Commentary

Stylianos E. Antonarakis

Laboratory of Human Molecular Genetics,
Department of Genetics and Microbiology,
University of Geneva, and Division of
Medical Genetics, Cantonal Hospital of
Geneva, Switzerland

Mapping by Sequence Homology

An important current preoccupation of the human geneticist is the mapping of disease phenotypes and genes. Linkage analysis using DNA polymorphic markers in appropriate families, and chromosomal abnormalities associated with pathologies were the two principal ways of assigning phenotypes to particular chromosomal regions. The menu for the mapping of genes or other DNA segments of interest contains many more choices: PCR or hybridization alternatives using somatic cell or radiation hybrids, or recombinant clones of physical maps, as well as fluorescent in situ hybridization and linkage analysis of polymorphic markers are the most common modalities by which a wealth of mapping information has already been accumulated. The Genome Database now contains (search of 15 March 1996) 84,609 mapped DNA 'objects'; among those 11,121 are DNA polymorphic markers, 33,276 are STS (sequence-tagged sites), 5,902 are genes (4,102 of those are cloned). The Online Mendelian inheritance in man (OMIM) now contains 8,002 entries, approximately half of which are disease phenotypes. There is still, however, a long way to go until the mapping of all human genes that number probably between 70,000 and 100,000 [1].

One of the most exciting recent developments in genomics research is the determination of approximately 400,000 partial sequences of human cDNAs (ESTs; expressed sequence tags) from different tissues and the availability of these sequences through the public databases [2]. These ESTs have been accumulated mainly from the Merck-Wash.U project (255,163 sequences on 13 February 1996, http://genome.wustl.edu/est) and the TIGR project [about 150,000 sequences, ref 3]. Other investigators in Europe, Japan and the US are also contributing to this fast-growing and useful resource. Besides the nucleotide sequence information, an important annotation of these ESTs is almost certainly their position in the growing physical maps of the human chromosomes. The current international mapping efforts for the ESTs include the use of PCR amplification in either YACs or radiation hybrids [4, 5].

Exon trapping [6] has proven to be an outstanding method of identifying partial gene sequences from cloned genomic DNA (or to be more accurate, sequences that start with a splice acceptor and end with a splice donor sequence). Several disease-causing genes have been identified with this method using clones that map in critical regions (examples include the huntingtin, presenilin and leptin genes) [7]. Furthermore, exon trapping is now extensively used to create transcription (genic) maps for several chromosomes using chromosome-specific cosmids, PACs, BACs, YACs as the starting material. After elimination of the false-positive sequences the overwhelming majority (nearly all) of the trapped sequences map back to their clones of origin. Comparison of the sequences of these trapped exons with the nucleotide sequences in the public databases often provides identities (or near identities) to one or more ESTs. This identity is obviously sufficient initial evidence for mapping the corresponding EST(s), and therefore their cDNA clones, to the chromosome or chromosomal region of the trapped sequence. This mapping by sequence homology/identity will probably become a popular and useful method of mapping in the coming years since sequencing is the ultimate method of exploration and characterization of the

genomes. Below, we refer to this mode of localizing cDNAs as mapping by sequence homology (MSH).

Let's look at some numbers. During the last 18 months, several investigators have used exon trapping from chromosome-specific cosmids to identify portions of transcription units. In our laboratory, 1,200 randomly picked chromosome 21-specific cosmids have been used and a total of 559 different 'exons' have been identified [8]. To date we have mapped 133 of them back to chromosome 21; no exon has been mapped elsewhere in the human genome. We are therefore confident that the majority of trapped sequences map to the chromosome of origin. Exons from 41% of the known chromosome 21 genes have been trapped, and interesting homologies with other genes have been found. Homology searches have also revealed identity or near identity (homology of 100 or >95%, respectively) of 49 trapped sequences (8.8%) with unmapped ESTs (more than 49 ESTs are identified because of the redundancy of the EST database). We propose that the vast majority of their corresponding cDNAs map to chromosome 21. In the laboratories of Buckler and McCormick [9], a similar experiment was done using chromosome 12-specific cosmids. From a total of 936 different exons that recognized 37% of the known chromosome 12 genes, 60 (6.4%) showed identity or near identity to ESTs and their corresponding cDNAs presumably map to chromosome 12. The experience with chromosome 22 from the Buckler laboratory is similar [10]. From a total of 603 different unique trapped sequences that recognized 35% of the known chromosome 22 genes, 57 (9.5%) showed identity (or near) to ESTs. All of these results were obtained by homology searches done around August/September 1995 and do not include searches in the EST database of TIGR. Thus approximately 8% of the trapped exons from the three different experiments (166 of 2,098) recognized and provisionally mapped cDNAs for which ESTs are available in the public databases. The homology computer search therefore provided considerable evidence for the localization of many cDNA fragments to a specific human chromosome or chromosomal segment. As with every mapping method, confirmation of this initial evidence by MSH is needed for the definitive mapping. It is well accepted in the genetic mapping community and a common practice over the years that two independent mapping methods should be used to definitely assign a mapping 'object' to a specific chromosomal region.

cDNA selection is an alternative method to isolate transcribed sequences from selected genomic regions [11, 12]. It has also been successfully used in the cloning of several disease-related genes. Recently, Lovett et al. [13] have applied this method to create chromosome-specific cDNA libraries by using chromosome-specific cosmids as a starting reagent. As with the experience with trapped exons, 9.8% of the selected cDNA clones showed identity to ESTs. After elimination of cDNA clones with repetitive elements, 78% (108/139) of the cDNAs map back to the appropriate chromosome [13]. This represents an outstanding enrichment for chromosome-specific cDNAs. It seems however that MSH after cDNA selection is not specific enough since there is one in five mapping missassignment. This is probably due to capture of cDNAs that belong to sequence-related gene families that map in different chromosomes. Thus, MSH from chromosome-specific cDNA selection needs to be confirmed by another mapping method.

How does the MSH compare with other existing mapping methods? In MSH after exon trapping there is no difference with the methods based on hybridization, since these latter depend on nucleic acid homology. Errors can be due to hybridization of related sequences scattered throughout the human genome; the same is true for the computer 'hybridization' in which a 95% homology can be due to sequencing errors or DNA polymorphisms as well as evolutionarily related sequences. There is no difference either from the PCR amplification based methods since these also depend on nucleotide matches, and amplification products of related targets may have the same size. In addition, false-positive amplification products of contamination of the PCR may result in erroneous mapping conclusions. Furthermore the mapping methods of linkage analysis and radiation hybrids provide a statistical statement for order and distance between DNA 'objects' and therefore as such give a best estimate which is always subject to changes with the availability of more data, elimination of genotyping errors (for linkage) or false-positive and negative results (for radiation hybrids), or alternative ways of statistical treatment of the data.

In summary, MSH may become a popular selection in the menu of mapping possibilities since it provides information that is not of inferior quality or accuracy from the other existing alternatives.

Acknowledgements

References

1 Fields C, Adams MD, White O, Venter JC: How many genes are in the human genome? Nature Genet 1994;7:345–347.

2 Boguski MS, Schuler GD: ESTablishing a human transcript map. Nature Genet 1995;10:369–371.

3 Adams MD, Kerlavange AR, Fleischman RD, Faldner RA, Bult CJ, et al: Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequences. Nature 1995;377:3–174. (For updates see web site: http://www.tigr.org/)

4 Berry R, Stevens TJ, Walter NAR, Wilcox AS, Rubano T, Hopkins JA, Weber J, Goold R, Bento Soares M, Sikela JM: Gene-based STS as the basis for a human gene map. Nature Genet 1995;10:415–423.

5 Stewart EA, Aggarwal A, Barrett E, Hebert S, Chu A, et al: Mapping human genes with respect to ordered meiotic linkage markers by using a panel of whole genome radiation hybrids (abstract). Am J Hum Genet 1995;57:A3.

6 Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE: Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 1991;88:4005–4009.

7 Collins FS: Positional cloning moves from perditorial to traditorial. Nature Genet 1995;9:347–350.

8 Chen HM, Chrast R, Rossier C, Morris MA, Lalioti MD, Antonarakis SE: Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping. Genome Res 1996;6:747–760.

9 Stotler CJ, Lawrence N, Rutter M, Deaven L, McCormick MK, Buckler AJ: Transcription and physical mapping of human chromosome 12 (abstract). Am J Hum Genet 1995;57:A56.

10 Troffater JA, Long KR, Murrell JR, Stotler CJ, Gusella JF, Buckler AJ: An expression-independent catalogue of genes from human chromosome 22. Genome Res 1995;5:214–224.

11 Lovett M, Kere J, Hinton LM: Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci USA 1991;88:9628–9632.

12 Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM: cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc Natl Acad Sci USA 1991;88:9623–9627.

13 Lovett M, Del Mastro R, Wang L, Gallardo T, Ashley J, Wasmuth J, McPherson J, Hilliard C, Simmons A, Clines G: Human chromosome-specific cDNA libraries: New tools for gene identification and genome annotation (abstract). Am J Hum Genet 1995;57:A55.