Evolutionary origins and maintenance of redundant gene expression during metazoan development

JONATHAN COOKE (j-cooke@nimr.mrc.ac.uk)
MARTIN A. NOWAK (martin.nowak@zoo.ox.ac.uk)
MAARTEN BOERLIJST
JOHN MAYNARD-SMITH

Various levels of redundancy in developmental gene function appear common in complex metazoans. There might be no apparent phenotype at many, or even any, of a gene's specific expression sites in homozygous null mutant embryos. Here we ask what underlies the origin of such arrangements. The generation of families of genes by duplication has clearly been important. Additionally, however, selection might have driven molecularly unrelated genes, which encode proteins of similar physiological function, to become expressed during the same sets of developmental events (times and places), even though each such gene might initially have evolved in connection with just one of these events.

A recent search of the Mouse Knockout Database[1] revealed about 15 genes, to date, whose homozygous null mutant phenotypes have proved to be too subtle for detection by the laboratories that engineered them. This can occur even for genes whose products have pronounced effects when experimentally over-expressed in an embryo *in vivo* or applied to tissue culture cells. There could be up to 10^5 bona fide genes (open reading frames) in the mouse genome, and many of these might turn out to earn their keep by optimizing the phenotype in ways too subtle for laboratory detection. Yet, genes with reported absence of null mutant phenotype include those with such crucial expected functions as glial fibrillary acidic protein, tenascin, vimentin, one of the collagens, proto-oncogenes believed to function in intracellular signal transduction, and components of the retinoid receptor and binding protein array[2–14]. The natural interpretation of these observations is that two or more different gene products perform the same essential role: if one gene is knocked out, the role is performed by another, and the phenotype appears to be normal.

A related phenomenon is that observed, for instance, in mouse follistatin, goosecoid and activin null mutants[15–17]. Specific features of the early expression patterns of these genes in vertebrate embryos, and the effects of ectopic expression of their products in *Xenopus* embryos, had led to the belief that particular early steps in body pattern formation would be compromised in the null phenotypes. In the event, while there is a definite phenotype in each case (which relates to other later expression sites of each gene and is quite adequate to account for its existence), the aspects of development corresponding to any role at the earlier expression sites appear to be untouched. It seems that these particular genes are redundant as far as their early expressions are concerned, but each has a later role in which it is uniquely effective. We are aware of additional striking examples that have not yet been published, and it is with mixed relief and disbelief that we finally read of the sonic hedgehog mouse mutant, whose complex phenotype corresponds almost exactly to what had been predicted on the basis of the specific early gene expressions, and the results of experimental embryological and ectopic gene expression experiments[18].

The aim of this review is to explain the evolutionary origin of such apparently redundant genes and redundant gene expression sites. Our own thoughts are centred on vertebrates, specifically the mouse, because one of us works in a medical research institute, but also because, as we shall see, systematic gene duplication might be particularly important in the vertebrate lineage. Nevertheless, it is clear that redundancy, in all of the forms that we discuss, is relatively widespread in other types of complex metazoan development (see examples in Refs 19, 20). Several other contributions to this topic have appeared, in this journal and elsewhere (e.g. Refs 19–21), whose help in shaping our own thoughts we gratefully acknowledge.

How can redundancy be maintained?

To explain the maintenance of redundant genes over evolutionary time, it is not sufficient to draw the analogy with engineering design and say that important functions must be backed up. Nothing could be more important for survival than the Krebs cycle, yet there is relatively little redundancy among the genes specifying the necessary enzymes. As we shall see, an important clue to the origins of redundancy might come from contrasting 'developmental' genes with those for universal cellular 'housekeeping' functions. It is useful to distinguish three types of process that might explain redundancy.

Redundant genes increase fitness in subtle ways

Lewis Wolpert recently remarked that putatively phenotypeless knockout mice should be taken from their laboratory cages and away from their investigators' microscopes to a test evening at the opera. The implication is that, in nature as opposed to the laboratory, an apparently redundant gene increases fitness, not only as backup when another gene fails, but also in more subtle ways when the other gene is functioning perfectly. Thomas[19] suggests that this could be true because two genes produce more of some product than one (e.g. rRNA genes), or increase fidelity, or because together they have some emergent function. It might also be, as Wolpert's remark implies, that the fitness advantage arises only in some environments. In other words, it is unclear whether, to date, it has been demonstrated that any genes are truly redundant.

The idea that redundancy is only apparent raises no theoretical difficulties. Even a very small selective advantage, of the same order as the germ line mutation rate, would be sufficient to maintain a second gene indefinitely. However, it can be shown that there are situations in which perfect redundancy may be maintained by selection.

 0168-9525/97/$17.00
PII: S0168-9525(97)01233-X

Redundant genes and expression sites are selected as backup against mutation

Suppose that two genes have identical developmental roles. Could each be maintained by selection in individuals mutant for the other? As Fisher[22] showed a long time ago, it is not easy. As explained in Fig. 1(a), if the two genes *A* and *B* each perform the same essential role perfectly, then selection will maintain only the gene with the lower mutation rate: the other will accumulate nonfunctional mutations. If mutation rates are exactly equal, then *A* and *B* will both survive for a long time although, ultimately, one or other will be eliminated by genetic drift. We return to this possibility in the following subsection.

There are, however, situations in which both genes can be maintained indefinitely. The two simplest cases are explained in Fig. 1(b). In case (i), two genes perform the same role R1, but with different efficiencies. The less efficient gene *B*, has a lower mutation rate, and will be maintained by selection in individuals mutant at the *A* locus. Case (ii) is more complex, but potentially interesting. Gene *A* performs one role only R1, and is maintained because it performs this function more efficiently than gene *B*. Gene *B* performs two roles R1 and R2, its product being expressed at the same time and site in development as that of *A* as well as at one or more other times and sites distinctive to itself. It is maintained as a bona fide gene because of its performance of R2. Why does it continue to perform R1, which it does less efficiently than *A*? The answer is that it is maintained by selection in individuals mutant for gene *A*. For this to work, it is necessary that mutations in gene *B* that cause the loss of its secondary role only, while preserving its primary role, should be less frequent than mutations causing loss of functional gene product and thus loss of both roles. This latter condition seems probable in most cases, because DNA encoding the specific biochemical function of a protein represents a larger mutational target than the regulatory motifs determining its selective synthesis at specific sites.

These mechanisms are interesting, because they suggest how more complex systems, involving genes performing more than one role, might arise and be evolutionarily stable. Fig. 1(c) shows an example with three genes and three roles. In fact, simulated evolution of regulatory systems shows that a system in which each gene has only one role, and each role can be performed by one gene only (i.e. one gene per expression site), can rather easily evolve into one in which there is extensive pleiotropy and redundancy (see also Ref. 23).

Redundant genes arise by duplication and last a long time

Suppose that a developmental gene is duplicated. There will then be two genes performing the same role or roles, and no loss of fitness if one gene is lost. At least initially, their mutation rates are likely to be equal, or nearly so. If so, both are likely to survive for millions of generations before one is lost by the random accumulation of mutations[24,25]. But the situation is not stable on the evolutionary time scales involved in, say, the adaptive radiation of vertebrates. However, it is also possible that one or both genes will acquire one or more additional roles following the duplication, and this seems especially likely for genes whose spatially restricted expressions underlie metazoan development. In such cases, stable

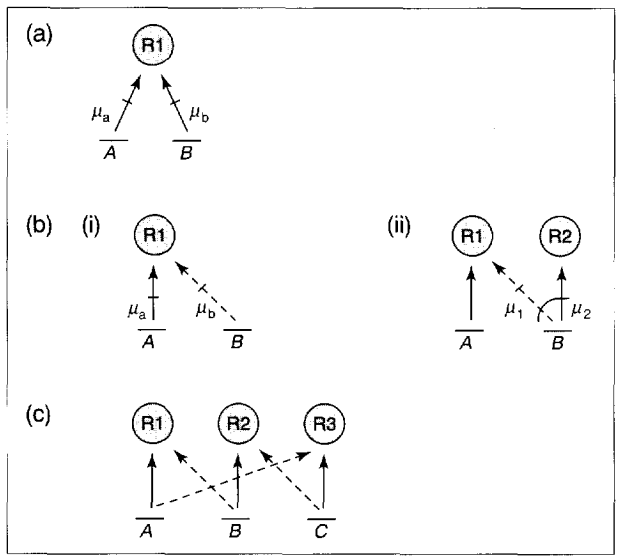

FIGURE 1. (a) Two genes, *A* and *B* perform a single role with equal efficiency. Either gene is as efficient on its own as are both acting together. μ_a are μ_b are the respective mutation rates to (recessive) nonfunctional alleles, *a* and *b*. The system is stable only if $\mu_a = \mu_b$ exactly. The reason is that at equilibrium, the rate at which nonfunctional alleles arise by mutation must exactly equal the rate at which they are eliminated by selection. Because selective elimination occurs only through double-mutant homozygotes, equal numbers of nonfunctional alleles must be lost for each gene. Therefore, for equilibrium to be possible, such alleles must arise at the same rate for each gene, a situation only likely to occur for genes of closely similar size and sequence characteristics, that is, recent duplicates. (b) Two examples of evolutionarily stable redundancy. In case (i), two genes perform the same role (R1), and *A* is more efficient than *B*. With complete recessivity, the relevant fitnesses are in the order *A/A B/B* = *A/A b/b* > *a/a B/B* >> *a/a b/b*. The system is stable only if $\mu_a > \mu_b$. In case (ii), *B* is maintained as a gene because of a second role, R2, but it also performs R1, although less efficiently than gene *A*. This redundant function is maintained by selection provided that $\mu_2 > \mu_1$, that is, mutations of *B* causing total loss of its function are more frequent than those causing loss of its performance of R1 only.
(c) A more complex system that can be evolutionarily stable.

redundancies resembling that in Fig. 1(bii) will evolve. So generally, following duplications, there will be a race between loss of redundancy by mutation, and its stabilization through acquisition of new roles.

Gene duplication and the origin of redundancy

Increasing sophistication in tracing common descent between protein structures, together with progressive storage of complete small genome sequences, is revealing reduplication and redundancy in the elaboration of the basic eukaryotic 'kit' of genes[26]. But in the vertebrate lineage in particular, duplication events involving whole genomes, or large parts of them, appear to have been a major mode of increase in complexity[27–29]. This has allowed maintenance of overlapping role sets (redundancy) because the duplicates retain shared, ancestral expression sites. Such conservatism is particularly likely because of the small targets for change by genetic drift, offered by the expression-control regions of genes, relative to those offered by the regions encoding the gene products. Crucial control elements (transcription-factor-binding sites) are short sequence motifs, acting

combinatorially or additively. Relative positions or distances along the DNA, and even polarity, of such sequences relative to the coding region are often not crucial to function. Thus, even if the now redundant role, hence expression site, of a gene is not selectively maintained, it might decay only very slowly by mutation and so be a sort of historical baggage.

Meanwhile, the acquisition of novel roles by members of such gene families can occur, by evolution of new expression sites. This is because control regions of genes can also evolve rapidly under positive selection (e.g. Ref. 30). Short sequence motifs can move readily between the neighbourhoods of different genes where, again, their precise positions of integration might not be crucial to their control function. Thus, for the reasons analysed in the previous section and Fig. 1(b), control inputs to genes and, hence, gene expression sites might be kept in place or might positively evolve, by selection, even if they are functionless in most individuals. We discuss possible examples of such coevolved, redundant gene function in the final section.

Without subscribing to naive recapitulationism – the idea that present-day embryos particularly resemble ancestral adult forms[31] – we can agree that the earliest-developing specific sites for gene expression in embryos will tend to be those associated with the evolutionarily oldest gene roles. For example, axial organization at the gastrular dorsal lip, or equivalent, occurs before subregionalization within a brain rudiment, or formation of limb buds; it also occurred in organisms whose development did not include the latter events. In view of the above scenario of gene duplications followed by partial diversification of expression, we would expect relatively more of the specific but redundant expression sites and, therefore, backed-up developmental roles among genes, to occur in earlier developmental stages. As yet, the database offers no clear test of this idea. The long-awaited screen of developmental zebrafish mutants[32], however, has produced less than the harvest of single mutant phenotypes affecting the earliest steps in vertebrate architecture, than might have been expected by direct analogy with its *Drosophila* predecessor. There is some evidence that rather than representing any transitional stage towards amniote vertebrates in its history of duplication events, the fish genome might have undergone all the events characterizing craniate vertebrates, plus some partial additional ones.

New roles keep genes alive

Any explanation of redundancy must explain the following observation. Although redundancy appears to be widespread among vertebrate and possibly other complex metazoan developmental genes, and among genes involved in the control of cell proliferation and, perhaps, fidelity of DNA replication, 'housekeeping' genes encoding the central metabolic machinery (the Krebs cycle, respiration, and so on) tend not to show survival of phylogenetically ancient, redundant duplicates. Why should this be? Duplicate copies of the genes encoding metabolic enzymes must have arisen, given the wholesale genome duplications already mentioned, but instead of redundancy, the result has been the decay of all but one duplicate. The natural explanation is that, for developmental genes, duplication offered an opportunity for the acquisition of new roles by recruitment to new specific sites of expression, coupled with partial redundancy by retention of the ancestral site. Fewer comparable opportunites have been open to metabolic genes, which are expressed throughout the body and function essentially identically everywhere, given the system properties whereby metabolism adapts to cells' varying microenvironments within the organism. Vertebrate exceptions to this generalization, that is, partly tissue-specific isotypes of enzymes of widespread function, tend to be results of more recent duplications in restricted parts of the evolutionary tree, where a version of an enzyme with special characteristics is adaptive for a particular lifestyle and/or tissue. Thus, overall, within complex metazoans, the major mechanism for retention of ancient gene duplicates would appear to have been the acquisition of novel expression sites for developmental genes, with its accompanying opportunity for new gene roles underlying the progressive extension of development itself. In much, although not all, of this duplication, the duplicate gene products retain functional interchangeability, in appropriate experiments (e.g. Ref. 33). In engineering terms therefore, the developmental complexity could as well be achieved by diversifying the set of specific expression sites in the one original gene. This process has surely been important also, but we record gene duplications and partial redundancies because they have been the means of survival of random gene duplicates in the blind evolutionary process.

Protection against 'developmental error'

There is one further reason why quasi-redundant genes, or gene expression sites, might survive or be positively selected in evolution. Suppose, as before, that two genes *A* and *B* perform a particular role with equal efficiency. If only one of the genes were present, the developmental step might occasionally fail as a result of a developmental accident – an effective failure in the normal processes of gene expression – rather than an absence of any gene function through mutation. With two separate genes, the likelihood of such a failure is greatly reduced. It turns out that a low frequency of such developmental failures is sufficient to maintain redundancy; in effect, the frequency of developmental expression failure for each gene must merely be higher than the germ line mutation rate to loss of function, or to loss of the relevant expression site in the other, which is a highly plausible assumption. But this hypothesis for redundancy phenomena must also recognize that, as mentioned above, they are frequent among genes with developmental but more exceptional among those with 'housekeeping' roles. There are at least two possible reasons. Firstly, mechanisms of spatiotemporally precise and limited gene expression in the embryo might be much more prone to non-genetic perturbation, causing failure within individuals, than are the mechanisms of universal gene expression. Secondly, correct function of most housekeeping genes (e.g. enzyme-catalysed reactions) might be more stringently specified than are some developmental functions (e.g. sequestering TGF-β ligands, see below, or adjustment of cell adhesivity/motility), so that certain molecularly unrelated genes might, nevertheless, have backup value for each other's roles in development.

Conclusions

Many processes that might maintain genetic redundancy have been discussed in this review. We can best summarize them in relation to particular cases.

A major source of gene redundancy and redundant expression has clearly been the incomplete role divergence between members of gene families arising by duplication. Many such families, such as the TGF-β-related and WNT (*Drosophila* wingless)-related genes in vertebrates[34,35] are substantial in size with complex patterns of partial role overlap. But a particularly clear example is afforded by the two *Drosophila engrailed*-related gene duplicates surviving in amniote vertebrates, *En1* and *En2* (Refs 33, 36) (the zebrafish appears to have retained, or added, a third member). These proteins share several distinctive structural domains, aside from the engrailed homeobox (DNA-binding domain), and are probably completely functionally interchangeable[33]. Both are expressed throughout a particular region of the brain rudiment that will form parts of the mid- and hind-brain, but *En1* is expressed earlier. The mouse null mutant phenotype, within this expression region, is more severe for *En1* alone. The *En2*-null mutant homozygote does, however, have a phenotype in this region, thought to be related to the fact that its expression extends into a later period. In this period there appears to be a graded requirement for *En* activity for somewhat different aspects of development. Each homozygous mutant shows a more severe phenotype if only one functional allele exists at the other locus, even though the double heterozygote is without apparent phenotype. Finally, *En1* alone exhibits several additional, quite different expression sites, some of which (in somite-derived structures) appear to be redundant, but two of which, in limb and sternum rudiments, show phenotype in the mutant[36].

A second potentially important type of redundancy is the apparent convergence of expressions of sequence-unrelated, but functionally similar genes, at particular sites. A topical example concerns vertebrate neural induction. The secreted proteins noggin, chordin, follistatin and flik (follistatin-like) are appropriately expressed in the early embryo for involvement in this intercellular signalling process, in which embryonic ectoderm is directed to form neural tissue[37–40]. At least the first three proteins can also perform this inductive role experimentally when ectopically expressed in the frog embryo[37–39]. Three unrelated sequence structures are involved, but the proteins have the common property of binding various members of the TGF-β family of small secreted proteins[41–43] to antagonize their action, while there is genetic evidence that the *Drosophila* homologue of Chordin (the short-gastrulation gene product) has this property as its sole function[44]. Why are up to four appropriate genes expressed at the site of induction when apparently one would do?

Because the process is very ancient and the proteins unrelated, it seems unlikely that we are looking at an ancestral function that has been retained redundantly and has not yet had time to decay. In fact, at least follistatin and noggin are revealed by their mutant mouse phenotypes to have necessary roles at one or more sites later in development (Ref. 15; A.P. McMahon, unpublished). Their existence as genes is, therefore, explained even if they are perfectly redundant within neural induction. But one must remember [Fig. 1(bii)] that this explanation for expressional or 'role' redundancy requires that mutations leading to total loss of function by a gene are more common than mutations leading to the loss of one (redundant) role only. A third explanation would be that the genes all have roles in initial neural induction, but subtly different ones, so that individuals with all genes functional have subtly superior neural rudiments than those with any of them inactive. To establish whether this is the case will require deeper understanding of what is going on at the molecular level, as well as more subtle anatomical and functional knowledge than is currently available. Finally, there is the possibility that two or more of the genes have essentially the same function in neural induction, and that redundancy has been selected as backup against somatic developmental error whereby one of these fails to express adequately at the appropriate site. As yet, there is insufficient evidence to enable us to evaluate the relative contributions of these evolutionary mechanisms to this example; what we have tried to do is to distinguish between them as carefully as we can, so that the relevant evidence can be recognized.

Acknowledgements

J.C. is grateful for stimulating and informative discussions with A. Furley, J. Sharpe and K. Hastings during the preparation of this article. We also thank P. Towers, J. Thomas and other, anonymous reviewers for constructive comments on drafts.

References

1 http://biomednet.com/cgi-bin/mko/mkohome.pl
2 Gomi, H. *et al.* (1995) *Neuron* 14, 29–41
3 Pekny, M. *et al.* (1995) *EMBO J.* 14, 1590–1598
4 Saga, Y. *et al.* (1992) *Genes Dev.* 6, 1821–1831
5 Steindler, D.A. *et al.* (1995) *J. Neurosci.* 15, 1971–1983
6 Colucci, G.E. *et al.* (1994) *Cell* 79, 679–694
7 Rosati, R. *et al.* (1994) *Nat. Genet.* 8, 129–135
8 Lowell, C.A., Soriano, P. and Varmus, H.F. (1994) *Genes Dev.* 8, 387–398
9 Umanoff, H. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 1709–1713
10 Stein, P.L. *et al.* (1994) *Genes Dev.* 8, 1999–2007
11 Gorry, P. *et al.* (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 9032–9036
12 Lufkin, T. *et al.* (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 7225–7229
13 Li, E. *et al.* (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 1590–1594
14 Mendelsohn, C. *et al.* (1994) *Dev. Biol.* 166, 246–258
15 Matzuk, M.M. *et al.* (1995) *Nature* 374, 360–363
16 Rivera-Perez, J.A. *et al.* (1995) *Development* 121(9), 3005–3012
17 Vassalli, A. *et al.* (1994) *Genes Dev.* 8, 414–427
18 Chiang, C.C. *et al.* (1996) *Nature* 383, 407–413
19 Thomas, J.H. (1993) *Trends Genet.* 9, 395–399
20 Tautz, D. (1992) *Bioessays* 14, 263–266
21 Brookfield, J. (1992) *Curr. Biol.* 2, 553–554
22 Fisher, R.A. (1935) *Am. Nat.* 69, 446–455
23 Nowak, M.A., Boerlijst, M., Cooke, J. and Maynard-Smith, J. (1997) *Nature* 338, 167–171
24 Kimura, M. and King, J.L. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 2858–2861
25 Takahata, N. and Maruyama, T. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 4521–4525

26 Wolfe, K.H. and Shields, D.C. (1997) *Nature* 387, 708–713
27 Lundin, L.O. (1993) *Genomics* 16, 1–19
28 Garcia-Fernandez, J. and Holland, P.W. (1994) *Nature* 370, 563–566
29 Storm, E.E. *et al.* (1994) *Nature* 368, 639–643
30 Li, X. and Noll, M. (1994) *Nature* 367, 83–87
31 De Beer, G. (1940) *Embryos and Ancestors*, Clarendon Press
32 Haffter, P. *et al.* (1996) *Development* 123, 1–36
33 Joyner, A. (1996) *Trends Genet.* 12, 15–20
34 Lyons, K.M., Jones, C.M. and Hogan, B.L.M. (1991) *Trends Genet.* 7, 408–412
35 Sidow, A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 5098–5102
36 Wurst, W., Auerbach, A.B. and Joyner, A.L. (1994) *Development* 120, 2065–2075
37 Lamb, T.M. *et al.* (1993) *Science* 262, 713–718
38 Sasai, Y. *et al.* (1995) *Nature* 376, 333–336
39 Hemmati-Brivanlou, A., Kelly, O.G. and Melton, D.A. (1994) *Cell* 77, 283–295
40 Patel, K. *et al.* (1996) *Dev. Biol.* 178, 327–342
41 Nakamura, T., Sugino, K., Titani, K. and Sugino, H. (1991) *J. Biol. Chem.* 266, 19432–19437
42 Zimmerman, L.B., De Jesus-Escobar, J.M. and Harland, R.M. (1996) *Cell* 86, 599–606
43 Piccolo, S., Sasai, B.L. and De Robertis, E.M. (1996) *Cell* 86, 589–598
44 Biehs, B., Francois, V. and Bier, E. (1996) *Genes Dev.* 10, 2922–2934

J. Cooke *is in the Division of Developmental Neurobiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK NW7 1AA.*
M.A. Nowak *and* ***M. Boerlijst*** *are in the Department of Zoology, South Parks Road, Oxford, UK OX1 3P.*
J. Maynard-Smith *is in the School of Biological Sciences, University of Sussex, Falmer, Brighton, UK BN1 9QG.*

Modular regulation of muscle gene transcription: a mechanism for muscle cell diversity

ANTHONY B. FIRULLI (firulli@utsw.swmed.edu)

ERIC N. OLSON (eolson@hamon.swmed.edu)

Skeletal, cardiac and smooth muscle cells express overlapping sets of muscle-specific genes, such that some muscle genes are expressed in only a single muscle cell type, whereas others are expressed in multiple muscle cell lineages. Recent studies in transgenic mice have revealed that, in many cases, multiple, independent cis-regulatory regions, or modules, are required to direct the complete developmental pattern of expression of individual muscle-specific genes, even within a single muscle cell type. The temporospatial specificity of these myogenic regulatory modules is established by unique combinations of transcription factors and has revealed unanticipated diversity in the regulatory programs that control muscle gene expression. This type of composite regulation of muscle gene expression appears to reflect a general strategy for the control of cell-specific gene expression.

Transcriptional regulation of tissue-specific genes is controlled by cell-type-specific promoters or enhancers that interpret unique combinations of transcription factors in different cell types. Studies of gene regulation in cultured cells have led to the identification of numerous tissue-specific promoters and enhancers. However, it is becoming increasingly clear that the regulatory DNA sequences that are important for transcription in tissue culture are often different from those that direct correct tissue specificity and temporospatial regulation *in vivo*. Recent studies of gene regulation in the skeletal, cardiac and smooth muscle cell lineages of transgenic mice have revealed that the complete developmental expression patterns of individual muscle genes are frequently dependent on composites of independent *cis*-acting regulatory regions, or modules, each of which directs a portion of the expression pattern of a gene. Thus, a regulatory module that directs transcription in a subset of muscle cells at a specific time and place in the embryo might be completely silent in other muscle cells of the same type. The strict temporospatial specificity of myogenic regulatory modules reveals surprising molecular heterogeneity among muscle cells within the same lineage, and suggests the existence of myogenic subprograms of gene expression that are established through combinatorial interactions between muscle-specific, positionally restricted and widely expressed transcription factors.

Here, we review the results of several recent analyses of muscle transgene expression and consider the implications of these studies for understanding the molecular mechanisms that generate muscle cell diversity. This type of modularity of *cis*-acting regulatory sequences provides a means of generating a multitude of patterns of gene expression from a finite number of regulatory elements and is emerging as a common theme in the developmental control of gene expression in other cell types in organisms ranging from *Drosophila* to humans[1].

Diversity of muscle cell types

Skeletal, cardiac and smooth muscle cells originate during embryogenesis from different mesodermal precursor cell populations. All skeletal muscle in vertebrates is derived from the somites, except for some muscles in the head, which appear to arise from cephalic mesoderm (reviewed in Ref. 2). The somites are transient structures that form in a rostrocaudal progression by segmentation of the paraxial mesoderm adjacent to the neural tube. Newly formed somites appear as paired epithelial spheres that become compartmentalized into a sheet of dorsal epithelial cells, known as the dermamyotome, which produces muscle precursors that migrate to the limbs and body wall. Immediately beneath the dermamyotome is the

 0168-9525/97/$17.00
PII: S0168-9525(97)01171-2