Data integration and network construction with muscle metabolome and meat quality data in pigs

Julia Welzenbach, Christine Große-Brinkhaus, Christiane Neuhoff, Karl Schellander, Christian Looft, Ernst Tholen

Institute of Animal Science
Animal Breeding and Husbandry / Genetics group
University of Bonn, Germany
Introduction
- meaning of drip loss in pork

1) technological meat characteristics
- reduced acceptance of processing industries

2) sensory meat characteristics
- reduced consumer acceptance

1 Fischer et al. 2007
Introduction
- meaning of drip loss in pork

- e.g. stress conditions before slaughtering
- high drip loss

- reduction in meat quality\(^1\) by impairment of...
- 1) technological meat characteristics
  - reduced acceptance of processing industries
- 2) sensory meat characteristics
  - reduced consumer acceptance

How to select for low drip loss despite low heritability!?^\(^1\)

\(^1\)Fischer et al. 2007
Introduction
- the way from genome to phenotype

1Bino et al., 2004
Introduction
- the way from genome to phenotype

- Genotype
- Transcriptome
- Proteome
- Metabolome

$\rightarrow$ Metabolites as more reliable phenotype$^1$

$^1$Bino et al., 2004
Introduction
- the way from genome to phenotype

- Metabolites as more reliable phenotype
- Fill in the gap between phenotype and genome

Bino et al., 2004
Objective of the study

Investigation of the relationship between metabolite profiles and drip loss

- get new insights in the complex biochemical processes of drip loss
- test different statistical approaches to reveal the most important metabolites
- detect metabolites as potential biomarkers
Animals and drip loss phenotyping

- 97 F₂ pigs of Duroc × Pietrain resource population
- Meat samples of *M. longissimus dorsi*
- Record of drip loss 24h pm
  - Bag-Methode of *Honikel* (1986)
  - drip loss range: 0,4 – 5,3%
Quantification and annotation of metabolites

- by gas – and liquid chromatography with mass spectrometry (GC-MS, LC-MS)

- „untargeted“ metabolomics profiling
  - Detection of the whole metabolome
  - 1993 metabolites detected

- functional annotation
  - Human Metabolome Database, Lipid Maps, METLIN
  - 400 of 1993 metabolites annotated
Pre-correction of phenotypes and metabolite profiles

- generalized lineare model (GLM) with slaughter weight and season

Statistical approaches to analyse the drip – metabolite associations

1. Correlation analysis
2. Principal component analysis (PCA)
3. Weighted network analysis (WNA)
   - using \textit{WGCNA}^{1}
4. Random forest regression\textsuperscript{2} (RFR)
   - using \textit{party}^{3}

\textsuperscript{1}Langfelder & Horvath 2008, \textsuperscript{2}Breimann 2001, \textsuperscript{3}Strobl et al. 2007
Statistical analysis I

- Pre-correctation of phenotypes and metabolite profiles
  - generalized linear model (GLM) with slaughter weight and season

- Statistical approaches to analyse the drip – metabolite associations
  1. Correlation analysis
  2. Principal component analysis (PCA)
  3. Weighted network analysis (WNA)
    - using R WGCNA\(^1\)
  4. Random forest regression\(^2\) (RFR)
    - using R party\(^3\)

\(^1\)Langfelder & Horvath 2008, \(^2\)Breimann 2001, \(^3\)Strobl et al. 2007
2. **PCA**
   - Condenses the metabolite profiles into representative, uncorrelated principle components (PCs)
   - Metabolites are quantified by their corresponding loadings

3. **Weighted network analysis (WNA)**
   - Generates biological interpretable modules based on a hierarchical clustering dendrogram
   - Metabolites are characterised by
     - **module membership (MM)**
       - Connectivity of the metabolites within a module
     - **metabolite significance (MS)**
       - Based on the trait ↔ metabolite correlation
4. Random Forest Regression (RFR)

- supervised learning tool using tree-based methods
- key characteristics of RFR:

- RFR is able to handle datasets with complex interaction structures and highly correlated variables

- RFR calculates variable importance (VI) values, that are based on prediction accuracy\(^1\)

\(^1\)Strobl et al. 2007
Results – Identification of biomarkers

1. Correlation analysis
   - 71 (5) metabolites positive (negative) correlated ($p \leq 0.05$)
   - Range: 0.24 to 0.28 respectively -0.24 to -0.23

2. PCA
   - First 3 PCs specify 46.9 % of metabolite expression variance
   - Loadings are very weak (range: -0.1 to 0.1) → not significant\(^1\)
     - PCA not used for biomarker identification
     - But: PCA was used to reduce the data set for RFR

\(^1\)DiLeo et al., 2011
Results
– Identification of biomarkers

3. Weighted Network analysis (WNA)
   - Clusters the metabolites into 10 modules
     - two modules significantly associated with drip loss

<table>
<thead>
<tr>
<th>trait</th>
<th>module</th>
<th>cor.</th>
<th>p-value</th>
<th>number metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>drip loss</td>
<td>'purple'</td>
<td>+ 0.21</td>
<td>p ≤ 0.04</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>'green'</td>
<td>+ 0.21</td>
<td>p ≤ 0.04</td>
<td>49</td>
</tr>
</tbody>
</table>

4. Random Forest Regression (RFR)
   - 293 metabolites with significant (p ≤ 0.05) variable importance (VI) for drip loss
Results – network construction

drip

color

pH1

pH24
**Results – network construction**

- **Glycero-3-phosphocholine**
- **Triacylglycerol**
- **2.3-Naphthalic acid**
- **3-Methyl-2-oxovaleric acid**
- **Glycerophospholipid**
- **Triacylglycerol**
- **Glycerophospholipid**
- **pH1**
- **Color**
- **pH24**
Results – network construction

- Glycero-3-phosphocholine
- 2.3-Naphthalic acid
- Allopurinol-1-ribonucleoside
- Triacylglycerol
- 3-Methyl-2-oxovaleric acid
- Cytidine
- 2-Naphthalic acid
- Glycerophospholipid
- His Ala Trp Trp
- Lys Ser Ile
- Lactic acid
- Phosphocreatine
- Glycero-3-phosphoserine
- pH1
- pH24
- drip
- color
Results – network construction

- Glycero-3-phosphocholine
- Triacylglycerol
- 2.3-Naphthalic acid
- Allopurinol-1-ribonucleoside
- Cytidine
- His Ala Trp Trp
- Lactic acid
- α-Hydroxybutyrate
- Stearoylcarnitine
- Heptadecanoyl carnitine
- Methyglyoxal
- 2.1-Stearoylcarnitine
- Glycero-3-phosphoserine
- drip
- Glycerophospholipid
- Lys Ser Ile
- Phosphocreatine
- Glucose
- Naphthalic acid
- 3-Methyl-2-oxovaleric acid
- Triacylglycerol
- Methyglyoxal
- 2.1-Stearoylcarnitine
- pH1
- pH24
Results – network construction

- Glycero-3-phosphocholine
- Triacylglycerol
- 2.3-Naphthalic acid
- Allopurinol-1-ribonucleoside
- Cytidine
- His Ala Trp Trp
- Lactic acid
- α-Hydroxybutyrate
- Stearoylcarnitine
- Heptadecanoyl carnitine
- Phosphoglycolic acid
- Inosine-5-monophosphate
- Methyglyoxal
- 2.1-Stearoylcarnitine
- Octulose-1.8-bisphosphate
- Nicotinamide adenine dinucleotide
Results – network construction

Dotted lines: connectivity between metabolites

- Glycero-3-phosphocholine
- Triacylglycerol
- 2.3-Naphthalic acid
- Allopurinol-1-ribonucleoside
- Cytidine
- His Ala Trp Trp
- Lactic acid
- Lys Ser Ile
- Naphthalic acid
- Phosphocreatine
- α-Hydroxybutyrate
- Stearoylcarnitine
- Heptadecanoyl carnitine
- Phosphoglyceric acid
- Inosine-5-monophosphate
- Octulose-1.8-bisphosphate
- Methyglyoxal
- 2.1-Stearoylcarnitine
- Glycero-3-phosphoserine
drip
- Triacylglycerol
- 3-Methyl-2-oxovaleric acid
- Glycerophospholipid
- Naphthalic acid
- Glycero-3-phosphocholine
- Glycerophospholipid
- Triacylglycerol
- Lys Ser Ile
- Cytidine
- His Ala Trp Trp
- Naphthalic acid
- Phosphocreatine
- α-Hydroxybutyrate
- Stearoylcarnitine
- Heptadecanoyl carnitine
- Phosphoglyceric acid
- Inosine-5-monophosphate
- Octulose-1.8-bisphosphate
- Methyglyoxal
- 2.1-Stearoylcarnitine
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results – network construction

Dotted lines: connectivity between metabolites

Arrows: directed connections metabolite → trait
Results - prediction accuracy of selected metabolites

- Overlap in “Top 30” significant metabolites for drip loss
- 20 metabolites identified by more than one method
- Multiple R²: 32.73 %

Correlation analysis

<table>
<thead>
<tr>
<th></th>
<th>WNA</th>
<th>11</th>
<th>16</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>RFR</td>
<td>26</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Stepwise Regression of the 20 metabolites
- Resulted in 5 important metabolites
- Multiple R²: 26.61 %

<table>
<thead>
<tr>
<th>Biomarker for drip loss</th>
<th>Regression coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycerophosphocholine</td>
<td>45.82*</td>
</tr>
<tr>
<td>C24:1 Sphingomyelin</td>
<td>-50.53*</td>
</tr>
<tr>
<td>Ubiquinon (Prenol Lipid)</td>
<td>532.20*</td>
</tr>
<tr>
<td>C26:1 Sphingomyelin</td>
<td>1870.00**</td>
</tr>
<tr>
<td>Not annotated</td>
<td>398.50***</td>
</tr>
</tbody>
</table>
Conclusion and perspective

- Correlation analysis, WNA and RFR are suitable in identification of predictors
- Glycero phosphopho- und Sphingolipids are the most promising biomarkers
- Selected set of metabolites has moderate prediction accuracy and also an effect on other meat quality traits
- Requirements of the development of reliable metabolite biomarkers
  - enhanced abilities of metabolite quantification and annotation
- Future perspective: usage of combined omics-profiles as more exact phenotype and in GWAS of identify 'real' powerful SNPs
Thank you for your attention
Random Forest Regression (RFR)

- supervised learning tool using tree-based methods with integrated permutation tests
- RFR procedure:

1. Sample Data using bootstrap draws
2. Train Data: Data set to grow the single trees.
3. Feature selection: Random selection of mtry predictors
4. Grow tree: Split data using the best predictors
5. Estimate OOB error by applying the tree to the OOB data
6. Repeat until specified number of trees is obtained
7. Variable importance measurement

- 97 samples (drip)
- 1084 metabolites
### Results - important biomarkers

- Ranking of metabolites in „top 30“ of correlation analysis, WNA and RFR

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Cor</th>
<th>MS</th>
<th>VI</th>
<th>pH1</th>
<th>Cor</th>
<th>MS</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>drip loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3-Naphthalic acid</td>
<td>23.</td>
<td>x</td>
<td>10.</td>
<td>His Ala Trp Trp</td>
<td>5.</td>
<td>4.</td>
<td>2.</td>
</tr>
<tr>
<td>Glycero-3-phosphoserine</td>
<td>x</td>
<td>28.</td>
<td>23.</td>
<td>Allopurinol-1-ribonucleoside</td>
<td>x</td>
<td>9.</td>
<td>25.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Cor</th>
<th>MS</th>
<th>VI</th>
<th>color</th>
<th>Cor</th>
<th>MS</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-Hydroxybutyrate</td>
<td>1.</td>
<td>1.</td>
<td>x</td>
<td>Octulose-1.8-bisphosphate</td>
<td>7.</td>
<td>1.</td>
<td>x</td>
</tr>
<tr>
<td>Heptadecanoyl carnitine</td>
<td>2.</td>
<td>2.</td>
<td>x</td>
<td>Fructose-6-phosphate</td>
<td>27.</td>
<td>9.</td>
<td>x</td>
</tr>
<tr>
<td>Stearoylcarnitine</td>
<td>3.</td>
<td>4.</td>
<td>x</td>
<td>Glucose-6-phosphate</td>
<td>23.</td>
<td>7.</td>
<td>x</td>
</tr>
<tr>
<td>Gle-cholesterol</td>
<td>x</td>
<td>x</td>
<td>2.</td>
<td>Inosine-5-monophosphate</td>
<td>28.</td>
<td>10.</td>
<td>x</td>
</tr>
<tr>
<td>Methyglyoxal</td>
<td>x</td>
<td>x</td>
<td>9.</td>
<td>Phosphoglycolic acid</td>
<td>11.</td>
<td>12.</td>
<td>x</td>
</tr>
<tr>
<td>Glucose</td>
<td>x</td>
<td>x</td>
<td>11.</td>
<td>Nicotinamide adenine dinucleotide</td>
<td>4.</td>
<td>x</td>
<td>2.</td>
</tr>
</tbody>
</table>
Results – potential biomarkers

consistent results of statistical approaches:
Glyceroosphospholipids (GPLs), Sphingolipids (SLs)

unfavourable slaughter conditions

post mortem:

pH decrease\(^1\), cell swelling, cell shrinkage\(^2\)

Degradation of cell membranes\(^3\), homeostatic balance ↓ membrane integrity ↓ spill of membrane lipids and proteins → drip loss

\(^1\) Ortenblad et al., 2003; \(^2\) Bertram et al., 2004; \(^3\) Lambert et al., 2001