Human Molecular Genetics, 2002, Vol. 11, No. 19 2341–2346

Functional substitution for $TAF_{II}250$ by a retroposed homolog that is expressed in human spermatogenesis

P. Jeremy Wang and David C. Page*

Howard Hughes Medical Institute, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA

Received May 28, 2002; Revised and Accepted July 15, 2002 DDBJ/EMBL/GenBank accession nos†

$TAF_{II}250$, the largest subunit of the general transcription factor TFIID, is expressed from the human X chromosome, at least in somatic cells. In male meiosis, however, the sex chromosomes are transcriptionally silenced, while the autosomes remain active. How then are protein-encoding genes transcribed during human male meiosis? Here we present a novel autosomal human gene, *TAF1L*, which is homologous to *$TAF_{II}250$* and is expressed specifically in the testis, apparently in germ cells. We hypothesize that during male meiosis, transcription of protein-encoding genes relies upon TAF1L as a functional substitute for $TAF_{II}250$. Like $TAF_{II}250$, the human TAF1L protein can bind directly to TATA-binding protein, an essential component of TFIID. Most importantly, transfection with human *TAF1L* rescued the temperature-sensitive lethality of a hamster cell line mutant in *$TAF_{II}250$*. *TAF1L* lacks introns and evidently arose by retroposition of a processed *$TAF_{II}250$* mRNA during primate evolution. The observation that TAF1L can functionally replace $TAF_{II}250$ provides experimental support for the hypothesis that during male meiosis, autosomes provide cellular functions usually supplied by the X chromosome in somatic cells.

INTRODUCTION

In mammalian spermatocytes (male meiotic cells), the activities of autosomal and sex-linked genes differ starkly. The autosomes are transcriptionally active. In contrast, the entireties of the X and Y chromosomes are condensed to form heterochromatin and are segregated into a special nuclear compartment, the 'XYbody', where Pol II (RNA polymerase II) is absent, as are critical splicing factors. There appears to be no transcription within this sex chromosome compartment (1–4). (This phenomenon is distinct from somatic 'X inactivation', which is the silencing of most but not all genes on one of the two X chromosomes in female cells.) In humans, sex chromosome silencing during male meiosis lasts roughly 15 days, during which time ongoing transcription of protein-encoding genes by Pol II is likely required (5). The largest of the essential factors engaged in Pol II transcription is $TAF_{II}250$, which, paradoxically, is encoded by the mammalian X chromosome (6–8). Since the *$TAF_{II}250$* gene is transcriptionally inactive, perhaps another protein acts as a substitute in male meiosis. Specifically, we wondered whether the autosomes encode a $TAF_{II}250$ isoform that enables or modulates Pol II transcription during male meiosis.

RESULTS

TAF1L is a retroposed homolog of *$TAF_{II}250$*

We examined the human genome, where the availability of extensive draft sequence facilitated an electronic search for $TAF_{II}250$ homologs (9,10). The *$TAF_{II}250$* gene (also known as *TAF1* or *CCG1*) spans 100 kb on the human X chromosome (6–8). Within the genome, we identified only one homologous locus, which we named *TAF1L*. *TAF1L* is autosomal, residing on chromosome 9. The *TAF1L* genomic locus exhibits >94% nucleotide identity to *$TAF_{II}250$*'s 38 exons, but it completely lacks the latter's 37 introns (Fig. 1) and thus spans only 6.5 kb. This suggested that *TAF1L* is a retroposed derivative of *$TAF_{II}250$*.

In the human genome, retroposed copies of genes are commonplace, but most are transcriptionally silent pseudogenes whose protein-coding regions have been disrupted by mutation (11). In contrast, we found that *TAF1L* is transcribed (Fig. 2) and that its protein-coding potential appeared intact. We confirmed this by sequencing *TAF1L* cDNA clones isolated from a human testis library. The *TAF1L* cDNA and genomic sequences proved to be co-linear, and the coding region is 96%

*To whom correspondence should be addressed. Tel: +1 6172585203; Fax: +1 6172585578; Email: page_admin@wi.mit.edu
†AF390562–AF390568 and G73370–G73375

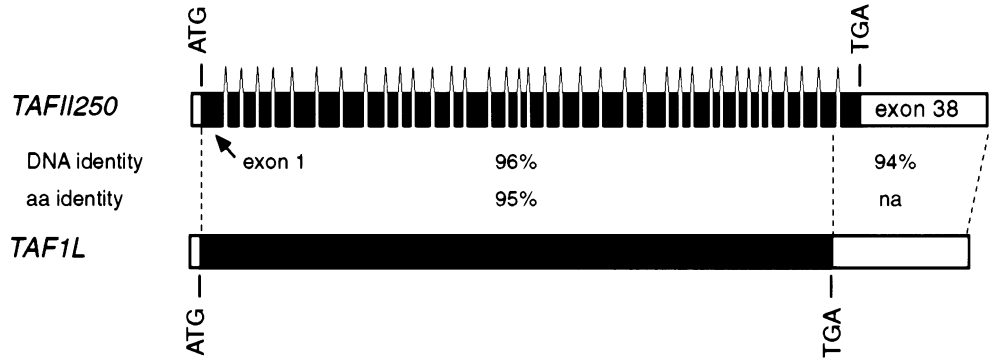


Figure 1. Comparison of exon/intron structures of human *TAF$_{II}$250* and *TAF1L* genes. Exons 1 and 38 in *TAF$_{II}$250* are labeled; there appear to be no introns in *TAF1L*. Coding regions are shown in black. Introns are not drawn to scale. Percentage identities in coding and 3′ untranslated regions (UTR) for both nucleotide and (where applicable) predicted amino acid sequences are indicated; na, not applicable. 5′-UTR display 91% nucleotide identity. In comparison with *TAF$_{II}$250*, the *TAF1L* coding region displays two small deletions (3 bp each) and three insertions. The first two insertions add 29 residues to the predicted TAF1L protein but do not shift the reading frame. The third insertion (10 bp), near the 3′ end, shifts the reading frame and thus the position of the termination codon, TGA, as shown. Note that we extended the published *TAF$_{II}$250* cDNA sequence (8) to include the 3′-UTR and AATAAA polyadenylation signal by electronically assembling overlapping expressed sequence tags (GenBank accession nos AU123960, AW959904, W05157 and W55984).

identical to that found in long *TAF$_{II}$250* transcripts (Fig. 1). [Alternative splice acceptor sites in exon 5 of *TAF$_{II}$250* yield two distinct transcripts, one being 63 nucleotides longer than the other (6).] We concluded that *TAF1L* was created by reverse transcription and subsequent autosomal integration of a properly spliced *TAF$_{II}$250* transcript. The result is a single-exon gene that encodes an 1826-residue protein with 95% amino acid identity to human TAF$_{II}$250 (Fig. 1). Interestingly, there is no apparent remnant of a poly(A) tail in the *TAF1L* gene, as is observed in many but not all retroposons.

TAF1L displays testis-specific expression

If TAF1L assumes the role of TAF$_{II}$250 in human spermatocytes, *TAF1L* should be transcribed in testes. Indeed, this is the only human organ in which we have detected *TAF1L* expression (Fig. 2A). More specifically, our model predicts that *TAF1L* should be expressed in testicular germ cells. Our data suggest that *TAF1L* expression may be restricted to such cells: *TAF1L* is transcribed in normal testis but not in testis lacking germ cells (Fig. 2B). (Alternatively, it is formally possible that *TAF1L* is expressed in the somatic protein of the testis in a germ-cell-dependent manner.) By contrast, the X-linked *TAF$_{II}$250* gene is expressed widely in human somatic tissues, including germ-cell-deficient testis (Fig. 2.).

Ideally, we would test our assumption that the X-linked *TAF$_{II}$250* gene is transcriptionally silenced in male meiotic cells by RNA *in situ* hybridization on human testis sections. Unfortunately, this experiment is not feasible because of the high nucleotide sequence identity between *TAF$_{II}$250* and *TAF1L* (96% in the coding region and 94% in the 3′-UTR; Fig. 1). RNA probes of sufficient length to detect these modest-abundance transcripts would likely lack the ability to discriminate between the two genes' transcripts.

TAF1L interacts with human TBP

We then tested whether the TAF1L and TAF$_{II}$250 proteins share functional characteristics. TAF$_{II}$250 has been shown to bind directly to TATA-binding protein (TBP) (6,7). To test whether TAF1L protein interacts with TBP, we performed a two-hybrid analysis in yeast cells. We fused the TAF1L protein to the DNA-binding domain of GAL4, and separately we fused human TBP to the activation domain of GAL4. Neither of the two fusion proteins alone activated a GAL4-responsive *lacZ* reporter gene. In combination, however, the two fusion proteins strongly activated the reporter gene (Fig. 3), indicating a direct interaction, in eukaryotic cells, of TAF1L and TBP. This supports the hypothesis that TAF1L functions as a TBP-associated factor (TAF).

TAF1L rescues the temperature-sensitive lethality of *TAF$_{II}$250* mutant cells

We then tested the ability of the human TAF1L protein to functionally replace TAF$_{II}$250 in mammalian cells. We took advantage of a hamster cell line, ts13, which is temperature-sensitive because of a single amino acid substitution (G690D) in TAF$_{II}$250 (12). When shifted to 39.5°C, ts13 cells arrest in the G_1 phase of the cell cycle and undergo apoptosis (13,14). The human *TAF$_{II}$250* gene had been cloned by transfecting ts13 cells with human genomic DNA and selecting for growth at 39.5°C (13). The human *TAF1L* gene that we report was not identified in this transfection screen, suggesting that either (i) the TAF1L protein is not functionally equivalent to TAF$_{II}$250 or (ii) the *TAF1L* gene's testis-specific promoter is inactive in ts13 cells, which derive from hamster kidney (8).

To distinguish between these two possibilities, we transfected ts13 cells at the permissive temperature (33.5°C) with a DNA construct in which a ubiquitously expressed viral promoter (cytomegalovirus, CMV) drove expression of human TAF1L coding sequences. We assayed the rescue capability of this pCMN–TAF1L construct by selection at 39.5°C. As controls, we conducted parallel experiments with pCMV alone, and also with a construct bearing a single amino acid substitution (G714D) in TAF1L—a mutation analogous to that in TAF$_{II}$250 in ts13 cells. At 33.5°C, transfections of ts13 with experimental and control constructs generated neomycin-resistant colonies at

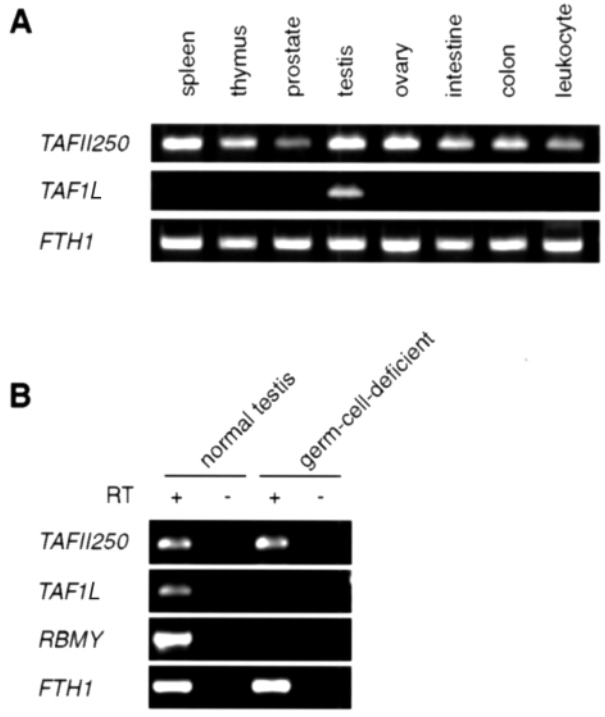


Figure 2. Expression of *TAF$_{II}$250* and *TAF1L* in human tissues assayed by RT–PCR. (**A**) Normal human tissues. (**B**) Normal and germ-cell-deficient human testis biopsies. *FTH1* (ferritin heavy chain; ubiquitously expressed) and *RBMY* (RNA-binding-motif protein; expressed only in male germ cells) served as controls (30). RT+ or RT− indicates presence or absence of reverse transcriptase in the reaction.

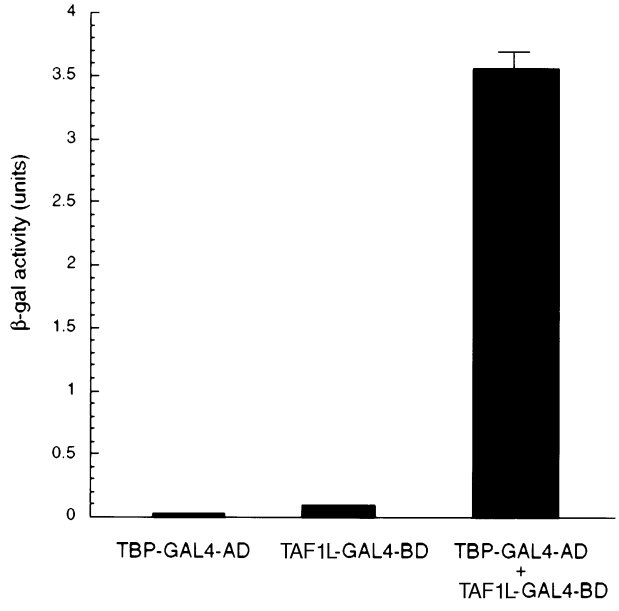


Figure 3. Human TAF1L protein interacts with human TBP protein in a yeast two-hybrid assay. Quantitative assays for β-galactosidase activity were performed with liquid cultures of yeast cells transformed with either TBP–GAL4-AD, or TAF1L–GAL4-BD or both. ONPG (*o*-nitrophenyl β-D-galactopyranoside; Sigma) was used as substrate. GAL4-AD, GAL4 activation domain; GAL4-BD, GAL4 DNA-binding domain.

similar efficiencies (Fig. 4B). At 39.5°C, neither the pCMV vector alone nor the mutant *TAF1L* construct yielded any viable cells, but wild-type *TAF1L* transfectants were viable and readily formed colonies (Fig. 4). By western blotting, we confirmed the presence of wild-type and mutant TAF1L proteins in their respective transfectants (Fig. 4C), indicating that the inability of the mutant TAF1L protein to complement the ts13 mutation is not due to protein degradation. Taken together, our results demonstrate that the human TAF1L protein is functionally interchangeable with TAF$_{II}$250 in mammalian cells.

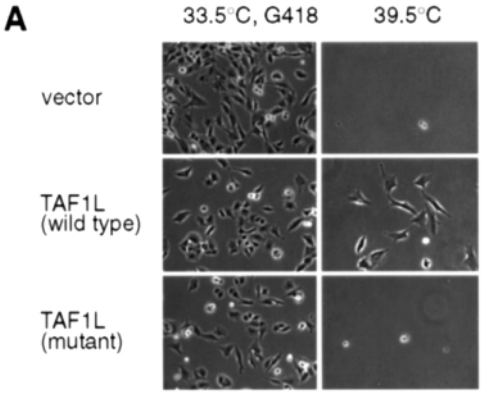
TAF1L retroposition occurred during primate evolution

To determine when in human evolution the *TAF1L* retroposition event occurred, we searched for *TAF1L* orthologs in the genomes of other mammals. We employed PCR primers that would likely amplify *TAF1L* but not *TAF$_{II}$250* genomic sequences. We found that *TAF1L* is present in apes and Old World monkeys but is absent in New World monkeys and rodents (Fig. 5A). We sequenced *TAF1L* orthologs in six apes and Old World monkeys (chimpanzee, gorilla, orangutan, gibbon, baboon and macaque), and found that the 5.5 kb open reading frame is conserved and intact in all six species, in each case displaying >97% nucleotide identity and >95% predicted protein sequence identity to human TAF1L. Analysis of rates of non-synonymous substitution (K_a) and synonymous substitution (K_s) in TAF1L orthologs revealed that the K_a/K_s ratio is consistently <1 (Fig. 5B), indicating that the TAF1L protein has been subject to functional constraint and purifying selection. Taken together, these results suggest two conclusions. First, the *TAF1L* retroposition probably occurred ~25–40 million years ago, prior to the radiation of extant Old World monkey lineages, but after their divergence from New World monkeys (15). Second, the functional integrity of the TAF1L protein apparently has been conserved among diverse Old World monkeys and apes, including humans.

DISCUSSION

In this study, we identified an autosomal retrogene that encodes a functional substitute for the product of its X-linked progenitor. The autosomal gene, *TAF1L*, appears to be expressed only in spermatogenic cells, whereas the X-linked homolog, *TAF$_{II}$250*, is widely expressed. Several other examples of active, testis-specific autosomal retrogenes with X-linked progenitors have been identified in human and/or mouse (16–24). In each known case, the autosomal retrogene is specifically expressed in testis, while the X-linked source gene is widely expressed. However, apart from TAF$_{II}$250 and TAF1L, functional comparisons of X- and autosome-encoded isoforms have yet to be conducted.

Our study raises questions about evolutionary pressures that may have favored the functional substitution of essential somatic factors by gametogenesis-specific homologs. It has been hypothesized that autosomal, testis-specific retrogenes evolved to compensate for inactivation of their X-linked source genes during male meiosis (16). Our discovery of the



B

cDNA clone transfected	number of colonies	
	33.5°C, G418	39.5°C
vector alone	268	0
TAF1L (wild type)	176	58
TAF1L (mutant)	188	0

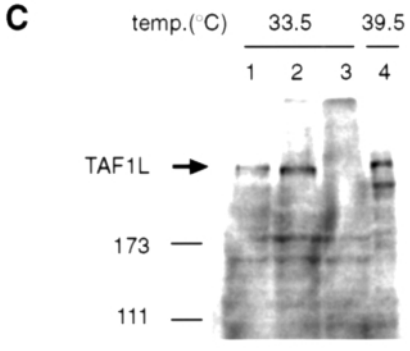
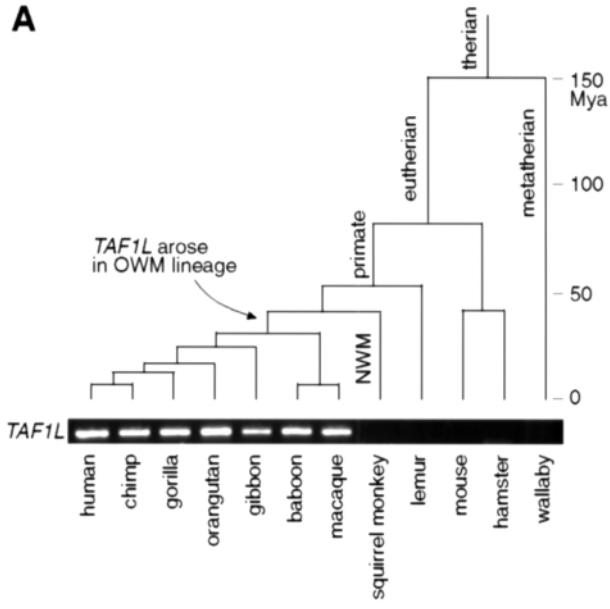


Figure 4. Rescue of temperature sensitivity of *TAF$_{II}$250*-mutant hamster cells (ts13) by human *TAF1L*. Following transfection with vector alone, or vector expressing wild-type human TAF1L, or vector expressing mutant human TAF1L (G714D), cells were split equally onto two plates: one incubated at the permissive temperature (33.5°C) with 400 µg/ml neomycin and one incubated at the non-permissive temperature (39.5°C) without neomycin. (**A**) Normarski images of cells 6 days after transfection. In control experiments with no DNA transfected, all cells died (not shown). (**B**) Numbers of colonies present 10 days after transfection. (**C**) Western blot of hemagglutinin (HA)-epitope-tagged TAF1L proteins expressed in transfected cells. Protein extracts from transfected cells grown at 33.5°C or 39.5°C were run on a 7.5% SDS–polyacrylamide gel. Mouse monoclonal antibody 12CA5 (Roche) was used to recognize the HA epitope. Lanes 1, 2 and 3: neomycin-resistant cells populations transfected with HA3-TAF1L (lane 1), mutant HA3-TAF1L (G714D) (lane 2) or pCMV–tag2 vector (Stratagene) (lane 3). Lane 4: HA3-TAF1L stable cell line isolated from a single colony at 39.5°C. The positions of TAF1L and molecular weight standards (in kDa) are indicated on the left.



B

	chimp	gorilla	orang	gibbon	baboon	macaque
human	0.52	0.50	0.52	0.52	0.41	0.43
chimp		0.69	0.61	0.60	0.45	0.48
gorilla			0.63	0.57	0.43	0.47
orang				0.48	0.38	0.39
gibbon					0.42	0.43
baboon						0.54

Figure 5. Orthologs of human *TAF1L* in apes and Old World monkeys. (**A**) *TAF1L* orthologs were sought by PCR using genomic DNA from 11 mammalian species, arrayed here phylogenetically (15). The 257 bp PCR products shown were obtained using primers selected from human *TAF1L* in regions of complete nucleotide sequence identity to human, mouse and hamster *TAF$_{II}$250* cDNAs. In the mammalian *TAF$_{II}$250* genomic loci, these conserved primers are interrupted by introns, and thus no amplification occurs. The results shown here were confirmed with two additional pairs of PCR primers (see Materials and Methods). NWM, New World monkeys; OWM, Old World monkeys; Mya, million years ago. (**B**) K_a/K_s ratios calculated from pairwise comparisons of *TAF1L* orthologs in seven species. *TAF1L* coding sequences were aligned using CLUSTAL W (31), and K_a and K_s values were calculated using Li's method as implemented in GCG software (32).

functional interchangeability of X-encoded TAF$_{II}$250 and autosome-encoded TAF1L is an important step in validating this hypothesis.

Human TAF$_{II}$250 joins a small group of widely expressed TFIID components for which tissue-selective homologs have been functionally characterized. In mice, a homolog of TAF$_{II}$130 is essential for ovarian follicle development (25). Spermiogenesis in mice requires TRF2, a TBP homolog (26), while spermatogenesis in *Drosophila* requires a testis-specific homolog of dTAF$_{II}$80 (27). To date, all functionally characterized, tissue-selective homologs of TFIID components appear to be engaged in gametogenesis. Among these examples, the case of human TAF$_{II}$250 is unique in that we are cognizant of specific adaptive pressures (stemming from X inactivation during male meiosis) that may have driven the evolution of the tissue-specific homolog, TAF1L.

Autosomal retropositioning of X-linked genes appears to have been a recurrent, if infrequent, event throughout mammalian evolution. Among the few molecularly documented events, the most ancient involves *PGK2*, which arose via retroposition at

least 130 million years ago, prior to the divergence of the marsupial and eutherian mammalian lineages (28). More recent events involving *PHDA2*, *G6pd2* and *Zfa* have been reported (18–20,29). Our present findings indicate that *TAF1L* arose during primate evolution, after our species' lineage diverged from that of New World monkeys (Fig. 5A). The coming availability of complete DNA sequences of the human, mouse and other mammalian genomes will provide unprecedented opportunity to assess systematically the history and extent of retroposition of all X-linked genes during mammalian evolution.

MATERIALS AND METHODS

cDNA sequencing by PCR screening of library subpools

Lambda phage lysates were prepared from 24 subpools (~80 000 clones each) of an amplified human testis cDNA library (Clontech). *TAF1L*-positive subpools were identified by PCR using primer pairs selected from the human *TAF1L* genomic locus (GenBank accession AL355811). Eleven *TAF1L*-positive subpools were identified. We then amplified 5′ and 3′ cDNA fragments separately from the positive subpools by two successive PCRs. In the first PCR (35 cycles), one *TAF1L*-specific primer and one vector primer were used. A 1 : 100 dilution of the first PCR product served as template in the second, nested PCR (25 cycles), where a second *TAF1L*-specific primer and a second vector primer were used to further amplify the cDNA fragment. The resulting PCR products were sequenced directly using the nested primers. Using these cDNA fragment sequences, a composite full-length cDNA sequence of *TAF1L* was assembled electronically.

Expression analysis by RT–PCR

Normal human tissue cDNA samples were purchased from Clontech. Human normal testis and germ-cell-deficient testis (Sertoli-cell-only-syndrome) biopsy samples were kindly provided by S. Silber (St Louis, MO). To construct bulk cDNA from testis biopsies, total RNA prepared using TRIzol reagent (Gibco BRL) was treated with DNase I to digest residual genomic DNA, and reverse-transcribed. Gene-specific PCR primers were designed to avoid cross-amplification between $TAF_{II}250$ and *TAF1L*. PCR conditions and primer sequences have been deposited at GenBank.

Yeast two-hybrid assay

Using the vector pAS2-1 (Clontech), the entire human TAF1L coding region was cloned in frame with the GAL4 DNA-binding domain, generating the construct TAF1L–GAL4-BD. The entire human TBP coding region was PCR-amplified from bulk human testis cDNA and fused with the GAL4 activation domain in the vector pACT2 (Clontech), generating the construct TBP–GAL4-AD. The human inserts of both resulting constructs were sequenced to confirm their integrity. Plasmids were transformed into yeast strain Y187 ($GAL1_{UAS}$–$GAL1_{TATA}$–*lacZ*). Quantitative liquid culture assays for β-galactosidase activity were performed as described by the manufacturer (Clontech).

Transfection constructs

The entire human *TAF1L* coding region was cloned into the pCMV–Tag2A vector (Stratagene), resulting in pCMV–TAF1L.

A second construct, pCMV–TAF1L (G714D), was derived from pCMV–TAF1L by oligonucleotide-directed mutagenesis. The following four primers were used: primer 1, CATCT CTGAAGAAGAGTCGAA; primer 2, TCTTGGTTG CCATGTCAACCTGCATCATTA; primer 3, TAATGATG CAGGTTGACATGGCAACCAAGA; and primer 4, AGGTTGTTCTCAAGTGCCTG. Primers 1 and 4 are outer primers. Primers 2 and 3 are inner primers; these two primers are complementary to each other and contain the desired nucleotide change (underlined). Two DNA fragments (501 and 172 bp) were amplified by PCR with primers 1 and 2 or primers 3 and 4, respectively. The overlapping PCR products were mixed and subjected to PCR amplification with primers 1 and 4. The resulting fragment (643 bp) harbors the desired G-to-A substitution at nucleotide 2141 of the *TAF1L* coding region. Restriction digestion yielded an *Eco*RI–*Pst*I fragment that was used to replace the corresponding portion of pCMV–TAF1L, resulting in pCMV–TAF1L (G714D).

Influenza hemagglutinin (HA) epitope-tagged versions of pCMV–TAF1L and pCMV–TAF1L (G714D) were then derived. DNA encoding three tandem copies of HA was PCR-amplified from pCU180 using primers GGAAGA TCTTTACCCATACGATGTTCCT and GGAAGATC TTGAGCAGCGTAATCTGGA. The amplified DNA was digested with *Bgl*II and ligated into *Bam*HI-restricted pCMV–TAF1L and pCMV–TAF1L (G714D), generating the constructs pCMV–HA3–TAF1L and pCMV–HA3–TAF1L (G714D), respectively.

Transfection of ts13 cells

Prior to transfection, hamster ts13 cells were cultured at 33.5°C in 10% CO_2 in Dulbecco's modified Eagle's medium with 10% fetal calf serum and penicillin–streptomycin (8). All transfections were performed using LIPOFECTIN reagent (Life Technologies), and in each case 3 μg plasmid DNA was used as suggested by the manufacturer. Two days after transfections, cells were split 1 : 12 and subjected to selection with neomycin or at 39.5°C.

TAF1L orthologs

To identify *TAF1L* orthologs, we performed PCR on genomic DNAs from diverse mammals using three different pairs of primers. PCR conditions and primer sequences have been deposited at GenBank: accession nos G73370 (shown in Fig. 5A), G73371 (not shown) and G73372 (not shown).

GenBank accession numbers

Primer sequences and RT–PCR conditions for human genes: $TAF_{II}250$, G73373; *TAF1L*, G73374; *FTH1*, G65764; and

RBMY, G73375. *TAF1L* sequences: human cDNA, AF390562; chimp genomic, AF390563; gorilla genomic, AF390564; orangutan genomic, AF390565; gibbon genomic, AF390566; baboon genomic, AF390567; and macaque genomic, AF390568.

ACKNOWLEDGEMENTS

We thank S. Silber for testis biopsy samples, R. Tjian for ts13 cells, T. Huffaker for the HA3 plasmid, Yerkes Primate Center for chimpanzee, gorilla, orangutan and gibbon DNA samples, Joerg Gromoll for a macaque DNA sample, K. Rice for a baboon DNA sample, N. Watson for advice on microscopy, and A. Arango, A. Baltus, J. Bradley, L. Brown, K. Kleene, F. Lewitter, D. Menke, J. Potash, S. Rozen and H. Skaletsky for comments on the manuscript. P.J.W. is the recipient of a Medical Foundation postdoctoral fellowship.

REFERENCES

1. Handel, M.A., Park, C. and Kot, M. (1994) Genetic control of sex-chromosome inactivation during male meiosis. *Cytogenet. Cell Genet.*, **66**, 83–88.
2. Solari, A.J. (1974). The behaviour of the XY pair in mammals. *Int. Rev. Cytol.*, **38**, 273–317.
3. Ayoub, N., Richler, C. and Wahrman, J. (1997) Xist RNA is associated with the transcriptionally inactive XY body in mammalian male meiosis. *Chromosoma*, **106**, 1–10.
4. Richler, C., Ast, G., Goitein, R., Wahrman, J., Sperling, R. and Sperling, J. (1994) Splicing components are excluded from the transcriptionally inactive XY body in male meiotic nuclei. *Mol. Biol. Cell*, **5**, 1341–1352.
5. Sharpe, R.M. (1993) Regulation of spermatogenesis. In Knobil, E. and Neill, J.D. (eds), *The Physiology of Reproduction*, Raven Press, New York, **1**, pp. 1363–1434.
6. Ruppert, S., Wang, E.H. and Tjian, R. (1993) Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature*, **362**, 175–179.
7. Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M. and Roeder, R.G. (1993) The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. *Nature*, **362**, 179–181.
8. Sekiguchi, T., Miyata, T. and Nishimoto, T. (1988) Molecular cloning of the cDNA of human X chromosomal gene (*CCG1*) which complements the temperature-sensitive G1 mutants, tsBN462 and ts13, of the BHK cell line. *EMBO J.*, **7**, 1683–1687.
9. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
10. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequencing of the human genome. *Science*, **291**, 1304–1351.
11. Vanin, E.F. (1985) Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.*, **19**, 253–272.
12. Hayashida, T., Sekiguchi, T., Noguchi, E., Sunamoto, H., Ohba, T. and Nishimoto, T. (1994) The *CCG1/TAFII250* gene is mutated in thermo-sensitive G1 mutants of the BHK21 cell line derived from golden hamster. *Gene*, **141**, 267–270.
13. Sekiguchi, T., Yoshida, M.C., Sekiguchi, M. and Nishimoto, T. (1987) Isolation of a human X chromosome-linked gene essential for progression from G1 to S phase of the cell cycle. *Exp. Cell Res.*, **169**, 395–407.
14. Sekiguchi, T., Nakashima, T., Hayashida, T., Kuraoka, A., Hashimoto, S., Tsuchida, N., Shibata, Y., Hunter, T. and Nishimoto, T. (1995) Apoptosis is induced in BHK cells by the tsBN462/13 mutation in the CCG1/TAFII250 subunit of the TFIID basal transcription factor. *Exp. Cell. Res.*, **218**, 490–498.
15. Kumar, S. and Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. *Nature*, **392**, 917–920.
16. McCarrey, J.R. and Thomas K. (1987) Human testis-specific *PGK* gene lacks introns and possesses characteristics of a processed gene. *Nature*, **326**, 501–505.
17. Dahl, H.H., Brown, R.M., Hutchison, W.M., Maragos, C. and Brown, G.K. (1990) A testis-specific form of the human pyruvate dehydrogenase E1 α subunit is coded for by an intronless gene on chromosome 4. *Genomics*, **8**, 225–232.
18. Hendriksen, P.J., Hoogerbrugge, J.W., Baarends, W.M., de Boer, P., Vreeburg, J.T., Vos, E.A., van der Lende, T. and Grootegoed, J.A. (1997) Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. *Genomics*, **41**, 350–359.
19. Ashworth, A., Skene, B., Swift, S. and Lovell-Badge, R. (1990) *Zfa* is an expressed retroposon derived from an alternative transcript of the *Zfx* gene. *EMBO J.*, **9**, 1529–1534.
20. Mardon, G., Luoh, S.W., Simpson, E.M., Gill, G., Brown, L.G. and Page, D.C. (1990) Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. *Mol. Cell. Biol.*, **10**, 681–688.
21. Das, B., McMahon, K.W., Jenkins, N.A., Gilbert, D.J., Copeland, N.G. and MacDonald, C.C. (2001) 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.
22. Sargent, C.A., Young, C., Marsh, S., Ferguson-Smith, M.A. and Affara, N.A. (1994) the glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. *Hum. Mol. Genet.*, **3**, 1317–1324.
23. Sedlacek, Z., Munstermann, E., Dhorne-Pollet, S., Otto, C., Bock, D., Schutz, G. and Poustka, A. (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.
24. Elliott, D.J., Venables, J.P., Newton, C.S., Lawson, D., Boyle, S., Eperon, I.C. and Cooke, H.J. (2000) An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. *Hum. Mol. Genet.*, **9**, 2117–2124.
25. Freiman, R.A., Albright, S.R., Zheng, S., Sha, W.C., Hammer, R.E. and Tjian, R. (2001) Requirement of tissue-selective TBP-associated factor $TAF_{II}105$ in ovarian development. *Science*, **293**, 2084–2087.
26. Zhang, D., Penttila, T.L., Morris, P.L., Teichmann, M. and Roeder, R.G. (2001) Spermiogenesis deficiency in mice lacking the *Trf2* gene. *Science*, **292**, 1153–1155.
27. Hiller, M.A., Lin, T.Y., Wood, C. and Fuller, M.T. (2001) Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.*, **15**, 1021–1030.
28. Boer, P.H., Adra, C.N., Lau, Y.F. and McBurney, M.W. (1987) The testis-specific phosphoglycerate kinase gene *Pgk-2* is a recruited retroposon. *Mol. Cell. Biol.*, **7**, 3107–3112.
29. Fitzgerald, J., Wilcox, S.A., Graves, J.A. and Dahl, H.H. (1993) A eutherian X-linked gene, *PDHA1*, is autosomal in marsupials: a model for the evolution of a second, testis-specific variant in eutherian mammals. *Genomics*, **18**, 636–642.
30. Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.J., Speed, R.M., Thomson, E.J., Jobling, M., Taylor, K. *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.
31. Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **23**, 403–405.
32. Li, W.H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, **36**, 96–99.