AD_____

Award Number: DAMD17-99-1-9032

TITLE: An Experimental System to Evaluate LOH in Prostate Cancer

PRINCIPAL INVESTIGATOR: William M. Strauss, Ph.D.

CONTRACTING ORGANIZATION: Beth Israel Deaconess Medical Center
Boston, Massachusetts 02215

REPORT DATE: January 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE | OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 2000	Annual (1 Jan 99 - 31 Dec 99)

4. TITLE AND SUBTITLE
An Experimental System to Evaluate LOH in Prostate Cancer

5. FUNDING NUMBERS
DAMD17-99-1-9032

6. AUTHOR(S)
William M. Strauss, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Beth Israel Deaconess Medical Center
Boston, Massachusetts 02215

E-MAIL: wstrauss@hihg.med.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The goal of this grant is to establish a new biological system for studying the progression of prostate cancer. We propose a technology we have previously developed to help define X-chromosome inactivation to increase our understanding of the molecular biology of prostrate cancer. Using a mouse model, our goal is to induce functional Loss of Heterozygosity (LOH) on a particular chromosome at various specified times during development or life span. Specifically, we plan to induce LOH only in mouse prostatic tissues. We are particularly interested in evaluating sites of allelic loss previously identified to be associated with prostate cancer (7q, 8p, 10p, 10q, 13q, 16q, 18q) in humans and understanding the effect of LOH on syntenic mouse chromosomes. We believe this approach has the potential to increase our knowledge of the acquisition and progression of prostate cancer. A systematic experimental approach for creating LOH can help define new prostate specific tumor suppressor genes. As human chromosome 8p is most frequently associated with prostate cancer (80% of primary and metastatic prostate cancers show this chromosomal variant, our mouse model system will initially focus on the development of those mouse chromosomes syntenic with human chromosome 8p. This technology would augment the positional cytogenetic approach to understanding the genetic complexity of prostate cancer.

14. SUBJECT TERMS
Prostate Cancer

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

X Where copyrighted material is quoted, permission has been obtained to use such material.

X Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

X Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

X In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

5-10-00

PI - Signature Date

Table of Contents

5)Introduction

The goal of this grant is to establish a new biological system for studying the progression of prostate cancer. We propose a technology we have previously developed to help define X-chromosome inactivation to increase our understanding of the molecular biology of prostrate cancer. Using a mouse model, our goal is to induce functional Loss of Heterozygosity (LOH) on a particular chromosome at various specified times during development or life span. Specifically, we plan to induce LOH only in mouse prostatic tissues. We are particularly interested in evaluating sites of allelic loss previously identified to be associated with prostate cancer (7q, 8p, 10p, 10q, 13q, 16q, 18q) in humans and understanding the effect of LOH on syntenic mouse chromosomes. We believe this approach has the potential to increase our knowledge of the acquisition and progression of prostate cancer. A systematic experimental approach for creating LOH can help define new prostate specific tumor suppressor genes. As human chromosome 8p is most frequently associated with prostate cancer (80% of primary and metastatic prostate cancers show this chromosomal variant, our mouse model system will initially focus on the development of those mouse chromosomes syntenic with human chromosome 8p. This technology would augment the positional cytogenetic approach to understanding the genetic complexity of prostate cancer.

6) Body

Background

Activation or inactivation of a gene may lead to carcinogenesis. Activation of genes refers to a dominant condition that results in stimulation of both growth and progression of cancer. Inactivation of genes refers to the phenomena where tumor supressor genes (genes that normally inhibit carcinogenesis) are inactivated resulting in a loss function. Classically, these mutations are due to lesions which alter the linear sequence of a particular gene. In addition, somatic dysregulation or inappropriate gene-silencing due to methylation (where the gene is present but nonfunctional), may have similar effects.

Inactivation of a gene may occur as a result of allelic deletion where one or both copies of a locus is lost (1-4). Most frequently one copy is lost and is detected as a loss of heterozygosity (LOH). When this loss involves a tumor suppression gene, carcinogenesis may occur. In prostate cancer, several common sites of allelic loss have been identified including 7q, 8p, 10p, 10q, 13q, 16q, and 18q (1-4). The 8p arm is most frequently lost as 80% of primary and metastatic prostate cancers show this chromosomal variant (10, 11). Most allelic deletions on 8p involve large chromosomal intervals. Working from large collections of clinical specimens, researchers are attempting to define a common overlapping chromosomal region. From this overlapping region, the goal is to then identify particular genes and evaluate their relationship to prostrate cancer.

This approach to allelic loss mapping is difficult if LOH is common and the lesions are complex or non-specific. The approach now used to solve this dilemma is to pursue a random search from available clinical specimens to try to strengthen the correlation of tumor phenotype and chromosome architecture. Another approach to accelerate gene identification in prostrate cancer would be to experimentally test the effect of LOH for particular chromosomal regions on the development of prostate cancer. To date, this approach has not been utilized although it has the potential to greatly speed up the process of prostrate cancer gene identification.

This accelerated approach would involve inactivating areas of the mouse chromosome known to contain sites of allelic loss previously identified to be associated with prostrate cancer. Despite the fact that most human genes have direct homologues in the mouse, the structure of the mouse chromosomes are quite different from the human. Apart from the fact that human chromosomes are acrocentric and the mouse chromosomes are telocentric, the mouse and human genomes show many dissimilarities in the linear arrangements of genes. After years of chromosome mapping by a variety of techniques, a comparative physical and genetic map of the human and mouse chromosomes has emerged (12, 13). It is now possible to draw a direct comparison between subregions of a human and mouse chromosome. For example, human chromosome 8p is distributed in discrete blocks among two mouse chromosomes (mouse ch8 and mouse ch14). These blocks, or syntenic regions, are stable heritable units of genes. Within each chromosomal block, the arrangement of genes is very similar if not identical between mouse and human. Inactivation of these "human chromosome" blocks on the mouse chromosome would allow for physical interval testing to determine their role in prostate cancer. Furthermore by experimentally inducing LOH in the mouse, these syntenic blocks could provide a experimental directed approach to test how a particular region of a human chromosome might operate in the pathogenesis of prostate cancer.

Lessons from X-chromosome inactivation: developmental LOH.

A new technology is required to direct the functional inactivation of a chromosome and create experimental LOH in transgenic mice. We propose to create this technology based upon our basic pioneering work to define the process of X-chromosome inactivation. X-chromosome inactivation is the only known example in mammals of a developmentally regulated functional loss of heterozygosity. It is an example of an epigenetic developmental program that begins anew in the development of every female and represents a unique aspect of an individuals characteristics. X-inactivation is a particular type of epigenetic program operating in female mammals for the purpose of gene dosage compensation between the heterogametic sexes. (14, 15). X-inactivation allows a female embryo to functionally appear as monosomic for the X-chromosome despite the presence of two X-chromosomes. If this process did not occur it would be catastrophic to the developing female embryo with twice the number of X-linked genes as the male.

Two facts are established regarding the mechanism of X-inactivation: (1) a gene which encodes a nontranslateable RNA called *Xist* is necessary for X inactivation (16, 17). Early in development, prior to X-inactivation both male and female cells exhibit a low level of *Xist* expression (18, 19). Subsequent to implantation in female cells one X-chromosome is chosen and exhibits a significant induction of *Xist* expression (18, 19). Soon after *Xist* induction, genes in cis to the actively transcribed *Xist* are repressed for the lifetime of the cell (19-21). If the structural portion of the *Xist* gene is interrupted by homologous recombination, the X-chromosome containing this interrupted allele is incapable of undergoing X-inactivation (16, 17). This demonstrates the necessity of *Xist* for X-inactivation.

A region on the X-chromosome not much larger than the *Xist* gene is sufficient to direct the choice of which X-chromosome undergoes inactivation (19, 21). This DNA interval was first cloned in the form of a yeast artificial chromosome (YAC) by our laboratory and introduced into male embryonic stem cell lines derived (ES cells) (19). This 450 kb YAC was sufficient to be counted as an X-chromosome and direct inactivation on any chromosome in which it was integrated (19). Similarly a 40 kb cosmid was shown to sufficient to cause autosomal inactivation when the cosmid was autosomally integrated in cis in male ES cells (21).

Based upon the observation that a YAC or cosmid spanning *Xist* is necessary and sufficient for X-chromosome inactivation, a vector harboring the *Xist* gene under conditional control could operate to inactivate any chromosome in which it was integrated. To this end we have created a full length cDNA of the murine *Xist* and used it to create a vector in which the cDNA is under control of the tetracycline inducible operator system (22). This vector once integrated into the chromosome of choice would inactivate this chromosome once the *Xist* gene was to be activated. The activation of the *Xist* gene would be controlled by the exogenous administration of tetracycline at any point in development.

The common mouse has been used as a model system for experimental prostate cancer research. Despite the fact that prostate cancer is rarely observed among rodents several approaches to experimental modeling have been developed in the mouse. Three approaches have been described including 1) androgenic hormone stimulation with carcinogen exposure, 2) retroviral transduction and organ reconstitution and 3) transgenic targeting (23). In this proposal we will only address the transgenic approach to prostate cancer modeling. There have been essentially two transgenic models described which show prostate changes characteristic of human disease. These two models involve two different dominant oncogene/protein and two different promoters. The first system involves the use of the MMTV promoter and the Int-2 oncogene. The MMTV promoter is a glucocorticoid responsive viral promoter with favored expression in the mammary tissue of the lactating female. However, it has been shown that male transgenic mouse lines expressing the Int-2 gene under MMTV control results in dramatic epithelial hyperplasia of the prostate (24). The TRAMP model (transgenic adenocarcinoma mouse prostate) has also been described (25, 26). The TRAMP system involves the rat probasin promoter directing expression of the SV40 large T gene in a prostate specific manner. Several reports indicate that this model system recapitulates the aggressive course of human prostatic cancer (25-28). Prostatic intraepithilial neoplasia is observed in male mice of 8 - 12 weeks age. These lesions appear to progress to adenocarcinoma by 30 weeks and finally to distant metastases (25-28).

Preliminary Data

At the time of our Phase I grant submission we had finished the construction of what we believed to be a full length inducible cDNA version of the gene *Xist*. Our construction was guided by the published structures for the *Xist* genomic locus and RNA (34,35). During the final quality control steps, prior to introduction of our construct into ES cells and mice we started an exhastive confirmation process to demonstrate not only that our Xist clone was identical to the published *Xist* structure, but that our clone was identical to the sequences found in the mouse germline. Much to our surprise (and dismay) despite the absolute identity of our clone the published structure, our clone contained discrepancies relative to the mouse genome. We struggled to discover the basis of these differences, and revealed that the published structure for *Xist* was in error and in need of revision. We discovered new structural data for the murine *Xist* gene. These data were published (36), and this paper demonstrates that the murine *Xist* transcript is at least 17.8 kb not 14.7 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.

Analyzing the human *XIST* structure has resulted in a strong structural correlation between the two organisms (37). Comparison of sequences from the genomic interval downstream to the 3′ end of the human *XIST* gene against the human EST database brought to light a number of human EST sequences which are mapped to the region. Furthermore, PCR-amplification of human cDNA libraries and RNA-Fluorescence *In Situ* Hybridization (RNA-FISH) demonstrate that the human *XIST* gene has additional 2.8 kb downstream sequences which have not been documented as a part of the gene. These data show that the full length *XIST* cDNA is in fact 19.3 kb, not 16.5 kb as previously reported. The newly defined region contains an intron that may be alternatively spliced and seven polyadenylation signal sequences. Sequences in the newly defined region show overall sequence similarity with the 3′ terminal region of mouse *Xist* and three subregions exhibit considerably high sequence conservation. Interestingly, the new intron spans the first two subregions that are absent in one of the two isoforms of mouse *Xist*. Taken together, we revise the structure of human *XIST* cDNA and compare cDNA structures between human and mouse *XIST/Xist*.

Finally, another paper has just been submitted documenting the structural explanation for the two RNA isoforms of murine *Xist* and the most reasonable mechanism for their production (38). To further define the molecular structures of the two *Xist* RNA isoforms, we performed northern blot analyses and RNAse protection assay (RPA). Consistent with previous data, our northern blot analyses show that majority of the two transcripts are directed by P2 promoter. Additionally, the northern probe spanning 853 base pairs sequence 3' of *Xist* gave only one band indicating the two isoforms are different at their 3' termini. Probes for the RPA spanned either originally defined 3′ terminus or two of the putative polyadenylation signals at the 3′ termini. Results of the RPA experiments clearly show that *Xist* does not end at the previously proposed site, and the two isoforms are different in their sizes which we called short (*S*) and long (*L*) forms. The *S* form ends at 17030 nucleotides from the +1 transcription start site while the *L* form ends at 17873 nucleotides of the *Xist* cDNA. Therefore the *S* form is 843 nucleotides shorter than the *L* form. The following lines of evidences suggest that the difference in length at the 3′ termini of the two *Xist* isoforms is due to differential polyadenylation, not splicing: 1) Only one band was detectable with the northern probes (pWS855, 859 and 860) spanning 3' of *Xist*. 2) RPA with P2 probe showed 3' termini of both *S* and *L* forms, and there are putative polyadenylation signals and hairpin structures close to these ends. 3) Analyses of splice site prediction program did not show any evidence of splicing in the sequence of *L* form. The extra sequence of the *L* form shares significant sequence similarity with our revision for the structure of the 3′ region of human *XIST*. This suggests that mouse *Xist* depends on differential polyadenylation to generate the two isoforms while human *XIST* may depend on alternative splicing in addition to differential polyadenylation. The newly revised structure of *Xist* isoforms may play essential roles in the stability of *Xist* and the process of X inactivation.

Clearly, the schedule for our prostate project was derailed by the important findings of inaccuracies in the *Xist* structure. Instead of being able to start our transgenic experiments immediately we have spent the last year defining the actual structure of the *Xist* gene and RNA. In addition to defining the true structure for mouse and Human *Xist/XIST* it was necessary to rebuild our cDNA constructs. We now report that full length cDNAs for mouse *Xist* (17.8 kb) have been made.

In addition, using the revised *Xist* cDNA three types of expression constructs have also been made. First a vector that expresses *Xist* in a constituitive manner. Second,

two types of inducible *Xist* constructs have been made, 1) a tetracycline regulated form, and 2) an interferon inducible form.

All of three of these *Xist* expressing constructs have been introduced into somatic and ES cells by random transfection for the purpose of expression testing. The somatic cells used for these experiments are NIH 3T3 cells. These immortalized cells have been successfully transfected with the constructs. In each case the *Xist* constructs expressed RNA which we could detect both by Northern and RNA-FISH. The RNA-FISH results were quite exciting as the ectopically derived *Xist* was observed to "coat" or localize on the transgenic chromosome.

A number of different ES cell lines which inducibly express ectopic *Xist* have been produced. These cell lines were characterized to determine the chromosome into which the transgene had integrated. Our current results show random integrations into mouse 4, 5, 8, a number of additional ES cell lines have yet to be characterized. The integration into distal chromosome 8 is especially exciting as this chromosome is directly relevant to our proposed prostate cancer model of LOH.

Experiments to functionally characterize the transfected constructs have been undertaken. Each of the ES cell lines with *Xist* integrations, into either chromosome 4, 5, 8, have characterized by *Xist* localization, and cis-inactivation of gene expression. For chromosome 4 gene specific assay for *c-jun*, *Tlr4*, and *CDC42* were evaluated by RNA FISH. For chromosome 5 gene specific assay for beta-actin, ketokinase, and *CENP-A* were evaluated by RNA FISH. For chromosome 8 gene specific assay for *EIF-4E* and *Aprt* were evaluated by RNA FISH. In the transfected ES cell cultures, when *Xist* is expressed in an inducible manner it localizes to the transgenic chromosome and result in silencing of the genes in cis to the construct.

7) Key Research Accomplishements

- Redefinition of murine and human Xist/XIST gene structure
- Redefinition of murine and human Xist/XIST RNA structure
- Construction of 2 inducible versions of the murine Xist gene.
- Transfection of these constructs into mouse somatic and ES cells.
- Conditional expression of the inducible version of murine Xist in ES cells.
- Demonstrations that Xist cDNA alone will accomplish cis-silencing.
- Targeting of muine Chromsome 8 with conditional Xist construct.

8)Reportable Outcomes

Publications

1. Hong, Y-K, S.D.Ontiveros, C.Chen, W.M. Strauss. A New Structure for the Murine *Xist* Gene and its Relationship to Chromosome Choice/Counting during X-chromosome Inactivation. *Proc. Natl. Acad, Sci. U.S.A.* **96**(12): 6829-6834(1999).

2. Hong, Y-K, S.D. Ontiveros, W.M. Strauss. A revision of the human *XIST* gene organization and structural comparison to mouse *Xist*. . *Mammalian Genome*. **11**:220-224(2000).

3. Memili, E. Y-K Hong, D. K. Kim, S.D. Ontiveros, W.M. Strauss. Murine *Xist* RNA isoforms are different at their 3'ends: a role for differential polyadenylation.. Submitted.

9) Conclusions

The scientific conclusions of this report are very optimistic. We have redefined the structure for mouse and human Xist/XIST gene and transcript. This transcript causes cis-inactivation of the chromosome from which it is expressed. Thus as we continue to construct the mouse strains harboring the Xist cDNA and the tetracycline transactivator under probasin promoter control we have confidence that the expression of Xist will cause the desired result.

10)References

1. Dong, J.-T., W. B. Isaacs, J. T. Isaacs. 1997. Molecular advances in prostate cancer. *Current Opinion in Oncology*. 9: 101-107

2. Heidenberg, H. B., J. J. Bauer, D. G. McLeod, J. W. Moul, S. Srivastava. 1996. The role of p53 tumor suppressor gene in prostate cancer: a possible biomarker. *Urology*. 48: 971-979

3. Bookstein, R., G. S. Bova, D. MacGrogan, A. Levy, W. B. Isaacs. 1997. Tumor-suppressor genes in prostatic oncogenesis: a positional approach. *British Journal of Urology*. 79 s1: 28-36

4. Koivisto, P. A. 1996. Molecular genetics of prostate cancer. *Critical Reviews in Oncogenesis*. 7: 143-150

5. Bookstein, R., D. MacGrogan, S. G. Hilsenbeck, F. Sharkey, D. C. Allred. 1993. p53 is mutated in a subset of advanced-stage prostate cancers. *Cancer Research*. 53: 3369-3373

6. Navone, N. M., P. Troncoso, L. Pisters, etal. 1993. p53 protein accumulation and gene mutation in the progression of human prostate carcinoma. *J. National Cancer Institute*. 85: 1657-1669

7. Aprikian, A. G., A. S. Sarkis, W. R. Fair, Z. F. Zhang, Z. Fuks, C. Cordon-Cardo. 1994. Immunohistochemical determination of p53 protein nuclear accumulation in prostatic adenocarcinoma. *Journal of Urology*. 151: 1276-1280

8. Carter, H. B., S. Piantadosi, J. T. Isaacs. 1990. Clinical evidence for and implications of the multistep development of prostate cancer. *J. Urol.* 143: 742-746

9. Carter, B. S., C. M. Ewing, W. S. Ward, e. al. 1990. Allelic loss of chromosomes 16q and 10q in human prostate cancer. *Proc. Natl. Acad. Sci. USA*. 87: 8751-8755

10. Cher, M. L., G. S. Bova, D. H. Moore, e. al. 1996. Genetic alterations in untreated prostate cancer metastases and androgene independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Research*. 56: 3091-3102

11. Emmert-Buck, M. R., D. D. Vocke, R. O. Pozzatti, D. P.H., S. B. Jennings, C. D. Florence, Z. Zhuang, D. G. Bostwick, L. A. Liotta, W. M. Linehan. 1995. Allelic loss on chromosome 8p12-21 in microdissected prostatic intraepithelial neoplasia. *Cancer Research*. 55: 2959-2962

12. Dietrich, W., H. Katz, S. E. Lincoln, H.-S. Shin, J. Friedman, N. C. Dracopoli, E. S. Lander. 1992. A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses. *Genetics*. 131: 423-447

13. Copeland, N. G., N. A. Jenkins. 1991. Development and applications of a molecular genetic linkage map of the mouse genome. *Trends in Genetics*. 7: 113-118

14. Lyon, M. F. 1961. Gene Action in the X-chromosome of the mouse (Mus musculus L.). *Nature*. 190: 372-373

15. Lyon, M. F. 1963. Attempts to test the inactive-X theory of dosage compensation in mammals. *Genet. Res. (Camb)*. 4: 93-103

16. Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff. 1996. Requirement for *Xist* in X chromosome inactivation. *Nature*. 379: 131-137

17. Marahrens, Y., B. Panning, J. Dausman, W. Strauss, R. Jaenisch. 1997. *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes and Development*. 11: 156-166

18. Beard, C., E. Li, R. Jaenisch. 1995. Loss of methylation activates *Xist* in somatic but not in embryonic cells. *Genes and Development*. 9: 2325-2334

19. Lee, J. T., W. M. Strauss, J. A. Dausman, R. Jaenisch. 1996. A 450 Kb Transgene displays properties of the mammalian X-inactivation center. *Cell*. 86: 83-94

20. Lee, J. T., R. Jaenisch. 1997. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature*. 386: 275-279

21. Herzing, L. B. K., J. T. Romer, J. M. Horn, A. Ashworth. 1997. *Xist* has properties of the X-chromosome inactivation centre. *Nature*. 386: 272-275

22. Gossen, M., H. Bujard. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. USA*. 89: 5547-5551

23. Buttyan, R., K. Slawin. 1993. Rodent models for targeted onogenesis of the prostate gland. *Cancer and Metastsis Reviews*. 12: 11-19

24. Muller, W. J., F. S. Lee, C. Dickson, G. Peters, P. Pattengale, P. Leder. 1990. The int-2 gene product acts as an epithelial growth factor in transgenic mice.. *EMBO Journal.* 9: 907-913

25. Greenberg, N. M., F. J. DeMayo, P. C. Sheppard, R. Barrios, R. Lebovitz, M. Finegold, R. Angelopoulou, J. G. Dodd, M. L. Duckworth, J. M. Rosen, e. al. 1994. The rat probasin promoter directs hormonally and developmentally regulated expression of a heterologous gene specifically to the prostate in transgenic mice. *Molecular Endocrinology.* 8: 230-239

26. Greenberg, N. M., F. DeMayo, M. J. Finegold, D. Medina, W. D. Tilley, J. O. Aspinall, G. R. Cunha, A. A. Donjacour, R. J. Matusik, J. M. Rosen. 1995. Prostate cancer in a transgenic mouse. *Proc. Natl. Acad. Sci. USA.* 92: 3439-3443

27. Foster, B. A., J. R. Gingrich, E. D. Kwon, C. Madias, N. M. Greenberg. 1997. Characterization of prostatic epithelial cell lines derived from transgenic adenocarcinoma of the mouse prostate (TRAMP) model. *Cancer Research.* 57: 3325-3330

28. Gingrich, J. R., R. J. Barrios, R. A. Morton, B. F. Boyce, F. J. Demayo, M. J. Finegold, A. R., J. M. Rosen, J. M. Greenberg. 1996. Metastatic prostate cancer in a transgenic mouse. *Cancer Research.* 56: 4096-4102

29. Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, D. C. Prasher. 1994. Green flourescent protein as a marker for gene expression. *Science.* 263: 802-805

30. Yan, Y., P. C. Sheppard, S. Kasper, L. Lin, S. Hoare, A. Kapoor, J. G. Dodd, M. L. Duckworth, R. J. Matusik. 1997. Large fragment of the probasin promoter targets high levels of transgene expression to the prostate of transgenic mice. *Prostate.* 32: 129-139

31. Johnson, C. V., R. H. Singer, J. B. Lawrence. 1991. Fluorescent detection of nuclear RNA and DNA: Implications for genome organization. *Methods in Cell Bio.* 35: 75-99

32. Lawrence, J. B., R. H. Singer, L. M. Marselle. 1989. Highly Localized Distribution of Specific Transcripts Within Interphase Nuclei Visualized by in situ Hybridization. *Cell.* 57: 493-502

33. Lawrence, J. B., R. H. Singer, J. A. McNeil. 1990. Interphase and Metaphase Resolution of Different Distances with the Human Dystrophin Gene. *Science.* 249: 928

34.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.

35.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* **351,** 329-331.

36.Hong, Y-K, S.D.Ontiveros, C.Chen, W.M. Strauss. A New Structure for the Murine *Xist* Gene and its Relationship to Chromosome

Choice/Counting during X-chromosome Inactivation. *Proc. Natl. Acad, Sci. U.S.A.* **96**(12): 6829-6834(1999).

37.Hong, Y-K, S.D. Ontiveros, W.M. Strauss. A revision of the human *XIST* gene organization and structural comparison to mouse *Xist*. . *Mammalian Genome*. **11**:220-224(2000).

38.Memili, E. Y-K Hong, D. K. Kim, S.D. Ontiveros, W.M. Strauss. Murine *Xist* RNA isoforms are different at their 3'ends: a role for differential polyadenylation.. Submitted

11)Appendicies

N/A

Mammalian Genome 11, 220–224 (2000).

Mammalian Genome
Incorporating Mouse Genome

A revision of the human XIST gene organization and structural comparison with mouse *Xist*

Young-Kwon Hong, Sara D. Ontiveros, William M. Strauss

Harvard Institute of Human Genetics, 4 Blackfan Circle, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA

Received: 3 August 1999 / Accepted: 15 November 1999

Abstract. The XIST gene plays an essential role in X Chromosome (Chr) inactivation during the early development of female humans. It is believed that the XIST gene, not encoding a protein, functions as an RNA. The XIST cDNA is unusually long, as its full length is reported to be 16.5 kilobase pairs (kb). Here, comparison of sequences from the genomic interval downstream to the 3′ end of the human XIST gene against the human EST database brought to light a number of human EST sequences that are mapped to the region. Furthermore, PCR amplification of human cDNA libraries and RNA fluorescence in situ hybridization (RNA-FISH) demonstrate that the human XIST gene has additional 2.8 kb downstream sequences which have not been documented as a part of the gene. These data show that the full-length XIST cDNA is, in fact, 19.3 kb, not 16.5 kb as previously reported. The newly defined region contains an intron that may be alternatively spliced and seven polyadenylation signal sequences. Sequences in the newly defined region show overall sequence similarity with the 3′ terminal region of mouse *Xist,* and three subregions exhibit quite high sequence conservation. Interestingly, the new intron spans the first two subregions that are absent in one of the two isoforms of mouse *Xist.* Taken together, we revise the structure of human XIST cDNA and compare cDNA structures between human and mouse XIST/*Xist.*

Introduction

X chromosome inactivation is an early developmental process occurring in female mammals to compensate for differences between male and female mammals in dosage of genes residing on the X Chr (Brockdorff 1998; Brown et al. 1992; Lee and Jaenisch 1997; Lyon 1961). This mammalian dosage compensation is achieved by the transcriptional silencing of genes on one of the two X Chrs in females (Brockdorff 1998; Gartler et al. 1972; Lee and Jaenisch 1997). The inactivated X Chr can be microscopically observed during interphase as a condensed body at the nuclear periphery (Barr and Bertram 1949). Consistent with the chromosomal level of transcriptional silencing, the inactive X Chr is both hypermethylated on CpG islands and hypoacetylated on histone H4, compared with the active X Chr (Jeppesen and Turner 1993; Keohane et al. 1998; Miller et al. 1974).

From the study of chromosomal translocations, an interval called the XIC (X inactivation center) of the X Chr has been identified to control the process of inactivation. Translocations containing this segment to an autosome direct the inactivation of the autosome. A gene expressed exclusively from the inactivate X Chr has been cloned from human and mouse and mapped to the XIC region (Borsani et al. 1991; Brockdorff et al. 1992; Brown et al. 1992). This gene, called XIST/*Xist* (X inactive specific transcript), shows several interesting features. First, both human and mouse XIST/*Xist* cDNA are unusually long, reportedly 16.5 kb and 17.8 kb, respectively (Brown et al. 1992; Hong et al. 1999). Second, the transcript does not seem to encode a protein, on the basis of the lack of a significant open reading frame, absence of the *Xist* RNA from polysomes, and localization of the transcript in the nucleus (Brockdorff et al. 1992; Brown et al. 1992). Third, the XIST/*Xist* RNA physically associates with, or 'coats,' the inactive X Chr (Brown et al. 1992; Clemson et al. 1996). Fourth, XIST/*Xist* transcripts can be observed as early as the four-cell stage, and upon the initiation of X-inactivation, the steady-state level of the transcript rises dramatically, apparently by stabilization of the RNA (Panning et al. 1997; Sheardown et al. 1997). Although the function of XIST/*Xist* is not known, deletion of the gene leads to failure of X-inactivation, and knock-out mice die around the gastrulation stage (Marahrens et al.1997; Penny et al. 1996).

In this report, we revise the structure of the human XIST cDNA and discuss structural features of the newly defined region.

Materials and methods

Reagents. Genomic DNA was purified from a human female placenta, as described (Strauss 1999), and used as genomic DNA control for all the PCR reaction. Human cDNA libraries were obtained as follows: female pancreas (CLONTECH Catalog No. HL 1163b), male colon (CLONTECH Catalog No. HL 1034b), male/female bone marrow (CLONTECH Catalog No. HL 5005b), male liver (CLONTECH Catalog No. HL 3006b), male/female pituitary gland (CLONTECH Catalog No. HL 1139b), male/female fetal brain (CLONTECH Catalog No. HL 5015b), and fetal heart (sex unknown, CLONTECH Catalog No. HL 1114b). DNA sequences of oligonucleotides used for this study are listed in Table 1.

Table 1. Names, locations and sequences of primers used for this study

Names of primers	Locations (bp)[a] A	B	Sequences
WS 925	−160	16321	ACCTTGACCTGGCCTACAGA
WS 926	520	17001	TTGTTCCTGTGTTTCCACCA
WS 927	339	16820	TTGCTCATTGGTCTGGCTTA
WS 928	1075	17556	CCATGCCCCTAACAAGAAAA
WS 929	1049	17530	TGGCTTGTTTTCTTGTTAGGG
WS 930	1759	18240	CCCACCCTCTGTGAGTGATT
WS 931	1659	18140	TTGGCCAAAATTGAAAGGAA
WS 932	2351	18832	CAGCTGAAGAAAGGGGTGTT
WS 933	2052	18533	AAAGCTGAAGCCAAAATATGC
WS 934	2754	19235	CCAACTCCCCAGTTTGTTTC
WS 935	1341	17822	TGAGCCACAATTGGTTTTGA
WS 936	2559	19040	AAGGACAATGACGAAGCCAC
GAPDH-F			GAAGGTGAAGGTCGGAGTC
GAPDH-R			GAAGATGGTGATGGGATTTC

[a] Locations of primers are shown as the distances from the previously published end (**A**) and from the transcriptional initiation site (**B**) of XIST (Brown et al. 1992).

Correspondence to: W.M. Strauss; e-mail: wstrauss@hihg.med.harvard.edu

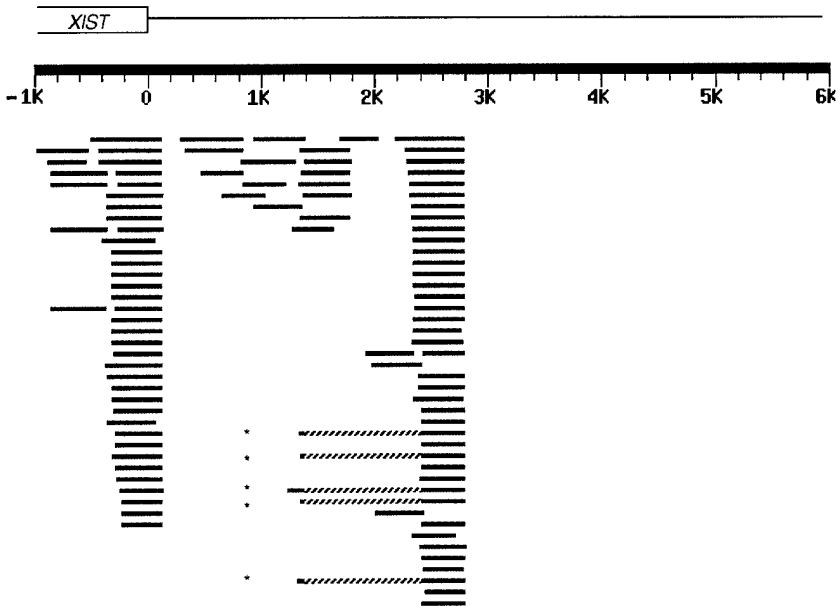


Fig. 1. Results of BLAST sequence similarity search of the 3′ end and downstream region of the human XIST gene with NCBI BLAST 2.0. Genomic sequences of 7000 bp from the human Xq13 region (GenBank Accession No. U80460) spanning the 3′ end of XIST and its downstream were searched against the human EST database with an Expect value, 0.0001, without filtering sequences of low compositional complexity, as of June 15, 1999 (www.ncbi.nlm.nih.gov). The previously reported end of the XIST gene is sketched on top, and relative locations from the end are shown in kilobases (K). Each bar represents an EST sequence registered in the human EST database. Gray bars represent segments in the XIST gene that have a similarity to the EST sequences above the threshold Expect value. A hatched region in the middle of a bar indicates a region of similarity below the Expect value. GenBank accession numbers of the EST sequences marked by asterisks are in order from the top N68599, AI268939, AI693632, N94964, and N93942.

PCR amplification and Southern blotting. Each PCR reaction was performed at 94°C 30 s, 58°C 30 s, and 72°C 30 s for 40 cycles (except 35 cycle for GAPDH). All the PCR products were analyzed in 2% agarose gels with the 100-bp size marker (Promega). The PCR products derived from the female pancreas cDNA library were cloned in pBSIISK+ (Stratagene) and sequenced with both T3 and T7 primers on an Applied Biosystems 377 sequencer in the Beth Israel Deaconess Medical Center sequencing facility. These clones were subsequently radiolabeled by the random hexamer method (Feinberg and Vogelstein 1983) in the presence of [^{32}P] dCTP and used as probes for Southern blot analysis as follows. The PCR products derived from either genomic DNA or cDNA libraries were electrophoresed on an 2% agarose gel and transferred to nylon membrane and hybridized with each probe for 12–24 h at 65°C in a buffer containing 7% SDS, 2 mM EDTA (pH 7.6), and 0.5 M sodium phosphate (pH 7.5).

RNA fluorescence in situ hybridization (FISH). Male and female human fibroblast cell lines were obtained from the NIGMS Repositories (Catalog numbers GM04033 and GM04281, respectively), grown on chamber slides, and fixed as described (Trask 1991). Two sets of DNA probes were used: the previously published XIST G1A (Clemson et al. 1998), ~10 kb genomic plasmid spanning from the fourth intron to the 3′ end of the human XIST, was labeled with biotin, and the plasmids containing the PCR fragments, WS927/928, WS929/930, WS931/932, and WS933/934 (Fig. 2), were pooled and labeled with digoxygenin. Probes were hybridized either separately or simultaneously to the interphase spreads of human fibroblasts. After washing, the slides were labeled with antidigoxygenin rhodamine or avidin-fluorescein. Images were collected with a NIKON E-800 microscope equipped with a Sensys (Photometrics, AZ) digital CCD camera. Grayscale images for either FITC, rhodamine, or DAPI filter sets were pseudocolored, and images were merged in the 12-bit format. The 12-bit data were compressed at eight-bit data during export to Photoshop 5.0 (Adobe) for final figure preparation.

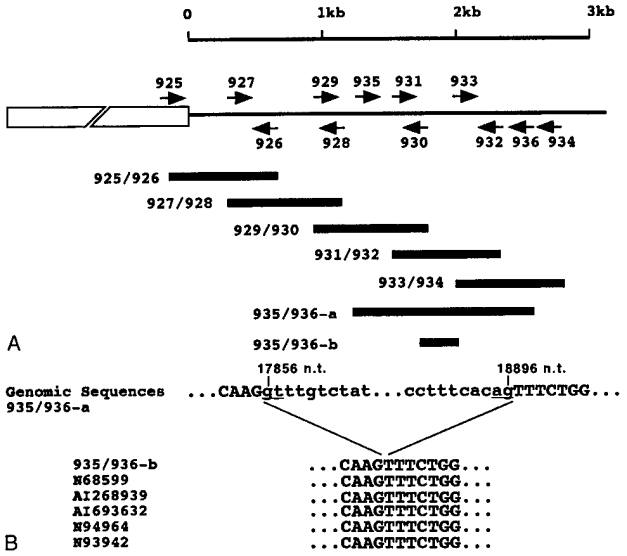


Fig. 2. **(A)** Relative locations of PCR primers (arrows) against the previously reported end of XIST and the overlapping PCR products (black bars) resulting from PCR reactions with each primer set. Names of primers used for the PCR reactions are shown next to the black bars. PCR amplification reaction with WS935 and WS936 yields products of two sizes, 935/936-a and 935/936-b. DNA sequences of the primers are listed in Table 1. **(B)** Sequence alignment of genomic DNA and 935/936-a to the PCR product 935/936-b, as well as the five human ESTs shown in Fig. 1 (GenBank Accession No. N68599, AI268939, AI693632, N94964, and N93942). Intron sequences are lower-cased, and sequences of splice donor and receptor are underlined. Nucleotide sequence positions of the 5′ and the 3′ ends of the intron are shown based on the numbering by Brown et al. (1992).

Results

We reviewed the human XIST gene structure by comparing the sequences encompassing the 3′ end and downstream of the gene against the human Expressed Sequence Tag (EST) databases. Surprisingly, this examination brought to light a number of ESTs (~100 sequences) mapped to, not only the 3′ end of the gene, but also further downstream of the gene (Fig. 1; Brown et al. 1992). This implied that the human XIST gene may encode additional information at the 3′ end. In order to test this, we designed DNA oligonucleotide primers spanning both the 3′ end and downstream of the gene (Table 1, Fig. 2A). Using these primers, we PCR-amplified the corresponding regions from several human cDNA libraries as well as human genomic DNA. The human cDNA libraries used for this purpose were derived either from male, female, or male/female pooled RNAs. We cloned all the PCR fragments amplified from the human female pancreas cDNA library into the pBSIISK+ vector. Sequencing analyses of the cloned fragments confirmed that the PCR-amplified products were indeed

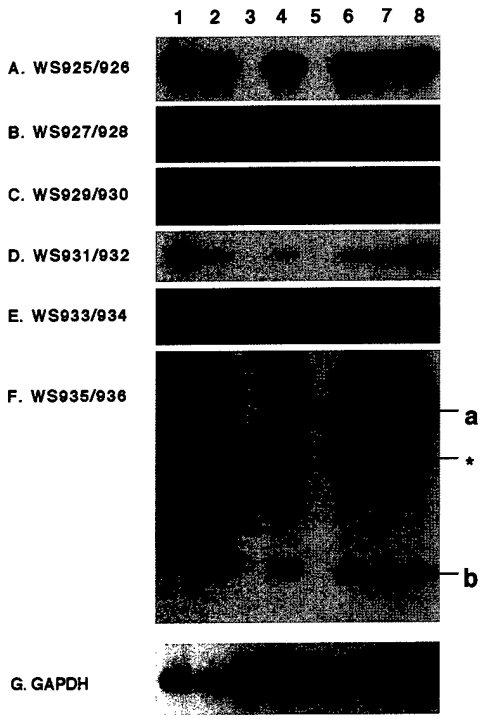


Fig. 3. Southern blot analyses of the PCR products. DNA fragments were PCR-amplified with each primer shown in Fig. 2A from either human genomic DNA or cDNA libraries (lane 1, female genomic DNA from placenta; lane 2, human female pancreas cDNA library; lane 3, human male colon cDNA library; lane 4, human male/female bone marrow cDNA library; lane 5, human male liver cDNA library; lane 6, human male/female pituitary gland cDNA library; lane 7, human male/female fetal brain cDNA library; lane 8, sex-unknown human fetal heart cDNA library). The PCR primer sets used for the reaction are shown next to each panel (A≈F). The PCR products were electrophoresed, transferred to a nylon membrane, and hybridized with probes. The probes were prepared from the cloned and sequenced PCR products, which were amplified with the corresponding primer set from the human female pancreas cDNA library (panels A through E) or from genomic DNA (panel F). In panel F, a and b indicate the PCR products, 1219 bp and 178 bp, amplified from unspliced and spliced cDNA templates, respectively, and *, nonspecific PCR product. The PCR amplification reactions are controlled by amplifying the human GAPDH gene (panel G).

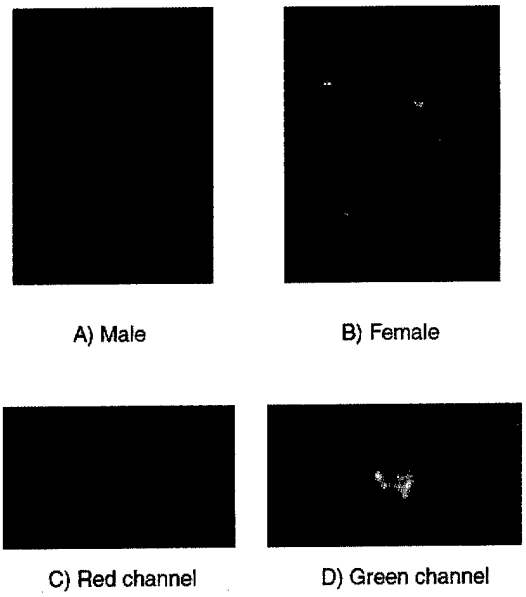


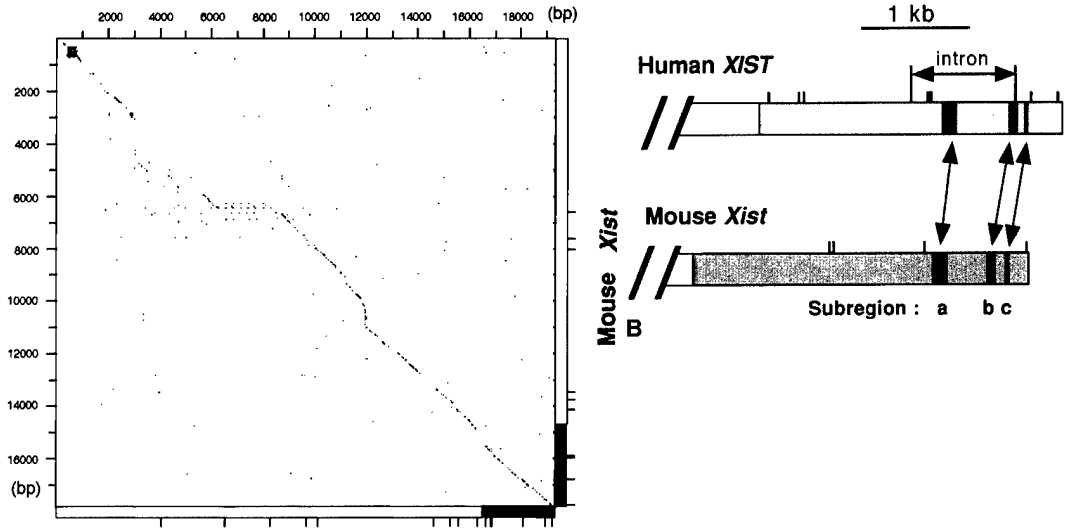
Fig. 4. Two-color RNA-FISH photography of human male **(A)** and female **(B)** interphase nuclei. XIST G1A (green), which spans the genomic sequence from the fourth intron to the previously documented end of the gene, and the newly defined region (red) are colocalized on the inactive X Chr in female (yellow). Enlarged gray-scale images of rhodamine (C) and fluorescein (D) signals of the upper right cell in panel (B) are shown.

mapped to the 3′ end of XIST and its downstream region (data not shown). Subsequently, the PCR products amplified from both human genomic DNA and human cDNA libraries were subjected to Southern blot analysis by using the cloned and sequenced PCR fragments as probes (Fig. 3, A–E). This Southern analysis demonstrates that only female-specific cDNA libraries yield the PCR products. Furthermore, the PCR products amplified from the female-specific cDNA libraries and human genomic DNA are of the same size and hybridize equally to the sequenced probes (Fig. 3A–E). These data imply that the corresponding sequences are expressed only in females, consistent with the female-specific expression of XIST. More importantly, the sequences downstream from the documented human XIST end, which have not been considered as a part of the gene, are transcribed and contiguous with the XIST RNA (Brown et al. 1992). Therefore, these data lead us to conclude that the full length of the human XIST RNA is at least 19.3 kb, not 16.5 kb as previously reported (Brown et al. 1992).

Sequence comparison of the 3′ downstream region of XIST to the EST databases also suggested the presence of an intron in the newly defined 2.8-kb region (Fig. 1). In order to prove the presence of an intron, we used a primer set (WS935 and WS936) that spans the potential intron and then PCR-amplified the region from both human genomic DNA and the human cDNA libraries. PCR products from human genomic DNA or cDNA library from female pancreas RNA were cloned and sequenced. Southern blot analysis shows two PCR fragments amplified from the female RNA-derived cDNA libraries, whereas only one fragment was amplified from the human genomic DNA (Fig. 3F). Sequence alignment between the XIST genomic DNA and the PCR product from the pancreas RNA-derived cDNA library, as well as the five ESTs marked in Fig. 1, demonstrate that the sequence (1041 bp) between nucleotide positions 17856 and 18896 is processed (Fig. 2B). The splice junctions match the consensus splice donor and acceptor sequences for vertebrates (Shapiro and Senapathy 1987). Therefore, we conclude that the newly identified region of the human XIST gene contains an intron.

RNA-fluorescence in situ hybridization (FISH) was employed to determine whether the newly identified region colocalized with the established sequences of XIST on the inactive X Chr in human female fibroblast cells. We used the previously characterized XIST genomic DNA probe, XIST G1A, which spans from the fourth intron to the formerly documented end of the gene and has been shown to paint the inactive X Chr in human females (Clemson et al. 1998). In Fig. 4, the DNA probes derived from the newly defined region colocalize with the XIST G1A probe on the inactive X Chr in a female-specific manner. This evidence clearly demonstrates that the newly defined region is a part of the human XIST transcript and correctly localizes on the inactive X Chr in human female cells.

Recently, we showed that the murine *Xist* gene has an additional 3.1 kb in sequence than was previously documented (Hong et al. 1999). In an attempt to explore the possible structure/functional conservation between the two homologs, we compared sequences of the human and murine XIST/*Xist* cDNA (Fig. 5A). The dot matrix sequence comparison of the two genes with a threshold of 80% sequence identity in a 21-base pair window sliding along the genes shows a substantial conservation between the two species. Importantly, the newly defined 3′ regions are



A **Human *XIST***

Subregion a

```
XIST:  18199 ctgggctttgacctttccgtttttaaatcactcacagagggtgggacaggaggaagagt 18258
             ||| |||||||||| ||||| |||||   ||||||||||||||||||||||||||  ||||
Xist:  16957 ctgtgctttgacccttccatttttctttcactcacagagggtgggacaggaggccgagt 17016

XIST:  18259 gaaggaaaaggtcaaacctgt 18279
             |||||||| |||| | |||||
Xist:  17017 gaaggaaagggtccagcctgt 17037
```

Subregion b

```
XIST:  18856 tggaaaatggaaagaagaaagtaactaaaagccttcctttcacagtttctggcatca 18912
             |||||  || |||| | ||| ||||| ||||||||||||||||||| ||| |||||
Xist:  17494 tggaatttgcaaaggaaaaaataactgaaagccttcctttcacagttactgccatca 17550
```

Subregion c

```
XIST:  18953 catcatctgctttattcacataaatgaa 18980
             |||||||||||||  ||||||||||||
Xist:  17579 catcatctgctttgatcacataaatgaa 17606
```

C

Fig. 5. **(A)** Dot matrix sequences comparison between the human XIST and mouse *Xist* cDNA sequences. With the DNA STRIDER 1.2 program, human and mouse XIST/*Xist* cDNA were compared by scoring sequence identity of at least 17 bp out of a 21-bp window moving along the genes. Therefore, a dot represents location for each gene that shares the similarity. Horizontal and vertical line diagrams at the bottom and the right side of the matrix depict the human and mouse XIST/*Xist* cDNA sequences, respectively. Shaded region of each line diagram shows the newly defined sequences of human XIST and mouse *Xist*, respectively (Hong et al. 1999). Short line ticks in the each diagram indicate locations of the consensus polyadenylation signal sequence. **(B)** Schematic representation of homologous regions between the 3′ ends of human and mouse XIST/*Xist*. Three subregions, a, b and c, are depicted that show a high sequence identity (a, 85% over 81 bp; b, 84% over 57 bp; c, 93% over 28 bp). Locations of polyadenylation signal sequences (ticks) and of the intron are shown. **(C)** Alignments of sequences from the three subregions. Nucleotide numbers are shown for both human XIST and mouse *Xist* cDNA sequences.

among regions that show the most significant similarity between the two genes. The newly defined region of human XIST also contains seven polyadenylation consensus signal sequences and three subregions with relatively high sequence similarity to mouse *Xist* (Fig. 5B, C). The newly identified intron in human XIST spans the first two subregions. A further examination of regions downstream of the new 3′ sequences did not show any significant sequence similarity.

Discussion

In this report, we demonstrate that the human XIST transcript is, in fact, 2.8 kb longer than the 16.5 kb previously documented (Brown et al. 1992). This finding is based on three independent approaches; the EST database analysis with the BLAST search algorithms, female specific-PCR amplification of XIST cDNA, and finally, co-localization of the newly defined region with XIST sequences previously shown to be associated with the inactive X Chr in female cells. Sequence comparisons of the 3′ end and the region downstream of XIST against the human EST databases reveal that a number of EST sequences are reported to map not only to the established 3′ end, but also further downstream of the gene (Fig. 1). Female-specific PCR-amplification of the 3′ end as well as the newly defined region of XIST from human cDNA library demonstrates that the additional 2.8-kb sequence is co-linear to the XIST transcript (Figs. 2, 3). Finally, the DNA probes containing the downstream region co-localize with the established XIST sequence on the inactive X Chr in human female cells (Fig. 4). On the basis of these data, we conclude the human XIST full-length cDNA is at least 19.3 kb long.

It is an interesting coincidence that the cDNA structures of both human and mouse XIST/*Xist* have been incorrectly characterized and reported to be shorter (Brockdorff et al. 1992; Brown et al. 1992). One possible explanation is the presence of adenosine-rich sequences in the genome sequence. We found a stretch of sequence with a high adenosine content downstream of the previously reported ends of both human XIST (25 A residues out of 27 nucleotides, GenBank Accession No. U80460) and mouse *Xist* (15

out of 17, GenBank Accession No. X99946). The adenosine-rich stretch is located 118 bp downstream from the previously reported ends of human XIST and immediately follows the previously defined end of the mouse *Xist* gene. This may have induced misprimings of the oligo dT primers for the first-strand synthesis during the construction of the cDNA libraries, from which the authors collected the XIST/*Xist* cDNA clones (Brockdorff et al. 1992; Brown et al. 1992). Consistent with this idea, the ends of a number of human EST sequences are mapped at the adenosine-rich stretch (Fig. 1). Nevertheless, we do not rule out the possible presence of shorter isoforms of XIST that may be formed by alternative usage of polyadenylation signals at the 3′ end. In fact, we have observed two major species of the *Xist* RNA that differ at their 3′ ends in the mouse (Hong et al. 1999).

It is our belief that the human XIST RNA does not extend further downstream than 2.8 kb, based on the following three lines of observation. First, even after we reduced the stringency of the BLAST search (Expected value = 0.01), we did not find any human EST sequences with a significant similarity to the region up to 10 kb downstream from the newly defined end (Fig. 1, data not shown, apart from a number of EST sequences with partial sequence similarity to the Alu repeats). Second, all the EST sequences that mapped close to the new 3′ end unanimously terminate at one location, 2.8 kb downstream from the previously documented end, which is not followed by any adenosine-rich sequence string (Fig. 1). Third, there is no significant sequence similarity between human and mouse sequences further downstream of the newly defined regions of the two genes (Fig. 5).

Two lines of data suggest that the intron in the newly defined region is likely to be alternatively spliced: PCR reactions using primers (WS935 and WS936) flanking the intron yielded DNA products amplified from unprocessed cDNA template as well as from spliced template, and PCR reactions using primers priming at the intron (WS929/930, WS931/932, WS933/934) also generated products, implying the presence of unprocessed cDNA template (Fig. 2). Interestingly, while the region shares a considerable sequence similarity between human and mouse, no intron has been found in the corresponding region of mouse *Xist,* to date. One implication may be that mouse *Xist* may depend only on differential polyadenylation to generate isoforms. Human XIST may utilize alternative splicing in addition to differential polyadenylation. Within the same context, it is noteworthy that the new intron in XIST spans the first two out of three subregions of high sequence similarity between mouse and human (Fig. 5B, C). Thus, it is formally possible that murine splice variants may be produced, but as of yet are undetected. We discovered that one of two isoforms of *Xist* ends before the first similarity subregion (Hong et al. 1999, data unpublished). Taken together, it is tempting to think that the two similarity subregions may provide important functional distinctions to the XIST/*Xist* isoforms. In order to validate this idea, additional study will be required.

Acknowledgments. 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 and Dr. J.B. Lawrence for the XIST G1A probe. This work was supported by a US Army Prostate Cancer Research Award (PC970479), the Harvard Nathan Shock Center for Biology of Aging Core C (5 P30-AG13314), the Harvard Milton Fund, and a Beth Israel Deaconess New Investigator Award, all awarded to W.M. Strauss.

References

Barr ML, Bertram EG (1949) A morphological distinction between neurons of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature 163, 676–677

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D et al. (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature 351, 325–329

Brockdorff N (1998) The role of Xist in X-inactivation. Curr Opin Genet Dev 8, 328–333

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP et al. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526

Brown CJ, Hendrich BD, Rupert JL, Lafreneire RG, Xing Y et al. (1992) The human *XIST* gene: analysis of a 17kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71, 527–542

Clemson CM, McNeil JA, Willard HF, Lawrence JB (1996) *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol 132, 259–275

Clemson CM, Chow JC, Brown CJ, Lawrence JB (1998) Stabilization and localization of xist RNA are controlled by separate mechanisms and are not sufficient for X inactivation [In Process Citation]. J Cell Biol 142, 13–23

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132, 6–13

Gartler SM, Chen SH, Fialkow PJ, Giblett ER, Singh S (1972) X chromosome inactivation in cells from an individual heterozygous for two X-linked genes. Nat New Biol 236, 149–150

Hong YK, Ontiveros SD, Chen C, Strauss WM (1999) A new structure for the murine xist gene and its relationship to chromosome choice/counting during X-chromosome inactivation [in Process Citation]. Proc Natl Acad Sci USA 96, 6829–6834

Jeppesen P, Turner BM (1993) The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell 74, 281–289

Keohane AM, Lavender JS, O'Neill LP, Turner BM (1998) Histone acetylation and X inactivation. Dev Genet 22, 65–73

Lee JT, Jaenisch R (1997) The (epi)genetic control of mammalian X-chromosome inactivation. Curr Opin Genet Dev 7, 274–280

Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). Nature 190, 372–373

Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis [see comments]. Genes Dev 11, 156–166

Miller OJ, Schnedl W, Allen J, Erlanger BF (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature 251, 636–637

Panning B, Dausman J, Jaenisch R (1997) X chromosome inactivation is mediated by Xist RNA stabilization. Cell 90, 907–916

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. Nature 379, 131–137

Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15, 7155–7174

Sheardown SA, Duthie SM, Johnston CM, Newall AE, Formstone EJ et al. (1997) Stabilization of Xist RNA mediates initiation of X chromosome inactivation [see comments]. Cell 91, 99–107

Strauss WM (1999) Preparation of genomic DNA from mammalian tissue. In *Current Protocols in Neuroscience,* vol Appendix 1H. (New York: John Wiley & Sons, Inc.) pp A.1H.1

Trask B (1991) Fluorescence in situ hybridization: applications in cytogenetics and gene mapping. Trends Genet 7, 149–154

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).

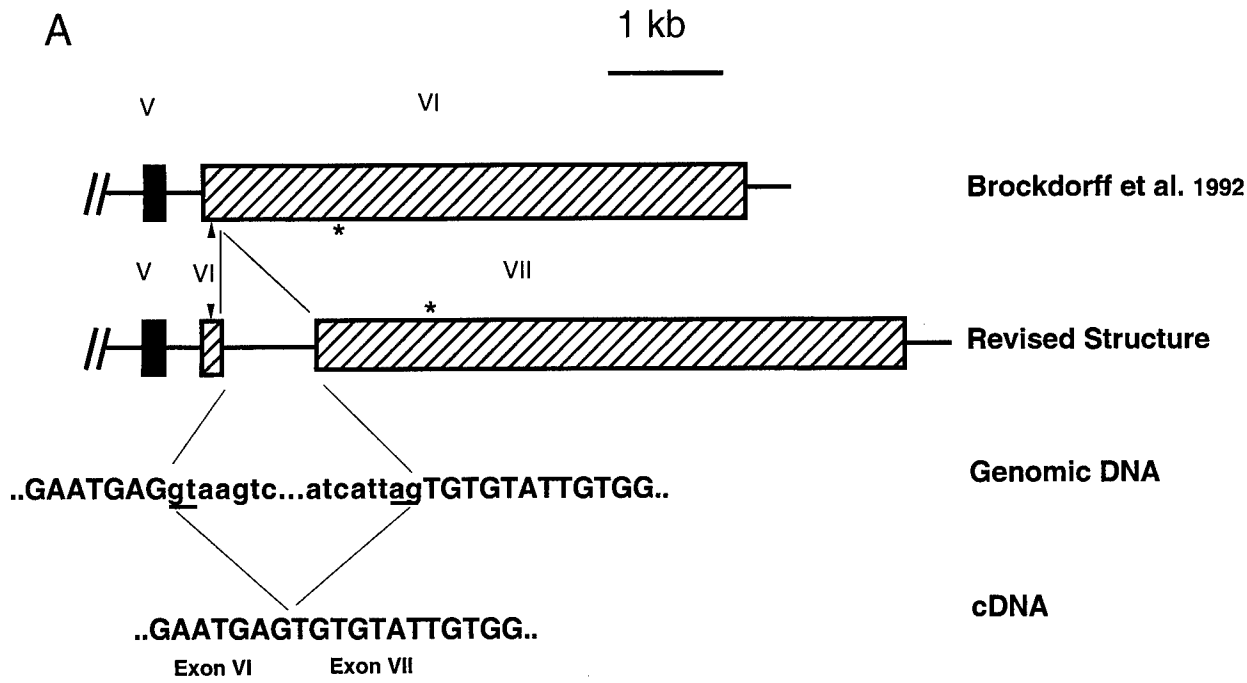
RESULTS

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



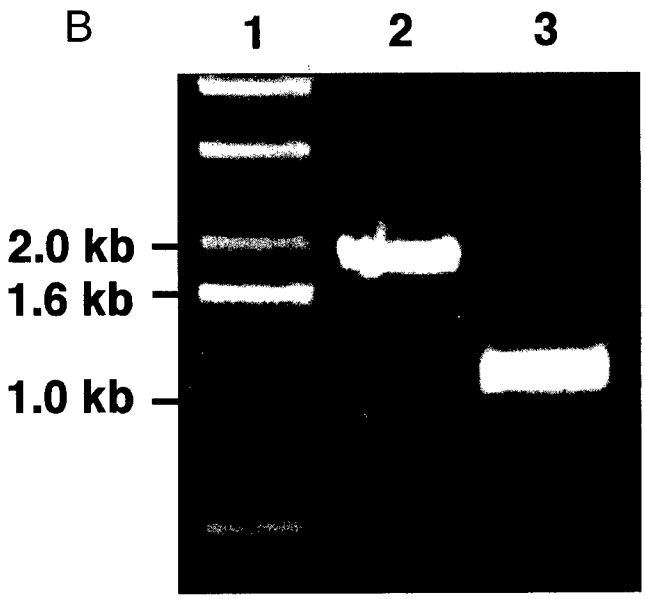


FIG. 1. The murine *Xist* gene consists of seven exons. (*A*) Revised genomic structure of *Xist* contains a 781-bp-long intron compared with the documented structure of exon VI by Brockdorff *et al.* (14). Exon VI/exon VII junction sequences are shown. Arrowhead, *Ava*I; asterisk, *Eco*RI. (*B*) Comparison 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

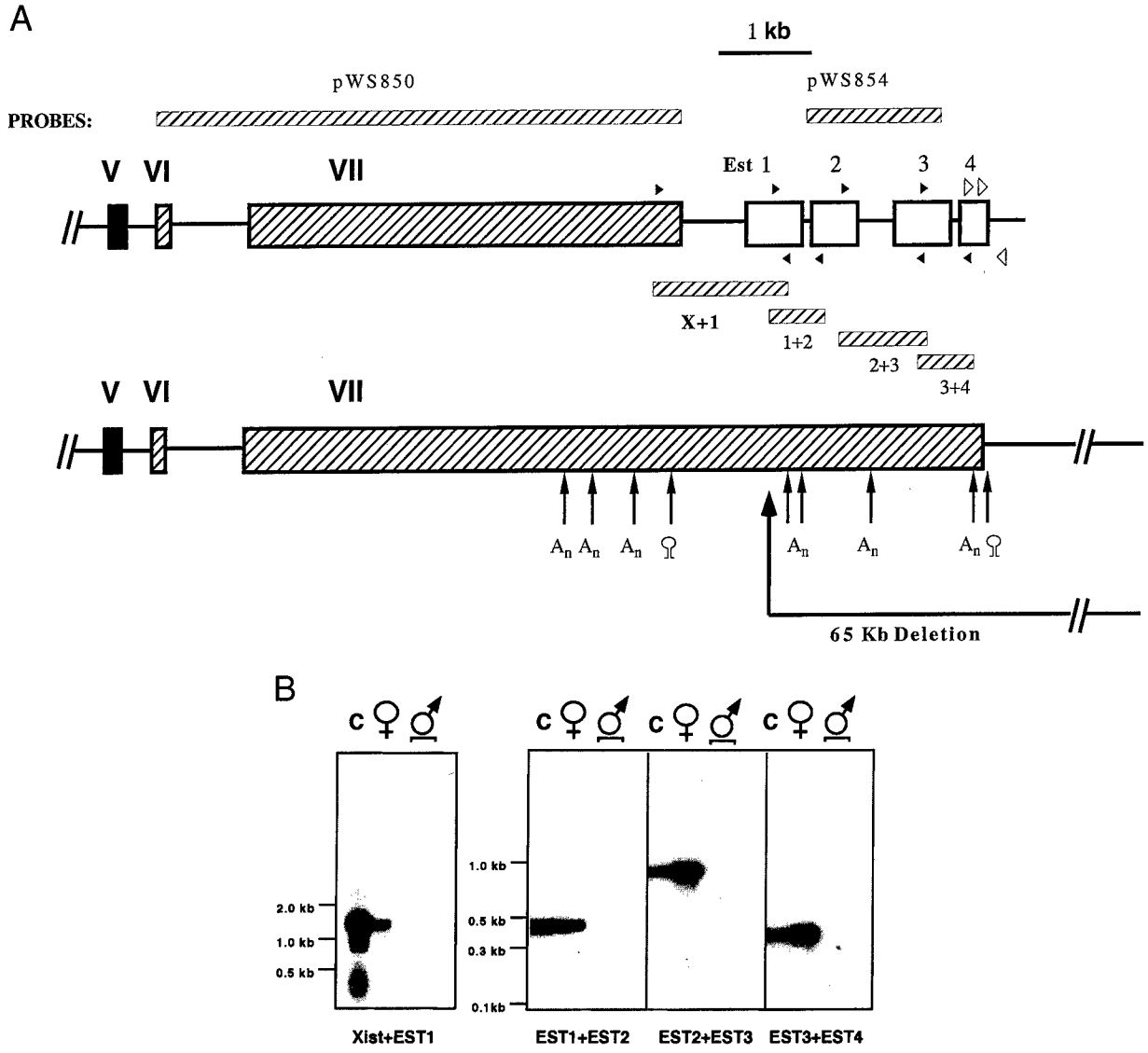


FIG. 2. New structure of the 3′end *Xist*. (*A*) The four ESTs (EST1≈4) are mapped relative to exon VII of *Xist* (GenBank accession nos. of EST1 ≈ 4: AA543875, AA221611, AA690387[§], and R74734, respectively). PCR fragments synthesized by using primers whose locations are marked by closed arrowheads 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

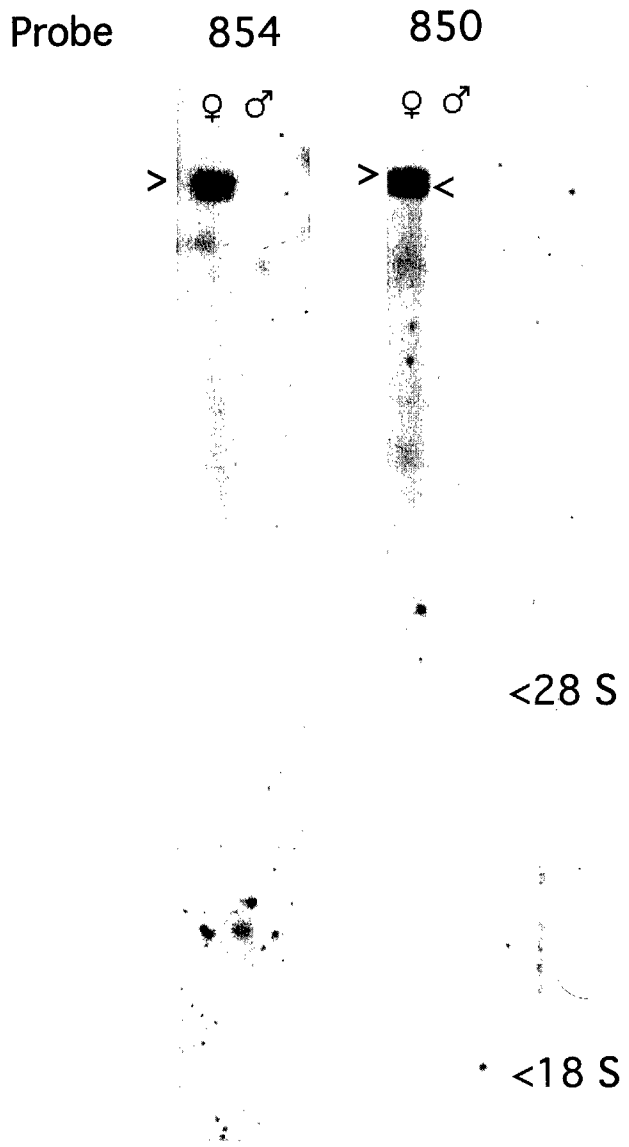


FIG. 3. Northern blot showing major and minor *Xist* species. Murine male (♀) and female (♂) kidney RNA was fractionated on a formaldehyde-agarose gel and transferred to positively charged nylon yielding duplicate strips. Duplicate lanes were hybridized to the individual probes: pWS850, pWS854, and mx8 (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′

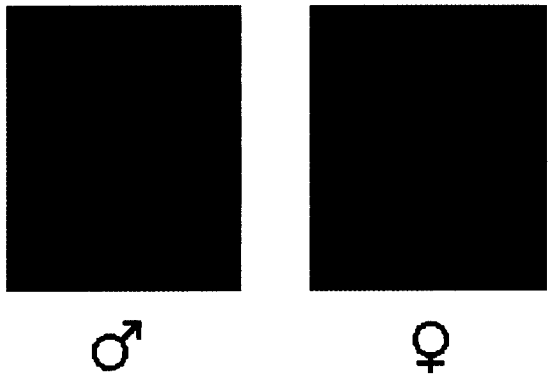


FIG. 4. RNA–FISH photomicrograph of male and female somatic cells. RNA–FISH was performed to visualize cytoplasmic/nuclear RNA that hybridized to *Xist* probes pWS850 (FITC) and pWS854 (rhodamine). Hybridizations of *Xist* probes were performed simultaneously and separate channels recorded; they were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