Rna-seq analyses in pigs: allelic expression - why make a choice?

Ole Madsen*, Djawad Radjabzadeh*, Hendrik-Jan Megens*, Mirte Bosse*, Laurent Frantz*, Yogesh Paudel*, Richard P.M.A. Crooijmans*, Laurie A. Rund#, Lawrence B. Schook#, Martien A.M. Groenen*

*Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

#Department of Animal Sciences, University of Illinois, Urbana, Illinois, USA

Introduction

- Mono-allelic gene expression is the phenomenon where only one of the parental alleles is expressed and the other is silenced.
- Different forms of mono-allelic expression have been described in mammals, including genomic imprinting, X-chromosome inactivation and allelic exclusion.
- The three types of mono-allelic expression are all regulated through epigenetic marks such as DNA methylation and histone modification, and by long non-coding RNAs (ncRNAs).
- RNA editing is the post-transcriptional chemical modification of the genome information in RNA bases.
- The most common type of RNA-editing is A → I, catalyzed by ADARs.
- The applicability of high-throughput sequencing methods for detailed discovery of individual genomic variation through whole genome (re-)sequencing and determination of individual variation in allelic expression through NGS transcriptome sequencing (RNA-seq) disclose excellent opportunities for detecting e.g. mono-allelic expressed genes and the level of RNA-editing
- Most studies on mono-allelic expression and RNA-editing has been done in human and mouse. Studies in other placental mammals is therefore essential to achieve a detailed understanding of the evolution of mono-allelic genes and RNA-editing in relation to their cellular function.

Aim

- The long term aim of this project is to detect the whole range of mono-allelic expressed and RNA-edited genes in pigs.
  - → Aim of this study is to make the first steps in developing the technical and analytic framework

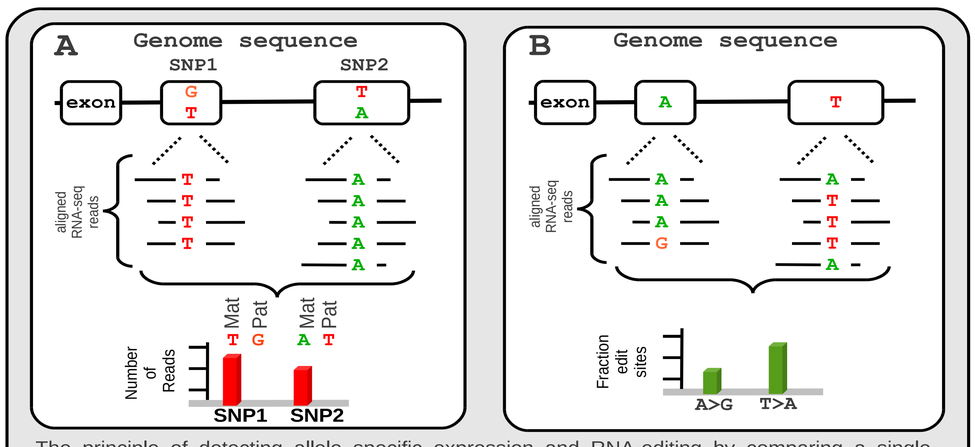
Material

- Illumina GAIIx and HighSeq paired-end sequencing of genomic DNA of different pig individuals and two family trio's (father, mother and offspring) at ~8-12x coverage.
- Illumina GAIIx and HighSeq paired-end sequencing of total-RNA cDNA libraries from different offspring/indivdual tissues (e.g. testis, placenta, brain and muscle at ~10-50x coverage for each sample)

Method

- TopHat[1] (with Bowtie implemented) will be used to align cDNA reads to the pig reference genome (v10.2).
- Genomic reads were aligned to the pig reference genome (v10.2) using Mosaik[2].
- SNPs from the transcriptome and genomic alignments will be scored using Samtools[3] and custom scripts.

The principle



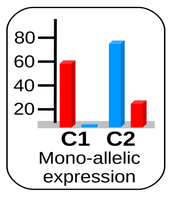
The principle of detecting allele specific expression and RNA-editing by comparing a single individuals genome/genomic variation with the variation in allelic expression in the same individuals transcriptome. RNA-seq sequences are aligned against a reference genome and the sequence variation in the RNA-seq sequences is compared to the variation present in the genome. A) Completely mono-allelic or imprinted expression. B) RNA-editing sites.

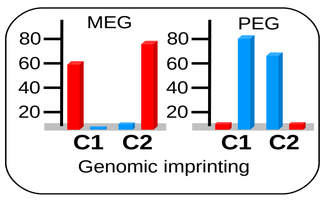
Note that genomic imprinting can only be detected if the genomic variations of both parents are available. (Pat is paternally expressed allele, Mat is maternally expressed allele).

Preliminary Results

Genomic Imprinting

- Number of genes observed as either mono-allelic expressed or genomic imprinted in three pig tissues at day 65 of fetal development.
- Most imprinted genes found in placenta, followed by brain and muscle.

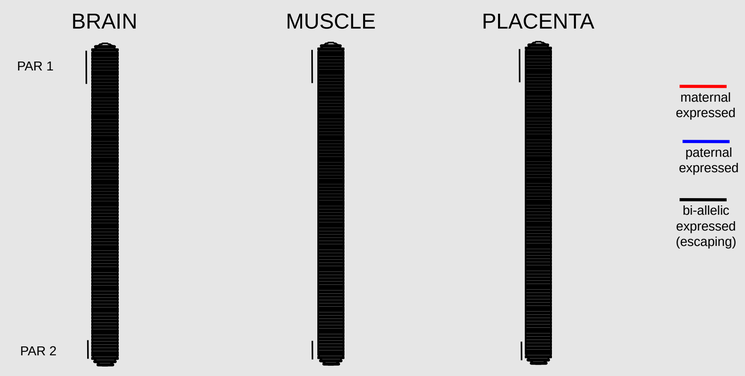




BRAIN	56	30	35
PLACENTA	53	46	27
MUSCLE	27	24	23
COMBINED	78	77	60

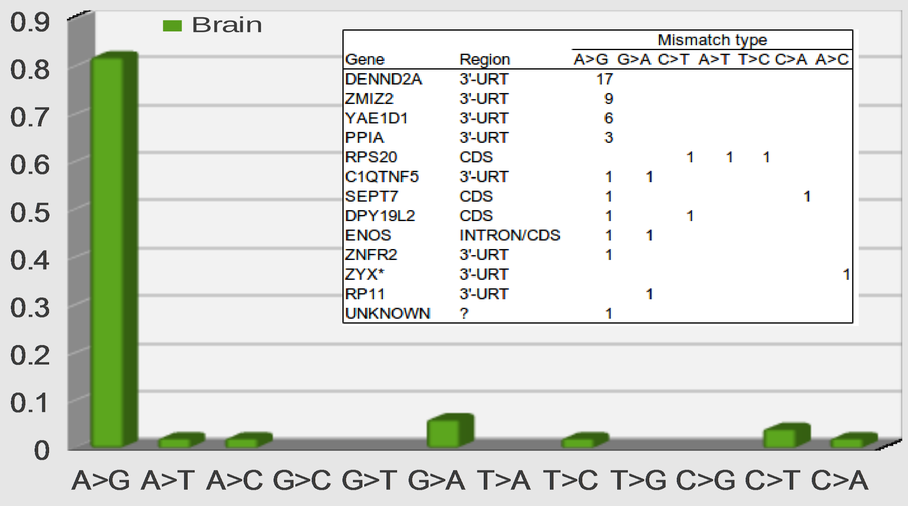
X-chromosome inactivation

- Schematic overview of genes escaping X-chromosome inactivation or change parental expression between the three examined tissues.



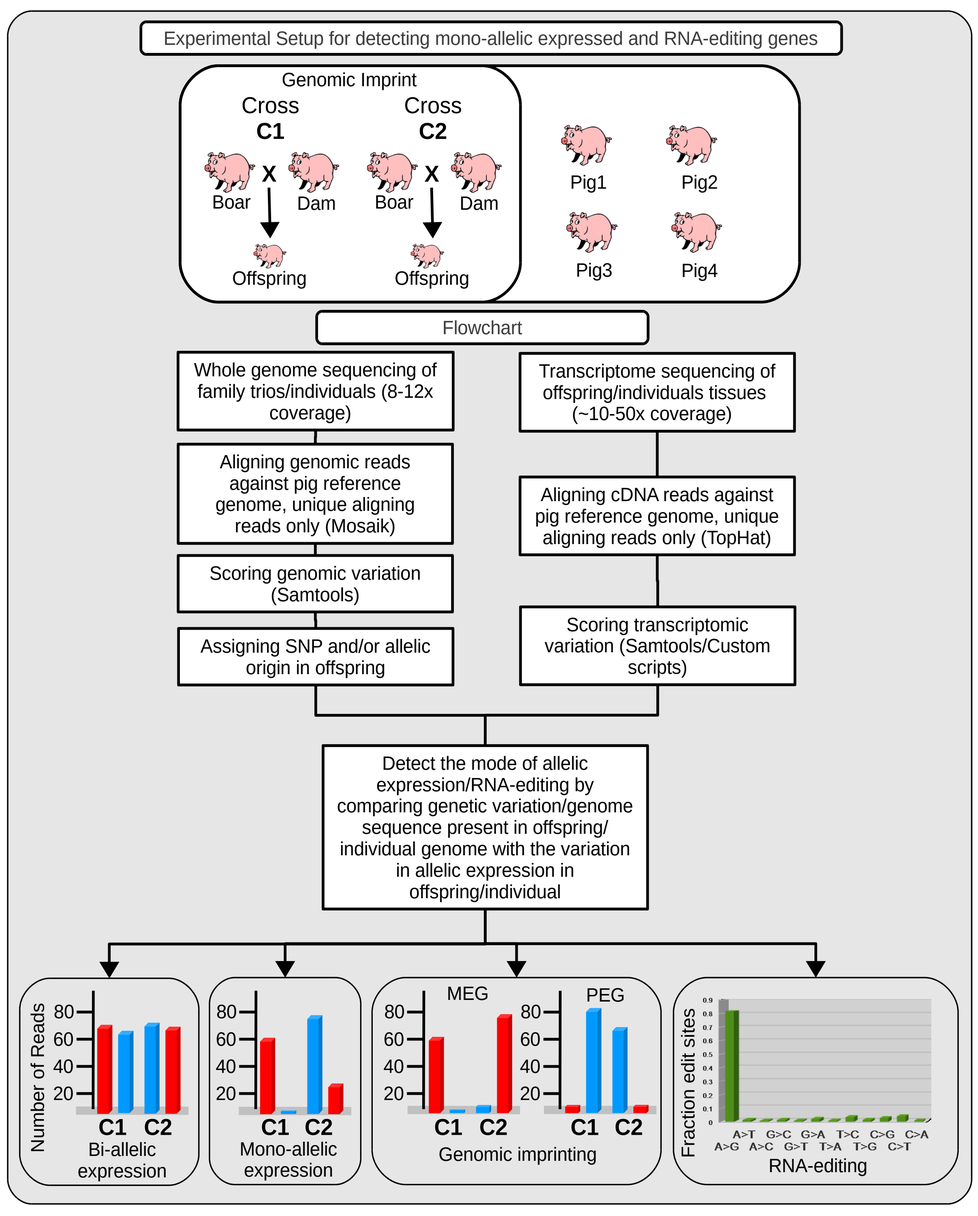
RNA-editing

- RNA-editing types on pig chromosome 18.
- Most common mismatch type is A>G, like in human and mouse
- The majority of RNA-editing sites are in 3'-UTRs



Gene	Region	Mismatch type A>G	G>A	C>T	A>T	T>C	C>A	A>C
DENND2A	3'-URT	17						
ZMIZ2	3'-URT	9						
YAE1D1	3'-URT	6						
PPIA	3'-URT	3						
RPS20	CDS			1	1	1		
C1QTNF5	3'-URT	1	1					
SEPT7	CDS	1					1	
DPY19L2	CDS	1		1				
ENOS	INTRON/CDS	1	1					
ZNFR2	3'-URT	1						
ZYX*	3'-URT							1
RP11	3'-URT		1					
UNKNOWN	?	1						

Pipeline Setup



Acknowledgement

This project was supported by National Research Initiative Grant no. 2009-34480-19875 and 2006-35216-16668 from the USDA Cooperative State Research, Education; and by the Dutch Technology Foundation (STW). We would like to thank IPG and HyPor for supplying the two family trios.

References: [1]Trapnell C, Pachter L, Salzberg SL. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**, 1105-11. [2]Li W and Strömberg M. (2010) MOSAIK. http://bioinformatics.bc.edu/marthlab/Mosaik. [3]Li H, et al. (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics **25**, 2078-9.