nature genetics

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

Julie Bradley[1,4], Andrew Baltus[1,4], Helen Skaletsky[1], Morgan Royce-Tolland[1], Ken Dewar[2,3] & David C Page[1]

We identified the gene carrying the juvenile spermatogonial depletion mutation (*jsd*), a recessive spermatogenic defect mapped to mouse chromosome 1 (refs. 1,2). We localized *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, *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 *Utp14b*. To our knowledge, *Utp14b* is the first protein-coding retrogene to be linked to a recessive mammalian phenotype. The X-linked progenitor of *Utp14b* is the mammalian ortholog of yeast Utp14, which encodes a protein required for processing of pre-rRNA[3] and hence for ribosome assembly. Our findings substantiate the hypothesis[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[5–7]. We find that *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 *jsd* mutation was previously mapped to a 0.4-cM region of mouse chromosome 1, spanning about 1.5 Mb[1,2]. By analyzing 4,826 meioses from two F_1 intercrosses (**Supplementary Fig. 1** online), we confirmed this localization and more precisely mapped *jsd* to a region of <0.1 cM (**Fig. 1a**). Sequencing of two wild-type BACs spanning the critical region showed that it measured ~272 kb.

The critical region contains only two genes that are conserved in mice and humans: *Acsl3* and *Kcne4*. (In humans, the genes *ACSL3* and *KCNE4* 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 *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 *jsd*/*jsd* mice to identify any differences from the wild-type strain (C57BL/6) on which the mutation arose spontaneously (sequence available on our website; see URLs). We found a sequence difference in ORF2.3 at nucleotides 306–307, where the dinucleotide GG in the wild-type chromosome was replaced by the hexanucleotide CTTTTC in the *jsd* chromosome (**Fig. 1c**). This mutation in ORF2.3 shifts the reading frame and introduces a stop codon at nucleotide 312, severely truncating the predicted protein. This mutation in ORF2.3 was genetically inseparable from the *jsd* mutation in all 4,826 meioses examined (**Supplementary Fig. 1** online).

We found no other substantial sequence differences between *jsd* and wild-type chromosomes. We detected single-unit differences in the lengths of two microsatellite arrays, both located well away from coding sequences. The coding and 5′ upstream regions of *Acsl3* and *Kcne4* were identical on *jsd* and wild-type chromosomes. We were unable to completely resequence 13 microsatellite or other low-complexity 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→CTTTTC mutation in ORF2.3 causes the spermatogenic failure observed in *jsd*/*jsd* male mice.

We then tested whether ORF2.3 is transcribed and, if so, whether its expression pattern is consistent with its being the gene defective in *jsd*/*jsd* males. Transplantation of wild-type spermatogonia into testes of *jsd*/*jsd* mice rescues spermatogenesis, implying that *jsd* functions in testicular germ cells[8,9]. We found that ORF2.3 was transcribed in testes by sequencing of wild-type testis cDNA clones (**Fig. 1b**) and northern-blot analysis (**Fig. 2a**). ORF2.3 transcription did not differ significantly in testes of wild-type versus *jsd*/*jsd* mice, in keeping with it carrying a frameshift mutation. ORF2.3 seemed to be expressed in the spermatogenic cells themselves, as transcripts were not found in germ cell–deficient testes (from W^v/W^v mice; **Fig. 2d**) but were detected in purified spermatogonia, spermatocytes and spermatids (by RT-PCR; data not shown). ORF2.3 was also transcribed in brain (**Fig. 2a**). Published spermatogonial transplantation findings[8,9] are compatible with expression of the gene containing the *jsd* mutation not only in testes but also elsewhere in the body. Thus, the transcription pattern of ORF2.3 is consistent with previous findings for *jsd*.

[1]Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [2]Whitehead Institute/MIT Center for Genome Research, Cambridge, Massachusetts 02142, USA. [3]Present address: McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada. [4]These authors contributed equally to this work. Correspondence should be addressed to D.C.P.

Published online 18 July 2004; doi:10.1038/ng1390

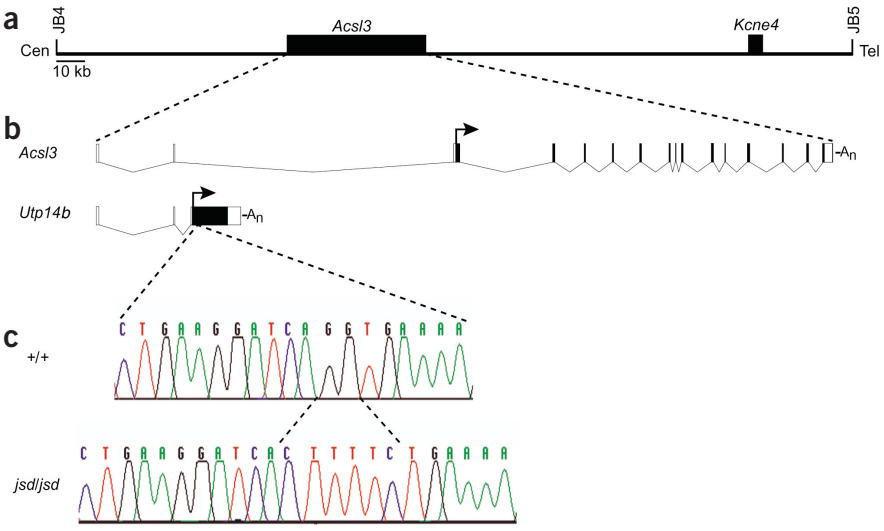

Figure 1 Insertion-deletion mutation in a transcribed ORF (ORF2.3) in the *jsd* critical region. (**a**) Sequence-based map of the *jsd* critical region, oriented with respect to the centromere (left) and telomere (right). The *jsd* mutation was mapped between genetic markers JB4 and JB5 (**Supplementary Fig. 1** online). Black boxes indicate the locations of two genes conserved between human and mouse. (**b**) *Acsl3* and *Utp14b* (ORF2.3) transcripts. Coding exons are shown in black, and translation start sites are indicated by arrows. A_n, poly-A tracts. (**c**) Sequence of nucleotides 295–313 of wild-type ORF2.3 and corresponding region of *jsd* chromosome.

Examination of the cDNA sequence containing ORF2.3, and comparison with the genomic DNA sequence, identified two 5′ untranslated exons, both of which also serve as 5′ untranslated exons in *Acsl3* (**Fig. 1b**). We called the gene that contains both ORF2.3 and its 5′ untranslated exons *Utp14b*. We conclude that *Acsl3* and *Utp14b* form a bicistronic unit with a shared promoter.

Because the *jsd* mutation falls in an intron of *Acsl3*, we considered the possibility that *Acsl3* itself, and not *Utp14b*, is the gene defective in *jsd/jsd* mice. If so, the mutation should disrupt transcription or splicing of *Acsl3*, as its coding sequence is unchanged. By northern-blot analysis, we observed a 3.5-kb *Acsl3* transcript in all tissues tested, with strongest expression in testis and brain (**Fig. 2b**,**c**). Thus, *Acsl3* was expressed in a wider range of tissues than *Utp14b*. Neither the size nor the abundance of this *Acsl3* transcript seemed to be altered in testes of *jsd/jsd* mice, suggesting that the mutation does not interfere with *Acsl3* transcription or splicing. We conclude that *Utp14b*, and not *Acsl3*, is the gene defective in *jsd/jsd* mice, and that truncation of the protein product of *Utp14b* causes spermatogenic failure.

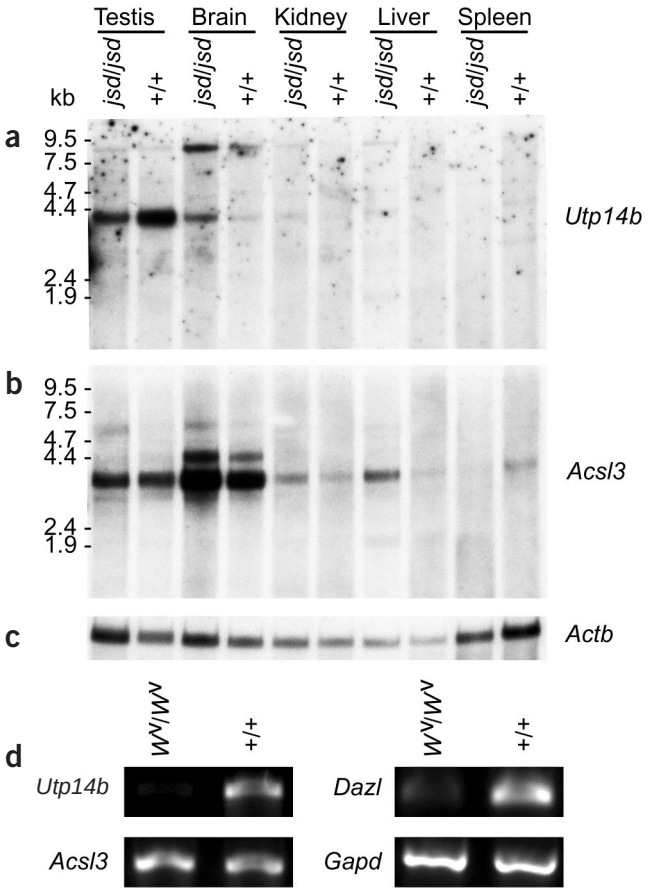

Figure 2 Transcription of *Utp14b* (ORF2.3) and *Acsl3*. (**a**) Northern-blot analysis of *Utp14b* (ORF2.3) in five tissues from *jsd/jsd* and wild-type 7-week-old male mice. (**b**) Hybridization of the same blot with an *Acsl3* probe. (**c**) Control hybridization of the same blot with an *Actb* probe. (**d**) RT-PCR analysis of poly(A)$^+$ RNA samples from testes of adult W^v/W^v (germ cell–depleted) and wild-type mice. *Dazl* (expressed only in germ cells) and *Gapd* (expressed ubiquitously) were used as controls.

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 *Acsl3*, whose function was preserved despite the enlistment of its promoter and 5′ untranslated exons in the service of the newly created gene *Utp14b*.

Other testis-expressed autosomal retrogenes with X-linked progenitors have been reported in mammals[4,12–23]. 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. *Zfa*, an autosomal retrogene derived from the X-chromosomal gene *Zfx*[18,19], is the only other protein-coding retrogene in which a loss-of-function mutation has been

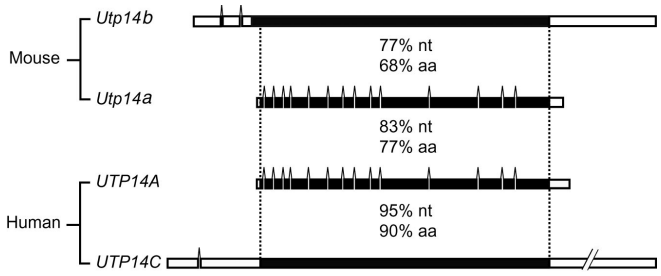

Figure 3 Schematic alignment of transcripts from mouse *Utp14b*, its X-linked homologs *Utp14a* and *UTP14A* in mouse and human, respectively, and its human autosomal homolog *UTP14C*. Coding regions are shown in black, and positions of introns are indicated. Percentage identities (nucleotide (nt) and amino acid (aa)) between coding sequences are given.

reported. Ablation of *Zfa* yielded no detectable phenotype[24]. To our knowledge, mouse *Utp14b* is also the first functional mammalian retrogene found to have parasitized an existing transcription unit.

Our findings substantiate the hypothesis[4] that mammalian spermatogenesis is dependent on autosomal retrogenes that evolved from X-linked progenitors to compensate for X-chromosome silencing during male meiosis. This hypothesis leads to three predictions regarding the *in vivo* function of X-to-autosome retrogenes, all of which are fulfilled by *Utp14b*. First, loss of retrogene function should perturb gametogenesis in males without otherwise affecting health or development in either sex. An early frameshift mutation in the *Utp14b* retrogene completely disrupts spermatogenesis. Second, retrogene function should be required only in the germ cells themselves, and not in the somatic cells of the testes. This is the case with *Utp14b*[8,9]. Third, retrogene function should be required only during meiosis; any spermatogenic defect should manifest at or after meiosis. As *jsd/jsd* XY mice age, premeiotic cell (spermatogonial) numbers and differentiation deteriorate, but this seems to be a secondary consequence of an altered hormonal environment in the testis, and not a direct result of the primary defect in the spermatogenic lineage[25]. The earliest defect observed in *jsd/jsd* deficient males is a marked reduction in the numbers of postmeiotic cells, or spermatids, at three weeks of age, during the first round of spermatogenesis[26]. This is consistent with the gene carrying the *jsd* mutation being required primarily, or exclusively, in meiotic cells.

Utp14a, the X-linked progenitor of *Utp14b*, encodes the mammalian ortholog of yeast Utp14, an essential protein required for processing of pre-rRNA (**Supplementary Fig. 3** online)[3]. X-to-autosome retrogenes are thought to serve essential housekeeping functions[4]. We postulate that this is the case for *Utp14b*, and specifically that it supports ribosome assembly and hence protein synthesis during male meiosis.

Utp14b exists in mice but not in humans. Might a similar but distinct functional retrogene exist in humans? This question is motivated by the facts that the X chromosome is inactivated during male meiosis in both humans and mice[5–7] and *Utp14a* is X-linked in both species. Electronic examination of the human genome identified a testis-transcribed gene, on chromosome 13, that shares the following properties with mouse *Utp14b*: an unmistakable similarity to the coding sequence of *Utp14a*, an intact and conserved ORF and an absence of coding-region introns (**Fig. 3** and ref. 27). This human retrogene, *UTP14C*, is not orthologous to mouse *Utp14b*. As indicated by comparison of mouse and human map locations (**Supplementary Fig. 4** online) and by phylogenetic sequence analysis (**Fig. 4**), the mouse *Utp14b* and human *UTP14C* retrogenes are products of two independent retroposition events.

We next searched for intronless derivatives of *Utp14a* 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 *Utp14* genes. We identified a *Utp14a*-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 *Utp14a* 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 preponderance of silent nucleotide substitutions when compared with mouse or human *Utp14a*. These findings suggest that most eutherians possess a *Utp14a*-derived retrogene that encodes a functional protein subject to purifying selection. We did not find *Utp14a*-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 *Utp14a*-derived retrogenes to determine whether they could all be accounted for by the two retroposition events that gave rise to mouse *Utp14b* 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 *Utp14a*-derived retrogenes observed in eutheria (**Fig. 4** and **Supplementary Fig. 5** online). Thus, the retrogenesis and fixation of *Utp14b* during rodent evolution was reproduced,

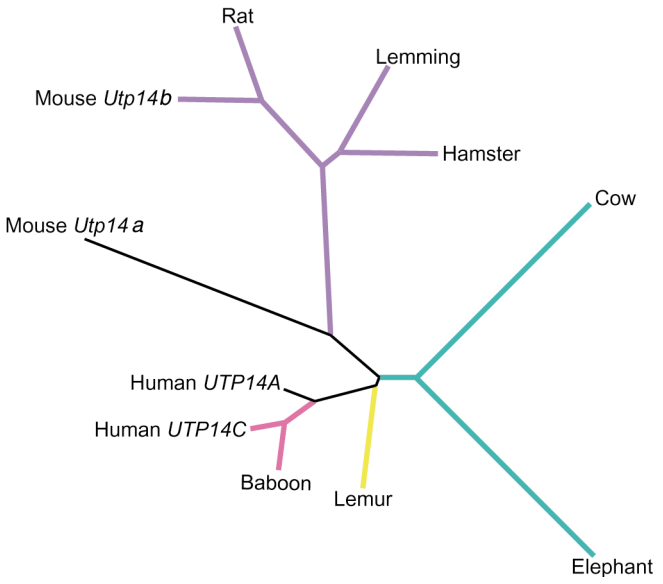

Figure 4 Phylogenetic comparison of *Utp14b* homologs in diverse eutherian mammals indicates that a minimum of four retroposition events (indicated by blue, red, yellow and green branches) occurred. This unrooted tree depicts phylogenetic relationships between the X-linked, intron-bearing *Utp14a* orthologs 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 retrogenes, 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 retroposed copies of protein-coding genes identified to date in mammalian genomes are functionless pseudogenes[28,29]. We found no *Utp14a* pseudogenes in either mouse or human. Ironically, we did find a processed pseudogene of *Utp14b* in the mouse (**Supplementary Fig. 5** online). The fixation and maintenance of *Utp14a*-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 *jsd*/+ male mice were a gift from W. Beamer (The Jackson Laboratory). We purchased C57BL/6J, DBA/2J, W^v/W^v and *Mus musculus castaneus*/Ei mice from The Jackson Laboratory.

Genetic mapping. We mated C57BL/6J *jsd*/*jsd* females with DBA/2J 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 *jsd*; see **Supplementary Table 1** for markers, PCR primers and assay conditions. We determined the *jsd* genotypes of critical recombinants (**Supplementary Fig. 1** online) by progeny testing, crossing to *jsd*/*jsd* or *jsd*/+ mice. We phenotyped adult males by dissecting out testes, which in *jsd*/*jsd* mice weighed no more than one-third as much as the testes of *jsd*/+ and +/+ 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 RPCI-22 and RPCI-23 libraries to construct a BAC contig of the *jsd* region of mouse chromosome 1. Two BACs spanning the critical *jsd* region, RPCI-22-292L24 and RPCI-23-395H12, were sequenced at the Whitehead Institute/MIT Center for Genome Research as described[30]. 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 *jsd*/*jsd* mice as starting material. We selected PCR primers using Primer 3. 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 *jsd*/*jsd* 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 5′ and 3′ regions of ORF2.3. We subcloned 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, pUtp14b3.5, included 5′ untranslated sequence and is depicted in **Figure 1**.

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

For RT-PCR, we used 1–5 μg of poly(A)$^+$ RNA in an oligo d(T)-primed first-strand cDNA synthesis reaction with Stratascript 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 *Utp14b*, mouse and human *UTP14A* and human *UTP14C*. We tested all primer pairs using mouse and human genomic DNAs as templates. We then applied satisfactory primer pairs to genomic DNAs from gorilla, chimpanzee, baboon, orangutan, rhesus monkey, squirrel monkey, lemur, 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 PHYLIP software (version 3.5c).

URLs. RepeatMasker is available at http://www.repeatmasker.org/. GENSCAN is available at http://genes.mit.edu/GENSCAN.html. BLAST search is available at http://www.ncbi.nlm.nih.gov/BLAST/. Primer 3 is available at http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi/. PHYLIP software is available at http://evolution.genetics.washington.edu/phylip.html. The sequence of the *jsd* critical region is available at http://jura.wi.mit.edu/page.

GenBank accession numbers. *jsd* critical region BACs: RPCI-22-292L24, AC079222; RPCI-23-395H12, AC079134. Mouse *Utp14b* cDNAs, AY316161 and AY316162; mouse *Utp14a* cDNA, AY316163; human *UTP14A* cDNA, BC001149; human *UTP14C* cDNA (KIAA0266), D87455. Genomic sequence of *Utp14b*-like retrogenes: baboon, AY316164; chimpanzee, AY316165; cow, AY316166; elephant, AY316167; gorilla, AY316168; guinea pig, AY316169; hamster, AY316170; lemming, AY316171; lemur, AY316172; orangutan, AY316173; rat, AY316174; rhesus monkey, AY316175; squirrel monkey, AY316176. Genomic sequence of *Utp14b*-derived pseudogene in mouse, AY641472.

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

ACKNOWLEDGMENTS

We thank W. Beamer for providing C57BL/6J *jsd*/+ mice; the Genome Sequencing group at the Whitehead/MIT Center for Genome Research for BAC sequencing; B. Birren and E.S. Lander for support; and A. Bortvin, J. Alfoldi, J. Koubova, J. Lange, J. Potash, J. Saionz, J. Wang, K. Kleene and S. Rozen 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/

1. Beamer, W.G., Cunliffe-Beamer, T.L., Shultz, K.L., Langley, S.H. & Roderick, T.H. *Juvenile spermatogonial depletion* (*jsd*): a genetic defect of germ cell proliferation of male mice. *Biol. Reprod.* **38**, 899–908 (1988).
2. Boettger-Tong, H. *et al.* Identification and sequencing the juvenile spermatogonial depletion critical interval on mouse chromosome 1 reveals the presence of eight candidate genes. *Biochem. Biophys. Res. Commun.* **288**, 1129–1135 (2001).
3. Dragon, F. *et al.* A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).
4. McCarrey, J.R. & Thomas, K. Human testis-specific *PGK* gene lacks introns and possesses characteristics of a processed gene. *Nature* **326**, 501–505 (1987).
5. Handel, M.A., Park, C. & Kot, M. Genetic control of sex-chromosome inactivation during male meiosis. *Cytogenet. Cell. Genet.* **66**, 83–88 (1994).
6. Solari, A.J. The behavior of the XY pair in mammals. *Int. Rev. Cytol.* **38**, 273–317 (1974).
7. Richler, C. *et al.* Splicing components are excluded from the transcriptionally inactive XY body in male meiotic nuclei. *Mol. Biol. Cell* **5**, 1341–1352 (1994).
8. Boettger-Tong, H.L., Johnston, D.S., Russell, L.D., Griswold, M.D. & Bishop, C.E. *Juvenile spermatogonial depletion* (*jsd*) mutant seminiferous tubules are capable of supporting transplanted spermatogenesis. *Biol. Reprod.* **63**, 1185–1191 (2000).
9. Ohta, H. *et al.* Defect in germ cells, not in supporting cells, is the cause of male infertility in the *jsd* mutant mouse: proliferation of spermatogonial stem cells without dif-

ferentiation. *Int. J. Androl.* **24**, 15–23 (2001).
10. Chen, H.H., Liu, T.Y., Li, H. & Choo, K.B. Use of a common promoter by two juxtaposed and intronless mouse early embryonic genes, *Rnf33* and *Rnf35*: implications in zygotic gene expression. *Genomics* **80**, 140–143 (2002).
11. Scanlan, M.J. *et al.* Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression. *Cancer Immun.* **1**, 4 (2001).
12. Dahl, H.-H.M., Brown, R.M., Hutchison, W.M., Maragos, C. & Brown, G.K. A testis-specific form of the human pyruvate dehydrogenase E1α subunit is coded for by an intronless gene on chromosome 4. *Genomics* **8**, 225–232 (1990).
13. Hendriksen, P.J.M. *et al.* Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. *Genomics* **41**, 350–359 (1997).
14. Dass, B. *et al.* The gene for a variant form of the polyadenylation protein CstF-64 is on chromosome 19 and is expressed in pachytene spermatocytes in mice. *J. Biol. Chem.* **276**, 8044–8050 (2001).
15. Sargent, C.A., Young, C., Marsh, S., Ferguson-Smith, M.A. & Affara, N.A. The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. *Hum. Mol. Genet.* **3**, 1317–1324 (1994).
16. Elliott, D.J. *et al.* An evolutionarily conserved germ cell-specific *hnRNP* is encoded by a retrotransposed gene. *Hum. Mol. Genet.* **9**, 2117–2124 (2000).
17. Sedlacek, Z. *et al.* 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 (1999).
18. Mardon, G. *et al.* Mouse *Zfx* protein is similar to *Zfy-2*: each contains an acidic activating domain and 13 zinc fingers. *Mol. Cell. Biol.* **10**, 681–688 (1990).
19. Ashworth, A., Skene, B., Swift, S. & Lovell-Badge, R. *Zfa* is an expressed retroposon derived from an alternative transcript of the *Zfx* gene. *EMBO J.* **9**, 1529–1534 (1990).
20. Boer, P.H., Adra, C.N., Lau, Y.F. & McBurney, M.W. The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon. *Mol. Cell. Biol.* **7**, 3107–3112 (1987).
21. Halford, S. *et al.* Characterization of a novel human opsin gene with wide tissue expression and identification of embedded and flanking genes on chromosome 1q43. *Genomics* **72**, 203–208 (2001).
22. Uechi, T., Maeda, N., Tanaka, T. & Kenmochi, N. Functional second genes generated by retrotransposition of the X-linked ribosomal protein genes. *Nucleic Acids Res.* **30**, 5369–5375 (2002).
23. Cremers, F.P. *et al.* An autosomal homologue of the choroideremia gene colocalizes with the Usher syndrome type II locus on the distal part of chromosome 1q. *Hum. Mol. Genet.* **1**, 71–75 (1992).
24. Banks, K.G. *et al.* Retroposon compensatory mechanism hypothesis not supported: *Zfa* knockout mice are fertile. *Genomics* **82**, 254–260 (2003).
25. Tohda, A. *et al.* Testosterone suppresses spermatogenesis in *juvenile spermatogonial depletion* (*jsd*) mice. *Biol. Reprod.* **65**, 532–537 (2001).
26. Kojima, Y. *et al.* Cessation of spermatogenesis in *juvenile spermatogonial depletion* (*jsd/jsd*) mice. *Int. J. Urol.* **4**, 500–507 (1997).
27. Nagase, T. *et al.* Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res.* **3**, 321-329, 341–354 (1996).
28. Vanin, E.F. Processed pseudogenes. Characteristics and evolution. *Biochim. Biophys. Acta* **782**, 231–241 (1984).
29. Zhang, Z., Harrison, P.M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13**, 2541–2558 (2003).
30. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).