Development 128, 1857-1868 (2001)
Printed in Great Britain © The Company of Biologists Limited 2001
DEV2681

A BAC transgenic analysis of the *Mrf4/Myf5* locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development

Jaime J. Carvajal*, David Cox*, Dennis Summerbell* and Peter W. J. Rigby*,‡

Division of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

*Present address: Section of Gene Function and Regulation, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

‡Author for correspondence (e-mail: p.rigby@icr.ac.uk)

Accepted 26 February; published on WWW 19 April 2001

SUMMARY

The muscle-specific transcription factors Myf5 and Mrf4 are two of the four myogenic regulatory factors involved in the transcriptional cascade responsible for skeletal myogenesis in the vertebrate embryo. *Myf5* is the first of these four genes to be expressed in the mouse. We have previously described discrete enhancers that drive *Myf5* expression in epaxial and hypaxial somites, branchial arches and central nervous system, and argued that additional elements are required for proper expression (Summerbell, D., Ashby, P. R., Coutelle, O., Cox, D., Yee, S. P. and Rigby, P. W. J. (2000) *Development* 127, 3745-3757). We have now investigated the transcriptional regulation of both *Myf5* and *Mrf4* using bacterial artificial chromosome transgenesis. We show that a clone containing *Myf5* and 140 kb of upstream sequences is sufficient to recapitulate the known expression patterns of both genes. Our results confirm and reinforce the conclusion of our earlier studies, that *Myf5* expression is regulated differently in each of a considerable number of populations of muscle progenitors, and they begin to illuminate the evolutionary origins of this complex regulation. We further show that separate elements are involved in the activation and maintenance of expression in the various precursor populations, reflecting the diversity of the signals that control myogenesis. *Mrf4* expression requires at least four elements, one of which may be shared with *Myf5*, providing a possible explanation for the linkage of these genes throughout vertebrate phylogeny. Further complexity is revealed by the demonstration that elements which control *Mrf4* and *Myf5* are embedded in an unrelated neighbouring gene.

Key words: Myf5, Mrf4, MRF, Myogenesis, Muscle, Branchial Arch, Epaxial, Hypaxial, Hyoid, Mandibular, Craniofacial, BAC, Transcriptional Regulation, Mouse

INTRODUCTION

The key events in the development of the first skeletal muscles of the vertebrate embryo are initiated and coordinated by the myogenic regulatory factors (MRFs), Myf5 (Braun et al., 1989), myogenin (Edmondson and Olson, 1989; Wright et al., 1989), MyoD (Davis et al., 1987) and Mrf4 (Rhodes and Konieczny, 1989; Braun et al., 1990; Miner and Wold, 1990), all of which are members of the basic helix-loop-helix super-family of transcription factors. *Myf5* is the first MRF to be expressed in birds (Hacker and Guthrie, 1998; Hirsinger et al., 2001) and mammals (reviewed by Buckingham, 1992). In mouse embryos, *Myf5* expression starts in the trunk at 8.0 days post coitum (dpc), followed by *myogenin* (*Myog* – Mouse Genome Informatics; 8.5 dpc), *Mrf4* (*Myf6* – Mouse Genome Informatics; 9.0 dpc) and *MyoD* (*Myod1* – Mouse Genome Informatics; 10.5 dpc). The first phase of *Mrf4* expression downregulates by 12.0 dpc, followed by a general upregulation in a second foetal phase (Bober et al., 1991; Hinterberger et al., 1991). In the developing limbs, *Myf5* is again first (10.5 dpc), followed by the coincident expression of *myogenin* and *MyoD*, and later of *Mrf4*. In the branchial arches, from which some of the facial musculature originates, *Myf5* is also expressed first (9.5 dpc), followed a day later by *MyoD* and *myogenin*; however, *Mrf4* is only expressed in craniofacial muscles during the later foetal phase. Analyses of mice mutant for the MRFs, both singly and in combination (reviewed by Arnold and Braun, 2000), are consistent with the notion that *Myf5* is the determination gene for skeletal muscle. Knowledge of how *Myf5* expression is induced and maintained is therefore central to understanding the mechanisms of skeletal myogenesis. The *Myf5* and *Mrf4* genes are linked in all vertebrates analysed (Braun et al., 1990; Patapoutian et al., 1993; Saitoh et al., 1993; O. Coutelle, C. Moreno de Barreda and P. W. J. R., unpublished), raising the question of whether these genes are coordinately controlled.

In vertebrate embryos, the somites are the source of the skeletal muscles of the trunk and limb. As somites mature, they

divide into a mesenchymal sclerotome, from which the axial skeleton originates, and an epithelial dermomyotome (reviewed by Christ and Ordahl, 1995). Cells of the dorsomedial lip of the dermomyotome migrate by involuting immediately under the dermomyotome and give rise to the dorsal myotome, which will form the epaxial muscles. In thoracic somites, the ventrolateral lip of the dermomyotome curves to form the somitic bud, which grows ventrally into the lateral body wall, giving rise to the thoracic hypaxial muscles. Muscle precursor cells detach from the medial edge of the somitic bud to form the ventral myotome, while at limb-bud levels, cells delaminate from the ventral dermomyotome and migrate as mesenchymal cells into the limb. In occipital somites, ventral dermomyotomal cells move into the hypoglossal cord and migrate rostrally to form pharyngeal and tongue muscles (Mackenzie et al., 1998). In the head, some muscles form in situ from prechordal and paraxial mesoderm (Couly et al., 1992), while other paraxial mesoderm-derived muscle precursor cells first migrate into the branchial arches from where they start a second migration to their final location in the head (Noden, 1983).

Considerable attention has been focused upon the analysis of regulatory elements controlling the MRF genes, as identification of the cognate transcription factors will be invaluable in elucidating the signals and mechanisms initiating myogenesis. We have previously shown that *Myf5* is regulated through at least four discrete enhancers dispersed throughout 14.2 kb of the *Mrf4*/*Myf5* locus (Summerbell et al., 2000). These enhancers individually control the initiation of expression in the epaxial and hypaxial domains of the somites, a subset of the facial muscles and in the central nervous system. However, they are unable, individually or together, to direct expression in some hypaxial (limb, ventral myotome and hypoglossal cord) and facial muscle populations, or to maintain expression correctly, and they cause inappropriate expression in the dermomyotome. Expression analyses of an *lacZ-Myf5* reporter gene in chimaeric mouse embryos, generated by transferring yeast artificial chromosomes (YACs) into embryonic stem cells, had shown that the sequences required for limb expression are situated far upstream of the gene (Zweigerdt et al., 1997). Hadchouel et al. have mapped this region to the –58 to –48 kb interval and shown that at least one other upstream region is required for *Myf5* expression (Hadchouel et al., 2000). The regulation of *Myf5* is thus markedly different from that of *myogenin* or *MyoD*, which are controlled by fewer elements that do not show the same specificity for individual precursor cell populations (Cheng et al., 1993; Yee and Rigby, 1993; Kablar et al., 1997; Kucharczuk et al., 1999). We have previously argued that the way in which *Myf5* is regulated is particularly suited to its role as the skeletal myogenic determination gene (Summerbell et al., 2000).

The regulation of *Mrf4* expression has been less intensively studied. Transgenic analyses have shown that a 6.5 kb sequence immediately upstream of the mouse *Mrf4* gene recapitulates the foetal phase of expression (Patapoutian et al., 1993), while 8.5 kb immediately upstream of rat *Mrf4* drives both a subset of the early somitic expression and the foetal phase (Pin et al., 1997).

In order to identify and characterise all of the elements that regulate the linked *Mrf4* and *Myf5* genes, we have isolated a series of overlapping bacterial artificial chromosomes (BACs) that constitute a de facto 5′ deletion series of the region. Using our modification of the protocol for homologous recombination based manipulation of BAC clones (Yang et al., 1997), alkaline phosphatase (AP) and *lacZ* reporter genes have been introduced into the *Mrf4* and *Myf5* genes, respectively. We show that transgenic animals carrying constructs containing 140 kb upstream of *Myf5* reproduce all known aspects of the temporal and spatial expression patterns of both genes. Using the 5′ BAC deletion series, we locate additional elements absent from our earlier plasmid-based constructs, which drive *Myf5* expression in individual branchial arches and different hypaxial muscle progenitor populations, and also sequences required for maintenance of expression. Furthermore, we show that the regulation of *Mrf4* expression is also complex, with at least four separate elements required to recapitulate the endogenous pattern. Our data are consistent with the hypothesis that at least one element operates on both genes.

MATERIALS AND METHODS

Isolation of BAC clones containing the *Mrf4*/*Myf5* locus

Twelve bacterial artificial chromosome (BAC) clones were isolated from a 129SV mouse library (Research Genetics, USA) by screening filters simultaneously with two genomic probes derived from the third exons of *Mrf4* (458 bp) and *Myf5* (435 bp). Six clones were confirmed as true positives by PCR amplification using the following primers: 5′-CAGCACAGCATAGCACAGGAG-3′ (Mrf4F), 5′-CTTTCACTT-GAGGTGGTGAGA-3′ (Mrf4R), 5′-CTGCAAAGTTTTACATCAG-3′ (Myf5F) and 5′-ACCGAAAAGCACGTATTCTGC-3′ (Myf5R).

Construction of shuttle vectors for homologous recombination

BAC clones were modified by an improved version (J. J. C., D. C. and P. W. J. R., unpublished) of another protocol (Yang et al., 1997). For the modification of *Myf5*, we used reporter construct #3 of Summerbell et al. (Summerbell et al., 2000), containing an *nlacZ* (nuclear localization signal and *lacZ*) reporter gene followed by an SV40 polyadenylation sequence cloned as a *Bam*HI cassette into the mutated translational start codon of *Myf5*. To construct the shuttle vector (Yang et al., 1997), an *Nhe*I/*Sma*I fragment containing 1828 bp and 1640 bp of homology 5′ and 3′, respectively, to the translational start codon of *Myf5*, and including the reporter cassette, was *Xho*I linkered and cloned into the *Sal*I site of the pSV1-RecA shuttle vector. The *AP-Mrf4* shuttle vector contains 1771 bp (*Hpa*I/*Nla*III) and 1546 bp (*Nla*III/*Xba*I) of homology 5′ and 3′, respectively, to the translational start codon of the *Mrf4* gene. The gene for human placental alkaline phosphatase containing a membrane localisation signal (gift from C. L. Cepko, Harvard Medical School, Boston, Massachusetts, USA) followed by a PGK polyadenylation sequence was cloned in frame into the translational start codon of the *Mrf4* gene. Full information on the cloning steps can be obtained on request.

Modification of BAC clones by homologous recombination

Chemically competent cells containing the appropriate BAC clone were transformed with 10-100 ng of the appropriate shuttle vector, plated and incubated at 30°C for 48 hours on LB-agar CAM/TET plates (12.5 μg/ml chloramphenicol; 10 μg/ml tetracycline). Six colonies were isolated and dispersed in 10 ml of LB and then 100 μl of each were plated onto LB-agar CAM/TET plates and incubated at 43°C overnight. Single colonies were transferred to 3 ml of LB (CAM/TET) and grown overnight at 43°C. Co-integrates were identified by restriction mapping and hybridisation. One co-integrate

colony was selected, streaked onto an LB-agar CAM-only plate and grown overnight at 43°C. Four colonies were selected, dispersed into 1 ml of LB and diluted 1/100 in LB. 100 μl of these dilutions were plated onto CAM/fusaric-acid plates (see Yang et al., 1997) and grown at 37°C for 3 days. From these plates, 24 colonies were analysed by digestion with *Eco*RI (for *lacZ*) or *Bam*HI (for AP) to identify fragments arising from the modification of the locus. Clones giving the expected restriction pattern were further analysed by digestion with a panel of enzymes (*Sal*I, *Hin*dIII, *Eco*RI, *Bam*HI), followed by Southern blotting and hybridisation with probes corresponding to the homology arms used in the appropriate shuttle vectors.

Generation of transgenic mice

BAC DNA was prepared using the endonuclease-free QIAgen maxiprep kit (QIAGEN Ltd., UK) following manufacturer's instructions with the following modifications: the volume of solutions P1 (resuspension), P2 (lysis) and P3 (neutralisation) was increased fivefold. DNA was eluted from the QIAgen column at 55°C using 15 ml of pre-heated elution buffer. After dialysis against microinjection buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA pH 8.0 and 100 mM NaCl), circular DNA was diluted to 1.5 ng/μl in microinjection buffer plus polyamines (final concentrations 30 μM spermine, 70 μM spermidine; Montoliu et al., 1995) and used to inject fertilised mouse eggs from CBA/Ca×C57Bl/10 crosses, as previously described (Yee and Rigby, 1993).

Histochemical staining

β-Galactosidase staining

Embryos were fixed overnight in Mirsky's fixative (National Diagnostics) at 4°C, washed three times in PBSA (Ca^{2+}, Mg^{2+}-free phosphate buffered saline)/0.02% Nonidet P-40 for 20 minutes at room temperature and placed in 10 ml of β-galactosidase staining solution (5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6 \cdot 3H_2O$, 2 mM $MgCl_2$, 0.02% Nonidet P-40, 0.4 mg/ml X-Gal in PBSA) for 2-20 hours (depending on the stage of the embryo) at room temperature, and post-fixed in Mirsky's fixative. In some instances, X-Gal was replaced by Rose-Gal, Magenta-Gal or Bluo-Gal (Diagnostic Chemicals Ltd., Canada).

AP staining

Embryos were fixed in 4% paraformaldehyde in PBSA overnight at 4°C, rinsed twice in PBSA +2 mM $MgCl_2$, and washed in PBSA +2 mM $MgCl_2$ for 10 minutes. Endogenous phosphatases were inactivated by incubation for 1 hour in PBSA +2 mM $MgCl_2$ at 65°C. Embryos were equilibrated in AP-buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 0.1% Tween 20) supplemented with 2 mM levamisol for 45 minutes at room temperature. Embryos were washed once in ice-cold AP-buffer, transferred to AP-staining buffer (AP-buffer plus 0.1 mg/ml of BCIP; 0.5 mg/ml NBT) and stained for 3-10 hours, depending on the stage of the embryo, in the dark at 4°C. The reaction was stopped by washing the embryos in ice-cold PBSA containing 2 mM EDTA in the dark at 4°C for 16-24 hours. Embryos were post-fixed in Mirsky's fixative adjusted to pH 5.0, in the dark at 4°C for 16 hours.

RESULTS

We have previously shown that *Myf5* is regulated by a number of distinct and discrete enhancers, dispersed throughout 14.2 kb of the *Mrf4*/*Myf5* locus. These enhancers separately initiate expression in the epaxial muscle precursors, some hypaxial precursors, some facial muscles and the central nervous system, and cause incorrect expression in the dermomyotome. Additional elements must be necessary to drive expression in the limbs, ventral myotome, hypoglossal cord and the remaining facial musculature, and to correctly maintain expression (Summerbell et al., 2000). Individual elements controlling *Mrf4* transcription have not been defined, and only a subset of the expression pattern has been reproduced using transgenic analyses (Patapoutian et al., 1993; Pin et al., 1997), again indicating that additional sequences are required.

In order to identify the additional regulatory elements required for proper expression of both genes, six overlapping BAC clones containing the locus were isolated from a murine library. A physical map of the BAC contig (Fig. 1) was generated using standard techniques. Clones were named according to the sequence length (in kb) upstream of the transcriptional start site of the *Myf5* gene: BAC59, BAC61, BAC81, BAC88, BAC140 and BAC195. Using the same convention we will henceforth refer to construct #1 of Summerbell et al. (Summerbell et al., 2000), containing 8.8 kb of upstream sequences, as p8.8Z. Sequences 3′ to the locus range from 39.6 to ~50 kb downstream of the *Myf5* gene (5.4 kb for p8.8Z). Release of a C57BL6/J mouse BAC clone sequence (GenBank Accession Number AC021642) overlapping our contig was used to refine the location of most of the BAC end-points to base-pair level. Analysis of this sequence revealed the presence of the gene encoding the protein tyrosine phosphatase RQ (*Ptprq* – Mouse Genome Informatics; PTP-RQ), transcribed in the same orientation as *Mrf4* and *Myf5*. We identified and positioned 25 exons of the gene representing 78.3% of the rat cDNA (GenBank Accession Number AF063249), while 5′ exons and upstream sequences were not found within this BAC; a consensus polyadenylation signal was located 31.1 kb upstream of *Myf5*. Other possible coding regions were also identified (Fig. 1). Of particular interest was EST AA060545 (GenBank) which contains two exons with 100% homology to the genomic sequence delimiting a putative intron flanked by consensus splice sites, suggesting that this EST represents a true cDNA clone. If so, the corresponding gene would overlap the gene for PTP-RQ and be transcribed in the opposite orientation.

Generation of transgenic mice using modified BAC clones

We modified the protocol for manipulating BACs by homologous recombination in *E. coli* (Yang et al., 1997) in order to increase the efficiency (J. J. C., D. C. and P. W. J. R., unpublished) and used it to introduce an *nlacZ* reporter cassette into *Myf5* (BAC-Z constructs) or both alkaline phosphatase (AP) and *nlacZ* into *Mrf4* and *Myf5*, respectively (BAC-APZ constructs). These constructs were injected into fertilised mouse eggs to generate transient embryos or transgenic lines. We observed no significant differences in the expression pattern of the *nlacZ-Myf5* transgene between BAC-APZ and BAC-Z lines from equivalent BAC clones, indicating that *AP-Mrf4* does not interfere with the expression of *nlacZ-Myf5*. Copy number was assessed in five BAC59Z (3-10 copies) and two BAC195Z (5-9 copies) lines and no correlation with staining intensity was observed. A summary of the main features of all the BAC transgenic lines is given in Table 1.

Identification of new *Myf5* regulatory elements

Transgene expression in both BAC195Z and p8.8Z lines (Fig. 2A,B) started before 8.5 dpc in the dorsal dermomyotome of

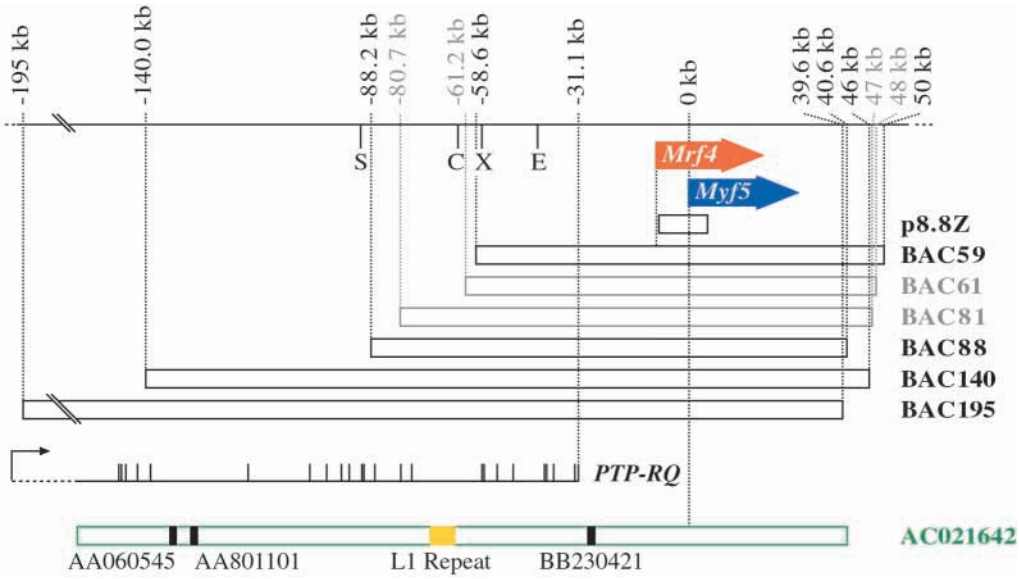

Fig. 1. Physical map of the *Mrf4*/*Myf5* BAC contig. Distances upstream and downstream from the transcriptional start site of *Myf5* are indicated. The position and orientation of the *Mrf4* and *Myf5* genes are indicated as red and blue arrows, respectively. The p8.8Z plasmid containing the proximal control elements of *Myf5* corresponds to construct #1 (Summerbell et al., 2000). Grey indicates BACs not used in this study. The sequenced BAC AC021642 is represented in green and shows the position of identified ESTs (black boxes) and a long L1 repeat (yellow). Mapped exons from the gene for protein tyrosine phosphatase RQ (PTP-RQ) are also represented. C, *Cla*I; E, *Eag*I; S, *Sal*I; X, *Xho*I.

Table 1. Summary of the main features for all the BAC transgenic lines

	nlacZ-Myf5					*AP-Mrf4*			
Line	Ventral tail somites	Ventralmost thoracic somites	Mandibular arch	Hyoid arch	Maintenance	Ventral tail somites	Ventralmost thoracic somites	Copy number	Other
BAC59Z.1	–	–	–	+	–	na	na	3	–
BAC59Z.2	Faint	–	Faint[1]	+	–	na	na	7	–
BAC59Z.3	–	–	–	+	–	na	na	4	–
BAC59Z.4	–	–	–	+	–	na	na	10	–
BAC59Z.5	–	–	–	+	–	na	na	5	a
BAC59Z.6	–	–	–	+	–	na	na	–	b
BAC88Z.1	–	–	+	+	+	na	na	–	–
BAC195Z.1	+	+	+	+	+	na	na	5	–
BAC195Z.2	+	+	+	+	+	na	na	9	–
BAC195Z.3	+	+	+	+	+	na	na	–	–
BAC195Z.4	–	–	+	+	–	na	na	–	b
BAC195Z.5	–	–	Faint	+	–	na	na	–	b
BAC195Z.6	+	+	+	+	+	na	na	–	–
BAC59APZ.1	–	–	–	+	–	–	–	–	–
BAC59APZ.2	–	–	–	+	–	–	–	–	–
BAC59APZ.3	–	–	–	+	–	–	–	–	b
BAC88APZ.1	–	–	+	+	+	–	–	–	a
BAC88APZ.2	–	–	+	+	+	ND	–	–	–
BAC88APZ.3	+	–	+	+	+	–	ND	–	a, c
BAC140APZ.1	+	+	+	+	ND	ND	ND	–	–
BAC140APZ.2	+	+	+	+	+	+	+	–	d[1]
BAC140APZ.3	+	+	+	+	+	+	+	–	–
BAC140APZ.4	Faint	Faint	Faint	Faint	–	+	+	–	b
BAC195APZ.1	–	–	–	–	–	ND	ND	–	b
BAC195APZ.2	+	Faint	Faint	Faint	ND	ND	ND	–	b
BAC195APZ.3	+	+	+	+	+	+	+	–	–
BAC195APZ.4	+	+	+	+	+	+	+	–	d[1]

Epaxial somitic and limb expression were seen in all lines.
[1]Variable within littermates; a, ectopic dorsal root ganglia; b, low expression levels; c, nasal placodes; d, ectopic apical limb bud; na, not applicable; ND, not done.

rostral somites, which is consistent with the epaxial enhancer defined by Summerbell et al. (Summerbell et al., 2000) correctly activating *Myf5* expression. From 9.5 dpc, differences between BAC195Z and p8.8Z became apparent. BAC195Z lines expressed the transgene in the dorsal lip and the myotome in all somites. In anterior thoracic somites we also saw expression in the most ventral region of the dermomyotome, corresponding to the early somitic bud (Fig. 2C, arrow). In

contrast, p8.8Z lines showed incorrect expression displaced caudally in the dermomyotome (Fig. 2D, arrow). BAC195Z lines expressed correctly in mandibular and hyoid arches (Fig. 2C, arrowheads), whereas p8.8Z activated the transgene only in the hyoid (Fig. 2D, arrowhead).

At 10.5 dpc BAC195Z lines, unlike p8.8Z, expressed the transgene in the forelimb (arrow), hypoglossal cord (arrowheads) and the ventral half of the dermomyotome of cervical and thoracic somites, corresponding to the somitic bud (compare Fig. 2E with 2F). At 11.5 dpc BAC195Z lines expressed the transgene in the presumptive intercostal muscles including the somitic bud (Fig. 2G, arrow) and derivatives of the mandibular and hyoid arches in the head. In contrast, p8.8Z lines (Fig. 2H) failed to maintain expression in the intercostal region and derivatives of the hyoid arch. At 12.5 dpc BAC195Z lines (Fig. 2I) expressed strongly in individual thoracic muscles (arrow), extraocular muscles (black arrowhead), mandibular and hyoid arch derivatives, and dorsal and ventral tail somites (red arrowhead), while p8.8Z lines (Fig. 2J) maintained expression only in dorsal and central myotome. By 13.5 dpc transgene expression was maintained strongly in all muscle masses of BAC195Z lines (Fig. 2K) and activated in the first row of snout muscles (Fig. 2K, arrow). In p8.8Z lines, the transgene was expressed very faintly in the dorsal domain of only a few tail somites (Fig. 2L, arrow). At 14.5 dpc BAC195Z lines expressed strongly in both body and head muscles, including all the snout musculature (Fig. 2M), while there was no expression in p8.8Z lines (data not shown). The last group of muscles to activate *Myf5* expression were those of the eyelid (by 15.5 dpc, data not shown).

In summary, BAC195Z drives and maintains transgene expression in dermomyotome and myotome, hypoglossal cord, branchial arches, and all epaxial, hypaxial, limb and head muscles. It contains all the elements identified as absent from p8.8Z (Summerbell et al., 2000) and recapitulates all known aspects of the expression pattern of *Myf5* (Tajbakhsh et al., 1996).

Analysis of *Myf5* expression in the BAC deletion series

In order to delimit the additional elements identified in BAC195Z, we generated transgenic lines using BAC140-, BAC88- and BAC59-based constructs. Deletion from BAC195APZ (or BAC195Z, Fig. 2) to BAC140APZ did not change reporter gene expression between 8.5 and 14.5 dpc (e.g. Fig. 3A,B; 10.0 dpc). We therefore have no evidence for control elements lying within this 55 kb interval.

Deletion from BAC140APZ to BAC88Z positioned an element involved in expression in the most ventral domain of the somites. This was best seen in the thoracic somites during the 12 hour window between 28 and 36 somites (Fig. 3C; dotted line between the arrows, compare with Fig. 3A,B), and in the tail at later stages (see below). This positions a new element in the −140 to −88 kb interval, setting the boundary for regulatory elements 140 kb upstream of the *Myf5* transcriptional start site, within the gene for PTP-RQ.

Comparison between BAC88Z and BAC59APZ (or

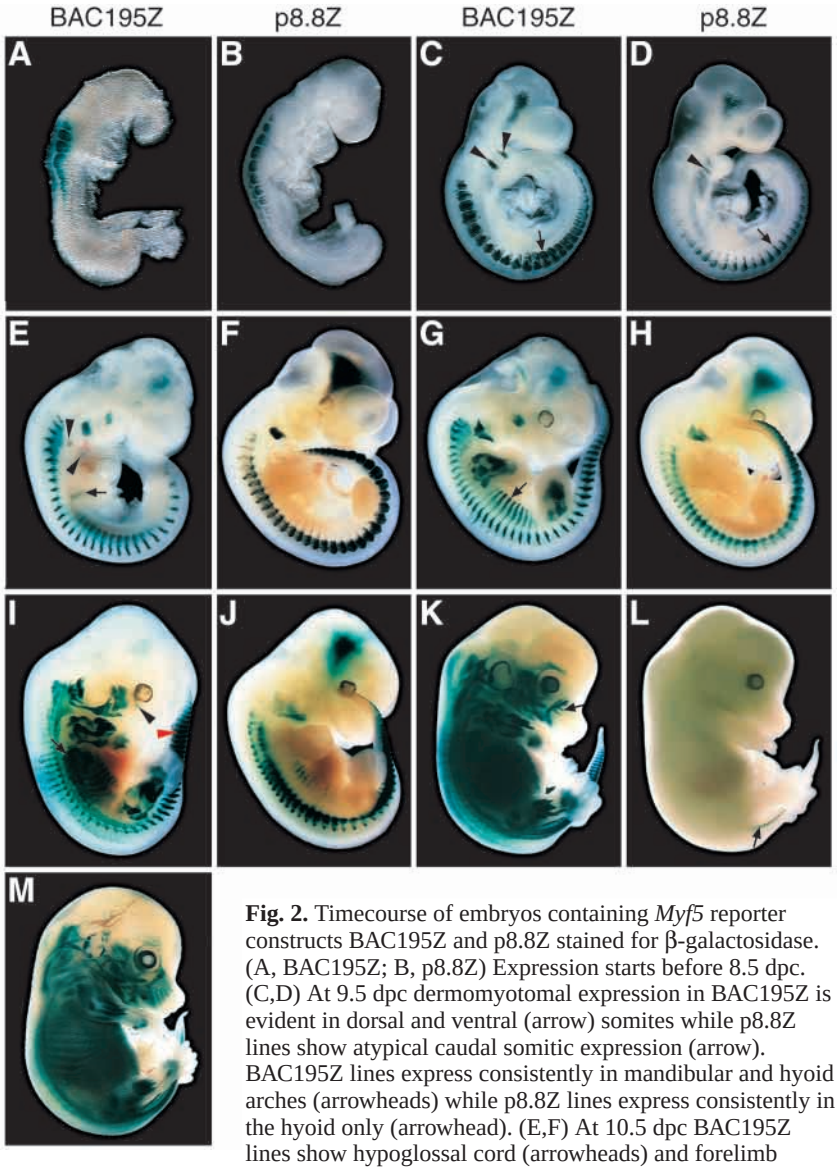

Fig. 2. Timecourse of embryos containing *Myf5* reporter constructs BAC195Z and p8.8Z stained for β-galactosidase. (A, BAC195Z; B, p8.8Z) Expression starts before 8.5 dpc. (C,D) At 9.5 dpc dermomyotomal expression in BAC195Z is evident in dorsal and ventral (arrow) somites while p8.8Z lines show atypical caudal somitic expression (arrow). BAC195Z lines express consistently in mandibular and hyoid arches (arrowheads) while p8.8Z lines express consistently in the hyoid only (arrowhead). (E,F) At 10.5 dpc BAC195Z lines show hypoglossal cord (arrowheads) and forelimb expression (arrow), absent in p8.8Z lines. (G,H) At 11.5 dpc expression in BAC195Z is stronger than in p8.8Z lines, extending ventrally into the somitic bud (arrow). Expression in hyoid arch derivatives starts to downregulate in p8.8Z lines. (I,J) At 12.5 dpc, thoracic (arrow) and extraocular muscles (black arrowhead) are visible in BAC195Z lines. Expression in the tail occupies the entire length of the somites (red arrowhead). Expression is only maintained in dorsal and central myotome in p8.8Z lines, with tail expression restricted to the dorsal domain. (K,L) By 13.5 dpc, transgene expression in BAC195Z lines expands in thoracic and facial domains, including some snout muscles (arrow). Downregulation in p8.8Z lines is complete except for residual tail expression (arrow). At 14.5 dpc (M) and later (not shown) expression is maintained in BAC195Z lines and all snout musculature is now expressing the transgene.

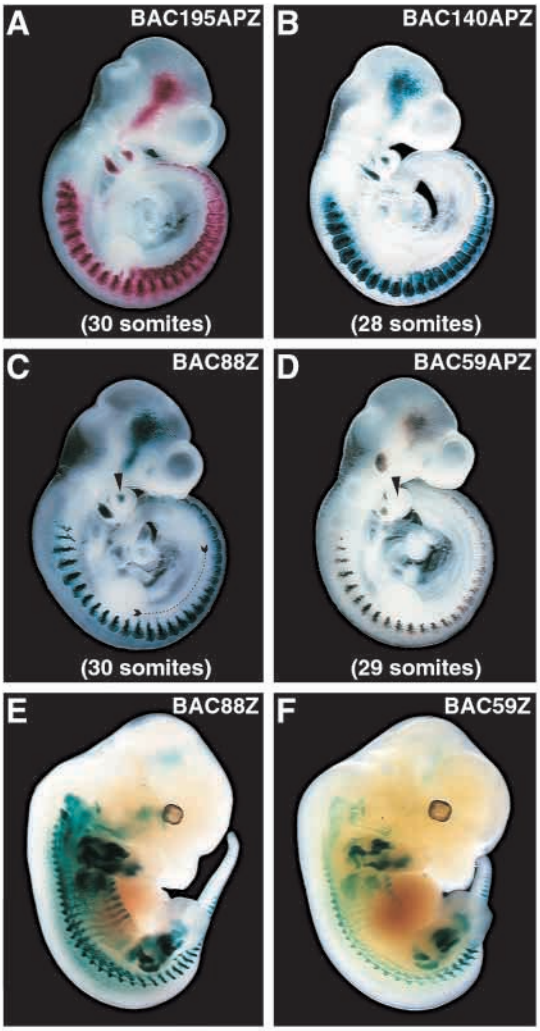

Fig. 3. Expression patterns in the BAC deletion series. (A,B) No differences in expression pattern can be distinguished between BAC195 (A) and BAC140 (B) lines. (C,D) Deletion to BAC88 (C) abolishes expression in the ventral thoracic somites (dotted line marks the ventral edge of the somitic bud at 10.0 dpc), while deletion to BAC59 (D) abolishes expression in the mandibular arch (arrowheads) and the hypoglossal cord. (E) At 12.5 dpc, transgene expression is missing from the ventral half of the tail somites in BAC88Z lines. (F) From 12.5 dpc, BAC59Z lines fail to maintain transgene expression in axial musculature.

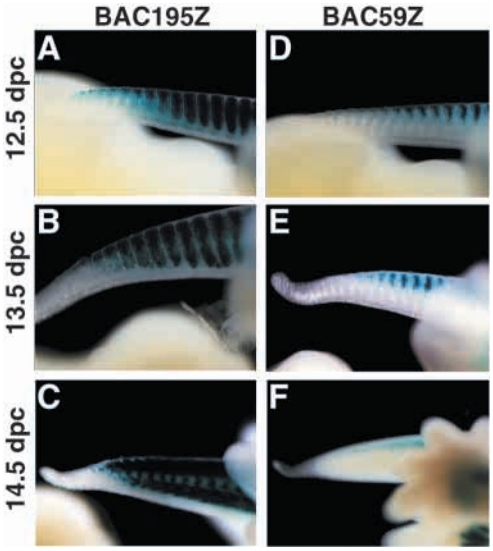

Fig. 4. Details of expression domains in the tail. BAC195Z lines (A-C) show dorsal expression in all somites which extends to the ventral edge as they mature, while in BAC59Z lines (D-F) initiation of expression is delayed in the dorsal domain, does not extend across the midline and is not maintained.

BAC59Z) lines positioned the element(s) involved in mandibular arch and hypoglossal cord expression, and in late maintenance. While hyoid arch expression was consistent in lines from both BACs, BAC59Z lines showed little or no expression in the mandibular arch at any stage (compare Fig. 3C with 3D; arrowheads) and hypoglossal cord expression was never seen. At 12.5 dpc (Fig. 3E), BAC88Z maintained expression, whereas BAC59 lines downregulated reporter expression in all muscles, except those of the limb (Fig. 3F). This indicates that sequences required for activation in the mandibular arch and the hypoglossal cord, and for maintenance of expression in the axial musculature are located in the −88 to −59 kb interval, again within the gene for PTP-RQ.

Deletion from BAC59APZ to p8.8Z positioned elements involved in dermomyotome, hyoid arch and limb expression. At 9.5 dpc, p8.8Z lines (Fig. 2D) expressed the transgene incorrectly in the caudal dermomyotome, whereas BAC59 lines did not (compare Fig. 2D (arrow) with Fig. 3D). Early branchial arch expression in BAC59Z lines was very similar to that in p8.8Z. However, while the latter downregulated arch expression from 11.5 dpc (Fig. 2H), in BAC59Z lines it was maintained through 13.5 dpc and then downregulated (data not shown). In p8.8Z lines, transgene expression was never activated in the limbs (Fig. 2H,J,L). These data indicate that sequences in the −59 to −8.8 kb interval correct the inappropriate dermomyotomal expression and drive expression in the limbs at appropriate times and in the hyoid arch at late stages.

The expression of *Myf5* in tail musculature

Comparison of BAC lines that contain the element regulating expression in the most ventral somitic domain (BAC195 and BAC140) and those lacking this region (BAC88 and BAC59) shows that *Myf5* expression is regulated differentially in ventral and dorsal tail somites. In BAC195 and BAC140 lines, the transgene was expressed in all tail somites at 12.5 dpc and 13.5 dpc (Fig. 4A,B). Somites I-VI expressed *nlacZ-Myf5* in the dorsal half, whereas older somites expressed it in both dorsal and ventral halves. By 14.5 dpc (Fig. 4C), when the first myofibres appeared in the tail, the transgene was strongly expressed in a characteristic pattern, revealing dorsal and ventral muscles with a central muscle population uniting the two components. Expression in the tail was fully maintained and could be detected in neonatal animals (data not shown). In contrast, in BAC88 and BAC59 lines at 12.5 and 13.5 dpc, the transgene was first expressed in somite VI or VII (Fig. 4D,E); expression was confined to the dorsal region and disappeared as myofibres arose in the tail (Fig. 4F).

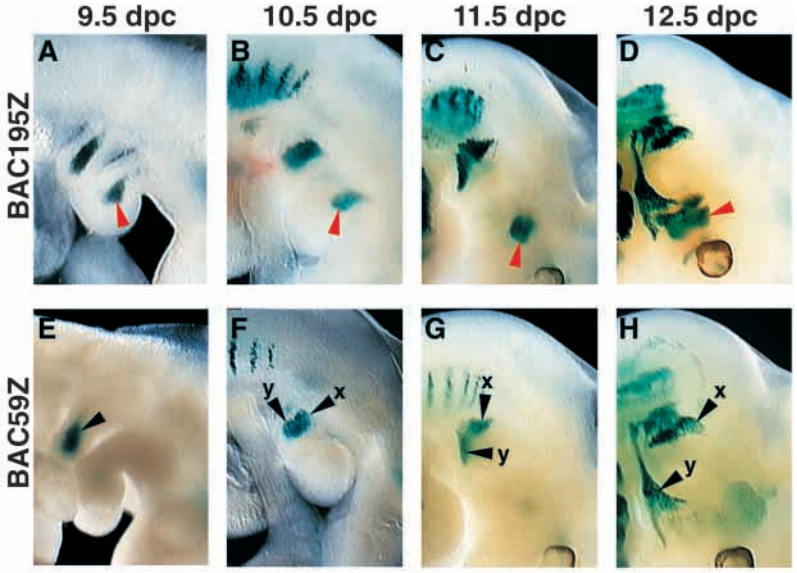

Fig. 5. Differences in branchial arch expression patterns. BAC195Z expresses in mandibular and hyoid arches and subsequently in all facial muscles (A-D) while BAC59Z expresses only in the hyoid arch and in the facial expression muscles at later stages (E-H). Comparisons at different stages allow the mapping of derivatives from mandibular (red arrowheads) and hyoid (black arrowheads) arches. (A) At 9.5 dpc expression is first seen in the central core of the mandibular arch. (B,C) By 10.5 dpc (B) this cell population starts to migrate into the facial region and by 11.5 dpc (C) arrests midway between the eye and the prospective ear. (D) From 12.5 dpc onwards this muscle mass divides and expands to give rise to the muscles of mastication. (E,F) At 9.5 dpc (E) the mesodermal core in the hyoid arch is already expressing the transgene and by 10.5 dpc (F) starts dividing into two domains (x and y) which begin migrating into the head region. (G) By 11.5 dpc the muscle mass separates into dorsal and ventral branches. (H) From 12.5 dpc the dorsal branch populates the region around the prospective ear, forming auricular muscles, while the ventral branch migrates into the face, forming muscles of facial expression.

Correlation between branchial arch and facial expression patterns

At 10.0 dpc, BAC195, BAC140 and BAC88 (Fig. 3A-C) constructs expressed the transgene in mandibular and hyoid arches whereas BAC59 lines only expressed reliably in the hyoid arch (Fig. 3D). Fig. 5A-D illustrates the ontogeny of the mandibular arch derivatives; Fig. 5E-H, by subtraction, illustrates that of the hyoid arch derivatives

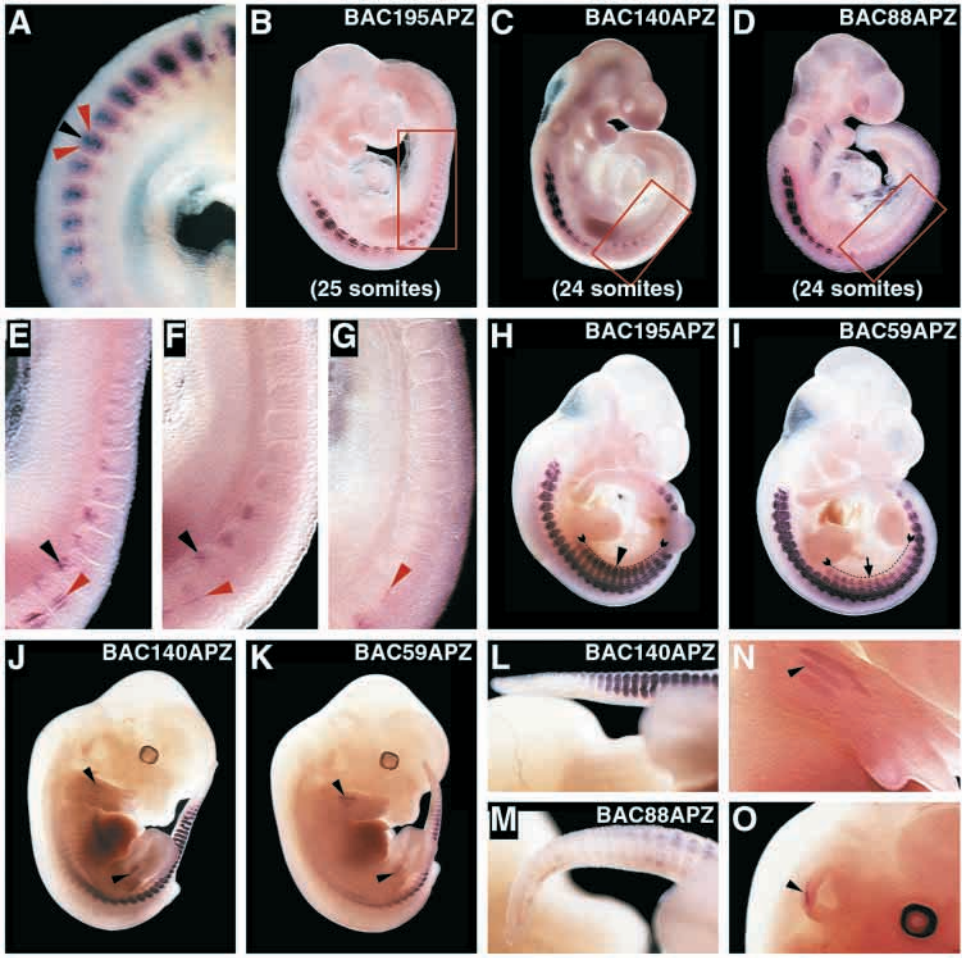

Fig. 6. Analysis of *Mrf4* expression using double reporter gene constructs. (A) Detail of a doubly stained 10.0 dpc embryo from a BAC59APZ line. Expression of both transgenes is visible in the myotome. The different cellular localisation signals of the two reporter genes allow the visualisation of *Mrf4* (AP, red arrowheads) and *Myf5* (β-galactosidase, black arrowhead). At 9.5 dpc an early hypaxial domain of *AP-Mrf4* expression is detected in BAC195APZ (B,E; black arrowhead) and BAC140APZ (C,F; black arrowhead) lines. Other somitic expression is restricted to spindle shaped myocytes in dorsal and ventral myotome (E,F; red arrowheads), which is the only expression domain in BAC88APZ lines (D,G). At 11.5 dpc, BAC195APZ lines (H) express the *AP-Mrf4* transgene in the most ventral part of the thoracic somites (arrowhead) which is not seen in BAC88APZ (not shown) or BAC59APZ (I, arrow) lines. In both cases the dotted line marks the ventral edge of the somites. (J,K) Expression is downregulated at 12.5 dpc in BAC140APZ (J) and BAC59APZ lines (K) and upregulated in limb buds (black arrowheads). (L,M) Expression in the tail occupies the entire myotome in BAC140APZ lines (L) but is restricted to the dorsal half in BAC59APZ (not shown) and BAC88APZ lines (M). At 13.5 dpc expression is maintained in limb muscles (N, arrowhead) and activated in the pinna of the ear (O, arrowhead) in BAC59APZ and BAC195APZ (not shown).

Mandibular arch muscle progenitors (Fig. 5A-D; red arrowheads) migrated as a single condensation to a position midway between the presumptive eye and the ear (11.5 dpc). The first fibres could be seen, running in a dorsoventral direction, as the cells separated into individual groups (12.5 dpc and later) which will form the muscles of mastication.

Hyoid arch muscle progenitors (Fig. 5E-H; black arrowheads) migrated out of the arch from 10.5 dpc, splitting into dorsal (x) and ventral (y) branches by 11.5 dpc. The dorsal domain divided further at 12.5 dpc and gave rise to the external muscles of the ear (x), while the ventral domain elongated rostrally and separated into dorsal and ventral branches (y) which did not divide into different muscle masses until 13.5 dpc (data not shown). These branches expanded towards the snout and the eye regions where they will form the muscles of facial expression.

Double reporter-gene BAC constructs

The introduction of two reporter genes into a single BAC enabled us to analyse and compare the expression patterns of *Mrf4* and *Myf5* in the same genomic context and in the same embryo. Double BCIP/NBT (red) and X-gal (blue) staining allowed us to clearly identify the locations of *AP-Mrf4* (membrane bound) and *nlacZ-Myf5* (nuclear localised) transgene expression. Fig. 6A shows the double expression pattern in thoracic somites in a BAC59APZ embryo at 10.0 dpc (32 somites). The *Myf5* reporter marked the nuclei of the myotomal cells, aligned at the centre of the myoblasts, while the *Mrf4* reporter stained the surrounding membrane, occupying the entire width of the myotome. Sectioning through doubly or singly stained embryos showed *Mrf4*-transgene expression restricted to the myotome, while the *Myf5*-transgene was expressed in myotome and dorsal dermomyotome (data not shown). In practice, we usually compared littermates stained independently for AP or β-galactosidase, which allows the detection of the more delicate details of the expression patterns of both transgenes.

Identification of a distal regulatory region driving ventral *Mrf4* expression

There are two phases of *Mrf4* expression during development: an early one restricted to embryonic myotome and a late one in foetal skeletal muscle (Bober et al., 1991; Hinterberger et al., 1991). Plasmid-based transgenic analyses have reproduced only a subset of the full expression pattern (Patapoutian et al., 1993; Pin et al., 1997).

AP-Mrf4 expression was first detected in all BAC transgenic lines at 9.0 dpc in the central part of the myotome of rostral somites; it extended rapidly in a rostrocaudal sequence along the body axis and both dorsally and ventrally within individual somites. We observed no differences in *AP-Mrf4* expression between BAC195 and BAC140 lines at any stage analysed. BAC195APZ (Fig. 6B,E) and BAC140APZ (Fig. 6C,F) lines revealed a new ventral domain of expression in anterior thoracic somites at 9.5 dpc (24-28 somite stage) and earlier, confirming our recent in situ hybridisation data (D. S., C. Halai and P. W. J. R., unpublished). Unlike *Mrf4*-expressing cells in dorsal myotome, these cells were not spindle shaped (Fig. 6E,F; compare red and black arrowheads). This ventral domain of expression was missing in BAC88APZ (Fig. 6D,G) and BAC59APZ (data not shown) lines. At the 38-44 somite stage (10.5 dpc-11.0 dpc), BAC195APZ (Fig. 6H) and BAC140APZ (data not shown) lines expressed the transgene throughout the myotome and the somitic bud while in BAC88APZ (data not shown) and BAC59APZ lines expression was absent in the most ventral portion of the thoracic somites (Fig. 6I). At 12.5 dpc, BAC195APZ and BAC140APZ (Fig. 6J,L) lines showed dorsal and ventral expression in tail somites, but in BAC88APZ (Fig. 6M) and BAC59APZ (Fig. 6K) lines this expression was restricted to the dorsal domain. Thus, a ventral domain control region is located in the −140 to −88.2 kb interval, overlapping an element that drives equivalent ventral *Myf5* expression (see above).

The first phase of *AP-Mrf4* expression downregulated from 12.0 dpc, in agreement with previous in situ hybridisation data (Bober et al., 1991; Hinterberger et al., 1991). However, the second phase of expression commenced earlier than reported; it started in both fore- and hindlimbs before 12.5 dpc (Fig. 6J,K; black arrowheads) and increased in intensity in subsequent stages (e.g. Fig. 6N, arrowhead). By 13.5 dpc expression could also be detected in the pinna of the ear muscles (Fig. 6O, arrowhead). There was a generalised upregulation of the transgene from 14.5 dpc (data not shown).

DISCUSSION

We have previously shown that *Myf5* expression is regulated by a number of discrete enhancers in a 14.2 kb region that spans the *Mrf4*/*Myf5* locus, and that each of these enhancers is specific for a particular population of muscle precursor cells (Summerbell et al., 2000). These data also made it clear that there must be other elements, remote from the locus, that control important aspects of the expression pattern of the gene. We have now turned to BAC transgenesis, using clones modified by homologous recombination in *E. coli*, to identify those missing elements and show that a BAC that contains 140 kb upstream of the *Myf5* gene is capable of recapitulating the complete expression patterns of both *Mrf4* and *Myf5*. The results confirm and reinforce the main conclusion of our plasmid-based studies, namely that *Myf5* expression is independently regulated in each of a large number of populations of muscle progenitors, and they begin to illuminate the evolutionary origins of this complex regulation. Further complexity is revealed by our demonstration that elements controlling *Mrf4* and *Myf5* are embedded in a neighbouring gene, which is apparently not expressed in muscle. In addition, *Mrf4* and *Myf5* are expressed at the same time in the ventral domain of the thoracic and tail somites, suggesting that Mrf4 acts earlier in muscle development than previously thought. Strikingly, we show that this particular aspect of the expression pattern of both genes is regulated by sequences in the same genomic interval, raising the possibility that a single element could control both genes.

The multiplicity of *Myf5* regulatory elements

Analysis of a number of BACs from the contig, which constitute a de facto 5′ deletion series, allowed us to localise sequences controlling particular aspects of the expression pattern of *Myf5*. In a parallel series of experiments, Hadchouel et al. used YAC transgenesis to analyse *Myf5* regulation (Hadchouel et al., 2000); the end-points of their YACs and our

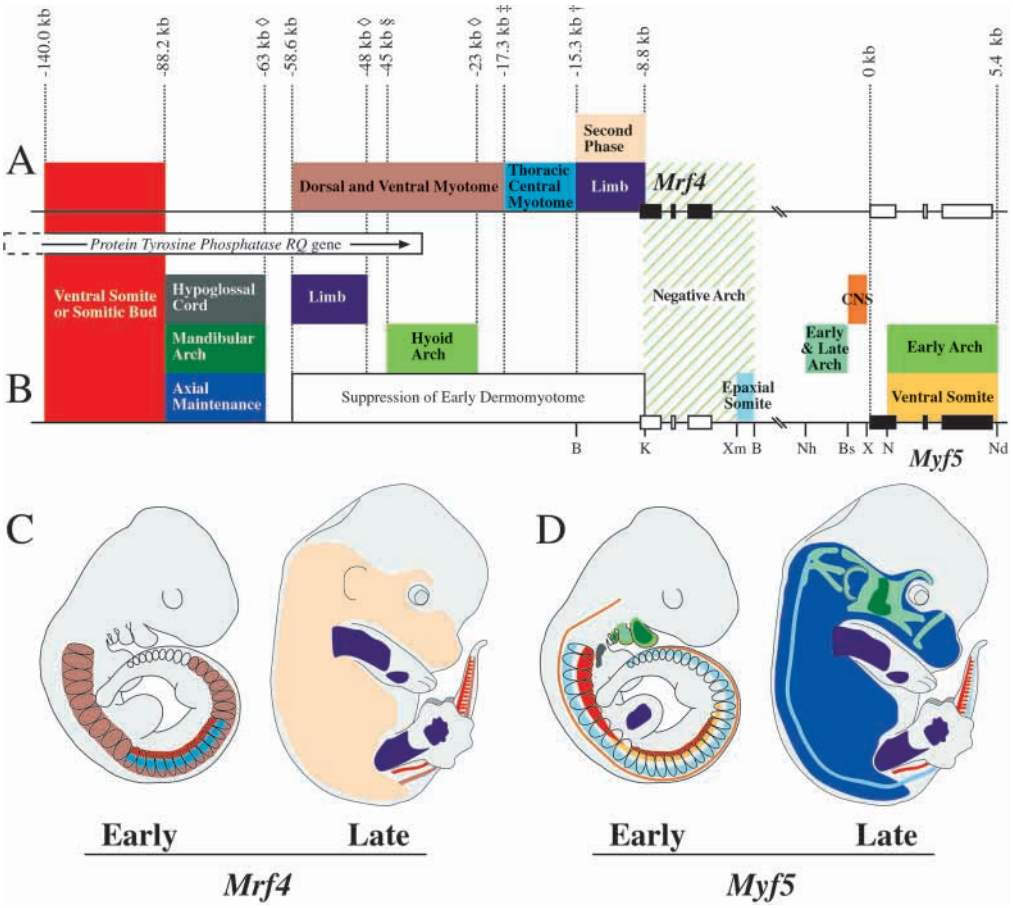

Fig. 7. Summary of the expression domains and regulatory regions for both *Mrf4* and *Myf5*. (A,B) Regulatory regions for *Mrf4* (A) and *Myf5* (B) are colour coded and their anatomical expression domains are represented below at early and late stages of development (C, *Mrf4*; D, *Myf5*).The ventral somite or somitic bud (red) region may contain a single element which controls both genes. The negative arch (green hatched) region suppresses late *Myf5* arch expression in p8.8Z lines but in the context of the locus may be involved in shielding the *Mrf4* promoter from the arch elements. The map is not to scale. §, approximate YAC ends (Zweigerdt et al., 1997); ◊, YAC ends (Hadchouel et al., 2000); ‡, rat genomic distance (Pin et al., 1997); †, restriction enzyme boundaries of plasmid constructs (Patapoutian et al. (1993); B, *Bam*HI; Bs, *Bsa*BI; K, *Kpn*I; N, *Nco*I; Nd, *Nde*I; Nh, *Nhe*I; X, *Xba*I; Xm, *Xmn*I.

BACs are usefully different. Thus, by comparing these data sets, and also using the earlier results of Zweigerdt et al. (Zweigerdt et al., 1997), we can further delimit these elements (Fig. 7).

The −140 to −88.2 kb region is required for *Myf5* expression in the ventral domain of tail somites and the most ventral component of thoracic somites, the developing somitic bud. Expression in the thoracic ventral domain is particularly dynamic, and it is necessary to focus on the 12 hour interval between 28 and 36 somites, while expression in the tail ventral domain is seen only from 12.5 dpc onwards. Hadchouel et al. suggest that all the sequences required for full *Myf5* expression are contained within 96 kb upstream of the gene, but they did not show data for these time intervals (Hadchouel et al., 2000). A YAC clone containing approximately 95 kb upstream of the *Myf5* gene was not able to drive this ventral somitic expression in the tail at 12.5 dpc (see Fig. 4B in Zweigerdt et al., 1997) but the missing expression domain could be due to the chimaeric nature of these embryos. Therefore the available data do not allow us to refine the position of this element further.

The −88.2 to −58.6 kb region directs consistent expression in the mandibular arch and hypoglossal cord, and is required to maintain expression levels in axial muscles from 11.5 dpc onwards. A YAC containing 63 kb of upstream sequences is not able to drive these aspects of the expression pattern (Hadchouel et al., 2000), localising the required element(s) to the −88.2 to −63 kb interval.

In the −58.6 to −8.8 kb region lie elements that direct expression in the limbs, block the suppression of late hyoid arch expression and suppress the inappropriate dermomyotomal expression seen in p8.8Z lines (Summerbell et al., 2000). Hadchouel et al. (Hadchouel et al., 2000) have localised the limb element to between −58 and -48 kb, a conclusion that is consistent with Zweigerdt et al. (1997) and our own data. Late hyoid arch expression is seen with a −45 kb YAC (Fig. 2D in Zweigerdt at al., 1997) but not with one of −23 kb (Fig. 6B in Hadchouel et al., 2000); thus by combining the data sets we can localise this element to between −45 and −23 kb.

The diversity of regulatory elements parallels the heterogeneity of somite-derived muscle precursors

Initiation of somitic expression

Proper expression in the somites requires at least four elements, one of which acts negatively. We previously defined two enhancers, one specific for the epaxial domain, one for the hypaxial (Summerbell et al., 2000). The epaxial enhancer acts in the dorsal lip of the dermomyotome and the dorsal-most myotome (Fig. 2B and Summerbell et al., 2000). None of our data indicate a requirement for an additional element in the initiation of epaxial expression. The hypaxial enhancer located within the *Myf5* gene drives a subset of the early hypaxial pattern; while the timing is correct the location within the ventral somite is not and the transgene is activated in caudal dermomyotome (Fig. 2D and Summerbell et al., 2000). This indicates that an element elsewhere within the locus normally prevents expression in this region. The present data show that this negatively acting element is within BAC59, but full

hypaxial expression is obtained only with BAC140, which carries the distal ventral element (see above). Thus, correct hypaxial expression probably requires the interaction of two positively acting elements that control initial activation, together with the negative element that prevents inappropriate dermomyotomal expression.

Hypaxial-derived muscle progenitors

The hypaxial muscle precursors give rise to a diversity of muscle populations, those of the limb, trunk, pharynx and tongue. In limb muscle progenitors myogenesis is delayed such that *Myf5* is not activated until after migration. *Myf5*$^{-/-}$ embryos show a 2.5 day delay in the activation of the myogenic programme in the axial musculature but the timing of limb myogenesis is not affected (Kablar et al., 1997). While in the trunk and the head multiple interacting elements are involved in initiation of expression in different muscle precursors, and an additional region(s) is required to cooperate with these in the maintenance of the expression levels, the available data do not subdivide the regulation of limb expression in the same manner and do not suggest interaction between its control region and others. This is consistent with the view that limb and axial muscles are controlled by separate circuits.

The progenitors of the ventral-most thoracic muscles express both *Mrf4* and *Myf5* early in myogenesis, and give rise to hypaxial musculature by elongation rather than migration. The first wave of myoblasts originates from ventral dermomyotome and elongation of the hypaxial myotome is then accomplished by the intercalation of second wave myoblasts derived from the four lips of the dermomyotome (Cinnamon et al., 1999). In this subset of hypaxial precursors the expression of both genes is controlled by an element, or elements, contained in the same genomic interval. In addition we observe a correlation between early ventral thoracic expression and later expression along the entire dorsoventral axis of the hypaxial domain, suggesting that this element(s) may be responsible for gene activation in second wave myoblasts.

Another group of hypaxial muscle precursors is formed by the progenitors of the pharyngeal and tongue muscles, which arise from the hypaxial domain of occipital somites that express *Myf5*, *MyoD* and *myogenin* (but not *Mrf4*) throughout their migration. *Myf5* expression in this population is regulated by a separate element in the −88.2 to −63 kb region.

There is thus a striking correlation between the heterogeneity of hypaxial muscle precursors and the diversity of elements required for proper *Myf5* regulation, suggesting that the emergence of the different hypaxial muscle groups during vertebrate evolution was paralleled by the acquisition of new regulatory sequences able to interpret the new signalling environments that initiate myogenesis.

Branchial arch expression of *Myf5* and the ontogeny of the facial musculature

The regulation of *Myf5* expression in the facial muscle precursors is also complex. We have previously defined a proximal enhancer between *Mrf4* and *Myf5* that is capable of driving consistent expression in the hyoid arch but only variable expression in the mandibular arch, and an intragenic enhancer, acting primarily in the hypaxial domain of the somites, that can also direct weak expression in the arches from 9.0 to 10.25 dpc. In the context of p8.8Z, expression is downregulated from 12.5 dpc and we defined this negative element as overlapping the epaxial enhancer and the *Mrf4* gene (Patapoutian et al., 1993; Summerbell et al., 2000). The activity of the intergenic enhancer is not maintained from 11.5 dpc by a YAC containing 23 kb of upstream DNA (Fig. 6C,D in Hadchouel et al., 2000) but late hyoid arch expression is restored in BAC59 lines and mandibular arch expression becomes reliable in BAC88 lines. We can thus define five elements involved in proper expression in the arches. The negative element that overlaps *Mrf4* antagonises the early arch enhancer at later times; the activity of this element would then be over-ridden in the mandibular and hyoid arches by elements located in the −88.2 to −63 kb and the −45 to −23 kb regions, respectively. This arrangement of multiple regulatory elements probably reflects the highly complex signalling environment present during facial development (for a review, see Francis-West et al., 1998).

The fact that *Myf5* expression is independently controlled in the different arches allows us to follow the development of muscles derived from the hyoid arch in BAC59 lines and, by subtraction, those derived from the mandibular arch in lines carrying the larger BACs. In the avian embryo, paraxial mesoderm migrates into the branchial arches and the contribution of these precursors to facial muscles is well documented (Hacker and Guthrie, 1998; Noden et al., 1999). Although it is assumed that mammalian craniofacial muscle development follows the same programme, few data are available, and these mostly focus on the first migration of paraxial mesoderm cells into the arches (Trainor et al., 1994; Trainor and Tam, 1995). We have now looked at those cells that express *Myf5* in the branchial arches and their derivatives at successive developmental stages and conclude that muscles involved in mastication derive from the mandibular, while muscles involved in facial expression originate in the hyoid. Thus their ontogeny is generally equivalent to that described by Noden et al. for the chick (Noden et al., 1999). However, we have identified an important difference in the behaviour of a subset of the cells that populate the hyoid arch. The population of mesodermal cells in the core of the arch splits into dorsal and ventral domains; the dorsal domain gives rise to the external muscles of the ear (Fig. 5E-H). Noden et al. also describe two groups of *Myf5*-expressing cells in the hyoid arch, the dorsal group located in close proximity to the otic vesicle (Noden et al., 1999). However, this cell population subsequently becomes incorporated into the larger ventral cell population of the arch, which will give rise to the mandibular depressor, among other muscles. During early evolution of mammals the retroarticular process and the attached mandibular depressor were lost (Köntges and Lumsden, 1996). It is interesting to speculate that this released the progenitor cells destined to form this muscle, which subsequently acquired the capacity to form the external musculature of the ear.

Mrf4 expression is regulated by elements interdigitated between those that control *Myf5*

By inserting a second reporter gene into the BACs we show that the regulation of *Mrf4* expression is also complex. Sequences between −15.3 and −8.8 kb are required for limb expression and for the second phase of gene activity (Patapoutian et al., 1993) and sequences between −17.3 and −15.3 kb drive early expression in the central myotome of thoracic somites (Fig. 2A in Pin et al., 1997).

In BAC-APZ lines, *AP-Mrf4* expression is first detected in the central part of the myotome, where it is confined to differentiated myocytes, and then expands dorsally and ventrally within the myotome, as the epaxial and hypaxial domains grow dorsomedially and ventrolaterally, respectively (Denetclaw and Ordahl, 2000). We also show that sequences within BAC59 drive strong expression in rostral, thoracic and caudal somites, indicating that myotomal expression is controlled by at least two separate elements: one proximal, driving only central myotome expression at thoracic levels, as previously described (Pin et al., 1997); and one distal (between −58.6 and −17.3 kb), driving the remaining somitic expression.

The −140 to −88.2 kb region is required for *Mrf4* and *Myf5* expression in the most ventral part of thoracic somites. There are striking similarities in the expression patterns of the two genes, and their regulation, in this set of precursors: (1) activation of expression in the ventral dermomyotome of thoracic somites is very similar temporally and spatially; (2) lack of early ventral expression correlates with the absence or reduction of transgene expression in the somitic bud at later stages; and (3) expression in tail somites is regulated independently in the dorsal and ventral domains. These observations raise the exciting possibility that a single element is acting simultaneously on both genes. The two genes have been linked throughout vertebrate evolution, and form a syntenic group in teleost fish (O. Coutelle, C. Moreno de Barreda and P. W. J. R., unpublished), birds (Saitoh et al., 1993) and mammals (Braun et al., 1990; Patapoutian et al., 1993). This linkage may have been maintained by co-regulation of this aspect of the expression patterns of the genes. The coincident activation of the two genes in this domain suggests that *Mrf4* may not be functionally downstream of *Myf5* and that in this case it might act earlier in myogenesis than previously thought.

Complexity in the regulation of the *Mrf4*/*Myf5* locus

Myf5 is the first myogenic regulatory factor gene to be activated during mouse development. We have previously proposed that the complex architecture of its regulatory elements reflects its special function and is required to interpret the distinct signalling environments that initiate myogenesis in the various progenitor populations (Summerbell et al., 2000). The present data strongly support this view. We have now demonstrated that the regulation of *Mrf4* is also complex, and that the elements that control the two genes are interspersed, raising the question of how they distinguish between the two promoters to give generally distinct expression patterns. This is complicated further by the observation that many of these regulatory elements are within the adjacent gene for PTP-RQ, which is expressed in the glomerular mesangial cells of the kidney (Wright et al., 1998) and does not appear to be expressed at appreciable levels in skeletal muscle precursors (J. J. C., C. Halai and P. W. J. R., unpublished). This indicates a requirement for transcriptional insulators to shield the promoter of the PTP-RQ gene from the regulatory elements controlling *Mrf4* and *Myf5* expression, and vice versa. It is, however, of considerable interest to note that kidney mesangial cells are myofibroblasts that express many genes normally thought of as muscle specific, including *MyoD* and *myogenin* (Mayer and Leinwand, 1997).

The promoters and the distal control elements must have evolved so that *Mrf4* expression cannot be activated by signals directed at *Myf5*, and vice versa, and the elucidation of the mechanisms underlying this specificity will be a major topic of future research. The multiplicity of elements controlling branchial arch expression will provide a good model because in the complex environment of the arches only *Myf5* is expressed, despite the fact that some of the necessary elements are physically closer to *Mrf4*. The negatively acting *Myf5* element that we have mapped as overlapping *Mrf4* may well be involved in shielding the latter from the arch elements but when we varied the context it interfered with *Myf5* expression in p8.8Z lines. Similarly, Hadchouel et al. have shown that sequences from −58 to −48 kb, when juxtaposed to the intergenic arch enhancer and a *Myf5* minimal promoter, can direct expression in limbs, arches, hypoglossal cord, and both epaxial and hypaxial components of the myotome, and conclude that many of the sequences important for proper *Myf5* regulation are within this 10 kb interval (Hadchouel et al., 2000). As we have previously shown that elements outwith this segment function as enhancers in the epaxial and hypaxial domains, this raises the possibility of redundancy. However, it is important to consider two aspects of the expression pattern driven by this 10 kb region (see Fig. 7B in Hadchouel et al., 2000): (1) in thoracic somites transgene expression is strong in the dorsal and ventral myotome but weaker in the centre, a characteristic of the *Mrf4* element that we have localised to the −58.6 to −17.3 kb interval; and (2) there is a delay in expression in tail somites, with only the first five caudal somites from the hindlimb expressing the transgene dorsally at 11.5 dpc, mimicking the pattern of *Mrf4* expression in caudal somites at this stage. It is thus possible that the −58 to −48 kb interval contains some *Myf5*-specific elements, e.g. the one that acts in the limb, together with elements that can act on *Myf5* when juxtaposed to it, but which in the context of the locus act on *Mrf4*. Equally, an element in this interval may act on both genes, like the upstream ventral element described above.

The presently available data open the way to a precise definition of all of the elements that control *Myf5* and thus to the identification of the cognate transcription factors and the signals that regulate them. Such work will reveal the rich diversity of signalling environments that are capable of initiating myogenesis in the multiple progenitor populations present in the embryo. It will be of interest to conduct parallel studies in other organisms in order to try to discern how this complex regulation evolved. However it arose, it will be a considerable challenge to understand how these two linked genes are controlled so that their expression patterns are sometimes distinct and sometimes overlapping.

We are grateful to the staff of the NIMR Biological Services Division, particularly Hannah Boyes and Jane Sealby, for expert animal husbandry; to Nat Heintz for helpful discussions; and to Marjorie Kipling and other members of the Molecular Embryology team at the Institute of Cancer Research for their critical reading of the manuscript. J. J. C. was supported by a Research Training Fellowship from the Medical Research Council (UK), which also paid for this work.

REFERENCES

Arnold, H. H. and Braun, T. (2000). Genetics of muscle determination and development. *Curr. Top. Dev. Biol.* **48**, 129-164.

Bober, E., Lyons, G. E., Braun, T., Cossu, G., Buckingham, M. and Arnold, H. H. (1991). The muscle regulatory gene, Myf-6, has a biphasic pattern of expression during early mouse development. *J. Cell Biol.* **113**, 1255-1265.

Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. and Arnold, H. H. (1989). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.* **8**, 701-709.

Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H. H. (1990). Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J.* **9**, 821-831.

Buckingham, M. (1992). Making muscle in mammals. *Trends Genet.* **8**, 144-149.

Cheng, T. C., Wallace, M. C., Merlie, J. P. and Olson, E. N. (1993). Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science* **261**, 215-218.

Christ, B. and Ordahl, C. P. (1995). Early stages of chick somite development. *Anat. Embryol.* **191**, 381-396.

Cinnamon, Y., Kahane, N. and Kalcheim, C. (1999). Characterization of the early development of specific hypaxial muscles from the ventrolateral myotome. *Development* **126**, 4305-4315.

Couly, G. F., Coltey, P. M. and Le Douarin, N. M. (1992). The developmental fate of the cephalic mesoderm in quail-chick chimeras. *Development* **114**, 1-15.

Davis, R. L., Weintraub, H. and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000.

Denetclaw, W. F. and Ordahl, C. P. (2000). The growth of the dermomyotome and formation of early myotome lineages in thoracolumbar somites of chicken embryos. *Development* **127**, 893-905.

Edmondson, D. G. and Olson, E. N. (1989). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev.* **4**, 628-640

Francis-West, P., Ladher, R., Barlow, A. and Graveson, A. (1998). Signalling interactions during facial development. *Mech. Dev.* **75**, 3-28.

Hacker, A. and Guthrie, S. (1998). A distinct developmental programme for the cranial paraxial mesoderm in the chick embryo. *Development* **125**, 3461-3472.

Hadchouel, J., Tajbakhsh, S., Primig, M., Chang, T. H., Daubas, P., Rocancourt, D. and Buckingham, M. (2000). Modular long-range regulation of Myf5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. *Development* **127**, 4455-4467.

Hinterberger, T. J., Sassoon, D. A., Rhodes, S. J. and Konieczny, S. F. (1991). Expression of the muscle regulatory factor MRF4 during somite and skeletal myofiber development. *Dev. Biol.* **147**, 144-156.

Hirsinger, E., Malapert, P., Dubrulle, J., Delfini, M. C., Duprez, D., Henrique, D., Ish-Horowicz, D. and Pourquie, O. (2001). Notch signalling acts in postmitotic avian myogenic cells to control MyoD activation. *Development* **128**, 107-116.

Kablar, B., Krastel, K., Ying, C., Asakura, A., Tapscott, S. J. and Rudnicki, M. A. (1997). MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. *Development* **124**, 4729-4738.

Köntges, G. and Lumsden, A. (1996). Rhombencephalic neural crest segmentation is preserved throughout craniofacial ontogeny. *Development* **122**, 3229-3242.

Kucharczak, K. L., Love, C. M., Dougherty, N. M. and Goldhamer, D. J. (1999). Fine-scale transgenic mapping of the MyoD core enhancer: MyoD is regulated by distinct but overlapping mechanisms in myotomal and non-myotomal muscle lineages. *Development* **126**, 1957-1965.

Mackenzie, S, Walsh, F. S. and Graham, A. (1998). Migration of hypoglossal myoblast precursors. *Dev. Dyn.* **213**, 349-358.

Mayer, D. C. and Leinwand, L. A. (1997). Sarcomeric gene expression and contractility in myofibroblasts. *J. Cell Biol.* **139**, 1477-1484.

Miner, J. H. and Wold, B. (1990). Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl. Acad. Sci. USA* **87**, 1089-1093.

Montoliu, L., Bock, C. T., Schutz, G. and Zentgraf, H. (1995). Visualization of large DNA molecules by electron microscopy with polyamines: application to the analysis of yeast endogenous and artificial chromosomes. *J. Mol. Biol.* **246**, 486-492.

Noden, D. M. (1983). The embryonic origins of avian cephalic and cervical muscles and associated connective tissues. *Am. J. Anat.* **168**, 257-276.

Noden, D. M., Marcucio, R., Borycki, A. G. and Emerson, C. P., Jr (1999). Differentiation of avian craniofacial muscles: I. Patterns of early regulatory gene expression and myosin heavy chain synthesis. *Dev. Dyn.* **216**, 96-112.

Patapoutian, A., Miner, J. H., Lyons, G. E. and Wold, B. (1993). Isolated sequences from the linked Myf-5 and MRF4 genes drive distinct patterns of muscle-specific expression in transgenic mice. *Development* **118**, 61-69.

Pin, C. L., Ludolph, D. C., Cooper, S. T., Klocke, B. J., Merlie, J. P. and Konieczny, S. F. (1997). Distal regulatory elements control MRF4 gene expression in early and late myogenic cell populations. *Dev. Dyn.* **208**, 299-312.

Rhodes, S. J. and Konieczny, S. F. (1989). Identification of MRF4: a new member of the muscle regulatory factor gene family. *Genes Dev.* **3**, 2050-2061.

Saitoh, O., Fujisawa-Sehara, A., Nabeshima, Y. and Periasamy, M. (1993). Expression of myogenic factors in denervated chicken breast muscle: isolation of the chicken Myf5 gene. *Nucleic Acids Res.* **21**, 2503-2509.

Summerbell, D., Ashby, P. R., Coutelle, O., Cox, D., Yee, S. P. and Rigby, P. W. J. (2000). The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors. *Development* **127**, 3745-3757.

Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M. (1996). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Dev. Dyn.* **206**, 291-300.

Trainor, P. A., Tan, S. S. and Tam, P. P. L. (1994). Cranial paraxial mesoderm: regionalisation of cell fate and impact on craniofacial development in mouse embryos. *Development* **120**, 2397-2408

Trainor, P. A. and Tam, P. P. L. (1995). Cranial paraxial mesoderm and neural crest cells of the mouse embryo: co-distribution in the craniofacial mesenchyme but distinct segregation in branchial arches. *Development* **121**, 2569-2581.

Wright, M. B., Hugo, C., Seifert, R., Disteche, C. M. and Bowen-Pope, D. F. (1998). Proliferating and migrating mesangial cells responding to injury express a novel receptor protein-tyrosine phosphatase in experimental mesangial proliferative glomerulonephritis. *J. Biol. Chem.* **273**, 23929-23937.

Wright, W. E., Sassoon, D. A. and Lin, V. K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**, 607-617.

Yang, X. W., Model, P. and Heintz, N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol.* **15**, 859-865.

Yee, S. P. and Rigby, P. W. J. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* **7**, 1277-1289.

Zweigerdt, R., Braun, T. and Arnold, H. H. (1997). Faithful expression of the Myf-5 gene during mouse myogenesis requires distant control regions: a transgene approach using yeast artificial chromosomes. *Dev. Biol.* **192**, 172-180.