GENOME ANALYSIS

Volume 2

Gene Expression and Its Control

KAY E. DAVIES &
SHIRLEY M. TILGHMAN
editors

GENE EXPRESSION and ITS CONTROL

Edited by

Kay E. Davies
University of Oxford

Shirley M. Tilghman
Princeton University

Volume 2 / GENOME ANALYSIS

Cold Spring Harbor Laboratory Press 1991

Genome Analysis Volume 2
Gene Expression and Its Control

Printed in the United States of America
ISBN 0-87969-359-2
ISSN 1050-8430

Cover and book design by Leon Bolognesi & Associates, Inc.

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory Press, 10 Skyline Drive, Plainview, New York 11803. Phone: 1-800-843-4388. In New York (516) 349-1930. FAX: (516) 349-1946.

Preface

Intensive efforts by many investigators are currently being applied to generate high-resolution genetic and physical maps of two mammalian species, human and mouse. These maps can be considered a compilation of road marks, whose usefulness depends on the average distance between the marks. At the highest level of resolution, they provide immediate molecular access to any region of the genome. Thus, one can realistically anticipate the day when the assignment of a gene or a trait to a chromosomal region will lead to its rapid cloning. Yet this will be a hollow victory if we do not develop in-parallel technologies for understanding the mammalian genome. The commitment of genome organizations throughout the world to the parallel analysis of the human and mouse genomes reflects an appreciation of this fact; i.e., the isolation of a human gene is just the beginning of a much longer process of understanding the function of the gene within the context of the organism. By general consensus, the mouse has become the model mammalian organism in which experimental tests of the function of specific genes will usually be conducted.

This volume is intended to provide a glimpse into the future for a look at how functional analyses of genes and genomic regions are likely to proceed and to emphasize the power of the comparative approach. The chapter by Frank Grosveld and his colleagues illustrates that patterns of transcription of human genes can be studied by transposing the genes into the mouse. The mechanisms underlying seemingly more complex phenomena such as genomic imprinting and X chromosome inactivation are beginning to be amenable to functional analysis as well, as illustrated in the chapters by Bruce Cattanach and Stephen Brown. Finally, to understand the function of a specific gene completely, it is necessary to generate specific mutations in it. Until recently, this has not been possible in the mouse, nor, in fact, in any multicellular eukaryote. The development of mouse embryonic stem cells which can shuttle between a tissue-culture dish and a mouse blastocyst coupled to the application of homologous recombination for specific gene replacement

marks the beginning of a new era in mouse genetic analysis, as described in the chapter by Martin Evans. The identification of mutations in completely new genes through insertional mutagenesis, reviewed by Michael Rudnicki and Rudolf Jaenisch, is likewise certain to yield fertile ground for understanding gene function.

We are grateful to all of the authors in this volume for their hard work and excellent contributions. We are also grateful to the staff of the Cold Spring Harbor Laboratory Press, especially Nancy Ford and her colleagues Dorothy Brown, Virginia Chanda, and Mary Cozza, who have been such enthusiastic partners in the publication of this series, and who have ensured that the series is produced in such a timely fashion.

Kay E. Davies
Shirley M. Tilghman
April 1991

GENE EXPRESSION and ITS CONTROL

Contents

Embryonic Stem Cells as a Route to an Experimental Mammalian Genetics

Martin Evans

The Wellcome Trust, Cancer Research Campaign
Institute of Cancer and Developmental Biology
and Department of Genetics
University of Cambridge
Cambridge CB2 1QR, England

Transgenesis by the use of embryonic stem cells, whereby specific gene alterations may be preselected in vitro, coupled with the more conventional microinjection of DNA into a zygote, now provides a route to a complete experimental genetics system in the mouse. This particular combination of molecular genetics, tissue culture of pluripotential embryonic cells, and whole-animal genetics has opened up a completely new route to exploration of genetic function and dissection of the complex developmental and physiological interactions in the whole animal.

This chapter discusses:

- ❑ embryonic stem cells from the mouse and other animals
- ❑ gene targeting in embryonic stem cells and methods for screening and selection of the cell clones carrying the desired genetic alteration
- ❑ examples of gene loci that have been targeted successfully and use of new mutations for the creation of animal models of genetic disease and for genetic analysis of development
- ❑ potential application of similar technologies to other species
- ❑ technical limitations and possible resolutions

Genome Analysis Volume 2: *Gene Expression and Its Control*
 0-87969-359-2/91 $3.00 + .00

INTRODUCTION

Embryonic stem cells

First isolated and described by Evans and Kaufman (1981), embryonic stem (ES) cells are tissue-culture cell lines of totipotential cells derived from an early embryonic epiblast. They maintain their totipotentiality in tissue culture and, upon reinoculation into a suitable embryo, give rise to chimeric mice in which all the tissues, including the germ cells, have contributions from the original ES cells (Bradley et al. 1984). Thus, a complete route is established between tissue culture and the whole-animal genome.

Germ-line transgenesis

To provide a route to genetic transformation of any specific tissue, it is necessary to transform a cell which is the progenitor of that tissue. Thus, to transform the germ-cell line, it is necessary to transform a cell in the germ-cell lineage. Transgenesis in mammals has therefore concentrated either on the early embryo at a stage when its cells precede any determination into somatic or germ-cell lines or upon transformation of a cell in the germ-cell lineage. In the intact embryo, either technique requires transforming cells that are present in small numbers and are inaccessible to experimental manipulation. They are, moreover, present only transiently in the process of development of the embryo. Of those cells that lead to the germ-cell lineage, to date only ES cells have been successfully maintained as tissue-culture cell lines.

A tissue-culture system has two distinct advantages over experimental systems using a normal embryo: (1) availability for unlimited cell numbers and (2) the protracted time over which manipulations and selections may be carried out. It is therefore possible to transform these cells genetically and to select or screen for infrequent events that lead to the production of the desired genetic alteration (Hooper et al. 1987; Kuehn et al. 1987). Coupled to this is the observation that homologous recombination may take place in mammalian cells between an incoming DNA construct and endogenous chromosomal genes (Smithies et al. 1985). This recombination allows the generation of specific gene alterations and, together with the more conventional procedure of zygote microinjection, has created a methodology for a completely experimental genetics system in the mouse. This work is being rapidly extended to other experimental and domestic animals.

Mammalian genetics

Notwithstanding the great theoretical and practical interest in mammalian genetics, the conventional genetic analysis of mammals has been

handicapped vis-à-vis other experimental animal systems because of relatively small litter sizes and long life cycles and in the case of human genetics, its nonexperimental nature. Mammalian genetics has, however, benefited from the study of human genetics. Although there is no experimental breeding in human systems, detailed observation of a very large population has allowed investigation both of polymorphisms and of very rare mutations, and statistical methods for analysis of data available from human genetic pedigrees have become highly refined (McKusick 1991).

Traditional genetics depends on mutations or preexisting genetic polymorphisms that are discovered in a species. Animal breeding depends on selection from suitable variation, either preexisting in or deliberately introduced into the stock. The major tools for genetic analysis are therefore breeding segregation studies and direct phenotypic analysis. The only experimental approach to widen the scope of genetic variants available for study is mutagenesis followed by specific screening or fortuitous recovery of relevant alleles. Added to this, methods for nonmeiotic genetic analysis—somatic cell genetics and, more recently, direct molecular biological analysis—have been very effectively applied as well as being combined with pedigree analysis so that mammalian genetic maps and knowledge of gene sequence data are advancing rapidly.

Our understanding of genetic function and the practical application of our genetic knowledge in experimental animals require the ability to modify the genome deliberately, preferably in a manner not entirely reliant upon screening the accidents of nature. Thus, the concept of a reverse mammalian genetics emerges, where the effect of specific genetic modification may be studied in the context of the intact organism. Genes of interest are now being identified from the results of intensifying mammalian genetic analysis, through molecular biology techniques, and by cross-homology with similar genes of other species. Both through the identification of important protein products and through direct gene mapping, many loci have been identified in the normal genome for which there are no known mutations and for which there is no known tested function. The only true test of the function in the whole organism is the genetic test of ablation or alteration.

Large classes of mutations in mammals are also systematically unavailable for study. A completely recessive embryonic lethal mutation is one that identifies a gene locus whose function is essential for normal development but which is not haplo-insufficient. Such entirely recessive embryonic lethal mutations are extremely unlikely to be discovered either in humans or in experimental mammals, and the only recessive embryonic lethals seen will be those that have an associated dominant nonlethal phenotype. One might expect that this class of mutation includes many of the most important control loci involved in the specification of the developmental process and that identification of these loci would be

necessary to provide a genetic analysis of development. A complete analysis would need to examine the effect of subtle changes in function at these loci and not merely identify any deletion of function. The technology established with ES cells has provided the opportunity for such direct experimental genetic analysis.

In a great number of cases, genes have been identified in mice, either by analogy with those of other species (e.g., human disease syndromes or *Drosophila* genetics) or from the biochemistry of their protein products, for which there is no mutation of protein structure or lack-of-function allele and thus no rigorous genetic test of function. Further functional studies can only be provided by creating such alleles. In the field of practical application to domestic farm animals, potential alteration of normal physiology may be desirable, and the deletion or modification of the function of controlling genes may be just as important as the overexpression of others.

Although work in this area is proceeding apace, it is still a demanding and time-consuming process, and as yet, a relatively small number of projects have come to fruition. In this chapter, I discuss the new experimental techniques and their applications, some of the results reported to date, and predictions about future directions.

AN EXPERIMENTAL MAMMALIAN GENETICS

ES cells from the mouse

The advent of murine ES cell lines and the subsequent demonstration of their functional totipotency expanded the range of techniques available for gene transfer and gene modification in mammals. In mice, these cells provide an alternative route for transgenesis to that of zygote pronuclear microinjection, and they have the advantage that a large range of in vitro genetic transformation and screening or selective techniques may be applied prior to reconstitution into an animal. In addition, techniques of somatic cell genetics can be used to introduce genetic manipulations into viable and fertile animals. This approach offers unique practical opportunities for experimental genetic manipulation of domestic animals, including targeted mutagenesis by selective inactivation or replacement of endogenous genes, the introduction of clonal alterations to the germ line, and improved control of expression of transgenes through screening the phenotypes of cells in culture.

Gene targeting

Both Smithies et al. (1985) and Thomas et al. (1986) were able to demonstrate that homologous recombination between incoming transfected DNA and chromosomal loci could occur in tissue culture of mammalian

cells. This allows specific genetic alteration of endogenous gene loci when used together with the technique of germ-line reconstitution from tissue culture using ES cells and provides an extremely powerful method for genetic manipulation of any chromosomal gene in the mouse.

To maintain the capacity for germ-line transformation, euploidy of the cells at least for the autosomes must not be grossly affected, and the introduction of other random mutations should not be so severe as to prevent germ-line chimerism and introduction of the transformed cell line into a breeding pool. Randomly cointroduced mutations may be segregated away from any desired mutation during meiosis. I have previously reviewed the potential genetic manipulation of mammals and the techniques available (Evans 1989).

General principles Gene targeting involves treating the cells with a DNA construct that is homologous to the chromosomal target DNA but which contains a region of nonhomology to provide an alteration in structure. In this way, the new desired mutation can be introduced and the modified stem cell clone can be screened or detected in culture. In most cases, this means that a positively selectable cassette (e.g., conferring G418 resistance) is included within the DNA construct, allowing selection for all cells that become stably transformed with the construct and leaving the problem as one merely of differentiating cell clones in which the incoming DNA has become integrated in the homologous site from those where it has become integrated at random.

Alternatives to the directly linked, positively selectable cassette are cotransfection and selection (Reid et al. 1991) and high-efficiency nonselective techniques such as cellular microinjection (Zimmer and Gruss 1989). DNA microinjection can result in up to 10% of the injected cells becoming stably transformed. Similarly, multiple reinfection with retroviral vector suspensions may result in all or the majority of the cells in a population becoming transfected (Robertson et al. 1986). Direct screening for the desired integration is clearly feasible with these two methods. Other methods, however, lead to transformation of very few cells in the population. Although widely different frequencies of transformation are reported, in our experience, electroporation rarely provides stable transformation of better than 1 in 10^4 ES cells treated, whereas the calcium phosphate method provides approximately one tenth of this rate. Lipofection, although potentially a very efficient method of DNA delivery, has in our hands only given about 1 in 10^4 stable transformants (M. Evans, unpubl.). Because of these low transformation efficiencies, it is necessary to use a dominantly selectable marker to isolate transformed cell clones. The most frequently used marker is the G418 resistance conferred by expression of the *neo*^{r} gene originally derived from the bacterial transposon Tn*5*. Selection for hypoxanthine

phosphoribosyltransferase (HPRT) function in HAT medium (hypoxanthine, aminopterin, thymidine) and selection for lack of HPRT function by the use of 6-thioguanine have also been demonstrated to be efficient and nondamaging selective techniques for use with ES cells. It is becoming clear that there is a need for the development of other suitable selection techniques using other dominant markers.

Isolating the targeted clone After genetic addition or targeted alteration is made to the cells in vitro, the required mutated cell clone must be isolated either by screening or by selection prior to reimplantation into the host blastocyst to produce chimeric animals. Although phenotypic screening may be possible in some specific cases, the final analysis must be at the DNA level because the final desired alteration of the cells reflects a specific alteration of the chromosomal DNA.

One of the most sensitive techniques for detection of such alterations is the polymerase chain reaction (PCR) (Saiki et al. 1985). As a screening method, PCR can be used to identify the desired alteration in multiclonal samples and can serve in sib selection techniques for selecting the desired cell clone (Kim and Smithies 1988; Joyner et al. 1989; Zimmer and Gruss 1989). In practice, there are associated problems with PCR both of oversensitivity giving rise to false positive reactions and of the difficulty of being certain of picking up the desired reaction in every case and in every sample. The technique always needs to be followed by verification by, for example, using Southern blotting.

Screening for the desired gene targeting in most cases depends on a PCR fragment of considerable length because this includes one side of the homology in the targeting construct. As the sensitivity and reliability of the PCR are to some extent inversely dependent on the size of the fragment being amplified, this is not an ideal situation. It may be desirable to design the targeting construct with PCR screening in mind and with a specific small indicator mutation near one of the ends of the construct.

Targeted mutagenesis of endogenous genes The first locus to which gene targeting was applied using the ES cell system which resulted in new alleles in the mouse gene pool was the *HPRT* locus. Null alleles at this locus had already been isolated in ES cells in vitro and reduced to the murine germ line, but the advantages of its being selectable in both the forward and reverse directions made this locus a target of choice for the development of gene targeting strategies by homologous recombination. Indeed, the first successful germ-line transmission of a locus introduced into ES cells in vitro by gene targeting was the correction of an

HPRT deletion mutation (Doetschman et al. 1987). Subsequently, gene targeting in ES cells has been used to create a number of mutations, some of which have now been reduced to the breeding germ line. Thus, the action of the mutated gene in vivo has been able to be explored.

A field of particular interest has been the developmental effects and consequences of mutation of various loci to a null phenotype. These have fallen into three natural groups: cellular oncogenes, immune-system-related loci, and loci with a developmental role. For example, cellular oncogenes that have been identified by the effects of their dominant mutations have in most cases no known null mutation, and therefore the normal functional effect of the cellular oncogene is difficult to assess. Because of the drastic effect of mutation and misexpression, it might be thought that such loci represent considerable hostages to fortune and that their maintenance and evolutionary conservation mean that they have essential, irreplaceable normal functions. Numbers of oncogenes have been targeted to null mutations in ES cells to provide a genetic test of their function. The phenotype of the homozygous null mutation has been reported for c-*src*, *Wnt*-1, c-*myb*, and N-*myc* proto-oncogenes (Charron et al. 1990; McMahon and Bradley 1990; Stanton et al. 1990; Thomas and Capecchi 1990; Mucenski et al. 1991; Soriano et al. 1991). c-*src* and *Wnt*-1 have been bred to homozygosity, and in both cases, live mice were born, demonstrating that all the complex processes of fetal cellular development are not dependent on expression of these genes. In the case of c-*src*, the surprising phenotype was that of failure of bone remodeling (osteopetrosis) and with *Wnt*-1, the abnormalities, although severe, were restricted to the midbrain and cerebellum. Both are areas in which the wild-type allele is expressed, but no simple correlation exists between normal gene expression and developmental perturbation.

Mucenski et al. (1991) have shown that c-*myb* null mutation mice develop normally to day 13, but become severely anemic as the hepatic hematogenesis falls (about day 15). That they develop normally this far is surprising in view of the widespread expression of c-*myb* (including ES cells; Dyson et al. 1989) and its postulated role in cell cycle control. These studies show that its critical function is in hepatic, but not yolk sac, hematopoiesis.

Genes involved in immune signaling and response have also provided a fertile field for investigation. Two examples of loci where gene targeting has been used to give experimental novel mutations are the β_2-microglobulin and immunoglobulin μ chain. β_2-microglobulin is an obligatory associate for expression of the major histocompatibility complex (MHC) class I genes on the cell surface, and loss of β_2-microglobulin expression might have been expected to cause severe problems of cell-cell recognition during development and immune system dysfunction. Results from two independent groups in fact show that mice with

no β_2-microglobulin develop normally, except for a deficit in cytotoxic T cells (Koller et al. 1990; Zilstra et al. 1990). Kitamura et al. (1991) introduced a disruption of the membrane exon of the immunoglobulin μ chain that, by preventing cell surface expression of IgM on pre-B cells as had been predicted, prevented the establishment of the B-cell arm of the immune system in the mice.

A third particularly active area of gene targeting is aimed at testing the function of loci with a putative developmental role, such as the homeobox loci. Chisaka and Capecchi (1991) have described an extensive syndrome of effects arising from homozygosity of a null mutation at *hox-1.5* that results in death shortly after birth. This series of defects is located in a region of the embryo that represents a region of *hox-1.5* expression but does not correlate in any simple way with the expression of *hox-1.5* in general; e.g., normal expression of the gene in lungs, stomach, and spleen is not reflected by a mutant phenotype in these organs. Joyner et al. (1989, 1991) have disrupted the *En-2* gene of mice—a locus that by homology with *Drosophila* was expected to be of great importance in developmental pattern formation—and have demonstrated that the resulting homozygous null mutant mice are phenotypically grossly normal but have a subtle cerebellar alteration. The scope of the phenotypic effect is much less than the field over which the wild-type gene is expressed. These results are unexpected and reemphasize the *experimental* nature of these investigations and their importance in resolution of gene function in mammalian development.

POTENTIAL APPLICATION TO OTHER SPECIES

There has been considerable effort to derive ES cell lines from a number of other species. ES cell lines with the proven ability to produce germline transmission from the chimera have so far been described only in mice, although putative ES cell lines have been described in hamster, pig, sheep, and cattle (Doetschman et al. 1988b; Notarianni et al. 1991). To date, none of the putative ES cell lines from other species have demonstrated the capacity to make germ-line chimeras, which is the essential property needed for transgenesis via these cells. In the larger farm animals, there is a considerable time constraint from the life cycle, and in nonmurine experimental animals, the technology of embryo manipulation and transfer still needs to be perfected. We are presently attempting to isolate ES cells from the laboratory rat, the experimental animal of choice for many pharmacological and physiological studies, where much of the experimental embryology is available.

It is expected that mice will remain the premier experimental genetic species, although adapting these techniques to larger domestic farm animals would open the way to a more carefully controlled trans-

genesis in these species. The results of Petitte et al. (1990), who were able to produce germ-line chimeric chickens by transfer of small numbers of transiently cultured epiblast cells, open the possibility that this technology may become applicable to avian species as well.

TECHNICAL LIMITATIONS AND POSSIBLE RESOLUTIONS

Although ES cell technology is clearly proving to be a very successful method for introducing experimental manipulations of endogenous gene sequences, it is at present laborious and requires expertise in three different laboratory skills: molecular biology, tissue culture, and experimental embryology. Fortunately, refinement of procedures in these areas promises to make the technology much more tractable.

I have previously reviewed the types of constructs that can be used (Evans 1989) and in particular, the technique of positive followed by negative selection for a construct which first goes into a targeted locus and then is removed either by the in-out mechanism described by Valancius and Smithies (1991) and Hasty et al. (1991) or by a second round of targeting where the new homology inserted in the first round is used to remove most of the construct leaving a specific mutation. These techniques lend themselves to construction of genomic clones with a specific series of cassettes all of which are becoming well established.

The second phase—that of ES cell culture and selection or screening of targeted clones—can be very time-consuming, especially when individual clones are grown in a large population prior to screening. Techniques that allow the use of multiwell plates such as those described for freezing (Chan and Evans 1991) and methods under development to allow rapid production of sufficient DNA for Southern blots (coupled with some degree of automated handling) should greatly improve this situation. The technique of immediate PCR analysis of clones (Joyner et al. 1989), or other techniques that can be applied to early screening of clones, will have significant advantages in that there is much less volume of tissue-culture work. In addition, the immediate identification of successfully targeted clones allows a shorter period of cell culture passage, which lessens the chance of loss of the clone's totipotency.

Experimental embryology techniques, including construction of chimeras and test breeding, also present difficulties at the moment. For example, Schwartzberg et al. (1989) reported the relative effectiveness of different host blastocysts. The C57BL/6 inbred mouse line is at the moment the preferred donor of host blastocysts, but unfortunately, these animals have a relatively low fertility rate. Techniques of introducing the ES cells into the embryos at the morula stage rather than blastocyst stage have been investigated with success by some researchers.

It is also clear that the maintenance of totipotentiality of the cells in culture and their ability to produce germ-cell chimeras are of critical importance. Although the various conditions for ES cell growth on STO feeders (Martin and Evans 1975), embryonic fibroblast feeders (Doetschman et al. 1988a), or without feeders but supplemented with either buffalo rat liver (BRL)-conditioned medium or recombinant LIF, have all been satisfactorily used (Smith et al. 1988; Williams et al. 1988), it is clear that any culture crisis may lead to aneuploidy and/or lack of good chimerization potential. Attention to exemplary tissue-culture conditions is clearly vital. Numbers of cell lines have been shown to be effective vectors for germ-line chimerism, and there is little reason to suppose that other such lines are not readily isolated and maintained. It is clear, however, that further improvements to the rigor of the tissue-culture protocols and establishment of well-characterized ES cell stocks will be of great utility. In all, I confidently anticipate that the procedures for the use of ES cells for selective gene modification in mice and other mammals will become much more of a routine process than it is today. It should become the method of choice for experimental genetic study in the complexity of the whole organism.

Acknowledgments

The author is supported by the Wellcome Trust.

References

Bradley, A., M. Evans, M.H. Kaufman, and E. Robertson. 1984. Formation of germ-line chimeras from embryo derived teratocarcinoma cells. *Nature* **309:** 225.

Chan, S.Y. and M.J. Evans. 1991. In situ freezing of embryonic stem cells in multiwell plates. *Trends Genet.* **7:** 76.

Charron, J., B.A. Malynn, E.J. Robertson, S.P. Goff, and F.W. Alt. 1990. High-frequency disruption of the N-myc gene in embryonic stem and preB cell lines by homologous recombination. *Mol. Cell. Biol.* **10:** 1799.

Chisaka, O. and M.R. Capecchi. 1991. Regionally restricted developmental defects resulting from targetted disruption of the mouse homeobox gene hox 1.5. *Nature* **350:** 473.

Doetschman, T., N. Maeda, and O. Smithies. 1988a. Targetted mutation of the Hprt gene in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **85:** 8583.

Doetschman, T., P. Williams, and N. Maeda. 1988b. Establishment of hamster blastocyst-derived embryonic stem (ES) cells. *Dev. Biol.* **127:** 224.

Doetschman, T., R.G. Gregg, N. Maeda, M.L. Hooper, D.W. Melton, S. Thompson, and O. Smithies. 1987. Targetted corrections of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330:** 576.