

# The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X

D. R. Bentley\*, P. Deloukas\*, A. Dunham\*, L. French\*, S. G. Gregory\*, S. J. Humphray\*, A. J. Mungall\*, M. T. Ross\*, N. P. Carter\*, I. Dunham\*, C. E. Scott\*, K. J. Ashcroft\*, A. L. Atkinson\*, K. Aubin\*, D. M. Beare\*, G. Bethel\*, N. Brady\*, J. C. Brook\*, D. C. Burford\*, W. D. Burrill\*, C. Burrows\*, A. P. Butler\*, C. Carder\*, J. J. Catanese†, C. M. Clee\*, S. M. Clegg\*, V. Cobley\*, A. J. Coffey\*, C. G. Cole\*, J. E. Collins\*, J. S. Conquer\*, R. A. Cooper\*, K. M. Culley\*, E. Dawson\*, F. L. Dearden\*, R. M. Durbin\*, P. J. de Jong‡, P. D. Dhami\*, M. E. Earthrowl\*, C. A. Edwards\*, R. S. Evans\*, C. J. Gillson\*, J. Ghorl\*, L. Green\*, R. Gwilliam\*, K. S. Halls\*, S. Hammond\*, G. L. Harper\*, R. W. Heathcott\*, J. L. Holden\*, E. Holloway\*, B. L. Hopkins\*, P. J. Howard\*, G. R. Howell\*, E. J. Huckler\*, J. Hughes\*, P. J. Hunt\*, S. E. Hunt\*, M. Izmajlowicz\*, C. A. Jones\*, S. S. Joseph\*, G. Laird\*, C. F. Langford\*, M. H. Lehvaslaiho\*, M. A. Leversha\*, O. T. McCann\*, L. M. McDonald\*, J. McDowall\*, G. L. Masien\*, D. Mistry\*, N. K. Moschonas‡, V. Neocleous§, D. M. Pearson\*, K. J. Phillips\*, K. M. Porter\*, S. R. Prathalingam\*, Y. H. Ramsey\*, S. A. Ranby\*, C. M. Rice\*, J. Rogers\*, L. J. Rogers\*, T. Sarafidou‡, D. J. Scott\*, G. J. Sharp\*, C. J. Shaw-Smith\*, L. J. Smink\*, C. Soderlund\*, E. C. Sotheran\*, H. E. Steingruber\*, J. E. Sulston\*, A. Taylor\*, R. G. Taylor\*, A. A. Thorpe\*, E. Tinsley\*, G. L. Warry\*, A. Whittaker\*, P. Whittaker\*, S. H. Williams\*, T. E. Wilmer\*, R. Wooster\* & C. L. Wright\*

\* The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

‡ Department of Biology, University of Crete and Institute of Molecular Biology and Biotechnology, PO Box 2208, 71409 Heraklion, Crete, Greece

§ Neurogenetic Laboratory, The Cyprus Institute of Neurology and Genetics, 6, International Airport Avenue, PO Box 23462, 1683 Nicosia, Cyprus

† Children's Hospital-BACPAC Resources, 747 52nd Street, Oakland, California 94609, USA

We constructed maps for eight chromosomes (1, 6, 9, 10, 13, 20, X and (previously) 22), representing one-third of the genome, by building landmark maps, isolating bacterial clones and assembling contigs. By this approach, we could establish the long-range organization of the maps early in the project, and all contig extension, gap closure and problem-solving was simplified by containment within local regions. The maps currently represent more than 94% of the euchromatic (gene-containing) regions of these chromosomes in 176 contigs, and contain 96% of the chromosome-specific markers in the human gene map. By measuring the remaining gaps, we can assess chromosome length and coverage in sequenced clones.

The task of sequencing the 3,200 megabase (Mb) human genome

can be subdivided into individual chromosome projects ranging in size from 263 Mb (chromosome 1)<sup>1</sup> to about 35 Mb (chromosomes 21q and 22q)<sup>2,3</sup>. Our strategy, in common with other groups<sup>4-6</sup>, was to map selected chromosomes individually, and then to combine the results with those of whole-genome mapping studies into a single map of the human genome<sup>7</sup>. Chromosome maps were constructed as follows (see Supplementary Information). First, we constructed a landmark map for each chromosome. Second, we identified bacterial clones (bacterial- or P1-derived artificial chromosomes (BACs or PACs)) from genomic libraries using the chromosome-specific landmarks, and assembled them into contigs on the basis of shared restriction enzyme fingerprints and landmark content. Third, contigs were extended and joined by chromosome walking. Walking was carried out using BAC end sequences generated in house from clones selected at the ends of contigs, or from the publicly available resources; joins were also made by identification of overlaps using genomic sequence data.

Clones that were representative of each contig were selected regularly for inclusion in the 'tiling path' (a set of minimally overlapping clones) for genomic sequencing<sup>8</sup>. Clones from the tiling path of each chromosome were deposited in the 'HumanMap' database at the Genome Sequencing Centre<sup>7</sup> (<http://genome.wustl.edu/gsc/human/mapping/>), for integration with the data obtained by whole-genome fingerprinting. New clones were identified from this integrated dataset to assist the extension and closure of the chromosome maps (Table 1). A detailed description of tiling paths and the underlying clone contigs is available as Supplementary Information and will continue to be updated at [www.sanger.ac.uk](http://www.sanger.ac.uk); all clones are publicly available.

A key question is the extent of coverage of each chromosome in the map. We analysed the coverage of the euchromatic regions on the basis of the assumptions given below and the approximate estimates that were available for chromosome length. Heterochromatic regions, which are estimated to comprise 3-15% of each chromosome (Table 1), are absent from the contigs analysed here. On the basis of the fingerprint bands in the maps (converted to Mb as described in Supplementary Information), 176 contigs represent 927 Mb, or 94% of the estimated total of 981 Mb euchromatin. Half of this (485 Mb) is in fifteen contigs of 22-62 Mb, that illustrates the extent of continuity obtained. We also analysed the representation of human gene markers in the map. The clone map contained 96% of the unique markers that were previously mapped to these chromosomes in GeneMap99 (ref. 9), and 90% were present in the genome sequence. As expected, this figure is higher for individual chromosomes 20 (99%) and 22q (98%), which are nearly or completely finished (Table 1).

Independent corroboration of the coverage of each chromosome required identification of the boundaries between euchromatic sequence and the centromeric, telomeric, and other heterochromatic repeat sequences, as well as measurement of the remaining gaps in the map. We have measured 52 of the 57 remaining gaps in

**Table 1 Status of chromosome maps**

Chromosome	Estimated length		Physical map			Gene map		
	Total (Mb)	Euchromatin (Mb)	Contigs (n)	Euchromatin covered (Mb)	(%)	Markers analysed (n)	Markers in contigs (%)	Markers in sequence (%)
1	263	238	78	226	95	4,769	94	87
6	183	178	13	170	96	2,942	96	92
9	145	123	21	113	92	1,771	98	90
10	144	139	14	124	89	2,022	99	91
13	114	98	7	102	105	1,028	99	90
20	72	67	6	61	91	1,159	99	99
22(q12-tel)	25.5	25.5	8	24.8	97	777	98	98
X(p21-q27)	115	112	29	106	94	925	93	88
Total	1,062	981	176	927	94	15,393	96	90

See methods in Supplementary Information for details.

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 telomere<sup>10</sup>. 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.

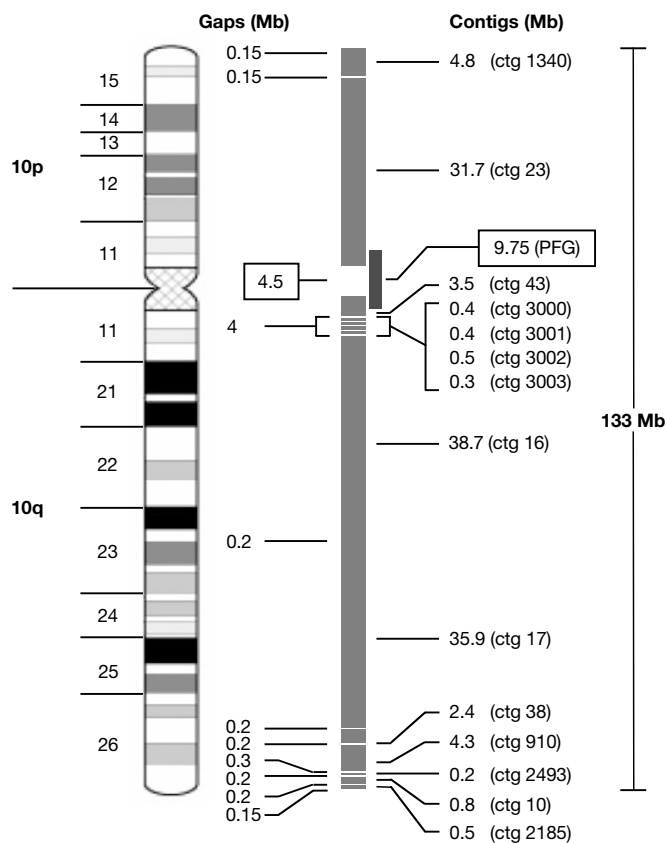
1. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
2. The chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
3. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **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 Mb<sup>11</sup>. 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<sup>1,2</sup>. 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 abundant<sup>3</sup>. 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 subtraction<sup>4</sup> and dissection of sequence family variants<sup>5</sup>. 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.