Genetic studies in the higher animals: new strategies

J. Gowrishankar and V. Radha
Centre for Cellular and Molecular Biology, Hyderabad 500 007, India

Recent advances—both conceptual and technical—have enabled the development of novel strategies for correlating phenotype with genotype in the higher animals. Included in these developments are the use of restriction-fragment-length polymorphisms (RFLPs) as markers for construction of genetic maps; methods for identifying genes responsible for human inherited disorders in which precise biochemical defects have not yet been ascertained; methods for mapping genes that affect the expression of quantitative (multigenic) traits; and 'reverse-genetic' approaches to studying the function of cloned genes.

AMONG the most powerful of approaches available for the general study of biological function is that offered by genetics, in which a comparison between the phenotype of the wild type and that of isogenic mutants defective in a particular gene is used to draw inferences as to the mechanism of action of the concerned gene product and its contribution to physiological function in the organism. Easily applied as it has been in the prokaryotes and other microorganisms, this approach has paid rich dividends in terms of providing us with basic understanding of the mechanisms of inheritance, of gene expression and its regulation, and of the various metabolic pathways and functions; indeed, the tools of modern biotechnology themselves represent the fruits of such lines of study.

Classical genetic analyses in the higher animals, on the other hand, have been hampered by the difficulties inherent in the generation and characterization of isogenic mutant lines. In this context, studies on the various inherited disorders of man have also been of special interest to the geneticist[1], not only because of their obvious clinical relevance but also because the affected individuals themselves serve as potentially valuable source material (as mutants in the formal sense) for the application of the genetic approach outlined above for understanding biological function. Ethical considerations, however, have precluded the use in humans of experimental strategies for gene transfer, mapping and so on, which have been profitably exploited in genetic studies in other organisms. Until recently, therefore, the dissection of function and mechanisms in the higher animals had been far less amenable to genetics than that in the microorganisms, and none but a very few of the inherited disorders had been understood at the molecular level.

In this review, we discuss the application of recombinant-DNA and related techniques in evolving new approaches for genetic studies in complex organisms, in particular the higher animals. The discussion is divided into two broad themes: (i) methods for identification and analysis of genes defined by pre-existing mutations (including the various inherited disorders of man), and (ii) mutational approaches to studying gene function after a gene has been cloned and isolated as a sequence of DNA—also called 'reverse genetics'. In both instances, our emphasis has been more on providing an outline of concepts and strategies with citations to illustrative examples, rather than on an exhaustive compilation of the list of genes and disorders that have been successfully studied in this manner.

Methods for molecular-genetic analysis of inherited disorders

Although the discussion below is primarily focused on human inherited disease, the principles therein are equally applicable to the study of mutant strains and lines in all categories of the higher animals. Depending upon how much prior knowledge exists (or existed until recently) with regard to the nature of deficiency in particular inherited diseases, one can group them into one or the other of the following three classes: (i) those in which the concerned gene product had not only been identified but its function also was known, as for example in sickle cell disease and the thalassaemias, Lesch–Nyhan syndrome, classical haemophilia, severe combined immunodeficiency, familial hypercholesterolaemia, and so on[1]; (ii) those in which the gene product had been identified as altered or absent in the affected individuals (for example, by two-dimensional gel electrophoresis or by antibody-based procedures) but whose function was not known; and, finally, (iii) those in which the disorder was (or even now is) understood only in terms of a gross phenotype, with no information on either the nature or function of the concerned gene product. This category includes diseases such as cystic fibrosis, Duchenne muscular dystrophy, Huntington's disease, achondroplasia and so on.

The methods for cloning and isolation of genes for

diseases belonging to the first two categories are relatively straightforward; they have also been extensively used for the cloning of genes not defined by any particular mutations but whose products have been identified by other means, e.g. various hormones and receptors, heat-shock proteins, and so on. The strategies available for identification of the cloned gene in these instances are only briefly outlined in this review, and include:
(i) assay for function of the cloned gene product, either by complementation in mutant cell lines[2] or after expression of the gene *in vitro* or in frog oocyte systems[3];
(ii) identification of the cloned gene product by gel-electrophoretic techniques, or by antibody-binding assays; the λgt11 vector-cloning system has been widely employed in the latter of these two strategies[4];
(iii) use of a method known as hybrid-selected translation[5], in which the cloned gene is identified by its ability to hybridize specifically to that species of messenger RNA which, when subsequently translated into protein *in vitro*, yields the product of interest;
(iv) identification of the cloned gene based on knowledge of its transcriptional regulation or tissue specificity of expression, so that the clones of interest are detected by their ability to exhibit differential hybridization to radio-labelled total messenger RNA isolated from tissue before and after induction, or from two different tissues[6-8];
(v) preparation of nucleic-acid probes based on available data on the amino-acid sequence of the gene product[9], or from cloned homologous genes from other organisms[10], which are then used to identify the corresponding gene clones by hybridization screening.

The third category of genetic disorders described above, in which neither structure nor function of the concerned gene product is known, is in some respects the most relevant to the discussion in this review, because it is here that the utility of the genetic approach in understanding physiology is clearly exemplified. This category, by definition, includes all those genes whose very existence has been surmised wholly and exclusively from the fact that mutants (that is, affected individuals) in these genes occur in the population. How does one proceed to identify and isolate a gene for which no other information is available?

RFLP analysis

The strategy to track down such genes uses a powerful and fundamental technique known as restriction-fragment-length-polymorphism (RFLP) analysis. The RFLP technique has been discussed in detail elsewhere[11,12], and is only briefly described here. RFLPs refer to small heritable variations in DNA sequences in different individuals of a population, which lead to the creation or abolition of recognition sites for one or the other of the various restriction endonucleases known. If such a polymorphism were to occur in a particular segment of the chromosome, it can be visualized by digesting total DNA from each of the individuals with an appropriate restriction enzyme, separating the mixture of fragments so generated according to size by gel electrophoresis, immobilizing the fragments (by blotting) on membrane filters, and then identifying the size of the fragment(s) from the region of interest by hybridization to a labelled probe of DNA cloned from that region; an RFLP would be manifest in such an experiment in the form of bands at different positions on the gel for different individuals. RFLPs have many of the attributes expected of ideal markers for the construction of a genetic map[13]: they are non-selective (impose no bias for or against their own inheritance in crosses); they exhibit codominance (heterozygotes can be distinguished from each of the two possible homozygous states); they are distributed relatively uniformly through the genome, and a single general strategy is sufficient to identify any one of them; they are amenable to progressively increasing saturation-mapping; and, finally, they can readily be integrated into a genetic map composed of other, more traditional markers.

The method used to map the genes involved in the third category of inherited disorders has been to prepare what may be called pedigree blots of DNA obtained from both affected and unaffected individuals in whole families and digested with each of a variety of restriction enzymes, and then to probe these blots with a series of mapped probes to try and identify RFLPs in any of the families that co-segregate with the disease (or the carrier) phenotype. Such co-segregation, once identified, is taken to imply the existence of genetic linkage between the locus corresponding to the concerned probe DNA and the gene affected in the particular inherited disorder. The subsequent steps are designed to enable one to cross the gap between these two loci, and include one or more of the following:
(i) identification of additional RFLP markers in the region that are increasingly closely linked to the disease locus, and also those that define the boundary from the opposite end;
(ii) chromosome walking, whereby successive contiguous stretches of DNA are cloned proceeding in both directions from the original RFLP-marker locus—in this context, the availability of chromosomal-translocation, inversion, insertion or deletion mutants in the vicinity enables one to establish the direction of the 'walk'[14];
(iii) chromosome jumping, which (like chromosome walking) also enables one to clone regions of DNA progressively further away from the original RFLP locus, but in a discontinuous fashion (therefore enabling

one to traverse larger DNA regions with fewer steps)[15]; (iv) searching in the cloned regions of DNA for features likely to be associated with the disease gene, such as by looking for HTF islands, sequence conservation during phylogeny, and so on[16].

The final proof for identification of the disease locus would come from demonstration of mutant alleles of the locus in the affected individuals. The strategy and steps described above have been successfully employed in the mapping and cloning of genes affected in cystic fibrosis[17,18] and multiple neurofibromatosis (von Recklinghausen's disease)[19-21], and are being used in ongoing attempts to clone the Huntington's-chorea gene[22].

A related approach has been followed in the cloning of genes in which the mutation contributing to a disease phenotype has itself been created by a gross cytogenetic alteration, such as a chromosome translocation, deletion or inversion; the strategy is simplified in that the near-random search for a co-segregating RFLP marker is replaced by a more directed one (using probes mapped to the vicinity of the rearrangement in a chromosome walk in normal and in affected individuals) to detect the junction point of the alteration. This approach has been used to identify the loci responsible for hereditary retinoblastoma[23,24] and Wilms' tumour-aniridia syndrome[25], and the translocations associated with several haematological malignancies such as Burkitt's lymphoma and chronic myelogenous leukaemia[26]. In a slightly modified form, it has also been successfully employed in the identification of two genes (ZFY and SRY) involved in expression of the maleness phenotype in humans[27,28], the latter of which is a likely candidate for being the much-sought-after testis-determining gene. In one instance where an inherited trait (Duchenne muscular dystrophy in association with chronic granulomatous disease and other abnormalities) in a patient was associated with a deletion, DNA from the patient was itself used as a reagent in an enrichment protocol (involving subtractive hybridization) to clone the wild-type sequences corresponding to the region of deletion[29,30].

Analysis of multigenic traits

The discussion above has focused on the identification and molecular analysis of inherited single-gene traits and disorders. It is clear, however, that many inherited traits (and certainly a large majority of the economically useful ones in both crop plants and domesticated animals) are governed not by the sole action of a single gene product but instead by a number of genes, with additional factors also from the environment. Such traits (which in humans are thought to include ones such as essential hypertension, maturity-onset diabetes, ischaemic heart disease, and schizophrenic illnesses) are said to show multifactorial or multigenic inheritance, and the sets of genes influencing their expression are referred to as quantitative-trait loci (QTLs). With the classical genetic techniques, it had been difficult even to determine the extent of genetic contribution to these multifactorial traits, let alone to identify and map the corresponding QTLs. Plant geneticists, however, have recently developed experimental and statistical computational techniques to analyse and map QTLs[31,32], and these techniques are also applicable in similar studies in the higher animals (Figure 1). The method involves crossing two independently inbred parental strains that are phenotypically widely different from one another in respect of several quantitative traits, and then analysing the genomes of a number (50 to 100) of the recombinant hybrids using a complete series of mapped and ordered RFLP probes to determine in each of them (the hybrids) which segments have been inherited from one parent and which from the other. The phenotype of each hybrid with respect to the quantitative trait is then compared with the RFLP-based genome-analysis data to determine statistically which segments of the genome contribute significantly to the ultimate phenotype and therefore are likely to harbour QTLs. Once identified,

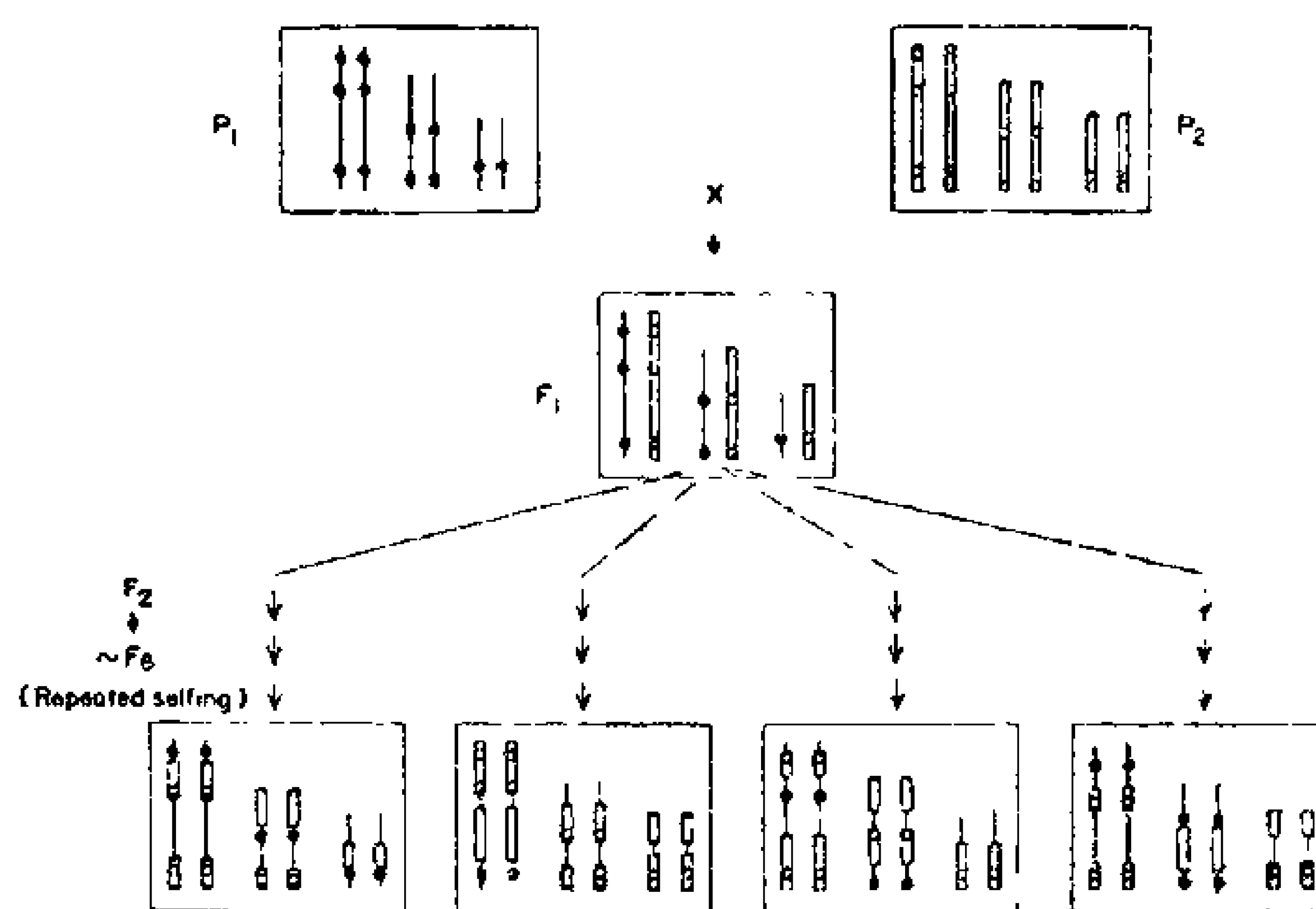

Figure 1. Schematic illustration of QTL mapping by RFLP analysis. P1 and P2 represent genetically distinct inbred diploid parental lines, whose chromosomes are indicated by the solid lines and open bars respectively. Each of several F1 individuals (or pairs) obtained in a cross between P1 and P2 is taken through a series of self-crosses over four or five generations (F2 to approximately F6) to generate recombinant inbred lines (RILs). Individuals in a given RIL would all be genetically identical and homozygous, but would differ from those of other RILs set up from the same pair of parents in the patchwork pattern of chromosomal contributions respectively from P1 and P2 (as shown in the panels at bottom of the figure, representing individuals from four RILs), the exact pattern for each RIL can be constructed with the aid of RFLP analysis, using a series of ordered probes covering the entire genome. Also shown in the figure are six hypothetical QTLs distributed on three chromosome-pairs (represented by the solid and open circles respectively in P1 and P2), and the differential pattern of their segregation in each of the RILs, the existence or otherwise of a QTL in any given chromosomal segment can be ascertained by statistical correlation in each RIL of phenotype for the given trait with the likelihood of inheritance of that segment from one or the other of the two parents

each QTL can be further analysed just as if it were a locus governing any monogenic trait.

The above strategy has successfully been used to identify and map the QTLs governing traits such as fruit mass and fruit pH in the tomato[31], and its principle is easily adopted for application in those animals in which inbred lines and defined crosses are possible[33]. On the other hand, this approach is not immediately applicable in human-genetic studies, because it is contingent upon the development of statistical techniques that can identify QTLs by analysis of a population taking individual families as units[34]. An alternative method for the detection of QTLs in man might be first to map QTLs for a given trait in the mouse, and then to look for the existence of corresponding loci in syntenic regions of the human genome[35].

Genetic approaches to study of function and regulation of cloned genes

After a putative gene has been cloned, one powerful approach to understanding its function is to use the cloned DNA to construct mutants altered in the structure or expression of the gene, and then to ask the question: How does such an alteration affect phenotype? This has been termed the reverse-genetic approach[36], because of the way in which it reverses the otherwise usual mode of experimental questioning in genetics: What is the mutation that underlies an observed (pre-existing) phenotype? Although the term has been applied also to studies of 'phenotype' following the introduction of the altered gene into the genome of cultured cells or in transient-expression assays following transfection, or even in *in vitro* reactions, this review focuses its attention primarily on those techniques of reverse genetics that permit phenotype studies in whole animals. The methods used to mutagenize a cloned DNA segment are also not discussed here.

The vast majority of reverse-genetic studies in higher animals have been done in the mouse[37]. The usual procedure for the construction of transgenic mice has been to introduce, by microinjection, DNA encoding the gene (i) into embryonic stem (ES)-cell lines and then to produce chimaeric embryos by transferring these cells into normal blastocysts, or (ii) into the male pronucleus of fertilized eggs just before the nuclear-fusion event[38-40]. The chimaeric embryos are allowed to develop to term in the uterus of foster mothers, and, from among the progeny, those animals are identified (by screening in the F2 generation) that have incorporated the microinjected gene into the germline. The integration of microinjected DNA almost always occurs by non-homologous recombination (so that the endogenous copy of the gene is not disrupted in the process), and the number of copies so integrated may vary from one to a few hundred; both these features limit the utility of this technique in undertaking genetic studies. It has, however, been used to delineate the circumstances in which aberrant expression of dominant oncogenes leads to the development of malignancies, or that of particular antigenic determinants or of antibody molecules in immune regulatory networks in the animals[40]. It has also been used to define the extent of *cis* sequences required to obtain tissue-specific expression of particular microinjected genes, and to ablate or destroy particular target tissues in the transgenic animal by effecting the tissue-specific expression of gene products that are lethal to cell metabolism[42]. One interesting and unexpected offshoot of these studies was the discovery that non-homologous integration of microinjected DNA itself can lead to the insertional inactivation of developmentally important genes, and several mutant strains have been obtained and characterized by this process[43,44].

One disadvantage of microinjection as a method for DNA delivery is the low frequency with which the transferred DNA is integrated into the genome; this may be overcome by the use, instead, of disabled retroviruses as vectors for transduction of DNA into target cells[38-40]. Much work at present is directed towards development and improvement of such vector systems, but all of them are designed more for gene transfer into cells *in vitro* than for the construction of transgenic organisms; in one instance, murine haematopoietic stem cells have been stably transduced with the human β-globin gene and have then been used successfully to reconstitute the marrow of lethally irradiated mice[45].

Generating null mutants through reverse genetics

The widespread application of DNA cloning and sequencing techniques has led to the identification of hitherto uncharacterized genes and open reading frames in respect of which no clues to their possible function are available, or even whether they represent functional genes at all. The number of such genes is bound to increase by several orders of magnitude as large-scale projects for systematic end-to-end sequencing of the human and other animal genomes are taken up. Reverse-genetic strategies could, in principle, be applied to obtain null mutants in each of these genes, the phenotypes of which would then enable one to surmise the function of the corresponding wild-type allele. Furthermore, these same techniques would also be expected to be useful even in the characterization of genes identified and cloned by the more conventional approaches.

A number of alternative conceptual strategies have been described to obtain mutants with null phenotype in a particular gene by reverse-genetic procedures. One

is to replace, by homologous recombination, the endogenous wild-type gene in an ES-cell line with a copy that has been mutated *in vitro*[37,46]. As mentioned above, the frequency of non-homologous integration of microinjected DNA vastly exceeds that of homologous recombination, but powerful selection procedures have been developed to enrich for and identify the rare events of the latter type. One example of such a selection relies on the principle that, in non-homologous integration, the entire segment of microinjected DNA is incorporated into the genome, whereas in homologous recombination, regions of DNA outside the region of homology are not so incorporated. Thus, by using a drug-resistance marker (for example G418^R) insertion within the region of homology and a negatively selectable marker (for example the gene *tk*, for thymidine kinase, from herpes simplex virus, whose expression would confer sensitivity to agents such as acyclovir) outside, it is possible to identify the desired class of recombinants that have acquired the drug-resistance marker gene but not *tk*, as opposed to those that have incorporated DNA by non-homologous integration and acquired both (Figure 2)[47]. Derivative strains of the ES-cell line that have acquired the mutation are then used to construct transgenic mouse strains. The progeny obtained from such an experiment would be hemizygous null mutants[48,49], pairs of which can subsequently be crossed to generate homozygous null mutants in the F2 generation. This method has successfully been employed to obtain transgenic mice homozygous for engineered insertions in the β_2-microglobulin gene[50], and in the *int-1* protooncogene[51]. It is also useful in the construction of animal models for particular inherited disorders of man, e.g. sickle cell disease model in the mouse[52], and in attempts at targeted gene therapy for such disorders[53,54].

The other strategies for obtaining null mutants are designed not to disrupt the structure of the endogenous wild-type gene but instead to interfere with either its expression or its function; the strains obtained in this way are therefore not true null mutants but phenocopy null mutants. In one method, the isolated gene is placed in inverted orientation downstream of a promoter so that 'antisense' RNA transcripts are made in the tissues of transgenic mice; these transcripts would interfere with the synthesis and/or expression of the 'sense' RNA transcripts from the endogenous gene, and therefore mimic a null-mutant phenotype[55]. In another method, the cloned gene is mutagenized in an attempt to obtain and identify dominant-negative mutations—that is, mutations that would exhibit a null phenotype even when the endogenous wild-type gene product is present in the same tissue. The rationale of this strategy is that the majority of cellular proteins function as oligomers, and that it is possible to obtain mutations, which, even when present in one subunit of the oligomer (all others being products of the wild-type gene and therefore normal), would render the protein inactive—thus, in effect, the mutated subunit molecules would titrate out the normal ones[56]. In a third method, information on the sequence of the cloned gene may be used to design a ribozyme gene (encoding a catalytic RNA molecule) whose expression in transgenic animals would serve to specifically destroy, by cleavage, RNA transcripts of the endogenous gene[57]. The ribozyme strategy has so far not been applied in studies on whole animals; on the other hand, the antisense-RNA approach has been used to examine the function of myelin basic protein in relation to the shiverer mutant phenotype in mice[58], and that involving dominant-negative mutations in studies to establish the role of the pro-α1(I) collagen gene in the disease osteogenesis imperfecta type II (ref. 59).

Recently, yet another novel method has been

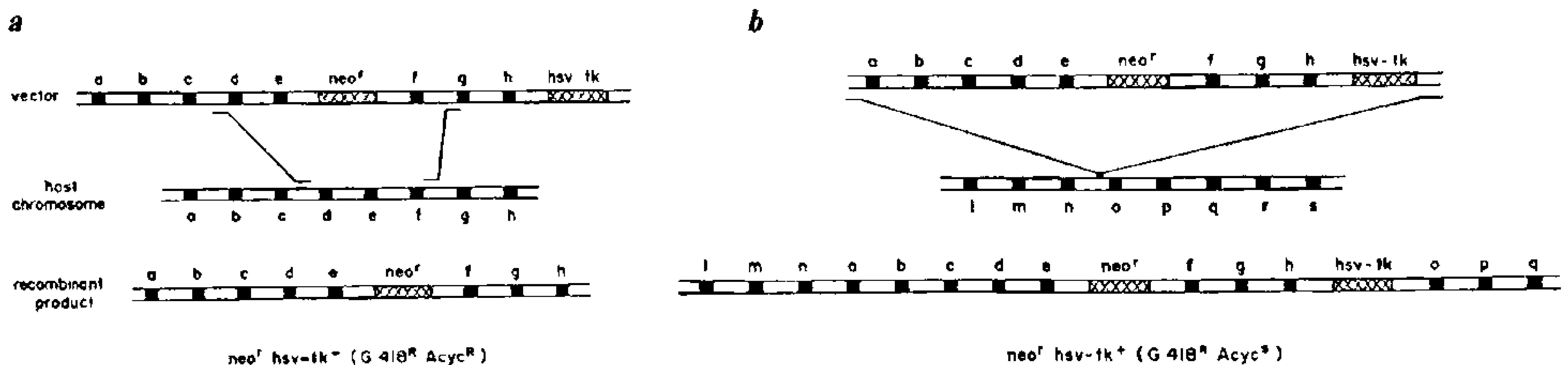

Figure 2. Positive negative selection strategy to identify homologous recombination events after DNA microinjection (gene targeting)[47]. The vector carries the *neo^r* insertion (encoding resistance to the antibiotic G418) in the gene situated between the loci *e* and *f*, and this insertion is sought to be introduced by homologous recombination into the corresponding position on the host chromosome. The vector also carries, in a region outside the segment of homology, the thymidine kinase gene of herpes simplex virus (hsv-tk), whose inheritance would confer on cells sensitivity to the nucleoside analogue acylovir (Acyc^S). Integration of microinjected vector DNA can occur either by homologous recombination, resulting in inheritance of the *neo^r* but not the hsv-tk marker (as in ***a***), or by non-homologous recombination resulting in inheritance of both (as in ***b***). Only the former class of recombinants would be expected to grow in media supplemented with both G418 and acyclovir, because the latter class would be G418^R Acyc^S, and the non-recombinants would be G418^S Acyc^R.

described that is potentially useful in determining the function of genes of higher animals incidentally identified during cloning and sequencing efforts. In this technique, the cloned gene is placed in an unstable plasmid downstream of a yeast promoter and is then introduced into a randomly mutagenized population of yeast cells; rare transformant clones are then sought to be identified (with the aid of a visually assayable reporter gene on the plasmid) in which the plasmid is stably maintained even in the absence of selection. The existence and identification of such colonies would indicate (i) that, in these strains, the cloned gene is able to complement mutations in a gene essential for yeast-cell viability, and (ii) that the power of yeast genetics can now be brought to bear in understanding its function through these mutations[60]. One expects that this would be a valuable approach to studying that large subset of genes in higher animals which have equivalent functional homologues in yeast.

Concluding remarks

The genomes of higher animals comprise roughly 100,000 genes, only five per cent of which have been identified and less than one per cent well studied. The techniques and strategies described above offer means hitherto unavailable to undertake detailed and systematic genetic analyses in these organisms, and to meaningfully integrate the wealth of information expected to be generated from genome-sequencing projects with studies of function, physiology and metabolism. One can with confidence also expect further improvements in the ease with which these techniques can be used and applied, especially in relation to reverse-genetic studies.

What are the opportunities, if any, available in this field in the special context of science and biotechnology in India? Three points are worthy of mention. One is the potential in animal husbandry for application of these techniques (particularly those involving QTLs)[33] in breeding of livestock of relevance to Indian needs. The second is the largely uncharted area of human inherited diseases and syndromes in various Indian races and populations. We identified the following publications reporting hitherto unknown inherited disorders in a survey of the Medline database from 1983 to 1990: two different deafness syndromes associated with other anomalies[61,62], osteogenesis imperfecta associated with blindness[63], a pedigree with bilateral preaxial polydactyly[64], familial angiolipomatosis associated with polycystic kidney disease[65], inherited degenerative chondropathy affecting the nose and larynx[66], early-onset hereditary spinocerebellar ataxia[67], and congenital hepatic fibrosis associated with facial abnormality[68], all of them in families of Indian origin. Each of these traits may be expected to define a new gene, and hence a new, as yet unidentified function, that may lead to a fruitful collaboration between clinic and basic science laboratory.

Finally, Lander and Botstein[69] have recently proposed an elegant and efficient strategy for the mapping of genes for recessive traits in humans, which would make use of a set of ordered RFLP probes to delineate such regions of the genome that are homozygous in a proband born of a consanguineous mating. They calculate that an unknown recessive 'disease' gene can be mapped by studying DNA from fewer than a dozen unrelated, affected inbred children, given a complete human RFLP linkage map. Their strategy is uniquely relevant to human-genetic studies in this country, because of the high incidence of consanguineous (first-cousin and uncle–niece) marriages in several Indian communities[70].

1. McKusick, V. A, *Mendelian Inheritance in Man*, 8th edn, Johns Hopkins University Press, Baltimore, USA, 1988.
2. Lee, M. G. and Nurse, P., *Nature*, 1987, **327**, 31.
3. Hollmann, M, O'Shea-Greenfield, A, Rogers, S. W. and Heinemann, S., *Nature*, 1989, **342**, 643.
4. Young, R. A. and Davis, R. W., *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 1194.
5. Old, R. W. and Primrose, S. B., *Principles of Gene Manipulation*, Blackwell Scientific Publications, London, 1981.
6. Schmidt, C. J. and Hamer, D. H., *Gene*, 1983, **24**, 137.
7. Lee, K. L., Isham, K. R, Stringfellow, L., Rothrock, R. and Kenney, F. T., *J. Biol. Chem.*, 1985, **260**, 16433.
8. Davis, M. M. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 2194.
9. Ebina, Y. *et al.*, *Cell*, 1985, **40**, 747.
10. Higashi, Y., Yoshioka, H, Yamane, M., Gotoh, O. and Yoshiaki, F.-K., *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 2841.
11. Botstein, D., White, R. L., Skolnick, M. and Davis, R. W., *Am. J. Hum. Genet.*, 1980, **32**, 314.
12. Donis-Keller, H. *et al.*, *Cell*, 1987, **51**, 319.
13. Tanksley, S. D., Young, N. D., Paterson, A. H. and Bonierbale, M. W., *Bio/Technology*, 1989, **7**, 257.
14. Steinmetz, M. *et al.*, *Nature*, 1982, **300**, 35.
15. Collins, F. S. and Weissman, S. M., *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 6812.
16. Orkin, S. H., *Cell*, 1986, **47**, 845.
17. Rommens, J. M. *et al.*, *Science*, 1989, **245**, 1059.
18. Riordan, J. R. *et al.*, *Science*, 1989, **245**, 1066.
19. Wallace, M. R. *et al.*, *Science*, 1990, **249**, 181.
20. Viskochil, D. *et al.*, *Cell*, 1990, **62**, 187.
21. Cawthon, R. M. *et al.*, *Cell*, 1990, **62**, 193.
22. Gusella, J. F. *et al.*, *Nature*, 1983, **306**, 234.
23. Dryja, T. P., Rapaport, J. M., Joyce, J. M. and Petersen, R. A., *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 7391.
24. Friend, S. H. *et al.*, *Nature*, 1986, **323**, 643.
25. Call, K. M. *et al.*, *Cell*, 1990, **60**, 509.
26. Rowley, J. D., *Cancer Res*, 1984, **44**, 3159.
27. Page, D. *et al.*, *Cell*, 1987, **51**, 1091.
28. Sinclair, A. H *et al.*, *Nature*, 1990, **346**, 240.
29. Kunkel, L. M, Monaco, A. P., Middlesworth, W., Ochs, H D. and Latt, S. A., *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 4778.
30. Royer-Pokora, B. *et al*, *Nature*, 1986, **322**, 32.
31. Paterson, A H. *et al.*, *Nature*, 1988, **335**, 721
32. Lander, F. S. and Botstein, D, *Genetics*, 1989, **121**, 185
33. Beckmann, J. S. and Soller, M, *Bio Technology*, 1987, **5**, 573.
34. Marx, J, *Science*, 1990, **247**, 1540.
35. Nadeau, J. H, *Trends Genet*, 1989, **5**, 82

36 Shortle, D., DiMaio, D and Nathans, D., *Annu Rev Genet*, 1981, **15**, 265
37 Van Brunt, J., *Bio Technology*, 1990, **8**, 725
38 Palmiter, R D and Brinster, R L., *Annu. Rev. Genet.*, 1986, **20**, 465
39 Jaenisch, R., *Science*, 1988, **240**, 1468
40 Hanahan, D., *Science*, 1989, **246**, 1265.
41 McNab, A R. *et al*, *Proc Natl Acad Sci. USA*, 1990, **87**, 6848.
42. Borrelli, E., Heyman, R., Hsi, M and Evans, R M., *Proc. Natl. Acad. Sci USA*, 1988, **85**, 7572.
43 Schnieke, A., Harbers, K. and Jaenisch, R., *Nature*, 1983, **304**, 315.
44 Woychik, R P., Stewart, T A., Davis, L. G., D'Eustachio, P and Leder, P., *Nature*, 1985, **318**, 36
45 Dzierzak, E. A., Papayannopoulou, T. and Mulligan, R. C., *Nature*, 1988, **331**, 35.
46 Capecchi, M R., *Science*, 1989, **244**, 1288.
47. Mansour, S L., Thomas, K. R. and Capecchi, M R., *Nature*, 1988, **336**, 348.
48 Zijlstra, M., Li, E, Sajjadi, F., Subramani, S. and Jaenisch, R., *Nature*, 1989, **342**, 435
49 Schwartzberg, P. L., Goff, S. P. and Robertson, E. J., *Science*, 1989, **246**, 799.
50 Zijlstra, M. *et al.*, *Nature*, 1990, **344**, 742.
51 Thomas, K. R. and Capecchi, M R., *Nature*, 1990, **346**, 847.
52. Ryan, T. M. *et al.*, *Science*, 1990, **247**, 566.
53 Thompson, S., Clarke, A. R, Pow, A. M, Hooper, M L. and Melton, D. W., *Cell*, 1989, **56**, 313.
54 Friedmann, T., *Science*, 1989, **244**, 1275
55 Helene, C. and Toulme, J-J., *Biochim. Biophys Acta*, 1990, **1049**, 99
56 Herskowitz, I., *Nature*, 1987, **329**, 219.
57. Cotten, M., *Trends Biotechnol.*, 1990, **8**, 174.
58 Katsuki, M. *et al*, *Science*, 1988, **241**, 593
59. Stacey, A. *et al*, *Nature*, 1988, **332**, 131
60. Kranz, J E. and Holm, C., *Proc. Natl Acad. Sci. USA*, 1990, **87**, 6629.
61. Sellars, S. and Beighton, P., *Clin. Genet.*, 1983, **23**, 376.
62. Cornell, J., Sellars, S and Beighton, P., *Clin Genet.*, 1984, **25**, 163.
63. Beighton, P., Winship, I. and Behari, D., *Clin. Genet*, 1985, **28**, 69.
64 Ray, A. K., *J. Genet. Hum*, 1987, **35**, 267.
65. Kumar, R., Pereira, B. J., Sakhuja, V. and Chugh, K. S., *Clin. Genet.*, 1989, **35**, 202
66. Kurien, M, Seshadri, M. S, Raman, R. and Sen Bhanu, T, *Arch Otolaryngol. Head Neck Surg*, 1989, **115**, 746.
67. Kumar, D and Blank, C. E, *Indian Pediatr*, 1989, **26**, 1014.
68 Thapa, B. R, Sahni, A. and Mehta, S., *Indian Pediatr.*, 1989, **26**, 82.
69. Lander, E. S and Botstein, D, *Science*, 1987, **236**, 1567.
70. Sundar Rao, P S S, Inbaraj, S G. and Kaliaperumal, V G., *Indian J. Med Res*, 1971, **59**, 294

ACKNOWLEDGEMENTS We thank D. P Kasbekar, Lalji Singh and H Sharat Chandra for their suggestions and comments on the manuscript.