Comparative analyses of miRNAomes between Tongcheng and Yorkshire Pigs

Yu Xi,Huijing Liu,Yuqiang Zhao,Ji Li,Wenchao Li,Guorong Liu,Jiayong Lin,Wanghong Liu,Jinlong Zhang,Minggang Lei,Debin Ni

Abstract

Tongcheng pigs (TC) and Yorkshire (YK) are two pig breeds with distinguished morphologies in muscle. A comprehensive study of porcine microRNAome (miRNAome) in longissimus muscle during 5 developmental stages (40, 55, 63, 70 and 90 dpc (days post coitum)) using Solexa sequencing technology, was carried out in order to compare miRNA expression involved in the differentiated regulation of skeletal muscle development between the TC and YK breeds. Known miRNAs and novel miRNAs were identified by alignment and annotation. MiRNA expression profiles were normalized by transcript per million (TPM). Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to cluster the samples based on the expression values of miRNAs. The short time-series expression miner (STEM) was conducted in order to further explore the temporal expression characteristics of miRNAs in the TC and YK breeds. Then the analysis of differentially expressed (DE) miRNAs was performed. Functional annotation of the predicted miRNA targets was performed based on Gene Ontology slim database (GO-Slim) and Kyoto Encyclopedia of Genes and Genomes database (KEGG). Ssc-miR-499-5p is the only multipoint (no less than 4 time points)-DE miRNA with considerable expression level during the time periods investigated. Combining the target prediction and dual luciferase reporter assay, we validated DSTN was the target gene of ssc-miR-499-5p. The results are expected to facilitate the understanding of the differences of miRNAs in myogenesis between the TC and YK breed and provide implications for the improvement of meat quality in animal production.

Citation: Yu Xi,Huijing Liu,Yuqiang Zhao,Ji Li,Wenchao Li,Guorong Liu,Jiayong Lin,Wanghong Liu,Jinlong Zhang,Minggang Lei,Debin Ni Comparative analyses of miRNAomes between Tongcheng and Yorkshire Pigs. **protocols.io**
dx.doi.org/10.17504/protocols.io.mpkc5kw
Published: 03 Feb 2018

Before start

All studies involving animals were conducted according to the relevant guidelines and regulations approved by the Ethics committee of Huazhong Agricultural University, Wuhan City, Hubei Province. All the experimental animal procedures were followed the Recommendations in the Hubei regulations for the Administration of Affairs Concerning Experimental Animals, 2005. All the animals were slaughtered after low voltage electrical stunning.

Protocol

Animal challenge, small RNA library construction and sequencing

Step 1.

Fifteen purebred TC gilts and fifteen purebred YK gilts with similar age, weight and genetic background wereobtained from the breeding pig farm of Huazhong Agricultural University (Wuhan, China). They were care and housing in the same condition according to the relevant guidelines and regulations mentioned in

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

the Ethics statement. The gilts were artificially inseminated with semen from the same purebred boars respectively. The pregnant sows were slaughtered after low voltage electrical stunning. They were at 40 dpc (days post coitus), 55 dpc, 63 dpc, 70 dpc and 90 dpc for each breed.Then we dissected the longissimus muscle tissue from the fetuses of these sows, respectively. The female fetuses were selected at 55, 63, 70 and 90 dpc and sex was ignored at 40dpc, for it had been reported that sex scarcely showed independent impact on muscle fiber development at early stage. A total of thirty muscle samples were obtained and prepared. All samples were immediately snap-frozen in liquid nitrogen and stored at −80°C until use. The total RNA from each sample was extracted using Trizol reagent (Invitrogen, Carlsbad, USA) following the manufacture's recommendations. RNA quality and quantity were evaluated using the Agilent 2100 Bioanalyzer. All RNA samples exhibited 28S/18S more than 1.5 and RNA integrity numbers (RIN) more than 8.0. Equal quantities of RNA isolated were pooled from three individual fetuses from different gilts at each time point. We constructed 10 small RNA libraries basing on the standard Illumina protocols and they were sequenced on an Illumina Genome Analyzer (Beijing Genomics Institute (BGI)). The 50nt sequence tags analysis based on Illumina HiSeqTM 2000 high-throughput sequencing uses the SBS-sequencing by synthesis.

Alignment and annotation of small RNA

Step 2.

The raw data were filtered for the following processes: A. discarding low-quality reads, B. trimming the adaptor sequences and C. eliminating sequences smaller than 18bp and reads with no insertion firstly. Then clean reads were obtained and they were mapped to genome by SOAP to analyze their expression and distribution on the genome. The reference database of the pig genome (Sscrofa10.2,ftp://ftp.ensembl.org/pub/release-76/fasta/sus_scrofa/dna/Sus_scrofa.Sscrofa10.2.dna_rm.toplevel.fa) was used for mapping.
SOAP: http://soap.genomics.org.cn/soap1/
Program and Parameters:
soap -v 0 -r 2 -M 0 -a clean.fa -D ref_genome.fa.index -o match_genome.soap
Input file:
clean.fa

A search against miRbase (version 20.00; www.mirbase.org) was conducted for all of the clean reads by BLAST.
Program and Parameters:
blastall -p blastn -F F -e 0.01
Input file:
clean.fa; hairpin_*.fa (reference); mirbase_*.txt (reference).

The reads that could not be mapped were subsequently annotated and classified by reference to exons/introns, rRNA, scRNA, snRNA, snoRNA, srpRNA and tRNA in the Genbank (www.ncbi.nlm.nih.gov) and Rfam (version 11.0; http://rfam.sanger.ac.uk) databases by BLAST.
Program and Parameters:
blastall -p blastn -F F -e 0.01
Input file:
clean.fa; *_ncgb_rRNAetc.fa (reference).

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Program and Parameters:
blastall -p blastn -F F -e 0.01
Input file:
clean.fa; Rfam.fa (reference).

To make every unique small RNA mapped to only one annotation, we followed the following priority rule: rRNAetc (in which Genbank > Rfam) > known miRNA > repeat > exon > intron. The remaining non-annotated sRNA sequences were analyzed by MIREAP (http://sourceforge.net/projects/mireap/) to predict novel miRNA candidates.
Mireap(http://sourceforge.net/projects/mireap)
Parameters for animal:
Minimal miRNA sequence length (18)
Maximal miRNA sequence length (26)
Minimal miRNA reference sequence length (20)
Maximal miRNA reference sequence length (24)
Minimal depth of Drosha/Dicer cutting site (3)
Maximal copy number of miRNAs on reference (20)
Maximal free energy allowed for a miRNA precursor (-18 kcal/mol)
Maximal space between miRNA and miRNA* (35)
Minimal base pairs of miRNA and miRNA* (14)
Maximal bulge of miRNA and miRNA* (4)
Maximal asymmetry of miRNA/miRNA* duplex (5)
Flank sequence length of miRNA precursor (10)
The sequences that were located in the porcine genome and could be folded into typical hairpin structures with nearby sequences were considered to be potential novel miRNAs. The novel miRNAs were aligned to mature miRNAs from other mammals in the miRbase by BLAST in order to obtain more information.

DATASET

sRNA Sequencing during Skeletal Muscle Development between Tongc

Expression profiles analysis and cluster analysis

Step 3.
MiRNA expression profiles were normalized by transcript per million (TPM) as the criteria previously reported. Normalization formula: TPM = Actual miRNA count/Total count of clean reads*1000000. If the normalized expression of a certain miRNA was zero, we revised its expression value to 0.01. The miRNAs with expression values of more than 28 TPM in at least one of the 10 libraries were selected for principal component analysis (PCA) and hierarchical cluster analysis (HCA). The expression values were normalized using the Z-score prior to their use for PCA and HCA. The analyses were conducted by using the R packages of gmodels.
The style of “input.txt”:
Sample1 Sample2 .....

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

```
miRNA-1	500	600
miRNA-2	400	800
.....
PCA: library(gmodels)
inname = "input.txt"
outname = "PCA.pdf"
mycolors <- c(rep("red",5), rep("blue",5))
expr <- read.table(inname, header=T, row.names=1)
data <- t(expr)
data.pca <- fast.prcomp(data)
a <- summary(data.pca)
tmp <- a[4]$importance
pro1 <- as.numeric(sprintf("%.3f",tmp[2,1]))*100
pro2 <- as.numeric(sprintf("%.3f",tmp[2,2]))*100
xmax <- max(data.pca$x[,1])
xmin <- min(data.pca$x[,1])
write.table(tmp,file="PCA result.xls",quote=FALSE,sep="\t")
write.table(data.pca$x,file="PCA values of samples.xls",quote=FALSE,sep="\t")
ymax <- max(data.pca$x[,2])
ymin <- min(data.pca$x[,2])
samples =rownames(data.pca$x)
pdf(outname)
plot(
data.pca$x[,1],
data.pca$x[,2],
xlab=paste("PC1","(",pro1,"%)",sep=""),
ylab=paste("PC2","(",pro2,"%)",sep=""),
main="PCA",
xlim=c(xmin*1.1,xmax*1.1),
ylim=c(xmin*1.1*pro1/pro2,xmax*1.1*pro1/pro2),
pch=16,col=mycolors,cex=0.33)
abline(h=0,col="gray")
abline(v=0,col="gray")
text(data.pca$x[,1],data.pca$x[,2],labels=samples,cex=0.7)
HCA: library(gplots)
data=read.table("input.txt",header=T,row.names=1)
dim(data)
tmp<-data[,1:10]
dim(tmp)
input = as.matrix(tmp)
heatmap.2(input,col=greenred(255),trace="none",density.info="none",scale="row",dendrogram="column
",margins=c(5,14),cexRow=0.2)
```

Time-series analysis

Step 4.

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Short Time-series Expression Miner v 1.3.8 (STEM, http://www.cs.cmu.edu/jernst/stem/) was used in order to cluster and visualize possible profiles of miRNA expressions varying over time. The Maximum Number of model Profiles was adjusted to 20 and the Maximum unit Change in model Profiles between Time Points was set to 1. The STEM clustering method was selected as the Clustering method and other options were set as default. The miRNA expression profiles were clustered based on statistically significant values (P-value<0.05).

Target prediction, GO and KEGG ontology analysis

Step 5.

The targets of miRNAs were extracted from three available target prediction programs, namely TargetScan (http://www.targetscan.org/), Mireap (http://sourceforge.net/projects/mireap) and miRanda (http://www.microrna.org/). The target genes identified by all the three softwares were considered to be the predicted target genes for each miRNA. Functional annotation of the predicted miRNA targets was performed based on Gene Ontology slim database (GO-Slim) and Kyoto Encyclopedia of Genes and Genomes database (KEGG) using PANTHER (http://www.pantherdb.org/) and KOBAS (http://kobas.cbi.pku.edu.cn/), respectively. Default setting was used for the statistical overrepresentation test of GO-slim and binomial test was used as the statistical method of KEGG Orthology enrichment. The enriched functional categories with a P-value <0.05 were defined as significantly enriched in the target gene candidates.

Differential expression of miRNAs

Step 6.

The analysis of differentially expressed (DE) miRNAs was performed as below: A. if the TPM of a certain miRNA is 28 or more than 28 at least one of the 10libraries, further differential expression analysis was conducted with this miRNA. B. calculate the absolute value of fold change and P-value. Fold change of miRNA expression between two samples were calculated with log 2 TPM (log 2 ratio). P-value formula:

$$p(x|y) = \left(\frac{N_2}{N_1}\right)^y \frac{(x+y)!}{x!\,y!\,\left(1+\frac{N_2}{N_1}\right)^{(x+y+1)}}$$

$$c(y \le y_{min}|x) = \sum_{y=0}^{y \le y_{min}} p(y|x)$$

$$D(y \ge y_{max}|x) = \sum_{y \ge y_{max}}^{\infty} p(y|x)$$

The x and y represent TPM of a given miRNA from the same time point between TC and YK, respectively. The N1 and N2 represent total count of clean reads in the time point betwent TC and YK, respectively.

Dual luciferase reporter assay

Step 7.

The sequence containing the miRNA binding from 3'UTR of DSTN was inserted into the pmirGLO Dual-Luciferase miRNA Target Expression Vector (Promega). The putative miRNA targeting site was varied by 4nt replaced using overlap-extension PCR. The primers used for constructing plasmid and overlap-extension PCR are listed:

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Primers for plasmid construction and point mutation in the dual luciferase reporter assay	
Primer	**Sequence (5'→3')**
DSTN 3'UTR- F	CGAGCTCACTGAGGGGAGCTGTCTTGT
DSTN 3'UTR-R	GCTCTAGACAAGTGCAGGAGCTGTTTGC
miRNA-499-DSTN-mut-a	CGAGCTCACTGAGGGGAGCTGTCTTGT
miRNA-499-DSTN-mut-b	TATTTTGGCCAATATCTGTTTCACACACACC
miRNA-499-DSTN-mut-c	GGTGTGTGTGAAACAGATATTGGCCAAAATA
miRNA-499-DSTN-mut-d	GCTCTAGACAAGTGCAGGAGCTGTTTGC
Underlined symbols are the cutting sites of *Sac I* or *Xba I* Restriction Enzyme.	

The complete plasmids were detected by double enzyme digestion and sequencing analysis. A total of 50 nM of miRNA mimics (GenePharama, Shanghai, China) or negative controls (NC) was co-transfected into PK-15 cells, in 24-well plates with wild type or mutated 3'UTR Dual Luciferase plasmid (200 ng) using X-tremeGENE HP DNA Transfection Reagent (Roche). The cells were harvested 24 hours following transfection and Dual Luciferase Reporter Assay System (Promega) was used for the luciferase activity assay. The relative luciferase activity (firefly luciferase activity/renilla luciferase activity) showed the degree of miRNA binding the 3'UTR of DSTN.

Interaction network of miRNA-mRNA

Step 8.

Four steps were done to construct interaction network of DE miRNAs and DE mRNAs.

Firstly, build a union of pairs of DE miRNAs and their targets of which the expression pattern were negative correlation in TC or YK (rho<0.5). The negative correlation were decided by Spearman's correlation coefficient for ranked data.

Secondly, pick up the differentially expressed genes (DEGs) between TC and YK from the same time points.

Thirdly, pick up the paris obtained in the first step containing DEGs.

Lastly, the interaction network was visualize and the topological property of the networks was procured by cytoscape software (http://www.cytoscape.org/). The parameter of degree in cytoscape was used to define the hub miRNA and mRNA. The gene expression and DEGs were obtained from our previous publication.

LINK:

http://www.ncbi.nlm.nih.gov/pubmed/29322263

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited