Available online at www.sciencedirect.com
ScienceDirect
Gene 384 (2006) 62–72
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Isolation and molecular characterization of the porcine transforming growth factor beta type I receptor (*TGFBR1*) gene☆

Kefei Chen [a], Laurie A. Rund [a], Jonathan E. Beever [a,b], Lawrence B. Schook [a,b,*]

[a] *Department of Animal Sciences, University of Illinois at Urbana–Champaign, 1201 W. Gregory Dr., Urbana, IL 61801, USA*
[b] *Institute for Genomic Biology, University of Illinois at Urbana–Champaign, 1201 W. Gregory Dr., Urbana, IL 61801, USA*

Received 4 May 2006; received in revised form 10 July 2006; accepted 11 July 2006
Available online 25 July 2006
Received by M. D'Urso

Abstract

The transforming growth factor beta (TGF beta) family is essential for normal growth and development of different organ systems. Here we describe the isolation and molecular characterization of the full-length cDNA and the determination of the genomic DNA sequence of the porcine *TGFBR1* gene. The full-length *TGFBR1* cDNA 1813 bp contains an open reading frame (ORF) of 1512 bp encoding a TGFBR1 protein of 503 amino acids with a calculated molecular weight (Mw) of 56.4 kDa. A BAC clone harboring the porcine *TGFBR1* gene was isolated and sequenced. The results of genomic and cDNA sequences of the porcine *TGFBR1* gene demonstrated that it spans a transcription unit of 62,182 bp consisting of nine exons ranging from 125 to 354 bp, and eight introns ranging from 1003 to 29,441 bp. A shorter porcine TGFBR1 isoform resulting from the alternative splicing of exon 7 in porcine *TGFBR1* cDNA was detected. The shorter TGFBR1 isoform contained a 1140 bp ORF encoding 379 amino acids with a calculated Mw of 41.7 kDa. The core promoter of porcine *TGFBR1* gene lacks a TATA box but contains GC boxes and CAAT boxes. Multiple transcription initiation and termination sites were identified in untranslated regions (UTR) resulting in the size of 5′-UTR varying from 15 to 62 bp, and the length of 3′-UTR varying from 169 to 228 bp. Quantitative real time PCR results showed that the *TGFBR1* transcript was ubiquitously expressed in all tissues examined (i.e. fat, adrenal, brain, spinal cord, muscle, mandibular lymph node, thymus, bone marrow, uterus, spleen, testis, kidney, liver, and ovary). A total of eighty-five gene polymorphisms (77 SNPs and 8 indels) were detected in the porcine *TGFBR1* gene by utilizing a panel of DNA from eight diversified pig breeds (Yorkshire, Chinese Meishan, Berkshire, Duroc, Hampshire, Landrace, Large White and Pietrain). The minor allele frequencies of these nucleotide variations varied from 0.13 to 0.5 with an average of 0.26. In addition, seventeen microsatellites were identified throughout the genomic sequence of the porcine *TGFBR1* gene.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Genomic organization; Splicing variants; SNP; Microsatellites; Expression; Pig

Abbreviations: AP, activating enhancer binding protein; BAC, bacterial artificial chromosome; bp, base pair; cDNA, complementary to RNA; CTF, CCAAT-box binding transcription factor; dNTP, deoxyribonucleoside triphosphate; FISH, fluorescence *in situ* hybridization; GCF, GC-binding factor; indel, insertion/deletion; kb, kilobase pair; kDa, kiloDalton; LINEs, long interspersed elements; LTR, long terminal repeat; MLTF, major late transcription factor; Mw, molecular weight; ORF, open reading frame; PCR, polymerase chain reaction; pI, isoelectric point; Q-PCR, quantitative PCR; QTL, quantitative trait loci; RACE, rapid amplification of cDNA end; RT, reverse transcription; SINEs, short interspersed nucleotide elements; SNPs, single nucleotide polymorphisms; Sp1, stimulatory protein 1; TGFBR1, transforming growth factor beta type I receptor; UCE, upstream control element; UTR, untranslated region(s).

☆ The nucleotide sequences reported in this paper were deposited in GenBank nucleotide sequence database under the accession numbers DQ519377–DQ519383.

* Corresponding author. University of Illinois at Urbana–Champaign, Department of Animal Sciences, 1201 W. Gregory Dr., Edward R. Madigan Laboratory 382, Urbana, IL 61801, USA. Tel.: +1 217 265 5326; fax: +1 217 244 5617.
E-mail address: schook@uiuc.edu (L.B. Schook).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.07.009

1. Introduction

Transforming growth factor beta type I receptor (TGFBR1, EC 2.7.1.37) is also known as activin receptor-like kinase 5 (ALK-5), and serine/threonine protein kinase R4 (SKR4). Transforming growth factor beta (TGF-β) belongs to a large superfamily of structurally related multifunctional proteins that includes activins, inhibins, bone morphogenetic proteins, myostatin, and Müllerian inhibiting substance (Kingsley, 1994; Moses et al., 1990). Three different isoforms of TGF-βs, named TGF-β1, TGF-β2, and TGF-β3, have been identified in various mammalian tissues (Massague, 1990). TGF-βs exert their effects through binding to specific cell surface receptors that consist of TGF-β type I, type II, and type III receptors (Massague, 1992). The type I and II receptors are transmembrane serine/threonine kinases which form heterodimers and are essential for signal transduction, whereas type III receptor is a proteoglycan which regulates the access of TGF-βs to the signaling receptors (Laiho et al., 1990; Lopez-Casillas et al., 1991; Segarini et al., 1989; Wang et al., 1991). TGF-βs bind directly to the TGFBR2 subunit of the heterodimer, which allows TGFBR2 to activate TGFBR1 by phosphorylation and subsequently results in downstream signaling generated by the type I receptors (Wrana et al., 1994). These secreted polypeptides play fundamental roles in development and tissue homeostasis by regulating cell proliferation, differentiation, apoptosis, and extracellular matrix formation (Chang et al., 2002; Massague et al., 2000). Heterozygous mutations in human *TGFBR1* and *TGFBR2* genes are linked to inherited disorders in cardiovascular, craniofacial, neurocognitive and skeletal development (Loeys et al., 2005). In particular, during skeletal development, TGF-βs and receptors have unique functions and act sequentially to modulate chondrocyte and osteoblast differentiation (Moses and Serra, 1996).

The human *TGFBR1* gene contains 9 exons spanning approximately 45 kb, and was mapped on chromosome 9q33-q34 by a hybrid cell DNA panel and fluorescence *in situ* hybridization (FISH) (Johnson et al., 1995) and 9q22 by FISH (Pasche et al., 1998). Recently, porcine *TGFBR1* was mapped between microsatellite markers *SW803* and *SW705* on SSC1 using linkage and RH mapping (Kopecny et al., 2004), and the cDNA sequence was reported (Shimanuki et al., 2005). As *TGFBR1* is located within a quantitative trait loci (QTL)-rich region on SSC1 for a variety of economically important traits (Hu et al., 2005), and its role in various developmental aspects in humans (Loeys et al., 2005); this suggests its potential role as a candidate gene for growth and development traits. In the present study, we report the genomic organization of the porcine *TGFBR1* gene determined by the complete cDNA and genomic DNA sequences, gene polymorphisms, one isoform of the *TGFBR1* cDNA with alternative splicing of the entire exon 7, multiple transcription initiation and termination sites at the UTR regions along with its expression patterns in a variety of porcine tissues.

2. Materials and methods

2.1. DNA sequencing

Purified plasmid DNAs were bidirectionally sequenced with the ABI Prism® BigDye™ Terminator Cycle Sequencing Kit (version 3.0) and M13 primers (rev.: 5′-AGC GGA TAA CAA TTT CAC ACA GGA-3′ and univ.: 5′-TAA TAC GAC TCA CTA TAG GG-3′). PCR products were sequenced with the respective gene-specific primers and 50–100 ng DNA-template directly after purification using the Qiagen MinElute 96 UF PCR purification kit (Qiagen, Valencia, CA) or ExoSAP-IT (USB Corporation, Cleveland, OH). All sequencing reactions were analyzed on an ABI 3730 DNA capillary sequencer (Applied Biosystems, Foster, CA).

2.2. Shotgun sublibrary construction and bioinformatics

The BAC clone RP44-329D7 containing the porcine *TGFBR1* gene was selected using an *in silico* tool (www.sanger.ac.uk/Projects/S_scrofa/mapping.stml) based on BAC fingerprints and end-sequences, and verified by PCR amplification using primer pairs TBR-E1F/E1R and TBR-E9F/E9R (Table 1) derived from porcine *TGFBR1* sequences (GenBank accession no: AB182258). BAC DNA was isolated by a NucleoBond plasmid purification kit (BD Biosciences, Palo Alto, CA). BAC DNA (5 μg) was sheared to 1.5 kb fragments by TOPO® shotgun subcloning kit (Invitrogen, Carlsbad, CA) and small molecules were removed using SizeSeq™ 400 Spun columns (Amersham

Table 1
Primers used for BAC clone identification and cDNA analysis: primer names, primer sequences, optimal annealing temperatures and amplification sizes

Primer name	Forward primer sequences (5′–3′)	Reverse primer sequences (5′–3′)	Size (bp)	Tm
TBR-E1F/E1R	GAGGCGAAGCTTGTTGAGG	GAGAAGGAGCGAGCCAGAG	231	58
TBR-E9F/E9R	TGCCGAAACAGAAACACTTG	TTGTAACCGTCCTGGGAAAG	471	58
TBR-5outer/5-RACE-outer	AGACTGGTCCAGCAATGACAGC	Supplied with the 5′-RACE kit	n/a	58
TBR-5inner/5-RACE-inner	TGTCTCGCGGAATTAGGTCAAT	Supplied with the 5′-RACE kit	333	58
TBR-3outer/3-RACE-outer	AGCTCGACGATGTTCCATTGGT	Supplied with the 3′-RACE kit	n/a	58
TBR-3inner/3-RACE-inner	CTCAGTCAGCAGGAAGGCATCA	Supplied with the 3′-RACE kit	300	58
TBRRT-AF/AR	ATGGACTCAGCTTTGGTTGG	CACTCTGTGGTTTGGAGCAA	304	58
TBRRT-BF/BR	CCAACCAAAGCTGAGTCCAT	ACTGTTGGAAAGCCACCATC	499	58
TBRRT-CF/CR	AGTGTTTCTGCCACCTTTGC	ATGTGAAGATGGGCAAGTCC	850	58
TBRRT-DF/DR	TGGAGAGGAGAGGAAGTTGC	CCTGGGTCCAAAGAAATCCT	1012	58
18S-rRNA-F/R	CGGCTACCACATCCAAGGAA	GCTGGAATTACCGCGGCT	198	58

Tm: Annealing temperature.

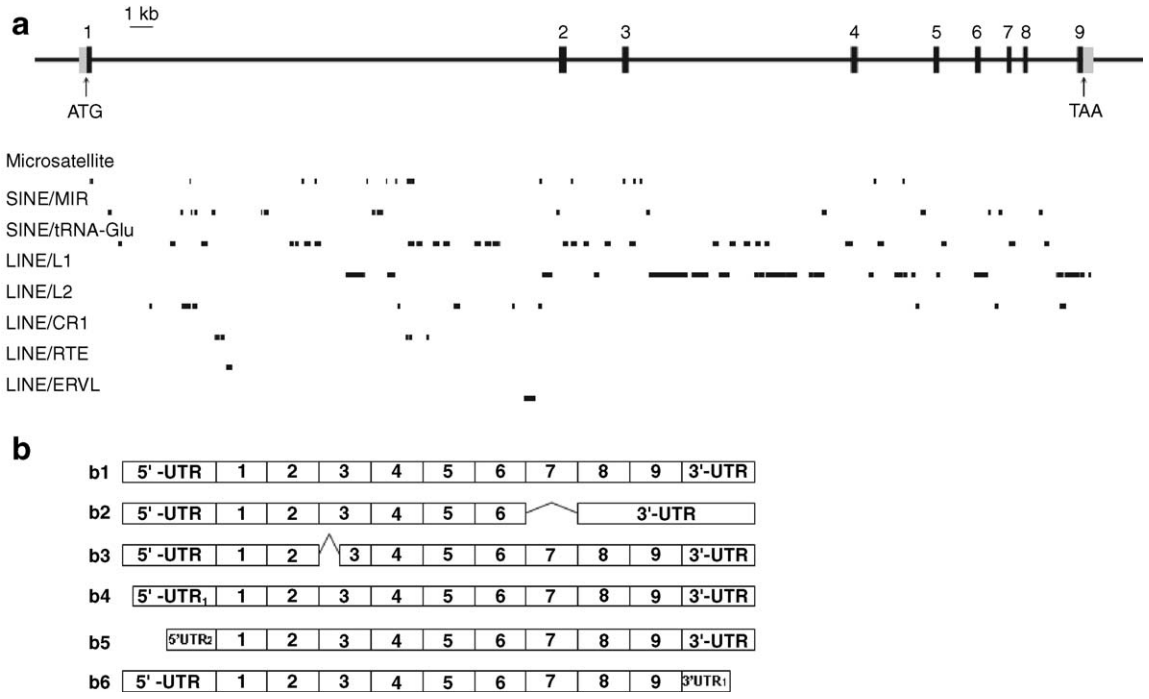

Fig. 1. Genomic structure and splicing variants of the porcine *TGFBR1* gene. The porcine *TGFBR1* gene spans 62,182 bp consisting of 9 exons (GenBank accession no: DQ519377). Genomic structure of the porcine *TGFBR1* gene and the locations of repetitive elements including microsatellites, SINEs, LINEs and LTRs are indicated (a). The thin line indicates non-coding regions. Exons are shown as closed boxes and are numbered. The 5′ and 3′ untranslated regions are illustrated as grey boxes. The start codon (ATG) and the stop codon (TAA) are indicated by arrows. The sizes are drawn to scale in Panel A. The cDNA structure of the porcine *TGFBR1* gene and alternative splicing variants along with multiple transcription initiation and termination sites are represented in Panel B. The fragment deletions are illustrated by curved lines. The specifics are: b1: the longest cDNA of 1813 bp with the transcription initiation site 62 bp upstream of the start codon, and the transcription termination site 228 bp downstream of the stop codon (GenBank accession no: DQ519378); b2: alternative splicing of the entire exon 7 resulting in the peptide sequences stopped at 2 amino acids following the sixth exon, and exon 8 along with exon 9 joining to 3′-UTR region (GenBank accession no: DQ519379); b3: alternative splicing of 12 bp at the beginning of exon 3 (GenBank accession no: DQ519380); b4–b5: alternative transcription initiation sites 56 bp and 15 bp upstream of the start codon, respectively (GenBank accession nos: DQ519381, DQ519382); b6: alternative transcription termination site 169 bp downstream of the stop codon (GenBank accession no: DQ519383).

Pharmacia, Piscataway, NJ). The sheared DNA fragments were blunt-end repaired by T4 DNA polymerase and Klenow DNA polymerase and dephosphorylated by Antarctic Phosphatase (New England Biolabs, Ipswich, MA). The blunt-end DNA was then subcloned into polylinker pCR4-TOPO® vector (Invitrogen, Carlsbad, CA). Recombinant plasmids were used to transform One Shot® TOP10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA). Shotgun sequences were assembled by phred/phrap/consed (http://www.phrap.org/phredphrapconsed.html). All sequences were trimmed for vectors and *E. coli* DNA sequences. After quality assessment ($Q>20$), sequences were masked for repetitive elements (http://www.repeatmasker.org/), and subjected to BLAST analysis for similarity to human genome sequences of HSA9 using NCBI-BLASTn (Altschul et al., 1997). An expectation value of e^{-5} was used as the threshold. Remaining sequence gaps were closed by primer walking until both strands of the plasmid DNAs were bidirectionally sequenced. Primers were designed using primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

Genomic organization was determined by the alignment of the genomic DNA sequence with the obtained cDNA sequence, and by comparative analysis with the human, mouse, rat and bovine *TGFBR1* gene. The putative promoter region was predicated with the Proscan software, version 1.7 (http://thr.cit.nih.gov/molbio/proscan/). To identify regulatory domains within the promoter, the 500-bp sequence upstream of the transcription initiation site of the *TGFBR1* in pig, human, mouse and rat was aligned using DNAMAN analysis software, version 5.0 (http://www.lynnon.com). Possible CpG islands were searched with the program CpGfinder on Softberry (http://www.softberry.com/berry.phtml). Protein patterns, profiles and motifs were predicted using available ExPASy analysis system (http://www.expasy.org/tools/scanprosite/). PFAM domains were detected by the simple modular architecture research tool (SMART) http://smart.embl-heidelberg.de/ and Pfam protein search at http://pfam.wustl.edu/hmmsearch.shtml. Comparative sequence alignment was performed by the online tool of the European Bioinformatics Institute, the ClustalW 1.82 program (http://www.ebi.ac.uk/clustalw/).

2.3. Isolation of RNA and expression analysis

Total RNA from 14 porcine organs (fat, adrenal, brain, spinal cord, muscle, mandibular lymph node, thymus, bone marrow, uterus, spleen, testis, kidney, liver, and ovary) was isolated with TRIZOL™ according to the manufacturer's instructions (Invitrogen, Carlsbad, CA). Concentration of total RNA was quantified by Eppendorf Biophotometer (Eppendorf, Westbury,

Table 2
Exon–intron junctions of the porcine *TGFBR1* gene

No.	Exon size (bp)	Splicing-donor	Intron size (bp)	Splicing-acceptor
1	147	147 ACGG**gt**gagcggcg… Thr A 393	29,441	148 …ctttttcc**ag**CATTA la Leu 394
2	246	GTTG**gt**aagttgtt… Val G 636	3818	…cctttttccag**ag**GTCCT ly Pro 637
3	243	TCAG**gt**aatgtatt… Ser G 867	13,924	…ttttttttt**ag**GTTTA ly Leu 868
4	231	AAAG**gt**cagtaata… Lys A 1035	4964	…ctacttac**ag**ATAAT sp Asn 1036
5	168	CAAG**gt**aagtcaaa… Gln G 1192	2207	…tcttcatt**ag**GAAAA ly Lys 1193
6	157	AAAAG**gt**acacttat… Lys Ar 1317	1950	…tcttttttt**ag**GTAC g Tyr 1318
7	125	GGTG**gt**aaatctcc… Gly G 1448	1003	…accaatgt**ag**GAATT ly Ile 1449
8	131	TGTGAA**gt**aagttgtt… Cys Glu 1802	3073	…cttttcc**ag**GCCTTA Ala Leu
9	354	GGTTA***C***aaatttgtcg…		

Coding sequences are shown in uppercase letters and non-coding regions in lowercase letters. The conserved GT/AG exon/intron junctions are shown in bold. Amino acid residues are indicated with respect to each boundary. For the last exon, the transcription termination is shown in bold italics instead of a splice donor. Number refers to the corresponding positions in the longest porcine *TGFBR1* cDNA transcription initiation point.

NY). Total RNA (2 μg) was used for reverse transcription PCR (RT-PCR) with the OmniscriptKit (Qiagen, Valencia, CA) using 10 μM Oligo-$(dT)_{24}$V, 10 U RNase inhibitor, 5 mM of each dNTP and 1 U of Omniscript RT in the supplemented RT-buffer at 37 °C for 1.5 h. Expression pattern of *TGFBR1* was analyzed in 14 different tissues by RT-PCR amplification using gene-specific primers TBRRT-AF/AR and TBRRT-BF/BR (Table 1). To further investigate quantitatively the *TGFBR1* expression, the gene-specific primers and the 18S rRNA as internal control were employed to perform real time RT-PCR on an ABI 7900HT fast real time PCR system (Applied Biosystems, Foster, CA) with SYBR® Green PCR Master Mix (Applied Biosystems, Foster, CA). The PCR profile of *TGFBR1* and 18S rRNA amplification contains: initial denaturation at 95 °C for 10 min, 40 cycles with 95 °C for 15 s, 60 °C for 1 min, with a final cycle of 95 °C 15 s, 60 °C 15 s for dissociation analysis. The threshold cycle (C_T) was detected by the amount of fluorescence reported as dR value. Three replicates were performed for each sample. Standard curves were constructed using purified gel-extracted RT-PCR products by 10-fold serial dilutions in $1{:}10^2$, $1{:}10^3$, $1{:}10^4$, $1{:}10^5$, $1{:}10^6$, $1{:}10^7$, $1{:}10^8$ along with a non-template control. The *TGFBR1* gene expression profiles obtained from the real time Q-PCR were subsequently normalized by the amount of 18S rRNA transcript per reaction to correct the differences in RNA amounts and reverse transcription reactions.

2.4. Cloning and sequencing of full-length porcine TGFBR1 cDNA

5′ and 3′-rapid amplification of cDNA end (RACE) products was generated using the Firstchoice™ RNA ligase-mediated (RLM)-RACE kit (Ambion, Austin, TX) with total liver RNA from an adult Duroc pig. RT-PCR to amplify the 5′-region was performed using nested gene-specific primers: TBR-5outer, TBR-5inner combined with the two supplied 5′-RACE-outer and 5′-RACE-inner primers, respectively. 3′-RACE was conducted using the nested gene-specific primers designated TBR-3outer and TBR-3inner with the counterparts of 3′-RACE-outer and 3′-RACE-inner primers, respectively. Primer pairs TBRRT-CF/CR and TBRRT-DF/DR were used to detect splicing variant throughout the coding region (Table 1). PCR conditions in a PTC-100™ Peltier Thermal Cycler (MJ research, Waltham, MA) were as follows: predenaturation at 95 °C for 15 min, 35 cycles with 95 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, and a final cycle with an extension at 72 °C for 8 min. 5′ and 3′-RACE products were cloned into the pCR®2.1-TOPO® vector system (Invitrogen, Carlsbad, CA).

2.5. SNP and microsatellites identification

SNPs in the porcine *TGFBR1* gene were detected by a direct sequencing approach throughout the *TGFBR1* coding and non-coding regions as well as the 5′ and 3′-flanking regions with DNAs from eight individuals (unrelated grandparents) as template from eight genetically divergent pig breeds containing Large White, Duroc, Hampshire, Landrace, Berkshire, Meishan, Yorkshire and Pietrain. Primers used for SNP detection are shown in Supplementary Materials (Table 1S). Amplifications were performed in 20-μl reactions containing 25 ng genomic DNA, 10 pmol of each primer, 200 μM of each dNTP and 1 U *Taq* DNA polymerase including 1× PCR buffer and 1× Q-solution (Qiagen, Valencia, CA). PCR products were purified and analyzed by 2% agarose gel electrophoresis and then sequenced by ABI 3730 automated DNA sequencer. Sequence comparison for SNP discovery was done by Phrap and Gap4 integration (http://staden.sourceforge.net/phrap.html). Microsatellite markers were identified by web-based programs repeatmasker (http://www.repeatmasker.org/) and Simple Sequence Repeat Identification Tool (SSRIT) (Temnykh et al., 2001) with the minimum number of repeat as five.

3. Results

3.1. Analysis of the genomic organization of the porcine TGFBR1 gene

In total, 75,340 bp of genomic sequence representing 7324 bp of the 5′-flanking region, 62 bp of 5′-UTR region, 1512 bp of coding region, 228 bp of 3′-UTR region, 60,380 bp of non-coding sequences along with 5834 bp of the 3′-flanking

Fig. 2. Sequence alignment of the regulatory regions of the porcine, mouse, rat and human *TGFBR1* gene. A 500-bp-long homologous nucleotide sequences of the 5′-flanking regions of the porcine, mouse, rat and human are aligned. Important putative promoter elements are indicated by boxes including CCAAT-box binding transcription factor (CTF), stimulatory protein 1 (Sp1), upstream control element (UCE.2), GC-binding factor (GCF), activating enhancer binding protein 2 (AP-2), and major late transcription factor (MLTF). These putative transcription factors are also shown proximal to the transcription initiation site.

sequence of the porcine *TGFBR1* gene (GenBank accession no: DQ519377) was obtained. Among the 62,182 bp of the porcine *TGFBR1* genomic sequence, 21,705 bp (34.91%) was interspersed repeats including 8433 bp short interspersed elements (SINEs, 13.56%), 12,505 bp-long interspersed elements (LINEs, 20.11%), 461 bp-long terminal repeat elements (LTR, 0.74%), and 306 bp DNA elements (0.49%). Forty-six SINEs, thirty LINEs, and one LTR repetitive element were identified. Twenty-four SINEs and fourteen LINEs in the first intron, five SINEs and one LINEs in the second intron, nine SINEs and five LINEs in the third intron, one SINE and two LINEs in the fourth intron, three SINEs and four LINEs in the fifth intron, two SINEs and one LINEs in the sixth intron, two SINEs and three LINEs in the eighth intron were detected. A schematic representation of the genomic organization and location of the repetitive elements of the porcine *TGFBR1* gene are illustrated in Fig. 1. The exon–intron boundaries and respective exon and intron sizes are shown in Table 2.

The porcine *TGFBR1* gene contains extremely large introns with 29,441 bp in intron 1 and 13,944 bp in intron 3. The porcine *TGFBR1* core promoter lacks a TATA box, but contains other transcriptional factors to initiate and facilitate the *TGFBR1* transcription including CCAAT-box binding transcription factor (CTF) at position −76 to −82 (AGCCAAT) and −173 to −178 (GATTGG), stimulatory protein 1 (Sp1) at position −91 to −96, −98 to −107, −108 to −113, −150 to −155, −204 to −213, −321 to −326, upstream control element (UCE.2: GGCCG) at position −123 to −127, GC-binding factor (GCF: GCGGGCC) at position −159 to −165, activating enhancer binding protein 2 (AP-2: CCCAGCGG) at position −162 to −169 and −283 to −291, and major late transcription factor (MLTF: CGTGAC) at position −204 to −209 (Fig. 2).

In addition, one CpG island with a high score was detected in the 5′ flanking region of the porcine *TGFBR1* gene. The putative CpG island at position −415 to 769 overlaps with the predicted promoter region and extends 769 bp downstream of

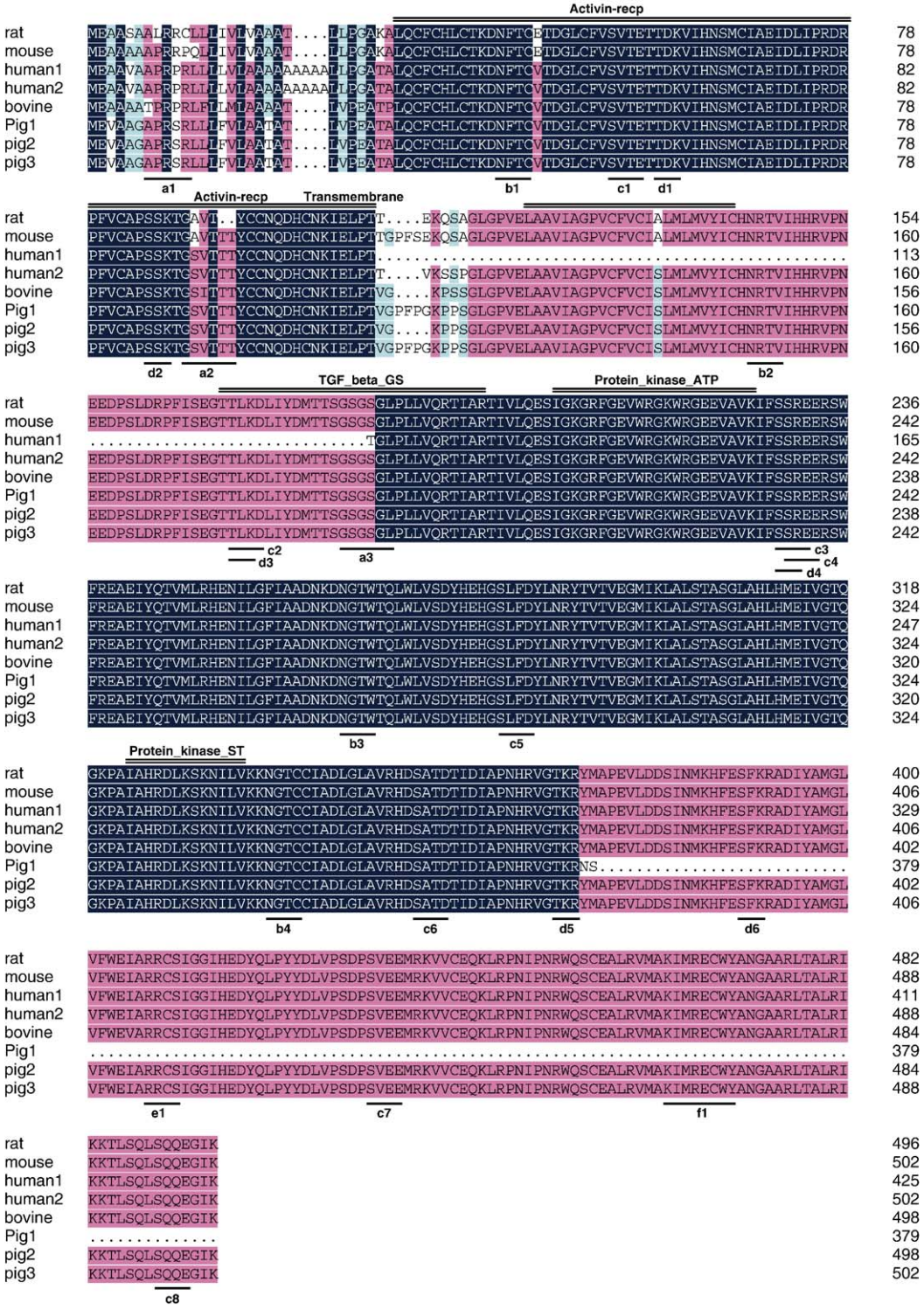

Fig. 3. Multiple alignment of different mammalian TGFBR1 peptide sequences. The deduced amino acid sequences of the porcine *TGFBR1* (pig1, pig2 and pig3, GenBank accession nos: DQ519380, DQ519379, and DQ519378) are shown aligned to those from rat (*Rattus norvegicus*, GenBank accession no: NP_036907), mouse (*Mus musculus*, GenBank accession no: NP_033396), human1 (*Homo sapiens*, GenBank accession no: AAH71181), human2 (*H. sapiens*, GenBank accession no: NP_004603), and bovine (*Bos taurus*, GenBank accession no: NP_777046). The pig2 and pig3 orthologs were also recently described (GenBank accession nos: AB182259 and AB182260). Identical amino acid residues among species are presented with a black background. The protein kinase domain (PS50011) was predicted at positions 201 to 491 of the porcine sequence. The putative domains, motifs and features are double-marked above the sequences, including the activin types I and II receptor domain (PF01064, activin_recp), transmembrane domain, TGF_beta type I GS motif (PF08515, TGF_beta_GS), protein kinase ATP-binding region (PS00107, protein_kinase_ATP), and serine/threonine protein kinase active site (PS00108, protein_kinase_ST). Underlined amino acids indicate patterns with a high probability of occurrence: a = N-myristoylation site (PS00008), b = N-glycosylation site (PS00001), c = casein kinase II phosphorylation site (PS00006), d = protein kinase C phosphorylation site (PS00005), e = cAMP- and cGMP-dependent protein kinase phosphorylation site (PS00004) and f = tyrosine kinase phosphorylation site (PS00007).

the transcription initiation site. The region has a GC content of 73.9%, compared to the average GC content of 39% in the reported 75,340 bp genomic sequences. The overall $CpG_{obs/exp}$ ratio is 0.106, but within the CpG island region at position −415 to 769, was 0.791.

3.2. Analysis of the porcine TGFBR1 cDNA and deduced protein sequence

A modified rapid amplification RACE protocol that allows the isolation of full-length cDNA was used to amplify the *TGFBR1* cDNA from total liver RNA. Repeated sequencing of independent 5′-RACE clones revealed that the transcription initiation point is 62 bp upstream of the start codon. 5′ and 3′-RACE products were cloned to determine the full-length cDNA sequence of 1813 bp (GenBank accession no: DQ519378). The porcine *TGFBR1* cDNA contains a complete ORF of 1512 bp encoding a protein of 503 amino acids with a calculated Mw of 56.4 kDa and a theoretical pI of 7.39. The 125-bp-long exon 7 was alternatively spliced resulting in a shorter ORF of 1140 bp encoding an isoform of 379 amino acids with a calculated Mw of 41.7 kDa and a theoretical pI of 6.70 (GenBank accession no: DQ519379). Multiple transcription initiation sites were detected at 62 bp, 56 bp and 15 bp upstream to the translation start codon (GenBank accession nos: DQ519378, DQ519381, DQ519382), and multiple transcription termination sites were detected at 228 bp and 169 bp after the stop codon (GenBank accession nos: DQ519378, DQ519383). In addition, we also detected a 12 bp alternative splicing event at the beginning of exon 3 of the porcine *TGFBR1* cDNA sequences (GenBank accession no: DQ519380). Twelve out of twenty-six recombinant plasmids with the subcloned RT-PCR products amplified by primer pair TBRRT-CF/CR showed that the 12 bp at the beginning of exon 3 was alternatively spliced out.

The TGFBR1 protein is highly conserved in mammals, and the porcine TGFBR1 protein sequence shows identities of 81.6%, 95.7%, 94.8%, 96.0% and 96.8% to human1 (GenBank accession no: AAH71181), human2 (GenBank accession no: NP_004603), rat (GenBank accession no: NP_036907), mouse (GenBank accession no: NP_033396) and bovine (GenBank accession no: NP_777046). The putative domains, motifs and features of TGFBR1 protein were predicted by ScanProsite, SMART and Pfam search tools. One profile of protein kinase domain (PS50011, residues 201–491), one activin type I and type II receptor domain (PF01064, residues 30–100), one transmembrane domain (residues 126–148), and one TGF beta type I GS motif (PF08515, residues 175–203) were observed in the porcine TGFBR1 protein. Two distinct patterns of one protein kinase ATP-binding region signature (PS00107, residues 211–232), and one serine/threonine protein kinase active site signature (PS00108, residues 329–341) were detected and are absolutely conserved in all mammal orthologs. In addition, twenty-three motifs by six distinct patterns with a high probability of occurrence were predicted including three N-myristoylation site (PS00008), four N-glycosylation site (PS00001), eight casein kinase II phosphorylation site (PS00006), six protein kinase C phosphorylation site (PS00005), one cAMP- and cGMP-dependent protein kinase phosphorylation site (PS00004), and one tyrosine kinase phosphorylation site (PS00007) (Fig. 3).

3.3. SNP and microsatellites identification

Forty-six primer pairs derived from the porcine *TGFBR1* gene were used to amplify 41,044 bp genomic sequences covering the entire coding region, UTR regions and the intronic regions except intron 1 and intron 3 (Table 1S). A total of seventy-seven SNPs and eight indels were detected throughout the *TGFBR1* gene: two regulatory SNPs (SNP_1: A^{-855}-G^{-855} and SNP_2: G^{-415}-C^{-415}) in the putative promoter region detected at the 5′ flanking region, one silent mutation (SNP_{22}) in exon 2, one SNP and one indel (SNP_{84}, SNP_{85}) in the 3′ flanking region, and two nonsynonymous SNPs (SNP_3: C^{84}-T^{84} and SNP_{75}: A^{57615}-G^{57615}) resulting in amino acid substitutions (SNP_3: Pro^8-Ser^8 and SNP_{75}: Val^{417}-Ile^{417}), and the other seventy-eight nucleotide variations detected in intronic sequences (Table 3).

Seventeen microsatellite markers were identified in the porcine *TGFBR1* genomic sequences including one interrupted tri-nucleotide $(CGG)_{12}$ microsatellite spanning in the first exon, ten microsatellites in the first intron, and two microsatellites in the second, third, and fourth introns, respectively. Among these seventeen microsatellites, ten microsatellites were dinucleotide repeat, three microsatellites were tri-nucleotide repeat, three microsatellites were tetra-nucleotide repeat, and one microsatellite was mononucleotide repeat. Number of repeats varied from 5 to 29 (Table 4).

3.4. Expression analysis of porcine TGFBR1 gene

RT-PCR analysis indicated that the porcine *TGFBR1* gene was ubiquitously expressed in the all tissues taken from Duroc adult pigs including fat, adrenal, brain, spinal cord, muscle, mandibular lymph node, thymus, bone marrow, uterus, spleen, testis, kidney, liver and ovary. Quantitative real time PCR results showed the lowest amount of *TGFBR1* expression in kidney, liver and thymus, moderate expression in fat, adrenal, muscle, mandibular lymph node, uterus, spleen and ovary, and the highest expression in brain, spinal cord and testis (Fig. 4).

4. Discussion

In the present study, we present the complete genomic DNA sequence and organization of the porcine *TGFBR1* gene. The analysis of the generated genomic sequence indicated that the porcine *TGFBR1* gene consists of 9 exons and 8 introns which confirms the results known from human, bovine and mouse orthologs (Roelen et al., 1998; Tomoda et al., 1994; Vellucci and Reiss, 1997). The genomic size of the porcine *TGFBR1* gene (62,182 bp) is similar to that of mouse (61,619 bp, Ensembl Gene ID: ENSMUSG 00000007613). In contrast, the genomic sizes of human *TGFBR1* gene (44,896 bp, Ensembl Gene ID: ENSG00000106799), and rat ortholog (53,018 bp, Ensembl Gene ID: ENSRNOG00000007036) are relatively

Table 3
Nucleotide polymorphisms within the porcine *TGFBR1* gene

No.	Position	Variation (M/N)	MAF	Pattern	No.	Position	Variation (M/N)	MAF	Pattern
1	−855	A/G	0.13	P1	44	51,592	T/C	0.38	P11
2	−415	G/C	0.31	P9	45	51,773	A/G	0.13	P1
3	84	T/C	0.31	P9	46	51,811	T/C	0.13	P1
4	1146	G/A	0.31	P9	47	51,867	A/T	0.13	P1
5	1350	C/G	0.31	P9	48	52,067	G/A	0.13	P1
6	1930	T/A	0.13	P1	49	52,142	T/C	0.13	P1
7	2107–8	*/T	0.13	P1	50	52,156	T/*	0.13	P1
8	2238	A/G	0.13	P1	51	52,463	T/C	0.31	P9
9	5189	C/T	0.25	P5	52	52,962–3	**/TA	0.50	P14
10	5483	T/C	0.19	P3	53	53,206	C/T	0.50	P14
11	5492	A/G	0.19	P3	54	53,460	G/A	0.25	P6
12	10,012	G/T	0.31	P9	55	53,498	T/C	0.13	P1
13	10,437	T/C	0.31	P9	56	53,631	G/T	0.13	P1
14	10,495–6	***/AAA	0.25	P6	57	54,013	A/C	0.31	P10
15	10,620	C/T	0.13	P1	58	54,103	T/C	0.31	P10
16	16,442	C/T	0.13	P1	59	54,265	C/A	0.31	P10
17	16,567	C/A	0.31	P9	60	54,274	T/C	0.31	P10
18	16,970	T/G	0.31	P9	61	54,410	A/G	0.31	P10
19	29,122	C/A	0.13	P1	62	54,587	G/A	0.31	P10
20	29,190	A/C	0.13	P1	63	55,736	G/C	0.31	P9
21	29,271	A/G	0.13	P1	64	56,211	G/A	0.19	P4
22	29,662	C/T	0.13	P1	65	56,240	G/A	0.31	P9
23	29,844–5	***/GTA	0.25	P7	66	56,485	A/G	0.31	P9
24	31,646	G/T	0.31	P9	67	56,546	C/T	0.31	P9
25	31,675	T/C	0.31	P9	68	56,901–2	***/CTT	0.31	P9
26	31,791	T/C	0.31	P9	69	57,134	A/G	0.31	P9
27	31,846	A/G	0.31	P9	70	57,277	G/A	0.31	P9
28	32,796	T/C	0.13	P1	71	57,282	T/C	0.31	P9
29	32,804	G/A	0.31	P9	72	57,288	C/T	0.31	P9
30	34,704	G/T	0.50	P13	73	57,290	T/C	0.31	P9
31	34,714	T/A	0.25	P8	74	57,407	C/A	0.25	P6
32	46,483	C/G	0.13	P1	75	57,615	A/G	0.13	P1
33	46,815	C/T	0.13	P1	76	57,653	G/T	0.31	P9
34	46,840	T/G	0.13	P1	77	58,866	A/G	0.31	P9
35	47,814	C/T	0.50	P14	78	59,229	T/C	0.31	P9
36	48,272	T/A	0.25	P6	79	59,340	G/A	0.44	P12
37	48,685	A/G	0.31	P9	80	59,621	C/T	0.13	P2
38	49,162	C/T	0.25	P6	81	59,868–9	AA/**	0.31	P9
39	49,255	G/C	0.31	P9	82	60,594	G/T	0.13	P1
40	49,497	T/C	0.31	P9	83	61,005	A/G	0.31	P9
41	49,522	T/C	0.31	P9	84	62,239	T/*	0.31	P9
42	50,844	G/A	0.38	P11	85	62,351	T/A	0.31	P9
43	51,499	G/A	0.38	P11					

+1 corresponds to the transcription initiation point of the longest porcine *TGFBR1* cDNA.
M = Major allele, N = Minor allele, MAF = Minor Allele Frequency.
SNP_3 (T^{84}-C^{84}) leading to amino acid exchange P^8-S^8 and SNP_{75} (A^{57615}-G^{57615}) leading to amino acid exchange V^{417}-I^{417}.
P1 = LW(MM), DU(MM), HA(MM), LA(MM), BE(NN), MS(MM); YO(MM), PI(MM); P2 = LW(MN), DU(MN), HA(MM), LA(MM), BE(MM), MS(MM), YO (MM), PI(MM); P3 = LW(MM), DU(MM), HA(MM), LA(MM), BE(NN), MS(MN); YO(MM), PI(MM); P4 = LW(MN), DU(MN), HA(MM), LA(MM), BE(MM), MS(MN), YO(MM), PI(MM); P5 = LW(MM), DU(NN), HA(NN), LA(MM), BE(MM), MS(MM); YO(MM), PI(MM); P6 = LW(MN), DU(MN), HA(MM), LA (MM), BE(NN), MS(MM), YO(MM), PI(MM); P7 = LW(MN), DU(MM), HA(MM), LA(MM), BE(NN), MS(MN); YO(MM), PI(MM); P8 = LW(MM), DU(MN), HA(MM), LA(MM), BE(NN), MS(MN), YO(MM), PI(MM); P9 = LW(MN), DU(MN), HA(MM), LA(MM), BE(NN), MS(MN); YO(MM), PI(MM); P10 = LW (NN), DU(MN), HA(MM), LA(MM), BE(NN), MS(MM), YO(MM), PI(MM); P11 = LW(NN), DU(MN), HA(MM), LA(MM), BE(NN), MS(MN); YO(MM), PI (MM); P12 = LW(MN), DU(MN), HA(MM), LA(MM), BE(NN), MS(MN), YO(MN), PI(MN); P13 = LW(NN), DU(NN), HA(NN), LA(MM), BE(NN), MS(MM); YO(MM), PI(MM); P14 = LW(NN), DU(NN), HA(MM), LA(MM), BE(NN), MS(NN), YO(MM), PI(MM).
LW = Large White, DU = Duroc, HA = Hampshire, LA = Landrace, BE = Berkshire, MS = Meishan, YO = Yorkshire, PI = Pietrain.

smaller. All *TGFBR1* genes have an extremely large intron 1 in mammals with 29,441 bp in pig, 23,334 bp in human, 30,352 bp in mouse, and 31,627 in rat, respectively. However, the first introns have less than 30% similarity between species except that mouse *TGFBR1* shares 82% identity with rat. The other non-coding regions of porcine *TGFBR1* gene also have very low similarity to human, mouse and rat orthologs, except that the 3′ flanking region (4500 bp downstream the stop codon) shares 80%, 65%, 64% identity with those of human, mouse and rat, respectively. The porcine *TGFBR1* gene has a relatively

short 3′-UTR (228 bp) compared with its orthologs in human (4887 bp, GenBank accession no: NM_004612), mouse (4216 bp, GenBank accession no: NM_009370), and rat (4160 bp, GenBank accession no: BC087035). The porcine *TGFBR1* cDNA contains a consensus polyadenylation signal (ATTAAA) in the 3′-UTR, and two additional polyadenylation signals (AATAAA) were observed 1586 nucleotides and 2806 nucleotides downstream of the transcription termination site. Therefore, the possibility of different lengths of 3′-UTR in porcine *TGFBR1* gene could not be ruled out in other tissues by using alternative polyadenylation signals. In addition, all of the exon–intron junctions follow the GT-AG rule for splice-donor and splice-acceptor sites (Table 2) which is also identified in other mammalian *TGFBR1* (Roelen et al., 1998; Tomoda et al., 1994; Vellucci and Reiss, 1997).

The core promoter of the porcine *TGFBR1* gene lacks a TATA box but contains a variety of other elements which demonstrate to provide promoter function. Surveys of promoter sequences suggest that 50% or more of promoters may be TATA-less (Lewin, 2004). The core promoter of porcine *TGFBR1* gene contains a CAAT box (CCAAT) at position −76 to −80, which plays a strong role in determining the efficiency of the promoter, but does not influence its specificity. Another common promoter element is GC boxes at positions −91 to −96, −98 to −107, −108 to −113, −150 to −155, −204 to −213 and −321 to −326. The GC boxes are often present in the promoter as multiple copies. The common promoter components were shown as highly conserved by multiple alignment of the promoter region in pig, human, mouse and rat *TGFBR1* genes (Fig. 2). The promoter of *TGFBR1* gene lacks a TATA box which is a crucial positioning element of the core promoter, therefore, the transcription initiation site altered at different locations. According to the 5′ and 3′-RACE PCR results, we cloned different fragment sizes of the 5′-UTR regions ranging from 15 to 62 bp suggesting that the porcine *TGFBR1* gene initiates the transcription at multiple sites. In comparison, the 5′-UTR region is 76 bp in human (GenBank accession no: NM_004612), 39 bp in rat (GenBank accession no: L26110) and as short as 7 bp in mouse (GenBank accession no: NM_009370), thus we conclude that divergent transcriptional regulation mechanisms exist for the *TGFBR1* gene in mammals.

Table 4
Microsatellite markers identified in the porcine *TGFBR1* gene

No.	Repeat motif	No. of repeats	Exon/ intron	Position in repeat	
				Beginning	End
1	CGG	12	Exon 1	113	195
2	TA	5	Intron 1	5864	5873
3	CA	21	Intron 1	12,863	12,904
4	TG	8	Intron 1	13,572	13,598
5	GA	5	Intron 1	17,127	17,136
6	TA	7	Intron 1	18,290	18,303
7	TG	6	Intron 1	18,498	18,509
8	TCCT	5	Intron 1	19,695	19,714
9	TCCC	9	Intron 1	19,731	19,766
10	TTCT	16	Intron 1	19,767	19,830
11	TG	13	Intron 1	28,408	28,437
12	TTA	5	Intron 2	30,591	30,616
13	CA	11	Intron 2	33,108	33,129
14	TTC	6	Intron 3	33,917	33,941
15	T	27	Intron 3	34,688	34,714
16	TG	11	Intron 4	48,537	48,596
17	TA	29	Intron 4	50,176	50,234

+1 corresponds to the transcription initiation point of the longest porcine *TGFBR1* cDNA.

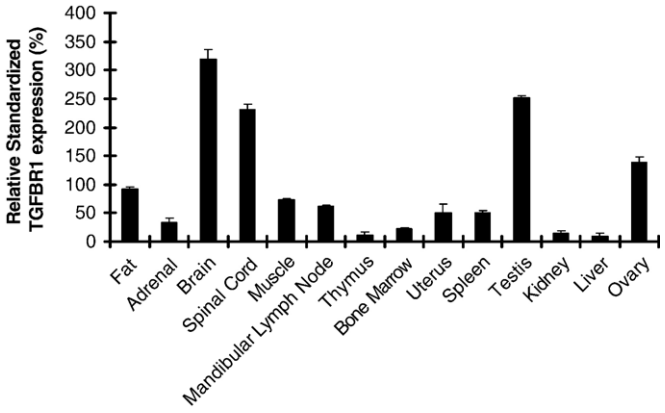

Fig. 4. Analysis of the porcine *TGFBR1* gene expression in different tissues using real time RT-PCR. Quantitative real time RT-PCR was performed using RNAs from different tissues including fat, adrenal, brain, spinal cord, muscle, mandibular lymph node, thymus, bone marrow, uterus, spleen, testis, kidney, liver, and ovary. mRNA levels of *TGFBR1* are normalized by 18S rRNA levels and indicated in the respective porcine tissues.

The porcine TGFBR1 peptide and the so far characterized mammalian TGFBR1 proteins including human, bovine, mouse and rat are highly conserved, and show high homologies to each other (*Sus scrofa*/*Homo sapiens*: 95.6%, *S. scrofa*/*Rattus norvegicus*: 94.8%, *S. scrofa*/*Mus musculus*: 96.0% and *S. scrofa*/*Bos taurus*: 96.8%; Fig. 3). All TGFBR1 of these five species encompass identical protein kinase domain profile with a length of 291 amino acids, a protein kinase ATP-binding region signature of 22 amino acids, and a serine/threonine protein kinase active site signature of 13 amino acids except the pig shorter isoform (pig1) of 179 amino acids. Multiple alignment revealed that activin types I and II receptor domain of 81 amino acids, transmembrane domain of 23 amino acids, and TGF beta type I GS motif of 29 amino acids, is highly conserved among all mammals. Activin types I and II receptor domain is a hydrophilic cysteine-rich ligand binding domain that contains an identical conserved 9-amino-acid cysteine box with the consensus of $CCX_{[4-5]}CN$ in mammals. The TGF beta type I GS motif has a highly conserved GSGSGLP signature in the cytoplasmic juxtamembrane region immediately preceding the protein's kinase domain. Point mutations in the GS motif modify the signaling activity of TGFBR1 (Feng and Derynck, 1997).

The porcine *TGFBR1* cDNA contains a complete ORF of 1512 bp encoding a predicted protein of 503 amino acids (Shimanuki et al., 2005). A single splice variant involving the removal of the first 12 nucleotides of exon 3 has also been reported (Shimanuki et al., 2005). We have confirmed the

presence of both the full-length transcript as well as the alternative splicing of exon 3. However, we have identified an additional splice variant missing the 125-bp-long exon 7 that results in a shorter isoform of 379 amino acids. This alternative splicing event not found by Shimanuki et al. could possibly arise since variant splice patterns are often specific to different stages of development, particular tissues or a disease state (Thanaraj et al., 2004). This shorter isoform contains protein kinase domain profile with only 175 amino acids, and shares relatively low homology of 75% with the short isoform of TGFBR1 peptide in human where exon 3 is alternatively spliced. It was estimated that one-to-two thirds of mammalian genes are alternatively spliced (Brett et al., 2002; Modrek and Lee, 2003), and the mammalian alternative splicing events frequently arise from the evolutionarily rapid loss or gain of exons from genomes (Modrek and Lee, 2003; Thanaraj et al., 2003; Nurtdinov et al., 2003; Sorek et al., 2004). Alternative splicing is one of the most important mechanisms regulating gene expression in multicellular organisms. The effects of alternative splicing range from a complete loss of function or acquisition of a new function to very subtle modulations, which are observed in the majority of cases reported such as binding properties, enzymatic activity, intracellular localization, protein stability, phosphorylation and glycosylation patterns (Stamm et al., 2005). Utilizing highly predictive computational method, over 11% of human and mouse alternative exons was estimated to be species-specific alternative splicing events (Pan et al., 2005).

TGFBR1 mRNA was expressed in the all porcine tissues tested using quantitative real time RT-PCR, being most abundant in brain, spinal cord and testis, and least abundant in liver, kidney and thymus. The expression patterns of *TGFBR1* are consistent with those in other species except for brain tissues (Franzen et al., 1993; Roelen et al., 1998; Tomoda et al., 1994). In human, mouse and bovine, the lowest level of gene expression was detected in brain by northern blot hybridization. The discrepancy of mRNA in brain tissues between pigs and other species might indicate a species-specific expression of *TGFBR1*, but it may also be due to differences in sensitivity of the detection methods. Moreover, RT-PCR results revealed that the shorter *TGFBR1* isoform was expressed ubiquitously in pig tissues, but less abundant than the longer ones (data not shown). Six out of twenty-six recombinant plasmids with subcloned RT-PCR products generated by primer pair TBRRT-DF/DR showing the alternative splicing of exon 7 also supported lower expression of the shorter isoform.

A total of 77 SNPs and 8 indels were detected in the porcine *TGFBR1* gene by comparative sequencing of genetically divergent pig breeds. The minor allele frequencies of these nucleotide variations varied from 0.13 to 0.5 with an average of 0.26. Most polymorphisms were detected in Berkshire and Meishan pigs as 79 out of these 85 nucleotide variations possess different alleles between them. Nonsynonymous SNPs at positions 84 and 57,615 lead to amino acid substitutions (Pro^{8}-Ser^{8} and Val^{417}-Ile^{417}) which were also identified recently in pig (Shimanuki et al., 2005). The nucleotide variations were mostly found in the intronic regions. In the human *TGFBR1* gene, both coding and non-coding mutations were correlated with relevant developmental diseases (Bian et al., 2005; Chen et al., 2004, 2006; Pasche et al., 2005).

Human genomic DNA sequences contain an average of one microsatellite every 2 to 4 kb (Brown, 2002). The 62 kb porcine *TGFBR1* genomic sequence contained 17 microsatellites with an average distance of 3.6 kb per microsatellite, that is in agreement with microsatellite occurrence in the human genome. The microsatellite markers with large number of repeat motifs are supposedly polymorphic and hold the potential for association analysis with economically important traits in pig. Six microsatellite repeat motifs located in the regions for SNP discovery (microsatellites 12–17), were observed informative in the SNP discovery panel DNAs (data not shown). The tri-nucleotide microsatellite marker at positions 113 to 195 partly located in the first exon could possibly lead to amino acid indel, however, we did not detect polymorphisms at this locus among the SNP discovery panel DNAs.

Among 161 QTL regions detected in pig chromosome one, the porcine *TGFBR1* gene was associated with approximately 50 putative QTL regions (Hu et al., 2005). With respect to its physiological function in growth and development, and its position in the QTL-rich region, microsatellite markers and SNPs as well as indels identified in the porcine *TGFBR1* gene could possibly contribute to dissect these QTLs or be the candidate gene/QTL itself and should thus be investigated in respective pedigrees.

Acknowledgements

This work was supported by grants from the USDA–National Research Initiative (Grant Nos. 2004-35205-14187 and 2005-4480-15939) and the USDA–Agricultural Research Service (Agreement No. 58-5438-2-313). We would like to thank the W.M. Keck Center for Comparative and Functional Genomics at UIUC for their assistance in sequencing and bioinformatics. Tissue samples for transcript analysis were provided by Dr. M. Rogatcheva.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.07.009.

References

Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bian, Y., et al., 2005. TGFBR1*6A may contribute to hereditary colorectal cancer. J. Clin. Oncol. 23, 3074–3078.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P., 2002. Alternative splicing and genome complexity. Nat. Genet. 30, 29–30.

Brown, T.A., 2002. Genome 2nd Edition: the Human Genome. Taylor and Francis Group, London (UK).

Chang, H., Brown, C.W., Matzuk, M.M., 2002. Genetic analysis of the mammalian transforming growth factor-beta superfamily. Endocr. Rev. 23, 787–823.

Chen, T., et al., 2004. An intronic variant of the TGFBR1 gene is associated with carcinomas of the kidney and bladder. Int. J. Cancer 112, 420–425.

Chen, T., et al., 2006. Int7G24A variant of transforming growth factor-beta receptor type I is associated with invasive breast cancer. Clin. Cancer Res. 12, 392–397.

Feng, X.H., Derynck, R., 1997. A kinase subdomain of transforming growth factor-beta (TGF-beta) type I receptor determines the TGF-beta intracellular signaling specificity. EMBO J. 16, 3912–3923.

Franzen, P., et al., 1993. Cloning of a TGF beta type I receptor that forms a heteromeric complex with the TGF beta type II receptor. Cell 75, 681–692.

Hu, Z.L., et al., 2005. A QTL resource and comparison tool for pigs: PigQTLDB. Mamm. Genome 16, 792–800.

Johnson, D.W., Qumsiyeh, M., Benkhalifa, M., Marchuk, D.A., 1995. Assignment of human transforming growth factor-beta type I and type III receptor genes (TGFBR1 and TGFBR3) to 9q33-q34 and 1p32-p33, respectively. Genomics 28, 356–357.

Kingsley, D.M., 1994. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev. 8, 133–146.

Kopecny, M., Stratil, A., Van Poucke, M., Bartenschlager, H., Geldermann, H., Peelman, L.J., 2004. PCR-RFLPs, linkage and RH mapping of the porcine TGFB1 and TGFBR1 genes. Anim. Genet. 35, 253–255.

Laiho, M., Weis, M.B., Massague, J., 1990. Concomitant loss of transforming growth factor (TGF)-beta receptor types I and II in TGF-beta-resistant cell mutants implicates both receptor types in signal transduction. J. Biol. Chem. 265, 18518–18524.

Lewin, B., 2004. Gene VIII. Pearson Education, Inc., New Jersey (USA), pp. 597–631.

Loeys, B.L., et al., 2005. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat. Genet. 37, 275–281.

Lopez-Casillas, F., Cheifetz, S., Doody, J., Andres, J.L., Lane, W.S., Massague, J., 1991. Structure and expression of the membrane proteoglycan betaglycan, a component of the TGF-beta receptor system. Cell 67, 785–795.

Massague, J., 1990. The transforming growth factor-beta family. Annu. Rev. Cell Biol. 6, 597–641.

Massague, J., 1992. Receptors for the TGF-beta family. Cell 69, 1067–1070.

Massague, J., Blain, S.W., Lo, R.S., 2000. TGFbeta signaling in growth control, cancer, and heritable disorders. Cell 103, 295–309.

Modrek, B., Lee, C.J., 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat. Genet. 34, 177–180.

Moses, H.L., Serra, R., 1996. Regulation of differentiation by TGF-beta. Curr. Opin. Genet. Dev. 6, 581–586.

Moses, H.L., Yang, E.Y., Pietenpol, J.A., 1990. TGF-beta stimulation and inhibition of cell proliferation: new mechanistic insights. Cell 63, 245–247.

Nurtdinov, R.N., Artamonova, I.I., Mironov, A.A., Gelfand, M.S., 2003. Low conservation of alternative splicing patterns in the human and mouse genomes. Hum. Mol. Genet. 12, 1313–1320.

Pan, Q., et al., 2005. Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet. 21, 73–77.

Pasche, B., et al., 1998. Type I transforming growth factor beta receptor maps to 9q22 and exhibits a polymorphism and a rare variant within a polyalanine tract. Cancer Res. 58, 2727–2732.

Pasche, B., et al., 2005. Somatic acquisition and signaling of TGFBR1*6A in cancer. JAMA 294, 1634–1646.

Roelen, B.A., et al., 1998. Molecular cloning, genetic mapping, and developmental expression of a bovine transforming growth factor beta (TGF-beta) type I receptor. Mol. Reprod. Dev. 49, 1–9.

Segarini, P.R., Rosen, D.M., Seyedin, S.M., 1989. Binding of transforming growth factor-beta to cell surface proteins varies with cell type. Mol. Endocrinol. 3, 261–272.

Shimanuki, S., Mikawa, A., Miyake, Y., Hamasima, N., Mikawa, S., Awata, T., 2005. Structure and polymorphism analysis of transforming growth factor beta receptor 1 (TGFBR1) in pigs. Biochem. Genet. 43, 491–500.

Sorek, R., Shamir, R., Ast, G., 2004. How prevalent is functional alternative splicing in the human genome? Trends Genet. 20, 68–71.

Stamm, S., et al., 2005. Function of alternative splicing. Gene 344, 1–20.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., McCouch, S., 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11, 1441–1452.

Thanaraj, T.A., Clark, F., Muilu, J., 2003. Conservation of human alternative splice events in mouse. Nucleic Acids Res. 31, 2544–2552.

Thanaraj, T.A., Stamm, S., Clark, F., Riethoven, J.J., Le Texier, V., Muilu, J., 2004. ASD: the alternative splicing database. Nucleic Acids Res. 32, D64–D69.

Tomoda, T., Kudoh, T., Noma, T., Nakazawa, A., Muramatsu, M., Arai, K., 1994. Molecular cloning of a mouse counterpart for human TGF-beta type I receptor. Biochem. Biophys. Res. Commun. 198, 1054–1062.

Vellucci, V.F., Reiss, M., 1997. Cloning and genomic organization of the human transforming growth factor-beta type I receptor gene. Genomics 46, 278–283.

Wang, X.F., Lin, H.Y., Ng-Eaton, E., Downward, J., Lodish, H.F., Weinberg, R.A., 1991. Expression cloning and characterization of the TGF-beta type III receptor. Cell 67, 797–805.

Wrana, J.L., Attisano, L., Wieser, R., Ventura, F., Massague, J., 1994. Mechanism of activation of the TGF-beta receptor. Nature 370, 341–347.