

# Increased prediction accuracy using a genomic feature model including information on QTLs and expression patterns

Pernille Sarup

*J. Jensen, M. Shirali, T. Ostersen, M. Henryon*

*K. Wimmers, H. Reyer, P. Varley and P. Sørensen*



# Genomic feature models

*From sequence ...*

*...to consequence*

- Premise of the GFBLUP model is that genomic features are enriched for causal variants affecting the traits.
  - Chromosomes/Genes
  - Biological Pathways
  - Gene or Sequence Ontologies
  - **Prior QTL regions**
  - **QTL regions from study population single marker/blocks**
  - **Expression** or Methylation patterns
  - Protein-Protein or Protein-Metabolite interactions .....



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Aim of study

*From sequence ...*

*...to consequence*

## **Genomic feature modeling approaches:**

Genomic feature best linear unbiased prediction model  
(GFBLUP)

## **Evaluated by:**

**Estimation of genomic parameters, and prediction ability of genomic feature models in two independent data sets.**



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# GFBLUP model

*From sequence ...*

*...to consequence*

$$\hat{y} = \mu + \mathbf{Z}_f \mathbf{f} + \mathbf{Z}_r \mathbf{r} + \mathbf{e}$$

$$\mathbf{f} \sim N(0, \mathbf{G}_f \sigma_f^2)$$

$$\mathbf{r} \sim N(0, \mathbf{G}_r \sigma_r^2)$$

$$\mathbf{e} \sim N(0, \mathbf{I} \sigma_e^2)$$

$$\mathbf{G}_f = (\mathbf{W}_f \mathbf{W}_f' / N_f)$$

$$\mathbf{G}_r = (\mathbf{W}_r \mathbf{W}_r' / N_r)$$

f, r, and e are assumed to be uncorrelated

# GBLUP derived set tests

*From sequence ...*

*...to consequence*

- Premise of the GFBLUP model is that genomic features are enriched for causal variants affecting the traits.
- The number, location and effect sizes of the true causal variants in the genomic feature are unknown.
- Results from set tests can be used in GFBLUP model

## **Covariance Association Test (CVAT) Identifies Genetic Markers Associated with Schizophrenia in Functionally Associated Biological Processes**

 Palle Duun Rohde, Ditte Demontis, Beatriz Castro Dias Cuyabano, The Genomic Medicine for Schizophrenia Group, Anders D. Børglum, Peter Sørensen  
*GENETICS August 1, 2016 vol. 203 no. 4 1901-1913; DOI: 10.1534/genetics.116.189498*

## **Genomic Prediction for Quantitative Traits Is Improved by Mapping Variants to Gene Ontology Categories in *Drosophila melanogaster***

Stefan M. Edwards, Izel F. Sørensen, Pernille Sarup, Trudy F. C. Mackay, Peter Sørensen  
*GENETICS August 1, 2016 vol. 203 no. 4 1871-1883; DOI: 10.1534/genetics.116.187161*



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Estimation of genomic parameters

From sequence ...

...to consequence

## Genomic heritability

$$h^2 = \frac{\sigma_f^2 + \sigma_r^2}{\sigma_f^2 + \sigma_r^2 + \sigma_e^2}$$

## Proportion of genetic variance explained by feature

$$h_f^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_r^2}$$

## Prediction ability

$$pa = cor(\hat{y}, \hat{y})$$



# Example 1

From sequence ...

...to consequence

RESEARCH ARTICLE | OPEN ACCESS

## Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs

[Pernille Sarup](#)  , [Just Jensen](#), [Tage Ostersen](#), [Mark Henryon](#) and

[Peter Sørensen](#)

*BMC Genetics* BMC series – open, inclusive and trusted 2016 17:11 | DOI: 10.1186/s12863-015-0322-9 | © Sarup et al. 2016

Received: 8 September 2015 | Accepted: 20 December 2015 | Published: 5 January 2016



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Duroc data

*From sequence ...*

*...to consequence*

## Genotypes for Duroc:

3085 boars, SNP60 BeadChip (33,029 SNPs)

## Corrected phenotypes for genotyped individuals:

EBVs + residuals from BLUP on ~34K individuals

Feed Efficiency **FE**

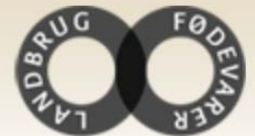
Lean Meat Percentage **LMP**

## Split into Training and Validation set:

Training set: 1814 Duroc boars born 1998-2010

Validation set: 1271 Duroc boars born 2012-2014

Danish Pig  
Research Centre



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.





# Defining genomic features

*From sequence ...*

*...to consequence*

- Chromosomes/Genes
- Biological Pathways
- Gene or Sequence Ontologies
- **Prior QTL regions**
- **QTL regions from study population single marker/blocks**
- Expression or Methylation patterns
- Protein-Protein or Protein-Metabolite interactions .....



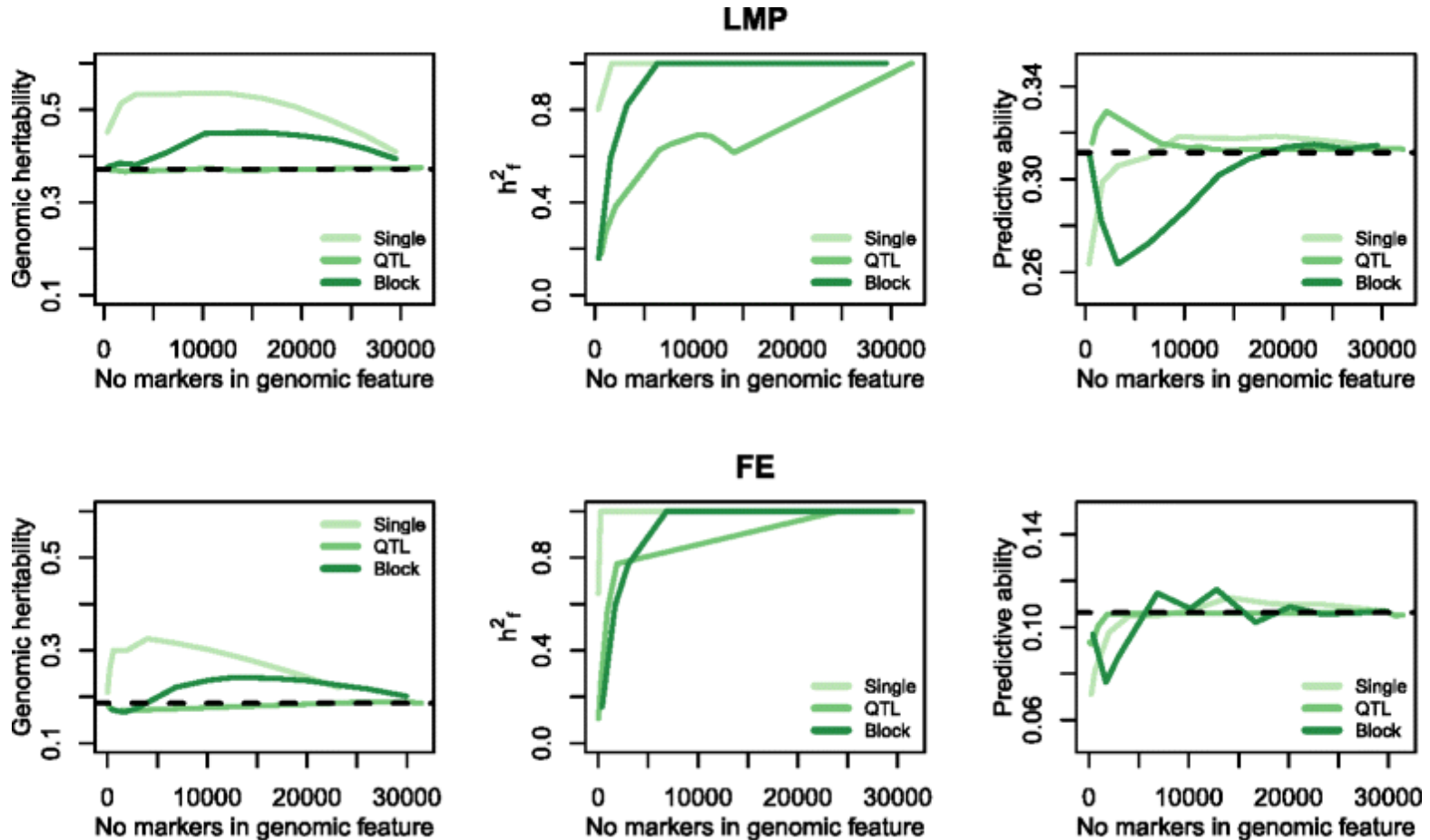
The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Results

From sequence ...

...to consequence



# Conclusion

*From sequence ...*

*...to consequence*

Increased PA for LMP when external data sources was incorporated.

Power to detect causal variants vital



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Example 2

*From sequence ...*

*...to consequence*

Genotypes for 2213 Hermitage MaxGro pigs  
from

~55K SNPs.

Corrected phenotypes for genotyped individuals

EBVs + residuals from BLUP on ~14K individuals



Total Feed Intake **TFI**

Residual Feed Intake **RFI**

50 random 10 fold validation sets.



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Genomic feature models

*From sequence ...*

*...to consequence*

**Genomic features are defined by grouping SNPs according to a certain classification scheme such as:**

- Chromosomes/Genes
- Biological Pathways
- Gene or Sequence Ontologies
- **Prior QTL regions**
- QTL regions from study population single marker/blocks
- **Expression** or Methylation patterns
- Protein-Protein or Protein-Metabolite interactions .....



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Defining genomic features

*From sequence ...*

*...to consequence*

Some of the SNPs was linked to genes that differ in gene expression between high RFI and low RFI pigs



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.

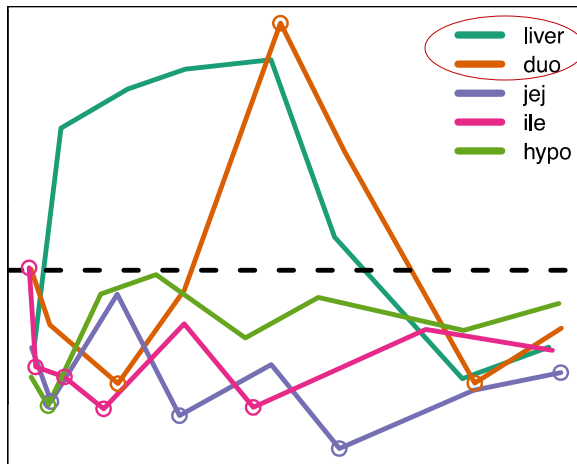


# results

From sequence ...

...to consequence

## Residual Feed Intake



## Total Feed Intake

GBLUP

○ p < 0.001

# Conclusions

*From sequence ...*

*...to consequence*

Prediction accuracies improved by ~10%.

Genes that are differential expressed in liver and duodenum most informative.



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.





# Conclusions

*From sequence ...*

*...to consequence*

**GFBLUP models can increase prediction ability and contribute knowledge on the biological background of complex traits**



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.



# Thank you for your attention

*From sequence ...*

*...to consequence*

## Genotype and phenotype data provided by:



*P. Varley*

## Gene expression experiments

*K. Wimmers, H. Reyer*



*M. Henryon, T. Ostersen*



## Model development and genomic prediction:

*P. Sørensen, P. Sarup, J. Jensen, M. Shirahata*

CENTER FOR QUANTITATIVE  
GENETICS AND GENOMICS



The ECO-FCE project is funded by the European Union Seventh Frame Programme (FP7 2007/2013) under grant agreement No. 311794.

