Gene 641 (2018) 367–375

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Variant calling from RNA-seq data of the brain transcriptome of pigs and its application for allele-specific expression and imprinting analysis

Maria Oczkowicz*, Tomasz Szmatoła, Katarzyna Piórkowska, Katarzyna Ropka-Molik

National Research Institute of Animal Production, Department of Genomics and Molecular Biology of Animals, Krakowska 1, 32-083 Balice, Poland

ARTICLE INFO

Keywords:
RNA-seq
Allele-specific expression
Imprinting
Pigs
Variant calling
SNPs

ABSTRACT

Identification of new polymorphic variants from RNA-seq data is difficult mainly because of the errors arising during bioinformatic analysis. Therefore, new experiments in this area are very profitable for improving new statistical methods. In our study of the porcine brain transcriptome, we have identified 10966 polymorphic variants, among which 7277 were single nucleotide polymorphisms (SNPs). Further, we have calculated allelic ratios for the SNPs identified and estimated that 52% of genes in porcine brain are subjected to allele-specific expression (ASE), a phenomenon in which one allele is preferentially expressed. Our investigation presents the first estimates of ASE in porcine brain. In addition, we have used the results of RNA-seq for the identification of SNPs in putatively imprinted genes. Finally, we have used these SNPs for the verification of the imprinted status of the *INPP5f variant 2*, *LRRTM1* and *HM13* genes in pigs by Sanger sequencing. We observed that *INPP5f variant 2* is paternally expressed, while *HM13* and *LRRTM1* are biallelically expressed in porcine brain. We have also confirmed maternal expression of the *MEG3* gene in pigs. Our results present how RNA-seq data may be used for imprinting studies without sequencing of parental genomes.

1. Introduction

RNA-seq is a cutting-edge technology that enables one to analyse gene expression level and identify different isoforms and polymorphic variants. It has also been used for the identification of new imprinted genes in mice (Wang et al., 2008; Gregg et al., 2010; DeVeale et al., 2012) and humans (Metsalu et al., 2014). However, variant calling from RNA-seq data is challenging despite rapid development of new methodologies for accurate identification of polymorphisms (Shen et al., 2013; Lu et al., 2017). Using different statistical approaches and different cut-offs during bioinformatic analysis of RNA-seq data leads to very high discrepancies between different experiments. For example, studies conducted by Gregg et al., in 2010 revealed that the number of imprinted genes in mice had been ~10-fold underestimated, while further experiments did not confirm these results and suggested that the number of imprinted genes was not dramatically greater (DeVeale et al., 2012). RNA-seq has also been used for investigation of genomic imprinting in chicken embryos (Frésard et al., 2014). Initially, 79 potentially imprinted genes were identified; however, none of them was confirmed by pyrosequencing. Absence of imprinting was also revealed by high throughput sequencing in one day old brains of chicken (Wang et al., 2015). Moreover, confusing results were observed for RNA editing sites. In human B-cells over 10,000 exonic editing events have been identified by RNA-seq (Li et al., 2011) Soon after it was showed that part of them – so called non-canonical editing is an artifact of current high-throughput sequencing technology (Gu et al., 2012; Kleinman and Majewski, 2012). Meanwhile, intensive attention has been put on ASE (Allele Specific Expression) – a process triggered by genetic or epigenetic variation in cis in which in heterozygous locus one of the alleles is significantly higher expressed than the other (Crowley et al., 2015). The situation when one of the alleles is completely silenced is called monoallelic expression and may be the result of imprinting, random monoallelic expression or X-inactivation (Metsalu et al., 2014). Recent studies show that very often both alleles are expressed in imprinted genes, but they differ in the level of expression (Bonthuis et al., 2015). It was showed, by RNA-seq, that there are hundreds of “noncanonical “imprinting effects in the mouse genome (Bonthuis et al., 2015).

Next-generation sequencing (NGS) techniques provide much new information on transcriptomes and may revolutionize our understanding of gene expression in the future. However, contradictory results and discordant estimations about the ASE and imprinting phenomena indicate an urgent need for further investigations in this area. Including farm animals in these investigations enables the

Abbreviations: cDNA, DNA complementary to RNA; ASE, Allele Specific Expression; RNA-seq, RNA-sequencing; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism

* Corresponding author.
E-mail address: maria.oczkowicz@izoo.krakow.pl (M. Oczkowicz).

http://dx.doi.org/10.1016/j.gene.2017.10.076
Received 27 June 2017; Received in revised form 19 October 2017; Accepted 26 October 2017
Available online 27 October 2017

determination of whether the ASE phenomenon is universal among mammalian species. In our study, we have used RNA-seq to identify 10,966 genetic variants and tried to assess the range of ASE in the porcine brain transcriptome. We have also collected RNA-seq data on several imprinted genes. Moreover, we have confirmed our results for the *INPP5F variant. 2*, *MEG3*, *HM13* and *LRRTM1* genes by Sanger sequencing.

2. Material and methods

2.1. Animals for the study

Twelve adult pigs were slaughtered at 100 kg of weight, and brain and blood samples were taken for RNA and DNA analyses, respectively. We have decided to investigate brain tissue since many ASE studies were performed on this tissue. The animals (4 males and 8 females) belonged to the Polish 990 synthetic line of pigs, which originated after the crossing of several breeds (Large White, Belgium Landrace, Duroc, German Landrace, Walsh Landrace and Hampshire) and were the offspring of three dams and two sires. Blood samples were taken from the parents for DNA analysis. The animals were maintained at the Testing Station of the National Research Institute of Animal Production in Pawłowice under the same housing and feeding conditions. Tissue samples were frozen in liquid nitrogen and kept at − 80 °C until RNA isolation.

2.2. RNA and DNA isolation

RNA from the brain was isolated using the PureLink RNA Isolation Kit (ThermoFisher Scientific, Waltham, USA). DNA was isolated from the blood of the offspring and parents using the Wizard DNA Genomic Purification Kit (Promega, Madison, USA). The quality and the quantity of RNA were evaluated using Tapestation (Agilent, Santa Clara, USA) and RNA ScreenTape 2200 (Agilent). The RIN (RNA Integrity Number) of all the samples was > 8.

2.3. cDNA library preparation

cDNA libraries were synthesized using 300 ng of total RNA using the TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, USA) according to the protocol for each sample. The concentrations of the cDNA libraries obtained were estimated by the Qubits 2.0 Fluorometer (ThermoFisher Scientific, Waltham, USA). Then, normalized sample libraries (diluted to 10 nM with Tris–Cl) were pooled and sequenced on a HiScanSQ System (Illumina, San Diego, USA). Two technical replicates were performed for each library. Flow-cell clustering was performed using the TruSeq SR Cluster Kit v3-cBot-HS (Illumina, San Diego, USA) on a cBot Instrument (Illumina, San Diego, USA). Sequencing by synthesis was performed using TruSeq SBS Kit v3- HS chemistry (50 single-end cycles) (Illumina, San Diego, USA).

2.4. cDNA synthesis

cDNA from the brain samples was synthesized using the cDNA Archive kit (ThermoFisher Scientific, Waltham, USA) and 1000 ng of RNA, according to the attached protocol.

2.5. Sanger sequencing

Several SNPs identified by RNA-seq technology were analysed by Sanger sequencing. DNA from the parents and from the offspring was used to perform PCR to amplify fragments of the *INPP5F* gene containing two SNPs identified by RNA-seq. Primers for the reaction are presented in Table 1. The PCR was performed using the AmpliTaq 360° MasterMix (ThermoFisher Scientific, Waltham, USA) according to the attached protocol. The sequencing reaction was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific, Waltham, USA) according to the attached protocol after the cleaning of the PCR products with ExoSAP (Affymetrix, Santa Clara, USA). Sequencing products were cleaned using BigDye XTerminator™ Purification Kit (ThermoFisher Scientific, Waltham, USA) and sequenced on a 3130 Genetic Analyser automatic sequencer (ThermoFisher Scientific, Waltham, USA).

Table 1
Sequences of primers used for Sanger sequencing.

MEG3_F_cggagtgctgtgggagaata
MEG3_R_tgcttctgccttgttgctct
H13RNA_F_agtggccaagtcctttgagg
H13RNA_R_cacccgctgggcattacat
H13DNA_F_tagccagtccctcctcaaga
H13DNA_R_ctcctgggcacccttaaagc
LRRTM1_F_gcagatccacaaggtggtca
LRRTM1_R_tcctgaagattgtggggtgg
INPP5FDNA_F_accaacctcaataagagttttcac
INPP5FDNA_R_acaaaaactgtgggcctgttg
INPP5FRNA_F_ttttctcgactgcctgttatg
INPP5FRNA_R_gaagactcgtggccttttcc

2.6. Read mapping and variant calling

After demultiplexing and generating reads (50 bp length) for each sample, quality control was performed using FastQC software (Andrews, 2010). Flexbar software (Dodt et al., 2012) was utilized to remove Illumina adapters, reads shorter than 35 bp and those that had quality under 20. TopHat software (Trapnell et al., 2009) was then used to map pruned sequences to the *Sus scrofa* genome (SusScr3). In addition, after obtaining the mapping results, Picard software was used to mark the duplicates and to add unique read groups to each bam file. To check the quality of the mapping results obtained, SAMStat software (Lassmann et al., 2011) was utilized. Variant calling was performed on a whole population using FreeBayes software (Garrison and Marth, 2012). Variant filtration and annotation were maintained with SnpEff and SnpSift software (Cingolani et al., 2012a; Cingolani et al., 2012b). We filtered variants that did not pass the proposed criteria read depth above 17 and quality and genotype quality above 20. As an additional filter, we used GATK software (McKenna et al., 2010) to filter clusters of at least 3 SNPs within a 35 bp window. Ensemble database was used as a reference of known variants. A chi square test was used to assess deviation from the expected 0.5 reference/alternate ratio. Variants with p value < 0.05 and allelic ratios as described in Results and discussion section were considered as allele specific expression variants.

3. Results and discussion

3.1. Mapping statistics

After demultiplexing, 16,110,531–25,216,613 raw reads per sample were filtered with the use of Flexbar, which led to the generation of 15,892,535 to 24,899,082 reads. Approximately, 80–85% of these reads were mapped to the *Sus scrofa* 3 genome.

Detailed information regarding raw reads per sample, filtered reads and mapped reads is presented in Table 2.

3.2. Identification of variants

FreeBayes software was used to call variants that were then filtered and annotated with the use of SnpSift and SnpEFF software. We identified 13,361 variants (5873 new and 7488 previously identified) across all samples. After applying a filter for eliminating SNP clusters, the number of variants decreased to 10,966 (4632 new and 6334 previously identified). Among them, 7277 (1082 new and 6195 previously

Table 2
Basic statistics of reads.

Samples	Statistics of reads					
	Number of raw reads	Number of filtered reads	Number of mapped reads	Number of reads mapped with mapping quality ≥ 30	Number of reads mapped to exons	Percent of reads mapped to exons (%)
1	20,987,763	20,712,545	18,566,584	16,270,602	10,536,968	64.7
2	16,110,531	15,892,535	14,394,146	12,607,619	7,913,985	62.8
3	19,096,078	18,844,532	17,224,696	15,262,118	9,776,716	64.1
4	25,004,607	24,684,538	22,160,640	19,235,242	12,800,326	66.5
5	24,801,364	24,494,039	21,852,736	18,926,248	12,334,485	65.2
6	16,353,732	16,171,656	14,509,773	12,461,188	8,033,002	64.5
7	22,990,662	22,700,465	20,628,752	18,068,312	11,525,222	64.1
8	20,033,717	19,775,392	17,890,152	15,637,867	10,092,559	64.5
9	21,595,388	21,328,806	19,262,316	16,846,730	11,233,974	66.7
10	25,216,613	24,899,082	22,528,082	19,747,276	12,666,232	64.1
11	23,199,782	22,906,090	20,771,872	18,336,028	11,946,922	65.2
12	22,061,996	21,788,372	19,793,992	17,353,786	11,049,271	64.0

identified) were single nucleotide polymorphisms (SNPs), 1745 were insertions (1728 new and 17 previously identified), and 1251 were deletions (1168 new and 83 previously identified). The remaining 693 were variants with several possible genotypes between samples. For further statistical analysis, we used only SNPs with one possible genotype between samples to decrease the risk of false-positive variants, and only this portion of the variants was submitted to dbSNP NCBI (available at https://www.ncbi.nlm.nih.gov/projects/SNP/snp_viewBatch.cgi?sbid=1062409).

3.3. Statistics for identified SNPs

We divided selected SNPs (7277 in 2510 genes) into five groups depending on genotype and allelic ratio (Table 3). To increase the reliability of the analysis, we made the assumptions that variants must be present in at least three individuals, and other types of variants or genotypes are allowed at this locus. The assumptions were the same for all groups of variants. The first group (heteroref) was the heterozygous variants in which the reference allele was predominant (ratio of reference to alternate allele was between 100 and 1.35). We identified 1751 such variants in 968 genes. The mean number of individuals for all variants was 1.83. Variants from the second group (heteroalt) were the heterozygous variants in which the alternate allele was predominant (ratio of reference to alternate allele was between 0.01 and 0.65). The number of variants in this group was much smaller (534 in 382 genes), as was the mean number of individuals with these variants (0.72). There were 46 variants common to the heteroref and heteroalt groups. The third group (50:50) was the “true heterozygotes” (allelic ratio was between 0.65 and 1.35). The number of this kind of variant was 2317 in 1055 genes. The number of variants with two alternate homozygotes was 4068 in 1685 genes for the reference homozygote and 2162 in 1048 genes for the alternate homozygote. Furthermore, we have presented the numbers of SNPs of each category in each of 12 samples regardless of the number of individuals in which they were observed in Table 4.

Table 3
Number of identified SNPs in each category with the assumptions that SNP must be present in at least three individuals. Heteroref - heterozygous variants in which reference allele was predominant (ratio of reference to alternate allele was between 100 and 1,35); heteroalt - heterozygous variants in which alternate allele was predominant (ratio of reference to alternate allele was between 0,01 and 0,65); and 50:50 - the “true heterozygotes” (allelic ratio was between 0,65 and 1,35). Ref - reference homozygote, alt – alternate homozygote, rs# - SNPs submitted previously to the NCBI database, new – newly identified SNPs.

	Heteroref	Heteroalt	Ratio heteroref/ heteroalt	50:50	ref	alt
No of variants	1751	535	3.27	2317	4068	2162
rs#	1271	469	2.71	2172	3552	1915
New	480	65	7.38	145	516	247
Ratio rs#/new	2.65	7.22	0.37	14.98	6.88	7.75
Genes	968	382	2.53	1055	1685	1048
Mean no of individuals	1.83	0.72		1.86	4.05	2.34

Allele-specific expression (ASE) has been studied intensively in a wide range of species, including mice (Crowley et al., 2015), humans (Wood et al., 2015; Metsalu et al., 2014), cattle (Chitwood et al., 2013; Chamberlain et al., 2015), and pigs (Wu et al., 2015)) and tissues. Numerous tools have been developed to better evaluate ASE (van de Geijn et al., 2015; Castel et al., 2015). Estimations of ASE frequency vary from 5% to 85%, depending on tissue type, species, developmental stage, genetic diversity and statistical methods used for the evaluation of ASE (Lagarrigue et al., 2013a, 2013b; Wood et al., 2015; Crowley et al., 2015). Interestingly, extensive variation in allele-specific expression has been observed between tissues, even when investigated in one individual (cattle) (Chamberlain et al., 2015). The highest proportion of ASE genes was noted in lung (71–82%), while the lowest was noted in thymus (8–16%). Authors estimated that the proportion of ASE genes in brain is 14–26%, which is much lower than the number predicted by Crowley et al. (2015) in mice (89%). Our estimates of ASE genes in porcine brain were 52% (968 heteroref genes + 382 heteroalt genes − 185 common genes / total number of genes [2510], Table 3). If we consider only validated SNPs (rs), the number of genes would decrease to 916 (~42% of all genes). It is difficult to compare these results with others since the results may differ even within one tissue type. In pigs, ASE was investigated in response to *Streptococcus suis* 2 infection in the spleen transcriptome (Wu et al., 2015). Authors observed significant differences between control and infected animals in 882 and 1096 SNPs in the Landrace and Enshi Black breeds, respectively. Moreover, Yang et al. (2016) observed 11,300 heterozygous ASE SNPs in prenatal pig muscle. Authors evaluated whether the SNPs associated with various economically important traits, previously identified in GWAS, are subjected to ASE and identified four such SNPs. Interestingly, two of the genes in which these SNPs were located (*SCD* and *PGM1*) are also subjected to ASE in our evaluations (Supplementary Table 1). Generally, we found 157 (7%) common ASE variants between our experiment and that of Yang et al. (2016) despite the fact that different tissues and developmental stages were analysed. Other interesting ASE genes identified in our study were *MGST3* and *GSTA4*, which are part of the glutathione S-transferase family. In mice, high variation in the expression of *Gsta4*, *Gstt2*, *Gstz1*, *Gsto1*, and *Mgst3* was observed among parental alleles, and it was shown that this variation is modulated by local expression QTLs (eQTLs) in several tissues (Lu et al., 2016).

Table 4
Number of identified SNPs in each category in each individual (1 – 12).

	1	2	3	4	5	6	7	8	9	10	11	12	Mean
Alt	1583	1592	1577	1224	1557	1531	1338	1478	1305	1560	1097	1171	1417.75
Heteroalt	492	397	495	438	445	414	410	427	376	444	374	560	439.33
50:50	1253	1254	1262	1031	1189	1094	1244	1150	1001	1107	912	1045	1128.5
Heteroref	1171	1170	1256	1035	1088	1031	1089	1122	939	1147	935	1352	1111.25
Ref	2778	2864	2687	2163	2462	2593	2507	2452	2411	2409	1941	2187	2454.5
Σ	7277	7277	7277	5891	6741	6663	6588	6629	6032	6667	5259	6315	6551.333

Our estimations of ASE frequency revealed strong bias towards the reference allele (968/382 genes), which at first glance suggests that the numbers may be overestimated. However, mean reference bias calculated as reference count/total count for all heterozygous SNPs was 0.56 (± 0.0003), which differs only ~10% from the expected 0.5. Reference bias has been observed previously by Degner et al. (2009) and further discussed by Stevenson et al. (2013). It was proposed that this bias is caused by the differentiating sites and that mapping to both parental genomes, considering all possible phasing of variants that may occur in the same read or performing mapping to the obtained haplotypes, may overcome this error (Shen et al., 2013; Satya et al., 2012). Another strategy was recommended by Stevenson et al. (2013), who proved that eliminating differentiating sites with at least as many neighbouring differentiating sites as the number of mismatches allowed provide comparable results to those obtained after mapping to both parental alleles. Our results were obtained after clustering (removing SNPs with > 3 differentiating sites within each 35 bp), a procedure with a similar effect as that suggested by Stevenson et al. (2013). We have performed a preliminary analysis without this procedure and identified 13,361 variants. Nevertheless, we failed to confirm the presence of 3 out of 6 SNPs by Sanger sequencing (data not shown). In contrast, all variants verified by Sanger sequencing and identified with the clustering procedure were present in chromatograms (Figs. 1–4). The clustering procedure also improved the reference count/total count ratio, which was 0.67 before clustering, and removed indels and MNPs. Therefore, we strongly support the need for removing clustered variants during variant identification in RNA-seq data as discussed in Best Practices for Variant Discovery in RNA-seq in GATK.

In our analysis, we have marked the reads that could not be mapped uniquely, but we did not eliminate the reads that overlapped by one or more insertions/deletions between alleles. This last action may be the source of the ~0.1% bias, which seems relatively low and suggests that our estimations are reliable (Stevenson et al., 2013). Other biases that have been postulated to affect RNA-seq variant calling are unidirectional strands and extremity read bias (Danecek et al., 2012; Lagarrigue et al., 2013b). Nevertheless, it should be noted that too much filtering may lead to the elimination of a large number of true variants and underestimation of ASE. It seems that applying strong quality and read depth filtering and having a high number of biological replicates are of great importance in these kinds of studies (Lagarrigue et al., 2013a, 2013b).

On the other hand, when we compare the ratio between the new and known variants (Table 3), the numbers are different among the heteroref, heteroalt and 50:50 groups. The highest proportion of previously confirmed variants was observed in the 50:50 group (14.98), while the lowest was observed in the heteroref group (2.65). If we assume that the ratio between known and new variants is an indicator of accuracy, the heteroref group had the highest error level. In our data,

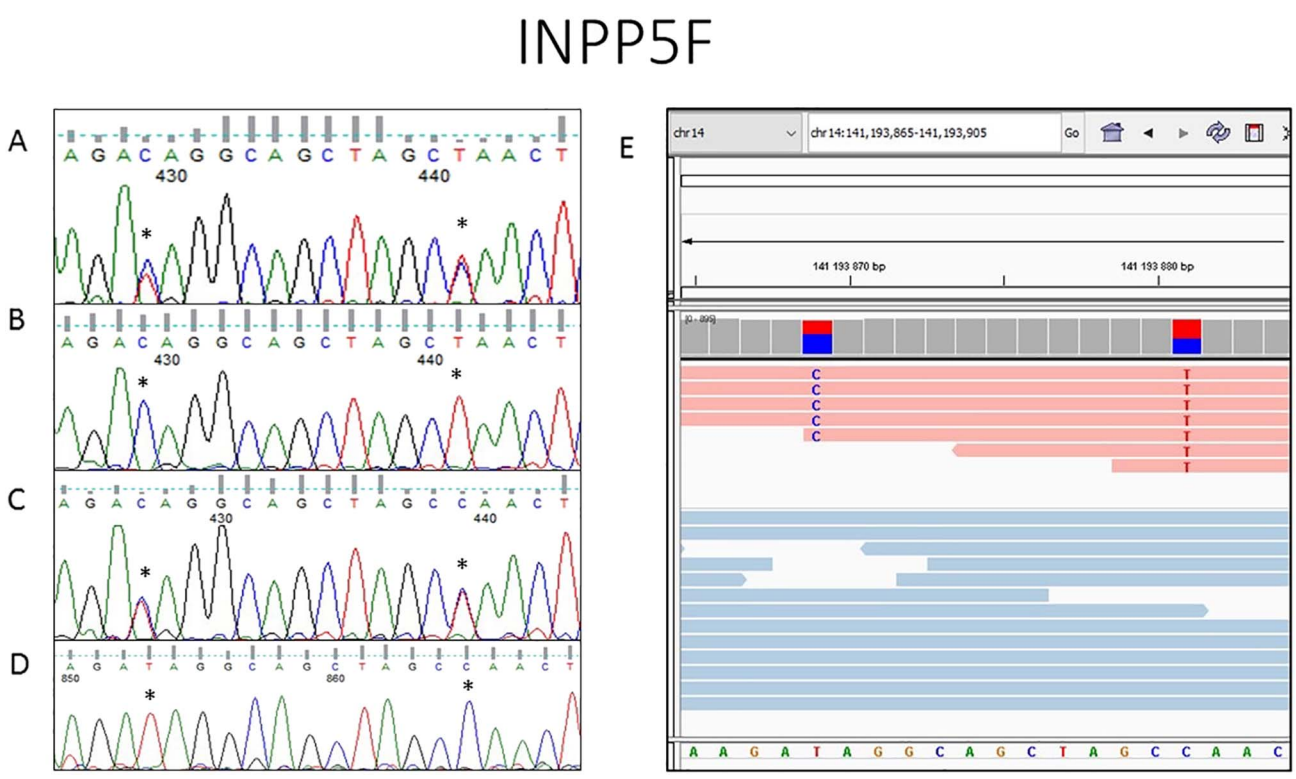

Fig. 1. Chromatograms of fragments of *INPP5F v2*, *MEG3*, *HM13* and *LRRTM1* gene sequence after Sanger sequencing A- offspring's DNA, B- mother's DNA, C- father's DNA, D- offspring's cDNA from the brain Asterisks indicate polymorphic SNPs.

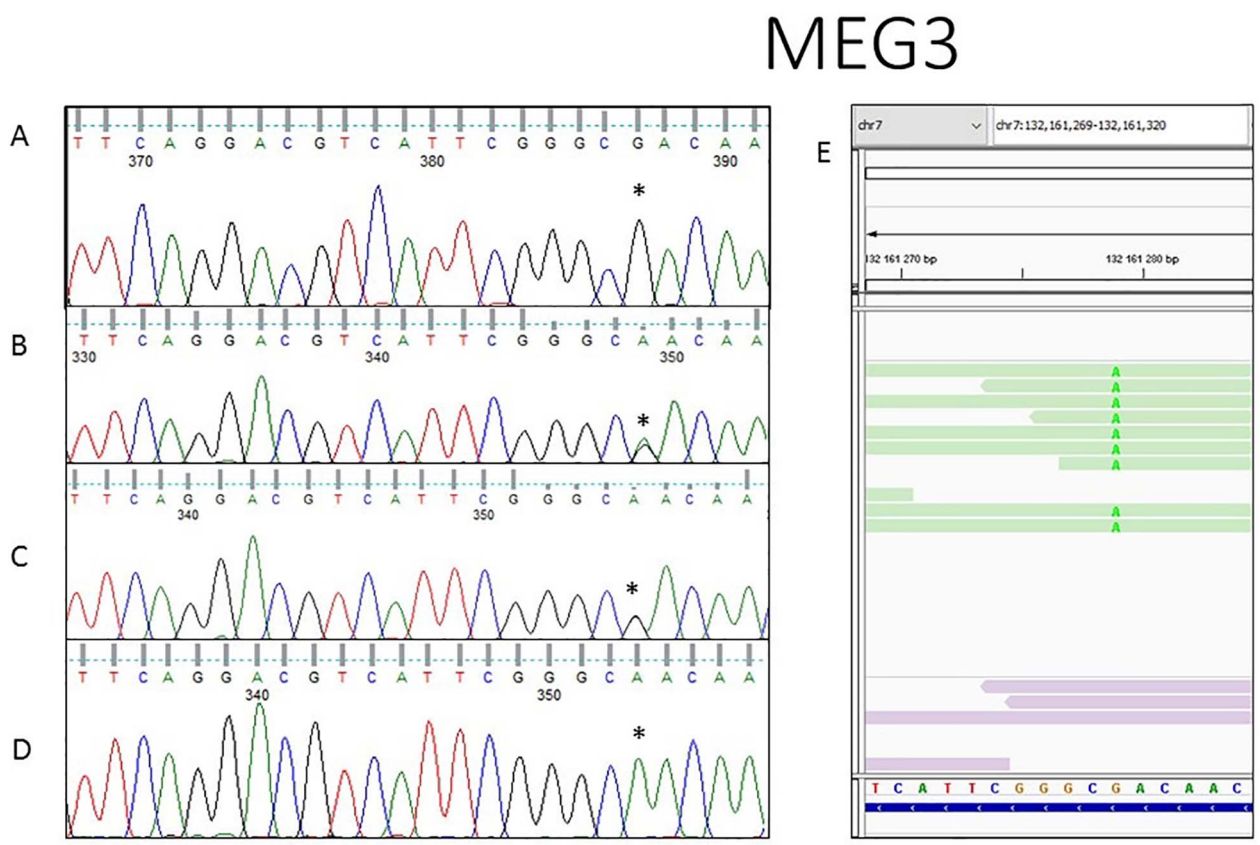

Fig. 2. Chromatograms of fragments of *INPP5F v2*, *MEG3, HM13* and *LRRTM1* gene sequence after Sanger sequencing A- offspring's DNA, B- mother's DNA, C- father's DNA, D- offspring's cDNA from the brain Asterisks indicate polymorphic SNPs.

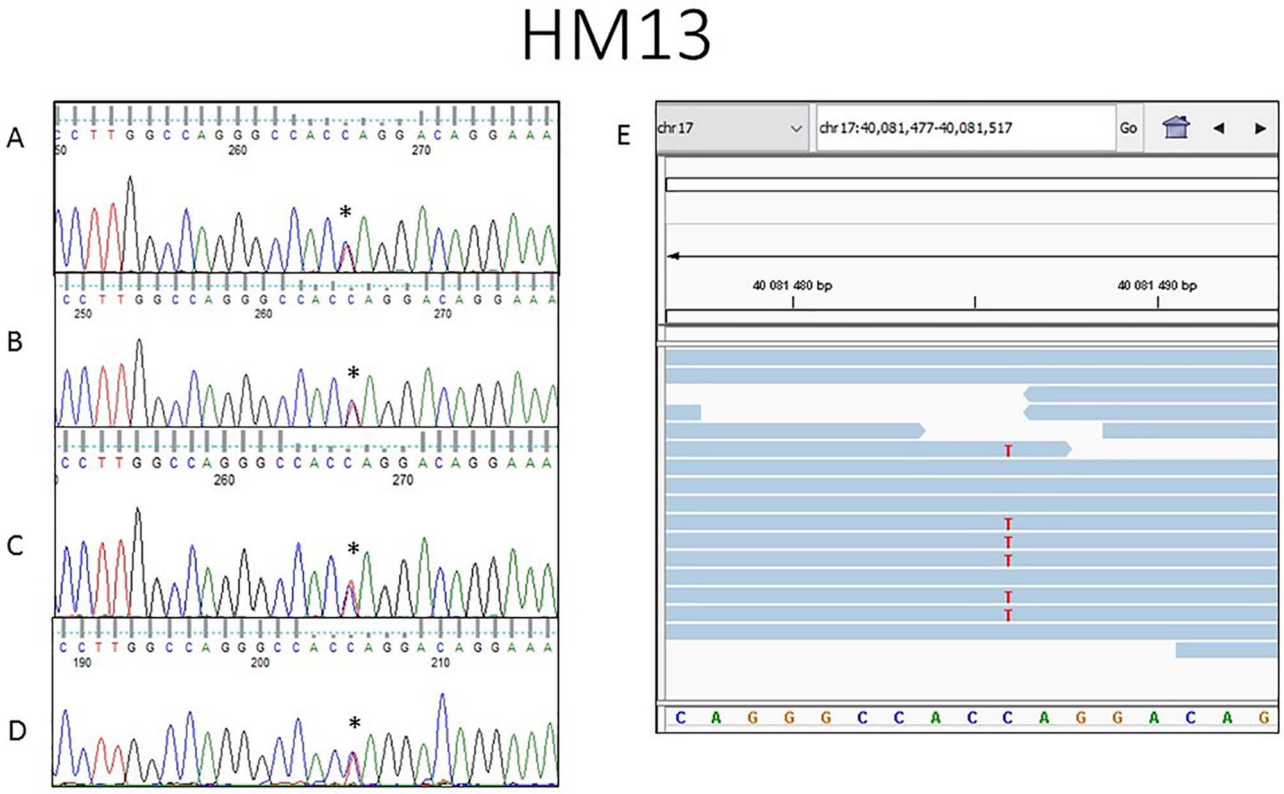

Fig. 3. Chromatograms of fragments of *INPP5F v2*, *MEG3, HM13* and *LRRTM1* gene sequence after Sanger sequencing A- offspring's DNA, B- mother's DNA, C- father's DNA, D- offspring's cDNA from the brain Asterisks indicate polymorphic SNPs.

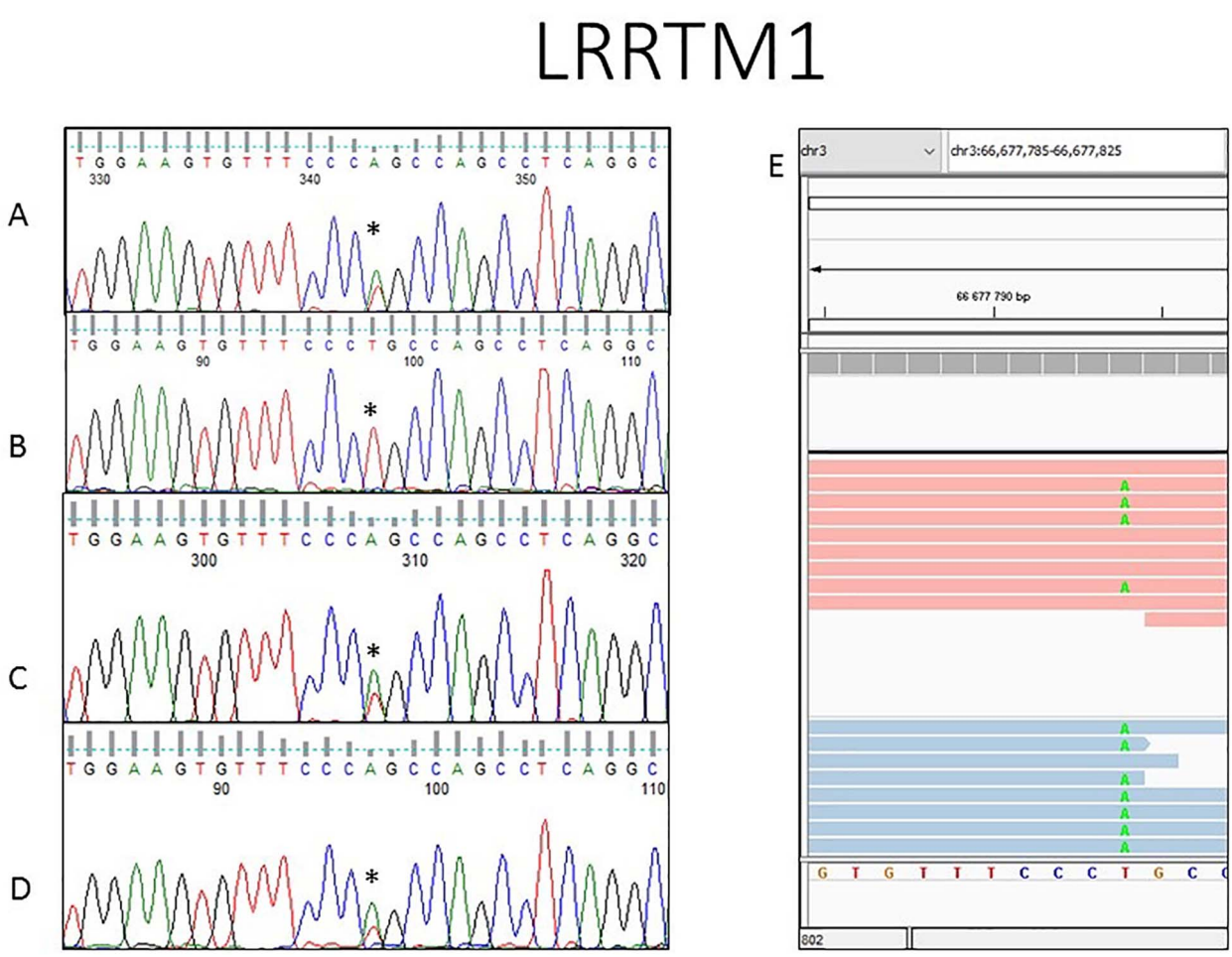

Fig. 4. Chromatograms of fragments of *INPP5F v2*, *MEG3, HM13* and *LRRTM1* gene sequence after Sanger sequencing A- offspring's DNA, B- mother's DNA, C- father's DNA, D- offspring's cDNA from the brain Asterisks indicate polymorphic SNPs.

the reference count/total count ratio was also improved (to 0.52) when only referenced variants were considered. What is more, the heteroref/heteroalt was much higher in new variants than in known variants (Table 3), but when we compare numbers of genes, the ratio was similar to that of the known variants. This result suggests that most of the new variants in heteroref are outside of genes.

The number of ASE variants identified in each gene varied between 1 and 16 (Supplementary Table 1). The highest number of variants was observed for the SLA-5 gene. We may increase the stringency of filtering of our results by eliminating genes with only one ASE variant. After this step, 101 genes with heteroalt variants and 364 genes with heteroref variants would remain, however we believe that this result is underestimated.

3.4. Allelic imbalance at the imprinted loci after RNA-seq

To evaluate allelic expression ratios at the imprinted loci, we have selected all genes available in the gene imprint database (www.geneimprint.com). There are 312 genes that are imprinted or predicted to be imprinted in at least one mammalian species in this database. We used this list to analyse imprinted genes in our dataset and found 49 SNPs located within 16 genes from the list (Table 5). There was no evidence for biallelic expression in 7 genes (*GNAS*, *FUCA1*, *GDAP1L1*, *INPP5F*, *NAP1L5*, *NNAT*, *MEG3*); all SNPs within these genes in all individuals were homozygotic or displayed very high allelic imbalance (Table 5). We have confirmed imprinting for two of these genes (*INPP5F var.2* and *MEG3*) by Sanger sequencing (discussed in the next paragraph). In addition, we have previously described imprinting of the *GNAS complex locus* in pigs (Oczkowicz et al., 2012; Oczkowicz et al., 2015). Moreover, imprinting of *NAP1L5* in pigs has been well documented in a wide range of tissues (Bishoff et al., 2009, Zhang et al., 2011, Jiang et al., 2011, Gu et al., 2011). The *FUCA1* gene was predicted by computational methods to be imprinted in humans (Luedi et al., 2007); however, to date, imprinting in this gene has not been found in any species. The *GDAP1L1* gene is paternally expressed in humans (Aziz et al., 2013) but has not been studied in pigs. Paternal expression of the *NNAT* gene is conserved in mammals and was shown in mouse (Kagitani et al., 1997), humans (Evans et al., 2001), pigs (Cheng et al., 2007), cattle (Zaitoun and Khatib, 2006) and rabbits (Duan et al., 2015).

On the other hand, there was at least one SNP with an allelic ratio between 1.35 and 0.65 in 9 genes *(CALCR*, *EGFL7*, *COMMD1*, *COPG2*, *HM13*, *LRRTM1*, *SNX14*, *NAP1L4*, *NTM*), which suggests biallelic expression. We confirmed lack of imprinting in two of these genes (*HM13 and LRRTM1*) in porcine brain by Sanger sequencing. Similarly, Bishoff et al. (2009) found no evidence for imprinting of *HM13* and *COMMD1*. Moreover, other studies revealed biallelic expression of *NAP1L4* in pigs (Li et al., 2011). On the other hand, it was suggested that *COPG2* was imprinted in porcine placenta but not in brain (Bishoff et al., 2009). In the *COMMD1* and *SNX14* genes, imprinting is restricted to neurons in mice (Huang et al., 2014), but imprinting of *SNX14* has not been studied in pigs to date. Similarly, imprinting of *NTM, CALCR* and *EGFL7* has not been studied in pigs. The *NTM* gene is known to be maternally expressed in human placenta (Barbaux et al., 2012), while *Calcr* is maternally expressed in the brain of mouse (Hoshiya et al., 2003). In addition, imprinting of *EGFL7* was computationally predicted in human

Table 5
Allelic ratio of SNPs localized within imprinted genes after RNA-seq analysis; ref. – reference homozygote, alt – alternate homozygote, shaded – ratio between 0,65 and 1,35, suggesting biallelic expression, bolded – SNPs used for verification of imprinting status by Sanger sequencing.

Gene short name	rs#	1	2	3	4	5	6	7	8	9	10	11	12
CALCR	rs81218963	ref	0.58	0.79	1.75	nd	0.67	ref	ref	ref	0.92	ref	0.33
CALCR	rs81218770	ref	1.82	1.63	2.27	ref	1.56	ref	ref	ref	1.36	ref	1.04
COMMD1	rs710841539	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	0.85
COMMD1	.	5.33	11	12	10	24	12	9	18	13	6.33	9	5.67
COPG2	.	ref	ref	ref	ref	nd	ref	ref	ref	ref	nd	nd	nd
COPG2	rs328324359	1.5	ref	0.28	0.85	0.35	nd	alt	0.43	nd	nd	nd	0.91
EGFL7	rs333089457	ref	ref	ref	ref	ref	ref	ref	0.57	ref	ref	ref	ref
EGFL7	rs343836979	ref	0.58	ref	nd	1.22	0.27	nd	ref	ref	ref	nd	nd
FUCA1	rs334497847	ref	ref	ref	nd	nd	nd	ref	nd	nd	nd	nd	nd
GDAP1L1	rs80874073	alt	alt	alt	alt	alt	alt	alt	alt	nd	alt	nd	alt
GNAS	rs80806327	18	ref	ref	20	ref	ref	6.67	ref	ref	ref	ref	ref
GNAS	rs333005482	alt	alt	0.08	alt	alt	alt	alt	0.08	alt	0.04	alt	0.06
GNAS	.	ref	ref	ref	43	ref	ref	ref	ref	ref	ref	ref	6.33
GNAS	rs321116562	ref	45	ref	ref	ref	ref	ref	ref	ref	ref	ref	3.78
HM13	**rs331705747**	alt	1.64	alt	nd	0.86	nd	0.43	0.5	0.58	0.64	nd	0.38
HM13	**rs346344633**	alt	0.78	alt	2.6	nd	1.18	0.57	1.4	1.33	1.17	nd	1
HM13	rs81210760	alt	0.15	alt	nd	nd	nd	0.25	nd	nd	0.19	nd	nd
INPP5F	rs339030505	ref	ref	ref	ref	ref	ref	ref	alt	10	ref	ref	17
INPP5F	**rs80962734**	alt	22	alt	8.43	19	15	alt	alt	alt	alt	15.67	alt
INPP5F	**rs80860688**	alt	ref	alt	8.13	14.75	16.75	alt	alt	alt	alt	25.5	0.02
INPP5F	rs322081108	alt	25	alt	30	15	74	0.02	alt	0.02	alt	11.2	0.04
INPP5F	rs336180445	alt	14.4	alt	8.25	21.67	12.17	alt	alt	alt	alt	8.67	alt
INPP5F	rs344477287	alt	20.67	alt	17.33	16.67	21	alt	ref	0.05	0.04	23.00	0.06
INPP5F	rs323058715	alt	14.25	alt	18.67	13.25	59	0.042	0.04	alt	alt	18	alt
INPP5F	rs328264947	alt	15	alt	5.88	20.33	ref	alt	alt	alt	alt	12.67	alt
INPP5F	rs343601614	alt	ref	alt	9.67	20.67	9.57	alt	alt	alt	alt	5.43	alt
INPP5F	rs329258420	0.01	35.5	alt	23	7.22	17.5	alt	alt	alt	alt	9.2	0.03
LRRTM1	**rs339128599**	1.31	1.35	ref	nd	ref	ref	nd	ref	nd	ref	nd	ref
LRRTM1	.	6	9	10	nd	7.33	5.2	nd	nd	nd	nd	nd	9
NAP1L4	rs337899675	ref	0.90	ref	ref	ref	ref	ref	37	ref	ref	ref	1.33
NAP1L4	rs320786902	ref	1.57	ref	nd	nd	ref	ref	ref	ref	ref	ref	nd
NAP1L4	rs711312081	1.07	1.14	alt	nd	nd	1.375	0.5	alt	nd	nd	nd	nd
NAP1L4	rs341219696	ref	1.22	ref	ref	ref	ref	ref	ref	27	ref	ref	0.88
NAP1L5	.	6.33	ref	ref	25	9	22	ref	ref	18	ref	nd	nd
NNAT	rs322935215	ref	ref	0.40	ref	ref	ref	ref	ref	ref	ref	ref	ref
NTM	rs341843203	1.13	0.88	0.96	0.04	alt	alt	alt	1.07	alt	alt	0.71	0.76
NTM	rs322733927	ref	ref	ref	ref	ref	ref	ref	2.75	ref	ref	nd	ref
NTM	rs334074130	1.64	0.71	0.74	alt	alt	alt	0.04	2.15	0.05	alt	1.25	34
NTM	rs344901949	1.33	ref	ref	nd	1.43	ref	alt	nd	nd	ref	nd	ref
NTM	rs318807781	ref	ref	ref	nd	ref	ref	ref	ref	19	ref	nd	1.08
NTM	rs326681234	0.78	1.18	0.55	0.04	alt	0.03	alt	0.6	nd	alt	nd	ref
NTM	rs337817366	0.81	0.47	1.43	alt	alt	alt	alt	1.36	0.06	alt	nd	ref
NTM	rs321183927	0.81	1.33	0.8	alt	alt	alt	alt	0.87	nd	alt	0.6	ref
NTM	rs343352318	1.46	1.13	0.85	nd	alt	alt	alt	0.78	alt	alt	1.38	34
SNX14	rs81001473	alt	1.17	alt	alt	alt	alt	alt	nd	nd	alt	nd	nd
SNX14	rs80897405	alt	alt	0.03	alt	alt	alt	alt	0.48	alt	alt	alt	alt
MEG3	**rs325797437**	ref	ref	alt	nd	nd	ref	ref	ref	nd	alt	ref	nd
MEG3	**rs81286030**	ref	ref	alt	nd	ref	nd	ref	ref	nd	alt	nd	nd
MEG3	.	ref	38	ref	nd	3.5	nd	20	33	21	11	ref	nd

but not proven experimentally (Luedi et al., 2007).

Our approach enabled us to exclude imprinting in loci where biallelic expression occurs; however, we were unable to find completely silent alleles in imprinted genes. Moreover, reciprocal crosses are recommended to confidently confirm imprinting patterns.

3.5. Verification of imprinting status of MEG3, INPP5F var.2, LRRTM1 and HM13 in porcine brain by Sanger sequencing

To verify the results obtained by RNA-seq we have used Sanger sequencing to sequence several SNPs located within genes previously

identified as putatively imprinted in pigs or other mammals (www.geneimprint.com). The maternally expressed gene 3 (*MEG3*) is imprinted in pigs and all other investigated mammals (Li et al., 2008, Bischoff et al., 2009). We have investigated two SNPs, *rs325797437* and *rs81286030*, within this gene. Two animals are heterozygous for these SNPs. In the cDNA from the brain of these animals, we observed only the maternal allele from both Sanger sequencing and RNA-seq (Fig. 2).

The *INPP5F* gene codes for inositol phosphatase and produces biallelically expressed and imprinted (*INPP5F variant 2*) transcripts in mice (Choi et al., 2005). In pigs, *INNP5F variant 2* has increased expression in biparental foetuses compared with parthenogenetic samples (Bischoff et al., 2009). We have designed variant-specific primers that amplify only variant 2 of *INPP5F* to investigate if this variant is also imprinted in porcine adult brain. The amplified fragment contained two SNPs, *rs80962734* and *rs80860688*. There were five individuals who were heterozygous (CT) at both loci among our samples. The mother of these individuals was homozygous (CC), while the father was heterozygous (CT). Sequencing of cDNA from the brain of these heterozygous offspring revealed that all investigated samples contained only the T allele, which must come from the father (Fig. 1). Our results show for the first time that *INPP5F variant 2* is paternally expressed in porcine brain.

Imprinting of the *HM13* gene was shown in mice (Wood et al., 2008), but not in pigs (Bischoff et al., 2009). Our results confirm that *HM13* is biallelically expressed in pigs since two SNPs, *rs331705747* and *rs346344633*, in both parental alleles were observed in all heterozygous animals investigated (Fig. 3).

Similarly, the *LRRTM1* gene appeared to be biallelically expressed in pigs (Fig. 4). We found two alleles of the *rs339128599* SNP in all investigated heterozygous samples of porcine cDNA. In contrast, *LRRTM1* is imprinted in humans (Francks et al., 2007).

4. Conclusions

Variant calling from RNA-seq data is still challenging; however, new methodologies have improved proper identification of variants (Shen et al., 2013; Lu et al., 2017). Therefore, all new experiments within this area provide new data, which may lead to improving the development of statistical software. This fact is especially true for the experiments performed on a species whose genome is not well known and for experiments spanning more complicated applications such as ASE analysis. Our study fulfils these two conditions. Although ASE has been known for some time, the results of different studies vary considerably. Our results suggest that more than half of genes are expressed differentially from both alleles of a locus in porcine brain. These results are the first estimates of ASE in this tissue in pigs. In the future we plan to examine other tissues from the same animals.

Moreover, in this paper, we present how RNA-seq data may be utilized for the analysis of imprinted genes without parental genome sequencing. Through the estimation of allelic ratios at SNPs in the genes in the gene imprint database, we selected informative SNPs for imprinting verification by Sanger sequencing. In this way, the imprinted status of genes may be verified in other tissues and at other developmental stages for which RNA-seq data are available.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2017.10.076.

Acknowledgements

This study was supported by the National Research Institute of Animal Production statutory activity, Research Project no. 04.009.1.

References

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aziz, A., Baxter, E.J., Edwards, C., Cheong, C.Y., Ito, M., Bench, A., Kelley, R., Silber, Y., Beer, P.A., Chng, K., Renfree, M.B., McEwen, K., Gray, D., Nangalia, J., Mufti, G.J., Hellstrom-Lindberg, E., Kiladjian, J.J., McMullin, M.F., Campbell, P.J., Ferguson-Smith, A.C., Green, A.R.J., 2013. Cooperativity of imprinted genes inactivated by acquired chromosome 20q deletions. Clin. Invest. 5, 2169–2182. http://dx.doi.org/10.1172/JCI66113.

Barbaux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.B., Pinard, A., Auer, J., Bessières, B., Barlier, A., Jacques, S., Simeoni, U., Dandolo, L., Letourneur, F., Jammes, H., Vaiman, D., 2012. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. Epigenetics 9, 1079–1090. http://dx.doi.org/10.4161/epi.21495.

Bischoff, S.R., Tsai, S., Hardison, N., Motsinger-Reif, A.A., Freking, B.A., Nonneman, D., Rohrer, G., Piedrahita, J.A., 2009. Characterization of conserved and nonconserved imprinted genes in swine. Biol. Reprod. 5, 906–920. http://dx.doi.org/10.1095/biolreprod.109.078139.

Bonthuis, P.J., Huang, W.C., Stacher Hörndli, C.N., Ferris, E., Cheng, T., Gregg, C., 2015. Noncanonical genomic imprinting effects in offspring. Cell Rep. 12, 979–991. http://dx.doi.org/10.1016/j.celrep.2015.07.017.

Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., Lappalainen, T., 2015. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 16, 195. http://dx.doi.org/10.1186/s13059-015-0762-6.

Chamberlain, A.J., Vander Jagt, C.J., Hayes, B.J., Khansefid, M., Marett, L.C., Millen, C.A., Nguyen, T.T., Goddard, M.E., 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics 16, 993. http://dx.doi.org/10.1186/s12864-015-2174-0.

Cheng, H.-C., Zhang, F.-W., Deng, C.-Y., Jiang, C.-D., Xiong, Y.-Z., Li, F.-E., Lei, M.-G., 2007. *NNAT* and *DIRAS3* genes are paternally expressed in pigs. Genet. Sel. Evol. 5, 599–607. http://dx.doi.org/10.1186/1297-9686-39-5-599.

Chitwood, J.L., Rincon, G., Kaiser, G.G., Medrano, J.F., Ross, P.J., 2013. RNA-seq analysis of single bovine blastocysts. BMC Genomics 14, 350. http://dx.doi.org/10.1186/1471-2164-14-350.

Choi, J.D., Underkoffler, L.A., Wood, A.J., Collins, J.N., Williams, P.T., Golden, J.A., Schuster Jr., E.F., Loomes, K.M., Oakey, R.J., 2005. A novel variant of Inpp5f is imprinted in brain and its expression is correlated with differential methylation of an internal CpG island. Mol. Cell. Biol. 25, 5514–5522. http://dx.doi.org/10.1128/MCB.25.13.5514-5522.2005.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012a. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program SnpSift. Front. Genet. 3, 35. http://dx.doi.org/10.3389/fgene.2012.00035.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6, 80–92. http://dx.doi.org/10.4161/fly.19695.

Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., Yun, Z., Bell, T.A., Buus, R.J., Calaway, M.E., Didion, J.P., Gooch, T.J., Hansen, S.D., Robinson, N.N., Shaw, G.D., Spence, J.S., Quackenbush, C.R., Barrick, C.J., Nonneman, R.J., Kim, K., Xenakis, J., Xie, Y., Valdar, W., Lenarcic, A.B., Wang, W., Welsh, C.E., Fu, C.P., Zhang, Z., Holt, J., Guo, Z., Threadgill, D.W., Tarantino, L.M., Miller, D.R., Zou, F., McMillan, L., Sullivan, P.F., Pardo-Manuel de Villena, F., 2015. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat. Genet. 47, 353–360. http://dx.doi.org/10.1038/ng.3222.

Danecek, P., Nellåker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M., Adams, D.J., 2012. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol. 13, 26. http://dx.doi.org/10.1186/gb-2012-13-4-r26.

Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y., Pritchard, J.K., 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics 25, 3207–3212. http://dx.doi.org/10.1093/bioinformatics/btp579.

DeVeale, B., van der Kooy, D., Babak, T., 2012. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet. 8, e1002600. http://dx.doi.org/10.1371/journal.pgen.1002600.

Dodt, M., Roehr, J., Ahmed, R., Dieterich, C., 2012. Flexbar — flexible barcode and adapter processing for next-generation sequencing platforms. Biology 1, 895–905. http://dx.doi.org/10.3390/biology1030895.

Duan, F., Chen, X., Yuan, L., Song, Y., Wang, A., Lv, Q., Li, Z., Lai, L., 2015. Conservation of imprinting of *Neuronatin* (*Nnat*) in rabbits. SpringerPlus 4, 257. http://dx.doi.org/10.1186/s40064-015-1054-z.

Evans, H.K., Wylie, A.A., Murphy, S.K., Jirtle, R.L., 2001. The neuronatin gene resides in a "micro-imprinted" domain on human chromosome 20q11.2. Genomics 77, 99–104. http://dx.doi.org/10.1006/geno.2001.6612.

Francks, C., Maegawa, S., Laurén, J., Abrahams, B.S., Velayos-Baeza, A., Medland, S.E., Colella, S., Groszer, M., McAuley, E.Z., Caffrey, T.M., Timmusk, T., Pruunsild, P., Koppel, I., Lind, P.A., Matsumoto-Itaba, N., Nicod, J., Xiong, L., Joober, R., Enard, W., Krinsky, B., Nanba, E., Richardson, A.J., Riley, B.P., Martin, N.G., Strittmatter, S.M., Möller, H.J., Rujescu, D., St Clair, D., Muglia, P., Roos, J.L., Fisher, S.E., Wade-Martins, R., Rouleau, G.A., Stein, J.F., Karayiorgou, M., Geschwind, D.H., Ragoussis, J., Kendler, K.S., Airaksinen, M.S., Oshimura, M., DeLisi, L.E., Monaco, A.P., 2007. LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. Mol. Psychiatry 12, 1129–1139. http://dx.doi.org/10.1038/sj.mp.4002053.

Frésard, L., Leroux, S., Servin, B., Gourichon, D., Dehais, P., Cristobal, M.S., Marsaud, N., Vignoles, F., Bed'hom, B., Coville, J.L., Hormozdiari, F., Beaumont, C., Zerjal, T., Vignal, A., Morisson, M., Lagarrigue, S., Pitel, F., 2014. Transcriptome-wide

investigation of genomic imprinting in chicken. Nucleic Acids Res. 42, 3768–3782. http://dx.doi.org/10.1093/nar/gkt1390.
Garrison, E., Marth, G., 2012. Haplotype-based Variant Detection From Short-read Sequencing arXiv Preprint arXiv:12073907 [q-bioGN].
Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., Dulac, C., 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science 329, 643–648. http://dx.doi.org/10.1126/science.1190830.
Gu, T., Zhao, S.H., Li, C.C., 2011. NAP1L5 is imprinted in porcine placenta. Anim. Genet. 42, 568–569. http://dx.doi.org/10.1111/j.1365-2052.2011.02214.x.
Gu, T., Buaas, F.W., Simons, A.K., Ackert-Bicknell, C.L., Braun, R.E., Hibbs, M.A., 2012. Canonical A-to-I and C-to-U RNA editing is enriched at 3′UTRs and microRNA target sites in multiple mouse tissues. Papavasiliou N, ed. PLoS One 7, e33720. http://dx.doi.org/10.1371/journal.pone.0033720.
Hoshiya, H., Meguro, M., Kashiwagi, A., Okita, C., Oshimura, M., 2003. Calcr, a brain-specific imprinted mouse calcitonin receptor gene in the imprinted cluster of the proximal region of chromosome 6. J. Hum. Genet. 4, 208–211. http://dx.doi.org/10.1007/s10038-003-0006-6.
Huang, H.S., Yoon, B.J., Brooks, S.M., Bakal, R., Berrios, J., Larsen, R.S., Wallace, M.L., Han, J.E., Chung, E.H., Zylka, M.J., Philpot, B.D., 2014. *Snx14* regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice. PLoS ONE 5, e98383. http://dx.doi.org/10.1371/journal.pone.0098383.
Jiang, C.D., Li, S., Deng, C.Y., 2011. Assessment of genomic imprinting of PPP1R9A NAP1L5 and PEG3 in pigs. Genetika 47, 537–542.
Kagitani, F., Kuroiwa, Y., Wakana, S., Shiroishi, T., Miyoshi, N., Kobayashi, S., Nishida, M., Kohda, T., Kaneko-Ishino, T., Ishino, F., 1997. Peg5/Neuronatin is an imprinted gene located on sub-distal chromosome 2 in the mouse. Nucleic Acids Res. 17, 3428–3432.
Kleinman, C.L., Majewski, J., 2012. Comment on “widespread RNA and DNA sequence differences in the human transcriptome”. Science 335, 1302. author reply 1302. https://doi.org/10.1126/science.1209658.
Lagarrigue, S., Martin, L., Hormozdiari, F., Roux, P.F., Pan, C., van Nas, A., Demeure, O., Cantor, R., Ghazalpour, A., Eskin, E., Lusis, A.J., 2013a. Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. Genetics 195, 1157–1166. http://dx.doi.org/10.1534/genetics.113.153882.
Lagarrigue, S., Hormozdiari, F., Martin, L.J., Lecerf, F., Hasin, Y., Rau, C., Hagopian, R., Xiao, Y., Yan, J., Drake, T.A., Ghazalpour, A., Eskin, E., Lusis, A.J., 2013b. Limited RNA editing in exons of mouse liver and adipose. Genetics 193, 1107–1115. http://dx.doi.org/10.1534/genetics.112.149054.
Lassmann, T., Hayashizaki, Y., Daub, C.O., 2011. SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27 (1), 130. http://dx.doi.org/10.1093/bioinformatics/btq614.
Li, X.P., Do, K.T., Kim, J.J., Huang, J., Zhao, S.H., Lee, Y., Rothschild, M.F., Lee, C.K., Kim, K.S., 2008. Molecular characteristics of the porcine DLK1 and MEG3 genes. Anim. Genet. 39, 189–192. http://dx.doi.org/10.1111/j.1365-2052.2007.01693.x.
Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., et al., 2011. Widespread RNA and DNA sequence differences in the human transcriptome. Science 333, 53–58.
Lu, L., Pandey, A.K., Houseal, M.T., Mulligan, M.K., 2016. The genetic architecture of murine glutathione transferases. PLoS One 11, e0148230. http://dx.doi.org/10.1371/journal.pone.0148230. (eCollection 2016).
Lu, Y., Boswell, M., Boswell, W., Kneitz, S., Hausmann, M., Klotz, B., Regneri, J., Savage, M., Amores, A., Postlethwait, J., Warren, W., Schartl, M., Walter, R., 2017. Molecular genetic analysis of the melanoma regulatory locus in *Xiphophorus* interspecies hybrids. Mol. Carcinog. 56, 1935–1944. http://dx.doi.org/10.1002/mc.22651.
Luedi, P.P., Dietrich, F.S., Weidman, J.R., Bosko, J.M., Jirtle, R.L., Hartemink, A.J., 2007. Computational and experimental identification of novel human imprinted genes. Genome Res. 17, 1723–1730. http://dx.doi.org/10.1101/gr.6584707.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. http://dx.doi.org/10.1101/gr.107524.110.
Metsalu, T., Viltrop, T., Tiirats, A., Rajashekar, B., Reimann, E., Kõks, S., Rull, K., Milani, L., Acharya, G., Basnet, P., Vilo, J., Mägi, R., Metspalu, A., Peters, M., Haller-Kikkatalo, K., Salumets, A., 2014. Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. Epigenetics 9, 1397–1409. http://dx.doi.org/10.4161/15592294.2014.970052.
Oczkowicz, M., Piestrzyńska-Kajtoch, A., Ropka-Molik, K., Rejduch, B., Eckert, R., 2012. Expression and imprinting analysis of the NESP55 gene in pigs. Gene Expr. Patterns 12, 18–23. http://dx.doi.org/10.1016/j.gep.2011.10.002.
Oczkowicz, M., Piestrzyńska-Kajtoch, A., Ropka-Molik, K., 2015. Guanine nucleotide binding protein (gnas complex locus) gene produces biallelically expressed and paternally expressed transcripts in pigs. Ann. Anim. Sci. 15, 867–877. http://dx.doi.org/10.1515/aoas-2015-0030.
Satya, R.V., Zavaljevski, N., Reifman, J., 2012. A new strategy to reduce allelic bias in RNA-Seq readmapping. Nucleic Acids Res. 40, e127. http://dx.doi.org/10.1093/nar/gks425.
Shen, Y., Garcia, T., Pabuwal, V., Boswell, M., Pasquali, A., Beldorth, I., Warren, W., Schartl, M., Cresko, W.A., Walter, R.B., 2013. Alternative strategies for development of a reference transcriptome for quantification of allele specific expression in organisms having sparse genomic resources. Comp. Biochem. Physiol. Part D Genomics Proteomics 8, 11–16. http://dx.doi.org/10.1016/j.cbd.2012.10.006.
Stevenson, K.R., Coolon, J.D., Wittkopp, P.J., 2013. Sources of bias in measures of allele-specific expression derived from RNA-sequence data aligned to a single reference genome. BMC Genomics 14, 536. http://dx.doi.org/10.1186/1471-2164-14-536.
Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25, 1105–1111. http://dx.doi.org/10.1093/bioinformatics/btp120.
Van de Geijn, B., McVicker, G., Gilad, Y., Pritchard, J.K., 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods 12, 1061–1063. http://dx.doi.org/10.1038/nmeth.3582.
Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., Clark, A.G., 2008. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS One 3, e3839. http://dx.doi.org/10.1371/journal.pone.0003839.
Wang, Q., Li, K., Zhang, D., Li, J., Xu, G., Zheng, J., Yang, N., Qu, L., 2015. Next-generation sequencing techniques reveal that genomic imprinting is absent in day-old *Gallus gallus domesticus* brains. PLoS One 10, e0132345. http://dx.doi.org/10.1371/journal.pone.0132345. (eCollection 2015).
Wood, A.J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C.V., Peters, J., Bourc'his, D., Oakey, R.J., 2008. Regulation of alternative polyadenylation by genomic imprinting. Genes Dev. 22, 1141–1146. http://dx.doi.org/10.1101/gad.473408.
Wood, D.L., Nones, K., Steptoe, A., Christ, A., Harliwong, I., Newell, F., Bruxner, T.J., Miller, D., Cloonan, N., Grimmond, S.M., 2015. Recommendations for accurate resolution of gene and Isoform allele-specific expression in RNA-Seq data. PLoS One 10, e0126911. http://dx.doi.org/10.1371/journal.pone.0126911.
Wu, H., Gaur, U., Mekchay, S., Peng, X., Li, L., Sun, H., Song, Z., Dong, B., Li, M., Wimmers, K., Ponsuksili, S., Li, K., Mei, S., Liu, G., 2015. Genome-wide identification of allele-specific expression in response to *Streptococcus suis* 2 infection in two differentially susceptible pig breeds. J. Appl. Genet. 56, 481–491. http://dx.doi.org/10.1007/s13353-015-0275-8.
Yang, Y., Tang, Z., Fan, X., Xu, K., Mu, Y., Zhou, R., Li, K., 2016. Transcriptome analysis revealed chimeric RNAs, single nucleotide polymorphisms and allele-specific expression in porcine prenatal skeletal muscle. Sci. Rep. 6, 29039. http://dx.doi.org/10.1038/srep29039.
Zaitoun, I., Khatib, H., 2006. Assessment of genomic imprinting of *SLC38A4*, *NNAT*, *NAP1L5*, and *H19* in cattle. BMC Genet. 7, 49. http://dx.doi.org/10.1186/1471-2156-7-49.
Zhang, F.W., Deng, C.Y., He, H.J., Gu, N., Han, Z.B., Chen, Y., Wu, Q., 2011. Molecular cloning mRNA expression and imprinting status of PEG3 NAP1L5 and PPP1R9A genes in pig. Genes Genet. Syst. 86, 47–52. http://dx.doi.org/10.1266/ggs.86.47.