Abstracts of the 35th Annual Meeting of the Japan Society of
Human Genetics, 1990

Plenary Lectures

PL-I. THE GENE MAP AND GENOME ORGANIZATION

Jean FRÉZAL

Hôpital des Enfants Malades, France

There are two general strategies for mapping genes which are referred to as physical mapping and genetic mapping.

Physical mapping. The physical approach aims at defining where a gene actually lies on a chromosome. The first method is based on the analysis of *interspecific hybrids* resulting from the fusion of a cell from a heteroploid rodent line with a diploid human fibroblast or lymphocyte (1,2). It takes advantage of the random loss of human chromosomes over time to study the correlation between the presence of a chromosome and that of a marker, either a protein, an enzyme, or an antigen and more recently a DNA fragment. This abolishes the requirement that the markers be expressed in the parental cell, which was central to somatic cell hybridization. In addition, thanks to hybrids with rearranged chromosomes, it is also possible to restrict the assignment of a gene to a small portion of a chromosome, a so-called smallest region of overlapping.

In situ hybridization is achieved by denaturation of a chromosome preparation similar to the preparation used for karyotyping (3,4). The denatured chromosomes are annealed with a radioactive probe, revealed by autoradiography and stained. The use of non-isotopically (biotinylated) labelled probes offers markedly improved speed and spatial resolution (5,6).

Cytogenetic mapping. The contribution of cytogenetics to gene mapping was initiated by the assignment of the gene for red cell acid phosphatase to the tip of chromosome 2 (7). *Deletion mapping* culminated with the localization of the retinoblastoma gene on 13q14 (8) and the cloning of the Duchenne muscular dystrophy gene on Xp21 (9). Now the field has been extended by the detection of submicroscopic deletions, for instance in the Miller-Dieker syndrome (10), which are only detectable by molecular techniques.

Translocation has also contributed to mapping, first with exceptional female cases of X-linked recessive disorders, such as Duchenne, Hunter's disease, choroideremia or Lowe syndrome. Secondly, another everexpanding field concerns the numerous acquired and specific translocations observed in malignancies, which hint at specific genes likely to be involved in cell proliferation, differentiation and their control.

Genetic mapping. Genetic mapping relies on segregation analysis for coinheritance of marker and disease genes in families. In some instances, chromosomal rearrangements can provide powerful clues and prompt investigations toward a particular chromosome.

In other instances, the analysis may be directed toward loci for a candidate gene, *i.e.* a gene which can be suspected to be involved in the causation of disease. This approach is well illustrated by the example of osteogenesis imperfecta and collagen mutations (11).

Finally, comparative gene mapping can be helpful, considering that there is a striking number of conserved syntenic segments between species (12).

In the absence of any clue, one has to engage in a blind study, testing the cosegregation of the disease gene against a battery of polymorphic markers hopefully covering the entire genome.

Theoretically some 150 to 200 spaced markers would give enough landmarks, each 20 cM apart from its neighbours to demonstrate linkage (13). Primary linkage maps have been built. In fact many difficulties remain since available markers are not equally spaced and not optimally informative (14).

The second prerequisite is relevant to the family material necessary for a successful study. The size of the sample depends on the mode of inheritance, penetrance, age at onset. It is preferable, but not always possible, to study large pedigrees rather than to sum data across several smaller families (15). The reason is that no matter how consistent the disease phenotype is, there is no a priori guarantee that different families have defects at the same locus.

Spinal muscular atrophies. We should like to illustrate the strategy of linkage, using small families by the work on spinal muscular atrophies achieved by J. Melki in our group (16).

Among a large samples, 24 multiplex families with chronic forms 2 and 3, selected on stringent clinical criteria, were chosen for linkage analysis. Blood samples were obtained from patients (at least two in a family) as well as from their unaffected sibs and parents. Lymphoblastoid cell lines were established for all individuals, DNA extracted and digested with six restriction enzymes (six digests per individual).

After Southern blotting, every sample was tested against labelled probes. The data were computerized and analysed with the linkage package program from GM Lathrop. In a first step, 60% of the genome was excluded as the site of the gene. Eventually, the 133rd probe p105-153 Ra (D5S39) gave an unequivocal indication of linkage with a lod score above 6.00 (*i.e.* an odds ratio of 10^6 in favour of linkage against 1 for independence) and no apparent heterogeneity between types 3 and 2.

Secondly, it has been demonstrated that the acute form of SMA (type 1) is allelic to the chronic 2 and 3 forms (17) and a multipoint analysis gave a maximum likelihood location for the locus (common to the three forms) on the proximal part of the long arm of the 5th in the interval between the two loci D5S39 and D5S6. It is worth pointing out that these results are fully consistent with those obtained independently by the joint group of T.C. Gilliam and Kay Davies (18,19).

Heterogeneity. In a sense, these results put an end to the longstanding controversy concerning the heterogeneity of the disease. They strongly suggest that most cases are due to mutations at the same locus. However, this does not rule out the possibility of unusual cases unlinked to 5q, on dominant mutations and/or phenocopies as suggested by the low segregation ratio observed by several authors, especially in the so-called intermediate form. It must be remarked that the mode of ascertainment through multiplex families born from normal parents results in masking this kind of heterogeneity. In any case, this has to be kept in mind when considering an antenatal diagnosis for these conditions.

As a matter of fact, the problem of heterogeneity is central in any study of linkage in human beings. From that point of view, SMA, peripheral neurofibromatosis, Friedreich

ataxia or Huntington's chorea contrast with the peripheral neuropathies of the Charcot-Marie-Tooth types (at least two autosomal and one X-linked locus) and the tuberous sclerosis (two autosomal loci). On the other hand, different mutations at the same locus may induce entirely distinct phenotypes as exemplified by mutations at the COL2A1 locus.

Retinitis pigmentosa epitomizes all kinds of heterogeneity: clinical with different age at onset, diffusion of retinal lesions, course, and associated symptoms like deafness in Usher's syndrome. All modes of inheritance have been found. There are many sporadic cases. Some may actually be of recessive type, isolated by chance. Others may represent new dominant mutations suggested by the high paternal age as demonstrated by J. Kaplan from our group (20).

Mapped genes. These are just a few examples of the 400 or so diseases which have been mapped up to now out of a total of 1600 mapped genes (anonymous segments excluded) (21, 22, 23). A majority of those genes have been localized without reference to their mutations by physical methods. Cytogenetic mapping accounts for a quarter of the total. Emphasis is currently put on linkage studies for clinical disorders without any identified biochemical abnormalities. Gene mapping now appears to be the privileged method for isolating, cloning and identifying the relevant gene(s). The power of this approach, referred to as reverse genetics, is already highlighted by several examples such as Duchenne, retinoblastoma, cystic fibrosis and now neurofibromatosis.

It is a long and hard road (24)! Fortunately, before the goal has been reached and the gene characterized, the primary localization allows the definition of closely linked polymorphic markers. These polymorphisms can thus be used to detect the defective gene in families.

Genome organization. On theoretical grounds, gene mapping and cloning open new insights into the genome. One of the most unexpected findings arising from these advances surely lies in the progressive unveiling of the complexity of genetic information which is altogether ambiguous and redundant. A number of genes can give rise to several products mainly by alternative splicing of the primary transcript. Most genes are not unique but repeated in a number of copies, usually below a dozen, but large families and superfamilies are already known, such as for immunoglobulin and cytochrome P-450.

The members of a family may be clustered on a chromosome, as exemplified by the MHC on 6p21.3 (25, 26) and the regulators of complement activator on 1q32 (27). They can also be randomly dispersed as are the genes for actin or ferritin. A compromise is reached with the dispersed clusters of globins, immunoglobins, T cell receptors and homeoboxes among others.

The rule is that the expression of clustered genes are co-regulated, either temporally or spatially, which does not exclude co-regulation for dispersed genes. It is also a general rule that the genes coding for the sequential enzymes involved in a metabolic pathway or a cascade of reactions are dispersed on different chromosomes. Tight linkage of sequential enzymes is strongly suggestive but not proof of multigene determination as exemplified by purine and pyrimidine biosynthesis (28).

Although the genetic information is largely redundant, the coding part of the genes only accounts for a few per cent of the genome and the significance of the enormous amount of extra DNA is a matter of intense speculation. Obviously splitting the genes in pieces considerably extends the length of DNA that they cover on a chromosome. A further increment is provided by the flanking regions which carry on the sequences needed

for their expression and its regulation.

A few of these flanking regions are now characterized, such as the activating region identified 5' to the non-alpha globin cluster or chromosome 11p15.5 (29). Other characteristics are in limbo, such as the attachment sites to the nuclear matrix which probably play an important part in the expression of genes and the replication of DNA.

Finally, as is well known, the human genome includes several families of repetitive DNA, either clustered at the centromere (alpha satellites) or at the telomere (30,31). Others, like the CpG-rich HTF islands are located upstream of many genes, notably, the housekeeping ones (32).

Other repetitive elements are interspersed like the Sine's (Alu) and Line's families. Their distribution is uneven. Alu families dominate in R bands which are GC rich and where most genes are concentrated. On the contrary, L1 is AT rich and dominates in Giemsa/quinacrine-positive bands.

The minisatellites also referred to as VNTR (variable number of tandem repeats) are more numerous in the vicinity of the telomeres. The newly discovered microsatellites consisting of stretches of alternating purines and pyrimidines (CA/GT) in variable number, are highly polymorphic and evenly dispersed on all parts (mainly the informative ones) of the genome (33).

Those sequences are already being used in conjunction with amplification by PCR to hammer the new tools for building physical as well as genetic maps of higher resolution.

They also account for a substantial part of variation, being mutation prone. Very briefly it was thought that the CpG pairs were not spots of mutations and this has already been confirmed for several disorders *i.e.* the hemophilias A and B (34), phenylketonuria (35,36), *etc.* resulting mostly in nonsense mutations.

The occurrence of small repeats in close vicinity may favour slippage during replication, a mechanism which may account for the short deletion in PKU (37), CF (38) and globins (39). In the same way, slippage or unequal crossing over may account for the high proportion of deletions observed in disorders such as DMD or steroid sulfatase deficiency for which convincing evidence exists (40).

The examples of Alu-mediated rearrangements, either partial deletions or duplications, are multiplying. Finally, insertions have been documented due to L1 at least in one case.

REFERENCES

- 1) Puck, T.T. and Kao, E.T. 1982. *Ann. Rev. Genet.* **16**: 225-271.
- 2) Van Cong, N. *et al.* 1986. *Ann. Genet.* **29**: 20-26.
- 3) Gerhard, D.S. *et al.* 1981. *Proc. Natl. Acad. Sci. USA* **78**: 3755-3759.
- 4) Zabel, B.U. *et al.* 1983. *Proc. Natl. Acad. Sci. USA* **80**: 6932-6936.
- 5) Lan, Y.F. 1985. *Cytogenet. Cell Genet.* **39**: 184-187.
- 6) Lichter, P. *et al.* 1990. *Science* **247**: 64-69.
- 7) Ferguson-Smith, M.A. *et al.* 1973. *Nature New Biol.* **243**: 271-274.
- 8) Yunis, J.J. and Ramsay, N. 1978. *Am. J. Dis. Child.* **132**: 161-163.
- 9) Koenig, M. *et al.* 1988. *Cell* **53**: 219-228.
- 10) Van Tuinen, P. *et al.* 1988. *Am. J. Hum. Genet.* **43**: 587-596.
- 11) Prockop, D.J. *et al.* 1990. *Adv. Hum. Genet.* **19**: 105-132.
- 12) Nadeau, J.H. 1989. *Trends Genet.* **5**: 82-86.
- 13) Bottstein, D. *et al.* 1980. *Am. J. Hum. Genet.* **32**: 314-331.

- 14) White, R. and Lalouel, J.M. 1988. *Ann. Rev. Genet.* **22**: 259-279.
- 15) Gusella, J.F. 1986. *Ann. Rev. Biochem.* **55**: 831-854.
- 16) Melki, J. *et al.* 1990. *Nature* **344**: 767-768.
- 17) Melki, J. *et al.* 1990. *Lancet* (in press).
- 18) Brzustowicz, L.M. *et al.* 1990. *Nature* **344**: 540-541.
- 19) Gilliam, T.C. *et al.* 1990. *Nature* **345**: 823-825.
- 20) Kaplan, J. *et al.* 1990. *Hum. Genet.* **85** (in press).
- 21) Frézal, J. and Klinger, H.P., eds. 1987. Ninth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.* **46**: 1-762.
- 22) Frézal, J. *et al.* 1989. *Genatlas*. INSERM/John Libbey, eds. London/Paris.
- 23) Kidd, K.K. *et al.* 1989. Tenth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.* **51**: 1-1147.
- 24) Dean *et al.* 1990. *Nucleic Acids Res.* **18**: 345-350.
- 25) Carroll, M.C. *et al.* 1987. *Proc. Natl. Acad. Sci. USA* **84**: 8535-8539.
- 26) Dunham, I. *et al.* 1987. *Proc. Natl. Acad. Sci. USA* **84**: 7237-7241.
- 27) Carroll, M.C. *et al.* 1988. *J. Exp. Med.* **167**: 1271-1280.
- 28) Jones, M.E. 1980. *Ann. Rev. Biochem.* **49**: 253-279.
- 29) Enver, T. *et al.* 1990. *Nature* **344**: 309-313.
- 30) Korenberg, J.R. and Rykowski, M.C. 1988. *Cell* **53**: 391-400.
- 31) Bernardi, G. 1989. *Ann. Rev. Genet.* **23**: 637-661.
- 32) Bird, A.D. 1986. *Nature* **321**: 209-213.
- 33) Weber, J.L. and May, P.E. 1989. *Am. J. Hum. Genet.* **44**: 388-396.
- 34) Antonarakis, S.E. 1988. *Adv. Hum. Genet.* **17**: 27-59.
- 35) Abadie, V. *et al.* 1989. *Genomics* **5**: 936-939.
- 36) John, S.W.M. *et al.* 1990. *Am. J. Hum. Genet.* **46**: 970-974.
- 37) Caillaud, C. 1990. MD Thesis, Paris.
- 38) Kerem, B.S. *et al.* 1989. *Science* **245**: 1073-1080.
- 39) Efstratiadis, A. *et al.* 1980. *Cell* **21**: 653-668.
- 40) Yen, P.H. *et al.* 1990. *Cell* **61**: 603-610.

PL-II. THE HUMAN GENOME PROJECT IN JAPAN: CURRENT STATUS AND PROSPECTS

Nobuyoshi SHIMIZU

Department of Molecular Biology, Keio University School of Medicine, Shinjuku-ku, Tokyo 160, Japan

The human genome consists of 3 billion basepairs (bp) of DNA which are distributed on 24 different chromosomes. Mapping of the estimated 50,000-100,000 genes and sequencing of the entire human genome has become a major interest in medical genetics and human molecular biology. To achieve this goal, a complete understanding of the genetic blue print of human beings, human genome projects have been established in several different countries and the international cooperation is under way. At this congress, I have presented a brief history of the development of the human genome program in Japan and a summary of the current status of the organization and administration of the Monbu-sho (the Ministry of Education, Science and Culture: MESC) component of the genome program. The MESC's current human genome program consists of five selected topics: 1. human genome mapping, 2. human genome function, 3. DNA technology, 4. bioinformatics, and 5. nonhuman genomes. I have commented on the relationship of this pro-