A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome

The Cancer Chromosome Aberration Project (CCAP) aims to integrate systematically the cytogenetic and physical maps of the human genome. This will be achieved by high-resolution fluorescence-*in-situ*-hybridization (FISH) mapping of colony-purified, STS-tagged, single-site hybridizing BAC clones spaced at 1–2-Mb intervals across the entire genome. In the opening phase of the CCAP initiative, we have generated the BAC clone sets that link the cytogenetic and physical maps of chromosomes 7 and 22. We have established a publicly accessible repository of CCAP clones and developed a web site (http://www.ncbi.nlm.nih.gov/CCAP) for data retrieval and cross-referencing of CCAP clones with other cytogenetic and genomic databases relevant to cancer.

On the basis of the established physical maps of human chromosomes 7 (ref. 1) and 22 (ref. 2), we identified and colony-purified BACs at approximately 1–2-Mb intervals, and subjected them to high-resolution[3], dual-colour FISH mapping[4] onto metaphase and prometaphase chromosomes. An example of the mapping procedure is shown (Fig. 1) for dual-colour FISH with clones from chromosome 22. All selected clones contained at least one mapped sequence-tagged site (STS).

On the basis of the estimated size of chromosome 7 (170 Mb), the localization of 84 BACs provides an average spacing of approximately 2 Mb between clones. For the purposes of defining relative clone order among clones mapping within the same chromosomal band, all bands were further arbitrarily subdivided into regularly spaced intervals labelled with lowercase letters. This subdivision was performed on the 850-band per genome ISCN (International System for Human Cytogenetic Nomenclature) ideogram[5]. For chromosome 7, 160 such intervals were defined (Fig. 2). For chromosome 22 (of approximately 35 Mb), 22 BACs and 62 intervals were defined (Fig. 2), providing, on average, a mapped clone for every 1.6 Mb. A complete list of STSs, clone names and high-resolution mapping positions for the set of BACs for chromosomes 7 and 22 is available (http://www.ncbi.nlm.nih.gov/CCAP and Tables 1 and 2, http://genetics.nature.com/supplementary_info/).

The collection of mapped BAC clones described here is offered as a shared resource to the scientific community for study of chromosome aberrations. CCAP BACs can be purchased through a separate repository (specify the BAC ID, chromosome number and catalogue #MB11200, http://www.resgen.com). With the availability of the genomic sequence of chromosome 7 and the sequence of the euchromatic part of human chromosome 22, several long, contiguous sequence segments (contigs) can be constructed[6,7]. BACs mapped by CCAP can often be localized within these contigs by alignment of BAC end sequences to the finished contig or because they are among the clones that actually contributed to the contig. Thus sequence data and pointers to additional clones will often be available for fine-scale analysis of genomic regions associated with chromosome aberrations. For example, for the first seven listed BACs beginning at the chromosome 22 centromere with bK115F6 and ending with bk992D9, the availability of sequence information allows us to delineate the distances between them to the nearest 5 kb as 1,195 kb, 1,120 kb, 1,420 kb, 1,800 kb, 1,085 kb and 1,745 kb, confirming the 1–2 Mb spacing we sought. We have made available a demonstration of the use of these seven BAC clones in the characterization of a translocation into the Igλ region in a malignant cell line derived from a patient with multiple myeloma (see Fig. 3, http://genetics.nature.com/supplementary_info/).

As of the date of submission, we have received additional clone sets and completed the initial phase of mapping for chromosomes 1 and 18 (BACs provided from the BAC/PAC Resources by P. de Jong and N. Nowak, http://bacpac.med.buffalo.edu/human/rfa_meeting.html), and chromosome 16 (V. Cheung). We are

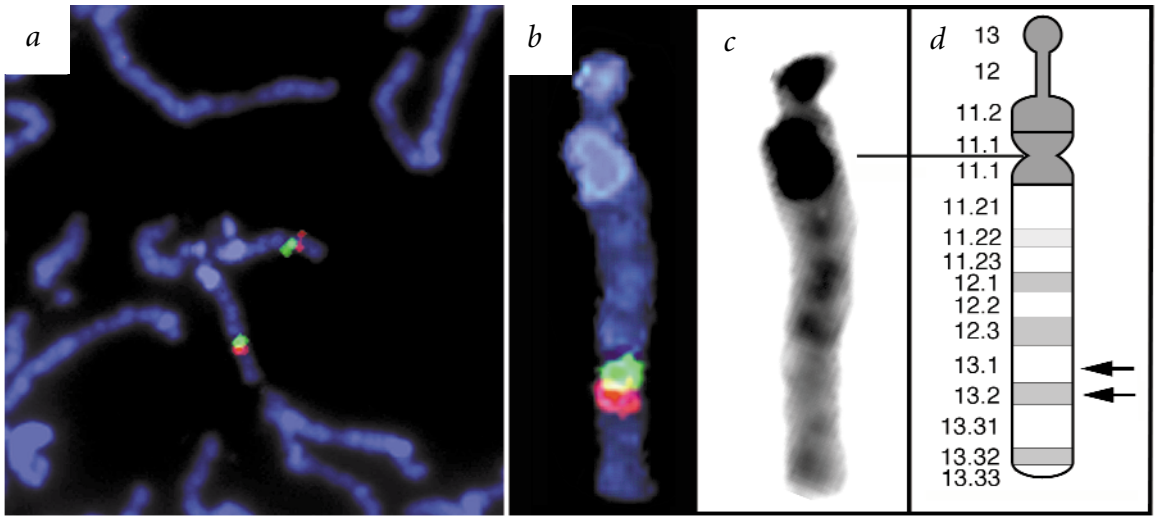
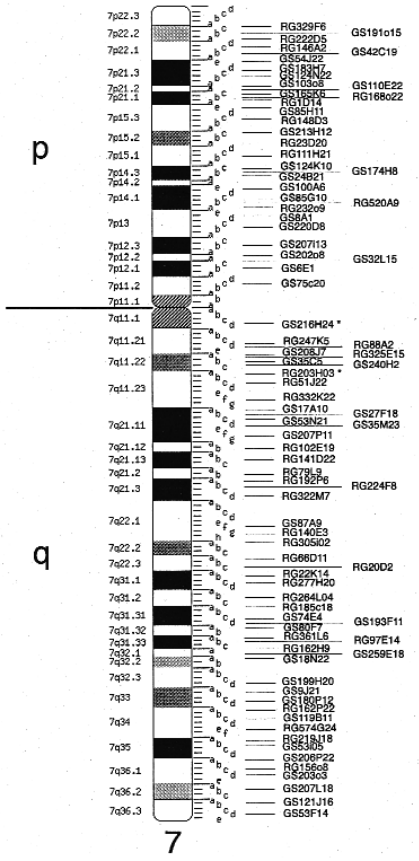


Fig. 1 Two-colour, high-resolution FISH mapping of chromosome 22 BACs. Metaphase chromosomes were prepared to provide a resolution of approximately 1,000 bands per genome. ***a***, FISH analysis performed with clone bK390B3 (green) and bK229A8 (red). To facilitate band assignment, suitable DAPI-stained chromosomes were extracted from the image (***b***), contrast inverted (***c***) and displayed next to a chromosome ideogram (***d***). Arrows (right) indicate the corresponding map positions.

© 2000 Nature America Inc. • http://genetics.nature.com

© 2000 Nature America Inc. • http://genetics.nature.com



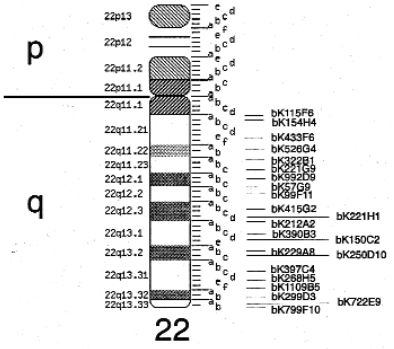


Fig. 2 Cytogenetic coordinate system for chromosomes 7 and 22 and localization of mapped BAC clones. To convert the cytogenetic position to a coordinate system, conventional chromosome bands were divided into arbitrarily defined intervals. This coordinate system allows for delineation of clone order even in the absence of other cytogenetic landmarks.

in the process of mapping clone sets for chromosomes 3 and 5 (N. Nowak), chromosome 14 (V. Cheung) and chromosome 12 (R. Kucherlapati). The use and power of this resource will certainly be aided by the related BAC development and mapping efforts of B. Trask and U.-J. Kim as well as the parallel and complementary efforts of the laboratory of J. Korenberg[8].

Ilan R. Kirsch[1], Eric D. Green[2], Raluca Yonescu[1], Robert Strausberg[3], Nigel Carter[6], David Bentley[6], Margaret A. Leversha[6], Ian Dunham[6], Valerie V. Braden[2], Eva Hilgenfeld[1], Greg Schuler[4], Alex E. Lash[4], Grace L. Shen[5], Maria Martelli[1], W. Michael Kuehl[1], Richard D. Klausner[3] & Thomas Ried[1]

[1]Genetics Department, Medicine Branch, National Cancer Institute; [2]Genome Technology Branch, National Human Genome Research Institute; [3]Office of the Director, National Cancer Institute; [4]National Center for Biotechnology Information, National Library of Medicine; and [5]Cancer Genetics Branch, National Cancer Institute, NIH, Bethesda, Maryland, USA. [6]Sanger Centre, Cambridge, UK. Correspondence should be addressed to I.R.K. (e-mail: kirschi@exchange.nih.gov).

1. Bouffard, G.G. *et al. Genome Res.* **7**, 673–692 (1997).
2. Collins, J.E. *et al. Nature* **377**, 367–379 (1995).
3. Yunis, J.J. & Chandler, M.E. *Prog. Clin. Pathol.* **7**, 267–288 (1978).
4. Ried, T., Baldini, A., Rand, T.C. & Ward, D.C. *Proc. Natl Acad. Sci. USA* **89**, 1388–1392 (1992).
5. *ISCN 1995 An International System for Human Cytogenetic Nomenclature* (ed. Mitelman, F.) (Karger, Basel, 1994).
6. Jang, W., Chen, H.C., Sicotte, H. & Schuler, G.D. *Trends Genet.* **15**, 284–286 (1999).
7. Dunham, I. *et al. Nature* **402**, 489–495 (1999).
8. Korenberg, J.R. *et al. Genome Res.* **9**, 994–1001 (1999).

© 2000 Nature America Inc. • http://genetics.nature.com

ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome

In recent years, the discovery of many regulatory elements within introns, the recognition of the regulatory potential of intronic and other non-protein coding RNAs, and the concept of a cellular 'ribotype' resulting from differences in RNA processing in different cells and tissues have led to increasing interest in the role of introns in enhancing eukaryotic genetic complexity, via alternative splicing and as both the recipient and donor of *cis*-acting and *trans*-acting elements[1–4].

To explore the evolution and function of introns in eukaryotes, we have developed an intron sequence information system (ISIS; http://isis.bit.uq.edu.au/) which contains information on over 170,000 spliceosomal introns. Data in ISIS version 1 is based on intron-containing sequences from GenBank release 111. ISIS contains phylogenetic and protein homology categories, information about individual sequences and various bioinformatic analyses of taxonomical groupings of sequences using non-redundant subsets of the data. The database is searchable by Blast, GenBank attributes and elements that we have annotated within introns, and gives graphical views of gene structure and elements such as alternative coding regions, EST matches and repetitive sequences.

During analysis of this database, we found many EST matches within sequences annotated as introns, indicating that there are many previously unrecognized alternatively spliced exons, especially as many of these exons are conserved between species. Alternative splicing was first predicted by Walter Gilbert[5], and subsequently verified by the discovery of cDNA isoforms exhibiting the addition or exclusion of whole or partial exons[6,7], although identification of such splice variants has largely occurred on an ad hoc basis. The development of large human EST (partial cDNA) sequence libraries over recent years, however, provides an opportunity to examine the incidence of alternative splicing globally by searching these libraries for exon skipping, exon truncation or inclusion of sequences currently described as intronic.

We examined the incidence of unrecognized exons in introns in 2,698 non-redundant human genes in ISIS which contain at least one complete sequenced intron and two flanking exons. We