His-1: A noncoding RNA implicated in mouse leukemogenesis

Fan Xu, Molly McFarland and David S. Askew*

Department of Pathology and Laboratory Medicine, University of Cincinnati College of Medicine, PO Box 670529, Cincinnati, OH 45267-0529, USA

The *His-1* gene is highly conserved among vertebrate species and is transcribed as a single spliced and polyadenylated cytoplasmic RNA that shares several features in common with the emerging class of untranslated RNAs. A role for the *His-1* gene in neoplastic transformation was first indicated by the identification of transcriptionally activated *His-1* genes in a series of mouse leukemias, and more recent studies with antisense *His-1* RNAs suggest that *His-1* is involved in an oncogenic pathway that controls cell cycle progression. The *His-1* gene is normally regulated in a developmental and tissue-specific manner, and functional analysis of the promoter region has identified gene silencing as one of the mechanisms responsible for the narrow tissue distribution of the RNA. Although *His-1* function is incompetely understood at the present time, the expression of the gene at an early stage of choroid plexus development suggests that future strategies to delete the *His-1* gene in the whole animal will provide insight into the role of this RNA in mammalian development.

ALTHOUGH the literature on protein-coding RNA transcripts is extensive, much less is known about a small group of RNAs which, although they are also transcribed by RNA polymerase II and undergo typical post-transcriptional processing events, do not serve as templates for the synthesis of a functional protein (reviewed in refs 1, 2). It is now clear that the existence of RNA-based mechanisms of gene function is an evolutionarily ancient phenomenon, and examples of noncoding RNAs have been reported in diverse species, including mammals[3–12], *Drosophila*[13–15], *Xenopus*[16], *C. elegans*[17,18], yeast[19] and bacteria[20]. The functions of these RNAs are broad, and sometimes incompletely understood, but several of them are providing important insights into new mechanisms of genomic imprinting, X-chromosome inactivation, germ cell formation, meiosis, heat shock, oxidative stress, and in some cases disease. The *His-1* RNA was the first example of a mammalian noncoding RNA to be implicated in the pathogenesis of mouse leukemias induced by the nonacute transforming retroviruses[21]. The purpose of this review is to summarize the current understanding of the structure and function of the *His-1* gene, and to outline the data supporting a role for the *His-1* RNA in oncogenesis.

*For correspondence. (e-mail: david.askew@uc.edu)

What is the evidence that the *His-1* gene transcribes a noncoding RNA?

The mouse *His-1* gene is located at the proximal end of chromosome 2, and its human homolog maps to a syntenic region on chromosome 2q14-q21 (ref. 21). The mouse gene is the most characterized of the two homologs, and has been shown to transcribe a single 3 kb polyadenylated RNA from 3 exons spanning 6 kb (Figure 1). The *His-1* gene is typical of most other eukaryotic structural genes. All known splice junctions conform to consensus sequences established for mammalian splice donors and acceptors, and each of the two major splice variants, differing in size by only 50 bp, are equally represented in cells that express the gene[12]. The 3' end of the *His-1* RNA contains a polyadenylation signal upstream of a poly(A) addition site, and this signal is preceded by a 19 bp GT-rich sequence that has been shown in previous studies to be important for the cleavage reaction[22]. Exon 3 contains nine copies of the mRNA destabilization motif, AUUUA. This motif interacts with the AU-binding factor and destabilizes mRNA[23], suggesting that *His-1* RNA levels may be regulated by mechanisms that involve RNA stability. In contrast to the unremarkable organization of the *His-1* gene, the sequence of the *His-1* cDNA reveals several unusual features that are shared by the small group of polyadenylated noncoding RNAs that are believed to function in the absence of an encoded protein product. Since the occurrence of noncoding RNAs in nature appears to be rare, it is worth considering what features define them as a group so that we can more readily distinguish these RNAs, and perhaps gain insight into sequence or structural features that are required for their functional activity. Although there is as yet no single unifying characteristic among these RNAs, they show several atypical sequence features that are often considered indicative of noncoding character. A discussion of

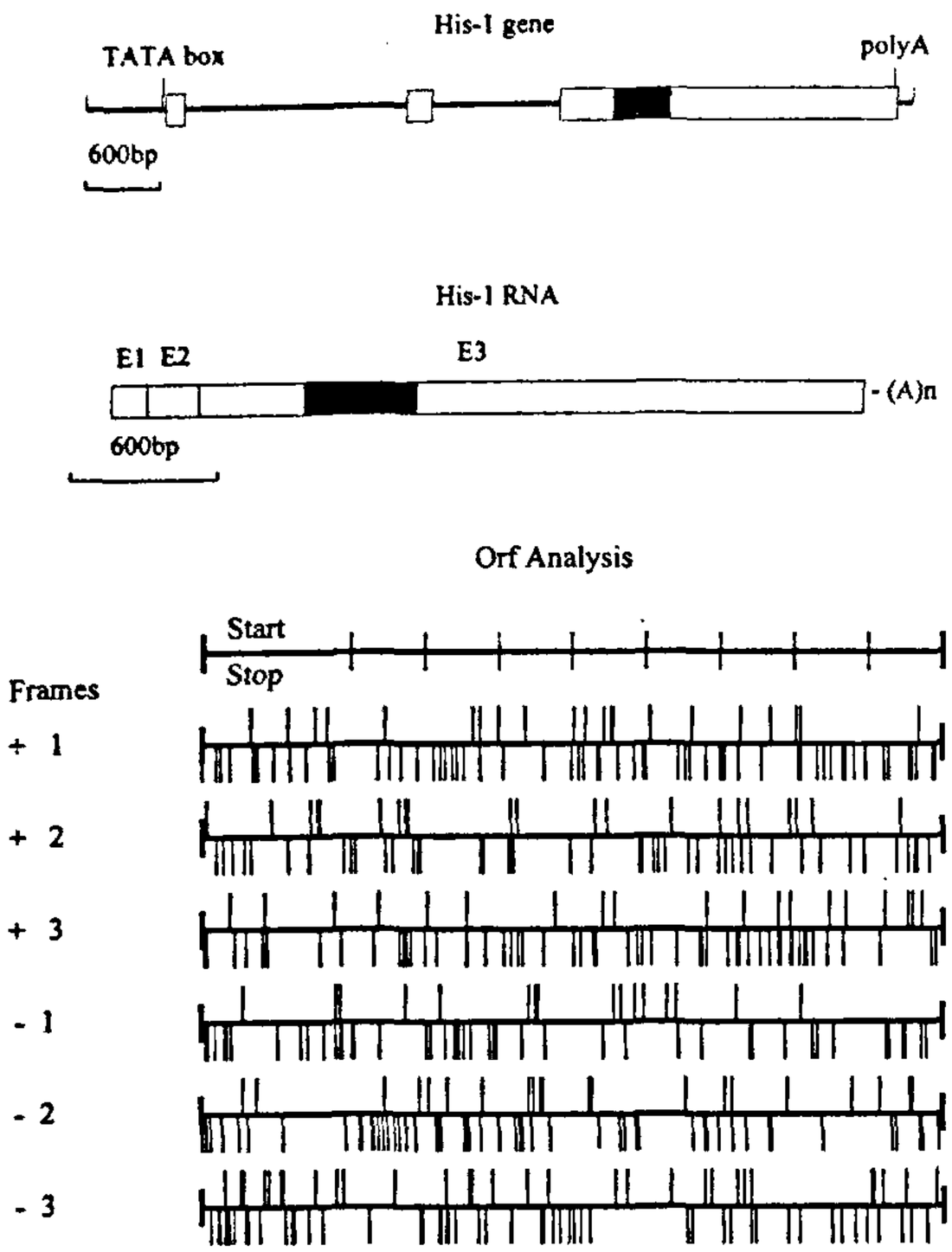

Figure 1. A schematic representation of the spliced and polyadenylated *His-1* RNA is shown below a map of the genomic organization of the *His-1* gene. Analysis of the orfs in all six reading frames is shown at the bottom of the figure (DNAMAN software, Lynnon Biosoft, Montreal). The location of every initiator codon (AUG) in the RNA sequence is shown by the vertical lines above the horizontal for each reading frame, and the location of each stop codon is indicated by the vertical lines below the horizontal.

how the *His-1* RNA shares each of these features with other examples of noncoding RNA is presented below.

The His-1 RNA lacks a functional orf

One of the most obvious differences between a typical protein-coding mRNA and noncoding RNAs such as *His-1* is the lack of a clearly identifiable single long open reading frame (orf). An extensive orf can not be identified in the *His-1* RNA because of a high density of stop codons in all three reading frames (Figure 1). Even if one takes into consideration the possibility that CTG is used as the initiator, the high density of stop codons in all known splice variants of the *His-1* RNA is incompatible with the presence of a single extended orf. None of the *His-1* orfs are longer than 213 bp in length, and the alternative splice donor used in exon 2 does not significantly lengthen the orfs which span that junction (Figure 1).

Although the presence of multiple small orfs in an RNA sequence is often considered the hallmark of noncoding RNA, the possibility that a functional peptide is made by a small orf must be taken into consideration before concluding that an RNA does not serve as a protein-coding template. Since it is practically impossible to unequivocally prove lack of coding function for every single orf in a large RNA, sequence analysis is used to identify the most likely candidate orfs. In the case of the *His-1* RNA, the longest orf would encode a peptide of 73 amino acids, but this orf contains a direct tandem repeat that is atypical of protein coding genes, and it does not translate a protein in rabbit reticulocyte or wheat germ lysates[12,24]. Moreover, this orf is located downstream of multiple ATGs, and the inhibitory effects of upstream ATGs on the usage of downstream ATGs makes it unlikely that this orf is functional *in vivo*[25].

GRAIL analysis predicts that the His-1 RNA is noncoding

Protein-coding exons in eukaryotic genomes share several sequence features that can be recognized by protein coding prediction algorithms. The absence of these features is therefore a useful marker for noncoding character, and this criterion can be effectively applied to the assessment of coding potential in both genomic and cDNA sequences[3,4,25]. Analysis of the *His-1* sequence with GRAIL, a neural network that recognizes features that are indicative of coding potential in mammalian genes[26], failed to identify significant protein coding regions in either the mouse or human *His-1* genes[12,24]. This finding strengthens the conclusion that the *His-1* RNA does not encode a protein, and underscores the atypical sequence information that is contained in this RNA. As more examples of noncoding RNAs are identified, it will be important to determine if these pattern-recognition programs can identify sequence features that are shared by other genes for noncoding RNAs, or provide insight into the mechanisms by which these RNAs function.

The His-1 gene lacks a conserved orf

Perhaps the strongest evidence that none of the *His-1* orfs are translated *in vivo* is provided by sequence comparisons between the mouse and human homologs. The rationale behind this approach is based upon the fact that highly conserved genes from different species will encode related proteins from a single conserved orf[11]. If there is no evolutionary pressure to restrict the divergence of a protein product, as would be expected for an untranslated RNA, the individual orfs can diverge significantly, even if the nucleic acid sequence is highly conserved. By applying this rationale to the analysis of

the mouse *His-1* gene, we were unable to identify any orf that would encode a related peptide in the human homolog[25]. These observations have led to the conclusion that, although evolutionary constraints have restricted divergence in the *His-1* nucleic acid sequence, it is unlikely that any of the *His-1* orfs are translated into a protein that has a conserved function *in vivo*.

Does the *His-1* gene play a role in oncogenesis?

Initial interest in the *His-1* RNA began with the discovery that the *His-1* gene is transcriptionally activated in mouse leukemias induced by the nonacute transforming retroviruses[12,21]. The mechanism by which these retroviruses cause leukemia involves integration of a DNA provirus into the host genome and the consequent activation of a flanking cellular protooncogene by either direct mutation or transcriptional activation (reviewed in refs 27, 28). Since retroviral integration is essentially a random event, the search for nonrandom viral insertion sites in multiple independently derived tumours has been a very successful strategy to identify cellular protooncogenes that contribute to the neoplastic process[27,28]. The *His-1* gene was identified by applying this insertional mutagenesis approach to a panel of mouse leukemias induced by Moloney-Murine Leukemia virus (MoMuLV)[12,21]. Retroviruses were shown to insert into the first intron of the *His-1* gene, thereby driving high levels of *His-1* RNA from the strong promoters located in the viral long terminal repeats (LTRs)[12]. A total of six leukemic cell lines have been identified to express the *His-1* RNA, two of which lack inserted retroviruses and therefore express the gene from the normal promoter. Since the *His-1* gene is normally silent in hematopoietic cells[29], its association with the leukemic phenotype suggests that inappropriate activation of this gene is causally related to the induction and/or progression of mouse leukemia.

Since noncoding RNAs are encountered infrequently in mammalian cells[1,2], it is not surprising that the vast majority of known retrovirally-activated proto-oncogenes encode proteins[27,28]. The discovery that the *His-1* gene is also a target for retrovirus-induced gene activation was the first to include a noncoding RNA in the large list of growth factors, growth factor receptors, protein kinases, GTP-binding proteins, and transcription factors that play a role in tumours induced by these retroviruses[27,28]. The recent identification of a second example of a retrovirally activated noncoding RNA in B-cell lymphomas induced by Avian Leukosis Virus (ALV) provides further evidence that noncoding RNA can influence oncogenic pathways[7]. The transcription of the *bic* gene was activated in these lymphomas by a similar retroviral insertional mutagenesis mechanism to that of the *His-1* gene, resulting in high levels of the noncoding *bic* RNA in tumours with high metastatic potential[7]. Although there are no shared sequence motifs between *His-1* and *bic*, it will be of interest to determine if the contribution of these RNAs to leukemogenesis involves an overlapping oncogenic mechanism.

Since most genes involved in leukemogenesis affect signaling pathways that control cell cycle regulation and/or differentiation[27,28], we have examined the *His-1* RNA for its ability to interfere with these processes in hematopoietic cells. To accomplish this, the *His-1* cDNA was inserted into a mammalian expression vector and introduced into hematopoietic cell lines that do not express the endogenous *His-1* gene. The constitutively expressed *His-1* cDNA had no detectable effect on growth rate however, and we have observed no alterations in the ability of the *His-1*-transfected cells to differentiate in response to cytokine stimulation[30]. These results indicate that *His-1* overexpression alone is insufficient to induce visible neoplastic change in hematopoietic cells, suggesting that additional cooperating mutations may be required to reveal the oncogenic effects of this RNA.

To determine if the *His-1* RNA is contributing to cell growth in leukemic cells that harbour activated *His-1* genes, we have used an antisense inhibition approach to interfere with the biological activity of the *His-1* RNA. An antisense construct was therefore designed to express a 500 bp region from exon 3 in the antisense orientation (Figure 1) under the control of the CMV promoter. We reasoned that enforced expression of the most evolutionarily conserved segment of exon 3 in the antisense orientation would inhibit *His-1* RNA function by creating an RNA–RNA hybrid in the region that is most likely essential for functional activity. Recent data has shown that the antisense *His-1* expression vector inhibits the growth of cells that express the endogenous *His-1* gene, suggesting that transcriptional activation of the *His-1* gene in hematopoietic cells is involved in an oncogenic pathway that affects the cell cycle[30]. We therefore speculate that overexpression of *His-1* disrupts normal cell cycle regulation in the leukemic cells. Current efforts are focussed on dissecting the mechanism by which this RNA affects the cell cycle machinery using an inducible antisense expression vector system.

The *His-1* gene has a narrow range of expression

Northern blot analysis and *in situ* hybridization have shown that the *His-1* RNA has a very limited distribution in normal mouse tissues[29]. These analyses were performed by hybridization to a *His-1* probe comprising the conserved 500 bp segment of exon 3 (Figure 1, blackened region). This probe is highly specific for the *His-* gene since it does not detect any related sequences in the mouse genome by genomic Southern blot analysis even under conditions of reduced stringency[12,25]. *In sit*

hybridization with the antisense *His-1* riboprobe was performed on sections of the developing mouse embryo from days 10 through 19 of gestation and on various adult mouse tissues[29]. The highest levels of expression were identified in the developing choroid plexus from a d.12 embryo, a time when the choroid plexus is just beginning to develop[31]. Hybridization was specific for the epithelial cells lining the developing choroid plexus, but was notably absent from the contiguous ependymal cells that line the ventricle. The fully developed choroid plexus in the adult animal is located within the ventricles of the brain, and it is responsible for the secretion of cerebrospinal fluid (CSF)[32]. No *His-1* transcripts were identified in the adult choroid plexus however, suggesting that *His-1* may have a role in the development of this organ. In the adult mouse, *in situ* hybridization detected *His-1* RNA in the epithelial cells of the prostate, seminal vesicle, and nonglandular portion of the stomach, but these levels were too low to be detectable by Northern blot analysis of the same tissues[29]. The *His-1* gene is therefore silent in the majority of mouse tissues, suggesting that the activity of its promoter is constrained by both developmentally regulated and tissue-specific factors. These observations are consistent with the concept that inappropriate activation of *His-1* in the hematopoietic cell lineage, either by retroviral insertional mutagenesis or some other mechanism, is one of several mutational events that contribute to leukemic transformation.

Regulation of the *His-1* promoter by positively and negatively acting factors

As a first step towards identifying functional domains in the *His-1* gene we have searched for regions that are the most evolutionary conserved across species. Cross hybridization studies have shown that *His-1* is conserved as a single-copy sequence in human, mouse, cat, pig, cow, dog, and chicken genomic DNA[12,25]. No related sequences were detectable in *Drosophila* however, suggesting that *His-1* is a feature of vertebrate species[25]. A comparison between the mouse and human genes reveals that both are A/T rich, with several regions of high sequence conservation interspersed with regions of divergence. The most extensive areas of homology are located in the 5′ flanking region, intron 1, intron 2, and exon 3 (ref. 25). Since the degree of evolutionary variance tolerated by intronic regions is usually quite high, it is striking that some of the most conserved segments in the *His-1* gene are located within introns. A sequence alignment between the two *His-1* homologs is consequently remarkably colinear, with no major deletions or insertions interrupting the two sequences[25]. Colinearity between species homologs is similar to what has been reported for the noncoding RNA *H19* (ref. 11), but is different from homologs of the noncoding RNA *Xist* where organizational conservation is more limited due to the presence of multiple insertions and deletions that have occurred during the evolution of the gene[3,4]. The most conserved *His-1* exonic region is found in a 500 bp sequence located at the 5′ end of exon 3 (Figure 1), and a more discontinuous region of homology is dispersed across 700 bp at the 3′ end of this exon[25]. Stop codons are arranged in the human sequence at a similar density to that of the mouse gene, suggesting that a high density of stop codons has functional relevance to each of these homologs. A known variant of the TATA box is located 36 bp upstream of the first of the three major transcription start points, and a consensus CCAAT box is located further upstream. Both of these sequences are present in the mouse and human genes, suggesting that they play a role in the function of this region in both species[25]. These sequence comparisons suggest that the *His-1* gene contains important regulatory sequences in the promoter and intronic regions, and that exon 3 contains a 500 bp domain that is likely to be relevant to the function of the transcribed RNA.

By cloning various segments of the *His-1* promoter upstream of a luciferase reporter gene we have defined the proximal *His-1* promoter as the 300 bp region upstream of the transcription start point, and we have identified a short positive regulatory element that is required for optimal activity of the promoter[33]. A potent silencer element was also identified in the first exon of the gene, and the inclusion of this sequence into the proximal promoter fragment was sufficient to silence the activity of the minimal promoter in cells that do not normally express the *His-1* RNA. A factor that binds to this silencer element in gel-shift assays has been identified in cells lacking *His-1* RNA, but this factor is notably absent in cell lines that express the *His-1* gene from the normal promoter[33]. These results indicate that cell type-specific negative regulation is an important mechanism used to restrict the transcription of the *His-1* gene, and implies that gene silencing effects can be overcome by retroviral promoters in leukemic cells that express *His-1* RNA.

What is the normal function of the *His-1* RNA?

One of the biggest challenges to studying any uncharacterized gene is elucidating its normal function. This is particularly difficult for *His-1* since it has no related family members in the sequence databases and there is no available data on the functional significance of RNA sequence motifs among other RNAs that lack coding potential. Furthermore, although recent antisense inhibition studies have suggested that *His-1* RNA may contribute to dysregulated growth in leukemic cells[30], the normal function of the *His-1* gene has remained elusive. To gain more insight into the biological role of the *His-1* gene, we have therefore taken a whole-animal approach. Ongoing studies are assessing the phenotypic

consequences of directing *His-1* expression to mouse tissues that are outside of the gene's normal range of expression. This has been accomplished by creating a *His-1* transgenic mouse that expresses the *His-1* cDNA from the widely expressed β-actin promoter[34,35]. Since the normal pattern of *His-1* gene expression is very narrow, we hypothesize that ectopic expression of the *His-1* RNA, particularly within the hematopoietic lineage, will allow us to more definitely characterize the biological function of this RNA and to understand its contribution to leukemogenesis. We have obtained ten *His-1* transgenic founders, and at the time of writing there is no obvious pathology associated with the presence of the *His-1* transgene. The availability of a *His-1* transgenic mouse provides a unique resource to address the long term consequences of overexpressing this RNA in mouse tissues, particulary with respect to alterations in predisposition to spontaneous or experimentally-induced neoplasms.

Based upon the extensive conservation of the *His-1* gene, and its expression in the developing choroid plexus epithelium, it is reasonable to hypothesize that the normal function of the RNA is related to the development and/or embryonic function of the vertebrate choroid plexus. To determine if the *His-1* gene is required for choroid plexus development, we are currently attempting to delete the *His-1* gene in embryonic stem cells and to subsequently create a mouse strain that harbours this same deletion. The deletion strategy will replace all 3 exons of the *His-1* gene, including about 600 bp of 5′ flanking sequence, with the neo^R gene. Similar strategies that involve extensive gene deletions provided important information about the functions of the noncoding RNAs *H19* and *XIST*[36,37]. Although the lack of functional data on the *His-1* RNA make it difficult to speculate on the phenotype of a *His-1* null mouse, we anticipate that the deletion of this gene will significantly advance our understanding of this highly conserved segment of chromosome 2.

Conclusion

Interest in the biology of untranslated RNA is continuing to expand as increasing numbers of such RNAs are identified and shown to have key roles in the regulation of cell function. The *His-1* RNA is now one of two noncoding RNA molecules that are activated by proviruses in leukemias induced by the nonacutely transforming retroviruses, suggesting that noncoding RNAs can also influence disease pathways when their normal functions are altered. Although much more work is needed to clarify the function of the *His-1* RNA, further studies on this and other noncoding RNAs has the potential to open new areas of investigation into pathways of cell function that may be relevant to human disease.

1. Askew, D. S. and Xu, F., *Histol. Histopathol.*, 1999, **14**, 235–241.
2. Lakhotia, S. C., *Indian J. Biochem. Biophys.*, 1996, **33**, 9
3. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., N P., Cooper, P. J., Swift, S. and Rastan, S., *Cell*, 1992, **71**, 515
4. Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. C Y., Lawrence, J. and Willard, H. F., *Cell*, 1992, **71**, 527–542.
5. Maxwell, E. S. and Fournier, M. J., *Annu. Rev. Biochem.* **64**, 897–934.
6. Rastinejad, F. and Blau, H. M., *Cell*, 1993, **72**, 903–917.
7. Tam, W., Ben-Yehuda, D. and Hayward, W. S., *Mol. Ce* 1997, **17**, 1490–1502.
8. Tycowski, K. T., Shu, M. D. and Steitz, J. A., *Nature* **379**, 464–466.
9. Velleca, M. A., Wallace, M. C. and Merlie, J. P., *M Biol.*, 1994, **14**, 7095–7104.
10. Wevrick, R., Kerns, J. A. and Francke U., *Hum. Mol.* 1994, **3**, 1877–1882.
11. Brannan, C. I., Dees, E. C., Ingram, R. S. and Tilghman, *Mol. Cell. Biol.*, 1990, **10**, 28–36.
12. Askew, D. S., Li, J. and Ihle, J. N., *Mol. Cell Biol.*, 19 1743–1751.
13. Lakhotia, S. C. and Sharma, A., *Genetica*, 1996, **97**, 339-
14. Nakamura, A., Amikura, R., Mukai, M., Kobayashi, Lasko, P. F., *Science*, 1996, **274**, 2075–2079.
15. Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. and R. L., *Cell*, 1997, **88**, 445–457.
16. Kloc, M. and Etkin, L. D., *Science*, 1994, **265**, 1101–110
17. Goodwin, E. B., Okkema, P. G., Evans, T. C. and Kim *Cell*, 1993, **75**, 329–339.
18. Lee, R. C., Feinbaum, R. L. and Ambros, V., *Cell*, 19 843–854.
19. Watanabe, Y. and Yamamoto, M., *Cell*, 1994, **78**, 487–49
20. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, Storz, G., *Cell*, 1997, **90**, 43–53.
21. Askew, D. S., Bartholomew, C., Buchberg, A. M., Va M. B., Jenkins, N. A., Copeland, N. G. and Ihle, J. N., *gene*, 1991, **6**, 2041–2047.
22. Wickens, M., *Trends Biochem. Sci.*, 1990, **15**, 277–324.
23. Akashi, M., Shaw, G, Hachiya, M., Elstner, E., Suzuki, Koeffler, P., *Blood*, 1994, **83**, 3182–3187.
24. Li, J., Rhodes, J. C. and Askew, D. S., *Gene*, 1997, **184**, 16
25. Kozak, M., *Mamm. Genom.*, 1996, **7**, 563–574.
26. Uberbacher, E. C. and Mural, R. J., *Proc. Natl. Acad. Sc* 1991, **88**, 11261–11265.
27. Askew, D. S., Bartholomew, C. and Ihle, J. N., *H Pathol.*, 1993, **7**, 1–22.
28. Jonkers, J. and Berns, A., *Biochim. Biophys. Acta.*, 1996 29–57.
29. Li, J., Witte, D. P., Van Dyke, T. and Askew, D. S., *Pathol.*, 1997, **150**, 297–1305.
30. McFarland, M., Xu, F. and Askew, D. S., unpublished dat
31. Kaufman, M. H., in *The Atlas of Mouse Development*, Ac Press, San Diego, 1992, pp. 173–174.
32. Nilsson, C., Lindvall-Axelsson, M. and Owman, C., *Bra Rev.*, 1992, **17**, 109–138.
33. Xu, F. and Askew, D. S., unpublished data.
34. Gunning, P., Leavitt, J., Muscat, G., Ng, Sun-Yu. and Ke *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4831–4835.
35. Deprimo, S. E., Stambrook, P. J. and Stringer, J. R., *Res.*, 1996, **5**, 459–466.
36. Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstr A. and Tilghman, S. M., *Nature*, 1995, **375**, 34–39.
37. Marahrens, Y., Panning, B., Dausman, J., Strauss, W Jaenisch, R., *Genes Dev.*, 1997, **11**, 156–166.

ACKNOWLEDGEMENTS. This work was supported by N Institutes of Health grant R29CA61909 to D.S.A and by an Am Cancer Society Junior Faculty Research Award to D.S.A.