Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 6829–6834, June 1999
Genetics

A new structure for the murine *Xist* gene and its relationship to chromosome choice/counting during X-chromosome inactivation

YOUNG-KWON HONG, SARA D. ONTIVEROS, CAIFU CHEN, AND WILLIAM M. STRAUSS*

Harvard Institute of Human Genetics, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA 02115

Edited by Thomas W. Cline, University of California, Berkeley, CA, and approved April 21, 1999 (received for review February 17, 1999)

ABSTRACT **In this report, we present structural data for the murine *Xist* gene. The data presented in this paper demonstrate that the murine *Xist* transcript is at least 17.4 kb, not 14.3 kb as previously reported. The new structure of the murine *Xist* gene described herein has seven exons, not six. Exon VII encodes an additional 3.1 kb of information at the 3′ end. Exon VII contains seven possible sites for polyadenylation; four of these sites are located in the newly discovered 3′ end. Consequently, it is possible that several distinct transcripts may be produced through differential polyadenylation of a primary transcript. Alternative use of polyadenylation signals could result in size changes for exon VII. Two major species of *Xist* are detectable by Northern analysis, consistent with differential polyadenylation. In this paper, we propose a model for the role of the *Xist* 3′ end in the process of X-chromosome counting and choice during embryonic development.**

Female mammals are mosaic for expression of X-linked genes with the clonal random silencing of gene expression from one X-chromosome (1–5). The process of murine X-chromosome inactivation is thought to involve not only the gene *Xist* (X inactive specific transcript) but also a locus called the *Xce* (X controlling element) (6, 7). Molecular, cellular, and transgenic evidence unequivocally proves that *Xist* is necessary for X-chromosome inactivation (8–11). The minimum genomic interval that recapitulates the phenomenon of X-chromosome counting, choice, and silencing can be encompassed on a 40-kb cosmid that contains 9 kb upstream and 6 kb downstream of *Xist* (10). In this report, we demonstrate that the widely disseminated structure for the murine *Xist* gene must be revised. The gene encodes a transcript that is larger than 14.3 kb. We further demonstrate that there are seven exons, not six. Finally, we show that an additional exon is found at the 3′ end and is located within the region shown to play a role in X-chromosome choice/counting (7, 12, 13).

Transgenic studies of X-chromosome inactivation have provided two conflicting lines of evidence. The first line of evidence would suggest that a cosmid spanning *Xist* encodes information to direct the process of X-chromosome inactivation (10). These complementation studies showed that in male embryonic stem cells, an ectopic *Xist* cosmid was essential for control of counting, choice, and silencing of a reporter gene. The other line of evidence, using the *cre/lox* deletional technology (12), would support the existence of another locus responsible for choice and counting. In these experiments, a 65-kb deletion downstream of *Xist* was shown to alter the transcript stability as well as the "choice" mechanism of X-chromosome inactivation. The *Xist* cosmid contained 6 kb of DNA downstream of the known 3′ end of *Xist*. The proximal *cre/lox* deletion site is within this 6-kb interval, 1 kb from the 3′ end of *Xist* (as described by Brockdorff *et al.,* ref. 14). The data from this report may reconcile these apparent contradictions by redefining the structure of *Xist*.

EXPERIMENTAL PROCEDURES

Reagents. YAC116 (9) was used as a genomic DNA control for all the PCR reactions. cDNA libraries were obtained as follows: female mouse lung (Stratagene, Catalogue no. 936307), female mouse brain (Stratagene, Catalogue no. 937319), female mouse uterus (CLONTECH, Catalogue no. ML1022B), male mouse brain (CLONTECH, Catalogue no. ML3000B), male mouse heart (CLONTECH, Catalogue no. ML1048B), mouse lung (CLONTECH, Catalogue no. ML1046B), and mouse testis (CLONTECH, Catalogue no. ML1020B). DNA sequences of the oligonucleotides used in this study are listed in Table 1.

PCR Amplification and Cloning. DNA fragments containing exons 6 and 7 were amplified either from YAC116 by using WS780 and WS758 or from the female mouse lung cDNA library by using WS780 and WS770. Each fragment was cloned into the pGEM-T vector (Promega), resulting in pWS850 and pWS873, respectively. Expressed Sequence Tag (EST) fragments were recovered from either YAC116, female mouse lung, male mouse brain, or male heart cDNA libraries by using the following primers: pWS811/WS813 for *Xist* + EST1, pWS812/WS815 for EST1 + EST2, pWS835/WS836 for EST2 + EST3, pWS837/WS838 for EST3 + EST4. These fragments were cloned into pBluescript II SK+ (Stratagene), resulting in pWS848, pWS849, pWS854, and pWS855, respectively. The individual ESTs were PCR amplified by using pWS812/WS813 (EST1), pWS814/WS815 (EST2), pWS816/WS817 (EST3), and pWS818/WS821 (EST4) and cloned into pBluescript II SK+, resulting in pWS857, pWS858, pWS859, and pWS860, respectively. All the cloned materials were sequenced by using T3 and T7 primers by the Beth Israel Deaconess Medical Center sequencing facility on an Applied Biosystems 377 sequencer.

DNA Blotting. Duplicate gels containing the PCR products with the primers, WS831/WS832 (*Xist* + EST1), WS833/WS834 (EST1 + EST2), WS835/WS836 (EST2 + EST3) and WS837/WS838 (EST3 + EST4) were transferred to nylon membrane and hybridized with probes spanning the 5′ and 3′ end of each fragment. Thus, *Xist* + EST1 was probed with both the *Xist* probe and the EST1 probe; EST1 + EST2, EST2 + EST3, and EST3 + EST4 fragments were analyzed similarly. The 270-bp product from a *Pst*I/*Bgl*II digestion of pWS850 was used as the probe for the 3′ region of *Xist*; probes for EST1, 2, 3, and 4 were prepared from *Bam*HI/*Sal*I digestions of pWS857, pWS858, pWS859, and pWS860, respectively. All the probes were radiolabeled by the random hexamer labeling method (15) in the presence of [^{32}P]dCTP and hybridized for

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: EST, expressed sequence tag; FISH, fluorescence *in situ* hybridization.
*To whom reprint requests should be addressed. e-mail: wstrauss@hihg.med.harvard.edu.

Table 1. Oligonucleotide sequences used in PCR analysis

Oligonucleotides	Location	DNA sequences (5′ to 3′)
WS758	Exon 7	CATTTCCTCATTGAAGTGAATTG
WS770	Exon 7	AAACAAGAATTCACTAAGGATAGAAGC
WS780	Exon 6	AATACACAATGCCATCTACCAAATATTA
WS812	EST1	TCCAGCCTTCTGAGTAAATATT
WS813	EST1	CCTCTTTTATTATTTCCACTCTA
WS814	EST2	GATGCTAAGTGCAACACAT
WS815	EST2	TCACAGTCATAGCTAAAATGG
WS816	EST3	GGGTGGGACAGGAGGCCG
WS817	EST3	AGATGATGGTAGGATGTGCTT
WS818	EST4	AGACTGGAGTCATCTTCCC
WS821	EST4	CAAACTACCCCACCCACTC
WS831	Old Xist end	CTTTGCTTTTATCCCAGGCA
WS832	EST1	CACTTTGGGTTGCATCCTTT
WS833	EST1	ATTTGTTCATTGCCTGGCTC
WS834	EST2	TCACACTGAGTGCCCTTTTG
WS835	EST2	GCTTTGTAGCAAGCCTGACC
WS836	EST3	TTCAACCTGGCTCCATCTTC
WS837	EST3	GAAGATGGAGCCAGGTTGAA
WS838	EST4	ATCCGTTCACAAAGTCCAGG
WS852	Downstream of EST4	GCTGGCCTACACGGGTATAA

12–24 hr at 65°C in a buffer containing 7% SDS, 2 mM EDTA (pH 7.6), and 0.5 M sodium phosphate (pH7.5).

Northern Blot Analysis. Total RNA was isolated by using the guanidinium thiocyanate method from kidneys of male and female mice (16). Total RNA was electrophoresed on 0.8% agarose gel containing 2.2 M formaldehyde for 16 hr at 100 V and then transferred to positively charged nylon, yielding duplicate matched lanes. Then, blots were hybridized with ^{32}P-labeled DNA fragments from either *Sac*II/*Sac*I digestion of pWS850 or *Cla*I/*Eco*RI of pWS854 (see Fig. 3).

RNA–Fluorescence *in Situ* Hybridization (FISH). Male and female fibroblasts were isolated from normal mice, grown on chamber slides, and fixed as described (17). Two plasmids were used as probes for RNA-FISH: EST 2/3 (pWS854, digoxygenin) or exons VI–VII (pWS850, biotin) (see Fig. 2*A*). Probes were hybridized either separately or simultaneously to the interphase spreads of murine male and female dermal fibroblasts. After washing, the spreads were labeled with anti-digoxygenin rhodamine or avidin-fluorescein. Images were collected with a Nikon E-800 microscope equipped with a Sensys (Photometrics, Tucson, AZ) digital CCD camera. Grayscale images for either FITC, rhodamine, or 4′,6-diamidino-2-phenylindole filter sets were pseudocolored and images were merged in the 12-bit format. The 12-bit data were compressed as 8-bit data during export to Photoshop 5.0 (Adobe Systems, Mountain View, CA) for final figure preparation (Fig. 4).

RESULTS

We examined the *Xist* transcript (GenBank accession no. L04961) and genomic sequence (GenBank accession no. U41394). This review brought to light discrepancies and required that the organization of the *Xist* gene be reevaluated. The predicted size and sequence of the established exon VI was in fact not in agreement with the apparent cDNA structure. The genomic sequence would have predicted that an additional 781 bp of sequence should be part of the cDNA. Cloning and sequencing the relevant regions from both the genomic and cDNA derived from PCR-amplified templates confirmed that this genomic sequence was absent in the cDNA (see Fig. 1 *A* and *B*). Consistent with this difference, splice donor and splice acceptor sequences could be identified. Despite the published (14) structure of *Xist*, the only logical conclusion was that exon VI was in fact two exons. These exons are therefore renamed exon VI and exon VII. Exon VI is 155 bp long.

A survey of the EST database was also conducted. During this survey, all mouse ESTs were mapped to the 94-kb sequence spanning the mapped location of the *Xce* (GenBank accession no. X99946). We determined that there were four ESTs that mapped between the published 3′ end of *Xist* and *Tsx* (see Fig. 2*A* and notes). In fact, all four of these female-specific ESTs were within 3.1 kb of the 3′ end of *Xist,* and no other EST sequences were observed to map to this interval. Further examination of these four female-specific ESTs revealed that they did not encode any significant ORFs as indicated by the DNA STRIDER program. In an effort to relate the significance of these ESTs to *Xist*, we constructed a series of PCR primers that spanned either the individual EST sequence, the 3′ end of *Xist* and EST1 (X + EST1), EST1 + EST2, EST2 + EST3, or EST3 + EST4. These primers were then used to screen cDNA libraries constructed from RNA derived from murine male or female soma. In particular, the female cDNA library was the one used to originally define the murine *Xist* structure (14). When primers spanning *Xist* and EST1 (X + EST1) were used, only a female-specific product was isolated (Fig. 2*B*). This result demonstrates the colinearity of *Xist* transcript with EST1. The isolation of female-specific PCR products for the combinations of primers termed EST1 + EST2, EST2 + EST3, EST3 + EST4 was also observed. These results demonstrate the colinearity of all the ESTs with each other and with *Xist* transcript. Our conclusion from these data was that the murine *Xist* transcript was in fact larger than originally defined, extending into the genomic region demarked by EST1–EST4.

We wished to determine whether sequences downstream of EST4 were also incorporated into the *Xist* transcript. Two sets of PCR primers were used in combination to attempt to recover cDNA material from three female as well as four male cDNA libraries (see Fig. 2*A* and data not shown). No products were detected in these experiments.

No size difference between genomic or cDNA template was observed with primers spanning *Xist* + EST1, EST1 + EST2, EST2 + EST3, EST3 + EST4. All PCR products were subject to complete sequence analysis. The results of this analysis confirmed that the PCR products were in fact derived from the region spanned by EST1–EST4. Several interesting structural features were observed: four additional polyadenylation signals, as well as two potential stem loops. This suggested that the *Xist* transcript was not only extending into the EST1–EST4

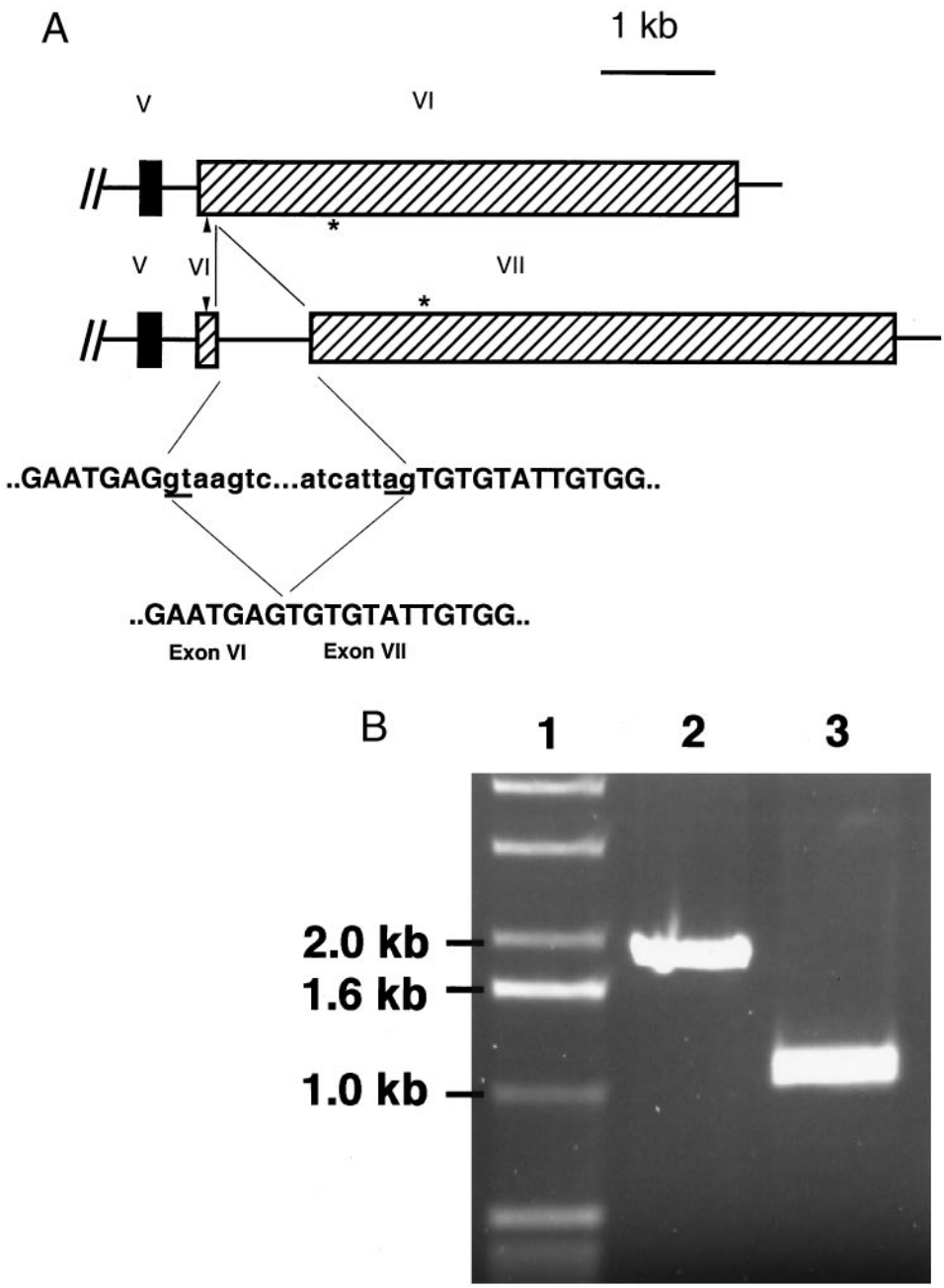

FIG. 1. The murine *Xist* gene consists of seven exons. (*A*) Revised genomic structure of *Xist* contains a 781-bp-long intron compared with the documented structure of exon VI by Brockdorff *et al.* (14). Exon VI/exon VII junction sequences are shown. Arrowhead, *Ava*I; asterisk, *Eco*RI. (*B*) Comparison of *Ava*I/*Eco*RI fragments from mouse genomic DNA, Y116 (lane 2), and female mouse lung cDNA (lane 3). PCR fragment spanning the region was cloned and sequenced. Lane 1 is 1-kb DNA size marker (GIBCO/BRL).

region, but that the portion of the transcript derived from this region could be differentially processed; however, no evidence for differential splicing was observed (see Fig. 2*B*).

To determine the size and complexity of the *Xist* transcript(s) encompassing the new 3′ end, Northern blots were used. Somatic RNA was extracted from murine female and male kidney. After fractionation on denaturing agarose gels, the resulting blots were hybridized either to a probe corresponding to exons VI–VII (pWS850) or to a probe corresponding to the region spanned by EST2 and 3 (pWS854) (see Fig. 3). The figure shows two major species of *Xist* using the pWS850 probe. The new 3′-end probe, pWS854, hybridizes disproportionately to the larger of the two major species of *Xist* RNA.

RNA–FISH was performed to determine whether the new 3′ end colocalized with the established sequences of the murine *Xist* transcript on the inactive X-chromosome found in female cells (see Fig. 4). The same two DNA probes used for Northern analysis were used in this experiment. The data from these experiments show that the probe (pWS854) corresponding to the 3′ end colocalizes with the rest of the *Xist* transcript (pWS850) on the inactive X-chromosome. Thus the new 3′ end of the murine *Xist* gene is associated with *Xist* molecules, which correctly localize in a functionally significant manner.

DISCUSSION

One conclusion of the data presented here is a redefinition of the exonic/intronic structure of *Xist*. The cDNA, as originally

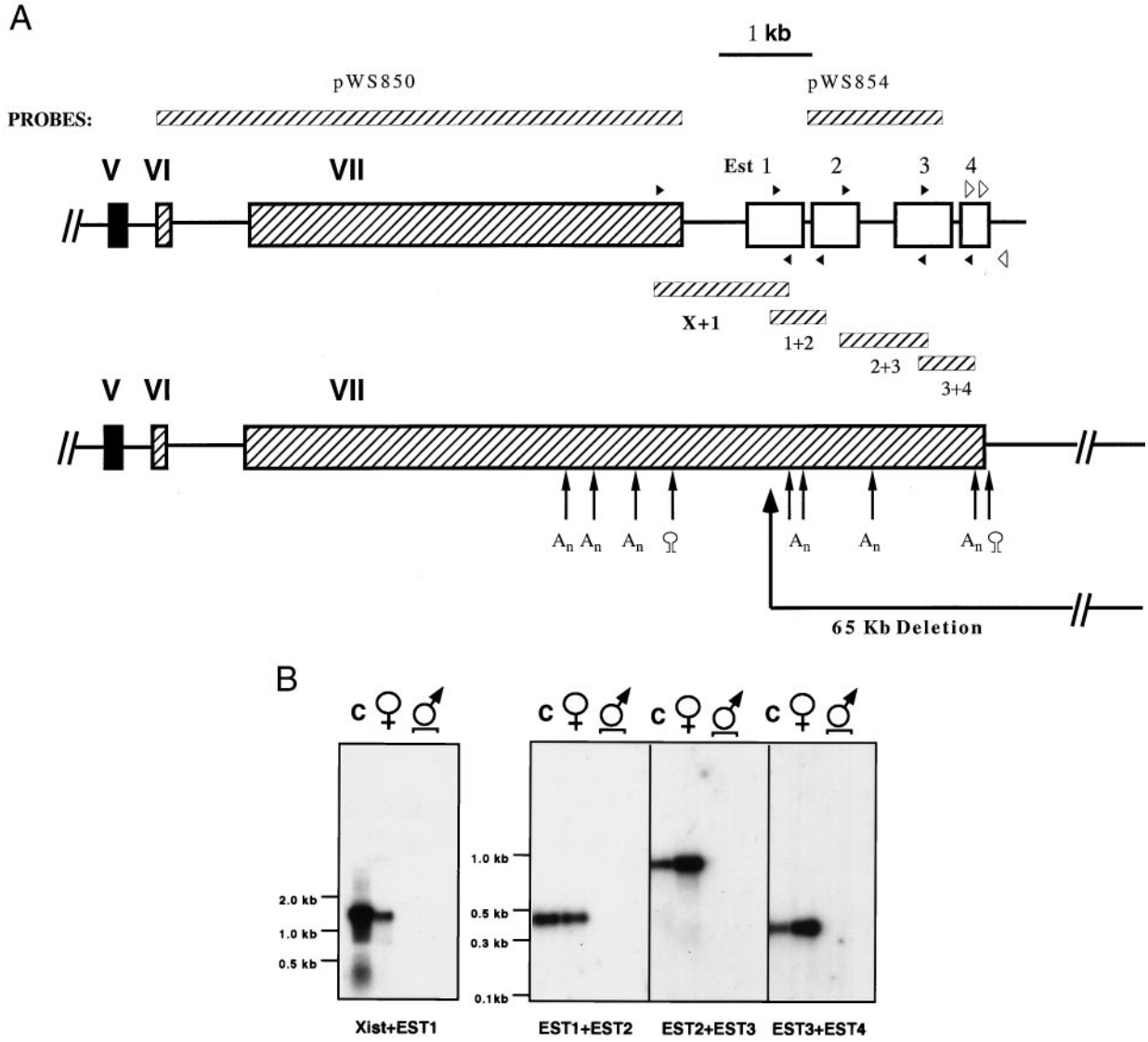

FIG. 2. New structure of the 3′end *Xist*. (*A*) The four ESTs (EST1≈4) are mapped relative to exon VII of *Xist* (GenBank accession nos. of EST1 ≈ 4: AA543875, AA221611, AA690387§, and R74734, respectively). PCR fragments synthesized by using primers whose locations are marked by closed arrowheads are shown to demonstrate the colinearity of all the ESTs. Locations of primers used to determine the end of the *Xist* transcript are marked by open arrowheads. Probes used for Northern blots and RNA–FISH (pWS854, pWS850) are also indicated. Consensus sequences for polyadenylation (A_n) and sequences of putative stem and loop structures are also localized. The 65-kb deletion created by Clerc and Avner (12) begins from the *Sca*I site (marked with heavy arrow) in the EST1. EST fragments were recovered as described in *Experimental Procedures*. (*B*) All the ESTs are colinear with *Xist*. All PCR products were sequenced. For the purpose of this figure, the PCR fragments for *Xist* + EST1 and for EST1 + EST2, EST2 + EST3, and EST3 + EST4 were electrophoresed in 0.8% and 2% agarose gels, respectively, transferred to nylon membranes, and hybridized with individual fragments (probes for this figure, *Xist*, EST1, EST2, and EST3, respectively). (C) YAC116 (genomic DNA); ♀, female mouse lung cDNA library; ♂, male mouse brain and male heart cDNA libraries. Approximate DNA sizes are marked by using either 1-kb marker (GIBCO/BRL) for 0.8% gel or 100-bp marker (NEB, Beverly, MA) for 2% gel. The top of each lane is the origin of migration.
§Note: Accession no. AA690387 is incorrectly identified as derived from a male mouse cDNA library in GenBank. It is correctly attributed to a female library on the I.M.A.G.E. home page (http://www-bio.llnl.gov/bbrp/image/image.html).

described (14), was thought to be composed of six exons. The genomic sequence deposited in GenBank shows splice donor/acceptor sites consistent with a seventh exon. Here we demonstrate that these putative signals are used, and the segment thought to encode exon VI is actually encoded by two exons that we have labeled VI and VII, the existence of this exon proven by comparison of cloned cDNA with cloned genomic DNA. The reassignment of exon VI into exons VI and VII does not greatly alter our understanding of *Xist*.

Another conclusion of the current report is the observation that the newly defined exon VII contains at least an additional 3.1-kb colinear sequence. In previous studies, EST and genomic sequence comparisons between the mouse and human *Xist* locus have suggested alternative structures for the murine *Xist* gene. One such study suggested that the mouse *Xist* gene contained a small distinct 3′ exon homologous to the human *Xist* eighth exon (18). In this same study, no evidence was found for a seventh exon. The data collected in the current study show no evidence for a distinct "eighth" exon in the major *Xist* transcript. Sequence analysis reveals at most seven polyadenylation sites 3′ of exon VI. Thus the structure of the *Xist* gene described here is consistent with a number of differential patterns of polyadenylation. Alternative use of polyadenylation signals could result in size changes for exon VII.

Consistent with our database and sequence evaluations, Northern analysis demonstrates that the major murine *Xist* transcript is longer than previously considered (14, 19). One of

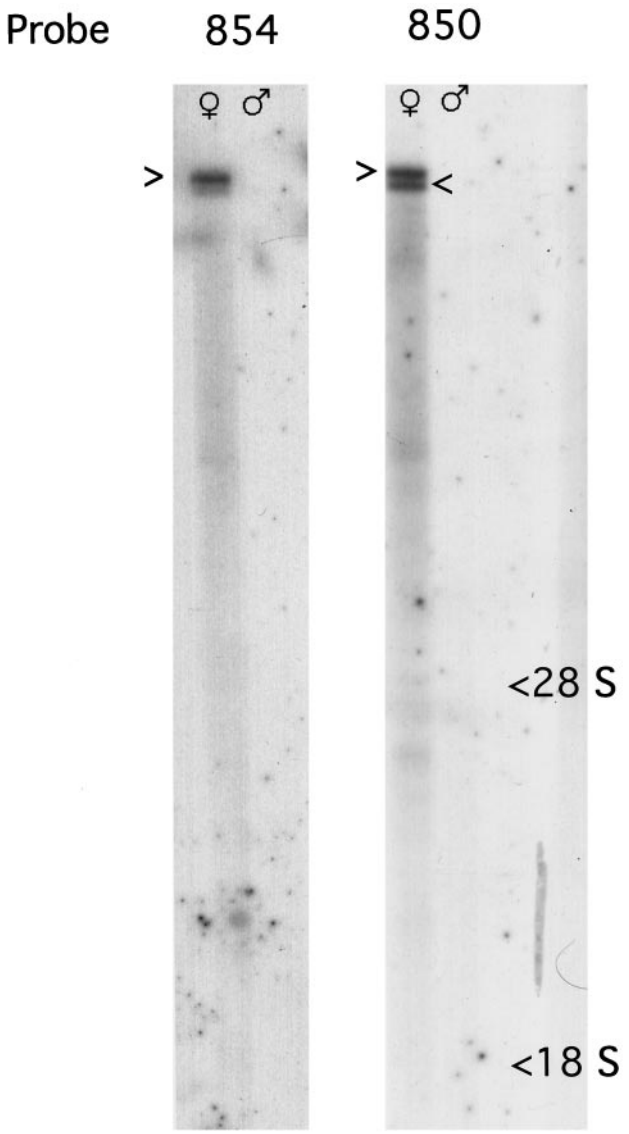

FIG. 3. Northern blot showing major and minor *Xist* species. Murine male (♀) and female (♂) kidney RNA was fractionated on a formaldehyde-agarose gel and transferred to positively charged nylon yielding duplicate strips. Duplicate lanes were hybridized to the individual probes: pWS850, pWS854, and mx8 (20). The lane containing mx8 is not shown, because there was no hybridization signal. *Xist* transcripts, both major and minor, are indicated by arrow symbols. Positions of ribosomal RNA are indicated to give an indication of relative mobility. The figure shows two major species of *Xist* using the pWS850 probe. The new 3′-end probe, pWS854, hybridizes disproportionately to the larger of the two major species of *Xist* RNA.

the two major *Xist* species hybridizes strongly to sequences previously identified with *Xist* (pWS850) as well as to the new sequences reported in this paper (pWS854). As expected, the 5′-probe, mx8 (20), failed to hybridize to the major murine somatic *Xist* transcripts (data not shown).

RNA–FISH experiments have produced results that show that transcripts containing the new 3′ end colocalize with transcripts containing the more 5′-exons. This colocalization supports the conclusion that the new 3′ end of the murine *Xist* gene is part of the functional transcript that has been demonstrated to be necessary for X-chromosome inactivation (8–11).

The current revision of *Xist* gene structure to include the addition of an enlarged exon VII alters the interpretation of the results from *cre/lox* deletional studies (12). The *Xist* proximal *loxP* (see Fig. 2*A*) site thought to lie distal to the 3′

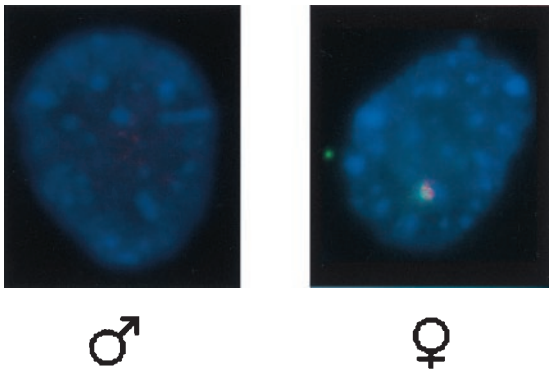

FIG. 4. RNA–FISH photomicrograph of male and female somatic cells. RNA–FISH was performed to visualize cytoplasmic/nuclear RNA that hybridized to *Xist* probes pWS850 (FITC) and pWS854 (rhodamine). Hybridizations of *Xist* probes were performed simultaneously and separate channels recorded; they were merged after recording. Micrograph shows the colocalization of the probes pWS850 and pWS854. (×600.)

end of *Xist* (14) in fact interrupts the *Xist* transcript in the EST1 region before the fourth polyadenylation signal (see Fig. 2*A*). Furthermore, it was established in these *cre/lox* transgenic studies that deletions altered *Xist* expression level in embryonic stem cells and somatic cells, as well as choice of *Xic* to undergo cis inactivation. In embryonic stem cells, expression of a 3′-deleted *Xist* allele is virtually undetectable despite the fact that the promoter region is untouched (20). This same allele is highly expressed and is always chosen in differentiated embryonic stem cells (12). Thus, exclusive choice of the deleted *Xist* allele may be caused by the loss of sequence at its 3′ end. This loss of sequence could alter mRNA stability. A change in *Xist* stability has been hypothesized to act as a trigger in the process of cis-inactivation (21, 22). The production of a transcript stabilized at the 3′ end may therefore be a rate-limiting step in both chromosome choice and initiation of cis-inactivation.

It is reasonable to consider that the observed transgenic phenomena are caused, at least in part, by the alteration in *Xist* genomic structure. However, other models are possible. In the reported deletion (12) there may be additional genes, regions, or elements critical for *Xist* stability and X-chromosome choice/counting. Only a revised functional analysis of the region 3′ to *Xist* will resolve the issue.

The authors acknowledge the extraordinary support of Dr. Fred Rosen and the Center for Blood Research, Harvard Medical School. We also thank Dr. Woo-Joo Song for his generous assistance. The authors also thank Drs. Philip Leder and George Church for critical review of this manuscript. This work was supported by a U.S. Army Prostate Cancer Research Award (PC970479), the Harvard Nathan Shock Center for Biology of Aging Core C, Harvard Medical School (5 P30 AG13314), the Harvard Milton Fund, and a Beth Israel Deaconess New Investigator Award, all awarded to W.M.S.

1. Goto, T. & Monk, M. (1998) *Microbiol. Mol. Biol. Rev.* **62,** 362–378.
2. Heard, E. & Avner, P. (1994) *Hum. Mol. Genet.* **3,** 1481–1485.
3. Heard, E., Clerc, P. & Avner, P. (1997) *Annu. Rev. Genet.* **31,** 571–610.
4. Lyon, M. F. (1994) *Eur. J. Hum. Genet.* **2,** 255–261.
5. Monk, M. (1995) *Dev. Genet.* **17,** 188–197.
6. Cattanach, B. M. & Williams, C. E. (1972) *Genet. Res.* **19,** 229–240.
7. Simmler, M.-C., Cattanach, B. M., Rasberry, C., Rougeulle, C. & Avner, P. (1993) *Mamm. Genome* **4,** 523–530.
8. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. (1996) *Nature (London)* **379,** 131–137.
9. Lee, J. T., Strauss, W. M., Dausman, J. A. & Jaenisch, R. (1996) *Cell* **86,** 83–94.

10. Herzing, L. B. K., Romer, J. T., Horn, J. M. & Ashworth, A. (1997) *Nature (London)* **386,** 272–275.
11. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. (1997) *Genes Dev.* **11,** 156–166.
12. Clerc, P. & Avner, P. (1998) *Nat. Genet.* **19,** 249–253.
13. Cattanach, B. M. (1972) *Mouse News Letter* **47,** 33.
14. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. & Rastan, S. (1992) *Cell* **71,** 515–526.
15. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
16. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
17. Trask, B. J. (1991) *Trends Genet*. **7,** 149–154.
18. Simmler, M. C., Cunningham, D. B., Clerc, P., Vermat, T., Caudron, B., Cruaud, C., Pawlak, A., Szpirer, C., Weissenbach, J., *et al*. (1996) *Hum. Mol. Genet.* **5,** 1713–1726.
19. Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D. & Rastan, S. (1991) *Nature (London)* **351,** 329–331.
20. Johnston, C., Nesterova, T. B., Formstone, E. J., Newall, A. E. T., Duthie, S. M., Sheardown, S. A. & Brockdorff, N. (1998) *Cell* **94,** 809–817.
21. Panning, B., Dausman, J. & Jaenisch, R. (1997) *Cell* **90,** 907–916.
22. Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E. T., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G.-C., Rastan, S. & Brockdorff, N. (1997a) *Cell* **91,** 99–107.