nature genetics
volume 2 no. 3 november 1992

Shake that 'body map'

A recent flurry of landmark developments in genetic linkage and physical mapping of the human genome[1-4] has once again underlined the enormous potential that genomic research surely holds for the future. Perhaps this is just as well, for, over the past two years, considerable excitement mixed with no little controversy has surrounded a complementary endeavour to survey the estimated 100,000 genes in the human genome. This approach, which involves the partial sequencing and characterization of comple-mentary DNA (cDNA) clones from tissue-specific libraries, to produce what many term 'expressed sequence tags' (ESTs), has yielded an abundance of valuable information, notably with respect to more than 2,500 human brain sequences[5,6] and close to 2,000 expressed genes from *Caenorhabditis elegans*[7,8].

Unfortunately, the debate about the merits of this approach has been obscured by the continuing furore over the patent application by the National Institutes of Health (NIH) on behalf of Dr Craig Venter's group, who published the brain cDNA studies and have since moved to a new, nonprofit facility in Maryland — the Institute for Genomic Research. The NIH is expected to appeal against the US Patent and Trademark Office's first refusal of the application in the next few months. For its part, the patent office is said to be keen to reach a firm conclusion on the issue.

Whatever the outcome, there is no doubt that the attraction of the EST strategy is winning many converts. In this issue of *Nature Genetics*, two groups, led by Kenichi Matsubara of Osaka University, and James Sikela at the University of Colorado Health Sciences Center, present new analyses of more than 2,000 ESTs[9,10]. While Sikela and coworkers, like Venter's group, have concentrated on brain cDNAs (as befits a pharmacology group primarily interested in neuropsychiatric disorders), Matsubara's team is pursuing an altogether more ambitious goal — the production of a 'body map' of expressed genes in various human tissues, which have been estimated to number around 200 in total.

In its report on page 173, the Japanese group describes the sequencing of the 3' ends of 982 cDNAs derived from the human liver cell line, HepG2. In contrast to earlier projects in which random-primed, normalized cDNA libraries were employed to standardize the relative proportions of clones corresponding to various transcripts, Matsubara and colleagues have deliberately chosen a nonbiased library, because they are as interested in the relative quantities of the RNA species as in their individual identity. This, they believe, will become a valuable basis for comparison with other EST datasets from different tissues prepared in the same manner. Their first choice of a HepG2 cDNA library contains short inserts derived from the most distal, 3' portions of the RNA, near the poly(A) tail. This approach has the advantage of minimizing the potential bias associated with variable cloning efficiencies, and facilitating

clone-to-clone comparison at the junction of the poly(A) tail, but the compromise is that little coding information is retained in the inserts.

The Japanese group acknowledges that the study of further cDNAs will be desirable to characterize the library in full, but analysis of some 1,000 clones provides interesting reading. Three genes of high abundance were each represented in more than 1% of the total number of clones analysed: elongation factor 1α, serum albumin and translationally controlled tumour protein. The next group of middle abundance genes contained 170 different species, only 35 of which were identified in the database. Finally, 468 clones (48%) were unique and 415 of these were apparently novel. Of identified clones, 27% were implicated in protein synthesis, a high figure which will probably not be emulated in other tissues. As a rough estimate, it is likely that about one third of expressed genes in a given tissue fall into this 'housekeeping' category, whereas another third are likely to be tissue-specific.

The 'body map' approach poses some difficult questions in spite of the interesting possibilities. It is far from clear how useful it will be to repeat the exercise for the 50–200 tissues in the human body given the expected redundancy of sequences —but then again, this may be useful information in itself. Matsubara and colleagues believe that characterization of hundreds of tissue-specific genes will have significant pharmaceutical implications, for example in comparing normal with malignant tissues (or normal tissues such as liver with transformed cell lines such as HepG2). Another benefit of their strategy may be to examine expression patterns during the course of development (such as cell differentiation in haematopoeisis) or in subregions of a complex organ, the obvious candidate being the brain.

The related paper in this issue[10] by Sikela and coworkers provides information on a further 1,024 brain ESTs (to complement the 2,500 or so reported so far[5,6]) using a more conventional procedure. A prescreening step removed many of the more abundant cDNAs to allow access to those less well represented clones. Further steps are under way to 'normalize' brain libraries still further — both Sikela and Venter's groups plan to make available several hundred separate brain clones to allow the preparation of a 'subtraction' library which will expose yet more genes.

A frequent question raised in consideration of the utility of such EST collections is how they will be converted to the genetic map — after all, how useful is a gene to a geneticist in the absence of its chromosomal assignment? Apart from the advantages of constructing near full-length cDNA libraries to pursue the function of some of these genes, Sikela and colleagues argue that longer clones facilitate their chromosomal mapping as well. Their cloning strategy also made use of the largely unique, 3' untranslated regions of the cDNA sequences, which facilitates both physical and genetic forms of chromosomal localization, the latter by virtue of identifying novel polymorphisms. In their paper, 20 clones were localized using either physical or genetic approaches; earlier this year, Venter and colleagues reported the localization of 46 of their human brain cDNAs[11]. (Both studies were performed in conjunction with Mihaelis Polymeropolous of the National Institute of Mental Health.) Given the thousands of genes that are being characterized, the need for improved gene-mapping technology is clear, but should be facilitated as more of the human genome is linked by yeast artificial chromosomes (YACs)[1,2]. By comparison, more than 40% of a recently published *C.elegans* cDNA collection were mapped to a gridded YAC genomic library[7].

Despite the controversy over the patent issue (the Japanese group will not be applying for patents, says Matsubara), the power of the EST technology is without question in identifying new genes faster than ever and providing new routes for investigators to tackle old problems. The work is currently being extended by many groups to other human tissues, such as T lymphocytes and skeletal muscle, and will soon embrace other model organisms too. □

1. Chumakov, I. *et al. Nature* **359**, 380–387 (1992).
2. Foote, S. *et al. Science* **258**, 60–66(1992).
3. NIH/CEPH collaborative mapping group *Science* **258**, 67–82 (1992).
4. Weissenbach, J. *et al. Nature* **359**, 794–799 (1992).
5. Adams, M.D. *et al. Science* **252**, 1651–1656 (1991).
6. Adams, M.D. *et al. Nature* **355**, 632–634 (1992).
7. Waterston, R. *et al. Nature Genet.* **1**, 114–123 (1992).
8. McCombie, W.R *et al. Nature Genet.* **1**, 124–131 (1992).
9. Okubo, K. *et al. Nature Genet.* **2**, 173–179 (1992).
10. Khan, A.S. *et al. Nature Genet.* **2**, 180–185 (1992).
11. Polymeropoulos, M.H. *et al. Genomics* **12**, 492–496 (1992).