558. Unravelling regulatory variants affecting gene expression in four porcine tissues

M. Gòdia[1*], M.F.L. Derks[1,2], B. Harlizius[2], O. Madsen[1] and M.A.M. Groenen[1]

[1]*Wageningen University & Research, Animal Breeding and Genomics, Wageningen, the Netherlands;*
[2]*Topigs Norsvin Research Center, Beuningen, the Netherlands;* marta.godia@wur.nl

Abstract

The aim of this work was to identify functional variants potentially associated to complex phenotypes, to improve genomic selection. In this study, we combined genomics, transcriptomics and epigenomics data. First, we genotyped 100 pigs with the high-density genotyping array (with 660K markers). Then, we generated RNA-seq data on these animals in 4 biologically divergent tissues: liver, spleen, lung and muscle. We performed an expression GWAS (eGWAS) resulting in the identification of over 150,000 significant associations in each of the tissues. We were able to detect over 1,300 genes that showed a significant eGWAS associations. We also found an enrichment of eGWAS hits within regulatory elements such as enhancers and promoters, which indicates their pivotal role as drivers of phenotypic variation and thus are of highly relevance towards understanding the genetics of complex traits.

Introduction

Unravelling the genetic basis of economically important traits in livestock remains a challenge for the animal breeding sector, and this knowledge is essential to be able to meet the increased demands for higher sustainability.

Genome-wide association studies (GWAS) have been widely used in the sector to detect genetic variants involved in relevant traits. However, this approach alone is insufficient to pinpoint the causal variants for complex traits. A major proportion of phenotypic variation is expected to be due to variation in gene expression. Hence, additional novel approaches such as RNA-seq, that quantify gene expression levels from a given tissue type is inevitable to link genomic variation to changes in gene expression. The integration of GWAS and RNA-seq data, known as expression GWAS (eGWAS) can be used to identify genetic variants associated to gene expression levels.

The eGWAS studies provide novel regulatory variation affecting gene expression. Results can be further supported by annotated regulatory elements provided by the Functional Annotation of Animal Genomes (FAANG) consortium (FAANG data portal: https://data.faang.org/home) using different epigenetic assays. These annotated regulatory elements (RE), include among others, promoters, enhancers and TSS in a variety of mammalian tissues.

The aim of this study was to perform an eGWAS to identify regulatory variants affecting gene expression in a commercial pig breeding line in four different tissues: liver, spleen, lung and muscle. Moreover, we integrated epigenomic data recently generated by the FAANG consortium to assess the likelihood that the significant variants are found within RE such as promoters and enhancers.

Materials & methods

RNA extraction, sequencing and analysis. Four different tissues: liver, spleen, lung and muscle were extracted from 100 crossbred animals. The crossbred animals are resulting from a F1 sow (Landrace×LargeWhite) mated to a Synthetic boar line. RNA extractions were done with the Rneasy kit (Qiagen) following manufacturer's instructions. RNA samples were subjected to quality control parameters

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_558, © M. Gòdia *et al.* 2022

and subjected to library preparation with NEBNext® Ultra™ Directional RNA Library Prep Kit (New England Biolabs). Samples were sequenced at 150 bp paired-end on an Illumina 6000 sequencing platform.

The bioinformatic analysis was followed as in Kern *et al.* (2021). Briefly, RNA-seq reads were trimmed using TrimGalore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Filtered reads were aligned to the *Sus scrofa* genome (Sscrofa11.1) with Ensembl version 104, using STAR v.2.7.8 (Dobin *et al.* 2013). Gene counts were assessed with htseq-count v.0.11.1 (Anders *et al.* 2014) and TMM normalized with EdgeR (Robinson *et al.* 2009).

The gene regulatory elements enhancers and promoters, specific for each of the tissues, were extracted from the FAANG database.

DNA extraction and genotyping. DNA from the 100 animals was extracted from spleen using standard phenol-chloroform protocol. After passing the quality control filter parameters, DNA was genotyped using the high-density Axiom™ Porcine Genotyping Array (Thermo Fisher Scientific) that queries 660K variant markers.

eGWAS and enrichment for regulatory regions. eGWAS included 87 samples with RNA-seq and genotype data that passed quality control filters. The RNA expression levels of the detected genes (average >1 CPM) were taken as quantitative traits and tested for association with the genotypes that passed quality control using a linear model in each of the tissues using GCTA v1.25.3 software (Yang *et al.,* 2011) as:

$$Y_i = \mu + SNP_i + kinship + e_i$$

where (Y_i) is the CPM gene expression as a function of the population mean (μ), fixed effect of each SNP (SNP_i), *kinship* based on the genomic relationship matrix (GRM) (Yang *et al.,* 2011 and a random residual effect (e_i). Only associations of FDR≤0.05 were considered.

To identify enrichment of eGWAS hits within RE, genomic regions of the eGWAS hits were iteratively randomized 100 times using BEDtools shuffle (Quinlan *et al.,* 2010). Then, the enrichment of the real eGWAS hit coordinates at gene regulatory regions was assessed by comparison with the overlap of the randomized locations with the regulatory regions using a one-sample T-test (one-tailed).

Results

RNA sequencing. On average, we obtained 44 M reads per sample and tissue. Of these, a mean of 99.9% passed quality filters and on average, 95.1, 92.0, 93.8 and 94.7% of the reads were mapped to the porcine genome in liver, spleen, lung, and muscle, respectively. Considering an expression average abundance >1 CPM in the 100 samples, the number of genes detected in liver, spleen, lung, and muscle, was 12,556, 12,568, 13,211 and 12,516, respectively. Yet, the overall profile was very different between tissues.

eGWAS. The eGWAS analysis resulted in several significant associations, with spleen having the highest number of associations (262,545), and lung the highest number of different genes (2,692) (Table 1).

The numbers of significant associations annotated as *cis*-regulatory variants (≤1 Mbp distance from the associated gene) were 40,211, 67,166, 44,870 and 32,070 in liver, spleen, lung, and muscle, respectively. We also identified several hotspot variants, where the variant was associated with the expression of several genes (≥10). The tissue with the highest number of hotspot variants was lung (3,820 variants), and the tissue with the lowest number was muscle (1,118 variants).

Table 1. Expression GWAS analysis results from the association between genotypes and RNAseq data in the four tissues.

Tissue	Number of significant associations	Number of variants involved	Number of genes involved
Liver	191,622	95,030	1,655
Spleen	262,545	124,353	1,301
Lung	241,047	87,664	2,693
Muscle	149,425	76,300	1,490

Some eGWAS peaks of interest included genes such as *RDH16* (Retinol Dehydrogenase 16) in liver and *TF* (Transferrin) in lung (Figure 1). *RDH16* is involved in vitamin A metabolism, which plays an important role in the regulation of feed efficiency in pigs by affecting energy metabolism, that mediates fatty acid biosynthesis and steroid hormone metabolism (Zhao *et al.*, 2016). *TF* has been linked to play a role in acute respiratory distress syndrome (Upton *et al.,* 2003).

Enrichment at regulatory elements. The location of significant eGWAS hits identified were queried for their overlap within RE regions. We found that eGWAS hits were enriched in promoters and enhancers regions (*P*-value<2.2E-16) in all the four tissues studied. In contrast, the number of eGWAS hit variants found in other REs including insulators, polycomb repressed or low signal elements, was significantly lower than the random set, supporting the role of certain regulatory elements in gene expression levels.

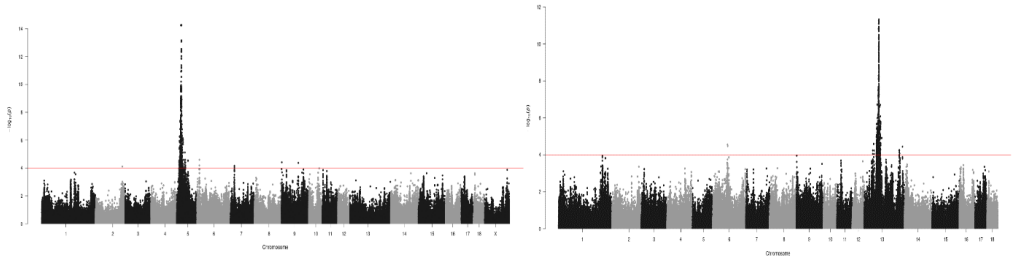

Figure 1. Manhattan plot of *RDH16* in liver (left) and *TF* in lung (right).

Discussion

In this study we report an analysis that compared the genome, transcriptome, and epigenome of biologically diverse tissues in pig. Results showed that the integrative approach of merging different techniques could help to identify variants or narrow down genomic regions involved in complex phenotypes of interest. The results can be integrated with existing QTL regions to identify potential causal variation underlying traits of interest and to understand and improve genomic selection in pig breeding programs.

Funding

This study was financially supported by the Netherlands Organization for Scientific Research (NWO) and Topigs Norsvin.

References

Anders S., Pyl P.T. and Huber W. (2015) Bioinformatics 31(2):166–169. https://doi.org/10.1093/bioinformatics/btu638

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1): 15-21. https://doi.org/10.1093/bioinformatics/bts635

Kern C., Wang Y., Xu X., Pan Z., Halstead M. *et al.* (2021) Nat Comm 12(1):1821. https://doi.org/10.1038/s41467-021-22100-8

Quinlan A.R., and Hall I.M. (2010). Bioinformatics 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Robinson M.D., McCarthy D.J., and Smyth G.K. (2009) Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Upton R.L., Chen Y., Mumby S., Gutteridge J.M. *et al.,* (2003) Eur Respir J 22(2):335–341. https://doi.org/10.1183/09031936.03.00075302

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am J Hum Genet 88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhao Y., Hou Y., Liu F., Liu A., Jing, L. *et al.,* (2016) G3 6(11):3615–3624. https://doi.org/10.1534/g3.116.032839