

# Gene mapping today: applications to farm animals

M Gillois

*Institut National de la Recherche Agronomique,  
Laboratoire de Génétique Cellulaire, BP 27, 31326 Castanet-Tolosan, France*

(Proceedings of the 9th European Colloquium on Cytogenetics of Domestic Animals;  
Toulouse-Auzeville, 10-13 July 1990)

chromosomes / gene mapping technologies

## INTRODUCTION

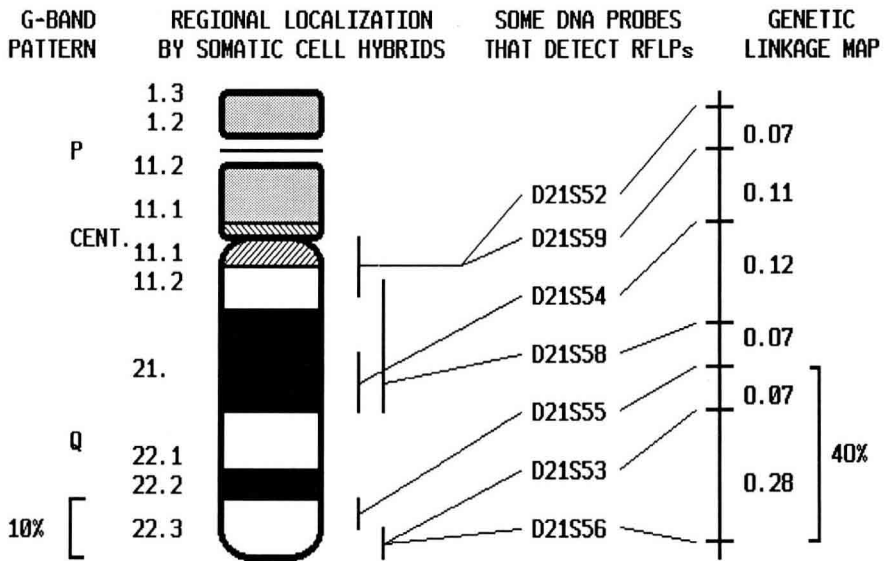
The genetic information contained in mammalian cells is spread over a set of nuclear chromosomes and the mitochondrial DNA. Converging evidence has made it possible to estimate that the mammalian genome may contain fifty thousand structural genes, which may be grouped in some three thousand to fifteen thousand clusters. Mapping the genome involves partitioning this number of genes into maps of individual linkage groups and determining their linear order on each of the chromosomes.

Dissection of the nuclear genome for mapping can be achieved by meiotic segregation in families, segregation in somatic cell hybrids, natural tagging of particular chromosomes in individuals with chromosome aberrations, by sorting of individual metaphase chromosomes, by analysis of subchromosomal fragments using somatic cell genetics and/or genomic or gene sequencing by molecular biology techniques.

During this lecture I wish to discuss the four following points: firstly, the identification of the relations which appear between the gene map and the chromosome structure; secondly, the evaluation of possibilities, complementarities and limitations of gene mapping techniques; thirdly, an overview of the status of the gene maps of cow, sheep, pig and rabbit; fourthly, the most important applications of gene mapping.

## RELATIONS BETWEEN THE CHROMOSOME STRUCTURE AND GENE MAP

It is classical to consider a genetic map obtained by Mendelian segregation analysis and a physical map obtained by manipulation and cytological observation of chromosomes (fig 1).



**Fig 1.** Physical and genetic map of human chromosome 21. RFLP: restriction fragment length polymorphism.

But the term chromosome is slightly ambiguous because sometimes we think about the double-stranded DNA and sometimes we think about the metaphase chromosome. The cytological structure of a metaphase or a pre-metaphase chromosome is very compact. The ratio of the length of the metaphase chromosome to the length of its DNA is about 1:8000.

A beautiful electron microscope preparation of a bare metaphase chromosome, published by Paulson and Laemmli (1977), shows a backbone of non-histone proteins that is responsible for the basic shape of metaphase chromosomes and the central scaffold organizing the DNA into loops throughout its length. These loops of DNA are anchored in the scaffolding and both ends of a loop appear to be anchored at the same place because the DNA strand emanates radially from the scaffold and returns to an adjacent point. Each loop contains from 30 000 to 100 000 base pairs (bp) of DNA (figs 2 and 3).

Using the technique of viscoelasticity, Kavenoff *et al* (1973) were able to demonstrate that the DNA of a chromosome is a single linear molecule. Consequently, we have to consider that along this DNA molecule there are specific sequence domains which are responsible for the anchorage of the DNA to the scaffold and for visible structures of chromosomes, such as the centromere or telomere. Curiously, repeated DNA sequences are situated in the region of contact between DNA and proteinaceous scaffold.

In a very interesting essay, Goldman (1988) considers that eukaryotic chromatin is organized into a series of loops approximately 20–330 kb in length, anchored to a protein matrix or chromosome scaffold at two points (the SAR or scaffold-attached regions). The SAR is typically a DNA sequence that is recognized by topoisomerase

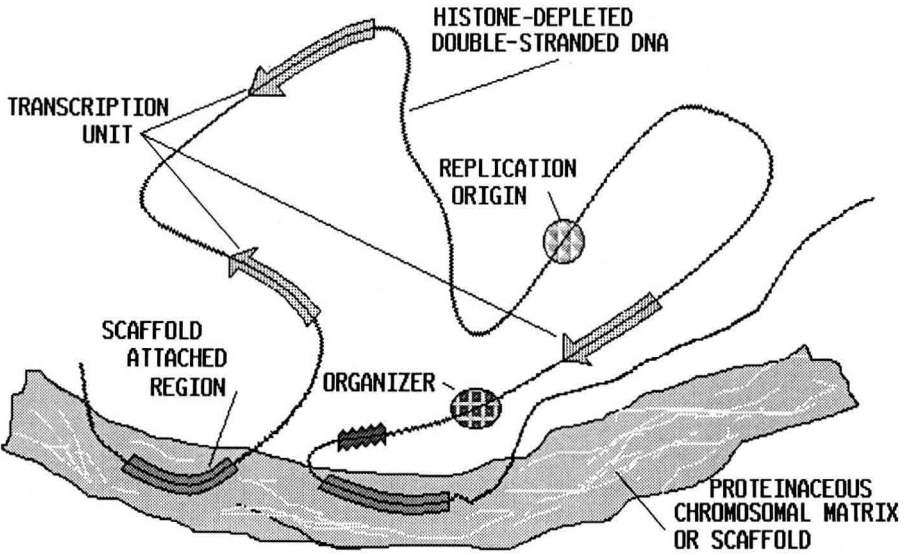


Fig 2. Diagram of a DNA loop or chromatin unit domain.

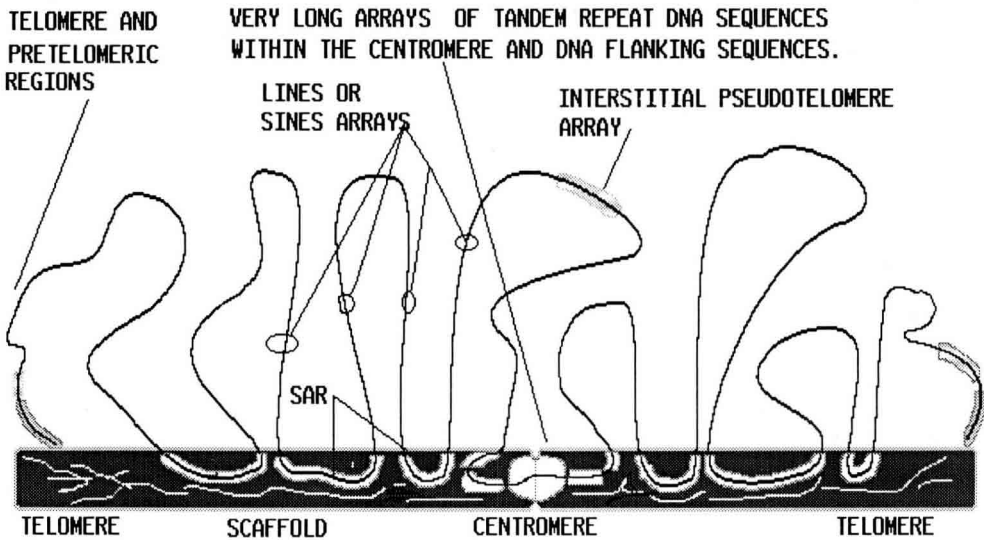


Fig 3. Diagram of a histone-depleted chromosome. LINES: long interspersed repeated DNA sequences; SINES: short interspersed repeated DNA sequences.

II. Each loop represents a unit of replication, having a single replication origin (RO) and a unit of supercoiling, in that its torsional state is independent of those of the surrounding loops. Each loop also represents a unit of function, in that a loop as a whole may be in a potentially active or repressed transcriptional state. There may be one or several transcription units in a loop. The organizer (ORG) is a DNA sequence that binds a transcription factor whose presence is required for assembly of active chromatin. The hypothesis has been advanced that, in mammals, transcriptionally competent domains replicate during the first half of the S phase, while transcriptionally incompetent domains replicate during the second half of the S phase. This explains the use of BrdU (5-bromo-2'-deoxyuridine) incorporation into DNA during chromosome dynamic banding.

Vogt (1990) in a recent paper points out that tandem repeated DNA sequence blocks which appear at many places in the DNA strand are responsible for the chromosome folding and share the chromatin codes.

The first code of the DNA sequence, detected by Lengyel *et al* in 1961, is of course the triplet code.

The second code of the DNA sequence was detected in 1980 by Trifonov and Sussman. They realized that pairs of adenine nucleotides occur with a 10.5 periodicity along any DNA sequence that interacts with the histone octamer unit and forms a nucleosome. This periodicity correlates well with that of the helical turn in the  $\beta$ -helical DNA secondary structure. This pattern of adenine nucleotides codes for a unidirectional curving of the DNA helix around the histone octamer complex. This code is the chromatin code.

Vogt enlarges on this idea, forming the concept of a 'chromatin-folding code'. He presents evidence that tandemly repetitive DNA sequences have the potential of expressing a locus-specific organization of their folding structures in the nucleus in which three characteristics are notable: 1) these repetitive sequences are able to attract specific nuclear proteins; 2) these repetitive sequences are able to develop a locus-specific, repetitive, higher order sequence structure because of their inherent possibilities of quick sequence alteration; 3) selection pressure created by the functional interaction between nuclear or chromosomal proteins and folded DNA stabilizes at each locus the molecular pattern of these specific repetitive sequences.

Salser *et al* (1976) define experimentally a locus-specific, repetitive DNA sequence structure whose sequence units have lost the ability to hybridize to related sequence units located at other chromosome positions.

Vogt introduces the chromatin code at the nucleosome level, the chromatin-folding code at the level of the centromere region, telomere region and scaffold region and he adds the gene expression code, the replication code and the recombination code.

In each species exists a so-called 'library code', which is a list of code sequences. The composition of the 'higher order' structure is species-specific and dictated by the combination of amounts of common basic sequence elements.

Two major types of tandem repeated DNA sequences are distributed in the genome (table I). For the first type, they are organized in large tandem block structures characterized by a repetitive head-to-tail arrangement of various sequence units, whose length can range from one to several thousand nucleotides. Their copy number is highly variable. For the second type, small variable tandem sequence

blocks, with a repetitive head-to-tail arrangement of the same basic sequence unit at one locus, can be clustered or interspersed in the genome and may contain related sequence elements.

**Table I.** Mammalian repetitive DNA sequence families.

---

Type I: long array

- 1.1 Satellite DNA sequence families (1-2-3-4)
- 1.2 Alphoid DNA sequence family
- 1.3 The Sau3A sequence family
- 1.4 The D22Z3 sequence family
- 1.5 The 71-7B sequence family
- 1.6 Low repetitive sequence families of the Y chromosome

Type II: short or medium array

- 2.1 The simple sequence families  $(T)_n/(A)_n$ ;  $(G)_n/(C)_n$
- 2.2 The simple sequence families with a trinucleotide sequence unit
- 2.3 The simple sequence families with a tetranucleotide sequence unit
- 2.4 The minisatellite sequence family  $(TTTTA)_n$
- 2.5 The telomeric minisatellite sequence family  $(TTAGGG)_n$
- 2.6 The minisatellite sequence 'super family' related to the chi-sequence of *E coli*
- 2.7 The minisatellite sequence families related to herpes virus genomes
- 2.8 The minisatellite sequence families related to adenovirus genomes
- 2.9 Other minisatellite sequence families

---

Now we are able to design a schematic representation of relations between the phase of DNA sequences and the histone-depleted chromosome (fig 3).

The current problem is how and why these various specific tandem repeat DNA sequences play a role in chromatin folding and gene expression.

Firstly, satellite DNA type I sequences are mostly located in the constitutive heterochromatin of the chromosome centromere region. Only the polymorphic heterochromatin of the long arm of the Y chromosome (region Y q12-Y qter, for the human chromosome) contains satellite III DNA fractions. The pattern of these repeats and the different sequence of units were extensively studied mainly by Prosser *et al* (1986). Some of these unit sequences of cores are chromosome-specific.

Secondly, the alphoid DNA sequence families are repetitive DNA sequences in the genome with a mean monomer length of 171 bp and clustered in a tandem repeat sequence organization. We know that each chromosome at its centromere region contains at least one specific alphoid super-repeat structure, often specified by a chromosome-specific tandemly repeated restriction site. The  $\alpha$ -protein in the nucleus is specifically attached to the alphoid DNA sequence structure supporting the specific phasing of nucleosomes in the centromeric chromatin.

Recently, Wevrick and Willard (1989) studied long-range organization of tandem arrays of  $\alpha$ -satellite DNA at the centromeres of human chromosomes and their meiotic transmission through three generations. They report that restriction site and array-length polymorphisms exist between individuals and appear to be meiotically stable on a megabase range. Similar results were reported by Jabs *et al* (1990).

Thirdly, Moyzis *et al* (1987) cloned a repeated sequence from human DNA that hybridized to human telomeres, and had the repeating units (TTAGGG). Since then Hastie and Allshire (1989) were able to give a schematic representation of the organization of the human telomere. Finally, sequences from the telomere are able to form guanine tetrads within an anti-parallel quadruplex, which may explain chromosome pairing (Sundquist and Klug, 1989). The recognition between homologous chromosomes could be ensured by specific flanking sequences of telomeres.

Interestingly, telomere-like repeat stretches may also occur in interstitial positions and may make the chromosome prone to breakage, fusion, fragile sites and recombination.

In conclusion, we have to identify DNA sequences which are responsible for the chromatin-folding code and locate them. For example, Collick and Jeffreys (1989) identified a 40 kDa mouse protein (Msbp-1) that bound to multiple copies of minisatellite core sequences. Similar proteins were found in a number of other species. Among these proteins, we can expect to identify couples of DNA sequences and proteins that are responsible for chromatin folding. Furthermore, we can expect to find reasons for the specific differences between chromosomes.

## EVALUATION OF GENE-MAPPING TECHNIQUES

### Use of somatic cell genetics in gene mapping

Gene mapping by somatic cell hybridization is based on a series of parasexual events involving fusion of somatic cells from different species, fusion of their nuclei and mitotic segregation of part of the genetic material of one of the parents during the proliferation of the fusion products.

In clones of cell hybrids retaining the complete genome of one parent (recipient genome) and one or several chromosomes of the segregating genome (donor genome), the resulting concurrent segregation of biochemical phenotypes (markers) indicates that their encoding genes are located on the same chromosome. It is the operational definition of synteny (fig 4).

Formation and isolation of somatic cell hybrids by selection and cloning, and techniques for molecular and cytogenetic analysis of the hybrid cells, are now routine methodologies.

A pragmatic approach has been adopted in order to create a multiple chromosome mapping panel. A number of independent hybrid clones are generated and, after a period of growth, analyzed for the presence of donor chromosomes by means of cytogenetic and molecular techniques. From this collection, a series of clones are chosen as a mapping panel based on various criteria, such as their donor chromosome content, their relative stability, the numerical relationship between their donor and recipient genomes and their growth characteristics. Clones which continue to segregate donor chromosomes at a high rate are discarded.

Some difficulties in this approach cannot be avoided. For example, specific combinations of donor chromosomes tend to co-segregate in hybrid cells, whereas other combinations are particularly rare. In addition, both donor chromosomes and recipient chromosomes can rearrange.

PRIMARY CELL = DONOR CELL

ESTABLISHED CELL = RECIPIENT CELL

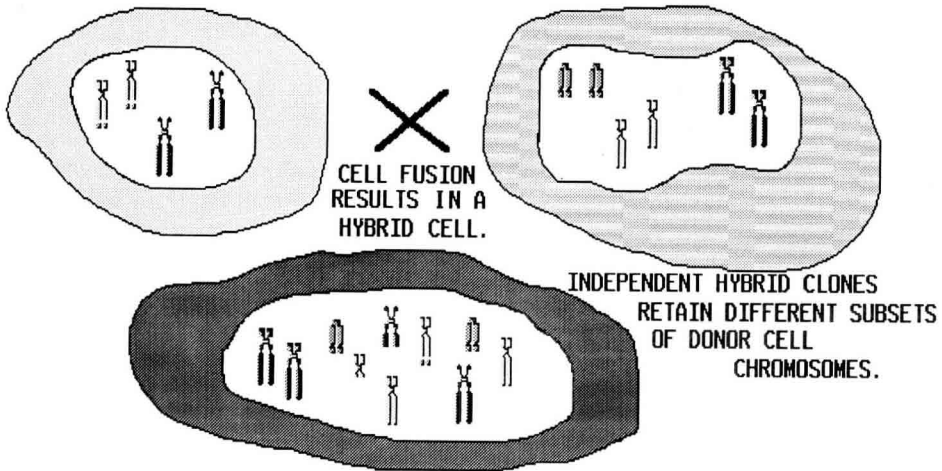


Fig 4. Somatic cell hybridization for synteny analysis.

Consequently, statistical rules concerning synteny or independence between markers are needed.

If two markers, X and Y, are borne by the same chromosome, a set of  $n$  hybrid clones is expected to be made up of  $a$  doubly positive clones (those carrying the corresponding chromosome) and of  $n - a$  doubly negative clones (those having lost the carrying chromosome). Conversely, if both markers are carried by two independent chromosomes, it is expected that all four classes of responses will be found in a set of hybrids. The numbers of each class are usually denoted by  $a$ ,  $b$ ,  $c$  and  $d$  (fig 5).

	$Y^+$	$Y^-$
$X^+$	$a$	$b$
$X^-$	$c$	$d$

**STATISTICAL DECISION RULES ARE NEEDED.**

Fig 5. Statistical decision rules concerning synteny or independence between markers. For  $n = a+b+c+d$ , expected values are such that  $b = c = 0$  for syntenic genes, or  $ad - bc = 0$  for independent segregation of chromosomes. In fact, these alternative biological hypotheses cannot be turned into a clear-cut mathematical statement: 1) some chromosome couples do not segregate independently; 2) discordant clones ( $X^+ Y^-$ ) and/or ( $X^- Y^+$ ) may occur at low frequencies for syntenic markers.

Expected values are such that  $b = c = 0$  for syntenic genes or  $ad - bc = 0$  for independent segregation of carrying chromosomes.

In fact, these alternative biological hypotheses cannot be turned into such clear-cut mathematical statements. Discordant clones may occur at low frequencies for syntenic pairs and a statistically significant departure from independence may occur for genes carried by distinct chromosomes. This is the reason why Chevalet and Corpet (1986) introduced a coefficient of correlation into the statistical framework accounting for association between chromosomes in their statement of decision rules.

Today, scientists are interested in the construction of somatic cell hybrids containing different overlapping deletions involving a particular chromosome. For example, Cowell and Mitchell (1989) have constructed a somatic hybrid cell mapping panel for regional assignment of human chromosome 13 DNA sequences. The assignment of chromosomal break points from karyotypic data alone is often subjective and alternative interpretations of G-banded appearances are possible. However, karyotypic analysis can exclude the presence of certain chromosomes in the hybrids (fig 6). To characterize the rearrangements more precisely, they used a set of chromosome 13-specific DNA probes.

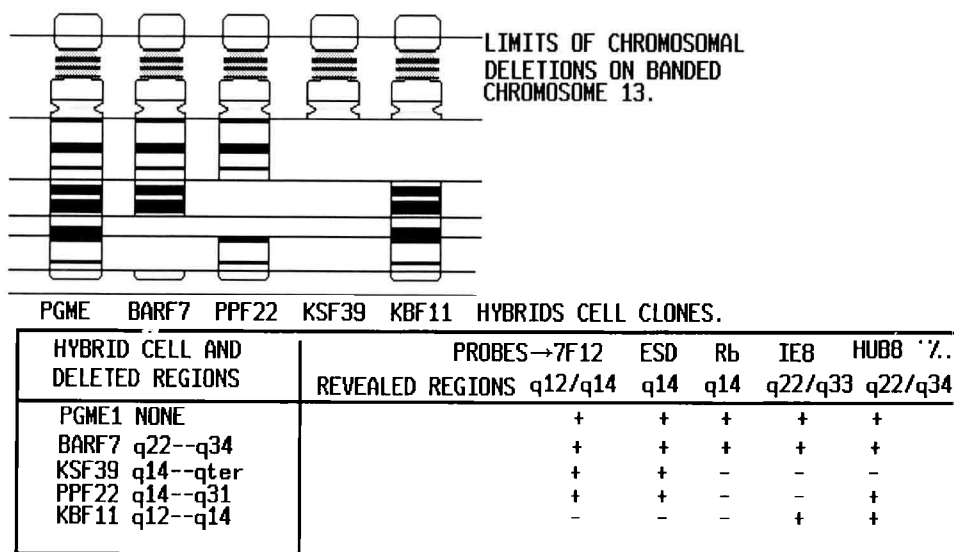


Fig 6. Human chromosome 13 mapping clone panel.

This type of approach, with natural or induced chromosome deletions, is useful to locate and furthermore to clone a gene responsible for a disease or a major inheritance trait.

The new trend in somatic cell genetics is to try to obtain single chromosome mapping panels. Hybrid clones containing single donor chromosomes can be generated by rare chance immediately after cell fusion or during the permanent process of chromosome loss from established cell line hybrids at a low frequency.



An alternative fusion procedure should produce hybrid cells with few or single donor chromosomes: microcell fusion. Metaphase nuclei are split into microcells or minisegregant cells which contain few or single chromosomes packed into cell membrane-derived envelopes. After fusion with a recipient cell, the production of hybrids with single donor chromosomes is obtained faster than with cell to cell fusion (Ege, 1980; Fournier and Ruddle, 1977; Tourian *et al*, 1978).

Unfortunately, it has turned out that, at some time during this procedure, donor chromosomes tend to be partially deleted or rearranged. Thus, the retention of intact chromosomes has to be carefully controlled. If hybrids with a single donor chromosome are obtained and maintained under selective pressure to retain this chromosome, fragmentation of the donor chromosome can occur. The consequence of these pitfalls is that it is difficult to obtain a complete single chromosome mapping panel.

For the near future, single chromosome mapping panels, with intact, deleted or rearranged chromosomes will be very useful for regional mapping using pulsed-field gel electrophoresis.

### Use of *in situ* hybridization in gene mapping

Molecular *in situ* hybridization involves preparing a probe of labeled DNA fragments corresponding to a gene or a part of a gene, previously isolated by a method of genetic engineering, and hybridizing this probe to the homologous DNA of the metaphase chromosomes on microscope slides. This technique is now well established, see the review of Gellin and Yerle (1989) and the following illustrations.

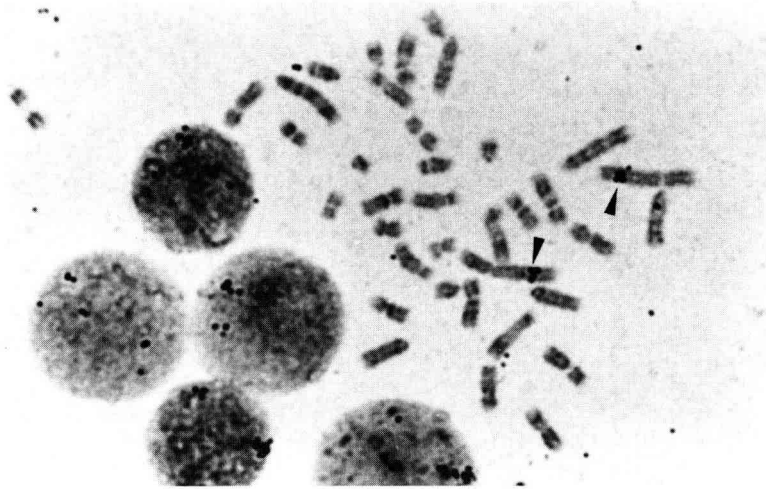
Figure 7 shows the *in situ* localization, performed by Yerle *et al* (1986) and Yerle and Gellin (1989), of the leukocyte-interferon gene in the q25 region of pig chromosome 1. Two spots are visible on the two homologues of chromosomes 1. A remarkable concentration of silver grains was located on the chromosome region 1q25 as seen on figure 8. A computer-workstation was used for *in situ* hybridization automatic quantitative analysis.

For non-isotopic *in situ* hybridization, two major applications have received considerable attention lately: gene mapping and interphase cytogenetics (fig 9).

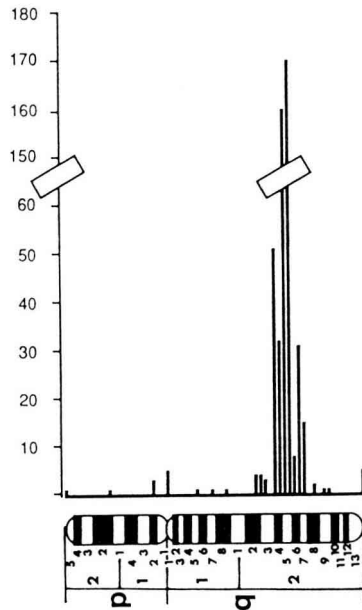
Several groups have reported the visualization of probes greater than one kilobase (kb) long using either conventional microscopy or digital-imaging microscopy. Probes of considerably greater genetic complexity have a high probability of containing interspersed repetitive sequences, usually widely distributed throughout the genome, which will result in a non-specific hybridization signal. However, protocols designed to suppress these unwanted signals with appropriate competitor DNA have been defined, and thus one can obtain highly specific delineation of the genome regions from which the probes have been derived.

With cosmid probes, nearly 90% of the target sequences in cell populations can be delineated by fluorescence and, on metaphase chromosome spreads, more than 80% exhibit specific signals on both chromatids of both homologous chromosomes.

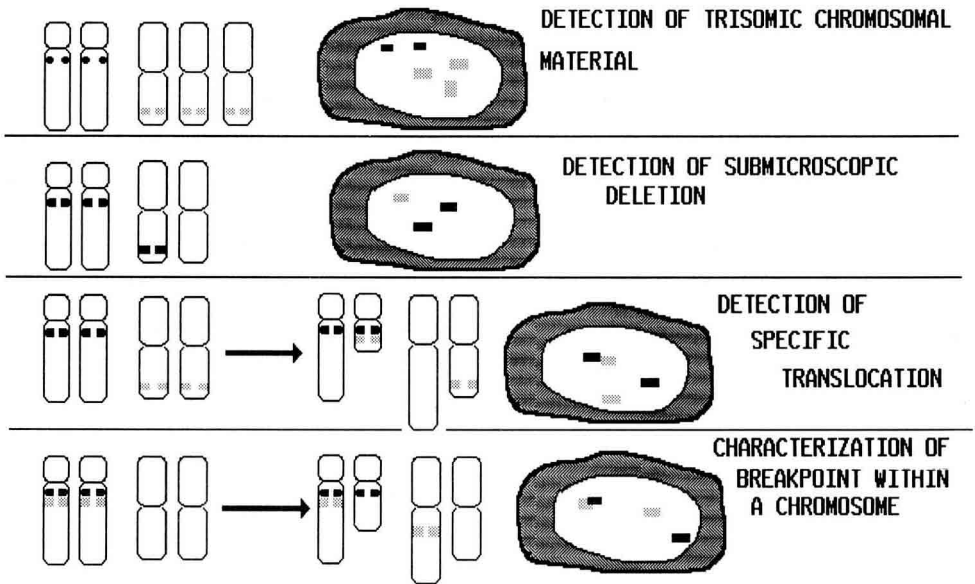
Chromosome assignment is not problematic, as several conventional chromosome banding methods are fully compatible with probe detection by fluorescence.



**Fig 7.** Pig metaphase spread after *in situ* hybridization with an  $\alpha$ -interferon probe. The arrowheads point to the grains on chromosome region 1q22-q27.



**Fig 8.** Distribution of silver grains on region q25 of pig chromosome 1.



**Fig 9.** Analyses possible with double-tagged fluorescent chromosomes.

Hybridization signals from two probes can be spatially resolved on metaphase chromosomes when the probes are only several hundred kilobases apart, but a minimum separation of one or two million base pairs is required to enable the establishment of their physical order.

To further improve the spatial resolution of gene mapping, co-hybridization studies with closely spaced probes from a single genomic region have been performed on interphase nuclei. Measurements of the distances between multiple pair probes, after hybridization to methanol-acetic acid-fixed flat nuclei, demonstrate a fairly linear relationship between physical distance and genome order over the range from thirty kilobase pairs to about one million base pairs.

The combination of metaphase and interphase nuclear mapping, particularly using multiple probes simultaneously, offers the opportunity to physically order genomic DNA segments with a resolution presently only achieved by gel electrophoresis methods and provides a new bridge to interrelate physical and genetic linkage information.

### Use of PFGE and YAC chromosome-cloning in gene mapping

Pulsed-field gel electrophoresis (PFGE) can resolve DNA molecules as large as several million base pairs in size. This is in contrast to conventional electrophoresis for which the practical upper limit ranges from 50 000 to 100 000 bp. This increased resolving power has especially important ramifications for the study of complex

genomes of large size. PFGE has made the cloning of large genes or groups of genes possible *via* the yeast artificial chromosome (YAC) method.

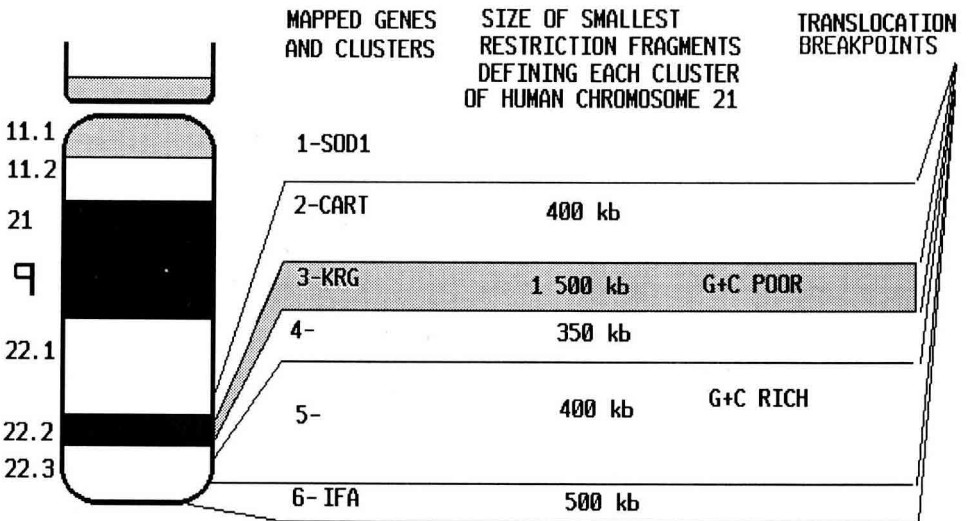
The essential technical innovation in PFGE is the use of electric fields that are not constant throughout the gel run, but that regularly alternate in direction. The frequency with which the change occurs, called the pulse time, dictates the size class of fragments to be resolved.

Because of the large size of molecules to be examined, DNA fragments for PFGE must be carefully prepared to prevent shearing. This problem has been neatly and quickly resolved by the practice of preparing DNA in agarose blocks or plugs.

The analysis of genomes of higher organisms requires restriction enzyme digestion. A special class of restriction enzymes is needed which must cut relatively infrequently, generating fragments in complete digests of between 100 and 1 000 kb in average size. For example, enzymes such as *NruI* (TCGCGA) or *NotI* (CCGGCCGG) give an excellent range of fragment sizes for pulsed-field work.

With an upper size limit of ten million base pairs, PFGE is obviously incapable of separating intact mammalian chromosomes. For example, the smallest human chromosome is about 50 million base pairs in size.

A good example of an application of this technique is the physical mapping of human chromosome 21, accomplished by Gardiner *et al* (1990). Human chromosome 21 at low resolution can be divided into the centromere-proximal half that is predominantly a Giemsa-dark band (q21) and the telomere-proximal half that is a Giemsa-light band (q22) (fig 10).



**Fig 10.** Pulse-field gel electrophoresis has increased by over 100-fold the size of DNA molecules that can be resolved.

DNA fragments from the q22 region digested by several enzymes (*NotI*, *BssHII*, *NruI*) average considerably less than 1 000 kb, whereas those from the q21 region average greater than 1 500 kb.

Furthermore, in more detailed analysis, it can be shown that the upper fragment's 1 500 kb from the q22 region map exclusively to the small Giemsa-dark band q22.2.

This observation is, in part, a reflection of fundamental differences between Giemsa-light and -dark bands. Dyes used in cytogenetic staining are base-specific and indicate that light bands are higher in G + C content than dark bands.

In addition, light bands are enriched for genes (or at least for those mapped so far) and, therefore, for the gene-associated CpG islands. It is logical that light bands, because of their different base composition, will contain more sites for rare cutting restriction enzymes.

Pulsed-field analysis can also provide information on both the sizes of mammalian genes and their organization. The entire long arm of human chromosome 21, 40 million base pairs, has been accounted for on a collection of 33 *NotI* restriction fragments, using some 50 unique sequence probes. Using complete digests, 70% of these probes have been physically linked in 13 separate groups.

Figure 10 shows data from Gardiner (1990) concerning the 6 physical linkage groups mapping to the distal third of the long arm. Together, these groups contain 14 of the 24 genes and 19 of 54 unique sequences used to construct the map. Consideration of the smallest restriction fragment that defines each group indicates that these genes are locally clustered, being contained within only 3 400 kb or approximately 8% of the long arm. Certainly, chromosome 21 contains many more than these two dozen genes, probably 500 or 1 000, but this present evidence strongly suggests that genes are not uniformly distributed.

In conclusion, pulsed-field gel electrophoresis has increased by over 100-fold the size of DNA molecules that can be quickly and easily resolved. Physical mapping using this technique yields new information on the relative proximity of individual genes, local clustering of groups of genes, and the molecular basis of cytogenetic banding patterns.

### Use of chromosome-crawling in gene mapping

Once the large-scale physical mappers have finished their hopping, jumping and walking on the chromosome, there is left the difficult task of establishing the content of large segments of genomic DNA. Although there has been much discussion about new methods for large-scale mapping, approaches designed to identify new expressed sequences have only recently been considered.

Two groups have now mounted detailed researches along large segments of the MHC (major histocompatibility complex), on the short arm of human chromosome 6, looking for expressed unknown sequences.

Spies *et al* (1989) probed Northern blots directly with cosmids, screening for the presence of transcripts in B cells, T cells, monocytes, epithelial cells and fibroblasts. Using this strategy, they identified five new transcribed loci and proceeded to obtain cDNA clones for each. All transcripts were constitutively expressed in all the tissues.

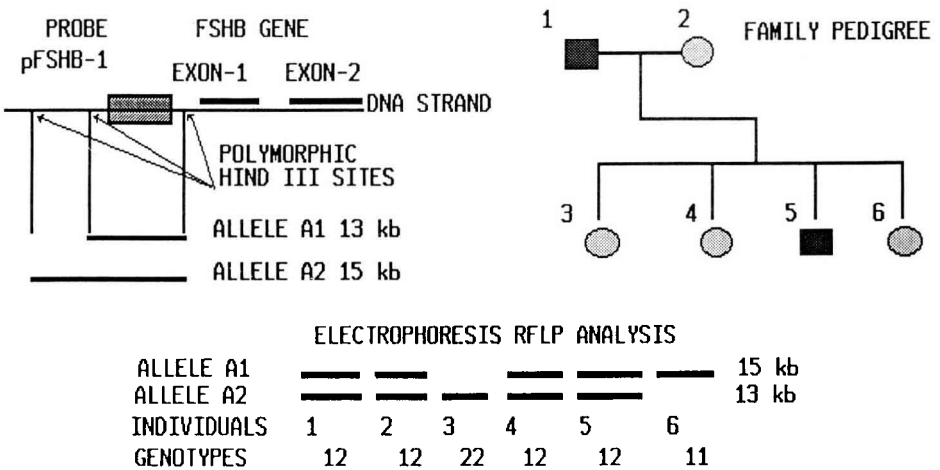
Sargent *et al* (1989) used a more thorough approach for identifying sequences. They first identified HTF (*HpaI* tiny fragment) islands within the cloned DNA. These CpG-rich regions were identified by the presence of clusters of restriction sites for enzymes *BssII*, *EagI* and *SacII*, and the lack of methylation of these sites in genomic DNA. Such sequences are commonly found adjacent to expressed

housekeeping genes. Sequences adjacent to these HTF islands were then used to probe Northern blots to identify the transcribed sequences. This approach revealed 12 new genes and established a density of such genes of one every 23 kb of DNA between the C4 and the tumor necrosis factor (TNF) loci.

Even if they have not identified the full complement of expressed loci in this region, these two studies are important because they provide an impression of the high density of expressed sequences in particular genomic regions. The MHC is found in a Giemsa-light or G + C-rich region of the genome, consistent with this very high frequency of housekeeping genes. Similar high gene density can be expected in other genomic regions associated with a Giemsa-light band or G + C-rich isochores.

**Use of DNA probes and pedigree analysis in gene mapping**

DNA probes become genetic markers when they are capable of detecting genetic variation between individuals at the level of their chromosome DNA sequences. The genetic variation is revealed as heritable DNA restriction fragment length patterns that differ between individuals when their DNA is digested with a restriction endonuclease and hybridized to the same specific DNA probe *via* the Southern blot technique (fig 11). The inheritance of these restriction fragment length polymorphisms (RFLPs) can be used to track the inheritance of small specific chromosome regions, once the RFLP-detecting DNA probe has been mapped to a chromosome. The RFLP-detecting DNA probe has to be assigned to a specific chromosome, and located in a unique locus, by somatic cell genetics and/or molecular *in situ* hybridization on the metaphasic chromosome.



**Fig 11.** Mendelian segregation of an RFLP associated with the human follicle-stimulating hormone  $\beta$ -subunit gene (FSHB).

Usually the band size differences on the electrophoretic gel are a direct result of variation in the DNA sequence comprising the enzyme recognition site. Point muta-

tions, or deletion or insertion of tracts of DNA sequence within the locus recognized by the DNA probe generate this Mendelian variation between individuals.

It is important to point out that two RFLPs of equal lengths provided by the same enzyme digestion can show internal genetic differences in each DNA sequence if they are digested again with another restriction enzyme. This internal mapping technique, developed by Jeffreys *et al* (1990) for DNA fingerprints, revealed genetic variation on an extraordinarily fine scale within a single locus and it is well suited to track the genealogy of a gene.

Figure 11 illustrates a simple type of RFLP associated with the gene for the  $\beta$ -subunit of the follicle-stimulating hormone (FSHB) located on the short arm of human chromosome 11 (11p). This is the common form of RFLP. It results from a single site mutation that alters the restriction endonuclease cleavage site, in this case for the enzyme *Hind*III. The FSHB DNA probe detects two different restriction fragment lengths, which represent the alleles. An individual who possesses both alleles has a different-sized FSHB restriction fragment on each chromosome of his chromosome 11 pair.

This approach is especially well suited for physically approaching or examining genes that are candidates for specific Mendelian inherited phenotypes, such as halothane sensitivity in pig or Booroola prolificacy gene in sheep. But first, it is necessary to find some RFLPs which are strongly correlated with the Mendelian phenotype.

In order to attain this aim, the realization of a low-density map of the genome is needed, such as is the case for the Booroola gene.

Sometimes, comparative gene mapping gives us indications which reduce the needed mapping to a region of a unique chromosome, which is the putative carrier. This is the case with the halothane gene, which may be located to the cen $\rightarrow$ q21 region of the pig chromosome 6, in the vicinity of five mapped clustered genes, as Yerle *et al* demonstrate in these proceedings.

The goal of using RFLPs to study the inheritance of a genetic disease or a clear Mendelian phenotype is to identify the specific chromosomal region that contains a particular gene.

Dozens or possibly hundreds of RFLP fragments may need to be tested in family pedigrees before linkage can be detected with the putative carrier chromosomal region.

Figure 12 shows a simple scheme of inheritance of two linked RFLPs. The mother's chromosomes, identified with filled circles for centromeres, are indistinguishable because each chromosome has the same allele for both the (A1,A1) or (B1,B1) locus. The father's chromosomes, centromeres shown as filled squares, can be distinguished because he is heterozygous (A1,A2), (B1,B2) at the two loci. In addition, the allele association pattern (or phase) of the father's chromosomes can be determined from his parents's chromosomes: the grandmother contributes an A1/B1 chromosome, the grandfather contributes an A2/B2 chromosome. A mating of this type, where one parent is doubly homozygous and the other is doubly heterozygous, is termed a double backcross. Recombinant (R) and non-recombinant (NR) chromosomes can be counted directly from double backcross matings as shown in the figure, when phases can be determined. This points out the usefulness of the three generation pedigrees.

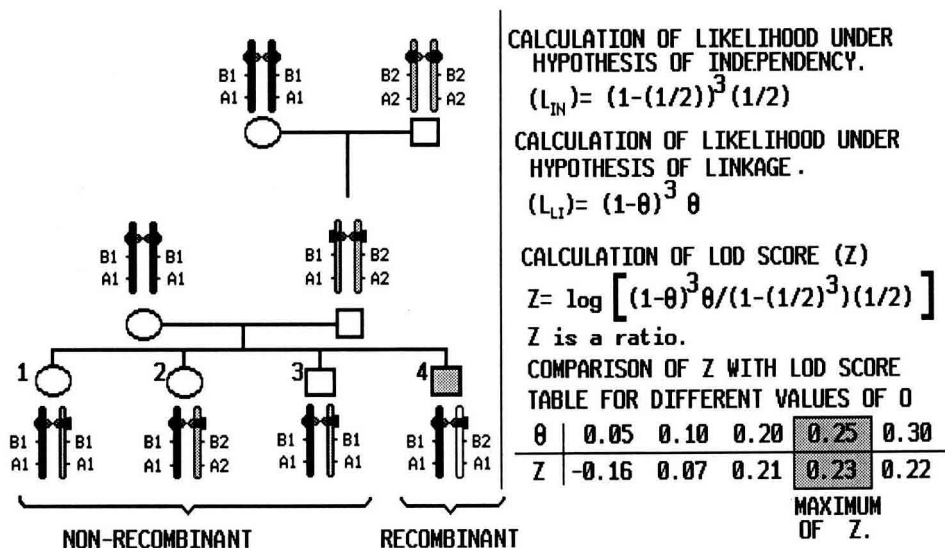


Fig 12. Rationale of linkage analysis with RFLP or variable number of tandem repeat (VNTR) probes.

Individuals 1, 2 and 3 have inherited NR chromosomes from their father. Individual 4 has inherited an R chromosome (A1/B2) from his father. The number of R chromosomes divided by the number of chromosomes yields the recombination fraction  $\theta$ , the probability with which a recombination occurs, and the measure of the genetic distance between markers: for this simple family pedigree,  $\theta = 1/4 = 0.25$ .

The LOD score method is conventionally used to quantify the data support for linkage; in this example the LOD score reflects the support for linkage between A and B loci. The LOD score is the logarithm (to the base 10) of the ratio of likelihoods ( $L$ ) calculated from the family pedigree data in two models: one model assumes linkage at different values of  $\theta$  [ $L(\theta)$ ], the other model assumes no linkage, free recombination, or  $\theta = 1/2$ ,  $L(1/2)$ .

The LOD score is calculated at incremental values of  $\theta$ . The recombination fraction ( $\theta$ ) associated with the maximum LOD score is accepted as the best estimate of the recombination between markers.

A lot of factors complicate the linkage analysis: numerous pedigrees, complexity of these pedigrees which includes numerous offspring, intricate consanguinity loops, and a number of individuals for whom the marker phenotype is unknown.

Fortunately, good computer programs are available, such as the program LINKAGE of Lathrop and Lalouel (1988) or the program MAPMAKER-QTL of Lander and Green (1987). Pairwise, or two-point, linkage analysis of sets of RFLPs that map to the same chromosome is first used to identify clusters of closely linked markers. The order of clusters of the chromosome or of closely linked RFLPs within a cluster is resolved with multilocus linkage analysis. Multilocus analysis, a method that examines the linkage relationships of multiple loci simultaneously, is a power-



ful technique that is used to resolve the order of markers on the map by assigning different probabilities to different marker orders.

There is another approach to increasing the informativeness of marker loci, the multiallelic multiloci variable number tandem repeat system (VNTR).

Variations in the number of short, tandemly repeated sequences found at a number of loci constitute an alternative class of DNA sequence polymorphisms in mammalian DNA. The number of copies of the tandem repeats can vary from one to several, to hundreds. The restriction fragment carrying the set of tandem repeats will probably appear in a number of different lengths, defining multiple distinguishable alleles at a single locus in a population sample. Heterozygosities over 90% have been identified so far.

A large number of VNTR loci have now been identified as a consequence of their partial homology with synthetic oligonucleotides, whose construction was based on DNA sequences from known VNTR loci or on a similar sequence.

Resolving the alleles of the system requires care, however, because the differences in fragment length among alleles can be as small as a single repeating unit. Units are known to range from 14 to 70 bp at different loci.

If at all possible, it is desirable to define the alleles at a VNTR locus in terms of the number of repeats carried, which makes it much easier, for example, to compare the alleles found in different branches of the same pedigree.

Scientists expect, with a unique VNTR DNA probe, to be able to study 15 to 30 loci simultaneously and to realize a low-density map of the genome. It is a necessary prerequisite to track a chromosomal region bearing an unlocated gene such as the Booroola gene. On the other hand, they expect also to track the genealogy of a gene in a single locus with the internal mapping technique developed by Jeffreys *et al* (1990).

### **Use of microdissection of metaphase chromosomes and sequence-independent amplification of DNA**

Many Mendelian genes, which are responsible for diseases, major traits or quantitative inheritance, are difficult to isolate by current cloning strategies, such as chromosome walking, even though the physical location of other closely linked genes has been determined. This is largely due to the difficulty of chromosome walking in the complex genomes of higher organisms which contain abundant repetitive DNA sequences. This same problem complicates the construction of complete physical maps and sequencing efforts.

Physical dissection of metaphase chromosomes is the most promising approach to this problem. Scalenghe *et al* (1981) were the first to dissect and clone DNA from *Drosophila* polytene chromosomes. Later, microdissection and microcloning were improved (Pirrotta *et al*, 1983), and completed by universal enzymatic DNA amplification (Ludecke *et al*, 1989).

Senger *et al* (1990) published a very well-documented review of the present state of the art. They point out that the conventional microdissection procedure has been improved mainly by the use of the pipette method for chromosomal harvesting, which avoids DNA contamination from broken cells; the dissection of GTG-banded chromosomes, which allows precise excision of single bands; and a modified

dissection technique, which does not use an oil chamber. These improvements have several advantages: straight needles and pipettes extended with the help of a simple pipette puller are sufficient; it is possible to dissect under the best directional and optical conditions and chromosomes can be placed perpendicular to the dissecting needle; banding pattern and thin needles required for dissection of single bands are much easier to visualize in air than in oil.

Ludecke *et al* (1990) show that to construct band-specific DNA libraries, 20 to 40 dissections per region are sufficient.

We can expect that this procedure will be performed soon for cloning bands in the cen-q12-q23 region of pig chromosome 6, in order to isolate the HAL-gene.

This technique seems more promising than the sorting of metaphase chromosomes, which provides very small amounts of DNA for analysis and needs expensive chromosome-cell sorter apparatus.

### **AN OVERVIEW OF GENE MAPS OF COW, SHEEP, PIG AND RABBIT**

Apart from the human gene map, only the gene map of the mouse is well advanced. For other species investigated, knowledge is still limited, particularly for farm animals, such as cattle, sheep, pig and rabbit. As shown on tables II-V, (Genetic Maps, 1990), a few linkage groups, 1-20 syntenies and a few loci assigned or regionally mapped are known for each of these species.

### **SOME REMARKS ON APPLICATIONS OF GENE MAPPING**

Eight applications are of some interest: 1) comparative gene mapping; 2) understanding relations between cytogenetics and molecular structure of chromosomes; 3) mapping a chromosome to find and to isolate a gene; 4) drawing real pedigrees of genes from pedigrees of zygotes; 5) genetic dissection of quantitative or complex inheritable characters; 6) genetic improvement based on selection of marker genes; 7) describing at the DNA level the molecular response to a selection pressure; 8) considering the maternal effect as a genetic imprint.

#### **Comparative gene mapping**

A beautiful illustration of comparative gene mapping has been published by Xu and Hardison (1989). Dutrillaux *et al* (1980) found considerable homology in the overall chromosome-banding patterns of rabbits and humans, which reflects a similarity in gene content in several regions, as shown by Soulie and de Grouchy (1982). In particular, Dutrillaux *et al* detected a correspondence between rabbit 1q and human 11 chromosomes, and these data confirm the presence of homologous genes in rabbit 1q and human 11p chromosome regions.

Another excellent example is presented by Yerle *et al* in these proceedings: the group of five genes: GPI, APOE, TGF $\beta$ , ENO1 and PGD, which are located in the cen-q25 region of pig chromosome 6, corresponds to two groups GPI, APOE, TGF $\beta$  located on the human 19 and murine 7 chromosomes and ENO, PGD located on the human 1 and murine 4 chromosomes.

**Table II.** The gene map of the cow (*Bos taurus*),  $2n = 60$ .

<i>Syntenic group</i>	<i>Chromosome</i>	<i>Gene locus</i>
U1		PGD, ENO1, AT3, ABLI
U2		SOD2, ME1, PGM3
U3	5	GAPD, LDHB, TPI1, PEPB, IFNG, A2M, 1NT1, HOX3, LALBA, KRAS2, GLI, PAH, NKNB, KRT2, GDH, LYS
U4	21	MPI, CYP11A, FES
U5		PKM2, NP, HEXA, FOS, KRT8L
U6		PGM1
U7		LDHA, TYR
U8		MDH2, ASL, PRM, GUSB, HAB
U9	18	GPI, DIA4
U10		SOD1, IFREC, PRGS, PAIS, CRYA1, SST, APP, ETS2, S1OOB, COL6A1, COL6A2, CBS, GAP43, PFKL, CD18, TF, CP
U11		1TPA, ADA, VIM
U12		ACY1, RHO
U13		HOX1, MET, COL1A2
U14		GSR, PLAT
U15	6	PGM2, PEPS, CASAS1, CASAS2, CASB, CASK, ABL, ASS, GRP78, LGB, J
U17	8	IDH1, FN1, CRYG
U18		ACO1, IFNA, IFNB, GSN, GGFB2, ALDOB, ALDH1, C5, ITI, NEFM
U19	15	CAT, A, PTH, HBB, CRYA2, FSHB
U20	23	GLO1, CYP21, BOLA-A, BOLA-B, BOLA-D, PRL, TCP-1, M, HSPA1
U21	19	GH, HOX2, KTR1, TK
U22		AMH, SPARC
U23		ALDH2, IL2, ILG
U24	14	TG, MOS, CA2, MYC, CYP11B
U25		PEPA
U26		GOT1, CYP17, ADRA2R
U27		POLR2, UMPH2
U28		MBP
U29		—
X	X	G6PD, HPRT, PGK1, GLA, F9, DMD
Y	Y	DYZB, DYZ1

Linkage groups not assigned to syntenic groups or chromosomes: [LGV], [LGVII], [ALB, GC], [S, P12].

These recent biochemical and molecular methods used in gene mapping have dramatically increased the number of homologous genes that have been mapped in more than one species. There are two well-documented reviews: one by Nadeau (1989) compares maps of linkage and syntenic homologies between mouse and man, and the other by O'Brien *et al* (1988) gives an evolutionary overview of mammalian genome organization.

**Table III.** The gene map of the sheep (*Ovis aries*),  $2n = 54$ : linkages, syntenies and assignments to chromosomes.

C, I	Linkage group I
PGD, ENO1, PEPC	U1
ACO1	U2
NP, PKM2	U3
PGM3, ME1	U8
MP1	U9
SOD1, PRGS, PAIS	U10
IDH1	U11
LDHA	U13
MDH2	U14
ADA, ITPA	U15
GSR	U16
HBG	U17
CASB, CASK	U18
AMH	U19
TPI, LDHB, PEPB, GAPD, LALBA, SHMT, KRTB(3q14-22), IGF1	chromosome 3
KRTA(11q24-28)	chromosome 11
GH	chromosome 11q25-qter
GPI	chromosome 14q22-q24
FSHB	chromosome 15q24-qter
OLA(20q12-23)	chromosome 20
RNR (NOR)	chromosomes 1, 2, 3, 4, 25
G6PD, HPRT, GLA, PGK	chromosome X
OTC, RCP, ARAF1, PLP, SYN	chromosome X

U is used for unknown chromosome. The syntenies were established by somatic cell hybridization. The RNR localizations were obtained using silver staining, and the regional mapping of genes by *in situ* hybridization.

An important application of comparative gene mapping involves predicting the location of genes in one species given their locations in another species. Locations of murine and human genes are the main operational guide used in domestic animal gene mapping.

### Understanding relations between cytogenetics and the molecular structure of chromosomes

Fifteen to twenty years ago, staining techniques were discovered which produced differential staining of parts of chromosomes. This revealed that euchromatin is heterogeneous and divided longitudinally into large domains or 'bands' which may have functional and structural significances. These chromosome-banding methods have been essential tools in the construction of physical gene maps of vertebrates. In this meeting, some examples are presented. Advances in molecular biology and gene mapping are beginning to point to features of genome organization that may

**Table IV.** The gene map of the pig (*Sus scrofa domestica* L),  $2n = 38$ : linkages, syntenies and assignments to chromosomes.

Group I	K, HPX
Group II	AM, I
Group III	PII, PO1, B, IGH1, IGH2, IGH3, IGH4
Group IV	CAS fractions
Group V	ELF, TF, CP
Group VI	SLB, L
Group VII	LP-B, LP-T, LP-V
Chromosome 1	IFNA(1q25)
Chromosome 2	MDH1
Chromosome 5	LDHB
Chromosome 6	GP $\beta$ , HAL, S, H, A1BG, PGD, APOE, TGFII, ENO1, CRC
Chromosome 7	SLA(7p14 $\rightarrow$ q12), CYP21, C4, J, C, MPI, PKM2, NP(q21-q22) PEPN, TNF $\alpha$
Chromosome 8	RNR(NOR)
Chromosome 9	SOD1
Chromosome 10	RNR(NOR)
Chromosome 12	GH, HOX2
Chromosome X	(SPL), TRAI, HPRT, G6PD, PGK, GLA

The linkage groups were established by family studies. Two of them were assigned to chromosomes: [SLA, J, C] Chr 7 and [GPI, HAL, S, H, PO2, PGD] Chr 6. For the others, the carrier chromosome is unknown.

**Table V.** The gene map of the rabbit (*Oryctolagus cuniculus* L),  $2n = 44$ : linkages, syntenies and assignments to chromosomes.

L1	Hb $\beta$ , Hq, Hx, Y, B, C
L2	I, du, En
L3	r <sub>1</sub> , r <sub>2</sub> ,
L4	a, Dw, Hg, w
L5	an, br, f, Hb
L6	e, Es.1, Es.2, Est1, Est2, Est4, Est6
L7	He, RLA
L8	Ab, Lpq-A, Lpq-B, Lpq-C, prt
L9	Ade, Aa
CH1	ACP-2, GALT, LDH-A
CH4	GAPD, TPI, LDH-B
CH6	CYP450-\3C1
CH9	ACY-1, GPX
CH13	PGM-1
CH15	GUK
CH16	KRAS2(p11 $\rightarrow$ q11)
CH17	ITP, NP, PEP-B, CYP450-PLM4, CYP450-PLM6
CH19	GSR, CKMN(q11 $\rightarrow$ q12)
CH20	CKB(q13 $\rightarrow$ qter)
CHX	G6PD, GLA, PGK, TRAI

play roles in chromosome banding as discussed by Comings (1978), Bickmore and Sumner (1989) and Rønne in this meeting.

### Mapping a chromosome to find a gene

We have discussed this point extensively during the second part of this lecture, and for more details see White and Lalouel (1988).

### Drawing real pedigrees of genes from pedigrees of zygotes

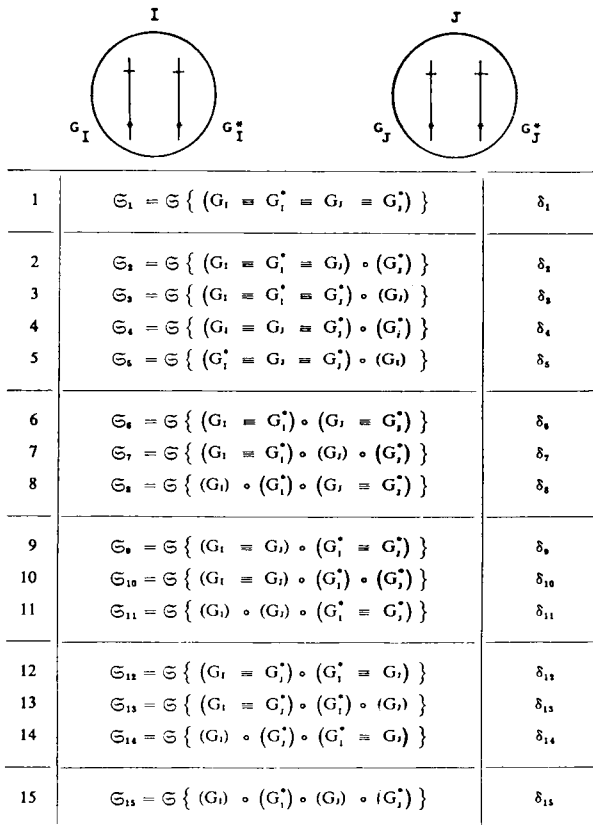
In 1963–1964, I introduced the notion of identity between genes in my book entitled, *La Relation d'Identité en Génétique (The Identity Relation in Genetics)*. This work is an axiomatic approach to genetics. Results are logical and a probabilistic description of genes in a small population when zygotes are related and consanguineous. These results have been applied to predictive genetics in the field of quantitative inheritance, mainly when relationships are known. RFLP and VNTR analyses open an exciting biochemical application of the theory of identity between genes. Now we can hope to be able to describe the actual state of identity between genes and the real pedigrees of genes (or genic pedigrees *versus* genotypic pedigrees), because, in the same experiment, we can test for identity and for non-identity between two, four or more homologous genes belonging to relatives.

The definition of the identity relation between genes is as follows: “two genes are linked by the identity relation if and only if they derive from the same ancestral gene, without mutation” (Gillois, 1963, 1964). Note, that Malecot (1948) in his book, *Les Mathématiques de l'Hérédité*, defines the identity relation between loci and not between genes. He uses this identity relation between loci to deduce joint probabilities of ‘allelic’ states of genes.

It is possible to extend this definition to a part or a segment of a chromosome: “Two chromosomal segments are linked by the identity relation if and only if they derive from the same ancestral chromosomal segment without mutation and recombination” (Gillois, 1963, 1964).

The important biochemical consequence of these definitions is that: “two genes, or two DNA strands, or two chromosomal segments which are identical carry the same pattern of identical restriction sites and give identical restriction fragments (RFLP).”

If each chromatogram of RFLP or VNTR is described by two criteria, the retention time and the surface of each peak (or its logarithm), more information is obtained. For example, VNTR DNA fragments which are in a haploid state fit with a linear regression, those in a diploid state fit with another linear regression and so on, as can be seen in figure 15. Consequently, we can decide what is the real identity situation realized among the 15 possibilities for four homologous genes, (see figures 13, 14 and 15). With RFLP for two loci or with VNTR, we can decide whether or not two loci are independent, if two loci are carried or not by the same chromosome, and the distance between two loci borne by the same chromosome.



**Fig 13.** Identity situations and identity coefficients between 4 homologous genes belonging to two related zygotes (from Gillois, 1963, 1964).

### Genetic dissection of quantitative or complex inheritable characters

The mathematical models of quantitative inheritance use a few of the following notions: contributions of ‘ghost’ genes to a character are independent, additive and have approximately the same magnitude. Often the additivity principle is transgressed, and interactions between ‘ghost’ genes, which may be described as dominance or epistasis, must be handled by the way of computational artifices that do not take real account of these phenomena.

These simple ideas, mainly developed by RA Fisher, enabled us not only to make universal a simplified conception of the genetics of heredity for continuous and discrete characters but, above all, defines *a priori* linear coefficients of correlation between relatives observed by statistical analysis. These coefficients form the essential tools of hereditary improvement techniques for crops and domestic animals.

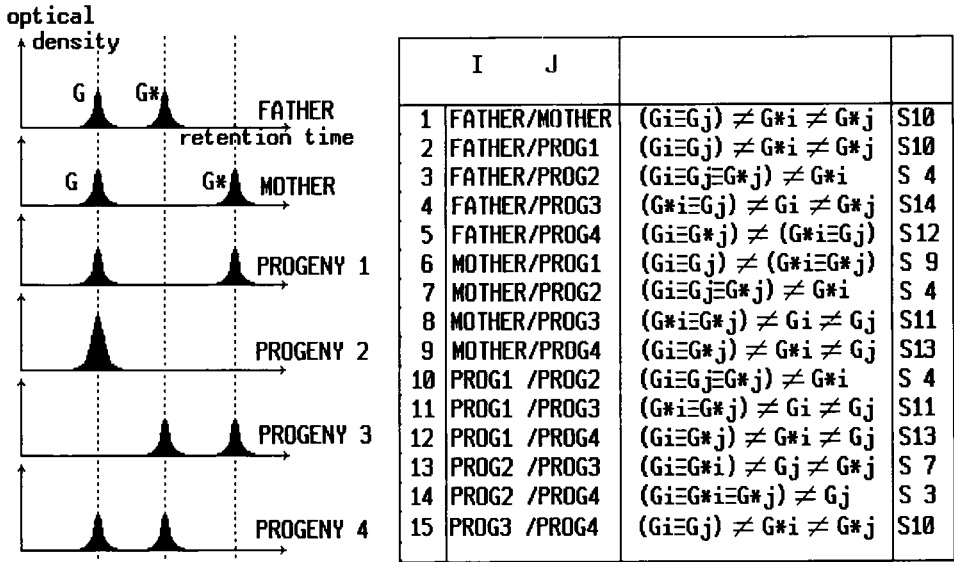


Fig 14. Identity coefficients can be deduced from RFLP chromatograms.

	FATHER		MOTHER		PROG. 1		PROG. 2		G IS THE GENE COMING FROM THE FATHER G* IS THE GENE COMING FROM THE MOTHER FIGURES IN THE TABLE ARE NUMBERS OF REPEATS
	G	G*	G	G*	G	G*	G	G*	
LOCUS 1	2	2	2	3	2	3	2	2	
LOCUS 2	6	4	4	4	4	4	6	4	
LOCUS 3	5	5	5	1	5	1	5	5	
LOCUS 4	8	10	10	10	10	10	8	10	

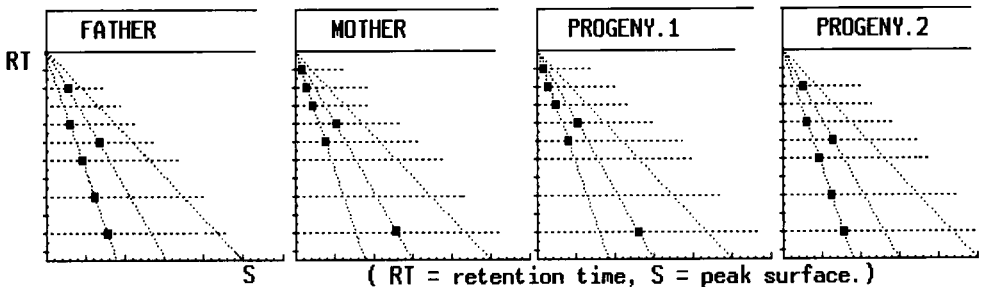


Fig 15. A simple illustration of a two-criteria analysis of a VNTR chromatogram: retention time versus peak surface.



These population improvement techniques have been increasingly polished, but have degenerated into statistical implements of production control. They are robust and frequently efficient tools, but became more and more irrelevant to genetic concepts in a strict and real sense.

Gene mapping provides an opportunity to obtain a real genetic interpretation of quantitative inheritance. Two alternative approaches are possible for genetic dissection of quantitative inheritance: the classical parametric approach and the non-parametric approach.

Paterson *et al* (1988, 1990) employed interval mapping, using the method of maximum likelihood. This procedure has been adapted for quantitative traits by Lander and Botstein (1989) from their own methods for analyzing complex traits in humans.

Interval mapping assesses the effects of each genomic segment, located between pairs of marker loci, rather than the effects associated with individual loci. This reduces the confounding effects of recombination between marker loci and quantitative trait loci (QTL), efficiently exploits the information from the RFLP linkage data and provides greater precision than was previously attainable. The results are depicted as QTL likelihood maps, a highly informative format to summarize large amounts of information. Likelihood maps represent something as the probability that one or more QTL lie at a particular point of the genome; this should not be confused with the magnitude of the contribution made by each region.

The non-parametric approach does not use the notion of an individual, quantitative, additive effect of genes. It considers two main ideas: 1) genes code sequences, particularly polypeptide chains. The heterogeneity of two allelic gene sequences ensures the variability of stereospecific mechanisms, as for example the variability of enzymatic activities; 2) genes do not work independently but according to one or many ordered steps in a dynamic network of actions. These dynamic networks can be described by systems of non-linear differential equations with delays as shown by Gillois *et al* (1978), Chevalet *et al* (1983), Corpet *et al* (1983) for the induction of the eleven genes comprising the *Lac* operon of *E. coli*.

In order to find the loci where a gene is located, I suggest the use of identity situations of loci or chromosomal fragments spread over the genome of two related zygotes. The main idea is that some identity situations are in relation with difference and sum of quantitative phenotypes of relatives. Loci or chromosomal fragments with high levels of identity in correlation with very small phenotypic differences are good candidates as closed markers or controlling loci.

The first step after this approach is to determine the nature of the whole or a large part of the graph of relations between genes which work together to build a phenotype. The second step is to establish the minimal dynamic non-linear differential equation system which describes the genes' concurrent actions building the complex inheritable phenotype. The necessary techniques are being developed, for example, the four Punnett decision tables which give the step orders of gene actions, or can be adapted from mathematical research in the field of economics, to give the differential equation system.

### Genetic improvement based on selection of marker genes

Genetic improvement based on selection of marker genes has been frequently advocated but remains untested for quantitative traits. Biometric approaches have been exceptionally successful in the absence of precise knowledge about the number, the location and the work relations of genes involved. Valid mapping experiments and marker-based selection are expensive and in many cases may not be justified by the technology currently available.

The most promising trend consists of altering the domestic animal genome by gene-targeting in pluripotential stem cell lines as a new route for genetic improvement. Interested readers are invited to read four pivotal reviews which have appeared in *Trends in Genetics*: Pluripotential stem cell lines as a route into the mouse germ line, by Robertson (1986); The new mouse genetics: altering the genome by gene targeting, by Cappechi (1989); Towards a molecular-genetic analysis of mammalian development, by Rossant and Joyner (1989); and Do multigene families regulate vertebrate development? by Dressler and Gruss (1988). Some preliminary results of experiments in which this strategy was applied to farm animals have been published by Rojas-Rousse and Gillois (1985) with rabbits and by McWhir *et al* (1988) with pigs.

The bottleneck in this field of research is in the ability to obtain perennialized totipotent cell lines and the ability to target the genome of these totipotent cells without the loss of their totipotency, which is clearly a job for somatic cell geneticists.

### Describing at the DNA level the molecular response to selection pressure

In a wonderful book entitled, *Microorganisms as Model Systems for Studying Evolution*, Mortlock (1984) writes: "The microorganisms present on the earth today possess a vast range of metabolic activities and are often able to demonstrate their surprising versatility by gaining both new enzyme activities and new metabolic pathways through mutations. It is generally assumed that the earliest microorganisms were very limited in their metabolic abilities, but as time passed they gradually expanded their range of enzymatic activities and increased both their biosynthetic and catabolic capacity. It is also believed that these primitive microorganisms increased the amount of genetic material they possessed by duplicating their existing genes and possibly by acquiring genetic material from other organisms. .../...

A small group of scientists has been exploring the means by which existing microorganisms are capable of mutating to expand their biochemical abilities."

Work reported in this book shows that microorganisms develop new metabolic activities, not by the acquisition of genetic material from external sources, but by altering in different ways their own existing genetic information under a light pressure of selection. The potential for genomic change depends upon the 'background phenotype' (in French, *bruit de fond phenotypique*), deregulation, amplification of pieces of DNA, borrowing of genes and differentiation of identical duplicated genes under selection due to the cost of the DNA replication or the necessity of homeostasis. These mechanisms must play in response to the selection

in plants or mammals (Gillois, 1983), and it seems of prime importance to track them as differences at the level of gene maps between unselected and improved plant or mammal lines.

### Considering the maternal effect as a genetic imprint

Maternal or paternal epigenetic effects are frequently advocated in selection or breeding plans. Embryonic transplantation has been performed for statistical studies. Papers by Reik (1989) and Monk (1990) show that it is possible to use the gene map and techniques developed for gene mapping to study the function of paternal and maternal chromosomes which can remember their parental origin. Observations suggest that maternal and paternal genomes have different epigenetic information imprinted on their chromosomes at different loci and that these differences can result in different developmental programs. If genomic imprinting can occur in a substantial proportion of the mammalian genome, it would be expected that penetrance or expressivity of genes in quantitative inheritance or in diseases would sometimes be related to their parental origin.

### REFERENCES

- Bickmore W, Sumner AT (1989) Mammalian chromosome banding: an expression of genome organization. *Trends Genet* 5, 144-148
- Cappechi MR (1989) The new mouse genetics: altering the genome by gene targeting. *Trends Genet* 5, 70-76
- Chevalet C, Corpet F (1986) Statistical decision rules concerning synteny or independence between markers. *Cytogenet Cell Genet* 43, 132-139
- Chevalet C, Corpet F, Gillois M, Micali A (1983) Modélisation dynamique de systèmes génétiques de régulation. I. L'induction de l'opéron lactose d'*Escherichia coli*: élaboration d'un modèle. *Génét Sél Evol* 15, 1-30
- Collick A, Jeffreys AJ (1990) Detection of a novel minisatellite-specific DNA-binding protein. *Nucleic Acids Res* 18, 625-629
- Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. *Annu Rev Genet* 12, 25-46
- Corpet F, Chevalet C, Gillois M, Micali A (1983) Modélisation dynamique de systèmes génétiques de régulation. II. L'induction de l'opéron lactose d'*Escherichia coli*: étude qualitative et numérique du modèle. *Génét Sél Evol* 15, 327-366
- Cowell JK, Mitchell CD (1989) A somatic cell hybrid mapping panel for regional assignment of human chromosome 13 DNA sequences. *Cytogenet Cell Genet* 52, 1-6
- Dressler GR, Gruss P (1988) Do multigene families regulate vertebrate development? *Trends Genet* 4, 214-219
- Dutrillaux B, Viegas-Pequignot E, Couturier J (1980) Très grande analogie de marquage chromosomique entre le lapin (*Oryctolagus cuniculus*) et les primates dont l'homme. *Ann Génét (Paris)* 23, 240-248
- Ege T (1980) Fusion of cell fragments as a method in cell genetics. In: *Transfer of Cell Constituents into Eukaryotic Cells*. (Celis JE, Graessman A, Loyter A, eds) Plenum Press, New York, 201-233
- Fournier REK, Ruddle FH (1977) Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster and human somatic cells. *Proc Natl Acad Sci USA* 74, 319-323
- Gardiner K (1990) Pulsed-field gel electrophoresis and investigations into mammalian genome organization. *J Cell Sci* 96, 5-8

- Gardiner K, Horisberger M, Krauss J, Tantravahi U, Korenberg J, Rao V, Reddy S, Patterson D (1990) Analysis of human chromosome 21: correlation of physical and cytogenetic maps; genes and CpG island distributions. *EMBO J* 9, 25-34
- Gellin J, Yerle M (1989) *In situ* hybridization, a technique for gene assignments. In: *Cytogenetics of Animals*. (Halnan CRE, ed) CAB International, Oxon, UK, 77-83
- Genetic maps (1990) *Locus maps of complex genomes*. 5th edn (O'Brien SJ, ed) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 6147 pp
- Gillois M (1963) *La Relation d'Identité en Génétique*. BCBQ Station Centrale de Génétique Animale, Paris, pp 294
- Gillois M (1964) La relation d'identité en génétique et la sélection. Thesis, Sorbonne, Paris
- Gillois M (1983) L'information génétique: du code génétique au message héréditaire *Génét Sél Evol* 15, 299-322
- Gillois M, Chevalet C, Micali A, Baron R, Costa R (1978) Biochemical and molecular approaches of quantitative genetics. *Ann Acad Brasil Cienc* 50, 227-237
- Goldman MA (1988) The chromatin domain as a unit of gene regulation. *BioEssays* 9, 50-55
- Hastie ND, Allshire RB (1989) Human telomeres: fusion and interstitial sites. *Trends Genet* 5, 326-331
- Jabs EW, Goble CA, Cutting GR (1990) Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. *Proc Natl Acad Sci USA* 86, 202-206
- Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 60, 473-485
- Kavenoff R, Klotz LC, Zimm BH (1973) On the nature of chromosome-sized DNA molecules. *Cold Spring Harbor Symp Quant Biol* 38, 1-8
- Lathrop GM, Lalonde JM (1988) Efficient computations in multilocus linkage analysis. *Am J Hum Genet* 42, 498-505
- Lander ES, Green P (1987) Construction of multilocus gene linkage maps in human. *Proc Natl Acad Sci USA* 84, 2363-2367
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-189
- Lengyel PJ, Speyer JF, Basilier C, Ochoa S (1961) Synthetic polynucleotides and the amino acid code. *Proc Natl Acad Sci USA* 47, 1936-1942
- Ludecke HJ, Senger G, Claussen U, Horsthemke B (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* 338, 348-350
- Malécot G (1948) *Les Mathématiques de l'Hérédité*. Masson, Paris, 64 pp
- McWhir J, Church RB, Evans MJ (1988) Polyoma early region transforms a porcine multipotential embryonic cell line, 35.21.11. Proceedings of the XVI International Congress of Genetics. Toronto, Canada
- Monk M (1990) Variation in epigenetic inheritance. *Trends Genet* 6, 110-114
- Mortlock RP (1984) *Microorganisms as Model Systems for Studying Evolution* Plenum Press, New York, 1-326
- Moyzis RK, Albright KL, Bartholdi MF, Cram LS, Deaven LL, Hildebrand CE, Joste NE, Longmire JL, Meyne J, Schwarzscher-Robinson T (1987) Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis. *Chromosoma* 95, 375-386
- Nadeau J (1989) Maps of linkage and synteny homologies between mouse and man. *Trends Genet* 5, 82-86
- O'Brien SJ, Seuanez HN, Womack JE (1988) Mammalian genome organization: an evolutionary view. *Annu Rev Genet* 22, 323-351

- Paterson AH, Lander ES, Hewitt JD, Paterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721-726
- Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* 124, 735-742
- Paulson JR, Laemmli UK (1977) The structure of histone-depleted metaphase chromosomes. *Cell* 12, 817-828
- Pirrotta V, Jackle H, Edstrom JE (1983) Microcloning of microdissected chromosome fragments. *Genet Eng Principles Methods* 5, 1-17
- Prosser J, Frommer M, Paul C, Vincent PC (1986) Sequence relationships of three human satellite DNAs. *J Mol Biol* 187, 145-155
- Reik W (1989) Genomic imprinting and genetic disorders in man. *Trends Genet* 5, 331-336
- Robertson EJ (1986) Pluripotential stem cell lines as a route into the mouse germ line. *Trends Genet* 2, 9-13
- Rojas-Rousse D, Gillois M (1985) Immortalisation de cellules de blastula de lapin par le gène des fonctions précoces du SV-40. In: *Aspect Industriel des Cultures Cellulaires d'Origine Animale et Végétale*. (Terre J, Petiard V, eds) Lyons, France 462-480
- Rossant J, Joyner AL (1989) Towards a molecular-genetic analysis of mammalian development. *Trends Genet* 5, 277-283
- Salser W, Bowen S, Browne D, El Addi F, Fedoroff N, Fry K, Heindell H, Paddock G, Poon R, Wallace B, Whitecome P (1976) Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Fed Proc* 35, 23-35
- Sargent CA, Dunham I, Cambell RD (1989) Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region. *EMBO J* 8, 2305-2312
- Scalenghe F, Turco E, Edstrom JE, Pirrotta V, Melli S (1981) Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma* 88, 205-216
- Senger G, Ludecke HJ, Horsthemke B, Claussen U (1990) Microdissection of banded human chromosomes. *Hum Genet* 84, 507-511
- Soulie J, de Grouchy J (1982) Of rabbit and man comparative gene mapping. *Hum Genet* 60, 172-175
- Spies T, Blanck G, Bresnahan M, Sands J, Strominger JL (1989) A new cluster of genes within the human major histocompatibility complex. *Science* 243, 214-217
- Sundquist WI, Klug A (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpairing loops. *Nature* 342, 825-829
- Tourian A, Johnson RT, Burg K, Nicolson SW, Sperling K (1978) Transfer of human chromosomes *via* human minisegregant cells into mouse cells and the quantification of the expression of hypoxanthine phosphoribosyltransferase in the hybrids. *J Cell Sci* 30, 193-209
- Trifonov EN, Sussman JLL (1980) The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc Natl Acad Sci USA* 77, 3816-3820
- Vogt P (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved 'chromatin-folding code'. *Hum Genet* 84, 301-336
- Wevrick R, Willard HF (1989) Long-range organization of tandem arrays of  $\alpha$ -satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc Natl Acad Sci USA* 86, 9394-9398
- White R, Lalouel JM (1988) Sets of linked genetic markers for human chromosomes. *Annu Rev Genet* 22, 259-279

- Xu J, Hardison RC (1989) Localization of the  $\beta$ -like globin gene cluster and the genes for parathyroid hormone and c-Harvey-ras 1 to region q14-q21 of rabbit chromosome 1 by *in situ* hybridization. *Cytogenet Cell Genet* 58, 157-161
- Yerle M, Gellin J (1989) Localization of leukocyte interferon gene in the q2.5 region of pig chromosome 1 by *in situ* hybridization. *Genet Sel Evol* 21, 249-252
- Yerle M, Gellin J, Echard G, Lefevre F, Gillois M (1986) Chromosomal localization of leukocyte interferon gene in the pig (*Sus scrofa domestica* L) by *in situ* hybridization. *Cytogenet Cell Genet* 42, 129-132