cells analysis[5].

The method also gives an uncertainty for the mean density of IRAS galaxies. The result for $r_{sm} = 20h^{-1}$ Mpc is 4.6%, and although this is an underestimate if there are significant correlations on scales greater than our smoothing scale, it suggests that this survey is approaching a fair sample of the Universe.

Conclusions

We have used the QDOT survey of IRAS galaxies to generate density-field maps of the local Universe. Many of the features seen were previously known, but many new superclusters and voids are seen in poorly surveyed parts of the sky. The root-mean-square density variation falls off with smoothing scale less rapidly than predicted by the standard cold dark matter theory of galaxy formation. We have detected skewness in the density field, even for smoothing radius $20h^{-1}$ Mpc, and the assumption of gaussian density fluctuations at the present epoch on these scales is no longer secure. □

Received 18 September; accepted 30 November 1990.

1. Kaiser, N. *Mon. Not. R. astr. Soc.* **219,** 785–790 (1990).
2. Rowan-Robinson, R., Saunders, W., Lawrence, A. & Leech, K. L. *Mon. Not. R. astr. Soc* (submitted).
3. Saunders, W. *et al. Mon. Not. R. astr. Soc.* **242,** 318–337 (1990).
4. Rowan-Robinson, M., *et al. Mon. Not. R. astr. Soc.* **247,** 1–18 (1990).
5. Efstathiou, G. *et al., Mon. Not. R. astr. Soc.* **247,** 10p–14p (1990).
6. Strauss, M. & Davis, M. in *Proc. 3rd IRAS Conf., Cornets to Cosmology* (ed. Lawrence, A.) 361–369 (Springer, Berlin, 1988).
7. Shapley, H. & Ames, A. *Ann. Harvard College Obs.* **88**(2) (1932).
8. Zwicky, F. *et al. Catalog. of Galaxies and Clusters of Galaxies* (California Institute of Technology, 1961–1968).
9. Shane, C. D. & Wirtanen, C. A. *Publ. Lick Obs.* **22,** 1 (1967).
10. Davis, M., Huchra, J., Latham, D. W. & Tonry, J. *Astrophys. J.* **253,** 423–445 (1982).
11. da Costa, L. N. *et al. Astrophys. J.* **327,** 544–560 (1989).
12. Geller, M. J. & Huchra, J. P. *Science* **246,** 897–903 (1990).
13. Loveday, J. D. thesis, Univ. Cambridge (1989).
14. Giovanelli, R. & Haynes, M. P. in *Large Scale Structures of the Universe, IAU Colloq. 130* (eds Audouze, J., Pelletan, M-C. & Szalay, A) 113–120 (Kluwer, 1988).
15. Tully, R. B. *Astrophys. J.* **303,** 25–38 (1986).
16. Tully, R. B. *Astrophys. J.* **323,** 1–18 (1987).
17. Bahcall, N. A. & Soneira, R. M. *Astrophys. J.* **277,** 27–37 (1984).
18. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
19. Hacking, P. B., Condon, J. J. & Houck, J. R. *Astrophys. J.* **316,** L15–L19 (1987).
20. Lonsdale, C. J., Hacking, P. B., Conrow, T. P. & Rowan-Robinson, M. *Astrophys. J.* **358,** 60–80 (1990).
21. Rowan-Robinson, M. *Space Sci. Rev.* **48,** 1–71 (1988).
22. Kaiser, N. *Mon. Not. R. astr. Soc.* **227,** 1–21 (1987).
23. Lahav, O., Rowan-Robinson, M. & Lynden-Bell, D. *Mon. Not. R. astr. Soc.* **234,** 677–701 (1988).
24. Kaiser, N. & Lahav, O. *Mon. Not. R. astr. Soc.* **237,** 129–162 (1989).
25. Cox, A. H. & Miller, R. J. *The Theory of Stochastic Processes* (1974).
26. Kendall, M. G. & Stuart, A. *The Advanced Theory of Statistics* Vol. 2, 4th edn (Griffin, 1979).
27. White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G. *Astrophys. J.* **313,** 505–516 (1987).
28. Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Astrophys. J.* **351,** 10–21 (1990).
29. Burstein, D. *et al.* in *Large Scale Structure and Motions in the Universe* (eds Mezzetti, M., Guiricin, G. T., Mardirossian, F. & Ramella, M.) (Kluwer, 1989).
30. Fairall, A. P. & Jones, A. *Publs Dept Astr. Cape Town* **10** (1988).
31. Scaramella, R., Baiesi-Pillastrini, G., Chincarini, G., Vettolani, G. & Zamorani, G. *Nature* **338,** 562–564 (1989).

ACKNOWLEDGEMENTS. We thank the staff at the AAT, and also those at the INT and WHT, operated by the Royal Greenwich Observatory at the Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canaries. We thank J. Huchra, M. Davis, M. Strauss, N. da Costa and R. Wolstencroft for supplying us with redshifts before publication. W.S. and A.L. acknowledge SERC postdoctoral and advanced fellowships, respectively.

A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome

Carolyn J. Brown, Andrea Ballabio*, James L. Rupert, Ronald G. Lafreniere, Markus Grompe*, Rossana Tonlorenzi* & Huntington F. Willard†

Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA
* Institute for Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

X-chromosome inactivation results in the *cis*-limited dosage compensation of genes on one of the pair of X chromosomes in mammalian females. Although most X-linked genes are believed to be subject to inactivation, several are known to be expressed from both active and inactive X chromosomes. Here we describe an X-linked gene with a novel expression pattern—transcripts are detected only from the inactive X chromosome (X_i) and not from the active X chromosome (X_a). This gene, called *XIST* (for X_i-specific transcripts), is a candidate for a gene either involved in or uniquely influenced by the process of X inactivation.

X-CHROMOSOME inactivation is a unique developmental regulatory mechanism, affecting the expression of genes on an entire chromosome in a *cis*-limited fashion—in the case of the human X chromosome, involving over 150 million base pairs (bp) of DNA and several thousand genes. X inactivation results in dosage equivalence between females who have two X chromosomes and males who have one X chromosome, by randomly inactivating one of the X chromosomes in females[1–3]. The mechanism of X inactivation remains unknown, although genetic evidence strongly suggests the existence of a specific locus required in *cis* for X inactivation to occur; this is the so-called X inactivation centre[4,5].

Many X-linked genes have been shown to be subject to X inactivation in both man and mouse on the basis of their mosaic or clonal expression[3,6–8]. Although only a few genes have been analysed for transcription[9–13], the basis for X inactivation is generally thought to be transcriptional. A growing number of genes have been described that escape X inactivation, both in or adjacent to the pairing segment with the Y chromosome[14,15], as well as more proximally on the short arm and on the long arm of the human X chromosome[9,10,16,17]. All of these genes are well expressed from both X_a and X_i chromosomes, although in some cases to a lesser extent from X_i.

To gain an insight into the molecular basis of X inactivation, we have searched for additional genes expressed from the X_i, particularly among those that map near the interval of the X inactivation centre (*XIC*) on the human X chromosome. We report here the isolation and characterization of a novel gene, *XIST*, expressed from X_i but not from X_a chromosomes, which by virtue of its localization on the X to the same interval as

† To whom correspondence should be addressed.

XIC, represents a candidate for a gene either involved in or strongly influenced by X inactivation.

Identification of X_i-specific transcripts

A 750-bp cDNA clone (8A11) was one of six clones previously obtained from a female placental complementary DNA library as part of a study to isolate cDNAs for human steroid sulphatase (ref. 18). This clone was used as a probe for the isolation of two related cDNA clones from the same library (see Fig. 1). On the basis of the localization of the gene encoding these cDNAs to the region of the X inactivation centre in Xq13 (ref. 19), we examined its expression by RNA slot blots, by northern blot analysis and by amplification of reverse-transcribed RNA using the polymerase chain reaction (RT-PCR). As shown in Fig. 1*a*, *XIST* cDNA probes hybridized to RNA prepared from female samples or from somatic cell hybrids containing an inactive human X chromosome, but not to RNA from males or from hybrids containing only an active human X chromosome. On

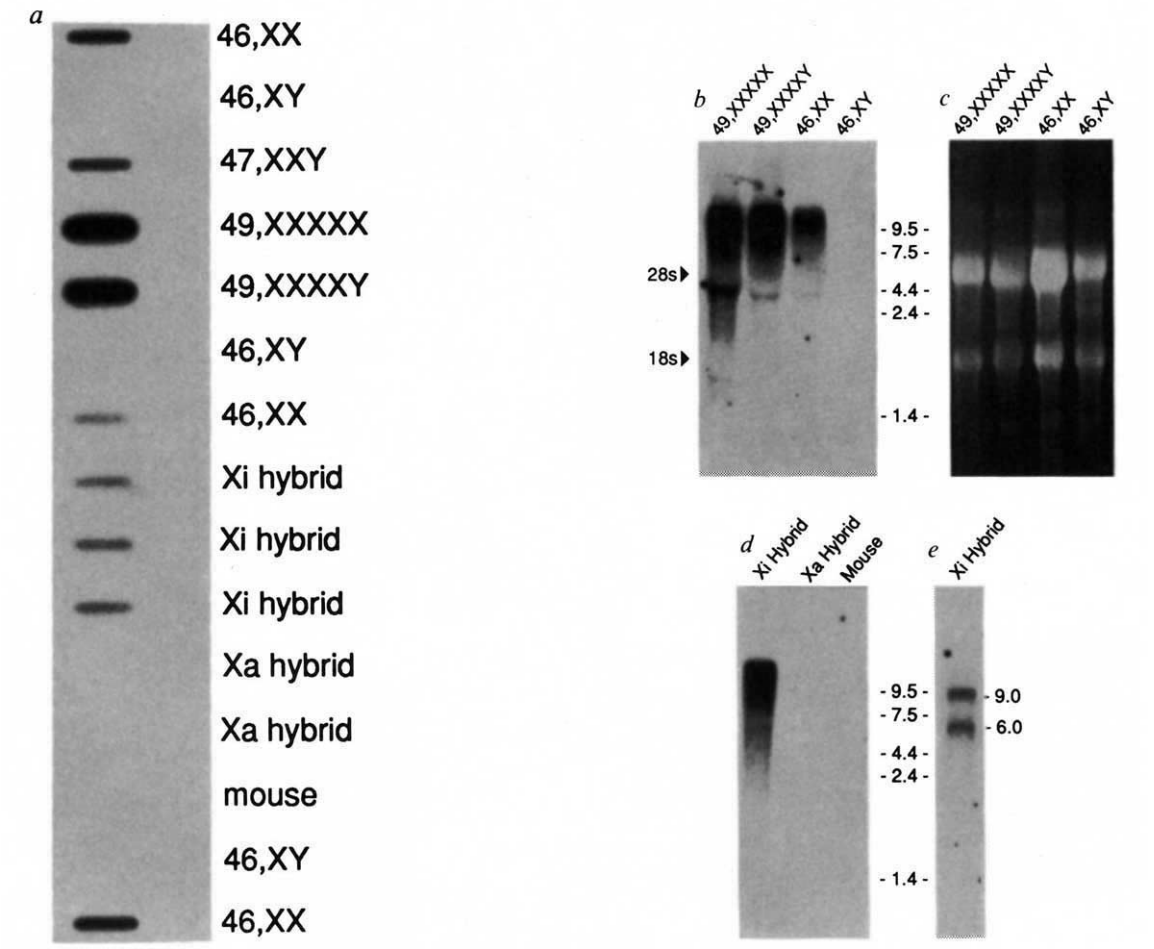


FIG. 1 Expression of the *XIST* gene in males, females, and somatic cell hybrids. *a*, Slot blot of total cellular RNA isolated from human lymphoblastoid cell lines or mouse–human somatic cell hybrids retaining either the active or inactive human X chromosome, hybridized with the 14A *XIST* cDNA probe. The probe hybridizes only to RNA samples from cell lines which contain an X_i. *b*, Northern blot of total cellular RNA from male and female cell lines, hybridized with the 14A cDNA probe. The probe detects a large, heterogeneous message only in cell lines which contain an X_i. Apparent bands below the position of 28S and 18S rRNA are created by the huge abundance of rRNA present. *c*, Ethidium bromide-stained RNA before northern transfer for the filter in *b*. *d*, Northern blot of poly(A)$^+$ RNA from human/mouse hybrids retaining either the X_a or the X_i and a mouse cell line control. The 14A *XIST* probe hybridizes to transcripts only in the hybrid with an inactive X chromosome. *e*, The X_i hybrid lane of *d* stripped and reprobed with a control cDNA probe (SB1.8) identifying a 9- and a 6-kb mRNA (ref. 17) to show that the mRNA is intact.

METHODS. Total cellular RNA was prepared using the guanidinium thiocyanate method[37]. The male (Y24.1, GM4391, GM7308) and female (Y24.2, HSC192, GM7011) lymphoblast cell lines (panel *a*, top to bottom in each case) were from karyotypically normal individuals. The X chromosome aneuploid lines were: 47, XXY, D64.0; 49, XXXXY, GM1202; 49, XXXXX, GM6061B. The X_a hybrids were t60-12 and AHA-11aB1, both of which contain the active X chromosome as the only human chromosome[38]. The X_i hybrids were t11-4Aaz5; t48-1a-1Daz4A and t86B1maz1B (ref. 38). 10 μg of total cellular RNA was denatured in formaldehyde/formamide and applied to a nitrocellulose membrane without vacuum for the slot blots. As controls, the RNA samples were extensively digested with RNase-free DNase I before applying to slot blots; the results were equivalent to those shown. To ensure equal RNA loading, the blot was stripped and reprobed with a cDNA for the human *PGK*1 gene; roughly equal signals were produced from each of the human RNA slots. For the northern blots, 20 μg total RNA or 5 μg poly(A)$^+$ RNA (purified by one or two passages through oligo(dT)) was electrophoresed on a 1.5% agarose gel in the presence of formaldehyde and transferred by capillarity to a nitrocellulose membrane. The blots were hybridized using standard methods[13] with ^{32}P-labelled 14A insert (*a*, *b*, *d*) or SB1.8 (ref. 17) (*e*), and washed at 42 °C with 0.1% SDS; 0.5×SSC. The 8A11 cDNA clone was one of six clones identified by immunoscreening of a placental λ gt11 cDNA expression library, using anti-steroid sulphatase antibodies[18]. The basis for the apparent immunoreactivity of this clone has not been investigated further; the clone has no significant DNA homology with the steroid sulphatase cDNAs reported[18]. The cDNA clones 14A and 7A were isolated by filter hybridization from the same library, using 8A11 as probe. The 14A cDNA clone had the longest insert (see Fig. 3) and was used as probe for all hybridizations.

northern blots, the same pattern of female (46, XX)- and X_i-specific hybridization was observed with a signal indicating a series of multiple transcripts showing extensive size and/or conformational heterogeneity, most >10 kb (Fig. 1*b*, *d*). Although in some experiments the hybridization pattern was suggestive of discrete bands, in most cases the signal was continuous, even when smaller cDNA fragments were used as probes (data not shown). Control experiments verified the integrity of the RNA samples tested (Fig. 1*e*), and because no hybridization was detected with the *XIST* cDNA probes in male RNA samples, hybridization artefacts could be discounted. In both slot blots and northern blots, the intensity of hybridization appeared to increase with the number of inactive X chromosomes present (Fig. 1*a*, *b*), although more extensive quantitation is required to confirm this. The northern blot data therefore suggest many heterogeneous *XIST* transcripts expressed from X_i but not from X_a chromosomes.

Because of the heterogeneity of transcripts detected with the cDNA probe, oligonucleotide primers were used for analysis by RT-PCR. PCR product was only observed in those reverse-transcribed RNA samples isolated from sources which contained an X_i. Karyotypically normal females were positive for the

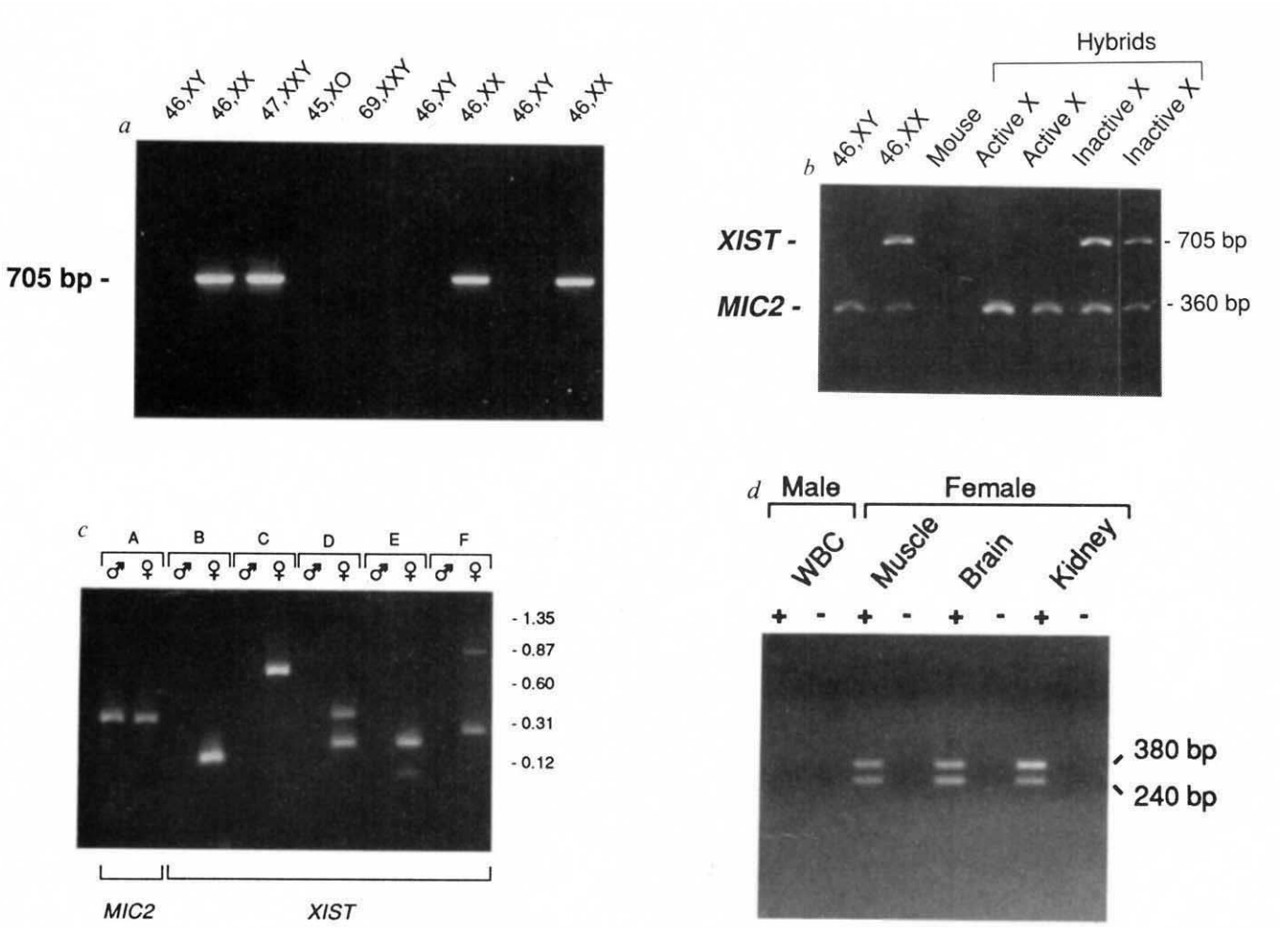


FIG. 2 RT-PCR analysis of *XIST* expression. *a*, RT-PCR of reverse-transcribed RNA (RT-RNA) from lymphoblastoid cell lines from karyotypically normal males and females, and individuals with X chromosome aneuploidies. The primers (sequence below, see also Fig. 3) amplified a 705-bp product only in RT-RNA from individuals with an X_i. *b*, RT-PCR of somatic cell hybrid lines retaining either X_a or X_i. As in *a*, the *XIST* primers amplified a 705-bp product only in those cell lines which contained an X_i. *MIC2* primers were included as a duplexed control for the presence of amplifiable cDNA, as *MIC2* is expressed from both the X_a and the X_i. The 360-bp human-specific *MIC2* product was observed in all hybrid lanes. *c*, RT-RNA from a male and female lymphoblast cell line was used in a PCR assay with five different *XIST* primer pairs (lanes B to F), and the *MIC2* primers as a control (lanes A). The primers are diagrammed in Fig. 3, and their sequences are given below. The 360-bp *MIC2* product was synthesized from both male and female RT-RNA, but the *XIST* products were observed only from the female RT-RNA. More than one product (only from the female RT-RNA) was observed for primer pairs D and F. Similar results are observed with hybrids containing the X_i (data not shown). The numbers to the left of the figure represent the size (in kb) of the *Hae*III-digested φX174 markers. *d*, RT-PCR analysis of *XIST* expression in samples from different tissues. RNA was prepared from male white blood cells (WBC) and from female brain, kidney, and muscle and was assayed both with (+) and without (−) reverse transcriptase.

METHODS. RNA (5 μg), isolated as described in Fig. 1, was reverse-transcribed with MMLV reverse transcriptase using random hexamer primers[13,39]. One-hundredth of the cDNA obtained was used for a PCR reaction with 1 μmol primers, for 30 cycles (94 °C, 1 min; 55 °C, 1 min; 72 °C, 4 min) with Promega or Cetus *Taq* polymerase and buffer. *d*, RNA was prepared from tissues obtained from an 18-week female fetus. Primers are listed 5′–3′ with the arrow designating orientation on the gene. *a*, Primers were: 1→: GAAGTCTCAAGGCTTGAGTTAGAAG; and 3←: GCCAGGCTCTAGAGAAAAATGT. *b*, *XIST* primers were as in *a* and the *MIC2* primers were as described previously[13]. PCR amplification of RNA samples which have not been reverse-transcribed does not yield the 705-bp *XIST* product or the 360-bp *MIC2* product (data not shown; see also *d*). *c*, Primer pairs were: A, *MIC2* primers, as described[13]; B, 1→(GAAGTCTCAAGGCTTGAGTTAGAAG) and 2←(AACCCAGGAGATAGGTAGATCCATC) to give a 180-bp product; C, as in *a* to give a 705-bp product; D, 3→(GCCAGGCTCTAGAGAAAAATGT) and 5←(ACAGACGTATTTTCGTCTAA) which give a 240- and a 380-bp product; E, 3→(GCCAGGCTCTAGAGAAAAATGT) and 4←(TGGCTCAAGTGTAGGTGGTT) which give a 260-bp product; F, 1→(GAAGTCTCAAGGCTTGAGTTAGAAG) and 5←(ACAGACGTATTTTCGTCTAA) which give a 300-, a 900- and a 1,040-bp product. Location of primers in the *XIST* gene are indicated in Fig. 3. *d*, Primer pair D was used.

TABLE 1 Summary of *XIST* expression

Cell-line	Karyotype	Expression RT-PCR	Slot blot	Northern
Chromosomally normal				
7 females	46, XX	+(7)	+(6)	+(6)
10 males	46, XY	−(10)	−(6)	−(6)
Chromosomally aberrant				
GM6061B	49, XXXXX	+	+	+
GM1202	49, XXXXY	+	+	+
106	45, XO	−	−	−
107	69, XXY	−	−	−
D64.0	47, XXY	+	+	+
GM10074 (ref. 19)	47, Y, t(X; 14) +der(14)	+	ND	ND
A.G. (ref. 19)	45, X/46, X, idic(Xp)	+	ND	ND
Somatic cell hybrids				
t60-12	Active X	−	−	−
AHA11aB1	Active X	−	−	−
t11-4Aaz5	Inactive X	+	+	+
t48-1a-1Daz4a	Inactive X	+	+	+
t81-az1D	Inactive X	+	+	+
t86-B1maz1b-3a	Inactive X	+	+	ND
LT23-IE2Buv5Cl26	Inactive X	+	ND	ND

The female RNAs analysed were from six lymphoblastoid cell lines and five different tissues from a single female fetus (see text and Fig. 2*d*). The male RNAs were isolated from lymphoblastoid cell lines (7 males), liver (2 males) and white blood cells (1 male). The '+' and '−' indicate whether expression was detectable for each assay (see Figs 1 and 2). 'ND' indicates that the assay was not performed for the given cell line. The numbers in parentheses after the '+' or '−' indicate the number of different female or male RNAs for which the specified assay was performed.

METHODS. RNA was isolated from the specified cells and analysed as described for Figs 1 and 2. All somatic cell hybrids were derived from female human cells and are described in Fig. 1, except for the X_i hybrids LT23-1E2Buv5Cl26, which was formed by fusion of mouse Ltk$^-$ cells with a 49, XXXXX human cell line, and t81-az1D, which was from a fusion of mouse tsA1S9 cells and the human cell line 81 (ref. 19). The four X_i hybrids t11-4Aaz5, t48-1a-1Daz4a, t81-az1D, and t86-B1maz1b-3a were under selection[16] for the X_i; these hybrids retain other human chromosomes, although no chromosome except the inactive X is common to all four lines. To eliminate any effect of other human chromosomes or any influence caused by selection, a fifth independent X_i-containing somatic cell hybrid was examined. This hybrid line, LT23-1E2Buv5, was not under selection for the X_i, and molecular cytogenetic analysis has identified that the only human material in this cell line is the X_i chromosome, present in about one-third of cells examined (V. E. Powers, data not shown).

705-bp product, whereas karyotypically normal males were negative (Fig. 2*a*). To determine if expression was sex-specific, in addition to X_i-specific, RNAs from a female without an inactive X (45, XO) and from males with one or more inactive X's (47, XXY; see also the 49, XXXXY sample in Fig. 1*a*) were examined. As shown in Fig. 2*a*, RT-PCR of RNA from the 45, XO female did not generate a product, but the RNA of the 47, XXY male did. RNA from a triploid cell line (69, XXY), in which both X chromosomes are active[20,21], was also examined. Despite the presence of two X chromosomes, no PCR product was detected. In total, samples from seven normal females, ten normal males, and seven chromosomally aberrant cell lines were analysed for *XIST* expression by PCR, northern blot and/or RNA slot blot (Table 1). In all cases, *XIST* expression was only observed in the presence of an X_i.

Additional evidence for X_i-limited expression came from the analysis of somatic-cell hybrids containing either X_i or X_a chromosomes (Fig. 2*b*). Hybrids with an X_a did not express *XIST*, whereas five independent hybrids with an X_i did (Table 1). As a control, all hybrids were shown to produce an RT-PCR product corresponding to the X-linked *MIC2* gene, consistent with previous demonstrations that *MIC2* escapes inactivation[13,15]. Taken together, the data in Figs 1 and 2 and Table 1 support the conclusion that expression of the *XIST* transcripts is absolutely dependent on the presence of an inactive X chromosome.

XIST gene structure and alternative transcripts

To examine expression of the *XIST* gene in more detail, several different oligonucleotide primers along the gene were synthesized for RT-PCR analysis. For each primer pair tested (see Fig. 3), a product was observed only in female or X_i-containing hybrid RNA samples, and never in male or X_a-containing hybrid RNA samples (Fig. 2*c*). A number of the *XIST* primer pairs gave more than one RT-PCR product, suggesting that alternatively spliced RNAs are present. Multiple transcripts were observed in all female tissues tested, including heart, muscle, brain, kidney, liver, fibroblasts and lymphoblasts (Fig. 2*d* and data not shown).

To identify portions of the *XIST* gene, two overlapping genomic clones were isolated which contained over 20 kb of genomic DNA corresponding to the *XIST* cDNAs (see Fig. 3). A genomic restriction map was derived from both Southern blot analysis of genomic DNA and mapping of the various clones. Both the intensity and pattern of hybridization on Southern blots was identical in normal males and in hybrids containing only a single X, at both normal and reduced stringency; thus all evidence is consistent with there being only a single copy of the *XIST* locus located on the X chromosome. Transcribed portions of the *XIST* gene were identified in cDNA clones isolated from two different cDNA libraries (see Fig. 3) and in various RT-PCR products isolated from both female cell lines and X_i-containing hybrid lines.

Several alternatively spliced RNA products have been identified and are diagrammed in Figs 3 and 4. The complete nucleotide sequence of the original cDNAs was determined, and all novel RT-PCR products not represented in the original cDNA clones (Fig. 3) were also sequenced. The sequences have been compared with the genomic *XIST* sequence across each putative splice junction (see Fig. 4). All putative donor and acceptor splice sites agree completely with consensus splice site sequences[22].

By using the RACE extension protocol[23] to identify 3′ poly(A) tails on RT-PCR products, a number of differentially spliced products were observed. The primary RACE product ended at the poly(A) site marked A1 on the map in Fig. 4. A second, less common product predicts the poly(A) site marked A2. The orientation of the RACE products, as well as the nature of the splice junctions, indicates the gene orientation shown in the figures. In total, at least five different splice products have been observed within this ~7-kb stretch of genomic DNA. This degree of alternative splicing, if representative of the entire gene, would explain the heterogeneous *XIST* RNA signal observed on northern blots (Fig. 1). Knowledge of the full extent of alternative splicing awaits isolation and analysis of the entire gene sequence, which the northern blot data suggest is likely to be at least 10 kb long.

Over 2.2 kb of transcribed *XIST* sequence has been determined (Fig. 5). Multiple stop codons were detected in all six reading frames and there were no sustained open reading frames >~300 bp in length for any of the predicted splice products observed to date. Furthermore, none of the sequences scored significantly in an algorithm designed to detect coding sequences[24]. It is likely, therefore, that the isolated portion of the *XIST* gene corresponds largely to untranslated material. Given the absence of an extended open reading frame, the fact that the original cDNA clone was identified by immunoscreening with anti-steroid sulphatase antibodies[18] remains unexplained. The entire sequence has been deposited in the GenBank/EMBL Data Library (accession no. X56196–X56199); no sequences in GenBank (August 1990 version), including steroid sulphatase

or any other X-linked gene for which sequence is available, show significant homology with any portion of the *XIST* sequence.

Discussion

The X inactivation centre (*XIC*) is a locus required in *cis* for an X chromosome to be inactivated[1–5]. The Xq13 region of the human X chromosome, which contains the *XIC*[25,26], is also the site of Barr-body condensation[27] and the site of a cytological bend on X_i chromosomes[28]. The region is homologous to a region on the mouse X chromosome, also required for X inactivation[29,30] and containing the *Xce* locus, alleles at which affect the choice of X chromosome to be inactivated[31,32]. The close association of these various inactivation-related

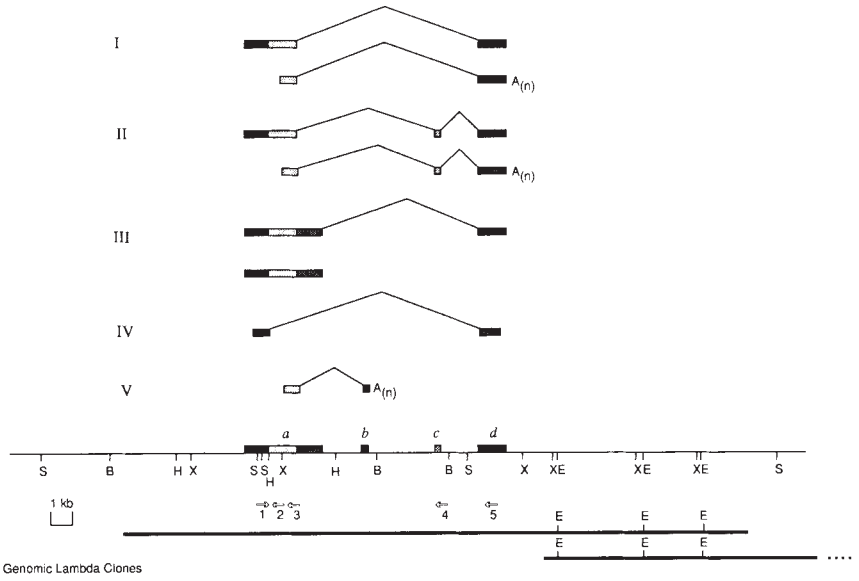
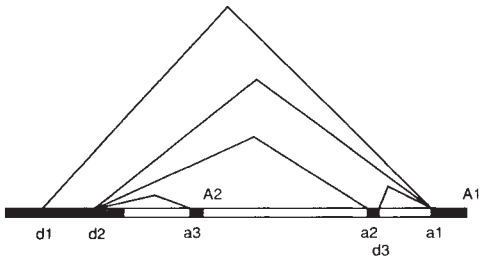


FIG. 3 Genomic and cDNA map of the *XIST* gene. The schematic drawings represent different transcribed products (I–V) from the *XIST* gene as recognized by analysis of cDNA clones or RT-PCR products. All products were X_i-specific. Product I is the 14A cDNA clone, and is consistent with the RT-PCR products from primers 1→ and 2← (primer pair B in Fig. 2*c*); 1→ and 3← (primer pair C); the lower product of 3→ and 5← (primer pair D); and the 900-bp product of 1→ and 5← (primer pair F). Product II is colinear with product I except for the addition of a 140-bp exon. This product has been identified as cDNA clone 7A (see Fig. 1) and is consistent with the RT-PCR products listed above, except that with primer pair D this product is the upper, 380-bp product, and with primer pair F it is the upper 1,040-bp product. Primer 4 is located within this additional 140-bp exon, and the observed 260-bp product from primers 3→ and 4← (primer pair E of Fig. 2*c*) is consistent with this product. Use of the RACE 3′ extension protocol with the 3→ primer yielded a number of products, of which the two major products are consistent with the splices of products I and II, with a 3′ extension including a poly(A) tail. These products are diagrammed below products I and II. Product III was isolated as a cDNA clone (pJR-c1) from a Stratagene human female fetal brain library (cat. no. 936206) by filter hybridization using 14A as probe. The 8A11 cDNA clone is identical to pJR-c1 up to the position of the splice, at which point 8A11 ends. Product III is also consistent with the RT-PCR products from primer pairs B and C. Product IV has only been observed as the RT-PCR product of 300 bp from the 1→ and 5← primers (primer pair F of Fig. 2*c*). Product V has been identified as a minor product of the 3′ RACE extension from primer 3→. Below these various splice products is a diagram of the *XIST* genomic DNA. This map is derived from Southern blotting of genomic DNA and is consistent with the map of the cDNAs described above, and the two genomic λ phage diagrammed at the bottom of the figure. The *a*, *b*, *c* and *d* above the transcribed portions of the *XIST* gene correspond to the sequenced segments (Fig. 5). The genomic phage clones were isolated by plaque hybridization with the 14A clone of the Los Alamos flow-sorted human X chromosome library LA0XN01 (obtained from the NIH Probe and Library Repository at the American Type Culture Collection, Rockville). The primer locations are represented by open arrows (not to scale). The arrow points in the direction in which the primer is 5′ to 3′. S, *Stu*I; B, *Bgl*II; H, *Hind*III; X, *Xta*I; E, *Eco*RI.



Splice donor	Splice acceptor	Product
d1 - AAG/gtatct	a1 - ctttcacag/TTT	IV
d2 - TAG/gtgagc	a1 - ctttcacag/TTT	I
d2 - TAG/gtgagc	a2 - ttttcctag/TCC	II
d3 - AAG/gtttgt	a1 - ctttcacag/TTT	II
d2 - TAG/gtgagc	a3 - ttttttccag/GTA	V
AG/gtaagt	$Py_{(6)}$ncag/N	Consensus

A1
genomic DNA CTTTCTGTAT GAATAAAACT TTCTGTAAGC TAGGTATTGT CTCTAC
RACE cDNA CTTTCTGTAT GAATAAAACT TTCTGTAAGC TAGAAAAAA AAAAAA

A2
genomic DNA TGCTATTGTA AAATAAAATA TACTGTTTTG AAAACTTTGT ATTTTGA
RACE cDNA TGCTATTGTA AAATAAAATA TACTGTTTTG AAAAAAAAA AAAAAA

FIG. 4 Alternatively spliced cDNA products, and sequence of the splice junctions and poly(A) addition sites. *a*, The different splice junctions diagrammed in Fig. 3 are summarized in this drawing of the *XIST* gene, and the splice junctions are designated as donors (d1–d3) or acceptors (a1–a3). The alternate polyadenylation sites as determined by RACE 3′ extensions are marked A1 and A2. A1 is a more abundant product than A2. *b*, The sequence of the different splice junctions. The '/' represents the splice site as determined by comparing the genomic and cDNA sequences. cDNA sequences are indicated in upper case. Genomic sequences not represented in cDNAs are indicated in lower case. All splice junctions agree with the consensus splice sequence[22] at the invariant bases (underlined), and generally agree at most of the favoured bases. The product designation refers to the products diagrammed in Fig. 3. *c*, Sequence of genomic DNA and cDNA before the polyadenylation sites A1 and A2. For both sites there is a potential polyadenylation signal (underlined) within 20 bp of the poly(A) tail.

phenomena establishes the primary importance of this region in X inactivation. As shown in the accompanying paper[19], the *XIST* gene is located in the portion of band Xq13 common to all structurally abnormal X_i chromosomes, and is therefore coincident with the *XIC* at the current level of mapping resolution.

The novel expression of *XIST*, its absolute dependence on X_i chromosomes and its mapping to the region of the *XIC* are strongly suggestive of some involvement in the X inactivation process. As the mechanism of X inactivation is not known, the role of the *XIC* and the potential involvement of *XIST* are necessarily conjectural. The involvement could be primary, in which case the *XIST* gene would be predicted to have a critical function in X inactivation, or secondary, in which case its expression might be determined as a result of other events occurring nearby which result in X inactivation; nonetheless, its X_i-specific and, among normal individuals, female-specific pattern of expression is novel among X-linked genes described to date.

X chromosome inactivation was first hypothesized by Lyon[3]. Much data have since accumulated supporting the theory that most X-linked genes are stably transcriptionally inactivated early in development; however, little is known either about the initiation of inactivation at the *XIC* early in embryogenesis or about the spreading of inactivation in *cis* along the length of the chromosome. All X chromosomes in excess of one are inactivated, suggesting that some signal prevents one X (the eventual X_a) from responding to a process which affects all unmarked X chromosomes (the eventual X_i's)[33]. It is reasonable to hypothesize that this signal acts on the *XIC*, as the *XIC* is the only region of the X which is required in *cis* for X inactivation to occur[19,25,26]. This signal would be responsible for blocking the marked X chromosome from responding to the developmental signal for inactivation[29,33,34].

If *XIST* expression has a primary role in X inactivation, then the hypothesized signal may interact directly with the *XIST* locus, thereby preventing its expression from active X chromosomes. Given the *cis*-limited nature of X inactivation[35], expressed products from the *XIST* gene on the X_i would be required to act in a similar *cis*-limited manner, if this gene is involved in X inactivation. One possibility may be that the *XIST* product is a *cis*-acting RNA molecule, perhaps involved structurally in formation of the heterochromatic Barr body. It is known, however, that at least some of the *XIST* RNA is polyadenylated (Fig. 1*d*), which would be more suggestive of a translatable mRNA. If *XIST* is a protein-encoding gene, its

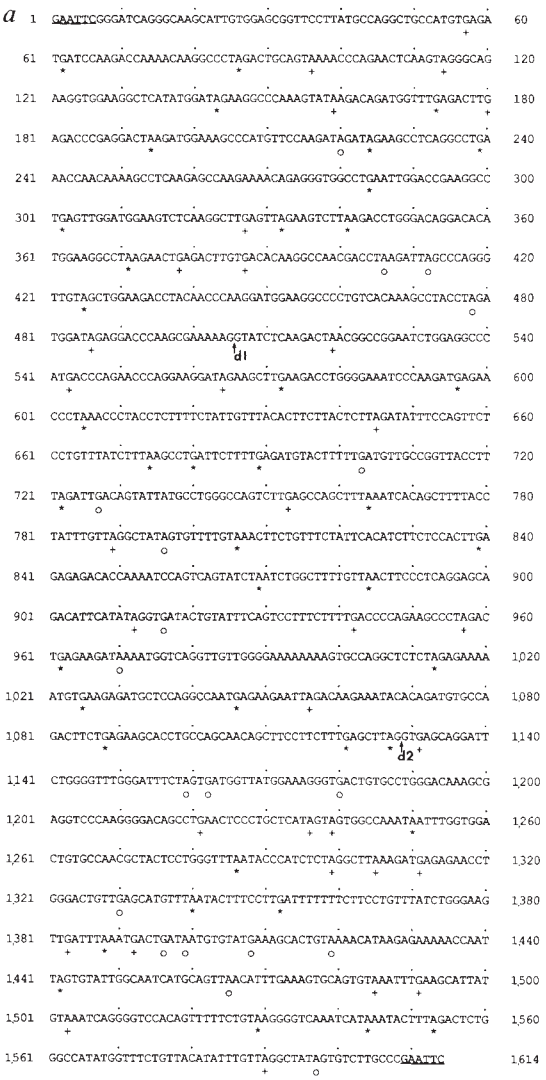


FIG. 5 Nucleotide sequence of transcribed portions of the *XIST* gene. Segments *a*, *b*, *c* and *d* correspond to the regions of the gene indicated in Fig. 3. In all sections, stop codons in the three forward reading frames are indicated by *, +, and 0. Each reading frame is interrupted many times. *a*, This 1,614-bp segment is colinear with genomic DNA and corresponds to sequences determined from the 14A and 8A11 cDNA clones and various RT-PCR products diagrammed in Fig. 3. The *Eco*RI restriction sites at the beginning and end (underlined) were introduced during construction of the placental cDNA library from which 14A and 8A11 were isolated[18]. They do not appear in genomic DNA at these positions. The arrows correspond to the d1 and d2 alternative splicing points as diagrammed in Fig. 4. *b*, This 160-bp segment corresponds to a transcribed region observed in splice product V (see Figs 3 and 4). Adjacent genomic sequence not observed in RT-PCR products is indicated in lower-case. *c*, This 146-bp segment corresponds to a transcribed region observed in splice product II (see Figs 3 and 4). Adjacent genomic sequence is indicated in lower-case. *d*, This 373-bp segment corresponds to the 3′ transcribed region observed adjacent to the major polyadenylation site in splice products I and II (see Figs 3 and 4). The adjacent genomic sequence is indicated in lower-case.

METHODS. All nucleotide sequences were determined manually on double-stranded templates using a modification[40] of the dideoxynucleotide termination method or automatically using an Applied Biosystems AB1370A fluorescent sequencer, as described by Gibbs *et al.*[41].

involvement in a *cis*-limited function is more difficult to explain. Its mRNA would have to leave the nucleus to be translated on cytoplasmic ribosomes. Upon return of the *XIST* protein to the nucleus, such a product would be required to distinguish between X_a and X_i chromosomes, including, possibly, distinguishing between different copies of the *XIC* and/or different copies of the *XIST* gene itself. Presumably, the protein-encoding capacity (or incapacity) of the *XIST* gene will become apparent once the complete gene is isolated and sequenced.

A number of models would be consistent with a primary role for *XIST* in X chromosome inactivation. First, the *XIST* product could be the molecule which acts in *cis* to cause X inactivation; second, *XIST* could be the *XIC* site which acts in conjunction with a different product to cause inactivation; or third, the *XIST* product could act upon the *XIC*, which would then produce or interact with the inactivating molecule(s). It is also possible that *XIST* is not the primary target of the developmental signal, but that an exogenous, developmentally regulated molecule interacts at the *XIC* to cause X inactivation, secondarily resulting in the expression of *XIST* from X_i chromosomes. In any case, the final result of X inactivation is that a *cis*-limited signal inactivates most genes on the X_i, although there are an increasing number of genes described on both the short and long arms of the X that escape X inactivation and thus do not receive a signal, do not respond to the signal, or cannot maintain their inactive state[9,10,14–17].

Examination of these and other possibilities concerning the *XIST* gene would be aided by experiments of two kinds. First, it will be critical to evaluate the expression of *XIST* in early development, at the time X inactivation occurs. Presumably, for *XIST* to have a role in X inactivation its expression must be modified in some critical way at the time of X inactivation. Second, it will be important to extend these observations to other species, most particularly to mouse, which offers a much more tractable system for developmental studies. Both genomic and cDNA *XIST* sequences are conserved at the level of DNA hybridization in a number of other mammalian species (including dog, cat, rabbit and cow and, at somewhat reduced stringency, mouse) (data not shown). If *XIST* expression is important in X inactivation, one would predict that homologous mouse sequences (*Xist*) should map to the homologous location on the mouse X chromosome[30,36], near the *Xce* locus, and should show the same pattern of X_i-restricted expression. The isolation of the homologous mouse gene would also allow for comparison of *Xist* expression and structure in strains carrying different alleles at the *Xce* locus.

Apart from its potential primary role in the X inactivation process itself, the *XIST* gene displays two novel properties, both of which are of considerable interest in their own right. First, expression is limited to inactive X chromosomes; even if *XIST* is not directly related to the *XIC* locus, it is thus strongly influenced by X inactivation and likely to provide an important molecular marker for X inactivation and an experimental foothold in the region. Second, expression of *XIST*, at least in the portion of the gene isolated to date, is normally female-specific. Identification of a potential gene product would be significant for examining developmental differences between the sexes and for analysing the potential effect of this gene's inappropriate expression in individuals with abnormalities in sex chromosome constitution, for example in Turner syndrome (45, X) and Klinefelter syndrome (47, XXY). □

Received 16 October; accepted 13 November 1990.

1. Lyon, M. F. *Biol. Rev.* **47,** 1–35 (1972).
2. Gartler, S. M. & Riggs, A. D. *A. Rev. Genet.* **17,** 155–190 (1983).
3. Lyon, M. F. *Nature* **190,** 372–373 (1961).
4. Russell, L. B. *Science* **140,** 976–978 (1963).
5. Cattanach, B. M. *A. Rev. Genet.* **9,** 1–18 (1975).
6. Migeon, B. R., der Kaloustian, V. M., Nyhan, W. L., Young, W. J. & Childs, B. *Science* **160,** 425–427 (1968).
7. Meyer, W. J., Migeon, B. R. & Migeon, C. J. *Proc. natn. Acad. Sci. U.S.A.* **72,** 1469–1472 (1975).
8. Davidson, R. G., Nitowsky, H. M. & Childs, B. *Proc. natn. Acad. Sci. U.S.A.* **50,** 481–484 (1963).
9. Schneider-Gadicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. *Cell* **57,** 1247–1258 (1989).
10. Fisher, E. *et al. Cell* **63,** 1205–1218 (1990).
11. Nadon, N., Dorn, N. & DeMars, R. *Somat. Cell molec. Genet.* **14,** 541–552 (1988).
12. Graves, J. A. M. & Gartler, S. M. *Somat. Cell. molec. Genet.* **12,** 275–280 (1986).
13. Brown, C. J., Flenniken, A., Williams, B. & Willard, H. F. *Nucleic Acids Res.* **18,** 4191–4195 (1990).
14. Shapiro, L. J., Mohandas, T., Weiss, R. & Romeo, G. *Science* **204,** 1224–1226 (1979).
15. Goodfellow, P., Pym, B., Mohandas, T. & Shapiro, L. J. *Am. J. hum. Genet.* **36,** 777–782 (1984).
16. Brown, C. J. & Willard, H. F. *Am. J. hum. Genet.* **46,** 273–279 (1990).
17. Brown, C. J., Barker, D. & Willard, H. F. *Am. J. hum. Genet.* **45,** A178 (1989).
18. Ballabio, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 4519–4523 (1987).
19. Brown, C. J. *et al. Nature* **349,** 82–84 (1991).
20. Migeon, B. R., Sprenkle, J. A. & Do, T. T. *Cell* **18,** 637–641 (1979).
21. Willard, H. F. & Breg, W. R. *Somat. Cell Genet.* **6,** 187–188 (1980).
22. Shapiro, M. B. & Senapathy, P. *Nucleic Acids Res.* **15,** 7155–7175 (1987).
23. Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85,** 8998–9002 (1988).
24. Fickett, J. W. *Nucleic Acids Res.* **10,** 5303–531 (1982).
25. Mattei, M. G., Mattei, J. F., Vidal, I. & Giraud, F. *Hum. Genet.* **56,** 401–408 (1981).
26. Summitt, R. L., Tipton, R. E., Wilroy, R. S. Jr., Martens, P. & Phelan, J. P. *Birth Defects: Orig. Art. Ser.* **14,** 219–247 (1978).
27. Therman, E., Sarto, G. E. & Patau, K. *Chromosoma* **44,** 361–366 (1974).
28. Fiejter, W. L., Van Dyke, D. L. & Weiss, L. *Am. J. hum. Genet.* **36,** 218–226 (1984).
29. Rastan, S. *J. Embryol. exp. Morph.* **78,** 1–22 (1983).
30. Rastan, S. & Robertson, E. J. *J. Embryol. exp. Morph.* **90,** 379–388 (1985).
31. Johnston, P. G. & Cattanach, B. M. *Genet. Res.* **37,** 151–160 (1981).
32. Rastan, S. *Genet. Res.* **40,** 139–147 (1982).
33. Brown, S. W. & Chandra, H. S. *Proc. natn. Acad. Sci. U.S.A.* **70,** 195–199 (1973).
34. Riggs, A. D. *Aust. J. Zool.* **37,** 419–441 (1990).
35. Lyon, M. F. *et al. J. Embryol. exp. Morph.* **97,** 75–85 (1986).
36. Keer, J. T. *et al. Genomics* **7,** 566–572 (1990).
37. Chirgwin, J. M. *et al. Biochemistry* **18,** 5294–5299 (1979).
38. Brown, C. J. & Willard, H. F. *Am. J. hum. Genet.* **45,** 592–598 (1989).
39. Grompe, M., Muzny, D. M. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **86,** 5888–5892 (1989).
40. Korneluk, R. G., Quan, F. & Gravel, R. A. *Gene* **40,** 317–323 (1985).
41. Gibbs, R. A., Nguyen, P.-N., McBride, L. J., Koepf, S. M. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **86,** 1919–1923 (1989).

ACKNOWLEDGEMENTS. We thank S. Cohen, G. Barsh, U. Francke, D. Botstein, D. H. Ledbetter, K. E. Davies and C. T. Caskey for discussions; F. Gonzales for the placental cDNA library; G. Persico for support; and R. Gibbs, D. Muzny and A. Civitello for assistance with DNA sequencing. M.G. is supported by the National Institutes of Health. This work was supported by the National Institutes of Health (H.F.W.).