

was extensively cross-checked against the radiation hybrid map and was further reinforced by restriction fingerprinting of all mapped BACs¹⁵.

The greatest challenges were posed by massive, NRY-specific amplified regions (or amplicons), which comprise about one-third of the euchromatic NRY. Of the 758 STSs on which the map is built, 136 are present at two or more locations in the NRY. Although we avoided such repetitive STSs in favour of single-copy STSs wherever possible, substantial portions of the euchromatic NRY contained little or no single-copy sequence. For many such amplicons, BACs derived from different copies could not be distinguished by STS content or restriction fingerprinting¹⁵. In many cases, we distinguished among amplicon copies (and the BACs corresponding to them) by typing 'sequence family variants' (SFVs)⁵. SFVs are subtle differences (for example, single-nucleotide substitutions or dinucleotide repeat length alterations) between closely related but non-allelic sequences. We were analysing BACs from only one male's Y chromosome, so these subtle sequence differences could not represent allelic variants. In general, we identified SFVs only after comparing the DNA sequences of BACs that originated from distinct copies, despite having similar STS content. Thus, mapping and sequencing were inseparable, iterative activities in amplicon-rich regions.

The euchromatic NRY amplicons are diverse in composition, size, copy number and orientation (Fig. 1), with some occurring as tandem repeats, others as inverted repeats, and still others dispersed throughout both arms of the chromosome. The euchromatic amplicons are well populated with testis-specific gene families that may be critical in spermatogenesis (see Fig. 1 in Supplementary Information)^{1,2}.

One pair of amplicons is of particular interest in the context of human variation. Highlighted in Fig. 1 (arrows) are two units, each 300 kb long, that exist in opposite orientations on the short arm. These inverted repeats bound a region of around 3.5 Mb that occurs in one orientation (Fig. 1 in Supplementary Information) in the male from whom the BAC library was constructed, but in the opposite orientation in the existing map of naturally occurring deletions⁶. This may reflect variation among men for a 3.5-Mb inversion^{1,16}, perhaps the result of homologous recombination between the 300-kb inverted repeats flanking the inverted segment. Large Y-chromosome inversions are postulated to have been crucial in the evolution of the human sex chromosomes¹², and this 3.5-Mb inversion may be one of many massive NRY variants that exist in modern populations. □

Received 16 November; accepted 21 December 2000.

- Vogt, P. H. *et al.* Report of the Third International Workshop on Y Chromosome Mapping 1997. *Cytogenet. Cell Genet.* **79**, 1–20 (1997).
- Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
- Foot, S., Vollrath, D., Hilton, A. & Page, D. C. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60–66 (1992).
- Reijo, R. *et al.* Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.* **10**, 383–393 (1995).
- Saxena, R. *et al.* Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* **67**, 256–267 (2000).
- Vollrath, D. *et al.* The human Y chromosome: A 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992).
- Hudson, T. J. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954 (1995).
- Gyapay, G. *et al.* A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**, 339–346 (1996).
- Stewart, E. A. *et al.* An STS-based radiation hybrid map of the human genome. *Genome Res.* **7**, 422–433 (1997).
- Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
- Lunetta, K. L., Boehnke, M., Lange, K. & Cox, D. R. Selected locus and multiple panel models for radiation hybrid mapping. *Am. J. Hum. Genet.* **59**, 717–725 (1996).
- Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
- Mumm, S., Molini, B., Terrell, J., Srivastava, A. & Schlessinger, D. Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res.* **7**, 307–314 (1997).
- Schwartz, A. *et al.* Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**, 1–11 (1998).

- The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
- Jobling, M. A. *et al.* A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.* **8**, 1391–1394 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Nusbaum and T. Hudson for materials and advice on radiation hybrid mapping; J. Crockett, J. Fedele, N. Florence, H. Grover, C. McCabe, N. Mudd, S. Sasso, D. Scheer, R. Seim and P. Shelby for technical contributions; and D. Berry, J. Bradley, Y. Lim, A. Lin, D. Menke, M. Royce-Tolland, J. Saionz and J. Wang for comments on the manuscript. Supported in part by NIH.

Correspondence and requests for materials should be addressed to D.C.P. (e-mail: dcp@wi.mit.edu).

A high-resolution map of human chromosome 12

Kate T. Montgomery*, **Eunice Lee***, **Ashley Miller***, **Stephanie Lau***, **Cecilia Shim***, **Jeremy Decker***, **Denise Chiu***, **Suzanne Emerling***, **Mandeep Sekhon†**, **Rachel Kim***, **Jack Lenz***, **Jinghua Han***, **Ilya Ioshikhes***, **Beatrice Renault***, **Ivonne Marondel***, **Sung-Joo Kim Yoon‡**, **Kyuyoung Song§**, **V. V. V. S. Murthy||**, **Steven Scherer¶**, **Raluca Yonescu#**, **Ilan R. Kirsch#**, **Thomas Ried#**, **John McPherson†**, **Richard Gibbs¶** & **Raju Kucherlapati***

*Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA

‡Catholic University Medical College Research Institutes of Medical Sciences, Seoul 137-404, Korea

§University of Ulsan College of Medicine, Seoul 138-736, Korea

||Department of Pathology, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA

¶Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

#Genetics Department, Medicine Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

†Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri 63108, USA

Our sequence-tagged site-content map of chromosome 12 is now integrated with the whole-genome fingerprinting effort^{1,2}. It provides accurate and nearly complete bacterial clone coverage of chromosome 12. We propose that this integrated mapping protocol serves as a model for constructing physical maps for entire genomes.

Human chromosome 12 has been estimated to be 125 megabases (Mb) in size. Several low-density clone maps, genetic linkage maps and radiation hybrid maps that include chromosome 12 have been described^{3–5}. The map presented here is based on large bacterial clones that include bacterial artificial chromosomes (BACs), P1 (phage) artificial chromosomes (PACs; <http://www.chori.org/bacpac>) and a few cosmids⁶. To construct the map, we divided the chromosome into 1–2 centiMorgan (cM) intervals and assembled all sequence-tagged site (STS) markers known to be in each interval. We used probes derived from the non-polymorphic markers in an interval to screen colony arrays representing six to ten genome equivalents from the RPCI libraries. To extend contigs and link adjacent contigs, we used STSs corresponding to end sequences of BACs at the outer boundaries of each contig to screen members of the contig from which they were derived as well as possible adjacent contigs. This process was reiterated until the region was covered by

overlapping clones⁷. The STS-content map was assembled manually and the data stored in a relational database. A representative portion of the map corresponding to 5 cM of 12p can be seen in Supplementary Information; the entire map is available on our website (<http://sequence.aecom.yu.edu/chr12/>).

Our map contains 3,090 STS markers. Of these, 438 are polymorphic, 1,836 are non-polymorphic and the remaining 816 are based on genes or expressed sequence tags (ESTs). These markers were obtained from gene maps⁸, linkage maps^{3,9,10} or the Whitehead Institute radiation hybrid map (<http://www.genome.wi.mit.edu>), or developed at the Albert Einstein College of Medicine Genome Center. If the size of chromosome 12 is 125 Mb, the STS markers provide an average resolution of 40 kb. There are more than 5,300 large-insert clones on the map, and each marker is present in an average of 8.1 bacterial clones. This depth provides a high level of confidence in the quality of the map. Originally, 1,154 tiling path and seeder clones were identified for sequencing chromosome 12; 1,115 of them are being sequenced. The map indicates that some of the seeder clones are redundant and do not need to be finished. The estimated minimal set of clones covering chromosome 12 is 1,025. On average, adjacent clones overlap by 30%. As the average size of the clones in the RPCI-11 library is 174 kb, we can estimate the physical size of chromosome 12 excluding the gaps (see Supplementary Information) to be 125 Mb.

We assessed the accuracy of the map in two ways. First, we analysed tiling path clones with any available sequence to determine whether they overlapped as predicted by the map. Using electronic polymerase chain reaction (ePCR), BLAST (<http://www.ncbi.nlm.nih.gov/>) or the annotation of finished clones, most overlaps were confirmed. Second, we analysed clones located at 1-Mb intervals along the chromosome using fluorescence *in situ* hybridization (FISH) as part of the Cancer Chromosome Aberration Project (CCAP)^{11,12}. We chose 120 clones, and their order proved to be consistent with the order in the integrated map. The names of the clones, their cytogenetic locations, and details of their distribution are available at <http://www.ncbi.nlm.nih.gov/CCAP/>.

The STS-content map of chromosome 12 has been compared and integrated with the whole-genome map generated by BAC clone fingerprinting. The RPCI-11 BAC library was the primary resource for both maps, so the two are fully compatible. The comparison validated both methods and revealed them to be complementary. When the comparison was initiated, the STS-content map contained 74 contigs anchored to the genetic map. The whole genome map contained 75 contigs anchored to the G4 radiation hybrid map⁴ and nine additional contigs that were assigned to chromosome 12 but unanchored.

The integrated map now consists of 18 contigs that are not linked by PCR or sequence data (see Supplementary Information). The genetic markers on the STS map provide strong evidence for the orientation and order of most of these 18 contigs. This permits the direct comparison of fingerprints of end clones that are adjacent to each other on the map. In two cases we find that the fingerprints overlap and consider these contigs to be manually closed²; in one case we know by sequence that the clones do not overlap and in one the fingerprints are difficult to interpret (see below). The remaining 12 pairs of end clones do not have similar fingerprints and additional clones may be required to achieve contiguity. By these criteria, we consider the chromosome 12 map to have 14 gaps (see Supplementary Information), excluding the centromere, in its bacterial clone coverage.

We previously reported a yeast artificial chromosome (YAC) map of chromosome 12 (ref. 5). When the YAC, BAC and whole-genome fingerprint maps are combined, YACs cover all but five of the gaps, so we expect that these gaps are less than 1 Mb. The gaps not covered by YACs include one in a repetitive region at 43.6 cM (described below), and four between 137.5 and 169.1 cM.

Some of the gaps differ, and may reflect significant genomic

features. Understanding these regions will benefit the mapping and sequencing of this and other genomes. For example, the gap at 13.9 cM is covered by a YAC, but we have been unable to identify a bacterial clone to fill it, despite screening the libraries RPCI-1, 3, 4, 5, 11 and 13 several times with different probes. Furthermore, sequence is currently available for most of the clones in the region, but there are no sequence hits in any of the databases for the ends of the gap. This is the only region of the chromosome that appears to be missing from the available bacterial clone libraries, suggesting that true deficiencies in the bacterial libraries may be rare.

Two gaps appear in the region corresponding to 43.6 cM on 12p. Screening of the 12X BAC library with markers in this region yielded a set of clones whose number far exceeded unique representation of this region in the genome. STS-content analysis of the clones showed that each marker was positive for as many as 30–40 clones and the order of the clones could not be determined logically. These clones had similar restriction patterns and fell into very dense fingerprint contigs that included clones containing markers on chromosomes 2, 3, 11, 12 and 14. The combined fingerprint and PCR results indicate that this region may contain large duplications reminiscent of those described on chromosome 22 (refs 13,14) and it may also represent a region duplicated in other parts of the genome. The fingerprints of the clones flanking one gap in this region have been examined and the data strongly support coverage. Neither fingerprint nor STS data can currently resolve the second gap. This repetitive region may represent one of the most difficult types of region to map and sequence in full, in both the human and other genomes.

This map is one of the most complete human chromosome maps available. The known marker order for each clone allows easy assembly of draft sequence and the map will be invaluable in directing efforts towards completing the sequencing of chromosome 12 as well as future functional genomic studies. □

Received 16 November; accepted 21 December 2000.

1. Marra, M. A. *et al.* High throughput fingerprint analysis of large-insert clones. *Genome Res.* **11**, 1072–1084 (1997).
2. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
3. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
4. Stewart, E. A. *et al.* An STS-based radiation hybrid map of the human genome. *Genome Res.* **7**, 422–433 (1997).
5. Krauter, K. *et al.* A second generation YAC contig map of human chromosome 12. *Nature* **377** (Suppl.), 321–334 (1995).
6. Montgomery, K. T. *et al.* Characterization of two chromosome 12 cosmid libraries and development of STSs from cosmids mapped by FISH. *Genomics* **17**, 682–693 (1993).
7. Renault, B. *et al.* A sequence-ready physical map of a region of 12q24.1. *Genomics* **45**, 271–278 (1997).
8. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
9. Murray, J. C. *et al.* A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049–2054 (1994).
10. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
11. Kirsch, I. R. *et al.* A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. *Nature Genet.* **24**, 339–340 (2000).
12. Ried, T., Baldini, A., Rand, T. C. & Ward, D. C. Simultaneous visualization of seven different DNA probes by *in situ* hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc. Natl Acad. Sci. USA* **89**, 1388–1392 (1992).
13. Edelmann, L. *et al.* A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Genet.* **8**, 1157–1167 (1999).
14. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature*. **402**, 489–495 (1999).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work is supported by the NIH and the Human Genetics program at Einstein. The Cancer Chromosome Aberrations Project (CCAP) is funded through the office of the director, NCI. We appreciate comments on the manuscript from F. Collins and the assistance of J. Goltz. This work was inspired by F. Ruddle.

Correspondence and requests for materials should be addressed to R.K. (e-mail: kucherla@aecom.yu.edu).