chromosome clone containing the 14q telomere (F.M., unpublished data). The distance to the alphoid centromeric repeats is unknown. However, the current most centromeric BAC already extends 1,200 kb beyond the most proximal marker of previously reported maps[11–13]. Interestingly, this clone contains two markers from our TNG map that exhibit extremely high retention rates in the hybrid lines, indicating that they may be close to the centromere.

The clone coverage of chromosome 14 that has been achieved using essentially an STC strategy is very satisfactory, and compares favourably with the coverage obtained for the human chromosomes that have been completely sequenced[3,4]. The two remaining gaps are located in the subtelomeric part of 14q; subtelomeric regions of many chromosomes are under-represented in most genomic libraries and hence often contain cloning gaps. The largest gap, estimated to be around 600 kb, was subsequently divided into two smaller gaps following the identification of clones through library screening with probes mapped within the gap. The second and more distal gap (20 kb) occurs in the immunoglobulin heavy-chain constant gene region, which contains a number of nearly identical genes and pseudogenes.

Considerations of the optimum design of an STC strategy should include theoretical aspects which correlate the effective depth of the clone end library and the number of seed points to the level of sequence redundancy[10,14], as well as practical aspects such as the sequencing capacity and costs[15], the time schedule for the project and the resolution of the available mapping data. However, a centralized repository of BAC end sequences is the only prerequisite for the construction of a tiling path based on the STC approach. Other mapping resources used in this project were auxiliary and provided useful information for seed selection and validation of map extension. Such a strategy is therefore generally portable to any large-scale sequencing project and is readily compatible with partitioning of the project. □

Received 20 October; accepted 21 December 2000.

1. Venter, J. C., Smith, H. O. & Hood, L. A new strategy for genome sequencing. *Nature* **381,** 364–366 (1996).
2. Mahairas, G. G. *et al.* Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. *Proc. Natl Acad. Sci. USA* **96,** 9739–9744 (1999).
3. Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402,** 489–495 (1999).
4. The chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. *Nature* **405,** 311–319 (2000).
5. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409,** 934–941 (2001).
6. Gyapay, G. *et al.* A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5,** 339–346 (1996).
7. Soderlund, C., Longden, I. & Mott, R. FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13,** 523–535 (1997).
8. Agarwala, R., Applegate, D. L., Maglott, D., Schuler, G. D. & Schaffer, A. A. A fast and scalable radiation hybrid map construction and integration strategy. *Genome Res.* **10,** 350–364 (2000).
9. Lange, K., Boehnke, M., Cox, D. R. & Lunetta, K. L. Statistical methods for polyploid radiation hybrid mapping. *Genome Res.* **5,** 136–150 (1995).
10. Roach, J. C., Siegel, A. F., van den Engh, G., Trask, B. & Hood, L. Gaps in the Human Genome Project. *Nature* **401,** 843–845 (1999).
11. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380,** 152–154 (1996).
12. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282,** 744–746 (1998).
13. Dear, P. H., Bankier, A. T. & Piper, M. B. A high-resolution metric HAPPY map of human chromosome 14. *Genomics* **48,** 232–241 (1998).
14. Batzoglou, S., Berger, B., Mesirov, J. & Lander, E. S. Sequencing a genome by walking with clone-end sequences: a mathematical analysis. *Genome Res.* **9,** 1163–1174 (1999).
15. Siegel, A. F. *et al.* Analysis of sequence-tagged-connector strategies for DNA sequencing. *Genome Res.* **9,** 297–307 (1999).

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 D. Cox, P. Dear and D. Cox for unpublished data and helpful discussions.943

Correspondence and requests for materials should be addressed to R.H. (e-mail: heilig@genoscope.cns.fr).

Integration of telomere sequences with the draft human genome sequence

H. C. Riethman*, Z. Xiang*, S. Paul*, E. Morse*, X.-L. Hu*, J. Flint†, H.-C. Chi‡, D. L. Grady‡ & R. K. Moyzis‡

* *The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA*
† *Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK*
‡ *Department of Biological Chemistry, College of Medicine, University of California, Irvine, California 92697, USA*

Telomeres are the ends of linear eukaryotic chromosomes. To ensure that no large stretches of uncharacterized DNA remain between the ends of the human working draft sequence and the ends of each chromosome, we would need to connect the sequences of the telomeres to the working draft sequence. But telomeres have an unusual DNA sequence composition and organization that makes them particularly difficult to isolate and analyse. Here we use specialized linear yeast artificial chromosome clones, each carrying a large telomere-terminal fragment of human DNA, to integrate most human telomeres with the working draft sequence. Subtelomeric sequence structure appears to vary widely, mainly as a result of large differences in subtelomeric repeat sequence abundance and organization at individual telomeres. Many subtelomeric regions appear to be gene-rich, matching both known and unknown expressed genes. This indicates that human subtelomeric regions are not simply buffers of nonfunctional 'junk DNA' next to the molecular telomere, but are instead functional parts of the expressed genome.

Telomeres are essential for genome stability and faithful chromosome replication. The chromatin structures associated with telomeric DNA mediate the many biological activities associated with telomeres, including cell-cycle regulation, cellular ageing, movement and localization of chromosomes within the nucleus, and transcriptional regulation of subtelomeric genes[1,2]. Specialized functions involving telomeric and subtelomeric DNA have evolved in several eukaryotes. For example, frequent subtelomeric gene conversion provides diversity for surface antigens in trypanosomes[3], and rapidly evolving subtelomeric gene families confer selective advantages for closely related yeast strains[4].

Human telomeres end with a stretch of the conserved simple repeat sequence $(TTAGGG)n$[5]. This tract is present at the end of all telomeres and therefore cannot be used to distinguish one telomere from another. To capture single-copy human DNA regions linked to telomeres that are useful for this purpose, we isolated large telomere-terminal fragments of human chromosomes using specialized yeast artificial chromosome (YAC) cloning vehicles called half-YACs[6]. Each half-YAC clone contains a large segment of subtelomeric DNA flanked by the cloning vector sequence at one end and the human telomere repeat sequence, which has been modified to operate as a functional yeast telomere *in vivo*, at the other. Characterization of these clones revealed low-copy subtelomeric repeats adjacent to the $(TTAGGG)n$ sequence[6,7]. Physical mapping experiments on a large group of these half-YAC clones showed that, in most cases, they can stably maintain faithful copies of human telomere-terminal DNA fragments in yeast[8]. By contrast, bacterial artificial chromosome (BAC) libraries used to prepare the human working draft sequence are not expected to contain sequences extending to the telomere, owing to the absence of restriction sites in $(TTAGGG)n$, the effects of length associated with the construction of size-selected DNA recombinant clones, and the genomic instability of these regions[9].

We used a combination of chromosome-specific single-copy sequences derived from the half-YAC clones and DNA end sequence derived from cosmid subclones of the half-YACs to connect most telomeres to the working draft sequence (Fig. 1). Our results show that the working draft sequence includes remarkably good coverage of human telomere regions. For the 24 human chromosomes, we analysed 46 telomere ends in all. The telomeres of the sex chromosome pair X and Y recombine meiotically, so these four telomeres are treated as two (designated the Xp/Yp pseudoautosomal telomere and the Xq/Yq pseudoautosomal telomere). We could integrate the working draft sequence with 32 telomere regions captured by half-YAC clones (blue dots). Of these 32 regions, 18 have working draft sequence coverage that includes DNA less than 50 kilobases (kb) from the telomere; for five of them, the sequence extends to the terminal (TTAGGG)*n* sequences[10–13] (see Supplementary Information). Although we were unable to capture two telomeres (5p and 20q) in half-YAC clones, we identified these regions in subtelomeric repeat-containing BAC clones and used them to connect to the working draft sequence (green dots).

We were unable to connect 7 of the remaining 12 telomere ends to the working draft, either because the working draft sequence does not yet extend into these regions (2q, 7p, 17p, 17q and Xp/Yp) or because unambiguous identification of overlapping working draft sequence was prevented by repeat sequences in the telomere clones (19p and 19q). BAC or cosmid clones connected to each of these seven telomeres were identified during construction of the fingerprint-based clone map of the human genome[14] (http://genome.wustl.edu/gsc/human/Mapping/index.shtml) and this is likely to facilitate the future integration of these telomeres into the working draft (see Methods). The five acrocentric chromosomes (13, 14, 15, 21 and 22) contain heterochromatic short arms comprising repeated DNA. We did not analyse these five short-arm telomeres (black rectangles) because their sequences were unstable in both yeast and bacteria, rendering them difficult to clone and characterize.

We have made available a detailed summary of the mapping experiments integrating telomeres with the working draft sequence, including specific telomere reagents, working draft contig designations, individual BAC clone accessions and accession numbers for our half-YAC-derived sequences (see Supplementary Information). For some chromosome ends, such as that of 11p (Fig. 2), we could precisely estimate the distance between the end of the working draft sequence and the telomere. However, this was not the case for many telomeric regions because much of the working draft sequence is still in small, unordered pieces.

As part of this study, around 1.1 megabases (Mb) of half-YAC-derived DNA sequence was acquired from cosmid end sequencing as well as from draft and finished sequencing of some subtelomeric cosmids (see Supplementary Information). In addition to defining the overlap relationship between the working draft sequence and the telomere clones, the half-YAC-derived sequences were useful for sampling the subtelomeric regions not yet included in the working draft sequence but captured in the half-YAC clones. Preliminary analysis of the half-YAC-derived sequences and the regions of the working draft sequence that overlapped with the half-YACs revealed several interesting features.

The sizes of subtelomeric repeat regions adjacent to the terminal (TTAGGG)*n* varied widely among individual telomeres, from 8 kb at the 7q telomere to ~300 kb at the 8p telomere. Large variations in subtelomeric repeat content have been detected near at least 18 telomeres[8,15,16]. Nonetheless, the scale of human genomic subtelomeric repeat content is now well defined, and it is clear that a significant part of the subtelomeric repeat region of the human genome is present in the working draft sequence.

Large subtelomeric repeat regions can cause false linkages in the

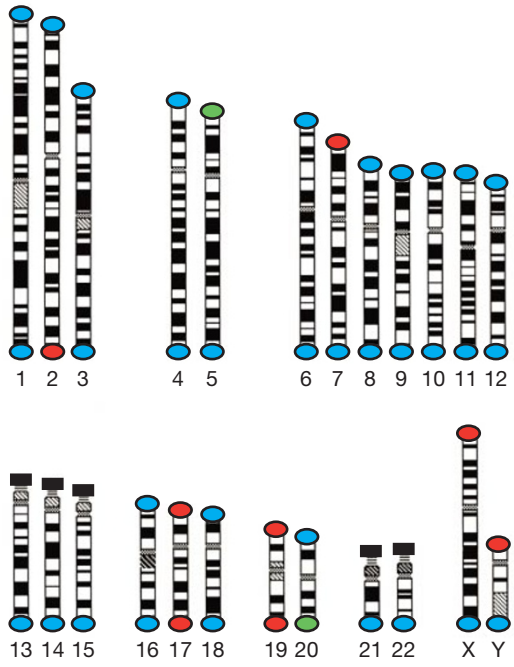

Figure 1 Summary of integration of telomeric DNA with working draft sequence. The two human pseudoautosomal telomere pairs (Xp/Yp and Xq/Yq) each recombine meiotically, so each pair is treated as a single telomere. Blue: the working draft sequence extends into these 32 telomere regions defined by half-YAC clones. Green: the working draft sequence ends within these subtelomeric repeat regions, but the distance to the molecular telomere is not known. Red: the telomeric DNA has not yet been integrated with working draft sequence. Telomere clones for the five acrocentric short-arm regions (black rectangles) have not been characterized.

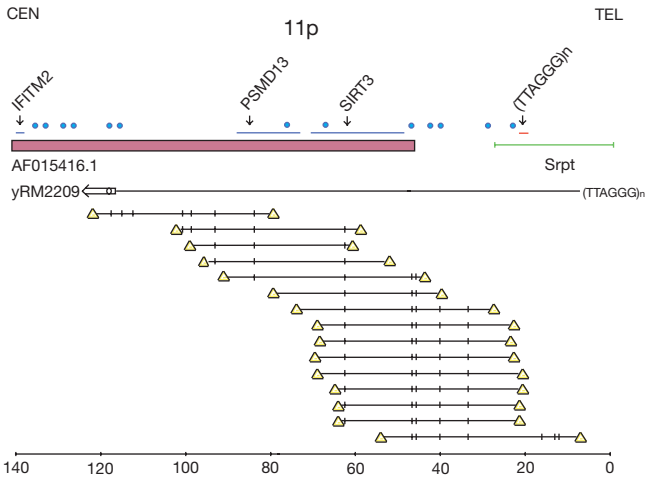

Figure 2 Connecting the 11p telomere to the working draft sequence. The end of the working draft sequence is represented by fragment AF015416 (magenta rectangle; a constitutent of working draft contig 18272). The half-YAC clone yRM2209 (black) is represented below. Cosmid subclones were derived from the half-YAC clone and the contig of these subclones, which encompasses most of the half-YAC clone, was aligned by *Eco*RI restriction sites (small vertical marks). The cosmid end sequences (yellow triangles) were screened to the working draft sequence to orient the clones relative to the 11p telomere. This physical mapping enabled us to determine the size and features of the 45-kb telomeric region missing from the working draft sequence. Features include a 25-kb region of subtelomeric repeat DNA (green), an internal (TTAGGG)*n* telomere repeat sequence (red) and four Unigene clusters of ESTs (blue dots). Within the 140-kb region extending from the telomere, we determined the location of three known genes, *IFITM2* (an interferon-induced transmembrane protein), *PSMD13* (a component of the proteasome 26S subunit) and *SIRT3* (a Sir-2-like histone deactylase implicated in telomere maintanance), as well as the positions of unigene clusters of ESTs that match 11ptel sequences and are distinct from known genes (blue dots).

BAC map and misassembly of working draft sequence. Large stretches of low-copy repeat DNA from subtelomeric repeat regions also localize to some pericentric chromosome regions, to the short-arm heterochromatin of acrocentric chromosomes and to a few loci in internal regions of chromosomes (for example, 1q42, 2q31, 4q28, 12p12 and Yq11.2). In previous iterations of the BAC map there were many instances of incorrect merges of subtelomeric repeat-containing BACs. To help identify these potentially problematic regions of the BAC map and working draft sequence, we have catalogued individual BAC clones containing segments of similarity with subtelomeric repeat regions (http://www.wistar.upenn. edu/Riethman). Inconsistencies between the current version of the BAC accession map (http://genome.wustl.edu:8021/pub/gsc1/ fpc_files/ freeze_2000_10_07/MAP/) and our telomere mapping studies are indicated in the Supplementary Information.

The abundance of low-copy repeat regions near telomeres is likely to make whole-genome shotgun assembly of subtelomeric regions virtually impossible. Indeed, previously characterized *Drosophila* subtelomeric repeat sequences are absent from its genome sequence[17]. By contrast, the entire sequence of yeast telomeric and subtelomeric regions was acquired using the half-YAC cloning strategy employed here[18].

Internal telomere-like sequences, each consisting of around 50–250 base pairs (bp) of a mixture of perfect and imperfect copies of (TTAGGG)*n*[11], were present in all subtelomeric repeat regions analysed. For example, multiple copies of internal telomere-like sequences were present in widely spaced parts of the 100-kb 18p subtelomeric repeat region, and were present in both orientations relative to the telomere. It is interesting to speculate that packaging of subtelomeric chromatin might involve interactions between the terminal (TTAGGG)*n* repeats and these internal telomere sequences. The TRF1 protein, which binds to (TTAGGG)*n in vivo* and can bind sequences corresponding to the short internal repeats *in vitro*[19], would be a good candidate for mediating such interactions.

Preliminary analysis of the potential gene content of the subtelomeric regions encompassed by the half-YAC-derived sequences and the overlapping portions of the subtelomeric working draft sequence was done by searching for sequence matches between the genomic DNA sequences and potential gene-derived complementary DNA and expressed sequence tag (EST) sequences in GenBank (http://www.wistar.upenn.edu/Riethman). Even this preliminary analysis reveals two features of subtelomeric regions. First, there are many sequence matches with genes and ESTs in most subtelomeric regions. We detected about 500 matches to transcripts identified by either a full-length cDNA or by a unigene cluster of expressed sequences in the 40 telomere regions analysed; 62 of these were found from half-YAC sequences mapping distal to the working draft sequence. Second, many of the genes and potential genes identified by sequence matches are members of gene families with many pseudogene copies. The sequence matches included around 100 known genes, both unique and members of gene families.

Human subtelomeric sequences have been proposed to serve as a buffer between the terminal (TTAGGG)*n* sequences, which are needed to protect chromosome ends from fusion and recombination, and vital internal chromosomal sequences[15]. However, the many expressed sequences throughout subtelomeric regions, extending almost to the molecular telomere, suggest that these regions may serve essential functions and are not simply dispensable junk DNA. □

Methods

We used a range of half-YAC-derived probes, including PCR- and cosmid subclone-derived probes and sequences (see Supplementary Information) and sets of collaboratively derived subtelomeric molecular and cytogenetic markers for specific telomeres[20–22], to connect specific cloned chromosome ends with flanking BAC contigs, either by DNA hybridization and PCR experiments or by computer-based matches (using BLAST2[23] sequence alignment programs) of sequenced subtelomeric DNA with working draft sequence.

Single-copy probes from three of the seven telomeres not connected to working draft sequence could be used to identify BAC clones from an 11× coverage RP11 BAC library, although fewer clones than expected were identified (singleton BACs from the 2q and the 17p telomeres, and three BACs from the 7p telomere). Low-copy repeat sequences at the 19p, 19q and 17q telomeres complicated attempted BAC library screens for these chromosome ends, but independent experiments have identified PAC and BAC clones connected to the 17q telomere[22] and a detailed physical map of chromosome 19 (http://greengenes.llnl.gov/genome/) exists to help guide closure of the 19p and 19q subtelomeric gaps, which occur in duplicated regions containing a family of zinc finger-encoding genes. The remaining telomere region (Xp/Yp) is encompassed by a 500-kb clone contig extending to within a few kb of the telomere[24].

Physical mapping experiments using a site-specific cleavage method (RARE cleavage[8,25]) have been done for 21 telomeres to demonstrate co-linearity of the half-YAC insert DNA with the cognate telomere. In the absence of RARE cleavage data, the presence of subtelomeric repeats adjacent to terminal (TTAGGG)*n* sequences in all of the designated half-YAC clones is taken as strong evidence for proximity to the telomere; this has been borne out by the RARE cleavage experiments carried out so far.

Half-YAC clones containing chromosome-specific DNA were not recovered from four chromosome ends. BAC and cosmid clones identified by virtue of their subtelomeric repeat content form the initial basis for the telomere linkages to 5p, 20q, 19q and Xp/Yp. The BAC clones used to mark the 5p and 20q telomeres and the cosmid used to mark the 19q telomere each contain an internal telomere repeat sequence and subtelomeric repeat sequences, and localize to telomeric ends of the BAC map (5p, 20q) and the chromosome 19 physical map (http://greengenes.llnl.gov/genome/). On the basis of the known sequence organization of other telomeres, only additional subtelomeric repeat sequence is likely to reside distal to the subtelomeric repeat segments contained in these clones, although the possibility of single-copy DNA distal to them cannot be formally excluded at present. A cosmid clone mapped to the Xp/Yp pseudoautosomal telomere using *Bal*31 exonuclease experiments[26] forms the telomeric boundary of a large cosmid contig[24] whose sequence is not yet available.

Received 15 November; accepted 18 December 2000.

1. Blasco, M. A., Gasser, S. M. & Lingner, J. Telomeres and telomerase. *Genes Dev.* **13,** 2353–2359 (1999).
2. deLange, T. & Jacks, T. For better or worse? Telomerase inhibition and cancer. *Cell* **98,** 273–275 (1999).
3. McCulloch, R., Rudenko, G. & Borst, P. Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur on variant surface glycoprotein expression sites lacking 70-bp repeat sequences. *Mol. Cell Biol.* **17,** 833–843 (1997).
4. Carlson, M., Celenza, J. L. & Eng, F. J. Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosomal telomeres. *Mol. Cell Biol.* **5,** 2894–2902 (1985).
5. Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc. Natl Acad. Sci. USA* **85,** 6622–6626 (1988).
6. Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T. & Olson, M. V. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc. Natl Acad. Sci. USA* **86,** 6240–6244 (1989).
7. Brown, W. R. A. *et al.* Structure and polymorphism of human telomere-associated DNA. *Cell* **63,** 119–132 (1990).
8. Macina, R. A. *et al.* Molecular cloning and RARE cleavage mapping of human 2p, 6q, 8q, 12q, and 18q telomeres. *Genome Res.* **5,** 225–232 (1995).
9. Doggett, N. A. *et al.* An integrated physical map of human chromosome 16. *Nature* **377** (Suppl.), 335–365 (1995).
10. Flint, J. *et al.* The relationship between chromosome structure and function at a human telomeric region. *Nature Genet.* **15,** 252–257 (1997).
11. Flint, J. *et al.* Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6,** 1305–1313 (1997).
12. Ciccodicola, A. *et al.* Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Hum. Mol. Genet.* **9,** 395–401 (2000).
13. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405,** 311–320 (2000).
14. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409,** 934–941 (2001).
15. Wilkie, A. O. M. *et al.* Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64,** 595–606 (1991).
16. Trask, B. J. *et al.* Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7,** 13–26 (1998).
17. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287,** 2185–95 (2000).
18. Louis, E. J. & Borts, R. A complete set of marked telomeres in *Saccharomyces cerevisiae* for physical mapping and cloning. *Genetics* **139,** 125–136 (1995).
19. Bianchi, A. *et al. TRF1* binds a bipartite telomeric site with extreme spatial flexibility. *EMBO J.* **18,** 5735–5744 (1999).
20. Ning, Y. *et al.* A complete set of human telomeric probes and their clinical application. *Nature Genet.* **14,** 86–89 (1996).
21. Rosenberg, M. *et al.* Characterization of short tandem repeats from thirty-one human telomeres. *Genome Res.* **7,** 917–923 (1997).
22. Knight, S. J. L. *et al.* An optimized set of human telomere clones for studying telomere integrity and architecture. *Am. J. Hum. Genet.* **67,** 320–332 (2000).
23. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
24. Gianfrancesco, F. A novel pseudoautosomal gene encoding a putative GTP-binding protein resides in the vicinity of the Xp/Yp telomere. *Hum. Mol. Genet.* **7,** 407–414 (1998).
25. Riethman, H., Birren, B. & Gnirke, A. in *Genome Analysis: A Laboratory Manual, Vol. 1, Analyzing DNA* (eds Birren, B., Green, E., Klapholz, S., Meyers, R. & Roskams,J.) 83–248 (Cold Spring Harbor Laboratory Press, New York, 1997).
26. Cooke, H. J. & Smith, B. A. Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harbor Symp. Quant. Biol.* **51,** 213–219 (1986).

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 J. Finklestein, N. Atigapramoj, E. Dabagyan, S. Huang, A. Ambriz, A. Harxhi and K. Sutton for their contributions to this work, which was supported by NIH and DOE.

Correspondence and requests for materials should be addressed to H.C.R. (e-mail: Riethman@wistar.upenn.edu).

Comparison of human genetic and sequence-based physical maps

Adong Yu*, Chengfeng Zhao*, Ying Fan*, Wonhee Jang†, Andrew J. Mungall‡, Panos Deloukas‡, Anne Olsen§, Norman A. Doggett‖, Nader Ghebranious*, Karl W. Broman¶ & James L. Weber*

** Center for Medical Genetics, Marshfield Medical Research Foundation, Marshfield, Wisconsin 54449, USA*
† National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894, USA
‡ The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
§ Joint Genome Institute, Walnut Creek, California 94598, USA
‖ Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
¶ Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205-2179, USA

Recombination is the exchange of information between two homologous chromosomes during meiosis. The rate of recombination per nucleotide, which profoundly affects the evolution of chromosomal segments, is calculated by comparing genetic and physical maps. Human physical maps have been constructed using cytogenetics[1], overlapping DNA clones[2] and radiation hybrids[3]; but the ultimate and by far the most accurate physical map is the actual nucleotide sequence. The completion of the draft human genomic sequence[4] provides us with the best opportunity yet to compare the genetic and physical maps. Here we describe our estimates of female, male and sex-average recombination rates for about 60% of the genome. Recombination rates varied greatly along each chromosome, from 0 to at least 9 centiMorgans per megabase (cM Mb^{-1}). Among several sequence and marker parameters tested, only relative marker position along the metacentric chromosomes in males correlated strongly with recombination rate. We identified several chromosomal regions up to 6 Mb in length with particularly low (deserts) or high (jungles) recombination rates. Linkage disequilibrium was much more common and extended for greater distances in the deserts than in the jungles.

All nucleated human cells contain two homologous copies of each chromosome, except for the sex chromosomes in males. During the formation of the sperm and egg cells, the number of each chromosome is reduced to one so that fertilization restores the normal diploid number in the next generation. The process of chromosome reduction, meiosis, is usually accompanied by exchange or recombination of DNA between the homologous parental chromosomes. Genetic maps, which are based on meiotic recombination, order and estimate distances between DNA sequences that vary between parental homologues (polymorphisms). The primary unit of distance along the genetic maps is the centiMorgan (cM), which is equivalent to 1% recombination.

The genetic maps used in our analysis were based upon the genotyping of 8,031 short tandem repeat polymorphisms (STRPs) from Généthon, the University of Utah and the Cooperative Human Linkage Center in eight reference CEPH families[5]. Excluding the sex chromosomes, the maps cover about 4,250 cM in females and 2,730 cM in males. The genetic maps are relatively marker dense, with an average of 2–3 STRPs per cM, but are also relatively low resolution because only 184 meioses (92 in each sex) were analysed.

The physical maps used were all DNA sequence assemblies. For chromosomes 21 and 22, we used the finished, published sequences[6,7]. For the other 20 autosomes and for the X chromosome, we used the public draft sequence assemblies, 5 September 2000 version (http://genome.cse.ucsc.edu)[4]. As we required relatively long stretches of sequence, we used only sequence assemblies that were over 1.5 Mb long (between terminal STRPs), contained more than three STRPs and had a marker order that agreed with published genetic and radiation hybrid maps. The amount of sequence used from each chromosome is shown in Fig. 1. Some chromosomes had much better coverage than others. We analysed 253 sequence assemblies ranging in length up to 70 Mb and spanning a total of 1,806 Mb (roughly 58% of the portion of the genome that is not highly repetitive). By far the most common reason for rejecting sequence assemblies was insufficient length; only seven assemblies were rejected for incompatible marker order.

Recombination rates varied greatly across the genome, from 0 to 8.8 cM Mb^{-1} (Table 1). Sex-average recombination rates (the average for males and females combined) did not vary as much as the sex-specific rates (for males and females considered separately) because male and female recombination rates at specific sites often differed substantially. We identified 19 recombination deserts up to 5 Mb in length with sex-average recombination rates below 0.3 cM Mb^{-1}, and 12 recombination jungles up to 6 Mb in length with sex-average recombination rates greater than 3.0 cM Mb^{-1} (see Supplementary Information). Wide variation in recombination rates across chromosomes has been observed previously for humans[8–11] and for other eukaryotic species[12–15], and is clearly the rule rather than the exception.

In an effort to identify the basis of differences in recombination rates, we compared the rates to several marker and sequence parameters. These parameters included GC content, STRP informativeness, position of the marker relative to the centromeres and telomeres, density of runs of various short tandem repeats, especially $(A)_n$, $(AC)_n$, $(AGAT)_n$, $(AAN)_n$ and $(AAAN)_n$ sequences, and the density of various interspersed repetitive elements, including

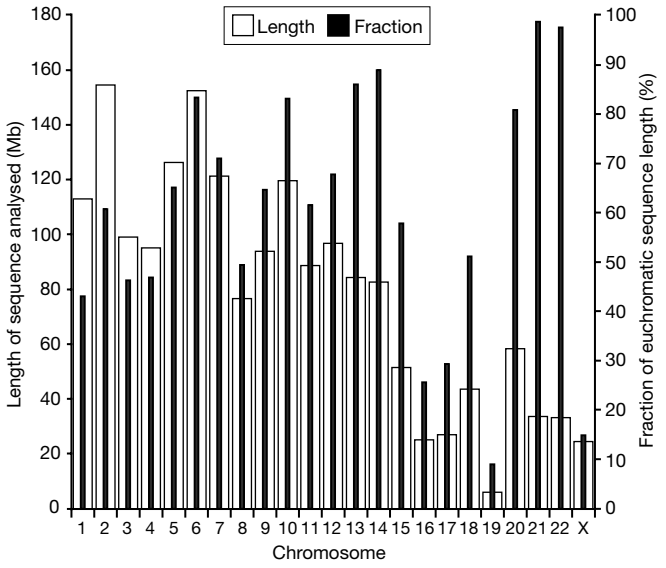

Figure 1 Sequence coverage for comparison of the genetic and physical maps. The total length of sequence used in the analysis (open bars) and the approximate percentage of the euchromatic sequence length (solid bars) are shown for each chromosome.