



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.



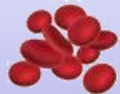
Goal: Study the genetic basis of the metabolic adaptability of dairy cows during early lactation

Data



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

Phenotypes



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

3 weeks ante-partum (-3W)

4 weeks post-partum (+4W)

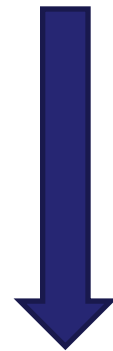
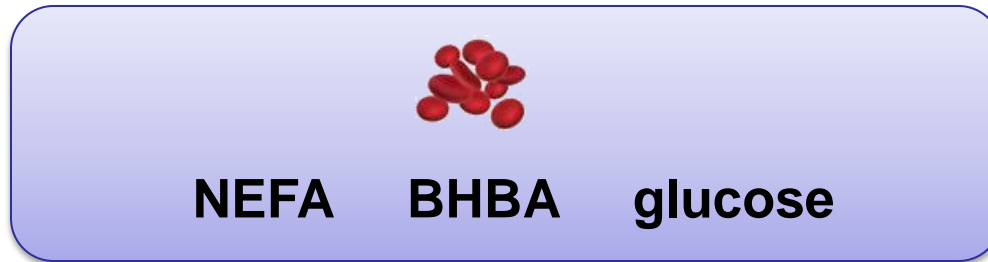
13 weeks post-partum (+13W)

Genotypes

601,455 SNPs
Illumina HD Bovine BeadChip

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

KEGG: 81 metabolic pathways
(6,376 genes)



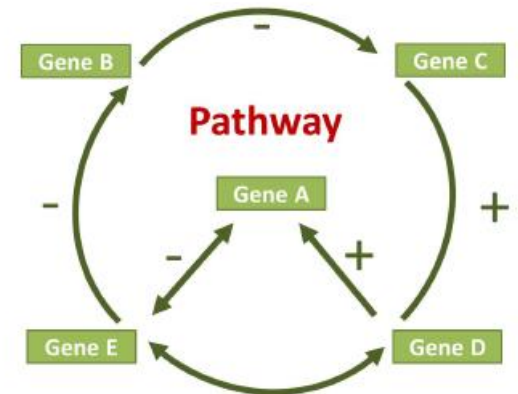
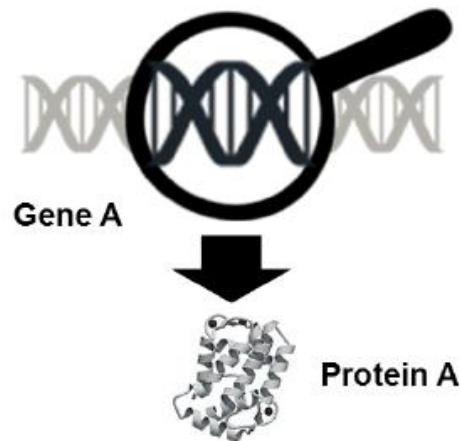
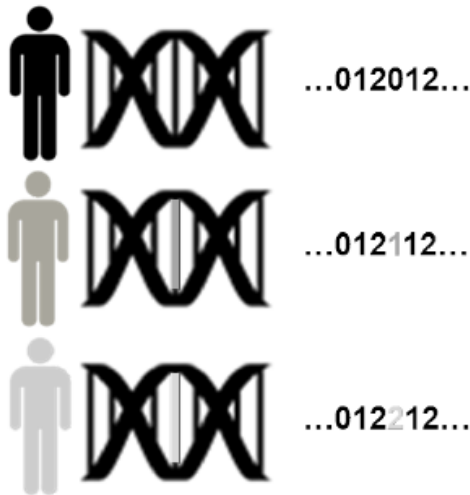
van Dorland et al. (2009)
Graber et al. (2010)
Gross et al. (2011)

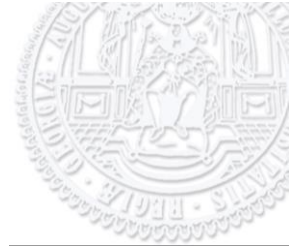
key factors characterizing the
metabolic adaptation of dairy cows

GWAS

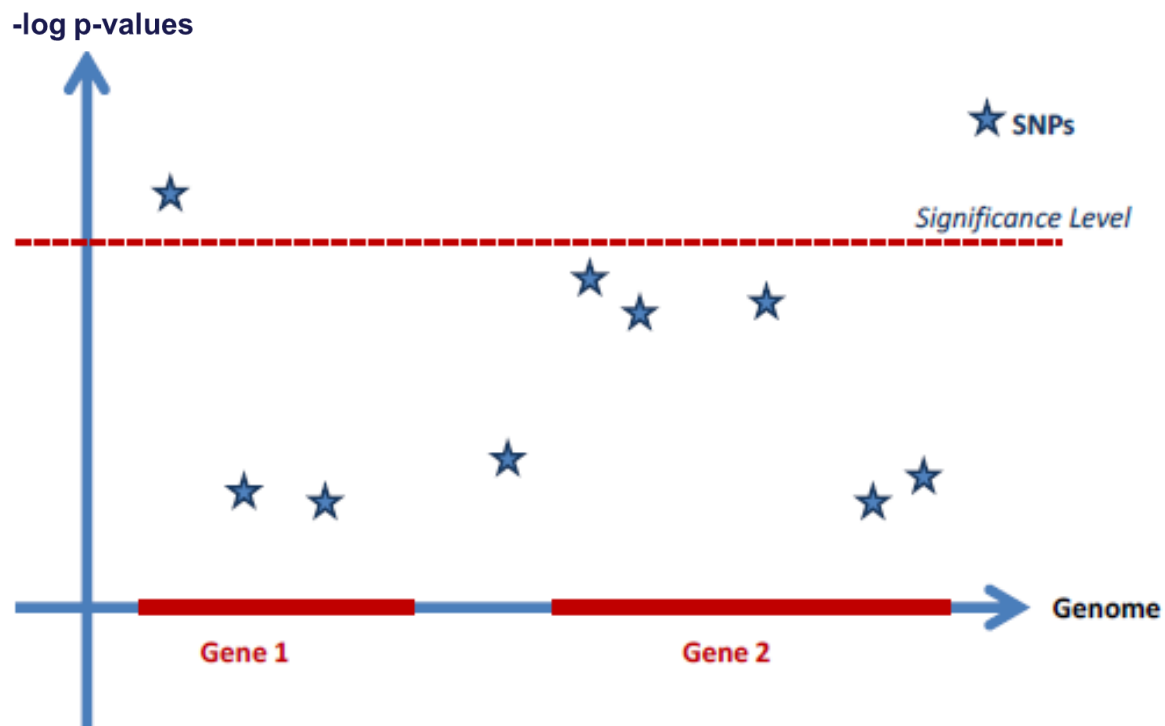


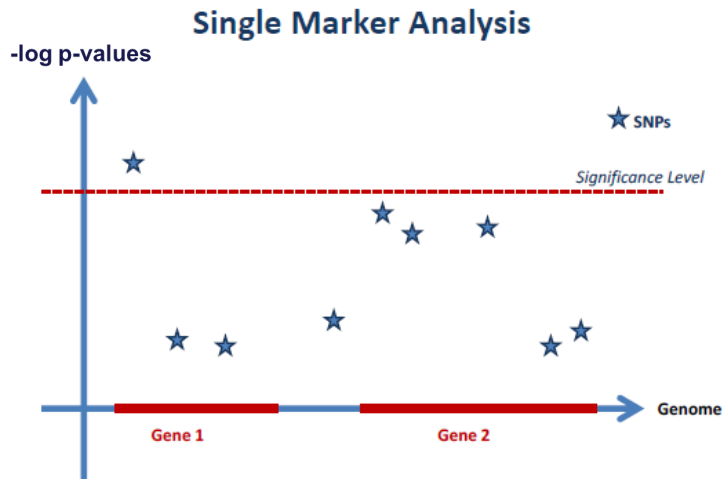
Genome-Wide Association Study (GWAS): Is the phenotype under consideration influenced by any genetic factors?





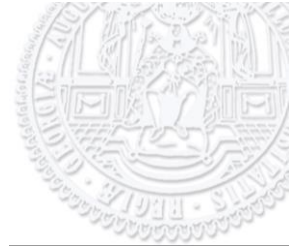
Single Marker Analysis



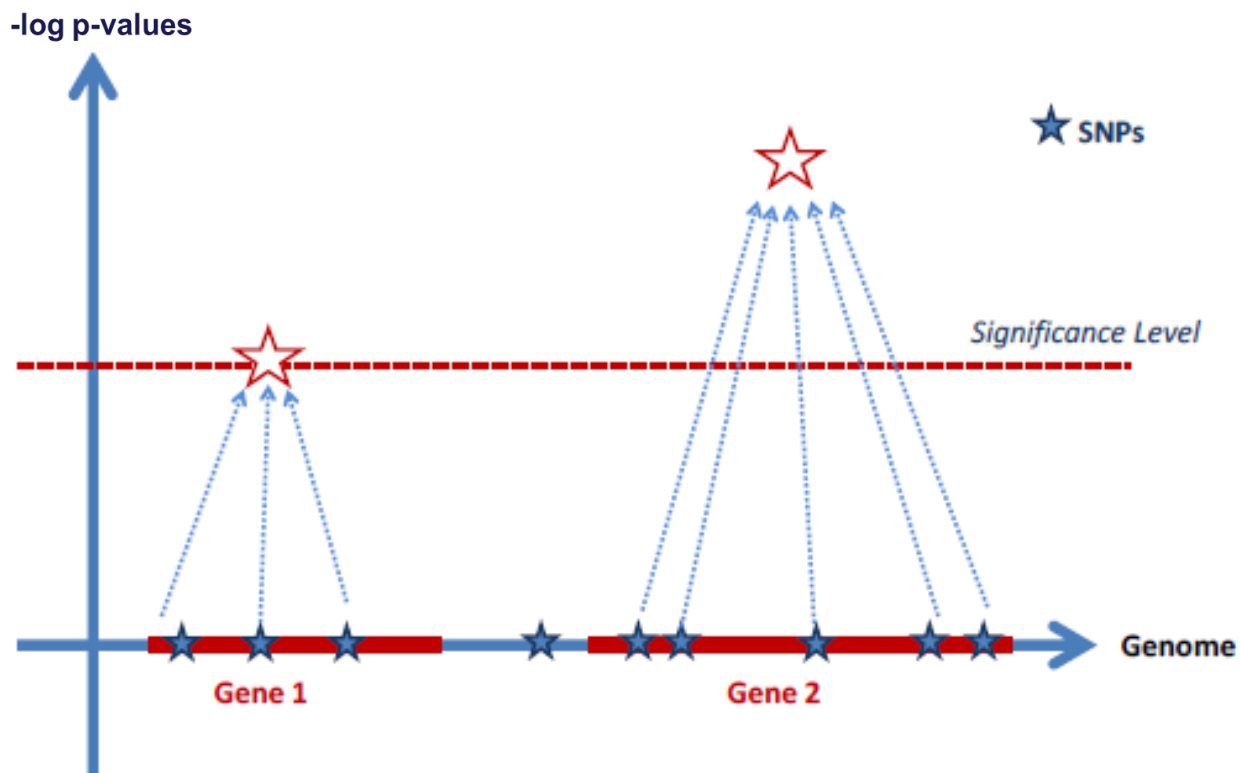


Disadvantages:

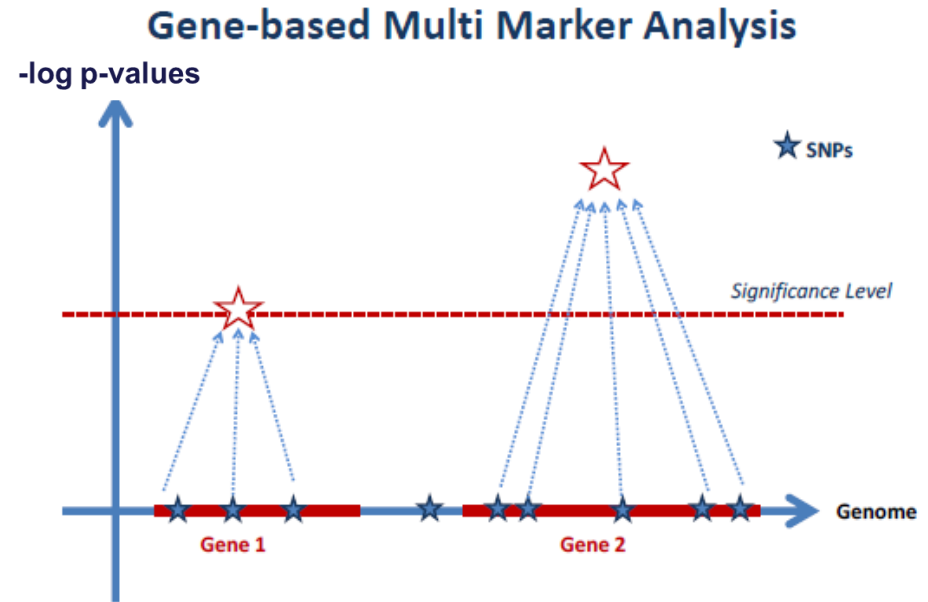
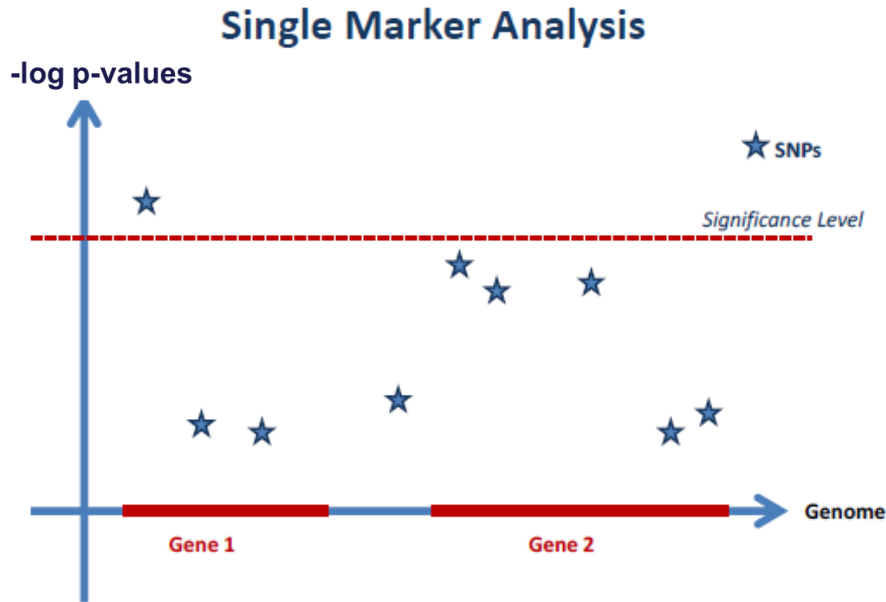
- high dimensional data (up to millions of SNPs)
- vast multiple testing problem
- low power
- correlation of SNPs (LD)
- limitation in biological interpretation



Gene-based Multi Marker Analysis



Methods – Gene-based Test



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:

$$y = \mu + \beta_1 X_1 + \dots + \beta_g X_g + \epsilon$$

$$\ell(\beta_1, \beta_2, \dots, \beta_g | \mathbf{y})$$

log-likelihood function

$$U_j = \frac{\partial \ell(\beta_1, \beta_2, \dots, \beta_g | \mathbf{y})}{\partial \beta_j}$$

score statistics of the SNPs $j=1,2,\dots,g$

$$s_j = \text{Var}(U_j)$$

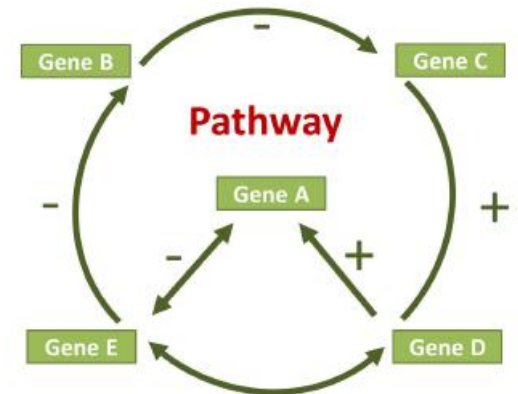
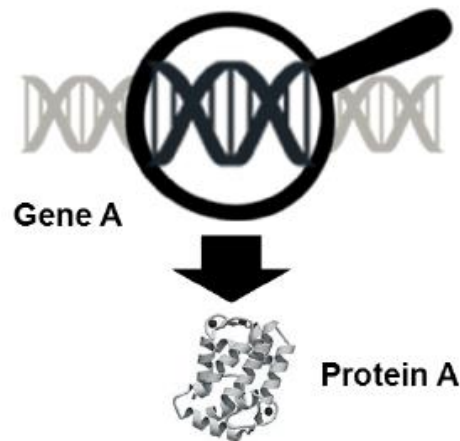
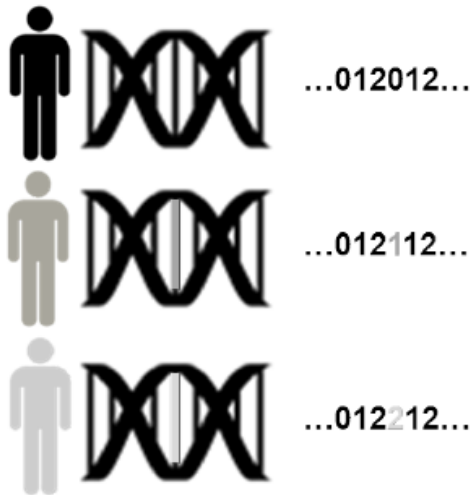
estimated variance of the score statistics

- test statistic according to Pan (2007): $T_S = \sum_{i=1}^g \frac{U_i^2}{s_i}$
- Zhang (2005): $T_S \sim a\chi_d^2 + b$ for certain numbers a, b and d if the null hypothesis $H_0 : \beta_1 = \beta_2 = \dots = \beta_g = 0$ is true.

GWAS



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

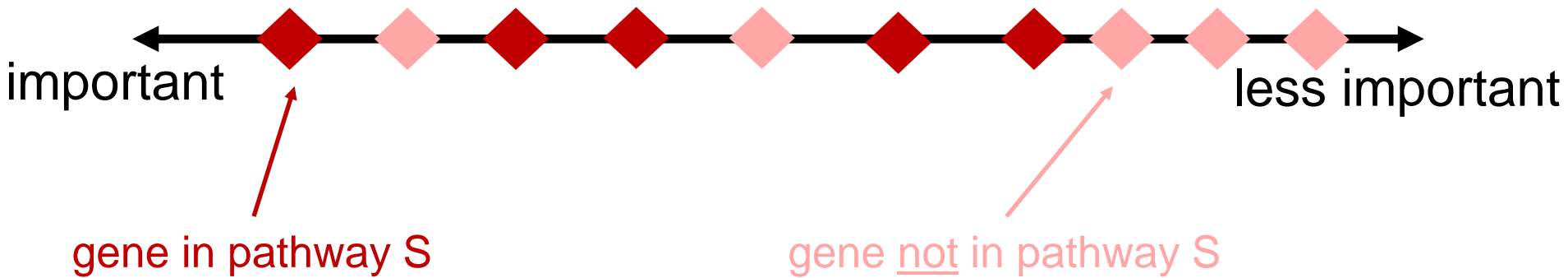


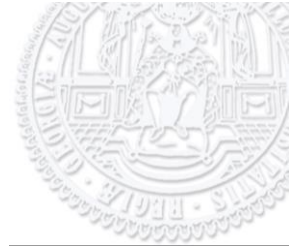


Gene-Set Enrichment Analysis (GSEA, Subramanian et al. 2005)

Inputs:

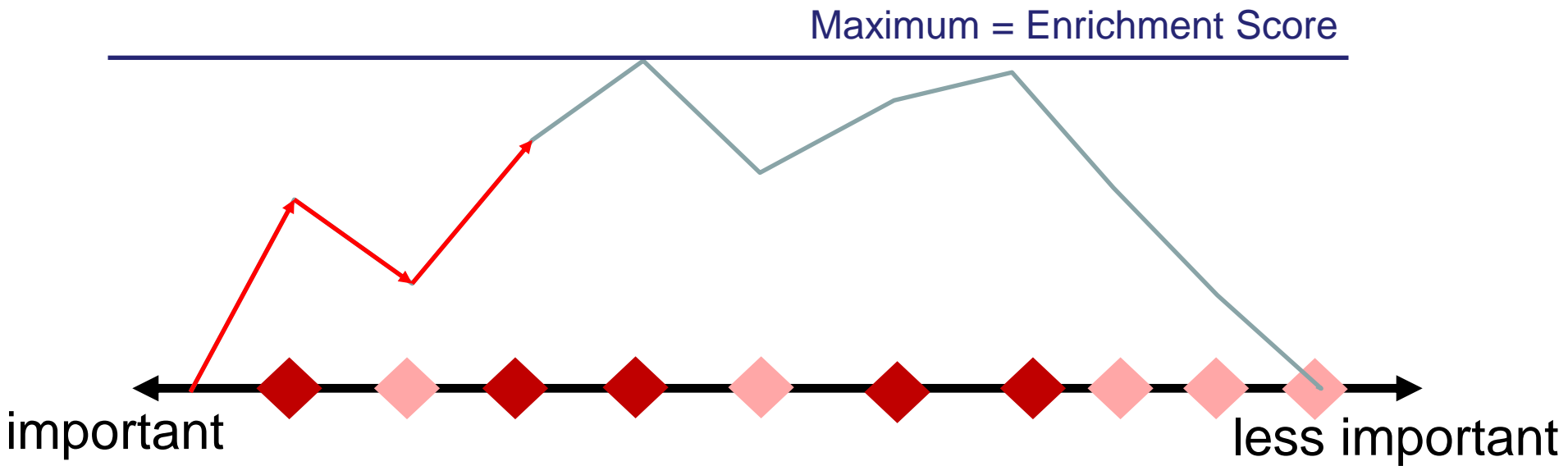
1. A list $L = \{g_1, g_2, \dots, g_n\}$ of n genes ordered according to a ranking metric $r(g_i) = r_i$ with $r_1 \geq r_2 \geq \dots \geq r_n \geq 0$.
 - ➔ r = 'importance' of a gene to a **phenotype**,
e.g. $r = -\log(p\text{-value})$
2. A gene set S with s genes, e.g. a pathway.





Procedure to test the association of the **phenotype** to the **pathway S**:

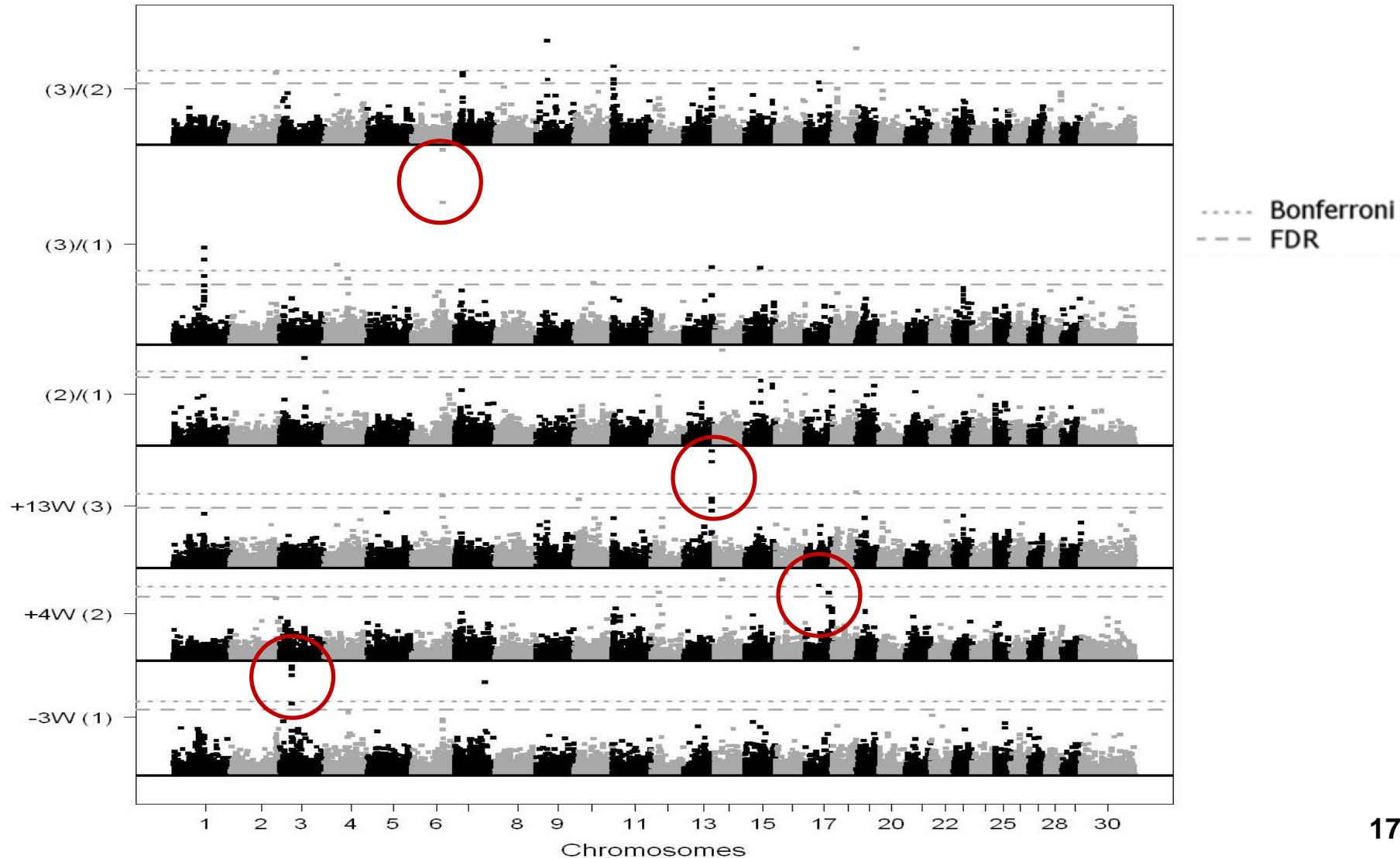
1. Start with a pathways score $Score(S) = 0$.
2. Go through the ordered list L from $i = 1, 2, \dots, n$ and
 - add $\frac{r_i}{n_p}$ with $n_p = \sum_{g_j \in S} r_j$ if the gene g_i is in the pathway S
 - or
 - subtract $\frac{1}{n-s}$ otherwise.
3. The enrichment score $E(S)$ of the pathway S is defined by the maximum value of the score $Score(S)$.
4. Permute the *phenotypes* to obtain the null distribution of $E(S)$.



Results – Gene-based Analysis



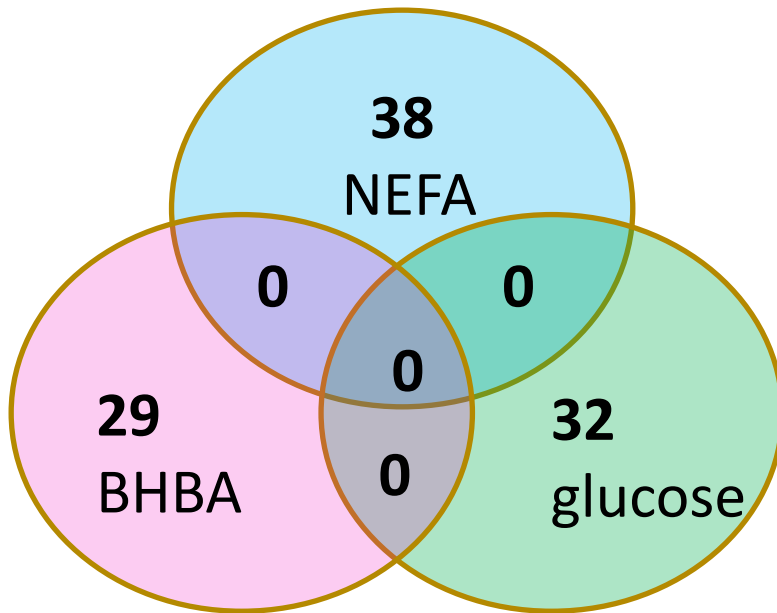
Gene-based Analysis for metabolite NEFA



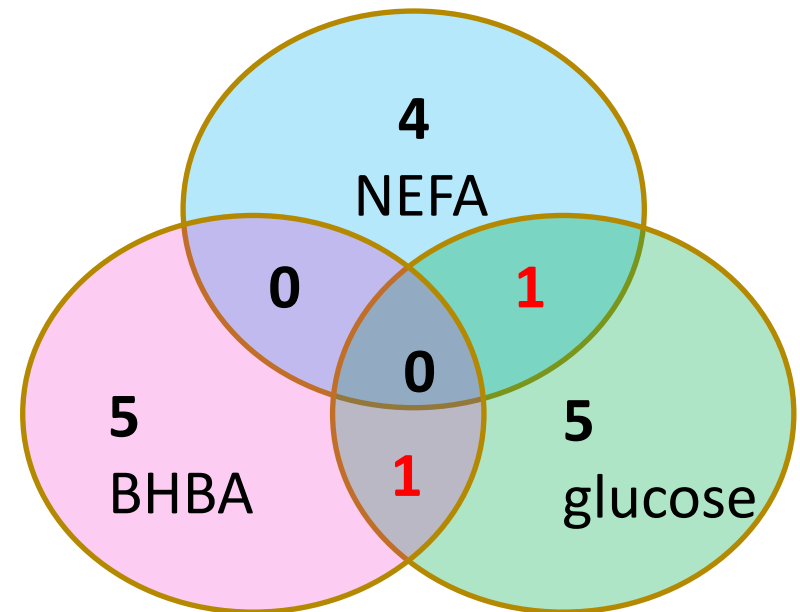
Results – Gene/Pathway Analysis



- Number of significant **genes** for the three metabolites (FDR < 5%):



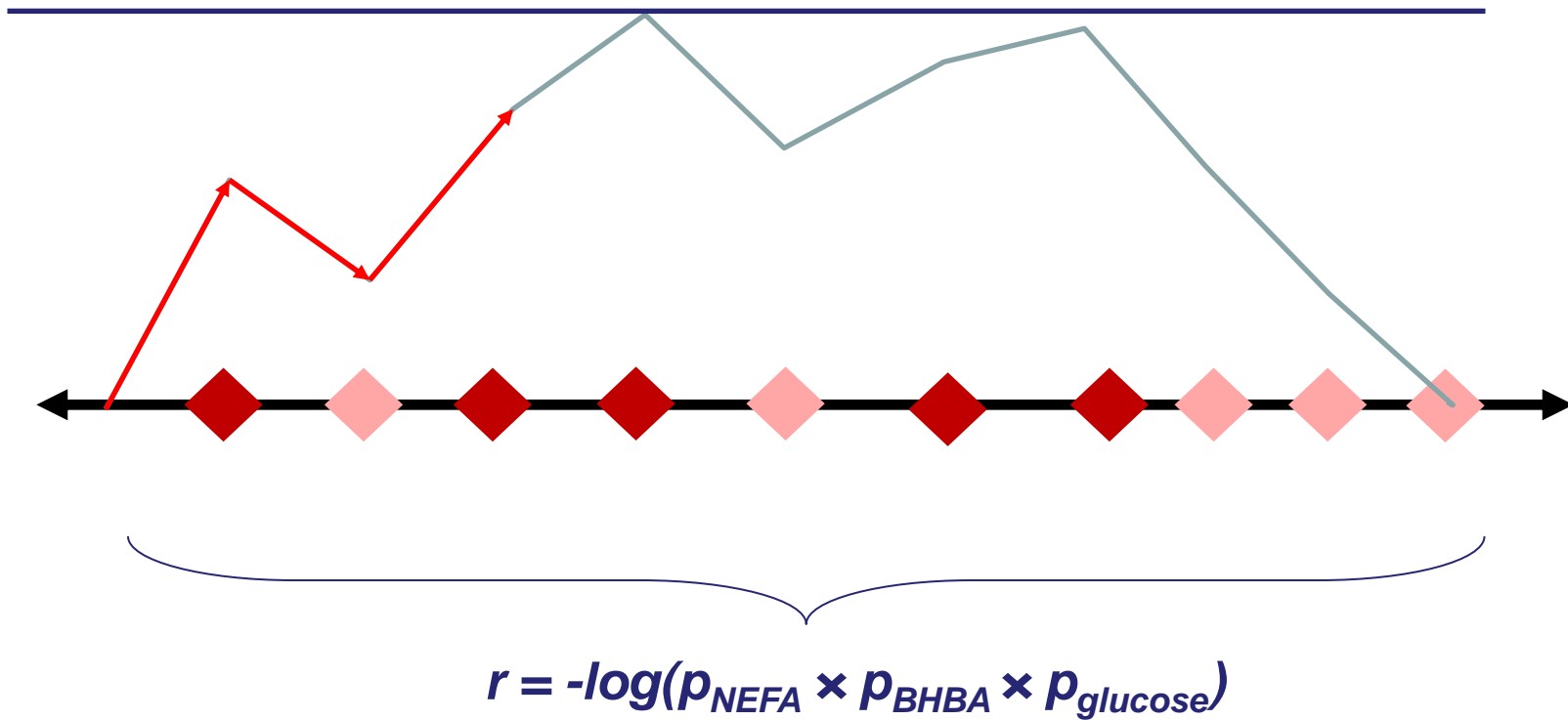
- Number of significant **pathways** for the three metabolites (FDR < 5%):



Results – Pathway Analysis



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

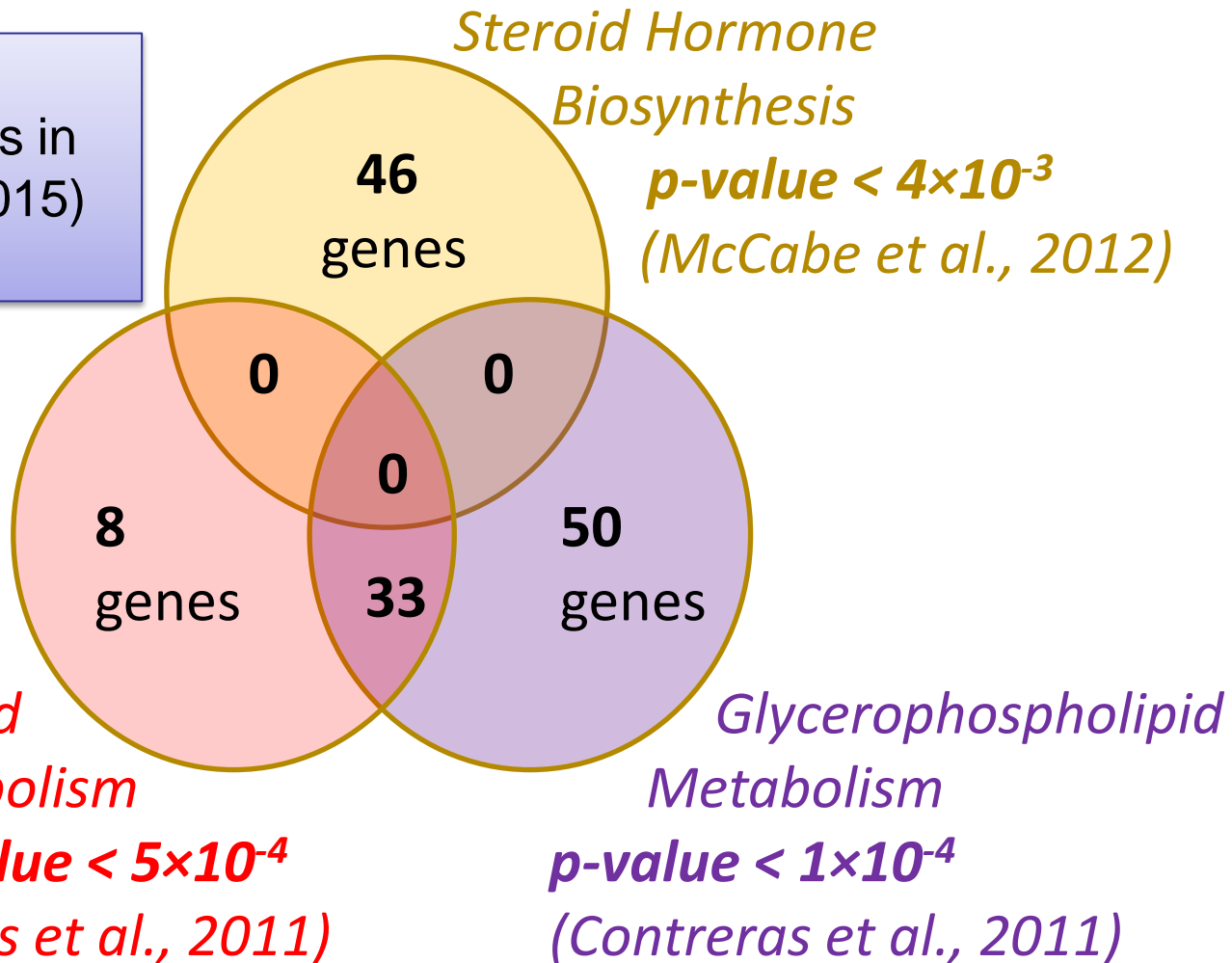


Results – Pathway Analysis



Significant pathways having a **joint** impact on the three metabolites

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 on 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'**

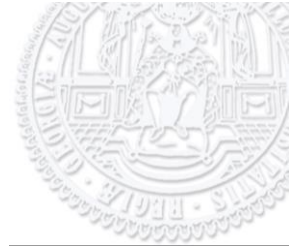


Step 2: Validation of the results on a **transcriptomic level** using RNA sequencing data



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

Acknowledgement



*Josef Johann Gross¹, Annette van Dorland¹, Jens Tetens³, Georg Thaller³,
Martin Schlather⁴, Rupert Bruckmaier¹, Henner Simianer²*

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

swissgenetics 



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Volkswirtschaftsdepartement EVD
Bundesamt für Berufsbildung und Technologie BBT
Förderagentur für Innovation KTI

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.

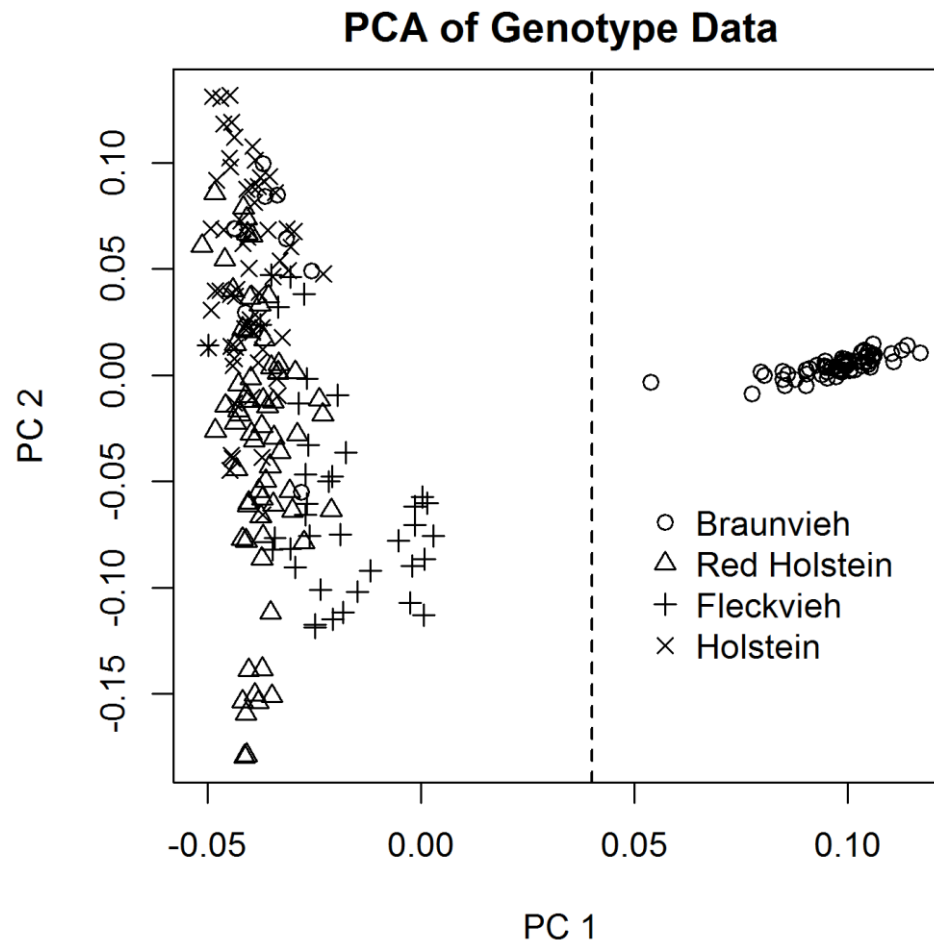


Thank you for your attention.

References



- Contreras, G Andres, und Lorraine M Sordillo. 2011. „Lipid Mobilization and Inflammatory Responses during the Transition Period of Dairy Cows“. *Comparative Immunology, Microbiology and Infectious Diseases* 34 (3): 281–89. doi:10.1016/j.cimid.2011.01.004.
- Drackley, James K., Heather M. Dann, G. Neil Douglas, Nicole A. Janovick Guretzky, Noah B. Litherland, John P. Underwood, und Juan J. Loo. 2005. „Physiological and pathological adaptations in dairy cows that may increase susceptibility to periparturient diseases and disorders“. *Growth* 7 (7.1).
- Graber, M, S Kohler, T Kaufmann, M G Doherr, R M Bruckmaier, und H A van Dorland. 2010. „A Field Study on Characteristics and Diversity of Gene Expression in the Liver of Dairy Cows during the Transition Period“. *Journal of Dairy Science* 93 (11): 5200–5215.
- Gross, J, H A van Dorland, R M Bruckmaier, und F J Schwarz. 2011. „Performance and Metabolic Profile of Dairy Cows during a Lactational and Deliberately Induced Negative Energy Balance with Subsequent Realimentation“. *Journal of Dairy Science* 94 (4): 1820–30.
- Ha, Ngoc-Thuy, Josef Johann Gross, Annette van Dorland, Jens Tetens, Georg Thaller, Martin Schlather, Rupert Bruckmaier, und Henner Simianer. 2015. „Gene-Based Mapping and Pathway Analysis of Metabolic Traits in Dairy Cows“. *PloS One* 10 (3): e0122325.
- Ingvartsen, K. L, R. J Dewhurst, und N. C Friggens. 2003. „On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper“. *Livestock Production Science* 83 (2–3): 277– 308.
- Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, u. a. 2005. „Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles“. *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50.



Results – Pathway Analysis



Significant pathways and references supporting the associations:

Phenotype	Time	Pathways	Literature
NEFA	T2	Histidine metabolism	Vanhatalo et al., 1999
		Sulfur metabolism	
	T2/T1	Glycerolipid metabolism	Contreras & Sordillo, 2011
		Glycerophospholipid metabolism	
		Taurine and hypotaurine metabolism	
BHBA	T2	Retinol metabolism	LeBlanc et al., 2004
		Tyrosine metabolism	
		Inositol phosphate metabolism	
		Steroid hormone biosynthesis	
		Synthesis and degradation of ketone bodies	
Tryptophan metabolism			
		Inositol phosphate metabolism	

Results – Pathway Analysis

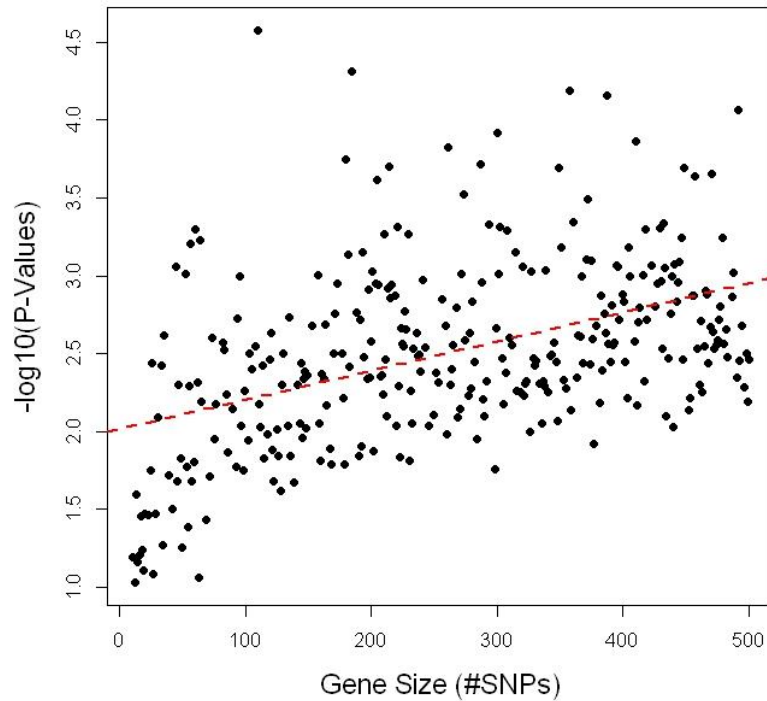


glucose	T2	Steroid biosynthesis	Marks & Banks, 1960
		Other glycan degradation	
		Fatty acid elongation	
		Ether lipid metabolism	
	T2/T1	Ether lipid metabolism	
		Starch and sucrose metabolism	Kanehisa et al., 2012
		Steroid hormone biosynthesis	Marks & Banks, 1960
		Glycerophospholipid metabolism	



Single Marker vs. Gene-based Analysis

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



Gene-based Score Test, R-Squared = 0.003

