Nucleic Acids Research, 1999, Vol. 27, No. 18 **3685–3689**

Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing

Theodore P. Rasmussen, Tracy Huang[1], Mary-Ann Mastrangelo, Janet Loring, Barbara Panning and Rudolf Jaenisch[1,*]

Whitehead Institute for Biomedical Research and [1]Department of Biology, M.I.T, Nine Cambridge Center, Cambridge, MA 02142, USA

Received May 28, 1999; Revised and Accepted July 16, 1999

ABSTRACT

Two protein isoforms of histone macroH2A1 (mH2A1) are found in mammalian cells. One isoform, mH2A1.2 is highly concentrated on the heterochromatinized inactive X chromosome (Xi) of female cells. mH2A1.2 protein is also present in male cells, but fails to form dense concentrations. Another protein isoform, mH2A1.1, differs from mH2A1.2 by a single short segment of amino acids. In this study, we cloned and characterized the genomic locus of the mouse mH2A1 gene and mapped it to chromosome 13. Two alternatively spliced transcripts derived from the mH2A1 locus are responsible for the generation of the two mH2A1 protein isoforms with mH2A1.2 mRNA being the most abundant spliced form in all tissues examined. The absolute amount of mH2A1 mRNA is similar in male and female cells for most tissues with the exception of testes where it is particularly abundant. Both spliced forms are present in all adult tissues analyzed as well as in female embryonic stem cells. In contrast, male embryonic stem cells expressed mH2A1.1 at low levels if at all. The relatively abundant expression of mH2A1 in both sexes suggests that mH2A1 has functions in addition to a possible involvement in X chromosome inactivation.

INTRODUCTION

The protein structure of histone macroH2A1 (mH2A1) resembles that of H2A core histones, except for the presence of a large C-terminal extension (the non-histone domain) of unknown function (1). Two protein isoforms of mH2A1 called mH2A1.1 and mH2A1.2 differ by a single short stretch of amino acids embedded within the non-histone domain (Fig. 1A). The 42 kDa mH2A1 protein co-sediments with mononucleosomes (1). mH2A1.2 protein preferentially associates with the inactive X chromosome (Xi) in female rat liver nuclei, where it forms a macrochromatin body (MCB). The association of mH2A1.2 protein with Xi is not exclusive, though, because additional diffuse staining is present in nuclei from both sexes (2).

Inactivation of one of two X chromosomes in female mammalian cells requires the action of Xist, a large untranslated RNA expressed from the X inactivation center (XIC) on the X chromosome. In differentiated female cells, Xist RNA is expressed only from the inactive X chromosome (3,4), where it remains associated with Xi in cis (5). Disruption of Xist results in failed X-inactivation and embryonic lethality (6,7). The mechanism by which Xist brings about X chromosomal silencing is unclear. Recently, it has been shown in mouse embryonic fibroblasts containing a conditional allele of Xist that the association of mH2A1 with Xi is disrupted upon removal of the Xist gene (8). Therefore, a ribonucleoprotein complex may exist that includes both Xist RNA and mH2A1 protein.

We show here that messages generated from the mH2A1 locus are alternatively spliced. In addition, we show that mature mH2A1 RNA is present at similar levels in adult male and female tissues. Histone macroH2A1 message is also particularly abundant in testes and is expressed in differentiating male and female embryonic stem cells. These patterns of expression suggest functions for mH2A1 in addition to X inactivation.

MATERIALS AND METHODS

Cloning and analysis of mH2A1 locus

A genomic bacterial artificial chromosome (BAC) library called RPCI-22 was obtained from the Roswell Park Cancer Institute. This library was generated from female 129/SvEvTACfBr mice. A probe derived from the non-histone region of a mouse EST (GenBank accession no. W33744) was used to identify genomic clones. Three independent BACs were identified and subcloned using *Hin*dIII or *Xba*I digests. Coding regions on subclones were sequenced with primers designed from cDNA sequence. Sequencing reactions extended into introns as judged by comparison of genomic DNA with cDNA clones. cDNA clones were obtained by screening a mouse cDNA library (Origene Technologies, Inc.) with PCR reactions specific for mH2A1.1 and mH2A1.2 spliced cDNA products. cDNAs including mH2A1.2 and mH2A1.1 were sequenced in their entirety and have been assigned GenBank acession numbers AF171080 and AF171081, respectively.

*To whom correspondence should be addressed. Tel: +1 617 258 5186; Fax: +1 617 258 6505; Email: jaenisch@wi.mit.edu

Chromosomal assignment was performed by fluorescent *in situ* hybridization (FISH) of mouse metaphase chromosome spreads with an mH2A1 BAC and P1 or BAC probes that label known mouse chromosomes (9). Using this approach, the mH2A1 locus was assigned to chromosome 13. Cytogenetic preparations were obtained using standard techniques. A FISH probe for chromosome 13 was biotinylated and a probe for mH2A1 was Cy3 labeled. Labeled probes were precipitated together with sonicated salmon sperm DNA, tRNA and COT-1 DNA, washed extensively with 70% ethanol, dried, resuspended in 50% formamide/2× SSC, and denatured for 10 min at 70°C. Hybridization was carried out at 37°C overnight on slides denatured at 70°C in 70% formamide/2× SSC pH 7.0 for 2 min, then dehydrated in ice-cold ethanol immediately prior to use. The slides were then washed extensively and the biotin-labeled probe was detected using avidin conjugated to FITC. The CY3-labeled probe was directly detected. DNA was counter-stained with DAPI at 10 ng/ml for 10 min.

ES cell culture

F1 ES cell lines 2–2 (female) and 2–3 (male) were prepared by standard methods from delayed blastocysts of a cross of a *Mus musculus casteneus*/ei male with a *Mus musculus domesticus* 129/Sv/Jae female (10). ES cells (without feeder cells) were grown in DMEM supplemented with 15% FBS and 1000 U/ml leukemia inhibitory factor (LIF). ES cells were induced to differentiate into embryoid bodies by passaging without feeders into the above medium minus LIF (11).

RNA methods

RNA was isolated from ES cells and tissues using RNazol with recommended methods (Teltest, Inc.) RNA was separated on 1% agarose gels with formaldehyde and northern blotting was performed using a restriction fragment from the non-histone region of a mouse EST clone (GenBank accession no. 33744) and a probe specific for GAPDH. Reverse transcriptase polymerase chain reaction (RT–PCR) was performed by annealing random hexamers to RNA, followed by first strand cDNA synthesis using Superscript reverse transcriptase (Gibco-BRL). PCR was then carried out for 30 cycles using standard conditions. Primers for mH2A1 RT–PCR have the following sequences: 1F: 5'-GAAGAAGAAATCCACCAA-GAC-3', 01F: 5'-CTAAGT-ATAGGATCGGAGTGG-3', 02F 5'-AAGAAGGGACGGGTCA-CAC-3':, 2F: 5'-ACAGGAGGC-AAGAAAGGGG-3', 3F: 5'-TTGCAAGTTGTTC-AGGCT-GAC-3', 4F: 5'-GAACCTTATTCACAGT-GAAATC-3', 5F: 5'-TAGGAAACACACTGGAGAAG-3', 1R: 5'-TTGGCGTC-CAGCTTGGCC-3', 2R: 5'-GACTATTACAGTGGATCACA-AAC-3', 5R: 5'-CTTC-TGGCCGAGGAAGAGG-3', 02R: 5'-TCCTCGCTTC-TTCGCTAGC-3', 01R: 5'-GC-TGCATT-GCCAGCCAGC-3'. GAPDH1: 5'-AAGG-GCTCATGAC-CACAGTCCATGCCAT-3', GAPDH2: 5'-ATATTCATTGT-CATACCAGGAAAT-3'

RESULTS

The mouse mH2A1 genomic locus

The sequences of a number of mouse ESTs were identified that are highly similar to the rat histone macroH2A1 complete cDNA sequence (1). Alignment of these ESTs showed that they fell into two classes that were identical except for a short region near the center of the sequences (Fig. 1A) that consisted of two alternative sequences similar to the rat mH2A1.1 and 1.2 cDNA sequences. The presence of two alternative sequences in mature cDNAs flanked by common identical sequences suggested that alternative splicing could explain the two classes of mouse mH2A1 cDNAs. We therefore amplified a set of overlapping RT–PCR products to scan for the presence of alternatively spliced exons that might give rise to RT–PCR products of novel size (Fig. 1B). RNA extracted from undifferentiated and differentiated male and female ES cells were used for first strand cDNA synthesis. A comprehensive set of overlapping RT–PCR products were obtained each of which had a mobility on agarose gels similar to that predicted from the EST contigs. Primer pairs 3F, 2R and 4F, 2R (which specifically amplify mH2A1.1 and 1.2 sequences, respectively) gave rise to specific RT–PCR products. No other alternatively spliced regions were detected.

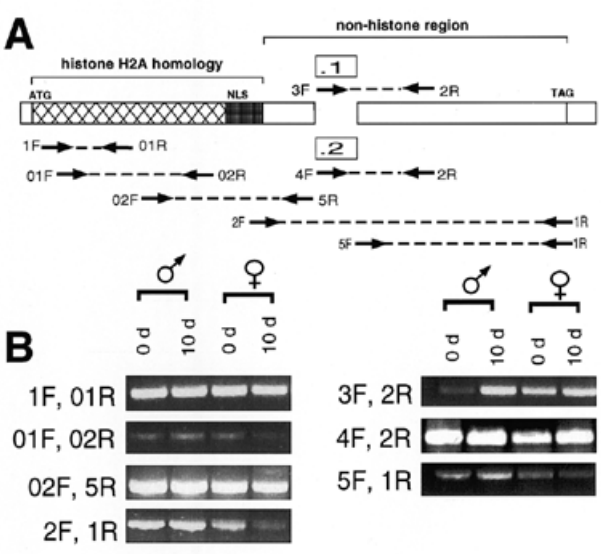

Figure 1. RT–PCR analysis to detect contiguous regions of mH2A1 mature spliced mRNA. (**A**) Predicted structures of mH2A1 RNAs derived by assembling ESTs into contigs. Arrows indicate the location of primers used for RT–PCR. Dotted lines represent the predicted RT–PCR products. Two alternative nucleotide sequences (designated .1 and .2) were observed among aligned ESTs, but these were always flanked by identical sequences. (**B**) RT–PCR analysis of RNA extracted from J1 (male) and 2–2 (female) embryonic stem cells either undifferentiated (0 d) or differentiated for 10 days (10 d). Primer pairs used are indicated and correspond to the positions shown in (A).

Three BACs from a genomic library derived from mouse strain 129/SvEvTACfBr were obtained using a mouse EST probe derived from the non-histone region of mouse mH2A1.2. BACs were subcloned into standard cloning vectors and sequenced using primers designed upon mouse EST sequences. cDNAs for mH2A1.1 and 1.2 messages were obtained from an E18 mouse embryo cDNA library using PCR assays with primers 3F and 2R to detect mH2A1.1 and 4F and 2R to detect mH2A1.2. Comparison of genomic sequence with DNA sequence revealed that the mouse mH2A1 locus contains

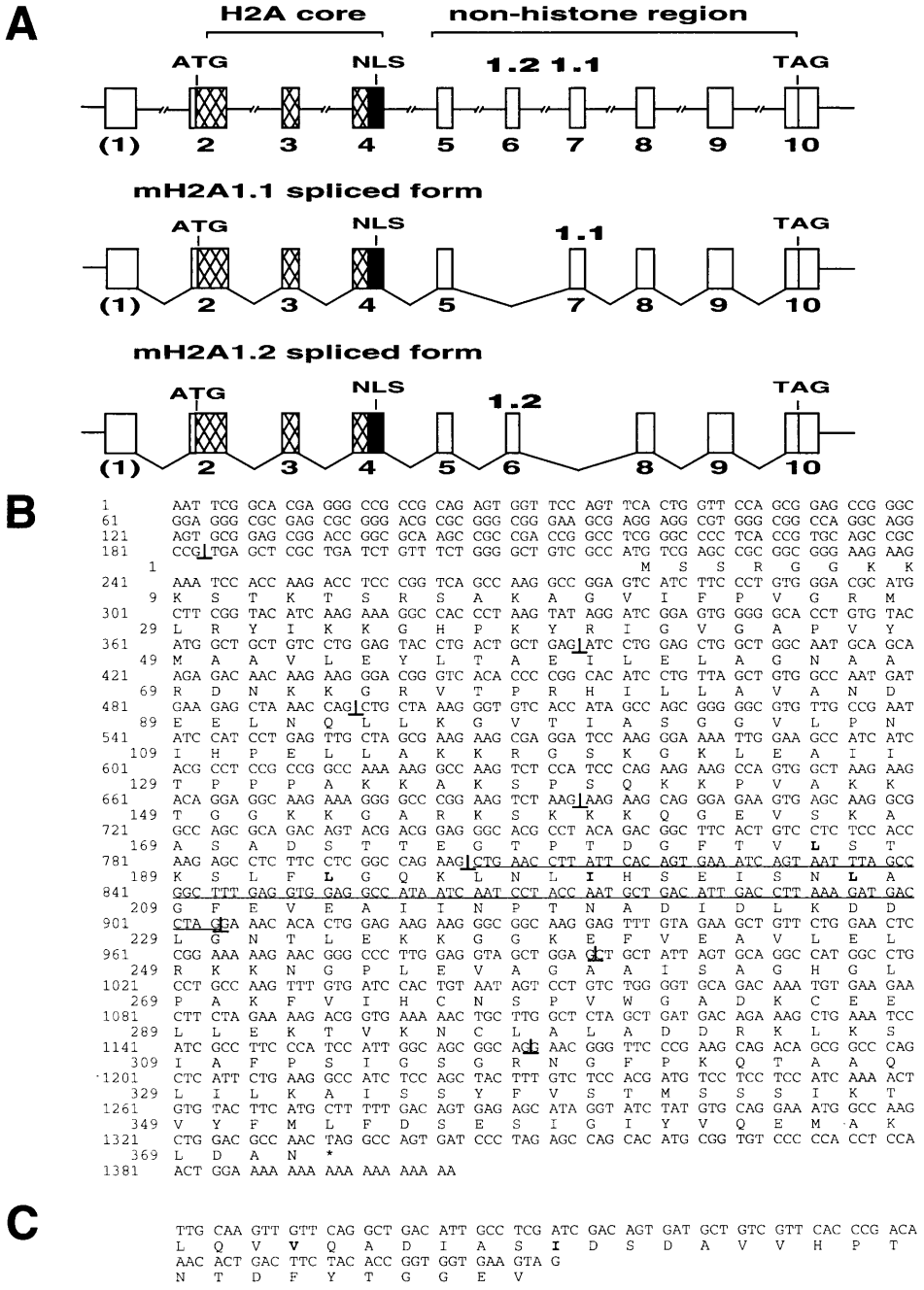

Figure 2. Genomic structure and cDNA sequences of mH2A1. (**A**) The mH2A1 locus contains nine coding exons. Introns are not shown to scale. Adjacent but alternatively spliced exons give rise to mRNAs encoding mH2A1.1 and mH2A1.2 mature spliced mRNAs. (**B**) cDNA sequence of mH2A1.2 mature spliced mRNA. Vertical bars indicate the positions from which introns are removed. The sequence specifying mH2A1.2 is underlined. Amino acid residues in the putative leucine zipper are shown in bold. (**C**) The sequence of exon 7, which occurs only in mH2A1.1-spliced message. Again, amino acid residues in the putative leucine zipper are shown in bold.

nine coding exons (Fig. 2A). At least one additional exon is present in the 5' untranslated region, though it was not identified in genomic subclones. Sequence encoding the mH2A1.2-specific region was found exclusively in exon 6, while sequence encoding the mH2A1.1-specific region was found exclusively in exon 7. The order of exons 6 and 7 was determined by restriction analysis and long-range PCR (data not shown). Therefore, the mouse mH2A1 locus contains two adjacent and alternatively spliced exons that specify two spliced isoforms of mature mH2A1 message. Exons 6 and 7 partially overlap a proposed leucine

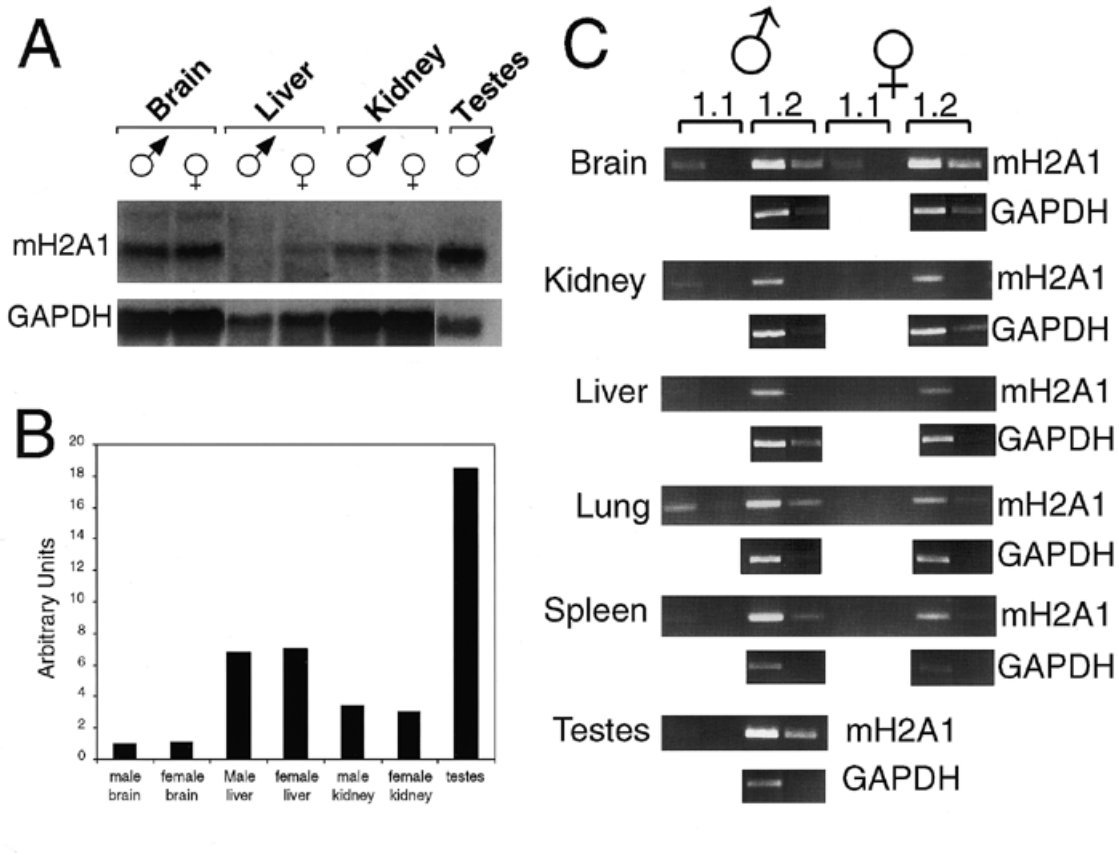

Figure 3. mH2A1 RNA expression and alternative splicing in adult tissues. (**A**) Northern blot of mH2A1 RNA extracted from male and female adult tissues. (**B**) Phoshorimager quantitation of northern blot shown in (A). mRNA abundance is shown as arbitrary units that resulted from standardization for loading to GAPDH mRNA. (**C**) Relative levels of alternatively spliced mH2A1.1 and mH2A1.2 mRNAs in RNA extracted from male and female tissues. Primers 3F and 2R were used to amplify mH2A1.1-specific products and 4F and 2R were used to amplify mH2A1.2-specific exons. Both primer pairs span introns in genomic DNA. Each assay was performed in duplicate, either with 2 µl of first-strand cDNA (left bands) or a 1/5 dilution of first-strand cDNAs (right bands). GAPDH-specific products from primers GAPDH1 and GAPDH2 show approximately equal loading for all samples.

zipper (1). The sequences of mH2A1 cDNAs are presented in Figure 2B and C. The predicted mouse mH2A1 protein sequences are nearly identical to the previously reported rat mH2A1 protein sequence (1).

A fluorescently labeled mH2A1 BAC and a P1 clone known to hybridize to chromosome 13 (Anton Wutz, personal communication) were shown to colocalize using FISH performed on mitotic chromosome spreads (data not shown). These results map the mH2A1 gene to chromosome 13.

Tissue-specific expression and alternative splicing of mH2A1

Northern blotting was used to determine the relative levels of mH2A1 RNA in various tissues using a probe that hybridizes to the non-histone region of mH2A1 (Fig. 3A). This probe cannot distinguish the two spliced forms of mH2A1 because of the similar small size of exons 6 and 7. We found that mH2A1 was expressed at nearly identical levels in both sexes in all tissues examined as judged by phosphorimager analysis using GAPDH for a loading standard (Fig. 3B). mH2A1 message was particularly abundant in testes. We performed isoform-specific semi-quantitative RT–PCR assays to determine the splicing profiles of various male and female tissues (Fig. 3C). All tissues analyzed contained mostly mH2A1.2 spliced RNA, though lower levels of mH2A1.1 RNA were also detected. No significant differences in the splicing patterns were observed between tissues from males and females.

We also analyzed mH2A1 expression profiles in more detail in differentiating male and female ES cells. The levels of mH2A1 message were similar in male and female cells over the course of differentiation as judged by northern blotting (Fig. 4A). The majority of mH2A1 message migrates at ~1.4 kb (which is similar to the length of the mH2A1.2 cDNA clone), but a minor band at 3.5 kb was detected. Although we cannot be sure that the 3.5 kb band represents a minor subclass of mH2A1 RNA, a human cDNA containing mH2A1.1 sequence and a long 3' untranslated region was recently identified from human tissues (12).

Semiquantitative analysis of spliced isoform content revealed that male and female ES cells contain predominantly mH2A1.2 spliced RNA. Female ES cells produce readily detectable levels of mH2A1.1 RNA, which remain rather constant over the course of differentiation. In contrast, two independent

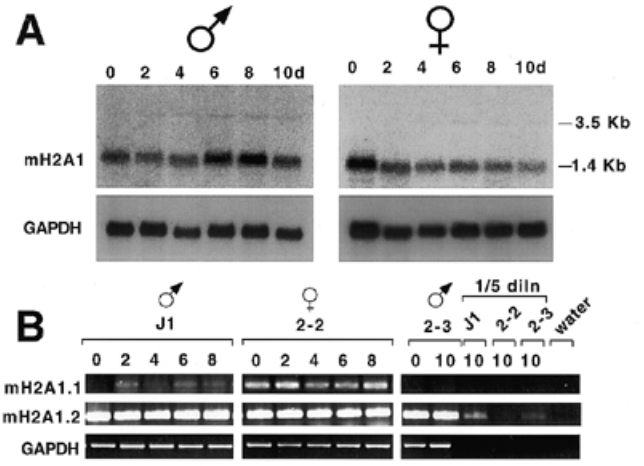

Figure 4. (**A**) Northern blot of mH2A1 RNA extracted from male and female ES cells differentiated from 0 to 10 days. GAPDH is shown as a loading control. The majority of mH2A1 is expressed as a 1.4 kb transcript. A less abundant transcript of 3.5 kb was also detected. (**B**) Changes in mH2A1 alternative splicing upon differentiation in male and female ES cells as judged by semi-quantitative RT–PCR. Primers 3F and 2R were used to amplify mH2A1.1-specific products and 4F and 2R were used to amplify mH2A1.2-specific exons. GAPDH-specific products were amplified with primers GAPDH1 and GAPDH2. Male ES cell line J1 is derived from *M.musculus* 129 blastocysts. Lines 2─2 and 2–3 are derived from *M.musculus*/*M.castaneus* hybrid blastocysts. 1/5 dilutions of first strand cDNAs gave rise to reduced amounts of RT─PCR products, indicating that assays were performed in a semi-quantitative range.

male ES cell lines contained low levels of mH2A1.1 message when grown under high LIF conditions to suppress differentiation. Upon removal of LIF, male cell lines J1 and 2–3 exhibited increased accumulation of the mH2A1.1 spliced RNA, though the amounts of this spliced form never reached those observed in female ES cells. Male ES cell line 2–3 (Fig. 4B) and female ES cell line 2–2 are derived from blastocysts from the same *M.musculus/M.castaneus* cross. Our results suggest that female ES cells may synthesize a higher level of mH2A1.1 message.

DISCUSSION

In this paper we show that the two previously described protein isoforms of mH2A1 originate by alternative splicing of primary transcripts generated from the mH2A1 locus that resides on chromosome 13. Because the mH2A1.2 isoform of the protein localizes to Xi (2), it seems possible that regulation of alternative splicing of mH2A1 could be involved in X chromosome inactivation. The alternatively spliced exons 6 and 7 include a putative leucine zipper (1) that begins in exon 5 suggesting that alternative splicing of the mH2A1 message may result in two protein variants with distinct leucine zipper binding specificities.

The finding that the absolute amounts of mH2A1 message are similar in both sexes suggests that mH2A1 is required for nuclear processes that are unrelated to X inactivation. For example, mH2A1 could be an important component of heterochromatin and it may be this role that causes it to localize to the inactive X chromosome. It is intriguing that mH2A1 is strongly expressed in testes which is the only male tissue where Xist is expressed and the X chromosome is inactivated (13,14). It seems possible that mH2A1 protein is recruited to the sites of heterochromatin formation during spermatogenesis in a process that might involve Xist RNA. The alternative splicing of mH2A1 in embryonic stem cells may differ in a sex-dependent fashion. We showed that mH2A1.2 (the most abundant form) is expressed at similar levels in ES cells of both sexes, but that the minor spliced mH2A1.1 form is expressed at a higher level in female as opposed to male cells. The significance of this observation is at present unclear. It has been shown that mH2A1 protein is expressed at a very low level in undifferentiated ES cells but is upregulated upon differentiation (15). In contrast, we observed that mH2A1 RNA is rather abundant in undifferentiated ES cells. This raises the possibility that mH2A1 RNA, while synthesized in undifferentiated ES cells, is not translated until differentiation occurs. However, we cannot exclude the possibility that the mH2A1 message we detected may be caused by a small subpopulation of differentiating cells present in our cultures.

mH2A1 has been suggested to play a role in X chromosome inactivation. The cloning and molecular elucidation of the mouse mH2A1 gene will enable us to directly test this hypothesis using gene targeting approaches.

ACKNOWLEDGEMENTS

We wish to thank Anton Wutz for supplying BAC and P1 probes for the chromosomal assignment of mH2A1 and Jessie Dausman for help with mouse methods. R.J. was supported by NIH/NCI grant 5-R35-CA44339.

REFERENCES

1. Pehrson,J.R. and Fried,V.A. (1992) *Science*, **257**, 1398–1400.
2. Costanzi,C. and Pehrson,J.R. (1998) *Nature*, **393**, 599–601.
3. Brown,C.J., Ballabio,A., Rupert,J.L., Lafreniere,R.G., Grompe,M., Tonlorenzi,R. and Willard,H.F. (1991) *Nature*, **349**, 38–44.
4. Brockdorff,N., Ashworth,A., Kay,G.F., Cooper,P., Smith,S., McCabe,V.M., Norris,D.P., Penny,G.D., Patel,D. and Rastan,S. (1991) *Nature*, **351**, 329–331.
5. Brown,C.J., Hendrich,B.D., Rupert,J.L., Lafreniere,R.G., Xing,Y., Lawrence,J. and Willard,H.F. (1992) *Cell*, **71**, 527–542.
6. Penny,G.D., Kay,G.F., Sheardown,S.A., Rastan,S. and Brockdorff,N. (1996) *Nature*, **379**, 131–137.
7. Marahrens,Y., Panning,B., Dausman,J., Strauss,W. and Jaenisch,R. (1997) *Genes Dev.*, **11**, 156–166.
8. Csankovszki,G., Panning,B., Bates,B., Perhson,J.R. and Jaenisch,R. (1999) *Nat. Genet.*, **22**, 323–324.
9. Shi,Y.P., Mohapatra,G., Miller,J., Hanahan,D., Lander,E., Gold,P., Pinkel,D. and Gray,J. (1997) *Genomics*, **45**, 42–47.
10. Robertson,E.J. (ed.) (1987) *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. The Practical Approach Series*. IRL Press, Oxford.
11. Lee,J.T., Strauss,W.M., Dausman,J.A. and Jaenisch,R. (1996) *Cell*, **86**, 83–94.
12. Lee,Y., Hong,M., Kim,J.W., Hong,Y.M., Choe,Y.K., Chang,S.Y., Lee,K.S. and Choe,I.S. (1998) *Biochim. Biophys. Acta*, **1399**, 73–77.
13. Richler,C., Soreq,H. and Wahrman,J. (1992) *Nature Genet.*, **2**, 192–195.
14. Salido,E.C., Yen,P.H., Mohandas,T.K. and Shapiro,L.J. (1992) *Nature Genet.*, **2**, 196–199.
15. Pehrson,J.R., Costanzi,C. and Dharia,C. (1997) *J. Cell Biochem.*, **65**, 107–113.