Institut für Tierwissenschaften, Abteilung Tierzucht und Tierhaltung
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Integration of genome wide association and expression profiling for investigating water holding capacity traits in a Duroc × Pietrain resource population

Inaugural - Dissertation

zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

von

Hanna Heidt

aus

Bretten (Baden)

Referent:	Prof. Dr. Karl Schellander
Korreferent:	Prof. Dr. Brigitte Petersen
Tag der mündlichen Prüfung:	25.10.2013
Erscheinungsjahr:	2013

Dedicated to my family

Meiner Familie

Integration of genome wide association and expression profiling for investigating water holding capacity traits in a Duroc × Pietrain resource population

Water holding capacity (WHC) is an important quality criterion for the consumers and the meat processing industries. Therefore, the aim of this study was to investigate WHC traits in the Duroc × Pietrain resource population (DuPi) using genome-wide association and genetical genomics approaches.

In the first step, 169 F_2 DuPi animals were genotyped using the porcine 60K SNP chip and four meat quality traits (drip loss, pH1, pH24 and pH decline) were used to investigate the genetic background of WHC. 49, 40, 9 and 33 significant SNP were observed ($P < 0.001$) for drip loss, pH1, pH24 and pH decline in loin, respectively. Analyses revealed 14 functional candidate genes significantly associated with drip loss. 26, 7 and 22 candidate genes were identified for pH1, pH24 and pH decline, respectively. The genes NELL1 and SOX6 located on SSC2 were significantly associated with drip loss and showed more than 3 point-mutations each with high linkage disequilibrium. The proportion of explained phenotypic variance ranged between 4.4 % and 8.43 % for identified SNP of all four traits.

In the second step, WHC was characterized by drip loss measured in *M. longissimus dorsi*. Performing expression analyses of transcriptional profiles for 132 F_2 DuPi animals revealed 1228 genes, which were significantly correlated with drip loss. A hyper geometric gene set enrichment test was performed and glycolysis/glyconeogenesis, pentose phosphate pathway and pyruvat metabolism were identified as most promising pathways. For 267 selected transcripts, eQTL analyses revealed 1541 significant associations in total. Because of positional accordance of the gene underlying transcript and the eQTL location, it was possible to identify 8 eQTL that could be assumed as *cis*-regulated. Comparing the results of gene set enrichment and the eQTL detection tests, molecular networks and potential candidate genes, which seem to play key roles in the expression of WHC, were detected.

In conclusion, applying a genome wide association analysis using the 60K porcine SNP panel allowed to investigate the genetic background of WHC traits in this study. Combing the genome-wide association analysis with the genetical genomics approach supports to identify WHC trait-associated SNP and to understand the biology of complex traits.

Integration von genomweiten Assoziations- und Expressionsanalysen zur Untersuchung von Merkmalen des Wasserbindevermögens in einer Duroc × Piétrain Ressourcenpopulation

Wasserbindungsvermögen (WHC) ist ein wichtiges Qualitätskriterium für die Verbraucher und die Fleischverarbeitungsindustrie. Das Ziel dieser Studie war es daher, die Merkmale des Wasserbindungsvermögens in der Duroc × Piétrain Ressourcenpopulation (DuPi) mit Hilfe der genomweiten Assoziations- sowie Ansätzen des „Genetical Genomics“ zu untersuchen.

Im ersten Teil der Studie wurden 169 F_2 DuPi Tier mittels des 60K Schweine SNP Chips genotypisiert und 4 Fleischqualitätsmerkmale (Tropfsaft, pH1, pH24 und pH Verlauf) erfasst, um den genetischen Hintergrund des Wasserbindevermögens zu untersuchen. Für das Merkmal Tropfsaftverlust im Kotelett wurden 49 signifikante SNP ($P < 0.001$) identifiziert, für pH1 40 SNP, für pH24 9 SNP sowie für den pH Verlauf 33 SNP. Die Analyse ergab 14 funktionelle Kandidatengene, die signifikant mit Tropfsaftverlust assoziiert waren. Für pH1, pH24 und den pH Verlauf wurden jeweils 26, 7 und 22 Kandidatengene identifiziert. Die Gene NELL1 und SOX6, die signifikant mit dem Tropfsaftverlust assoziiert waren, befanden sich auf Chromosom 2 und zeigten mehr als 3 Punktmutationen, die sich in einem hohen Kopplungsungleichgewicht zueinander befanden, Der Anteil der erklärten phänotypischen Varianz durch einzelne SNP lag zwischen 4,4 und 8,43 % für alle 4 Merkmale.

Im zweiten Teil dieser Studie wurde WHC durch den Tropfsaftverlust, der im *M. longissimus dorsi* gemessen wurde, charakterisiert. Mittels einer Expressionsanalyse der Transkriptionsprofile von 132 F_2 DuPi Tieren wurden 1228 Gene gefunden, die mit dem Tropfsaftverlust signifikant korreliert waren. Ein hypergeometrischer „Gene set enrichment“ Test wurde durchgeführt und der Glykolyse/Glykoneogenese, der Pentose Phosphat Pathway sowie der Pyruvat Metabolismus als viel versprechenste Pathways identifiziert. Eine eQTL Analyse wurde mit 267 ausgewählten Transkripten durchgeführt, die insgesamt 1542 signifikante Assoziationen ergaben. Auf Grund der positionellen Übereinstimmung zwischen dem Gen, das dem Transkript zu Grunde lag, und dem eQTL, konnten 8 mögliche *cis* eQTL identifiziert werden. Durch den Vergleich der „Gene set enrichment“ und der eQTL Studie konnten molekulare Netzwerke sowie potenzielle Kandidatengene, die während der Ausprägung des WHC eine Schlüsselrolle spielen, entdeckt werden.

Schlussfolgernd lässt sich sagen, dass in dieser Studie durch die Anwendung der genomweiten Assoziationsanalyse mittels des 60K Schweine SNP Chip eine Untersuchung des genetischen Hintergrunds der WHC Merkmale möglich wurde. Durch die Kombination der genomweiten Assoziationsanalyse und dem „Genetical Genomics“ Ansatz konnten WHC Merkmal-Assoziierte SNP identifiziert und Einblicke in die Biologie der komplexen Merkmale gewonnen werden.

Contents page

List of figures

Appendix:

List of tables

Appendix:

List of abbreviations

ABC19	:	*ATP-binding cassette, sub-family A*
ABCA12	:	*ATP-binding cassette, sub-family A (ABC1), member 12*
ADAMTS3	:	*a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 3*
AHNAK	:	*AHNAK nucleoprotein*
AKR7A2	:	*aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)*
ALG14	:	*asparagine-linked glycosylation 14 homolog*
AMBP	:	*alpha-1-microglobulin/bikunin precursor*
ATP	:	adenosine triphosphate
ATP2A1	:	*ATPase, Ca^{2+} transporting, cardiac muscle, fast twitch 1*
BCMO1	:	*beta-carotene 15,15'-monooxygenase*
BP	:	biological processes
Ca^{2+}	:	calcium ion
CASR	:	*calcium-sensing receptor*
CC	:	cell components
CCDN1	:	*cyclin-dependent kinase inhibitor 2A*
cDNA	:	copy deoxyribonucleic acid
CFH	:	*complement factor H*
cM	:	centi Morgan
CMAHP	:	*cytidine monophospho-N-acetylneuraminic acid hydroxylase*

CMD	:	Congenital muscular dystony
CNV	:	copy number variants
C5orf35	:	*SET domain containing 9*
CPA6	:	*carboxypeptidase A6*
CPT1B	:	*carnitine palmitoyltransferase 1B (muscle)*
cRNA	:	Copy ribonucleic acid
CYP2A19	:	*cytochrome P450 2A19*
CYP2C	:	*cytochrome P450, family 2, subfamily C*
CYP3A	:	*cytochrome P450, family 3, subfamily A*
DNA	:	deoxyribonucleic acid
DNase	:	deoxyribonuklease
DuPi	:	Duroc × Pietrain resource population
e.g.	:	exempli gratia
ENPP1	:	*ectonucleotide pyrophosphatase/phosphodiesterase 1*
eQTL	:	expression quantitative trait loci
ER	:	endoplasmatic reticulum
eSNP	:	expression single nucleotide polymorphism
ESR1	:	*estrogen receptor 1*
ETS1	:	*v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)*
EXD1	:	*exonuclease 3'-5' domain containing 1*
F_1	:	filial generation 1

F_2	:	filial generation 2
FAM65B / C6orf 32	:	*family with sequence similarity 65, member B*
FGF3, 4, 19	:	*fibroblast growth factor 3, 4 and 19*
FDR	:	false discovery rate
FTO	:	*fat mass and obesity associated*
FUT2	:	*fucosyltransferase 2*
G6PT	:	*glucose-6-phosphate transporter*
G6P	:	*glucose-6-phosphatase*
glm	:	general linear model
GO	:	gene ontology
GWAS	:	genome-wide association study
h	:	hour
h^2	:	heritability
HD	:	high density
HGMCS1	:	*3-hydroxy-3 hethyl-glutaryl-CoA synthase 1*
HHEX	:	*hematopoietically expressed homeobox*
HSD17B14	:	*hydroxysteroid (17-beta) dehydrogenase 14*
HTR2A	:	*5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled*
HWE	:	Hardy-Weinberg equilibrium
IBK	:	Infectious bovine keratoconjunctivitis

IDE	:	*insulin-degrading enzyme*
IF	:	ichthyosis fetalis
IGF2	:	*insulin-like growth factor 2*
IGFBP5	:	*insulin-like growth factor binding protein 5*
kb	:	kilo bases
KEGG	:	Kyoto encyclopedia of genes and genomes database
KITLG	:	*KIT ligand*
KLF4	:	*Kruppel-like factor 4 (gut)*
LD	:	linkage disequilibrium
M. long. dorsi	:	*Musculus longissimus dorsi*
MAF	:	minor allele frequency
MAS	:	marker assisted selection
Mb	:	mega bases
MC4R	:	*melanocortin 4 receptor*
MCR1	:	*NADH-cytochrome b5 reductase*
MF	:	molecular functions
MFN2	:	*mitofusin 2*
MIFT	:	*Microphthalmia-assocaited transcription factor 1*
MMP13	:	*matrix metallopeptidase 13*
mRNA	:	messenger ribonucleic acid
MYC	:	*myelocytomatosis oncogene*

NBAS : *neuroblastoma amplified sequence*

N_e : effective population size

NELL1 : *NEL-like 1*

NEURL : *neuralized homolog (Drosophila)*

n.n. : not named

ORAOV1 : *oral cancer overexpressed 1*

P : parental

pH1 : pH 45 min post mortem in *M. long. dorsi*

pH 24 : pH 24 h post mortem in *M. long. dorsi*

PITX3 : *paired-like homeodomain 3*

PKC : *protein kinase C*

PLXNA2 : *plexin A2*

p.m. : post mortem

PRKAG3 : *protein kinase, AMP-activated, gamma 3 non-catalytic subunit*

PSE : pale soft and exudative

PTPRA : *protein tyrosine phosphatase, receptor type, A*

QFAM : Family-based association tests for quantitative traits

QTLdb : quantitative trait loci database

QTL : quantitative trait loci

QTN : quantitative trait nucleotide

r^2 : linkage disequilibrium

r_G	:	genetic correlation
r_P	:	phenotypic correlation
RFN	:	reddish, firm and nonexudative
RNA	:	ribonucleic acid
RNase	:	ribonuclease
ROS	:	reactive oxygen species
RSE	:	reddish soft and exudative
RYR1	:	*ryanodine receptor 1*
SALL4	:	*sal-like 4*
SD	:	standard deviation
SEMA5A	:	*sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A*
SHANK1	:	*SH3 and multiple ankyrin repeat domains 1*
SH2B1	:	*SH2B adaptor protein 1*
SLC6A5	:	*solute carrier family 6 (neurotransmitter transporter, glycine), member 5*
SLC37A4	:	*solute carrier family 37 (glucose-6-phosphate transporter), member 4*
SNP	:	single nucleotide polymorphism
SOX6	:	*SRY (sex determining region Y)-box 6*
SP-D	:	*surfactant protein D*
SRR	:	*serine racemase*

SSCD	:	Squamous cell carcinoma of the digit
SSC	:	*Sus scrofa* chromosome
STPOs	:	Standard Poodles
SULT2A1	:	*sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1*
T2D	:	type 2 diabetes
TCF3, 4	:	*transcription factor 3, 4*
TCF7L2	:	*transcription factor 7-like 2 (T-cell specific, HMG-box)*
TOX3	:	*TOX high mobility group box family member 3*
TPI1	:	*triosephosphate isomerase 1*
TXNDC12	:	*thioredoxin domain containing 12 (endoplasmic reticulum)*
TYROBP	:	*TYRO protein tyrosine kinase binding protein*
µg	:	mikro gram
VTN	:	*vitronectin*
WBSF	:	Warner Bratzler shear force
WHC	:	Water holding capacity
ZYX	:	*zyxin*

Chapter 1 General introduction

1.1 Complex traits

The term “complex trait” refers to any phenotype that does not exhibit classical Mendelian recessive or dominant inheritance associated to a single gene locus. In general, complexities occure when the simple correspondence between genotype and phenotype breaks down, either because same genotype can result in different phenotypes or different genotypes can result in the same phenotype (Figure 1) (Lander and Schork 1994). Different factors such as specific modifier genes, the genetic background, epigenetic mechanisms, stochastic effects in morphogenesis and influences of the environment as could caused this phenotypic variation (Wolf 1997). Often, it is impossible to find a genetic marker that shows perfect co-segregation with a complex trait. The reasons for this can be reduce to a few basic problems like incomplete penetrance and phenocopy, genetic heterogeneity, high frequency of causing allele and other transmission factors (Wolf 1997).

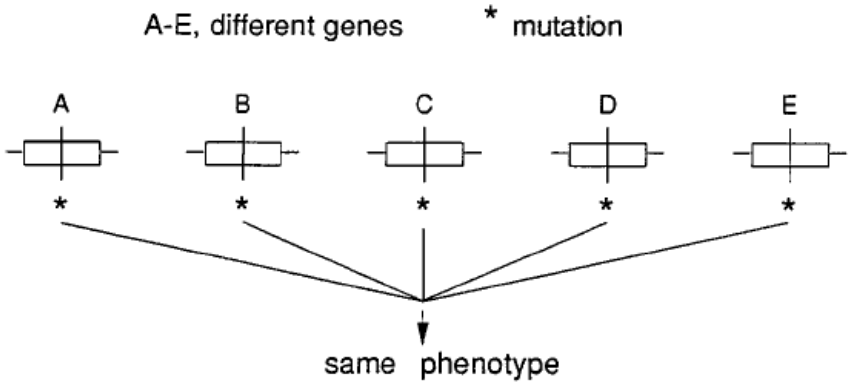

Figure 1: Complex traits: different genotypes result in the same phenotype (modified from Wolf 1997)

Many economically important traits in livestock are characterized by a complex inheritance (Andersson 2007; Andersson and Georges 2004). A chromosomal region that contains one or more genes, which influence a multifactorial (complex) trait is known as quantitative trait loci (QTL) (Andersson 2001; Mackay 2001). The challenge with complex traits lies not in detecting QTL, but in discovering the genes that underlie them (Andersson and Georges 2004). Identifying genes with varying expression linked to a variation in the QTL will help to decipher underlying processes. However, these variations do not identify

the genes controlling the trait but provide informations about the molecular background, which is involved (Rothschild 2004).

Pork is one of the most widely consumed meats worldwide and meat quality is one of the most frequently investigated complex traits in swine research (Lobjois *et al.* 2008). Many QTL for meat quality traits have already been identified in different intercrosses (Rothschild 2004). However, only a few genes, explaining the major proportion of the phenotypic variance, have been already identified for meat quality such as malignant hyperthermia (*ryanodine receptor 1*, RYR1) (Fujii et al. 1991) and glycogen content in skeletal muscle (*protein kinase, AMP-activated, gamma 3 non-catalytic subunit*, PRKAG3) (Milan *et al.* 2000).

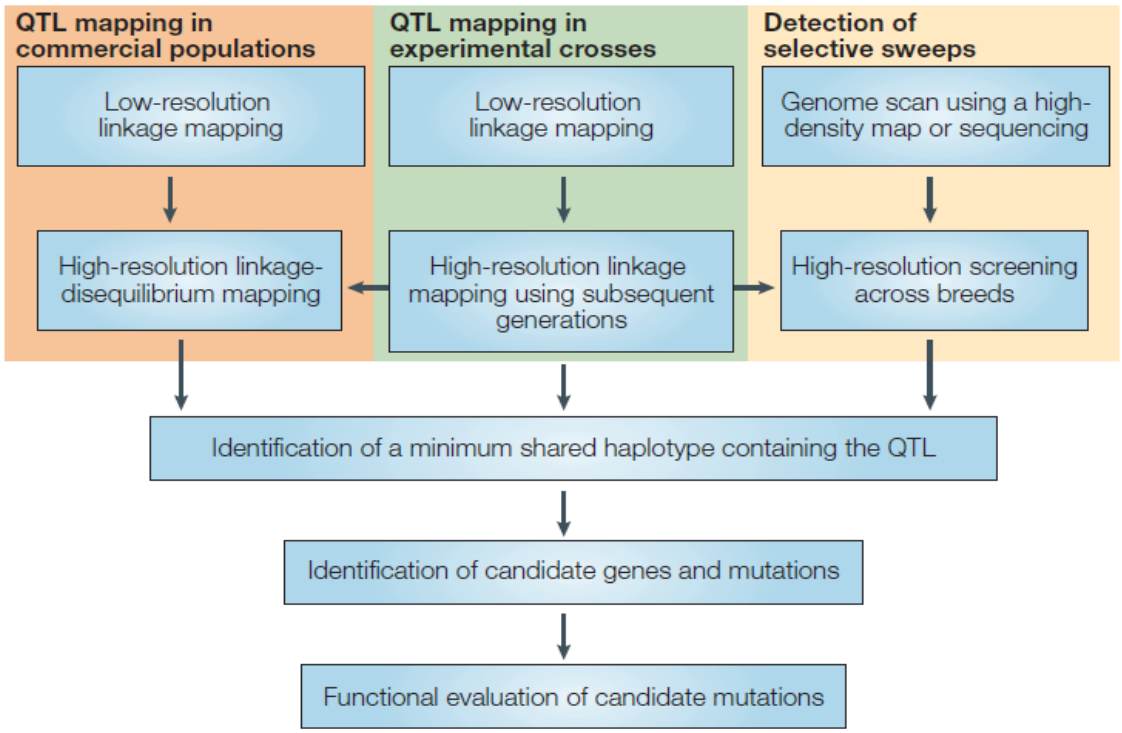

Figure 2: Approaches for mapping and positional cloning of QTL in domestic animals (Andersson and Georges 2004).

The identification of genes and mutations that underlie the QTL is problematic for several reasons. First, it remains difficult to determine the exact chromosomal location of a QTL. Second, most QTL have a small phenotypic effect, so the mutations that cause them are difficult to distinguish from neutral polymorphisms. Another factor that complicates the

identification of QTL mutations is that a good proportion of these mutations are regulatory. And the ability to evaluate functionally important mutations in non coding regions is poorly developed (Andersson and Georges 2004). So, it is clear that decoding the genetic basis of complex traits presents a analytical challenge. For this reasons, the system in which we have more power to detect QTL and the mutation that underlie them are of particular interest (Andersson and Georges 2004). Because of the less genetic heterogeneity within breeds due to the limited population size domestic animals are such a system (Andersson 2001; Nezer *et al.* 2003). The availibility of dense marker maps such as single nucleotide polymorphism (SNP) chips open up the possibility for a new approach for QTL detection and allow potential of using domestic animals for decoding the genetic basis of complex traits (Figure 2) (Andersson and Georges 2004).

1.2 Meat quality traits

1.2.1 Muscle composition and post mortem conversation into meat

Muscle is composed of approximately 75 % water. The other main components include protein (nearly 20 %), lipids (approximately 5 %), carbohydrates (nearly 1 %) and vitamins as well as minerals (round about 1 %). In fact, nearly 85 % of the water in muscle is held within the myofibrils and the cell membrane (sarcolemma) and between the muscle cells and muscle bundles (Huff-Lonergan and Lonergan 2005; Offer and Cousins 1992; Offer and Knight 1988).

During the post mortem conversion of muscle to meat, many changes occur, including:

1. A gradual depletion of available energy
2. A shift from aerobic to anaerobic metabolism favouring the production of lactic acid, resulting in the pH of the tissue declining from near neutrality to 5.4 -5.8
3. A rise in the ionic strength, because of the inability of ATP dependent calcium, sodium, and potassium pumps to function
4. An increasing inability of the cell to maintain reducing conditions (Huff-Lonergan and Lonergan 1999; Huff-Lonergan and Lonergan 2005; Huff-Lonergan et al. 1996).

Once pH has reached the isoelectric point, positive and negative electrical charge on the proteins are equal. These positive and negative groups within the protein attract each other

and cause a space reduction within the myofibrils (Huff-Lonergan and Lonergan 2005). Additionally, rapid pH decline is leading to an ultimate pH (pH at 24 h) while muscle is still warm. This results in denaturation of many proteins including those involved in binding of cellular water (Huff-Lonergan and Lonergan 2005). This it is accompanied by leakage of muscle cells and loss of water, ions and proteins (Greaser 2001; Offer and Knight 1988). There is also a variation of fluid released from the muscle during conversion of muscle to meat, which are caused by various environmental effects for example stress prior to slaughtering and different genotypes (Greaser 2001).

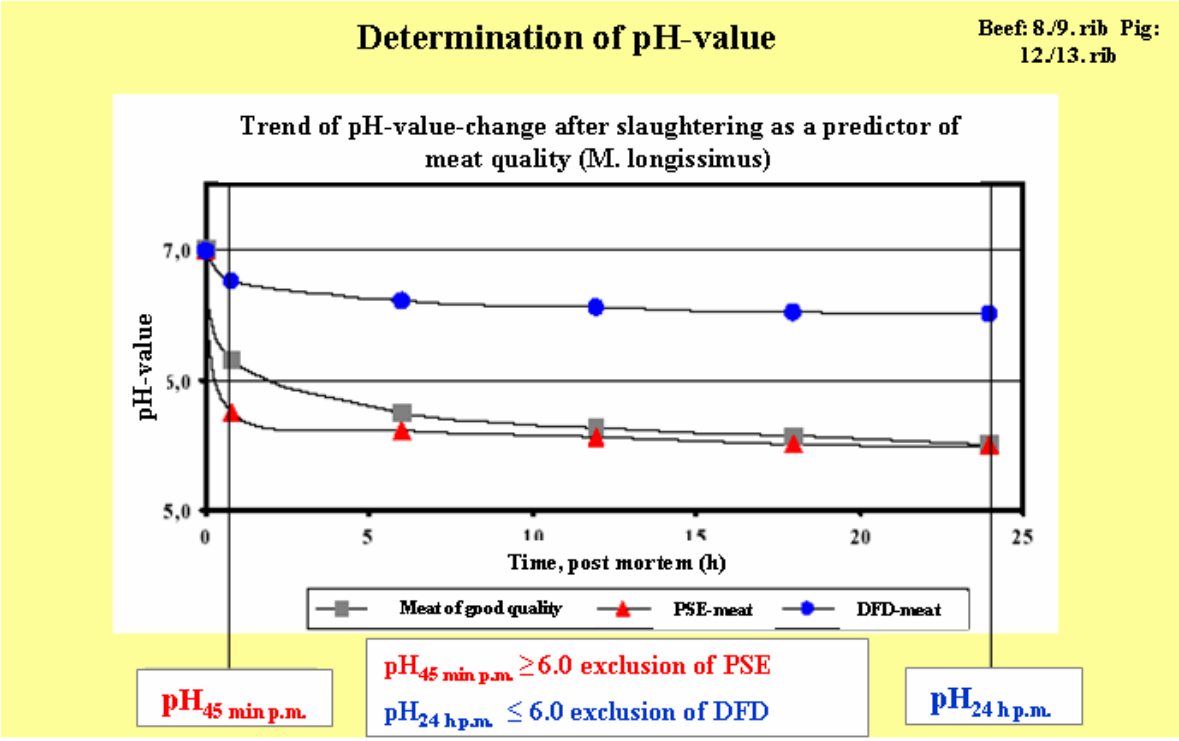

Figure 3: Determination of post mortem pH-value

1.2.2 The role of water holding capacity in meat quality

Water holding capacity (WHC) develops as a result of early post-mortem biochemical and biophysical processes that occur in muscle (Huff-Lonergan and Lonergan 2005). Therefore, WHC is the capacity of meat to retain its water during application of external forces (Hamm 1985), which is an important quality criterion for the meat processing industries and the consumers. WHC affects the financial output, nutritional value, sensorial and technological properties of porcine meat. WHC can be predicted by measuring drip

loss using gravitational techniques. Drip loss is a fluid consisting of water and protein expelled from the meat surface without any mechanical force other than gravity (Offer and Knight 1988). The highest drip loss is often found in pale, soft and exudative (PSE) meat from pigs that have inherited a mutation on the ryanodine receptor / calcium release channel (RYR1) gene (Fujii et al. 1991). With the help of a commercial test for this mutation, the German pig production has mostly eliminated this mutation in pig populations used for fattening (Tholen et al. 2005; ZDS 2006). However there are other factors that cause PSE meat for example the short term stress before slaughter, which causes a rapid pH decline, protein denaturation and higher drip loss (Rosenvold and Andersen 2003). Another observed meat quality is reddish, soft and exudative (RSE) meat, which is an untypical deviation and probably a mild form occurrence of PSE (Fischer 2007; Kauffman et al. 1993). RSE reveals the colour of red, firm and nonexudative (RFN) pork but the exudation of PSE meat (van Laack and Kauffman 1999). The risk for RSE is increased by a high glycolytic potential as well as by a low ultimate pH value (van Laack and Kauffman 1999).

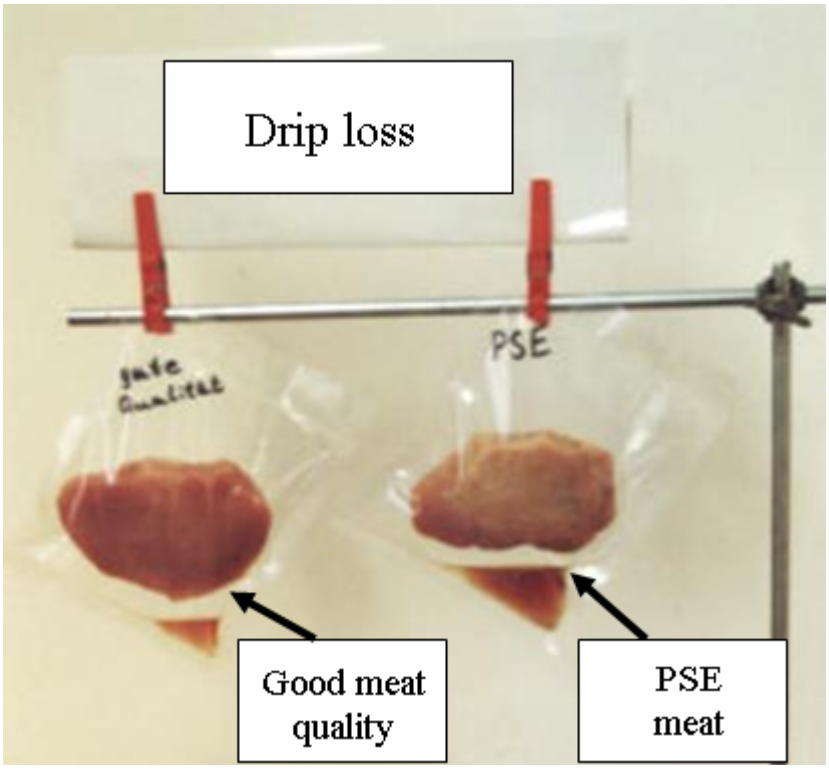

Figure 4: Drip loss shown in different meat qualities. Good meat quality: low drip loss; bad meat quality (pale, soft, exudative (PSE) meat): high drip loss

1.2.3 The genetic basis of water holding capacity

Drip loss is genetically and phenotypically correlated with early pH (pH1) (r_G = -0.91 and r_P = -0.67, respectively) and late pH (pH24) (r_G = -0.72 and r_P = -0.51, respectively) in loin (Borchers *et al.* 2007). Heritability estimates for water holding capacity traits are low: pH1 (h^2 = 0.14), pH24 (h^2 = 0.20) and drip loss (h^2 = 0.14) (Borchers *et al.* 2007). The presented heritabilities and correlation estimations are an indication of a polygenic background that is possibly involved in the expression of the examined traits. The identification of genes that regulate meat quality traits will assist in efficient meat production and facilitate the resolution of existing production problems or so-called marker-assisted selection (MAS) (Brunner *et al.* 2012). Both candidate gene and QTL mapping strategies have been used in domestic animals for the discovery of genetic markers suitable for MAS (Rothschild *et al.* 2007).

Until now, 6397 QTL covering 578 phenotypic traits were identified and submitted into pig QTLdb (www.animalgenome.org). In total, 1072 QTL were identified for drip loss, pH1, pH24 and pH decline in loin in different pig populations (www.animalgenome.org, release 11, June 2013) (Hu *et al.* 2010). These QTL were mostly identified via genome scanning based on linkage analysis and microsatellite genotyping.

In order to identify chromosomal regions, which were influenced by drip loss the quantitative trait loci (QTL) approach was used in many different studies. Liu et al. (2007) found 24 significant and 47 suggestive QTL for meat quality traits and carcass composition traits in a F_2 Duroc × Pietrain resource population. For drip loss several QTL on SSC2, SSC3, SSC5 and SSC18 were identified. In the study of Edwards et al. (2008), 94 QTL regions for meat quality were observed in another F_2 Duroc × Pietrain resource population, but only a QTL on SSC9 was detected for drip loss. The QTL on SSC2 and SSC5 (Liu *et al.* 2007) as well as the QTL on SSC9 (Edwards *et al.* 2008) were identified in the study of Thomsen et al. (2004), where a Berkshire × Yorkshire population was used. Consequently, by using low density of microsatellite markers, QTL are often mapped to a large interval of 20 centimorgans (cM) or more. Only a few causative mutations, which are called quantitative trait nucleotides (QTN), have been identified based on results of complex traits in pigs via QTL fine mapping analysis (Ren et al. 2011; Van Laere et al. 2003). Applying a genome wide association analyses by using the current 60K porcine SNP panel provides more dense genotypic markers than microsatellite markers which helps to

improve accuracy in finding the exact QTL locations and candidate genes for complex traits (Luo *et al.* 2012b).

1.3 Genome-wide association studies (GWAS)

Genome wide association studies (GWAS) are defined as studies in which a dense array of genetic markers, which capture a substantial proportion of common variation in genome sequence, is typed in a set of DNA samples that are informative for a trait of interest. The aim is to map the effects for a trait of interest through the detection of association between genotype frequency and trait status (McCarthy *et al.* 2008). For genotyping commercial "SNP chips", exist for cattle (777,000 SNP; Illumina Bovine HD BeadChip), dogs (170,000 SNP, Illumina Canine HD BeadChip), sheep (56,000 SNP, Illumina Ovine 50K BeadChip), horse (54,602 SNP, Illumina Equine 50K BeadChip) and pig (64,232 SNP, Illumina Porcine 60K BeadChip). For chicken a 600K SNP chip (580,961SNP, Axiom® Genome-Wide Chicken Genotyping Array, Affymetrix) is designed. Nevertheless, the chips for domestic animals contain less SNP than the Human SNP chip with more than 1 million SNP (www.illumina.com, Illumina, Inc., San Diego, USA, www.affymetrix.com, Affymetrix, Santa Clara, USA).

The basic design of a GWAS is that a sample of individuals are recorded for a trait of interest and characterized for a genome-wide panel of markers in order to detect statistical associations between the trait and any of the markers. The design parameters include the choice and number of individuals and markers. Most commonly, the GWAS data are analyzed by the examination of one single nucleotide polymorphism (SNP) at a time using simple linear models. The genomic sequence is available for human, mouse, rat and several domestic species, which is a requirement to establish SNP chips (Goddard and Hayes 2009). A typical GWAS could be divided in 5 parts.

1. At first, the number of individuals used in the study must be determined because the number of individuals applied for GWAS depends on the size of the effects that one wishes to detect. The crucial parameter is the proportion of the variance explained by the SNP. This parameter combines the allele frequency with the mean difference between the SNP genotypes (Goddard and Hayes 2009). Therefore, the number of individuals depends on the number for which both genotypes and phenotypes have

been directly measured. In domestic animals studies, the number can be reduced by using animals that have been progeny tested so that the mean of the progeny can be used instead of their own phenotypic value (Goddard and Hayes 2009).

2. Secondly, the number of SNP, which have to be analyzed, should be calculated. Due to the fact that the number of SNP applied, depends on the distance over which linkage disequilibrium (LD) exists. Which means the non-random allocation of alleles at nearby variants to individual chromosome as a result of recent mutation, genetic drift or selection, manifest as correlations between genotypes at closely linked markers (Goddard and Hayes 2009; McCarthy et al. 2008). If the SNP are too far apart from each other, a QTL may not be in an acceptable LD with the markers and could not be detected. Therefore, increasing the SNP density will increase the power to detect QTL (Goddard and Hayes 2009). However, the distribution of the marker effects shows that most SNP have small effects (contribute random noise), whereas markers in regions in which the causative mutations lies have much larger effects. Although the markers with the largest associated effects from a genomic analysis may not perfectly track the causative mutation, they are potentially useful tools to identify the chromosomal region (Cole *et al.* 2009).

3. Subsequently the source of bias in livestock studies has to be considered. An important source of false positive associations is admixture in the sample of individuals used, which means that a population or sample of individuals derived from more than one breed and that have not undergone random mating (Goddard and Hayes 2009). One of the biggest problem would be if a sample consists of a mixture of breeds. Another form of admixture is the relationships among the animals, because livestock are usually bred in half-sib families (cattle) or full-sib families (pig). Therefore, relationships among animals in the sample cause LD between loci even if they are unlinked (Goddard and Hayes 2009).

4. Afterwards the design of the GWA study has to be chosen. There are five different approaches (case-control, cohort, trio, family-based association and DNA pooling), which can be used for GWAS (Table1) (McCarthy et al. 2008; Pearson and Manolio 2008). The most frequently used study design is the case-control design, in which allele frequencies of individuals with the trait of interest are compared to those in a comparison group. These studies are often easier and less expensive to conduct than studies using other designs such as a family-based association study. Therefore, many

studies use multistage designs to reduce the number of false positives while minimizing the number of genome-wide scans in order to keep the statistical power (Hirschhorn and Daly 2005).

5. After selection of individuals with the trait of interest and a suitable comparison group, the DNA isolation is performed in order to genotype theses individuals. Subsequently a quality control of the genotyping data has to be performed to ensure high quality of the data. Therefore, association between SNP, which passed the quality thresholds and the trait of interests, has to be calculated using GWAS approaches. At least a replication of identified associations in an independent population or examination of functional background should be performed (Pearson and Manolio 2008).

Table 1: Study designs used in GWA studies (modified from McCarthy et al. 2008; Pearson and Manolio 2008)

	Case- Control	Cohort	Trio	Family-based association	DNA pooling
Assumptions	▪ Case and control are drawn from the same population ▪ Genomic and phenotypic data are collected similarly in cases and controls ▪ Cases are representatives of all cases of the trait / disease	▪ Individuals under study are more representative of the population from which they are drawn	▪ Disease-related design alleles are transmitted in excess of 50% to affect offspring from heterozygous parents	▪ Association testing is performed within families ▪ Protection from population substructure effects	▪ Using estimates of allele frequencies derived from pools of DNA ▪ DNA pools are compiled from multiple subjects than individual DNA samples
Advantages	▪ Short time frame ▪ Large number of cases and controls can be assembled ▪ Optimal epidemiologic design for studying rare diseases	▪ Cases are incident and free of survival bias ▪ Direct measure of risk ▪ Fewer bias than case-control studies	▪ Controls for population structure ▪ Immune to population stratification ▪ Allows check for Mendelian inheritance	▪ Robust strategy for dealing with population stratification ▪ Correction for population substructure effects ▪ Restriction of high-density scanning to a reduced subset of pedigree members	▪ More economical genome-wide surveys for association
Disadvantage	▪ Prone to number of biases including population stratification ▪ Overestimation of the relative risk for common diseases/ traits	▪ Large sample size are needed for genotyping if incidence is low ▪ Expensive ▪ Poorly suited for studying rare diseases	▪ Maybe difficult to assemble both parents and offspring ▪ Highly sensitive to genotyping error	▪ Reduced power	▪ Reduced power ▪ Loss of individual genotype data ▪ Difficulties ensuring equimolar representation of samples

1.3.1 GWAS performed for Mendelian traits

Spontaneous mutants in domestic animals provided insights into genotype-phenotype correlations, which are relevant for biomedical research (Patterson et al. 1988), because of the strong phenotypic selection or specific behavioural and morphological traits. One consequence of the breeding programs used to propagate lineages with such strong phenotypic homogeneity is the increased incidence of disease (Graw 2003; Verma and FitzPatrick 2007). In dogs, Karlsson et al. (2007) observed the mutations in the gene *Microphthalmia-associated transcription factor 1* (MITF) causing white spotting using nine solid Boxers and ten white Boxers. Another recent study is the identification of the dominant mutation causing the hair ridge in Rhodesian and Thai Ridgeback dogs (Andersson 2009; Salmon Hillbertz et al. 2007). The mutation is a 133 kb duplication that includes three *fibroblast growth factor* genes (FGF3, FGF4 and FGF19), *oral cancer over expressed 1* (ORAOV1) gene and 3' part of CCDND1 observed by GWAS using 10 cases and 10 controls (Salmon Hillbertz *et al.* 2007). In the study of Karyadi et al. (2013) a copy number variant (CNV) at the *KIT ligand* (KITLG) locus was observed causing squamous cell carcinoma of the digit (SCCD) in Standard Poodles (STPOs) using GWAS with 31 cases and 34 controls. In a second GWAS, they compared 24 black and 24 light coloured STPOs, Karyadi et al.(2013) suggested that a compensatory mutation within the *melanocortin 1 receptor* (MC1R) locus likely protects lighted coloured STPOs from disease. These findings highlight how studies of breed-limited diseases are useful for disentangling multigene disorders. Additionally, dogs are diagnosed with nearly all of the same cancers in human (Merlo *et al.* 2008), and the underlying pathology and treatment response is typically the same as for human (Dorn 1976), suggesting that canine cancer genetic studies are a useful way to advance the understandings in human disease (Cadieu and Ostrander 2007; Karyadi et al. 2013; Khanna et al. 2006; Ostrander 2012).

1.3.1.1 GWAS performed in livestock animals for Mendelian traits

Strong inbreeding in the livestock population increased the risk of the occurrence of genetic diseases. The most common mode of transmission of genetic defects is monogenetic autosomal recessive inheritance. Progenies with recessive defects are typically the consequence of inbreeding. Recessive genetic diseases become apparent many years after the initial mutation event (Charlier et al. 2008; Drögemüller et al. 2011).

In cattle, most of the known recessive defects became apparent 5 to 10 generations after the founder animal, which corresponds to the time when female and male descendants of the original carrier are mated. During the latent phase, the deleterious allele might have been widely spread throughout the population explaining sudden outbreaks with many affected animals appearing simultaneously (Charlier et al. 2008; Drögemüller et al. 2011). Charlier et al. (2008) identified the causal mutations and the molecular basis for congenital muscular dystony (CMD) type 1 and 2 in Belgian Blue cattle and ichthyosis fetalis (IF) in Italian Chianina cattle using GWAS. To map the gene *ATPase, Ca^{2+} transporting, cardiac muscle, fast twitch 1* (ATP2A1) causing CMD1 they used 12cases and 14 controls of Belgian Blue cattle. The gene *Solute carrier family 6, member 5* (SLC6A5) was identified in 7 cases compared with 24 controls of Belgian Blue cattle causing CMD2 (Charlier et al. 2008). *ATP-binding cassette, sub-family A (*ABC19*), member 12* (ABCA12) was observed in 3 cases and 9 controls of Italian Chianina cattle determining IF (Charlier et al. 2008). In the study of Drögemüller et al. (2011) an unusual splicing defect in the *Mitofusin 2* (MFN2) gene is significantly associated with bovine progressive degenerative myeloencephalopathy (weaver syndrome) in Tyrolean Grey Cattle.

In Texel sheep a missense mutation in the gene *Paired-like homeodomain 3* (PITX3) was associated with Microphthalmia using 23 cases and 23 controls in GWAS (Becker *et al.* 2010).

In pig, there are very few studies dealing with hereditary defects. However, most of these traits have a polygenetic background, for example the inverted teat defect with several QTL and their positional candidate genes on SCC3, 4, 6 and 11 (Jonas et al. 2008). Therefore, the identification of the causal mutation is very difficult.

These studies were particularly efficient because of the low N_e in livestock. If animals are suffering from a fatal, recessive disorder and are homozygous for a large chromosomal segment containing the causative gene, allows the detection of the causative gene using lower number of animals and moderately dense markers. For complex traits much larger numbers of animals is needed to detect the causative genes (Goddard and Hayes 2009).

1.3.2 Investigating quantitative traits using GWAS

For quantitative traits the results generally indicate many mutations, suggesting that each individual mutation has a small effect (Goddard and Hayes 2009). In human, Levy et al.

(2007) identified in the Framingham Heart Study 100K Project with 4464 people 7 SNP significantly associated with blood-pressure and 5 SNP associated with arterial-stiffness. In the study of Hazra et al. (2008), 3 SNP of the gene *fucosyltransferase 2* (FUT2) were significantly associated with the plasma vitamin B_{12} level in 1658 women and 1059 independent replications from the Nurses' Health Study. Measures of obesity, weight and body mass index were significantly associated with 29 variants including these close to the *fat mass and obesity associated* (FTO), *melanocortin 4 receptor* (MC4R), *brain-derived neurotrophic factor* (BDNF) and *SH2B adaptor protein 1* (SH2B1) genes in 25342 Icelanders (Thorleifsson *et al.* 2009). O'Seaghdha et al. (2010) observed common variants in the *calcium-sensing receptor* (CASR) gene in a large meta-analysis of GWAS of serum calcium levels, including data from 20611 individuals of European ancestry.30 SNP were associated with serum uric acid in island population of the Adriatic coasts of Croatia (Karns *et al.* 2012). Ali et al.(2013) identified five candidate genes (TCF7L2, HHEX, IDE, ENPP1 and FTO) significantly associated with type 2 diabetes (T2D) in three unrelated Indian populations. Additionally, Pueyo et al. (2013) observed genetic variants in the *surfactant protein-D* (SP-D), which were also significantly associated with insulin resistance and T2D.

1.3.2.1 Investigating quantitative traits in livestock using GWAS

In livestock, Kolbehdari et al. (2008) identified 196 significant SNP affecting conformation and functional traits in Canadian Holstein dairy cattle. In the same population, 144 SNP were associated with production, functional and reproduction traits (Daetwyler *et al.* 2008). Lillehammer et al. (2009) found significant SNP for a genotype × environment interaction for milk yield at the level of herd production in Australian Holstein dairy cattle. In Angus cattle several regions on the chromosomes 2, 12, 13 and 21 were identified causing infectious bovine keratoconjunctivitis (IBK) (Kizilkaya *et al.* 2013).

A GWAS in chickens reported 21 SNP related to 19 genes, which were significantly associated with resistance to *Salmonella enterica* colonization (Goddard and Hayes 2009; Hasenstein et al. 2008). In another study in chickens several SNP were identified, which were associated with the resistance to *Campylobacter jejuni* (Connell et al. 2013).

In pig, several GWA studies were performed, conducted for quantitative traits in pigs such as meat quality (Duijvesteijn et al. 2010; Luo et al. 2012b; Ponsuksili et al. 2011; Ramos et al. 2011), reproduction (Onteru et al. 2011; Uimari et al. 2011), immune traits (Luo et al. 2012a; Wang et al. 2012), growth with soundness (Fan et al. 2011) and feed conversion ratio (FCR) (Sahana et al. 2013). For example, in the study of Duijvesteijn et al. (2010) 987 pigs divergent for androstenone levels in fat tissue in a commercial Duroc-based sire line were genotyped. The association analysis revealed 37 SNP on *sus scrofa* chromosome (SSC) 1 and 6 significantly associated with androstenone. Among them, the 5 most significant SNP explained together 13.7% of the genetic variance in androstenone. On SSC6, a larger region of 10 Mb was shown to be associated with androstenone covering several candidate genes (*cytochrome P450 A19* (CYP2A19), *sulfotransferases* SULT2A1 and SULT2B and *hydroxysteroid-dehydrogenases* (*HSD17B14*)) potentially involved in the synthesis and metabolism of androgens (Duijvesteijn et al. 2010).

However, the distribution of marker effects showed that most SNP had small effects, which could be described as random noise, whereas markers in regions harbouring genes with causative mutations had much larger effects. Although the markers with the largest associated effects from genomic analysis might not track perfectly the causative mutations, they were potentially useful tools for identifying chromosomal regions (Cole *et al.* 2009). Therefore, fine mapping techniques such as RNA sequencing are needed for further investigating the possible candidate genes.

1.4 Genetical Genomics

A genome scan is the most general approach to identify genomic regions showing quantitative trait loci (QTL). Such QTL regions are generally large and can contain thousands of genes. Most of them are candidate loci for the trait (Wayne and McIntyre 2002). Quantitative expression studies such as microarray technology, can indicate regulatory variation in genes for complex traits (Wayne and McIntyre 2002). By combining QTL mapping and microarray analyses, it is possible to identify regulatory networks underlying the quantitative trait of interest and localize genomic variation, the so-called genetical genomics approach (Jansen 2003; Jansen and Nap 2001). Furthermore, QTL analyses of expression levels of gene identify genomic regions, which are likely to

contain at least one causal gene with the regulatory effect on the expression level, termed expression QTL (eQTL). The use of eQTL analyses has been demonstrated as a promising tool for narrowing the gap between detected phenotypic QTL regions and confirmed causative variations for the pig species (Rothschild et al. 2007; Steibel et al. 2011). Detected eQTL can be classified into a locus, which is located close to a gene (*cis*-regulation) or distant from the gene (*trans* regulation) (Jansen and Nap 2001). Differentially expressed genes, where eQTL mapping indicates *cis*-regulation, are more likely to represent the "cause", for example the genetic background of the trait of interest. Whereas differential expressed genes revealing *trans*-regulation represents the "effect", for example pathways that are affected by causal variation (Wimmers et al. 2010). The largest and most significant reported eQTL are often *cis*-regulated. However, some *trans*-regulated eQTL seem to control expression in several or many genes spread across the genome. These findings provide general information on the organization of the control of expression (Haley and de Koning 2006).

The key advantage of eQTL mapping is that it connects variation at the level of RNA expression to vary at the level of DNA. Only latter provides versatile tools for breeding whereas the first reveals information on the biology of a trait and directs to new candidate genes. In summary, integration of information on QTL for a trait of interest in breeding with analyses of trait correlated expression and with mapping of eQTL for the corresponding trait-dependent-regulated genes facilitates the identification of genes and pathways with cumulative evidence of their involvement in the biology of the traits of interest and enable to built priority lists of candidate genes (Ponsuksili *et al.* 2010; Wimmers *et al.* 2010)

However, there are also some issues that limit of the use of genetical genomics approach, in particular the resolution of the genetic maps that is depending on the number of markers and animals used, the structure of the population used and artefacts caused by the limited sensitivity and specificity of microarray experiments (Ponsuksili *et al.* 2010; Verdugo *et al.* 2010).

1.4.1 Genetical genomic approaches applied to complex traits in human

A number of eQTL studies have been published to date on a variety of tissues and cells (Kabakchiev and Silverberg 2013). Some of these include monocytes (Zeller *et al.* 2010),

liver tissue (Schadt *et al.* 2008) and brain tissue (Gibbs *et al.* 2010). At least 3 public databases accessed the significant results reported in a few published papers. Combining these data is challenging, not only because scientists investigate eQTL analysis with different statistical frameworks, but also because experimental techniques vary considerably (Kabakchiev and Silverberg 2013). In most instances, microarray technology is used to measure gene expression, but RNA sequencing techniques are more and more applied (Babak et al. 2010; Lalonde et al. 2011; Majewski and Pastinen 2011; Pickrell et al. 2010). Further comparison across studies is difficult because genotyping platforms by different manufacturers, or across variations of the same platform, provide dissimilar coverage of genetic markers. Nevertheless, at least 30 % of eQTL appear to be stable between tissues and cell types (Kabakchiev and Silverberg 2013). However, some eQTL seem to be tissue specific and it is important to identify these. Furthermore, for the majority of disorders that affects a single organ or a limited number of tissues, eQTL have a specific effect on the phenotype and could be undetectable at a different anatomic location. Additionally, some eQTL in multiple tissues such as blood, liver and skeletal muscles have been shown an opposite allelic effect depending on the cell type (Figure 5) (Fu *et al.* 2012).

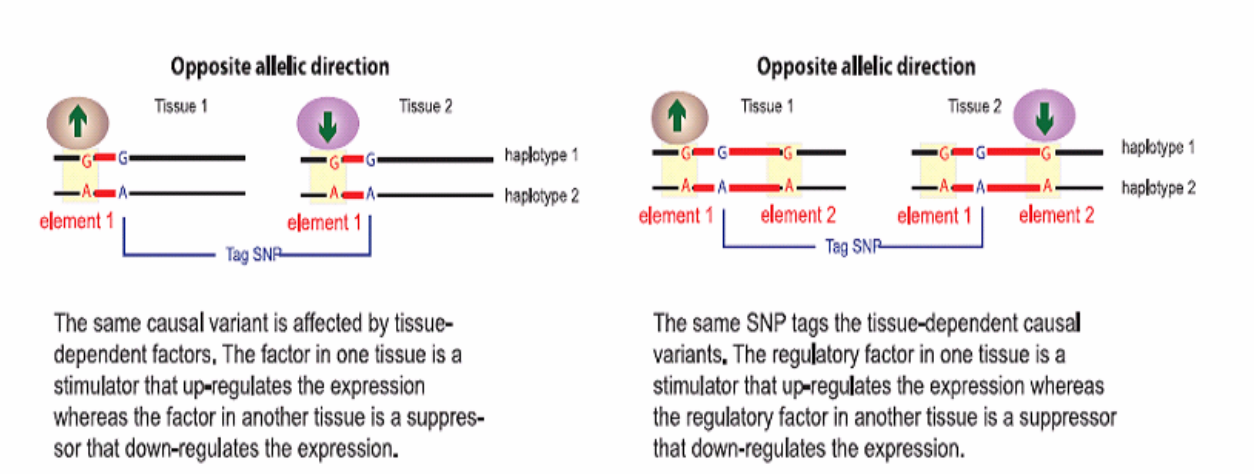

Figure 5: Molecular model of tissue dependent *cis*-regulation with opposite allelic direction (modified from Fu et al. 2012)

Thus far, four genome-wide eQTL studies have been performed using human livers, where many liver eQTL have been found to be reproducible and a proportion of these could be specific for liver transcripts (Glubb *et al.* 2012). Schadt et al. (2008) identified 1350 *cis*-

regulated and 242 *trans*-regulated eQTL in post mortem tissues and resection from donor livers of Caucasians. Additionally, Schroder et al. (2011) observed 1179 *cis* and 47 *trans* eQTL in normal tissues resection during surgery for liver cancer in Caucasians. The highest number of eQTL (7902 *cis*-regulated and 785 *trans*-regulated) in Caucasian were found in the study of Greenawalt et al. (2011). In the study of Innocenti et al. (2011) 1787 *cis* and 353 *trans* eQTL were observed in post mortem tissue and resections from donor livers of European Americans and African Americans.

In the study of Stranger et al. (2012) 5691 eQTL were identified in different human populations like Asians, European-admixed and African subpopulations. Li et al. (2013) found 1359 eQTL significantly associated with breast cancer, whereas 689 (50.7 %) genes were *cis*-regulated and 670 (49.3 %) *trans* genes. Three significant *cis* associations mapping to breast cancer risk loci were identified at chromosome 2 (*insulin-like growth factor binding protein 5* (IGFBP5)), 5 (*SET domain containing 9* (C5orf35)) and16 (*TOX high mobility group box family member 3* (TOX3)). They also observed three *trans* associations at chromosome 6 (*estrogen receptor 1* (ESR1)), 9 (*Kruppel-like factor 4* (KLF4)) and 8 *(myelocytomatosis oncogene* (MYC)). These findings provide a more comprehensive picture of gene expression determinants in breast cancer as well as insights into the underlying biology of breast cancer risk loci (Li *et al.* 2013).

1.4.2 The importance of genetical genomics in livestock

The genetics underlying production traits has been studied and exploited for genetic improvement of livestock through selective breeding for decades. For many traits, regions of the genome that affect these traits have been identified and in some cases even the molecular polymorphism underlying the QTL has been identified (Andersson and Georges 2004; de Koning et al. 2005). A combination of proven approaches in QTL detection and emerging technologies in gene transcription analysis can provide a fast track for unravelling the genetic network underlying differences in production traits (de Koning *et al.* 2005). However, the genetical genomic approach in livestock was only used in chicken and pig and was not so common like in human. In cattle, until now no genetical genomics studies were performed. One reason might be the population structure, because it is not possible to build recombinant inbred lines for livestock species (de Koning et al. 2005) and there are also no reciprocal backcrosses (F_2 population) in cattle.

In chicken, Le Mingon et al. (2009) identified three candidate genes *3-hydroxy-3 hethyl-glutaryl-CoA synthase 1* (HGMCS1), *transcription factor 3* (TCF3) and *sal-like 4* (SALL4) in the QTL region on chromosome 5 for abdominal fatness. Le Bihan-Duval et al. (2011) observed the candidate gene *beta-carotene 15, 15'-monooxygenase* (BCMO1) as *cis*-regulated for chicken breast meat colour. These findings indicated that higher expression of BCMO1 gene was linked with lower meat yellowness (Le Bihan-Duval *et al.* 2011). In the study of Ka et al. (2013) the *carnitine palmitoyltransferase 1B* (CPT1B) expression was found to be influenced by a *cis*-acting eQTL in muscle of two chicken lines selected for high and low body weight. The increased expression in hypothalamus and reduced expression in muscle is consistent with an increased food intake in the HWS line and at the same time reduced fatty acid oxidation in muscle yielding a net accumulation of energy intake and storage (Ka *et al.* 2013).

In pig, Lobjois et al. (2008) identified 63 differentially expressed genes on SSC 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 17 and 18 for different ranges of Warner-Bratzler shear force (WBSF) in a commercial F_2 population. In the study of Ponsuksili et al. (2008b) found 789 differentially expressed genes between high and low drip loss in a Duroc × Pietrain resource population. 10 genes *vitronectin* (VTN), *alpha-1-microglobulin/bikunin precursor* (AMBP), *serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1* (SERPINA1), *cytochrome P450, family 2, subfamily C* (CYP2C*), cytochrome P450, family 3, subfamily A* (CYP3A), *TYRO protein tyrosine kinase binding protein* (TYROBP), *AHNAK nucleoprotein* (AHNAK), *insulin-like growth factor 2* (IGF2*), zyxin* (ZYX) were selected for validation (Ponsuksili *et al.* 2008b). In the same population 104 eQTL significantly associated with water holding capacity were revealed with 96 *trans* acting eQTL and 8 *cis*- acting eQTL (Ponsuksili *et al.* 2008a). The eight candidate genes with *cis* eQTL were located on SSC2, 3, 4 and 6 (Ponsuksili *et al.* 2008a). In a third study, a principle component analysis was used to identify 85 candidate genes exhibiting *cis* eQTL, which were associated with different meat quality traits such as drip loss, pH1, ph24 and meat colour (Ponsuksili *et al.* 2010). Steibel et al. (2011) revealed 62 eQTL and 3 gene networks enriched with genes involved in lipid metabolism, DNA replication and cell cycle regulation in loin muscle tissue of Duroc × Pietrain F_2 population. 2 candidate genes (*aldo-keto reductase 7A2* (AKR7A2) and *thioredoxin domain containing 12* (TXNDC12)), which were part of the lipid metabolism, were located on SSC6 (Steibel *et al.* 2011). In the study of Cánovas et al. (2012), 396 *trans*-regulated and 59 *cis*-regulated eQTL were

observed in the *Gluteus Medius* muscle in purebred Duroc population, whereas 11 eQTL hotspots were mapped on SSC1, 3, 5, 6, 7, 12 and 18.

When comparing the studies using genetical genomics approach in human and livestock, the percentage of *cis*- and *trans*-regulated eQTL were different between the species. In human the proportion of *cis* eQTL ranged between 83 % and 96 % (Greenawalt et al. 2011; Innocenti et al. 2011; Schadt et al. 2008; Schroder et al. 2011). In contrast, in livestock the proportion of *cis*-regulated eQTL ranged between 10 % and 13 % (Cánovas et al. 2012; Ponsuksili et al. 2008a; Ponsuksili et al. 2010). However, considerable heterogeneity of *cis* eQTL effects is possible between different tissues: A recent study reported that the proportion of heritability due to gene expression attributable to *cis*-regulation differs between tissues (37 %in blood and 24 % in adipose tissue) (Price *et al.* 2011). By comparing the overlap of significant *cis* eQTL at a predefined threshold, estimates on the tissue dependence of *cis* eQTL were between 30 % (liver, adipose tissue) and 70 % - 80 % (fibroblasts, T cells, lymphoblastoid cell lines) (Dimas et al. 2009; Emilsson et al. 2008; Gerrits et al. 2009; Heap et al. 2009).

1.5 Genetical genomics approach using GWAS

Over the last few years, GWAS were used in identifying numerous loci related to complex traits (Kabakchiev and Silverberg 2013). Since the first GWAS paper was published in 2005, linking age-related macular degeneration to SNP in the *complement factor H* (CFH) gene (Klein *et al.* 2005), the field of genetic research has seen a proliferation in the application of this approach. The catalogue of published GWAS, curate by the National Institute of Health, list more than 1200 studies spanning more than 600 phenotypic traits (Kabakchiev and Silverberg 2013). These include diverse disorders such as asthma (Moffatt *et al.* 2007), autism (Wang *et al.* 2009) and Parkinson's disease (Maraganore *et al.* 2005). Also quantitative traits like body weight (Thorleifsson *et al.* 2009) and blood pressure (Levy *et al.* 2007) could be found. In these situations, QTL analysis has been applied successfully in correlating levels of a trait of interest with genotype. Possibly the most natural trait to be associated with variations in the genome is the immediate product of the transcribed genes: messenger RNA (mRNA). Such approaches combine 2 genome-

wide technologies together for a system biology treatment of physiologic problems (Kabakchiev and Silverberg 2013).

The core hypothesis behind eQTL analysis is that polymorphic sites in the genome, such as SNP, could have tangible effect on gene regulation by altering the coding or promoter sequences of genes, their splicing junctions, or other regulatory elements. All of these regions affect the rate at which genes are transcribed, which isoforms are preferentially expressed and how stable the final mRNA product is. Thus, SNP, which were suspected to affect gene expression (eSNP) could be tested with associative statistics (Figure 5) (Kabakchiev and Silverberg 2013).

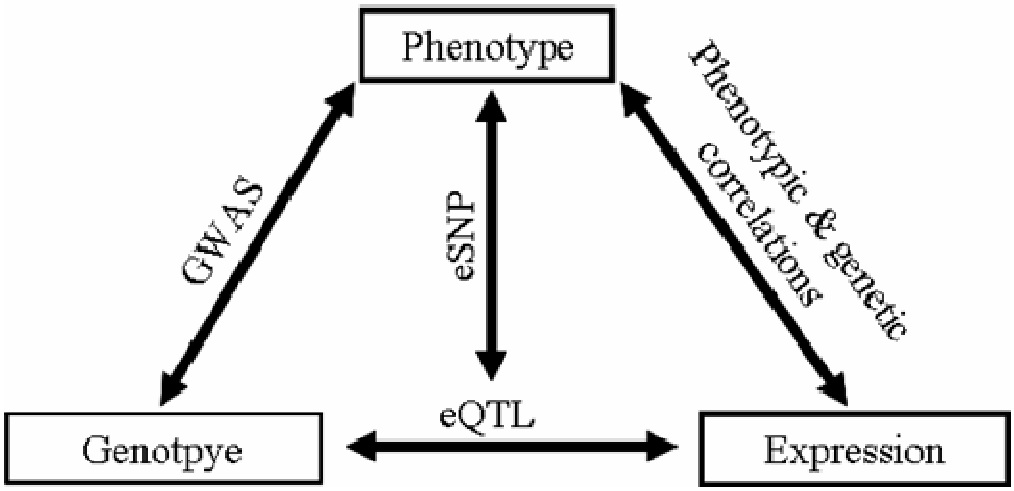

Figure 6: Combining genome-wide association studies with the genetical genomics approach (modified from Powell et al. 2012)

A comprehensive analysis of the GWAS data spanning many different published studies indicated that trait-associated loci, especially those pertaining to complex phenotypes, were enriched for being eQTL as well (Kabakchiev and Silverberg 2013; Nicolae et al. 2010). Nicolae et al. (2010) estimated in their study that approximately 17 % of Crohn's disease associated SNP could be eQTL in lymphoblastoid cell lines. Gaffney et al. (2012) observed in their study that in about 80 % of significant eQTL there is at least an additional SNP. For schizophrenia, Kim et al. (2012) identified four candidate genes *5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled* (HTR2A), *plexin A2* (PLXNA2), *serine racemase* (SRR) and *transcription factor 4* (TCF4), which were significantly associated

with *cis* SNP in at least one brain region of prefrontal cortex, hippocampus, temporal cortex, thalamus and cerebellum. In the study of Kabakchiev and Silverberg (2013) 15 % of the SNP associated with the intestinal tissue could be *cis*- regulated eQTL and only 2 SNP seemed to be *trans* regulated eQTL. For autism, Cheng et al. (2013) identified 12 SNP near the gene *sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A* (SEMA5A) as *cis* eQTL and 920 SNP as *trans*-regulated, which could be divided in 245 eQTL cluster.

In livestock, until now only two recent studies in pig are known combining genome-wide association studies with a genetical genomics approach. In the study of Ponsuksili et al. (2011), 150 crossbred pigs (Pietrain × (German Large White × German Landrace)) were used to identify 448 *cis* eQTL corresponded to 71 genes and 3297 *trans* eQTL were related to 408 genes associated with fatness traits. In another recent study of Ponsuksili et al. (2012), 207 muscle- and 150 liver samples of pigs from a commercial crossbred Pietrain × (German Large White × German Landrace) were analyzed. In muscle, 2001 *cis* and 1663 *trans* eQTL corresponding to 593 genes correlated with plasma cortisol level. In liver, 1019 *cis* eQTL and4873 *trans* eQTL were found, corresponding to 116 and 927 genes, respectively. However, in muscle a higher proportion of cis eQTL was observed (Ponsuksili *et al.* 2012).

1.6 Scope of the study

The Bonner Duroc × Pietrain resource population (DuPi) is well established for investigating quantitative traits such as carcass or meat quality. In the studies of Liu et al. (2007; 2008) QTL analyses in585 F_2 DuPi animals were performed revealing 137 QTL for 35 traits of growth, carcass composition and meat quality. Especially for drip loss, pH1 and pH24 in loin 11 QTL were found on SSC 1, 2, 3, 4, 5, 6, 7, 15 and 18 (Liu *et al.* 2007; Liu *et al.* 2008). For these QTL analyses 106 microsatellites, which were spread across the 18 autosomes and spanned 1987 Kosambi cM, were used.

Furthermore, Ponsuksili et al. (2008a) selected out of 585 F_2 DuPi pigs 74 animals based on their phenotype of drip loss and pH 24. Gene expression profiles in *Musculus longissimus dorsi* were obtained using microarray technology. The relationship between the phenotype and the gene expression profiles was determined using the Pearson correlation coefficient. Additionally, based on approaches of 'genetical genomic', expression QTL (eQTL) were detected to characterize the candidate genes as *cis*- or *trans*-regulated (Ponsuksili *et al.* 2008a). 104 eQTL significantly associated with water holding capacity were identified with 96 *trans* acting eQTL and 8 *cis* acting eQTL. The eight candidate genes, which were *cis*- regulated, were located on SSC2, 3, 4 and 6 (Ponsuksili *et al.* 2008a).

The aim of this complementary study was to investigate water-holding capacity in the DuPi population using a genome-wide association technique. In the first part (Chapter 2), a genome-wide association analysis was performed using the 60K SNP chip of Illumina. QTL regions and their promising candidate genes were identified and their explained phenotypic variance were calculated. In the second part (Chapter 3), candidate genes related to drip loss were observed by (1) analyzing the correlation of drip loss and transcript abundance and (2) combining genome-wide expression profiling and genotyping.

Chapter 2 A genome-wide association study for water holding capacity in a Duroc × Pietrain resource population

Hanna Heidt[1], Christine Große-Brinkhaus[1], Sudeep Sahadevan[1], Christian Looft[1], Heinz Jüngst[1], Dawit Tesfaye[1], Ernst Tholen[1], Karl Schellander[1] and Mehmet Ulas Cinar [1§]

[1]Institute of Animal Science, Group of Animal Breeding and Genetics, University of Bonn, Endenicher Allee 15, 53115 Bonn, Germany

2.1 Abstract

Background: Low water-holding capacity (WHC) results in a fluid consisting of water and protein expelled from the meat surface, so-called drip loss. It has a low heritability and is influenced by many different genes in the genome. Although several candidate genes have been identified with other methods such as QTL, many more genes influencing the trait of interest are present in the pig genome. The genome-wide association (GWA) approach allows dissecting the genome and finding polymorphisms with influence on the observed phenotypic variation. Therefore, the aim of this study was to identify candidate genes related to WHC by performing a GWA study using the Illumina 60K porcine SNP chip. One hundred sixty nine F_2 animals of a Duroc × Pietrain (DuPi) resource population were genotyped and four meat quality traits (drip loss, pH1, pH24 and pH decline) were used to investigate the genetic background of WHC.

Results: 49 significant SNP were observed ($P < 0.01$) for drip loss, 40 SNP for pH1, 9 SNP for pH24 and 33 SNP for pH decline in loin. In order to clarify the annotation and function of each SNP the Biomart software was used. Analyses revealed 14 putative functional candidate genes significantly associated with drip loss. Twenty-six, 7 and 22 candidate genes were identified for pH1, pH24 and pH decline, respectively. The genes NELL1 and SOX6 located on SSC2 were significantly associated with drip loss and showed more than 3 point-mutations each with high linkage disequilibrium. The proportion of explained phenotypic variance for all four traits ranged between 4.4 % and 8.43 %.
Conclusion: Applying a genome wide association analysis using the 60K porcine SNP panel allowed investigating the genetic background of WHC traits in this study. Several regions in the genome were identified affecting drip loss, pH1, pH24 and pH decline in the DuPi population.

Keywords: pig, Illumina 60K SNP chip, water-holding capacity, drip loss, pH traits

2.2 Background

Water holding capacity (WHC), i.e. the capacity of meat to retain its water during application of external forces (Hamm 1985), is an important quality criterion for the meat processing industries and the consumers. WHC affects the financial output, nutritional value, sensorial and technological properties of porcine meat. During the post-mortem conversion of muscle to meat, lactic acid produced in the tissue leading to a reduction in pH of the meat (Huff-Lonergan and Lonergan 2005). Once pH has reached the isoelectric point, positive and negative electrical charge on the proteins are equal. These positive and negative groups within the protein attracted each other and cause a space reduction within the myofibrils (Huff-Lonergan and Lonergan 2005). Additionally, rapid pH decline resulting in ultimate pH (pH at 24 h) while muscle is still warm results in denaturation of many proteins including those involve in binding of cellular water (Huff-Lonergan and Lonergan 2005). Consequently, a fluid consisting of water and protein expelled from the meat surface without any mechanical force other than gravity, the so called drip loss (Offer and Knight 1988). WHC can be predicted by measuring drip loss using gravitational techniques. Drip loss is genetically and phenotypically correlated with early pH in loin (pH1) (r_G = -0.91 and r_P = -0.67, respectively) and late pH in loin (pH24) (r_G = -0.72 and r_P = -0.51, respectively) (Borchers *et al.* 2007). Heritability estimates for water holding capacity traits were low: pH1 (h^2 = 0.14), pH24 (h^2 = 0.20) and drip loss (h^2 = 0.14) (Borchers *et al.* 2007). The presented heritability and correlation estimations are an indication of a polygenic background that is possibly involved in the expression of the examined traits. The identification of genes that regulate meat quality traits will assist in efficient meat production and facilitate the resolution of existing production problems or so-called marker-assisted selection (MAS) (Brunner *et al.* 2012). Both candidate gene and QTL mapping strategies have been used in domestic animals for the discovery of genetic markers suitable for MAS (Rothschild *et al.* 2007).

Until now, 6818 QTL covering 585 phenotypic traits were identified and submitted into pig QTLdb. In total, 1040 QTL were identified for drip loss, pH1, pH24 and pH decline in loin in different pig populations (www.animalgenome.org, Release 17, July 2012) (Hu *et al.* 2010). These QTL were mostly identified via genome scanning based on linkage analysis and microsatellite genotyping.

Liu et al. (2007; 2008) found 11 QTL for drip loss and pH values in loin muscle in the Duroc × Pietrain F_2 resource population by using different statistical models. In another

Duroc and Pietrain F_2 resource population 9 drip loss and pH value QTL were identified on different porcine chromosomes (Edwards *et al.* 2008). Consequently, by using low density of microsatellite markers, QTL are often mapped to a large interval of 20 centimorgans (cM) or more. Only a few causative mutations, which are called quantitative trait nucleotides (QTN), have been identified based on results for complex traits in pigs via QTL fine mapping analysis (Ren et al. 2011; Van Laere et al. 2003). Applying a genome wide association analyses by using current 60K porcine SNP panel provides more dense genotypic markers than microsatellite markers which helps to improve accuracy in finding the exact QTL locations and candidate genes for complex traits (Luo *et al.* 2012b).

Although, there are several GWA studies, conducted for quantitative traits in pigs such as meat quality (Duijvesteijn et al. 2010; Luo et al. 2012b; Ponsuksili et al. 2011; Ramos et al. 2011), reproduction (Uimari *et al.* 2011), immune genetic (Luo *et al.* 2012a; Wang *et al.* 2012) and growth with soundness (Fan *et al.* 2011), no study was devoted to water-holding capacity related traits in pigs. Therefore, the aim of this study was to identify candidate genes related to WHC by performing a whole genome association analyses by using 60K porcine SNP chip genotyping. To the best of our knowledge, this is the first study, which shows functional mutations, related to drip loss and pH traits in pigs.

2.3 Results

In this study, the F_2 DuPi animals were used to perform GWAS for water-holding capacity. Descriptive statistics for the phenotypes as means, standard deviations, minimum and maximum values of the traits measured the current experiment were given in table 2. The animals were slaughtered at an average of 183.7 days and with an average carcass weight of 86 kg. The mean of drip loss was 2.01 %, and the average values of pH1, pH24 and pH decline were 6.54, 5.52 and 1.02, respectively. Further details of the animals and the phenotype recoding can be found in Materials and Methods as well as in Liu et al (2007; 2008).

The GWA analyses were performed using PLINK software (Purcell *et al.* 2007). The analyses were based on the family-based association tests for quantitative traits (QFAM) approach in order to avoid population stratification. Quality control was applied using PLINK and R software. The following criteria for quality control were chosen: call rate of

95 %, minor allele frequency (MAF) > 5 % and missing rate per SNP 2 %. In total, 49 SNP were found to be significantly ($p < 0.001$) associated with drip loss, 29 SNP with pH1, 8 SNP with pH24 and 29 SNP with pH decline. The 49 SNP associated with drip loss were located at several regions on the genome, especially on SSC2, SSC12 and SSC17 (Figure 7). These significant SNP were further annotated to ENSEMBL database using Biomart (R – package) and 14 candidate genes associated with drip loss were identified (Durinck *et al.* 2012). Among these candidates, three genes *NEL-like 1* (NELL1), *sex determining region Y-box 6* (SOX6) and *protein tyrosine phosphatase, receptor type, A* (PTPRA) showed high number of significant SNP (8, 4 and 3, respectively) (Table 3). The phenotypic variance explained by a single SNP varied between 4.4 % and 5.87 % with a standard error of 0.14 and 0.24 (Table 3). Whereas NELL1, SOX6 and PTPRA showed a phenotypic variance of 5.54 %, 4.53 % and 5.87 % with standard errors of 0.23, 0.14 and 0.14, respectively (Table2). The additive effects for the significant SNP were also calculated and ranged from -0.53 to 0.71 (Table 3). For pH1, 25 candidate genes were found across the whole genome. Furthermore, out of 25 genes 12 were identified on SSC1 and SSC6 (Table 4). The phenotypic variance explained by a single SNP varied between 5.63 % and 8.76 % with a standard error ranged from 0.12 – 0.02 (Table 4). The additive effects for pH1 varied between -0.11 and 0.92 (Table 4). Twenty-two genes were identified for pH decline, which were spread all over the genome, especially on SSC1 and SSC7 (Figure 10 and Table6). The phenotypic variance explained by pH decline ranged from 5.63 % to 8.10 % with the standard error of 0.02 to 0.04 while the additive effects of pH decline varied between -0.15 and 0.09. The SNP of gene *family with sequence similarity 65, member B* (FAMB65B), located on SSC7 were significantly associated with pH1 as well as with pH decline. The explained phenotypic variance of FAMB65B for pH1 and pH decline ranged between 5.63 % (0.02) and 6.17 % (0.02), respectively. The additive effect of FAMB65B showed the same trend for both pH1 (-0.05) and pH decline (-0.06). For pH24, seven genes were located mainly on SSC7 and SSC13 (Figure 9; Table 5). The phenotypic variance explained by a single SNP varied from 5.63 % to 6.80% with a standard error ranged from 0.01 to 0.02 (Table 5). The additive effects of pH24 ranged from 0.04 to 0.07 (Table 5). Furthermore, the *avian v-ets erythroblastosis virus E23 oncogene homolog* 1 (ETS1) located on SSC9 and significantly associated with pH24, revealed biological important functions related to WHC. This gene showed a phenotypic variance of 6.1 % with a standard error of 0.01 and the additive effect was 0.05.

Two genes *SH3 and multiple ankyrin repeat domains 1* (SHANK1) and *neuroblastoma amplified sequence* (NBAS) were found to be pleiotropic for pH1 and drip loss traits (Table 3 and Table 4). For pH1 and pH decline traits eight pleiotropic loci were detected: *estrogen receptor 1* (ESR1), *exonuclease 3'-5' domain containing 1* (EXD1), *asparagine-linked glycosylation 14 homolog* (ALG14), *carboxypeptidase A6* (CPA6), *cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene* (CMAHP), FAM65B, *a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 3* (ADAMTS3) and *matrix metallopeptidase 13* (MMP13) (Table 4 and Table 6).

Linkage disequilibrium (LD) was calculated for two regions on SSC2 using Haploview (Barrett *et al.* 2005). The regions of SSC2 have been chosen for further investigations because these regions contained 12 markers associated significantly with drip loss in the F_2 population (Table 3). The construction of the haplotype blocks followed the criterions of Gabriel et al. (2002) (Figure 11 and Figure 12). In figure 11A, the LD of the first region including 4 haplotype blocks between 40.6 Mb and 41.5 Mb was identified. The largest block (Block3) containing 8 SNP belonged to NELL1 gene, which was significantly associated with drip loss and showed a very high LD of $r^2 = 1$ (Figure 11A and Table 3). Additionally, in the block 3, the SNP MARC0075421 was not in LD ($r^2 = 0.06$) with any other SNP in this region (Figure 11A). This might be due to the wrong position of the SNP because of a mistake in the build10 assembly of ENSEMBL pig genome database (Ramos *et al.* 2011). In this animal population (DuPi), all the significant SNP of block 3 were presented in 4 haplotypes (Figure 11B). The second region was located between 45.6 Mb and 46.6 Mb containing 2 LD blocks (Figure 12A). The four significant SNP, which were related the candidate gene SOX6, were found in the second haplotype block and showed a high LD with $r^2 = 0.98$ (Figure 12A and Table 3). In the second block there was also a SNP MARC0054900, which was not in LD ($r^2 = 0.18$) with the other SNP in this region (Figure 12A) like the SNP in block 3 (Figure 11A). Furthermore, the second block including the SNP, which were significantly associated with drip loss, showed 4 different haplotypes in the F_2 population (Figure 12B).

2.4 Discussion

2.4.1 Filtering of SNP data

Quality control of the SNP was based on the call rate, MAF and pedigree errors. Deviations from Hardy-Weinberg equilibrium (HWE) were not considered in the quality control step, because the HWE is underpowered to detect genotyping errors (Cox and Kraft 2006). Because of a high degree of relationship within the F_2 crosses (full sib and half sib families) two discordant full sibs of each family were selected showing extreme values of drip loss (see Table 2). A deviation from HWE could be expected due to the discordant allele frequencies in the parental breeds. Similarly in the studies of Duijvesteijn et al. (2010) and Ponsuksili et al. (2011) the HWE for controlling pedigree errors was also not considered.

2.4.2 QTL-areas

In this study 9 QTL on six different porcine chromosomes were significantly associated with drip loss. Some of the identified QTL have been already reported in the DuPi population. Liu et al. (2007; 2008) identified QTL on SSC2, SSC3 and SSC6 related to drip loss in the same DuPi population. Because of the repeated detection of these QTL, it can be assumed that they play an important role in the expression of drip loss. We identified two QTL on SSC2 which were located in the confidence interval of the QTL identified by Liu et al. (2007) and have been described in other cross bred populations as well (Kim et al. 2005; Liu et al. 2007; Thomsen et al. 2004; van Wijk et al. 2006). For 14 detected QTL affecting pH-traits, only two of these QTL located on SSC1 and SSC7 have been reported earlier in the DuPi population (Liu *et al.* 2007; Liu *et al.* 2008). In the recent analysis, the QTL on SSC1 was found for pH1 and pH decline traits, however, Liu et al. (2007) identified a significant QTL for pH24 in loin. Vidal et al. (2006) identified a QTL for pH1 on SSC1 in a non-inbred maternal Landrace line. Additionally an epistatic QTL study in the DuPi population revealed several epistatic QTL pairs associated with meat quality traits in the same region (Grosse-Brinkhaus *et al.* 2010). Four QTL on SSC2, 3, 8 and 13 for pH1 were identified considering epistasis (Grosse-Brinkhaus *et al.* 2010). In our study, it was possible to identify these epistatic QTL independently from a second QTL, because of the high number of genetic markers provided by the SNP chips.

Until now, we discussed only QTL, which have been described once in the DuPi populations (Liu *et al.* 2007; Liu *et al.* 2008). In general, 17 of the identified QTL have been detected in other pig populations, whereas six QTL are presumably novel positions and not mentioned in the pig QTLdb before (Hu *et al.* 2010).

A QTL on SSC15 affecting drip loss has been identified by Bertram et al. (2000) in a commercial Danish pig population. Markljung et al. (2008) found a QTL on SSC16 for drip loss in a Hampshire × Finnish Landrace cross. The position on SSC17, explaining 5 % of the phenotypic variance, has been described in the DuPi population and in different commercial breeds related to candidate genes (Wimmers *et al.* 2007). Edwards et al. (2008) investigated also a Duroc × Pietrain resource population and detected a QTL for pH1 on SSC3, which explained 6.27 % of the phenotypic variance. Yue et al. (2003) detected a QTL on SSC6 for pH1 and on SSC7 for pH24 in crosses of Meishan, Pietrain and European Wild Boar. In our analyses these QTL were characterized by SNP belonging to 7 different genes on SSC6 and 3 genes on SSC7. Evidently, more than one gene leads to the QTL effect and Yue et al. (2003) reported large confidence interval in the same regions.

On SSC8 and SSC11 QTL for pH1 were found in a Duroc × Berlin Miniature pig population, located close the QTL detected in our study (Wimmers *et al.* 2006). For pH24 two additional QTL regions on SSC9 were determined in this study which were also found in the studies of de Koning et al. (2001) and Ponsuksili et al. (2005). Putative QTL for pH decline on SSC4, 7 and 9 were found in a White Duroc × Erhualian cross (Duan *et al.* 2009) and confirmed our observations. QTL regions on SSC4, 7, 9 and 10 for pH1 and on SSC1 and on SSC5 for pH decline were not mentioned in a research article, which was submitted to pig QTLdb previously. They could be assumed novel for WHC traits (Table 4 and Table 6; Figure 8 and Figure 10).

2.4.3 Candidate genes

In summary 69 candidate genes containing 131 SNP were identified for the four WHC traits. Four genes were chosen for further discussion because of their biological importance related to WHC, high linkage between the particular SNP as well as multiple detection of the same gene in different traits.

For drip loss, two major QTL regions of interest were detected containing possible candidate genes (Table 3). The regions on SSC2 showed a high number of significant SNP which were linked with the two candidate genes NELL1 and SOX6. Both regions showed high LD between the SNP of the two candidate genes, which may be cause of assembled inheritance of SNP. This means for both regions each haplotype block explains a part of the effect of the whole region but due to the high LD between the SNP they capture the same variation. Therefore, both areas remain relevant for the determination of drip loss. NELL1 is a thrombospondin 1 like protein that is strongly expressed in neuronal tissue of human and mice (Kuroda and Tanizawa 1999). Kuroda and Tanizawa. (1999) identified NELL1 in their study as a protein kinase C (PKC) interacting protein, which is a calcium / phospholipid dependent protein kinase with specific physiological function in various cellular activities like cell growth, differentiation and positive regulation of actin filament polymerization by interacting with their own target proteins. Thus NELL1 seemed to influence the muscle contraction indirectly via PKC which inhibits the myogenesis (Vaidya *et al.* 1991). SOX6 was identified on SSC2 as a second important QTL with higher LD. SOX6 is a member of the evolutionarily highly conserved SOX transcription factor family and is expressed in various cell types. It is implicated in the regulation of more than one gene especially in the skeletal muscle (Connor *et al.* 1995; Hagiwara *et al.* 2005). In the studies of Hagiwara et al. (2005; 2007). SOX6 was required for normal development and / or function of the skeletal muscle in human and mice. SOX6 protein was expressed as well in slow muscle fibres as in fast muscle fiber types. However, in the absence of SOX6 transcription factor, the expression of fast skeletal isoforms genes, including myosin heavy chain and troponin genes became distorted. Furthermore, SOX6 regulated both sarcomeric and calcium regulatory components that contributes to the contractile properties of slow myofibres (Quiat *et al.* 2011). Wimmers et al. (2006) described in their study that drip loss at cellular level was associated with high proportion of fast twitch glycolytic muscle fibres, low vascularisation and reduced mitochondrial activity. Therefore, SOX6 seemed to be a good candidate gene for drip loss.

In addition, seven genes were identified for pH1 and pH decline and two related to drip loss and pH1. It can be assumed that pleitropy are important for the relationship between the traits. Out of the nine pleiotropic genes, one gene (FAM65B) will be discussed further because of its impact in the expression of WHC. On SSC 7, the candidate *gene family with sequence similarity 65, member B* (FAM65B) also known as C6orf32 was identified,

which was significantly associated with both pH1 and pH decline and seemed to have a pleiotropic background. C6orf32 is up regulated during muscle cell differentiation and myofiber repair (Cerletti *et al.* 2006). Yoon et al. (2007) described in their study that the down regulation of the C6orf32 expression *in vitro* resulted with decreased muscle cell differentiation and fusion. The authors suggest that C6orf32 also play an important role in cytoskeletal rearrangement during fusion of myoblast into multinucleated myotubes (Yoon *et al.* 2007). Myoblast fusion is important not only for skeletal muscle formation during development, but also for postnatal muscle growth and regeneration of skeletal muscle. Myoblast fusion follows an ordered set of cellular events, including elongation, cell migration recognition/adhesion and membrane fusion (Pavlath 2010). The ratios of different fiber types vary in the diverse muscles depending on developmental stage and muscle function (Brocks *et al.* 2000). Change in fiber type composition can resulted in higher proportion of glycolytic fibres and increase in fiber diameter in pigs (Brocks *et al.* 2000). At the cellular level, high proportion of glycolytic muscle fibres, large fiber diameters, low vascularisation and reduced mitochondrial activity are associated with reduced water-holding capacity (Wimmers *et al.* 2006).

For late pH in loin, only one gene (ETS1) out of seven showed a biological function related with WHC. The *avian v-ets erythroblastosis virus E23 oncogene homolog 1* (ETS1) gene is located on SSC9 ETS1 proto-oncoprotein is a member of the ETS family of transcription factors that share a unique DNA binding domain. It is expressed in various cell types like endothelial cells, vascular smooth muscle cells, skeletal muscle cells and epithelial cells and had both normal and pathological functions (Dittmer 2003). ETS1 is involved in immune / defence response of cytokines and oxidative stress in skeletal muscle, because it is a key mediator of extracellular signal induced activation (Baron *et al.* 2011). The term oxidative stress is frequently used to define only a ‘pathological’ condition in which the production of reactive oxygen species (ROS) (Brigelius-Flohe 2009) which act also as important signalling molecules in muscle contraction and adaptation (Musaro *et al.* 2010).

An increase in ROS production has been demonstrated under osmotic cell swelling, muscle contraction, anoxia, and sepsis, which may cause pH decline and drip loss formation in the skeletal muscle (Ortenblad *et al.* 2003).

2.4.4 Additive effects and explained phenotypic variance

The additive effects as well as the proportion of explained phenotypic variance were calculated for drip loss, pH1, pH24 and pH decline via PLINK software. The calculated additive effects and the explained phenotypic variance for drip loss were quiet high in comparison to other QTL studies done in the same pig population (Edwards et al. 2008; Liu et al. 2007). In the study of Liu et al. (2007) the additive effect for drip loss on SSC2 was -0.18 and the proportion of explained phenotypic variance was 3.75 %. In this study, the additive effects for NELL1 and SOX6 on SSC2 were estimated 0.71 and -0.41 respectively. The phenotypic variance explained by NELL1 was 5.54 % and by SOX6 4.53 %. The additive effects calculated for pH1 (-0.11 to 0.09) were less than for drip loss, but in the studies of Liu et al. (2007), Edwards et al. (2008) and Wimmers et al. (2006) the additive effects were nearly in the same size (-0.04, -0.06 and 0.04, respectively). However, the explained phenotypic variance of pH1 was higher (5.63 % to 8.73 %) than in the studies of Liu et al. (2007) (5.4 %) and Wimmers et al. (2006) (1.6 %). For pH24, the additive effects in the study of Liu et al.(2007) (-0.04 to 0.05) showed the same trend as in this study (0.07 - -0.04), but in the study of Wimmers et al. (2006) the additive effects as well as the proportion of phenotypic variance were less (-0.03 to -0.02, 1.3 % to 1.8%, respectively). The calculated additive effects for pH decline (-0.15 to 0.09) were higher than in the study of Edwards et al.(2008) (-0.06 to 0.07). Reasons for the higher proportion of explained phenotypic variance could be the fine mapping approach of the SNP chips and the higher specificity that the other QTL studies used microsatellites, which were only partly spread across the genome.

In this study, the size of the SNP effects for the pH traits were small, whereas markers in regions in which causative mutations lie have much larger effects, like the SNP, which are significantly associated with drip loss. However, the markers with the largest associated effects from a genomic analysis may not track perfectly the causative mutations; on the other hand, they are potentially useful tools for identifying chromosomal regions associated with trait (Cole *et al.* 2009).

2.5 Conclusion

Applying a genome wide association analysis using the 60K porcine SNP panel allowed investigating the genetic background of WHC traits in this study. Several regions in the genome were identified affecting drip loss, pH1, pH24 and pH decline in the DuPi population. Especially, two regions of SSC2 showed high LD and were clustered in two spans of 0.5 Mb. In these clusters, only two genes were located NELL1 and SOX6, which were associated with drip loss. Most of the candidate genes were functionally related to WHC traits. However, there are several genes, which were not involved in the expression of muscle properties. It can be assumed that these genes have high LD to candidate genes, which were not covered by the SNP chip. Nevertheless, it could be also possible that the significant SNP associated with these traits were annotated in the wrong chromosomal region, which means that the candidate genes are possibly not the right ones. In general, further fine mapping and next generation sequencing technologies are requested to identify the causative mutations of the underlying QTL regions. Validations of candidate SNP are warranted in other pig populations to prove their role in meat quality traits.

2.6 Materials and Methods

2.6.1 Animals and phenotypic traits

In this study, 214 animals of a reciprocal cross of the Duroc (Du) and Pietrain (Pi) breeds were used. These animals consisted of 169 F_2, 39 F_1 and 9 parental (P) animals, which came from 59 full and half sib families. The F_1 generation was produced by mating Duroc boars to Pietrain sows and Pietrain boars to Duroc sows. The F_1 animals were reciprocally crossed to produce the F_2 generation. All F_2 animals were kept and performance tested at the Frankenforst experimental farm of the University of Bonn. The phenotypes were recorded in a commercial slaughterhouse according to the rules of German performance stations (ZDS, 2003)

In total, 4 meat quality traits (drip loss, pH1, pH24 and pH decline) were analyzed. Drip loss was determined using the bag method of Honikel (1987). Therefore, the samples from Musculus longissimus dorsi were collected 24 h post mortem (p.m.), weighted and wrapped in plastic bag. After storage for 48 hours at 4 C the samples were re-weighted and drip loss were calculated as a percentage of weight loss based on the start weight of a

sample. Muscle pH were measured at two different time points after slaughtering using star-series equipment (Rudolf Mathaeus Company, Germany). pH1 was measured between spinous processes of the 13.-14. thoracic vertebra after 45 minutes p.m. pH24 was measured at the same position 24 hours later. pH decline was calculated by subtracting pH1 from pH24.

2.6.2 Sample preparation, genome-wide genotyping and quality control

DNA were extracted from M. longissimus dorsi using Genomic DNA Purification Kit of Fermentas Life Science. DNA concentration was measured using NanoDrop ND-8000 (Thermo-Scientific) and concentration was adjusted to 100 ng/µl by using double distilled RNase and DNase free water. Illumina bead array technology (Porcine SNP 60K Bead Chip) was used for genotyping the samples (Illumina, Inc., San Diego, CA) in accordance with the protocol for SNP Infinium HD assay (http://Illumina.com). At first 200 ng DNA was used for genome-wide amplification and fragmentation. After hybridization to the 62,163 locus-specific 50mers, the DNA was covalently linked to the beads, which were distributed on the surface of the microarray. Single-base extension of the oligos on the BeadChip was implemented using the captured DNA as template, absorbing detectable labels on the Bead Chip. The Illumina iScan detected the signals of each wavelength and converted them to intensity data. In order to normalize the intensity data for each SNP and to specify a cluster position and a genotype, the GenomeStudio software (Illumina, Inc., San Diego, USA) was used. A quality score for each genotype was generated. Because of missing phenotype information, the P and F_1 generation were excluded from further analyses. In the final step, 169 F_2 samples were used for the genome-wide association analyses. The quality of the data was measured using PLINK software (Purcell *et al.* 2007). For the quality control, following measurement parameters were chosen: samples with a call rate less than 95 %; markers with a low minor allele frequency (MAF) < 5 % and SNP with a missing rate of more than 2 % were removed. After the quality check of the data, 153 animals as well as 46.964 SNP remained in the data set.

2.6.3 Genome-wide association analyses (GWA)

For the genome-wide association study, the phenotype data were corrected for environmental effects by using a general linear model (R-software). The model used

following fixed effects: gender (n = 2), slaughterhouse (n = 2) and slaughter season (6 seasons in 3 years) as well as the co-variables age at slaughtering and slaughter weight.

The genome-wide association analyses was conducted with PLINK software, which is based on the family association tests for quantitative traits (QFAM) and performed permutation techniques to account for the dependence between related animals. The within-sib-ship test of QFAM is robust for population stratification. Nominal scores were permuted to receive an empirical p-value while obtaining familial correlation between phenotype and genotype. The permutation procedure applied by QFAM corrected for relatedness within families and was performed 1,000,000 times. Genetic effects and the proportion of explained phenotypic variance were also calculated using PLINK (Purcell *et al.* 2007).

Significant SNP ($p < 0.001$) were annotated using tools of Biomart (R-software) based on porcine Ensembl database (build 10.2). Additionally, based on gene ontology (GO) information the functions of the candidate genes were investigated. SNP, which were not annotated, were characterized using orthologous human-porcine genes.

Linkage disequilibrium (LD) between SNP was calculated on all the animals of the GWA study using Haploview version 4.2 (Barrett *et al.* 2005) and the LD blocks were defined by the criteria of Gabriel et al. (2002).

2.7 Acknowledgments

Authors are grateful to Prof. Dr. med. Markus M. Nöthen from Institute of Human Genetics, University of Bonn, Germany for genotyping the samples via Illumina 60K porcine SNP chip.

This work was part of the DRIP phase II (Genetic- functional background of the water-holding capacity in pork) project and supported by the German Research Foundation (DFG), Germany, grant nr: SCHE562/13-1.

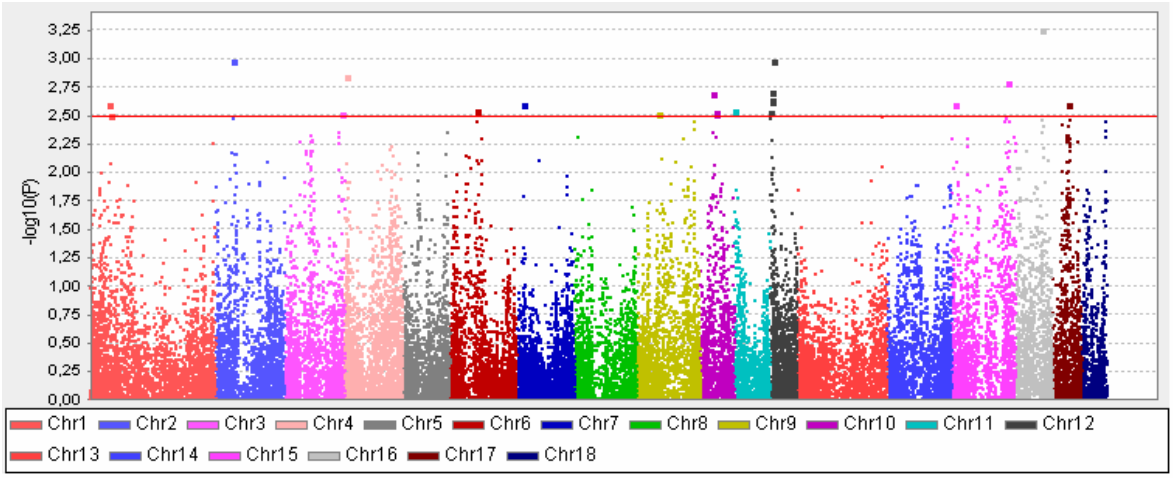

Figure 7: Association between drip loss and 46,964 SNP across autosomes with a threshold of $p < 0.001$.

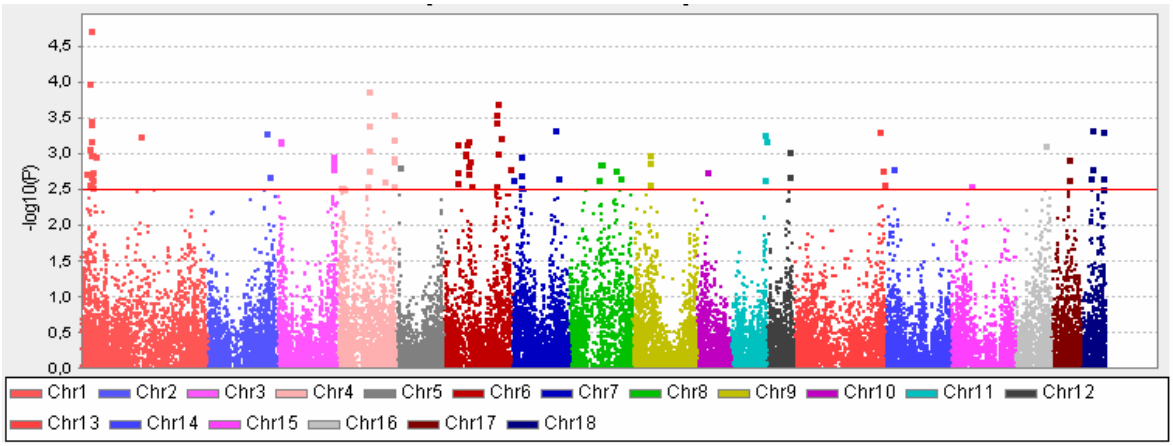

Figure 8: Association between pH1 and 46,964 SNP across autosomes with a threshold of $p < 0.001$.

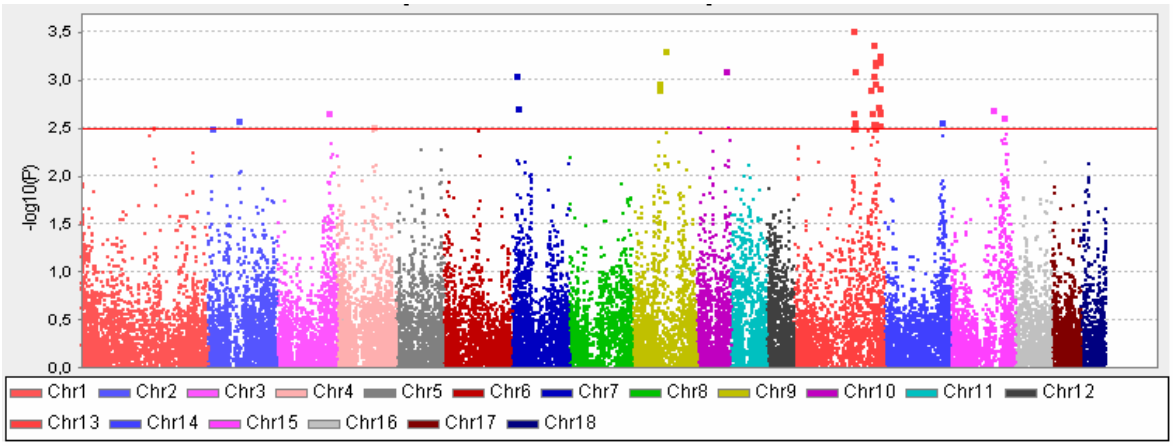

Figure 9: Association between pH24 and 46,964 SNP across autosomes with a threshold of $p < 0.001$.

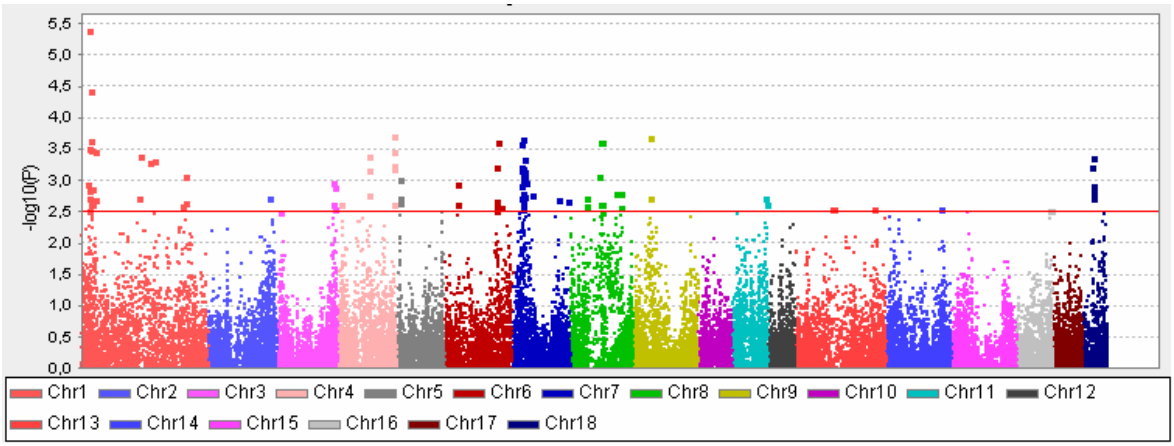

Figure 10: Association between pH decline and 46,964 SNP across autosomes with a threshold of $p < 0.001$.

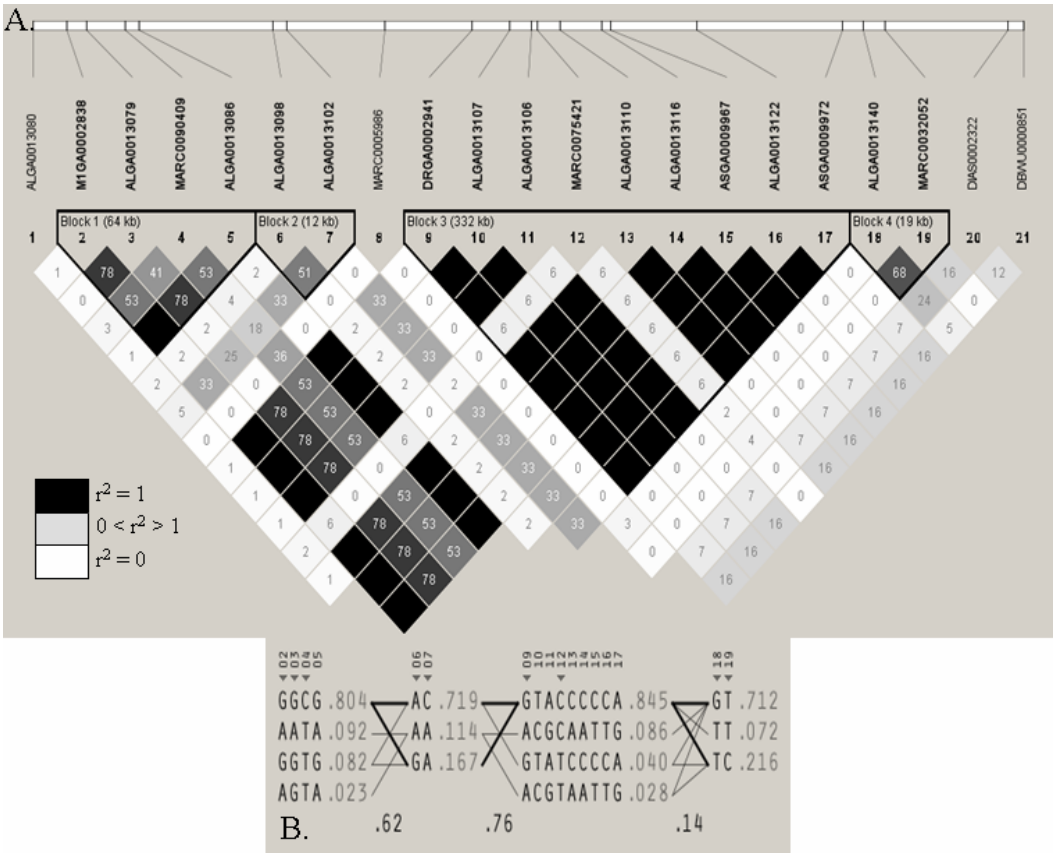

Figure 11: Linkage disequilibrium plot for the region 40.6 Mb and 41.5 Mb on SSC2.

(A) The values in the boxes are pair wise SNP correlations (r2) and the colour shows the degree of correlation. (B) Haplotypes of all SNP from the LD block. Each line represents a haplotype and the frequency of them in the population is given at the end of the line.

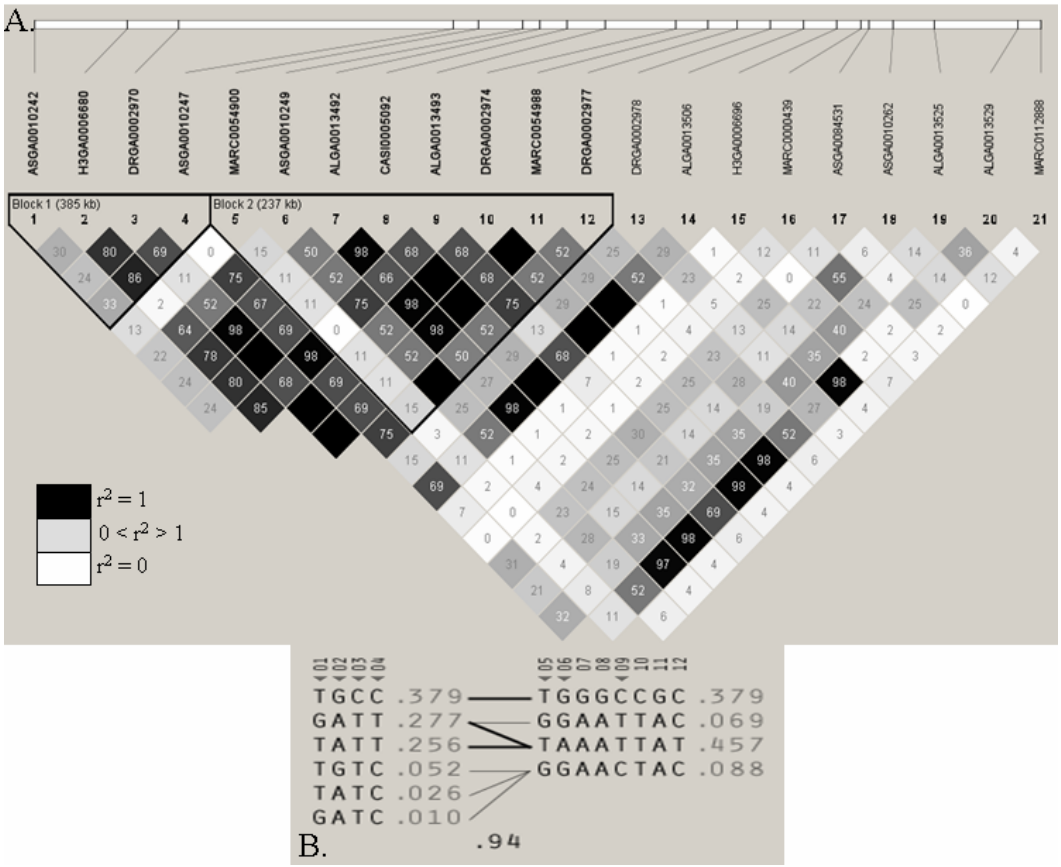

Figure 12: Linkage disequilibrium plot for the region 45.7 Mb and 46.6 Mb on SSC2.

(A) The values in the boxes are pair wise SNP correlations (r2) and the colour shows the degree of correlation. (B) Haplotypes of all SNP from the LD block. Each line represents a haplotype and the frequency of them in the population is given at the end of the line.

Table 2: Phenotypic traits

Trait	N	mean	SD	min	max
Slaughter age (days)	169	183.7	17.117	150	236
Slaughter weight (kg)	169	86	6.043	71.85	104.2
Drip loss (%)	169	2.011	1.384	0.4	5.3
pH1 h p.m.	169	5.518	0.118	5.3	6.06
pH24 h p.m.	169	6.538	0.022	5.89	7.01
pH decline	169	1.02	0.232	0.04	1.49

Table 3: Candidate genes for drip loss: With their locations, number of SNP and additive effects, proportion of explained phenotypic variance and standard error

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
NELL1	SSC2	35177819-35952290	8	ALGA0013106	35514396	0.7085	5.54%	0.2381
				ALGA0013107	35495067			
				ALGA0013110	35570391			
				ALGA0013116	35644749			
				ALGA0013122	35675812			
				ASGA0009967	35601001			
				ASGA0009972	35752040			
				DRGA0002941	35461144			
SOX6	SSC2	39826251-40247509	4	ALGA0013492	39995080	-0.4059	4.53%	0.1485
				CASI0005092	40020472			
				DRGA0002974	40072638			
				MARC0054988	40169024			
NBAS	SSC3	113396698-113507557	2	ALGA0021460	113572582	0.524	5.13%	0.1832
				H3GA0010861	113478423			
SHANK1	SSC6	38985752-39116792	1	ALGA0122867	39016808	0.6538	4.42%	0.2473

Table 3 continued:

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
CARD14	SSC12	1260008-1270313	1	MARC0058650	1263505	-0.5273	5.70%	0.1745
GAA	SSC12	1315074-1451389	1	M1GA0015797	1451828	-0.5233	5.60%	0.1755
SGSH	SSC12	1202710-1363070	1	MARC0058650	1263505	-0.5273	5.70%	0.1745
NEB	SSC15	133973904-134062152	1	ASGA0071803	134011523	0.5707	4.80%	0.1763
CDH12	SSC16	9383414-9692211	1	MARC0097282	9547711	0.4081	4.40%	0.1549
ODZ2	SSC16	53069753-53282894	1	H3GA0053095	53095399	-0.3946	5.21%	0.1369
CSNK2A1	SSC17	36504785-36557774	2	ALGA0094690	36518164	-0.4911	5.10%	0.1722
				ASGA0076651	36550434			
PTPRA	SSC17	34211675-34370646	3	ASGA0076450	34372013	-0.4446	5.87%	0.1449
				DIAS0004607	34236959			
				MARC0090754	34292057			
TCF15	SSC17	36433889-36439851	1	ALGA0094684	36430579	-0.4911	5.10%	0.1725
TRIB3	SSC17	36716172-36752493	1	ASGA0076661	36715251	-0.4911	5.09%	0.1725
VPS16	SSC17	34376484-34398823	1	ASGA0076450	34372013	-0.4551	5.37%	0.1555

SSC Sus scrofa chromosome; 1 add. Effects: calculated additive effects for each gene; 2 VAR: explained phenotypic variance for each gene; 3 SE: standard error of the calculated variance

Table 4: Candidate genes for pH1: With their location, number of SNP, additive effects, proportion of explained phenotypic variance and standard error

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
ESR1	SSC 1	15672455-15672871	1	ASGA0001214	15282738	-0.1024	8.76%	0.02573
EXD1	SSC 1	136534077-136584755	1	ALGA0005984	136563631	-0.07129	6.92%	0.02029
ITPKA	SSC 1	136277482-136285872	1	ASGA0004539	136279728	-0.07129	6.93%	0.02029
PHACTR2	SSC 1	22569930-22640591	1	H3GA0001101	22641924	-0.06466	6.29%	0.01936
MEGF10	SSC 2	117876711-118003906	1	ASGA0011890	117919789	0.0728	7.06%	0.0205
NBAS	SSC 3	113396698-113507557	1	DRGA0004285	113408351	0.07068	6.27%	0.02126
ALG14	SSC4	126917074-127023023	2	DRGA0005306	126961901	-0.1054	7.72%	0.02836
				DRGA0005309	126990548			
CPA6	SSC4	69822261-69901082	2	ASGA0020008	69856928	-0.08021	8.43%	0.02052
				INRA0014602	69833962			
TMEM56	SSC4	126853819-126880070	1	ASGA0022794	126884310	-0.09878	6.12%	0.03004
ASPDH	SSC6	49897788	1	ALGA0122867	39016808	-0.1044	6.28%	0.0313
KIR2DL1	SSC6	41046734-41057508	1	ASGA0028271	41057960	-0.08507	6.68%	0.02468

Table 4 continued:

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
MMEL1	SSC6	43145753-43183292	1	MARC0016324	43187702	-0.08544	6.77%	0.0246
NCR1	SSC6	40934296-40940693	1	ALGA0035381	40932540	-0.08507	6.68%	0.02468
POLD1	SSC6	51096716-51120965	1	ALGA0122867	39016808	-0.1044	6.28%	0.0313
PTGFR	SSC6	124888115-124931460	1	DRGA0006827	93817943	0.08074	6.38%	0.024
SHANK1	SSC6	38985752-39116792	1	ALGA0122867	39016808	-0.1044	6.28%	0.0313
SSU72	SSC6	58158518-58185386	1	ALGA0118516	43014111	-0.06469	5.95%	0.01996
CMAHP	SSC 7	25081295-25138620	1	ASGA0031709	20936416	0.07203	6.26%	0.02162
FAM65B	SSC 7	20647367-20873629	1	H3GA0020222	20868671	-0.05982	5.63%	0.01901
KIAA0247	SSC 7	101649403-101736004	2	ALGA0043784	101712093	-0.0794	7.16%	0.0222
				ASGA0035373	101686817			
ADAMTS3	SSC 8	72590275-72885005	2	M1GA0027035	58286355	-0.07574	6.00%	0.02326
				MARC0059165	58321296			
MMP13	SSC 9	33198282-33209282	1	UMB10000140	33208949	-0.08589	6.31%	0.02569
KIF26B	SSC 10	20383677-20756612	1	ALGA0057607	17923023	0.09223	5.74%	0.02899
RAB20	SSC 11	77468650-77468946	1	H3GA0032794	77473867	0.07631	6.77%	0.02197
KCNJ15	SSC 13	211979422-212022415	1	ASGA0060081	142443122	0.0587	5.80%	0.01836

SSC Sus scrofa chromosome; 1 add. Effects: calculated additive effects for each gene; 2 VAR: explained phenotypic variance for each gene; 3 SE: standard error of the calculated variance

Table 5: Candidate genes for pH 24: With their location, number of SNP, additive effects, proportion of explained phenotypic variance and standard error

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
EDN1	SSC7	9074488-9080825	1	ALGA0038455	9083961	0.0409	6.44%	0.0121
RANBP9	SSC7	10143401-10261724	1	DRGA0007123	10141135	0.07115	5. 63 %	0.02259
SIRT5	SSC7	9985118-10185983	2	DRGA0007121	10123678	0.07115	5.64%	0.02259
				DRGA0007123	10141135			
ETS1	SSC9	53988258-54053051	1	H3GA0027462	54017672	0.04581	6.10%	0.01396
CHODL	SSC13	131781216-131802686	1	DIAS0002702	131807191	0.06108	6.80%	0.01756
PCYT1A	SSC13	100603224-100649430	1	H3GA0037319	100646053	0.04578	6.59%	0.01338
TMPRSS15	SSC13	131807103-131950631	1	DIAS0002702	131807191	0.06108	6.79%	0.01756

SSC Sus scrofa chromosome; 1 add. Effects: calculated additive effects for each gene; 2 VAR: explained phenotypic variance for each gene; 3 SE: standard error of the calculated variance

Table 6: Candidate genes for pH decline: With their location, number of SNP, additive effects, proportion of explained phenotypic variance and standard error

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
ESR1	SSC1	15672455-15672871	2	ALGA0001212	15324789	-0.06267	5.67%	0.01981
				ASGA0001214	15282738			
EXD1	SSC1	136534077-136584755	1	ALGA0005984	136563631	-0.08076	7.42%	0.02243
ITPKA	SSC1	136277482-136285872	1	ASGA0004539	136279728	-0.08076	7.24%	0.02243
MEGF11	SSC1	170966774-171161116	1	MARC0081465	170982070	-0.1087	7.01%	0.03061
PLA2G4E	SSC1	135687454-135706137	1	MARC0029067	135699206	-0.07025	5.68%	0.02223
RAB11A	SSC1	181463316-181492724	1	MARC0081465	170982070	-0.1087	7.06%	0.03061
RMND1	SSC1	15924855-15968565	1	ASGA0001260	15962275	-0.1059	7.61%	0.02874
SHPRH	SSC1	20526290-20598245	1	ALGA0001564	20538317	-0.1503	7.96%	0.03992
TMX3	SSC1	161049204-161088890	1	DIAS0000468	161087201	-0.1464	7.04%	0.04128
ALG14	SSC4	126917074-127023023	2	DRGA0005306	126961901	-0.1202	8.10%	0.03151
				DRGA0005309	126990548			
CPA6	SSC4	69822261-69901082	2	ASGA0020008	69856928	-0.08636	7.30%	0.02389
				INRA0014602	69833962			
TMEM56	SSC4	126853819-126880070	1	ASGA0022794	126884310	-0.1165	6.94%	0.03312

Table 6 continued:

Gene Name	Chromosome	Gene Position	SNP	SNP Name	SNP Position	Add. Effects[1]	VAR[2]	SE[3]
KDELR3	SSC5	6786249-6796827	1	ALGA0105509	5872514	-0.06538	5.64%	0.02076
ATF6B	SSC7	27981853-27992535	1	DIAS0000302	28004848	-0.06691	5.74%	0.02104
CMAHP	SSC7	25081295-25138620	1	ASGA0031709	20936416	0.08867	7.74%	0.02376
DHX16	SSC7	26985174-27002559	1	DIAS0000349	26758347	-0.09366	7.16%	0.02617
FAM65B	SSC7	20647367-20873629	1	H3GA0020222	20868671	-0.06934	6.17%	0.021
LRRC16A	SSC7	21351576-21536235	1	MARC0114328	21456047	-0.07253	6.35%	0.02161
NOTCH4	SSC7	28070737-28095772	2	ASGA0032063	28086571	-0.07849	6.27%	0.02354
				M1GA0009803	28099092			
ZNF193	SSC7	23642080-23646550	1	M1GA0009707	23640996	-0.08869	5.79%	0.02777
ADAMTS3	SSC8	72590275-72885005	2	M1GA0027035	58286355	-0.09553	7.78%	0.02552
				MARC0059165	58321296			
MMP13	SSC9	33198282-33209282	1	UMB10000140	33208949	-0.0899	5.63%	0.02855
IQUB	SSC18	22063004-22112561	1	MARC0001905	22112584	-0.1217	7.21%	0.03389

SSC Sus scrofa chromosome; 1 add. Effects: calculated additive effects for each gene; 2 VAR: explained phenotypic variance for each gene; 3 SE: standard error of the calculated variance

Chapter 3 A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population

Hanna Heidt[1], Mehmet Ulas Cinar[1], Muhammad Jasim Uddin[1], Christian Looft[1], Heinz Jüngst[1], Dawit Tesfaye[1], Astrid Becker[2], Andreas Zimmer[2], Siriluck Ponsuksili[3], Klaus Wimmers[4], Ernst Tholen[1], Karl Schellander[1] and Christine Große-Brinkhaus[1*]

[1] Institute of Animal Science, University of Bonn, Endenicher Allee 15, D-53115 Bonn, Germany

[2] Institute of Molecular Psychiatry, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

[3] Research Institute for the Biology of Farm Animals (FBN), Research Group Functional Genomics, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

[4] Research Institute for the Biology of Farm Animals (FBN), Research Unit Molecular Biology, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

Modified version is published in Mammalian Genome (2013) **24**: 416-26

3.1 Abstract

In this study lean meat water holding capacity (WHC) of a Duroc × Pietrain (DuPi) resource population with corresponding genotypes and transcriptomes were investigated using the approaches of genetical genomics. WHC was characterized by drip loss measured in *M. longissimus dorsi*. The 60K Illumina SNP chips identified genotypes of 169 F_2 DuPi animals. Whole genome transcriptomes of muscle samples were available for 132 F_2 animals using the Affymetrix 24K Genechip Porcine Genome Array. Performing genome-wide associations studies (GWAS) of transcriptional profiles, which are correlated with phenotypes allows elucidating *cis*- and *trans*-regulation. Expression levels of 1228 genes were significantly correlated with drip loss and were further analyzed for enrichment of functional annotation groups as defined by Gene Ontology and KEGG pathways. A hyper geometric gene set enrichment test was performed and revealed glycolysis/ glyconeogenesis, pentose phosphate pathway and pyruvat metabolism as the most promising pathways. For 267 selected transcripts, eQTL analyses was performed and revealed 1541 significant associations in total. Because of positional accordance of the gene-underlying transcript and the eQTL location, it was possible to identify 8 eQTL that can be assumed as *cis*-regulated. Comparing the results of gene set enrichment and the eQTL detection tests, molecular networks and potential candidate genes, which seemed to play key roles in the expression of WHC, were detected. The *α-1-microglobulin / bikunin precursor* (AMBP) gene was assumed to be *cis*-regulated and was part of the gycolyse metabolism. This approach supports to identify trait-associated SNP and to understand the biology of complex traits.

3.2 Introduction

Water holding capacity (WHC), i.e. the capacity of meat to retain its water during application of external forces (Hamm 1985) is an important quality criterion for the meat processing industries and the consumers. WHC can be predicted by measuring drip loss, which is a fluid consisting of water and protein expelled from the meat surface without any mechanical force other than gravity and is influenced by shrinkage of the myofibrils, pH value and temperature post mortem (Offer and Knight 1988).

The structural components and the biological processes of WHC and its associated traits have been extensively investigated (Greaser 2001; Huff-Lonergan and Lonergan 2005, 2007; Huff Lonergan et al. 2010; Offer and Cousins 1992; Rosenvold and Andersen 2003). However, the genetic mechanisms underlying in WHC during the conversion of muscle to meat are not fully understood. The presented heritability for drip loss in different breeds varied between $h^2 = 0.01$ and 0.31 depending on the different measurement methods (e.g. Borchers et al. 2007; Ciobanu et al. 2011) and demonstrate the complexity of this trait.

A genome scan is the most general approach to identify genomic regions showing quantitative trait loci (QTL), classically for complex phenotypic characteristics. Such QTL regions are generally large and contain thousands of putative genes, which are all candidate loci for the trait (Wayne and McIntyre 2002). Until now, 6397 QTL covering 578 phenotypic traits were identified and submitted into pig QTLdb [www.animalgenome.org, Release April 12, 2013]. In total, 953 QTL were identified for drip loss in loin in different pig populations (Hu et al. 2010). These QTL were mostly identified via genome scan based on linkage analyses and microsatellite genotyping. Liu et al. (2007; 2008) found 7 QTL for drip loss (*Sus scrofa* chromosome (SSC) 2, 3, 4, 5, 6, 18) in loin muscle in the Duroc × Pietrain (DuPi) F_2 resource population via applying different statistical methods.

Quantitative expression studies such as microarray technology, can indicate regulatory variation in genes for complex traits (Wayne and McIntyre 2002). By combining QTL mapping and microarray analyses, it is possible to identify regulatory networks underlying the quantitative trait of interest and localizing genomic variation, the so-called genetical genomics approach (Jansen 2003; Jansen and Nap 2001). Furthermore, a QTL analysis of expression levels of gene identifies genomic regions, which are likely to contain at least one causal gene with a regulatory effect on the expression level, termed expression QTL (eQTL). The use of eQTL analyses has been demonstrated as a promising tool for

narrowing the gap between detected phenotype related QTL regions and confirmed causative variations for the pig species (Rothschild et al. 2007; Steibel et al. 2011). Detected eQTL can be classified into a locus, which is located close to a gene (*cis*-regulation) or distant from the gene (*trans*-regulation) (Jansen and Nap 2001). Differential expressed genes, where eQTL mapping indicates *cis*-regulation, are more likely to represent the "cause", for example the genetic background of the trait of interest. Whereas differential expressed genes revealing *trans*-regulation represents the "effect", for example pathways that are affected by causal variation (Wimmers et al. 2010). The largest and most significant reported eQTL are often *cis*-regulated. However, some *trans*-regulated eQTL seem to control expression in several or many genes spread across the genome. These findings provide general information on the organization of the control of expression (Haley and de Koning 2006).

Previous results showed eight *cis*-regulated eQTL significantly associated with drip loss, which were located on SSC 2, 3, 4 and 6. Additionally, eight candidate genes were identified from these results (Ponsuksili *et al.* 2008a). Lobjois et al. (2008) showed in their study that from 63 genes associated with the Warner-Bratzler shear force only 22 were mapped in the pig genome and 12 were located in the areas previously associated with tenderness (SSC2, 6 and 13). In the study of Steibel et al. (2011) expression profiles from loin muscle tissue were combined with linkage analyses of 124 microsatellites showing 62 eQTL including 22 *cis*- regulated eQTL. As shown in these studies, differences in gene expression can be used in reverse genetic studies to generate well-defined hypotheses regarding downstream effects on molecular, cellular and functional networks, and finally at the phenotype level (Ciobanu et al. 2011). However, genetical genomic studies are rarely applied to investigate the complex genetic structure and gene regulation of drip loss in pigs. Therefore, the aim of this study was to identify candidate genes related to drip loss by (1) analyzing the correlation of drip loss and transcript abundance and (2) combining genome-wide gene expression profiling and genotyping.

3.3 Materials and methods

3.3.1 Animals and phenotypic traits

In this study, 214 animals of a reciprocal cross of the Duroc and Pietrain breeds were used. These animals consisted of 169 F_2, 39 F_1 and 9 parental (P) animals, which came from 59 full and half sib families. All F_2 DuPi animals were kept and performance tested at the Frankenforst experimental farm of the University of Bonn. The phenotypes were recorded in a commercial slaughterhouse according to the rules of German performance stations (ZDS 2003). Further information can be found in the study of Liu et al. (2007).

Drip loss was measured using the bag method of Honikel (1986). Therefore, the samples from *M. longissimus dorsi* were collected 24 h post mortem (p.m.), weighted and wrapped in plastic bag. After storage for 48 hours at 4 °C the samples were re-weighted and drip loss were calculated as a percentage of weight loss based on the start weight of a sample.

3.3.2 Sample preparation, genome-wide genotyping and quality control

DNA was extracted from *M. longissimus dorsi* using Genomic DNA Purification Kit of Fermentas Life Science. DNA concentration was measured using NanoDrop ND-8000 (Thermo-Scientific) and concentration was adjusted to 100 ng/µl by using double distilled RNase and DNase free water. Illumina bead array technology (Porcine SNP 60K Bead Chip) was used for genotyping the samples (Illumina, Inc., San Diego, CA) in accordance with the protocol for SNP Infinium HD assay (http://Illumina.com). At first 200 ng DNA was used for genome-wide amplification and fragmentation. After hybridization to the 62163 locus-specific 50mers, the DNA was covalently linked to the beads, which were distributed on the surface of the microarray. Single-base extension of the oligos on the BeadChip was implemented using the captured DNA as template, absorbing detectable labels on the Bead Chip. The Illumina iScan detected the signals of each wavelength and converted them to intensity data. In order to normalize the intensity data for each SNP and to specify a cluster position and a genotype, the GenomeStudio software (Illumina, Inc., San Diego, USA) was used. A quality score for each genotype was generated. Because of missing phenotype information, the P and F_1 generation were excluded from further analyses. In the final step, 169 F_2 samples were used for the genome-wide association analyses. The quality of the data was measured using PLINK software (Purcell et al. 2007). For the quality control following measurement, parameters were chosen: samples with a

call rate less than 95 %; markers with a low minor allele frequency (MAF) < 5 % and SNP with a missing rate of more than 2 % were removed. After the quality check of the data, 153 animals as well as 46964 SNP remained in the data set.

3.3.3 Whole-genome expression profiling

Gene expression profiling of the M. *longissimus dorsi* was conducted with 100 out of the 169 F_2 animals. In brief, total RNA of the loin was isolated using TRI Reagent (Sigma, Taufkirchen, Germany) according to the manufacture's protocol. The RNA was cleaned up using the RNeasy Kit (Quiagen, Hilden, Germany). RNA concentration was measured using NanoDrop ND-8000 (Thermo-Scientific). The integrity and the absence of contamination was check using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Waghäusel - Wiesental, Germany). According to Affymetrix protocols, 500 ng of total RNA were reversely transcribed into cDNA, transcribed into cRNA and labelled using the Affymetrix One Cycle Synthesis and Labelling Kit (Affymetrix, UK) to prepare antisense biotinylated RNA targets. Muscle expression pattern were assessed using GeneChip Porcine Genome Array (Affymetrix) which contains 23937 probe sets that interrogate 23256 transcripts in pig, which represents 20201 genes. Preparation of target products, hybridization and scanning using the GeneChip scanner 3000 were performed according to Affymetrix protocols using 5 µg of total RNA to prepare antisense biotinylated RNA. The quality of hybridization was assessed in all samples following the manufacturer's recommendations. Data were analysed with Affymetrix GCOS 1.1.1 software using global scaling to a target signal of 500.

3.3.4 Microarray data processing

In order to increase the number of muscle transcriptome data, additional arrays from previous investigations were added. To combine these two data sets, it was necessary to remove outliers and correct for the significant batch effect (Oldham et al. 2008). The whole microarray processing is described in Supplementary Methods (Appendix). After microarray processing 132 F_2 DuPi animals were left for further investigation.

3.3.5 Correlation between traits and expression levels

Individual phenotypes of drip loss and expression levels were adjusted for systematic effects using a general linear model (glm) of R-software (www.r-project.org). The linear model contained as fixed effects “gender”, “season” and “place of slaughter” and as covariable “age of slaughter” and “weight of slaughter”. Following, Pearson correlation coefficient were calculated using the residuals of the expression intensities and the trait drip loss. Genes that showed a significant correlation ($P \leq 0.05$) were further analyzed.

3.3.6 Gene set enrichment and pathway analyses

The array contained 23937 probe sets, which represented 12307 UniGene (annotation from November 2012) to measure the global transcripts. The Porcine Genome Array annotation available from NetAffx™ Analyses Centre (Porcine.na30.annot.csv) was used. Additional and updated annotation was obtained from the Ensemble database, using the biomaRt package (version 2.14.0) in R (www.r-project.org). In total 8059 probes were able to be assigned to the group Biological Processes (BP), 8179 probes to Cellular Components (CC) and 8867 probes to Molecular Functions (MF) of the Gene Ontology (GO; http://www.geneontology.org/) database.

A hyper geometric gene set enrichment test (GOstats package version 2.24.0) was performed based on the clusters identified in correlation analyses. Overrepresentation of gene sets defined by the particular groups of BP, CC or MF in the GO database or the Kyoto Encyclopedia of Genes and Genomes database (KEGG; http://www.genome.jp/kegg/) was tested using Fisher's exact test. For this test, only the significant genes, which were annotated, with an Entrez gene ID were included. When a gene had a duplicate on the array, only a single gene ID was used. A gene-set was considered significant if $P \leq 0.05$.

3.3.7 Genome-wide association analyses (GWA)

For the genome-wide association study, the gene expression values were corrected for environmental effects by using a glm of R-software. The model contained the same effects used for the correlation analyses.

The genome-wide association analyses was conducted with PLINK software, which is based on the family association tests for quantitative traits (QFAM) and performed permutation techniques to account for the dependence between related animals. The within-sib-ship test of QFAM is robust for population stratification. Nominal scores were permuted to receive an empirical p-value while obtaining familial correlation between phenotype and genotype. The permutation procedure applied by QFAM corrected for relatedness within families and was performed 100,000 times. Genetic effects and the proportion of explained variance were also calculated using PLINK (Purcell et al. 2007).

Additionally, a false discovery rate (FDR) ≤ 0.1 were used in order to correct for multiple testing (Benjamini and Hochberg 1995).

3.4 Results

Expression profiling and eQTL analyses were performed on 132 animals of the DuPi population with 100 animals of our study and 32 animals from previous studies (Ponsuksili *et al.* 2008a; Ponsuksili *et al.* 2010). The animals were chosen because of a high degree of relationship within the F_2 crosses and extreme values of drip loss. The animals were slaughtered at an average of 183.7 days with an average carcass weight of 86 kg. The mean of drip loss was 2.0 % with a minimum of 0.4 % and a maximum of 5.3 %. Using Affymetrix Porcine Genome Array 24123 expression measurements were performed from each *M. longissimus dorsi* RNA samples of the 132 F_2 animals and were corrected for the batch effects using "ComBat"(Johnson *et al.* 2007). The preselected data were further analyzed with the hybrid algorithm of PLIER (AffymetrixTechnicalNote 2005) resulting in 20733 probe sets. Pearson correlation was calculated for each 20733 probe sets and the drip loss phenotypes. A total of 1228 transcripts were significantly correlated with drip loss ($P \leq 0.05$), with 406 negative and 822 positive correlated genes.

3.4.1 Biological pathways associated with drip loss

We tested the list of significantly positive and negative correlated genes ($P \leq 0.05$) for enrichment in functional annotation groups in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) and Gene Ontology (GO) databases. This

dataset was represented by 1227 probe sets, which represent 688 annotated genes. Eight significant KEGG pathways ($P \leq 0.05$) including “Glycolysis/Glyconeogenesis”, “Pentose phosphate pathway” and “Pyruvat metabolism” for all genes significantly associated with drip loss were found (Table A2). For the negative correlated genes 19 significant KEGG pathways were observed, which were mostly disease pathways (Table 7). Additionally, 10 significant KEGG pathways were identified for positive correlated genes (Table 8).

GO analyses were divided in three parts: BP, CC and MF. For the biological processes, 114 GO categories of genes with significant expression levels were found, whereas the GO categories “Hexose catabolic process”, “Primary metabolic process”, “Gluconeogenesis”, ”Glycolysis”, “Monosaccharide biosynthetic process”, “Protein modification process”, “Cellular metabolic process”, “Cellular carbohydrate catabolic processes” and “Carbohydrate catabolic process” showed most of the genes with a significant association of drip loss (Table A3).

For the genes, which were negatively correlated with drip loss, 235 GO terms were identified compared to the genes positively correlated with drip loss with 129 GO terms. For the cellular components, only 28 GO categories were observed for all significantly associated genes (Table A4). The positive and negative correlated genes revealed 45 GO terms mainly consisting of intracellular components and 21 GO terms representing extracellular components, respectively.

The molecular functions indicated 46 GO categories for all significant associated genes, whereas four GO terms “Phosphorylase activity”, “Oxidoreductase activity, acting on the aldehyde or oxo group of donors”, “Fructose 1, 6-bisphosphate 1-phosphatase activity” and “Catalytic activity” showed the genes, which were significantly associated with drip loss (Table A5). For the genes, which were positively correlated with drip los 77 GO categories were identified compared to the genes negatively correlated with drip loss with 58 GO terms.

3.4.2 Whole-genome association analyses for transcripts correlated with drip loss (eQTL)

A total of 246 genes, which were highly correlated with drip loss ($P \leq 0.01$) were selected for eQTL analyses. Additionally, 21 genes were selected from the gene set enrichment analyses because of their functional role in the development of drip loss and were added

for eQTL analyses. In total, 46946 SNP were tested for association to each of the 267 expression transcripts significantly associated with drip loss. The analyses revealed for 30 transcripts 1451 eQTL with a nominal P-value less than 0.0001 and a genome-wide FDR value of 0.1. The average value for each transcript was 50 eQTL with a minimum of one eQTL per transcript and a maximum of 943 eQTL per transcript. Four transcripts had only one eQTL (Table 9). Their calculated correlations with drip loss ranged between -0.25 and 0.30. From these 1451 eQTL, 466 eQTL were located on the same chromosome like the transcripts. In this study, only 8 eQTL were assumed as potentially *cis*-regulated, because they were located close to the transcript (Table 10). Their explained variance ranged from 10 % up to 30 %. The *neuralized homolog (Drosophila)* (NEURL) gene revealed two potential *cis* eQTL, which were located within the gene. The gene TPI1 which is located on SSC5 showed the highest significance ($P < 2.58E^{-07}$) on SSC12.

3.5 Discussion

3.5.1 Trait dependent expression analyses

The association between the quantitative trait drip loss and the gene expression was analyzed using the Pearson correlation. This approach was similarly used in many studies like Qackenbush (2001), Kraft et al. (2003) and Ponsuksili et al. (2008a). In the study of Kraft et al. (2003), a within-family correlation analyses was utilized to remove the effect of family stratification. In contrast, we used a general linear model to correct both drip loss phenotype and expression levels for environmental effects in the correlation analyses. In the further analyses, we used the family approach of PLINK (Purcell et al. 2007) to avoid the family stratification. For the pathway and GO term analyses, we considered correlation significant at $P < 0.05$ like the studies of Blalock et al. (2004) and Ponsuksili et al. (2008a).

In order to control the problem of multiple testing, a genome wide Bonferroni correction may result in high false negatives (Han et al. 2009). Thus instead of a genome wide Bonferroni correction, which might be overly conservative, we set a more relaxed threshold using a correction following Benjamini and Hochberg with a FDR < 0.10 (Benjamini and Hochberg 1995). Since Storey and Tibshirani (2003) propose an adaptation of the false discovery rate (FDR) targeted at genome-wide experiments to provide a better balance between statistical stringency and power to detect true effects. However, for some

genes even this method seems to be too stringent, e.g. for the gene solute carrier family 37, member 4 (SLC37A4). This gene is located in a region on SSC 9 and showed eQTL on SSC 18 where several promising candidate gene influencing drip loss are discussed (Jennen et al. 2007).

3.5.2 Gene set enrichment of significant correlated genes with drip loss

The genetic background of WHC during the conversion of muscle to meat is not yet completely understood. One possible explanation for these variation obtained in the structure of the muscle itself (Huff-Lonergan and Lonergan 2005). In this study, transcript level of muscle at slaughter were correlated with drip loss after slaughter in order to find biological processes, which are relevant in the development of drip loss. Functional annotation analyses are essentially based on the extrapolation of pathway information and gene ontology data of the pig (Ponsuksili *et al.* 2008a). In our study, changes were observed in the genomic regulations of different cellular pathways, which were correlated with drip loss. The genes positively correlating with drip loss were genes of the group belonging to the energy metabolism (Table 8), the same results were observed in the GO term analyses. For example, glycolytic metabolism is a process characteristic of skeletal muscle (Hamill et al. 2012). Each muscle consists of three main fibre types, slow-twitch oxidative, fast-twitch oxido-glycolytic and fast-twitch glycolytic fibres (Peter et al. 1972). The fast-twitch glycolytic fibres are associated with drip loss, low vascularisation, reduced mitochondrial activity and higher glycogen content (Wimmers et al. 2006). The negative correlated transcripts were enriched mostly in disease pathways same results were found in GO analyses.

Taken together, analyses of trait correlated expression revealed that the complex relationships between biological processes taking place in skeletal muscle and meat quality like drip loss are driven by the energy reserves in the muscle and their metabolisation (Ponsuksili *et al.* 2008a).

3.5.3 Candidate genes for drip loss and their regulation patterns

Drip loss is a complex trait, which is genetically controlled by many different genes. With the growing knowledge of genome sequences and gene annotation, the eQTL analyses give insight into the architecture of regulatory networks (Ponsuksili et al. 2010). 8 eQTL of 7

transcripts were observed as putative *cis*-regulated following the criteria used in the study of Ponsuksili et al. (2011). In our study, 24 % of the significantly associated transcripts were *cis*-regulated and 76 % were *trans*-regulated. This proportion of *cis*-regulated genes is higher in comparison to the studies of Ponsuksili et al. (2008a; 2011; 2010) where 10 % of all transcripts were *cis*-regulated.

The *cis*-regulated genes are of great interest, because the underlying genes are expected to harbour the genetic variants that influence their own expression level and influence the physiological traits of interest (Göring *et al.* 2007).

For example, the polypeptide *α-1-microglobulin / bikunin precursor* (AMBP), which was located on SSC1, was already known to be involved in the formation of drip loss (Cinar et al. 2012; Ponsuksili et al. 2008b) and was also found to be significantly associated with drip loss in our study (Figure 13). AMBP is the precursor of Bikunin, which plays an important role in the stabilization of the extra cellular matrix (Tyagi et al. 2002). Additionally AMBP is important for cell growth, development, metabolism, immune response and the level of intracellular calcium (Grewal et al. 2005). The extracellular matrix is reported to influence meat quality (Velleman 2000), because signal transduction from the extracellular matrix to the myoblast plays a significant role in muscle formation and growth (Velleman 2002). Furthermore the extracellular matrix consist of proteins including collagens and proteogycans, which are contributed to increase the WHC in tissue (Velleman 2002). In the studies of Ponsuksili et al. (2008a; 2008b), AMBP was differentially expressed in pig muscles with high vs. low drip loss and low vs. high pH. These results were confirmed in the study of Cinar et al. (2012). AMBP is mapped on SSC1 where QTL for meat quality traits are reported in different pig breeds and crosses (Hu et al. 2010). In contrast to our results, Cinar et al. (2012) identified AMBP as *trans*-regulated. One reason might be that in the present study a higher number of genetic markers were used. However, this high number of genetic markers may not track perfectly the causative mutations but they are potentially useful tools for identifying chromosomal regions associated with the trait (Cole et al. 2009). The functions of the other potential *cis*-regulated genes (Table 10) were not yet understood in porcine skeletal muscle.

Many loci are *trans*-acting modulators of gene expression (Jansen and Nap 2004) and are also of great interest because these genes are likely to play a role in explaining trait variation (Wang and Nettleton 2006). For example, the gene *Insulin -like growth factor 2* (IGF2) located on SSC2 (Figure 14), which was also found to be differentially expressed in

pigs with high and low drip loss (Ponsuksili *et al.* 2008b) and the gene *triosephosphate isomerase1* (TPI1) located on SSC5 (Figure 15).

IGF2 promotes growth and plays a role in regulating proliferation, differentiation and apoptosis of cells in many different tissues, such as skeletal muscle (Pavelic et al. 2002; van Wyk and Smith 1999). The IGF2 gene is imprinted in most mammalian tissues and is exclusively expressed from the paternal allele (Jeon et al. 1999; Nezer et al. 1999). Additionally, IGF2 is involved in the myogenesis and controls primarily the muscle mass and fat deposition (Jeon et al. 1999; Nezer et al. 1999). It is responsible of 15-30 % of the phenotypic variation in muscle mass (Van Laere et al. 2003).

TPI1 gene is a glycolysis enzyme, which play an important role in energy generation for muscle cells (Solem et al. 2008). In the study of Kwasiborski et al.(2008) a protein analyses was performed and TPI1 was positively correlated with drip loss in a Large White population. The same results ($r = 0.27$ with $P < 0.001$) was found in our study using the DuPi population.

In this study, the presumable *trans*-regulated gene *solute carrier family 37, member 4* (SLC37A4), which is located on SSC 9 is a promising candidate gene for drip loss but failed the FDR level of 0.1 (Figure 16). SLC37A4 also known as glucose-6-phosphate transporter (G6PT) is expressed in several tissues like heart, brain and skeletal muscle (Lin et al. 2000). G6PT translocates glucose-6-phosphate (G6P), the product of gluconeogenesis and glycogenolysis, from the cytoplasm to the lumen of the endoplasmatic reticulum (ER) (Arion et al. 1980). Inside the ER, G6Pase catalyzes the conversion of G6P to glucose. Therefore, G6PT and G6Pase work in both in the glucose homeostasis (Lin et al. 2000). G6PT plays a role in the antioxidant protection, because a defective G6P transport leads to an increase of reactive oxygen species (ROS) (Leuzzi et al. 2003), which act also as important signalling molecules in muscle contraction and adaptation (Musaro et al. 2010). An increase in ROS production has been demonstrated under osmotic cell swelling, muscle contraction, anoxia, and sepsis, which may cause pH decline and drip loss formation in the skeletal muscle (Ortenblad et al. 2003).

3.6 Conclusion

Mapping quantitative traits and unravelling transcriptional control are challenging, when applied to one phenotype at a time. In studies of typical quantitative porcine traits like WHC, strong effects are rarely found. Here we have coupled genomic technologies for expression profiling with genome-wide genetic mapping using SNP markers, and shown that specific chromosomal regions contain functional candidate genes. These approaches and results allowed to investigate and to dissect the genetic contribution to natural variation in porcine gene expression. Beside known and already confirmed genes, additional candidate genes, located in promising regions were identified. These promising candidate genes need further validation in other pig populations and the gene regulation networks have to be closer investigation.

3.7 Acknowledgments

Authors are grateful to Prof. Dr. med. Markus M. Nöthen from Institute of Human Genetics, University of Bonn, Germany for genotyping the samples via Illumina 60K porcine SNP chip.

This work was part of the DRIP phase II (Genetic- functional background of the water-holding capacity in pork) project and supported by the German Research Foundation (DFG), Germany, grant nr: SCHE562/13-1.

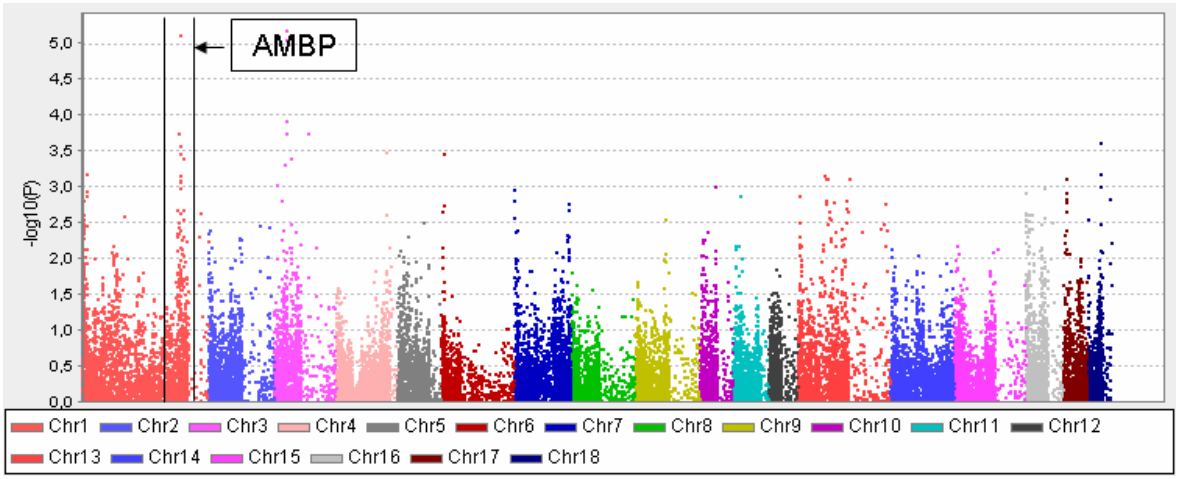

Figure 13: eQTL of polypeptide α-1-microglobulin / bikunin precursor (AMBP) across the autosomes associate with drip loss and is located on SSC 1 (position: 285708391 – 285725500, ensembl build 10.2)

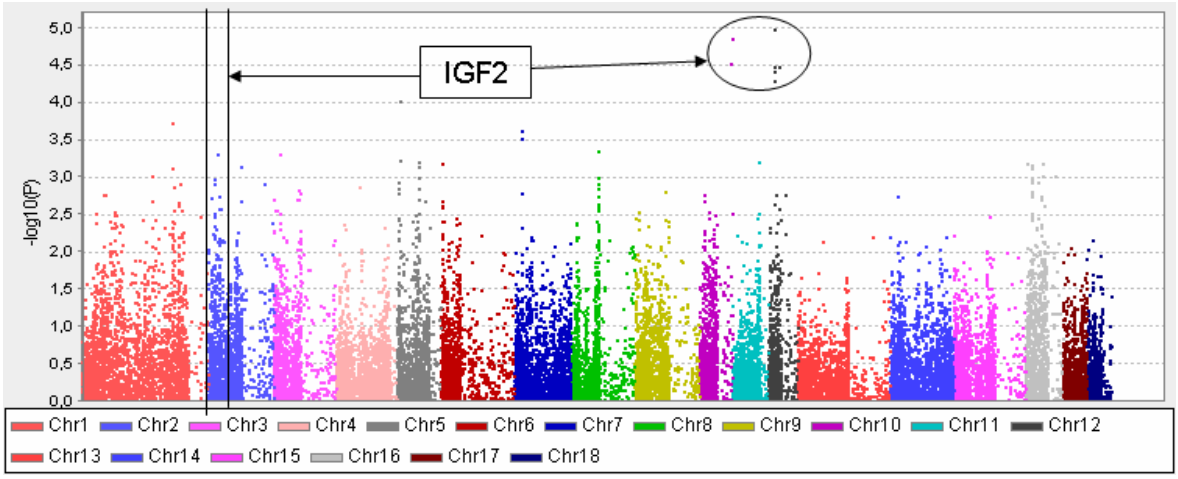

Figure 14: eQTL of gene Insulin -like growth factor 2 (IGF2) across the autosomes associate with drip loss and is located on SSC 2 (position: 2107672 – 2107381, ensembl built 10.2) but showed most significant eQTL on SSC 10 and 12

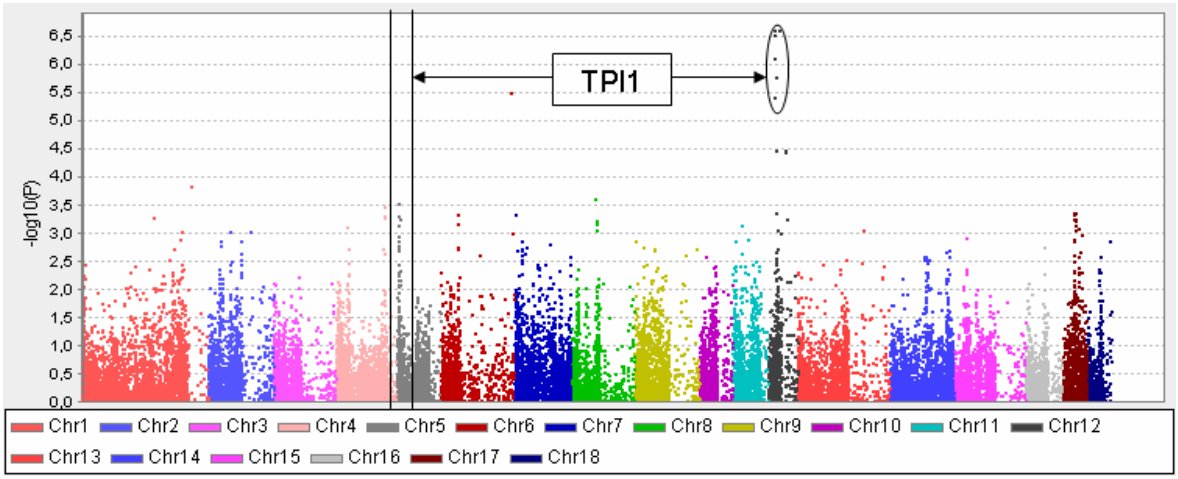

Figure 15: eQTL of gene triosephosphate isomerase1 (TPI1) across the autosomes associate with drip loss and is located on SSC 5 (position: 66274876 – 66278367, ensembl built 10.2) but showed most significant eQTL on SSC 12

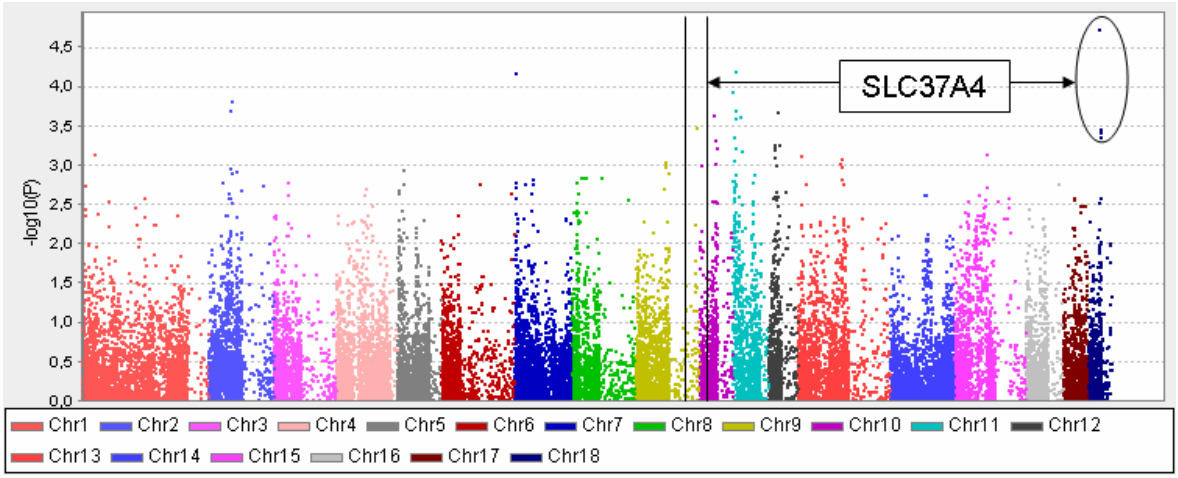

Figure 16: eQTL of gene solute carrier family 37, member 4 (SLC37A4) across the autosomes associate with drip loss and is located on SSC 9 (position: 51282372-51287727, ensembl built 10.2)

Table 7: Significant KEGG identifiers detected based on negative correlated transcripts with drip loss

KEGGID	P value	Odds Ratio	Exp Count	Count (Size)	Term
4650	0.0043	5.2191	1.0893	5 (66)	Natural killer cell mediated cytotoxicity
4920	0.0055	4.8926	1.1554	5 (70)	Adipocytokine signalling pathway
5160	0.0057	4.0985	1.6505	6 (100)	Hepatitis C
4660	0.0083	4.4085	1.2709	5 (77)	T cell receptor signalling pathway
5220	0.0115	5.0224	0.8913	4 (54)	Chronic myeloid leukaemia
4012	0.0123	4.9226	0.9078	4 (55)	ErbB signalling pathway
4115	0.0194	4.2459	1.0398	4 (63)	p53 signalling pathway
4380	0.0227	3.3566	1.6340	5 (99)	Osteoclast differentiation
5213	0.0257	5.1639	0.6437	3 (39)	Endometrial cancer
5214	0.0274	5.0230	0.6602	3 (40)	Glioma
4662	0.0293	4.8895	0.6767	3 (41)	B cell receptor signalling pathway
4370	0.0293	4.8895	0.6767	3 (41)	VEGF signalling pathway
4610	0.0342	3.5168	1.2379	4 (75)	Complement and coagulation cascades
5221	0.0372	4.4190	0.7427	3 (45)	Acute myeloid leukaemia
4940	0.0393	4.3151	0.7592	3 (46)	Type I diabetes mellitus
5215	0.0419	3.2810	1.3204	4 (80)	Prostate cancer
600	0.0421	6.8051	0.3301	2 (20)	Sphingolipid metabolism
650	0.0421	6.8051	0.3301	2 (20)	Butanate metabolism
4010	0.0445	2.4945	2.6078	6 (158)	MAPK signalling pathway

A hyper geometric gene set test was performed based on negative correlated transcripts with drip loss. Over representation of gene sets defined by the KEGG database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Table 8: Significant KEGG identifiers detected based on positive correlated transcripts with drip loss

KEGGID	P value	Odds Ratio	Exp Count	Count (Size)	Term
4120	0.0002	3.8293	3.6610	12 (102)	Ubiquitin mediated proteolysis
10	0.0003	6.5513	1.3280	7 (37)	Glycolysis / Gluconeogenesis
620	0.0058	4.9405	1.1844	5 (33)	Pyruvat metabolism
3040	0.0068	2.7345	4.0558	10 (113)	Spliceosome
4130	0.0096	5.5038	0.8614	4 (24)	SNARE interactions in vesicular transport
310	0.0236	4.0691	1.1127	4 (31)	Lysine degradation
4115	0.0243	2.9112	2.2612	6 (63)	p53 signalling pathway
30	0.0330	4.8240	0.7178	3 (20)	Pentose phosphate pathway
230	0.0335	2.2907	3.7687	8 (105)	Purine metabolism
4720	0.0385	3.4286	1.2921	4 (36)	Long-term potentiation

A hyper geometric gene set test was performed based on positive correlated transcripts with drip loss. Over representation of gene sets defined by the KEGG database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Table 9: Transcripts with their number of potential cis- and trans-regulated eQTL and their correlation coefficient

Probe set id	Gene name	Chromosome[1]	Estimated correlation[2]	Number of eQTL
Ssc.11793.1.A1_at	ZNF79	1	0.255**	1
Ssc.1801.2.S1_at	FNBP1	1	-0.251**	1
Ssc.18261.1.S1_at	PSTPIP2	1	0.275***	45
Ssc.1894.1.S1_at	AMBP	1	-0.178*	4
Ssc.19564.1.S1_at	PCDH9	1	0.242**	943
Ssc.3339.1.S1_at	BCKDHB	1	0.252**	15
Ssc.16250.1.S2_at	IL1RN	2	-0.228**	5
Ssc.20525.1.S1_at	IGF2	2	-0.194*	14
Ssc.30674.1.S1_at	PTBP1	2	0.242**	12
Ssc.11624.1.A1_at	ATL2	3	0.239**	10
Ssc.14375.1.A1_at	RRM2B	4	0.223**	11
Ssc.1589.1.A1_at	JTB	4	0.235**	37
Ssc.1297.1.S1_at	TPI1	5	0.267**	17
Ssc.15905.1.A1_at	GDF11	5	0.240**	11
Ssc.16770.1.S1_at	EMG1	5	0.237**	59
Ssc.14396.1.S1_at	TAF12	6	0.226**	22
Ssc.8727.1.A1_at	RPRD1A	6	0.255**	21
Ssc.30641.1.S1_at	ATP5L	9	0.249**	10
Ssc.10209.1.A1_at	CHURC1	10	0.232**	12
Ssc.21622.1.A1_at	FBP2	10	0.172*	1
Ssc.3766.1.S1_at	RAB18	10	0.222**	4
Ssc.6339.1.A1_at	DHTKD1	10	0.230**	1
Ssc.7225.1.A1_at	BAG1	10	0.245**	7
Ssc.7523.1.A1_at	PHB	12	0.307***	5

Table 9 continued:

Probe set id	Gene name	Chromosome[1]	Estimated correlation[2]	Number of eQTL
Ssc.11661.1.A1_at	PPP1CC	14	0.210*	46
Ssc.17499.1.S1_at	NEURL	14	0.244**	86
Ssc.30435.1.A1_at	BAG3	14	0.273***	38
Ssc.1473.1.S1_at	SMARCD3	18	0.227**	3
Ssc.7292.1.S1_at	NA	18	0.234**	10

[1] Number of *Sus scrofa* chromosomes where the gene is located

[2]The potential candidate genes with their number of significant transcripts and their calculated correlation coefficient of Pearson with the phenotypic trait drip loss. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$

Table 10: Potential *cis*-regulated eQTL and their explained variance

Chromosome[1]	Probe set id	Gene name	Position of the transcript	SNP	Position of SNP	*P*-Value[2]	Variance[3]
1	Ssc.11793.1.A1_at	ZNF79	301852485 - 301870268	ALGA0010677	303486951	5.10E-07**	0.1745
1	Ssc.1894.1.S1_at	AMBP	285708391 - 285725500	ASGA0007079	287975124	7.85E-06*	0.1409
1	Ssc.19564.1.S1_at	PCDH9	86706491 - 86706515	ALGA0004442	86848196	0.0001701**	0.1019
4	Ssc.1589.1.A1_at	JTB	104615009 - 104617469	DIAS0000629	109076487	4.78E-05*	0.1181
5	Ssc.16770.1.S1_at	EMG1	66181482 - 66187680	ASGA0025815	66243536	1.18E-05*	0.1359
10	Ssc.3766.1.S1_at	RAB18	44073199 - 44073223	ALGA0058578	44475243	1.56E-06*	0.1609
14	Ssc.17499.1.S1_at	NEURL	124318583 - 124396585	ASGA0066211	124326965	5.73E-12***	0.3027
14	Ssc.17499.1.S1_at	NEURL	124318583 - 124396585	H3GA0042143	124382934	3.71E-05*	0.1214

[1]*Sus scrofa* chromosome; [2]Candidate genes with their potential *cis*-regulated eQTL, their explained variance and the calculated *P*-Value. *: FDR < 0.1, **: FDR < 0.05, ***: FDR < 0.01; [3]Gene expression variation explained by the associated SNP

Chapter 4 General discussion and conclusion

The impact of water-holding capacity traits and their biological background were investigated in a Duroc × Pietrain resource population. It was possible to identify several chromosomal regions and their potential candidate genes significantly associated with WHC using GWAS and genetical genomics approaches. In this study 9 QTL on six different porcine chromosomes (SSC2, 3, 6, 12, 15, 16 and 17) were significantly associated with drip loss.

In GWAS, until now the prediction of the total genetic values considers mainly additive effects (Meuwissen et al. 2001). However, numerous epistatic QTL pairs and dominance effects influencing meat quality traits such as WHC are identified, indicating that the genomic control of meat quality is a complex process involving numerous QTL as well as a complex network of gene interactions (Carlborg and Haley 2004; Duthie et al. 2010). Therefore, a successful implementation of epistatic effects leads to an increased squared coefficient of determination (Hu et al. 2011). Nevertheless, in our study, it was possible to identify epistatic QTL regions on SSC2, 3, 8 and 13 for pH1, which have been reported earlier in our population considering epistatic effects (Grosse-Brinkhaus *et al.* 2010) independently from a second QTL, because of the high number of genetic markers provided by the SNP chips. In the studies of Liu et al. (2007; 2008) and Edwards et al. (2008) dominance effects play in addition to additive effects an important role. However, in this study, dominance effects were not detected using GWAS like in other GWAS studies (Duijvesteijn et al. 2010; Ramos et al. 2011).

The additive effects as well as the proportion of explained phenotypic variance were calculated for drip loss, pH1, pH24 and pH decline via PLINK software. The calculated additive effects and the explained phenotypic variance for drip loss were quiet high in comparison to other QTL studies done in the same pig population (2.4.4) (Edwards et al. 2008; Liu et al. 2007). Reasons for the higher proportion of explained phenotypic variance could be the marker density of the SNP chips and the higher specificity in comparison other QTL studies using microsatellites, which were only partly spread across the genome. In this study, the size of the SNP effects for the pH traits were small, whereas markers in regions in which causative mutations lie have much larger effects, like the SNP, which are significantly associated with drip loss. However, the markers with the largest associated effects from a genomic analysis may not track perfectly the causative mutations; on the other hand, they are potentially useful tools for identifying chromosomal regions associated with traits (Cole *et al.* 2009).

Another reason for the small size of the SNP effects calculated for pH traits might be the population size of 169 animals. It has been shown that QTL mapping in small populations confound the effects of QTL with statistical artefacts caused by sampling. Therefore, the genetics of complex traits should be studied in large populations (Beavis 1994, 1998). Additionally, almost all QTL mapping procedures can detect QTL with large effects, but not all can detect QTL with intermediated and small effects. However, quantitative traits are defined as traits controlled primarily by intermediate and small effects (Xu 2003). Especially in GWAS, the number of animals depends on the size of the effects that one wishes to detect. The crucial parameter is the proportion of the variance explained by the SNP this parameter combines the allele frequency with the mean difference between the SNP genotypes (Goddard and Hayes 2009). In practice, some SNP explain more than 4 %of the genetic variance, and so a smaller experiment would be sufficient but, in fact, most SNP associated with complex traits explain less than 4 % of the genetic variance, and so over 1800 animals are required (Goddard and Hayes 2009; Visscher 2008).

In the second study, eQTL for drip loss were identified on several chromosomal regions (SSC1, 2, 3, 4, 5, 6, 9, 10, 12, 14 and 18), whereas 8 eQTL of 7 transcripts were observed as *cis*-regulated and found on SSC1, 4, 5, 10 and 14. The eQTL on SSC1, 2, 3, 4, 6, and 18 validated the results of Ponsuksili et al. (2008a; 2008b). These eQTL seemed to play an important role in the development of drip loss. In contrast the eQTL on SSC5, 9, 10, 12 and 14 were presumably novel. This could be due to the fact of the combination of the different techniques such as GWAS and genetical genomics.

The recent study is based on 132 F_2 DuPi animals. In comparison to other published studies in livestock (Le Mignon et al. 2009; Liaubet et al. 2011; Ponsuksili et al. 2008a; Ponsuksili et al. 2008b; Ponsuksili et al. 2010) the number of animals used in this study is quite high or in the same range (Ponsuksili et al. 2011; Steibel et al. 2011). However, the number of animals used for eQTL studies are still too small in comparison to QTL studies and providing less power to detect individual eQTL, because of the higher costs (de Koning et al. 2005; Haley and de Koning 2006). That implies a minimum population size of 200 F_2 to detect major QTL effects, while an experiment of 400 animals gives good statistics to detect also smaller effects. Therefore, by combining the two methods GWAS and genetical genomics the power could be increased, because of the marker density of both chips.

With the development of high-throughput sequencing and genotyping technologies, the numbers of markers collected in genetic association studies is growing rapidly, increasing the importance of methods for correcting for multiple hypotheses testing (Han et al. 2009). While the Bonferroni correction provides the simplest way to correct for multiple testing by assuming independence between markers, permutation testing is widely considered the golden standard for accurately correcting for multiple testing (Westfall and Young 1993). However, the Bonferroni correction ignores correlation between markers due to linkage equilibrium and leads to an overly conservative correction of false positives, which is exacerbated as the marker density increases and affects the statistical power (Han et al. 2009). Therefore, a correction for multiple testing was not performed in the first study because of high linkage disequilibrium of the SNP on SSC2 significantly associated with the candidate genes NELL1 and SOX6. These QTL have been identified earlier in the DuPi population. Liu et al. (2007; 2008) identified QTL on SSC2, SSC3 and SSC6 related to drip loss. Because of the repeated detection of these QTL, it can be assumed that they play an important role in the expression of drip loss. In the second study, instead of a genome wide Bonferroni correction, which might be overly conservative, a more reluctant threshold using a correction following Benjamini and Hochberg (1995) with a FDR < 0.10 was set. Since Storey and Tibshirani (2003) propose an adaptation of the false discovery rate (FDR) targeted at genome-wide experiments to provide a better balance between statistical stringency and power to detect true effects. However, for some genes even this method seems to be too stringent, e.g. for the gene solute carrier family 37, member 4 (SLC37A4). This gene is located in a region on SSC 9 and showed eQTL on SSC 18 where several promising candidate gene influencing drip loss are discussed (Jennen et al. 2007).

Mapping quantitative traits and unravelling transcriptional control are challenging, when applied to one phenotype at a time. In studies of typical quantitative porcine traits like WHC, strong effects are rarely found. At first, applying a genome wide association study using the 60K porcine SNP panel allowed investigating the genetic background of WHC traits. Several regions in the genome were identified affecting drip loss, pH1, pH24 and pH decline in the DuPi population. Especially, two regions of SSC2 showed high LD and were clustered in two spans of 0.5 Mb. In these clusters, only two genes were located NELL1 and SOX6, which were associated with drip loss. Most of the candidate genes were functionally related to WHC traits. However, there are several genes, which were not involved in the expression of muscle properties. It can be assumed that these genes have

high LD to candidate genes, which were not covered by the SNP chip. Nevertheless, it could be also possible that the significant SNP associated with these traits were annotated in the wrong chromosomal region, which means that the candidate genes are possibly not the right ones. In the second step, genomic technologies for expression profiling with genome-wide genetic mapping using SNP markers were coupled, and shown that specific chromosomal regions contain functional candidate genes. These approaches and results allowed to investigate and to dissect the genetic contribution to natural variation in porcine gene expression. Beside known and already confirmed genes like AMBP and IGF2, additional candidate genes (TPI1 and SLC37A4), located in promising regions were identified. (Ponsuksili et al. 2010). 7 of 29 transcripts were observed as putative *cis*-regulated following the criteria used in the study of Ponsuksili et al. (2011). Therefore, 24 % of the significantly associated transcripts were *cis*-regulated and 76 % were *trans*-regulated.

In the studies of Christensen (2003) and Otto et al. (2004) it was shown that the proportion of drip loss varies between the different measurement positions in *M. long. dorsi*. Drip loss decreased from caudal end to the cranial end of the *M. long. dorsi*. In other muscle groups such variations in drip loss could also be observed. Therefore, it could be helpful to validate the results of this study in other muscle such as *Musculus semimembranosus* in order to characterize trait specific mutations.

The validation and confirmation of the recent results in other pig populations is indispensable. The candidate genes AMBP and IGF2 have been already validated (Cinar et al. 2012; Ponsuksili et al. 2008b) and might be promising for a marker-assisted selection of WHC traits.

In general, further fine mapping and next generation sequencing technologies are requested to identify the causative mutations of the underlying QTL regions (Cole et al. 2009). Validations of candidate genes are warranted in other pig populations to prove their role in meat quality traits and the gene regulation networks have to be closer investigated. Additionally, proteomic profiling of WHC traits in loin could be used to identify metabolites and biological background of WHC.

Chapter 5 Summary

Water holding capacity (WHC), i.e. the capacity of meat to retain its water during application of external forces (Hamm 1985) is an important quality criterion for the meat processing industries and the consumers. The structure and the biological processes of WHC and its associated traits have been extensively investigated (Greaser 2001; Huff-Lonergan and Lonergan 2005, 2007; Huff Lonergan et al. 2010; Offer and Cousins 1992; Rosenvold and Andersen 2003). However, the genetic mechanisms underlying in WHC during the conversion of muscle to meat are not fully understood.

Therefore, two studies were performed to investigate WHC traits in the DuPi population. The aim of the first study was to detect QTL regions and promising candidate genes for WHC traits using a genome-wide association approach. The proportion of explained phenotypic variance determined the importance of the identified QTL regions. In the second part, the correlation between drip loss and the expression level was calculated and the biological background of WHC was observed.

In the first step, 169 F_2 DuPi animals were genotyped and four meat quality traits (drip loss, pH1, pH24 and pH decline) were used to investigate the genetic background of WHC. 49 significant SNP were observed ($P < 0.001$) for drip loss, 40 SNP for pH1, 9 SNP for pH24 and 33 SNP for pH decline in loin. Analyses revealed 14 putative functional candidate genes significantly associated with drip loss. 26, 7 and 22 candidate genes were identified for pH1, pH24 and pH decline, respectively. The genes NELL1 and SOX6 located on SSC2 were significantly associated with drip loss and showed more than 3 point-mutations each with high linkage disequilibrium. The proportion of explained phenotypic variance ranged between 4.4 % and 8.43 % for identified SNP of all four traits. Applying a genome wide association analysis using the 60K porcine SNP panel allowed investigating the genetic background of WHC traits. Several regions in the genome were identified affecting drip loss, pH1, pH24 and pH decline in the DuPi population. In GWAS, the most natural trait to be associated with the variation in the genome should be the product of the transcribed genes (mRNA). Such an approach combines GWAS with genetical genomics for investigating the biological background of the complex traits such as WHC.

Therefore, in the second step, WHC was characterized by drip loss measured in *M. longissimus dorsi*. Performing genome wide associations studies (GWAS) of transcriptional profiles for 132 F_2 DuPi animals, which were correlated with phenotypes allows elucidating *cis*- and *trans*-regulation. Expression levels of 1228 genes were

significantly correlated with drip loss and were further analyzed for enrichment of functional annotation groups as defined by gene ontology and KEGG pathways. A hyper geometric gene set enrichment test was performed and revealed glycolysis/glyconeogenesis, pentose phosphate pathway and pyruvat metabolism as most promising pathways. For 267 selected transcripts eQTL analyses was performed and revealed 1541 significant associations in total. Because of positional accordance of the gene underlying transcript and the eQTL location, it was possible to identify 8 eQTL that can be assumed to be *cis*-regulated. Comparing the results of gene set enrichment and the eQTL detection tests molecular networks and potential candidate genes, which seem to play key roles in the expression of WHC, were detected. The *α-1-microglobulin / bikunin precursor* (AMBP) gene was assumed to be *cis*-regulated, is part of the gycolyse metabolism, and has been identified in previous studies.

In conclusion, applying a genome wide association analysis using the 60K porcine SNP panel allowed investigating the genetic background of WHC traits in this study. Several regions in the genome were identified affecting drip loss, pH1, pH24 and pH decline in the DuPi population. Combing the genome-wide association analysis with the genetical genomics approach supports to identify trait-associated SNP and to understand the biology of complex traits. In this study, several promising candidate genes (NELL1, SOX6 and AMBP) could be identified affecting WHC. However, further investigations such as RNA deep sequencing will still needed to validate the results of this study.

Chapter 6 References

Affymetrix Inc. CJM plier: Implements the Affymetrix PLIER algorithm. .

AffymetrixTechnicalNote (2005) Guide to probelogarithmic intensity error (PLIER) estimation. Affymetrix, Santa Clara, CA

Ali S, Chopra R, Manvati S, Singh YP, Kaul N, Behura A, Mahajan A, Sehajpal P, Gupta S, Dhar MK, Chainy GBN, Bhanwer AS, Sharma S, Bamezai RNK (2013) Replication of type 2 diabetes candidate genes variations in three geographically unrelated Indian population groups. PLoS One 8, e58881

Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. 2, 130-138

Andersson L (2007) The molecular basis for phenotypic changes during pig domestication. In Pigs and humans: 10,000 years of interaction, U. Albarella, K. Dobney, A. Ervynck, P. Rowley-Conwy, eds. (New York: Oxford Universtiy Press), pp 42-54

Andersson L (2009) Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. Genetica 136, 341-349

Andersson L, Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. Nat. Rev. Genet. 5, 202-212

Arion WJ, Lange AJ, Walls HE, Ballas LM (1980) Evidence for the participation of independent translocation for phosphate and glucose 6-phosphate in the microsomal glucose-6-phosphatase system. Interactions of the system with orthophosphate, inorganic pyrophosphate, and carbamyl phosphate. J. Biol. Chem. 255, 396-406

Babak T, Garrett-Engele P, Armour CD, Raymond CK, Keller MP, Chen R, Rohl CA, Johnson JM, Attie AD, Fraser HB, Schadt EE (2010) Genetic validation of whole-transcriptome sequencing for mapping expression affected by cis-regulatory variation. BMC Genomics 11, 473

Baron D, Magot A, Ramstein G, Steenman M, Fayet G, Chevalier C, Jourdon P, Houlgatte R, Savagner F, Pereon Y (2011) Immune response and mitochondrial metabolism are commonly deregulated in DMD and aging skeletal muscle. PLoS One 6, e26952

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265

Beavis WD (1994) The power to detect and deceit of QTL experiments: lessons from comparitive QTL studies. In Forty-Ninth Annual Corn & Sorghum Industry Research Conference (Washington, DC.: Seed Trait Association), pp 250-266

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In Molecular Dissection of Complex Traits, A.H. Paterson, ed. (New York: CRC Press), pp 145-162

Becker D, Tetens J, Brunner A, Burstel D, Ganter M, Kijas J, Drögemüller C, Consortium ISG (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. PLoS One 5, e8689

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 57, 289-300

Bertram HC, Petersen JS, Andersen HJ (2000) Relationship between RN- genotype and drip loss in meat from Danish pigs. Meat Sci. 56, 49-55

Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc. Natl. Acad. Sci. U. S. A. 101, 2173-2178

Borchers N, Otto G, Kalm E (2007) Genetic relationship of drip loss to further meat quality traits in purebred Pietrains. Arch. Tierz.-Arch. Anim. Breed. 50, 84-91

Brigelius-Flohe R (2009) Commentary: oxidative stress reconsidered. Genes Nutr 4, 161-163

Brocks L, Klont RE, Buist W, de Greef K, Tieman M, Engel B (2000) The effects of selection of pigs on growth rate vs leanness on histochemical characteristics of different muscles. J. Anim. Sci. 78, 1247-1254

Brunner RM, Srikanchai T, Murani E, Wimmers K, Ponsuksili S (2012) Genes with expression levels correlating to drip loss prove association of their polymorphism with water holding capacity of pork. Mol. Biol. Rep. 39, 97-107

Cadieu E, Ostrander EA (2007) Canine genetics offers new mechanisms for the study of human cancer. Cancer Epidemiol. Biomarkers Prev. 16, 2181-2183

Cánovas A, Pena RN, Gallardo D, Ramírez O, Amills M, Quintanilla R (2012) Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population. PLoS One 7, e35583

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat. Rev. Genet. 5, 618-625

Cerletti M, Molloy MJ, Tomczak KK, Yoon S, Ramoni MF, Kho AT, Beggs AH, Gussoni E (2006) Melanoma cell adhesion molecule is a novel marker for human fetal myogenic cells and affects myoblast fusion. J. Cell Sci. 119, 3117-3127

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet JC, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genet. 40, 449-454

Cheng Y, Quinn JF, Weiss LA (2013) An eQTL mapping approach reveals that rare variants in the SEMA5A regulatory network impact autism risk. Hum. Mol. Genet., Epub ahead of print

Christensen LB (2003) Drip loss sampling in porcine m. longissimus dorsi. Meat Sci. 63, 469-477

Cinar MU, Kayan A, Uddin MJ, Jonas E, Tesfaye D, Phatsara C, Ponsuksili S, Wimmers K, Tholen E, Looft C, Jungst H, Schellander K (2012) Association and expression quantitative trait loci (eQTL) analysis of porcine AMBP, GC and PPP1R3B genes with meat quality traits. Mol. Biol. Rep. 39, 4809-4821

Ciobanu DC, Lonergan SM, Huff-Lonergan EJ (2011) Genetics of meat quality and carcass traits. In The Genetics of the Pig, M.F. Rothschild, A. Ruvinsky, eds. (New York: CAB International)

Cole JB, VanRaden PM, O'Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR (2009) Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92, 2931-2946

Connell S, Meade KG, Allan B, Lloyd AT, Downing T, O'Farrelly C, Bradley DG (2013) Genome-wide association analysis of avian resistance to Campylobacter jejuni colonization identifies risk locus spanning the CDH13 gene. G3 3, 881-890

Connor F, Wright E, Denny P, Koopman P, Ashworth A (1995) The Sry-related Hmg box-containing gene Sox6 is expressed in the adult testis and developing nervous-system of the mouse. Nucleic Acids Res. 23, 3365-3372

Cox DG, Kraft P (2006) Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. Hum. Hered. 61, 10-14

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J. Dairy Sci. 91, 3225-3236

de Koning DJ, Carlborg O, Haley CS (2005) The genetic dissection of immune response using gene-expression studies and genome mapping. Vet. Immunol. Immunopathol. 105, 343-352

de Koning DJ, Harlizius B, Rattink AP, Groenen MAM, Brascamp EW, van Arendonk JAM (2001) Detection and characterization of quantitative trait loci for meat quality traits in pigs. J. Anim. Sci. 79, 2812-2819

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Arcelus MG, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325, 1246-1250

Dittmer J (2003) The biology of the Ets1 proto-oncogene. Mol. Cancer 2, 29

Dorn CR (1976) Epidemiology of canine and feline tumors. Comp Cont Educ Pract Vet 12, 307-312

Drögemüller C, Reichart U, Seuberlich T, Oevermann A, Baumgartner M, Boghenbor KK, Stoffel MH, Syring C, Meylan M, Muller S, Muller M, Gredler B, Solkner J, Leeb T (2011) An unusual splice defect in the mitofusin 2 gene (MFN2) is associated with degenerative axonopathy in Tyrolean Grey cattle. PLoS One 6, e18931

Duan YY, Ma JW, Yuan F, Huang LB, Yang KX, Xie JP, Wu GZ, Huang LS (2009) Genome-wide identification of quantitative trait loci for pork temperature, pH decline, and glycolytic potential in a large-scale White Duroc x Chinese Erhualian resource population. J. Anim. Sci. 87, 9-16

Duijvesteijn N, Knol EF, Merks JWM, Crooijmans RPMA, Groenen MAM, Bovenhuis H, Harlizius B (2010) A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet. 11, 42

Durinck S, Huber W, Davis S (2012) Interface to BioMart databases (e.g. Ensembl, Wormbase and Gramene). (Bioconductor http://bioconductor.org)

Duthie C-A, Simm G, Doeschl-Wilson A, Kalm E, Knap PW, Roehe R (2010) Quantitative trait loci for meat quality traits in pigs considering imprinting and epistatic effects. Meat Sci. 87, 394-402

Edwards DB, Ernst CW, Raney NE, Doumit ME, Hoge MD, Bates RO (2008) Quantitative trait locus mapping in an F-2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. J. Anim. Sci. 86, 254-266

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher JR, Reitman ML, Kong A, Schadt EE, Stefansson K (2008) Genetics of gene expression and its effect on disease. Nature 452, 423 - 428

Fan B, Onteru SK, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF (2011) Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PLoS One 6, e14726

Fischer K (2007) Drip loss in pork: influencing factors and relation to further meat quality traits. J. Anim. Breed. Genet. 124, 12-18

Fu JY, Wolfs MGM, Deelen P, Westra HJ, Fehrmann RSN, Meerman GJT, Buurman WA, Rensen SSM, Groen HJM, Weersma RK, van den Berg LH, Veldink J, Ophoff RA, Snieder H, van Heel D, Jansen RC, Hofker MH, Wijmenga C, Franke L (2012) Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 8, e1002431

Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, O'Brien PJ, MacLennan DH (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 253, 448 - 451

Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S (2007) Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm. Genome 18, 463-472

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R,

Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296, 2225-2229

Gaffney D, Veyrieras J-B, Degner J, Pique-Regi R, Pai A, Crawford G, Stephens M, Gilad Y, Pritchard J (2012) Dissecting the regulatory architecture of gene expression QTLs. Genome Biol. 13, R7

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge YC, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang JH (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80

Gerrits A, Li Y, Tesson BM, Bystrykh LV, Weersing E, Ausema A, Dontje B, Wang XS, Breitling R, Jansen RC, de Haan G (2009) Expression quantitative trait loci are highly sensitive to cellular differentiation state. PLoS Genet. 5, e10000692

Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 6, e10000952

Glubb D, Dholakia N, Innocenti F (2012) Liver expression quantitative trait loci: a foundation for pharmacogenomic research. Frontiers in Genetics 3, 153

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10, 381-391

Göring HHH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JBM, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nature Genet. 39, 1208-1216

Graw J (2003) The genetic and molecular basis of congenital eye defects. Nat. Rev. Genet. 4, 876-888

Greaser ML (2001) Postmortem muscle chemistry In Meat science and applications, Y.H. Hui, W.-k. Nip, R.W. Roger, O.A. Young, eds. (New York: Marcel Dekker, Inc.)

Greenawalt DM, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J, Zhang B, Castro V, Zhu J, Sieberts SK, Wang S, Molony C, Heymsfield SB, Kemp DM, Reitman ML, Lum PY, Schadt EE, Kaplan LM (2011) A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. 21, 1008-1016

Grewal JS, Tsai JY, Khan SR (2005) Oxalate-inducible AMBP gene and its regulatory mechanism in renal tubular epithelial cells. Biochem. J. 387, 609-616

Grosse-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, Jungst H, Looft C, Schellander K, Tholen E (2010) Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc x Pietrain population. Genet. Sel. Evol. 42

Hagiwara N, Ma B, Ly A (2005) Slow and fast fiber isoform gene expression is systematically altered in skeletal muscle of the Sox6 mutant, p(100H). Dev. Dyn. 234, 301-311

Hagiwara N, Yeh M, Liu A (2007) Sox6 is required for normal fiber type differentiation of fetal skeletal muscle in mice. Dev. Dyn. 236, 2062-2076

Haley C, de Koning DJ (2006) Genetical genomics in livestock: potentials and pitfalls. Anim. Genet. 37 Suppl 1, 10-12

Hamill RM, McBryan J, Mcgee C, Mullen AM, Sweeney T, Talbot A, Cairns MT, Davey GC (2012) Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. Meat Sci. 92, 440-450

Hamm R (1985) Wasserbindugsvermögen des Fleisches - Aspekte eines wichtigen Qualitätsmerkmals. Mitteilungsblatt *BAFF* 88, 6383-6387

Han B, Kang HM, Eskin E (2009) Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. PLoS Genet. 5, e10000456

Hasenstein JR, Hassen AT, Dekkers JCM, Lamont SJ (2008) High resolution, advanced intercross mapping of host resistance to Salmonella colonization. In AGAH, pp 213-218

Hazra A, Kraft P, Selhub J, Giovannucci EL, Thomas G, Hoover RN, Chanock SJ, Hunter DJ (2008) Common variants of FUT2 are associated with plasma vitamin B-12 levels. Nature Genet. 40, 1160-1162

Heap GA, Trynka G, Jansen RC, Bruinenberg M, Swertz MA, Dinesen LC, Hunt KA, Wijmenga C, vanHeel DA, Franke L (2009) Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC Med. Genomics 2, 1

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95-108

Honikel KO (1987) Water binding-capacity of meat. Fleischwirtschaft 67, 418

Honikel KO, Kim CJ, Hamm R, Roncales P (1986) Sarcomere shortening of prerigor muscles and its influence on drip loss. Meat Sci. 16, 267-282

Hu Z, Park CA, Fritz ER, Reecy JM (2010) QTLdb: A comprehensive database tool building bridges between genotypes and phenotypes. In The 9th World Congress on Genetics Applied to Livestock Production (Leipzig, Germany)

Hu ZQ, Li YG, Song XH, Han YP, Cai XD, Xu SZ, Li WB (2011) Genomic value prediction for quantitative traits under the epistatic model. BMC Genet. 12, 15

Huff-Lonergan E, Lonergan SM (1999) Postmortem mechanisms of meat tenderization: the roles of the structural proteins and calpain system. In Quality attributes of muscle foods, Y.L. Xiong, C.-T. Ho, F. Shahidi, eds. (New York: Kluwer Academic / Plenum Publishers), pp 229-251

Huff-Lonergan E, Lonergan SM (2005) Mechanisms of water-holding capacity of meat: The role of postmortem biochemical and structural changes. Meat Sci. 71, 194-204

Huff-Lonergan E, Lonergan SM (2007) New frontiers in understanding drip loss in pork: recent insights on the role of postmortem muscle biochemistry. J. Anim. Breed. Genet. 124, 19-26

Huff-Lonergan E, Mitsuhashi T, Beekman DD, Parrish FC, Olson DG, Robson RM (1996) Proteolysis of specific muscle structural proteins by mu-calpain at low pH and temperature is similar to degradation in postmortem bovine muscle. J. Anim. Sci. 74, 993-1008

Huff Lonergan E, Zhang W, Lonergan SM (2010) Biochemistry of postmortem muscle - Lessons on mechanisms of meat tenderization. Meat Sci. 86, 184-195

Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu WQ, Lin YS, Moloney C, Aldred SF, Trinklein ND, Schuetz E, Nickerson DA, Thummel KE, Rieder MJ, Rettie AE, Ratain MJ, Cox NJ, Brown CD (2011) Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 7, e1002078

Jansen RC (2003) Studying complex biological systems using multifactorial perturbation. Nat. Rev. Genet. 4, 145-151

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet. 17, 388-391

Jansen RC, Nap JP (2004) Regulating gene expression: surprises still in store. Trends Genet. 20, 223-225

Jennen DGJ, Brings AD, Liu G, Jungst H, Tholen E, Jonas E, Tesfaye D, Schellander K, Phatsara C (2007) Genetic aspects concerning drip loss and water-holding capacity of porcine meat. J. Anim. Breed. Genet. 124, 2-11

Jeon JT, Carlborg O, Tornsten A, Giuffra E, Amarger V, Chardon P, Andersson-Eklund L, Andersson K, Hansson I, Lundstrom K, Andersson L (1999) A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nature Genet. 21, 157-158

Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118-127

Jonas E, Schreinemachers H-J, Kleinwächter T, Ün C, Oltmanns I, Tetzlaff S, Jennen D, Tesfaye D, Ponsuksili S, Murani E, Juengst H, Tholen E, Schellander K, Wimmers K (2008) QTL for the heritable inverted teat defect in pigs. Mamm. Genome 19, 127-138

Ka S, Markljung E, Ring H, Albert FW, Harun-Or-Rashid M, Wahlberg P, Garcia-Roves PM, Zierath JR, Denbow DM, Paabo S, Siegel PB, Andersson L, Hallbook F (2013) Expression of carnitine palmitoyl-CoA transferase-1B is influenced by a cis-acting eQTL in two chicken lines selected for high and low body weight. Physiol. Genomics 1, 367-376

Kabakchiev B, Silverberg MS (2013) Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology 144, 1488-1496

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27-30

Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kampe O, Hedhammar A, Lander ES, Andersson R, Andersson L, Lindblad-Toh K (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nature Genet. 39, 1321-1328

Karns R, Zhang G, Sun GY, Indugula SR, Cheng H, Havas-Augustin D, Novokmet N, Rudan D, Durakovic Z, Missoni S, Chakraborty R, Rudan P, Deka R (2012) Genome-wide association of serum uric acid concentration: replication of sequence variants in an island population of the Adriatic coast of Croatia. Ann. Hum. Genet. 76, 121-127

Karyadi DM, Karlins E, Decker B, vonHoldt BM, Carpintero-Ramirez G, Parker HG, Wayne RK, Ostrander EA (2013) A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. PLoS Genet. 9, e1003409

Kauffman RG, Sybesma W, Smulders FJM, Eikelenboom G, Engel B, van Laack RLJM, Hoving-Bolink AH, Sterrenburg P, Nordheim EV, Walstra P, van der Wal PG (1993) The effectiveness of examining early post-mortem musculature to predict ultimate pork quality. Meat Sci. 34, 283-300

Khanna C, Lindblad-Toh K, Vail D, London C, Bergman P, Barber L, Breen M, Kitchell B, McNeil E, Modiano JF, Niemi S, Comstock KE, Ostrander E, Westmoreland S, Withrow S (2006) The dog as a cancer model. Nat. Biotechnol. 24, 1065-1066

Kim JJ, Zhao HH, Thomsen H, Rothschild MF, Dekkers JCM (2005) Combined line-cross and half-sib QTL analysis of crosses between outbred lines. Genet. Res. 85, 235-248

Kim S, Cho H, Lee D, Webster MJ (2012) Association between SNPs and gene expression in multiple regions of the human brain. Transl Psychiatry 2, e113

Kizilkaya K, Tait R, Garrick D, Fernando R, Reecy J (2013) Genome-wide association study of infectious bovine keratoconjunctivitis in Angus cattle. BMC Genet. 14, 23

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308, 385-389

Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS (2008) A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein Bulls. J. Dairy Sci. 91, 2844-2856

Kraft P, Schadt E, Aten J, Horvath S (2003) A family-based test for correlation between gene expression and trait values. Am. J. Hum. Genet. 72, 1323-1330

Kuroda S, Tanizawa K (1999) Involvement of epidermal growth factor-like domain of NELL proteins in the novel protein-protein interaction with protein kinase C. Biochem. Biophys. Res. Commun. 265, 752-757

Kwasiborski A, Sayd T, Chambon C, Sante-Lhoutellier V, Rocha D, Terlouw C (2008) Pig Longissimus lumborum proteome: part II: relationships between protein content and meat quality. Meat Sci. 80, 982-996

Lalonde E, Ha KCH, Wang ZB, Bemmo A, Kleinman CL, Kwan T, Pastinen T, Majewski J (2011) RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res. 21, 545-554

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265, 2037-2048

Le Bihan-Duval E, Nadaf J, Berri C, Pitel F, Graulet B, Godet E, Leroux SY, Demeure O, Lagarrigue S, Duby C, Cogburn LA, Beaumont CM, Duclos MJ (2011) Detection of a cis eQTL controlling BMCO1 gene expression leads to the Iidentification of a QTG for chicken breast meat color. PLoS One 6, e14825

Le Mignon G, Desert C, Pitel F, Leroux S, Demeure O, Guernec G, Abasht B, Douaire M, Le Roy P, Lagarrigue S (2009) Using transcriptome profiling to characterize QTL regions on chicken chromosome 5. BMC Genomics 10, 575

Leuzzi R, Banhegyi G, Kardon T, Marcolongo P, Capecchi PL, Burger HJ, Benedetti A, Fulceri R (2003) Inhibition of microsomal glucose-6-phosphate transport in human neutrophils results in apoptosis: a potential explanation for neutrophil dysfunction in glycogen storage disease type 1b. Blood 101, 2381-2387

Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, Vasan RS, Mitchell GF (2007) Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med. Genet. 8 Suppl 1, 53

Li Q, Seo J-H, Stranger B, McKenna A, Pe'er I, LaFramboise T, Brown M, Tyekucheva S, Freedman MatthewÂ L (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell 152, 633-641

Liaubet L, Lobjois V, Faraut T, Tircazes A, Benne F, Iannuccelli N, Pires J, Glenisson J, Robic A, Le Roy P, SanCristobal M, Cherel P (2011) 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

Lillehammer M, Hayes BJ, Meuwissen THE, Goddard ME (2009) Gene by environment interactions for production traits in Australian dairy cattle. J. Dairy Sci. 92, 4008-4017

Lin BC, Pan CJ, Chou JY (2000) Human variant glucose-6-phosphate transporter is active in microsomal transport. Hum. Genet. 107, 526-529

Liu G, Jennen DGJ, Tholen E, Juengst H, Kleinwachter T, Holker M, Tesfaye D, Un G, Schreinemachers HJ, Murani E, Ponsuksili S, Kim JJ, Schellander K, Wimmers K (2007) A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Anim. Genet. 38, 241-252

Liu GS, Kim JJ, Jonas E, Wimmers K, Ponsuksili S, Murani E, Phatsara C, Tholen E, Juengst H, Tesfaye D, Chen JL, Schellander K (2008) Combined line-cross and half-sib QTL analysis in Duroc-Pietrain population. Mamm. Genome 19, 429-438

Lobjois V, Liaubet L, SanCristobal M, Glenisson J, Feve K, Rallieres J, Le Roy P, Milan D, Cherel P, Hatey F (2008) A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. Anim. Genet. 39, 147-162

Luo W, Chen S, Cheng D, Wang L, Li Y, Ma X, Song X, Liu X, Li W, Liang J, Yan H, Zhao K, Wang C, Zhang L (2012a) Genome-wide association study of porcine hematological parameters in a Large White x Minzhu F2 resource population. Int J Biol Sci 8, 870-881

Luo W, Cheng D, Chen S, Wang L, Li Y, Ma X, Song X, Liu X, Li W, Liang J, Yan H, Zhao K, Wang C, Zhang L (2012b) Genome-wide association analysis of meat quality traits in a porcine Large White x Minzhu intercross population. Int J Biol Sci 8, 580-595

Mackay TFC (2001) Quantitative trait loci in Drosophila. Nat. Rev. Genet. 2, 11-20

Majewski J, Pastinen T (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends Genet. 27, 72-79

Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. 77, 685-693

Markljung E, Braunschweig MH, Karlskov-Mortensen P, Bruun CS, Sawera M, Cho IC, Hedebro-Velander I, Josell A, Lundstrom K, von Seth G, Jorgensen CB, Fredholm M, Andersson L (2008) Genome-wide identification of quantitative trait loci in a cross between Hampshire and Landrace II: Meat quality traits. BMC Genet. 9, 22

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JP (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9, 356-369

Merlo DF, Rossi L, Pellegrino C, Ceppi M, Cardellino U, Capurro C, Ratto A, Sambucco PL, Sestito V, Tanara G, Bocchini V (2008) Cancer incidence in pet dogs: Findings of the animal tumor registry of Genoa, Italy. J. Vet. Intern. Med. 22, 976-984

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829

Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundstrom K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P, Andersson L (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science 288, 1248 - 1251

Moffatt MF, Kabesch M, Liang LM, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SAG, Wong KCC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WOC (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 448, 470-473

Musaro A, Fulle S, Fano G (2010) Oxidative stress and muscle homeostasis. Curr Opin Clin Nutr Metab Care 13, 236-242

Nezer C, Collette C, Moreau L, Brouwers B, Kim JJ, Giuffra E, Buys N, Andersson L, Georges M (2003) Haplotype sharing refines the location of an imprinted quantitative trait locus with major effect on muscle mass to a 250-kb chromosome segment containing the porcine IGF2 gene. Genetics 165, 277-285

Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M (1999) An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nature Genet. 21, 155-156

Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888

O'Seaghdha CM, Yang Q, Glazer NL, Leak TS, Dehghan A, Smith AV, Kao WHL, Lohman K, Hwang S-J, Johnson AD, Hofman A, Uitterlinden AG, Chen Y-DI, The GC, Brown EM, Siscovick DS, Harris TB, Psaty BM, Coresh J, Gudnason V, Witteman JC, Liu YM, Kestenbaum BR, Fox CS, Köttgen A (2010) Common variants in the calcium-sensing receptor gene are associated with total serum calcium levels. Hum. Mol. Genet. 19, 4296-4303

Offer G, Cousins T (1992) The mechanism of drip production - formation of 2 compartments of extracellular-space in muscle postmortem. J. Sci. Food Agric. 58, 107-116

Offer G, Knight P, eds. (1988) The structural basis of water-holding capacity in meat. Part 1: general principles and water uptake in meat processing. (New York: Elsevier Applied Science)

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. Nat. Neurosci. 11, 1271-1282

Onteru SK, Fan B, Nikkila MT, Garrick DJ, Stalder KJ, Rothschild MF (2011) Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89, 988-995

Ortenblad N, Young JF, Oksbjerg N, Nielsen JH, Lambert IH (2003) Reactive oxygen species are important mediators of taurine release from skeletal muscle cells. Am J Physiol Cell Physiol 284, C1362-1373

Ostrander EA (2012) Both ends of the leash - the human links to good dogs with bad genes. N. Engl. J. Med. 367, 636-646

Otto G, Roehe R, Looft H, Thoelking L, Kalm E (2004) Comparison of different methods for determination of drip loss and their relationships to meat quality and carcass characteristics in pigs. Meat Sci. 68, 401-409

Patterson DF, Haskins ME, Jezyk PF, Giger U, Meyerswallen VN, Aguirre G, Fyfe JC, Wolfe JH (1988) Research on genetic-diseases - reciprocal benefits to animals and man. AVMA 193, 1131-1144

Pavelic K, Bukovic D, Pavelic J (2002) The role of insulin-like growth factor 2 and its receptors in human tumors. Mol. Med. 8, 771-780

Pavlath GK (2010) Spatial and functional restriction of regulatory molecules during mammalian myoblast fusion. Exp Cell Res 316, 3067-3072

Pearson TA, Manolio TA (2008) How to Interpret a Genome-wide Association Study. JAMA 299

Peter JB, Gillespi.Ca, Barnard RJ, Stempel KE, Edgerton VR (1972) Metabolic profiles of 3 fiber types of skeletal-muscle in Guinea-pigs and rabbits. Biochemistry 11, 2627-2633

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768-772

Ponsuksili S, Chomdej S, Murani E, Blaser U, Schreinemachers HJ, Schellander K, Wimmers K (2005) SNP detection and genetic mapping of porcine genes encoding enzymes in hepatic metabolic pathways and evaluation of linkage with carcass traits. Anim. Genet. 36, 477-483

Ponsuksili S, Du Y, Murani E, Schwerin M, Wimmers K (2012) Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. Genetics 192, 1109-1122

Ponsuksili S, Jonas E, Murani E, Phatsara C, Srikanchai T, Walz C, Schwerin M, Schellander K, Wimmers K (2008a) Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. BMC Genomics 9, 367

Ponsuksili S, Murani E, Brand B, Schwerin M, Wimmers K (2011) Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. J. Lipid Res. 52, 668-678

Ponsuksili S, Murani E, Phatsara C, Jonas E, Walz C, Schwerin M, Schellander K, Wimmers K (2008b) Expression profiling of muscle reveals transcripts differentially expressed in muscle that affect water-holding capacity of pork. J. Agric. Food Chem. 56, 10311-10317

Ponsuksili S, Murani E, Schwerin M, Schellander K, Wimmers K (2010) Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. BMC Genomics 11:572

Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, Whitfield JB, Dermitzakis ET, Martin NG, Visscher PM, Montgomery GW (2012) The Brisbane systems genetics study: genetical genomics meets complex trait genetics. PLoS One 7, e35430

Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K (2011) Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. PLoS Genet. 7, e1001317

Pueyo N, Ortega FJ, Mercader JM, Moreno-Navarrete JM, Sabater M, Bonàs S, Botas P, Delgado E, Ricart W, Martinez-Larrad MT, Serrano-Rios M, Torrents D, Fernández-Real JM (2013) Common genetic variants of surfactant protein-D (SP-D) are associated with type 2 diabetes. PLoS One 8, e60468

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559-575

Quackenbush J (2001) Computational analysis of microarray data. Nat. Rev. Genet. 2, 418-427

Quiat D, Voelker KA, Pei JM, Grishin NV, Grange RW, Bassel-Duby R, Olson EN (2011) Concerted regulation of myofiber-specific gene expression and muscle performance by the transcriptional repressor Sox6. Proc. Natl. Acad. Sci. U. S. A. 108, 10196-10201

Ramos AM, Duijvesteijn N, Knol EF, Merks JWM, Bovenhuis H, Crooijmans RPMA, Groenen MAM, Harlizius B (2011) The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. BMC Genet. 12, 35

Ren J, Duan Y, Qiao R, Yao F, Zhang Z, Yang B, Guo Y, Xiao S, Wei R, Ouyang Z, Ding N, Ai H, Huang L (2011) A missense mutation in PPARD causes a major QTL effect on ear size in pigs. PLoS Genet. 7, e1002043

Rosenvold K, Andersen HJ (2003) Factors of significance, for pork quality - a review. Meat Sci. 64, 219-237

Rothschild MF (2004) Porcine genomics delivers new tools and results: this little piggy did more than just go to market. Genet. Res. 83, 1 - 6

Rothschild MF, Hu ZL, Jiang Z (2007) Advances in QTL mapping in pigs. Int J Biol Sci 3, 192-197

Sahana G, Kadlecová V, Hornshø H, Nielsen B, Christensen OF (2013) A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. J. Anim. Sci. 91, 1041-1050

Salmon Hillbertz NHC, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, Wade CM, Von Euler H, Gustafson U, Hedhammar K, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G (2007) Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nature Genet. 39, 1318-1320

Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6, e107

Schroder A, Klein K, Winter S, Schwab M, Bonin M, Zell A, Zanger UM (2011) Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. Pharmacogenomics J 13, 12-20

Solem C, Koebmann B, Jensen PR (2008) Control analysis of the role of triosephosphate isomerase in glucose metabolism in Lactococcus lactis. IET Syst. Biol. 2, 64-72

Steibel JP, Bates RO, Rosa GJ, Tempelman RJ, Rilington VD, Ragavendran A, Raney NE, Ramos AM, Cardoso FF, Edwards DB, Ernst CW (2011) Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. PLoS One 6, e16766

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100, 9440-9445

Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET (2012) Patterns of *cis* regulatory variation in diverse human populations. PLoS Genet. 8, e1002639

Tholen E, Jüngst H, Schultze-Langenhorst C, Schellander K (2005) Genetic foundation of meat quality traits of station tested slaughter pigs in North Rhine-Westphalia (Germany). Arch. Tierz.-Arch. Anim. Breed. 48, 123-130

Thomsen H, Lee HK, Rothschild MF, Malek M, Dekkers JCM (2004) Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. J. Anim. Sci. 82, 2213-2228

Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I, Jonsdottir T, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Jonsson F, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Lauritzen T, Aben KK, Verbeek ALM, Roeleveld N, Kampman E, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Becker DM, Gulcher J, Kiemeney LA, Pedersen O, Kong A, Thorsteinsdottir U, Stefansson K (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nature Genet. 41, 18-24

Tyagi S, Salier JP, Lal SK (2002) The liver-specific human alpha(1)-microglobulin/bikunin precursor (AMBP) is capable of self-association. Arch. Biochem. Biophys. 399, 66-72

Uimari P, Sironen A, Sevon-Aimonen ML (2011) Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed. Genet. Sel. Evol. 43, 42

Vaidya TB, Weyman CM, Teegarden D, Ashendel CL, Taparowsky EJ (1991) Inhibition of myogenesis by the H-Ras oncogene - implication of a role for protein-kinase-C. J. Cell Biol. 114, 809-820

van Laack RL, Kauffman RG (1999) Glycolytic potential of red, soft, exudative pork longissimus muscle. J. Anim. Sci. 77, 2971-2973

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425, 832-836

van Wijk HJ, Dibbits B, Baron EE, Brings AD, Harlizius B, Groenen MAM, Knol EF, Bovenhuis H (2006) Identification of quantitative trait loci for carcass composition and pork quality traits in a commercial finishing cross. J. Anim. Sci. 84, 789-799

van Wyk JJ, Smith EP (1999) Commentary - Insulin-like growth factors and skeletal growth: Possibilities for therapeutic interventions. J. Clin. Endocrinol. Metab. 84, 4349-4354

Velleman SG (2000) The role of the extracellular matrix in skeletal development. Poult. Sci. 79, 985-989

Velleman SG (2002) Role of Extracellular Matrix (ECM) in Growth and Development: Role of the extracellular matrix in muscle growth and development. J. Anim. Sci. 80, E8-E13

Verdugo RA, Farber CR, Warden CH, Medrano JF (2010) Serious limitations of the QTL/Microarray approach for QTL gene discovery. BMC Biol. 8, 96

Verma AS, FitzPatrick DR (2007) Anophthalmia and microphthalmia. Orphanet J. Rare Dis. 2, 47

Vidal O, Varona L, Oliver MA, Noguera JL, Sanchez A, Amills M (2006) Malic enzyme 1 genotype is associated with backfat thickness and meat quality traits in pigs. Anim. Genet. 37, 28-32

Visscher PM (2008) Sizing up human height variation. Nature Genet. 40, 489-490

Wang D, Nettleton D (2006) Identifying genes associated with a quantitative trait or quantitative trait locus via selective transcriptional profiling. Biometrics 62, 504-514

Wang JY, Luo YR, Fu WX, Lu X, Zhou JP, Ding XD, Liu JF, Zhang Q (2012) Genome-wide association studies for hematological traits in swine. Anim. Genet. 44, 34-43

Wang K, Zhang HT, Ma DQ, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PMA, Kim CE, Hou CP, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Retuerto AIA, Herman EI, Dong HM, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SFA, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459, 528-533

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: An approach to candidate gene identification. Proc. Natl. Acad. Sci. U. S. A. 99, 14903-14906

Westfall PH, Young SS (1993) Resampling-based multiple testing. (New York: Wiley)

Wimmers K, Fiedler I, Hardge T, Murani E, Schellander K, Ponsuksili S (2006) QTL for microstructural and biophysical muscle properties and body composition in pigs. BMC Genet. 7, 15

Wimmers K, Murani E, Ponsuksili S (2010) Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat performance in pigs. Brief. Funct. Genomics 9, 251-258

Wimmers K, Murani E, tePas MFW, Chang KC, Davoli R, Merks JWM, Henne H, Muraniova M, da Costa N, Harlizius B, Schellander K, Boll I, Braglia S, de Wit AAC, Cagnazzo M, Fontanesi L, Prins D, Ponsuksili S (2007) 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

Wolf U (1997) Identical mutations and phenotypic variation. Hum. Genet. 100, 305-321

Xu S (2003) Theoretical basis of the Beavis effect. Genetics 165, 2259-2268

Yoon S, Molloy MJ, Wu MP, Cowan DB, Gussoni E (2007) C60RF32 is upregulated during muscle cell differentiation and induces the formation of cellular filopodia. Dev. Biol. 301, 70-81

Yue G, Stratil A, Kopecny M, Schroffelova D, Schroffel J, Hojny J, Cepica S, Davoli R, Zambonelli P, Brunsch C, Sternstein I, Moser G, Bartenschlager H, Reiner G, Geldermann H (2003) Linkage and QTL mapping for Sus scrofa chromosome 6. J. Anim. Breed. Genet. 120, 45-55

Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Mennerich D, Rust W, Perret C, Proust C, Nicaud V, Loscalzo J, Hubner N, Tregouet D, Munzel T, Ziegler A, Tiret L, Blankenberg S, Cambien F (2010) Genetics and beyond - the transcriptome of human monocytes and disease susceptibility. PLoS One 5, e10693

Zentralverband der Deutschen Schweineproduktion (ZDS) (2003) Richtlinie für die Stationsprüfung auf Mastleistung, Schlachtkörperwert und Fleischbeschaffenheit beim Schwein, 10.12.2003. (Bonn, Germany)

Zentralverband der Deutschen Schweineproduktion (ZDS) (2006) Stressstabilität in der Tierzucht. http://www.zds-bonn.de/stressstabilitaet_in_der_schweinezucht.html

Chapter 7 Appendix

Method supplement for Chapter 3:

A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population

Hanna Heidt, Mehmet Ulas Cinar, Jasim Uddin, Christian Looft, Heinz Jüngst, Dawit Tesfaye, Ernst Tholen, Astrid Becker, Andreas Zimmer, Karl Schellander and Christine Große-Brinkhaus

Here we present additional details on the microarray data pre-processing steps performed prior to the analysis of co-expression between transcripts and the phenotype drip loss as well as the GWAS using gene expression abundance. The pre-processing steps follow the description of Oldham et al. (2008).

In order to increase the number of microarray samples, previous obtained data were added to the 100 muscle expression profiles generated in this study. In order to combine these two datasets, first outliers of each had to be removed. In a second step the jointed data was corrected for a batch effect. Gene expression analysis, especially in the case of correlation calculations, is particular sensitive to the presence of outlier samples and systematic biases in microarray data (FULLER *et al.* 2007; OLDHAM *et al.* 2008).

It was not necessary to scale both datasets to the same average intensity, because both datasets were scaled to a target signal of 500. Expression values were generated for each data set in R using the “expresso” function of the “affy” package (GENTLEMAN *et al.* 2004, www.bioconductor.org) with “mas” settings and no normalization.

The main statistical criteria for the identification of potential outlying samples was the inter-array correlation (IAC), which was defined as the Pearson correlation coefficient of the expression levels for a given pair of microarrays. In general, samples with an average IAC < 2.0 standard deviations below the mean IAC for the dataset were removed. Samples were also hierarchically clustered using average linkage and 1-IAC as a distance metric to identify outliers. This process was repeated for each dataset until no outliers were evident. In general this approach was described as an unbiased method for the identification and removal of samples with aberrant gene expression levels (OLDHAM *et al.* 2008).

In the following parts the different steps of the outlier removal of each dataset and the correction of the batch effect are described. An overview how many samples were left after each step is given in table A1.

Table A1: Overview of the number of expression profile samples used in this study

	Arrays	no. samples before pre-processing	no. samples after pre-processing	after take out duplicate animals	SNP data available for no array
Dataset 1 (study data)	Porcine Genome Array	100	97	93	93
Dataset2 (Ponsuksili *et al.* 2008a)	Porcine Genome Array	74	72	64	39

The outlier removal was performed for dataset 1 that contained before the analysis 100 samples. In a first step the distribution of the IACs of dataset 1 was examined (Figure A1). It can be seen that the distribution is not normal distributed and has a long tail on the left. This indicates the presence of outliers.

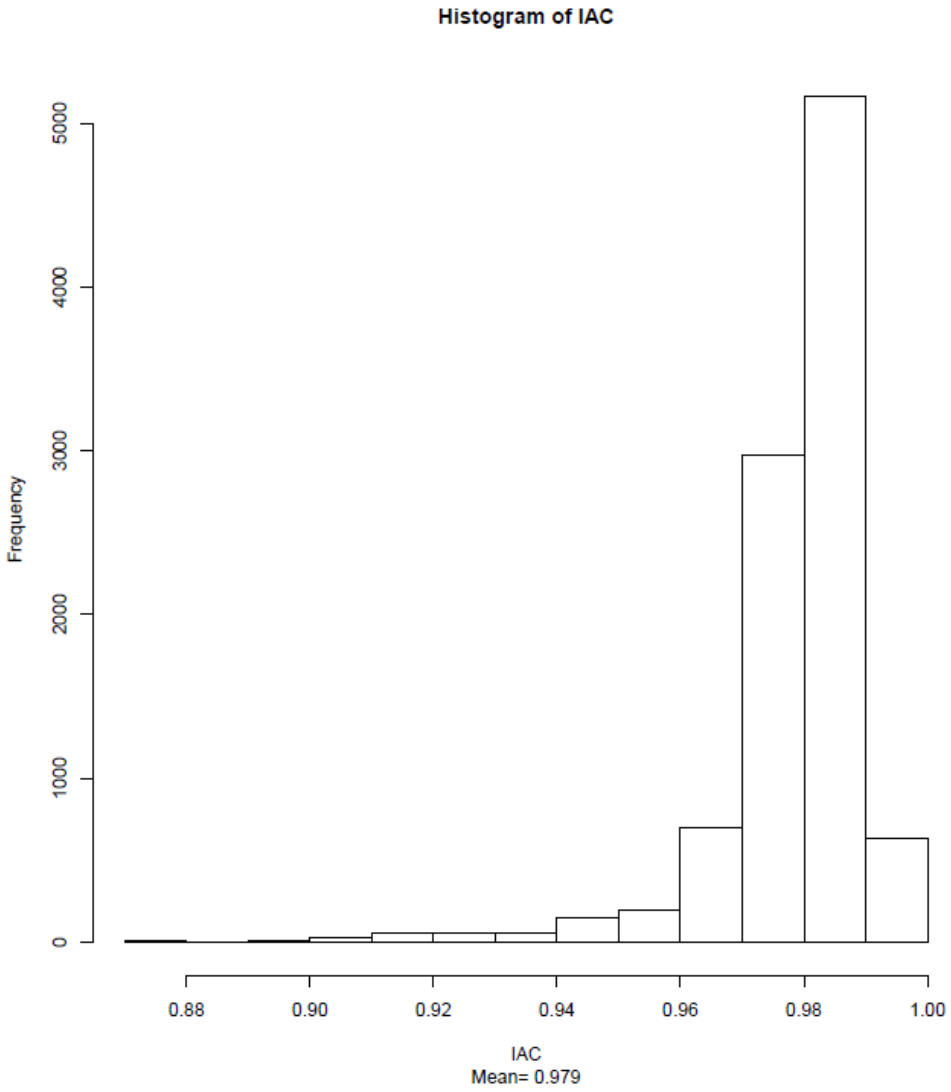

Figure A1: Histogram of IAC of dataset 1 before outlier removal

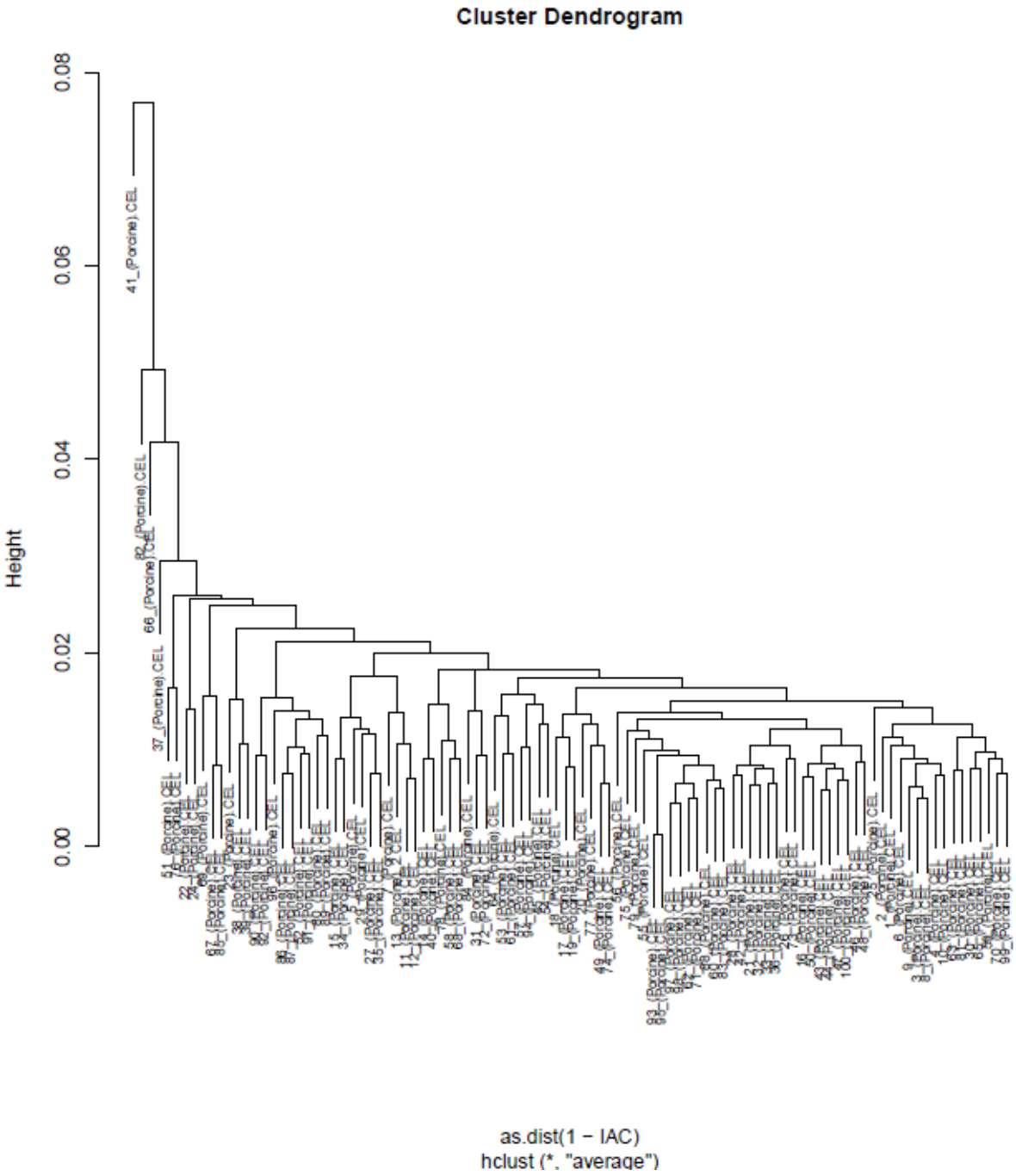

Figure A2: Cluster dendrogram of raw samples, before outlier removal (dataset 1)

The hierarchical clustering (average linkage) using the 1-IAC as a distance metric confirmed possible outliers (Figure A2). The samples are labelled from 1 to 100. There are three clear outliers at left (41, 82 and 66). An additional way to visualize outliers is to calculate the mean IAC for each array and investigate this distribution. In figure 3 one can

see that the three outliers depicted above are the same outliers identified in the initial dendrogram.

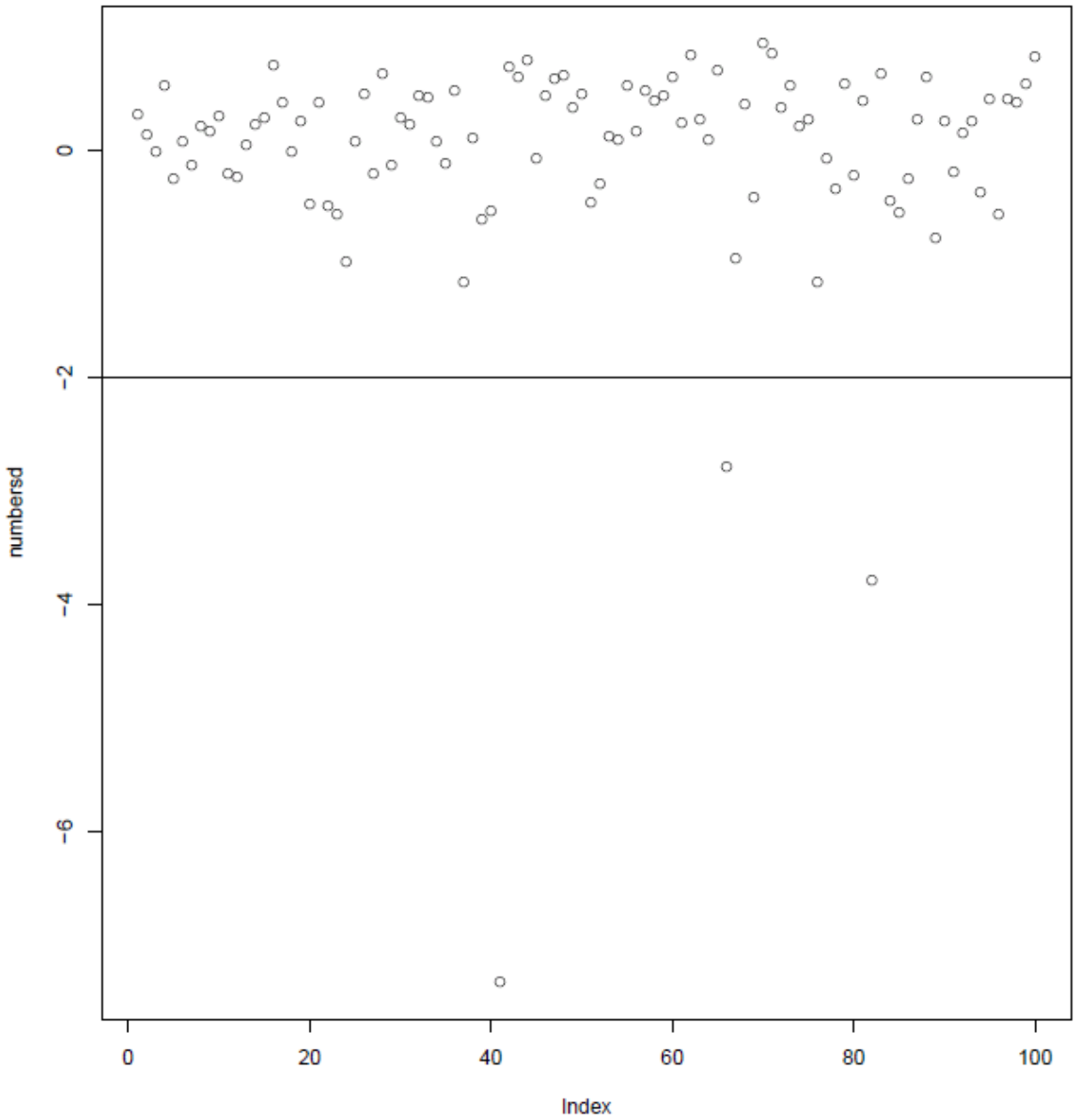

Figure A3: Plot of the number of samples deviated from the mean IAC (dataset A1)

There three outliers are removed and a new IAC matrix is calculated. The distribution of IACs is not anymore strong right shifted and the mean IACs have improved (Figure A4). In addition, the hierarchical cluster dendrogram does not reveal any outliers (Figure A5).

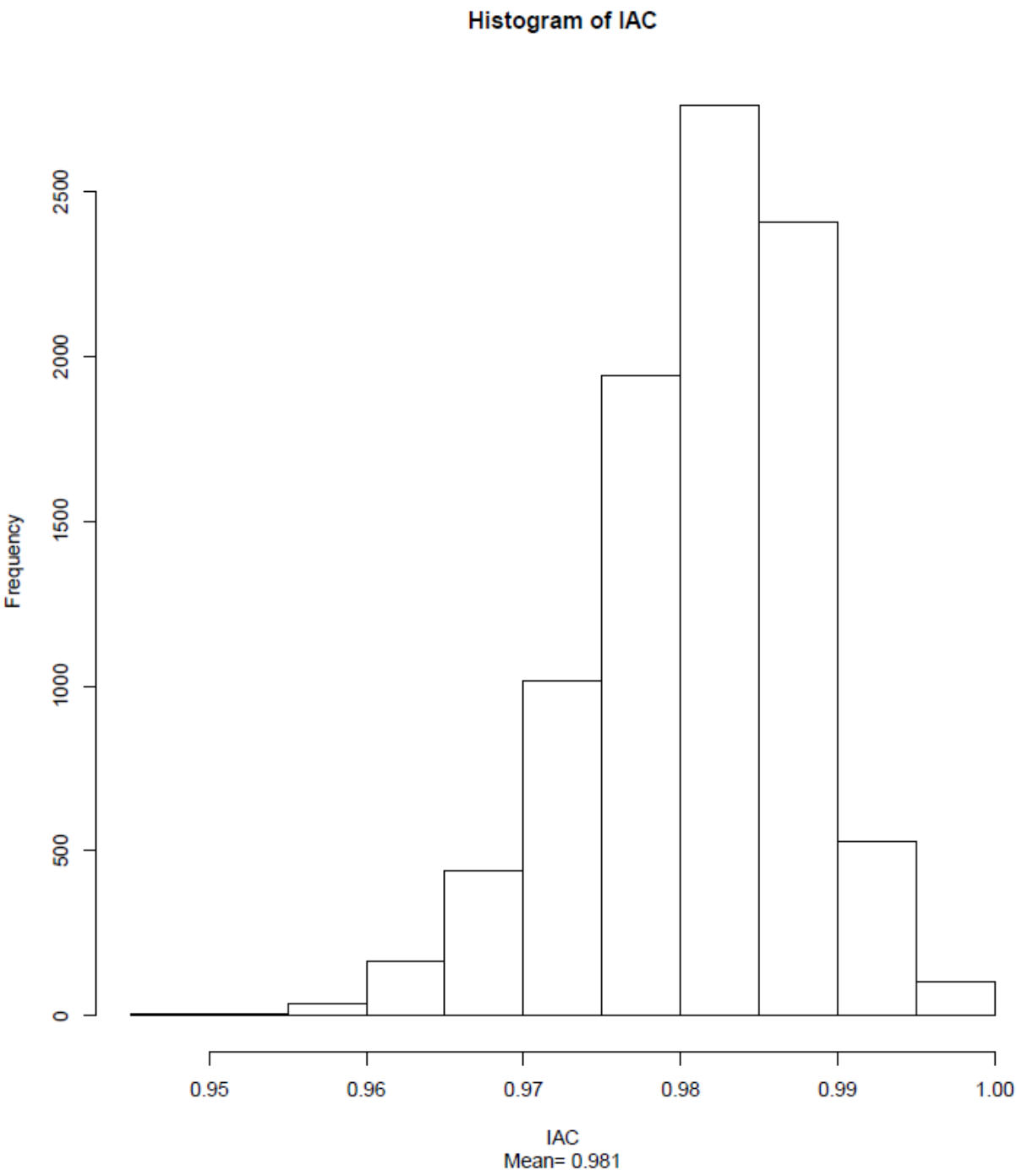

Figure A4: Histogram of AIC of dataset 1 after outlier removal

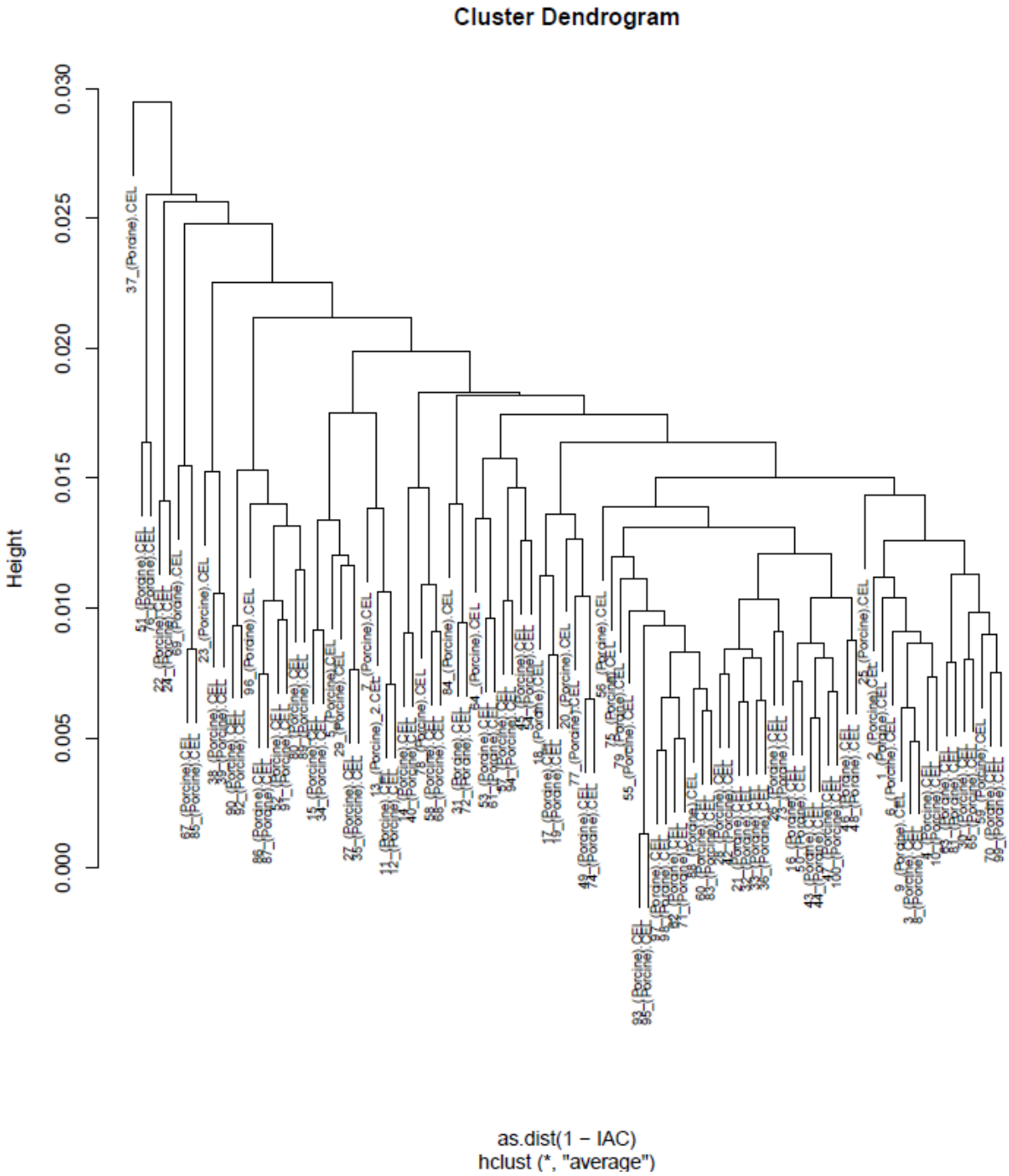

Figure A5: Cluster dendrogram after removal of the outliers (dataset 1)

The same procedures for outlier detection and removal were performed for dataset 2 comprising the expression profiles from an earlier study (Ponsuksili *et al.* 2008a). Before the analysis dataset 2 contained 74 samples.

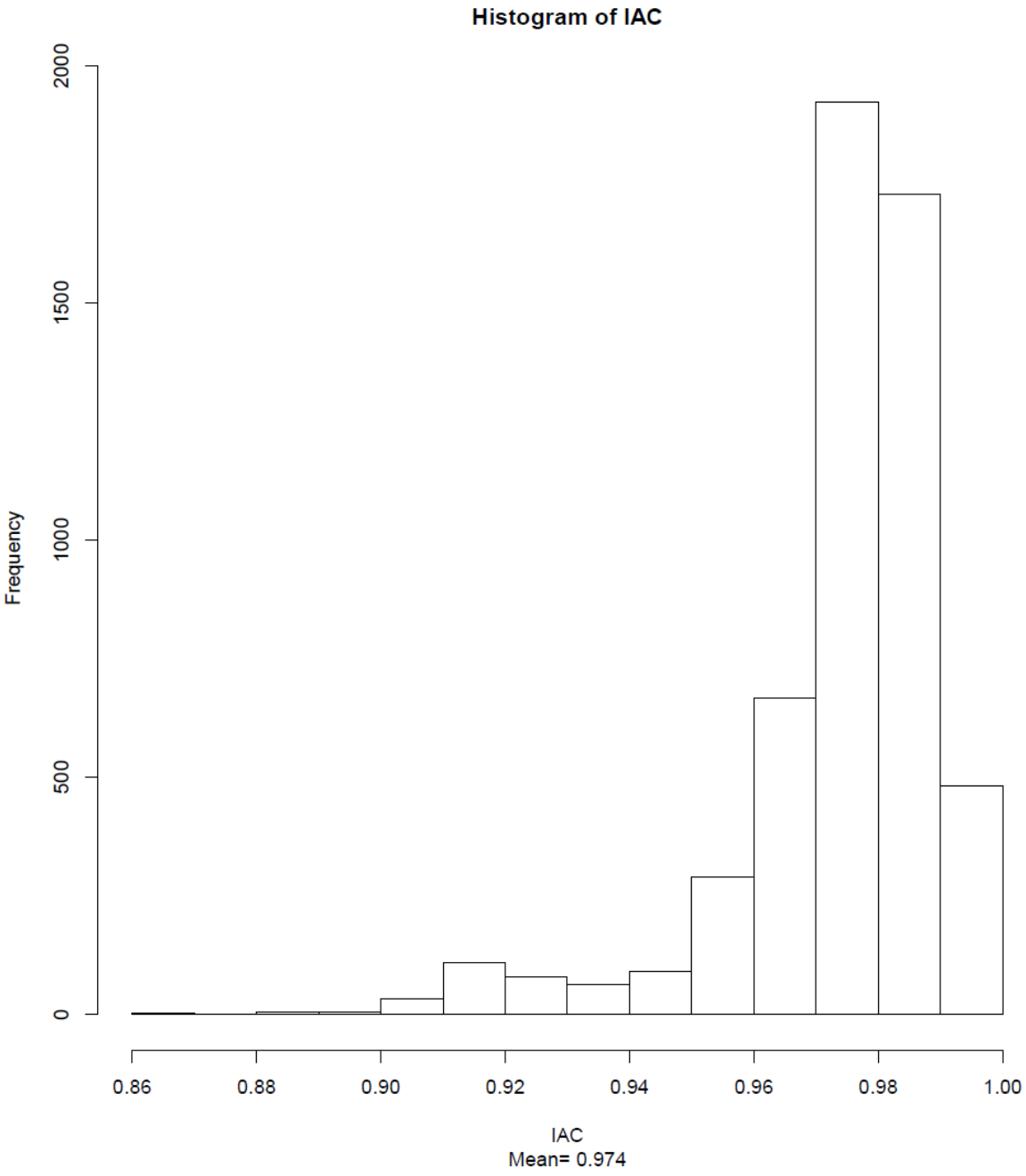

Figure A6: Histogram of AIC of dataset 2 before outlier removal

Comparable to dataset 1, dataset 2 also showed a long tail on the left of the distribution (Figure A6). In the hierarchical cluster dendrogram two possible outliers were detected (DRIP_114_G5 und DRIP26pHhigh (Figure A7).

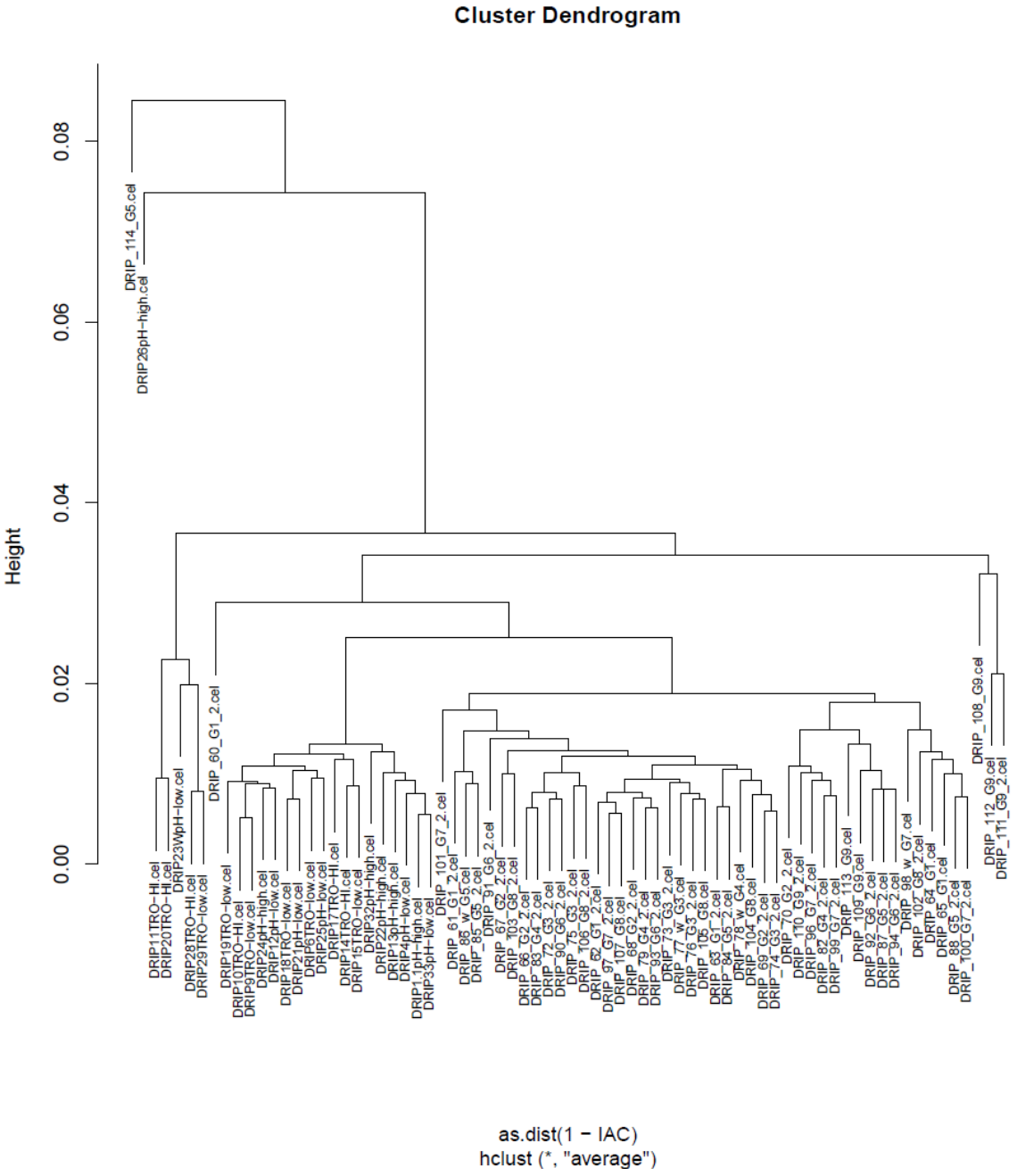

Figure A7: Cluster dendrogram of raw samples, before outlier removal (dataset 2)

The same result revealed the plot of the mean IACs (Figure A8). These two outliers were removed and the distribution and the mean IAC improved (Figure A9). At this point no clear outliers were detectable in the final hierarchical cluster dendrogram (Figure A10).

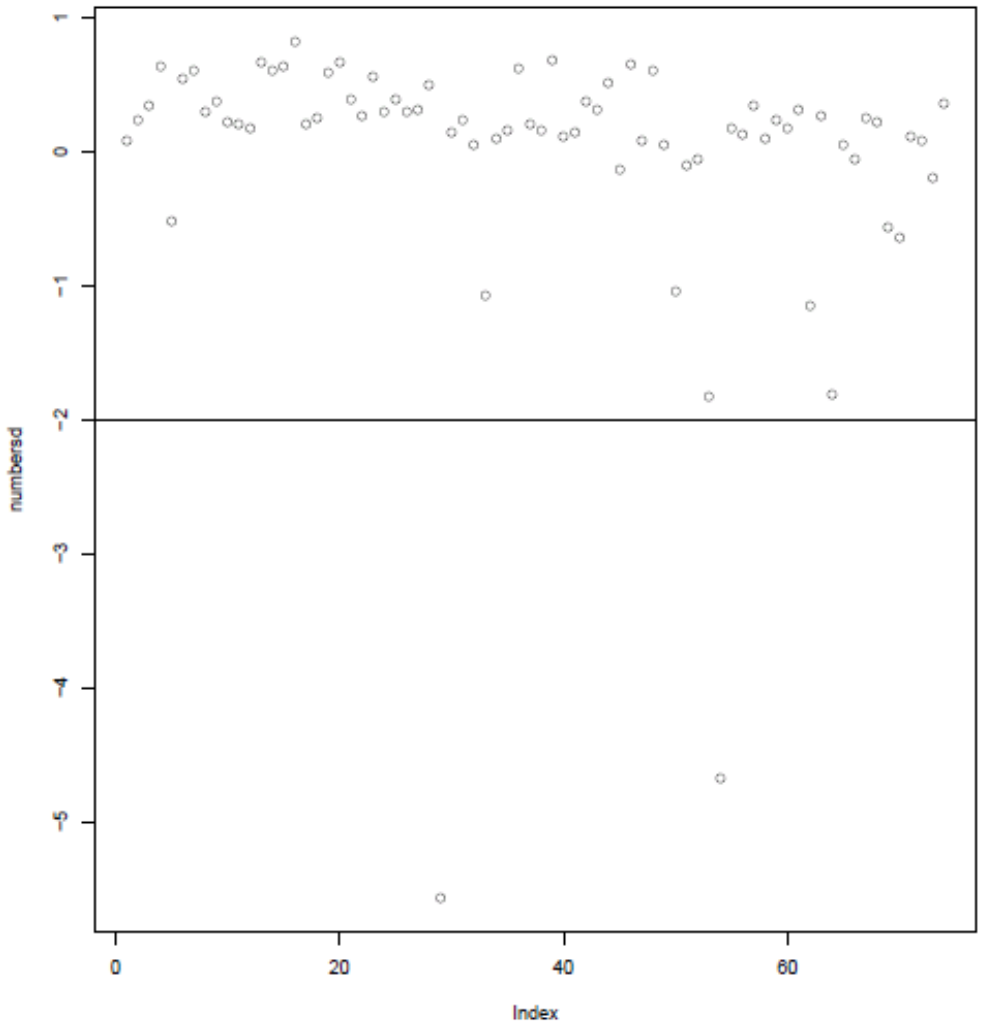

Figure A8: Plot of the number of samples deviated from the mean IAC (dataset 2)

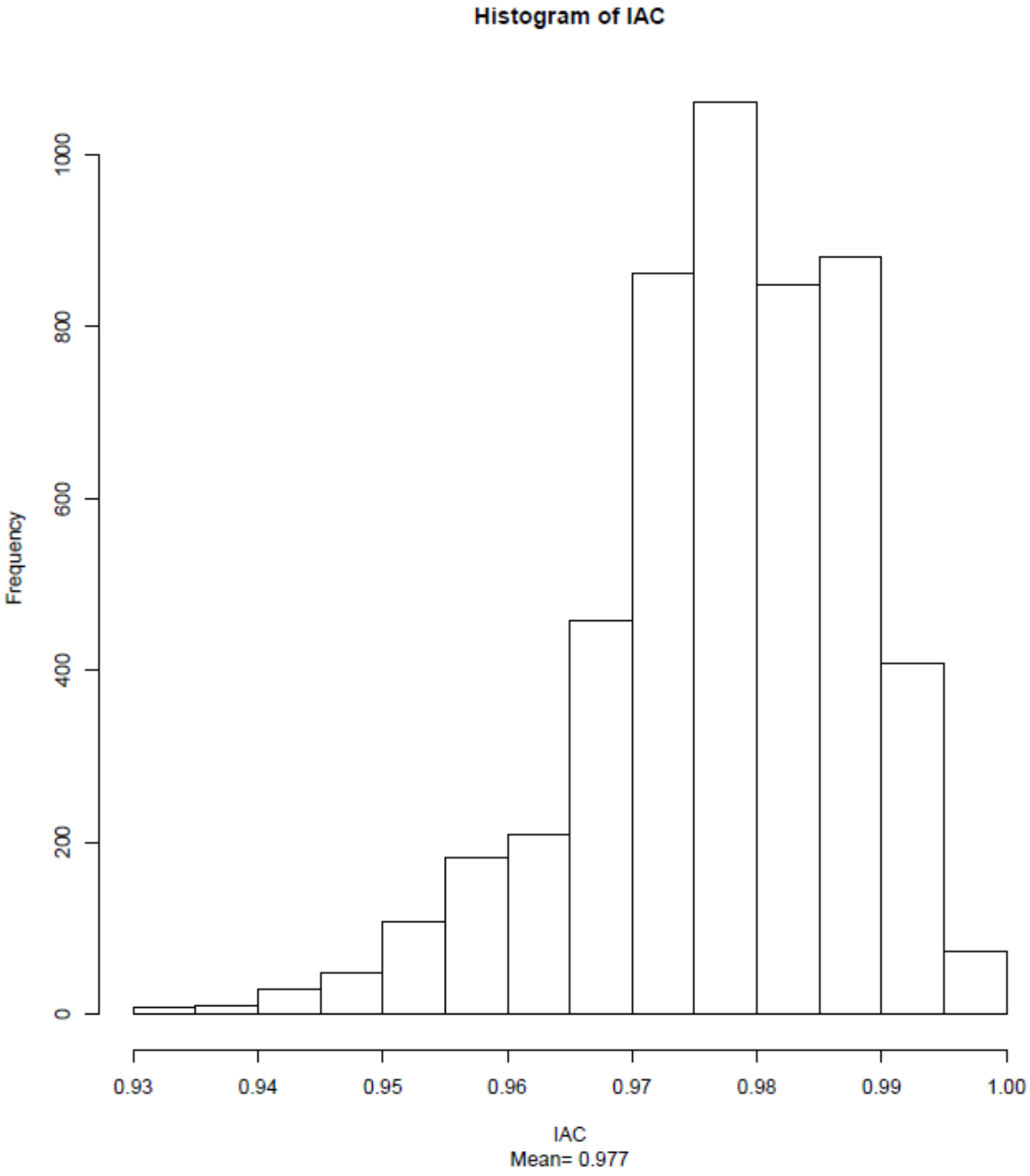

Figure A9: Histogram of AIC of dataset 1 after outlier removal

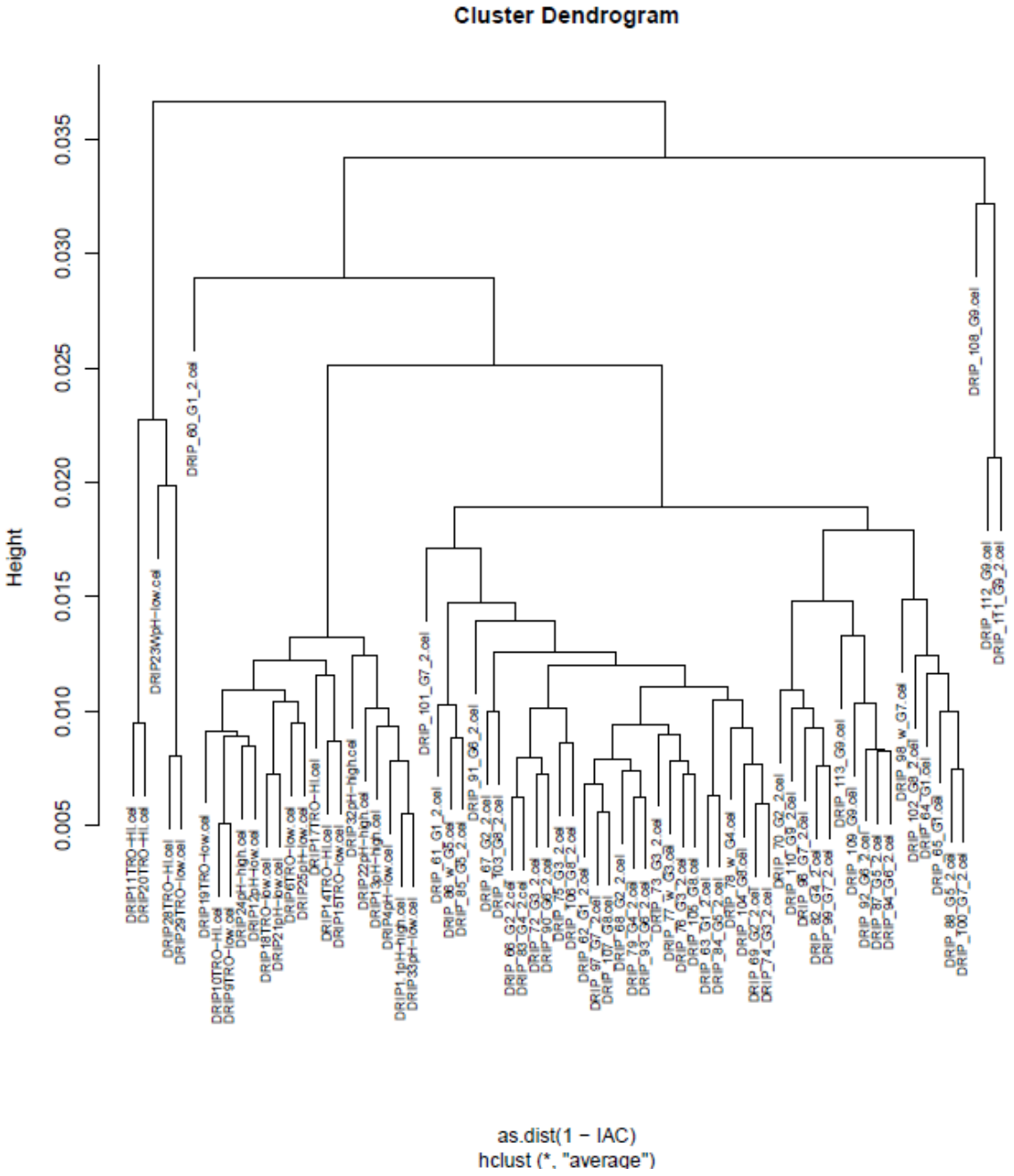

Figure A10: Cluster dendrogram after removal of the outliers (dataset 2)

Normalization and correction of the batch effect

Following outlier removal, quantile normalization was performed for each dataset using the function “plier ” in R (AFFYMETRIX INC.).

Average linkage hierarchical clustering using 1-IAC as a distance metric revealed that most samples clustered by study, indicating the presence of significant batch effects between the two datasets (Figure A11). To correct for the batch effects, additional normalization was performed using the R package “Combat” (Johnson et al. 2007, http://statistics.byu.edu/johnson/ComBat/). Normalization procedures do not adjust the data for batch effects, so when combining batches of data, particular batches that contain large batch-to-batch variation, normalization is not sufficient for adjusting for batch effects and other procedures must be applied. Each dataset was assigning a single batch number. ComBat successfully eliminated batch effects in each dataset as evidenced by hierarchical clustering and significant improvement of mean IAC.

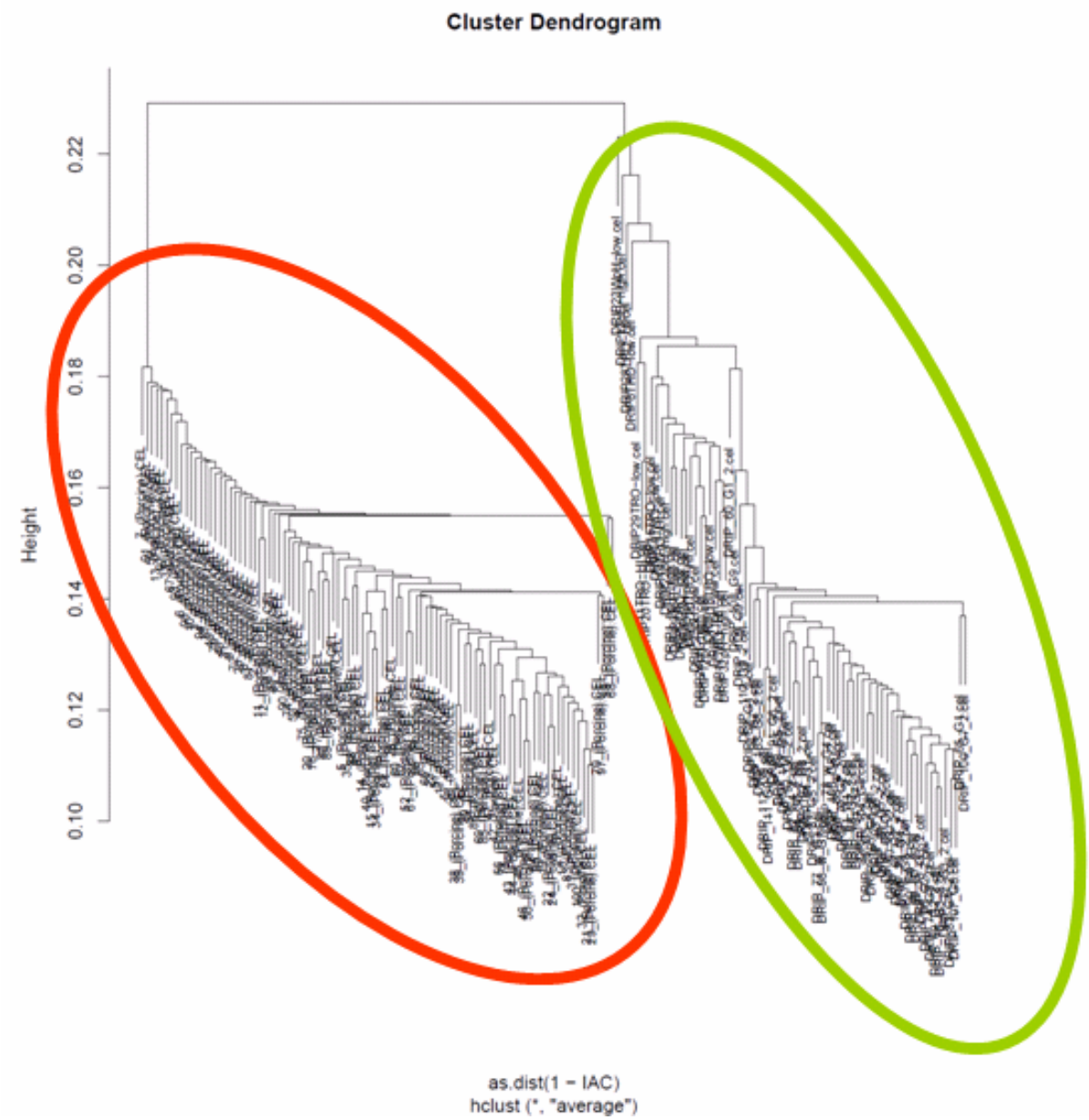

Figure A11: Cluster dendrogram after normalization, before batch correction (red circle dataset 1 and green circle dataset 2).

After correcting for the batch effect, the hierarchical cluster dendrogram revealed one homogeneous dataset (Figure A12). In a last step, 12 additional samples were removed because these belonged to the same animals.

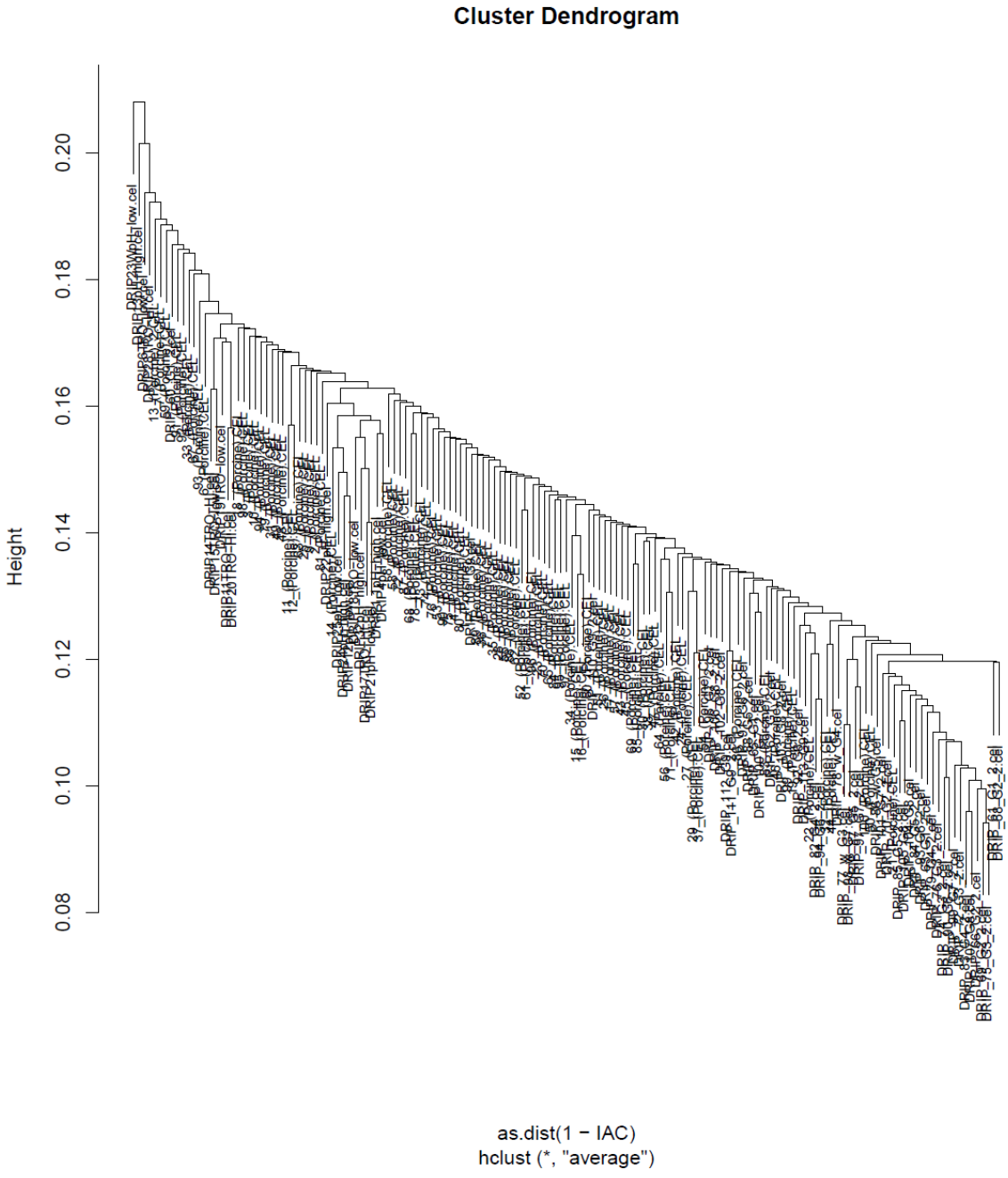

Figure A12: Cluster dendrogram after batch correction.

Table A2: Significant KEGG identifiers detected based on significantly correlated transcripts with drip loss

KEGGID	P value	Odds Ratio	Exp Count	Count (Size)	Term
10	0.0005	5.1552	1.9387	8 (37)	Glycolysis / Gluconeogenesis
4115	0.0014	3.5392	3.3010	10 (63)	p53 signalling pathway
4120	0.0023	2.7558	5.3445	13 (102)	Ubiquitin mediated proteolysis
30	0.0183	4.5931	1.0479	4 (20)	Pentose phosphate pathway
4650	0.0207	2.5568	3.4582	8 (66)	Natural killer cell mediated cytotoxicity
620	0.0269	3.2866	1.7291	5 (33)	Pyruvat metabolism
4130	0.0341	3.6704	1.2575	4 (24)	SNARE interactions in vesicular transport
5322	0.0437	2.1746	3.9822	8 (76)	Systemic lupus erythematosus

A hyperactive geometric gene set test was performed based on significantly correlated transcripts with drip loss. Over representation of gene sets defined by the KEGG database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Table A3: Significant GO Term identifiers for the biological processes (BP) detected based on significantly correlated transcripts with drip loss

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0009438	0.00001	Inf	0.22781	4 (4)	methylglyoxal metabolic process
GO:0017014	0.00004	14.35539	0.74037	6 (13)	protein nitrosylation
GO:0035606	0.00012	16.71779	0.56952	5 (10)	peptidyl-cysteine S-trans-nitrosylation
GO:0018198	0.00016	10.04508	0.91123	6 (16)	peptidyl-cysteine modification
GO:0019320	0.00019	3.47941	4.67005	14 (82)	hexose catabolic process
GO:0044238	0.00025	1.42179	313.86142	350 (5511)	primary metabolic process
GO:0006094	0.00026	4.11709	3.18930	11 (56)	gluconeogenesis
GO:0006096	0.00035	3.95634	3.29135	11 (58)	glycolysis
GO:0046364	0.00036	3.94093	3.30320	11 (58)	monosaccharide biosynthetic process
GO:0070367	0.00070	49.97352	0.22781	3 (4)	negative regulation of hepatocyte differentiation
GO:0060999	0.00070	49.97352	0.22781	3 (4)	positive regulation of dendritic spine development
GO:0060359	0.00109	8.35378	0.85428	5 (15)	response to ammonium ion
GO:0016458	0.00115	3.97602	2.67674	9 (47)	gene silencing
GO:0030422	0.00168	24.98371	0.28476	3 (5)	production of siRNA involved in RNA interference
GO:0036211	0.00312	1.36388	94.71086	119 (1663)	protein modification process
GO:0001833	0.00323	16.65377	0.34171	3 (6)	inner cell mass cell proliferation
GO:0021957	0.00324	Inf	0.11390	2 (2)	corticospinal tract morphogenesis

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0030091	0.00324	Inf	0.11390	2 (2)	protein repair
GO:0006513	0.00351	4.77693	1.53770	6 (27)	protein monoubiquitination
GO:0021955	0.00356	8.33878	0.68342	4 (12)	central nervous system neuron axonogenesis
GO:0050773	0.00512	4.36048	1.65160	6 (29)	regulation of dendrite development
GO:0034770	0.00541	12.48880	0.39866	3 (7)	histone H4-K20 methylation
GO:0044237	0.00543	1.29025	310.15956	337 (5446)	cellular metabolic process
GO:0044275	0.00576	3.01732	3.36016	9 (59)	cellular carbohydrate catabolic process
GO:0007190	0.00658	6.66939	0.79733	4 (14)	activation of adenylate cyclase activity
GO:0051642	0.00829	9.98982	0.45561	3 (8)	centrosome localization
GO:0060056	0.00829	9.98982	0.45561	3 (8)	mammary gland involution
GO:0070570	0.00829	9.98982	0.45561	3 (8)	regulation of neuron projection regeneration
GO:0097061	0.00829	9.98982	0.45561	3 (8)	dendritic spine organization
GO:0019852	0.00858	6.06234	0.85428	4 (15)	L-ascorbic acid metabolic process
GO:0045862	0.00859	3.04377	2.96149	8 (52)	positive regulation of proteolysis
GO:0016052	0.00915	2.77597	3.60302	9 (64)	carbohydrate catabolic process
GO:0035988	0.00934	33.24797	0.17086	2 (3)	chondrocyte proliferation
GO:0032864	0.00934	33.24797	0.17086	2 (3)	activation of Cdc42 GTPase activity
GO:0051081	0.00934	33.24797	0.17086	2 (3)	nuclear envelope disassembly

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0043170	0.01017	1.26110	208.81628	233 (3762)	macromolecule metabolic process
GO:0006333	0.01059	2.54007	4.32834	10 (76)	chromatin assembly or disassembly
GO:0007029	0.01093	5.55646	0.91123	4 (16)	endoplasmic reticulum organization
GO:0031098	0.01122	2.00987	8.54277	16 (150)	stress-activated protein kinase signalling cascade
GO:0010388	0.01631	7.13384	0.56952	3 (10)	cullin deneddylation
GO:0008038	0.01631	7.13384	0.56952	3 (10)	neuron recognition
GO:0018022	0.01631	7.13384	0.56952	3 (10)	peptidyl-lysine methylation
GO:0061037	0.01631	7.13384	0.56952	3 (10)	negative regulation of cartilage development
GO:0005980	0.01672	3.79160	1.53770	5 (27)	glycogen catabolic process
GO:0031281	0.01679	4.76152	1.02513	4 (18)	positive regulation of cyclase activity
GO:0051349	0.01679	4.76152	1.02513	4 (18)	positive regulation of lyase activity
GO:0031344	0.01788	1.93914	8.25801	15 (145)	regulation of cell projection organization
GO:0048619	0.01799	16.62195	0.22781	2 (4)	embryonic hindgut morphogenesis
GO:0015760	0.01799	16.62195	0.22781	2 (4)	glucose-6-phosphate transport
GO:0015748	0.01799	16.62195	0.22781	2 (4)	organophosphate ester transport
GO:0045820	0.01799	16.62195	0.22781	2 (4)	negative regulation of glycolysis
GO:0030423	0.01799	16.62195	0.22781	2 (4)	targeting of mRNA for destruction involved in RNA interference

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0030819	0.01943	3.62630	1.59465	5 (28)	positive regulation of cAMP biosynthetic process
GO:0018410	0.02034	4.44354	1.08208	4 (19)	C-terminal protein amino acid modification
GO:0034728	0.02105	6.29442	0.62153	3 (11)	nucleosome organization
GO:0090103	0.02149	6.24134	0.62647	3 (11)	cochlea morphogenesis
GO:0046685	0.02149	6.24134	0.62647	3 (11)	response to arsenic-containing substance
GO:0031050	0.02149	6.24134	0.62647	3 (11)	dsRNA fragmentation
GO:0060070	0.02159	2.14343	5.52433	11 (97)	canonical Wnt receptor signalling pathway
GO:0006349	0.02241	3.47478	1.65160	5 (29)	regulation of gene expression by genetic imprinting
GO:0003407	0.02431	4.16531	1.13904	4 (20)	neural retina development
GO:0006505	0.02431	4.16531	1.13904	4 (20)	GPI anchor metabolic process
GO:0044257	0.02526	1.57328	17.37030	26 (305)	cellular protein catabolic process
GO:0030810	0.02567	3.33538	1.70855	5 (30)	positive regulation of nucleotide biosynthetic process
GO:0070979	0.02567	3.33538	1.70855	5 (30)	protein K11-linked ubiquitination
GO:0021795	0.02747	5.54718	0.68342	3 (12)	cerebral cortex cell migration
GO:0048557	0.02747	5.54718	0.68342	3 (12)	embryonic digestive tract morphogenesis
GO:0051865	0.02871	3.91981	1.19599	4 (21)	protein autoubiquitination
GO:0045761	0.02871	3.91981	1.19599	4 (21)	regulation of adenylate cyclase activity
GO:0072520	0.02885	11.07995	0.28476	2 (5)	seminiferous tubule development

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0030970	0.02885	11.07995	0.28476	2 (5)	retrograde protein transport, ER to cytosol
GO:0000185	0.02885	11.07995	0.28476	2 (5)	activation of MAPKKK activity
GO:0046689	0.02885	11.07995	0.28476	2 (5)	response to mercury ion
GO:0046639	0.02885	11.07995	0.28476	2 (5)	negative regulation of alpha-beta T cell differentiation
GO:0031054	0.02885	11.07995	0.28476	2 (5)	pre-miRNA processing
GO:0031122	0.02885	11.07995	0.28476	2 (5)	cytoplasmic microtubule organization
GO:0060575	0.02885	11.07995	0.28476	2 (5)	intestinal epithelial cell differentiation
GO:0006344	0.02885	11.07995	0.28476	2 (5)	maintenance of chromatin silencing
GO:0032042	0.02885	11.07995	0.28476	2 (5)	mitochondrial DNA metabolic process
GO:0030801	0.02922	3.20670	1.76551	5 (31)	positive regulation of cyclic nucleotide metabolic process
GO:0001824	0.02922	3.20670	1.76551	5 (31)	blastocyst development
GO:1900544	0.02922	3.20670	1.76551	5 (31)	positive regulation of purine nucleotide metabolic process
GO:0000096	0.02961	2.54165	3.01845	7 (53)	sulfur amino acid metabolic process
GO:0050821	0.03241	2.48726	3.07540	7 (54)	protein stabilization
GO:0070647	0.03242	1.47583	21.98340	31 (386)	protein modification by small protein conjugation or removal
GO:0007528	0.03424	4.99185	0.74037	3 (13)	neuromuscular junction development

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0009057	0.03482	1.34268	40.49274	52 (711)	macromolecule catabolic process
GO:0006334	0.03539	2.43515	3.13235	7 (55)	nucleosome assembly
GO:0042493	0.03573	1.51692	17.93982	26 (315)	response to drug
GO:0010951	0.03584	2.04042	5.23957	10 (92)	negative regulation of endopeptidase activity
GO:0019941	0.03607	1.54322	16.28822	24 (286)	modification-dependent protein catabolic process
GO:0070936	0.03722	2.97692	1.87941	5 (33)	protein K48-linked ubiquitination
GO:0007254	0.04066	1.91779	6.09384	11 (107)	JNK cascade
GO:0044267	0.04123	1.21469	109.88622	126 (1973)	cellular protein metabolic process
GO:0072498	0.04167	8.30894	0.34171	2 (6)	embryonic skeletal joint development
GO:0051597	0.04167	8.30894	0.34171	2 (6)	response to methylmercury
GO:0043501	0.04167	8.30894	0.34171	2 (6)	skeletal muscle adaptation
GO:0070389	0.04167	8.30894	0.34171	2 (6)	chaperone cofactor-dependent protein refolding
GO:0000389	0.04167	8.30894	0.34171	2 (6)	nuclear mRNA 3'-splice site recognition
GO:0060306	0.04167	8.30894	0.34171	2 (6)	regulation of membrane repolarization
GO:0060236	0.04167	8.30894	0.34171	2 (6)	regulation of mitotic spindle organization
GO:0046633	0.04167	8.30894	0.34171	2 (6)	alpha-beta T cell proliferation

Table A3 continued:

GOBPID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0006734	0.04167	8.30894	0.34171	2 (6)	NADH metabolic process
GO:0061001	0.04167	8.30894	0.34171	2 (6)	regulation of dendritic spine morphogenesis
GO:0042518	0.04167	8.30894	0.34171	2 (6)	negative regulation of tyrosine phosphorylation of Stat3 protein
GO:0007216	0.04167	8.30894	0.34171	2 (6)	G-protein coupled glutamate receptor signalling pathway
GO:0048680	0.04167	8.30894	0.34171	2 (6)	positive regulation of axon regeneration
GO:0050854	0.04178	4.53749	0.79733	3 (14)	regulation of antigen receptor-mediated signalling pathway
GO:0016311	0.04276	1.64951	10.82084	17 (190)	dephosphorylation
GO:0021537	0.04450	2.15532	3.98663	8 (70)	telencephalon development
GO:0048522	0.04720	1.19766	120.85174	137 (2122)	positive regulation of cellular process
GO:0051402	0.04814	1.85898	6.26470	11 (110)	neuron apoptotic process
GO:0031329	0.04918	1.46433	18.50934	26 (325)	regulation of cellular catabolic process

A hyper geometric gene set test was performed based on significantly correlated transcripts with drip loss. Over representation of gene sets defined by the GO database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Table A4: Significant GO Term identifiers for the cell components (CC) detected based on significantly correlated transcripts with drip loss

GOCCID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0030176	0.0047	2.9073	3.8608	10 (67)	integral to endoplasmic reticulum membrane
GO:0061202	0.0086	9.8656	0.4610	3 (8)	clathrin sculpted gamma-aminobutyric acid transport vesicle membrane
GO:0000780	0.0086	9.8656	0.4610	3 (8)	condensed nuclear chromosome, centromeric region
GO:0044424	0.0112	1.3465	393.7287	413 (6934)	intracellular part
GO:0031300	0.0143	2.0557	7.3182	14 (127)	intrinsic to organelle membrane
GO:0043227	0.0169	1.2330	308.2287	331 (5349)	membrane-bounded organelle
GO:0000786	0.0175	3.7443	1.5558	5 (27)	nucleosome
GO:0001741	0.0184	16.4157	0.2305	2 (4)	XY body
GO:0005736	0.0184	16.4157	0.2305	2 (4)	DNA-directed RNA polymerase I complex
GO:0005634	0.0189	1.2209	185.6056	208 (3221)	nucleus
GO:0000151	0.0290	1.9662	6.5115	12 (113)	ubiquitin ligase complex
GO:0030062	0.0295	10.9424	0.2881	2 (5)	mitochondrial tricarboxylic acid cycle enzyme complex
GO:0005954	0.0295	10.9424	0.2881	2 (5)	calcium- and calmodulin-dependent protein kinase complex
GO:0016272	0.0295	10.9424	0.2881	2 (5)	prefoldin complex
GO:0031527	0.0295	10.9424	0.2881	2 (5)	filopodium membrane
GO:0032589	0.0298	3.8710	1.2101	4 (21)	neuron projection membrane
GO:0016581	0.0298	3.8710	1.2101	4 (21)	NuRD complex
GO:0005789	0.0340	1.5004	19.5344	28 (339)	endoplasmic reticulum membrane

Table A4 continued:

GOCCID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0005681	0.0345	1.8550	7.4334	13 (129)	spliceosomal complex
GO:0017109	0.0426	8.2058	0.3457	2 (6)	glutamate-cysteine ligase complex
GO:0005688	0.0426	8.2058	0.3457	2 (6)	U6 snRNP
GO:0000506	0.0426	8.2058	0.3457	2 (6)	glycosylphosphatidylinositol-N-acetylglucosaminyltransferase (GPI-GnT) complex
GO:0016442	0.0426	8.2058	0.3457	2 (6)	RNA-induced silencing complex
GO:0071565	0.0426	8.2058	0.3457	2 (6)	nBAF complex
GO:0005669	0.0430	4.4811	0.8067	3 (14)	transcription factor TFIID complex
GO:0060198	0.0430	4.4811	0.8067	3 (14)	clathrin sculpted vesicle
GO:0010008	0.0481	1.8034	7.0301	12 (122)	endosome membrane

A hyper geometric gene set test was performed based on significantly correlated transcripts with drip loss. Over representation of gene sets defined by the GO database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Table A5: Significant GO Term identifiers for the molecular function (MF) detected based on significantly correlated transcripts with drip loss

GOMFID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0004365	0.0001	16.6473	0.5715	5 (10)	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity
GO:0035605	0.0001	16.6473	0.5715	5 (10)	peptidyl-cysteine S-nitrosylase activity
GO:0004462	0.0002	Inf	0.1715	3 (3)	lactoylglutathione lyase activity
GO:0004645	0.0003	22.1586	0.4001	4 (7)	phosphorylase activity
GO:0016903	0.0018	4.6685	1.8289	7 (32)	oxidoreductase activity, acting on the aldehyde or oxo group of donors
GO:0042132	0.0033	Inf	0.1143	2 (2)	fructose 1,6-bisphosphate 1-phosphatase activity
GO:0004040	0.0033	Inf	0.1143	2 (2)	amidase activity
GO:0015152	0.0033	Inf	0.1143	2 (2)	glucose-6-phosphate transmembrane transporter activity
GO:0070012	0.0033	Inf	0.1143	2 (2)	oligopeptidase activity
GO:0004719	0.0033	Inf	0.1143	2 (2)	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity
GO:0048273	0.0033	Inf	0.1143	2 (2)	mitogen-activated protein kinase p38 binding
GO:0042799	0.0055	12.4387	0.4001	3 (7)	histone methyltransferase activity (H4-K20 specific)
GO:0016893	0.0067	6.6422	0.8001	4 (14)	endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters
GO:0003824	0.0083	1.2607	177.4927	203 (3169)	catalytic activity
GO:0016595	0.0084	9.9498	0.4572	3 (8)	glutamate binding

Table A5 continued:

GOMFID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0035256	0.0094	33.1175	0.1715	2 (3)	G-protein coupled glutamate receptor binding
GO:0048408	0.0094	33.1175	0.1715	2 (3)	epidermal growth factor binding
GO:0004525	0.0094	33.1175	0.1715	2 (3)	ribonuclease III activity
GO:0030160	0.0094	33.1175	0.1715	2 (3)	GKAP/Homer scaffold activity
GO:0004842	0.0098	1.9936	9.1444	17 (160)	ubiquitin-protein ligase activity
GO:0016874	0.0107	1.6917	16.9171	27 (296)	ligase activity
GO:0090079	0.0120	8.2905	0.5144	3 (9)	translation regulator activity, nucleic acid binding
GO:0016881	0.0169	1.7960	11.2232	19 (197)	acid-amino acid ligase activity
GO:0004521	0.0170	4.7422	1.0287	4 (18)	endoribonuclease activity
GO:0051287	0.0226	2.7085	2.8576	7 (50)	NAD binding
GO:0032947	0.0227	3.4605	1.6574	5 (29)	protein complex scaffold
GO:0048306	0.0227	3.4605	1.6574	5 (29)	calcium-dependent protein binding
GO:0030170	0.0248	2.9335	2.2861	6 (40)	pyridoxal phosphate binding
GO:0003725	0.0260	3.3217	1.7146	5 (30)	double-stranded RNA binding
GO:0004722	0.0260	3.3217	1.7146	5 (30)	protein serine/threonine phosphatase activity
GO:0005165	0.0277	5.5251	0.6858	3 (12)	neurotrophin receptor binding
GO:0019203	0.0290	11.0366	0.2858	2 (5)	carbohydrate phosphatase activity
GO:0010340	0.0290	11.0366	0.2858	2 (5)	carboxyl-O-methyltransferase activity

Table A5 continued:

GOMFID	P value	Odds Ratio	Exp Count	Count (Size)	Term
GO:0017176	0.0290	11.0366	0.2858	2 (5)	phosphatidylinositol N-acetylglucosaminyltransferase activity
GO:0016757	0.0305	1.8412	8.0585	14 (141)	transferase activity, transferring glycosyl groups
GO:0016866	0.0335	3.0749	1.8289	5 (32)	intramolecular transferase activity
GO:0070742	0.0346	4.9720	0.7430	3 (13)	C2H2 zinc finger domain binding
GO:0031014	0.0419	8.2765	0.3429	2 (6)	troponin T binding
GO:0004095	0.0419	8.2765	0.3429	2 (6)	carnitine O-palmitoyltransferase activity
GO:0034235	0.0419	8.2765	0.3429	2 (6)	GPI anchor binding
GO:0004357	0.0419	8.2765	0.3429	2(6)	glutamate-cysteine ligase activity
GO:0008195	0.0419	8.2765	0.3429	2 (6)	phosphatidate phosphatase activity
GO:0005047	0.0419	8.2765	0.3429	2 (6)	signal recognition particle binding
GO:0010181	0.0422	4.5195	0.8001	3 (14)	FMN binding
GO:0005246	0.0451	3.3172	1.3717	4 (24)	calcium channel regulator activity
GO:0004222	0.0462	2.2815	3.3148	7 (58)	metalloendopeptidase activity

A hyper geometric gene set test was performed based on significantly correlated transcripts with drip loss. Over representation of gene sets defined by the GO database was tested using Fisher's exact test. The gene set was considered significant if $P < 0.05$

Acknowledgement - Danksagung

Thank you – Danke schön

Mein herzlicher Dank gilt Prof. Dr. Karl Schellander für die Überlassung des Themas der vorliegenden Dissertation, sowie Prof. Dr. Brigitte Peterson für die Übernahme des zweiten Gutachters.

Des Weiteren bedanke ich mich bei …

… Prof. Dr. med. Markus M. Nöthen vom Institut für Humangenetik der Universität Bonn für das Genotypisieren der Proben mittels des Illumina 60K porcine SNP Chips.

…Prof. Dr. Andreas Zimmer und Dr. Astrid Becker vom Institut für Molekulare Psychiatrie der Universität Bonn für die Bearbeitung der Affymetrix Chips

… Dr. Ernst Tholen vom Institut für Tierwissenschaften der Universität Bonn für die Hilfe bei allen statistischen Fragen sowie bei allen großen und kleinen Problemen.

… Prof. Dr. Christian Looft vom Institut für Tierwissenschaften der Universität Bonn fürs Korrekturlesen meiner Arbeit.

…Dr. Christine Große-Brinkhaus vom Institut für Tierwissenschaften der Universität Bonn für die Unterstützung während meiner Doktorarbeit. Sie stand mir stets mit Rat und Tat zur Seite, sei es bei statistischen, technischen oder rhetorischen Fragestellungen.

…Dr. Ulas Cinar vom Institut für Tierwissenschaften der Universität Bonn für die Unterstützung während meiner Doktorarbeit. Er hat mir vor allem bei der Planung und der Durchführung des Projektes geholfen sowie beim Korrekturlesen meiner Arbeit.

… Dr. Jassim Uddin vom Institut für Tierwissenschaften der Universität Bonn fürs Korrekturlesen meiner Arbeit.

… Helga Brodeßer, Nadine Leyer, Birgit Koch-Fabritius, Jessy Gonyer und den Azubis, ohne die die Isolierung der ganzen Proben nicht machbar gewesen wäre.

… Peter Müller für die Hilfe mit allen computertechnischen Problemen.

... Bianca Peters und Ulrike Schroeter, die mir immer bei bürokratischen Angelegenheiten weitergeholfen haben.

… Stephan Knauf, für die schnelle und kompetente Bearbeitung jedes meiner technischen und handwerklichen Anliegen.

… allen Doktoranden und Post-Doktoranden aus dem Institut für Tierwissenschaften der Universität Bonn, speziell Heiko Buschbell, Anke Brings, Felix Austermann, Dr. Ahmed Gad, Christiane Neuhoff, Maren Pröll und Sarah Bergfelder für die tolle Zeit zusammen und eure Hilfe bei Fragen fachlicher, technischer oder rhetorischer Natur. Den Kollegen Ralph Schemmer, Dr. Katrin Gerlach, Frederike Hippenstiel, Christian Böttger, Dr. Eva Findeisen, Dr. Christian Cox, Dr. Patrick Steuer, Anna Rauen und Isabel Gussek danke ich für die sowohl ernsthaften als auch lustigen Gespräche und das manchmal rettende Feierabendbier.

… all meinen Freunden, die auf die eine oder andere Art zum Gelingen der Arbeit beigetragen haben.

Mein besonderer Dank gilt meiner Familie, sowohl dem brettener und lingener Teil als auch dem Teil aus Luxemburg. Bei meinen Eltern will ich mich einfach für Alles bedanken, was sie mir bis jetzt in meinem Leben ermöglicht haben. Bei Lena und Hugo, Hans und Margot, Christoph und Anke, Lena, Johanna, Marie, Hermann und Marianne, Heinz und Edeltraut und natürlich meiner Oma Anneliese möchte ich mich bedanken, weil ich bei euch meine Batterien wiederaufladen konnte und ihr mich immer wieder motiviert habt meine Arbeit zu beenden.

Auch dem luxemburgischen Teil meiner Familie danke ich sehr, speziell Dan, Mady, Lynn, Tom und Finn, die dafür gesorgt haben, dass ich den Kopf wieder frei bekommen habe.

Zuletzt möchte ich mich noch bei meinem Freund Luc bedanken, dass er mich die ganze Zeit unterstützt, meine Launen ertragen, mich mit Essen versorgt hat, wenn der Hunger mal wieder riesig war und mir immer das Gefühl gegeben hat, dass ich alles schaffen kann. Ich freu mich schon riesig auf die gemeinsame Zeit, die noch vor uns liegt!

Publications

Heidt H., Cinar M.U., Uddin M.J., Looft C., Jüngst H., Tesfaye D., Becker A., Zimmer A., Ponsuksili S., Wimmers K., Tholen E., Schellander K. and Große-Brinkhaus C.(2013) A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. Mammalian Genome (2013) **24**: 416-26.

Heidt H., Neuhoff C., Cinar M.U., Tesfaye D., Tholen E., Schulze Langenhorst C., Looft C., Schellander K. & Große-Brinkhaus C. (2012) Teilprojekt 4.3: Merkmalsassoziation der Biomarker bei verschiedenen Rinder- und Schweinerassen in unterschiedlichen Produktionsverfahren In: Zwischenevaluierung des AgroClustErs PHÄNOMICS (ed. By Rostock U). BMBF, Dummerstorf.

Heidt H., Sahadevan S., Grosse-Brinkhaus C., Cinar M.U., Looft C., Jüngst H., Tesfaye D., Tholen E., Nöthen M. & Schellander K. (2011a) A genome-wide association study for drip loss in a Duroc × Pietrain resource population. In: 4th International Symposium on Animal Functional Genomics, Dublin, Ireland.

Heidt H., Sahadevan S., Grosse-Brinkhaus C., Cinar M.U., Looft C., Jüngst H., Tesfaye D., Tholen E., Nöthen M. & Schellander K. (2011b) Genomweite Assoziationsstudie (60K-SNP Chip) für das Merkmal Tropfsaftverlust in der Bonner Duroc x Pietrain Ressourcenpopulation unter Berücksichtigung der Familienstruktur. In: Vortragstagung der Deutschen Gesellschaft für Züchtungskunde e.V. (DGfZ) und der Gesellschaft für Tierzuchtwissenschaften e.V. (GfT), p.B01, Freising-Weihenstephan

Heidt H., Salilew-Wondim D., Gad A., Große-Brinkhaus C., Tesfaye D., Hoelker M., Looft C., Tholen E. & Schellander K. (2010) Expression pattern in the endometrium of genes involved in calcium signaling pathway during bovine estrous cycle. In: 61st Annual Meeting of the European Association for Animal Production (EAAP), Heraklion, Kreta, Greece

Neuhoff C., Cinar M.U., Große-Brinkhaus C., Tholen E., Tesfaye D., Looft C., Wimmers K., Ponsuksili S., **Heidt H.,** Jüngst H., Phatsara C. & Schellander K. (2010) Erstellung von Expressionsprofilen 5 funktioneller Kandidatengene (CCL28, STAT5A, STAT5B, LIPE und EGF) des Erbfehlers Stülpzitze beim Schwein. In: Vortragstagung der Deutschen Gesellschaft für Züchtungskunde e.V. (DGfZ) und der Gesellschaft für Tierzuchtwissenschaften e.V. (GfT),Tagungsband, p. C27 (Proc), Kiel, Deutschland.

Neuhoff C., Große-Brinkhaus C., **Heidt H.,** Tesfaye D., Tholen E., Kellerman J., Looft C., Schellander K. & Cinar M.U. (2012a) Teilprojekt 3.3: Integrative ‚OMICS'-Profile ausgewählter relevanter Gewebe von Schweinen mit differenter Wachstumsleistung und Muskelfleischzusammensetzung In: Zwischenevaluierung desAgroClustErs PHÄNOMICS (ed.by Rostock U). BMBF, Dummerstorf.

Neuhoff C., Pröll M., Grosse-Brinkhaus C., **Heidt H.,** Tesfaye D., Tholen E., Looft C., Schellander K.& Cinar M.U. (2012b) Proteome analysis of skeletal muscle in high and low drip loss Duroc × Pietrain F2 pigs. In: 33rd Conference of the International Society for Animal Genetics (ISAG),Cairns, Australia.