

the maps of chromosomes 6, 9, 10, 13 and 20 by fluorescent *in situ* hybridization (FISH), in addition to the gaps previously measured on chromosome 22. Clones immediately flanking each gap were hybridized to extended DNA fibres, interphase nuclei or metaphase chromosome spreads.

An example of this analysis is given for chromosome 10 (Fig. 1). Fourteen contigs account for 124.4 Mb of DNA. They contain pericentromeric satellite sequences, and also a subtelomeric sequence on the short arm which is <0.1 Mb from the telomere; the most distal clone on the long arm is <0.15 Mb from the telomere¹⁰. From FISH analysis, the total extent of euchromatic gaps is ≤ 4.2 Mb. From these measurements, the estimated coverage of euchromatin in the map can be revised from 89% (based on Morton's previous estimate; Table 1) to 96.7% (124.4 of 128.6 Mb). A 9.75-Mb restriction map that spans the centromere has been constructed for chromosome 10 (ref. 11). By anchoring this to the clone maps on both chromosome arms, we estimate that the size of the gap across the centromeric region is around 4.5 Mb. This analysis provides a new estimate of 133 Mb for the total length of the chromosome (Fig. 1), compared with the previous value of 144 Mb. In a similar analysis of chromosome 13, seven contigs span 102 Mb and the remaining gaps in the euchromatic region total a maximum of 5 Mb. This leads to a new, increased estimate of the size of the euchromatic region of 107 Mb, and the current coverage in the map is 95%.

Our strategy may form the basis for finishing the map and sequence of every human chromosome. Accurate measurement of all gap sizes assists our continued efforts to bridge the remaining gaps in clones, using other cloning systems (for example yeast artificial chromosomes (YACs), plasmids and bacteriophage

lambda), and characterization of the centromeric satellites and the remaining heterochromatic regions, including the short arms of the acrocentric chromosomes (13, 14, 15, 21 and 22), will enable us to determine the true physical extent of DNA in the human genome. □

Received 28 November; accepted 21 December 2000.

1. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
2. The chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
3. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
4. Tilford, C. A. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
5. Montgomery, K. T. A high-resolution map of human chromosome 12. *Nature* **409**, 945–946 (2001).
6. Bruls, T. *et al.* A physical map of human chromosome 14. *Nature* **409**, 947–948 (2001).
7. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
8. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
9. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282**, 744–746 (1998).
10. Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
11. Jackson, M. S., See, C. G., Mulligan, L. M. & Lauffart, B. F. A 9.75-Mb map across the centromere of human chromosome 10. *Genomics* **33**, 258–270 (1996).

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 M. Sekhon, A. Chinwalla, J. McPherson and staff of the Genome Sequencing Centre, St. Louis for their assistance; the many collaborators who have contributed reagents and information to assist map construction on individual chromosomes; M. Jackson for helpful discussion; T. Cox, R. Pettett and the web team; and the Wellcome Trust for support.

Correspondence and requests for material should be addressed to D.R.B (e.mail: drb@sanger.ac.uk).

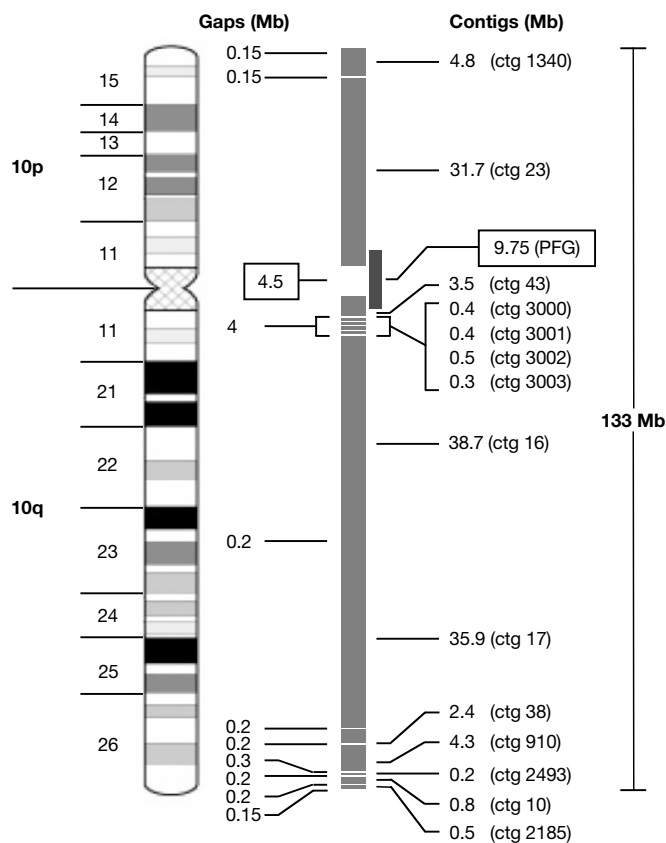


Figure 1 Physical map of chromosome 10. Clone contigs cover 124.4 Mb, euchromatic gaps cover 4.2 Mb, and the gap across centromeric satellites is 4.5 Mb¹¹. The separation between contigs 43 and 16 was determined as ~4 Mb on the basis of FISH of metaphase chromosome spreads, from which the sum of contigs 3000–3003 (1.6 Mb) was subtracted for a net gap coverage of 2.4 Mb in this region of 10q11.

A physical map of the human Y chromosome

Charles A. Tilford*, Tomoko Kuroda-Kawaguchi*, Helen Skaletsky*, Steve Rozen*, Laura G. Brown*, Michael Rosenberg*, John D. McPherson†, Kristine Wylie†, Mandeep Sekhon†, Tamara A. Kucaba†, Robert H. Waterston† & David C. Page*

* Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

† Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, Missouri 63108, USA

The non-recombining region of the human Y chromosome (NRY), which comprises 95% of the chromosome, does not undergo sexual recombination and is present only in males. An understanding of its biological functions has begun to emerge from DNA studies of individuals with partial Y chromosomes, coupled with molecular characterization of genes implicated in gonadal sex reversal, Turner syndrome, graft rejection and spermatogenic failure^{1,2}. But mapping strategies applied successfully elsewhere in the genome have faltered in the NRY, where there is no meiotic recombination map and intrachromosomal repetitive sequences are abundant³. Here we report a high-resolution physical map of the euchromatic, centromeric and heterochromatic regions of the NRY and its construction by unusual methods, including genomic clone subtraction⁴ and dissection of sequence family variants⁵. Of the map's 758 DNA markers, 136 have multiple locations in the NRY, reflecting its unusually repetitive sequence composition. The markers anchor 1,038 bacterial artificial chromosome clones, 199 of which form a tiling path for sequencing.

A low-resolution physical map of the Y chromosome was previously assembled by testing naturally occurring deletions and yeast artificial chromosome (YAC) clones for the presence or absence of 182 Y-chromosomal sequence-tagged sites (STSs)^{3,6}. These STS markers were generated from Y DNA sequences selected at random, which promoted representative sampling of the entire chromosome⁶. Nonetheless, most randomly selected sequences proved unusable as map landmarks because they corresponded either to interspersed repetitive elements found throughout the genome or to male-specific repetitive sequences dispersed to many locations in the NRY.

To construct a high-resolution map, we generated additional STSs in a directed manner. To enrich for the single-copy sequences most useful as map landmarks, we systematically applied genomic clone subtraction, whereby a 'tracer' clone's DNA is depleted of sequences shared with a set of 'driver' clones⁴. We identified a tiling path of 57 YACs that collectively spanned the euchromatic NRY³, and then carried out, in parallel, 57 subtractions, each employing one YAC as tracer and the remaining YACs (minus those overlapping the tracer YAC) as drivers. We sequenced a random sample of products from each of the 57 subtractions, identifying 308 additional STSs that proved useful in map assembly.

We used radiation hybrid mapping to integrate and order the random and subtraction-derived STSs. Large-fragment radiation hybrid panels offering long-range connectivity have been used to assemble human genome maps at 500–1,000-kilobase (kb) resolution^{7–10}. To obtain greater resolution in the NRY, where the number of STSs appeared sufficient to cover the euchromatic region at an

average spacing of 50 kb, we used a small-fragment radiation hybrid panel that had been used to build detailed maps of limited autosomal segments¹¹. We tested this panel for all random and subtraction-derived NRY STSs. Nascent radiation hybrid linkage groups were ordered and oriented with respect to the centromere by positioning selected STSs on the existing map of natural deletions⁶. Additional STSs generated in our project's later phases were also tested. Ultimately, 513 STSs were positioned, at a resolution of around 50 kb, on a radiation hybrid map encompassing nearly the entire euchromatic NRY.

To prepare for sequencing the NRY, we used the radiation hybrid map as a scaffold for assembling contigs of bacterial artificial chromosome (BAC) clones. We screened a BAC library of human male genomic DNA with hybridization probes derived from NRY STSs. Through subsequent polymerase chain reaction tests of STS content, we assembled 1,038 BACs into contigs that, except for four small gaps, represented the whole NRY (see Fig. 1 in Supplementary Information). Many portions of this BAC map could be assembled only after the sequence of selected BACs had been determined and compared. Also, many NRY genes and extragenic sequences are known to have closely related counterparts on the X chromosome¹². In many cases, it was initially unclear whether BACs identified using X-homologous NRY STSs, especially those from a 4-Mb region of 99% X–Y identity^{13,14}, derived from the NRY or from the X chromosome. We resolved these ambiguities by resequencing STSs from the BACs in question and comparing them to X- or Y-derived reference sequences. The resulting map of overlapping BACs and ordered STSs (Fig. 1 in Supplementary Information)

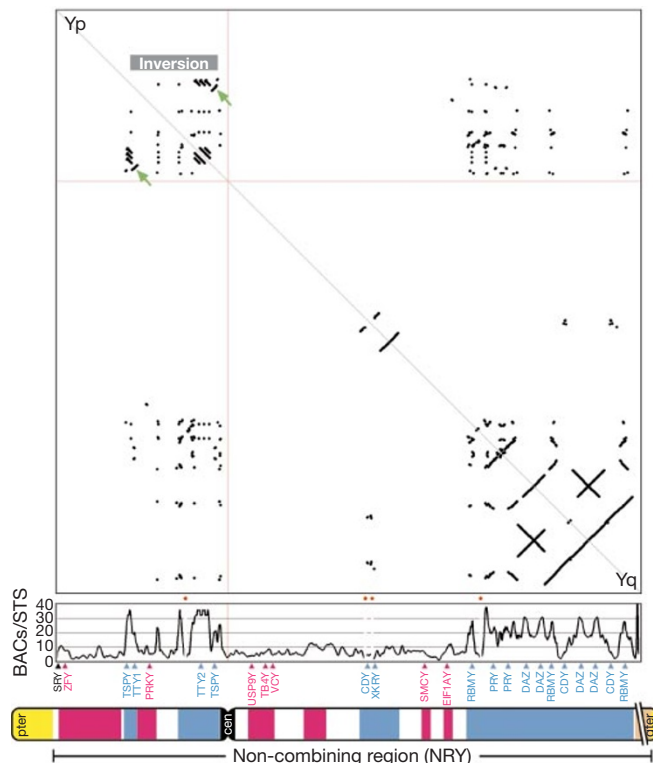


Figure 1 Repetitive structure of euchromatic NRY. Bottom, schematic of the Y chromosome, comprising large NRY flanked by pseudoautosomal regions (yellow). NRY is divided into euchromatic and heterochromatic (tan, shown truncated) portions, roughly 24 and 30 Mb, respectively. pter, short-arm telomere; cen, centromere; qter, long-arm telomere. Within euchromatic NRY, regions rich in NRY-specific amplicons (blue) or sequence similarity to X chromosome (red) are shown. Above chromosome schematic are positions of some NRY genes; most are found in amplicons (blue) or have X-linked homologues (red). Above genes is a plot of the average number of NRY BACs that contain each of the 758 STSs mapped (136 of these STSs at two or more locations) along euchromatic NRY. As expected, STSs in amplicon regions tended to be present in more BACs than STSs in X-homologous or

unshaded regions. (Plotted values are local averages within sliding window of five consecutive STSs; values reflect all NRY BACs containing those STSs, not just BACs assigned to site indicated.) Some amplicon regions were under-represented in the BAC library; four gaps remain (red diamonds; ≤ 100 kb each) in BAC coverage of NRY. Top, STS-based dot plot of euchromatic NRY. Each dot reflects occurrence of a particular STS at two points in map (complete map shown in Fig. 1 in Supplementary Information). Dots fall almost exclusively within amplicon regions. Many repeats of entire groups of STSs are apparent, with lines parallel to light grey diagonal indicating direct repeats and lines perpendicular to light grey diagonal indicating inverted repeats. Green arrows, inverted repeats flanking 3.5-Mb inversion (see text). Pale red lines, centromere.

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: dcpage@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