Genome analysis

Genome Mapping and Sequencing

Cold Spring Harbor Laboratory, New York, NY, USA, 10–14 May 1995

Stephen W. Scherer

Beutenberg Symposium on Genome Analysis: Strategies, Medical and Industrial Applications

Jena, Germany, 28–30 June 1995

Jörg D. Hoheisel

The goal of the human genome project (HGP) is to construct detailed genetic and physical maps of the chromosomes, to identify and map all of the estimated 50 000–100 000 genes, and ultimately to complete the nucleotide sequence of human DNA.

The annual *Cold Spring Harbor Genome Meeting* brings together an international cast of researchers to review their progress towards this goal, and to discuss the development of new mapping, sequencing and information technologies. The major theme that emerged from this year's meeting was that the emphasis in human genome research would be moving from genetic and physical mapping to transcriptional (gene) mapping and large-scale genomic sequencing.

The initial five-year goal for the US HGP (1990–1995) was to establish a fully connected genetic and physical map with markers spaced on average every two to five centimorgans (cM), and 100 kb, respectively. Généthon (Evry, France) reported the final version of their human genetic-linkage map, which encompassed over 5300 $(CA)_n$ microsatellite repeats. The Co-operative Human Linkage Center (CHLC) (USA) have integrated 1000 tri- and tetra-nucleotide repeat markers and 400 gene-specific polymorphisms, to come up with a comprehensive 1 cM map. These two studies alone have supplied enough markers for a 1 cM genetic map.

YACs, STS and ESTs

Physical mapping of the human genome has been based largely on the ordering of yeast artificial chromosome (YAC) clones into groups (contigs) based on the regions of overlap that they share with each other. The most common method to demonstrate overlap between YACs is to show that they contain common DNA markers [genes, sequence tagged-sites (STSs), expressed sequence tagged-sites (ESTs) and genetic markers].

Chromosome-specific maps incorporating all of these types of markers into 'integrated' maps were presented at the meeting for chromosomes 2, 3, 4, 5, 7, 10, 11, 12, 13, 14, 16, 19, 21, 22 and X. The CEPH-Généthon group (France) and the Massachussetts Institute of Technology Genome Centre (The Whitehead Institute, Cambridge, MA, USA) reported on the progress of their 'whole genome' mapping effort, using primarily STS-content mapping of mega-YACs (YACs with an average insert size of 1000 kb), and estimated that they had covered approximately 75% and 90% of the genome, respectively.

Important, but possibly overlooked, results were presented on a poster by Kathleen Gardiner *et al.* (Eleanor Roosevelt Institute, Denver, CO, USA) that compared the YAC-based map of chromosome 21 (one of the best-mapped chromosomes) with a pulsed-field gel electrophoresis (PFGE) restriction enzyme map of the same chromosome. A PFGE map is derived using genomic DNA that has not been cloned. The results suggested that there were at least three regions (encompassing about 10 Mb of DNA, or 20% of the chromosome) that were present in the PFGE map but not in the YAC map. Similar 'gaps' were described on other chromosomes, suggesting that the estimates of YAC coverage of the genome may not be the most conservative. A study of chromosome 7 by Lap-Chee Tsui's group (The Hospital for Sick Children, Toronto, Canada), however, indicated that, to date, almost every marker not present in YAC libraries could be found in bacterial base-cloning systems such as cosmid, P1, PAC and BAC. Fluorescent *in situ* hybridization (FISH) and long-range PCR were other valuable techniques used to analyze problematic regions.

Radiation hybrid mapping

Radiation hybrid (RH) mapping was demonstrated to be an increasingly more popular method and the groups of Jean Weissenbach (Généthon), Peter Goodfellow (University of Cambridge, Cambridge, UK) and David Cox (Stanford Human Genome Center, Stanford, CA, USA) described their progress to order DNA markers with this technique. A large fraction of the markers used are being integrated with other physical mapping efforts including YAC mapping, and the data indicate that this technique is reliable and may be particularly useful for ordering DNA markers in 'difficult to map' regions and for rapid EST mapping.

Transcriptional maps

A major emphasis of the HGP is to place all of the genes on the map [Stewart, A. (1995) *Mol. Med. Today* 1, 53]. Various groups and consortia described different, but generally complementary, strategies to accomplish this and most involved mapping ESTs and cDNA fragments (transcripts) or CpG islands against YACs and/or RH panels. The Merck Expressed Gene Initiative, which involves groups at Merck (NJ, USA), Washington University (St Louis, MO, USA) and the Lawrence Livermore National Laboratory (Livermore, USA) aims by mid-1996 to have, in the public domain, the sequence of the 5' and 3' ends of 200 000 cDNA fragments from various cDNA libraries, providing a rich resource for gene mapping and discovery.

DNA sequencing

The argument that it is time to re-focus efforts and put forth a major initiative towards sequencing the human genome, quietly dominated the meeting. As outlined by the 'Sulston–Waterston plan' [Marshall, E. (1995) *Science* 267, 783–784], the human genome would be sequenced using existing automated gel-based technology, instead of waiting for new developments that may eventually increase speed and reduce cost.

Impressive amounts of DNA sequence were presented, including 16 Mb of the 100 Mb *Caenorhabditis elegans* genome (on target for the finish date of 1998), by the Sanger Centre (Hinxton Hall, Cambridge, UK)/Washington

University (St Louis, MO, USA) collaboration, 1 Mb of the Huntington's disease locus in a year by the Sanger Centre, and about 1 Mb of the DiGeorge Syndrome critical region on chromosome 22 (University of Oklahoma, Norman, OK, USA). Based on these and other successful pilot projects in yeast and *Drosophila*, it has been estimated that the complete human genome sequence could be obtained with over 99% precision by 2001 (five years ahead of schedule). 'The opposition' to this plan argued silently that the proposal has not been seriously scrutinized, and that no guidelines have been established to define standards for cost, completeness or accuracy. Genome cartographers and technologists alike, muttered concerns that the investment in a single strategy would shift funds away from technology development too soon, and that only a few large (existing) centres were capable of competing. The salvation for 'mappers' is that YAC contigs are not suitable reagents for DNA sequencing and 'sequence ready' clones (in bacterial vectors such as cosmids that are currently available for only a few chromosomal regions) would have to be generated. In fact 'DNA sequencers' who are usually absent from mapping poster sessions, were seen shopping around for 'next month's template'.

The final word seemed to belong to DNA sequencing 'mogul' John Sulston (Sanger Centre, Cambridge, UK). In reminding genome scientists that the complete human DNA sequence would represent the ultimate map containing all of the restriction sites, STSs, ESTs and genes, the only question that was left unanswered was 'how the rest of the community could get a piece of the [funding] pie'.

The first *Beutenberg Symposium on Genome Analysis* followed hard on the heels of the German Federal Ministry's announcement of the German Human Genome Program and emphasized the rapidly growing interest shown in this line of research in Germany, not only by scientists, but also by industry.

Although the meeting was attended by a number of distinguished scientists from all over the world, the proceedings retained a distinctly German flavour. Without a poster session, scientifically, the meeting was more a display of the current state of research in genome analysis, rather than a broad presentation of new developments.

Technologies

Jean Weissenbach (Généthon, Paris, France) and Hans Lehrach (Max Planck Institut für Molekulare Genetik, Berlin, Germany) gave an insight into the progress of their respective projects: the creation and use of a high-resolution genetic map by the former, and a highly relational, broad analysis on many levels by the latter. Later in the session, more specific, technical developments were presented. Aaron Bensimon (Institut Pasteur, Paris, France) described the process of 'molecular combing', by which even very large DNA molecules (Mb) can be attached at one end to a solid surface and subsequently stretched and aligned by a receding air–water interface, making them a potentially important tool for analysis. Jörg Hoheisel discussed the usefulness of sorting template DNAs by hybridization prior to sequencing, as attempted as part of the European Yeast Sequencing Project.

Molecular medicine

Hans Hilger Ropers (University Hospital Nijmegen, The Netherlands) presented his results on genetic predisposition to aggressive behaviour (scientifically sound, but controversially interpreted), while Anthony Monaco (Institute of Molecular Medicine, Oxford, UK) described how positional cloning has moved on from studies of relatively rare monogenic diseases towards the analysis of various polygenic disorders (which have a much higher incidence in the population). The presentation by Kay Davies (Institute of Molecular Medicine, Oxford, UK) used the example of Duchenne muscular dystrophy to illustrate how laborious and complex the analysis of a disease can become following the identification of the relevant gene.

Magarete Hohe (MDC, Berlin, Germany) presented a multiplex PCR sequencing approach for identifying sequence variations over large numbers of candidate DNA segments, and Hans Neubauer (Virchow Klinikum, Berlin, Germany) concentrated on clinical aspects of typing human leukaemia. Claire Huxley (St Mary's Hospital Medical School, London, UK) reported on the promising progress being made in the development of a mammalian artificial chromosome system, which has the potential to revolutionize gene therapy once human systems are available.

DNA sequencing

As at other recent meetings on this subject, there was a strong emphasis on large-scale sequencing. Progress reports by Mark Vaudin (Washington University, St. Louis, MO, USA), Andre Rosenthal (IMB, Jena, Germany), Karen Thomas and Sam Aparicio (Sanger Centre, Cambridge, UK) indicated the breath-taking progress in genomic sequencing, while Annemarie Poustka (DKFZ, Heidelberg, Germany) talked about the benefit of systematic gene identification.The concentration on existing sequencing methods was highlighted by the fact that there was not much talk about new technologies, apart from improvements in automated systems (Wilhelm Ansorge, EMBL, Heidelberg, Germany).

An overview by Werner Mewes (MIPS, Munich, Germany) on the sequencing of yeast which will be the first eukaryotic genome to be sequenced completely (expected early in 1996). The consequences of this, such as sequence-based functional analyses by the European Functional Analysis Network, fuelled discussion on how accurate a genomic sequence should be in order to be useful, with no obvious winner of the argument.

Costs of sequencing

An interesting part of the meeting were presentations by those industrial companies that are active in the field. Apart from slightly too much self-advertisement, some interesting points of view were raised.

A provocative statement was made by Thomas Pohl (GATC, Constance, Germany), who claimed that his company was able to sequence at a cost of only DM0.25 running costs per final base, if somebody would cover the other 85% of his company's expenses. His remark highlights the irrelevance of putting forward and comparing such numbers, irrelevant because no standard of accounting is set.

Social aspects of genome analysis

In this session, Ernst Peter Fischer (Constance University, Germany) discussed some of the issues relating to personal insurance that have been raised by the Human Genome project.

Overall, the first Beutenberg Symposium was a worthwhile meeting, which has the potential to develop into an influentional, international conference on genome analysis in the future.

Stephen W. Scherer is at the Department of Genetics, Rm 9102, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada, M5G 1X8.
Tel: +1 416 813 8140
Fax: +1 416 813 4931
e-mail: steve@genet.sickkids.on.ca

Jörg D. Hoheisel is at the Department of Molecular Genetic Genome Analysis, German Cancer Research Centre, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.
Tel: +49 6221 42 4680
Fax: +49 6221 42 4682
e-mail: j.hoheisel@dkfz-heidelberg.de