Understanding the Genetic Basis of Complex Traits (SFA1)

Peter Sørensen

Center for Quantitative Genetics and Genomics (QGG)

Department of Molecular Biology and Genetics
Aarhus University
Denmark

Today's Talks

- **Genomic feature models**
  Senior Scientist Peter Sørensen

- **Gene mapping in cattle: lessons learnt from genome-wide variants**
  Senior Scientist Goutam Sahana, QGG, AU

- **Understanding gene by environment interactions in ryegrass**
  Professor Torben Asp, Dept. of Molecular Biology & Genetics, AU

- **Building genomic resources for breeding in a new species (mink)**
  Postdoc Zexi Cai, QGG, AU (Goutam Sahana)

Genomic Feature Models

.....from understanding to prediction

Genomic Feature Models

Based on a simple hypothesis:

1. Causal mutations are clustered in regions on the genome defined by **genomic features** such as:
   - biological pathways
   - gene or sequence ontologies
   - prior QTL regions (QTL, eQTL, mQTL, pQTL, ...)
   - expression or methylation patterns
   - protein-protein or protein-metabolite interactions .....

2. If we use a **statistical model that quantifies the effect of a set of genetic variants** defined by a genomic feature we can

   - increase detection power for causal variants with small effects
   - increase prediction accuracy of complex trait phenotypes

Genomic Feature Models

Prediction Models

- Mapping Variants to Gene Ontology Categories Improves Genomic Prediction for Quantitative Traits in *Drosophila melanogaster*. Edwards SM, Sørensen IF, Sarup P, Mackay TF, Sørensen P. (2016). Genetics 203 (4): 1871-1883.
- Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. Fang L, Sahana G, Ma P, Su G, Zhang S, Yu Y, Lund MS, Sørensen P. 2018. *BMC Genomics* 18(1):604

Marker Set Association Models

- MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle. Fang L, Sørensen P, Sahana G, Panitz F, Su G, Zhang S, Yu Y, Li B, Ma L, Liu G, Lund MS, Thomsen B. 2018. *Sci Rep* 8:1–14
- Multiple trait covariance association test identifies gene ontology categories associated with chill coma recovery time in *Drosophila melanogaster*. Sørensen IF, Edwards SM, Rohde PD, Sørensen P. 2017. *Sci Rep* 7:2413.

Implemented in R (psoerensen.github.io/qgg)

- qgg: an R package for large-scale quantitative genetic analyses. Rohde PD, Sørensen IF, Sørensen P. Bioinformatics 2019

Genomic Feature BLUP Models

Predictions in GFBLUP

GBLUP same weights on genomic relationships

$$\mathbf{g}_{\text{pred}} = \left(\mathbf{G}_{\text{vt}} \cdot \sigma_{\text{g}}^2\right)\left[\left(\mathbf{G}_{\text{tt}} \cdot \sigma_{\text{g}}^2\right) + \mathbf{I} \cdot \sigma_{\text{e}}^2\right]^{-1}\left(\mathbf{y}_{\text{t}} - \mathbf{Xb}\right)$$

$$\mathbf{g}_{\text{pred}} = \left(\mathbf{G}_{\text{f}_{\text{vt}}} \cdot \sigma_{\text{f}}^2 + \mathbf{G}_{\text{r}_{\text{vt}}} \cdot \sigma_{\text{r}}^2\right)\left[\mathbf{G}_{\text{f}_{\text{tt}}} \cdot \sigma_{\text{f}}^2 + \mathbf{G}_{\text{r}_{\text{tt}}} \cdot \sigma_{\text{r}}^2 + \mathbf{I} \cdot \sigma_{\text{e}}^2\right]^{-1}\left(\mathbf{y}_{\text{t}} - \mathbf{Xb}\right)$$

GFBLUP different weights on genomic relationships
=> differential shrinkage

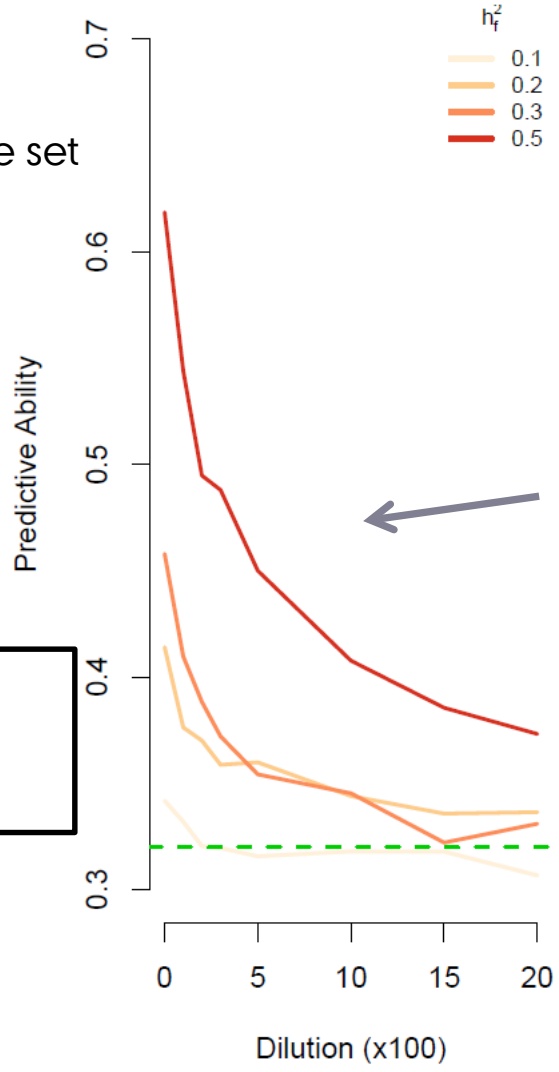
Edwards et al. 2016

Genomic Feature BLUP Models

Effect of dilution
- adding non-causal SNPs to feature set

Cross validation
- 10% test and 90% train
- predictions across lines
- repeated 50 times

$$PA = \frac{1}{50}\sum_{i=1}^{50} Corr(y_{pred}, y_{obs})$$



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

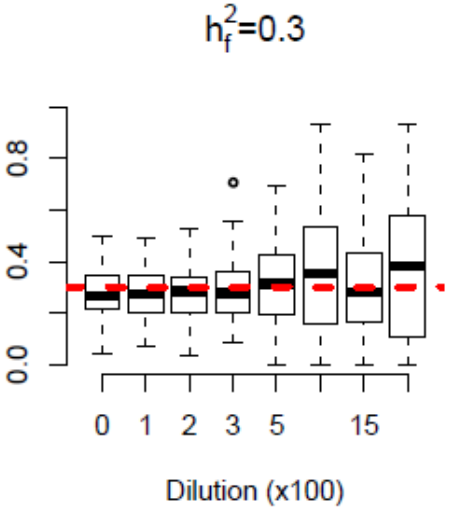
GFBLUP can increase accuracy

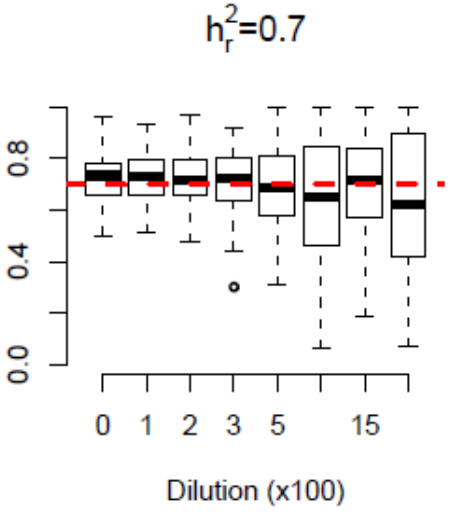
GBLUP

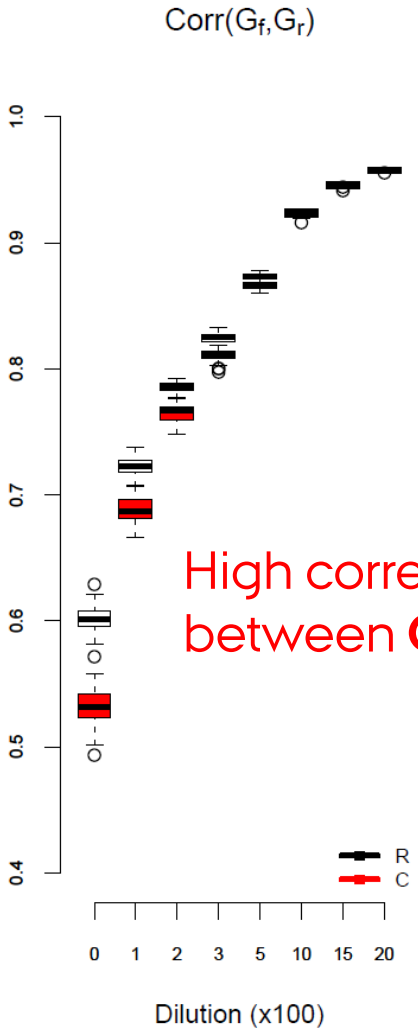
Edwards et al. 2016.

Genomic Feature BLUP Models

If “diluted” difficult to reliably estimate genomic parameters => lower accuracy







High correlation between $\mathbf{G}_f$ and $\mathbf{G}_r$

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

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

Edwards et al. 2016

Genomic Feature BLUP Models

- Increased prediction accuracy provided genomic feature is enriched for causal variants and highly dependent on:
    - heritability (h^2 and h_f^2)
    - relatedness/LD
    - dilution

- Small increase (1-3%) within breeds (Danish Duroc/Danish Holstein/Danish Jersey)
- Larger increase (5-20%) across breeds (Danish Holstein -> Danish Jersey)
- Largest increase (10-50%) in populations of unrelated individuals (Drosophila (DGRP) and humans (UK Biobank))

Edwards et al 2016, Sarup et al. 2016, Fang et al. 2017, Rohde et al 2018

Marker Set Association Models

- GFBLUP (and Bayesian GF) models are computationally intensive

- Patterns derived from single-marker statistics can reveal associations between a set of genetic markers (genomic feature) and a complex trait.

- Marker set tests are computationally fast and powerful modelling approaches that allow us to rapidly analyze many different layers of genomic features

- Improved inference and prediction accuracy of GFBLUP may be achieved by identifying genomic regions enriched for causal genetic variants.

Rohde et al. 2016,2018, Fang et al 2017, Sørensen et 2017

Marker Set Association Models

Marker set statistics from the single marker statistics such as:

1. Count number of single marker test statistics above a certain threshold

$$T_{count} = \sum_{i=1}^{n_F} I(t_i > t_0)$$

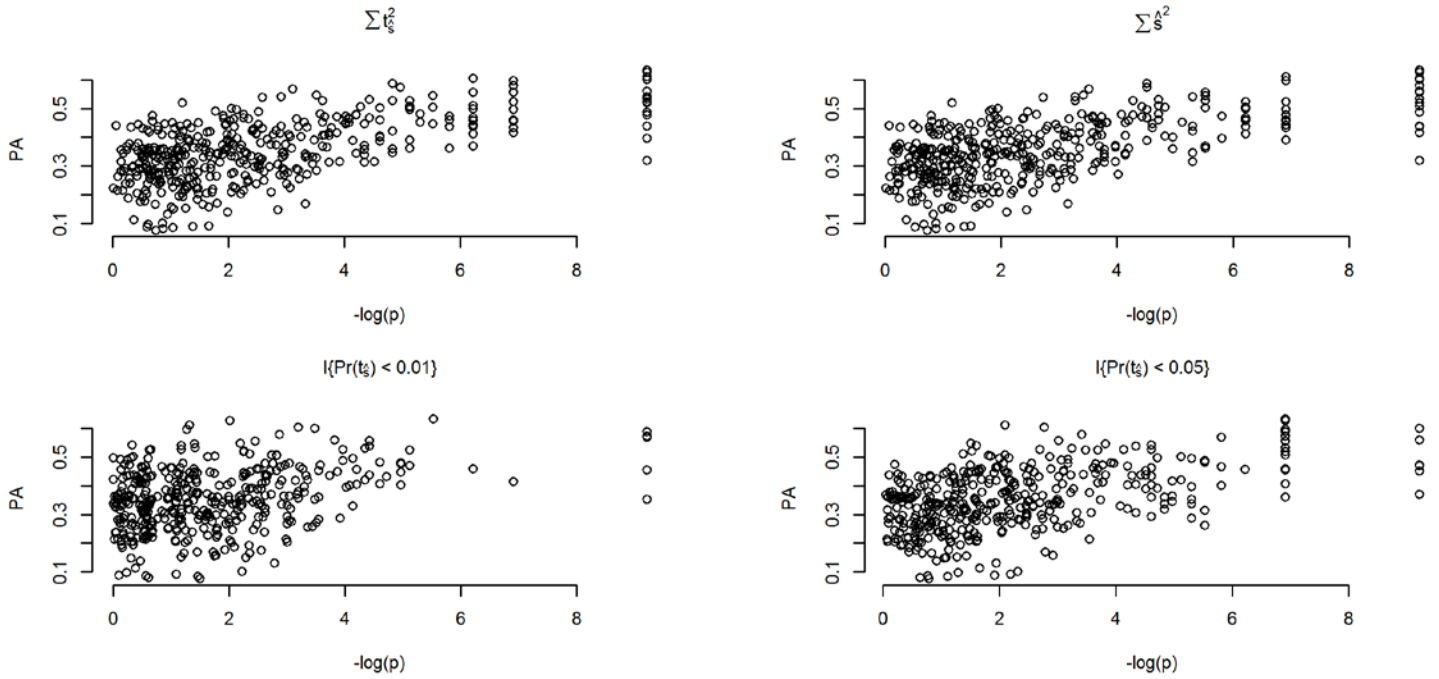
2. Sum of all single marker test statistics

$$T_{sum} = \sum_{i=1}^{n_F} t_i$$

Rohde et al. 2016, Sørensen et 2017

Marker Set Test Association Models

Simulated data example



Strong relationship between enrichment score (-log(p)) and prediction accuracy (PA) of GFBLUP model

Genomic Feature BLUP Models

GFBLUP models in dairy cattle – study 1

Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds

Fang et al. 2017 *Genet Sel Evol* 49:1–18.

- Compare GBLUP and GFBLUP models
- Within and across breed prediction
- Danish Holstein/Danish Jersey dairy cattle breeds
- Gene Ontology terms used as genomic features (SNP->gene->GO term)
- Milk production and mastitis traits

Genomic Feature BLUP Models

Top-ranking GO terms

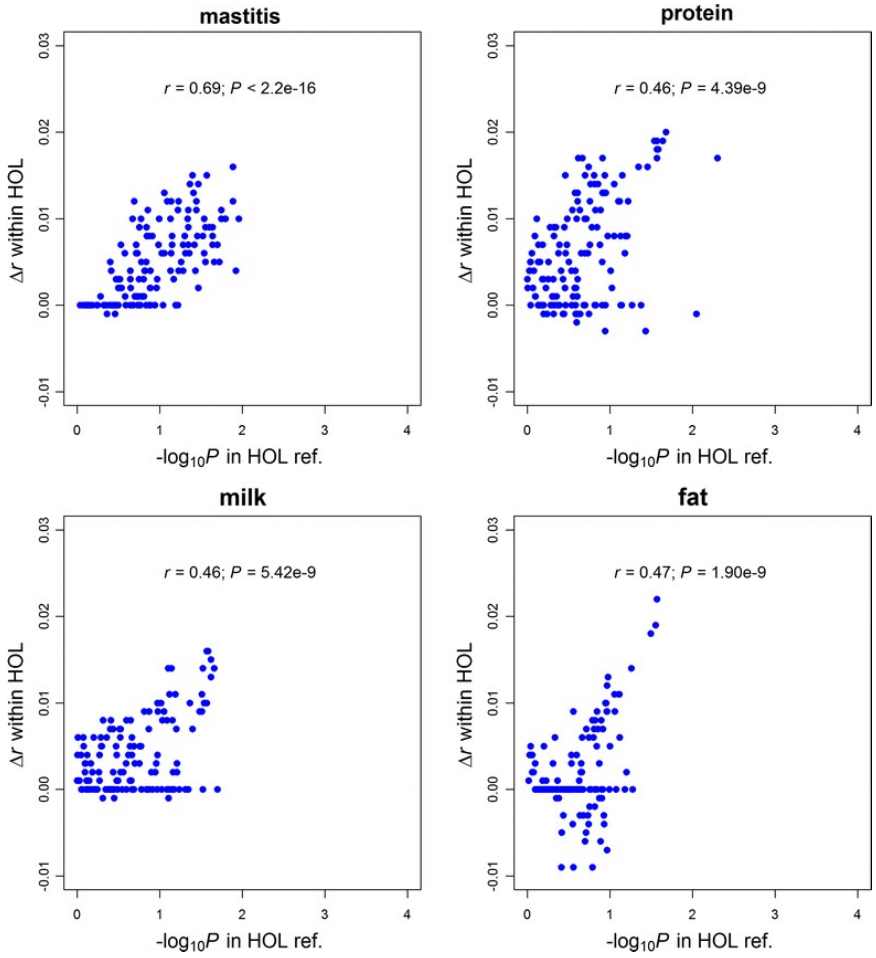
Table 2 Top five Gene Ontology (GO) terms with GFBLUP between breeds for the four traits

Trait	GO ID	$r^a_{\_GFBLUP}$	bias[b]	Δr^c	$(H_f^2)^d$	Nsets[e]	GO term
Milk	GO:0040018	0.360	0.826	0.200	0.103	962	Positive regulation of multicellular organism growth
	GO:0042572	0.342	0.808	0.182	0.171	678	Retinol metabolic process
	GO:0034605	0.336	0.805	0.176	0.178	1621	Cellular response to heat
	GO:0045944	0.331	0.805	0.171	0.190	11,185	Positive regulation of transcription from RNA polymerase II promoter
	GO:0032496	0.325	0.798	0.165	0.129	1702	Response to lipopolysaccharides
Mastitis	GO:0043066	0.077	0.277	0.135	0.064	8831	Negative regulation of apoptotic process
	GO:0032496	0.067	0.176	0.125	0.020	1702	Response to lipopolysaccharides
	GO:0032091	0.045	0.171	0.103	0.032	702	Negative regulation of protein binding
	GO:0043280	0.018	0.178	0.076	0.003	583	Positive regulation of cysteine-type endopeptidase activity involved in apoptotic process
	GO:0071346	0.014	0.115	0.072	0.020	3494	Cellular response to interferon-gamma

Fang et al. 2017

Marker Set Association Models

Relationship between enrichment score and prediction accuracy



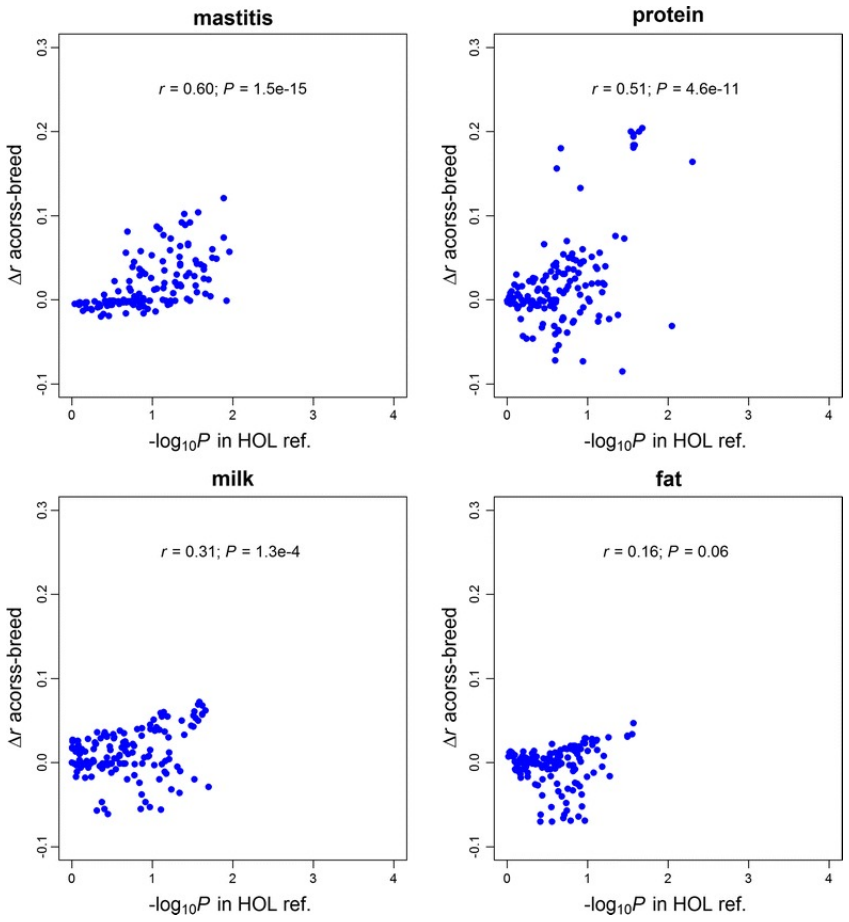
Within breed prediction

- Danish Holstein
- GBLUP<->GFBLUP
- Small increase (<2%)
- Similar trend across 4 traits

Fang et al. 2017

Marker Set Association Models

Relationship between enrichment score and prediction accuracy



Across breed prediction

- Danish Holstein => Danish Jersey
- GBLUP<->GFBLUP
- Larger increase (<20%)

Fang et al. 2017

Genomic Feature BLUP Models

GFBLUP models in dairy cattle – study 2

MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle.

Fang et al. 2018 *Sci Rep* 8:1–14.

MicroRNAs (miRNA) are key modulators of gene expression and so act as putative fine-tuners of complex traits phenotypes.

Hypothesis: Causal variants of complex traits are enriched in miRNAs and miRNA-target networks

Genomic Feature BLUP Models

- Enrichments analysis of association signals in miRNAs and their miRNA-target networks
  - 750 bovine autosome miRNA genes expressed in different tissues (miRbase)
  - SNPs 5kb up-/downstream
  - In silico prediction of miRNA targets

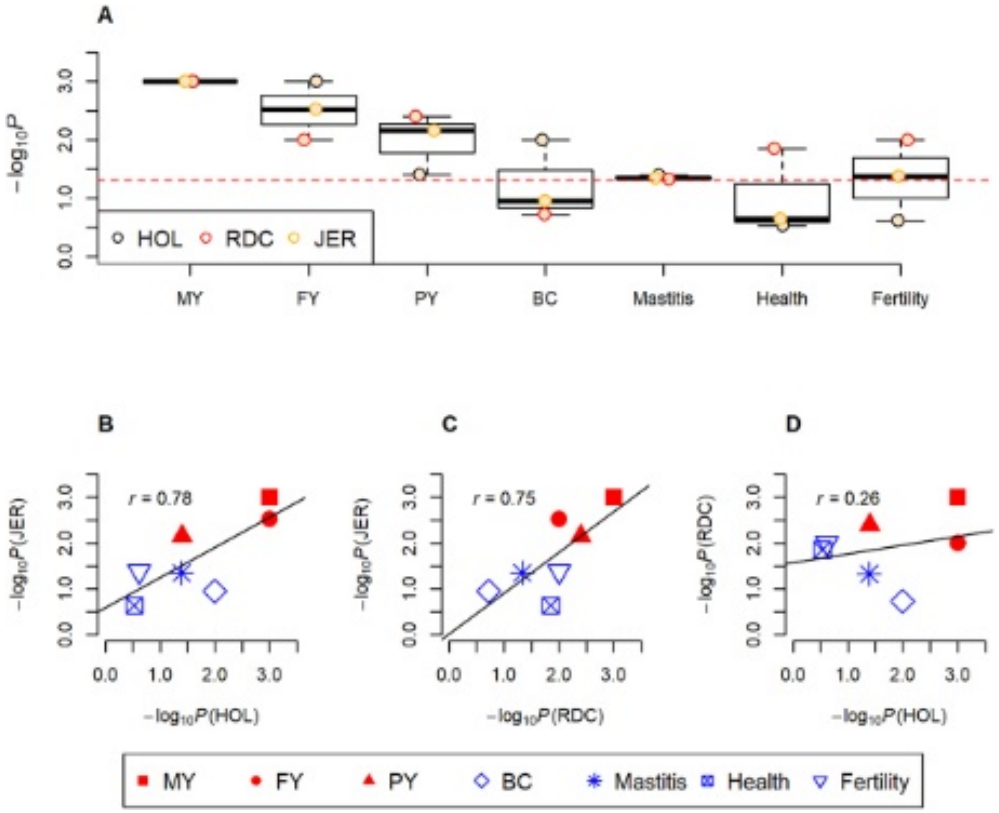
- Genome-wide association study (GWAS) for seven functional and milk production traits

- Imputed sequence variants (13~15 million)

- >10,000 animals from three dairy cattle breeds, i.e., Danish Holstein (HOL), Nordic Red Cattle (RDC) and Danish Jersey (JER).

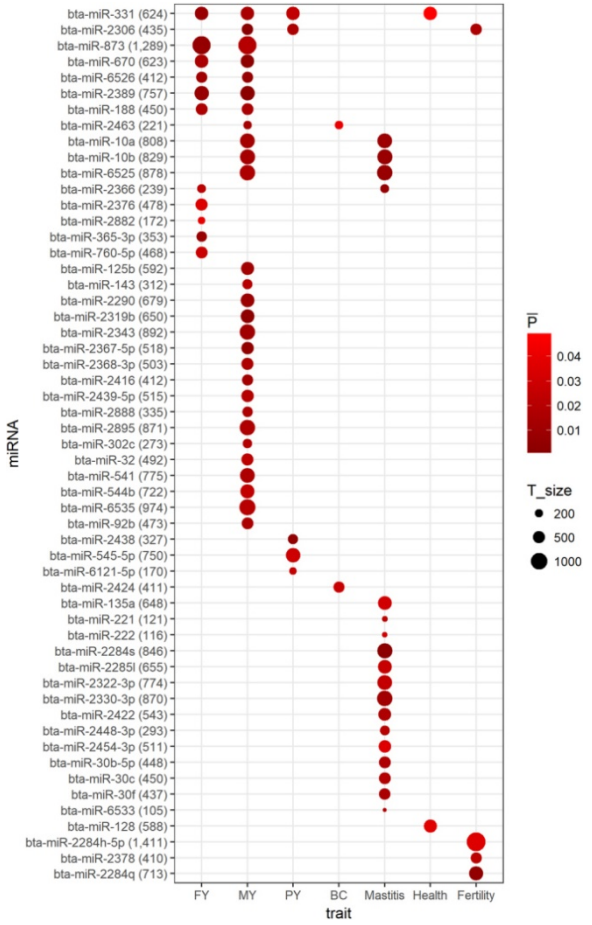
Marker Set Association Models

Genomic regions harboring miRNA genes significantly (P < 0.05) enriched with GWAS signals for milk production traits and mastitis



Marker Set Association Models

Genomic regions harboring miRNA-target gene networks significantly (P < 0.05) enriched with GWAS signals



- 55 significant miRNA-target networks were detected for seven traits
- 12 miRNAs were involved in several traits
- genes differentially expressed in response to mammary gland infections enriched in the miRNA-target networks associated with mastitis.
- findings consistent across three breeds.

Summary

- **GF prediction models** can increase the accuracy of genomic predictions provided feature is enriched for causal variants

- **Marker Set Association Models** can reveal genomic features associated with complex traits and can be used to improve GF prediction models

- **GF models applied to a range of features and traits**
  - GF models in dairy cattle (GO, GWE, miRNA) (Fang et al. 2017, 2018)
  - GF models in pigs (Animal QTLdb) (Sarup et al. 2016)
  - GF models in in fly (GO, ppi, QTL) (Rohde et al. 2016/3017/2018; Sørensen et al.2017)
  - GF models in humans (GO, ppi, QTL) (Rohde et al. 2016/2018; Sørensen et al.)
  - GF models for GxG and GxE in fly (Rohde et al. 2017, Morgante et al.)

- **Models implemented in R (psoerensen.github.io/qgg)**

Acknowledgements

Lingzhao Fang

Palle Duun Rohde

Bo Thomsen

Stefan McKinnon Høj-Edwards

Pernille Sarup

Izel Fourie Sørensen

Beatriz Castro Dias Cuyabano

Trudy Mackay NC State

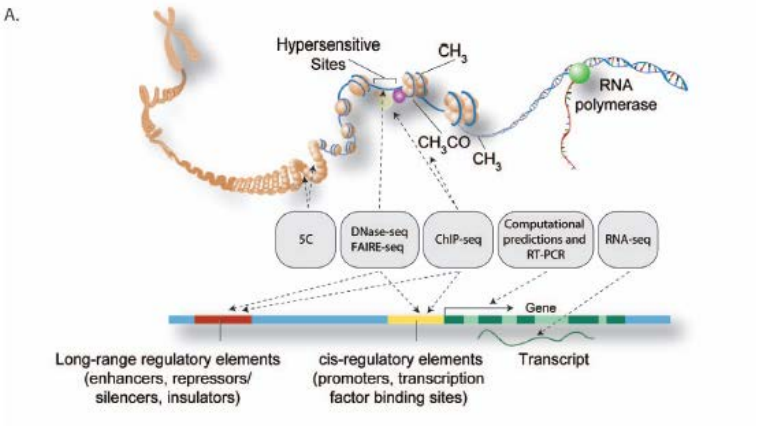
Fabio Morgante NC State

Danish Strategic Research Council

GenSAP: Centre for Genomic Selection in Animals and Plants, contract no. 12-132452

Data

Whole-genome sequences and **multiple novel trait phenotypes** from large numbers of individuals from **multiple populations**



Encyclopedia of DNA Elements (ENCODE)

FAANG

- **Molecular phenotypes** (e.g. transcriptome, proteome,metabolome, methylome) associated to the traits/diseases of interest
- **Molecular-interaction maps** that provide insight into the structural and functional organization of the genomes

Research Focus

Develop statistical models that can **use prior biological information**

- increase prediction accuracy of phenotypes or genetic predisposition
- provide novel insights into the genetic basis of the traits

Phenotype = Genome + Metabolome +Transcriptome+...+ residual

- Rapid accumulation of biological information in database
- Genetic architecture (few large, many small effects, gene by gene,.....)

"Easy" to detect using methods that allow for differential shrinkage (e.g. Bayesian mixture models).

Difficult to detect

Research Focus

Want to better understand genetic architecture of complex traits

- disentangle genetic variation
- disentangle genetic correlation

What is the contribution of different types of variants?

- rare versus common versus structural variants
- functional variants

What is the importance of type of effect?

- additive versus non-additive variance (GxG, GxE)

How do we reliably quantify these contributions?

- which factors influence inference?
- can we partition genomic (co)variance?
- what happens under selection?