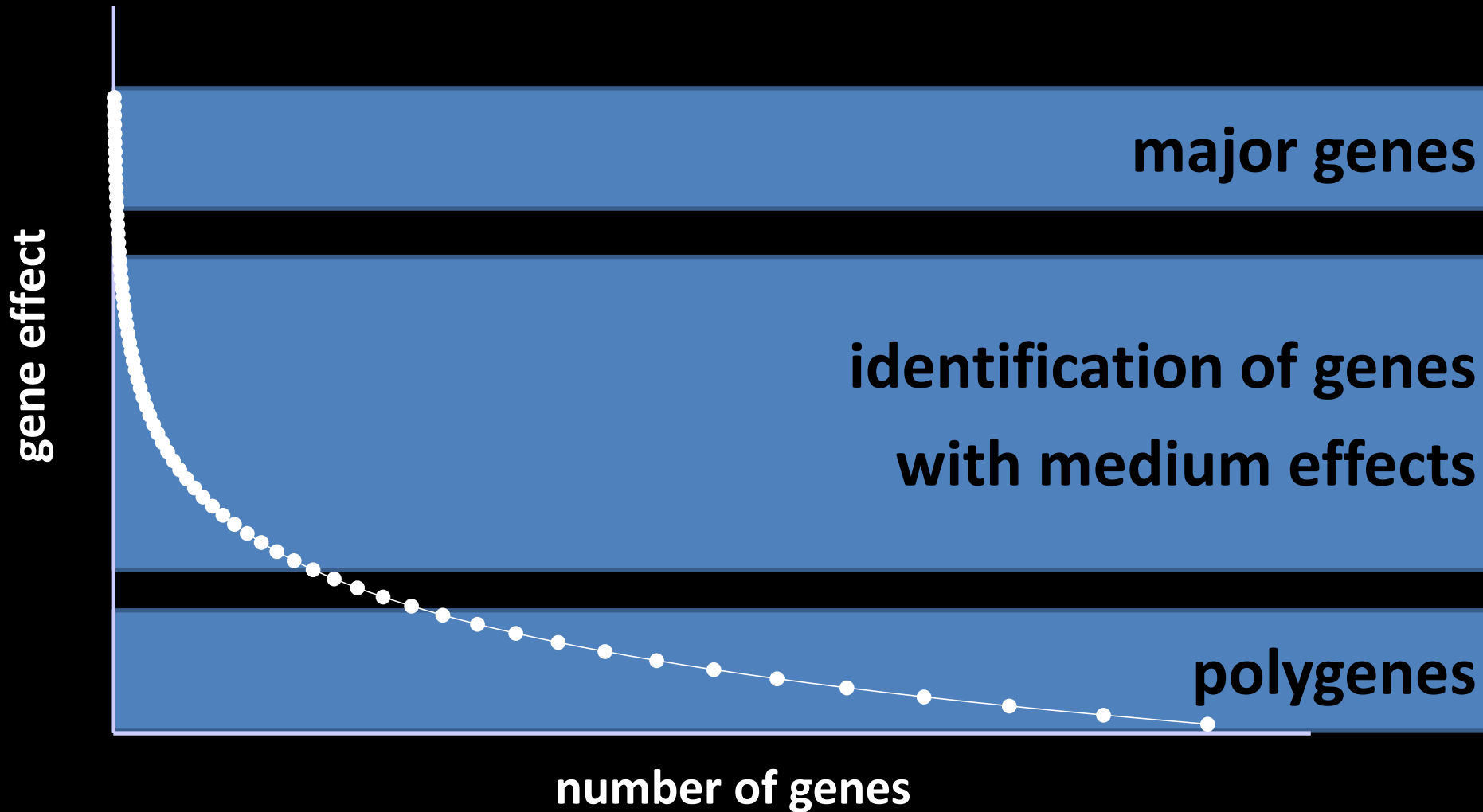


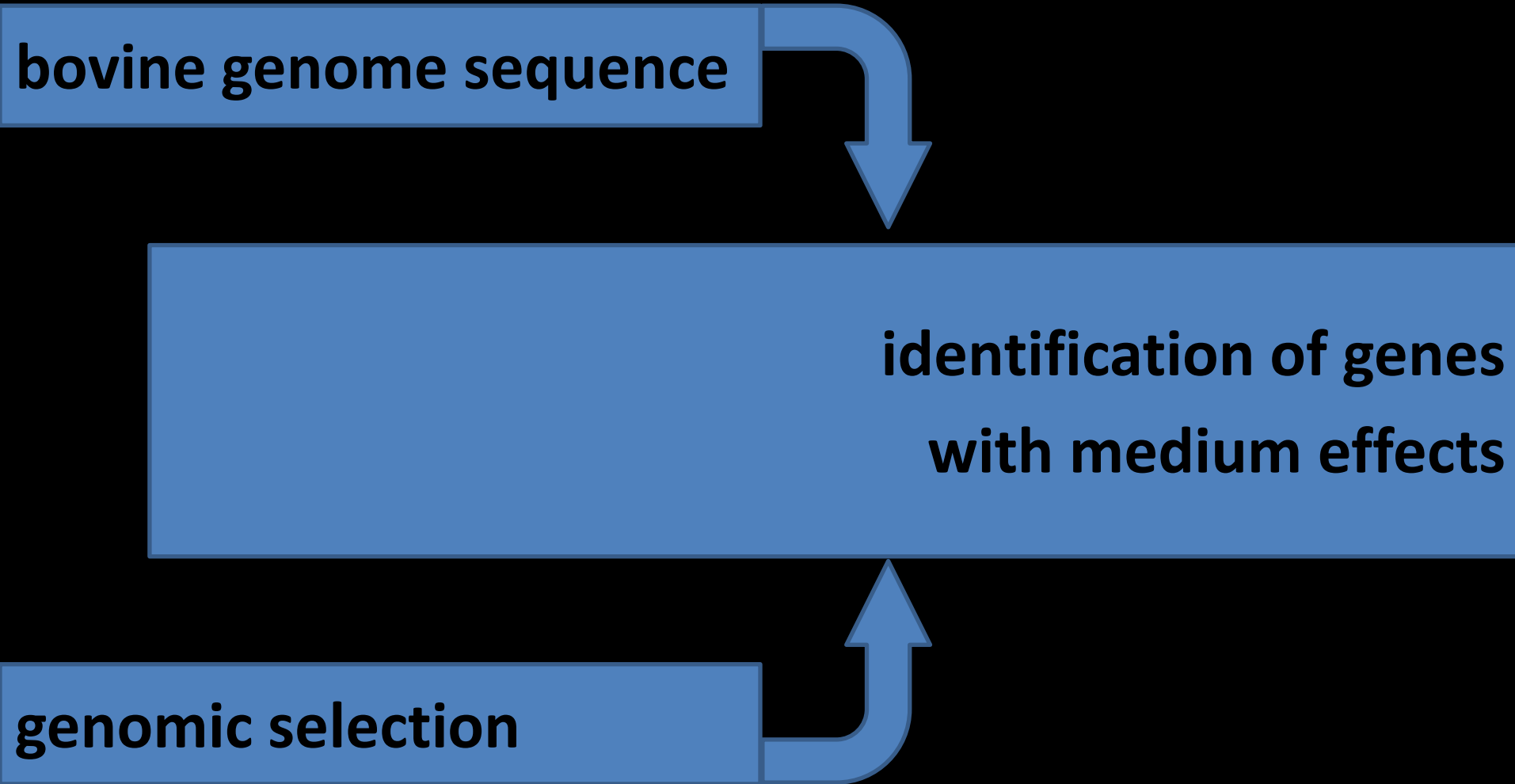


# Constructing gene networks underlying traits routinely recorded in dairy cattle

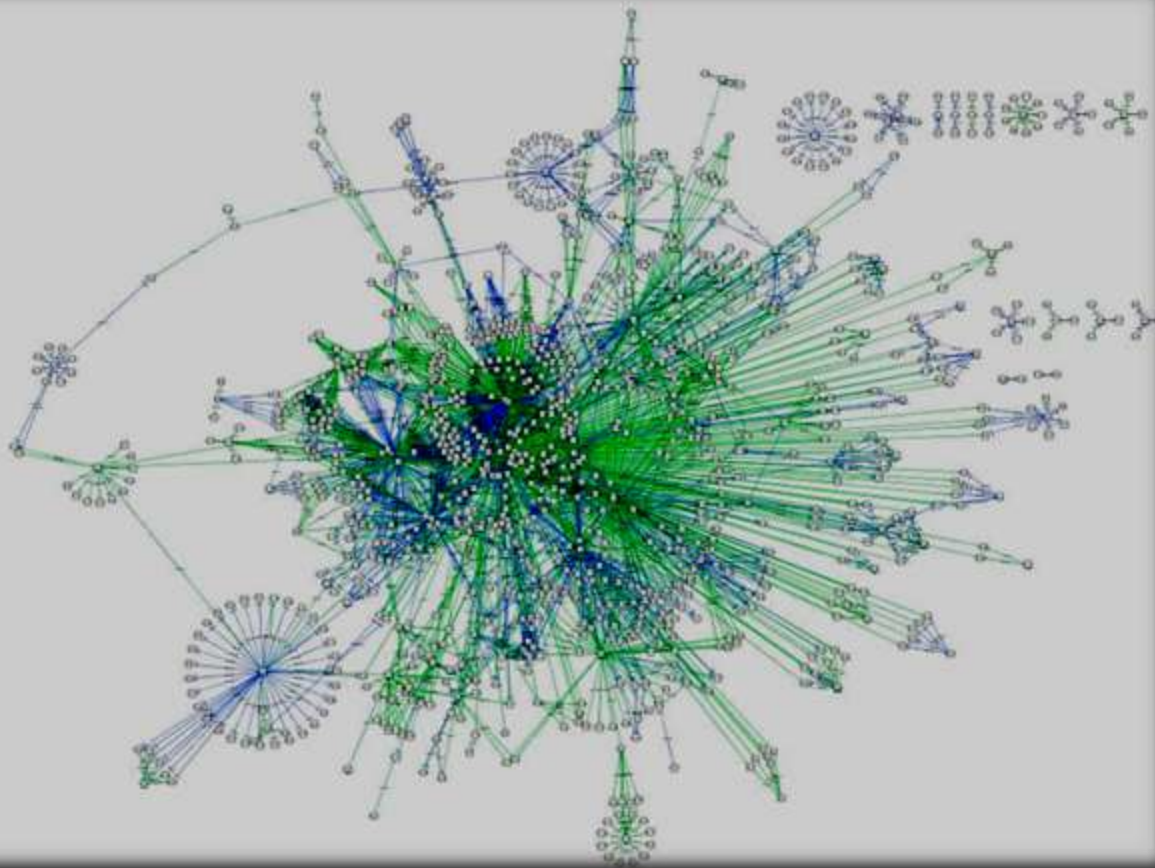
Joanna Szyda Stanisław Kamiński Andrzej Żarnecki

- 1. Motivation**
- 2. Data**
- 3. Gene selection**
- 4. Gene network**
- 5. Results**
- 6. Conclusions**





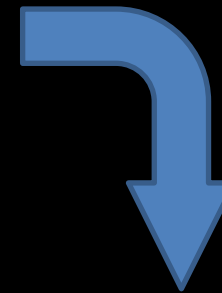
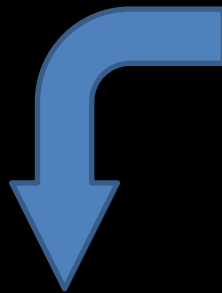
gene regulatory network  
anchored on known genes



**protefat yield**



**2461 HF bulls**



**genotypes:**  
**Illumina 50K chip**

**phenotypes:**  
**dEBV for 29 traits**

# Gene selection procedure: 1. SNP effect estimation

$$y = \mu + Zq + e$$

- $y$  EBV
- $\mu$  general mean
- $q$  **SNP**  $\sim N(0, \mathbf{I}\sigma_q^2)$
- $Z \in \{ -1, 0, 1 \}$
- $e$  residual



**estimates of 46 267  
additive effects  
á 29 traits**

# Gene selection procedure: 2. SNP selection

$$q \text{ SNP} \sim N(0, \mathbf{I}\sigma_q^2)$$



**CV for  $\alpha=0.20$**



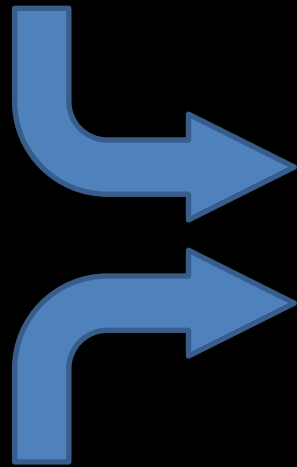
**Bonferroni**

**N=46 267**

# Gene selection procedure: 3. SNP annotation

**annotation file**

ENSEMBL v. 63 27.06.2011



**editing**

SNP  $\pm$  1000 bp  
from a gene



**significant  
genes**

**significant SNPs**

genomic evaluation 04.2011

# Gene network construction

gene list



**Cytoscape**

Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization



**Bisogenet**

*Martin et al. 2010 BMC Bioinformatics*

- NCBI
- UniProt
- KEGG pathway
- Gene Ontology
- protein interaction
- gene interaction



no cattle data



human data used




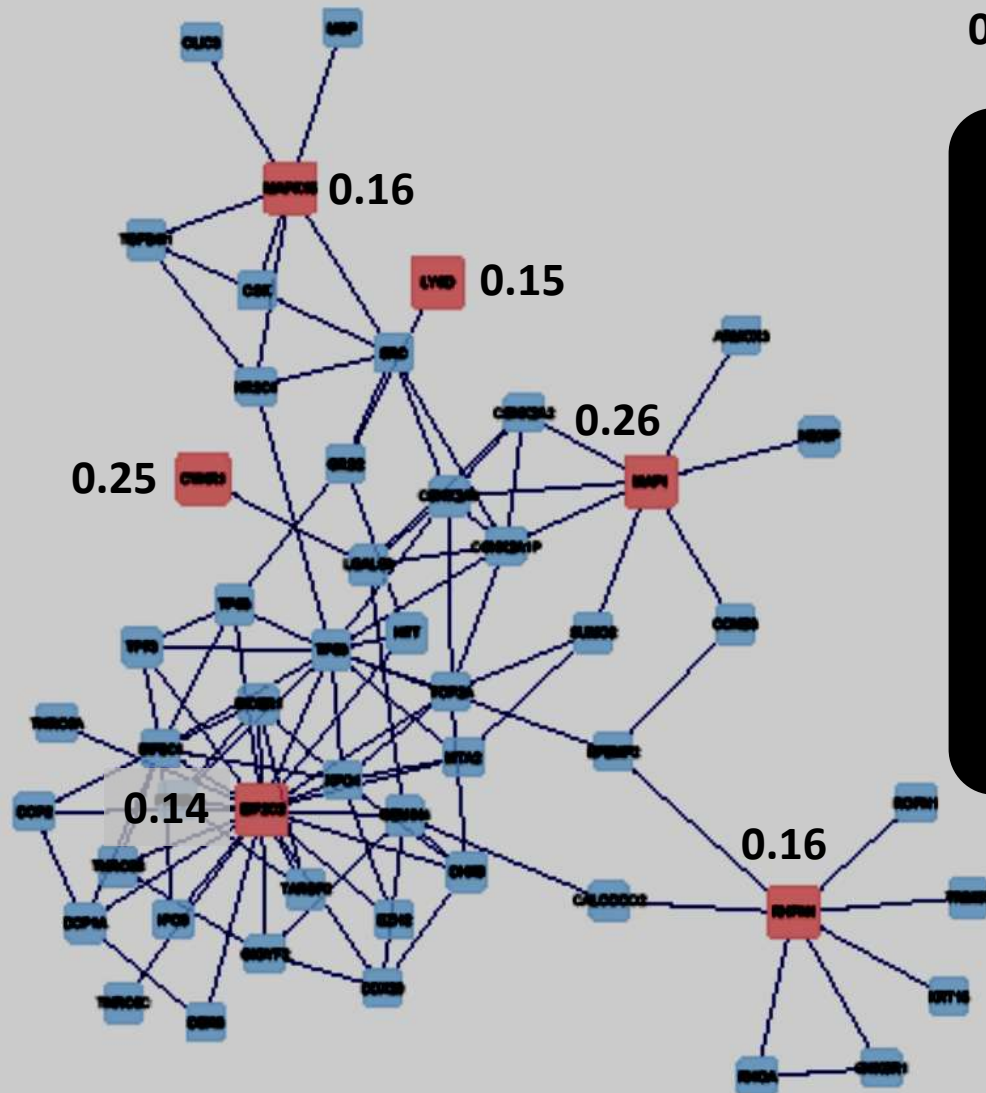
- **5 significant SNP identified**
- **BTA: 1, 5, 6, 25**
- **2 SNPs annotated within genes**
  - **1 locus unknown**
  - **1 locus no interactions present**



- **34 significant SNP identified**
- **BTA: 2, 3, 7, 14, 26**
- **9 SNPs annotated within genes**
  - **2 loci unknown**
  - **1 locus no interactions (DGAT1)**
  - **6 loci interactions found**

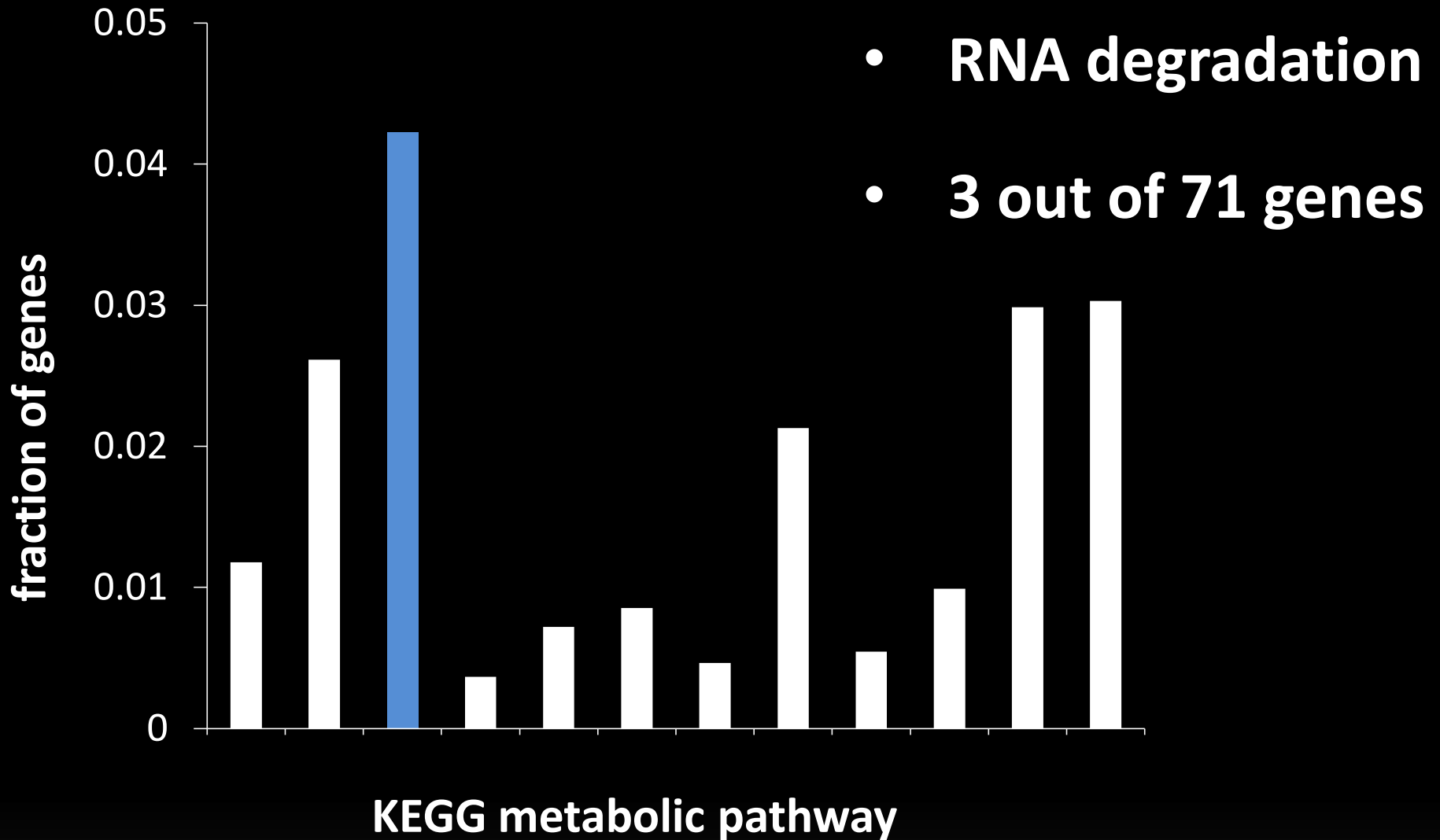
# Results: fat yield

0.29 



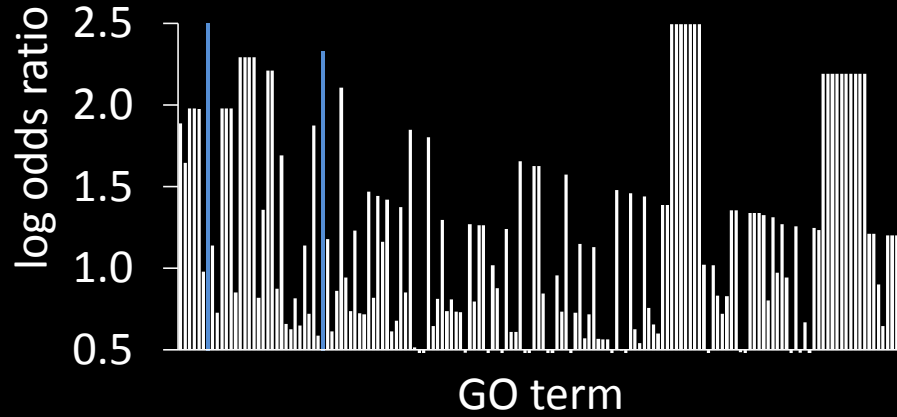
- 54 genes
- BTA: 1-7, 10-11, 13-16, 18-19, 21-25, 29
- 20 genes - metabolic pathways known
- 51 genes - GO terms known

# Results: fat yield, metabolic pathways



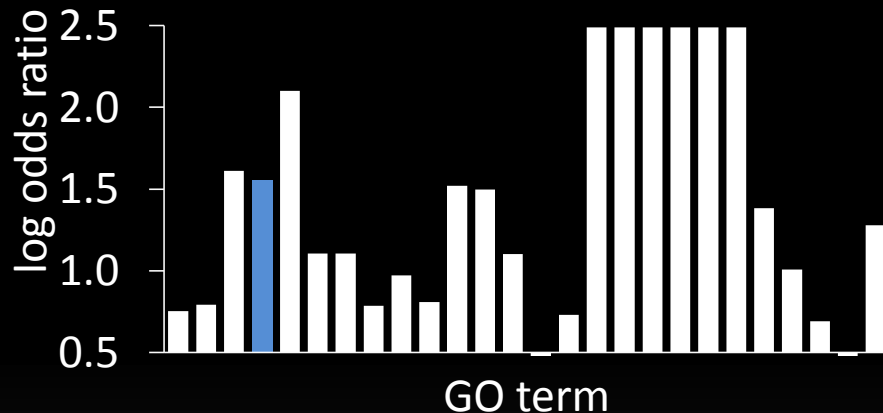
# Results: fat yield, gene ontology

## Biological Process



- 150 GO with  $P < 0.05$
- gene silencing
- 8 out of 38 genes  $P < 3^{-10}$
- RNA interference
- 3 out of 6 genes / 2 out of 4 genes

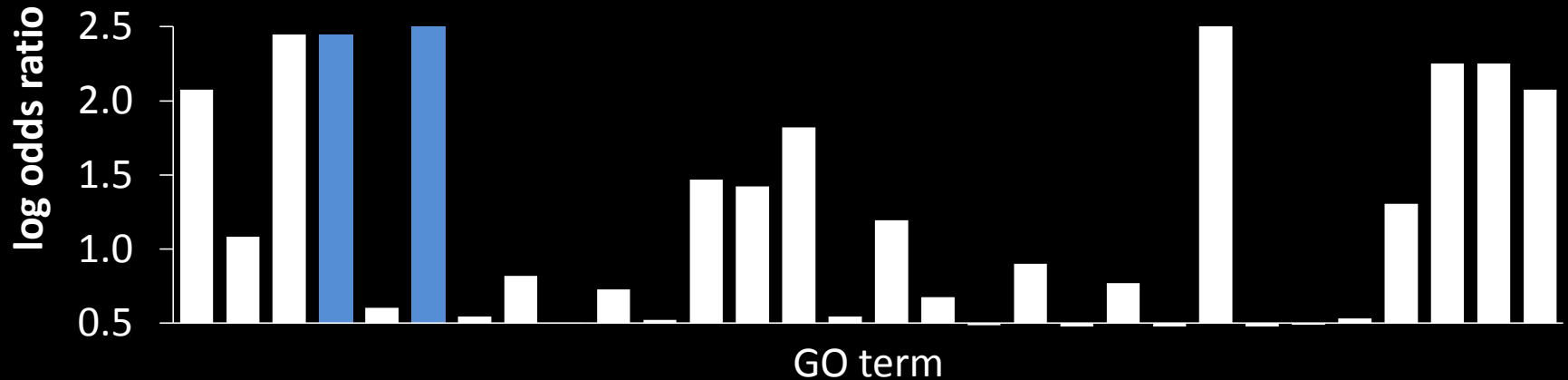
## Molecular Function



- 26 GO with  $P < 0.05$
- protein binding
- 44 out of 51 genes  $P < 9^{-5}$
- small interfering RNA binding
- 2 out of 7 genes

# Results: fat yield, gene ontology

## Cellular Component



- 29 GO with  $P < 0.05$
- cytoplasmic mRNA processing body
- 8 out of 38 genes  $P < 5^{-11}$
- RNA-induced silencing complex / micro-ribonucleoprotein complex
- 5 out of 10 genes / 2 out of 4 genes

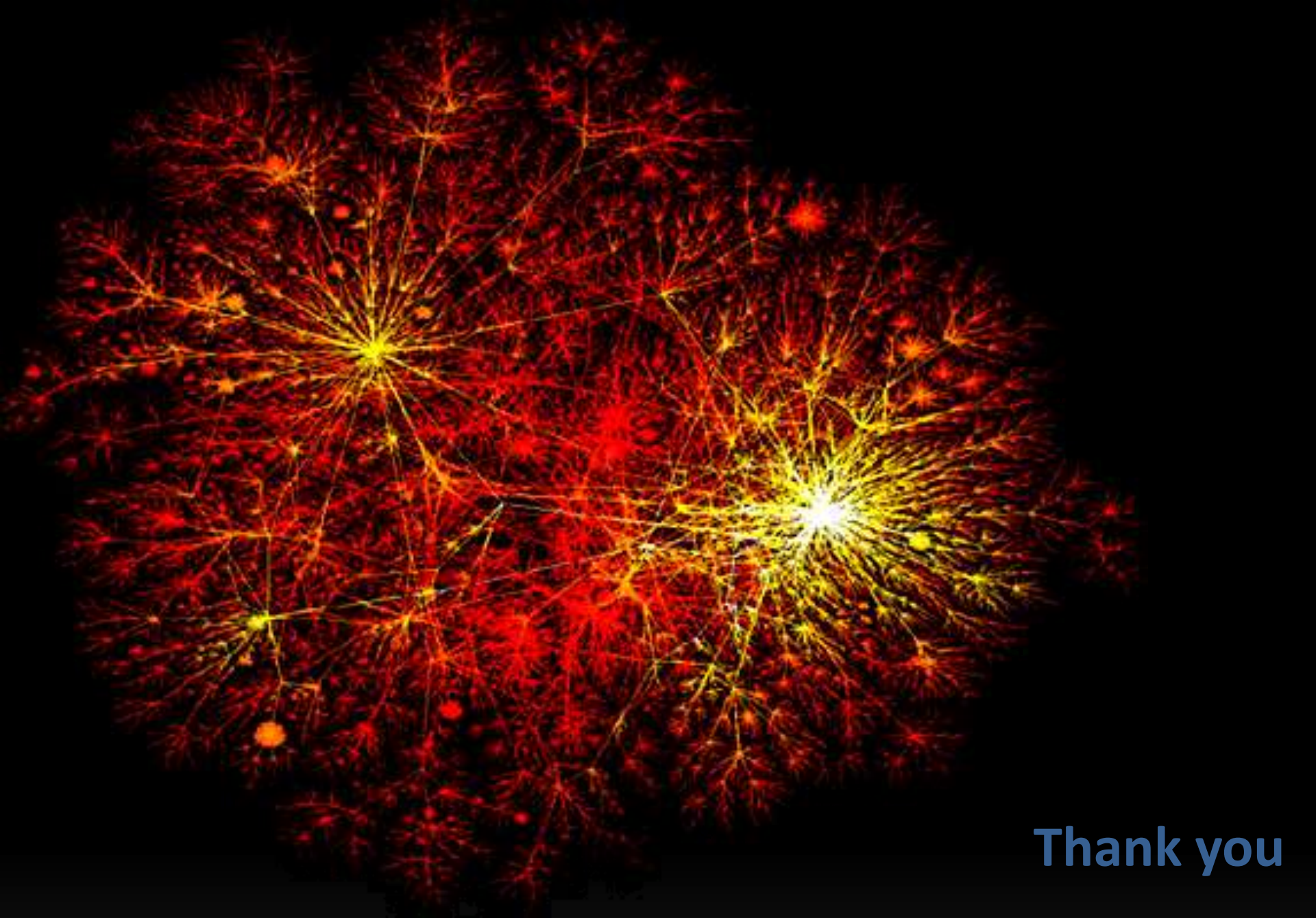
- 1. Fat yield regulated by RNA genes ???**
- 2. Poor resolution with 50K chip**
  - many SNPs outside genes
  - 777K better
- 3. No data for *Bos taurus***

## 1. SNP choice

- how many SNPs
- consider LD
- $\alpha$
- correct for multiple testing

## 2. Network validation

- individual genotyping - expensive
- Individual sequencing - even more expensive
- permutation
- replication across different (national) data sets



Thank you