View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Alternative Splicing and Expression Profile Analysis of Expressed Sequence Tags in Domestic Pig

Liang Zhang[1,2#†], Lin Tao[1,2#], Lin Ye[1,2], Ling He[1,2], Yuan-Zhong Zhu[1,2], Yue-Dong Zhu[1,2], and Yan Zhou[1,2*‡]

[1] *James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310008, China;* [2] *Hangzhou Genomics Institute, Hangzhou 310008, China.*

[†] *Current address: Zhejiang Hisun Pharmaceutical Co. Ltd. (Shanghai Office), Shanghai 200233, China.*

[‡] *Current address: School of Life Sciences, Fudan University, Shanghai 200433, China.*

Domestic pig (*Sus scrofa domestica*) is one of the most important mammals to humans. Alternative splicing is a cellular mechanism in eukaryotes that greatly increases the diversity of gene products. Expression sequence tags (ESTs) have been widely used for gene discovery, expression profile analysis, and alternative splicing detection. In this study, a total of 712,905 ESTs extracted from 101 different non-normalized EST libraries of the domestic pig were analyzed. These EST libraries cover the nervous system, digestive system, immune system, and meat production related tissues from embryo, newborn, and adult pigs, making contributions to the analysis of alternative splicing variants as well as expression profiles in various stages of tissues. A modified approach was designed to cluster and assemble large EST datasets, aiming to detect alternative splicing together with EST abundance of each splicing variant. Much efforts were made to classify alternative splicing into different types and apply different filters to each type to get more reliable results. Finally, a total of 1,223 genes with average 2.8 splicing variants were detected among 16,540 unique genes. The overview of expression profiles would change when we take alternative splicing into account.

Key words: alternative splicing, expression profile, EST, clustering

Introduction

Domestic pig (*Sus scrofa domestica*), a reference model of Artiodactyla and a major source of red meat supplies to the world, is one of the most important mammals to humans. With the similar physiological characteristics to human beings, pig is becoming a useful target for human disease research and organ transplantation. Researches for the domestic pig will bring benefits to both economy and health fields.

The domestic pig originates from the Eurasian wild boar (*Sus scrofa*), and the domestication is reported to have started ~9,000 years ago (*1*). It seems that the domestication has occurred independently from wild boar subspecies in Europe and Asia, as the sequencing and analysis results of mitochondrial DNA and nuclear genes from wild and domestic pigs in Asia and Europe gave clear evidence that their ancestors diverged ~500,000 years ago (*1*). Today there exist hundreds of domestic pig breeds in the world. In our large-scale survey of expression sequence tags (ESTs) of the domestic pig, we selected several European breeds and one Asian breed to construct libraries.

Alternative splicing is a cellular mechanism in eukaryotes that greatly increases the diversity of gene products (*2*). Basically, there are five ways that a gene can be internally alternatively spliced (*3*), and three types of alignment results could be observed based on EST/cDNA data (Figure 1). For a particular gene, more than one alternative splicing possibility may apply simultaneously, and the combination gives birth to tremendous variations of gene products. About 40% of human genes are estimated to have two or more alternative splicing products (*4–7*). Different alternative splicing forms have been associated with different physiological and biological phenotypes (*8*), and tissue-specific alternative splicing has been of more interest recently (*9*).

EST sequencing, a technology that could be traced back to the early 1990s, has been widely used for gene

#Contributed equally to this work.
***Corresponding author.**
E-mail: zhouy@chgc.sh.cn

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

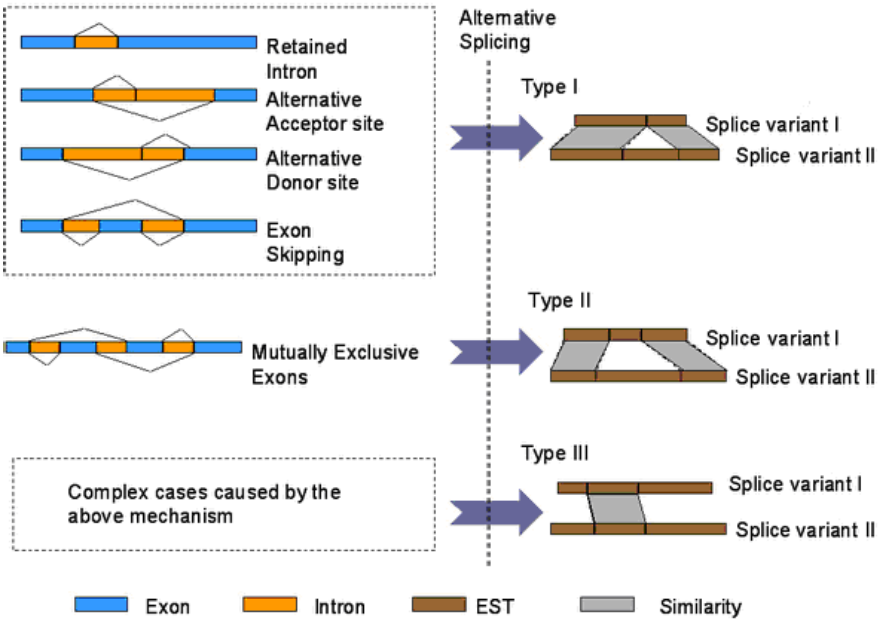

Fig. 1 Different types of alternative splicing. Five models of alternative splicing and three types of EST/cDNA alignement results are shown. Model: I, retained intron; II, alternative acceptor site; III, alternative donor site; IV, exon skipping; V, mutually exclusive exons. Type I represents the results of models I–IV; Type II represents the results of model V; whereas the causes of Type III results are not for sure up to date.

discovery, expression profile analysis, and more interestingly, alternative splicing detection, although recently gene chips (*10*) and serial analysis of gene expression (SAGE) (*11*, *12*) have been proved to be successful large-scale techniques to survey expression profiles. EST/cDNA clustering (*13*, *14*) and alignment to genome sequences for obtaining more reliable results (*9*, *15*) used to be the only approach to identify alternative splicing forms, until recently gene chips and SAGE were introduced into alternative splicing detection (*16–18*).

The most important previous work on porcine EST analysis would be the TIGR Porcine Gene Index (http://www.tigr.org/tdb/tgi/ssgi/, SsGI Release 9.0, May 14, 2004), which included 284,853 porcine EST sequences representing 84,858 unique genes, among which 468 clusters were predicted to have 2 or more alternative splicing forms. In our study, a total of 712,905 ESTs extracted from 101 different non-normalized EST libraries of the domestic pig in a project supported by the collaboration between Denmark and China were analyzed. These EST libraries cover the nervous system, digestive system, immune system, and meat production related tissues from embryo, newborn, and adult pigs, making contributions to the analysis of alternative splicing variants as well as expression profiles in various stages of tissues. A modified approach of two-step clustering with split and mergence was designed to cluster and assemble large EST datasets, aiming to detect alternative splicing together with EST abundance of each splicing variant.

Researchers were curious about the characteristics of splice sites in the early 1980s when only several tens of sequence data were available (*19*); however, the question soon moved onto the regulation of alternative splicing, since this would be a better way to fully understand the splicing mechanism. Another interesting result in human alternative splicing research is that the exons containing Alu sequences would be more frequently alternatively spliced in premature mRNA (*20*). We then searched all of the known repeat sequences in constitutively spliced exons and alternatively spliced exons predicted in pig unique genes. These results may help to find the difference of splicing regulation mechanisms between human and human's close old friend, domestic pig.

Results

Sequence quality evaluation

Contamination sequences were masked using the Cross_Match program (http://www.phrap.org), including 9,670 human mitochondrial genome sequences, 75,780 Repbase sequences (*21*), 150,799

vector sequences, and 1,372,308 low-complexity sequences. The length and location distribution of the masked segments on ESTs is shown in Figure 2. A total of 743,856 porcine EST sequences from 101 different libraries were originally prepared for our study. Among them, 30,951 short EST sequences with lengths less than 100 bp were screened out. As a result we got 712,905 high-quality porcine EST sequences with an average length of 395 bp and an average quality score of 38.5 for further analysis (Figure 3).

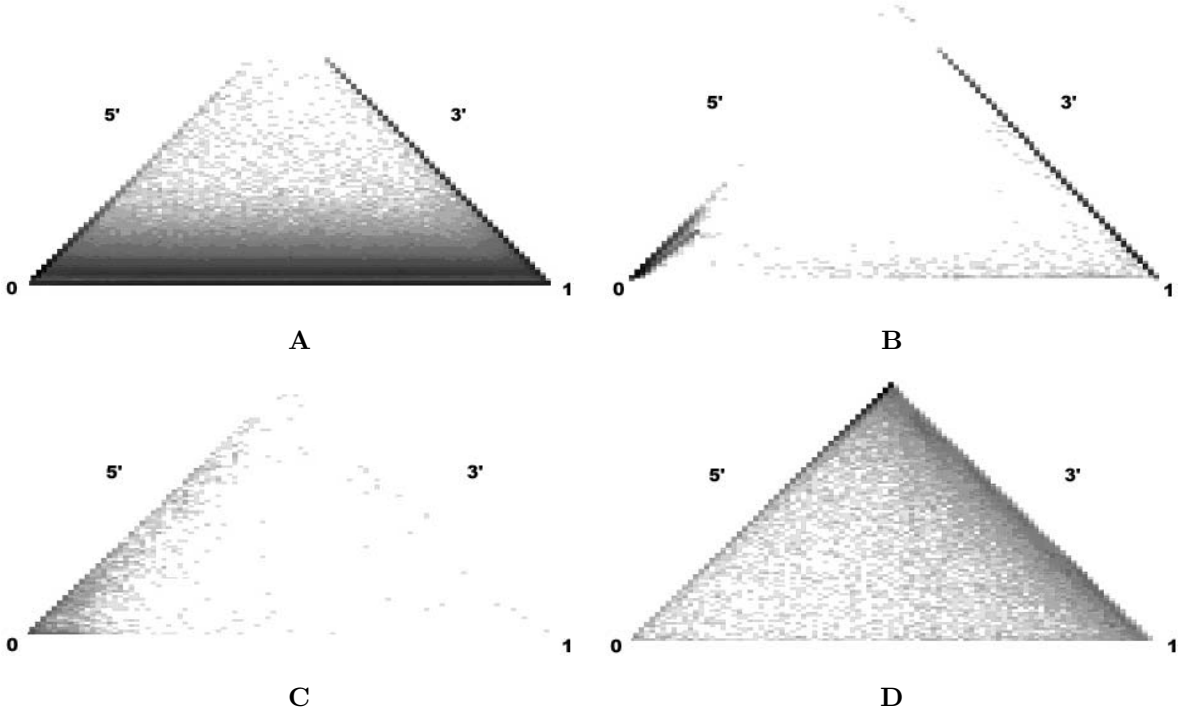

Fig. 2 Length and location distribution of the masked segments on ESTs. Each point in the figure corresponds to a masked segment. The distance between each point and the 5′ (3′) line of the triangle shows the distance between the 5′ (3′) end of the masked segment and the 5′ (3′) end of ESTs. The height of each point to the bottom line is linear to the length of the masked segment. Regions with higher density of points are in deeper color. **A**. Distribution of the total 1,450,000 masked segments by Cross_Match, among which 1,372,308 were low-complexity sequences, that is, small and simple sequences such as AAAAA, CCCCC, GGGGG, or TTTTT, existing on ESTs randomly. **B**. Distribution of 150,799 masked vector sequences, most of which are close to the 5′ or 3′ terminal of ESTs. **C**. Distribution of 9,670 masked segments with homology to the human mitochondrial genome, most of which are close to the 5′ terminal of ESTs. **D**. Distribution of 75,780 masked segments with high similarity to Repbase sequences.

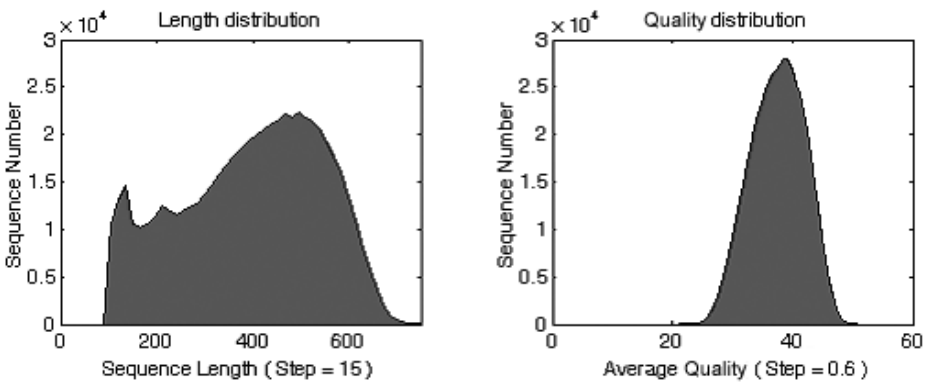

Fig. 3 Length and quality distribution of 712,905 EST sequences. The average length is 395 bp and the average quality score is 38.5.

Two-step clustering and assembly

The high-quality ESTs were then assembled into unique genes. The mathematically defined repeats (MDRs) (*22*, *23*), which appear at least 100 times in the 712,905 cleaned-up EST sequences with a minimal length of 20 bp, were trimmed temporarily. There were 433,802 ESTs with lengths more than 100 bp after trimming MDRs, and we run the PHRAP program (http://www.phrap.org) using the original sequences to get 31,050 contigs and 33,666 singletons.

The original sequences of the trimmed 279,103 ESTs before MDR trimming and 33,666 singletons were compared with the contigs by similarity analysis. ESTs of the contigs, together with the similar ESTs among the 279,103 ESTs and 33,666 singletons with an E-value threshold of 1E-15, were classified into one cluster. For each cluster, we run PHRAP again using the original sequences of the ESTs. Finally there were 53,139 contigs and 47,382 singletons.

Split and mergence to get reliable prediction

The general assumption in our alternative splicing prediction is that the ESTs expressed from the same gene will be clustered into the same cluster, and the contigs of the cluster represent different splicing variants. To get more reliable alternative splicing detection results, we designed some filters to split and merge the clusters (see Materials and Methods). As a result, the alternative splicing dataset I contained 14,492 unique genes, in which 892 clusters (genes) contained more than two alternative splicing forms with average 2.2 splicing variants (1,990 contigs). The alternative splicing dataset II contained 2,048 unique genes, including 331 clusters with average 4.3 splicing variants (1,419 contigs). To sum up, the 579,428 ESTs in alternative splicing datasets I and II represented 16,540 unique genes, with 1,223 genes having all together 3,409 splicing variants, and totally there were 18,726 contigs/singletons (Figure 4).

We then tested the splitting and merging filters on the alternative splicing dataset I and compared the consensus features between clusters that had two or more consensus sequences before and after splitting/merging (Figure 5). The figure indicates that the number of short contigs is relative low after splitting/merging (Figure 5A). The number of clusters having more than three contigs is also relatively low (Figure 5C). Contigs in the same cluster after filtering

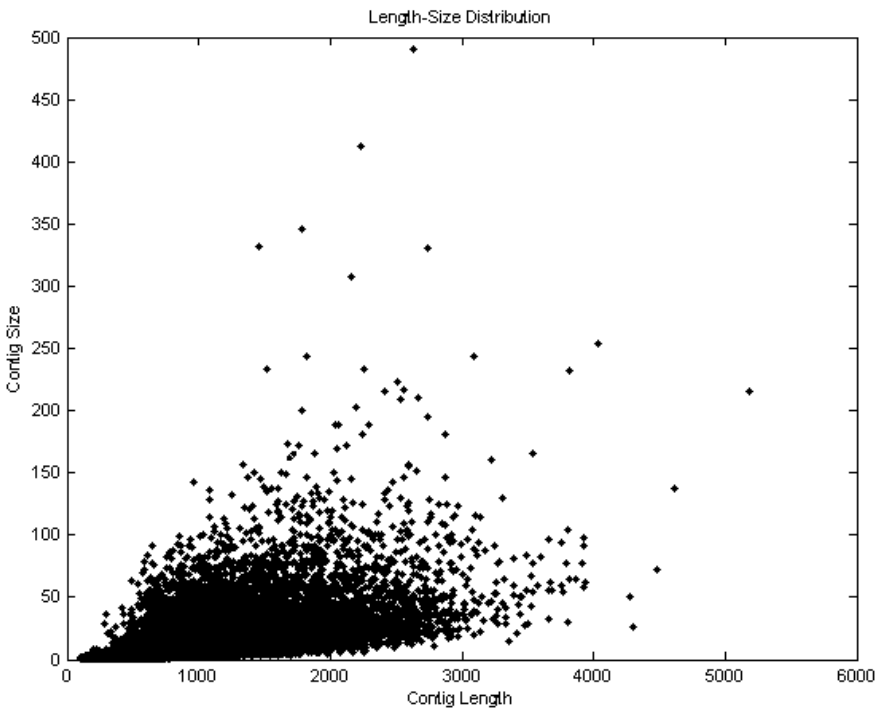

Fig. 4 Length and size distribution of totally 18,726 contigs/singletons after clustering/assembly and splitting/merging. Each point in the figure represents a contig/singleton with the value on x-axis for its length (bp) and the value on y-axis for its size.

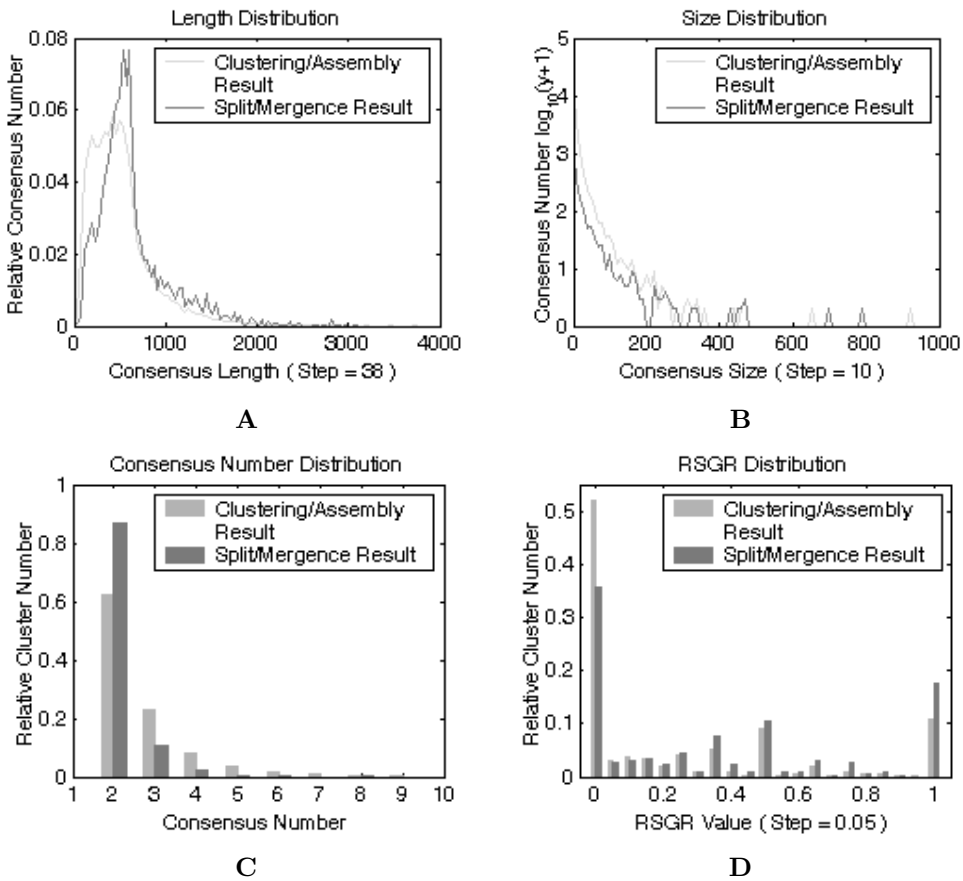

Fig. 5 Comparison of clusters for the contig consensus length (**A**), size (**B**), number (**C**), and RSGR value (**D**) before and after splitting/merging. Note that only clusters having more than two contigs in the alternative splicing dataset I are shown here. The RSGR value is in the range from 0 to 1. The value 1 indicates that the contigs in the same cluster share all of their assigned genomic reads, consequently having more chance to come from the same loci of the genome.

seem to be more likely come from the same gene by the rate of shared genomic reads (RSGR) (Figure 5D), which is used to indicate the possibility of the contigs in a cluster to be from one gene in the genome when we do not have enough genomic reads to assemble genomic sequences (see Materials and Methods).

We further used full-length porcine cDNAs and BAC sequences to validate the alternative splicing prediction results. An amount of 3,409 consensus sequences from 1,223 genes having 2 or more alternative splicing forms were aligned to 656 porcine full-length cDNA sequences (downloaded from NCBI on Oct 29, 2002). Among the 1,223 genes, 14 genes (1.1%) had all of their splicing variants aligned (similarity threshold of 80%) to the same cDNAs, whereas before splitting and merging, only 0.6% of the total contigs in the alternative splicing dataset I were aligned to the same cDNAs. The results showed that after splitting and merging, it was more reliable that the sequences in the same cluster were expressed from the same gene.

Mechanism classification of alternative splicing events

After clustering and assembly analysis, we obtained 1,223 candidate alternatively spliced genes with average 2.8 alternative splicing variants. They were classified into Type I (26%), Type II (11%), and Type III (63%) alternative splicing events as described in Introduction.

Functional classification of expressed genes

The obtained 18,726 contigs/singletons were annotated as described in Materials and Methods. As a result, there were 4,068 (21.7%) contigs/singletons annotated by BLASTX to the Swiss-Prot database, 8,631 (46.1%) annotated by BLASTN to the NCBI nt database, and 6,457 (34.5%) annotated by BLASTX to the NCBI nr database. Totally there were 9,149 contigs/singletons (48.9%) annotated.

Then the contigs/singletons were classified into Molecular Function and Biological Process catalogues of Gene Ontology (http://www.geneontology.org/) using the annotation information with the priority from Swiss-Prot to nr and nt databases. There were 1,284 contigs/singletons classified into the two catalogues by the annotation from Swiss-Prot, 2,826 by the annotation from nr, and 3,354 by the annotation from nt. A total of 4,796 contigs/singletons were successfully classified (Figure 6).

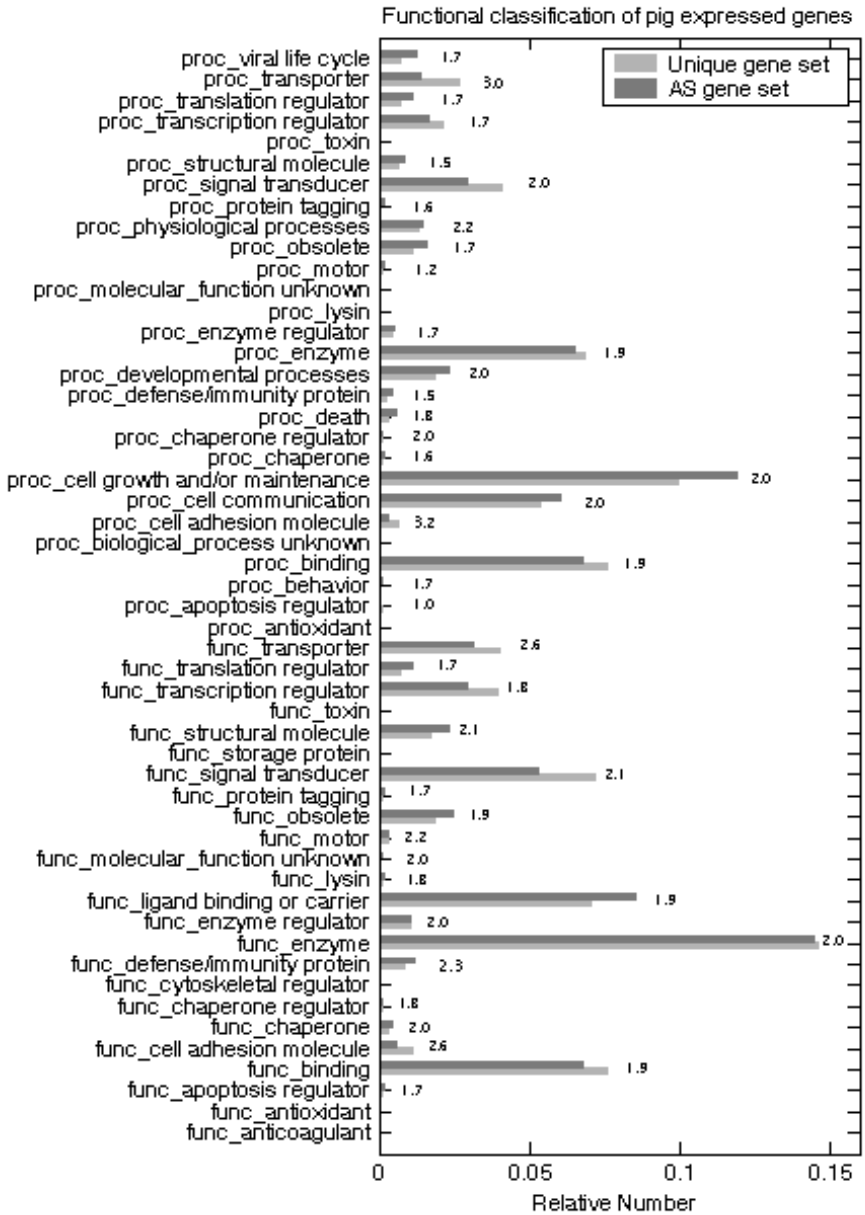

Fig. 6 Gene Oncology classification results of the unique genes (unique gene set) and the alternative splicing genes (AS gene set). The gray bar represents the relative number of the unique genes classified into a certain catalogue to all the unique genes successfully classified. The dark bar represents the relative number of the predicted alternative splicing genes classified into a certain catalogue to all the predicted alternative splicing genes successfully classified. The number on the right of each catalogue is the average number of splicing variants of the genes classified in this catalogue.

Alternative splicing rates in different tissues during development

We then compared the alternative splicing rates in different tissues during three different development stages (Table 1). It shows that the alternative splicing rate in brain is relatively higher than that in other tissues during the embryo and the newborn stages, and the alternative splicing rate in brain and thymus decreases during the grown-up process. In contrast, the alternative splicing rates in lung, intestine, and liver are higher in adult tissues than in immature tissues. This would give an overview of the relationship between alternative splicing regulation and development stages. Detailed analysis of individual genes in individual systems may be more appropriate for the follow up papers.

Expression profile analysis with alternative splicing turned on

Given the high rate of alternative splicing events in animals, we would not get clear knowledge of expression profiles until we take a close look at the expression level of each splicing variant (data not shown).

Characteristic analysis of splice sites

Finally, it would be interesting to have a glance at the splice site property after we have more than one thousand candidate alternative splicing genes in hand. We firstly did a repeat sequence survey in transcripts (Table 2). Repeats were searched in types I and II alternative splicing genes in the alternative splicing dataset I by Repeatmasker (*21*). A total of 92 contigs was found to bear repeats on them. Table 2 shows the distribution rate of repeats on alternatively spliced exons and constitutive exons. Our result indicates that the majority of repeats have no significant difference in alternatively spliced exons and constitutive exons, except LINE/L1, which is more frequently found in constitutive exons.

It was reported that some short oligonucleotide segments with specific sequences would enhance pre-mRNA splicing in human genes when located in exons, but prevent splicing when located in introns. These short sequences were called exonic splicing enhancers (ESEs) (*24*). We analyzed the distribution and frequency of some ESEs on porcine genomic reads (Figure 7). The result shows that there are no significant differences of ESE existence rate between

Table 1 Average number of splicing variants in different tissues under three development stages

Stage	Brain	Lung	Intestine	Liver	Muscle (semitendinosus)	Thymus
Embryo	1.32	1.21	1.24	1.20	1.21	1.27
Newborn	1.32	1.15	1.17	1.20	1.22	1.19
Adult	1.24	1.32	1.51	1.25	No data	No data

Table 2 Distribution rate of repeats on alternatively spliced exons (AS exons) and constitutive exons (Non-AS exons)*

Repeat name	Non-AS exon (Total 28,544 bp)	AS exon (Total 14,997 bp)
DNA/MER1_type	2	0
DNA/MER2_type	1	0
LINE/L1	18	4
LINE/L2	2	2
Low_complexity	7	3
LTR/MaLR	1	0
Other/Pig	4	1
Simple_repeat	9	3
SINE/MIR	2	1
SINE/Pig	30	21
LTR/MER4-group	1	2

*Note that the total length of all constitutive exons is 28,544 bp, about 2 folds to that of alternatively spliced exons (14,997 bp). If the repeat sequences exist randomly on exons, the expected proportion between the numbers of the two columns will be 2:1.

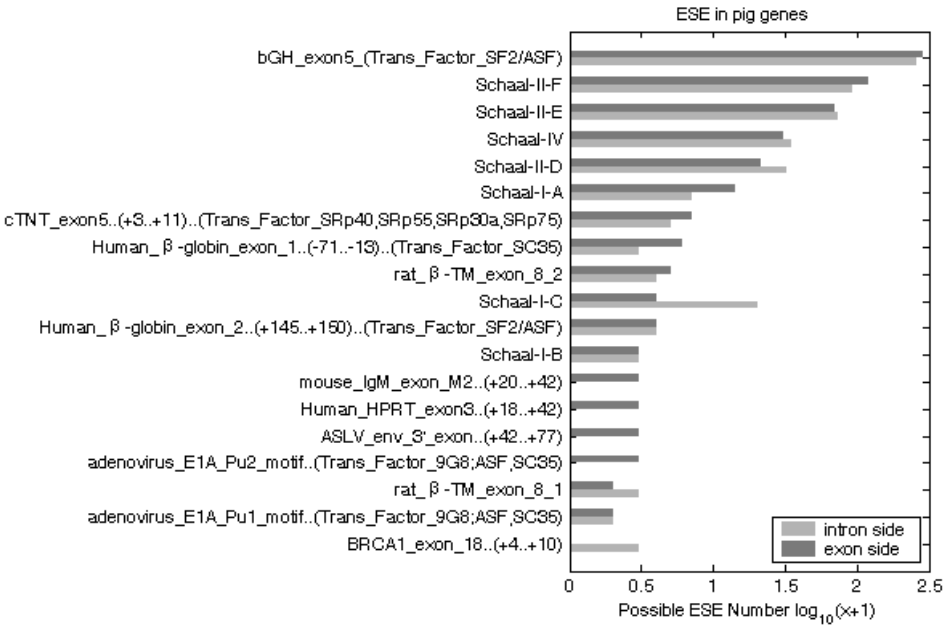

Fig. 7 Distribution and frequency of ESE sequences on porcine genomic reads. A total of 40 ESEs of human and other species were used here. We used BLASTN to find the best EST-genome read pair, and then used SIM4 to detect the splicing sites on genomic reads; only 1,378 splice sites that complied with the GT-AG rule at acceptor and donor sites of introns were used for ESE detection. For each splice site, sequences of 150 bp were extracted from both exon side and intron side to align with the ESEs by Cross_Match. As a result 19 ESEs were found to have matches on porcine genomic reads. Gray bars represent the logarithmic values of ESE matches within the intronic regions on porcine genomic sequences, and the dark bars represent those of the exonic regions.

intronic and exonic regions (except the ESE Schaal-I-C that appears more frequently in intronic regions), implying that the alternative splicing regulation signals evolve quite quickly in eukaryotes.

Discussion

Since the pig genome data have not been completed yet, the prediction and evaluation of alternative splicing forms are not reliable enough. Recent SAGE tag analysis (*11*) indicated some novel human genes, or novel transcripts, to be more precise (*17*). This is a powerful experimental method to find 3′ alternative splicing, which is complementary to our detection method based on directional cloning and 5′ sequencing strategy (see Materials and Methods). However, splicing variants located in the 5′ region of the gene would be more likely to be missed due to the extent limitation of cDNA libraries near the 5′ end of mRNA. Using 892 mRNA sequences of *S. scrofa* containing complete coding sequences, we found that our ESTs distributed evenly along the mRNA sequences. This result encouraged us that there were few 5′/3′ biases in our EST data coverage, which implies that there were few 5′/3′ biases in our alternative splicing detection.

The splicing site turns out to be one of the vaguest signals in eukaryotes (*25*). Different computational algorithms based on artificial neural networks (*26*), maximal dependence decomposition (*27*) and so on have been developed to detect and predict alternative splicing signals from coding and non-coding genomic regions (*27*, *28*). Recently, some ESEs have been successfully predicted through comprehensive statistic analysis on 6-mer oligos among exons and introns, providing experimental proof after nearly one decade of hard bench-work searching (*29–31*). Using our ESTs, we have predicted the splice sites on our porcine genomic reads. The statistic results of the frequencies that the human ESEs present on the exon and intron sides of the porcine genomic reads are listed in this study, which may help to find out whether the splicing mechanisms between human and pig are the same.

For some of the porcine genes with different splicing variants, we will design specific probes for each exon and validate the expression frequencies of different exons in porcine tissues with the gene chip technique in our future studies.

Materials and Methods

Preparation of EST sequences

The EST sequences were obtained by running Cross_Match (http://www.phrap.org) to the contamination database, including human mitochondrial genome sequences, Repbase sequences, vector sequences, and low-complexity sequences as described above.

Clustering

MDRs appearing at least 100 times with a minimal length of 20 bp were trimmed temporarily. For the ESTs with lengths more than 100 bp after trimming MDRs, we run the PHRAP program (http://www.phrap.org) using the original sequences to get contigs and singletons. The rest ESTs were analyzed by running BLASTN to get the contigs (E-value less than 1E-30). ESTs of the contigs, together with the similar ESTs with an E-value threshold of 1E-30, were classified into one cluster. For each cluster, we run PHRAP again using the original sequences of the ESTs.

For each pair of contigs/singletons with over 40% similarity, they would be considered as alternative splicing if there was at least one un-matched gap more than 30 bp in the middle of one sequence, or both $5'$ or $3'$ terminal un-matched segments of the two sequences exceeded 30 bp or 10% of their sequence length. Contigs/singletons considered as splicing variants of the same gene would be put into the same cluster. For the contigs/singletons with over 40% similarity but could not be considered as alternative splicing due to the short consensus of the contigs, they would be merged into the contigs having longer consensus, and the final contig size would be the sum of these two contigs. If a contig/singleton could neither be merged nor be considered as a splicing variant, it would be split from the cluster and be regarded as a unique gene.

Porcine genomic sequences (~1,800,000 genomic reads sequenced by Beijing Institute of Genomics) having more than 30% similarity were assigned to each contig in the clusters. We calculated the RSGR value for each cluster by the formula RSGR $= [(M/N) - 1]/(n - 1)$, where M is the sum of the read number aligned to each contig of the cluster, N is the non-duplicate read number among the total of M reads, and n is the contig number of the cluster.

Annotation

The contigs/singletons were annotated by running BLASTX to the Swiss-Prot database (E-value less than 1E-10), BLASTN to the NCBI nt database (E-value less than 1E-15), and BLASTX to the NCBI nr database (E-value less than 1E-10). For each of the contigs/singletons, the description of the best hit in the database with alignment more than 30% was given as its annotation. Then the contigs/singletons were classified into Molecular Function and Biological Process catalogues of Gene Ontology (http://www.geneontology.org/) using the annotation information with the priority from Swiss-Prot to nr and nt databases.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 30300068).

Authors' contributions

LZ and YZ designed the analyzing project, set up the principle and parameters of clustering, and drafted the manuscript. LT and LH composed the Perl program for the two-step clustering, annotation, and classification of the porcine EST sequences. LY participated in the analysis of the alternative splicing variants. YZZ carried out the sequence quality evaluation. YDZ created some of the figures with the analysis results. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Giuffra, E., *et al.* 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154: 1785-1791.
2. Sunyaev, S., *et al.* 2000. Individual variation in protein-coding sequences of human genome. *Adv. Protein Chem.* 54: 409-437.
3. McKeown, M. 1992. Alternative mRNA splicing. *Annu. Rev. Cell. Biol.* 8: 133-155.
4. Mironov, A.A., *et al.* 1999. Frequent alternative splicing of human genes. *Genome Res.* 9: 1288-1293.

5. Modrek, B., *et al.* 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 29: 2850-2859.
6. Brett, D., *et al.* 2000. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.* 474: 83-86.
7. Nigumann, P., *et al.* 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 79: 628-634.
8. Hanke, J., *et al.* 1999. Alternative splicing of human genes: more the rule than the exception? *Trends Genet.* 15: 389-390.
9. Xu, Q., *et al.* 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 30: 3754-3766.
10. Suchyta, S.P., *et al.* 2003. Development and testing of a high-density cDNA microarray resource for cattle. *Physiol. Genomics* 15: 158-164.
11. Velculescu, V.E., *et al.* 1995. Serial analysis of gene expression. *Science* 270: 484-487.
12. Yamamoto, M., *et al.* 2001. Use of serial analysis of gene expression (SAGE) technology. *J. Immunol. Methods* 250: 45-66.
13. Huang, Y.H., *et al.* 2002. PALS db: Putative Alternative Splicing database. *Nucleic Acids Res.* 30: 186-190.
14. Miller, R.T., *et al.* 1999. A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Res.* 9: 1143-1155.
15. Kan, Z., *et al.* 2001. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.* 11: 889-900.
16. Chen, J., *et al.* 2002. High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3′ complementary DNAs. *Genes Chromosomes Cancer* 33: 252-261.
17. Chen, J., *et al.* 2002. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc. Natl. Acad. Sci. USA* 99: 12257-12262.
18. Hu, G.K., *et al.* 2001. Predicting splice variant from DNA chip expression data. *Genome Res.* 11: 1237-1245.
19. Mount, S.M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10: 459-472.
20. Sorek, R., *et al.* 2002. Alu-containing exons are alternatively spliced. *Genome Res.* 12: 1060-1067.
21. Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 9: 657-663.
22. Wang, J., *et al.* 2002. RePS: a sequence assembler that masks exact repeats identified from the shotgun data. *Genome Res.* 12: 824-831.
23. Yu, J., *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79-92.
24. Fairbrother, W.G., *et al.* 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297: 1007-1013.
25. Stamm, S., *et al.* 2000. An alternative-exon database and its statistical analysis. *DNA Cell Biol.* 19: 739-756.
26. Hebsgaard, S.M., *et al.* 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res.* 24: 3439-3452.
27. Pertea, M., *et al.* 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* 29: 1185-1190.
28. Davis, C.A., *et al.* 2000. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* 28: 1700-1706.
29. Tacke, R. and Manley, J.L. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* 14: 3540-3451.
30. Coulter, L.R., *et al.* 1997. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* 17: 2143-2150.
31. Liu, C., *et al.* 2000. Alternative splicing of the histamine H(3) receptor mRNA at the third cytoplasmic loop is not detectable in humans. *Brain Res. Mol. Brain Res.* 83: 145-150.