Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 3789–3794, April 1997
Developmental Biology

Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells

(antisense/reporter gene/transplantation)

BRIAN P. ZAMBROWICZ*†, AKIRA IMAMOTO*‡, STEVE FIERING*, LEONARD A. HERZENBERG§, WILLIAM G. KERR§¶, AND PHILIPPE SORIANO*‖

*Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and §Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

Communicated by R. L. Brinster, University of Pennsylvania, Philadelphia, PA, January 3, 1997 (received for review August 28, 1996)

ABSTRACT **The ROSAβgeo26 (ROSA26) mouse strain was produced by random retroviral gene trapping in embryonic stem cells. Staining of ROSA26 tissues and fluorescence-activated cell sorter-Gal analysis of hematopoietic cells demonstrates ubiquitous expression of the proviral βgeo reporter gene, and bone marrow transfer experiments illustrate the general utility of this strain for chimera and transplantation studies. The gene trap vector has integrated into a region that produces three transcripts. Two transcripts, lost in ROSA26 homozygous animals, originate from a common promoter and share identical 5′ ends, but neither contains a significant ORF. The third transcript, originating from the reverse strand, shares antisense sequences with one of the noncoding transcripts. This third transcript potentially encodes a novel protein of at least 505 amino acids that is conserved in humans and in *Caenorhabditis elegans*.**

Gene traps provide a general strategy to identify genes exhibiting discrete patterns of expression during development and differentiation (1). The trap vectors contain a reporter gene, typically β-galactosidase (β-gal), that is not expressed unless it integrates into an intron or exon of a transcription unit. The integration results in reporter gene expression that reflects the expression pattern of the endogenous gene or is influenced by nearby transcriptional regulatory elements. The reporter gene also provides a molecular tag for cloning the trapped gene. By using these techniques, novel genes have been identified with a diversity of reporter gene expression patterns, providing invaluable tools as lineage markers for the study of normal development and to better understand the developmental consequences of specific mutations. One mouse gene trap line, ROSA26, displays ubiquitous expression of the reporter gene during embryonic development and, therefore, has been useful as a marker line in chimera experiments (2).

In this work, we have extended these observations by demonstrating ubiquitous β-gal activity in various tissues and in hematopoietic cells and by illustrating the use of this strain for bone marrow transfers. Because of its general utility, it was important to characterize the region in which the reporter gene had integrated. We show herein that both DNA strands of the ROSA26 genomic region are transcribed producing convergent and antisense RNAs.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/943789-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

MATERIALS AND METHODS

Genotyping and 5-Bromo-4-Chloro-3-Indolyl β-D-Galactoside (X-Gal) Staining. Mice were maintained on C57BL/6 × 129Sv, 129Sv congenic, and C57BL/6J congenic backgrounds and are available from the Induced Mutant Resource at The Jackson Laboratory. X-Gal staining was carried out as described (3). Mouse genomic DNA was digested with *Stu*I and electrophoresed on a 0.7% agarose gel. The gel was blotted onto Hybond N^+ and probed with the 5′ RACE (rapid amplification of cDNA ends) product. The probe hybridizes to ≈7- and 12-kb bands corresponding to the wild-type and mutant alleles, respectively. PCR genotyping was done with the following three primers: 5′-GGCTTAAAGGCTAACCTGATGTG-3′; 5′-GCGAAGAGTTTGTCCTCAACC-3′; and 5′-GGAGCGGGAGAAATGGATATG-3′. The sizes of the wild-type and mutant fragments were 374 and 1146 bp, respectively.

Multiparameter Fluorescence-Activated Cell Sorter-Gal (FACS-Gal) Analysis. Mononuclear cells prepared from spleen, bone marrow, thymus, and the peritoneal cavity were first subjected to hypotonic loading with fluorescein di-β-D-galactopyranoside (FDG), returned to isotonicity at 4°C, and stained with antibodies for specific surface determinants as described (4). For antibody staining of cells "loaded" with FDG, the cells were kept at 4°C at all steps in the staining procedure, including centrifugation in prechilled rotors and adapters. Antibody stains and the staining medium used in the procedure also were kept on ice throughout the duration of the procedure.

5′ and 3′ RACE and cDNA and Genomic Cloning. 5′ RACE was carried out as described by Chen (5), and 3′ RACE was done as described by Frohman (6). Plasmids pR26–10 and pR26–9 contain subclones of 3′ RACE products from transcripts 1 and 2, respectively. The pR26–10 insert was used to identify eight clones from an embryonic day (E) 11.5 oli-

Abbreviations: ES cell, embryonic stem cell; FACS-Gal, fluorescence-activated cell sorter-Gal; β-gal, β-galactosidase; X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; FDG, fluorescein di-β-D-galactopyranoside; RACE, rapid amplification of cDNA ends; E, embryonic day; AS, antisense; SA, splice acceptor; RT–PCR; reverse transcription-coupled PCR; BM, bone marrow; PGK, phosphoglycerate kinase-1.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U83173, U83174, U83175, and U83176).

†Present address: Lexicon Genetics, 4000 Research Forest Park, The Woodlands, TX 77381.
‡Present address: Ben May Institute for Cancer Research, University of Chicago, Chicago, IL 60637.
¶Present address: Department of Molecular and Cellular Engineering, University of Pennsylvania Medical School, Philadelphia, PA 19104.
‖To whom reprint requests should be addressed. e-mail: psoriano@fhcrc.org.

go(dT)-primed mouse embryo cDNA library. The pR26–9 insert was used to probe the E11.5 and an E16.5 mouse oligo(dT)-primed embryo cDNA library. A single clone was identified in the E16.5 library. The same probe was used to screen a mouse randomly primed embryonic stem (ES) cell cDNA library, and eight transcript antisense (AS) clones were obtained. The longest of these was used to reprobe the ES cell cDNA library and 25 additional clones were obtained.

The pR26–10 insert was also used to screen a mouse 129Sv genomic library. Three clones were obtained and one contained genomic sequence on both the 5′ and 3′ ends of the ROSAβgeo integration site. A partial *Eco*RI fragment of this clone (G19) was subcloned into the *Eco*RI site of pBSIIKS (Stratagene), resulting in plasmid pR26G19, which was used to map the ROSA26 region. The 5′ end of the ROSAβgeo insertion was amplified by PCR from ROSA26 homozygous mouse DNA with an exon 1-specific primer (r265′f, 5′-TGCGTTTGCGGGGATGG-3′) and a splice acceptor (SA)-specific primer (SAR, 5′-GCGAAGAGTTTGTCCTCAAC-3′).

Northern Blot Analysis. Northern blots were made using 20 μg of total RNA per lane on a 1.4% agarose gel. The *Eco*RI–*Hin*dIII fragment of pR26–10 (nucleotides 98–1162 of transcript 1) was used as a probe for transcript 1, and the *Xho*I fragment of transcript AS cDNA-1 (nucleotides 887 to ≈1600 of transcript AS) was used as a probe for transcript AS.

Reverse Transcription-Coupled PCR (RT–PCR). The RT–PCRs were carried out with kidney total RNA and the 3′ RACE protocol (6). The primers for detecting transcript 1 are R26GSP0 and Q_0 followed by Rosa263′ (5′-GCCGTTCTGT-GAGACAG-3′) and 575–695R (5′-AAATGTTCTGGA-CAAACACTTC-3′) and result in a 533-bp product. Primers for detecting transcript 2 are R26GSP0 and Q_0 followed by R26B (5′-CGCACTGCTCAAGCCTTTGTTC-3′) and Rosa263′ and result in a 217-bp product. Primers for detecting transcript AS are R26alt2 (5′-TAACTCCAGTTCTA-GGGGG-3′) and Q_0 followed by R26B and Rosa26i2-F1 (5′-GGTCAAGCAGTGTAACCTG-3′) and result in a 188-bp product.

Testing of Promoter Fragments. Several putative promoter fragments 5′ of exon 1 of transcripts 1 and 2 were placed upstream of β-gal. A Kozak ATG exists in exon 1 just 5′ of the *Not*I site that could affect translation of β-gal, so it was mutagenized to a *Bam*HI site by using primer rosa265′-mutR (5′-CGGATCCCCGCAAACGCACCAA-3′). These fragments were subcloned into the *Hin*dIII site of pSAβ-gal (2) after the removal of the SA site, and the resulting constructs were electroporated into ES cells. After selection with G418, resistant colonies were pooled (≈1000 per construct), grown up, and used to produce cell extracts. β-gal activity was measured using *o*-nitrophenyl β-D-galactopyranoside as a substrate (7), and Bio-Rad protein assays were done on each extract to determine the protein concentrations.

RESULTS

The Trapped Gene Is Expressed Ubiquitously. The ROSA26 mutant line was produced by infection of ES cells with the ROSAβgeo retrovirus (2). Heterozygotes did not display an overt phenotype and were recovered in expected numbers from heterozygous fathers (47%; $n = 147$) or mothers (46%; $n = 84$) bred to wild type. Significantly fewer than expected homozygotes were recovered from crosses between two heterozygous parents (11%; $n = 114$; $P < 0.01$, χ^2 test), but these homozygotes did not display an overt phenotype and were fertile. ROSA26 was one of several gene trap lines that exhibited widespread β-gal expression (2), starting at the morula–blastocyst stage. Examination of serial sections through E9.5 embryos demonstrated blue staining in all cells (ref. 8 and data not shown). Most tissues are formed by birth, so we also examined expression in neonates (Fig. 1). Ubiquitous

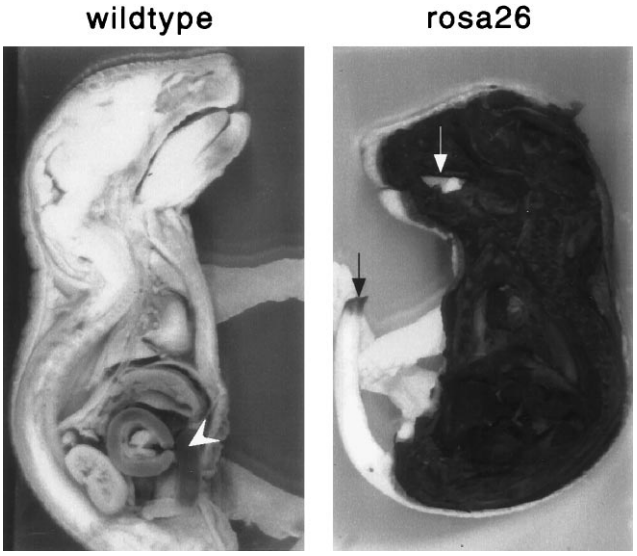

FIG. 1. Ubiquitous expression of β-gal in newborns. A wild-type (*Left*) and a heterozygous (*Right*) newborn were fixed with paraformaldehyde and cross-sectioned prior to X-Gal staining. The arrowhead indicates background staining in the intestines. The white arrow indicates the tongue, in which the mucus membrane prevented penetration of the X-Gal stain. The black arrow shows the depth of X-Gal penetration from the cut surface in the tail. Note that skin makes a good barrier.

staining was found in the following tissues: brain, bone marrow, cartilage, heart, intestine, kidney, liver, lung, pancreas, muscle (skeletal and smooth), skin (dermis and epidermis), spleen, submandibular gland, thymus, trachea, and urinary bladder. Because the staining was superficial even when tissues were cut open, histological sections were examined only in layers that contained stained cells to confirm ubiquitous staining. Frozen sections generally provided much weaker signals than paraffin sections. Moreover, ubiquitous expression has been found in adult testis (G. R. MacGregor, personal communication), and the brain exhibits ubiquitous β-gal expression except for olfactory bulb granule cells (R. L. Davis, personal communication).

***β*-Gal Expression in the Hematolymphoid Compartment and Hematopoietic Transplantation.** Nucleated cells in spleens from ROSA26 and two other strains, ROSA11 and ROSA27, that also exhibited apparently ubiquitous βgeo expression in E12 embryos (2), were analyzed for expression of β-gal by multiparameter FACS-Gal analysis. Only ROSA26 showed ubiquitous expression in the nucleated cells in spleen (Fig. 2*A*). In addition, all major hematolymphoid lineages express β-gal ubiquitously. Ubiquitous expression of β-gal was found in B cells ($B220^+$), T cells ($CD5^+$), and myeloid cells ($Mac\text{-}1^+$) in the spleen and their relative proportion was comparable to that found in normal animals (e.g., C57BL/6J), indicating that development of these various lineages is not impaired in mice homozygous for the gene trap integration. When other hematolymphoid tissues were analyzed, ubiquitous expression was also observed in nucleated cells, including bone marrow (BM), thymus, peritoneal cavity, and peripheral blood (Fig. 2*A* and data not shown). Because nucleated erythrocytes progenitors are present in BM cell suspensions, they should also express β-gal because all BM cells express β-gal in ROSA26 (Fig. 2*A*) whereas nonnucleated definitive erythrocytes present in peripheral blood of ROSA26 mice do not express β-gal (data not shown). Lack of expression in mature erythrocytes might be due to the long life of these cells after enucleation, during which time the βgeo protein is degraded.

Because of the ubiquitous expression of β-gal, ROSA26 mice might be useful to monitor engraftment of transplanted

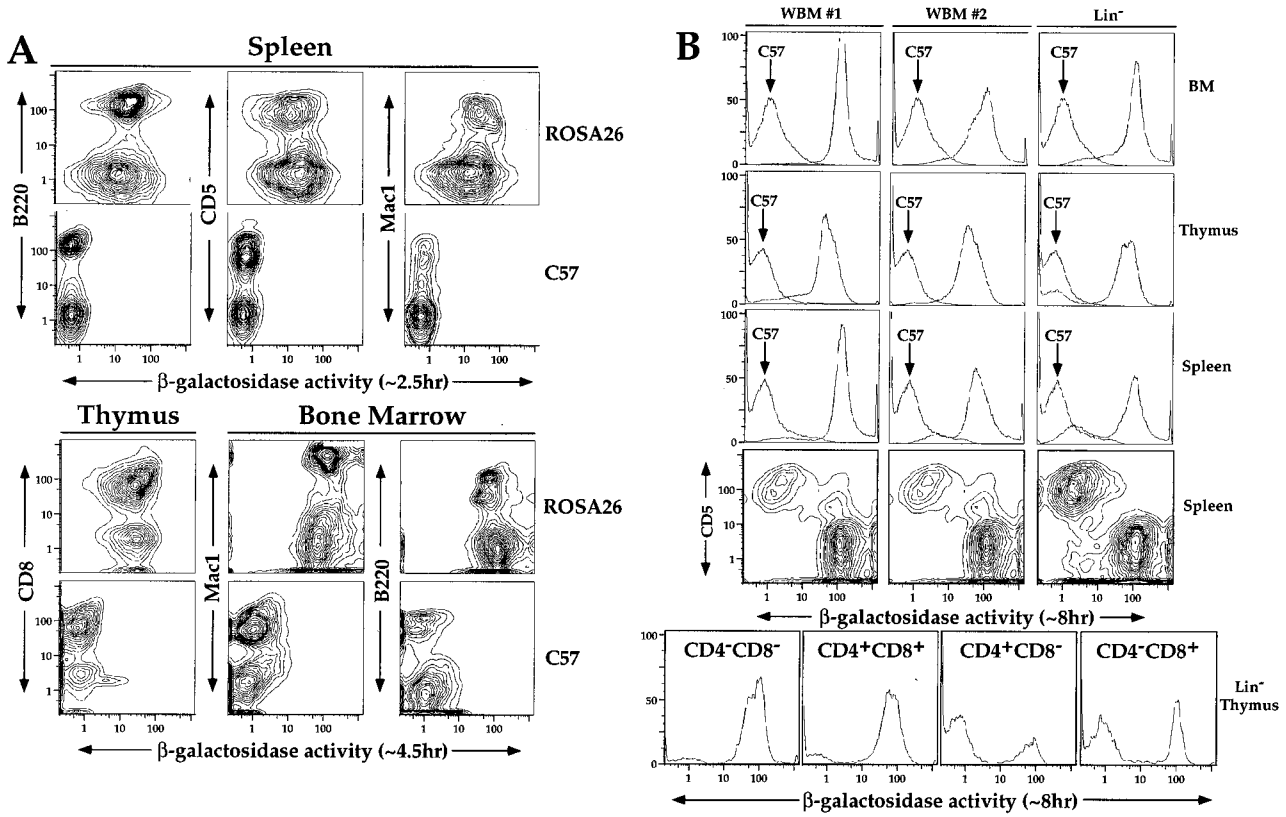

FIG. 2. Multiparameter FACS-Gal analysis of hematolymphoid organs in ROSA26 mice and mice transplanted with ROSA26 bone marrow. (*A*) Multiparameter FACS-Gal analysis of adult hematolymphoid organs in ROSA26 homozygous mice. Nucleated cell suspensions from spleen, thymus, and BM of ROSA26 homozygous mice were loaded with FDG, stained with antibody markers specific for certain hematopoietic lineages (B cells, B220; T cells, CD5 or CD8; myeloid cells, Mac1) and analyzed by FACS. To determine background fluorescence in the FACS-Gal assay, fluorescence levels of ROSA26 mice were compared with C57BL/6J controls (contour plots labeled C57 on right). (*B*) FACS-Gal analysis of C57BL/6J mice transplanted with ROSA26 BM cells. After lethal irradiation, mice were transplanted with either whole BM (WBM #1 or #2) or cells sorted to be negative for a panel of hematolymphoid lineage markers (Lin^-). Four weeks after irradiation and transplantation, single cell suspensions were prepared from BM, thymus, and spleen of the transplant recipients; loaded with FDG; stained with various antibody markers; and analyzed on the FACS. To assess the degree of reconstitution by the ROSA26 bone marrow cells, the fluorescence histogram of an identically prepared cell suspension from C57BL/6J age-matched mice is denoted by an arrow (C57). To demonstrate that the remaining host cells found in the spleen are T cells (high CD5 expressing cells), we show two-color probability plots for CD5 vs. β-gal staining of splenocytes. To determine which stages of thymocyte development still contain a significant number of host cells, we show FACS-Gal histograms electronically gated for various patterns of CD4 and CD8 expression in the thymus of the mouse reconstituted with Lin^- cells (Lin^- Thymus).

hematolymphoid cells, whether they are primitive stem/progenitor cell populations or mature end-stage cells. To this end, we performed several BM transplantations into lethally irradiated (750 rad; 1 rad = 0.1 Gy) recipient C57BL/6J mice, with either whole BM (2×10^6 cells) or cells partially enriched for hematopoietic stem/progenitor activity by sorting for cells that do not express antigens present on lineage-committed hematopoietic cells (1×10^5 Lin^- cells). These cells were isolated from heterozygous ROSA26 mice backcrossed three generations to C57BL/6J and sorted to be $CD5^-$ $Mac1^-$ $B220^-$ $CD4^-$ $CD8^-$ $Gr\text{-}1^-$ (Lin^-). β-gal expression in the hematolymphoid compartment of these mice showed that 4 weeks after transplantation, BM-derived progenitor cells could reconstitute all major hematolymphoid lineages as evidenced by the high proportion of $\beta\text{-gal}^+$ cells found in nucleated cells of BM, spleen, and thymus (Fig. 2*B*). FACS-Gal analysis done in combination with antibody stains to delineate the various lineages showed that while there had been nearly complete donor cell reconstitution of B lineage cells ($B220^+$) and myeloid lineage cells ($Mac1^+$) in the periphery (data not shown), there had not yet been significant contribution of donor-derived progenitor cells to the peripheral T cell compartment (high $CD5^+$) as evidenced by the overwhelming proportion of $\beta\text{-gal}^-$ (host origin) T cells (high $CD5^+$) in the spleen of mice reconstituted with either whole BM or Lin^- cells (Fig. 2*B*).

The majority of mature T cells (high $CD5^+$) in the spleens of all the reconstituted mice 4 weeks after transplantation were $\beta\text{-gal}^-$ and, therefore, of host origin (C57BL/6J). To test whether thymic progenitor cells of host origin could be giving rise to these peripheral T cells, we analyzed the major developmental stages of T lymphopoiesis for β-gal expression as a marker of donor vs. host origin (Fig. 2*B*). This analysis showed that in the animals reconstituted with whole BM, nearly all cells present in the major stages of T lymphopoiesis ($CD4^-8^-$, $CD4^+8^+$, $CD4^+8^-$, $CD8^+4^-$) were $\beta\text{-gal}^+$ and, therefore, of donor origin suggesting that the host-derived T cells might be derived from long-lived radio-resistant T cells (data not shown). However, in thymocytes of the animal reconstituted with Lin^- BM cells (see Fig. 2*B*, Lin^- thymus), while the overwhelming majority of the more immature T cell progenitor populations ($CD4^+8^+$, $CD4^-8^-$) are $\beta\text{-gal}^+$ and, therefore, of donor origin, there was still significant host contribution to single positive ($CD4^+8^-$, $CD8^+4^-$) thymocytes. This suggests that peripheral T cells of host origin could be derived from residual thymic progenitors, as well as radio-resistant mature T cells. The successful engraftment of lethally irradiated animals by $\beta\text{-gal}^+$ ROSA26-derived BM progenitor cells as monitored by multiparameter FACS-Gal analysis points to the utility of the ROSA26 strain for studies of BM transplantation and delineation of the developmental potential of stem/progenitor cell populations.

The ROSA26 Region Produces Three Transcripts. 5′ RACE was employed to identify exons from the trapped gene. The RACE product contained 130 bp of unique sequence. To confirm that this sequence was derived from the trapped gene, it was used as a probe on Southern blots of *Stu*I-digested DNA from wild-type, heterozygous, and homozygous ROSA26 mice (data not shown). The probe identified an restriction fragment length polymorphism (wild-type band of ≈7 kb and mutant band of ≈12 kb) that cosegregates with the ROSA26 allele.

The unique sequence was used to design primers for 3′ RACE. Two classes of products were obtained. Both contained identical 5′ ends but diverged at their 3′ ends. These 3′ RACE products were used as probes on cDNA libraries. Multiple cDNAs were obtained for one of the two RACE products and their sequence was used to piece together a 1,170-nt cDNA referred to as transcript 1 (Fig. 3*A*) that contains a poly(A) addition signal 20 nt 5′ of its 3′ poly(A) tail. The second 3′ RACE product was used as a probe on several cDNA libraries but only one 412-nt cDNA, referred to as transcript 2, was obtained (Fig. 3). Further 3′ RACE identified transcript 2 messages as long as 2.1 kb (data not shown). Searches of both transcripts 1 and 2 revealed no significant ORFs nor any similarities to known sequences.

A

```
GGCTC CTCAG AGAGC CTCGG CTAGG TAGGG GATCG GGACT CTGGC GGGAG    50
GGCGG CTTGG TGCGT TTGCG GGGAT GGGCG GCCGC GGCAG GCCCT CCGAG   100
CGTGG TGGAG CCGTT CTGTG AGACA GCCGG ATCAT TCCTT GAGGA CAGGA   150
CAGTG CTTGT TTAAG GCTAT ATTTC TGCTG TCTGA GCAGC AACAG GTCTT   200
CGAGA TCAAC ATGAT GTTCA TAATC CCAAG ATGTT GCCAT TTATG TTCTC   250
AGAAG CAAGC AGAGG CATGA TGGTC AGTGA CAGTA ATGTC ACTGT GTTAA   300
ATGTT GCTAT GCAGT TTGGA TTTTT CTAAT GTAGT GTAGG TAGAA CATAT   350
GTGTT CTGTA TGAAT TAAAC TCTTA AGTTA CACCT TGTAT AATCC ATGCA   400
ATGTG TTATG CAATT ACCAT TTTAA GTATT GTAGC TTTCT TTGTA TGTGA   450
GGATA AAGGT GTTTG TCATA AAATG TTTTG AACAT TTCCC CAAAG TTCCA   500
AATTA TAAAA CCACA ACGTT AGAAC TTATT TATGA ACAAT GGTTG TAGTT   550
TCATG CTTTT AAAAT GCTTA ATTAT TCAAT TAACA CCGTT TGTGT TATAA   600
TATAT ATAAA ACTGA CATGT AGAAG TGTTT GTCCA GAACA TTTCT TAAAT   650
GTATA CTGTC TTTAG AGAGT TTAAT ATAGC ATGTC TTTTG CAACA TACTA   700
ACTTT TGTGT TGGTG CGAGC AATAT TGTGT AGTCA TTTTG AAAGG AGTCA   750
TTTCA ATGAG TGTCA GATTG TTTTG AATGT TATTG AACAT TTTAA ATGCA   800
GACTT GTTCG TGTTT TAGAA AGCAA AACTG TCAGA AGCTT TGAAC TAGAA   850
ATTAA AAAGC TGAAG TATTT CAGAA GGGAA ATAAG CTACT TGCTG TATTA   900
GTTGA AGGAA AGTGT AATAG CTTAG AAAAT TTAAA ACCAT ATAGT TGTCA   950
TTGCT GAATA TCTGG CAGAT GAAAA GAAAT ACTCA GTGGT TCTTT TGAGC  1000
AATAT AACAG CTTGT TATAT TAAAA ATTTT CCCCA CAGAT ATAAA CTCTA  1050
ATCTA TAACT CATAA ATGTT ACAAA TGGAT GAAGC TTACA AATGT GGCTT  1100
GACTT GTCAC TGTGC TTGTT TTAGT TATGT GAAAG TTTGG GAATA AACCT  1150
ATGTC CTAAA T
```

B

```
TTGTATATTTTATTTTAAAGCAGATGATAAAACTGTAGATCTTAAGGGATACTTCTGCTTC   60
AACATATAAATAAAATTTCGTCTACTATTTTGACATCTAGAATTCCCTATGAAGACGAAG

TGAGATGATACAAAGAATTTAGACCATAAAACAGTAGGTTGCACAAGCAATAGAATATGG  120
ACTCTACTATGTTTCTTAAATCTGGTATTTTGTCATCCAACGTGTTCGTTATCTTATACC

CCTAAAGTGTTCTGACACTTAGAAGCCAAGCAGTGTAGGCTTCTTAAGAAATACCATTAC  180
GGATTTCACAAGACTGTGAATCTTCGGTTCGTCACATCCGAAGAATTCTTTATGGTAATG

AATCACCTTGCTAGAAATCAAGCATTCTGGAGTGGTCAAGCAGTGTAACCTGTACTGTAA  240
TTAGTGGAACGATCTTTAGTTCGTAAGACCTCACCAGTTCGTCACATTGGACATGACATT

GTTACTTTTCTGCTATTTTTCTCCCAAAGCAAGTTCTTTATGCTGATATTTCCAGTGTTA  300
CAATGAAAAGACGATAAAAAGAGGGTTTCGTTCAAGAAATACGACTATAAAGGTCACAAT
                                                  exon 3
GGAACTACAAATATTAATAAGTTGTCTTCACTCTTTTTCTTTACCAAGGAGGGTCTCTTCC  360
CCTTGATGTTTATAATTATTCAACAGAAGTGAGAAAAGAAATGGTTCCTCCCAGAGAAGG

TTCATCTTGATCTGAAGGATGAACAAAGGCTTGAGCAGTGCGCTTTAGAAGATAAACTGC  420
AAGTAGAACTAGACTTCCTACTTGTTTCCGAACTCGTCACGCGAAATCTTCTATTTGACG

AGCATGAAGGCCCCCGATGTTCACCCAGACTACATGGACCTTTCGCCACACATGTCCCAT  480
TCGTACTTCCGGGGGCTACAAGTGGGTCTGATGTACCTGGAAAGCGGTGTGTACAGGGTA

TCCAGATAAGGCCTGGCACACACAAAAAACATAAGT
AGGTCTATTCCGGACCGTGTGTGTTTTTTGTATTCA
```

FIG. 3. ROSA26 sequences. (*A*) Transcript 1 cDNA. The boldface **A** indicates the first base of exon 2. The splice donor used in transcript 2 is underlined and the poly(A) addition signal is boxed. (*B*) Genomic sequences containing the start of exon 3 of transcript 2 (top sequence) and the final exon of transcript AS (bottom sequence). Splice acceptor sites for the two exons are underlined. The start of exon 3 of transcript 2 is indicated by a solid arrow while the start of the final exon of transcript AS is indicated by an open arrow. The stop codon for transcript AS is in boldface type and the poly(A) addition signal for transcript AS is boxed. Note that the sequence between nucleotides 348 and 492 is shared by transcripts 2 and 3 with a sense–antisense relationship.

While probing cDNA libraries for transcript 2, multiple cDNAs were obtained with identity to transcript 2 sequences but were transcribed from the complementary strand. Sequencing of these cDNAs identified a 2-kb cDNA. This cDNA, referred to as transcript AS, contains an ORF of at least 1605 nt that begins at the 5′ end. The cDNA sequence contains a poly(A) addition signal 25 nt from the 3′ poly(A) tail. 5′ RACE was used to find additional 5′ sequence but identified the same 5′ end, suggesting the cDNA may be full length. The most 5′ in-frame ATG codon is in the context of a Kozak start site (9) and may encode the translation initiation site. Searches of the database with an amino acid sequence deduced from the ORF identified one gene cloned by the *Caenorhabditis elegans* genome project with an overall similarity of 59.5% and identity of 40.5%. In addition, three human expressed sequence tags were found that had sequence identities with transcript AS ranging from 77 to 86%. Transcript AS contains no overlap with transcript 1 but overlaps with transcript 2 over 381 nt that are all contained within the transcript AS coding sequence (Fig. 3*B*). Thus the ROSA26 region encodes three transcripts, two noncoding and one coding transcript that is highly conserved in evolution.

Loss of the Noncoding Transcripts in ROSA26 Homozygotes. To determine the effect of the retroviral integration on the expression of the three transcripts, Northern blot and RT–PCR studies were carried out. Transcript 1 was present in all wild-type tissues, but absent in all mutant tissues (Fig. 4*A*). By contrast, reprobing the same Northern blot revealed that transcript AS is present in all tissues of both wild-type and mutant mice and confirmed the integrity of RNA. The ≈1.4- and 2.0-kb sizes of transcripts 1 and AS, respectively, suggest that the cDNAs isolated must be nearly full length. Probing of polysomal RNA showed that transcript AS was present on polysomes whereas transcript 1 was not. Furthermore, *in vitro* translation of transcript 1 failed to produce a protein product (data not shown).

RT–PCR was carried out on kidney RNA from wild-type and ROSA26 mutant mice to examine the effects of retroviral integration on expression of transcript 2, as it could not be detected by Northern blots (Fig. 4*B*). This analysis confirmed the previous results on expression of transcript 1 and AS and showed that transcript 2 expression was eliminated in mutant mice.

Mapping of the ROSA26 Genomic Region. A transcript 1 probe was used to screen a 129Sv genomic library to clone the ROSA26 genomic region. PCR was used to amplify the ROSA26 mutant genomic DNA. The wild-type ROSA26 genomic region was mapped and transcripts 1, 2, and AS exon sequences and the ROSAβgeo integration site were identified (Fig. 5). The ROSAβgeo provirus has integrated into the first intron of transcript 1 and 2 with the SA sequence oriented in a position to trap both transcript 1 and 2. Transcript 1 and 2 share exon 1 and the 5′ end of the second exon, but transcript 1 continues to be transcribed through the genomic DNA whereas transcript 2 splices to a third exon. This third exon of transcript 2 overlaps with the final exon of transcript AS resulting in a sense–antisense relationship between these transcripts. Transcript AS is in reverse orientation to transcripts 1 and 2 and thus cannot be trapped by the SA sequence of ROSAβgeo.

Backcross panel mapping was used to identify the chromosomal location of ROSA26. Southern blots of mouse genomic DNA were used to identify a *Msp*I restriction fragment length

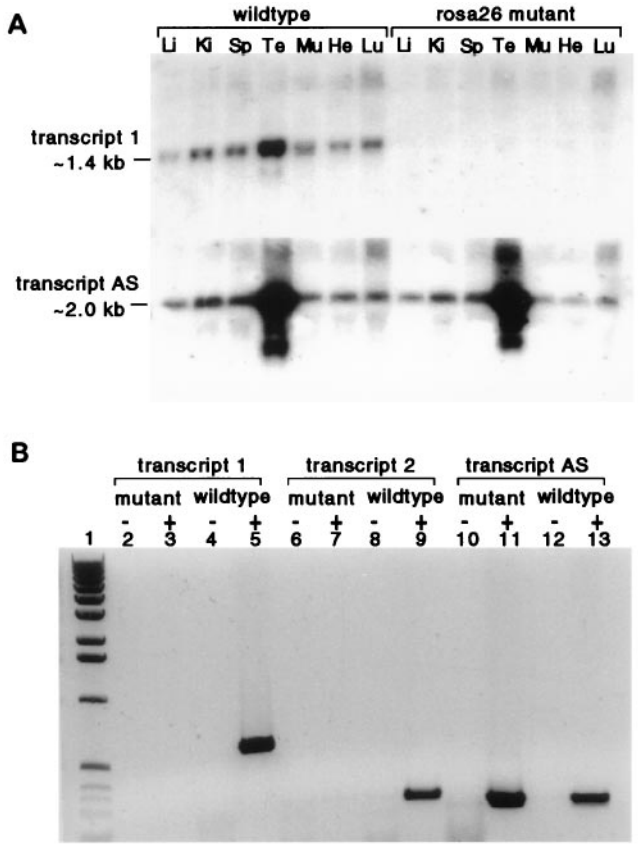

FIG. 4. Effect of the ROSA26 mutation on expression of the three transcripts. (*A*) Multiple adult tissue RNAs from wild-type and ROSA26 homozygous mutant mice were used on a Northern blot that was probed first for transcript 1 and then for transcript AS. He, heart; Ki, kidney; Li, liver; Lu, lung; Mu, muscle; Sp, spleen; Te, testis. (*B*) RT–PCR assays were done on adult kidney RNA samples obtained from wild-type or ROSA26 homozygous mutants as indicated. The RT reaction was carried out in the presence (+) or absence (−) of reverse transcriptase and with three different sets of primers that could specifically detect transcripts 1 (lanes 2–5), 2 (lanes 6–9), or 3 (lanes 10–13). The molecular size marker in lane 1 is the 1-kb ladder (GIBCO/BRL).

polymorphism between *Mus musculus* and *Mus spretus* DNA. Backcross panel DNAs were obtained from The Jackson Laboratory and *Msp*I blots were probed to demonstrate that the ROSA26 region maps to mouse chromosome 6 with no crossovers with the marker D6Mit10.

Rosa26 Transcript Promoter Identified. Because ROSA26 may be a useful region for targeting ubiquitous expression of various genes, we tested whether sequences 5′ of exon 1 of transcripts 1 and 2 had promoter activity. Primer extension was used to identify three transcription start sites (see GenBank accession no. U83173) and GC and CAAT boxes, but this region lacks a TATAA sequence. These features are common for housekeeping gene promoters. To identify the promoter, various fragments were fused to a β-gal reporter gene (Fig. 6*A*). A potential translation start site within exon 1 was mutated to a *Bam*HI site, as it might prevent proper translation of β-gal. A wild-type fragment containing the potential translation start site was also fused to β-gal as was the phosphoglycerate kinase-1 (PGK) promoter as a positive control and no promoter as a negative control. All constructs also contained a PGK promoter directing the expression of the neomycin-resistance gene for positive selection. Constructs were electroporated into ES cells and after G418 selection, β-gal activity was determined on extracts from pooled colonies (Fig. 6*B*). The PGK promoter produced the highest β-gal activity

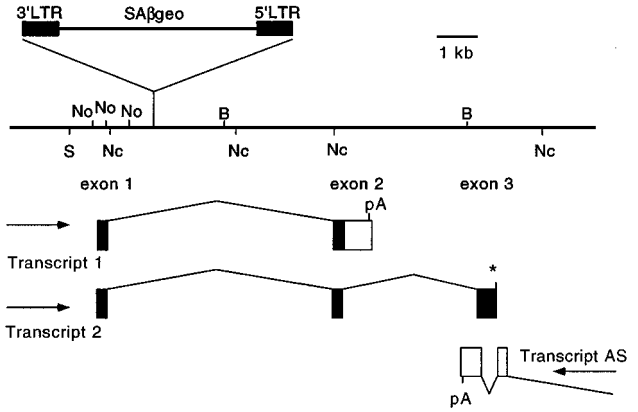

FIG. 5. Map of the ROSA26 genomic region. The line indicates the genomic DNA drawn with the scale and restriction enzyme sites as indicated. The position and orientation of the ROSAβgeo retroviral integration is indicated as are the positions of the exons of the three transcripts that have been mapped. Exons are indicated by rectangles, arrows indicate the direction of transcription of the three transcripts, and lines indicate the splicing patterns for the three transcripts. The 3′ end of exon 2 of transcript 1 and exon 3 of transcript 2 have not yet been determined. B, *Bgl*II; Nc, *Nco*I; No, *Not*I; S, *Sal*I; pA, poly(A) addition signal.

and the promoterless construct and mock-electroporated ES cells produced almost no β-gal activity. All ROSA26 promoter fragments tested had promoter activity in ES cells albeit at lower levels than observed with the PGK promoter. This might be due to position effects on integration of the transgene, as ES cells isolated from ROSA26 mice were found to have 3-fold more β-gal activity than was observed with the PGK promoter. Removal of the potential translation start site improved the expression of β-gal. Moreover, the 1-kb promoter-containing fragment has been found to direct high-level widespread expression of a reporter gene in transgenic mice (E. P. Sandgren, personal communication).

DISCUSSION

The βgeo reporter gene is ubiquitously expressed during embryonic development and in all hematopoietic cells of ROSA26 mice. This makes a useful marker in chimera and transplantation experiments. As described (10), cytoplasmic β-gal activity provides an excellent marker for the *in situ* labeling of expressing cells and has been used to achieve broad expression in transgenic mice (10, 11). However, it is not uncommon with transgenic mice to observe variation or lack of penetrance of transgene expression (12). We have observed virtually complete penetrance and no variation of β-gal expression in ROSA26 mice. This may be due to the trapping of an endogenous gene programmed to be expressed ubiquitously. Confidence in the fidelity of ROSA26 βgeo expression makes it an excellent choice as a marker line or for achieving ubiquitous expression of other genes or cDNAs.

Because of the ubiquitous expression of β-gal in hematolymphoid cells, which can be monitored in conjunction with antibody stains on the FACS, these mice offer a powerful new tool for the study of hematolymphoid stem/progenitor cell populations. Most studies of these cell types have used the Ly5 allotype congenic mouse strains, Ly5.1 and Ly5.2. However, ROSA26 mice may offer some advantages, as β-gal expression is also found in erythroid cells, except for mature erythrocytes. In addition, β-gal may also be expressed in other cell types (e.g., dendritic cells) that have only recently begun to be appreciated as having hematopoietic origin or whose origins do not lend themselves to analysis on the basis of a Ly5 allotype. Hence, the ROSA26 strain should permit donor–host

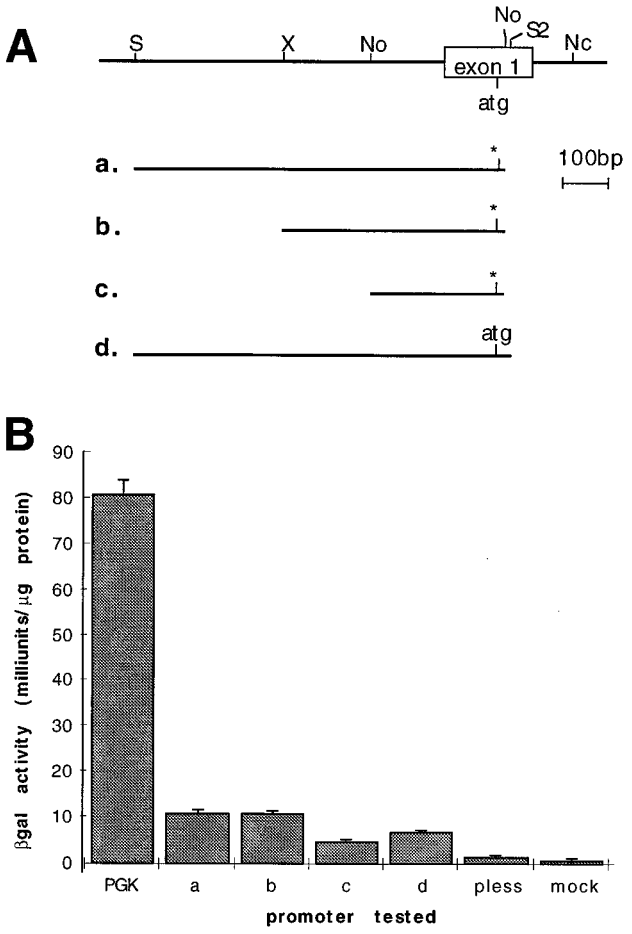

FIG. 6. Characterization of the ROSA26 promoter. (*A*) Map of the genomic DNA at the 5′ end of exon 1 of transcripts 1 and 2. The relative position of fragments used for promoter activity assay is shown. (*a–c*) A potential translation start site in exon 1 (atg) has been converted to a *Bam*HI site (asterisk). Nc, *Nco*I; No, *Not*I; S, *Sal*I; S2, *Sac*II; X, *Xho*I. (*B*) The various promoter fragments (bars a–d) were fused to *lacZ*, the constructs were electroporated into ES cells, and β-gal activity was assayed. The PGK promoter (PGK) was used as a positive control and no fragment (pless) and mock-transfected cells (mock) were used as negative controls.

origin discrimination via FACS analysis as well as via X-Gal staining of tissue sections.

Although there is no apparent defect in heterozygous or homozygous ROSA26 mice, homozygotes were recovered in fewer numbers. The higher level of βgeo expression in homozygotes may result in this poor recovery. Alternatively, the noncoding transcripts may provide a function, either in regulating the expression of transcript AS RNA or protein or by providing a cellular function completely separate from any regulation of the coding transcript or protein. In prokaryotes, convergent and overlapping antisense transcripts originating from a single genomic locus may allow the encoding of more genetic information into small genomes or to regulate gene expression (13). The pattern of ROSA26 transcription is similar to a growing number of eukaryotic genes that produce natural convergent and antisense transcripts (14–20). While there is limited data suggesting a role for these transcription patterns in gene regulation (19, 20), in many cases it remains unclear whether antisense transcripts have any functional significance (21). Determining whether antisense transcription is present and if the transcripts are conserved at the human ROSA26 region may provide some insight into their importance.

We thank Ron Davis, Grant MacGregor, and Eric Sandgren for communicating unpublished results. B.P.Z. was the recipient of a National Institutes of Health postdoctoral fellowship. This work was supported by grants from the National Institute of Child Health and Human Development and the Markey Molecular Medicine Center to P.S.

1. Gossler, A. & Zachgo, J. (1993) in *Gene Targeting: A Practical Approach*, ed. Joyner, A. L. (Oxford Univ. Press, New York), pp. 181–227.
2. Friedrich, G. & Soriano, P. (1991) *Genes Dev.* **5,** 1513–1523.
3. MacGregor, F. R., Zambrowicz, B. P. & Soriano, P. (1995) *Development (Cambridge, U.K.)* **121,** 1487–1496.
4. Kerr, W. G., Nolan, G. P., Serafini, A. T. & Herzenberg, L. A. (1989) *Cold Spring Harbor Symp. Quant. Biol.* **54,** 767–776.
5. Chen, Z. (1996) *Trends Genet.* **12,** 87–88.
6. Frohman, M. A. (1994) *PCR Methods and Applications* **4,** S40–S58.
7. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
8. Chen, Z.-F. & Behringer, R. R. (1995) *Genes Dev.* **9,** 686–699.
9. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
10. Beddington, R. S. P., Morgenstern, J., Land, H. & Hogan, A. (1989) *Development (Cambridge, U.K.)* **106,** 37–46.
11. Suemori, H., Kadodawa, Y., Goto, K., Araki, I., Kondoh, H. & Nakatsuji, N. (1990) *Cell Differ. Dev.* **29,** 181–186.
12. Palmiter, R. D. & Brinster, R. L. (1986) *Annu. Rev. Genet.* **20,** 465–99.
13. Simons, R. W. & Kleckner, N. (1988) *Annu. Rev. Genet.* **22,** 567–600.
14. Khochbin, S. & Lawrence, J.-J. (1989) *EMBO J.* **8,** 4107–4114.
15. Kimelman, D. & Kirschner, M. W. (1989) *Cell* **59,** 687–696.
16. Murphy, P. R. & Knee, R. S. (1994) *Mol. Endocrinol.* **8,** 852–859.
17. Kindy, M. S., McCormack, J. E., Buckler, A. J., Levine, R. A. & Sonenshein, G. E. (1987) *Mol. Cell. Biol.* **7,** 2857–2862.
18. Krystal, G. W., Armstrong, B. C. & Battey, J. F. (1990) *Mol. Cell. Biol.* **10,** 4180–4191.
19. Hildebrandt, M. & Nellen, W. (1992) *Cell* **69,** 197–204.
20. Tasheva, E. & Roufa, D. J. (1995) *Genes Dev.* **9,** 304–316.
21. Miyajima, N., Horiuchi, R., Shibuya, Y., Fukishige, S., Matsubara, K., Toyoshima, K. & Yamamoto, T. (1989) *Cell* **57,** 31–39.