Somatic Cell Genetics and Gene Families

Peter D'Eustachio and Frank H. Ruddle

Summary. The utility of somatic cell genetic analysis for the chromosomal localization of genes in mammals is well established. With the development of recombinant DNA probes and efficient blotting techniques that allow visualization of single-copy cellular genes, somatic cell genetics has been extended from the level of phenotypes expressed by whole cells to the level of the cellular genome itself. This extension has proved invaluable for the analysis of genes not readily expressed in somatic cell hybrids and for the study of multigene families, especially pseudogenes dispersed in different chromosomes throughout the genome.

In the past 25 years, enormous effort has gone into the construction of genetic maps in humans and mice. What information concerning gene function derives from such an anatomization of mammalian genomes? Extensive changes in gene order and chromosome structure are possible even between closely related species. At the same time, conservation of linkage groups between species has also been found. For example, 23 groups conserved between humans and mice have been described (*1*). While functional relationships among members of a linkage group are not always clear, preservation of the linkage over the evolutionary time separating humans and mice suggests that a functional relationship exists. If the mammalian genome is a patchwork, understanding its evolution and present function must be grounded in knowing the contents and the boundaries of the patches.

Mendelian genetic analysis in mammals, especially humans, is difficult, however. Generation times are long, numbers of progeny are small, and matings involving genetic polymorphisms of interest can be found only by inspection. The genetic analysis of somatic cells in culture has provided a useful alternative approach. The experimental strategy is straightforward. Hybrid cell lines that carry various numbers and combinations of chromosomes derived from two different mammalian species are constructed in culture. Determining which species-specific markers such a somatic cell hybrid expresses as a function of which chromosomes from each species it retains often allows the chromosomal locations of the genes encoding the markers to be deduced.

This strategy is limited to phenotypes detectable in individual cells in culture, such as constitutively expressed enzymes and cell surface markers—for example, major histocompatibility antigens (*2, 3*). It is therefore no surprise that recombinant DNA methodology has revolutionized the genetic analysis of somatic cell hybrids. Any gene—indeed, any DNA sequence—for which a recombinant DNA probe exists can now be mapped. In this article, after a brief review of the basic methodology, we discuss the kinds of results obtained by DNA-level analysis of somatic cell hybrids. These have been especially striking in the cases of oncogenes and of multigene families.

Gene mapping in somatic cell hybrids. The experimental strategy is diagrammed in Fig. 1 (*4*). A mixture of two different cell types is treated with an agent that induces membrane fusion, such as polyethylene glycol or inactivated Sendai virus. The nuclei of the fused cells can themselves fuse to yield a single nucleus carrying all of the chromosomes from the parental cells. Hybrids containing one or two sets of chromosomes from each parental cell type are recovered by culturing the fused cell mixture under selective conditions. If the parental cells contain complementing auxotrophic markers, the fusion mixture is cultured in minimal medium. If parental cells contain dominant drug-resistance markers, those drugs are added. Unfused cells and fusions of like cells die, while hybrids between unlike cells survive. Fusions of multiple cells of both types typically are viable but unable to divide.

The crucial point is that, as the hybrid cells grow in culture, chromosomes of one parental type (by convention, the donor) are lost while a complete set of chromosomes from the other parent (the recipient) is retained. Loss of donor chromosomes is largely random (*5*). The physiological basis of this phenomenon is poorly understood, but the phenomenon is reproducible. Hybrids between human and rodent cells tend to lose human chromosomes. Those between mouse and Chinese hamster tend to lose mouse chromosomes. The chromosomal content of a hybrid cell clone at any point in its growth can be assessed in two ways. First, it can be scored for donor-specific isoenzyme markers whose chromosomal location is known, a rapid but qualitative assay. Second, karyotypic analysis, by standard banding procedures to identify chromosomes in hybrid cells in metaphase allows the exact frequency of each donor chromosome to be determined (*2, 3*).

Panels of hybrid cell clones carrying various combinations of human or mouse chromosomes on constant mouse or Chinese hamster backgrounds, respectively, have thus been assembled. By scoring panel members for the presence or absence of a donor cell phenotype and the presence or absence of the various donor chromosomes, phenotypic markers are assigned to specific chromosomes.

Mapping recombinant DNA probes in somatic cell hybrids. Here, the marker is a donor-specific DNA sequence. The technique was first applied to map α- and β-globin genes in human × mouse somatic cell hybrids. These were detected by DNA hybridization in solution. Human α- and β-globin complementary DNA (cDNA) probes reanneal to form stable duplexes with the corresponding human genomic DNA sequences, and do not cross-react with each other or with mouse globin genes. The presence or absence of duplex material in a hybridization reaction containing human globin cDNA and hybrid cell genomic DNA thus indicates the presence or absence of the human globin gene in the hybrid. Scoring panels of hybrids in this way, α-

Peter D'Eustachio is an assistant professor in the Department of Biochemistry, New York University Medical Center, 550 First Avenue, New York 10016. Frank H. Ruddle is a professor in the Department of Biology, Yale University, Kline Biology Tower, New Haven, Connecticut 06511.

globin genes were assigned to chromosome 16 and β-globin genes to chromosome 11 (*6*).

Liquid hybridization has two drawbacks. First, it requires large amounts of DNA. More important, if the donor gene probe cross-reacts with the homologous recipient cell gene, interpreting the assay is difficult. The range of genes assayable by liquid hybridization is thus limited. The blotting procedure of Southern (called Southern blotting) avoids these difficulties (*7*). Genomic DNA isolated from somatic cell hybrids is digested with restriction endonucleases to yield specific DNA fragments. These are fractionated according to size by means of agarose gel electrophoresis. The fractions are denatured, transferred to nitrocellulose filters, and hybridized on the filter with isotopically labeled DNA probes (*7, 8*). Noncoding sequences flanking and intervening in a structural gene diverge rapidly, so that even genes highly conserved between two species can usually be resolved as DNA fragments of different size. In this way, the chromosomal locations of numerous genes and gene clusters in humans (Table 1) and mice (Table 2) have been determined.

Southern blotting is as sensitive as other methods for gene mapping in somatic cell hybrids, for example, isoenzyme assays. Donor genes present at dosages of 0.2 per pseudodiploid hybrid cell can reliably be detected. This sensitivity varies only slightly from gene to gene, and from laboratory to laboratory. The technique is especially useful for comparative gene mapping experiments within a species. Once a panel of somatic cell hybrids is assembled, the small amount of DNA used for any one mapping experiment (30 micrograms or less) allows extensive, internally consistent mapping experiments to be carried out with material obtained from a somatic cell hybrid population of moderate size. Purified genomic DNA is stable, and therefore experiments can be carried out over a period of years on material derived from a hybrid at a single stage of donor chromosome segregation (*9*).

These mapping procedures are equally applicable to DNA sequences whose functions are unknown. Fragments chosen at random from genomic DNA libraries have been systematically screened to identify those containing a single-copy sequence DNA derived from the murine X chromosome (*10*) and from human chromosomes 10 and 22 (*11*).

Is gene organization on a donor chromosome faithfully preserved in somatic cell hybrids? The mapping procedure itself is a sensitive assay for gene rearrangement. Any mutation at a restriction site and any insertion or deletion of more than a few hundred base pairs of DNA would yield restriction fragments of altered mobility. No rearranged genes have been found. The fractions of the human and murine genomes tested in this way so far are small. Nevertheless, the results are consistent with the idea that donor DNA sequences are preserved unaltered in somatic cell hybrids even in the absence of any identifiable selective pressure for their conservation.

Cloned DNA markers from somatic cell hybrids. A DNA cloning strategy based on the use of somatic cell hybrids has been developed that allows a large number of DNA markers to be localized to a particular chromosome or chromosomal region. Genomic DNA fragments isolated from a somatic cell hybrid containing a single human donor chromosome or chromosome fragment are cloned into bacteriophage to generate a library. Species-specific reiterated sequences are widely dispersed throughout both the human and rodent genomes (*12*), and a typical recombinant phage contains several. The few percent of recombinant phage containing human DNA is identified by transferring DNA from phage plaques onto a nitrocellulose filter and hybridizing the filter with an isotopically labeled probe corresponding to human reiterated DNA sequences (*13*). This procedure allowed DNA fragments from human chromosomes 2, 3, 5, 6, 11, 12, 14, 16, 20, and X to be isolated (*13–16*). We now have data suggesting that the same technique can be used to isolate mouse DNA fragments from mouse × Chinese hamster somatic hybrid cell lines. Application of these techniques to hybrid cell lines generated by chromosome- or DNA-mediated gene transfer, which typically contain small donor chromosome fragments incorporating a selectable marker gene (*17*), has led to the cloning of such markers (*18*). It should be possible to work out the structure of the surrounding chromosomal regions at a level of resolution not previously attainable in higher eukaryotes.

In situ hybridization. The procedures discussed so far allow a gene to be localized to a chromosome or chromosomal region indirectly, as a result of a multistep analysis. In situ hybridization techniques permit genes to be visualized directly on metaphase chromosomes, after hybridization with labeled probes. In mammals, repeated DNA sequences occurring in tandem arrays such as ribosomal RNA genes have been localized (*19*). Recent results suggest that single-copy DNA sequences may likewise be detectable. Harper and Saunders mapped a cloned 14.9-kilobase (kb) human genomic DNA fragment of un-

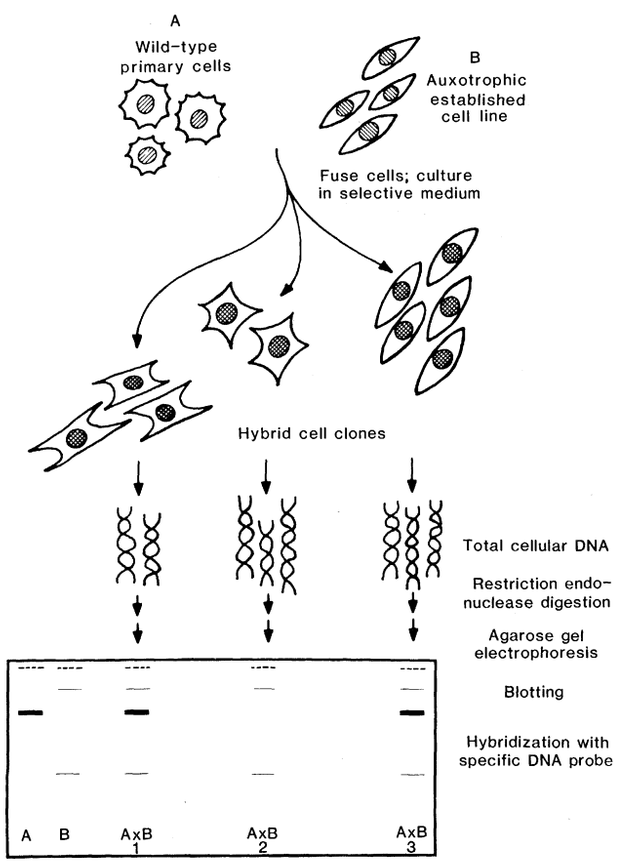

Fig. 1. Gene assignment by restriction mapping in somatic cell hybrids. Human genes assigned to chromosomes in this way, with human (A) × rodent (B) hybrids are listed in Table 1; mouse genes assigned with mouse (A) × Chinese hamster (B) hybrids are listed in Table 2.

known function to the telomeric region of the short arm of chromosome 1 (*20*). Wahl *et al.*, using a cloned 8.6-kb probe have mapped aspartate transcarbamylase to the short arm of chromosome B9 in the Syrian hamster (*21*).

The technique is subject to two limitations. First, long stretches of single-copy sequence DNA are generally required as probes. Second, dextran sulfate, needed to enhance hybridization sensitivity, interferes with chromosome banding, needed for precise gene localization. Nevertheless, the technique has been used to confirm chromosomal assignments of several human genes (Table 3). The technique may prove most powerful in conjunction with somatic genetic and Mendelian analysis. Analysis of somatic cell hybrids provides a rapid chromosomal assignment; in situ hybridization allows confirmation and more precise localization; and with this information in hand appropriate material can be sought for family studies and pedigree analysis. Human immunoglobulin heavy chain constant region genes are a paradigm, having been mapped in this way to band q32.3 of chromosome 14 (*22*).

Recombinant DNA techniques have led, then, to several general and rapid gene mapping procedures. Beyond expediting the process of gene cataloging, they provide access to genes not readily studied otherwise. Products of differentiated cells are easily mapped, as are highly conserved genes and DNA sequences not expressed in normal tissues. Gene mapping carried out at the level of phenotypic markers is ambiguous in that polymorphism in structural genes cannot be distinguished from polymorphism in a regulatory gene acting on a constant set of structural genes. DNA-level mapping can detect structural genes unambiguously. The possibility thus arises not only of mapping structural genes, but of then constructing breeding experiments, or studies of gene expression in somatic cell hybrids, to define and map these regulatory genes as well. The complementary mapping strategy—the use of recombinant DNA probes derived from somatic cell hybrids to define markers of particular chromosomes—makes the entire genomes of humans and mice accessible for detailed genetic analysis.

The mapping data alone can give some hints as to gene function. Oncogenes provide a striking example. These genes were originally identified in murine and avian RNA tumor viruses, where they encode the malignant transformation of the infected host cells. These viral oncogene sequences have seen shown to be homologous to sequences in the germline genomes of humans and mice. Several of

Table 1. Human genes assigned by restriction analysis. The capital letters in parentheses are the gene names tentatively suggested for adoption (*1*).

Name	Chromosomal location	Type	DNA polymorphism	Reference
Single-copy genes (S) and compact families (C)				
C-fos homolog (CFOS)	2	(S)	–	(*26*)
Ig κ light chains (IGK)	2	(C)	–	(*49*)
Proopiocortin (POC)	2p		–	(*50*)
C-myb homolog (CMYB)	6	(S)	–	(*26*)
Prolactin (PRL)	6	(C)	–	(*50*)
Collagen, type I, α2 (COL1A2)	7		–	(*51*)
C-mos homolog (CMOS)	8	(S)	–	(*26*)
C-myc homolog (CMYC)	8	(S)	–	(*26*)
α-Interferon (IFL)	9pter–q12	(C)	–	(*36*)
β-Interferon (IFF)	9pter–q12	(S)	–	(*36*)
C-abl homolog (CABL)	9	(S)	–	(*26*)
C-ras^{H} homolog (CRASH)	11p13	(S)	+	(*26*)
Insulin (INS)	11p15.5–p13	(S)	+	(*52*)
β-Globins (HBB)	11p1208–p1205	(C)	+	(*6, 53*)
C-ras^{k} homolog (CRASK)	12	(S)	–	(*26*)
Ig heavy chains (IGH)	14q32.3	(C)	+	(*22*)
C-fes homolog (CFES)	15	(S)	–	(*26*)
α-Globins (HBA)	16pter–p11	(C)	+	(*6, 54*)
Chymotrypsinogen B (CTRB)	16		–	(*55*)
Growth hormone (GH), chorionic somatomammotropin (CSH), and placental lactogen (PL)	17q21–qter	(C)	–	(*56*)
Collagen, type I, α1 (COL1A1)	17q21–qter	(S)	–	(*57*)
C-src homolog (CSRC)	20	(S)	–	(*26*)
C-sis homolog (CS1C)	22q11–qter	(S)	–	(*26*)
Dispersed families				
Argininosuccinate synthetase (ASS)	6,9,X,Y,*		+	(*32*)
Immunoglobulin λ light chains (IGL)	*22*,*		+	(*46, 49*)
β-Tubulin	*		–	(*44*)

*Family members reside on chromosomes in addition to the ones named, but cannot be definitively assigned on the basis of available data. The chromosome carrying the functional gene is in italics.

Table 2. Murine genes assigned by restriction analysis.

Name	Chromosomal location	Type	DNA polymorphism	Reference
Single-copy gene (S) and compact families (C)				
Proto-oncogene *c-abl*	2	(S)	–	(*23*)
Major urinary protein (*Mup*)	4	(C)	–	(*58*)
Proto-oncogene *c-mos*	4	(S)	–	(*24*)
Kidney cDNA (that is, pMKA11, locus RP1)	4	(S)	+	(*59*)
α-, β-, γ-caseins; casein-like cDNA (that is, pCM 40)	5	(C)	–	(*37*)
J protein	5	(S)	–	(*60*)
α-Fetoprotein (*Afp*), albumin (*Alb*)	5	(C)	–	(*38*)
Ig κ light chains (*Igk*)	6	(C)	–	(*34*)
β-Actin	6	?	–	(*47*)
Myosin light chain 2	7	(S)	–	(*47*)
Kidney cDNA (pMK208,908; locus RP2)	7	(S)	+	(*59*)
Myosin heavy chain	11	(C)	–	(*47*)
Ig heavy chains (*Igh*)	12C2→ter	(C)	+	(*35*)
Nonimmunoglobulin-associated rearranging DNA fragments (2)	15	?	–	(*25, 27*)
Proto-oncogene *c-myc*	15	(S)	–	(*25*)
λ light chains (*Igl*)	16cen→B5	(C)	–	(*33*)
Dispersed families				
Ribosomal protein S16	5*		–	(*30*)
Ribosomal protein L18	7,12*		–	(*30*)
Ribosomal protein L19	6*		–	(*30*)
Ribosomal protein L30	3,12*		–	(*30*)
α-Globins	*11*,15,17		–	(*31*)
Pro-opiomelanocorticotropin (POMC)	*12*,19		–	(*67*)
Kidney cDNA pMKA17	14*		+	(*59*)
α-Actin	2,*3*,17*		–	(*47*)

*Additional family members are present on chromosomes other than the one or those named, but cannot be definitively mapped on the basis of available data. Chromosomes carrying functional genes are indicated in italics, if known.

Table 3. Human genes and DNA fragments assigned by in situ hybridization.

Name	Chromosomal location	Reference
DNA fragment D1S1	1p36	(*20*)
Immunoglobulin κ light chains (IGK)	2p	(*61*)
Histones (H1, H2A, H2B, H3, H4)	7q21 → 36	(*62*)
C-mos homolog (CMOS)	8q22	(*26*)
C-myc homolog (CMYC)	8q24	(*26*)
α- and β-Interferons (IFL, IFF)	9p21 → pter	(*66*)
Insulin (INS)	11p15	(*63*)
γ-Interferon	12q24.1	(*66*)
Immunoglobulin heavy chains (IGH)	14q32	(*22*)
α-Globin (HBA)	16	(*54, 64*)
Collagen, type I, α1 (COL1A1)	17q21 → q22	(*57*)
Growth hormone (GH) gene cluster	17q22 → q24	(*65*)

these cellular sequences have well-characterized RNA and protein products. In no case, however, has a normal cell function been associated with any of these genes. The murine and human oncogene homologs are widely scattered throughout their respective genomes (*23–27*). Strikingly, three of them, human C-CIS and C-ABL, and murine *c-myc*, lie in chromosomal regions associated with specific translocations of lymphoid neoplasms. Two other mouse DNA sequences, discovered because of their association with aberrantly rearranged immunoglobulin heavy chain genes in myeloma cell lines, have been shown by mapping experiments to contain subregions derived from chromosome 15 in one case, and from chromosome 6 and 15 in the other covalently attached to subregions derived from the *Igh* region of chromosome 12. These fragments may be the junction sequences formed in myeloma-specific translocations involving chromosomes 6, 12, and 15 (*25*).

Intraspecies polymorphisms in the mapped DNA sequences are crucial in these mapping analyses. They provide the link between the physical maps generated by the techniques discussed so far and the linkage maps of Mendelian genetics. The data in Tables 1 and 2 probably underestimate the frequency of polymorphism. In many cases, limited searches for polymorphic variants were made. More important, the DNA sequences listed may not be representative of the genome as a whole. Many were isolated because they contain unusually extended regions of single-copy sequence DNA. Many human and murine DNA polymorphisms, however, appear to be due to insertions, deletions, or rearrangement of reiterated sequences (*28, 28a*). The calculation of Botstein *et al.* (*29*) that about 15 percent of all restriction fragments in both humans and mice should contain polymorphic regions may prove correct.

Multigene families. The most striking results to emerge from these studies to date concern multigene families. The existence of families of related DNA sequences, such as those encoding immunoglobulins, globins, major histocompatibility antigens, and so forth, is well established. It is likewise well established that many of these families include pseudogenes—closely homologous DNA sequences altered so as to prevent transcription of functional messenger RNA's (mRNA) or translation of functional proteins (or both). The surprising result yielded by DNA-level mapping of these families was that many were widely dispersed in the genome. Members of four murine ribosomal protein gene families have been mapped. All are dispersed to two or more chromosomes (*30*). Likewise, the two functional murine adult α-globin genes and one embryonic α-globin gene form a tightly linked cluster on chromosome 11. Two pseudo-α-globin genes have been identified that clearly do not map to chromosome 11, but to chromosomes 15 and 17 (*31*). At least ten DNA sequences homologous to an argininosuccinate synthetase cDNA probe exist in the human genome, dispersed over at least eight autosomes and both sex chromosomes (*32*) (Fig. 2).

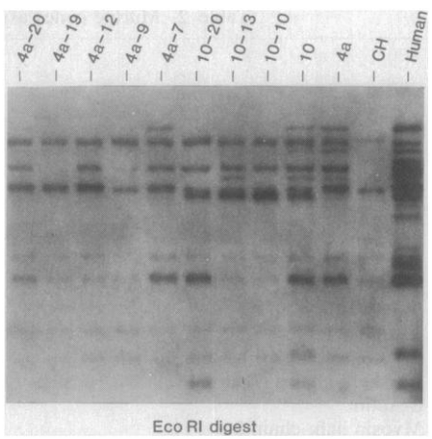

Fig. 2. Analysis of the human argininosuccinate synthetase multigene family. Human × Chinese hamster somatic cell hybrids were generated as outlined in Fig. 1. Total genomic DNA extracted from each of ten hybrid clones (tracks 4a–20, . . . , 4a), Chinese hamster cells (*CH*), and human cells was digested to completion with Eco RI restriction endonuclease, fractionated by size in an agarose gel (direction of migration is from top to bottom as shown), blotted, and hybridized with a cDNA probe corresponding to human argininosuccinate synthetase. Different hybrids contain different subsets of the complex of DNA fragments found in the human parent, indicating wide dispersion of this gene family over the human genome. [Courtesy of *Cell*; from Beaudet *et al.*, in (*32*)]

Not all gene families are dispersed. In the mouse, all immunoglobulin λ light chain genes so far detected—three functional constant region (C) λ genes, one pseudo C λ gene, and two functional variable region (V) λ genes—are clustered on the proximal portion of chromosome 16 (*33*). All κ light chain genes are clustered on mouse chromosome 6 (*34*). Likewise, all murine heavy (H) chain C genes, and all heavy chain V genes so far examined, form a linked cluster on the distal portion of chromosome 12 (*35*). Attempts to find V_H-like DNA sequences elsewhere in the genome have so far been unsuccessful. All detectable members of the human α-interferon gene family reside on chromosome 9 (*36*). Three murine casein genes, and one casein-like gene, are clustered on chromosome 5 (*37*), as are the pair of evolutionarily and functionally linked genes α-fetoprotein and albumin (*38*). Similarly, all human and murine DNA fragments detected by cDNA probes corresponding to major histocompatibility complex genes map to human chromosome 6 and mouse chromosome 17, respectively (*39, 40*).

The possibility remains that genes are dispersed from these families to novel locations on the same chromosome, although mechanisms for such specific dispersion are hard to envision. Chromosome walking experiments directly support the notion of clustering for substantial parts of the murine gene families encoding immunoglobulin heavy chains and the major histocompatibility complex (*40, 41*). More data are needed, but it is attractive to speculate that gene families differ significantly from one another in their susceptibility to dispersion.

Mechanisms for gene dispersion. On the basis of recent structural studies of dispersed genes four mechanisms seem relevant to account for the presence of differentially mobile genes. The first is rescue of a host cell gene by recombination with a retroviral sequence (*42*). Inte-

gration of the recombinant retrovirus into the genome of a germ cell would give rise to a dispersed copy of the original host cell gene. The copy would lack intervening sequences and be flanked by retroviral long terminal repeats. All host cell sequences could be dispersed in this way. Nevertheless, no dispersed gene exactly fitting this description has been found, although the murine $\alpha\psi 3$ globin gene differs from it only in the orientation of the flanking long terminal repeat sequences with respect to one another (*28a*).

Second, DNA sequences with the organization of transposons might mediate the generation of additional exact copies of themselves widely dispersed over the genome. The interspersed short repeated sequences characteristic of mammalian genomes, such as the Alu family, may themselves form transposons (*12, 43*). Any stretch of DNA flanked by such repeated elements could be moved unaltered, although it would be flanked in its new location by an additional small directly repeated DNA sequence (*43*). No case of gene dispersion by this mechanism in mammals has been described.

The third mechanism, like the first, involves an RNA intermediate. Here, however, a cellular RNA species is presumed to program synthesis of a cDNA molecule that is inserted into a chromosome. Several dispersed human β-tubulin pseudogenes have a cDNA-like structure. Their homology to the functional gene begins at a point corresponding to the normal poly(A) (polyadenylate) addition site. In the pseudogene, this site is followed by an A-rich nucleotide stretch. Intervening sequences found in the functional gene are precisely deleted in the pseudogenes (*44*). Weiner and colleagues have shown that the human small nuclear RNA U3 can take on a secondary structure that, in vitro, programs the synthesis of a cDNA corresponding to its 5′ 72-nucleotide residues. This 72-bp nucleotide is found widely dispersed in the human genome (*45*). DNA sequences transcriptionally active in germ cells should be susceptible to dispersion by this mechanism. It is striking that gene families such as immunoglobulins, the major histocompatibility complex, globins, and α-interferons, which largely lack dispersed members, are also transcribed only at very low levels, if they are transcribed at all, in germ cells (*46*). Extensively dispersed families such as human β-tubulin and argininosuccinate synthetase are transcriptionally active in germ cells. A crucial test of the generality of this model will come with the chromosomal mapping and structural analysis of pseudogenes from families some functional members of which are expressed in germ cells and some members of which are expressed only in somatic tissue (*47*).

The fourth mechanism is gene conversion. Originally described in fungi, gene conversion has been invoked as a mechanism to explain very highly homologous regions within human globin gene clusters. It may have a general role in generating sequence homogeneity within gene clusters, and closely related events could be responsible for variation in the sizes of these clusters. The apparent requirement for regions of perfect homology flanking the two sequences involved in the conversion event, however, appears incompatible with gene conversion being a mechanism for exporting sequences from such a cluster to unrelated, distant sites in the genome (*48*).

All these models are based on examination of a tiny fraction of the genome. Nevertheless, the gene mapping techniques described in the first part of this article provide a rapid, general method for identifying compact and dispersed gene families and for dissecting apart the members of the latter. In conjunction with DNA cloning and sequencing approaches, they should allow the structures of these families to be worked out in detail and thus begin to provide an insight into their evolutionary origins and present-day functions.

References and Notes

1. P. L. Pearson, T. H. Roderick, M. T. Davisson, P. A. Lalley, S. J. O'Brien, in (*2*), p. 208.
2. *Human Gene Mapping, 6th International Workshop* (Karger, Basel, 1982).
3. H. L. Foster, J. D. Small, J. G. Fox, Eds., *The Mouse in Biomedical Research* (Academic Press, New York, 1981), vol. 1.
4. F. H. Ruddle, *Nature (London)* **294**, 115 (1981); _____ and R. Creagan, *Annu. Rev. Genet.* **9**, 407 (1975); N. R. Ringertz and R. E. Savage, *Cell Hybrids* (Academic Press, New York, 1976).
5. Exceptions to the pattern of random donor chromosome loss include the failure of mouse × Chinese hamster hybrids to retain mouse chromosome 11 and the preferential loss of human chromosome 9 from human × mouse hybrids (but not from human × Chinese hamster hybrids). For any particular combination of parental cell types, such phenomena are reproducible. Their physiological basis is unclear.
6. A. Deisseroth *et al.*, *Cell* **12**, 205 (1977); A. Deisseroth, A. Nienhuis, J. Lawrence, R. Giles, P. Turner, F. H. Ruddle, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1456 (1978).
7. E. Southern, *J. Mol. Biol.* **98**, 503 (1975).
8. G. M. Wahl, M. Stern, G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3683 (1979).
9. These procedures do not alter the phenomenon of continuing donor chromosome loss from somatic cell hybrids, but only increase the amount of mapping data obtainable from a hybrid cell panel at one point in its history. Constructing hybrids that stably retain small numbers of donor chromosomes remains a problem. Directly selectable marker genes are available for three mouse and human chromosomes (hypoxanthine phosphoribosyltransferase (E.C. 2.4.2.8)—mouse X and human Xq26–q27; adenine phosphoribosyltransferase (E.C. 2.4.2.7)—mouse 8 and human 16q12–q22; thymidine kinase (E.C. 2.7.1.21)—mouse 11 and human 17q21–q22). Laboratory strains of mice have been developed that together carry each mouse chromosome except 6 and Y translocated onto a selectable one [M. C. Green, Ed., *Genetic Variants and Strains of the Laboratory Mouse* (Fischer Verlag, Stuttgart, 1981)]. Human primary fibroblast lines carrying translocations involving these selectable markers and various other chromosome fragments together cover about two-thirds of the human genome [*Human Genetic Mutant Cell Repository* (NIH publication 820-2011, Bethesda, Maryland, ed. 8, 1982)]. Microcell-mediated chromosome transfer allows individual translocation chromosomes from these sources to be introduced into recipient cells [R. E. K. Fournier, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6349 (1981)]. Several hybrid lines have been constructed, but a systematic dissection of either the human or the murine genome into single-chromosome somatic cell hybrids remains to be done.
10. S. F. Wolf, C. E. Mareni, B. R. Migeon, *Cell* **21**, 95 (1980); S. F. Wolf and B. R. Migeon, *Nature (London)* **295**, 667 (1982).
11. F.-T. Kao, J. A. Hartz, M. L. Law, J. N. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 865 (1982).
12. C. W. Schmid and W. R. Jelinek, *Science* **216**, 1065 (1982); W. Jelinek and C. Schmid, *Annu. Rev. Biochem.*, in press.
13. J. F. Gusella, C. Keys, A. Varsani-Breiner, F.-T. Kao, C. Jones, T. T. Puck, D. Housman, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2829 (1980).
14. A. Y. Sakaguchi, S. L. Naylor, J. F. Gusella, D. Housman, R. Schmickel, T. B. Shows, *Am. J. Hum. Genet.* **33**, 148A (1981).
15. M. L. Law, F.-T. Kao, D. Patterson, J. N. Davidson, *J. Cell Biol.* **87**, 109a (1981).
16. M. H. Skolnick and U. Francke, in (*2*), p. 194.
17. L. A. Klobutcher and F. H. Ruddle, *Annu. Rev. Biochem.* **50**, 553 (1981); G. A. Scangos, K. M. Huttner, D. K. Juricek, F. H. Ruddle, *Mol. Cell. Biol.* **1**, 111 (1981).
18. M. Goldfarb, K. Shimizu, M. Perucho, M. Wigler, *Nature (London)* **296**, 404 (1982); C. Shih and R. A. Weinberg, *Cell* **29**, 161 (1982).
19. H. J. Evans, R. A. Buckland, M. L. Pardue, *Chromosoma* **48**, 405 (1974); V. G. Dev, R. Tantravahi, D. A. Miller, O. J. Miller, *Genetics* **86**, 389 (1977); S. Elsevier and F. H. Ruddle, *Chromosoma* **52**, 219 (1975).
20. M. E. Harper and G. F. Saunders, *Chromosoma* **83**, 431 (1981).
21. G. M. Wahl, L. Vitto, R. A. Padgett, G. R. Stark, *Mol. Cell. Biol.* **2**, 308 (1982).
22. D. W. Cox, V. D. Markovic, I. E. Teshima, *Nature (London)* **297**, 429 (1982); I. R. Kirsch, C. C. Morton, K. Nakahara, P. Leder, *Science* **216**, 301 (1982); M. J. Hobart *et al.*, *Ann. Hum. Genet.* **45**, 331 (1981); C. M. Croce *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3416 (1979).
23. S. Goff *et al.*, in *RNA Tumor Viruses* (Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y., 1981), p. 147; S. Goff, P. D'Eustachio, F. H. Ruddle, D. Baltimore, *Science* **218**, 1317 (1982).
24. D. Swan, M. Oskarsson, D. Keithley, F. H. Ruddle, P. D'Eustachio, G. F. vande Woude, *J. Virol.* **44**, 752 (1982).
25. L. J. Harris, P. D'Eustachio, F. H. Ruddle, K. Marcu, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6622 (1982); K. Calame, S. Kim, P. Lalley, R. Hill, M. Davis, L. Hood, *ibid.*, p. 6994.
26. R. Dalla-Favera, G. Franchini, S. Martinotti, F. Wong-Staal, R. C. Gallo, C. M. Croce, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4714 (1982); N. Heister-Kamp *et al.*, *Nature (London)* **299**, 747 (1982); R. Dalla-Favera, R. C. Gallo, A. Giallongo, C. M. Croce, *Science* **218**, 686 (1982); D. Forman and J. Rowley, *Nature (London)* **300**, 403 (1982); B. G. Neel, S. G. Jhanwar, R. S. K. Chaganti, W. S. Hayward, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7842 (1982); R. Dalla-Favera, M. Bregni, J. Erikson, D. Patterson, R. C. Gallo, C. M. Croce, *ibid.*, p. 7824; P. E. Barker, I. M. Verma, F. H. Ruddle, unpublished; A. Y. Sakaguchi, S. L. Naylor, T. B. Shows, J. J. Toole, M. McCoy, R. A. Weinberg, *Science* **219**, 1081 (1983).
27. B. G. VanNess *et al.*, *Nature (London)* **301**, 425 (1983); K. Marcu, personal communication.
28. B. Calabretta, D. L. Robberson, H. A. Barrera-Saldana, T. P. Lambrou, G. F. Saunders, *Nature (London)* **296**, 219 (1982); K. B. Marcu, J. Banerji, N. A. Penncavage, R. Lang, N. Arnheim, *Cell* **22**, 187 (1980).

28a. K. Lueders, A. Leder, P. Leder, E. Kuff, *Nature (London)* **295**, 426 (1982).

29. D. Botstein, R. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
30. P. D'Eustachio, O. Meyuhas, F. H. Ruddle, R. P. Perry, *Cell* **24**, 307 (1981).
31. A. Leder, D. Swan, F. H. Ruddle, P. D'Eustachio, P. Leder, *Nature (London)* **293**, 196 (1981); R. A. Popp, P. A. Lalley, J. B. Whitney,

W. F. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6362 (1981).
32. A. L. Beaudet, T.-S. Su, W. E. O'Brien, P. D'Eustachio, P. E. Barker, F. H. Ruddle, *Cell* **30**, 287 (1982); S. P. Daiger, R. S. Wilden, T.-S. Su, *Nature (London)* **298**, 682 (1982).
33. U. Francke, B. deMartinville, P. D'Eustachio, F. H. Ruddle, *Cytogenet. Cell Genet.* **33**, 267 (1982); B. Blomberg, A. Traunecker, H. N. Eisen, S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3765 (1981).
34. D. Swan, P. D'Eustachio, L. Leinwand, J. Seidman, D. Keithley, F. H. Ruddle, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2735 (1979).
35. P. D'Eustachio, D. Pravtcheva, K. Marcu, F. H. Ruddle, *J. Exp. Med.* **151**, 1545 (1980).
36. D. Owerbach, W. J. Rutter, T. B. Shows, P. Gray, D. V. Goeddel, R. M. Lawn, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3123 (1981); D. L. Slate *et al.*, *J. Exp. Med.* **155**, 1019 (1982); T. B. Shows, A. Y. Sakaguchi, S. L. Naylor, D. V. Goeddel, R. M. Lawn, *Science* **218**, 373 (1982).
37. P. Gupta, J. M. Rosen, P. D'Eustachio, F. H. Ruddle, *J. Cell Biol.* **93**, 199 (1982).
38. P. D'Eustachio, R. S. Ingram, S. M. Tilghman, F. H. Ruddle, *Somat. Cell Genet.* **7**, 289 (1981).
39. D. H. Margulies, G. A. Evans, L. Flaherty, J. G. Seidman, *Nature (London)* **295**, 168 (1982).
40. M. Steinmetz, A. Winoto, K. Minard, L. Hood, *Cell* **28**, 489 (1982).
41. A. Shimizu, N. Takahashi, Y. Yaoita, T. Honjo, *Cell* **28**, 499 (1982); A. L. M. Bothwell, M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, D. Baltimore, *ibid.* **24**, 625 (1981); M. Steinmetz, K. Minard, S. Horvath, J. McNicholas, J. Frelinger, C. Wake, E. Long, B. Mach, L. Hood, *Nature (London)* **300**, 35 (1982); R. Goodenow, M. McMillan, M. Nicholson, B. T. Sher, K. Eakle, N. Davidson, L. Hood, *ibid.*, p. 231.
42. S. P. Goff, E. Gilboa, O. Witte, D. Baltimore, *Cell* **22**, 777 (1980).
43. M. P. Calos and J. Miller, *ibid.* **20**, 579 (1980).
44. C. D. Wild, C. E. Crowther, T. P. Cripe, M. G.-S. Lee, N. Cowan, *Nature (London)* **297**, 83 (1982); N. Cowan, personal communication.
45. L. Bernstein, S. M. Mount, A. M. Weiner, *Cell* **32**, 461 (1983).
46. Two α-globin pseudogenes have been described in the mouse. One, α3, has a retrovirus-like structure (*42*). The other, α4, has retained the introns found in the functional α-globin gene and thus appears not to be cDNA-like. The one dispersed immunoglobulin pseudogene described to date, a human λ-chain pseudogene, has a cDNA-like structure. It corresponds, however, not to a λ-chain mRNA, but to an anomalous immunoglobulin RNA species found at low levels in some lymphoid cells. Its transcription in these cells is thought to be mediated by an aberrant promoter. It is difficult to exclude the notion that this promoter could likewise be active, rarely and at low levels, in germ cells [G. F. Hollis, P. A. Hieter, O. W. McBride, D. Swan, P. Leder, *Nature (London)* **296**, 321 (1982); R. P. Perry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1937 (1980)].
47. H. Czosnek, U. Nudel, M. Shani, P. Barker, F. H. Ruddle, D. Yaffe, *EMBO (Eur. Mol. Biol. Org.) J.*, in press.
48. T. D. Petes and H. L. Klein, in *The Recombination of Genetic Material*, K. B. Low, Ed. (Academic Press, New York, in press).
49. O. W. McBride, P. A. Hieter, G. F. Hollis, D. Swan, M. C. Otey, P. Leder, *J. Exp. Med.* **155**, 1480 (1982).
50. D. Owerbach *et al.*, in (*2*), p. 306; D. Owerbach, W. J. Rutter, N. E. Cooke, J. A. Martial, T. B. Shows, *Science* **212**, 815 (1981).
51. C. Junien *et al.*, *Am. J. Hum. Genet.* **34**, 381 (1982).
52. D. Owerbach, G. I. Bell, W. J. Rutter, T. B. Shows, *Nature (London)* **286**, 82 (1980); B. deMartinville, J. Leary, A. Ullrich, U. Francke, in (*2*), p. 265; G. I. Bell, M. J. Selby, W. J. Rutter, *Nature (London)* **295**, 31 (1982).
53. J. Gusella, A. Varsanyi-Breiner, F.-T. Kao, C. Jones, T. T. Puck, C. Keys, S. Orkin, D. Housman, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5239 (1979); S. Orkin *et al.*, *Nature (London)* **296**, 627 (1982).
54. R. Barg *et al.*, in (*2*), p. 252.
55. A. Y. Sakaguchi, S. L. Naylor, C. Quinto, W. J. Rutter, T. B. Shows, in (*2*), p. 313.
56. D. Owerbach, W. J. Rutter, J. A. Martial, J. D. Baxter, T. B. Shows, *Science* **209**, 289 (1980).
57. C. Huerre *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6627 (1982).
58. K. L. Bennett, P. A. Lalley, R. K. Barth, N. D. Hastie, *ibid.*, p. 1220; K. Krauter, L. Leinwand, P. D'Eustachio, F. H. Ruddle, J. E. Darnell, *J. Cell Biol.* **94**, 414 (1982).
59. R. Elliott, *Mouse Newsl. No. 64* (1981).
60. M. Yagi, P. D'Eustachio, F. H. Ruddle, M. Koshland, *J. Exp. Med.* **155**, 647 (1982).
61. S. Malcolm, P. Barton, D. L. Bentley, M. A. Ferguson-Smith, C. S. Murphy, T. H. Rabbitts, in (*2*), p. 296.
62. M. E. Chandler, L. H. Kedes, R. H. Cohn, J. J. Yunis, *Science* **205**, 908 (1979).
63. M. E. Harper, A. Ullrich, G. F. Saunders, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4458 (1981).
64. D. S. Gerhard, E. S. Kawasaki, F. C. Bancroft, P. Szabo, *ibid.*, p. 3755.
65. M. E. Harper, H. A. Barrera-Saldana, G. F. Saunders, *Am. J. Hum. Genet.* **34**, 227 (1982).
66. J. M. Traut, S. Olson, R. M. Lawn, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7809 (1982).
67. E. Herbert and M. Uhler, *Cell* **30**, 1 (1982); M. Uhler, E. Herbert, P. D'Eustachio, F. H. Ruddle, *J. Biol. Chem.*, in press.
68. Supported by NIH grant GM-09966 (F.H.R.), an ACS institutional award (P.D.), and by the Leukemia Society of America, Inc. We thank Drs. Peter Barker and Chester Partridge for their suggestions.

RESEARCH ARTICLE

Genetic Expression in the Developing Brain

Nirupa Chaudhari and William E. Hahn

Abstract. *The adult mouse brain contains complex populations of polyadenylated [poly(A)$^+$] and nonpolyadenylated [poly(A)$^-$] messenger RNA's (mRNA's). These mRNA's are separate sequence populations, similar in complexity, and in combination are equivalent to ~ 150,000 different mRNA sequences, of average length. Essentially all of the "adult" poly(A)$^+$ mRNA's are present in the brain at birth. In contrast, most of the poly(A)$^-$ mRNA's are absent. Brain poly(A)$^-$ mRNA's begin to appear soon after birth, but the full adult complement is not reached until young adulthood. This suggests that these poly(A)$^-$ mRNA's specify proteins required for the biological capabilities of the brain that emerge during the course of postnatal development.*

A major fraction of the DNA that is thought to code for protein (*1*) is expressed as messenger RNA (mRNA) in the mammalian (rodent) brain (*2, 3*). It seems likely that as many, or perhaps more, structural genes are required for the formation and function of the brain than are necessary for all other tissues and organs combined. The complexity of whole brain polysomal RNA is equivalent to about 150,000 different sequences, 1500 nucleotides in length (*2, 3*). The diversity of mRNA molecules in the brain is biologically consistent with the extensive cellular heterogeneity and microdifferentiation of the various cell types that occur in this organ. In addition, the greater complexity of brain RNA might be due to the inherent complexity of neurons. Studies on tumor cells of neural origin and nuclear RNA prepared from the neuronal fraction indicate that neurons contain RNA of higher sequence complexity than do other types of mammalian cells (*4*).

Measurements of the complexity of brain polysomal RNA have shown that nonhomologous populations of polyadenylated [poly(A)$^+$] and nonpolyadenylated [poly(A)$^-$] mRNA are present. The sequence complexity of these mRNA populations is about the same (*2, 3*). Also, in *Drosophila* a complex poly(A)$^-$ mRNA population is present, many species of which are present in the head of the fly (*5*).

Although poly(A)$^-$ mRNA's have been identified in various systems (*6*), a highly complex population of poly(A)$^-$ mRNA's appears to be characteristic of the brain since, in other mammalian cells, tissues, and organs that have been examined, essentially all of the sequence complexity is present in the poly(A)$^+$ RNA fraction of polysomal RNA (*7*). Most of the rare, complex class mRNA species in the brain are present at much less than one copy per cell (*3*), and, although little is known regarding anatomical localization of specific mRNA's, we can suppose that many of the rare copy species are restricted to given cell types, or subpopulations of cells, where they are present in relatively high copy frequency (*8*). The translational efficiency in cell free systems of brain poly(A)$^-$ mRNA's,

Nirupa Chaudhari is a postdoctoral fellow and William E. Hahn is a professor in the Department of Anatomy, University of Colorado, School of Medicine, Denver 80262.