SCIENTIFIC REPORTS

OPEN

Quantification of allelic differential expression using a simple Fluorescence primer PCR-RFLP-based method

Changzhi Zhao[1], Shengsong Xie[1,2], Hui Wu[1], Yu Luan[1], Suqin Hu[1], Juan Ni[1], Ruiyi Lin[1], Shuhong Zhao[1,2], Dingxiao Zhang[1,2] & Xinyun Li[1,2]

Allelic differential expression (ADE) is common in diploid organisms, and is often the key reason for specific phenotype variations. Thus, ADE detection is important for identification of major genes and causal mutations. To date, sensitive and simple methods to detect ADE are still lacking. In this study, we have developed an accurate, simple, and sensitive method, named fluorescence primer PCR-RFLP quantitative method (fPCR-RFLP), for ADE analysis. This method involves two rounds of PCR amplification using a pair of primers, one of which is double-labeled with an overhang 6-FAM. The two alleles are then separated by RFLP and quantified by fluorescence density. fPCR-RFLP could precisely distinguish ADE cross a range of 1- to 32-fold differences. Using this method, we verified *PLAG1* and *KIT*, two candidate genes related to growth rate and immune response traits of pigs, to be ADE both at different developmental stages and in different tissues. Our data demonstrates that fPCR-RFLP is an accurate and sensitive method for detecting ADE on both DNA and RNA level. Therefore, this powerful tool provides a way to analyze mutations that cause ADE.

In diploid organisms, the expression level of two alleles are equal for most of the genes, whereas some of them are significantly different. This phenomenon is named allelic differential expression (ADE)[1–3]. Nowadays, ADE occurs because of several reasons, including genomic imprinting[4], X-chromosome inactivation[5], and mutation of regulatory elements[6]. ADE plays an important role in phenotypic determination[7–10]. Therefore, detection of ADE and associated casual mutations in the regulatory elements is important for uncovering the molecular mechanisms of economic traits of farm animals or diseases of human beings[11].

Recently, many approaches have been applied to determine ADE, such as high-throughput DNA/RNA sequencing[12–18], DNA microarray[19–21], Taqman PCR[22,23], allele specific real-time PCR[24], and digital PCR[25]. These methods lead to the detection of major genes associated with human disease, causal mutations in cis-regulatory elements, loss of function alleles, and others[26–29]. Although various bioinformatics tools, including IDP-ASE[30], cisASE[31] and QuASAR[32], have been developed to detect ADE based on high-throughput sequencing data, many factors can still hamper the accuracy of ADE, such as statistical confidence[33], mapping bias[34,35], and the initial reference alignment issues[36]. In particular, mapping bias gives rise to a preferential alignment to the reference allele, which creates a major obstacle in ADE analysis[37]. Microarray can be used to detect ADE, but the noise and the quality of the chip may confound the results. In addition, a standard threshold to estimate significant ADE is still lacking in DNA microarray method, which may produce a certain proportion of false positive results. Furthermore, high-throughput methods have some limitations, including false positive and false negative issues, high consumption of time and cost, and that the potential ADE still needs to be validated by PCR. Current PCR-based methods often use two pairs of primers with difference at the 3′-end nucleotide, which may cause bias due to variation in amplification efficiency. Previously, SNaPshot was frequently used to detect ADE due to its low background noise and high sensitivity[38,39]. This method includes a first multiplexed PCR step, followed

[1]Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction of the Ministry of Education & Key Lab of Swine Genetics and Breeding of Ministry of Agriculture and Rural Affairs, Huazhong Agricultural University, Wuhan, 430070, P.R. China. [2]The Cooperative Innovation Center for Sustainable Pig Production, Huazhong Agricultural University, Wuhan, 430070, P.R. China. Correspondence and requests for materials should be addressed to D.Z. (email: Zhangdingxiao1980@yahoo.com) or X.L. (email: xyli@mail.hzau.edu.cn)
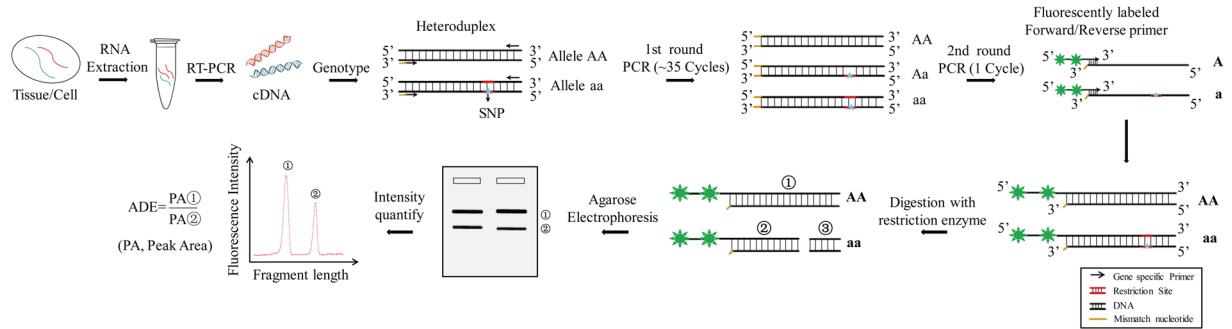
Figure 1. Schematic diagram of Fluorescence PCR-RFLP (fPCR-RFLP) method. Total RNA was extracted from tissues or cells. cDNAs were synthesized by RT-PCR. Heteroduplex was validated by genotyping and then subjected to the first round of PCR (about 35 cycles) using a gene-specific primer. Next, the PCR product was used as template for the second round of PCR with one cycle using one fluorescently labeled primer. This step generated two alleles with over-hang labeling. The labeled PCR products were digested by restriction enzyme. The digested products were separated by agarose gel electrophoresis. Finally, only the fluorescently labeled fragments were detected by a laser detection system. The ratio of peak area represented the ADE.

by a single-base extension reaction by the use of an extension primer with different lengths directly adjacent to a SNP. The one nucleotide extension with distinct fluorescent dideoxynucleotides reflects the SNP allele. Subsequently, alleles were analyzed by capillary electrophoresis and DNA sequencer. By analyses of sequencing data via GeneMapper® software[40,41], the peaks of two alleles representing ADE are recorded. Consequently, SNaPshot is a method with many steps and requires unusual laboratory equipment (e.g., capillary electrophoresis and DNA sequencer), leading to limited dissemination of this technique. Thus, development of a simple efficient, and cost-effective methods is still needed to detect ADE.

Here we have developed a simple and robust method combining RT-PCR and RFLP for detection of ADE, which is named fluorescence primer PCR-RFLP quantitative method (fPCR-RFLP). This method uses only one pair of primers with one of them is double-labeled with an overhang 6-FAM at its 5′ terminal. Two alleles are then separated by RFLP and agarose electrophoresis, followed by gel imaging analysis with a scanner detecting FAM signals. The ratio of ADE is quantified based on the fluorescent intensity of corresponding bands. Recently, we have resequenced a total of 68 pig genomes of four breeds representing Chinese domestic pigs, Western domestic pigs, Asian wild boars, and Western wild pigs, and identified numerous SNPs based on alignment against Swine reference genome (Sus scrofa 10.2). To illustrate the utility of our method, we validated two ADE genes, *PLAG1* and *KIT*, expressed in different tissues and at different developmental stages. Therefore, we establish fPCR-RFLP as a simple, sensitive and accurate method, which could provide a powerful way to detect mutations that cause ADE in many species.

Results

Development of Fluorescence primer PCR-RFLP (fPCR-RFLP) for quantifying ADE.
In the fPCR-RFLP method, the first round of PCR with cDNA as template was performed using gene specific primers. The resulting PCR product was then amplified using a fluorescently labeled primer in the second round of PCR. These PCR products were digested with an appropriate restriction enzyme. Subsequently, restriction fragments were separated by agarose gel electrophoresis. The labeled fragments can be analyzed by peak area ratio.

The detailed workflow of fPCR-RFLP method was illustrated in Fig. 1. The major steps were as follows. To detect ADE, cDNAs were first synthesized by RT-PCR. Second, genotyping was performed to identify heterozygous types. Third, the target fragment of heterozygosity was amplified firstly using universal PCR. There were three types of PCR products, including two kinds of homodimer (AA and aa) and one heterodimer (Aa). Then, the second round of PCR was performed with one cycle using FAM labeled primers and the products of first round PCR as template. Two alleles labeled with FAM (AA-FAM and aa-FAM) were generated after the second round of PCR. Afterward, the two alleles were digested using an appropriate restriction enzyme, followed by gel electrophoresis to detect the AA-FAM and aa-FAM. Fluorescence density of the bands was analyzed using software Fujifilm FLA-5100 Scanner, and the corresponding peak area was considered as the expression level of the two alleles. The ADE was then calculated as the peak area ratio of the two bands.

The fluorescence density is related to the copy number of DNA strand and the ways of FAM labeling.
For the fPCR-RFLP method, the fluorescence density was very important. First, it should be strong enough for detection. Second, it could be representative of the copy number of DNA strand. Therefore, the fluorescence density was analyzed via different ways. First, one FAM (1× FAMer) and two FAM (2× FAMer) labeled primers were serial diluted (from 16× to 1/320× or 1/640×). Serial dilution assay showed a strong correlation between fluorescence density and the copy number of DNA strand ($R^2 = 0.9962$ for 1× FAMer; $R^2 = 0.9969$ for 2× FAMer) (Fig. 2A–C). The raw and normalized fluorescence density was shown in Supplementary Table S2. The raw value was normalized using Log_2. Furthermore, variations in fluorescence density of 8 different types of FAM labeling were tested (Fig. 2D). A plasmid with normal and mutant *GAPDH* genes was used as template (Fig. 2E). The products with two FAM-labeled primer were brighter than those with one FAM-labeled primer. The overhang FAM-labeled primers
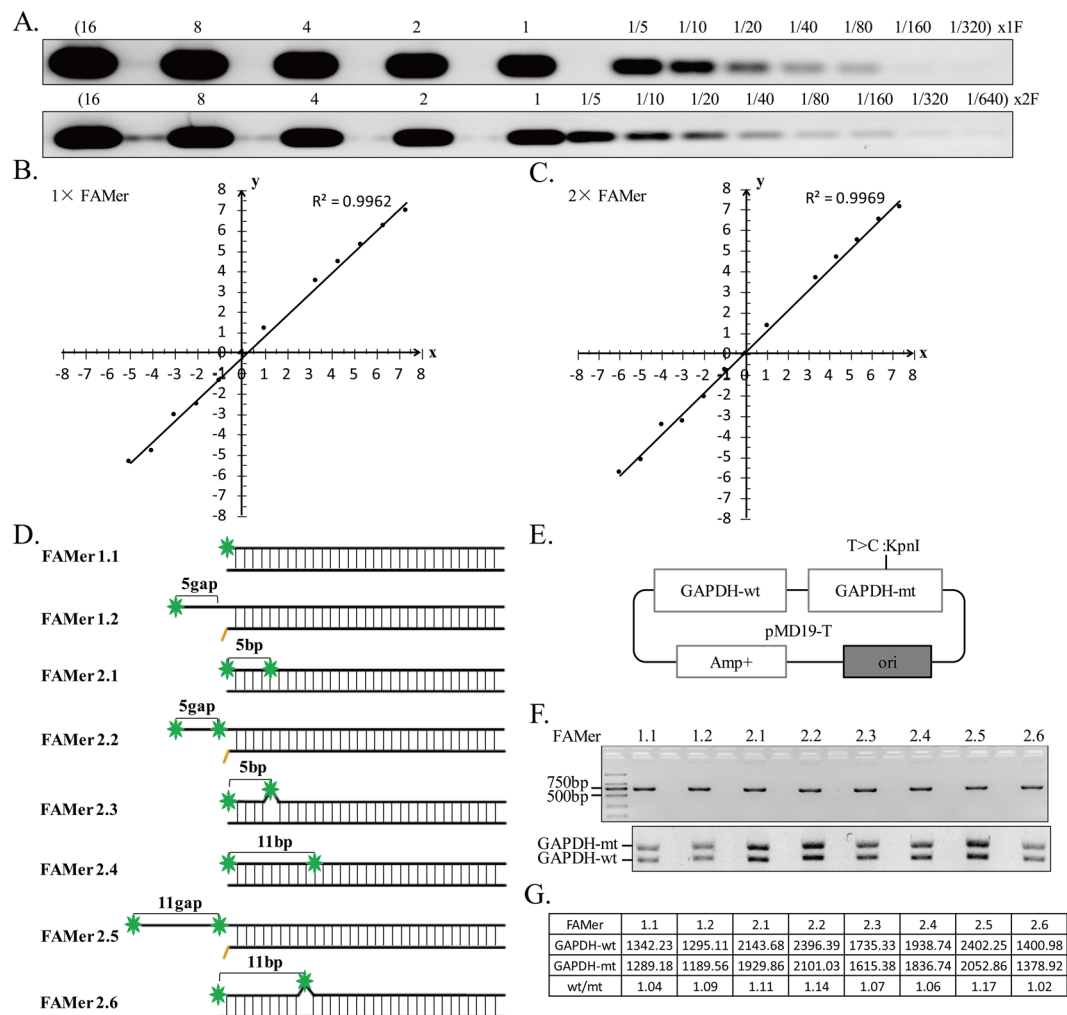
Figure 2. Evaluation of fluorescence density of the PCR products amplified with different FAM labeled primers. (**A**–**C**) Fluorescence density analysis of serial diluted 1× FAMer or 2× FAMer-labeled primers. (**D**) The diagram of PCR products with 8 different types of FAM-labeled primers. Green asterisk represents the 6-FAM-labeled nucleotide. (**E**) A vector containing normal and mutant GAPDH DNA copy was used as PCR template for the fluorescence density analysis. (**F**) The fPCR-RFLP results of 8 different FAM-labeled primers. (**G**) The fluorescence density of wild type (wt) and mutant (mt) GAPDH copies revealed that the ratio of wt/mt was approximately 1.0. Full-length gels are presented in Supplementary Fig. 3.

of FAMer 2.2 and FAMer 2.5 were the brightest two (Fig. 2F,G). FAMer 2.5 had longer overhang, FAMer 2.2 was thus chosen for subsequent experiments. Furthermore, three and five FAM-labeled primers were also tested. The results showed that the fluorescence density was weakened when the space of two FAM was less than 3 nucleotides (see Supplementary Fig. S1). We recommended a 5-bp space between two FAM-labeled nucleotides.

Assessment of the accuracy of fPCR-RFLP method. To assess the accuracy of the fPCR-RFLP method, plasmids containing one normal and one mutant copy of *GAPDH* or *SRY* gene were constructed (Fig. 3A). Then, the DNA level of normal and mutant gene was detected using fPCR-RFLP method. The *Kpn* I and *BamH* I were used to distinguish the normal and mutant alleles. The ratios of normal/mutant allele were 1.02 to 1.08 for *SRY* gene and 1.06 to 1.23 for *GAPDH* (Fig. 3B,C). Furthermore, plasmids with A allele or G allele of *PAPPA2* were constructed and mixed with ratios of 32:1, 16:1, 8:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:8, 1:16, and 1:32 (Fig. 3D,E). The raw and normalized fluorescence densities from three independent experiment were shown in Supplementary Table S3. The fPCR-RFLP results showed a linear relationship ($R^2 = 0.9958$) across a range of 32:1 to 1:32 (Fig. 3F). The raw value was normalized using Log_2. We next used genomic DNA as input to assess the accuracy of fPCR-RFLP. Two genes, *PAPPA2* and *miR-155*, located at the autosome were randomly selected. The DNA copy number of the two alleles of eight heterozygotes was detected using the fPCR-RFLP method. As expected, the ratio of two alleles of *PAPPA2* and *miR-155* gene was 0.92 to 1.04 and 1.00 to 1.19, respectively (Fig. 3G,H). To further validate our results, the second round of PCR products were cloned into PMD19-T vector and 96 bacteria colonies were picked randomly for sequencing. The result showed that the ratio of AA and GG genotypes was 1.05 (see Supplementary Fig. S2A, D). Therefore, we conclude that fPCR-RFLP is an efficient method for detection of ADE genes.
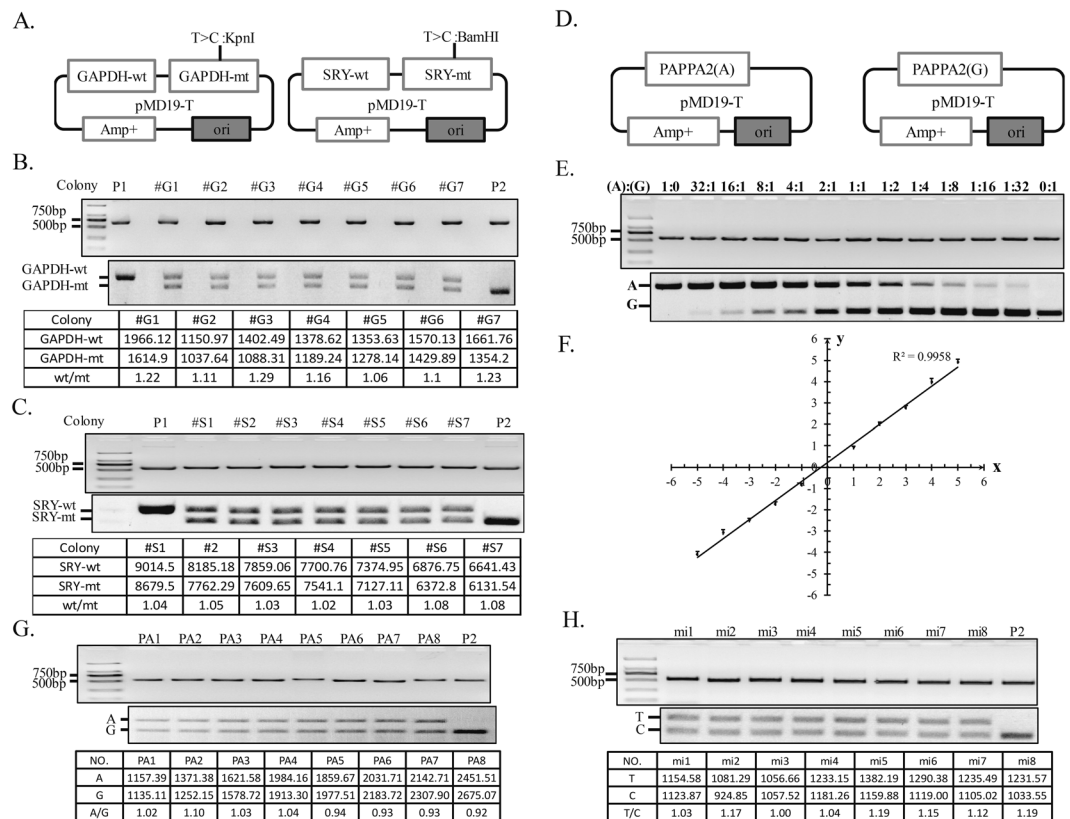
3

Figure 3. Assessment of the accuracy of fPCR-RFLP method. (**A**) The map of two vectors contain normal and mutant GAPDH or SRY gene. (**B,C**) The fPCR-RFLP results of GAPDH and SRY using the constructed plasmids as template. Lanes #G1~#G7 represent 7 independent colonies of GAPDH gene; lanes #S1~#S7 represent 7 independent colonies of SRY gene. (**D**) The map of vectors containing A or G alleles of the PAPPA2 gene. (**E**) The fPCR-RFLP results of PAPPA2 gene with different input ratios of A allele and G allele. (**F**) The linear range of fPCR-RFLP method analyzed based on the results of PAPPA2 gene. Error bars represent S.D. from $n = 3$ replicates. (**G,H**) The results of fPCR-RFLP of PAPPA2 gene and miR-155 using genomic DNA as template. Lanes PA1~PA8 represent 8 independent individuals for PAPPA2 gene; lanes mi1~mi8 represent 8 independent individuals for miR-155 gene; P2 is the positive control of restriction enzyme digestion. Full-length gels are presented in Supplementary Fig. 3.

Validation of ADE gene using fPCR-RFLP method. *KIT* and *PLAG1* are two allelic differentially expressed genes in pigs that were identified through RNA-Seq[42,43]. The expression of the A allele of *PLAG1* gene was significantly higher than that of G allele in spleen and muscle, and was slightly higher in brain tissue of 95-day embryo (Fig. 4A). Similarly, the expression of T allele of the *KIT* gene was significantly higher than that of the C allele in spleen, muscle, and liver, and was slightly higher in brain tissue of 95-day embryo; T allele copy was also overexpressed in adipose, muscle, and spleen tissues of 6-month-old adult pigs (Fig. 4B). We aimed to validate these ADEs using fPCR-RFLP method. Firstly, the tissue distribution of *PLAG1* and *KIT* genes was unraveled using RT-PCR. *KIT* and *PLAG1* were expressed in all 10 tissues examined (Fig. 4C). Then, the mRNA expression of the two alleles and their genomic DNA copy numbers were detected by using the fPCR-RFLP method. Tissues were collected from different individuals. The expression patterns of the two alleles of *PLAG1* gene were different in spleen, muscle, and brain, with expression level of the A allele being about 1.5-fold to 9-fold higher than that of the G allele in these tissues in adult pig (Fig. 4D, upper panel). Similarly, it was 1.3-fold to 2-fold higher in spleen, muscle, and brain tissues of 95-day embryo in three individuals, respectively (Fig. 4D, middle panel). However, the genomic DNA copy number ratio of A and G alleles was 1.00 to 1.08 in eight individuals (Fig. 4D, bottom panel). Statistically, the ratio of A and G alleles in adult and embryo cDNA level is significantly higher than DNA level, indicating ADE (Fig. 4F). For the *KIT* gene, T allele expression was about 2.5-fold to 7-fold higher than that of the C allele at mRNA level in spleen, muscle, and adipose tissues of adult pig in four individuals. It was 1.6-fold to 3-fold in liver, spleen, and muscle tissues of 95-day embryo in three individuals. The DNA copy number of T/C allele was less than 1.33-fold in nine individuals (Fig. 4E). The ratio of T and C alleles in adult and embryo cDNA level is significantly higher than DNA level, again suggesting ADE (Fig. 4G). Furthermore, we validated the results by sequencing of bacterial clones derived from PCR products and revealed a ratio of 3.87 for T/C alleles of *KIT* gene and a ratio of 1.82 for A/G alleles of *PLAG1* gene (see Supplementary Fig. S2B–D).
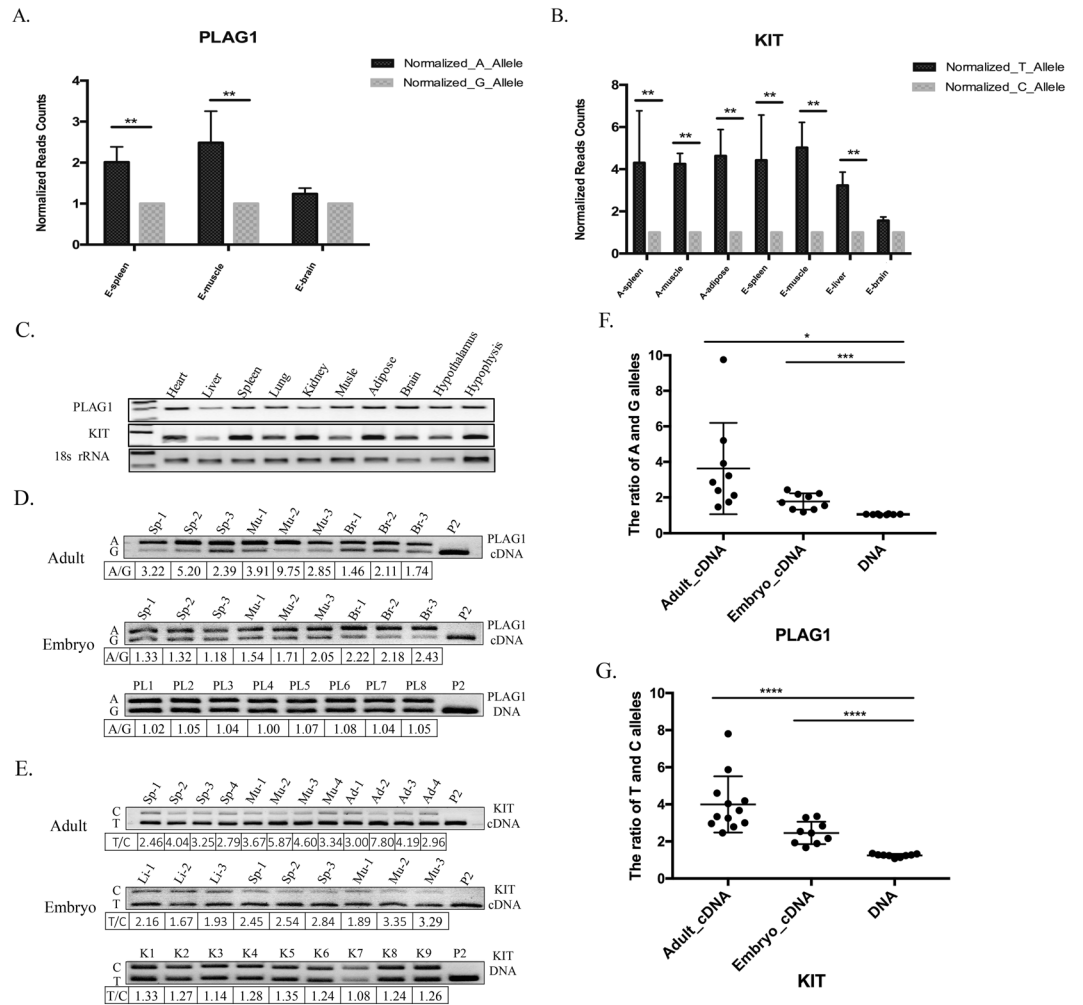
Figure 4. Validation of ADE of PLAG1 and KIT gene using fPCR-RFLP method. (**A,B**) The expression difference of two alleles of PLAG1 and KIT genes detected by RNA-seq in several pig tissues at different developmental stages. A, adult; E, embryo. (**C**) The tissue distribution of PLAG1 and KIT genes. 18 s rRNA used as internal control. (**D,E**) Detection of two ADE genes (PLAG1 and KIT) using fPCR-RFLP from three individuals. Sp, spleen; Mu, Muscle; Br, brain; Li, liver; Ad, adipose. Lanes PL1~PL8 represent the fPCR-RFLP results of PLAG1 gene in 8 independent individuals. Lanes K1~K9 represent the fPCR-RFLP results of KIT gene in 9 independent individuals. P2 is the positive control of restriction enzyme digestion. (**F**) The ratio of A and G allele of PLAG1 gene at cDNA and DNA level. Data were pooled from D, and t test was used to check statistics. (**G**) The ratio of T and C allele of KIT gene at cDNA and DNA level. Data were pooled from E, and t test was used to check statistics. Full-length gels are presented in Supplementary Fig. 3.

Discussion

In this study, we have developed a robust method for detecting ADE. This method is simple, highly accurate, and sensitive. In this method, the quantitative analysis was performed after the PCR reaction. A previous quantitative analysis was mainly performed at early or middle stages of PCR reaction[44]. Moreover, in this method, only one pair of primers was used, which could avoid potential bias caused by amplification efficiency rendered by different primers. Particularly, a second round of PCR with only one cycle was employed to eliminate interference from the heterodimer DNA molecule. Classical RFLP methods were used to distinguish the two alleles. For the SNP sites that could not be recognized by any restriction enzyme, a non-FAM labeling mismatch primer was used to artificially create a restriction site in the amplified product to allow the following RFLP analysis[45,46]. To increase the fluorescence density, we designed an overhang labeling primers to reduce the cross fluorescence interference between two FAM. These are the major differences between our method and previous quantitative methods[44,47].

Our method exhibits a perfectly positive linear relationship between fluorescence density and the number of DNA molecule copies. This confirms that the number of molecules could be quantified using fluorescence density, and increasing the fluorescence density of a single molecule may improve the sensitivity of this method. Therefore, we tested different types of labeling to search for increasing fluorescence density. Interestingly, fluorescence interference was observed when two FAMs had less than three nucleotides span (see Supplementary Fig. S1). Also, we found that fluorescence density was significantly increased after heat denaturation (see Supplementary Fig. S1C).

A possible reason for that is that heat denaturation creates distance between two FAMs to reduce fluorescence interference. Therefore, the over-hang labeling primers were designed, and as expected, the fluorescence density was stronger than other labeling types. For over-hang labeling, the 3′ terminal of the template DNA in the second round of PCR contains at least one nucleotide mismatch with the FAM labeling primer (Fig. 1). A 5-bp gap between two FAM labeled nucleotides was recommended in this study. Our data showed that it could precisely quantify the ADE within a range of ratio of 1:32 to 32:1.

Taqman PCR and allele-specific real-time PCR are commonly used for ADE detection[48,49]. However, these methods suffer a major problem, i.e., it is difficult to distinguish two alleles on the same locus[22–24]. Besides, two pairs of primers are used for these methods, which cause bias due to different amplification efficiency. Comparatively, our method can easily distinguish two alleles based on PCR-RFLP. Also, only one pair of primer is used, which could avoid the bias of amplification by using different primers.

Further, the allelic expressions of PLAG1 and KIT were studied. They are two candidate genes associated with growth and MCH traits based on GWAS studies[42,43]. Our fPCR-RFLP results showed that the ratio of the two alleles is 1:1 at DNA level, whereas it is significantly different at mRNA level, confirming that they are ADE genes, at least in our analysed individuals. This illustrates fPCR-RLFP as a reliable method to detect ADE gene. Of note, we notice that the ADE results in the same tissue type differ in different individuals (Fig. 4D,E). This would highlight that population heterogeneity (i.e., biological variation) is likely the main cause of variation in results, but other factors that impact on gene expression and splicing may also contribute, such as the abundance of cis-acting loci[50], DNA methylation[51], inheritance of allele-specific epigenetic marks[52], and different matrix[53]. Further studies are needed to investigate whether these SNPs are casual mutations located within regulatory regions and the ADE of PLAG1 and KIT are responsible for variations linked to the traits.

Methods and Materials

Plasmid construction.   Vectors with normal and mutant copies of GAPDH or SRY gene were constructed by In-Fusion® HD Cloning Kit (Clontech). The normal copy sequences were amplified by PCR using pig genomic DNA. The mutant copy sequences were generated by point mutation strategy. In-Fusion PCR primers were designed by an online tool (http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Selection_Guides/Online_In-Fusion_Tools). Recombination products were ligated to PMD19-T vector by TA cloning. Then, ligation mixture was transformed into Trans1-T1 competent cells (Transgen). The plasmids were extracted by using TaKaRa MiniBEST Plasmid Purification Kit (TaKaRa). All gene specific primers and 6-FAM labeled primers are diluted to a final concentration of $10\,\mu M$ and listed in Supplementary Table S1.

Total RNA extraction and reverse transcription.   Tissues were disrupted in TransZol Up (Transgen) with a homogenizer. RNA was extracted according to the manufacturer's instructions. cDNAs were synthesized using the PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa). The cDNA products were used as PCR template to identify heterozygous individuals and to quantify allele expression.

Fluorescence PCR-RFLP.   The first round of PCR was performed with $\sim$200 ng of cDNA template and $10\,\mu M$ of gene-specific forward and reverse primers in a $50\,\mu L$ reaction using Premix Taq™ (Takara). PCR reaction was performed in a Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific) programmed for one cycle of 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C, with a final extension step for 5 min at 72 °C. PCR products ($23\,\mu L$) were used as template for the second round of PCR in a $50\,\mu L$ reaction with $25\,\mu L$ Premix Taq™ and $15\,\mu M$ FAM-labeled primers. The PCR reaction program was set as one cycle of 5 min at 95 °C, followed by one cycle of 30 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C, with a final extension step for 5 min at 72 °C. Fluorescence PCR products ($25\,\mu L$) were digested with 1 unit of the appropriate restriction enzyme in a $30\,\mu L$ reaction. All digestion products were subjected to agarose gel electrophoresis with $6\times$ loading buffer (TaKaRa) without nucleic acid dye. Finally, fluorescent image analysis was performed using Fujifilm FLA-5100 Scanner (Fujifilm). The Fluorescent density of each bands was analyzed with image analysis software Multi Gauge (Fujifilm).

Ethics approval and consent to participate.   Animals care and all the experimentation in this study were carried out in accordance with the pre-approved guidelines from the Standing Committee of Hubei People's Congress. All experimental protocols were approved by the Ethics Committee of Huazhong Agricultural University, Wuhan City, Hubei Province, P. R. China.

References

1. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nat Genet* **32**, 261–266 (2002).
2. Enard, W. *et al*. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
3. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297**, 1143 (2002).
4. Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* **3** (2011).
5. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).
6. Sadee, W. Measuring cis-acting regulatory variants genome-wide: new insights into expression genetics and disease susceptibility. *Genome Medicine* **1**, 116 (2009).
7. Wang, D., Guo, Y., Wrighton, S. A., Cooke, G. E. & Sadee, W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J* **11**, 274–286 (2011).
8. Smith, R. M. *et al*. Multiple regulatory variants modulate expression of 5-hydroxytryptamine 2A receptors in human cortex. *Biol Psychiatry* **73**, 546–554 (2013).
9. Sullivan, D. *et al*. Dopamine transporter DAT and receptor DRD2 variants affect risk of lethal cocaine abuse: a gene-gene-environment interaction. *Transl Psychiatry* **3**, e222 (2013).

10. Barrie, E. S. *et al*. Regulatory polymorphisms in human DBH affect peripheral gene expression and sympathetic activity. *Circ Res* **115**, 1017–1025 (2014).

11. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat Rev Genet* **7**, 862–872 (2006).

12. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res* **21**, 1728–1737 (2011).

13. Xu, X. *et al*. Next-generation DNA sequencing-based assay for measuring allelic expression imbalance (AEI) of candidate neuropsychiatric disorder genes in human brain. *BMC Genomics* **12**, 518 (2011).

14. Leon-Novelo, L. G., McIntyre, L. M., Fear, J. M. & Graze, R. M. A flexible Bayesian method for detecting allelic imbalance in RNA-seq data. *BMC Genomics* **15**, 920 (2014).

15. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).

16. Reuter, J. A., Spacek, D. V., Pai, R. K. & Snyder, M. P. Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat Methods* **13**, 953–958 (2016).

17. Benitez, J. A., Cheng, S. & Deng, Q. Revealing allele-specific gene expression by single-cell transcriptomics. *Int J Biochem Cell Biol* **90**, 155–160 (2017).

18. Jiang, Y., Zhang, N. R. & Li, M. SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol* **18**, 74 (2017).

19. Bjornsson, H. T. *et al*. SNP-specific array-based allele-specific expression analysis. *Genome Res* **18**, 771–779 (2008).

20. Serre, D. *et al*. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet* **4**, e1000006 (2008).

21. Liu, R. *et al*. Allele-specific expression analysis methods for high-density SNP microarray data. *Bioinformatics* **28**, 1102–1108 (2012).

22. Ferstl, B. *et al*. Allele-specific quantification of HLA-DQB1 gene expression by real-time reverse transcriptase-polymerase chain reaction. *Genes Immun* **5**, 405–416 (2004).

23. Fogarty, M. P., Xiao, R., Prokunina-Olsson, L., Scott, L. J. & Mohlke, K. L. Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. *Hum Mol Genet* **19**, 1921–1929 (2010).

24. Pan, N. *et al*. Quantification of classical HLA class I mRNA by allele-specific, real-time polymerase chain reaction for most Han individuals. *HLA* **91**, 112–123 (2018).

25. Wiseman, D. H. & Somervaille, T. C. P. Nanofluidic Allele-Specific Digital PCR Method for Quantifying IDH1 and IDH2 Mutation Burden in Acute Myeloid Leukemia. *Methods Mol Biol* **1633**, 235–255 (2017).

26. Kukurba, K. R. *et al*. Allelic expression of deleterious protein-coding variants across human tissues. *PLoS Genet* **10**, e1004304 (2014).

27. Buil, A. *et al*. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47**, 88–91 (2015).

28. MacArthur, D. G. *et al*. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

29. Maroilley, T. *et al*. Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* **18**, 967 (2017).

30. Deonovic, B., Wang, Y., Weirather, J., Wang, X. J. & Au, K. F. IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. *Nucleic Acids Res* **45**, e32 (2017).

31. Liu, Z. *et al*. cisASE: a likelihood-based method for detecting putative cis-regulated allele-specific expression in RNA sequencing data. *Bioinformatics* **32**, 3291–3297 (2016).

32. Harvey, C. T. *et al*. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics* **31**, 1235–1242 (2015).

33. Heap, G. A. *et al*. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet* **19**, 122–134 (2010).

34. Kraig, R. S., Joseph, D. C. & Patricia, J. W. Sources of bias in measures of allelespecific expression derived from RNA-sequence data aligned to a single reference genome. *BMC Genomics* **14**, 536 (2013).

35. Munger, S. C. *et al*. RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. *Genetics* **198**, 59–73 (2014).

36. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**, 1061–1063 (2015).

37. Nikolaos, I. P., Maria, G. A., Emmanouil, T. D. & Tuuli., L. Allelicmapping bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome biol* **15**, 467 (2014).

38. Wilkins, J. M. *et al*. Extreme context specificity in differential allelic expression. *Hum Mol Genet* **16**, 537–546 (2007).

39. Ciobanu, D. C. *et al*. Detection, validation, and downstream analysis of allelic variation in gene expression. *Genetics* **184**, 119–128 (2010).

40. Dias-Santagata, D. *et al*. Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Mol Med* **2**, 146–158 (2010).

41. Collier, P. *et al*. Validation of standards for quantitative assessment of JAK2 c.1849G > T (p.V617F) allele burden analysis in clinical samples. *Genet Test Mol Biomarkers* **17**, 429–437 (2013).

42. Jung, E. J. *et al*. Genome-wide association study identifies quantitative trait loci affecting hematological traits in an F2 intercross between Landrace and Korean native pigs. *Anim Genet* **45**, 534–541 (2014).

43. Zhu, J. *et al*. A systems genetics study of swine illustrates mechanisms underlying human phenotypic traits. *BMC Genomics* **16**, 88 (2015).

44. Germer, S., Holland, M. J. & Higuchi, R. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res* **10**, 258–266 (2000).

45. Hingorani, A. D. & Brown, M. J. A simple molecular assay for the C1166 variant of the angiotensin II type 1 receptor gene. *Biochem Biophys Res Commun* **213**, 725–729 (1995).

46. Russ, A. P., Maerz, W., Ruzicka, V., Stein, U. & Gross, W. Rapid detection of the hypertension-associated Met235—>Thr allele of the human angiotensinogen gene. *Hum Mol Genet* **2**, 609–610 (1993).

47. Knapp, S., Meghjee, N., Cassidy, S., Jamil, K. & Thursz, M. Detection of allele specific differences in IFNL3 (IL28B) mRNA expression. *BMC Med Genet* **15**, 104 (2014).

48. Livak, K. J., Marmaro, J. & Todd, J. A. Towards fully automated genome-wide polymorphism screening. *Nat Genet* **9**, 341–342 (1995).

49. Ranade, K. *et al*. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res* **11**, 1262–1268 (2001).

50. Yeo, S. *et al*. The abundance of cis-acting loci leading to differential allele expression in F1 mice and their relationship to loci harboring genes affecting complex traits. *BMC Genomics* **17**, 620 (2016).

51. Kerkel, K. *et al*. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* **40**, 904–908 (2008).

52. McDaniell, R. *et al*. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).

53. McDermaid, A., Monier, B., Zhao, J., Liu, B. & Ma, Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief Bioinform* (2018).

Acknowledgements

Author Contributions

X.L., C.Z., D.Z. and S.Z. conceived of the fPCR-RFLP method. X.L., S.Z. and C.Z. planned experiments. C.Z., H.W., S.H. and J.N. performed experiments. Y.L. and S.X. performed bioinformatics analysis of the data of RNA-Seq C.Z., X.L., S.X., R.L., D.Z. and S.Z. wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-42815-5.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.