

Peri-Implantation Lethality in Mice Lacking the Sm Motif-Containing Protein Lsm4

EMILIO HIRSCH,^{1†} TOSHITAKA OOHASHI,¹ MARIANNE AHMAD,² STEFAN STAMM,³
AND REINHARD FÄSSLER^{1,2*}

Max Planck Institut für Biochemie¹ and Max Planck Institut für Neurobiologie,³ 82152 Martinsried, Germany, and
Department of Experimental Pathology, Lund University, 221 85 Lund, Sweden²

Received 1 October 1999/Accepted 13 October 1999

Small nuclear ribonucleoproteins (snRNPs) are particles present only in eukaryotic cells. They are involved in a large variety of RNA maturation processes, most notably in pre-mRNA splicing. Several of the proteins typically found in snRNPs contain a sequence signature, the Sm domain, conserved from yeast to mammals. By using a promoter trap strategy to target actively transcribed loci in murine embryonic stem cells, a new murine gene encoding an Sm motif-containing protein was identified. Database searches revealed that it is the mouse orthologue of Lsm4p, a protein found in yeast and human cells and putatively associated with U6 snRNA. Introduction of the *geo* reporter gene cassette under the control of the murine *Lsm4* (*mLsm4*) endogenous promoter showed that the gene was ubiquitously transcribed in embryonic and adult tissues. The insertion of the *geo* cassette disrupted the *mLsm4* allele, and homozygosity for the mutation led to a recessive embryonic lethal phenotype. *mLsm4*-null zygotes survived to the blastocyst stages, implanted into the uterus, but died shortly thereafter. The early death of *mLsm4p*-null mice suggests that the role of *mLsm4p* in splicing is essential and cannot be compensated by other Lsm proteins.

Spliceosomal U small nuclear ribonucleoproteins (snRNPs) of the Sm class are a group of RNA-protein complexes which are the major components of the spliceosome, the macromolecular complex that catalyzes the pre-mRNA splicing reaction (13, 17, 18, 22). Spliceosomal U snRNPs contain five different snRNAs, U1, U2, U4, U5, and U6. Each is found in different U snRNP particles which display specific functions in splicing. Excluding U6, all snRNAs share a 5'-terminal trimethylguanosine (m₃G) cap and an Sm site. This latter structure consists of a single-stranded U-rich sequence whose consensus is PuA(U_n)G_nPu, where $n > 3$. Proteins that bind to individual U snRNAs can be classified into specific and common proteins depending on whether they can associate with one snRNA or with all of them, respectively. At least eight common proteins, termed B, B', D1, D2, D3, E, F, and G, are known so far to associate with all U snRNPs with the exception of U6 snRNP. All these proteins are a major target for autoantibodies in the human disease systemic lupus erythematosus (26) and have a conserved motif named the Sm domain. Sequence comparison shows two Sm consensus sites, termed Sm1 and Sm2 (4, 14, 25), which are linked by a loop-folded spacer and thus form a single protein domain (15). Searches of the sequence databases indicate that the Sm motifs are highly conserved during evolution and are present in proteins (called Sm-like or Lsm) which have no clear counterpart among the eight components of the canonical Sm protein complex (23, 25).

The precise role of U snRNA binding Sm proteins in splicing reactions is not clear. The only known function of Sm proteins is their crucial role in the biogenesis of U snRNPs. During this process, the nuclear encoded 5'-terminal 7-monomethylguanosine-capped (m₇G) U snRNA is transiently transferred to the cytoplasm. Thereafter, the Sm proteins bind in a highly

ordered manner to the Sm site and form the Sm core domain. The recent resolution of the crystal structure of two Sm protein complexes suggests that seven snRNP common proteins might assemble in a ring shape that encloses the snRNA in the center (15). The correct assembly of the core domain is a prerequisite for the subsequent trimethylation of the m₇G cap of the U snRNA. Due to the presence of two nuclear localization signals in the m₃G cap of the RNA and in the Sm core domain (9), the assembled U snRNP is imported to the nucleus, where it exerts its function in pre-mRNA maturation.

A central unit of the nuclear pre-mRNA splicing machinery is the 25S [U4/U6 · U5] tri-snRNP. This complex, by joining the U1 and U2 snRNP-containing prespliceosome, completes the assembly of the spliceosome and drives U6 in a process involving structural rearrangements needed for the splicing reaction. U6 differs from the other snRNAs because it contains a γ -monomethyl cap and lacks an Sm site (3). Whereas U6 does not complex with the eight common Sm proteins, it is well established that three yeast Sm-like proteins, Lsm4p/Uss1p (4), Lsm3p (formerly denoted Smx4p) (25), and Lsm8p (21) can associate with U6 as well as U6-containing particles. These data have recently been extended by the finding that at least four other Sm-like proteins (Lsm2p and Lsm5 to Lsm7p) can associate with the U6 snRNP (12, 19, 23). The emerging scenario suggests a reciprocal interaction of these proteins with one another to form a U6-containing complex(es) (19, 23). It is also becoming evident that these proteins are required for the stability and maintenance of normal U6 snRNA levels (19, 23) and that their depletion blocks pre-mRNA splicing (19). In particular, genetic ablation of the yeast Lsm4p reveals that it is essential for pre-mRNA splicing and that its *in vivo* depletion causes the arrest of cell growth (4). Recent results indicate that *Lsm* genes are a highly conserved: Lsm4p, for example, is also present in humans, where it appears to retain identical functional properties of the yeast orthologue (23).

Here we show the molecular and developmental analysis of a mouse mutant in which a gene trap vector has inserted in the murine orthologue of yeast *Lsm4*. The mutation completely

* Corresponding author. Mailing address: Department of Experimental Pathology, Lund University, S-22185 Lund, Sweden. Phone: 46-46-173-553. Fax: 46-46-158-202. E-mail: reinhard.fassler@pat.lu.se.

† Present address: Dipartimento di Genetica, Biologia e Biochimica, Università di Torino, Turin, Italy.

eliminates the expression of murine Lsm4p (mLsm4p) and leads to a peri-implantation lethal phenotype.

MATERIALS AND METHODS

Gene trap strategy. Construction of the targeting vector and isolation of electroporated R1 embryonic stem (ES) cell clones have been described previously (7). Colonies were screened individually by cleaving genomic DNA with *Bam*HI and probing Southern blots with an internal probe present in the targeting vector. This probe consisted of an *Eco*RI-*Bst*EII 300-bp fragment containing sequences from intron I of the gene coding for β_1 -integrin and allows us to test the number of random integrations per ES cell clone. ES cells with a single-copy integration and random integration were differentiated in vitro and stained for β -galactosidase activity as described previously (2). Briefly, ES cells were trypsinized and resuspended at 3×10^4 cells/ml in differentiation medium. Several drops (20 μ l, corresponding to 600 cells) were put onto the lid of a bacterial petri dish filled with phosphate-buffered saline (PBS). After inversion of the lid, the cells were incubated in hanging drops at 37°C under 5% CO₂ for 2 days. This culture method leads to ES cell aggregates (embryoid bodies [EBs]), where cells start to differentiate into derivatives of all three germ layers (2). Then EBs were transferred to fresh bacteriological dishes and incubated in differentiation medium for a further 8 days, resulting in a total incubation period of 10 days in suspension. Finally, EBs were plated on gelatin-coated glass coverslips, where they were allowed to adhere, incubated for a further 10 days, fixed in 4% paraformaldehyde, stained for β -galactosidase activity overnight (6), covered with another coverslip, and analyzed under an Axiophot microscope (Zeiss).

Cloning of G101 cDNA and the G101 gene. Total RNA was prepared from the ES cell clone G101 by using RNeasy columns (Qiagen, Hilden, Germany). The G101 cDNA was isolated by using 5' and 3' rapid amplification of cDNA ends (RACE). The 5' RACE product was generated with GIBCO-BRL kit 18374-025 as specified by the manufacturer. Briefly, cDNA was synthesized with an oligonucleotide based on the β -galactosidase 5' sequence (5'-AGTAAACAACCCGT CGGATTC-3'). A first-round PCR and a nested PCR were performed with Expand High Fidelity thermostable DNA polymerase mixture (Boehringer, Mannheim, Germany). The primers for the nested PCR were derived from upstream sequences of the β -galactosidase gene (5'-GGAACAAACGGCGGA TTGAC-3' and 5'-TGGGATAGGTTACGTTGGTG-3', respectively). The 3' RACE product was generated with GIBCO-BRL kit 18373-019, and the primers were derived from the 5' extended sequence of the β -galactosidase fusion cDNA (5'-GTGGGAGCGCGTGTGTCTGTGCCG-3' and 5'-GTGCCGCGCGGA AGTTATCCC-3', respectively). The 3' RACE was performed with a mixture of Expand Long Template and High Fidelity thermostable DNA polymerases (Boehringer).

The genomic PCR was carried out with the Expand Long Template PCR system (Boehringer) with one primer pair (5'-GGAACAAACGGCGGATTGA C-3' and 5'-AGTAAACAACCCGTGCGGATTC-3') to detect the G101 allele and another primer pair (5'-GGAACAAACGGCGGATTGAC-3' and 5'-GCTGTC TTCAGCAGCGACAAGGG-3') to detect the wild-type mLsm4 allele. All amplified fragments were cloned in the pCRII TA vector (Invitrogen, San Diego, Calif.) and sequenced with the ABI Prism Dye Terminator kit (Applied Biosystems, Foster City, Calif.). The sequences were analyzed with a 373A automatic sequencer (Applied Biosystems). Sequence alignments were performed with the BLAST program (1) and the CLUSTALW package (27).

Cellular localization of an EGFP-mLsm4 fusion protein in transfected cells. The full-length mLsm4 cDNA was cloned into plasmid pEGFP-C1 (Clontech, Palo Alto, Calif.) in frame with the C terminus of the enhanced green fluorescence protein (EGFP). Lipofectamine (GIBCO-BRL, Karlsruhe, Germany) was used to transiently transfect the expression construct into either NIH 3T3, COS7, or HEK293 cells plated on Lab-Tek chambers (Nunc, Wiesbaden, Germany). Control cells were transfected with a wild-type pEGFP-C1 vector.

Cell nuclei were immunolabeled with a mouse anti-3-methylguanosine monoclonal antibody (Oncogene Science). First, cells were fixed in PBS containing 2% paraformaldehyde, washed in PBS, permeabilized for 15 min in PBS containing 1% Triton X-100, washed in PBS, and incubated for 1 h at room temperature with primary antibody (1:200 dilution). Then the cells were labeled with a rabbit anti-mouse CY3 antibody (Jackson ImmunoResearch Laboratories, West Grove, Pa.; dilution 1:1,000), covered with a coverslip, and analyzed under a Leitz confocal laser-scanning microscope.

Western blots of transfected cell lysates were analyzed with a polyclonal EGFP antibody (Clontech) as previously described (20).

Generation and analysis of G101 mutant mice. ES cells from clone G101 were expanded and injected into C57BL/6 blastocysts as described previously (6). Chimeric mice were mated with C57BL/6 females to test for germ line transmission or with 129sv females to obtain inbred lines.

Organs from 8-week-old G101 heterozygous mice, implantation chambers at 6.5 days postcoitum (E6.5), and embryos at E7, E8.5, and E10.5 were dissected and fixed for 2 h at 4°C in PBS containing 4% paraformaldehyde. Embryos were washed in PBS and stained for β -galactosidase expression by published procedures (6). For light microscopic examination, pieces of tissue were frozen in dry ice-cold isopentane and cut with a Leitz cryostat. Sections (6 μ m thick) were collected on glass slides (Shandon, Frankfurt, Germany), stained with hematox-

ylin-eosin, and analyzed under a Zeiss Axiopod microscope. Whole-mount preparations of embryos were photographed under an Olympus stereomicroscope.

Genotypes of adult mice derived from heterozygous crosses were determined by Southern blot analysis. Genomic DNA was digested with *Eco*RI, gel separated, blotted, and probed with a 1.4-kb *Bam*HI fragment derived from intron 1 of the mLsm4 wild-type locus.

Northern blot analysis was performed on 15 μ g of kidney poly(A)⁺ mRNA purified on oligo(dT) spin columns (Pharmacia, Uppsala, Sweden) and hybridized with the whole mLsm4 cDNA. The blots were reprobated with an α -actin cDNA fragment (7).

For the genotyping of blastocysts derived from heterozygous crossings, uteri of pregnant females were flushed 3.5 days postcoitum. Single blastocysts were lysed for 30 min at 60°C in 10 μ l of 1 \times PCR buffer (Boehringer) containing 1 mg of proteinase K (Sigma, Deisenhofer, Germany) per ml. Following proteinase heat inactivation, PCRs were carried out in the presence of three primers, which allowed us to distinguish between wild-type (with primers 1 and 2, which are derived from sequences located in wild-type mLsm4 intron 1 and exon 2, respectively, and hence are around the insertion site of the β -galactosidase-containing targeting vector) and mutant (primers 1 and 3; the latter is located in the β -galactosidase gene of the targeting vector) strains. The primers (manufactured by GIBCO-BRL) had the following sequences: primer 1, 5'-GGCTCACAT CTGTAGAATGGG-3'; primer 2, 5'-TCTGACTACCCACGAATG-3'; and primer 3, 5'-GCTGTCTTCAGCAGCGACAAGGG-3'.

Nucleotide sequence accession number. The gene sequence of the Lsm4 DNA was submitted to the EMBL data bank under accession no. AJ249439.

RESULTS

Generation of mutant ES cells using a promoterless β -galactosidase-neomycin gene. The gene trap-type targeting vector was constructed to inactivate the β_1 -integrin gene (7). Briefly, a promoterless β -galactosidase-neomycin (*geo*) gene was cloned in frame with the start codon of the β_1 -integrin and electroporated into ES cells. Neomycin-resistant colonies were screened by a Southern assay with an internal probe corresponding to sequences present in intron 1 of the β_1 -integrin gene. Of 104 G418-resistant ES cell clones analyzed, 46 (44%) showed no recombination in the β_1 -integrin gene (7) and thus had randomly integrated into active sites of the ES cell genome. To confirm the copy number of integrated targeting vectors, all the clones were analyzed by a Southern assay with a β -galactosidase cDNA probe (results not shown). ES cell lines showing the integration of one copy were expanded, differentiated in vitro into EBs, and tested for β -galactosidase expression. In EBs derived from ES cell clone G101, all the cells expressed β -galactosidase (results not shown). This finding suggested that the targeting vector had disrupted a house-keeping gene which may be important for mouse development. Therefore, ES cell clone G101 was chosen for further investigation.

Cloning of G101 cDNA. By using 5' RACE from the known β -galactosidase sequence, the 5' untranslated region of the targeted gene was cloned. In accordance with the name of the ES cell clone, the gene and cDNA is called G101. Sequencing of the 5' RACE product indicated a fusion of the reporter gene to a 59-bp unknown sequence, containing an ATG codon spliced in frame to the ATG of the β -galactosidase-neomycin cDNA. To obtain the full-length cDNA, the sequence that extended the β -galactosidase was used to design primers for a 3' RACE with a proofreading thermostable DNA polymerase. The amplified product was 791 bp with an open reading frame of 411 bp coding for a protein of 137 amino acids. The protein has a theoretical molecular mass of 15 kDa and an isoelectric point of pH 10.85.

A homology search in the protein data bank revealed high homology (92% identity) to the human Lsm4 protein (23). Sequence comparison indicates that the gene is evolutionarily conserved through distant species, such as *Schizosaccharomyces pombe*, *Nicotiana tabacum*, and *Caenorhabditis elegans* (Fig. 1A). Alignment of the six genes displaying the highest similarity indicates that the homology extends beyond the Sm sites,

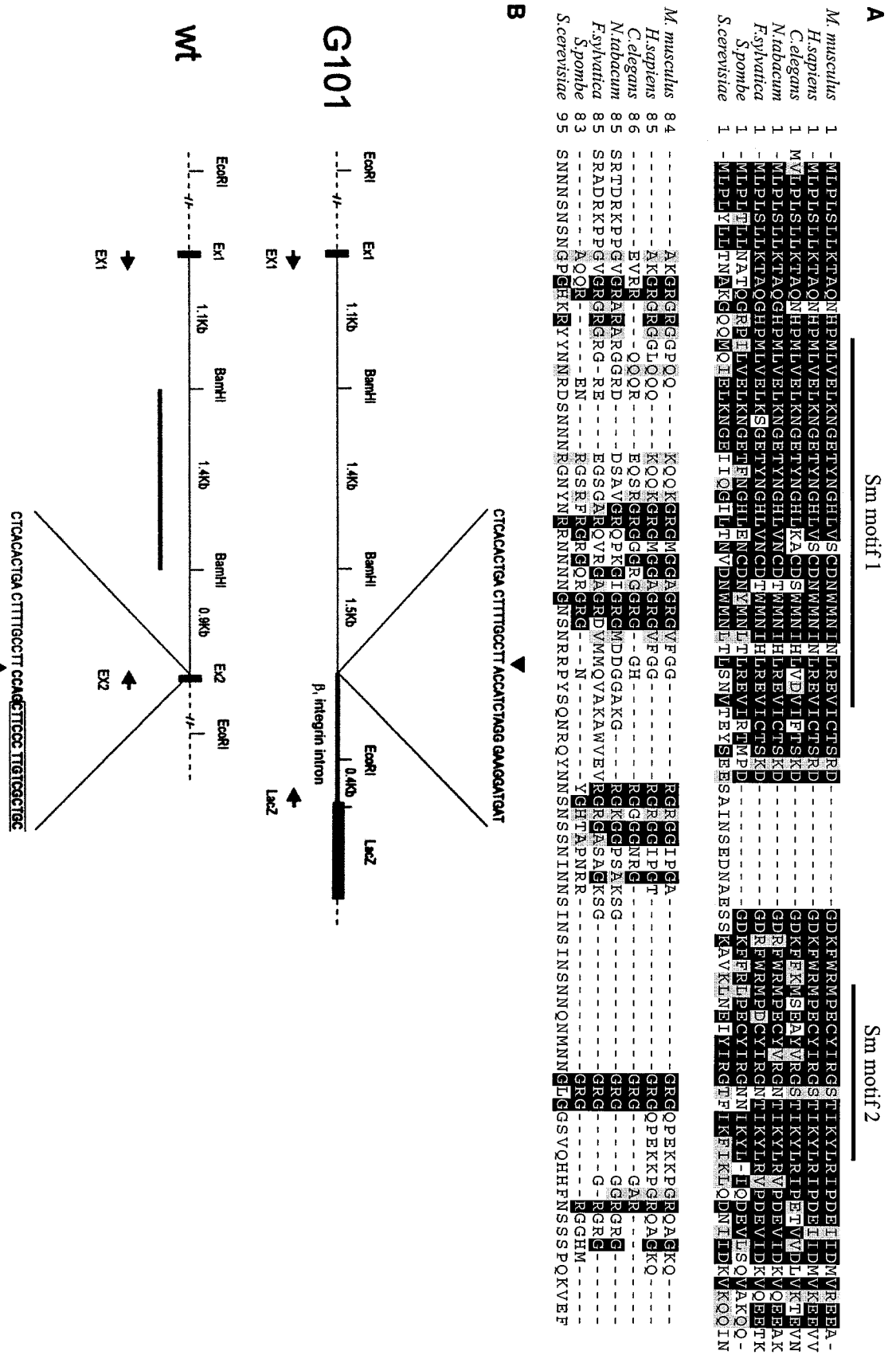


FIG. 1. Identification of the mLsm4 cDNA and characterization of the gene-trapped genomic locus. (A) Sequence alignment of Lsm4p-like proteins. Putative homologues of Lsm4p were identified by BLAST searches (<http://www.ncbi.nlm.nih.gov/egibm/BLAST/nph-blast?form=0>) and aligned with CLUSTALW program (27). Identities and similarities are highlighted by using BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html). The positions of Sm motifs 1 and 2 are indicated. White on black represents amino acid identity in the majority of the seven sequences; black on grey represents conservation of the nature of the amino acid at that site. Accepted groupings were M = I = V = L, K = R = H, F = Y = W, S = T, E = D, A = G, and Q = N. The accession numbers of the proteins are as follows: *Homo sapiens*, CAB45867; *C. elegans*, S54169; *N. tabacum*, S54169; *Faus sylvatica*, CAA10233; *S. pombe*, CAB10801; *S. cerevisiae*, P40070. (B) Genomic structure of the G101 locus and its wild-type (wt) counterpart. The black triangle indicates the targeting-vector insertion site. Sequences downstream of the integration site corresponding to β -integrin and wild-type loci are shown in bold. The 5' end of the wild-type mLsm4 exon 2 is boxed. Arrows indicate the position of the LacZ primer used for 5' RACE and the position of the Ex1 primer used for 3' RACE. The black bar shows the probe used for Southern blot genotyping of mouse tail DNA.

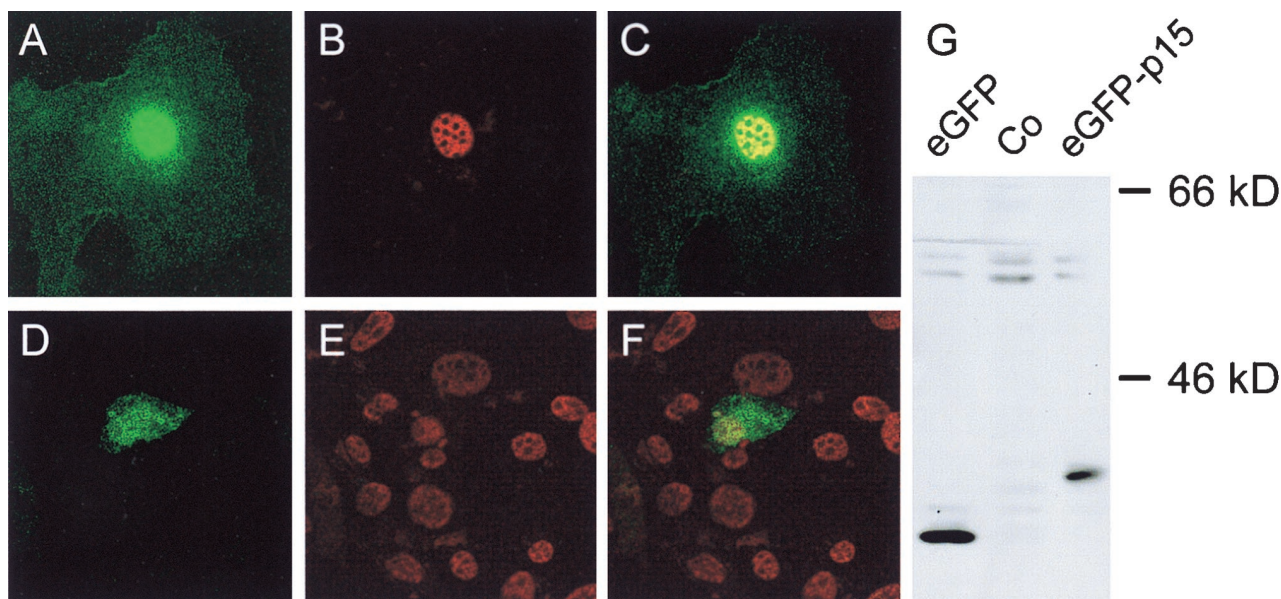


FIG. 2. Cellular distribution of the EGFP-mLsm4 fusion protein. (A) NIH 3T3 cell expressing the mLsm4-GFP fusion analyzed under vital conditions by confocal laser-scanning microscopy. Fluorescent labeling is partly cytoplasmic but is concentrated in the nucleus. (B) Immunostaining of the same cell as that in panel A with anti-3-methylguanosine monoclonal antibodies, showing the speckled distribution of the antigen. (C) Superimposition of panels A and B. EGFP-mLsm4 does not distribute in 3-methylguanosine-containing speckles. (D) Control NIH 3T3 cell expressing wild-type EGFP. Expression is distributed all over the cell. (E) Staining of the same cell with anti-3-methylguanosine antibodies. (F) Superimposition of panels D and E. (G) Western blot analysis with anti-EGFP antibodies of protein extracts derived from EGFP-mLsm4- or wild-type EGFP-expressing cells. Fusion of mLsm4-specific amino acids shifts EGFP immunoreactivity to a predicted molecular mass of approximately 40 kDa.

suggesting that the gene products form a group of orthologue proteins (Fig. 1A).

Cloning and characterization of the G101 allele. To test whether the gene encoding mLsm4 was disrupted by the integration of the targeting vector, the 5' regions of the mutant and wild-type alleles were cloned, sequenced, and compared. The mutant DNA between the exon identified by 5' RACE (referred to below as exon 1) and the β -galactosidase-neomycin cassette was amplified by PCR with genomic DNA from the G101 ES cell clone as template. The wild-type intron between exons 1 and 2 was amplified with primers designed on the basis of the mLsm4 cDNA. The length of the PCR product was 4.4 kb from the mutant allele and 3.4 kb from the wild-type allele. Sequence comparisons of the mutant and wild-type PCR products revealed sequence identity up to the last 4 nucleotides of intron 1 (Fig. 1B). Downstream of this point, the mutant product showed DNA sequences derived from the targeting vector. These data indicate that the first 2 exons of the G101 allele are normally separated by a 3.4-kb intron and that the mutation resulted from the integration of the targeting vector 4 bp upstream of exon 2, leading to an abnormal splice product composed of exon 1 and the β -galactosidase gene (Fig. 1B).

mLsm4p is located predominantly in the cell nucleus but also in the cytoplasm. To characterize the intracellular localization of mLsm4, the cDNA was inserted in frame with the 3' end of an EGFP reporter gene and transfected into NIH 3T3, COS7, and HEK293 cells. Transient-transfection products were analyzed by confocal laser-scanning microscopy 24 to 48 h after lipofection. In transfected cells that appeared adherent and well spread, the fluorescent signal was strong in the nucleus and weak in the perinuclear cytoplasm (Fig. 2A). A comparison of EGFP-mLsm4 distribution with anti-3-methylguanosine immunolabeling (Fig. 2B and C) revealed overlapping expression of EGFP-mLsm4p and anti-3-methylguanosine in the nucleus but not in the perinuclear cytoplasm,

where only weak staining of EGFP-mLsm4 was detected. The staining of EGFP-mLsm4 was evenly distributed throughout the nucleus (Fig. 2A and C) and never showed the speckled signal typical of spliceosomal components (Fig. 2B). In control experiments, cells were transfected with EGFP alone. The fluorescent signal appeared evenly distributed in the whole cytoplasm and nucleus (Fig. 2D and F). The presence in the EGFP-mLsm4-transfected cells of the correct fusion protein was confirmed by immunoblotting with an antibody recognizing EGFP (Fig. 2G).

The mLsm4 gene is ubiquitously transcribed. To characterize the role of mLsm4 during mouse development, G101 ES cells were injected into blastocysts to generate germ line chimeras. Heterozygous mice showed no morphological abnormalities, were fertile, and had a normal life span. Since in these mice β -galactosidase was transcribed under the control of the mLsm4 promoter, LacZ expression was analyzed at different developmental stages. The expression of the reporter gene was ubiquitously detectable in both embryonic and extraembryonic tissues at E7 (Fig. 3A) and throughout the embryo at later stages of development (Fig. 3B to D).

To better understand the in vivo expression pattern of mLsm4 in adult mice, β -galactosidase activity was additionally examined in sections of several organs dissected from heterozygous G101 animals. LacZ expression was ubiquitous in all organs tested, including the lungs, thymus, trachea, liver, kidneys, brain, cerebellum (Fig. 3E), heart (Fig. 3F), skeletal muscles (Fig. 3G), and skin (Fig. 3H). Stained serial sections through these organs demonstrated LacZ expression in all cell types, supporting the view of a housekeeping function of the mLsm4 gene in adult mice.

Genetic ablation of the mLsm4 gene causes peri-implantation death. To test whether the function of the mLsm4 gene was rate limiting during development or could be compensated for by other Sm-containing proteins, mice heterozygous for the

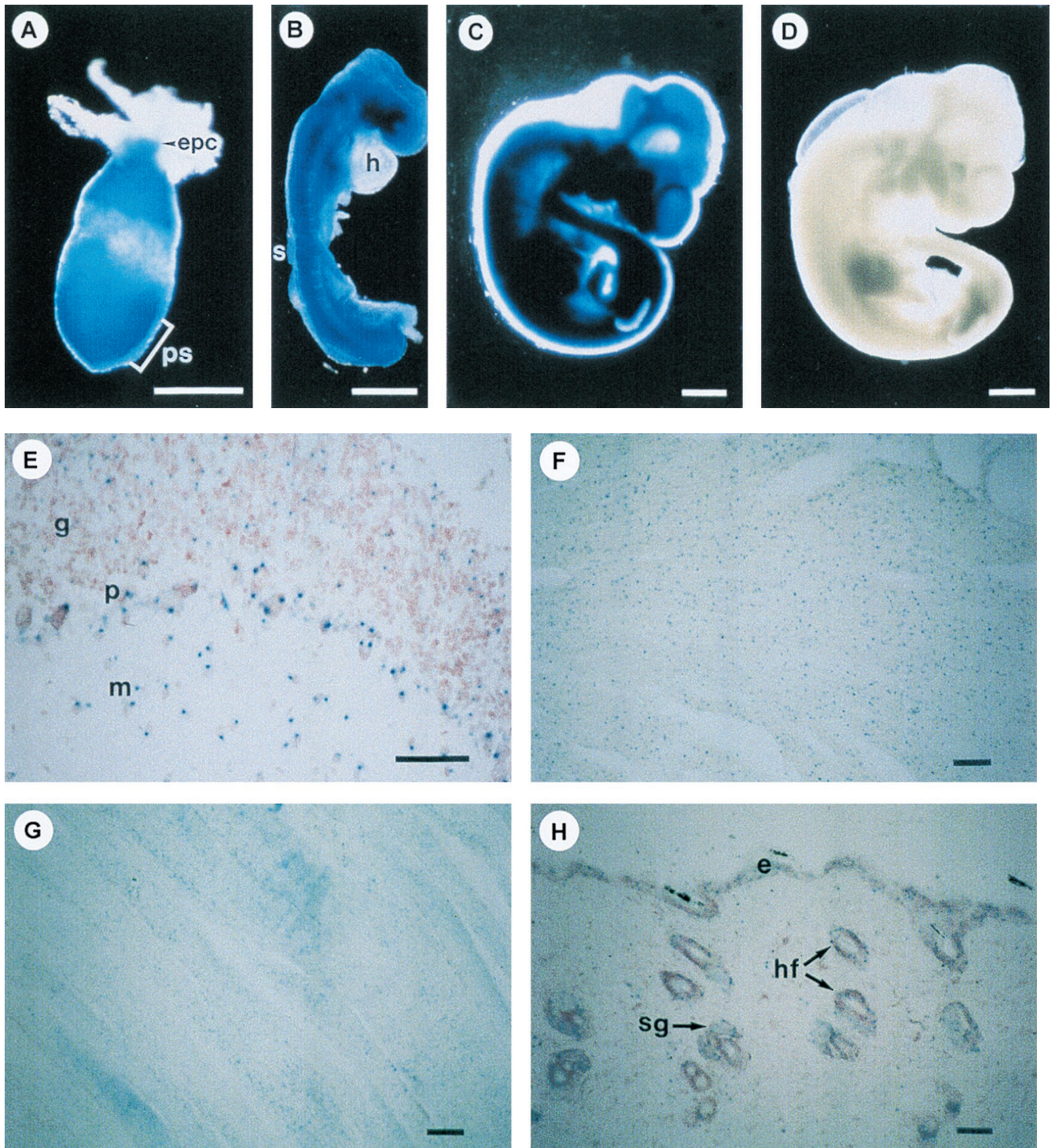


FIG. 3. Whole-mount preparation of LacZ-stained G101 heterozygous embryos at different developmental stages (A to D) and localization of LacZ expression in tissue sections derived from adult heterozygous mice (E to H). (A) E7 embryo still contained in its extraembryonic membranes. 5-Bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) staining appears ubiquitous, labeling both embryonic (ps) and extraembryonic tissues. epc, ectoplacental cone; ps, primitive streak. (B) E8.5 embryo (12 somites) dissected out of the yolk sac and amnion. LacZ activity strongly labels the central nervous system and mesodermal derivatives. The swollen structure of the heart makes it difficult to appreciate its staining. h, heart; s, somites. (C) Dissected E10.5 embryo. Tissues derived from all three germ layer clearly express the *geo* reporter gene. (D) E10.5 wild-type control embryo from the same litter as that shown in panel C. (E) Cerebellar cortex. LacZ staining is evident in the cytoplasm of Purkinje cells (p) but also of cells in the granular (g) and molecular (m) layers. (F) Myocardium. (G) Skeletal muscle. (H) Skin. The epidermis (e) and the dermis both express LacZ. Staining also labels LacZ activity in the hair follicles (hf) and sweat glands (sg). Bar, 500 μ m (A to D) and 50 μ m (E to H).

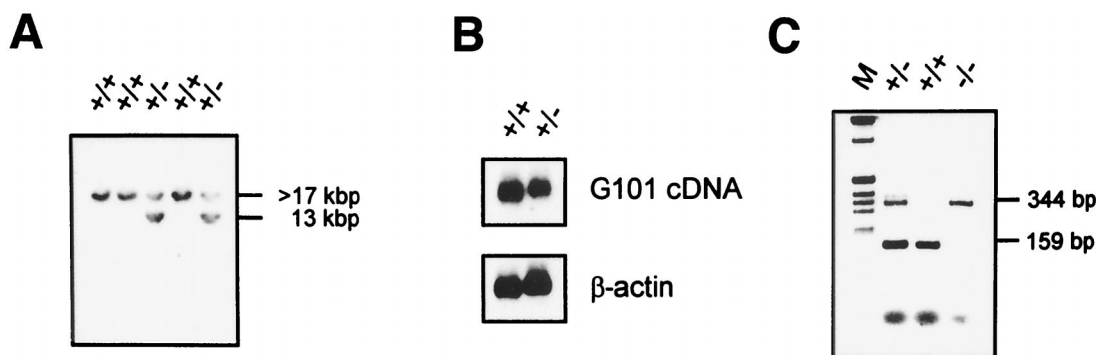


FIG. 4. G101 heterozygous mouse crossings: correlation of the genotype with the mLsm4 mRNA expression level. (A) Southern blot analysis of a typical litter obtained by crossing G101 heterozygous mice. The probe corresponds to the 1.4-kb fragment shown in Fig. 1B. The 17- and 13-kb bands identify the wild-type and mutant alleles, respectively. (B) Typical result of Northern blot detection of mLsm4 mRNA in wild-type and heterozygous G101 mice. As expected for a null allele, the G101 mutation leads to a twofold reduction of mLsm4 mRNA in heterozygous mice. (C) Genotyping of blastocysts derived from a G101 heterozygous mouse crossing. The 159- and 344-bp bands identify the wild-type and mutant alleles, respectively. M, 1-kb ladder from GIBCO-BRL.

G101 mutation were intercrossed to produce homozygous animals. Adult progeny were genotyped by Southern blot assay (Fig. 4A). Of 252 offspring analyzed, none was homozygous, indicating that the mutation is a recessive lethal. To further prove that the G101 locus encoded a null allele and that the embryonic death was due to the lack of mLsm4 protein, the mLsm4 cDNA was used to hybridize mRNA derived from wild-type and heterozygous G101 mice. As expected, the amount of mLsm4 mRNA was reduced to half in heterozygous animals (Fig. 4B).

To determine the developmental stage at which the lack of mLsm4 causes death, blastocysts were isolated from heterozygous crosses, individually photographed, and genotyped by PCR. Of 34 embryos examined, none showed morphological abnormalities. After genotype analysis, five embryos were found to be homozygous, indicating that lack of mLsm4 does not impair preimplantation development (Fig. 4C).

Next, the exact time at which mLsm4-deficient embryos die was assessed at different developmental stages. Decidual swellings derived from heterozygous crossings were dissected at E6.5, E7.5, and E10.5. Of 26 E6.5 implantation sites analyzed by hematoxylin and LacZ staining, 5 contained wild-type, LacZ-negative embryos (Fig. 5A and D). Of the LacZ-positive embryos, 15 appeared normal (Fig. 5B and E) and 6 contained only few traces of LacZ-positive trophoblastic cells and signs of maternal blood infiltration (Fig. 5C and F). At stage E7.5, 6 of 19 decidual swellings contained no embryo. At E10.5, 5 of 17 decidual swellings were without embryos. PCR analysis of tissue derived from the normal embryos at all stages analyzed indicated that they were either wild type or heterozygous but never homozygous for the G101 mutation. The presence of approximately 25% empty decidual swellings at E7.5 and E10.5 indicated that homozygous embryos were dying between implantation and E6.5.

DISCUSSION

Proteins with an Sm motif play important roles in snRNP biogenesis and transport and in the splicing of nuclear pre-mRNA. By using a gene trap approach in ES cells, we identified the mouse orthologue of the human Sm-like protein hLsm4p (23). mLsm4p was found to be ubiquitously expressed in embryonic and adult tissues, and its genetic deletion appeared to cause a peri-implantation embryonic death.

Sequence comparisons revealed that mLsm4p has high homology to at least five proteins expressed in distant species.

This confirms previous observations indicating that Sm-like proteins share a deep evolutionary origin (23). As shown for another set of Sm-like proteins (Lsm2p homologues [19]), the grouping of Lsm4p putative homologues is corroborated by the identification of extensive homology beyond the Sm sites. Interestingly, all putative Lsm4 proteins, excluding *Saccharomyces cerevisiae* Lsm4p/Uss1p, contained at the C terminus one or more partially conserved RGG repeats, which are known to bind RNA (16). The specific function of Lsm proteins has not yet been completely defined, and this observation might support the previous hypothesis that they can play multiple roles in RNA processing (19, 23).

Recent evidences indicate that most Sm-like proteins, including Lsm4p, are part of the [U4/U6 · U5] tri-snRNP (12, 19, 23). Lsm4p was found to associate with U6 snRNA in both human and yeast cells (23), and, given the high homology between human and murine Lsm4p, these data strongly suggest that mLsm4p has an identical function in the mouse.

Yeast mutants lacking Lsm proteins such as Lsm4p show reduced levels of U6 snRNA, suggesting that these proteins are essential for the formation of a snRNP complex which protects U6 snRNA from degradation (4, 19). The assembly of U snRNP complexes follows the transient transport to the cytoplasm of the nucleus-encoded U snRNA. Binding of Sm proteins to the U snRNA results in the formation of the Sm core domain, a highly stable RNP complex which, excluding U6 snRNP, is common to all U snRNPs. Transport to the nucleus of U1, U2, U3, U4, and U5 snRNAs can be accomplished only after the association with Sm proteins. It is currently hypothesized that in the U6 snRNP, a certain subset of Lsm proteins (namely, Lsm2p to Lsm8p) might share similar functions, resulting in the stabilization of U6 snRNA.

Transport studies with oocytes indicated that U6 snRNA does not leave the nucleus after transcription. In contrast, experiments with mouse fibroblasts show that U6 free from U4 snRNA is present in the cytoplasm, where it is matured before nuclear import (11). Mayes et al. (19) suggest that the Lsm complex might act as part of the nuclear localization signal of the U6 snRNP. Visualization of transfected GFP-tagged mLsm4p showed that it localizes in the nucleus and in the perinuclear cytoplasm. This observation suggests that mLsm4p may indeed serve as a chaperone involved in the transport of U6 snRNP to the nuclear compartment. Interestingly, the C-terminal region of mLsm4 harbors a sequence that resembles the M9 transport signal of hnRNP1 (amino acids 83 to 136) (10). Further studies are required to determine whether

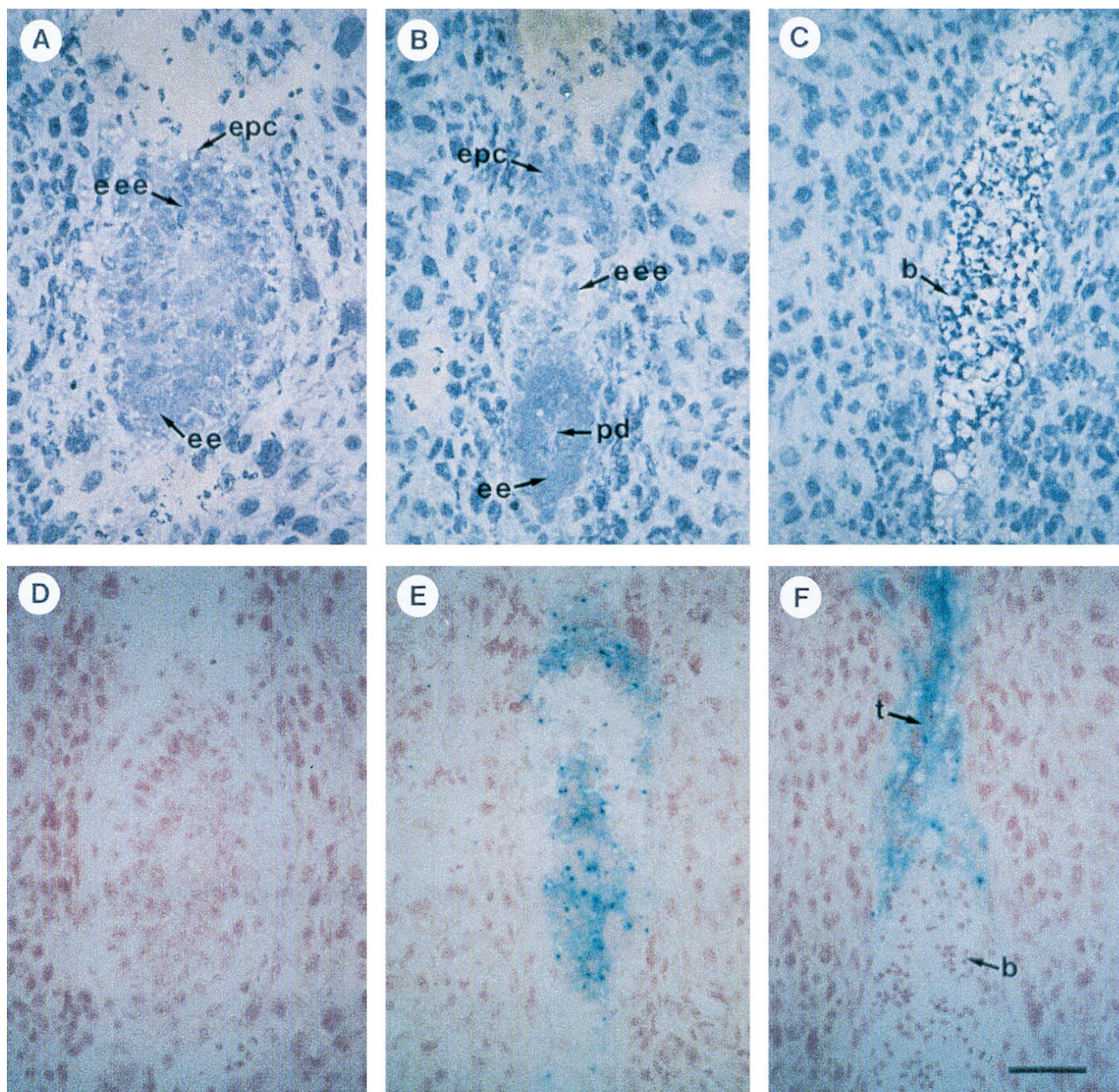


FIG. 5. Parasagittal sections of E6.5 decidua derived from a heterozygous cross. Adjacent sections from three embryos (A and D, B and E, and C and F) were stained either with hematoxylin (A to C) or with X-Gal and Nuclear Fast Red (D to F). Adjacent sections in panels A and D show that the embryo is wild type: morphology is normal but cells are *geo* negative. The sections in panels B and E show a presumably heterozygous embryo that appears normal and *geo* positive. The embryo in panels C and F shows the hallmarks of a putative homozygous embryo: accumulation of maternal blood cells, no detectable structure of the embryo proper, and X-Gal-stained trophoblast remnants. b, maternal blood cells; ee, embryonic ectoderm; eee, extraembryonic ectoderm; epc, ectoplacental cone; p, proamniotic cavity; t, trophoblast cells. Bar, 50 μ m.

mLsm4p has a shuttling activity. Nonetheless, whereas other spliceosomal U snRNPs show a speckled nuclear pattern, the cellular distribution of mLsm4 was scattered throughout the nucleus, suggesting that mLsm4p might be highly mobile and able to relocate dynamically to areas where it fulfills its function.

So far, the function of specific Lsm proteins has been analyzed only in mutant yeast strains. Under these experimental conditions, most of the Lsm proteins appear to be essential for cell viability. Mutation of *S. cerevisiae* Lsm4p/Uss1p blocks pre-mRNA splicing and results in a lethal phenotype (4). The

identification of a mLsm4-null allele in mice opens the possibility of studying the *in vivo* function of Lsm proteins in higher eukaryotes. Lsm4-null mice reach the blastocyst stage, hatch, become adhesion competent, and implant but die shortly thereafter, thus showing an absolute requirement of mLsm4 for normal development. At present we do not know whether mLsm4 is crucial for the survival of mammalian cells or critical for differentiation of early cell lineages and consequently for the arrest of further development. Several indirect observations suggest that the disruption of the mLsm4 gene may be lethal to the cell. First, mLsm4 is similar to the yeast protein

Lsm4/Uss1, which is absolutely required for yeast cell survival (4). Second, we have cultured mLsm4-null blastocysts in vitro and found that they rapidly deteriorate. Interestingly, however, inner cell mass cells are affected much earlier by the absence of mLsm4 than are the trophectodermal cells, which can survive for some days. Third, the mLsm4 protein is expressed in all cells of developing embryos and adult mice. The ubiquitous expression pattern suggests a housekeeping function of mLsm4 which may well be important for the general cell survival. Fourth, the survival of mLsm4-null embryos to the peri-implantation stage could well be due to the presence of maternal mLsm4 mRNA, which may become depleted at this stage, leading to the rapid death of all cells of the embryo. Several pre- and peri-implantation lethal phenotypes are caused by depletion of maternal mRNA. For example, mice lacking the survival-of-motor-neurons (SMN) gene die at the morula stage. The early lethal phenotype is due to the depletion of maternal mRNA coding for SMN, resulting in massive cell death and inhibition of development beyond the morula stage (24). SMN is important for snRNP assembly and activity, and mutations in SMN in humans cause spinal muscular atrophy (8). An explanation for the significantly longer survival of mLsm4-null mice than of SMN-null mice arises from previous observations which suggested that snRNAs and snRNPs accumulate in the oocyte and that their de novo expression starts after the morula stage (5). This may indicate that a large pool of snRNPs of maternal origin is required to splice newly synthesized mRNAs as soon as transcriptional activation occurs in the two-cell embryo and that mLsm4 represents a vital part of this pool.

The availability of the mLsm4-deficient mouse strain provides a valuable tool for a further characterization of mLsm4. In particular, it will be possible to introduce cDNAs encoding mutant forms of mLsm4 into an mLsm4-null background. Such genetic complementation experiments can be used to identify the interaction site(s) of mLsm4 with other proteins and to identify the minimal requirements for in vivo rescue.

ACKNOWLEDGMENTS

We thank Stefan Benkert for expert technical assistance and Ray Boot-Handford and Cord Brakebusch for critically reading the manuscript.

E.H. was supported by a Max Planck fellowship, and R.F. was supported by the Hermann and Lilly Schilling Stiftung, Deutsche Forschungsgemeinschaft and the Swedish National Research Foundation.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Bloch, W., E. Forsberg, S. Lentini, C. Brakebusch, K. Martin, H. W. Krell, U. H. Weidle, K. Addicks, and R. Fässler. 1997. $\beta 1$ integrin is essential for teratoma growth and angiogenesis. *J. Cell Biol.* **139**:265–278.
- Branlant, C., A. Krol, J. P. Ebel, E. Lazar, B. Haendler, and M. Jacob. 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J.* **1**:1259–1265.
- Cooper, M., L. H. Johnston, and J. D. Beggs. 1995. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.* **14**:2066–2075.
- Dean, W. L., A. C. Seufert, G. A. Schultz, R. S. Prather, C. Simerly, G. Schatten, D. R. Pilch, and W. F. Marzluff. 1989. The small nuclear RNAs for pre-mRNA splicing are coordinately regulated during oocyte maturation and early embryogenesis in the mouse. *Development* **106**:325–334.
- Fässler, R., and M. Meyer. 1995. Consequences of lack of $\beta 1$ integrin gene expression in mice. *Genes Dev.* **9**:1896–1908.
- Fässler, R., M. Pfaff, J. Murphy, A. A. Noegel, S. Johansson, R. Timpl, and R. Albrecht. 1995. Lack of $\beta 1$ integrin gene in embryonic stem cells affects morphology, adhesion, and migration but not integration into the inner cell mass of blastocysts. *J. Cell Biol.* **128**:979–988.
- Fischer, U., Q. Liu, and G. Dreyfuss. 1997. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell.* **90**:1023–1029.
- Fischer, U., V. Sumpter, M. Sekine, T. Satoh, and R. Lührmann. 1993. Nucleocytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J.* **12**:573–583.
- Fischer, U., W. M. Michael, R. Lührmann, and G. Dreyfuss. 1996. Signal-mediated nuclear export pathways of proteins and RNAs. *Trends Cell Biol.* **6**:290–293.
- Fury, M. G., and G. W. Zieve. 1996. U6 snRNA maturation and stability. *Exp. Cell Res.* **228**:160–163.
- Gottschalk, A., G. Neubauer, J. Banroques, M. Mann, R. Lührmann, and P. Fabrizio. 1999. Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* **18**:4535–4548.
- Green, M. J. 1991. Biochemical mechanism of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* **7**:559–599.
- Hermann, H., P. Fabrizio, V. A. Raker, K. Foulaki, H. Hornig, H. Brahm, and R. Lührmann. 1995. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interaction. *EMBO J.* **14**:2076–2088.
- Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Lührmann, J. Li, and K. Nagai. 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**:375–387.
- Kiledjian, M., and G. Dreyfuss. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**:2655–2664.
- Krämer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**:367–409.
- Mattaj, J. W., D. Tollervey, and B. Seraphin. 1993. Small nuclear RNAs in messenger RNA and ribosomal RNA processing. *FASEB J.* **7**:47–53.
- Mayes, A. E., L. Verdone, P. Legrain, and J. D. Beggs. 1999. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.* **18**:4321–4331.
- Nayler, O., C. Cap, and S. Stamm. 1998. Human transformer-2-beta gene: complete nucleotide sequence, chromosomal localisation and generation of a tissue specific isoform. *Genomics* **53**:191–202.
- Pannone, B. K., D. Xue, and S. L. Wolin. 1998. A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.* **17**:7442–7453.
- Rio, D. C. 1993. Splicing of pre-mRNA: mechanism, regulation and role in development. *Curr. Opin. Genet. Dev.* **3**:574–584.
- Salgado-Garrido, J., E. Bragado-Nilsson, S. Kandels-Lewis, and B. Seraphin. 1999. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.* **18**:3451–3462.
- Schrank, B., R. Götz, J. M. Gunnensen, J. M. Ure, K. V. Toyka, A. G. Smith, and M. Sendtner. 1997. Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl. Acad. Sci. USA* **94**:9920–9925.
- Seraphin, B. 1995. Sm and Sm like proteins belong to a large family: Identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14**:2089–2098.
- Tan, E. M. 1989. Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv. Immunol.* **44**:93–151.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.