

EXPRESSION QUANTITATIVE TRAIT LOCI AND ALLELE-SPECIFIC EXPRESSION  
EXHIBITING JOINT ASSOCIATION WITH POLYGENIC TRAIT PHENOTYPES IN PIGS

By

Deborah Velez-Irizarry

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfilment of the requirements  
for the degree of

Animal Science–Doctor of Philosophy

2018

## ABSTRACT

### EXPRESSION QUANTITATIVE TRAIT LOCI AND ALLELE-SPECIFIC EXPRESSION EXHIBITING JOINT ASSOCIATION WITH POLYGENIC TRAIT PHENOTYPES IN PIGS

By

Deborah Velez-Irizarry

Significant genetic gain in pork production has been achieved in the past 30 years. Advancements in sequencing technology, improvements in the annotation of the pig genome, and development of quantitative genetic models were instrumental in these efforts. Several quantitative trait locus (QTL) have been identified for growth, meat quality and carcass composition phenotypes, however, the biological mechanisms underlying most QTL remain unknown. Functional genomic analysis can reveal insights on the genetic architecture of complex traits, and transcriptomic profiling of skeletal muscle during the initial steps leading to the conversion of muscle to meat can identify key regulators of meat quality and carcass phenotypes. In this study, we aimed to identify potential candidate genes and molecular markers regulating phenotypic traits using an F2 Duroc x Pietrain pig resource population. Gene transcripts obtained with RNA-seq of *longissimus dorsi* muscle from 168 F2 animals were used to estimate gene expression variation subject to genetic control by mapping expression QTL (eQTL), and identifying allele-specific expression (ASE). A total of 334 eQTL were mapped ( $FDR \leq 0.01$ ) with 187 exhibiting local acting regulation. Joint association of eQTL with phenotypic QTL (pQTL) segregating in our population revealed 16 genes significantly associated with 21 pQTL for meat quality, carcass composition and growth traits. Ten of these pQTL were for meat quality phenotypes that co-localized with one eQTL on SSC2 (8.8Mb region) and a putative hotspot associated with 11 gene transcripts on SSC15 (121Mb region). Biological processes identified for co-localized eQTL genes associated with meat quality traits included calcium signaling

(*MRLN*, *PKP2* and *CHRNA9*), energy metabolism (*SUCLG2* and *PFKFB3*) and redox hemostasis (*NQO1* and *CEP128*).

Allele specific expression (ASE) analysis facilitates the identification of cis-acting regulation of transcript abundance, which can be associated with a measurable phenotypic difference. In this study, we tested for ASE in 69,502 coding SNP (cSNP) called directly from *longissimus dorsi* transcriptomes. A total of 18,234 cSNP with significant ASE were identified ( $\text{FDR} \leq 0.01$ ). A meta-analysis merging cSNP p-values per gene identified 4,170 genes with significant allele-specific effects ( $\text{FDR} \leq 0.01$ ). A gene-wise conditional analysis fitting all ASE cSNP per gene for each phenotype identified 60 genes associated with growth, carcass composition and meat quality phenotypes. Ring finger and Zinc finger transcription factors were associated with 45-min pH, drip loss and 10<sup>th</sup>-rib backfat, and allelic expression bias for these genes was confirmed with pyrosequencing. Six genes exhibiting significant cis-acting effects and two genes associated with both cis and trans action were key regulators of the PI3K-Akt-mTOR signaling pathway. PI3K-Akt-mTOR plays an important role in skeletal muscle response to acute hypoxia, regulates cellular hypertrophy, and has been implicated in glycolytic metabolism. Results support an important role for activation of the PI3K-Akt-mTOR signaling pathway during the initial conversion of muscle to meat.

**Key words:** eQTL, ASE, skeletal muscle, pig

Copyright by  
DEBORAH VELEZ-IRIZARRY  
2018

This thesis is dedicated to my boys Onyx and Adriel, and husband Tua. Thank you for giving me the support and courage to complete my PhD.

## ACKNOWLEDGEMENTS

The completion of this dissertation was made possible thanks to the support and guidance of many people. I would like to first thank Dr. Cathy Ernst for your attention to detail and mentorship. I would like to acknowledge the rest of my committee, Dr. Juan Steibel and Dr. Robert Templeman for the assistance in learning statistical models and R programming, Dr. Ronald Bates for the insights in animal breeding and Dr. Hans Cheng for his expert advice in allele specific expression. A special thanks to the graduate students in the Breeding and Genetics Group and Animal Molecular Genetics Lab whom made this journey one of mutual respect and collaboration. Thank you Nancy Raney and Laurie Molitor for all the assistance and dedicated work in the lab. My research would not have been possible without your aid. I would also like to acknowledge the graduate student coordinators of the Department of Animal Science, Barb Sweeney and Steve Bursian, your aid in academic affairs has been instrumental; thank you for treating me like family. A special thanks to my parents Awilda and Mariano for their support and my husband Tua for tackling this journey head on with me. You encouraged me to continue and made the impossible possible. Last but not least, I would like to thank my sons, Onyx and Adriel, you are my main drivers and I encourage you to follow your dreams as you have helped me follow mine.

## TABLE OF CONTENTS

|                                                                                                                                                 |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------|----|
| LIST OF TABLES .....                                                                                                                            | ix |
| LIST OF FIGURES .....                                                                                                                           | x  |
| KEY TO ABBREVIATIONS .....                                                                                                                      | xi |
| CHAPTER ONE .....                                                                                                                               | 1  |
| Introduction .....                                                                                                                              | 1  |
| REFERENCES.....                                                                                                                                 | 7  |
| CHAPTER TWO .....                                                                                                                               | 13 |
| Genetic control of <i>longuissimus dorsi</i> muscle gene expression and joint association with phenotypic quantitative trait loci in pigs ..... | 13 |
| ABSTRACT .....                                                                                                                                  | 13 |
| INTRODUCTION.....                                                                                                                               | 14 |
| MATERIALS AND METHODS .....                                                                                                                     | 15 |
| Pig population and phenotype collection .....                                                                                                   | 15 |
| Genotyping .....                                                                                                                                | 16 |
| RNA extraction and RNA sequencing .....                                                                                                         | 17 |
| RNA-seq count normalization and transformation.....                                                                                             | 18 |
| Heritability of phenotype and gene expression .....                                                                                             | 19 |
| Genome wide association .....                                                                                                                   | 20 |
| Local and distant regulators.....                                                                                                               | 21 |
| Co-localization analysis.....                                                                                                                   | 22 |
| RT-qPCR.....                                                                                                                                    | 23 |
| RESULTS .....                                                                                                                                   | 24 |
| Identification of eQTL .....                                                                                                                    | 24 |
| Local versus distant regulators of gene expression .....                                                                                        | 26 |
| Heritability of gene expression.....                                                                                                            | 28 |
| Phenotypic QTL .....                                                                                                                            | 30 |
| Co-localization of phenotypic QTL with expression QTL .....                                                                                     | 31 |
| RT-qPCR confirmation of CHRNA9.....                                                                                                             | 41 |
| DISCUSSION .....                                                                                                                                | 42 |
| CONCLUSION .....                                                                                                                                | 52 |
| SUPPLEMENTARY MATERIALS .....                                                                                                                   | 54 |
| REFERENCES.....                                                                                                                                 | 55 |
| CHAPTER THREE.....                                                                                                                              | 63 |
| Allele-specific expression in <i>longuissimus dorsi</i> muscle transcriptomes associated with phenotypic traits in pigs .....                   | 63 |
| ABSTRACT .....                                                                                                                                  | 63 |
| INTRODUCTION.....                                                                                                                               | 64 |
| MATERIALS AND METHODS.....                                                                                                                      | 65 |

|                                                                      |     |
|----------------------------------------------------------------------|-----|
| RNA extraction and RNA-seq bioinformatics pipeline .....             | 65  |
| cSNP calling and unbiased allele specific read mapping .....         | 67  |
| Allele-specific expression analysis.....                             | 69  |
| Confirmation of ASE cSNP .....                                       | 70  |
| Kegg pathways and gene ontology enrichment.....                      | 72  |
| Effects of ASE cSNP on trait phenotypes.....                         | 72  |
| Phenotypic QTL mapped with cSNP.....                                 | 73  |
| RESULTS .....                                                        | 74  |
| Identification of cSNP .....                                         | 74  |
| Allele-specific expression .....                                     | 75  |
| Pyrosequencing to confirm cSNP with allele-specific expression ..... | 77  |
| Gene ontology and Kegg pathway enrichment .....                      | 78  |
| Effects of cSNP on trait phenotypes .....                            | 80  |
| DISCUSSION .....                                                     | 83  |
| CONCLUSION .....                                                     | 90  |
| SUPPLEMENTARY MATERIALS .....                                        | 91  |
| REFERENCES.....                                                      | 92  |
| CHAPTER FOUR.....                                                    | 98  |
| Conclusions .....                                                    | 98  |
| FUTURE RESEARCH DIRECTIONS.....                                      | 102 |
| REFERENCES .....                                                     | 106 |

## LIST OF TABLES

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Table 1.1 Review of eQTL studies conducted in pig populations .....                  | 6  |
| Table 2.1 eQTL summary among regulator types.....                                    | 25 |
| Table 2.2 Heritability summary for all genes and genes with an associated eQTL ..... | 30 |
| Table 2.3 Phenotypic QTL co-localized with expression QTL.....                       | 32 |
| Table 2.4 Expression QTL significantly associated with phenotypic traits .....       | 33 |
| Table 2.5 Comparison of RT-qPCR and RNA-seq for CHRNA9 gene expression .....         | 42 |
| Table 3.1 cSNP selected for pyrosequencing confirmation .....                        | 78 |
| Table 3.2 ASE genes enriched in Kegg pathways.....                                   | 80 |
| Table 3.3 ASE genes enriched in GO terms for biological processes .....              | 81 |
| Table 3.4 Phenotypic QTL mapped with cSNP .....                                      | 82 |
| Table 3.5 Gene-wise conditional analysis cSNP showing ASE .....                      | 82 |

## LIST OF FIGURES

|                                                                                                                                                             |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2.1 Manhattan plots illustrating the classification of different types of gene expression regulation based on eQTL position.....                     | 27 |
| Figure 2.2 eQTL map.....                                                                                                                                    | 28 |
| Figure 2.3 Heritability of transcript profiles .....                                                                                                        | 29 |
| Figure 2.4 Manhattan plots of meat quality and carcass composition pQTL co-localized with eQTL .....                                                        | 34 |
| Figure 2.5 Manhattan plots of growth pQTL co-localized with eQTL.....                                                                                       | 35 |
| Figure 2.6 Pearson correlations among phenotypic traits with an associated pQTL .....                                                                       | 37 |
| Figure 2.7 Proportion of variance explained by pQTL SNP for phenotypes (blue) and transcript abundance (green).....                                         | 38 |
| Figure 2.8 Proportion of phenotypic variance explained by PRKAG3 T30N SNP and putative hotspot SNP H3GA0052416 for meat quality traits mapped to SSC15..... | 41 |
| Figure 3.1 RNA-seq bioinformatics pipeline for cSNP calling .....                                                                                           | 67 |
| Figure 3.2 Filtering pipeline to remove potentially erroneous cSNP calls .....                                                                              | 69 |
| Figure 3.3 Number of cSNP called from RNA-seq and allelic ratios .....                                                                                      | 76 |
| Figure 3.4 Comparison of gene transcripts exhibiting significant ASE and associated with an eQTL .....                                                      | 77 |
| Figure 3.5 Histograms of ARs obtained with pyrosequencing for nine ASE cSNP .....                                                                           | 79 |
| Figure 3.6 Manhattan plots for pQTL mapped using cSNP called directly from the <i>longissimus dorsi</i> transcriptome of 168 animals.....                   | 83 |

## KEY TO ABBREVIATIONS

ASE = allele specific expression

AR = allelic ratio

BIN1 = bridging integrator 1

CEP128 = centrosomal protein 128

CHRNA9 = cholinergic receptor nicotinic alpha 9 subunit

CIT = citron Rho-interacting serine/threonine kinase

cSNP = coding SNP

eQTL = expression quantitative trait locus

FBXW7 = F-box and WD repeat domain containing 7

FDR = false discover rate

FRMD8 = FERM domain-containing 8 gene

GBLUP = genomic best-linear unbiased prediction

GWAS = genome wide association study

$h^2$  = heritability

HIF-1 = hypoxia inducible factor 1

HPD = hepsin

IGF1 = insulin like growth factor 1

IGFBP5 = insulin like growth factor binding protein 5

l. dorsi = *longuissimus dorsi*

LD = linkage disequilibrium

MRLN = myoregulin

MSUPRP = Michigan State University Pig Resource Population

NAMPT = nicotinamide phosphoribosyltransferase

NQO1 = NADPH quinone oxidoreductase -1

OBSL1 = obscurin like 1

PFKFB3 = 6-phosphofructo-2-kinase / fructose-2-,6-biphosphatase 3

PI3K-Akt-mTOR = phosphatidylinositol-3-kinase/Akt/mammalian target of rapamycin

PKP2 = plakophilin

PPARGC1B = PPARG coactivator 1 beta

pQTL = phenotypic quantitative trait locus

PRKAG3 = protein kinase AMP-activated non-catalytic subunit gamma3

QTL = quantitative trait locus

RNF141 = ring finger protein 14

RNF150 = ring finger binding protein 150

SNP = single nucleotide polymorphism

SSC = *Sus scrofa* chromosome

SSX21P = synovial sarcoma X breakpoint 2 interacting protein

SUCLG2 = succinate-COA ligase GDP-forming beta subunit

TMM = trimmed mean M-values

TOR1B = torsin B

TYW3 = TRNA-YW synthesizing protein 3 homolog

VEGF = vascular endothelial growth factor

WASP = allele-specific pipeline for unbiased read mapping

WBS = Warner Bratzler shear force

## CHAPTER ONE

### Introduction

The use of genomic improvement techniques by swine breeders has significantly advanced pork production<sup>1</sup>. Genomic regions harboring single nucleotide polymorphisms (SNP) contributing significant portions of phenotypic variation have been observed for economically important trait phenotypes in swine populations. These genomic regions are known as quantitative trait locus (QTL). Our group has used an F2 pig resource population over the past decade to identify QTL for growth, body composition and meat quality traits<sup>2-9</sup>. The Michigan State University Pig Resource Population (MSUPRP) was developed from an outcross between Duroc and Pietrain to detect candidate variants associated with quantitative traits. These breeds were selected for their tendency to differ in growth, leanness and meat quality phenotypes<sup>10</sup>. QTL have been identified in this population using linkage mapping<sup>2-5</sup> and a high density SNP panel (ProcineSNP60 bead chip)<sup>6-8</sup>. Efforts to fine map the identified QTL genomic regions have been pursued using different approaches, such as increasing the number of microsatellite markers surrounding QTL<sup>4,5</sup>, performing meta-analysis combining independent genome wide association (GWA) studies<sup>6,7</sup> and restricting analysis to 2 Mb regions surrounding the QTL marker with the lowest p-value<sup>8,9</sup>.

The extent of LD<sup>11</sup> and small effective population size in pigs<sup>12,13</sup> limits the resolution to identify candidate variants. Large effect QTL under selection tend to cluster among LD blocks, usually spanning large genomic regions<sup>14</sup>. Because the MSUPRP is an F2 cross, the number of recombination events is reduced, which consequently limits the resolution of QTL intervals resulting in large QTL regions encompassing numerous SNP in close linkage disequilibrium (LD) with a causative variant. This LD structure while beneficial for genomic selection

complicates the identification of candidate gene influencing QTL regions. Expression QTL (eQTL) analysis aims to identify genes whose expression is subject to genetic control by modeling gene expression as a response variable. A gene's function can be defined by its product and pattern of expression, which is regulated by a large number of functional elements at a given developmental period and environmental condition<sup>15</sup>. The goal of eQTL studies is to prioritize variants with functional relevance in biological processes conferring measurable fitness in economically important trait phenotypes. The joint association of eQTL regions with phenotypic QTL (pQTL) regions in a single population can aid in the identification of candidate genes whose expression is transcriptionally regulated by SNP associated with phenotypic variation.

Early eQTL maps of the swine genome were constructed with microarray gene expression data and microsatellite markers<sup>16–21</sup>. These early studies reduced the number of candidate genes obtained through QTL mapping by identifying positional candidate cis-acting eQTL coinciding with QTL regions. Cis-acting regulators of gene expression identify candidate locus directly influencing the expression of the associated gene, and thus infer direct cause of variation in gene expression. In contrast trans-acting regulators and regulatory hotspots may affect the expression of distant genes through gene-gene interactions. Initial eQTL studies were of low resolution due to the limited coverage of few microsatellite markers across the genome (typically 115 – 170 markers)<sup>16–21</sup>. Our group has previously mapped eQTL for the MSUPRP<sup>21</sup> using microarray gene expressions for *longuissimus dorsi* (l. dorsi) tissue and 124 microsatellite markers.

Microarrays are known to have technical issues with hybridization and quantification of genes with low transcript abundance<sup>22,23</sup>. The application of next generation sequencing technologies overcomes these limitations, and RNA-seq data has been shown to outperform

microarrays for evaluating both known and novel genes, and allows better quantification of lowly expressed transcripts<sup>22,23</sup>. In this project, we build on our previous work<sup>21,24</sup> to increase the resolution of our eQTL map for the l. dorsi transcriptome using high density SNP genotypes and RNA-seq data to increase the genome-wide coverage of gene expression regulation. The integration of pQTL and eQTL analysis for the same population increases our scope of inference to elucidate the biological architecture driving differences between divergent trait phenotypes. Such approaches have identified candidate genes and gene networks regulating meat quality traits<sup>18,19,21,25–31</sup>, disease resistance<sup>20,32–36</sup> and stress response<sup>17,37</sup> in swine populations (Table 1.1).

The overall goal of this dissertation research is to elucidate functional variants and candidate genes associated with variation in polygenic traits in pigs by identifying positional candidate eQTL and cis-acting regulators of gene expression associated with pQTL regions. We have implemented two approaches to meet this goal. One approach is to map eQTL using statistically proven QTL models adapted to fit gene expressions as response variables. The extent of LD, however, limits the differentiation between cis and trans action, specifically for eQTL mapping to the same chromosome as the associated gene position. Due to this limitation we define cis-action as ‘local’ and trans-action as ‘distant’ when referring to our eQTL analysis. Co-localization of identified eQTL with known pQTL for the same population identify not only local regulators of gene expression, but also distant factors influencing transcriptional variation. The second approach is to identify cis-acting regulators of gene expression through allele-specific expression analyses (ASE) using RNA-seq data<sup>38–40</sup>. Different functional categories are involved in transcriptional regulation including enhancers, silencers, insulators, and promoters, among other architectural elements<sup>15,41</sup>. A cis-acting variant could be located within any one of these

functional elements affecting transcription factor binding sites, mRNA stability or microRNA binding sites<sup>42</sup>. For instance, a regulatory sequence in the DNA containing a SNP may affect the affinity of trans-acting regulators causing allele-specific expression because it only affects the allele containing the variant. Distant acting variants indirectly affect transcription by altering a gene that regulates the expression of a target gene such as transcription factors or microRNA, and therefore affects the expression of both copies of the target gene in a diploid organism. In order to detect allelic imbalance for cis-acting variants we must quantify the allele-specific expression of polymorphic locus by studying heterozygous samples<sup>38,40,43–45</sup>.

Accurately quantifying ASE from RNA-seq data is challenging because such data is prone to technical artifacts including genotyping error and mapping bias, which lead to inaccurate estimates of ASE and inflated false positive determinations of allelic imbalance<sup>46–49</sup>. To address this issue, we have implemented a robust unbiased allele-specific read mapping protocol<sup>46</sup> to control for technical bias when estimating allelic imbalance. Only a few studies of allelic imbalance have been performed in livestock species<sup>34,39,40,50–52</sup>. The ASE analyses reported to date for pigs have been limited to small sample sizes (only 4 animals in Wu *et al.*<sup>40</sup>, 12 in Oczkiewicz *et al.*<sup>52</sup>, and 38 in Maroilley *et al.*<sup>34</sup>). However, these studies have increased our understanding of cis-regulatory elements influencing immune capacity<sup>34</sup>, prenatal skeletal muscle growth<sup>51</sup> and the adult brain transcriptome<sup>52</sup> in pigs. The ASE analysis reported in this dissertation utilized transcriptomic data from l. dorsi muscle for 168 F<sub>2</sub> MSUPRP animals. This represents a considerably larger sample size than any previous reports of ASE in pigs, allowing detection of a higher number of heterozygous coding SNP with low read coverage, and providing novel cis-acting variants regulating mRNA transcript abundance. In addition, we assessed the effect of cis-acting variants on trait phenotypes, since SNP called directly from transcriptomic

data provide increased marker coverage of coding regions. We also applied pyrosequencing to verify selected polymorphic locus that exhibited significant allelic imbalance and that explained a portion of phenotypic variance.

This dissertation research has two important implications: (1) The discovery of genomic regions directly influencing expression of single genes (local and distant acting variants), and multiple genes (regulatory hotspots) to reveal the functional significance of pQTL within the swine genome. (2) The localization of cis-acting regulators of gene expression that account for a significant portion of phenotypic variation providing insights into potential architectural elements regulating economically important traits in pigs.

The aims of this dissertation research include:

1. Identify potential candidate genes and molecular markers regulating phenotypic traits using an F2 Duroc x Pietrain pig resource population.
  - a. Map eQTL for the MSUPRP using RNA-seq of *l. dorsi* to identify local and distant regulators of transcript abundance.
  - b. Identify eQTL co-localizing with pQTL and estimate peak pQTL SNP effect on eQTL significance using a conditional analysis.
2. Perform an ASE analysis to confirm cis acting variants found with the previous eQTL analysis and identify novel polymorphic sites with allelic imbalance.
  - a. Estimate the effect of ASE cSNP on growth, body composition and meat quality trait phenotypes.
  - b. Confirm select ASE markers associated with phenotypic traits using pyrosequencing.

**Table 1.1** Review of eQTL studies conducted in pig populations.

| Phenotype                                 | Tissue <sup>a</sup> | Animals | Breed <sup>b</sup> | Platform      | eQTL  | Year               |
|-------------------------------------------|---------------------|---------|--------------------|---------------|-------|--------------------|
| Water holding capacity                    | l. dorsi            | 74      | D x P              | Microarray    | 897   | 2008 <sup>18</sup> |
| Meat Quality                              | l. dorsi            | 74      | D x P              | Microarray    | 9,180 | 2010 <sup>19</sup> |
| Cellular stress                           | l. lumbrun          | 57      | Multiple breeds    | Microarray    | 272   | 2011 <sup>17</sup> |
| Obesity                                   | liver               | 150     | P x (LW x L)       | Microarray    | 4,727 | 2011 <sup>33</sup> |
| Meat quality and carcass merit            | l. dorsi            | 176     | D x P              | Microarray    | 62    | 2011 <sup>21</sup> |
| Lipid metabolism                          | g. medius           | 105     | D                  | Microarray    | 613   | 2012 <sup>16</sup> |
| Sense and antisense transcript expression | liver               | 497     | D x E              | DGE           | 370   | 2012 <sup>53</sup> |
|                                           | l. dorsi            | 589     |                    |               | 399   |                    |
| Plasma cortisol level                     | l. dorsi            | 207     | P x (LW x L)       | Microarray    | 593   | 2012 <sup>37</sup> |
| Serum lipids                              | liver               | 497     | D x E              | DGE           | 643   | 2013 <sup>36</sup> |
| Drip loss                                 | l. dorsi            | 132     | D x P              | Microarray    | 30    | 2013 <sup>25</sup> |
| Fatty acid composition                    | l. dorsi            | 102     | I x L              | Microarray    | 13    | 2013 <sup>26</sup> |
| Glycolytic potential                      | l. dorsi            | 497     | D x E              | DGE           | 7     | 2014 <sup>27</sup> |
| Meat quality                              | l. dorsi            | 207     | P x (LW x L)       | Microarray    | 7     | 2014 <sup>28</sup> |
| Response to Actinobacillus infection      | lung                | 100     | H x L              | Microarray    | 193   | 2014 <sup>20</sup> |
| Obesity                                   | adipose             | 36      | Da x G             | RNAseq        | 1,060 | 2015 <sup>32</sup> |
| Fat deposition and muscularity            | l. dorsi            | 176     | D x P              | Microarray    | 7     | 2015 <sup>24</sup> |
| Meat quality                              | l. dorsi            | 114     | I x L              | Dynamic Array | 19    | 2016 <sup>29</sup> |
| Meat quality                              | g. medius           | 104     | D                  | Microarray    | 3     | 2017 <sup>30</sup> |
| PRRSV infection                           | blood               | 44      | LW x L<br>D x L/Y  | RNAseq        | 869   | 2017 <sup>35</sup> |
| Fatness and Yield                         | l. dorsi            | 102     | I x L              | Microarray    | 63    | 2017 <sup>31</sup> |
| Immune capacity                           | blood               | 243     | LW                 | Microarray    | 1,901 | 2017 <sup>34</sup> |

<sup>a</sup> *longuissimus dorsi* (l. dorsi), *longuissimus lumbrun* (l. lumbrun), *glutes medius* (g. medius) <sup>b</sup>Duroc (D), P (Pietrain), Large White (LW), Landrace (L), Erhualian (E), Iberian (I), Hampshire (H), Danish (Da), Göttingen (G), Yorkshire (Y) <sup>c</sup>Digital gene expression (DGE)

## REFERENCES

## REFERENCES

1. Steen, H. A. M. Van Der, Prall, G. F. W. & Plastow, G. S. Application of genomics to the pork industry. *J. Anim. Sci.* **83**, E1–E8 (2005).
2. Edwards, D. B. *et al.* Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: I. Growth traits. *J. Anim. Sci.* **86**, 241–253 (2008).
3. Edwards, D. B. *et al.* Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. *J. Anim. Sci.* **86**, 254–66 (2008).
4. Choi, I. *et al.* Application of alternative models to identify QTL for growth traits in an F2 Duroc x Pietrain pig resource population. *BMC Genet.* **11**, 97 (2010).
5. Choi, I. *et al.* Identification of carcass and meat quality QTL in an F2 Duroc × Pietrain pig resource population using different least-squares analysis models. *Front. Genet.* **2**, 18 (2011).
6. Bernal Rubio, Y. L. *et al.* Implementing meta-analysis from genome-wide association studies for pork quality traits 1. *J. Anim. Sci.* **93**, 5607–5617 (2015).
7. Bernal Rubio, Y. L. *et al.* Meta-analysis of genome-wide association from genomic prediction models. *Anim. Genet.* **47**, 36–48 (2016).
8. Gualdrón Duarte, J. L. *et al.* Refining genomewide association for growth and fat deposition traits in an F 2 pig population. *J. Anim. Sci.* **94**, 1387–97 (2016).
9. Casiró, S. *et al.* Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. *J. Anim. Sci.* **95**, 554–558 (2017).
10. Edwards, D. B., Bates, R. O. & Osburn, W. N. Evaluation of Duroc- vs . Pietrain-sired pigs for carcass and meat quality measures. *J. Anim. Sci.* **81**, 1895–1899 (2003).
11. Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C. & Steibel, J. P. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* **13**, 24 (2012).
12. Uimari, P. & Tapio, M. Extent of linkage disequilibrium and effective population size in finnish landrace and finnish yorkshire pig breeds. *J. Anim. Sci.* **89**, 609–614 (2011).
13. Zhang, C. & Plastow, G. Genomic diversity in pig (*Sus scrofa*) and its comparison with human and other livestock. *Curr. Genomics* **12**, 138–146 (2011).
14. Dekkers, J. C. M. & Hospital, F. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* **3**, 22–32 (2002).

15. Spitz, F. Gene regulation at a distance : From remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
16. Cánovas, A. *et al.* Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population. *PLoS One* **7**, e35583 (2012).
17. Liaubet, L. *et al.* Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics* **12**, 548 (2011).
18. Ponsuksili, S. *et al.* Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* **9**, 367 (2008).
19. Ponsuksili, S., Murani, E., Schwerin, M., Schellander, K. & Wimmers, K. Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. *BMC Genomics* **11**, 572 (2010).
20. Reiner, G. *et al.* Pathway deregulation and expression QTLs in response to *Actinobacillus pleuropneumoniae* infection in swine. *Mamm. Genome* **25**, 600–617 (2014).
21. Steibel, J. P. *et al.* Genome-Wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. *PLoS One* **6**, e16766 (2011).
22. Kratz, A. & Carninci, P. The devil in the details of RNA-seq. *Nat. Biotechnol.* **32**, 882–884 (2014).
23. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**, e78644 (2014).
24. Peñagaricano, F. *et al.* Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. *BMC Syst. Biol.* **9**, 58 (2015).
25. Heidt, H. *et al.* A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. *Mamm. Genome* **24**, 416–426 (2013).
26. Muñoz, M. *et al.* Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. *BMC Genomics* **14**, 845 (2013).
27. Ma, J. *et al.* A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Biol.* **10**, e1004710 (2014).

28. Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M. & Wimmers, K. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *Int. J. Biol. Sci.* **10**, 327–337 (2014).
29. Puig-Oliveras, A. *et al.* Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci. Rep.* **6**, 31803 (2016).
30. González-Prendes, R. *et al.* Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Sci. Rep.* **7**, 39830 (2017).
31. Martínez-Montes, A. M. *et al.* Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics. *Mamm. Genome* **28**, 130–142 (2017).
32. Kogelman, L. J. A., Zhernakova, D. V, Westra, H., Cirera, S. & Fredholm, M. An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. *Genome Med.* **7**, 105 (2015).
33. Ponsuksili, S., Murani, E., Brand, B., Schwerin, M. & Wimmers, K. Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J. Lipid Res.* **52**, 668–678 (2011).
34. Maroille, T. *et al.* Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* **18**, 967 (2017).
35. Kommadath, A. *et al.* Genetic architecture of gene expression underlying variation in host response to porcine reproductive and respiratory syndrome virus infection. *Sci. Rep.* **7**, 46203 (2017).
36. Chen, C. *et al.* Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics* **14**, 848 (2013).
37. Ponsuksili, S., Du, Y., Murani, E., Schwerin, M. & Wimmers, K. Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. *Genetics* **192**, 1109–1122 (2012).
38. MacEachern, S., Muir, W. M., Crosby, S. D. & Cheng, H. H. Genome-wide identification and quantification of cis - and trans -regulated genes responding to Marek ' s disease virus infection via analysis of allele-specific expression. *Front Genet* **2**, 113 (2012).

39. Cheng, H. H. *et al.* Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek ' s disease is predominantly determined by transcriptional regulation. *BMC Genomics* **16**, 816 (2016).
40. Wu, H., Gaur, U., Mekchay, S., Peng, X. & Li, L. Genome-wide identification of allele-specific expression in response to *Streptococcus suis* 2 infection in two differentially susceptible pig breeds. *Anim. Genet.* **56**, 481–491 (2015).
41. Mora, A., Sandve, G. K. & Gabrielsen, O. S. In the loop : promoter – enhancer interactions and bioinformatics. *Brief. Bioinform.* **17**, 980–995 (2015).
42. Wang, X. & Clark, A. G. Using next-generation RNA sequencing to identify imprinted genes. *Heredity (Edinb)*. **113**, 156–166 (2014).
43. Ernst, C. W. & Steibel, J. P. Molecular advances in QTL discovery and application in pig breeding. *Trends Genet.* **29**, 215–224 (2013).
44. MacEachern, S. *et al.* Genome-wide identification of allele-specific expression ( ASE ) in response to Marek ' s disease virus infection using next generation sequencing. *BMC Proc.* **5**, S14 (2011).
45. Maceachern, S., Muir, W. M., Crosby, S. D. & Cheng, H. H. Genome-Wide Identification and Quantification of cis- and trans-Regulated Genes Responding to Marek's Disease Virus Infection via Analysis of Allele-Specific Expression. *Front. Genet.* **2**, 113 (2011).
46. Geijn, B. Van De, Mcvicker, G., Gilad, Y. & Pritchard, J. K. WASP : allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
47. Degner, J. F. *et al.* Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. **25**, 3207–3212 (2009).
48. Satya, R. V., Zavaljevski, N. & Reifman, J. A new strategy to reduce allelic bias in RNA-Seq readmapping. *Nucleic Acids Res.* **40**, e12 (2012).
49. Castel, S. E., Levy-moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195 (2015).
50. Perumbakkam, S., Muir, W. M., Black-pyrkosz, A., Okimoto, R. & Cheng, H. H. Comparison and contrast of genes and biological pathways responding to Marek ' s disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. *BMC Genomics* **14**, 64 (2013).

51. Yang, Y. *et al.* Transcriptome analysis revealed chimeric RNAs, single nucleotide polymorphisms and allele-specific expression in porcine prenatal skeletal muscle. *Sci. Rep.* **6**, 29039 (2016).
52. Oczkowicz, M., Szmatoła, T., Piórkowska, K. & Ropka-Molik, K. Variant calling from RNA-seq data of the brain transcriptome of pigs and its application for allele-specific expression and imprinting analysis. *Gene* **641**, 367–375 (2018).
53. Chen, C. *et al.* A genome-wide investigation of expression characteristics of natural antisense transcripts in liver and muscle samples of pigs. *PLoS One* **7**, e52433 (2012).

## CHAPTER TWO

### Genetic control of *longissimus dorsi* muscle gene expression variation and joint association with phenotypic quantitative trait locus in pigs

Vélez-Irizarry, D., S. Casiro, R.O. Bates, N.E. Raney, J.P. Steibel and C.W. Ernst

#### ABSTRACT

Economically important growth and meat quality traits in pigs are controlled by cascading molecular events occurring during development and continuing throughout the conversion of muscle to meat. Evaluating transcriptomic profiles of skeletal muscle during the initial steps leading to the conversion of muscle to meat can identify key regulators of polygenic phenotypes. In this study, we aim to identify potential candidate genes and molecular markers regulating phenotypic traits using an F2 Duroc x Pietrain pig resource population. Gene transcripts obtained with RNA-seq of *longissimus dorsi* muscle from 168 F2 animals were used to estimate gene expression variation subject to genetic control by mapping expression QTL (eQTL). A total of 334 eQTL were mapped ( $FDR \leq 0.01$ ) with 188 exhibiting local acting regulation. Joint association of eQTL with phenotypic QTL (pQTL) segregating in our population revealed 16 genes significantly associated with 21 pQTL for meat quality, carcass composition and growth traits. Ten of these pQTL were for meat quality phenotypes that co-localized with one eQTL on SSC2 (8.8Mb region) and 11 on SSC15 (121Mb region). Biological processes identified for co-localized eQTL genes include calcium signaling (*FERM*, *MRLN*, *PKP2* and *CHRNA9*), energy metabolism (*SUCLG2* and *PFKFB3*) and redox hemostasis (*NQO1* and *CEP128*), and results support an important role for activation of the PI3K-Akt-mTOR signaling pathway during the initial conversion of muscle to meat.

**Keywords:** expression QTL, skeletal muscle, RNA-seq, pig

## INTRODUCTION

Applications of genomic improvement techniques have significantly advanced livestock breeding. Genomic regions harboring single nucleotide polymorphisms (SNP) accounting for a significant portion of phenotypic variation for economically important traits have been identified and implemented in marker assisted selection<sup>1-3</sup>. In pigs, these efforts have identified candidate genes affecting meat quality (e.g. *CRC1*, *PRKAG3*, *CAST*), weight gain (e.g. *MC4R*) and litter size (e.g. *ESR*)<sup>4</sup>. However, we still do not fully understand the molecular mechanisms underlying the variability observed in pork traits. For meat quality traits, cascading molecular events starting before exsanguination and continuing throughout the conversion of muscle to meat play a critical role in determining the eating quality of pork. By studying the transcriptomic profile of the initial steps leading to the conversion of muscle to meat we can elucidate key regulators of polygenetic trait phenotypes. Specifically, we can identify gene transcripts subject to genetic control that potentially regulate complex traits by mapping expression QTL (eQTL), and testing their co-localization with phenotypic QTL (pQTL). In this study, we use an F2 Duroc x Pietrain resource population developed at Michigan State University<sup>5,6</sup> (the MSUPRP) to identify eQTL significantly associated with pQTL for meat quality, carcass composition and growth traits.

Meat quality traits are highly correlated. During the conversion of muscle to meat,  $\text{Ca}^{2+}$  ions are released from the sarcoplasmic reticulum and the anaerobic production of ATP leads to the accumulation of lactic acid that reduces muscle pH<sup>7</sup>. The rate of pH decline and release of  $\text{Ca}^{2+}$  directly influences water holding capacity, meat color and the rate of proteolytic activity that leads to meat tenderization<sup>7</sup>. While these molecular processes have been extensively studied with numerous QTL identified for tenderness, drip loss, pH, meat color and enzyme activity<sup>8</sup>, we know little of the genetic architecture regulating these traits. This is likely due to the high

variability of meat quality traits that are known to be heavily influenced by both genetic and environmental factors such as antemortem handling<sup>9–11</sup>. Regulators of gene expression have been used to study the molecular bases of polygenetic phenotypic differences in swine populations<sup>12–15</sup>. Expression QTL maps provide a foundation to study divergent molecular processes in livestock species<sup>2,16</sup>. This approach has been successful in identifying candidate genes, causative variants and molecular networks regulating phenotypic traits in swine, including back fat<sup>17</sup>, drip loss<sup>18</sup>, glycolytic potential<sup>15</sup>, plasma cortisol levels<sup>12</sup> and lipid metabolism<sup>19</sup>.

In this study we use a GBLUP-based GWA model to map eQTL. With this model, we can elucidate both local and distant acting regulators of gene expression, and narrow sense heritability ( $h^2$ ) can be estimated for each gene. Joint analysis of pQTL and eQTL can identify potential genetic regulators of phenotypic traits and give insights into the genetic architecture of complex traits. Putative hotspots are of particular interest where a single marker is associated with the expression of multiple genes, serving as a potential master regulator that can account for a significant portion of phenotypic variation. In this study, we aim to map eQTL for *longissimus dorsi* muscle of the well characterized MSUPRP to identify local and distant regulators of transcript abundance. A joint-association of eQTL with pQTL may reveal novel insights into the genetic architecture of meat quality, carcass composition and growth traits.

## **MATERIALS AND METHODS**

### *Pig population and phenotype collection*

Animal housing and care protocols were evaluated and approved by the Michigan State University All University Committee on Animal Use and Care (AUF # 09/03-114-00). The MSUPRP was developed from 4 Duroc boars and 15 Pietrain sows<sup>5,6</sup>. From the F1 progeny, 56 animals (6 males and 50 females) were retained to produce the F2 generation, which included

1,259 animals from 142 litters. A total of 67 phenotypic traits were collected for the F2 generation<sup>5,6</sup>. A subset of the F2 pigs were selected for this study using a selective profiling scheme based on extremes in loin muscle area and backfat thickness phenotypes within litter (44 litters) and sex<sup>20</sup>. Summary statistics for the 67 phenotypic traits (29 growth traits, 20 carcass composition traits and 18 meat quality traits) in the F2 population, and the subset of animals used for this study are shown in Supplementary Table 2.S1.

### *Genotyping*

SNP genotypes for the MSUPRP were available from prior studies<sup>21,22</sup>. Genotyping was performed by Neogen Corporation - GeneSeek Operations (Lincoln NE) using the Illumina PorcineSNP60 BeadChip<sup>23</sup> for the F0, F1 and ~1/3 of the F2 population and the GeneSeek Genomic Profiler for Porcine Low Density (GGP-Porcine LD) for the remaining F2 pigs<sup>21,22</sup>. Missing genotypes were imputed with an accuracy of 0.97<sup>21,22</sup>. Monomorphic markers and non-autosomal markers were eliminated from further analysis, as were those showing divergence from Mendelian inheritance rules. An updated genomic map for SNPs on the Sscrofa11.1 genome assembly was obtained from Neogen (Lincoln NE). Additional filtering was performed to exclude markers with a minor allele frequency lower than 0.01 and reduce the degree of correlation between adjacent markers (i.e. if a pair of neighboring markers had a correlation of allelic dosage greater than 0.95, one of the two markers was eliminated; this filtering was performed only for the eQTL analysis). Filtering resulted in 38,679 markers for the eQTL analysis and 43,130 for the pQTL analysis. Two coding SNPs in the protein kinase AMP-activated non-catalytic subunit gamma 3 (*PRKAG3*) gene, I199V and T30N<sup>24,25</sup>, were also genotyped in the MSUPRP as previously described in Casiro *et al.*<sup>26</sup>.

### *RNA extraction and RNA sequencing*

Tissue samples were taken immediately post mortem from the *longuissimus dorsi* muscle, flash frozen in liquid nitrogen and stored at -80°C until processing. RNA extraction was performed with the miRNeasy Mini Kit (Qiagen, Germantown, MD) following the manufacturer's protocol. Quality and quantity of extracted total RNA were determined using the Agilent 2100 Bioanalyzer (RIN  $\geq$  7). Sequencing was performed at the Michigan State University Research Technology Support Facility. Libraries for 24 samples were prepared using the Illumina TruSeq RNA Library Prep Kit v2, and sequenced on the Illumina HiSeq 2000 platform (2 x 100bp paired-end reads). The remaining 152 libraries were prepared using the Illumina TrueSeq Stranded mRNA Kit, and sequenced on the Illumina HiSeq 2500 platform (2 x 125bp, paired-end reads). Base calling was performed with the Illumina Real Time Analysis v1.18.61 software, and the Illumina Bc12fastq v1.8.4 was used for conversion to FastQ format. A total of 96 sequence files (741Gb) consisting of ~63 million short-reads per library were obtained from the HiSeq 2000 platform and 1,218 sequence files (~2Tb) of ~23 million short-reads per library were obtained from the HiSeq 2500 platform. Eight samples were removed from further analysis due to low sequence quality, leaving a total of 168 samples for subsequent analyses. Sequence data has been deposited in the NCBI Sequence Read Archive accession number PRJNA403969.

Raw RNA sequence reads were first filtered for adapter sequences using Trimmomatic<sup>27</sup> followed by quality trimming using Condetri where the first 6 bases at the 3' end and low quality reads were filtered out retaining reads with a minimum length of 75 bases. The quality of each sequenced nucleotide was evaluated on adapter filtered and quality trimmed RNA-seq reads using the FASTX toolkit<sup>28</sup> and a mean Phred quality score of  $37.01 \pm 0.99$  was obtained. After

adaptor and quality filtering, RNA-seq reads were mapped to the reference genome assembly *Sus scrofa* 11.1 using the splice aware aligner Tophat2<sup>29</sup>. Sample-specific transcriptomes were assembled using Cufflinks and merged with the reference genome to create a set of known and novel isoforms using Cuffmerge<sup>30</sup>. A total of 30,723 full length transfrags were identified. Alignment statistics and base coverage were obtained with SAMtools<sup>31</sup>. Samples showed on average 92.4% of sequencing reads mapping to the reference genome and 73.3% were unique and properly paired with their complementary sequence. Total gene expression abundance was quantified using unique and properly paired reads using HTseq<sup>32</sup>. Genes with total count abundance less than 168 were removed from further analysis to reduce the number of genes with low expression, leaving 16,121 gene transcripts for eQTL analysis.

#### *RNA-seq count normalization and transformation*

Expressed gene counts were normalized using the trimmed mean of M-values (TMM) to reduce systematic technical biases of sequenced transcripts<sup>33</sup>. TMM normalization has been shown to control false positive associations<sup>34</sup>. The normalized gene counts were then transformed to follow an approximately Gaussian distribution by calculating the log counts per million (log-cpm) as described in Law *et. al.*<sup>35</sup>. Briefly, a linear model was fit to obtain the expected log-cpm for each gene,  $E(y) = x\beta$ , where  $y$  are the log-cpm,  $x$  is a vector of ones and  $\beta$  is a vector of estimated regression coefficients. The residual standard deviations for each gene and their calculated average log-cpm were used to estimate the mean variance trend,  $\hat{w}$ , by fitting a LOWESS curve<sup>35</sup>. Variance coefficients were standardized to keep similar scales for residual variance and additive variance:

$$\widehat{\mathbf{w}}_{std} = \frac{\frac{1}{\sqrt{\widehat{\mathbf{w}}}}}{\frac{1}{n} \sum \frac{1}{\sqrt{\widehat{\mathbf{w}}}}} \quad (1)$$

where,  $\widehat{\mathbf{w}}_{std}$  are the variance coefficients,  $n$  the total number of animals, and  $\widehat{\mathbf{w}}$ , the estimated mean variance trend. The normalized log-cpm were used as the response variable,  $\mathbf{y}$ , and the variance coefficients,  $\widehat{\mathbf{w}}_{std}$ , to model heterogeneity of error variance in the eQTL scan. This approach accounts for the mean variance relationship of each gene expression instead of assuming equal variance for all observations.

#### *Heritability of phenotype and gene expression*

A genomic best-linear unbiased prediction (GBLUP) model<sup>21,22</sup> was used to estimate the heritability of each phenotype and gene expression by fitting the following equation:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{a} + \mathbf{e}, \quad (2)$$

where,  $\mathbf{y}$  is a vector with measurements of a phenotype for each animal when estimating phenotypic heritability, and a vector with normalized log-cpm gene expression when estimating the heritability of gene expression.  $\mathbf{X}$  is an incidence matrix of fixed effects including sex and additional covariates unique to each phenotype<sup>26,36</sup>, and includes the transcriptional profiling selection scheme (i.e. within litter and sex extreme for loin muscle area or back fat thickness) when analyzing gene expression. The vector  $\mathbf{b}$  contains the estimated fixed effect,  $\mathbf{a}$  is a vector of random additive genetic effects and  $\mathbf{e}$  is a vector of random residual errors. The additive genetic effects are assumed  $\mathbf{a} \sim N(0, G\sigma_a^2)$  with the genomic relationship matrix<sup>37</sup>,  $G = \mathbf{Z}\mathbf{Z}'$ .  $\mathbf{Z}$  is a matrix of normalized SNP genotypes, with elements:

$$Z = \frac{M-2p}{\sqrt{\Sigma(2p(1-p))}}, \quad (3)$$

where,  $\mathbf{M}$  is the matrix of SNP genotypes and  $\mathbf{p}$  is a vector with the frequency of each reference allele. The error term is  $\mathbf{e} \sim N(0, \sigma_e^2 \text{diag}(\widehat{\mathbf{w}_{std}}))$  with a variance inversely proportional to the variance coefficients,  $\widehat{\mathbf{w}_{std}}$ . These variance coefficients account for the heteroskedasticity across genes with different expression. The heritability of gene expressions were calculated by taking the ratio of the variance of the additive genetic effects to the total phenotypic variance,  $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ .

Statistical significance of heritability was determined using a likelihood ratio test,  $LR = 2[\log L(\hat{\theta}) - \log L(\widehat{\theta}_0)]$ , comparing the likelihood of the model represented in Eq. 1 ( $L(\hat{\theta})$ ) and the likelihood of a null model that does not include the genetic additive effect ( $L(\widehat{\theta}_0)$ ). Testing the null hypothesis  $\sigma_a^2 = 0$  is equivalent to testing  $h^2 = 0$ . The likelihood ratios were compared to a chi-squared distribution with one degree of freedom and the resulting p-value divided by 2 to account for the asymptotic distribution of the likelihood ratios that tend to follow a mixture of chi-square distributions with different degrees of freedom<sup>38</sup>. Multiple test corrections were performed using a FDR of 0.01<sup>39</sup>. Differences in heritability between local and distant eQTL were determined with Wilcoxon rank sum test<sup>40</sup>.

### *Genome wide association*

The SNP effects,  $\hat{\mathbf{g}}$ , and their variances  $Var(\hat{\mathbf{g}})$  were estimated as a linear transformation of the BLUP breeding values,  $\hat{\mathbf{a}}$ , from Eq. 2<sup>41,42</sup>. A test statistic for the association of each marker with each phenotype or gene expression measure is computed by standardizing the SNP effects:

$$\mathbf{T} = \frac{\hat{\mathbf{g}}}{\sqrt{Var(\hat{\mathbf{g}})}}, \quad (4)$$

The p-values associated with this  $\mathbf{T}$  test statistic were calculated using the Gaussian cumulative distribution function,  $\Phi$ , as follows:

$$p - value = 2[1 - \Phi(|\mathbf{T}|)], \quad (5)$$

and subject to multiple test corrections per each gene expression ( $FDR \leq 0.01$ )<sup>39</sup>.

It has been demonstrated<sup>41,42</sup> that the  $\mathbf{T}$  test statistics and p-values resulting from Eq. 4 and 5 are equivalent to those obtained from fitting a single marker model, specifically the Efficient Mixed-Model Association (EMMA) model<sup>43</sup>.

#### *Local and distant regulators*

Due to low SNP density and long-range linkage disequilibrium in this pig population, distinguishing local versus distant regulation of gene expression is difficult. We applied the following algorithm to classify putative eQTL as local or distant regulators of a gene's expression:

- 1) An eQTL was defined as any gene with at least one marker association surpassing the significance threshold ( $FDR \leq 0.01$ ).
- 2) The plausible position range of each eQTL was defined by the position of the first significant marker at the beginning of the QTL and last significant marker at the end of the QTL. If the eQTL had only one marker association the position of the marker was used.
- 3) Given the mapped position of the gene profile (start and end position of the transcript) there are several possibilities
  - a. The associated eQTL plausible position range overlaps totally or partially: Local eQTL

- b. The associated eQTL is on a different chromosome: Distant eQTL
- c. The associated eQTL is on the same chromosome but does not overlap:
  - i. There are non-significant SNP ( $FDR \geq 0.01$ ). between the mapped position of the gene profile and its associated eQTL range: Distant eQTL
  - ii. There are no SNP between gene and eQTL range (including the filtered SNP due to high LD): Plausible Local

### *Co-localization analysis*

The genomic positions of the mapped eQTL were co-localized with pQTL previously identified for the F2 MSUPRP for growth, carcass composition and meat quality traits. An eQTL was considered co-localized if its QTL position overlapped the mapped position of a pQTL. The statistical significance of each co-localized eQTL with pQTL was determined through a conditional analysis that tested the effect of the most significant marker associated with the pQTL on the co-localized eQTL gene expression, as follows:

$$\mathbf{y} = X\mathbf{b} + \mathbf{Z}_{SNP}b_{SNP} + \mathbf{a} + \mathbf{e}, \quad (6)$$

where,  $\mathbf{y}$  is the expression of the co-localized eQTL gene. The  $X$ ,  $\mathbf{b}$ ,  $\mathbf{a}$  and  $\mathbf{e}$  were previously described in Eq. 2.  $\mathbf{Z}_{SNP}$  is a vector of standardized marker genotypes for the pQTL peak marker, co-localized with eQTL gene, and  $b_{SNP}$  the estimated marker effect. Type I error rate of 0.05 and Bonferroni p-value cutoff based on the number of tests performed ( $p\text{-value} \leq 5.952e-04$ ) was used to determine SNP effect significance. We also considered the effect the peak pQTL marker had on the eQTL peak by performing a linear transformation of the BLUP breeding values from Eq. 6 to estimate the individual SNP effects and tested their significance as described in Eq. 4 and 5. Multiple test corrections were performed using an  $FDR \leq 0.01^{39}$ . If

fitting the top pQTL marker completely eliminated the eQTL peak, the two QTL were considered to be significantly co-localized. The proportion of variance explained by the peak pQTL markers for each co-localized eQTL was estimated as described in Casiro *et al.*<sup>26</sup>. Briefly, the variance associated with the co-localized peak pQTL marker,  $\sigma_{SNP}^2$ , was estimated as:

$$\widehat{\sigma_{SNP}^2} = b^2 \text{var}(Z_{SNP}), \quad (7)$$

where,  $b^2$  is the calculated peak pQTL marker effect from Eq. 6, and the proportion of gene expression variance accounted for by the co-localized pQTL peak SNP is

$\widehat{\sigma_{SNP}^2} / (\widehat{\sigma_{SNP}^2} + \widehat{\sigma_a^2} + \widehat{\sigma_e^2})$ . The estimated additive genetic variance,  $\sigma_a^2$ , and error variance,  $\sigma_e^2$ , is obtained after fitting equation 6. Equations 6 and 7 were also used to estimate the proportion of gene expression variance explained by the *PRKAG3* T30N SNP for all identified eQTL to uncover eQTL significantly associated with *PRKAG3* and the proportion of phenotypic variance explained for meat quality phenotypes with an associated pQTL on SSC15.

#### RT-qPCR

To verify the expression of *CHRNA9*, 28 animals were selected based on the genotypes of the peak eQTL SNP (10 animals per genotype equally weighted by sex except for the AA genotype that had only 8 animals, 4 per sex). Total RNA was extracted from the longissimus muscle samples as described above, and 2µg was reverse transcribed using the High Capacity cDNA Reverse Transcriptase Kit with RNase inhibitor (Applied Biosystems, Foster City, CA). A custom Taqman Gene Expression Assay (ThermoFisher Scientific, Waltham, MA) was designed for *CHRNA9* using pig RNA sequence to span exons 4 and 5 (determined based on the structure of the human *CHRNA9* gene, Accession No. AC118275). The GeNorm<sup>44</sup> algorithm was used to select two reference genes, *PPIA* (ThermoFisher Scientific Assay No. Ss03394781\_g1) and

*SDHA* (ThermoFisher Scientific Assay No. Ss03376909\_u1), with the highest gene-stabilizing measure to normalize the expression of *CHRNA9*. RT-qPCR was performed in triplicate using 50 ng cDNA and TaqMan Gene Expression Master Mix for a final volume of 20  $\mu$ l. Assays were run on a StepOnePlus Real-Time PCR System (Applied Biosystems). The cycling conditions were 52°C for 2 min, 95°C for 10 min followed by 50 cycles of 95°C for 15 s and 60°C for 1 s.  $\Delta$ Ct values were calculated as the mean difference between the geometric mean of the reference genes and the target genes. To verify the RNA-seq results, the effect of the peak eQTL marker for *CHRNA9* was measured using Eq. 6 with the response variable being the  $\Delta$ Ct transcript abundance. Analysis of variance with a type I error rate of 0.05 was used to determine significant additive and dominance effects of the peak *CHRNA9* SNP.

## RESULTS

### *Identification of eQTL*

A genome wide association study (GWAS) was conducted using 23,162 SNP markers and 15,223 transcript abundance profiles for 168 F2 pigs. The GWAS identified 334 eQTL (3,094 significant gene marker associations; whole genome FDR  $\leq$  0.01 per gene, p-value  $\leq$   $2.04\text{e-}04 \pm 3.86\text{e-}04$ ) for 321 gene transcripts and 2,523 molecular markers (Supplementary Table 2.S2). The number of SNP associated with variation in transcript abundance was on average  $9.26 \pm 15.14$ , and the size of each eQTL peak was on average  $12.04 \pm 22.90$  Mb (Table 2.1).

All autosomes had associated eQTL, with SSC9 containing the most associations (42 eQTL). Two chromosomes contained a putative hotspot; SSC9 (ASGA0044684; SSC9:125.0 Mb) and SSC15 (H3GA0052416; SSC15:121.8 Mb). A putative hotspot is defined as a single marker associated with multiple gene expressions, and we considered a single marker associated

with more than ten genes to be a putative hotspot. ASGA0044684 was associated with 25 transcripts, and H3GA0052416 with 11 transcripts ( $\text{FDR} \leq 0.01$ ). Both putative hotspots mapped to non-coding regions, an intron variant of the ral guanine nucleotide dissociation stimulator like 1 (*RGL1*) gene on SSC9, and an intergenic variant on SSC15.

**Table 2.1.** eQTL summary among regulator types.

| Gene Regulator                                                            | N <sup>1</sup> | Min <sup>2</sup> | Max <sup>3</sup> | Mean <sup>4</sup> | SD <sup>5</sup> |
|---------------------------------------------------------------------------|----------------|------------------|------------------|-------------------|-----------------|
| <i>Average length of eQTL plausible position range<sup>a</sup></i>        |                |                  |                  |                   |                 |
| All regulators                                                            | 334            | 0                | 175.20           | 12.04             | 22.90           |
| Local                                                                     | 166            | 0                | 175.20           | 22.51             | 28.19           |
| Plausible Local                                                           | 22             | 0                | 11.44            | 1.43              | 2.82            |
| Distant Same Chromosome                                                   | 59             | 0                | 25.55            | 2.10              | 5.40            |
| Distant                                                                   | 87             | 0                | 69.76            | 1.47              | 7.92            |
| <i>Average distance from eQTL to gene transcript position<sup>a</sup></i> |                |                  |                  |                   |                 |
| All regulators                                                            | 334            | 1.75e-3          | 104.80           | 3.64              | 12.23           |
| Local                                                                     | 166            | 1.75e-3          | 25.12            | 1.92              | 3.88            |
| Plausible Local                                                           | 22             | 5.70e-3          | 1.52             | 0.24              | 0.41            |
| Distant Same Chromosome                                                   | 59             | 8.23e-3          | 104.80           | 9.78              | 23.25           |
| Distant                                                                   | 87             | -                | -                | -                 | -               |
| <i>Number of SNP associations</i>                                         |                |                  |                  |                   |                 |
| All regulators                                                            | 334            | 1                | 105              | 9.26              | 15.14           |
| Local                                                                     | 166            | 1                | 105              | 16.77             | 18.60           |
| Plausible Local                                                           | 22             | 1                | 14               | 2.95              | 3.12            |
| Distant Same Chromosome                                                   | 59             | 1                | 17               | 2.01              | 2.37            |
| Distant                                                                   | 87             | 1                | 5                | 1.46              | 0.97            |
| <i>Heritability</i>                                                       |                |                  |                  |                   |                 |
| All regulators                                                            | 334            | 5.47e-10         | 0.97             | 0.32              | 0.23            |
| Local                                                                     | 166            | 5.47e-10         | 0.97             | 0.419             | 0.22            |
| Plausible Local                                                           | 22             | 0.04             | 0.63             | 0.32              | 0.17            |
| Distant Same Chromosome                                                   | 59             | 1.19e-09         | 0.74             | 0.27              | 0.22            |
| Distant                                                                   | 87             | 1.34e-09         | 0.76             | 0.17              | 0.17            |

<sup>a</sup>Values shown in mega bases. <sup>1</sup>Number of eQTL. <sup>2</sup>Minimum value. <sup>3</sup>Maximum value.

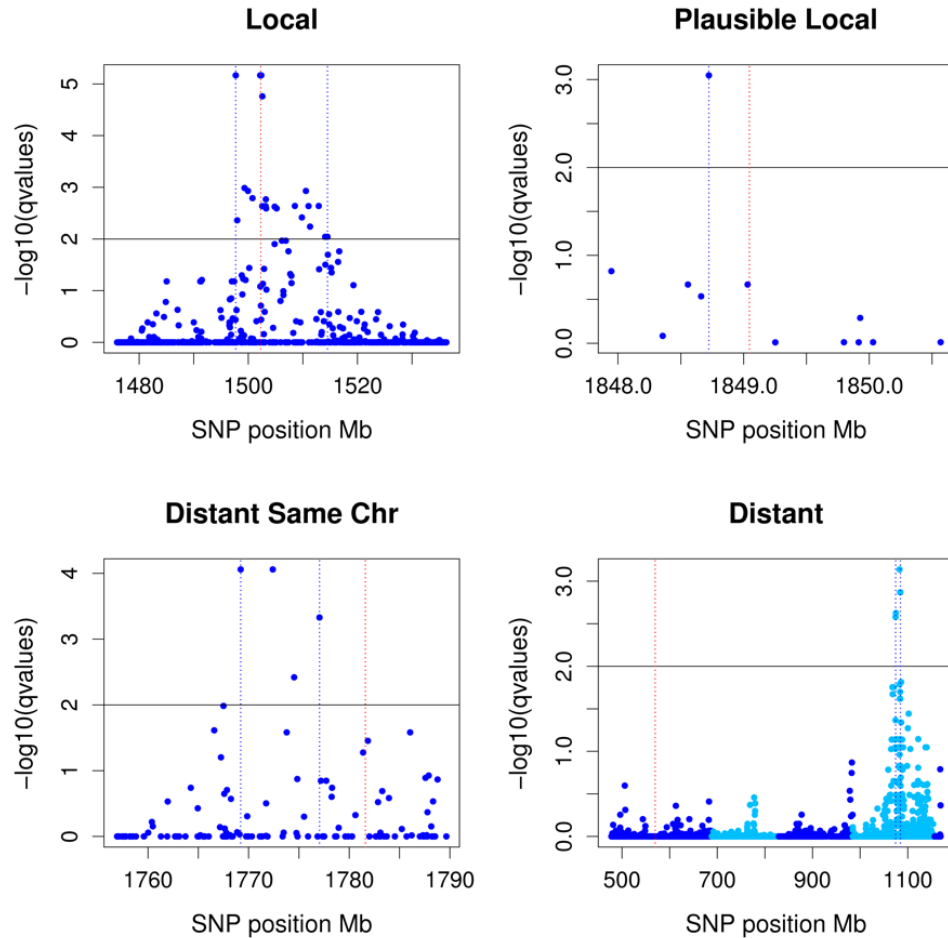
<sup>4</sup>Average value. <sup>5</sup>Standard deviation of value.

### *Local versus distant regulators of gene expression*

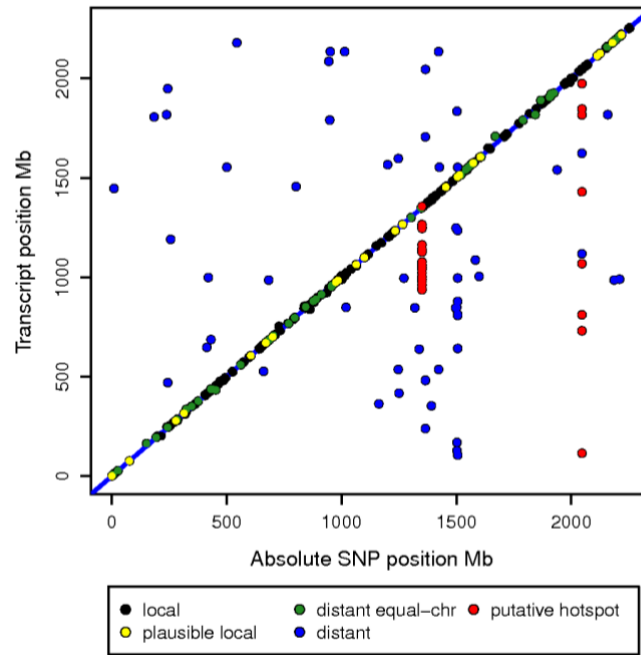
For each of the eQTL peaks, a plausible position range delimited by the first and last significant marker ( $\text{FDR} \leq 0.01$ ) was identified and compared to the mapped position of the associated gene transcript to distinguish between local and distant regulators of gene expression (Figures 2.1 and 2.2). A classification of local acting regulator of gene expression was determined if the position of the associated gene transcript overlapped the eQTL plausible position range (Figure 2.1). We identified 166 local regulators of gene expression (Figure 2.2, black associations)

The average distance from the mid gene position and peak eQTL SNP for local regulators was  $1.92 \pm 3.88$  Mb, however, due to the large plausible position range for some local eQTL (up to 175 Mb) the maximum distance for a local regulator was 25 Mb (Table 2.1). If the gene mapped to the same chromosome but fell outside the range of its associated eQTL with markers below the significance threshold between the gene and eQTL positions, the eQTL was considered to be a distant regulator on the same chromosome as the associated gene (Figure 2.1). A total of 59 distant regulators on the same chromosome as the associated gene were identified (Figure 2.2, green associations) with their eQTL peak at an average distance of  $9.78 \pm 23.25$  Mb from the associated gene position (Table 2.1). However, in situations where the area between the eQTL range and the associated gene transcript was found to be devoid of markers, the eQTL was considered to be a plausible local regulator (Figure 2.1). Under this classification, 22 plausibly local regulators of gene expression were identified (Figure 2.2, yellow associations) with their eQTL peak at an average distance of  $0.24 \pm 0.41$  Mb from the associated gene position (Table 2.1). An eQTL mapped to a different chromosome than its associated gene transcript was classified as a distant regulator (Figure 2.1). We observed 87 distant regulators of gene

expression (Figure 2.2, blue associations). A non-parametric test showed local eQTL had significantly higher numbers of associated SNP than distant eQTL ( $p\text{-value} \leq 2.20\text{e-}16$ ).



**Figure 2.1** *Manhattan plots illustrating the classification of different types of gene expression regulation based on eQTL position.* The x-axis represents the absolute genomic position of the marker and the y-axis the significance of the association with the gene transcript,  $-\log_{10}$  q-value. The two blue vertical dotted lines delimit the eQTL plausible position range (eQTL-PPR), and the vertical red dotted line indicates the absolute position of the gene transcript. Local-acting regulator: the position of the gene transcript falls within or overlaps the eQTL-PPR. Plausible local regulator: the eQTL-PPR does not contain or overlap the gene transcript and the density of SNP in the region separating the two is zero. Distant-acting regulator on the same chromosome: the position of the gene transcript falls outside the specified eQTL-PPR but on the same chromosome and the SNPs between the genomic position of the gene and the eQTL-PPR do not surpassing the significance threshold. Distant-acting regulator: the eQTL-PPR is on a different chromosome than the genomic position of the associated gene transcript.

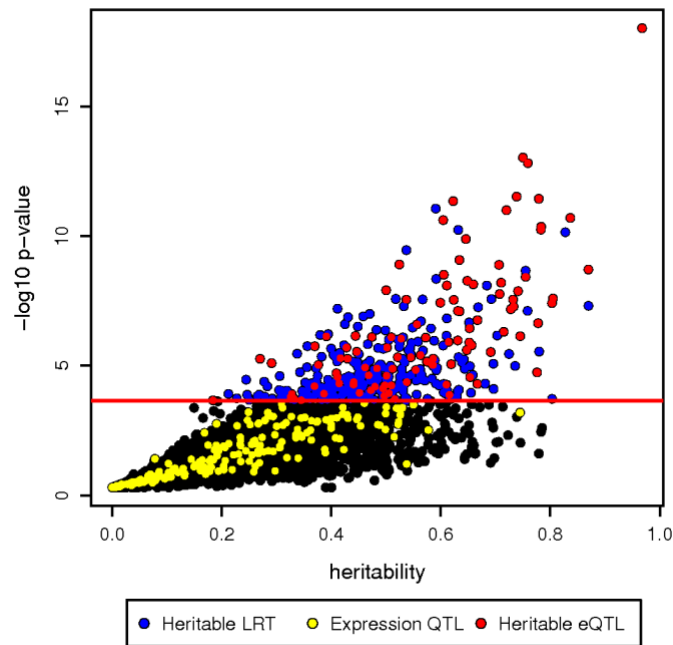


**Figure 2.2 eQTL map.** The y-axis represents the absolute genomic position of the gene and the x-axis represents the absolute genomic position of its associated SNP marker. Associations aligning on the diagonal are eQTL found on the same chromosome as the gene. A plausible position range was identified for each eQTL peak based on the peak's flanking markers, and local regulation determined when the gene position overlapped this range, shown in black. Plausible local regulators of gene expression (described in Figure 2.1) are shown in yellow. The eQTL peaks shown in green are distant regulators that map to the same chromosome as their associated gene. Distant regulators mapping to a different chromosome than the associated gene are shown in blue. The eQTL shown in red are potential putative hotspots on SSC9 and SSC15.

### *Heritability of gene expression*

Heritability ( $h^2$ ) was estimated for all gene transcripts with 344 exhibiting significantly heritable expression ( $FDR \leq 0.01$ ,  $p\text{-value} \leq 2.27e-04$ ). The mean  $h^2$  for significantly heritable transcripts was  $0.51 \pm 0.13$ , whereas the mean  $h^2$  for other transcripts was  $0.09 \pm 0.12$  (Table 2.2). The relationship between the estimated  $h^2$  of gene expression and its significance is shown in Figure 2.3. A significant enrichment of genes associated with an eQTL was observed for the significantly heritable gene transcripts ( $p\text{-value} \leq 2.2e-16$ ; shown in red, Figure 2.3). The  $h^2$  of genes with an associated eQTL that were not significantly heritable was on average  $0.21 \pm 0.16$

(shown in yellow, Figure 2.3 and summarized in Table 2.2), whereas the group of significantly heritable genes associated with an eQTL had a mean  $h^2$  of  $0.57 \pm 0.15$  (Table 2.2). Mean heritability among the different regulator types was higher in the group of eQTL associated with local acting regulation,  $0.42 \pm 0.22$ , and lowest in eQTL with distant acting regulation,  $0.17 \pm 0.17$  (Table 2.1). Non-parametric test showed a significant difference between local and distant heritabilities ( $p\text{-value} \leq 1.08e-14$ ).



**Figure 2.3** *Heritability of transcript profiles.* Heritability of genes is shown on the x-axis and p-values from the likelihood ratio test (LRT) for significant heritable expression are on the y-axis. A total of 344 gene expression transcripts were found to be heritable (shown in blue and red,  $FDR \leq 0.01$ ). A significant enrichment of genes with associated eQTL was observed among the heritable genes (103 genes;  $p\text{-value} \leq 2.2e-16$ ; shown in red). The 218 genes associated with an eQTL that did not surpass the threshold for significant heritability are shown in yellow.

**Table 2.2.** Heritability summary for all genes and genes with an associated eQTL.

| Significant h <sup>2</sup> | N      | Heritability (h <sup>2</sup> ) |       |       |       |  |
|----------------------------|--------|--------------------------------|-------|-------|-------|--|
|                            |        | Min                            | Max   | Mean  | SD    |  |
| <i>All Genes</i>           |        |                                |       |       |       |  |
| Yes <sup>1</sup>           | 344    | 0.184                          | 0.968 | 0.508 | 0.133 |  |
| No                         | 14,879 | 2.210e-19                      | 0.785 | 0.091 | 0.123 |  |
| <i>eQTL Genes</i>          |        |                                |       |       |       |  |
| Yes <sup>1</sup>           | 103    | 0.184                          | 0.968 | 0.574 | 0.147 |  |
| No                         | 218    | 5.475e-10                      | 0.745 | 0.206 | 0.165 |  |

<sup>1</sup>FDR  $\leq$  0.01*Phenotypic QTL*

Genomic regions significantly associated with growth<sup>36</sup>, meat quality and carcass composition<sup>26</sup> traits have been previously identified in our MSUPRP. However, these analyses used an earlier assembly of the pig genome (Sscrofa10.2), therefore, we reanalyzed the 67 phenotypic traits for the F2 population (960 animals) following previous methods<sup>21,22</sup> to generate an updated QTL map using the most current genome assembly (Sscrofa 11.1). Our QTL analysis of 29 growth traits identified 14 pQTL (Supplementary Table 2.S3, FDR  $\leq$  0.05, p-value  $\leq$  2.50e-04) for which seven were confirmed from Duarte *et al.*<sup>36</sup> and five exhibited a different peak SNP, in part because one of the SNP on SSC6 (ALGA0122657) did not have a genomic position in the new genome build. We were unable to confirm two pQTL on SSC2 for 10<sup>th</sup> rib backfat at 16-weeks and last rib backfat at 19-weeks, and one pQTL on SSC3 for birth weight that were reported in Duarte *et al.*<sup>36</sup>. However, we identified two new pQTL for loin muscle area at 16-weeks on SSC6 and last rib backfat at 10-weeks on SSC12. Our QTL analysis for carcass composition and meat quality traits identified 29 pQTL (Supplementary Table 2.S3, FDR  $\leq$

0.05). Fourteen pQTL were confirmed from Casiro *et al.*<sup>26</sup> and eight exhibited a different peak SNP, in part because three SNP (SSC6: ALGA0122657, SSC11: M1GA0015491 and SSC15: MARC0047188) did not have genomic positions in the new genome build. Seven new pQTL were identified for cook yield (SSC5 and SSC8), last lumbar backfat (SSC4, SSC9 and SSC10), dressing percent (SSC11) and loin weight (SSC11; Supplementary Table 2.S3, FDR  $\leq$  0.05). In total, 43 pQTL were mapped using the Sscrofa11.1 genome assembly, including six QTL for 10<sup>th</sup> rib backfat from 13 to 22 weeks of age, seven QTL for last rib backfat from 13 to 22 weeks of age, one QTL for loin muscle area at 16 weeks of age, 13 QTL for carcass composition traits and 16 QTL for meat quality traits.

#### *Co-localization of phenotypic QTL with expression QTL*

The association of eQTL co-localized with pQTL was performed through conditional analysis of transcript abundance, which fixed the peak pQTL SNP, to elucidate eQTL significantly associated with phenotypic traits. Manhattan plots of eQTL co-localized with pQTL are shown in Figure 2.4 for meat quality and carcass composition traits, and Figure 2.5 for growth traits. The conditional analysis tested 53 eQTL (orange associations) co-localized with 34 pQTL (blue associations) for ten growth and 11 meat quality and carcass composition traits (Figures 2.4 and 2.5, Table 2.3 and Supplementary Table 2.S4). A total of 16 eQTL were significantly associated with 21 pQTL, where conditioning upon the peak pQTL marker resulted in the complete removal of eQTL significance (p-value  $\leq$  5.95e-04 for SNP effect and FDR  $\leq$  0.01 for eQTL significance; black associations in Figures 2.4 and 2.5; Table 2.4 and Supplementary Table 2.S4). Three pQTL regions common among correlated phenotypes co-localized with eQTL, resulting in eQTL significantly associated with variation for multiple phenotypes.

**Table 2.3** Phenotypic QTL co-localized with expression QTL.

| Phenotype <sup>1</sup> | SSC | Peak SNP <sup>2</sup> | E <sup>3</sup> | VS <sup>4</sup> | h <sup>2</sup> | N <sup>5</sup> |
|------------------------|-----|-----------------------|----------------|-----------------|----------------|----------------|
| 10th-Rib BF            | 1   | ALGA0010839           | +              | 0.03            | 0.45           | 1              |
| WBS                    | 2   | M1GA0002229           | -              | 0.04            | 0.26           | 1*             |
| SP Tenderness          | 2   | H3GA0005676           | -              | 0.05            | 0.28-0.29      | 1*             |
| Last Lumbar BF         | 4   | ASGA0092651           | -              | 0.04            | 0.41           | 4              |
| Last-Rib BF 16-wk      | 5   | ALGA0031990           | +              | 0.03            | 0.47           | 1              |
| Cook Yield             | 5   | MARC0036560           | +              | 0.03            | 0.31           | 1              |
| Loin Muscle Area 16-wk | 6   | ASGA0105067           | +              | 0.04            | 0.29           | 4*             |
| Growth and Carcass BF  | 6   | ALGA0104402           | -              | 0.04-0.07       | 0.35-0.57      | 6*             |
| 10th-Rib BF            | 6   | M1GA0008917           | -              | 0.12            | 0.45           | 6*             |
| Loin Weight, Growth BF | 6   | ASGA0029651           | -/+            | 0.06            | 0.30-0.41      | 4*             |
| Number of Ribs         | 7   | ALGA0043983           | +              | 0.12            | 0.36           | 10             |
| Cook Yield             | 8   | DRGA0008986           | -              | 0.03            | 0.31           | 1              |
| Dressing %             | 11  | M1GA0014839           | +              | 0.03            | 0.24           | 2              |
| Loin Weight            | 11  | ALGA0060368           | -              | 0.03            | 0.30           | 2*             |
| Last-Rib BF10-wk       | 12  | ASGA0054658           | -              | 0.02            | 0.35           | 2              |
| Meat Quality, Protein  | 15  | MARC0093624           | -/+            | 0.06-0.21       | 0.19-0.38      | 22*            |
| Meat Quality           | 15  | H3GA0052416           | +              | 0.04-0.07       | 0.07-0.29      | 16*            |

<sup>1</sup>Phenotypes associated with pQTL. BF is backfat. SP Tenderness includes sensory panel tenderness and overall tenderness. Growth BF includes ultrasound last-rib backfat at 10, 13 and 22 weeks and Carcass BF includes carcass 10<sup>th</sup>-rib and last-rib backfat. Meat Quality includes the phenotypes for sensory panel juiciness, tenderness and overall tenderness, Warner Bratzler Shear Force, Cook Yield, Drip Loss and 24-hour pH. Protein is protein percent.

<sup>2</sup>Peak pQTL SNP (FDR  $\leq$  0.05) <sup>3</sup>Effect of B allele for peak pQTL SNP on phenotype, positive increases phenotypic trait. <sup>4</sup>Proportion of phenotypic variance explained by peak SNP. <sup>5</sup>Number of eQTL co-localized with the pQTL;

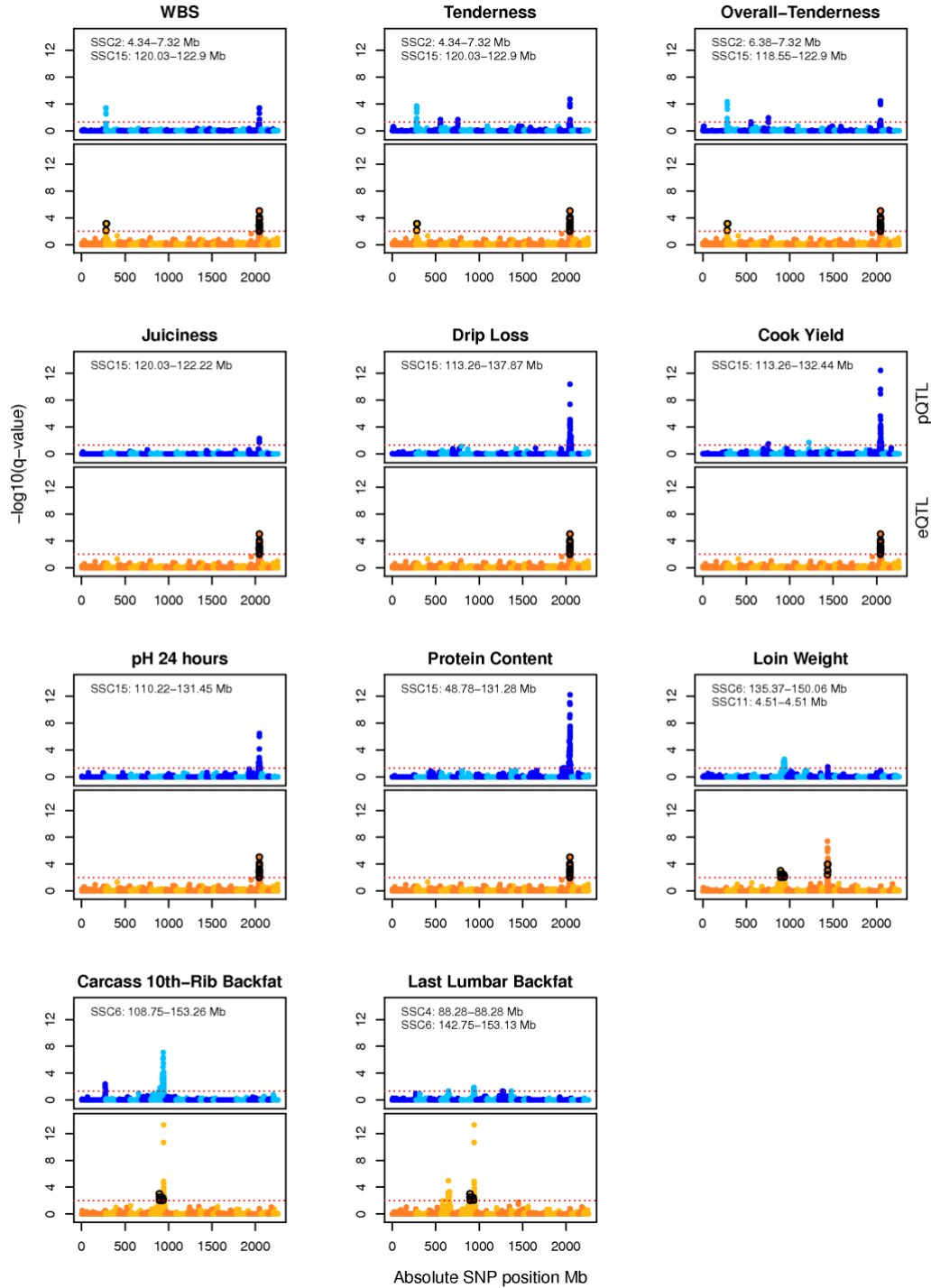
\*Contains at least one eQTL significantly associated with the phenotype.

**Table 2.4** Expression QTL significantly associated with phenotypic traits.

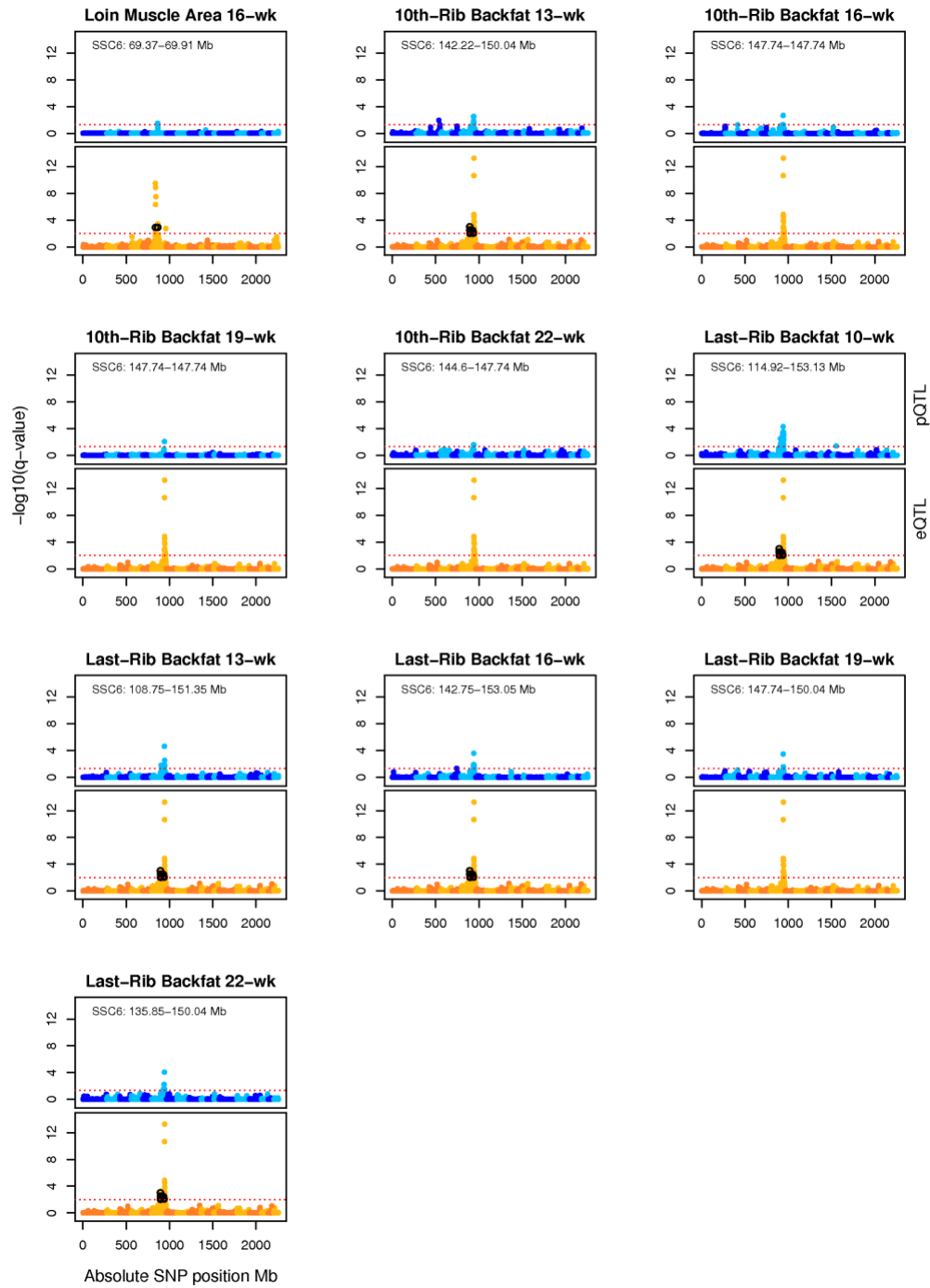
| Gene                     | SSC Gene | SSC eQTL | Regulator <sup>2</sup> | E <sup>3</sup> | Phenotype <sup>4</sup>             |
|--------------------------|----------|----------|------------------------|----------------|------------------------------------|
| TEX9                     | 1        | 15       | Distant                | +              | Meat Quality, Protein              |
| FRMD8                    | 2        | 2        | Local                  | –              | Tenderness                         |
| PKP2                     | 5        | 15       | Distant                | –              | Meat Quality, Protein              |
| NQO1                     | 6        | 15       | Distant                | +              | Meat Quality, Protein              |
| HPN                      | 6        | 6        | Local                  | +              | Loin Muscle Area 16-wk             |
| SSC6:104.08 <sup>1</sup> | 6        | 6        | Local                  | –              | Carcass and Growth BF, Loin Weight |
| SSX2IP                   | 6        | 6        | Local                  | +              | Carcass BF, Loin Weight            |
| CEP128                   | 7        | 15       | Distant                | +              | Meat Quality, Protein Percent      |
| CHRNA9                   | 8        | 15       | Distant                | -              | Meat Quality, Protein              |
| PFKFB3                   | 10       | 15       | Distant                | –              | Meat Quality, Protein              |
| SSC11:2.19 <sup>1</sup>  | 11       | 11       | Local                  | –              | Loin Weight                        |
| SUCLG2                   | 13       | 15       | Distant                | –              | Meat Quality, Protein              |
| CIT                      | 14       | 15       | Distant                | +              | Meat Quality, Protein              |
| CCDC60                   | 14       | 15       | Distant                | +              | Meat Quality, Protein              |
| MRLN                     | 14       | 15       | Distant                | –              | Meat Quality, Protein              |
| SSC15:48.94 <sup>1</sup> | 15       | 15       | Distant Same SSC       | +              | Meat Quality, Protein              |

<sup>1</sup>Novel gene transcripts: *Sus Scrofa* chromosome and start position. <sup>2</sup>Regulator type for eQTL, local is for eQTL containing mapped position of gene transcript, distant is for eQTL on a different SSC than associated transcript position and distant same SSC are eQTL on the same SSC as associated eQTL but not contained (Figure 2.1).

<sup>3</sup>Effect of B allele for peak eQTL marker on the gene's expression: positive increases and negative decreases the gene expression. <sup>4</sup>Phenotypes significantly associated with the gene's expression. Tenderness includes WBS and sensory panel tenderness and overall tenderness. Meat Quality includes the phenotypes for sensory panel juiciness, tenderness and overall tenderness, Warner Bratzler Shear Force, Cook Yield, Drip Loss and 24-hour pH. Protein is protein percent. Growth BF includes ultrasound last-rib backfat at 10, 13 and 22 weeks and Carcass BF includes carcass 10<sup>th</sup>-rib and last-rib backfat.



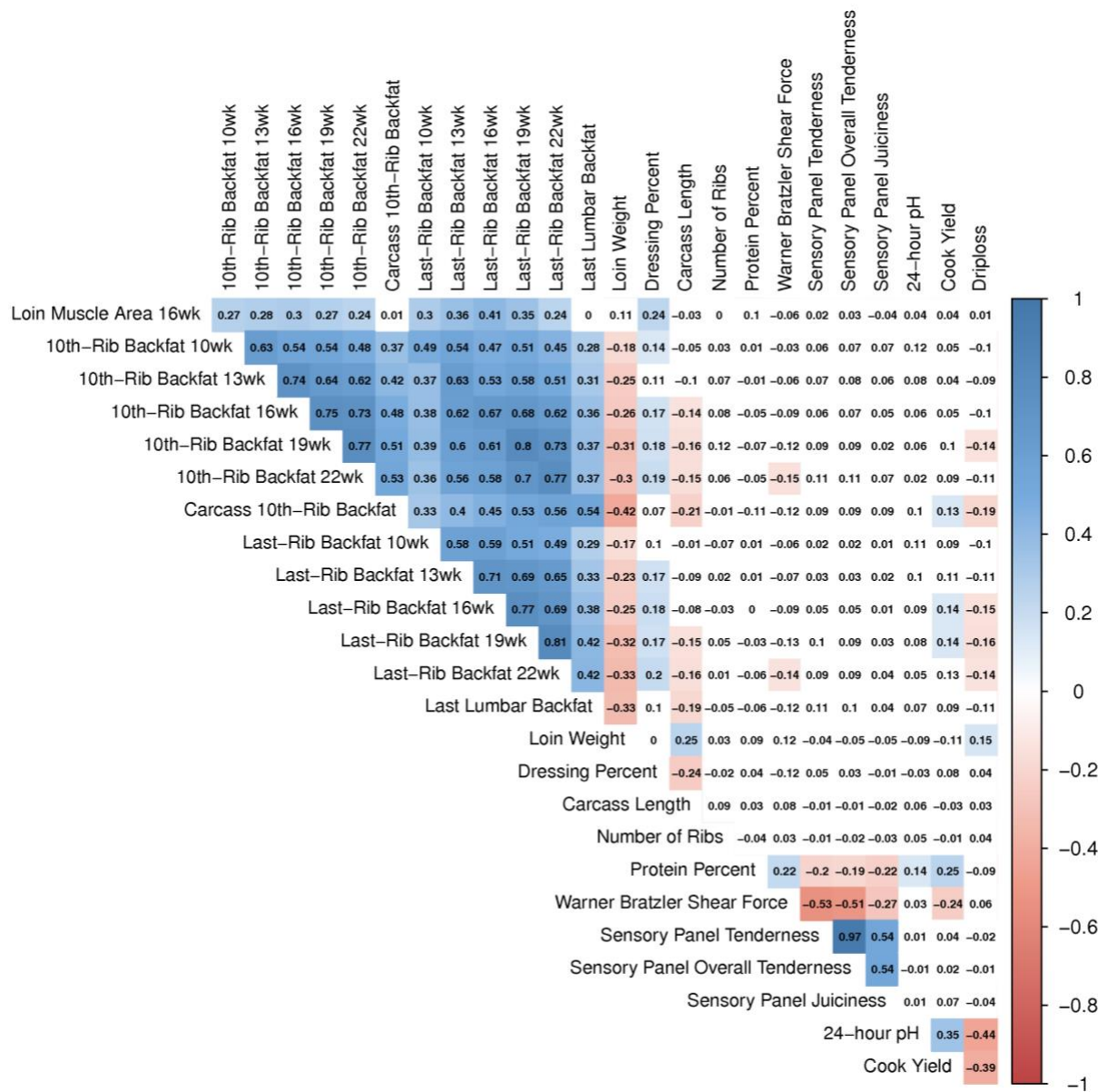
**Figure 2.4** Manhattan plots of meat quality and carcass composition pQTL co-localized with eQTL. The x-axis is the absolute genome position in mega bases. The y-axis is the negative base 10 logarithm of q-values, with the red line representing the significance threshold. Manhattan plots in shades of blue are for the pQTL ( $FDR \leq 0.05$ ) and those in shades of orange are for the eQTL ( $FDR \leq 0.01$ ). SNP associated with an eQTL co-localizing with a pQTL, and whose association is no longer significant after performing the conditional analysis are shown in black.



**Fig 2.5** Manhattan plots of growth pQTL co-localized with eQTL. The x-axis is the absolute genome position in mega bases. The y-axis is the negative base 10 logarithm of q-values, with the red line representing the significance threshold. Manhattan plots in shades of blue are for the pQTL ( $FDR \leq 0.05$ ) and those in shades of orange are for the eQTL ( $FDR \leq 0.01$ ). SNP associated with an eQTL co-localizing with a pQTL, and whose association is no longer significant after performing the conditional analysis are shown in black.

Meat quality traits exhibited phenotypic correlations as expected for these traits. WBS was negatively correlated with sensory panel scores (i.e. juiciness, tenderness and overall-tenderness) and cook yield, and positively correlated with protein percent (p-value  $\leq 8e-05$ , Figure 2.6). Cook yield was negatively correlated with drip loss and positively correlated with 24-hour pH and protein percent (p-value  $\leq 8e-05$ , Figure 2.6). Phenotypes related to tenderness were associated with QTL on SSC2, and all eight of the aforementioned correlated meat quality phenotypes were associated with QTL mapped to SSC15 (Figure 2.4). A similar trend was observed for growth and carcass composition traits related to fat deposition and muscle weight where serial ultrasound measures for 10<sup>th</sup> and last rib backfat were positively correlated with carcass 10<sup>th</sup>-rib and last lumbar backfat, and negatively correlated with loin weight (p-value  $\leq 8e-05$ , Figure 2.6), and these traits were associated with QTL on SSC6 (Figures 2.4 and 2.5).

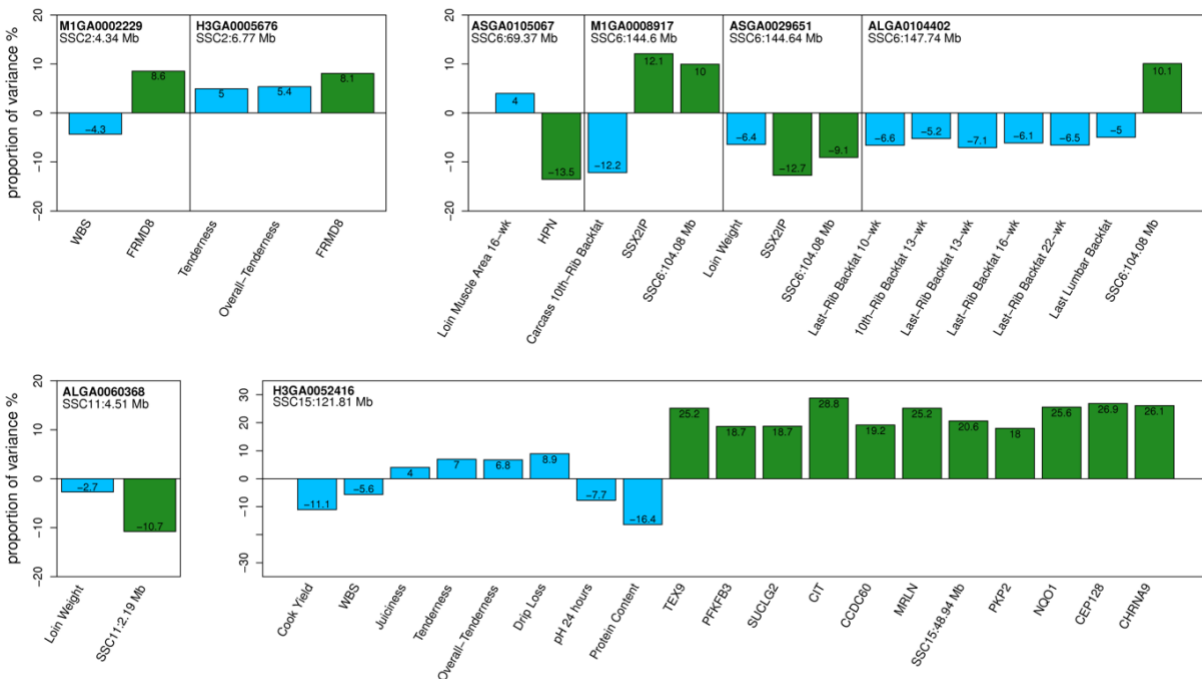
Phenotypic QTL for growth and carcass composition traits associated with eQTL on SSC6 revealed two genomic regions. A 28.82 Mb region (SSC6:43.819-72.625 Mb) was associated with the hepsin gene (*HSN*) and loin muscle area at 16 weeks. A 53.33 Mb region (SSC6:99.932-153.261 Mb) was associated with a novel transcript (SSC6:104.08) and serial ultrasound measures of last rib backfat (at 10, 13, 16 and 22 weeks of age), 10<sup>th</sup> rib backfat at 13 weeks of age, and carcass last lumbar backfat. The peak pQTL marker for loin muscle area at 16 weeks of age, ASGA0105067, accounted for 4% of the phenotypic variance and 13.5% of the gene expression variance with increased loin muscle area associated with decreased expression of the *HPN* gene (Figure 2.7). The pQTL marker for backfat deposition, ALGA0104402, accounted for 5-7.1% of the phenotypic variance, and 10.1% of the gene expression variance, with increased expression of the novel transcript SSC6:104.08 associated with reduced backfat deposition (Figure 2.7).



**Figure 2.6.** Pearson correlations among phenotypic traits with an associated pQTL. Significant correlations are shaded in color, p-value  $\leq 8e-05$ , with shades of red depicting negative correlations and shades of blue depicting positive correlations.

Two additional pQTL for carcass composition phenotypes (carcass 10<sup>th</sup> rib backfat and loin weight) also mapped to the 53.33 Mb region on SSC6 and were significantly associated with SSC6:104.08 and SSX2IP. The peak pQTL marker for carcass 10<sup>th</sup> rib backfat (M1GA0008917)

accounted for 12.2% of the phenotypic variance with increased expression of SSC6:104.08 and *SSX2IP* associated with reduced 10<sup>th</sup> rib backfat. For loin weight, the peak pQTL marker (ASGA0029651) was associated with reduced loin weight and reduced expression of SSC6:104.08 and *SSX2IP*, accounting for 6.4% of the phenotypic variance and up to 12.7% of the transcript expression variance (Figure 2.7). A second pQTL for loin weight was mapped on SSC11 and was significantly associated with a novel transcript (SSC11:2.19), which coincides with the uncharacterized locus LOC110255792. The peak pQTL marker for loin weight on SSC11 (ALGA0060368) accounted for 2.7% of the phenotypic variance and 10.7% of the gene expression variance. Reduced loin weight was associated with reduced expression of the SSC11:2.19 transcript (Figure 2.7).



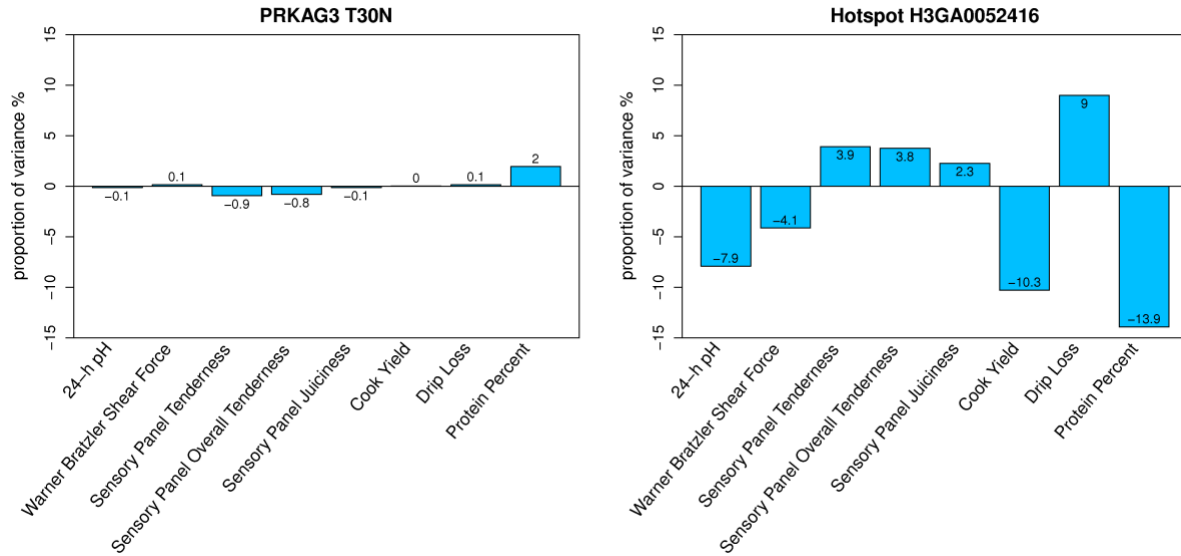
**Figure 2.7.** Proportion of variance explained by peak pQTL SNP for phenotypes (blue) and gene transcript abundance (green). Traits are shown on the x-axis, and the proportion of phenotypic variance explained by the SNP marker is shown on the y-axis. Directionality of bar plots indicates SNP effect on phenotype or gene expression (i.e., increase or decrease).

Considering the pQTL for meat quality and carcass composition traits with their associated eQTL reveals two genomic regions of particular note. A 7.90 Mb region on SSC2:4.341-12.242 Mb was associated with the FERM domain-containing 8 gene (*FRMD8*) and WBS, sensory panel tenderness and overall tenderness phenotypes, and a 110.21 Mb region on SSC15:27.666-137.874 Mb was associated with 11 genes and eight meat quality or carcass composition phenotypes (Tables 2.3 and 2.4). Significant negative correlations were observed between WBS and all three sensory panel phenotypes as expected for these traits ( $r = -0.44 \pm 0.14$ ,  $p\text{-value} \leq 8e-05$ , Figure 2.7); more force needed to break myofibers (i.e., higher shear force values) was correlated with lower meat tenderness based on subjective scores evaluated by a trained sensory panel. The peak pQTL markers, M1GA0002229 and H3GA0005676, for meat quality traits on SSC2 accounted for approximately five percent of the phenotypic variance and eight percent of *FRMD8* gene expression variance (Figure 2.7) with increased expression of *FRMD8* associated with increased sensory panel tenderness and overall tenderness scores and decreased WBS. High LD was observed between the two SNP ( $r = 0.64$ ).

Eleven of the eQTL significantly associated with phenotypes were distant regulators of gene expression, and all of these were also associated with the putative hotspot within the 110.21 Mb region on SSC15. The SSC15 putative hotspot marker H3GA0052416 was the peak pQTL marker for sensory panel juiciness, tenderness and overall tenderness (Tables 2.3), as well as the peak eQTL marker for seven gene transcripts (Tables 2.4). The peak pQTL marker for WBS, 24-hour pH, cook yield, drip loss and protein percent on SSC15 (MARC0093624) is in high LD with the putative hotspot marker (Pearson correlation 0.89). These results suggest a potential candidate variant(s) on SSC15 accounting for a significant portion of phenotypic variation for meat quality and carcass composition phenotypes, as well as individual gene expression

variation. Since these two markers are in high LD, the proportion of phenotypic and gene expression variance was estimated for the putative hotspot marker for all eight phenotypes and eleven gene transcripts including *CCDC60*, *CEP128*, *CHRNA9*, *CIT*, *MRLN*, *NQO1*, *PFKFB3*, *PKP2*, *SUCLG2*, *TEX9*, and a novel transcript SSC15:48.94 Mb (mapped to the uncharacterized locus LOC110257028). The H3GA0052416 marker accounted for 4-16% of the phenotypic variance and approximately 23% of the gene expression variance (Figure 2.7). Increased expression of the eleven genes associated with the B allele of the putative hotspot was also associated with an increase in sensory panel scores and drip loss, and a decrease in WBS, 24-hour pH, cook yield and protein percent (Figure 2.7).

The gene protein kinase AMP-activated non-catalytic subunit gamma 3 (*PRKAG3*) maps to this region of SSC15, and variants of *PRKAG3* have been implicated as affecting meat quality phenotypes<sup>24,25</sup>. We genotyped all F2 animals for two *PRKAG3* coding SNP<sup>26</sup> and included these SNP in our GWAS, however, the eQTL scan did not reveal associations with either of the *PRKAG3* markers. To further assess the effect of *PRKAG3*, we performed a conditional analysis to estimate the significance of these markers on identified eQTL (Supplementary Table 2.S5). One gene, *NQO1*, was significantly associated with the *PRKAG3* T30N SNP ( $FDR \leq 0.01$ ), where T30N accounted for up to 12% of the gene expression variance. Given the high signal of the putative hotspot on SSC15 for various genes and meat quality traits, we estimated the proportion of phenotypic variance explained by both the putative hotspot and the *PRKAG3* T30N marker for meat quality and carcass composition traits (Figure 2.8). The *PRKAG3* T30N marker accounted for 0.1-2% of phenotypic variance for meat quality traits, whereas the putative hotspot marker accounted for 2-14%. This analysis shows the putative hotspot accounts for a greater proportion of phenotypic variance than the *PRKAG3* T30N SNP.



**Figure 2.8** Proportion of phenotypic variance explained by *PRKAG3* T30N SNP and putative hotspot SNP H3GA0052416 for meat quality traits mapped to SSC15. Traits are shown on the x-axis, and the proportion of phenotypic variance explained by the SNP marker is shown on the y-axis. Directionality of bar plots indicates the SNP effect on the phenotype.

#### RT-qPCR confirmation of *CHRNA9*

The GBLUP-based GWA model identified 24 eQTL mapped to a 125 Mb region on SSC15. Eleven of these eQTL co-localized with pQTL for meat quality and carcass composition traits, and among these the *CHRNA9* gene was selected for verification using RT-qPCR (Figure 2.7). *CHRNA9*, is implicated in catecholamine secretion and the adaptive response to chronic stress<sup>45</sup>, and is essential for muscle contraction<sup>46</sup>. The genomic position of the *CHRNA9* gene is on SSC8: 31.44-31.51Mb, and the eQTL associated with this gene mapped to SSC15, therefore exhibiting distant acting regulation of *CHRNA9* gene expression. RT-qPCR was performed to confirm the expression pattern of the *CHRNA9* gene in *longissimus dorsi* muscle. Pearson correlations between the  $\Delta$ Ct and RNA-seq log-cpm for *CHRNA9* transcript abundance was -0.58. The marker DIAS0000678 was significantly associated with both RNA-seq and  $\Delta$ Ct for

*CHRNA9* (p-value  $\leq 4.23\text{e-}06$ ), exhibiting a significant dominant effect with the B allele associated with increased *CHRNA9* transcript abundance (p-value  $\leq 0.05$ , Table 2.5).

**Table 2.5** Comparison of RT-qPCR and RNA-seq for *CHRNA9* gene expression.

|                            | Estimate | Standard Error | Test Statistic | p-value  |
|----------------------------|----------|----------------|----------------|----------|
| <b>RT-qPCR<sup>1</sup></b> |          |                |                |          |
| DIAS0000678                | 150.89   | 30.63          | 4.93           | 8.37e-07 |
| Contrasts                  |          |                |                |          |
| A vs B                     | 2.19     | 0.67           | 3.26           | 1.12e-03 |
| AA vs AB                   | 0.86     | 0.68           | 1.26           | 2.08-e01 |
| AA vs BB                   | 3.32     | 0.71           | 4.69           | 2.66e-06 |
| AB vs BB                   | 2.46     | 0.73           | 3.37           | 7.60e-04 |
| <b>RNA-seq<sup>2</sup></b> |          |                |                |          |
| DIAS0000678                | -141.55  | 30.77          | -4.60          | 4.23e-06 |
| Contrasts                  |          |                |                |          |
| A vs B                     | -135     | 0.66           | -2.03          | 4.27e-02 |
| AA vs AB                   | -0.89    | 0.56           | -1.58          | 1.13-e01 |
| AA vs BB                   | -3.07    | 0.70           | -4.07          | 1.05e-05 |
| AB vs BB                   | -2.17    | 0.63           | -3.46          | 5.38e-04 |

<sup>1</sup> $\Delta$ Ct values for *CHRNA9* gene expression obtained with RT-qPCR. <sup>2</sup>Log-cpm for *CHRNA9* gene expression obtained with RNA-seq.

## DISCUSSION

For this study, we identified 334 eQTL for longissimus dorsi muscle transcripts from pigs in an F2 resource population. We declared local versus distant eQTL effects based on LD structure, identifying 188 local and 146 distant regulators of gene expression. Heritability of gene expression was estimated in this study with 344 gene transcripts exhibiting significant heritable expression. A joint analysis of eQTL with pQTL showed four genomic regions associated with variation in gene transcript abundance (N=16) and variation in phenotypes for growth (SSC6), carcass composition (SSC6, SSC11 and SSC15) and meat quality traits (SSC2 and SSC15). Most eQTL associated with pQTL were distant regulators of gene expression (69%). These distant regulators mapped to a putative hotspot on SSC15 associated with meat quality and carcass

composition traits. The remaining three genomic regions associated with variation in gene transcript abundance and trait phenotypes contained local regulators of gene expression.

When an eQTL and the associated gene are located on the same chromosome, the low resolution of the swine genome due to long range linkage disequilibrium<sup>47,48</sup> limits the ability to distinguish between cis-acting and trans-acting eQTL. Most eQTL association studies use a fixed distance threshold between the position of the eQTL peak and the gene transcript to define cis-acting (i.e., local) versus trans-acting (i.e., distant) regulation. For instance, distance thresholds between 1 Mb and 10 Mb have been used in recent pig eQTL maps<sup>12,15,49–52</sup>. Human eQTL scans have used more conservative distance thresholds of 100kb - 500kb between gene position and eQTL to declare local regulation<sup>53,54</sup>. A shorter local threshold is logical for human eQTL studies because they typically show higher resolution due to increased SNP density (millions of genotyped markers<sup>53</sup>), and the extent of LD is much more limited than in livestock populations due to greater genetic diversity in human populations<sup>54</sup>. In this study, we present an alternative to the use of a fixed distance for declaring local versus distant eQTL effects. This is important because the range of a mapped eQTL will depend on the LD pattern at the QTL genomic position. Building upon previous approaches to determine local regulation<sup>13,17,55,56</sup> in eQTL linkage maps, this study considered the significance of each individual marker surrounding the plausible position range of the eQTL peak to distinguish between local and distant modes of action. In cases where there are no genotyped markers between the plausible position of the eQTL peak and the position of the associated gene, there is not sufficient information to determine local versus distant; here we consider this scenario as plausible local regulation. We note that in our study the median distance between plausible local eQTL regulators and their associated gene was 24kb, which is a shorter distance than eQTL designated as local for other

pig eQTL mapping studies<sup>12,15,49–52</sup>. Therefore, it is feasible that most of these regulators may be acting locally, since cis-acting transcription factor binding sites have been found located ~100kb from the mapped position of a gene transcript<sup>57</sup>. However, without a more dense SNP set and/or a larger population size, we cannot definitively identify the mode of action of these eQTL. A potential way to further investigate if these eQTL are acting locally or distantly would be through allele-specific expression analyses<sup>16</sup>.

Heritability of gene expression contributes to our understanding of the inheritance of gene expression regulation. Estimating the heritability of gene expression is common in human eQTL studies to elucidate the genetic contribution of gene expression variation and its influence on the divergence of complex traits<sup>53,54,58,59</sup>. Human studies have shown higher heritability estimates for housekeeping genes and genes with local eQTL, whereas genes with distant eQTL tend to exhibit lower heritability<sup>53,58,59</sup>. Bryois *et al.*<sup>58</sup> suggested a fraction of missing heritability may be due to common variants with both local and distant effects on gene expression, with the latter being of small effect size. Examples of local eQTL with large distant effects in human studies include variants influencing expression of transcription factor genes or histone methyltransferase genes<sup>58</sup>. Heritability of gene expression has not been emphasized in pig eQTL studies, with the exception of one report where heritability was used as a filtering criteria to prioritize genes<sup>56</sup>. In this study, we estimated narrow sense heritability for all gene expression profiles and determined significance with likelihood ratio tests. Among all transcripts, only 2% exhibited significant heritable expression ( $FDR \leq 0.01$ ). However, the significantly heritable transcripts were enriched among eQTL, with 35% of eQTL exhibiting significantly heritable expression. Consistent with previous studies in humans, the observed heritabilities for genes with distant eQTL were significantly lower than for locally regulated genes<sup>53</sup>. This trend is consistent

with previous findings where genes influenced by many distant factors of small effect tend to exhibit lower heritability than genes with local regulation. Testing for significant additive genetic effects of transcript abundance in outbred animal populations requires large sample size to increase power to detect smaller effects. In our GWA scan, we were able to capture the variance associated with gene transcripts subject to genetic control with low heritability. A previous eQTL scan performed with 57 muscle tissue samples from an F2 swine population observed an average heritability of 0.45 for eQTL genes<sup>56</sup>. While this value is greater than the average heritability observed in our study (0.32), Liaubet *et al.*<sup>56</sup> limited the eQTL scan to gene transcripts with heritability greater than 0.05. The use of a heritability threshold to filter genes in eQTL studies may miss potential associations, especially those of low effect such as distant eQTL, which we show to have lower average heritability estimates.

We identified three gene transcripts that were associated with pQTL for fat deposition and carcass composition traits on SSC6. One of these eQTL genes, synovial sarcoma X breakpoint 2 interacting protein (*SSX2IP*), was significantly associated with pQTL for carcass 10<sup>th</sup> rib backfat and loin weight. An eQTL was previously identified for this gene on SSC6 using microarray data from the same animals used in this study, and consistent with our results Peñagaricano *et al.*<sup>60</sup> reported a negative causal effect of increased expression of *SSX2IP* on backfat thickness<sup>60</sup>. Interestingly, *SSX2IP* has been associated with waist to hip ratio, a common measure of body fat distribution, in women of African descent<sup>61</sup>.

Genes associated with pQTL for tenderness phenotypes on SSC2 or meat quality phenotypes on SSC15 share biological processes known to directly influence the organoleptic properties of meat, including calcium signaling (*FRMD8*, *MRLN*, *PKP2* and *CHRNA9*), energy metabolism (*SUCLG2* and *PFKFB3*), redox hemostasis (*NQO1* and *CEP128*) and cytoskeletal

structure (*CIT* and *CCDC60*). One of the genes related to calcium signaling is the FERM domain containing 8 (*FRMD8*) gene associated with pQTL for WBS, and sensory panel tenderness and overall tenderness on SSC2. Two independent GWAS, one in a crossbred commercial pig population<sup>62</sup> and another in a multigenerational Landrace-Duroc-Yorkshire composite population<sup>63</sup>, reported QTL for slice shear force (a procedure similar to WBS) in the same genomic region as this study. Zhang et al.,<sup>62</sup> identified *FRMD8* to be one of four genes in the region to play a role in pork tenderization and the peak SNP reported by Nonneman et al.<sup>63</sup>, was the same peak SNP identified in our analysis (H3GA0005672). We showed with our conditional analysis that increased expression of *FRMD8* was associated with improvements in pork tenderness. *FRMD8* is a member of the FERM (Four-point-one, Ezrin, Radixin, Meosin) protein superfamily known to possess both structural and signaling functions including numerous protein-binding interactions mainly in the cytoskeleton of cells<sup>64</sup>. This includes interactions with transmembrane ion channels and membrane lipids including the phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>). PIP<sub>2</sub> is the precursor of inositol 1,4,5-triphosphate (IP<sub>3</sub>) involved in Ca<sup>2+</sup> signaling<sup>65–67</sup> and IP<sub>3</sub> has been suggested as a potential indicator of meat tenderness in beef cattle<sup>68</sup>. The activation of the PIP<sub>2</sub> Ca<sup>2+</sup> signaling system controls diverse cellular processes in numerous tissues<sup>69</sup>. In skeletal muscle the sarcoplasmic reticulum ryanodine receptor is the Ca<sup>2+</sup> release channel, however PIP<sub>2</sub> has been localized to the transverse tubular membrane and IP<sub>3</sub> receptors have been found in differentiated muscle fibers, and implicated in excitation-contraction coupling (for review see Csernoch et al.<sup>70</sup>). *FRMD8* may play a role in Ca<sup>2+</sup> signaling and excitation-contraction coupling of skeletal muscles through interactions with PIP<sub>2</sub>.

Similar to *FRMD8*, the *MRLN* gene is also implicated in muscle contraction. *MRLN* encodes myoregulin, a micropeptide inhibitor of the sarco/endoplasmic reticulum Ca<sup>2+</sup> ATPase

(SERCA). SERCA regulates relaxation after muscle contraction, specifically, it pumps  $\text{Ca}^{+2}$  back to the sarcoplasmic reticulum. Binding myoregulin to SERCA lowers its affinity to  $\text{Ca}^{+2}$ , reducing the rate of  $\text{Ca}^{+2}$  reuptake into the sarcoplasmic reticulum<sup>71</sup>. Increased expression of *MRLN* was associated with improvements in pork tenderization, decreased 24-hour pH and increased drip loss in our study. The observed effect of *MRLN* gene expression on meat quality phenotypes may be due to its involvement in regulating muscle contractility and calcium signaling which have a direct effect on postmortem proteolysis.

Additional genes implicated in calcium signaling and associated with meat quality phenotypes and the putative hotspot were the *PKP2* and *CHRNA9* genes. *PKP2* encodes a plakophilin protein known to localize to cell desmosomes and nuclei and play a role in linking cadherins to intermediate filaments in the cytoskeleton. In mouse cardiac muscle, *PKP2* has been shown to regulate the transcription of genes controlling intercellular calcium homeostasis, and reduced expression of *PKP2* decreases the expression of several calcium signaling genes including the cardiac muscle ryanodine receptor<sup>72</sup>. In this study, increased expression of *PKP2* was associated with improvements in pork tenderization and decreases in 24-hour pH, protein percent and cook yield suggesting a role for this gene in modulating skeletal muscle calcium signaling during the conversion of muscle to meat. The *CHRNA9* gene is one of sixteen subunits of the nicotinic acetylcholine receptor (AChR). These ligand-gated ion channels permit the transmission of presynaptic acetylcholine release and postsynaptic excitatory potential. Found only in neuronal tissue, *CHRNA9* is one of three AChR containing only  $\alpha$  subunits<sup>46</sup> ( $\alpha 9$ -AChR), and in neuromuscular junctions AChR are essential for muscle contraction<sup>46</sup>. Since  $\alpha 9$ -AChR possess higher calcium permeability, they play an important role in catecholamine secretion and the adaptive response to chronic stress<sup>45</sup>. In this study, increased expression of *CHRNA9* was

associated with improved tenderness scores, increased drip loss, and decreased cook yield, protein percent and 24-hour pH. In addition, we verified the expression of *CHRNA9* in skeletal muscle with RT-qPCR and confirmed a significant dominance effect of the peak eQTL SNP (DIAS0000678) on *CHRNA9* gene expression. Changes in the expression of *CHRNA9* may potentially regulate the postsynaptic excitatory potential during the conversion of muscle to meat thereby influencing  $\text{Ca}^{2+}$  release to the cytoplasm, apoptotic mitochondrial changes and proteolytic enzymatic activity.

Additional genes associated with meat quality traits on SSC15 (*PFKFB3*, *CEP128*, *NQO1* and *SUCLG2*) were implicated in biological processes related to redox homeostasis and energy metabolism. The *PFKFB3* gene regulates the synthesis and degradation of fructose-2, 6-bisphosphate and fructose-6-phosphate in the process of glucose metabolism. The promoter of the *PRKFB3* gene contains hypoxia-inducible factor-1 (HIF-1) binding sites<sup>73</sup>. The transcription factor HIF-1 is a master regulator of oxygen homeostasis by activating several downstream pathways including the mitogen-activated protein kinase (MAPK), mammalian target of rapamycin (mTOR), phosphoinositide 3-kinase-protein kinase B (PI3K-Akt), vascular endothelial growth factor (VEGF) and calcium signaling pathways as well as anaerobic metabolism. *PFKFB3* is consistently overexpressed in many tumor cells and knockdown of *PFKFB3* promotes apoptosis of tumor cells<sup>73</sup>. Rapidly proliferating tumor cells have the ability to increase glucose uptake by using anaerobic glycolysis as the primary source of energy, known as the Warburg effect. Taken together *PFKFB3* is critical for cell proliferation and survival by regulating glucose metabolism and prevents apoptosis through the activation of cyclin-dependent kinases<sup>73,74</sup>. No reports have suggested a role for *PRKFB3* in meat quality. However, in our

study, increased expression of *PRKFB3* was associated with increased pork tenderness. Thus, *PRKFB3* may be involved in postmortem glycolytic potential similar to *PRKAG3*.

The *CEP128* gene is related to the PI3K-Akt-mTOR signaling pathway. Centrosomal protein 128 (*CEP128*) is part of the centrosomal protein family including *CEP55* which have been implicated in cancer progression<sup>75</sup>. Mutations within *CEP128* have been associated with an aggressive type of lymphoma, the diffuse large B-cell lymphoma (DLBCL)<sup>76</sup>. Functional gene studies have not been performed for *CEP128*, however mutations identified in refractory DLBCL patients, including those in *CEP128*, were associated with PI3K-Akt-mTOR signaling pathways and increased mitochondrial oxidative phosphorylation, and play an important role in treatment resistance<sup>76</sup>. The PI3K-Akt-mTOR pathway is upregulated in cancer cells, controlling the survival and proliferation of these cells. In our study, increased expression of *CEP128* was associated with improved tenderness scores and may play a role in PI3K/Akt/mTOR signaling. In addition, the Edomucin (*EMCN*) gene associated with a local acting eQTL on SSC8 plays a critical role in angiogenesis. Angiogenesis is the process of new blood vessel formation with its key regulator, vascular endothelial growth factor (VEGF), triggering downstream signaling cascades including MAPK-ERK1/2, PI3k/Akt and p38-MAPK pathways<sup>77</sup>. These signaling pathways promote endothelial cell migration, proliferation, and survival and are activated by HIF-1 which induces VEGF expression<sup>78</sup>. While this eQTL is not directly associated with a phenotype in our population, it is connected to the pathways regulated by the genes associated with the putative hotspot on SSC15.

The remaining two genes, *NQO1* and *SUCLG2*, were associated with improvements in meat tenderization and pH decline. The nuclear erythroid-2-p45-related factor-2 (Nrf2) is a transcription factor known to regulate redox homeostasis and anti-inflammatory response by

controlling the expression of Phase I and Phase II anti-oxidant enzymes containing the antioxidant response element (ARE; cis-acting regulatory or enhancer sequence) in their promoter region. *NQO1* (NADPH quinone oxidoreductase-1) is one of these enzymes whose expression is induced by Nrf2 in several tissues<sup>79–82</sup>. Consequently, knockdown of Nrf2 has been reported to significantly decrease expression of *NQO1* in both mouse skeletal muscle<sup>81</sup> and C2C12 mouse myotubes<sup>82</sup>. In early postmortem muscle, the antioxidant defense system is speculated to influence proteolysis and thereby meat tenderization<sup>7</sup>. Increased expression of *NQO1* in this study was associated with several meat quality traits including tenderness, pH and drip loss phenotypes implying a significant role in post-mortem proteolysis. The succinate-CoA ligase GDP-forming beta subunit (*SUCLG2*) has been implicated in the *SUCLG1*-related mitochondrial DNA depletion syndrome affecting brain and skeletal muscle tissues. Individuals affected by this syndrome present an array of symptoms including spasmodic muscle contractions, contracture or destruction of muscle cells and hypoglycemia<sup>83</sup>. Knockdown of the *SUCLG2* gene in fibroblasts was reported to decrease mitochondrial DNA, mitochondrial nucleoside diphosphate kinase and cytochrome c oxidase activities<sup>84</sup>. These results highlight the critical role *SUCLG2* plays in mitochondrial DNA maintenance and ATP production. In our study, increased expression of *SUCLG2* was associated with improvements in meat quality traits suggesting a potential role in regulating ATP production and postmortem pH decline.

In addition to genes involved in specific biological functions, genes encoding structural proteins were also observed to be associated with the putative hotspot on SSC15 (*CIT* and *CCDC60*). *CIT*, citron Rho-interacting serine/threonine kinase, is considered to be a scaffold protein that binds to several mitotic proteins, and knockout of *CIT* leads to cytokinetic defects. One such protein-protein interaction involves the two-pore channel 1 (*TPC1*) which Horton et.

al,<sup>85</sup> reported to cause disruption in myosin light chain phosphorylation (pMLC). In skeletal muscle pMLC has been associated with age related muscle dysfunction<sup>86</sup>, and decreased pMLC is associated with reduced fraction of myosin heads interacting with thin filaments<sup>86</sup>. Thus, increased expression of *CIT* could potentially increase muscle breakdown, which is consistent with our findings where higher expression of *CIT* was associated with improvements in pork tenderization, and reduced protein content and cook yield. *CCDC60* is a coil-coil domain protein, which are believed to act as “cellular velcro” holding together molecules, cellular structures and tissues<sup>87</sup>. The biological function of *CCDC60* is unknown, but recent GWAS have associated this gene with the neurological disorder schizophrenia in humans<sup>88</sup>. A proteomic analysis of post-mortem pre-frontal cortex of schizophrenia patients and non-schizophrenia individuals identified differentially expressed proteins involved in calcium homeostasis, cytoskeleton assembly and energy metabolism<sup>89</sup>. It is feasible that similar functions may occur in skeletal muscle tissue. In this study increased expression of *CCDC60* was associated with tenderness, pH, cook yield and drip loss phenotypes implicating the role of this gene in the conversion of muscle to meat.

Eleven eQTL genes were enriched in pQTL for meat quality traits on SSC15; *PFKFB3*, *SUCLG2*, *CIT*, *CCDC60*, *MRLN*, *PKP2*, *NQO1*, *CEP128*, *CHRNA9*, *TEX9* and a novel transcript SSC15:48.94. The novel transcript mapped to an uncharacterized locus, LOC110257028, on SSC15. The other ten gene transcripts mapped to different chromosomes than their associated eQTL. These results illustrate the advantage of the joint association of gene expression profiles and trait phenotypes to uncover the genetic architecture of polygenic traits. In this study, increased expression of the 11 genes was associated with improvements in meat quality phenotypes. Moreover, this QTL region harbors a putative hotspot (H3GA0052416) regulating

the expression of all 11 gene transcripts. Breitling et al. reported the high false positive rate associated with hotspot discovery, in order to mitigate this we used a higher threshold of significance to detect eQTL. The hotspot discovered on SSC15 was also associated with the most significant marker for multiple meat quality phenotypes. The high correlation observed between the 11 gene expressions, and between the eight meat quality phenotypes raises the question if these associations are due to a master regulator on SSC15. The *PRKAG3* gene has been suggested as such a regulator of meat quality traits in pigs. *PRKAG3* regulates glycogen potential, which has a cascading effect in postmortem metabolism. The SNP map used in this study does not have sufficient coverage of the *PRKAG3* gene. To address this our F2 population was genotyped for two known *PRKAG3* SNPs, I199V and T30N<sup>26</sup>, however, *PRKAG3* did not explain the relationship observed in the putative hotspot. A missense polymorphism within the *PRKAG3* gene, T30N SSC15:120.865 Mb, was significantly associated with just one of the 11 genes, *NQO1*, despite showing significant association with all eight meat quality phenotypes in this population<sup>26</sup>.

## CONCLUSION

In summary, the joint analysis of pQTL with eQTL from our well characterized pig resource population identified molecular markers significantly associated with both economically important phenotypes and gene transcript abundance. This approach revealed both local and distant acting regulators of gene expression influencing meat quality, carcass composition and growth traits. These phenotypic traits are correlated, and we show how correlated phenotypes exhibit correlated gene expression measured through a putative hotspot contained within QTL regions for both expression and phenotypic traits. We highlight novel candidate genes with specific roles in cytoskeletal structure and signaling pathways regulating

meat quality phenotypes including redox hemostasis (*NQO1* and *CEP128*), energy metabolism (*SUCLG2* and *PRKFB3*),  $\text{Ca}^{2+}$  signaling (*FRMD8*, *MRLN*, *PKP2* and *CHRNA9*) and cytoskeletal structure (*CIT* and *CCDC60*) during the initial conversion of muscle to meat. Taken together the identified genes and their associated functions and pathways increase our knowledge of the genomic architecture of meat quality phenotypes.

## SUPPLEMENTARY MATERIALS

Supplementary tables available at <https://velezdeb84.wixsite.com/deborahvelezirizarry>

**Supplementary Table 2.S1** Summary statistics for phenotypic traits for the MSUPRP F2 population and the subsample used in this study.

**Supplementary Table 2.S2** Expression quantitative trait locus (eQTL) mapped for longissimus dorsi muscle transcripts from the MSUPRP (n=168).

**Supplementary Table 2.S3** Phenotypic QTL identified in the F2 MSUPRP.

**Supplementary Table 2.S4** Results of conditional analysis for expression QTL co-localized with phenotypic QTL.

**Supplementary Table 2.S5** Conditional Analysis: PRKAG3 SNP effect on eQTL genes expression.

## REFERENCES

## REFERENCES

1. Dekkers, J. C. M. Commercial application of marker- and gene-assisted selection in livestock : Strategies and lessons. *J. Anim. Sci.* **82**, E313–E328 (2004).
2. Kadarmideen, H. N., von Rohr, P. & Janss, L. L. G. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mamm. Genome* **17**, 548–64 (2006).
3. Goddard, M. E. & Hayes, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **10**, 381–391 (2009).
4. Steen, H. A. M. Van Der, Prall, G. F. W. & Plastow, G. S. Application of genomics to the pork industry. *J. Anim. Sci.* **83**, E1–E8 (2005).
5. Edwards, D. B. *et al.* Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: I. Growth traits. *J. Anim. Sci.* **86**, 241–253 (2008).
6. Edwards, D. B. *et al.* Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. *J. Anim. Sci.* **86**, 254–66 (2008).
7. England, E. M., Schef, T. L., Kasten, S. C., Matarneh, S. K. & Gerrard, D. E. Exploring the unknowns involved in the transformation of muscle to meat. *Meat Sci.* **95**, 837–843 (2013).
8. Hu, Z., Park, C. A., Reecy, J. M., Animal, T. & Database, Q. T. L. Developmental progress and current status of the Animal QTLdb. *Nuc* **44**, D827–D833 (2016).
9. Wan, X. *et al.* Elucidating a molecular mechanism that the deterioration of porcine meat quality responds to increased cortisol based on transcriptome sequencing. *Sci. Rep.* **6**, 36589 (2016).
10. Rocha, L. M., Velarde, A., Dalmau, A., Saucier, L. & Faucitano, L. Can the monitoring of animal welfare parameters predict pork meat quality variation through the supply chain ( from farm to slaughter )? *J. Anim. Sci.* **94**, 359–376 (2016).
11. Hao, Y. *et al.* Transcriptome analysis reveals that constant heat stress modifies the metabolism and structure of the porcine longissimus dorsi skeletal muscle. *Mol. Genet. Genomics* **291**, 2101–2115 (2016).
12. Ponsuksili, S., Du, Y., Murani, E., Schwerin, M. & Wimmers, K. Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. *Genetics* **192**, 1109–1122 (2012).

13. Ponsuksili, S. *et al.* Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* **9**, 367 (2008).
14. Wimmers, K., Murani, E. & Ponsuksili, S. Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat performance in pigs. *Brief. Funct. Genomics* **9**, 251–258 (2010).
15. Ma, J. *et al.* A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Biol.* **10**, e1004710 (2014).
16. Ernst, C. W. & Steibel, J. P. Molecular advances in QTL discovery and application in pig breeding. *Trends Genet.* **29**, 215–224 (2013).
17. Muñoz, M. *et al.* Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. *BMC Genomics* **14**, 845 (2013).
18. Heidt, H. *et al.* A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. *Mamm. Genome* **24**, 416–426 (2013).
19. Steibel, J. P. *et al.* Genome-Wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. *PLoS One* **6**, e16766 (2011).
20. Cardoso, F. F. *et al.* Selective transcriptional profiling and data analysis strategies for expression quantitative trait loci mapping in outbred F2 populations. *Genetics* **180**, 1679–90 (2008).
21. Gualdrón Duarte, J. L. *et al.* Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels. *BMC Genet.* **14**, 38 (2013).
22. Badke, Y. M. *et al.* Methods of tagSNP selection and other variables affecting imputation accuracy in swine. *BMC Genet.* **14**, 8 (2013).
23. Ramos, A. M. *et al.* Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* **4**, e6524 (2009).
24. Ciobanu, D. *et al.* Evidence for new alleles in the protein kinase adenosine monophosphateactivated gamma3-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. *Genetics* **159**, 1151–1162 (2001).
25. Milan, D. *et al.* A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).

26. Casiró, S. *et al.* Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. *J. Anim. Sci.* **95**, 554–558 (2017).
27. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
28. Gordon, A. & Hannon, G. J. Fastx-toolkit. Computer program distributed by the author. (2010). at <[http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)>
29. Kim, D. *et al.* TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
30. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
32. Anders, S., Pyl, P. T. & Huber, W. HTSeq A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).
33. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
34. Dillies, M.-A. *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* **14**, 671–83 (2013).
35. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
36. Gualdrón Duarte, J. L. *et al.* Refining genomewide association for growth and fat deposition traits in an F 2 pig population. *J. Anim. Sci.* **94**, 1387–97 (2016).
37. Vanraden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
38. Visscher, P. M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res. Hum. Genet.* **9**, 490–495 (2006).
39. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
40. Bauer, D. F. Constructing confidence sets using rank statistics. *J. Am. Stat. Assoc.* **67**, 687–690 (1972).

41. Gualdrón Duarte, J. L. *et al.* Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* **15**, 246 (2014).
42. Bernal Rubio, Y. L. *et al.* Implementing meta-analysis from genome-wide association studies for pork quality traits 1. *J. Anim. Sci.* **93**, 5607–5617 (2015).
43. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).
44. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 34–1 (2002).
45. Guérineau, N. C., Desarménien, M. G., Carabelli, V. & Carbone, E. Functional chromaffin cell plasticity in response to stress: Focus on nicotinic, gap junction, and voltage-gated Ca<sup>2+</sup> channels. *J. Mol. Neurosci.* **48**, 368–386 (2012).
46. Di Cesare Mannelli, L. *et al.*  $\alpha$ -Conotoxin RgIA protects against the development of nerve injury-induced chronic pain and prevents both neuronal and glial derangement. *Pain* **155**, 1986–1995 (2014).
47. Wimmers, K. *et al.* Associations of functional candidate genes derived from gene-expression profiles of prenatal porcine muscle tissue with meat quality and muscle deposition. *Anim. Genet.* **38**, 474–484 (2007).
48. Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C. & Steibel, J. P. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* **13**, 24 (2012).
49. Chen, C. *et al.* A genome-wide investigation of expression characteristics of natural antisense transcripts in liver and muscle samples of pigs. *PLoS One* **7**, e52433 (2012).
50. Kogelman, L. J. A., Zhernakova, D. V, Westra, H., Cirera, S. & Fredholm, M. An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. *Genome Med.* **7**, 105 (2015).
51. Ponsuksili, S., Murani, E., Brand, B., Schwerin, M. & Wimmers, K. Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J. Lipid Res.* **52**, 668–678 (2011).
52. Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M. & Wimmers, K. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *Int. J. Biol. Sci.* **10**, 327–337 (2014).
53. Yang, S. *et al.* Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. *BMC Genomics* **15**, 13 (2014).

54. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat Genet.* **39**, 1202–7. (2007).
55. Cánovas, A. *et al.* Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population. *PLoS One* **7**, e35583 (2012).
56. Liaubet, L. *et al.* Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics* **12**, 548 (2011).
57. Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).
58. Bryois, J. *et al.* Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* **10**, e1004461 (2014).
59. Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–7 (2014).
60. Peñagaricano, F. *et al.* Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. *BMC Syst. Biol.* **9**, 58 (2015).
61. Ng, M. C. Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African ancestry anthropometry genetics consortium. *PLoS Genet.* **13**, e1006719 (2017).
62. Zhang, C. *et al.* Genome wide association studies (GWAS) identify QTL on SSC2 and SSC17 affecting loin peak shear force in crossbred commercial pigs. *PLoS One* **11**, e0145082 (2016).
63. Nonneman, D. J. *et al.* Genome-wide association of meat quality traits and tenderness in swine. *J. Anim. Sci.* **91**, 4043–4050 (2013).
64. Tepass, U. FERM proteins in animal morphogenesis. *Curr. Opin. Genet. Dev.* **19**, 357–367 (2009).
65. Chang, C. L. *et al.* Feedback regulation of receptor-induced  $Ca^{2+}$  signaling mediated by e-syt1 and nir2 at endoplasmic reticulum-plasma membrane junctions. *Cell Rep.* **5**, 813–825 (2013).
66. Chang, C. L. & Liou, J. Phosphatidylinositol 4, 5-bisphosphate homeostasis regulated by Nir2 and Nir3 proteins at endoplasmic reticulum-plasma membrane junctions. *J. Biol. Chem.* **290**, 14289–14301 (2015).

67. Chang, C. L. & Liou, J. Homeostatic regulation of the PI(4,5)P<sub>2</sub>-Ca<sup>2+</sup> signaling system at ER-PM junctions. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1861**, 862–873 (2016).
68. Kim, N. K. *et al.* Proteins in longissimus muscle of Korean native cattle and their relationship to meat quality. *Meat Sci.* **80**, 1068–1073 (2008).
69. Berridge, M. J. Inositol trisphosphate and calcium signalling mechanisms. *Biochim. Biophys. Acta - Mol. Cell Res.* **1793**, 933–940 (2009).
70. Csernoch, L. & Jacquemond, V. Phosphoinositides in Ca<sup>2+</sup> signaling and excitation–contraction coupling in skeletal muscle: an old player and newcomers. *J. Muscle Res. Cell Motil.* **36**, 491–499 (2015).
71. Anderson, D. M. *et al.* Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* **9**, ra119 (2016).
72. Cerrone, M. *et al.* Plakophilin-2 is required for transcription of genes that control calcium cycling and cardiac rhythm. *Nat. Commun.* **8**, 106 (2017).
73. Lu, L., Chen, Y. & Zhu, Y. The molecular basis of targeting PFKFB3 as a therapeutic strategy against cancer. *Oncotarget* **8**, 62793–62802 (2017).
74. Yalcin, A. *et al.* 6-Phosphofructo-2-kinase (PFKFB3) promotes cell cycle progression and suppresses apoptosis via Cdk1-mediated phosphorylation of p27. *Cell Death Dis.* **5**, e1337 (2014).
75. Li, F., Jin, D., Tang, C. & Gao, D. CEP55 promotes cell proliferation and inhibits apoptosis via the pi3k/akt/p21 signaling pathway in human glioma u251 cells. *Oncol. Lett.* **15**, 4789–4796 (2018).
76. Park, H. Y. *et al.* Whole-exome and transcriptome sequencing of refractory diffuse large B-cell lymphoma. *Oncotarget* **7**, 86433–86445 (2016).
77. Park-Windhol, C. *et al.* Endomucin inhibits VEGF-induced endothelial cell migration, growth, and morphogenesis by modulating VEGFR2 signaling. *Sci. Rep.* **7**, 17138 (2017).
78. Lin, C., McGough, R., Aswad, B., Block, J. & Terek, R. Hypoxia induces HIF-1-alpha and VEGF expression in chondrosarcoma cells and chondrocytes. *J. Orthop. Res.* **22**, 1175–1181 (2004).
79. Lampiasi, N. & Montana, G. An in vitro inflammation model to study the Nrf2 and NF-κB crosstalk in presence of ferulic acid as modulator. *Immunobiology* **223**, 339–355 (2017).
80. Chen, H. *et al.* A ROS-mediated mitochondrial pathway and Nrf2 pathway activation are involved in BDE-47 induced apoptosis in Neuro-2a cells. *Chemosphere* **184**, 679–686 (2017).

81. Miller, C. J. *et al.* Disruption of Nrf2/ARE signaling impairs antioxidant mechanisms and promotes cell degradation pathways in aged skeletal muscle. *Biochim. Biophys. Acta* **1822**, 1038–1050 (2012).
82. Horie, M., Warabi, E., Komine, S., Oh, S. & Shoda, J. Cytoprotective role of Nrf2 in electrical pulse stimulated C2C12 myotube. *PLoS One* **10**, e0144835 (2015).
83. El-Hattab, A. W. & Scaglia, F. Gene Reviews: SUCLG1-Related Mitochondrial DNA Depletion Syndrome, Encephalomyopathic Form with Methylmalonic Aciduria. (University of Washington, Seattle, 2017). at <https://www.ncbi.nlm.nih.gov/books/NBK425223/>
84. Miller, C., Wang, L., Ostergaard, E., Dan, P. & Saada, A. The interplay between SUCLA2, SUCLG2, and mitochondrial DNA depletion. *Biochim. Biophys. Acta* **1812**, 625–629 (2011).
85. Horton, J. S., Wakano, C. T., Speck, M. & Stokes, A. J. Two-pore channel 1 interacts with citron kinase, regulating completion of cytokinesis. *Channels* **9**, 21–29 (2015).
86. Gregorich, Z. R. *et al.* Top-down targeted proteomics reveals decrease in myosin regulatory light-chain phosphorylation that contributes to sarcopenic muscle dysfunction. *J. Proteome Res.* **15**, 2706–2716 (2016).
87. Rose, A., Schraegle, S. J., Stahlberg, E. A. & Meier, I. Coiled-coil protein composition of 22 proteomes - Differences and common themes in subcellular infrastructure and traffic control. *BMC Evol. Biol.* **5**, 66 (2005).
88. Kirov, G. *et al.* A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol. Psychiatry* **14**, 796–803 (2009).
89. Martins-De-Souza, D. *et al.* Prefrontal cortex shotgun proteome analysis reveals altered calcium homeostasis and immune system imbalance in schizophrenia. *Eur. Arch. Psychiatry Clin. Neurosci.* **259**, 151–163 (2009).

## CHAPTER THREE

### Allele-specific expression in *longissimus dorsi* muscle transcriptomes associated with phenotypic traits in pigs

Vélez-Irizarry, D., S. Casiro, R.O. Bates, N.E. Raney, H. Cheng, J.P. Steibel and C.W. Ernst

#### ABSTRACT

Advancements in sequencing technology, improvements in the annotation of the pig genome, and development of quantitative genetic models have contributed to an increased rate of genetic gain for economically important pig production traits. Several quantitative trait locus (QTL) have been identified, however, the biological mechanisms underlying most QTL remain unknown. Allele-specific expression (ASE) analysis facilitates the identification of cis-acting regulation of transcript abundance, which can be associated with a measurable phenotypic difference. In this study, we tested for ASE in 69,502 *longissimus dorsi* coding SNP (cSNP), which were called directly from RNA-seq data. A total of 18,234 cSNP with significant ASE were identified ( $FDR \leq 0.01$ ) using a Quasibinomial model. A meta-analysis merging cSNP p-values per gene identified 4,170 genes with significant allele-specific effects ( $FDR \leq 0.01$ ). A gene-wise conditional analysis fitting all ASE cSNP per gene for each phenotype identified 60 genes associated with growth, carcass composition and meat quality phenotypes. Ring finger and Zinc finger transcription factors were associated with 45-min pH, drip loss and 10<sup>th</sup>-rib backfat, and allelic expression bias for these genes was confirmed with pyrosequencing. Results support an important role for the activation of the PI3K-Akt-mTOR signaling pathway on meat quality traits.

**Key Words:** ASE, skeletal muscle, RNA-seq, pig

## INTRODUCTION

Genes exhibit specific patterns of expression finely modulated by spatial and temporal specificity, environmental conditions and allelic variation. A series of architectural elements cause this modulating effect. At the transcriptional level these include promoters, enhancers, silencers, and insulators, among others, and are collectively known as cis regulatory elements<sup>1,2</sup>. Polymorphism residing in cis regulatory elements can directly affect the transcription of a gene. For instance, a sequence motif containing a single nucleotide polymorphism (SNP) may affect the affinity of trans-acting regulators resulting in allele-specific expression because it affects only one of the alleles. Cis regulatory elements within coding regions are thus susceptible to nonsynonymous, synonymous and splice junction mutations that can lead to phenotypic consequences. For instance, an intergenic enhancer containing a variant associated with HIV-1 acquisition produces a shift in promotor use resulting in allele-specific isoform expression conferring susceptibility to HIV infection<sup>3</sup>. Imprinting occurs when methylation status of the parental copy of a gene is passed on to the offspring and can produce mono-allelic expression, where only one allele is expressed. For example, the IGF2 gene is regulated by an imprinting control region and the expression of IGF2 is transcribed mainly from the maternal allele regulating fetal development and postnatal growth<sup>4</sup>. Knowledge of cis-regulatory elements is expected to improve our understanding of phenotypic diversity in livestock species. Through allele-specific expression (ASE) analysis we can identify cis-acting variants by estimating the relative transcript abundance of each allele at a single heterozygous locus, and test for bias in allelic expression. This bias is observed as a departure from the expected equal expression ratio.

High-throughput sequencing provides in-depth coverage of polymorphic locus allowing estimation of allele-specific transcript abundance.

In pigs, numerous QTL have been identified for growth, carcass composition and meat quality traits<sup>5-13</sup>, however, the biological mechanisms regulating these QTL are largely unknown. Through functional genomic studies such as ASE analysis, the genetic architecture of important phenotypes can be evaluated. Previous ASE studies in pigs have used blood<sup>14</sup>, prenatal skeletal muscle<sup>15</sup> and brain<sup>16</sup> to elucidate locus exhibiting ASE and overlapping known QTL regions for growth and immune-related phenotypes. These studies have identified biomarkers for immune capacity<sup>14</sup>, chimeric RNA<sup>16</sup> and imprinted genes<sup>16</sup>. The aim of this study is to 1) elucidate ASE in the *longissimus dorsi* muscle transcriptome, and 2) identify genes with cis-acting effects associated with growth, carcass composition and meat quality traits. This work contributes toward unraveling the genetic architecture driving variation in economically important phenotypes.

## **MATERIALS AND METHODS**

### *RNA extraction and RNA-seq bioinformatic pipeline*

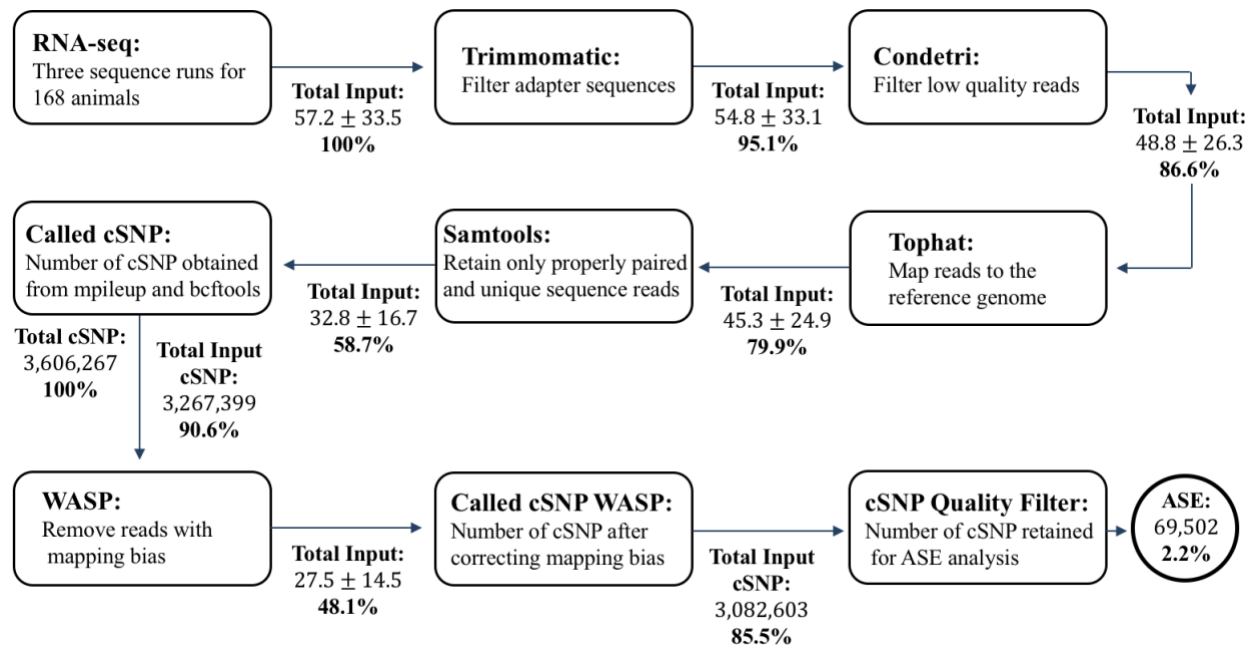
Tissue samples were collected post mortem from the *longissimus dorsi* muscle, flash frozen in liquid nitrogen and stored at -80°C until processing. Total RNA extraction was performed with the miRNeasy Mini Kit (Qiagen, Germantown, MD) following the manufacture's protocol. RNA quality and quantity were determined using the Agilent 2100 Bioanalyzer (RIN  $\geq$  7).

Sequencing was performed at the Michigan State University Research Technology Support Facility. Libraries for 24 samples were prepared using the Illumina TrueSeq RNA Library Prep Kit v2, and sequenced on the Illumina HiSeq 2000 platform (2 x 100bp paired-end

reads). The remaining 152 libraries were prepared using the Illumina TrueSeq Stranded mRNA Kit, and sequenced on the Illumina HiSeq 2500 platform (2 x 125bp, paired-end reads). Base calling was performed with the Illumina Real Time Analysis v1.18.61 software, and the Illumina Bc12fastq v1.8.4 was used for conversion to FastQ format. A total of 96 sequence files (741Gb) consisting of ~63 million short-reads per library were obtained from the HiSeq 2000 platform and 1,218 sequence files (~2Tb) of ~23 million short-reads per library were obtained from the HiSeq 2500 platform.

The bioinformatic pipeline used in this study first filtered RNA-seq reads for adapter sequences using Trimmomatic<sup>17</sup> followed by quality trimming using CondDeTri<sup>18</sup> where the first 6 bases at the 3' end, low quality reads (reads with 20% base quality scores < 25) and low quality bases (quality scores < 10) were filtered out retaining reads with a minimum length of 75 bases. This step is critical to remove sequencing errors with low quality scores<sup>18</sup>. The quality of each sequenced nucleotide was evaluated on adapter filtered and quality trimmed RNA-seq reads using the FASTX toolkit<sup>19</sup>. A mean Phred quality score of  $37.01 \pm 0.99$  was observed for sequenced nucleotides. The percentage of retained reads from each step in the bioinformatics pipeline is represented in Figure 3.1. On average 87% of reads were retained after adapter and quality filtering, eight samples were removed from further analysis due to low sequence quality, leaving a total of 168 samples for subsequent analyses. After adapter and quality filtering, RNA-seq reads were mapped to the reference genome assembly *Sus scrofa* 11.1 using the splice aware aligner Tophat<sup>20</sup>, on average 92% ( $45.3 \pm 24.9$  million short reads) mapped to the reference genome *Sus scrofa* 11.1. Sequence reads not mapping uniquely to the reference genome were removed from further analysis to eliminate duplicate read counts when calling cSNP<sup>21</sup>, on average 73% of mapped reads ( $32.8 \pm 16.7$  million short reads) were unique and properly paired

with its complementary sequence (Figure 3.1). Uniquely mapped reads were obtained with SAMtools<sup>22</sup>. Unfiltered sequence files for 168 animals has been deposited in the NCBI Sequence Read Archive accession number PRJNA403969.



**Figure 3.1** RNA-seq bioinformatics pipeline for cSNP calling.

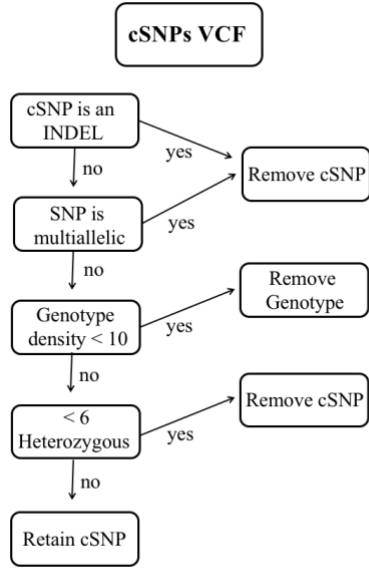
### *cSNP calling and unbiased allele-specific read mapping*

Allele-specific read counts were determined with a two-step procedure. First cSNP were called using SAMTools<sup>22</sup> mpileup to obtain the sequence of individual bases from each aligned transcript and bcftools to call the cSNP and genotypes for each animal<sup>23</sup>. Approximately 59% of sequence reads were retained for variant calling, Figure 3.1. Twenty VCF files (variant call format), one for each chromosome (18 autosomes and two sex chromosomes) were obtained. The genomic coordinates and observed nucleotides for each called cSNP were identified using an R package developed by our group, editTools<sup>24</sup>, <https://github.com/funkhou9/editTools>. The genomic coordinates and observed nucleotides for each called cSNP, excluding multiallelic cSNP and insertion deletions (INDEL), were used as input for WASP, an unbiased allele-specific

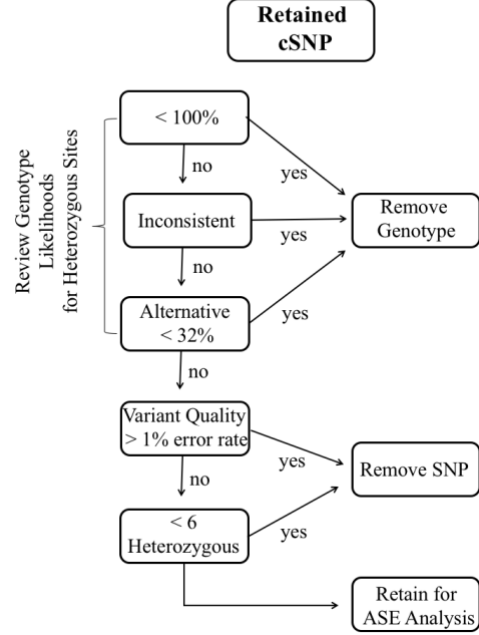
read mapper<sup>25,26</sup>. Briefly, sequence reads aligning with a polymorphic site are copied and modified so that the polymorphic site is switched to contain the alternative nucleotide in the position. The modified reads are then remapped to the reference genome using the same procedure described above. Modified reads are retained only if they map to exactly the same genomic position as the original read. The output of WASP are alignment files containing all reads correctly mapping to the genome. cSNP are called once more using the same procedure described above.

Additional filters were applied to ensure the removal of potentially erroneous SNP calls (Figure 3.2). Two filtering steps were performed. The first filter eliminates cSNP that are INDEL or multiallelic, since allelic imbalance cannot be accurately determined. cSNP with low read coverage, < 10 reads overlapping the polymorphic site, and low heterozygous genotype frequency, < 6 heterozygous samples were discarded from the analysis (Figure 3.2). The second filter ensures that monoallelic cSNP called heterozygous are removed from further analysis. This is achieved by flagging sites with low or inconsistent genotype likelihoods (Figure 3.2). The probability of erroneous ascertainment of variant allele was used to retain only high-quality variants. Sensitivity, accuracy and type I error rate of called cSNP from RNA-seq was estimated for heterozygous genotypes by comparing the overlap of cSNPs and SNPs ascertained from the Porcine SNP60 BeadChip<sup>13</sup>, assuming the genotypes for chip SNPs as the true genotype. Sensitivity of called cSNP was estimated as the ratio of true heterozygous calls from RNA-seq and total heterozygous genotypes from the Porcine SNP60 chip for overlapping SNP. Accuracy was estimated as the ratio of true heterozygous calls and total heterozygous calls from RNA-seq. Type I error rate is the ratio of total missed heterozygous calls and total heterozygous genotypes from the Porcine SNP60 chip.

### First Filter



### Second Filter



**Figure 3.2** Filtering pipeline to remove potentially erroneous cSNP calls.

### Allele-specific expression analysis

To test for significant allelic imbalance of cSNPs, a Quasibinomial model<sup>27,28</sup> was fit on a SNP by SNP basis for heterozygous samples as follows:

$$P(X = k_i) = \binom{n}{k_i} p_i (p_i + k_i \phi)^{k_i - 1} (1 - p_i - k_i \phi)^{n_i - k_i} \quad (8)$$

where,  $k_i$  is the number of reads mapping back to the non-reference allele of the cSNP in question for sample  $i$ ,  $p_i$  is the probability of observing a read for the non-reference allele given  $n$  total number of reads found mapping to the cSNP for sample  $i$ ,  $\hat{p}_i = k_i/n_i$ . The variance of  $\hat{p}_i$  is  $Var(\hat{p}_i) = \phi n_i \hat{p}_i (1 - \hat{p}_i)$ . Lastly,  $\phi$  is the overdispersion parameter calculated as:

$$\phi = \frac{1}{r-1} \sum_i \frac{(k_i - n_i \hat{p}_i)^2}{n_i \hat{p}_i (1 - \hat{p}_i)} \quad (9)$$

where,  $r$  are the degrees of freedom. The overall allelic (population-average) expression ratio,  $AR$ , for the cSNP is denoted as  $AR = h(p) = e^p / 1 + e^p$ , where  $h$  is the inverse of the log link function. The logit scale was used to ensure the allelic expression ratio is  $0 < p < 1$ . A t-test was used to test the hypothesis of significant ASE,  $H_0: p = 0.5$  versus  $H_a: p \neq 0.5$ , and genome-wide multiple test correction was performed<sup>29</sup> with  $FDR \leq 0.01$  considered significant.

Each ASE cSNP was mapped to gene transcripts using the pig genome assembly *Sus Scrofa* 11.1, in order to summarize gene-wise allele-specific expression. A potential limitation to this approach is gene-wise heterogeneity of ASE ratios and significance. For instance, alternative splicing, cis-trans interactions and antagonistic relationships between gene-wise ASE cSNP can make the interpretation of ASE difficult<sup>30</sup>. To circumvent this problem, a meta-analysis of gene-wise p-values was used to combine p-values from all cSNP mapping to a gene into a single significance measure. A robust approach to meta-analysis is the Simes method<sup>31</sup>, which adjusts all p-values on a gene-wise basis so that the minimum p-value can be selected for each gene, and multiple test correction performed ( $FDR \leq 0.01$ ).

#### *Confirmation of ASE cSNP*

To further assess ASE of cSNP, we selected nine cis-acting variants with empirical evidence of phenotypic regulation to confirm the observed allelic imbalance using pyrosequencing. The protocol used for the pyrosequencing assay is described in Kwok *et al.*<sup>32</sup>. Briefly, primers were designed to amplify the genomic region surrounding each of the nine cSNP using PyroMark Assay Design Software 2.5.8, including forward and reverse primers for polymerase chain reaction (PCR) and a sequencing primer for allele quantification (Supplementary Table 3.S1). Either the forward or reverse primer was biotinylated using Biotin-TEG and HPLC purification (IDT, Coralville, IA). PCR was performed for pigs exhibiting

heterozygous genotypes for each cSNP using total *longuissimus dorsi* muscle RNA (described above). Three negative controls were run for each cSNP, a no template control PCR reaction (examines primer heteroduplexing), a sequencing primer control (no PCR reaction, examines duplexing with sequencing primer) and template only control (no sequencing primer, examines self-priming of biotinylated primer). The positive controls were prepared from pools of total RNA from four homozygous animals for the AA and BB genotypes for each cSNP. A total of six positive controls were prepared for each cSNP as ratios of homozygous AA and BB pools (AA:BB = 0:100, 20:80, 40:60, 60:40, 80:20 and 100:0). The PyroMark OneStep RT-PCR Kit was used following the manufacturer's protocol and amplification was performed in a DNA Engine Peltier Thermal Cycler (Bio-Rad, Hercules, CA). Cycling conditions were 50°C for 30 min for reverse transcription and 95°C for 15 min for initial PCR activation, followed by 45 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s, with a final extension of 72°C for 10 min. PCR products, 25ul, were diluted in 11ul of 18.2 mΩ dd H<sub>2</sub>O and mixed with 40ul of the master mix containing 4ul of streptavidin-coated sepharose beads and 40ul of binding buffer (10mM Tris-HCL, 2M NaCl, 1mM EDTA and 0.1% Tween<sup>TM</sup> 20 pH 7.6) for a total volume of 80ul. This solution was agitated on a Monoshaker for at least five minutes. Immobilized PCR products were captured using a vacuum prep tool, washed and denatured to remove unbound primers and unbiotinylated strands using three solutions (i.e. 70% ethanol, denaturing solution containing 0.2M NaOH and wash buffer containing 10mM Tris-Acetate pH 7.6). Only the template strands remained bound after the washing steps. Sepharose beads with bound strands were diluted in a solution containing 0.2ul of sequencing primer and 38.8ul annealing buffer (20 mM Tris-Acetate, 5 mM MgAc<sub>2</sub> pH 7.6) and placed on a 96 sample thermoplate at 80°C for 2 minutes for annealing before samples were placed in the pyrosequencer PSQ 96MA machine.

PyroMark Gold Q96 reagents containing the enzymes, substrate and dNTPs for pyrosequencing were used in quantities recommended by the PyroMark AQ 2.5.8 software for each pyrosequencing assay analyzed. Relative levels of allele-specific expression were determined by the differing number of nucleotides incorporated at the cSNP site with the PyroMark AQ 2.5.8 software <sup>32</sup>.

#### *Kegg pathway and gene ontology enrichment*

Biological pathways and processes enriched with genes exhibiting significant ASE effects provide insights into gene expression networks regulated by genetic variation in our study population. Genes found with significant cis-acting effects were subjected to pathway analysis using the R package clusterProfiler<sup>33,34</sup>. The background gene list used in enrichment analysis consisted of all autosomal gene transcripts found expressed in *longissimus dorsi* for our population (15,249 transcripts). The gene symbols were converted to ENTREZ IDs using the human annotation<sup>35</sup>, and gene ontology for biological processes and Kegg pathway enrichment performed and significance determined after multiple test correction ( $FDR \leq 0.05$ ).

#### *Effects of ASE cSNP on trait phenotypes*

We selected cSNP with significant ASE for each of the genes identified through the meta-analysis as having significant cis-acting effects and tested their effects on variation in trait phenotypes. A gene-wise conditional analysis was performed for 67 phenotypes including growth, carcass composition and meat quality traits to estimate cSNP effects on phenotypic variation. A GBLUP model<sup>13,36,37</sup> was fit on a gene-by-gene basis for each phenotype as follows:

$$y_i = Xb + \sum_{l=1}^L R_{il}s_l + a_i + e_i, \quad (3)$$

where,  $y_i$  is the phenotypic data for sample  $i$ ,  $Xb$  the estimated fixed effects of overall mean and additional covariates specific to each phenotype<sup>7,13</sup>,  $s_l$  is the estimated cSNP effect for genotype

$l$  and  $R_{il}$  is the standardized allelic dosage of cSNP  $l$  for animal  $i$ . The R matrix was calculated as  $R = U / \sqrt{\sum (2\mathbf{p}(1 - \mathbf{p}))}$ , where U is a matrix of cSNP genotypes and  $\mathbf{p}$  a vector with the frequency of non-reference allele. The additive genetic effects,  $\mathbf{a}$ , were assumed to be  $\mathbf{a} \sim N(0, G\sigma_a^2)$  and the residual errors,  $\mathbf{e}$ , were assumed to be  $\mathbf{e} \sim N(0, I\sigma_e^2)$ . The genomic relationship matrix,  $\mathbf{G}$ , was previously calculated for our eQTL analysis using genotypes obtained for the 168 animals from the PorcineSNP60 BeadChip<sup>38</sup>. Multiple test correction was performed with a false discovery rate of 0.10 to determine significant cSNP effect. We estimated the proportion of variance explained by cis-acting variants for a single trait phenotype using methods described in Casiro *et al.*<sup>13</sup>. Briefly, the variance associated with each cSNP,  $\widehat{\sigma}_{s_l}^2$ , was estimated as  $\widehat{\sigma}_{s_l}^2 = s_l^2 \text{var}(Z_l)$ , where,  $s_l^2$  is the estimated effect of cSNP  $l$  and  $\text{var}(Z_l)$  the variance associated with the standardized allelic dosage of cSNP  $l$ . The proportion of phenotypic variance accounted for by each cSNP was  $\widehat{\sigma}_{s_l}^2 / (\sum \widehat{\sigma}_{s_l}^2 + \widehat{\sigma}_a^2 + \widehat{\sigma}_e^2)$ . The estimated additive genetic variance,  $\sigma_a^2$ , and error variance,  $\sigma_e^2$ , was obtained after fitting equation 3.

#### *Phenotypic QTL mapped with cSNP*

Calling cSNP directly from the *longuissimus dorsi* transcriptomes of the 168 animals increases genetic coverage to identify potential QTL segregating in our population, and distinguishes cSNP with ASE significantly associated with a phenotypic trait. First, we selected cSNP with less than 5% missing call rate and minor allele frequency greater than 0.01, resulting in 46,428 cSNP including 11,947 exhibiting significant ASE. Missing genotypes were imputed using BEAGLE 4.1<sup>39</sup>, a hidden Markov model that finds the most likely haplotype pairs to reconstruct missing genotypes, using the codeGeno function in the R package synbreed<sup>40</sup>. QTL were identified first using the GBLUP model described in equation 3 excluding fixed effects of individual cSNP,  $\sum_{l=1}^L Z_{il}s_l$ , to estimate the individual animal effects,  $\hat{\mathbf{a}}$ . This was followed by a

genome wide association analysis (GWA) as described in Duarte et al.<sup>36</sup>. Briefly, the individual cSNP effects,  $\hat{g}$ , and their variances,  $\widehat{\sigma_g^2}$ , were estimated as a linear transformation of the GBLUP animal effects,  $\hat{a}$ , from equation 3. A test statistic for the association of each cSNP with phenotype was computed by standardizing the SNP effects,  $= \hat{g}/\sqrt{\widehat{\sigma_g^2}}$ , and p-values associated with this **T** test statistic calculated using the Gaussian cumulative distribution function, p-value  $= 2[1 - \Phi(|T|)]$ . Significant cSNP effects were determined after multiple test corrections using a threshold of  $FDR \leq 0.05$ .

## RESULTS

### *Identification of cSNP*

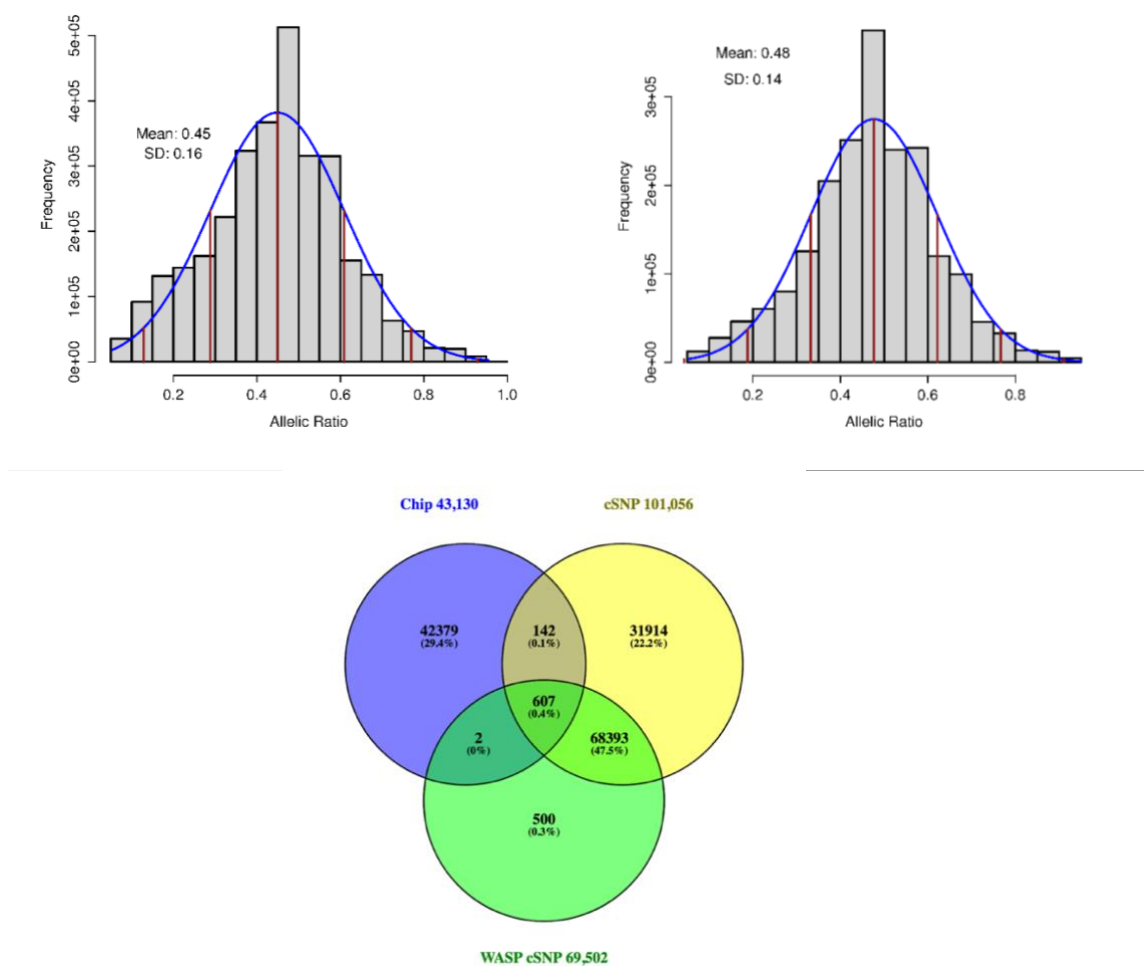
RNA sequencing of *longuissimus dorsi* muscle for 168 F2 animals generated a total of 3,606,267 identifiable polymorphic sites, less than 1% were multialleleic (5,800) and 9.2% were INDEL (313,776). The WASP algorithm corrects for bias towards the reference genome and genotyping errors when calling cSNP from RNA-seq in order to reduce bias in the estimation of allelic abundance<sup>25,26</sup>. The WASP algorithm identified  $11.3 \pm 5.7$  million reads overlapping a polymorphic site, from which 29.4% were considered biased towards the reference allele and 16.4% were duplicate reads resulting from amplification. cSNP were subsequently called after removing biased reads and quality filtered for heterozygous cSNP with sufficient coverage (10 reads) and number of heterozygous animals ( $> 6$ ), resulting in the retention of 69,502 cSNP for ASE analysis (Supplementary Table 3.S2). The allelic ration (AR) of the non-reference allele increased from  $0.45 \pm 0.16$  to  $0.48 \pm 0.14$  after applying the WASP algorithm (Figure 3.3). A comparison of overlapping cSNP and SNP ascertained with the Porcine SNP60 BeadChip for the same population identified 609 common SNP (Figure 3.3). Assuming the chip SNP as the true genotype, the sensitivity to detect a heterozygous genotype from RNA-seq was estimated as 0.99

$\pm 0.05$  and accuracy was  $0.99 \pm 0.02$ . The type 1 error rate of called heterozygous sites was 0.005.

### *Allele-specific expression*

The 69,502 cSNP were evaluated for ASE using a Quasibinomial logistic regression with overdispersion, followed by a meta-analysis to summarize gene-wise ASE. The ASE analysis identified 18,234 cSNP with significant allelic imbalance ( $\text{FDR} \leq 0.01$ , Supplementary Table 3.S2) and the meta-analysis identified 4,151 genes exhibiting cis-acting effects ( $\text{FDR} \leq 0.01$ , Supplementary Table 3.S3) from the 7,535 genes containing cSNP. On average  $10.92 \pm 12.96$  cSNP mapped per gene and the 4,151 genes exhibiting cis-acting effects contained on average  $5.20 \pm 6.97$  cSNP with ASE (Supplementary Table 3.S3). A subset of ASE cSNP (2,705) showed a narrow allelic bias falling within  $\pm 5\%$  contained within 2,705 of ASE cSNP and 176 of the genes with cis-acting effects.

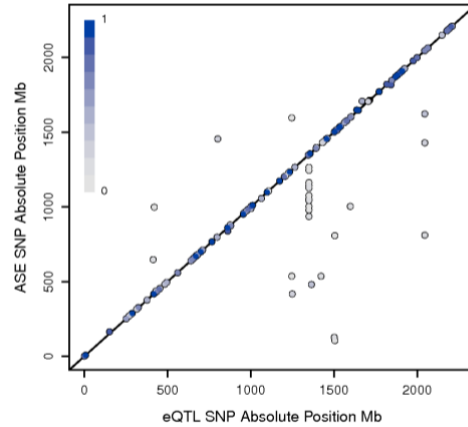
An eQTL study previously performed for the same population identified 188 local acting regulators of gene transcript abundance (Chapter 2). In this study, 91 transcripts with local eQTL contained ASE cSNP. Correlations between the most significant cSNP for an ASE gene and the peak eQTL marker indicates the extent of LD for the two candidate markers. Pearson correlations for the extent of LD were significant for 70 of the 91 genes ( $\text{FDR} \leq 0.01$ ), where correlations between the associated markers averaged  $R = 0.71 \pm 0.22$  (Figure 3.4). For the eQTL analysis, 59 genes were determined to be distant regulators residing on the same chromosome as the position of the associated gene transcript. Twenty-six of these eQTL genes were also associated with significant ASE, with 77% exhibiting significant LD between the ASE cSNP and the eQTL marker ( $R = 0.70 \pm 0.21$ , Pearson correlation  $\text{FDR} \leq 0.01$ ). Finally, 24 genes exhibiting ASE were associated with a distant eQTL (Figure 3.4).



**Figure 3.3** Number of cSNP called from RNA-seq and allelic ratios. Histograms of the allelic ratios of non-reference alleles are shown before (left) and after (right) applying the WASP algorithm. The Venn diagram illustrates comparison of called cSNP from RNA-seq before and after correcting for bias in genotype calls (yellow and green, respectively) with genotypes obtained using the Porcine SNP60 BeadChip for the 168 F2 animals.

A putative hotspot on SSC15 associated with meat quality traits was identified in the eQTL analysis (Chapter 2). Two of the genes associated with the hotspot also exhibited ASE, *PFKFB3* (AR = 0.20, 10-64777250-A-G) and *NQO1* (AR = 0.70, 6-17299064-G-T). Another gene, *OSBL1*, contained a cSNP in high LD with the putative hotspot ( $R = 0.78$ , 15-121563981-T-C), however, this cSNP did not exhibit ASE. Another *OSBL1* cSNP that did show ASE (AR=0.43,

15-121571895-C-A) mapped 234 Kb upstream of the putative hotspot and was significantly correlated with the hotspot SNP ( $R = 0.24$ ;  $p\text{-value} \leq 0.001$ ).



**Figure 3.4** Comparison of gene transcripts exhibiting significant ASE and associated with an *eQTL*. The x-axis represents the absolute position of the peak *eQTL* marker in Mb and the y-axis the absolute position of the cSNP with the most extreme allelic bias for each gene. Correlations among *eQTL* and ASE marker are color coded with a light gray color indicating low correlation, and the color intensifying to a darker blue for higher correlations. Markers aligning with the diagonal exhibit cis-acting effects and those on the off-diagonal are markers aligning to genes associated with both cis-acting and distant effects on transcript abundance.

#### *Pyrosequencing to confirm cSNP with allele-specific expression*

A total of nine cSNP exhibiting both ASE and an association with a phenotypic trait were selected for confirmation using pyrosequencing (Table 3.1). Six of these genes were confirmed to show similar allelic imbalances (Pearson correlation  $R = 0.81$ ) as was observed using RNA-seq (Figure 3.5). Four genes selected for confirmation showed higher frequency of the non-reference allele (*ZNF79*, *RNF141*, *RNF150*, and *TYW3*). Three of these genes were confirmed with pyrosequencing, *ZNF79* (RNA-seq AR=0.61, Pyrosequencing AR=0.59), *RNF141* (RNA-seq AR=0.64, Pyrosequencing AR=0.79) and *RNF150* (RNA-seq AR=0.66, Pyrosequencing AR=0.62). The AR of *TYW3* was 0.55 with RNA-seq, however, pyrosequencing of the *TYW3* cSNP indicated an AR of 0.51 for the non-reference allele, therefore not confirming ASE for this

cSNP. *TYW3* was one of the three cSNP exhibiting a narrow bias (AR of  $0.5 \pm 0.05$ ), but still considered significant in the RNA-seq ASE analysis. The other two cSNP exhibiting a narrow bias were the *NUDT3* and *NAMPT* cSNP, *NUDT3* was confirmed as exhibiting ASE with pyrosequencing (RNA-seq AR=0.47, Pyrosequencing AR=0.44), whereas *NAMPT* was not confirmed (RNA-seq AR = 0.47, Pyrosequencing AR = 0.51). While the direction of apparent allelic bias for the *PPARGC1B* cSNP was the same on both platforms (RNA-seq AR=0.30, Pyrosequencing AR=0.46), the ASE observed by RNA-seq was not confirmed by pyrosequencing.

**Table 3.1** cSNP selected for pyrosequencing confirmation.

| Phenotype           | SSC | Pos. <sup>1</sup> | Genes    | Het. <sup>2</sup> | cSNP <sup>3</sup> | AR <sup>4</sup> | PV <sup>5</sup> | q-value <sup>6</sup> |
|---------------------|-----|-------------------|----------|-------------------|-------------------|-----------------|-----------------|----------------------|
| 45-min pH           | 1   | 267.9             | ZNF79    | 32                | 9                 | 0.61            | 0.15            | 8.75E-03             |
| Drip Loss           | 2   | 49.0              | RNF141   | 59                | 9                 | 0.64            | 0.52            | 9.34E-03             |
| 10th-Rib Backfat    | 8   | 86.3              | RNF150   | 60                | 31                | 0.66            | 0.47            | 7.18E-02             |
| Protein Percent     | 15  | 25.4              | BIN1     | 75                | 99                | 0.24            | -0.41           | 2.16E-03             |
| Protein Percent     | 15  | 120.9             | PRKAG3   | 60                | 145               | 0.44            | -0.80           | 8.01E-04             |
| Carcass Length      | 7   | 30.3              | NUDT3    | 66                | 27                | 0.47*           | -0.14           | 2.41E-02             |
| 10th-Rib Backfat    | 6   | 138.4             | TYW3     | 66                | 3                 | 0.55*           | 0.13            | 6.81E-02             |
| Last-Lumbar Backfat | 6   | -                 | TYW3     | -                 | -                 | -               | 0.13            | 9.60E-03             |
| Marbling            | 6   | -                 | TYW3     | -                 | -                 | -               | 0.13            | 4.67E-02             |
| WBS                 | 9   | 106.1             | NAMPT    | 78                | 75                | 0.47*           | -0.83           | 6.36E-04             |
| Loin Muscle Area    | 2   | 150.8             | PPARGC1B | 13                | 26                | 0.30            | -0.20           | 7.99E-02             |

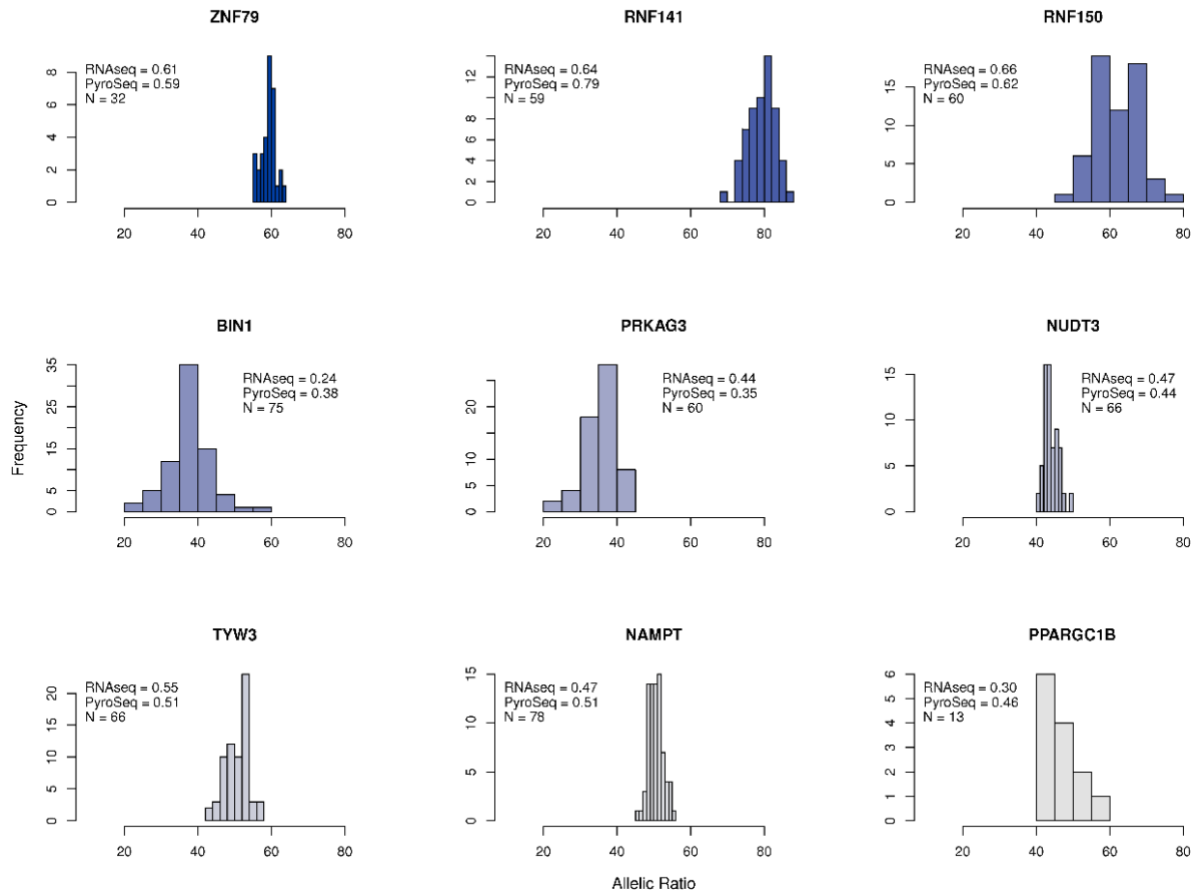
<sup>1</sup>Position of cSNP in Mb. <sup>2</sup>Number of heterozygous animals analyzed. <sup>3</sup>Number of cSNP mapped to the gene.

<sup>4</sup>Allelic ratio for cSNP. <sup>5</sup>Proportion of phenotypic variance accounted for by cSNP. <sup>6</sup>Estimated q-value for conditional analysis. \*cSNP with narrow bias, within  $0.5 \pm 0.05$ .

### Gene ontology and Kegg pathway enrichment

Genes showing significant cis-acting effects were enriched in five Kegg pathways related to energy metabolism, protein processing, focal adhesion and fatty acid degradation (FDR  $\leq 0.05$ , Table 3.2). Gene set enrichment for biological processes showed 219 enriched gene ontology (GO) terms (Table 3.3, top 12 GO terms; Supplementary Table 3.S4). Several muscle specific GO terms were enriched including terms associated with energy deprivation and

anaerobic respiration consistent with what is expected for the tissue (i.e., skeletal muscle) and time point of collection (i.e., immediately postmortem).



**Figure 3.5** Histograms of ARs obtained with pyrosequencing for nine ASE cSNP. The x-axis represents the AR of the alternative allele for the ASE cSNP, and the y-axis the frequency observed for the ratio. Displayed within the graph for each gene are the average AR of the alternative allele obtained from the two sequencing platforms (i.e. RNAseq and Pyrosequencing).

**Table 3.2** ASE genes enriched in Kegg pathways.

| <b>Kegg ID</b>               | <b>Description</b>                          | <b>Genes<sup>1</sup></b> | <b>Background<sup>2</sup></b> | <b>p-value</b> | <b>q-value</b> |
|------------------------------|---------------------------------------------|--------------------------|-------------------------------|----------------|----------------|
| hsa01200                     | Carbon metabolism                           | 46                       | 81                            | 9.05e-06       | 0.002          |
| hsa00020                     | Citrate cycle                               | 16                       | 22                            | 1.67e-04       | 0.024          |
| hsa04141                     | Protein processing in endoplasmic reticulum | 58                       | 121                           | 4.60e-04       | 0.035          |
| hsa04510                     | Focal adhesion                              | 71                       | 155                           | 6.00e-04       | 0.035          |
| hsa00071                     | Fatty acid degradation                      | 20                       | 32                            | 6.19e-04       | 0.035          |
| <i>Total number of genes</i> |                                             | <i>1409</i>              | <i>4244</i>                   | -              | -              |

<sup>1</sup>Number of genes exhibiting ASE enriched in Kegg pathway compared to background genes.

<sup>2</sup>Number of genes expressed in our skeletal muscle samples (background) connected to Kegg pathway.

### *Effects of cSNP on trait phenotypes*

We tested the effects of cSNP on phenotypic traits using two approaches. For both analyses only cSNP with less than five percent missing genotypes were considered, resulting in 28,328 cSNP with 6,293 showing significant ASE mapping to 3,352 genes. The first approach consisted of a GWAS to map phenotypic QTL using called cSNP. This cSNP-GWAS identified 108 cSNP associated with 5 phenotypic QTL for backfat, carcass length, number of ribs and protein percent ( $FDR \leq 0.05$ ; Table 3.4, Figure 3.6 and Supplementary Table 3.S5). The cSNP associated with QTL mapped to 35 gene transcripts showing significant cis-acting effects as determined by the gene-wise meta-analysis of ASE cSNP for 33 genes. A total of 33 ASE cSNP were associated with QTL for 10<sup>th</sup>-rib backfat, carcass length or protein percent.

The second approach estimated the genotypic effect of cSNP with ASE (i.e. 6,293 cSNP) on phenotypic variation by performing a gene-wise ASE conditional analysis (i.e. 3,352 genes) for all 67 trait phenotypes. This conditional analysis identified 57 cSNP associated with 25 phenotypes and 60 gene transcripts ( $FDR \leq 0.1$ ,  $p\text{-value} \leq 6.49\text{e-}05$ ; Table 3.5 and Supplementary Table 3.S6).

**Table 3.3** ASE genes enriched in GO terms for biological processes.

| GO ID <sup>1</sup>           | Description                                           | Genes <sup>2</sup> | Background <sup>3</sup> | p-value  | q-value  |
|------------------------------|-------------------------------------------------------|--------------------|-------------------------|----------|----------|
| 0003012                      | Muscle system process                                 | 145                | 291                     | 2.80E-11 | 1.30E-07 |
| 0010608                      | Posttranscriptional regulation of gene expression     | 170                | 377                     | 1.07E-08 | 8.16E-06 |
| 0006936                      | Muscle contraction                                    | 113                | 230                     | 1.23E-08 | 8.16E-06 |
| 0022613                      | Ribonucleoprotein complex biogenesis                  | 149                | 324                     | 1.86E-08 | 1.08E-05 |
| 0015980                      | Energy derivation by oxidation of organic compounds   | 90                 | 175                     | 2.46E-08 | 1.14E-05 |
| 0010927                      | Cellular component assembly involved in morphogenesis | 47                 | 76                      | 4.32E-08 | 1.83E-05 |
| 1903311                      | Regulation of mRNA metabolic process                  | 97                 | 195                     | 6.12E-08 | 2.37E-05 |
| 0009060                      | Aerobic respiration                                   | 31                 | 44                      | 1.18E-07 | 4.21E-05 |
| 0006091                      | Generation of precursor metabolites and energy        | 140                | 309                     | 1.54E-07 | 5.10E-05 |
| 0031032                      | Actomyosin structure organization                     | 69                 | 132                     | 4.72E-07 | 1.46E-04 |
| 0006099                      | Tricarboxylic acid cycle                              | 17                 | 20                      | 1.12E-06 | 3.06E-04 |
| 0042692                      | Muscle cell differentiation                           | 118                | 261                     | 1.65E-06 | 3.84E-04 |
| <i>Total number of genes</i> |                                                       | 3071               | 9762                    | -        | -        |

<sup>1</sup>Top 12 enriched GO terms are presented, for the complete list of 255 GO terms refer to Supplementary Table 3.S4

<sup>2</sup>Number of genes exhibiting ASE enriched in GO term compared to background genes.

<sup>3</sup>Number of genes expressed in our skeletal muscle samples (background) connected to GO term.

Six cSNP with ASE mapped to five genes were observed to be associated with phenotypic traits in both the cSNP GWAS and conditional analysis for carcass 10<sup>th</sup>-rib backfat (TYW3), carcass length (BRD2, DST and NUTD3) and protein percent (PRKAG3). The TYW3 gene was significantly associated with carcass 10<sup>th</sup> rib backfat, marbling scores and last lumbar backfat with the ASE cSNP exhibiting an AR of 0.55 for the non-reference allele on SSC6:138.43 Mb accounting for 13% of phenotypic variance. The cSNP SSC15:120858205-A/G mapped to the PRKAG3 gene showed an AR of 0.52 (non-reference allele) and accounted for

79% of protein percent variance. Three genes observed with both the cSNP GWAS and the gene-wise conditional analysis were significantly associated with variation in carcass length, exhibiting an AR of 0.58, 0.55 and 0.47 for BRD2, DST and NUDT3, respectively. The three cSNP mapped to a 5Mb region on SSC7 and accounted for 13, 22 and 14 percent of phenotypic variance for BRD2, DST and NUDT3, respectively.

**Table 3.4** Phenotypic QTL mapped with cSNP.

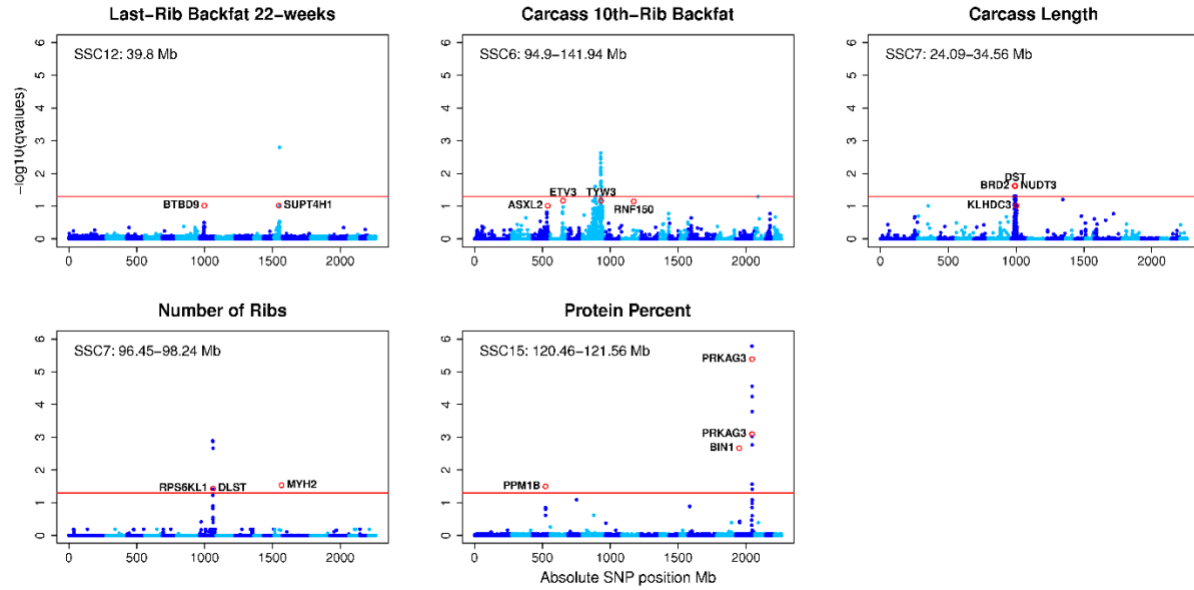
| Phenotype                     | SSC | Range Peak Mb   | cSNP <sup>1</sup> | ASE cSNP <sup>2</sup> | Genes <sup>3</sup> | Genes Meta Analysis <sup>4</sup> |
|-------------------------------|-----|-----------------|-------------------|-----------------------|--------------------|----------------------------------|
| 10 <sup>th</sup> -Rib Backfat | 6   | 94.90 - 141.94  | 30                | 11                    | 10                 | 7                                |
| Carcass Length                | 7   | 24.09 - 34.55   | 59                | 20                    | 25                 | 19                               |
| Number of Ribs                | 7   | 96.45 - 98.24   | 4                 | 0                     | 4                  | 0                                |
| Last-Rib Backfat 22-wk        | 12  | 39.80           | 1                 | 0                     | 1                  | 1                                |
| Protein %                     | 15  | 120.45 - 121.56 | 14                | 2                     | 11                 | 8                                |

<sup>1</sup>Number of associated cSNP (FDR  $\leq 0.05$ ). <sup>2</sup>Number of associated cSNP with significant ASE (FDR  $\leq 0.01$ ).

<sup>3</sup>Number of gene transcripts containing cSNP associated with QTL. <sup>5</sup>Number of gene transcripts containing cSNP associated with QTL and showing significant cis-acting effects (Meta-analysis FDR  $\leq 0.01$ ).

**Table 3.5** Gene-wise conditional analysis of ASE cSNP.

| Category                | Phenotypes | cSNP | Genes | Proportion Phenotypic Variance |
|-------------------------|------------|------|-------|--------------------------------|
| Growth Weight           | 1          | 2    | 2     | 0.11-0.27                      |
| Growth Backfat          | 5          | 8    | 10    | 0.07-0.46                      |
| Growth Loin Muscle Area | 2          | 5    | 7     | 0.09-0.43                      |
| Backfat                 | 5          | 13   | 13    | 0.10-0.79                      |
| Carcass                 | 6          | 11   | 13    | 0.10-0.51                      |
| Meat Quality            | 6          | 19   | 17    | 0.10-0.83                      |
| <i>Total</i>            | 25         | 57   | 60    | -                              |



**Figure 3.6** Manhattan plots for pQTL mapped using cSNP called directly from the *longissimus dorsi* transcriptomes of 168 animals. The x-axis represents the absolute position of each cSNP, alternating blue tones highlight each chromosome. The y-axis illustrates the negative logarithm of the calculated q-values from the GWAS. Red circles highlight cSNP for ASE genes significantly associated with a phenotypic trait; determined through a conditional analysis testing the effect of cSNP with ASE per gene on phenotypic variation ( $FDR \leq 0.1$ ).

## DISCUSSION

ASE analysis facilitates the identification of functional genomic regions regulated by cis-acting effects, and through joint association of ASE sites with phenotypic traits we can elucidate the genetic architecture of the trait. In this study, we observed 26% (18,234) of called cSNP showing significant allelic bias resulting in 55% (4,151) of genes expressed in *longissimus dorsi* muscle exhibiting allele-specific expression ( $FDR \leq 0.01$ ). A study performed in brain tissue of pigs looking at genes showing ASE found 52% of genes biased in their allelic expression<sup>16</sup> consistent with the results observed in this study. A subset of ASE cSNP (15%) did, however, show a narrow allelic bias falling within  $\pm 5\%$ . This observation had a minimal impact

on the number of genes exhibiting significant cis-acting effects because frequently additional ASE cSNP within a gene showed more extreme allelic bias.

A comparison of our previous eQTL study (Chapter 2) with the ASE analysis showed an overlap of 136 genes with associated eQTL and ASE (42% of eQTL genes). From the 188 eQTL classified as either local or plausible local, 48% (91 eQTL) showed ASE. The correlation between the peak eQTL SNP and top significant ASE cSNP corresponding to the gene for the 91 local eQTL was significant for 70 of these ( $R = 0.71 \pm 0.22$ ), suggesting the peak eQTL is in high LD with the ASE cSNP. The ASE analysis showed more precision in the identification of cis-acting effects than the genome-wide eQTL analysis, however, both approaches provide valuable information on the regulation of transcript abundance. For instance, we observed 24 genes associated with distant eQTL (trans effects) and exhibiting ASE. Two of these genes (*SUCLG2* and *NQO1*) were associated with a putative hotspot on SSC15:121.8Mb and may play a role in meat quality and carcass phenotypic diversity.

Biological processes enriched among ASE cSNP related to *SUCLG2* and *NQO1* and other genes associated with both eQTL and ASE include energy derivation by oxidation of organic compounds (*SUCLG2*, *ACO1*, *PPP1CB* and *UQCRC2*) and regulation of cellular ketone metabolic process (*NQO1* and *PSMC1*). Both of these processes are related to mitochondrial oxidative phosphorylation postmortem and ATP production for maintaining cellular homeostasis in anaerobic conditions, and have been implicated in the development of pale, soft and exudative meat<sup>41</sup>. Additional biological processes related to genes associated with eQTL and containing cSNP with ASE include cytoskeleton organization (*TBCD*, *RND3*, and *LIMK1*), muscle hypertrophy in response to stress (*CAMTA2*), ATP metabolic process (*PFKFB3*) and proteasome-mediated ubiquitin-dependent protein catabolic process (*FBXW7*).

A recent study of *longissimus dorsi* transcriptome differences between Duroc and Pietrain breeds have shown several genes differentially expressed between these breeds<sup>42</sup>. The *FBXW7* (F-box and WD repeat domain containing 7) and *SUCLG2* (succinate-CoA ligase) were reported as differentially expressed and upregulated in Duroc pigs<sup>42</sup>. The *FBXW7* is one of four subunits of an E3 ubiquitin protein ligase complex involved in the proteasomal degradation of target proteins<sup>43</sup>. Expression of one isoform for this gene (*FBXW7β*) has been implicated in muscle atrophy by upregulating MYOG (myogenin), FBXO32 (F-box protein 32) and TRIM63 (tripartite motif containing 63)<sup>44,45</sup>. In this study, the *FBXW7* gene contained two cSNP with ASE (AR = 0.43; SSC8:76637796-G/T and SSC8:76637801-A/G), and was associated with a trans-acting eQTL on SSC9:125.04 and a putative hotspot marker (ASGA0044684). Two of the muscle specific atrogenes regulated by *FBXW7* were not only expressed in our samples, but also showed ASE. The *TRIM63* gene is a muscle specific RING finger protein. This gene contained 11 cSNP with ASE and AR ranging from 0.18 to 0.63 for the non-reference allele. The *FBXO32* gene contained 42 cSNP with ASE, and AR ranging from 0.18 to 0.70 for the non-reference alleles. The high genetic diversity observed for *TRIM63* and *FBXO32*, and the different ASE effects suggest large variability in the expression of these genes, and indicate that these genes may play an important role in meat quality through proteasomal degradation of myofibrils. The *SUCLG2* gene contains a cSNP (SSC13:48824575-T/C) showing significant ASE with an AR of 0.59. This gene plays an important role in mitochondrial DNA maintenance and ATP production and has been implicated in human disorders related to muscle atrophy and infantile lactic acidosis<sup>46</sup>. While none of these ASE genes were found to be associated with meat quality phenotypes in the conditional analysis, these results suggest cis-acting, and to some degree trans-

acting, effects may regulate the expression of these genes during the conversion of muscle to meat.

A gene-wise conditional analysis aimed to estimate the effects of ASE cSNP on variation at the phenotypic level. Significant associations were observed for 25 phenotypes including growth, carcass composition and meat quality traits. Meat quality traits associated with ASE cSNP included WBS (*NAMPT*), drip loss (*RNF141*), pH at 45-min (*ZNF79* and *TOR1B*) and marbling score (*TYW3*). The *NAMPT* (nicotinamide phosphoribosyltransferase) gene plays an important role in oxidative stress and mitochondrial biogenesis and is required for the metabolic adaptation associated with calorie restriction<sup>47</sup>. In pigs this gene is highly expressed in intramuscular fat<sup>48</sup>. In this study, a cSNP mapped to *NAMPT* (SSC9:106120529-G/A) showed significant ASE with a narrow bias of 0.47 for the non-reference allele. This cSNP was significantly associated with WBS, with the non-reference allele accounting for 83% of the phenotypic variance and associated with a reduction in WBS. While this allele appears to be strongly associated with WBS, a pyrosequencing assay for this *NAMPT* cSNP did not confirm significant allelic expression bias. *NAMPT* was one of 61 ASE genes enriched in the oxidoreduction coenzyme metabolic process along with *IGF1*, *PRKAG2* and *PRKAA2*. In this study *IGF1* (insulin like growth factor 1) showed an extreme allelic bias for cSNP SSC5:81853529-G/A (AR = 0.15). *IGF1* is known for its hypertrophic activity through the activation of the phosphoinositide 3-kinase (PIK3)/Akt signaling pathway which can block mediators of skeletal muscle atrophy<sup>49</sup> such as *TRIM63* and *FBXO32*. Similarly, *PRKAA2* (protein kinase AMP-activated catalytic subunit alpha 2) has been previously associated with the PI3K/Akt signaling pathway in *longuissimus dorsi* of pigs<sup>50</sup>. The activity of these genes may regulate the rate of postmortem metabolism during the initial conversion of muscle to meat.

Drip loss is a measure of the water holding capacity of meat affected by pH decline. Four genes were significantly associated with drip loss in this study (*AMPD3*, *ITGB1*, *SDC4* and *RNF141*). The *RNF141* (ring finger protein 141) gene has previously been shown to be upregulated in Duroc pigs compared to Pietrain pigs<sup>42</sup>. In this study, the non-reference allele of the *RNF141* cSNP, SSC2-49033433-G/A, was associated with an increase in drip loss accounting for 51% of the phenotypic variance and a significant allelic imbalance was confirmed by pyrosequencing (AR=0.79). The *SDC4* (syndecan 4) gene was enriched in actin cytoskeleton organization pathway along with *NF2* (neurofibromin 2) and *OBSL1* (obscurin like 1), all showing significant cis-acting effects. Interestingly, *NF2* is a transcription factor implicated in sensing environmental stress, and increased expression of this gene activates the PI3K/Akt/mTOR pathway<sup>51</sup>. The insulin-like growth factor binding protein 2 (*IGFBP2*) on SSC15 has been previously associated with growth, carcass composition and meat quality traits in our pig population<sup>52</sup>. The *OBSL1* gene interacts with protein anchoring myosin filaments, and mutations within this gene modulate the expression of *IGFBP2* and *IGFBP5*<sup>53</sup>. In this study, the *OBSL1* cSNP, SSC15:121567503-C/G, showed significant ASE with an AR of 0.21. ASE cSNP of *OBSL1* were not directly associated with meat quality traits in the conditional analysis, however, another cSNP within *OBSL1* showed high correlation with the putative hotspot (R=0.78, 15-121563981-T-C) and this cSNP was associated with protein percent in the cSNP GWAS. One of the *OBSL1* ASE cSNP (AR=0.43, 15-121571895-C-A) was significantly correlated with the putative hotspot (R = 0.24, p-value  $\leq$  0.001). These results support *OBSL1* as a candidate gene for meat quality traits on SSC15.

In this study, five cSNP were associated with pH at 45-min postmortem. Two of these mapped to genes on SSC1 (*ZNF79* and *TOR1B*). *ZNF79* (zinc finger protein 79) is involved in

nucleic acid binding, and *TOR1B* (torsin B) is an ATPase found in the endoplasmic reticulum<sup>54</sup>. The cSNP (SSC1:267942146-G/T) for *ZNF79* accounted for 12% of the phenotypic variance for 45-min pH with the non-reference allele associated with increased pH. Significant allelic bias for the non-reference allele was confirmed with pyrosequencing (AR=0.59). The cSNP for *TOR1B* (SSC:1-269972250-G/C) is in high LD with the *ZNF79* cSNP ( $R = 0.76$ ) and showed an AR of 0.40 for the non-reference allele. *TOR1B* expression has previously been shown to be upregulated in Pietrian versus Duroc<sup>42</sup>. The enrichment analysis of genes with cis-regulation showed *TOR1B* to be involved in chaperone-mediated protein folding along with several other genes in the heat shock protein (HSP) family (*HSPH1*, *HSPB1*, *HSPB6* and *HSPA8*). Hsp70 chaperons (*HSPH1* and *HSPA8*) have been known to regulate protein folding and protein degradation via ATP dependent reaction during stressful conditions to maintain homeostasis<sup>55</sup>. *ZNF79*, *TOR1B* and the HSP genes may therefore play a role in post-mortem pH decline by maintaining protein stability.

Carcass composition traits and fatness traits associated with allelic imbalance include protein percent (*BIN1* and *PRKAG3*), loin muscle area (*PPARGC1B*) and carcass 10<sup>th</sup>-rib backfat (*TYW3*). The non-reference alleles of cSNP in *BIN1* (bridging integrator 1) and *PRKAG3* (protein kinase AMP-activated non-catalytic subunit gamma 3) on SSC15 were associated with reduced protein percent. The *PRKAG3* gene regulates glycogen potential and is associated with meat quality traits in pigs<sup>8,13,56</sup>. *BIN1* was enriched in the muscle cell differentiation pathway along with the proteases *CAPN2* (calpain 2, SSC10) and *CAPN3* (calpain 3, SSC1). The calpain system is an endogenous proteolysis system involved in protein degradation, and that plays an important role in meat tenderization<sup>41,57</sup>. *BIN1* activates a caspase-independent apoptotic process and promotes synaptic vesicle endocytosis for synaptic vesicle recycling<sup>58</sup>. Interestingly,

*CALPN3* was previously shown to be upregulated in Pietrain<sup>42</sup> *longissimus dorsi* muscle, and both *BIN1* and *CAPN3* were shown to be highly expressed in intramuscular adipose tissue<sup>48</sup>. In this study, the allelic bias of *BIN1* and *PRKAG3* was confirmed with pyrosequencing.

The *PPARGC1B* (PPARG coactivator 1 beta) gene on SSC2 contained a cSNP with ASE, with the non-reference allele associated with a reduction in loin muscle area. An important paralog of this gene is *PPARGC1A* previously suggested to play a role in energy metabolism specific to muscle fiber type, and shown to be up-regulated in *longissimus dorsi* of Duroc compared to Pietrain pigs<sup>42</sup>. In this study, *PPARGC1B* allelic expression bias was not confirmed with pyrosequencing, however, only a small number of pigs in our population were heterozygous for this cSNP.

The *TYW3* (TRNA-YW synthesizing protein 3 homolog) gene contained two cSNP with ASE showing a narrow bias of 0.55. This gene was significantly associated with 10<sup>th</sup>-rib backfat, last-lumbar backfat and marbling score accounting for 13% of phenotypic variance for all three phenotypes. Pyrosequencing of the *TYW3* cSNP, SSC6:138435089-A/G, did not confirm significant allelic bias. The *CRYZ* gene also showed significant ASE in our analysis with an AR of 0.57 (SSC6:138460416-G/A). Both *TYW3* and *CRYZ* are associated with resistin gene expression, and circulating resistin levels have been implicated in insulin resistance and obesity<sup>59</sup>. *CRYZ* has NADPH-dependent quinone reductase activity and encodes a protein that binds to adenine-uracil rich elements in 3'-UTR of mRNA, acting as a trans-acting factor<sup>60</sup>. The *CRYZ* gene was not associated with fatness traits in the conditional analysis, but five cSNP mapping to this gene (including 6-138460416-G-A) were associated with 10<sup>th</sup>-rib backfat in the cSNP pQTL analysis. These results suggest *CRYZ* and *TYW3* may play an important role in subcutaneous fatness traits through the regulation of resistin levels. Two additional genes,

*ACSL3* and *RNF150*, with allelic imbalance were associated with 10<sup>th</sup>-rib backfat. *ACSL3* (long-chain acyl-CoA synthetase 3) on SSC15 was associated with 22wk 10<sup>th</sup>-rib backfat, and this gene plays a role in mitochondrial oxidation of fatty acids<sup>42</sup>. *RNF150* (ring finger protein 150) is associated with carcass 10<sup>th</sup>-rib backfat accounting for 47% of the phenotypic variance with the non-reference allele associated with increased backfat. The allelic imbalance observed for *RNF150* was confirmed by pyrosequencing.

## CONCLUSION

This study provides new information on the complex regulation of the pig longissimus muscle transcriptome, and direct or indirect relationships with economically important phenotypic traits. Several genes identified in this study are involved in the PI3K/Akt/mTOR signaling pathway, regulating postmortem metabolism, apoptosis, calcium homeostasis, and insulin signaling. We observed several genes with ASE within this pathway suggesting a potential role for PI3K/Akt/mTOR signaling on meat quality and carcass composition traits. A high degree of overlap was observed for genes and pathways identified through the ASE analysis of our F2 Duroc x Pietrain population, and differentially expressed genes reported between the parental breeds<sup>42</sup>. These results suggest phenotypic divergence between breeds can be attributed to cis-acting effects regulating important biological processes.

## **SUPPLEMENTARY MATERIALS**

Supplementary tables available at <https://velezdeb84.wixsite.com/deborahvelezirizarry>.

**Supplementary Table 3.S1** Primer sequences for pyrosequencing array.

**Supplementary Table 3.S2** cSNP retained for ASE analysis.

**Supplementary Table 3.S3** Gene-wise meta-analysis of cSNP mapping to a gene transcript.

**Supplementary Table 3.S4** Gene ontology terms enriched for genes with significant ASE.

**Supplementary Table 3.S5** Phenotypic QTL using cSNP.

**Supplementary Table 3.S6** Gene-wise conditional analysis of cSNP with ASE.

## **REFERENCES**

## REFERENCES

1. Mora, A., Sandve, G. K. & Gabrielsen, O. S. In the loop : promoter – enhancer interactions and bioinformatics. *Brief. Bioinform.* **17**, 980–995 (2015).
2. Spitz, F. Gene regulation at a distance : From remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
3. Palstra, R.-J. *et al.* Allele-specific long-distance regulation dictates IL-32 isoform switching and mediates susceptibility to HIV-1. *Sci. Adv.* **4**, e1701729 (2018).
4. Lahbib-Mansais, Y. *et al.* Expressed alleles of imprinted IGF2, DLK1 and MEG3 colocalize in 3D-preserved nuclei of porcine fetal cells. *BMC Cell Biol.* **17**, 35 (2016).
5. Edwards, D. B. *et al.* Quantitative trait loci mapping in an F2 Duroc x Pietrain resource population: I. Growth traits. *J. Anim. Sci.* **86**, 241–253 (2008).
6. Choi, I. *et al.* Application of alternative models to identify QTL for growth traits in an F2 Duroc x Pietrain pig resource population. *BMC Genet.* **11**, 97 (2010).
7. Gualdrón Duarte, J. L. *et al.* Refining genomewide association for growth and fat deposition traits in an F 2 pig population. *J. Anim. Sci.* **94**, 1387–97 (2016).
8. Choi, I., Bates, R. O., Raney, N. E., Steibel, J. P. & Ernst, C. W. Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population. *Meat Sci.* **92**, 132–138 (2012).
9. Edwards, D. B. *et al.* Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. *J. Anim. Sci.* **86**, 254–66 (2008).
10. Bernal Rubio, Y. L. *et al.* Meta-analysis of genome-wide association from genomic prediction models. *Anim. Genet.* **47**, 36–48 (2016).
11. Sanchez, M. *et al.* A genome-wide association study of production traits in a commercial population of Large White pigs : evidence of haplotypes affecting meat quality. *Genet. Sel. Evol.* **46**, 12 (2014).
12. Nonneman, D. J. *et al.* Genome-wide association of meat quality traits and tenderness in swine. *J. Anim. Sci.* **91**, 4043–4050 (2013).
13. Casiró, S. *et al.* Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. *J. Anim. Sci.* **95**, 554–558 (2017).

14. Maroille, T. *et al.* Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* **18**, 967 (2017).
15. Yang, Y. *et al.* Transcriptome analysis revealed chimeric RNAs, single nucleotide polymorphisms and allele-specific expression in porcine prenatal skeletal muscle. *Sci. Rep.* **6**, 29039 (2016).
16. Oczkowicz, M., Szmatoła, T., Piórkowska, K. & Ropka-Molik, K. Variant calling from RNA-seq data of the brain transcriptome of pigs and its application for allele-specific expression and imprinting analysis. *Gene* **641**, 367–375 (2018).
17. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
18. Smeds, L. & Kunstner, A. ConDeTri - A content dependent read trimmer for Illumina data. *PLoS One* **6**, e26314 (2011).
19. Gordon, A. & Hannon, G. J. Fastx-toolkit. Computer program distributed by the author. (2010). at <[http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)>
20. Kim, D. *et al.* TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
21. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* **21**, 1728–1737 (2011).
22. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
23. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–93 (2011).
24. Funkhouser, S. A. *et al.* Evidence for transcriptome-wide RNA editing among *Sus scrofa* PRE-1 SINE elements. *BMC Genomics* **18**, 1–9 (2017).
25. Geijn, B. Van De, Mcvicker, G., Gilad, Y. & Pritchard, J. K. WASP : allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
26. Castel, S. E., Levy-moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195 (2015).

27. McCullagh, P. & Nelder, J. A. *Generalized Linear Model*. (Capman and Hall, 1989).
28. Consul, P. C. On some properties and applications of quasi—binomial distribution. *Commun. Stat. - Theory Methods* **19**, 477–504 (1990).
29. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
30. Maceachern, S., Muir, W. M., Crosby, S. D. & Cheng, H. H. Genome-Wide Identification and Quantification of cis- and trans-Regulated Genes Responding to Marek's Disease Virus Infection via Analysis of Allele-Specific Expression. *Front. Genet.* **2**, 113 (2011).
31. Simes, R. J. An improved Bonferroni-type procedure for multiple tests of significance. *Biometrika* **73**, 751–754 (1986).
32. Kwok, C. & Hitchins, M. P. Pyrosequencing: Methods and Protocols, Methods in Molecular Biology: Allele quantification pyrosequencing at designated SNP sites to detect allelic expression imbalance and loss-of-heterozygosity. **1315**, (Springer Science+Buisness Media New York, 2015).
33. Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: An R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).
34. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
35. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. *R Packag. version 3.6.0*. (2018).
36. Gualdrón Duarte, J. L. *et al.* Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* **15**, 246 (2014).
37. Bernal Rubio, Y. L. *et al.* Implementing meta-analysis from genome-wide association studies for pork quality traits 1. *J. Anim. Sci.* **93**, 5607–5617 (2015).
38. Ramos, A. M. *et al.* Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* **4**, e6524 (2009).
39. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
40. Wimmer, V., Albrecht, T., Auinger, H.-J. & Schön, C.-C. synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* **28**, 2086–2087 (2012).

41. England, E. M., Schef, T. L., Kasten, S. C., Matarneh, S. K. & Gerrard, D. E. Exploring the unknowns involved in the transformation of muscle to meat. *Meat Sci.* **95**, 837–843 (2013).
42. Liu, X. *et al.* Muscle transcriptional profile based on muscle fiber, mitochondrial respiratory activity, and metabolic enzymes. *Int. J. Biol. Sci.* **11**, 1348–1362 (2015).
43. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. & Pavletich, N. P. Structure of a Fbw7-Skp1-Cyclin E complex: Multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* **26**, 131–143 (2007).
44. Shin, K., Ko, Y. G., Jeong, J. & Kwon, H. Skeletal muscle atrophy is induced by Fbxw7 $\beta$  via atroгене upregulation. *Cell Biol. Int.* **41**, 213–220 (2017).
45. Shin, K. *et al.* Fbxw7 $\beta$ , E3 ubiquitin ligase, negative regulation of primary myoblast differentiation, proliferation and migration. *Anim. Sci. J.* **88**, 712–719 (2017).
46. Miller, C., Wang, L., Ostergaard, E., Dan, P. & Saada, A. The interplay between SUCLA2, SUCLG2, and mitochondrial DNA depletion. *Biochim. Biophys. Acta* **1812**, 625–629 (2011).
47. Song, J. *et al.* Nicotinamide phosphoribosyltransferase is required for the calorie restriction-mediated improvements in oxidative stress, mitochondrial biogenesis, and metabolic adaptation. *Journals Gerontol.* **69**, 44–57 (2014).
48. Sun, W. X. *et al.* Global comparison of gene expression between subcutaneous and intramuscular adipose tissue of mature Erhualian pig. *Genet. Mol. Res.* **12**, 5085–5101 (2013).
49. Glass, D. J. *PI3 Kinase Regulation of Skeletal Muscle Hypertrophy and Atrophy. Current Topics in Microbiology and Immunology* **19(5)**, (Springer, Berlin, Heidelberg, 2010).
50. Puig-Oliveras, A. *et al.* Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci. Rep.* **6**, 31803 (2016).
51. Bendavit, G., Aboukassim, T., Hilmi, K., Shah, S. & Batist, G. Nrf2 transcription factor can directly regulate mTOR: Linking cytoprotective gene expression to a major metabolic regulator that generates redox activity. *J. Biol. Chem.* **291**, 25476–25488 (2016).
52. Prasongsook, S., Choi, I., Bates, R. O., Raney, N. E. & Ernst, C. W. Association of Insulin-like growth factor binding protein 2 genotypes with growth, carcass and meat quality traits in pigs. *J. Anim. Sci. Technol.* **57**, 31 (2015).
53. Edouard, T. *et al.* OBSL1 mutations in 3-M syndrome are associated with a modulation of IGFBP2 and IGFBP5 expression levels. *Hum. Mutat.* **31**, 20–26 (2010).

54. Rose, A. E., Zhao, C., Turner, E. M., Steyer, A. M. & Schlieker, C. Arresting a Torsin ATPase reshapes the endoplasmic reticulum. *J. Biol. Chem.* **289**, 552–564 (2014).
55. Edkins, A. *CHIP: A Co-chaperone for Degradation by the Proteasome*. In *The Networking of Chaperones by Co-c.* (Springer, Cham, 2015).
56. Milan, D. *et al.* A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).
57. Lian, T., Wang, L. & Liu, Y. A new insight into the role of calpains in post-mortem meat tenderization in domestic animals: A review. *Asian-Australasian J. Anim. Sci.* **26**, 443–454 (2013).
58. Elliott, K., Ge, K., Du, W. & Prendergast, G. C. The c-Myc-interacting adaptor protein Bin1 activates a caspase-independent cell death program. *Oncogene* **19**, 4669–4684 (2000).
59. Qi, Q. *et al.* Genome-wide association analysis identifies TYW3/CRYZ and NDST4 loci associated with circulating resistin levels. *Hum. Mol. Genet.* **21**, 4774–4780 (2012).
60. Lapucci, A. *et al.* zeta-Crystallin is a bcl-2 mRNA binding protein involved in bcl-2 overexpression in T-cell acute lymphocytic leukemia. *FASEB J.* **24**, 1852–1865 (2010).

## CHAPTER FOUR

### Conclusions

The overall goal of this dissertation research is to reduce the number of candidate genes obtained through QTL mapping by identifying positional candidate eQTL associated with pQTL regions. In particular, we aimed to characterize the prevalence of local and distant acting variants of gene expression, by conducting expression QTL (eQTL) and allele specific expression (ASE) analyses using mRNA extracted from the *longuissimus dorsi* muscle of pigs from our F2 Duroc x Pietrain resource population (MSUPRP) and estimate their effect on phenotype. Transcription is a spatially and temporally controlled process regulating mRNA production, with mRNA transcripts subsequently translated into protein, the central dogma of molecular biology. Several cis-acting elements and trans-acting factors, including epigenetic markers, and environmental influences impact transcription<sup>1,2</sup>. eQTL maps reveal gene networks that can increase our knowledge of the genetic architecture of complex traits. In a well-characterized and phenotyped population like our MSUPRP, querying the co-localization of such eQTL with pQTL reveals candidate genes affecting multiple trait phenotypes. Genetic variation in the form of ASE is observed when one allele is preferentially expressed at a higher degree relative to the alternative allele, deviating from the 1:1 allelic ratio expected in biallelic expression of heterozygous locus. ASE analysis provides a means of confirming cis acting regulators, and ASE coding SNP (cSNP) associations with phenotype identify candidate markers with functional relevance.

Our eQTL scans for variants associated with total transcript abundance shed light on both local and distant regulators of gene expression. The latter include regulatory hotspots regarded as a single marker associated with variation in multiple gene transcripts. In our study, a putative hotspot on SSC15 (intergenic variant, H3GA0052416) was associated with eight meat quality

and carcass composition phenotypes, and eleven genes expressions. This genomic region being associated with variation at the transcriptional and phenotypic level (i.e. eQTL co-localized with pQTL) reveals functional variation influencing phenotypic divergence. The majority of genes associated with the putative hotspot (10) were associated with trans-acting regulation, and the other gene was a novel transcript mapped 73Mb upstream of the putative hotspot. The association of this genomic region with multiple meat quality traits has been demonstrated in GWAS performed by our group<sup>3,4</sup> and in independent studies<sup>5,6</sup>.

The *PRKAG3* (protein kinase AMP-activated gamma 3 non-catalytic subunit) gene has been implicated as the candidate gene in this genomic region<sup>7-9</sup>, however, our studies show that variants within this gene and previously implicated in regulating meat quality traits and glycolytic potential<sup>3,10,11</sup> do not account for a significant portion of phenotypic variance for meat quality traits, suggesting another gene or group of genes may be involved. Our ASE analysis identified two candidate gene in this region, the *IGFBP5* (insulin-like growth factor binding protein 5; 3Mb upstream of the putative hotspot) gene, and a modulator of *IGFBP5* expression<sup>12</sup>, *OBSL1* (obscurin like 1; 234 Kb upstream), both exhibiting significant cis-effects. Mutations identified in *OBSL1* have previously been associated with abnormal IGFBP2 and IGFBP5 expression and suggested to be a disease locus associated with heterogeneity in the 3-M growth retardation syndrome in humans<sup>12</sup>. Our findings suggest *OBSL1* as a candidate gene for the putative hotspot on SSC15 associated with meat quality traits. While ASE cSNP of *OBSL1* were not directly associated with meat quality traits in the conditional analysis, a cSNP within *OBSL1* showed high correlation with the putative hotspot ( $R=0.78$ , 15-121563981-T-C) and was associated with protein percent in the cSNP GWAS. In addition, the ASE cSNP of *OSBL1*

(AR=0.43, 15-121571895-C-A) was significantly correlated with the putative hotspot ( $R = 0.24$ ,  $p\text{-value} \leq 0.001$ ).

Insulin-like growth factor 1 (*IGF1*) is the upstream regulator of the phosphatidylinositol-3-kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR) signaling pathway, and *IGFBP5* is a strong inhibitor of *IGF1* signaling<sup>13</sup>. *MTOR* (mechanistic target of rapamycin kinase) has been shown to regulate a feedback inhibition of *IGF1* signaling through *HIF1A* (hypoxia-inducible factor 1-alpha) dependent expression of *IGFBP5*<sup>13</sup>. This is an important finding since the conversion of muscle to meat is governed by anaerobic processes that control postmortem energy metabolism, mainly the degradation of glycogen and accumulation of lactate<sup>14</sup>. Lactate accumulation in turn reduces pH, causing dysregulation of calcium homeostasis leading to increased  $\text{Ca}^{2+}$  release from the sarcoplasmic reticulum compromising mitochondrial integrity and increasing pro-apoptotic factors<sup>15</sup>. The rate of postmortem energy metabolism is the major factor influencing meat quality development, therefore, knowing *IGFBP5* and *OBSL1* exhibit significant cis-acting effects, are in close proximity to the putative hotspot on SSC15, and are important mediators of PI3K signaling, it is reasonable to assume these genes play an important role in post mortem metabolism. For instance, *HIF1A* is an important transcriptional regulator of the glycolytic pathway during hypoxic stress<sup>16–18</sup> and it is influenced by high fat diets in pigs<sup>18</sup>. *HIF1A* dependent expression of *IGFBP5* promotes *IGF1* inhibition with a feedback loop involving various genes found to exhibit cis-acting effects in our study including *IRS1*, *GRB10*, *MTOR*, *IGF1*, *IGFBP5*<sup>13</sup> and *NRF2*<sup>19</sup>. Therefore, by merging results from our eQTL, pQTL and ASE analyses we provide new insights on the complex architecture driving variation in important pig production traits.

In this study, we aimed to characterize ASE in the *longuissimus dorsi* transcriptome in pigs. Overall, 55% of expressed genes exhibited ASE, and over 50 cSNP accounted for a significant portion of phenotypic variance for growth, carcass composition and meat quality phenotypes in our MSUPRP. A 36% overlap was observed between genes exhibiting significant ASE in our study, and differentially expressed genes reported for an independent study evaluating differences in *longuissimus dorsi* transcript abundance between Duroc and Pietrain breeds<sup>20</sup>. These results suggest phenotypic divergence between breeds can be attributed to cis-acting effects regulating important production traits. Duroc breed pigs are known for their fast growth and backfat deposition, whereas Pietrain breed pigs are characterized for their leanness<sup>21</sup>. The PI3K/Akt/mTOR signaling pathway contained several genes exhibiting significant allelic imbalance with some showing extreme allelic ratios of the non-reference allele ( $< 0.20$ ; *IGF1*, *IGFBP5*, *HIF1AN*, *TRIM63* and *FBXO32*) and others exhibiting both cis and trans acting effects (*NQO1* and *PFKFB3*). PI3K/Akt/mTOR plays an important role in skeletal muscle response to acute hypoxia<sup>22</sup>, regulates cellular hypertrophy by blocking transcriptional mediators of atrophy<sup>23</sup> (i.e. *TRIM63* and *FBXO32*), and has been implicated in intramuscular fatty acid content in pork<sup>24</sup>. The transcriptional regulation of genes implicated in this pathway may explain some of the phenotypic differences observed between Pietrain and Duroc breeds. For example, both *PRKAA2* (protein kinase AMP-activated catalytic subunit alpha 2) and *PPARGC1A* (PPARG coactivator 1 alpha) genes were upregulated in Duroc *longuissimus dorsi*<sup>20</sup>. *PRKAA2* activates the PI3K/Akt pathway implicated in intramuscular fatty acid content<sup>24</sup> and *PPARGC1A* increases mitochondriogenesis via activation of AMPK that blocks mTOR<sup>25</sup>, consequently, *MTOR* gene expression was upregulated in Pietrain *longuissimus dorsi*<sup>20</sup>. *PPARGC1A* has also been implicated in fiber type conversion through increased mitochondrial respiration<sup>25</sup> consistent

with the higher number of slow oxidative fibers in Duroc breed pigs<sup>20</sup>. Both *PRKAA2* and *MTOR* genes exhibited significant ASE in our study. An important paralog of *PPARGC1A* is *PPARGC1B* found to be significantly associated with loin muscle area in our study, however, the allelic bias for this gene was not confirmed with pyrosequencing.

Candidate markers identified through eQTL and ASE analyses that are associated with phenotypic variation for economically important pig production traits or implicated in signaling pathways known to play an important role in postmortem metabolism improve our understanding of the genetic architecture of these traits. Through this study, we shed light on potential cis-acting effects for several genes implicated in the activation of the PI3K/Akt/mTOR signaling pathway in response to hypoxic stress and suggest this pathway plays a crucial role in regulating postmortem energy metabolism of the *longissimus dorsi* muscle, resulting in divergence of important phenotypic traits in pigs. The cSNP identified in this study provide valuable information on gene networks implicated in the regulation of meat quality and growth traits. Of more importance are candidate markers with ASE not found in commercial SNP arrays since they may have functional relevance for phenotypic variation and breed divergence.

## **FUTURE RESEARCH DIRECTIONS**

An application for results obtained in this study is the use of cSNP associated with growth, carcass composition and meat quality phenotypes, or implicated in influential gene networks, in SNP arrays for genomic selection or for genome-wide association studies to estimate individual SNP effects in resource and commercial populations. Targeted research on genes identified in this study may demonstrate mechanisms driving phenotypic variability and breed divergence with potential for biotechnological applications to meet breeding challenges and consumer needs. Of particular importance is the assessment of genes with ASE within the

PIK3/Akt/mTOR pathway and how variation in the expression of these genes alter phenotypic divergence. While this study suggests *IGFBP5* and *OBSL1* play an important role in postmortem metabolism and PIK3/Akt/mTOR signaling, several questions arise. For instance, how does ASE affect protein production for the key mediators of the pathway (*IGF1*, *MTOR*, *IGFBP5* and *OBSL1*)? What is the driver of ASE, is it the methylation pattern of these genes or is imprinting a contributing factor? Are other epigenetic regulators involved such as long-non coding or micro RNA (miRNA)? Is ASE influencing transcription factor binding since HIF-1 is an important transcription factor for this pathway?

Several approaches can be taken to address these questions. ELISA (enzyme-linked immunosorbent assays) assays can quantify protein expression for *IGF1*, *MTOR* and *IGFBP5* and transcription factor activity for HIF-1 in animals genotyped for the ASE cSNP and exhibiting extreme phenotypic differences in meat quality, carcass composition and/or growth phenotypes. With these assays, we can test the hypothesis that ASE alters protein production or HIF-1 transcription factor binding leading to variation at the phenotypic level. Methylation patterns can be assessed with relative ease (since we know the genes of interest) using bisulfite conversion and pyrosequencing<sup>26</sup> of genes exhibiting ASE and implicated in the PIK3/Akt/mTOR pathway to identify differentially methylated regions (DMR) and test the hypothesis that ASE is a result of DMR. Imprinting effects can be assessed within our population by genotyping the F1 generation for the ASE cSNP of interest and testing the hypothesis that ASE results from parent of origin effects. Furthermore, ASE cSNP influencing variation in genes implicated in the PIK3/Akt/mTOR pathway can be characterized across breeds and populations of pigs, in order to test the hypothesis that breed differences arise from cis-acting effects. Our group is currently characterizing miRNA expression and its influence in phenotypic divergence

using the same tissue and population that we used for this study. Several miRNA have been implicated in regulating hypoxia-inducible factors like HIF-1 via the RNA interference pathway<sup>27</sup>. A closer examination of correlations between the expression of miRNA and genes exhibiting ASE could reveal important insights into the regulation of postmortem metabolism.

The data generated through this analysis can be used to elucidate *longuissimus dorsi* transcriptome complexity and its influence on phenotypic divergence. For instance, our data has the potential to facilitate study of alternative splicing events through exon-specific expression to identify differential exon usage such as exon skipping and intron retention rates per gene. Combined with the ASE results (Chapter Three) we can gain insights on ASE induced alternative splicing and potential ASE isoforms. Similar to our eQTL analysis we can also map splice QTL to discover variants influencing alternative splicing patterns and provide deeper insights into functional and regulatory roles these variants exert on variations observed among gene expression profiles. This has been shown before in kidney renal clear cell carcinomas where a genome wide association analysis of alternative splicing patterns identified 915 cis and trans acting sQTL, some of which were previously associated with susceptibility locus for cancer<sup>28</sup>. Given that alternative splicing increases transcriptome complexity significantly, it has the potential to account for a greater amount of variability in gene expressions which can translate to variability in phenotypes. Merging eQTL, pQTL, ASE and sQTL can reveal potential insights on the genetic architecture of important phenotypes for pig production and reveal functional variants with commercial application.

In the past 30 years advancements in sequencing technology, improvements in the annotation of the pig genome, and development of quantitative genetic models has driven significant genetic gain in the pork industry. This dissertation research enhances our

understanding of the genetic architecture of pig production traits by identifying potential drivers and biological mechanisms controlling phenotypic variation.

## REFERENCES

## REFERENCES

1. Mora, A., Sandve, G. K. & Gabrielsen, O. S. In the loop : promoter – enhancer interactions and bioinformatics. *Brief. Bioinform.* **17**, 980–995 (2015).
2. Spitz, F. Gene regulation at a distance : From remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
3. Casiró, S. *et al.* Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. *J. Anim. Sci.* **95**, 554–558 (2017).
4. Bernal Rubio, Y. L. *et al.* Implementing meta-analysis from genome-wide association studies for pork quality traits 1. *J. Anim. Sci.* **93**, 5607–5617 (2015).
5. Liu, X. *et al.* Genome-wide association analyses for meat quality traits in Chinese Erhualian pigs and a Western Duroc × (Landrace × Yorkshire) commercial population. *Genet. Sel. Evol.* **47**, 44 (2015).
6. Zhang, C. *et al.* Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. *BMC Genet.* **16**, 33 (2015).
7. Ryan, M. T. *et al.* SNP variation in the promoter of the PRKAG3 gene and association with meat quality traits in pig. *BMC Genet.* **13**, 66 (2012).
8. Uimari, P. & Sironen, A. A combination of two variants in PRKAG3 is needed for a positive effect on meat quality in pigs. *BMC Genet.* **15**, 29 (2014).
9. Choi, I., Bates, R. O., Raney, N. E., Steibel, J. P. & Ernst, C. W. Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population. *Meat Sci.* **92**, 132–138 (2012).
10. Ciobanu, D. *et al.* Evidence for new alleles in the protein kinase adenosine monophosphateactivated gamma3-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. *Genetics* **159**, 1151–1162 (2001).
11. Milan, D. *et al.* A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).
12. Edouard, T. *et al.* OBSL1 mutations in 3-M syndrome are associated with a modulation of IGFBP2 and IGFBP5 expression levels. *Hum. Mutat.* **31**, 20–26 (2010).
13. Ding, M., Bruick, R. K. & Yu, Y. Secreted IGFBP5 mediates mTORC1-dependent feedback inhibition of IGF-1 signalling. *Nat. Cell Biol.* **18**, 319–327 (2016).

14. England, E. M., Schef, T. L., Kasten, S. C., Matarneh, S. K. & Gerrard, D. E. Exploring the unknowns involved in the transformation of muscle to meat. *Meat Sci.* **95**, 837–843 (2013).
15. Guo, B. *et al.* Disorder of endoplasmic reticulum calcium channel components is associated with the increased apoptotic potential in pale , soft , exudative pork. *MESC* **115**, 34–40 (2016).
16. Majmundar, A. J., Wong, W. J. & Simon, M. C. Hypoxia-Inducible Factors and the Response to Hypoxic Stress. *Mol. Cell* **40**, 294–309 (2010).
17. Hasawi, N. Al, Alkandari, M. F. & Luqmani, Y. A. Phosphofructokinase: A mediator of glycolytic flux in cancer progression. *Crit. Rev. Oncol. Hematol.* **92**, 312–321 (2014).
18. Li, Y. *et al.* Effects of dietary energy sources on early postmortem muscle metabolism of finishing pigs. *Asian-Australasian J. Anim. Sci.* **30**, 1764–1772 (2017).
19. Bendavit, G., Aboukassim, T., Hilmi, K., Shah, S. & Batist, G. Nrf2 transcription factor can directly regulate mTOR: Linking cytoprotective gene expression to a major metabolic regulator that generates redox activity. *J. Biol. Chem.* **291**, 25476–25488 (2016).
20. Liu, X. *et al.* Muscle transcriptional profile based on muscle fiber, mitochondrial respiratory activity, and metabolic enzymes. *Int. J. Biol. Sci.* **11**, 1348–1362 (2015).
21. Edwards, D. B., Tempelman, R. J. & Bates, R. O. Evaluation of Duroc- vs. Pietrain-sired pigs for growth and composition. *J. Anim. Sci.* **84**, 266–275 (2006).
22. Gan, Z. *et al.* Transcriptomic analysis identifies a role of PI3K–Akt signalling in the responses of skeletal muscle to acute hypoxia in vivo. *J. Physiol.* **595**, 5797–5813 (2017).
23. Glass, D. J. *PI3 Kinase Regulation of Skeletal Muscle Hypertrophy and Atrophy. Current Topics in Microbiology and Immunology* **19(5)**, (Springer, Berlin, Heidelberg, 2010).
24. Puig-Oliveras, A. *et al.* Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci. Rep.* **6**, 31803 (2016).
25. Eggerman, M. A. & Glass, D. J. Signaling pathways controlling skeletal muscle mass. *Crit. Rev. Biochem. Mol. Biol.* **49**, 59–68 (2014).
26. Kurdyukov, S. & Bullock, M. DNA Methylation Analysis: Choosing the Right Method. *Biology* **5**, 3 (2016).
27. Serocki, M. *et al.* miRNAs regulate the HIF switch during hypoxia: a novel therapeutic target. *Angiogenesis* **21**, 183–202 (2018).

28. Kettering, M. & Kingdom, U. Integrative genome-wide analysis of the determinants of rna splicing in kidney renal clear cell carcinoma. *Pacific Symp. Biocomput.* **2015**, 44–55 (2015).