Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 8709–8714, July 1998
Genetics

Chromosomal basis of X chromosome inactivation: Identification of a multigene domain in Xp11.21-p11.22 that escapes X inactivation

ANDREW P. MILLER* AND HUNTINGTON F. WILLARD†

Department of Genetics and Center for Human Genetics, Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, OH 44106

Communicated by Shirley M. Tilghman, Princeton University, Princeton, NJ, May 22, 1998 (received for review January 30, 1998)

ABSTRACT **A number of genes have been identified that escape mammalian X chromosome inactivation and are expressed from both active and inactive X chromosomes. The basis for escape from inactivation is unknown and, *a priori*, could be a result of local factors that act in a gene-specific manner or of chromosomal control elements that act regionally. Models invoking the latter predict that such genes should be clustered in specific domains on the X chromosome, rather than distributed at random along the length of the X. To distinguish between these possibilities, we have constructed a transcription map composed of at least 23 distinct expressed sequences in an ≈5.5-megabase region on the human X chromosome spanning Xp11.21-p11.22. The inactivation status of these transcribed sequences has been determined in a somatic cell hybrid system and correlated with the position of the genes on the physical map. Although the majority of transcribed sequences in this region are subject to X inactivation, eight expressed sequences (representing at least six different genes) escape inactivation, and all are localized to within a region of less than 370 kb. Genes located both distal and proximal to this cluster are subject to inactivation, thereby defining a unique multigene domain on the proximal short arm that is transcriptionally active on the inactive X chromosome.**

X chromosome inactivation is the process whereby one of the two X chromosomes in normal diploid female cells is inactivated to compensate for the dosage difference of X-linked genes between males and females (1). One of the most intriguing aspects of X inactivation in humans is that certain genes have been found that escape inactivation and are expressed from both X chromosomes (2–4). Although the basis for the expression of these genes from the inactive X chromosome is unclear at present, their study is likely to be informative for understanding the chromosomal mechanisms involved in X inactivation, implying the existence of local and/or chromosomal signals that distinguish genes that escape inactivation from those that are subject to inactivation.

Both gene-specific and chromosome-wide models have been invoked to explain the behavior of these genes and their ability to be expressed from otherwise inactive X chromosomes. One possibility is that each X-linked gene has associated with it specific sequences that are able to bind both activator and repressor proteins and that the combination of these sites for any single gene determines whether it will be subject to inactivation (5). However, analysis of human X;autosome translocations indicates that such putative gene-specific elements must not be restricted to X-linked genes, because autosomal genes that are translocated to the X chromosome can either remain active (6–8) or be inactivated (7, 9–12). Further, expression analysis of autosomal transgenes inserted into the mouse X chromosome also implies that such gene-specific elements must be reasonably widespread in mammalian genomes (13, 14). A similar model acknowledges the well established chromosomal control of X inactivation, such that physical linkage to the X chromosome usually is sufficient for any given gene to be inactivated (2, 14). Against this background, however, certain X-linked genes may escape inactivation because they contain a specific sequence, such as a unique promoter or enhancer element, that allows that particular gene to be transcribed even from the context of the otherwise inactive X. Under these models, one would expect genes that escape X inactivation to be randomly distributed on the chromosome with respect to genes that are subject to inactivation.

In contrast to these gene-specific models, there may be specific subchromosomal domains on the X chromosome that are regulated differentially along the length of the X (14). Genes that escape inactivation would do so, not because of their particular features, but because of their specific physical location on the chromosome, i.e., within an active subchromosomal domain. This model predicts that genes that escape X inactivation should be clustered (although not necessarily exclusively) within these domains. This model also implies that such domains might be flanked by DNA sequences such as boundary elements (15) that would prohibit regulatory signals from influencing genes in neighboring domains on the chromosome. A refinement of this domain model proposes that the inactivation signal spreads outward from the center of each domain, such that genes that are immediately adjacent to the origin of the signal would be completely inactivated and genes that lie at the edge of the domain could escape inactivation because of increased distance from the inactivation signal. This model suggests that genes mapping an intermediate distance between the center and the edge of the domain might be partially inactivated, with expression from the inactive X being reduced compared with the active allele. This might be the case for loci such as the human *STS* or mouse *Xe169* genes, which are expressed from the inactive X (16–18), but only at a fraction of the level seen from the active X (19–21).

To differentiate between these different models requires analysis of the distribution of a number of transcribed sequences across a defined region of the human X chromosome. Four independent lines of evidence suggest that the proximal portion of the short arm of the human X chromosome is transcriptionally active on the inactive X chromosome and thus is suitable for such an analysis. First, replication studies have demonstrated that this

Abbreviations: Mb, megabase; YAC, yeast artificial chromosome; EST, expressed sequence tag; STS, sequence-tagged site; RT-PCR, reverse transcription–PCR.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF069134–AF069138).

*Present address: Sequana Therapeutics, 11099 North Torrey Pines Road, Suite 160, La Jolla, CA 92037. e-mail: andrew_miller@sequana.com.

†To whom reprint requests should be addressed at: Department of Genetics, Case Western Reserve University School of Medicine, 2109 Adelbert Road, Cleveland, OH 44106-4955. e-mail: hfw@po.cwru.edu.

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.

region replicates early during S phase on the inactive X chromosome (22), suggesting that it might contain euchromatic sequences. Further, analysis of the structure of the Barr body in isodicentric Xq chromosomes suggests that this region lies outside the Barr body and, therefore, is not heterochromatic (23). Additionally, studies of histone acetylation on the inactive X chromosomes have demonstrated that this region is acetylated on the inactive X chromosome, in contrast to most of the rest of the chromosome, which is underacetylated (24). Finally, phenotype/genotype correlations in patients with abnormal X chromosomes and features of Turner syndrome suggest that this region contains genes that are involved in this phenotype (25). Such genes are predicted to escape X inactivation and to have functional Y homologues (26).

We have reported previously the construction of an ≈5.5-megabase (Mb) yeast artificial chromosome (YAC) contig spanning the most proximal portion of the short arm of the X chromosome, Xp11.21-p11.22, thus providing a substrate for the construction of a transcription map spanning this region (27). In this report, we have examined the X inactivation status of 23 expressed sequences mapping to this contig and provide evidence strongly supporting chromosomal domain models of X inactivation.

MATERIALS AND METHODS

Direct cDNA Selection. Direct cDNA selection was carried out by using cDNA from a mouse/human somatic cell hybrid, AHA-llaB1 (28), which contains an active X chromosome as its only human component. Total RNA was isolated by using RNAzol (Cinna/Biotecx Laboratories, Friendswood, TX). Approximately 5 μg of poly(A)$^+$ RNA was converted to double-stranded cDNA by using a cDNA synthesis system (GIBCO/BRL). The resulting cDNA was digested with *Mbo*I before the addition of a cDNA linker primer, amplified with the linker primer *Mbo*Ib in a 100-μl PCR (29), and blocked with COT1 DNA (GIBCO/BRL) for 4 hr at 60°C.

Cosmid DNA or total Alu-PCR products from YAC clones were biotinylated by using the BioNick Labeling System (GIBCO/BRL). Approximately 100 ng of biotinylated DNA was then denatured and allowed to hybridize to 1 μg of blocked cDNA in a total volume of 20 μl in 120 mM $NaPO_4$ for 60 hr at 60°C under mineral oil. After hybridization, the biotinylated DNA was captured on streptavidin-coated magnetic beads (Dynal) in 100 μl of binding buffer (1 M NaCl/10 mM Tris, pH 7.4/1 mM EDTA) for 20 min at room temperature with constant rotation. After washing, the bound cDNA was eluted with 50 μl of 0.1 M NaOH for 10 min followed by neutralization with 50 μl 1 M Tris, pH 7.4. PCR was performed on the eluate with a linker primer. This amplified primary selected cDNA was blocked with 1 μg of COT1 DNA at 60°C for 1 hr, followed by a second round of hybridization to 100 ng of the appropriate genomic DNA under the same conditions as the first round of selection. PCR-amplified secondary-selected cDNA was cloned by using the TA cloning system (Invitrogen). PCR products from the resulting colonies were analyzed on a 1.5% agarose gel, transferred to nylon membranes, and hybridized with probes corresponding to the genomic DNA used for the selection. PCR was used to localize the selected transcribed sequences on a YAC contig spanning Xp11.21-p11.22 (27).

X Inactivation Analysis. The somatic cell hybrid panel used to determine expression from active and inactive X chromosomes has been described (2, 3, 30). This panel contains two active X-containing hybrids (AHA-11aB1 and t60–12), both of which retain the X chromosome as the only human component. In this study, seven independent, inactive X-containing hybrids were used to examine expression from the inactive X. Cell culture, RNA preparation, and reverse transcription (RT)–PCR were performed as described (3). In several cases, selective restriction enzyme digestion was used to distinguish human and mouse RT-PCR products.

Northern Blot Analysis. Multiple tissue Northern blots were obtained from CLONTECH. Probe labeling and hybridization were carried out according to the supplied protocol. All solutions to come in contact with the probes or blots were treated with diethyl-pyrocarbonate.

DNA Sequence Analysis. DNA sequencing was performed by using an Applied Biosystems 373 automated fluorescence sequencer. One microgram of plasmid DNA or 50 ng of PCR product was sequenced by using *Taq* polymerase cycle sequencing according to the Applied Biosystems protocol, using 3.2 pmol of the appropriate primer. Sequence comparisons with GenBank were accomplished by using the BLAST function available through the World Wide Web (WWW): http://www.ncbi.nlm.nih.gov/Recipon/blast_search.html (31). Identification of protein motifs was achieved by using the BLOCKS function available on the WWW: http://www.blocks.fhcrc.org (32).

RESULTS

Identification of Expressed Sequences. A number of previously characterized genes have been localized within the Xp11.21-p11.22 contig (27) and formed a starting point for the transcription map. Several different strategies were used to identify additional transcribed sequences within this region.

As a first approach to identifying novel transcribed sequences within the contig, YACs from the contig spanning Xp11.21-p11.22 were used to query the physical mapping database at the MIT Genome Center (http://www-genome.wi.mit.edu/), containing the results of both sequence-tagged site (STS) and expressed sequence tag (EST) content analysis on a subset of CEPH YACs. This approach allowed us to identify six distinct ESTs mapping to the region of interest (Table 1). Primers for these loci were used to confirm the location of these ESTs on the X chromosome and to localize them on the YAC contig (Fig. 1*A*).

Three other genes or ESTs that mapped to the pericentromeric region of the X chromosome were identified through the literature. One of these ESTs, WI-7092, was developed from the 3′ untranslated region of the mitochondrial ubiquinone-binding protein, a nuclear-encoded component of ubiquinol-cytochrome *c* oxidoreductase (33). Although this sequence appears to map to both Xp11.21 and to chromosome 8, RT-PCR analysis suggests that only the X chromosome locus is expressed (data not shown). IB772 is an EST that has homology to the murine upstream regulatory element binding protein (UreB1) (34). Trophinin is a novel human transcript expressed on the trophectoderm surface of blastocysts and is suggested to be involved in embryo implantation (35). This gene originally was mapped to proximal Xp (36) and was subsequently fine-mapped in this analysis.

One expressed sequence in the region was identified by the cross-species hybridization of the corresponding genomic DNA (data not shown). Mapping of this genomic fragment suggests that it corresponds to a duplicated locus on the X, one locus on either side of an X;17 translocation breakpoint (6, 37) (Figs. 1 and 2*A*). Northern blot analysis suggests that this transcript is expressed only in skeletal muscle tissue (Fig. 2*B*). Sequence analysis and PCR of a skeletal muscle cDNA library demonstrated that only the proximal locus, FIJG-5, is expressed (data not shown). This expressed sequence is a plausible candidate gene for a severe form of X-linked arthrogryposis (MIM 301830) that has been mapped to the most proximal portion of Xp (38).

To identify additional expressed sequences from Xp11.21-p11.22, direct cDNA selection (29) was carried out by using cDNA derived from a mouse/human somatic cell hybrid that retains the active X chromosome as the only human component (2, 28). Because two genes that escape X inactivation already had been localized to the very distal portion of the YAC contig (27), direct cDNA selection experiments were concentrated in this region, using a partial cosmid and P1 contig (Fig. 1*B*) as the genomic target DNA. Fig. 2*C* shows representative results of these selections. Alu-PCR products derived from YAC clones

Table 1. Expressed sequences mapping to Xp11.21-p11.22

Expressed sequence	Properties	Ref.
SSX2	Synovial sarcoma; gene family with duplicated loci	27, 44, 45
IB3700	EST*	
	F-ggggtattgcaacatttaca	
	R-tgtccatctctcttcccttc (289 bp)	
DXS6672E	EST*	
DXS1008E	EST*	
XE169/SMCX	XY-homologous gene, escapes inactivation in human and mouse	17, 18, 20, 21, 27
ADS56	Selected cDNA	
	F-tcagagaagcccaaggtcc	
	R-gctgtgtgtcctaactcctctg (148 bp)	
ADS20/DXS1000E	Selected cDNA; contains expressed $(CA)_n$ from DXS1000E	
	F-cactgaccagctgtgtgagc	
	R-agatgttttgctcctggctg (171 bp)	
DXS423E	Chromosome condensation, escapes inactivation	43
ADS9	Selected cDNA; homology to short-chain alcohol dehydrogenases	40
	F-agcgtgtgttttctctgcct	
	R-gaaagtgtgcaacttcttggc (181 bp)	
ADS13	Selected cDNA	
	F-cactctggtagccagtgcag	
	R-ttcgtatctgtgccctccat (150 bp)	
FIJG-5	Skeletal muscle-specific transcript; duplicated locus	
IB772	EST; similar to murine transcription activator *UreB1*	34
	F-gtttcagtctcaatgcagc	
	R-cctgtttcctcaccttctct (163 bp)	
DXS1013E	EST*	
FGD1	Guanine nucleotide exchange factor; Aarskog–Scott syndrome, faciogenital dysplasia gene	
DXS7159E	EST*; MAGE/Necdin homology	
	F-agaaatggcaaatagatttaatgc	
	R-gggatgtcaattgacccttt (177 bp)	
Trophinin	Involved in embryo implantation	35, 36
	F-taaattgcagtagtccttccca	
	R-acagcacacagaaagcaaaca (152 bp)	
ALAS2	Sideroblastic anemia; erythroid-specific 5-aminolevulinate synthase	
NIB1132	EST*; MAGE/Necdin homology	
	F-ctatgtcctgcagtaactaagg	
	R-tgaagaaggatggaggattt (111 bp)	
ADS40/NIB723	Selected cDNA, EST; unigene cluster; Krueppel-like Zn-finger	41
	F-agtgtgtgatttctcatcatgc	
	R-aactcaaggctgtgaacaaaca (700 bp)	
QP-C/WI-7092	Mitochondrial ubiquinone-binding protein of complex III	33
	F-ttgaagtctgtggatgcagc	
	R-ttcaaaaactccagccattaca (125 bp)	
ADS37/DXS579E	Selected cDNA	
	F-caggtattgctggtaggaaatg	
	R-atgagtggattggtgggttc (112 bp)	
ZXDB	Zinc finger protein, duplicated locus	2, 42
ZXDA	Zinc finger protein, duplicated locus	2, 42

Expressed sequences were identified from several sources, as described in the text. Novel cDNAs isolated by direct selection are indicated, with forward (F) and reverse (R) primers used in RT-PCR assay. Length of the amplifed genomic or cDNA fragment is given.

*ESTs identified from query of the MIT Genome Center database.

covering Xp11.21 also were used in additional selection experiments.

Eight different expressed sequences were obtained in this manner (Table 1 and data not shown). Two of these were selected with Alu-PCR products, one with a P1 clone containing *XE169*, and five with cosmid clones. Two of the selected clones, ADS20 and ADS37, matched known STS/EST markers from the region, DXS1000E and DXS579E, respectively. Clone ADS9 detected a single transcript of ≈1.0 kb expressed in multiple tissues (Fig. 2*D*). Three EST clones were obtained that allowed us to build a sequence contig of 968 bp that includes a potential ATG start codon within a consensus Kozak sequence (39), a stop codon, a polyadenylation site, and a poly(A) tail. Sequence analysis suggests that this transcript encodes a protein of 261 aa related to the short-chain alcohol dehydrogenase family (Prosite database, accession no. PS00061) (40). Another selected cDNA, ADS40, mapped adjacent to an EST, NIB723, within the YAC contig (Fig. 1). An RT-PCR connection strategy using primers from both ADS40 and NIB723 identified them as being part of the same transcript. This sequence then was used to identify a UNIGENE cluster that has significant homology to Krueppel-like Zn-finger genes (41).

X Inactivation Transcription Map. To develop the transcription map shown in Fig. 1, gene-specific primers were used to localize and order 23 transcribed sequences along the Xp11.21-p11.22 region by analyzing the marker content of individual YAC

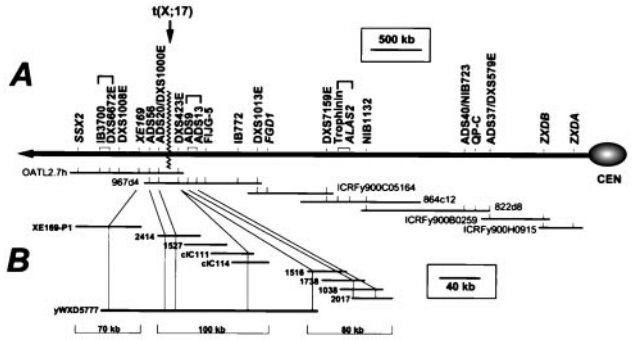

FIG. 1. Map of transcribed sequences in Xp11.21-p11.22. (*A*) Transcribed sequences have been placed along the chromosome based on their map position on a complete YAC contig spanning Xp11.21-p11.22. The names and positions of seven YACs that span the region are given (27). Three pairs of expressed sequences cannot be ordered relative to each other, and their locations on the map are indicated by brackets. The position of a t(X;17) breakpoint (6, 37) is indicated by the arrow. (*B*) P1, YAC, and cosmid contig of the region spanning the t(X;17) breakpoint. Bracketed regions span at least 250 kb, but <370 kb, as described in the text.

clones spanning the region (27). The X inactivation status of some genes in the region was known before this study. *ZXDA* and *ZXDB* were shown previously to be subject to X inactivation (2, 42), whereas *XE169* (*SMCX*) and DXS423E were known to escape inactivation (17, 18, 43). *SSX* was shown previously to be expressed from the inactive X chromosome (27), but the recent identification of multiple copies of *SSX*-related sequences in Xp11.21–11.23 (44, 45) has complicated the X inactivation analysis of any particular member of this multigene family. For this reason, *SSX2* sequences were not included in the subsequent analysis. In addition, *ALAS2* and FIJG-5 (Fig. 2*B*) show tissue-specific expression and are not expressed in fibroblasts; thus, they were not considered further.

Sixteen newly identified expressed sequences within the region were analyzed for their inactivation status by using an RT-PCR approach based on mouse/human somatic cell hybrid lines that retain either an active or an inactive X chromosome (2, 3). Two active X-containing hybrids and seven inactive X-containing hybrids were assayed along with the parental mouse cell line and female genomic DNA as negative and positive controls, respectively. All of the gene-specific primers used in this study amplify the same sized products from both DNA and cDNA, so cDNA synthesis reactions without reverse transcriptase were used to control for DNA contamination in the RNA samples. Data for a subset of these genes are shown in Fig. 3. Eleven of the expressed sequences were expressed only from active X hybrids and appear, on that basis, to be subject to X inactivation. An additional sequence, IB772, was expressed from the active X hybrids and one inactive X hybrid, but not from the remaining six inactive X hybrids. As observed previously (3), this pattern suggests that IB772 also is subject to X inactivation. Finally, four transcripts were expressed both from active X hybrids and at least five inactive X-containing hybrids and thus escape X inactivation.

A Domain That Escapes X Inactivation. All expressed sequences that we have examined that escape X inactivation are clustered within the distal portion of the Xp11.21-p11.22 YAC contig. In addition, two other selected clones in the same region also escape inactivation by this analysis (data not shown). However, because their map positions are indistinguishable from ADS9 and ADS20, we cannot conclude at present that these represent distinct transcription units. Combined with previous data (17, 18, 42, 43), therefore, at least six (and perhaps as many as eight) distinct expressed sequences define a domain that escapes X inactivation (Fig. 4).

Analysis of contigs spanning this region suggests that these eight sequences are located within a region of no more than 370

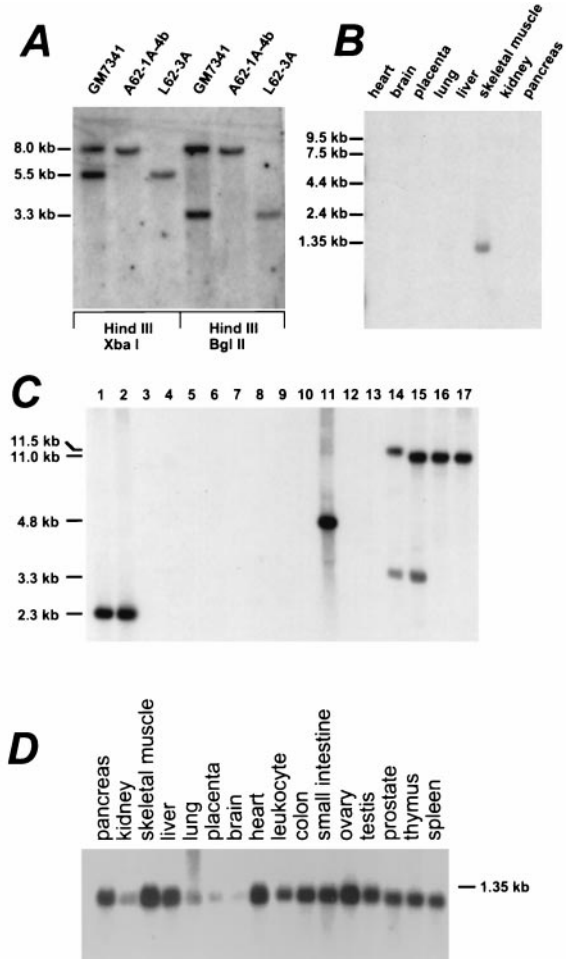

FIG. 2. Southern and Northern analysis of expressed X-linked sequences and cDNA clones. (*A*) A 2.2-kb genomic fragment, FIJG, was hybridized to a Southern blot containing DNA digested with *Hin*dIII and either *Xba*I or *Bgl*II. The lanes correspond to female genomic DNA (GM7341), and two somatic cell hybrids retaining reciprocal portions of a t(X;17) (6). A62–1A-4b contains Xp11.21-qter, and L62–3A contains Xp11.21-pter (37). This probe detects two loci in genomic DNA, one of which maps distal to the t(X;17) breakpoint (5.5- and 3.3-kb bands, referred to as FIJG-4), whereas the other maps proximal to the breakpoint (8.0-kb band, referred to as FIJG-5). The probe is colinear with genomic DNA and does not contain *Hin*dIII, *Xba*I, or *Bgl*II restriction sites. (*B*) Expression patterns were examined on a multiple-tissue Northern blot (CLONTECH) by using the same 2.2-kb genomic probe. A transcript of ≈1.3 kb was detected in skeletal muscle tissue only. (*C*) Secondary selected cDNA products from a typical direct cDNA selection experiment were used as a complex probe and hybridized to *Eco*RI digests of a partial cosmid contig covering the region that was selected (lanes 1–17). Hybridizing fragments correspond to genomic fragments homologous to selected cDNA clones. (*D*) The selected cDNA clone ADS9 is widely expressed and detects on Northern blots a single RNA species of ≈1.0 kb in various tissues.

kb (Figs. 1*B* and 4). Cosmid and P1 clones cover approximately 250 kb of DNA with two gaps in the contig. The 3′ end of DXS423E and DXS1000E are in one cosmid contig of ≈100 kb, and the 5′ end of DXS423E, ADS9, and ADS13 lie in another cosmid contig of ≈80 kb. *XE169* is represented by a single P1 clone (≈70 kb). Additionally, YAC clone yWXD5777 (120 kb) is deleted internally but is positive for markers on both sides of the two gaps in the cosmid/P1 contig (Fig. 1*B*), indicating that the combined size of both gaps is no larger than 120 kb. This places an upper limit of 370 kb on the size of the region containing these genes. The actual distance may be smaller, but because we have been unable to close these two gaps in the cosmid/P1 contig, it is not possible to determine the exact distance. Three transcribed

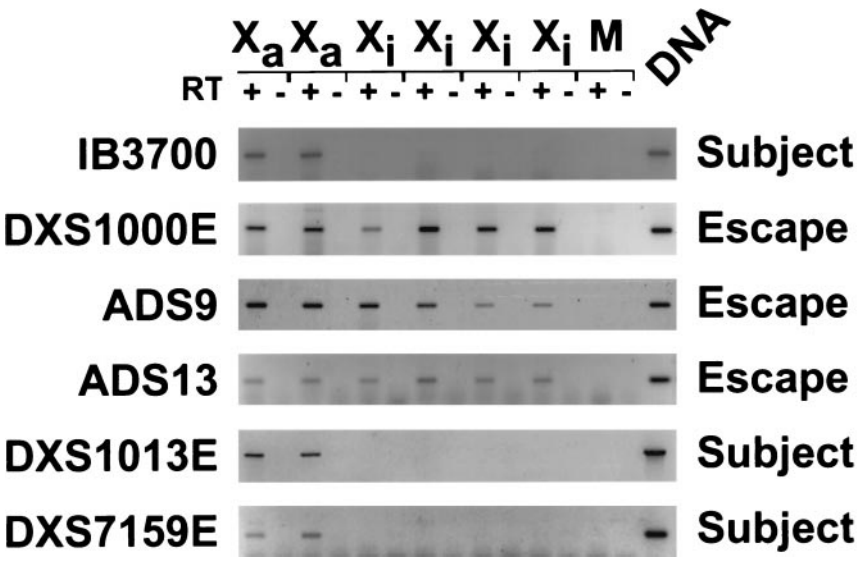

FIG. 3. Expression analysis of transcribed sequences in Xp11.21-p11.22. Negative images of ethidium bromide-stained agarose gels show the expression analysis of transcribed sequences from the active and inactive X chromosomes. cDNA synthesis reactions without reverse transcriptase (RT) were used to control for DNA contamination. DNA from a human female cell line (GM7002) was used as a positive control, and cDNA from the mouse parental cell line, M, was used as a negative control. The hybrids are (left to right): X_a, AHA-11aB1; X_a, t60–12; X_i, t11–4Aaz5; X_i, t48–1a-1Daz4A; X_i, t75–2 maz34–4a; X_i, t86-B1 maz1b-3a. On the basis of data from these and three additional inactive X hybrids, IB3700, DXS1013E, and DXS7159E appear to be subject to X inactivation, whereas DXS1000E, ADS9, and ADS13 escape X inactivation.

sequences located immediately distal to this cluster and 11 transcripts located between this domain and the centromere (a region of ≈5 Mb) are subject to X inactivation (Fig. 4), although we cannot exclude the possibility that the domain varies in size among different X chromosomes, because some genes demonstrate heterogeneous expression patterns in different hybrids (3).

DISCUSSION

To examine the chromosomal basis of X inactivation, we have constructed a transcription map of the most proximal portion of the human X chromosome short arm, in Xp11.21-p11.22, and have analyzed the expression of 23 distinct transcribed sequences from the active and inactive X chromosomes. All of the sequences that were found to be expressed from both active and inactive X chromosomes lie within a region that is no larger than 370 kb. The total number of genes represented by these expressed sequences is somewhat uncertain, because it is possible that some of the selected clones correspond to different fragments from the same transcript. Further, it is conceivable that some of the expressed sequences represent unspliced heteronuclear RNA products that were enriched during the cDNA selection process. Nonetheless, based on genes of known complete structure and on Northern blot analysis (Fig. 2 and data not shown), the cluster must comprise no fewer than six independent genes. Regardless of the exact number and identity of genes in this region, it is noteworthy that no sequences were identified within this region of ≈370 kb that are subject to X inactivation. This suggests that the region corresponds to a novel chromosomal domain that is coordinately regulated to escape X chromosome inactivation.

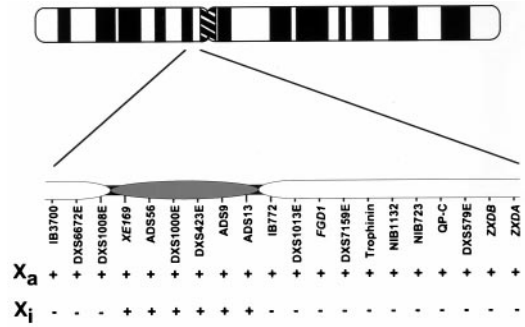

FIG. 4. Gene expression (+, −) from active (X_a) and inactive (X_i) X chromosomes and model for the chromosomal basis for genes escaping X inactivation. Six genes that escape inactivation map exclusively within a domain (denoted by the shaded oval) that is transcriptionally active on the inactive X chromosome. Genes mapping both distal and proximal to this domain are subject to inactivation, implying that they lie in different domains that are transcriptionally silent on the inactive X chromosome.

The data presented here extend less comprehensive data describing the clustering of both pseudoautosomal and other non-pseudoautosomal genes that escape X inactivation (2, 4). For example, *UBE1* and *PCTK1*, both of which escape X inactivation, map within 5 kb of each other (46). In addition, a novel gene, *INE2*, was shown to escape inactivation (47) and maps adjacent to *GRPR* (48), which also escapes inactivation (49). These two regions of the X chromosome may, therefore, contain multigene domains that escape inactivation, and systematic identification of additional genes within these regions would allow testing of this hypothesis.

Some, but not all, genes that escape X inactivation have homologues on the Y chromosome (4). Of those genes analyzed in this study, *XE169* (*SMCX*) is known to have a Y-linked homologue (17, 18), whereas DXS423E does not (43). ADS9 and ADS20 did not reveal any Southern blot differences between male and females (data not shown), suggesting that they, too, do not have an identifiable Y homologue. However, some cDNAs tested from this region may be too small to provide definitive data on this point in the absence of complete gene isolation and characterization.

In light of these data, we can review the different models that have been invoked to explain the behavior of genes that escape X inactivation. The finding of a chromosomal domain composed exclusively of genes expressed from both X chromosomes appears to be inconsistent with gene-specific models, but rather supports a domain model for genes that escape X inactivation. Two related domain models have been proposed, and the current data are broadly consistent with both. One model posits that the X inactivation signal spreads outward from specific sequences along the length of the X [e.g., "booster elements" (50)] and that genes that lie a large distance away from such elements thereby may escape inactivation. Such a model predicts that genes within a single domain will be inactivated in a gradient, such that genes near the center of the domain are inactivated completely, and genes some distance from the center are only partially inactivated (and thus partially expressed from the inactive X). Although some genes have been identified that are expressed from the inactive X at a fraction of the level from the active X (19–21), there currently is no evidence that this depends on the physical location of the gene within a specific domain. Although the current studies were not intended to be quantitative, cDNA dilution experiments have shown that all of the genes within this domain are well expressed from the inactive X, at levels near those from the active X (data not shown). Further experiments to more precisely quantitate the level of inactive X expression of

the transcribed sequences examined here would be required to address definitively this prediction of the model. This model further predicts that genes that escape inactivation because of their distance from such elements would map along the length of the X chromosome. However, based on the currently available information, this does not appear to be the case. Nonetheless, the resolution of the current X chromosome transcription map and the lack of X inactivation data for a sufficient amount of genes and ESTs along the entirety of the chromosome might obscure this observation.

A more general domain model, in which genes are partitioned into subchromosomal domains that determine whether genes within each domain are likely to be subject to or escape from inactivation, also appears to fit our observation of a cluster of genes that escape inactivation. This model, although not necessarily mutually exclusive with the domain gradient model, focuses attention on possible boundary elements to delimit the domain, similar to the *scs* elements flanking the *hsp70* locus in *Drosophila* and the 5′ elements in the chicken β-globin domain (51, 52). These sequences prevent regulatory elements within one domain from affecting gene expression in an adjacent domain (reviewed in ref. 15). The location of such elements now can be predicted, based on the X inactivation map presented here and using the physical mapping resources described here and previously (27).

The identification of a domain that escapes X inactivation has interesting implications for X;autosome translocations. The breakpoint in a cytogenetically balanced X;17 translocation (6, 37) lies within this domain, between the 3′ end of DXS423E and DXS1000E (Fig. 1*A*). Although the normal X was inactivated in most cells in this patient, in approximately 25% of her cells, the derivative X was inactivated, but with no spreading of inactivation into the autosomal material (6). We raise the possibility that X inactivation does not spread into the adjacent autosomal material because the breakpoint lies within a domain that normally escapes X inactivation. The lack of documented spread of inactivation even at a greater distance away from the translocation breakpoint may be consistent with recent data suggesting that some autosomal DNA may not contain the sequence elements required to fully nucleate inactivation domains (53). Several other X;autosome translocations also do not demonstrate spreading of inactivation into the autosome (6–8, 54); these breakpoints may also lie within domains that escape X inactivation, a prediction that can be tested once the breakpoints are mapped precisely. In contrast, spreading of X inactivation into autosomal regions has been reported for some translocations with breakpoints in other regions of the X chromosome (9–12, 54, 55). We predict that these breakpoints lie in regions of the X chromosome that are subject to inactivation, thus allowing inactivation to spread into the autosomal portion.

The identification of a coordinately regulated domain such as the one reported here on the X chromosome has implications for other regions of the genome that demonstrate coordinately regulated gene expression. Genomic imprinting is defined by the presence of transcripts that are expressed exclusively from either the paternally or maternally inherited allele (56). X inactivation and imprinting appear to be similar in that one allele of an imprinted gene is silenced on an otherwise active chromosome, whereas genes that escape X inactivation are expressed from an otherwise inactive X chromosome (57). The most relevant comparison for this work is the finding that imprinted genes are clustered, similar to the results presented here for X inactivation, although imprinted genes within the same cluster can be expressed exclusively from either allele (58–62). Together, the evidence supporting subchromosomal domain models for both X inactivation and genomic imprinting provides functional correlates for the chromosomal loop model of genome organization (63).

We thank members of the Willard laboratory for helpful discussions, L. Carrel for discussions and comments on the manuscript, D. Ledbetter for communicating results before publication, A. Cottle for assistance, and A. Cadle for providing data on the *FGD1* gene. This work was performed in partial fulfillment of the requirements for the Doctor of Philosophy from Stanford University to A.P.M. The research was supported in part by a grant from the National Institutes of Health to H.F.W. (GM45441).

1. Lyon, M. F. (1961) *Nature (London)* **190,** 372–373.
2. Willard, H. F., Brown, C. J., Carrel, L., Hendrich, B. & Miller, A. P. (1993) *Cold Spring Harbor Quant. Symp. Biol.* **58,** 315–322.
3. Brown, C., Carrel, L. & Willard, H. F. (1997) *Am. J. Hum. Genet.* **60,** 1333–1343.
4. Disteche, C. (1995) *Trends Genet.* **11,** 17–22.
5. McBurney, M. W. (1988) *BioEssays* **9,** 85–88.
6. Disteche, C. M., Swisshelm, K., Forbes, S. & Pagon, R. A. (1984) *Hum. Genet.* **66,** 71–76.
7. Hagemeijer, A., Hoovers, J., Smit, E. M. E. & Bootsma, D. (1977) *Cytogenet. Cell Genet.* **18,** 333–348.
8. Palmer, C. G., Hubbard, T. W., Henry, G. W. & Weaver, D. D. (1980) *Am. J. Hum. Genet.* **32,** 179–187.
9. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. (1982) *Am. J. Hum. Genet.* **34,** 811–817.
10. Couturier, J., Dutrillaux, B., Garber, P., Raoul, R., Croquette, M. F., Fourlinnie, J. C. & Maillard, E. (1979) *Hum. Genet.* **49,** 319–326.
11. Summitt, R. L., Martens, P. R. & Wilroy, R. S. (1974) *J. Pediatr.* **84,** 539–546.
12. Leisti, J. T., Kaback, M. M. & Rimoin, D. L. (1975) *Am. J. Hum. Genet.* **27,** 441–453.
13. Krumlauf, R., Chapman, V. M., Hammer, R. E., Brinster, R. L. & Gartler, S. M. (1986) *Nature (London)* **319,** 224–226.
14. Goldman, M. A. (1988) *BioEssays* **9,** 50–55.
15. Kioussis, D. & Festenstein, R. (1997) *Curr. Opin. Genet. Dev.* **7,** 614–619.
16. Shapiro, L. J., Mohandas, T., Weiss, R. & Romeo, G. (1979) *Science* **204,** 1224–1226.
17. Agulnik, A. I., Mitchell, M. J., Mattei, M., Borsani, G., Avner, P. A., Lerner, J. L. & Bishop, C. E. (1994) *Hum. Mol. Genet.* **3,** 879–884.
18. Wu, J., Salido, E. C., Yen, P. H., Mohandas, T. K., Heng, H. H. Q., Tsui, L.-C., Park, J., Chapman, V. M. & Shapiro, L. J. (1994) *Nat. Genet.* **7,** 491–496.
19. Migeon, B. R., Shapiro, L. J., Norum, R. A., Mohandas, T., Axelman, J. & Dabora, R. L. (1982) *Nature (London)* **299,** 838–840.
20. Sheardown, S., Norris, D., Fisher, A. & Brockdorff, N. (1996) *Hum. Mol. Genet.* **5,** 1355–1360.
21. Carrel, L., Hunt, P. A. & Willard, H. F. (1996) *Hum. Mol. Genet.* **5,** 1361–1366.
22. Willard, H. F. & Latt, S. A. (1976) *Am. J. Hum. Genet.* **28,** 213–227.
23. Therman, E., Sarto, G. E., Disteche, C. & Denniston, C. (1976) *Chromosoma* **59,** 137–145.
24. Jeppesen, P. & Turner, B. M. (1993) *Cell* **74,** 281–289.
25. Therman, E. & Susman, B. (1990) *Hum. Genet.* **85,** 175–183.
26. Ferguson-Smith, M. (1965) *J. Med. Genet.* **2,** 142–155.
27. Miller, A. P., Gustashaw, K., Wolff, D. J., Rider, S. H., Monaco, A. P., Eble, B., Schlessinger, D., Gorski, J. L., van Ommen, G. J., Weissenbach, J. & Willard, H. F. (1995) *Hum. Mol. Genet.* **4,** 731–739.
28. Dorman, B. P., Shimizu, N. & Ruddle, F. H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2363–2367.
29. Morgan, J. G., Dolganov, G. M., Robbins, S. E., Hinton, L. M. & Lovett, M. (1992) *Nucleic Acids Res.* **20,** 5173–5179.
30. Brown, C. J. & Willard, H. F. (1989) *Am. J. Hum. Genet.* **45,** 592–598.
31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
32. Henikoff, S. & Henikoff, J. G. (1994) *Genomics* **19,** 97–107.
33. Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M. & Ozawa, T. (1988) *Biochem. Biophys. Res. Commun.* **156,** 987–994.
34. Gu, J., Ren, K., Dubner, R. & Iadarola, M. J. (1994) *Mol. Brain Res.* **24,** 77–88.
35. Fukuda, M. N., Sato, T., Nakayama, J., Klier, G., Mikami, M., Aoki, D. & Nozawa, S. (1995) *Genes Dev.* **9,** 1199–1210.
36. Pack, S. D., Tanigami, A., Ledbetter, D. H., Sato, T. & Fukuda, M. N. (1998) *Cytogenet. Cell Genet.* **79,** 123–124.
37. Lafrenière, R. G., Brown, C. J., Powers, V. E., Carrel, L., Davies, K. E., Barker, D. F. & Willard, H. F. (1991) *Genomics* **11,** 352–363.
38. Kobayashi, H., Baumbach, L., Cox-Matise, T., Schiavi, A., Greenberg, F. & Hoffman, E. P. (1995) *Hum. Mol. Genet.* **4,** 1213–1216.
39. Kozak, M. (1986) *Cell* **44,** 283–292.
40. Zhuchenko, O., Bailey, J., Wehnert, M. & Lee, C. (1996) *Am. J. Hum. Genet.* **59,** A928 (abstr.).
41. Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H. & Fujii-Kuriyama, Y. (1993) *Nucleic Acids Res.* **21,** 1527–1532.
42. Greig, G. M., Sharp, C. B., Carrel, L. & Willard, H. F. (1993) *Hum. Mol. Genet.* **2,** 1611–1618.
43. Brown, C. J., Miller, A. P., Carrel, L., Rupert, J. L., Davies, K. E. & Willard, H. F. (1995) *Hum. Mol. Genet.* **4,** 251–255.
44. Chand, A., Clark, J., Cooper, C. & Craig, I. (1995) *Genomics* **30,** 545–552.
45. de Leeuw, B., Balemans, M., Olde Weghuis, D. & Geurts van Kessel, A. (1995) *Hum. Mol. Genet.* **4,** 1097–1099.
46. Carrel, L., Clemson, C., Dunn, J., Miller, A., Hunt, P., Lawrence, J. & Willard, H. (1996) *Hum. Mol. Genet.* **5,** 391–401.
47. Esposito, T., Gianfrancesco, F., Ciccodicola, A., D'Esposito, M., Nagaraja, R., Mazzarella, R., D'Urso, M. & Forobosco, A. (1997) *Genomics* **43,** 183–190.
48. Alitalo, T., Francis, F., Kere, J., Lehrach, H., Schlessinger, D. & Willard, H. (1995) *Genomics* **25,** 691–700.
49. Ishikawa-Brush, Y., Powell, J., Bolton, P., Miller, A., Francis, F., Willard, H., Lehrach, H. & Monaco, A. (1997) *Hum. Mol. Genet.* **6,** 1241–1250.
50. Gartler, S. M. & Riggs, A. D. (1983) *Annu. Rev. Genet.* **17,** 155–190.
51. Kellum, R. & Schedl, P. (1992) *Mol. Cell. Biol.* **12,** 2424–2431.
52. Chung, J. H., Whiteley, M. & Felsenfeld, G. (1993) *Cell* **74,** 505–514.
53. White, W. M., Willard, H. F., VanDyke, D. L. & Wolff, D. J. (1998) *Am. J. Hum. Genet.*, in press.
54. Keitges, E. A. & Palmer, C. G. (1986) *Hum. Genet.* **72,** 231–236.
55. Allderdice, P. W., Miller, O. J., Miller, D. W. & Klinger, H. P. (1978) *Am. J. Med. Genet.* **2,** 233–240.
56. Nicholls, R. D. (1994) *Am. J. Hum. Genet.* **54,** 733–740.
57. Tilghman, S. M. & Willard, H. F. (1995) in *Chromatin Structure and Gene Expression*, ed. Elgin S. (Oxford Univ. Press, Oxford), pp. 197–222.
58. Zemel, S., Bartolomei, M. S. & Tilghman, S. M. (1992) *Nat. Genet.* **2,** 61–65.
59. Glenn, C. C., Porter, K. A., Jong, M. T. C., Nicholls, R. D. & Driscoll, D. J. (1993) *Hum. Mol. Genet.* **2,** 2001–2005.
60. Wevrick, R., Kerns, J. A. & Francke, U. (1994) *Hum. Mol. Genet.* **3,** 1877–1882.
61. Villar, A. J. & Pedersen, R. A. (1994) *Nat. Genet.* **8,** 373–379.
62. Guillemot, F., Caspary, T., Tilghman, S. M., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Anderson, D. J., Joyner, A. L., Rossant, J. & Nagy, A. (1995) *Nat. Genet.* **9,** 235–241.
63. Saitoh, Y. & Laemmli, U. K. (1993) *Cold Spring Harbor Quant. Symp. Biol.* **58,** 755–765.