J. Biochem. 125, 1104-1114 (1999)

Retrovirus Integration Site *Mintb* Encoding the Mouse Homolog of hnRNP U[1]

Minoru Matsui,* Walter C. Breau,†,2 Shiho Iwasaki,* Sadahiro Hagiwara,* Yoshitaka Tamai,‡ Chisato Mori,§,3 Michael L. Bloom,†,4 Mary B. Jerry,†,5 Edward M. Eddy,§ and Makoto M. Taketo*,†,‡,6

*Laboratory of Biomedical Genetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033; †The Jackson Laboratory, Bar Harbor, ME 04609, USA; ‡Banyu Tsukuba Research Institute (Merck), Tsukuba 300-2611; and §Laboratory of Reproductive and Developmental Toxicology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

Received January 29, 1999; accepted March 17, 1999

Retroviral genes are not usually expressed in mouse embryonal carcinoma (EC) cells, but they are readily expressed upon differentiation of these cells. We previously reported the isolation of EC cell lines that express a neomycin resistance (*neo*) gene introduced by a recombinant transducing Moloney murine leukemia virus from specific integration sites, *Minta*, *Mintb*, *Mintc*, or *Mintd*. In some of these clones, the entire 5′ long terminal repeat (LTR) was deleted, and the *neo* gene was expressed by read-through transcription from upstream cellular promoters in a "promoter-trap" fashion. One such promoter ("promoter B" at the *Mintb* locus) was found in a CpG island, associated with an upstream enhancer ("enhancer B"). Although enhancer B caused expression of the *neo* gene in the transductant EC cell line, no endogenous transcription from promoter B was detected in the parental EC or NIH3T3 cells. In contrast, we found a strong counter-flow endogenous transcription unit ("R" for reverse), which apparently interfered with transcription from promoter B. Promoter R turned out to have a bidirectional activity in transfection assays. In normal tissues, promoter R activates gene R, which encodes an 800-residue protein that is highly homologous to the rat and human heterogeneous nuclear ribonucleoprotein U (hnRNP U). Northern and *in situ* hybridization analyses revealed that gene R was abundantly expressed in the testis, especially in the pachytene spermatocytes and round spermatids.

Key words: embryonal carcinoma cell, retrovirus, ribonucleoprotein, spermatogenesis, testis.

Mouse embryonal carcinoma (EC) cells, undifferentiated stem cells derived from teratocarcinoma, do not express the genes of viruses such as papovaviruses and retroviruses (*1, 2*). We previously reported clonal EC cells that expressed a bacterial neomycin resistant (*neo*) gene from a recombinant transducing retroviral vector of the Moloney murine leukemia virus (Fvc-12 virus) because of cellular enhancers and/or promoters adjacent to the integration sites *Minta*, *Mintb*, *Mintc*, or *Mintd* (*3-7*): *i.e.*, in an "enhancer trap" or a "promoter trap" fashion. Structural analysis revealed that the *neo* gene was expressed only when part of the proviral genome that included the splicing donor site was removed. The deletion often extended upstream to include the entire 5′ LTR and the flanking cellular sequences (*5*). From these transductant EC cell lines, we cloned the proviral genomes and their flanking cellular sequences into a bacteriophage λ vector. The 5′ LTR of the proviral genome was lost from the *Mintb* locus in transductant cell line B, resulting in a recombination of cellular DNA with the proviral sequence immediately upstream of the *neo* gene (*6*). Expression of the *neo* gene suggested that the upstream cellular sequence contained a promoter that was active in EC cells, and that the proviral *neo* gene was accordingly transcribed by a read-through mechanism.

Here, we report such a "trapped" promoter ("B"), which is located in an intron of another gene ("R" for reverse) transcribed in the opposite direction. We further analyzed gene "R" transcribed by promoter R. Gene R encodes an

[1] This work was supported in part by grants from the Council for Tobacco Research (CTR 1828 and 1828A) and the National Institute of Health (CA39652) to M.M.T. The nucleotide sequences for gene R and promoter B have been deposited in the GenBank under Accession Numbers X51436 and X51437, respectively.

Present addresses: [2] Elms College, 291 Springfield St. Chicopee, MA 01013, USA; [3] Department of Anatomy, Faculty of Medicine, Kyoto University, Sakyo, Kyoto 606, Japan; [4] National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA; [5]Department of Veterinary and Animal Sciences, University of Massachusetts, Amherst, MA 01003-5720, USA.

[6] To whom correspondence should be addressed at present address: Laboratory of Biomedical Genetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033. Phone: +81-3-5841-4859, Fax: +81-3-5841-4778, E-mail: taketo@mol.f.u-tokyo.ac.jp

Abbreviations: hnRNP, heterogeneous nuclear ribonucleoprotein; EC, embryonal carcinoma; *neo*, neomycin resistance; LTR, long terminal repeat; Rec, deletion-recombination point; SAR, scaffold attachment region; MAR, matrix attachment region; NLS, nuclear localization signal.

© 1999 by The Japanese Biochemical Society.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

800-residue protein, whose amino acid sequence is highly homologous to those of the rat and human heterogeneous nuclear ribonucleoprotein U (hnRNP U). The hnRNP U protein is the largest (120 kDa) member of hnRNPs and an abundant nucleoplasmic phosphoprotein. It is implicated in the processing of hnRNA to mRNA (for review, see Ref. *8*). Human hnRNP U was identified originally as the unique hnRNP containing an RGG box known as an RNA binding motif (*9*). In human cells, this protein was identified also as scaffold attachment factor A (SAF-A), which binds the AT-rich sequences known as scaffold attachment region (SAR) elements or matrix attachment region (MAR) elements (*10, 11*). Recently, SAF-A was suggested to be involved in apoptotic nuclear breakdown, because its proteolytic degradation eliminated the binding to the SAR elements and resulted in loss of chromatin structure (*12*). In the rat, cDNA clones for the hnRNP U homolog were cloned (*13*). Proteins of 120 kDa with a binding activity to the SAR element were also identified in the *Xenopus* (*14*), chicken (*15*), mouse (*14*), and bovine (*14*). These hnRNP U homologs are suggested to alter DNA conformation through binding to SAR/MAR elements and to be involved in such cellular processes as transcriptional activation and repression (*13-15*). Here, we report Northern and *in situ* hybridization analyses of the mouse hnRNP U and suggest its role in a testis-specific function.

MATERIALS AND METHODS

Transfection of DNA into EC and NIH3T3 Cells—The procedures for permanent transfection assay were described earlier (*5*). Transfections for transient expression assays using luciferase gene constructs were performed essentially in the same way except that 5 μg of plasmid DNA together with 15 μg of calf thymus DNA (carrier) were added to a plate of $5\text{-}10\times10^5$ cells inoculated a day earlier.

Luciferase Gene Constructs—Promoter B plasmids pXP21 and pXP22 were constructed by placing the promoter fragment (*Hin*dIII-*Eco*RI) into promoter-less luciferase plasmids, pXP1 and pXP2 (*16*) digested with the two enzymes. Partial promoter plasmid pXP23 was constructed by insertion of the blunt-ended *Sma*I-*Eco*RI fragment into *Sma*I site of pXP1. Enhancer-promoter B constructs pXP22 and pXP222 were constructed by inserting the enhancer-containing *Hin*dIII fragment into pXP21. Promoter R constructs pXP31 and pXP32 were constructed by insertion of the promoter fragment (*Xho*I-*Pst*I) into luciferase vector pXP2 as an *Xho*I-*Hin*dIII fragment that had been cloned into BluescriptKSII− (Stratagene, La Jolla, CA). Control plasmid pRSVL containing RSV-LTR (*17*) was obtained from Dr. Helinsky (UCSD, San Diego, CA), and the promoterless luciferase vectors (*16*) were gifts from Dr. Nordeen (Univ. Colorado, Denver, CO).

Luciferase Assays—Luciferase activities were determined according to the protocol of DeWet *et al.* (*17*) except that a Bioluminat LB9500 luminometer (Berthold Analytical Instruments, Nashua, NH) was used. Integrated photon counts were recorded for 1 min and corrected for protein concentrations, which varied from 4.5-7.0 mg/ml and 1.5-3.4 mg/ml for EC and NIH3T3 cell extracts, respectively. Data in Fig. 1 are presented as counts/mg protein.

Isolation of the Cellular Genomic Sequence Lost by Deletion—From a mouse genomic library generated by partial digestion of 129/Sv male DNA with *Sau*3A and cloned into the EMBL3 vector (*18*), three bacteriophage λ clones were isolated using the 956-bp *Hpa*I fragment probe of the promoter region (gray box in Fig. 2). Inserts of these clones carried overlapping sequences of 10 to 17 kb long. We characterized clone 8-1.2 (Fig. 2A), which contained the probe fragment in the center, leaving downstream about 5 kb of the cellular genomic DNA. Restriction mapping coupled with Southern hybridization analysis revealed that about 1.6 kb had been deleted in clone B together with a segment of recombinant retroviral genome including the 5′ LTR (*6*).

Isolation of cDNA Clones for the Gene R Transcripts—The cDNA library prepared from F9 EC cell RNA (*19*) was kindly provided by Dr. L. Gudas (Dana-Farber Cancer Institute, Boston, MA). The library was screened with the 0.9-kb *Pst*I fragment probe (hatched box in Fig. 2, A and C).

RNA Analyses—Northern blot hybridization analyses were performed as described by Brown (*20*), using ^{32}P-labeled pBC234 insert DNA (2,059-bp *Xho*I-*Xba*I fragment of the gene R cDNA) as a probe. In order to discriminate the two polyadenylated forms, we used the pBC238 insert DNA (316-bp *Xba*I-*Eco*RI fragment) as a probe. Hybridization mapping of the 5′ ends of the transcripts from promoter B and R was performed as described earlier (*5, 21*), using *Pst*I-*Xho*I fragment as probes (Fig. 4).

DNA Sequencing and Computer-Assisted Statistical Analyses—For sequencing, overlapping subclones were constructed in appropriate sizes in the pBS+, BluescriptKSII−, or BluescriptSK+ vectors. The dideoxynucleotide sequencing method was employed using double-stranded DNA and modified T7 polymerase (Sequenase, US Biochemical, Cleveland, OH), and AmpliTaq® DNA polymerase (FS, Perkin-Elmer Corporation, Foster City, CA). In Fig. 4, only the top strand in the original orientation is shown, although both strands were sequenced for the entire regions presented.

In Situ Hybridization—The procedures were described elsewhere (*22*). The probe is the pBC262 insert DNA (887-bp *Bgl*II-*Sac*I fragment of the gene R cDNA).

RESULTS

Structures and Transfection Activities of the Plasmid Subclones Derived from Phage Clone λB—In order to localize the *cis*-acting elements responsible for the *neo* gene expression in transductant EC cell line B, the 17-kb insert of bacteriophage clone λB (*5*) was recloned into pUC19 (pBU1), and a series of subclones was constructed. Upon luciferase reporter assay using these plasmids, we found that about 300 bp of the cellular sequence immediately upstream of the *neo* gene contained an essential component for the cellular promoter (data not shown), which was provisionally named promoter "B" (for clone B).

Analysis of the neo mRNA—Radiolabeled cRNA probes anti-parallel to the direction of the *neo* gene transcription were prepared and hybridized with polyA$^+$ RNAs isolated from parental F9tk$^-$ and clone B transductant EC cell lines. The *neo* gene probe (pDG1 insert shown in Fig. 2B) hybridized to a single band of about 2.6 kb in clone B but not in F9tk$^-$ (Fig. 1A). The upstream or downstream cellular

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

probes adjacent to the integrated proviral vector hybridized no mRNAs in either clone B or F9tk$^-$ (data not shown), indicating the absence of endogenous polyA$^-$ transcripts in the vicinity of this promoter in the same direction of transcription as the proviral *neo* gene.

Analysis of Cellular Enhancer-Promoter B Region—To characterize promoter B and its possible enhancer in detail, we employed the firefly luciferase reporter assay (*17*) with various expression vectors (Fig. 1B). The resultant luciferase activities in F9tk$^-$ and NIH3T3 cell lines are shown in Fig. 1, C and D, respectively. As a positive control, we used the Rous sarcoma virus LTR (pRSVL; *17*). The results indicated that promoter B was active in both undifferentiated EC cells and NIH3T3 fibroblasts, and that the 888-bp *Hin*dIII fragment contained an enhancer that was equally effective in both EC and NIH3T3 cells.

We determined the DNA sequence of the 5′-flanking cellular region, which encompassed promoter B, including about 2 kb upstream of the *neo* gene (sequence data not shown; available at GenBank Accession Number X51437). The GC content in this region is very high, and there is no CpG dinucleotide suppression: between the *Sma*I and *Xho*I sites (1,853 bp), the CpG island contains 204 CpGs and 238 GpCs with the average GC content of 72% (Fig. 2C). In the promoter B sequence, there are an Sp1 binding site (CCCCGCCGCC; 169–160 bp upstream of Rec) and a cAMP-responsive element (TGACGTCA; 113–106 bp upstream of Rec) but no TATA box or YY binding site. We could not identify any known repressor or silencer elements which might account for the inactivity of promoter B. Upon a similarity search by the FASTA algorithm, the DNA sequence of the 133 bp fragment in promoter B was found to be highly similar to those of the antisense strands of human hnRNP U and rat SP120 cDNAs (91.9 and 95.6% identity, respectively).

Isolation and Analysis of the Cellular Genomic Sequence Lost by Deletion We isolated contiguous fragments of mouse genomic DNA that hybridized with a probe from the promoter region (Fig. 2A) as described in "MATERIALS AND METHODS." Restriction mapping, Southern hybridization, and sequencing analyses of the clone revealed that a DNA fragment of about 1.6 kb had been deleted in EC cell line B together with a segment of the recombinant provirus (Fig. 2B).

Analysis of Transcripts around the Proviral Integration Site—Because another gene, "gene R (R for reverse)," was expected on the other strand, we used RNA probes transcribed in both orientations from a genomic fragment that had been deleted in transductant EC cell line B (hatched boxes in Fig. 2, A and C), and analyzed total cellular RNA and polyA$^+$ RNA from F9tk$^-$ and NIH3T3 cells. We could not evaluate the non-polyadenylated RNA transcribed by promoter B because of a high background mainly due to a cross-hybridization with rRNAs (Fig. 3; lanes 1, 2, 5, and 6). The cRNA probe anti-parallel to promoter B hybridized with no mRNA (lanes 3 and 4). On the other hand, the RNA probe in the same orientation as promoter B hybridized two RNA fragments of about 3.4 and 3.8 kb (lanes 7 and 8). These results indicated the expression of gene R, which was transcribed in the opposite direction to the *neo* gene.

Isolation and Characterization of cDNA for the Gene R mRNA—To identify gene R, we isolated candidate cDNA

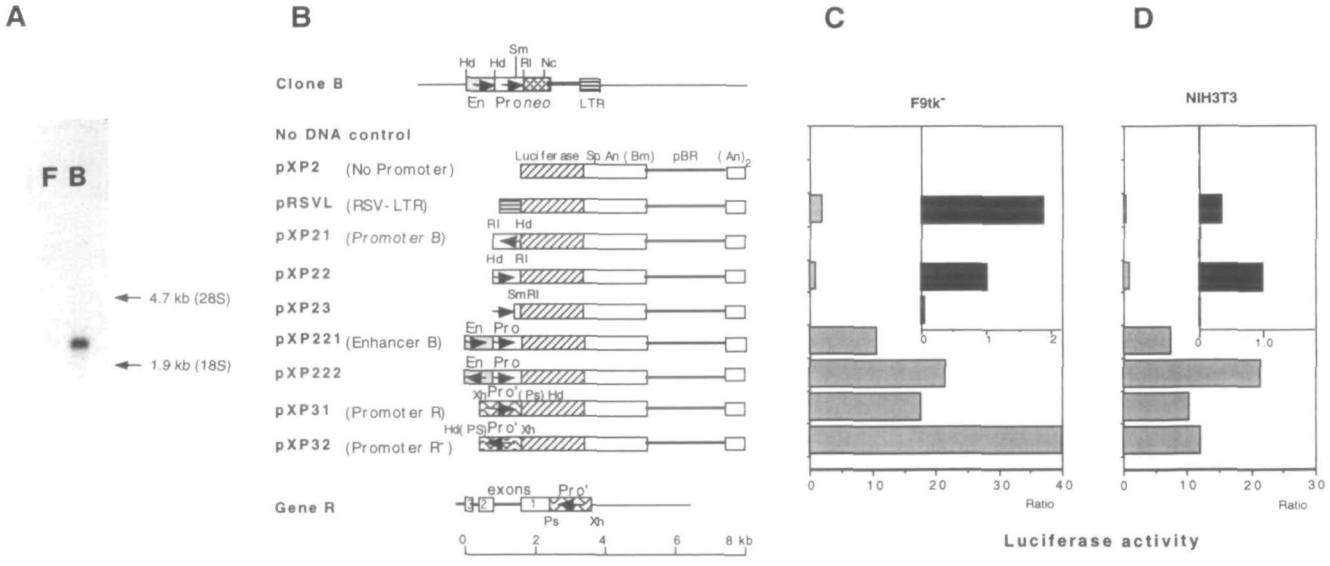


Fig. 1. **A: Northern blot hybridization analysis of *neo* mRNA.** Two micrograms of polyA$^+$ RNA isolated from F9tk$^-$ (lane F) or clone B (lane B) EC cells were loaded per lane. The filter was hybridized with a cRNA probe transcribed from a pDG1 plasmid (shown in Fig. 2B). **B, C, and D: Luciferase activities of test constructs in transient expression assays.** The test plasmids were constructed and their luciferase activities were determined by transient expression assays as described in "MATERIALS AND METHODS." B: Structures of the plasmid constructs. Abbreviations Pro and En indicate cellular promoter B (*Hin*dIII *Eco*RI fragment) and enhancer B (*Hin*dIII fragment), respectively, placed in front of the promoter-less structural gene for the firefly luciferase (hatched box), whereas Pro′ indicates cellular promoter R (*Xho*I *Pst*I fragment; Figs. 2 and 4). On top and bottom of the luciferase constructs are maps of the genomic DNA in transductant EC cell line B and non-transductant cells, respectively. C and D: Luciferase activities are shown as average ratios to the activity of pXP22, which had about 9,600 and 640,000 counts in the EC and NIH3T3 cell lines, respectively. The large difference in the activities between the two cell lines is apparently due to that of transfection efficiencies. Insets in C and D show the same data in an expanded scale for constructs with lower activities.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

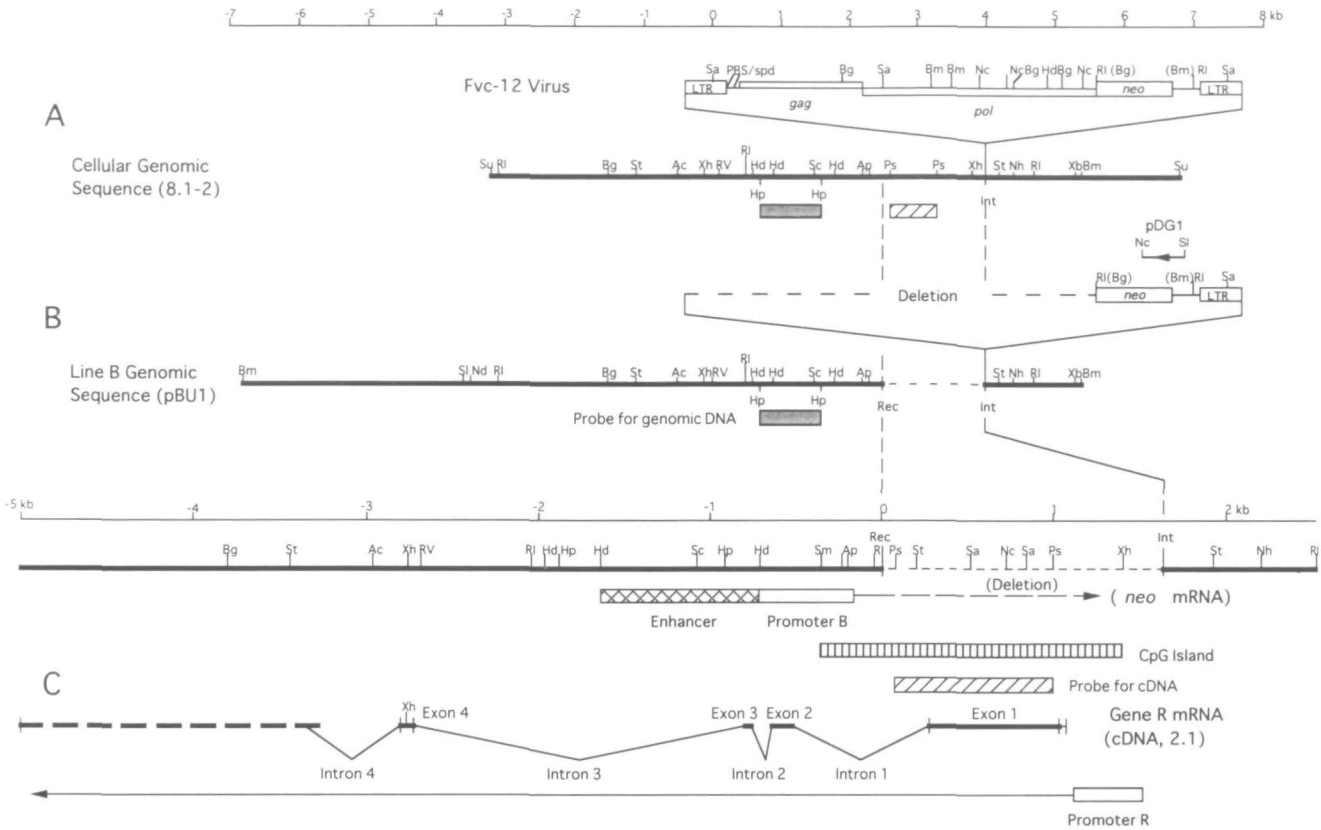


Fig. 2. **A: Restriction map of the cellular genomic sequence of EC cell clone B into which the recombinant retrovirus Fvc-12 (*4, 6*) had integrated prior to the selection for *neo* gene expression. B: Restriction map of the same cell line after selection for *neo* gene expression.** The regions containing the negative elements in the proviral genome that block viral *neo* gene expression have been deleted upstream of the *neo* gene together with the 5′-flanking cellular sequence. **C: Schematic diagrams of two promoters in the CpG island transcribing in opposite directions.** Gray and hatched boxes indicate the probes used for isolation of the bacteriophage clones covering the contiguous genomic sequence and the gene R cDNA, respectively. The box with vertical lines indicates the CpG island. Int and Rec indicate the proviral integration site and deletion-recombination point, respectively. Only relevant restriction sites are shown. Abbreviations for restriction sites are: Ac, *Acc*I; Ap, *Apa*I; Bm, *Bam*HI; Hd, *Hin*dIII; Hp, *Hpa*I; Nc, *Nco*I; Nd, *Nde*I; Nh, *Nhe*I; RI, *Eco*RI; RV, *Eco*RV; Sa, *Sac*I; Sc, *Sca*I; Sm, *Sma*I; St, *Stu*I; Su, *Sau*3AI; Xh, *Xho*I.

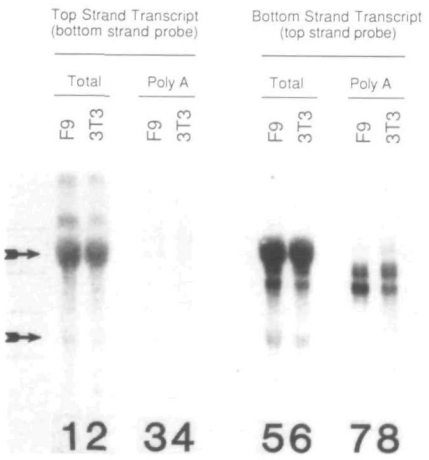


Fig. 3. **Northern analysis of the transcripts around the provirus integration site.** Total cellular RNA (lanes 1, 2, 5, and 6) and polyA$^+$ RNA (lanes 3, 4, 7, and 8) isolated from F9tk EC cells (lanes 1, 3, 5, and 7) and NIH3T3 fibroblasts (lanes 2, 4, 6, and 8) were electrophoresed in an agarose gel. Blotted filters were then hybridized with cRNA probes transcribed from the *Pst*I fragment of the cellular genomic sequence (hatched box in Fig. 2). Lanes 1 to 4 show hybridization with the bottom strand probe (*i.e.* complementary to the top strand, which is in the same direction as the proviral *neo* gene in EC cell line B). Lanes 5 to 8 show hybridization with the top strand probe (*i.e.* complementary to the bottom strand, which is in the opposite direction to the proviral *neo* gene). Arrows indicate the positions of 28S (4.7 kb) and 18S (1.9 kb) rRNAs.

clones as described in "MATERIALS AND METHODS." The clone 2.1 contained the longest insert of 3,380 bp cDNA, whose open reading frame encoded 800 amino acid residues (Fig. 5), leaving untranslated regions of about 190 and 800 bp on the 5′ and 3′ ends, respectively (Fig. 4). The sequence immediately upstream of the initiation codon agrees with the consensus for a translation initiation site described by Kozak (*23*). Analysis of the 3′ untranslated region revealed two copies of the polyadenylation signal sequence (AATAAA), 377-bp apart. The cDNA domains corresponding to the first four exons of gene R were precisely assigned based upon the genomic DNA sequence information (Fig. 2C).

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

These data indicated that the deleted region of the cellular genomic DNA in EC cell line B included exon 1 and promoter R, whereas promoter B and its enhancer region, which activated the proviral *neo* gene, corresponded to introns 1, 2, and 3 of gene R together with two short exons (exons 2 and 3). Only exon 1 lies in the CpG island, leaving the other exons outside.

Characterization of Promoter R—To investigate cellular promoter R, the DNA sequence upstream of the 5′ end of clone 2.1 was determined and the contiguous genomic sequence was reconstituted (Fig. 2). As shown in Fig. 4, there was a TATA box sequence (−244 to −238), and two possible Sp1 binding sites were identified upstream of the TATA box sequence (−309 to −300, −258 to −248). We further determined the mRNA initiation site using an RNase protection assay (Fig. 6). A single major protected band indicated that the gene R transcript started about 35 bp downstream from the TATA box (Fig. 4; −208 to

```
-493                                                    ccaggtgaggact  -481
-480  gaaggcgggtggaggagagcctccgccagacacagaaggcagagcgccggatggagccga  -421
-420  cggctcaagagcgtaaaggcgcgcactgcggctgttcctgcgtttaaccgaaaggtggct  -361
-360  tctaagagcctccatttgcgcggtccggcccagcgcggaggtcgtgcagtgggggcgggc  -301
-300  gggcctgcgcgcgcagcgctctagcgggcctcgcctcggcagggggcgggcgcctata  -241
-240  ttacgtgccggcgccgccccctgcgagaggacgttgcgtgctgcctcgcgccgggcgagt  -181
-180  ctccgcgtcttgcctcgcgaactcggtgaaggaattggcgccgttcgacaccaggcgggt  -121
-120  ccgttctgcagcagcactcagccatctccagccgcagccgccgcccgggccgaccagcag   -61
 -60  cagccgccgccgccacaggccgagggagccagcagtcagcggggaaccggggccctcaac    -1
   1  ATGAGTTCTTCGCCTGTTAATGTCAAGAAGCTGAAGGTGTCGGAGCTGAAGGAGGAGCTC    60
  61  AAGAAGCGGCGCCTCTCCGACAAGGGCCTCAAGGCCGATCTCATGGATCGACTCCAGGCC   120
 121  GCGCTGGACAACGAGGCAGGAGGCCGCCCCGCCATGGAGCCCGGGAACGGCAGTCTCGAC   180
 181  CTAGGTGGCGATGCGGCCGGGCGCTCGGGAGCGGGCCTAGAGCAGGAGGCCGCGGCTGGC   240
 241  GCCGAAGACGACGAGGAGGAGGAGGGCATCGCCGCTCTGGACGGCGACCAGATGGAGCTG   300
 301  GGCGAGGAGAACGGGGCGGCGGGGGCGGCTGACGCCGGCGCGATGGAGGAAGAGGAGGCA   360
 361  GCGTCGGAAGACGAGAACGGCGACGACCAGGGCTTCCAGGAGGGGGAAGACGAGCTCGGC   420
 421  GACGAGGAGGAGGGCGCGGGCGACGAGAACGGTCACGGGGAGCAGCAGTCCCAACCGCCG   480
 481  GCAGCGGCGGCGCAGCAGCAGCCTTCCCAGCAGCGTGGTGCCGGCAAGGAGGCCGCGGGC   540
 541  AAGAGCAGCGGCCCCACCTCGCTCTTCGCGGTGACGGTGGCGCCGCCAGGGGCGAGGCAG   600
 601  GGCCAACAGCAGGCGGGAGGAGACGGCAAAACAGAACAGAAAGGTGGAGATAAAAAGAGA   660
 661  GGCGTTAAAAGACCGCGAGAAGATCATGGCCGAGGGTATTTTGAGTACATCGAAGAAAAC   720
 721  AAGTACAGCAGAGCCAAGTCTCCTCAGCCACCTGTTGAAGAAGAAGACGAACACTTCGAT   780
 781  GACACAGTGGTTTGTCTTGATACTTATAATTGTGATCTGCATTTTAAAATCTCGAGAGAC   840
 841  CGTCTGAGTGCTTCTTCCCTTACTATGGAGAGTTTTGCTTTCCTGTGGGCTGGAGGAAGA   900
 901  GCTTCCTACGGTGTGTCAAAAGGCAAAGTCTGCTTTGAGATGAAGGTAACAGAGAAGATT   960
 961  CCAGTAAGACACTTATATACAAAAGATATTGATATACATGAAGTTCGGATTGGCTGGTCA  1020
1021  CTAACCACAAGTGGAATGTTGCTTGGTGAAGAAGAATTTTCTTACGGGTATTCTCTGAAA  1080
1081  GGAATAAAAACATGCAACTGTGAGACAGAAGATTATGGGGAGAAGTTTGATGAAAATGAT  1140
1141  GTGATTACATGCTTTGCTAACTTTGAAACTGATGAAGTTGAACTCTCTTATGCGAAGAAT  1200
1201  GGACAAGATCTTGGTGTTGCCTTTAAGATCAGTAAGGAAGTTCTTGCTGACCGGCCACTA  1260
1261  TTCCCACATGTTCTCTGCCATAACTGTGCAGTTGAATTTAATTTCGGTCAAAAGGAAAAG  1320
1321  CCATACTTTCCAATACCTGAAGACTGTACTTTTATCCAAAATGTCCCCTTAGAGGACCGA  1380
1381  GTTAGAGGACCAAAAGGACCTGAAGAGAAGAAGGATTGTGAGGTTGTAATGATGATTGGC  1440
1441  TTGCCAGGAGCTGGAAAAACTACCTGGGTTACTAAACATGCAGCTGAAAACCCTGGGAAA  1500
1501  TACAACATTCTTGGAACAAATACGATTATGGACAAGATGATGGTGGCAGGTTTTAAGAAG  1560
1561  CAAATGGCAGATACTGGAAAACTGAACACACTGTTGCAGAGAGCCCCACAGTGTCTTGGC  1620
1621  AAGTTTATTGAAATTGCTGCCCGTAAGAAGCGAAATTTTATTCTGGATCAGACAAATGTG  1680
1681  TCTGCTGCTGCCCAGAGAAGAAAAATGTGCCTGTTTGCAGGCTTCCAGCGGAAAGCTGTT  1740
1741  GTAGTGTGCCCAAAAGATGAAGACTATAAGCAGAGGACACAGAAGAAGGCAGAAGTAGAG  1800
1801  GGGAAGGACCTACCAGAACATGCTGTCCTCAAGATGAAAGGAAACTTCACCCTTCCAGAG  1860
1861  GTTGCAGAATGCTTTGATGAAATAACCTATGTTGAACTTCAGAAAGAGGAAGCCCAAAAG  1920
1921  CTTTTGGAGCAATATAAAGAAGAAAGCAAAAAGGCACTGCCACCAGAAAAGAAGCAAAAC  1980
1981  ACTGGCTCAAAGAAAAGCAATAAGAATAAGAGTGGCAAGAACCAGTTCAACAGAGGTGGT  2040
2041  GGCCATAGAGGCCGTGGAGGATTCAATATGCGAGGTGGCAATTTCAGAGGAGGAGCTCCT  2100
2101  GGGAATCGTGGTGGATATAATAGGAGAGGCAACATGCCACAGAGAGGTGGTGGCGGTGGA  2160
2161  AGTGGTGGAATTGGCTATCCATACCCACGTGGCCCTGTTTTTCCTGGCCGAGGTGGTTAC  2220
2221  TCAAACAGAGGGAATTACAACAGAGGTGGAATGCCCAACAGAGGGAACTATAACCAGAAC  2280
2281  TTCAGAGGACGAGGAAATAATCGTGGCTACAAAAATCAATCTCAGGGCTACAATCAGTGG  2340
2341  CAGCAGGGTCAATTCTGGGGTCAGAAGCCATGGAGTCAGCATTATCACCAAGGATATTAT  2400
2401  TGAatacccaaataaaacgaactgatacatatttctccaaaaccttcacaagaagtcgac  2460
2461  tgttttctttagtaggctaacttttaaacattccacaagaggaagtgcctgcgggttcc  2520
2521  ttttttagaagctttgtgggttgatttttttttctttctttttgtacatttttaattg  2580
2581  cagttttaaagtgaatcgtaagagaacctcagcattgtgcacgataagagaatgtgtcag  2640
2641  tatttcagggttctacattttatctgtaaaatgtgacttttttttttatcacaacaaaag  2700
2701  taaaatgttgctttgtacctggtgtcttttattaagcatttactcccccatttctcacag  2760
2761  agaataacagtcgggagtcattgtcacaataaaatagcaatgttagcagccagattcatg  2820
2821  gaaggactaaggggtcctcatgaattgcattaagattctgtactgctcatgatacactcc  2880
2881  atcctctctagactgcctgccgggtaatagtggacgggtaatctctgacaaaacgggaag  2940
2941  gctatttttttctgtttgacaatggaattggcataattgggaatgaagataaaatttgg  3000
3001  aaccaagattgagaagatggagtgtatgtagaagggctgttcaaaaatgtaaacttggtt  3060
3061  gcattatatgtggaggctcaaacttgtgaaggttaaataccataaaatttccatttgtt  3120
3121  ctgcattttgattctgaaaagaaagctgctttgcccatttcttattaaaaaaacttgttg  3180
3181  taaaaaaaaaaaaaaaaaaaa                                        3202
```

Fig. 4. **DNA sequence of gene R.** The amino acid coding region is shown in uppercase letters, and others in lowercase. Characteristic sequences are indicated: two possible Sp1 binding sites, boxed and bold; TATA box sequence, dotted underline; transcription initiation site, boxed (because it was estimated from the sequence ladder after correction of the mobility difference between DNA and RNA, the base of initiation could not be determined precisely; see "MATERIALS AND METHODS"); 5′ end of clone 2.1 insert, c at −178 in italic and bold; consensus sequence for translation initiation, italic and underline; initiation and stop codons, capital and bold; two polyA signals, lowercase and bold.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

−197).

Partial fragments of promoter R were tested for the promoter activity by a luciferase reporter assay. As shown at the bottom of Fig. 1, C and D, the constructs showed strong promoter activities not only in the original (pXP31) but also in the reverse (pXP32) orientation (designated R^-). Furthermore, promoter R was active in both types of cells and its activity was 10 to 40 times stronger than that of promoter B.

DNA Sequence Analysis of Gene R—The deduced amino acid sequence of the putative coding region of gene R (Fig. 5) is highly homologous to those of human hnRNP U [96.3% identity (*9*), GenBank Accession Number X65488] and rat SP120/hnRNP U [99.3% identity, (*13*), GenBank Accession Number D14048]. A putative nuclear localization signal (NLS) exists between the amino acids 218 and 226, consistent with the finding that hnRNP U and SP120 possess an NLS in similar positions and that both proteins are localized in the nucleus (*9, 13*). An RGG box was found (position, 691 to 710), which appears to be necessary for the

```
hn   1 MSSSPVNVKKLKVSELKEELKKRRLSDKGLKAELMERLQAALDDEEAGGRPAMEPGNGSL  60
ms   1 MSSSPVNVKKLKVSELKEELKKRRLSDKGLKADLMDRLQAALDNE-AGGRPAMEPGNGSL  59
rt   1 MSSSPVNVKKLKVSELKEELKKRRLSDKGLKADLMDRLQAALDNE-AGGRPAMEPGNGSL  59

hn  61 DLGGDSAGRSGAGLEQEAAAGGDEEEEEEEEEEEGISALDGDQMELGEENGAAGAADSGP 120
ms  60 DLGGDAAGRSGAGLEQEAAAGA-----EDDEEEEGIAALDGDQMELGEENGAAGAADAGA 114
rt  60 DLGGDAAGRSGAGLEQEAAAGA-----EDDEEEEGISALDGDQMELGEENGAAGAADAGA 114

hn 121 MEEEEAASEDENGDDQGFQEGEDELGDEEEGAGDENGHGEQQPQPPATQ-QQQPQQQRGA 179
ms 115 MEEEEAASEDENGDDQGFQEGEDELGDEEEGAGDENGHGEQQSQPPAAAAQQQPSQQRGA 174
rt 115 MEEEEAASEDENGDDQGFQEGEDELGDEEEGAGDENGHGEQQSQPPAAA--QQASQQRGP 172

hn 180 AKEAAGKSSGPTSLFAVTVAPPGARQGQQQAGGDGKTEQKGGDKKRGVKRPREDHGRGYF 239
ms 175 GKEAAGKSSGPTSLFAVTVAPPGARQGQQQAGGDGKTEQKGGDKKRGVKRPREDHGRGYF 234
rt 173 GKEAAGKSSGPTSLFAVTVAPPGARQGQQQAGGDGKTEQKAGDKKRGVKRPREDHGRGYF 232

hn 240 EYIEENKYSRAKSPQPPVEEEDEHFDDTVVCLDTYNCDLHFKISRDRLSASSLTMESFAF 299
ms 235 EYIEENKYSRAKSPQPPVEEEDEHFDDTVVCLDTYNCDLHFKISRDRLSASSLTMESFAF 294
rt 233 EYIEENKYSRAKSPQPPVEEEDEHFDDTVVCLDTYNCDLHFKISRDRLSASSLTMESFAF 292

hn 300 LWAGGRASYGVSKGKVCFEMKVTEKIPVRHLYTKDIDIHEVRIGWSLTTSGMLLGEEEFS 359
ms 295 LWAGGRASYGVSKGKVCFEMKVTEKIPVRHLYTKDIDIHEVRIGWSLTTSGMLLGEEEFS 354
rt 293 LWAGGRASYGVSKGKVCFEMKVTEKIPVRHLYTKDIDIHEVRIGWSLTTSGMLLGEEEFS 352

hn 360 YGYSLKGIKTCNCETEDYGEKFDENDVITCFANFESDEVELSYAKNGQDLGVAFKISKEV 419
ms 355 YGYSLKGIKTCNCETEDYGEKFDENDVITCFANFETDEVELSYAKNGQDLGVAFKISKEV 414
rt 353 YGYSLKGIKTCNCETEDYGEKFDENDVITCFANFETDEVELSYAKNGQDLGVAFKISKEV 412

hn 420 LAGRPLFPHVLCHNCAVEFNFGQKEKPYFPIPEEYTFIQNVPLEDRVRGPKGPEEKKDCE 479
ms 415 LADRPLFPHVLCHNCAVEFNFGQKEKPYFPIPEDCTFIQNVPLEDRVRGPKGPEEKKDCE 475
rt 413 LADRPLFPHVLCHNCAVEFNFGQKEKPYFPIPEDCTFIQNVPLEDRVRGPKGPEEKKDCE 473

hn 480 VVMMIGLPGAGKTTWVTKHAAENPGKYNILGTNTIMDKMMVAGFKKQMADTGKLNTLLQR 539
ms 475 VVMMIGLPGAGKTTWVTKHAAENPGKYNILGTNTIMDKMMVAGFKKQMADTGKLNTLLQR 534
rt 473 VVMMIGLPGAGKTTWVTKHAAENPGKYNILGTNTIMDKMMVAGFKKQMADTGKLNTLLQR 532

hn 540 APQCLGKFIEIAARKKRNFILDQTNVSAAAQRRKMCLFAGFQRKAVVVCPKDEDYKQRTQ 599
ms 535 APQCLGKFIEIAARKKRNFILDQTNVSAAAQRRKMCLFAGFQRKAVVVCPKDEDYKQRTQ 594
rt 533 APQCLGKFIEIAARKKRNFILDQTNVSAAAQRRKMCLFAGFQRKAVVVCPKDEDYKQRTQ 592

hn 600 KKAEVEGKDLPEHAVLKMKGNFTLPEVAECFDEITYVELQKEEAQKLLEQYKEESKKALP 659
ms 595 KKAEVEGKDLPEHAVLKMKGNFTLPEVAECFDEITYVELQKEEAQKLLEQYKEESKKALP 654
rt 593 KKAEVEGKDLPEHAVLKMKGNFTLPEVAECFDEITYVELQKEEAQKLLEQYKEESKKALP 652

hn 660 PEKKQNTGSKKSNKNKSGKNQFNRGGGHRGRGGLNMRGGNFRGGAPGNRGGYNRRGNMPQ 719
ms 655 PEKKQNTGSKKSNKNKSGKNQFNRGGGHRGRGGFNMRGGNFRGGAPGNRGGYNRRGNMPQ 714
rt 653 PEKKQNTGSKKSNKNKSGKNQFNRGGGHRGRGGFNMRGGNFRGGAPGNRGGYNRRGNMPQ 712

hn 720 RGGGGGGSGGIGYPYPRAPVFPGRGSYSNRGNYNRGGMPNRGNYNQNFRGRGNNRGYKNQ 779
ms 715 RGGGGG-SGGIGYPYPRGPVFPGRGGYSNRGNYNRGGMPNRGNYNQNFRGRGNNRGYKNQ 773
rt 713 RGGGGG-SGGIGYPYPRGPVFPGRGGYSNRGNYNRGGMPNRGNYNQNFRGRGNNRGYKNQ 771

hn 780 SQGYNQWQQGQFWGQKPWSQHYHQGYY 806
ms 774 SQGYNQWQQGQFWGQKPWSQHYHQGYY 800
rt 772 SQGYNQWQQGQFWGQKPWSQHYHQGYY 798
```

Fig. 5. **Amino acid sequence alignment of the gene R product (mouse), hnRNP U (human), and SP120 (rat).** Nuclear localization signals are shown in bold, and RGG-boxes in bold and italic letters. The species are shown on the left margin with abbreviations: hn, human; ms, mouse; and rt, rat. The symbols plus (+) and minus (−) represent 'similar' and 'not-similar' amino acids, respectively.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

RNA-binding property of human hnRNP U (*9*). The gene R product and human hnRNP U also share the features such as putative *N*-glycosylation sites (codons: 56-59, 559-562, 615-618, 669-672 and 772-775), putative phosphorylation sites (codons: 23-26 for cAMP-/cGMP-dependent protein kinase phosphorylation site; 26-28, 317-319, 358-360, 525-527, 593-595, 649-651, 663-665, 666-668, 671-673, and 741-743 for protein kinase C phosphorylation sites; 122-125, 508-511, 617-620, and 629-632 for casein kinase II phosphorylation sites), and abundant acidic amino acids

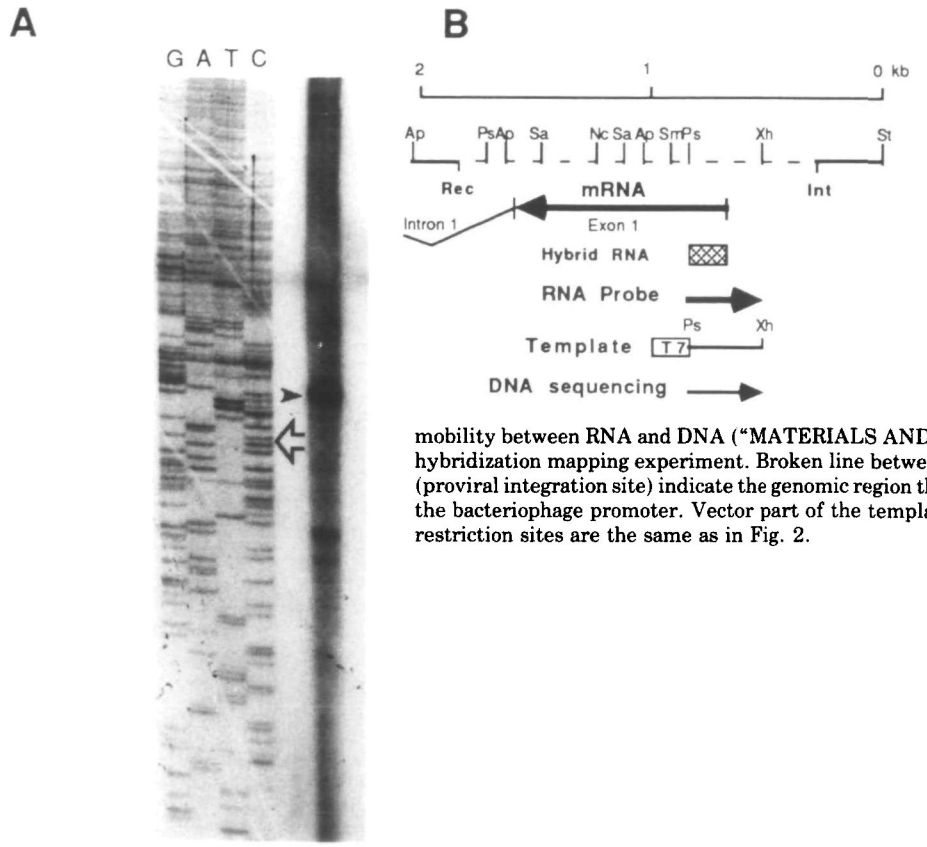


Fig. 6. **Hybridization mapping of the transcription start site of gene R.** A: Right: Autoradiograph of labeled cRNA probe fragments protected from RNase digestion after hybridization at 30 C with 20 μg of total cellular RNA from F9tk$^-$ EC cell line. Left: Autoradiograph of sequencing mixtures using the same DNA template and the primers as those used for labeling the cRNA probe, and electrophoresed simultaneously. The small arrowhead indicates the position of the protected fragment, whereas the open arrow below shows the position of the DNA sequence corresponding to the RNA size after correction for the difference in the electrophoretic mobility between RNA and DNA ("MATERIALS AND METHODS"). B: Schematic diagram of the hybridization mapping experiment. Broken line between Rec (deletion-recombination site) and Int (proviral integration site) indicate the genomic region that was deleted in EC cell line B. T7 indicates the bacteriophage promoter. Vector part of the template plasmid is not shown. Abbreviations for restriction sites are the same as in Fig. 2.

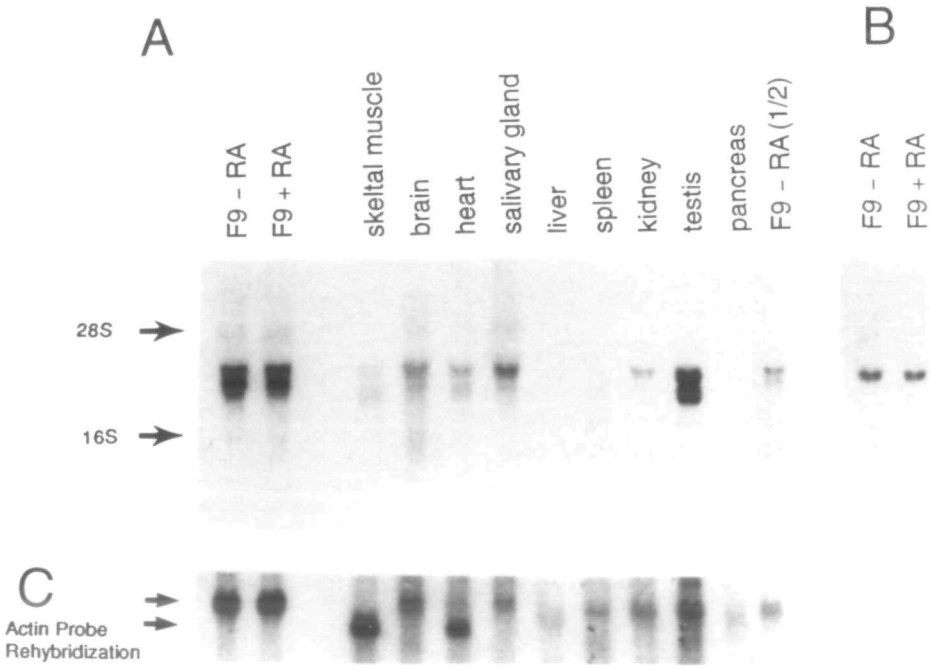
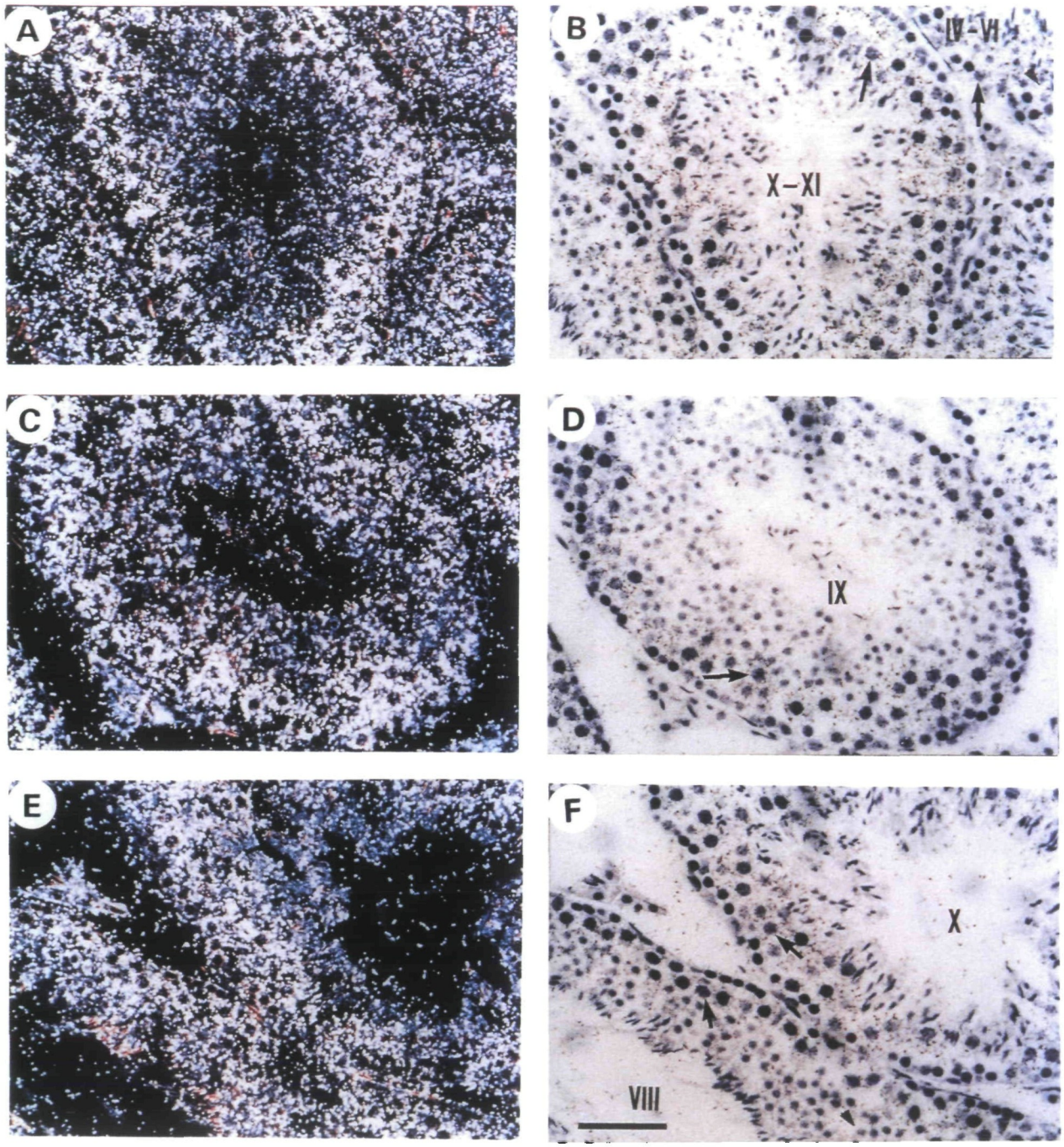


Fig. 7. **Detection of the gene R expression by Northern analysis in various tissues.** A: Two bands were detected in each tissue, reflecting two polyadenylated forms of the gene R transcript. The gene R was expressed in virtually all tissues, but band intensity was the highest in the testis. B: The probe covering the 3′ untranslated region between the first and the second polyA signals hybridized with the upper (larger) fragment only, suggesting both polyA signals were used in intact cells. +RA and −RA indicate that EC cells were 'treated' and 'untreated' with retinoic acid, respectively. The procedure for the retinoic acid treatment was reported previously (*4*). The amount of loaded RNA sample was reduced to 1/2 on the lane of "F9-RA (1/2)." C: A β-actin probe was used as a control. The bands with different sizes reflected the mRNAs for actin isoforms.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012



G

13	14	14	– 15	– 15	– 15	16	– 16				
1	2	3	+ 4	+ 5	+ 6	7	++ 8	± 9	– 10	– 11	12
P	P	P	+ P	+ P	+ P	P	+ P	++ P	++ P	++ D	2°
	In	In	– In	– B	– B	Pl	± Pl	± L	± L	± Z	Z
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII

Fig. 8. ***In situ* localization of the gene R mRNA in seminiferous tubules of an adult mouse.** Three sections are shown in dark-field (A, C, and E) and bright-field photomicrographs (B, D, and F), respectively. The silver grains (white dots in the dark-field photographs) indicate the abundant gene R mRNA. The slides were counterstained with hematoxylin. The Roman numbers indicate the stages of the spermatogenic cycle. G: Gene R expression and the spermatogenic cycle. The symbols + +, +, ±, and – indicate very strong, strong, weak, and no expression of the gene R transcripts, respectively. The Roman numbers show the stages of the spermatogenic cycle, whereas the Arabic numbers (1–16) indicate the steps of spermatid development (*24*). Arrows and arrowheads indicate representative pachytene spermatocytes and round spermatids (step 4–6 in B and step 8 in F), respectively. Abbreviations are: In, intermediate spermatogonia; B, type B spermatocytes; Pl, pre-leptotene spermatocytes; L, leptotene spermatocytes; Z, zygotene spermatocytes; P, pachytene spermatocytes; Di, diplotene spermatocytes; 2°, secondary spermatocytes. Scale bar, 60 μm.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

in the N-terminus. The remarkable similarity in the amino acid sequence and the presence of the well-conserved motifs strongly suggest that the gene R product is the orthologous mouse homolog of hnRNP U.

Analyses of Gene R Expression—The antisense DNA probe for the gene R coding region hybridized with two bands of different sizes (3.4 and 3.8 kb) in nearly equal intensity in RNAs from various tissues (Fig. 7A). The gene R transcripts were found in various tissues, with the highest expression in the testis. When a probe spanning the 3′ untranslated region between the first and second polyadenylation signals was used, only the upper band was detected (Fig. 7B), probably reflecting the presence of two polyadenylated forms of the gene R transcript.

To further investigate the expression of gene R in the testis, we performed *in situ* hybridizations with the pBC262 insert as a probe (Fig. 8). In the seminiferous tubule at each stage of spermatogenesis, specific spermatocytes and spermatids are found corresponding to each stage (*24*). Abundant gene R mRNA (a dense distribution of silver grains in Fig. 8, A, C, and E) was observed in various stages of seminiferous tubules. In stages X–XI, a high-level transcription of gene R was detected in pachytene spermatocytes (arrow), which were located in the periphery of the seminiferous tubule, whereas the signals in more centrally-located step 10 and 11 spermatids were as weak as in the background (Fig. 8, A, B, E, and F). In stages IV–VI, strong signals were observed in both pachytene spermatocytes (arrow) and round spermatids (step 4–6, arrowhead), but there was little hybridization in the step 15 spermatids (Fig. 8, A and B, right upper). The stage IX seminiferous tubules showed a high level of gene R expression in the pachytene spermatocytes (arrow) and a low level of expression in step 9 spermatids (Fig. 8, C and D). In stage VIII, a strong signal was detected in both pachytene spermatocytes (arrow) and step 8 spermatids (arrowhead), but expression was detected little in step 16 spermatids (Fig. 8, E and F, left lower). In summary, gene R was expressed in pachytene spermatocytes and round spermatids (Fig. 8G).

DISCUSSION

In the transductant EC cell line, we found three promoter sequences (B, R, and R^-) at the proviral integration site. Of these, only promoter R was active in the parental EC and NIH3T3 cells. Promoter B was situated downstream of the promoter R in the reverse orientation. Because promoter B was much weaker than promoter R, it was active for the proviral *neo* gene only in the transductant cell line, where promoter R was spontaneously deleted. The remaining allele of gene R in the clone B is intact and expressed (data not shown), and gene R appears to be essential because homozygous knockout is embryonically lethal (unpublished data).

Several mammalian genomic sequences in CpG islands have bidirectional promoter activities and transcribe two genes of head-to-head configuration (*25–32*). Although transfected promoter R^- showed significant activity by the reporter assay, we could not detect its natural transcript in the parental EC cells. Similar bidirectional promoters driving only unidirectional transcripts *in vivo* include those of c-Ha-*ras* (*33*), transcription factor USF (*34*), and dipeptidylpeptidase IV (*35*). The *in vivo* significance of promoters B and R^- needs to be investigated further in various cell types.

As seen in the Northern analysis (Figs. 3 and 7), two different sizes of the gene R mRNA were detected in various tissues, most likely due to the two alternative polyadenylation signals 376 bp apart. Two forms of the hnRNP U transcripts were reported in the human as well (*11*). A number of genes have multiple polyadenylated forms of mRNAs, and some of them are developmentally regulated (for review see Ref. *36*). Both polyadenylation signals of gene R were used equally in the adult tissues investigated, and it remains to be determined whether they play different roles during development.

Although the hnRNP U gene is a housekeeping gene, and the encoded protein helps process hnRNA to mRNA, the gene R mRNA was expressed at a wide variety of levels: at a high level in the testis and at significantly lower levels in the liver, spleen and pancreas. Our observation that gene R is highly expressed in the testis in pachytene spermatocytes and round spermatocytes (Figs. 7 and 8) is consistent with a previous electron microscopic finding using antibodies against hnRNPs (*37*). The timing of gene R expression during the course of spermatogenesis coincides well with that of the hnRNA synthesis: highest in late pachytene spermatocytes and significantly lower in the other stages (*38, 39*). Furthermore, it is possible that the hnRNP U protein is one of several RNA-binding proteins that play important roles in spermatogenesis (for review see Ref. *40*). The human and rat hnRNP U proteins bind to AT-rich elements, SAR/MAR. At the same time, proteins such as topoisomerase II, SATB1, and yeast RAP1 also bind to the elements and regulate cellular functions including transcription (*41–45*). Of these, yeast RAP1 has been reported to stimulate meiotic recombination at the *HIS4* locus (*45*). A testis-specific gene, *Tsx*, is located in AT-rich sequences (*46*), and it is conceivable that the gene R product regulates its expression through SAR/MAR.

It has been reported that human hnRNP U is associated with the glucocorticoid receptor in the nucleus and interferes its signal transduction (*47*). The amino acid sequence of the C-terminus portion of androgen receptor is significantly similar to that of glucocorticoid receptor (51% identity and 80% similarity in 366 amino acids). Because the C-terminus domain is sufficient for the hnRNP U binding, it is conceivable that the gene R product is associated with androgen receptor as well.

The human hnRNP U/SAF-A binds to hnRNA through its RGG-box at the C-terminus. In contrast, it has been reported that a double-stranded DNA-binding motif at the N-terminus mediates its apoptosis-inducing activity (*12*). It remains to be determined whether the testis-specific function of hnRNP U is mediated by either of these domains.

In conclusion, we cloned gene R, which is the hnRNP U homolog in the mouse and likely to play an important role in spermatogenesis.

We thank Dr. T. Kadesh for providing pSV2Agpt, Drs. D. Helinski and S. Nordeen for luciferase constructs, Dr. H. Lehrach for mouse genomic DNA library, and Dr. L. Gudas for EC cell cDNA library. We thank Drs. G. Dreyfuss, T. Ikeyama, and W. Balkan for fruitful discussions.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

REFERENCES

1. Swartzendruber, D.E. and Lehman, J.M. (1975) Neoplastic differentiation: interaction of simian virus 40 and polyoma virus with murine teratocarcinoma cells *in vitro*. *J. Cell. Physiol.* **85**, 179-187
2. Teich, N.M., Weiss, R.A., Martin, G.R., and Lowy, D.R. (1977) Virus infection of murine teratocarcinoma stem cell lines. *Cell* **12**, 973-982
3. Feuer, G., Taketo, M.M., Hanecak, R.C., and Fan, H. (1989) Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. *J. Virol.* **63**, 2317-2324
4. Taketo, M.M., Gilboa, E., and Sherman, M.I. (1985) Isolation of embryonal carcinoma cell lines that express integrated recombinant genes flanked by the Moloney murine leukemia virus long terminal repeat. *Proc. Natl. Acad. Sci. USA* **82**, 2422-2426
5. Taketo, M. and Tanaka, M. (1987) A cellular enhancer of retrovirus gene expression in embryonal carcinoma cells. *Proc. Natl. Acad. Sci. USA* **84**, 3748-3752
6. Taketo, M. and Shaffer, D.J. (1989) Deletions in a recombinant retrovirus genome associated with its expression in embryonal carcinoma cells. *J. Virol.* **63**, 4431-4433
7. Taketo, M., Howard, T.A., and Seldin, M.F. (1992) Mapping of recombinant retrovirus integration sites that cause expression of the viral genome in murine embryonal carcinoma cells. *Mammalian Genome* **2**, 240-245
8. Dreyfuss, G., Matunis, M.J., Piñol-Roma, S., and Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289-321
9. Kiledjian, M. and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**, 2655-2664
10. Fackelmayer, F.O. and Richter, A. (1994) hnRNP-U/SAF-A is encoded by two differentially polyadenylated mRNAs in human cells. *Biochim. Biophis. Acta* **1217**, 232-234
11. Fackelmayer, F.O. and Richter, A. (1994) Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity. *Biochemistry* **33**, 10416-10422
12. Göhring, F., Schwab, B.L., Nicotera, P., Leist, M., and Fackelmayer, F.O. (1997) The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *EMBO J.* **16**, 7361-7371
13. Tsutsui, K., Tsutsui, K., Okada, S., Watarai, S., Seki, S., Yasuda, T., and Shohmori, T. (1993) Identification and characterization of a nuclear scaffold protein that binds the matrix attachment region DNA. *J. Biol. Chem.* **268**, 12886-12894
14. Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U., and Richter, A. (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J.* **11**, 3431-3440
15. von Kries, J.P., Buck, F., and Strätling, W.H. (1994) Chicken MAR binding protein p120 is identical to human heterogeneous nuclear ribonucleoprotein (hnRNP) U. *Nucleic Acids Res.* **22**, 1215-1220
16. Nordeen, S.K. (1988) Luciferase reporter gene vectors for analysis of promoters and enhancers. *BioTechniques* **6**, 454-458
17. De Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D., and Subramani, S. (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.* **7**, 725-737
18. Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**, 827-842
19. LaRosa, G.J. and Gudas, L.J. (1988) An early effect of retinoic acid: cloning of an mRNA (Era-1) exhibiting rapid and protein synthesis-independent induction during teratocarcinoma stem cell differentiation. *Proc. Natl. Acad. Sci. USA* **85**, 329-333
20. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman J.G., Smith, J.A., and Struhl, K. (eds.) (1986) *Current Protocols in Molecular Biology*, Unit 4.9, Wiley Interscience, Brooklin, NY
21. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984) Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035-7056
22. Mori, C., Welch, J.E., Sakai, Y., and Eddy, E.M. (1992) *In situ* localization of spermatogenic cell-specific glyceraldehyde 3-phosphate dehydrogenase (Gapd-s) messenger ribonucleic acid in mice. *Biol. Reprod.* **46**, 859-868
23. Kozak, M. (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8132
24. Russell, L.D., Ettlin, R.A., SinhaHikim, A.P., and Clegg, E.D. (1990) *Histological and Histopathological Evaluation of the Testis*, 1st ed., Cache River Press, Clearwater, FL
25. Williams, T.J. and Fried, M. (1986) The MES-1 murine enhancer element is closely associated with the heterogeneous 5′ ends of two divergent transcription units. *Mol. Cell. Biol.* **6**, 4558-4569
26. Lavia, P., Macleod D., and Bird, A. (1987) Coincident start sites for divergent transcripts at a randomly selected CpG-rich island of mouse. *EMBO J.* **6**, 2773-2779
27. Linton, J.P., Yen, J.-Y.J., Selby, E., Chen, Z., Chinsky, J.M., Liu, K., Kellems, R.E., and Crouse, G.F. (1989) Dual bidirectional promoters at the mouse *dhfr* locus: cloning and characterization of two mRNA classes of the divergently transcribed *Rep-1* gene. *Mol. Cell. Biol.* **9**, 3058-3072
28. Heikkilä, P., Soininen, R., and Tryggvason, K. (1993) Directional regulatory activity of *cis*-acting elements in the bidirectional alpha 1(IV) and α 2(IV) collagen gene promoter. *J. Biol. Chem.* **268**, 24677-24682
29. Wright, K.L., White, L.C., Kelly, A., Beck, S., Trowsdale, J., and Ting, J.P.Y. (1995) Coordinate regulation of the human TAP1 and LMP2 genes from a shared bidirectional promoter. *J. Exp. Med.* **181**, 1459-1471
30. Byrd, P., Cooper, P.R., Stankovic, T., Kullar, H.S., Watts, G.D.J., Robinson, P.J., and Taylor, A.M.R. (1996) A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies. *Hum. Mol. Genet.* **5**, 1785-1791
31. Brenner, V., Nyakatura, G., Rosental, A., and Platzer, M. (1997) Genomic organization of two novel genes on human Xq28: compact head to head arrangement of IDHγ and TRAPδ is conserved in rat and mouse. *Genomics* **44**, 8-14
32. Schmiedeknecht, G., Büchler, C., and Schmitz, G. (1997) A bidirectional promoter connects the p14.5 gene to the gene for RNase P and RNase MRP protein subunit hPOP1. *Biochem. Biophys. Res. Commun.* **241**, 59-67
33. Lowndes, N.F., Paul, J., Wu, J., and Allan, M. (1989) c-Ha-*ras* gene bidirectional promoter expressed *in vitro*: location and regulation. *Mol. Cell. Biol.* **9**, 3758-3770
34. Lin, Q., Luo, X., and Sawadogo, M. (1994) Archaic structure of the gene encoding transcription factor USF. *J. Biol. Chem.* **269**, 23894-23903
35. Qvist, H., Sjöström, H., and Norén, O. (1998) The TATA-less, GC-rich porcine dipeptidylpeptidase IV (DPPIV) promoter shows bidirectional activity. *Biol. Chem.* **379**, 75-81
36. Edwards-Gilbert, G., Veraldi, K.L., and Milcarek, C. (1997) Alternative poly (A) site selection in complex transcription units: means to and end? *Nucleic Acids Res.* **25**, 2547-2561
37. Biggiogera, M., von Schack, M.-L., Martin, T.E., Gordon, J., Müller, M., and Fakan, S. (1993) Immunoelectron microscope localization of snRNP, hnRNP, and ribosomal proteins in mouse spermatogenesis. *Mol. Reprod. Dev.* **35**, 261-271
38. Monesi, V. (1965) Synthetic activities during spermatogenesis in the mouse. *Exp Cell Res.* **39**, 197-224
39. Kierszenbaum, A.L. and Tres, L.L. (1974) Transcription sites in spread meiotic prophase chromosomes from mouse spermatocytes. *J. Cell. Biol.* **63**, 923-935
40. Eddy, E.M. and O'Brien, D.A. (1998) Gene expression during mammalian meiosis. *Curr. Top. Dev. Biol.* **37**, 141-200
41. Liu, J., Bramblett, D., Zhu, Q., Lozano, M., Kobayashi, R., Ross,

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

S.R., and Dudley, J.P. (1997) The matrix attachment region-binding protein SATB1 participates in negative regulation of tissue-specific gene expression. *Mol. Cell. Biol.* **17**, 5275-5287
42. Vignais, M.-L. and Sentenac, A. (1989) Asymmetric DNA bending induced by the yeast multifunctional factor TUF. *J. Biol. Chem.* **264**, 8463-8466
43. Hofmann, J.F.-X., Laroche, T., Brand, A.H., and Gasser, S.M. (1989) RAP-1 factor is necessary for DNA loop formation *in vitro* at the silent mating type locus HML. *Cell* **57**, 725-737
44. Devlin, C., Tice-Baldwin, K., Shore, D., and Arndt, K.T. (1991) RAP1 is required for BAS1/BAS2- and GCN4-dependent transcription of the yeast *HIS4* gene. *Mol. Cell. Biol.* **11**, 3642-3651
45. White, M.A., Wierdl, M., Detloff, P., and Petes, T.D. (1991) DNA-binding protein RAP1 stimulates meiotic recombination at the *HIS4* locus in yeast. *Proc. Natl. Acad. Sci. USA* **88**, 9755-9759
46. Simmler, M.-C., Cunningham, D.B., Clerc, P., Vermat, T., Caudron, B., Cruaud, C., Pawlak, A., Szpirer, C., Weissenbach, J., Claverie, J.-M., and Avner, P. (1996). A 94 kb genomic sequence 3′ to the murine *Xist* gene reveals an AT rich region containing a new testis specific gene *Tsx*. *Hum. Mol. Genet.* **5**, 1713-1726
47. Eggert, M., Michel, J., Schneider, S., Bornfleth, H., Baniahmad, A., Fackelmayer, F.O., Schmidt, S., and Renkawitz, R. (1997) The glucocorticoid receptor is associated with the RNA-binding nuclear matrix protein hnRNP U. *J. Biol. Chem.* **272**, 28471-28487

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012