Development 122, 2769-2778 (1996)
Printed in Great Britain © The Company of Biologists Limited 1996
DEV4726

MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation

Yumiko Saga[1,3,*], Naomi Hata[1], Satoru Kobayashi[2,3], Terry Magnuson[4], Michael F. Seldin[5,†] and Makoto M. Taketo[1,6]

[1]Banyu Tsukuba Research Institute (Merck) Tsukuba, Ibaraki, 300-26 Japan
[2]Department of Biology and [3]Center for Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, Tsukuba, Ibaraki, 305 Japan
[4]Department of Genetics, Case Western Reserve University, Cleveland, OH 44106, USA
[5]Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA
[6]Department of Biomedical Genetics, Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo 113, Japan

*Author for correspondence at Biomedical Research Laboratories, Banyu Tsukuba Research Institute, 3 Okubo, Tsukuba, Ibaraki 300-26, Japan (e-mail: sagay@banyu.co.jp)
[†]Present address: Department of Biological Chemistry, University of California, Davis, CA 95616, USA

SUMMARY

A subtractive hybridization strategy was used to isolate putative genes involved in the development of mouse primordial germ cells (PGC). Complimentary DNA was amplified on RNA isolated from the base of the allantois where PGC are located in the 7.5 days post coitum (dpc) mouse embryo. It was then subtracted by hybridization with cDNA amplified on RNA of the anterior region where PGC are absent. A novel gene thus isolated is designated as *Mesp1* and encodes a possible transcription factor MesP1 containing a basic helix-loop-helix motif. Its earliest expression was observed at the onset of gastrulation, as early as 6.5 dpc, in the nascent mesodermal cells that first ingressed at the end of the primitive streak. These expressing cells in the lateral and extraembryonic mesoderm showed a wing-shaped distribution. Its initial expression was soon down-regulated at 7.5 dpc before the completion of gastrulation, except at the proximal end of the primitive streak which included the extraembryonic mesoderm and the base of allantois. At 8 dpc, the expression at the base of the allantois moved laterally. This distribution between 7.0 and 8.0 dpc was similar to that of PGC detected by the alkaline phosphatase activity. However, the expression of *Mesp1* was down-regulated thereafter, when PGC entered in the migration stage. After birth, *Mesp1* expression was detected only in mature testes, but in a different isoform from that expressed in the embryo. *Mesp1* was mapped to the mid region of chromosome 7, near the mesodermal deficiency gene (*mesd*). However, a Southern hybridization study clearly showed that *Mesp1* was distinctly different from *mesd.* The amino acid sequence and its expression pattern suggest that MesP1 plays an important role in the development of the nascent mesoderm including PGC.

Key words: gastrulation, primordial germ cells, mesoderm, mouse, *Mesp1*

Accession number in GenBank: D83674
Accession number in GDB: MGD-CREX-520

INTRODUCTION

The process of gastrulation establishes not only the basic body plan of the animal but also the orderly distribution and differentiation of the definitive fetal tissues (reviewed in Tam and Beddington, 1992). In the mouse, gastrulation begins with the formation of the primitive streak, a local thickening in the embryonic ectoderm through which cells pass on their way to forming the embryonic and extraembryonic mesoderm, as well as the definitive endoderm (Batten and Haar, 1979; Hashimoto and Nakatsuji, 1989). The site of streak formation marks the posterior end of the embryo, and defines its anterior-posterior axis, accordingly. Studies of the cell fates in the primitive streak have shown that it is regionalized so that a fate map can be drawn that is topologically similar to those of other vertebrates. However, the fate of epiblasts of prestreak and midstreak-stage embryos was not fixed and cell mixing occurs (Lawson and Pedersen, 1992). This suggests that cells may be determined as they pass through the streak regarding their ultimate fate within the embryo. In addition to the formation of three germ layers, the earliest distinction between the somatic and germ cell lineages also becomes evident during gastrulation. In the mouse, primordial germ cells (PGC) are first distinguishable from other cells of the embryo at 7 days post coitum (dpc) by their high alkaline phosphatase activity, although it is now known that the activity itself is not involved in the development of PGC (MacGregor et al., 1995). These cells do not appear as a discrete tissue, but as a relatively scattered population of cells (Ginsburg et al., 1990), localized to the extraembryonic mesoderm at the base of the allantois,

near the endoderm that will give rise to the hind gut. This region invaginates then, and as a result, the germ cells are swept into the embryo and become embedded in the hindgut wall where they are found at 9 dpc. Before 7 dpc, there are no markers for the mouse germ line. However, single-cell microinjection studies on 6-6.5 dpc embryos indicated that allocation to the germ line takes place during the first 16 hours of gastrulation (Lawson and Hage, 1994). The epiblast precursors of the PGC reside in the most proximal part of the presumptive extraembryonic mesoderm before gastrulation, and are not yet restricted to the germline. It is not known whether the proximity to the extraembryonic ectoderm and the spatial separation from the presumptive embryonic somatic lineage is important for the germ line to develop. It is conceivable that both the timing and location of the PGC lineage restriction help seclude the germline from the influence of genes controlling regional specification (Dixon, 1994), because cell allocations take place in the extraembryonic mesoderm.

In order to isolate the genes possibly involved in the differentiation of mouse PGC, we employed a regional subtraction hybridization strategy. Complementary DNA was prepared from RNA extracted from the base of the allantois where the putative precursor of PGC is located, and subtracted with cDNA derived from the anterior embryo at 7-7.5 dpc. We report here one of the isolated genes, which shows an interesting expression pattern that is restricted in the early mesoderm cells including PGC. The gene encodes a protein with a basic helix-loop-helix motif and is likely to be one of the transcription factors involved in cell fate determination and differentiation.

MATERIALS AND METHODS

Embryo isolation

Embryos were obtained from naturally mated ICR mice (CLEA, Japan). The day when the vaginal plug was detected was designated as 0.5 dpc. For construction of the PCR-amplified cDNA, about 40 7.5 dpc embryos and 30 8.5 dpc were dissected from uteri in phosphate-buffered saline (PBS). Embryos were dissected from the decidua and the Reichert's membrane was removed. These embryos were then separated into posterior (including the base of the allantois) and anterior regions (Fig. 1A). The separated regions of the embryos were pooled, respectively, and the total RNA was isolated using Isogen (Nippon Gene, Japan).

cDNA amplification

Synthesis and amplification of cDNA were performed according to the method described by Abe (1992). 1 μg of total RNA isolated from the posterior or anterior region of mouse embryos at either 7.5 dpc or 8.5 dpc was mixed with 100 ng of *Not*I-oligo(dT) primer (5′AATTCGCGGCCGCTTTTTTTTTTTTTT-3′), heated at 65°C for 3 minutes and placed on ice. First strand cDNA was synthesized in a 25 μl-reaction mixture containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 500 μM each of dNTPs and 40 units of Moloney murine leukemia virus RNaseH$^-$ reverse transcriptase (Superscript, BRL) at 45°C for 1 hour. Second-strand cDNA was synthesized in a mixture of 50 mM Tris-HCl (pH 7.6), 100 mM KCl, 5 mM $MgCl_2$, 100 mM β-NAD, 10 mM $(NH_4)_2SO_4$, 5 mM DTT, 230 units/ml DNA polymerase I, 200 units/ml *E. coli* DNA ligase and 200 units/ml RNaseH at 12°C for 2 hours. Double-stranded cDNA was treated with T4 DNA polymerase at 12°C for 10 minutes. One twentieth of the synthesized, blunt-ended cDNA was ligated to phosphorylated 'lone linkers' LLsal 1A & B for cDNA derived from the anterior region and LLsal 2A & B for cDNA from the posterior region, for 30 minutes using a ligation kit (TAKARA, Japan).

```
LLsal1A: 5´-pATTGACGTCGACTATCCAGG-3´
LLsal1B: 3´-    CTGCAGCTGATAGGTCCp-5´
LLsal2A: 5´-pTCGAGTCGACTATATGTACC-3´
LLsal2B: 3´-    TCAGCTGATATACATGGp-5´
```

Excess linkers were removed by filtration through Microcon-100 (Amicon) and one half of the ligated cDNA was used for subsequent amplification steps. Lone linker PCR (LL-PCR) amplification was performed in a 100 μl mixture containing 10 mM β-mercaptoethanol, 200 μM each of dATP, dCTP, dGTP and dTTP, 0.01% gelatin, 100 pmol of LLsal1A (for anterior) or LLsal2A (for posterior) primers, the cDNA and 2 units of AmpliTaq (Perkin Elmer/Cetus). After 14 cycles of amplification (94°C for 45 seconds, 53°C for 2 minutes, 72°C for 7 minutes), one-tenth of the reaction product was subjected to an additional five to ten cycles of amplification in 100 μl reaction buffer under the same conditions. Using this protocol of amplification, about 1-2 μg cDNA was obtained.

Subtraction hybridization and library construction

To enrich the posterior region-specific cDNA population, two successive subtractions were performed by the method used for cDNA equalization described by Ko et al. (1990). 200 ng of the posterior cDNA amplified with LLsal2A was mixed with 20 μg of the anterior cDNA amplified with LLsal1A, coprecipitated, dissolved in 10 μl of distilled water and combined with 10 μl of 2× hybridization solution [0.24 M NaH_2PO_4 (pH 6.8), 1.64 M NaCl, 2 mM EDTA, 0.2% SDS]. After denaturation (95°C 5 minutes), association was performed at 65°C for 16 hours. The mixture was applied on a hydroxyapatite column and chromatographed at 65°C. Single-strand cDNA molecules were eluted with 0.15 M phosphate buffer. The eluate was concentrated using a Centricon and amplified using primer LLsal2A to obtain the posterior cDNA. After the second round subtraction, cDNA was digested with *Sal*I (partially filled with dTTP and dCTP) and subcloned in Bluescript SK(+) vector (Stratagene) digested with *Bam*HI (partially filled with dATP and dGTP).

Dot blot analysis

About 100 ng of cDNA amplified using either LLsal1A or LLsal2A were blotted on Hybond-N(+) nylon membranes (Amersham) and used for hybridizations. For screening, cDNAs amplified from the posterior and anterior regions were blotted on the same membranes and the intensities of hybridizations were compared.

In situ hybridization

Embryos isolated at 6.0-9.5 dpc and testes (2 weeks and 3 months) were fixed in 4% paraformaldehyde in Mg^{2+}/Ca^{2+}-free PBS for 1 hour to overnight, depending on their size. The embryos were then dehydrated in a graded series of 25, 50, 75 and 100% methanol in PBT (PBS with 0.1% Tween 20). For cryostat sections, testes were incubated in 30% sucrose for 6 hours after fixation and embedded in OTC compound.

The protocol used for whole-mount in situ hybridization was described by Barnes et al. (1994) except for using maleic acid buffer for washing. Briefly, the embryos were bleached for 1-4 hours in methanol/30% hydrogen peroxide (4/1) and rehydrated. They were then treated for 5 minutes in 15 μg/ml proteinase K in PBT and the reaction was stopped by two washes with 2 mg/ml glycine in PBT for 5 minutes each. Embryos were refixed for 20 minutes in 0.2% glutaraldehyde/4% paraformaldehyde in PBT. Riboprobes labeled with digoxigenin-conjugated UTP were generated in the sense and antisense orientations, respectively. Prehybridization and hybridization solution contained 50% formamide and 5× SSC (pH 5.0), 50 μg/ml yeast total RNA, 1% SDS, 50 μg/ml heparin. Embryos were prehybridized for 1 hour at 70°C followed by hybridization overnight at 70°C in the fresh hybridization solution containing 1 μg probe/ml.

Wash solutions were as follows: solution 1, 50% formamide, 5× SSC (pH 5.0), 1% SDS; solution 2, 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5), 0.1% Tween 20 and solution 3, 50% formamide, 2× SSC (pH 5.0). Washes were carried out as follows: solution 1 for 30 minutes at 70°C; solution 1/solution 2 (1:1) for 10 minutes at 70°C twice; solution 2 for 5 minutes at 37°C three times; 100 μg/ml RNase A in solution 2 for 1 hour at 37°C; and solution 3 for 30 minutes at 65°C twice.

Embryos were then washed with MBST (Maleic acid-buffered saline with 0.1% Tween 20) with 2 mM levamisole for 5 minutes three times, and then incubated for 1 hr in 10% heat-inactivated lamb serum, 2 mM levamisole in MBST. Then, embryos were incubated at 4°C overnight in 500 μl of 1:2000 diluted preabsorbed anti-DIG Fab fragments conjugated with alkaline phosphatase (Boehringer Mannheim). Embryos were washed extensively in MBST containing 2 mM levamisole at RT, and then in NTMT [100 mM NaCl, 100 mM Tris (pH 9.5), 50 mM $MgCl_2$, 0.1% Tween 20] twice at RT for 20 minutes. Embryos were incubated in NTMT with 2 mM levamisole that contained per ml; 4.5 μl of 100 mg/ml NBT and 3.5 μl of 50 mg/ml of BCIP. Reactions were carried out at RT in the dark. For sectioning of whole-mount stained embryos, specimens were dehydrated and embedded in paraffin wax. Sections were cut at 5-7 μm and photographed.

In situ hybridization of testis sections were performed as described (Tsukamoto et al., 1991).

Detection of alkaline phosphatase activity

The detection method for the alkaline phosphatase activity to localize PGC in the whole embryos and in sections were described by Ginsburg et al. (1990).

Northern analysis and RT-PCR

Northern blotting was performed as described previously (Takeda et al., 1994). For RT-PCR, reverse transcription was performed by the same method as used for the first strand cDNA synthesis above, using random primers on the total RNA derived from embryos; 7.5 dpc posterior (600 ng) and testis (1 μg) RNAs as templates with or without reverse transcriptase. After reactions, samples were treated with RNaseA, and the cDNAs were precipitated and resuspended in 100 μl TE. 5 μl of each sample were used for the subsequent PCR reaction. The reaction mixture (50 μl) included 5% DMSO in addition to the buffer described above. After 25 cycles of amplification (94°C for 30 seconds, 55°C for 1 minute, and 72°C for 1 minute), 10 μl of the reaction product were subjected to the agarose gel electrophoresis and transferred to the nylon membrane. The cDNA probe was radiolabeled and used for the detection of specific PCR bands.

RNase mapping

The *Mesp1* genomic DNA fragment spanning from the *Bam*HI site at base number 164 to the *Xho*I site at 575 (Fig. 3A) was subcloned in the Bluescript SK(+) vector and linearized at the *Bam*HI site. RNA probe preparation and subsequent RNase mapping were performed according to the instruction manual supplied by the manufacturer (Ambion).

cDNA and genomic DNA screening, and DNA sequencing

A testis cDNA library constructed in the pAP3neo vector derived from the pCD1 vector (Okayama and Berg, 1983) kindly provided by Nishina (Osaka University, Japan) was screened with the *Mesp1* PCR product. Genomic DNA library constructed from TT2 (B6/CBA) ES cell genomic DNA in the λ Fix II phage vector (Gibco, BRL) was screened with the *Mesp1* PCR product as a probe. Genomic DNA fragments were subsequently subcloned in the pBluescript II SK vector and sequenced using either Sequenase II (USB) or in a Perkin-Elmer Model ABI377 sequencer, and analysed using a GCG sequence analysis software package.

Chromosomal mapping

Interspecific backcross mice were generated by mating C3H/HeJ-*gld* and (C3H/HeJ-*gld* × *Mus spretus*)F_1. A total of 114 N_2 mice were used to map the *Mesp1* locus with a mouse MesP1-specific cDNA probe as described previously (Watson et al., 1992). A description of the probes and restriction fragment length polymorphisms for the loci linked to *Mesp1*, including insulin-like growth factor 1 receptor (*Igf1r*), tyrosinase (*Tyr*) and feline sarcoma oncogene (*Fes*), has been presented (Saunders and Seldin, 1990).

RESULTS

Isolation of the *Mesp1* gene cDNA by a differential expression strategy

A subtraction hybridization strategy was employed to enrich cDNAs expressed specifically at the base of the allantois (see Fig. 1A; Materials and Methods). However, many non-specific sequences were also amplified by the PCR procedure. Such ubiquitously expressed cDNA clones were eliminated by hybridization with the unsubtracted cDNA population, which contained many repetitive sequences. The remaining

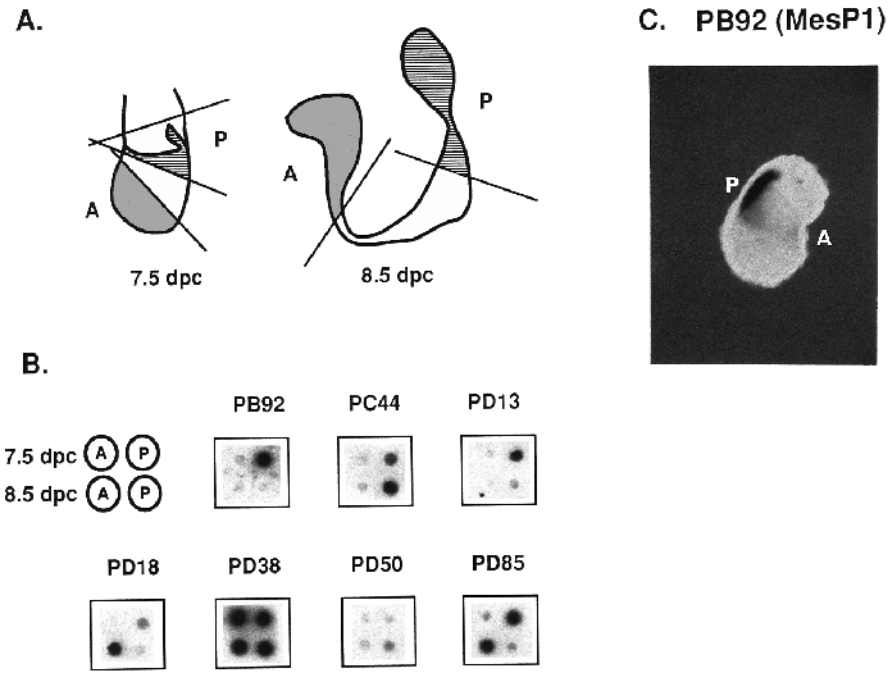

Fig. 1. (A) A schematic drawing of the embryonic portions used for RNA isolation. P (striped); the posterior region containing the base of the allantois. A (grey): the anterior region. (B) Dot blot analysis showing the expression pattern of each candidate clone. Several typical patterns are presented. Note that clone PB92 (MesP1) probe hybridized only to the posterior cDNA at 7.5 dpc. (C) Expression of MesP1 was localized at the base of the allantois in a 7.5 dpc embryo demonstrated by a whole-mount in situ hybridization. A, anterior, P, posterior.

candidate clones were screened further by a dot blot analysis as follows. Complementary DNAs were amplified with lone linkers from RNAs isolated from the posterior region including the base of the allantois, and from the anterior region, respectively. They were then blotted on the same nylon membrane and hybridized with each radio-labeled probe of the candidate cDNA clone. Depending on the clones isolated, different expression patterns were obtained as shown in Fig. 1B. We further investigated those clones that were expressed specifically at the base of the allantois at 7.5 dpc, and in the posterior region at 8.5 dpc. In total, more than 20 out of 500 independent clones were isolated by these criteria. Finally, we determined the expression pattern of each clone by a whole-mount in situ hybridization in the 7.5 dpc embryos. Only one clone (PB92 that we designated as MesP1; for 'mesoderm, posterior 1'; gene symbol, *Mesp1*) showed a clear expression at the base of the allantois (Figs 1C, 2). Most of the other clones showed no staining at all. The expression pattern of MesP1 was quite similar to that of the alkaline phosphatase (ALP) staining for PGC (see below).

Mesoderm-specific expression of MesP1 revealed by a whole-mount in situ analysis

The spatial pattern of MesP1 expression during the early postimplantation development was determined by a whole-mount in situ hybridization, and compared with that of the PGC distribution revealed by the ALP activity. Before gastrulation, no clear staining was observed. The earliest expression was detected at the junction of the epiblast and extraembryonic ectoderm, which is known to be the initiation site for gastrulation (Fig. 2A). Around 6.5-6.75 dpc (the early streak stage), very strong expression of MesP1 transcripts became evident as gastrulation progressed (Fig. 2A). Transverse sections through the embryo at this stage revealed a hybridizable signal present in the mesoderm, including the invaginating primitive streak cells, but neither in the ectoderm nor in the endoderm (Fig. 2C,D). This staining pattern strongly suggested that all nascent mesodermal cells, including those of the extraembryonic lineage, expressed the MesP1 transcripts. However, no specific ALP was detected at this early stage (6.5-6.75 dpc). A dramatic change in the MesP1 expression pattern was observed before the completion of gastrulation. The anterior limit of its expression never extended beyond the node, and most expression in the mesoderm cells was down-regulated except at the base of the allantois and the lateral margin of the primitive streak (Fig. 2E-G). Thus, after the down-regulation of the initial expression, MesP1 was localized at 7.5 dpc at the base of the allantois (Fig. 2E-G) where the ALP activity became detectable (Fig. 2H). Transverse sections of these specimens revealed a similar staining pattern. Namely, the mesodermal cells in the allantois were positive for both MesP1 (Fig. 2I) and ALP (Fig. 2J), although we have not succeeded in a double staining with MesP1 and ALP using the same specimen. During the next few hours of development, MesP1 and ALP expression patterns showed similar changes, reflecting the cell movement in the embryos (Fig. 2K,L). PGC are known to migrate out from the base of the allantois to the edges of the presumptive tailbud and then into the endoderm of the hindgut. MesP1 expression also moved from the base of the allantois to the edges of the tailbud.

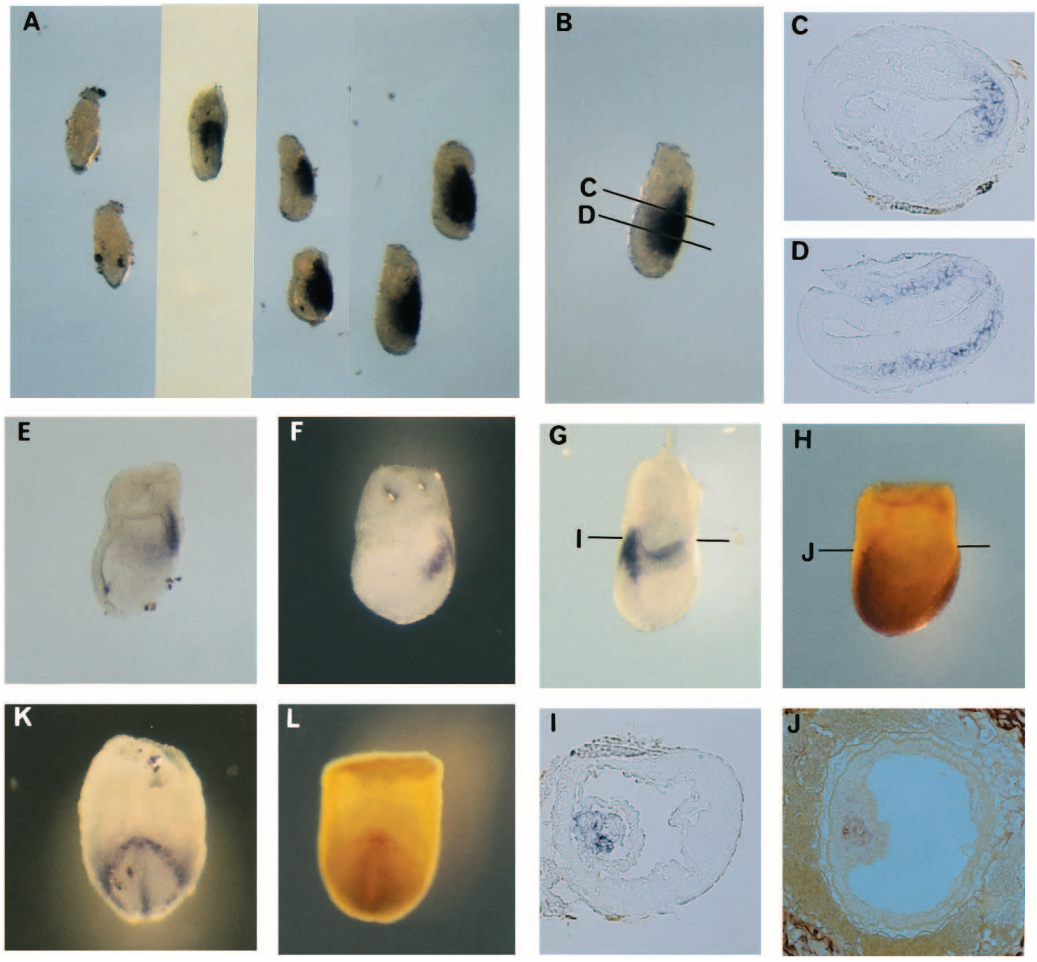

Fig. 2. Expression of MesP1 in early mouse embryos and comparison with PGC localization revealed by ALP activity. (A) Whole-mount in situ hybridization showing the onset of MesP1 expression during gastrulation at 6.5-6.75 dpc. (B) A-P positions of the transverse sections in C and D are marked. (C,D) Transverse sections of a stained embryo showing the expression restricted in the nascent mesoderm and ingressing primitive streak cells, but not in the ectoderm layer. (E-L) 7.5-8.0 dpc early neural fold stage embryos stained with either with the MesP1 probe (E,F,G,I,K) or for ALP (H,J,L). (I,J) Transverse sections of the stained embryos probed with MesP1 (I) and stained for ALP (J).

However, almost all of its expression disappeared before 8.5 dpc. In contrast, the ALP activity of PGC became stronger and served as a stable marker for PGC. After 8.5 dpc, the two faint stained bands appeared on the both sides of the node and remained at least for one more day (data not shown).

A basic helix-loop-helix motif in MesP1

The MesP1 cDNA initially isolated by PCR contained 768 nucleotides (nt) with an open reading frame starting from the first nucleotide and ending at ntd 420 (Fig. 3A). Apparently, this clone lacked an ATG initiation codon. In the 3′ region, a typical polyadenylation signal AATAAA appeared 16 nt upstream of a poly(A) addition site. A computer-assisted similarity search revealed that MesP1 contained a possible helix-loop-helix (HLH) motif similar to those found in several cDNAs containing the basic HLH motif (Benezra et al., 1990). However, our MesP1 clone lacked the basic region, because of the missing 5′ region. Thus, a genomic DNA library was screened and a fragment was isolated that was contiguous with the upstream region of the MesP1 cDNA clone PB92. Fig. 3A shows the genomic organization of the *Mesp1* gene, which consists of 2 exons separated by a 272 bp intron. The genomic DNA for exon 1 was sequenced and compared with the PCR-amplified cDNA. From the 5′ end of the cDNA fragment, the open reading frame continued upstream where three in-frame methionine codons were found. A stop codon was found 37 nt further upstream of the first methionine. The sequence around the first ATG perfectly matched with the Kozak consensus sequence. If this ATG represents the translation initiation codon, the size of the transcript should be more then 1.1 kb and it should encode 243 amino acids.

Fig. 3B shows a comparison of the amino acid sequence of the MesP1 basic HLH region with those of other basic HLH proteins. A strong similarity was observed in the helix II region. The most striking similarity in helix II was observed with mouse Twist and recently isolated paraxis (meso-1) (Burgess et al., 1995; Blanar et al., 1995), which are also expressed in the early mesoderm lineage. However, the similarity of the amino acid sequences of MesP1 within this region is not high enough to be classified in the same subfamily. Accordingly, it appears that MesP1 belongs a novel basic helix-loop-helix subfamily (see Discussion). The protein sequence outside the bHLH domain showed no remarkable homology with other proteins reported.

Another isoform of MesP1 expressed in adult testes

To determine the expression of MesP1 in ontogeny, we performed a northern blot analysis using RNA isolated from

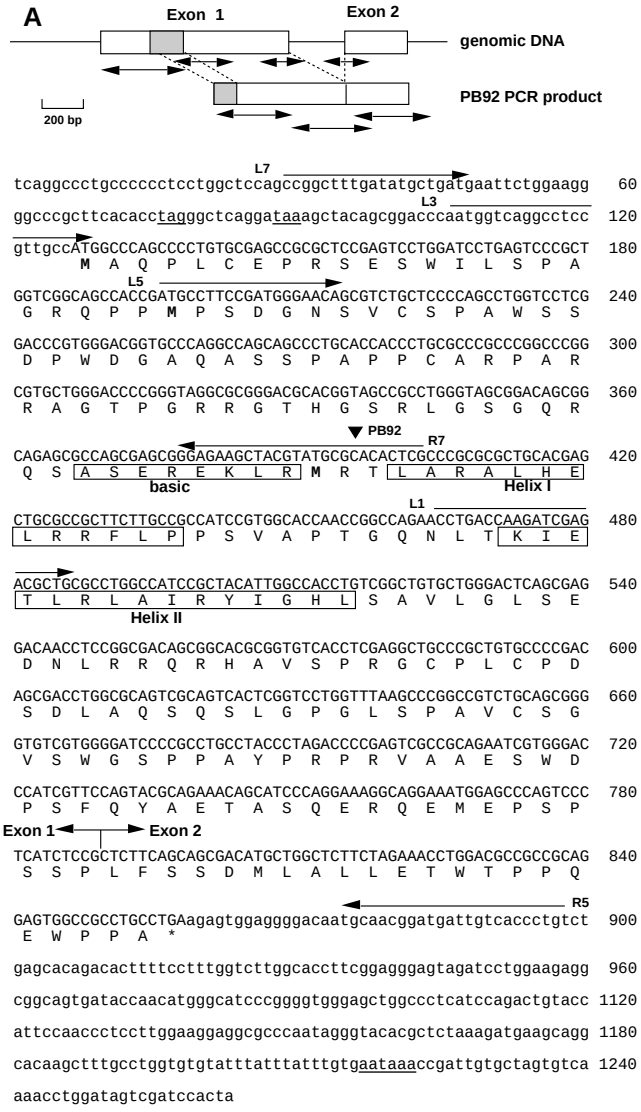

Fig. 3. (A) Genomic organization of *Mesp1*, and its nucleotide and deduced amino acid sequences. Sequenced segments are indicated by arrows. The sequence of intron 1 is not shown. The 5′ end of the original PCR product from embryos is indicated as ▼ PB92. The amino acids corresponding to the basic helix-loop-helix motif are boxed. Three methionine residues that may serve as translation initiation codons are indicated in boldface. Two in-frame stop codons in the 5′ upstream region are underlined. The possible polyadenylation signal in the 3′ untranslated region is also underlined. (B) The bHLH motif in MesP1 is compared with those of other bHLH proteins, including M-twist, meso-1, HEN1 (Brown et al., 1992) and MyoD. Shaded amino acids indicate residues that match the consensus (cons) derived from the known bHLH family (Cai and Davis, 1990). $=L, I, V or M; &=F, L, I or Y.

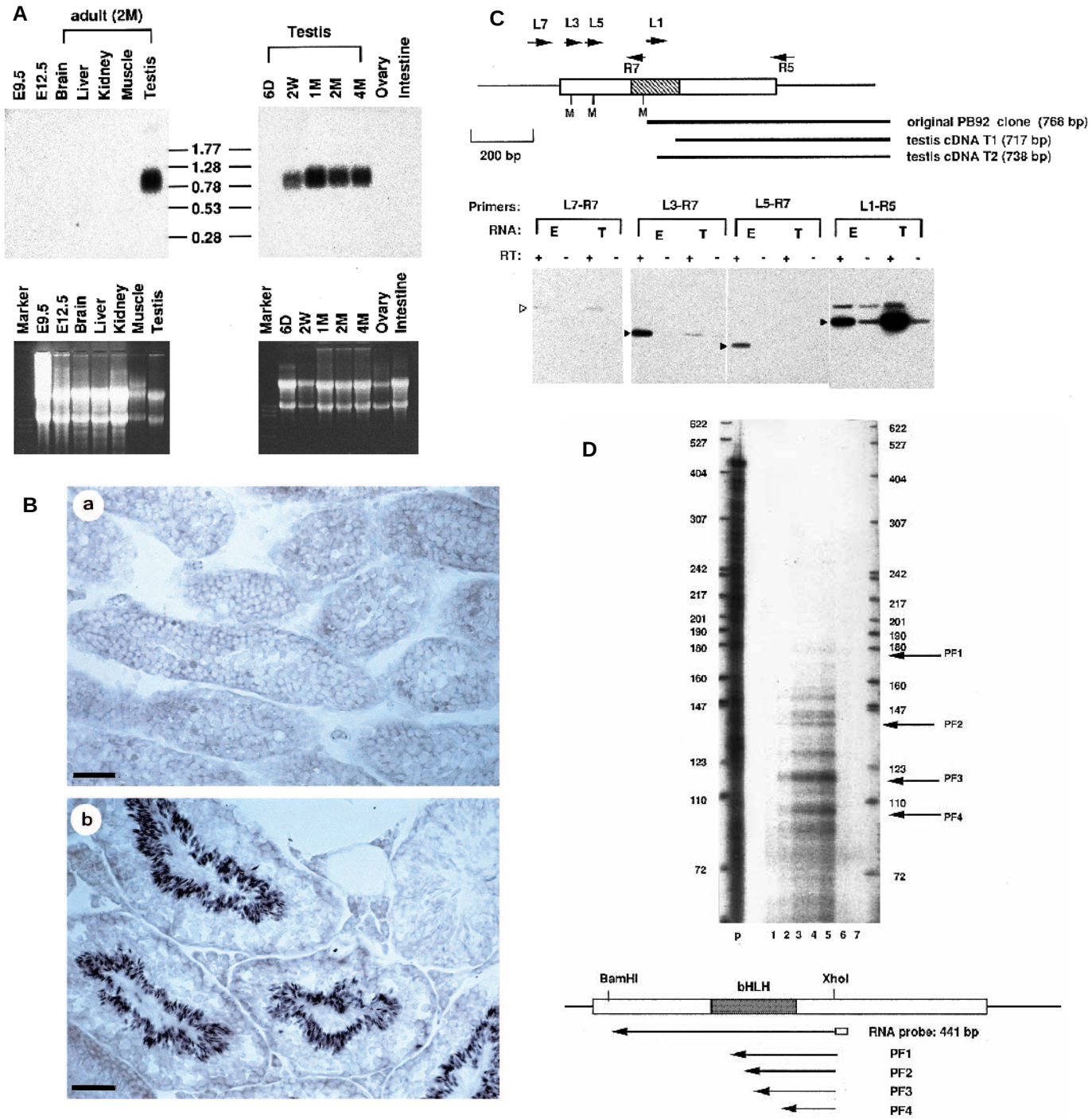

Fig. 4. (A) MesP1 expression in the testis demonstrated by northern analysis. Total RNA (20 μg) prepared from 9-12.5 dpc embryos and adult tissues was analysed. RNA size markers were used to estimate the transcript length. Before transfer, gels were stained with ethidium bromide to determine the quality and quantity of each RNA preparation (lower panel). (B) MesP1 expression in mature sperm observed by in situ hybridization. A strong and specific signal was detected from the 3-month-old mature testis (b) but not from 2-week-old inmature testis (a). Bar indicates 50 μm. (C) RT-PCR analysis of different MesP1 mRNA isoforms expressed in the early embryo and testis. The coding region is shown as a box where its bHLH region is hatched. The precise location and sequence of each primer is shown in Fig. 3A. The MesP1 cDNA originally isolated by PCR (clone PB92) and that from a testis cDNA library (T1 and T2) are shown as solid lines with their lengths. After PCR, the products were separated on an agarose gel and transferred to a nylon membrane. A DNA fragment spanning from L7 to R5 was used as the hybridizing probe. The PCR bands representing expected sizes are indicate by arrowheads. Abbreviations: E, embryo RNA; T, testis RNA; RT, (with or without) reverse transcription. (D) An RNase mapping result representing the multiple transcriptional initiation sites for the MesP1 mRNA in the testis. The RNA probe was hybridized with the following RNA samples: (1) 20 μg yeast tRNA, (2-5) testis total RNA; 2 μg (2), 5 μg (3), 10 μg (4), 20 μg (5), (6) 20 μg liver total RNA (7) 20 μg kidney total RNA. Lane (p) indicates the undigested RNA probe. The sizes of the probe and the protected fragments are shown under the schematic representation of MesP1 cDNA. The box represents the putative coding region with the bHLH region shaded.

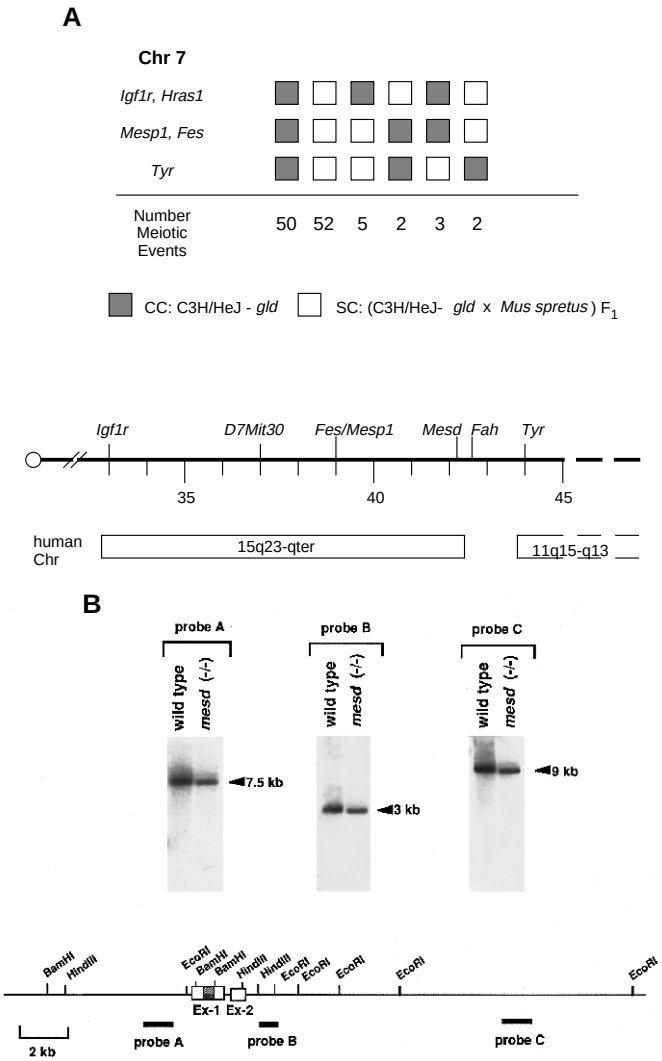

Fig. 5. (A) Segregation of *Mesp1* and neighboring markers on mouse chromosome 7 in (C3H/HeJ-*gld* × *Mus.spretus*) F_1 × C3H/HeJ-*gld* interspecific backcross mice. Shaded boxes represent the homozygous C3H pattern (CC) and open boxes represent the F_1 pattern (SC). The informative RFLVs for *Mesp1* were defined using restriction endonuclease *Taq*I. The C3H allele showed a band of 1.2 kb whereas that of *Mus. spretus* was 7.5 kb. The RFLVs and segregation of *Igf1r, Fes and Tyr* in this cross have been described previously (Saunders and Seldin, 1990). The number of offspring carrying each type of chromosome is listed at the bottom of each column. A partial chromosome 7 linkage map showing the location of *Mesp1* in relation to linked gene is shown below. Recombination distances between loci are shown in centimorgans to the left of the chromosome. Corresponding map for the human chromosome is also shown (Brilliant et al., 1994). (B) Southern hybridizations showing that *Mesp1* is not deleted in *mesd* (–/–) mutant ES cell DNA. DNAs were digested either with *Hin*dIII (for probe A) or *Eco*RI (for probe B and C). A restriction map of *Mesp1* and its flanking region is illustrated on the bottom. The probes used for the Southern hybridizations are also shown: A, 5′ flanking; B, transcribed region; C, 3′ flanking.

several stages of embryos and adult tissues. As shown in Fig. 4A, no expression of MesP1 was detected at any stages of embryos after 8.5 dpc or any adult tissues except the testis. The MesP1 expression in the testis was regulated developmentally. Its mRNA became visible 2 weeks after birth, and increased thereafter, reaching the full level expression in mature testes. As shown by in situ hybridization (Fig. 4B), only mature sperms expressed strong and specific MesP1 transcripts as expected.

The estimated length of the testis mRNA by northern analysis was only 0.8 kb, whereas the transcript length estimated by the genomic sequence analysis was more than 1.1 kb (Fig. 3A). In order to investigate the possibility that the MesP1 mRNA expressed in the early mesoderm is an isoform different from that expressed in the testis, we performed RT-PCR and compared their transcripts. Various combinations of forward and reverse primers were used on cDNA prepared from the embryo and testis. As shown in Fig. 4C, with primers L3 and L5 paired with R7, strong cDNA bands were amplified from the embryo, although faint bands were visible for the testis cDNA. With downstream primers L1 and R5 used as a control pair, however, transcripts derived from the testis showed a much stronger amplification. Therefore, the major transcripts expressed in the testis lacked the upstream sequences expressed in the embryonic mesoderm. In order to confirm this result, and to possibly determine the transcription initiation site in the testis, an RNase mapping experiment was performed. As shown in Fig. 4D, multiple protected bands were observed only for the RNA derived from the testis, but not from the liver or kidney where no MesP1 was expressed. Moreover, almost all estimated initiation sites were mapped within the bHLH region. This result strongly suggests that the transcripts expressed in the testis do not encode any functional proteins as the bHLH protein.

Because no significant amplification was detected with the 5′-most primer L7 from either embryo or testis transcripts compared with L3 primed amplification, it is likely that the transcription initiation site for the embryonic MesP1 isoform is located between L7 and L3 (about 60 nt), although we cannot rule out the possibility that a minor transcription starts from a more upstream region both in the embryo and testis as suggested by the faint bands.

Chromosomal localization of *Mesp1*

To determine if *Mesp1* maps near any known developmental mutations, particularly those with phenotypes consistent with its expression pattern, we mapped *Mesp1* using a panel of DNA samples from an interspecific backcross that has been characterized for over 1000 genetic markers throughout the genome. The genetic markers included in this map span between 50 and 85 centi-Morgans (cM) on each mouse autosome and chromosome (Chr) X (for example, see Watson et al., 1992 or view entire cross at the following internet address: http://www/informatics.jax.org/usr/seldin). DNAs from two parental mice C3H/HeJ-*gld* and (C3H/HeJ-*gld* × *Mus spretus*)F_1 were digested with various restriction endonucleases and hybridized with a *Mesp1* cDNA probe to determine RFLVs and to allow haplotype analyses. Informative Taq1 RFLVs were detected: C3H/HeJ-*gld*, 1.2 kb; *Mus spretus*, 7.5 kb. In each of the backcross mice, either the C3H/HeJ-*gld* parental band or both bands (*Mus spretus* and a half-intensity C3H/HeJ-*gld* bands) were observed, indicating a single locus.

Comparison of the haplotype distribution of the RFLV allowed *Mesp1* to be mapped to a specific region of mouse Chr 7 with respect to the reference loci (Fig. 5A). The best gene

order (Bishop, 1985) ± the standard error (Green, 1981) indicated that *Mesp1* was located 6.1±2.2 cM distal to the insulin-like growth factor1 receptor gene (*Igf1r*) and 4.4±1.9 cM proximal to the tyrosinase gene (*Tyr*). *Mesp1* cosegregated with the feline sarcoma oncogene (*Fes*) among 114 meiotic events analyzed in this panel. This region of the mouse Chr 7 is homologous to two regions in the human genome. While the human homologues of *Igf1r* and *Fes* map to 15q25-qter and 15q26.1, respectively, the human tyrosinase gene (*TYR*) maps to 11q21. Accordingly, it is very likely that the human homologue of *Mesp1* maps to 15q26 if linked to *FES*, although a remote possibility remains that it is located in 11q. There are some morphogenetic mouse mutations in the vicinity of the *Mesp1* locus (Holdener-Kenny et al., 1992). Among them, mesoderm deficiency (*mesd*) (Holdener et al., 1994) appeared to be the most likely candidate for a *Mesp1* mutation. *mesd* is one of loci identified as a result of extensive genetic complimentation studies between radiation-induced deletions removing the albino gene. Mouse embryos lacking *mesd* do not produce mesoderm but have well-defined extraembryonic and thickened embryonic ectoderm. To investigate the possibility that *Mesp1* is included in the deleted region of *mesd*, DNA prepared from an *mesd* ES cell line (Ail-6) established from one of the blastocysts obtained by an intercross of c^{1DthWb}(Holdener et al., 1994) was analyzed. As shown in Fig. 5B, Southern hybridization results clearly showed a band for the *Mesp1* coding region in *mesd* (–/–) mutant DNA. To further exclude a possibility that the regulatory region of *Mesp1* might be affected in the *mesd* (–/–) mutant by deletion, 7 kb upstream and 15 kb downstream regions were analyzed by Southern hybridization. Again, the results showed no obvious deletions within these regions as expected for *mesd*. Accordingly, we conclude that lacking *Mesp1* is not likely to be responsible for the *mesd* phenotype caused by deletions in the albino(*c*) locus.

DISCUSSION

Differentiation of the mesoderm is one of the critical events in the establishment of the vertebrate body plan. In the mouse, mesodermal tissues are formed from the embryonic ectoderm cells that ingress through the primitive streak. *Mesp1* is particularly interesting because it is the first gene reported to be expressed almost exclusively in the early nascent mesoderm.

Our initial aim was to isolate genes responsible for formation and development of the primordial germ cells (PGC). PGC appear in ontogeny just after gastrulation, at the base of the allantois, and are derived from mesodermal cells in the most posterior primitive streak region. We employed a regional subtractive hybridization strategy taking the advantage of the specific localization of the PGC. By the subsequent dot blot screening, we eliminated all non-specific clones expressed in other regions. This screening appears to be sensitive enough to distinguish the levels of expression. Most candidate clones selected by dot blot analysis showed no staining by whole-mount in situ hybridization. It is possible that the levels of expression were below the detection limit by whole-mount in situ hybridization. Interestingly, the expression of one of the resulting cDNA clones, encoding a protein MesP1, was detected at the base of the allantois in the 7.5 dpc embryo. This result indicates that our method is practically useful in isolating stage- and region-specific cDNAs.

The in situ hybridization study indicates that *Mesp1* is expressed in the mesodermal cells at the base of the allantois where PGC are localized at 7.5 dpc. However, the earliest expression of *Mesp1* (at 6.5-6.75 dpc) was not restricted to the PGC population, but apparently all nascent mesodermal cells expressed this gene. Thereafter, *Mesp1* expression did not spread over all mesodermal cell; it was seen only in a fraction of cells, particularly those in the early emergence group destined to the extraembryonic mesoderm and the most posterior part of the tailbud. The expression remained for a while at the base of the allantois, and then spread out to the lateral margins of the tail bud mesoderm. This pattern is similar to that of the initial migration of PGC. Therefore, the expression observed at the lateral parts of the tailbud might reflect the cell migration from the base of the allantois, and possibly of the PGC themselves. Shortly thereafter, however, *Mesp1* expression disappeared quickly as the PGC started to migrate into the endoderm of the hind gut. Thus, it is possible that the *Mesp1* expression is associated with the development of the presumptive PGC population.

No *Mesp1* expression was observed in later development or adult tissues except the testis, although our analysis did not cover all tissues or stages of development. Its expression in the testis was age and maturation dependent; starting 2 weeks after birth and reaching the maximum at about 1 month. Although the expression was strong and specific to the sperm, it is unlikely to encode a functional protein because it lacks the 5′ region, which is transcribed in the embryo and encodes a functional basic HLH domain. We screened a testis cDNA library and isolated several clones that hybridized to the *Mesp1* cDNA clone PB92. The testis sequences were identical to that of PB92 in the 3′ region and we did not find any novel sequences suggesting an alternative splicing in the 5′ region. We also tried 5′RACE (rapid amplification of the cDNA end). However, the 5′ end was not extended beyond that of cDNA clones isolated from the cDNA library. These results, including the northern data, which revealed a shorter size for the testis transcript (Fig. 4A), and RNase mapping data, indicating multiple initiation sites within the bHLH region (Fig. 4D), are consistent with a gene structure composed of different promoters functioning in the embryo from the testis. Such a possibility is currently investigated using transgenic mice. We are also determining whether any MesP1 protein is expressed in the testis or not.

Tissue-specific HLH proteins play important roles in cell lineage determination and differentiation, including the neuronal, lymphoid and mesodermal lineages. Especially, functions of bHLH proteins such as the MyoD family in the mesoderm lineage are well investigated and their regulatory network has been established by the studies of knockout mice (reviewed in Olson and Klein, 1994). In addition, using a yeast two-hybrid system, several cell-type-specific bHLH proteins that dimerize selectively with the ubiquitous class bHLH proteins were isolated. They include paraxis, scleraxis (Cserjesi et al., 1995a) and dermo-1 (Li et al., 1995), which appear to be associated with the differentiation of the somitic mesodermal cell lineages such as myotome, sclerotome and dermatome. In contrast, the expression of MesP1 started earlier than any other bHLH proteins expressed in the embryonic mesodermal lineage (cf. eHAND/Hxt, MASH-2 in trophoblast: Cross et al., 1995;

Cserjesi et al., 1995b; Guillemot et al., 1994). Its transient expression and down-regulation before somitogenesis suggest that MesP1 may help determine non-somitic early mesodermal lineages including the extraembryonic mesoderms.

Mesp1 appears to be one of the earliest genes responsive to the signal for the early nascent mesoderm formation. In *Xenopus*, various members of the fibroblast growth factor (FGF) and transforming growth factor-β (TGF-β) families have been demonstrated to play important roles in mesoderm induction and patterning (reviewed in Kessler and Melton, 1994). Also in the mouse, several signaling molecules of the TGF-β and FGF families have been implicated in the early gastrulation stage leading to the mesoderm formation; e.g., nodal (Conlon et al., 1994), BMP-4 (Winner et al., 1995), and FGF-3 and 4 (Niswander and Martin, 1992). The functional analysis of several mutant mice generated by either insertional mutagenesis (413.d) (Zhou et al., 1993), radiation-induced deletion (*mesd*) and disruption by homologous recombination (FGFr-1, BMP-4, Brachyury etc.) (Yamaguchi et al., 1994; Deng et al., 1994; Beddington et al., 1992) demonstrated that these molecules are involved in the early mesoderm formation and the subsequent mesoderm patterning. The most striking phenotype in the mesoderm formation was reported for the *mesd* mutants, which produce no mesodermal cells either in vivo or in vitro (Holdener et al., 1994). Because the gene mapping data strongly suggested a close linkage between *mesd* and *Mesp1*, we investigated their possible identity. However, the *mesd* deletion did not overlap with the *Mesp1* gene including the 6 kb upstream and 15 kb downstream regions of the *Mesp1* coding sequence. A remote possibility remains, however, that the regulatory region of *Mesp1* localized further away from its coding region may be deleted in the *mesd* deletion.

Although our strategy was designed to isolate the cDNA clones involved in the PGC development, *Mesp1* is more likely to be involved in early mesoderm formation. Its relationship with the PGC development remains to be investigated further. A novel subfamily of bHLH proteins can be defined that is characterized by an amino acid similarity within the bHLH domain. In this connection, we recently isolated by genomic DNA screening, another gene *Mesp2*, encoding a protein MesP2, which shares 92% identity with MesP1 in the bHLH domain. A preliminary study indicated an expression pattern only partially overlapping with that of MesP1. However, it is also expressed in the mesodermal cells at a slightly later stage than MesP1 (data not shown). We are constructing both *Mesp1* and *Mesp2* knockout mutants by homologous recombination and hope to unravel the functions of these genes in early embryonic development including the PGC differentiation.

We thank Dr Nishina for providing a testis cDNA library, Dr Abe for the valuable suggestion for the lone linker-aided amplified cDNA synthesis and Ms Rochelle for an excellent technical assistance. This work was supported in part by NIH Grant HG00734. This research was also supported partly by TARA project (University of Tsukuba, Japan).

REFERENCES

Abe, K. (1992). Rapid isolation of desired sequences from lone linker PCR amplified cDNA mixtures: Application to identification and recovery of expressed sequences in cloned genomic DNA. *Mammalian Genome* **2,** 252-259.

Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. and Hogan, B. L. (1994). Embryonic expression of Lim-1, the mouse homolog of Xenopus Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev. Biol.* **161,** 168-178.

Batten, B. E. and Haar, J. L. (1979). Fine structural differentiation of germ layers in the mouse at the time of mesoderm formation. *Anat. Rec.* **194,** 125-142.

Beddington, R. S. P., Rashbass, P. and Wilson, V. (1992). *Brachyury*-a gene affecting mouse gastrulation and early organogenesis. *Development* **1992 Supplement**, 157-165.

Benezra, R., Davis, R. L., Lookshon, D., Turner, D. L. and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61**, 49-59.

Bishop, D. T. (1985). The information content of phase-known matings for ordering genetic loci. *Genet. Epidemiol*. **2,** 349-361.

Blanar, M. A., Crossley, P. H. Peters, K. G., Steingrimasson, E., Copeland, N. G., Jenkins, N. A., Martin, G. R. and Rutter, W. J. (1995). Meso-1. a basic-helix-loop-helix protein involved in mammalian presomitic mesoderm development. *Proc, Natl. Acad. Sci. USA*, **92,** 5870-5874.

Brilliant, M. H., Williams, R. W., Conti, C. J., Angel, J. M., Oakey, R. J. and Holoderner, B. C. (1994) Mouse chromosome 7. *Mamm. Genome* **5**, S104-S123.

Brown, L., Espinosa III, R., Le Beau, M. M. and Siciliano, M. J. (1992). *HEN1* and *HEN2*: A subgroup of basic helix-loop-helix genes that are coexpressed in a human neuroblastoma. *Proc. Natl. Acad. Sci. USA*, **89,** 8492-8496.

Burgess, R., Cserjesi, P., Ligon, K., and Olson, E. N. (1995). Paraxis: A basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol*. **168,** 296-306.

Cai, M, and Davis, R. W. (1990). Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell* **61,** 437-446.

Conlon, F. L., Lyons, K. M. Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for *nodal* in the formation and maintenance of the primitive streak in the mouse. *Development* **120,** 1919-1928.

Cross, J. C., Flannery, M. L., Blanar, M. A., Steingrimsson, E., Jenkins, N. A., Copeland, N. G., Rutter, W. J. and Werb, Z. (1995). Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development* **121**, 2513-2523.

Cserjesi, P., Brown, D., Ligon, K. L., Lyons, G. E., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. and Olson, E. N. (1995a). Scleraxis:a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. *Development* **121,** 1099-1110.

Cserjesi, P., Brown, D., Lyons, G. E. and Olson, E. N. (1995b). Expression of the novel basic helix-loop-helix gene eHAND in neural crest derivatives and extraembryonic membranes during mouse development. . *Dev. Biol.* **170,** 664-678**.**

Deng, C.-X., Wynshaw-Boris, A., Shen, M. M., Daugherty, C., Ornitz, D. M. and Leder, P. (1994). Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes Develop.* **8,** 3045-3057.

Dixon, K. E. (1994). Evolutionary aspects of primordial germ cell formation. *In Germline development. Ciba Foundation Symposium* **182,** 92-120.

Ginsburg, M., Snow, M. H. L. and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* **110,** 521-528.

Green, E. L. (1981). Linkage, recombination and mapping. In *Genetics And Probability In Animal Breeding Experiments.* (ed. E. Green). New York: MacMillan Press.

Guillemot, F. Nagy, A. Auerbach, A. Rossant, J. and Joyner, L. A. (1994). Essential role of Mash-2 in extraembryonic development . *Nature* **371,** 333-336

Hashimoto K, and Nakatsuji, N. (1989). Formation of the primitive streak and mesoderm cells in mouse embryo-detailed scanning electron microscopical study. *Dev. Growth & Differ*. **31,** 209-218

Holdener, B. C., Faust, C., Rosenthal, N. S. and Magnuson T. (1994). *msd* is required for mesoderm induction in mice. *Development* **120,** 1335-1346.

Holdener-Kenny, B., Sharan, S. K. and Magnuson, T. (1992). Mouse albino-deletions:From genetics to genes in development. *BioEssays* **14,** 831-839.

Kessler, D. S. and Melton, D. A. (1994). Vertebrate embryonic induction: mesodermal and neural patterning. *Science* **266,** 596-604.

Ko, M. S. H. (1990). An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs. *Nucleic Acid Res.* **18,** 5705-5711.

Lawson, K. A. and Hage, W. (1994). Clonal analysis of the origin of

primordial germ cells in the mouse. In *Germline Development. Ciba Foundation Symposium* **182,** 68-91.

Lawson, K. A. and Pedersen R. A. (1992). Clonal analysis of cell fate during gastrulation and early neurulation in the mouse. In *Postimplantation Development in the Mouse. Ciba Foundation Symposium* **165,** 3-26.

Li, L., Cserjesi, P. and Olson, E. N. (1995) Dermo-1: A novel twist-related bHLH protein expressed in the developing dermis. *Dev. Biol.* **172,** 280-292.

MacGregor, G. R., Zambrowicz, B. P. and Soriano, P. (1995). Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* **121,** 1487-1496.

Niswander, L. and Martin, G. R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* **114,** 755-768.

Okayama, H. and Berg, P. (1983). A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell Biol.* **3**, 280-289.

Olson, E. N. and Klein, W. H. (1994). bHLH factors in muscle development: Dead lines and commitments, what to leave in and what to leave out. *Genes Develop.* **8,** 1-8.

Saunders, A. M. and Seldin, M. F. (1990). A molecular genetic linkage map of mouse chromosome 7. *Genomics* **8**, 525-535.

Takeda, H., Matsuzaki, T., Oki, T., Miyagawa, T. and Amanuma, H. (1994). A novel POU domain gene, zebrafish *pou2*: expression and roles of two alternatively spliced twin products in early development. *Genes Develop.* **8**, 45-59.

Tam, P. P. and Beddington, R. S. P. (1992). Establishment and organization of germ layers in the gastrulating mouse embryo. In *Postimplantation Development in the Mouse. Ciba Foundation Symposium* **165**, 27-49.

Tsukamoto, T., Kusakabe, M. and Saga Y. (1991). In situ hybridization with non-radioactive digoxigenin-11 UTP-labeled cRNA probes: Localization of developmentally regulated mouse tenascin mRNAs. *Int. J. Dev. Biol.* **35**:25-32.

Yamaguchi, T. P., Harpal, K., Henkemeyer, M., and Rossant, J. (1994). *fgfr-1* required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Develop.* **8,** 3032-3044.

Watson, M. L., D'Eustachio, P. D., Mock, B. A., Steinberg, A. D., Morse III, H. C., Oakey, R. J., Howard, T. A., Rochelle, J. M., and Seldin, M. F. (1992). A linkage map of mouse chromosome 1 using an interspecific cross segregating for the *gld* autoimmunity mutation. *Mamm. Genome* **2**, 349-361.

Winner, G., Blessing, M., Labosky, P. A. and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Develop.* **9,** 2105-2116.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B. L. M. and Kuehn, M. R. (1993). *Nodal* is a novel TGF-β-like gene expressed in the mouse node during gastrulation. *Nature* **361,** 543-547.

(Accepted 20 June 1996)