

SEQUENCE CHARACTERIZATION OF INSULIN-LIKE GROWTH FACTOR II (*IGF2*) PROMOTER 2 IN PHILIPPINE NATIVE PIGS (*Sus scrofa* Lin.)

John Vincent A. Garcia^{1*} and Consuelo Amor S. Estrella¹

ABSTRACT

Insulin-like Growth Factor 2 (*IGF2*) is known to be one of the critical drivers of animal growth. Part of the *IGF2* was sequenced to determine if there are DNA variants that differ between the two largest and two smallest Philippine native pig breeds. Samples (n=30) were taken from native pigs reared in Institutional herds from Quezon, Marinduque, Kalinga, and Eastern Samar. The promoter 2 (P2) region was amplified via PCR and sent for sequencing. Seven distinct single-nucleotide polymorphisms were detected. Allelic and genotypic frequencies were computed per genetic and body size groupings. Using Fisher's Exact Test of Independence, the proportion of alleles in SNP g.652 C>G was seen to be significantly different among the four Philippine native pig genetic groups (p-value = 0.0134), and between large and small-sized pig groups (p-value = 0.0168). P2 sequence motif characterization was performed using JASPAR, revealing fourteen (14) transcription factor (TF) binding sites which interact with seven (7) distinct TFs found in pigs, suggesting that base mutations within the proximity of these TF binding sites may influence P2-mediated *IGF2* expression in Philippine native pigs.

Keywords: Philippine native pigs, *IGF2*, single-nucleotide polymorphisms, transcription factors

INTRODUCTION

Philippine native pigs are popularly used for lechon because of their distinct taste and meat quality. However, one of the drawbacks of raising native pigs is their slow growth. At 6 months, the average weights of native pigs from Quezon and Marinduque are 36 kg and 28 kg, respectively, while those from Kalinga and Eastern Samar are lighter at 15 kg and 20 kg, respectively (DOST-PCAARRD, 2022). To sustain the demand for native pigs, it is imperative to improve their body weight and growth, without losing their excellent meat quality characteristics.

One key regulator affecting growth and cell division in many livestock and poultry species is the insulin-like growth factor (IGF) family. It is a group of peptides that includes *IGF1* and *IGF2*, its receptors, and binding proteins. While the IGFs are implicated in both prenatal and postnatal growth, *IGF2* was seen to be an important mitogenic agent in fetal development (Nedbal *et al.*, 2000; Li *et al.*, 2008). Moreover, *IGF2* is maternally imprinted,

¹Institute of Animal Science, College of Agriculture and Food Science, University of the Philippines Los Baños, College Los Baños 4031 Laguna, Philippines (*email: jagarcia17@up.edu.ph)

meaning, only the *IGF2* allele inherited from the father would be expressed due to the epigenetic suppression of the maternal allele (Van Laere *et al.*, 2003).

IGF2 has been identified as a candidate gene for a major QTL affecting muscle mass and deposition in commercial pig breeds (Nezer *et al.*, 1999). A single-nucleotide polymorphism (SNP) 3072 A>G in the 3rd intron of the gene was seen to be associated with increased levels of *IGF2* mRNA in muscular tissues, which in turn affects carcass composition (Van Laere *et al.*, 2003; Oczkiewicz *et al.*, 2009; Stickens *et al.*, 2010). It was reported that the G allele for this polymorphism is associated with increased adipose tissue content in belly and lard carcass cuts, as well as improved intramuscular fat in ham cuts of carrier pigs (Burgos *et al.*, 2012). Variations were also detected on the four promoter sequences of the *IGF2* gene, which were associated with increased intramuscular fat (IMF) deposition in Large White (Aslan *et al.*, 2012). Structural variations in these regions may affect the transcription levels of various genes through the creation of new (or loss of existing) putative transcription factor (TF) binding sites, or by changing the structural properties of the DNA (e.g. duplex stability and minor groove width) which subsequently influences the TF affinity and recognition of the gene (Yella *et al.*, 2018).

The association of SNPs in the *IGF2* gene with growth and carcass quality, therefore, warrants the characterization of its DNA sequence in the two largest and two smallest breeds of Philippine native pigs. Thus, the objectives of the study are to identify structural variants in one of the promoter sequences (promoter 2 or P2), of the *IGF2* that was identified by previous studies as a region with SNPs associated with economically important carcass traits (Aslan *et al.*, 2012) in Philippine native pigs from Quezon, Marinduque, Kalinga and Eastern Samar; Additionally, the study aimed to determine if allelic frequencies in these loci differ among the four genetic groups and between two size classifications. Furthermore, prediction of TF binding site motifs present in the *IGF2* P2 region was also performed to gain insights on how this could impact the growth of Philippine native pigs through potential variations in gene expression as affected by SNPs in the sequence.

MATERIALS AND METHODS

All tissue samples were collected following the approved IACUC Protocol No. CAFS-2018-004. Molecular genetic techniques were performed in the Animal Biotechnology Laboratory (ABL) of the Institute of Animal Science, College of Agriculture and Food Science, University of the Philippines Los Baños.

Tissue samples (ear notch) were collected from Philippine native pigs (unrelated individual males and females ranging from two to three years old) coming from four (4) Institutional native pig breeder farms in Quezon Province (n=10), Marinduque (n=6), Kalinga (n=7), and Eastern Samar (n=7). The pig populations were selected on the basis that they represent the phenotypic extremes of the Philippine Native pigs (Quezon and Marinduque had the largest native pig breeds, while Eastern Samar and Kalinga had the smallest native pig breeds). The pool of animals used in the study was the most distantly related individuals per breed. The tissue samples were kept in absolute ethanol until DNA extraction was performed.

The stored ear notch samples were washed three times with 10x Phosphate buffered saline (PBS). Genomic DNA was extracted using the GF-1 Tissue DNA Extraction Kit (Vivantis, Malaysia), following the manufacturer's instructions. Quantity and quality

confirmation of the stock gDNA was done via UV spectrophotometry (Abs260/Abs280) and agarose gel electrophoresis (1% agarose, 100 volts), respectively.

After quantifying the extracted genomic DNA, a working stock with a concentration of 10 ng/ μ L was derived from the stock solution and used as the template DNA for PCR amplification. Using *Primer3 Plus*, two primer pairs were designed to target different areas in the promoter 2 region (A and B) of *IGF2*. Details regarding the two designed primer pairs for IGF2 P2 are shown in Table 1, while the composition of the separate PCR cocktail mixes used for IGF2_P2A and IGF2_P2B primers is shown in Table 2.

Table 1. PCR primer pairs used for IGF2 P2 fragment amplification.

PRIMER NAME	SEQUENCE (5' TO 3')	LENGTH (bp)	T _m (°C)	PRODUCT SIZE (bp)
IGF2_P2A_F	CCCTGACTTTATGACCCTTGC	21	55.7	540
IGF2_P2A_R	GGCTGATTTCCACGGTTTAG	20	53.8	
IGF2_P2B_F	GAAGCCACCACAACACTACGG	19	56	482
IGF2_P2B_R	GCAAGTCAGAGCAGCCTGAT	20	57.6	

Table 2. Optimized PCR cocktail mix for IGF2_P2A and IGF2_P2B primer pairs.

COMPONENTS	INITIAL CONCENTRATION	IGF2_P2A FINAL CONCENTRATION	IGF2_P2B FINAL CONCENTRATION
SNP water	-	-	-
Buffer A	10x	1x	1x
MgCl2	50 mM	1.7 mM	1.5 mM
dNTPs	25 mM	0.2 mM	0.2 mM
Forward primer	20 μ M	0.5 μ M	0.5 μ M
Reverse primer	20 μ M	0.5 μ M	0.5 μ M
Taq Polymerase	5 U/ μ L	0.4 U	0.4 U
gDNA	10 ng/ μ L	20 ng	20 ng

The PCR reactions for the IGF2_P2A primers follow an initial denaturation step for 5 minutes at 95°C, a 30-second denaturation step at 95°C, an annealing step of 30 seconds at 54°C, an extension step of 30 seconds at 72°C, and a final extension step of 10 minutes at 72°C. The thermal profile for IGF2_P2B primers followed a similar program except that the annealing step was adjusted to 23 seconds at 60°C. All reactions were carried out using a Veriti™ Thermal Cycler 96-well (Thermo Scientific). The PCR products that were amplified during this process were then subjected to agarose gel electrophoresis (2% agarose gel at 100V) for approximately 30 minutes and then visualized via Ultraviolet transillumination (Bio-Rad Gel Doc™ XR+ Systems) to determine the quality of the products.

TriPLICATE PCR runs per sample were pooled to a total volume of approximately

60 µL and sent for Sanger nucleotide sequencing with PCR clean-up services to Apical Scientific Sdn. Bhd. (Selangor, Malaysia). Once the sequence data of all 30 samples (and a reference sequence from a Large White breed pig) were obtained, the chromatograms were assessed for genetic variants or SNPs.

The allelic frequencies of the native pig samples per breed and per size classification were calculated. Using this data, a Fisher's Exact Test of Independence was performed to determine whether there is a significant difference in the proportion of alleles at any polymorphic loci in the *IGF2* P2 region among the four breeds and between the size classifications.

The sequence data obtained was also used to predict the different transcription factor binding sites or motifs, and putative CpG islands. This was done by uploading the gene sequence and maintaining default parameters in the JASPAR 2018 online database (Khan *et al.*, 2018) and the EMBOSS CpGplot online tool, respectively.

RESULTS AND DISCUSSION

Allelic frequencies of SNPs detected in *IGF2* P2

Seven distinct single-nucleotide polymorphisms were detected across all native pig samples (Table 3). Fisher's Exact Test of Independence revealed a significant difference in the proportions of alleles at SNP g.652 among the four breeds (p -value = 0.0134), with the native pigs from Marinduque having a higher frequency of G allele (0.75) at this locus compared to the other genetic groups (0.14 – 0.40). At SNP g.583, the same trend (P =0.0707) was seen as in SNP g.652. Furthermore, Eastern Samar native pigs had equal frequencies of C and T at SNP g.695, while the majority of native pigs had higher C frequencies at this locus. In both Kalinga and Eastern Samar native pig breeds, the C allele was fixed (frequency of 1.0) at locus g.691. Although recorded to be of low allelic frequency, both C and T alleles were still present in Quezon and Marinduque native pigs, thereby showing a distinct, albeit minimal, genetic distinction from the small genetic group. The geographical isolation of the native pig populations may have contributed to these differences, that is, three of which came from different regions in Luzon (Kalinga, Quezon, and Marinduque groups) that are remote from each other, and one genetic group was sourced from the Visayas region (Eastern Samar).

Further classification of the allelic frequencies of the animals into body size revealed that the gene frequencies at SNP g.652 were different (p = 0.0168) between the two groups, while frequencies at SNP g.583 tended to differ (p = 0.0694) between sizes. The 'small-sized' pig group had higher frequencies of G and C at g.652 and g.583 loci, respectively, compared to the 'large-sized' group. Whether the native pigs are classified into genetic groups or size, SNP g.652 was seen to be a significant marker.

Promoter sequence motif characterization

An analysis of the P2 sequence was performed, where putative CpG islands were first identified to determine whether the P2 region is a potential site for epigenetic regulation of *IGF2* in Philippine native pigs. CpG islands on the promoter were predicted using EMBOSS CpGplot (Rice *et al.*, 2000), which, under default parameters of the program, revealed no putative CpG islands in the course of the *IGF2* P2 region. This means that there is a low probability that the P2 region itself is a site for imprinting regulatory mechanisms (such as

Table 3. Allelic frequencies of Philippine native pig samples by geographical origin and size groups.

NO.	LOCUS	GENOTYPE	ALLELIC FREQUENCIES							
			QUEZON	MARINDUQUE	KALINGA	EASTERN S.	P-value	LARGE	SMALL	P-value
1	g.206	T	0.75	0.92	0.86	0.65	0.3725	0.81	0.75	0.7546
		C	0.25	0.08	0.14	0.35		0.19	0.25	
2	g.216	A	0.75	0.92	0.86	0.65	0.3725	0.81	0.75	0.7546
		C	0.25	0.08	0.14	0.35		0.19	0.25	
3	g.255	T	0.65	0.92	0.58	0.65	0.2433	0.75	0.61	0.2752
		G	0.35	0.08	0.42	0.35		0.25	0.39	
4	g.583	G	0.60	0.25	0.72	0.71	0.0707	0.47	0.71	0.0694
		A	0.40	0.75	0.28	0.29		0.53	0.29	
5	g.652*	C	0.60	0.25	0.86	0.71	0.0134	0.47	0.79	0.0168
		G	0.40	0.75	0.14	0.29		0.53	0.21	
6	g.691	C	0.95	0.92	1.00	1.00	0.6836	0.93	1.00	0.4938
		T	0.05	0.08	0.00	0.00		0.07	0.00	
7	g.695	C	0.75	0.83	0.71	0.5	0.3244	0.78	0.61	0.1675
		T	0.25	0.17	0.29	0.5		0.22	0.39	

*P<0.05 is considered significant.

methylation). However, a CpG scan of the entirety of porcine *IGF2* (using reference sequence from GenBank, AY242108) revealed that a putative CpG island is located downstream of the P2 region, which can influence the activity of promoter 2. A CpG island map generated by EMBOSS and the location of the P2 region relative to the predicted CpG Islands is shown (Figure 1). Deletion of this CpG island can elucidate the role of this site in epigenetic regulation of *IGF2*.

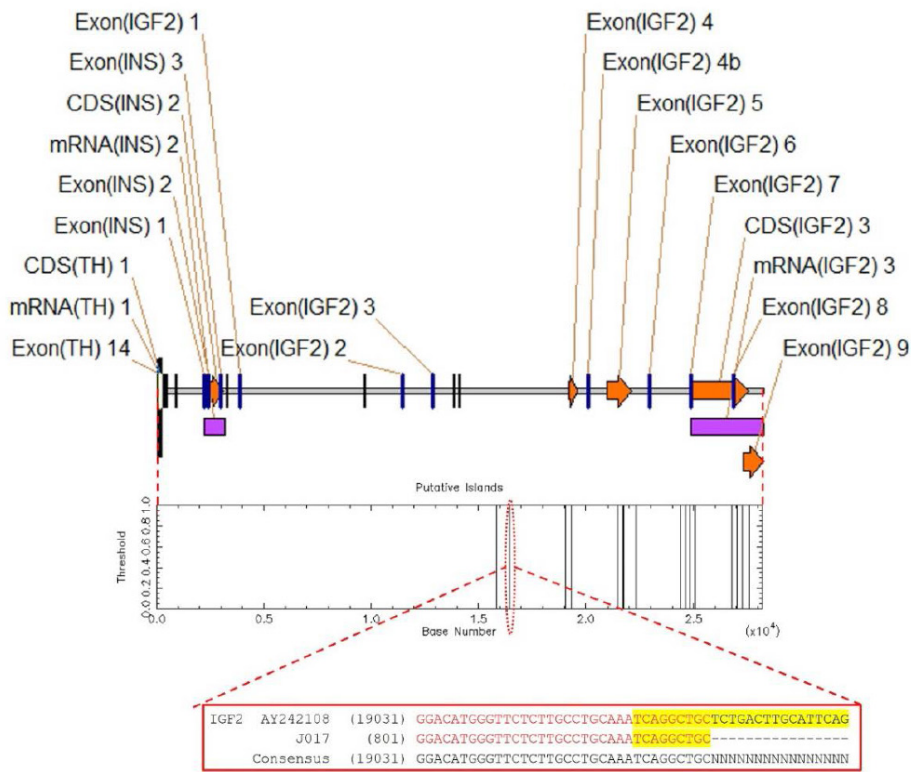


Figure 1. Annotated map of the porcine INS-IGF2 genomic sequence from AY242108 side by side with CpG islands predicted by EMBOSS using default program parameters, and the tail-end of the targeted P2 region (J017) shown as a part (highlighted sequences) of the CpG region that was predicted by EMBOSS placing the CGI immediately downstream of the P2 sequence.

In addition to the nucleotide base variations that were identified from the native pig sequences, and due to the target sequence being a promoter region, the DNA sequence data were also subjected to transcription factor (TF) binding site motif analysis using the online TF database JASPAR (Khan *et al.*, 2018). Transcription factors are proteins that regulate gene expression by binding to specific transcription factor binding sites, where a transcription complex may form before the transcription mechanism is initiated. Transcription factor binding sites were predicted for the Philippine native pig samples by uploading the sequence data and scanning it for domain sequences that are present in the JASPAR CORE Vertebrata database (meaning that these TFs have already been experimentally determined) at an 85% relative profile score threshold (the acceptable threshold when there is limited information

about the gene). At this threshold, 14 transcription factor binding sites were predicted to be present in the *IGF2 P2* region of Philippine native pigs on both the forward and reverse strands (Figure 2).

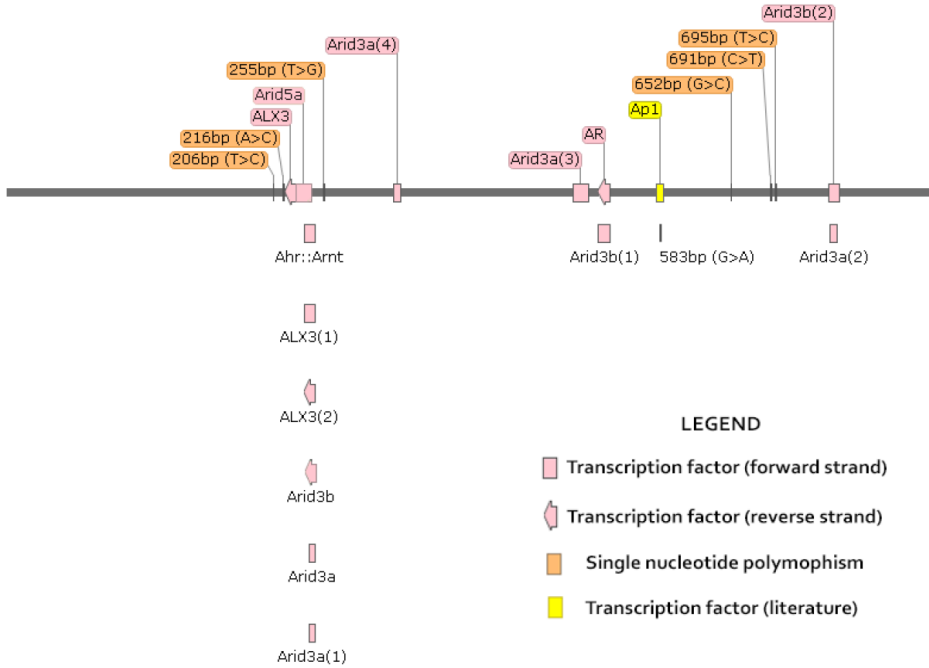


Figure 2. Feature map of *IGF2 P2* region with the relative positions of SNPs and TF binding sites (drawn to scale).

The seven (7) TFs identified through JASPAR to bind with *IGF2 P2* were cross-examined with existing information in GenBank on *Sus scrofa* genomic sequence data and have been determined to be present in the pig genome. These TFs perform important functions in gene regulation. For instance, three of the detected TFs (ARID3a, ARID3b, and ARID5a) belong to the ARID (AT-rich interaction domain) family motifs, which are conserved DNA-binding elements with an important regulatory role in embryogenic development and control of the cell cycle (Patsialou *et al.*, 2005). Specifically, ARID3a is involved in human and mouse placental development (Rhee *et al.*, 2017). In chick and mouse embryos, the transcriptional regulatory function of ARID3b has been associated with normal embryonic limb development (Casanova *et al.*, 2011).

The significance of these predicted transcription factor binding sites, as well as the detected single-nucleotide polymorphisms, can be attributed to transcription factor binding site affinity. Mutations on, or those flanking, TF binding sites along the DNA sequence of the *IGF2* promoter region may lead to variable DNA transcription of the gene by altering the DNA-binding behavior of transcription factors and other regulatory elements either through loss of a binding site, creation of a new site, or through alteration of the conformation and stability of the DNA structure itself (Yella *et al.*, 2018). For instance, in the *IGF2* promoter region, the presence of the A allele (SNP g.583) created an AP-1 TF binding site that is

otherwise absent with the alternative G allele. The creation of this new TF binding site may influence the overall transcription of *IGF2*, although the specific pathway or mechanism through which this is made possible is still unclear (Aslan *et al.*, 2012). In mice, AP-1 transcription factor was responsible for the transactivation of *IGF2* promoters (particularly P3), and *in vitro* mutations in the AP-1 binding sites accounted for the diminished regulatory function of the transcription factor, which consequently led to reduced promoter activity (Caricasole and Ward, 1993).

Given that there are few researches dedicated to elucidating the regulatory elements of porcine genes, and due to the predictive nature of this study, it is difficult to assume the possible effects, or lack thereof, of the detected mutations in the promoter 2 region of *IGF2*, on the pre-natal development and/or intramuscular fat deposition (traits which are highly associated with *IGF2*) of Philippine native pigs. Further studies on the interactions of these regulatory elements in the presence of nucleotide mutations, such as expression level studies of mutant genes compared to wild types or DNA-binding assays, are necessary to determine whether these nucleotide variations are significant in the expression of *IGF2* in the animals which could potentially manifest as a distinct phenotypic trait.

CONCLUSION

This study established the presence of seven (7) single-nucleotide polymorphisms in the promoter 2 region of *IGF2* in four Philippine native pig breeds. The proportions of variants in SNP g. 652 are significantly different among the four native pig populations (p -value = 0.0134) and between the two size classifications (p -value = 0.0168). Furthermore, an analysis of the sequence predicted 14 transcription factor binding sites that correspond to seven (7) distinct transcription factors on the reverse and forward strand of the *IGF2* P2 region, three of which (ARID3a, ARID3b, and ARID5a) are associated with pre-natal development. These findings further enhance the level of understanding of the genetic characteristics of Philippine native pigs about *IGF2*. Specifically, the significant polymorphisms identified in this study may serve as a foundation for future research on the growth and development of Philippine native pigs. However, future studies should involve a larger sample size of native pigs than the present study to provide a more comprehensive understanding of the SNP signatures associated with growth and carcass-related genes in native pigs.

ACKNOWLEDGMENT

The authors would like to extend their sincere appreciation for the help of the BAI-NSPRDC and state university collaborators (Marinduque State College, Kalinga State University, and Eastern Samar State University) from which all animal samples were sourced for this study, as well as the assistance and support of UPLB researchers, Ms. Katrina Umali-Aquino and Ms. Renzen Marie Jamisola.

REFERENCES

- Aslan O, Hamill RM, Davey G, McBryan J, Mullen A, Gispert M, and Sweeney T. 2012. Variation in the *IGF2* gene promoter region is associated with intramuscular fat content in porcine skeletal muscle. *Mol Biol Rep* 39(4):4101–4110.

- Burgos C, Galve A, Moreno C, Altarriba J, Reina R, Garcia C, and Lopez-Buesa P. 2012. The effect of two alleles of *IGF2* on fat content in pig carcasses and pork. *Meat Sci* 90(2):309–313.
- Caricasole A, and Ward A. 1993. Transactivation of mouse insulin-like growth factor II (IGF-II) gene promoters by the AP-1 complex. *Nucleic Acids Res* 21(8):1873–1879.
- Casanova J, Uribe V, Badia Careaga C, Giovinazzo G, Torres M, and Sanz-Ezquerro J. 2011. Apical ectodermal ridge morphogenesis in limb development is controlled by *Arid3b*-mediated regulation of cell movements. *Development* 138(6):1195–1205.
- Department of Science and Technology – Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development [DOST-PCAARRD]. 2022. Philippine native pigs. Retrieved on 09 Apr 2022 from <https://pab-is.pcaarrd.dost.gov.ph/nativepigs/>.
- Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, and Mathelier A. 2018. JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 46(D1):D260–D266.
- Li C, Bin Y, Curchoe C, Yang L, Feng D, Jiang Q, O’neill M, Tian C, and Zhang S. 2008. Genetic imprinting of H19 and *IGF2* in domestic pigs (*Sus scrofa*). *Anim Biotechnol* 19(1):22–27.
- Nedbal S, Zink N, Lahm H, Hoefflich A, and Wolf E. 2000. Functional dissection of the insulin-like growth factor (IGF) system – prospects for animal breeding. *Arch Tierz* 43(3):223–230.
- Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, and Georges M. 1999. An imprinted QTL with major effects on muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nat Genet* 21(2):155–156.
- Oczkowicz M, Tyra M, Walinowicz K, Rozycki M, and Rejduch B. 2009. Known mutation (A3072G) in intron 3 of the *IGF2* gene is associated with growth and carcass composition in Polish pig breeds. *J Appl Genet* 50(3):257–259.
- Patsialou A, Wilsker D, and Moran E. 2005. DNA-binding properties of ARID family proteins. *Nucleic Acid Res* 33(1):66–80.
- Rhee C, Edwards M, Dang C, Harris J, Brown M, Kim J, and Tucker H. 2017. ARID3A is required for mammalian placenta development. *Dev Biol* 422 (2):83–91.
- Rice P, Longden I, and Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277.
- Stickens A, Mathur P, Janssens S, Bruggeman V, Onagbesan O, Schroyen M, Spincemaille G, Decuypere E, Georges M, and Buys N. 2010. Indirect effect of *IGF2* intron3 g.3072A>G mutation on prolificacy in sows. *Anim Genet* 41(5):493–498.
- Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collete C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, and Andersson L. 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 425(6960):832–836.
- Yella V, Bhimsaria D, Ghoshdastidar D, and Rodr A. 2018. Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. *Nucleic Acid Res* 46(22):11883–11897.