

An Integrated Physical Map for the Short Arm of Human Chromosome 5

Ellen T. Peterson,¹ Robert Sutherland,^{1,2} Donna L. Robinson,¹ Leslie Chasteen,¹ Meryl Gersh,³ Joan Overhauser,³ Larry L. Deaven,¹ Robert K. Moyzis,^{1,4} and Deborah L. Grady^{1,4,5}

¹Life Sciences Division and Center for Human Genome Studies, ²Software Design and Development Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA; ³Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107 USA; ⁴Department of Biological Chemistry, College of Medicine, University of California-Irvine, Irvine, California 92697 USA

The short arm of human chromosome 5 contains ~48 Mb of DNA and comprises 1.5% of the genome. We have constructed a mega-YAC/ STS map of this region that includes 436 YACs anchored by 216 STSs. By combining and integrating our map with the 5p maps of other groups using the same recombinant DNA library, a comprehensive map was constructed that includes 552 YACs and 504 markers. The YAC map covers >94% of 5p in four YAC contigs, bridges the centromere, and includes an additional 5 Mb of 5q DNA. The average marker density is 95 kb. This integrated 5p map will serve as a resource for the continuing localization of genes on the short arm of human chromosome 5 and as a framework for both generating and aligning the DNA sequence of this region.

One of the primary initial goals of the Human Genome Project was to establish a physical map representing every chromosome of the human genome (Botstein et al. 1990). Advancements toward this end include detailed maps of chromosomes Y (Foote et al. 1992), 3 (Gemmill et al. 1995), 11 (Qin et al. 1996), 12 (Krauter et al. 1995), 16 (Doggett et al. 1995) 19 (Ashworth et al. 1995), 21 (Chumakov et al. 1992; Nizetic et al. 1994), 22 (Bell et al. 1995; Collins et al. 1995), 7 (Bouffard et al. 1997), and X (Crollius et al. 1996; Nagaraja et al. 1997) as well as maps consisting of large segments of the entire genome (Cohen et al., 1993; Chumakov et al. 1995; Hudson et al. 1995; Gyapay et al. 1996; Schuler et al. 1996; Deloukas et al. 1998). These contributions will be critical for correctly aligning future DNA sequences (Venter et al. 1998) and the ultimate understanding of human biology and disease (Collins et al. 1998).

One of the best studied and more common (1: 50,000 births; Niebuhr 1978) human deletion syndromes is cri-du-chat (Lejuene et al. 1963; Niebuhr 1978). This syndrome results from a deletion of all or part of the short arm of chromosome 5. Previous efforts have yielded YAC contigs (Goodart et al. 1994; Gersh et al. 1997) covering the gene regions considered to account for some of the primary features of this syndrome (Overhauser et al. 1994; Church et al. 1995; Gersh et al. 1995). Cri-du-chat breakpoints are known to occur throughout most of the 5p region (Overhauser

et al. 1994). A detailed physical map of the entire chromosome arm would provide a valuable resource for future studies. In this paper, we describe an integrated chromosome YAC/STS physical map of 5p, which provides nearly complete coverage of this 48-Mb arm.

RESULTS AND DISCUSSION

Overview of the Project

The integrated 5p YAC/STS map was constructed by a multitiered approach as follows: (1) Nucleation was obtained with 84 regionally ordered 5p STS markers generated from flow-sorted chromosomes for this study (Grady et al. 1996; approximately one STS/0.6 Mb). The Genethon physical map of the human genome (Cohen et al. 1993; Chumakov et al. 1995) was examined to identify tier one to three mega-YACs likely to be derived from 5p. All 84 STSs were used to screen this subset of mega-YAC clones producing YAC contigs spanning 40.8 Mb or ~85% of the short arm of chromosome 5. (2) Direct screening of the entire mega-YAC library (Methods) was then conducted in regions with a limited number of YACs from the initial nucleation, including STSs from the ends of contigs. (3) Next, an additional 132 STSs generated by other genome-wide mapping efforts [Whitehead, Genethon, Stanford, and Cooperative Human Linkage Center (CHLC)] were tested independently by our laboratory against our YAC contigs. Again, STSs from regions of low coverage were used to identify additional YACs by screening the entire mega-YAC library. At this point, 216 STSs and 436 YACs had been unambiguously localized to 5p,

⁵Corresponding author.
E-MAIL dgrady@uci.edu; FAX (949) 824-3403.

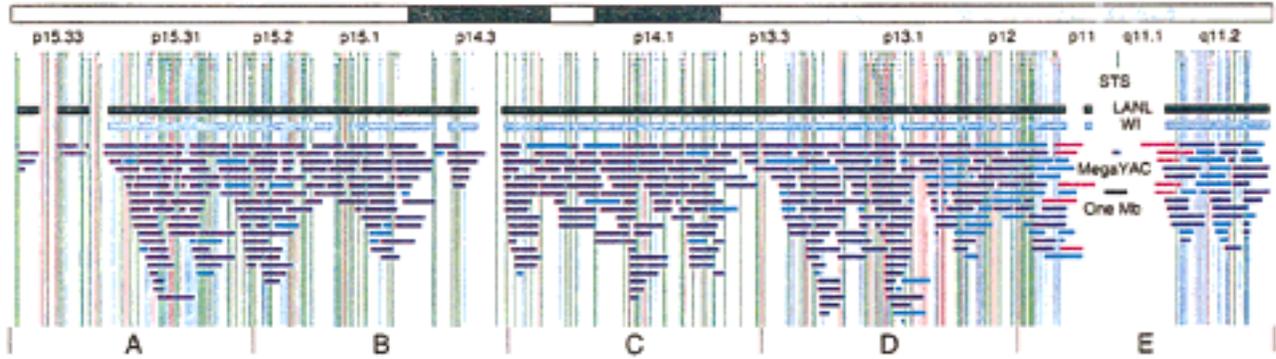


Figure 1 Overview of the physical map of the short arm of human chromosome 5. The overall YAC coverage reported in this study (black bar) and that of the Whitehead Institute (turquoise bar) is summarized with respect to the chromosome 5p ideogram. The narrow black and gray bars represent individual YACs that have been mapped and are shown in detail in Fig. 2. Labels A–E indicate the page of Fig. 2 on which the detailed information for that fraction of the map can be found.

covering ~44 Mb or 92% of the region. (4) Finally, an additional 288 STSs and 116 YACs mapped by other laboratories and not tested independently by our laboratory were incorporated into the overall map with SIGMA (System for Integrated Genome Map Assembly; Doggett et al. 1995). This integration did not significantly improve the YAC coverage, but helped add to and refine marker order. The final 5p map, then, represents an independently constructed YAC/STS map that experimentally extends, integrates, and confirms maps generated by the genome-wide mapping efforts (Chumakov et al. 1995; Hudson et al. 1995). We believe it is one of the first maps to attempt such an integration.

The Map

The map consists of four separate contigs totaling 45 Mb, crosses the centromere, and extends into the long

arm for ~5 Mb (Figs. 1, 2). The SIGMA depiction of the data was assembled in a manner similar to that described by Doggett et al. (1995). Briefly, all clones, STSs, somatic cell hybrids, etc., are entered into SIGMA as elements in the map and the relationships between these elements (contains, overlaps, left of, and so on), are entered as objectives. Elements are then moved manually in the map to minimize unfulfilled objectives. The SIGMA representation integrates the ideogram for 5p and the genetic map with the STS markers and YACs used to construct the physical map.

The map contains 522 YACs and 504 STSs (Fig. 2). There are an average of 4.6 STSs per YAC and 6.5 YACs per STS. Whereas all 522 YACs were used to construct the map, only the positions of 461 YACs (84%), that best represent the contigs, are depicted in Figure 2. In general, YACs known to be chimeric (either by in situ hybridization or STS content from other chromo-

Figure 2 Integrated physical map of human chromosome 5p. The map is represented on 10 consecutive pages and includes all of the short arm (p) of chromosome 5, the centromeric region, and the proximal 4.5 Mb of the long arm (q). The map is arranged from *top* to *bottom* in tiers depending on the type of information that is provided. The uppermost tier represents the genetic map of human chromosome 5p, with the genetic markers (D5S no.) indicated along the map in a centimorgan (cM) scale. Lines connecting the genetic positions of the markers and physical positions indicate the correspondence between the two maps. The scale for the physical map is 3 cm = 1 Mb of DNA. The next tier depicts the ideogram of the chromosome. Below the ideogram, the somatic cell hybrid breakpoints are shown with orange vertical lines continuing through the remainder of the map and arrows to indicate the region of the chromosome represented in a given somatic cell hybrid. Breakpoints that have been anchored cytogenetically are indicated by registration lines extending upward to specific cytogenetic bands. The bin names (pA–qA) delineated by the breakpoints are shown in the next tier. The level below depicts the STSs and ESTs found here for the chromosome. Registration lines for these markers are color coded to represent anonymous markers (green), ESTs (red), and genes (violet), tested in this study, and integrated STSs tested by the Whitehead Institute (blue). Below the STS tier is a black bar summarizing YAC coverage. Gaps in this bar indicate potential discontinuities in mega-YAC coverage. The turquoise bars below the low resolution coverage are an overlay of the Whitehead YAC contig coverage (Hudson et al. 1995). Beneath the Whitehead contigs are mega YACs (names beginning with My) and a few smaller YACs made by collaborating labs. The CEPH mega YACs are drawn to scale according to their size reported by CEPH or determined by our laboratory. YACs of unknown length were sized according to their STS coverage. Dark blue, light blue, and violet red are used to indicate YACs tested for this study, Whitehead contig YACs untested for this study, but integrated into the SIGMA map, and tested YACs that span the centromere, respectively. The STS content data generated here for each YAC is available from the authors. Reported disease regions and genes located through Online Mendelian Inheritance in Man (OMIM), indicated by lime green bars, are depicted in the level below the YACs. The next tier shows genes (green), pseudo or homologous genes (orange), and cDNAs (turquoise), which have been localized minimally to the cytogenetic map. The horizontal length of the line corresponds to the ambiguity of the element's location; some were localized cytogenetically whereas others were confined to YAC members containing the sequence. Registration lines from markers intersect genes in which they are contained. Those genes we localized more precisely by PCR for this study are indicated by purple registration lines. The next tier depicts those genes that have also been mapped to mouse chromosomes. The corresponding mouse syntenic loci, mouse syntenic chromosome regions, and mouse cM location are shown in the bottom tier of the integrated map.

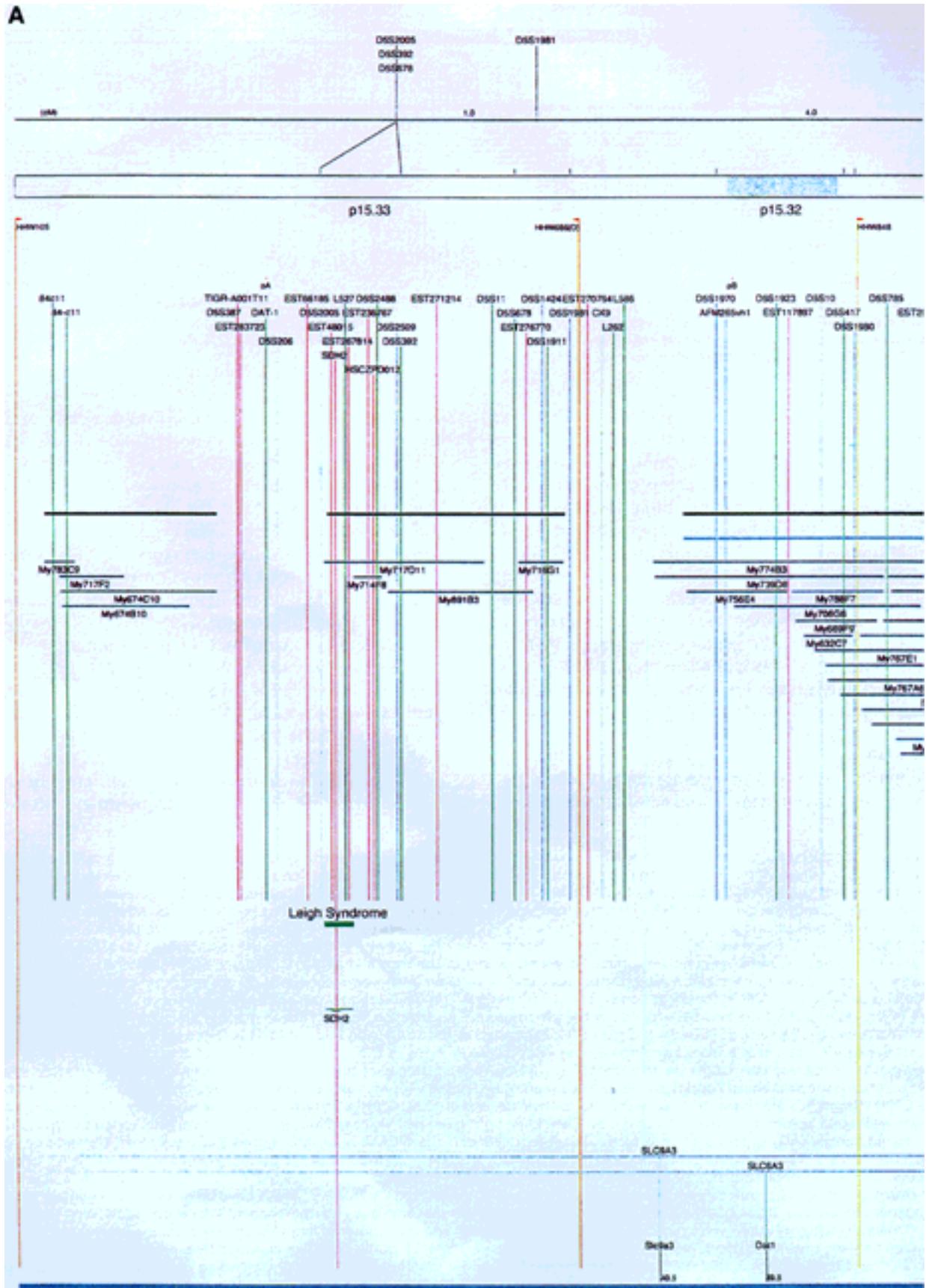
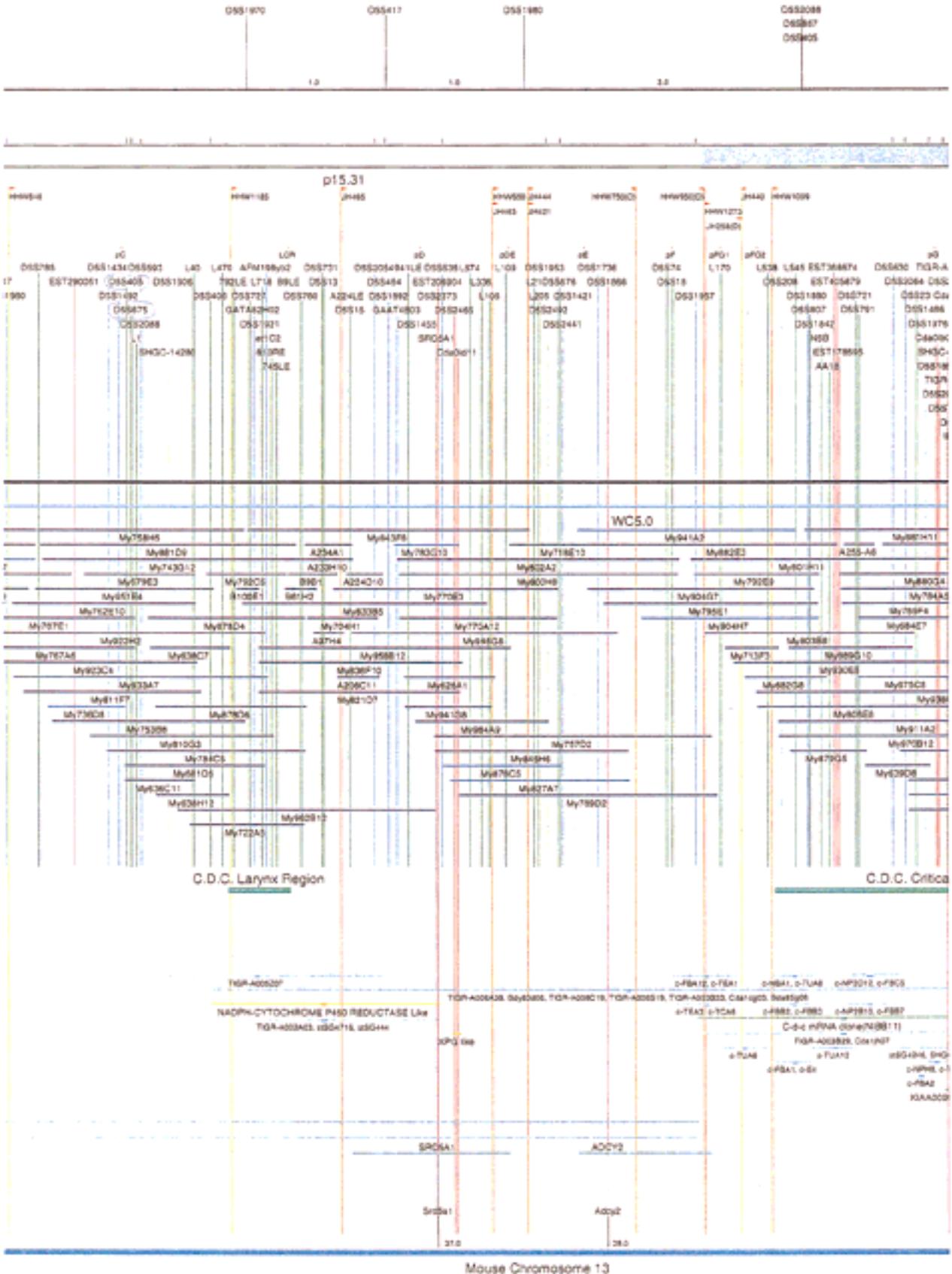
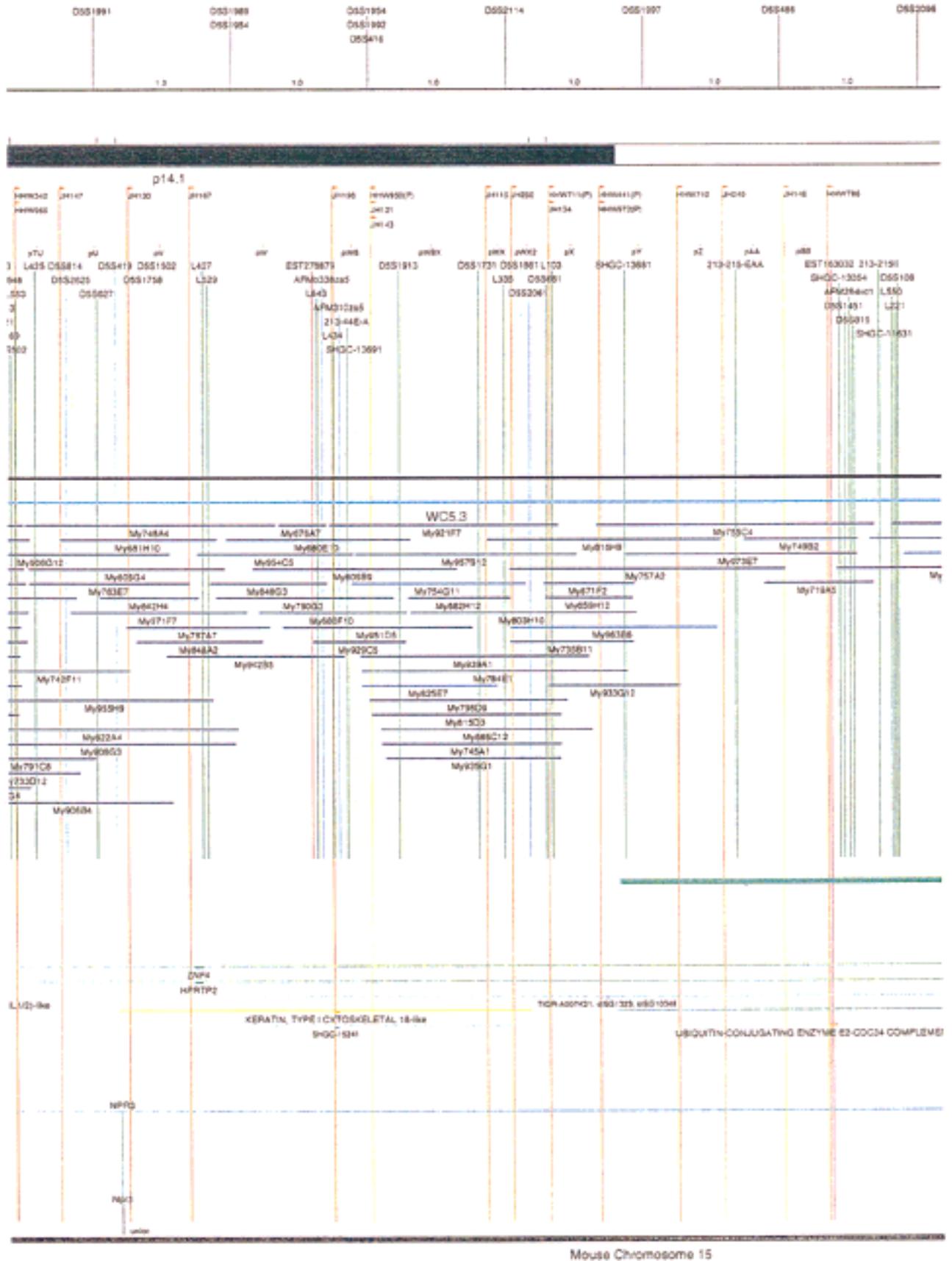
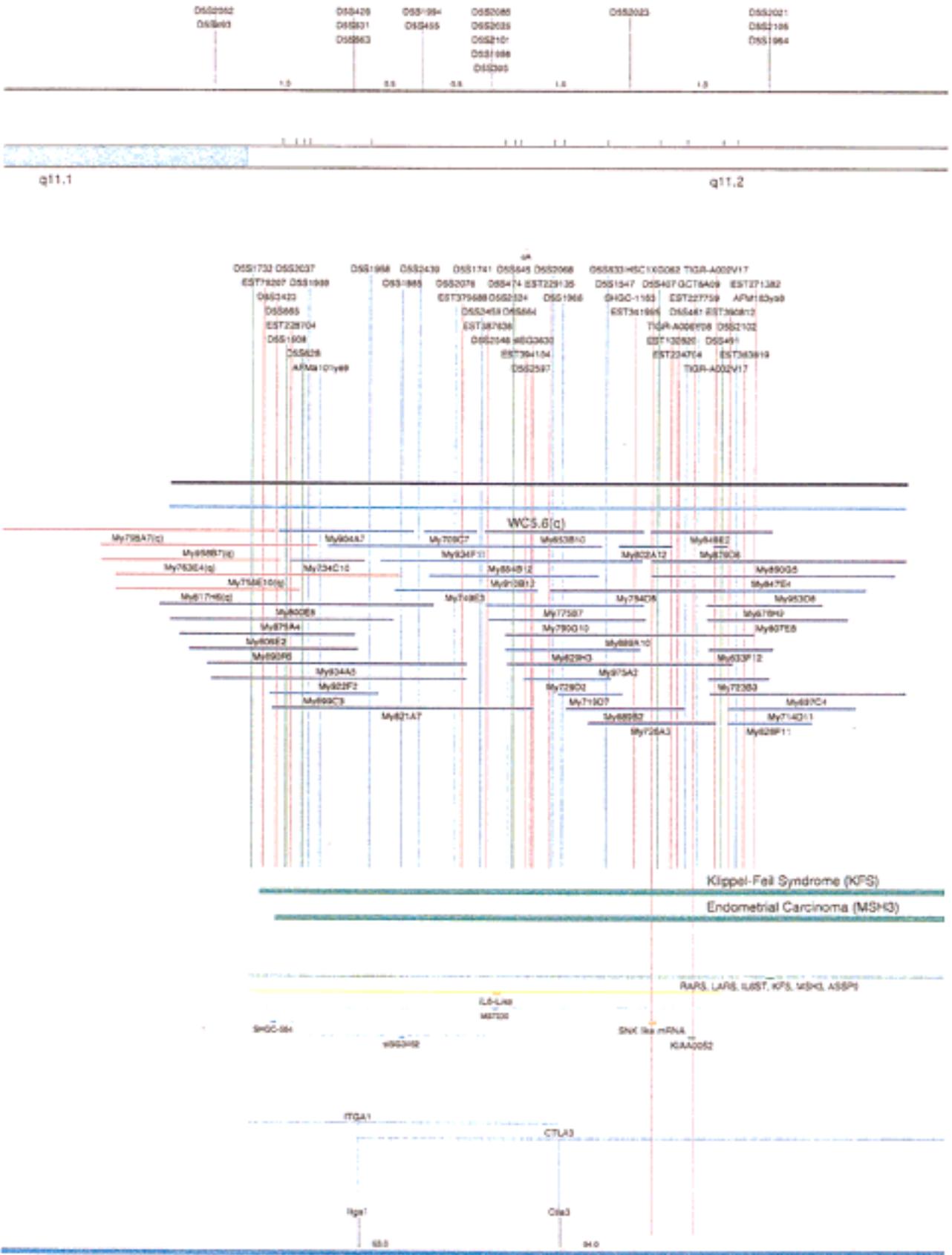


Figure 2 (See preceding page for legend.)









somes) are not shown on Figure 2 unless they added unique information. As indicated in Figure 2, the SIGMA map includes 382 YACs localized independently by our laboratory and 79 YACs localized by Hudson et al. (1995).

The scale of the map was determined by an estimate of 48 Mb for the short arm of chromosome 5 (194 Mb total, Morton 1991; Grady et al. 1996). The distance between markers and contig length was estimated with the constraints of YAC lengths either reported by Genethon, or for a small subset (~10%), independently determined by our laboratory. SIGMA and SEGMAP (Green and Green 1991) were then used to optimize marker/clone order/distance as described (Doggett et al. 1995). Chromosome breakpoints for a set of somatic cell hybrids containing 5p deletions were used to define bins such that the physical map could be anchored to the cytogenetic map (Overhauser et al. 1994; Gersh et al. 1995). The position of the centromere was identified by a number of criteria. First, YACs considered to reside close to the centromere were localized by in situ hybridization (Riethman et al. 1989; McCormick et al. 1993). YAC clones 924H10 and 889A10 clearly localized noncentromerically to the p and q arm, respectively. YAC 748C1 hybridized on the p arm close to the centromere (data not shown). The chromosome 5 centromeric region has been estimated to contain ~3 Mb of α -satellite and satellite II/III DNA (Moyzis et al. 1987; Meyne et al. 1989; Grady et al. 1992; data not shown). This is consistent with the standard cytogenetic ideogram length of the centromere shown in Figures 1 and 2. No YACs isolated for this study contained large blocks of centromeric α -satellite or satellite II/III repeat DNA. Five YACs, however, contained STS markers from both the p and q arms, and hybridized in situ to the centromeric region (Fig. 2). None of these YACs are long enough to bridge the presumed size of the centromere. The independent isolation of 5 different YACs containing both p and q arm proximal markers suggests that rather than being chimeras, either (a) they originated as much larger YACs that deleted portions of centromeric DNA, or (b) there exists a duplication of the pericentric region of chromosome 5 at the p and q junctions. Although it has been reported that both α -satellite and satellite II/III DNA are difficult to maintain in both yeast and *Escherichia coli* cells (Neil et al. 1990; Peterson et al. 1996), the probability that the pericentric region of chromosome 5 contains duplicated material is high (see below).

The telomeric end of the map is less well defined, because a telomeric half YAC for 5p has not been isolated yet, despite numerous attempts (Riethman et al. 1989; data not shown). The telomeric regions of most human chromosomes are under-represented in the mega-YAC library, presumably due to the well-known

end effects inherent in the production of size-selected recombinant DNA libraries, and the GC richness of many of these regions (Doggett et al. 1995). The extensive somatic cell hybrid panel map of 5p, however, allows for a reasonable approximation of the location of the telomere (Figs. 1,2; Overhauser et al. 1994; Gersh et al. 1995).

Map Construction and Integration

The initial STS marker/YAC order was experimentally and computationally determined in our laboratory by analyzing the STS content data for 216 STSs and 436 YACs. These YACs (dark blue) and STSs (green) are indicated in Figures 1 and 2 (see overview of the project, steps 1–3 above). The pictured STS order assumed YAC deletion was minimal and that consistency with the somatic cell hybrid and genetic map order should be maintained (Doggett et al. 1995). In the small number of discrepancies observed, it was assumed that marker order obtained by STS/YAC mapping was correct, and this is the order indicated on the SIGMA map (Fig. 2). Attempts at gap closure by rescreening the entire mega YAC library with STS markers residing at the ends of the contigs yielded no new clones. At this stage of the map, ~92% (44 Mb) of chromosome 5p was covered in YAC/STS contigs.

To attempt an integration of our experimentally determined physical map with the whole-genome YAC/STS maps generated by other groups (Chumakov et al. 1995; Hudson et al. 1995), SIGMA was used to computationally compare results. Information from the Genethon and Whitehead whole genome projects could be compared directly, because the same recombinant DNA library was used in these studies (Chumakov et al. 1995; Hudson et al. 1995). The Genethon map of 5p, on the basis of genetic markers, contains 338 YACs anchored by 58 STSs in 1 of 29 bins (Chumakov et al. 1995). Similar to what was described by Doggett et al. (1995), we were able to confirm much of the Genethon level 1 data obtained by STS content mapping, but found the YACs related by overlaps in levels 2 and above to be less likely to be derived from chromosome 5p. Only 34.3% of the Chumakov et al. (1995) 5p YACs are present on our integrated map (Fig. 2), mostly because of their tier 1 data.

The genome map reported by the Whitehead Institute group on the other hand, is very consistent with our map. No major discrepancies were observed between the two maps. Data in Hudson et al. (1995) (Data Release II) indicate that the part of their whole genome map representing 5p and the proximal portion of 5q includes 7 contigs, comprising 563 YACs and 190 STSs. Most of these clones are the same as those on our map, so a direct integration is possible. By use of SIGMA, YAC clones mapped in this study (dark blue in Fig. 2) that also were mapped by Hudson et al. (1995),

have had additional STS data integrated from that study (blue STSs in Fig. 2), without confirmation from our laboratory. Although our contig coverage near the telomere was more complete than the Whitehead mapping effort (Fig. 2), the Whitehead map near the centromere was more extensive (Hudson et al. 1995). Seventy-nine Whitehead-identified YACs, not identified by our study, are integrated in Figure 2 (light blue), using SIGMA to minimize discrepancies between the data sets.

Our initial map of 5p contained 436 YACs anchored with 216 STSs, and represented 92% (44 Mb) of chromosome 5p (Fig. 2). The map of Hudson et al. (1995) contained 563 YACs anchored with 190 STSs, and represented 86% (41 Mb) of 5p. By comparing and combining these independently derived physical maps, a map containing 552 YACs and 504 STSs was obtained (Fig. 2). The four contigs (Fig. 1) span between 1 and 25 Mb of DNA and represent ~94% of chromosome 5p (Fig. 1). Four gaps, defined as regions not believed to be represented by YAC DNA, appear to exist at the telomeric end, within Bin A, between Bins A and B, and within Bin R. These gaps are estimated to contain <4 Mb of unmapped DNA. An additional contig of ~5 Mb represents the centromere proximal region of the q arm (Figs. 1,2).

Map Completeness

The localization of 504 STS markers along an estimated 48 Mb of human chromosome 5p DNA, for an average STS spacing of 95 Kb, has met one of the initial goals of the Human Genome Project (Botstein et al. 1990). The completeness of the map can also be assessed by a completeness at resolution graph (Olson and Green 1993; Doggett et al. 1995) (Fig. 3). This measure conveys the fraction of the chromosome that is found within a certain distance of a discernible point on the map, including YAC ends, markers, and breakpoints. Because the SIGMA map contains 504 markers, 922 YAC ends, and 56 breakpoint bins, a total of 1482 discrete elements can be used to position newly derived 5p sequences. The graph indicates that the completeness of the STS/YAC end/breakpoint map, at a resolution of 100 kb, is 0.81, or that 81% of the chromosome is within 50 kb of one of these mapped elements. The completeness of the STS map alone at a resolution of 100 kb is 0.56. The discrepancy between the curves representing the predicted and actual values exists because the markers are not randomly distributed, many being generated by targeted approaches (i.e., Simmons et al. 1995), whereas the predicted curve is based on a random distribution.

Integration of Genes, ESTs, and Mouse Homologs

From the hundreds of gene markers and ESTs mapped to 5p by other laboratories (Schuler et al. 1996; De-

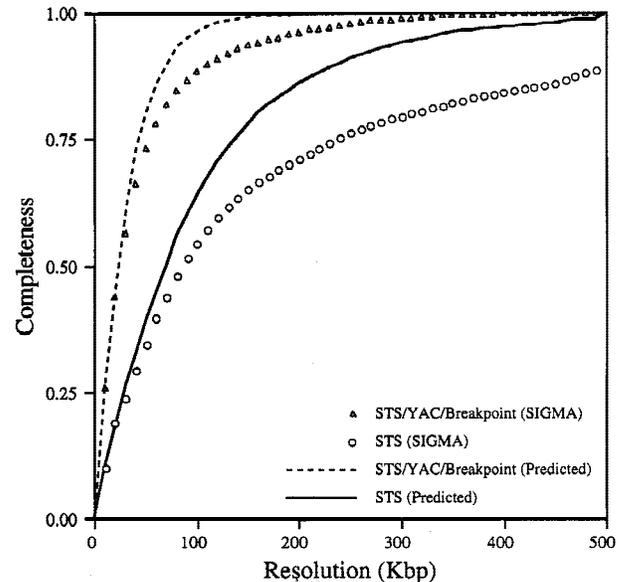


Figure 3 Completeness at resolution graph for the YAC/STS/Hybrid breakpoint (BP) map. The location of each STS, YAC end, and hybrid breakpoint (data points) was extracted from the SIGMA map. A window of given size was moved that distance down the chromosome. For each interval, it was determined whether any data points were contained within the window. The result of different window sizes was evaluated at regular intervals to create the plot. (▲) The curve generated by use of STSs, YAC ends, and BPs; (○) the curve for the STSs alone. The predicted completeness at resolution curves (broken line for STS/YAC/BP and solid line for STS alone) were obtained by randomly placing the same number of sites on the map.

loukas et al. 1998), Figure 2 shows a subset of 53 gene sequences and disease regions useful as landmarks. In most cases, this information was collected from the National Center for Biotechnology Information (NCBI), Genbank, and the Genome Database (GDB). Of the identified genes that had been reported previously to exist on 5p by cytogenetic analysis, 15 have been localized more precisely to YAC members of our contig by PCR with primers directed to the gene sequence (Fig. 2; Methods). Because the mouse homologs for many of the genes found on 5p have been mapped (Mouse Genome Database), the syntenic chromosome information is also displayed in Figure 2. BLAST searches with marker sequences permitted us to identify additional elements, including putative pseudo-genes and other regions displaying homology to reported sequences. Markers fitting this description are indicated in Figure 2. Finally, anonymous cDNAs captured from the cri-du-chat critical region (Simmons et al. 1995, 1997) are shown as well.

Comparison with the Genetic and Radiation Hybrid Maps

A direct comparison to the 5p Genetic map can be made, because numerous markers are present on both

maps (Gersh et al. 1994; Dib et al. 1996), and SIGMA allows us to estimate physical distances directly. Interestingly, the physical map differs significantly from the genetic map. Whereas the short arm of chromosome 5 is ~25% of the total chromosome on the basis of its cytogenetic size, it accounts for about one-third of the genetic distance. Therefore, there is more recombination on the short arm than would be expected for its relative size. Furthermore, by plotting genetic versus physical distance over the length of the map (Fig. 4), local recombinationally inert and active regions are revealed. Relatively inactive regions include the 5p telomere, centromere, the portion of 5p immediately proximal to the centromere, and dark band regions including p13.1, p14.1, and p14.3. The remainder of the chromosome, including the mapped material on the other side of the centromere, is fairly recombinationally active.

A direct comparison with radiation hybrid maps of 5p is also possible by use of markers shared by both studies (McPherson et al. 1997). Whereas the experimentally determined linear order of markers is, in general, the same between the two maps, the number of markers on the physical map (Fig. 2) is nearly an order of magnitude greater. Like the genetic map, there is a

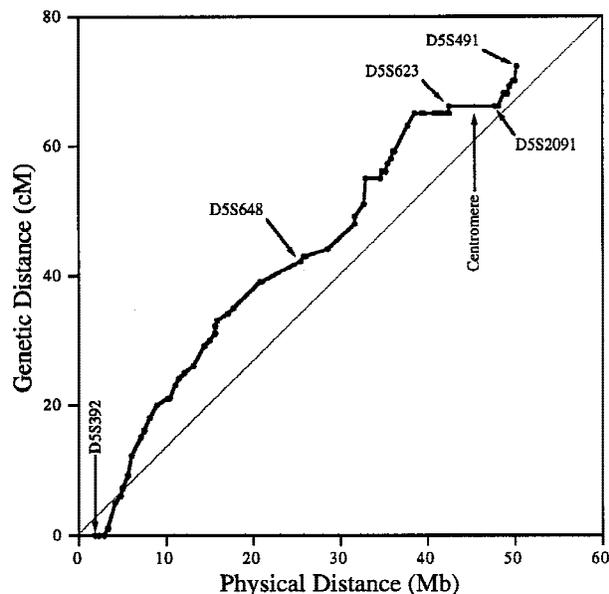


Figure 4 Correlation between genetic and physical distances for genetic markers. Genetic distances between markers, taken from Gersh et al. (1994), Murray et al. (1994), and Dib et al. (1996) for chromosome 5p, are shown in the first tier of Fig. 2. Physical distances between markers were determined from the location of these markers in the SIGMA map. All genetic markers were localized to the mega-YAC contigs, whereas a subset were localized to breakpoint intervals. Genetic markers D5S1996, D5S631, D5S430, and D5S665 were removed from the plot because of ordering discrepancies with respect to the physical map. Selected genetic markers and the location of the centromere are indicated on the graph for reference.

greater than fivefold (14.8–81.5 cR/Mb observed, McPherson et al. 1997) difference in the ratio of cR/Mb between various mapped markers, making direct comparisons imprecise. In the few cases in which radiation hybrid mapping could split markers that physical mapping could not (i.e., 675,405,5pC; 502,648 5pI; 663,426 5pM), this information did not assist in the unambiguous definition of the STS order.

Map Quality

Whereas YACs containing long stretches of contiguous DNA are very useful for generating first-generation physical maps, contributing critical contig information, the integrity of the YAC DNA is not always strictly maintained. It is clear by close examination of the map that some YACs that have STSs missing have likely undergone deletions or contain DNA from non-contiguous regions of the genome. According to available data, ~90 of the YACs reported on this map have been observed to contain DNA (i.e., STSs) from chromosomes other than 5; this chimera fraction is calculated to be 33%, similar to previous reports (Chumakov et al. 1995; Doggett et al. 1995; Hudson et al. 1995).

The mega-YAC physical map of 5p appears to be nearly complete. It has been estimated that ~2% of the genome will not be represented among the mega-YAC library (Chumakov et al. 1995). Because 34,560 YAC clones from the Centre d'Etude du Polymorphisme Humain (CEPH) Mark II and mega-YAC libraries (10X coverage) were used to generate the Genethon YAC contig map of the human genome (Chumakov et al. 1995), it is estimated that ~520 of these YACs (1.5%) map to 5p. Because our map of 5p, resulting from the collective effort of several laboratories, includes 510 CEPH mega YACs (with an additional 42 on the proximal portion of 5q), nearly all of the Mark II and mega YACs expected to reside on 5p have been tested and most of them have been localized on the integrated map.

Map Peculiarities

There are several notable features of the 5p map that suggest rearrangements have occurred within the DNA close to the centromere. Because one would expect a fairly even distribution of YACs to cover the chromosome, the relatively high number of YACs localized to the proximal region of 5p suggests that this region of the chromosome may represent a tandem duplication. However, *in situ* hybridization of a small sample of YAC clones from this region to metaphase chromosome spreads gives no indication of a large tandem duplication (data not shown). We have drawn the map assuming this region is not tandemly duplicated, but, it remains a formal possibility that should be considered. Another peculiarity is that some markers found 10–20 Mb apart on opposite sides of the centromere (i.e., L88 and D5S491) appear to be present together on

multiple YACs. Furthermore, the order of several genes in this region is rearranged with respect to the homologous genes clustered on mouse chromosome 15. This also suggests that there is duplicated material close to the centromere. Interestingly, two independent laboratories have identified gene sequences (*ZNF4*–Habeebu et al. 1989, *PMCLH*–Pedeutour et al. 1994) that hybridize to both sides of the centromere. A duplication and rearrangement scheme close to the centromere has been proposed to occur on chromosome 5 by others (Pedeutour et al. 1994; Sargent et al. 1994) as well as on other chromosomes (Tunnacliffe et al. 1993; Eichler 1999). Only further high-resolution mapping and sequencing will resolve the organization in this region.

The map represented here is our current view of the chromosome, considering all of the available data. Errors in maps this complex are perhaps inevitable, owing to the presence of low abundance repeats, chimeric clones, deleted clones, and chromosomal duplications. We would like to hear of any potential errors in this map and any relevant data that could refine it. Correspondence concerning the integrated 5p map can be sent electronically to dgrady@uci.edu.

METHODS

YAC Contig Construction

DNA from microtiter plates (372) of the CEPH mega-YAC library were pooled into 672 pools from an array of 11X17X2 plates by a row and column-pooling scheme. From adjacent pools of the original 672, 95 primary level pools, each containing 6 columns (column half-plates), or 8 rows (row plates), were created. In this design, a positive for one YAC in the library should result in five positive pools (three rows and two columns) at the primary level of screening. Primary through tertiary level screens with chromosome 5p STSs were conducted as described previously (Doggett et al. 1995).

A sublibrary of YACs (tier 1 to tier 3) previously localized to chromosome 5p (Cohen et al. 1993; Chumakov et al. 1995) was also assembled for STS screening. To provide contig nucleation points, 84 STSs generated from flow-sorted chromosome 5 DNA were tested against this pool. All 84 STSs had been localized previously to 1 of 51 bins (average size 1 Mb) on 5p, by somatic cell breakpoint analysis (Overhauser et al. 1994; Grady et al. 1996; Gersh et al. 1997).

Two primer sets (labeled 84c11 and 84-c11) were derived from the cosmid 84c11 (Overhauser et al. 1994) sequence to generate additional contig information near the telomeric end. These primers are as follows: 84-c11: 5'-TGGTCTAGTAATCGTTCAGGATTTC-3' and 5'-GGAATCACTATTCTCACTTCTG-3' (product 285 bp); and 84c11: 5'-ACGCATGTGTCAGTACTAATTTTGG-3' and 5'-GGGTGCAAACCTTGT-TATAAAG-3' (product 109 bp). Additional STSs were identified from previously mapped λ -phage clones D5S19, D5S32, D5S712, and D5S730 (Overhauser et al. 1994). Single-copy DNA fragments from each λ -phage clone were sequenced and PCR primers were designed. The PCR primers used are as follows: D5S19 (213-44E-A), 5'-ATGTGACCTGACCTTGGACA-3' and 5'-GGATCATCAGTCTCAATAGC-3' (PCR product 340

bp); D5S32 (213-92H-A), 5'-TTGTCAACAGCAGCGACTCA-3' and 5'-GAGCTGTTGCGTGTTAATTC-3' (PCR product 270 bp); D5S712 (D5S1521), 5'-AGCAGCAGCCGTTAGTGTAG-3' and 5'-GGACGATAGACTTAGAAGTC-3' (PCR product 192 bp); D5S730 (213-215-EAA), 5'-TCAAGAAGACTACCA-CATGA-3' and 5'-CTCTCTTATTCCAAGATTC-3'. An additional 132 STSs localized to 5p and generated by other genome-wide mapping efforts (Whitehead, Genethon, Stanford, and CHLC) were synthesized and tested independently by our laboratory against our nucleated YAC contigs. STSs from regions of low coverage were used to identify additional YACs by screening the entire mega-YAC library. YAC/STS maps were constructed as described (Doggett et al. 1995), with SIGMA. SIGMA allows sized DNA segments to be placed on a map with discrete objectives (relationships) to other mapped objects. Once all of the objects are placed on the map and all of the possible objectives are met, a picture of that region will be graphically represented. Because SIGMA uses scaled objects, coverage of the chromosome region is easily calculated. SIGMA was also used to computationally compare our mapping results with the physical maps generated by Hudson et al. (1995) and Chumakov et al. (1995). All STS markers used for map assembly in this study have been published (Overhauser et al. 1994; Grady et al. 1996; Gersh et al. 1997; Dib et al. 1996; Gyapay et al. 1996; Hudson et al. 1995, this work), and most can be accessed from the appropriate website (below).

Gene and mRNA Mapping

The indicated genes, putative genes, and sequence homologies were localized to YAC members of our contig in the following way. Primer sets directed to genes were either found in the literature or designed from the published gene sequence. These primers were synthesized and used to screen the relevant YAC contig region by PCR. The primer sets for retinal mRNA, TARS, HMGCS1, PMCHL1, LIFR, SKP2, IL7R, PTGER2, GLAST1, and ZNF131 were designed for this study and their sequences are as follows: retinal mRNA, 5'-TCAAGGTC-AAGCGTTAGTCG-3' and 5'-CTGTGAGTTTGTGGGGAAG-3' (product 99 bp); TARS, 5'-AGGAGGAGTTTGGCAAAGTC-3' and 5'-CTCCACATTACACTAAAGC-3' (product 185 bp); HMGCS1 5'-ATATTCCAAGCCCTGCCAAG-3' and 5'-TGA-AGTCTTGCACCTCACAG-3' (product 115 bp); PMCHL1, 5'-ATGGTCTGTCACTGAATCTG-3' and 5'-AGCTGAGTTTTCTTCATCCC-3' (product 137 bp); LIFR, 5'-TAACAGTGCACCGTGTAC-3' and 5'-ATACTTCACAGGATCCCTCC-3' (product 103 bp); SKP2, 5'-CAGGGA-AAATAGGCAGGAAG-3' and 5'-TGGGCTCCCTAGT-ATACTTG-3' (product 101 bp); IL7R, 5'-TGACCACTCTCCTGAGTTC-3' and 5'-AGCTGAATCATTTGGGTCACC-3' (product 105 bp); PTGER2, 5'-GCGAGTAAGAAATCCAGCAC-3' and 5'-CAGTGGCTGTCGACTTTTGG-3' (product 115 bp); GLAST1, 5'-CTTAAAGTGCCCATCCAGG-3' and 5'-TCAAA-GAATCTCTCAGGGC-3' (product 222 bp); and ZNF131, 5'-CACTGAGTCATTGCCATCTG-3' and 5'-TGTACTTTACGGACTGCTTGG-3' (product 134 bp). Mouse syntenic genes were identified through GDB and NCBI.

Computer Resources

Four computer resources were invaluable to this process, internet resources, relational database processing, laboratory tracking, and SIGMA map assembly. World Wide Web (WWW) access to external laboratories facilitated the collection of large amounts of recently generated data. The primary

WWW sites that provided mapping data were the Whitehead Institute, CEPH/Genethon, GDB, and NCBI. A relational database was used to manage, analyze, assimilate, and report on the many available data sets. Laboratory tracking used a 4D database to record all of the experimental results and multiple Excel spread sheets to build an STS/YAC matrix.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Michelle Valdez, who contributed to this work and died tragically before it's completion. We appreciate the technical support of L. Sue Thompson for DNA oligomer synthesis, Lynne Goodwin, Linda Meincke, and Judy Tesmer for library screening, Evelyn Campbell for in situ hybridization, and Norman Doggett for comments on the manuscript and assistance with figure preparation. Funded by the United States Department of Energy under contract W-7405-ENG-36 and Grant DE-FG03-97ER62485.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Ashworth, L.K., M.A. Batzer, B. Brandriff, E. Branscomb, P. deJong, E. Garcia, J.A. Ganes, L.A. Gordon, J.E. Lamerdin, G. Lennon et al. 1995. An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11**: 422-427.

Bell, C.J., M.L. Budarf, B.W. Nieuwenhuisen, B.L. Barnoski, K.H. Buetow, K. Campbell, A.M. Colbert, J. Collins, M. Daly, P.R. Desjardins et al. 1995. Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers. *Hum. Mol. Genet.* **4**: 59-69.

Botstein, D., C. Cantor, J. Carbonell, A. Carrano, T. Caskey, F. Collins, U. Francke, E. Lander, L. Lerman, R. Moyzis et al. 1990. Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years. FY 1991-1995 No. DOE/ER-0452P, NIH No. 90-1590.

Bouffard, G.G., J.R. Idol, V.V. Braden, L.M. Iyer, A.F. Cunningham, L.A. Weintraub, J.W. Touchman, R.M. Mohr-Tidwell, D.C. Peluso, R.S. Fulton et al. 1997. A physical map of human chromosome 7: An integrated YAC contig map with average STS spacing 79 kb. *Genome Res.* **7**: 673-692.

Centre d'Etude du Polymorphisme Humain (CEPH)-Genethon: <http://www.cephb.fr/ceph-genethon-map.html>

Chumakov, I.M., I. Le Gall, A. Billault, P. Ougen, P. Soularue, S. Guillou, P. Rigault, H. Bui, M.F. De Tand, E. Barillot et al. 1992. Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nat. Genet.* **1**: 222-225.

Chumakov, I.M., P. Rigault, I. Le Gall, C. Bellanne-Chantelot, A. Billault, S. Guillou, P. Soularue, G. Guasconi, E. Poullier, I. Gros et al. 1995. A YAC contig map of the human genome. *Nature* **377**: (Suppl.)175-297.

Church, D.M., U. Bengtsson, K.V. Nielsen, J.J. Wasmuth, and E. Niebuhr. 1995. Molecular definition of deletions of different segments of distal 5p that result in distinct phenotypic features. *Am. J. Hum. Genet.* **56**: 1162-1172.

Cohen, D., I. Chumakov, and J. Weissenbach. 1993. A first-generation physical map of the human genome. *Nature* **366**: 698-701.

Collins, J.E., C.G. Cole, L.J. Smink, C.L. Garrett, M.A. Leversha, C.A. Soderlund, G.L. Maslen, L.A. Everett, K.M. Rice, A.J. Coffey et al. 1995. A high-density YAC contig map of human chromosome 22. *Nature* **377**: (Suppl)367-379.

Collins, F.S., A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, L. Walters, and members of the DOE and NIH planning groups. 1998. New goals for the U.S. Human Genome Project:1998-2003.

Science **282**: 682-689.

Cooperative Human Linkage Center (CHLC): <http://www.chlc.org/HomePage.html>

Crollius, H.R., M.T. Ross, A. Grigoriev, C.J. Knights, E. Holloway, J. Misfud, K. Li, M. Playford, S.G. Gregory, S.J. Humphray et al. 1996. An integrated YAC map of the human X chromosome. *Genome Res.* **6**: 943-955.

Deloukas, P., G.D. Schuler, G. Gyapay, E.M. Beasley, G. Soderland, P. Rodriguez-Tome, L. Hui et al. 1998. A physical map of 30,000 human genes. *Science* **282**: 744-746.

Dib, C., S. Faure, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.

Doggett, N.A., L.A. Goodwin, J.G. Tesmer, L.J. Meincke, D.C. Bruce, L.M. Clark, M.R. Altherr, A.A. Ford, H.C. Chi, and B.L. Marrone et al. 1995. An integrated physical map of human chromosome 16. *Nature* **377**: (Suppl)335-365.

Eichler, E.E. 1999. Repetitive conundrums of centromere structure and function. *Hum. Mol. Genet.* **8**: 151-155.

Footo, S., D. Vollrath, A. Hilton, and D.C. Page. 1992. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. *Science* **258**: 60-66.

Gemmill, R.M., I. Chumakov, P. Scott, B. Waggoner, P. Rigault, J. Cypser, Q. Chen, J. Weissenbach, K. Gardiner, and H. Wang et al. 1995. A second-generation YAC contig map of human chromosome 3. *Nature* **377**: (Suppl)299-319.

GenBank Database Query Form, National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/genbank/query_form.html

Genome Data Base (GDB), University School of Medicine, Johns Hopkins University, Baltimore, MD, 1996.: <http://gdbwww.gdb.org/>

Gersh, M., S.A. Goodart, and J. Overhauser. 1994. Physical mapping of genetic markers on the short arm of chromosome 5. *Genomics* **24**: 577-579.

Gersh, M., S.A. Goodart, L.M. Pasztor, D.J. Harris, L. Weiss, and J. Overhauser. 1995. Evidence for a distinct region causing a cat-like cry in patients with 5p deletions. *Am. J. Hum. Genet.* **56**: 1404-1410.

Gersh, M., D. Grady, K. Rojas, M. Lovett, R. Moyzis, and J. Overhauser. 1997. Development of diagnostic tools for the analysis of 5p deletions using interphase FISH. *Cytogenet. Cell Genet.* **77**: 246-251.

Goodart, S.A., A.D. Simmons, D. Grady, K. Rojas, R.K. Moyzis, M. Lovett, and J. Overhauser. 1994. A yeast artificial chromosome contig of the critical region for cri-du-chat syndrome. *Genomics* **24**: 63-68.

Grady, D.L., D.L. Robinson, M. Gersh, E. Nickerson, J. McPherson, J.J. Wasmuth, J. Overhauser, L.L. Deaven, and R.K. Moyzis. 1996. The generation and regional localization of 303 new chromosome 5 sequence-tagged sites. *Genomics* **32**: 91-96.

Green, E.D. and P. Green 1991. Sequence-tagged site (STS) content mapping of human chromosomes: Theoretical considerations and early experiences. *PCR Methods Applic.* **1**: 77-90.

Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillet, D. Muselet, J.F. Prud'Homme, C. Dib, C. Auffray et al. 1996. A radiation hybrid map of the human genome. *Hum. Mol. Gen.* **5**: 339-346.

Habeebu, S.S.M., J.E. Gibson, N.A. Affara, and M.A. Ferguson-Smith. 1989. Localization of two zinc finger protein genes to (a) two loci on chromosome 5 at 5p13-p14 and 5q12-q13 and to (b) the long arm of the X at Xq13-q21. *Cytogenet. Cell Genet.* **51**: 1009.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu et al. 1995. An STS-based map of the human genome. *Science* **270**: 1945-1954. with supplementary data from the Whitehead Institute/MIT Center for Genome Research, Human Genetic Mapping Project, Data Release 11 (October 1996).

Krauter, K., K. Montgomery, S.J. Yoon, J. LeBlanc-Straceski, B. Renault, I. Marondel, V. Herdman, L. Cupelli, A. Banks, J.

- Lieman et al. 1995. A second-generation YAC contig map of human chromosome 12. *Nature* **377**: (Suppl)321–333.
- Lejuene, J., J. Lafourcade, R. Berger, J. Vialatte, M. Boeswillwald, P. Seringe, and R. Turpin. 1963. Tois cas de deletion partielle du bras court d'un chromosome 5. *C.R. Acad. Sci. (Paris)* **257**: 3098–3102.
- McCormick, M.K., A. Buckler, W. Bruno, E. Campbell, K. Shera, D. Torney, L. Deaven, and R. Moyzis. 1993. Construction and characterization of a YAC library with a low frequency of chimeric clones from flow-sorted human chromosome 9. *Genomics* **18**: 5538.
- McPherson, J.D., B. Apostol, C.B. Wagner-McPherson, S. Hakim, R. Del Mastro, N. Aziz, E. Baer, G. Gonzales, M.C. Krane, R. Markovich, P. Masny et al. 1997. A radiation hybrid map of human chromosome 5 with integration of cytogenetic, genetic and transcript maps. *Genome Res.* **7**: 897–909.
- Meyne, J., L.G. Littlefield, and R.K. Moyzis. 1989. Labeling of human centromeres using an alphoid DNA consensus sequence: Application to the scoring of chromosome aberrations. *Mut. Res.* **226**: 75–79.
- Morton, N.E. 1991. Parameters of the human genome. *Proc. Natl. Acad. Sci.* **88**: 7474–7476.
- Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. World Wide Web (URL:<http://www.informatics.jax.org/>) (October 1996).
- Moyzis, R.K., K.L. Albright, M.F. Bartholdi, L.S. Cram, L.L. Deaven, C.E. Hildebrand, N.G. Joste, J.L. Longmire, J. Meyne, and T. Schwartzacher-Robinson. 1987. Human chromosome-specific repetitive DNA sequences: Novel markers for genetic analysis. *Chromosoma* **95**: 375–386.
- Murray, J.C., K.H. Buetow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk et al. 1994. A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science* **265**: 2049–2054.
- Nagaraja, R., S. MacMillan, J. Kere, C. Jones, S. Griffin, M. Schmatz, J. Terrell, M. Shomaker, C. Jermak, C. Hott et al. 1997. X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. *Genome Res.* **7**: 210–222.
- National Center for Biotechnology Information (NCBI), National Library of Medicine, Bethesda, MD, 1996.: <http://www.ncbi.nlm.nih.gov/>
- Neil, D., A. Villasante, R. Fisher, D. Vetrie, B. Cox, and C. Tyler-Smith. 1990. Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res.* **18**: 1421–1428.
- Niehbuh, E. 1978. The cri du chat syndrome. *Hum. Genet.* **44**: 227–275.
- Nizetic, D., L. Gellen, R.M. Hamvas, R. Mott, A. Grigoriev, R. Vatcheva, G. Zehetner, M.L. Yaspo, A. Dutriaux, C. Lopes et al. 1994. An integrated YAC-overlap and "cosmid-pocket" map of the human chromosome 21. *Hum. Mol. Genet.* **3**: 759–770.
- Olson, M.V. and P. Green. 1993. Criterion for the completeness of large-scale physical maps of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 349–355.
- Online Mendelian Inheritance in Man (OMIM), Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, <http://www3.ncbi.nlm.nih.gov/omim/>
- Overhauser, J., X. Huang, M. Gersh, W. Wilson, J. McMahon, U. Bengtsson, K. Rojas, M. Meyer, and J.J. Wasmuth. 1994. Molecular and phenotypic mapping of the short arm of chromosome 5: Sublocalization of the critical region for the cri-du-chat syndrome. *Hum. Mol. Genet.* **3**: 247–252.
- Pedeutour, F., C. Szpirer, and J.L. Nahon. 1994. Assignment of the human pro-melanin-concentrating hormone gene (PMCH) to chromosome 12q23-q24 and two variant genes (PMCH1 and PMCHL2) to chromosome 5p14 and 5q12-q13. *Genomics* **19**: 31–37.
- Peterson, E.T., D.L. Grady, and R.K. Moyzis. 1996. High frequency of mutational changes during the cloning of a human centromeric repeat. In *Biological structure and dynamics, proceedings of the ninth conversation, State University of New York. Albany, NY 1995*, (ed. R.H. Sarma and M.H. Sarma), pp. 209–215. Adenine Press, Albany, NY.
- Qin, S., N.J. Nowak, J. Zhang, S.N. Sait, P.G. Mayers, M.J. Higgins, Y. Cheng, L. Li, D.J. Munroe, D.S. Gerhard et al. 1996. A high-resolution physical map of human chromosome 11. *Proc. Natl. Acad. Sci.* **93**: 3149–3154.
- Riethman, H.C., R.K. Moyzis, J. Meyne, D.T. Burke, and M.V. Olson. 1989. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc. Natl. Acad. Sci.* **86**: 6240–6244.
- Sargent, C.A., I.J. Chalmers, M. Leversha, and N.A. Affara. 1994. A rearrangement on chromosome 5 of an expressed human beta-glucuronidase pseudogene. *Mamm. Genome* **5**: 791–796.
- Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome, A. Aggarwal, E. Bajorek et al. 1996. A gene map of the human genome. *Science* **274**: 540–546.
- Simmons, A.D., S.A. Goodart, T.D. Gallardo, J. Overhauser, and M. Lovett. 1995. Five novel genes from the cri-du-chat critical region isolated by direct selection. *Hum. Mol. Genet.* **4**: 295–302.
- Simmons, A.D., J. Overhauser, and M. Lovett. 1997. Isolation of cDNAs from the cri-du-chat critical region by direct screening of a chromosome 5-specific cDNA library. *Genome Res.* **7**: 118–127.
- Stanford Human Genome Center, Stanford University, Stanford, CA. 1996.: <http://www-shgc.stanford.edu/>
- Tunnacliffe, A., L. Liu, J.K. Moore, M.A. Leversha, M.S. Jackson, L. Papi, M.A. Ferguson-Smith, H.J. Thiesen, and B.A. Ponder. 1993. Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: Evidence for a pericentric inversion during primate evolution. *Nucleic Acids Res.* **21**: 1409–1417.
- Venter, J.G., M.D. Adams, G.G. Sutton, A.R. Kerlavage, H.O. Smith, and M. Hunka piller. 1998. Shotgun Sequencing of the Human Genome. *Science* **280**: 1540–1542.
- Whitehead Institute for Biomedical Research, MIT Center for Genome Research Cambridge, MA, 1996.: http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map

Received June 17, 1999; accepted in revised form August 23, 1999.