Takano, K., Noguchi, Y., Yamamoto, Y., Suzuki, O., Matsuda, J., and Ogura, A	unguiculatus) embryos cryopreserved by vitrification				
Miki, H., Inoue, K., Kohda, T., Honda, A., Ogonuki, N., Yuzuriha, M., Mise, N., Matsui, Y., Baba, T., Abe, K., Ishino, F., and Ogura, A	Birth of Mice Produced by Germ Cell Nuclear Transfer	Genesis	41	81-86	2005
Kanatsu-Shinohara, M., Miki, H., Inoue, K., Ogonuki, N., Toyokuni, S., Ogura, A., and Shinohara, T	Long-term culture of mouse male germline stem cells under serum- or feeder-free conditions	Biol. Reprod	in press		2005
Abe K, Hazama M, Katoh H, Yamamura K, Suzuki M.	Establishment of an efficient BAC transgenesis protocol and its application to functional characterization of the mouse brachyury locus.	Exp Anim	53	311-320	2004
Oyanagi, M., Kato, A., Yamada, Y. K., and Sato, N. L.	Detection of MHV-RNAs in mouse intestines and in filter dust in mouse room ventilation duct by modified RT-nested PCR	Experimental Animals	53	37-41	2004

Available online at www.sciencedirect.com
SCIENCE DIRECT®
General and Comparative Endocrinology 136 (2004) 406–410
GENERAL AND COMPARATIVE ENDOCRINOLOGY
www.elsevier.com/locate/ygcen

Communication in Genomics and Proteomics

Sequence analysis of cDNA encoding follicle-stimulating hormone and luteinizing hormone β-subunits in the Mongolian gerbil (*Meriones unguiculatus*)

Minako Koura, Hiroko Handa, Yoko Noguchi, Kaoru Takano, Yoshie Yamamoto, Junichiro Matsuda, and Osamu Suzuki*

Department of Veterinary Science, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

Received 12 November 2003; revised 23 January 2004; accepted 26 January 2004

Abstract

To examine the molecular basis for efficient superovulation in the Mongolian gerbil, the cDNA sequences of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) β-subunits were determined and compared with those of other mammals. FSHβ and LHβ cDNAs were 1637 and 507 bp long, respectively, from the 5′-end to putative polyA sites. The deduced sequences of the FSHβ and LHβ precursor proteins were 129 and 141 amino acids in length, respectively. The amino acid sequences of both Mongolian gerbil hormone subunits showed overall similarity to those of other rodents, confirming that the combination of equine chorionic gonadotropin (eCG) and human chorionic gonadotropin (hCG) should be effective for induction of superovulation in Mongolian gerbils, as in mice and rats. However, the use of hCG might need to be re-evaluated owing to its low homology to rodent LH.

Keywords: Mongolian gerbil; FSHβ; LHβ; cDNA; Phylogeny

1. Introduction

The Mongolian gerbil (*Meriones unguiculatus*) is frequently used as a rodent model in the study of epilepsy (Loskota et al., 1974) and gastric infection with *Helicobacter pylori* (Hirayama et al., 1996). Embryo manipulation techniques, such as cryopreservation of embryos/gametes and transgenic technologies, need to be employed to improve the usability of laboratory animals, but these techniques are not fully applicable to gerbils yet. For these applications, many embryos/oocytes are required, making induction of superovulation an essential technique. In general, superovulation is induced by injection of gonadotropins. In particular, the combination of equine chorionic gonadotropin (eCG) and human chorionic gonadotropin (hCG) is widely used for induction of superovulation in many species. However, this method is not efficient in all species. In rabbits, multiple injections of FSH are more effective than a single injection of eCG (Hirabayashi et al., 2000). In guinea pigs, human menopausal gonadotropin (hMG) is a better inducer of superovulation than are the chorionic gonadotropins (Suzuki et al., 2003). In Mongolian gerbils, the combination of eCG plus hCG is effective for superovulation, but a problem arises in that a considerable number of oocytes remain trapped within the corpora lutea (Fischer and Fisher, 1975). A careful choice of gonadotropins is still needed for practical superovulation in this animal.

In this paper, we describe the cDNA sequences of the FSH and LHβ-subunits in order to develop a molecular basis for efficient superovulation in the Mongolian gerbil. Phylogenetic analyses of the deduced proteins were also performed. The possibility of using this sequence information as a criterion for selection of gonadotropins for superovulation of Mongolian gerbils is discussed.

* Corresponding author. Fax: +81-3-5285-1179.
E-mail address: osuzuki@nih.go.jp (O. Suzuki).

0016-6480/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygcen.2004.01.012

2. Materials and methods

2.1. RNA extraction, RT-PCR, and sequencing

The full sequences of FSHβ and LHβ were determined by a combination of both 5′- and 3′-rapid amplification of cDNA ends (RACE) with the SMART RACE cDNA amplification kit (BD Biosciences Clontech, Palo Alto, CA), followed by direct sequencing of the RACE products. In brief, two cDNA libraries for RACE reactions were prepared from total RNA extracted from the pituitary glands of male Mongolian gerbils. RACE reactions were performed using these libraries, a DNA polymerase mixture (HotStarTaq Master Mix, QIAGEN GmbH, Hilden, Germany), anchor primers (universal primer mix included in the kit) and gene-specific primers under the following thermal conditions: one cycle at 94 °C for 15 min (denature and enzyme activation), and 40 cycles of 94 °C for 2 s, 68 °C for 5 min carried out with a Hybaid PCR Express thermal cycler (Thermo Bioanalysis Japan KK, Tokyo, Japan). To determine the FSHβ sequence, nested PCR amplifications for primer walking were carried out with primary PCR products (diluted 1:50), a DNA polymerase mixture, nested adaptor primers (included in the kit) and gene-specific primers under the following thermal conditions: one cycle of 94 °C for 15 min; and 30 cycles of 94 °C for 2 s, 68 °C for 5 min. Gene-specific primers were as follows (see Fig. 1 for positions): for the FSHβ cDNA sequence, 5′-RACE was first conducted with primer 5Fp (TTC CTT CAT TTC ACT GAA GGA GCA GTA). Then, primer walking was performed in 3′-RACE with a total of four primers: 3Fp1 (GAA GGA AGA GTG TCG TTT CTG CGT AAG), 3Fp2 (GGA CCC AGC CAG ACC CAA TAC CCA GAA), 3Fp3 (CCA CCG CTC ATC CCT CTA TCC ATG CTG), and 3Fp4 (GCT CAG AAT TTC CAA GCT ATT GAC ACC). For the LHβ sequence, 3′-RACE was first conducted with primer 3Lp (GGC TGC TGC TGA GCC CAA GTG TGG TGT), followed by 5′-RACE with primer 5Lp (GCT GAG GGC CAC AGG GAA GGA AAC CAT). Neither 5Fp nor 3Lp was perfectly identical to the target cDNA sequence since they were designed on the basis of mouse sequences (NP_032071 and NM_008497 from GenBank for FSHβ and LHβ, respectively) using Primer3 software (Rozen and Skaletsky, 1998). Other primers matched perfectly to targets because they were designed using Primer3 with already-determined partial sequences. The obtained RACE products were gel-purified and then directly sequenced using a DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Piscataway, NJ) with a DNA sequencer (RISA384, Shimadzu Biotech, Kyoto, Japan). Full sequences were obtained by combining overlapping sequences of the 5′- and 3′-RACE products. Both 5′- and 3′-ends of the cDNA sequences were judged by the appearance of SMART II Oligo and putative polyA-sequences, respectively, in the determined sequences.

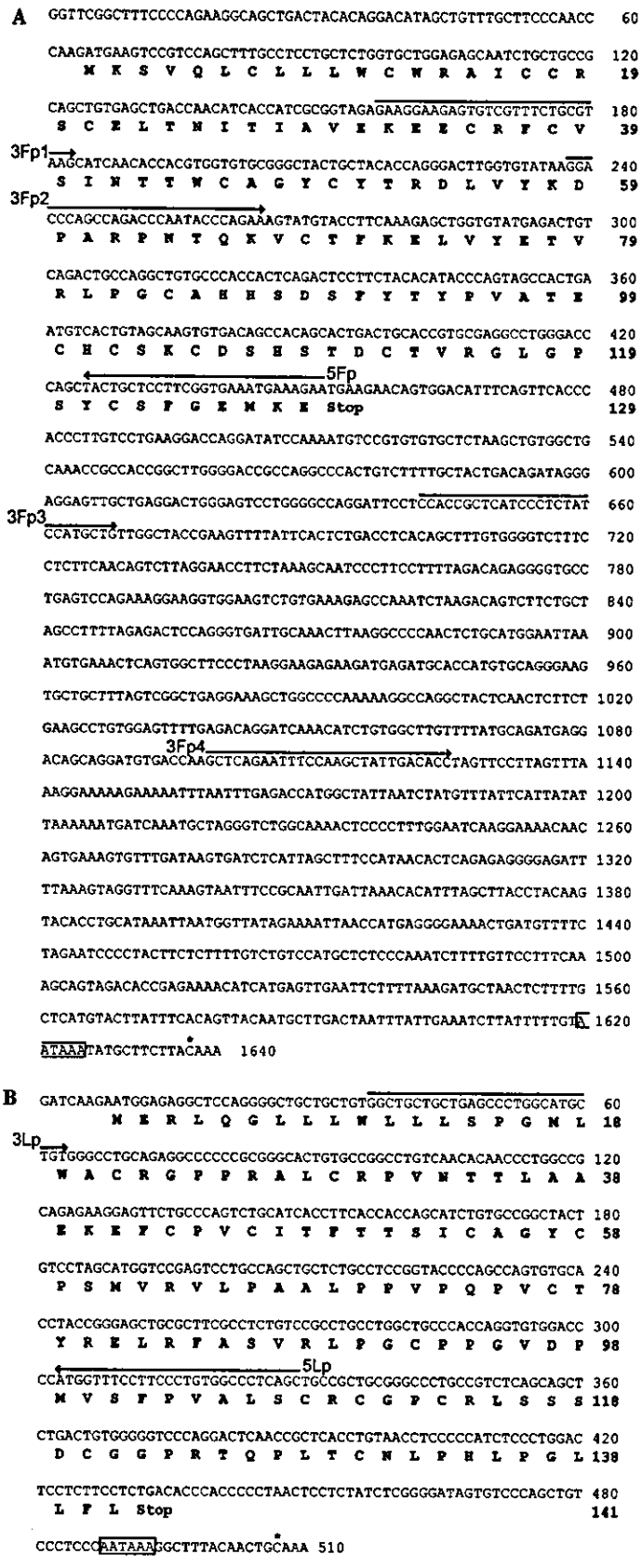
2.2. Molecular phylogenetic analysis

Additional mammalian FSHβ and LHβ sequences were retrieved from GenBank as follows (Accession Nos. are shown in parentheses): mouse (NP_032071 and NM_008497), rat (P18427 and J00749), pig (P01228 and D00579), cattle (NP_776485 and M10077), sheep (P01227 and X52488), horse (P01226 and S41704), human (P01225 and NM_000894), and Djungarian hamster (*Phodopus sungorus*, AF106914 and AF106915; Bernard et al., 2000). The amino acid sequence of hCG was also retrieved from GenBank (J00117). Amino acid sequences of these species, including the Mongolian gerbil, were aligned using Clustal X, version 1.81 (Thompson et al., 1997). Neighbor-joining (NJ) trees that contained FSHβ and LHβ protein sequences of eight species (the Djungarian hamster was omitted because both AF106914 and AF106915 were still partial sequences) were constructed using the MEGA, version 2.1, program package (Kumar et al., 2001), based on the number of amino acid substitutions with 5000 bootstrap iterations.

3. Results and discussion

The full-length of FSHβ and LHβ cDNA sequences are shown in Fig. 1. Both sequences have been deposited in GenBank with the Accession Nos. AY376457 and AY369077, respectively. The FSHβ and LHβ cDNAs were 1637 and 507 bp long, respectively, from the 5′-end to putative polyA sites. Both sequences contained putative polyA signal sequences (AATAAA) starting at position 1620 and 488, respectively, for FSHβ and LHβ. The deduced amino acid sequences of the FSHβ and LHβ precursor proteins were 129 and 141 in length, respectively. Both amino acid sequences contained the unique β-subunit sequence, CAGY, which is thought to be a key structure for binding to common α-subunits (Gharib et al., 1990). Positions of half-cysteine residues and putative N-glycosylation sites were well conserved in comparison with those of other mammals. The Mongolian gerbil FSHβ cDNA shares the common features of a long 3′-UTR with other mammals.

Alignments of the amino acid sequences indicate overall similarity of both Mongolian gerbil subunits to those of other rodents (Fig. 2). Phylogenetic analysis of each protein also supports the similarity of these two hormones among rodents (Fig. 3). Interestingly, there are two variations of the N-terminus of the FSHβ precursor protein in rodents: an MM-type and an M-type. Mice and rats have the former type of FSHβ precursor protein,


A
3Fp1
3Fp2
3Fp3
3Fp4
5Fp
B
3Lp
5Lp

```
A M. Gerbil  -MKSVQLCLL LWCWRAICCR SCELTNITIA VEKEECRFCV SINTTWCAGY CYTRDLVYKD PARPNTQKVC TFKELVYETV [ 80]
  D. Hamster -..LI...I. F.........Q G.......... ..........I ..........  .......... ........I. .........I [ 80]
  Mouse      M..LI...I. F.........H .........S .........I ..........  .......... .......... .......... [ 80]
  Rat        M...I...I. ...L..V..H .........S .........I .......E..  .......... .......... .........I [ 80]
  Cattle     -.....F.F. FC........ .........T ......G..I ..........  ........R. .....I..T. .......... [ 80]
  Sheep      -.....F.F. FC........ .........T ......S..I ..........  .......... .....I..A. .......... [ 80]
  Pig        -...L.F.F. FC..K....N .........T ......N..I ..........  .......... .....I..T. .......... [ 80]
  Horse      -.....F.F. FC..K.V..N .......... ......G..I ..........  .......... .....I..T. .......... [ 80]
  Human      -..TL.FFF. FC..K....N .......... I.........I ..........  .......... ....KI..T. .......... [ 80]

  M. Gerbil  RLPGCAHHSD SFYTYPVATE CHCSKCDSHS TDCTVRGLGP SYCSFGEMKE [130]
  D. Hamster .......... .......... ...G....D. .......... .......*   [127]
  Mouse      ......R... .L........ ...G....D. .......... .....S....  [130]
  Rat        ......R... .L........ ...G....D. .......... ..........  [130]
  Cattle     KV......A. .L........ ........D. .......... .....R.I..  [130]
  Sheep      KV......A. .L........ ...G...RD. .......... .....SDIR.  [130]
  Pig        KV......A. .L........ ...G....D. .......... .....S....  [130]
  Horse      KV......A. .L.......A ...G..N.D. .......... ......D...  [130]
  Human      .V......A. .L.......Q ...G....D. .......... ..........  [130]

B M. Gerbil  MERLQGLLLW LLLSPGMLWA CRGPPRALCR PVNTTLAAEK EFCPVCITFT TSICAGYCPS MVRVLPAALP PVPQPVCTYR [ 80]
  D. Hamster .......... .....SVV.. S...L.P... ...A.....N .A.......S .......... .......... .........H [ 80]
  Mouse      .......... .....SVV.. S...L.P... ...A.....N .......... .......... .......... .......... [ 80]
  Rat        .......... .....SVV.. S...L.P... ...A.....N .......... .......... .......... .......... [ 80]
  Cattle     ..MF...... ...GVAGV.. S...L.P..Q .I.A...... .A........ .......... .K....VI.. .M..R....H [ 80]
  Sheep      ..M....... ...GVAGV.. S...L.P..Q .I.A...... .A........ ........L. .KQ...VI.. .M..R....H [ 80]
  Pig        ..M....... ....VAGV.. S...L.P... .I.A.....N .A........ .......... .......... .......... [ 80]
  Horse      ..T....... M...V.GV.. S...L.P... .I.A...... .A..I..... .......... ....M..... AI........ [ 80]
  Human      ..M......L ....M.GA.. S.E.L.PW.H .I.AI..V.. .G......VN .T.......T .M...Q.V.. .L..V..... [ 80]
  hCG        ..MF.....L ....M.GT.. SKE.L.PR.. .I.A...V.. .G......VN .T.......T .T...QGV.. AL..V..N.. [ 80]

  M. Gerbil  ELRFASVRLP GCPPGVDPMV SFPVALSCRC GPCRLSSSDC GGPRTQPLTC NLPH------ LPGLLFL--- ---------- [160]
  D. Hamster ..H....... .......... .......... ......T... .......M*                                    [128]
  Mouse      ..A....... ........I. .......... .......... .......MA. D...------ .....L.--- ---------- [160]
  Rat        .......... ........I. .......... .......... .......M.. D...------ .....LF--- ---------- [160]
  Cattle     .......... .......... ........H. .......T.. ........A. DH.P------ ..DI...--- ---------- [160]
  Sheep      .......... .......... ........H. .......T.. ........A. DH.P------ ..DI...--- ---------- [160]
  Pig        ..S...I... ........T. ........H. .......... ....A...A. DR.L------ .......--- ---------- [160]
  Horse      ......I... .......... ........H. ...QIKTT.. .VF.D...A. APQASSSSKD P.SQPLTSTS TPTPGASRRS [160]
  Human      DV..E.I... ...R....V. .......... ....R.T... ...KDH.... DH.Q------ .S.....--- ---------- [160]
  hCG        DV..E.I... ...R..N.V. .YA.....Q. AL..R.TT.. ...KDH.... DD.RFQDSSS SKAPPPS..S PSR.PGPSDT [160]

  M. Gerbil  --------- [169]
  D. Hamster
  Mouse      --------- [169]
  Rat        --------- [169]
  Cattle     --------- [169]
  Sheep      --------- [169]
  Pig        --------- [169]
  Horse      SHPLPIKTS [169]
  Human      --------- [169]
  hCG        PILPQ---- [169]
```

Fig. 2. Alignments of amino acid sequences of FSHβ (A) and LHβ (B) subunits in nine mammalian species. hCG is also included in (B). A dot indicates an amino acid residue identical to the corresponding site in the Mongolian gerbil sequence (M. Gerbil, top row). Dashes in the sequences indicate gaps inserted by Clustal X software. Asterisks (*) indicate ends of currently obtained partial sequences of the Djungarian hamster (D. Hamster).

whereas Mongolian gerbils have the latter type, which is similar to Djungarian hamsters. So far, FSHβ and LHβ cDNA sequences have been reported for only a few rodent species. More research is needed to define more precisely the rodent-specific features of these proteins.

Ovulation in Mongolian gerbils can be induced by the combination of eCG and hCG injections, as is done for superovulation of mice and rats. The similarity of the response in gerbils, rats, and mice may be accounted for in part by the similarity of the β-subunits of the gonadotropins. However, hCG should ideally be replaced with native or a more similar LH to solve the problem of the considerable number of oocytes that are trapped within the corpora lutea after superovulation of Mon-

Fig. 1. Sequences of FSHβ (A) and LHβ (B) subunit cDNAs. Deduced amino acid sequences of both sequences are also shown in bold face. Putative polyadenylation signal sequences are boxed and putative polyadenylation sites are indicated by asterisks (*). Arrows indicate primer positions (5Fp, 3Fp1, 3Fp2, 3Fp3, 3Fp4, 5Lp, and 3Lp).

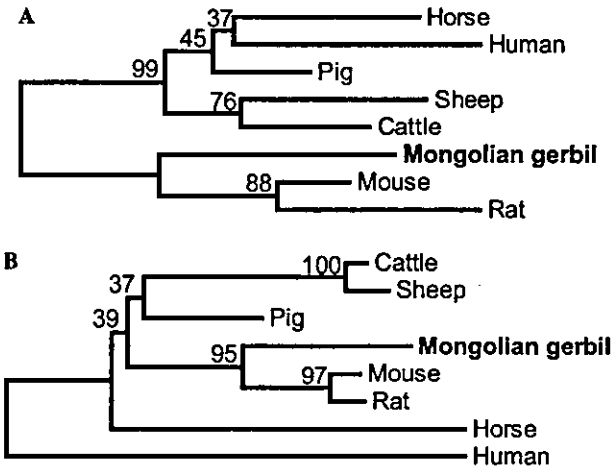


Fig. 3. Phylogenetic trees constructed from inferred amino acid sequences of FSHβ (A) and LHβ (B) subunits of eight mammalian species. The figures at the nodes represent bootstrap values based on 5000 iterations.

golian gerbils (Fischer and Fisher, 1975), since the species-specificities of LHs are reportedly higher than those of FSHs (Ishii and Kubokawa, 1984). In addition, the efficacy and species-spectrum of hCG as an ovulation inducer should be re-evaluated, since the amino acid sequence of the hCG β-subunit shows low homology to other luteinizing hormones (Fig. 2). Receptor homology should also be considered when selecting hormones for superovulation. We previously reported that hMG was a better inducer of follicular development in the guinea pig than eCG (Suzuki et al., 2003). We speculated that this difference might be attributable mainly to the homology of the FSH-receptors, rather than that of the ligands (FSH). The guinea pig FSH receptor is more homologous to the human receptor than to receptors of other mammals, whereas both CGα (Suzuki et al., 2002) and FSHβ (unpublished data) subunits have higher homologies to their pig counterparts than to humans. For the next step in the analysis of the selection of effective hormones for induction of superovulation in the Mongolian gerbil, we should compare the FSH and LH receptor sequences to those of other mammals.

In summary, we determined the cDNA sequences of FSHβ and LHβ subunits in the Mongolian gerbil. This information should provide a molecular basis for understanding hormonal control of ovulation, as well as improving superovulation technologies in the Mongolian gerbil.

References

Bernard, D.J., Merzlyak, I.Y., Horton, T.H., Turek, F.W., 2000. Differential regulation of pituitary gonadotropin subunit messenger ribonucleic acid levels in photostimulated Siberian hamsters. Biol. Reprod. 62 (1), 155–161.

Fischer, T.V., Fisher, D.L., 1975. Effect of gonadotropins on ovulation and ovarian histology in the immature Mongolian gerbil. Am. J. Anat. 142 (3), 391–396.

Gharib, S.D., Wierman, M.E., Shupnik, M.A., Chin, W.W., 1990. Molecular biology of the pituitary gonadotropins. Endocr. Rev. 11 (1), 177–199.

Hirabayashi, M., Hirao, M., Hochi, S., Kimura, K., Hirasawa, K., Ueda, M., 2000. Effects of superovulatory protocols on the recovery and subsequent developmental potential of rabbit zygotes (In Japanese). J. Exp. Anim. Tech. 35 (1), 23–28.

Hirayama, F., Takagi, S., Yokoyama, Y., Iwao, E., Ikeda, Y., 1996. Establishment of gastric Helicobacter pylori infection in Mongolian gerbils. J. Gastroenterol. 31 (Suppl. 9), 24–28.

Ishii, S., Kubokawa, K., 1984. Avian gonadotropin receptors: a comparative view. J. Exp. Zool. 232 (3), 431–434.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17 (12), 1244–1245.

Loskota, W.J., Lomax, P., Rich, S.T., 1974. The gerbil as a model for the study of the epilepsies. Seizure patterns and ontogenesis. Epilepsia 15 (1), 109–119.

Rozen, S., Skaletsky, H.J., 1998. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Suzuki, O., Mochida, K., Yamamoto, Y., Noguchi, Y., Takano, K., Matsuda, J., Ogura, A., 2002. Comparison of glycoprotein hormone α-subunits of laboratory animals. Mol. Reprod. Dev. 62 (3), 335–342.

Suzuki, O., Koura, M., Noguchi, Y., Takano, K., Yamamoto, Y., Matsuda, J., 2003. Optimization of superovulation induction by human menopausal gonadotropin in guinea pigs based on follicular waves and FSH-receptor homologies. Mol. Reprod. Dev. 64 (2), 219–225.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25 (24), 4876–4882.

Exp. Anim. 53(2), 103–111, 2004

Chromosomal Mapping and Zygosity Check of Transgenes Based on Flanking Genome Sequences Determined by Genomic Walking

Akira NOGUCHI, Naho TAKEKAWA, Thorbjorg EINARSDOTTIR, Minako KOURA, Yoko NOGUCHI, Kaoru TAKANO, Yoshie YAMAMOTO, Junichiro MATSUDA, and Osamu SUZUKI

Department of Veterinary Science, National Institute of Infectious Diseases, 1–23–1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

Abstract: *Transgenes can affect transgenic mice via transgene expression or via the so-called positional effect. DNA sequences can be localized in chromosomes using recently established mouse genomic databases. In this study, we describe a chromosomal mapping method that uses the genomic walking technique to analyze genomic sequences that flank transgenes, in combination with mouse genome database searches. Genomic DNA was collected from two transgenic mouse lines harboring pCAGGS-based transgenes, and adaptor-ligated, enzyme restricted genomic libraries for each mouse line were constructed. Flanking sequences were determined by sequencing amplicons obtained by PCR amplification of genomic libraries with transgene-specific and adaptor primers. The insertion positions of the transgenes were located by BLAST searches of the Ensembl genome database using the flanking sequences of the transgenes, and the transgenes of the two transgenic mouse lines were mapped onto chromosomes 11 and 3. In addition, flanking sequence information was used to construct flanking primers for a zygosity check. The zygosity (homozygous transgenic, hemizygous transgenic and non-transgenic) of animals could be identified by differential band formation in PCR analyses with the flanking primers. These methods should prove useful for genetic quality control of transgenic animals, even though the mode of transgene integration and the specificity of flanking sequences needs to be taken into account.*

Key words: *chromosomal mapping, flanking primers, genomic walking, zygosity check*

Introduction

Transgenes can exert effects in transgenic animals by transgene expression and by the so-called positional effect. Transgene expression often varies among multiple lines that have been derived from different founders. This variation may result from the position of the transgene integration site as well as the copy

(Received 22 August 2003 / Accepted 4 November 2003)

Address corresponding: O. Suzuki, Department of Veterinary Science, National Institute of Infectious Diseases, 1–23–1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

number of the transgene. A transgene integrated into the X chromosome would exhibit sex-linked inheritance. Side effects, such as unexpected gene knockouts, may be caused by the disruption of an endogenous gene due to transgene insertion [e.g., 11, 14]. Thus, information on transgene position is important for analyses of transgenic animals.

Genes are typically localized by *in situ* hybridization methods, such as FISH [e.g., 3, 6]. In recent years, mouse genome databases have expanded rapidly and offer highly integrated records of genomic sequences, chromosome maps, and related information. These databases can be used to map transgenes on mouse chromosomes with great accuracy if the flanking sequences of the transgene can be determined.

Here, we describe a rapid, simple method using genomic walking [12] to determine genomic sequences that flank pCAGGS-based transgenes (Fig. 1A) [8]. The strategy of the genomic walking used in this study is outlined in Fig. 1B. Genomic sequences flanking the transgenes were determined by two consecutive PCR amplifications with multiple enzyme-digested, adaptor-ligated genomic libraries of transgenic animals (Fig. 1B). Determination of the nucleotide sequences flanking the transgene enables the transgene to be mapped in chromosomes by searching mouse genome databases. In addition, the sequence information can be used to construct locus-specific zygosity check systems with flanking primer methods (Fig. 1C). However, our experience suggests that modes of transgene integration should be considered as well as the flanking genomic sequences.

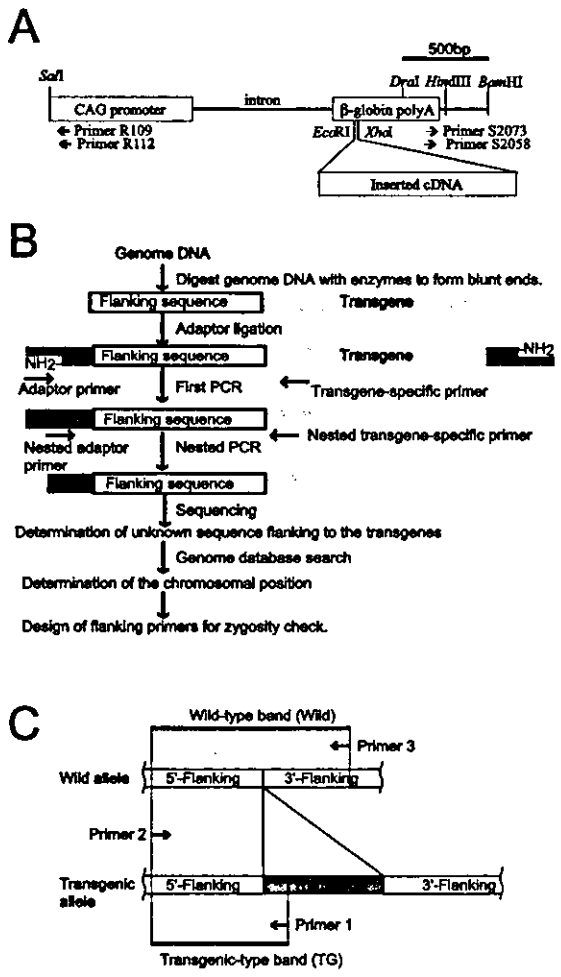


Fig. 1. Transgene construct (A) and strategies for chromosome mapping (B) and zygosity check (C) of the transgene. Transgenes were excised from pCAGGS-plasmids that contained mouse cDNA. Two sets of primers, R109 and R112, and S2058 and S2072, were used as transgene-specific primers for genomic walking to 5'-flanking and 3'-flanking regions, respectively. Figure 1B summarizes the procedure of the genomic walking in this study (modified from a chart in a manual of the Universal GenomeWalker kit). In brief, six enzyme-digested, adaptor-ligated genomic libraries were constructed from transgenic mice. Fragments flanking the transgenes were amplified by two consecutive PCR amplifications with adaptor and transgene-specific primers. Genomic sequences flanking the transgenes were determined by sequencing the fragments. Integration sites of the transgenes were determined by genome database searches. Genomic sequences flanking the other end of the transgene were determined by genomic walking from the 5'-flanking region to the 3'-flanking region with genomic libraries from non-transgenic mice. Flanking primers for a zygosity check were designed on the basis of the genomic sequences at both ends of the integration sites (See text for details). Figure 1C summarizes the strategy for zygosity check. Zygosity was judged by the differential bands produced with flanking primers, which were designed to distinguish between wild-type (Wild) and transgenic (TG) alleles by amplicon length.

Materials and Methods

Transgenic animals

This study used two lines of transgenic mice, 4c30 and CK35, at the eighth and sixth generations, respectively, and their non-transgenic parent strain, C57BL/6CrSlc (B6). The transgenic lines were produced in

our laboratory by zygote microinjection of transgene constructs based on a pCAGGS plasmid (Fig. 1A) [8]. Line 4c30 received a plasmid that contained mouse α2,3-sialyltransferase type II [5] cDNA, whereas line CK35 received a plasmid that contained human acid β-galactosidase [9] cDNA. Constructs containing a cDNA insert at the multi-cloning site (MCS) were excised from *Sal*I to *Bam*HI (approximately 3.5 kb long) and from *Sal*I to *Hin*dIII (approximately 4.5 kb long) from the plasmids and used for production of 4c30 and CK35, respectively. Southern blot analyses confirmed integration of the transgenes in the two lines. Animal experiments were performed according to the Guides for Animal Experiments Performed at NIID.

Determination of flanking sequences using genomic walking for chromosomal mapping of the transgene

The genomic sequences that flanked the transgenes were determined by genomic walking using the Universal GenomeWalker™ kit (BD Bioscience Clontech, Palo Alto, CA) with a slight modification. Adaptor-ligated genomic DNA libraries of the transgenic lines were constructed with tail DNA digested with six restriction enzymes: *Dra*I, *Sca*I, *Pvu*II, *Eco*RV, *Ssp*I, and *Stu*I (Takara Bio Inc., Tokyo, Japan). The genomic walk consisted of two PCR amplifications. The primary PCR amplification was performed with an outer transgene-specific primer (R112, 5'-CCA GGC GGG CCA TTT ACC GTA AGT TAT-3') and the outer adaptor primer provided in the kit. The primary PCR mixture was diluted and used as the template for a nested PCR amplification with a nested transgene-specific primer (R109, 5'-GGC GGG CCA TTT ACC GTA AGT TAT GT-3') and the nested adaptor primer provided in the kit. All PCR amplifications were performed with a hot-start DNA polymerase (HotStarTaq; Qiagen K.K., Tokyo, Japan) in a Hybaid PCR Express Thermal Cycler (Thermo Hybaid, Ashford, Middlesex, UK). The hot-start DNA polymerase requires a 15-min incubation at 95°C for activation as the first step of PCR amplification. The primary PCR amplification was performed at 95°C for 15 min, followed by 40 cycles at 94°C for 2 s and 68°C for 5 min. The nested PCR amplification was performed at 95°C for 15 min, followed by 30 cycles at 94°C for 2 s and 68°C for 5 min. The nested PCR products were separated by electrophoresis on a 2% agarose gel in TAE buffer. The nested PCR bands were visualized with ethidium bromide staining and extracted from the gels using a MinElute Gel Extraction kit (QIAGEN). The nucleotide sequences of the PCR bands were determined by direct sequencing using a dye-primer method (Thermo Sequenase Primer Cycle Sequencing kit with 7-deaza GTP; Amersham Biosciences, Piscataway, NJ) with a DSQ-2000L DNA Sequencer (Shimadzu Corp., Kyoto, Japan). Among the nested PCR products, the DNA fragment containing the 5'-flanking sequence of the transgene was determined by identifying a nucleotide sequence that was not found in the transgene and that continued to the 5'-transgene sequence. Other than the adaptor primers provided in the kit, all primers were designed using the Primer3 program [10]. The transgene insertion sites in chromosomes were determined using a BLAST search of the flanking sequences in the Ensembl genome database [2] via the Internet (http://www.ensembl.org).

Genomic walking analyses that proceeded backwards from the tails of the transgene were performed with two primers S2058 (5'-GTA TAT GAA ACA GCC CCC TGC TGT CCA-3') and S2072 (5'-CCC CTG CTG TCC ATT CCT TAT TCC ATA G-3') in combination with adapter primers provided in the kit.

Zygosity check of the transgene by a flanking primer method

For locus-specific zygosity checks, we performed PCR analyses using flanking primers with sequences based on the flanking sequences of the transgenes (Fig. 1C). The flanking sequence around a transgene was determined by performing two consecutive genomic walking experiments. The first walk from the transgene to the 5'-flanking region was performed with genomic libraries of the transgenic line. A second walk from the 5'- to 3'-flanking region was performed with non-transgenic genomic libraries to determine the 3'-flanking sequence of the transgene. Two flanking primers, designated as Primer 2 and Primer 3 in Fig. 1C, were designed with the Primer3 program [10] using 5'- and 3'-flanking sequences at the transgene insertion site of transgenic lines 4c30 and CK35. Primer R109 was used as Primer 1, as shown in Fig. 1C. The zygosity of these transgenic animals was determined using PCR analyses with their tail DNA and the flanking primers under the following thermal conditions: 95°C for 15 min, followed by 30 cycles of 94°C for 2 s, 55°C for 15 s, and 72°C for 1

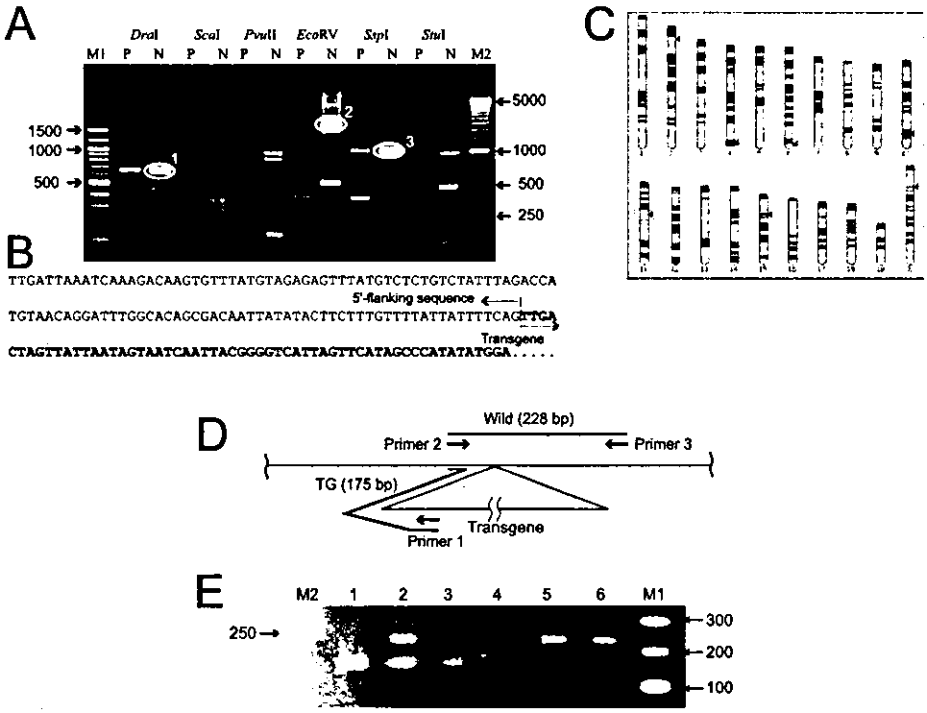


Fig. 2. Chromosomal mapping and zygosity check in transgenic mouse line 4c30. The products of primary (P) and nested (N) PCR amplification with 4c30 genomic libraries were separated by agarose gel electrophoresis (A). Sequence information revealed that bands #1 and #2 in Fig. 2A were derived from the tandem repeat of the transgene. The sequence (B) flanking the transgene was determined from the ~1,000-bp *Stu*I band (band #3 in Fig. 2A). A search of the Ensembl genome database with the flanking sequence suggested that the transgene is located on chromosome 11 (boxed in panel C). Primers were designed for the zygosity check of 4c30 (D). Zygosity of six mice was determined by the differential bands on a gel (E) . Lanes 1 and 3: homozygous transgenic; Lanes 2 and 4: hemizygous; Lanes 5 and 6: non-transgenic; M1: 100-bp DNA ladder; M2: 250-bp DNA ladder.

min. The PCR products were separated by electrophoresis on a 2% agarose gel in TAE buffer, and bands were detected by ethidium bromide staining with ultraviolet illumination. The zygosities of all the animals used for the zygosity check in this study were confirmed in advance using Southern blot analysis, breeding tests, and zygosity-specific symptoms (homozygous-specific cardiac hypertrophy) in 4c30, and the enzyme activity of the transgene products in CK35 tissues.

Results

Mapping and genotyping of mouse line 4c30

Figure 2A shows the electrophoretic patterns of primary and nested PCR amplification products from the first genomic walking experiment. The gel contained three major bands consistent with both primary (P) and nested (N) PCR amplifications: ~650 bp of *Dra*I library, ~1,500 bp of *Eco*RV, and ~1,000 bp of *Ssp*I library. Sequencing analysis revealed that the first two bands were derived from the tandem repeat of the transgene, and that only the last band contained the 5'-flanking sequence (Fig. 2B). A BLAST search with the 5'-flanking sequence revealed that the transgene was integrated into Chromosome 11 (Fig. 2C).

After the second genomic walking experiment, two primers, labeled Primer 2 and Primer 3 in Fig. 1C, were designed so that the primers in combination with Primer R109 (labeled Primer 1 in Fig. 1A) would produce wild-type and transgenic alleles as 228-bp (Wild) and 175-bp (TG) bands, respectively (Fig. 2D). As shown in Fig. 2E, analysis with these primers revealed

a clear genotyping of six line 4c30 mice as homozygous (lanes 1 and 3) with a Wild band only, wild-type (lanes 5 and 6) with a TG band only, and hemizygous (lanes 2 and 4) with both bands. These genotypings matched those determined using the Southern blot analysis perfectly.

Mapping and genotyping of mouse line CK35

More bands were detected in the electrophoretic patterns of nested PCR products from mouse line CK35, as compared with the patterns observed for line 4c30 (Fig. 3A). Of these CK35 bands, only one band of ~700 bp in the *Ssp*I library contained the 5'-flanking genomic sequence (Fig. 3B). A BLAST search of this sequence in the Ensembl genome databases indicated that the transgene was mapped to chromosome 3 (Fig. 3C).

Flanking primers 2 and 3 for the mouse line CK35 were designed from two consecutive genomic walking experiments, as described for mouse line 4c30 (Fig. 3D). The expected Wild (329-bp) and transgene (213-bp) bands are shown as solid lines in Fig. 3D. A zygosity check was performed using PCR amplification with these primers and R109 as Primer 1. The results revealed unexpected bands of about the same length as the anticipated Wild band in homozygous mice (lanes 1 and 2 in Fig. 3E), in which zygosity had been confirmed by Southern blot analysis, yet only single bands were produced when primer pairs for wild-type (Primer 2 and Primer 3) and transgenic (Primer 1 and Primer 2) alleles were used separately (Fig. 3F). Sequencing of the unexpected PCR products revealed that the nucleotide sequence of the transgene at the 3'-flanking region was identical to the nucleotide sequence at the 5'-flanking region. Since Primer 1 could bind to two regions, one additional band, shown as a dotted line in Fig. 3D, was produced when all three primers were simultaneously used for PCR amplification. Based on these facts, new flanking primers were designed so that two TG bands derived from 5'- and 3'-junctions would move to about the same position on a gel and yet be clearly distinct from a Wild band (Fig. 3G). With these new primers, clear genotyping of line CK35 was achieved as shown in Fig. 3H. As expected, two TG bands formed one bold band on a gel (lanes 1 and 2 in Fig. 3H).

Backward genomic walking in mouse line 4c30

Figure 4 shows electrophoretic patterns of nested PCR products obtained by backward genomic walking with B6 and 4c30 strains. Note that a considerable number of amplicons were found even in genomic libraries of non-transgenic mice as well as in 4c30 libraries.

Discussion

Our results with two examples show that the chromosomal mapping and design of flanking primers for a zygosity check can be achieved by the genomic walking technique in combination with a genome database search. While this method is quite powerful, the mode of transgene integration and specificity of flanking sequences need to be taken into account.

Figures 2 and 3 summarize two examples of successful chromosomal mapping using genomic walking. The locations of transgene insertions were determined by our method on chromosomes 11 and 3 in 4c30 mice and CK35 mice, respectively (Figs. 2C and 3C). Transgenes are often integrated into a host genome in tandem. In 4c30, the transgenes contained a *Dra*I site in the pCAGGS backbone (Fig. 1A) and an *Eco*RV site in the inserted cDNA sequence. If head-to-tail tandem repeats of the transgene exist in a host genome, then amplicons are produced in *Dra*I- and *Eco*RV-digested genome libraries. As expected, approximately 650-bp *Dra*I (band #1) and 1,700-bp *Eco*RV (band #2) bands were seen in the example (Fig. 2A). Tandem repeats were also found in CK35 mice (band #1 in *Dra*I library). Thus, genomic walking can be used for detection of tandem repeats, although this should be confirmed by sequencing.

Successful zygosity checks were also accomplished in both examples by PCR amplification with the flanking primers and tail DNA from transgenic and non-transgenic mice. The genotypes determined using the flanking primers matched those determined in advance using the Southern blot analysis. As shown in Fig. 2D, three primers for the 4c30 line were designed so that amplification would produce wild-type and transgenic alleles as 228 bp (Wild) and 175 bp (TG) bands, respectively (Fig. 2D). Analyses with these primers clearly provided the genotypes of the six mice (Fig. 2E). As shown in Fig. 3, flanking primer design for the CK35 line was also achieved, although certain elaborations were needed, because the mode of transgene integration in CK35 was more complicated than in 4c30. This topic will be discussed in detail

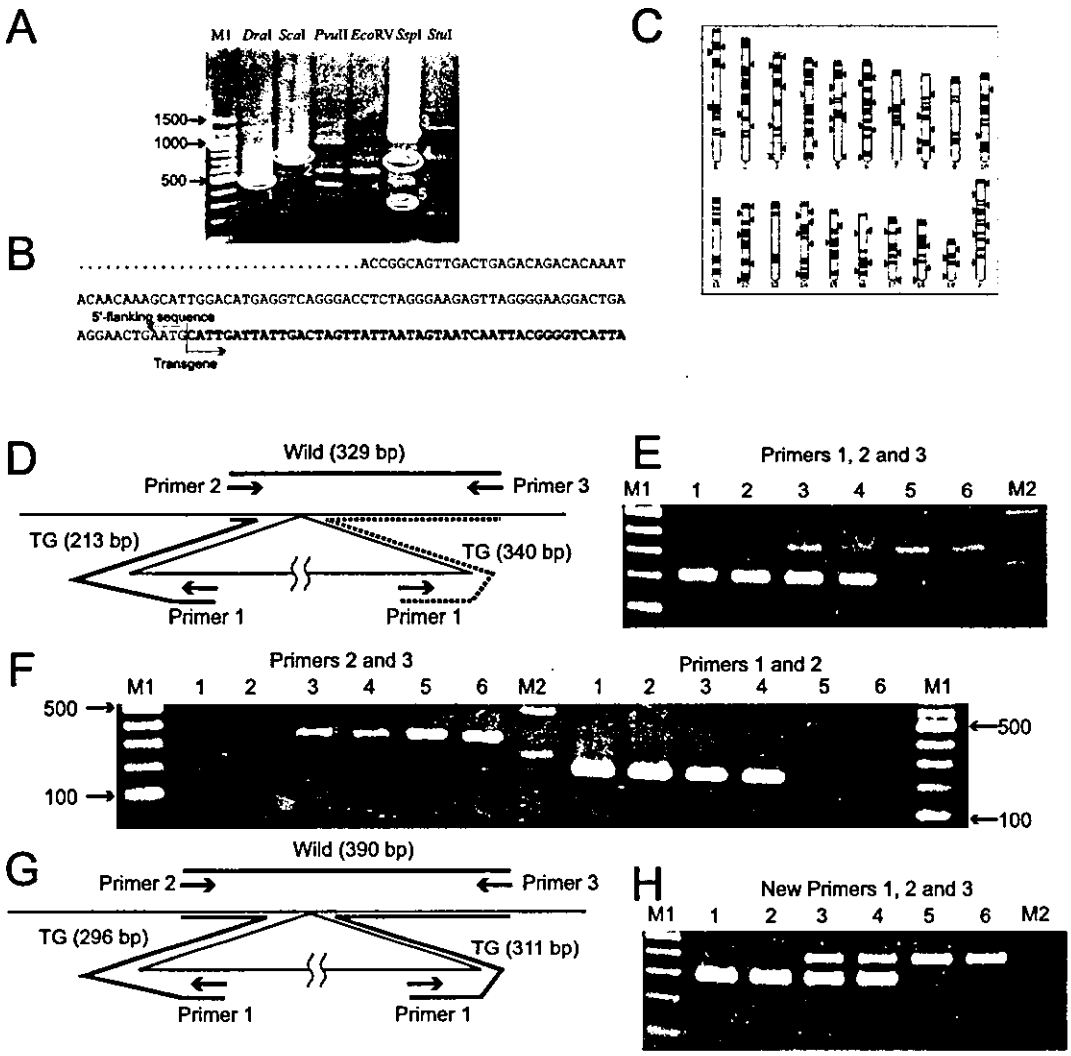


Fig. 3. Chromosomal mapping and zygosity check in transgenic mouse line CK35. The products of the nested PCR amplification of genomic libraries from CK35 genomic DNA were separated by agarose gel electrophoresis (A). Band #1 was derived from the tandem repeat of the transgene. Bands #2, #3 and #5 in Fig. 3A were not related to the transgene. The sequence flanking the transgene (B) was determined from the ~700-bp *Ssp*I band (band #4 in Fig. 3A). A search of the Ensembl genome database with the flanking sequence suggested that the transgene is located on chromosome 3 (boxed in panel C). Flanking primers for zygosity check of CK35 were designed as follows (panels D-H). DNA samples for Lanes 1 and 2 had been pre-confirmed as homozygous transgenic, lanes 3 and 4 as hemizygous transgenic, and lanes 5 and 6 as non-transgenic by Southern blot analysis. D) Flanking primer design for the zygosity check in line CK35 at the first attempt. Three primers were designed so that PCR amplification produced results in 329-bp bands (Wild) for wild-type alleles and 213-bp bands (TG) for transgenic allele, as indicated by solid lines. The dotted line shows unexpected TG bands produced by Primer 1 and Primer 3. E) The zygosity check of transgenic line CK35 with primers shown in panel D. Band patterns indicated that zygosities seen in lanes 1 and 2 were not relevant to our Southern blot analysis, while the patterns in the other lanes were the same. Sequence analysis of bands at a position similar to Wild bands in lanes 1 and 2 revealed that the bands were produced from the 3'-junction of the transgene by primer pair 1 and 3 as shown in a dotted line in panel D. F) The zygosity check of transgenic line CK35 with two sets of primer pairs in panel D. TG and wild-type alleles were clearly observed using primer pairs separately. G) The principle of the zygosity check by PCR with re-designed flanking primers for line CK35. The new primers were designed so that two TG bands, 296 bp and 311 bp long, derived from the 5'- and 3'-junctions of the transgene, respectively, would move to about the same position in the gel and yet be clearly distinguishable from a Wild band (390 bp). H) The zygosity check of transgenic line CK35 with new primers shown in panel G. Zygosity was clearly determined by band patterns, as already confirmed by Southern blot analysis. In panels E, F and H, M1 and M2 indicate 100-bp and 250-bp DNA ladders, respectively.

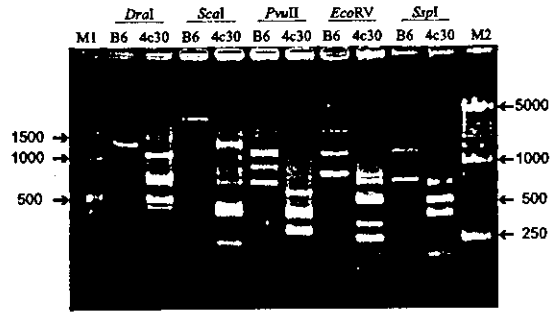


Fig. 4. Electrophoretic patterns of nested PCR products from backward genomic walking with B6 and 4c30 strains. Note that a considerable number of amplicons were found even in non-transgenic mouse genomic libraries as well as in 4c30 libraries. M1: 100-bp DNA ladder; M2: 250-bp DNA ladder.

below. PCR amplification with flanking primers has the potential to identify the transgene copy number from PCR product length using Primer 2 and Primer 3 shown in Fig. 1C. Nevertheless, in some cases this might be impossible if only a few repeats of the transgene exceed the length of DNA that routine PCR can amplify.

The mode of transgene integration can complicate flanking sequence determination. As shown in Fig. 4, our experiments revealed that the 5'-flanking region of pCAGGS-based transgenes was easier to determine than the 3'-flanking region, because primers with greater specificity could be designed for the 5'-flanking sequence. Primers for the CMV enhancer region are highly specific for detecting transgenes in the mouse genome, because this region is derived from a non-mammalian genome. Conversely, primers for the β-globin polyadenylation signal sequence residing downstream from the MCS of a pCAGGS-based transgene produced many non-specific bands (Fig. 4), probably because many highly homologous regions exist in the mouse genome. Efficient 3'-end genomic walking may require specific tag sequences at the 3'-end of the transgene. In addition, the mode of tandem repeat of the transgene should be considered when determining flanking sequences. Figure 5 illustrates the three patterns of transgene integration into the host genome. Using 5'-end genomic walking, one and two

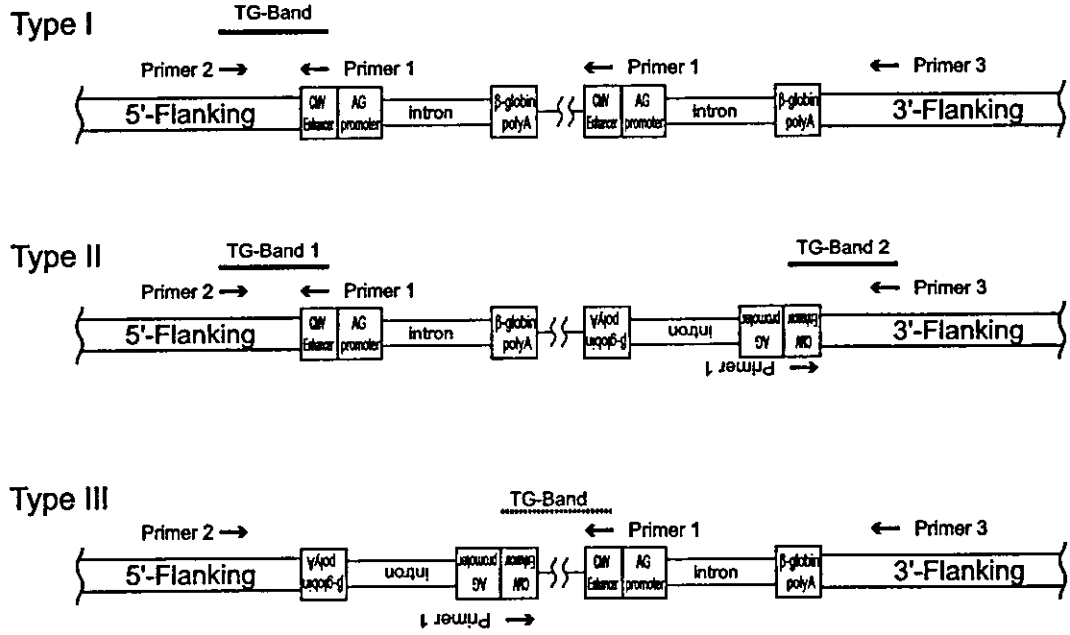


Fig. 5. Integration patterns of transgenes into a host genome. The three possible types of transgene integration at 5'- and 3'-junctions are: Type I, head and tail; Type II, head and head; Type III, tail and tail. In type I, only one expected band (TG-band) is produced by PCR amplification, while two bands (TG-bands 1 and 2) are amplified in type II. No band or an unexpected band (TG-band shown by a dotted line) may be produced by PCR in type III. See discussion for details.

PCR amplicons are obtained in types I and II, respectively. In type III, a total of three amplicons (two TG and one Wild band) should be considered for a zygosity check. In our study, the CK35 line turned out to be an example of the type II integration of the transgene (Fig. 3G). In type III (i.e., tail and tail orientation), a genome-specific sequence cannot be determined by our method, which captures the 5'-flanking sequence of pCAGGS-based transgenes. PCR amplification with Primer 1 might produce a band seen as a "TG-band" in type III, if a head-to-head concatenation of transgenes has occurred, but the band would not be informative with respect to the flanking genomic sequence. In addition, unknown sequences, such as fragmented copies of the transgene, might also be incorporated in the insertion site. Therefore, careful judgment should be exercised when determining which PCR bands contain sequences flanking the transgene.

The redundancy of genomic sequences also complicates localization of the insertion position of the transgene. Higher animals like mice have large genomes with complex nucleotide sequence conformations, such as gene duplications, pseudogenes, interspersed-type repeats such as SINE and LINE, etc. Consequently, similar nucleotide sequences are disseminated throughout the mouse genome. If the transgenes are integrated in or near such redundant sequences, it will be quite difficult to locate the transgene at a single location based only on the flanking sequence information. Our preliminary experiments showed that a sequence more than 200 bases long was insufficient to determine a single site in the mouse genome in certain cases, because many identical sequences were found in multiple chromosomes in the genome database (data not shown). Thus, sequence information is a powerful tool, but not always sufficient for practical genotyping.

Our mapping method has both advantages and disadvantages in comparison with the widely used cytogenetic mapping using the FISH technique. Mapping with FISH is more versatile, since it can be applied to any transgenes and does not depend on the flanking sequence information or the mode of integration of the transgenes. The resolution of mapping with FISH is lower than the resolution of our method, which can locate transgenes at the sequence level, whereas FISH does so at the chromosome band level. Although our method involves many steps, it can be carried out in three days using commercially available kits, which, in our opinion, is comparable to genomic mapping using FISH in terms of ease of performance.

Transgenes are not always integrated at a single site [6]. Our method can be applied to mice harboring transgenes in multiple sites, since multiple PCR products derived from multiple integration sites can be separated in a gel. Another possible strategy is to map transgenes in descendants in later generations, in which transgenes integrated at multiple sites are segregated. No triple-loci integration was observed in EGFP transgenic mice [6], suggesting that only two integration sites need be considered using descendants, even when the founder mice have been judged to have multiple transgene integrations using methods such as Southern blot analysis. The original composition of the transgene integration sites in founder mice could be re-constituted afterwards using a hereditary analysis of multiple sublines. Chromosome translocation was also observed in transgenic mice [6]. If chromosomes are rearranged at transgene integrations, flanking sequence information would be inappropriate for locating the position of the site, depending on the scale of the rearrangement. A zygosity check system could still be designed from flanking sequences determined using our method, but our method would not be applicable for precise mapping of the transgene.

A zygosity check by PCR amplification with flanking primers is powerful, because it is site-specific. Zygosity checking in transgenic animals is a critical step for analysis of transgene effects. All transgenic mice should be genotyped before experiments are performed. For production and maintenance of transgenic mouse lines, essentially all animals used as breeding pairs should be genotyped as well. Thus, large-scale genotyping should be routinely employed, and this requires a simple and rapid genotyping method. Several techniques are used for this purpose [reviewed in, e.g., 1]. Although Southern and dot blot analyses are commonly used, these methods are technically demanding and often provide ambiguous results. Moreover, the use of radioisotopes may cause problems in some laboratories. Although genotyping by FISH [4, 7] and real-time quantitative PCR amplification [13] have been reported, these methods do not offer the simplicity required for practical genotyping on a large scale. PCR amplification with flanking primers would be a rapid,

simple, and position-specific method for zygosity checking of the transgene. As shown in Fig. 1C, PCR amplification of a mouse genome with three primers enables zygosity to be judged by differential band formation. A drawback of the method is the need for nucleotide sequences that flank the transgene, which typically has an unpredictable insertion site. As described in this paper, the problem can be solved by genomic walking in most cases and, therefore, the flanking primer method can be practically applied. In addition, site-specific genotyping with flanking primers enables one to check the zygosity of each transgenic allele even when animals inherit transgenes from different transgenic lines with the same transgenes in different chromosomal positions (i.e., alleles). Since this checking system distinguishes between multiple transgenic lines containing the same transgene(s) inserted at different loci, multi-cross hybrids between lines can be produced with confirmed zygosity at individual sites. This results in the production of mice harboring more copies of the transgene than the original lines, a strategy that is useful if the gene dosage is a major concern.

Chromosome position and the zygosity of the transgene are important factors in the analysis of transgene expression in transgenic animals. In particular, the presence and chromosomal position of the transgene should be monitored for quality control of transgenic animals as genetically certified transgenic lines. As described above, determination of the flanking sequence of the transgene by genomic walking is useful, not only for chromosome mapping, but also for constructing zygosity check systems. Genomic walking is a powerful tool for genomic studies, especially with rapidly expanding genome databases like Ensembl [2]. For example, the second genomic walk for a zygosity check may be replaced by a genome database search. Thus, the rapid, simple method described here should facilitate the study of transgene effects.

References

1. Hogan, B., Beddington, R., Constantini, F., and Lacy, E. 1994. Identifying homozygous transgenic mice or embryos, pp. 305–308. *In*: Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor Laboratory Press, Woodbury.
2. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. 2002. The Ensembl genome database project. *Nucleic Acids Res.* 30: 38–41.
3. Ichikawa, S., Ozawa, K., and Hirabayashi, Y. 1998. Assignment1 of a UDP-glucose:ceramide glucosyltransferase gene (*Ugcg*) to mouse chromosome band 4B3 by in situ hybridization. *Cytogenet. Cell Genet.* 83: 14–15.
4. Kulnane, L.S., Lehman, E.J., Hock, B.J., Tsuchiya, K.D., and Lamb, B.T. 2002. Rapid and efficient detection of transgene homozygosity by FISH of mouse fibroblasts. *Mamm. Genome* 13: 223–226.
5. Lee, Y.C., Kurosawa, N., Hamamoto, T., Nakaoka, T., and Tsuji, S. 1993. Molecular cloning and expression of Galβ1,3GalNAc β2,3-sialyltransferase from mouse brain. *Eur. J. Biochem.* 216: 377–385.
6. Nakanishi, T., Kuroiwa, A., Yamada, S., Isotani, A., Yamashita, A., Tairaka, A., Hayashi, T., Takagi, T., Ikawa, M., Matsuda, Y., and Okabe, M. 2002. FISH analysis of 142 EGFP transgene integration sites into the mouse genome. *Genomics* 80: 564–574.
7. Nishino, H., Herath, J.F., Jenkins, R.B., and Sommer, S.S. 1995. Fluorescence in situ hybridization for rapid differentiation of zygosity in transgenic mice. *Biotechniques* 19: 587–590, 592.
8. Niwa, H., Yamamura, K., and Miyazaki, J. 1991. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108: 193–199.
9. Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H., and Suzuki, Y. 1988. Cloning, sequencing, and expression of cDNA for human β-galactosidase. *Biochem. Biophys. Res. Commun.* 157: 238–244.
10. Rozen, S. and Skaletsky, H.J. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html. 1998.
11. Schnieke, A., Harbers, K., and Jaenisch, R. 1983. Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene. *Nature* 304: 315–320.
12. Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23: 1087–1088.
13. Tesson, L., Heslan, J.M., Menoret, S., and Anegon, I. 2002. Rapid and accurate determination of zygosity in transgenic animals by real-time quantitative PCR. *Transgenic Res.* 11: 43–48.
14. Xiang, X., Benson, K.F., and Chada, K. 1990. Mini-mouse: disruption of the pygmy locus in a transgenic insertional mutant. *Science* 247: 967–969.

DNA RESEARCH 11, 371–379 (2004)

A Mutation in the Serum and Glucocorticoid-Inducible Kinase-Like Kinase (*Sgkl*) Gene is Associated with Defective Hair Growth in Mice

Kentaro MASUJIN,[1] Taro OKADA,[2] Takehito TSUJI,[1] Yoshiyuki ISHII,[2] Kaoru TAKANO,[3] Junichiro MATSUDA,[3] Atsuo OGURA,[4] and Tetsuo KUNIEDA[1,*]

Graduate School of Natural Science and Technology, Okayama University, Tsushima-naka, Okayama, 700-8530, Japan,[1] Graduate School of Agricultureal and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0857, Japan,[2] National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan,[3] and RIKEN Bioresource Center, Tsukuba, Ibaraki 305-0074, Japan[4]

(Received 5 August 2004; revised 7 October 2004)

Abstract

YPC is a mutant mouse strain with defective hair growth characterized by thin, short hairs and poorly developed hair bulbs and dermal papillae. To identify the gene associated with the phenotype, we performed genome-wide linkage analysis using 1010 backcross progeny and 123 microsatellite markers covering all chromosomes. The mutant locus (*ypc*) was mapped to a 0.2-cM region in the proximal part of mouse chromosome 1. This 0.2-cM region corresponds to a 450-kb region of genome sequence that contains two genes with known functions and five ESTs or predicted genes with unknown functions. Sequence analysis revealed a single C-to-A nucleotide substitution at nucleotide 1382 in the *Sgkl* gene, causing a nonsense mutation at codon 461. *Sgkl* encodes serum and glucocorticoid-inducible kinase-like kinase (SGKL), which belongs to a subfamily of serine/threonine protein kinases and has been suggested to have a role downstream of lipid signals produced by activation of phosphoinositide 3-kinase (PI3K). In the mutant SGKL, a serine residue in the C-terminal end of the protein (Ser486), which is indispensable for activation of SGKL upon phosphorylation, is abolished by premature termination. Specific expression of the *Sgkl* gene in the inner root sheath of growing hair follicles was also identified by *in situ* hybridization. Therefore, we concluded that the nucleotide substitution in the *Sgkl* gene is the causative mutation for defective hair growth in the *ypc* mutant mouse and that the signaling pathway involving SGKL plays an essential role in mammalian hair development.

Key words: Hair follicle; SGKL/SGK3/CISK; WNT signaling; Mutant mouse; IRS

1. Introduction

Hair follicle morphogenesis and the hair growth cycle are complex processes dependent on a series of mesenchymal-epithelial interactions in skin.[1] Reciprocal exchange of signals between dermal and epidermal cells of skin regulates the formation of hair placodes during embryonic development, and it also regulates cyclic transformation of the growth (anagen), regression (catagen), and quiescent (telogen) phases in the hair cycle in adult skin. As these processes show a high degree of organization and self-renewal, hair follicle development and hair cycling are thought to be excellent models for investigating the molecular mechanisms of mesenchymal-epithelial interactions.

Numerous growth factors and cytokines have been shown to be involved in morphogenesis and cycling of hair follicles. WNT,[2,3] TGFα,[4,5] BMPs,[6,7] and FGFs[8] in particular, as well as their signal transduction molecules,[9,10] play essential roles in these processes. Experiments with transgenic mice or those with knockout mutations in these genes have demonstrated a number of abnormalities in morphogenesis and cycling of hair follicles, including a short-hair phenotype and cyclical balding in transgenic mice overexpressing the *Wnt3* gene in skin[3] and abnormally long hair in *Fgf5* knockout mice, which is caused by defective regulation of the hair cycle.[8] On the other hand, spontaneous mutant mouse strains showing abnormalities in hair morphogenesis have also provided useful information on the molecular mechanisms of these processes. For example, the hairless (*hr*) mutant, which

Communicated by Michio Oishi

* To whom correspondence should be addressed. Tel. +81-86-251-8314, Fax. +81-86-251-8388, E-mail: tkunieda@cc.okayama-u.ac.jp

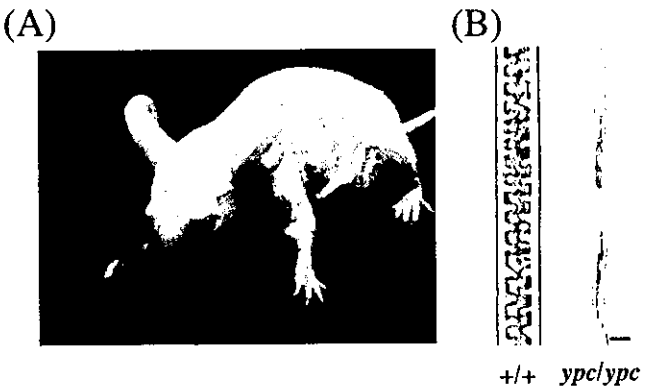


Figure 1. Phenotype of the *ypc/ypc* mouse. External appearance of a 4-week-old *ypc/ypc* male mouse showing poorly developed hairs and wavy vibrissae (A). Stereomicroscopic appearance of the hair shafts of *ypc/ypc* mice showing a disorganized hair medulla (B). Bar: 10 μm.

shows complete loss of hair and degeneration of hair follicles in the first regression phase, is caused by a retroviral insertion in the gene encoding a putative zinc-finger transcription factor,[11] and the waved-2 (*wa-2*) mouse phenotype, characterized by curly whiskers, waved hairs, and open eyelids at birth, is due to a single-base mutation in the EGF/TGFα receptor gene (*Egfr*).[12]

YPC is a mutant mouse strain established from a Swiss albino mouse colony at the National Institute of Health, Japan, which shows defective hair growth controlled by an autosomal single recessive gene (*ypc*). Homozygous *ypc/ypc* mice show thin and short hairs, wavy vibrissae (Fig. 1), and poorly developed hair bulbs and dermal papillae, but structures of the epidermis, dermis, and the sebaceous glands are normal.[13] These phenotypes of *ypc/ypc* mice are unique and distinct from other mutant mice with hair abnormalities. Therefore, the YPC strain could be an excellent model for investigating the molecular mechanisms underlying the morphogenesis and cycling of hair follicles, and identification of the gene responsible for the *ypc* mutation will provide new insight regarding genes involved in these processes and the mesenchymal-epithelial interaction.

In the present study, we mapped the *ypc* locus to the proximal region of mouse chromosome 1 by linkage analysis, sequenced genes in the critical region, and finally identified the causative mutation in a gene associated with intracellular transduction of signals implicated in the morphogenesis and cycling of hair follicles.

2. Materials and Methods

2.1. Animals

In the present study, we used mice of YPC and JF1/Msf strains maintained at Okayama University and those of an ICR colony. The YPC strain was provided by the National Institute for Infectious Diseases. The JF1/Msf strain, used as mating partners in the linkage analysis, was provided by the National Institute of Genetics. As the JF1/Msf strain is derived from *Mus musculus molossinus* and has distance from most of the inbred laboratory strains, this strain is useful for linkage analysis. The mice of the ICR colony, which were used as normal controls for sequence and histological analyses, were purchased from Japan CLEA Inc.

2.2. Histological analysis

Mid-dorsal skin samples of the mutant and normal mice at postnatal day 0 (P0), P3, and P5 were fixed in 10% neutral-buffered formalin, dehydrated through an alcohol gradient, and embedded in paraffin. Longitudinal sections of hair follicles 4 μm thick were stained with hematoxylin and eosin (HE). For high-resolution histology, small pieces of the skin were fixed in 2.5% glutaraldehyde and 2% paraformaldehyde (PFA) in 0.1 M phosphate buffer (PB), postfixed in 1% osmium tetroxide, embedded in Epon, and sectioned at a thickness of 1 μm. These sections were stained with toluidine blue.

2.3. Experimental cross and linkage analysis

A total of 1010 backcross progeny, including 496 affected and 514 unaffected mice obtained by crossing the YPC and JF1/Msf strains, were used for linkage analysis. Homozygous *ypc/ypc* mice of the YPC strain were mated with +/+ JF1/Msf mice and the resultant F_1 (*ypc*/+) mice were mated with homozygous *ypc/ypc* mice of the YPC strain to obtain the 1010 backcross progeny. Genomic DNA samples of the backcross progeny were obtained from livers by phenol/chloroform extraction and the genotypes of 123 microsatellite markers on mouse chromosomes were determined. Aliquots of 20 ng of genomic DNA were subjected to PCR amplification in 10-μl reaction mixtures, containing 1.5 mM Mg^{2+}, with 40 cycles consisting of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 60 sec. The PCR products were typed by polyacrylamide or agarose gel electrophoresis. The nucleotide sequences of the primers for these microsatellite markers were obtained from the Mouse Genome Informatics sequence database (http://www.informatics.jax.org/). The recombination fractions were calculated and the order of loci was determined using MapManager QT Ver. 3.0.

2.4. New microsatellite markers on the critical region

New microsatellite markers were obtained by searching GA/TC repeats on the sequence data of the region of mouse chromosome 1 close to the *ypc* locus. The sequence data were obtained from the mouse genome sequence in the NCBI database (http://www.ncbi.nlm.nih.gov). Among the GA/TC repeats found in this region, we chose the GA/TC repeats located inside or close to functional genes, and a total of 14 new microsatellite markers

Table 1. Microsatellite markers for the *ypc* critical region.

Microsatellite markers	Primer sequences		Product size (bp)
	Forward 5′-3′	Reverse 5′-3′	
D1Mok1	ACTAATAAGAAGAGTAGCAGCTGCC	AGGATTTACTGGTAGCCTACAAAGC	283
D1Mok2	GGTGCAAAACAAAGTTCCTAAGAGG	TGTGTGAGTGTAATGCACAGTAGG	252
D1Mok3	TCTTGAAATCTGGAAAAGAGATATCC	GTTCTTATAGTATTTGTCTTCTCTGG	280
D1Mok4	CCTCATCTTGCTGTAACCATAGAGG	GGTTTAGTATCTGGAATCATAGGTCC	276
D1Mok5	GACCACACCATGACTACCACACC	GCTTTCCTGGAGTTGGCTGAGG	256
D1Mok6	AGTAACAGAGACAGCTTTTGATCCC	GTCTCAGGAGACCACATGACAGG	254
D1Mok7	TAACGTTTCTTGACCAGGAAATAGC	TTTGCTGATCTGGCACACATTTTCC	241
D1Mok8	CAAGGCCAACCACAGTTACATAGG	TTCAGGGATTTATTTGGTTGATCTGG	302
D1Mok9	GGCCAATGACACCCTCTCCT	TGCACAGCTGCATTATAGGC	355
D1Mok10	GACATGGAATTCATGGTAGTAAGAGC	ACTCAAAGGAGACTGATCTCTGCC	316
D1Mok11	AACAAATCCTTAGCCATCAGAGTCC	GGGAGAATATTAGAGGAACTTATACC	375
D1Mok12	GGGGTGGGTACTCCCCATAGCC	AATGCATGTTGGAGCTTGTTCCACC	418
D1Mok13	TGGAACATCCAGGCAGTCTATAGC	TTGGGATGGAGACCCTGGTTACC	216
D1Mok14	CGTCACTGTGAGACTCTGACAGG	GAACTTTCACTATGCAGTTTCCACG	361

(*D1Mok1 to D1Mok14*) were obtained. The nucleotide sequences of the primers for these markers are shown in Table 1. The genotypes of these microsatellite markers in the backcross progeny were determined with the same conditions as described above.

2.5. *Cloning and sequence analysis*

The entire coding regions of two genes and an EST were amplified from skin RNA of normal and mutant mice by RT-PCR. Total RNA samples were obtained by the acid guanidinium-phenol-chloroform (AGPC) method. First-strand cDNA was synthesized from 10 μg of total RNA, using oligo d(T) primer and Superscript II reverse transcriptase (Gibco-BRL, Gaithersburg, MD) and used for PCR amplification with primer pairs for these genes. The nucleotide sequences of these primers obtained from EMBL/GenBank databases are shown in Table 2. PCR amplification was carried out for 30 to 40 cycles, consisting of 94°C for 45 sec, 50–58°C for 45 sec, and 72°C for 120 sec in 10-μl reaction mixtures containing 1.5 mM Mg^{2+}. The amplified fragments were cloned into the pGEM-T Easy Vector system (Promega, Madison, WI) and their nucleotide sequences were determined by the dideoxy chain termination method with a Hitachi SQ5500 automated DNA sequencer. To distinguish authentic base substitutions from PCR errors, at least five independent clones were completely sequenced in both orientations.

2.6. *Genotyping of the* Sgkl *gene*

To confirm the correlation between the nucleotide substitution and the phenotype, genomic DNAs obtained from mice of the YPC strain, the backcross progeny, and other inbred strains including JF1/Msf, C57BL/6J, BALB/c, and ICR were used for genotyping the *Sgkl* gene. The 380-bp region flanking the substitution was amplified by PCR using a pair of primers 5′-TGAAAGGAAGGTAAGGTGAA-3′ and 5′-GCCCTATTTCTTGCATACAG-3′. The PCR products were digested with *Mse* I restriction endonuclease and the digests were electrophoresed through 2% agarose gel in TBE buffer and visualized with ethidium bromide staining.

2.7. In situ *hybridization*

For *in situ* hybridization, DIG-labeled sense and antisense *Sgkl* riboprobes were synthesized using T7 and SP6 RNA polymerases (Roche Diagnostics, Mannheim, Germany) from the vector pCRII-TOPO (Invitrogen, Carlsbad, CA) containing a 618-bp fragment of the mouse *Sgkl* gene (nt 311 to 928, NM_133220). *In situ* hybridization was carried out as described previously[14] with some modifications. In brief, 4% PFA-fixed mid-dorsal skin paraffin sections 8-μm thick were acetylated, treated with 0.2 M HCl, digested with 10 μg/ml proteinase K for 20 min, and then fixed with 4% PFA. Following prehybridization, sections were hybridized with DIG-labeled sense or antisense probes at 57°C overnight, and then the sections were washed in 50% formamide/SSC, digested with 20 μg/ml RNAse A, and re-washed in SSC.

Table 2. Primers for genes in the *ypc* critical region.

Genes	Fragments	Primer sequences		Product size (bp)
		Forward 5'-3'	everse 5'-3'	
Sgkl	Sgkl-1	GAATTCAGGGCTTCACAGGA	GCCCACAGAGACCAGAACT	340
	Sgkl-2	ACTACAAGGAGAGCTGCCCA	TGAGGTAGAATGTGGCTTCG	410
	Sgkl-3	GGACAGCCCAAGACATCAGT	AACCTCGCTCTGGGTTCAGG	438
	Sgkl-4	CCTCCAAAGGGAAAGGTCTT	ACTCCTGGTCTCAAGTTTAAGGGC	399
	Sgkl-5	GCCGAGATGTTGCTGAAATG	GGTTTCTGAATGTCAAAGTG	425
Mybl1	Mybl1-1	CTCTCGGTCACCTGAGGGAA	GTCATCCTCATCTCTTGTCC	260
	Mybl1-2	CATGATTATGAAGTACCTCAAC	GTCCAGGAAGACTTCTTCAC	374
	Mybl1-3	GTGCAGAGAAAGATGGCA	TACCCAGGGATCTGAACAGG	327
	Mybl1-4	TGGACCATTTGCAAACCCAG	GTGGGTGAATTCTGCTGAGC	367
	Mybl1-5	TGGAGGAACACACTACTGAG	ATAGCTTCACTGTCTCCACC	349
	Mybl1-6	GTCAGTCTTGTACTTGAAGGG	GATCTTTGGGGGTTGCTTCC	367
	Mybl1-7	CAACCCCAATTTGTGGGCAG	CCAGCTCTCTAAGGCAAGGG	368
	Mybl1-8	CTCTGCATCTGTGAAGAAGG	GGGTAGGGCATTTTATCAGC	443
1700011J18Rik	18Rik-1	GGAGGAGCCTGGCACCAGGA	GGTTTCTGCAAAGAGAGAGG	307
	18Rik-2	CTTGACACAATGCCTGAAGAA	GAGGAGCCTCTGACTTTCA	322
	18Rik-3	CTGGGAGGCAATTACATTGC	CATGTAAATACTTTACGTCGAC	284

Following antigen blocking, the sections were incubated with alkaline phosphatase conjugated anti-DIG antibody (1:500, Roche Diagnostics) at 4°C overnight. Positive signals were visualized by BCIP (5-bromo-4-chloro-3-indoyl phosphate) and NBT (nitroblue tetrazolium).

3. Results

3.1. Histological examinations

As shown in Fig. 1, the homozygous *ypc/ypc* mice showed poorly developed hairs and wavy vibrissae. The hair shafts of *ypc/ypc* mice were thinner and shorter than those of normal mice and had a disorganized hair medulla. To investigate the precise defects in postnatal hair follicle development in the *ypc/ypc* mouse, we performed histological examinations of mid-dorsal skin of *ypc/ypc* and +/+ mice at P0, P3, and P5. While no histological differences were observed between *ypc/ypc* and +/+ mice at P0 (Figs. 2A and 2B), the hair follicles in the *ypc/ypc* mice at P3 showed irregular structures characterized by small hair bulbs, narrow and immature hair shafts, and lack of uniform orientation (Figs. 2C and 2D). These features were also observed at P5, and the length of hair follicles of *ypc/ypc* mice was shorter than that in +/+ mice at this stage (Figs. 2E and 2F). At high magnification, toluidine blue-stained sections showed a reduced number of matrix cells, disorganized hair medulla, increased thickness of outer root sheath (ORS), and decreased thickness of inner root sheath (IRS) (Figs. 2G and 2H) in *ypc/ypc* mice. In conclusion, histological examination of *ypc/ypc* mice revealed abnormal postnatal development of the hair follicles, including small hair

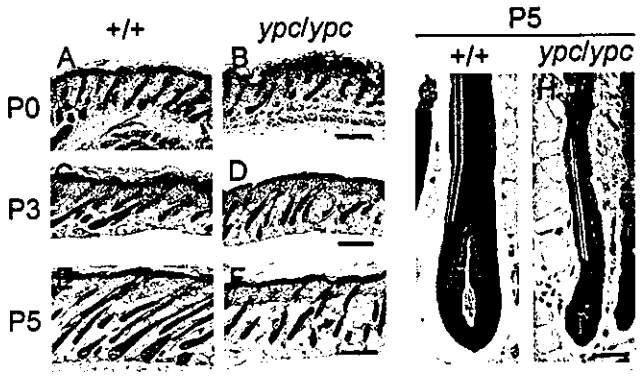


Figure 2. Histological findings of postnatal day 0 (P0), P3, and P5 *ypc/ypc* mice. At P0, no morphological differences were observed between *ypc/ypc* and +/+ mice (A, B). At P3, smaller hair bulbs, disorganized hair shafts, and a lack of uniform orientation were observed in *ypc/ypc* (C, D). At P5, follicles of *ypc/ypc* mice still showed the structural abnormalities, and the length of the hair follicles became shorter than that in ICR (E, F). Toluidine blue-stained sections of *ypc/ypc* mice with high magnification showed a reduced number of matrix cells, disorganized hair medulla, increased thickness of outer root sheath (ORS), and decreased thickness of inner root sheath (IRS). Outer and inner yellow lines denote the boundary between ORS and IRS, and IRS and hair shaft, respectively. Bars: 200 μm (A tó F) and 50 μm (G and H).

bulbs, disorganized hair shafts, and lack of uniform orientation.

3.2. *Linkage analysis*

To localize the *ypc* locus on mouse chromosomes, initial linkage analysis was performed using 45 backcross progeny, including 17 affected and 28 normal mice. By typing the 100 microsatellite markers covering the entire mouse genome, significant linkage was observed between the *ypc* locus and the microsatellite marker *D1Mit4* on the proximal region of chromosome 1. No significant linkage was observed between the *ypc* locus and any of the other 99 markers examined. These findings indicated that *ypc* is localized in the proximal region of mouse chromosome 1. We next performed fine linkage mapping of the *ypc* locus using a total of 1010 backcross progeny and 23 microsatellite markers in this region, including 14 new microsatellite markers (*D1Mok1* to *D1Mok14*) obtained by searching for GA/TC repeats in the sequence data of this region. The segregation of alleles of these markers in the F_2 progeny is shown in Fig. 3A. Fine linkage mapping revealed that the *ypc* locus is positioned in a 0.2-cM region between microsatellite markers *D1Mok5* and *D1Mok8*, and no recombination was observed between *ypc* and *D1Mok6* or *D1Mok7*. As shown in Fig. 3B, comparison of the linkage map with the published sequence of the mouse genome showed that the 0.2-cM region corresponds to a 450-kb region containing two genes with known function (*mybl1* and *Sgkl*), three ESTs (311003E14Rik, 5730538E15Rik, and 1700011J18Rik), and two sequences predicted to be coding sequences of genes (LOC329092 and 6030422M02).

3.3. *Sequence analysis of the genes located in the critical region*

We sequenced the entire coding regions of the *mybl1* and *Sgkl* genes as well as 1700011J18Rik in *ypc/ypc* and +/+ mice by RT-PCR and cloning of the amplified fragments. As the features of the other two expressed sequence tags (ESTs) and two predicted genes were unlikely to be functional and these sequences are not conserved between mouse and human, they were not sequenced. Comparison of the nucleotide sequences of these genes and EST between *ypc/ypc* and +/+ mice revealed a single C-to-A nucleotide substitution in the *Sgkl* gene, which encodes serum and glucocorticoid-inducible kinase-like kinase (SGKL). The nucleotide substitution occurred at nucleotide 1382, and resulted in premature termination at codon 461 (Fig. 4). Therefore, the SGKL protein in *ypc/ypc* mutant mice lacks the 35 amino acid residues of the C-terminal end of the protein. The C-terminal end is highly conserved among mouse and human and contains a serine residue for phosphorylation, which is indispensable in activation of SGKL. No mutation in the nucleotide sequence affecting the amino acid sequence of the proteins, including missense, nonsense, and frame shift mutations, were observed in the remaining gene or EST. These findings clearly indicate that the nonsense mutation of the *Sgkl* gene is associated with the mutant mouse phenotype.

3.4. *Association between the nucleotide substitution and the phenotype*

To confirm the association between the nonsense mutation and the phenotype, the genotype of the *Sgkl* gene was examined in various mice of the YPC strain, the back cross progeny, and other inbred strains. A 380-bp region including the nucleotide substitution was amplified by PCR. Since the nucleotide substitution is located in a recognition sequence of *Mse* I endonuclease, the amplified fragment of the mutant allele has an *Mse* I cleaving site, while that of the wild-type allele has no *Mse* I cleaving site. Digestion of the amplified fragment with *Mse* I demonstrated that all mice of the YPC strain and all affected mice of the backcross progeny gave digested fragments, all phenotypically normal mice of the backcross progeny gave both the digested and undigested fragments, and all mice of the other strains gave the undigested fragment. Therefore, the nucleotide substitution of the *Sgkl* gene was perfectly associated with the phenotype in these mice.

3.5. *Expression of* Sgkl *in developing hair follicles*

To investigate the function of the *Sgkl* gene in hair development, we performed *in situ* hybridization of the *Sgkl* gene using skin sections of normal mice at different stages of hair cycle. The *in situ* hybridization showed no positive signal at P0 (Fig. 5A) and strong positive signals at P3, P5, and P7 (Fig. 5B, C, and D). In these stages, *Sgkl* mRNA was detected in the cells of the IRS, which surrounds the hair shaft and plays an important role in formation of the hair shafts. The *Sgkl* mRNA was also detected in a subpopulation of matrix cells at P3. Since the cells forming the IRS, which express the *Sgkl* gene, first appear around P3 in the normal mouse and the morphological defect in the *ypc/ypc* mouse also appears at P3, the expression pattern of *Sgkl* in the normal mouse was correlated with the appearance of the morphological defect in the mutant mice.

4. Discussion

SGKL (also termed SGK3 or CISK; Cytokine-Independent Survival Kinase) has been identified as a homolog of serum- and glucocorticoid-inducible kinase (SGK1),[15–17] a serine/threonine protein kinase belonging to the AGC (protein kinase A, protein kinase G, and protein kinase C) subfamily of protein kinases.[18] SGKs are activated by 3-phosphoinositide-dependent protein