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

Systems biology analysis of meat quality in Duroc × Pietrain pigs based on integrated omics approaches

Inaugural – Dissertation

Zur Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

von

Julia Welzenbach

aus

Bad Honnef

Referent:	Prof. Dr. Karl Schellander
Korreferent:	Dr. Ernst Tholen
	Prof. Dr. Jens Léon
	Prof. Dr. Reiner Roehe
Tag der mündlichen Prüfung:	02.06.2017
Erscheinungsjahr:	2017

Für Tobias

Abstract

Water holding capacity and drip loss are important commercially interesting meat quality parameters with low heritabilities and complex genetic and metabolic background. The aim of this thesis was the application of different statistical approaches to integrate the omics levels genome, transcriptome, proteome, metabolome and phenotype in 100 Duroc × Pietrain pigs to elucidate the genetic and metabolic background of meat quality, paying special attention to drip loss. The pigs were genotyped and performance tested. The metabolome, proteome and transcriptome were profiled in muscle samples of the animals.

In the first study, metabolite profiles were analysed based on different statistical procedures to elucidate the underlying biochemical processes and to identify potential metabolite indicators for drip loss, pH1, pH24 and meat color. In case of drip loss, the procedure 'Random forest regression' was the most suitable method to identify reliable biomarkers. Based on a systems biological approach, in a second step different omics levels were integrated to increase the information density for the analysis of meat quality and carcass composition traits. The omics analyses were applied as promising alternatives to standard genetic association studies and the metabolic phenotypes were used as more accurate traits to characterise underlying functional pathways and candidate genes. Therefore, metabolite and protein profiles were used to perform an enrichment analysis that revealed the sphingolipid metabolism with significant influences on drip loss. Based on the identified pathways, metabolites and proteins were selected as 'intermediate phenotypes' for drip loss.

In the third study, the most promising metabolic traits for drip loss and other meat quality and carcass composition traits were picked using a network analysis that integrated all quantified transcripts, metabolites and proteins. In the following, besides the conventional production traits, the selected 'intermediate phenotypes' (single metabolites/proteins or combined metabolic traits) were analysed in genome-wide association studies (GWAS). As a result, several highly interesting candidate genes for drip loss and carcass composition on *Sus scrofa* chromosomes 5 and 18 were identified.

Due to the higher information density between genotype and phenotype, we hypothesize that GWAS based on intermediate phenotypes are able to improve the statistical power in the identification of reliable candidate genes and to avoid false positive, redundant results. In conclusion, our omics approaches provide comprehensive insights in the genetic variation of genes directly involved in the metabolism of production traits.

Zusammenfassung

Das Wasserbindungsvermögen und der Tropfsaftverlust (TSV) sind ökonomisch relevante Fleischqualitätsparameter mit geringer Heritabilität und komplexer genetischer Fundierung. Es wird angenommen, dass die Analyse verschiedener Omics Ebenen bei der Untersuchung komplexer Merkmale sehr zielführend sein kann. Das Ziel dieser Arbeit war die Anwendung unterschiedlicher statistischer Verfahren, um die Omics Ebenen Genom, Transkriptom, Proteom, Metabolom und Phänotyp zu integrieren und in 100 Duroc × Pietrain Schweinen die genetische und metabolische Grundlage vom Merkmal TSV und anderen Fleischqualitätsparametern zu untersuchen.

In der ersten Studie wurden Metabolit-Profile genutzt, um auf Basis verschiedener statistischer Prozeduren die zugrundeliegenden Stoffwechselprozesse und potentielle Biomarker für die Merkmale TSV, pH-Wert und Fleischfarbe aufzudecken. Für TSV war ‚Random Forrest Regression' die geeignetste Methode um zuverlässige Biomarker zu identifizieren. Mittels eines systembiologischen Ansatzes wurden im nächsten Schritt mehrere Omics Ebenen verknüpft, um die Informationsdichte bei der Analyse verschiedener Merkmale der Fleischqualität und Schlachtköperzusammensetzung zu erhöhen. Somit wurden in der zweiten Studie Metabolit- und Protein-Profile genutzt, um eine *Pathway*-Analyse durchzuführen. Diese ergab u.a., dass der Sphingolipid-Metabolismus einen signifikanten Einfluss auf den TSV hat. Basierend auf den identifizierten *Pathways* wurden einzelne Stoffwechselkomponenten als ‚intermediäre Phänotypen' für den TSV ausgewählt. In der dritten Studie wurden die vielversprechendsten intermediären Phänotypen für TSV und andere Merkmale mit Hilfe einer Netzwerkanalyse ausgewählt. Dazu wurden Transkripte, Proteine und Metabolite zu Modulen kombiniert. Neben den konventionellen Leistungsmerkmalen wurden ausgewählte ‚intermediäre Phänotypen' (einzelne Metabolite/Proteine bzw. Module der Netzwerkanalyse) in genomweiten Assoziationsanalysen (GWAS) untersucht. Im Rahmen dieser Arbeit wurden einige interessante Kandidatengene für TSV auf *Sus scrofa* Chromosom (SSC) 18 und die Schlachtköperzusammensetzung auf SSC5 detektiert.

Durch die erhöhte Informationsdichte zwischen Geno- und Phänotyp ist anzunehmen, dass GWAS von intermediären Phänotypen zu einer verbesserten statistischen Aussagekraft bei der Detektion von Kandidatengenen führen und falsch-positive Assoziationen vermieden werden können. Dies erlaubt, die genetische Fundierung komplexer Stoffwechselvorgänge aufzuklären.

Table of contents

List of figures

List of tables

List of abbreviations

AHCYL2	Adenosylhomocysteinase-like 2
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BFT	Backfat thickness
BLUP	Best linear unbiased prediction
BP	Biological Processes
CC	Cellular Components
CIF	Conditional inference forest
CPU	Central processing unit
CREB3L2	cAMP responsive element binding protein
CV	Coefficient of variation
DEG	Differentially expressed genes
DFD	Dark, firm and dry (meat)
Du × Pi	Duroc × Pietrain
EBV	Estimated breeding values
ENET	Elastic Net
ENPP3	Ectonucleotidepyrophosphatase/phosphodiesterase 3
eQTL	Expression quantitative trait loci
eSNP	Expression single nucleotide polymorphism
EXOC4	Exocyst complex component 4
F6P	Fructose-6-phosphate
FAs	Fatty acids
FBPase	Fructose-1,6-biphosphatase-2
FDR	False discovery rate
GBP4	Guanylate-binding protein 4
GC-MS	Gas chromatography-mass spectrometry
GG	Genetical genomics
Glycerone-p	Dihydroxyacetone phosphate
GO	Gene ontology
GPL	Glycerophospholipid

GS	Genomic selection
GWAS	Genome-wide association studies
h^2	Heritability
ICPL	Isotope-coded protein labeling
IMF	Intramuscular fat content
IPF-LASSO	Integrative 'Least Absolute Shrinkage and Selection Operator' (LASSO) with Penalty Factors
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
LC-QTOF/MS	Liquid chromatography - quadrupole time of flight - mass spectrometry
LD	Linkage disequilibrium
LMC	Lean meat content; in belly (LMC_{belly}) or calculated with Bonner formula (LMC_{bonn})
LR	Landrace
LRGUK	Leucine-rich repeats and guanylate kinase domain containing
LW	Large White
MAF	Minor allele frequency
MAR	Maximum adjacency ratio
MDA	Mean decrease in accuracy
ME	Module eigenvalue
MF	Molecular Functions
MFR	Meat fat ratio
mGWAS	GWAS based on metabotypes
MLD	*Musculus longissimus dorsi*
MLL	*Musculus longissimus lomborum*
MM	Module membership
MR-GSE	Mean-rank gene-set enrichment
MS	Metabolite significance
MSM	*Musculus semimembranosus*
mRNA	Messenger ribonucleotide acid
NMR	Nuclear magnetic resonance

OBS	Omics based selection
OOB	Out-of-bag
PC	Principle component
PCA	Principle component analysis
PGAM2	Phosphoglycerate mutase 2
pHu	Ultimate pH-value
PIK3C3	Phosphatidylinositol 3-kinase, catalytic subunit type 3
PKM	Pyruvate kinase (muscle)
PKN2	Protein kinase N2
p.m.	*Post mortem*
PPP	Pentose phosphate pathway
pQTL	Protein QTL
PSE	Pale, soft and exudative (meat)
PTPRT	Protein tyrosine phosphatase, receptor type
QTL	Quantitative trait loci
R^2	Coefficient of determination
RFR	Random forest regression
RN	Rendement napole (gene)
RMSE	Root mean square error
RSE	Reddish-pink, soft and exudative (meat)
SAMD4A	Sterile alpha motif domain containing 4a
SD	Slaughter date
SNP	Single nucleotide polymorphism
SRM	Selected reaction monitoring
SSC	*Sus scrofa* chromosome
SW	Slaughter weight
TOM	Topological overlap matrix
TPI1	Triose phosphate isomerase 1
UTR	3’ untranslated region
VI	Variable importance
WBSF	Warner-Bratzler shear force
WGCNA	Weighted gene Co-expression network analysis

WHC	Water holding capacity
WNA	Weighted network analysis

Chapter 1. General introduction

In livestock breeding, most quantitative production traits with high heritability are comprehensively investigated. Many major genes are detected and already established in genomic selection and the underlying formulas for calculation of genomic breeding values are validated in large reference populations.

Consequently, livestock populations are on a remarkable high genetic level in many economically important production traits like milk yield in cattle or carcass lean content in pigs. On the other hand, there are many low heritable traits, which gain an increasing interest in livestock breeding not only because of economic, but also of sustainability aspects like animal health and welfare or resource efficiency. These traits usually have a complex genetic foundation and are strongly influenced by a variety of environmental factors. Thus, the elucidation of the genetic background is associated with many challenges. To obtain significant associations between genetic markers and phenotypes, a large number of individuals is required and the phenotypes have to be quantified in a very precise way and under defined standardized conditions. In recent years, substantial progress in genomic data recording and statistical methodologies could be observed. Usually, the association between single nucleotide polymorphisms (SNPs) and phenotype is analysed by genome-wide association studies (GWAS). However, these analyses include DNA information but frequently ignore other omics levels like metabolome or proteome layer. Hence, the one omics layer GWAS approach may not be sufficient to decode the complex biological mechanisms and genomic architecture of low heritable traits [1]. As a consequence, it can be expected that many detected quantitative trait loci (QTL), particularly found in the analysis of low heritable traits, are presumably false positive results [2].

The novel scientific field of omics analyses provide promising approaches to uncover the genetic background of complex traits with higher accuracy. These analyses can be seen as a part of inter-disciplinary systems biology science that focus on complex interactions of all cellular and biochemical components within a cell or organism, to get a holistic understanding of biological processes. The major reason for the increasing use of omics approaches may be attributed to progress in molecular biology, particularly in genome sequencing and other high-throughput measurements. These technologies enable us to collect comprehensive data sets on systems performance and gain information on the underlying molecular biology [3]. The omics levels comprise all stages of genetic and metabolic regulation in organisms including the genome, transcriptome, proteome,

metabolome and phenotype. Based on the different quantified omics levels, the selection of metabolic phenotypes that can be used as more accurate indicators for classical performance traits is a promising concept and could be the next landmark in advancement of genetic analyses.

Within this thesis, we applied different statistical approaches to investigate the genetic and metabolic foundation and regulation of the low heritable meat quality parameter water holding capacity (WHC) or drip loss in Duroc × Pietrain (Du × Pi) pigs. For this purpose, based on different concepts to handle big data sets, we selected the most promising metabolites, proteins and transcripts and used them as bio indicators or metabolic phenotypes to increase the information density between candidate genes and traits of interest in genome-wide association analyses.

1.1 Meat quality traits in pigs

With above 110 millions of tons of worldwide pork consumption, in 2015 pigs (*Sus scrofa domesticus*) were among the most important animals for meat production (www.statistika.com). To meet the expectations of consumers, in the last decade, the meat production industry was increasingly focused on high quality pork products [4]. In general, quality includes intrinsic and extrinsic attributes [5]. Kauffman et al. [6] described the intrinsic meat quality as the 'sum of all quality factors in terms of sensory, nutritive, hygienic and toxicological and technological properties'. Sensory meat parameters include tenderness, flavour/smell and color while nutritive factors include fat, proteins and connective tissue content. The technological meat quality refers to several parameters like WHC, pH-value, intensity and homogeneity of color and firmness that affect the suitability for storage and processing in the respective production processes [7]. Deficiencies in pork quality lead to economic losses in meat production and reduced consumer acceptance. The attractiveness of pork to consumers particularly depends on its intrinsic, sensory characteristics, such as leanness, taste, odour, color, tenderness and juiciness [8].

1.1.1 Muscle composition and post mortem conversation into meat

Muscle composition and metabolic processes that occur within the muscle tissue before, during and after slaughter lead to the development of specific meat quality attributes [9]. With regard to a better understanding of the differences in pork quality a clear

comprehension of the tissue formation and the proceeding metabolic processes is necessary. Muscle tissue is mainly composed of the principle components water (~75%), proteins (~19%) and lipids (~4%), and the minor components (~1%) like carbohydrates, vitamins and minerals. Most of the water is bonded within the myofibrils and the cell membrane (sarcolemma) and between the muscle cells and muscle bundles [10,11].

Scheffler and Gerrard [9] described that "muscle tissue is able to keep homeostatic conditions by adapting its metabolic activity appropriate to changes in energy need and content of stored energy". In muscle cells, energy is generated in form of adenosine triphosphate (ATP) that is provided by fatty acids (FAs) "from adipose tissue, ketone bodies from the liver, glucose from the blood or stored glycogen and phosphocreatine in the muscle", (Nelson et al. [12]). Under the conditions of regular energy intake and rest, blood glucose levels are high and sufficient ATP is generated in the muscle cells to maintain the cell metabolism; glycogen is synthesized by glycogenesis and accumulated in the cells until energy is needed [13]. If required, muscle glycogen becomes available for metabolic processes like glycogenolysis and glycolysis [12]. In anaerobic conditions, the processes lead to the formation of lactate that reduces the pH value of the muscle. In normal physiological conditions, the original state is rebuilt while lactate is degraded in the liver. Within this process, lactate is transformed to pyruvate and in the next step, glucose is build up by gluconeogenesis [14].

After slaughter, metabolism in muscle tissue changes and leads to the conversion of muscle to meat. In the early *post mortem* (p.m.), with the absence of respiration and blood flow, the homeostatic conditions in the muscle cells break down. According to this, there is a shift from aerobic to anaerobic metabolism and within the muscle cells the availability of energy may become the limited factor with respect to muscle metabolism. Consequently, the muscle cells lack oxygen and the anaerobic glycolysis results in decreased glycogen and ATP levels in muscle. Correspondingly, lactic acid is accumulated and accelerates the rate of pH decline of the tissue from near neutrality to 5.4-5.8 (Fig. 1, A). Once the pH-value has reached the isoelectric point, positive and negative electrical charge of the proteins are equal. These positive and negative elements within the protein attract each other and cause a space reduction within the myofibrils [11]. Additionally, muscle pH declines until either muscle glycogen content is depleted or the function of glycogenolytic, glycolytic or fermentative enzymes is stopped while muscle is still warm [13,15]. This process results in

the denaturation of many proteins that are essential for many cellular functions and proteins involved in binding of cellular water [9,11]. This is accompanied by leakage of muscle cells and loss of water, ions and proteins [10]. It is the combined effect of protein denaturation and pH decrease due to the metabolism of muscle glycogen p.m. that results in the conversion of muscle tissue to the pork end product, determining the quality of the pork product [15].

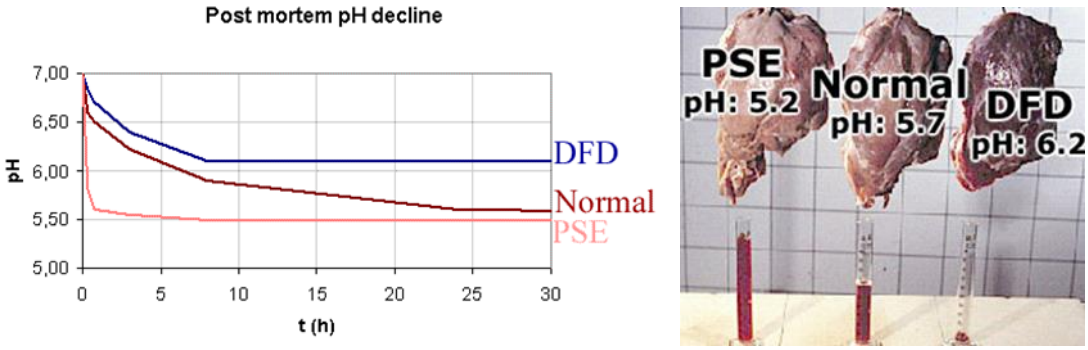


Figure 1: (left) Determination of *post mortem* (p.m.) pH-value in *Musculus longissimus dorsi*. DFD = dark, firm, dry; PSE = pale, soft, exudative (http://qpc.adm.slu.se); (right) Variation in color, pH-value 24 hours p.m. and drip loss of PSE, normal and DFD meat, measured in ham (http://qpc.adm.slu.se).

1.1.2 The role of water holding capacity in meat quality

One important meat quality parameter is water holding capacity (WHC) that is defined as the ability of meat to hold all or part of its intrinsic water [16]. WHC is the result of biochemical and biophysical processes, occurring in the early p.m. muscle to meat conversion [11]. Based on weight losses and restricted processing suitability, WHC is both an economic relevant meat characteristic as well as crucial for the consumer acceptance. Pork with a low WHC is unappealing to consumers because of the negative visual impression caused by liquid exudate [9]. On the other hand, meat with very high WHC is

also disadvantageous due to its dry and sticky nature [9]. Accordingly, WHC is strongly associated with meat quality parameters juiciness and tenderness. Thus, WHC affects the financial output, nutritional value and sensorial and technological properties of meat. The loss of inherent water retained in muscle cells that occurs due to the effects of pre-slaughter and p.m. processes, the drip loss phenomena is the main parameter of WHC and causes the incidence of unacceptable tenderness of pork.

The highest drip loss is often found in 'Pale, soft and exudative' (PSE) meat. PSE meat has a lightness score of L^*>50, a drip loss value of >6%, a pH-value two hours p.m. (pH2) of <5.8 and a pH24-value of between 5.3 and 5.7 (cf. Fig. 1) [17]. Besides possible genetic predisposition of PSE meat, there are also other factors causing exudative meat. Major factors affecting the variation of WHC in pork include temperature and pH p.m., which define the dimension of myofibrillar protein denaturation and the net charge of the myofibrils [18]. High temperatures and low pH values in meat are associated with high levels of proteolysis, whilst expended pH decrease is related with lower net burden of myofibrils [9,18]. When the cytoskeletal (myofibrillar) proteins denature, a reduction in myosin head length within the muscle fibres leads to closer contact between thick and thin filaments [18]. This leads to expulsion of water [18]. Alternatively, Scheffler et al. [9] describe that with reduced net charge of myofibrils, myofilaments are attracted towards each other forcing water out of the muscle fibre lattice. Ultimately, both leads to an increase in drip loss and tenderness and reduces the WHC of pork. Rapid pH decline, protein denaturation and high drip loss can be caused by short term stress before slaughter [19].

In contrast to PSE meat, 'Dark, firm and dry' (DFD) pork is characterised by an increased WHC. DFD meat that has a pH24-value of >6, a drip loss value of <2% and an L^* value of <43, shows a raised tendency to for microbial deterioration and bad dry curing qualities (cf. Fig 1).

A wide variety of methods are used for indirect measurement of WHC in meat. These methods can be differentiated based on the definition of what kind of fluid extracting out of the muscle tissue is measured. Depending on the force that is applied to the meat sample more or less fluid is expelled from the muscle cells. For the estimation of WHC, forces such as pressure (filter paper press method) or suction (filter paper method) has to be applied to the meat sample and the amount of released water is determined. The most

common method in WHC recording is the filter paper press method due to Grau and Hamm [20]. This method is applied in meat inspection, although it is labour-intensive because of cutting, homogenising and weighing procedures [21]. The filter paper method is cheap, easy, very fast and no special facilities are needed for the visual scoring [22]. The third described method for the indirect measurement of WHC is the drip loss method. In the drip loss method, pressure is exerted by gravity and shrinking during storage [21]. Offer and Knight [23] defined drip loss as a fluid consisting of water and proteins, expelled from the meat surface without any mechanical force other than gravity that is quantified as proportion of weight loss due to fluid exudation using gravitational techniques (Fig. 1, B). Diverse techniques were developed for the direct measurement of drip loss: the loose bound water method, the capillary volumeter method, the tray method, the bag method of Honikel [16], the EZ-DripLoss method [24] and the centrifugation method [25][25]. Particularly, the bag method is simply to implement and needs no special equipment, but takes at least 48h to obtain results.

Tab. 1 indicates that the quantification of WHC or drip loss strongly depends on the performed method. The correlation between filter paper press, filter paper and drip loss (bag) methods are rather low, whereas a high correlation can be observed comparing the bag and EZ-DripLoss method. These results were to be expected because different (drip loss, filter paper, filter paper press) or equal (bag, EZ-Driploss) forces were applied to remove intrinsic water of meat. In following, unless otherwise stated, we set the measurement of drip loss based on the bag method as the method of choice (golden standard) to quantify WHC. This is also stated by Honikel [16] and Otto et al. [24].

Besides the described methods for the indirect quantification of WHC, using filter paper press, filter paper or drip loss method, there are several other parameters, which can be used as indicators for WHC: pH-value in meat 1h p.m. and light scattering and reflectance of meat (meat color) are classical on-line techniques predicting WHC. In contrast to most WHC traits, these indirect parameters can be measured on-line on intact carcasses in the slaughterhouse in a cost effective manner. As can be seen in Tab. 1, drip loss is strongly correlated with pH1 (r_p=-0.67) and pH24 (r_p=-0.51) and to a minor degree to meat color (r_p=-0.30). The magnitude of these correlations can be explained by the physiological connection to WHC given in section 1.1.1 (p. 3).

Table 1: Phenotypic correlation coefficients (r_p) between different techniques to estimate water-holding capacity and drip loss. Modified by van Oeckel et al. [21], Otto et al. [24], Sellier [26] and Borchers et al. [25].

Method	Filter paper	Drip loss	pH1	pH24
Filter paper press	0.10	0.35		
Filter paper		0.20		
EZ-DripLoss		0.86		
pH1		-0.67		
pH24		-0.51	0.33	
Color		0.41	-0.50	-0.13

- The filter paper press method based on a modification of the method of Grau and Hamm [20]: samples of 300mg homogenised meat were placed on a filter paper between two cover glasses under a pressure of 1kg for 5min. The difference between the areas of the pressed meat and the wet area on the filter paper was determined by planimeter.
- The filter paper method was carried out as described by Kauffman et al. [22]: A filter paper of known weight was applied to the meat sample with a 90g rubber plug, the sample was first exposed to the air for 15min. After 2s the filter paper was removed and reweighed to give weight of absorbed water.
- Drip loss method based on the bag method of Honikel [16]: 150g meat samples (free of external fat and connective tissue) were hung by a nylon cord in a plastic bag at 4°C for 48h, ensuring the meat had no contact with the juice in the bag. The difference in weight of meat sample (after superficially wiping dry), before and after the hanging, divided by sample weight ×100 yields the % drip loss.
- EZ-DripLoss method: 10g meat samples were placed in pre-weighed drip loss containers. After 48h, each container was weighed including meat and drip loss and once again for only drip loss.
- All traits measured in *Musculus longissimus dorsi*; pH was measured 1h (pH1) and 24h (pH24) *post-mortem* (p.m.); Meat color quantifies the light reflectance (CIE L* value) measured 24h p.m.

1.1.3 Genetic foundation of water holding capacity

Drip loss, as an indicator for WHC, has a low to medium heritability (0.01 to 0.31) and is genetically and phenotypically correlated with meat quality traits pH1, pH24 and color [26]. The heritabilities of these traits are very heterogeneous (cf. Tab. 2). While pH24 has a maximal heritability of 0.39, the heritability of meat color reaches a maximum of 0.56. It might be reasonably assumed that the wide range of heritability estimators is the result of the different recording systems. Different techniques for quantification are more or less prone for measurement errors due to random environmental factors like imprecise sample collection or preparation or not standardized testing environment. In all these cases, the measurement errors lead to a higher phenotypic trait variation and complicate the clarification of the associated genetic background of the traits.

Table 2: Heritabilities and mean genetic correlation coefficients between meat quality parameters drip loss, pH1, pH24 and meat color.

Trait	Drip loss	pH1	pH24	Color
drip loss	**0.01 to 0.31**	-0.55 to 0.01	-0.99 to -0.50	0.49
pH1		**0.04 to 0.41**	0.49	-0.38
pH24			**0.07 to 0.39**	-0.65 to -0.38
color				**0.15 to 0.57**

Drip loss (based on bag method), pH1, pH24 and color (light reflectance, *CIE L** value) measured in *Musculus longissimus dorsi;* diagonal (**bold**) = heritability (h^2); upper triangle = mean genetic correlation (r_g); extracted from Sellier [26] and Borchers et al. [25].

Drip loss has medium to high negative genetic correlations to pH1 and pH24. High drip loss is correlated with strong pH decline in muscle 1h and 24h p.m. The genetic correlation between drip loss and meat color has a medium level. Consequently, high drip loss is correlated with high light reflectance (pale color) in muscle 24h p.m.

The presented heritabilities and correlation coefficients are reliable estimators of the genetic foundation of drip loss and other WHC related meat quality traits, under the assumption of a polygenic determination. Due to the continuous normal distribution of the trait expression, it can be expected that many genes are involved in the genetic determination, so that the polygenic assumption can be justified. The low heritability of traits related to WHC implicates that many environmental factors are involved in the phenotypic expression of meat quality. The contribution of many genes and environmental conditions complicate the identification of single genes with significant effect on meat quality [27]. Nevertheless, several major genes are widely investigated and discussed:

(1) PSE meat is genetically caused by a mutation on the ryanodine receptor / calcium release channel (RYR1) or halothane gene on *Sus scrofa chromosome* (SSC) 6, that regulates Ca^{++}transport across muscle cell membranes [28]. The recessive mutation (R614C missense mutation) at the RYR1 locus causes susceptibility to stress (porcine stress syndrome), malignant hyperthermia syndrome (MHS) and a 90 to 95% incidence of PSE. However, with the help of a commercial available test for the RYR1 mutation, today, the German pig production has mostly eliminated this mutation in pig populations used for fattening [ZDS 29,30];
(2) The rendement napole (RN) gene, identified in the Hampshire breed, is responsible for increased glycogen content of the ‘white’ (fast-glycolytic) muscle types. A dominant mutation in the adenosine monophosphate (AMP)-activated protein kinase, y-3 subunit (PRKAG3), also known as RN gene, is associated with lower Napole yield and leaner carcasses. As a result, the meat quality is reduced caused by a lower pH based on p.m. degradation of glycogen. The meat is also called ‘acid meat’ [31]. Other SNP variants in PRKAG3 have positive effects on pork quality traits including ultimate pH, meat color, WHC, drip loss, tenderness, and cooking loss [32];
(3) The CAST gene, coding for the calpastin pathway, is correlated with several WHC traits [33], juiciness [34], pH24 and color [35];
(4) Insulin-like growth factors IGF-1 and 2 and leptin and leptin receptor LEPTIN and LEPR [30,36,37];
(5) Other genes, for which associations with drip loss and pH have been reported, include myogenin [38], pyruvate kinase muscle isozyme 2 (PKM2) [39], and troponin I [40].

Regarding economically important traits including meat quality, the dominant aim of pig breeding programs is selection towards reduced frequency of specific mutations with negative effect and enrichment of positive allele variants in commercial pig populations. However, it must be taken into consideration that there are some genetically antagonistic relations between carcass composition and meat quality traits resulting from antagonistic pleiotropic or linked gene effects (see Tab. 3). Carcasses with low lean meat contents (LMC) are associated with high pH24-values that induce DFD meat. On the other hand, carcasses with low meat fat ratio (MFR) (fatty meat) are associated with low pH1 in meat that increases the risk of PSE meat. Moreover, low MFR leads to tender and juicy meat. These relations should be considered in formation of a balanced breeding goal that mostly comprises high slaughter yields of high quality meat with favourable meat to fat composition.

Table 3: Genetic correlation coefficients of meat quality parameters water holding capacity, drip loss, pH1, pH24, meat color and tenderness with carcass composition traits leanness and fatness.

Trait	Carcass leanness	Carcass fatness
WHC	0.10	0.26
drip loss	-0.13	0.15
pH1	0.16	-0.21
pH24	-0.19	0.02
color	0.05	-0.10
tenderness	-0.20	0.24

WHC – water holding capacity; Ciobanu et al. [32].

In recent years most important genes affecting WHC were detected via less effective marker technologies like microsatellite genotyping. Due to progress in automated recording of meat quality and development of high throughput genotyping, nowadays it is possible to identify QTL or genes for meat quality even with smaller effects. Up to now, the Pig QTLdb has collected 13,030 QTL for 663 traits from 477 publications (update: October

2016; http://www.genome.iastate.edu/cgi-bin/QTLdb/index). Fig. 2 shows the detected QTL for drip loss across all porcine autosomes so far. These QTL were mainly identified via genome scanning based on linkage analysis and microsatellite genotyping. QTL studies have been greatly enhanced by development of GWAS, which use porcine SNP chips to genotype above 60,000 SNPs distributed across the whole pig genome.

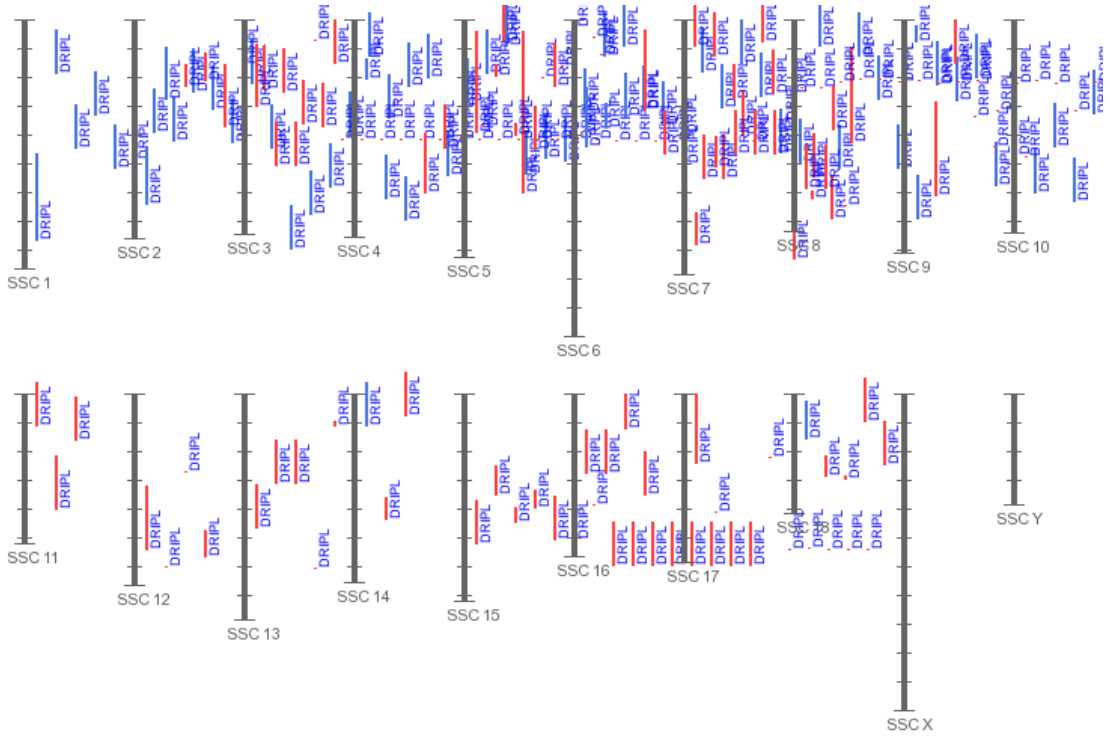


Figure 2: Quantitative trait loci (QTL) for drip loss in the pig genome. Red QTL lines represent for significant and light blue lines for suggestive statistical evidence. (http://www.animalgenome.org; 5 December, 2016).

Besides obvious progress in genotyping technologies, nowadays more advanced and accurate methods in recording meat quality and its underlying physiological parameters are available. These novel phenotypes are derived within all intermediate levels of the genome to phenotype axis (transcripts, proteins and metabolites). They can be used as biomarkers in selection or as so-called metabotypes in genetic analysis to increase the accuracy and reliability of GWAS.

1.2 The different omics levels

In the last two decades, many biology related scientific fields have transformed from primarily empirical and observational fields towards systems-level understanding. In order to summarise the role of different omics levels, the definition of Liu [41] and Schadt et al. [42] are given. Liu [41] pointed out that in different omics approaches the biology of living organisms is explained by structure and dynamics of cellular and organismal functions rather than by characteristics and effects of isolated parts of a cell or organism. Accordingly, Liu [41] stated that "systems biology requires the ability to digitalise biological output, the computational power to analyse comprehensive and massive data sets, and the capacity to integrate heterogeneous data from multiple experimental sources using interdisciplinary tools, the so-called 'omics' technologies into a usable knowledge format. Thus, systems biology can be described as 'integrative biology' with the ultimate goal of being able to predict *de novo* biological outcomes". As a consequence, Liu [41] and Hood et al. [43] believe that integrative, systems approaches provide unbiased and complete data sets about a biological system and interconnections of metabolic components. Moreover, according to Liu [41], systems biology allows the ranking of the often large number of involved molecules and pathways with respect to their importance for specific trait related processes. In a similar way, Schadt et al. [42] defined that omics technologies refer to a group of high-throughput research tools, including genomics, transcriptomics, proteomics, and metabolomics (Fig. 3). These tools are based on comprehensive analyses of all levels of genetic and metabolic expression. Schadt et al. [42] posit that by integration of a diversity of data like DNA variation (SNPs), gene expression (transcripts), protein-protein interaction, metabolite expression, and other types of molecular phenotype data, like epigenetic markers, more comprehensive networks of genes and metabolic components, so-called metabotypes, both within and between tissues and cells can be constructed to present a more detailed picture of the molecular processes underlying physiological states associated with a given phenotypic expression.

Until a few years ago, a general paradigm in molecular biology was the unidirectional information flow from genes to transcripts, which are translated into proteins. In the next step, enzymes regulate metabolic pathways and thereby affect the final phenotype of an individual (see Fig. 4, A) [44,45]. Based on new findings in the field of systems biology, Hollywood et al. [46] postulate that today it is known that the traditional 'linear' thinking is

often misleading. Actually, biochemical processes at the level of a cell or organism are intimately networked with many feedback-loops (see Fig. 4, B) [46]. The levels of genes, transcripts, proteins and metabolites are mutually dependent and the mechanisms controlling the interactions are very dynamic and complex. Moreover, in particular at the levels of transcripts and proteins, many internal processes like alternative splicing of messenger ribonucleotide acid (mRNA) lead to an increasing variation of metabolic intermediates (see Fig. 4, B and C) [46]. The reconstruction, visualisation and interpretation of such relations are big challenges for the understanding of metabolic processes effecting complex traits like meat quality in livestock [44,46].

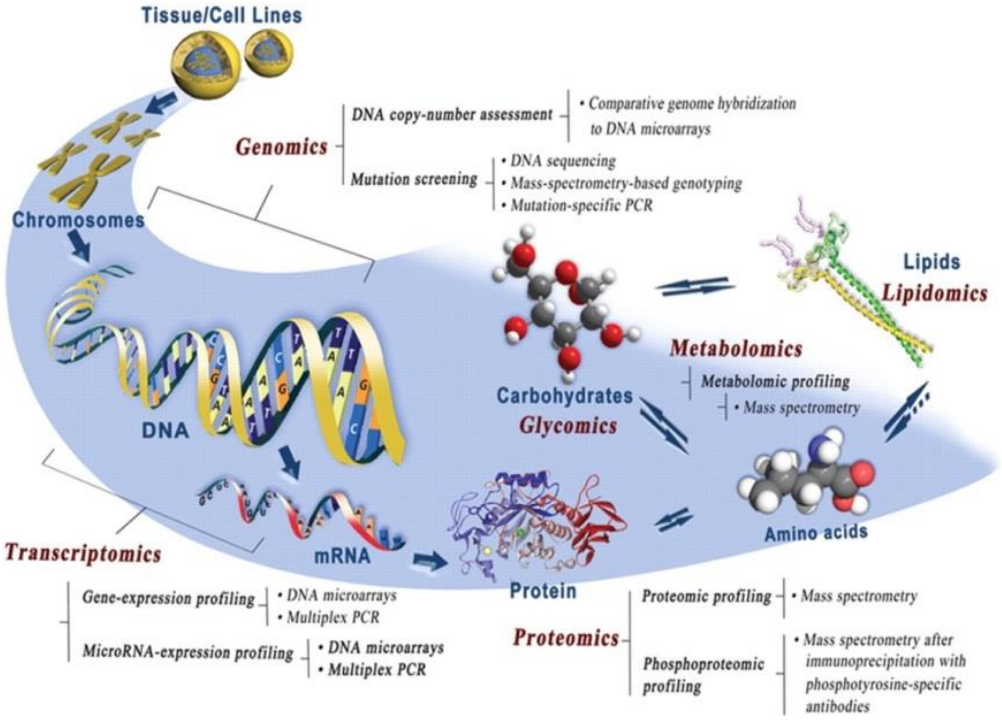


Figure 3: Schema of omics technologies, their corresponding analysis targets, and assessment methods. DNA is first transcribed to mRNA and translated into protein which can catalyse reactions and give rise to metabolites, which include glycoproteins, carbohydrate and lipids (Fig. from Wu et al. [47] based on Sawyers [48]).

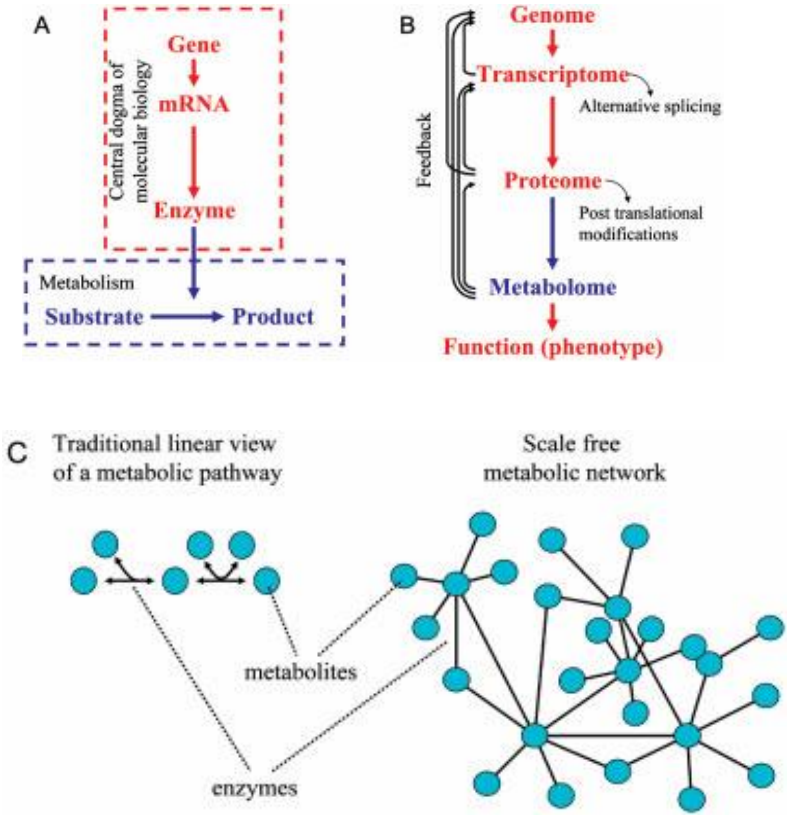


Figure 4: (A) Traditional paradigm of molecular biology. (B) The omics organization in systems biology. (C) Traditional linear view of a metabolic pathway and the now accepted view of scale-free connections in a metabolite neighbourhood; nodes = metabolites, connections = enzymatic action; Hollywood et al. [46].

1.2.1 Genomics

As mentioned in section 1.1.3 (p. 9), livestock breeding is focused on modifying the genetic constitution of important livestock species to make them better matching consumers claims [49]. Kadarmideen et al. [49] described that "the approach of selecting parents of a next generation is based on their (or relatives) phenotypes in economically important traits". Best linear unbiased prediction (BLUP) combines the recording of individual performance traits and the phenotypes of relatives into estimates of breeding values (EBV) [50]. From 1980 onward, progress in molecular genetics led to more genetic improvement compared to selection based on phenotypic records [49]. In the 1980s, marker-assisted selection detected

and (fine) mapped genes (QTL) influencing traits of interest and included the QTL information into the BLUP-EBV [51]. In the early 1990s, QTL mapping projects spurred a lot of enthusiasm [49]. According to Meuwissen et al. [51], QTL mapping studies resulted in huge QTL databases. Consequently, the Animal QTL database (http://www.animalgenome.org/QTLdb/) today reports several thousands of QTL for major livestock species, with increasing tendency. However, the reproducibility of the QTL maps is low, i.e. QTL positions move/(dis)appear from one to the other investigation [51]. As possible explanation for this phenomenon, Xu [52] presented that the bulk of QTL have minor effects. Combined with the situation that a huge number of genetic markers is tested, due to Xu [52] there is a marked 'Beavis effect' in QTL mapping studies. This phenomenon leads to overestimation of significant markers and is still challenging in genetic analyses.

Moreover, most of the QTL were detected using microsatellite markers covering large regions of the genome that contain a variety of genes [53]. Therefore, it was difficult to detect the causative genes of QTL variation and the fine mapping of the initially mapped QTL was a persisting challenge [54]. Subsequently, the release of whole genome sequences of major livestock species has led to a paradigm shift in availability of thousand up to one million genetic markers today. These markers are genotyped using high-throughput Affymetrix or Illumina genotyping platforms (DNA arrays or SNP chips) [53]. Such high-density SNP information was the basis for the introduction of advanced statistical and computational genetic procedures in terms of GWAS [49].

1.2.1.1 Genome-wide association studies

In the first phase, high-throughput genotyping data was predominantly used in GWAS, which test the association of SNPs and observed phenotypes and provide estimates for hundreds of thousands of genetic markers on each phenotype considered [53]. The aim is to detect statistical significant associations between phenotype and specific genetic markers [55]. Most commonly, the data set is analysed by the investigation of one SNP at a time using simple linear models. It is defined to a fundamental principle that with increasing number of individuals applied for GWAS, the statistical power to detect potential associations rises [56]. Based on the low population-wide LD between SNPs and causal variants in livestock and the very high level of significance (e.g. $p<1\times10^{-7}$) that is needed to overcome the multiple testing corrections, GWAS require a larger number of tested individuals in contrast to linkage-QTL studies [53]. Besides the number of individuals, also

the population structure, as possible source of bias in livestock studies, has to be considered. Relationships among animals cause LD between loci even if the loci are functionally unlinked [56]. For this reason, population stratification, which means that the individuals derive from mixture of breeds or families, is an important source of false positive associations. As second key principle can be invoked that, with increasing SNP density, the power to detect significant associations between SNPs and trait expression will increase [56]. Goddard and Hayes [56] concluded that most SNPs have small effects which might lead to a random noise, whereas markers in regions in which the causative mutations are located have much larger effects and are potentially useful tools to localise the chromosomal region of the causative mutation.

To avoid incorrect associations, based on minor allele frequency (MAF) and Hardy-Weinberg disequilibrium, a quality control of the genotyping data has to be performed. Only effects and samples, which passed the quality thresholds, are considered in GWAS. Finally, a repeated analysis of identified associations in an independent population or the examination of the functional background should be performed [57]. While the standard GWA procedure typically examines one SNP at a time, it is also promising to perform GWAS based on haplotypes. Due to high LD or physical proximity, SNPs often do not segregate independently and are transmitted in haplotype blocks that comprise clusters of three to five tightly linked genetic markers [53]. Once a chromosomal region with significant SNPs is identified by GWAS, local haplotypes can be defined and their effects can be estimated and used in predictions for complex phenotypes [53].

Regardless of what type of GWAS is performed (either single SNP- or haplotype-based) there are still major technical and analytical challenges in GWAS (e.g. the ‘Large p, small n’–problem) [53]:

- Multiple testing (e.g. Bonferroni correction) leads to very conservative thresholds and may result in missing biologically relevant loci;
- An insufficient/small number of analysed animals reduces the power to detect loci of small effects and, on the other hand, there is an increased risk for overestimation of specific SNP (haplotype) effects;
- Population stratification may lead to spurious associations;

- For some species the SNP coverage (density) of the chips is still low increasing the likelihood that the causative mutation will not be identified;
- Poor model fitting (e.g. unaccounted epistatic and genotype-environmental interaction effects, inappropriate fixed and random effects) will lead to incorrect associations;
- Due to bad quality control concepts or other filtering mechanisms, rare variants and undetected copy number variation effects are excluded.

Regarding the last aspects, Kadarmideen [53] pointed out that bad quality control or filtering mechanisms describes the likely, most significant limitations of GWAS. Due to statistical problems, a fraction of the genetic information (for example rare variants) is rejected and thereby the GWAS are not able to explain the full genetic variation in complex traits. This situation was first described as 'missing genetic variation or heritability', after GWAS on complex traits, by Manolio et al. [58], Clarke and Cooper [59] and Gibson [60]. Several studies have demonstrated that a mixed or random model GWAS can picture a much larger proportion of the actually present genetic variation (e.g. for human height [61–63]). It can be assumed that SNP chips that only comprise common variants, but no rare variants, might be the cause of hidden heritability [64]. For this reason, when defining the thresholds in quality control, it should be taken into account that not too much of the genetic information is rejected.

1.2.1.2 Genome-wide association studies in pork quality

As a precondition of GWAS application, a SNP chip for the respective species has to be available. In pigs, the 'PorcineSNP60 v1 Genotyping BeadChip' © from Illumina offers 64,232 SNP markers. Hence, SNP chips for pigs contain less SNPs than human SNP chips, containing more than one million SNPs (www.illumina.com, Illumina, Inc., San Diego, USA, www.affymetrix.com, Affymetrix, Santa Clara, USA). Nevertheless, the porcine SNP chips are effectively applied in livestock breeding to improve performance traits and other trait complexes like disease resistance and behaviour traits [53]. In pigs, several GWAS were successfully performed for meat quality and revealed promising candidate genes for different complex traits. Based on a GWA study of 987 commercial end product Du pigs, Duijvesteijn et al. [65] identified clusters of candidate genes on SSC1 and 6 for the pork quality characteristic androstenone. In a GWA study performed by Luo et al. [66] in 455 pigs of a Large White (LW) x Minzhu intercross population, 45 SNPs showed significant

associations with one or various of the meat quality traits intramuscular fat content (IMF), marbling, moisture, and color. Out of the 45 SNPs, 36 SNPs are located on SSC12. These significantly associated SNPs were related with, or were located with minimal gap to, previously reported QTL and some were located within introns of already detected candidate genes [67]. Ma et al. [68] applied a GWA analysis in two different pig populations, 434 Sutai pigs and 933 F_2 White Du × Erhualian pigs, for the meat quality traits pH, color, drip loss, moisture content, protein content and IMF, marbling and firmness scores in *Musculus longissimus dorsi* (MLD) and *Musculus semimembranosus* (MSM). The authors detected in total 127 chromosome-wide significant SNPs for pH (on SSC3), drip loss (on SSC3 and 15), and IMF (on SCC 9 and X).

GWAS in pork quality traits were also successfully performed to confirm already known associations. In 2015, Zhang et al. [69] identified a QTL close to the PRKAG3 gene, affecting meat pH and colour, based on GWAS in 1943 crossbred commercial pigs. As already described in 1.1.3 (p. 9), the AMP-activated protein kinase, y-3 subunit, coded by the PRKAG3, has achieved attention as RN gene affecting the Hampshire effect that is associated with reduced Napole yield and leaner carcasses [31].

Comprehensive QTL analyses and GWAS were also performed in the Bonner Du × Pi population with the scope to investigate quantitative performance traits such as meat or carcass quality parameters. In the investigations of Liu et al. [27], QTL analyses in 585 F_2 Du × Pi animals were performed revealing 137 QTL for 35 traits of growth, meat quality and carcass composition. Especially for WHC, drip loss, pH1, pH24 and meat color in MLD 11 QTL were found on SSC1, 2, 3, 4, 5, 6, 7, 15 and 18 [27,70]. For these QTL analyses, 106 microsatellites across the 18 porcine autosomes were used. Other scientific groups applied QTL approaches in Berkshire × Yorkshire pig and identified several QTL for drip loss on SSC2, 3, 5, 9 and 18 [71,72]. These studies give evidence that due to the low density of microsatellite markers, QTL are often mapped to a large interval of two Mb or more. Consequently, it is a big challenge to identify the causative mutations or so-called quantitative trait nucleotides that are directly associated with complex traits [73,74]. Today, GWAS based on the current 60K porcine SNP panel of Illumina® provide a promising opportunity to receive more dense genetic markers than microsatellite markers, which helps to improve accuracy in finding candidate genes for complex traits and the underlying causative mutations [66].

However, the distribution of marker effects showed that most SNPs affecting meat quality traits 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 SNPs with the largest associated effects from genomic analysis might not track perfectly the causative mutations, these are potentially useful tools for identifying promising chromosomal regions [75]. Therefore, fine mapping techniques such as RNA sequencing (transcriptomics) or 'genetical genomics' approaches are needed for further investigation of possible candidate genes.

1.2.2 Transcriptomics

In contrast to genomic approaches that focus the static aspects of genomic information (SNP state), transcriptomic approaches investigate the gene expression profiles to identify genes that differ in expression between experimental samples [76]. Due to Cassar-Malek et al. [77], "the sequencing of the genome and the transcriptome as well as the increasing availability of different types of microarray analysing platforms in the last decade provided new possibilities to clarify the molecular background of physiological and productive functions in livestock species and their regulation". By technical progress, commercial oligonucleotide chips (www.affymetrix.com; www.agilent.com) or oligonucleotide sets (www.operon.com) became available for many livestock species [77]. Based on this development, many gene expression studies in several fields relevant for livestock like nutrigenomics and characterisation of production systems or meat quality were performed [77]. Today, high-throughput microarray technologies enable the simultaneous measurement of up to tens of thousands transcripts at the same time. Most transcriptomics experiments are focused on detection and annotation of differentially expressed genes (DEG) and co-expressed genes as well as construction of gene networks and to elucidate the relevant genetic regulation and interactions under the given environmental conditions [76].

1.2.2.1 Transcriptomic approaches in pork quality

Transcriptomic analyses are based on the ability to simultaneously analyse hundreds or thousands of transcripts (mRNA) on specific arrays. In transcriptomics of meat science, it is

the final objective to identify meat quality transcript biomarkers that are quantifiable on living animals or early p.m. on the carcass, in order to orientate meat production towards the most adapted processes in meat processing [78,79]. The quantification of these expression biomarkers will allow predicting the final quality level of meat from growth to slaughter period. With the application of biomarkers such evaluation can be performed within hours compared to classical methods that are performed days after slaughter or even after meat aging and are often expensive and/or invasive. Furthermore, the use of transcript biomarkers allows to understand the interactions between genetic and environmental factors that affect the manifestation of complex phenotypes like meat quality [1]. This knowledge will help to adapt production systems and breeding efforts of animals to their individual meat potential and to better assignment of meat to a suitable branch of production due to its characteristics.

As in other livestock animals, pork quality results from complex interactions between animal genetic background, rearing and slaughter conditions, and carcass and meat processing. According to Picard et al. [1], "even though many factors influencing pork quality have been highlighted, its variability remains high". The identification of meat quality biomarkers has been conducted by comparing contrasted groups for a given trait and today great efforts are undertaken to reveal promising transcript predictors for one or several meat quality traits, simultaneously.

As an example, one of the very early transcriptomic analyses was performed by Damon et al. [80], who tends to highlight transcript biomarkers for the meat quality defect PSE. For a better understanding of the PSE phenomenon, they took MSM samples 20 minutes p.m. and identified different up-regulated transcript profiles encoding myofibrillar proteins involved in actin-myosin interaction and sarcomere integrity and enzymes of glycolytic pathway in PSE muscles [80]. Transcriptomic approaches have also been performed to expose the underlying molecular mechanisms and to highlight transcript biomarkers of IMF that is crucial for eating quality and acceptability of pork. By comparing two groups of related pigs, but with contrasted IMF levels in *Musculus longissimus lomborum* (MLL) at 110 kg slaughter weight (SW) (1.36% vs 4.58%), Liu et al. [81] highlighted differential transcript expression levels of genes associated with glucose, lipid and protein metabolic processes.

Regarding meat quality traits, the microarray transcriptomic analysis in MLL of two pig breeds (French local Basque, LW, respectively n=50) showed that breed differences in muscle physiology and meat quality were associated to differences in metabolic processes, skeletal muscle structure, extracellular matrix, and proteolysis [82]. The results allowed the identification of transcript biomarkers of ultimate pH (pHu), drip loss, meat color, IMF content, Warner-Bratzler shear force (WBSF) and tenderness [83]. An external validation of the selected biomarkers was performed in 100 commercial Du × Landrace (LR) × Yorkshire pigs [84]. Among the potential expression biomarkers is FOS (FBJ murine osteosarcoma viral oncogene homolog) that is involved in response to calcium transport and that is positively associated to pHu and negatively associated to drip loss and meat color [83].

Altogether, this short overview showed that transcript biomarkers of pork quality have been successfully identified and validated. However, due to Alessandro and Zolla [2], further work is needed to improve the predictive capacity of transcript biomarkers in order to foresee the development of control tools for pork industry.

1.2.3 Proteomics

Proteomics analyses measure the entire complement of proteins, also called proteome, in terms of its presence and relative abundance. The definition of the proteome was characterised by Wilkins et al. [85] and comprises the total amount of proteins expressed at a certain time point. The term proteomics was established as an analogy with genomics or transcriptomics. According to remarkable enhancements in the accuracy, sensitivity, speed and throughput of mass spectrometry and development of powerful analytical software, scopes and efficiency of quantitative proteomics has grown enormously [86]. Following genomics and transcriptomics, proteomics is one of next levels of a systems biological understanding. Due to Mullen et al. [87], besides the systematic determination of protein sequences, proteomics also considers quantity, modification status, interaction partners, activity, subcellular localisation, and structure of proteins in a given cell type. While the genome contains the information about which genes are available, the proteome contains the information about which genes are actually being expressed [88]. Because genes might be present, but not transcribed and because the number of mRNA copies does not always

reflect the number of functional proteins that are present [89], neither the genomic DNA code nor the amount of mRNA that is expressed for each protein, yields an accurate picture of the current metabolic state of the individual or cell. For this reason, the analysis of the proteome is useful to complete the picture of the underlying metabolic processes and it can be seen as molecular link between the genome and the phenotype [88]. Moreover, Bendixen [90] defined that the aim of proteomics is to receive insight into cellular protein expression, post-transcriptional modification, subcellular localisation, turnover and interaction with other proteins and hence to reveal the function of genes. Finally, it is the purpose of proteomics to explain how genetic and environment interact to control cellular functions and metabolic processes [90].

Bendixen [90] predicted that, in contrast to prokaryotes, in higher organisms the number of proteins significantly exceeds the number of genes according "to alternative splicing and post-translational modifications like glycosylation or phosphorylation". From this, Bendixen [90] concluded that in contrast to the static structure of the genome, the proteome is continuously changing due to factors influencing on either protein synthesis or degradation. Thus, proteomics approaches can be viewed as taking a snapshot of the proteomic background of a biological system at a defined time point and under given environmental conditions. However, as described by Ghaemmaghami et al. [91], who successfully determined the whole proteome of yeast, a global proteomics approach is a challenging task. Due to the multitude and diversity of proteins in higher organisms, a wide range of technologies must be used to prepare, separate and quantify the relative expression levels of thousands of proteins in parallel [90]. Typically, mammalian tissue samples contain between 10,000 and 30,000 different protein species [90]. In one experimental setup, in practice only a few hundred to several thousand proteins, presenting just a part of the entire proteome, can be analysed. Mullen et al. [87] pointed out that a proteomics profiling approach generates a huge amount of data and that processing and interpretation of the proteomic data is challenging. In contrary, profiling of a selected set of proteins is far easier and depending on the regarding scientific issue, targeted protein profiling is as beneficial as profiling the whole proteome, but less elaborate and expensive [87]. For example, with regard to proteins influencing meat quality, working with myofibrillar, exudate or sarcoplasmatic extracts may be more manageable than targeting to profile the whole proteome [92].

In general, Jayarsi et al. [93] differentiate between two approaches for the "proteome characterisation, namely comparative proteomics and mapping proteomics". Similar to genome databases resulting from genome sequencing projects, mapping proteomics has the objective to characterise 'cellular proteomes' and to collect them in comprehensive databases [90]. Due to the complex variety of modifications and constantly changes over time and in physiological state, such approaches are a huge task [94]. Based on the fact that genetic variations may cause phenotypic differences that can be studied using proteomics, comparative proteomic studies investigate the biochemical processes that constitute the link between phenotype of interest and underlying genotypes [95]. From this, Hunter et al. [95] concluded that moment-by-moment snapshots of cellular responses at the protein level are the consequence of transcriptional and translational events.

1.2.3.1 Application of proteomics in pork quality improvement

Over the last decades proteomic studies in livestock animals have developed quickly. In pigs, the investigations mainly focused muscle growth, breed differences and meat quality traits, such as tenderness, juiciness, flavour, color and WHC [90]. Moreover, Mullen et al. [87] summarised that knowledge gained from proteomic approaches might be advantageous in recording of environmental factors and optimising of management and husbandry systems for reliable, high quality standards and in adapting quality to consumer and market needs.

In the field of meat quality traits, proteomic studies investigate the p.m. biochemistry of skeletal muscle, often in a breed-specific way. Due to Paredi et al. [96], the major goals are "to deepen the understanding of physiological changes, taking place at the protein level," to light the black box of the 'muscle to meat conversion' and to identify reliable protein biomarkers [90,96]. Proteins tend to be promising biomarkers for meat quality because most parameters, like WHC, pH drop and Minolta values – lightness, redness and yellowness – that are related to meat color are intrinsically affected by muscle protein composition. As an example from Paredi et al. [97], red muscle relies on oxygen for their metabolic functions and displays higher concentrations of the heme group-carrying protein myoglobin. PSE meat that is characterised by deficiencies in exudation, color and texture is induced by a reduced proteolytic rate of proteins troponin T, myosin light chain and α-crystallin, and the lack of heat shock protein 27 [96]. These proteins might be potential biomarkers for PSE meat [97]. Furthermore, in pigs with a mutation in the HAL gene there

is a higher risk for the generation of PSE zones [98]. In homozygous pigs, it was found a lower quantity of i) proteins involved in the ATP synthesis, ii) antioxidant proteins, including glutathione peroxidase, glutathione transferase and five proteins related to the aldehyde dehydrogenase family, and iii) chaperone related proteins, such as HSP27, alpha crystalline B chain, Hsc71 and Hsc70/Hsp90 [96]. The absence of these proteins was linked to low pH and high temperature in muscle, causing an increased proteolytic rate of the glycolytic enzymes [10].

In a comparative investigation in breeds of Casertana (low lean meat content) and LW (high lean meat content) pigs, D´Alessandro et al. [99] investigated breed-specific differences in the protein expression. These authors identified breed-specific protein profiles related to growth performance, fat accumulation and p.m. performance through a direct influence on the forcedly anaerobic behaviour of pig muscles after slaughter. In Casertana, but not in LW, higher levels of glycolytic enzymes (e.g. glycerol-3-phosphate dehydrogenase) and lactate accumulation were associated with slow pH drop, although it did not produce lower ultimate pH in LW. On the other hand, pyruvate kinase M1 and tropomyosin levels were related to WHC and meat color after slaughter in LW [99]. In Casertana pigs, different levels of glycolytic enzymes, heat shock proteins (HSPB6) and anti-oxidant enzymes (SOD1, glutaredoxin and lipoxygenase) were correlated to higher proteolysis and modestly lower WHC [99]. An overview of the most important protein biomarkers related to meat quality traits in pigs is presented in Tab. 4.

Table 4: Proteins involved in the variability of pork quality traits.

Trait	Muscle	Protein biomarker	Ref.
tenderness	MLD	actin, MHCK, MLCK II, TPI1	[100]
tenderness	MLD	desmin, actin, MHCK, MLCK, troponin T, tropomyosin α1 and α4, thioredoxin, capZ	[101]
PSE (WHC, color)	MSM	troponin T, MLCK 1, α-ß-crystalline	[102]
PSE (pH decline rate)	MLD	pyrute kinase, TPI1	[103]
color, WHC	MLD	cofilin 2, troponin T, α-ß-crystalline, HSP27, group chain A-ALDH, G3P-DH, hemoglobin α-chain, DJ-1 protein	[104]
WHC	MLT, MLL	HSP70	[105]
drip loss	MLD	creatine phosphokinase M-type, desmin	[106], [107]
color	MSM	ATPase beta subunit, NADH-DH, succinate-DH, hemoglobin, HSP27, α-crystalline, enolase 1 and 3, G3P-DH, transferrin	[108]
drip loss	MLD	lactate-DH, antichymotrypsin, myosin, isocitrate-DH	[109]
ultimate pH		14-3-3 protein, MLCK	

WHC – water holding capacity; PSE – pale, soft, exudative meat; MLD – *Musculus longissimus dorsi*; MSM - *Musculus semimembranosus*; MLT - *Musculus longissimus thoracis*; MLL - *Musculus longissimus lumborum*; MHC - myosin heavy chain kinase; MLC - myosin light chain kinase; TPI1 - triose phosphate isomerase 1; HSP - heat shock protein, G3P-DH - glycerol 3 phosphate dehydrogenase, ALDH - aldehyde dehydrogenase, DH – dehydrogenase; ref – reference.

1.2.4 Metabolomics

Similar to transcriptomics and proteomics, Fontanesi [110] defined the role of metabolomics as follows: “Metabolomics reveals the molecular readout of the biochemical state of a biological system that can be depicted by the metabolic species and their level that is derived by a specific combination of gene expression activity and environmental factors”. Whilst metabolomics is applied as complementary tool to transcriptomics and proteomics, it may be seen to have special advantages, which were described by Hollywood et al. [46]: “Since the traditional ‘linear’ information flow from one gene to one transcript to one protein is no longer true, it seems obvious that cellular processes are in reality intimately networked with many feedback-loops” (see Fig. 4B, p. 15). As a consequence of this paradigm shift, it can be assumed that although changes in the levels of individual enzymes may be expected to have only a small impact on metabolic fluxes, they still can have substantial effects on the concentrations of a multitude of metabolic components [46]. In addition, ter Kuile and Westerhoff [111] pointed out that as the ‘downstream’ result of gene expression changes in the metabolome are amplified relatively to variation in the transcriptome and proteome expression, which is likely to lead to increased sensitivity. Finally, it is known that metabolic fluxes are not regulated only by gene expression but by post-transcriptional and posttranslational events and as such, the metabolome can be considered closer to the phenotype compared to the transcriptome and proteome [111].

The entirety of all existing metabolic species constitutes the metabolome [110]. The metabolome comprises the quantitative complement of all low-molecular weight molecules (typically 3000 *m/z*) present in cells which are in a particular physiological or developmental status [46,112,113]. Compared to the proteome, the analysis of the metabolome is a more challenging task since it considers all the metabolites, regardless of their chemical nature, i.e. amino acids, antibodies, aptamers, small biomolecules, etc. [114]. In general, metabolites can be distinguished into endogenous metabolites and xenobiotics. While endogenous metabolites are directly produced by the organism, xenobiotics are chemical compounds that are present in an organism, but not generated by the metabolism of the individual itself or derived by external metabolic components that could be metabolised partly in the organism, like pharmaceutical products and their metabolites, environmental substances such as pollutants, etc. [115]. Fontanesi [110] divides endogenous metabolites into primary metabolites like sugar phosphates, amino acids,

nucleotides, and organic acids, and secondary metabolites (derived by primary metabolites, like small hormones, lipids, phytochemicals, etc.). Based on these classifications, it is possible to determine if a metabolite is a metabolism-originated organic compound and thereby does directly result from gene expression [115].

Metabolite profiling approaches are commonly practiced in two fundamental ways. Targeted metabolite profiling with the help of internal standards and mass spectrometry aims at quantifying previously selected metabolites. In contrast, untargeted metabolite profiling largely focuses on discovery to detect and quantify any small molecule (< 1500 Da) or the whole metabolome to understand their biochemical functions, changes and interactions within a biological system [116]. Untargeted profiling approaches, or so-called metabolomics, have the aim to identify, quantify and characterise the whole metabolome in a biological sample, preferably by one single experimental process [110]. Against this background, Fontanesi [110] summarised that metabolomics is a multidisciplinary approach that combines analytical and technological tools, data mining procedures and biostatistics, biochemistry, and bioinformatics for the interpretation of the omics data.

1.2.4.1 Metabolite bio indicators in meat quality

A very active research area in metabolomics is the discovery of metabolites that are indicative of disease or special phenotypes. In such approaches, metabolite profiling is used to generate quantitative lists of metabolites from control populations and to test subjects that show a desired or unwanted phenotype. Data analysis is then used to annotate the metabolites and to determine which metabolites are discriminatory for the phenotype [46]. With the exception of extreme examples in all higher organisms, Fontanesi [110] postulated that most metabolites have high variation among individuals that is affected by their genetic background and the given and environmental conditions. This variation constitutes a big challenge in metabolite biomarker discovery. Patti et al. [117] statet, that “even if many developments occurred in defining metabolic differences and identifying key biochemical mechanisms affecting important phenotypes, a complete characterisation of all metabolites produced in complex biological systems, i.e. in a complex organism like an animal or a plant, in many different conditions, is not yet possible”. Nevertheless, costs of metabolite profiling are declining while the precision of the analytical tools is constantly increasing. Metabolite profiles may be used in practice to obtain targeted metabolic information for identified biomarkers or to predict phenotypes of economic interest [110].

Generally, metabolomic-based animal selection or phenotype prediction is of practical interest in livestock production and breeding, especially when the traits cannot be recorded directly on selection candidates because the measurement requires the slaughtering of the individuals (e.g. meat quality) or if phenotyping is expensive (e.g. feed efficiency) [118]. One advantage of metabolomic-based selection it that samples (e.g. blood or urine) for metabolite profiling can be collected during an animal´s life, depending on the target phenotype (e.g., linked to growth during breeding period or meat quality near slaughter time). Moreover, Rohart et al. [118] stated that phenotypic measurement performed on the animals itself rather than on its relatives, like it is the current process, would provide more accurate predictions of the genetic constitution. Furthermore, sample collection for metabolomic profiling is usually non-invasive and metabolite profiling is cheaper than the maintenance of a conventional breeding program with performance testing of a huge number of animals [118].

As pioneers in metabolomics in meat science, Bertram et al. [119] used nuclear magnetic resonance (NMR)-based metabolomics to investigate the effects of pre-slaughter exercise stress on the plasma metabolite profile, at time of slaughter, with effect on WHC. Based on 40 pigs and three pre-slaughter stress levels, the study revealed a clear increase in plasma lactate due to exercise. Lactate was found to be the metabolite of importance for the association between metabolome and pH, temperature and WHC in muscle p.m. [119]. The aim of Muroya et al. [120] was to determine key metabolites and pathways associated with pork quality in fast- and slow-type muscles at different aging times. They assumed that the decrease of p.m. muscle glycolysis and pH value is accompanied by several changes in metabolite expression and protein degradation and denaturation [120]. These changes affect meat quality traits color and WHC. As a consequence, Muroya et al. [120] postulated that global understanding of variation in p.m. muscle metabolites provides a lot of information regarding how to monitor and control key players for the development of meat quality.

Metabolomics profiling was performed with capillary electrophoresis-time of flight mass spectrometry and revealed that hydrophilic amino acids and β-alanine-related compounds are associated with the muscle type while glycolytic and ATP degradation products contributed to aging time. Beneath others, crucial factors for meat quality development in the two porcine muscle types were the rate of inosine-5’-monophosphate accumulation by adenylate kinase 7 and 5’-nucleotidase and the rate of hypoxanthine accumulation [120].

Besides Muroya et al. [120], also Choe and Kim [121] investigated potential associations between muscle fibre types and meat quality traits. To this end, they profiled the metabolites glucose, lactate, cortisol and their transformation products in 111 Yorkshire pigs. Their results implied that the glucose level, recorded at the time point of exsanguination, might be an indicator for the early p.m. glycolytic rate and might be useful to identify pork with undesirable quality traits [121]. In a further analysis, Choe et al. [122] compared the glucose, glycogen and lactate profiles, recorded after exsanguination, with metabolite profiles recorded 45min and 24h p.m. in the same animals. This study confirmed that the ultimate lactate content in muscle via p.m. glycolysis is the main factor influencing pork quality (color, WHC) under standard slaughter conditions [122].

Rohart et al. [118] went one step further and investigated the suitability of metabolite profiles, recorded with NMR, in prediction of commonly used production phenotypes in growing pigs. To test the prediction power, different statistical approaches were applied. Based on the fact that the blood samples for metabolite profiling were taken some time before slaughter, it was not possible to predict meat quality traits very well. By contrast, carcass composition traits like LMC and growth traits like ‘average daily feed intake’ were well predicted using metabolomic data.

Beneath the usage of metabolite biomarkers in prediction or as indicators for specific performance traits, metabolomic studies are also performed to figure breed-specific differences in the metabolite profiles to elucidate the metabolic background of performance traits depending on different breeds. Straadt et al. [123] performed a NMR-based metabolomics approach, to investigate (I) breed-specific differences in metabolite profiles and (II) to investigate associations of specific metabolites to technological and sensory properties of the meat. Amongst others, Straadt et al. [123] concluded that high carnosine contents in meat are associated with many unfavourable sensory capacities like meat flavour and taste.

Although, until today, significant efforts have been directed toward the investigation of metabolites as reliable biomarkers in the field of pork quality, currently metabolites are not yet used in selection of livestock products of animal selection. The costs for metabolome profiling are still too high for the standardized application of metabolite profiles in animal production.

1.3 Integrated omics approaches

The primary aim of all various aspects of livestock genomics is the identification of causal and regulatory gene variants and the application of predictive genetic markers or biomarkers for complex traits and diseases. GWAS provided useful insights into the genetic architecture of complex traits and elucidated a lot of promising SNPs and candidate genes. However, the result of these GWAS does not enhance the comprehension of molecular pathways that are involved in diseases and complex traits. Therefore, it is the task and challenge of post-GWAS to consider multi-omics data [124]. Such association analyses focus on a holistic investigation of all omics levels instead of only one omics layer. Two or more high-throughput omics technologies can potentially be applied to the same sample. Therefore, it is important to assess how the information content of these diverse data sets at different biological levels can be combined to exhaust the full potential for a holistic understanding of the biological pathways of a specific trait [114].

The higher information content of multi-omics approaches can be used to confirm or verify the importance of detected genomic regions and candidate genes, revealed by genetic analyses of conventional livestock production traits. In this context, the meaning of specific candidate genes can be substantiated by positional or functional overlap of SNPs, detected for conventional phenotypes and related metabolic traits like transcripts, proteins, metabolites or combined omics traits. In case of positional overlap, in a certain chromosomal region, identical significant SNPs are detected both for the target trait as well as for associated metabolic phenotypes [125]. The higher information density in this approach gives evidence that the causative mutation affecting the phenotype, in fact, is located in the regarding genomic region. Functional overlap describes the accordance of SNPs detected for target traits and metabolic phenotypes caused by similar functional effects and metabolic pathways [125]. The consistent or similar functional annotation of different genetic markers detected for target and metabolic phenotypes confirmed the importance of the corresponding metabolic background affecting the target phenotype. In consequence, it can be assumed that the detected candidate genes play an important role in regulation of metabolic processes influencing phenotypes of interest.

1.3.1 Genetical genomics – the combination of genomics and transcriptomics

The core hypothesis behind gene expression or ‘expression quantitative trait loci’ (eQTL) analysis is that polymorphic sites in the genome, such as SNPs, could have concrete effects

on gene regulation by altering the coding or promoter sequences of genes, their splicing junctions, or other regulatory elements [126]. 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, SNPs, which were suspected to affect gene expression (eSNPs), can be tested with associative statistics (Fig. 5) [126]. Based on this functional relationship, 'genetical genomics' (GG) approaches were developed.

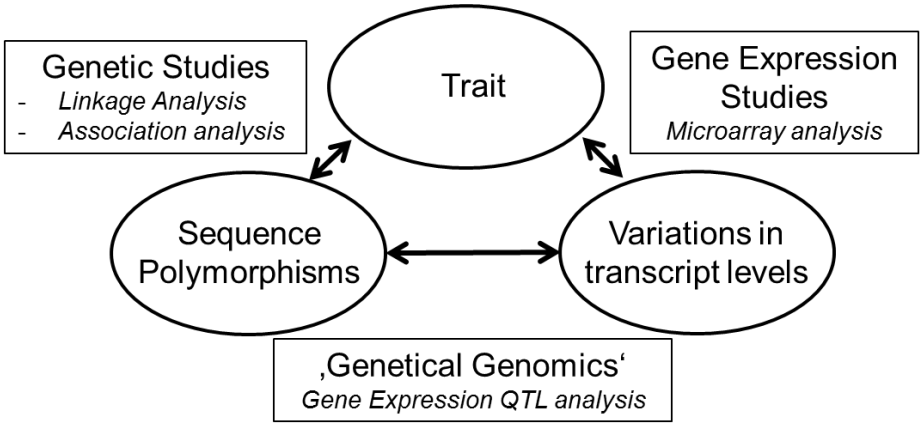


Figure 5: The genetical genomics concept brings together the traditional genetic analysis and the gene expression studies by directly characterising the genetic influence of gene expression. Modified from Li and Burmeister [127].

The concept of GG, or GWAS of expression data, constitutes a combined approach of QTL mapping and microarray expression (transcriptomics) analysis and allows the identification of regulatory networks, underlying the quantitative trait of interest and localisation of genomic variation [128]. In the underlying statistical model of genetic association analyses, expression levels of each gene are related as quantitative dependent variables to genetic marker information as explanatory variables. Individual differences in gene expression patterns reflect related genomic sequence differences (e.g. SNPs) between individuals. The resulting eQTL are mapped either within (*cis*-eQTL) or beyond (*trans*-eQTL) the corresponding gene. Expression QTL analyses have proven to be a promising tool for filling the gap of knowledge between phenotypic traits and their associated QTL and

confirmed causative mutations in pigs [129,130]. A GG approach combined with SNP mapping in the same population exposes not only *cis*- or *trans*-eQTL but also identifies SNPs for expression differences, causing changes in expression phenotype (called *cis*-SNPs or *trans*-SNPs) [49].

Trans-regulated DEG are representatives of pathways that are affected by causal variation. In contrast, genes indicated by *cis*-eQTL are more likely to represent the causative allele variant of the trait of interest [131]. Whereas most significant reported eQTL are usually *cis*-regulated, some *trans*-eQTL seem to control the expression in several genes spread across the porcine genome. In conclusion, the integration of QTL and eQTL information facilitates the identification of candidate genes and associated metabolic pathways with high evidence of their involvement in the biology of traits of interest [49,131,132]. However, the benefit of GG approaches is limited by the resolution of genetic maps that depends on the number of available markers as well as available animals, the structure of the population and the sensitivity and specificity of microarray experiments [132].

1.3.1.1 Genetical genomic approaches in pork quality traits

Until now, different responsible chromosomal regions have been identified for different performance traits in pigs. Moreover, in some cases the underlying causative molecular polymorphism has been identified. However, QTL regions prevalently are large and contain several putative causal genes. As stated above, the combination of approved QTL linkage and microarray expression analyses offers new options for the understanding of the biology, including the underlying genetic factors, affecting a complex trait. This has been proven by many experimental eQTL studies that have successfully applied GG approaches in meat quality and carcass composition traits of pigs.

In order to clarify the underlying biological processes of drip loss and to identify candidate genes for drip loss, Ponsuksili et al. [133] applied a GG approach based on transcriptome profiles recorded by the Affymetrix Porcine Genome Array in MLL samples of six divergent Du × Pi sib pairs. They identified 789 DEG between high and low drip loss animals. While up-regulated genes were associated with membrane proteins, signal transduction, cell communication, response to stimulus, and cytoskeleton, down-regulated genes regulated oxidoreductase activity, lipid metabolism and electron transport. The meaning of the ten selected 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) and zyxin (ZYX) was validated in a separate F_2 Du × Pi resource population [133]. In the same population (n=74; F_2 Du × Pi), based on a linkage analysis, Ponsuksili et al. [134] revealed 1,279 transcripts with trait correlated expression to WHC and 104 eQTL coinciding with QTL regions for WHC, with 96 *trans*-eQTL and eight *cis*-eQTL. The eight *cis*-eQTL regulating candidate genes were located on SSC2, 3, 4 and 6 [134].

In later approaches of Ponsuksili et al. [135], 150 crossbred pigs (Pi × (German LW × German LR)) were used to identify 448 *cis*-eQTL corresponding to 71 genes and 3,297 *trans*-eQTL that were related to 408 genes associated with fatness traits. In another GG study of Ponsuksili et al. [136], 207 muscle- and 150 liver samples of pigs from a commercial crossbred Pi × (German LW × German LR) pigs were analysed. In muscle, 2,001 *cis*- and 1,663 *trans*-eQTL corresponding to 593 genes were correlated with plasma cortisol level. In liver, 1,019 *cis*-eQTL and 4,873 *trans*-eQTL were found, corresponding to 116 and 927 genes, respectively. However, in muscle a higher proportion of *cis*-eQTL was observed.

Lobjois et al. [137] and Hamill et al. [138] investigated the parameter WBSF and IMF content as measurable indicators of palatability and mouthfeel such as tenderness, flavour and juiciness, respectively. Through their transcriptome analysis of MLL samples of 30 commercial F_2 pigs, Lobjois et al. [137] found a relationship between gene expression variability and tenderness of the cooked meat samples. The gene expression variability was caused by 63 DEG on almost all porcine autosomes. The identified genes were involved in the functional processes cell cycle, energy metabolism and muscle development. Some of the detected genes, located on SSC2, 6 and 13, confirmed the findings of previous studies investigating meat tenderness so that these genes seemed to be potential positional genes. Hamill et al. [138] performed transcriptomic profiling in MLL samples with lower or higher IMF (n=8) and WBSF values on day 1 p.m. (n=8). They identified 101 DEG in relation to WBSF, whereas 160 genes were associated with differences in IMF.

For the comprehensive analysis of different meat quality (pH1, meat color, curing yield and drip loss) and meat composition traits (IMF, content of several FAs (C16:0, C18:0, C18:1, and C18:2), ratio of saturated/ mono unsaturated/poly unsaturated FAs, and protein and humidity contents) and to target genes for selection, Cánovas et al. [139] performed a GG approach in MLL samples from 59 Du × LR/LW pigs. Their findings indicate that the number of DEG related to meat quality (n=506) exceeds the number of DEG for meat composition traits (n=279). The candidate genes among the up-regulated were associated with muscle development and functionality and repair mechanisms [139]. A similar complex investigation was performed by Yu et al. [140] in two different breeds, Lantang, a Chinese indigenous breed, and LR, bred for a high LMC. The meat of Lantang pigs is characterised by higher meat lightness values and IMF content and lower pH45min, pH24h and shear force in MLD in contrast to meat of LR pigs. Moreover, a FA analysis demonstrated a lower mono unsaturated FAs and higher poly unsaturated FAs percentage in MLD of Lantang compared to MLD of LR ($p<0.05$). Using a transcriptome profiling technology, Yu et al. [140] observed 586 transcripts as differentially expressed, of which 267 transcripts were highly expressed in Lantang pigs. Validation by polymerase chain reaction, revealed 13 candidate genes for muscle FA composition.

1.3.2 Proteomics and genomics

As proteins or enzymes represent the main actors in biological systems, any in-depth genotype–phenotype map requires detailed analyses of protein expression and their modifications. To fill the gap of knowledge between phenotype and corresponding genes, the objective is to consider selected proteins as intermediate phenotypes in genetic association studies [141]. Although it is possible to map QTL associated with the amount of a specific protein, so-called ‘proteinQTL’ or ‘pQTL’, currently, only a small proportion of all proteins can be detected and quantified using high-throughput proteomic approaches, like mass spectrometry and immunoassays [141,142]. With state-of-the-art technology, the most commonly used analytical tools are incapable of quantifying very small molecules.

In cattle, studies have been already performed replacing eQTL approaches by ‘pQTL’ approaches that aim at locating QTL involved in the control of the abundance of proteins. This strategy e.g. was applied to a set of proteins that have been identified as relevant

biomarkers for meat quality and, in particular, for tenderness in beef [1]. Indeed, a multitude of proteins was detected and quantified simultaneously in the experimental populations of the so-called Qualvigene project. In this project, over one thousand animals of three cattle meat breeds were performance tested in meat tenderness and genotyped with the bovine 50 K SNP chip [1]. But until now, unlike in the field of GG in pork quality, there were no pQTL approaches or protein-based GWAS in pork quality.

1.3.3 Metabolomics and genomics

Based on the fact that the metabolome is positioned between the genome and the final phenotypes (see Fig. 3, p. 14), metabolites might be used as so-called intermediate phenotypes [143]. According to Fontanesi [110], intermediate phenotypes or metabotypes are investigated to get insights in the metabolic processes leading, to economically relevant traits in animal production. Furthermore, metabotypes can also be combined with other omics expression data (e.g., transcriptomics and proteomics), to complete the characterisation and the elucidation of complex phenotypes [110]. To clarify the genetic background of the biological processes influencing livestock performance traits, metabotypes are used as novel phenotypes in GWAS. In accordance with the hypothesis that single or combined metabolite profiles are very precise indicators for relevant metabolic processes and related economically important traits (see section 1.2.4.1., p. 28), the GWAS for metabotypes should result in true associations between genetic marker and phenotype. Kühn et al. [113] stated that metabolite phenotypes provide a very detailed picture of diverse genetically modulated metabolic processes compared to the conventionally measured phenotypes that are the result of a many metabolic processes [113]. Several genetic studies have demonstrated that the result of any single omics analysis, e.g. a standard GWA procedure, might not be sufficient to decode extremely complex biological mechanisms [116]. Therefore, the combination of the omics layers genome, metabolome and final phenotype may be promising to clear up the complex regulation mechanisms and genetic background of the traits of interest. In particular in complex traits with cost-intensive phenotyping and high error rate, genetic analyses of metabotypes have a good potential for the successful identification of reliable genetic markers and candidate genes [144]. Metabolic processes respond very quick and dynamic to environmental changes and strongly depend on constitution and performance condition of the individuals. Therefore, any analysis of the metabolome has to be performed carefully

to exclude non-genetic factors that might dilute the signal coming from genetic variation [113]. Despite of these challenges, until today, several studies in the field of animal breeding postulated that metabolite phenotypes can increase the statistical power of GWAS for end-point phenotypes [145].

In livestock production, metabolomics in is well-established for many scopes of application that are not related to genetics, e.g. for the detection of drug abuse and product origin of food and to control the embryo and oocyte quality in productive processes [146]. Kühn et al. [113] described that in the last few years, there is an increasing interest in integrating metabolomics with genomics/genetics due to its capability to provide metabolic phenotypes for a detailed characterisation of the physiological state of individuals. As pioneers in integrating genomics and metabolomics analyses in livestock production, Widmann et al. [147] combined two concepts, systems biology and metabolic analyses, to detect candidate genes and functional pathways that control the complex trait 'growth at the beginning of puberty' in male cattle. Their study showed, for the first time in cattle, that the combination of genetic, metabolic and phenotypic data in a systems biological approach, using networks concepts, contributed to an advanced understanding of metabolic and gene interactions and gene-trait relationships [113,147]. Fontanesi et al. [148] presented the first genomics-metabolomics approach in pigs. Based on 900 Italian LW pigs, the authors performed GWAS for performance and carcass traits and 186 plasma metabolites (metabotypes). For most metabotypes, significant SNPs were located close to or within genes directly involved in the regarding metabolic pathways of the related metabolites. A few of these markers were also associated with some production and carcass traits [110]. These markers have good prospects to indicate promising candidate genes for economically relevant traits. With a similar study design, Zhang et al. [145] selected 33 FAs as metabotypes and used them as biomarkers of meat quality and tasting flavour in five divergent pig lines. The GWAS of the FAs did not only confirm seven already reported QTL for sensory meat characteristics, but also revealed some novel population-specific loci. Due to Zhang et al. [145], the results "can be immediately transited into breeding practice for beneficial FA composition".

1.3.4 Whole-omics data: Integration in systems genomics

As a subfield of systems biology, the 'systems genetics' or 'systems genomics' concept in an animal genetics context was established by Kadarmideen et al. [49] and aims at the integrated analysis of all different omics levels. A typical omics integration procedure goes

from genome to transcriptome, to proteome, to metabolome, to phenotype [114] (cf. Fig 6). Due to Suravajhala et al. [114], “systems genetics approaches are used to identify causal and regulatory gene networks and predictive markers for complex traits”. Furthermore, it is the aim to close the gap of knowledge and to light up the complex regulatory relationships that exist between genotypes and phenotypes [149]. The benefit of systems genetics approaches is that they allow the investigation of molecular interactions in a context that considers the dynamic in a biological system in the best way [141]. In contrast to a single-omics analysis, a holistic omics approach takes natural perturbation on all omics levels into account. Moreover, it should also be stated that such systems genetic approaches require much smaller sample size (around 40) to achieve a statistical power of 80% in detecting key genes and biomarkers [141].

QTL mapping studies of intermediate phenotypes, e.g. based on combined QTL, eQTL, mQTL and pQTL profiles, generate high dimensional data that need to be classified, annotated, stored and analysed to comprehend their function in the variation of the intermediate phenotypes and the correlated traits of interest [49,53]. To eliminate sources of errors, due to Morrison et al. [150], comprehensive quality control of the omics data based high-throughput technologies is mandatory to avoid redundant and false-positive data.

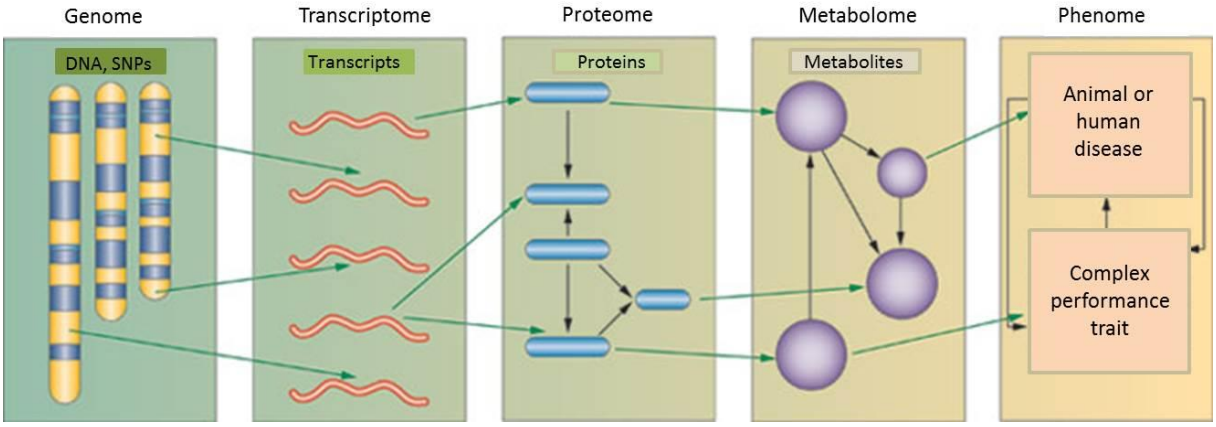


Figure 6: Illustration of integrative systems genetics approaches that integrate genomic data and other omics data types with diseases and phenotypic traits. Modified from MacLellan et al. [151].

According to Suravajhala et al. [114], typical results from multi-omics analyses include eQTL maps, eSNPs, mQTL or mSNPs, directed gene regulatory networks (using SNP data), protein-protein-interaction networks build by using multi-omics data as evidence, co-expression networks of genes that are mapped as *cis*- and *trans*-acting eQTL or of those that are differentially expressed. The most intuitive way to integrate and analyse multi-omics data are network studies and pathway-based enrichment analyses. Until today, many software solutions and web-based applications, like the InCroMap software of Eichner et al. [152] and the 3Omics program of Kuo et al. [153], were developed for the integration, analysis and visualization of systems genetics data. The analysis of relationships or networks provides deeper insights into interactions within and across cells and tissues as well as information on 'hub' regulatory genes/transcripts/proteins/metabolites that can be used as bio indicators [114]. As it is discussed in this thesis, many scientific groups take the view that multi-omics approaches have the potential to revolutionise livestock breeding practice, by divert away from a basically 'black box' approach toward an approach that considers holistic regulatory networks and pathways underlying phenotypes of interest [53,114]. However, it can be assumed that omics technologies will not replace the other approaches of quantitative genetics in livestock breeding [49,53]. Instead, omics approaches will expand and complete the range of already existing genetic analyses [53].

Applications of multi-omics studies involve genomic prediction and selection, using functional, regulatory and causal variants and the development of very accurate assays for performance trait improvement or disease prevention/diagnosis [114].

In the field of human genetics, due to Civelek and Lusis [141], systems genetic studies that addressed classic questions about the underlying molecular genetic architecture of complex traits, like 'How does information flow from DNA to phenotype?' and 'What is the nature of gene-by-environment (G × E) interactions?', received much attention during the last decade. Moreover, multi-omics approaches are used in prediction of disease risk and progression. As an example, Vazquez et al. [154] successfully used whole-genome multi-omics profiles to increase the proportion of explained variance and the prediction accuracy of breast cancer patients. In contrast, systems biology and systems genetics in livestock are still evolving fields and, to my knowledge, there are no true systems genetics approaches, combining all omics levels, applied in livestock, until now. However, in the field of meat science there are various efforts to combine different omics levels with the aim to gain deeper insights into the complex interplay of genetic constitution, gene expression, metabolic processes and regulation and environmental stimuli involved in the development of meat quality [87]. For example, D'Alessandro et al. [99] combined the omics levels proteomics and metabolomics in MLL, to analyse growth performance, fat accumulation and meat quality in Casertana and LW pigs. A similar approach was performed in Chianina beef cattle, to investigate the metabolic background of beef tenderness, based on an integrated proteomics, interactomics (investigates the molecular interactions within cells) and metabolomics analysis in MLD [155]. By the integrative transcriptomics and proteomics analysis in MLD of Canadian double-muscled LW pigs, Liu et al. [156] provided new insight into genetic mechanisms of pig double muscle traits and identified major genes and proteins involved in muscle hyperplasia and hypertrophy. Despite of the efforts in integrating various omics levels (e.g. proteomics+metabolomics, transcriptomics+proteomics), until now, there is a lack of real systems biology or systems genetics studies in meat quality in pigs.

1.4 Scope of the study

Omics analyses are part of an emerging scientific field in livestock production and quantitative genetics. Since the most genes causing hereditary defects (e.g. RYR1 gene affecting PSE meat in pigs) and the most genes with large effects causative for performance traits (e.g. DGAT1 gene for milk fat content in cattle) were detected and are successfully established in breeding programs, the focus in animal breeding increasingly is on getting a deeper understanding of the metabolic processes underlying the performance traits, investigating genetic × environment interactions, identifying genetic markers for low heritable traits also with small effects. For all of these objectives, omics analyses constitute a promising alternative to standard genetic association studies.

The basis for this analysis was the Bonner Du × Pi population, which is well established for investigating quantitative traits such as carcass or meat quality. After the QTL analyses of Liu et al. [27] in 585 F_2 Du × Pi animals that revealed QTL for growth and meat quality traits and carcass composition, gene-, protein- and metabolite expression profiles were recorded. Based on a GG approach, Ponsuksili et al. [134] and Heidt et al. [157] identified eight *cis*-regulated candidate genes significantly associated with WHC, respectively. By analysing meat quality parameter WHC also from the viewpoint of proteomics, previous examinations revealed nine proteins up-regulated and 14 proteins down-regulated in low drip animals compared to pigs with high drip loss.

The purpose of this thesis was an omics analysis, investigating and integrating the levels of the genome, transcriptome, proteome, metabolome and phenotype to elucidate the genetic and metabolic background of meat quality and carcass composition traits, paying special attention to the low heritable, complex meat quality trait drip loss. Fig. 7 gives an overview of all performed analysis steps. In the first part (Chapter 2), based on different statistical procedures, the metabolite profiles were analysed to elucidate the underlying biochemical processes and to identify potential metabolite bio indicators for drip loss, pH1, pH24 and meat color. To reveal promising candidate genes for WHC, we selected several metabolites and proteins as more accurate phenotypes and analysed them as intermediate phenotypes in GWAS (Chapter 3). In the final step, we integrated transcripts, proteins and metabolites in a network analysis to get comprehensive insights in the metabolic processes related to meat quality and carcass composition and used single networks as phenotypes in GWAS (Chapter 4).

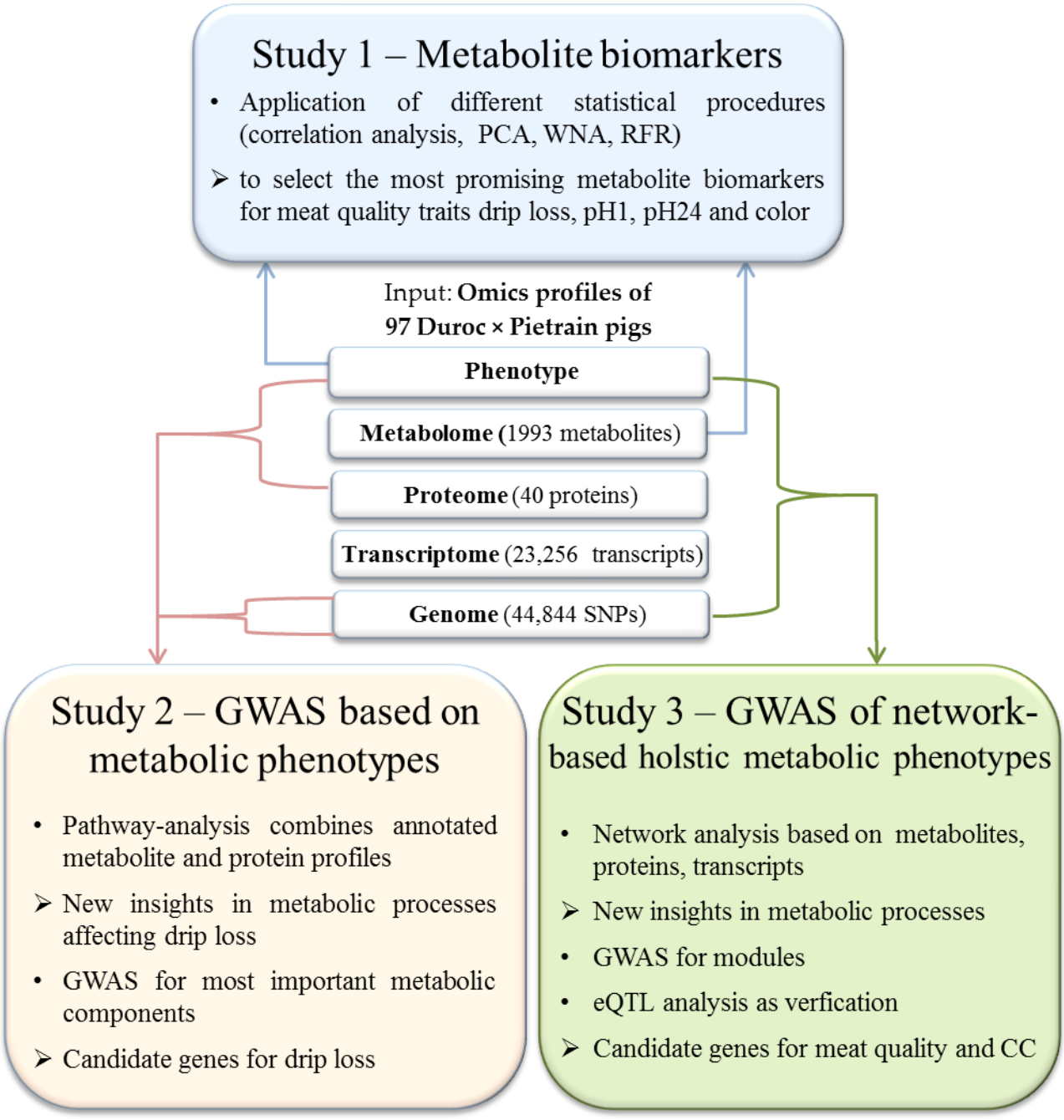


Figure 7: Work flow of the thesis. PCA – Principal component analysis; WNA – Weighted network analysis; RFR – Random forest regression, GWAS – Genome-wide association analysis; eQTL – expression quantitative trait loci; CC – carcass composition.

Chapter 2. Different Statistical Approaches to Investigate Porcine Muscle Metabolome Profiles to Highlight New Biomarkers for Pork Quality Assessment

Short title: Metabolite-Profiles as biomarker for Pork Quality Using Different Statistical Approaches

Julia Welzenbach[1], Christiane Neuhoff[1], Christian Looft[1], Karl Schellander[1], Ernst Tholen[1] Christine Große-Brinkhaus[1]*

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

*Corresponding author: Institute of Animal Science, University of Bonn, Endenicher Allee 15, 53115 Bonn, Germany, cgro@itw.uni-bonn.de, +49228733386

Email:jwel@itw.uni-bonn.de (J.W.), cneu@itw.uni-bonn.de (C.N.), cloo@itw.uni-bonn.de (C.L.), ksch@itw.uni-bonn.de (K.S.), etho@itw.uni-bonn.de (E.T), cgro@itw.uni-bonn.de (C.G.)

Published in PLOSONE; DOI:10.1371/journal.pone.0149758; February 26, 2016

Abstract

The aim of this study was to elucidate the underlying biochemical processes to identify potential key molecules of meat quality traits drip loss, pH1, pH24 and meat color. An untargeted metabolomics approach detected the profiles of 393 annotated and 1,600 unknown metabolites in 97 Duroc × Pietrain pigs. Despite obvious differences regarding the statistical approaches, the four applied methods namely correlation analysis, principal component analysis, weighted network analysis (WNA) and random forest regression (RFR), revealed mainly concordant results. Our findings lead to the conclusion that meat quality traits pH1, pH24 and color are strongly influenced by processes of *post-mortem* energy metabolism like glycolysis and pentose phosphate pathway, whereas drip loss is significant associated with metabolites of lipid metabolism. In case of drip loss, RFR was the most suitable method to identify reliable biomarkers and to predict the phenotype based on metabolites. On the other hand, WNA provides the best parameters to investigate the metabolite interactions and to clarify the complex molecular background of meat quality traits. In summary, it was possible to attain findings on the interaction of meat quality traits and their underlying biochemical processes. The detected key metabolites might be better indicators of meat quality especially of drip loss than the measured phenotype itself and potentially might be used as bio indicators.

Keywords: meat quality, water holding capacity, pork, metabolite profiling, random forest, weighted network analysis

2.1 Introduction

Sensory and technological quality characteristics of meat products are essential for acceptance of consumers and manufacturing industries. The variability of meat quality is high and the regulation of muscle properties influencing meat quality traits is still unclear [8]. One important commercially interesting meat quality parameter is the ability of meat to retain water also known as WHC. In order to characterise WHC in pork, drip loss is measured. High drip loss leads to significant reduction of meat quality resulting in monetary losses and reduced acceptance of consumers and meat-processing companies. Regularly drip loss in MLD is around 1 to 5% [18]. Drip loss is affected significantly by the structure of the muscle and the muscle cell itself and by unfavorable slaughtering conditions. Drip loss in particular is influenced negatively by too short rest periods and stress before slaughter that is associated with the rate and extent of muscular pH decline [11]. Furthermore, meat quality attributes are controlled by genetic effects as well, although the heritability for some traits is low. Genetic studies revealed several quantitative trait loci and candidate genes. However, the underlying mechanisms leading to the variation in all meat quality traits need to be better understood [27,134,158].

Some studies suggest that the levels of metabolites are helpful in order to understand the complex biological mechanisms of the underlying meat quality traits [159]. In this regard, metabolomics is a useful technique to identify candidate biomarkers that influence and indicate complex traits [81], improve preventive health care and enable early recognition of diseases [160]. In animal breeding biomarkers might be used for prediction of economical attractive phenotypes. For example Te Pas et al. [161] and Rohart et al. [118] investigated the suitability of metabolite profiles in prediction of meat quality traits in pigs. Furthermore, investigating metabolites as new phenotypes might allow uncovering the biochemical processes leading to aberrant meat quality. In general, metabolites are closer to the target phenotype compared to the level of the transcriptome or genome. In a current study, Muroya et al. [120] used this characteristic of metabolites to reveal metabolic pathways in different porcine muscle types.

In order to identify reliable metabolite biomarkers and metabolic pathways, eligible approaches of metabolite quantification and annotation are needed. A promising procedure is the untargeted metabolite profiling using mass spectrometry and subsequent data base query. In this situation, caused by the possibility of quantitative high-throughput analysis of

biological samples, the number of measured metabolites is usually much larger than the number of available biological samples. This case is also known as the 'large p, small n' problem or rather overfitting [162]. Several methods have been described that are able to handle data sets with a large number of variables [163, 164].

Therefore, the main objective of this study was to analyse the relationships between muscle metabolite profiles and meat quality traits through an untargeted metabolomic approach in order to predict their potential as biomarker and to investigate the underlying molecular structures and processes of meat quality. In regard to the 'large p, small n' problem, four different statistical methods, namely correlation analysis, PCA, RFR and WNA, were applied. Whereas correlation analysis and PCA are appropriate and commonly used methods to investigate the relationship between different variables, RFR and WNA hold several advantages in the analysis of highly multivariate, complex data. The construction of biological networks based on metabolites allows the identification of molecular interactions because they do not only quantify the correlations between pairs of metabolites, but also the extent to which these molecules are connected with other expressed metabolites.

2.2 Material and methods

Animals, tissue collection, phenotyping

This study is based on 97 performance-tested F_2 animals of a reciprocal crossbreed Du × Pi. The animals were selected within F_2 family and based on their extreme high or low values of drip loss. The animals were kept and performance tested under standardized conditions at the Frankenforst experimental farm of the University of Bonn from 2002 until 2007. Data recording and sample collection were conducted strictly in line with the German law on animal welfare. The entire experiment, including applied standard operating procedures, was approved by the veterinary and food inspection, Siegburg, Germany (No. 39600305-547/15). All animals were slaughtered at an average of 180.5 days and average carcass weight of 86.5kg. The phenotypes were recorded in a commercial slaughterhouse according to the rules of German performance stations [ZDS, 29]. Further information can be found in Liu et al. [27].

In brief, sample collection was performed thoroughly after exsanguination. About 10 min p.m. tissue samples were rapidly dissected, snap-frozen in liquid nitrogen and stored at –80°C. For further examination we choose the meat quality traits drip loss, meat color, pH1

and pH24 in MLD. Drip loss was measured using the bag method of Honikel and Kim [165]. The samples from MLD between 13th/14th rib were collected 24h p.m., weighed, and wrapped in a plastic bag. After storage for 48h at 4°C, the samples were reweighed and drip loss were calculated as a percentage of weight loss based on the initial weight of a sample. Muscle color was measured at 24h p.m. by Opto-Star (Matthaeus, Klausa, Germany). Opto-Star measures the light reflection of the meat and gives it as meat color value. High light reflectance factor stands for pale meat; low reflectance describes dark red meat color. The traits pH1 and pH24 were measured 1 and 24h p.m. in MLD. To describe the relationship between meat quality traits we performed a phenotypic correlation analysis.

Metabolite profiling

The samples metabolite spectra in MLD were measured by Metabolomic Discoveries GmbH (Potsdam, Germany; www.metabolomicdiscoveries.com) via gas chromatography - mass spectrometry (GC-MS) and liquid chromatography - quadrupole time of flight - mass spectrometry (LC-QTOF/MS).

For metabolite extraction frozen muscle tissue was mechanically disrupted in a ball mill in liquid nitrogen. 40mg of homogenate was mixed with 500µl 80% (v/v) methanol and incubated for 15 min in a thermo shaker (1000 rpm) at 70°C. Cellular debris was removed by centrifugation. 10µl of the extract were dried and subsequently used for the analysis on GC-MS. For LC-MS 1µl was injected. Derivatisation and analyses of metabolites by a GC-MS 7890. A mass spectrometer (Agilent, Santa Clara, USA) were carried out as described [166]. The LC separation was performed using hydrophilic interaction chromatography with a ZIC-HILIC 3.5µm, 200 A column (Merck Sequant, Umeå Sweden), operated by an Agilent 1290 UPLC system (Agilent, Santa Clara, USA). The LC mobile phase was a linear gradient from 90% to 70% acetonitrile over 15 min, followed by linear gradient from 70% to 10% acetonitrile over 1 min, 3 min washed with 10% acetonitrile and 3 min reequilibration with 90% acetonitrile. The flow rate was 400 µl/min. Hyphenated mass spectrometry was performed using a 6540 QTOF/MS Detector (Agilent, Santa Clara, USA). The measured metabolite concentration was normalized to the internal standard.

GC-MS and LC-QTOF/MS are used for untargeted metabolite profiling and facilitate the identification and robust quantification (accurate molar mass) of a few hundred metabolites in a single tissue sample. Chromatography followed by mass spectrometry has a relatively

broad coverage of compound classes, including organic and amino acids, sugars, sugar alcohols, phosphorylated intermediates and lipophilic compounds. With the combination of both methods it is possible to detect metabolites in a range of 50-1700 Dalton, with a precision of 1-2ppm and a solution of mass/Δmass=40.000 (Report METABOLOMIC DISCOVERIES GmbH). For details on the methods see Lisec et al. [166]. Metabolites were identified and annotated in comparison to Metabolomic Discoveries' databases, which resort to Human Metabolome Database (HMDB, www.hmdb.ca), METLIN (www.metlin.scripps.edu/) Lipid Maps (www.lipidmaps.org/). Annotation of metabolites was based on mass assignment, retention behavior and structure information. Non-annotated metabolites are characterised by their accurate mass and retention time.

2.3 Statistical analysis

Processing/correction of phenotype and metabolite data

Individual phenotypes of meat quality traits and metabolite expression levels were corrected for systematic effects using a fixed, generalized linear model of R software (www.r-project.org). The linear model contained besides population average μ and random residuum e, the effect 'season' (S, 3-month classes) and SW as a linear covariable.

$$Y_{ij} = \mu + b(SW_{ij}) + S_i + e_{ij} \quad (1)$$

All further statistical analysis methods were carried out using the calculated residuals of metabolite expression intensities and meat quality characteristics.

Association between metabolite profiles and meat quality traits

To investigate associations between metabolite profiles and meat quality traits we applied four different statistical approaches: 1) Correlation analysis, 2) PCA, 3) WNA and 4) RFR. These methods have different properties in order to handle the specific statistical problems ('large p, small n', high dimensionality and distinct correlation between variables) of the metabolomics data set. All statistical methods of analysis were performed with R (http://www.r-project.org).

2.3.1 Correlation analysis

In a first step simple Pearson correlation coefficients were estimated to investigate the relationship between paired samples of metabolites and meat quality traits. Significant correlations ($p \leq 0.05$) were considered for further biological interpretations.

2.3.2 Principal component analysis

The PCA is an unsupervised method which condenses the large number of metabolites into a set of representative, uncorrelated principle components (PCs) by means of their variance covariance structure [167]. Only PCs which explain more than 1.5% of the entire metabolite expression variance were considered for further analysis. The relevance of each metabolite within each PC was quantified by their corresponding loadings.

2.3.3 Weighted network analysis

Similar to the PCA, the WNA procedure [164] tries to reduce the dimensionality of the metabolic information. Simple network statistics were used to generate a limited number of biological interpretable modules. Pearson correlation matrix (adjacency matrix) of all bivariate metabolite comparisons is used to calculate the distances between the metabolites, corresponding to the differential metabolite expression. By raising the absolute value of the Pearson correlation to a power $\beta \geq 1$ (soft thresholding), the weighted network construction emphasizes large correlations at the expense of low correlations [168]. The distances between metabolites are integrated into a topological overlap matrix (TOM) which is used to cluster the variable expression profiles hierarchically. The results are visualized by a dendrogram with hierarchically arranged branches and connected nodes. The branch position of each metabolite indicates the actual connectivity (topology) in the network. The metabolite located at the end of the branch is the most connected nodes (‘hubs’), which play an important role in influencing the co-regulation patterns of other nodes in the network. Moreover these hubs may act as linking nodes for communication and interaction between different networks [169].

For further evaluation the branches were clustered into separate co-regulated modules, which are visualized by different colors in the corresponding dendrogram. The mathematical delimitation of each module was obtained through semi-automated, adaptive pruning of the hierarchical clustering dendrogram. Based on the distance matrix of all

metabolites (dissimilarity of TOM) and the hierarchical clustering dendrogram the function produces a vector of numerical labels giving assignment of objects to modules [164].

In a next step the metabolite expression profiles for each module are decomposed via a singular value decomposition to form module eigenvalues (MEs). This procedure is closely linked to a PCA within a module, where the MEs resemble the first PC. The importance of each metabolite for its module (Module membership, MM) is quantified by the correlation between MEs and metabolite expression profiles. Moreover the significance of each module specific metabolite (Metabolite significance, MS) for the response traits is expressed by the Pearson correlation coefficient. The MS values correspond to the Pearson correlation coefficients between metabolites and meat quality traits.

Another method to classify metabolites as key indicators of a metabolic network is the concept of maximum adjacency ratio (MAR). MAR is a function of connectivity that is calculated across all metabolites. Thereby, MAR describes the relativeness of the entire metabolite network. In coexpression networks, MAR is a useful parameter since it allows to determine whether a node forms moderate relationships with a lot of features ($MAR_i<0.5$) or very strong relationships with relative few features ($MAR_i>0.5$) [170]. From the viewpoint of network analysis MAR differs from MM because is not a module based parameter, but is able to indicate strong linked-up metabolites, that are involved in many metabolic pathways [171, 172].

The WNA procedure used in our study is implemented in the package 'Weighted Gene Co-expression Network Analysis' *(WGCNA)* in R [164]. As an optional feature of *WGCNA*, the user is allowed to assess the minimum number of metabolites contained in each module. To construct an interpretable number of modules, we used the standard thresholding parameter (β) and a minimum of 30 metabolites per module in our analysis [172]. The mathematical delimitation of each module was obtained through pruning of the hierarchical clustering dendrogram implemented in the function 'cutreedynamic' of *WGCNA.*

2.3.4 Random forest regression

RFR is a supervised learning tool that estimates the associations between metabolites and response variables (meat quality traits) using tree-based methods with integrated permutation tests [173]. As it has been shown by Strobl et al. [174] and Nicodemus et al. [175], the random forest algorithm is believed to successfully identify relevant predictor

metabolites even in high dimensional settings involving complex interaction structures and highly correlated variables. The bootstrapping algorithm implemented in RFR involves two layers of random sampling: response values and metabolite profiles. The RFR procedure is described in detail in Breiman [163] (cf. S1, p. 193).

Because of important pitfalls of the traditional RFR algorithm by Breiman [163], implemented in R package *randomForest* [171], in this study the RFR routine was calculated based on an alternative class of decision trees developed by Hothorn et al. [176] and Strobl et al. [173]. In this enhanced RFR procedure, tree construction and variable importance (VI) estimation is addressed through the principle of non-parametric conditional hypothesis testing (cf. S2, p. 194).

Essentially, the conditional RFR has the following advantages: the procedure uses the 'conditional inference forest' (CIF) methodology as splitting criterion. At each splitting node, each predictor is globally tested for its association with the trait of interest and a p-value is computed. Hence, CIF splitting is based on an essentially unbiased splitting criterion that automatically adjusts for different marginal distributions of the predictors and thus does not share the pitfall of Breiman´s RFR. Moreover the resampling scheme in conditional RFR based on subsampling instead of bootstrap sampling and Strobl et al. [173] recommend to systematically using sampling without replacement to prevent biases in VI measurement. Finally the aggregation procedure in CIFs works by averaging the observation weights extracted from each of the trees and not by averaging predictions directly (majority voting). As a result, even in case of high correlated predictors or variables with wide scale of measurement, modifications of the standard RFR procedure lead to less biased forest construction and VI calculation.

VI calculation based on the permutation principle of 'mean decrease in accuracy' (MDA). The so-called 'MDA importance' or 'permutation importance' is directly based on the prediction accuracy rather than on the splitting criterion (see Gini importance in Breiman [163]). The MDA importance describes the difference between out-of-bag (OOB) errors after random permutation of the relevant predictor where the OOB error results from validation of the original tree. Substitution of a considerable predictor is expected to decrease the OOB error. Therewith high MDA values indicate metabolites with distinct effect on the observed trait. The MDA, that is given by a particular predictor is determined

during the OOB error calculation phase whereas the resulting VI value is conditional in the sense of beta coefficients in regression models, but represents the effect of a predictor in both main effects (metabolite-trait-association) and interactions (metabolite-metabolite-interaction) [174].

The mean MDA of each predictor based on the aggregated forest can be used to rank the predictors. In order to reduce the number of metabolites to a manageable size, a permutation test of Hothorn et al. [176] was performed. We set the threshold of the permutation test to $p \leq 0.1$ which rejects uninformative predictors and enables the selection of predictor variables with significant importance. Hereby the risk of too many wrongly believed predictive predictors is reduced [175].

The root mean square error (RMSE) of RFR is calculated as the square root of the difference between the realized (y_i) and the predicted observation ($\hat{y}_i^{\mathrm{OOB}}$) within the OOB data after permuting each predictor variable in the training data set divided by the number of trees (n).

$$RMSE_{OOB} = \sqrt{n^{-1} + \sum_{n=1}^{n} \{y_i - \hat{y}_i^{OOB}\}^2} \tag{2}$$

RMSE is calculated at each splitting step in the trees just as averaged over the whole forest. The coefficient of determination (R^2) of RFR is computed as

$$R^2{}_{OOB} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i^{OOB})^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

The enhanced RFR approach of Hothorn et al. [176] is implemented in the R package *party* and its subroutines ‘cforest’ and ‘varimp’ by Strobl et al. [173] were used in this study. All needed settings are realized by the activation of the specifications ‘controls = cforest_unbiased’ in the tree building function ‘cforest’ and ‘conditional = TRUE’ in the VI calculating function ‘varimp’. Because *party* does not provide the OOB error estimation by default, the function ‘postResample’ within R package *caret* was used to calculate RMSE and R^2 based on the conditional forest learned by ‘cforest’.

RFR calculation, in particular using function ‘varimp’ of *party*, is regarding central processing unit (CPU) time and random-access memory (RAM) capacity particular in the situation of our large (1993) amount of independent variables very time demanding. To

reduce the CPU time of RFR, through a previous selection step, we removed a portion of the apparently uninformative predictors.

Finally, RFR was applied on a preselected set of 3×400 metabolites, which were most important in the first three PCs according to their absolute loadings. After removing duplicates, 1,084 metabolites remained in the final data set. According to Strobl et al. [177] the number of decision trees ('ntree' parameter') was set to 1,084 and the number of candidate predictors at each split ('mtry' parameter) was set to 361 ('ntree'/3). The remaining parameters were set to default.

2.3.5 Prediction of response variables using aggregated metabolite profiles

Accuracies in prediction of the meat quality response variables using metabolites profiles were calculated for each applied method via multiple regression analysis. The statistical regression models comprised as independent variables either the first ten PCs of PCA, ten modules of WNA or ten metabolites with highest VI values identified by RFR.

In addition, the results of all analysis were used in a joint analysis in order to identify important biological interpretable networks of metabolites or interactions between promising metabolites and meat quality traits. In this context, the subjective selection of metabolites for the joint analysis based on following conditions: a) metabolite is ranked within the top 30 variables according to their importance indicators (absolute correlation coefficient, absolute loading of PCA, MS of WNA and VI of RFR) in at least one of the applied statistical methods, b) metabolite is annotated. These importance parameters were used to identify metabolites with high meaning for the observed traits. Based on the selected metabolites (six metabolites for each trait) correlation coefficients between metabolites and between metabolites and traits were calculated to construct a network. The software package *pajek* [178] was used to visualize the complex network of all pre-selected metabolites and meat quality traits via arrows and connection lines.

2.4 Results

2.4.1 Meat quality traits and metabolic profiling

The raw values of the performance data, given in Tab. 5, reflect the normal range of meat quality in commercial crossbred pig population. With the exception of the correlation between pH1 and pH24 measured in MLD, all correlation coefficients between meat quality parameters were significant different from zero and had the expected sign (Tab. 5).

Table 5: Descriptive statistics and phenotypic correlations between meat quality traits.

Trait	Mean	S	Min	Max	pH1	pH24	Color
drip loss, %	1.90	1.39	0.40	5.30	-0.314**	-0.350***	- 0.371***
pH1	6.53	0.22	5.89	6.94		-0.024	+0.272**
pH24	5.52	1.12	5.32	6.06			+0.638***
color	72.5	7.25	56.00	92.00			

S = standard deviation; Min = minimum; Max = maximum; *** ($p \leq 0.001$), ** ($p \leq 0.01$).

Untargeted metabolite profiling detected 1,993 different metabolites in 97 meat samples. Using Metabolomic Discoveries' databases, 393 metabolites were successfully assigned to a biological function along with a tagged name and description (first choice). In case of 128 annotated metabolites, the *Kyoto Encyclopedia of Genes and Genomes* (KEGG)-IDs were also available. Non-annotated metabolites were characterised by their available exact mass information. In a further step, we tried to annotate the most important metabolites manually by using the METLIN database (second choice). Based on the known accurate mass, neutral charge and a maximal tolerance of +/-10ppm a potential functional annotation was assumed for the unknown metabolites.

2.4.2 Correlation analysis

The correlation analysis revealed 77, 436, 155 and 235 metabolites significantly correlated with drip loss, pH1, pH24 and color respectively (Tab. 6). The correlation coefficients ranged from -0.46 to 0.44. As can be seen from the Venn-diagram in Fig. 8 most of the relationships were trait specific, whereas only 152 of the 903 significant correlated

metabolites showed a significant correlation with more than one meat quality trait. In case of trait meat color more than half of the significant metabolites are also significant correlated with other meat quality trait.

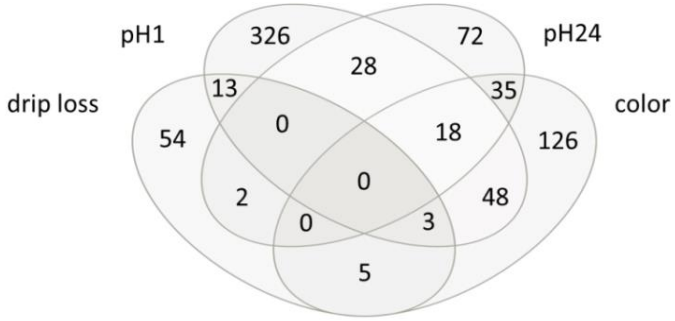


Figure 8: Venn-diagram of significant correlated metabolites.

Table 6: Results of the correlation analysis for traits and metabolites.

Trait	No. of positive correlated	Range	No. negative correlated	Range
drip loss	72	0.20 to 0.25	5	-0.21 to -0.24
pH1	212	0.20 to 0.44	224	-0.20 to -0.46
pH24	99	0.20 to 0.41	56	-0.20 to -0.32
color	162	0.20 to 0.35	73	-0.20 to -0.37

significance threshold $p \leq 0.05$.

2.4.3 Principal component analysis

PCA was used to condense expression profiles of all metabolites in a reasonable number of PCs. As shown in Fig. 9, the first three PCs already specified 46.9% of the observed variance. These proportion increases with diminishing response of additional PCs from 60% using six PCs up to 70% using more than ten PCs.

In order to identify significant metabolites, we focused on the first three PCs as it has been proposed by DiLeo et al. [172]. In these PCs the loadings of all metabolites were in a range of -0.1 to +0.1. According to the criteria to rank loadings in PCs [172], in our study the

metabolites do not reach significant eigenvalues of >0.2 or <-0.1. Furthermore beneath the possibility to rank the variables, a general biological characterisation of the first PCs is hardly possible.

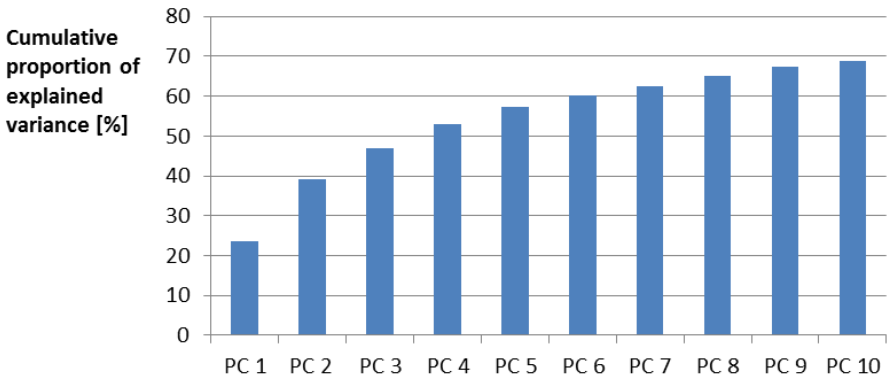


Figure 9: Cumulative proportion of explained variance by principal component one to ten. PC = principal component.

2.4.4 Weighted network analysis

WNA allowed the entire data set of 1993 probe sets of metabolites to be utilized in the construction of the weighted co-expression network. The hierarchical clustering algorithm and the following pruning process resulted in ten modules (see S3). The number of metabolites per module ranged between 776 (module ‘blue’) and 31 (module ‘salmon’). Four metabolites were not assigned to any module, and were labeled with color ‘gray’.

The relationships between meat quality traits and modules are given as correlation coefficients between traits and MEs (Fig. 10). Drip loss was significant positively correlated with modules ‘purple’ and ‘greenyellow’. Meat color and pH1 showed a significant negative correlation with the module ‘magenta’. MEs of module ‘black’ were significantly correlated to pH24 and pH1, but these coefficients were controversial in sign (Tab. 7, Fig. 10).

MAR values were calculated using metabolites of the entire data set. However, regarding the MAR calculation of a specific metabolite, it can be expected that the metabolites which

belong to the same module provide the most valuable information due to their high intramodular connectivity. The majority of the metabolites had MAR values below 0.2 and only 88 metabolites had MAR values above 0.3.

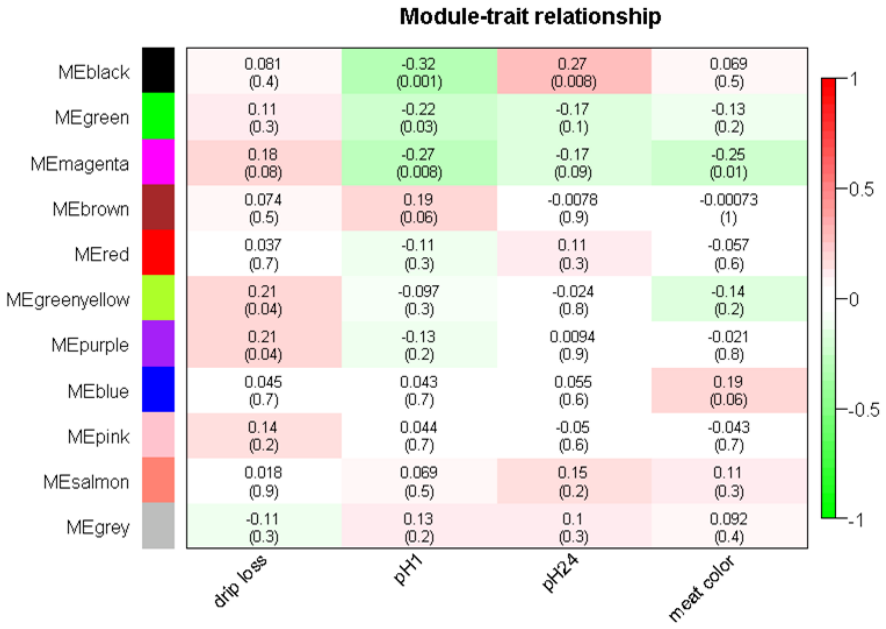


Figure 10: Correlation coefficients and corresponding p-values of module-trait relationship. Correlations of traits drip loss, pH1, pH24 and meat color to modules are characterised by color range from red (‘1’ - positive correlation) to green (‘-1’ - negative correlation); in parenthesis below correlation coefficients the p-value is given; ME = module eigenvalue.

Table 7: Selection of significant modules for meat quality traits in weighted network analysis.

Trait	Module	Cor.	p-value	Number metabolites
drip loss	‚purple‘	+0.21	$p \leq 0.04$	52
drip loss	‚green-yellow‘	+0.21	$p \leq 0.04$	49
pH1	‚magenta‘	- 0.27	$p \leq 0.008$	53
pH1	‚black‘	- 0.32	$p \leq 0.001$	73
ph24	‚black‘	+0.28	$p \leq 0.008$	73
color	‚magenta‘	- 0.25	$p \leq 0.01$	53

Cor. = Correlation.

Of particular interest were metabolites with high MM, MAR and MS. We used both the 'not module based' parameter MAR as well as the 'module based' parameters MS and MM to select metabolites that are important from different perspectives. Within the significant modules the metabolite qualifiers MM showed in many cases positive correlation coefficients to MS and MAR estimators (Tab. 8). For example, in the significant module 'magenta' (for trait color) the correlation coefficients MM:MS=0.39 and MM:MAR=0.60 indicated the high information content of the MM qualifier not only for the module specific connectivity but also for the response variable and the relativeness of the entire network (Fig. 11). Likewise in module 'black' (for trait pH1) there were significant positive correlations between MM:MS and MM:MAR. Particular in these modules it can be expected to find a reasonable number of potential key metabolites for meat quality influencing pathways [179].

Table 8: Correlation between metabolite significance, module membership and maximum adjacency ratio for modules of weighted network analysis.

	'purple' drip loss	**'greenyellow' drip loss**	**'black' pH1**	**'black' pH24**	**'magenta' pH1**	**'magenta' color**
MM : MS	0.00	0.36**	0.55**	-0.13	0.29*	0.39**
MM : MAR	-0.18	0.81**	0.29**	0.29**	0.60**	0.60**
MAR : MS	0.18	0.25	-0.22*	0.50**	-0.14	0.40

MS = metabolite significance; MM = module membership; MAR = maximum adjacency ratio; ** ($p \leq 0.01$), * ($p \leq 0.05$).

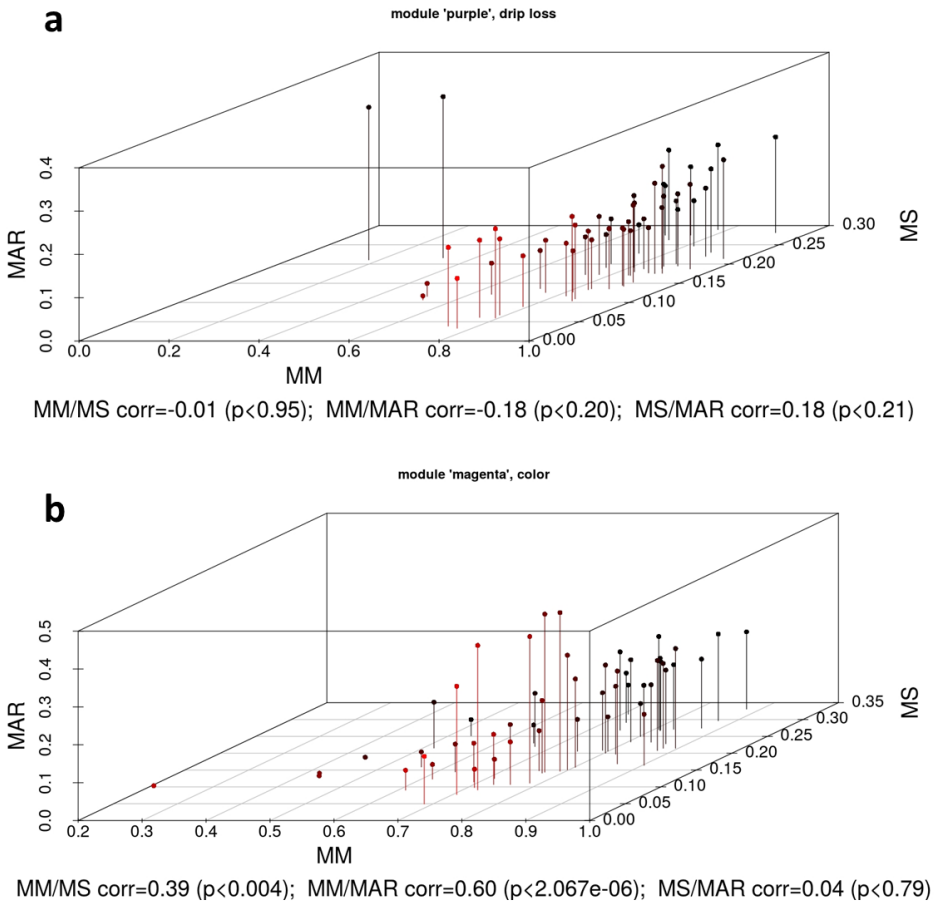


Figure 11: Scatterplot of parameters metabolite significance, module membership and maximum adjacency ratio of the modules 'purple' (a) and 'magenta' (b) that are significantly correlated with meat quality traits drip loss (a) and meat color (b). MS = metabolite significance; MM = module membershio; MAR = maximum adjacency ratio.

Regarding the relationship MS:MAR a clear tendency were observed only in the modules 'black' for trait pH24 and 'magenta' for color where the correlation coefficient exceeded a value 0.4. In all other modules this relationships were negative or close to zero. To demonstrate the relationship of MM, MS and MAR, the scatterplots in Fig. 11 visualize the relations exemplarily for modules 'purple' and 'magenta', that were significantly associated with drip loss and meat color. The plots for the remaining module-trait associations can be found in S4.

2.4.5 Random forest regression

In contrast to the previous approaches, RFR is a supervised learning method characterising the relationship between trait and metabolites using decision trees. Due to computational problems of 1993 available metabolites only the probably most important 1084 metabolites were used in RFR. These metabolites were selected based on their absolute loading values in PC1 to PC3 in PCA as described above. By this procedure the data set was reduced from 1993 to 1084 metabolites. Regarding the different meat quality traits, diverse conditional RFR accuracy parameters (RMSE, R^2 and coefficient of variation (CV)) of the prediction based on metabolite profiles were estimated. R^2 values ranged between 0.4 (pH24) and 0.55 (pH1). CV values for pH1 (2.21%) and color (6.95%) were below 10%, whereas CV values for pH24 (17.57%) and particular drip loss (51.15%) indicated a weak accuracy of RFR for these traits (Tab. 9).

Table 9: Accuracy parameters and number of metabolites with significant variable importance (VI) and maximal VI per trait according to random forest regression.

trait	RMSE	R^2	CV [%]	Max. VI	significant metabolites
drip loss	0.97	0.41	51.15	0.012	293
pH1	0.14	0.55	2.21	0.002	401
pH24	0.97	0.40	17.57	0.013	317
color	5.04	0.47	6.95	1.658	332

RMSE - root mean square error; R^2 - coefficient of determination; CV - coefficient of variation.

Despite these partly unsatisfying accuracies, a considerable number (293 to 401) of metabolites with significant impact on various meat quality traits (Tab. 9, Fig. 12) were detected. Significance of VI values was tested via a permutation test with a threshold of $p \leq 0.1$. As shown in Fig. 13, there is a large number of metabolites identified for more than one trait. For example, 14 and 110 (34+29+26+21) metabolites had a significant impact on all or at least on three meat quality traits.

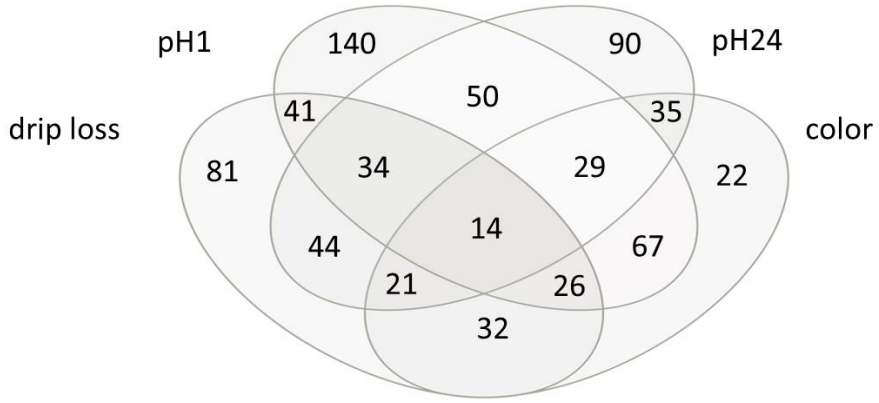


Figure 12: Venn-diagram of significant metabolites by random forest regression of Strobl et al. (2009).

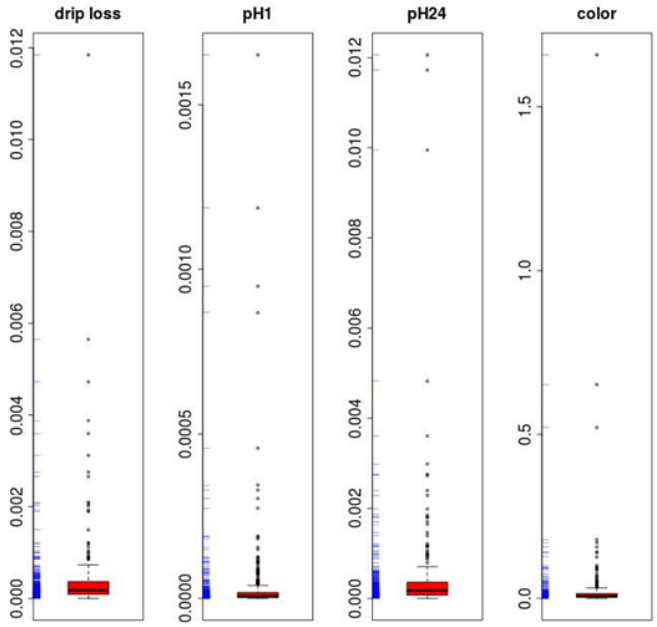


Figure 13: Variable importance boxplot of important metabolites by random forest regression of Strobl et al. (2009).

2.4.6 Evaluation of applied statistical methods in prediction of meat quality traits

In a final step, the potential of all applied statistical models to predict meat quality was quantified via a trait specific multiple regression analysis. Regarding the statistical procedures PCA, WNA and RFR the first ten PCs, all WNA modules or ten highest RFR VI values were used as independent variables. Tab. 10 shows the corresponding accuracy parameters of these analyses. In general, prediction based on metabolite profiles was very challenging in case of drip loss and worked best for pH1. Regarding the statistical methods in most analyses RFR showed the highest accuracy. Only for pH24, the first ten PCs and the modules of WNA resulted in higher R^2 compared to RFR (Tab. 10).

Table 10: Predictive power of principal component analysis, weighted network analysis and random forest regression in drip loss, pH1, pH24 and meat color based on a multiple regression model.

Trait	Multiple correlation coefficients								
	Ten principal components of PCA			Ten modules of WNA			Ten metabolites with highest variable importance of RFR		
	RMSE	R^2	CV	RMSE	R^2	CV	RMSE	R^2	CV
drip loss	1.13	0.07	59.75	1.10	0.18	57.94	1.10	0.32	58.13
pH1	0.43	0.35	6.53	0.44	0.30	6.69	0.43	0.37	6.64
pH24	0.32	0.27	5.73	0.32	0.27	5.78	0.34	0.12	6.13
color	2.54	0.23	3.50	2.56	0.21	3.53	2.51	0.37	3.46

PCA = principal component analysis; WNA = weighted network analysis; RFR = random forest regression; RMSE - root mean square error; R^2 - coefficient of determination; CV - coefficient of variation [%].

2.4.7 Joint analysis of significant associated metabolites

With the exception of PCA, the applied methods revealed significant metabolites for the observed meat quality traits. PCA resulted in weak loading values that prohibited the identification of important metabolites. In contrast, using the results of the correlation analysis and the methods WNA and RFR, it was possible to detect significant associated metabolites for meat quality traits. Comparing these methods by summarising the results presented in the Tab. 6, 7 and 8, the number of detected significant trait specific metabolites varied to a large extent. For example the number of significant metabolites for drip loss ranged from 76 (correlation analysis) to 293 (RFR). On the other hand, a considerable overlapping of significant metabolites identified by different statistical methods was detected and is presented in Fig. 14. In general, it can be assumed that metabolites, whose importance is confirmed by different methods, can be used as reliable predictors for meat quality traits.

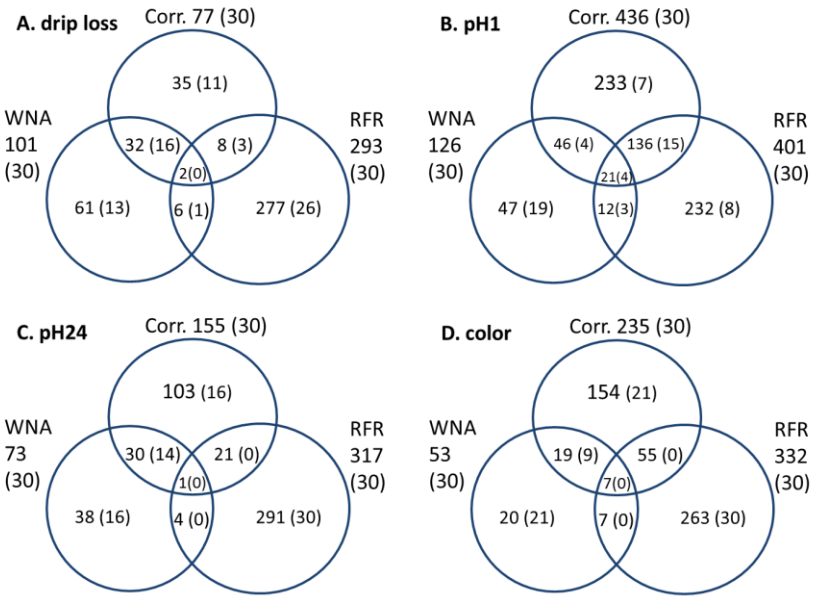


Figure 14: Venn-diagram of the entire significant metabolites for drip loss, pH1, pH24 and color identified by correlation analysis, weighted network analysis and random forest regression and of the selection of 30 metabolites with highest absolute correlation coefficients, metabolite significance and variable importance values in brackets. Corr. = correlation analysis; WNA = weighted network analysis; RFR = random forest regression.

In order to get a more comprehensive overview about the complex biological architecture of meat quality traits, the most important metabolites that were identified by the three methods were used to set up a network via virtualization tool *pajek*. Importance of metabolites was characterised by the parameters a) correlation to meat quality, b) MS in significant modules of WNA and c) VI in RFR. According to these parameters the most important 30 metabolites per method were preselected. The final joint network analysis comprised only metabolites which were annotated and identified by at least two methods. Following this rule, six metabolites were identified for pH1, drip loss and color, whereas three metabolites had an impact on pH24. For pH24, this initial subset did not contain results from the RFR analysis, so that the list was extended by three annotated metabolites which had the highest VI value (Tab. 11).

Based on the 24 selected metabolites in Tab. 11, a metabolomic network was created which comprise the meat quality traits drip loss, pH1, pH24 and color (Fig. 15). In the network the dotted lines represent connections between traits and between metabolites whereas the arrows stand for directed effects of metabolites on the observed traits. Directed and undirected connections are displayed in case of significance ($p \leq 0.05$) and absolute correlation ≥ 0.5. Fig. 15 indicates that the metabolites found by different statistical methods were highly interconnected. As a general tendency, different procedures identified similar or related chemical substances for a specific trait.

Table 11: Selection of metabolites for joint analysis based on their ranking in top 30 metabolites in correlation analysis, metabolite significance of weighted network analysis and variable importance of random forest regression.

Drip loss	**Cor**	**MS**	**VI**	**pH1**	**Cor**	**MS**	**VI**
2.3-Naphthalic acid	23.	×	10.	Histidine-alanine-tryptophan-tryptophan	5.	4.	2.
Glycero-3-phosphocholine	8.	×	7.	Cytidine	25.	8.	12
Glycero-3-phosphoserine	×	28.	23.	Allopurinol-1 ribonucleoside	×	9.	25
Glycerophosholipid	22.	14.	×	Lactic acid	24.	×	10
Triacylglycerol	19.	12.	×	Lysine-serine-isoleucine	19.	×	6.
3-Methyl-2-oxovaleric acid	21.	13.	×	Phosphocreatine	26.	×	21
pH24	**Cor**	**MS**	**VI**	**color**	**Cor**	**MS**	**VI**
α-Hydroxybutyrate	1.	1.	×	Octulose-1.8-bisphosphate	7.	1.	×
Heptadecanoyl carnitine	2.	2.	×	Fructose-6-phosphate	27.	9.	×
Stearoylcarnitine	3.	4.	×	Glucose-6-phosphate	23.	7.	×
Gle-cholesterol	×	×	2.	Inosine-5-monophosphate	28.	10.	×
Methylglyoxal	×	×	9.	Phosphoglycolic acid	11.	12.	×
Glucose	×	×	11.	Nicotinamide adenine dinucleotide	4.	×	2.

Cor = correlation analysis; MS = metabolite significance; VI = variable importance; × - Metabolite was not ranked in top 30 of the respective importance values; Gle = gallic acid-linoleic acid ester.

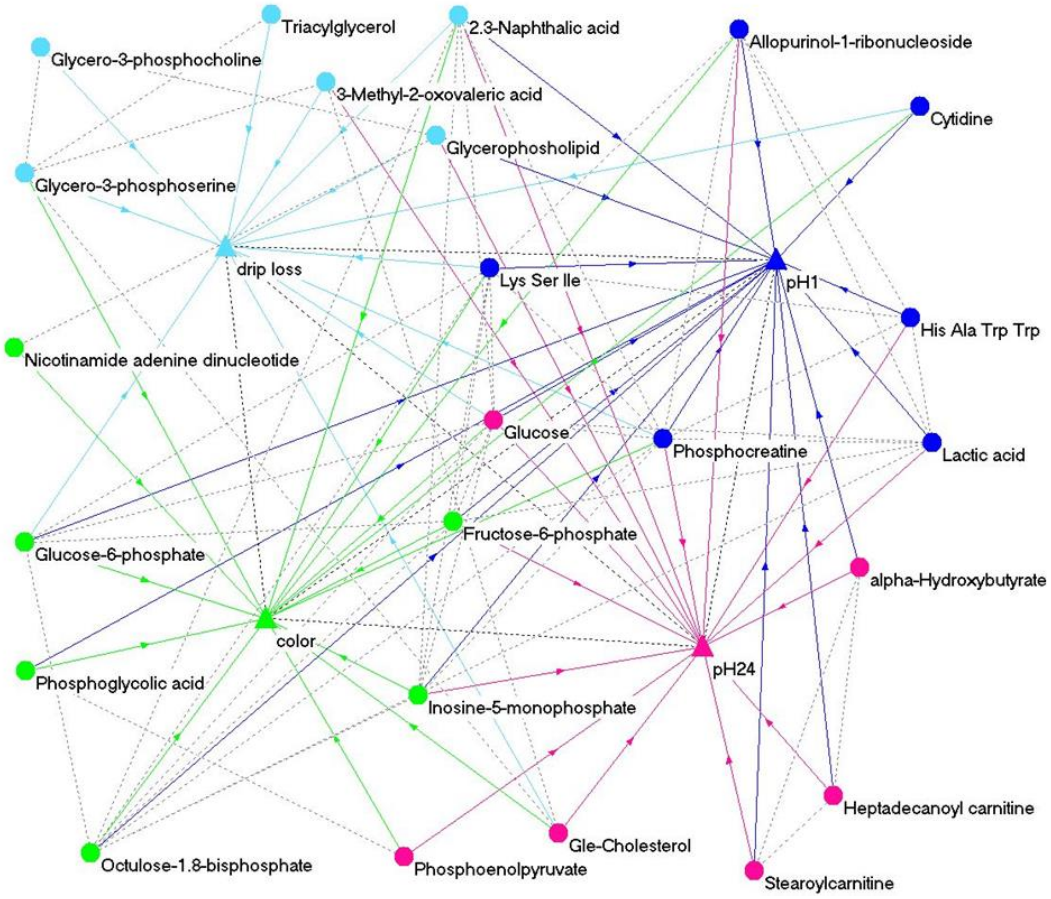


Figure 15: Metabolomic network of traits drip loss, pH1, pH24 and color and 24 strongly associated metabolites. dotted lines: undirected connections between metabolites and between traits; arrows: directed relations between metabolites and traits; triangles - traits; ellipses – metabolites; light blue – drip loss; blue – pH1; magenta – pH24; green – meat color.

2.5 Discussion

Challenges in metabolomics and functional analysis

Based on an untargeted metabolomic approach the main objective of our study was to identify key metabolites which play an important role in the complex biological architecture of meat quality traits. Moreover, these metabolites can be used as informative predictive biomarkers of meat quality of pigs. Similar objectives are pursued in a few studies reported in literature [160, 99]. Recently D'Alessandro et al. [99] successfully used metabolomics to compare highly phenotypically differentiated pig breeds. Rohart et al. [118] investigated the prediction power of metabolomic profiles for commonly used production phenotypes in pig breeds and in a current study Muroya et al. [120] tried to reveal characteristic metabolic pathways in different porcine muscle types. These studies used up to 188 well known metabolites to characterise different targets traits. In contrast to these studies, our approach tried to cover the entire metabolome of pigs expressed in meat samples.

Using a GC-MS technology, in our study 1,993 different metabolites were identified. Wishart et al. [180] showed that this method is the most efficient way for metabolite detection. However, as it has been reported by Hollywood et al. [46] metabolomics approaches by GC-MS only disclosure about 10% of the metabolome. From this follows, that the set of metabolites found in our study reflects only a small percentage of anticipated count of metabolites.

Besides this limitation, the corresponding annotation step provides only a fragmented picture because only a small amount of physiological or biochemical functions of metabolites are stored in available public data bases. In our study only 393 out of 1,993 metabolites were annotated. According to Chagoyen and Pazos [181], reasons for these fragmentary information might be the of lack of scientific fundamentals and principles of physiological and biochemical processes of higher life forms. In addition, functional analysis of high dimensional omics data is a big challenge in systems biology studies as it can be seen by the different, non-standardized statistical methods which were used here and elsewhere to analyse metabolomics data [172].

Potential and abilities of statistical methods

In order to quantify the consequences of missing statistical standards in a first step of our analysis we evaluated different statistical methods with respect to their relevant theoretical statistical properties and their consequences regarding the final results. All applied methods tried to solve the problem of the 'large p, small n' situation of the metabolomics data set used in our study.

The correlation analysis is a useful method to get a first overview. In the last decade, in many scientific fields we registered an increasing number of available variables. New techniques were proposed to address these challenging tasks involving many irrelevant and redundant variables and often comparably few training examples. Selecting the most relevant variables is a challenge for building a reliable predictor, particularly if the variables are redundant. Conversely, a subset of useful variables may exclude many redundant, but relevant, variables. Correlation methods belong to the category of ranking criteria defined for individual variables, independently of the context of others. This leads to the consequence that some variables may have a low rank because they are redundant and yet be highly relevant [182]. In our study correlation analysis nevertheless induced concordant results with RFR and WNA. The latter finding was not surprising because the module generating process in WNA, MS is calculated based on Pearson correlation coefficients as well as the simple correlation analysis. In conclusion, correlation analysis is a comprehensible procedure to get a first idea of what variables might be potential bio indicators. However, because of the described weaknesses, beneath correlation analysis, other methods based bootstrap or Bayesian procedures should be applied to validate or disprove the first results.

The PCA approach tries to condense the information content of the independent variables into a set of PC. This method is promising, in particular because bivariate correlation analysis revealed significant relationships between metabolites and meat quality traits in a range of maximal -0.20 to -0.46 and +0.20 to +0.44. However, using PCA only weak loading values were estimated within the first PC and no significant metabolites were identified based on the thresholds described by DiLeo et al. [172]. Consequently the analytic tool did not give comprehensive insight in interactions between metabolome and phenotypic traits.

WNA addresses the challenge of the ‘large p, small n’ situation by summarising a network of modules to reduce the complexity of a data set, which is thereby analysed with greater statistical power [164]. The investigation and interpretation of the ‘trait to metabolite’ associations in WNA is focused on the most highly-connected ‘hub’ metabolites with high MM, MS and MAR within the significant modules.

As described above, the parameter MS quantifies the importance of a metabolite for meat quality traits. Therefore, MS is the most eligible parameter to select promising metabolite biomarkers for a particular trait. In contrast, the parameters MM or MAR are indicators for the connectivity of metabolites and are able to indicate potential key players in the regulatory network regulating the trait of interest and between the metabolic pathways. MM quantifies the importance of a metabolite for the specific module, whereas the parameter MAR provides information about the relatedness of each metabolite within the whole network. According to Langfelder and Horvath [164] MAR values below 0.5 indicate components with many, weak connections to the network neighbors instead of few strong associations. In our study only four metabolites in the modules ‘black’ and ‘magenta’ showed MAR values above 0.4, which indicated a more important role of these metabolites regarding the network connectivity. In these modules the correlation MS:MAR also revealed a clear tendency of 0.5 (module ‘black’, pH24) and 0.4 (module ‘magenta’, color), respectively (Tab 7). In other significant modules the MS:MAR correlations were negligible weak. This finding indicates that an intensively connected metabolite does not necessarily provide important information for the expression of the response variable.

Within significant modules, MM values of almost all metabolites were highly expressed in a range between 0.5 to one (Fig. 11, S4). This result can be expected because of the underlying cluster algorithm. In contrast, the MS values within all significant modules were much lower and were almost equally distributed between zero and 0.3 for drip loss, color and pH1 (in ‘magenta’) and zero to 0.45 for pH1 (in ‘black’) and pH24 (Tab. 7). As visualized in Fig. 13, in module ‘purple’, glucosylceramide (d18:1/24:1(15Z)) and another unannotated metabolite with low MM but high MAR values do not fit in our expectations derived from a positive correlation between MM and MAR qualifiers. Theoretically, components with low MM but high MAR probably have a high connectivity across the whole network. In this context, glucosylceramide (d18:1/24:1(15Z)) is involved in many pathways for example in sphingolipid-, ceramide glucosyl- and lipopolysaccharide

metabolism so that this metabolite can be considered as a nodal point between different modules or metabolic pathways. On the other hand, in module ‘magenta’ there are some metabolites with both high MM and MAR qualifiers, but MS values close to zero (Fig. 11). These metabolites might be key players in the underlying biological pathways of module ‘magenta’, but on the other side they do not play an important role for the expression of the response variable color by its own.

According to Muroya et al. [120] module construction and MM/MS/MAR calculation is reasonable, because it can be expected that the biology of meat ageing process is regulated by a number of key factors in several key metabolic pathways. Module construction has the advantage that the function of a large amount of non-annotated metabolites can be inferred from their better-annotated neighbors within the modules. This advantage is particular important in the analysis of our data set, because only 20% of the metabolites were annotated.

In a PCA the meaning of metabolites is calculated purely by their statistical correlation (covariance) to all other metabolites. This means, that the significance of metabolites with high regulatory importance, but no directly connected to the trait of interest (weak loadings), is not detected by a PCA. In contrast, in the framework of the WNA analysis metabolites which have a central position within a regulatory network have a higher probability to be identified [172]. Varying the minimum number of metabolites within a module has an impact on the module sizes and the total number of modules identified. This option allows the user to consider biological background knowledge. Choosing a lower number of variables per module allows the user to investigate the underlying biological pathways more in detail. In conclusion, these attributes of WNA provided evidence that the procedure was an appropriate method for analyzing metabolomics data in a system biology approach.

As a final result, the WNA procedure leads to differentiable modules with similar expression profiles within the modules. From a biological point of view the intra module similarity can be interpreted as a distinct co-regulation of the module metabolites. Moreover, the indicators MM and MAR enables to identify key players in regulatory network which is possibly linked to the specific module, whereas metabolites strongly influencing specific traits are characterised by high MS [164]. Nowadays weighted co-

expression network analyses are applied in a wide scientific field in order to estimate the relationships, connectivity and dependency of different variables in biological systems. In metabolomics approaches the combined abilities of WNA to cluster and select variables are also very useful. For example, DiLeo et al. [172] and MacLennan el al. [183] successfully used WNA to select metabolite biomarkers in tomatoes and transcripts as biomarkers in mice.

In RFR, VI is usually used for selection of (a) causal variables highly related to the response variable for explanatory and interpretation purposes and (b) of a small number of relevant predictor variables. It was shown in test runs using all independent variables (results not shown), many of the metabolites had very little importance in the trees and therefore in prediction of the trait of interest. Despite the expectation that the RFR procedure is able to handle high dimensional data with redundant and unimportant variables, the analysis ran more robust and in acceptable running time based on the reduced data set with 1,084 instead of 1,993 metabolites.

To deal with the impurity's bias for selecting split variables towards uncorrelated predictor variables, VI values were calculated with an enhanced RFR procedure that guarantees unbiased tree algorithms for reliable prediction and interpretability in both individual trees and forests. In standard RFR, by Breiman [163], the VI based ranking of the predictors says nothing about the significance of the top-ranked predictors and the procedure always outputs a ranking – even if all predictors are uninformative in the prediction. In contrast, in Hothorn´s conditional RFR [176], VI computation is based on an implemented permutation test which analysed the significance of the respective metabolite. This selection step leads to a reduced number of explanatory variables in the model that avoids overfitting and ensures a smaller prediction error [118]. Generally, the VI parameter in RFR can be interpreted similar to MS values in WNA. In contrast to WNA, which determines MM of each metabolite within a module, RFR does not estimate the relative similarity among metabolites. This limitation of the RFR procedure makes it difficult to assign metabolites to different functional pathways. Moreover, RFR approaches partially produce 'odd unexpected results' in some specific cases [184]. Even in the enhanced conditional RFR procedure, the risk of biased VI values in case of specific data structure or predictor type cannot be overlooked completely. As well as the pretended advantage of RFR, the absence of a specific underlying stochastic model, is also a challenge in the sense that it is difficult

to understand how the prediction within the variety of decision trees works exactly [184]. Nevertheless, RFR has become a major analysis tool in many fields of bioinformatics due to its high flexibility and in-build VI calculation. Also in prediction of various characteristics based on metabolomics data, RFR has been used successfully [185].

Accuracy of prediction of meat quality traits by applied methods

To evaluate the prediction ability of meat quality traits by PCA, WNA and RFR a linear (multiple) regression model was used. Suitability of the methods in prediction on basis of metabolite profiles was different regarding the different traits (Tab. 10). Compared to drip loss, pH1 and color prediction performed better using the ten selected metabolites by RFR, whereas pH24 prediction based on PCs or WNA modules resulted in higher R^2. According to the studies of Rohart et al. [118], who also used RFR for phenotypic prediction based on metabolomic data, prediction accuracy depends strongly on the observed trait. In our study, prediction worked best for the trait pH1 and worst for drip loss, considering R^2. This result might correspond to the genetic foundation of these traits. It has been summarised by Ciobanu et al. [32] that the lowest heritability estimates were found for drip loss whereas pH1, pH24 and color showed higher values (h^2 drip loss = 0.31, h^2 pH1 = 0.41, h^2 pH24 = 0.39, h^2 color = 0.57). That means, drip loss is stronger influenced by environmental effects which might complicate the prediction accuracy by metabolite information.

Joint analysis

A network of metabolites and meat quality traits is represented in Fig. 14. Trait pH1 was the most cross-linked trait in our study and several metabolites like 2,3-naphthalic acid and glucose were significantly associated with all respected traits. Moreover glucose, selected due to high importance for trait meat color, was connected to eight other metabolites, amongst others to IMP, lactic acid and 2,3-naphtalic acid. Besides its influence on meat color, the metabolite IMP that is involved in purine metabolism and biosynthesis of alkaloids derived from histidine and purine, also showed significant associations to drip loss, pH1 and pH24. Taking into account the significant phenotypic correlation among the four traits as well (Tab. 8), all observations indicated that meat quality traits were highly interconnected and influenced by similar biochemical processes.

Regarding the different statistical approaches it can be summarised that the applied procedures all in all identified similar or related chemical substances as important for a specific meat quality trait. For example in regard to of drip loss, correlation analysis, WNA and RFR revealed several glycerophospholipids (GPL) and glycerolipids that are involved in lipid metabolism and arise from degradation of membrane structures. Moreover, similar to the findings of Hidalgo et al. [186], different acids, like 2,3-naphthalic acid and the α-keto acid 3-methyl-2-oxovaleric acid, that are associated with lipid oxidation were identified by different methods as important for drip loss. For example 2,3-naphthalic acid is part of the pathway 'degradation of aromatic compounds' that directly leads to generation of pyruvic acid and other compounds that are involved in energy metabolism like acetic acids. Most important metabolic processes in muscle and meat are energetic processes like glycolysis/gluconeogenesis, citrate cycle and pentose phosphate pathway (PPP), which verifiable are responsible for muscle physiology and meat quality [14, 9].

In hypoxic tissues after slaughtering anaerobe metabolic processes predominate and in glycolysis glycogen is released via glucose to pyruvic acid. Under aerobic conditions pyruvic acid is metabolized in citrate cycle and oxidative phosphorylation [9]. In case of stress before slaughtering or a too short resting period before stupefaction in hypoxic tissues, the rate of oxidative processes like glycolysis is increased and pyruvic acid do not flow into glycolysis but is transferred to lactic acid [165]. Accumulation of lactic acid goes along with pH decrease to 5.6 [165]. The coincidence of low pH1 and high temperature in muscle lead to partial denaturation of proteins and reduction of intercellular space. Thereby, lipids are dissolved from membranes, permeability of membranes is increased and drip loss is the result [11]. Based on this background, the meaning of e.g. 2,3-naphthalic acid, glucose and several GPLs, sterol lipids and FAs for meat quality characteristics drip loss, pH1, pH24 and color is traceable. These metabolites are indicators for complex metabolic processes and are characteristic of the specific occurrence of meat quality traits. Selected metabolites potentially may be used as universal bio indicators for prediction of special traits. Availability of such 'multiple applicable' biomarkers would reduce effort and cost of phenotyping in breeding programs and commercial meat processing. Regarding the associations between the metabolites, it was observed that some metabolites were significantly correlated with many other components. This finding suggested that some strongly networked metabolites are the key players of metabolic processes responsible for the large complex of meat quality traits in pigs. Intense investigation of these important

metabolites might lead to a deeper understanding of the underlying biological pathways and the causal reasons of development of quality traits.

Key metabolites and apparently significant associated pathways related to meat quality traits

The different applied methods resulted in several key metabolites mainly belonging to the family of lipids (GPLs, sterol lipids, prenol lipids). In addition to lipids, the statistical analysis also detected other compounds like the naphthalene 2,3-Naphthalic acid and the α-keto acid 3-methyl-2-oxovaleric acid with strong association to drip loss. GPLs are the major lipids in mammalian cell membranes [187]. Preslaughter stress results in increased rate and extent of pH decline, decomposition of membrane structures and cell swelling and shrinkage. In this way dissolved lipids and lipid compounds run off the cells into the extramyofibrillar compartment. This process of lipid decomposition also is accompanied by lipid oxidation that results in increasing concentration α-keto acids. Therefore the relation between drip loss and associated lipids and acids can be explained and have been already described by Lambert et al. [187] and Poulsen et al. [188].

Examination of compounds with significant association to pH1 resulted in metabolites of purine and pyrimidine metabolism (nucleotide metabolism), glycolysis and PPP. PH1 is a major indicator for PSE meat, which is characterised by low pH1. The higher the rate of glycolysis, PPP and related metabolic processes like lactic acid - and nucleotide metabolism, the lower is pH1 in meat. The onset of rigor mortis at low pH1 and high temperature causes the denaturation of around 20% of the sarcoplasmic and myofibrillar proteins [165]. This explains the significant meaning of polypeptides like histidine-alanine-tryptophan-tryptophan and lysine-serine-isoleucine.

Trait pH24 was significant associated with metabolites of pyruvate metabolism, glycolysis, PPP and purine metabolism. Moreover, pH24 was significant associated to metabolites resulting in the course of protein degradation (e.g. polypeptide glutamine-histidine-alanine) and metabolites of lipid metabolism, like GPs, sterol lipids and FA esters (e.g. stearoylcarnitine), and hydroxy acids like α-hydroxybutyrate (ketone body). The meat quality parameter pH24 is an indicator for DFD meat, which characteristically leads to a pH ultimate value >6. High ultimate pH results in relative little protein degradation, high WHC, dark meat and early spoilage of the meat. Meat spoilage follows from microbial

reduction, natural autoxidation of lipids and autolytic enzymatic processes [165]. Toldra and Flores [17] reported the significance of FAs and ketones and polypeptides (products of autolytic enzymatic spoilage) for pH24 in meat. The degradation of free FAs to ketone bodies in liver is one option to generate energy for muscle cells. With empty glycogen stores p.m. energy is mainly supplied by mobilization of lipid stores and transformation of released FAs into ketone bodies [165]. Because these anaerobic processes lead to reduced pH decline p.m. several FAs and ketone bodies (e.g. α-hydroxybutyrate) might be good indicators for pH24. Relevance of p.m. energy metabolisms like glycolysis, PPP and pyruvic acid metabolism, indicated by metabolites like glucose and phosphoenolpyruvate, for meat quality traits in pigs also has been described by Scheffler et al. [9].

Analysis of pork color resulted in different significant associated metabolites (phosphates, pyruvic acid) of glycolysis, PPP and pyruvic acid metabolism. This means high rate of glycolysis and activated PPP and pyruvate metabolism results in high meat color value (Opto value, scattering effect), because glucose is metabolized to glycogen and finally to lactic acid. This goes along with acidification and pale meat color and explains the determined significant meaning of phosphates and downstream products of glycolysis like octulose-1,8-bisphosphate and phosphoglycerate in our study. Muroya et al. [120] and D'Alessandro et al. [99] who investigated characteristic metabolic pathways of meat quality in pigs could confirm these results. They also indicated significant correlation coefficients between meat color indicators (L*, a*, b*) and higher rate of glycolysis.

2.6 Conclusion

In this study untargeted metabolite profiling of muscle samples of 97 Du × Pi pigs was used to identify underlying biochemical processes and potential key molecules affecting meat quality traits. Because of limited technical capabilities of GC-MS and a lack of basic knowledge about biochemical processes of higher life forms only detection and annotation of a small percentage of metabolites influencing meat quality was possible. To get deeper insights in the involved biological pathways we applied and evaluated different statistical methods, namely correlation analysis, PCA, WNA and RFR. Although the methods based on different statistical approaches and in spite of differences between the parameters and requirements of the particular methods to achieve statistical significance, they revealed similar results. Using the described methods for analysis of the holistic metabolite profiling we were able to detect both metabolites with already known meaning for meat quality as

well as metabolites whose influence on meat quality traits not yet has been described. As expected, the applied methods revealed metabolites as important, that are involved in p.m. glycogen degradation and energy consumption under the exclusion of oxygen like glucose, GPLs and different phosphates. On the other hand the meaning of several metabolites like e.g. the polypeptides histidine-alanine-tryptophan-tryptophan and lysine-serine-isoleucine for trait pH1 has not yet been described in literature.

The consistent results lead to the conclusion that meat quality traits pH1, pH24 and color are strongly influenced by processes of p.m. energy metabolism like glycolysis, PPP, pyruvic acid metabolism and associated processes. Drip loss in particular is significant associated with different glycerophospho-, sterol- and prenol lipids and compounds involved in lipid metabolism which are products of membrane degradation. In summary, it was possible to attain findings on the interaction of meat quality traits and their underlying biochemical processes. The detected key molecules will be used in further investigations in order to clarify the complex molecular structures underlying drip loss. Furthermore, these selected metabolites might be better indicators of meat quality especially of drip loss than the measured phenotype itself and potentially might be used as bio indicators. For this purpose the validation of the candidate bio indicators in another set of pigs is desirable.

Chapter 3. Integrative Analysis of Metabolomic, Proteomic and Genomic Data to Reveal Functional Pathways and Candidate Genes for Drip Loss in Pigs

Julia Welzenbach[1], Christiane Neuhoff[1], Hanna Heidt[1,2], Mehmet Ulas Cinar[1,3], Christian Looft[1], Karl Schellander[1], Ernst Tholen[1] and Christine Große-Brinkhaus[1,*]

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

jwel@itw.uni-bonn.de (J.W.); cneu@itw.uni-bonn.de (C.N.); heidt@ibla.lu (H.H.); mucinar@erciyes.edu.tr (M.U.C.); cloo@itw.uni-bonn.de (C.L.); ksch@itw.uni-bonn.de (K.S.); etho@itw.uni-bonn.de (E.T.)

[2]Institute for Organic Agriculture Luxembourg, Association sans but lucratif (A.S.B.L)., 13 Rue Gabriel Lippmann, L-5365 Munsbach, Luxembourg

[3]Department of Animal Science, Faculty of Agriculture, Erciyes University, Talas Bulvari No. 99, 38039 Kayseri, Turkey

*****Correspondence: cgro@itw.uni-bonn.de; Tel.: +49-228-733-386

Published in International Journal of *Molecular Science*; DOI: 0.3390/ijms17091426; August 30, 2016

Abstract

The aim of this study was to integrate multi omics data to characterise underlying functional pathways and candidate genes for drip loss in pigs. The consideration of different omics levels allows elucidating the black box of phenotype expression. Metabolite and protein profiling was applied in *Musculus longissimus dorsi* samples of 97 Duroc × Pietrain pigs. In total, 126 and 35 annotated metabolites and proteins were quantified, respectively. In addition, all animals were genotyped with the porcine 60 k Illumina beadchip. An enrichment analysis resulted in ten pathways, amongst others, sphingolipid metabolism and glycolysis/gluconeogenesis, with significant influence on drip loss. Drip loss and 22 metabolic components were analysed as intermediate phenotypes within a genome-wide association study. We detected significantly associated genetic markers and candidate genes for drip loss and for most of the metabolic components. On chromosome 18, a region with promising candidate genes was identified based on single nucleotide polymorphisms associated with drip loss, the protein ‘phosphoglycerate mutase 2‘ and the metabolite glycine. We hypothesize that association studies based on intermediate phenotypes are able to provide comprehensive insights in the genetic variation of genes directly involved in the metabolism of performance traits. In this way, the analyses contribute to identify reliable candidate genes.

Keywords: drip loss; pork quality; metabolomics; proteomics; enrichment analysis; genome-wide association study (GWAS); candidate genes

3.1 Introduction

Pork quality is the result of complex interactions between genetic and environmental effects like rearing and slaughtering conditions, and carcass and meat processing. One important commercially interesting pork quality parameter is the ability of meat to retain water, also known as WHC. In order to characterise WHC in pork, drip loss is measured. This fluid, mainly from muscle cells, resigns from the meat surface without any mechanical force other than gravity and is influenced by shrinkage of the myofibrils, pH-value, and temperature p.m. [10,119]. Average drip loss in MLD is around 1% to 5% [18]. Heritability estimates of WHC vary to a large extent between 0.01 and 0.31 [25]. This wide range could be explained by breed effects and large measurement errors of drip loss due to the multifactorial environmental effects [32]. Structural causes of drip loss concerning the muscle fibres and the biological processes associated with pork quality have been largely investigated and comprehended [11,19,189]. However, genetic mechanisms and interactions between different levels of metabolic regulation underlying drip loss are not fully understood [27,133,158].

Genetic studies, using standard approaches to identify candidate genes, already revealed several QTL and candidate genes for drip loss in pigs [27,70,190]. However, it can be expected that GWAS based on functional, metabolic phenotypes or metabotypes reduce the risk to detect false-positive associations [191]. Several studies have demonstrated that the results of any single omics analysis, like an association analysis of SNPs and phenotypic expression as implemented by GWAS, may not be sufficient to decode extremely complex biological mechanisms [1]. In the case of multifactorial traits, metabotypes can be used in order to improve the accuracy of the phenotypic measurement. The combined analysis of different omics levels provides a promising tool to increase the information density between genome and phenotype. Thereby, integrative approaches for overall analysis of the entire cascade of genome and metabolic levels (transcriptome, metabolome and proteome) provide a potential prospective to identify reliable biomarkers (transcripts, proteins, metabolites) and genetic markers (SNP, QTL, candidate genes) [116]. The knowledge of functional associated omics variables/markers including interactions between genetic and environmental factors may provide a comprehensive new insight into underlying biological processes in muscle growth and meat quality [1].

In recent years, the innovative technologies to record hundreds or thousands of omics profiles simultaneously and to analyse their relation to different traits were extensively developed further and established in many meat production sectors [2,88].

For meat scientists, the final objective is to identify meat quality genetic markers (like SNPs) or biomarkers which are quantifiable on live animals or early p.m. on the carcass in order to orientate meat production towards the most adapted processes in meat processing or distribution circuits [1]. For this purpose, until now, a variety of genetic approaches was applied (see [27,70,190]). In the last decade, several scientific groups investigated different omics levels or integrated two or more omics levels to identify candidate transcripts or genes. For example, Te Pas et al. [161], Rohart et al. [118], Muroya et al. [120] and Welzenbach et al. [192] investigated the suitability of metabolite profiles and metabolic pathways in prediction of pork quality traits. Heidt et al. [157] applied a combined genomics and transcriptomics approach to reveal candidate genes for drip loss. The investigation of metabolic components, like metabolites and proteins, as new, more reliable phenotypes is a research focus in enhancement of meat quality traits [87]. D´Alessandro et al. [99] used a combined metabolomic and proteomic analysis to investigate the biochemical background of breed-specific meat quality differences. Apart from a few exceptions (see [81]), there have been very few studies combining more than two omics levels to identify candidate genes and QTL for pork quality, until now.

The aim of this study is the integration of omics levels genome, proteome and metabolome to elucidate underlying functional pathways and corresponding candidate genes for drip loss. Based on the increased information density due to the consideration of proteome and metabolome, we expect that our GWAS approaches based on metabolic traits contribute to identify true candidate genes with higher accuracy.

3.2 Results

In this study, metabolite and protein profiling, Enrichment analysis and GWAS were performed on 97 F_2 Du × Pi pigs. The mean drip loss was 1.97%, with a minimum of 0.4% and a maximum of 5.3% (Tab. 12). In total, 1993 metabolites in each MLD sample were quantified, using GC-MS and LC-QTOF/MS. However, out of these, only 128 metabolites were matched to their related KEGG IDs. According to the results of previous examinations, 40 proteins with expected significance for drip loss were quantified in the

tissue samples. In the case of 35 proteins, we were able to annotate those with entrez gene identifiers.

Table 12: Descriptive statistics and phenotypic correlations between drip loss and metabotypes.

Traits	Mean ± SD[1]	Min[2]	Max[3]	Correlation to drip loss[4]
drip loss, %	1.97 ± 1.40	0.40	5.30	1
pH1	6.53 ± 0.22	5.89	6.94	−0.31 **
pH24	5.52 ± 1.12	5.32	6.06	−0.35 ***
PKM	26,454.10 ± 17,829.55	13.47	88,251.64	−0.20 *
PGAM2	5600.37 ± 4985.98	−10.77	32,935.16	−0.19
FBPase	27,407.08 ± 20,231.70	809.35	114,192.30	−0.11
TPI1	1754.68 ± 1526.65	32.13	7802.84	−0.21 *
pyruvic acid	$4.32 \times 10^{-2} \pm 3.62 \times 10^{-2}$	6.16×10^{-3}	2.11×10^{-1}	0.22 *
lactic acid	$6.49 \times 10^{-1} \pm 3.28 \times 10^{-1}$	1.88×10^{-1}	1.64	0.08
glucose	$9.02 \times 10^{-3} \pm 1.32 \times 10^{-2}$	1.21×10^{-4}	8.41×10^{-2}	0.19
phosphoenol pyruvate	$5.59 \times 10^{-2} \pm 8.95 \times 10^{-2}$	1.80×10^{-3}	0.53	0.13
glycerone-p	1.86 ± 1.10	2.48×10^{-1}	5.85	0.07
DG3P	$2.56 \times 10^{-1} \pm 4.09 \times 10^{-1}$	2.61×10^{-3}	2.61	0.14
fumaric acid	$2.67 \times 10^{-3} \pm 1.25 \times 10^{-3}$	5.50×10^{-4}	7.23×10^{-3}	0.12
succinic acid	$1.38 \times 10^{-2} \pm 5.02 \times 10^{-3}$	3.23×10^{-3}	3.23×10^{-2}	−0.02
malic acid	$6.03 \times 10^{-3} \pm 2.92 \times 10^{-3}$	8.85×10^{-4}	1.64×10^{-2}	0.11
methylglyoxal	$9.62 \times 10^{-3} \pm 5.44 \times 10^{-3}$	2.61×10^{-4}	2.89×10^{-2}	0.22 *
glycine	$8.59 \times 10^{-2} \pm 2.39 \times 10^{-2}$	4.84×10^{-2}	1.62×10^{-1}	0.11
hydroxypyruvic acid	$1.06 \times 10^{-2} \pm 6.81 \times 10^{-3}$	1.76×10^{-3}	4.98×10^{-2}	0.02
F6P	$2.17 \times 10^{-2} \pm 3.43 \times 10^{-2}$	2.91×10^{-4}	2.25×10^{-1}	0.12
serine	$6.04 \times 10^{-3} \pm 2.99 \times 10^{-3}$	1.76×10^{-3}	2.15×10^{-2}	−0.01
glycerone	$1.41 \times 10^{-1} \pm 8.36 \times 10^{-2}$	2.17×10^{-2}	4.37×10^{-1}	0.20
ceramide	$1.68 \times 10^{-4} \pm 1.24 \times 10^{-3}$	2.33×10^{-6}	6.59×10^{-4}	0.05
glucosylceramide	$2.46 \times 10^{-3} \pm 4.72 \times 10^{-3}$	1.69×10^{-4}	2.72×10^{-2}	0.21 *
phosphoethanol-amine	$8.57 \times 10^{-4} \pm 5.01 \times 10^{-4}$	2.28×10^{-4}	3.52×10^{-3}	0.12

[1]Mean and standard deviation (SD); [2]minimum (Min); [3]maximum (Max); [4]calculation of correlation coefficients based on residuals; Mean, SD, Min and Max of proteins are based

Continued explanations of Tab. 12:

on signal dependent intensities of ion fragments (in mass) relative to time, so-called selection reaction monitoring (SRM) intensities; The units of the metabolite profiles are based on mass intensities, recorded by GC-MS and LC-QTOF/MS, normalized to an internal standard; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; glycerone-p = dihydroxyacetone phosphate; PGAM2 = phosphoglycerate mutase 2 (muscle); PKM = pyruvate kinase (muscle); FBPase = fructose-1,6-bisphosphatase 2; TPI1 = triose phosphate isomerase 1; DG3P = D-glycerate-3-phosphate; F6P = fructose-6-phosphate; bold: proteins.

3.2.1 Biological pathways involved in the metabolite and protein abundance

In total, 163 metabolic components (128 metabolites and 35 proteins) were assigned to 219 KEGG pathways that potentially are involved in muscle to meat conversion and manifestation of meat quality characteristics. Based on the Wilcoxon rank sum test in ten out of 219 KEGG pathways, the metabolites and proteins were significantly enriched ($p \leq 0.05$) due to functional connectivity (Tab. 13). These pathways comprised in total 18 metabolites and four proteins and can be roughly distinguished into energy-relevant processes like ‘Glycolysis/gluconeogenesis’ and ‘Pyruvate metabolism’ and into pathways associated with different metabolic diseases like ‘Type II diabetes mellitus’ and ‘NAFLD’ (Non-alcoholic fatty liver disease). ‘Sphingolipid metabolism’ (p=0.014), that comprised four metabolites, was the most significantly enriched composition of metabolic components. Most metabolites and proteins were assigned to a single pathway. Of particular importance were across pathway components that might be indicators of key regulators with a strong impact on drip loss. As an example, the metabolites glucose and pyruvic acid are participants in six and five different pathways, respectively. The strongest overlapping induced by the metabolites glucose and pyruvic acid can be observed between glycolysis, methane and pyruvate metabolism showing that these pathways are closely linked. In contrast, the most significant pathway sphingolipid metabolism has only one link to the methane metabolism due to overlapping metabolite serine, whereas the remaining involved metabolites are exclusively members of sphingolipid metabolism.

Table 13: Significant KEGG pathways for drip loss.

Pathway	KEGG	p-value*	Involved metabolites/proteins
Sphingolipid metabolism	00600	0.014	ceramide, glucosylceramide, phosphoethanolamine, serine
Type II diabetes mellitus	04930	0.018	pyruvic acid, glucose, **PKM**
Methane metabolism	00680	0.020	glycine, pyruvic acid, hydroxypyruvic acid, F6P, malic acid, serine, phosphoenol pyruvate, glycerone-p, glycerone, DG3P
Renal cell carcinoma	05211	0.027	fumaric acid, malic acid
Insulin secretion	04911	0.043	pyruvic acid, glucose
Meiosis yeast	04113	0.045	glucose
NAFLD	04932	0.045	glucose
Glycolysis/ Gluconeogenesis	00010	0.045	pyruvic acid, lactic acid, glucose, phosphoenol pyruvate, glycerone-p, DG3P, **FBPase**, **TPI1**, **PKM**, **PGAM2**
Pyruvate metabolism	00620	0.053	fumaric acid, pyruvic acid, succinic acid, lactic acid, malic acid, phosphoenol pyruvate, methylglyoxal, **PKM**
Steptomycin biosynthesis	00521	0.056	glucose, myo-inositol

The enrichment analysis was performed based on 129 metabolites and 35 proteins. Overrepresentation of metabolic pathways defined by the KEGG database regarding to drip loss was tested using Wilcoxon's rank sum test; * The pathway was considered significant if $p \leq 0.05$; Kyoto Encyclopaedia of Genes and Genomes (KEGG)-ID = KEGG pathway ID; NAFLD = Non-alcoholic Fatty liver disease; glycerone-p = dihydroxyacetone phosphate; PGAM2 = phosphoglycerate mutase 2 (muscle); PKM = pyruvate kinase (muscle); FBPase = fructose-1,6-bisphosphatase 2; TPI1 = triosephosphate isomerase 1; DG3P = D-glycerate-3-phosphate; F6P = fructose-6-phosphate; bold: proteins.

Regarding the target trait drip loss, five metabolic components were significantly ($p \leq 0.05$) correlated (Tab. 12). Metabolites pyruvic acid, methylglyoxal and glucosylceramide were significantly positive correlated while the proteins pyruvate kinase (muscle) (PKM) and triose phosphate isomerase 1 (TPI1) were negative correlated with drip loss. However, the correlation coefficient was not above a value of 0.22 in any case (Tab. 12).

3.2.2 Whole-genome association analysis for drip loss and metabolites and proteins of selected biological pathways

Beneath the meat quality trait drip loss itself, 22 metabotypes (18 metabolites and four proteins) were analysed within a GWA study. In total, 44,844 SNPs were tested for association with at least one of the 22 metabolic traits or meat quality trait drip loss itself. In order to ensure the statistical power and accuracy of GWAS possible population stratification was considered [99]. In this context, PCs, which condensed the genetic relationship between animals, were considered in the statistical model as covariates. Depending on the investigated trait, between two and ten PCs were considered in order to avoid negative effects of population stratification on the validity of the GWAS (Tab. 14). In most traits, the genomic inflation factor λ was close to one with a range of one to 1.05. Accordingly to the λ-thresholds (1.05) suggested by Price et al. [193] our correction was sufficient to remove disturbing population stratification. Only in the case of phosphoethanolamine the λ value (1.08) was slightly too high (Tab. 14).

Applying a moderate significance threshold with a false discovery rate (FDR) of $q \leq 0.10$, the GWAS revealed 871 (without double counting) significant associations for 15 traits, including drip loss, three proteins and 11 metabolites. These SNP were distributed over almost all porcine chromosomes. Four hundred thirty one SNPs showed a chromosome-wide significance levels of $q \leq 0.05$ but no SNP was detected as genome-wide significant ($q \leq 0.01$).

The average number of significant SNPs per trait was 66, with a minimum of two SNPs (for glucose) and a maximum of 249 SNPs for fructose-1,6-biphosphatase-2 (FBPase). The majority of the SNPs was significant at a moderate chromosome-wide level ($q \leq 0.1$). The highest proportion of explained variance was observed for SNPs that affected glucosylceramide, dihydroxyacetone phosphate (glycerone-p) and D-glycerate-3-phosphate (DG3P). The most significant SNPs were detected for metabolite hydroxylpyruvic acid ($q \leq 2.19 \times 10^{-2}$).

The average number of detected SNPs per chromosome is 67 and the highest numbers of significant SNPs were detected on SSC14, 17 and 18. For drip loss, we detected SNPs on SSC16 and 18, which explain a maximum of variance proportion of 8.8%. Based on the distance to neighboring significant SNPs on the chromosome (1Mb), we condensed the

SNPs into 330 important QTL regions with an average of 29 QTL per chromosome (Tab. 14).

On several chromosomes we identified 126 (45) SNPs (QTL) which were significant for more than one trait. These SNPs are located on SSC1, 7, 8, 14, 17 and 18. As presented in Fig. 16 the most overlapping exists between metabolites hydroxypyruvic acid and succinic acid on SSC14. Moreover, the overlapping on SSC18 is of particular interest, because it indicates a metabolic process comprising glycine and phosphoglycerate mutase 2 (PGAM2) that influences drip loss (Fig. 16). On SSC7, there was only one overlapping SNP of glucose and fructose-6-phosphate (F6P). In contrast, on SSC1 and 8, we indeed detected significant SNPs for two traits but the QTL are located in distant chromosomal regions.

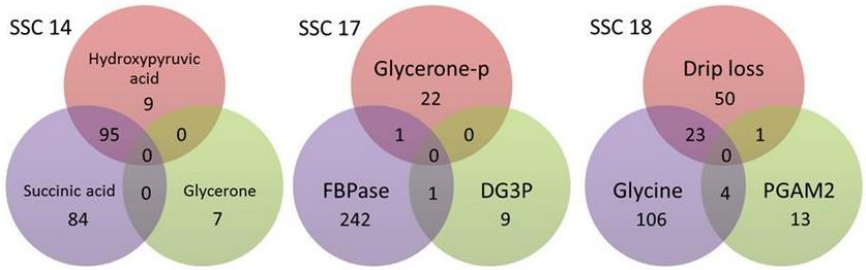


Figure 16: Overlapping SNPs at *Sus scrofa* chromosomes (SSC) 14, 17 and 18. GWAS procedures resulted in varying numbers of significant SNPs (q ≤ 0.1) per trait. On some chromosomes there are overlapping SNPs with meaning for two traits; Glycerone-p = dihydroxyacetone phosphate; PGAM2 = phosphoglycerate mutase 2 (muscle); FBPase = fructose-1,6-bisphosphatase 2; DG3P = D-glycerate-3-phosphate.

Table 14: Results of association analyses. Continued on next page.

Trait	PC[1]	λ^2	Number of Significant SNP/QTL per Porcine Chromosome[3]													∑SNP[4]	Min p^5	Min q^6	Max σ^2_y [7]
			1	2	3	4	6	7	8	10	13	14	16	17	18				
drip loss	10	1.007											4/4		74/20	78 #	6.58	6.26	8.8
PKM	10	1				33/13										33 #	10.7	7.84	14.3
PGAM2	10	1.06													18/7	18 #	19.9	8.67	13.9
FBPase	10	1									5/1			244/92		118 *, 131 #	1.98	2.27	16.6
glucose	10	1						2/2								2 #	5.86	8.80	15.3
glycerone-p	10	1.046	4/1						7/4					23/10		34 #	2.35	5.07	17.3
DG3P	10	1												10/5		2 *, 8 #	1.50	2.19	17.3
succinic acid	2	1.03										179/64				122 *, 57 #	29.3	5.07	13.3
glycine	10	1.05		97/41						2/2					133/48	102 *, 130 #	3.39	4.67	17.1
hydroxyl-pyruvic acid	10	1										104/28				76 *, 28 #	3.44	1.88	16.1
F6P	10	1						12/9								12 #	8.00	7.69	14.8
glycerone	10	1										7/4				7 #	7.95	8.56	14.8
ceramide	4	1.006							20/8							20 #	11.8	8.02	14.4
glucosyl-ceramide	10	1.012	3/3		1/1											4 #	1.59	6.64	17.4

Table 14: Continued.

Trait	PC[1]	λ[2]	Number of Significant SNP/QTL per Porcine Chromosome[3]													∑SNP[4]	Min p^5	Min q^6	Max σ^2_y [7]
			1	2	3	4	6	7	8	10	13	14	16	17	18				
phosphor-ethanol-amine	10	1.08					15/8									11 *, 4 #	15.4	3.81	14.5
∑SNP/QTL excluding double counting			7/4	97/41	1/1	33/13	15/8	13/10	27/12	2/2	5/1	195/80	4/4	275/100	197/54				
∑overlapping SNP/QTL [8]								1/1				95/16		2/7	28/21				

[1]Number of principal components (PCs) considered in GWAS; [2]λ = inflation factor; [3]number of chromosome-wide significant associated SNPs and QTL per traits and chromosome (at least $q \leq 0.1$); [4]sum of significant associated SNPs per traits (* $q \leq 0.05$; # $q \leq 0.1$); [5]minimal empirical *p*-value (times 10^{-5}); [6]minimal *q*-value (times 10^{-5}), based in the false discovery rate concept; [7]maximal proportion of explained variance (%); [8]sum of overlapping SNP/QTL with meaning for two traits; glycerone-p = dihydroxyacetone phosphate; PGAM2 = phosphoglycerate mutase 2 (muscle); PKM = pyruvate kinase (muscle); FBPase = fructose-1,6-bisphosphatase 2; DG3P = D-glycerate-3-phosphate; F6P = fructose-6-phosphate; bold: proteins.

The functional annotation of the 871 significantly associated SNPs revealed 1,430 genes that are located in a distance of ≤1Mb to the SNPs and thereby are in linkage disequilibrium to our significant SNPs (Tab. 15). 257 SNPs are localized in an intron region of a specific gene. These genes, which are mainly located on SSC14, 17 and 18, might be important potential candidate genes for drip loss and associated metabolic traits and processes (Tab. 15).

Table 15: Functional annotation of significant SNPs associated with drip loss and metabolic traits.

SSC[1]	**1**	**2**	**3**	**4**	**6**	**7**	**8**	**10**	**13**	**14**	**16**	**17**	**18**	∑
Genes[2]	30	148	4	65	31	48	70	15	12	375	13	367	252	1430
SNP[3]	2/ 7	30/ 97	-/ 1	15/ 33	2/ 15	-/ 13	2/ 27	1/ 2	5/ 5	83/ 195	-/ 4	54/ 275	63/ 197	257/ 871

[1]*Sus scrofa* chromosomes; [2]number of genes that are located in a distance of ≤1Mb **to** the significant SNPs revealed by GWAS; [3]number of intronic SNPs in relation to the total number of significant SNPs per chromosome (without double counting of overlapping SNPs).

For the identification of potential candidate genes, we concentrated on the most important QTL regions with a high density of significant SNPs for different traits. These SNPs were selected based on the following three criteria: The SNPs had to be:

1. chromosome-wide significant (at least $q \leq 0.1$);
2. within the 'Top 10' or 'Top 25' of significant SNPs for metabolic traits or drip loss;
3. exonic or intronic.

Using these criteria we identified 23 potential candidate genes for drip loss and nine associated metabolic components (Tab. 16). SSC18 is of particular interest, because on this chromosome we identified candidate genes for drip loss, glycine and PGAM2. The number of detected genes for a single trait varied between one and six. On SSC4 six genes in a range of 20Mb were detected for protein PKM. The importance of each candidate gene is

indicated by one to five significant intronic SNPs. Five genes (*ZNHIT6, HLCS, ANK3, RASGEF1A* and *LRGUK*) harbor more than one intronic SNPs. Based on the QTL comprising five intronic SNPs in a small range of 0.29Mb, it might be reasonably assumed that HLCS is a very promising candidate gene for FBPase. Most significant intronic SNPs with highest proportion of explained variance in a range of 15.28% to 17.44% were detected for glucosyl-ceramide, glycerone-p and glycine (Tab. 16).

For drip loss, five candidate genes were identified on SSC18 (Tab. 16 and 17). The most significant SNPs (Var_{max}=8.82%; $p_{min} \leq 6.58 \times 10^{-5}$) associated with drip loss were detected on SSC16, but these SNPs do not fulfill the previously described conditions to detect potential candidate genes (Tab. 17). Distributed over four regions, SSC18 harbors two genes for PGAM2, four genes for drip loss and one gene (*LRGUK*) significantly associated with drip loss and glycine. Because '*Leucine-rich repeats and guanylate kinase domain containing*' (*LRGUK)* is in linkage disequilibrium with *EXOC4* that was associated with drip loss as well, this region ranging from 15.9Mb to 16.1Mb is of particular interest. From 12.2Mb to 12.9Mb there is a second interesting region with two candidate genes, for PGAM2 and drip loss, respectively. The Manhattan plot of SSC18 is presented in Fig. 17. Moreover, the Manhattan plots of SSC1, 4, 6, 10, 13, 14 and 17 are shown in S5 (p. 199).

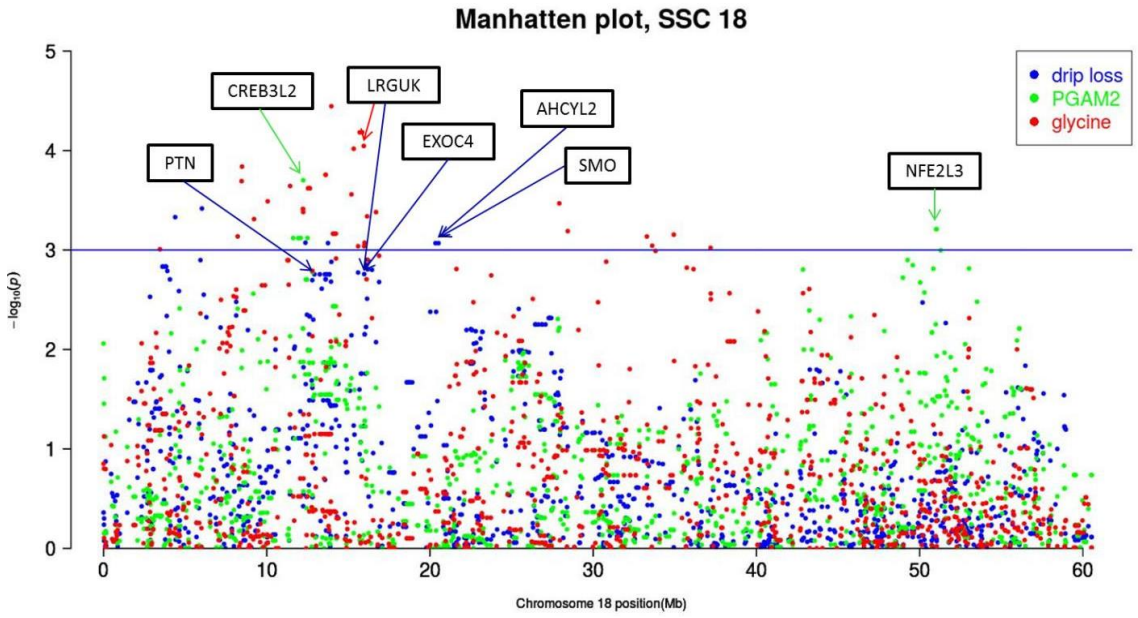


Figure 17: Chromosome-wide Manhattan plot of *Sus scrofa* chromosome (SSC) 18. PGAM2 = phosphoglycerate mutase 2; the declaration of gene symbols (in black boxes) can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/genegenes.

Table 16: Annotation of most promising SNPs for drip loss and associated metabolic components. Continued on next page.

SSC[1]	Trait	Gene[2]	SNP[3]	Position[4]	Mut[5]	MAF[6]	eEff (se)[7]		Chi2	emp. p[8]	q [9]	*Var* [10]
1	glycerone-p	*ENPP3*	INRA0001633	35387799	G/A	0.47	-4.00×10^{-2}	(1.00×10^{-2})	18.68	0.22	5.07	17.35
	glucosyl-ceramide	*SAMD4A*	ALGA0007238	204522804	C/A	0.47	-9.32×10^{-5}	(2.15×10^{-5})	18.80	0.16	6.64	17.44
4	**PKM**	*NTNG1*	INRA0016801	123080603	G/A	0.27	-9.21×10^{2}	(2.57×10^{2})	12.88	3.32	7.84	13.21
		GBP4	ASGA0023322	139599066	G/A	0.38	-6.43×10^{2}	(1.77×10^{2})	13.26	2.71	7.84	12.72
		PKN2	M1GA0006779	139861416	C/A	0.43	6.76×10^{2}	(1.89×10^{2})	12.88	3.32	7.84	12.40
		ZNHIT6	ALGA0029718	142789911	A/G	0.46	8.52×10^{2}	(2.20×10^{2})	15.01	1.07	7.84	14.29
			ALGA0029732	142739989	G/A	0.39	9.23×10^{2}	(2.49×10^{2})	13.70	2.14	7.84	13.09
			ALGA0029741	142730172	G/A	0.46	8.13×10^{2}	(2.15×10^{2})	14.20	1.64	7.84	13.50
		DDAH1	ASGA0023626	143204232	A/G	0.40	9.05×10^{2}	(2.43×10^{2})	13.86	1.97	7.84	13.21
		WDR63	INRA0018033	143449789	A/G	0.40	9.05×10^{2}	(2.43×10^{2})	13.86	1.97	7.84	10.77
6	phosphor-ethanolamine	*PIK3C3*	DRGA0006746	118055075	G/A	0.26	2.91×10^{-5}	(7.54×10^{-6})	14.93	1.76	3.81	14.36
		TTLL5	INRA0022204	120225026	C/A	0.26	2.91×10^{-5}	(7.54×10^{-6})	14.93	1.76	3.81	14.36
10	glycine	*AKT3*	MARC0098464	18065301	C/A	0.34	-1.55×10^{-3}	(3.80×10^{-4})	16.56	0.69	5.11	15.69
13	**FBPase**	*HLCS*	MARC0019610	210504370	G/A	0.49	6.54×10^{2}	(1.70×10^{2})	14.71	1.25	8.64	13.92
			MARC0005075	210516458	A/C	0.49	6.54×10^{2}	(1.70×10^{2})	14.71	1.25	8.64	13.92
			ASGA0089689	210516937	G/A	0.49	6.54×10^{2}	(1.70×10^{2})	14.71	1.25	8.64	13.92
			ASGA0089950	210531047	A/G	0.49	6.54×10^{2}	(1.70×10^{2})	14.71	1.25	8.64	13.92
			ASGA0097399	210534054	G/C	0.49	6.54×10^{2}	(1.70×10^{2})	14.71	1.25	8.64	13.92
14	succinic acid	*ANK3*	MARC0033238	68550413	G/A	0.52	1.69×10^{-4}	(4.59×10^{-5})	13.60	2.93	2.82	13.26
			ASGA0064107	68604989	A/G	0.52	1.69×10^{-4}	(4.59×10^{-5})	13.60	2.93	2.82	13.26
		RASGEF1A	ALGA0078235	66284845	G/A	0.52	1.69×10^{-4}	(4.59×10^{-4})	13.60	2.93	2.82	13.26
			ALGA0078240	66320818	A/C	0.52	1.69×10^{-4}	(4.59×10^{-5})	13.60	2.93	2.82	13.26
			ALGA0078243	66332408	G/A	0.52	1.69×10^{-4}	(4.59×10^{-5})	13.60	2.93	2.82	13.26

Table 16: Continued.

SSC[1]	Trait	Gene[2]	SNP[3]	Position[4]	Mut[5]	MAF[6]	eEff (se)[7]		Chi2	emp. p[8]	q[9]	*Var*[10]
17	DG3P	*PTPRT*	MARC0016232	50694545	A/G	0.41	-1.96×10^{-2}	(5.27×10^{-3})	13.88	1.94	6.53	13.49
		VAPB	H3GA0049968	65818274	A/G	0.48	1.71×10^{-2}	(4.79×10^{-3})	12.78	3.51	6.53	12.55
18	PGAM2	*CREB3L2*	ALGA0107449	12234417	G/A	0.41	1.88×10^{2}	(4.90×10^{1})	14.79	1.99	8.67	13.98
	drip loss	*PTN*	ALGA0097051	12921061	A/G	0.25	-7.81×10^{-2}	(2.49×10^{-2})	9.87	17.6	6.26	5.61
	glycine	*LRGUK*	ASGA0079000	15942579	A/G	0.31	-1.63×10^{-3}	(4.07×10^{-4})	16.06	0.90	1.66	15.28
	drip loss		ALGA0097170	15969549	G/A	0.45	-4.34×10^{-2}	(1.38×10^{-2})	9.87	17.5	6.26	5.61
		EXOC4	DIAS0001125	16179365	G/A	0.48	4.15×10^{-2}	(1.31×10^{-2})	10.08	15.6	6.26	5.72
		AHCYL2	H3GA0050495	20338092	A/G	0.28	-7.16×10^{-2}	(2.14×10^{-2})	11.21	8.54	6.26	6.32
		SMO	ASGA0079098	20520014	G/A	0.30	-7.16×10^{-2}	(2.14×10^{-2})	11.21	8.54	6.26	6.32
	PGAM2	*NFE2L3*	ASGA0100894	51012467	C/A	0.42	1.89×10^{2}	(5.35×10^{1})	12.53	6.18	8.67	12.10

The SNP order complies with number of chromosomes and position on the chromosome; Selection of promising SNPs based on the criteria, that they are (1) chromosome-wide significant (at least $p < 0.1$); (2) within the 'Top 10' significant SNPs per metabolic trait or 'Top 25' for drip loss and (3) located within an annotated gene; [1]*Sus scrofa* chromosomes (SSC); [2]The declaration of gene symbols can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/gene; [3]None of the SNPs is located in an exon region of the regarding candidate gene; [4]position in BP (base pairs); [5]mutation (Mut); [6]minor allele frequency (MAF); [7]eEff = substitution effect and se = standard error; [8]empirical *p*-value, times 10^{-4}; [9]*q*-value (based on the false discovery rate (FDR) concept), times 10^{-2}; [10]Var = proportion of the explained variation [%]; glycerone-p = dihydroxyacetone phosphate; PGAM2 =

Table 17: 'Top 25' significant SNPs identified for drip loss and potential candidate genes. Continued on next page.

SNP	SSC[1]	Position[2]	Mut[3]	MAF[4]	eEff (se)[5]		Chi2	Emp. p[6]	q[7]	*Var*[8]	Located within a Gene[9]
ALGA0089069	16	11629284	C/A	0.08	2.26×10^{-1}	(5.65×10^{-2})	16.05	6.58×10^{-5}	7.02×10^{-2}	8.82	×
CASI0008411	16	23115634	G/A	0.10	1.89×10^{-1}	(4.86×10^{-2})	15.03	1.12×10^{-4}	7.02×10^{-2}	8.30	×
MARC0097282	16	10946289	G/A	0.33	7.45×10^{-2}	(1.95×10^{-2})	14.65	1.38×10^{-4}	7.02×10^{-2}	8.15	×
ASGA0072217	16	9183890	A/G	0.34	7.25×10^{-2}	(1.93×10^{-2})	14.16	1.78×10^{-4}	7.02×10^{-2}	7.90	×
ALGA0111681	18	6026724	G/A	0.15	1.35×10^{-1}	(3.78×10^{-2})	12.71	3.83×10^{-4}	6.26×10^{-2}	7.11	×
ASGA0104044	18	4388048	A/C	0.15	1.27×10^{-1}	(3.61×10^{-2})	12.34	4.68×10^{-4}	6.26×10^{-2}	6.92	×
MARC0003904	18	12368984	G/A	0.35	-6.37×10^{-2}	(1.90×10^{-2})	11.23	8.45×10^{-4}	6.26×10^{-2}	6.33	×
ASGA0078921	18	13751595	G/A	0.29	-7.46×10^{-2}	(2.23×10^{-2})	11.21	8.53×10^{-4}	6.26×10^{-2}	6.33	×
H3GA0050495	18	20338092	G/A	0.30	-7.16×10^{-2}	(2.14×10^{-2})	11.21	8.54×10^{-4}	6.26×10^{-2}	6.33	*AHCYL2*
ASGA0079098	18	20520014	A/G	0.30	-7.16×10^{-2}	(2.14×10^{-2})	11.21	8.54×10^{-4}	6.26×10^{-2}	6.33	*SMO*
ALGA0105391	18	5935981	G/A	0.31	6.69×10^{-2}	(2.07×10^{-2})	10.48	1.26×10^{-3}	6.26×10^{-2}	5.94	×
INRA0055248	18	13959002	G/A	0.47	-4.24×10^{-2}	(1.31×10^{-2})	10.40	1.32×10^{-3}	6.26×10^{-2}	5.90	×
MARC0036783	18	16113241	A/G	0.47	-4.23×10^{-2}	(1.32×10^{-2})	10.27	1.41×10^{-3}	6.26×10^{-2}	5.82	×
ASGA0098607	18	3614625	A/G	0.38	-5.35×10^{-2}	(1.68×10^{-2})	10.20	1.47×10^{-3}	6.26×10^{-2}	5.79	×
ALGA0104874	18	3620895	A/G	0.38	-5.35×10^{-2}	(1.68×10^{-2})	10.20	1.47×10^{-3}	6.26×10^{-2}	5.79	×
ASGA0088995	18	3741888	G/G	0.38	-5.35×10^{-2}	(1.68×10^{-2})	10.20	1.47×10^{-3}	6.26×10^{-2}	5.79	×
H3GA0050278	18	3808173	A/G	0.38	-5.35×10^{-2}	(1.68×10^{-2})	10.20	1.47×10^{-3}	6.26×10^{-2}	5.79	×
ASGA0078689	18	3833808	G/A	0.38	-5.35×10^{-2}	(1.68×10^{-2})	10.20	1.47×10^{-3}	6.26×10^{-2}	5.79	×

Table 17: Continued.

SNP	SSC[1]	Position[2]	Mut[3]	MAF[4]	eEff (se)[5]	Chi2	Emp. p[6]	q[7]	*Var*[8]	Located within a Gene[9]
DIAS0001125	18	16179365	G/A	0.48	4.15×10^{-2} (1.31×10^{-2})	10.08	1.56×10^{-3}	6.26×10^{-2}	5.72	*EXOC4*
ALGA0097186	18	16444813	G/A	0.47	4.20×10^{-2} (1.32×10^{-2})	10.06	1.58×10^{-3}	6.26×10^{-2}	5.71	×
ALGA0096804	18	3907848	G/A	032	5.58×10^{-2} (1.77×10^{-2})	10.01	1.63×10^{-3}	6.26×10^{-2}	5.69	×
ALGA0116114	18	15594213	A/G	0.46	-4.22×10^{-2} (1.34×10^{-2})	9.94	1.69×10^{-3}	6.26×10^{-2}	5.65	×
ALGA0097170	18	15969549	G/A	0.45	-4.34×10^{-2} (1.38×10^{-2})	9.88	1.75×10^{-3}	6.26×10^{-2}	5.62	*LRGUK*
ALGA0097051	18	12921061	A/G	0.25	-7.81×10^{-2} (2.49×10^{-2})	9.87	1.76×10^{-3}	6.26×10^{-2}	5.61	*PTN*
ALGA0097067	18	13674866	A/G	0.25	-7.81×10^{-2} (2.49×10^{-2})	9.87	1.76×10^{-3}	6.26×10^{-2}	5.61	×

The SNP order complies with raising *p*-value; [1]*Sus scrofa* chromosomes (SSC); [2]position in base pairs (BP); [3]mutation (Mut); [4]minor allele frequency (MAF); [5]substitution effect and standard error (se); [6]empirical *p*-value and significant thresholds, [7]*q*-value (based in the false discovery rate (FDR) concept); [8]Var = proportion of the explained variation [%]; [9]The declaration of gene symbols can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/gene, '×', SNP is not located within a gene, none of the SNPs is located in an exon region of the regarding gene.

3.3 Discussion

Systems Biological Approach or Integrated Analysis of Genome, Proteome and Metabolome to Elucidate the 'Muscle to Meat' Black Box

Based on the multitude of possible post-transcriptional events, the genetic information flow from SNPs to phenotypic variations is not linearly dispersed in living organisms and samples collected p.m. [194]. This situation describes the black box between genes and phenotypes that needs to be opened to detect genetic variation influencing complex traits. Several studies have demonstrated that the results of single omics analysis, like standard GWA procedure, may not be sufficient to decode extremely complex biological mechanisms [1]. A possible solution is to integrate different omics levels in genetic analyses and to analyse the entire cascade of metabolic levels. The omics levels proteome and metabolome were chosen for our analysis because we expected that these metabotypes are the final products of specific pathways and thereby are closely connected with classical target phenotypes routinely measured in animal production [118,161]. While the genome (SNP information) contains the information on which allele variants exist, the other omics levels indicate which genes are actually being expressed and which pathways are active. Therefore, metabolites and proteins constitute essential links between genetic information and phenotypical expression of complex traits and might be used in genetic association studies to improve the statistical power and to reveal less false positive, redundant results [118]. The concentration of metabolites and proteins in muscle and blood compared to drip loss is less influenced by environmental effects and thereby can be used as more accurate phenotype to identify candidate genes. This means, intermediate phenotypes might be more appropriate to estimate the genetic potential of the individuals than the performance trait itself. For example, a pig with excellent genetic potential for high meat quality and low drip loss might show high drip loss caused by bad environmental factors and management effects. In this case, drip loss is a poor indicator for the effective genetic potential of the individual.

To elucidate biological pathways affecting a trait, the consideration of the proteome is advantageous compared to the transcriptome. This can be assumed because the amount of proteins is not only regulated by a constant level of transcript expression but also by many possible genetic interacting mechanisms of protein regulation/modification and connected activation of other pathways [147]. In a similar context, Ala-Korpela et al. [195],

Kadarmideen [53] and Widmann et al. [147] have stated that systems biological approaches are valuable and powerful in identifying key causal and highly predictive genetic variants for complex traits as well as in building up complex genetic regulatory networks.

Impact of metabolic pathways and involved metabolites and proteins for drip loss

In this study, metabolite profiling was based on an untargeted metabolomics approach to uncover the whole metabolome. Compared to that, proteins were profiled more specific by means of a targeted proteomics approach using the absolute quantification of 40 proteins that have been shown as important indicators for drip loss in previous investigations. For the final enrichment analysis 128 annotated metabolites and 35 proteins were used. Five proteins were rejected because of missing entrez gene identifier. The drastic reduction of the number of metabolites from 1,865 to only 128 is a severe bottleneck, so that it is highly probable that even metabolites with strong influence on drip loss were excluded. This situation is caused by the fragmentary information of biochemical functions of metabolites that is stored in metabolome databases. According to Chagoyen and Pazos [181], this lack of scientific fundamentals and principles of physiological and biochemical processes of higher life forms is a big challenge in systems biology studies. In a similar way, Chagoyen and Pazos [181] argued that there is a need of more accurate profiling tools for omics phenotypes in order to get a more comprehensive insight into the metabolic processes.

Our enrichment analysis considered all available annotated metabolome and proteome information and revealed ten functional KEGG pathways with significant ($p \leq 0.05$) enriched components. The applied test mean-rank gene-set enrichment (MR-GSE) statistic is based on Pearson's correlation coefficients between metabotypes and drip loss and averages the ranks of the applied statistics instead of the statistics themselves. This procedure makes the results less influenced by individual components in the set of variables [196] and is the main difference to other usually applied testing procedures, like the Tktest of Tian et al. [197]. Further details are given by Ackermann and Strimmer [198].

In summary, it can be expected that the underlying function of our applied enrichment test has enough power to detect overrepresented groups of variables (e.g., genes or metabotypes), even if the effects are very small or the amount of data is not sufficient to detect the important variables individually [196]. This argument can be used to explain, why our enrichment analysis has resulted in functional sets of metabotypes although

correlation coefficients between individual metabotypes and drip loss do not significantly deviate from zero (Tab. 12, p. 81).

In our study, we observed particularly pathways and corresponding key regulators which affect muscle metabolism related to meat quality traits. Glycolysis, pyruvate and methane metabolism are strongly connected and belong to the most important energetic processes that influence the muscle to meat conversion [9,199]. Because drip loss strongly depends on p.m. energetic processes in muscle, the meaning of glycolysis and pyruvate metabolism is obvious. After slaughtering, in muscle tissues, anaerobe metabolic processes predominate and, in glycolysis, glycogen is released via glucose to pyruvic acid. Under aerobic conditions, pyruvic acid is metabolized in citrate cycle and oxidative phosphorylation [9]. In the case of stress before slaughtering, in hypoxic tissues the rate of oxidative processes like glycolysis is increased and pyruvic acid does not flow into glycolysis but is transferred to lactic acid. Accumulation of lactic acid goes along with pH decrease to 5.6 [165]. The meaning of metabolic processes associated with energy metabolism for drip loss is confirmed by a multitude of studies. Among others, Binke [14], Scheffler and Gerrard [9] and D'Alessandro et al. [99] allocated the relevance of glycolysis and pyruvate metabolism for meat quality. The coincidence of low early pH values and high temperature in muscle lead to partial denaturation of proteins and reduction of intercellular space. Thereby, lipids are dissolved from membranes, permeability of membranes is increased and drip loss is the result [189]. In cell exudate dissolved lipids clarify the connection between drip loss and activity of sphingolipid metabolism that includes the metabolisation of ceramides, phosphoethanolamine and serines. The relation between drip loss and associated lipids and acids has been already described by Lambert et al. [187] and Poulsen et al. [188].

As a result of our enrichment analysis, the metabolite glycine is associated with drip loss. In methane metabolism the enzyme glyoxylate transaminase catalyzes the metabolisation of metabolite glyoxylate into glycine or hydroxypyruvic acid (www.genome.jp). High glycine contents indicate a higher rate of glycolytic processes. A high glycolytic potential is known to be related with high drip loss. The link between drip loss and glycine was already described by Lim et al. [200], who observed higher drip loss in the case of higher glycine level in porcine skeletal muscle cells.

The meaning of PKM that is involved in pathways glycolysis/gluconeogenesis, pyruvate metabolism and type II diabetes mellitus (Tab. 13) was already clarified by several studies. For example, D'Alessandro et al. [99] confirmed that the PKM level appeared to be highly related to many meat quality criteria (WHC, meat color). Beneath PKM, PGAM2 and DG3P are also involved in glycolysis/gluconeogenesis and pyruvate metabolism. Under anaerobic conditions PGAM2 catalyzed the degradation of DG3P to 2-phosphoglycerates (see S6). Because high levels of glycolytic enzymes like phosphoglycerates are associated with increased drip loss [97], PGAM2 might be considered as an appropriate indicator for drip loss [103]. In addition, Davoli et al. [201] appreciated that the corresponding gene PGAM2, is a potential candidate gene for drip loss. The non-essential α-amino acid glycine is also product of catabolism of DG3P and is thus part of the same metabolic process as PGAM2.

Another section of glycolysis/gluconeogenesis illustrates the interactions of the enzymes FBPase and TPI1 and the metabolite dihydroxyacetone phosphate (glycerone-p). In gluconeogenesis FBPase converts fructose-1,6-biphosphate to F6P and in glycolysis phosphofructokinase catalyzes the metabolisation of F6P to fructose-1,6-bisphophate. In the following process of glycolysis, the enzyme fructose-bisphosphate aldolase converts fructose-1,6-bisphophate to glycerone-p. In the next step, glycerone-p is metabolized to glyceraldehyde-3-phosphate catalyzed by TPI1 (see S6, p. 202). Laville et al. [202] revealed a significant correlation between high TPI1 and tender meat with low drip loss. The meaning of FBPase for meat quality in pigs was described by Nam et al. [203]. They detected a lower FBPase expression in pigs with high drip loss and weak pH decrease p.m. [203].

Beneath metabolic processes whose activity directly depends on the individual energy resources, also sphingolipid metabolism is significantly associated with drip loss. With a p-value of 0.014, metabolic compounds in sphingolipid metabolism are the most strongly enriched metabolites and proteins in our study and thereby have an obvious effect on drip loss. According to Heidt et al. [157] there is a negative correlation between drip loss in Du × Pi pigs and transcripts associated with sphingolipid metabolism. According to our analysis, the metabolites ceramide, glucosylceramide, phosphoethanolamine and serine are involved in sphingolipid metabolism. Ceramides are lipid signaling molecules that activate proliferative or apoptotic pathways. They are products of the metabolism of free FAs to

long-chain fatty acyl-CoAs (LCACoAs). LCACoAs can either be used for energy production through β-oxidation or undergo conversion to various signaling molecules, such as ceramide and diacylglycerol [204]. In the analysis of differentially expressed transcripts in Du × Pi pigs, Ponsuksili et al. [133] concluded that low drip loss is associated with ceramide pathways. Especially, drip loss is associated with ceramides as lipid signaling molecules that can activate proliferative or apoptotic pathways. The ceramide biosynthesis is part of the sphingolipid metabolism and ceramides arise from the conversion of complex sphingolipids such as glucosylceramides. According to Dobrowsky and Kolesnick [205], the levels of ceramides and glucosylceramides and the enzymes regulating their metabolism are associated with the cells response to stress. The degradation of membranes accompanies with cell stress and as a consequence drip loss has a relation to metabolites that indicate cell stress. This connection explains the relationship between drip loss and transformation products of sphingolipid metabolism.

The metabolic processes and their involved components and overlapping are presented in S6–S9. Several metabolic components, such as glucose and pyruvic acid are involved in five of ten pathways relevant for drip loss. The connective position of these metabolites confirms their specific role as metabolic key players in the regulation of meat quality. The meaning of the disease related pathways (e.g., type II diabetes mellitus) and other processes (meiosis in yeast) for drip loss (Tab. 13, p. 83) are based on the strong influence of specific involved metabolic components like glucose and pyruvic acid. It is not to be expected that there is in fact a physiological connection between meiosis in yeast and meat quality in pigs.

Significant markers and candidate genes for drip loss and associated metabolic traits

Drip loss is a complex trait that is genetically controlled by a variety of different genes [133] and is influenced by interaction of metabolic processes and participants like genes, transcripts, proteins and metabolites [11]. Against this background, it is problematic to identify genes with a strong influence on drip loss using classical GWAS approaches. Moreover, statistical problems like stratification within the investigated population increase the risk of false positive results. In order to adjust for population stratification we included PCs as fixed effects into the model of the GWAS procedures as suggested by Aulchenko et al. [206] and applied among others by Becker et al. [207] and Utsunomiya et al. [208].

Depending on the investigated trait (drip loss, protein, metabolite) the models contain two to ten PCs, which lead to λ-values close to one. From these results we conclude a sufficient elimination of population stratification without unacceptable reduction of the genetic variation.

Instead of a Bonferroni correction, that favors the occurrence of false negative associations [209], we used the q-value which based on the FDR to correct for multiple testing. Storey and Tibshirani [210] suggested including the FDR in GWAS to provide a better balance between statistical significance and power to detect true effects. As it has been recommended by Benjamini and Hochberg [211], we set a relaxed significant threshold of $q \leq 0.10$.

The performed GWAS procedures resulted in a varying number of significant SNPs for drip, 11 metabolites and three proteins. The total of 871 significant SNPs are spread across the entire porcine genome, but concentrated on SSC14, 17 and 18. For drip loss itself, promising candidate genes are located on SSC18. This region has been earlier described by Jennen et al. [212] and Liu et al. [27]. In the region around 12Mb, the meaning of '*Sus scrofa* pleiotropic factor beta' (PTN) ($q \leq 6.26 \times 10^{-2}$) is highlighted by the direct neighborhood of gene 'cAMP responsive element binding protein' (CREB3L2). CREB3L2 was identified by the GWAS of the protein PGAM2, which revealed an intronic SNP (ALGA0107449) as one of the most significant marker (Tab. 16, p. 91). The family of cAMP response element binding proteins is crucial for a variety of cellular processes including cell proliferation, differentiation, apoptosis, extra-stimuli and stress response [213]. Although the meaning of CREB3L2 so far was not precisely described for meat quality, our results suggest that this gene seems to have a relevant influence in energy metabolism in skeletal muscle that is indicated by its interacting effect on PGAM2, glycine and drip loss (Fig. 18).

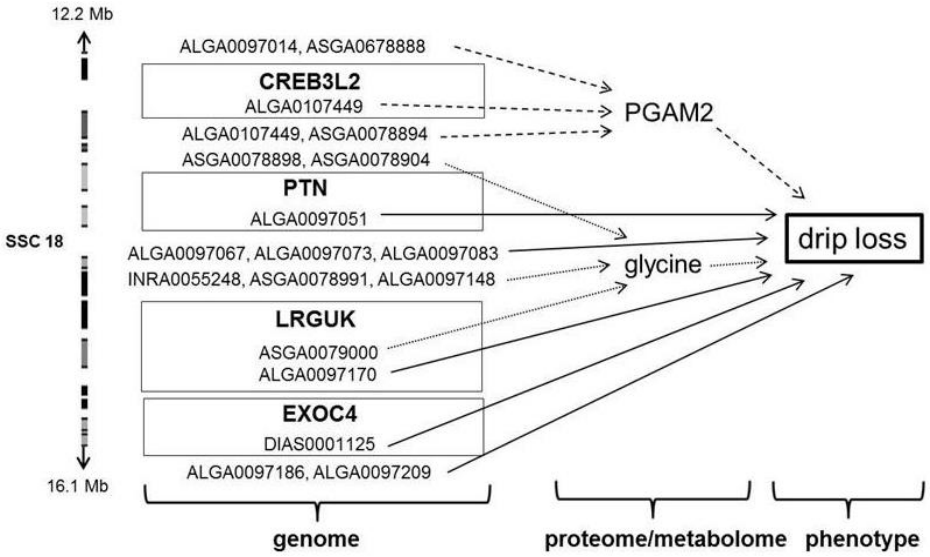


Figure 18: Region on Sus scrofa chromosome SC18 with potential candidate genes for drip loss and associated metabolic traits phosphoglycerate mutase 2 and glycine. PGAM2 = phosphoglycerate mutase 2; fat solid arrows = direct relation between SNPs and drip loss; thin solid arrows = indirect relation between SNPs and drip loss via metabolite glycine; dotted arrow = indirect relation between SNPs and drip loss via protein PGAM2; genes in boxes: CREB3L2 = cAMP responsive element binding protein 3 like 2; PTN = *Sus scrofa* pleiotropic factor beta; LRGUK = leucine-rich repeats and guanylate kinase domain containing; EXOC4 = exocyst complex component 4.

In the second interesting region on SSC18 from 15.9 to 16.1Mb, two intronic SNPs located in the gene LRGUK were found. These SNPs are ranked in the 'Top 10 list' for drip loss as well as for glycine. The nearby gene 'Exocyst complex component 4' (EXOC4) is also significantly associated with drip loss. EXOC4 is part of the exocyst complex (Exo70), which is involved in insulin-stimulated glucose transport. Due to Laramie et al. [214], in humans polymorphisms near EXOC4 and LRGUK on chromosome 7 are associated with type 2 diabetes and fasting glucose. The metabolic pathway that is regulated by the polymorphisms near EXOC4 and LRGUK potentially is also relevant for drip loss in pork, because fasting glucose also effects the pH decrease in muscle p.m. and drip loss. The

investigations of Leheska et al. [215] demonstrated that fasting before slaughtering yielded in a significant lower glucose level and weaker pH decrease in muscle p.m. and in less drip loss. In the third interesting region on SSC18 around 20Mb, directly next to each other genes 'Adenosylhomocysteinase-like 2' (AHCYL2) and 'Smoothened, frizzled class receptor' (SMO) are located and significantly associated with drip loss. Just like the polymorphism between EXOC4 and LRGUK, AHCYL2 is associated with type 2 diabetes [216]. Until now, there is no further evidence that this chromosomal region has an influence on meat quality. The effect of gene 'Nuclear factor, erythroid 2-like 3' (NFE2L3) at 51Mb, associated with protein PGAM2, fits into the same metabolic background like the previously described genes [217]. In summary, the multitude of significant SNPs detected for drip loss and associated metabotypes gives an ambiguous indication that in the described regions on SSC18 promising candidate genes for drip loss can be expected.

In this study, the most significant SNPs were detected on SSC1. Two SNPs ($p \leq 2.23 \times 10^{-5}$ and $p \leq 1.59 \times 10^{-5}$) associated with glycerone-p and glucosylceramide, are located within the genes 'Ectonucleotidepyrophosphatase/phosphodiesterase 3' (ENPP3) and 'Sterile alpha motif domain containing 4a' (SAMD4A). ENPP3 is associated with lipid and FA metabolism and it has been reported by to Liu et al. [218] that this gene affects fat deposition and skeletal muscle growth in pigs. SAMD4A is also associated with lipid metabolism [219] and influences the metabolisation of glucosylceramides that is part of sphingolipid metabolism. Combining biological knowledge found in literature and the highly significant results of our enrichment analysis leads to the conclusion that the sphingolipid metabolism is one of the most important metabolic pathways associated with drip loss.

Beneath glucosylceramides, phosphoethanolamines are also key players in sphingolipid metabolism. Two genes significantly associated with this metabolite were detected on SSC6 (Tab. 16). 'Phosphatidylinositol 3-kinase, catalytic subunit type 3' (PIK3C3) is involved in the regulation of hepatic glucose output, glycogen synthase, and antilipolysis in typical insulin target cells such as those in the liver, muscle and fat tissue [220]. Among others, PIK3C3 influences the cellular response to glucose starvation (GO term: 0042149). This biological process describes the change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of deprivation

of glucose. According to Kim and Volsky [220] and Hirose et al. [221] a polymorphism in PIK3C3 is associated with body weight and carcass fat in LR and Du pigs.

Moreover, we identified potential candidate genes for several metabolic components involved in glycolysis/gluconeogenesis. The protein PKM is one of the most prominent members of these pathways. The activity of PKM is decreased in the case of low glucose availability in muscle that is positive correlated with anabolic cellular processes. During the conversion of muscle to meat, the metabolic processes change into the catabolic range and if glucose is used up very early, the PKM level is significantly associated with the aberrant glycolysis leading to PSE development [9]. In our analysis, it was shown that PKM is influenced by six candidate genes on SSC4. In the chromosomal region of 139Mb, genes 'Guanylate-binding protein 4' (GBP4) and 'Protein kinase N2' (PKN2) are located. Zhao et al. [222] have identified GBP4 as a significant QTL for LMC of pigs by comparing two divergent pig breeds with respect to carcass composition traits. Fontanesi et al. [223] have reported markers close to PKN2 that were associated with back fat thickness. SSC4 harbors two genes (ZNHIT6, DDAH1) within a region of 142–143Mb which were significantly associated with average daily gain in LW pigs [223]. These polymorphisms seem to have a strong impact on the metabolic rate and the deposition of skeletal muscle mass.

Two SNPs on SSC17 give evidence that 'Protein tyrosine phosphatase, receptor type' (PTPRT) and 'VAMP (vesicle-associated membrane protein)—associated protein B and C' (VAPB) are candidate genes that affect the metabolite DG3P. The protein encoded by PTPRT is a signaling molecule that regulates a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. In humans, PTPRT is strongly associated with high-fat diet-induced obesity and insulin resistance [224,225]. Moreover, in beef cattle, Tizioto et al. [226] identified PTPRT as candidate gene for shear force. With respect to the negative correlation between intramuscular fat content and shear force both studies came to homogeneous results. The importance of PTPRT is additionally indicated by the fact that the most important intronic SNP of PTPRT is an overlapping SNP that is also significantly associated with protein FBPase (Fig. 16, p. 91). DG3P and FBPase are strongly connected in glycolysis/gluconeogenesis and PTPRT might be a key player in regulation of glycolysis and thus a promising candidate gene for several meat quality traits.

Challenges and perspectives

As it has been postulated by Fiehn [143] and Krastanov [227], the development and performance of omics approaches have revolutionized the collection of biological data. Detection, quantification and annotation of hundreds of thousands of variables in tissue or blood samples presupposes enormous progress in chip technology, technical profiling/screening method, expansion of biological databases and handling of high dimensional data sets. From the statistical point of view, there are some unsolved questions, how to weight or to integrate the different omics levels in a statistical model. Genomic selection tools provide solution to weight large amount of SNP information in the case of a limited number of animals [51]. In a similar way 'Omics based selection' (OBS) methods try to weight genetic, transcriptional and metabolic information in an optimal manner. Under the condition of a successful weighting of metabotypes and the correct consideration of exogenous factors and the time point of profiling, OBS has the perspective to be an effective strategy in animal breeding, monitoring of state of health and supply status (e.g., nutritional metabolomics) and early disease detection (e.g., molecular epidemiology). Finally it should be mentioned that the profiling of metabotypes is non-invasive and may be performed in living organisms [228]. However, because of the complex interaction of genes, transcripts, proteins and metabolites these methods are conceptually very demanding and generally accepted methods are still missing. Moreover, beneath not standardized statistical methods to integrate omics data, the possibilities of metabolite and protein annotation are limited due to the fragmentary information of regarding databases. As a solution, network analyses might be valuable for the integration of multi omics data and the indirect annotation of unknown omics components based on the functional connectivity within a module of the network. A further difficulty is the dynamics of metabotypes in dependence of environmental effects and processing conditions. Biochemical processes response very quickly and dynamic to changes in exogenous factors. While the genetic information remains constant during the lifetime of an individual, the expression of transcripts, proteins and metabolites is very dynamic and regulated by a large number of factors. Thus, proteomic and metabolomic approaches can be viewed as recording of the metabolic status at a specific time point in a system of steady dynamic nature. Consequently, in estimation of performance traits the time point of metabolite and protein profiling has to take into account precisely.

3.4 Materials and methods

3.4.1 Animals, tissue collection, phenotyping

This study is based on 97 animals of a reciprocal Du × Pi crossbreed. The animals were selected from F_2 families and based on their extreme high or low values of drip loss [157]. The animals were kept and performance tested under standardized conditions at the Frankenforst experimental farm of the University of Bonn from 2002 until 2007. Data recording and sample collection were conducted strictly in line with the German law on animal welfare. The entire experiment, including applied standard operating procedures, was approved by the veterinary and food inspection, Siegburg, Germany (No. 39600305-547/15). All animals were slaughtered at an average of 180.5 days (range 151–223 days) and average carcass weight of 86.5kg (range 73.0–101.8kg). The phenotypes were recorded in a commercial slaughterhouse according to the rules of German performance stations (ZDS, 29). Further information can be found in Liu et al. [27] and Heidt et al. [157].

Sample collection was performed about ten min p.m., immediately after exsanguination. Tissue samples were rapidly dissected, snap-frozen in liquid nitrogen and stored at −80°C. Drip loss was measured in MLD using the bag method of Honikel and Kim [165]. The samples from MLD between 13th/14th rib (one chop per individual) with a thickness of 2.5–3.0cm were collected 24h p.m., weighed, and suspended in a plastic bag. After storage for 48h at 4°C, the samples were reweighed and drip loss were calculated as a percentage of weight loss based on the initial weight of a sample. In the tested animals drip loss ranged between 0.4% and 5.3%, whereby 49 pigs have drip loss values of lower 1.5% and the remaining 48 pigs show drip loss values of ≥1.5 %.

3.4.2 Untargeted metabolite profiling

For metabolite profiling we choose an untargeted approach to screen the entire metabolome. The metabolite spectra in the MLD samples of 97 Du × Pi pigs were measured by Metabolomic Discoveries GmbH (Potsdam, Germany; www.metabolomicdiscoveries.com) via gas GC-MS and LC-QTOF/MS. GC-MS and LC-QTOF/MS facilitate the identification and quantification of a few hundred metabolites in a single tissue sample. Chromatography followed by mass spectrometry has a relatively broad coverage of compound classes, including organic and amino acids, sugars, sugar alcohols, phosphorylated intermediates and lipophilic compounds. With the combination of

both methods it is possible to detect metabolites in a range of 50–1,700Dalton, with a precision of 1–2 part per min (ppm) and a solution of mass/Δmass=40.000 (Report METABOLOMIC DISCOVERIES GmbH). For details on the LC-QTOF/MS method see Lisec et al. [166]. Metabolites were identified and annotated in comparison to Metabolomic Discoveries' databases, which resort to Human Metabolome Database (HMDB, www.hmdb.ca), METLIN (www.metlin.scripps.edu/) and Lipid Maps (www.lipidmaps.org/). Annotation of metabolites was based on mass assignment, retention behavior and structure information. Metabolites, which could not be annotated, are characterised by their accurate mass and retention time. For details in metabolite quantification and annotation see Welzenbach et al. [144]. Only metabolites with known KEGG-ID were used for further analysis. KEGG-IDs were obtained using R packages KEGGREST, biomaRt and AnnotationDbi of Bioconductor (https://www.bioconductor.org) based on HMDB-IDs.

3.4.3 Targeted protein profiling

For the protein quantification and annotation we applied a two-step procedure. In the first step, an untargeted proteome profiling approach via isotope-coded protein labeling (ICPL) was used to determine the whole proteome (holistic approach) in MLD samples of 42 Du × Pi pigs selected based on their extreme phenotypes of drip loss. In the second step (validation step), a set of 40 selected proteins was quantified in the 97 Du × Pi pigs of this study (targeted protein profiling approach).

The ICPL procedure, which based on stable isotope labeling, combined with mass spectrometry has emerged as a powerful tool to identify and relatively quantify thousands of proteins within complex protein mixtures [229]. In contrast to traditional proteomics approaches e.g., by 2D-gel quantified proteins was based on mass spectra and database query amongst others with the ICPL-Quant software. Based on the holistic ICPL approach and literature research, 40 proteins with expected meaning for drip loss were selected. These proteins were validated via selected reaction monitoring (SRM) in the 97 DuPi pigs of this study. Using a triple quadrupole mass spectrometer, targeted SRM offers high selectivity, sensitivity and a wide dynamic range in the quantitative analysis of small molecules [81]. The ICPL and SRM analyses were performed by TOPLAP GmbH (Munich, Germany). For a more detailed description of the ICPL and SRM application in our samples, see Kellermann [79] and Gallien et al. [81]

Based on the available entrez gene ID or ensemble peptide ID, R packages KEGGREST, biomaRt and AnnotationDbi of Bioconductor (https://www.bioconductor.org) were used to identify the corresponding KEGG-IDs of the proteins.

3.4.4 Genome profiling

DNA was extracted from MLD using a Genomic DNA Purification Kit (Fermentas Life Science, Thermo Fisher Scientific, Waltham, MA, USA). DNA concentration was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and concentration was adjusted to 100ng/µL by using double-distilled RNase and DNase free water. Illumina bead array technology (Porcine SNP 60 K Bead Chip) was used for genotyping the samples (Illumina, Inc., San Diego, CA, USA) in accordance with the protocol for SNP Infinium HD assay (http://Illumina.com). DNA (200 ng) was used for genome-wide amplification and fragmentation. A quality score for each genotype was generated. Sample preparation and genotyping has been described by Heidt et al. [157].

3.4.5 Statistical methods

Quality Control and Annotation of Genetic Data

Quality control was performed as implemented in R package GenABEL [206]. SNPs were excluded from further analysis under the following conditions: (a) Minor allele frequency (MAF) < 1%; (b) Call rate < 95 % and (c) strong deviation from the Hardy-Weinberg-Equilibrium ($p < 10^{-3}$). After checking the quality of the data, 97 animals and 44,844 SNPs remained in the data set.

Pig Sscrofa 10.2 (International Swine Genome Sequencing Consortium) [230] was used to annotate all investigated SNPs. In order to detect biologically relevant genes being in linkage disequilibrium with significant associated SNPs, the R package biomaRt [231] was used. This procedure of functional annotation filtered genes in a distance of up to 1Mb around the significant SNP regions. We chose this window, because in our assumption there is an association between SNP and potential candidate gene if the distance is ≤1 Mb.

Metabolite and protein enrichment and pathway analysis

In order to investigate the overrepresentation of specific metabolite and protein sets in different KEGG pathways, an enrichment analysis was performed based on corresponding annotated metabolites and proteins and the target trait drip loss. For assignment of metabolites and proteins to relevant metabolic pathways, R package biomaRt was applied [231].

The enrichment analysis was performed as implemented in R package limma [228]. The underlying test procedure of limma, called MR-GSE, was developed by Michaud et al. [196] and refers to a Wilcoxon rank-sum test. The test statistic ranks the sets of metabolic components based on Pearson's correlation coefficients between paired samples of metabolites/proteins and drip loss. The result is a list of ranked compositions containing a varying number of metabolites and proteins. It was assumed that significantly ($p \leq 0.05$) enriched sets of metabolic components represent specific functional pathways that might be associated with muscle metabolism and meat quality traits. The procedure computes a *p*-value to test the hypothesis that a set of variables (metabolites and proteins) tends to be more highly ranked in terms of a given test statistic compared to randomly chosen variables. The calculated *p*-value indicates whether a set of variables is statistically independent that means that the variables are on average less or equally correlated than randomly chosen variables (H_0 hypothesis), or whether a set of metabolites and proteins is enriched because of functional connectivity (H_1 hypothesis). In the following step, metabolites and proteins of significant enriched functional pathways were analyed in GWAS.

Genome-wide association analysis

The GWAS for pork quality parameter drip loss and metabolites/proteins of significant pathways was applied using the R package GenABEL [206]. The phenotypic traits (drip loss, metabolite/protein expression values) were corrected for SD and SW:

$$y_{ijk} = \mu + SD_j + \beta_s SW_i + \beta_g g_{ik} + e_{ijk} \quad (1)$$

where y_{ij} is the phenotype of the *i*-th individual. Fixed effect *SD* and as covariable *SW* with regression coefficient β_s are implemented in the model. Genetic effects were estimated via a fixed covariable 'genotype' (g_{ik}) and corresponding regression coefficient (β_g). The

significance of each SNP was tested using a fast score test. In order to verify potential stratification in our F_2 DuPi population, the inflation factor λ, which depends on the squared original test statistic of the *i*-th SNP (T_i^2) was calculated as

$$\lambda = \frac{Median(T_i^2)}{0.4549} \quad (2)$$

Aulchenko et al. [52] and Price et al. [193] showed that an inflation factor λ in the range of 1.0 to 1.05 is an indicator of a sufficiently corrected population stratification which can be analysed with an acceptable risk of false positive results. Preliminary results of our analysis showed that λ deviates slightly from this optimum. This implies that some population stratification exist within our F_2 DuPi pigs. In order to correct for this problem, within the fast score test, PCs estimated from the genomic kinship (EIGENSTRAT) [193,206], were included as fixed covariables. The genomic kinship matrix was used to reveal the PCs reflecting the axes of genetic variation and describing the stratification of the populations involved in this study. The number of PCs used in this step is variable and depends on the ability to correct different levels of population stratifications. The number of PCs was increased stepwise from one to ten PCs and the final number of PCs was chosen so that the inflation factor λ was nearest to one.

The correction of phenotypes, the estimation of the PCs and the association analysis was performed with the function 'egscore' as implemented in the R-package GenABEL. In order to reduce the risk of false-positive associations, the SNP significance tests were corrected for multiple testing based on the *q*-value calculation. This approach is a significance measurement based on the false discovery rate (FDR) concept [210]. We chose a significance threshold of $q \leq 0.1$. The variance explained by the respective SNP was calculated using following formula:

$$Var\ (\%) = \frac{\tilde{\chi_{1df}}}{(\mathrm{n} - 2 + \chi_{1df}^2)} \quad (3)$$

where χ^2 is the result of the score test as implemented in GenABEL and n the number of individuals. This formula resulted from the transformation of a Student's *t*-distribution into a *z*-distribution [232]. Based on a similar MAF, a similar allele substitution effect and a similar proportion of explained variance we assumed that SNPs within a distance of <1Mb to each other belong to one QTL.

3.5 Conclusion

Systems biological approaches utilize the information content of all available omics variables (SNPs, transcripts, proteins, and metabolites) in order to clarify the physiological, biochemical and genetic background of complex traits. Theoretically, across omics utilization is advantageous in comparison to classical genetic approaches, which merely investigate associations between SNPs and phenotype.

In this context, Picard et al. [1] and van der Van der Sijde et al. [124] have stated that there is an increasing interest to combine all the omics levels in a holistic omics approach to investigate the complexity of the molecular events beyond expected biological functions and to identify important genes. It can be expected that meat quality traits are influenced by a high number of interacting genes that are unknown or involved in unexpected functions, so that the across omics level approach used in our study is particularly useful. Based on the described integrated analysis of the omics levels genome, proteome, metabolome and phenotype, we increased the information density between genes and trait of interest to decode the complex biological mechanisms influencing drip loss and to reveal promising candidate genes. At least some of these genes have not been detected based on a standard GWAS procedure. The most promising candidate genes were located on SSC18 where we detected several, partly overlapping QTL for drip loss itself and the intermediate phenotypes PGAM2 and glycine. These candidate genes need further investigations to identify underlying functional mutations affecting drip loss and the related metabolic processes.

Based on the results of this study, it was possible to confirm the already known findings about the importance of energy related metabolic processes influencing meat quality and particularly drip loss. On the other hand, this study also provides novel insights into the underlying biochemical pathways of drip loss. According to our findings, the sphingolipid metabolism is of particular importance for drip loss manifestation. The involved metabolites glucosylceramide and phosphoethanolamine are promising intermediate phenotypes for drip loss and revealed promising candidate genes on SSC1. It can be expected that such integrated omics approaches might be successfully applied to clarify the biochemical and genetic background also in more complex traits than meat quality. Against this background, our study may be considered as a model investigation to test one possible procedure to combine different omics levels.

Chapter 4. Systems biology Analysis of Porcine Muscle Profiles Reveals Underlying Biochemical Processes of Pork Quality and Composition Traits and Highlights Potential Candidate Genes

Short title: Systems biological Analysis to Investigate Biochemical Processes and Candidate Genes Affecting Pork Quality and Composition.

Julia Welzenbach[1], Christiane Neuhoff[1], Christian Looft[1], Hanna Heidt[2], Karl Schellander[1], Ernst Tholen[1], Christine Große-Brinkhaus[1]*

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

Email: jwel@itw.uni-bonn.de (J.W.), cneu@itw.uni-bonn.de (C.N.), cloo@itw.uni-bonn.de (C.L.), ksch@itw.uni-bonn.de (K.S.), etho@itw.uni-bonn.de (E.T.), cgro@itw.uni-bonn.de (C.G.)

[2]Institute for Organic Agriculture Luxembourg, Association sans but lucratif (A.S.B.L)., 13 Rue Gabriel Lippmann, L-5365 Munsbach, Luxembourg

*Corresponding author: Institute of Animal Science, University of Bonn, Endenicher Allee 15, 53115 Bonn, Germany, cgro@itw.uni-bonn.de, +49228733386

Not yet published.

4.1 Introduction

Aim of this study was to characterise underlying functional pathways and corresponding candidate genes for meat quality and particularly drip loss based on a systems biology approach. Pork quality is the result of complex interactions between genetic and environmental effects like rearing and slaughtering conditions, and carcass and meat processing. One important commercially interesting pork quality parameter is drip loss. Based on reduced acceptance of processing industries and consumers, high drip loss leads to monetary losses in pork production. Heritability estimates of drip loss vary to a large extent between 0.01 and 0.31. This wide range could be explained by breed effects and large measurement errors of drip loss due to the multifactorial environmental effects [32]. While structural and physiological causes of drip loss concerning the muscle fibres and cells have been largely investigated and comprehended, genetic mechanisms and interactions between different layers of metabolic regulation underlying drip loss are not fully understood [27,133,158].

Several studies have demonstrated that the results of any single omics analysis, like an association analysis of SNPs and phenotypic expression as implemented in GWAS, might not be sufficient to decode extremely complex biological mechanisms [1]. Moreover, in high dimensional omics data sets the number of measured omics variables is usually much larger than the number of available biological samples. This statistical, critical point is also known as the ‘large p, small n’ problem or overfitting that might lead to random noise like a high percentage of false-positive associations in GWAS [164].

In a systems biology approach different omics profiles are used as so-called intermediate phenotypes in order to increase the information density between genome and phenotype and as more accurate phenotypes compared to the conventional performance traits. Thereby, particularly in case of multifactorial traits like drip loss, an integrative analysis of the entire cascade of different omics levels provides a promising tool to identify reliable biomarkers (transcripts, proteins, metabolites) and genetic markers (SNP, QTL, candidate genes) [116]. The knowledge of functional associated omics variables/markers including interactions between genetic and environmental factors might provide comprehensive new insights into underlying biological processes in muscle growth and the black box of ‘muscle to meat conversion’ [1].

In the integration of multiple omics data sets, network analyses constitute a promising perspective. In these analyses, high dimensional data sets are condensed with the consequence that overfitting is reduced. In this study, we performed a WNA that is a powerful tool to reduce our omics data to the most important metabolic processes (modules of the network) correlated with meat quality. We used selected modules as metabolic phenotypes in GWAS to confirm the relationship between phenotype and genetic markers. By analysing different meat quality and carcass composition traits and several modules as auxiliary characteristic in GWAS, instead of the conventional GWAS approach for drip loss directly, our methodical procedure is beneficial in the genetic analysis of complex, multifactorial traits. In this context, it can be expected that GWAS for metabolic processes related to drip loss yield in reduced risk to detect false-positive associations and in higher accuracy in the identification of candidate genes [191,233].

The purpose of this study was the characterisation of underlying functional pathways and corresponding candidate genes for meat quality and particularly drip loss based on a systems biology approach. A WNA was used to integrate transcript, protein and metabolite profiles by modulating modules that are associated with metabolic pathways. Moreover, this approach was selected to reduce overfitting. The following GWAS based on modules were performed to increase the information density between phenotype and genotype and to detect reliable genetic markers for drip loss.

4.2 Material, methods and statistical analysis

To figure the whole metabolic background of drip loss we investigated the meat quality traits pH1, pH24 and color and additionally the related carcass composition traits meat fat ratio (MFR), backfat thickness (BFT) and lean meat content in belly (LMC_{belly}) and carcass lean meat content calculated with the Bonner formula (LMC_{bonn}). The whole omics cascade was quantified in MLD samples of 90 F_2 Du × Pi pigs. For details in the measurement of meat quality parameters see Welzenbach et al. [192, pp. 2-3]. BFT [cm] is described as average value of measured rind thickness at withers (thickest part), middle of the back (thinnest part) and loin. The parameter MFR is the quotient of weight-specific fat area [cm²] to weight-specific muscle area [cm²] of the back. The LMC in belly [%] is calculated by the formula of Grub and weighted the BFT in loin [cm], the fat thickness alongside [cm], the meat area [cm²] and the fat area [cm²]. Another parameter indicating the LMC is calculated by the Bonner formula. The regression equation calculating the LMC_{bonn} for

crossbreed pigs weights seven parameters related to meat and fat content. See ZDS [29] for more details in stationary performance testing.

Fig. 19 gives an overview of the available data basis and the performed statistical procedures. In the first step, a WNA was used to form combined metabolic phenotypes, so-called modules, consisting of transcripts, metabolites and proteins. These modules were annotated (Step 2) and treated as more accurate phenotypes compared to the conventional meat quality and carcass composition parameters and were analysed in GWAS (Step 3). Finally, as verification of the identified candidate genes, we performed eQTL analyses for the hub players of the modules (Step 4).

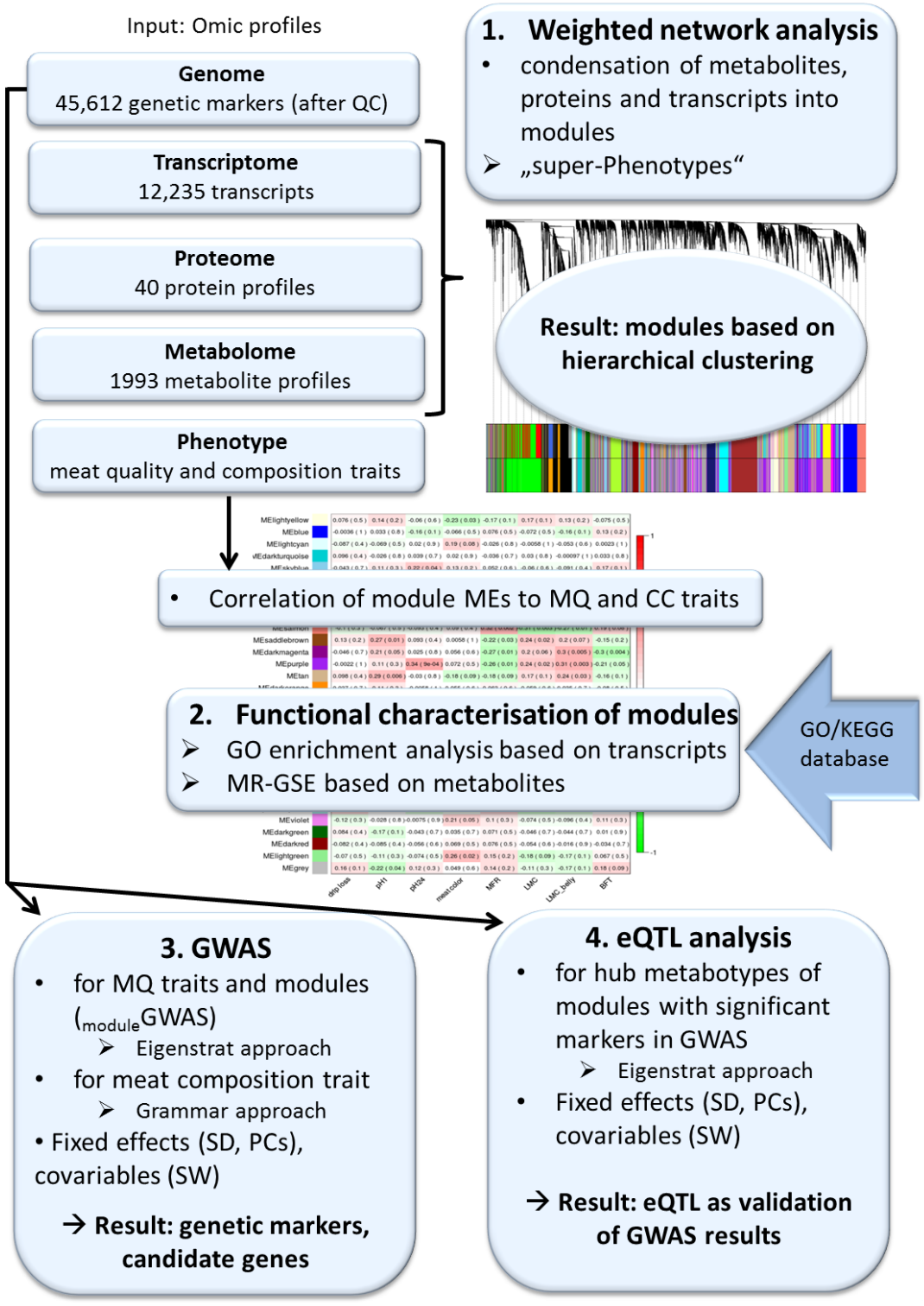


Figure 19: Workflow of the analysis. QC – quality control; ME – module eigenvalues; MQ – meat quality, CC – carcass composition; MR-GSE – mean-rank gene set enrichment; SD – slaughter date; SW – slaughter weight; PCs – principal components.

4.2.1 Weighted network analysis

This methodical approach is beneficial in handling the 'large p, small n' problem in various ways. First, the WNA is a powerful tool to condense the omics data set to the most important modules correlated with meat quality. Thereby, the procedure results in a reduction of the number of dependent variables 'p'. Moreover, the network construction based on transcripts, proteins and metabolites allows the identification of molecular interactions because they do not only quantify the correlation between traits and omics variables, but also the extent to which the omics components are connected among themselves. In this context, network analyses allow the indirect annotation of unknown metabolic components based on their annotated neighbours within a module of the network.

In our study, in the first step transcripts (n=12,235), proteins (n=40) and metabolites (n=1,993) were condensed into modules based on a WNA. The omics profiling of metabolites and proteins has been described by Welzenbach et al. [144, sections 4.2 and 4.3, p. 17]. Gene expression profiling in MLD samples was performed by Heidt et al. [157] according to the Affymetrix protocols (Affymetrix, Santa Clara, CA, USA). Muscle expression patterns were assessed using the GeneChip® Porcine Genome Array (Affymetrix), which contains 23,937 probe sets that interrogate 23,256 transcripts which represent 20,201 genes in pig. Hybridization and scanning using the GeneChip® Scanner 3000 (Affymetrix) were performed according to the manufacturer's protocols. Data were analysed with Affymetrix's GCOS 1.1.1 software using global scaling to a target signal of 500. In order to reject abundant transcripts, we filtered for the transcripts with highest expression and variance. Accordingly, only transcripts with variance >0.5 and mean expression value of >6.0 were considered in further process. For details in the construction of a weighted network using R package WGCNA by Langfelder and Horvath [164].

The modules were related to the phenotypic traits and functional annotation based on gene ontology (GO) enrichment analyses and/or MR-GSE analyses. Based on module specific transcripts, hypergeometric gene set enrichment tests (GOstats package ver. 2.24.0) were performed to investigate the overrepresentation of transcript sets defined by the groups Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF) in the GO database (GO; http://www.geneontology.org/) or by KEGG terms, using Fisher's exact test. A gene set was considered as significant if $p \leq 0.05$. In case of KEGG-/entrez gene ID annotated metabolites/proteins within a module, the annotation was performed

using a MR-GSE procedure, based on a Wilcoxon rank sum test. Further details have been described in section 3.2.1 (p. 82). Aim of the WNA and the module annotation was the detection of metabolic pathways associated with drip loss and related traits.

4.2.2 Imputation

In order to get the greatest possible number of pigs with recorded profiles of all omics levels (genome, transcriptome, proteome, metabolome and phenotype), we imputed missing genotypes of seven Du × Pi animals, from which transcriptomic, proteomic and metabolomic information were available. Imputation was performed by FImpute [234]. FImpute uses an overlapping sliding window approach to efficiently exploit relationships and haplotype similarities between target and reference individuals. The pedigree of our animals with missing or inaccurate genotypes comprises 321 genotyped Du × Pi pigs that were used as reference individuals. After imputation of the genotypes of seven animals, a total of 90 Du × Pi pigs with profiling of the whole omics cascade were available.

4.2.3 GWAS for classical and metabotype indicators for meat quality and carcass composition traits

In order to identify reliable genetic markers and candidate genes for meat quality and carcass composition traits several GWAS analysis were performed. These analyses comprised genomic SNP data as independent variable and various classical and composite indicators of meat quality or carcass composition as dependent variables. Classical indicators were BFT, LMC_{bonn}, LMC_{belly}, MFR, pH1, pH24 and meat color. As composite indicators, important WNA based modules were used, where the module eigenvalues showed a high correlation to at least one of the target carcass composition or meat quality traits. The eigenvalues of selected modules were used as dependent phenotypes.

Quality control and annotation of the genetic data has been described in detail by Welzenbach et al. [144, section 4.5.1., p. 18]. In brief, the SNPs were checked for call rate of <0.95, MAF>0.01 and Hardy-Weinberg disequilibrium $p<10^{-3}$. The annotation was performed with R package biomaRt [231] based on the Pig Sscrofa 10.2 database.

As described in the publication of Welzenbach et al. [144], the GWAS analysis was problematic due to genetic stratification in the analysed population. In order to remove the negative consequences of population stratification different GWAS approaches (EIGENSTRAT, GRAMMAR) implemented in the R package GenABEL [193,206] were realized.

EIGENSTRAT

For meat quality traits and modules existing population stratification were corrected by means of a specific number of PCs (*) estimated from the genomic kinship. This proceeding is implemented in the EIGENSTRAT approach in GenABEL [193,206]. In general the following underlying statistical GWAS model was used:

$$y_{ijk} = \mu + SD_j + \beta_s SW_i + \beta_g g*_{ik} + e_{ijk}$$

where y_{ijk} is the phenotype of the i-th individual. As relevant environmental fixed effect the model contains the 'slaughter day' (*SD)* and as a covariable the 'slaughter weight' (*SW)* with regression coefficient β_s. Genetic effects were estimated via a fixed covariable 'genotype' (g_{ik}) and corresponding regression coefficient (β_g). To correct for population stratification the covariable 'genotype' and the phenotype were linear adjusted for a specific number of PCs (*), estimated from the genomic kinship.

In order to verify the success of the correction, the genomic inflation factor λ was calculated. More detailed information of the calculation of λ are given in the publication of Welzenbach et al. [144]. Price et al. [193] showed that the correction to remove disturbing population was sufficient if the resulting λ was in a range of 1.0 to 1.05. A series of GWAS were realized using an increasing number of PCs which were included step by step into the statistical model. After each round λ was calculated. This stepwise GWAS approach was stopped if the optimum range of λ was reached.

GRAMMAR

In the selected Du × Pi animals, the population stratification regarding the carcass composition traits was lower than in meat quality traits. For this reason, the correction of population stratification using the EIGENSTRAT approach was too stern in carcass composition traits. Preliminary tests showed that the GRAMMAR approach of Aulchenko

et al. [235] was best suited to analyse carcass composition traits. The GRAMMAR approach is a fast approximate association tests between a trait and genetic polymorphisms in samples with genetic sub-structure. In a first step, phenotypes are corrected for relevant fixed effects and a random polygenetic animal effect, which reflects the genetic variance and genomic relations between animals, is calculated. The genomic relationship is estimated by the genomic kinship relationship between all animals. In this step the following statistical model was used:

$$y_{ijk} = \mu + S_j + Sexj + \beta_S SD_j + \beta_s SW_i + a_i + e_{ijk} \quad (1)$$

where y_{ijk} is the phenotype of the i-th individual. As fixed effects *SD*, *Sex* and season (*S*) and as covariable *SW* with regression coefficient β_s are implemented in the model. Moreover, the polygenetic component a_i maximises the likelihood of the phenotypic data.

In a second step of the GRAMMAR approach genetic effects were estimated by means of a simple model, which comprises the population mean μ and the fixed covariable 'genotype' (g_{ik}) and corresponding regression coefficient (β_g). As dependent variables y_{ik} the residuals e_{ijk} of model step 1 are used.

$$y_{ik} = \mu + \beta_g g_{ik} + e_{ik} \quad (2)$$

Finally, we corrected for multiple testing based on the q-value calculation (see Welzenbach et al. [144], section 4.5.3., pp. 18-19 for details).

4.2.4 Verification step: eQTL analysis for hub metabotypes

In a verification step, GWAS (eQTL analyses) were performed by using the most important hub metabotypes (transcripts, metabolites and proteins) of the selected modules. According to Heidt et al. [157], "a QTL analysis of expression levels of genes identifies genomic regions that are likely to contain at least one causal gene with a regulatory effect on the expression level, termed eQTL." The positions of the resulting significant eQTL were compared with the location of candidate genes (or the associated SNPs), which was detected in the GWAS for classical phenotypes and composite module traits. This approach was motivated by the expectation, that in case of overlapping QTL regions the risk of false positive results is reduced and the biological relevance of these regions can be confirmed with higher accuracy.

For this validation step it was necessary to identify hub metabotypes, which are the most important single values within a module. Most important key or hub metabotypes of each module were identified by three parameters: MM, MS and connectivity. The importance of each metabotype for its module (MM) is quantified by the correlation between MEs and omics expression profiles. The MS values correspond to the Pearson correlation coefficients between metabotypes and response traits (meat quality and carcass composition traits). The intramodular soft connectivity was defined as

$$K_i = \sum_{u \neq i} a_{iu}$$

which is the sum of all pairwise adjacencies of a metabotype to all other metabotypes, a_{iu}, in the module. Within each module the metabotypes could be ranked using the absolute values of both MS, MM and connectivity to identify the key players, or so-called hub metabotypes, of a metabolic network.

According to these indicators, all module members were ranked according to their importance and stored into a sorted MM, MS and connectivity list. All metabotypes which were listed in the upper quartile of at least one of the ranking lists were used as phenotypes in the eQTL analyses described above. These metabotypes are regarded as key members of the underlying metabolic process. In the majority of cases, transcripts were the most important metabotypes of the modules.

For each important metabotype an eQTL analysis was performed which was based on the EIGENSTRAT approach with ten PCs to remove possible population stratification. Based on a genome-wide FDR value of $q \leq 0.05$ it was possible to identify significant eQTL. After eQTL analysis, corresponding genes could be identified by database query in the Ensembl genome browser 84 (www.ensembl.org).

In the detected eQTL it can be differentiated between regions, which are located close to a gene (*cis*-regulation) or distant (*trans*-regulation) [128]. *Cis*-regulated eQTL are more likely to represent the causative genomic region, whereas *trans*-regulated eQTL represent the ‘effect’, e.g. pathways that are affected by causal variations [131]. Although the most significant reported eQTL are often *cis*-regulated, there are some evidence that trans-regulated eQTL also might be decisive in controlling of gene expression [236].

4.3 Results

4.3.1 Weighted network analysis

The WNA allows investigating the entire data set using the profiles of metabolites, proteins and transcripts for the construction of a weighted co-expression network. The hierarchical clustering algorithm and the following pruning process condensed the metabotypes into 30 modules. Metabotypes that were not assigned to any module (n=112), were labelled with color grey.

The relationship between meat quality, carcass composition traits and modules is given by correlation coefficients between traits and MEs (Fig. 20). For the further investigation we selected modules that showed significantly ($p \leq 0.05$) strong ($r \geq |0.25|$) correlation coefficients to one or several meat quality or carcass composition traits (Tab. 18). Drip loss, meat color, BFT and LMC_{bonn} were significantly correlated only to one module, whereas pH24, pH1, MFR and LMC_{belly} were associated with three or four different modules. The number of metabotypes per module ranged between 1,683 (module purple) and 60 (module sienna3). Most of the metabotypes within the modules belong to the class of transcripts. However, in modules sienna3 and darkolivegreen the number of metabolites and transcripts is almost balanced and module white even comprises mainly metabolites (Tab. 18).

Module purple that includes the most metabotypes is the only module, which comprises proteins (Tab. 18). While module salmon is positively correlated with MFR and negatively correlated with LMC, the correlation coefficients have reversed signs in modules darkmagenta and purple. In a similar way, adverse results can be observed for the relation between the MEs of different modules and the meat quality indicator pH24. Regarding this relationship, significant positive correlations were observed in modules purple, sienna3 and midnightblue, whereas a significant negative correlation was found in module darkgrey (Fig. 20).

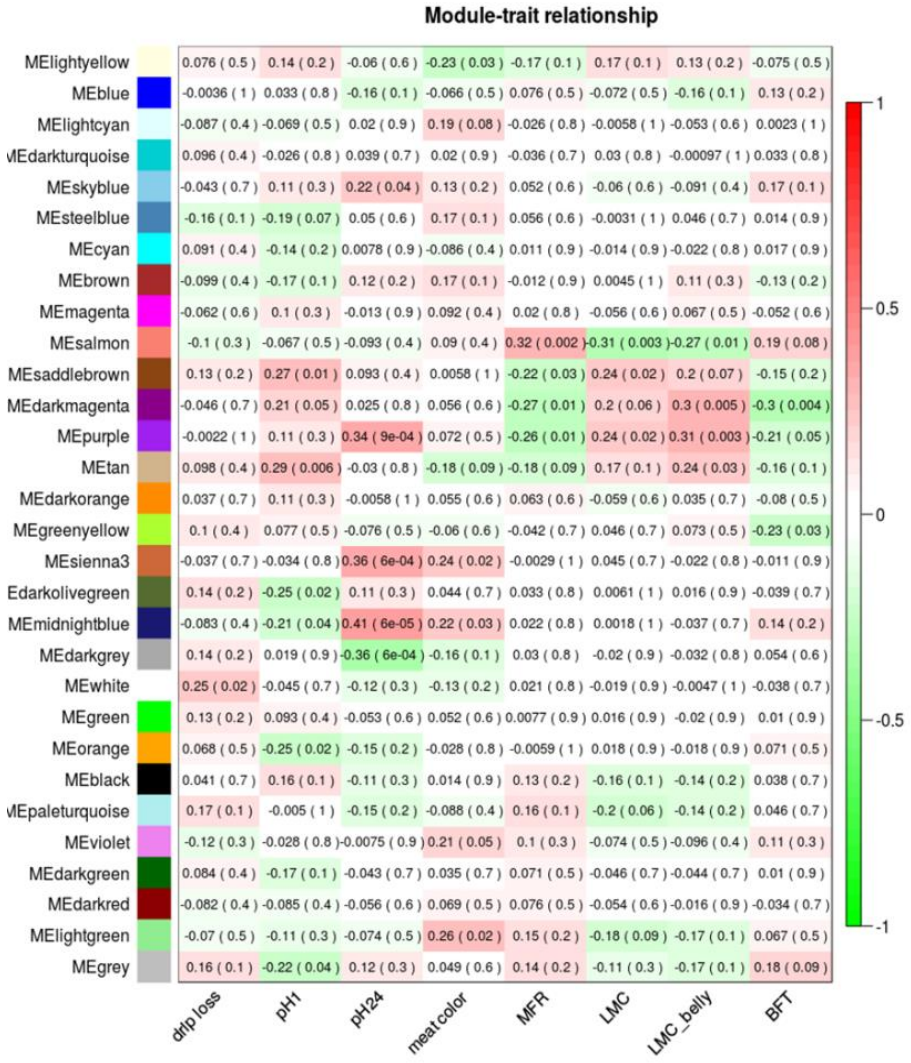


Figure 20: Correlation coefficients and corresponding p-values of module-trait relationship. Correlations of traits drip loss, pH1, pH24 and meat color to modules are characterised by color range from red (‘1’ - positive correlation) to green (‘-1’ - negative correlation). In parenthesis below correlation coefficients, the p-value is given. BFT – backfat thickness, LMC – lean meat content measured by formula of Grub in belly (LMC_belly) and by Bonner formula (LMC), MFR – meat fat ratio. ME = module eigenvalues.

Table 18: Composition of significantly correlated modules of weighted network analysis to meat quality and composition traits.

Module	No. transcripts	No. metabolites (annotated/ KEGG annotated)	No. proteins (entrez gene annotated)	∑	Correlated traits
salmon	385	-	-	385	MFR, LMC_{belly}, LMC_{bonn}
saddlebrown	112	2 (-)	-	114	pH1
darkmagenta	62	-	-	65	MFR, LMC_{belly}, BFT
purple	1,675	1 (-)	9 (9)	1,683	MFR, $LMC_{belly,}$ pH24
tan	1,442	-	-	1,442	pH1
sienna3	31	29 (6/-)	-	60	pH24
darkolivegreen	39	47 (14/8)	-	86	pH1
midnightblue	317	-	-	317	pH24
darkgrey	158	2 (1/-)	-	160	pH24
white	46	83 (25/11)	-	129	drip loss
lightgreen	247	1 (-)	-	248	meat color

BFT – backfat thickness, LMC – lean meat content measured by formula of Grub in belly (LMC_{belly}) and by Bonner formula (LMC_{bonn}), MFR – meat fat ratio.

4.3.2 Module annotation

For each module three to 22 significant ($p \leq 0.05$) KEGG pathways were detected. Tab. 19 shows a maximum of five most significant KEGG pathways with highest number of metabotypes in the module that can be assigning to the KEGG pathway (count). Although many of the detected KEGG pathways describe diseases and pathways related to metabolic defects (e.g. Pathways in cancer, Type I diabetes mellitus, Parkinson's disease), the annotation enabled the allocation of the modules to metabolic pathways and molecular functions correlated with muscle growth, muscle to meat conversion and meat ageing and quality. For example KEGG pathways ‘Protein processing in endoplasmic reticulum’, ‘Metabolic pathway’, ‘Oxidative phosphorylation’ and ‘Citrate cycle (TCA cycle)’ are strongly associated with energy metabolism and anabolic processes effecting muscle and fat tissue growth [131]. Some modules are annotated with the identical KEGG IDs, indicating that module connected metabolic processes, like ‘MAPK signalling pathway’, ‘Dilated

cardiomyopathy',' Antigen processing and presentation' and 'Protein processing in endoplasmic reticulum' are likely to play an important role in effecting meat quality and carcass composition traits. Beneath the annotation with KEGG terms, we also assigned the modules to related Gene Ontologies (GO) terms 'Biological Process (BP)', 'Cellular Component (CC)' and 'Molecular Function (MF)' (Results not shown).

Table 19: Functional annotation of the 11 significantly correlated modules by selection of most important KEGG identifiers with highest count. Continued on next page.

Module	KEGG	p-value	Count	Size	KEGG term
salmon	5200	2.15E-03	11	162	Pathways in cancer
	4722	1.57E-05	9	61	Neurotrophin signalling pathway
	4010	4.27E-03	8	104	MAPK signalling pathway
	5220	3.41E-05	7	38	Chronic myeloid leukaemia
	4630	5.12E-03	6	64	Jak-STAT signalling pathway
saddle-brown	5414	4.88E-07	6	47	Dilated cardiomyopathy
	5410	1.47E-05	5	48	Hypertrophic cardiomyopathy (HCM)
	4020	1.20E-03	4	69	Calcium signalling pathway
	4260	3.65E-03	3	45	Cardiac muscle contraction
	4530	5.21E-03	3	51	Tight junction
dark-magenta	4612	2.69E-12	9	41	Antigen processing and presentation
	5332	9.78E-09	6	24	Graft-versus-host disease
	5330	4.20E-08	6	30	Allograft rejection
	4940	6.35E-08	6	32	Type I diabetes mellitus
	5320	6.35E-08	6	32	Autoimmune thyroid disease
purple	4810	3.38E-04	24	96	Regulation of actin cytoskeleton
	4145	3.17E-03	20	87	Phagosome
	4142	4.71E-04	19	70	Lysosome
	5140	5.93E-04	15	50	Leishmaniosis
	5414	3.09E-03	13	47	Dilated cardiomyopathy
tan	4510	6.46E-05	21	98	Focal adhesion
	4810	1.22E-03	18	96	Regulation of actin cytoskeleton
	4141	1.53E-03	17	90	Protein processing in endoplasmic reticulum
	4512	6.15E-07	16	46	ECM-receptor interaction
	4670	9.97E-03	11	58	Leukocyte transendothelial migration
sienna3	4662	1.24E-02	1	31	B cell receptor signalling pathway
	4664	1.28E-02	1	32	Fc epsilon RI signalling pathway
	4666	1.71E-02	1	43	Fc gamma R-mediated phagocytosis

Table 19 continued.

Module	KEGG	p-value	Count	Size	KEGG term
sienna3	4380	2.59E-02	1	65	Osteoclast differentiation
	4650	1.91E-02	1	48	Natural killer cell mediated cytotoxicity
darkolive-green	310	3.54E-02	1	18	Lysine degradation
	20	3.73E-02	1	19	Citrate cycle (TCA cycle)
	380	3.93E-02	1	20	Tryptophan metabolism
midnight-blue	4141	2.17E-03	7	90	Protein processing in endoplasmic reticulum
	4010	1.90E-02	6	104	MAPK signalling pathway
	4612	1.34E-03	5	41	Antigen processing and presentation
	4920	3.01E-03	5	49	Adipocytokine signalling pathway
	3040	1.22E-02	5	68	Spliceosome
darkgrey	4740	4.42E-03	2	10	Olfactory transduction
	4140	1.43E-02	2	18	Regulation of autophagy
	4970	4.23E-02	2	32	Salivary secretion
	460	3.08E-02	1	3	Cyanoamino acid metabolism
	232	4.08E-02	1	4	Caffeine metabolism
white	670	1.90E-02	1	8	One carbon pool by folate
	350	3.30E-02	1	14	Tyrosine metabolism
	51	4.00E-02	1	17	Fructose and mannose metabolism
light-green	1100	8.70E-13	43	569	Metabolic pathways
	5012	2.79E-23	27	88	Parkinson's disease
	5016	9.66E-22	27	99	Huntington's disease
	190	4.20E-19	24	89	Oxidative phosphorylation
	5010	2.79E-16	23	103	Alzheimer's disease

Count - number of metabotypes in the module that can be assigning to the KEGG pathway; size - number of all associated metabotypes of the KEGG pathway; KEGG database information.

The annotation procedure performed in the modules darkolivegreen, white and purple and their corresponding KEGG annotated metabolites and entrez gene annotated proteins did not reveal any significant ($p \leq 0.05$) KEGG pathway annotations.

4.3.3 GWAS of meat quality and carcass composition traits and significantly correlated modules

GWAS were performed for a total of 19 traits (four meat quality traits, four meat composition traits and the eigenvalues of 11 modules) as dependent variables and 45,616 SNPs, which passed the SNP quality control filter, as independent genetic variables. In a first step, the data was tested for possible population stratification. Based on the GWAS results of model 1 the parameter λ was calculated. In case of meat quality parameters and modules, the parameter λ did not fall into the optimal range 1.00 to 1.05 given by Price et al. [193]. Consequently, the EIGENSTRAT method was used to correct for population stratification. The corresponding statistical GWAS models include a variable number of PCs for each trait. As it has been pointed out by Welzenbach et al. [144], the PCs condense the genetic relationship between animals and can be used to remove undesirable population stratification. Depending on the investigated trait between one to 24 PCs were required in order to avoid negative effects of population stratification on the validity of the GWAS (Tab. 20). After this correction step, for all investigated traits the genomic inflation factor λ was one or close to one. Therewith, due to Price et al. [193], our correction was sufficient to remove disturbing population stratification (Tab. 20).

Applying a moderate significance threshold with a FDR of $q \leq 0.10$, the GWAS revealed 874 (without double counting) significant associations for 17 traits, including the meat quality and carcass composition traits and nine modules (Tab. 20). These SNPs were distributed over almost all porcine chromosomes. For modules purple and darkolivegreen it was not able to determine significant SNPs under the defined conditions. While 151 SNPs showed a chromosome-wide significance levels of $q \leq 0.05$, eight SNPs were detected as genome-wide significant (at least $q \leq 0.01$). The average number of significant SNPs per trait is 41, with a minimum of four SNPs (BFT, sienna3) and a maximum of 208 SNPs for lightgreen. The majority of the SNPs are significant at a moderate chromosome-wide level ($q \leq 0.1$). For LMC_{belly} and LMC_{bonn}, four genome-wide significant SNPs were identified, respectively. SNPs for modules lightgreen and darkmagenta showed highest proportions of explained variance (var=19.96% and 19.38%). The average number of detected SNPs per chromosome is 43 and the highest numbers of significant SNPs were detected on SSC7, 13, 15 and 18. For drip loss, we detected SNPs on SSC2 and 16, which explain a maximum of variance proportion of 8.83%.

On several chromosomes we identified SNPs significantly associated with more than one trait. In total, 17 overlapping (identical) SNPs for different traits are located on SSC5, 6, 8, 12, 15 and 18. As presented in Fig. 21, the most overlapping exists between modules midnightblue and tan on SSC18. Moreover, the overlapping on SSC5, 6 and 12 is of particular interest because one SNP significantly affects both LMC_{belly}, LMC_{bonn}, BFT and MFR. The overlapping SNPs appear to be central genetic markers for carcass composition traits and related physiological processes. In contrast, on SSC4, 7 and 17 we indeed detected significant SNPs for two or more traits but the markers are located in distant chromosomal regions and according to this, there are no overlapping SNPs.

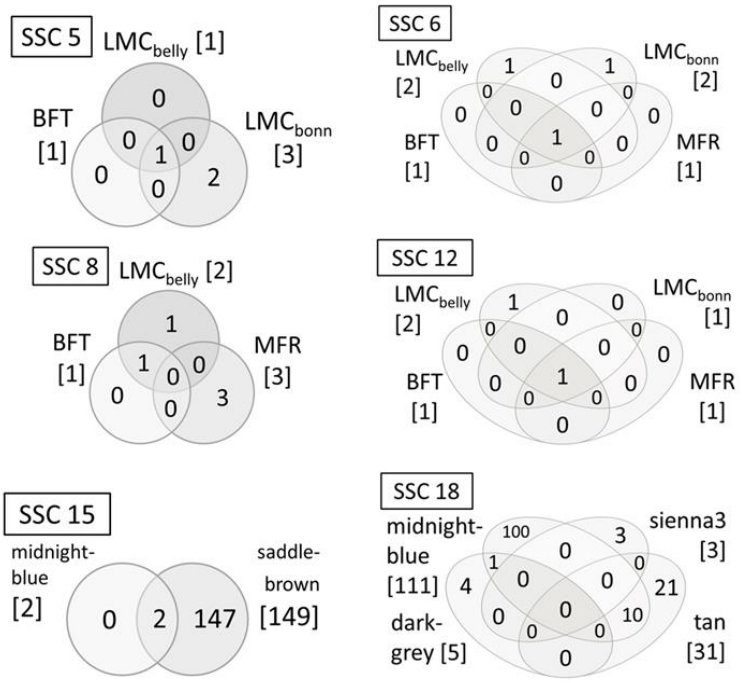


Figure 21: Venn-diagrams of overlapping SNPs on *Sus scrofa* chromosomes (SSC) 5, 6, 8, 12, 15 and 16. GWAS resulted in varying number of significant SNP per trait,. On the presented chromosomes there are overlapping SNPs with meaning for two or more traits. The total number of detected SNPs on the chromosome is given in square bracket.

Table 20: Results of association analyses. Continued on next page.

Trait[1]	PC[2]	Λ[3]	number of significant SNP per porcine chromosome[4]																	∑SNP[5]	min p-value[6]	min q-value[7]	max σ^2_y[8]
			1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18				
drip loss	9	1.00		16													3			*2; #17	4.39E-05	3.64E-02	8.83
pH1	1	0.99	34																	*2; #32	1.14E-05	4.38E-02	10.01
pH24	4	1.00							15											#15	7.73E-05	5.56E-02	8.63
color	1	0.99							4		4									*2; #6	2.98E-05	4.40E-02	9.15
BFT	-	0.99					1	1		1			1							*2; #2	2.95E-06	7.27E-03	7.11
MFR	-	1.00					2							187						*2; #187	2.31E-05	2.06E-02	10.11
LMC_{belly}	-	1.00				1	1	2		2			2					4		***3;**1; *1; #7	2.29E-09	6.63E-06	10.20
LMC_{bonn}	-	0.99	1		3		3	2				2	5							***4;*3; #9	2.93E-16	8.49E-13	12.82
darkmagenta	9	1.00	3														1	1		*3; #2	7.33E-06	2.07E-02	18.96
darkgrey	16	1.00			26														5	*3; #28	9.56E-06	2.18E-02	19.38
lightgreen	4	1.00				3			205											*5; #203	4.91E-05	3.59E-02	19.96
salmon	15	1.00					9													#9	5.83E-05	5.89E-02	15.78
tan	24	1.00							1						12				31	*12; #32	1.51E-05	1.72E-02	17.61
white	4	1.00							27											*10; #17	2.12E-05	2.88E-02	17.23
midnightblue	16	1.00														2			111	*37, #77	2.28E-05	2.26E-02	17.14
saddlebrown	19	1.00														149				*63, #86	6.47E-06	1.12E-02	19.01

Table 20: Continued.

Trait[1]	PC[2]	Λ[3]	number of significant SNP per porcine chromosome[4]																	∑SNP[5]	min p-value[6]	min q-value[7]	max σ^2_y[8]
			1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18				
sienna3	13	1.00																1	3	*4	6.71E-05	3.96E-02	15.31
∑SNP excluding double counting			38	16	29	4	14	3	252	5	4	4	6	187	12	149	4	8	139				
∑ overlapping SNP[9]			-	-	-	-	1	1	-	1	-	-	1	-	-	2	-	-	11				

[1]BFT – backfat thickness, LMC – lean meat content measured by formula of Grub in belly (LMC_{belly}) and by Bonner formula (LMC^{bonn}), MFR – meat fat ratio, modules estimated by weighted network analysis are indicated by colors; [2]number of principal components (PCs) considered in genome-wide association studies for meat quality traits and modules (EIGENSTRAT approach); [3]λ = inflation factor; [4]number of chromosome-wide significant associated SNPs per traits and chromosome (at least $q \leq 0.1$); [5]sum of significant associated SNPs per traits (***$q \leq 0.001$; **$q \leq 0.01$; *$q \leq 0.05$; #$q \leq 0.1$); [6]minimal empirical p-value; [7]minimal q-value, based in the false discovery rate concept; [8]maximal proportion of explained variance [%]; [9]sum of overlapping SNP with meaning for more than one trait.

The functional annotation of the 874 significantly associated SNPs revealed 1,631 genes that are located in a distance of ≤1Mb to the SNPs and thereby potentially are in linkage disequilibrium to the significant SNPs of the present study. One quarter of the detected SNPs (n=145) is localised in an intron region of a specific gene. These genes that are mainly located on SSC7, 13 and 18 might be important potential candidate genes for meat quality and carcass composition traits and associated metabolic processes.

For the identification of potential candidate genes, we concentrated on the most important QTL regions with a high density of significant SNPs for different traits. These SNPs were selected based on the following three criteria: The SNPs had to be (1) chromosome-wide significant (at least q≤0.1), (2) within the 'Top 10' of significant SNPs for meat quality/carcass composition/modules, (3) exonic or intronic. Using these criteria, we identified 28 potential candidate genes for all investigated meat quality and carcass composition traits and for seven of the nine modules (Tab. 21). The SNPs indicating the candidate genes are intronic with exception of three SNPs that are located within the 3' untranslated region (UTR) of a gene. In the 3' UTRs may be located different regulatory sequences. These sequences can be binding sites for miRNAs or proteins that affect the stability or transport of mRNA. The number of detected genes for a single trait varied between one and three. On SSC1, three genes in a range of 0.5Mb were detected for pH1. In the most cases the importance of a candidate gene is indicated by one significant intronic SNP. Only candidate gene LGR4 on SSC2 harbours two intronic SNPs, which emphasizes the meaning of this candidate gene for drip loss. SSC5 is of particular interest, because on this chromosome we identified one candidate gene (BTG1) for BFT, LMC_{belly} and LMC_{bonn} and three candidate genes for module salmon that is strongly correlated with carcass composition traits. One candidate gene for module salmon (EEA1) is located in close proximity with BTG1. Based on these two candidate genes in a small range of 0.62Mb, it might be reasonably assumed that in the region around 94.5Mb to 95.0Mb, a functional mutation for carcass compositions traits or related metabolic processes is located. Moreover, SNPs 'DRGA0006183' (in gene BTG1) and 'ALGA0031253' (in gene ESYT1) are located within a 3' UTR that are characterised by their regulatory functions. This is further evidence for the meaning of BTG1 and ESYT1 as candidate genes for the regarding traits. Most significant intronic SNPs with highest proportion of explained variance (17.74% and 17.14%) were detected for modules darkmagenta (SSC17) and midnightblue (SSC18) (Tab. 21).

For drip loss, two candidate genes were identified on SSC2 (Tab. 21). The most significant SNPs (Var_{max}=8.83 %; $p_{min} \leq$4.39E-05) associated with drip loss were detected on SSC16, but these SNPs do not fulfil the previously described conditions to detect potential candidate genes (Tab. 21). On SSC2, the three SNPs associated with drip loss have a proportion of explained variation of 7.30% and two SNPs with a distance of 32k base pairs are located in gene LGR4.

4.3.4 Selection of hub metabotypes and eQTL analysis

For seven of the 11 investigated modules, GWAS revealed significant genetic markers and candidate genes (Tab. 21). As a verification of our GWAS results, we analysed the hub metabotypes of the corresponding modules by means of an eQTL analysis. Within each module, based on the parameters MS, MM and connectivity, between one and 20 metabotypes, which are obviously key regulators/members of the regarding module/metabolic relation, were selected. Altogether, we performed eQTL analyses for 113 hub metabotypes. In most modules, the selected hub metabotypes were transcripts. Only in modules white and sienna3 the most important members of the metabolic process pertain to the class of metabolites. In module white the most important hub metabotype is the metabolite 1-octadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine, which is the only selected annotated metabolite. This metabolite belongs to the class of glycerophospholipids. In case of 71 of the 113 selected hub metabotypes it was possible to identify significant eQTL (Tab. 22). For the most transcript hub metabotypes, annotation information was available and it was possible to attach the corresponding gene that is affected by the transcript. See Tab. 22 for the names of the associated genes and their location. For metabolites in modules white and sienna3 and some transcript of the other modules, assignment of corresponding genes was not possible. The average number of eQTL for each metabotype was seven, with a range of one to 113 eQTL. In case of 14 metabotypes, the performed eQTL analyses resulted in only one eQTL (Tab. 22). In case of seven transcripts, the associated eQTL are located on the same chromosome as the corresponding gene associated with the transcript. Although there is no positional accordance of eQTL and underlying ‘causative’ transcript of the gene (*cis*-regulated), the eQTL are *trans*-regulated and it might be that these eQTL are involved in the control of gene expression [236].

Table 21: Annotation of most promising SNPs for meat quality and carcass composition traits and associated modules. Continued on next page.

SSC[1]	Trait	Gene[2]	SNP	Position[3] (location)	Mut[4]	eEff5 (se)	Chi2	emp. p-value[6]	q-value[7]	Var[8]
1	dark-magenta	LMBRD1	INRA0002232	56104185 (intr)	G/A	4.39E-03 (1.07E-03)	16.75	5.22E-05	8.73E-02	15.99
	pH1	ITPKA	ASGA0004539	144941329 (intr)	C/A	-1.48E-02 (3.48E-02)	18.17	2.02E-05	4.38E-02	9.51
		PPP1R14D	ALGA0005986	145646235 (intr)	G/A	-9.61E-03 (2.48E-03)	15.05	1.05E-04	7.53E-02	8.00
		RMDN3	ALGA0005991	145736477 (intr)	G/A	-9.92E-03 (2.55E-03)	15.15	9.95E-05	7.53E-02	8.05
2	drip loss	CD44	ASGA0104263	28330606 (intr)	G/A	4.11E-01 (1.11E-01)	13.63	2.28E-04	6.11E-02	7.30
		LGR4	ALGA0012966	35471499 (intr)	A/G	4.11E-01 (1.11E-01)	13.63	2.28E-04	6.11E-02	7.30
			ALGA0012967	35503490 (intr)	A/G	4.11E-01 (1.11E-01)	13.63	2.28E-04	6.11E-02	7.30
5	salmon	ESYT1	ALGA0031253	23002609 (3' UTR)	A/G	-3.66E-03 (9.01E-04)	16.49	5.83E-05	5.89E-02	15.78
		NTN4	INRA0020248	91953489 (intr)	T/C	-3.39E-03 (9.28E-04)	13.35	2.99E-04	6.95E-02	13.17
		EEA1	H3GA0017095	94466720 (intr)	C/T	-3.04E-03 (8.32E-04)	13.36	2.97E-04	6.95E-02	13.18
	BFT	BTG1	DRGA0006183	95090653 (3' UTR)	T/G	2.05E-01 (4.73E-02)	10.31	3.73E-05	7.37E02	5.63
	LMC_{belly}					-3.00E-01 (7.57E-02)	10.10	9.71E-08	1.79E-04	8.30
	LMC_{born}					-4.15E-06 (9.69E-07)	18.29	4.10E-12	8.55E-09	9.56
7	pH24	CDKAL1	ASGA0031511	17472981 (intr)	G/A	4.28E-03 (1.11E-03)	14.80	1.69E-04	5.56E-02	7.88
	white	DST	DRGA0007481	33588453 (intr)	T/C	-9.61E-03 (2.30E-03)	17.40	3.41E-05	2.88E-02	16.51
	lightgreen	PTPN21	DRGA0008163	116876922 (intr)	C/T	-5.24E-03 (1.28E-03)	16.72	4.91E-05	3.59E-02	15.96
		TTC8	DRGA0008172	117224903 (intr)	A/G	-5.67E-03 (1.44E-03)	15.61	8.77E-05	3.59E-02	15.06
		TRIP11	ALGA0044982	120176020 (intr)	G/A	-9.21E-03 (2.39E-03)	14.89	1.27E-04	4.18E-02	14.47
8	color	UBE2T	SIRI0001503	28853192 (intr)	G/A	9.72E-01 (2.70E-01)	12.98	3.15E-04	8.81E-02	6.98
		PPP1R12B	ALGA0057938	28938242 (intr)	T/C	9.72E-01 (2.70E-01)	12.98	3.15E-04	8.80E-02	6.98

Table 21: Continued.

SSC[1]	Trait	Gene[2]	SNP	Position[3] (location)		Mut[4]	eEff5 (se)		Chi2	emp. p-value[6]	q-value[7]	Var[8]
12	LMC_{belly}	KRT15	DIAS0003753	21492055	(intr)	G/A	-4.29E-01	(1.50E-01)	8.21	1.12E-04	6.99E-02	4.53
13	MFR	TAMM41	ALGA0070793	75170684	(intr)	G/A	-1.53E-08	(5.23E-09)	8.54	5.06E-03	8.45E-02	4.70
		ACAD11	ALGA0070916	81362489	(intr)	T/C	-7.78E-09	(2.35E-09)	10.94	1.50E-03	8.46E-02	5.95
		BFSP2	ALGA0070925	82085380	(intr)	A/G	-7.78E-09	(2.35E-09)	10.94	1.50E-03	8.46E-02	5.95
17	sienna3	FAT1	DRGA0016533	9400936	(intr)	T/G	6.95E-03	(1.74E-03)	15.90	6.71E-05	9.97E-02	15.31
	LMC_{belly}	XRN2	DRGA0016692	33090432	(intr)	G/A	-1.81E-01	(5.69E-01)	10.10	1.84E-05	1.17E-02	5.51
	dark-magenta	RAD21L1	ALGA0110267	38603020	(intr)	C/A	-5.54E-03	(1.27E-03)	18.98	1.66E-05	2.47E-02	17.74
18	midnight-blue	GIMAP2	SIRI0000131	6733601	(intr)	C/T	-5.50E-03	(1.29E-03)	18.20	2.28E-05	2.26E-02	17.14
		TMEM213	ALGA0097013	11325435	(3' UTR)	C/A	-5.11E-03	(1.38E-03)	13.67	2.43E-04	2.85E-02	13.44
		GRM8	ALGA0097418	23067778	(intr)	G/A	3.78E-03	(9.56E-04)	15.64	8.65E-05	2.85E-02	15.09
	tan	BMPER	H3GA0050799	43325649	(intr)	T/C	2.56E-03	(7.19E-04)	12.68	3.79E-04	4.44E-02	12.60

The SNP order complies with number of chromosomes and position on the chromosome; Selection of promising SNPs based on the criteria, that they are (1) chromosome-wide significant (at least $q<0.1$), (2) within the 'Top10' significant SNPs per metabolic trait and (3) located within an annotated gene (exonic or intronic); [1]sus scrofa chromosomes (SSC); [2]The declaration of gene symbols can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/gene; [3]position in Mb (base pairs), intr – intronic, 3'UTR - 3' untranslated region; [4]mutation (Mut); [5]substitution effect and standard error (se); [6]empirical p-value; [7]q-value (based on the false discovery rate concept); [8]proportion of the explained variation (Var, %).

Table 22: Results of eQTL-analysis of hub metabotypes of the modules. Continued on next page.

Module	Hub metabotype[1]	Gene[2]	(SSC)[3]	No. eQTL[4]	(SSC)[5]
dark-magenta	Ssc.13780.4.S1_x_at	-	7	1	16
darkgrey	Ssc.3282.1.S1_at	NR2F1	2	2	9
	Ssc.4973.1.S1_at	-	15	3	3, 4
	Ssc.8075.2.S1_at	TMEM88B	6	3	6, 13
	Ssc.10751.1.A1_at	-	2	1	18
	Ssc.12151.1.A1_at	-	14	6	2, 6, 18
	Ssc.15905.1.A1_at	-	5	2	1
	Ssc.17792.1.S1_at	-	7	2	2, 18
	Ssc.17797.1.A1_at	-	7	12	2, 6
	Ssc.28050.1.A1_at	-	7, 18	1	4
	Ssc.29484.1.A1_at	PIAS4-201	2	23	2, 9, 18
lightgreen	Ssc.886.1.S1_at	-	4	35	7
	Ssc.1092.1.A1_at	COX5B	3	3	7
	Ssc.1725.1.S1_at	SLC25A12	15	6	5
	Ssc.2294.1.S1_a_at	COQ6	7	113	7
	Ssc.5035.1.S1_at	-	7, 15	9	3, 7
	Ssc.5334.1.S1_at	COQ9	6	1	7
	Ssc.5389.1.S1_at	IDH3A	7	17	2
	Ssc.5790.1.S1_at	UQCRC1	13	4	7, 9
	Ssc.7201.1.A1_at	NDUFS2	4	7	2, 5
	Ssc.9742.1.S1_s_at	COX5A	7	8	7
	Ssc.10949.1.S1_at	-	3	1	7
	Ssc.10953.1.A1_at	NDUFS1	15	1	7
	Ssc.14462.1.S1_a_at	CKMT2	2	2	5, 14
	Ssc.15103.1.S1_at	NDUFS6	16	7	7
salmon	Ssc.1584.1.A1_at	MEF2D	4	2	9
	Ssc.4086.1.S1_at	YWHAG	3	3	13
	Ssc.5825.1.S1_at	RPRD1B	17	1	4
	Ssc.6483.1.S1_at	-	14	5	2, 6
	Ssc.11541.1.A1_at	-	8	3	1, 9
	Ssc.12247.1.A1_at	-	6	2	8
	Ssc.15611.1.S1_at	-	5	1	8
	Ssc.16047.2.S1_at	STAT5A	12	4	10, 14
	Ssc.18457.1.S1_at	DGCR2	14	2	1, 10
	Ssc.23514.1.A1_at	-	2	11	3, 4, 10
	Ssc.23986.1.S1_at	-	1	1	8
tan	Ssc.1228.1.S1_at	MARCKS	1	3	5, 16, 17
	Ssc.4141.1.A1_at	-	11	3	5, 17
	Ssc.5524.1.S1_at	-	9	1	2
	Ssc.7106.1.S1_at	CDO1	2	3	18
	Ssc.7117.1.A1_a_at	CHODL	13	2	5

Table 22 continued.

Module	Hub metabotype[1]	Gene[2]	(SSC)[3]	No. eQTL[4]	(SSC)[5]
tan	Ssc.7756.1.A1_at	-	4	19	18
	Ssc.8609.1.A1_at	PDCD4	14	2	9, 13
	Ssc.22086.1.A1_at	MAGED1	X	2	8, 9
	Ssc.28502.1.S1_at	-	2	17	8, 9, 12
white	1-octadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine	-	-	23	7
	X314.986.8.33	-	-	1	1
	X392.9623.8.33	-	-	27	7
	X734.5549.3.21	-	-	16	7
	X825.6364.3.05	-	-	51	7
	X919.5007.1.99	-	-	5	1, 8
midnight-blue	Ssc.4121.1.S1_at	-	3	3	18
	Ssc.6354.1.A1_at	TGIF1	6	2	18
	Ssc.7957.1.A1_at	-	17	1	18
	Ssc.9439.1.A1_a_at	FUBP1	6	16	18
	Ssc.13649.1.S1_at	BAG3	14	9	2, 18
	Ssc.14224.1.A1_at	MAFF	5	2	4
	Ssc.17309.1.S1_at	HSPA4	2	1	18
	Ssc.17671.1.S1_at	-	17	15	8
	Ssc.25077.1.S1_at	-	Y,17	1	2
	Ssc.25783.1.S1_at	WDR26	10	1	5
	Ssc.26179.1.S1_at	MIDN	2	2	2, 11
saddle-brown	Ssc.14406.1.A1_at	MRPS6	13	1	18
sienna3	X415.3296.3.6	-	-	7	2, 6, 17
	X425.3514.3.11	-	-	1	13
	X427.366.3.11	-	-	1	3
	X439.3289.3.5	-	-	6	2
	X449.3499.3.08	-	-	8	2
	X451.366.3.08	-	-	10	2, 8
	X479.3966.3.03	-	-	29	2

[1]Selected hub metabotypes of each module based on the parameters module membership, metabotypes significance and connectivity; [2]corresponding gene of the transcript. In case of metabolites as module specific hub metabotypes, it was not possible to assign corresponding genes; [3]*Sus scrofa* chromosome (SSC) on which the gene is located; [4]number of expression quantitative trait loci (eQTL) that are associated with the hub metabotypes; [5]location of the eQTL.

4.3.5 Verification of candidate genes by eQTL analysis

For further validation purposes, the positions of significant SNPs, which result from the different GWAS were compared with the results of the eQTL analyses. As dependent variables, the GWAS comprise various dependent phenotypes (module eigenvalues, meat quality or carcass composition traits), whereas the eQTL analyses include expression of several module hub metabotypes. The validity is increased if the position of the candidate genes found in GWAS and eQTL analyses are identical or at least within the same chromosomal location (<1Mb distance).

As an example, where the results of GWAS and eQTL analysis were not consistent, the following section comprises the findings linked to module darkmagenta. Darkmagenta is related with MFR and LMC_{belly} (see Tab. 18) and is significantly influenced by the SNPs 'INRA0002232' (SSC1) and 'ALGA0110267' (SSC17). These SNPs are intronic parts of the candidate genes LMBRD1 and RAD21L1, respectively. For darkmagenta, the most important (one out of 62) transcript 'Ssc.13780.4.S1_x_at' (Tab. 22) was selected as a hub metabotype and expression phenotype in the eQTL analysis (Tab. 18, 22). The SNP 'ASGA0074268' located on SSC16 was detected as a significant eQTL. But in contrary to our expectations, this eQTL or rather neighboured SNPs were not consistent with the significant SNPs identified by the GWAS of module darkmagenta.

Examples, where GWAS and eQTL analyses show consistent results, are listed in Tab. 23. In total, eight SNPs have been detected as significant within the GWAS as well as in the eQTL analyses. A promising result is the significant SNP 'DRGA0007481' located on SSC7, which is an intronic part of the candidate gene DST. This marker is significant for module white, which is correlated with drip loss. Four (out of six) of the most important hub metabotypes for module white (1-octadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine, X392.9623.8.33, X734.5549.3.21, X825.6364.3.05) are also significantly influenced by SNP 'DRGA0007481'. This result underlines the importance of this SNP and the linked DST gene for drip loss. Other examples of promising candidate genes where GWAS findings were confirmed by eQTL results are: a) PTPN21, TTC8 and TRIP which are relevant for the phenotype meat color, b) GIMAP2, TMEM213 and GRM8 relevant for module midnightblue and c) BMPER relevant for module tan.

Table 23: Accordance of SNPs identified for a module correlated with a meat quality trait and eQTL identified for different hub metabotypes of the module. Some genetic markers are both significant SNP for a module and likewise eQTL for hub metabotypes of the regarding module.

Trait	GWAS				eQTL analysis		
	Module	Candidate gene	SNP	SSC	Selected hub metabotypes of the modules	No. eQTL	Consistent SNP/eQTL (q-value)
drip loss	white	DST	DRGA0007481	7	glycero-3-phosphoserine	23	DRGA0007481 (0.014)
					X392.9623.8.33	27	DRGA0007481 (0.041)
					X734.5549.3.21	16	DRGA0007481 (0.044)
					X825.6364.3.05	51	DRGA0007481 (0.009)
color	lightgreen	PTPN21	DRGA0008163	7	Ssc.886.1.S1_at	37	DRGA0008163, DRGA0008172, ALGA0044982
					Ssc.2294.1.S1_a_at	113	DRGA0008172, ALGA0044982
		TTC8	DRGA0008172		Ssc.5035.1.S1_at	9	DRGA0008163, DRGA0008172, ALGA0044982
					Ssc.7201.1.A1_at	7	DRGA0008172
		TRIP	ALGA0044982		Ssc.9742.1.S1_s_at	8	DRGA0008163, DRGA0008172, ALGA0044982
					Ssc.15103.1.S1_at	7	DRGA0008163, DRGA0008172, ALGA0044982
pH24	midnight-blue	GIMAP2	SIRI0000131	18	Ssc.4121.1.S1_at	3	SIRI0000131, ALGA0097418
					Ssc.6354.1.A1_at	2	SIRI0000131
		TMEM213	ALGA0097013		Ssc.9439.1.A1_a_at	16	SIRI0000131, ALGA0097013, ALGA0097418
					Ssc.13649.1.S1_at	9	SIRI0000131, ALGA0097418
		GRM8	ALGA0097418		Ssc.17309.1.S1_at	1	ALGA0097418
pH1	tan	BMPER	H3GA0050799	18	Ssc.7756.1.A1_at	19	H3GA0050799

SSC– *Sus scrofa* chromosome.

4.4 Discussion

Systems biology approach: An integrated analysis of genome, proteome, transcriptome and metabolome to elucidate the 'muscle to meat' black box

Meat quality traits are affected by complex interactions of diverse metabolic processes that are controlled by several genetic mechanisms. However, there is still a knowledge gap regarding the biochemical processes and the interaction of metabolic components, which are involved in the muscle to meat conversion. Therefore, the genetic background of meat quality is not completely understood. Furthermore, a variety of post-transcriptional events and metabolic regulation processes, responding very quickly to environmental changes, complicate the elucidation of the relationship between genes and phenotypes. In this study, we used a systems genetics approach to describe the general architecture of production traits. To uncover genetic variation, we jointly analysed the five underlying omics levels: genome based on polymorphism data, transcriptome, proteome, metabolome and phenome [49]. While the genome (SNP information) contains the information of existing allele variants, the other omics levels indicate, which genes are actually being expressed and which pathways are active. Metabolites, proteins and transcripts are essential links between genetic information and phenotypic expression. Metabolic traits can be used as more accurate phenotypes compared to the conventional performance traits in genetic association studies. Thus, the statistical power in the identification of candidate genes is improved and less false positive as well as less redundant results are observed [118]. Ala-Korpela et al. [195], Kadarmideen [53] and Widmann et al. [147] already stated that systems biology approaches are valuable and powerful in identifying key-causal and highly predictive genetic variants for complex traits and allow building up complex genetic regulatory networks.

Benefit of network analysis in elucidation of metabolic background of meat quality

According to Fontanesi [110], network reconstruction methodologies that based on systems genetics concepts are useful to clarify the complexity of multi omics information and to identify the linking of omics components of different metabolic stages among themselves. Following this assessment [110], we applied a network analysis to handle our multi omics data set and to address the challenge of the 'large p, small n' situation by summarising a network of modules to reduce the complexity of a data set, which is thereby analysed with

greater statistical power [164]. Appling hierarchical clustering, sets of different metabotypes are assigned to modules and a subsequent database query allows the functional annotation of the modules to KEGG pathways and GO terms. Based on this, a module can be considered as a condensed metabolic phenotype. By choosing several module eigenvalues as intermediate independent phenotypes, the multi omics data set was adjusted for irrelevant and redundant variables and the procedure weakened the 'large p, small n' problem by decreasing the number of dependent variables (p). Thereby, focusing the network analysis on modules (and their highly connected intramodular hub metabotypes), contribute to a biologically meaningful data reduction scheme.

Based on the identification of modules using a hierarchical clustering process, it can be assumed that the metabotypes within a module are co-regulated and might correspond to the same biological pathways. This functional connection is useful for the annotation of unknown metabotypes based on their annotated neighbours within a network [183,237]. For example, the simple metabolite search of database METLIN gives some evidence that several non-annotated metabolites of module white also belong to the category of 'sphingolipids' just like the most important hub metabolite of module white (Tab. 22). On the other hand, the module annotation is only based on the annotated metabolic components. It can be expected that module annotation considering all quantified metabolic components would give a more precise view into the relevant metabolic background.

Another critical point in the functional annotation of the modules, estimated with connectivity-based approach WNA, is that the procedure does not allow overlapping metabolic components. This fact is contrary to the assumption that metabolic components are very likely actors in many different pathways [238]. In contrast to WNA, the independent component analysis (ICA) [239], a method of extraction of expression patterns in order to identify sets of co-regulated genes, allows overlapping genes between modules. For this reason, ICA is likely to better reflect biological reality than methods that partition genes into distinct clusters like WNA [238]. Moreover, the WNA calculated with R package WGCNA does not allow the illustration of causal connection between the metabotypes within a module. Although the parameters MM and connectivity give evidence about the importance of the single metabotypes, the procedure does not allow predictions whether component A affects component B or if A is the result of the metabolisation of B. The knowledge of causal relationships between the metabotypes is

essential for the elucidation of metabolic networks [240]. To get a deeper understanding of the causal relations and interactions of the metabolic components, further investigations are needed. For this purpose, it will probably be necessary to profile the metabolic components over a longer time frame in the individuals.

New insight in metabolic processes influencing drip loss

As indicated in Fig. 20, modules sienna3, darkolivegreen, midnightblue, darkgrey and white are predominantly significant correlated with one or more meat quality traits and show no relation to investigated carcass composition traits. Moreover, Tab. 18 demonstrates that these modules predominantly include metabolites in contrast to the other modules, which mainly comprise transcripts. In contrast to transcripts, metabolites represent final products of metabolic processes and thereby are more likely related to meat quality traits than to carcass composition traits [118,120]. From a physiological perspective, it can be speculated that these modules describe biochemical processes during meat aging, which are influenced by many environmental effects particular at the time of slaughtering.

Based on the KEGG and GO databases it was possible to allocate the modules to different metabolic pathways and molecular functions. At first glance, several identified pathways are associated with diseases and metabolic defects, but are rarely directly related to biochemical processes involved in muscle to meat conversion or meat quality. This can be expected, because the KEGG and GO databases contain a large number of humane diseases, which are linked to energy metabolisms in muscle like ‘Protein processing in endoplasmic reticulum’ and ‘Oxidative phosphorylation’. However, it should be mentioned that, despite of the existence of such functional links, the transfer of physiological aspects of human diseases to meat quality is not always plausible.

In our analysis, module white, which is correlated with drip loss, is enriched for KEGG pathways ‘One carbon pool by folate’, ‘Tyrosine metabolism’ and ‘Fructose and mannose metabolism’. The most significant identified GO terms of module white are ‘intrinsic apoptotic signalling pathway in response to osmotic stress’ (BP), ‘sarcoplasmic reticulum membrane’ (CC) and ‘methylenetetrahydrofolate dehydrogenase (NADP+) activity’ (MF). Folate or folic acid, a B vitamin, is an essential component of the one-carbon metabolic pathway, which provides carbon groups for nucleic acid synthesis and for numerous methylation reactions, like DNA methylation [241]. The KEGG pathway ‘One carbon pool

by folate' describes the metabolism of cofactors and vitamins. Methylenetetrahydrofolate dehydrogenase (NADP+) is a central enzyme in this metabolism catalysing the metabolisation of different vitamins (http://www.genome.jp/dbget-bin/www_bget?map00670). Together with selenium, folic acid has an antioxidant effect and is used as nutritional supplement for a positive influence on the antioxidative status of meat and on meat quality [242]. A number of studies have evaluated effects of folic acid and selenium on drip loss and meat color [243,244]. It is speculated that organic selenium reduces drip loss by maintaining integrity of the cell membrane and therefore leads to increasing water holding capacity of the muscle. According to Traore et al. [245], higher drip loss is associated with protein oxidation that can be explained as covalent modification of a protein induced either directly by reactive oxygen species or indirectly by reaction with secondary products of oxidative stress. However, the not-essential alpha-amino acid tyrosine or rather the KEGG pathway 'Tyrosine metabolism' is involved in the protein oxidation affecting drip loss.

In case of modules saddlebrown, tan, sienna3, midnightblue and darkgrey, correlated with meat quality traits pH1, pH24 and meat color, there is only little overlap in functional annotation. Only modules tan and midnightblue are consistently enriched for genes and metabotypes of the KEGG pathway 'Protein processing in endoplasmic reticulum'. *Post-mortem* processes in the endoplasmatic reticulum like lipid and protein oxidation are known to produce off-odours, discoloration and unacceptable flavors associated with rancidity [246]. These deficits in meat quality are indicated by parameters color and pH in meat. Moreover, further modules are characterised by the same KEGG pathways. For example, two modules are associated with the metabolic processes 'MAPK signaling pathway', 'Dilated cardiomyopathy' and 'Antigen processing and presentation', respectively. Therefore, these pathways are likely to play an important role in effecting meat quality and carcass composition traits.

Genome-wide association analysis of pathway related modules

There are various reasons for applying the GWAS for modules presenting condensed metabolic information instead of single omics variables: (1) the reduced 'large p, small n' problem by analyzing a smaller number of 'metabolic traits' and (2) redundant results are avoided and computing capacity can be saved. In humans, several metabolite or protein

based GWAS have demonstrated that, in most cases, markers in protein encoding genes are associated with many metabolic components (usually belonging to the same family or included in the same pathway in targeted bio fluids) [110]. Thereby, it is beneficial to summarise metabolic components that belong to the same category or metabolic process into modules instead of analyzing hundreds of thousands of 'traits' independently and producing a plethora of partly redundant associations [49]. GWAS for network based modules, so-called metabotypeGWAS (mGWAS), save computing capacity, reduce the complexity of the systems biological data set and allow calculating genetic associations for metabolic pathways instead of single omics variables. Based on mGWAS it is possible to identify direct links between the function of genes and metabolic pathways, even if the functional interpretation of the results can be complicated by the 'level' in which the relevant metabotype information is placed (transcript, protein or metabolite). The involved genetic variants usually explain a higher fraction of the observed genetic variance for the investigated metabolic process than for low heritable production traits. Nevertheless, the variants are also associated with complex traits like drip loss (even if it is not as strongly as for the specific metabolic pathway) and provide information to understand the etiology of the target trait [247–249].

In mGWAS in livestock, significant markers have been reported even if a lower number of individuals was analysed compared to common GWAS performed in humans. This might be due to the fact, that in animals it is usually easier to control environmental factors and population stratification/pedigree as well as to identify potential sources of variability to be included in the models. These assumptions confirmed the perspective that it is promising to combine multi omics approaches with genomic analyses. In this way, complex omics data are useful to dissect traditional production traits and to describe new traits close to the metabolism of individuals [110].

Until now, a lot of research was done in the field of systems genetics in livestock to increase the information density between genome and phenotype by consideration of additional omics levels. However, most studies analyse only one further omics level (either transcripts, proteins or metabolites). For example, population-based metaboliteGWAS have been reported so far in performance-tested pigs, using targeted metabolomics on plasma [223], and in dairy cattle, using untargeted metabolomics on milk [250]. Related to this, Ponsuksili et al. [247] formed modules based on a network analysis of expression data

(transcripts) and performed eQTL analyses for the modules to identify common regulators of muscle and meat properties in pigs.

In our GWAS approach of the meat quality traits and modules, we adjusted for population stratification by including PCs as fixed effects into the GWAS model (EIGENSTRAT) as suggested by Aulchenko et al. [206] and applied among others by Becker et al. [207], Bergfelder-Drüing et al. [251] and Utsunomiya et al. [208]. Depending on the investigated phenotype (meat quality traits or modules), the models contain one up to 24 PCs, which lead to λ-values close to one. From these results, we conclude a sufficient elimination of population stratification. However, it remains questionable if a large number of PCs in the statistical model will lead to a substantial reduction of useful genetic variation. The Du × Pi population used in this analysis was formed in order to provide a distinct genetic variation in drip loss. In consequence, the selected animals showed a large variation in all meat quality traits.

In contrast to meat quality phenotypes and modules, for carcass composition traits the disturbing effects of population stratification were less important. Regarding the resulting inflation factors λ, the GRAMMAR approach was sufficient.

In the final step of both GWAS approaches (GRAMMAR, EIGENSTRAT), the q-value statistics, which is based on the FDR, was applied in order to correct for multiple testing. Storey and Tibshirani [210] have demonstrated that the FDR provides an acceptable balance between statistical significance and statistical power to detect true effects. In the present study the relaxed significance threshold was set to $q \leq 0.10$, as it has been recommended by Benjamini und Hochberg [211].

New candidate genes for meat quality and carcass composition

The main objective of this study was to identify candidate genes for meat quality and carcass composition traits using both conventional production traits as well as enriched metabolic modules as phenotypes in GWAS. In total, 28 intronic SNPs were found (Tab. 21, p. 132). In order to concentrate this information on the most reliable or promising results, the following section is focussed on candidate genes and genomic regions whose importance was confirmed by overlapping GWAS results using enriched and conventional phenotypes. Moreover, only modules were used, which were significantly correlated with at

least one conventional phenotype. Relevant candidate genes can be arised from congruent genomic regions (positional overlap) or from connected pathways (functional overlap). As an additional verification step, enriched metabolic modules can be characterised by their most important hub metabotypes, which were used as phenotypes in eQTL analyses.

For meat quality traits, genomic overlapping arised exclusively from connected pathways of genes detected for conventional or enriched metabolic phenotypes.

For drip loss and the related metabolic processes presented by module white, three candidate genes on SSC2 (LGR4 and CD44, associated with drip loss) and SSC7 (DST, associated with module white) were detected (Tab. 21). Despite of the positional differences, the overlapping functional annotation of these genes leads to the assumption that processes and characteristics responsible for interactions between cells and the conformation of membranes and the extracellular matrix play an important role in the manifestation of drip loss.

The meaning of gene ‘Leucine-Rich Repeat Containing G Protein-Coupled Receptor 4’ (LGR4) is of particular meaning because two of the most significant SNPs of drip loss are located within intronic regions of this gene (Tab 21). LGR4 is expressed in proliferating cells of diverse tissues, including adult stem cells and progenitor cells. This gene is coding for the transmembrane protein LGR4 that serves as receptor binding the ligand R-spondin [252]. Transmembrane signaling as implemented by receptor proteins like LGR4 plays a critical role in development of the male reproductive tract, eyelids, hair and bone and in different diseases. Furthermore, Glinka et al. [252] provided evidence for an important role of LGR4 on endocytosis and membrane permeability. Wimmers et al. [131] stated that transcripts being up-regulated at high drip loss in pigs belong to groups of genes functionally categorized as genes of membrane proteins, signal transduction, cell communication and response to stimulus. This assumption goes along with our observation of the promising candidate gene LGR4 affecting drip loss. CD44, the second candidate gene for drip loss located on SCC2, also codes a transmembrane glycoprotein acting as cell surface adhesion molecule involved in cell-cell and cell-matrix interaction [253]. Until now, there are no other studies presenting the meaning of LGR4 and CD44 as candidate genes for drip loss or meat quality in general.

Module white is significantly ($p \leq 0.02$) positive correlated ($r=0.25$) with drip loss and the functional annotation of the detected gene 'dystonin' (DST) on SSC7 shows clear accordance with genes LGR4 and CD44. DST encodes a member of the plakin protein family of adhesion junction plaque proteins and is expressed beneath others in neural and muscle tissue (www.genecards.org). According to Damon et al. [254], DST is enriched in the cluster 'extracellular matrix part' of the CC GO terms characterised in MLL in LW pigs compared to Basque pigs. DST is involved in matrix assembly, and its targeted expression strongly affects the collagen network [255]. Dalpe et al. [256] showed that DST deficient mice exhibited weak skeletal muscle cytoarchitecture. Thus, all these biological processes are in accordance with the lean meat properties of the LW compared to the Basque pigs, namely their elevated collagen content and shear force value, and lower tenderness score [257]. The meaning of DST as potential candidate gene is confirmed by the successful verification of DST by eQTL analysis of hub metabotypes of module white (see Tab. 23). The significant intronic SNP in DST 'DRGA0007481' was also identified as significant eQTL for four hub metabolites of module white.

Meat color is a second meat quality trait, where functional overlapping was observed. For this trait and the related module lightgreen, two and three candidate genes on SSC8 and 7 were detected. The importance of candidate genes 'Ubiquitin Conjugating Enzyme E2 T' (UBE2T) and 'protein phosphatase 1 regulatory subunit 12B' (PPP1R12B) detected for meat color, is clarified by the functional overlapping to the genes 'Protein Tyrosine Phosphatase, Non-Receptor Type 21' (PTPN21) and 'Tetratricopeptide Repeat Domain 8' (TTC8) associated with module lightgreen. All these genes are involved in signal transduction and code for different phosphatase proteins or related components responsible for the conformation or stability of these proteins (www.genecards.org). According to Herault et al. [258], gene PPP1R12B that is involved in muscle contraction, is overexpressed in MSM in pigs and is associated with GO BP terms 'Regulation of muscle contraction' and 'Signal transduction'. MSM muscle is characterised by higher proportion of type A myofiber and lower proportion of Type IIb myofiber and higher oxidative capacity compared to MLD [258]. The myofiber composition and the oxidative potential impact the meat color. Based on the small distance of 85,050 base pairs, we assumed that genes PPP1R12B and UBE2T are potentially in LD on SSC8 and constitute a promising location influencing meat quality and the associated cellular and metabolic processes. Genes 'Protein Tyrosine Phosphatase, Non-Receptor Type 21' (PTPN21) and

‘Tetratricopeptide Repeat Domain 8’ (TTC8), detected for module lightgreen, are located in a distance of 347,981 base pairs on SSC7. The protein encoded by the gene PTPN21 is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation (www.genecards.org).

Meat color is significantly correlated with drip loss (r=-0.4) so that connected pathways between these two meat quality traits can be expected. As already described, drip loss (via module white) is related with KEGG pathway ‘Tyrosine metabolism’ (see Tab. 19). This connection is reflected by gene PTPN21 that is involved in tyrosine metabolism and that was identified based on the GWAS of module lightgreen. The importance of the candidate genes identified for module lightgreen and indirectly for meat quality was verified by the eQTL analysis of six hub players of lightgreen. The three SNPs ‘DRGA0008163’, ‘DRGA0008172’ and ‘ALGA0044982’, intronic located in genes PTPN21, TTC8, TRIP, were also revealed as significant eQTL by eQTL analysis of six hub metabotypes of module lightgreen (see Tab. 23, p. 137).

In contrast to meat quality indicators, for carcass composition traits positional as well as functional overlap was found.

Potential candidate genes for carcass composition traits and the related modules were mainly identified on SSC5, 13 and 17. As can be seen in Fig. 22, the chromosomal location on SSC5 is the most interesting region because candidate gene ‘B-cell translocation gene’1 (BTG1) is significantly associated with BFT, LMC_{belly} and LMC_{bonn} and the corresponding intronic SNP ’DRGA0006183‘ is potentially in LD with a second promising SNP. This SNP (‘H3GA0017095’) indicates candidate gene ‘Early endosome antigen 1‘ (EEA1) that is associated with module salmon, which is strongly correlated with all investigated carcass composition traits. In that regard, salmon is a module which condenses all metabolic processes, which are related to carcass composition traits and is directly or indirectly related to genes BTG1 and neighboured gene EEA1. In contrast to standard GWAS, combining all the omics levels given in Fig. 22 increases the information content and thereby the statistical power of the analysis. Consequently, it can be concluded that the genomic region harbouring the genes EEA1 and BTG1 is a very reliable genomic indicator for carcass composition traits within our Du x Pi population.

Genes of the BTG family are involved in cell growth, differentiation and survival rate [259]. Comparing LW vs. Meishan pigs, Mo et al. [260] identified significant differences in expression of BTG2. Moreover, a GWA study in F_2 LW × Meishan pigs resulted in significant associations between BTG2 and carcass composition traits BFT, LMC and MFR [260]. Sun et al. [261] also reported that the BTG2 and BTG3 expression is strongly upregulated in subcutaneous fat tissue in Chinese Erhualian pigs. In contrast to this observation, in a catabolic state the BTG1 expression in fat tissue is significantly upregulated in Yorkshire pigs [262]. Gene EEA1 plays a role in signaling pathways of phagosomes and is involved in endocytosis, those activity varies widely in different breeds (Du, Pi, LR and LW) [263]. As shown in Tab. 19 (p. 124), module salmon is characterised by several signaling pathways, like the 'Mitogen-activated protein kinase (MAPK) signaling pathway' and 'Neutrophin signaling pathway' KEGG terms. This pathways control cell growth and differentiation and the manifestation of meat to fat ratio in many pig breeds [264]. Therewith, the functional annotation of module salmon goes along with the physiological context of the identified candidate genes of module salmon.

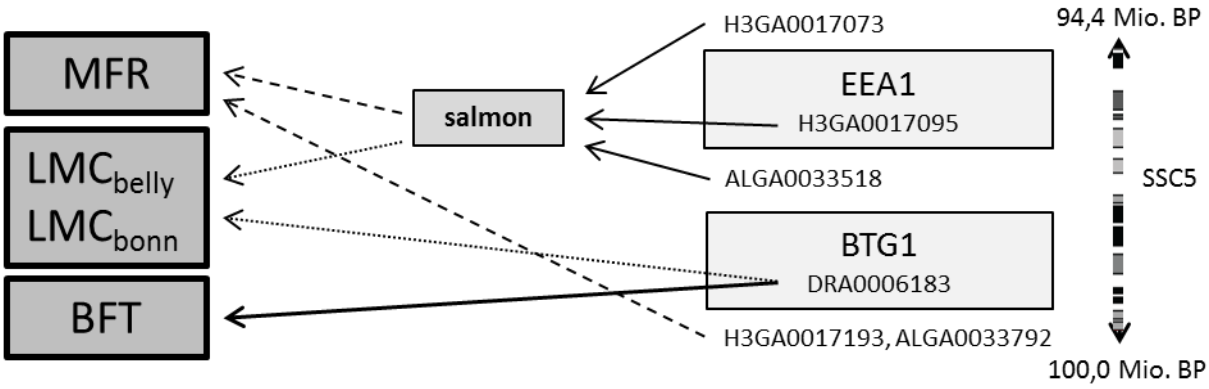


Figure 22: QTL region on *Sus scrofa* chromosome 5 with promising candidate genes for carcass composition traits. MFR - meat fat ratio; BFT - backfat thickness; LMC - lean meat content measured by formula of Grub in belly (LMC_{belly}) and by Bonner formula (LMC_{bonn}); BP – base pairs; SSC5 – *Sus scrofa* chromosome 5; The declaration of gene symbols in the light grey boxes can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/gene.

4.5 Conclusion

As has been shown for meat quality or carcass composition traits, the reflection of functional and positional overlapping GWAS results of conventional traits and enriched metabolic modules as phenotypes is a useful instrument to improve the reliability and the physiological plausibility of detected QTL. This approach can be extended by eQTL for hub metabotypes, which are the key regulators in the metabolic modules. It has been demonstrated for meat quality traits that the importance of candidate genes can be validated by positional overlap of QTL (conventional and enriched module phenotypes) and eQTL of module hub metabotypes.

In case of carcass composition traits, functional as well as positional overlap was relevant. However, there was no accordance between intronic SNPs indicating candidate genes (result of GWAS) and eQTL for hub metabotypes of specific modules (result of eQTL analyses). One reason might be that the underlying Du x Pi population was particular selected for divergent phenotypes for drip loss. Although there is a moderate correlation to meat quality, the variation in carcass composition traits is expressed to a lesser extent.

This might be one reason why the results of GWAS and eQTL analyses did not overlap. Nevertheless, regarding carcass composition traits, the meaningful functional and positional overlap of QTL results for conventional and enriched module traits provides reliable information for the physiological importance of genes and genomic regions.

Chapter 5. General discussion and conclusion

Introduction in systems genetics and aims of the thesis

The muscle to meat conversion is a complex physiological process which strongly affects meat quality traits in pigs. Although meat quality traits are strongly processed in pigs, today there is still a lack of knowledge concerning the biochemical processes particular during early meat aging. Besides external environmental factors like slaughter conditions, feeding system and other management effects, mainly the individual genetic background is responsible for metabolic efficiency of meat conversion under given conditions. In addition, the complex underlying biological pathways of these metabolic processes are also responsible for the expression of related meat quality and production traits.

Systems biological approaches have the aim to get holistic insights in complex traits. For this purpose, the system is perturbed, the responses are monitored, multi-level data are collected and analysed and hypotheses are formulated, which describe the whole metabolic process with mathematical models [55]. In general, a systems biology approach provides a more comprehensive picture of the processes effecting complex traits [1]. Due to Fontanesi [110], "the complexity of biological systems requires unconventional approaches to describe the interactions among the different levels of biological information and their dynamics". Against this background, omics analyses have the potential to revolutionize quantitative genetics in livestock production by providing a deeper understanding of the metabolic processes underlying the performance traits, investigating genetic × environment interactions and identifying genetic markers for low heritable performance traits based on omics phenotypes [53,114].

After comprehensive preliminary investigations in the Bonner Du × Pi population, the present doctoral thesis has the objective to integrate all omics levels in order to elucidate the genetic and metabolic background of meat quality and carcass composition traits, paying special attention on drip loss. In the first step, a metabolite study (Chapter 2) with the aim to identify potential metabolite bio indicators for drip loss, pH1, pH24 and meat color was performed. In the following studies, the number of omics levels was extended leading to an increased biological and statistical complexity. In the second study (Chapter 3), the omics levels genome, proteome and metabolome were combined and analysed as intermediate phenotypes in GWAS to reveal promising candidate genes for WHC. In the final study (Chapter 4), a holistic omics approach was performed, combining genome,

transcriptome, proteome and metabolome, to get comprehensive insights into the metabolic processes related to meat quality and carcass composition and to identify genetic markers based on GWAS for combined metabolic traits.

The underlying basis of this thesis was a data set that characterises the whole metabolic background of meat quality and carcass composition in F_2 Du × Pi pigs. To my knowledge, such a comprehensive omics data set is unique in the genomic research area of pork quality. This applies especially for the untargeted metabolite profiling data set in study 1, which comprises a very large number of quantified metabolites.

Against this background, it is not surprising that mGWAS lead to the identification of many interesting genes and genomic regions, which are important for the target traits. Beyond the single GWA approaches described in chapter 1, we focused in the following chapters 3 and 4 on various aspects how to combine the different omics information in order to find reliable QTLs for the target traits. In this regard, statistical tools and informational external database were available, which were used be used for data analysis and biological interpretation of the results. However, particular in the construction of enriched phenotypes that condense the large amount of expression data and the biological interpretation of these intermediate traits no generally accepted standards are available. Finally, some details of the applied methods can be regarded as heuristic or prototype approaches, which need further verification.

Selection of the best metabolite indicators for drip loss based on different statistical approaches

In the first published study (Chapter 2), different statistical approaches (correlation analysis, PCA, WNA and RFR) were applied to handle the ‘large p, small n’ problem and to determine the most predictive metabolite biomarkers for different meat quality traits. Despite various available statistical tools, it is still a challenge to handle such large data sets with 1,993 detected and quantified metabolites but only a limited number of 97 animals without the risk of unacceptable overfitting.

Besides univariate statistical analyses (e.g. ANOVA), there is a wide variety of machine learning methods (e.g. ensemble methods, support vector machines, partial least squares-discriminant analysis, linear discriminant analysis) that are useful in the analysis of omics

variables. In our studies, the machine learning and network based methods RFR and WNA were used. RFR belongs to the family of ensemble methods (e.g. bagging [163] and boosting [265]) and became a very popular approach in classification and discrimination in various applications. For example, RFR is widely used in untargeted omics approaches for the identification of reliable bio indicators [266,266,267]. Regarding the RFR procedure, in study 1, the Hothorn´s conditional RFR [176] was applied that calculates the VI significance based on an implemented permutation test according to the principle of MDA (see Gini importance in [163]). As already described in Welzenbach et al. [192], leads this selection step to a reduced number of explanatory variables in the model that avoids overfitting and ensures a smaller prediction error. Important disadvantages of RFR in comparison to WNA are the limited possibilities to quantify network parameters that allow investigating metabolite-metabolite interactions. RFR provides the parameter VI, which can be used as a statistical indicator in order to characterise the prediction accuracy of this parameter with respect to the target traits. Although the parameters are based on different statistics (sampling vs. Pearson correlations), VI values in RFR have a similar validity and meaning than MS values in WNA. However, both parameters do not allow assigning metabolites to functional pathways. In order to lower this problem, WNA offers the parameter MM, which reflects the importance of a metabolite, taking into account metabolite-metabolite interactions and the role and position of specific metabolites within the network.

Tested by trait specific multiple regression analyses, the machine learning method RFR selected the most predictive ten metabolite biomarkers. Moreover, the prediction accuracy also depends on the investigated trait. In general, prediction based on metabolite profiles is very challenging especially in case of drip loss and it worked best for pH1 (Tab. 10, p. 62).

In addition, RFR procedures partly generate ‘odd unexpected results’ in certain cases. Even in the conditional RFR, Boulesteix et al. [184] assess that in case of specific data structure or predictor type, there is an increased risk of biased VI values. In conclusion, the usage of RFR in biomarker discovery in metabolomics seems to be very promising, but due to the complex resampling schemes in construction of the decision trees the physiological interpretation of the most predictive variables is critical.

Besides the statistical methods, the choice of the appropriate tissue for metabolite profiling is crucial for the successful identification of bio indicators. It is important to record metabolite profiles in tissues or fluids that are responsible for the manifestation of the target trait. In that regard, metabolite profiling in the muscle tissue of important carcass cuts like MLD, as we have it done in our studies, is an obvious procedure to characterise meat quality. In a similar context, Muroya et al. [120] stated that in fast and slow type muscles in pigs, different metabolites are the best indicators for meat quality parameters. In addition to the choice of the appropriate tissue, the time point of metabolite profiling is of great importance. Because the metabolome will undergo a rapid change, it is essential to assess the time point of metabolite profiling as precisely as possible. That means the metabolome should be profiled at that time point, in which crucial metabolic processes affecting the trait of interest occur. Rohart et al. [118] performed the metabolite profiling a few weeks before slaughtering and did not identified predictive metabolite indicators for meat quality in the investigated pigs. Later, these authors conceded a mistake and traced their bad results to the too early time point of metabolite profiling. In contrast, in the present study, metabolites were quantified in MLD samples, which were collected and snap-frozen immediately after exsanguination. At that time, the rate of glycolysis and oxidative processes and the quantity of related metabolites may be used to derive ultimate meat quality parameters.

Consequently, in our study metabolite profiling is based on a reasonable experimental design regarding the tissue and time point of measurement. However, there are still two critical points remaining. First, the metabolite profiling procedure, based on GC-MS and LC-QTOF/MS performed by Metabolomic Discoveries GmbH leaves open the question if the metabolites are endogenous or xenobiotics (see section 1.2.4, p. 27). Only endogenous metabolites that are produced directly by the organism provide insights in the underlying metabolic processes of specific traits. Xenobiotics have their origin in the environment of the individuals and are absorbed by the organism. Thereby, these metabolites might be environmental indicators but distort the view into the metabolic programming independent of the environmental conditions. The manual annotation of one non-annotated metabolite with the molar mass of 362.0154.4.99 in METLIN showed with a negligible small tolerance of two ppm that this metabolite figures the chemical compound Coumaphos. This substance has an antiparasitic effect and is applied as insecticide against ectoparasites in swine. The metabolite Coumaphos was selected as predictive biomarker for drip loss (identified by correlation analysis) and pH1 (identified based on MS in WNA). Because the manual

annotation in METLIN is affected by some risk of error, there is a need to check if the described metabolite actually is the substance Coumaphos. Xenobiotics provide a huge field of new applications in the investigation of individual × environment interaction. But in accordance with the objective to analyse the metabolic background of performance traits, a distinction should be made between xenobiotics and endogenous metabolites.

In general, the limitations in metabolite annotation are an important unsolved problem in evaluation of metabolomics analyses. In the present study, only 20% of the detected and quantified metabolites were annotated. In fact, the biological function of the major part of the selected, most predictive metabolites remains unknown. Thereby, it can be expected that many metabolic pathways, which are implicated in the manifestation of meat quality were not detected. However, it can be speculated that the highly ranked non-annotated metabolites are linked to annotated metabolites so that the amount of undetected pathway is reduced. Nevertheless, it can be postulated that enhanced technical opportunities for the metabolite annotation would lead to substantial improvements in the detection of biological pathways.

Another bottleneck effect in metabolomics is caused by the limited technical capabilities in metabolite detection and quantification. Although currently GC-MS and LC-QTOF/MS are the most precise applications in metabolite profiling, it can be assumed that only a small percentage of the actually present metabolic components can be measured. In particular, small molecules might be hidden by bigger components and thereby often stay undetected. This is another bottleneck effect in metabolomics, which limits the possibility to uncover all relevant or at least the most predictive metabolites.

GWAS based on metabolic traits

In the second and third study (Chapter 3 and 4) the focus was on the disclosure of the genetic and metabolic background of drip loss and other meat quality parameters as well as carcass composition traits. As already mentioned in Welzenbach et al. [144], the information flow from genotype to phenotype is not linearly dispersed pursuant to the variety of possible post-transcriptional events. Moreover, metabolic regulation processes responding very quickly to environmental changes complicate the elucidation of the genetic background. Kadarmideen [53] and Widmann et al. [147] clarified that holistic omics

approaches are beneficial and powerful in the identification of reliable candidate genes in complex traits. Therefore, in line with omics approaches, the GWAS in studies 2 and 3 were not only applied for standard performance traits like drip loss and LMC but also for metabotypes, like proteins, metabolites and transcripts or combined metabolic traits. It can be expected that metabotypes are products of certain metabolic processes and hence are strongly associated with conventional phenotypes continuously recorded in animal production [118,161]. Therefore, metabolic traits constitute essential links between genetic information and phenotypical expression of complex performance traits and might be used in GWAS to improve the statistical power and to reveal less false positive, redundant results [144]. This means, metabotypes potentially are more convenient in the estimation of the individual genetic potential than the conventional target phenotypes itself. Moreover, the mGWAS resulted in genetic variants that provide information to comprehend the etiology of the traits of interest [110].

It can be hypothesized that both single metabolites and proteins as well as enriched metabolic traits, as they have been analysed in study 2 and 3, are appropriate traits to investigate the genetic background and to understand the etiology of complex traits. In study 2, metabolites and proteins were selected based on their belonging to specific pathways that are related to drip loss. The pathway analysis was performed based on KEGG annotated metabolites and entrez gene ID annotated proteins. This step led to a drastic reduction of adaptable metabolites and proteins, because their biological functions and IDs are still unknown to a large extent. Because of this severe bottleneck, the number of utilizable metabolites drops from 1,865 to only 128 and it can be expected that many metabolites with strong influence on drip loss were excluded. The protein profiling was based on a targeted approach. For this reason the number of quantified proteins was relatively small right from the beginning of the analysis. Finally, in the enrichment analysis 35 entrez gene ID annotated proteins were included.

One important statistical problem of the enrichment analysis is the detection of overrepresented groups of metabotypes with small effects. Based on simulation results, Michaud et al. [196] verified that the underlying function of the applied enrichment test in study 2 (Function geneSetTest; Package limma version 3.30.8) has presumably enough

power to detect overrepresented groups of metabotypes, even if the effects are very small. This argument can be used to explain, why the enrichment analysis has resulted in functional sets of metabotypes although correlation coefficients between individual metabotypes and drip loss did not significantly deviate from zero (Tab. 12, p. 81).

Based on the significant enriched pathways, we used the related metabolites and proteins as metabolic traits in GWAS to elucidate the genetic background of drip loss. As mentioned above, the potential of these GWAS in study 2 is constrained because of the reduced number of metabolites and proteins, which were annotated. Because of this preselection, the GWAS results presumably reflect only a fragmentary picture of the complex metabolic processes.

This constraint was partly removed in study 3, where all detected and quantified metabolites, proteins and transcripts, regardless of the annotation, were condensed to modules. The modules were annotated and the most important modules were selected as metabolic traits for meat quality and carcass composition traits. There are various reasons for applying the GWAS for enriched modules reflecting condensed metabolic information instead of single omics variables:

1. The entire metabolic information was used to figure out the metabolic and genetic background of meat quality and carcass composition.
2. By analyzing a smaller number of metabolic traits, the ‘large p, small n’ problem was reduced, redundant results were avoided and computing capacity was preserved.
3. The WNA also provided information about the relationships within and between the modules. Even if the parameters MM and connectivity did not provide directed information about the interactions of the metabolic components, the parameters allowed the identification of the most important actors of a pathway.

Based on the mGWAS approach it was possible to identify direct links between the function of genes and metabolic pathways. However, because of the complex interactions between metabotypes of the different omics levels, a detailed functional interpretation of the results is hardly possible. Nevertheless, the WNA based on different data sets of holistic

omics profiling that pictured the underlying metabolism to a large extent. To my knowledge, there are no other investigations creating metabolic traits based on 12,235 transcripts, 1,993 metabolites and 40 proteins with the aim to elucidate the genetic and metabolic background of production traits in livestock. In addition, the network approach provided a solution to handle the missing annotation problem. That means, based on the connectivity within a module, the function of non-annotated metabotypes was partly uncovered by clustering these components to metabotypes with known functions.

Important metabotypes/pathways for drip loss

Based on all applied omics approaches it was possible to confirm the meaning of energetic processes like glycolysis/gluconeogenesis, citrate cycle and PPP, which regulate muscle physiology and determine the final meat quality [9,199]. The enrichment analysis (Chapter 3) revealed that a higher glycine content indicates a higher rate of glycolytic processes. A high glycolytic potential is known to be related with high drip loss [144]. The relationship between elements of glycolysis/gluconeogenesis like glycine and drip loss in meat was already illustrated by Lim et al. [200]. They described the observation of higher drip loss in pork in the case of increased glycine level in porcine skeletal muscle cells. Moreover, the enzymes FBPase and TPI1 and the metabolite glycerone-p are also agents in glycolysis/gluconeogenesis and thereby are linked to drip loss. Beneath glycolysis, the enrichment analysis also highlighted the meaning of pyruvate and methane metabolism that affect the muscle to meat conversion and directly affect drip loss [9,199]. The proteins PKM and PGAM2 and the metabolite DG3P are equally key players of glycolysis/gluconeogenesis and pyruvate metabolism and promising indicators for drip loss.

In contrast to the glycolysis related components that are responsible for the whole complex of meat quality traits, several lipids seemed to be much more specific for drip loss. The various performed statistical procedures in study 1 (Pearson correlation analysis, PCA, RFR and WNA) identified a few key metabolites that mainly pertain to the family of lipids (GPLs, sterol lipids, prenol lipids). In mammalian cell membranes, GPLs are the most important lipids [187]. The enrichment analysis in study 2 also revealed the sphingolipid metabolism that belongs to the GPL metabolism, as most important pathway for drip loss. In cell fluid dissolved GPLs and sphingolipids prove the link between drip loss and rate of

processes of sphingolipid metabolism [187,188]. Hub actors of this pathway are ceramides, phosphoethanolamines and serines.

The results of study 3 cannot be integrated in the known metabolic background of drip loss so easily. According to the module annotation, drip loss was associated with KEGG pathways 'One carbon pool by folate (folic acid)', 'Tyrosine metabolism' and 'Fructose and mannose metabolism'. While the metabolism of fructose and mannose still fitted in the physiological background of energetic processes, the tyrosine metabolism proved another factor that affects drip loss. The not-essential alpha-amino acid tyrosine is involved in protein oxidation. According to Traore et al. [245], higher drip loss is associated with protein oxidation that can be explained as covalent modification of a protein induced either directly by reactive oxygen species or indirectly by reaction with secondary products of oxidative stress.

Candidate genes for drip loss

Beneath the elucidation of the metabolic background of drip loss, the second main objective of this thesis was the detection of reliable candidate genes and genomic regions based on multi omics approaches. The most promising genomic region and candidate genes for drip loss based on mGWAS were detected on SSC18. This region also has been earlier described by Jennen et al. [212] and Liu et al. [27], performing a QTL study in the Du × Pi animals. In the region around 12 Mb, the genes PTN and CREB3L2 are located in close proximity. The family of cAMP response element binding proteins (associated with gene CREB3L2) is crucial for a variety of cellular processes including cell proliferation, differentiation, apoptosis, extra-stimuli and stress response [213]. Although the meaning of CREB3L2 has not been described precisely for meat quality so far, the results suggest that this gene seems to have a relevant influence in energy metabolism in skeletal muscle that is indicated by its interacting effect on PGAM2, glycine and drip loss (Fig. 18, p. 101). Further promising genes in this region on SSC18 were LRGUK and EXOC4.

Unfortunately, the meaning of these candidate genes for drip loss was not confirmed by the GWAS in study 3. The GWA study of drip loss directly resulted in two candidate genes, LRG4 and CD44, on SSC2. Glinka et al. [252] provided evidence for an important role of LGR4 on endocytosis and membrane permeability. CD44 coded for a transmembrane glycoprotein acting as cell surface adhesion molecule involved in cell-cell and cell-matrix

interaction [253]. Moreover, a third gene potentially effecting drip loss was identified on SSC7 based in the GWA study of module white. The gene DST encodes a member of the plakin protein family of adhesion junction plaque proteins and is expressed beneath others in neural and muscle tissue. DST is involved in matrix assembly, and its targeted expression strongly affects the collagen network [255]. There are many studies verifying the association of DST to collagen content, shear force and tenderness [83,257]. Moreover, the meaning of DST as potential candidate gene was also confirmed by the successful verification of DST by eQTL analysis of hub metabotypes of module white (Tab. 23, p. 137).

One reason for the missing overlapping in identified candidate genes might be the differences in information density and statistical power in study 2 and 3. As described in section 4.2.2 (p. 117), missing genotype data was imputed in study 3, so that the number of utilizable SNPs and the number of pigs with records available across all omics levels was increased. These changes in data structure might be responsible for the result that only three SNPs (‘ALGA0089069’, ‘CASI0008411’ and ‘ASGA0072217’) located on SSC16 were found in both studies (see Tab. 17, p. 93). However, these SNPs are intronic or not located in a 1Mb distance to adjacent functional genes. For this reason, the SNPs are not presented in results of study 2. In conclusion, although found in both studies, these SNPs located on SCC16 seemed not to be suitable as potential candidate genes for drip loss.

Challenges and perspectives

Enormous progress in detection technology has resulted in a continuous expansion of omics databases filled with biological data. Along with new approaches in statistical evaluation, these informative data have enabled and revolutionized the analysis of high dimensional omics data sets [143,227]. Despite the enhancements in omics technologies, the so far performed omics analyses are exemplary approaches and still far away from proving a holistic view into complex systems. Problems in the holistic profiling of an omics level like the metabolome or the proteome mainly result from still fragmentary functional databases leading to an incomplete annotation. Also from the statistical perspective, some problems remain unresolved. For example, it is a big challenge how to correctly combine and rate omics variables of various origins into a statistical model. The integration of omics data is associated with many challenges caused by the data heterogeneity, the ‘large p, small n’

situation and multicollinearity. To be able to handle these facts, some new, innovative statistical approaches have been developed. For example, penalized regression methods based one 'The Least Absolute Shrinkage and Selection Operator' (LASSO) proposed by Tibshirani [268] or 'Elastic Net' (ENET) proposed by Zou and Hastie [269], are able to combine various types of omics data and to handle multicollinearity and overfitting. Simultaneously, in contrast to similar procedures, LASSO and ENETs are characterised by a reduced required computation time [270]. Today, LASSO and ENETs are applied in genetics [271,272] and omics analyses [273]. In the so-called IPF-LASSO (Integrative LASSO with Penalty Factors), the variables (e.g. transcripts, metabolites) are allocated to different penalty factors for feature selection and prediction. The results of Boulesteix et al. [274] showed that IPF-LASSO performs better than standard LASSO when the variables differ in importance with respect to outcome prediction. Based on penalized regression methods, in the future, information from different omics levels could be weighted in an ideal way for new genomic selection procedures, like 'Omics based selection' (OBS). By taking into account environmental aspects and the time point of omics quantification, OBS might be beneficial used in animal genetics and monitoring of health, welfare and disease.

Besides the progress in animal selection procedures, omics profiles also have the potential to revolutionize association studies. Present GWAS can associate phenotypes only with genetic factors, and not with metabolic factors that reflect environmental stimuli. Therefore, only a small proportion of heritability for multifactorial can be explained by conventional association studies. GWAS did not indicate direct biochemical interactions between genotype and phenotype. In contrast, globally integrated association studies that reflect both genomic and metabolic information leading to molecular networks, are promising in the analysis of complex traits and the identification of related molecular mechanisms. Yugi et al. [275] proposed so-called 'trans-ome-wide association studies' (trans-OWAS) including all omics levels as the future of association studies in humans. It can be expected that this trend will also be transferred to livestock breeding (see Fig. 23, p 137).

Although my omics approach already based on a comprehensive metabolic data set, for an effective holistic view into the metabolism behind meat quality traits, also omics levels like the epigenome (complete collection of changes to the DNA and histone proteins), metagenome (complete collection of genetic material contained in an environmental sample) and the functome (complete collection of functions described by all the

complementary members in living organisms) should be considered in subsequent investigations [114]. Furthermore, it can be expected that the interest in the omics level microbiome, that pictures the relationship between the genetic of the host and the microbes e.g. in its gastrointestinal tract, also will strongly increase in the next years. Finally, it can be noted that the research in omics is still at the very beginning and, until today, about the entire application potential can only be speculated

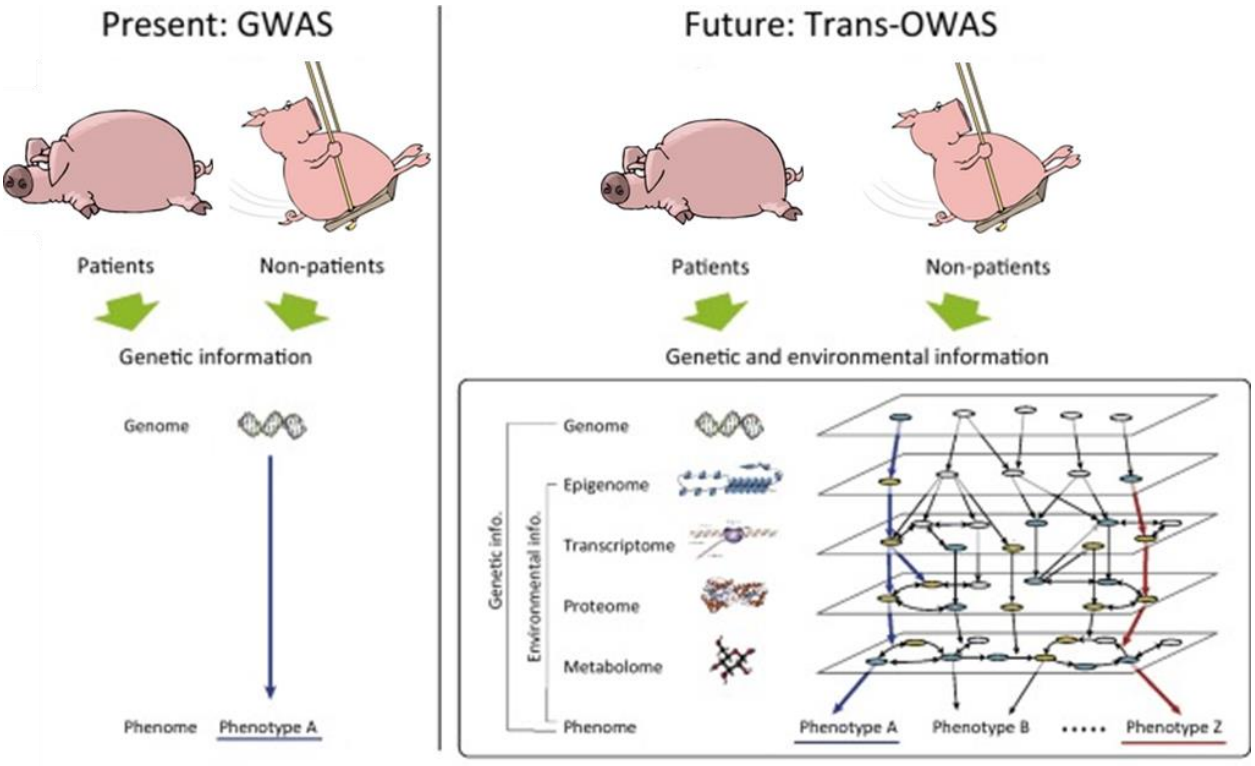


Figure 23: From gnome-wide association study (GWAS) to trans-omics-wide association study (Trans-OWAS). Modified from Yugi et al. [222].

Chapter 6. Summary

WHC and its phenotypic indicator drip loss are important commercially interesting meat quality parameters, which are essential for the acceptance of consumers and manufacturing industries. Today, many genes causing hereditary defects and genes with large effects causative for performance traits are known and successfully established in breeding programs. Nevertheless, the variability of drip loss is high and the regulation of muscle properties influencing meat quality traits is still unclear. For this reason, animal breeding increasingly focusses on a deeper understanding of the underlying metabolic processes of performance traits, on investigating genetic × environment interactions, or on identifying genetic markers for low heritable traits also with small effects. Several research groups stated that systems biological approaches are valuable and powerful in identifying key causal and highly predictive genetic variants for complex traits as well as in building up complex genetic regulatory networks.

In the Bonner Du × Pi population, the quantitative traits meat quality and carcass composition were comprehensively investigated and based on different omics levels several promising QTL, eQTL and proteins for meat quality parameters were identified. To continue these investigations, the purpose of this thesis was the application of different statistical approaches to analyse and integrate the omics levels genome, transcriptome, proteome, metabolome and phenotype. Besides the investigation of the suitability of the methods, the aim was to elucidate the genetic and metabolic background of meat quality and carcass composition traits with special attention to drip loss. For this purpose, nearly 100 animals of the Bonner Du × Pi population were genotyped with the porcine 60 k Illumina beadchip and the metabolome, proteome and transcriptome was profiled in samples of the MLD.

In the first study (Chapter 2), based on different statistical procedures, namely correlation analysis, PCA, WNA and RFR, metabolite profiles were analysed to elucidate the underlying biochemical processes and to identify potential metabolite bio indicators for drip loss, pH1, pH24 and meat color. By an untargeted metabolomics approach the profiles of 393 annotated and 1,600 unknown metabolites were detected in 97 Du × Pi pigs. Despite obvious differences in the underlying statistics, the four applied methods revealed mainly concordant results regarding the identification of key metabolites for meat quality

parameters. The findings led to the conclusion that meat quality traits pH1, pH24 and color are strongly influenced by processes of p.m. energy metabolism like glycolysis and PPP, whereas drip loss is significant associated with metabolites of lipid metabolism. In case of drip loss, RFR was the most suitable method to identify reliable biomarkers and to predict the phenotype based on metabolites. On the other hand, WNA provided the best parameters to investigate the metabolite interactions and to clarify the complex molecular background of meat quality traits. In summary, it was possible to attain findings on the interaction of meat quality traits and their underlying biochemical processes. It can be assumed that the detected key metabolites are better indicators of meat quality especially for drip loss than the measured phenotype itself and potentially might be used as promising bio indicators.
In the sense of a systems biology approach, the next step, the investigation of the metabolic background was used to identify key causal and highly predictive genetic variants for meat quality and carcass composition traits based on metabolic and proteomic information (Study 2, Chapter 3). In the last step of the analysis, combined networks of metabolites, proteins and transcripts were used to elucidate the genetic and metabolic determination of production traits (Study 3, Chapter 4). In both studies, the omics analyses were applied as promising alternatives to standard genetic association studies. Accordingly, the metabotypes were used as more accurate phenotypes to characterise underlying functional pathways and candidate genes for drip loss and other meat quality parameters and carcass composition traits.

In study 2, profiles of 126 KEGG annotated metabolites and 35 entrez gene ID annotated proteins in 97 Du × Pi pigs were considered. In the first step, an enrichment analysis resulted in ten pathways, amongst others, in sphingolipid metabolism and glycolysis/gluconeogenesis, with significant influence on drip loss. In the following, drip loss and 22 metabolic components were analysed as intermediate phenotypes within a GWA study. Based on the GWAS it was possible to identify significantly associated genetic markers and candidate genes for drip loss and for most of the metabolic traits. On SSC18, a region with promising candidate genes was identified based on SNPs associated with drip loss, the protein PGAM2 and the metabolite glycine.

In study 3, based on a network analysis that integrated the profiles of 12,235 transcripts, 1,993 metabolites and 40 proteins in 90 Du × Pi pigs, most promising metabolic traits were selected. To figure the whole metabolic background of drip loss, the meat quality traits pH1, pH24 and color and additionally the related carcass composition traits MFR, BFT, LMC_{belly} and LMC_{bonn} were investigated. The meat quality, carcass composition and metabolic traits were analysed in GWAS to identify reliable candidate genes for drip loss and related performance traits. Finally, as verification of the identified candidate genes, eQTL analyses for the hub players of the modules were performed. Based on this systems biological approach it was possible to confirm already known functional pathways related to drip loss like the 'Tyrosine metabolism' that is associated with protein oxidation. Moreover, the study provided some new insights into the metabolic and genetic background of drip loss. For example, the genes LGR2, CD44 and DST seemed to be promising candidate genes for drip loss and related metabolic processes. In case of MFR, BFT and LMC the holistic omics approach was particularly expedient and the GWAS revealed a highly interesting genomic region for this trait complex on SSC5. The importance of these regions is based on the high density of significant markers for the carcass composition traits and the associated metabolic phenotype 'module salmon'.

In general it can be hypothesized that association studies based on intermediate phenotypes are able to provide comprehensive insights in the genetic variation of genes directly involved in the metabolism of performance traits. Therefore, the statistical power in the identification of reliable candidate genes is improved and false positive, redundant results are avoided.

Chapter 7. References

1. Picard B, Lebret B, Cassar-Malek I, Liaubet L, Berri C, Le Bihan-Duval E, et al. Recent advances in omic technologies for meat quality management. Meat Sci. 2015; 109: 18–26. doi: 10.1016/j.meatsci.2015.05.003.
2. D'Alessandro A, Zolla L. Meat science: From proteomics to integrated omics towards system biology. J Proteomics. 2013; 78: 558–577. doi: 10.1016/j.jprot.2012.10.023.
3. Kitano H. Systems biology: A brief overview. SCIENCE. 2002; 295: 1662–1664.
4. Cassens RG. Historical perspectives and current aspects of pork meat quality in the USA. Food Chem. 2000; 69: 357–363.
5. Hocquette J-F, Renand G, Levéziel H, Picard B, Cassar-Malek I. The potential benefits of genetics and genomics to improve beef quality-a review. Animal Science Papers and Reports. 2006; 24: 173–186.
6. Kauffman RG, Cassens RG, Scherer A, Meeker DL. Variations in pork quality. History, definition, extent, resolution. Des Moines, Iowa: National Pork Producers Council; 1992.
7. Hofmann K. Der Begriff Fleischqualität, Definition und Anwendung. Fleischwirtschaft. 1987; 67: 4.
8. Verbeke W, Perez-Cueto, F. J. A., de Barcellos, M. D., Krystallis A, Grunert KG. European citizen and consumer attitudes and preferences regarding beef and pork. Meat Sci. 2010; 84: 284–292. doi: 10.1016/j.meatsci.2009.05.001.
9. Scheffler TL, Gerrard DE. Mechanisms controlling pork quality development. The biochemistry controlling postmortem energy metabolism. Meat Sci. 2007; 77: 7–16. doi: 10.1016/j.meatsci.2007.04.024.
10. Offer G. Modelling of the formation of pale, soft and exudative meat: Effects of chilling regime and rate and extent of glycolysis. Meat Sci. 1991; 30: 157–184. doi: 10.1016/0309-1740(91)90005-B.
11. Huff-Lonergan E, Lonergan SM. Mechanisms of water-holding capacity of meat. The role of postmortem biochemical and structural changes. Meat Sci. 2005; 71: 194–204. doi: 10.1016/j.meatsci.2005.04.022.
12. Nelson DL, Lehninger AL, Cox MM. Lehninger principles of biochemistry: Macmillan; 2008.

13. Choe JH, Choi YM, Lee SH, Nam YJ, Jung YC, Park HC, et al. The relation of blood glucose level to muscle fiber characteristics and pork quality traits. Meat Sci. 2009; 83: 62–67. doi: 10.1016/j.meatsci.2009.03.011.
14. Binke R. From muscle to meat. Fleischwirtschaft. 2004; 84: 224–227.
15. Pösö AR, Puolanne E. Carbohydrate metabolism in meat animals. Meat Sci. 2005; 70: 423–434. doi: 10.1016/j.meatsci.2004.12.017.
16. Honikel KO. How to measure the water-holding capacity of meat? Recommendation of standardized methods. Evaluation and control of meat quality in pigs: Springer; 1987. pp. 129–142.
17. Toldra F, Flores M. The use of muscle enzymes as predictors of pork meat quality. Food Chem. 2000; 69: 387–395. doi: 10.1016/S0308-8146(00)00052-2.
18. Schafer A, Rosenvold K, Purslow PP, Andersen HJ, Henckel P. Physiological and structural events postmortem of importance for drip loss in pork. Meat Sci. 2002; 61: 355–366. doi: 10.1016/S0309-1740(01)00205-4.
19. Rosenvold K, Andersen HJ. Factors of significance for pork quality - a review. Meat Sci. 2003; 64: 219–237.
20. Grau R, Hamm R. Eine einfache Methode zur Bestimmung der Wasserbindung in Fleisch. Fleischwirtschaft. 1952; 4: 295–297.
21. van Oeckel MJ, van Warnants N, Boucqué CV. Comparison of different methods for measuring water holding capacity and juiciness of pork versus on-line screening methods. Meat Sci. 1999; 51: 313–320.
22. Kauffman RG, Eikelenboom G, van der Wal PG, Merkus G, Zaar M. The use of filter paper to estimate drip loss of porcine musculature. Meat Sci. 1986; 18: 191–200.
23. Offer G, Knight P. structural basis of water-holding in meat. 2. Drip losses. Developments in meat science. 1988.
24. Otto G, Roehe R, Looft H, Thoelking L, Kalm E. Comparison of different methods for determination of drip loss and their relationships to meat quality and carcass characteristics in pigs. Meat Sci. 2004; 68: 401–409.
25. Borchers N, Otto G, Kalm E. Genetic relationship of drip loss to further meat quality traits in purebred Pietrains. Arch Tierzucht. 2007; 50: 84–91.
26. Sellier P. Genetics of meat and carcass traits. Genetics of the pigs, CAB International New York.

27. Liu G, Jennen, D. G. J., Tholen E, Juengst H, Kleinwachter T, Holker M, et al. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Anim Genet. 2007; 38: 241–252. doi: 10.1111/j.1365-2052.2007.01592.x.
28. Fujii J, Otsu K, Zorzato F, Leon S de, Khanna VK, Weiler JE, et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science. 1991; 253: 448–451.
29. Zentralverband der Deutschen Schweineproduktion (ZDS). Richtlinien für die Stationsprüfung auf Mastleistung, Schlachtkörperwert und Fleischbeschaffenheit beim Schwein; 10.12.2003.
30. Tholen E. Genetic foundation of meat quality traits of station tested slaughter pigs in North Rhine-Westphalia (Germany). A status report. Arch Tierzucht. 2005; 48: 123.
31. Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, et al. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science. 2000; 288: 1248–1251.
32. Ciobanu DC, Lonergan SM, Huff-Lonergan EJ. Genetics of Meat Quality and Carcass Traits. In: Rothschild MF, Ruvinsky A, editors. The Genetics of the Pig: CAB International; 2011. p. 358.
33. Ciobanu DC, Bastiaansen, J W M, Lonergan SM, Thomsen H, Dekkers, J C M, Plastow GS, et al. New alleles in calpastatin gene are associated with meat quality traits in pigs. J Anim Sci. 2004; 82: 2829–2839.
34. Casas E, White SN, Wheeler TL, Shackelford SD, Koohmaraie M, Riley DG, et al. Effects of calpastatin and mu-calpain markers in beef cattle on tenderness traits. J Anim Sci. 2006; 84: 520–525.
35. Reardon W, Mullen AM, Sweeney T, Hamill RM. Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine M. longissimus and M. semimembranosus. Meat Sci. 2010; 86: 270–275. doi: 10.1016/j.meatsci.2010.04.013.
36. Purslow PP, Mandell IB, Widowski TM, Brown J, deLange, C. F. M., Robinson, J. A. B., et al. Modelling quality variations in commercial Ontario pork production. Meat Sci. 2008; 80: 123–131.

37. Borchers N. Genetic relationship of drip loss to further meat quality traits in purebred Pietrains. Archiv für Tierzucht. 2007; 50: 84.
38. Kim JM, Choi BD, Kim BC, Park SS, Hong KC. Associations of the variation in the porcine myogenin gene with muscle fibre characteristics, lean meat production and meat quality traits. J. Anim. Breed. Genet. 2009; 126: 134–141. doi: 10.1111/j.1439-0388.2008.00724.x.
39. Sieczkowska H, Kocwin-Podsiadla M, Zybert A, Krzecio E, Antosik K, Kaminski S, et al. The association between polymorphism of PKM2 gene and glycolytic potential and pork meat quality. Meat Sci. 2010; 84: 180–185. doi: 10.1016/j.meatsci.2009.08.045.
40. Yang H, Xu ZY, Lei MG, Li FE, Deng CY, Xiong YZ, et al. Association of 3 polymorphisms in porcine troponin I genes (TNNI1 and TNNI2) with meat quality traits. J Appl Genet. 2010; 51: 51–57. doi: 10.1007/BF03195710.
41. Liu ET. Systems biology, integrative biology, predictive biology. Cell. 2005; 121: 505–506.
42. Schadt EE, Zhang B, Zhu J. Advances in systems biology are enhancing our understanding of disease and moving us closer to novel disease treatments. Genetica. 2009; 136: 259–269.
43. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. SCIENCE. 2004; 306: 640–643.
44. Barabsi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5: 101–113.
45. Fell D, Cornish-Bowden A. Understanding the control of metabolism: Portland press London; 1997.
46. Hollywood K, Brison DR, Goodacre R. Metabolomics. Current technologies and future trends. Proteomics. 2006; 6: 4716–4723. doi: 10.1002/pmic.200600106.
47. Wu RQ, Zhao XF, Wang ZY, Zhou M, Chen QM. Novel molecular events in oral carcinogenesis via integrative approaches. Journal of dental research. 2011; 90: 561–572.
48. Sawyers CL. The cancer biomarker problem. Nature. 2008; 452: 548–552.
49. Kadarmideen HN, Rohr P von, Janss, Luc L G. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. Mamm Genome. 2006; 17: 548–564. doi: 10.1007/s00335-005-0169-x.

50. Henderson CR. Best linear unbiased prediction of breeding values not in the model for records. J Dairy Sci. 1977; 60: 783–787.
51. Meuwissen T, Hayes B, Goddard M. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 2016; 6: 6–14.
52. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003; 165: 2259–2268.
53. Kadarmideen HN. Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. Livest Sci. 2014; 166: 232–248. doi: 10.1016/j.livsci.2014.04.028.
54. Georges M. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. Annu Rev Genomics Hum Genet. 2007; 8: 131–162. doi: 10.1146/annurev.genom.8.080706.092408.
55. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis, John P. A., et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008; 9: 356–369. doi: 10.1038/nrg2344.
56. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009; 10: 381–391. doi: 10.1038/nrg2575.
57. Pearson TA, Manolio TA. How to interpret a genome-wide association study. J Am Med Assoc. 2008; 299: 1335–1344. doi: 10.1001/jama.299.11.1335.
58. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461: 747–753. doi: 10.1038/nature08494.
59. Clarke AJ, Cooper DN. GWAS: heritability missing in action. Eur J Hum Genet. 2010; 18: 859.
60. Gibson G. Hints of hidden heritability in GWAS. Nat Genet. 2010; 42: 558–560.
61. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010; 11: 446–450.
62. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42: 565–569.

63. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011; 43: 519–525.
64. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012; 13: 135–145.
65. Duijvesteijn N, Knol EF, Merks JWM, Crooijmans RP, Am Groenen M, Bovenhuis H, et al. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet. 2010; 11: 1.
66. Luo W, Cheng D, Chen S, Wang L, Li Y, Ma X, et al. Genome-wide association analysis of meat quality traits in a porcine Large White× Minzhu intercross population. Int J Biol Sci. 2012; 8: 580.
67. Malek M, Dekkers JC, Lee HK, Baas TJ, Rothschild MF. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm Genome. 2001; 12: 630–636. doi: 10.1007/s003350020018.
68. Ma J, Yang J, Zhou L, Zhang Z, Ma H, Xie X, et al. Genome-wide association study of meat quality traits in a White Duroc× Erhualian F2 intercross and Chinese Sutai pigs. PLoS ONE. 2013; 8: e64047.
69. Zhang C, Wang Z, Bruce H, Kemp RA, Charagu P, Miar Y, et al. Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. BMC Genet. 2015; 16: 1.
70. Liu G, Kim JJ, Jonas E, Wimmers K, Ponsuksili S, Murani E, et al. Combined line-cross and half-sib QTL analysis in Duroc-Pietrain population. Mamm. Genome. 2008; 19: 429–438. doi: 10.1007/s00335-008-9132-y.
71. Edwards DB, Ernst CW, Raney NE, Doumit ME, Hoge MD, Bates RO. Quantitative trait locus mapping in an F2 Duroc x Pietrain resource population: II. Carcass and meat quality traits. J Anim Sci. 2008; 86: 254–266. doi: 10.2527/jas.2006-626.
72. Thomsen H, Lee HK, Rothschild MF, Malek M, Dekkers, J C M. Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. J Anim Sci. 2004; 82: 2213–2228.

73. Ren J, Duan Y, Qiao R, Yao F, Zhang Z, Yang B, et al. A missense mutation in PPARD causes a major QTL effect on ear size in pigs. PLoS Genet. 2011; 7: e1002043. doi: 10.1371/journal.pgen.1002043.
74. van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature. 2003; 425: 832–836. doi: 10.1038/nature02064.
75. Cole JB, VanRaden PM, O'Connell, JR, van Tassell CP, Sonstegard TS, Schnabel RD, et al. Distribution and location of genetic effects for dairy traits. J Dairy Sci. 2009; 92: 2931–2946.
76. Kadarmideen, H. N. and Reverter A. Combined Genetic, Genomic and Transcriptomic Methods in the Analysis of Animal Traits: Cab Rev; 2007.
77. Cassar-Malek I, Picard B, Bernard C, Hocquette J-F. Application of gene expression studies in livestock production systems: a European perspective. Anim Prod Sci. 2008; 48: 701–710.
78. Chiofalo B, Lo Presti V, Savoini G, D'Alessandro E, Chiofalo V, Liotta L. Nucleotides in Broiler Chicken Diet. Effect on Breast Muscles Quality. Czech J Food Sci. 2011; 29: 308–317.
79. D'Alessandro A, Rinalducci S, Marrocco C, Zolla V, Napolitano F, Zolla L. Love me tender: An Omics window on the bovine meat tenderness network. J Proteomics. 2012; 75: 4360–4380. doi: 10.1016/j.jprot.2012.02.013.
80. Damon M, Louveau I, Lefaucheur L, Lebret B, Vincent A, Leroy P, et al. Number of intramuscular adipocytes and fatty acid binding protein-4 content are significant indicators of intramuscular fat level in crossbred Large White x Duroc pigs. J Anim Sci. 2006; 84: 1083–1092.
81. Liu JS, Damon M, Guitton N, Guisle I, Ecolan P, Vincent A, et al. Differentially-expressed genes in pig longissimus muscles with contrasting levels offFat, as identified by combined transcriptomic, reverse transcription PCR, and proteomic analyses. J Agr Food Chem. 2009; 57: 3808–3817. doi: 10.1021/Jf8033144.
82. Damon M, Wyszynska-Koko J, Vincent A, Hérault F, Lebret B. Comparison of muscle transcriptome between pigs with divergent meat quality phenotypes identifies genes related to muscle metabolism and structure. PLOS ONE. 2012; 7: e33763. doi: 10.1371/journal.pone.0033763.

83. Damon M, Denieul K, Vincent A, Bonhomme N, Wyszynska-Koko J, Lebret B. Associations between muscle gene expression pattern and technological and sensory meat traits highlight new biomarkers for pork quality assessment. Meat Sci. 2013; 95: 744–754. doi: 10.1016/j.meatsci.2013.01.016.
84. Lebret B, Ecolan P, Bonhomme N, Méteau K, Prunier A. Influence of production system in local and conventional pig breeds on stress indicators at slaughter, muscle and meat traits and pork eating quality. Animal. 2015; 9: 1404–1413. doi: 10.1017/S1751731115000609.
85. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev. 1996; 13: 19–50.
86. Abdallah C, Dumas-Gaudot E, Renaut J, Sergeant K. Gel-based and gel-free quantitative proteomics approaches at a glance. Int J Plant Genomics. 2012; 2012: 494572. doi: 10.1155/2012/494572.
87. Mullen AM, Stapleton PC, Corcoran D, Hamill RM, White A. Understanding meat quality through the application of genomic and proteomic approaches. Meat Sci. 2006; 74: 3–16. doi: 10.1016/j.meatsci.2006.04.015.
88. Hollung K, Veiseth E, Jia XH, Faergestad EM, Hildrum KI. Application of proteomics to understand the molecular mechanisms behind meat quality. Meat Sci. 2007; 77: 97–104. doi: 10.1016/j.meatsci.2007.03.018.
89. Celis JE, Kruhøffer M, Gromova I, Frederiksen C, Ostergaard M, Thykjaer T, et al. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. Febs Lett. 2000; 480: 2–16.
90. Bendixen E. The use of proteomics in meat science. Meat Sci. 2005; 71: 138–149. doi: 10.1016/j.meatsci.2005.03.013.
91. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature. 2003; 425: 737–741. doi: 10.1038/nature02046.
92. Sierra, V., O'Reilly, K., White, A., Mullen, A. M., & Troy, D. J., editor. Proteolytic fragments in bovine exudate as potential tenderisation markers; 2005.
93. Jayasri K, Padmaja K, Prasad PE. Proteomics in animal health and production. IOSR J Agric Vet Sci. 2014; 7: 50–56.

94. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol. 2003; 21: 255–261. doi: 10.1038/nbt0303-255.
95. Hunter TC, Andon NL, Koller A, Yates JR, Haynes PA. The functional proteomics toolbox: methods and applications. J Chromatogr B Analyt Technol Biomed Life Sci. 2002; 782: 165–181.
96. Paredi G, Raboni S, Bendixen E, Almeida AM de, Mozzarelli A. "Muscle to meat" molecular events and technological transformations: The proteomics insight. J Proteomics. 2012; 75: 4275–4289.
97. Paredi G, Sentandreu MA, Mozzarelli A, Fadda S, Hollung K, de Almeida, A. M. Muscle and meat. New horizons and applications for proteomics on a farm to fork perspective. J Proteomics. 2013; 88: 58–82. doi: 10.1016/j.jprot.2013.01.029.
98. Silveira A, Freitas, P. F. A., César, A. S. M., Cesar, A. S. M., Antunes RC, Guimarães EC, et al. Influence of the halothane gene (HAL) on pork quality in two commercial crossbreeds. Genet Mol Res. 2011; 10: 1479–1489.
99. D'Alessandro A, Marrocco C, Zolla V, D'Andrea M, Zolla L. Meat quality of the longissimus lumborum muscle of Casertana and Large White pigs. Metabolomics and proteomics intertwined. J Proteomics. 2011; 75: 610–627. doi: 10.1016/j.jprot.2011.08.024.
100. Lametsch R, Karlsson A, Rosenvold K, Andersen HJ, Roepstorff P, Bendixen E. Postmortem proteome changes of porcine muscle related to tenderness. J Agric Food Chem. 2003; 51: 6992–6997. doi: 10.1021/jf034083p.
101. Lametsch R, Roepstorff P, Møller HS, Bendixen E. Identification of myofibrillar substrates for μ-calpain. Meat Sci. 2004; 68: 515–521. doi: 10.1016/j.meatsci.2004.03.018.
102. Laville E, Sayd T, Sante-Lhoutellier V, Morzel M, Labas R, Franck M, et al. Characterisation of PSE zones in semimembranosus pig muscle. Meat Sci. 2005; 70: 167–172. doi: 10.1016/j.meatsci.2004.12.008.
103. Huang H, Larsen MR, Karlsson AH, Pomponio L, Costa LN, Lametsch R. Gel-based phosphoproteomics analysis of sarcoplasmic proteins in postmortem porcine muscle with pH decline rate and time differences. Proteomics. 2011; 11: 4063–4076. doi: 10.1002/pmic.201100173.

104. Hwang IH, Park BY, Kim JH, Cho SH, Lee JM. Assessment of postmortem proteolysis by gel-based proteome analysis and its relationship to meat quality traits in pig longissimus. Meat Sci. 2005; 69: 79–91. doi: 10.1016/j.meatsci.2004.06.019.
105. Di Luca A, Mullen AM, Elia G, Davey G, Hamill RM. Centrifugal drip is an accessible source for protein indicators of pork ageing and water-holding capacity. Meat Sci. 2011; 88: 261–270. doi: 10.1016/j.meatsci.2010.12.033.
106. van de Wiel, D. F. M., Zhang WL. Identification of pork quality parameters by proteomics. Meat Sci. 2007; 77: 46–54. doi: 10.1016/j.meatsci.2007.04.017.
107. Kristensen L, Purslow PP. The effect of ageing on the water-holding capacity of pork: role of cytoskeletal proteins. Meat Sci. 2001; 58: 17–23.
108. Sayd T, Morzel M, Chambon C, Franck M, Figwer P, Larzul C, et al. Proteome analysis of the sarcoplasmic fraction of pig semimembranosus muscle: implications on meat color development. J Agric Food Chem. 2006; 54: 2732–2737. doi: 10.1021/jf052569v.
109. Te Pas, M. F. W., Kruijt L, Pierzchala M, Crump RE, Boeren S, Keuning E, et al. Identification of proteomic biomarkers in M. Longissimus dorsi as potential predictors of pork quality. Meat Sci. 2013; 95: 679–687. doi: 10.1016/j.meatsci.2012.12.015.
110. Fontanesi L. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. Animal Frontiers. 2016; 6: 73–79.
111. ter Kuile, B H, Westerhoff HV. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. Febs Lett. 2001; 500: 169–171.
112. Hasim A, Ma H, Mamtimin B, Abudula A, Niyaz M, Zhang LW, et al. Revealing the metabonomic variation of EC using H-1-NMR spectroscopy and its association with the clinicopathological characteristics. Mol Biol Rep. 2012; 39: 8955–8964. doi: 10.1007/s11033-012-1764-z.
113. Kühn C, Weikard R, Widmann P. Metabolomics: A pathway for improved understanding of genetic modulation of mammalian growth and tissue deposition. Manuscript. 2014.
114. Suravajhala P, Kogelman LJA, Kadarmideen HN. Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. Genet Sel Evol. 2016; 48: 1.

115. Junot C, Fenaille F, Colsch B, Bécher F. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. Mass Spectrom Rev. 2014; 33: 471–500. doi: 10.1002/mas.21401.

116. Madhavan S, Gauba R, Gusev, Robert Clarkeand Yuriy. Integrative analysis workflow for untargeted metabolomics in translational research. Metabolomics (Los Angel). 2014; 2014.

117. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Bio. 2012; 13: 263–269. doi: 10.1038/nrm3314.

118. Rohart F, Paris A, Laurent B, Canlet C, Molina J, Mercat MJ, et al. Phenotypic prediction based on metabolomic data for growing pigs from three main European breeds. J Anim Sci. 2012; 90: 4729–4740. doi: 10.2527/jas.2012-5338.

119. Bertram HC, Oksbjerg N, Young JF. NMR-based metabonomics reveals relationship between pre-slaughter exercise stress, the plasma metabolite profile at time of slaughter, and water-holding capacity in pigs. Meat Sci. 2010; 84: 108–113. doi: 10.1016/j.meatsci.2009.08.031.

120. Muroya S, Oe M, Nakajima I, Ojima K, Chikuni K. CE-TOF MS-based metabolomic profiling revealed characteristic metabolic pathways in postmortem porcine fast and slow type muscles. Meat Sci. 2014; 98: 726–735. doi: 10.1016/j.meatsci.2014.07.018.

121. Choe JH, Kim BC. Association of blood glucose, blood lactate, serum cortisol levels, muscle metabolites, muscle fiber type composition, and pork quality traits. Meat Sci. 2014; 97: 137–142. doi: 10.1016/j.meatsci.2014.01.024.

122. Choe JH, Choi MH, Ryu YC, Go GW, Choi YM, Lee SH, et al. Estimation of pork quality traits using exsanguination blood and postmortem muscle metabolites. Asian Austal J Anim. 2015; 28: 862–869. doi: 10.5713/ajas.14.0768.

123. Straadt IK, Aaslyng MD, Bertram HC. An NMR-based metabolomics study of pork from different crossbreeds and relation to sensory perception. Meat Sci. 2014; 96: 719–728. doi: 10.1016/j.meatsci.2013.10.006.

124. van der Sijde, Marijke R, Ng A, Fu J. Systems genetics: From GWAS to disease pathways. Biochim Biophys Acta. 2014; 1842: 1903–1909. doi: 10.1016/j.bbadis.2014.04.025.

125. Andersson L. Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet. 2001; 2: 130–138.

126. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology. 2013; 144: 1488-96, 1496.e1-3. doi: 10.1053/j.gastro.2013.03.001.
127. Li J, Burmeister M. Genetical genomics: combining genetics with gene expression analysis. Hum Mol Genet. 2005; 14 Spec No. 2: R163-9. doi: 10.1093/hmg/ddi267.
128. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. Trends Genet. 2001; 17: 388–391.
129. Rothschild MF, Hu Z-l, Jiang Z. Advances in QTL mapping in pigs. Int J Biol Sci. 2007; 3: 192–197.
130. Steibel JP, Bates RO, Rosa, Guilherme J. M., Tempelman RJ, Rilington VD, Ragavendran A, et al. Genome-Wide Linkage Analysis of Global Gene Expression in Loin Muscle Tissue Identifies Candidate Genes in Pigs. PLOS ONE. 2011; 6. doi: 10.1371/journal.pone.0016766.
131. Wimmers K, Murani E, Ponsuksili S. Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat performance in pigs. Briefings in functional genomics. 2010; 9: 251–258.
132. Ponsuksili S, Murani E, Schwerin M, Schellander K, Wimmers K. Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. BMC Genomics. 2010; 11: 572. doi: 10.1186/1471-2164-11-572.
133. Ponsuksili S, Murani E, Phatsara C, Jonas E, Walz C, Schwerin M, et al. Expression Profiling of Muscle Reveals Transcripts Differentially Expressed in Muscle That Affect Water-Holding Capacity of Pork. J Agr Food Chem. 2008; 56: 10311–10317. doi: 10.1021/Jf800881y.
134. Ponsuksili S, Jonas E, Murani E, Phatsara C, Srikanchai T, Walz C, et al. Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. BMC Genomics. 2008; 9. doi: 10.1186/1471-2164-9-367.
135. Ponsuksili S, Murani E, Schwerin M, Schellander K, Wimmers K. Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits. BMC Genomics. 2010; 11: 572. doi: 10.1186/1471-2164-11-572.

136. Ponsuksili S, Du Y, Murani E, Schwerin M, Wimmers K. Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. Genetics. 2012; 192: 1109–1122.
137. Lobjois V, Liaubet L, SanCristobal M, Glénisson J, Fève K, Rallières J, et al. A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. Anim Genet. 2008; 39: 147–162. doi: 10.1111/j.1365-2052.2007.01695.x.
138. Hamill RM, McBryan J, McGee C, Mullen AM, Sweeney T, Talbot A, et al. Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. Meat Sci. 2012; 92: 440–450. doi: 10.1016/j.meatsci.2012.05.007.
139. Cánovas A, Varona L, Burgos C, Galve A, Carrodeguas JA, Ibáñez-Escriche N, et al. Early postmortem gene expression and its relationship to composition and quality traits in pig Longissimus dorsi muscle. J Anim Sci. 2012; 90: 3325–3336. doi: 10.2527/jas.2011-4799.
140. Yu K, Shu G, Yuan F, Zhu X, Gao P, Wang S, et al. Fatty acid and transcriptome profiling of longissimus dorsi muscles between pig breeds differing in meat quality. Int J Biol Sci. 2013; 9: 108–118. doi: 10.7150/ijbs.5306.
141. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014; 15: 34–48.
142. Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, et al. Defining the consequences of genetic variation on a proteome-wide scale. Nature. 2016; 534: 500–505.
143. Fiehn O. Metabolomics - the link between genotypes and phenotypes. Plant Mol Biol. 2002; 48: 155–171. doi: 10.1023/A:1013713905833.
144. Welzenbach J, Neuhoff C, Heidt H, Cinar MU, Looft C, Schellander K, et al. Integrative Analysis of Metabolomic, Proteomic and Genomic Data to Reveal Functional Pathways and Candidate Genes for Drip Loss in Pigs. Int J Mol Sci. 2016; 17: 1426.
145. Zhang W, Yang B, Zhang J, Cui L, Ma J, Chen C, et al. Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations. Scientific reports. 2016; 6.

146. Kühn C. Metabolomics in animal breeding. Genetics Meets Metabolomics: Springer; 2012. pp. 107–123.
147. Widmann P, Reverter A, Fortes, Marina R. S., Weikard R, Suhre K, Hammon H, et al. A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. BMC Genomics. 2013; 14. doi: 10.1186/1471-2164-14-798.
148. Fontanesi L, Bovo S, Mazzoni G. L., Samore AB, Schiavo G, Scotti E, et al. Genome wide perspective of genetic variation in pig metabolism and production traits. 10th World Congress on Genetics Applied to Livestock Production. 2014.
149. Cole JB, Lewis RM, Maltecca C, Newman S, Olson KM, Tait RG. Systems Biology in Animal Breeding: Identifying Relationships among Markers, Genes, and Phenotypes [Breeding and Genetics Symposium]. 2013.
150. Morrison N, Cochrane G, Faruque N, Tatusova T, Tateno Y, Hancock D, et al. Concept of sample in OMICS technology. Omics. 2006; 10: 127–137.
151. MacLellan WR, Wang Y, Lusis AJ. Systems-based approaches to cardiovascular disease. Nature Reviews Cardiology. 2012; 9: 172–184.
152. Eichner J, Rosenbaum L, Wrzodek C, Häring H-U, Zell A, Lehmann R. Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software. JChromatogr. 2014; 966: 77–82.
153. Kuo T-C, Tian T-F, Tseng YJ. 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. BMC systems biology. 2013; 7: 1.
154. Vazquez AI, Veturi Y, Behring M, Shrestha S, Kirst M, Resende MFR, et al. Increased Proportion of Variance Explained and Prediction Accuracy of Survival of Breast Cancer Patients with Use of Whole-Genome Multi-omic Profiles. Genetics. 2016: genetics. 115.185181.
155. D'Alessandro A, Marrocco C, Rinalducci S, Mirasole C, Failla S, Zolla L. Chianina beef tenderness investigated through integrated Omics. J Proteomics. 2012; 75: 4381–4398.

156. Liu S, Han W, Jiang S, Zhao C, Wu C. Integrative transcriptomics and proteomics analysis of longissimus dorsi muscles of Canadian double-muscled Large White pigs. Gene. 2015. doi: 10.1016/j.gene.2015.11.016.
157. Heidt H, Cinar MU, Uddin MJ, Looft C, Juengst H, Tesfaye D, et al. A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. Mamm. Genome. 2013; 24: 416–426. doi: 10.1007/s00335-013-9473-z.
158. Terenina E, Babigumira BM, Le Mignon G, Bazovkina D, Rousseau S, Salin F, et al. Association study of molecular polymorphisms in candidate genes related to stress responses with production and meat quality traits in pigs. Domest Anim Endocrin. 2013; 44: 81–97. doi: 10.1016/j.domaniend.2012.09.004.
159. Te Pas, M. F. W., Hoekman, A. J. W., Smits MA. Biomarkers as management tools for industries in the pork production chain. J on Chain and Network Sci. 2011; 10: 155–166.
160. Bertram HC, Oksbjerg N, Young JF. NMR-based metabonomics reveals relationship between pre-slaughter exercise stress, the plasma metabolite profile at time of slaughter, and water-holding capacity in pigs. Meat Sci. 2010; 84: 108–113. doi: 10.1016/j.meatsci.2009.08.031.
161. Te Pas M, Kruijt L, Smits M, editors. Use of biomarkers as tools for tracking and tracing meat and meat products and to predict and monitor meat quality; 26.09.2012.
162. Gromping U. Variable Importance Assessment in Regression. Linear regression versus random forest. Am Stat. 2009; 63: 308–319. doi: 10.1198/tast.2009.08199.
163. Breiman L. Random forests. Mach Learn. 2001; 45: 5–32. doi: 10.1023/A:1010933404324.
164. Langfelder P, Horvath S. WGCNA. an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9. doi: 10.1186/1471-2105-9-559.
165. Honikel KO, Kim CJ. Causes of the development of PSE pork. Fleischwirtschaft. 1986; 66: 349–353.
166. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc. 2006; 1: 387–396. doi: 10.1038/nprot.2006.59.

167. Jolliffe IT, Trendafilov NT, Uddin M. A modified principal component technique based on the LASSO. J Comput Graph Stat. 2003; 12: 531–547. doi: 10.1198/1061860032148.
168. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm Genome. 2007; 18: 463–472. doi: 10.1007/s00335-007-9043-3.
169. Dong J, Horvath S. Understanding network concepts in modules. BMC Syst Biol. 2007; 1. doi: 10.1186/1752-0509-1-24.
170. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. Plos Comput Biol. 2008; 4. doi: 10.1371/journal.pcbi.1000117.
171. Horvath S. Weighted network analysis. Applications in genomics and systems biology. New York: Springer; 2011.
172. DiLeo MV, Strahan GD, den Bakker M, Hoekenga OA. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. PLoS ONE. 2011; 6. doi: 10.1371/journal.pone.0026683.
173. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures. Illustrations, sources and a solution. BMC Bioinformatics. 2007; 8. doi: 10.1186/1471-2105-8-25.
174. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008; 9. doi: 10.1186/1471-2105-9-307.
175. Nicodemus KK, Malley JD, Strobl C, Ziegler A. The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinformatics. 2010; 11. doi: 10.1186/1471-2105-11-110.
176. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. J Comput Graph Stat. 2006; 15: 651–674. doi: 10.1198/106186006X133933.
177. Strobl C, Hothorn T, Zeileis A. Party on! R J. 2009; 1: 14–17.
178. Batagelj V, Mrvar A. Pajek - Analysis and visualization of large networks. Lect Notes Comput Sc. 2002; 2265: 477–478.

179. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A. 2006; 103: 17973–17978. doi: 10.1073/pnas.0605938103.

180. Wishart DS. Advances in metabolite identification. Bioanalysis. 2011; 3: 1769–1782. doi: 10.4155/Bio.11.155.

181. Chagoyen M, Pazos F. Tools for the functional interpretation of metabolomic experiments. Brief Bioinform. 2013; 14: 737–744. doi: 10.1093/bib/bbs055.

182. Guyon I, Nikravesh M, Gunn S, Zadeh LA. Feature Extraction. Foundations and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006.

183. MacLennan NK, Dong J, Aten JE, Horvath S, Rahib L, Ornelas L, et al. Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. Mol Genet Metab. 2009; 98: 203–214. doi: 10.1016/j.ymgme.2009.05.004.

184. Boulesteix A-L, Janitza S, Kruppa J, Koenig IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov. 2012; 2: 493–507. doi: 10.1002/widm.1072.

185. Acharjee A, Kloosterman B, de Vos, R. C. H., Werij JS, Bachem, C. W. B., Visser, R. G. F., et al. Data integration and network reconstruction with similar to omics data using Random Forest regression in potato. Anal Chim Acta. 2011; 705: 56–63. doi: 10.1016/j.aca.2011.03.050.

186. Hidalgo FJ, Navarro JL, Delgado RM, Zamora R. Determination of alpha-keto acids in pork meat and Iberian ham via tandem mass spectrometry. Food Chem. 2013; 140: 183–188. doi: 10.1016/j.foodchem.2013.02.052.

187. Lambert IH, Nielsen JH, Andersen HJ, Ortenblad N. Cellular model for induction of drip loss in meat. J Agr Food Chem. 2001; 49: 4876–4883. doi: 10.1021/Jf010121y.

188. Poulsen KA, Young JF, Theil P, Kolko M, Oksbjerg N, Lambert IH. Role of phospholipase A(2) in the induction of drip loss in porcine muscle. J Agr Food Chem. 2007; 55: 1970–1976. doi: 10.1021/Jf062341n.

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

190. Otto G, Roehe R, Looft H, Thoelking L, Knap PW, Rothschild MF, et al. Associations of DNA markers with meat quality traits in pigs with emphasis on drip loss. Meat Sci. 2007; 75: 185–195.
191. Bernal Rubio, Y. L., Gualdrón Duarte, J. L., Bates RO, Ernst CW, Nonneman D, Rohrer GA, et al. Meta-analysis of genome-wide association from genomic prediction models. Anim Genet. 2015. doi: 10.1111/age.12378.
192. Welzenbach J, Neuhoff C, Looft C, Schellander K, Tholen E, Große-Brinkhaus C. Different Statistical Approaches to Investigate Porcine Muscle Metabolome Profiles to Highlight New Biomarkers for Pork Quality Assessment. PLoS ONE. 2016; 11: e0149758. doi: 10.1371/journal.pone.0149758.
193. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 2006; 38: 904–909. doi: 10.1038/ng1847.
194. Zhu M, Yu M, Zhao S. Understanding quantitative genetics in the systems biology era. Int J Biol Sci. 2009; 5: 161–170.
195. Ala-Korpela M, Kangas AJ, Inouye M. Genome-wide association studies and systems biology: together at last. Trends Genet. 2011; 27: 493–498. doi: 10.1016/j.tig.2011.09.002.
196. Michaud J, Simpson KM, Escher R, Buchet-Poyau K, Beissbarth T, Carmichael C, et al. Integrative analysis of RUNX1 downstream pathways and target genes. BMC Genomics. 2008; 9: 363. doi: 10.1186/1471-2164-9-363.
197. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering statistically significant pathways in expression profiling studies. Proc. Natl. Acad. Sci. U.S.A. 2005; 102: 13544–13549. doi: 10.1073/pnas.0506577102.
198. Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. BMC Bioinformatics. 2009; 10: 47.
199. Binke R, Schwagele F. The quantification of animal and plant species in meat products - Development of an efficient method by means of PCR. Fleischwirtschaft. 2003; 83: 135–139.
200. Lim DG, Jo C, Seo KS, Nam KC. Comparison of meat quality of loins and butts in different two-way crossbred pigs. Livest Sci. 2014; 161: 210–217. doi: 10.1016/j.livsci.2013.12.015.

201. Davoli R, Bigi D, Fontanesi L, Zambonelli P, Yerle M, Zijlstra C, et al. Mapping of 14 expressed sequence tags (ESTs) from porcine skeletal muscle by somatic cell hybrid analysis. Anim Genet. 2000; 31: 400–403.
202. Laville E, Sayd T, Terlouw C, Chambon C, Damon M, Larzul C, et al. Comparison of sarcoplasmic proteomes between two groups of pig muscles selected for shear force of cooked meat. J Agric Food Chem. 2007; 55: 5834–5841. doi: 10.1021/jf070462x.
203. Nam J, Lee D-G, Kwon J, Choi C-W, et al. Comparative Proteome Analysis of Porcine Longissimus dorsi on the Basis of pH24 of Post-mortem Muscle. J Agr Sci. 2012.
204. Corcoran MP, Lamon-Fava S, Fielding RA. Skeletal muscle lipid deposition and insulin resistance: effect of dietary fatty acids and exercise. Am J Clin Nutr. 2007; 85: 662–677.
205. Dobrowsky RT, Kolesnick RN. Analysis of sphingomyelin and ceramide levels and the enzymes regulating their metabolism in response to cell stress. Methods Cell Biol. 2001; 66: 135–165.
206. Aulchenko YS, Ripke S, Isaacs A, Van Duijn, Cornelia M. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23: 1294–1296. doi: 10.1093/bioinformatics/btm108.
207. Becker D, Wimmers K, Luther H, Hofer A, Leeb T. A genome-wide association study to detect QTL for commercially important traits in Swiss Large White boars. PLoS ONE. 2013; 8: e55951. doi: 10.1371/journal.pone.0055951.
208. Utsunomiya YT, do Carmo, Adriana S, Carvalheiro R, Neves, Haroldo H R, Matos MC, Zavarez LB, et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 2013; 14: 52. doi: 10.1186/1471-2156-14-52.
209. Han B, Kang HM, Eskin E. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. PLoS Genet. 2009; 5: e1000456. doi: 10.1371/journal.pgen.1000456.
210. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U.S.A. 2003; 100: 9440–9445. doi: 10.1073/pnas.1530509100.
211. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful appraoch to multiple testing. J R Stat Soc Series B Stat Methodol. 1995; 57: 289–300.

212. Jennen DG, Brings AD, Liu G, Jüngst H, Tholen E, Jonas E, et al. Genetic aspects concerning drip loss and water-holding capacity of porcine meat. J. Anim. Breed. Genet. 2007; 124: 2–11.
213. Qi YM, Lei T, Zhou L, Chen XD, Long QQ, Long H, et al. Genomic organization, alternative splicing and tissues expression of porcine CREB3L4 gene. Mol. Biol. Rep. 2009; 36: 1881–1888. doi: 10.1007/s11033-008-9394-1.
214. Laramie JM, Wilk JB, Williamson SL, Nagle MW, Latourelle JC, Tobin JE, et al. Polymorphisms near EXOC4 and LRGUK on chromosome 7q32 are associated with Type 2 Diabetes and fasting glucose; the NHLBI Family Heart Study. BMC Med Genet. 2008; 9: 46. doi: 10.1186/1471-2350-9-46.
215. Leheska JM, Wulf DM, Maddock RJ. Effects of fasting and transportation on pork quality development and extent of postmortem metabolism. J Anim Sci. 2002; 80: 3194–3202.
216. Hamet P, Tremblay J, Seda O, Macmahon S, Chalmers J. Genetic component of complications in type 2 diabetes: Google Patents; 2009.
217. Casas-Agustench P, Arnett DK, Smith CE, Lai C-Q, Parnell LD, Borecki IB, et al. Saturated fat intake modulates the association between an obesity genetic risk score and body mass index in two US populations. J Acad Nutr Diet. 2014; 114: 1954–1966. doi: 10.1016/j.jand.2014.03.014.
218. Liu Q, Zhang J, Guo W, Zhao Y, Hu X, Li N. Identifying lipid metabolism genes in pig liver after clenbuterol administration. Front Biosci (Elite Ed). 2012; 4: 2605–2616.
219. Griffiths B, Lewis CA, Bensaad K, Ros S, Zhang Q, Ferber EC, et al. Sterol regulatory element binding protein-dependent regulation of lipid synthesis supports cell survival and tumor growth. Cancer Metab. 2013; 1: 3. doi: 10.1186/2049-3002-1-3.
220. Kim S-Y, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC Bioinformatics. 2005; 6: 144. doi: 10.1186/1471-2105-6-144.
221. Hirose K, Ito T, Fukawa K, Arakawa A, Mikawa S, Hayashi Y, et al. Evaluation of effects of multiple candidate genes (LEP, LEPR, MC4R, PIK3C3, and VRTN) on production traits in Duroc pigs. J Anim Sci. 2014; 85: 198–206.
222. Zhao X, Liu Z-Y, Liu Q-X. Gene coexpression networks reveal key drivers of phenotypic divergence in porcine muscle. BMC Genomics. 2015; 16: 50. doi: 10.1186/s12864-015-1238-5.

223. Fontanesi L, Schiavo G, Galimberti G, Calò DG, Russo V. A genomewide association study for average daily gain in Italian Large White pigs. J Anim Sci. 2014; 92: 1385–1394. doi: 10.2527/jas.2013-7059.
224. Feng X, Scott A, Wang Y, Wang L, Zhao Y, Doerner S, et al. PTPRT regulates high-fat diet-induced obesity and insulin resistance. PLoS ONE. 2014; 9: e100783. doi: 10.1371/journal.pone.0100783.
225. Scott A. Role of PTPRT in obesity and its substrate paxillin tyrosine-88 in colorectal cancer. Cleveland, Ohio: Doctoral Thesis, Case Western Reserve University; 2014.
226. Tizioto PC, Decker JE, Taylor JF, Schnabel RD, Mudadu MA, Silva FL, et al. Genome scan for meat quality traits in Nelore beef cattle. Physiol Genomics. 2013; 45: 1012–1020. doi: 10.1152/physiolgenomics.00066.2013.
227. Krastanov A. Metabolomics - the State of Art. Biotechnol Biotec Eq. 2010; 24: 1537–1543. doi: 10.2478/V10133-010-0001-Y.
228. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer; 2005. pp. 397–420.
229. Kellermann J. ICPL—isotope-coded protein label. 2D Page: Sample Preparation and Fragmentation. pp. 113–123.
230. Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen, Martien A. M., Harlizius B, et al. Pig genome sequence - analysis and publication strategy. BMC Genomics. 2010; 11. doi: 10.1186/1471-2164-11-438.
231. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic data sets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4: 1184–1191. doi: 10.1038/nprot.2009.97.
232. Stuart A, Ord JK, Kendall MG. Kendall's advanced theory of statistics. 5th ed. London: Edward Arnold; 1991.
233. Fan H, Wu Y, Zhou X, Xia J, Zhang W, Song Y, et al. Pathway-Based Genome-Wide Association Studies for Two Meat Production Traits in Simmental Cattle. Sci Rep. 2015; 5: 18389. doi: 10.1038/srep18389.
234. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014; 15: 478. doi: 10.1186/1471-2164-15-478.

235. Aulchenko YS, Koning D-J de, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics. 2007; 177: 577–585.
236. Haley C, Koning DJ de. Genetical genomics in livestock: potentials and pitfalls. Anim Genet. 2006; 37: 10–12.
237. Saris CGJ, Horvath S, van Vught, Paul WJ, van Es, Michael A, Blauw HM, Fuller TF, et al. Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. BMC Genomics. 2009; 10: 405.
238. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet. 2011; 7: e1002367.
239. Hyvarinen A, Oja E. Independent component analysis: algorithms and applications, vol. 13. Neural networks. 2000.
240. Peñagaricano F, Valente BD, Steibel JP, Bates RO, Ernst CW, Khatib H, et al. Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. BMC systems biology. 2015; 9: 58.
241. Pilsner JR. Metabolic interactions between arsenic, folate, and selenium and their impact on genomic DNA methylation: COLUMBIA UNIVERSITY; 2007.
242. Zhan X, Wang M, Zhao R, Li W, Xu Z. Effects of different selenium source on selenium distribution, loin quality and antioxidant status in finishing pigs. Anim Feed Sci Tech. 2007; 132: 202–211.
243. Torrent J. Selenium yeast and pork quality. Biotechnology in the Feed Industry. 1996; 1996: 161.
244. Mahan DC, Cline TR, Richert B. Effects of dietary levels of selenium-enriched yeast and sodium selenite as selenium sources fed to growing-finishing pigs on performance, tissue selenium, serum glutathione peroxidase activity, carcass characteristics, and loin quality. J Anim Sci. 1999; 77: 2172–2179.
245. Traore S, Aubry L, Gatellier P, Przybylski W, Jaworska D, Kajak-Siemaszko K, et al. Higher drip loss is associated with protein oxidation. Meat Sci. 2012; 90: 917–924.
246. Warner RD, Dunshea FR, Ponnampalam EN, Cottrell JJ. Effects of nitric oxide and oxidation in vivo and postmortem on meat tenderness. Meat Sci. 2005; 71: 205–217.

247. Ponsuksili S, Siengdee P, Du Y, Trakooljul N, Murani E, Schwerin M, et al. Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. PLOS ONE. 2015; 10: e0123678.
248. Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nature genetics. 2012; 44: 841–847.
249. Kogelman LJA, Zhernakova DV, Westra H-J, Cirera S, Fredholm M, Franke L, et al. An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. Genome medicine. 2015; 7: 105.
250. Buitenhuis B, Poulsen NA, Gebreyesus G, Larsen LB. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC genetics. 2016; 17: 114.
251. Bergfelder-Drüing S, Grosse-Brinkhaus C, Lind B, Erbe M, Schellander K, Simianer H, et al. A genome-wide association study in large white and landrace pig populations for number piglets born alive. PLOS ONE. 2015; 10: e0117468.
252. Glinka A, Dolde C, Kirsch N, Huang Y, Kazanskaya O, Ingelfinger D, et al. LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling. EMBO reports. 2011; 12: 1055–1061.
253. Goodison S, Urquidi V, Tarin D. CD44 cell adhesion molecules. Mol Pathol Lib. 1999; 52: 189.
254. Damon M, Denieul K, Vincent A, Bonhomme N, Wyszynska-Koko J, Lebret B. Associations between muscle gene expression pattern and technological and sensory meat traits highlight new biomarkers for pork quality assessment. Meat Sci. 2013; 95: 744–754. doi: 10.1016/j.meatsci.2013.01.016.
255. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV. Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. The Journal of cell biology. 1997; 136: 729–743.
256. Dalpé G, Mathieu M, Comtois A, Zhu E, Wasiak S, Repentigny Y de, et al. Dystonin-deficient mice exhibit an intrinsic muscle weakness and an instability of skeletal muscle cytoarchitecture. Developmental biology. 1999; 210: 367–380.

257. Damon M, Wyszynska-Koko J, Vincent A, Herault F, Lebret B. Comparison of muscle transcriptome between pigs with divergent meat quality phenotypes identifies genes related to muscle metabolism and structure. PLOS ONE. 2012; 7: e33763.
258. Herault F, Vincent A, Dameron O, Le Roy P, Cherel P, Damon M. The longissimus and semimembranosus muscles display marked differences in their gene expression profiles in pig. PLOS ONE. 2014; 9: e96491.
259. Winkler GS. The mammalian anti-proliferative BTG/Tob protein family. J Cell Physiol. 2010; 222: 66–72. doi: 10.1002/jcp.21919.
260. Mo XY, Lan J, Jiao QZ, Xiong YZ, Zuo B, Li FE, et al. Molecular characterization, expression pattern and association analysis of the porcine BTG2 gene. Mol Biol Rep. 2011; 38: 4389–4396. doi: 10.1007/s11033-010-0566-4.
261. Sun WX, Wang HH, Jiang BC, Zhao YY, Xie ZR, Xiong K, et al. Global comparison of gene expression between subcutaneous and intramuscular adipose tissue of mature Erhualian pig. Genet Mol Res. 2013; 12: 5085–5101. doi: 10.4238/2013.October.29.3.
262. Lkhagvadorj S, Qu L, Cai W, Couture OP, Barb CR, Hausman GJ, et al. Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant. Physiol Genomics. 2009; 38: 98–111. doi: 10.1152/physiolgenomics.90372.2008.
263. SanCristobal M, Rohart F, Lascor C, Bouffaud M, Trouilh L, Martin, Pascal G P, et al. Exploring transcriptomic diversity in muscle revealed that cellular signaling pathways mainly differentiate five Western porcine breeds. BMC Genomics. 2015; 16: 1055. doi: 10.1186/s12864-015-2259-9.
264. Zhao X, Mo D, Li A, Gong W, Xiao S, Zhang Y, et al. Comparative analyses by sequencing of transcriptomes during skeletal muscle development between pig breeds differing in muscle growth rate and fatness. PLOS ONE. 2011; 6: e19774. doi: 10.1371/journal.pone.0019774.
265. Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 1997: 119–139.
266. Gromski PS, Xu Y, Correa E, Ellis DI, Turner ML, Goodacre R. A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data. Analytica chimica acta. 2014; 829: 1–8.

267. Chen T, Cao Y, Zhang Y, Liu J, Bao Y, Wang C, et al. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. Evidence-Based Complementary and Alternative Medicine. 2013; 2013.
268. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996: 267–288.
269. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005; 67: 301–320.
270. Pineda S, Real FX, Kogevinas M, Carrato A, Chanock SJ, Malats N, et al. Integration analysis of three omics data using penalized regression methods: An application to bladder cancer. PLoS Genet. 2015; 11: e1005689.
271. Cho S, Kim K, Kim YJ, Lee J, Cho YS, Lee J, et al. Joint Identification of Multiple Genetic Variants via Elastic-Net Variable Selection in a Genome-Wide Association Analysis. Annals of human genetics. 2010; 74: 416–428.
272. Pineda S, Milne RL, Calle ML, Rothman N, de Maturana, Evangelina López, Herranz J, et al. Genetic variation in the TP53 pathway and bladder cancer risk. a comprehensive analysis. PLOS ONE. 2014; 9: e89952.
273. Mankoo PK, Shen R, Schultz N, Levine DA, Sander C. Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. PLOS ONE. 2011; 6: e24709.
274. Boulesteix A-L, Bin R de, Jiang X, Fuchs M. IPF-LASSO: integrative L1-penalized regression with penalty factors for prediction based on multi-omics data. 2015.
275. Yugi K, Kubota H, Hatano A, Kuroda S. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple ‘Omic’Layers. Trends in biotechnology. 2016; 34: 276–290.

Chapter 8. Appendix

S1: The random forest regression procedure of Breiman can be subdivided into a series of six steps.

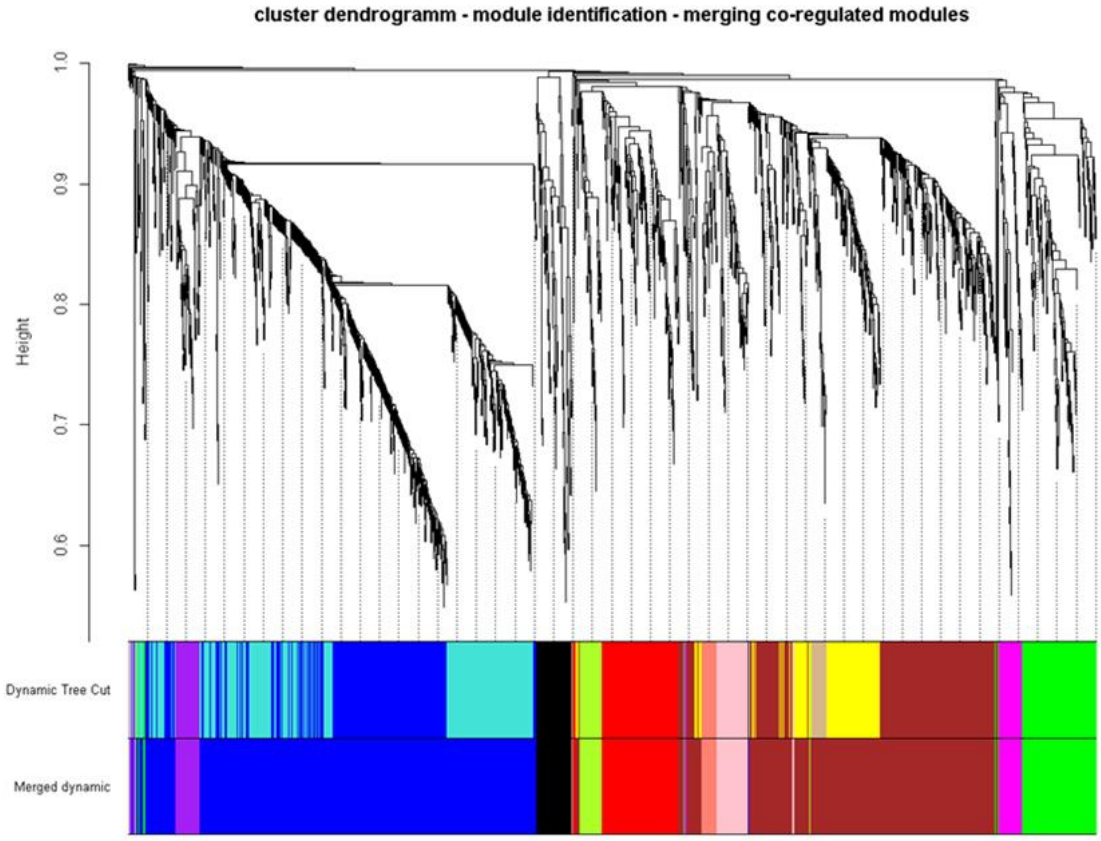
1. Sampling randomly selected subsets of meat quality observations and subsets of metabolite profiles via bootstrapping. Two-thirds of the data goes into the train data to construct the tree; One-third of the data ('Out-of-bag' (OOB) data) is used to estimate the OOB error of the grown tree (validation strep)
2. At each split randomly selecting a subset of predictors ('mtry') from the train data.
3. Growing a single regression trees by recursively splitting the subset of metabolites in the subset of predictors. At each node, split the data using the best predictor out of the subset of 'mtry' predictors. Tree construction is stopped when tree growth stopping criteria are fulfilled.
4. Estimating the OOB error by applying the tree to the OOB data. The resulting parameters 'coefficient of determination' (R^2) and 'root mean square error' (RMSE) express the suitability of the tree for prediction of meat quality in independent samples.
5. Generating a random forest as collection of trees by repeating the steps 1-4 'ntree' times.
6. Aggregation of the trees and, based on the entire forest, measurement of the final variable importance (VI) values of the metabolites. Additionally, R^2 and RMSE are averaged over the forest and represent the final parameters of prediction accuracy.

S2: Differences between the traditional random forest regression of Breiman and conditional inference forests.

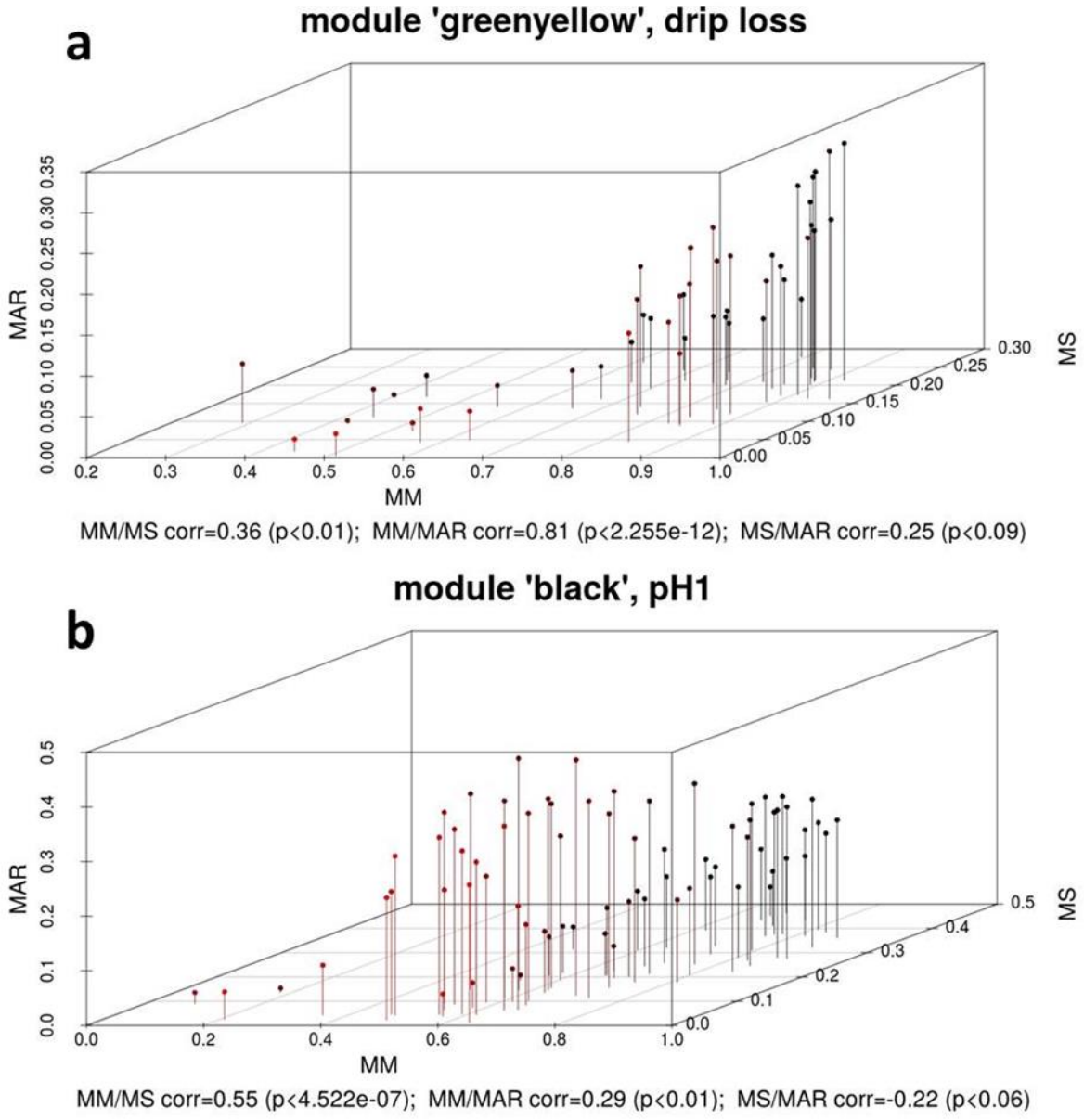
In essence, random forest regression (RFR) algorithm of Breiman [163] and Hothorn et al. [176] differs with respect to the a) splitting criteria, b) the resampling scheme and c) the way the predictions of each tree are aggregated to produce a coincident prediction.

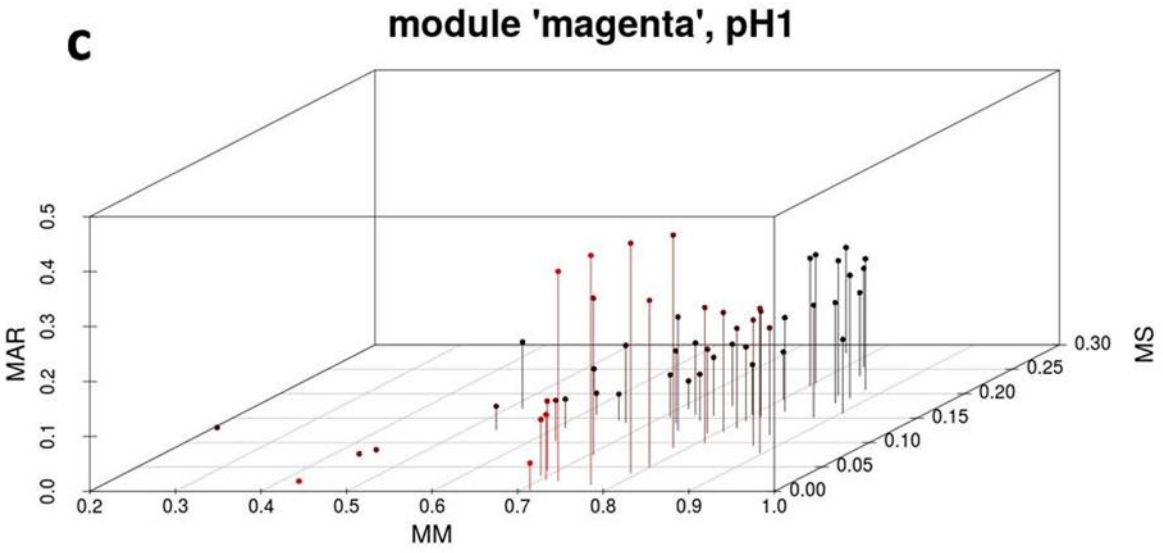
a) Splitting criteria: In contrast to Breiman´s RFR, the conditional RFR according to Hothorn uses the 'conditional inference forest' (CIF) methodology as splitting criterion. At each splitting node, each predictor is globally tested for its association with the trait of interest and a p-value is computed. Hence, CIF splitting is based on an essentially unbiased splitting criterion that automatically adjusts for different marginal distributions of the predictors and thus does not share the pitfall of Breiman´s RFR.
b) Resampling scheme: The resampling scheme in conditional RFR based on subsampling instead of bootstrap sampling and Strobl et al. [173] recommend to systematically use sampling without replacement to prevent biases in VI measurement.
c) Aggregation procedure: In the conditional RFR it works by averaging the observation weights extracted from each of the trees and not by averaging predictions directly (majority voting).

S3: Module identification in weighted network analysis based on a cluster dendrogram and merging of co-regulated modules.

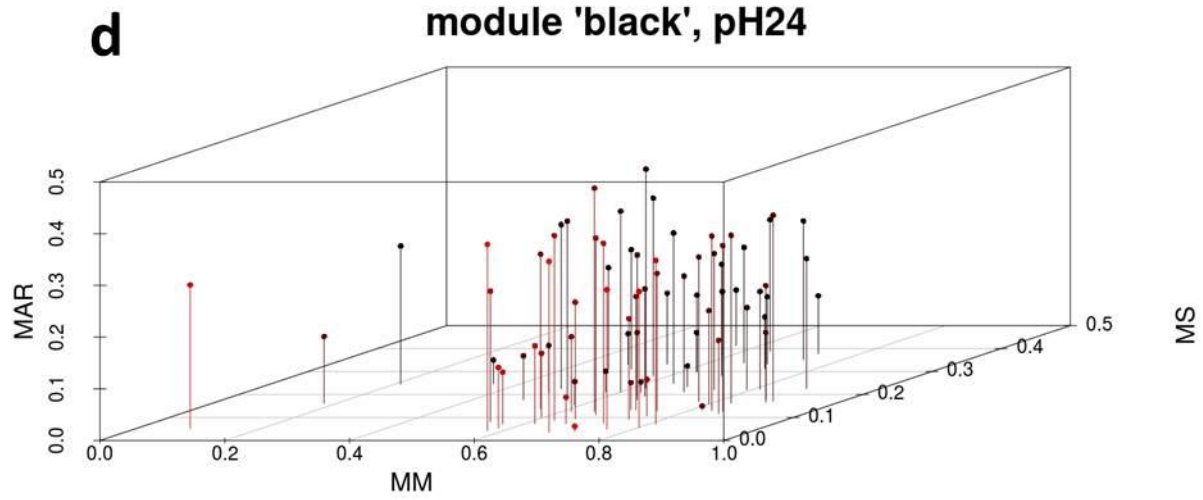


S4: Scatterplot of parameters metabolite significance, module membership and maximum adjacency ratio of the modules 'greenyellow' (a), 'black' (b, d) and 'magenta' (c) that are significantly correlated with meat quality traits drip loss (a), pH1 (b, c) and pH24 (d). Metabolite significance = MS; module membership = MM; maximum adjacency ratio = MAR; corr = Pearson correlation coefficient; drip loss measured in *Musculus longissimus dorsi* (MLD) 24 h post-mortem (p.m.); pH1 measured in LD 45 minutes p.m.; pH24 measured in MLD 24 h p.m.



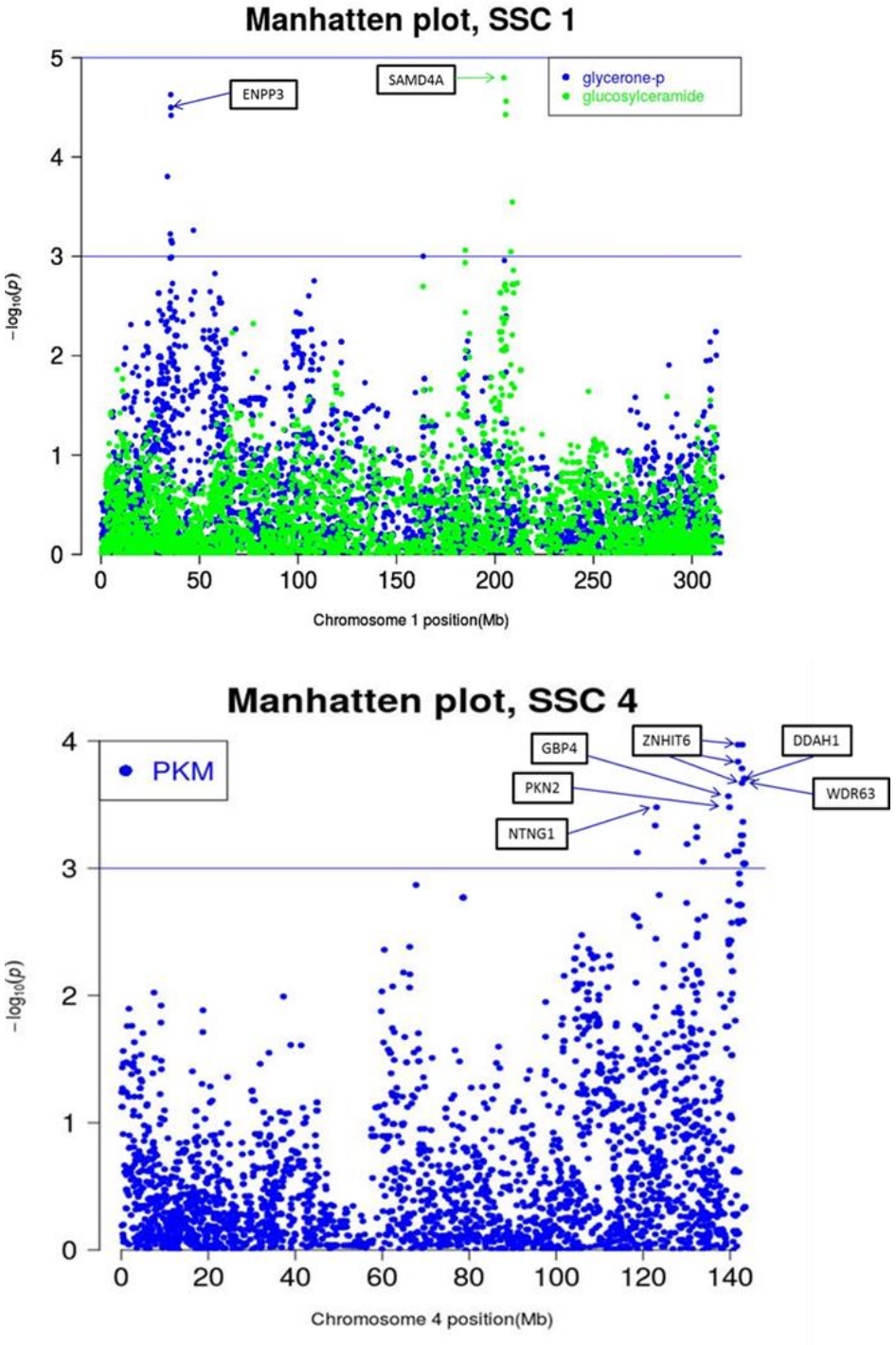


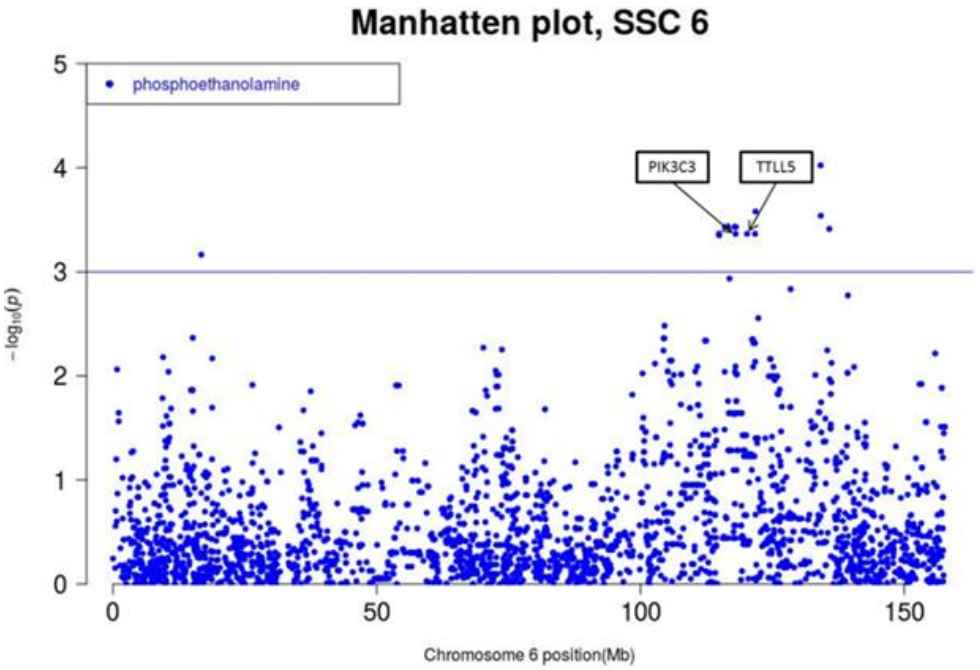
MM/MS corr=0.26 (p<0.06); MM/MAR corr=0.60 (p<2.067e-06); MS/MAR corr=-0.14 (p<0.32)

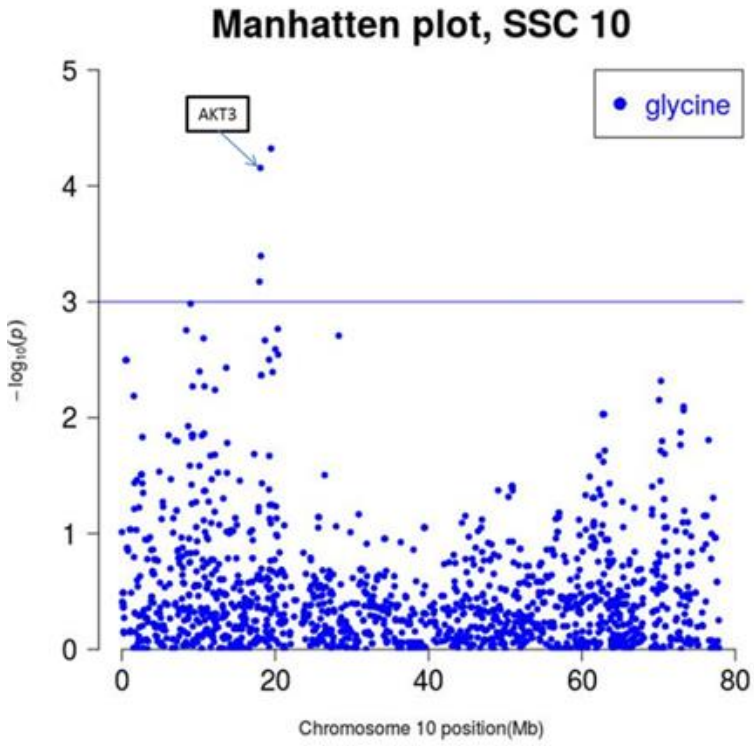


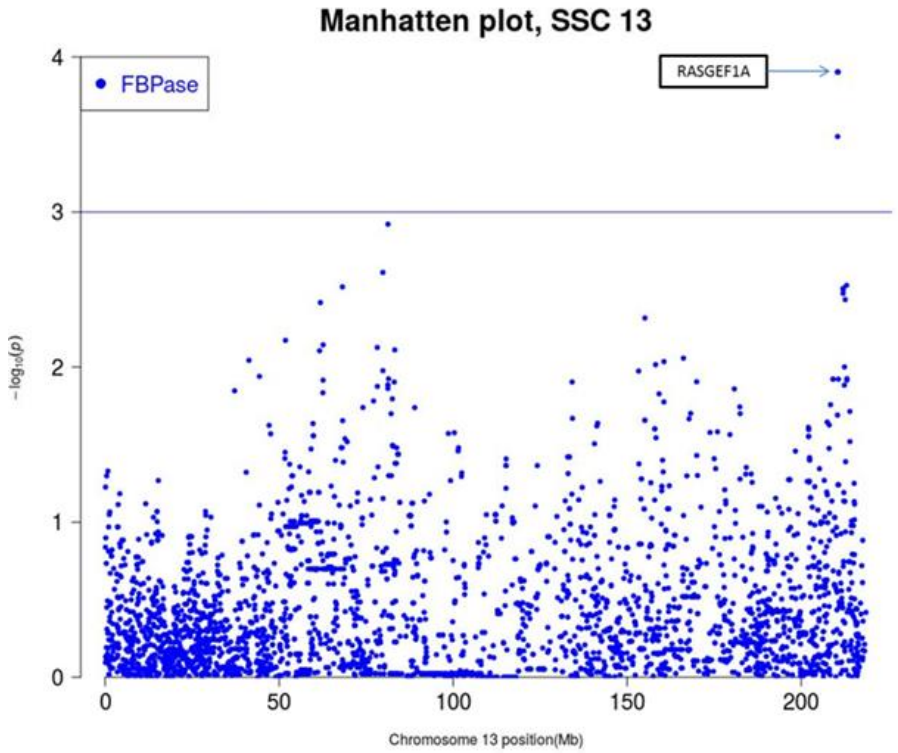
MM/MS corr=-0.13 (p<0.29); MM/MAR corr=0.29 (p<0.01); MS/MAR corr=0.50 (p<5.883e-06)

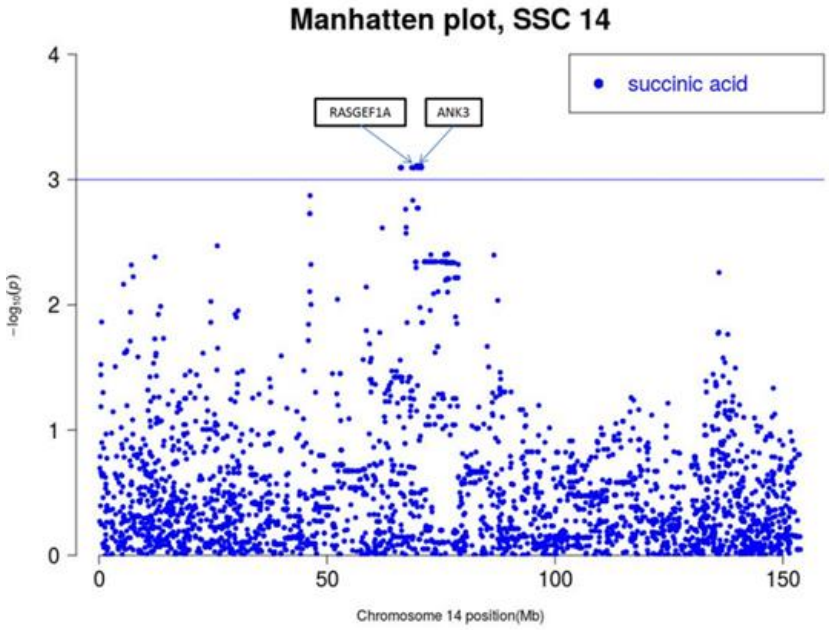
S5: Chromosome-wide Manhattan plots of *sus scrofa* chromosomes (SSC) 1, 4, 6, 10, 13, 14 and 17. Glycerone-p = dihydroxyacetone phosphate; DG3P = d-glycerate-3-phosphate; PKM = pyruvate kinase (muscle); FBPase = fructose-1,6-bisphosphatase 2; the declaration of gene symbols (in black lettering) can be obtained from Ensembl or http://www.ncbi.nlm.nih.gov/genegenes; Continued on following pages.

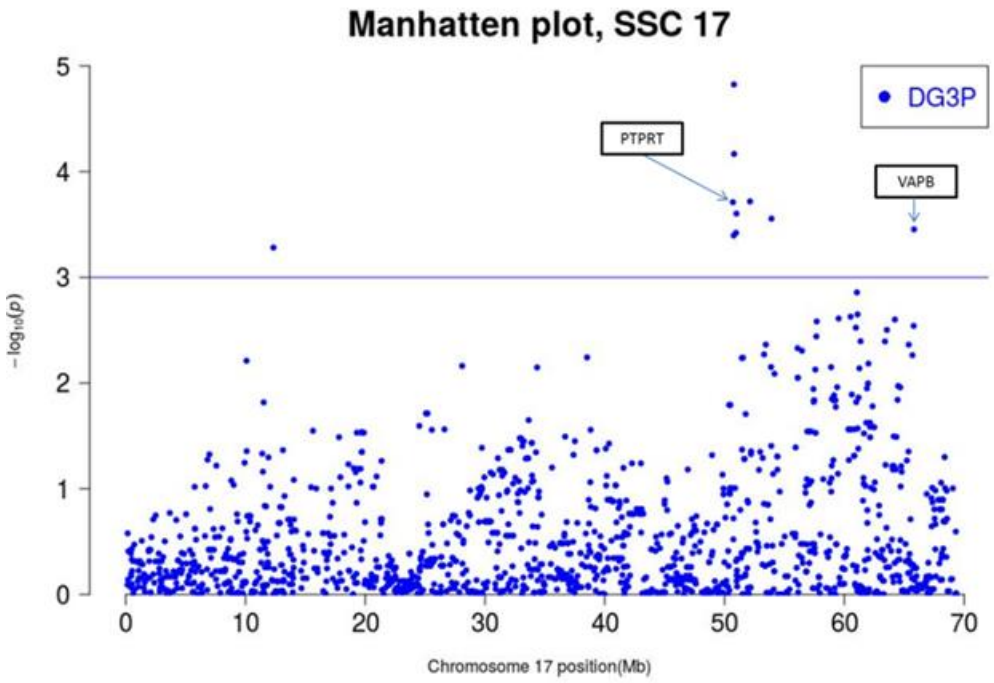



Manhatten plot, SSC 6
phosphoethanolamine
PIK3C3
TTLL5
$-\log_{10}(p)$
Chromosome 6 position(Mb)

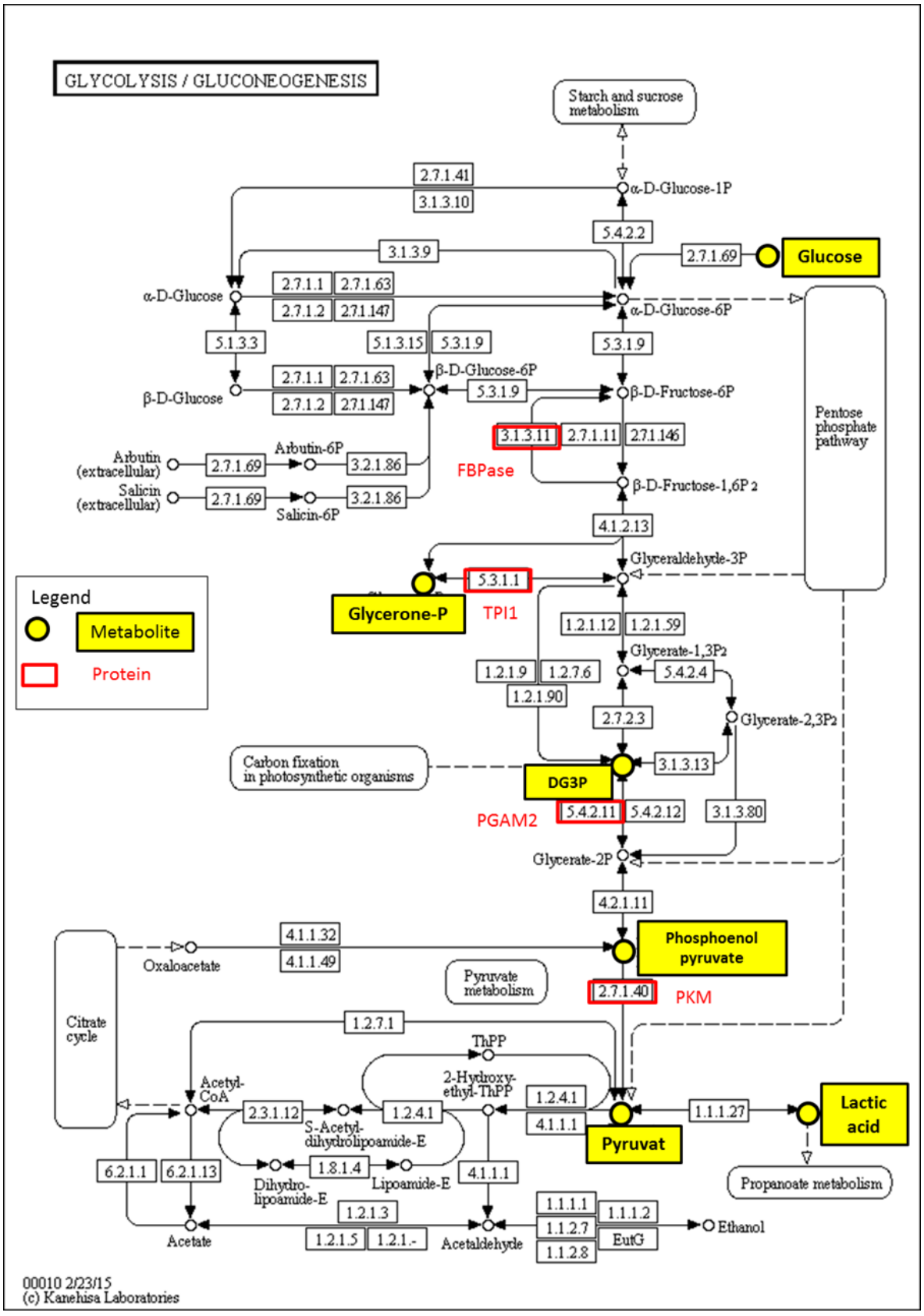

Manhatten plot, SSC 10
glycine
AKT3
$-\log_{10}(p)$
Chromosome 10 position(Mb)


Manhatten plot, SSC 13
FBPase
RASGEF1A
$-\log_{10}(p)$
0
1
2
3
4
0
50
100
150
200
Chromosome 13 position(Mb)

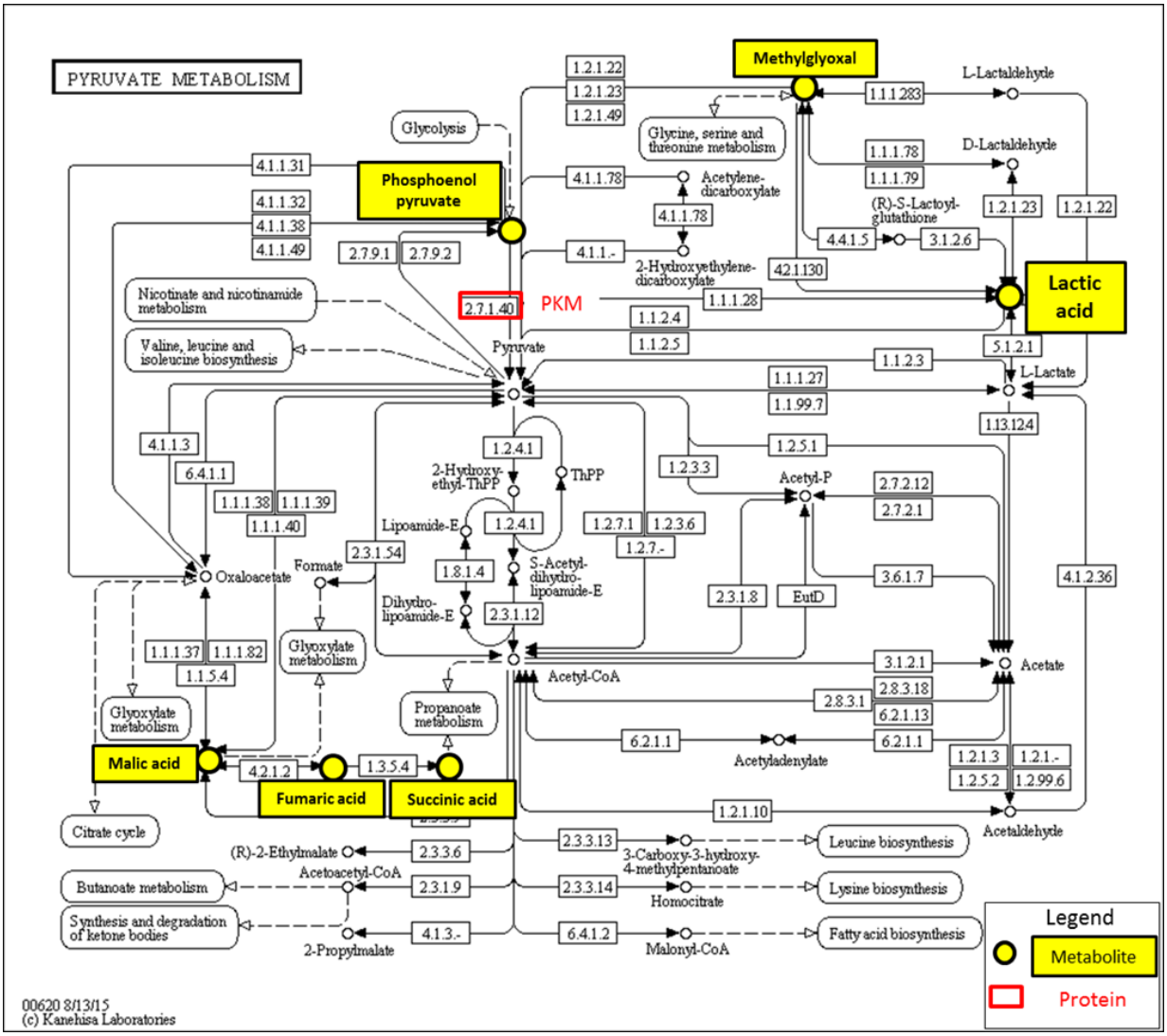

Manhatten plot, SSC 14
succinic acid
RASGEF1A
ANK3
$-\log_{10}(p)$
0
1
2
3
4
0
50
100
150
Chromosome 14 position(Mb)


Manhatten plot, SSC 17
DG3P
PTPRT
VAPB
5
4
3
2
1
0
0
10
20
30
40
50
60
70
Chromosome 17 position(Mb)

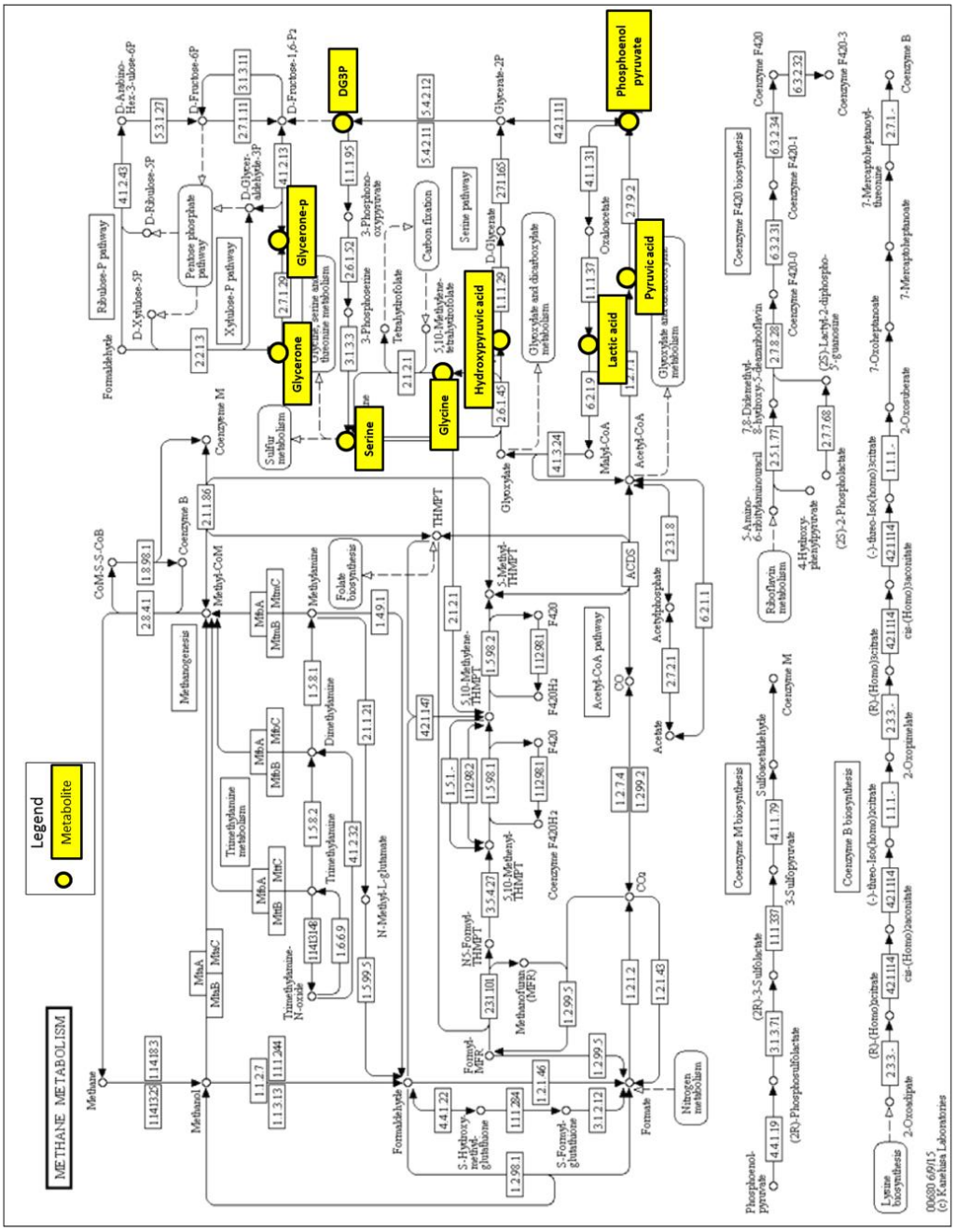
S6: KEGG pathway 'Glycolysis/Glyconeogenesis' (map00010). Metabolites and proteins revealed through the enrichment analysis are illustrated in yellow and red; Glycerone-p = dihydroxyacetone phosphate; PGAM2 = phosphoglycerate mutase 2 (muscle); PKM = pyruvate kinase (muscle); FBPase = fructose-1,6-bisphosphatase 2; TPI1 - triosephosphate isomerase 1; DG3P = d-glycerate-3-phosphate; The declaration of the other metabolites and proteins codes (in boxes) can be obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG)-database (http://www.genome.jp/kegg/).


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
α-D-Glucose
α-D-Glucose-6P
β-D-Glucose
β-D-Glucose-6P
β-D-Fructose-6P
β-D-Fructose-1,6P2
Glucose
Pentose phosphate pathway
Arbutin (extracellular)
Arbutin-6P
Salicin (extracellular)
Salicin-6P
FBPase
Glyceraldehyde-3P
Glycerone-P
TPI1
Glycerate-1,3P2
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
DG3P
PGAM2
Glycerate-2P
Phosphoenol pyruvate
Oxaloacetate
Pyruvate metabolism
PKM
Citrate cycle
ThPP
2-Hydroxy-ethyl-ThPP
Acetyl-CoA
S-Acetyl-dihydrolipoamide-E
Dihydro-lipoamide-E
Lipoamide-E
Pyruvat
Lactic acid
Propanoate metabolism
Acetate
Acetaldehyde
Ethanol
EutG
Legend
Metabolite
Protein
00010 2/23/15
(c) Kanehisa Laboratories

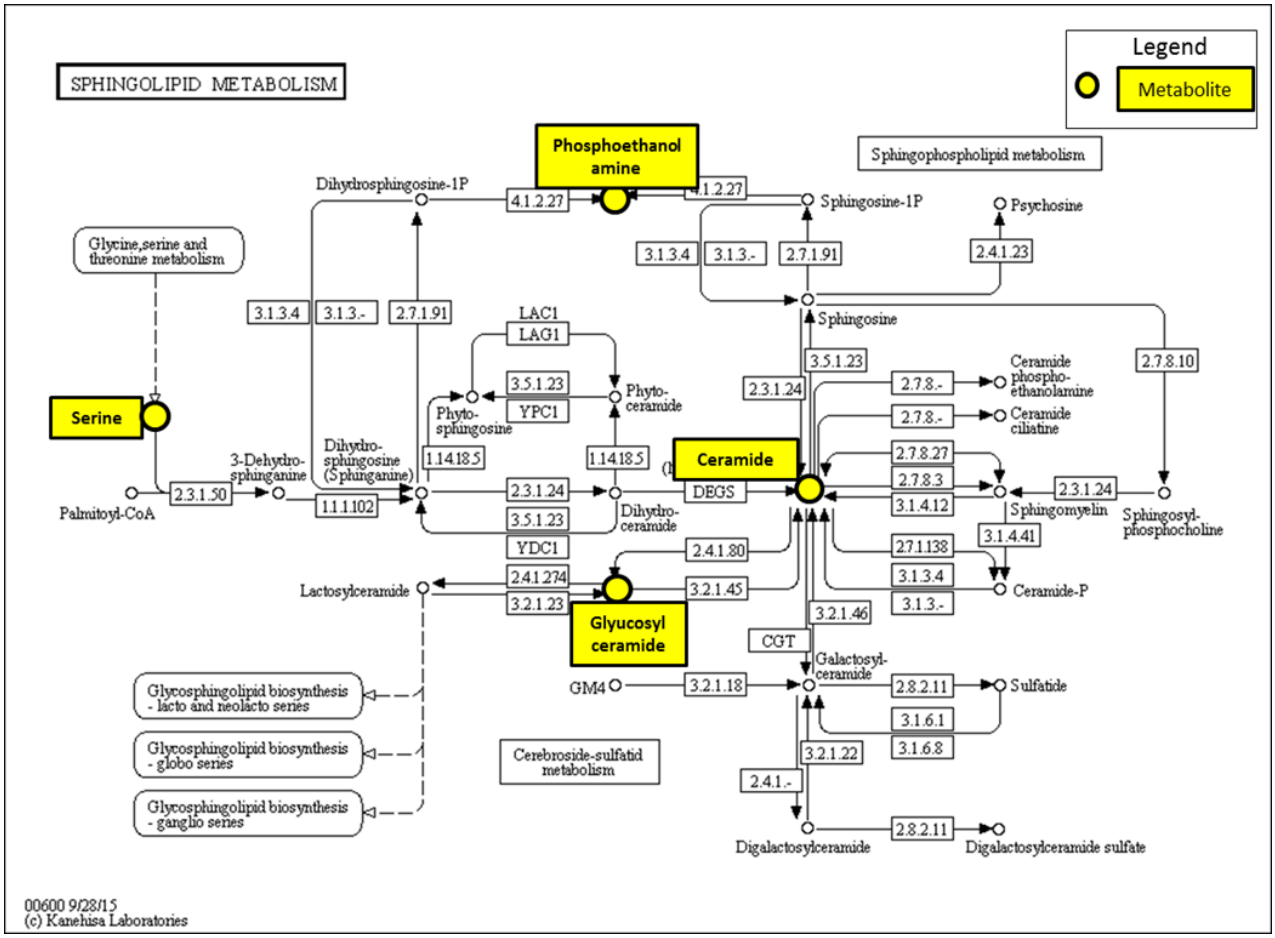
S7: KEGG pathway 'Pyruvate metabolism' (map00620). Metabolites and proteins revealed through the enrichment analysis are illustrated in yellow and red; PKM = pyruvate kinase (muscle); The declaration of the other metabolites and proteins codes (in boxes) can be obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG)-database (http://www.genome.jp/kegg/).



S8: KEGG pathway 'Methane metabolism' (map00680). Metabolites and proteins revealed through the enrichment analysis are illustrated in yellow and red; Glycerone-p = dihydroxyacetone phosphate; DG3P = d-glycerate-3-phosphate. The declaration of the other metabolites and proteins codes (in boxes) can be obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG)-database (http://www.genome.jp/kegg/).



S9: KEGG pathway 'Sphingolipid metabolism' (map00600). Metabolites and proteins revealed through the enrichment analysis are illustrated in yellow and red; The declaration of the other metabolites and proteins codes (in boxes) can be obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG)-database (http://www.genome.jp/kegg/



Acknowledgement – Danksagung

Dankeschön!

Herzlich bedanken möchte ich mich bei Herrn Prof. Dr. Karl Schellander für die Möglichkeit am Institut für Tierwissenschaften zu promovieren und für die Überlassung des Themas meiner Dissertation.

Mein großer Dank gilt der wundervollen Arbeitsgruppe Haustiergenetik:

Ich möchte mich bei meinem Doktorvater Dr. Ernst Tholen bedanken, der mich täglich bestärkt und motiviert hat, die mitunter herausfordernden und „nervenaufreibenden“ Aufgaben zu bewältigen. Ernst, danke für dein Vertrauen in mich und deine Hartnäckigkeit mir auszureden, "Keine Ahnung" zu sagen. Ich bin sehr dankbar für alles, was ich von dir gelernt habe – sowohl fachlich als auch für's Leben.

Oft werde ich an die tolle Zeit in unserem "Kranich-Nest" mit Dr. Christine Große Brinkhaus und Esther Heuß zurückdenken.

...Tine, tausend Dank für deine Unterstützung und Geduld mit mir. Du standest immer hinter mir und ich habe jeden Tag etwas von dir gelernt. Deinen berühmten Satz, "Alles wird gut!", habe ich mittlerweile verinnerlicht und du bist maßgeblich dafür verantwortlich, dass ich mit einem guten Gefühl in die Zeit als Postdoc starten kann.

...Esther, meine Leidensgenossin, ohne dich wäre die Zeit am ITW nur halb so schön gewesen! Danke für deinen Zuspruch, deine Unterstützung und die Tatsache, dass wir immer für einander da waren. Ich hoffe wir verlieren uns nie aus den Augen und können gelegentliche Kostbar-Café Orange-Pausen beibehalten.

Danke Peter, dass du in technischen "Notfällen" stets zur Stelle warst und für all die schönen Kaffeepausen, die ich sehr vermissen werde.

Liebe Bianca, vielen Dank für deinen Einsatz und deine Hilfe in allen bürokratischen Fragen.

Vielen Dank Dr. Sarah Bergfelder-Drühing, du hast mir den Start am ITW erleichtert und mir vom ersten Tag an beigestanden. Ines Brinke, für deine moralische Unterstützung in der finalen Phase meine Doktorarbeit möchte ich mich herzlich bedanken.

Dr. Maren Pröll und Dr. Christiane Neuhoff, danke für eure Hilfe und Unterstützung, die schönen Abende bei gutem Essen und die vielen Tagungen, die ich mit euch erleben durfte. Die ISAG in Salt Lake City und unser Roadtrip durch Kalifornien war ein großer Spaß mit euch!

Ich möchte mich bei allen Kollegen im Haus bedanken, die dazu beigetragen haben, dass ich mich (fast) jeden Tag gefreut habe, ins ITW zu fahren.

Ich möchte mich bei all meinen Freunden bedanken, die mich stets in meinem Vorhaben zu promovieren unterstützt haben und dafür gesorgt haben, dass man am Ende langer Arbeitstage den nötigen Ausgleich hatte. Danke Nina, Tine und Julia für eure wundervolle Freundschaft. Ihr seid die besten!

Ein großes Dankeschön möchte ich auch meiner Familie aussprechen. Papa und Mama, ihr habt mich zu dem Menschen gemacht, der ich heute bin. Danke für eure Unterstützung und all die Kleinigkeiten im Leben, die ihr für mich gemacht habt und immer noch macht. Liebe Oma, ich glaube kaum jemand hat so eine tolle Oma wie ich. Danke für dein stets offenes Ohr, deinen Zuspruch und deine Ratschläge - egal um welches Problem es sich auch handelt.

Der größte Dank gilt jedoch meinem Freund Tobi. Du hast mich jeden Abend aufgebaut, wenn die Statistik mich wahnsinnig gemacht hat. Danke für deine Liebe, dein Verständnis und deine Gelassenheit, mit der du mich immer wieder auf den Boden der Tatsachen geholt hast. Ich freue mich auf unsere gemeinsame Zukunft. Diese Arbeit ist dir gewidmet!

Publications and presentations

Welzenbach. J.; Neuhoff, C.; Heidt, H.; Cinar, M. U.; Looft, C.; Schellander, K.; Tholen, E. and Große-Brinkhaus, C. (2016) Integrative Analysis of Metabolomic, Proteomic and Genomic Data to Reveal Functional Pathways and Candidate Genes for Drip Loss in Pigs. *International Journal of Molecular Sciences*. 2016; 17: 1426. doi:10.3390/ijms17091426.

Welzenbach, J.; Neuhoff, C.; Looft, C.; Schellander, K.; Tholen, E. and Große-Brinkhaus, C. (2016) Different Statistical Approaches to Investigate Porcine Muscle Metabolome Profiles to Highlight New Biomarkers for Pork Quality Assessment. *PLoS ONE*. 2016; 11: e0149758. doi: 10.1371/journal.pone.0149758.

Welzenbach. J.; Neuhoff, C.; Heidt, H.; Cinar, M. U.; Looft, C.; Schellander, K.; Tholen, E. and Große-Brinkhaus, C. (2016) Integrative Analysis of Metabolome, Proteome and Genome to Reveal Functional Pathways and Candidate Genes for Drip Loss in Pigs. Oral presentation in: 35th International Society for Animal Genetics Conference (ISAG), 7.23.2016 – 7.27.2016, Salt Lake City, Utah, United States.

Welzenbach, J.; Große-Brinkhaus, C.; Neuhoff, C.; Pröll, M.; Schellander, K.; Looft, C. and Tholen, E. (2016) Systembiologische Analyse der Fleischqualität und Schlachtkörperzusammensetzung beim Schwein. In: Vortragstagung der Deutschen Gesellschaft für Züchtungskunde e.V. (DGfZ) und der Gesellschaft für Tierzuchtwissenschaften e.V. (GfT), p.B08, Hannover.

Welzenbach, J.; Große-Brinkhaus, C.; Neuhoff, C.; Schellander, K.; Looft, C. and Tholen, E. (2015) Data integration and network construction with muscle metabolome and meat quality data in pigs. Oral presentation in: 66th Annual Meeting of the European Association for Animal Production (EAAP), 8.31.2015 – 9.4.2015, Warsaw, Poland.

Welzenbach, J.; Große-Brinkhaus, C.; Neuhoff, C.; Schellander, K.; Looft, C. and Tholen, E. (2015) Data integration and network construction with muscle metabolome and meat quality data in pigs to highlight new biomarkers for pork quality assessment. Poster presentation in: 6th International Symposium on Animal Functional Genomics (ISAFG), Piacenza, Italy.

Welzenbach, J.; Große-Brinkhaus, C.; Neuhoff, C.; Schellander, K.; Looft, C. and Tholen, E. (2015) Funktionale Analyse des Tropfsaftverlustes beim Schwein unter Nutzung von metabolomischen, proteomischen und genomischen Informationen. In: Vortragstagung der Deutschen Gesellschaft für Züchtungskunde e.V. (DGfZ) und der Gesellschaft für Tierzuchtwissenschaften e.V. (GfT), Berlin.

Welzenbach, J.; Große-Brinkhaus, C.; Neuhoff, C.; Schellander, K.; Looft, C. and Tholen, E. (2014) Metaboliten-Profile als Biomarker für Tropfsaftverluste beim Schwein. In: Vortragstagung der Deutschen Gesellschaft für Züchtungskunde e.V. (DGfZ) und der Gesellschaft für Tierzuchtwissenschaften e.V. (GfT), p.D09, Dummerstorf.