FIGURE 3.—Analyses of *Skt* transcripts. (A–C) RT–PCR analyses using E10.5 embryos to detect *Skt* transcripts in wild-type (+/+) and *Skt*^Gt/Gt embryos. The transcripts containing nucleotide sequences upstream of the insertion site of the trap vector were detected in both wild-type and *Skt*^Gt/Gt embryos (A). The transcripts containing nucleotide sequences downstream of the *Skt* sequence were not detected in *Skt*^Gt/Gt embryos (B and C). M, molecular marker. (D) Northern blot analyses to detect *Skt* mRNA in the wild-type ES cells, E10.5 embryo, and 8-week-old mice, using the *Skt*-specific probe in the 5'-region (see Figure 2B). Total RNA (10 µg) from TT2 ES cells, wild-type E10.5 embryos, mRNA (5 µg) from wild-type organs, and a *Skt* RNA probe were used for Northern blotting. (E) Northern blot analyses to detect *Skt* and β-*geo* fusion transcripts. Total RNA (20 µg) from TT2 and heterozygous (Gt/+) ES cells, wild-type (+/+), heterozygous (Gt/+), and homozygous (Gt/Gt) adult brains was used for Northern blotting. The *Skt* RNA probe or *lacZ* RNA probe was used in the left and right panels, respectively. Br, brain; H, heart; K, kidney; T, testis; Li, liver; Lu, lung; I, intestine.

(amino acid residues 626–656) as determined by analysis using Lupas's algorithm (LUPAS *et al*. 1991). An ATG codon is located at nucleotide positions 559–561. This codon is most likely the initiation codon, because there is a Kozak sequence surrounding this ATG codon (KOZAK 1996): *i.e.*, there is a G residue following the ATG codon and an A residue three nucleotides upstream. A BLAST search of the amino acid sequence deduced from the *Skt* cDNA sequence revealed 80.6% homology with an uncharacterized human protein, KIAA 1217 (accession no. AB033043), and no other evolutionarily conserved protein was identified.

Sequence comparison of *Skt* cDNA with the murine genome sequence in the public Mouse Genome Resources (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/) revealed that the *Skt* gene consists of 19 exons spanning >300 kb and that the ATG codon is located in the second exon (Figure 2B). The trap vector was integrated in the 14th intron, resulting in the disruption of the Skt protein at position 998 (arrowhead in Figure 2, A and B). Sequence analyses of fusion transcripts revealed the presence of two fusion transcripts with the β-*geo* sequence: *Skt-a*, containing the 1st–14th exons, and *Skt-b*,

lacking 33 bp of the 13th exon from *Skt-a* (Figure 2B). In any case, a truncated protein lacking 355 amino acids encoded by exons 15–19 in the C-terminal region is expected to be produced from the *Skt*^a allele.

RT–PCR and Northern blot analyses: We performed RT–PCR using wild-type and *Skt*^Gt/Gt embryos at E10.5. RT–PCR with primers a and b located within the 5'-region of the integration site and detected the expected band in both embryos (Figure 3A). Using RT–PCR with two primer pairs—c in the 5'-region and d in the 3'-region of the integration site and primers e and f within the 3'-region of the integration site—we could not detect any product in *Skt*^Gt/Gt embryos (Figure 3, B and C), indicating the absence of *Skt* mRNA containing exons 15–19 in *Skt*^Gt/Gt embryos.

To analyze the size and expression pattern of *Skt* mRNA, we performed Northern blot analysis using the *Skt*-specific probe in the 5'-region (see Figure 2B). As shown in Figure 3D, a major band of 8 kb and a minor band of 6.5 kb were detected in wild-type ES cells. In wild-type whole embryos at E10.5, the 8- and 6.5-kb bands were also detected, although the expression levels were low. Surprisingly, Northern blot analysis using mRNA

from wild-type adult organs revealed the presence of four different mRNA transcripts, 5.5, 6.5, 7.0, and 8.0 kb. The faint 7.0-kb and strong 6.5-kb bands in the brain, the 7.0- and 5.5-kb bands in the heart, the 7.0-kb band in the testis, the 8.0- and the 5.5-kb bands in the lung, and the 8.0-kb band in the intestine were detected (Figure 3D). We then examined the presence of the *Skt* and β-*geo* fusion mRNA (Figure 3E). In $Skt^{Gt/+}$ ES cells, an expected 10.5-kb band representing the fusion transcript containing β-*geo* was detected with both the *Skt* and *lacZ* probes, although the intensity was weaker than that of the 8-kb band of the endogenous transcript. In the $Skt^{Gt/+}$ and $Skt^{Gt/Gt}$ adult brains, two bands of 10.5 and 9.5 kb were detected with the *Skt* probe (Figure 3E, left), and these bands were also hybridized with the *lacZ* probe (Figure 3E, right), confirming that both transcripts corresponded to the fusion mRNA containing β-*geo*. This result indicates the existence of alternative splicing in the upstream region of the insertion site. Northern blot analysis using mRNA from wild-type adult heart revealed the presence of two different transcripts, the 7- and 5.5-kb bands when the *Skt* probe was used. However, in the Skt^{Gt} adult heart, one 9.5-kb band was detected with the *lacZ* probe (data not shown). This result indicates that alternative splicing in the heart may occur in the downstream region of the insertion site, leading to the production of the 7- and 5.5-kb bands. The largest size of mRNA is 8 kb, which is larger than that of the predicted cDNA sequence comprising 5930 nucleotides (accession no. AB125594). As described below, the anti-Skt antibodies recognized only an ~150-kDa protein corresponding to the predicted molecular weight of 147 kDa in extracts from the nucleus pulposus of the caudal IVDs (see lane 4 in Figure 6A). Thus, the 8-kb mRNA may contain untranslated regions at both the 5'- and 3'-ends and splicing may occur in each untranslated region. Further study will be required to determine the untranslated regions.

Formation of vertebral column and β-*geo* expression in Skt^{Gt} mice: We examined formation of the vertebral column in Skt^{Gt} mice during the embryonic, fetal, and postnatal periods (E8.0, E8.5, E9.0, E9.5, E10.5, E11.5, E12.5, E13.5, E14.5, 15.5, E16.5, E17.5, E18.5, E19.5, newborn, and 2 weeks of age) by histological analysis with and without X-gal staining. Before E8.0, the β-*geo* gene was expressed in the chorion, but not in the embryo (Figure 4A). At E8.5 (Figure 4B) and E9.0 (Figure 4C), intense staining was detected in the notochord. At E11.5, the notochord and the mesonephros expressed β-*geo* strongly (Figure 4, D–F). In addition, sections of E11.5 embryos showed β-*geo* expression in the epithalamus sulcus, roof of the neopallial cortex, lens vesicle, inner layer of retina, heart (atrium and ventricle), surface of hepatic primordium, infundiblum, surface ectoderm, hind gut, and mesenchyme of the limb bud (data not shown). No abnormality was found in both vertebral and intervertebral regions up to E16.5. In the $Skt^{Gt/Gt}$

embryo at E17.5, compression was observed in some IVDs (Figure 4, I and J) of the $Skt^{Gt/Gt}$ embryo, but not in the wild-type embryo (Figure 4, G and H). In the $Skt^{Gt/Gt}$ neonate, the X-gal-positive cells in the nucleus pulposus were shifted to the periphery and the alignment of the vertebral bodies was undulated (Figure 4, M and N). At this stage, sizes of IVDs are the same as those of wild-type mice. In 2-week-old $Skt^{Gt/Gt}$ mice, the X-gal-positive notochord cells were dislocated to the left or right side and the nuclei pulposi were smaller and eccentric, resulting in the disappearance of normal IVDs (Figure 4, Q and R). In the $Skt^{Gt/+}$ embryo and mice, the X-gal-positive cells were positioned in the center of the vertebrae at any stage (Figure 4, G, K, and O). Interestingly, the expression of the reporter gene was much lower in the upper IVDs than in the fifth to seventh caudal IVD, coinciding with a relationship between the phenotype of the $Skt^{Gt/Gt}$ and strong expression in the caudal region of the tail.

Histological analyses of IVDs of $Skt^{Gt/Gt}$ and *Sd* mutant adult mice: In $Skt^{Gt/Gt}$ adults, the β-*geo* gene was also expressed in many tissues such as the corpus callosum in the brain, uriniferous tubules in kidney, cardiac muscle, Sertoli's cells in testes, and basal cells and outer root sheaths of hair follicles in skin (data not shown).

In normal adult mice, the nucleus pulposus was located in the center of IVDs at all levels of the vertebral column (Figure 5, A, D, G, and J). However, the nucleus pulposus in the IVDs of the tail region of $Skt^{Gt/Gt}$ mice contained an aggregation of notochord-like cells with fewer vacuoles than normal and were dislocated to the periphery (Figure 5, H and K). The nucleus pulposus in the upper regions appeared normal (Figure 1, B and E). Although histochemical analyses of embryos and new-borns in the *Sd* mutant were reported by PAAVOLA *et al.* (1980) and THEILER (1988), the histochemical analyses of the IVDs in *Sd* mutant adult mice have not previously been performed in detail. Thus, we analyzed the IVDs of the whole spine in *Sd* +/+ + mutant mice. Surprisingly, the IVDs were totally occupied by peripheral fibers similar to those seen in the annulus fibrosus and no nucleus pulposus was found within the IVDs (Figure 5, C and F). The degeneration of the nucleus pulposus in the center of caudal IVDs was occasionally observed (Figure 5, I and L). Although tail kinks of $Skt^{Gt/Gt}$ mice were restricted to the 20–25th caudal vertebrae (Figure 1B), an irregular boundary with direct contact between the nucleus pulposus and annulus fibrosus (Figure 5, O and P) was observed in the 5–25th caudal IVDs of $Skt^{Gt/Gt}$ mice, compared to the sharp boundary and fixed space observed between the nucleus pulposus and the fibrous layers of the annulus fibrosus in normal mice (Figure 5, M and N). In addition to the abnormalities of the nucleus pulposus in IVDs, the annulus fibrosus development was also impaired in $Skt^{Gt/Gt}$ mice as demonstrated by the thin fibrous layers of annulus fibrosus and the failure of fibrous adhesion in the vertebral bodies (Figure 5, Q and R). We also examined whether similar abnormalities were
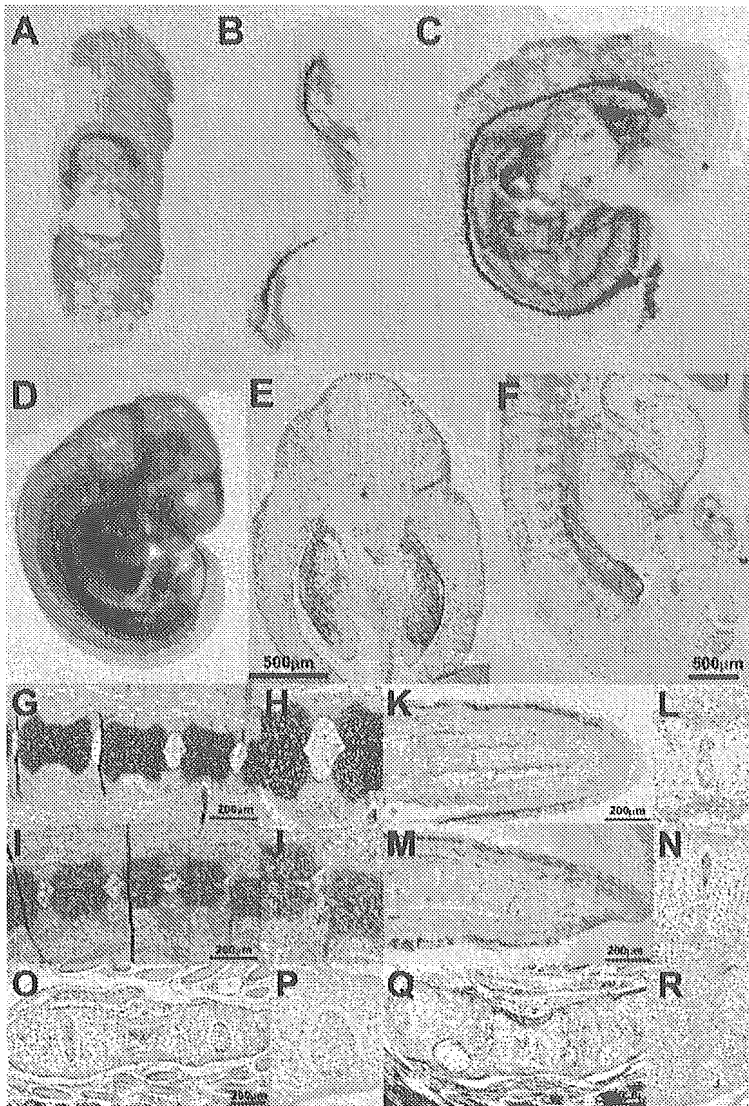
FIGURE 4.—β-*gal* expression and histological analyses in *Skt*^Gt^ mice. At E7.5 (A), the chorion was stained with X-gal, but the midline region of the embryo was not stained. At E8.5 (B) and E9.0 (C) intense staining was detected in the notochord. At E11.5 (D–F), the notochord and the mesonephros expressed β-*geo* strongly in whole-mount X-gal staining (D), the frontal section (E), and the sagittal section (F). Sagittal sections of the tail bud of *Skt*^Gt/+^ (G and H) and *Skt*^Gt/Gt^ (I and J) embryos at E17.5. At E17.5, some IVDs were compressed in the tail bud of *Skt*^Gt/Gt^ (J). Sagittal sections of the tail tips of newborn *Skt*^Gt/+^ (K and L) and *Skt*^Gt/Gt^ (M and N) mice. In the *Skt*^Gt/Gt^ neonate, the vertebral body alignment was undulated (M and N). Sagittal sections of the tail tips of 2-week-old *Skt*^Gt/+^ (O and P) and *Skt*^Gt/Gt^ (Q and R) mice. In the 2-week-old *Skt*^Gt/Gt^ mice, the X-gal-positive nuclei pulposi were dislocated to the periphery (Q and R). (H, J, L, N, P, and R) Higher magnification of the area indicated by the boxes in G, I, K, M, O, and Q, respectively. Sections (G–J) were stained with alcian blue and sections of X-gal staining (E, F, and K–R) were counterstained with Nuclear Fast red. Bars, 200 μm.

observed in the regions that did not have kinks in *Skt*^Gt/Gt^ mice (Table 1). Surprisingly, histological analysis of externally normal tails of *Skt*^Gt/Gt^ mice revealed the presence of similar IVD abnormalities such as dislocation of the nucleus pulposus and thin fibrous layers of the annulus fibrosus (Figure 5S) as found in kinked tails (Figure 5, Q and R). This IVD phenotype was observed over 10 generations in the progeny backcrossed to C57BL/6. Thus, the primary phenotype of *Skt*^Gt/Gt^ mice is the deformity of IVDs in the tail region. These histological pictures were clearly distinct from those of *Sd* mice.

Analysis of Skt protein using anti-Skt antibody: The anti-Skt antibodies recognized an ~150-kDa protein corresponding to the predicted molecular weight of 147 kDa in the lysates of BMT10 cells transiently expressing the *Skt* cDNA by the CAG promoter as well as in extracts from the nucleus pulposus of the caudal IVDs. No band was detected in extracts from untreated BMT10 cells and BMT10 cells transfected with mock expression vector

(Figure 6A). This is consistent with the notion that ATG at positions 559–561 and TAA at positions 4615–4617 in the *Skt* cDNA are the start codon and termination codon, respectively. As expected, the amount of Skt protein was decreased in *Skt*^Gt/+^ mice and was below the detectable level in *Skt*^Gt/Gt^ mice (Figure 6A). After immunohistochemical staining, the nucleus pulposus cells from upper caudal vertebral discs were positively stained in wild-type mice (Figure 6B, a and c), but not in *Skt*^Gt/Gt^ mice (Figure 6B, b and d). At higher magnification, the staining was observed mainly in the cytoplasm of the nucleus pulposus cells, indicating cytoplasmic localization of the Skt protein. Since we could not produce antibodies against the N terminus, the expression of the truncated protein was not confirmed in *Skt*^Gt/Gt^ mice.

Relationship of Skt with Etl4^lacZ^ or Sd loci: Both the expression pattern of the *Skt* gene and the phenotype of *Skt*^Gt^ mutant mice were quite similar to those of *Etl4*^lacZ^ mutant mice (ZACHGO *et al.* 1998). Furthermore, by examining
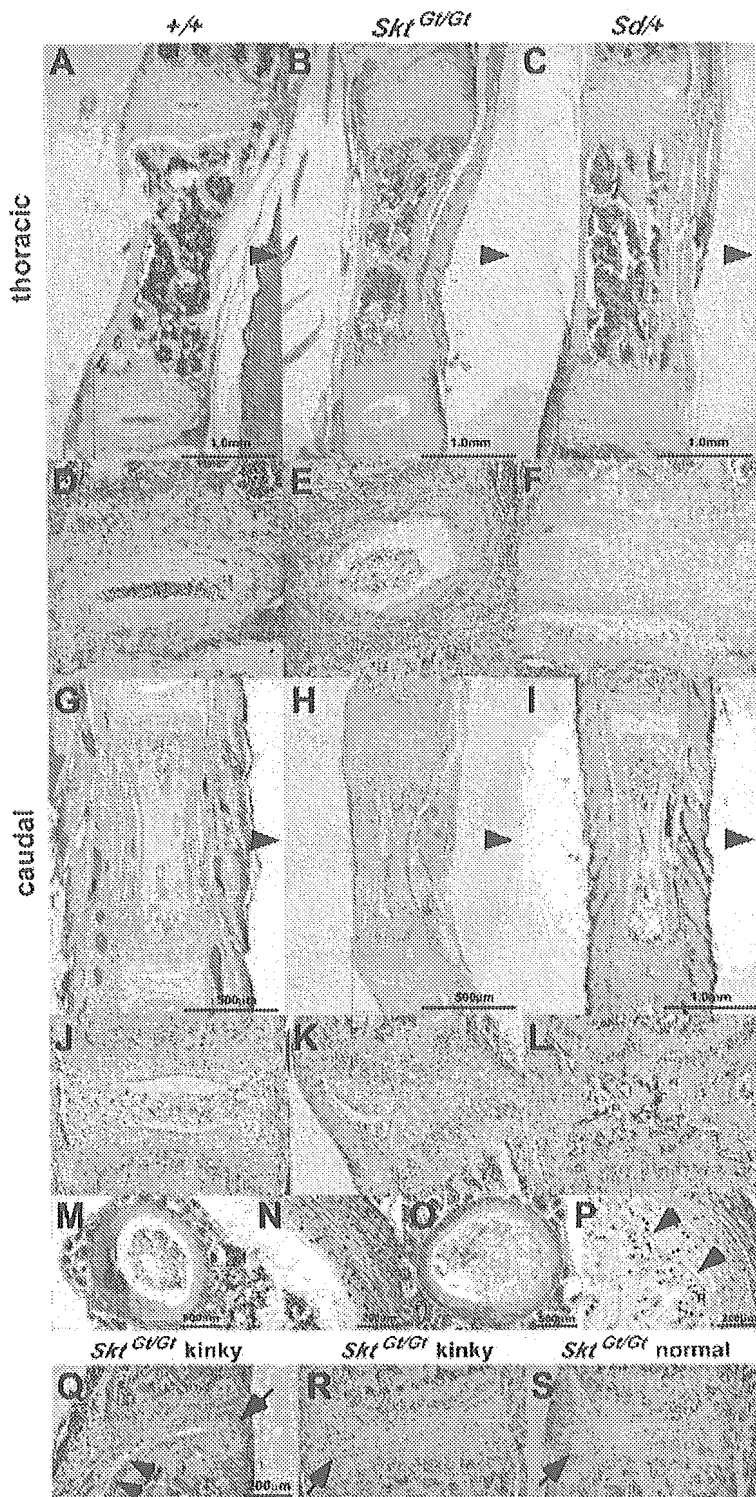
FIGURE 5.—Histological analyses of IVDs of $Skt^{Gt/Gt}$ and Sd mutant adult mice. Sagittal sections of the thoracic and caudal IVD from 8-week-old adult wild-type (A, D, G, and J), $Skt^{Gt/Gt}$ (B, E, H, and K), and heterozygous Sd mice (C, F, I, and L). (D, E, F, J, K, and L) Higher magnification of the area indicated by the boxes in A, B, C, G, H, and I, respectively. Arrowheads indicate the dorsal side. Axial sections of the upper caudal IVD in wild-type (M and N) and $Skt^{Gt/Gt}$ (O and P) 8-week-old mice. (N and P) Higher magnification of the area indicated by the boxes in M and O, respectively. The arrowheads in P indicate an irregular boundary with close contact between the nucleus pulposus and annulus fibrosus. (Q–S) The 20–25th caudal IVDs of $Skt^{Gt/Gt}$ 8-week-old mice. Impaired development of the annulus fibrosus in $Skt^{Gt/Gt}$ mice was demonstrated by the thin fibrous layers of annulus fibrosus (Q and R) and the failure of fibrous adhesion to the vertebral bodies (arrowhead in Q). Similar IVD abnormalities such as dislocation of the nucleus pulposus (arrow in Q–S) and impaired growth of the annulus fibrosus were observed in the non-kinked regions (S) and in the kinked regions (Q and S). Haematoxylin and eosin (HE) staining was used. Bars, 200 μm.

the reported primer sequences of the $Etl4^{lacZ}$ locus (MAATMAN *et al.* 1997), we found that the $Etl4^{lacZ}$ locus is located in the third intron of the Skt gene (Figure 2B) and that the distance between the integration sites of $Etl4^{lacZ}$ and Skt^{Gt} was 237 kb.

To examine the genetic distance and interaction between the Skt^{Gt} and Sd locus, we crossed Skt^{Gt} mice with Sd mice to produce the compound heterozygote ($Sd +/+$ Skt^{Gt}; *trans*-configuration). Then, these compound heterozygotes were used to analyze the recombination rate
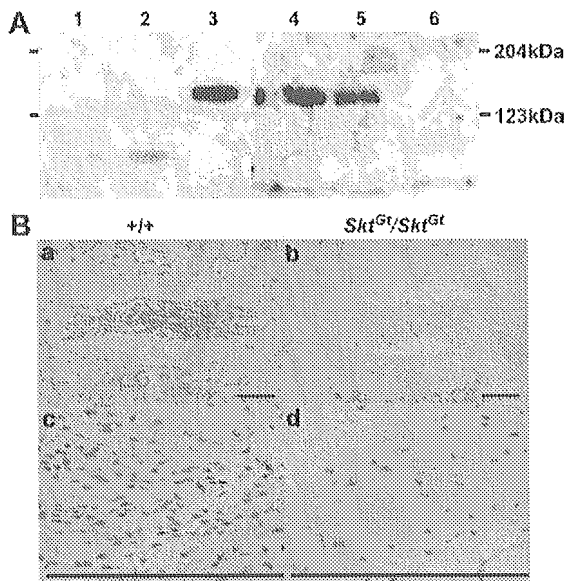
FIGURE 6.—Detection of the Skt protein. (A) Western blot analysis to detect Skt protein using extracts from untreated BMT10 cells (lane 1), BMT10 cells transfected with vector (lane 2), BMT10 cells transfected with the *Skt* expression vector (lane 3), and extracts of the nucleus pulposus of caudal IVDs from 8-week-old wild-type mice (lane 4), $Skt^{Gt/+}$ mice (lane 5), and $Skt^{Gt/Gt}$ mice (lane 6). An ~150-kDa protein corresponding to the predicted molecular weight of 147 kDa was detected in lanes 3, 4, and 5, but not in lane 6. The amount of Skt protein was reduced in the $Skt^{Gt/+}$ mutant (lane 5) and was below the detectable level in $Skt^{Gt/Gt}$ (lane 6). (B) Immunohistochemistry of frontal sections of the nucleus pulposus in upper caudal IVDs from adult 8-week-old mice using purified anti-Skt antibodies. Skt protein was detected in the cytoplasm of nucleus pulposus cells in wild-type (a and c), but not $Skt^{Gt/Gt}$ mice (b and d). (c and d) Higher magnification of the area indicated by the boxes in a and b, respectively. Bars, 200 μm.

between the two loci. As shown in Table 2, one compound heterozygote carrying the *Sd* mutation and the Skt^{Gt} insertion on the same chromosome ($Sd Skt^{Gt}/+ +$; *cis*-configuration) and two wild-type mice were obtained among 249 mice obtained by mating the *trans* compound heterozygote with the wild-type C57BL/6 mice, demonstrating that the *Sd* and Skt^{Gt} mutations were genetically separated. The genetic distance was calculated to be ~0.95 cM by combining the data from mating the [$Sd Skt^{Gt}/+ +$] (*cis*) compound heterozygote with the wild-type mice (Table 2).

To evaluate the effect of the Skt^{Gt} mutation on the tail phenotype in heterozygous *Sd* mice in either the *trans*- or the *cis*-configuration, the number of vertebrae was determined by X-ray analysis. Heterozygous *Sd* [$Sd +/+ +$] mice ($n = 22$) had a variable number of vertebrates and the vertebral columns were truncated at the sixth caudal vertebral body on average. Both *trans* [$Sd +/+ Skt^{Gt}$] ($n = 23$) and *cis* [$Sd Skt^{Gt}/+ +$] ($n = 10$) compound heterozygous mice had shorter tails, in which the vertebral columns were truncated at the second and third caudal vertebral body on average (Figure 7A). We examined whether the phenotype of [$Sd Skt^{Gt}/+ Skt^{Gt}$] neonatal mice is more severe than those of *trans* [$Sd +/+ Skt^{Gt}$] and *cis* [$Sd Skt^{Gt}/+ +$] compound heterozygous mice. Approximately 80% of [$Sd Skt^{Gt}/+ Skt^{Gt}$] neonatal mice died within 48 hr of birth and all mutant mice died within 2 weeks. In addition, as shown by making alcian blue/alizarin red whole-mount preparations of [$Sd Skt^{Gt}/+ Skt^{Gt}$] neonatal mice, [$Sd Skt^{Gt}/+ Skt^{Gt}$] neonatal mice had shorter tails than those of *trans* [$Sd +/+ Skt^{Gt}$] and *cis* [$Sd Skt^{Gt}/+ +$] compound heterozygous mice, in which the vertebral columns were truncated at the fourth sacral vertebral body on average, suggesting a cumulative effect of the Skt^{Gt} mutation on the *Sd* mutant.

To examine the pathologic effect of the *Skt* mutation on the *Sd* phenotype, we carried out histological analyses on the IVDs of *Sd* $+/+ Skt^{Gt}$ and *Sd* $Skt^{Gt}/+ +$ mutant mice. Both compound heterozygous mice showed IVD histology similar to that seen in *Sd* $+/+ +$ mutant mice. The nucleus pulposus was totally absent and replaced by peripheral fibers similar to those seen in the annulus fibrosus in all IVDs (Figure 7B, a–h). These results suggest that the *Skt* mutation did not affect the histological picture of IVD in the *Sd* mutation. In addition, the expression pattern of the β-*geo* gene in the tail notochord was the same in both *trans* [$Sd +/+ Skt^{Gt}$] and *cis* [$Sd Skt^{Gt}/+ +$] compound heterozygous embryos at E9.5 and E13.5 (Figure 7C). The notochord of both *trans* and *cis* compound heterozygous embryos was thin and clearly stained at E9.5 (Figure 7C, a and d), and the X-gal-positive notochord was similarly fragmented in both *trans* and *cis* compound heterozygous embryos at E13.5 (Figure 7C, b, c, e, and f). This suggested that the *Sd* mutation did not affect *Skt* expression. In addition, this *trans–cis* test demonstrated that the double heterozygotes exhibited indistinguishable phenotypes in

TABLE 2

Distribution of the haplotypes among the 315 offspring of the backcross *Sd* $+/+ Skt^{Gt}$ or *Sd* $Skt^{Gt}/+ +$ × C57BL/6

	Sd $+/+ Skt^{Gt}$ × C57 BL/6				*Sd* $Skt^{Gt}/+ +$ × C57 BL/6		
$+ +/+ Skt^{Gt}$	*Sd* $+/+ +$	*Sd* $Skt^{Gt}/+ +$	$+ +/+ +$	$+ Skt^{Gt}/+ +$	*Sd* $+/+ +$	*Sd* $Skt^{Gt}/+ +$	$+ +/+ +$
144	102	1	2	0	0	24	42

Sd, Skt^{Gt} compound heterozygotes were generated from crosses with an inbred laboratory strain (C57BL/6)
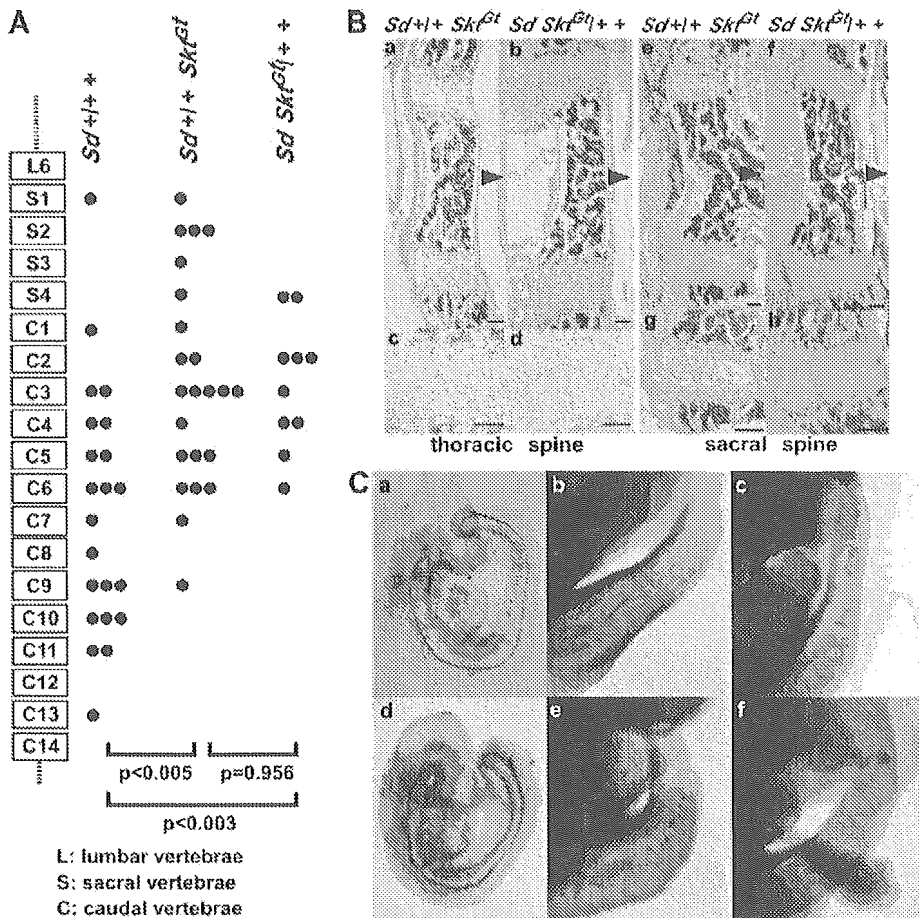
A

	Sd+/+	*Sd+/+ Skt^{Gt}*	*Sd Skt^{Gt}+ +*
L6			
S1	●	●	
S2		●●●	
S3		●	
S4		●	●●
C1	●	●	
C2		●●	●●●
C3	●●	●●●●●	●
C4	●●	●	●●
C5	●●	●●●	●
C6	●●●	●●●	●
C7	●	●	
C8	●		
C9	●●●	●	
C10	●●●		
C11	●●		
C12			
C13	●		
C14			

$p<0.005$ $p=0.956$

$p<0.003$

L: lumbar vertebrae
S: sacral vertebrae
C: caudal vertebrae

B *Sd+/+ Skt^{Gt} Sd Skt^{Gt}+ +*   *Sd+/+ Skt^{Gt} Sd Skt^{Gt}+ +*

thoracic spine   sacral spine

C

FIGURE 7.—(A) Schematic of axial levels and severity of vertebral malformations in compound mutant 8-week-old mice. A single solid circle indicates the level of the terminal vertebral body for a single mouse. In both *trans* and *cis* compound mutant mice, the degree of vertebral malformation is more severe than that in heterozygous *Sd* mice (*trans*, $P < 0.005$; *cis*, $P < 0.003$, Mann–Whitney *U*-test). There was no significant difference between *trans* [*Sd* +/+ *Skt^{Gt}*] ($n = 23$) and *cis* [*Sd Skt^{Gt}/+* +] ($n = 10$) mice ($P = 0.956$, Mann–Whitney *U*-test). (B) Histological analyses of vertebral columns in the *Sd* +/+ *Skt^{Gt}* and *Sd Skt^{Gt}/+* + mutant mice. HE staining of midsagittal sections in thoracic spines of *Sd* +/+ *Skt^{Gt}* (a and c) and *Sd Skt^{Gt}* /+ + (b and d) mice and in sacral spines of *Sd* +/+ *Skt^{Gt}* (e and g) and *Sd Skt^{Gt}/+* + (f and h) mice. Arrowheads indicate dorsal sides. (c, d, g, and h) Higher magnification of the areas indicated by the boxes in a, b, e, and f, respectively. (C) Whole-mount X-gal staining in the *Sd* +/+ *Skt^{Gt}* and *Sd Skt^{Gt}/+* + mutant embryos. The β-*gal* expression in the tail notochord of *trans*-heterozygous embryos [*Sd* +/+ *Skt^{Gt}*] at E9.5 (a) and E13.5 (b and c) and of *cis*-heterozygous embryos [*Sd Skt^{Gt}/+* +] at E9.5 (d) and E13.5 (e and f).

regard to notochord degradation whether *Sd* and *Skt^{Gt}* were in a *trans-* or in a *cis*-configuration.

DISCUSSION

We established a new recessive trap line, B6;CB-*Skt^{GtAyu802UMEG}*, which had a deformity in caudal IVDs. The insertion site of the trap vector is in the 14th intron of the novel gene *Skt*, located on chromosome 2, near the locus for *Danforth's short tail*. In addition, we found that the enhancer trap locus *Etl4^{lacZ}*, which was previously reported to be an allele of *Sd*, was located in the third intron of the *Skt* gene.

Structure, expression, and function of the *Skt* gene: The sequence of the trapped gene, *Skt*, was obtained using the gene-trap ES clone. The *Skt* gene contains 19 exons encoding a novel protein of 1352 amino acids with a proline-rich region in the N terminus (amino acid positions 298–364) and a coiled-coil domain in the middle (amino acid positions 626–656). Although the role of the proline-rich region in many proteins is not

clear yet, a proline-rich region located in the amino-terminal region has been shown to be important for proper folding in cytochrome P450s (mitochondrial, microbial, and microsomal P450s) (KUSANO *et al.* 2001a,b). Thus, the proline-rich region in the Skt protein may have a similar function in protein folding. The coiled-coil motif was first described by CRICK (1952) and by PAULING and COREY (1953) as the main structural element of a large class of fibrous proteins that included keratin, myosin, and fibrinogen, mediating dimerization, heterodimer formation, or trimerization (for a review see LUPAS 1996). These proteins provide a scaffold for regulatory complexes such as tropomyosin and a protective surface for pathogens. As the Skt protein is localized in the cytoplasm, it is conceivable that the Skt protein may provide structural elements or act as a scaffold for regulatory complexes.

Insertion of the gene-trap vector into the 14th intron of the *Skt* gene, downstream of the coiled-coil domain and the proline-rich region, resulted in production of two fusion transcripts with the β-*geo* sequence: one, *Skt-a*,

containing only exons 1–14 and the other, *Skt-b*, lacking 33 bp of the 13th exon from *Skt-a* (see Figure 2B). Northern blot analysis showed that the amount of fusion transcripts from the trapped allele was lower than that of the wild-type allele. Therefore, it is also possible that the insertion of the gene-trap vector resulted in decreased mRNA stability. As tail phenotypes are recessive, insertion of the trap vector may cause a hypomorphic or null mutation, but not a dominant-negative mutation. Although we detected four types of mRNA transcribed from the wild-type *Skt* allele, further analysis will be required to elucidate the function of these mRNA.

The expression patterns of the *Skt* gene can be monitored by X-gal staining, because the expression of the reporter gene, β-*geo*, is under control of the regulatory region of the *Skt* gene. At E11.5, β-*geo* was expressed mainly in the notochord and mesonephros at high levels, and in other tissues as described in the RESULTS. In the *Etl4^{lacZ}* line, the *lacZ* reporter gene was also expressed in two main tissues, the notochord and the mesonephros (ZACHGO *et al.* 1998). In both lines, *lacZ* expression was first detected at E8.5 in the presumptive notochord cells. In the *Etl4^{lacZ}* line, *lacZ* expression was detected in the future IVDs at E13.5 and persisted up to E14.5. However, no *lacZ* expression was detected in the IVDs of newborn and adult mice. On the other hand, *lacZ* expression in the IVDs persisted to adult stage in the *Skt^{Gt}* line. These results suggest that the enhancer located near the *Etl4^{lacZ}* locus is not sufficient to express the *lacZ* gene in the IVDs of adult mice.

Histochemical analyses of the vertebral column revealed the differences between *Sd* mice and *Skt^{Gt}* mice. As histological data on *Etl4^{lacZ}* mice were not described by ZACHGO *et al.* (1998), it is not clear whether the IVD histology is similar to that seen in *Sd* or *Skt^{Gt}* mice. As reported previously, the development of both the vertebral column and the urogenital system is affected in *Sd* mutation, suggesting that the *Sd* gene is required for formation of derivatives from both the paraxial and the intermediate mesoderm. In addition, both vertebral bodies and IVDs are affected in *Sd* mice. The notochord shows discontinuities as early as E9.5, resulting in the total absence of the nucleus pulposus at all levels and is replaced by peripheral fibers similar to those of the annulus fibrosus in *Sd* adult mice. All vertebral bodies are reduced in a dorso-ventral direction and the number of tail vertebrae is reduced, leading to shortening or absence of the tail. However, in *Skt^{Gt}* mice, the compression and dislocation of the nucleus pulposus were first observed at E17.5 and were restricted to the tail region. Interestingly, the size of the nucleus pulposus was similar to that in wild-type mice until birth. After birth, the nucleus pulposus did not expand and was dislocated to the periphery, resulting in a kinky-tail phenotype in adults. These observations suggest that *Sd* acts at an early stage of mesoderm development involving both sclerotome and notochord development and that *Skt* acts during

the fetal period and at the later stage involving growth and hypertrophy of the nucleus pulposus. Although massive apoptosis in notochord cells was observed in embryos with targeted disruption of *Jun* (BEHRENS *et al.* 2003) or *Sox5^{-/-}/Sox6^{-/-}* (SMITS and LEFEBVRE 2003), no apoptotic notochord cells were observed in the *Skt^{Gt/Gt}* embryos at E16.5 and in neonatal mice. Thus, apoptosis is not the cause of compression or dislocation of the nucleus pulposus in *Skt^{Gt}* mice. As the Skt protein contains the coiled-coil motif that is involved in the formation of mechanically rigid structures, it is possible that the nucleus pulposus lacking Skt protein may not be capable of sustaining mechanical loads, leading to compression or dislocation of the nucleus pulposus.

Although the IVD is formed from two components of developmentally different origins, the nucleus pulposus and the annulus fibrosus, interaction of the nucleus pulposus and annulus fibrosus is not clear yet. In *Sd* mice, disappearance of the notochord cells occurs at early stages of development. THEILER (1988) reported that the fibers of the annulus fibrosus are reduced in *Sd* heterozygous embryos and newborns. However, our results suggest that the annuli fibrosi are not reduced in *Sd* heterozygous adult mice. Thus, the annulus fibrosus may be able to completely compensate for the loss of the nucleus pulposus during growth after birth. In *Skt^{Gt}* mice, the thin annulus fibrosus was observed together with abnormalities in the nucleus pulposus. At present, it is not known whether the thin annulus fibrosus is caused by the direct effect of *Skt* deficiency or indirectly caused by defects in the nucleus pulposus.

The relationship among the *Skt^{Gt}*, *Etl4^{lacZ}*, and *Sd* mutations: Our breeding studies revealed that the genetic distance between the *Skt* and *Sd* loci was 0.95 cM. We believe that the *Skt* gene is distinct from the *Sd* gene for the following reasons. First, anti-Skt antibody detected the predicted size of protein deduced from the *Skt* gene. If the *Skt* was part of the *Sd* gene, we should have detected a larger protein by Western blot analysis. But, we could not detect any such protein. Second, our *trans-cis* test demonstrated that both double heterozygotes exhibited indistinguishable phenotypes, regardless of whether *Sd* and *Skt^{Gt}* were in a *trans-* or a *cis*-configuration. On the basis of the results by ZACHGO *et al.* (1998), *Sd* is a gain-of-function mutation. As the *Skt* is a recessive mutation, producing a hypomorphic or null allele, the *Sd* phenotype is expected to be attenuated in the *cis*-configuration. Therefore, the phenotype in double heterozygotes in the *trans*-configuration might be more severe than that in the *cis*-configuration if the *Skt* gene is part of the *Sd* gene.

We have shown that the *Etl4^{lacZ}* locus is located in the third intron of the *Skt* gene. Interestingly, a sequence that has 93% homology to consensus sequence 3 (CS3) in node/notochord enhancers of *Hnf3β* (NISHIZAKI *et al.* 2001) was found in the fourth intron of the *Skt* gene, located 106 kb downstream of the insertion site of

Etl4^{lacZ}. As the expression patterns of *Etl4^{lacZ}* are quite similar to those in *Skt^{ct}* mice and as *Etl4^{lacZ}* mice have similar kinks in the tail region (ZACHGO *et al.* 1998), the *lacZ* gene in the enhancer trap vector may be expressed under the control of this possible node/notochord enhancer. ZACHGO *et al.* (1998) reported that the *Etl4^{lacZ}* locus is localized ~0.75 cM distal to *Sd* and that the *Sd* phenotype is attenuated when *Etl4^{lacZ}* is present in *cis*. These results suggest that the possible enhancer sequence in the fourth intron of the *Skt* gene functions also as the enhancer for the *Sd* gene and that the insertion of an enhancer trap vector in the third intron, as found in the *Etl4^{lacZ}* line, may eliminate the enhancer function for the *Sd* gene, resulting in the attenuation of the *Sd* phenotype.

Future study on the functions of the Skt and Sd proteins may provide important clues for understanding the mechanisms for the development of notochordal cells and their differentiation into the nucleus pulposus cells.

LITERATURE CITED

ALFRED, J. B., K. RANCE, B. A. TAYLOR, S. J. PHILLIPS, C. M. ABBOTT *et al.*, 1997 Mapping in the region of Danforth's short tail and the localization of tail length modifiers. Genome Res. 7: 108–117.

ALLEN, N. D., D. G. CRAN, S. C. BARTON, S. HETTLE, W. REIK *et al.*, 1988 Transgenes as probes for active chromosomal domains in mouse development. Nature 333: 852–855.

ANG, S., and J. ROSSANT, 1994 HNF-3β is essential for node and notochord formation in mouse development. Cell 78: 561–574.

ARAKI, K., T. IMAIZUMI, T. SEKIMOTO, K. YOSHINOBU, J. YOSHIMUTA *et al.*, 1999 Exchangeable gene trap using the Cre/mutated lox system. Cell. Mol. Biol. 45: 737–750.

ASZODI, A., D. CHAN, E. HUNZIKER, J. F. BATEMAN and R. FASSLER, 1998 Collagen II is essential for the removal of the notochord and the formation of intervertebral discs. J. Cell Biol. 143: 1399–1412.

BEHRENS, A., J. HAIGH, F. MECHTA-GRIGORIOU, A. NAGY, M. YANIV *et al.*, 2003 Impaired intervertebral disc formation in the absence of *Jun*. Development 130: 103–109.

CHIANG, C., Y. LITINGTUNG, E. LEE, K. E. YOUNG, J. CORDEN *et al.*, 1996 Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature 383: 407–413.

CRICK, F. H., 1952 Is alpha-keratin a coiled coil? Nature 170: 882–883.

DUNN, L. C., S. GLUECKSOHN-SCHOENHEIMERL and V. BRYSON, 1940 A new mutation in the mouse affecting spinal column and urogenital system. J. Hered. 31: 343–348.

GERARD, R. D., and Y. GLUZMAN, 1985 New host cell system for regulated simian virus 40 DNA replication. Mol. Cell. Biol. 5: 3231–3240.

GOSSLER, A., A. L. JOYNER, J. ROSSANT and W. C. SKARNES, 1989 Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244: 463–465.

GRUNEBERG, H., 1953 Genetical studies on the skeleton of the mouse. VI. Danforth's short-tail. J. Genet. 51: 317–326.

GRUNEBERG, H., 1958 Genetical studies on the skeleton of the mouse. XXII. The development of Danforth's short-tail. J. Embryol. Exp. Morphol. 6: 124–148.

HOGAN, B., R. BEDDINGTON, F. COSTANTINI and E. LACY, 1994 *Manipulating the Mouse Embryo: A Laboratory Manual*, pp. 379–381. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

KORN, R., M. SCHOOR, H. NEUHAUS, U. HENSELING, R. SONININEN *et al.*, 1992 Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. Mech. Dev. 39: 95–109.

KOZAK, M., 1996 Interpreting cDNA sequences: some insights from studies on translation. Mamm. Genome 7: 563–574.

KUSANO, K, N. KAGAWA, M. SAKAGUCHI, T. OMURA and M. R. WATERMAN, 2001a Importance of a proline-rich sequence in the amino-terminal region for correct folding of mitochondrial and soluble microbial p450s. J. Biochem. 129: 271–277.

KUSANO, K., M. SAKAGUCHI, N. KAGAWA, M. R. WATERMAN and T. OMURA, 2001b Microsomal p450s use specific proline-rich sequences for efficient folding, but not for maintenance of the folded structure. J. Biochem. 129: 259–269.

LANE, P. W., and C. S. BIRKENMEISER, 1993 Urogenital syndrome (*us*): a developmental mutation on chromosome 2 of the mouse. Mamm. Genome 4: 481–484.

LANGMAN, J., 1969 *Medical Embryology*, pp. 134–135. Williams & Wilkins, Baltimore.

LUPAS, A., 1996 Coiled coils: new structures and new functions. Trends Biochem. Sci. 21: 375–382.

LUPAS, A., M. VAN DYKE and J. STOCK, 1991 Predicting coiled coils from protein sequences. Science 252: 1162–1164.

MAATMAN, R., J. ZACHGO and A. GOSSLER, 1997 The Danforth's short tail mutation acts cell autonomously in notochord cells and ventral hindgut endoderm. Development 124: 4019–4028.

NISHIZAKI, Y., K. SHIMAZU, H. KONDOH and H. SASAKI, 2001 Identification of sequence motifs on the node/notochord enhancer of Foxa2 (Hnf3β) gene that are conserved across vertebrate species. Mech. Dev. 102: 57–66.

NIWA, H., K. YAMAMURA and J. MIYAZAKI, 1991 Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108: 193–200.

PAAVOLA, L. G., D. B. WILSON and E. M. CENTER, 1980 Histochemistry of the developing notochord, perichordal sheath and vertebrae in Danforth's short-tail (*Sd*) and normal C57BL/6 mice. J. Embryol. Exp. Morphol. 55: 227–245.

PAULING, L., and R. B. COREY, 1953 Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. Nature 171: 59–61.

RUFAI, A., M. BENJAMIN and J. R. RALPHS, 1995 The development of fibrocartilage in the rat intervertebral disc. Anat. Embryol. 192: 53–62.

SMITS, P., and V. LEFEBVRE, 2003 Sox5 and Sox6 are required for notochord extracellular matrix sheath formation, notochord cell survival and development of the nucleus pulposus of intervertebral discs. Development 130: 1135–1148.

THEILER, K., 1988 Vertebral malformations. Adv. Anat. Embryol. Cell Biol. 112: 1–99.

WILSON, V., L. MANSON, W. C. SKARNES and R. S. BEDDINGTON, 1995 The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. Development 121: 877–886.

ZACHGO, J., R. KORN and A. GOSSLER, 1998 Genetic interactions suggested that Danforth's short tail (Sd) is a gain-of-function mutation. Dev. Genet. 23: 86–96.

Communicating editor: C. A. KOZAK

The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse

Alex S. Nord[1,3], Patricia J. Chang[1], Bruce R. Conklin[2], Antony V. Cox[3], Courtney A. Harper[1], Geoffrey G. Hicks[4], Conrad C. Huang[1], Susan J. Johns[1], Michiko Kawamoto[1], Songyan Liu[4], Elaine C. Meng[1], John H. Morris[1], Janet Rossant[5], Patricia Ruiz[6], William C. Skarnes[3], Philippe Soriano[7], William L. Stanford[8], Doug Stryke[1], Harald von Melchner[9], Wolfgang Wurst[10], Ken-ichi Yamamura[11], Stephen G. Young[12], Patricia C. Babbitt[1] and Thomas E. Ferrin[1,*]

[1]University of California, San Francisco, 600 16th Street, San Francisco, CA 94143-2240, USA, [2]Gladstone Institute of Cardiovascular Disease, University of California San Francisco Department of Medicine and Pharmacology, 1650 Owens Street, San Francisco, CA 94158, USA, [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK, [4]Manitoba Institute of Cell Biology, University of Manitoba, 675 McDermot Avenue, Winnipeg, Manitoba, Canada R3E 0V9, [5]The Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8, [6]Center for Cardiovascular Research, Charité Universitätsmedizin and Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany, [7]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA, [8]University of Toronto, 4 Taddle Creek Road, Toronto, Ontario, Canada M5S 3G9, [9]Department of Molecular Hematology , University of Frankfurt Medical School, 60590 Frankfurt am Main, Germany, [10]GSF Research Center for Environment and Health, Institute for Developmental Genetics, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany, [11]Institute of Molecular Embryology and Genetics, Kumamoto University, 2-2-1 Honjo, Kumamoto 860-0811, Japan and [12]University of California, Los Angeles, 650 Charles E. Young Dr So., Los Angeles, CA 90095, USA

ABSTRACT

Gene trapping is a method of generating murine embryonic stem (ES) cell lines containing insertional mutations in known and novel genes. A number of international groups have used this approach to create sizeable public cell line repositories available to the scientific community for the generation of mutant mouse strains. The major gene trapping groups worldwide have recently joined together to centralize access to all publicly available gene trap lines by developing a user-oriented Website for the International Gene Trap Consortium (IGTC). This collaboration provides an impressive public informatics resource comprising ~45 000 well-characterized ES cell lines which currently represent ~40% of known mouse genes, all freely available for the creation of knockout mice on a non-collaborative basis. To standardize annotation and provide high confidence data for gene trap lines, a rigorous identification and annotation pipeline has been developed combining genomic localization and transcript alignment of gene trap sequence tags to identify trapped loci. This information is stored in a new bioinformatics database accessible through the IGTC Website interface. The IGTC Website (www.genetrap.org) allows users to browse and search the database for trapped genes, BLAST sequences against gene trap sequence tags, and view trapped genes within biological pathways. In addition, IGTC data have been integrated into major genome browsers and bioinformatics sites to provide users with outside portals for viewing this data. The development of the IGTC Website marks a major advance by providing the research community with the data and tools necessary to effectively use

---

*To whom correspondence should be addressed. Email: tef@cgl.ucsf.edu

public gene trap resources for the large-scale characterization of mammalian gene function.

INTRODUCTION

The large and continually growing number of genome sequencing projects provides an opportunity to greatly advance our understanding of genetics and disease. One of the keys to realizing this goal is the development of genomic resources to elucidate functional characteristics of these genes, especially in mammalian genomes. The mouse is an especially useful mammalian model system, providing an excellent subject for studies of gene function because of its short generational span, ease of handling, and the close structural and functional similarity of its genome to that of humans (1,2). Furthermore, using this organism, scientists have access to a wide range of procedures for genetic manipulation, including the use of embryonic stem (ES) cells to create mice with defined single-gene mutations using gene targeting and gene trapping techniques (3).

Gene trapping is a high-throughput method of creating mutagenized ES cells for use in generating knockout and other mutant mouse strains for research in functional genomics (4). Second generation gene trap vectors have recently enhanced the value of the method by offering the potential for creating conditional and other desired alleles using site-specific recombination (5–7). Major scientific initiatives are currently underway in North America and Europe to knock out every mouse gene in ES cells in order to characterize gene function and provide insight into systems associated with human disease (8,9).

A number of gene trap projects have already made notable progress toward this goal by generating resources of gene trap mouse ES cell lines harboring well-characterized insertional mutations (6,7,10–14), although until now the individual gene trap projects have been isolated, providing only details about their own cell lines. The International Gene Trap Consortium (IGTC) is a collaboration representing the major public gene trap resources worldwide, whose mission is to offer the scientific community access to all publicly available gene trap cell lines on a non-collaborative basis for nominal handling fees (15). The centralization of gene trap resources provides many advantages to the research community, allowing more effective utilization of the experimental opportunities offered by gene trap cell lines through standardized protocols for the identification and annotation of sequences from trapped loci, and the increased availability of experimental protocols. As

reported here, the release of the IGTC Website (www.genetrap.org) marks a major advance, generating a standardized informatics pipeline and providing in one place both easy access to all publicly available gene trap cell lines and sophisticated tools for analysis of resource data. Gene trap centers currently involved in this effort are listed in Table 1.

IGTC: RESOURCE OVERVIEW

The IGTC Website centralizes access to all publicly available gene trap cell line data for the first time. This repository was created to address the needs of the international gene trap community by providing researchers with the data and informatics tools necessary to find gene trap cell lines with mutations in genes and loci of interest. IGTC member projects produce gene trap cell lines and directly submit gene trap sequence tags to the Genome Survey Sequences Database (dbGSS) division of GenBank at NCBI (16). Data from all publicly available cell lines are downloaded from dbGSS and subjected to the IGTC identification and annotation pipeline, which then automatically populates the MySQL database used to generate the annotation information presented on the Website. The IGTC Website has been designed to provide easy user access to the extensive array of assembled gene trap informatics data. Interface options include homology searches using BLAST, search and browse capabilities, and viewing trapped genes within biological pathways. The project also includes the integration of gene trap data at major genome browsers and other informatics data sites in order to offer a variety of outside portals to this data. Cell line requests from the IGTC site are forwarded to the originating gene trap resource, where the cell line is removed from cryogenic storage and sent to the user for experimental analysis. The IGTC site also provides useful documentation, on-line tutorials and scientific overviews on gene trapping and the use of gene trap cell lines.

GENE TRAP IDENTIFICATION AND ANNOTATION PIPELINE

Gene trap mutations are characterized through a process of sequencing, identification and annotation. This process involves obtaining cDNA or genomic sequence upstream or downstream of the insertion site and identifying and annotating the locus at which the insertion occurs. A full identification and annotation protocol has been developed for the IGTC that integrates genomic and transcript-based identification approaches and adds information from other major informatics

Table 1. IGTC members

| IGTC members	Cell lines	Website
Baygenomics (USA)	9848	www.baygenomics.ucsf.edu/
Centre for Modelling Human Disease (Toronto, Canada)	4137	www.cmhd.ca/genetrap/
Embryonic Stem Cell Database (University of Manitoba, Canada)	8559	www.EScells.ca/
Exchangeable Gene Trap Clones (Kumamoto University, Japan)	49	egtc.jp/show/index
German Gene Trap Consortium (Germany)	13031	www.genetrap.de/
Sanger Institute Gene Trap Resource (Cambridge, UK)	7354	www.sanger.ac.uk/PostGenomics/genetrap/
Soriano Lab Gene Trap Database (FHCRC, Seattle, USA)	1627	www.fhcrc.org/science/labs/soriano/trap.html
Telethon Institute of Genetics and Medicine–TIGEM (Naples, Italy)	1435	core.tigem.it/genetrap/public/
TOTAL	44605	
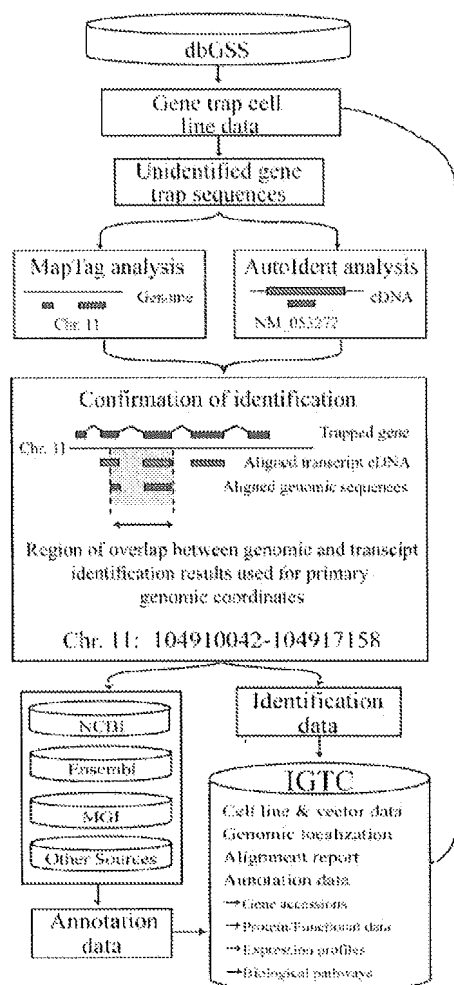
Figure 1. IGTC identification and annotation pipeline. IGTC members submit gene trap cell line data to dbGSS. The first step of the IGTC pipeline is the download of all publicly available gene trap cell line data from dbGSS. Gene trap sequences are then processed through the dual identification protocol based on genomic localization and transcript alignment using MapTag and AutoIdent, respectively. Returned homologous genomic regions and transcripts are aligned to genomic sequence to generate confirmed overlapping sequence regions, which are used as the primary identification data. Confirmed genomic coordinates are queried against major informatics databases to obtain annotation data for the genomic locus identified, and the returned data are entered into the IGTC database.

resources, such as genome browsers and specialized informatics sites to annotate the identified loci (Figure 1). Use of both genomic and transcript-based annotation methods were incorporated to increase confidence in identification and localization on the genome of trapped loci, providing a significant improvement over identification strategies previously in use. This pipeline was designed to be both robust and flexible, allowing the incorporation of multiple methods of identification and mapping and the future integration of new genomic data resources that may be developed.

Gene trap cell line sequences are submitted by each individual gene trap resource to dbGSS along with relevant ancillary information about the cell line, including its original source and the vector used in the trapping experiment. Within

dbGSS, IGTC sequences are grouped using LinkOut (17) (http://www.dlib.org/dlib/march02/03inbrief.html KWAN), which links gene trap dbGSS entries to the IGTC Website. Requiring gene trap cell lines to be entered into dbGSS ensures that all IGTC lines are in the public domain, and that all data are consistent, transparent and freely accessible. Researchers can access additional information about methodology and protocols from the original trapping experiment via links to the individual gene trap project sites (Table 1).

The two complementary programs, MapTag (15) and AutoIdent (12), are used to locate the trapped gene on the mouse genome by analyzing the similarity between the unidentified gene trap sequence and genomic and transcript sequences, respectively. The minimum region of genomic overlap serves as the primary identification data, from which all subsequent annotation is derived. Finally, the identified map coordinates are annotated with gene features obtained from major genomic and informatics databases.

Genomic localization: MapTag

MapTag was developed as an automated method of identifying homologous genomic regions for gene trap sequences using the Ensembl database (18) and SSAHA algorithm (19). MapTag identifies matches in genomic sequence and assembles individual related stretches of genomic similarity using a basic splice model that takes into account exon boundaries, allowing the processing of nucleotide sequences of cDNA or genomic origin. The program filters alignment results, applying simple heuristics using the overall match length, percent identity and exon coverage, and filtering to remove pseudogenes, to determine the best match. As gene trap sequence tags are typically short and imperfect, they can be difficult to identify to a unique locus. The protocol used by MapTag greatly improves the ability to differentiate between matches that show comparable levels of similarity by correctly selecting the result that exhibits a correspondence to the insertion site. The program returns the identified genomic region and an estimate of the match confidence.

Transcript identification: AutoIdent

The pipeline also uses transcript identification to provide an independent and orthogonal identification method. AutoIdent, an automated protocol developed by BayGenomics, uses the BLAST (20) algorithm to identify the most similar sequences in the GenBank non-redundant nucleotide database at NCBI. When the program identifies many high-scoring matches to very similar (synonymous) genes, AutoIdent adds steps to filter results and condense synonymous transcripts to obtain a single result. The program applies stringent criteria for acceptance of a high-quality gene identification but allows more relaxed criteria to identify multiple matching sequences or to a homologous sequence in another species. At the end of the process, AutoIdent returns the best match transcript along with alignment data.

Identification reconciliation and confirmation

Genomic and transcript identification data are stored in the IGTC database. The pipeline then compares results from the two identification protocols, using overlap between the

genomic coordinates from MapTag and the genomic localization of transcripts returned by AutoIdent to confirm the identification. Gene trap sequences are assigned as localized when the genomic and transcript map coordinates overlap, or when only one protocol returns map coordinates. If the identified coordinates conflict or neither protocol returns map data, the gene trap sequence is classified as unlocalized and does not go through the rest of the annotation pipeline. This reconciliation step assures that each identified cell line is mapped to a genomic insertion locus with high confidence and in a manner that is fully documented.

Annotation of gene trap cell lines

Gene trap cell line annotation is based on the confirmed genomic coordinates as the primary identification data. Map coordinates are used to query the Ensembl and Entrez (21) databases to obtain gene features for the identified locus. These accessions are used as primary keys to further query Ensembl, Entrez and the Mouse Genome Informatics (MGI) resource (22) for secondary annotation data associated with the trapped genes, including major gene accession systems, Gene Ontology classifications (GO), protein domain, structure and function, PubMed and phenotype data, homology and orthology data, and microarray probesets. In addition to supporting an extensive array of informatics data, the IGTC site also includes tissue-specific expression data from the SymAtlas project (23) and biological pathway and GO hierarchy diagrams with trapped genes marked, produced using the GenMAPP program (24).

The IGTC database

The IGTC uses the open source MySQL database platform (www.mysql.com). The database is populated in an automated process using results from the MapTag and AutoIdent identification protocols and is structured to optimize information access via Web queries. New gene trap cell line entries in dbGSS are downloaded weekly and run through the IGTC pipeline. Identification results from MapTag are updated with each Ensembl build and AutoIdent is programmed to regularly BLAST sequences in the database to update accession numbers and other changes in the information provided by GenBank. Annotation data generation is synchronized with the Entrez, MGI and Ensembl databases, and the IGTC database is updated by downloading information from these sites as necessary. The information in the IGTC database is available upon request in a tab-delimited or database compatible format.

ACCESSING GENE TRAP DATA

The IGTC site provides a user-friendly approach to gene trap data, allowing researchers to access the gene trap database from a sequence, accession number or ID, expression or pathway perspective using a variety of interfaces for searching and viewing gene trap data. The site is organized around the cell line annotation page, where the user can view all annotation data for a selected cell line (Figure 2). Primary identification and annotation data appear at the top of the page with a link to detailed identification results. This is followed by expandable lists of secondary annotation data, which provide information

useful in the selection and analysis of cell lines of potential interest. Below this is a section containing details about the cell line and gene trap vector, including primer sequences and a description of the vector properties. To aid in the comparison of insertion sites of different cell lines that trap the same gene, the bottom of the cell line annotation page contains a diagram showing the gene trap sequence aligned to the transcripts returned by AutoIdent. The Website provides a similar annotation page for all trapped genes, organized by gene ID. Researchers can also access gene trap data via links from other major informatics resources, including the genome browsers and primary accession pages at NCBI (25), Ensembl (26) and UCSC (27).

The IGTC Website offers users diverse ways to find gene trap cell lines, ranging from searches using protein data, microarray probesets or nucleotide sequences, to screening for traps placed in the context of biological pathways or in genes that demonstrate a particular expression profile. Figure 3 lists the ways in which gene trap data can be accessed, categorized by access type, details about available data type and data access point. Users can search the IGTC database by accession number, ID or keyword and chromosomal location. Results from database searches are displayed as a list of cell line IDs or gene symbols, which link to the individual cell line or gene annotation page. These lists can be exported as tab-delimited files for use in spreadsheet programs or custom databases. Users can also browse the database by MGI Marker Symbol, Gene Name or chromosome location. BLAST analysis can be performed using nucleotide sequence for a gene or locus of interest to search against gene trap sequences or genomic sequence of trapped genes. Trapped genes can be viewed with a pathway perspective using biological pathway diagrams and functional GO groupings, which are colored by the number of cell lines available for each gene. Users can also search for traps in genes with a designated expression profile by selecting a tissue of interest and choosing the expression level of the gene relative to the median tissue expression.

Finally, users can browse gene trap cell lines displayed at major external informatics sites, such as genome browsers and gene pages. IGTC cell lines have been mapped to genomic sequence using the NCBI map viewer, UCSC genome browser and Ensembl genome browser. IGTC gene traps are maintained at these sites either as standard map tracks or as user-configured tracks. (Users should follow site-specific directions to view gene trap data using these browsers.) In addition, gene trap cell lines are listed on some of the major gene pages and in mouse strain resource databases, and the IGTC maintains a full list of partnering sites. By integrating gene trap data into the larger context provided by these bioinformatics resources, the IGTC can reach more potential users who are interested in genomic resources and functional genetics.

SUMMARY

The establishment of the IGTC database and Website marks a major advance in making large-scale mouse knockout resources available to the scientific community. Through the collaboration of gene trap projects worldwide, a standardized

Figure 2. IGTC Website cell line annotation page. The image provides an example of the main data page for gene trap cell lines. Users can view primary identification and annotation data, with a link to detailed reports from the IGTC pipeline. An extensive amount of related annotation data is also presented to give researchers more information about the trapped locus. Finally, the page shows details about the cell line and vector and an image showing all gene trap cell line sequences aligned to the trapped gene. From this page, users are directed to the IGTC member site for cell line requests.

ICON	Access Type	Data Access Details	Access Point
?	Search	Search the IGTC database for trapped genes by keyword, accession, or cell line features	IGTC Website
	Browse	Browse trapped genes by MGI Marker Symbol, gene name, or chromosome	IGTC Website
	BLAST	Search for homologous gene trap sequences using BLAST alignment	IGTC Website
	Biological Pathways	View trapped genes visually using biological pathway diagrams and GO hierarchies	IGTC Website
	Gene Expression	Search for trapped genes that match a desired tissue-specific expression profile	IGTC Website
ACTGA CACGT GCTCA	Nucleotide Sequence	Search the Genome Survey Sequence database at NCBI for gene trap sequences	dbGSS
	Genome Resources	View gene trap cell line data integrated at major bioinformatics websites	Ensembl, Entrez, MGI
	Genome Browsers	Visually browse genomic sequence with gene trap data mapped on to chromosomes	NCBI,UCSC, Ensembl
	Mouse Resources	Find available gene trap cell lines at mutant mouse resource websites	Mouse strain resources

Figure 3. Gene trap data access. Several methods for users to find gene trap cell lines of interest are illustrated.

identification and annotation pipeline has been developed to analyze gene trap data and offer the data access tools necessary for maximum resource utility. For the first time, researchers investigating gene function in the mouse can query all publicly available gene trap ES cell lines at a single site. The IGTC Website contains approximately 45 000 cell lines harboring mutations in nearly 40% of mouse genes (15), including many genes with gene trap cell lines representing multiple mutant alleles. Researchers can search and browse the resource based on accession numbers or IDs, keywords, sequence data, tissue expression and biological pathways, and can also access the IGTC site from other major informatics sites. Furthermore, they can easily request IGTC cell lines for functional characterization of gene function and disease models. For example, BayGenomics cell lines are available from the Mutant Mouse Regional Resource Center (MMRRC) at the University of California Davis (www.mmrrc.org). The MMRRC-UC Davis also offers to microinject ES cells to derive knockout mice for investigators. Soon, the Soriano and TIGEM collections will also be available from the MMRRC. The IGTC Website will grow and continue to develop new tools and features as the diversity of trapped genes and experimental options for using gene trap cell lines expand.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Waterston,R.H., Chinwalla,A.T., Cook,L.L., Delehaunty,K.D., Fewell,G.A., Fulton,L.A., Fulton,R.S., Graves,T.A., Hillier,L.W., Lindblad-Toh,K. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
2. Sung,Y.H., Lee,H.-W. and Song,J. (2004) Functional genomics approach using mice. *J. Biochem. Mol. Biol.*, **37**, 122–132.
3. Nolan,P.M. (2000) Generation of mouse mutants as a tool for functional genomics. *Pharmacogenomics*, **1**, 243–255.
4. Stanford,W.L., Cohn,J.B. and Cordes,S.P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nature Rev. Genet.*, **2**, 756–768.
5. Branda,C.S. and Dymecki,S.M. (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell*, **6**, 7–28.
6. Schnutgen,F., De-Zolt,S., Van Sloun,P., Hollatz,M., Floss,T., Hansen,J., Altschmied,J., Seisenberger,C., Ghyselinck,N.B., Ruiz,P. *et al.* (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA*, **102**, 7221–7226.
7. Cobellis,G., Nicolaus,G., Iovino,M., Romito,A., Marra,E., Barbarisi,M., Sardiello,M., Di Giorgio,F.P., Iovino,N., Zollo,M. *et al.* (2005) Tagging genes with cassette-exchange sites. *Nucleic Acids Res.*, **33**, e44.
8. Austin,C.P., Battey,J.F., Bradley,A., Bucan,M., Capecchi,M., Collins,F.S., Dove,W.F., Duyk,G., Dymecki,S., Eppig,J.T. *et al.* (2004) The knockout mouse project. *Nature Genet.*, **36**, 921–924.
9. Auwerx,J., Avner,P., Baldock,R., Ballabio,A., Balling,R., Barbacid,M., Berns,A., Bradley,A., Brown,S., Carmeliet,P. *et al.* (2004) The European dimension for the mouse genome mutagenesis program. *Nature Genet.*, **36**, 925–927.
10. Hicks,G.G., Shi,E.-G., Li,X.-M., Li,C.-H., Pawlak,M. and Ruley,H.E. (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet.*, **16**, 338–344.
11. Wiles,M.V., Vauti,F., Otte,J., Fuchtbauer,E.-M., Ruiz,P., Fuchtbauer,A., Arnold,H.-H., Lehrach,H., Metz,T., Von Melchner,H. *et al.* (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nature Genet.*, **24**, 13–14.

12. Stryke,D., Kawamoto,M., Huang,C.C., Johns,S.J., King,L.A., Harper,C.A., Meng,E.C., Lee,R.E., Babbit,P.C., Ferrin,T.E. *et al.* (2003) BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. *Nucleic Acids Res.*, **31**, 278–281.

13. Hansen,J., Floss,T., Wurst,W., Van Sloun,P., Schnütgen,F., Von Melchnert,H., Füchtbauer,E.-M., Vauti,F., Arnold,H.-H. and Ruiz,P. (2003) A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA*, **100**, 9918–9922.

14. To,C., Epp,T., Reid,T., Lan,Q., Yu,M., Li,C.Y., Ohishi,M., Hant,P., Tsao,N., Casallo,G. *et al.* (2004) The Centre for Modeling Human Disease Gene Trap resource. *Nucleic Acids Res.*, **32**, D557–559.

15. Skarnes,W.C., Nord,A.S., Cox,T., von Melchner,H., Wurst,W., Hicks,G., Young,S.G., Conklin,B.R., Ruiz,P., Soriano,P. *et al.* (2004) A public gene trap resource for mouse functional genomics. *Nature Genet.*, **36**, 543–544.

16. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2005) GenBank. *Nucleic Acids Res.*, **33**, D34–D38.

17. Kwan,Y.K. (2002) LinkOut: Explore beyond PubMed and Entrez. *D-Lib Mag.*, **8**.

18. Hubbard,T., Andrews,D., Caccamo,M., Cameron,G., Chen,Y., Clamp,M., Clarke,L., Coates,G., Cox.T., Cunningham,F. *et al.* (2005) Ensembl 2005. *Nucleic Acids Res.*, **33**, D447–D453.

19. Ning,Z., Cox,A.J. and Mullikin,J.C. (2001) SSAHA: a fast search method for large DNA databases. *Genome Res.*, **11**, 1725–1729.

20. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

21. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.

22. Eppig,J.T., Bult,C.J., Kadin,J.A., Richardson,J.E., Blake,J.A., Anagnostopoulos,A., Baldarelli,R.M., Baya,M., Beal,J.S., Bello,S.M. *et al.* (2005) The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res.*, **33**, D471–475.

23. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

24. Dahlquist,K.D., Salomonis,N., Vranizan,K., Lawlor,S.C. and Conklin,B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genet.*, **31**, 19–20.

25. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S., Helmberg,W. *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **33**, D39–D45.

26. Stalker,J., Gibbins,B., Meidl,P., Smith,J., Spooner,W., Hotz,H.-R. and Cox,A.V. (2004) The Ensembl Web site: mechanics of a genome browser. *Genome Res.*, **14**, 951–955.

27. Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.

Impaired expression of importin/karyopherin β1 leads to post-implantation lethality

Katsutaka Miura [a,1], Kumiko Yoshinobu [b], Takashi Imaizumi [a,1], Kyoko Haruna [a,c], Yoichi Miyamoto [d], Yoshihiro Yoneda [d], Naomi Nakagata [b], Masatake Araki [b], Taihei Miyakawa [e], Ken-ichi Yamamura [a], Kimi Araki [a,*]

[a] *Institute of Molecular Embryology and Genetics, Kumamoto University, Honjo 2-2-1, Kumamoto 860-0811, Japan*
[b] *Institute of Resource Development and Analysis, Kumamoto University, Honjo 2-2-1, Kumamoto 860-0811, Japan*
[c] *TransGenic Inc., 12-32 Hanabata-cho, Kumamoto 860-0806, Japan*
[d] *Department of Frontier Biosciences, Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan*
[e] *Department of Neuropsychiatry, Kumamoto University School of Medicine, Honjo 1-1-1, Kumamoto 860-8556, Japan*

Abstract

Importin β1 (Impβ)/karyopherin β1 (Kpnb1) mediates the nuclear import of a large variety of substrates. This study aimed to investigate the requirement for the *Kpnb1* gene in mouse development, using a gene trap line, B6-CB-$Ayu8108^{Gtgeo1MEG}$ ($Ayu8108^{geo}$), in which the trap vector was inserted into the promoter region of the *Kpnb1* gene, but in reverse orientation of the *Kpnb1* gene. $Ayu8108^{geo/geo}$ homozygous embryos could develop to the blastocyst stage, but died before embryonic day 5.5, and expression of the *Kpnb1* gene in homozygous blastocysts was undetectable. We also replaced the *βgeo* gene with Impβ cDNA through Cre-mediated recombination to rescue Impβ expression. Homozygous mice for the rescued allele $Ayu8108^{Impβ/Impβ}$ were born and developed normally. These results demonstrated that the cause of post-implantation lethality of $Ayu8108^{geo/geo}$ homozygous embryos was impaired expression of the *Kpnb1* gene, indicating indispensable roles of Impβ1 in early development of mice.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Importin β1; Gene trap; Cre/lox site-specific recombination; Embryonic lethal

Active nuclear import of proteins is a highly selective process involving specific recognitions between nuclear localization signals (NLS) and suitable receptors [1]. Importin/karyopherin β1 (Impβ) is a key player in nuclear protein import and mediates targeting of a canonical NLS substrate bound to an adaptor protein, importinα [2,3]. On the nuclear membrane, Impβ interacts with nuclear envelope-localized nuclear pore complexes (NPCs) and carries the importinα/cargo complex from the cytoplasm into the cell nucleus. In the nucleus,

RanGTP, which exists predominantly in the nucleus, binds to Impβ and induces the release of import cargoes. Thereafter, individual importins are recycled back to the cytoplasm. Thus, Impβ is a critical component in mechanisms involved in targeting of the NLS substrate into the nucleus [4–6].

In addition to nuclear transport, Impβ has also been shown to play a role in regulation of spindle formation and of aster promoting activity (APA) during mitosis [7–10]. Two microtubule organizing components, NuMA and TPX2, which are retained in the nucleus during interphase, bind to Impβ via importinα during mitosis and are kept away from chromatin. RanGTP releases NuMA and TPX2 from importin α/β heterodimer around chromosomes and promotes spindle formation.

Impβ is the prototype of the karyopherin family including more than 20 members in mammals [11]. Ten of these members have been identified to have a role in nuclear import. In general, such importinβ family proteins function without the need of an adaptor protein and can directly interact with their substrates. To the best of our knowledge, only Impβ requires an adaptor, importinα, to bind to its cognate cargoes.

Extensive studies using cultured cells have clarified molecular mechanisms of nuclear transport. However, roles and functional redundancy of Impβ family members in vivo have been poorly studied. In *Drosophila*, Impβ mutant strains (*Ketel* mutations) have been isolated and analyzed [12–14]. Homozygous larvae for the loss-of-function *Ketel* allele die during the second larval instar, but on the other hand, homozygous somatic clones induced by mitotic recombination are viable in the follicle epithelium in wings and tergites. Since the *Ketel* gene is not expressed in larval and adult cells that are mitotically inactive, another import pathway might substitute for the function of Impβ in *Drosophila*.

In order to identify the requirement for the *importin/karyopherinβ1* (Kpnb1) gene in mouse development, we analyzed a *Kpnb1* mutant mouse line established by gene trapping [15]. The trap vector was inserted in the promoter region of the *Kpnb1* gene in the reverse orientation of the transcription of the *Kpnb1* gene, and expression of the *Kpnb1* gene from the mutated allele was severely decreased. Homozygous embryos could develop to the blastocyst stage, but died before embryonic day (E) 5.5, indicating that Impβ1 played an indispensable role in early development of mice.

Materials and methods

Isolation of ES clones and establishment of mutant mouse lines. The gene trap vector pU-hachi and isolation of trap clones were described previously [15]. The vector contained a splice acceptor region (SA) from the mouse *En-2* gene, *lox*71, the internal ribosomal entry site (IRES) from the encephalomyocarditis virus (ECMV), the *β-galactosidase/neomysin phosphotransferase fusion* gene (*βgeo*), *lox*P, the SV40 polyadenylation sequence (pA), and pUC19 (Fig. 1A). In order to replace the *βgeo* gene with the Impβ cDNA sequence, a replacement vector carrying *lox*66-Impβ cDNA–phosphoglycerate kinase 1 (PGK)–*puromycin resistant* (*Pac*) gene–*lox*P–pSP73 was constructed. Site-specific integration in trap clones mediated by Cre-recombinase was performed as described previously [15]. Chimeric mice were produced by aggregation of ES cells with eight-cell embryos of ICR mice (Nippon Clea, Tokyo, Japan), as described previously [16]. Chimeric male mice and their heterozygous progeny were backcrossed for five to eight generations to a C57BL/6J background.

Molecular cloning of flanking genomic region by plasmid rescue. Plasmid rescue was performed as described previously [15]. Genomic DNA of Ayu8108 ES cells was digested with *Pst*I or *Sph*I, followed by self-ligation, and introduction into *Escherichia coli* cells by electroporation. The recovered plasmids were mapped by restriction enzymes and sequenced using the dideoxy chain termination method using a BigDye Terminator Cycle Sequencing (Perkin-Elmer, Foster City, CA).

Genotyping of mice. Genomic DNA was isolated by proteinase K digestion, phenol–chloroform extraction, and ethanol precipitation from tail biopsies of newborn, E10.5 embryos, and E7.5 embryos.
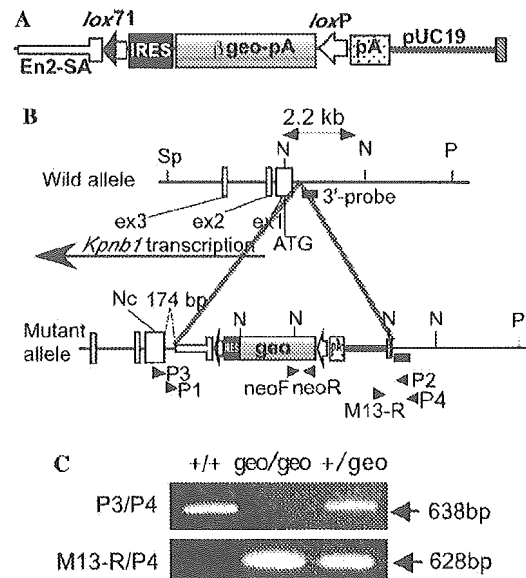

Fig. 1. Gene trap events in the Ayu8108 clone. (A) Structure of the trap vector, pU-hachi. The pU-hachi vector contains an IRES-βgeo-pA cassette flanked by *lox*71 and *lox*P to exchange the βgeo cassette into any DNA sequence through recombination by Cre. En2, the mouse *Engrailed 2* gene; SA, splice acceptor; pA, polyadenylation signal. (B) Integration pattern of the trap vector. Open boxes on the maps represent the first three exons of the *Kpnb1* gene. The trap vector was inserted 174 bp upstream from the first exon. The start codon and direction of transcription of the *Kpnb1* gene are indicated. The solid bar with "3'-probe" shows the probe used in Southern blotting for genotyping. Arrowheads indicate primers used for genotyping and RT-PCR. N, *Nco*I; Sp, *Sph*I; P, *Pst*I. (C) Genotyping by PCR. The primer pairs neo-F/neo-R and P1/P2 were used to detect the *Ayu8108*^{geo} and wild-type alleles, respectively.

For newborn mice and E10.5 embryos, Southern blot analysis was carried out. Seven micrograms of genomic DNA was digested with an appropriate enzyme, electrophoresed on 1% agarose gel, and blotted onto a nylon membrane (Roche Molecular Biochemicals, Mannheim, Germany). Probe preparation and hybridization were performed with the DIG DNA Labeling and Detection Kit (Roche).

For E7.5 embryos, PCR analysis was carried out. To identify the mutant allele, the primers neo-F (5'-AGAGGCTATTCGGCTATGAC-3') and neo-R (5'-CACCATGATATTCGGCAAGC-3') were used. PCR conditions were 30 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 60 s, using 0.5 U AmpliTaq DNA polymerase (Roche). To identify the wild allele, the primers P1 (5'-ATTGGCCGGACGGCTAGCGT-3') and P2 (5'-GCTGAGCCCGAAGGCCTTTA-3') were used. PCR conditions were 30 cycles of 94 °C for 30 s, 58 °C for 60 s, and 72 °C for 60 s, using 0.5 U LA *Taq* DNA polymerase (TaKaRa, Shiga, Japan).

For blastocysts and cultured embryos, individual embryos were lysed for 5 min in 2 µl of 0.005% SDS, 0.035 N NaOH, at 100 °C. After neutralization by adding 36 µl water, PCR was carried out using 5 µl of the extract. Primer pairs M13-R (5'-AGGAAACAGCTATGACCATGA-3') and P3(5'-GGTCACCACAAGCCCATTCA-3'), P3 and P4 (5'-GAAC TCCTCGCTTCAGTTCT-3') were used to identify mutant and wild alleles, respectively. PCR conditions were 40 cycles of 94 °C for 30 s, 62 °C for 30 s, and 72 °C for 30 s, using 0.5 U LA *Taq* polymerase (TaKaRa).

Quantitative analysis of expression of the Kpnb1 gene. Ten micrograms total RNA was reverse-transcribed using Superscript II (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Real-time quantitative reverse-transcription polymerase chain reaction (RT-PCR) was performed with a LightCycler instrument, using Faststart DNA

Master SYBR Green I Kit (Roche). PCRs were performed using B1 (5′-TAACCATCCTCGAGAAGACC-3′) and B2 (5′-ATCCCCTGGATT ATGGCAGT-3′) primers, and 40cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s.

Isolation and in vitro culture of blastocysts. Heterozygous female mice were superovulated and mated with heterozygous male mice, and 2-cell stage embryos were isolated and cultured in KSOM [17] medium for 2 days to the blastocyst stage. Individual blastocysts were transferred to single wells coated with gelatin in 15% FCS–DMEM. Plates were incubated and kept in a humid chamber at 37 °C, 5% CO_2 for 3 days.

Histological analysis of E5.5 embryos. Heterozygous and wild-type females were superovulated and mated with heterozygous males. Fertilized eggs were collected from the oviducts and transferred to foster mothers on the same day. Decidual swellings containing E5.5 embryos were dissected, fixed in 4% paraformaldehyde solution, and sectioned. Serial sections were stained with hematoxylin and eosin (H/E), and observed under a microscope.

RT-PCR analyses. Total RNA was isolated from individual blastocysts using the RNeasy Mini Kit (Qiagen, Valencia, CA) and suspended in 30 μl water. First-strand cDNA was synthesized using 9 μl RNA solution with oligo(dT) primers in a ThermoScript RT-PCR System (Invitrogen, Carlsbad, CA). One-twentieth volume of the first-strand reaction was used for PCR amplification. Expressions of the *βgeo* and *Kpnb1* genes were detected using primers neo-F and neo-R, and primers B3 (5′-CTCTTCA GAATGTTCTCCGG-3′) and B4 (5′-GATCTCCGCCCTTCAGTTAA-3′), respectively. Pericentriolar material-1 (PCM1) was used as internal positive control. Primers for PCM1 were 5′-GCGTTACCCAACTT AATC-3′ and 5′-TGTGAGCGAGTAACAACC-3′. Both PCR conditions were 40 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s using 0.5 U LA *Taq* DNA polymerase (TaKaRa). In the case of adult mice, 10 μg of total RNA was used for first-strand cDNA synthesis using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). To detect expressions of the inserted Impβ cDNA and the endogenous *Kpnb1* gene, primers E3 (5′-GTTCGAGCTTGGAATTCATG-3′) and B5 (5′-CCGT CGAGCATTAGCATCAA-3′), and B3 and B6 (5′-CCTCTC ATTCCCAAGCATTC-3′) were used, respectively. PCR conditions were 35 cycles of 94 °C for 30 s, 58 °C for 60 s, and 72 °C for 60 s using 0.5 U AmpliTaq polymerase (Perkin-Elmer).

5-rapid amplification of cDNA ends (RACE) and 3′RACE analyses. Total RNA was prepared using Sepazol reagent (Nacalai Tesque, Kyoto, Japan), and mRNA was purified using the Oligotex-dT30 Super mRNA Purification Kit (TaKaRa). Reverse-transcription was performed with Thermoscript (Invitrogen) using primer E1 (5′-AGCAGTGAA GGCTGTGC-3′) for 5′RACE. The 5′RACE System for Rapid Amplification of cDNA Ends Reagent Assembly Ver. 2.0 (Invitrogen) was used according to the manufacturer's instructions, using the primer E2 (5′-CTT TGTTAGGGTTCTTCTTC-3′). Products were electrophoresed and subjected to Southern blotting using digoxigenin-ddUTP-labeled oligonucleotide probes for the SA sequence in the trap vector, prepared using the DIG Oligonucleotide 3′-End Labeling Kit (Roche). Hybridization with the probes was carried out overnight at 60 °C.

The downstream cDNA fragment containing a poly(A) stretch was obtained using 3′RACE System for Rapid Amplification of cDNA Ends (Invitrogen). First-strand cDNA was synthesized with a 3′RACE Adapter Primer using the Thermoscript RT-RCR System (Invitrogen) and two nested primers P5 (5′-CATTGGCCTGCGGCTTCA-3′) and P6 (5′-CGG CTTCAAGGTCCGGTT-3′). RACE products were cloned into pGEM-T vectors (Promega, Madison, WI) and sequenced.

Results

Analysis of the insertion event in Ayu8108 trap clone

Ayu8108 ES clone cells were isolated using gene trap screening with the pU-Hachi trap vector (Fig. 1A), which was designed for the exchangeable gene trap [15]. Southern

blot analysis with a probe for the pUC vector fragment revealed a single copy integration of the vector (data not shown). Genomic DNA fragments flanking both the 5′ and 3′ ends of the integrated vector were obtained by the plasmid rescue method after removal of the *βgeo* sequence by Cre-mediated recombination [18]. Sequence analysis of the flanking genomic DNA and homology search in the GenBank database revealed that the trap vector was integrated into the promoter region of the *Kpnb1* gene. Although the transcription initiation site of the *Kpnb1* gene was not yet determined, several expression sequence tag (EST) sequences covering the 5′-region of the *Kpnb1* gene show 5′-ends at 290–323 bp upstream of the start codon of the *Kpnb1* gene. Therefore, in our study, the 5′-end of the BY740663 EST sequence, which has the most extended 5′-end, was considered as the *Kpnb1* transcription initiation site. The trap vector was integrated at 225 bp upstream of the first exon, however, direction of transcription of the *βgeo* was opposite to that of the *Kpnb1* gene (Fig. 1B). Deletion of genomic DNA at the integration site was only of 46 bp, and no deletion was found in the integrated trap vector.

Chimeric mice were produced by aggregation of Ayu8108 ES cells with morulae, and the trap mouse line designated as B6-CB-*Ayu8108^{GtgeoIMEG}* (*Ayu8108^{geo}*) was established.

Identification of the trapped transcript in the Ayu8108 line

The fact that the direction of transcription of the *βgeo* was opposite to that of the *Kpnb1* gene (Fig. 1) indicated the existence of another gene in the promoter region of the *Kpnb1* gene. In fact, ubiquitous expression of the trapped gene was observed by X-gal staining in E9.5 and E12.5 embryos (data not shown). Moreover, in adult mice, transcripts of the *βgeo* were detected in the brain, heart, lung, kidney, and testis using RT-PCR (data not shown).

To identify endogenous transcripts fused to the *βgeo* gene, we performed 5′RACE and obtained a clear single band (Fig. 2A). Sequence analysis revealed that transcription of the fusion transcripts started at 50 bp upstream from the insertion site, which corresponded to 124 bp upstream of the first exon of the *Kpnb1* gene, and spliced at a cryptic splice-donor site (position 52 in the pU-hachi sequence, GenBank Accession No. AB242616) within the intron of the trap vector to fuse the authentic splice acceptor (Fig. 2B). Then, 3′RACE was performed to obtain the whole transcript of the trapped gene, and a 1.3-kb product was obtained (GenBank Accession No. AB242615). As shown in Fig. 2C, the sequence was identical to the 3′-flanking genomic sequence of the trap vector. Homology search using BLAST programs [19] identified several mouse ESTs with high homology (>95%). However, in the 1.3-kb transcript, multiple stop codons appeared in all three frames, and no apparent ORF was found. Northern blot analysis was performed using both total and poly(A) RNAs from adult tissues and an RNA probe for
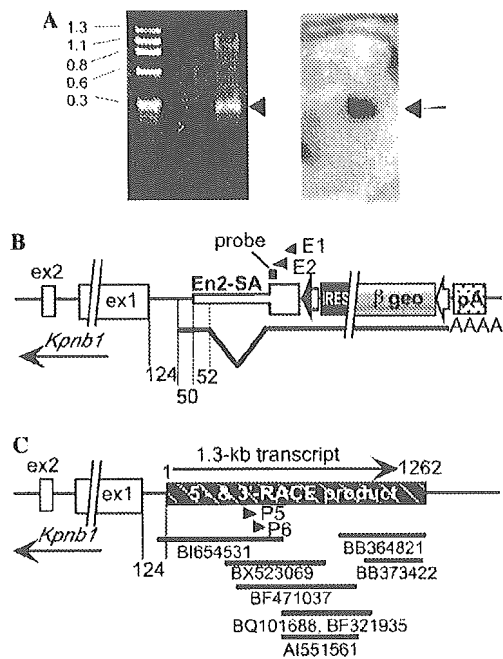
Fig. 2. Identification of the transcript trapped in the Ayu8108 clone. (A) 5'-RACE analysis of the Ayu8108 ES line. A single band (arrowhead) was obtained with E1 and E2 primers (left), and the product hybridized with an oligoprobe from the SA sequence (right, arrow), showing successful amplification of the fusion transcript with the SA sequence of the trap vector. (B) Schematic structure of transcription of the fusion transcript from the *Ayu8108^{geo}* mutant allele. Transcription started 124 bp upstream from the first exon of the *Kpnb1* gene and was spliced at a cryptic splice donor site in the intron sequence of the trap vector. (C) Transcription of the trapped gene in the wild allele. The 1.3-kb transcript identified by 5'- and 3'RACE is indicated by a thick box. The arrow on the box indicates the direction of transcription. Primers used for 3'-RACE are indicated by arrowheads. EST sequences from GenBank are shown under the map with their respective accession numbers.

the 1.3-kb product. However, only a faint signal was detected with total RNA, and no signal was obtained with poly(A) RNA (data not shown). Thus, in the Ayu8108 clones, we could not detect endogenous transcripts from the trapped gene. However, it was possible that a non-coding RNA gene was trapped.

Expression of the Kpnb1 gene is impaired in Ayu8108^{+/geo} mice

Since the integration site of the trap vector was quite close to the first exon of the *Kpnb1* gene, it was expected that insertion would result in impaired expression of the *Kpnb1* gene, although direction of transcription was opposite. In order to examine expression levels of the *Kpnb1* gene in *Ayu8108^{+/geo}* heterozygous mice, we performed real-time RT-PCR quantification in liver, kidneys, and testes of adult mice. As shown in Fig. 3, expression levels of the *Kpnb1* mRNA in *Ayu8108^{+/geo}* mice were significantly lower (about 60–80%) than those of wild-type, indicating that integration of the trap vector produced a hypomorphic allele of the *Kpnb1* gene.
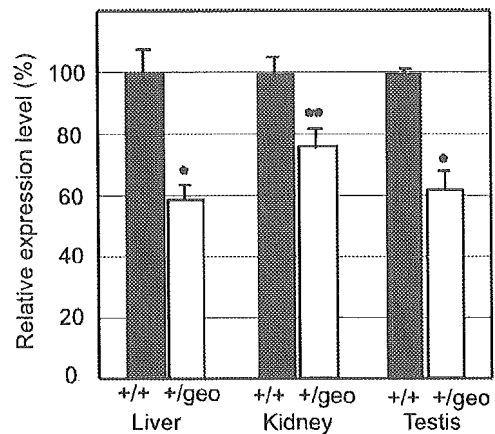

Fig. 3. Comparison of expression levels of the *Kpnb1* gene between wild-type and *Ayu8108^{+/geo}* heterozygous mice by Real-time RT-PCR. Total RNAs from liver ($n = 6$), kidneys ($n = 6$), and testes ($n = 5$) of wild-type (+/+, black bars) and *Ayu8108^{+/geo}* (+/geo, open bars) adult mice were used. In each tissue, relative expression level against average amount in wild-type mice is indicated with standard deviation. $*p < 0.01$; $**p < 0.05$ (Student's t-test).

Ayu8108^{geo/geo} homozygous embryos were lethal after implantation

Ayu8108^{+/geo} heterozygous mice were healthy in appearance and fertile. To investigate the phenotype of *Ayu8108-geo/geo* homozygous mice, heterozygous mice were crossed, and neonates, E10.5 and E7.5 embryos, were genotyped by Southern blotting or PCR analysis (Fig. 1C). As shown in Table 1, no homozygous progeny was identified. Out of 62 deciduas obtained from E7.5, 27.4% of deciduas were empty, likely to correspond to missing homozygotes (Table 1), indicating that decreased expression of the *Kpnb1* gene caused embryonic lethality. To score rates of abortion in the uteri, heterozygous and negative littermate females were superovaluated and mated with heterozygous males. Fertilized eggs were then collected, and transferred into foster mothers on the same day, and then E5.5 deciduas were histologically analyzed. As shown in Table 2, in experiments with wild-type and heterozygous females, frequency of resorptions was 8.3% and 36.2%, respectively.

Table 1
Genotyping distribution of embryos and pups derived from heterozygous intercrosses

Age stage	+/+ (%)	+/geo, geo/+ (%)	geo/geo (%)	Resorbed (%)	Total
Blastocyst	6 (33.3)	9 (50.0)	3 (16.7)	—	18
Outgrowth	5 (20.0)	10 (40.0)	5 (20.0)	—	20
E7.5	18 (29.1)	27 (43.5)	0	17 (27.4)	62
E10.5	15 (37.5)	25 (62.5)	0	ND	40
Newborn	12 (33.3)	24 (66.7)	0	—	36

Embryos and newborn mice were genotyped by Southern blot analysis (newborn and 10.5 dpc), or by PCR analysis using P1/P2 and neo-F/R primers (7.5 dpc) or with M13-R, P3, and P4 primers (blastocysts and outgrowth blastocysts). ND, Not determined: although empty deciduas existed, their numbers were not counted.

Table 2
Frequency of resorptions of E5.5 embryos

Cross	Total No. of deciduas	No. of resorptions	Resorption frequency (%)
+/+ X +/geo	48	4	8.3
+/geo X +/geo	47	17	36.2

Fertilized eggs from the indicated crosses were transferred into oviducts. Then, at E5.5, deciduas were dissected, sectioned, and stained with hematoxylin and eosin to examine morphology of embryos. All viable embryos showed normal morphology.

The difference of 27.9% was expected to correspond to the rate of homozygotes, and no embryonal tissue was observed in the empty deciduas (data not shown). These results strongly suggested that all $Ayu8108^{geo/geo}$ homozygous embryos were lethal before E5.5.

To investigate whether homozygous embryos could survive to the blastocyst stage, two-cell stage embryos from heterozygous crosses were collected and cultured up to the blastocyst stage. All embryos showed normal morphology without any developmental delay. Each embryo was lysed and subjected to PCR for genotyping. At this stage, homozygous embryos were identified (Table 1). Then, growth potential of homozygous embryos was examined by in vitro culture on gelatin-coated culture slides for 3 days. All blastocysts obtained from heterozygous crosses hatched out of the zona pellucida and attached to the slides. Genotyping of explants identified 5 (20%) homozygotes (Table 1). These results suggested that $Ayu8108^{geo/geo}$ homozygous embryos could develop to the blastocyst stage (E3.5), and hatch out, but they died shortly after implantation by E5.5.

In order to examine expression of the *Kpnb1* gene at the blastocyst stage, RT-PCR analysis using individual blastocysts obtained from heterozygous crosses was performed. Fig. 4 depicts results of a typical experiment. In total, 70 blastocysts were examined, and 21 blastocysts (30%) showed wild-type expression pattern [βgeo(−)/Kpnb1(+)], 34 (49%) showed heterozygous expression pattern [βgeo(+)/Kpnb1(+)], and 15 (21%) showed homozygous expression pattern [βgeo(+)/Kpnb1(−)]. Results of genotyping and RT-PCR indicated that expression of the *Kpnb1*
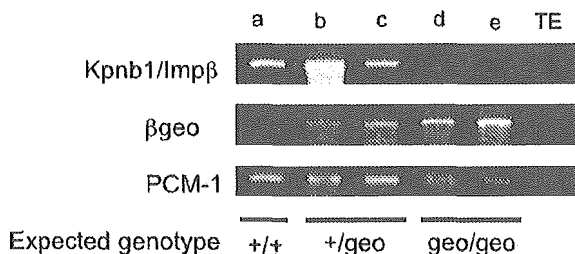

Fig. 4. RT-PCR analysis of individual blastocysts. Total RNAs from individual blastocysts (a–e) were subjected to RT-PCR to detect expression of the endogenous *Kpnb1* gene (upper panel) and the *βgeo* gene (middle panel). Detection of PCM-1 mRNA was performed as positive control (lower panel).

gene was undetectable by RT-PCR in $Ayu8108^{geo/geo}$ homozygous embryos.

Expression of the Kpnb1 gene rescued post-implantation lethality

In $Ayu8108^{geo/geo}$ homozygous mice, expressions of the *Kpnb1* gene and of the 1.3-kb transcript were severely impaired by insertion of the trap vector. In order to determine what was responsible for lethality, we replaced the βgeo gene of the trap vector with Impβ cDNA, using the Cre/mutated *lox* recombination system as outlined in Fig. 5A. The inserted Impβ cDNA should be expressed under the control of the promoter for the 1.3-kb transcript, and production of the 1.3-kb transcripts should remain interrupted. After introduction of the replacement vector and of the Cre-expression vector [20] into Ayu8108 ES clones, 24 colonies were picked up, and 22 of 24 (90%) clones were shown to have the replaced alleles, $Ayu8108^{+/Imp\beta}$, by Southern blotting and PCR analyses (Fig. 5B). Then, the $Ayu8108^{+/Imp\beta}$ mouse line was established from one of the replaced clones through production of germline chimeras. Heterozygous $Ayu8108^{+/Imp\beta}$ mice were crossed, to obtain and examine phenotype of $Ayu8108^{Imp\beta/Imp\beta}$ homozygous mice. Genotype analysis (Figs. 1B and 5B) of 63 4-week-old offsprings identified 13 (21%) wild-type mice, 35 (56%) heterozygous mice, and 11 (17%) homozygous mice for the replaced allele, and all $Ayu8108^{Imp\beta/Imp\beta}$ homozygous mice appeared healthy. Expression of the endogenous *Kpnb1* gene and of the inserted Impβ cDNA was analyzed by RT-PCR using specific primer pairs for each transcript. As shown in Fig. 5C, expression of the integrated Impβ cDNA was detected in heterozygous and homozygous mice. Unexpectedly, the endogenous *Kpnb1* gene was expressed in $Ayu8108^{Imp\beta/Imp\beta}$ homozygous mice. This demonstrated that recovered Impβ expression from the inserted cDNA and endogenous *Kpnb1* gene rescued early embryonic lethality, and lethality was caused by impaired expression of the *Kpnb1* gene, but not of the 1.3-kb transcript.

Discussion

A hypomorphic allele of the mouse *Kpnb1* gene was generated by gene trap mutagenesis using ES cells. In heterozygous adult mice, expression level of the *Kpnb1* gene was reduced to about 60–70%, and in homozygous blastocysts, transcripts were not detected by RT-PCR. Homozygous blastocysts were able to grow on gelatin-coated slides, but homozygous implanted embryos died before E5.5 in vivo. These results indicated that Impβ protein was indispensable for the development of early stage embryos, and that any other importinβ family members could not compensate for nuclear import activity.

The developmental function of Impβ has been analyzed only in mutant *Drosophila* strains. In *Drosophila*, homozygous mutants could develop to the second larval instar