

# Genomic prediction and GWAS with sequence information versus HD or 50k SNP chips

Roel Veerkamp, Rianne van Binsbergen, Mario Calus,  
Chris Schrooten, Aniek Bouwman

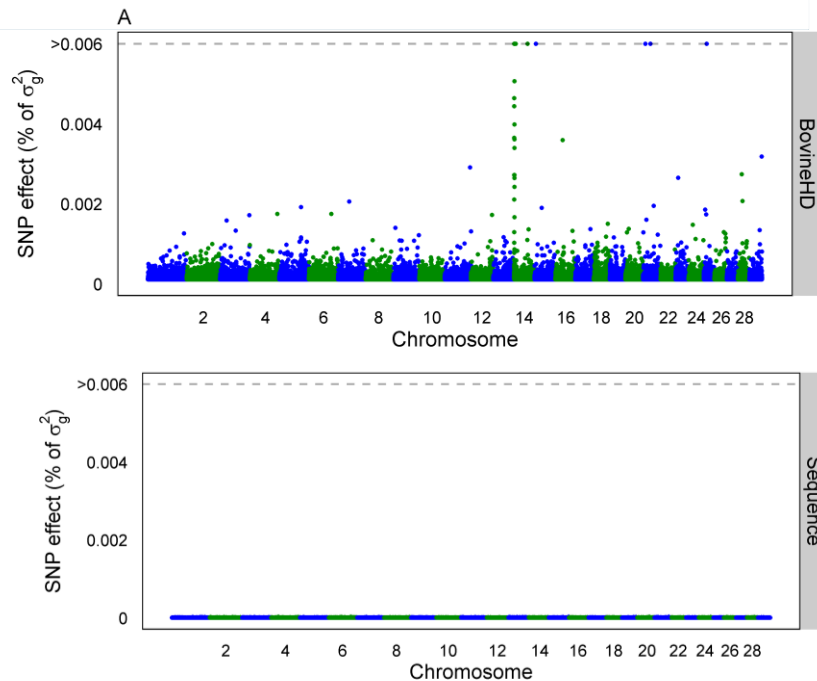
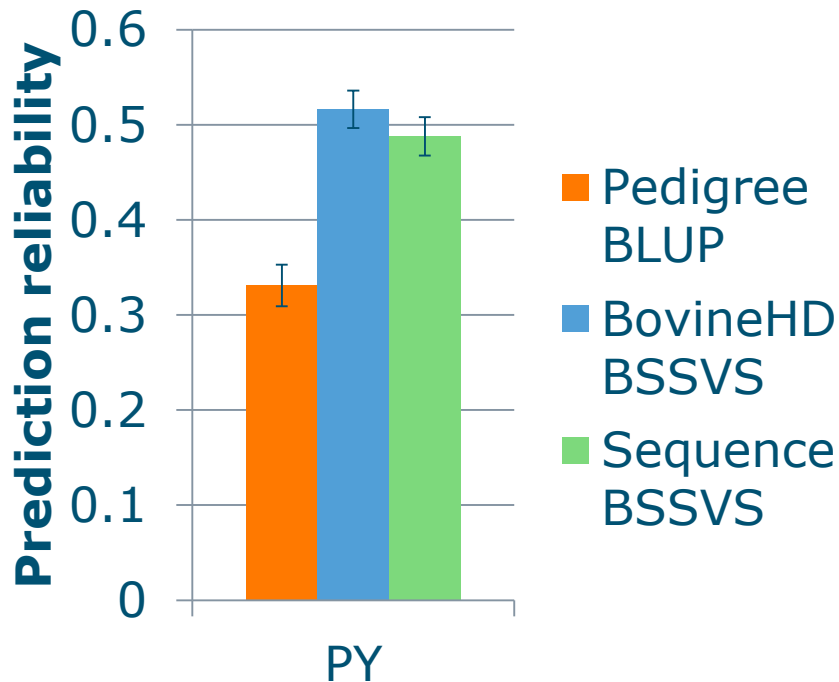


# Background

- Whole genome sequence data expected to perform better in GWAS and GS, ore persistent across generations / breeds
  - Causal mutation (QTN) is included
  - No dependency on LD between SNP and QTL

# Identifying QTN with GS?

HD SNP:  
631,428



SEQ SNP:  
12,590,056

# Objective

- The potential benefit of using sequence data for
  - QTL detection
- i.r.t.
  - Explaining total genetic variation
  - Prediction accuracy GS

# Method (1): Imputation to sequence

5556 Holstein Friesian bulls CRV

777K SNP genotypes (Illumina BovineHD BeadChip)

1147 animals (multiple breeds)

28M SNP (whole-genome sequence data)

1000 bull genomes project Run 4

Minimac  
v2

5556 Holstein Friesian bulls with phenotype (PY) and imputed sequence

3469 bulls used for discovery and training & 2287 bulls used for validation

# Method (2): statistical models

- GWAS using single SNP regression (GCTA)
  - GWAS using single SNP regression
  - Include GRM based on HD SNP set
  - MAF >0.01
- **Conditional and joint GWAS (COJO)**
  - Stepwise selection of SNP explaining additional variance

nature  
genetics

Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits

Jian Yang<sup>1,2</sup>, Teresa Ferreira<sup>3</sup>, Andrew P Morris<sup>3</sup>, Sarah E Medland<sup>1</sup>, Genetic Investigation of Anthropometric Traits (GIANT) Consortium<sup>4</sup>, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium<sup>4</sup>, Pamela A F Madden<sup>5</sup>, Andrew C Heath<sup>5</sup>, Nicholas G Martin<sup>1</sup>, Grant W Montgomery<sup>1</sup>, Michael N Weedon<sup>6</sup>, Ruth J Loos<sup>7</sup>, Timothy M Frayling<sup>6</sup>, Mark I McCarthy<sup>3,8</sup>, Joel N Hirschhorn<sup>9-13</sup>, Michael E Goddard<sup>14,15</sup> & Peter M Visscher<sup>1,2,16</sup>

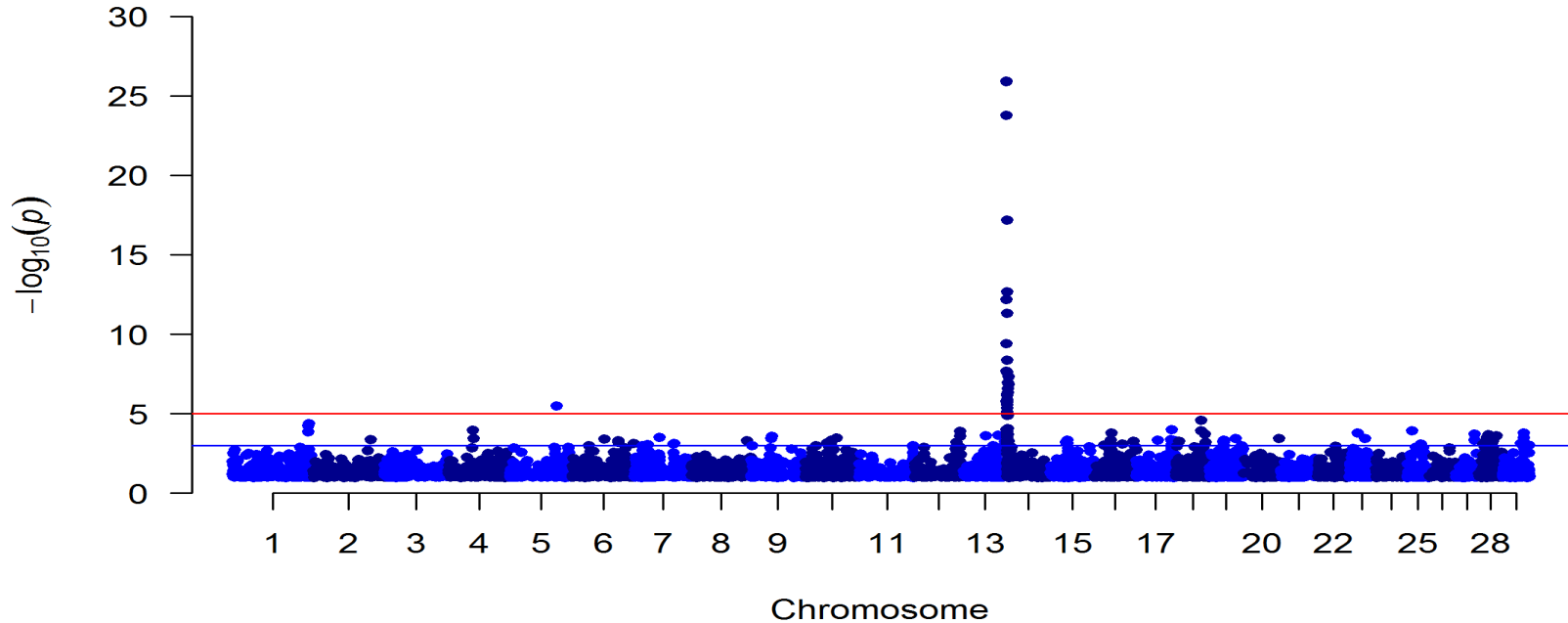


# Method (3): Two validation methods

SNP set selection from GWAS:

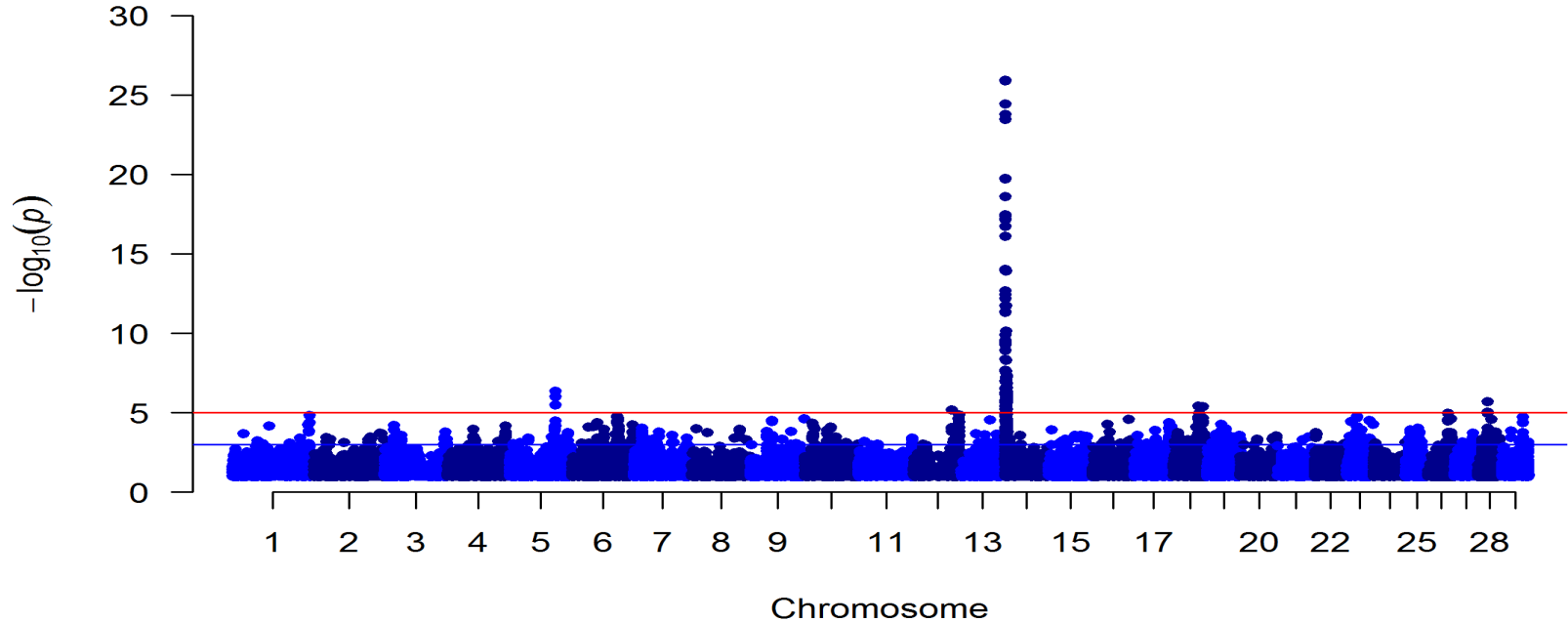
1. Estimate heritability in validation animals using GRM based on selected sets of SNP
2. Train GRM (based on selected SNP sets) on discovery animals, back solve SNP and predict DGV for 2287 validation animals. Correlate DGV with phenotypes.

# Results GWAS: 50K

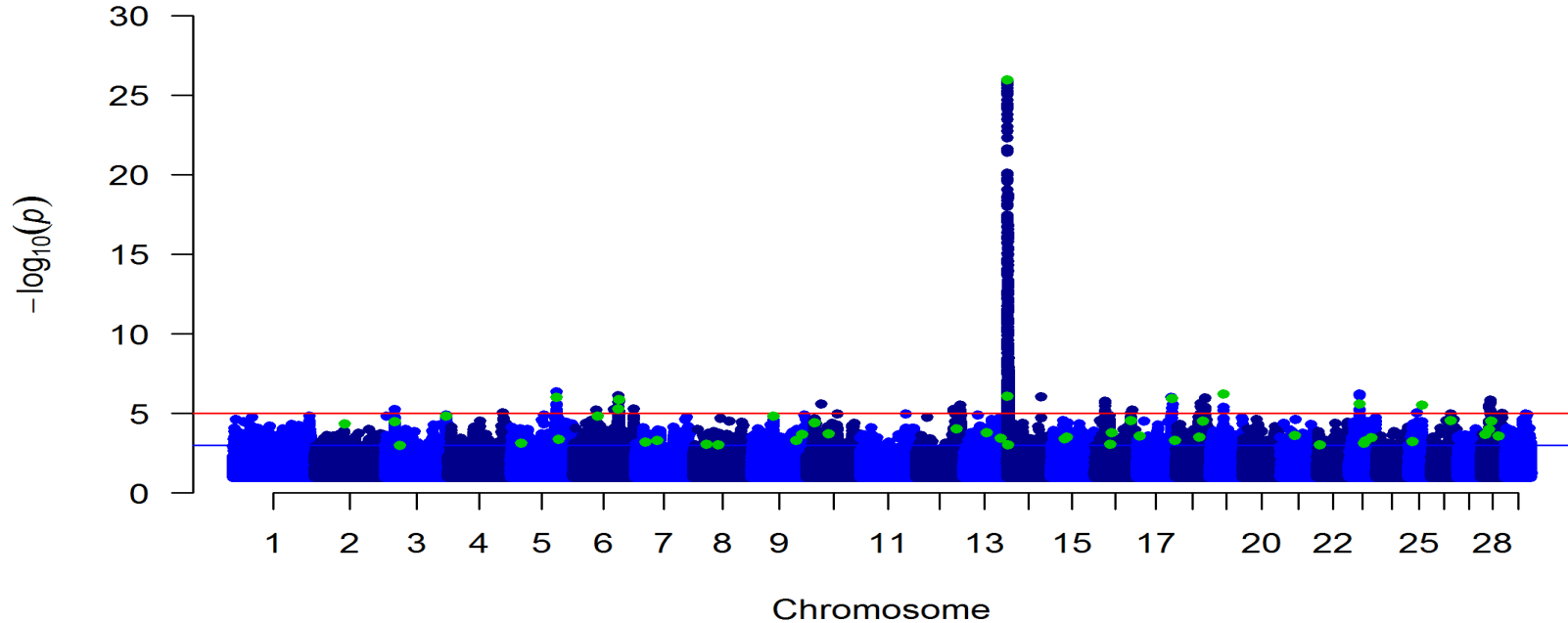




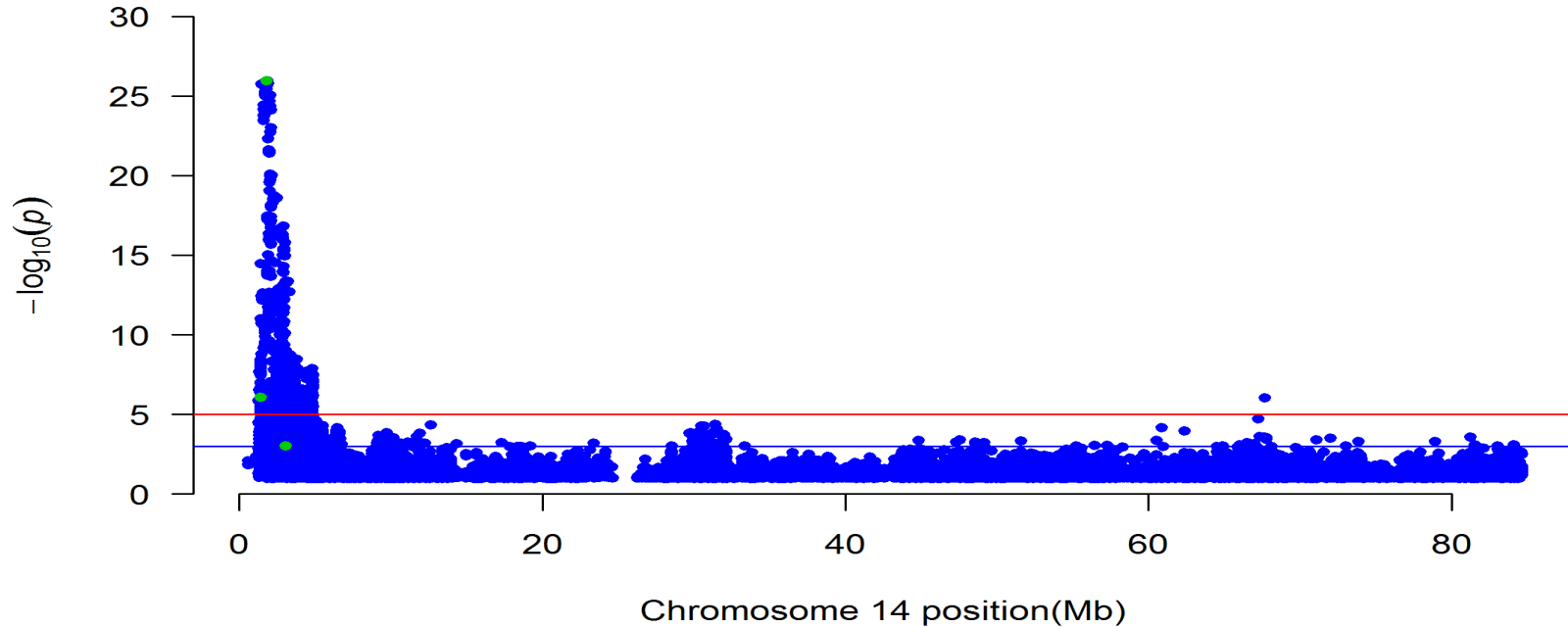
# Results GWAS : HD



# Results GWAS Sequence + cojo5



# Results GWAS: Cojo5 on Chr14 (DGAT)



# Results: SNP set selection GWAS

Number of SNP selected in the different SNP sets:

	Sequence	HD	50k	COJO
All	13,789,029	656,044	49,580	
$-\log(p) > 3$	24,387	1,238	120	119
$-\log(p) > 5$	2,194	159	27	49

# Results: Heritability

$h^2$ , phenotypic variance explained by GRM using selected SNP sets

GRMs	Sequence	HD	50k	COJO
All	0.83	0.82	0.81	
$-\log(p) > 3$	0.53	0.40	0.22	0.24
$-\log(p) > 5$	<del>0.60</del> *	<del>0.43</del> *	<del>0.22</del> *	0.16

\* Scale problems with GRM when estimating variances

# Results: Genomic prediction

Correlation between genomic breeding value and phenotype

GRMs	Sequence	HD	50k	COJO
All	0.68	0.68	0.68	
$-\log(p) > 3$	0.58	0.56	0.42	0.38
$-\log(p) > 5$	0.39	0.30	0.28	0.31

# Conclusions

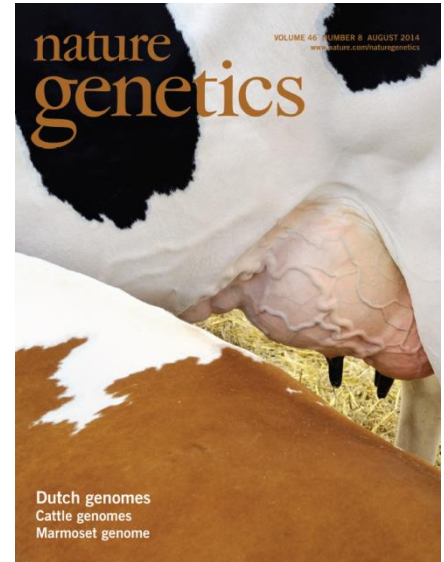
- Simply using sequence within Holstein population, unlikely to improve GS, but helps QTL detection.
- Subsets of selected SNP always poorer  $h^2$  and GS
  - Full seq. accuracy GS of 0.68 and  $h^2 = 0.83$
  - 51 SNPs accuracy GS of 0.31 and  $h^2 \approx 0.16$  (DGAT!)
- Good way to get realistic expectations from QTL.

# Acknowledgements



1000 bull genomes consortium

[www.1000bullgenomes.com](http://www.1000bullgenomes.com)



LIVESTOCK RESEARCH  
WAGENINGEN UR