

Novel SNPs and InDels discovered in two promoter regions of porcine pregnancy-associated glycoprotein 2-like subfamily (*pPAG2-Ls*) in crossbreed pigs

Martyna Bieniek-Kobuszewska^{1,2} · Grzegorz Panasiewicz¹ · Aleksandra Lipka¹ · Marta Majewska³ · Bożena Szafranska¹

Received: 10 March 2016 / Revised: 31 August 2016 / Accepted: 5 September 2016 / Published online: 5 October 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This is a pioneer study of single nucleotide polymorphisms (SNPs) within the entire promoter region (1204 bp) of the dominant *pPAG2-L* subfamily in the pig. The *pPAG2-L* subfamily was sequenced/examined using genomic deoxyribonucleic acid (gDNA) templates of crossbreed pigs (Landrace x Large White), and compared to two bacterial artificial chromosome (BAC) clones containing gDNA of the Duroc breed (as the positive controls). Our analysis of the *pPAG2-L* promoter identified 31 SNPs and one InDel mutation in crossbreed pigs. Among 42 SNPs identified in two BAC clones, 24 SNPs had not been previously detected in crossbreed pigs. The sequence alignment of *pPAG2-L* promoter, performed with Lasagne-Search 2.0, Cluster Bluster and MatInspector software, revealed a total of 28 transcription factor binding sites (TFBS) and 10 TFBS (AP-1, CCAAT, CHOP:C, FOXP1, LSF, MRF-2, Myc, NF1, NF-Y, TGIF) within SNPs in the core sequences. It was noted that TFBS (NF1) was found to be unique to the *pPAG2* promoter sequence containing SNPs: g.-1100G>A(R), g.-1101T>C(Y),

represented by GA and TC genotypes ($p_x = 0.12$). Our broad-based novel database thus provides an SNP *PAG2-L* pattern for modern genotyping of female and male progenitors. This is required for further studies of various potential correlations between guiding SNP genotypes of the *pPAG2-L* subfamily in the sows of many breeds, in which the most economically important reproductive traits are properly documented on each farm.

Keywords PAG · Polymorphism · Promoter · SNP · TBSF

Introduction

Multiple pregnancy-associated glycoprotein family (PAGs) belongs to placental aspartic peptidases that have been identified mainly by complementary DNA (cDNA) cloning in some eutherian taxa only (see Xie et al. 1997; Szafranska et al. 2006; Wallace et al. 2015). In the domestic pig (Szafranska et al. 1995; Panasiewicz et al. 2004), among eight trophoblastic cloned and identified cDNAs (GenBank: L34360–1, AF315377, AF272734, AY188554, AF272735, AY373029 and AY775784), five are classified as catalytically active porcine (p) *PAG2-L* subfamily (*pPAG2*, -4, -6, -8 and -10) or potentially inactive *pPAG1-L* subfamily (*pPAG1*, -3 and -5). In the pig (*Cetartiodactyla*), both subfamilies have been identified in trophoblastic cells (TR) during the peri-implantation period and in trophoblast cells (TRD – chorionic epithelium) after post-placental development (Szafranska and Panasiewicz 2002; Szafranska et al. 2006).

The *PAG* family encodes multiple chorionic polypeptide precursors that start expression during peri-implantation, the period of the highest embryonic mortality in many eutherians (Xie et al. 1997; Szafranska et al. 2006; Wallace et al. 2015). All identified *PAG* cDNAs permitted genomic DNA (gDNA)

Martyna Bieniek-Kobuszewska and Grzegorz Panasiewicz contributed equally to this work.

✉ Grzegorz Panasiewicz
panasg@uwm.edu.pl

- ¹ Department of Animal Physiology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, Oczipowskiego Str 1A, 10-719 Olsztyn-Kortowo, Poland
- ² Department of Dermatology, Sexually Transmitted Diseases and Clinical Immunology, Faculty of Medical Sciences, University of Warmia & Mazury in Olsztyn, Wojska Polskiego Str 30, 10-229 Olsztyn, Poland
- ³ Department of Human Physiology, Faculty of Medical Sciences, University of Warmia & Mazury in Olsztyn, Warszawska Str 30, 10-082 Olsztyn, Poland

cloning and initial exon-intron structure organization discoveries of nine exons and eight introns in the cattle and the pig (Xie et al. 1995; Szafranska et al. 2001). To date, 11 promoter sequences of the *PAG* family have been identified: bovine – *bPAG1* (Xie et al. 1995), *bPAG2*, 3, 5, 8, 11, 12, 15, and 18 (Telugu et al. 2009), porcine – *pPAG2* (Szafranska et al. 2001), and equine – *ePAG* (Green et al. 1999). In the pig, the single nucleotide polymorphisms (SNPs) within the promoter of the *pPAG2* and either *pPAG2-L* or *pPAG1-L* subfamilies have not yet been identified.

Despite all available information concerning the promoter sequences of the *PAG* family, there is no information on polymorphism or the potential influence of SNPs on the regulation of transcription. The present study was conducted to identify polymorphisms in the promoter region of the *pPAG2-L* subfamily and to determine the genotype frequency in crossbreed and Duroc pigs (as controls). Such examined SNPs in the crossbreed pigs will provide the main pattern for the genotyping of multiparous sows of many breeds, in which reproductive traits are known, which is economically important in the livestock industry.

Materials and methods

Animals and genomic DNA templates

Blood samples (40 ml in to the tubes with K₂EDTA) were collected from female ($n = 6$) and male ($n = 11$) crossbreed pigs (Landrace x Large White) slaughtered under commercial conditions. All gDNA templates were isolated by the Sherlock AX procedure (A&A Biotechnology, Poland) and used for various amplicon productions by PCR with primer sets (F:GGCTTATCTGTCCCCACTGG and R:AGTAAGACACAGGCAGTC or F:GACTGCCTGTGTCTTACT and R:GACTGTCAGGAATGATGGCA), specific for the *pPAG2* promoter (Acc. No. U39198/GenBank; Szafranska et al. 2001). The PCR conditions were initial denaturation at 94 °C/3 min; then 40 cycles 94 °C/1 min, 62 °C/1 min, 72 °C/2 min, and the final extension at 72 °C/7 min. All PCR mixes (10 µl) contained 0.42 µl dNTP; 0.4 µl 25 mM MgCl₂; 2 µl 10× buffer; 0.4 µl JumpStart™ Taq DNA Polymerase (Sigma-Aldrich, USA); 0.7 µl of each primer (100 ng/µl); and various gDNA templates of the crossbreed pigs (200 ng), parallel to gDNA bacterial artificial chromosome (BAC) clones (5 ng) specific only for the Duroc breed (CH242-60C13 and CH242-294016; BACPAC Resources, CHORI, Children's Hospital Oakland Research Institute, USA) – as the positive controls.

Preparation of the positive controls (gDNA) – BAC clones

Two commercial BAC clones (CH242-60C13 and CH242-294016) were propagated in transformed *Escherichia coli*

strain DH10B bacteria containing gDNA inserts (279 330 and 97 794 bp, respectively) in the pTARBAC1.3 vectors (13 462 bp). The DH10B bacteria were grown in NZY Broth medium supplemented with chloramphenicol (12.5 mg/ml medium), and plasmid DNA (pDNA) was harvested by standard alkaline lysis. Isolated pDNAs were spectrophotometrically assessed and used for PCR amplifications, as described above.

Sequencing and SNP identification within *pPAG2-L* promoters

All *pPAG2-L* amplicons were separated in 1 % agarose gels, purified according GenElute™ Gel Extraction Kit procedure (Sigma-Aldrich, USA) and sequenced in both sense and anti-sense directions by 3130 Genetic Analyzer using the BigDye® Terminator v.3.1 Cycle sequencing procedure (Applied Biosystems, USA). The obtained chromatograms were initially examined by Sequencing Analysis Software (Applied Biosystems, USA), and all sequences were verified by FinchTV (Geospiza, Inc., USA) and aligned by DNASIS v.3.0 (Hitachi Software Engineering Co. Ltd., Japan) and the National Center for Biotechnology Information Basic Local Alignment Search Tool (NCBI BLASTn) using discontinuous Megablast or Blastn in the GenBank. All identified SNPs were named according to the International Union of Pure and Applied Chemistry (IUPAC) codes.

Computer analysis

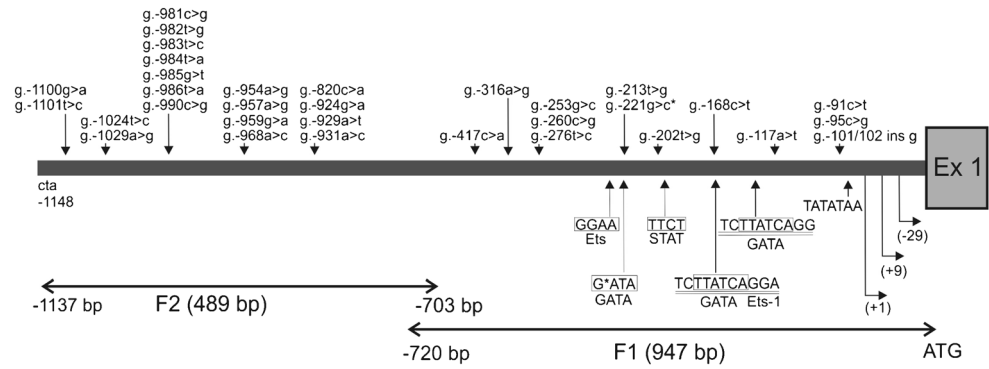
In silico analysis of the *pPAG2-L* promoter sequences for a presence of putative TFBS was performed by Cluster Bluster, MatInspector™ and Lasagna-Search 2.0 with TRANSFAC matrices. The analyses were carried out for all possible TFBS according to the individual settings of each software: Cluster Bluster (Gap Parameter 35; Cluster Score Treshold 2; Motif Score Treshold 2; Residue Abundance Range 100, Pseudocount 0.375), MatchInspector (minimize false positives), Genomatix RegionMiner tool for overrepresentation of TFBSs (Genomatix Software GmbH), and Lasagna-Search 2.0 ($p \leq 0.001$).

Results

Identifications of sequences and novel SNPs within the *pPAG2-L* promoter in the crossbreeds

In total, 31 novel SNPs located from *g.-91C>T(Y)* to *g.-1101T>C(Y)* plus one InDel mutation (*g.-100/101InsG*) upstream ATG were identified in the promoter region of the *pPAG2-L* subfamily (Fig. 1, Table 1). This provides a novel major pattern of the largest genetic variation of the porcine

Fig. 1 Schematic localization of the SNPs in the promoter sequence (1204 bp upstream from ATG) of the *pPAG2-L* subfamily examined in the crossbreed pigs. This figure includes the transcription start site (+1, +9, -29); potential binding sites for transcription factors - Ets, GATA, STAT (boxed); TATA-box (TATATAA); unique tandem repeats (double underlined); the occurrence of SNP (*p*.-168 g > c*) in the coding sequence for GATA



genome due to various crossbreeds. All SNPs were submitted to the dbSNP/NCBI database and analyzed according to the 1204 bp of the *pPAG2* promoter (Acc. No. U39198; GenBank). The SNPs were identified within two promoter fragments, including F1) 947 bp proximal regulatory region (-720 bp upstream ATG) and F2) 489 bp flanking distal region (from -703 to -1137 bp upstream ATG). Among 32 SNPs/InDel, 13 SNPs were identified in the F1 and 19 SNPs in the F2. Within the F1, one insertion (*g*.-100/101InsG) and 12 SNPs (4 transitions – TRNs and 8 transversions – TRVs) were identified, while among 19 SNPs in the F2, we detected 9 TRNs and 10 TRVs.

Genotype frequency varied in homo- and heterozygotes in a range $p_x = 0.55-1$. Among 31 SNPs and one InDel, 18 SNPs dominated in homozygotes ($p_x = 0.56-0.88$), whereas 14 SNPs existed in heterozygotes ($p_x = 0.55-1$). Only one SNP (*g*.-221G > C) localized in the F1, dominated in heterozygotes ($p_x = 0.86$), changing the sequence of GATA, while other neighboring SNPs did not affect TFBS – GATA. Most of the 19 SNPs identified in the F2 were closely localized in three major groups: (1) 820–931, (2) 954–968, and (3) 981–990 bp, while 4 other SNPs existed in tandems until -1101 bp upstream ATG. Interestingly, among the 32 identified SNPs/InDel, only three identified SNPs in homozygotes (Table 1): 12) *g*.-316a > g (GG); 14) *g*.-820C > A (AA); and 21) *g*.-968A > C (CC) had genotypes existing with frequencies (p_x) ranging from 0.06 to 0.6.

All original chromatograms of 31 SNPs and one InDel identified within both F1 and F2 regions of the *pPAG2-L* promoter (Table 1), including the monoallelic (homozygotes) and biallelic (heterozygotes) visualized by the Finch TV (Fig. 2), revealed sequence differences compared to the only available consensus sequence of the *pPAG2* promoter for various porcine breeds (U39198; Szafranska et al. 2001). In addition, we identified that in the BAC clones (CH242-60C13 and CH242-294016) used as the only available commercial control sequences (for *pPAG3* and *pPAG6*), a surprisingly large diversity was identified for various members of the entire *pPAG* family, including the *pPAG1-L* and *pPAG2-L* subfamilies.

Identification of sequences and SNPs within the *pPAG2-L* promoter in control BAC clones

Sequencing of the commercial BAC clones (CH242-60C13 and CH242-294016), used as the major positive controls containing only gDNA specific for the Duroc breed, revealed very high sequence diversity, including the presently identified 36 SNPs and 6 InDels (Table 2). Conversely, our parallel broad-based *in silico* analysis of both BAC clone sequences revealed very huge multiplicity and diversity of the entire *pPAG2-Ls* and/or additionally numerous and various fragments (Panasiewicz et al. unpublished data).

Surprisingly, 42 SNPs with the Duroc gDNA template were discovered, including 24 novel SNPs (Table 2), not identified in crossbreed pigs. It was also found that 18 SNPs occurred in both Duroc and crossbreed pigs (underlined in Table 2). Among 42 SNPs, five deletions, *g*.-950DelA; *g*.-956DelG; *g*.-974DelG; *g*.-975 DelG, and *g*.-976DelG, were identified, 17 TRNs and 19 TRVs and one insertion (*g*.-101_-102InsG), which was also discovered in all crossbreed pigs (Table 2). Further evidence of the *pPAG2-L* genetic diversity provided a comparative analysis of the SNPs, which were identified in the promoter sequences of BAC clones (Duroc) and crossbreed pigs (Table 2).

In silico identification of various TFBSs in the *pPAG2-L* promoter sequence of the crossbreeds

The investigations for the vertebrate-specific promoter elements by Lasagne-Search 2.0 (using 259 motifs from TRANSFAC® application) in the *pPAG2-L* promoter sequence revealed 11 various TFBSs (Table 3, Fig. 3). Among of all identified motifs, only two were identified on the sense strand, whereas nine were on anti-sense strand (p value = 0, e value = 0, and score from 8.14 to 18.64). It was confirmed that allele G (*g*.-221G > C; according to IUPAC numbering in Table 1) created the GATA core motif, while the C allele was unable to create such TFBS (Fig. 1). Three SNPs, *g*.-1101T > C(Y), *g*.-924G > A(R), and *g*.-202T > G(K), appeared in the TFBS core sequences of NF-Y, CHOP:C, and MRF-2,

Table 1 Genotype frequencies (p_x) of the identified SNPs in the promoter of the *pPAG2-L* gene subfamily in crossbreed pigs

SNP locus (IUPAC code) ^a	SNP locus acc. DNASIS ^b	Genotypes and frequencies (p_x)						
		Genotype	p_x	Genotype	p_x	Genotype	p_x	
F1 region								
1)	<i>g.-91C>T (Y)</i>	<i>g.1111C>T</i>	CC	0.45	CT	0.55	TT	0.0
2)	<i>g.-95C>G (S)</i>	<i>g.1107C>G</i>	CC	0.45	CG	0.55	GG	0.0
3)	<i>g.-101/102->INSG</i>	<i>g.1100/1101->InsG</i>	GG	1.0	—	0.0	—	0.0
4)	<i>g.-117A>T (W)</i>	<i>g.1085A>T</i>	AA	0.57	AT	0.43	TT	0.0
5)	<i>g.-168C>T (Y)</i>	<i>g.1034C>T</i>	CC	0	CT	1.0	TT	0.0
6)	<i>g.-202T>G (K)</i>	<i>g.1000T>G</i>	TT	0.56	TG	0.44	GG	0.0
7)	<i>g.-213T>G (K)</i>	<i>g.989T>G</i>	TT	0.14	GT	0.86	GG	0.0
8)	<i>g.-221G>C (S)</i>	<i>g.981G>C</i>	GG	0.14	CG	0.86	CC	0.0
9)	<i>g.-253G>C (S)</i>	<i>g.949G>C</i>	GG	0.17	CG	0.83	CC	0.0
10)	<i>g.-260C>G (S)</i>	<i>g.942C>G</i>	CC	0.0	CG	1.0	GG	0.0
11)	<i>g.-276T>C (Y)</i>	<i>g.926T>C</i>	TT	0.0	CT	1.0	CC	0.0
12)	<i>g.-316A>G (R)</i>	<i>g.886A>G</i>	AA	0.2	AG	0.2	GG	0.6
13)	<i>g.-417C>A (M)</i>	<i>g.785C>A</i>	CC	0.0	AC	1.0	AA	0.0
F2 region								
14)	<i>g.-820C>A (M)</i>	<i>g.382C>A</i>	CC	0.2	AC	0.4	AA	0.4
15)	<i>g.-924G>A (R)</i>	<i>g.278G>A</i>	GG	0.71	AG	0.29	AA	0.0
16)	<i>g.-929A>T (W)</i>	<i>g.273A>T</i>	AA	0.71	AT	0.29	TT	0.0
17)	<i>g.-931A>C (M)</i>	<i>g.271A>C</i>	AA	0.71	AC	0.29	CC	0.0
18)	<i>g.-954A>G (R)</i>	<i>g.248A>G</i>	AA	0.0	AG	1.0	GG	0.0
19)	<i>g.-957A>G (R)</i>	<i>g.245A>G</i>	AA	0.0	AG	1.0	GG	0.0
20)	<i>g.-959G>A (R)</i>	<i>g.243G>A</i>	GG	0.0	AG	1.0	AA	0.0
21)	<i>g.-968A>C (M)</i>	<i>g.234A>C</i>	AA	0.12	AC	0.82	CC	0.06
22)	<i>g.-981C>G (S)</i>	<i>g.221C>G</i>	CC	0.76	CG	0.24	GG	0.0
23)	<i>g.-982T>G (K)</i>	<i>g.220T>G</i>	TT	0.76	GT	0.24	GG	0.0
24)	<i>g.-983T>C (Y)</i>	<i>g.219T>C</i>	TT	0.76	CT	0.24	CC	0.0
25)	<i>g.-984T>A (W)</i>	<i>g.218T>A</i>	TT	0.76	AT	0.24	AA	0.0
26)	<i>g.-985G>T (K)</i>	<i>g.217G>T</i>	GG	0.76	GT	0.24	TT	0.0
27)	<i>g.-986T>A (W)</i>	<i>g.216T>A</i>	TT	0.0	AT	1.0	AA	0.0
28)	<i>g.-990C>G (S)</i>	<i>g.212C>G</i>	CC	0.82	CG	0.18	GG	0.0
29)	<i>g.-1024T>C (Y)</i>	<i>g.178T>C</i>	TT	0.82	CT	0.18	CC	0.0
30)	<i>g.-1029A>G (R)</i>	<i>g.173A>G</i>	AA	0.82	AG	0.18	GG	0.0
31)	<i>g.-1100G>A (R)</i>	<i>g.102G>A</i>	GG	0.88	AG	0.12	GG	0.0
32)	<i>g.-1101T>C (Y)</i>	<i>g.101T>C</i>	TT	0.88	CT	0.12	CC	0.0

^a Numbering of SNPs submitted in dbSNP/NCBI database

^b Numbering according to promoter sequence U39198, GenBank (Szafranska et al. 2001)

respectively. The other SNPs occurred within AP-2rep, MyoD, HNF-1, FAC1, and AREB6, but STAT5A occurred outside of the TFBS core sequences.

The Cluster Buster revealed 12 TFBS in three clusters (C1–C3) including 11–188 bp (C1), 1050–1081 bp (C2), and 812–884 bp (C3) of the *pPAG2* promoter (Table 3). The score of motifs was 4.12–8.18 for sense strand and 4.46–8.86 for the anti-sense strand. Some of TFBS were identified in various regions, e.g., GATA (C1: 11–23; C2: 1050–1062 and 1069–1081; C3: 872–884) or NF-1 (C1:

100–117, 128–145, C3: 812–829 and 854–871). Only two SNPs, *g.-1100G>A (R)* and *g.-1101 T>C (Y)*, were identified within NF1 and AP-1 motifs.

The MatInspector/Genomatix confirmed eight major TFBSs (Table 3). The SNP (*g.-929A>T(W)/g.273A>T*) was identified within core AACAA (272–275) of the Fork head domain factors (FOXP1; 265–281), whereas two SNPs, *g.-931A>C(M)* and *g.-924G>A(R)*, were outside the core sequence. Furthermore, two SNPs *g.-95C>G(S)* and *g.-91C>T (Y)* were identified outside the GTCA core sequence (1114–

Table 2 SNP identification in promoter sequence of the *pPAG2-L* gene subfamily within BAC clones: CH242-60C13 and CH242-294016

SNP localization in 1204-bp promoter				BAC clones (Duroc)				Crossbreeds
SNP locus (IUPAC code) ^b	SNP locus acc. DNASIS ^c	<i>pPAG2</i>	CH242-60C13 ^d		CH242-294016 ^e			
			Amplicon 1	Amplicon 2	Amplicon 3	Amplicon 4		
F1 region								
1)	<u><i>g.-101-102InsG</i></u>	<i>g.1100_1101InsG</i>	<i>DelG</i>	GG	—	GG	—	GG
2)	<u><i>g.-168C>T(Y)</i></u>	<i>g.1033C>T</i>	CC	TT	—	TT	—	CC
3)	<u><i>g.-316A>G(R)</i></u>	<i>g.886 A>G</i>	AA	AG	—	AG	—	GG/AG
4)	<u><i>g.-417C>A(M)</i></u>	<i>g.785C>A</i>	CC	AC	—	AC	—	AA
F2 region								
5)	<i>g.-740G>T(K)^a</i>	<i>g.462G>T</i>	GG	GG	TT	GG	TT	GG
6)	<i>g.-810A>G(R)^a</i>	<i>g.392A>G</i>	AA	AA	GG	AA	GG	AA
7)	<u><i>g.-820C>A(M)</i></u>	<i>g.382C>A</i>	CC	AC	CC	AC	CC	AA/CC
8)	<i>g.-826G>A(R)^a</i>	<i>g.376G>A</i>	GG	GG	AA	GG	AA	GG
9)	<i>g.-827G>C(S)^a</i>	<i>g.375G>C</i>	GG	GG	CC	GG	CC	GG
10)	<i>g.-828G>A(R)^a</i>	<i>g.374G>A</i>	GG	GG	AA	GG	AA	GG
11)	<i>g.-836C>A(M)^a</i>	<i>g.366C>A</i>	CC	CC	AA	CC	AA	CC
12)	<i>g.-839G>A(R)^a</i>	<i>g.363G>A</i>	GG	GG	AA	GG	AA	GG
13)	<i>g.-870T>G(K)^a</i>	<i>g.332T>G</i>	TT	TT	GG	TT	GG	TT
14)	<i>g.-890C>T(Y)^a</i>	<i>g.312C>T</i>	CC	CC	TT	CC	TT	CC
15)	<u><i>g.-924G>A(R)</i></u>	<i>g.278G>A</i>	GG	GG	AA	GG	AA	GG
16)	<u><i>g.-929A>T(W)</i></u>	<i>g.273A>T</i>	AA	AA	TT	AA	TT	AA
17)	<u><i>g.-931A>C(M)</i></u>	<i>g.271A>C</i>	AA	AA	CC	AA	CC	AA
18)	<i>g.-950DelA^a</i>	<i>g.252DelA</i>	AA	AA	DelA	AA	DelA	AA
19)	<i>g.-956DelG^a</i>	<i>g.246DelG</i>	GG	GG	DelG	GG	DelG	GG
20)	<u><i>g.-968A>C(M)</i></u>	<i>g.234A>C</i>	AA	AC	CC	AC	CC	AC/CC
21)	<i>g.-973G>A(R)^a</i>	<i>g.229G>A</i>	GG	GG	AA	GG	AA	GG
22)	<i>g.-974DelG^a</i>	<i>g.228DelG</i>	GG	GG	DelG	GG	DelG	GG
23)	<i>g.-975DelG^a</i>	<i>g.227DelG</i>	GG	GG	DelG	GG	DelG	GG
24)	<i>g.-976DelG^a</i>	<i>g.226DelG</i>	GG	GG	DelG	GG	DelG	GG
25)	<i>g.-977G>T(K)^a</i>	<i>g.225G>T</i>	GG	GG	TT	GG	TT	GG
26)	<i>g.-978G>T(K)^a</i>	<i>g.224G>T</i>	TT	TT	GG	TT	GG	TT
27)	<u><i>g.-981C>G(S)</i></u>	<i>g.221C>G</i>	CC	CC	GG	CC	GG	CG
28)	<u><i>g.-982T>G(K)</i></u>	<i>g.220T>G</i>	TT	TT	GG	TT	GG	TG
29)	<u><i>g.-983T>C(Y)</i></u>	<i>g.219T>C</i>	TT	TT	CC	TT	CC	CT
30)	<u><i>g.-984T>A(W)</i></u>	<i>g.218T>A</i>	TT	TT	AA	TT	AA	AT
31)	<u><i>g.-985G>T(K)</i></u>	<i>g.217G>T</i>	GG	GG	TT	GG	TT	GT
32)	<u><i>g.-986T>C(Y)</i></u>	<i>g.216T>C</i>	TT	TT	CC	TT	CC	AT
33)	<u><i>g.-990C>G(S)</i></u>	<i>g.212C>G</i>	CC	CC	GG	CC	GG	CG
34)	<u><i>g.-1024T>C(Y)</i></u>	<i>g.178T>C</i>	TT	CT	CC	CT	CC	CT
35)	<u><i>g.-1029A>G(R)</i></u>	<i>g.173A>G</i>	AA	AG	GG	AG	GG	AG
36)	<i>g.-1042C>A(M)^a</i>	<i>g.160C>A</i>	CC	CC	AA	CC	AA	CC
37)	<i>g.-1060C>T(Y)^a</i>	<i>g.142C>T</i>	CC	CC	TT	CC	TT	CC
38)	<i>g.-1070G>A(R)^a</i>	<i>g.132G>A</i>	GG	GG	AA	GG	AA	GG
39)	<i>g.-1077C>T(Y)^a</i>	<i>g.125C>T</i>	CC	TT	—	TT	—	CC
40)	<i>g.-1093G>A(R)^a</i>	<i>g.109G>A</i>	GG	AA	—	AA	—	GG
41)	<i>g.-1095G>T(K)^a</i>	<i>g.107G>T</i>	TT	GG	—	GG	—	TT
42)	<i>g.-1096C>A(M)^a</i>	<i>g.106C>A</i>	CC	AA	—	AA	—	CC

SNPs occur in Duroc and crossbreed pigs are underlined

^a SNPs not detected in crossbreed pigs are italicized

^b Numbering of SNPs submitted in dbSNP/NCBI database

^c Numbering according to promoter sequence U39198, GenBank (Szafranska et al. 2001)

^d BAC clone (CH242-294016) containing *pPAG3* gene sequences

^e BAC clone (CH242-60C13) containing *pPAG3* and *pPAG6* gene sequences

Table 3 Prediction of transcription factors (TF) binding sites with the context of novel SNPs [in brackets] identified *in silico* in the promoter of the *pPAG2-L* gene with the use of Lasagne-Search 2.0, Cluster Buster, and MatInspector

TF motif	Localization	Strand	Score	TF sequence
Lasagne-Search 2.0				
AP-2rep	23–29	–	8.14	CAGTGGG ^b
<u>NF-Y</u>	93–103	+	13.05	TGACCAA [T/C][G/A]TG
MoyD	188–199	+	11.99	ACACAGGTGCTT
<u>CHOP:C/EBPalpha</u>	275–297	–	13.2	ATGTGAAATC[C/T]CC
HNF-1	404–420	–	14.2	AGTCAATGAATGGCCTG
FAC1	574–587	–	10.21	ATCCAAAACATGTT
AREB6	667–678	–	11.89	TGACACCTGGGG
<u>MRF-2</u>	995–1005	–	18.64	CAGAAAT[A/C]CAGA
GATA-2	1051–1060	–	11.32	CCTGATAAGA
GATA-2	1070–1079	–	11.32	CCTGATAAGA
STAT5A	1158–1165	–	8.85	GAGTTCCTG
Cluster Buster				
GATA	11–23	–	6.87	GACAGATAAGCCA
CCAAT	37–52	+	6.16	CTTGACAAATGGAAGG
Ets	47–57	+	8.18	GGAAGGAAAAG
AP-1	52–62	–	4.49	TGTGACTTTTC
Myc	57–66	–	7.12	ACCATGTGAC
<u>CCAAT</u> ^a	92–107	+	4.28	TTTGACCAA[T/C][G/A]TGGCT
<u>Myc</u>	97–106	+	4.58	CCAA[T/C][G/A]TGGC
<u>LSF</u>	101–115	+	5.26	[T/C][A/G]TGGCTGGAACCTC
<u>LSF</u>	102–116	–	5.85	GGAGGTCCAGCCA[T/C]
NF-1	128–145	+	4.65	TCTTGACTCCCACCCCA
AP-1	129–139	+	4.63	CTTGACTCCCA
Sp1	133–145	–	7.06	TGGGGGTGGGAGT
<u>LSF</u>	173–187	+	5.45	[A/G]CAGG[T/C]TGAATCCAG
LSF	174–188	–	8.86	GCTGGATTCAACCTG
<u>NF1</u> ^c	100–117	+	5.25	A[C/T][A/G]TGGCTGGAACCTCCT
<u>AP-1</u>	177–187	+	4.12	GC[T/C]GAATCCAG
NF-1	812–829	–	7.85	TCTTGGCAGCCACTTTGT
SRF	838–850	+	5.33	AACTATGAAATGC
TATA	840–854	+	4.45	CTATGAAATGCCAAG
SRF	848–860	+	5.9	TGCCAAGCATGGC
NF-1	854–871	+	5.03	GCATGGCCCCCAGCACTT
Myf	864–875	–	4.46	AAATAAGTGCTG
Mef-2	868–879	+	7.84	ACTTATTTTTAT
GATA	872–884	–	6.1	ATGTGATAAAAAT
GATA	1050–1062	–	6.94	TCCTGATAAGATC
GATA	1069–1081	–	8.43	ACCTGATAAGAAA
MatInspector				
MZF1.01	18–28	–	1.0	GTGGGGACAGA
<u>FOXP1 ES.01</u> ^a	265–281	+	1.0	ATTTGA[A/C]A[A/T]CAGG[G/A]GAT
SMAD3.01	293–303	–	1.0	GAAGTCTGGAT
FOXP1 ES.01	568–584	+	1.0	TGAATAAAAACATGTTTT
FOXP1 ES.01	573–589	–	1.0	CATCCAAAACATGTTTT
TBX20.01	797–819	–	1.0	CACTTTGTGAGGTGTGCTTTCT
GATA.01	1069–1081	–	1.0	ACCTGATAAGAAA
<u>TGIF01</u> ^d	1106–1122	+	1.0	A[C/G]AGG[C/T]ATGTCAGAGCA

^a Underlined motifs are the TFBS with SNPs in the core sequence^b Bolded letters are the core nucleotides^c Bolded and underlined motifs are the new TFBS that appeared as a result of SNP occurrence^d Underlined and italic motifs are the TFBS with SNP outside the core sequence

1117) of the TALE homeodomain class recognizing TG motifs (TGIF). There was no prevalence of SNPs in the sequences coding other TFBSs: MZF1 (18–28), SMAD3 (293–303), FOXP1 (573–589), TBX20 (797–819), and GATA (1069–1081).

Discussion

In total, 31 SNPs and one InDel (Table 1, Fig. 1) were identified in the *pPAG2-L* promoter (up to -1100 upstream ATG), including 19 SNPs in the F2 (15 SNPs in three clusters, and another four SNPs occurring in tandems). In F1, 13 SNPs were identified, including one InDel (*g.-101/102InsG*) inside the GATA sequence ($p_x = 0.86$).

Currently, the 32 polymorphisms of the *pPAG2-L* promoter (from *g.-91C>T(Y)* to *g.-1101T>C(Y)*) identified in crossbreed pigs can only be compared to the five deposited promoter sequences: *pPAG2* (Szafranska et al. 2001), *bPAG1*, *ePAG*, *bPAG2* (Xie et al. 1995; Green et al. 1999; Telugu et al. 2009), and *fPAG* (Ensembl). However, a comparison of the SNPs in the *pPAG2-L* promoter sequence is impossible, because the SNPs of *bPAG1*, *ePAG*, and *bPAG2* have not been studied.

Surprisingly, a comparative analysis of identified BAC clone sequences revealed 97.0–99.3 % homology to the *pPAG2* promoter (Szafranska et al. 2001), which suggests SNPs among different breeds. The sequence diversity of two BAC clones originating from Duroc (used as gDNA control templates) revealed 42 polymorphisms (36 SNPs and 6 InDel; Table 2), among which novel 24 SNPs have also been identified in the crossbreed pigs (Table 1). It should be noted that some identified SNPs in crossbreed pigs were identical as in the BAC clones (Duroc), which undoubtedly indicates that the currently-tested crossbreed pigs were interbreeding hybrids between Duroc with other breeds. Thus, the identified SNPs are evidence of duplication and positive selection of the *pPAG2-L* subfamily in different breeds.

Previously, specific sequences of the various TFBSs were identified in porcine *pPAG2*: Ets, Ets-1, GATA, GATA-like, and STAT (Szafranska et al. 2001). Presently, the location of SNPs identified in the *pPAG2-L* promoter regulatory F1 region suggests importance due to the potential impact on the TFBSs and, consequently, transcription activation. *In silico* analysis using three programs (Table 3) revealed a total of 28 various TFBSs. In the F1, we found conserved sequences for TATATAA box (from -73 to -79 bp), AP1 (activator protein 1) and CCAAT (enhancer binding protein (C/EBP). The Lasagne 2.0 software was able to detect SNPs (*g.-1101T>C*, *g.-924G>A*, and *g.-202T>G*), which diminish binding sites for NF-Y, CHOP:C, and MRF-2, which may have an influence on the *PAG2-L* expression in these three heterozygote genotypes: CT ($p_x = 0.12$), AG ($p_x = 0.29$), and

TG ($p_x = 0.44$), respectively. Cluster Buster also confirmed that two SNPs, *g.-1000G>A(R)* and *g.-1101T>C(Y)*, may affect the *PAG2-Ls* expression by NF1 and AP-1.

Furthermore, the presence of AP-2 transcription factor was detected, which was also found in the promoters of some placental bovine genes, especially *bPAG1* and *bPAG17*. This suggests that the AP-2 family is a major factor regulating genes depending on cytochrome P-450 involved in the production of steroid hormones in the binucleated cells (Ushizawa et al. 2007). The significant evidence of the *PAG* family involvement in the regulation of pregnancy maintenance has provided commercial bovine microarrays containing 1780 genes, including 30 expressed genes (25–250 dpc), mainly in the bovine two-nucleated placental cells (Ushizawa et al. 2007). Moreover, *Affymetrix* microarrays showed a significant correlation of the *bPAG11* with prostaglandins (PG) synthesis: PGE synthase ($r = 0.76$), cytosolic PGE synthase 3 ($r = 0.69$), and endoperoxide synthase 2 ($r = 0.86$), suggesting an important role of the *bPAG11* in the PG cascade activation (Thompson et al. 2011). It is necessary to underline that two presently identified SNPs in the *pPAG2-L* promoter (*g.-117A>T* and *g.-168C>T*) were localized within, or almost nearby, 10 nt unique tandem repeats (TCTTATCAGG located at -94 to -103 and -113 to -122 upstream of ATG), which are specific in the activation of the *PAG* gene family in pigs (Szafranska et al. 2001) or/and cattle (Telugu et al. 2009), respectively. Both of these SNPs identified in crossbreed pigs are very close to the conservative Ets sequence with proximity to the GATA within the *pPAG2* promoter recognized previously (Szafranska et al. 2001). In cattle, the Ets analyzed in eight known *bPAG* promoters maintained conservative sequences (Telugu et al. 2009). Thus, these SNPs may affect placental development and pregnancy maintenance in both species.

Although there was no prevalence of SNPs in the GATA sequence of the *pPAG2-L* promoter, the Cluster Buster revealed that allele A (SNP *g.-1100G>A*) creates NF1, while allele G (genotype GG) does not create this TFBS. However, the frequency of heterozygote genotype GA and TC was, in both cases, $p_x = 0.12$; the frequency of homozygote genotypes GG and TT (which do not determine TFBS) was $p_x = 0.88$. The SNPs *g.-1100G>A(R)* that appeared inside the core determined the occurrence of a new TFBS (NF-1).

In contrast, the MatInspector revealed sequences characteristic for three variously located *FOXP1 ES.01*, although only one SNP (*g.-929A>T(W)/g. 273A>T*) was localized in the core AACA sequence of this TFBS (265–281 bp). Previous studies have shown that *FOXP1* deletion has an embryonic lethal defect that affects a variety of organs, including cardiac (Wang et al. 2004) and lung development (Shu et al. 2007), and B cell differentiation (Hu et al. 2006). The SNPs within the *pPAG2-Ls* promoter (from *g.-990C>G* to *g.-954A>G*; *g.-820C>A*; *g.-417C>A*; *g.-276T>C* to *g.-213T>G*; *g.-168C>T*;



Fig. 2 Chromatograms of identified SNPs (*arrows*) in homozygotes and heterozygotes within the promoter sequence region (from g.-91C>T to g.-1029A>G) of the *pPAG2-L* subfamily in crossbreed pigs. [For an interpretation of the references to color in this figure legend, the reader is referred to the web version of this article]

and g.-117A>T to g.-91C>T) did not influence or affect the appearance of TFBSs. The other TFBSs identified *in silico* is Tbx20 – a transcription factor that is essential for proper heart development in a growing fetus (Song et al. 2006).

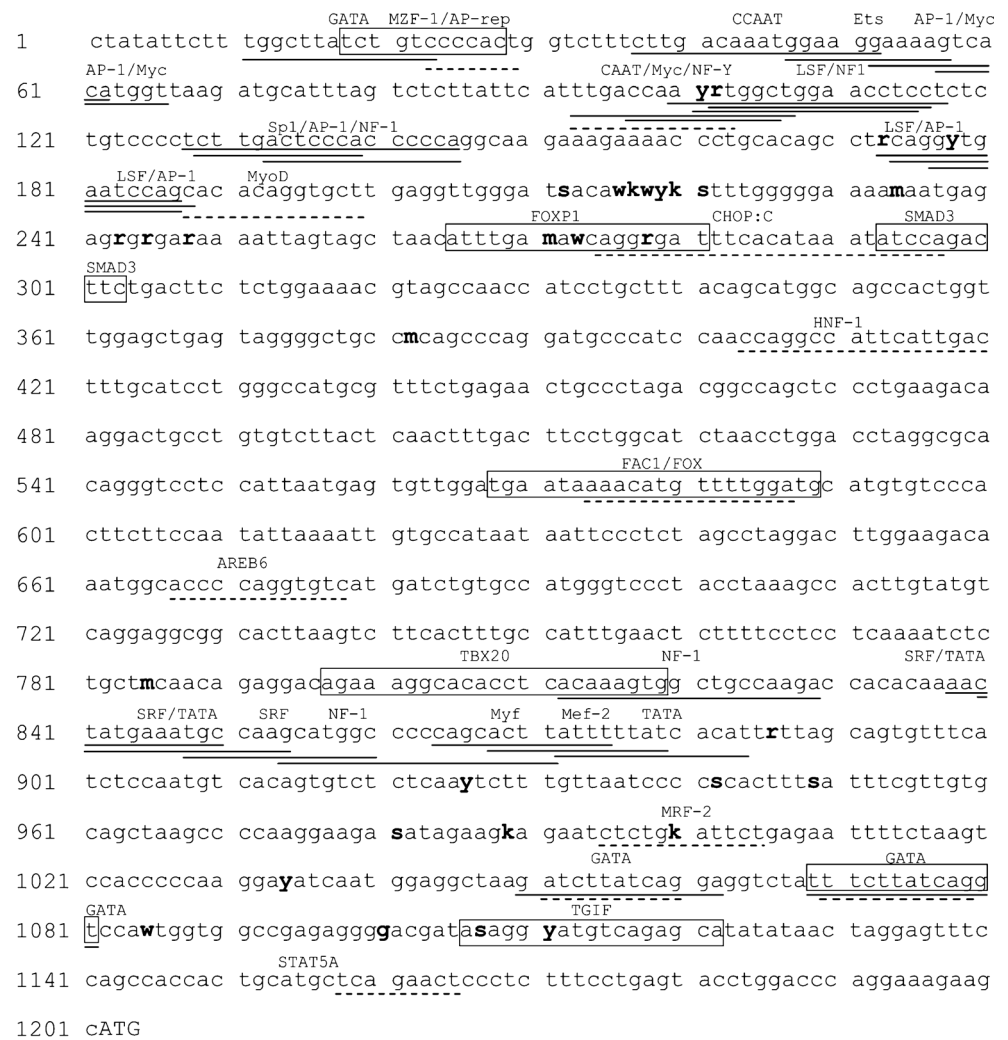
The participation of many transcription factors involved in *bPAG* activation has been shown by EMSA (electrophoretical mobility shift assay), and the most important was assigned to the Ets2 and DDVL – drosophila dorsal ventral factor (Telugu et al. 2009), as well as in the regulation of transcription of chorionic genes, e.g., IFN τ in cattle (Ezashi et al. 1998) or hCG in women (Ghosh et al. 2003). Presently, in crossbreed pigs, no SNPs were identified in the Ets2 binding site, although the identified SNPs in the GATA suggest the possibility of disturbance during *pPAG2-L* subfamily activation. In addition, the bovine microarrays (Kumar et al. 2010) indicate

a strong influence of the STAT Pax-2 (signal transducer and activator of transcription; paired homeobox 2) on the promoters of genes that are expressed in the placenta, e.g., *bPAG2*, *PTGS2* (*COX2*, *PG – endoperoxide synthase 2*) and *LSG 34F*, as a homologue of a secretory vesicle protein in the male (*SSLP-1*; *seminal vesicle protein secreted*).

The investigation of SNP spreading in the selected population requires an important parameter – MAF (minor allele frequency) at the level of >0.1 in commercial pig breeds (McLaren et al. 2013). Although in our study, the MAF was not specified, among 32 SNPs, we are able to identify the genotype frequency $p_x = 0.56–0.88$ for 18 SNPs, which indicates the dominance of homozygotes, while in the case of 14 SNPs it indicates the dominance of heterozygotes ($p_x = 0.55–1$).

It is well known that the smallest polymorphism results from homozygosity of different domestic pig breeds. The occurrence of 1.2 SNPs/kbp in the genome of European Large White breed and only 0.05 SNPs/kbp in the genome of *Sus barbatus* indicates the inbreeding of pigs from the isolated populations, as a result of natural barriers or a human

Fig. 3 Promoter sequence (1204 bp with ATG) of the *pPAG2* gene (U39198; GenBank) containing identified SNPs (acc. IUPAC code; *bolded line*) and TFBS predicted with the use of Lasagne-Search 2.0 (*dashed line*), Cluster Buster (*solid line*), and MatInspector (*boxed*)



economic activity (Ferreira et al. 2008). The regions of homozygosity (ROHs) in European, Asian, and wild pigs (60 K SNP microarray) vary about 778.8 ± 349 ROHs/genome (one ROHs range between 10 kbp to 83.6 Mbp; average 1.11 Mbp), which represents about 23 % of the porcine genome (Bosse et al. 2012). A higher level of ROHs occurs in genomes of wild and domestic European pigs, while the lowest ROHs level (and the largest polymorphism) is present in the genome of wild Asian pigs. Furthermore, $1733 \text{ SNP} \pm 0.57/\text{kbp}$ occur in the porcine genome, but only $2.49 \text{ SNP} \pm 0.57/\text{bp}$ are in regions outside the ROHs (Bosse et al. 2012). This may suggest that the greatest heterozygosity of the *pPAG2-L* promoter occurs within various areas potentially located outside the ROHs in the crossbreed pigs.

Conclusion

This study provides pioneering information on polymorphism and hints at the discovery of 32 SNPs/InDel identified within the regulatory proximal and flanking distal regions of the *pPAG2-L* promoter subfamily in crossbreed pigs and 42 SNPs/InDels identified in the Duroc breed (as inserts in BAC clones used as controls). Many of the *pPAG2-L* SNPs were identified in various TFBSs (at least 8 to 26, due to the three softwares used), which suggests the high importance of allelic (homo- and heterozygotic) diversity and meaningful influence on transcriptional regulation of the *pPAG2-L* subfamily expression.

Since this is the first study describing the *pPAG2-L* subfamily diversity in the genome of crossbreed pigs, it therefore also increased/extended the general knowledge on the last version of the domestic pig genome (Ss10.2). The results present a broad-based novel database (main pattern) – as the widest genotyping prototype, which is required for further investigations of various potential correlations between guiding SNP genotypes of the *pPAG2-L* subfamily in the sows of many breeds, in which the most economically important reproductive traits are properly documented on each farm of female and male progenitors.

Acknowledgments This study was financed by the Polish Ministry of Science and Higher Education/National Science Centre (NN311/066237 granted to G. Panasiewicz).

Authors' contributions MB-K performed experimental work and drafted the manuscript; GP assisted in the study design, coordinated all experimental work, and had important input in writing the manuscript; AL and MM participated in all research stages, and BS conceived the study design and helped in writing the final version of the manuscript. All authors have approved the final version of the article.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bosse M, Megens HJ, Madsen O, Paudel Y, Frantz LAF, Schook LB, Crooijmans RPMA, Groenen MAM (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics* 8(11):e1003100. doi:10.1371/journal.pgen.1003100
- Ezashi T, Early AD, Ostrowski MC, Roberts RM (1998) Control of interferon τ gene expression by Ets-2. *Proc Natl Acad Sci U S A* 95: 7882–7887
- Ferreira E, Souto L, Soares AMVM, Fonseca C (2008) Genetic structure of the wild boar population in Portugal: evidence of a recent bottleneck. *Mamm Biol* 74:274–285
- Ghosh D, Ezashi T, Ostrowski MC, Roberts RM (2003) A central role for Ets-2 in the transcriptional regulation and cyclic adenosine 5'-monophosphate responsiveness of the human chorionic gonadotropin-beta subunit gene. *Mol Endocrinol* 17:11–26
- Green J, Xie S, Szafranska B, Newman A, Gan X, McDowell K, Roberts RM (1999) Identification of a new aspartic proteinase expressed by the outer chorionic cell layer of the equine placenta. *Biol Reprod* 60: 1069–1077
- Hu H, Wang E, Borde M et al (2006) Foxp1 is an essential transcriptional regulator of B cell development. *Nat Immunol* 7:819–826
- Kumar CG, Everts RE, Loor JJ, Lewin HA (2010) Functional annotation of novel lineage-specific genes using co-expression and promoter analysis. *BMC Genomics* 11:161–176
- McLaren DG, Cleveland MA, Deeb N, Forni S, Mileham AJ, Newman S, Southwood OI, Wang L (2013) Genome-wide selection for reproductive traits in swine. In: Rodriguez-Martinez H, Soede NM and Flowers WL (eds) Control of pig reproduction IX, pp 227–242
- Panasiewicz G, Majewska M, Szafranska B (2004) Trophoblastic cDNA cloning of porcine pregnancy-associated glycoprotein genes (pPAG) and *in silico* analysis of coded polypeptide precursors. *Reprod Biol* 4:131–141
- Shu W, Lu MM, Zhang Y et al (2007) Foxp2 and Foxp1 cooperatively regulate lung and esophagus. *Development* 134:1991–2000
- Song MR, Shirasaki R, Cai CL et al (2006) T-Box transcription factor Tbx20 regulates a genetic program for cranial motor neuron cell body migration. *Development* 133:4945–4955
- Szafranska B, Panasiewicz G (2002) The placental expression of the porcine pregnancy-associated glycoprotein (pPAG) gene family examined *in situ* and *in vitro*. *Anim Reprod Sci* 72:95–113
- Szafranska B, Xie S, Green J, Roberts RM (1995) Porcine pregnancy associated glycoproteins: new members of the aspartic proteinase gene family expressed in the trophoblast. *Biol Reprod* 53:21–28
- Szafranska B, Miura R, Ghosh D, Ezashi T, Xie S, Roberts RM, Green J (2001) Gene for porcine pregnancy-associated glycoprotein 2 (poPAG2): its structural organization and analysis of its promoter. *Mol Reprod Dev* 60:137–146
- Szafranska B, Panasiewicz G, Majewska M (2006) Biodiversity of multiple pregnancy-associated glycoprotein (PAG) family: gene cloning and chorionic protein purification in domestic and wild eutherians (Placentalia)—a review. *Reprod Nutr Dev* 5:481–502

- Telugu BP, Walker AM, Green JA (2009) Characterization of the bovine pregnancy-associated glycoprotein gene family—analysis of gene sequences, regulatory regions within the promoter and expression of selected genes. *BMC Genomics* 24(10):185. doi:10.1186/1471-2164-10-185
- Thompson IM, Cerri RLA, Kim IH, Ealy AD, Hansen PJ, Staples CR, Thatcher WW (2011) Effects of lactation and pregnancy on metabolic and hormonal responses and expression of selected conceptus and endometrial genes of Holstein dairy cattle. *J Dairy Sci* 95:5645–5656
- Ushizawa K, Takahashi T, Hosoe M, Hiroko I, Keneyama K, Kizaki K, Hashizume K (2007) Global gene expression analysis and regulation of the principal genes expressed in bovine placenta in regulation to the transcription factor AP2-family. *Reprod Biol Endocrinol* 5:17. doi:10.1186/1477-7827-5-17
- Wallace RM, Pohler KG, Smith MF, Green JA (2015) Placental PAGs: gene origins, expression patterns, and use as markers of pregnancy. *Reproduction* 149:R115–R126
- Wang B, Weidenfeld J, Lu MM et al (2004) Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. *Development* 131:4477–4487
- Xie S, Green J, Beckers JF, Roberts RM (1995) The gene encoding bovine pregnancy-associated glycoprotein-1, an inactive member of the aspartic proteinase family. *Gene* 159:193–197
- Xie S, Green J, Bixby JB, Szafranska B, Demartini JC, Hecht S, Roberts RM (1997) The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes. *Proc Natl Acad Sci U S A* 94:12809–12816