

Profiling novel alternative splicing within multiple tissues puts insight into the porcine genome annotation

Jian-Feng Liu (✉ liujf@cau.edu.cn)

China Agricultural University <https://orcid.org/0000-0002-5766-7864>

Wen Feng

College of Animal Science and Technology, China Agricultural University

Pengju Zhao

China Agricultural University

Xianrui Zheng

China Agricultural University

Research article

Keywords: Alternative splicing, transcript, protein, SNP

Posted Date: October 12th, 2019

DOI: <https://doi.org/10.21203/rs.2.15981/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background Alternative splicing (AS) is a process that mRNA precursor splices intron to form the mature mRNA. AS plays important roles in contributing to transcriptome and proteome diversify. However, to date there is no research about pig AS in genome-wide level by RNA sequencing.

Results To characterize the AS in pigs, herein we detected genome-wide transcripts and events by RNA sequencing technology (RNA-seq) in 34 different tissues in Duroc pigs. In total, we identified 138,403 AS events and 29,270 expressed genes. We found alternative donor site was the most common AS form, which is accounted for 44% of the total AS events. The percentage of the other 3 AS forms are all around 19%. The results showed that the most common AS events (alternative donor site) can produce different transcripts or different proteins which affect the biological process. Among these AS events, 109,483 were novel AS events, and the number of alternative donor splice site has increased the most (Accounting for 44% of the novel AS events).

Conclusions The expression of gene with tissue specific AS events showed that the functions of these genes were consistent with the tissue function. AS increased proteome diversity and resulted in novel proteins that gained and lost important functional domains. In summary, these findings extend genome annotation and highlight roles that AS acts in tissue identity in pig. **Key words:** Alternative splicing; transcript; protein; SNP

Background

Alternative splicing (AS) is a regulated process, which can generate multiple transcripts from a single gene and therefore plays an extremely important role in expanding protein diversity. AS has effects on 95% of multi-exonic genes in human and it also presented a high AS proportion in other animals [1–3]. AS have four basic modes, which are exon skipping, alternative donor site, alternative acceptor site and intron retention (IR), of which the most common is exon skipping in animal [4]. In addition to these four forms, there are two other main mechanisms by which different mRNAs may be generated from the same gene, multiple promoters [5] or multiple polyadenylation sites [6].

Many alternatively spliced isoforms play important roles in the biological timing and development of tissue [7–10]. It is becoming clear that a large amount of AS contribute to the acquisition of adult tissue functions and identity in human during tissue development [11]. AS can also reveal new regulatory mechanisms when genes simultaneously express multiple isoforms, like causing human diseases [12]. Protein coding genes always have several alternatively spliced isoforms, which reveals the importance of AS's function on gene expression [13]. Protein generated by different AS also have the potential to gain or lose domain, which may change significantly. Those protein products might have the same, similar or even opposite biological functions, then effect the performance of phenotypes [14]. Many point mutations like single nucleotide polymorphism (SNP), which may be the result of changes in alternatively splicing [15–17], have been shown to have substantial phenotypic variation affect pre-mRNA splicing [18]. AS could lead to early termination of transcription by premature stop codons [19]. AS regulation plays roles in multiple eukaryotic biological processes, including cell growth [20], lung cancer [21], chromatin modification [22] and tissue development [23].

As the best medical model for human disease, pig has similar anatomical and physiological structure with human [24]. Various research groups have attempted to understand the role of AS in the pig by using expressed sequence tags (EST). A total of 1,223 genes with average 2.8 splicing variants were detected among 16,540 unique genes via EST data [25]. It was reported that more AS could be produced by RNA-seq than EST technology in human and plants [1, 26]. AS in pigs found to play an essential role in regulation of gene expression in genital tissues [27, 28]. However, to our best knowledge, there is no research about pig AS in genome-wide level by RNA sequencing.

In this study, we constructed a profiling AS within multiple tissues to extend the porcine AS genome annotation. To achieve the aim, we performed genome-wide transcripts identification in 34 normal tissues of the pig using over 340 Gb of sequence data generated from 116 RNA sequencing (RNA-seq) libraries. We identified 2,486,239 transcripts and 2,430,911 novel transcripts. Novel and known transcripts were examined for tissue-specific expression patterns and new isoforms were investigated to determine how their divergence from annotated transcripts affects their protein coding and regulation by SNP. The data reported here provide a valuable resource for enhancing the understanding and utilization of the pig AS.

Results

Identification of Novel Transcripts in 34 different pig tissues

To discover and map novel transcripts, we performed RNA-sequencing in 34 different pig tissues. An average of 48.48 million reads per tissue were sequenced from 116 strand-specific and paired-end cDNA libraries (*Table S1*); of these sequences, an average of 43.97 million reads (90.7%) per sample passed the strict quality control (QC). The 1,495 million high-quality reads (376.7 Gb, 135-fold genome coverage) were aligned to the pig genome (*Sus scrofa* 11.1) and 1,223 million mapped fragments (310.1 Gb, 110.8-fold genome coverage) with an average alignment rate of 88.29% were recovered (*Table S2*).

A total of 2,486,239 transcripts with 29,270 genes were produced across all tissues. In these transcripts, 60,578 transcripts with 20,401 coding genes and 3,486 non-coding genes were annotated in the pig Ensembl database. The remaining 2,281,529 novel transcripts were from 26,493 loci without any annotated information. Full details for each tissue are available in *Table S3*. We determined known protein coding genes with FPKM (fragments per kilobase of exon per million fragments mapped) >0.1 among these non-redundant transcript isoforms at least one tissue. These novel transcripts enhance pig non-coding gene annotation and increase the number of average transcripts of each gene in pig.

Classification of Alternative Splicing Types

AS events are classified into twelve basic types, among them, we counted 4 basic types of AS events including exon skipping, alternative acceptor 5' site, intron retention and alternative acceptor 3' site (Fig. 1A). We detected 138,403 AS events and 29,270 genes across all 34 tissues. Moreover, a gene has 4.73 AS events on average. Alternative donor sites was the most common AS event type (60,851, 44%), the other 3 types share the similar percentages (Exon skipping, 19%; Alternative acceptor site, 19%; Intron retention, 18%) (Fig. 1A). The number of novel AS events increased 1.5 to 9.4 times compared to original ones. The most difference was showed in novel alternative donor sites in terms of quantity, which is accounting for 44% in all novel AS events (Fig. 1B).

The novel AS divides into two types, one is the novel AS of known genes and the other one is the novel transcript of unknown genes. The number of novel transcripts in different tissues was showed in Table S4. We found that much more novel transcripts of unknown genes were detected than those of known genes. We further compared the number of original and novel transcripts of per annotated genes. Fig. 2A showed that the more annotated isoforms per gene, the more novel isoforms. Compare to the original isoforms, novel isoforms did not increase too much, may be because of the limited isoforms of per gene. The number would not change much even with improved methods.

The number of AS events among different genes could be various. In our analysis, the AS events number of *AHNAK* gene is up to 391. However, thousands genes only have one AS event. In order to explain the difference existing in these genes, we performed cluster analysis on them. We considered the genes whose number of AS events is greater than 20 belong to a group, and those equals 1 belong to another group. Then we perform KEGG analysis on these two group (Fig. 2B, Table S5). For genes with AS events >20, altogether 24 pathways were significantly enriched, including phosphatidylinositol signaling system (P value is $6.30E-09$, the number of genes is 59), inositol phosphate metabolism (P value is $2.80E-06$, the number of genes is 43), glycerophospholipid metabolism (P value is $1.30E-04$, the number of genes is 49) and so on. These genes were mostly involved in lipid metabolism. Other significant pathway mainly related with the cancer and reproduction. However, for genes with ASE = 1, only 7 pathways were significantly enriched. These pathways mainly involve in neuroactive and some immune-related diseases. The second significant pathway (olfactory transduction, P value is $4.9E-145$) is enriched with 639 genes.

Then we further compared the genes distribution with one AS events and more than 20 AS events in the different tissues (Fig. 2C). Average 1.33 genes with one AS events were expressed in one tissue. However, there were altogether 5,857 genes in 34 tissues, which means many genes were expressed in multiples tissues simultaneously. For the genes with more than 20 AS events, average 27 were expressed in one tissue with 2,131 genes altogether.

Tissue-specificity of Alternative splicing in different tissues

The tissue specific AS events in pig transcriptome were detected by Jekroll-splicingexpress [29]. We defined those were found to be expressed above 0.1 FPKM in only one tissue type, with expression of <0.1 FPKM in all others to be tissue specific transcripts. The number of specific transcripts found varied significantly in different tissues, with cell possessing the most unique isoforms (1, 633) and pancreas possessing the least (133; Table 1). Further examination of these tissue-specific transcripts showed that 86% of the genes that encode them were also only expressed in a single tissue with above 0.1 FPKM, revealing that these transcripts' tissue specificity occurs at the transcriptional level. The remaining 14% of tissue-specific transcripts had other isoforms expressed in other different tissues, indicating that their tissue specificity likely results from AS (Fig. 3A).

In order to identify developmentally regulated changes in AS, Stringtie was used to quantify transcript expression across all tissues. To assess the relationship between the different tissues and their splicing profiles, we have computed the number of isoforms and related genes in different tissues (Table S3). Totally, 7,595 novel isoforms from 2,421 known non-coding genes and 136,537 novel isoforms from 15,385 coding genes were identified. Herein we considerate that the annotated transcripts are those with Ensemble IDs. Adult pig testes had the highest number isoforms when compared with other tissues, followed by thymus, which is similar to humans [30]. The number of AS in adult pig testis is 2,459 more than that in infant testis. This difference may arise from sexual maturity of pig since testis is the genital organ and generates sperm.

We then analyzed the expression of novel and known transcripts in different tissues (Table S6). The average level of expression for novel transcripts in pancreas, liver, longissimus, dorsi, spleen, prostate, adrenal gland and breast these tissues have a higher expression (FPKM>10) than known transcripts. By comparing the difference between the highest expression and the lowest expression in the genes with different AS numbers, it can be found that the more isoforms of a gene, the greater difference in expression (Fig. 3B). Our results found that 116,439 genes were tissue specific, of which 109,773 genes were new detected. In the meanwhile, 5,683 genes were found to be expressed in all tissues, of which 110 are first detected (Table S7). Our results will provide useful resource to improve transcript abundance.

Potential Effect of Novel Alternative Splicing on the Pig Proteome

Translated DNA sequences of novel transcripts were aligned to UniProt database by DIAMOND software [31]. In order to determine their possible effect on the pig proteome, computationally predicted proteins encoded by novel isoforms of known genes were compared to the annotated proteins generated by their similar known transcript. Total of 1,184 AS were mapped to 361 new proteins (Table S8). Our results have improved the annotation of unknown proteins.

HMMER3 was used to identify conserved domains within novel and known proteins, which were then compared in a pairwise manner. One novel transcript of APOBEC3B in tissue of uterus was mapped to a protein (UniProt identifier: D3U1S2_PIG). By comparing their most similar known (UniProt identifier: F1SNY0_PIG) and novel transcript, transcript length changed from 3083 bp to 3072 bp. The annotated protein contains two APOBEC-like N-terminal domains, while a new isoform lacks one APOBEC-like N-terminal domain potentially altering its function significantly (Fig. 4A). For coding exons, the number of novel transcript was less 3 than known one (Fig. 4C). It has been showed that DNA ligase 3 (LIG3) proteins have been proposed to active leukemia via transcription error [32]. Our results found that one novel transcript (UniProt identifier: I3L9T2_PIG) of LGI3 have gained a domain (DNA ligase 3 BRCT domain) than the most similar known transcript (UniProt identifier: I3LN18_PIG; Fig. 4B). On the other hand, the novel transcript has one less the coding exons than known one

(Fig. 4D).. In the meanwhile, we have predicted the protein conformations of these four transcripts (Fig. 4E&4F).. From the protein conformations, we can also see the domains' loss or gain.

Conserved AS between pig and human

The pig is generally considered a promising medical model for human disease and pig orthologs of many human disease-associated genes have been identified. In order to investigate the orthologous relationship between humans and pigs at the protein level, we aligned the detected pig proteins with the human proteins using blast software and the criteria of a minimum length ≥ 50 bp and identity $\geq 80\%$. We identified 4,168 (56.97%) pig protein with orthology to human proteins. We used protein MX2 as an example (Fig. 4H), whose homology is up to 70.01%. The protein sequence was aligned to the human proteins. Six conserved domains were found, including DLP_1, DYNc, Dynamin_M, Dynamin_N, GED and CrfC (Fig. 4G; Table S9).. These conserved proteins provide an indication of genome evolution and allow us to further explore the potential functional similarity between human and pig genomes.

Effect of SNP on Pig Alternative Splicing

Deep RNA-seq sequencing can help us detect sequence variations associated with AS. Herein we grouped these AS into 3 types: canonical splicing with a GT-AG intron (i.e., GT and AG splicing signals at donor and acceptor sites, respectively); Semi-canonical splicing with GC-AG or AT-AC introns; Novel splicing without GT-AG introns. The semi-canonical and novel splicing are also called aberrant splicing together. SNP occurs in different regions of a gene would change the expression of transcript, then change protein abundance, especially when SNP occurs in the exon region (Table S10).. The change of exon would cause synonymous or nonsynonymous mutation in proteins. Our results showed some aberrant splicing in the exon lead to non-synonymous mutations in codons, more seriously, some translations are terminated in advanced in some tissues (Table S11)..

Related SNP, genes and codons were described in the Table 2 when an aberrant splicing occurs more than 20 tissues. Our result showed that 6 novel splicing cause non-synonymous mutation (Table 2).. Particularly, we found a missense SNV (exon3:c.T376A:p.S126T) within Synapse associated protein 1 (SYAP1) gene that generated a Ser -to- Thr substitution, leading to a change in the SYAP1 protein conformation space.

Validation of known and novel transcripts by RT-PCR

The reliability of transcripts identification in this study was verified by RT-PCR of 8 randomly selected novel and 2 known transcripts in 6 tissues (Fig. 5A).. For transcripts identified in all tissues, the RT-PCR analysis showed 5 (TCONS_00011100, TCONS_01815428, TCONS_01410294, TCONS_00028775 and TCONS_02428049) were expressed in all tissues, and 1 other (TCONS_01245051) were detected in 5 tissues. TCONS_01413087 and TCONS_01387293 were successfully validated as tissue-specific transcripts in kidney and thyroid respectively. These results confirm the accuracy of our identification and characterization of the pig transcripts.

Discussion

In the present study, we profiled novel AS from 34 different tissues in Duroc pigs via deep RNA sequencing, which provided insight into the porcine genome annotation. A total of 138,403 AS events and 29,270 expressed genes were detected, 4.73 events in each gene on average. The average number of AS events in each gene has improved 0.69 times compared with 2.8 splicing variants by EST data [11]. Previous studies have revealed that exon skipping, alternative donor site, alternative acceptor splice site, and intron retention are the four basic AS types. Exon skipping is thought to be the most prevalent AS type in animals [4]. In our study, alternative donor site is the most common AS form, which is accounted for 44% of the total AS events. The percentage of the other 3 AS forms are all around 19%. The results showed that the most common AS events (alternative donor site) can produce different transcripts or different proteins which affect the biological process. Among these AS events, 109, 483 were novel AS events, and the number of alternative donor splice site has increased the most (Accounting for 44% of the novel AS events). Regardless of the splicing mechanism, this large increase in new alternatively spliced transcripts will improve the diversity of pig proteome dramatically.

We found genes with less AS events have more chance to express in the tissue-specific manner. Genes with only one AS event were majoring located on the pathway of lipid metabolism. For those with more than 20 AS events were belonged to pregnancy or disease related signal pathways. Whether these genes in different kinds of pathways is related to the number of their AS events need further research.

Based on our AS profiling analysis, we detected a total of tissue specific 15,606 AS events were in 3,743 genes. We found the function of genes with tissue specific AS is nearly consistent with their related tissues. In addition, we observed the expression of the transcript in different tissues could be various. For example, the expression of the Gutathione peroxidase 4 (GPX4) gene is ranging from 20 to over 400 FPKM. And its transcript expression in the adult testes are much higher than in other tissues (Fig. 5B). GPX4 gene has been translated into specific enzymes in the testes of adult rats and the enzymes was reported to be important structural proteins in mature sperm [33, 34].

It has been reported that genes expressed in some tissues are considered to be tissue-enriched genes with tissue-specific related functions in multicellular organisms. In the present study, we found tissues with complex biological processes usually have more tissue-enriched genes, and these genes closely related to the function of the corresponding tissues. For example, the RHO gene enriched in the retina plays an important role in the deposition of retinal pigments [35]. These specific tissue-enriched genes not only confirms the biological properties of known genes, but can also be used to predict the potential function of unknown genes in the pig genome.

In general, majority of the novel splicing events lead to new proteins. We revealed that novel AS often contained domain losses or gains relative to their most similar known isoform, then even lead to opposite function [14]. In our analysis, we observed unknown proteins database was computationally predicted for all novel transcripts. 1184 AS were mapped to 361 new proteins. And we took gene *APOBEC3B* and *LIG3* as examples. By comparing the known and novel transcript, transcript length had changed. The novel transcript of *APOBEC3B* and *LIG3* loss and gain a domain predicted by HMMER3, respectively (Fig. 5A&D).

SNPs affect protein-binding sites (or mutations in the binding proteins themselves) and contribute to aberrant splicing. Some nonsynonymous mutations even cause transcription to stop in advance. We found a missense SNV (exon3:c.T376A:p.S126T) within Synapse associated protein 1 (*SYAP1*) gene that generated a Ser -to-Thr substitution. We can see the protein conformation of these two AS predicted by Swiss-model have changed from the Serine to Threonine.

Conclusion

In summary, our analysis greatly expanded the pig transcriptome to include more than 2,281,529 newly annotated transcripts. Many genes were also found to have more AS events, the greater difference in highest and lowest expression. The expression of gene with tissue specific AS events show that, the function of these genes are consistent with the tissue. This expansion increased proteome diversity and resulted in novel proteins that gained and lost important functional domains. Point mutation like SNP in the AS could lead to new AS. Taken together, these findings extend genome annotation and highlight role that AS plays in tissue identity in pig.

Methods

Samples collection and sequencing

PBMC and 33 different tissues were removed from 9 unrelated healthy Duroc pigs of Shenzhen Jinxinnong Technology Co., Ltd. in this study (Table S2). The whole sample collection and treatment were conducted in strict accordance with the protocol approved by the Institutional Animal Care and Use Committee (IACUC) of China Agricultural University and Shenzhen Jinxinnong Technology Co., Ltd. Tissues and cell then were used for RNA extraction, mRNA-seq library preparation and RNA sequencing.

All qualified genomic RNA were sequenced using Illumina HiSeq 2500 sequencing system at Novogene (Beijing, China). A minimum of 10G clean reads were generated for each sample (corresponding to 126 bp paired-end reads). The sequenced RNA-Seq raw data of these 34 samples is available from NCBI Sequences Read Archive (BioProject number: PRJNA392949).

Read alignment to the reference *Sus scrofa* 11.1 genome

The RNA-Seq raw data were trimmed based on the quality control for downstream analyses by following steps: BBmap automatically detected the adapter sequence of reads and removed reads containing illumina adapters; the Q20, Q30 and GC content of the clean data were calculated by FASTQC for quality control and filtering; homopolymer trimming to 3' end of fragments and removed the N bases of 3' end were carried out by Fastx toolkit (version 0.0.14). The resulting sequences then were mapped to reference genome sequence *Sus scrofa* 11.1 by Hisat 0.1.6-beta 64-bit. Ensemble *Sus scrofa*11.1 version 91 (http://ftp.ensembl.org/pub/release-91/gff3/sus_scrofa/) annotation was used as the transcript model reference for the alignment, splice junctions finding as well as for all protein-coding gene and isoform expression-level quantifications. Beside, Stringtie 1.0.4 (Linux_x86_64) calculated the

FPKM (fragments per kilobase of exon model per million mapped reads [controlling for fragment length and sequencing depth]) values. A gene or transcript was defined as expressed if it was measured above 0.1 FPKM across at least one tissue.

Alternative splicing events in pig transcriptome

Asprofile (v1.0.4) software was used to classify and count the AS events in each sample [36], whose results consists of twelve AS events types. In our analysis, only exon skipping, alternative donors site, alternative acceptor site and intron retention were included. Tissue specific AS events in pig transcriptome were detected by jekroll-splicingexpress software (<https://bitbucket.org/jekroll/splicingexpress/>).

Comparison of Novel and Known Protein Domains

Translated DNA sequences of novel transcripts were aligned to UniProt database by DIAMOND software. HMMER3 (Eddy, 2011; <https://www.ebi.ac.uk/Tools/hmmer/>) was used to determine conserved protein domains for novel isoforms and their most similar known transcripts. The domain changes in novel and known proteins were carried out in a pairwise manner. Protein conformation were predicted by Swiss-model (<https://swissmodel.expasy.org/>).

SNP compared with Alternative Splicing Variations

Here we mainly use the pvaas software to detect the SNV mutation associated with the novel AS, which is mainly to determine the correlation between the SNV and AS by Fisher exact test. Its reliability were assessed by P value after correcting by the Benjamini-Hochberg procedure. Further filtering were performed

in order to ensure the accuracy of SNV: The minimum mutation frequency should be greater than 5% (the minimum mutation reads number accounts for more than 5% of AS); the mutant reads number ≥ 5 ; the corrected P value <0.001 ; the reads number of new AS ≥ 10 .

Known and Novel AS Validation by PCR

RNAs from 6 pig tissues including lung, kidney, brain, gut, thyroid and lymph were transcribed into cDNA using PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa Bio) for PCR reaction. Two reverse transcribed reaction system were conducted. DNA-free contained 1.0 µg RNA, 2.0 µL 5×gDNAEraser Buffer, 1.0 µL gDNA Eraser 7 µL RNase-free dH₂O. After 5 min standing, Reverse Transcribed Reaction System contained 4.0 µL 5×PrimeScript® Buffer 2, 1.0 µL PrimeScript® RT Enzyme Mix 1.0 µL RT Primer Mix, 4.0 µL RNase-free dH₂O with 10.0 µL the liquid of Reverse transcribed reaction system, whose reacting temperature is 37°C for 15 min, followed by 85°C for 5s. The primers for PCR amplification were designed by Primer-blast and confirmed by Oligo 7.0. The details about the primers are listed in *Table S12*. PCR was performed on 25 µL volumes containing 22 mL premix (Golden Star T6 Super PCR Mix, Beijing TsingKe Biotech Co., Ltd), 1 µL forward and 1 µL reverse primers (10 µmol/L) 0.5 µL, and 1 µL cDNA. PCR conditions were 95°C for 8 min, followed by 35 cycles of 95°C for 30 s, 55~60°C for 30 s, 72°C for 30s and 8min at 72°C.

Availability of Data and Materials

The sequenced RNA-Seq raw data for 34 pig tissues is available from NCBI Sequences Read Archive with the BioProject number: PRJNA392949.

Abbreviations

AS: Alternative splicing

EST: Expressed sequence tags

FPKM: Fragments per kilobase of exon model per million mapped reads APOBEC3B: Apolipoprotein B mRNA editing enzyme catalytic subunit 3B

LIG3: DNA ligase 3

MX2: MX dynamin like GTPase 2

SYAP1: Synapse associated protein 1

AHNAK: AHNAK nucleoprotein

GPX4: Gutathione peroxidase 4

References

1. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet* 2008, *40*(12):1413–1415.
2. Chacko E, Ranganathan S: *Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. BMC Genomics* 2009, *10 Suppl 3*:S11.
3. Liu S, Zhou X, Hao L, Piao X, Hou N, Chen Q: *Genome-Wide Transcriptome Analysis Reveals Extensive Alternative Splicing Events in the Protoscoleces of Echinococcus granulosus and Echinococcus multilocularis. Front Microbiol* 2017, *8*:929.
4. Sammeth M, Foissac S, Guigo R: *A general definition and nomenclature for alternative splicing events. PLoS Comput Biol* 2008, *4*(8):e1000147.
5. Koenigsberger C, Chicca JJ, 2nd, Amoureux MC, Edelman GM, Jones FS: *Differential regulation by multiple promoters of the gene encoding the neuron-restrictive silencer factor. Proc Natl Acad Sci U S A* 2000, *97*(5):2291–2296.
6. Di Giarmartino DC, Nishida K, Manley JL: *Mechanisms and consequences of alternative polyadenylation. Mol Cell* 2011, *43*(6):853–866.
7. Emrich SJ, Barbazuk WB, Li L, Schnable PS: *Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research* 2007, *17*(1):69–73.
8. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, Bryois J, Padioleau I, Romano L, Planchon A *et al*: *Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet* 2015, *11*(1):e1004958.
9. Ruhl C, Stauffer E, Kahles A, Wagner G, Drechsel G, Ratsch G, Wachter A: *Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes. Plant Cell* 2012, *24*(11):4360–4375.
10. Staiger D, Brown JWS: *Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. The Plant Cell* 2013, *25*(10):3640–3656.

11. Baralle FE, Giudice J: *Alternative splicing as a regulator of development and tissue identity. Nature reviews Molecular cell biology* 2017, 18(7):437–451.
12. Scotti MM, Swanson MS: *RNA mis-splicing in disease. Nat Rev Genet* 2016, 17(1):19–32.
13. Duque P: *A role for SR proteins in plant stress responses. Plant Signal Behav* 2011, 6(1):49–54.
14. Sun X, Yang Q, Deng Z, Ye X: *Digital inventory of Arabidopsis transcripts revealed by 61 RNA sequencing samples. Plant Physiol* 2014, 166(2):869–878.
15. Jiang F, Guo M, Yang F, Duncan K, Jackson D, Rafalski A, Wang S, Li B: *Mutations in an AP2 transcription factor-like gene affect internode length and leaf shape in maize. PLoS One* 2012, 7(5):e37040.
16. Li X, Zhu C, Yeh CT, Wu W, Takacs EM, Petsch KA, Tian F, Bai G, Buckler ES, Muehlbauer GJ *et al*: *Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res* 2012, 22(12):2436–2444.
17. Xing A, Williams ME, Bourett TM, Hu W, Hou Z, Meeley RB, Jaqueth J, Dam T, Li B: *A pair of homoeolog ClpP5 genes underlies a virescent yellow-like mutant and its modifier in maize. Plant J* 2014, 79(2):192–205.
18. Krawczak M, Reiss J, Cooper DN: *The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet* 1992, 90(1–2):41–54.
19. Lewis BP, Green RE, Brenner SE: *Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A* 2003, 100(1):189–192.
20. Kalsotra A, Cooper TA: *Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet* 2011, 12(10):715–729.
21. Misquitta-Ali CM, Cheng E, O'Hanlon D, Liu N, McGlade CJ, Tsao MS, Blencowe BJ: *Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. Mol Cell Biol* 2011, 31(1):138–150.
22. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ: *Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Bio* 2013, 14(3):153–165.
23. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: *Alternative isoform regulation in human tissue transcriptomes. Nature* 2008, 456(7221):470–476.
24. Bassols A, Costa C, Eckersall PD, Osada J, Sabria J, Tibau J: *The pig as an animal model for human pathologies: A proteomics perspective. Proteom Clin Appl* 2014, 8(9–10):715–731.
25. Zhang L, Tao L, Ye L, He L, Zhu YZ, Zhu YD, Zhou Y: *Alternative splicing and expression profile analysis of expressed sequence tags in domestic pig. Genomics Proteomics Bioinformatics* 2007, 5(1):25–34.
26. Thatcher SR, Zhou W, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao X, Baumgarten A, Li B: *Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. Plant Cell* 2014, 26(9):3472–3487.
27. Tang LT, Ran XQ, Mao N, Zhang FP, Niu X, Ruan YQ, Yi FL, Li S, Wang JF: *Analysis of alternative splicing events by RNA sequencing in the ovaries of Xiang pig at estrous and diestrous. Theriogenology* 2018, 119:60–68.
28. Zhang Y, Cui Y, Zhang X, Wang Y, Gao J, Yu T, Lv X, Pan C: *Pig StAR: mRNA expression and alternative splicing in testis and Leydig cells, and association analyses with testicular morphology traits. Theriogenology* 2018, 118:46–56.
29. Kroll JE, Kim J, Ohno-Machado L, de Souza SJ: *Splicing Express: a software suite for alternative splicing analysis using next-generation sequencing data. PeerJ* 2015, 3:e1419.
30. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A *et al*: *Proteomics. Tissue-based map of the human proteome. Science* 2015, 347(6220):1260419.
31. Buchfink B, Xie C, Huson DH: *Fast and sensitive protein alignment using DIAMOND. Nature methods* 2015, 12(1):59–60.
32. Muvarak N, Kelley S, Robert C, Baer MR, Perrotti D, Gambacorti-Passerini C, Civin C, Scheibner K, Rassool FV: *c-MYC Generates Repair Errors via Increased Transcription of Alternative-NHEJ Factors, LIG3 and PARP1, in Tyrosine Kinase-Activated Leukemias. Mol Cancer Res* 2015, 13(4):699–712.
33. Michaelis M, Gralla O, Behrends T, Scharpf M, Endermann T, Rijntjes E, Pietschmann N, Hollenbach B, Schomburg L: *Selenoprotein P in seminal fluid is a novel biomarker of sperm quality. Biochemical and biophysical research communications* 2014, 443(3):905–910.
34. Conrad M: *Transgenic mouse models for the vital selenoenzymes cytosolic thioredoxin reductase, mitochondrial thioredoxin reductase and glutathione peroxidase 4. Biochim Biophys Acta* 2009, 1790(11):1575–1585.
35. Yu X, Shi W, Cheng L, Wang Y, Chen D, Hu X, Xu J, Xu L, Wu Y, Qu J *et al*: *Identification of a rhodopsin gene mutation in a large family with autosomal dominant retinitis pigmentosa. Sci Rep* 2016, 6:19759.

Declarations

Acknowledgements

We thank Beijing Qiyuan Gene Biotechnology Co., Ltd. for offering professional assistance in data analysis to this study and Shenzhen Jinxinnong Technology Co., Ltd for offering 9 Duroc pigs.

Funding

This work was supported by the National Natural Science Foundations of China (31661143013, 31790414), the Beijing Natural Science Foundation (65199011). The funders had no role in study design, data collection and analysis, and interpretation of decision to publish, or preparation of the manuscript.

Author information

Affiliations

National Engineering Laboratory for Animal Breeding; Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture; College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China.

Wen Feng, Pengju Zhao, Xianrui Zheng, Jian-Feng Liu

Contributions

WF and JFL designed the experiments. PZ performed population analyses. WF contributed to computational analyses. XZ collected samples and prepared for sequencing. WF wrote and JFL revised the paper. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Jian-Feng Liu.

Ethics declarations

Ethics approval and consent to participate

Shenzhen Jinxinnong Technology Co., Ltd gave permission for their pigs to be used in this research. The whole sample collection and treatment were conducted in strict accordance with the protocol approved by the Institutional Animal Care and Use Committee (Number:DK1023).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Tables

Due to technical limitations, tables are only available as a download in the supplemental files section

Figures

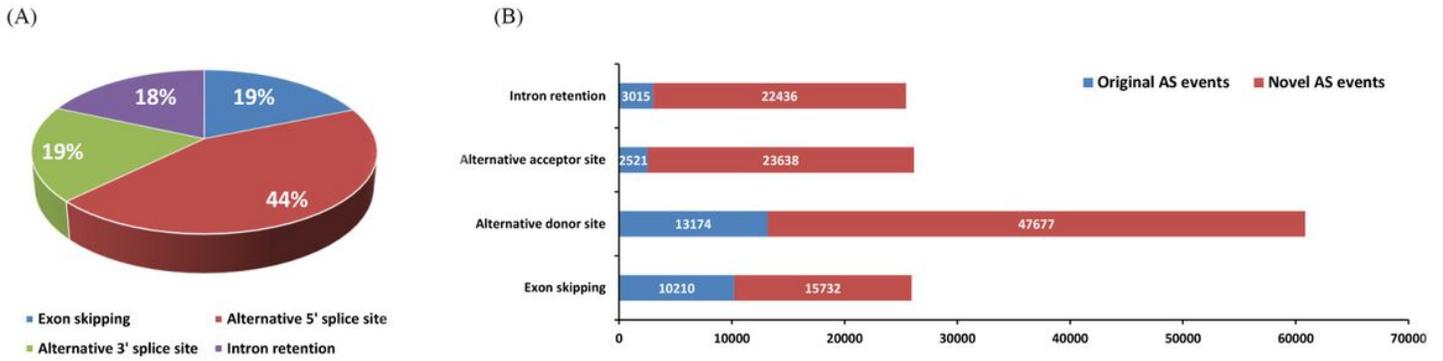


Figure 1
 Four Modes of Alternative Splicing. A: The proportion of exon skipping, alternative donor site, alternative acceptor site and intron retention. B: The number of known and novel alternative splicing events.

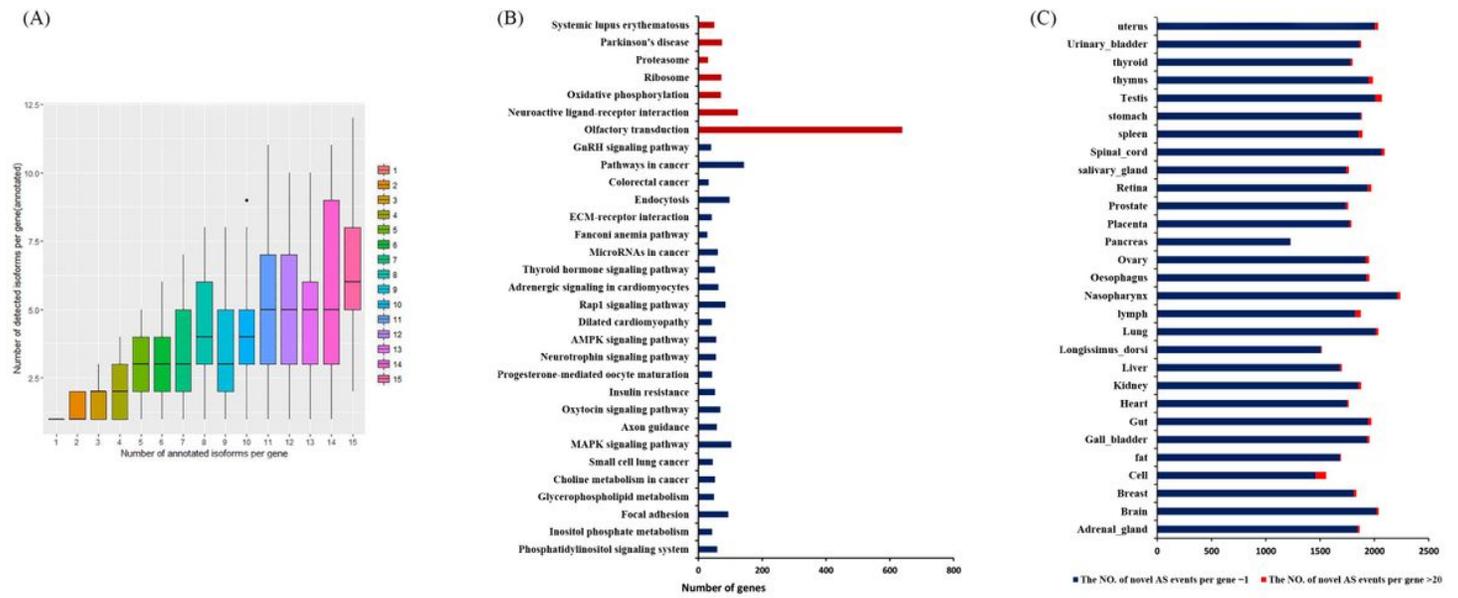


Figure 2
 The Difference between Different Numbers of Alternative Splicing Related Genes. A: The relationship of the number of annotated isoforms per gene and its novel isoforms. B: The KEGG pathways of genes with a larger and smaller number of AS events. The larger group (Blue bar): The number of AS events is greater than 20 per gene. The smaller group (Red bar): The number of AS events equals 1 belong per gene. C: The number of genes in different tissues with 1 ASE (Blue bar) and >20 ASE (Red bar).

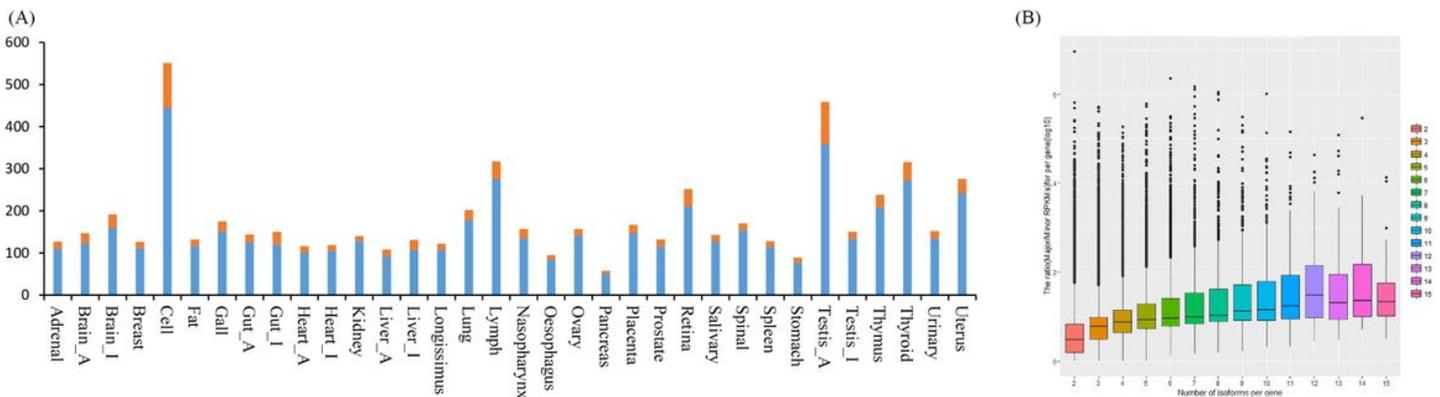


Figure 3

A: Single Tissue-Specific Transcripts. Number of known and novel transcripts that are expressed at or above 0.1 FPKM in only one tissue and <0.1 FPKM in all others. Red areas represent the transcripts and related genes are both only expressed in one and same tissue (gene expression dependent). Blue areas represent the transcripts that only expressed in one tissue, with other isoforms present in other tissues (alternative splicing dependent). B: The expression of gene with different numbers of Alternative Splicing. The Y-axis represents the ratio of the major /minor FPKM, which means the greater ratio, the greater difference between highest and lowest expression of different transcript per gene.

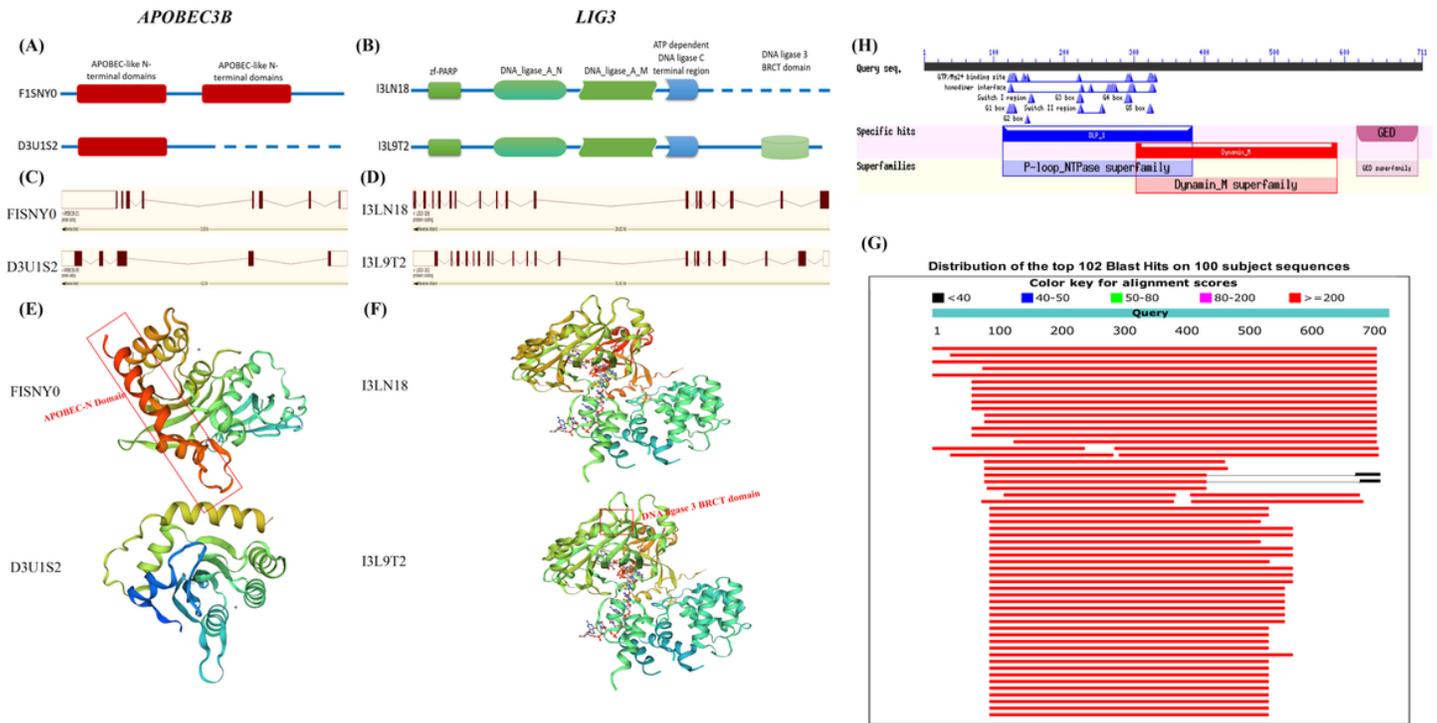


Figure 4
 Examples of Proteins Encoded by Known Transcripts (Top) Compared with Those Encoded by Novel Transcripts (Bottom). A: APOBEC3B's loss of an APOBEC-like N-terminal domain; B: LIG3's gain of a DNA ligase 3 BRCT domain; C&D: The number of coding exons of APOBEC3B and LIG3; E&F: The predicted protein conformations of APOBEC3B and LIG3. G: The conserved domains of protein MX2 between human and pig. H: The distribution of the top 102 Blast Hits on 100 subject sequences. G and H were downloaded from NCBI blast.

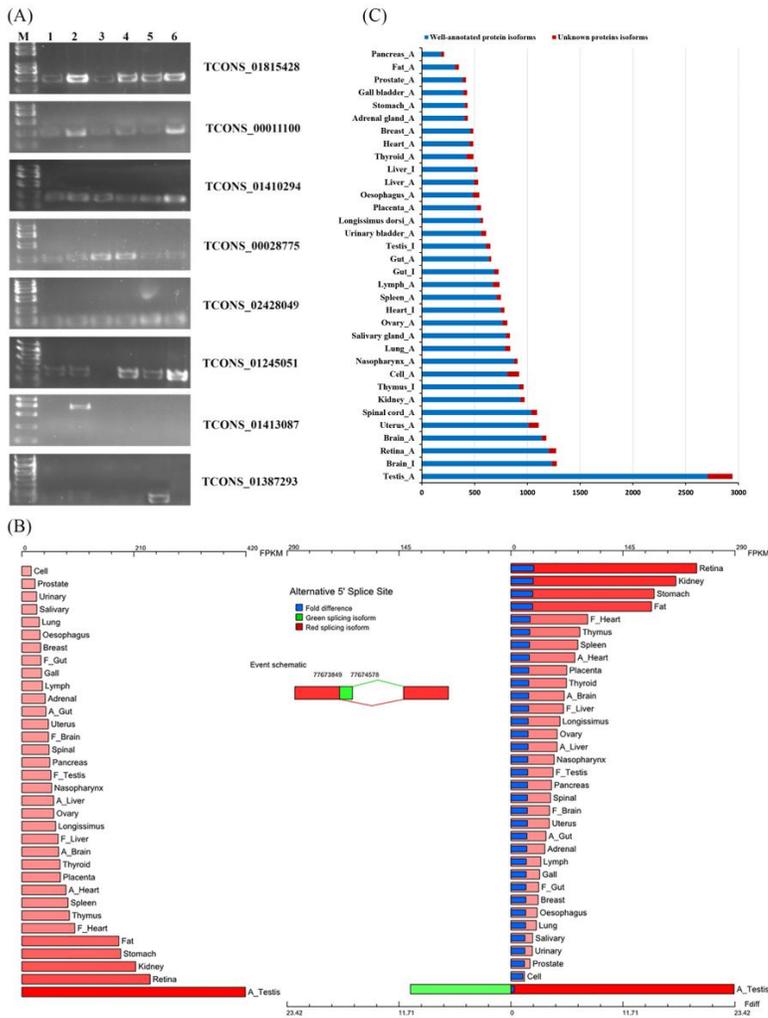


Figure 5

A: The validation of transcripts using RT-PCR. M, marker; and 1–6 represent lung, kidney, brain, gut, thyroid and lymph, respectively. B: The specific AS events of the GPX4 gene in the testis of adult pigs. Bar plots in the left side show the expression levels of GPX4 gene for each tissue. Bar plots in the right side show expression levels of GPX4 transcript for each tissue. Green and red splicing isoform represent the novel and canonical splicing, respectively. The blue bar represents fold difference (expression levels of canonical splicing / expression levels of gene). C: Numbers of tissue-enriched isoforms for known and unknown proteins. Blue bars and red bars represent the number of isoforms of well annotated proteins and unknown proteins, independently.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.xlsx](#)
- [SupplementaryTables.xlsx](#)