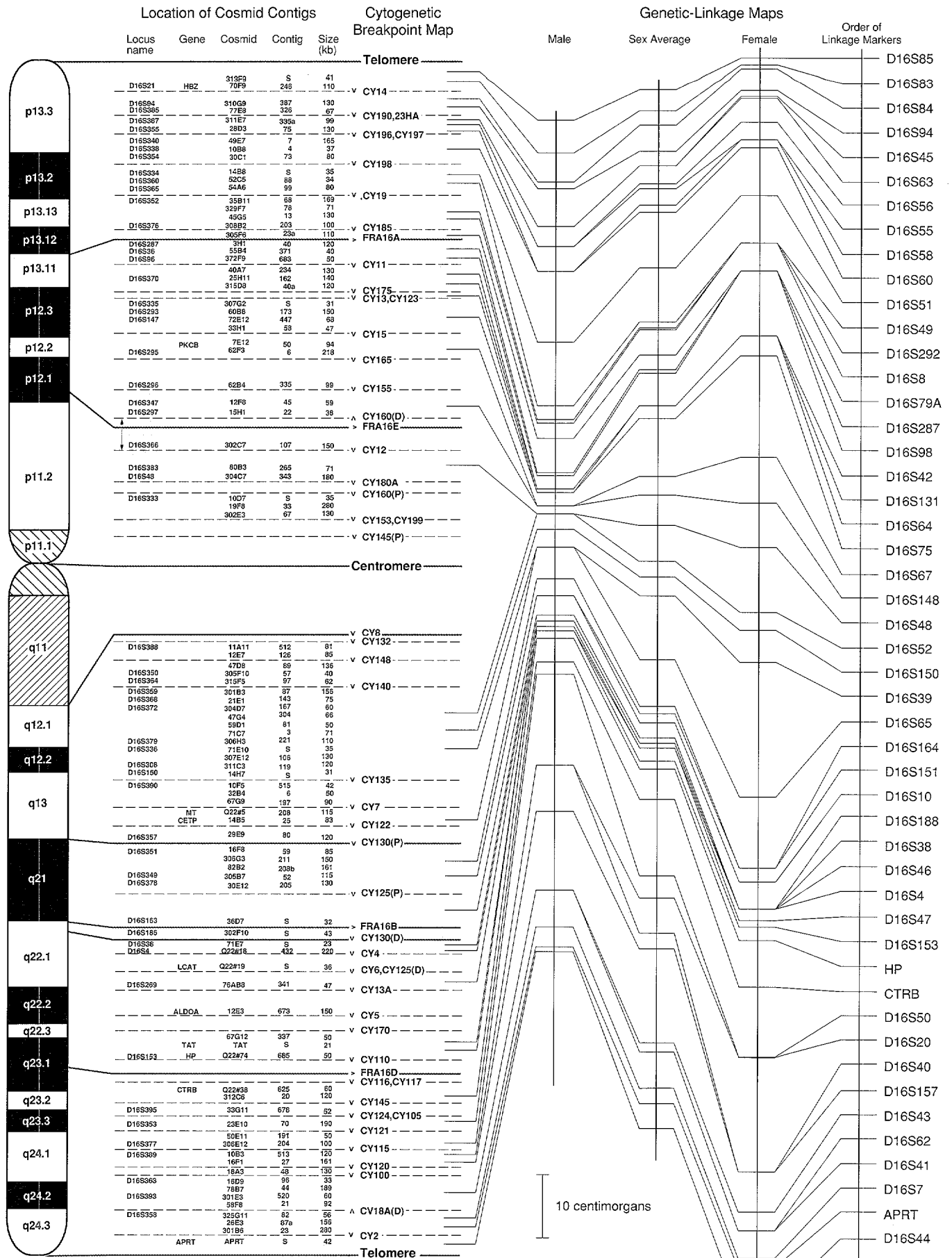

(b) YAC-cosmid contig

Figure 10. Confirmation of YAC-Cosmid Overlaps by YAC Subcloning

Hybridization experiments using inter-Alu PCR products from YAC clone Y16.3 suggested that this clone bridges the gaps between four cosmid contigs. To confirm that result, Y16.3 was subcloned into cosmid vectors and, as shown in (a), the resulting cosmid subclones were fingerprinted and assembled into a contig spanning the YAC. Five cosmid contigs from our chromosome 16 map were then aligned with the YAC subclone contig, based on their repetitive-sequence fingerprints. These results confirm the overlaps deduced from the hybridization experiments. Overlap of Y16.3 with an additional contig, C688, was detected by repetitive-sequence fingerprinting of the cosmid subclones of Y16.3. Thus, four out of five cosmid contigs that overlap this YAC were detected by the hybridization of inter-Alu PCR products to the high-density arrays of fingerprinted cosmid clones. (b) Other hybridization experiments have indicated overlaps with two other YACs and a sixth cosmid contig. Together these three YACs and six cosmids cover over 600,000 base pairs in chromosome 16.

Location of Cosmid Contigs

Locus name	Gene	Cosmid	Contig	Size (kb)
D16S21	HBZ	313F9 / 70F9	S / 248	41 / 110
D16S94		310G9	387	130
D16S385		77E8	326	67
D16S387		311E7	335a	99
D16S355		28D3	75	130
D16S340		49E7	7	165
D16S338		10B8	4	37
D16S354		30C1	73	80
D16S334		14B8	S	35
D16S360		52C5	88	34
D16S365		54A6	99	80
D16S352		35B11	68	169
		329F7	78	71
		45G5	13	130
D16S376		308B2	203	100
		305F6	23a	110
D16S287		3H1	40	120
D16S36		55B4	371	40
D16S86		372F9	683	50
		40A7	234	130
D16S370		25H11	162	140
		315D8	40a	120
D16S335		307G2	S	31
D16S293		60B8	173	150
D16S147		72E12	447	68
		33H1	58	47
	PKCB	7E12	50	94
D16S295		62F3	6	218
D16S296		62B4	335	99
D16S347		12F8	45	59
D16S297		15H1	22	38
D16S366		302C7	107	150
D16S383		60B3	265	71
D16S48		304C7	343	180
D16S333		10D7	S	35
		19F8	33	280
		302E3	67	130
D16S388		11A11	512	81
		12E7	126	85
		47D8	89	136
D16S350		305F10	57	40
D16S364		315F5	97	62
D16S359		301B3	87	155
D16S368		21E1	143	75
D16S372		304D7	167	60
		47G4	304	66
		59D1	81	50
		71C7	3	71
D16S379		306H3	221	110
D16S336		71E10	S	35
D16S308		307E12	105	130
D16S150		311C3	119	120
		14H7	S	31
D16S390		10F5	515	42
		32B4	6	50
		67G9	197	90
	MT	Q22#5	208	115
	CETP	14B5	25	83
D16S357		29E9	80	120
D16S351		16F8	59	85
		306G3	211	150
		82B2	208b	161
D16S349		305B7	52	115
D16S378		30E12	205	130
D16S163		36D7	S	32
D16S185		302F10	S	43
D16S36		71E7	S	23
D16S4		Q22#18	432	220
	LCAT	Q22#19	S	36
D16S269		76AB8	341	47
	ALDOA	12E3	673	150
		67G12	337	50
	TAT	TAT	S	21
D16S153	HP	Q22#74	685	50
	CTRB	Q22#38	625	60
		312C6	20	120
D16S395		33G11	676	52
D16S353		23E10	70	190
		50E11	191	50
D16S377		305E12	204	100
D16S389		10B3	513	120
		16F1	27	161
		18A3	48	130
D16S363		16D9	96	33
		78B7	44	189
D16S393		301E3	520	60
		58F8	21	92
D16S358		325G11	62	56
		26E3	87a	156
		301B6	23	280
	APRT	APRT	S	42

Cytogenetic Breakpoint Map

Telomere

CY14
CY190,23HA
CY196,CY197
CY198
CY19
CY185
FRA16A
CY11
CY175
CY13,CY123
CY15
CY165
CY155
CY160(D)
FRA16E
CY12
CY180A
CY160(P)
CY153,CY199
CY145(P)

Centromere

CY8
CY132
CY148
CY140
CY135
CY7
CY122
CY130(P)
CY125(P)
FRA16B
CY130(D)
CY4
CY6,CY125(D)
CY13A
CY5
CY170
CY110
FRA16D
CY116,CY117
CY145
CY124,CY105
CY121
CY115
CY120
CY100
CV18A(D)
CY2

Telomere

Genetic-Linkage Maps

Male Sex Average Female

Order of Linkage Markers

D16S85
D16S83
D16S84
D16S94
D16S45
D16S63
D16S56
D16S55
D16S58
D16S60
D16S51
D16S49
D16S292
D16S8
D16S79A
D16S287
D16S98
D16S42
D16S131
D16S64
D16S75
D16S67
D16S148
D16S48
D16S52
D16S150
D16S39
D16S65
D16S164
D16S151
D16S10
D16S188
D16S38
D16S46
D16S4
D16S47
D16S153
HP
CTRB
D16S50
D16S20
D16S40
D16S157
D16S43
D16S62
D16S41
D16S7
APRT
D16S44

10 centimorgans

Chromosome 16 bands: p13.3, p13.2, p13.13, p13.12, p13.11, p12.3, p12.2, p12.1, p11.2, p11.1, q11, q12.1, q12.2, q13, q21, q22.1, q22.2, q22.3, q23.1, q23.2, q23.3, q24.1, q24.2, q24.3

probe was derived. Consequently, the cosmid clone is highly likely to overlap the YAC clone.

To date we have generated inter-Alu PCR products from 411 YAC clones and hybridized those products to the arrays of fingerprinted cosmid clones. As shown in Figure 9, the inter-Alu products yield intense hybridization signals. The hybridization results enabled us to incorporate 334 YAC clones into our cosmid contig map. The PCR products from 133 YAC clones showed overlap with only a single cosmid contig and therefore extended those contigs but did not close any gaps in the map. Other YACs were shown to overlap as many as six separate cosmid contigs. Table 2 (page 199) lists the number of YACs whose PCR products hybridized to clones in one, two, three, four, five, or more than five cosmid contigs. The hybridization results also enabled us to join 203 singletons into the YAC-cosmid contigs. The number of YAC cosmid contigs in our map is now 462, and the average contig size has grown from 100,000 base pairs to 218,000 base pairs. The total number of "islands" in

the map (462 YAC-cosmid contigs plus 54 YAC singletons) cover 94 percent of chromosome 16. Overlaps between YAC and cosmid clones were detected by hybridization of YAC inter-Alu PCR products to cosmid clones.

Verification of YAC-Cosmid Contigs

Our implicit assumption in the discussion above was that if the inter-Alu PCR products from a YAC hybridize to a cosmid clone, the human DNA insert in the YAC clone overlaps the human insert in the cosmid clone, and thus the two are from the same region of chromosome 16. However, we have discovered that chromosome 16 contains a number of low-abundance repetitive sequences (see "What's Different about Chromosome 16?"). Those repetitive sequences would not have been masked by annealing the PCR products with $Cot1$ repetitive DNA prior to hybridization. Therefore, if the inter-Alu PCR products from a YAC clone contain those low-abundance repeats, they would hybridize to cosmid

clones that did not necessarily overlap the YAC clone. Consequently, we used an independent method to confirm the inferred overlaps between YACs and cosmid contigs.

Our procedure involved subcloning the DNA insert in each of seven YAC clones into cosmid vectors, generating a repetitive-sequence fingerprint for each of the resulting cosmid subclones, and comparing the fingerprints of the subclones to each other and to the fingerprints of the original set of fingerprinted cosmid clones to detect overlaps. Figure 10(a) shows how the cosmid subclones of YAC 16.3 overlapped among themselves and linked up with members of our original set of fingerprinted cosmid clones. Hybridization of inter-Alu PCR products had indicated that the YAC 16.3 clone overlapped four cosmid contigs. The results of subcloning the YAC confirmed the hybridization results. Two more YAC clones were found to overlap this region based on hybridization of their inter-Alu PCR products. This YAC-cosmid contig currently contains two of these YACs and six cosmid contigs, and

Figure 11. The Integration of Physical and Genetic-Linkage Maps of Chromosome 16

Physical and genetic-linkage mapping data presently available for chromosome 16 are summarized in the figure on this spread. Together they provide a resource for isolating a variety of genes on the chromosome. At right are three genetic-linkage maps showing genetic distances (in centimorgans) of 49 polymorphic DNA markers derived from male, female, and sex-averaged linkage data. These data were compiled by the Second International Workshop on Human Chromosome 16 and are based on analysis of pedigrees in CEPH (Centre d'Etude du Polymorphisme Humain). The coordinates of the physical mapping data are defined by (1) the cytogenetic map showing the dark and light Giemsa-stained bands of chromosome 16; and (2) the cytogenetic breakpoint map, the set of fifty horizontal lines that are positioned along the chromosome bands at the fifty breakpoints of chromosome 16 determined from the mouse/human hybrid-cell panel (see Figure 6). A cosmid clone from our contig map can be localized to a region or interval between two breakpoints by showing that it is present in the DNA of hybrid cells containing the chromosomal region corresponding to that interval but absent in the DNA of hybrid cells lacking that region. Each of 140 cosmids, and thus the contigs in which they reside, have now been placed into one of the 50 intervals. The YACs that overlap those 140 contigs are thereby regionally localized as well. The DNA in the cosmid contigs and YACs that have been located on the breakpoint map covers 21 million base pairs, or about 21 percent of the chromosome. In a separate effort, polymorphic DNA markers from the linkage map have been located onto the breakpoint map thereby integrating the linkage map with the cytogenetic breakpoint map and with the cosmid contigs located on the breakpoint map. We have also integrated the linkage map directly with our cosmid contigs by hybridizing 50 gene and polymorphic DNA markers to our high-density arrays of fingerprinted clones and identifying which clones contain those genes and markers. Shown in red are cosmids that have been both regionally localized and shown to contain a marker from the linkage map.

it spans a region over 600,000 base pairs long [see Figure 10(b)]. In most instances the overlaps inferred from the hybridization of YAC inter-Alu PCR products were confirmed by the analysis of YAC subclones. In one instance, the inter-Alu PCR products contained a low-abundance repeat and produced a false YAC-cosmid overlap. Such false overlaps can be avoided by mapping the locations of these low-abundance repeats. Additional experiments showed that 85 to 90 percent of the cosmids that overlap a YAC are identified by the hybridization of YAC inter-Alu PCR products. In general, our verification experiments suggest that YAC inter-Alu PCR products provide convenient and reliable probes for integrating YACs into cosmid contig maps.

Integration of the Physical Map with the Genetic-Linkage Map

As discussed in Part I of "Mapping the Genome," genetic-linkage analysis with polymorphic DNA markers is often the only way to find the approximate location of genes that cause inherited disorders. The polymorphic DNA markers that are tightly linked to, or usually co-inherited with, certain diseases are located close to the causative gene (see "Modern Linkage Mapping"). To find the gene, those markers must be located on a contig map and the cloned DNA in the neighborhood of the markers can then be searched for the causative gene. In other words, the genetic-linkage map must be integrated with the physical map.

Although our contig map is not yet complete, we have been locating previously cloned genes and polymorphic DNA markers on our cosmid contigs. Here, again, the high-density arrays of fingerprinted cosmid clones are an invaluable resource. Gene and DNA-marker probes are radioactively labeled and hybridized to these arrays to determine which cosmids contain those genes or markers. Alternatively, if a gene or marker exists in a cosmid

from another library, we can fingerprint that clone to integrate it with our existing contigs. Using both of these approaches, we have now located more than 50 genes and DNA markers on cosmid contigs, thereby integrating our cosmid contigs with genetic-linkage maps.

Earlier we mentioned that 140 contigs have also been localized to intervals between breakpoints on the cytogenetic breakpoint map of chromosome 16 derived from the panel of 50 mouse/human hybrid cells (see Figure 5). In addition, hybridization experiments show that inter-Alu products from 82 YAC clones overlap those localized contigs. The YAC clones and cosmid contigs now localized to intervals on chromosome 16 cover 21.4 million base pairs. Figure 11 summarizes the integration achieved so far between the linkage maps, the cytogenetic breakpoint map, and our cosmid contig map.

Application of the Map toward the Isolation of Disease Genes. The integrated maps provide potent resources to identify, isolate, and sequence regions

Figure 12. Chromosome-16 Maps at Different Levels of Resolution

Maps of chromosome 16 are being made by several different techniques and at a wide range of resolutions. The figure shows only a few of the landmarks on each map and also indicates the level of resolution presently available for each. The three low-resolution maps include a cytogenetic map, the hybrid-cell, or cytogenetic-breakpoint map, and the genetic-linkage map. At higher resolution is the cosmid contig map, which presently consists of separate contigs that are being connected by YAC clones. At the highest level of resolution, which is the sequences of bases in the genome, STSs are being generated to serve as unique physical landmarks. These landmarks can be located on all physical maps at all levels of resolution. The position of STS N16Y1-10 is traced from one level of resolution to another. It can be amplified, or duplicated millions of times, by the polymerase chain reaction using the two unique primers shown at the bottom of figure. At the top of the figure is shown the position of the STS (red) determined by in-situ hybridization to cosmid clones from which the STS was derived. In-situ hybridization localizes the STS to a region 3 to 4 million bases in length in bands 16q12.1 and 16q12.2. The STS is also shown regionally localized to the interval between breakpoints CY7 and CY8 on the hybrid-cell cytogenetic map. The intervals on this map have an average size of 1.6 million bases. The particular STS shown is not polymorphic, and therefore it cannot be located on the genetic-linkage map through linkage analysis. However, the DNA markers on the linkage map have been regionally localized on the hybrid-cell map. The alignment between the two maps allows us to infer that the STS lies between markers 16AC6.5 and D16S150 on the genetic-linkage map. The next higher level of resolution is provided by contig maps of overlapping cloned fragments. The figure shows a YAC clone containing the STS as well as a cosmid contig from which the STS was derived. The YAC clone must overlap the cosmid contig because they both contain the same STS. The position of the YAC relative to the cosmid contig is known because all inter-Alu PCR products from the YAC clone hybridized to all clones in the cosmid contig. The STS was derived from the right end clone of the cosmid contig. The information at the highest level of resolution is the base sequence of the STS determined in the process of developing the PCR protocol that recognizes and amplifies this sequence whenever it appears in a DNA sample.

Chromosome 16

←————— 95 million base pairs —————→

p   q

Low resolution (2–4 million base pairs)

13.3 13.2 13.13 13.12 13.11 12.3 12.2 12.1 11.2 11.1   11.2 12.1 12.2 13.0 21.0 22.1 22.2 22.3 23.1 23.2 23.3 24.1 24.2 24.3

Cytogenetic map
(Resolution of in situ hybridization 3–4 megabases)

CY14 23HA CY19 CY11 CY180 CY13 CY15 CY165 CY12   CY8 CY7 FRA16B CY4 FRA16D CY2

Cytogenetic breakpoint map
(Resolution 1.6 megabases; not all breakpoints are shown)

D16S85 D16S60 D16S159 D16S48 16AC6.5 D16S150 D16S149 D16S160 D16S40 D16S144

Genetic-linkage map
(Resolution 3–5 centimorgans; not all linkage markers are shown)

500 times expansion

YAC N16Y1

←————— 150,000 base pairs —————→

High resolution (1–100,000 base pairs)

310C4
N16Y1-29
N16Y1-18
N16Y1-13
N16Y1-14
N16Y1-12
N16Y1-16
N16Y1-30
5F3
312F1
309G11
N16Y1-19
N16Y1-10

▼ = $(GT)_n$

Cosmid Contig 211

Physical map of overlapping cosmid clones
(Resolution 5–10 kilobases)

500 times expansion

STS N16Y1-10

Primer

3'-AGTCAAACGTTTCCGGCCTA-5'

5'   5'-GATCAAGGCGTTACAGTGA   TCAGTTTGCAAAGGCCGGAT 3'
3'   3'-CTAGTTCCGCAATGTACT   AGTCAAACGTTTCCGGCCTA-5' 5'

Sequence-tagged site
(Resolution 1 base)

5'-GATCAAGGCGTTACATGA-3'

Primer

←————— 219 base pairs —————→

These maps include the overlap relationships between cosmid contigs and YACs, the regional localization of contigs, YACs, and STSs, and the integration of the genetic-linkage map with the physical contig map.

Figure 12 illustrates the levels of resolution at which information about chromosome 16 is available and also illustrates how STSs serve to integrate the various types of information and levels of resolution. These mapping data, in combination with the resources used to generate the data (the high-density arrays of cosmid clones, the pooled YAC library, the STSs, and the hybrid-cell panel), are already proving useful for the isolation of disease genes and other important regions on chromosome 16. For example, these resources were used to complete the map for the metallothionein gene family, to isolate the chromosome 16 microdeletion region associated with Rubenstein-Taybi syndrome, and to identify chromosome 16-specific repetitive DNA sequences associated with rearrangements of this chromosome that accompany a type of acute nonlymphocytic leukemia.

Several national and international collaborative efforts (described in the accompanying box) are now underway to isolate a variety of disease genes on chromosome 16. Each of these efforts takes advantage of the physical mapping progress on chromosome 16, and collectively they illustrate how the physical mapping of the human genome already has far-reaching significance in the field of medicine.

Completing the Map and Looking toward the Future

In line with the mapping goals stated in the Human Genome Project's Five-Year Plan, the completed map of chromosome 16 will have at most 100 contigs with lengths of between 1 and 2 million base pairs. The contigs will be ordered along the chromosome and represent at least 99 percent of the DNA within it. Moreover the map will be dotted with STS markers at intervals of 100,000 to 200,000 base pairs. Every region of the chromosome will then be rapidly accessible by STS screening of a genomic YAC library.

To complete this final map, we will be making a second YAC library of chromosome 16 by using a restriction enzyme whose restriction sites have a distribution pattern different from those of Cla1 (which was the restriction enzyme used in the construction of the first YAC library). A directed approach will then be used to screen this library (and a total genomic library if necessary) for YACs that extend the current YAC-cosmid contigs. We expect that most of the remaining gaps can be closed in this manner. The ongoing development of STSs from the original 576 cosmid contigs will provide the framework for an STS map at a resolution between 100,000 and 200,000 base pairs.

The approach we used to map chromosome 16 is resulting in a high-resolution map of this chromosome. The repetitive-sequence fingerprinting of cosmid clones, the subsequent assembly of contigs, and the evaluation of contig accuracy and chromosome coverage through hybridization experiments have produced a robust map with information on sizes, ordering, and sequence complexity of DNA restriction fragments. Mapping data of this type are invaluable for interrelating chromosome structure with function. Already the chromosomal distribution of $(GT)_n$ repeats has been determined from those data.

With the advent of YAC and PCR technologies, it is now possible to rapidly produce a lower-resolution map of an entire chromosome. YAC clones are 10 to 20 times larger than cosmid clones, so far fewer are needed to create a complete contig map. The assembly of contigs by STS-content mapping is relatively efficient and straightforward. Although physical maps constructed from YAC clones and STS markers will not be as useful for elucidating the structure-function relationships of chromosomes as those made from cosmid clones, the YAC maps still permit immediate access to genes or regions of medical and scientific importance. Consequently, in developing a strategy to map a second chromosome, chromosome 5, we chose to exploit the new technologies. Deborah Grady at our Laboratory and John Wasmuth at the University of California, Irvine, have begun a collaborative effort using chromosome-specific STSs and YAC libraries to rapidly generate a relatively low-resolution map of chromosome 5. Their strategy and some early data are presented in "Mapping Chromosome 5." ■

Collaborations on the Isolation of Disease Genes on Chromosome 16

Polycystic Kidney Disease (PKD1). Polycystic kidney disease is a common dominant single-gene disorder (affecting at least 1 in 1000 Caucasians) that is responsible for cystic kidneys, accompanied by hypertension and renal failure. The principal locus for the genetic defect, PKD1, has been assigned to chromosome band 16p13.3 by genetic linkage with polymorphic DNA markers shown to reside in that band.

Steve Reeders (Yale University School of Medicine), Anna-Maria Frischauf (Imperial Cancer Research Fund), and collaborators have constructed both a long-range restriction map (covering 1 million base pairs) and an ordered contig map (covering 75,000 base pairs) that span the entire PKD1 region. Construction of the contig map by cosmid walking from multiple start sites within the region was greatly aided by the use of two chromosome 16-specific cosmid libraries constructed at Los Alamos. A gene-by-gene search is now being carried out in the region to identify candidate disease genes (genes that are expressed in the kidney and that have alleles that are specific to affected individuals). This effort will probably soon lead to the identification of the gene that is responsible for the disease.

Batten's Disease (CLN3). Batten's disease is a juvenile-onset neurodegenerative disease with incidence rates of up to 1 in 25,000 live births. It is characterized by the accumulation of autofluorescent fatty pigments in neurons. The responsible locus (CLN3) is inherited in an autosomal recessive pattern. That is, the defective allele must be present on both chromosomes in order for the disease to be manifested. The gene responsible for this disease has been mapped to the region between two polymorphic markers in the chromosomal band 16p12.

We have found thirteen cosmid contigs and one YAC clone from our physical map that lie in this same interval, and in collaboration with groups in London (Mark Gardiner), the Netherlands (Martijn Breuning), and Australia (David Callen), we are developing new polymorphic DNA markers from these contigs in an attempt to find markers that are closer to the disease locus. We have used prior knowledge of the repetitive-sequence fingerprint of four of these cosmid clones to develop STSs containing GT-repeat sequences present on these clones. Since GT repeats tend to be variable in length, we expect these STSs to be polymorphic and therefore useful for linkage analysis. We are now evaluating their informativeness in linkage studies. (Genetic-linkage markers for the remaining cosmids are being developed by the other laboratories with the aid of the fingerprint data.) The development of these new genetic-linkage markers in the Batten's-disease region will allow the disease gene to be localized to a manageable region (approximately 1 million bases). Then construction of a detailed physical map starting from the existing contigs and YACs in the region can be completed. The availability of the Los Alamos clones in the Batten's region has substantially reduced the extensive work that would have been required to find genetic-linkage markers from this region and to construct a complete map of the region.

Familial Mediterranean Fever (FMF). FMF is an autosomal recessive form of arthritis that is characterized by acute attacks of fever with inflammation of the lining of the abdominal cavity (peritonitis), pleural cavity (pleurisy), and joints (synovia). The gene frequency among non-Ashkenazic Jews, Armenians, Turks, and Middle Eastern Arabs is comparable to the gene frequency for cystic-fibrosis defects among Caucasians (1 in 25). As with Batten's disease, genetic-linkage markers flanking

the disease locus have been identified by researchers led by Dan Kastner at the National Institutes of Health. We are working with that group to identify contigs and YACs that lie within this region so that additional genetic-linkage markers can be developed.

Rubenstein-Taybi Syndrome (RTS). RTS is characterized by abnormal facial features, broad thumbs and big toes, and mental retardation. RTS is a rare disorder that accounts for an estimated 1 in 500 institutionalized cases of mental retardation. Almost all cases seem to arise from spontaneous mutations. Three patients with RTS have been found to have translocations involving the short arm of chromosome 16. Using fluorescence in-situ hybridization, Martijn Breuning (Leiden University) was able to pinpoint the location of breakpoints in two of these patients relative to cosmids that he had ordered in the region in his group's effort to map breakpoints associated with ANLL M4. One of these cosmids, RT1, appeared to be very close to the breakpoints and was found to be deleted in 6 out of 24 patients. By screening our gridded arrays of chromosome 16 cosmids with RT1, Breuning identified one cosmid, 316H7, that overlapped RT1 by 10 kilobases. This overlapping cosmid was also hybridized to metaphase chromosomes from the two patients with RTS. In both cases, Breuning found three signals, one on the normal chromosome 16, a second signal on the aberrant chromosome 16, and a third on the chromosome that the p arm of 16 had translocated to. These results indicated that cosmid 316H7 spanned both translocation breakpoints in these RTS patients. Since the gene(s) responsible for RTS is likely to be disrupted by these breakpoints, the identification of cosmid 316H7, which spans the breakpoints, opens the door for identification of the gene(s) that causes this syndrome.

Acute Nonlymphocytic Leukemia (ANLL). In contrast to PKD1, CLN3, and FMF, which follow a Mendelian pattern of inheritance, acute nonlymphocytic leukemia is a polygenic trait, that is, it involves the interaction of several genes. A high frequency of rearrangements (inversions and translocations involving both the p and q arms) of chromosome 16 is associated with a specific subtype of acute nonlymphocytic leukemia known as ANLL subtype M4 (see "What's Different about Chromosome 16?"). This association suggests that chromosome 16 may contain at least one of the genes involved in the progression of the disease state and that the chromosomal rearrangements disrupt the functioning of that gene. We are collaborating with groups in the United States, Australia, and the Netherlands to isolate the chromosomal breakpoint regions associated with ANLL. Our prior identification of chromosome 16-specific repeats that map near these regions is aiding the search for the breakpoint regions. Genes that are disrupted as a result of the chromosomal rearrangements will be candidates for having a role in ANLL.

Breast Cancer. Like ANLL, breast cancer appears to be a polygenic trait involving specific alterations of chromosome 16 in addition to alterations in other genes. Deletions in the q22 region of chromosome 16 that are not always detectable at the gross microscopic level occur at a relatively high frequency in the malignant cells of breast tumors. These deletions are readily detectable using fluorescence in-situ hybridization by noting the absence of a positive hybridization signal from a probe that usually hybridizes to the deleted region and the presence of a signal from a second probe that hybridizes to the centromere. We have sent cosmid clones from the q arm of chromosome 16 to Joe Gray (UCSF), who is attempting to pinpoint the region of deletion associated with breast cancers. A gene-by-gene search through the deleted region will presumably lead to the identification of a gene whose function suppresses the development of cancer (tumor-suppressor gene). ■

Further Reading

D. T. Burke, G. F. Carle, and M. V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806–812.

D. C. Schwartz and C. R. Cantor. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37:67–75.

R. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.

R. Saiki, D. H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. Erlich. 1988. Primer directed amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491.

M. Olson, L. Hood, and D. Botstein. 1989. A common language for physical linkage mapping of the human genome. *Science* 245:1434–1435.

C. L. Smith, J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor. 1987. A physical map of the *Escherichia coli* K12 genome. *Science* 236:1448–1453.

D. L. Daniels and F. R. Blattner. 1987. Mapping using gene encyclopaedias. *Nature* 325:831–832.

Y. Kohara, K. Akiyama, and K. Isono. 1987. The physical map of the whole E. coli chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 50:495–508.

M. V. Olson, J. E. Dutchik, M. V. Graham, G. M. Brodeur, M. Frank, M. MacColin, R. Scheinman, and T. Frank. 1986. Random-clone strategy for genomic restriction mapping yeast. *Proceedings of the National Academy of Sciences of the United States of America* 83:7826–7830.

A. Coulson, J. Sulston, S. Brenner, and J. Karn. 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 83:7821–7825.

A. V. Carrano, J. Lamerdin, L. K. Ashworth, B. Watkins, E. Branscomb, T. Slezak, M. Raff, P. J. De Jong, D. Keith, L. McBride, S. Meister, and M. Kronick. 1989. A high-resolution, fluorescence-based, semiautomated method for DNA fingerprinting. *Genomics* 4:129–136.

S. Brenner and K. J. Livak. 1989. DNA fingerprinting by sampled sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 86:8902–8906.

G. A. Evans and K. A. Lewis. 1989. Physical mapping of complex genomes by cosmid multiplex analysis. *Proceedings of the National Academy of Sciences of the United States of America* 86:5030–5034.

A. G. Craig, D. Nizetic, J. Hoheisel, G. Zehetner, and H. Lehrach. 1990. Ordering of cosmid clones covering the Herpes simplex virus type I (HSV-I) genome: A test case for fingerprinting by hybridization. *Nucleic Acids Research* 18:2653–2660.

L. L. Deaven, M. A. Van Dilla, M. R. Bartholdi, A. V. Carrano, L. S. Cram, J. C. Fuscoe, J. W. Gray, C. E. Hildebrand, R. K. Moyzis, and J. Perlman. 1986. Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *Cold Spring Harbor Symposia on Quantitative Biology* 51:159–167.

N. E. Morton. 1991. Parameters of the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 88:7474–7476.

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. *Chromosome* 95:375–386.

D. F. Callen, E. Baker, H. J. Eyre, and S. A. Lane. 1990. An expanded mouse-human hybrid cell panel for mapping human chromosome 16. *Annals of Genetics* 33:190–195.

C. F. Callen, C. E. Hildebrand, and S. Reeders. Report of the second international wokshop on human chromosome 16. Submitted for publication.

E. S. Lander and M. S. Waterman. 1988. Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* 2:231–239.

R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, and J.-R. Wu. 1988. A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 85:6622–6626.

J. Meyne, R. L. Ratliff, and R. K. Moyzis. 1989. Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* 86:7049–7053.

F. Rouyer, A. de la Chapelle, M. Andersson, and J. Weissenbach. 1990. An interspersed repeated sequence specific for human subtelomeric regions. *The EMBO Journal* 9:505–514.

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

J. L. Weber and P. E. May. 1989. Abundant class of human polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44:388–396.

A. J. Jeffreys, V. Wilson, and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* 314-:67–73.

M. F. Singer. 1982. Highly repeated sequences in mammalian genomes. *International Review of Cytology* 76:67–112.

B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, and H. H. Kazazian, Jr.. 1991. Isolation of an active human transposable element. *Science* 254:1805–1808.

H. F. Willard. 1989. The genomics of long tandem arrays of satellite DNA in the human genome. *Genome* 31-737–744.

R. J. Britten and D. E. Kohne. 1968. Repeated sequences in DNA. *Science* 161:529–540.

R. K. Moyzis, D. C. Torney, J. Meyne, J. Buckingham, J. M. Meyne, J.-R. Wu, C. Burks, K. M. Sirotkin, and W. B. Goad. 1989. The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* 4:273–289.

R. L. Stallings, D. C. Torney, C. E. Hildebrand, J. L. Longmire, L. L. Deaven, J. H. Jett, N. A. Doggett, and R. K. Moyzis. 1990. Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proceedings of the National Academy of Sciences of the United States of America* 87:6218–6222

E. M. Southern. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98:503–517.

A. P. Feinberg and B. Vogelstein. 1983. A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* 132:6–13.

T. M. Cannon, R. J. Koskela, C. Burks, R. L. Stallings, A. A. Ford, P. E. Hempfner, H. T. Brown, and J. W. Fickett. 1991. A program for computer-assisted scoring of southern blots. *Biotechniques* 10:764–767.

D. J. Balding and D. C. Torney. 1991. Statistical analysis of DNA fingerprint data for ordered clone physical mapping of human chromosomes. *Bulletin of Mathematical Biology* 53:853–879.

J. W. Fickett and M. J. Cinkosky. A genetic algorithm for assembling chromosome physical maps. Submitted for publication.

D. F. Callen, N. A. Doggett, R. L. Stallings, L. Z. Chen, S. A. Whitmore, S. A. Lane, J. K. Nancarrow, S. Apostolou, A. D. Thompson, N. M. Lapsys, H. J. Eyre, E. G. Baker, Y. Shen, R. I. Richards, K. Holman, H. Phillips, and G. R. Sutherland. High resolution cytogenetic-based physical map of human chromosome 16. Accepted for publication in *Genomics*.

R. L. Stallings, N. A. Doggett, C. Callen, S. Apostolou, P. Harris, H. Michison, H. Breuning, J. Sarich, C. E. Hildebrand, and R. K. Moyzis. Evaluation of a cosmid contig physical map of human chromosome 16. Accepted for publication in *Genomics*.

M. K. McCormick, J. H. Shero, M. C. Cheung, Y. W. Kan, P. A. Hieter, and S. E. Antonarakis. 1989. Construction of human chromosome 21-specific yeast artificial chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 86:9991–9995.

E. D. Green, R. M. Mohr, J. R. Jones, J. M. Buckingham, L. L. Deaven, R. K. Moyzis, and M. V. Olson. 1991. Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* 11:548–564.

E. D. Green and M. V. Olson. 1990. Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: A model for human genome mapping. *Science* 250:94–98.

M. K. McCormick, E. Campbell, L. Deaven, and R. Moyzis. Non-chimeric yeast artificial chromosome libraries from flow sorted human chromosomes 16 and 21. Submitted for publication.

E. D. Green and M. V. Olson. 1990. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the United States of America* 87:1213–1217.

D. L. Nelson, S. A. Ledbetter, L. Corbo, M. F. Victoria, R. Ramirez-Solis, T. D. Webster, D. H. Ledbetter, and C. T. Caskey. 1989. Alu polymerase chain reacgtion: A method for rapdi isolation of human-specific sequences from complex DNA sources. *Proceedings of the*

National Academy of Sciences of the United States of America 86:6686–6690.

P. A. Medvick, R. M. Hollen, R. S. Roberts, D. Trimmer, and T. J. Beugelsdijk. 1992. Automated DNA hybridization array construction and database design for robotic control and for source determination of hybridization responses. *International Journal of Genome Research* 1:17–23.

R. L. Stallings, N. A. Doggett, D. Bruce, M. K. McCormick, A. Ford, D. C. Torney, J. Dietz-Band, C. E. Hildebrand, and R. K. Moyzis. Approaching closure of a human chromosome 16 contig physical map using inter-Alu PCR products from a chromosome specific YAC library. Submitted for publication.

G. G. Germino, D. Weinstat-Saslow, H. Himmelbauer, G. A. J. Gillespie, S. Somlo, B. Wirth, N. Barton, K. L. Harris, A.-M. Frischauf, and S. T. Reeders. 1992. The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. *Genomics* 13:144–151.

M. Gardiner, A. Sandford, M. Deadman, J. Poulton, W. Cookson, S. Reeders, I. Jokiaho, L. Peltonen, H. Eiberg and C. Julier. 1990. Batten Disease (Spielmeyer-Vogt disease, juvenile onset neuronal ceroidlipofuscinosis) gene (CLN3) maps to human chromosome 16. *Genomics* 8:387–390.

E. Pras, I. Aksentijevich, L. Gruberg, J. E. Balow, L. Prosen, M. Dean, A. D. Steinberg, M. Pras, and D. L. Kastner. 1992. Mapping of a gene causing familial mediterranean fever to the short arm of chromosome 16. *New England Journal of Mediciane* 326:1509–1513.

M. H. Breuning, H. G. Dauwerse, G. Fugazza, J. J. Saris, L. Spruit, H. Wijnen, M. Tommerup, C. B. van der Hagen, K. Imaizumi, Y. Kuroki, M.-J. van den Boogaard, J. M. de Pater, E. Mariman, B. Hamel, H. Himmelbauer, A.-M. Frischauf, R. L. Stallings, G.-J. B. van Ommen, and R. C. M. Hennekam. 1992. Submicroscopic deletions of chromosome 16 in patients with Rubenstein-Taybi syndrome. Submitted for publication.

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

T. Sato, F. Akiyama, G. Sakamoto, F. Kasumi, and Y. Nakamura. 1991. Accumulation of genetic alteration and progression of primary breast cancer. *Cancer Research* 51:5794–5799.

# 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 non-staining 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.
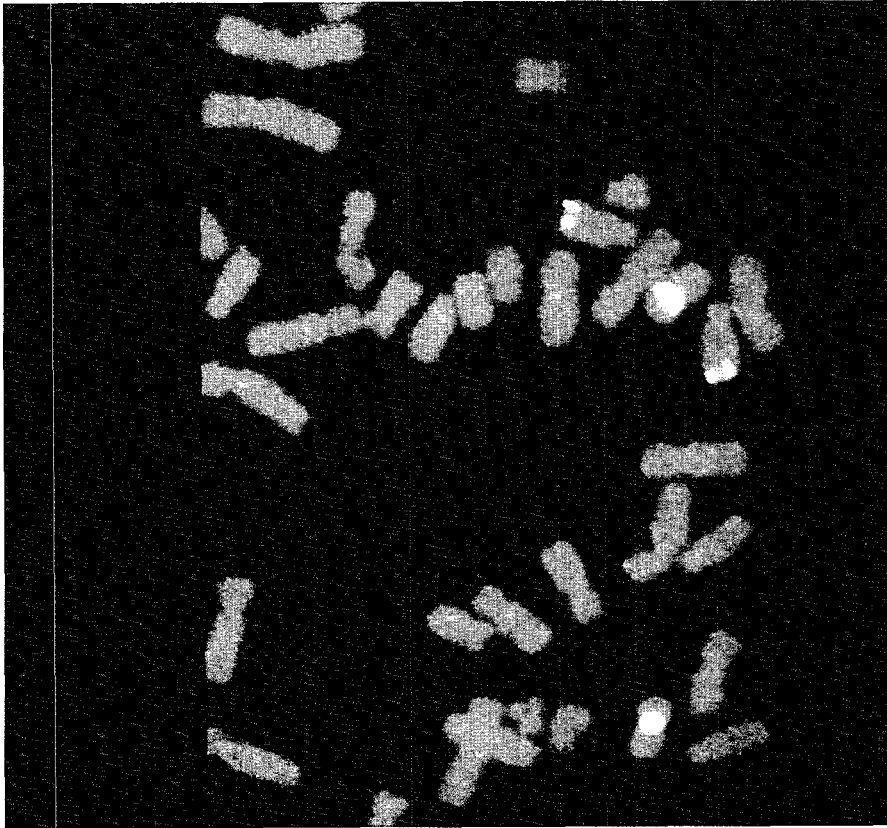
Photograph courtesy of David Ward,
Yale University School of Medicine

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

Chromosome 16 (Giemsa banding pattern)
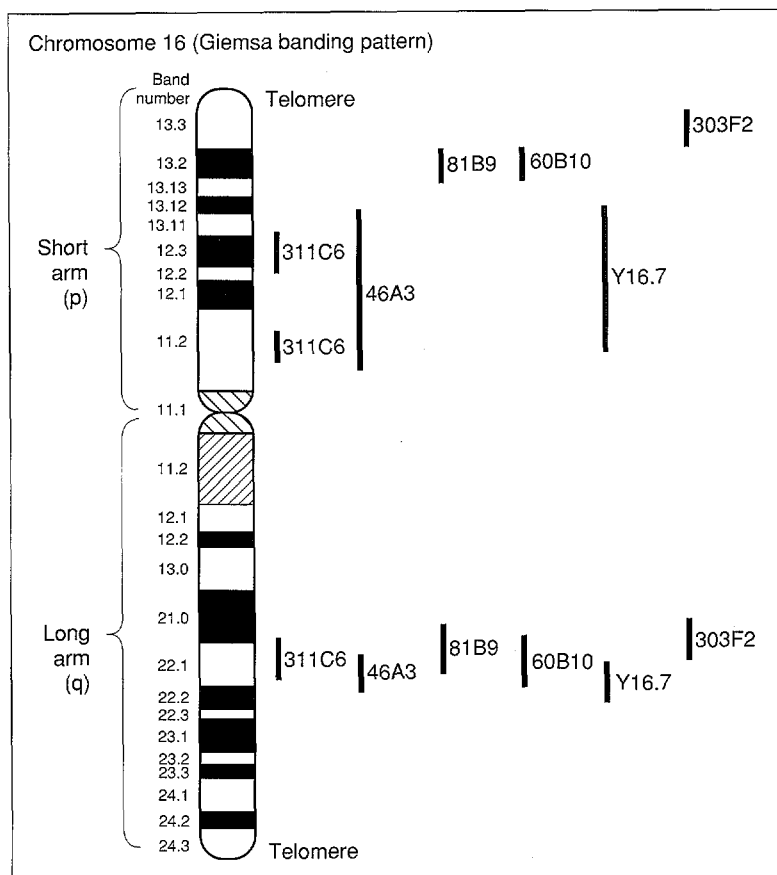
any of the five satellite sequences (α, β, I, II, III) that are commonly found in the centromeric region of all human chromosomes. Previous work at the Laboratory had revealed that a large block of chromosome-specific, satellite-II-variant DNA occurs at the pericentromeric region of the long arm of chromosome 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.
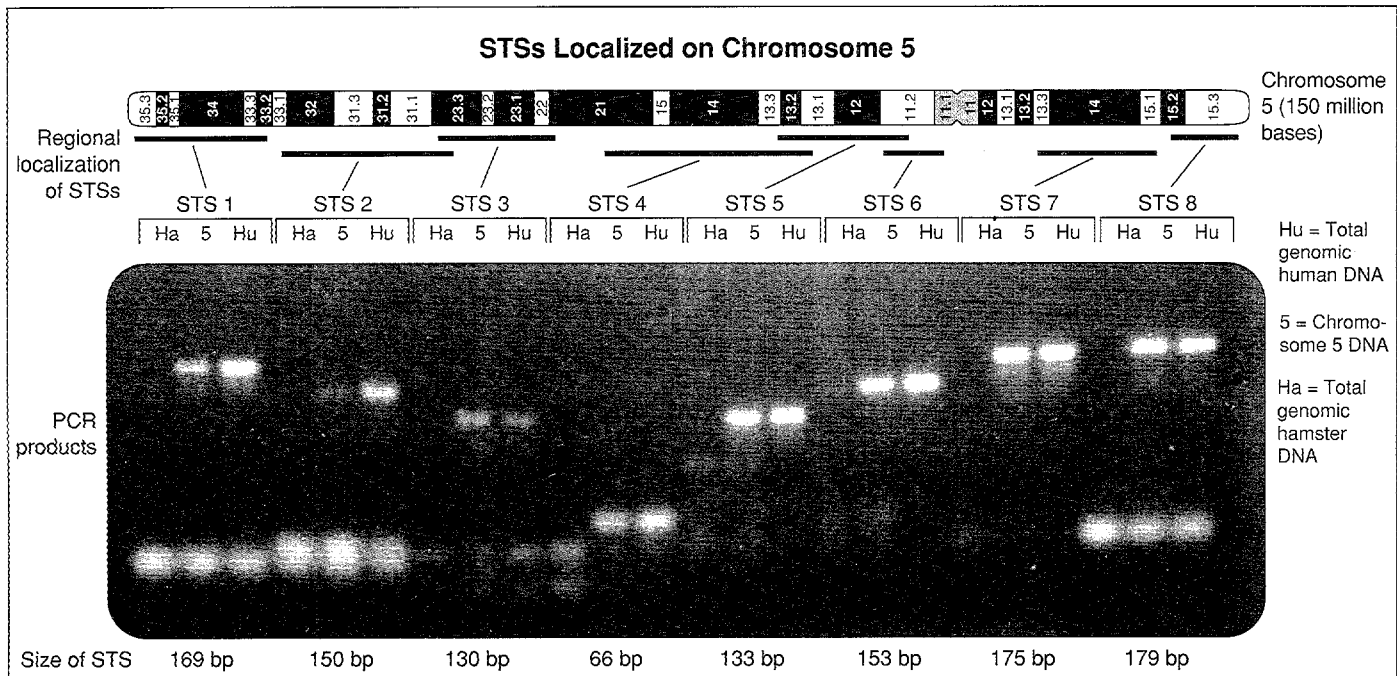
Mapping Chromosome 5 *Deborah Grady*

Constructing physical maps of complex genomes relies on the ability to isolate DNA segments for detailed analysis and to position those segments along the genome by identifying physical landmarks within them. The chromosome-16 physical map, now nearing completion, is a high-resolution map of DNA segments that have been isolated through cloning in cosmid and YAC vectors. The cloned fragments have been assembled into contigs and positioned along the chromosome based on detailed information about the positions of restriction sites, repetitive sequences, and the unique physical landmarks called STSs, or sequence-tagged sites. The chromosome-16 contig map provides information at a resolution of about 10,000 base pairs and will prove useful in studying chromosomal structure and organization.

In view of the need to complete physical maps of other chromosomes both rapidly and efficiently, we are adopting a different approach in mapping a second chromosome, chromosome 5. The goal is to construct a lower-resolution map consisting of (1) a series of STSs spaced evenly across the chromosome; and (2) YAC contigs assembled and ordered along the chromosome on the basis of their STS content. The project is being carried out in collaboration with John Wasmuth of the University of California at Irvine.

Our starting strategy utilizes the Los Alamos technologies for constructing chromosome-specific libraries to rapidly build a map covering 60 percent of the chromosome. The first step is to create a "framework" map of STS markers spaced at intervals of 0.5 to 1 million bases along chromosome 5. Given the statistics associated with generating STS markers at random and the fact that chromosome 5 is 194 million bases long, we will have to generate at least 400 STS markers to produce an STS map with a resolution of 1 million base pairs. We are developing the STS markers from a chromosome 5-specific library of M13 clones constructed at Los Alamos specifically for this purpose. Generating an STS involves sequencing a short cloned fragment of genomic DNA and identifying unique primer pairs from that sequence, which, when used in the polymerase chain reaction (PCR), will amplify a unique site in the genome. (See "The Polymerase Chain Reaction and Sequence-tagged Sites.")

Wasmuth is localizing the position of each STS to one of the intervals along human chromosome 5 defined by a panel of 30 hamster/human hybrid cells each containing various portions of chromosome 5. This localization is accomplished by determining through PCR screening which hybrid cells contain the STS and which do not. This method allows regional localization at a resolution of between 5 and 10 million base pairs. Plans are being made to refine the localization to a resolution of 200,000 base pairs using radiation-hybrid mapping. This mapping technique is analogous to genetic-linkage mapping in that distances are measured by how often two markers on the same chromosome become separated from one another. In linkage studies the separation is due to crossing over during meiosis, and the frequency of crossing over, the so-called genetic distance, is not necessarily proportional to the physical distance. In radiation-hybrid mapping the separation occurs through radiation-induced chromosome breakage, and the frequency of the radiation-induced breakage between two markers is linearly proportional to the physical distance separating the markers. Moreover, the technique is readily applied to any unique markers, in particular, to STSs.

Once generated and regionally localized on the chromosome, each STS will be "anchored," or located, on a non-chimeric YAC clone from a chromosome 5-specific YAC library, which has been constructed at Los Alamos. The cloning technique used to construct non-chimeric clones from flow-sorted chromosomes is discussed in "Libraries from Flow-sorted Chromosomes."

STSs Localized on Chromosome 5

Size of STS: 169 bp, 150 bp, 130 bp, 66 bp, 133 bp, 153 bp, 175 bp, 179 bp

Hu = Total genomic human DNA

5 = Chromosome 5 DNA

Ha = Total genomic hamster DNA

Chromosome 5 (150 million bases)

The non-chimeric YACs, localized along chromosome 5 by their STS content, will provide a solid base on which to build YAC contigs covering the chromosome. At Los Alamos, we will concentrate on mapping the short arm of chromosome 5 (52 million base pairs). Special emphasis will be placed on the region of chromosome 5 involved in the Cri du chat syndrome, one of the most common terminal-deletion syndromes in humans.

The figure (above) illustrates our early work on STS generation and regional localization. The upper portion shows the regional localization along chromosome 5 of eight STSs generated from our chromosome 5-specific M13 library. The regional localization (indicated with bars) will be reduced to intervals of 5 to 10 million bases once all available hybrid cells are screened for the presence of each STS.

The photograph in the lower portion of the figure shows the results of testing for the existence and uniqueness of each STS. The three gel lanes for each STS show the PCR products generated from total-genomic human DNA (right lane), chromosome-5 DNA (middle lane), and total-genomic hamster DNA (left lane) using the primer pairs that operationally define each STS. The PCR products from the three reactions were separated in parallel in a 3 percent agarose gel and stained with ethidium bromide to visualize the DNA. In all cases a single PCR amplification product of the same size resulted from the total-genomic human DNA and the chromosome-5 DNA. The hamster DNA served as a control to ensure that a positive signal from the chromosome-5 DNA did not represent a spurious signal arising from hamster DNA. In all cases, the hamster DNA yielded no PCR product. The test also shows that human/hamster hybrid cells can be screened for an STS without concern that false positive signals will arise from the hamster DNA in the hybrid cell. The PCR results demonstrate the existence of each STS as a unique landmark on chromosome 5 and the specificity of the PCR protocol defining each STS. The size of each STS is given at the bottom of the figure. ∎