

Map of Man

It is now technically feasible to construct a physical map of the human genome represented in a real ordered library of cloned DNA segments. Such a library would be of great value for the whole range of molecular studies in human biology and medicine and is essential for research in human genetics. It is also the first and necessary step that needs to be taken for the far more grandiose project of sequencing the human genome.

We are constructing such a library for the nematode, C.elegans, which has a haploid genome content of 7×10^7 base pairs. The method we have developed is based on cloning random fragments of nematode DNA in cosmids. These 45 kb fragments are then 'fingerprinted'. The DNA is first cut with an enzyme with a 6 base pair recognition site which on the average yields ten fragments. The ends are labelled by the incorporation of a radioactive nucleotide and the DNA is then cut with an enzyme recognizing a 4 base pair site, to give an average of 20 fragments with an average length of 128 nucleotides. These small fragments are separated by electrophoresis under conditions which allow the accurate measurement of their lengths and the resulting pattern is stored in a computer. All new clones are matched against the growing database for coincident bands which links the clones into 'contigs'. If the clones represent a true random sample of the genome, then a map with very few gaps could be obtained with 20,000 clones, which is ten times the coverage, ie the ratio of genome to clone length. Our current experience suggests that there are biases in the cloning procedure which will make it necessary to introduce non-random procedures using preselected clones to close the gaps as the map saturates into a

collection of islands. Nevertheless, we now have reasonable benchmarks to consider what would be required to construct a similar library for the human genome.

Human cells have a haploid genome content of 3 to 4×10^9 base pairs, some 50 times larger than that of the nematode. As a first approximation we have simply scaled the nematode work by this factor to estimate the resources required for the human genome map. About a million clones would need to be studied. At the moment, about 100 clones can be handled by one person per week; this includes making the recombinants, growing them up, fingerprinting them, entering the data into the computer and editing and checking the contigs. At 5,000 clones a year, the nematode genome required 4 man-years so the human genome would be about 200 man-years using the present technology and organization of the work. We think that quite attainable small modifications of the procedures, together with a proper organization of the work, could reduce this to 150 man-years bailed on biochemical workers.

We wish to propose that a reference laboratory be set up to construct this ordered library of segments of DNA. Initially, such a laboratory will devote most of its resources to generating the database, but as this becomes built up some resources would need to be added or diverted to positioning clones derived from other sources in the library and to providing such users with the flanking sequences from matched contigs. Our experience with the nematode is that the reference laboratory could offer a valuable and much desired service. In addition, the availability of the database and the ordered library will ensure collaboration and the flow of material into the reference laboratory. We do not plan to work on locating the clones or chromosomes by in situ

hybridization, for example. This can be done by collaborations with other laboratories doing cytogenetics. In any event chromosomal linkage would be established by placing premapped clones into the data base, including the sequences identifying mapped restriction length polymorphisms.

Reference Laboratory

To fix ideas we consider a reference laboratory with a group of 6 technicians doing the mapping. It needs to operate as separate organization and be managed by, and related, to a research laboratory. The group of 6 biochemical technicians would need to be supported by 4 other people who would undertake entering the data into a computer and other work required for the maintenance of the data base and the increasing work of communicating with outside users and dealing with collaborations. Using present technology with random clones such a laboratory could achieve about one fifth of the map in 5 years, doing 40,000 clones per year. With more technicians this could be increased but we believe this is an easily manageable way to start. However, at the same time a research and development project needs to be supported in the associated research group to improve the methods used, and to study how best to achieve the final stage of cloning the map. Improvements would be sought in the following areas:

- (a) methods of improving randomness of the clones,
- (b) methods of cloning larger fragment of DNA,
- (c) automation of the biochemical operations and automatic entry of the results to enhance rate of throughput,
- (d) improvements in computer methods of searching for matches in the data base. As the latter grows this will come to be the slowest step and parallel processors will be required.

Any improvement will be immediately fed into the reference laboratory to realise its immediate benefit; it will not nullify what has already been done, but simply accelerate the rate at which saturation of the random clones is achieved.

The costs of the reference laboratory can be estimated as follows:

1. Capital costs:

The space required for the reference laboratory would be:

1,200 sq ft of biochemical space

300 sq ft to house a computer

300 sq ft for data entry, communications, etc.

and ultimately of space at -70° to house the million clones. This will need to be duplicated at two separate sites for security. We estimate that this could easily be achieved with 10m^3 of cold space. If we set this initially to 200 sq ft, this sets the space needed at a total of 2,000 sq ft.

The centre should have its own computer which with associated peripherals and specialized input and display devices might cost £250,000.

Equipment for the biochemical work: £200,000.

2. The running costs may be estimated as follows:

Biochemical expenses (1£ per clone): £40,000 pa

Computer maintenance, communications and other materials: £30,000 pa

Salaries are estimated at £15,000 per person pa. £150,000 pa

To this we need to add maintenance (? rental) of the space.