

Promoter identification and analysis of key glycosphingolipid biosynthesis-globo series pathway genes in piglets

W.Y. Qin¹, L.N. Gan¹, R.W. Xia¹, W.H. Dong¹, S.Y. Sun¹, G.Q. Zhu², S.L. Wu¹ and W.B. Bao¹

¹Key Laboratory for Animal Genetics, Breeding, Reproduction and Molecular Design of Jiangsu Province, College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu, China

²College of Veterinary Medicine, Yangzhou University, Yangzhou, Jiangsu, China

Corresponding author: W.B. Bao
E-mail: wbbao@yzu.edu.cn

Genet. Mol. Res. 16 (2): gmr16029574

Received December 9, 2016

Accepted May 15, 2017

Published June 28, 2017

DOI <http://dx.doi.org/10.4238/gmr16029574>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. Glycosphingolipid biosynthesis-globo series pathway genes (*FUT1*, *FUT2*, *ST3GALI*, *HEXA*, *HEXB*, *B3GALNT1*, and *NAGA*) play an important regulatory role in the defense against *Escherichia coli* F18 in piglets. In this study, we identified the transcription initiation site and promoter of this gene cluster by mined previous RNA-seq results using bioinformatics tools. The *FUT1* transcription initiation region included five alternative splicing sites and two promoter regions, whereas each of the six other genes had one promoter. Dual luciferase reporter results revealed significantly higher transcriptional activity by *FUT1* promoter 2, indicating that it played a more important role in transcription. The promoters of glycosphingolipid biosynthesis genes identified contained a CpG island within the first 500 bp, except for the *B3GALNT1* promoter which included fewer CpG sites. These

results provide a deeper insight into methylation and the regulatory mechanisms of glycosphingolipid biosynthesis-globo series pathway genes in piglets.

Key words: Piglet; Glycosphingolipid biosynthesis-globo series pathway; *E. coli* F18; Promoter region; CpG island

INTRODUCTION

Porcine post-weaning diarrhea (PWD), primarily caused by *Escherichia coli* F18, is one of the worst diseases affecting the pork industry in recent years (Boldin, 2008). *E. coli* F18 adheres to the intestinal epithelial cells through interactions between the bacterial pili and receptors on the brush border of small intestinal epithelial cells (Bertin and Duchet-Suchaux, 1991). This interaction triggers the production of bacterial enterotoxins that cause diarrhea. Therefore, the pathogenesis is dependent on the expression of the required mammalian protein receptors (Boldin, 2008; Li et al., 2012). Glycosphingolipids are important components of the cell membrane. These compounds facilitate cell adhesion, growth, proliferation and differentiation, signal transduction, and other processes (Hoetzel et al., 2007; Lahiri and Futerman, 2007). The four types of glycosphingolipids (ganglio-, globo/isoglobo-, lacto/neolacto-, and gala/neogala- series glycosphingolipids) have different monosaccharide components (Inagaki, 2008). Previously, we used two-color microarrays to analyze differential gene expression in duodenal tissues collected from full-sib pairs of Suta pigs (Duroc × Meishan) with both resistant and sensitive adhesion phenotypes. This analysis identified seven key glycosyltransferase genes [α (1,2)fucosyltransferase 1 (*FUT1*); α (1,2)fucosyltransferase 2 (*FUT2*); ST3 β -galactosidase α -2,3-sialyltransferase 1 (*ST3GAL1*); β -N-acetylgalactosamine A (*HEXA*); β -N-acetylgalactosamine B (*HEXB*); β -1,3-N-acetylgalactosaminyltransferase 1 (*B3GALNT1*); alpha-N-acetylgalactosaminidase (*NAGA*)] from the glycosphingolipid biosynthesis-globo series pathway. This pathway might play important regulatory roles against *E. coli* F18 infection in weaning piglets (Bao et al., 2012). RNA polymerase was specifically bound to DNA in the promoter region of the identified gene cluster, indicating that this region may regulate gene expression (Christensen et al., 2004; Zhang et al., 2015). CpG methylation of promoter regions can affect chromatin structure, DNA conformation, stability, and DNA-protein interactions, thereby regulating transcription (Bender, 2004; Li et al., 2014). Although the complete porcine genome sequence has been determined and deposited in GenBank, transcription initiation sites and promoter regions of this and many other gene clusters remain unclear. In order to identify the promoter region, we compared all the seven porcine gene sequences mentioned above with their human orthologs, and mined previous RNA-seq results of Meishan pigs using Integrative Genomics Viewer (IGV). Additionally, dual luciferase reporter gene technology and bioinformatics were applied to investigate the transcriptional activity of multiple promoter regions and CpG islands, respectively.

MATERIAL AND METHODS

Ethics statement

The study proposal was approved by the Institutional Animal Care and Use Committee

(IACUC) of the Yangzhou University Animal Experiments Ethics Committee [permit No.: SYXK (Su) IACUC 2012-0029]. All experimental procedures involving piglets were performed in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals approved by the State Council of People's Republic of China.

Experimental animals and sample collection

Meishan pigs were obtained from the Meishan Pigs Breeding Conservation Co., Ltd. (Kunshan, China). All the piglets were maintained under same piggery conditions, the environmental temperature was controlled at 25°-28°C, they were fed *ad libitum* with a commercial-type compound feed for weaned piglets containing 21.7% crude protein, and free from antimicrobial additives and organic acids. No additional food was provided to the experimental animals. This experiment was conducted in the Key Laboratory for Animal Genetics, Breeding, Reproduction and Molecular Design of Yangzhou University. Piglets from healthy full-sib populations, with similar birth weight, weaning weight, and body shape and color, aged 35 days, were used in the present study. Finally, six piglets were selected and sacrificed by electrical stunning; approximately 1 g ear tissue was collected in 1.5-mL nuclease-free Eppendorf tubes, frozen immediately in liquid nitrogen, and stored at -80°C until further study. Genomic DNA was extracted from ear notches according to a modified phenol/chloroform method (Sambrook et al., 1989). The quality and integrity of DNA were determined using the Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The pGL3-basic and pGL3-control vectors were purchased from Invitrogen (Carlsbad, CA, USA). Intestinal epithelial cells (IPEC-J2) were given by Professor Schifferli (College of Veterinary Medicine, University of Pennsylvania, USA).

Bioinformatic analysis

Comparison of porcine and human variable promoter regions

The promoter regions of the seven genes were compared using Transcriptome Viewer software (<http://www.genomatix.de/>) (Scherf et al., 2000).

Determination of promoter regions using transcriptome sequencing

Transcriptome sequencing results (GenBank accession No. PRJNA271310) were mined using IGV to identify the transcription initiation sites and promoters of all the seven genes.

Determination of promoter CpG islands and prediction of putative transcription factor binding sites (TFBSs)

Promoter CpG islands within a 2-kb region, upstream of the seven-gene cluster, were identified and analyzed using MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) (Li and Dahiya, 2002), and putative TFBSs were predicted using Alibaba (<http://www.gene-regulation.com/pub/programs/alibaba2/index.html>) (Grabe, 2002). The following conditions were applied: pairs in known sites: 50; mat. width in bp: 10; min. number of sites: 4; min. mat. conservation: 80% (high); sim. of seq to mat.: 1%; factor class level: 4.

Transcriptional activity of *FUT1* promoters 1 and 2

To measure the transcriptional activity of *FUT1* promoters, we performed polymerase chain reaction (PCR). Each 20- μ L reaction volume contained 100 ng template DNA, 1 μ L each of the primers (10 pM), 10 μ L PCR Master Mix (Tiangen Biotech, Beijing, China), and sterilized distilled water. The primers, synthesized by Invitrogen Biotechnology Co., Ltd., Shanghai, China, are shown in Table 1. The PCR conditions were as follows: denaturation at 95°C for 5 min, 35 cycles of 95°C for 30 s, 63°C for 30 s, and 72°C for 50 s, followed by a final extension at 72°C for 10 min. The PCR products were stored at 4°C and electrophoresed on 1% agarose gels.

Table 1. Primers used for *FUT1* promoters.

Region	Primer (5'→3')	Length (bp)	Position (from translation start site)
Promoter 1	F: ATTCACAGCTGACAGGAACACC	1138	-1633-2770 bp
	R: AAGGTCCAATCGGAGCTGCG		
Promoter 2	F: GGAAGGCTCGCATCCGCTT	636	-580-1215 bp
	R: CAGATTCCAGTCATAAGTGCAGC		

RESULTS

Characterization of the gene cluster

The results of identification and characterization of seven porcine glycosphingolipid biosynthesis-globo series pathway genes are shown in Table 2. *FUT1* gene and *FUT2* gene on the same chromosome, and the position of *FUT1* on chromosome is extremely close to *FUT2*, *FUT1* and *NAGA* are shown in the reverse direction (genetic code from right to left), whereas the rest are in the forward direction.

Table 2. Porcine glycosphingolipid biosynthesis-globo series gene information.

Gene	Chromosome	Position	Transcriptional orientation
<i>FUT1</i>	6	49824783-49827868	-
<i>FUT2</i>	6	49781887-49791227	+
<i>ST3GAL1</i>	4	7703297-7776990	+
<i>HEXA</i>	7	65420490-65455601	+
<i>HEXB</i>	2	85331775-85372032	+
<i>B3GALNT1</i>	13	109170833-109175561	+
<i>NAGA</i>	5	3970212-3977625	-

Comparison of porcine and human variable promoters

A comparison with human promoter sequences from the human database revealed that all the seven porcine genes include several alternative splice sites (Figure 1). Further, human *FUT1* and *B3GALNT1* include an alternative promoter (Figure 1), and we predict the same for the porcine genes.

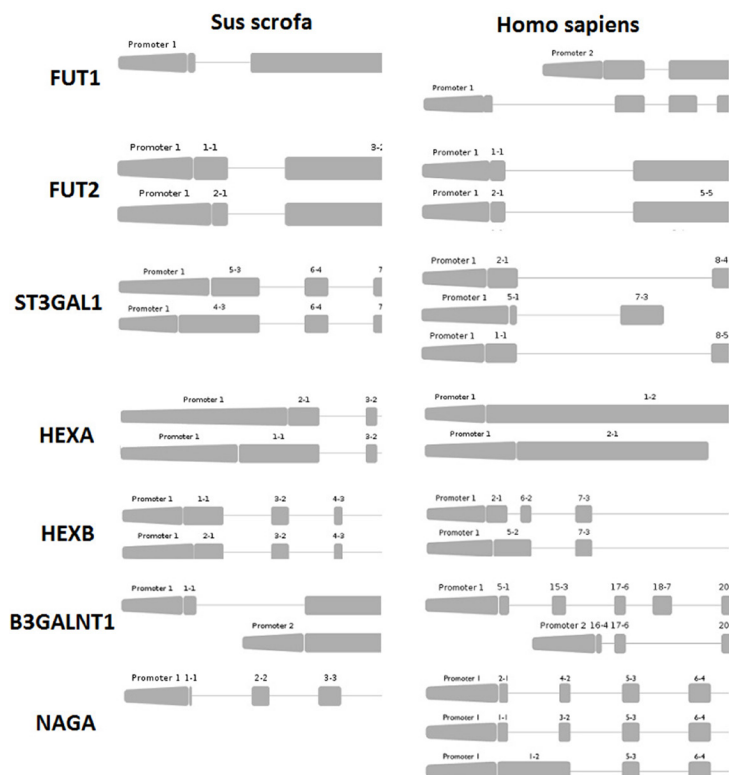


Figure 1. Comparison of porcine and human glycosphingolipid biosynthesis-globo series pathway genes. The left side of the gray strips: promoter; the gray bars on the right: exons; straight line: intron.

Analysis of promoters and transcriptome sequencing

Except *FUT1* and *B3GALNT1*, the Meishan pig duodenal transcriptome sequences revealed the locations of exons in five genes, and expression levels and transcription start sites were in accordance with the porcine genome database [GenBank accession No.: *FUT2* (U70881.2), *ST3GAL1* (M97753.1), *HEXA* (EU442572.1), *HEXB* (AB529531.1), and *NAGA* (EU442573.1)]. The *FUT1* (GenBank accession No.: U70883) gene included five alternative splicing sites and two promoter regions, whereas all the other genes had only one promoter. Moreover, the *B3GALNT1* (GenBank accession No.: NM_214351.1) exon was incomplete in the porcine genome database; however, the new upstream exon identified, which included the promoter, is shown in Figure 2.

Identification of CpG islands

MethPrimer showed that both the *FUT1* promoters contained multiple CpG islands, whereas *FUT2*, *ST3GAL1*, *HEXA*, *HEXB*, and *NAGA* promoters included a single CpG island within a 500-bp region. The absence of CpG islands, and presence of fewer CpG sites in the *B3GALNT1* promoter than that in the other genes, is shown in Figure 3.

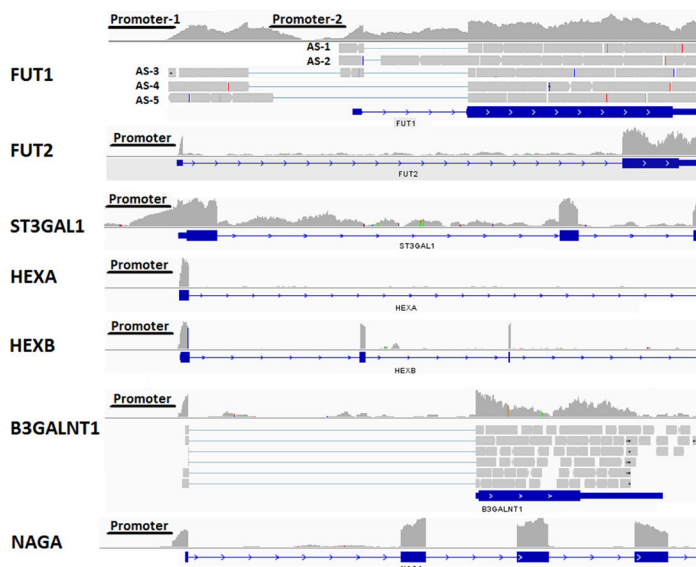


Figure 2. Analysis of promoter regions using Integrative Genomics Viewer (IGV). The peaks represent different expression levels from the different exon regions. The long blue bars represent genome sequence information from the pig genome database. The gray bars indicate transcriptome sequencing reads after combining.

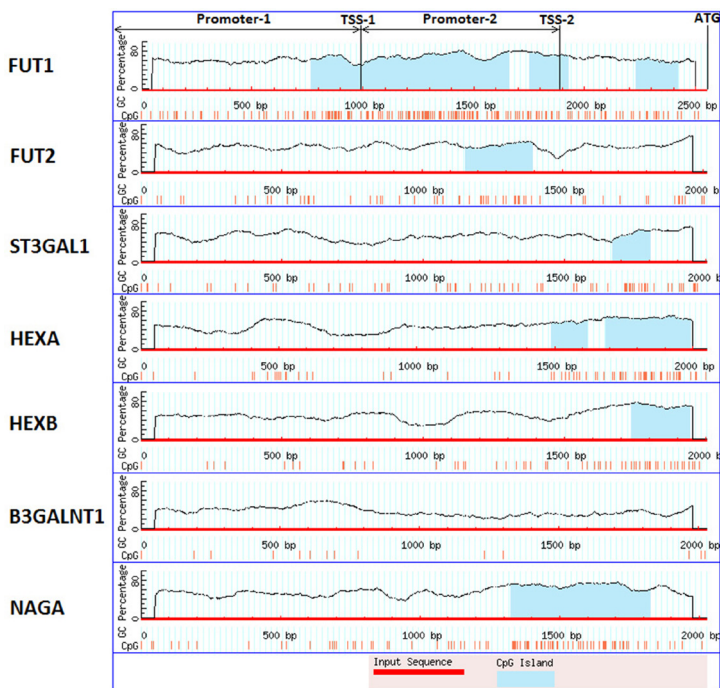


Figure 3. Detection of CpG islands in the promoter regions. Criteria used: Island size > 100, GC Percent > 50.0, Obs/Exp > 0.6.

Analysis of CpG islands

Analysis of the six identified CpG islands using Alibaba version 2 revealed 20, 11, 14, 18, 23, and 18 TFBSs in the *FUT1*, *FUT2*, *HEXA*, *HEXB*, *NAGA*, and *ST3GAL1* CpG islands, respectively (Figures S1 to S6).

Dual luciferase analysis of *FUT1* promoters

IPEC-J2 cells were transfected with recombinant plasmids pGL3-P1 and pGL3-P2 for 24 h (Zi et al., 2013), and dual luciferase assays were performed. The products were subjected to electrophoresis on a 1% agarose gel. Figure 4 shows that the size of the amplified fragments corresponded with the expected PCR product size. These results suggested the transcriptional activity of promoter 2 was 2.75-fold higher than that of promoter 1 ($P < 0.01$) (Figure 5).

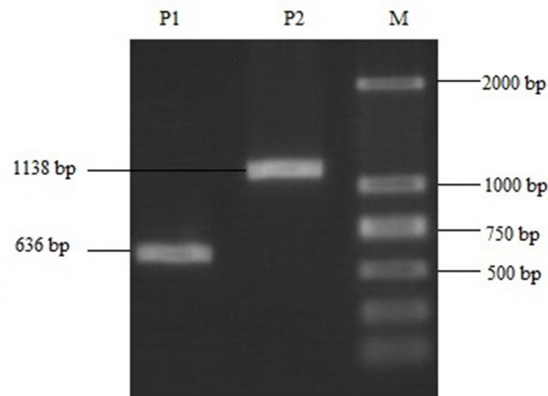


Figure 4. Agarose gel electrophoresis of PCR products obtained by the amplification of *FUT1* promoters 1 and 2. M: DL200 marker; P1: promoter 1; P2: Promoter 2.

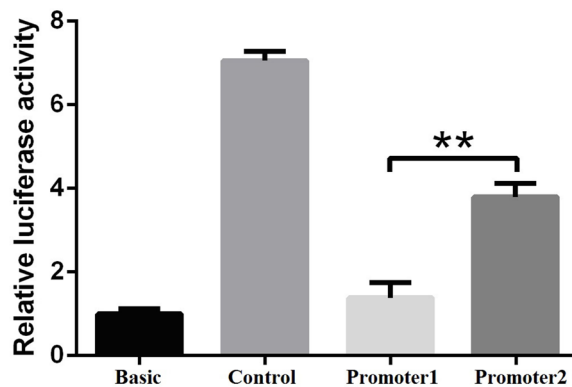


Figure 5. Comparison of the relative transcriptional activity of *FUT1* promoters 1 and 2. Basic: pGL3-basic; Control: pGL3-control; Promoter 1: pGL3-Promoter 1; Promoter 2: pGL3-Promoter 2; **difference was extremely significant.

DISCUSSION

E. coli F18 relies on its fimbriae to bind to the F18 receptor on porcine small intestinal brush border epithelial cells for infecting the cells and causing the disease, and these receptors may be glycoproteins or glycolipids (Karlsson, 1989). Using established full-sib pairs of Sutai pigs (Duroc × Meishan) with *E. coli* F18-resistant and -sensitive adhesion phenotypes, we analyzed gene expression in duodenal tissues. The differentially expressed genes identified were found to be involved in a variety of biological processes, including extracellular modification (glycosylation), immune response, cell adhesion and signal transduction, all of which are related to the anabolic metabolism of glycolipids, and inflammation- and immune-associated pathways (Bao et al., 2012).

We identified a seven-gene cluster of the glycosphingolipid biosynthesis-globo series pathway, which might play an important regulatory role against *E. coli* F18 infection in weanling piglets. Glycosphingolipids are not the primary products of this gene cluster. Rather, this gene cluster encodes specific glycosyltransferases that combine the glycosphingolipid precursor, ceramide, with the appropriate oligosaccharide (Hakomori, 2000). Therefore, the expression of the glycosphingolipid biosynthesis-globo series pathway enzymes may indirectly affect the synthesis of the glycosphingolipid products, thereby affecting resistance to *E. coli* F18. Transcription is initiated by binding of an appropriate RNA polymerase to the promoter region; this process is often stringently regulated for controlling gene expression (Christensen et al., 2004; Li et al., 2015). Analysis of *FUT1* gene promoter region revealed that the *FUT1* transcription initiation region had five alternative splicing sites and two promoters, but other genes in the glycosphingolipid biosynthesis-globo series pathway were not studied (Gan et al., 2016). Moreover, other regions of the two promoters can affect gene expression; thus, further studies are required to gain a more complete understanding of the regulation of the entire seven-gene cluster.

The complete porcine genome has been published, but the specific functional areas of many genes, including promoter regions, remain unclear. Additional information, such as that obtained from 5'rapid amplification of cDNA ends (5'-RACE) experiments, is needed to determine transcription start sites and the exact promoter region boundaries, whereas the human genome is much more complete. The promoter sequences of the seven-gene cluster are available in the Eukaryotic Promoter Database (EPD). We used this resource to compare pig and human gene sequences to predict promoter information for the porcine genes. The results suggest that all the porcine genes include different alternative splicing sites. Additionally, both *FUT1* and *B3GALNT1* sequences in humans include alternative promoters, and we expect the same in gene orthologs of pigs.

To determine the promoter region, identification of the transcription start site is necessary. Various methods exist for elucidating the 5'-end, including full-length cDNA library construction, chromosome walking, and 5'-RACE. Full-length cDNA libraries can be useful even if the 5'-end of the gene is unknown, but the process is complicated, technically challenging, and expensive (Zhu et al., 2009). Chromosome walking techniques are commonly used for cloning the known flanking sequence using inverse, ligation-mediated, and specific PCR primers, and the experimental procedures involved differ depending on the sequence features present (Liu et al., 2006; Liang et al., 2009). The third method, 5'-RACE is widely used to obtain the 5'-boundary, but the kits are expensive and can be difficult to use. In addition, a major issue with this approach is that the success rate is low for genes with low expression levels (Luo et al., 2011).

The rapid development of low-cost next-generation sequencing (NGS) methods has made it possible to obtain sequence information for the different functional regions of each gene, including those with lower expression levels (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). Our transcriptome sequencing results showed that the locations of exons in highly expressed genes were in close proximity to the transcription start site, in accordance with the porcine genome database. Moreover, the combination of the dual luciferase reporter results elucidated that expression level of *FUT1* promoter 2 was higher, indicating that it might play the leading role in regulating transcription.

Methylation of the promoters in CpG islands is an important mechanism of regulating gene expression (Tate and Bird, 1993). Bioinformatic analysis of the promoter regions of all the seven genes revealed CpG islands in promoter 1 of *FUT1*, and in the promoters of all of the other genes, except *B3GALNT1*, which included fewer CpG sites and absence of methylation. DNA methylation regulates transcription by affecting the binding of transcription factors (TFs). Therefore, we searched for the promoter regions for TFBSs. Methylation at these specific loci can lead to gene silencing (Mitchell and Tjian, 1989; Latchman, 1997; Lee and Young, 2000; Xie et al., 2012). The CpG islands of *FUT1*, *FUT2*, *HEXA*, *HEX*, *NAGA*, and *ST3GAL1* were found to contain 20, 11, 14, 18, 23, and 18 putative TFBSs, respectively. Methylation of TFBSs for SP1, CREB, Myc, USF-1, CTCF, GATA-1, and AP-2 decreases the affinity of protein-DNA interactions, thereby lowering the transcription rate (Clark et al., 1997; Perini et al., 2005; Kim and Leonard, 2007; Uhm et al., 2012). We found SP1 TFBSs in all CpG islands identified in the present study. Therefore, we predict that SP1 is important for regulating transcription of the glycosphingolipid biosynthesis-globo series pathway genes, and expect that methylation of these loci might inhibit SP1 binding and gene expression.

CONCLUSIONS

Among the seven porcine glycosphingolipid biosynthesis-globo series pathway genes identified in this study, the promoter 2 of *FUT1* gene might plays the leading role in regulating transcription. *FUT1* promoter 1 and the promoters of *FUT2*, *ST3GAL1*, *HEXA*, *HEXB*, and *NAGA* contain CpG islands within the first 500 bp, *B3GALNT1* gene may not become the choice of the methylation detection, in that the *B3GALNT1* promoter has no CpG island. The CpG islands include numerous TFBSs, including the SP1 binding site, and we expect that methylation of these loci might inhibit SP1 binding and gene expression.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by grants from the National Natural Science Funds (#31472066, #31372285), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (#14KJA230003), the Genetically Modified Organisms Technology Major Project (#2014ZX0800601B), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

REFERENCES

- Bao WB, Ye L, Pan ZY, Zhu J, et al. (2012). Microarray analysis of differential gene expression in sensitive and resistant pig to *Escherichia coli* F18. *Anim. Genet.* 43: 525-534. <https://doi.org/10.1111/j.1365-2052.2011.02287.x>
- Bender J (2004). DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* 55: 41-68. <https://doi.org/10.1146/annurev.arplant.55.031903.141641>
- Bertin AM and Duchet-Suchaux MF (1991). Relationship between virulence and adherence of various enterotoxigenic *Escherichia coli* strains to isolated intestinal epithelial cells from Chinese Meishan and European large white pigs. *Am. J. Vet. Res.* 52: 45-49.
- Boldin B (2008). Persistence and spread of gastro-intestinal infections: the case of enterotoxigenic *Escherichia coli* in piglets. *Bull. Math. Biol.* 70: 2077-2101. <https://doi.org/10.1007/s11538-008-9348-8>
- Christensen MA, Zhou W, Qing H, Lehman A, et al. (2004). Transcriptional regulation of BACE1, the β -amyloid precursor protein β -secretase, by Sp1. *Mol. Cell. Biol.* 24: 865-874. <https://doi.org/10.1128/MCB.24.2.865-874.2004>
- Clark SJ, Harrison J and Molloy PL (1997). Sp1 binding is inhibited by (³mCp) CpG methylation. *Gene* 195: 67-71. [https://doi.org/10.1016/S0378-1119\(97\)00164-9](https://doi.org/10.1016/S0378-1119(97)00164-9)
- Gan LN, Dong WH, Wang J, Sun SY, et al. (2016). Determination and activity analysis of pig (*Sus scrofa*) *FUT1* gene promoter region. *J. Agric. Biotechnol.* 24: 168-175.
- Grabe N (2002). AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol. (Gedruckt)* 2: S1-S15.
- Hakomori S (2000). Traveling for the glycosphingolipid path. *Glycoconj. J.* 17: 627-647. <https://doi.org/10.1023/A:1011086929064>
- Hoetzel S, Sprong H and van Meer G (2007). The way we view cellular (glyco)sphingolipids. *J. Neurochem.* 103 (Suppl 1): 3-13. <https://doi.org/10.1111/j.1471-4159.2007.04721.x>
- Inagaki M (2008). [Structure and biological activity of glycosphingolipids from starfish and feather stars]. *Yakugaku Zasshi* 128: 1187-1194. <https://doi.org/10.1248/yakushi.128.1187>
- Karlsson KA (1989). Animal glycosphingolipids as membrane attachment sites for bacteria. *Annu. Rev. Biochem.* 58: 309-350. <https://doi.org/10.1146/annurev.bi.58.070189.001521>
- Kim HP and Leonard WJ (2007). CREB/ATF-dependent T cell receptor-induced *FoxP3* gene expression: a role for DNA methylation. *J. Exp. Med.* 204: 1543-1551. <https://doi.org/10.1084/jem.20070109>
- Lahiri S and Futerman AH (2007). The metabolism and function of sphingolipids and glycosphingolipids. *Cell. Mol. Life Sci.* 64: 2270-2284. <https://doi.org/10.1007/s00018-007-7076-0>
- Latchman DS (1997). Transcription factors: an overview. *Int. J. Biochem. Cell Biol.* 29: 1305-1312. [https://doi.org/10.1016/S1357-2725\(97\)00085-X](https://doi.org/10.1016/S1357-2725(97)00085-X)
- Lee TI and Young RA (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34: 77-137. <https://doi.org/10.1146/annurev.genet.34.1.77>
- Li C, Yang X, Xu M, Zhang J, et al. (2014). Association between *GSTP1* CpG methylation and the early phase of lead exposure. *Toxicol. Mech. Methods* 24: 111-115. <https://doi.org/10.3109/15376516.2013.859195>
- Li HY, Tong SM and Yan Q (2015). Correlation between *FUT4* expression and its promoter methylation in HaCaT cells. *Hereditas* 37: 48-54.
- Li LC and Dahiya R (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18: 1427-1431. <https://doi.org/10.1093/bioinformatics/18.11.1427>
- Li WF, Zhu HY, Yu DY, Fu LQ, et al. (2012). Expression of F18 fimbrial adhesion F antigen gene (*fedF*) in *Escherichia coli* and its immunoprotection in BALB/c Mice (*Mus musculus*). *J. Agric. Biotechnol.* 20: 1065-1071.
- Liang CZ, Zhang R and Guo SD (2009). Progress of chromosome walking. *Biotechnol. Bull.* 10: 75-82.
- Liu B, Su Q, Tang MQ, Yuan XD, et al. (2006). [Progress of the PCR amplification techniques for chromosome walking]. *Yi Chuan* 28: 587-595.
- Luo C, He XH, Chen H, Wei YL, et al. (2011). A high-efficient method of Race technique for obtaining the gene 5' end. *Chih Wu Sheng Li Hsueh T'ung Hsun* 47: 409-414.
- Mitchell PJ and Tjian R (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245: 371-378. <https://doi.org/10.1126/science.2667136>
- Perini G, Diolaiti D, Porro A and Della Valle G (2005). *In vivo* transcriptional regulation of N-Myc target genes is controlled by E-box methylation. *Proc. Natl. Acad. Sci. USA* 102: 12117-12122. <https://doi.org/10.1073/pnas.0409097102>
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29: 24-26. <https://doi.org/10.1038/nbt.1754>
- Sambrook J, Fritsch EF and Maniatis T (1989). Molecular cloning: a laboratory manual. *Cold spring harbor laboratory press.*

- Scherf M, Klingenhoff A and Werner T (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* 297: 599-606. <https://doi.org/10.1006/jmbi.2000.3589>
- Tate PH and Bird AP (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* 3: 226-231. [https://doi.org/10.1016/0959-437X\(93\)90027-M](https://doi.org/10.1016/0959-437X(93)90027-M)
- Thorvaldsdóttir H, Robinson JT and Mesirov JP (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14: 178-192. <https://doi.org/10.1093/bib/bbs017>
- Uhm TG, Lee SK, Kim BS, Kang JH, et al. (2012). CpG methylation at GATA elements in the regulatory region of *CCR3* positively correlates with *CCR3* transcription. *Exp. Mol. Med.* 44: 268-280. <https://doi.org/10.3858/emm.2012.44.4.022>
- Xie XL, Yu Y, Yuan ZF, Yang J, et al. (2012). [Comparative analysis on content and distribution of CpG sites in milk production traits and mastitis-related genes in dairy cattle]. *Yi Chuan* 34: 437-444. <https://doi.org/10.3724/SP.J.1005.2012.00437>
- Zhang J, Zhang WQ, Ding YL, Xu P, et al. (2015). Correlation between type 2 diabetes and DNA methylation and mRNA expression of APN in abdominal adipose tissues in Xinjiang Uygur population. *Hereditas* 37: 269-275.
- Zhu LJ, Zhangsun DT and Luo SL (2009). Methodology of constructing the full length cDNA library and its application. *Natural Sci. J. Hainan University* 27: 185-190.
- Zi C, Wu Z, Wang J, Huo Y, et al. (2013). Transcriptional activity of the *FUT1* gene promoter region in pigs. *Int. J. Mol. Sci.* 14: 24126-24134. <https://doi.org/10.3390/ijms141224126>

Supplementary material

[Figure S1](#). Prediction of transcription factor binding sites (TFBSs) in the *FUT1* CpG islands.

[Figure S2](#). Prediction of transcription factor binding sites (TFBSs) in the *FUT2* CpG islands.

[Figure S3](#). Prediction of transcription factor binding sites (TFBSs) in the *HEXA* CpG islands.

[Figure S4](#). Prediction of transcription factor binding sites (TFBSs) in the *HEXB* CpG islands.

[Figure S5](#). Prediction of transcription factor binding sites (TFBSs) in the *NAGA* CpG islands.

[Figure S6](#). Prediction of transcription factor binding sites (TFBSs) in the *ST3GAL1* CpG islands.