Mammalian Genome 12, 538–545 (2001).
DOI: 10.1007/s00335001-0003-z

Mammalian Genome
Incorporating Mouse Genome

Expression and conservation of processed copies of the RBMX gene

Patricia A. Lingenfelter,[1] Margaret L. Delbridge,[2] Sushma Thomas,[1] Hopi E. Hoekstra,[3] Michael J. Mitchell,[4] Jennifer A. Marshall Graves,[2] Christine M. Disteche[1]

[1]Department of Pathology, BOX 357470, University of Washington, Seattle, Washington 98195-7470, USA
[2]Department of Genetics, La Trobe University, Melbourne, Victoria 3083, Australia
[3]Department of Zoology, University of Washington, Seattle, Washington, 98195, USA
[4]INSERM Unité de Génétique Médicale et Développement, Faculté de Médecine, Marseille 13385, France

Received: 19 January 2001 / Accepted: 14 March 2001

Abstract. RBMX and RBMY are members of an ancient pair of genes located on the sex chromosomes that encode RNA-binding proteins involved in splicing. These genes have differentiated and evolved separately on the X and Y Chromosomes. RBMY has acquired a testis-specific function, whereas, as shown here, RBMX is ubiquitously expressed and is subject to X inactivation. We have also found that multiple processed copies of RBMX are present in the human genome. RBMX-like sequences (RBMXLs) located on human Chrs 1, 4, 6, 9 (9p13 and 9p24), 11, 20, and X lack introns and thus probably result from retroposition events. We found RBMXLs to be conserved in primates and great apes at corresponding chromosomal locations, indicating that they arose prior to the divergence of human. Some of the RBMXLs show insertions, deletions, and stop codons, which would probably result in nonfunctional proteins. The RBMXL on Chr 20 is deleted in some individuals. Two of the largely intact RBMXLs, located on Chrs 1 and 9p13, are expressed in different tissues and may encode novel proteins involved in splicing in a tissue-specific manner. The RBMXL located at 9p13 is specifically expressed in testis, and to a lesser extent in brain, and may therefore play a role in testis function. This autosomal, testis-specific copy of RBMX could potentially compensate for RBMX that is presumably inactivated in male germ cells, in a manner analogous to autosomal retroposed copies of other X-linked genes.

Introduction

Mammalian sex chromosomes are thought to have evolved from an ancestral homologous pair of chromosomes (Ohno 1967). Once a gene on the Y Chr (SRY) determined male sex, recombination was suppressed between the proto-sex chromosomes, which resulted in progressive attrition of the Y (Charlesworth 1991; Graves 1995). Although many Y-linked genes were lost, some X/Y gene pairs persisted, including genes advantageous to males (Rice 1996). A few novel genes have been acquired on the Y by translocation or by retroposition from autosomal genes (Saxena et al. 1996; Lahn and Page 1999b). However, the major pathway for accumulation of male-advantage genes on the Y has been the acquisition of a male-specific function for the Y-partner of an ancestral pair of X/Y genes. Together with the differentiation or loss of Y-genes, X-partners of X/Y gene pairs usually became subject to X inactivation and upregulation (of the active X Chr), presumably to maintain balance of expression, both between males and females and with the autosomes (Graves et al. 1998).

We recently reported that the testis-specific RBMY gene (RNA-binding motif gene, Y chromosome, formerly RBM or YRRM), had a homolog, RBMX (formerly heterogeneous nuclear ribonucleoprotein G, hnRNPG or HNRPG), located on the human X Chr at band Xq26 (Delbridge et al. 1999). An X-linked homolog of the mouse *RBMY* gene was also reported (Mazeyrat et al. 1999). This implies that the RBMX/RBMY gene family, like other X-Y shared genes, originated from homologs on the sex chromosomes.

Although there is evidence that RBMY plays a role in spermatogenesis, the organization and the function of this gene family are not well understood. The human RBMY gene family contains at least 30 copies on the Y Chr, only a few of which are functional (Ma et al. 1993; Yen et al. 1995; Chai et al. 1998; Elliott et al. 1998). RBMY genes encode RNA-binding proteins with N-terminal RNA-binding motifs, and a C-terminal auxiliary domain containing between one and four repeated segments with a high proportion of serine, arginine, glycine, and tyrosine residues (SRGY boxes), a composition typical of many RNA splicing proteins (Fu 1995). Expression of human RBMY is confined to the developing germ cell lineage in adult testis, but the gene may also have a role in germ cell development (Elliott et al. 1997). RBMY sequences are candidate spermatogenesis genes, implicated in male fertility by the presence of deletions in the long arm of the Y Chr of patients with azoospermia (Ma et al. 1993; Pryor et al. 1997; Kostiner et al. 1998; Kent-First et al. 1999).

RBMX has an intron/exon structure similar to that of RBMY (Delbridge et al. 1999). Like RBMY, RBMX encodes an RNA-binding protein involved in splicing (Soulard et al. 1993; Delbridge et al. 1999; Venables et al. 2000). Our previous mapping of the RBMX gene by fluorescence in situ hybridization (FISH) revealed a series of homologous RBMX-like sequences (RBMXLs) dispersed in the human genome.

Here we report the map position, organization, expression, and evolution of human RBMX and RBMXLs. RBMX was found to be ubiquitously expressed and subject to X inactivation. All RBMXLs lacked introns, so probably resulted from retroposition. Three of the RBMXLs were expressed in different tissues. Furthermore, these sequences were remarkably conserved, and most were present in primates, suggesting that they arose prior to primate divergence and retained a specific function. One RBMXL, located on the short arm of Chr 9, was specifically expressed in testis and might therefore represent a splicing factor involved in testis function.

Materials and methods

Sequence analyses. PCR fragments were cloned into a pcrIITOPO vector (Invitrogen, Carlsbad, CA). Sequencing was done with M13 forward and reverse primers by using an ABI377 sequencer (Department of Biochemistry, University of Washington, Seattle, Wash.). Sequence analysis was done with GCG version 10.0 (Genetics Computer Group, University of Wisconsin, Madison, Wis.). Database searches were conducted in Gen-

Correspondence to: C.M. Disteche; *E-mail:* cdistech@u.washington.edu

Table 1. Primer sequences.

Primer name	Sequence	Product size in bp	RBMX sequence
hnRPGX2	ctgccctctcgtagagatgtttatttgtct	210[a], 410[b]	*RBMX*
hnRPGX3	cacgaccatatccatctctatcgctatatt		
hnRPGchr1f	aactcacgtagtgctccact	167	*RBMXL1*
hnRPGchr1r	gccaaccctgtcacaacttg		
hnRPG4f	gttatgcaccaccaccaccg	409	*RBMXL4*
hnRPG4r	ggaatcacgtggaggatttc		
hnRPG6f	gatctgcacctcctaagaga	268	*RBMXL6*
hnRPG6r	atgaactggaatgactgtaa		
hnRPGchr9f	ggtcattccagttcatgtgt	390	*RBMXL9*
hnRPGchr9r	cagcgcttggtgctccgcac		
hnRPGchr11f	gggaaccgctgccctctcac	342	*RBMXL11*
hnRPGchr11r	gcttccgccgtaagctggca		
RBMXL20f	tggaccgcctccacctccaagaagtagaggcccttt	263	*RBMXL20*
RBMXL20r	ccattccactgctactgcgaactggc		
hnRPGdXf	ctgccctcttgtagagagct	266	*RBMXLX*
hnRPGdXr	caccacctcttggtgttcca		
HSXF481	gtgggggtcctcctcctaag	~775	*RBMXL1, RBMXL4, RBMXL6, RBMXL11, RBMXLX*
HSXR1256	tttttccttttagtagtccttgggtag		
HSXR614[c]	gacaaataaacatctctacga	150[c]	*RBMX, RMBXL1, RBMXL6*

[a] Product size from cDNA.
[b] Product size from genomic DNA.
[c] Product size in combination with primer HSXR1256.

Bank (National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md, http://www.ncbi.nlm.nih.gov/entrez) and the Sanger Centre (Cambridge, UK, http://www.sanger.ac.uk).

Phylogenetic trees were constructed with the program PAUP 4.0 (Swofford 1997). The neighbor-joining method (Saitou and Nei 1987) with an HKY85 model of sequence evolution was the primary algorithm used to determine gene topologies.

Fluorescence in situ hybridization. Metaphase chromosome spreads were prepared from human peripheral blood samples by standard methods. Cynthia Friedman and Barbara Trask (University of Washington, Seattle, Wash.) kindly provided metaphase chromosome preparations from primate cell lines (from chimpanzee, gorilla, and orangutan). FISH was carried out as described by Edelhoff et al. (1994). A 2-kb genomic DNA clone that included exons 6-9 of RBMX was labeled with biotin by nick-translation.

Analysis of hybrid cell lines. Monochromosomal hybrid cell lines (NIGMS human/rodent somatic cell hybrid mapping panel # 2a) and a hybrid cell line that retained a deleted copy of Chr 9 (GM 12448) were obtained from the Coriell Institute (Camden, NJ). PCR analysis was done with the primers described in Table 1. PCR reactions were set up with 30 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C. Products were separated on 2% agarose gels.

X inactivation analysis. RT-PCR was done on RNA prepared from somatic cell hybrids retaining a copy of the active or inactive human X Chr. Hybrid cell line Y.162.11C retained a copy of the active X and a copy of Chrs 12 and 13 (Hansen et al. 1996). Hybrid cell line 8121-TGR-D retained a copy of an inactive X deleted for bands q27 and q28 and copies of Chrs 6, 7, 14, and 21 (Ledbetter et al. 1991). Hybrid cell line X8-6T2H1 retained a copy of the inactive X and copies of Chrs 4, 5, 6, 12, and 21 (Hansen et al. 1996). Hybrid cell line HY.70C4T3 (from M. Rocchi, Universita di Bari, Italy) retained a copy of the inactive X and a copy of Chr 21 (D'Esposito et al. 1996). Hybrid cell line THX88 (from A. Ballabio, TIGEM, Italy) retained a copy of the inactive X and other unidentified chromosomes (Ellison et al. 1993). These hybrid cell lines have been extensively used to determine the activity of X-linked genes (Hansen et al. 1996; Hornstra and Yang 1994; Agulnik et al. 1994; Esposito et al. 1997). After RT-PCR reactions set up as described above, products were separated on agarose gels.

Expression analyses. Multiple tissue cDNA panels I and II were obtained from Clontech. PCR amplification was done with 30 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C. Products were separated on agarose gels.

Results

Genomic organization and expression of RBMX. The RBMX coding region spans eight exons (Fig. 1). The start of the coding sequence is in exon 1, and the stop codon is in exon 8. We cannot exclude that there is an additional exon at the 5′ end, which would correspond to exon 1 of RBMY. The predicted amino acid sequence shows RNA-binding domains including two RNP domains and three RGG boxes (Fig. 2). In RBMX, only a single exon (exon 7) corresponds to the four repeated exons (exons 8–11) encoding the SRGY box in RBMY. As in RBMY, this region of the RBMX protein is rich in SR dipeptides. The complete genomic sequence of RBMX including the intronic sequences was deposited in GenBank (AF266720, AF266721, AF266722, AF266723). We noted that the RBMX sequence originally deposited in GenBank (NM002139) (Soulard et al. 1993) showed base differences at nt 624, 776, and 777 in comparison with the genomic sequence we obtained. One of these substitutions at nt 777 (A instead of G) coded for glutamic acid (E) in the original hnRPG sequence (GenBank NM002139), in contrast to the glycine (G) residue encoded by our sequence. All three base pair substitutions were present only in the original RBMX sequence (GenBank NM002139), but not in any of the ESTs (GenBank AW410104, AI751439, AA126135, AA081146), nor in the genomic sequence that we determined, suggesting that these substitutions represented rare polymorphisms or sequencing errors.

Expression of RBMX was examined by PCR by using multiple tissue cDNA panels I and II from Clontech and RBMX-specific primers hnRPGXf and hnRPGXr (Table 1). The specificity of these primers was verified with a panel of monochromosomal hybrid cell lines, as described below for the autosomal copies of RBMX. RBMX was expressed in all tissues examined including lung, brain, heart, skeletal muscle, liver, pancreas, kidney, small intestine, colon, spleen, thymus, leukocytes, prostate, placenta, ovary, and testis (Fig. 3A). The ubiquitous expression of RBMX contrasts with that of RBMY, which is confined to the testis (Elliott et al. 1997).

X inactivation status of RBMX. The X inactivation status of RBMX was determined by RT-PCR amplification of the gene in human × hamster hybrid cell lines retaining either an active or inactive human X Chr, and in control human and hamster RNA samples. These hybrid cell lines have been extensively character-

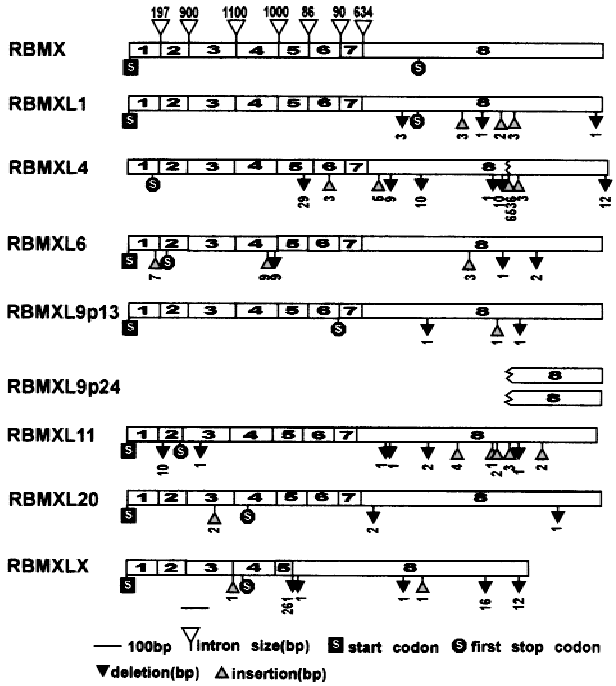

Fig. 1. Genomic organization of RBMX and RBMXLs. Exons are indicated as boxes. The size of exon 1 is 110 bp; exon 2, 109 bp; exon 3, 117 bp; exon 4, 156 bp; exon 5, 115 bp; exon 6, 121 bp; exon 7, 100 bp; and exon 8, 892 bp. Introns are indicated as triangles on top, with the size indicated in bp. Start codons, first stop codons, deletions and insertions are indicated below the sequences.

ized (Agulnik et al. 1994; Hansen et al. 1996). Primers hnRPGXf and hnRPGXr (Table 1) were selected on the basis of the sequencing data obtained for RBMX and RBMXLs (see below) in order to specifically amplify the functional X-linked gene. RT-PCR products of RBMX were obtained only in the hybrid cell line retaining an active human X Chr (Xa) and from control human RNA, whereas there was no product in the four cell lines retaining an inactive X Chr (Xi) or from the control hamster RNA (Fig. 3B). This was consistent with a lack of expression of the human RBMX gene from the inactive X Chr and confirmed that this gene is inactivated in the human.

Control RT-PCR amplification of HPRT (hypoxanthine phosphoribosyl transferase), a gene known to be subject to X inactivation in the human, resulted in a product only from the hybrid cell line retaining the active X and from control human RNA (Fig. 3B). Control RT-PCR amplification of MIC2, a gene that escapes X inactivation in the human, gave a product in all samples, except from the control hamster RNA (Fig. 3B).

Mapping of RBMX-like sequences on human chromosomes. We performed FISH, using a clone containing about 2 kb of genomic sequence including exons 5–8 of RBMX for in situ hybridization to human metaphase chromosomes. This resulted in signals not only on the X Chr at band Xq26, but also on several autosomes (Fig. 4A). The distribution of 130 signals in 105 metaphase cells in summarized in Fig. 4B. In addition to the expected signals at Xq26 where RBMX lies, signals were consistently seen on Chr 1p22-31, 4q25-27, 6p12, 9p13, 9p24, and 11q13, with a weaker signal occasionally seen at Xq13. These results suggest the presence of multiple autosomal loci, and of a potential second locus on the X Chr, in addition to RBMX.

We confirmed the location of RBMXLs using a panel of somatic cell hybrid lines (NIGMS human/rodent somatic cell hybrid mapping panel # 2a), each containing a single human chromosome on a rodent background. PCR products with homology to RBMX were amplified from the cell lines containing human Chr X, 1, 4, 6, 9, and 11, confirming the FISH results (Fig. 4C). The HSXF481/HSXR1256 primer pair (Table 1) resulted in amplification of sequences from Chr X, 1, 4, 6, 9, and 11. The PCR products amplified from the X were smaller (533 bp) than those amplified from the autosomes (about 775 bp), suggesting the presence of a truncated copy of RBMX on the X, in addition to the gene itself (Fig. 4C). With this latter primer pair, a 2.3-kb product (too large to detect under our PCR conditions) would be expected from the gene at Xq26, owing to the presence of intronic sequences between the primers. Additional primer pairs were used to confirm the presence of processed RBMXLs in the monochromosomal hybrid cell lines. However, not all primers derived from the RBMX gene sequence resulted in amplification of products from all RBMXLs, reflecting sequence differences between copies (see below). For example, the HSXF481/HSXR614 primer pair (Table 1) amplified sequences from Chr X, 1, and 6. The X-products were larger (1.1 kb) than those amplified from the autosomes (150 bp), consistent with the absence of introns in the RBMXLs. RBMXL copies on Chr 4, 9, and 11 and the truncated copy on the X were not amplified with these primers (data not shown).

To distinguish the RBMXLs on Chr 9, we used a hybrid cell line that retained a deleted Chr 9 on a hamster background (GM 12448). The deleted portion of Chr 9 in this cell line extended from band 9p22 to 9pter. Other human chromosomes retained in this cell line were Chr 5 and 22 and a portion of Chr 17, none of which contained RBMXLs. Using this hybrid line, we developed PCR primer pairs specific for each locus to confirm the presence of RBMXLs both at 9p13 and 9p24 (see below, Fig. 5A).

Characterization of the RBMX-like sequences. A combination of our own sequence analysis of PCR products obtained from the hybrid cell lines, and of database searches (GenBank and Sanger Centre) showed that all RBMXLs had between 92% and 96% nucleotide identity to the exons of RBMX. The highest nucleotide identity with RBMX was observed for the RBMXLs on Chr 1 and 9p13 (96%). The autosomal RBMXLs and the truncated X Chr copy all lacked introns, indicating that they represented retroposons of RBMX (Fig. 1). Interestingly, all retroposed copies showed a consistent but silent base pair substitution (G to C) at nt 777. The mouse and marsupial RBMX sequences also showed a C at this position (Delbridge et al. 1999; Mazeyrat et al. 1999).

The complete sequence of RBMXL1 was found within a PAC clone mapped to Chr 1p31 (GenBank AL139416), which confirmed our FISH results. Complete identity was found with the partial sequence (775 bp) we had obtained from the Chr 1 hybrid cell line. Examination of RBMXL1 sequence revealed 96% nucleotide identity with RBMX and an open reading frame encoding 390 aa with no major mutations within the coding region (Figs. 1 and 2), suggesting that the copy on Chr 1 could encode a functional protein.

RBMXL4 was found to be embedded within an 8398-bp genomic sequence mapped to 4q25 and deposited in GenBank (AC004051). Sequence comparisons revealed 94% homology to RBMX. Complete identity was observed between the partial sequence (775 bp) we obtained from the Chr 4 hybrid cell line and the GenBank sequence. A large insert of 6536 bp was found further downstream at nt 1358 (of the corresponding RBMX sequence), which was confirmed by PCR analysis of genomic DNA (data not shown). This insert, which differed in its sequence and position from any of the introns found in the RBMX gene, must therefore have been inserted into RBMXL4 after the retroposition event. Since this insert lies within the 3′ untranslated region, it would not affect the coding region. However, a deletion of 29 nucleotides at nt 611 created a frameshift, and no significant open

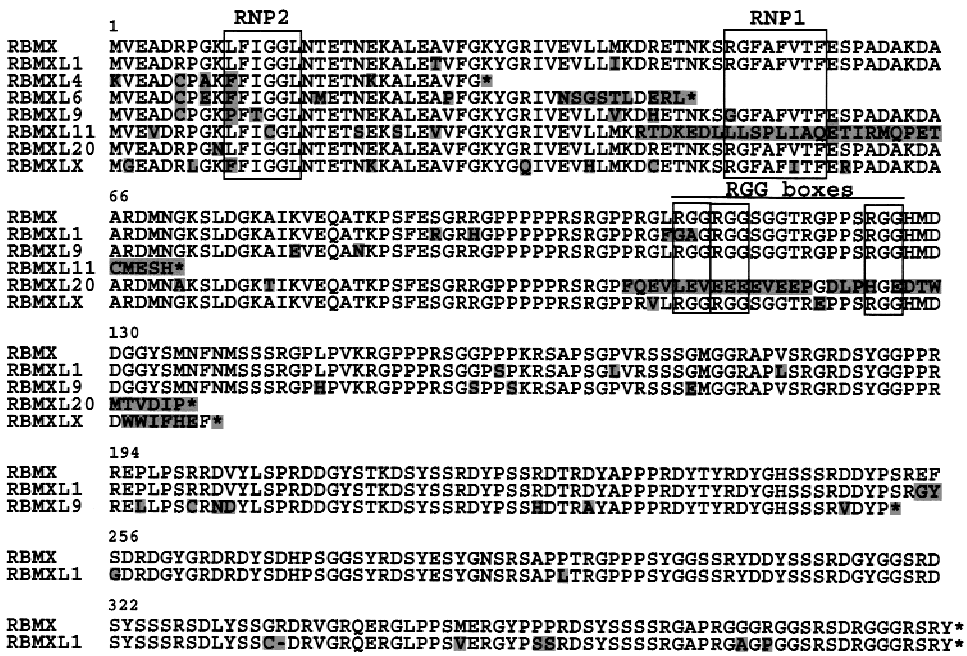

Fig. 2. Amino acid alignment of RBMX and RBMXLs. Amino acid differences are highlighted. RNA binding domains, RNP1 and RNP2, and RGG boxes are boxed. The amino acid sequence for RBMXL4 is hypothetical since the DNA sequence does not include a start codon.

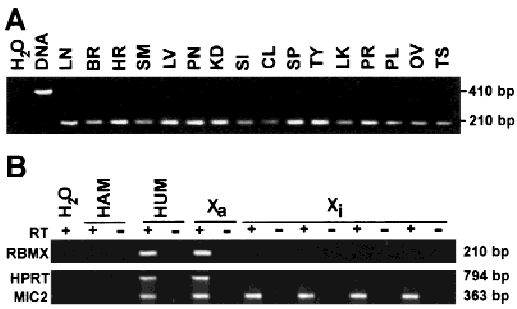

Fig. 3. A. Expression of RBMX as determined by PCR using primers specific for the RBMX sequence (Table 1) on a cDNA panel derived from multiple tissues as indicated above the lanes. A 210-bp product was obtained in all tissues. Control amplification of genomic DNA produced a larger fragment of 410 bp, as expected from the presence of introns between the primers. **B.** X inactivation of RBMX as determined by RT-PCR amplification of RBMX and control genes MIC2 (escapes) and HPRT (subject to X inactivation) from hybrid cell lines that retain an active (Xa) or inactive X (Xi) chromosome. Control amplifications of human and hamster RNA are also shown. Each PCR was done with (+) and without (–) reverse transcriptase (RT).

reading frame was found, indicating that the Chr 4 copy was probably nonfunctional (Figs. 1 and 2).

The complete 1560-bp sequence we obtained from Chr 6 had 94% homology to the coding region of RBMX and contained a-7-bp insertion which would cause a shift of reading frame within the first 120 bp, resulting in premature termination of translation (Delbridge et al. 1999). Thus, RBMXL6 is probably a nonfunctional pseudogene (Figs. 1 and 2).

The RBMXL at 9p13 (hereafter called RBMXL9) was entirely contained within clone bA255N4 (GenBank AL390844) and showed 96% nucleotide identity with RBMX. An open reading frame was found which encodes a putative protein of 216 aa, with a stop codon at nt 299, which would preserve the RNA binding domains (Figs. 1 and 2). No other mutations were found, suggesting that this copy could potentially encode a functional protein. In addition, an EST (GenBank G25315) corresponding to RBMXL9 was identified. A second site on Chr 9, at 9p24, contained sequences homologous to the 3′end of RBMX, starting at nt 1420, beyond the open reading frame (Fig. 1). This truncated sequence was embedded within two different BAC clones mapped to 9p24 (GenBank AL354707, AL162411), suggesting that two independent copies were present in the human genome. Extending this sequence by genomic walking and database searches did not reveal any further regions of homology to RBMX.

RBMXL11 on Chr 11q13 was also found in GenBank (GI8117428). It was identical to the sequence we had established from the hybrid cell line. The complete RBMXL11 sequence showed 92% identity to RBMX. However, two deletions causing frameshifts in the coding sequence and a stop codon at nt 992 suggested that this copy was probably nonfunctional (Figs. 1 and 2).

Database searches revealed an additional RBMXL20 on Chr 20 (GenBank AL390202), which neither FISH nor PCR amplification from hybrid cell lines (see above) had detected. To confirm that presence or absence of this copy in the human genome was polymorphic, we tested samples of human genomic DNA, using two different primer pairs based on the Chr 20 sequence, and found RBMXL20 to be absent in 8 of 18 Chrs 20 tested (data not shown). Examination of RBMXL20 revealed 96% homology with RBMX. A 1-bp insertion caused a frameshift at nt 315 as well as a stop codon, suggesting that this copy was probably nonfunctional (Figs. 1 and 2).

In addition to the RBMX gene located at Xq26, the FISH studies and PCR analyses of monochromosomal hybrid cell lines had suggested that the X Chr also contained a deleted processed copy of RBMX at Xq13 (see above). Partial sequence analysis of this copy indicated that it corresponded to a genomic sequence deposited in GenBank (AL03497) and mapped to Xq12-13.3. The large deletion from nt 612 to 880 of the corresponding RBMX sequence included many of the sequences contained in the FISH probe, which would explain the presence of a fainter signal occasionally seen at Xq13. Because of this large deletion, this RBMXLX was probably nonfunctional (Figs. 1 and 2).

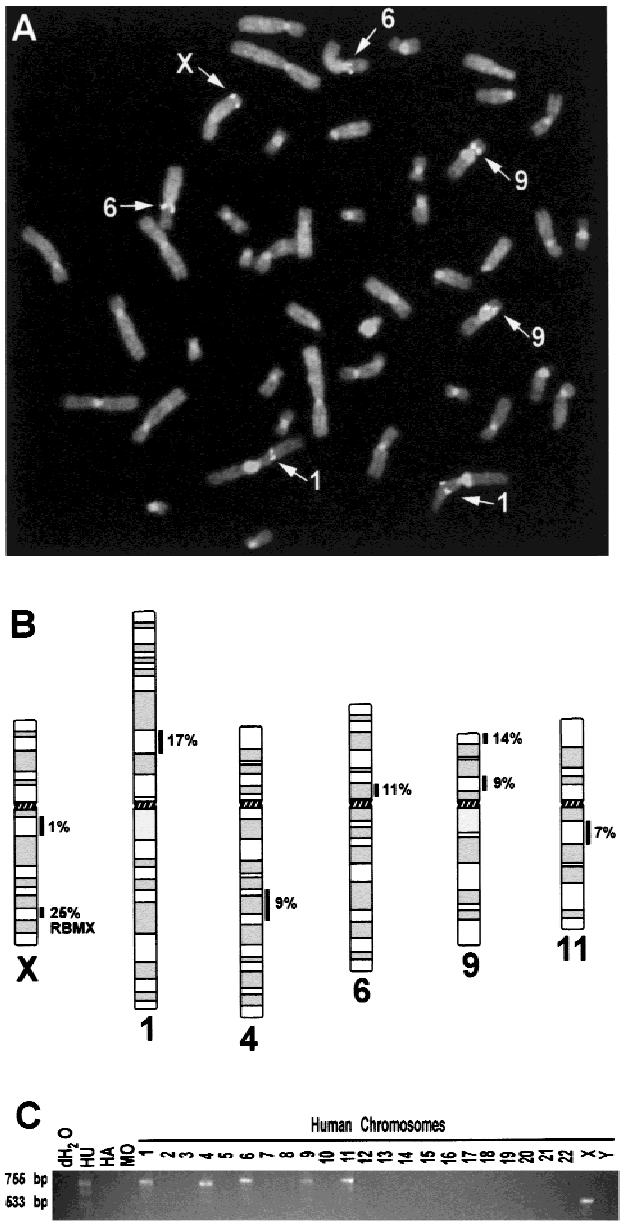

Fig. 4. A. Example of FISH of a genomic RBMX probe to human chromosomes. In the cell shown here, signals were detected on the long arm of the X Chr at Xq26, on the short arm of Chr 1 (1p22–31), the short arm of Chr 6 (6p12), and on the short arm of Chr 9 (9p13). Note that the lower Chr 9 has a pericentric inversion of the heterochromatin, representing a polymorphism in the human population. Other cells showed signals on the long arm of Chr 4 (4q25–27), on a second site on the short arm of Chr 9 (9p24), and on the long arm of Chr 11 (11q13) (not shown). **B.** Summary of the FISH data on human chromosome ideograms with the frequency of signals shown in %. **C.** PCR amplification with primers HSXF481 and HSXR1256 (Table 1) of RBMXLs from hybrid cell lines, each retaining a single human chromosome as indicated above the lanes. Parental rodent lines are in lane MO (mouse) and HA (hamster). Human control DNA is in lane HU.

Expression of the RBMX-like sequences. Expression of the retroposed copies of RBMX was examined by PCR by using multiple tissue cDNA panels I and II (Clontech). This was done after verification that each primer pair (Table 1) selectively amplified only one processed copy, as shown by amplification of products from DNA derived from the hybrid cell lines, except for RBMXL20, which is absent in the Chr 20 hybrid line (Fig. 5A). To distinguish expression of the two Chr 9 copies, we designed primers that

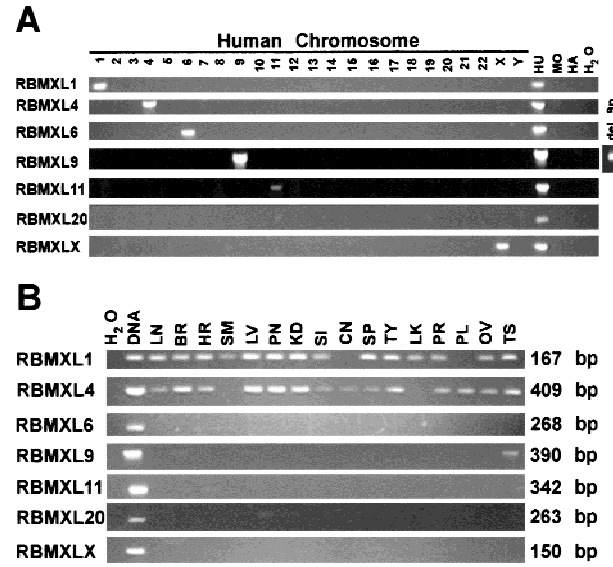

Fig. 5. A. PCR amplifications of each chromosome-specific RBMXL (indicated at left). Primer sets (Table 1) were designed to amplify a single copy in the somatic cell hybrids. Differentiation between the two copies on Chr 9 is shown at the right by amplification from a hybrid cell line that retains a deleted copy of Chr 9. Chromosomes contained in the hybrids are shown on top. Controls include human (HU), hamster (HA), and mouse (MO) DNA. **B.** PCR using the primer sets developed in (A) to examine expression of each chromosome specific RBMXL (indicated at left) by amplification of products from a panel of cDNAs derived from multiple tissues as indicated on top of the lanes. Control DNA amplification products are in the second lane. The size of the products is at right. (LN lung, BR brain, HR heart, SM skeletal muscle, LV liver, PN pancreas, KD kidney, SI small intestine, CN colon, SP spleen, TY thymus, LK peripheral blood leukocytes, PR prostrate, PL placenta, OV ovary, TS testis).

amplified each copy, as verified by amplification from the hybrid line retaining a deleted Chr 9.

PCR of the multiple tissue panels with these chromosome-specific primer pairs revealed that RBMXL1 was expressed in all tissues except colon and placenta; RBMXL4, in all tissues except skeletal muscle and leukocytes; and RBMXL9, only in testis and brain (low expression) (Fig. 5B). The RBMXLs on Chr 6, 11, and 20 and the deleted copy on the X Chr did not appear to be expressed (Fig. 5B). However, in the case of RBMXL20, we cannot exclude that all the cDNA samples examined were from an individual with no copy of this sequence.

RBMX-like sequences in other species. The conservation of RBMXLs in other species was investigated by FISH and PCR analyses. FISH analyses with a human RBMX probe revealed signals on chimpanzee, gorilla, and orangutan chromosomes. Signals were observed at the expected locus of the RBMX gene on the X Chr of these species. Signals were also detected on several autosomes, at loci corresponding to the human chromosomes that contained retroposed copies of RBMX (Table 2). PCR analyses using the chromosome-specific primers developed for human RBMXLs (see above) generated same-size products in primates as compared with humans for several of the RBMXLs, confirming the presence of autosomal retroposons in these species (Table 2). The absence of PCR products in some instances where the FISH analyses had shown a signal probably reflected base pair differences between species. This analysis indicated that the retroposition of copies of RBMX probably occurred prior to the divergence of the human in the primate lineage.

Phylogenetic trees of the different RBMX/RBMY-related sequences were constructed (Fig. 6). The DNA sequences included in the analysis were: human RBMX and RBMXLs described here,

Table 2. Summary of FISH and PCR data in other species.

	Human		Chimpanzee		Gorilla		Orangutan		Baboon		Gibbon	
	FISH	PCR	FISH	PCR	FISH	PCR	FISH	PCR	FISH	PCR	FISH	PCR
RBMX	+	410[a]	+[b]	410	+	410	+	410	+	410	+	410
RBMXL1	+	167	+	167	+	167	+	[c]		–		167
RBMXL4	+	409	+	409		409	+	[c]		409		[c]
RBMXL6	+	268	+	–	+	268	+	[c]		–		–
RBMXL9	+	390	+	–		–		[c]		–		–
RBMXL11	+	342	+	342	+	342	+	342		–		342
RBMXL20	–	263	+	263		–				–		–
RBMXLX	+	266		[c]		–	+	266		–		266

[a] Size of products in bp.
[b] A + indicates presence of signals on the primate chromosome corresponding to the human chromosome. Because of the difficulty in identifying all primate chromosomes corresponding to the human chromosomes, only consistently positive FISH signals on clearly identified primate chromosomes were recorded.
[c] A PCR product of a different size from that obtained in human was observed.

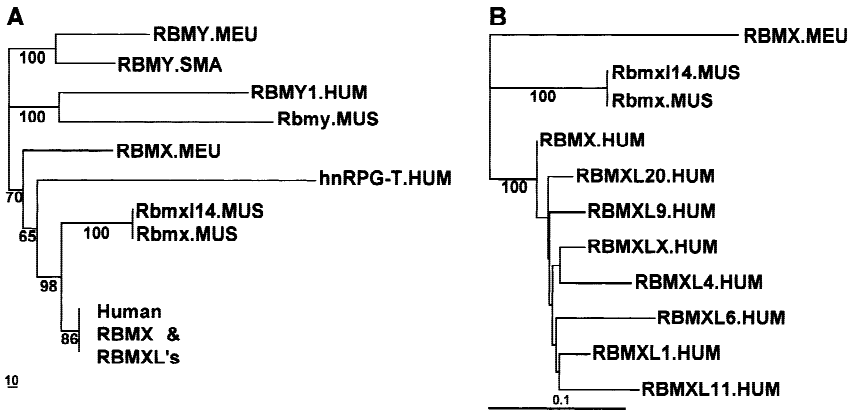

Fig. 6. Neighbor-joining phylogenies showing the relationships of RBMX/RBMY sequences from marsupial (MEU = *Macropus eugenii,* SMA = *Sminthopsis macroura*), mouse (MUS) and human (HUM). Numbers below branches indicate percentage bootstrap support (1000 replicates). Scale is indicated in percentage nucleotide divergence. **A.** Relationships of RBMY1, hnRNPG-T, RBMX, and RBMXLs. The tree was rooted with the marsupial RBMY (MEU and SMA). Only one member of the human RBMY1 gene subfamily was included in the analysis. The SGRY repeats regions of human RBMY1 were excluded from the analyses. **B.** Higher resolution showing the relationships of RBMX and RBMXLs (with the chromosome origin indicated). The tree was rooted with the RBMX gene from *M. eugenii*. hnRNPG-T is not shown (see A).

mouse *Rbmx* and an *Rbmx*-like sequence from mouse Chr 14 (Delbridge et al. 1998), marsupial *RBMX* (from *Macropus eugenii*), a novel human RBMX-related gene, hnRNPG-T (Venables et al. 2000), human RBMY1, mouse *Rbmy,* and marsupial *RBMY* (from *Macropus eugenii* and *Sminthopsis macroura*). In human, RBMY sequences form a complex set of genes, some arrayed in tandem and some dispersed on the Y (Chai et al. 1997, 1998). Interestingly, human RBMY1, RBMY2, and hnRNPG-T, and mouse *Rbmx* and *Rbmy* showed a C at nt 777 found in RBMXLs but not in RBMX (G) (Chai et al. 1997; Venables et al. 2000; Mazeyrat et al. 1999).

Trees resulting from neighbor joining, parsimony, and maximum likelihood methods all produced similar topologies. All RBMX genes (from marsupial, mouse, and human), RBMXLs (from mouse and human), and the hnRNPG-T gene clustered in a clade of the tree suggesting that the autosomal retroposed copies (RBMXLs and hnRNPG-T) were probably derived from RBMX and not from RBMY (Fig. 6A and B). This analysis also confirmed that RBMXLs described here evolved from RBMX, which was the basal gene in humans. Because the *Rbmxl* gene of *Mus musculus* did not cluster with the RBMXLs of human, this suggested that retroposition events were independent and recent in the primate and mouse lineages. Short branch lengths (and low bootstrap support) within the human RBMX/RBMXLs clade suggested relatively rapid retroposition of the genes.

Discussion

We have shown that the RBMX/RBMY gene family represents a complex set of genes with multiple copies in the human genome. The original pair of X-borne and Y-borne genes, RBMX and RBMY, have a classical intron-exon structure, whereas the dispersed copies of RBMX-like sequences in the human genome are intronless, indicating that they arose by retroposition. The RBMX/RBMY genes are thought to represent an ancient X/Y pair that was located on the original homologous pair of proto-sex chromosomes. Indeed, the RBMX/RBMY gene pair is located on the sex chromosomes of marsupials (Delbridge et al. 1999), and RBMX is located in the most ancient stratum of the human X Chr (Lahn and Page 1999a). RBMY was independently amplified in different mammalian lineages, perhaps as a result of selection to counteract mutations and loss of function (Graves et al. 1998).

RBMX/RBMY represents an example of evolution of an X/Y gene pair in human, in which the Y-gene has acquired a testis-specific function, whereas the X-gene has retained ubiquitous expression in both sexes and has become subject to X inactivation. X/Y gene pairs have undergone differing degrees of differentiation in mammalian lineages (Graves et al. 1998; Jegalian and Page 1998). Some Y partners of X/Y gene pairs have remained ubiquitously expressed and thus may have retained a similar function, in which case the X homolog escapes X inactivation, presumably to keep expression levels similar in both sexes (Disteche 1999). In the case of RBMX, our present results confirm a survey of human genes in terms of their X inactivation status done prior to the findings of multiple RBMXLs; this survey (Carrel and Willard 1999) included RBMX (listed as HNRNPG), as subject to X inactivation. The development of RBMX-specific primers in the present study allowed us to verify that it is the active X-borne gene rather than the truncated pseudogene on the X that is subject to X inactivation.

The widespread expression of RBMX in all 16 tissues tested in the present study contrasts with the testis-restricted expression of RBMY. RBMY has been specifically implicated in spermatogenesis. It is a candidate for azoospermia because RBMY deletions (in region AZFb) are associated with male infertility (Ma et al. 1993). RBMY expression is specific to male germ cells, but its timing

differs between species. In human, RBMY protein is present in spermatogonia and spermatocytes, but not in elongated spermatids, whereas in mouse the protein is associated with spermatogonia and elongated spermatids, but not spermatocytes (Elliott et al. 1998; Mahadevaiah et al. 1998). With yeast two-hybrid methods, it has recently been shown that both RBMX and RBMY interact with other proteins to perform pre-mRNA-splicing functions (Venables et al. 2000). Thus, while RBMX appears to be involved in these functions in a ubiquitous manner, RBMY may be involved in germ cell-specific splicing by providing a tissue-specific factor (Elliott et al. 2000).

Retrotransposition from RBMX is the likely mechanism of dispersion of the intronless RBMXLs on human Chr 1, 4, 6, 9, 11, 20, and X described here. The presence of RBMXLs on the corresponding primate and great-ape chromosomes is consistent with dispersion of these copies prior to the divergence of humans. Although RBMXL4, RBMXL6, RBMXL11, and RBMXL20 (which is not present in all individuals) have base changes that probably render them nonfunctional, the high degree of conservation of RBMXLs in humans and in primates suggests that other copies may be functional. Consistent with this hypothesis is the finding that three of the RBMXLs are transcribed. Interestingly, we found that their expression was restricted to specific tissues. Some transcribed retroposons are inefficiently translated into proteins, which could possibly be the case for RBMXLs (Matsumoto et al. 1998). In addition, the insertion of sequences at the 3'end of genes, such as the one found in RBMXL4, would further repress translation. Whether any of the RBMXLs reported here play a role in constitutive or alternative splicing of specific targets in a tissue-specific manner remains to be determined.

We found that the RBMXL9 located at 9p13 was specifically expressed in testis and, to a lesser extent, in brain. The presence of a stop codon in this sequence would result in a truncated protein, but that protein would still retain the RNA-binding domains. Alternatively, the sequence might be read through (McCaughan et al. 1995). Although retroposed copies of genes can be permissively expressed in testis with no apparent function, some retroposons appear to have acquired specific roles (Kleene and Mastrangelo 1999). One possibility is that RBMXL9 represents a novel gene involved in testis function.

Since RBMX is subject to X inactivation, a role for the testis-expressed RBMXL9 could be to provide compensatory expression when the X Chr becomes inactivated during spermatogenesis. Several genes subject to X inactivation have homologous intronless copies on autosomes, presumably to provide such compensatory expression. Genes of this type include phosphoglycerate kinase 1 (PGK1, human and mouse), glucose-6-phosphate dehydrogenase (*G6pd,* mouse), glycerol kinase (GKD, human), pyruvate dihydrogenase E1-alpha subunit (PDHA1, human and mouse) and XAP-5 (EST of unknown function, human and mouse) (Boer et al. 1987; McCarrey and Thomas 1987; Dahl et al. 1990; Hendriksen et al. 1997; Sargent et al. 1994; Sedlacek et al. 1999).

A novel RBMX-related sequence called hnRNPG-T has been reported to interact with the same splicing activator Tra2 beta as RBMX and RBMY (Venables et al. 2000). The hnRNPG-T gene is an intronless sequence that resembles RBMX but has a lower level of sequence homology to RBMX than any of the RBMXLs described here. The identity between RBMX and hnRNPG-T is only about 69% at the nucleotide level and 73% at the amino acid level, whereas we found identity of nucleotide sequences in the 92–96% range for the RBMXLs. This would have precluded us from detecting hnRNPG-T by FISH, given our stringent washing conditions. It is intriguing that hnRNPG-T is selectively expressed in testis and in brain (Venables et al. 2000), like RBMXL9. It will be interesting to determine whether hnRNPG-T interacts with the RBMXL9 in these tissues. Like RBMXL9, hnRNPG-T gene could also provide RBMX compensatory function in spermatocytes when RBMX is inactivated (Venables et al. 2000).

It is remarkable that several other X/Y gene pairs have dispersed, often processed copies of the X-linked homologs in their genome (Ashworth et al. 1990; Ehrmann et al. 1998). This retroposition may have been necessary as part of the evolution of X/Y genes which included Y gene loss or acquisition of a male-specific function and incorporation of the X gene into the X inactivation system (Graves et al. 1998). In the case of RBMX/RBMY, multiple retroposition and duplication events may have occurred at different times, possibly leading to differentiation of the copies into tissue-specific splicing factors. The evolutionary pathway of the RBMX gene family (Fig. 6) resembles that of the mouse *Pgk1* gene family, which contains a large number of retroposed copies including the testis-specific gene *Pgk2*. Some of the retroposed copies of *Pgk1* appear to have arisen millions of years ago, and others have continued to arise as recently as the divergence of the laboratory mouse strains (Adra et al. 1988). Sequence analysis of other X-linked genes reveals that some testis-specific retroposed copies occurred in an ancestral species and persisted in divergent lineages (Sedlacek et al. 1999), while others evolved independently in separate lineages (Hendriksen et al. 1997). It is probable that hnRNPG-T represents an ancient RBMX retroposon that has diverged in its sequence, whereas the highly conserved RBMXLs described here have arisen more recently. Examination of the organization and function of the different copies of RBMX in multiple species will help sort out their evolutionary history.

Acknowledgments. We thank Scott Hansen (University of Washington, Seattle) for the human × rodent hybrid cell lines that retained an active or inactive X Chr and for his helpful advice. We are grateful to Barbara Trask and Cynthia Friedman (University of Washington, Seattle) for the primate chromosome preparations and for their help with karyotype identification. This work was supported by grant GM 46883 to C.M.Disteche from the National Institutes of Health, and grants to J.A.M. Graves from the Australian Research Council and the National Health and Medical Research Council.

References

Adra CN, Ellis NA, McBurney MW (1988) The family of mouse phosphoglycerate kinase genes and pseudogenes. Somat Cell Mol Genet 14, 69–81

Agulnik AI, Mitchell MJ, Mattei MG, Borsani G, Avner PA et al. (1994) A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. Hum Mol Genet 3, 879–884

Ashworth A, Skene B, Swift S, Lovell-Badge R (1990) Zfa is an expressed retroposon derived from an alternative transcript of the Zfx gene. EMBO J 9, 1529–1534

Boer PH, Adra CN, Lau YF, McBurney MW (1987) The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon. Mol Cell Biol 7, 3107–3112

Carrel L, Willard HF (1999) Heterogeneous gene expression from the inactive X chromosome: An X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. Proc Natl Acad Sci USA 96, 7364–7369

Chai NN, Phillips A, Fernandez A, Yen PH (1997) A putative human male infertility gene DAZLA: genomic structure and methylation status. Mol Hum Reprod 3, 705–708

Chai NN, Zhou H, Hernandez J, Najmabadi H, Bhasin S et al. (1998) Structure and organization of the RBMY genes on the human Y chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. Genomics 49, 283–289

Charlesworth B (1991) The evolution of sex chromosomes. Science 251, 1030–1033

Dahl HH, Brown RM, Hutchison WM, Maragos C, Brown GK (1990) A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. Genomics 8, 225–232

Delbridge ML, Ma K, Subbarao MN, Cooke HJ, Bhasin S et al. (1998) Evolution of mammalian HNRPG and its relationship with the putative azoospermia factor RBM. Mamm Genome 9, 168–170

Delbridge ML, Lingenfelter PA, Disteche CM, Graves JAM (1999) The

candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. Nat Genet 22, 223–224
D'Esposito M, Ciccodicola A, Gianfrancesco F, Esposito T, Flagiello L et al. (1996) A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. Nat Genet 13, 227–229
Disteche CM (1999) Escapees on the X chromosome. Proc Natl Acad Sci USA 96, 14180–14182
Edelhoff S, Ayer DE, Zervos AS, Steingrimsson E, Jenkins NA et al. (1994) Mapping of two genes encoding members of a distinct subfamily of MAX interacting proteins: MAD to human chromosome 2 and mouse chromosome 6, and MXI1 to human chromosome 10 and mouse chromosome 19. Oncogene 9, 665–668
Ehrmann IE, Ellis PS, Mazeyrat S, Duthie S, Brockdorff N et al. (1998) Characterization of genes encoding translation initiation factor eIF-2gamma in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. Hum Mol Genet 7, 1725–1737
Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F et al. (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. Proc Natl Acad Sci USA 94, 3848–3853
Elliott DJ, Oghene K, Makarov G, Makarova O, Hargreave TB, et al. (1998) Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. J Cell Sci 111, 1255–1265
Elliott DJ, Bourgeois CF, Klink A, Stevenin J, Cooke HJ (2000) A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection. Proc Natl Acad Sci USA 97, 5717–5722
Ellison KA, Roth EJ, McCabe ER, Chinault AC, Zoghbi HY (1993) Isolation of a yeast artificial chromosome contig spanning the X chromosomal translocation breakpoint in a patient with Rett syndrome. Am J Med Genet 47, 1124–1134
Esposito T, Gianfrancesco F, Ciccodicola A, D'Esposito M, Nagaraja R et al. (1997) Escape from X inactivation of two new genes associated with DXS6974E and DXS7020E. Genomics 43, 183–190
Fu XD (1995) The superfamily of arginine/serine-rich splicing factors. RNA 1, 663–680
Graves JAM (1995) The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. Bioessays 17, 311–320
Graves JAM, Disteche CM, Toder R (1998) Gene dosage in the evolution and function of mammalian sex chromosomes. Cytogenet Cell Genet 80, 94–103
Hansen RS, Canfield TK, Fjeld AD, Gartler SM (1996) Role of late replication timing in the silencing of X-linked genes. Hum Mol Genet 5, 1345–1353
Hendriksen PJ, Hoogerbrugge JW, Baarends WM, de Boer P, Vreeburg JT et al. (1997) Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. Genomics 41, 350–359
Hornstra IK, Yang TP (1994) High-resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5′ region on the active and inactive X chromosomes: correlation with binding sites for transcription factors. Mol Cell Biol 14, 1419–1430
Jegalian K, Page DC (1998) A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. Nature 394, 776–780
Kent-First M, Muallem A, Shultz J, Pryor J, Roberts K, et al. (1999) Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. Mol Reprod Dev 53, 27–41
Kleene KC, Mastrangelo MA (1999) The promoter of the Poly(A) binding protein 2 (*Pabp2*) retroposon is derived from the 5′-untranslated region of the *Pabp1* progenitor gene. Genomics 61, 194–200
Kostiner DR, Turek PJ, Reijo RA (1998) Male infertility: analysis of the markers and genes on the human Y chromosome. Hum Reprod 13, 3032–3038
Lahn BT, Page DC (1999a) Four evolutionary strata on the human X chromosome. Science 286, 964–967
Lahn BT, Page DC (1999b) Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. Nat Genet 21, 429–433
Ledbetter SA, Schwartz CE, Davies KE, Ledbetter DH (1991) New somatic cell hybrids for physical mapping in distal Xq and the fragile X region. Am J Med Genet 38, 418–420
Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE et al. (1993) A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell 75, 1287–1295
Mahadevaiah SK, Odorisio T, Elliott DJ, Rattigan A, Szot M et al. (1998) Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. Hum Mol Genet 7, 715–727
Matsumoto K, Wassarman KM, Wolffe AP (1998) Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. EMBO J 17, 2107–2121
Mazeyrat S, Saut N, Mattei MG, Mitchell MJ (1999) RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. Nat Genet 22, 224–226
McCarrey JR, Thomas K (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326, 501–505
McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP (1995) Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc Natl Acad Sci USA 92, 5431–5435
Ohno S (1967) Sex chromosomes and sex linked genes. In *Monographs on Endocrinology,* Londhardt A, Mann T, Samuels LT, Zander J (Springer Berlin: Verlag)
Pryor JL, Kent-First M, Muallem A, Van Bergen AH, Nolten WE et al. (1997) Microdeletions in the Y chromosome of infertile men. N Engl J Med 336, 534–539
Rice WR (1996) Evolution of the Y sex chromosome in animals. BioScience 46, 331–343
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4, 406–425
Sargent CA, Young C, Marsh S, Ferguson-Smith MA, Affara NA (1994) The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. Hum Mol Genet 3, 1317–1324
Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H et al. (1996) The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet 14, 292–299
Sedlacek Z, Munstermann E, Dhorne-Pollet S, Otto C, Bock D et al. (1999) Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of an ancient functional retroposon differentially expressed in testis. Genomics 61, 125–132
Soulard M, Della Valle V, Siomi MC, Pinol-Roma S, Codogno P et al. (1993) hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Res 21, 4210–4217
Swofford DL (1997) Phylogenetic analysis using parsimony (PAUP). (Sunderland, Mass: Sinauer Press)
Venables JP, Elliott DJ, Makarova OV, Makarov EM, Cooke HJ et al. (2000) RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with tra2beta and affect splicing. Hum Mol Genet 9, 685–694
Yen PH, Chai NN, Henandez J, Najmabadi H, Bhasin S (1995) Characterisation of several YRRM genes in deletion interval of the human Y chromosome. Am J Hum Mol Genet 57, A154