Cell, Vol. 71, 527–542, October 30, 1992, Copyright © 1992 by Cell Press

# The Human *XIST* Gene: Analysis of a 17 kb Inactive X-Specific RNA That Contains Conserved Repeats and Is Highly Localized within the Nucleus

**Carolyn J. Brown,*† Brian D. Hendrich,*† Jim L. Rupert,*† Ronald G. Lafrenière,*† Yigong Xing,‡ Jeanne Lawrence,‡ and Huntington F. Willard*†**
*Department of Genetics
Stanford University
Stanford, California 94305
‡Department of Cell Biology
University of Massachusetts Medical School
Worcester, Massachusetts 01655

## Summary

**X chromosome inactivation in mammalian females results in the cis-limited transcriptional inactivity of most of the genes on one X chromosome. The *XIST* gene is unique among X-linked genes in being expressed exclusively from the inactive X chromosome. Human *XIST* cDNAs containing at least eight exons and totaling 17 kb have been isolated and sequenced within the region on the X chromosome known to contain the X inactivation center. The *XIST* gene includes several tandem repeats, the most 5′ of which are evolutionarily conserved. The gene does not contain any significant conserved ORFs and thus does not appear to encode a protein, suggesting that *XIST* may function as a structural RNA within the nucleus. Consistent with this, fluorescence in situ hybridization experiments demonstrate localization of *XIST* RNA within the nucleus to a position indistinguishable from the X inactivation-associated Barr body.**

## Introduction

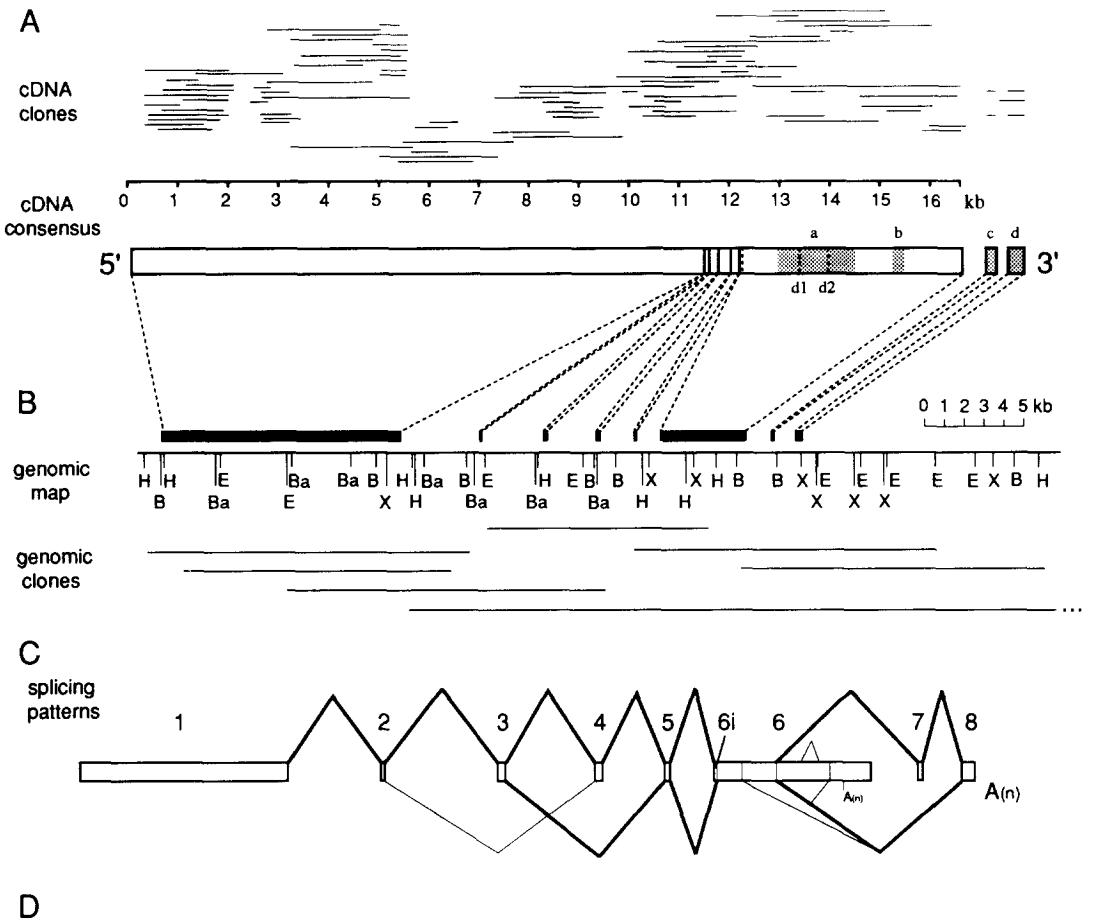
X inactivation occurs early in mammalian development to transcriptionally silence one of the pair of X chromosomes in females, thereby achieving dosage equivalence with males (Lyon, 1961). The X inactivation process is a unique cis-limited regulatory event that affects nearly an entire chromosome and that can be envisaged as involving three components, initiation, promulgation, and maintenance (Brown and Willard, 1992; Gartler and Riggs, 1983; Grant and Chapman, 1988). The initial choice of chromosome to be inactivated is random (Lyon, 1961); however, in individuals with multiple X chromosomes, all X chromosomes in excess of one are inactivated (Grumbach et al., 1963), arguing that the initiation event in X inactivation is a marking of the single X that is to remain active, rendering it unavailable or unresponsive to the X inactivation signal (Gartler and Riggs, 1983; McBurney, 1988; Rastan and Robertson, 1985). Because of its chromosomal nature, inactivation likely requires a spreading step that results in the cis-limited transcriptional inactivation of most, but not all, genes on the unmarked X chromosome. Finally, once an X has been inactivated, it must be stably maintained in the inactive state throughout somatic cell divisions. DNA methylation at the CpG-rich islands of some X-linked genes has been implicated in this maintenance step (Cattanach, 1975; Gartler and Riggs, 1983; Riggs, 1990a).

A number of sites and factors are undoubtedly involved in each of these steps, including a region on the proximal long arm of the X chromosome, called the X inactivation center (*XIC*), which is required in cis for inactivation of that chromosome (Russell, 1963; Lyon, 1971; Cattanach, 1975; Mattei et al., 1981; Brown et al., 1991a). This same region is further implicated in the processes of X inactivation as the site of condensation of the Barr body (Therman et al., 1974), the heterochromatic region found at the periphery of the nucleus that corresponds to the inactive X chromosome (Barr and Carr, 1962; Daly et al., 1977). The *XIC* could plausibly be involved in initiation, promulgation, and/or maintenance of X inactivation. The homologous *Xic* region on the mouse X chromosome (Rastan, 1983; Keer et al., 1990) has also been described to be the location of a locus (*Xce*), alleles of which determine the probability of an X chromosome being subject to X inactivation (Cattanach et al., 1969; Johnston and Cattanach, 1981), suggesting an involvement of *XIC/Xic* in at least the initiation step of X inactivation.

We recently described a gene, *XIST*, that maps to the *XIC* region (Brown et al., 1991b). In contrast with the majority of X-linked genes, which are believed to be subject to X inactivation and expressed only from the active X chromosome (Lyon, 1962; Brown and Willard, 1992), and in contrast with the growing number of genes that have been described that escape X inactivation and are therefore expressed from both the active and inactive X chromosome (Mohandas et al., 1977; Goodfellow et al., 1984; Schneider-Gädicke et al., 1989; Brown and Willard, 1989; Fisher et al., 1990; Franco et al., 1991; Yen et al., 1992), *XIST* is expressed only from the inactive X chromosome (Brown et al., 1991b). The mouse *Xist* gene has also been identified and is similarly expressed only from the inactive X chromosome and maps to the *Xic* region (Borsani et al., 1991; Brockdorff et al., 1991). The unique pattern of expression of *XIST* and its localization to a key region for X inactivation in both human and mouse suggest that it is either involved in, or directly influenced by, the process of X inactivation.

We have now isolated and sequenced 17 kb of *XIST* cDNA within eight different exons. Expression of all portions of the gene remains specific to the inactive X; among karyotypically normal individuals, *XIST* expression is, therefore, detected only in females. The lack of any extended conserved open reading frames (ORFs) within the sequence, as well as the nuclear localization of the majority of the transcripts, suggests that *XIST* may function as a structural RNA within the nucleus.

† Present address: Department of Genetics, Center for Human Genetics, Case Western Reserve University, Cleveland, Ohio 44106.




D

| EXON No. | EXON SIZE (bp) | SPLICE | SEQUENCE of 5' EXON | Intron Sequence | SEQUENCE of 3' EXON |
|---|---|---|---|---|---|
| 1 | 11,364 | 1/2 | TTAGAATACT | gtaagtacta....tgtgttccag | GATCCCATTG |
| 2 | 64 | 2/3 | GCTCCTCTTG | gtaatgacag....tcttcttaag | GACATTCTGA |
| 3 | 137 | 3/4 | GGAGAAAAAG | gtagtttggg....ctcttttgag | ATCTTCCTCA |
| 4 | 209 | 4/5 | ACACGTCAAG | gtgcgtaa.....ttttatag | CTCTTCATTG |
| 5 | 164 | 5/6 | GCTGAATGAA | gtaagttgtt.....nd | TGTGTATTTC |
| 6 | 4,543# | 5/6i | GCTGAATGAA | gtaagttgtt.....tatctaaag^ | TGTGTCTTAC |
| 7 | 146 | 7/8* | AAGCGAAAAG | gtttgtctat.....ctttcacag | TTTCTGGCAT |
| 8 | 377 | | | | |

Figure 1. Organization of the *XIST* Gene

(A) A consensus cDNA contig was derived from the over 70 cDNA clones shown at the top of the figure. The complete cDNA extends almost 17 kb and spans eight exons with additional internal alternative splice junctions (dotted vertical lines within the consensus). Previously published sequences (Brown et al., 1991b) are stippled (a–d), and previously described alternative splice donor sites are marked d1 and d2. The clones shown represent only the *XIST* portion of the coligated cDNA clones, and the scale prohibits visualizing some of the alternative splices summarized in (C).

(B) A restriction enzyme map for the genomic region containing the *XIST* exons was derived from analysis of genomic DNA and genomic clones from the region. The enzymes shown: E, EcoRI; H, HindIII; B, BglII; Ba, BamHI; X, XbaI. The lower genomic clone is a cosmid that extends further 3′.

(C) A summary of the alternative splicing events that have been observed, showing splicing events observed in cDNA clones with thicker lines, while the thinner lines indicate events observed only by RT-PCR analysis. The exons are numbered from the 5′ end. A(n) denotes polyadenylation sites. Exon 4 was observed in 9 of 11 clones (82%) that extended through the region, while exon 6i was present in 4 of 8 clones (50%) within that region. An alternative splice within exon 6, which results in the utilization of exons 7 and/or 8 (Brown et al., 1991b), was observed in clones from both heart and brain cDNA libraries.

(D) Sequence of the splice junctions is listed along with the size of the eight exons. nd, the intron sequence adjacent to exon 6 has not been determined. Number sign, the size of exon 6 is given; if the 6i splice is used, the exon is 56 bp shorter. Carat, the intron given here is transcribed if exon 5 splices to the beginning of exon 6. Asterisk, the sequence of exons 7 and 8 was previously published as XISTc and XISTd, respectively (Brown et al., 1991b). Exon 8 has been extended by 5 bp in the 3′ direction (see Experimental Procedures). The previously published alternative splice sites that are now within exon 6 (see [C]) are not included; however, the splice between exons 7 and 8 is included.

## Results

### *XIST* Transcripts Are 17 kb in Length and Are Alternatively Spliced

Over 70 cDNA clones were isolated from female cDNA libraries using successive 5′ ends of previously identified cDNA clones (Brown et al., 1991b) as well as genomic DNA fragments. These clones were arranged into the contig shown in Figure 1A by sequence alignment, polymerase chain reaction (PCR) analysis, and restriction mapping (see Experimental Procedures), yielding a consensus cDNA of 17 kb. In addition, overlapping genomic phage and cosmid clones were isolated, yielding the restriction map shown in Figure 1B. There were no restriction site differences observed between the genomic phage map and the map from human male or female DNA or from human–rodent somatic cell hybrids retaining the active or inactive human X chromosome (data not shown).

The presence of multiple alternative splicing events was observed in both cDNA clones and by reverse transcription PCR (RT-PCR) analysis, extending the limited data presented previously (Brown et al., 1991b). Multiple cDNA clones were isolated that excluded exons 4 or 6i (Figure 1). These and additional alternative splices were also detected by RT-PCR analysis (see Figure 2; data not shown) and are summarized in Figure 1C. The precise location of exons in the cDNA was determined by the lack of continuity between genomic DNA and cDNAs as demonstrated by PCR and sequence analysis, as well as by the location of alternative splices in different cDNA clones. All splice junctions (Figure 1D) closely match the consensus splice junction sequences for vertebrates (Shapiro and Senapathy, 1987), corroborating our assignment of the 5′ to 3′ orientation of the *XIST* gene based on polyadenylation and strand-specific reverse transcription (data not shown). Two of the exons are unusually large, ~11 kb (exon 1) and ~4.5 kb (exon 6). The remaining exons range from 64 bp to 372 bp in length, within the reported size range for typical exons (Blake, 1983).

### *XIST* Is Expressed Exclusively from the Inactive X Chromosome

Expression of *XIST* from the active and inactive X chromosome was analyzed using RT-PCR analysis with oligonucleotide primer pairs designed along the length of the gene. For these analyses, cDNA from normal male and female lymphoblastoid cell lines was PCR amplified with primers for the X-linked *MIC2* gene (a control gene previously shown to escape X inactivation; Goodfellow et al., 1984) or with a series of primers for the *XIST* gene. The *MIC2* primers amplified product from both male and female cDNA, whereas *XIST* primers amplified product only from cDNA derived from female RNA (Figure 2). Similarly, human-specific *XIST* primer pairs amplified product only from inactive X-containing hybrids (data not shown). The inactive X-specific expression and alternative splicing are consistent with Northern blot analyses that show hybridization to a large, heterogeneous transcript only in RNAs from cells containing an inactive X chromosome (data not shown; Brown et al., 1991b). The expression of *XIST* RNA exclusively from the inactive X chromosome is further supported by fluorescence in situ hybridization experiments (see Figure 7).

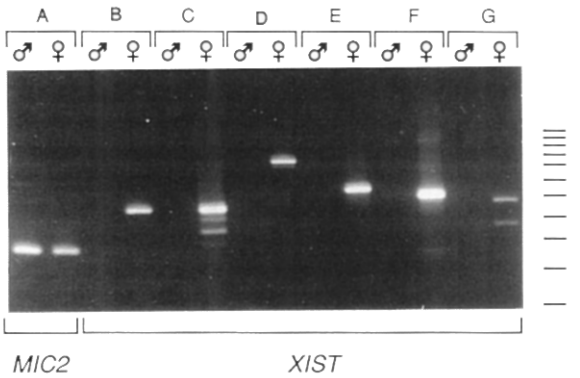



Figure 2. Inactive X-Specific Expression of *XIST* Demonstrated by RT-PCR Analysis

PCR amplification of cDNA derived from RNA from male and female lymphoblastoid cell lines with primers: A, MIC2; B, 6 + 7r; C, 8 + 11r; D, 11 + 13r; E, 1 + 3r; F, 14 + 15r; G, 3 + 18r (sequence given in Experimental Procedures). The size standard indicated to the right is a series of (n × 123) bp DNA markers. More than one band is observed in lanes C and G as a result of alternative splicing.

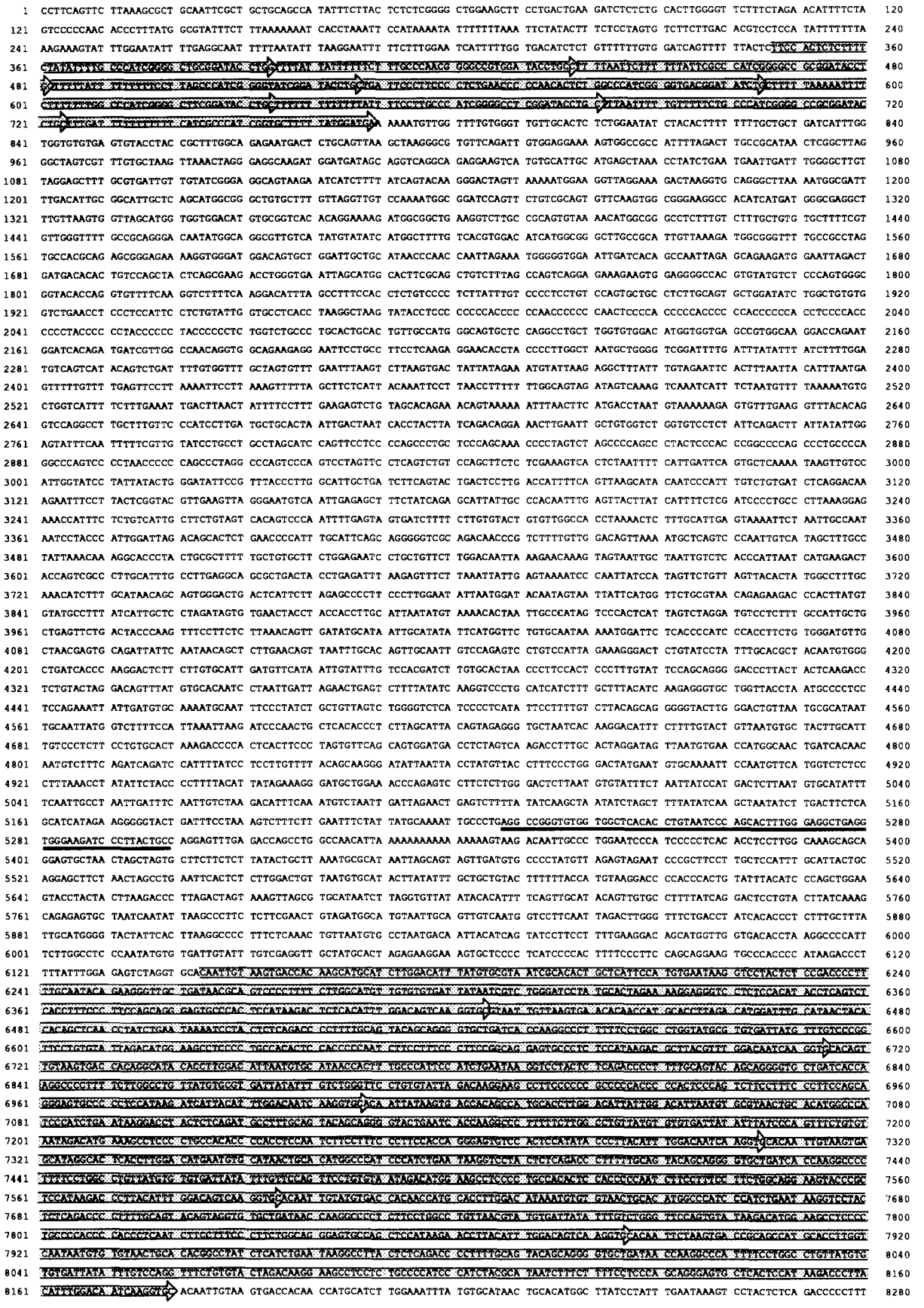
### Complete Sequence of the *XIST* cDNA

A consensus sequence of 16,481 bp was assembled by contig analysis of sequences from 67 of the *XIST* cDNA clones (Figure 3). Approximately 1.8 kb of this sequence has been previously reported (Brown et al., 1991b) along with the sequence of exons 7 and 8 (146 bp and 377 bp, respectively). Combining the sequence shown in Figure 3 with that of exons 7 and 8, the complete *XIST* sequence spans 17,004 bp. Comparisons of the complete *XIST* sequence with sequences in GenBank (V.71) revealed no significant sequence similarities, except for the region between position +5225 and +5300 that contains a portion of an Alu family repeat (Deininger, 1989) (Figure 3).

Three complementary approaches were used to determine the region of transcription initiation. First, the 5′ rapid amplification of cDNA ends (RACE) protocol (Frohman et al., 1988) was used with primers 29r and 20r (Figure 4B). A broad band, containing multiple *XIST*-hybridizing products of ~250–500 bp, was observed after amplification with primer 29r, suggesting heterogeneity in the site of initiation within the region between positions +1 and +250. Primer 20r amplified products of approximately 700 bp, indicating other transcription start sites slightly more 3′ than those identified with primer 29r. In both cases, RACE was only observed with female and not male RNA (Figure 4B).

Second, to localize the 5′-most transcription start, primers were created from sequences located in the region 500 bp upstream of primer 29r and analyzed for their ability to amplify cDNA in conjunction with a primer from within the cDNA sequence. As shown in Figure 4, primer 30 was able to amplify cDNA from inactive X-containing cell lines, while the more 5′ primer 31 was unable to amplify cDNA significantly. These data suggest that transcription initiates in

Cell
530



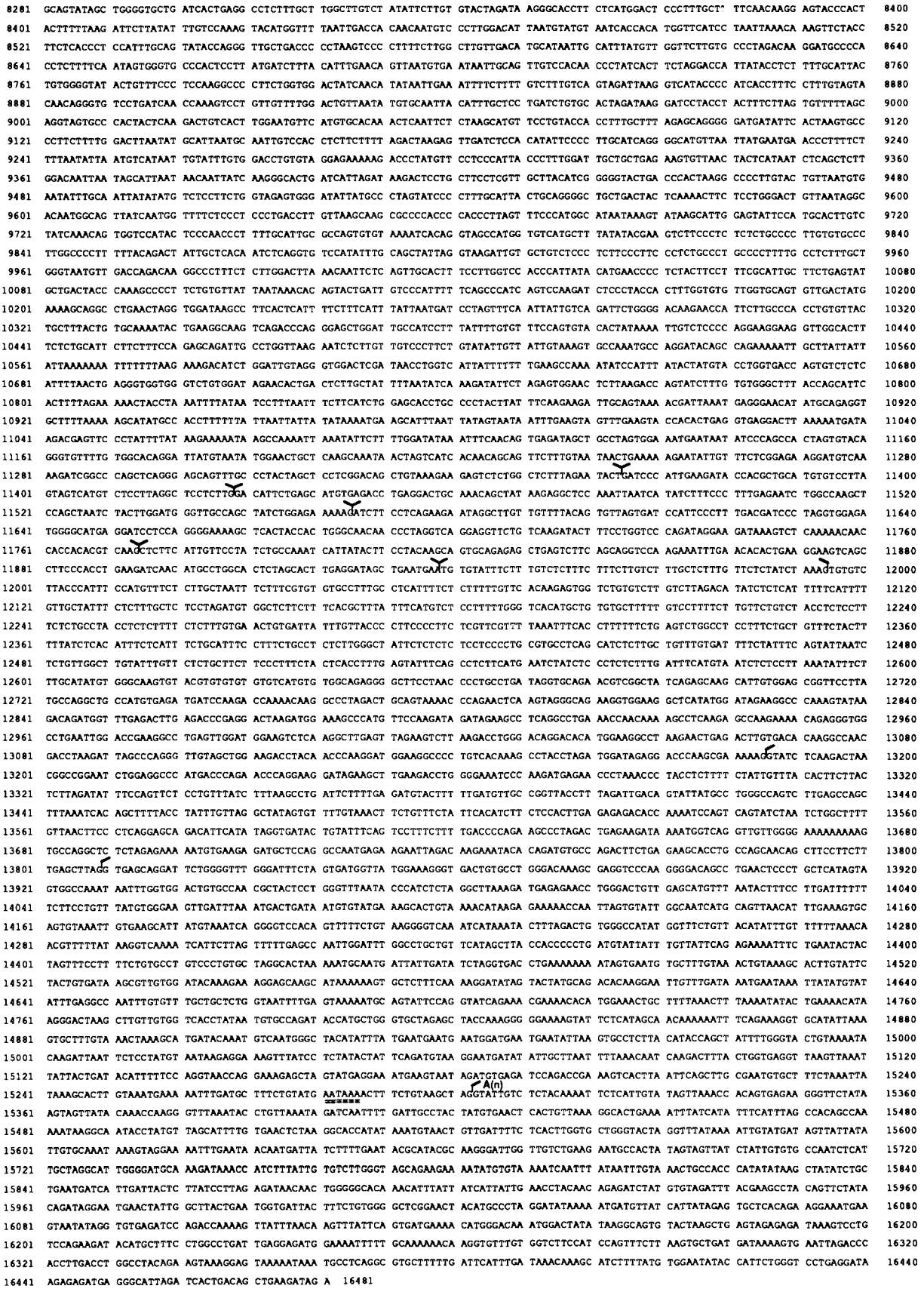
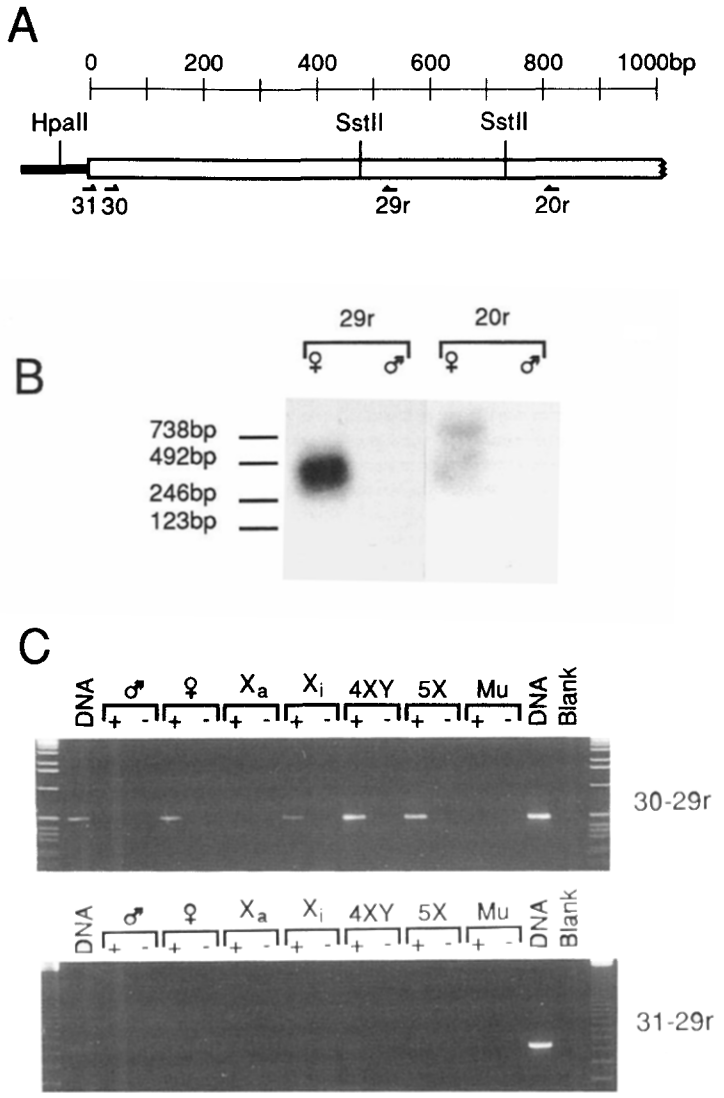


Figure 3. Nucleotide Sequence of the Human *XIST* Gene

The sequence of 16,481 bp of *XIST* cDNA is given. The two regions of tandem repeats discussed in the text are indicated by stippled arrows. Splice junctions between different exons (see Figure 1) are indicated by a branched tree, while sites that are only splice donors or acceptors are marked by the right or left branches of the tree, respectively. The previously described polyadenylation signal is underlined by a double line and followed by A(n) at the polyadenylation site. A region homologous to a portion of the human Alu repeat consensus is underlined by a single line.



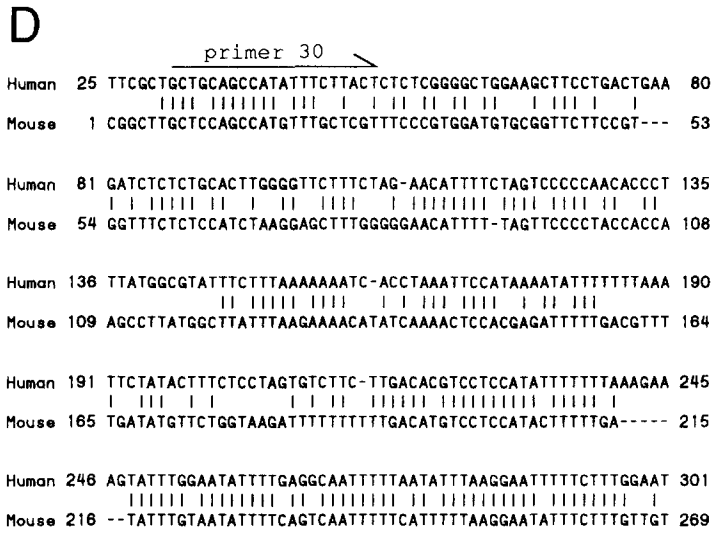




Figure 4. Analysis of the *XIST* Transcriptional Start Sites

(A) A schematic diagram of the 5′ end of the *XIST* gene showing the location of the primers used in (B) and (C), as well as methylation-sensitive restriction enzyme sites discussed in the text.

(B) Southern blot of a 5′ RACE analysis of RNA from male and female lymphoblasts using the 29r and 20r primers and hybridized with a probe from the 5′ end of the *XIST* gene. Size markers are shown to the left of the figure.

(C) Primers derived from genomic sequence (30 and 31) and cDNA sequence (20r and 29r) were used in the combinations shown at the right to amplify genomic DNA or total RNA with (plus) or without (minus) reverse transcription. The genomic DNA was derived from a normal female cell line, while RNAs were from normal male and female lymphoblasts, somatic cell hybrids retaining the active ($X_a$) or inactive ($X_i$) human X chromosome, a 49,XXXXY lymphoblast line (4XY), a 49,XXXXX lymphoblast line (5X), and the parental mouse line for the somatic cell hybrids (Mu). The blank control reactions (Blank) contained all PCR components except template to control for contamination. Since this region of the *XIST* gene does not contain any introns, cDNA products are colinear with genomic DNA, and RT-PCR reactions were compared with ones performed in the absence of reverse transcriptase as a control for possible DNA contamination of the initial RNA (lanes marked minus). *PGK1* primers were used as a control to demonstrate the presence of amplifiable cDNA in the male and active X-containing somatic cell hybrids (data not shown).

(D) The sequence of the 5′ end of human *XIST* is shown aligned with the sequence of the most 5′ mouse *Xist* cDNA clone that we have identified (see Experimental Procedures). Primer 30 is marked. Mouse/human base pair identities are indicated by a hash mark between the human (upper) and mouse (lower) sequences.

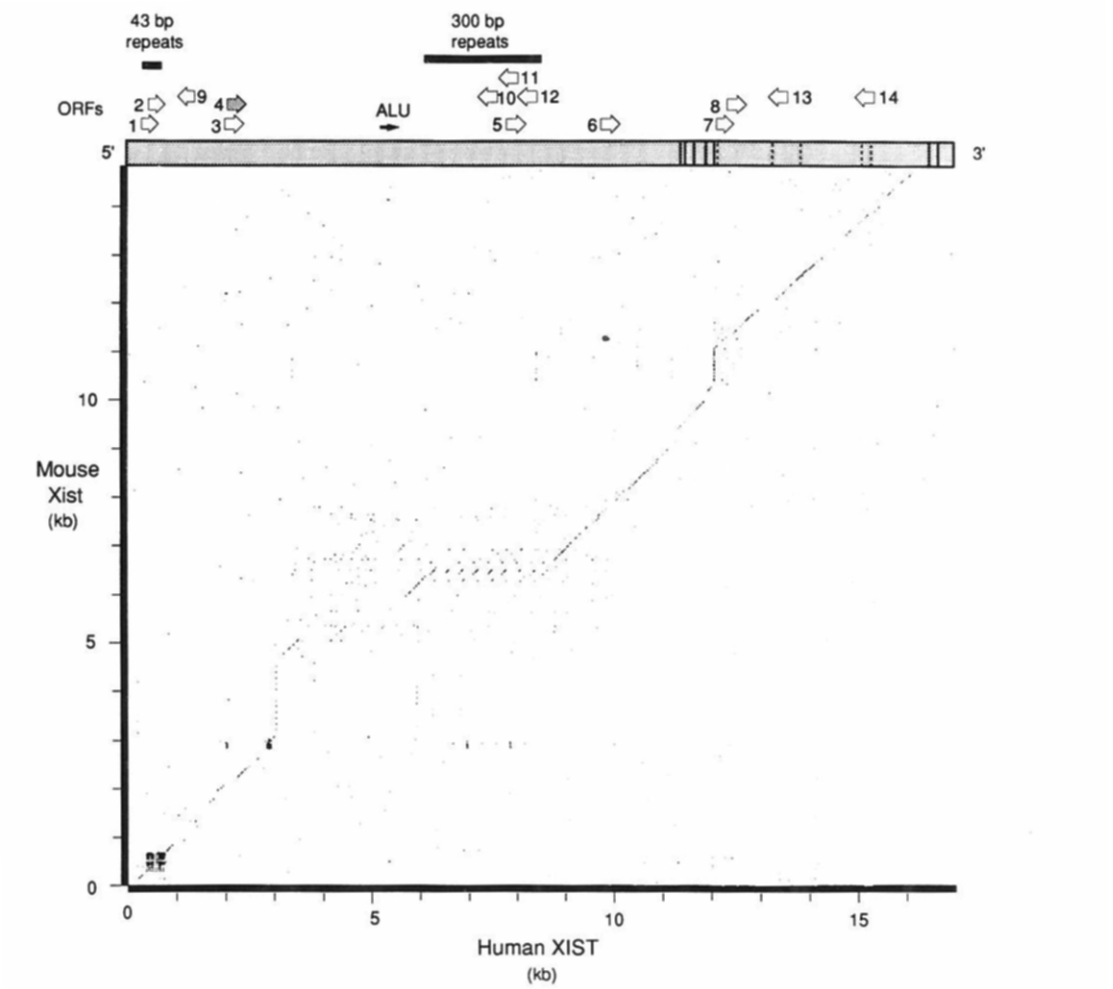



Figure 5. Dot Plot Comparison between the Human *XIST* and Mouse *Xist* cDNA Sequences

The human *XIST* sequence (horizontal) is compared with the mouse *Xist* sequence with 16 of 21 identical base pairs resulting in a dot (dot plot program; Devereux et al., 1984). The line diagram at the top of the comparison shows a number of the features of the human *XIST* gene described in the text. There are two tandem repeat regions marked as solid bars. The 14 longest ORFs of the gene are shown as numbered arrows. ORF4 is shaded to indicate its high positive GRAIL score that indicates likely protein-coding potential.

the region between the 5′ end of primer 31 (designated position +1 in the sequence contig; see Figure 3) and the 3′ end of primer 30 (position +52), although it does not rule out the possibility that a number of other (minor) start sites could exist further 5′ or 3′ of this region.

Third, RNAase protection experiments revealed a major transcriptional start at position +40 (B. D. H., unpublished data). This is consistent with the fact that a primer spanning position +40 and a primer located downstream of this site are able to amplify cDNA well (primers 30 and 33, respectively; Figure 4; data not shown), while a primer located upstream of +40 (primer 31; Figure 4) cannot amplify cDNA significantly.

Thus, in combination, the three approaches identify a series of transcription start sites at the 5′ end of the *XIST* cDNA sequence. The localization of transcriptional start site(s) between positions +1 and +40 also corresponds well with the 5′ end of a mouse *Xist* cDNA clone isolated by screening a mouse cDNA library with the 5′-most cDNA probe from the human gene (see Experimental Procedures). The alignment of the 5′ end of the human and mouse sequences is shown in Figure 4D. Despite local sequence conservation between the two genes of up to ~80%, overall sequence similarity is limited (~60%). Within the region of transcription initiation identified above (see also Brockdorff et al., 1992 [this issue of *Cell*]), there was no obvious strong interspecies conservation (Figure 4D).

### Homology with the Mouse *Xist* Gene

The complete cDNA sequences of the human (see Figure 3; Brown et al., 1991b) and mouse (Brockdorff et al., 1992; Borsani et al., 1991) genes were compared to identify conserved regions or features. A dot plot comparison of the two sequences is presented in Figure 5. The stringency for the comparison shown was 76% sequence identity in sliding 21 bp windows. As shown in Figure 4 and by Brockdorff et al. (1992), the two genes initiate transcription at approximately the same point and show detectable homol-

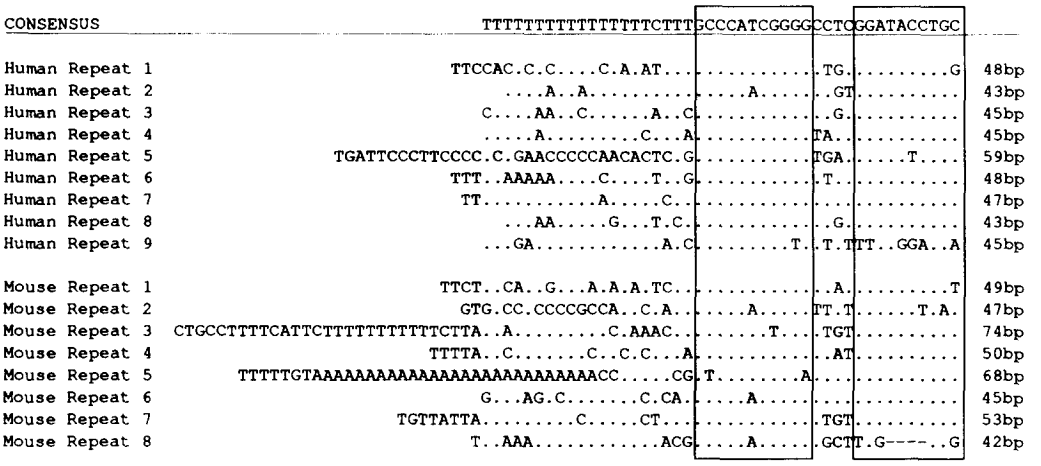



Figure 6. Alignment of Tandem Repeats within the *XIST* Gene

Human and mouse 5′ tandem repeats are aligned, with the first human repeat beginning at base 347 of the *XIST* cDNA and the first mouse repeat beginning at base 311 of the mouse *Xist* cDNA (see Figure 5). Repeats are listed 5′ to 3′ with no gaps between repeats. The consensus represents the most common base at that position. Boxes indicate two highly conserved 11 bp and 10 bp cores within the overall total repeat.

ogy over much of their length, although the human sequence extends significantly further 3′ than the mouse sequence (as human exons 7 and 8 have no detected counterparts in mouse). Within the gene, homology between human and mouse is interrupted multiple times both by blocks of sequence present in the human but not found in the mouse and by sequences present in the mouse but not in the human cDNAs, some of which are coincident with the junctions between exons. Although extended homology between the two genes is obvious, it should be emphasized that extensive gapping was required to maintain alignment of the two sequences. This finding suggests that divergence of the *XIST* and *Xist* sequences has not been constrained by a requirement for rigid maintenance of colinearity.

Between positions +350 and +770 in the human sequence there are blocks of strong sequence identity between the two genes, corresponding to a series of nine 43–59 bp direct repeats (aligned in Figure 6). Individual copies of the repeat vary from 43 to 59 bp in human and from 42 to 74 bp in mouse. Most of the length heterogeneity is due to variation within a T-rich stretch that lies between copies of a more homogeneous 25 bp GC-rich core region, within which two highly conserved 10 bp and 11 bp boxes are apparent (Figure 6).

The human sequence between positions +6000 and +8300 contains a series of ~300 bp direct repeats (indicated in Figure 3) that are 70% homologous to a single, partial copy of the repeat in the mouse sequence (mouse positions 6300–6500; Brockdorff et al., 1992). The eight copies of this repeat (plus a portion of a ninth copy) are approximately 90% identical in sequence to each other. The two sequences show more consistent homology 3′ to these repeats, interrupted by short (150–200 bp) regions within the mouse sequence that have no corresponding sequence in humans. At human position +11,990 there is an insertion of 725 bp in the mouse sequence, coincident with the start of human exon 6. The RNA sequence in this region is very A poor and U rich. The genes show substantial similarity 3′ to this region, as previously noted (Borsani et al., 1991).

### Assessment of Potential Coding Sequences within the *XIST* Gene

The initial characterization of the *XIST* sequence suggested that the 3′-most 3 kb of sequence was untranslated (Brown et al., 1991b). However, the previously described mouse *Xist* cDNA contained an almost 900 bp ORF at the 5′ end of that sequence (Borsani et al., 1991). To analyze the coding potential of the complete *XIST* gene, the 14 longest ORFs (those longer than 300 bp) have been analyzed in detail (Table 1; see Figure 5). Although the transcriptional orientation of *XIST* has been clearly established (by analysis of polyadenylation, consensus splice junctions, and reverse transcription with strand-specific primers across the length of the gene), we analyzed ORFs on both strands, since it is possible that the *XIST* gene may overlap one or more other genes transcribed in the opposite direction. The longest ORF identified is only 483 bp long, about 3% of the length of the complete cDNA sequence. None of the ORFs extends through an entire *XIST* exon.

The complete *XIST* sequence was analyzed by the GRAIL computer program, which is trained to identify protein-coding regions in human DNA (Uberbacher and Mural, 1991). Only two regions registered scores over 0.5, identified by the program as possible coding exons. One of these is ORF4 (see Figure 5), which registered a score greater than 0.5 for seven consecutive 10 bp shifts of a 100 bp window, peaking with a score of 0.94 (excellent coding potential). The other GRAIL-positive region is a short (69 bp) ORF on the opposite strand to the *XIST* sequence; this ORF registered a marginal score and is not included in Table 1.

Each ORF was analyzed for the presence of a suitable translation initiation sequence (Kozak, 1987) and com-

Table 1. Longest ORFs for the Human *XIST* Gene

| ORF[a] | Location | Size (bp) | Frame | GRAIL[b] | Initiator Sequence[c] | | | Mouse/Human[d] DNA Identity (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | | Site | Sequence | Score | |
| 1 | 177–501 | 324 | 1 | – | None | | | 71 |
| 2 | 314–646 | 333 | 3 | – | None | | | 49 |
| 3 | 1,842–2,258 | 417 | 3 | – | None | | | 34 |
| 4 | 1,927–2,319 | 393 | 1 | + | 2098 | TGC**ATG**G | 0.85 | 45 |
| 5 | 7,635–8,042 | 408 | 2 | – | 7635 | A**ATG**T | NS | 48 |
| | | | | | 7926 | AA**ATG**T | 0.74 | |
| 6 | 9,567–9,998 | 432 | 2 | – | 9795 | TGT**ATG**C | 0.72 | 61 |
| 7 | 11,938–12,267 | 330 | 3 | – | 11,938 | A**ATG**T | NS | 57 |
| | | | | | 12,175 | TTT**ATG**T | 0.66 | |
| 8 | 12,147–12,548 | 402 | 2 | – | 12,147 | CTA**ATG**T | 0.72 | 45 |
| 9 | 956–590 | 366 | 6 | – | 948 | AGTT**ATG**C | 0.68 | 74 |
| 10 | 7,138–6,754 | 384 | 5 | – | 7081 | TGG**ATG**G | 0.74 | 48 |
| 11 | 7,572–7,185 | 387 | 5 | – | 7352 | AGCT**ATG**C | 0.68 | 46 |
| 12 | 7,951–7,468 | 483 | 6 | – | 7498 | TTC**ATG**T | 0.74 | 47 |
| 13 | 13,083–12,720 | 363 | 6 | – | 13,041 | TTC**ATG**T | 0.74 | 47 |
| 14 | 14,843–14,465 | 378 | 5 | – | None | | | 58 |

[a] The ORF column assigns a number to each of the 14 longest ORFs in the *XIST* gene, based on their location from the 5′ end of the gene, with those ORFs (1–8) extending 5′→3′ listed first. Listed next are those ORFs (9–14) that extend from the 3′ end of the gene.

[b] The GRAIL score is positive for the one ORF listed that was identified as being potentially protein coding by the GRAIL computer program (Uberbacher and Mural, 1991).

[c] The initiator sequence column lists the location in the gene of the nearest initiator to the start of the ORF. The ATG is shown in bold. The scores for these consensus start sites are derived from the Geneid program (Guigo et al., 1992) with higher scores indicating greater confidence in the site. NS means that the site does not register, in which case the next potential start is also listed.

[d] The homology of the sequence to the mouse sequence at the nucleic acid level was evaluated using the GAP program (Devereux et al., 1984); all ORFs except ORF8 included frame disruptive gaps to generate this homology.

pared with the mouse *Xist* sequence (Brockdorff et al., 1992). DNA sequence conservation within the ORFs is generally low, <50% for 10 of the 14 ORFs. This is also reflected in the dot plot of human/mouse sequences (see Figure 5) in that the ORFs are not located in the regions of maximal sequence conservation. Conceptual translation of the ORFs in those cases where there was an overlapping mouse ORF resulted in short predicted protein sequences with similarities that were usually lower than the nucleic acid similarities. While unidentified transcripts, either from developmental stages or tissues not analyzed, may encode a protein, the data are most consistent with the absence of any coding potential within the currently available 17 kb *XIST* cDNA sequence.

### Subcellular Localization of *XIST* Transcripts

The subcellular location of the *XIST* transcripts was analyzed to assess whether *XIST* was associated with the translational machinery in the cytoplasm. In contrast with actively translated X-linked genes such as *PGK1* and *RPS4X*, our RT-PCR analyses of cDNA prepared from nuclear and cytoplasmic fractions have shown *XIST* RNA localized predominantly to the nucleus (data not shown).

To attempt to localize *XIST* RNA within the nucleus, fluorescence in situ hybridization analysis was performed using a digoxygenin-labeled *XIST* probe. A 9 kb genomic probe was hybridized to diploid female human fibroblasts (WI-38) and aneuploid human fibroblasts (47,XXX and 49,XXXXX) under conditions that allow RNA but not DNA hybridization (see Experimental Procedures). Consistent with the subcellular fractionation experiments described above, hybridization signal clearly above background was predominantly nuclear, concentrating at a single subnuclear location (Figure 7). Strikingly, these nuclear hybridization signals were highly localized to a site corresponding precisely with the position of the heterochromatic Barr body (detected by DAPI staining) at the periphery of the nucleus (Figure 7). Since the Barr body specifically corresponds to the inactive X chromosome (Barr and Carr, 1962), these data confirm the specific association of *XIST* transcripts with inactive X chromosomes. In 46,XX female cells, the two homologous X chromosomes are usually widely separated within the nucleus (Lawrence et al., 1990; D. Wolff and H. F. W., unpublished data), and just one site of hybridization (coincident with the Barr body) was observed in over 97% of nuclei (Figures 7A, 7C, and 7E), in contrast with an autosomal control gene (fibronectin) that consistently showed two signals (Figure 7K).

In normal male cells (which have no inactive X chromosome), no *XIST* hybridization could be detected (data not shown). However, in aneuploid cells, in which the number of inactive X chromosomes is one less than the total number of X chromosomes (Grumbach et al., 1963), the number of labeled nuclear sites was always one less than the total number of X chromosomes. Thus, in situ hybridization showed *XIST* transcripts localized at two nuclear sites in a 47,XXX cell line (~91% of nuclei; Figures 7G and 7H) and at four sites in the 49,XXXXX cell line (63% of nuclei; Figures 7I and 7J). In both cell lines, these sites were consistently coincident with Barr bodies visualized by DAPI staining.

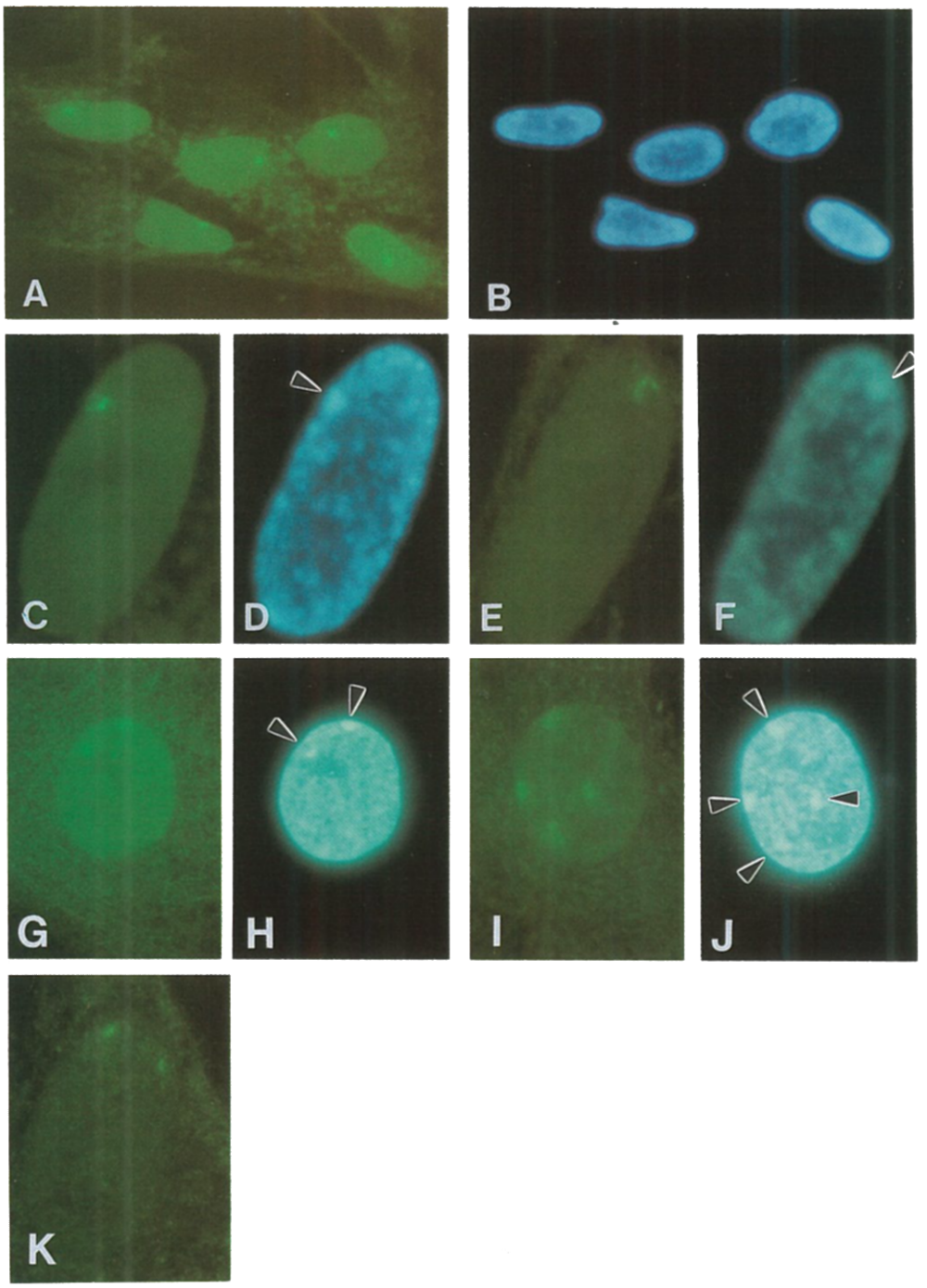



Figure 7. Fluorescence In Situ Hybridization of *XIST* RNA within Nuclei of Normal and Aneuploid Fibroblasts

Digoxigenin-labeled probes for *XIST* RNA were hybridized in situ to nondenatured cells, and specific hybridization was detected with fluorescein-conjugated anti-digoxigenin antibody (green). Nuclei were stained with DAPI (blue). (A) shows a low magnification view of several WI-38 cells showing the consistent detection of a single site of concentrated *XIST* RNA within each nucleus. Homogeneous fluorescence throughout the nucleus and cytoplasm is not appreciably above background levels observed in negative control samples. (B) shows DAPI staining of (A) ([A] and [B], 400×). (C) and (E) show higher magnification views ([C] and [D], 1200×; [E] and [F], 1080×) of *XIST* RNA signals within individual WI-38 nuclei, demonstrating that these colocalize with the condensed Barr body (arrows) evident by DAPI staining as shown in (D) and (F), respectively. Close examination of the *XIST* RNA signal in (C) shows one region of more intense fluorescence and, in addition, a dimmer fluorescence broadly distributed over the Barr body. (G) and (I) show *XIST* RNA distribution in the nucleus of 47,XXX cells and 49,XXXXX cells, respectively. The location of Barr bodies revealed by DAPI staining is apparent in (H) and (J). (K) shows detection of fibronectin RNA in the nucleus of WI-38 cells. Two signals are apparent, reflecting transcription from the two autosomal copies of the fibronectin gene ([G]–[K], 1080×).

## Discussion

Much of the currently available information about mammalian X inactivation reflects the well-established chromosomal nature of the phenomenon. The inactive X becomes heterochromatic, forming the Barr body in interphase nuclei (Barr and Carr, 1962; Therman et al., 1974; Walker et al., 1991), late replicating in S phase (Morishma et al., 1962; Willard and Latt, 1976), and methylated at the CpG islands associated with housekeeping genes that are subject to inactivation (for review see Grant and Chapman, 1988), in addition to expressing *XIST* (Brown et al., 1991b). Little is known about the actual mechanisms that result in the cis-limited inactivation of one of a pair of homologous chromosomes within the same nucleus. Thus, it is unclear which, if any, of the features mentioned above has a causal role in the inactivation process and which are secondary events resulting from the inactivation itself and/or required for maintenance of the inactive state (Riggs, 1990a; Brown and Willard, 1992). That *XIST*, both in humans and in mice, is expressed only from the inactive X chromosome and has been mapped genetically and physically to the smallest region defined for the localization of the *XIC* suggests that *XIST* is tightly associated with the events of X inactivation (Brown et al., 1991a, 1991b; Borsani et al., 1991; Brockdorff et al., 1991).

### Organization and Features of the *XIST* Gene

The 17 kb complete *XIST* cDNA is one of the longest cDNAs described (cf. Koenig et al., 1987; Wallace et al., 1990). Most of the length of the cDNA is derived from exons 1 and 6, which are 11.4 kb and 4.5 kb, respectively. The major transcriptional start site(s) has been localized to between positions +1 and +40 by several independent procedures (Figure 4), consistent with similar data obtained for the mouse gene (Brockdorff et al., 1992). Characterization of the promoter in both human and mouse will be important for understanding the events involved in X inactivation, since expression of *XIST* clearly distinguishes active and inactive X chromosomes. As posited by some models of *XIST* action in X inactivation (see below), interactions at the *XIST* promoter could be a critical determinant of whether an X becomes inactive or not. Preliminary DNA methylation analyses, using methylation-sensitive restriction enzymes, indicate that sites at the 5′ end of the gene are unmethylated on the inactive X but methylated on the active X in both human and mouse (B. D. H., unpublished data), consistent with the previously observed correlation between expression and methylation for X-linked housekeeping genes (for review see Grant and Chapman, 1988). However, a more thorough analysis of all potentially methylated sites, using sequence-based assays (Pfeifer et al., 1990; Frommer et al., 1992), will be necessary to determine how extensively the 5′ end of the *XIST* gene is hypomethylated on inactive X chromosomes.

The *XIST* RNA size heterogeneity detected by Northern analysis (Brown et al., 1991b; data not shown) can be at least partially explained by extensive alternative splicing in addition to heterogeneity of the 3′ and 5′ ends. Our analysis is necessarily limited to the length of products generated by reverse transcriptase (both for creation of cDNA libraries and for the RT-PCR analysis) so that the entire 17 kb RNA is never analyzed at one time. While most of the alternative splices have been seen in different libraries and in different tissues by RT-PCR, the possibility of tissue-specific or novel splicing patterns during development cannot be excluded.

The significance of alternative splicing in the *XIST* gene is unclear, especially since extensive alternative splicing has not been detected for the mouse gene (Brockdorff et al., 1992). It is possible that, other similarities between the human and mouse genes aside, the specific sequence of final *XIST* or *Xist* transcripts is not under rigorous selective pressure.

### Does *XIST* Encode a Protein?

To analyze the possible function of the *XIST* gene, 17 kb of *XIST* cDNA has been sequenced. Within this 17 kb of cDNA, the longest potential ORF is less than 500 bp. Analysis of the longest ORFs in the cDNA does not support a clear candidate for an *XIST* protein. The most convincing protein-coding region is ORF4, which spans 393 bp within the large first exon and scores well on computer analysis with the GRAIL program, which correctly predicts coding potential 94% of the time for other sequences with scores comparable to this ORF (Uberbacher and Mural, 1991). Further, ORF4 is associated with a strong Kozak initiation sequence (Kozak, 1987). However, this sequence is not well conserved in mouse (with both base changes and insertions/deletions), arguing against its coding potential and/or its critical involvement in X inactivation. While it is conceivable that the human and mouse genes have diverged widely, it seems inexplicable why a 17 kb messenger RNA (mRNA) would be required to encode a relatively short *XIST* peptide.

The data presented here and by Brockdorff et al. (1992) are reminiscent of studies on the *H19* gene, a genetically imprinted autosomal locus expressed, like *XIST*, from only one of the two copies of the gene in each cell (Bartolomei et al., 1991). Also like *XIST/Xist*, the human and murine *H19* genes do not share an ORF, suggesting that the *H19* product may function as an RNA (Brannan et al., 1990). The *H19* transcripts are also spliced and polyadenylated but appear to be transported out of the nucleus and become associated with a cytoplasmic particle (Brannan et al., 1990). In contrast, the bulk of *XIST* RNA appears to be located in the nucleus (Figure 7), providing additional evidence that *XIST* may not be translated. Nonetheless, it is possible that *XIST* is only translated at a specific developmental time period (e.g., at the blastocyst stage when X inactivation is believed to occur) or only in specific tissues.

Fluorescence in situ analysis indicated that the bulk of the *XIST* transcripts are confined to the nucleus, notably localized to the region of the Barr body (Figure 7). It has been observed for several other genes that the number of sites at which the RNA accumulates within the nucleus correlates with the number of actively transcribed genes (Lawrence et al., 1989; Y. X., C. V. Johnson, P. Dobner, and J. L., unpublished data; Coleman and Lawrence, 1991). The number of sites of *XIST* RNA accumulation

correlates with the number of inactive X chromosomes (0 for 46,XY cells; 1 for 46,XX cells; 2 for 47,XXX cells; and 4 for 49,XXXXX cells). It has been directly demonstrated that both the transcription and splicing of the RNA occurs within the highly concentrated RNA focus or track (Y. X., C. V. Johnson, P. Dobner, and J. L., unpublished data). Hence, the experiments described here do not formally distinguish between *XIST* RNA at the site of transcription or, possibly, *XIST* RNA at a site of deposition. However, since the *XIST* transcripts are relatively stable within the nucleus (at least in vitro), may not code for protein, and appear to localize over a broader region of the inactive X than the concentrated transcriptional focus (Figure 7), we favor the suggestion that *XIST* transcripts, after synthesis and processing, may become stably associated with the Barr body. However, further analysis will be required to confirm this and to clarify whether *XIST* RNA is associated with the entire inactive X chromosome or whether it is confined to a particular region on the X, perhaps the *XIST* locus itself (acting in a potentially autoregulatory fashion), and/or elsewhere within the critical *XIC* region. Future experiments will be required to examine the detailed distribution of *XIST* RNA and to compare it in the same cells with other RNAs produced from other transcriptionally active genes, especially including those X-linked genes that escape inactivation and are transcribed from both active and inactive X chromosomes.

### Implications of Repeated Regions within the *XIST* Gene

All X chromosomes in excess of one in a cell are inactivated, which suggests that a developmental event marks one X (presumably at the *XIC/Xic*) to remain active, all others becoming inactivated subsequently. This model predicts an absolute requirement for the existence of two copies of the *XIC/Xic* for X inactivation to occur, as documented experimentally in mouse (Rastan and Robertson, 1985). Suggestions for this marking event include interactions with a single nuclear attachment site (Comings, 1967) or an activator molecule produced in a limited fashion that binds in a highly cooperative manner to multiple binding sites (Gartler and Riggs, 1983; McBurney, 1988; Riggs, 1990a).

A repeat region at the *XIC/Xic* would be a good candidate for such a binding site, and thus it is particularly notable that the human and mouse *XIST* genes contain a number of internally repetitive regions. The 5′ 43–59 bp repeats are well conserved between species (Figure 6) and therefore may represent the best candidates for a role in X inactivation. Two 10 bp and 11 bp conserved core regions have been detected within these repeats and could represent binding sites for a DNA- or RNA-binding protein. Other repeated regions in *XIST/Xist* are very poorly conserved in number and/or sequence. It has been suggested that the behavior of different alleles at the *Xce* locus in mouse, which determine the probability that an X chromosome carrying that allele will be active or inactive (Cattanach et al., 1969), reflects differences in binding of an activator molecule (Riggs, 1990a). Since the expression of *Xist* appears to be inversely correlated with the strength of particular *Xce* alleles (Brockdorff et al., 1991), it will be very interesting to search for sequence differences between different *Xce* alleles in the *Xist* 5′ repeat region.

### Models for *XIST* Function in X Inactivation

While the human and mouse *XIST* genes are clearly homologous, gaps are consistently required to maintain the human–mouse alignment (Figure 5). Comparison of the human and mouse sequences demonstrates that there are regions that are well conserved spread throughout the gene. This observation, in addition to the overall similarity in organization of the two genes, suggests that the entire gene (with the possible exception of human exons 7 and 8) may be important for whatever function is served by *XIST*. In addition, the widespread conservation argues against such a function being to generate a very small protein encoded by only a limited region within the gene. Analysis of conserved regions in other species could direct our understanding of which regions of the *XIST* gene are potentially important for its function.

That *XIST* expression is completely concordant with X inactivation (Brown et al., 1991b; Figures 2 and 7) clearly suggests that *XIST* is either involved in, or directly influenced by, the process of X inactivation. This hypothesis is strengthened by the fact that *XIST* maps to the *XIC* critical region on the human X chromosome (Brown et al., 1991a) and by the observation that *XIST* transcripts are localized to the Barr body in interphase nuclei (Figure 7). It is important to emphasize that these data do not prove a role for *XIST*, causal or otherwise, in X inactivation. They do, however, help to establish *XIST* as a strong candidate for a role as at least one component of the *XIC*.

According to conventional models of gene structure and function, *XIST* may encode a protein involved in X inactivation. While none of the available data presented in this paper or by Brockdorff et al. (1992) supports such a model, it is possible that one of the small ORFs (possibly ORF4) does encode a protein. Such a protein, to explain the apparent importance of restricting expression to the inactive X chromosome, could be involved in feedback regulation of one or more genes located at the *XIC*, perhaps including the *XIST* gene itself.

Alternatively, *XIST* may function in X inactivation as a structural RNA molecule, consistent with the lack of a significant conserved ORF, with the lack of strict sequence conservation between human and mouse, and with the predominantly nuclear localization of *XIST* transcripts. Indeed, the cis-limited nature of X inactivation is much more readily explained by an RNA with limited capacity for diffusion than by a protein that would have to be translated in the cytoplasm and then transported back into the nucleus. Such an RNA could serve a number of plausible roles: to induce facultative heterochromatin formation as part of the cis-limited spreading of inactivation (by analogy with position-effect variegation) (Henikoff, 1990; Riggs, 1990b); to facilitate critical protein–DNA interactions on the inactive X (either early in development at the time of initiation of X inactivation or constitutively in somatic cells, perhaps involved in the maintenance of the inactive state through interaction with methylated DNA-binding proteins [Boyes

and Bird, 1991; Lewis et al., 1992]); or to sequester the inactive X within the nucleus, for example, as part of the Barr body.

However, the generation of *XIST* transcripts need not implicate *XIST* RNA directly in any function. For example, *XIST* expression could be the consequence of the transcriptional activity (or inactivity) of a second gene in the *XIC* region that is important for X inactivation. According to this model, expression of *XIST* would be a by-product of expression or repression of another gene, perhaps sharing a bidirectional promoter. If regulated by transcriptional interference (e.g., Corbin and Maniatis, 1989), then one might hypothesize that the upstream *XIC* gene would be expressed only from active X chromosomes. Alternatively, the two genes might be coordinately regulated (i.e., both limited in expression to the inactive X). This model is similar to the situation noted within the cis-regulatory region of the Drosophila bithorax complex, in which a number of heterogeneous, alternatively spliced transcripts (with no convincing coding potential) are expressed during development (Lipshitz et al., 1987; Cumberledge et al., 1990).

Finally, *XIST* expression may be either the cause or consequence of an altered chromatin conformation in the *XIC* region on the inactive X chromosome. Since in such a model it is the act of transcription rather than the product that is important, the sequence of *XIST* need not necessarily be under rigid evolutionary constraints (Ballabio and Willard, 1992).

Whatever the role of *XIST*, it is highly likely that a series of developmentally regulated and chromosomally promulgated cues, involving a number of interacting genes and their products, are required to carry out the events of X chromosome inactivation. *XIST* is an attractive candidate for one of the players involved, and its isolation and characterization should facilitate identification of the others.

## Experimental Procedures

### Clone Isolation

Screening of the λ library followed standard techniques (Maniatis et al., 1982). Phage were plated to a density of $10^4$ per 100 mm plate and transferred to nitrocellulose filters for hybridization to probes labeled by random hexamer priming (Feinberg and Vogelstein, 1983). Final wash stringency was 0.1% SDS and 0.1 × SSC at 65°C. Genomic λ clones were isolated from the American Type Culture Collection (ATCC) flow-sorted X chromosome library (LA0XNL01) by screening with previously isolated probes (Brown et al., 1991b). Similarly, a cosmid (ICRFc100h0130) was isolated from the Imperial Cancer Research Fund flow-sorted X chromosome reference library (Lehrach et al., 1990). cDNA clones were identified at a frequency of $1 \times 10^{-4}$ to $1 \times 10^{-5}$ from human heart (#936207, Stratagene, La Jolla, California; 17-year-old female heart) and fetal brain (#936206, Stratagene, La Jolla, California; 17–18 week gestation female fetal brain) cDNA libraries that were generated by oligo(dT) and random priming of RNA.

Mouse *Xist* cDNA clones were isolated from a female mouse (B6 × CBA)$F_1$ lung cDNA library generated by oligo(dT) and random priming (#936307, Stratagene, La Jolla, California). Two independent mouse cDNA clones were isolated by screening approximately $10^7$ primary plaques with the 5′-most human *XIST* cDNA probe (Hbc1a) at a final wash stringency of 0.5% SDS, 50 mM Tris–HCl (pH 8.6), and 0.5 M NaCl at 65°C. Fourteen additional overlapping *Xist* clones were obtained by screening the same plating of the library at a final wash stringency of 0.1% SDS and 0.1 × SSC at 65°C with the 5′-most mouse clone identified above. For all cDNAs, the cDNA phagemids were transformed into Escherichia coli to recover Bluescript plasmids containing cDNA inserts for further analysis.

### Contig Generation and Sequence Determination and Analysis

The cDNA clones were aligned by restriction mapping, PCR amplification with vector and gene-specific primers, and partial sequence analysis. cDNA clones from the fetal brain library frequently contained coligations of non-*XIST* cDNAs to *XIST* cDNAs. These events were detected when portions of a clone did not align with the cDNA contig by sequence comparison, restriction mapping, or PCR analysis or when portions of a clone did not map to the appropriate region of the X chromosome, Xq13, using the panel of hybrids described previously (Brown et al., 1991a). Primers were derived from the sequence of four of these coligated cDNAs. No amplification of female cDNA was observed with these primers (in combination with *XIST* primers), suggesting that the events had likely occurred during the creation of the library and were not biologically significant. Such coligation events are therefore not included in Figure 1. Hybridization of the cDNA clones to both human genomic DNA digests and the genomic clones allowed the identification of genomic fragments containing expressed regions of the *XIST* gene, which were then further defined by sequence comparisons between cDNA and genomic clones. Heterogeneity has been demonstrated for the 3′ end of the gene by 3′ RACE analysis (Frohman et al., 1988) as shown in Figure 1. One cDNA clone was isolated that contains a poly(A) tail at the end of exon 8, 5 bp further 3′ than the previously reported sequence and within 8 bp of the 3′ terminus of three other clones from three different libraries, confirming the major polyadenylation site described previously.

Multiple cDNA clones were identified and sequenced for most (>90%) of the *XIST* sequence. The first 360 bp of sequence was derived from genomic clones as cDNA clones were not isolated that extended as far 5′ as RT-PCR demonstrated transcription (Figure 4). There were three regions for which only a single cDNA clone was isolated. One of these is the CG-rich region (positions 1975–2069), which is 82% CG and does not reverse transcribe, PCR amplify, or sequence well; sequence was only obtained from a genomic clone covering this region. The other two regions have both been amplified in female cDNA with flanking primer pairs. One corresponds to positions 7042–7538, where it is difficult to identify the ends of coligation events because amplification with gene-specific primers is hindered by the repetitive nature of the sequence. Sequence of this region was derived from nested deletions of a genomic clone (colinear with cDNA), generated using an exonuclease III–mung bean nuclease deletion kit (Stratagene, La Jolla, California). The last region that is present in only one cDNA clone is a 130 bp stretch of cDNA (positions 9580–9710) that spans two clusters of clones.

Nucleotide sequence of the clones was determined on double-stranded templates using vector and gene-specific primers either manually (Korneluk et al., 1985) or automatically using an Applied Biosystems fluorescent sequencer (ABI model 373A with V1.1.1 Sequence analysis software) at the Stanford University Medical Center Protein and Nucleic Acid Facility. Over 85% of the complete *XIST* cDNA sequence has been determined in both directions; where this was impractical or impossible, the sequence was determined in multiple clones and/or multiple times in a single direction.

Contig assembly was accomplished with the DNASTAR software package. The GCG series of programs was used for data base searching, folding analyses, comparative dot plot analyses, and compositional analyses (Devereux et al., 1984). Similarities were calculated with the GAP program using standard parameters of gap = 5 and gap length weight = 0.3.

### Cell Lines and DNA Analysis

Karyotypically normal or abnormal human male and female lymphoblast or fibroblast cell lines were obtained from the Camden Cell Repository (Camden, New Jersey) and were grown at 37°C in α minimal essential medium supplemented with 15% fetal bovine serum. Mouse–human somatic cell hybrids retaining the active X chromosome as their only human chromosome (t60-12 and AHA–11aB1) or retaining the inactive X chromosome without the active X chromosome (t11–4Aaz5 and t48-1a–1Daz4a) have been previously described (Brown and Willard, 1989; Brown et al., 1991b). The use of the parental mouse cell line tsA1S9az31b (for all hybrids except AHA–11aB1) allowed for se-

Table 2. Primers for PCR Analyses

| Primer | Primer Sequence | Location |
|---|---|---|
| 1 | GAAGTCTCAAGGCTTGAGTTAGAAG | 12,991–13,015 |
| 2r | TTGGGTCCTCTATCCATCTAGGTAG | 13,151–13,175 |
| 3 | GCCAGGCTCTCTAGAGAAAAATGT | 13,681–13,704 |
| 3r | ACATTTTTCTCTAGAGAGCCTGGC | 13,681–13,704 |
| 4r | TGGCTCAAGTGTAGGTGGTT | 102–121 of c |
| 5r | TGTCTGCATAAAAGCAGATT | 92–111 of d |
| 6 | TCAGTTGCACTTTCCTTGGT | 10,008–10,027 |
| 7r | ACCCTACAATCCAGATGTC | 10,583–10,601 |
| 8 | AGCTCCTCGGACAGCTGTAA | 11,316–11,335 |
| 11 | TCCCACCTGAAGATCAACA | 11,882–11,900 |
| 11r | TGTTGATCTTCAGGTGGGA | 11,882–11,900 |
| 13r | TAGTCCTCGGGTCTCAAGTCT | 12,853–12,873 |
| 14 | GAAAGTGCAGTGTAAATTGTGAAGCA | 14,152–14,178 |
| 15r | CAGCATGGTATCTGGCACAT | 14,789–14,809 |
| 17r | GAGATTGGCACACAATAGA | 15,699–15,717 |
| 18r | TAGCAACCAACTCCCCAGTT | 325–344 of d |
| 20r | AGAGAGTGCAACAACCCACA | 783–802 |
| 29r | ATCAGCAGGTATCCGATACC | 509–528 |
| 30 | GCTGCAGCCATATTTCTTACT | 28–48 |
| 31 | CCTTCAGTTCTTAAAGCGCT | 1–20 |

Primers are listed 5′ to 3′; for those marked with r, the 3′ end of the primer is directed to the 5′ end of the *XIST* gene. Orientation inconsistencies in the previously published primers (1–5; Brown et al., 1991b) have now been corrected. The location in the *XIST* gene is given 5′ to 3′ on the gene, regardless of primer orientation.

lection for the X chromosome (either active or inactive) by growth at 39°C.

DNA was isolated from somatic cells grown in culture or from tissues by phenol extraction. Conditions for restriction enzyme digestion, gel electrophoresis, Southern blotting, prehybridization, and hybridization were as previously described (Willard et al., 1983).

### PCR Conditions

PCR amplification consisted of 30 cycles in a Ericomp thermocycler with 200 mM nucleotides, 1 mM primers, 50 mM KCl, 10 mM Tris–HCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 0.1% Triton X-100, and 2.5 U Promega Taq polymerase, with each cycle consisting of a 1 min denaturation at 94°C, a 1 min annealing at 54°C, and a 4 min elongation at 72°C. Products were analyzed after electrophoresis on a 2% agarose gel and staining with ethidium bromide. Conditions for 5′ and 3′ RACE PCR were as described (Frohman et al., 1988).

### Expression Analysis

For analysis of gene expression by the RT-PCR, 5 μg of RNA was reverse transcribed with 200 U of M-MLV reverse transcriptase as previously described (Brown et al., 1990). Approximately 50 ng of this reaction was used for PCR assays with the primers listed in Table 2. The previously described *MIC2* (Brown et al., 1990) and *PGK1* (Franco et al., 1991) primers that span splice junctions and therefore do not amplify DNA were used as controls for the presence of amplifiable cDNA. 3′ RACE PCR was performed as described (Frohman et al., 1988) using primer 3. 5′ RACE PCR was also performed as described (Frohman et al., 1988) using primers 20r and 29r. The products of RACE PCR analyses were always analyzed after Southern transfer and hybridization to ensure the products were from the correct region.

RNA was isolated from cells grown in tissue culture by guanidinium thiocyanate extraction (Chirgwin et al., 1979) or RNAzol (Biotecx) treatment. RNA was divided into poly(A)$^+$ and poly(A)$^-$ fractions based on binding to oligo(dT)–cellulose columns (Pharmacia). Nuclear and cytoplasmic RNA preparations were made as described for the isolation of mRNA from mammalian cells (Maniatis et al., 1982).

### Fluorescence In Situ Hybridization Analysis

A 9 kb genomic clone labeled with digoxigenin dUTP (Boehringer Mannheim) by nick translation was used for detection of RNA by in situ hybridization. Hybridization and detection procedures were essentially as described previously (Lawrence et al., 1989; for review see Johnson et al., 1991) and therefore will only be summarized in brief here. Monolayer cells grown on glass coverslips were permeabilized on ice with Triton X-100 in RNA-preserving CSK buffer (Fey et al., 1986; Carter et al., 1991) for 30 s prior to fixation in 4% paraformaldehyde for 5 min and storage in 70% ethanol. Previously conditions were defined whereby hybridization to RNA is promoted and DNA hybridization eliminated (Lawrence et al., 1989). Hybridization was to nondenatured cells (so that cellular DNA was not accessible for hybridization) overnight at 37°C in 50% formamide, 2× SSC, using a probe concentration of 5 μg/ml. Slides were rinsed and hybridization detected using anti-digoxigenin antibody conjugated to fluorescein (Boehringer Mannheim). The Barr body was identified by staining nuclei with 0.1 μg/ml DAPI for 30 s.

### Acknowledgments

We would like to thank members of the lab (in particular Laura Carrel) for their support, assistance, and numerous discussions. We also thank Drs. A. Ballabio, P. Avner, and S. Rastan for helpful discussions and Drs. S. Rastan, A. Ashworth, and N. Brockdorff for making available unpublished *Xist* sequence. This research was supported by research grants from the National Institutes of Health (GM 45441 and HG 00013 to H. F. W. and HG 00251 to J. L.) and by a predoctoral training grant from the National Institutes of Health (B. D. H.).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received June 16, 1992; revised August 25, 1992.

### References

Ballabio, A., and Willard, H. F. (1992). Mammalian X chromosome inactivation and the *XIST* gene. Curr. Opin. Genet. Dev. *2*, 439–447.

Barr, M. L., and Carr, D. H. (1962). Correlations between sex chromatin and sex chromosomes. Acta Cytol. *6*, 34–45.

Bartolomei, M. S., Zemel, S., and Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153–155.

Blake, C. (1983). Exons: present from the beginning? Nature *306*, 535–537.

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H. F., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*, 325–328.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123–1134.

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the *H19* gene may function as an RNA. Mol. Cell. Biol. *10*, 28–36.

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P. J., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. Nature *351*, 329–331.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (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*, this issue.

Brown, C. J., and Willard, H. F. (1989). Noninactivation of a selectable human X-linked gene that complements a murine temperature-sensitive cell cycle defect. Am. J. Hum. Genet. *45*, 592–598.

Brown, C. J., and Willard, H. F. (1992). Molecular and genetic studies of human X chromosome inactivation. Adv. Dev. Biol., in press.

Brown, C. J., Flenniken, A. M., Williams, B. R. G., and Willard, H. F. (1990). X chromosome inactivation of the human TIMP gene. Nucl. Acids Res. *18*, 4191–4195.

Brown, C. J., Lafrenière, R. G., Powers, V. E., Sebastio, G., Ballabio,

A., Pettigrew, A., Ledbetter, D. H., Levy, E., Craig, I. W., and Willard, H. F. (1991a). Localization of the X inactivation centre on the human X chromosome in Xq13. Nature *349*, 82–84.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafrenière, R. G., Grompe, M., Tonlorenzi, R., and Willard, H. F. (1991b). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Carter, K. C., Taneja, K. L., and Lawrence, J. B. (1991). Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. J. Cell Biol. *115*, 1191–1202.

Cattanach, B. M. (1975). Control of chromosome inactivation. Annu. Rev. Genet. *9*, 1–18.

Cattanach, B. M., Pollard, C. E., and Perez, J. N. (1969). Controlling elements in the mouse X-chromosome 1: interaction with the X-linked genes. Genet. Res. *14*, 223–235.

Chirgwin, J. M., Pryzybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Coleman, J. R., and Lawrence, J. B. (1991). Onset of dystrophin and myosin gene expression in muscle detected by fluorescence in situ hybridization to nuclear transcripts. J. Cell Biol. (Suppl.) *115*, 454a.

Comings, D. E. (1967). The rationale for an ordered arrangement of chromatin in the interphase nucleus. Am. J. Hum. Genet. *23*, 440–460.

Corbin, V., and Maniatis, T. (1989). Role of transcriptional interference in the *Drosophila melanogaster Adh* promoter switch. Nature *337*, 279–282.

Cumberledge, S., Zaratzian, A., and Sakonju, S. (1990). Characterization of two RNAs transcribed from the cis-regulatory region of the abd-A domain within the *Drosophila* bithorax complex. Proc. Natl. Acad. Sci. USA *87*, 3259–3263.

Daly, R. F., Patau, K., Therman, E., and Sarto, G. E. (1977). Structure and Barr body formation of an Xp+ chromosome with two inactivation centers. Am. J. Hum. Genet. *29*, 83–93.

Deininger, P. L. (1989). SINEs short interspersed repeated DNA elements in higher eucaryotes. In Mobile DNA, M. Howe and D. Berg, eds. (Washington, DC: American Society for Microbiology Press), pp. 619–636.

Devereux, J., Haeverli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. *12*, 387–395.

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

Fey, E. G., Krochamalnic, G., and Penman, S. (1986). The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell Biol. *102*, 1654–1665.

Fisher, E. M. C., Beer-Romero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R., and Page, D. C. (1990). Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and implications for Turner syndrome. Cell *63*, 1205–1218.

Franco, B., Guioli, S., Pragliola, A., Incerti, B., Bardoni, B., Tonlorenzi, R., Carrozzo, R., Maestrini, E., Pieretti, M., Taillon-Miller, P., Brown, C. J., Willard, H. F., Lawrence, C., Persico, M. G., Camerino, G., and Ballabio, A. (1991). A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. Nature *353*, 529–536.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA *85*, 8998–9002.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. USA *89*, 1827–1831.

Gartler, S. M., and Riggs, A. D. (1983). Mammalian X-chromosome inactivation. Annu. Rev. Genet. *17*, 155–190.

Goodfellow, P., Pym, B., Mohandas, T., and Shapiro, L. J. (1984). The cell surface antigen locus, MIC2X, escapes X-inactivation. Am. J. Hum. Genet. *36*, 777–782.

Grant, S., and Chapman, V. (1988). Mechanisms of X chromosome regulation. Annu. Rev. Genet. *22*, 199–233.

Grumbach, M. M., Morishma, A., and Taylor, H. (1963). Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization. Proc. Natl. Acad. Sci. USA *49*, 581–589.

Guigo, R., Knudsen, S., Drake, N., and Smith, T. (1992). Prediction of gene structure. J. Mol. Biol. *226*, 141–157.

Henikoff, S. (1990). Position-effect variegation after 60 years. Trends Genet. *6*, 422–426.

Johnston, C. V., Singer, R. H., and Lawrence, J. B. (1991). Fluorescent detection of nuclear RNA and DNA: implications for genome organization. Meth. Cell Biol. *25*, 73–99.

Johnston, P. G., and Cattanach, B. M. (1981). Controlling elements in the mouse. IV. Evidence of non-random X-inactivation. Genet. Res. *37*, 151–160.

Keer, J. T., Hamvas, R. M. J., Brockdorff, N., Page, D. C., Rastan, S., and Brown, S. D. M. (1990). Genetic mapping in the region of the mouse X-inactivation center. Genomics *7*, 566–572.

Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., and Kunkel, L. M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509–517.

Korneluk, R. G., Quan, F., and Gravel, R. A. (1985). Rapid and reliable dideoxy sequencing of double-stranded DNA. Gene *40*, 317–323.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15*, 8125–8148.

Lawrence, J. B., Singer, R. H., and Marselle, L. M. (1989). Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization. Cell *57*, 493–502.

Lawrence, J. B., Singer, R. H., and McNeil, J. A. (1990). Interphase and metaphase resolution of different distances within the human dystrophin gene. Science *249*, 928–932.

Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A. P., Nizetic, D., Zehetner, G., and Poutska, A. (1990). Hybridization fingerprinting in genome mapping and sequencing. In Genome Analysis, Vol. 1: Genetic and Physical Mapping, K. E. Davies and S. M. Tilghman, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 39–81.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Lipshitz, H. D., Peattie, D. A., and Hogness, D. S. (1987). Novel transcripts from the ultrabithorax domain of the bithorax complex. Genes Dev. *1*, 307–332.

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

Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. Am. J. Hum. Genet. *14*, 135–145.

Lyon, M. F. (1971). Possible mechanisms of X chromosome inactivation. Nature New Biol. *232*, 229–232.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mattei, M. G., Mattei, J. F., Vidal, I., and Giraud, F. (1981). Structural anomalies of the X chromosome and inactivation center. Hum. Genet. *56*, 401–408.

McBurney, M. W. (1988). X chromosome inactivation: a hypothesis. Bioessays *9*, 85–88.

Mohandas, T., Sparkes, R. S., Hellkuhl, B., Grzeschik, K. H., and Shapiro, L. J. (1977). Expression of an X-linked gene from an inactive human X chromosome in mouse–human hybrid cells: further evidence for the noninactivation of the steroid sulfatase locus in man. Proc. Natl. Acad. Sci. USA *77*, 6759–6763.

Morishima, A., Grumbach, M. M., and Taylor, J. H. (1962). Asynchro-

nous duplication of human chromosomes and the origin of sex chromatin. Proc. Natl. Acad. Sci. USA *48*, 756–763.

Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D., and Riggs, A. D. (1990). In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev. *4*, 1277–1287.

Rastan, S. (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos: location of the inactivation centre. J. Embryol. Exp. Morphol. *78*, 1–22.

Rastan, S., and Robertson, E. J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. J. Embryol. Exp. Morphol. *90*, 379–388.

Riggs, A. D. (1990a). Marsupials and mechanisms of X chromosome inactivation. Aust. J. Zool. *37*, 419–441.

Riggs, A. D. (1990b). DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. Phil. Trans. Roy. Soc. (Lond.) *326*, 285–297.

Russell, L. B. (1963). Mammalian X-chromosome action: inactivation limited in spread and in region of origin. Science *140*, 976–978.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R., and Page, D. C. (1989). *ZFX* has a gene structure similar to *ZFY*, the putative human sex determinant, and escapes X inactivation. Cell *57*, 1247–1258.

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

Therman, E., Sarto, G. E., and Patau, K. (1974). Center for Barr body condensation on the proximal part of the human Xq: a hypothesis. Chromosoma *44*, 361–366.

Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc. Natl. Acad. Sci. USA *88*, 11262–11265.

Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M., and Migeon, B. R. (1991). The Barr body is a looped X chromosome formed by telomere association. Proc. Natl. Acad. Sci. USA *88*, 6191–6195.

Wallace, M. R., Marchuk, D. A., Andersen, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H., and Collins, F. S. (1990). Type I neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. Science *249*, 181–186.

Willard, H. F., and Latt, S. A. (1976). Analysis of deoxyribonucleic acid replication in human X chromosomes by fluorescence microscopy. Am. J. Hum. Genet. *28*, 213–227.

Willard, H. F., Smith, K. D., and Sutherland, J. (1983). Isolation and characterization of a major tandem repeat family from the human X chromosome. Nucl. Acids Res. *11*, 2017–2033.

Yen, P. H., Ellison, J., Salido, E., Mohandas, T., and Shapiro, L. (1992). Isolation of a new gene from the distal short arm of the human X chromosome that escapes X-inactivation. Hum. Mol. Genet. *1*, 47–52.

**GenBank Accession Numbers**

The accession numbers for the sequences reported in this paper are M97168 for the human *XIST* gene and M97167 for the mouse *Xist* gene.