Study of the Mutant Mice Produced by Gene Trapping Using Mouse Embryonic Stem Cells.

Akinori Miyashita

2000
Doctoral Program in Biosystem Science
Graduate School of Science and Technology
Niigata University

TABLE OF CONTENTS

Mutant organisms with interesting phenotypes have contributed enormously to our understanding of gene function. Historically, the mouse has been the mammal of choice for genetic analysis primally because of its short gestation period and large litter sizes, the availability of inbred strains, and the ability to perform controlled matings. Therefore, it is now commonplace for the study of gene function to use the mutant mouse.

Forward genetics

Mutations of mice, either arisen spontaneously or generated by radiation or chemical mutagens, have been collected and genetically analyzed for many years (Lyon et al., 1996). With the distinct improvement of genetic mapping and sequencing techniques, an increasing number of previously identified mutations have been assigned to cloned genes. This methodology is called "forward genetics" (from phenotype to gene). For example, using this approach, the obese gene (*ob*) regulating energy balance (Zhang et al., 1994) and the circadian clock gene (*Clock*) (Kimg et al., 1997; Antoch et al., 1997) were cloned respectively; the former was found out as a spontaneous mutation (Herberg, L. et al., 1977) and the latter was identified in the progeny of mice treated with the alkylating agent *N*-ethyl-*N*-nitrosourea (Vitaterna et al., 1994).

Both the development of high-resolution genetic

linkage maps (for review, see Copeland et al., 1993) and the assemblage of sequence data (such as expressed sequence tags, ESTs) are rapidly progressing in the mouse. In this context, the forward genetic approach will be more available in future to identify mutated genes (Takahashi et al., 1994).

Reverse genetics

Recently, "reverse genetics" (from gene to phenotype) has been shown to be very effective in understanding gene's functions of mice. Transgenesis, a process of introducing foreign DNA into the genome of organisms, has opened the way for the reverse genetic approach to the study of gene function. Because foreign DNA integrates into the host genome, it plays a role of insertional mutagen. The advantage of this mutagenesis is to be capable of isolating genes responsible for mutated phenotypes easily, since the transgene can serve as a molecular tag for cloning a mutated allele.

As the first example of the transgenic mutagenesis, Jaenisch et al. showed that Moloney Murine Leukemia virus worked as insertional mutagen in the Mov13 mouse strain and disrupted the α1(I) collagen gene (for review, see Jaenisch, 1988; Chen et al., 1996). It has been also reported that mutant mice which lost the function of a gene were obtained in generating transgenic mice lines. For example, the *inv* (*inversion of embryonic turning*) mutation, causing the reversion

Abbreviations

bp, base pair(s)

Cd68, murine macrosialin gene

cDNA, complementary DNA

Eif4a1, eukaryotic protein synthesis initiation factor 4A1 gene

ES, embryonic stem

ethidium bromide, EtdBr

Etl-1, mouse enhancer trap locus 1 gene

FLII, human homolog of *Drosophila melanogaster flightless-I* (*fliI*) gene

Fliih, mouse homolog of *FLII*

Fxr2h, fragile X-related gene 2 homolog

GT, gene trap

kb, kilobase pair(s) or 1000 bp

LLGL, human homolog of the murine Llglh gene

Llglh, mouse homolog of Drosophila lethal(2) giant larvae gene

Mb, megabase pair(s)

Nab2, mouse NGFI-A binding protein 2 gene

nt, nucleotide(s)

PCR, polymerase chain reaction

RACE, rapid amplification of cDNA ends

R-PTP-κ, mouse receptor-linked protein-tyrosine phosphatases-kappan gene

SL15, suppressor of Lec15 gene

Sox15, *Sry*-like HMG box gene 15

SOX20, *Sry*-related HMG-box gene 20

Stat6, signal transducers and activators of transcription gene 6

Supl15h, mouse homolog of suppressor of Lec15 (C. griseus) gene

Tiam-1, T-lymphoma invasion and metastasis gene 1

UTR, untranslated region(s)

Abstract

We characterized a region of the mouse genome disrupted by integration of a gene trap (GT) vector in ES cells. On 5' rapid amplification of cDNA ends analysis of the fusion transcripts containing the GT vector we identified the eukaryotic protein synthesis initiation factor 4A1 gene (*Eif4a1*) as a promoter-trapped gene. Plasmid rescue was used to show that the other end of the integrated vector disrupted the murine homolog of the human fragile X mental retardation syndrome-related protein 2 gene (*Fxr2h*). Structural analysis of P1 clones, isolated from the wild type mouse genome by PCR with *Eif4a1* specific primers, indicated that the integration of the GT vector was accompanied by the deletion of about 35 kb of genomic DNA and that the disrupted region also included three genes, *Cd68*, *Supl15h* and *Sox15*, the latter two of which are transcribed in opposite directions with overlapping 3' ends. These five different genes at least, *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h*, are clustered in a 40 kb region. The chromosomal location of this region was mapped by means of interspecific backcross panel DNAs to the central part of mouse chromosome 11, exhibiting a known region of synteny with human chromosome 17.

Introduction

The GT strategy in mouse ES cells was recently developed to identify unknown genes involved in embryogenesis (Gossler et al., 1989; Friedrich and Soriano, 1991; Skarnes et al., 1992). This strategy depends on the random integration of a promoterless GT vector, followed by its expression only when the vector is inserted into an active transcription unit of cellular genes. The random mutation complements the direct approach of gene targeting (Capecchi, 1989) since novel genes can be accidentally disrupted by the random integration of a GT vector.

The GT vector is delivered via retrovirus infection or electroporation. There are no reports to show a deletion at the insertion site by using retrovirus vector. Niwa et al. (1993) showed that a GT vector electroporated into ES cells caused large genomic deletions or rearrangements spanning more than 10 kb at the site of integration. In some cases, the fortuitous insertion of a GT vector occurred between two genes, *Fgf3* and *Fgf4* (Carlton et al., 1998), and in the locus which produces three transcripts (Zambrowicz et al., 1997). Several gene-cluster regions, where some genes share the promoter region or 3' UTRs, have been reported, and the gene expression may be affected with each other in such region. The GT events with a deletion of host genomic DNA may occasionally cause a dramatical or an unexpected mutant phenotype. Therefore, it is important to analyze the wild-type and

mutant allele of the GT vector insertion site in order to avoid misunderstanding of the gene function.

The GT3-11 mouse line was established by means of the GT strategy (Shirai et al., 1996). Genotype analyses suggested that the phenotype of the homozygotes showed lethality in the early embryogenesis. To determine which gene is responsible for the prenatal lethality we used 5' RACE analysis and 3' plasmid rescue to determine the integration region. We found that a GT vector was integrated into two functional genes, *Eif4a1* (Nielsen et al., 1985, Nielsen and Trachsel, 1988; Reddy et al., 1988) and *Fxr2h* (Zhang et al., 1995). Unexpectedly, between the two genes we found three different genes, *Cd68* (Holness et al., 1993; Li et al., 1998; Jiang et al., 1998), *Supl15h* (Ware and Lehrman, 1996, 1998; Mao et al., 1998), and *Sox15* (van de Wetering and Clevers, 1993).

In this report, we describe the structural organization of the five genes in a cluster that spans about 40 kb on mouse chromosome 11.

Materials and methods

Gene trap mouse line and genotyping

An ES cell line, GT3-11, into which a GT vector is introduced in a single copy, was established and transmitted into a germ line as described previously (Shirai et al., 1996). The GT3-11 mouse line was maintained on C57BL/6 X 129/terSv mixed genetic background. For genotyping of pups and embryos from heterozygote matings, genomic DNA was prepared according to Maniatis et al. (1989) and subjected to Southern blot analysis with a XbaI fragment of *Fxr2h* (nt positions 233 to 488; Accession No. AB025264) and/or PCR (Figs. 1 and 4A). Two pairs of primers that recognize the wild-type allele of *Eif4a1* (GT3-11-24/-25) or the trapped allele (LacZ-1/-2) were used. Amplified products were analyzed by electrophoresis on 1.2% agarose gels (GIBCO BRL) using EtdBr staining. Pups and embryos were scored as wild type (positive with only GT3-11-24/-25 PCR), heterozygous (positive with both PCR) or homozygous (positive with only LacZ-1/-2 PCR).

PCR

The primers used in this study are listed in Table 1. PCR was carried out under the following conditions; 93°C for 1 min, 60°C for 1 min, and 72°C for 2 min for 30 cycles, in a 10 ml solution of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 2.5 mM each of dNTPs, 10 mM each of primers and 0.4 U of Taq polymerase (TAKARA).

Table 1

Sequences of the primers used in this study.

Name	Sequence	Accession number (nt positions)
GT3-11-3	5' CGTCTAGAGTAACTGGAACGAGATTG 3'	AB011595 (2022 - 2040)
GT3-11-6	5' GCGTCGACTTGTACTGGACTAGATTGG 3'	AB011595 (2162 - 2180)
GT3-11-9	5' CTTCATGACATCAGTAACAG 3'	AB011595 (2418 - 2437)
GT3-11-10	5' TTTGGTCGTAAGGGTGTGGC 3'	AB011595 (5163 - 5182)
GT3-11-17[a]	5' CAAGAATTACGACTGTGCACCT 3'	
GT3-11-21[a]	5' CACTATCCACACTACAGTAGAC 3'	
GT3-11-24	5' TTTGGAGTGGTAGAACAGAGC 3'	AB011595 (1918 - 1938)
GT3-11-25	5' CTGCTGAGCCAATTCACGTG 3'	AB011595 (2617 - 2636)
GT3-11-26	5' GAGGGGCTGTCCTGCGA 3'	AB011595 (5287 - 5303)
GT3-11-27	5' TTTCCTGAATCAATAAATACTAT 3'	AB011595 (5763 - 5785)
GT3-11-28[a]	5' AAGTGCCTACCTGGATGATCC 3'	
MFXR2-1	5' TGTAAAGGATGTCCATGAAG 3'	AB025264 (974 - 993)
MFXR2-2	5' CTCTACTTCATCTCCTTCTG 3'	AB025264 (1873 - 1892)
MFXR2-3	5' AGTGGAGGGGTGTGTAACAG 3'	AB025311 (2641 - 2659)
MFXR2-4	5' CCGTCTTCCACAGTGATGAG 3'	AB025311 (2877 - 2896)
neo1[b]	5' CGAGATCTGTTGTGCCCAGTCATAG 3'	
neo2[b]	5' GTCCAGATCATCCTGATCGACA 3'	
LacZ-1[c]	5' GTTGGCAATTTAACCGCCAGT 3'	
LacZ-2[c]	5' ACCGTCGATATTTCAGCCATG 3'	
KS[d]	5' TCGAGGTCGACGGTATC 3'	
XT-18	5' CCCTCGAGGTCGACGGTATCGTTTTTTTTTTTTTTTTT 3'	

[a], Designed on the basis of *the Eif4a1* sequence previously reported by Reddy et al. (1988).

[b], *neo*-specific primer of GT vector

[c], *LacZ*-specific primer of GT vector

[d], Stratagene[R]

Analyses of the 5' and 3' flanking regions of the GT vector

For cloning of fusion transcripts with the GT vector, 5' RACE was performed using total RNA from GT3-11 ES cells as previously described (Ito et al., 1999). The first strand cDNA was synthesized with the *neo*-specific primer, neo2, and dATPs were added at the 3' end of the first strand cDNAs with terminal deoxynucleotidyl transferase (TAKARA). The dA-tailed cDNA was subjected to two rounds of nested PCR amplification. For the first round, using 10 pmol each of neo2 primer and oligo(dT)18 adaptor primer, XT-18, one cycle of PCR (94°C for 1 min, 42°C for 2 min, 72°C for 5 min) and then 2 more cycles (94°C for 1 min, 48°C for 2 min, 72°C for 5 min) were carried out in a total volume of 50 μl. After the 3 cycles of PCR, 15 pmol of neo2 primer and 50 pmol of KS primer (Stratagene) were added, followed by 30 cycles of PCR (94°C for 1 min, 65°C for 2 min, 72°C for 5 min) in a final volume of 100 μl. For the second round PCR, an aliquot of the first-round PCR products was added to a solution containing 50 pmol of the gene-specific primer, neo1, and 50 pmol of KS primer under the same conditions for the first-round PCR. The PCR products were digested with *Bgl*II and *Sal*I, subcloned into the pBluescript II KS(-) vector (Stratagene), and then sequenced. The

first-round PCR products were also amplified using neo1 and GT3-11-1 primers (Fig. 2A). For the amplified products, both agarose gel electrophoresis followed by EtdBr staining and sequencing were carried out.

To determine the 5' flanking region of the GT vector, genomic DNA of a heterozygote was amplified by PCR through the use of the GT3-11-3 primer for exon 3 of *Eif4a1* and the neo1 primer. The GT3-11-3/neo1 PCR product was ligated to the pCR2.1 vector (Invitrogen) and sequenced. Plasmid rescue was carried out to analyze the 3' flanking region of the GT vector integration site. Genomic DNA was prepared from GT3-11 ES cells and digested with *Eco*RI. Each *Eco*RI fragment was circularized, and then transfected into *E. coli.*, followed by selection on agar plates containing ampicillin. Transfectants with a DNA fragment containing the ampicillin-resistance gene of the GT vector and its 3' flanking region were selected. Plasmids derived from the transfectants were prepared and sequenced.

Two specific pairs of primers, the GT3-11-6 and GT3-11-9 for *Eif4a1* and the MFXR2-1 and MFXR2-2 for *Fxr2h*, were used for screening a genomic P1 library from the mouse 129 strain (Genome Systems). P1 plasmids were isolated according to the manufacturer's protocol, subcloned into pBluescript II SK(-) (Stratagene), and then sequenced.

Southern analysis

Genomic DNAs of GT3-11 wild-type mice were digested with *Bam*HI, *Eco*RI and *Xba*I, separated on a 0.7 % agarose gel, transferred, and then hybridized with [^{32}P]-labeled probes. The boxes marked with a-f under the map (Fig. 4B) represent the probes used for Southern blot analyses: probes a and b for *Eif4a1*, GT3-11-6/-9 (nt 2162 to 2437 of Accession No. AB011595) and GT3-11-26/-27 (nt 5287 to 5785 of Accession No. AB011595) PCR products, respectively; probe c for *Cd68*, *Bam*HI-*Kpn*I fragment (nt 2046 to 2353 of Accession No. AB009287); probe d for *Supl15h*, *Hind*III-*Sma*I fragment (nt 483 to 716 of Accession No. AB014471); probe e for *Sox15*, *Pst*I-*Sac*I fragment (nt 48 to 491 of Accession No. AB014474); and probe f for *Fxr2h*, MFXR2-3/-4 PCR product (nt 2641 to 2896 of Accession No. AB025311).

Genetic mapping

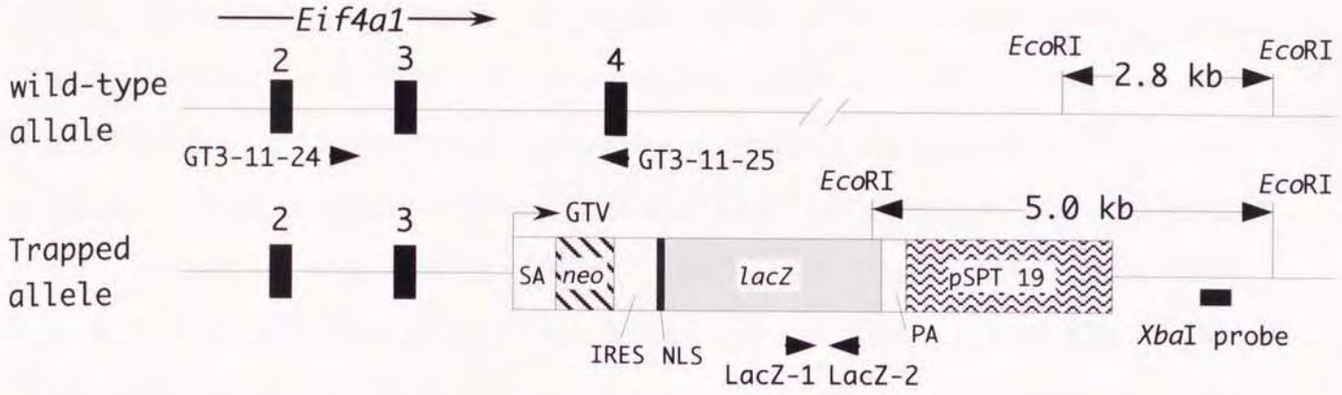
Two DNA panels from (C57BL/6J x M. spretus) F1 x C57BL/6J backcross mice (BSB panel) and (C57BL/6JEi x SPRET/Ei) F1 x SPRET/Ei backcross mice (BSS panel) were obtained from the Jackson Laboratory. PCR, with the GT3-11-3 and GT3-11-9 primers, was performed to define the locus polymorphism. The results were transmitted back to the database of Jackson Laboratory for comparison with the existing map.

Results

Embryonic lethality of GT3-11 homozygotes

GT3-11 heterozygotes are fertile. Genotype analysis of postnatal mice derived from heterozygous parents was carried out by Southern blotting using the *Xba*I probe. The sizes of the wild-type and mutant *Eco*RI fragments were 2.8 kb and 5.0 kb, respectively (Fig. 1C). There were no homozygotes among 163 mice (Table 2). These results suggested that the embryonic lethality was associated with homozygosity for the integration of a GT vector. To determine the time of embryonic lethality, timed matings were established, and the resulting embryos were dissected free of maternal tissues and genotyped by PCR (Table 2). Homozygotes were not obtained from E8.5 to E14.5. At E3.5, homozygous embryos that were morphologically normal were recovered in the expected Mendelian ratio. These results indicate that in GT3-11 the GT vector insertion causes early embryonic lethality of homozygotes between E3.5 and E8.5 of gestation.

A



B

		Genotype		
		W	Ht	Hm
PCR	GT3-11-24◄ ►GT3-11-25	+	+	-
	LacZ-1◄ ►*LacZ*-2	-	+	+
Southern blotting(kb)		2.8	2.8/5.0	5.0

W,wild-type; Ht, heterozygous; Hm, homozygous

C

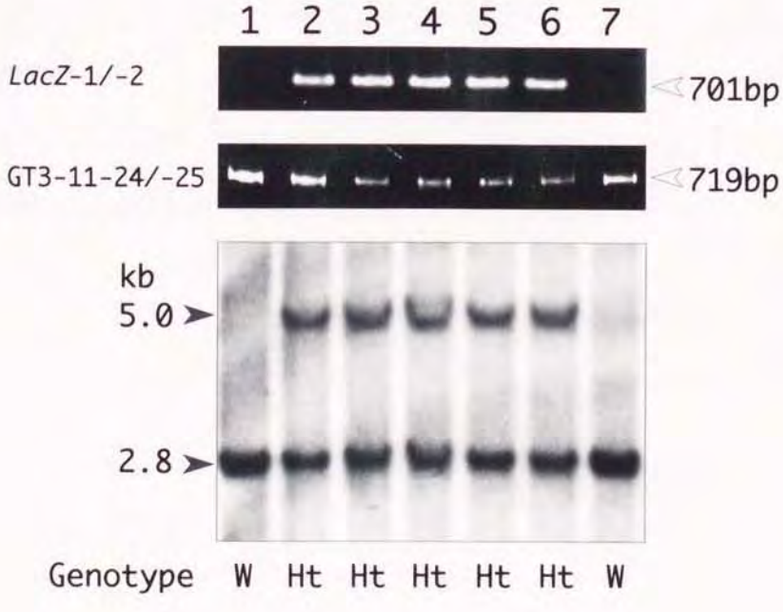


Fig. 1. Genotyping of GT3-11 mice

(A) The schema for identification of genotypes. PCR with GT3-11-24 and -25 and with LacZ-1 and -2 recognize wild-type allele and mutant allele, respectively. The filled box (*Xba*I probe) under the trapped allele shows the probe for Southern blot analysis that recognize the 2.8-kb *Eco*RI fragment of wild-type allele and the 5.0-kb fragment of the trapped allele. Exons of *Eif4a1* are indicated by the filled boxes with numbers. Arrows indicate the transcription orientations of GT vector (GTV) and *Eif4a1*. SA, splice acceptor site of E-cadherin gene; *neo*, *E. coli*. neomycin phosphotransferase gene; IRES, internal ribosomal entry site from encephalomyocarditis virus; NLS, nuclear localization signal from SV40 large T antigen; *LacZ*, *E. coli*. β-galactosidase gene; pA, polyadenylation signal of thymidine kinase gene; pSPT19, a plasmid purchased from Pharmacia Biotech.

(B) The predicted pattern of genotyping analyzed by PCR and Southern blotting.

(C) PCR and Southern blot analyses of genotypes. Lanes 1-7 show the results of genotyping in the individual postnatal mouse derived from a female.

Table 2.
Genotypes of offspring derived from heterozygous parents

stage	litters (no.)	embryos (no.)	genotype			
			W	Ht	Hm	N.D.
embryo						
E3.5	5	42	11	22	8	1
E8.5	5	47	6	24	0	17
E10.5	4	36	10	16	0	10
E14.5	1	14	2	8	0	4
Total	15	139	29	70	8	
postnatal[a]						
Male			35	58	0	
Female			20	50	0	
Total		163	55	108	0	

[a] Postnatal days 21 to 56
W ; Wild-type, Ht ; Heterozygous, Hm ; Homozygous
N.D.; Genotype could not determined.

Analyses of the insertion sites of a GT vector

We attempted to identify the gene related to the early embryonic lethality of homozygotes. Analysis of 5' RACE was used to determine the break point of a fusion protein with a GT vector. This analysis revealed that two fusion transcripts, clone L and S, were produced in the trapped allele of the GT3-11 ES cells (Fig. 2). It was found that the clone L contained cDNA corresponding to exons 1, 2 and 3 of *Eif4a1* (Nielsen et al., 1985, Nielsen and Trachsel, 1988; Reddy et al., 1988), and the clone S was the transcript in which exon 2 of *Eif4a1* was alternatively spliced out. The alternative splicing of the exon 2 was not detect in the wild-type allele of the GT3-11 ES cells (data not shown).

To determine whether or not the GT vector was actually integrated into the *Eif4a1* locus, the sequence between the splicing acceptor site of the GT vector and exon 3 of *Eif4a1* was determined through the use of PCR with the GT3-11-3 and neo1 primers and genomic DNA from GT3-11 heterozygote. An about 950 bp fragment was amplified and sequenced. The 5' end of the GT vector was flanked by 263 bp of intron 3 of *Eif4a1* (Fig. 3).

To identify the 3' insertion site of the GT vector, we carried out genomic cloning of the flanking region by means of plasmid rescue. An approx. 5.0 kb *Eco*RI fragment, containing the 3' part of the GT vector and an about 2.0 kb fragment of the ES cell genome, was

kb fragment of the ES cell genome, was obtained from ampicillin-resistant clones and its sequence was determined (Fig. 3). A search of available DNA and protein data bases revealed that the rescued plasmid contained exons 2 and 3 corresponding to nt positions 309 to 361 (identity, 94%) and 362 to 455 (identity, 92%) of human *FXR2* cDNA (Zhang et al., 1995), respectively. These exons conform to the gt-ag boundary rule. The amino acids sequences deduced from the sequences of these exons were shown to be identical with amino acid residues 28 to 76 of the human FXR2 protein. The transcriptional orientation of *Fxr2h* was opposite to those of the GT vector and *Eif4a1* (Figs. 3 and 4). These results indicated that the GT vector was integrated between intron 3 of *Eif4a1* and intron 3 of *Fxr2h*.

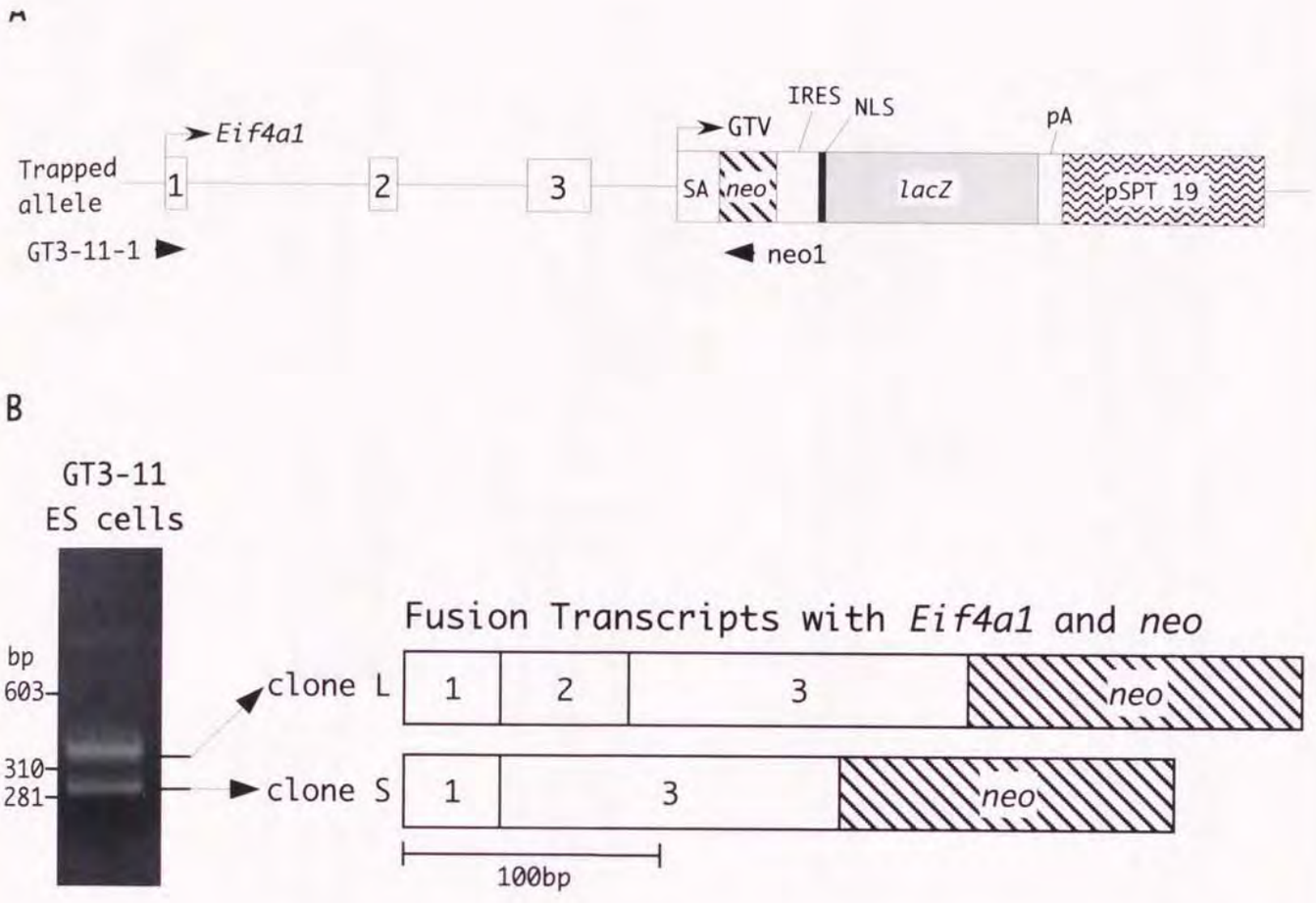


Fig. 2. Two fusion transcripts with the GT vector.
(A) Schematic diagram of the 5' region of the GT vector (GTV) in trapped allele. Open boxes with numbers indicate the exons of *Eif4a1*. Horizontal arrowheads represent primers (GT3-11-1 and neo1) used in 5' RACE. Transcription orientations of *Eif4a1* and GTV are indicated by horisontal arrows.
(B) It was found by 5' RACE that the GTV was fused with *Eif4a1* in the trapped allele of the GT3-11 ES cells. The clones L (containing exons 1, 2 and 3 of *Eif4a1*) and S (containg exons 1 and 3 of *Eif4a1*) show the fusion transcript with the GTV. The sizes of the clones L and S are 348 bp and 299 bp, respectively. ϕX174-*Hae*III digests (TOYOBO) were used as standards of DNA fragment length (shown in left of the lane).

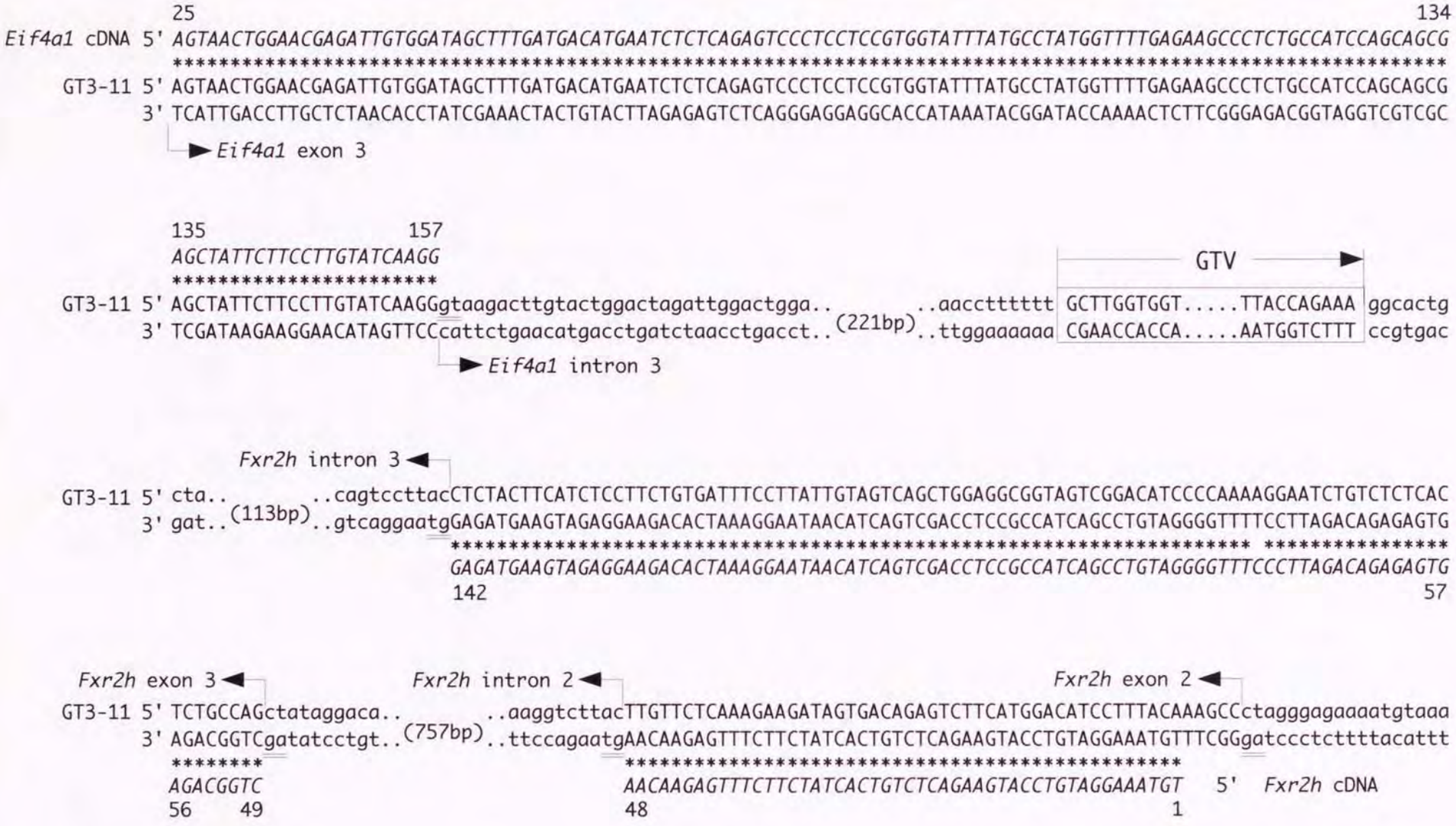

25
134
Eif4a1 cDNA 5' AGTAACTGGAACGAGATTGTGGATAGCTTTGATGACATGAATCTCTCAGAGTCCCTCCTCCGTGGTATTTATGCCTATGGTTTTGAGAAGCCCTCTGCCATCCAGCAGCG
GT3-11 5' AGTAACTGGAACGAGATTGTGGATAGCTTTGATGACATGAATCTCTCAGAGTCCCTCCTCCGTGGTATTTATGCCTATGGTTTTGAGAAGCCCTCTGCCATCCAGCAGCG
3' TCATTGACCTTGCTCTAACACCTATCGAAACTACTGTACTTAGAGAGTCTCAGGGAGGAGGCACCATAAATACGGATACCAAAACTCTTCGGGAGACGGTAGGTCGTCGC
Eif4a1 exon 3
135
157
AGCTATTCTTCCTTGTATCAAGG
GTV
GT3-11 5' AGCTATTCTTCCTTGTATCAAGGgtaagacttgtactggactagattggactgga..
..aacctttttt GCTTGGTGGT.....TTACCAGAAA ggcactg
3' TCGATAAGAAGGAACATAGTTCCcattctgaacatgacctgatctaacctgacct..
(221bp)
..ttggaaaaaa CGAACCACCA.....AATGGTCTTT ccgtgac
Eif4a1 intron 3
Fxr2h intron 3
GT3-11 5' cta..
..cagtccttacCTCTACTTCATCTCCTTCTGTGATTTCCTTATTGTAGTCAGCTGGAGGCGGTAGTCGGACATCCCCAAAAGGAATCTGTCTCTCAC
3' gat..
(113bp)
..gtcaggaatgGAGATGAAGTAGAGGAAGACACTAAAGGAATAACATCAGTCGACCTCCGCCATCAGCCTGTAGGGGTTTTCCTTAGACAGAGAGTG
GAGATGAAGTAGAGGAAGACACTAAAGGAATAACATCAGTCGACCTCCGCCATCAGCCTGTAGGGGTTTCCCTTAGACAGAGAGTG
142
57
Fxr2h exon 3
Fxr2h intron 2
Fxr2h exon 2
GT3-11 5' TCTGCCAGctataggaca..
..aaggtcttacTTGTTCTCAAAGAAGATAGTGACAGAGTCTTCATGGACATCCTTTACAAAGCCctagggagaaaatgtaaa
3' AGACGGTCgatatcctgt..
(757bp)
..ttccagaatgAACAAGAGTTTCTTCTATCACTGTCTCAGAAGTACCTGTAGGAAATGTTTCGGgatccctcttttacattt
AGACGGTC
AACAAGAGTTTCTTCTATCACTGTCTCAGAAGTACCTGTAGGAAATGT
5' Fxr2h cDNA
56
49
48
1

Fig. 3. Nucleotide sequences of the insertion site.
Italicized upper-case letters represent the sequences of *Eif4a1* and *Fxr2h* cDNAs. Nt of *Eif4a1* cDNA (25-157) shown here is numbered based on the sequence reported previously by Nielsen et al. (1985). The number under *Fxr2h* cDNA (1-142) is from the sequence of accession No. AB025269. Identical nt between the sequences of GT3-11 and *Eif4a1* cDNA, and GT3-11 and *Fxr2h* cDNA are denoted by asterisks. The sequences of exons and GT vector (GTV) are shown in upper-case letters and those of introns in lower-case ones. The dinucleotides, gt and ag, at splice sites are double-underlined. Arrows indicate the transcription orientation.

Cluster of five different genes

In order to analyze the pre-integration region between *Eif4a1* and *Fxr2h* in wild-type ES cells, two P1 clones, P1-1 and P1-2, were isolated from the P1 library (Fig. 4B). The P1-2 clone included the complete region spanning *Eif4a1* to *Fxr2h*. Restriction analyses and partial nt sequencing revealed that the integration of the GT vector was accompanied by the deletion of about 35 kb of endogenous DNA, and that the deleted region was found to include three different genes, *Cd68*, *Supl15h* and *Sox15*. The entire exon/intron structures of these five different genes were determined (Fig. 4B).

Eif4a1 consists of 11 exons, which are identical with the results for *Eif4a1* reported by Reddy et al. (1988) and Nielsen and Trachsel (1988). However, the sequences of introns of *Eif4a1* in GT3-11 were slightly different from those of *Eif4a1* reported by Reddy et al. (1988). To verify the sequences reported by Reddy et al. (1988), PCR was performed using three pairs of primers (GT3-11-3/-17, GT3-11-10/-21 and GT3-11-28/-17) designed on the basis of their sequences and genomic DNAs of C57BL/6 and ICR as templates. No purposive PCR products, however, were detected. On the other hand, we obtained 396 bp and 547 bp PCR products with GT3-11-3/-9 and GT3-11-6/-9, respectively (data not shown). The GT3-11-6 and -9 primers were shown to be unique for the *Eif4a1* sequence described in this study. This

in this study. This discrepancy of the sequences between both clones may reflect the difference of libraries used for genomic cloning.

In the course of structural analysis and sequencing of an about 10 kb-*Xba*I fragment containing *Eif4a1*, we found another gene, *Cd68*. Exon 1 of *Cd68* is located 782 bp downstream of the polyA addition site of the *Eif4a1* mRNA long form (GenBank Accession No. X03040). *Cd68* consists of 6 exons and its transcriptional orientation is the same as that of *Eif4a1*.

To reach the *Fxr2h*, another site of the integration, we subcloned several restriction fragments from the P1 clones and sequenced them. Unexpectedly, the transcriptional start site of *Supl15h* was found to lie about 1.7 kb downstream of the polyA addition site of *Cd68*. *Supl15h* consists of 7 exons, which were determined on comparison with the *Supl15h* cDNA sequence (Accession No. AB024713). It was transcribed in the same direction as *Cd68*. The 3' end of the *Supl15h* cDNA was determined by alignment with mouse EST sequences (GenBank Accession Nos. W53527 and AA140375) and Chinese hamster *SL15* cDNA (Ware and Lehrman, 1996, 1998). Furthermore, *Sox15*, consisting of two exons, exists within about 2.0 kb downstream of *Supl15h*. *Supl15h* and *Sox15* were found to share the 3' ends of the transcripts in the opposite transcription orientation. A conserved sequence (consensus YGTGTTYY; Y = C or T; McLauchlan et al., 1985) is located approx. 30 bp downstream from the polyadenylation signal. There

downstream from the polyadenylation signal. There were TGTGTTTA sequence 34 bp downstream of the polyadenylation signal of *Supl15* and TGTATTCT 32 bp dowmstream of that of *Sox15* (Fig. 5).

The polyA addition site of *Fxr2h* lay about 2.0 kb upstream of the transcription start site of *Sox15*. The transcriptional orientation of *Fxr2h* was the same as that of *Sox15*. *Fxr2h* consisted of the 17 exons, which was determined in comparison with mouse *Fxr2h* cDNA (Accession No. AB0252639).

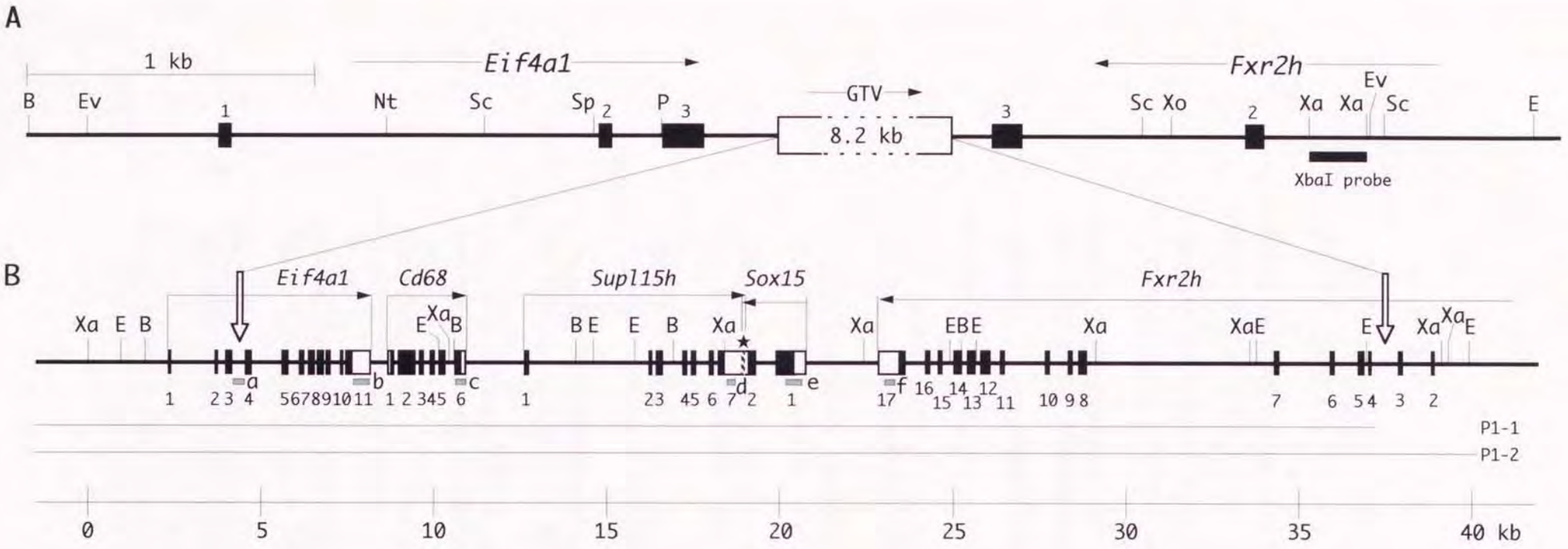


Fig. 4. Genomic organization of the integration region (A) and pre-integration region (B) of the transgene.

Five different genes, *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h*, are disrupted on the integration of the GT vector (GTV). The exons of each genes are represented by boxes with numbers. The 5' and 3' UTRs are represented open boxes. B, E, Ev, Nt, P, Sc, Sp, Xa and Xo indicate *Bam*HI, *Eco*RI, *Eco*RV, *Not*I, *Pst*I, *Sac*I, *Spe*I, *Xba*I and *Xho*I, respectively. The overlapping region of *Supl15h* and *Sox15* is indicated by the shaded box marked with a filled star, and its sequence is shown in Fig. 5. The transcription orientations of individual genes and GTV are indicated by horizontal arrows. Open vertical arrows indicate the insertion sites of GTV (B). The *Xba*I probe in (A) is used in genotyping (Fig. 1A).

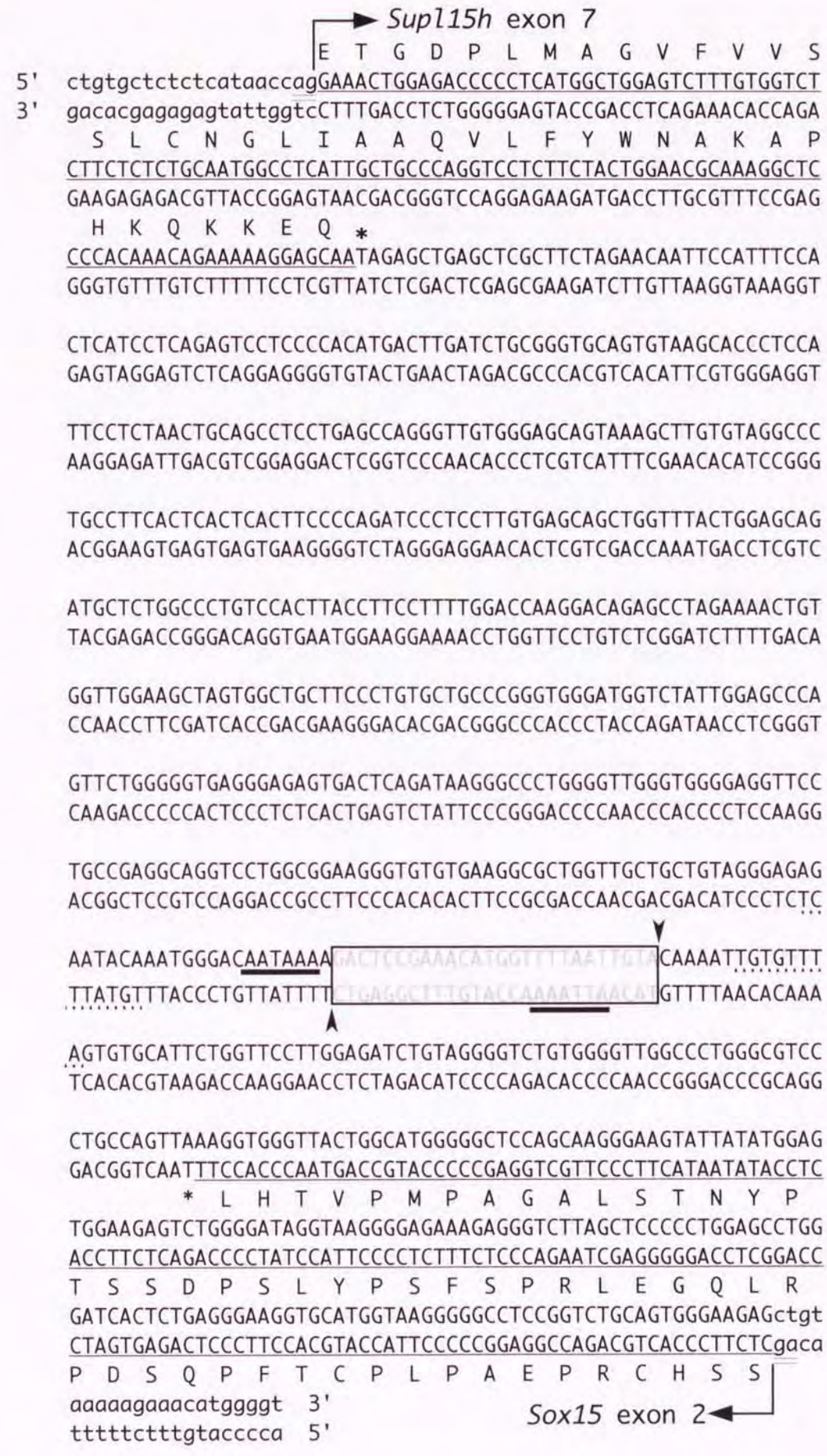

Sup115h exon 7
E T G D P L M A G V F V V S
5' ctgtgctctctcataaccagGAAACTGGAGACCCCCTCATGGCTGGAGTCTTTGTGGTCT
3' gacacgagagagtattggtcCTTTGACCTCTGGGGGAGTACCGACCTCAGAAACACCAGA
S L C N G L I A A Q V L F Y W N A K A P
CTTCTCTCTGCAATGGCCTCATTGCTGCCCAGGTCCTCTTCTACTGGAACGCAAAGGCTC
GAAGAGAGACGTTACCGGAGTAACGACGGGTCCAGGAGAAGATGACCTTGCGTTTCCGAG
H K Q K K E Q *
CCCACAAACAGAAAAAGGAGCAATAGAGCTGAGCTCGCTTCTAGAACAATTCCATTTCCA
GGGTGTTTGTCTTTTTCCTCGTTATCTCGACTCGAGCGAAGATCTTGTTAAGGTAAAGGT
CTCATCCTCAGAGTCCTCCCCACATGACTTGATCTGCGGGTGCAGTGTAAGCACCCTCCA
GAGTAGGAGTCTCAGGAGGGGTGTACTGAACTAGACGCCCACGTCACATTCGTGGGAGGT
TTCCTCTAACTGCAGCCTCCTGAGCCAGGGTTGTGGGAGCAGTAAAGCTTGTGTAGGCCC
AAGGAGATTGACGTCGGAGGACTCGGTCCCAACACCCTCGTCATTTCGAACACATCCGGG
TGCCTTCACTCACTCACTTCCCCAGATCCCTCCTTGTGAGCAGCTGGTTTACTGGAGCAG
ACGGAAGTGAGTGAGTGAAGGGGTCTAGGGAGGAACACTCGTCGACCAAATGACCTCGTC
ATGCTCTGGCCCTGTCCACTTACCTTCCTTTTGGACCAAGGACAGAGCCTAGAAAACTGT
TACGAGACCGGGACAGGTGAATGGAAGGAAAACCTGGTTCCTGTCTCGGATCTTTTGACA
GGTTGGAAGCTAGTGGCTGCTTCCCTGTGCTGCCCGGGTGGGATGGTCTATTGGAGCCCA
CCAACCTTCGATCACCGACGAAGGGACACGACGGGCCCACCCTACCAGATAACCTCGGGT
GTTCTGGGGGTGAGGGAGAGTGACTCAGATAAGGGCCCTGGGGTTGGGTGGGGAGGTTCC
CAAGACCCCCACTCCCTCTCACTGAGTCTATTCCCGGGACCCCAACCCACCCCTCCAAGG
TGCCGAGGCAGGTCCTGGCGGAAGGGTGTGTGAAGGCGCTGGTTGCTGCTGTAGGGAGAG
ACGGCTCCGTCCAGGACCGCCTTCCCACACACTTCCGCGACCAACGACGACATCCCTCTC
AATACAAATGGGACAATAAAA CAAAATTGTGTTT
TTATGTTTACCCTGTTATTTT GTTTTAACACAAA
AGTGTGCATTCTGGTTCCTTGGAGATCTGTAGGGGTCTGTGGGGTTGGCCCTGGGCGTCC
TCACACGTAAGACCAAGGAACCTCTAGACATCCCCAGACACCCCAACCGGGACCCGCAGG
CTGCCAGTTAAAGGTGGGTTACTGGCATGGGGGCTCCAGCAAGGGAAGTATTATATGGAG
GACGGTCAATTTCCACCCAATGACCGTACCCCCGAGGTCGTTCCCTTCATAATATACCTC
* L H T V P M P A G A L S T N Y P
TGGAAGAGTCTGGGGATAGGTAAGGGGAGAAAGAGGGTCTTAGCTCCCCCTGGAGCCTGG
ACCTTCTCAGACCCCTATCCATTCCCCTCTTTCTCCCAGAATCGAGGGGGACCTCGGACC
T S S D P S L Y P S F S P R L E G Q L R
GATCACTCTGAGGGAAGGTGCATGGTAAGGGGGCCTCCGGTCTGCAGTGGGAAGAGctgt
CTAGTGAGACTCCCTTCCACGTACCATTCCCCCGGAGGCCAGACGTCACCCTTCTCgaca
P D S Q P F T C P L P A E P R C H S S
aaaaagaaacatggggt 3'
ttttctttgtacccca 5'
Sox15 exon 2

Fig. 5. Overlapping region of *Supl15h* and *Sox15*.

The sequence of exons and introns are shown in upper-case letters and lower-case letters, respectively. The coding sequence of both *Supl15h* and *Sox15* are underlined. The amino acids sequences, depicted in one letter notation, are shown above the coding sequence of *Supl15h* and below that of *Sox15*. The splice acceptor sites are double underlined. Asterisks show stop codons of both genes (TAG for *Supl15h* and TAA for *Sox15*). Polyadenylation signals (AATAAA for *Supl15h* and ATTAAA for *Sox15*) and polyA additional sites are indicated by thick lines and arrowheads, respectively. The transcription termination region (26 nt) shared by *Supl15h* and *Sox15* is presented in gray upper-case letters in a box. Dotted lines indicate the conserved sequences located approx. 30 bp downstream from the polyadenylation signal (McLauchlan et al., 1985).

Chromosomal mapping

Although *Eif4a1* exists as a multi-copy gene including pseudogenes (Nielsen and Trachsel, 1988), a single hybridization band was observed on Southern blots when probe a (PCR product with GT3-11-6/-9) was used (Fig. 7a and b). The pre-integration region of the GT vector was mapped using the BSB and BSS panels from the Jackson Laboratory. *Eif4a1* was mapped to the central part of mouse chromosome 11 (Fig. 8). Raw data (code 587) are available from the World Wide Web site: http://lena.jax.org/resources/documents/cmdata/bkmap/BSS.(or BSB.)html. *Eif4a1* and *Cd68* have already been mapped to mouse chromosome 11 by Nielsen et al. (1993) and Jiang et al. (1998), respectively. *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* are each present in a single copy in the mouse genome (Fig. 7c, d and e). Therefore, judging from the results of Southern blot and sequence analyses of this locus described above together, it is obvious that not only *Eif4a1* and *Cd68* but also *Supl15h*, *Sox15* and *Fxr2h* map to mouse chromosome 11.

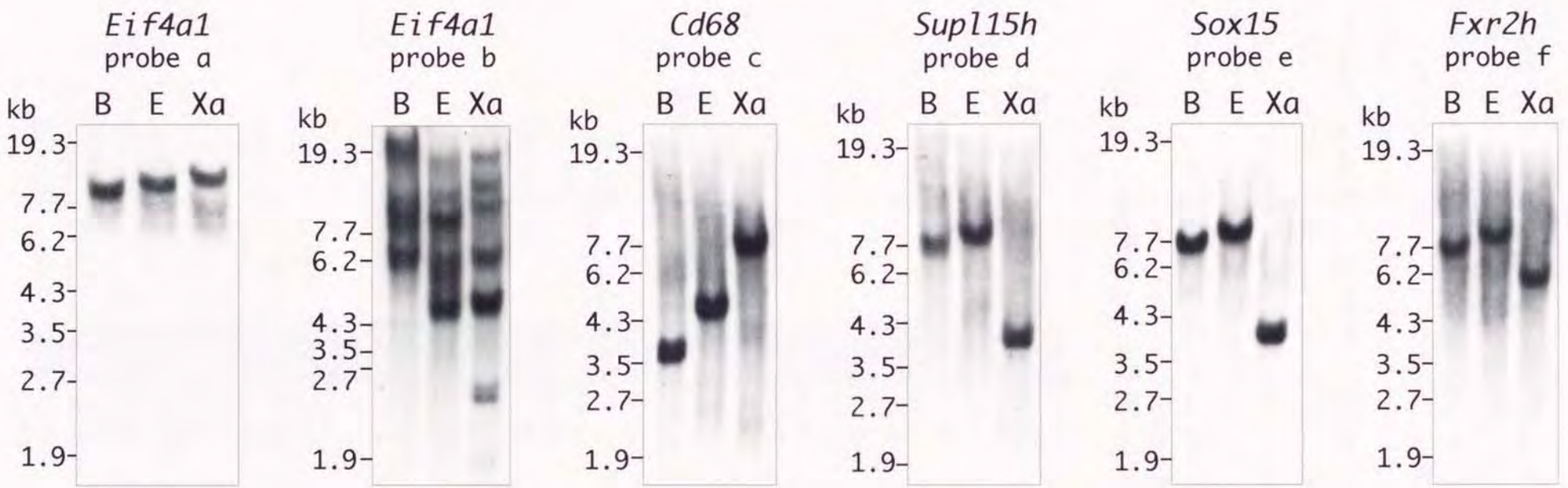


Fig. 6. Genomic Southern analysis.
Genomic DNAs from GT3-11 wild-type mice were cleaved with *Bam*HI (B), *Eco*RI (E) and *Xba*I (Xa), fractionated on agarose gels, blotted onto nitrocellulose filters, and hybridized to ^{32}P-labeled probes as described in ***Materials and Methods***. The positions of each probes (probe a - f) are shown in Fig. 4B. λ-*Eco*T22I digests (TAKATA) were used as standars of DNA fragment length.

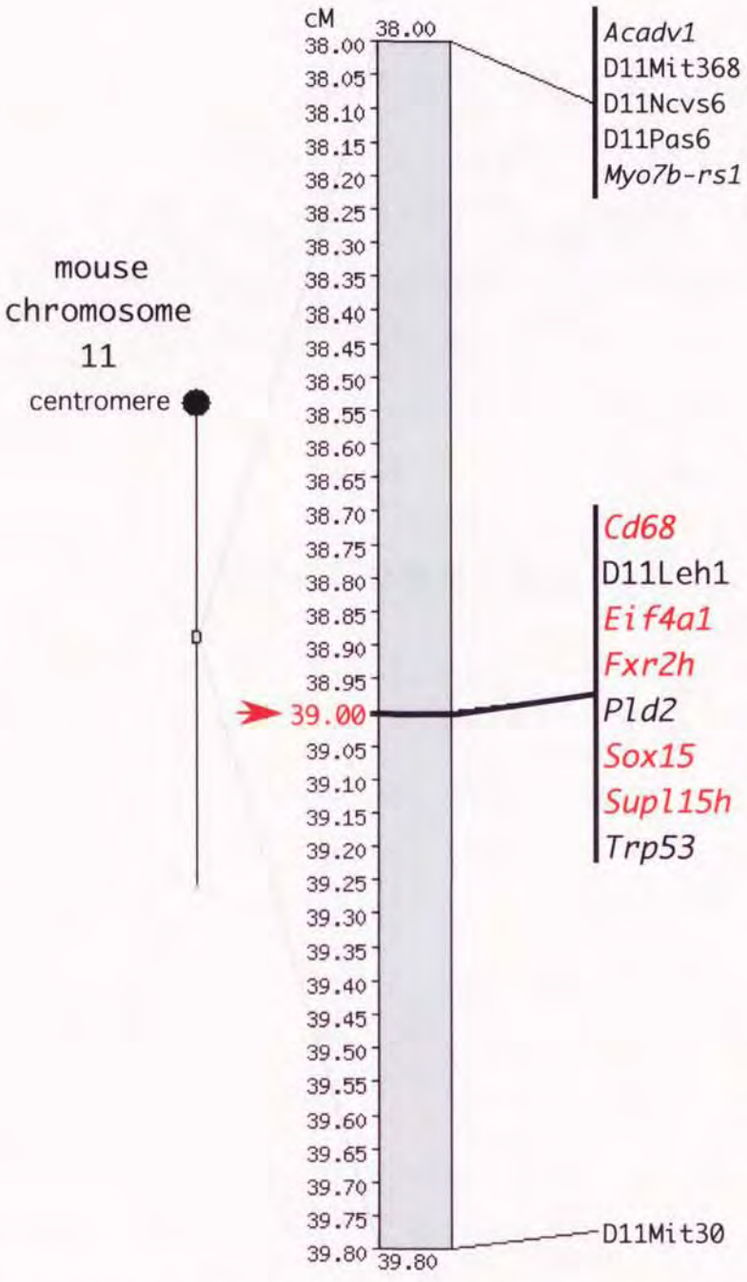


Fig. 7. Chromosomal location of *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h*.

The integration locus of the GT vector in GT3-11 was mapped to 39.00 cM from the centromere of mouse chromosome 11. The red arrow indicates the locus where *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* (shown in red letters) are clustered. This locus was found to be very close to *Trp53* coding mouse p53 protein, well-known as a cellular tumour antigen. This figure was obtained from the World Wide Web (http://www.informatics.jax.org/searches/linkmap.cgi?chromosome=11&midpoint=39.0&cmrange=1.0&dsegments=1&syntenics=0) which is the site of the Jackson Laboratory, and partially modified.

Discussion

We characterized the transgene integration region that is associated with the homozygous lethality of the GT3-11 mouse line. GT vector integration generated a genomic DNA deletion that spans a distance of about 35 kb on chromosome 11 and excludes the five independent genes, *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* .

Hiraoka et al. (1998) have reported that the SOX20 protein encompasses an SRY-type HMG box and a C-terminal domain exhibiting strong homologies to those of the Sox15 protein, and both genes have a close evolutionary relationship. It has been shown that *SOX20* is localized on human chromosome 17p13 and is within 35 kb of a new fragile X mental retardation 1 gene (*FMR1*) homolog different from the fragile X-related 1 gene (*FXR1*) (Siomi et al., 1995; Meyer et al., 1996). More precise chromosomal localization of *SOX20* at 17p13.1 has been reported by Vujic et al. (1998). *FXR2* has been mapped to human chromosome 17 at 17p13.1 (Zhang et al., 1995). The human *SL15* gene has been mapped to chromosome 17p12-13.1 (Mao et al., 1998), and shares the 3' end of a transcript with *SOX20* in the opposite direction (Accession No. AB025355). Taken these findings and our present data together, it is likely that mouse *Sox15* is a homolog of human *SOX20*. In this context, we can be fairly certain that the region including *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* in the mouse genome is strongly conserved even in the human genome.

The genomic structure, sequence and chromosomal

The genomic structure, sequence and chromosomal location of *Cd68* were recently reported (Li et al., 1998; Jiang et al., 1998). *Eif4a1* lies upstream of *Cd68* (GenBank Accession No. AF045554), which is conserved in the human genome on chromosome 17p13 (Jones et al., 1998). Our results coincide with these findings. Li et al. (1998) have also shown that the region located between -7.0 and -2.5 kb of the translation initiation site of *Cd68* exhibits strong enhancer activity in macrophages and repressor activity in nonmyeloid cells. Surprisingly, *Eif4a1* is obviously within this region. Their observations and our findings suggest that gene-targeting/trapping in a gene-cluster region could affect the expression of not only a transgene but also neighboring genes.

In addition we found *Supl15h* and *Sox15* are transcribed in opposite directions and share the 3' end of the transcription termination region. Like these of *Supl15h* and *Sox15*, the transcripts of *Nab2* and *Stat6* converge, such that the 3' ends of the *Stat6* and *Nab2* mRNAs overlap by 58 bp (Svaren et al., 1997). The 3' end of the *FLII* transcript overlaps that of the *LLGL* transcript, and the corresponding mouse genes, *Fliih* and *Llglh*, also overlap (Campbell et al., 1997). Overlapping transcripts are disrupted in the ROSA26 mouse strain created by random retroviral gene trapping, and the ROSA26 genomic region produces convergent and overlapping antisense transcripts (Zambrowicz et al., 1997). In rat and man, *FXR2* and

and man, *FXR2* and androgen-binding protein/sex hormone-binding globulin genes (*ABP/SHBG*) overlap with the 5' region of each gene transcribed in the opposite direction (Joseph, 1998; GenBank Accession No. AF044263). Human *ABP/SHBG* has been mapped to chromosome 17p13.1 (Joseph, 1998), and mouse *ABP/SHBG* to chromosome 11 (Joseph et al., 1991). Therefore it is possible that *Fxr2h* and mouse *ABP/SHBG* may also share regulatory elements of these genes. We do not know whether or not additional genes, including *ABP/SHBG*, adjacent to this region exist. Mutations in these overlapping regions may affect the expression of both genes, although it remains unclear whether or not this characteristic structure has any functional significance as to transcriptional regulation in a cell.

Although the GT strategy essentially depends on the random integration of a vector, there are several examples of the same genes being trapped independently: *jumonji* (Takeuchi et al., 1995; Baker et al., 1997; Voss et al., 1998), *α-enolase* (Chowdhury et al., 1997; Couldrey et al., 1998), *R-PTP-κ* (Skarnes et al., 1995; Chowdhury et al., 1997), *Etl-1* (Korn et al., 1992; Voss et al., 1998), *Eif4a1* in a B-lineage cell line (Kerr et al., 1996), and *Tiam-1* (Voss et al., 1998) in BW5147 T-lymphoma cells (Habets et al., 1994). The repeated trapping of these genes with different integration procedures, i.e. retrovirus infection or electoporation

of GT constructs, suggests the existence of integration hot spots. A larger amount of sequencing information on trapped genes will be necessary to address this issue.

As described above, it is essential to the GT strategy that the roles of a trapped gene are analyzed with due regard to genomic deletions or rearrangements and intergation sites in the mouse genome like GT3-11. In this paper, a cluster of *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* on mouse chromosome 11 in the pre-integration region of a transgene has been reported. It will be important to analyze the expression of *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h* in different development stages to determine the candidate gene(s) for the embryonic death of GT3-11 homozygotes.

General discussion

With the advancement of genetic and physical mapping in mammalian genomes, comparative genomics in mammals is now progressing (O'Brine et al., 1999). A sequencing of the human genome is predicted to be completed by end of 2003 (Collins et al., 1998). A sequence information in the form of ESTs dramatically increases, and over half of the 70,000 to 100,000 human ESTs are already mapped, and even more gene products have sequence representation in mouse and human EST databases (O'Brine et al., 1999). Therefore, gene trapping using mouse ES cells is thought to be a comprehensive approach to an experimental genetics that will illuminate the function of genes *in vivo* (Evans, 1998).

Role of gene trapping for the future.

The GT strategy in mouse ES cells is now expected as the valuable method to facilitate the study of gene function on a large scale. Wurst et al. (1995) used the GT strategy in mouse ES cells to screen for insertional mutations in genes developmentally regulated at 8.5 days of embryogenesis on a large scale for the first time. Hicks et al. (1997) showed by large-scale gene trapping that many of the 10,000-20,000 genes are expressed in mouse ES cells. Zambrowicz et al. (1998) reported that a high-throughput mutagenesis method based on the usual GT strategy trapped and mutated 2,000 genes regardless of their expression status in ES

kept in the library called Omnibank (http://www.lexgen.com/aboutomnibank.php3), which now includes about 70,000 mouse genetic clone (Dalton, 1999).

The dramatic increase in sequence information of both ESTs and genomic DNA has created a ' gene function gap '. Namely, the identification of novel genes far outpaces the rate at which their function can be identified. Therefore, it is obvious that mutant mice created by the GT strategy on a large scale are valuable in the functional genomics.

Through the analysis of GT3-11 mouse line, we revealed that five different genes, *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h*, are clustered in a 40 kb region of mouse chromosome 11 (Miyashita et al., 1999). This is the first report of gene trapping in which integration of the GT vector caused about 35 kb of genomic deletion and disruption of the five genes, *Eif4a1*, *Cd68*, *Supl15h*, *Sox15* and *Fxr2h*. There are some reports showing that microdeletion (< 5 Mb) of an autosomal chromosome gives rise to a contiguous gene syndrome (for review, see Strachan and Read, 1996). To analyze the function of large chromosomal region, Su et al. (2000) and Zheng et al. (2000) have exploitted a system based on Cre-loxP-mediated recombination to generate chromosomal deletions with sizes extending from a few kb to several Mb in ES cells. In this cotext, the GT strategy associated with genomic deletion such as GT3-11 mouse line is very meaningful

gene(s) resposible for mutant phenotypes.

Acknowledgment

I wish to express my cordial thanks to Dr. Shoji Odani (Faculty of Science, Niigata University) and Dr. Ryozo Kuwano (Research Laboratory for Molecular Genetics, Niigata University) for many helpful suggestions and encouragements throughout the present study. I am very grateful to Dr. Hitoshi Sasai (Japan Tobacco Inc.) for the helpful discussion, and the Jackson Laboratory for providing the interspecific backcross DNA panels. Finally, but not least, my thanks are extended to the members of the Research Laboratory for Molecular Genetics (Niigata University) for their encouragements.

*This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan (to R.K).

References

Antoch, M.P., Song, E.J., Chang, A.M., Vitaterna, M.H., Zhao, Y., Wilsbacher, L.D., Sangoram, A.M., King, D.P., Pinto, L.H., Takahashi, J.S., **1997**. Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. Cell 89, 655-667.

Baker, R.K., Haendel, M.A., Swanson, B.J., Shambaugh, J.C., Micales, B.K., Lyons, G.E., **1997**. In vitro preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. Dev. Biol. 185, 201-214.

Carlton, M.B., Colledge, W.H., Evans, M.J., **1998**. Crouzon-like craniofacial dysmorphology in the mouse is caused by an insertional mutation at the Fgf3/Fgf4 locus. Dev. Dyn. 212, 242-249.

Campbell, H.D., Fountain, S., Young, L.G., Claudianos, C., Hoheisel, J.D., Chen, K-S., Lupskis, J.R., **1997**. Genomic structure, evolution, and expression of human FLII, a gelsolin and leucine-rich-repeat family member: overlap with LLGL. Genomics 42, 46-54.

Capecchi, M.R., **1989**. Altering the genome by homologous recombination. Science 244, 1288-1292.

Chen, Z., Friedrich, G.A., Soriano, P., **1996**. Genetic variants and strains of the laboratory mouse. Oxford Univ. Press. 3rd Ed. pp1756-1773.

Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A., Gruss, P., **1997**. Evidence of stochastic integration of gene trap vector into the mouse germline. Nucleic Acids Res. 25, 1531-1536.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., **1998**. New goals for the U.S. Human Genome Project: 1998-2003. Science 282, 682-689.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., Lincoln, D.S., Nadeau, J.H., Lander, E.S., **1993**. A genetic linkage map of the mouse: current applications and future prospects. Science 262, 57-66.

S.E., Steen, R.G., Lincoln, D.S., Nadeau, J.H., Lander, E.S., **1993**. A genetic linkage map of the mouse: current applications and future prospects. Science 262, 57-66.

Couldrey, C., Carlton, M.B., Ferrier, J., Colledge, W.H., Evans, M.J., **1998**. Disruption of murine alpha enolase by a retroviral gene trap results in early embryonic lethality. Dev. Dyn. 212, 284-292.

Dalton, R., **1999**. Terms of access to cloned mice comes under researchers' fire. Nature 401, 515-516.

Evans, M.J., Kaufman, M.H., **1981**. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156.

Evans, M.J., **1998**. Gene trapping-A preface. Dev. Dyn. 212, 167-169.

Friedrich, G., Soriano, P., **1991**. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. 5, 1513-1523.

Gossler, A., Joyner, A.L., Rossant, J., Skarnes, W.C., **1989**. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244, 463-465.

Habets, G.G., Scholtes, E.H., Zuydgeest, D., van der Kammen, R.A., Stam, J.C., Berns, A., Collard, J.G., **1994**. Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. Cell 77, 537-549.

Herberg, L., Coleman, D.L., **1977**. Laboratory animals exhibiting obesity and diabetes syndromes. Metabolism 26, 59-99.

Hiraoka, Y., Ogawa, M., Sakai, Y., Taniguchi, K., Fujii, T., Umezawa, A., Hata, J., Aiso, S., **1998**. Isolation and expression of a human SRY-related cDNA hSOX20. Biochim. Biophys. Acta 1396, 132-137.

Hicks, G.G., Shi, E.G., Li, X.M., Li, C.H., Pawlak, M., Ruley, H.E., **1997**. Functional genomics in mice by tagged sequence

Hicks, G.G., Shi, E.G., Li, X.M., Li, C.H., Pawlak, M., Ruley, H.E., **1997**. Functional genomics in mice by tagged sequence mutagenesis. Nat. Genet. 16, 338-344.

Holness, C.L., da Silva, R.P., Fawcett, J., Gordon, S., Simmons, D.L., **1993**. Macrosialin, a mouse macrophage-restricted glycoprotein, is a member of the lamp/lgp family. J. Biol. Chem. 268, 9661-9666.

Ito, K., Ishii, N., Miyashita, A., Tominaga, K., Kuriyama, H., Maruyama, H., Shirai, M., Naito, M., Arakawa, M., Kuwano, R., **1999**. Molecular cloning of a novel 130-kDa cytoplasmic protein, ankhzn, containing ankyrin repeats hooked to a zinc finger motif. Biochem. Biophys. Res. Commun. 257, 206-213.

Jaenisch, R., **1988**. Transgenic animals. Science 240, 1468-1474.

Jiang, Z., Shih, D.M., Xia, Y.R., Lusis, A.J., de Beer, F.C., de Villiers, W.J., van der Westhuyzen, D.R., de Beer, M.C., **1998**. Structure, organization, and chromosomal mapping of the gene encoding macrosialin, a macrophage-restricted protein. Genomics 50, 199-205.

Jones, E., Quinn, C.M., See, C.G., Montgomery, D.S., Ford, M.J., Kolble, K., Gordon, S., Greaves, D.R., **1998**. The linked human elongation initiation factor 4A1 (EIF4A1) and CD68 genes map to chromosome 17p13. Genomics 53, 248-250.

Joseph, D.R., Adamson, M.C., Kozak, C.A., **1991**. Genetic mapping of the gene for androgen-binding protein/sex hormone-binding globulin to mouse chromosome 11. Cytogenet. Cell Genet. 56, 122-124.

Joseph, D.R., **1998**. The rat androgen-binding protein (ABP/SHBG) gene contains triplet repeats similar to unstable triplets: evidence that the ABP/SHBG and the fragile X-related 2 genes overlap. Steroids 63, 2-4.

Kerr, W.G., Heller, M., Herzenberg, LA., **1996**. Analysis of lipopolysaccharide-response genes in B-lineage cells demonstrates that they can have differentiation stage-restricted

lipopolysaccharide-response genes in B-lineage cells demonstrates that they can have differentiation stage-restricted expression and contain SH2 domains. Proc. Natl. Acad. Sci. USA 93, 3947-3952.

King, D.P., Zhao, Y., Sangoram, A.M., Wilsbacher, L.D., Tanaka, M., Antoch, M.P., Steeves, T.D., Vitaterna, M.H., Kornhauser, J.M., Lowrey, P.L., Turek, F.W., Takahashi, J.S., **1997**, Positional cloning of the mouse circadian clock gene. Cell 89, 641-653.

Korn, R., Schoor, M., Neuhaus, H., Henseling, U., Soininen, R., Zachgo, J., Gossler, A., **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.

Li, A.C., Guidez, F.R., Collier, J.G., Glass, K., **1998**. The macrosialin promoter directs high levels of transcriptional activity in macrophages dependent on combinatorial interactions between PU.1 and c-Jun. J. Biol. Chem. 273, 5389-5399.

Lyon, M.F., Rastan, S., Brown, S.D.M., **1996**. Genetic variants and strain of the laboratory mouse. 3rd edition. Oxford Univ. Press.

Mao, M., Fu, G., Wu, J.S., Zhang, Q.H., Zhou, J., Kan, L.X., Huang, Q.H., He, K-L., Gu, B. W., Han, Z.G., Shen, Y., Gu, J., Yu, Y.P., Xu,S.H., Wang, Y.X., Chen, S.J., Chen, Z., **1998**. Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proc. Natl. Acad. Sci. USA 95, 8175-8180.

Maniatis, T., Fritsch, E.F., Sambrook, J., **1989**. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Martin, G.R., **1981**. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA 78, 7634-7638.

McLauchlan, J., Gaffney, D., Whitton, J.L., Clements, J.B., **1985**. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. 13, 1347-1368.

Meyer, J., Wirth, J., Held, M., Schempp, W., Scherer, G., **1996**. SOX20, a new member of the SOX gene family, is located on chromosome 17p13. Cytogenet. Cell Genet. 72, 246-249.

Miyashita, A., Shimizu, N., Endo, N., Hanyuu, T., Ishii, N., Ito, K., Itoh, Y., Shirai, M., Nakajima, T., Odani, S., Kuwano, R., **1999**. Five different genes, Eif4a1, Cd68, Supl15h, Sox15 and Fxr2h, are clustered in a 40 kb region of mouse chromosome 11. Gene 237, 53-60.

Nielsen, P.J., McMaster, G.K., Trachsel, H., **1985**. Cloning of eukaryotic protein synthesis initiation factor genes: isolation and characterization of cDNA clones encoding factor eIF-4A. Nucleic Acids Res. 13, 6867-6880.

Nielsen, P.J., Trachsel, H., **1988**. The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed. EMBO J. 7, 2097-2105.

Nielsen, P.J., Rochelle, J.M., Seldin, M.F., **1993**. The functional genes for protein synthesis initiation factor 4AI and 4AII map to mouse chromosomes 11 and 16. Mamm. Genome 4, 185-186.

Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S., Yamamura, K., **1993**. An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. J. Biochem. 113, 343-349.

O'Brien, S. J., Menotti-Raymond, M., Murphy, W. J., Nash, W. G., Wienberg, J., Stanyon, R., Copeland, N. G., Jenkins, N. A., Womack, J. E., Marshall, Graves, J. A., **1999**. The promise of comparative genomics in mammals. Science 286, 458-462, 479-481.

Reddy, N.S., Roth, W.W., Bragg, P.W., Wahba, A.J., **1988**. Isolation and mapping of a gene for protein synthesis initiation factor 4A and its expression during differentiation

and its expression during differentiation of murine erythroleukemia cells. Gene 70, 231-243.

Shirai, M., Miyashita, A., Ishii, N., Itoh, Y., Satokata, I., Watanabe, Y.G., Kuwano, R., **1996**. A gene trap strategy for identifying the gene expressed in the embryonic nervous system. Zoolog. Sci. 13, 277-283.

Siomi, M.C., Siomi, H., Sauer, W.H., Srinivasan, S., Nussbaum, R.L., Dreyfuss, G., **1995**. FXR1, an autosomal homolog of the fragile X mental retardation gene. EMBO J. 14, 2401-2408.

Skarnes, W.C., Auerbach, B.A., Joyner, A.L., **1992**. A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev. 6, 903-918.

Skarnes, W.C., Moss, J.E., Hurtley, S.M., Beddington, R.S., **1995**. Capturing genes encoding membrane and secreted proteins important for mouse development. Proc. Natl. Acad. Sci. USA 92, 6592-6596.

Strachan, T., Read, A.P., **1996**. Human molecular genetics. pp421-448. BISO Sci. Publ. Ltd.

Su, H., Wang, X., Bradley, A., **2000**. Nested chromosomal deletions induced with retroviral vectors in mice. Nat. Genet. 24, 92-95.

Svaren, J., Apel, E.D., Simburger, K.S., Jenkins, N.A., Gilbert, D.J., Copeland, N.A., Milbrand, J., **1997**. The Nab2 and Stat6 genes share a common transcription termination region. Genomics 41, 33-39.

Takahashi, J.S., Pinto, L.H., Vitaterna, M.H., **1994**. Forward and reverse genetics approaches to behavior in the mouse. Science 264, 1724-1733.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., Higashinakagawa, T., **1995**. Gene trap

a novel mouse gene, jumonji, required for neural tube formation. Genes Dev. 9, 1211-1222.

van de Wetering, M., Clevers, H., **1993**. Sox 15, a novel member of the murine Sox family of HMG box transcription factors. Nucleic. Acid Res. 21, 1669.

Vitaterna, M.H., King, D.P., Chang, A.M., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., Takahashi, J.S., **1994**. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science 264, 719-725.

Voss, A.K., Thomas, T., Gruss, P., **1998**. Efficiency assessment of the gene trap approach. Dev. Dyn. 212, 171-180.

Vujic, M., Rajic, T., Goodfellow, P.N., Stevanovic, M., **1998**. cDNA characterization and high resolution mapping of the human SOX20 gene. Mamm. Genome 9, 1059-1061.

Ware, F.E., Lehrman, M.A., **1996**. Expression cloning of a novel suppresser of the Lec15 and Lec35 glycosylation mutations of Chinese hamster ovary cells. J. Biol. Chem. 271, 13935-13938.

Ware, F.E., Lehrman, M.A., **1998**. Addition and corrections to expression cloning of a novel suppressor of the Lec15 and Lec35 glycosylation mutations of Chinese hamster ovary cells. J. Biol. Chem. 273, 13366.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D.P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A., Ang. Siew-Lan., **1995**. A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. Genetics 139, 889-899.

Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., Soriano, P., **1997**. Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proc. Natl. Acad. Sci. USA 94, 3789-3794.

Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg,

hematopoietic cells. Proc. Natl. Acad. Sci. USA 94, 3789-3794.

Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C., Sands, A. T., **1998**. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392, 608-611.

Zhang, Y., O'Connor, J.P., Siomi, M.C., Srinivasan, S., Dutra, A., Nussbaum, R.L., Dreyfuss, G., **1995**. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. EMBO J. 14, 5358-5366.
Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M., **1994**. Positional cloning of the mouse obese gene and its human homolog. Nature 372, 425-432.

Zheng, B., Sage, M., Sheppeard, E.A., Jurecic, V., Bradley, A., **2000**. Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. Mol. Cell Biol. 20, 648-655.