



Data in Brief

Profiling and characterization of a *longissimus dorsi* muscle microRNA dataset from an F₂ Duroc × Pietrain pig resource population



Kaitlyn R. Daza^a, Juan P. Steibel^{a,b}, Deborah Velez-Irizarry^a, Nancy E. Raney^a, Ronald O. Bates^a, Catherine W. Ernst^{a,*}

^a Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA

^b Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Keywords:

Pig
Skeletal muscle
Small RNA sequencing
miRNA

ABSTRACT

To elucidate the effects of microRNA (miRNA) regulation in skeletal muscle of adult pigs, miRNA expression profiling was performed with RNA extracted from *longissimus dorsi* (LD) muscle samples from 174 F₂ pigs (~5.5 months of age) from a Duroc × Pietrain resource population. Total RNA was extracted from LD samples, and libraries were sequenced on an Illumina HiSeq 2500 platform in 1 × 50 bp format. After processing, 232,826,977 total reads were aligned to the *Sus scrofa* reference genome (v10.2.79), with 74.8% of total reads mapping successfully. The miRDeep2 software package was utilized to quantify annotated *Sus scrofa* mature miRNAs from miRBase (Release 21) and to predict candidate novel miRNA precursors. Among the retained 295 normalized mature miRNA expression profiles sscmiR1, sscmiR133a3p, sscmiR378, sscmiR206, and sscmiR10b were the most abundant, all of which have previously been shown to be expressed in pig skeletal muscle. Additionally, 27 unique candidate novel miRNA precursors were identified exhibiting homologous sequence to annotated human miRNAs. The composition of classes of small RNA present in this dataset was also characterized; while the majority of unique expressed sequence tags were not annotated in any of the queried databases, the most abundantly expressed class of small RNA in this dataset was miRNAs. This data provides a resource to evaluate miRNA regulation of gene expression and effects on complex trait phenotypes in adult pig skeletal muscle. The raw sequencing data were deposited in the Sequence Read Archive, BioProject PRJNA363073.

Specifications	location
Organism/cell line/tissue	<i>Sus scrofa</i> /skeletal muscle
Sex	Male and female
Sequencer or array type	Illumina HiSeq 2500
Data format	Raw
Experimental factors	Pigs for transcriptional profiling selected from larger sample set based on phenotypes for loin muscle area and 10th rib backfat thickness
Experimental features	Small RNA sequencing using RNA extracted from skeletal muscle of 176 F ₂ Duroc × Pietrain pigs, and quantification and characterization of miRNAs
Consent	N/A
Sample source	Michigan State University, East Lansing, MI, USA
	1. Direct link to deposited data
	The deposited sequencing data can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA363073
	2. Experimental design, materials and methods
	2.1. Sample collection, RNA isolation and sequencing
	A subset of 176 F ₂ pigs from 44 litters (two males and two females per litter) of the Michigan State University Duroc × Pietrain Pig Resource Population (MSUPRP) were selected for transcriptional profiling [1–3]. Samples of <i>longissimus dorsi</i> (LD) tissue were collected from each animal at slaughter and frozen at –80 °C. Animal care and experimental protocols were approved by the Michigan State University

* Corresponding author at: 474 S. Shaw Lane, Room 1205, East Lansing, MI 48824, USA.
E-mail address: ernstc@msu.edu (C.W. Ernst).

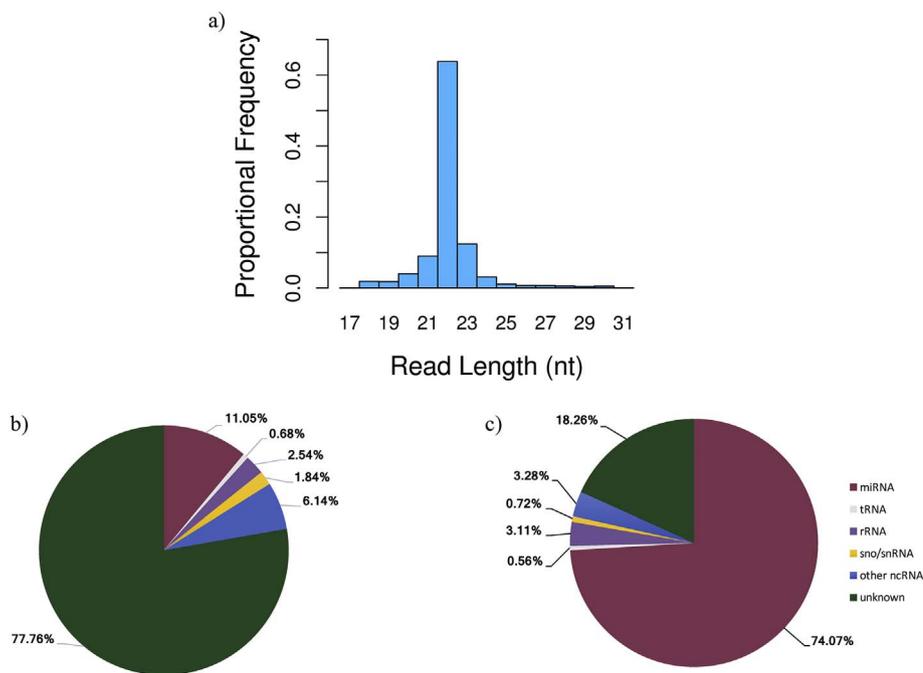


Fig. 1. Characterization of small RNA sequencing data. a) Sequence read length distribution; composition of small RNA classes present in the small RNA sequencing dataset based on b) unique sequences, and c) total read counts.

All University Committee on Animal Use and Care (AUF# 09/03-114-00). Total RNA was isolated from LD samples using the miRNeasy Mini Kit (QIAGEN, California, USA), and small RNA library construction and sequencing was performed at the MSU Research Technology Support Facility. Samples were prepared for sequencing utilizing the Bioo Scientific NEXTFlex™ Small RNA Sequencing Kit (v2; Bioo Scientific, Austin, TX, USA). Libraries were barcoded and multiplexed for sequencing on the Illumina HiSeq 2500 platform (Illumina, Inc.; California, USA) in 50 bp, single-end format. Two libraries failed to produce acceptable sequencing output and were removed from further analysis, yielding 174 files of 50 nt short read sequences in fastq format (86 males, 88 females).

2.2. Analysis of sequencing data

The 3' adaptor sequences were trimmed from raw short reads using cutadapt (cutadapt/1.4.1; [4]), and trimmed reads of length 26–38 nt were filtered for quality using FASTX toolkit (FASTX/0.0.14; http://hannonlab.cshl.edu/fastx_toolkit), removing any reads in which > 50% of the nucleotides had Phred score < 30. The selected sequence length retained an additional 4 nt on both the 5' and 3' ends of each read to accommodate the randomized adaptors ligated during library preparation, which allowed the identification of PCR duplicates. Distribution of sequence read lengths was assessed, and trimmed, filtered reads were collapsed using FASTX toolkit (FASTX/0.0.14; http://hannonlab.cshl.edu/fastx_toolkit) into unique expressed sequence tags containing a sequence identifier and the numerical count of each read (read count). PCR duplicates and random adaptor sequences were removed using the ShortRead package of R prior to genome alignment and quantification [5].

2.3. Characterization of small RNAs

To characterize classes of small RNA present in the dataset, unique expressed sequence tags underwent multiple queries utilizing BLAST+ (v2.2.30) prior to read alignment and quantification of known miRNAs. Unique sequences expressed at least two times were included in this analysis. Sequences were sequentially queried against: the *Sus scrofa* mature miRNA miRBase database (Release 21) [6], the Ensembl *Sus scrofa* ncRNA database (Release 84), and the *Sus scrofa*, *Homo sapiens*,

and *Mus musculus* Rfam databases (version 11) using blastn. The blast-short parameter was implemented as it is optimal for queries of sequences < 50 nt, and an e-value threshold of 1×10^{-5} was used to declare a significant hit. Results of each query were filtered as follows to retain unique sequence read hits: hits were retained that obtained 100% sequence identity, followed by at least 96% sequence identity, and finally hits with the maximum bitscore were retained. If multiple BLAST+ hits remained for a given sequence, the hits most identical to the sequence were retained (based on percent identity). If multiple significant hits persisted with equal sequence identity, a single hit was retained for each sequence. This yielded the composition of classes of small RNA present in the small RNA sequencing dataset, based on both the unique expressed sequence tags and total read counts.

The miRDeep2 software package (v0.0.5) was used to align high-quality reads to the *Sus scrofa* reference genome (v10.2.79; Ensembl), and to quantify *Sus scrofa* mature microRNAs (miRNAs) obtained from miRBase (Release 21) [6,7]. The average abundance of each mature pig miRNA was adjusted for differences in sequencing depth between libraries by converting the read counts to counts per million (cpm) with the edgeR package of R, incorporating TMM normalization factors [8]. miRNAs expressed at < 1 cpm in ≥ 44 libraries were removed from the dataset prior to calculation of the normalization factors.

miRDeep2 [7] was also utilized to predict candidate novel miRNAs. The software was provided the known *Sus scrofa* mature and precursor miRNA sequences, and the known *Homo sapiens* mature miRNA sequences from miRBase (Release 21) [6] to search for sequence homology for novel miRNA prediction. The resulting candidate novel precursors were filtered based on: the miRDeep2 score output by the algorithm (≥ 10), the estimated probability of the novel miRNA being a true positive ($\geq 91 \pm 2\%$), a significant Randfold *p*-value, and a minimum read count of 10 reads for both mature and star (complementary) sequences. Retained sequences were converted to DNA alphabet and FASTA format (ShortRead (R)) [5]. BLAST+ (v2.2.30) was utilized to further characterize the candidate novel miRNA precursors by searching for sequence homology to the database of known human precursor and mature miRNAs in miRBase (Release 21) [6], utilizing a stringent e-value threshold of 1×10^{-5} resulting in high-confidence homologous sequences.

All code used to analyze this dataset is publicly available and can be found at <https://github.com/perrykai>.

3. Data description

Determining underlying mechanisms controlling complex trait phenotypes such as growth, carcass composition and meat quality in pigs is important for achieving continued genetic improvement. One genetic mechanism involved in regulating these traits is the silencing of gene expression via miRNAs. Previous studies report that miRNAs exhibit dynamic expression in pig skeletal muscle across developmental stages, physiological states, and breeds [9–15]. Moreover, the signaling pathways enriched for genes that these miRNAs target are shown to be involved in myogenesis and muscle regeneration [11]. Due to the influential role of miRNAs on economically-important complex traits, it is necessary to continue characterizing miRNAs in pig skeletal muscle. Thus, the objective of this work was to profile and characterize the expression of miRNAs in LD muscle of market-age (approximately 5.5 months of age) F₂ pigs from the MSUPRP [1,2].

Assessment of sequence read length distribution revealed 63.8% of the total reads to be 22 nt long, which is the common length of mature miRNAs. This indicates the success of small RNA sequencing at isolating miRNAs (Fig. 1a), and concurs with results in Xie et al. [12] where 43.7% of total small RNA sequencing reads were 22 nt in length. In total, 232,826,977 high-quality reads (mean 1,338,086 reads per library) were aligned to the *Sus scrofa* reference genome (v10.2.79; Ensembl), with 74.8% of reads successfully mapping. Moreover, 158,672,792 reads were quantified as miRNAs from the 174 samples, corresponding to 91.2% of the mapped reads.

After filtering for abundance across libraries, 295 miRNA expression profiles were obtained (Table S1). The five most abundant miRNAs represented 47.9% of the total cpm in the dataset including ssc-miR-1, ssc-miR-133a3p, sscmiR-378, ssc-miR-206, and ssc-miR-10b. These miRNAs have previously been identified in pig skeletal muscle, including Nielsen et al. [11], where ssc-miR-1 and ssc-miR-206 were the two most highly abundant miRNAs identified from sequencing LD samples from seven 1.5–2-year-old Danish Landrace/Yorkshire crossbred pigs. miR-1, miR-206, and miR-133a are also considered the “myomiRs”, and have been well-characterized for their roles in mammalian skeletal muscle myogenesis and regeneration (for review, see [16]).

Results of the small RNA characterization analysis are shown in Fig. 1b, considering unique expressed sequence tags, and Fig. 1c, considering the total read count of each sequence. Most of the unique expressed sequence tags were not annotated in any of the five databases queried (“unknown”, 77.8%; Fig. 1b). Additionally, 11.1% of the unique expressed sequence tags (where each unique sequence was counted one time) were classified as miRNAs, and 6.1% of the unique expressed sequence tags consisted of other non-coding RNAs (Fig. 1b). When considering the sequences based on total read count, 74.1% of the sequences were identified as miRNAs, while 18.3% of the total reads were not annotated (Fig. 1c). These results show that miRNAs are the most abundant small RNA expressed in this dataset, and are consistent with results obtained in Xie et al. [12], where miRNAs were the most abundant class of small RNAs found in a pooled RNA sample from 16 pig tissues.

There were 132 candidate novel miRNAs predicted from miRDeep2 that passed filtering steps. After BLAST + query, there were 27 unique candidate novel miRNA precursors homologous to a human miRNA precursor (Table 1; additional annotation information in Table S2). Eight candidate novel precursors had more than one homologous human miRNA precursor, ranging from 2 to 6 hits per candidate novel sequence. The consensus secondary structures for the 27 unique candidate novel precursors predicted by miRDeep2 can be found in Fig. S1.

In summary, 295 known *Sus scrofa* mature miRNA expression profiles were obtained from next-generation sequencing of LD RNA from 174 F₂ Duroc × Pietrain pigs of the MSUPRP. Twenty-seven unique candidate novel pig miRNA precursors were predicted, and the composition of classes of small RNA present in the dataset were

characterized to reveal that miRNAs were the most abundant small RNA class expressed in the skeletal muscle of these pigs. This work contributes to the further characterization of microRNAs in pig skeletal muscle, which will lead to a more complete picture of the genetic regulation of complex economically-important pig production traits.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gdata.2017.07.006>.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors wish to thank Scott Funkhouser for his R programming support. Computer services were provided by the Michigan State University High Performance Computing Center.

Funding: This research was supported by Agriculture and Food Research Initiative Competitive Grant no. 2014-67015-21619 from the USDA National Institute for Food and Agriculture and by National Needs Fellowships from the National Institute of Food and Agriculture, Award No. 2012-38420-30199.

References

- [1] D.B. Edwards, C.W. Ernst, R.J. Tempelman, G.J.M. Rosa, N.E. Raney, M.D. Hoge, R.O. Bates, Quantitative trait loci mapping in an F₂ Duroc × Pietrain resource population: I. Growth traits, *J. Anim. Sci.* 86 (2008) 241–253, <http://dx.doi.org/10.2527/jas.2006-625>.
- [2] D.B. Edwards, C.W. Ernst, N.E. Raney, M.E. Doumit, M.D. Hoge, R.O. Bates, Quantitative trait locus mapping in an F₂ Duroc × Pietrain resource population: II. Carcass and meat quality traits, *J. Anim. Sci.* 86 (2008) 254–266, <http://dx.doi.org/10.2527/jas.2006-626>.
- [3] J.P. Steibel, R.O. Bates, G.J.M. Rosa, R.J. Tempelman, V.D. Rillington, A. Ragavendran, N.E. Raney, A.M. Ramos, F.F. Cardoso, D.B. Edwards, C.W. Ernst, Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs, *PLoS One* 6 (2011) e16766, <http://dx.doi.org/10.1371/journal.pone.0016766>.
- [4] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet journal* 17 (10) (2011), <http://dx.doi.org/10.14806/ej.17.1.200>.
- [5] M. Morgan, S. Anders, M. Lawrence, P. Aboyoun, H. Pagès, R. Gentleman, ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data, *Bioinformatics* 25 (2009) 2607–2608, <http://dx.doi.org/10.1093/bioinformatics/btp450>.
- [6] S. Griffiths-Jones, H.K. Saini, S. Van Dongen, A.J. Enright, miRBase: tools for microRNA genomics, *Nucleic Acids Res.* 36 (2008) D154–8, <http://dx.doi.org/10.1093/nar/gkm952>.
- [7] M.R. Friedländer, S.D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, *Nucleic Acids Res.* 40 (2012) 37–52, <http://dx.doi.org/10.1093/nar/gkr688>.
- [8] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, *Genome Biol.* 11 (2010) R25, <http://dx.doi.org/10.1186/gb-2010-11-3-r25>.
- [9] T.-H. Huang, M.-J.J. Zhu, X.-Y.Y. Li, S.-H.H. Zhao, Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development, *PLoS One* 3 (2008) e3225, <http://dx.doi.org/10.1371/journal.pone.0003225>.
- [10] T.G. McDaniel, T.P.L. Smith, M.E. Doumit, J.R. Miles, L.L. Coutinho, T.S. Sonstegard, L.K. Matukumalli, D.J. Nonneman, R.T. Wiedmann, MicroRNA transcriptome profiles during swine skeletal muscle development, *BMC Genomics* 10 (2009) 77, <http://dx.doi.org/10.1186/1471-2164-10-77>.
- [11] M. Nielsen, J.H. Hansen, J. Hedegaard, R.O. Nielsen, F. Panitz, C. Bendixen, B. Thomsen, MicroRNA identity and abundance in porcine skeletal muscles determined by deep sequencing, *Anim. Genet.* 41 (2009) 159–168, <http://dx.doi.org/10.1111/j.1365-2052.2009.01981.x>.
- [12] S.-S. Xie, X.-Y. Li, T. Liu, J.-H. Cao, Q. Zhong, S.-H. Zhao, Discovery of porcine microRNAs in multiple tissues by a Solexa deep sequencing approach, *PLoS One* 6 (2011) e16235, <http://dx.doi.org/10.1371/journal.pone.0016235>.
- [13] L. Qin, Y. Chen, X. Liu, S. Ye, K. Yu, Z. Huang, J. Yu, X. Zhou, H. Chen, D. Mo, Integrative analysis of porcine microRNAome during skeletal muscle development, *PLoS One* 8 (2013) e72418, <http://dx.doi.org/10.1371/journal.pone.0072418>.
- [14] P. Siengdee, N. Trakooljul, E. Murani, M. Schwerin, K. Wimmers, S. Ponsuksili, Transcriptional profiling and miRNA-dependent regulatory network analysis of longissimus dorsi muscle during prenatal and adult stages in two distinct pig breeds, *Anim. Genet.* 44 (2013) 398–407, <http://dx.doi.org/10.1111/age.12032>.
- [15] Z. Cai, L. Zhang, X. Jiang, Y. Sheng, N. Xu, Differential miRNA expression profiles in the longissimus dorsi muscle between intact and castrated male pigs, *Res. Vet. Sci.* (2014) 99–104, <http://dx.doi.org/10.1016/j.rvsc.2014.12.012>.
- [16] M. Horak, J. Novak, J. Bienertova-Vasku, Muscle-specific microRNAs in skeletal muscle development, *Dev. Biol.* 410 (2016) 1–13, <http://dx.doi.org/10.1016/j.ydbio.2015.12.013>.