Profiling the landscape of transcription, chromatin accessibility and chromosome conformation of cattle, pig, chicken and goat genomes [FAANG pilot project ‘’FR-AgENCODE’’]

Foissac S., Djebali S., Acloque H., Bardou P., Blanc F., Cabau C., Derrien T., Drouet F., Esquerre D., Fabre S., Gaspin C., Gonzalez I., Goubil A., Klopp C., Laurent F., Marthey S., Marti M., Mompart F., Munyard K., Muret K., Pollet S., Queré P., Rau A., Robelin D., San Cristobal M., Tixier-Boichard M., Tosser-Klopp G., Villa-Vialaneix N., Vincent-Naulleau S., Zytnicki M., Pinard-Van der Laan MH., Lagarrigue S., Giuffra E.

Functional annotation of livestock genomes is a critical and obvious next step to derive maximum benefit for agriculture, animal science, animal welfare and human health. The aim of the Fr-AgENCODE project is to generate multi-species functional genome annotations by applying high-throughput molecular assays on three target tissues/cells relevant to the study of immune and metabolic traits. An extensive collection of stored samples from other tissues is available for further use (FAANG Biosamples “FR-AGENCODE”). From each of two males and two females per species (pig, cattle, goat, chicken), strand-oriented RNA-seq and chromatin accessibility ATAC-seq assays were performed on liver tissue and on two T cell types (CD3+CD4+ & CD3+CD8+) sorted from blood (mammals) or spleen (chicken). Chromosome Conformation Capture (*in situ* Hi-C) was also carried out on liver. Sequencing reads from the 3 assays were processed using standard processing pipelines. While most (50-70%) RNA-seq reads mapped to annotated exons, thousands of novel transcripts and genes were found, including extensions of annotated protein-coding genes and new lncRNAs (see abstract #69857). Consistency of ATAC-seq results was confirmed by the significant proportion of called peaks in promoter regions (36-66%) and by the specific accumulation pattern of peaks around gene starts (TSS) vs. gene ends (TTS). Principal Component Analyses for RNA-seq (based on quantified gene expression) and ATAC-seq (based on quantified chromatin accessibility) highlighted clusters characterized by cell type and sex in all species. From Hi-C data, we generated 40kb-resolution interaction maps, profiled a genome-wide Directionality Index and identified from 4,100 (chicken) to 12,100 (pig) topologically-associating domains (TADs). Correlations were reported between RNA-seq and ATAC-seq results (see abstract #XXXX). In summary, we present here an overview of the first multi-species and -tissue annotations of chromatin accessibility and genome architecture related to gene expression for farm animals.

Submitted version

(with limitation of 15 authors, 3500 characters and 5 keywords)

Profiling the landscape of transcription, chromatin accessibility and chromosome conformation of cattle, pig, chicken and goat genomes [FAANG pilot project ‘’FR-AgENCODE’’]

S. Foissac[1], S. Djebali*[1], H. Acloque[1], FR-AgENCODE consortium[2], M.H. Pinard-Van der Laan[2], S. Lagarrigue[3], E. Giuffra[2]; GenPhySE, INPT, ENVT, INRA, Université de Toulouse, Toulouse, France[1], GABI, AgroParisTech, INRA, Université Paris Saclay, Paris, France[2], PEGASE, INRA, AGROCAMPUS OUEST, Rennes, France[3]

Functional annotation of livestock genomes is a critical and obvious next step to derive maximum benefit for agriculture, animal science, animal welfare and human health. The aim of the Fr-AgENCODE project is to generate multi-species functional genome annotations by applying high-

throughput molecular assays on three target tissues/cells relevant to the study of immune and metabolic traits. An extensive collection of stored samples from other tissues is available for further use (FAANG Biosamples "FR-AGENCODE"). From each of two males and two females per species (pig, cattle, goat, chicken), strand-oriented RNA-seq and chromatin accessibility ATAC-seq assays were performed on liver tissue and on two T cell types (CD3+CD4+ & CD3+CD8+) sorted from blood (mammals) or spleen (chicken). Chromosome Conformation Capture (*in situ* Hi-C) was also carried out on liver. Sequencing reads from the 3 assays were processed using standard processing pipelines. While most (50-70%) RNA-seq reads mapped to annotated exons, thousands of novel transcripts and genes were found, including extensions of annotated protein-coding genes and new lncRNAs (see abstract #69857). Consistency of ATAC-seq results was confirmed by the significant proportion of called peaks in promoter regions (36-66%) and by the specific accumulation pattern of peaks around gene starts (TSS) vs. gene ends (TTS). Principal Component Analyses for RNA-seq (based on quantified gene expression) and ATAC-seq (based on quantified chromatin accessibility) highlighted clusters characterized by cell type and sex in all species. From Hi-C data, we generated 40kb-resolution interaction maps, profiled a genome-wide Directionality Index and identified from 4,100 (chicken) to 12,100 (pig) topologically-associating domains (TADs). Correlations were reported between RNA-seq and ATAC-seq results (see abstract #71581). In summary, we present here an overview of the first multi-species and -tissue annotations of chromatin accessibility and genome architecture related to gene expression for farm animals.

KEYWORDS
Multispecies
Functional Annotation of Animal Genomes (FAANG)
ATAC-seq
RNA-seq
Hi-C