Gene and Pathway Analysis of Metabolic Traits in Dairy Cows

Ngoc-Thuy Ha
Animal Breeding and Genetics Group
Department of Animal Sciences
Georg-August-University Goettingen, Germany
Motivation – Background

- Selective breeding of high-yielding dairy cows
  - up to 45 kg milk per day
- High energy demand can not be fully covered by food intake
  - negative energy balance during their early lactation
- Mobilization of body fat, protein and mineral stores
  - adaptation of the hepatic metabolism

Successful metabolic adaptation without any disorder occurrences

Development of production-related disorders, such as ketosis and fatty liver
Motivation – Metabolic Adaptation

Why does the success of adaptation differ substantially between cows – even under the same conditions and similar production levels?

Ingvartsen et al. (2003)
Drackley et al. (2005)
Graber et al. (2010)

This metabolic 'robustness' has a genetic basis.

Goal: Study the genetic basis of the metabolic adaptability of dairy cows during early lactation
**Data**

178 dairy cows (field study, Graber et al., 2010)

**Phenotypes**

- NEFA (non-esterified fatty acid)
- BHBA (beta-hydroxybutyrate)
- glucose

**Genotypes**

- 601,455 SNPs
  - Illumina HD Bovine BeadChip

- **Ensembl**: 22,025 genes
  - (231,712 intragenic SNPs)

- **KEGG**: 81 metabolic pathways
  - (6,376 genes)

<table>
<thead>
<tr>
<th>Phenotypes</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 weeks ante-partum (-3W)</td>
<td>601,455 SNPs</td>
</tr>
<tr>
<td>4 weeks post-partum (+4W)</td>
<td>Illumina HD Bovine BeadChip</td>
</tr>
<tr>
<td>13 weeks post-partum (+13W)</td>
<td><strong>Ensembl</strong>: 22,025 genes</td>
</tr>
<tr>
<td></td>
<td>(231,712 intragenic SNPs)</td>
</tr>
<tr>
<td></td>
<td><strong>KEGG</strong>: 81 metabolic pathways</td>
</tr>
<tr>
<td></td>
<td>(6,376 genes)</td>
</tr>
</tbody>
</table>
Data

van Dorland et al. (2009)  
Graber et al. (2010)  
Gross et al. (2011)

key factors characterizing the metabolic adaptation of dairy cows
GWAS

Genome-Wide Association Study (GWAS): Is the phenotype under consideration influenced by any genetic factors?
**Methods** – **SMA**

**Single Marker Analysis**

- log p-values

SNPs

Significance Level

Gene 1

Gene 2

Genome
Methods – **SMA**

**Disadvantages:**
- high dimensional data (up to millions of SNPs)
- vast multiple testing problem
- low power
- correlation of SNPs (LD)
- limitation in biological interpretation
Methods – Gene-based Test

Gene-based Multi Marker Analysis

- log p-values
Advantages of the gene-based analysis:

- less multiple testing
- able to account for the correlation (LD) of the SNPs
- able to detect genes with many small or medium-sized genetic effects
**Gene-based Score Test (GBST)**

Regression model for a gene with g SNPs:

\[ y = \mu + \beta_1 x_1 + \cdots + \beta_g x_g + \epsilon \]

- **log-likelihood function**
  \[ \ell(\beta_1, \beta_2, \ldots, \beta_g \mid y) \]

- **score statistics of the SNPs** \( j=1,2,\ldots,g \)
  \[ U_j = \frac{\partial \ell(\beta_1, \beta_2, \ldots, \beta_g \mid y)}{\partial \beta_j} \]

- **estimated variance of the score statistics**
  \[ s_j = \text{Var}(U_j) \]

- **test statistic according to Pan (2007):**
  \[ T_S = \sum_{i=1}^{g} \frac{U_i^2}{s_i} \]

- **Zhang (2005):**
  \[ T_S \sim a \chi^2_d + b \text{ for certain numbers } a, b \text{ and } d \text{ if} \]

the null hypothesis \( H_0 : \beta_1 = \beta_2 = \ldots = \beta_g = 0 \text{ is true.} \)
GWAS

**Genome-Wide Association Study (GWAS):** Is the phenotype under consideration influenced by any genetic factors?
Methods – Pathway Analysis

Gene-Set Enrichment Analysis (GSEA, Subramanian et al. 2005)

Inputs:
1. A list $L = \{g_1, g_2, \ldots, g_n\}$ of $n$ genes ordered according to a ranking metric $r(g_i) = r_i$ with $r_1 \geq r_2 \geq \cdots \geq r_n \geq 0$.
   
   $r$ = ‘importance’ of a gene to a phenotype,

   e.g. $r = -\log(p\text{-value})$

2. A gene set $S$ with $s$ genes, e.g. a pathway.
Procedure to test the association of the phenotype to the pathway S:

1. Start with a pathways score \( \text{Score}(S) = 0 \).

2. Go through the ordered list \( L \) from \( i = 1, 2, \ldots, n \) and

   - add \( \frac{r_i}{n_p} \) with \( n_p = \sum_{g_j \in S} r_j \) if the gene \( g_i \) is in the pathway \( S \)
   - or

   - subtract \( \frac{1}{n-s} \) otherwise.

3. The enrichment score \( E(S) \) of the pathway \( S \) is defined by the maximum value of the score \( \text{Score}(S) \).

4. Permute the phenotypes to obtain the null distribution of \( E(S) \).
GSEA – Subramanian et al. (2005)

Maximum = Enrichment Score
Results – Gene-based Analysis

Gene-based Analysis for metabolite NEFA

![Gene-based Analysis Diagram]
**Results – Gene/Pathway Analysis**

- Number of significant **genes** for the three metabolites (FDR < 5%): 
  - NEFA: 38 genes
  - BHBA: 29 genes
  - Glucose: 32 genes

- Number of significant **pathways** for the three metabolites (FDR < 5%): 
  - NEFA: 4 pathways
  - BHBA: 5 pathways
  - Glucose: 5 pathways
Results – Pathway Analysis

Are there pathways that have a joint impact on the three metabolites?

\[ r = -\log(p_{NEFA} \times p_{BHBA} \times p_{glucose}) \]
**Results – Pathway Analysis**

Significant pathways having a joint impact on the three metabolites

- **Steroid Hormone Biosynthesis**
  - 46 genes
  - *p*-value < $4 \times 10^{-3}$
  - (McCabe et al., 2012)

- **Ether Lipid Metabolism**
  - 8 genes
  - *p*-value < $5 \times 10^{-4}$
  - (Contreras et al., 2011)

- **Glycerophospholipid Metabolism**
  - 50 genes
  - *p*-value < $1 \times 10^{-4}$
  - (Contreras et al., 2011)

More results in Ha et al. (2015)
Discussion

- Detected several biologically sensible significant genes and pathways associated with candidate metabolites transition period
  
  → evidence for genetic basis

- Many genes are only significant at certain points of times
  
  → time-dependency of the genetic basis

  → potential candidate genes that become active in early lactation

- Three pathways were that are involved in the metabolism of lipids and steroids and have a joint impact an all three phenotypes
  
  → complexity of the genetic basis of the metabolic adaptation
Outlook

Step 1: Identification of candidate genetic factors (SNPs, genes, pathways) for the metabolic ‘robustness’

Step 2: Validation of the results on a transcriptomic level using RNA sequencing data

Step 3: Using the results to develop an SNP-chip optimized for the breeding of more ‘robust’ dairy cows
Acknowledgement

Josef Johann Gross\textsuperscript{1}, Annette van Dorland\textsuperscript{1}, Jens Tetens\textsuperscript{3}, Georg Thaller\textsuperscript{3}, Martin Schlather\textsuperscript{4}, Rupert Bruckmaier\textsuperscript{1}, Henner Simianer\textsuperscript{2}

\textsuperscript{1} Veterinary Physiology, Vetsuisse Faculty University of Bern, Switzerland
\textsuperscript{2} Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Goettingen, Germany
\textsuperscript{3} Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, Germany
\textsuperscript{4} Chair of Mathematical Statistics, University of Mannheim, Germany

This study was supported by a grant of the Swiss Commission for Technology and Innovation CTI (project no. 13948.2 PFLS-LS) and Swissgenetics, Zollikofen, Switzerland.
Thank you for your attention.

Drackley, James K., Heather M. Dann, G. Neil Douglas, Nicole A. Janovick Guretzky, Noah B. Litherland, John P. Underwood, und Juan J. Loor. 2005. „Physiological and pathological adaptations in dairy cows that may increase susceptibility to periparturient diseases and disorders“. Growth 7 (7.1).


### Results – Pathway Analysis

#### Significant pathways and references supporting the associations:

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Time</th>
<th>Pathways</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEFA</td>
<td>T2</td>
<td>Histidine metabolism</td>
<td>Vanhatalo et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sulfur metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2/T1</td>
<td>Glycerolipid metabolism</td>
<td>Contreras &amp; Sordillo, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glycerophospholipid metabolism</td>
<td>Contreras &amp; Sordillo, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taurine and hypotaurine metabolism</td>
<td></td>
</tr>
<tr>
<td>BHBA</td>
<td>T2</td>
<td>Retinol metabolism</td>
<td>LeBlanc et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tyrosine metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inositol phosphate metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Steroid hormone biosynthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2/T1</td>
<td>Synthesis and degradation of ketone bodies</td>
<td>Kanehisa et al., 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tryptophan metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inositol phosphate metabolism</td>
<td></td>
</tr>
</tbody>
</table>
### Results – Pathway Analysis

<table>
<thead>
<tr>
<th>Glucose</th>
<th>T2</th>
<th>Steroid biosynthesis</th>
<th>Marks &amp; Banks, 1960</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Other glycan degradation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fatty acid elongation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2/T1</td>
<td>Ether lipid metabolism</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Starch and sucrose metabolism</td>
<td>Kanehisa et al., 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Steroid hormone biosynthesis</td>
<td>Marks &amp; Banks, 1960</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glycerophospholipid metabolism</td>
<td></td>
</tr>
</tbody>
</table>
Methods – Size Bias

Single Marker vs. Gene-based Analysis

SMA with Minimum P-Value, R-Squared = 0.204

Gene-based Score Test, R-Squared = 0.003