Annotation of novel and conserved microRNA genes in the build 10 *Sus scrofa* reference genome and determination of their expression levels in ten different tissues

Bo Thomsen[1], Mathilde Nielsen[1], Jakob Hedegaard[1], Henrik Hornshøj[1], Rasmus Nielsen[1], Martien Groenen[2], Alan Archibald[3], Laurie Rund[4], Lawrence Schook[4], Jan Gorodkin[5], Christian Bendixen[1]

[1]Department of Genetics and Biotechnology, Aarhus University, Denmark, [2]Animal Breeding and Genomics Centre, Wageningen University, The Netherlands,[3]The Roslin Institute and Royal School of Veterinary Studies, University of Edinburgh, UK, [4]Department of Animal Sciences, University of Illinois, Urbana, Illinois, [5]Center for non-coding RNA in Technology and Health, University of Copenhagen, Denmark

INTRODUCTION.

MicroRNAs are short single-stranded RNA molecules that regulate gene expression post-transcriptionally by binding to complementary sequences in the 3' untranslated region of target messenger RNAs. Vertebrate genomes encode in the order of hundreds of microRNAs, each of which may regulate several mRNAs, suggesting that microRNAs participate in the regulation of essentially all biological pathways. Consequently, there is an increasing interest in determining the genomic positions of microRNA genes and characterizing their tissue-specific expression profiles. The DNA template used in the pig genome sequencing project was provided by a Duroc pig named TJ Tabasco. In an effort to annotate microRNA genes in the reference genome we have conducted deep sequencing to determine the miRNA transcriptomes in ten different tissues isolated from Pinky, a genetically identical clone of TJ Tabasco.

RESULTS

Sequence and Expression analysis of Conserved microRNAs.

The primary purpose was to generate miRNA sequences that are highly homologous to the reference genome sequence, which along with computational prediction will improve confidence in the genomic annotation of microRNA genes. Illumina Genome Analyzer IIx sequencing of ten small RNA libraries yielded 40.6 million reads of which 99% matched perfectly to the Build 10 assembly. The majority of reads (~97%) matched known microRNAs in miRBase. BLASTN identified more than 600 conserved miRNA/miRNA*, which is a significant increase relative to the 211 porcine miRNA/miRNA* deposited in the current version of miRBase. Figure 1A shows a two-way hierarchical clustering based on the Top100 most highly expressed microRNAs. Figure 1B shows the identity and abundance of the Top5 microRNAs in each of the ten tissues.

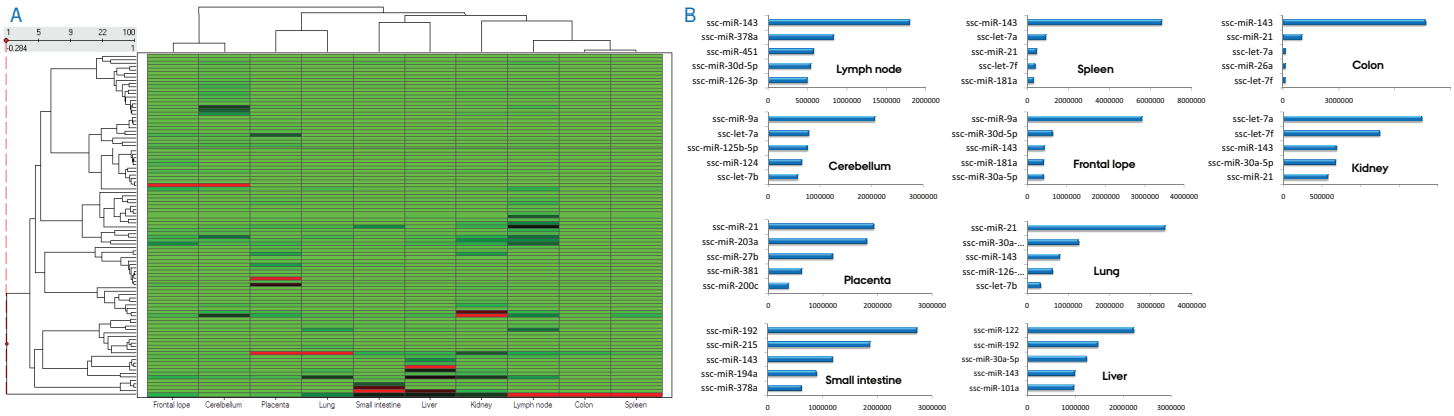

Figure 1. A. Heatmap of clustering according to expression of conserved microRNAs in ten tissues (columns, tissues; rows ,genes). B. Five most highly expressed microRNAs in each tissue.

Discovery of novel pig-specific microRNA genes.

The full complement of microRNAs of an organism comprises both conserved genes and more recently evolved species-specific genes. For *de novo* discovery of such novel, non-conserved microRNAs in the build 10 reference genome, we used miRDeep to help distinguish microRNAs from other RNA species and fragments (Friedlander et al., 2008, Nature Biotechnology). This analysis identified the genomic locations and expression levels of approximately one hundred novel microRNAs in our data set. Because these novel microRNAs are not sequence-conserved, they will remain undetected in efforts to identify pig microRNAs based on homology searches for orthologs to known microRNAs. Non-conserved, species-specific microRNAs may soon outnumber the conserved microRNAs as deep sequencing becomes more broadly applied. Figure 2 shows an example of a novel microRNA/microRNA* on chromosome 7, which is expressed at low levels in six different tissues.

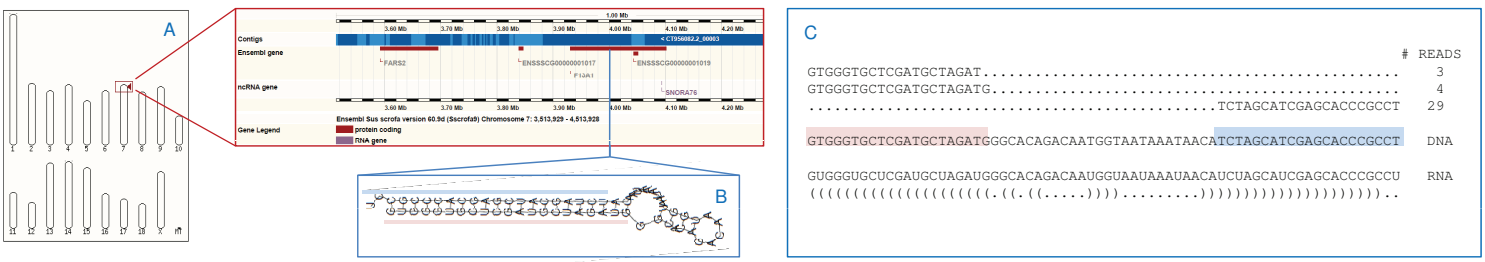

Figure 2. Novel pig-specific microRNA gene: A. Genomic position; B. stem-loop structure of the precursor; C. precursor sequence with read count numbers of mature and star sequences

DEPARTMENT OF GENETICS AND BIOTECHNOLOGY
FACULTY OF AGRICULTURAL SCIENCES
AARHUS UNIVERSITY