An X-to-autosome retrogene is required for spermatogenesis in mice

Julie Bradley\textsuperscript{1,4}, Andrew Baltus\textsuperscript{1,4}, Helen Skalesky\textsuperscript{1}, Morgan Royce-Tolland\textsuperscript{1}, Ken Dewar\textsuperscript{2,3} & David C. Prince\textsuperscript{1}

We identified the gene carrying the juvenile spermatogonial depletion mutation (\textit{jsd}), a recessive spermatogenic defect mapped to mouse chromosome 1 (refs. 1, 2). We localized \textit{jsd} to a 272-kb region and resequenced this area to identify the underlying mutation: a frameshift that severely truncates the predicted protein product of a 2.3-kb genomic open reading frame. This gene, \textit{Utp14b}, evidently arose through reverse transcription of an mRNA from an X-linked gene and integration of the resulting cDNA into an intron of an autosomal gene, whose promoter and 5' untranslated exons are shared with \textit{Utp14b}. To our knowledge, \textit{Utp14b} is the first protein-coding retrogene to be linked to a recessive mammalian phenotype. The X-linked progenitor of \textit{Utp14b} is the mammalian ortholog of yeast Utp14, which encodes a protein required for processing of pre-rRNA\textsuperscript{3} and hence for ribosome assembly. Our findings substantiate the hypothesis\textsuperscript{4} that mammalian spermatogenesis is supported by autosomal retrogenes that evolved from X-linked housekeeping genes to compensate for silencing of the X chromosome during male meiosis\textsuperscript{5-7}. We find that \textit{Utp14b}-like retrogenes arose independently and were conserved during evolution in at least four mammalian lineages. This recurrence implies a strong selective pressure, perhaps to enable ribosome assembly in male meiotic cells.

The \textit{jsd} mutation was previously mapped to a 0.4-cM region of mouse chromosome 1, spanning about 1.5 Mb\textsuperscript{1,2}. By analyzing 4,826 meioses from two \textit{F\textscript{2}} intercrosses (Supplementary Fig. 1 online), we confirmed this localization and more precisely mapped \textit{jsd} to a region of <0.1 cM (Fig. 1a). Sequencing of two wild-type BACs spanning the critical region showed that it measured \textasciitilde272 kb.

The critical region contains only two genes that are conserved in mice and humans: \textit{Acsl3} and \textit{Kcnb4}. (In humans, the genes \textit{ACSL3} and \textit{KCNB4} are also located adjacent to one another, on chromosome 2, in a region orthologous to mouse chromosome 1.) In addition, GenScan analysis of the critical region sequence identified a 2.3-kb open reading frame (ORF) in the second intron of \textit{Acsl3} (Fig. 1b) that was absent from the homologous region of the human gene. We refer to this ORF as ORF2.3.

Using the wild-type BAC sequence as a reference, we resequenced the critical region in genomic DNA from \textit{jsd/jsd} mice. We found a sequence difference in ORF2.3 at nucleotide 312: the dinucleotide GG in the wild-type chromosome was replaced by TTTTC in \textit{jsd} (Fig. 1c). This induced a frameshift in ORF2.3 and introduced a premature termination codon at nucleotide 312, severely truncating the predicted protein in ORF2.3. The \textit{jsd} mutation was genetically inseparable from the \textit{jsd} gene.

We found no other substantial sequence differences between \textit{jsd} and wild-type chromosomes. We detected single-unit differences in the lengths of two microsatellite arrays, both located upstream of \textit{Acsl3} and \textit{Kcnb4}. These coding and 5' upstream regions of \textit{Acsl3} and \textit{Kcnb4} were identical in \textit{jsd} and wild-type chromosomes. We were able to completely resequence 13 microsatellite or other long arrays, but none of these fell in coding or putative regulatory regions. The absence of substantial sequence differences elsewhere in the critical region indicates that the GG\textsuperscript{\textrightarrow}TTTTC mutation in ORF2.3 is responsible for the spermatogenic failure observed in \textit{jsd/jsd} males.

We then tested whether ORF2.3 is transcribed in \textit{jsd/jsd} males. Transcription of wild-type spermatagonia of \textit{jsd/jsd} mice rescues spermatogenesis, implying that \textit{jsd/jsd} spermatogonia are testicular germ cells\textsuperscript{8,9}. We found that ORF2.3 was transcribed in testes from \textit{jsd/jsd} males (Fig. 2a), in keeping with the frameshift mutation. ORF2.3 seemed to be expressed in testes of wild-type mice, in keeping with the frameshift mutation. ORF2.3 seemed to be expressed only in testes of wild-type mice (Fig. 2b). ORF2.3 was also transcribed in testes of \textit{jsd/jsd} mice, in keeping with the frameshift mutation. ORF2.3 seemed to be expressed only in testes of wild-type mice (Fig. 2b).

Published online 18 July 2004; doi:10.1038/ng1390
Examination of the cDNA sequence containing ORF2.3 and comparison with the genome DNA sequence identified two 5' untranslated exons, both of which also serve as 5' untranslated exons in Acs3 (Fig. 1b). We called the gene that contains both ORF2.3 and its 5' untranslated exons Utp14b. We conclude that Acs3 and Utp14b form a bicistronic unit with a shared promoter.

Because the jsd mutation lies in an intron of Acs3, we considered the possibility that Acs3 itself, and not Utp14b, is the gene defective in jsd/jsd mice. If so, the mutation should disrupt transcription or splicing of Acs3, as its coding sequence is unchanged. By northern blot analysis, we observed a 3.5-kb Acs3 transcript in all tissues tested, with strongest expression in testis and brain (Fig. 2b,c). Thus, Acs3 was expressed in a wider range of tissues than Utp14b. Neither the size nor the abundance of this Acs3 transcript seemed to be altered in testes of jsd/jsd mice, suggesting that the mutation does not interfere with Acs3 transcription or splicing. We conclude that Utp14b, and not Acs3, is the gene defective in jsd/jsd mice, and that truncation of the protein product of Utp14b causes spermatogenic failure.

At a molecular level, mouse Utp14b is atypical not only in that it forms part of a larger, bicistronic unit, which are rare in mammalian genomes,10 but also because it has no ortholog in the human genome, and its coding sequence is not interrupted by introns. Electronic searches of the mouse and human genomes showed that the X chromosome carries a widely expressed gene (Utp14a) whose coding sequence is homologous to that of mouse Utp14b but is interrupted by 14 introns (Fig. 3 and Supplementary Fig. 2 online).11 The existence of a conserved, intron-bearing homolog suggested a molecular evolutionary explanation for Utp14b. After the mouse lineage diverged from that of humans, a processed mRNA derived from the X-chromosomal gene Utp14a was probably reverse-transcribed. The resulting cDNA was integrated into an intron of the autosomal gene Acs3, whose function was preserved despite the elimination of its promoter and 5' untranslated exons in the service of the newly created gene Utp14b.

Other tests-expression autosomal retrogenes with X-linked progenitors have been reported in mammals.12-31 In each case, as with Utp14b, the X-linked source gene is expressed widely, whereas the autosomal retrogene is expressed most prominently in testis. Mouse Utp14b is the first protein-coding retrogene in mammals to which a recessive phenotype has been ascribed. Zfg, an autosomal retrogene derived from the X-chromosomal gene Zfg,16,19, is the only other protein-coding retrogene in which a loss-of-function mutation has been
Figure 3 Schematic alignment of transcripts from mouse Up14b, its X-linked homologs Up14a and UTP14A in mouse and human, respectively, and its human autosomal homolog UTP14C. Coding regions are shown in black, and positions of introns are indicated. Percentage identities (nucleotide (nt) and amino acid (aa)) between coding sequences are given.

We next searched for intronless derivatives of Up14a in 16 other eutherian species. We carried out PCR on genomic DNAs from these 16 species using primers from regions of nucleotide-sequence conservation among the mouse and human Up14 genes. We identified a Up14a-derived retrogene in 13 species, including such diverse eutherians as hamster, lemur, elephant and cow. In all 13 cases, we confirmed both strong coding-sequence similarity to Up14a and the absence of coding-region introns by sequencing of PCR products. In every case, the sequenced region (at least 750 nucleotides, and typically >1,500 nucleotides) showed an intact, conserved ORF, with no frameshift mutations, no premature stop codons and a predominance of silent nucleotide substitutions when compared with mouse or human Utp14a. These findings suggest that most eutherians possess a Up14a-derived retrogene that encodes a functional protein subject to purifying selection. We did not find Up14a-derived retrogenes in dog, horse or rabbit, but this negative evidence is inconclusive. Sequence divergence might account for the failure of our PCR primers to amplify a product in these species.

Finally, we carried out a sequence-based phylogenetic analysis of the known Up14a-derived retrogenes to determine whether they could all be accounted for by the two retroposition events that gave rise to mouse Up14b and human UTP14C. This analysis indicated that at least two (and perhaps three) additional, independent retroposition events are required to account for the array of Up14a-derived retrogenes observed in eutheria (Fig. 4 and Supplementary Fig. 5 online). Thus, the retrogene and fixation of Utp14b during rodent evolution was reproduced.

Figure 4 Phylogenetic comparison of Utp14b homologs in diverse eutherian mammals indicates that a minimum of four retroposition events (indicated by blue, red, yellow and green branches) occurred. This unrooted tree depicts phylogenetic relationships between the X-linked, intron-bearing Up14a orthologs in mouse and human and their identified intronless derivatives in nine eutherians. Mouse Utp14b and its retrogene homologs in rodents seem to be orthologs, (i.e., the result of a single retroposition event; blue branch). An independent event (red branch) accounts for the human and baboon retrogene, which are orthologs. Another event (yellow branch, node well-separated from the human/baboon node) accounts for the lemur retrogene. At least one and perhaps two events (green branch) account for the cow and elephant retrogenes. The lengthy branches to the cow and elephant retrogenes indicate that the responsible event(s) occurred well before that responsible for the lemur retrogene, whose branch is relatively short.
presumably with independent sites of insertion, in other eutherian lineages. This is notable, as most retrotransposed copies of protein-coding genes identified to date in mammalian genomes are functionless pseudogenes. We found no UpI promoters in other mouse or human. Interestingly, we did find a processed pseudogene of UpI-Ib in the mouse (Supplementary Fig. 5 online). The fixation and maintenance of UpI-Ib-derived retrogenes in multiple eutherian lineages implies a persistent and widespread selective pressure. That selective pressure could arise from the need to assemble ribosomes and thereby support spermatogenesis in the face of meiotic X chromosome inactivation.

METHODS
Mice. C57Bl/6j jol+ male mice were a gift from W. Beam (The Jackson Laboratory). We purchased C57Bl/6J, DBA2J, P5V4 and Mus musculus castaneus/Ei mice from The Jackson Laboratory.

Genetic mapping. We mated C57Bl/6j jol+ females with DBA2J or Mus musculus castaneus/Ei males and intercrossed the offspring. We genotyped the resulting progeny for polymorphic DNa markers located on chromosome 1 and flanking jol; see Supplementary Table 1 for markers, PCR primers and assay conditions. We determined the jol genotypes of critical recombinants (Supplementary Fig. 1 online) by progeny testing, crossing to jol+ and jol− mice. We phenotyped adult males by dissecting testes, which in jol+ mice weighed no more than one third as much as the testes of jol+ and 1/2 littermates. These experiments were approved by the Committee on Animal Care at The Massachusetts Institute of Technology.

BAC sequencing and electronic gene prediction. We screened both the RPCL-22 and RPCL-23 libraries to construct a BAC contig of the jol region of mouse chromosome 1. Two BACs spanning the critical jol region, RPCL-22 292124 and RPCL-23 295112, were sequenced at the Whitehead Institute/MIT Center for Genome Research as described. We searched for known and previously unidentified genes in the critical region by analyzing its sequence using RepeatMasker, GENSCAN and a BLAST search of the mouse and human expressed-sequence tag and nonredundant segments of GenBank.

Resequencing. We generated a series of overlapping fragments, each about 1 kb in length, that collectively span the 272 kb critical region (Fig. 1) by PCR using genomic DNA from C57Bl/6j jol+ mice as starting material We selected PCR primers using Primer3. We purified PCR products on Sephacryl-S300 columns and sequenced them using fluorescent-dye-terminator cycle sequencing protocols (BigDye kit, Amersham). Primers used in PCR generation of sequencing templates were also used as sequencing primers, and we selected additional sequencing primers at sites internal to the PCR-generated templates. Whenever the C57Bl/6j jol+ sequence differed from the BAC reference sequence, we resequenced C57Bl/6j wild-type genomic DNA, which resolved all discrepancies except those described in the text.

cDNA and transcriptional analysis. We screened a mouse testis cDNA library (Stratagene) by hybridization using probes corresponding to 3′ and 5′ regions of ORF2,3. We selected two cDNA inserts into plasmids and sequenced them in their entirety. Both cDNA clones contained the 3′ untranslated region and were polyadenylated. The longer clone, 1pL14h15.5, included 5′ untranslated sequence and is depicted in Figure 1.

We isolated total RNAs from mouse tissues homogenized in TRIzol reagent (GibcoBRL) and poly(A)1 RNAs from total RNAs using the MicroPoly(A) Purist kit (Ambion). We prepared northern blots using 2 mg of purified poly(A)1 RNA per lane and the NorthernMax-Gly (Ambion) glyoxal-based system. We hybridized northern-blots membranes to probes (radioactively labeled by random priming in ExpressHyb (Clontech) hybridization solution at 68 °C, washed them with 2x saline sodium citrate and 0.1% SDS at room temperature and 0.2x saline sodium citrate and 0.1% SDS at 50 °C and exposed them to film at -80 °C for 12-48 h. Before hybridizations, we stripped northern-blots membranes in 1% SDS, 10 mM Tris and 1 mM EDTA for 15 min at 80 °C, equilibrated them in 2x saline sodium citrate and 0.1% SDS for 2 min and then exposed them to control film.

For RT-PCR, we used 1-5 mg of poly(A)1 RNA in an oligo (dT)-primed first-strand cDNA synthesis reaction with Stratagene reverse transcriptase (Stratagene).

PCR amplification of homologs in diverse mammals. We chose primers from regions of at least 15 nucleotides of sequence identity among mouse Up14b, mouse and human UTP14A and human UTP14C. We tested all primer pairs using mouse and human genomic DNA as templates. We then applied satisfactory primer pairs to genomic DNA from gorilla, chimpanzee, baboon, orangutan, rhesus monkey, squirrel monkey, lemurs, cow, elephant, horse, dog, rabbit, guinea pig, hamster, lemming and rat. We purified and sequenced the resulting PCR products as described earlier. We obtained the sequences used in phylogenetic analysis (Fig. 4 and Supplementary Fig. 5 online) using species-specific primers to amplify and sequence a single PCR product from each species.

Phylogenetic analysis. We built phylogenetic trees using the neighbor-joining method with Kimura 2-parameter distance as implemented in PHYLP software (version 3.5c).


GenBank accession numbers. jol critical region BACs RPCL-22 292124, AC079222; RPCL-23 295112, AC079134, Mouse Utp14b cDNAs, AI61616 and AI616162 mouse Utp14a cDNA, AI616163; human UTP14A cDNA, BC011449; human UTP14C cDNA (KIAA0266), D87455. Genomic sequence of Utp14b-like retrogenes baboon, AI316164; chimpanzee, AI316165; cow, AI316166; elephant, AI316167; gorilla, AI316168; guinea pig, AI316169; hamster, AI316170; lemming, AI316171; lemur, AI316172; orangutan, AI316173; rat, AI316174; rhesus monkey, AI316175; squirrel monkey, AI316176. Genomic sequence of Utp14b-derived pseudogene mice, AI614172.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
We thank W. Beam for providing C57Bl/6j jol+ mice; the Genome Sequencing group at the Whitehead/MIT Center for Genome Research for PCR sequencing; R. Bilen and R.S. Lander for support; A. Botvinick, J. Alfoldi, J. Koushov, I. Lange, I. Posada, J. Saizón, T. Wang, K. Kleene and S. Rosen for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Received 3 February, accepted 3 June 2004

Published online at http://www.nature.com/naturegenetics/

2. Boteng-Eng, H. et al. Identification and sequencing the juvenile spermatogonial depletion group at the Whitehead/MIT Center for Genome Research for PCR sequencing. R. Bilen and R.S. Lander for support; A. Botvinick, J. Alfoldi, J. Koushov, I. Lange, I. Posada, J. Saizón, T. Wang, K. Kleene and S. Rosen for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
We thank W. Beam for providing C57Bl/6j jol+ mice; the Genome Sequencing group at the Whitehead/MIT Center for Genome Research for PCR sequencing; R. Bilen and R.S. Lander for support; A. Botvinick, J. Alfoldi, J. Koushov, I. Lange, I. Posada, J. Saizón, T. Wang, K. Kleene and S. Rosen for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Received 3 February, accepted 3 June 2004

Published online at http://www.nature.com/naturegenetics/


