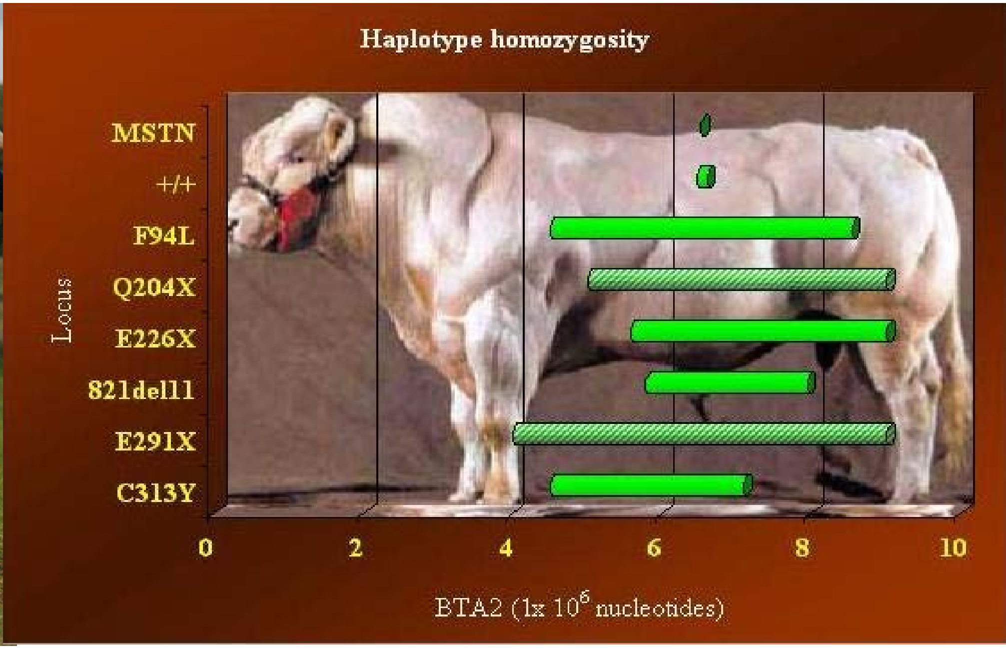
Across breed GEBVs



Michael Goddard

Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC3010, Australia.

Agriculture Victoria, Agriculture Research Division, AgriBio, Centre for AgriBioscience, Bundoora, VIC3083, Australia.

This talk

What are they?

Why do we want them?

Why are they hard to get?

Solutions

THE UNIVERSITY OF
MELBOURNE

AGRICULTURE VICTORIA

What are they?

Prediction equations that work for multiple breeds

Why do we want them?

Training populations within breed are too small

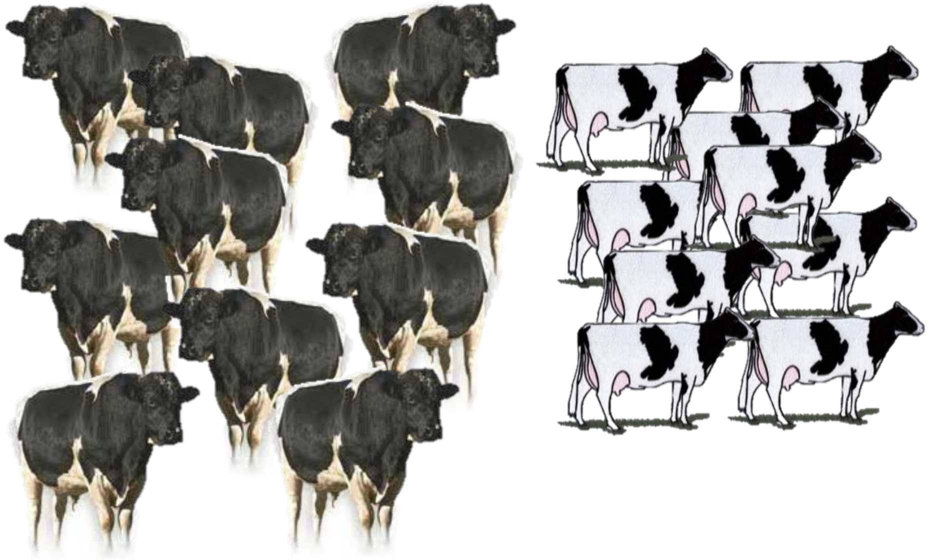
- numerically small breed
- Hard to measure traits eg FCE

Therefore, use multi-breed training population

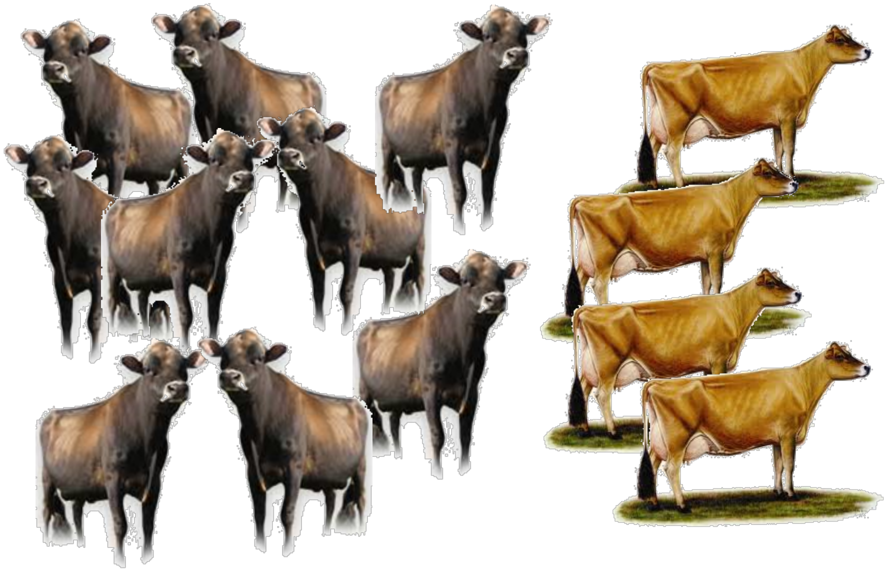
Aim = Accurate GEBVs for a breed with a small training population based on a multi-breed training population

Why are they hard?

Training in a different breed leads to low accuracy GEBVs



Holstein 4000 bulls, 10023 cows

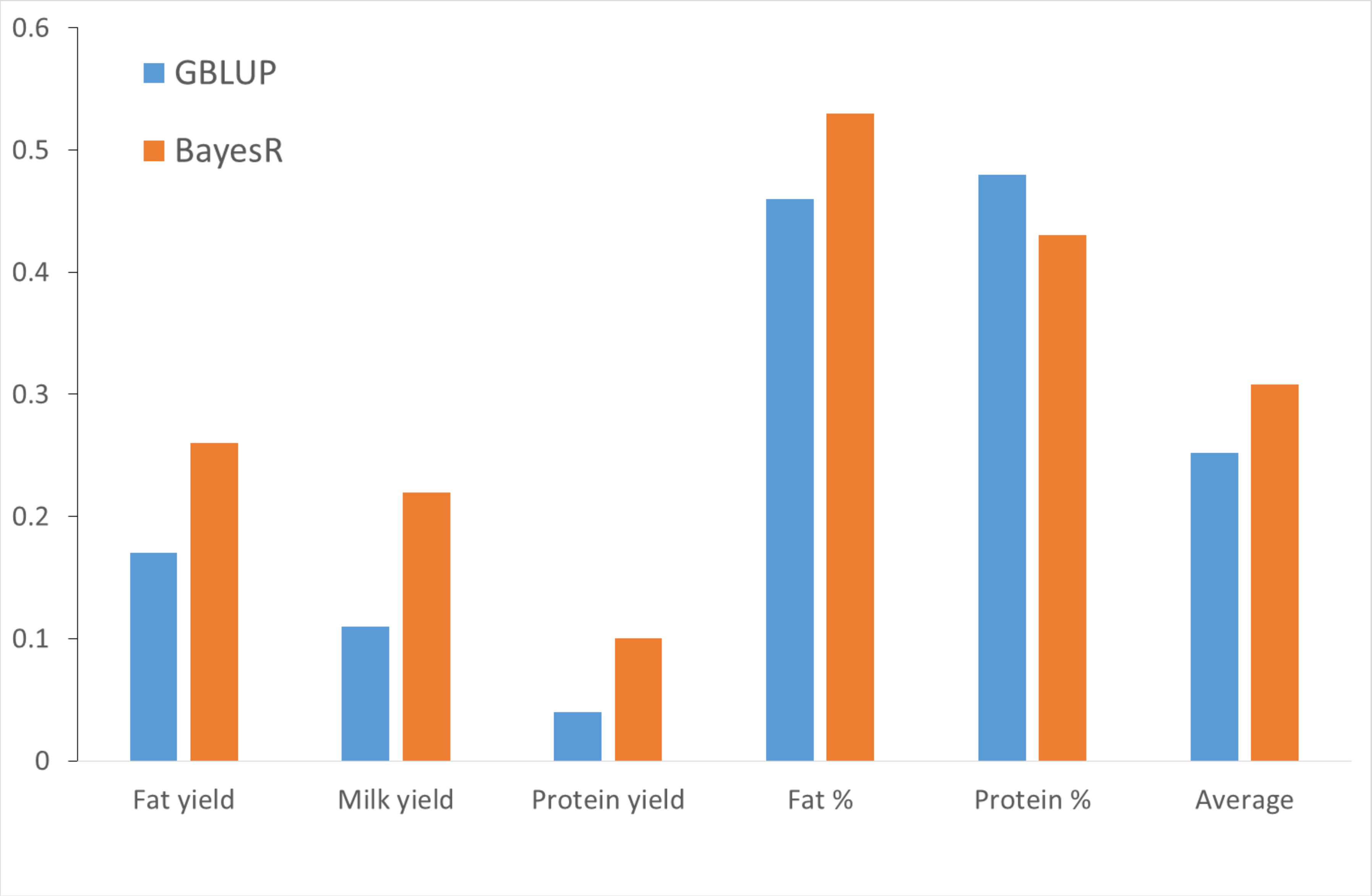


Jersey 1044 bulls, 4232 cows

Aussie Reds 114 Bulls

Real or imputed 630K SNP for all individuals

Accuracy r(DGV,DTD) in Aussie Red Bulls



Why are they hard?

SNP x breed interactions
  - differences in LD phase between breeds
  - QTL x breed interactions
    - Due to non-additive gene action
      - typically small variances
    - equivalent to sire x breed interactions
      - typically small
    - Low accuracy even in simulation

Differences in allele frequency
  - F_{ST} is low
  - QTL segregate across breeds

Why are they hard?

LD phase differs between breeds

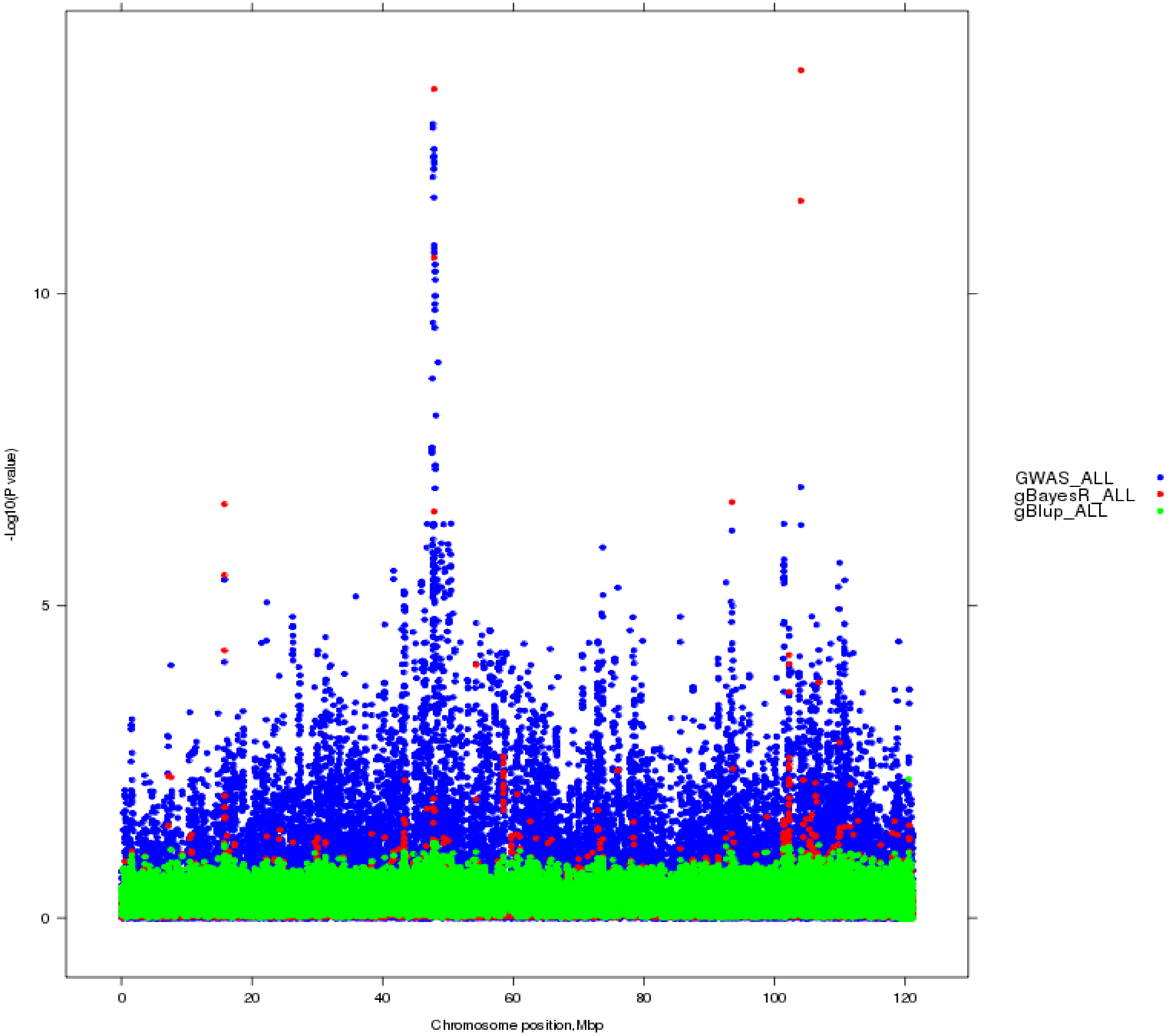
Within breed GEBVs estimate the effect of large chromosome segments

This works due to LD within a breed

Effective number of chromosome segments = 5000

That is, segments 600 kb long

PW_lwt_chr 5


-Log10(P value)
10
5
0
0
20
40
60
80
100
120
Chromosome position,Mbp
GWAS_ALL ●
gBayesR_ALL ●
gBlup_ALL ●

Why are they hard?

Within breed GEBVs estimate the effect of large chromosome segments

This works due to LD within a breed

Effective number of chromosome segments = 5000

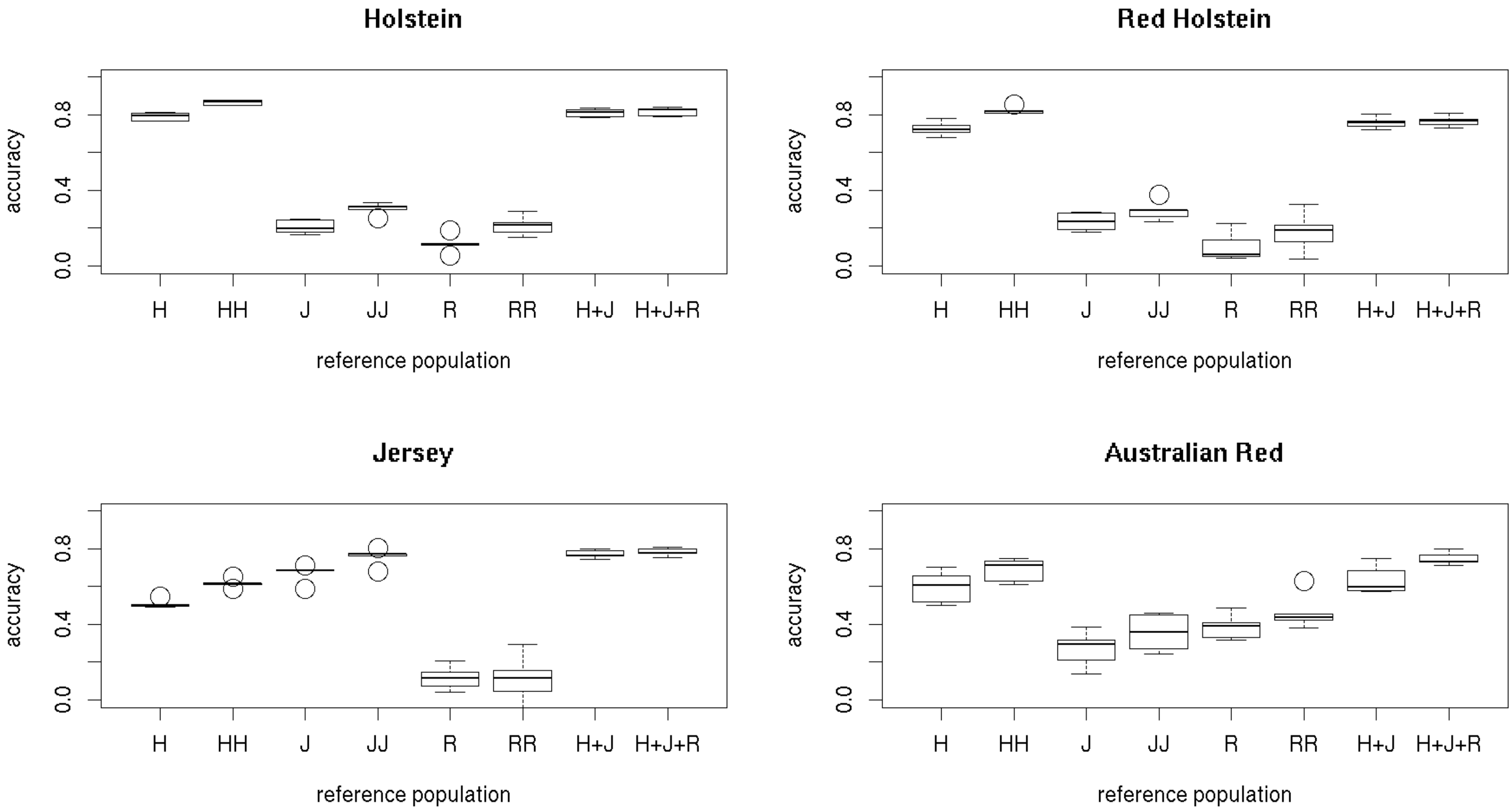
That is, segments 600 kb long

Across breeds conserved segments are much smaller (x10 smaller)

Solutions

Include target breed in training population

Accuracy of Bayes R (Irene van den Berg)

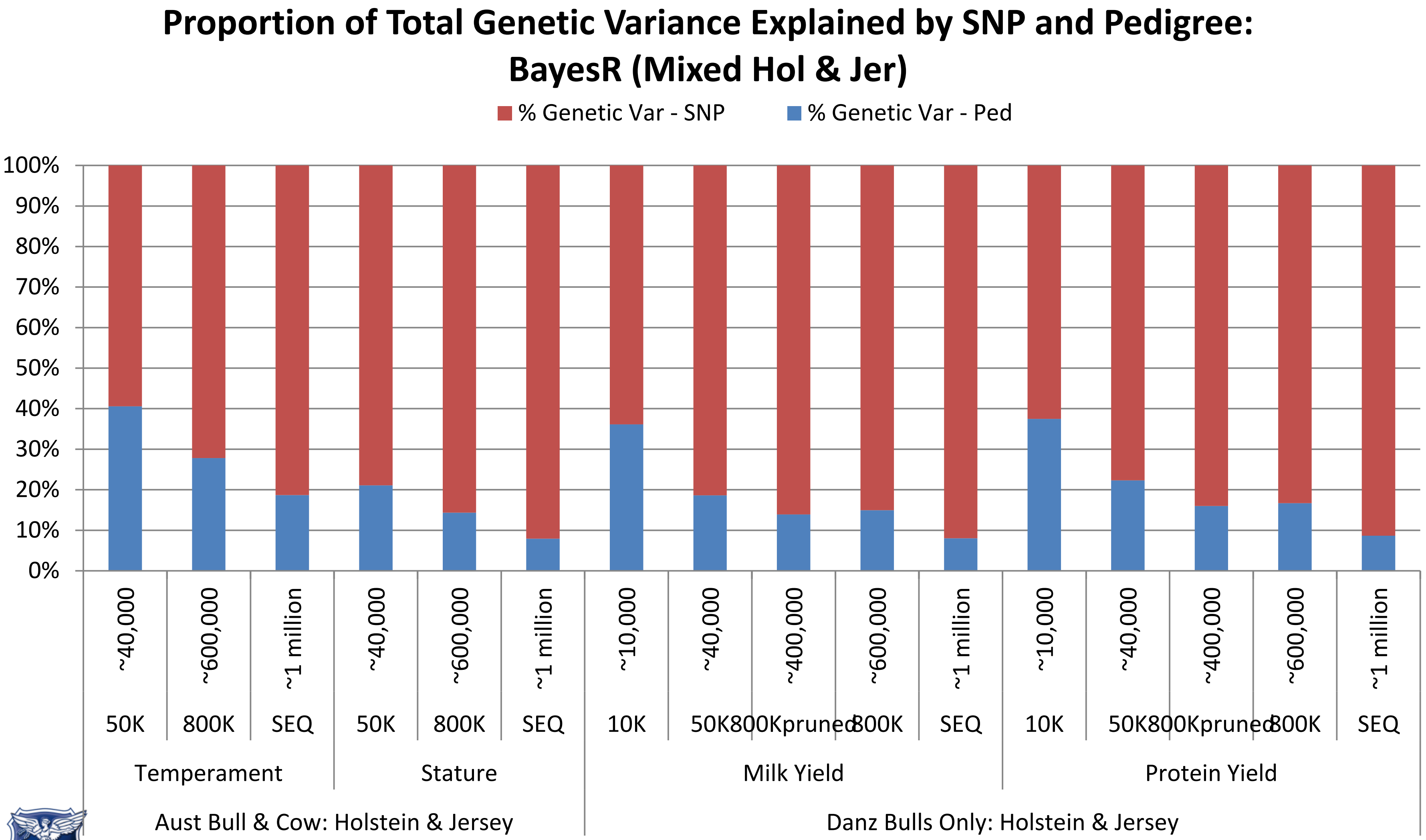


Solutions

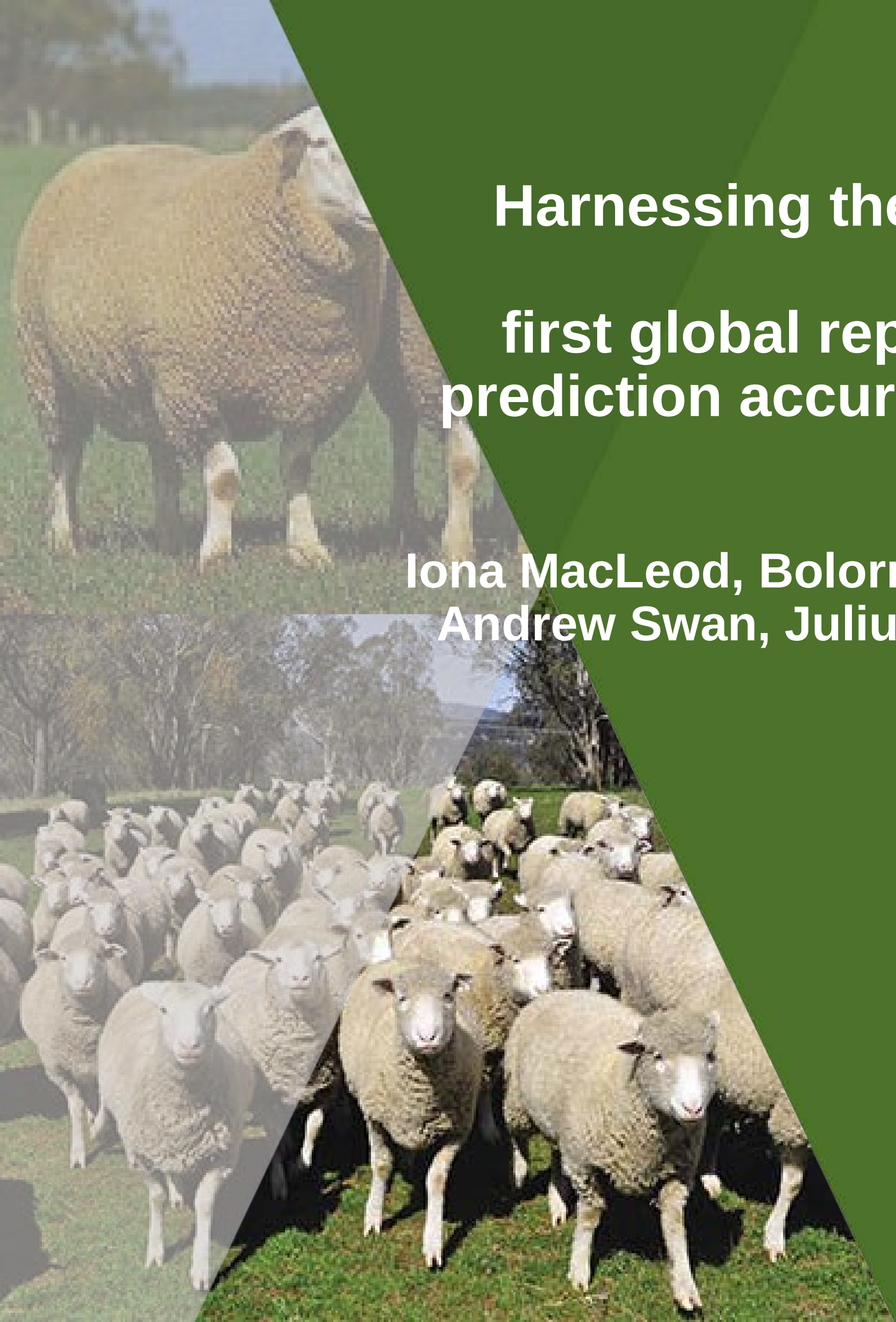
Include target breed in training population

Use denser SNP panels or sequence

Variance explained by SNPs and sequence (Iona Macleod)



Proportion of Total Genetic Variance Explained by SNP and Pedigree: BayesR (Mixed Hol & Jer)



Economic Development,
Jobs, Transport
and Resources

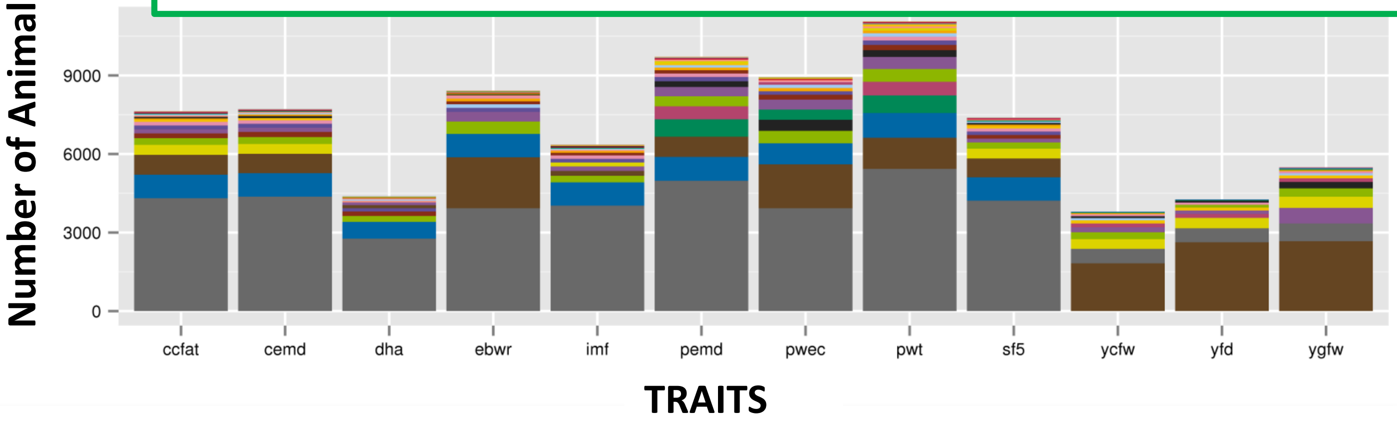
Harnessing the power of whole-genome sequence: first global report of improved genomic prediction accuracy using sequence data in sheep

Iona MacLeod, Bolormaa Sunduimijid, Majid Khansefid, Andrew Swan, Julius van der Werf & Hans Daetwyler

SHEEP CRC

AGRICULTURE VICTORIA

Composition of animals in Genomic Prediction Reference Set



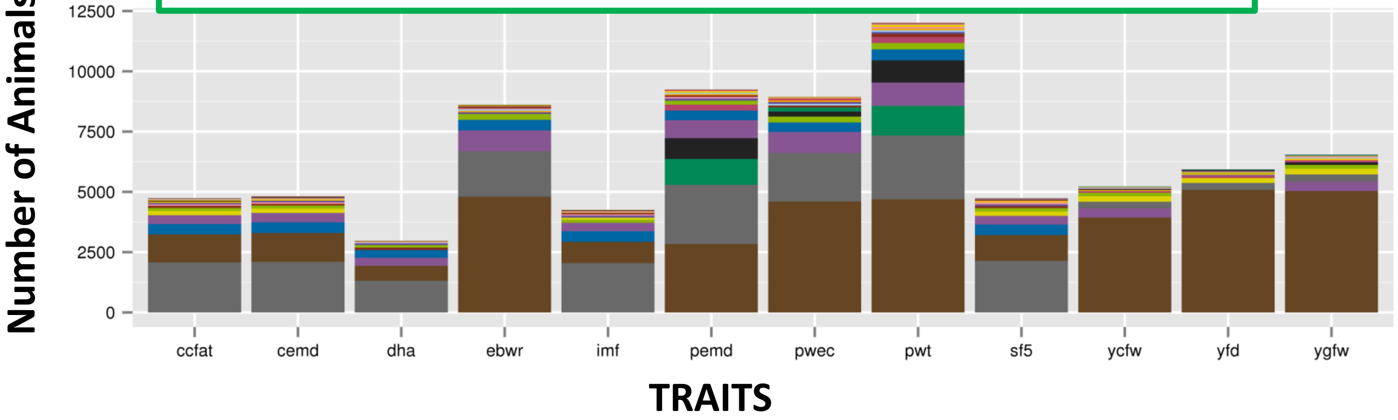
Breed

- ■ BorderLeicester_Composite_crossbred
- ■ BorderLeicester_Merino_crossbred
- ■ BorderLeicester_crossbred
- ■ BorderLeicester_pure
- ■ Composite_crossbred
- ■ Coopworth_pureANDcrossbred
- ■ Corriedale_pureANDcrossbred
- ■ DorperANDWhiteDorper_pureANDcrossbred
- ■ MaternalBreeds_pureANDcrossbred
- ■ MeatBreeds_pureANDcrossbred
- ■ Merino_pure
- ■ PollDorset_Merino_crossbred
- ■ PollDorset_pure
- ■ PollDorset_pureANDcrossbred
- ■ Research_pureANDcrossbred
- ■ SammANDDohneMerino_pureANDcrossbred
- ■ Southdown_pureANDcrossbred
- ■ Suffolk_pureANDcrossbred
- ■ Texel_pureANDcrossbred
- ■ WhiteSuffolk_pureANDcrossbred

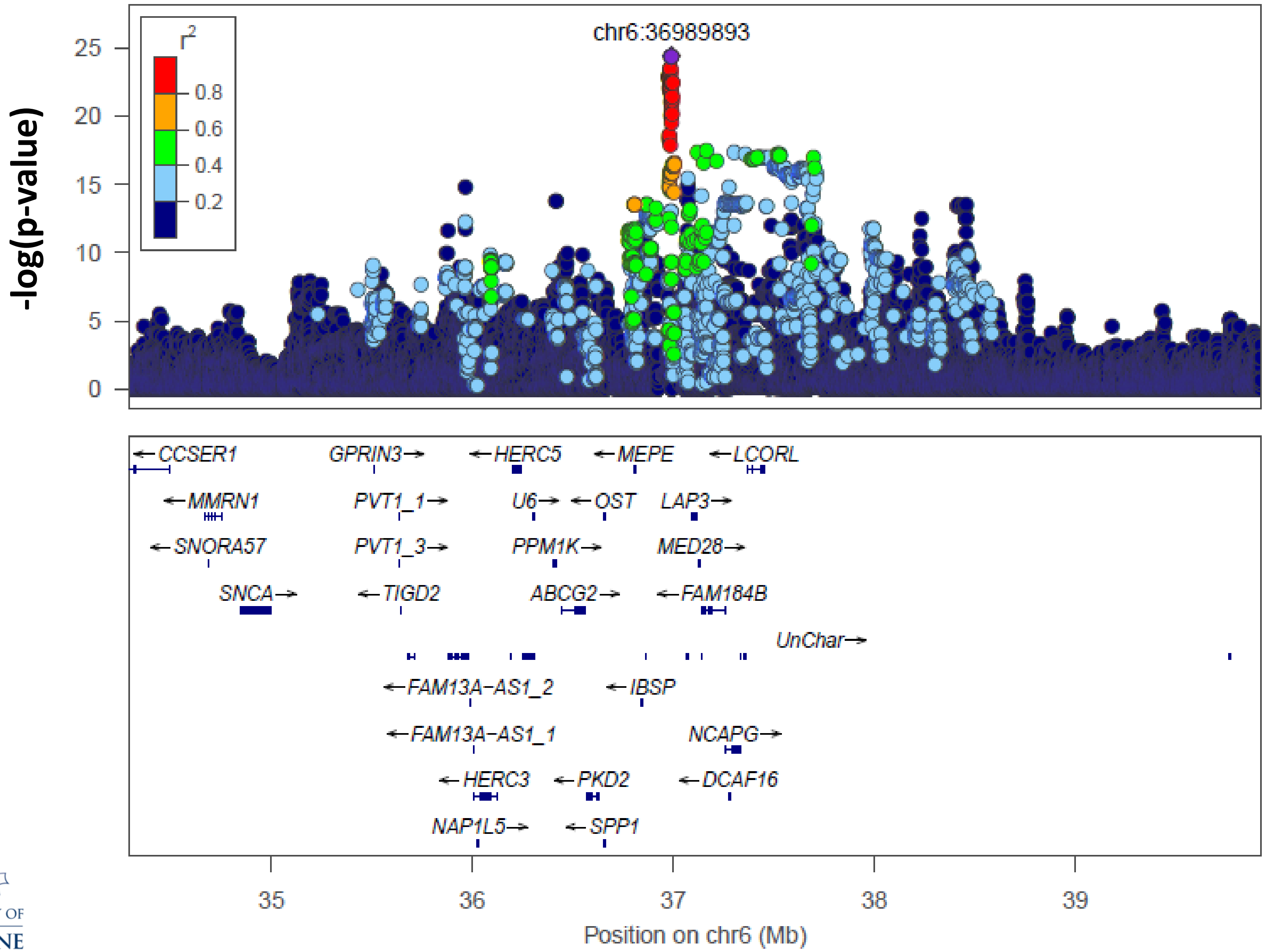
Validation sets - low relationships with Ref.:

1. **Merino**
2. **Merino x Border Leicester F1**

Composition of animals in Discovery GWAS Set



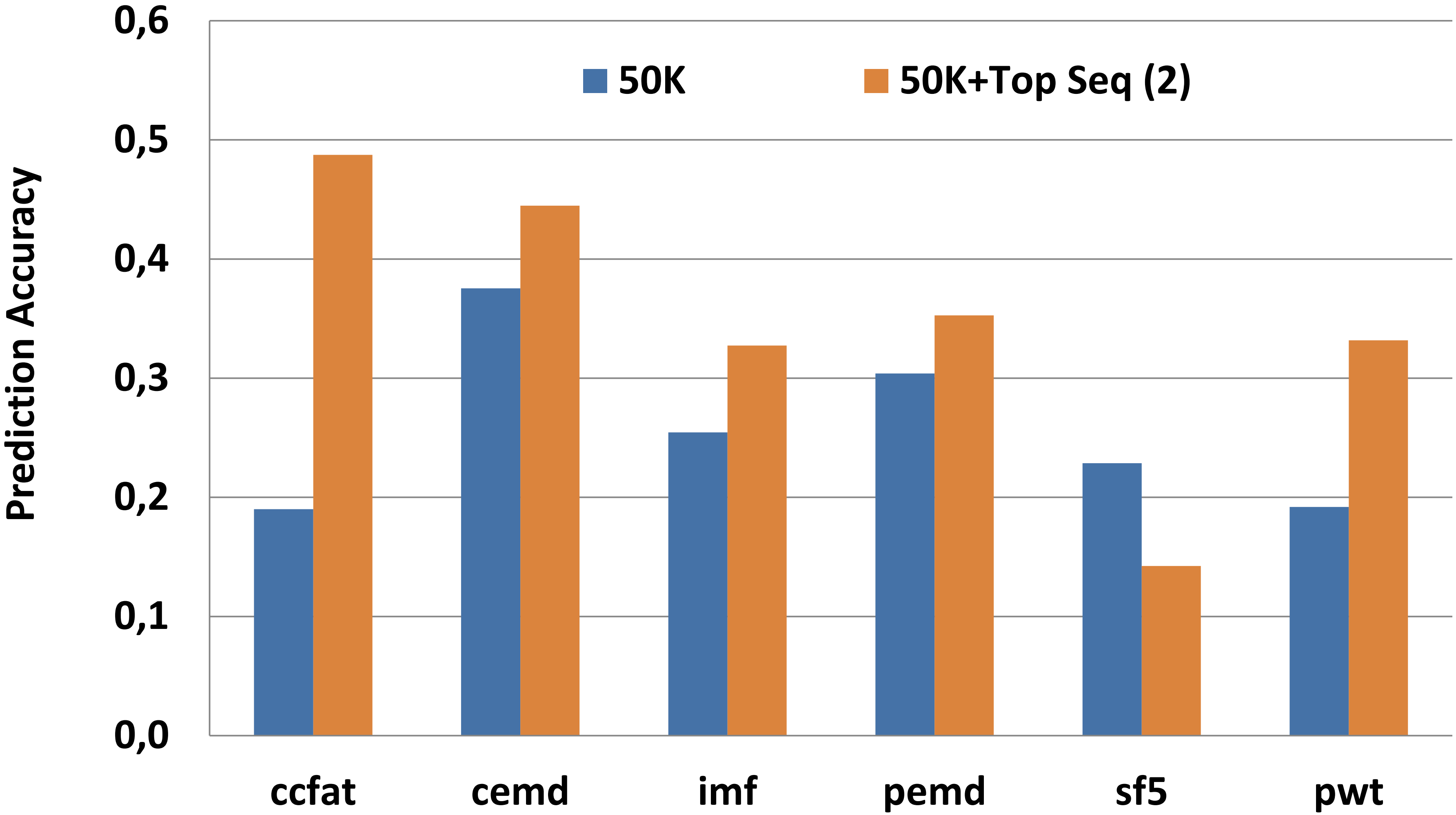
GWAS – Carcass Fat Depth (ccfat)



THE UNIVERSITY OF MELBOURNE

Meat Traits:

GBLUP Accuracy - Merino x Border Leicester



Solutions

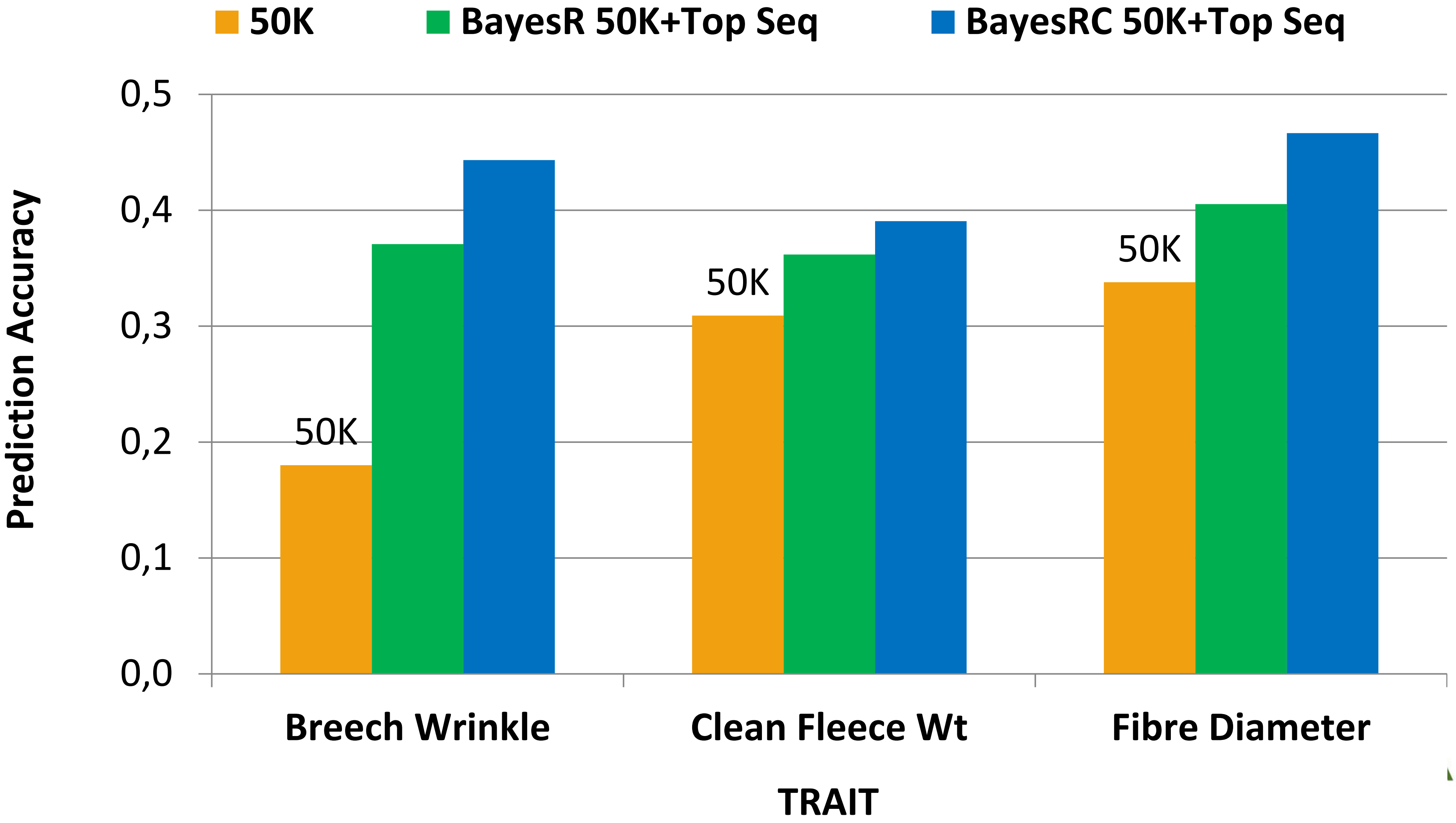
Include target breed in training population

Use denser SNP panels or sequence

Use Bayesian statistical method not GBLUP

Wool Traits:

Prediction Accuracy in Merinos



Solutions

Include target breed in training population

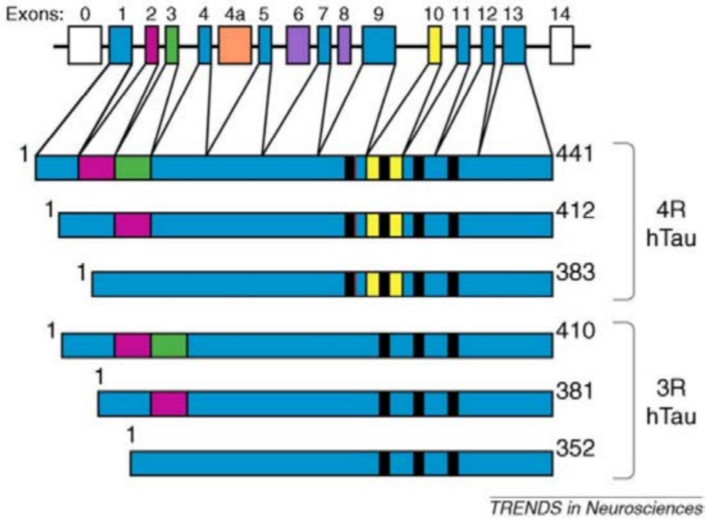
Use denser SNP panels or sequence

Use Bayesian statistical method not GBLUP

Multi-trait analysis e.g. gene expression data
- gene cis eQTL
- splicing cis eQTL
- exon cis eQTL

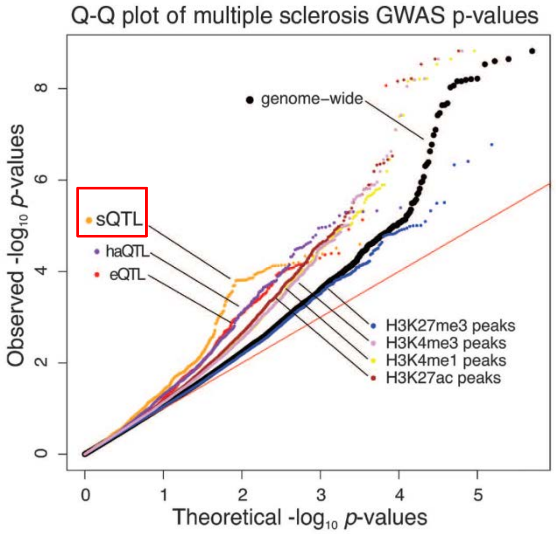
Phenotypic differences due to splicing

- Human Tau gene splicing related to the Alzheimer's disease



TRENDS in Neurosciences

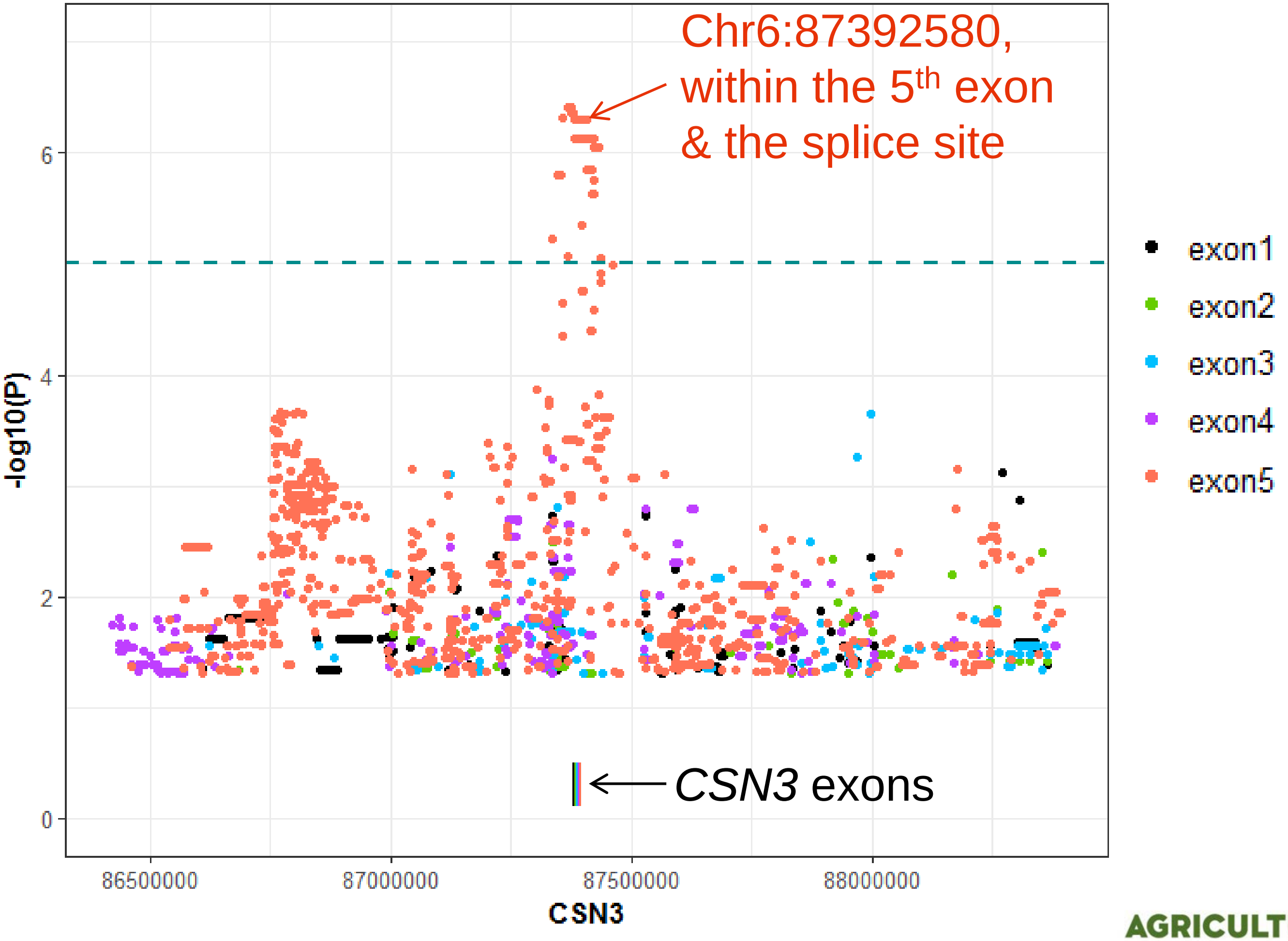
- Many genome variants affecting gene splicing, sQTL contribute to human diseases



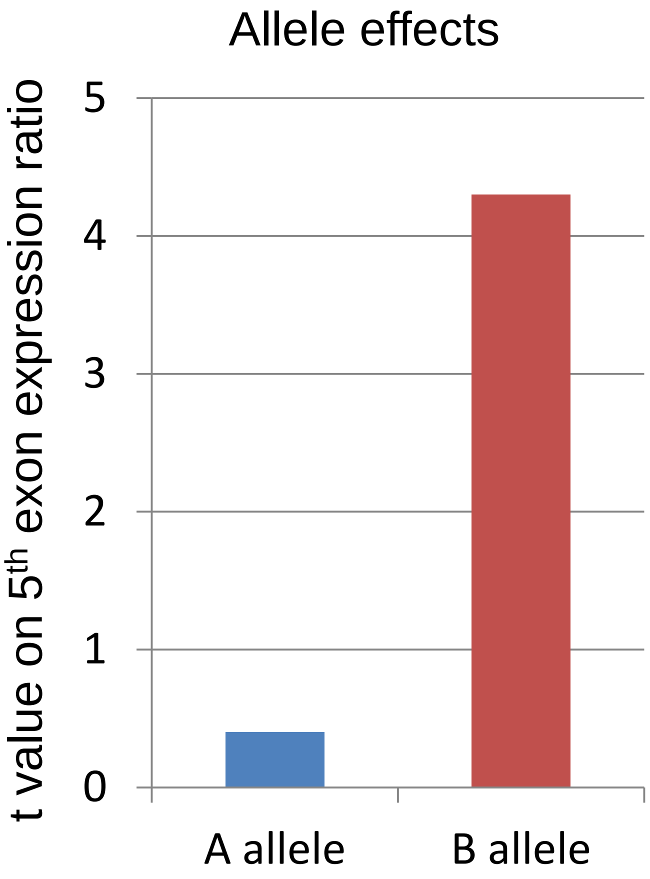
Li et al., 2016

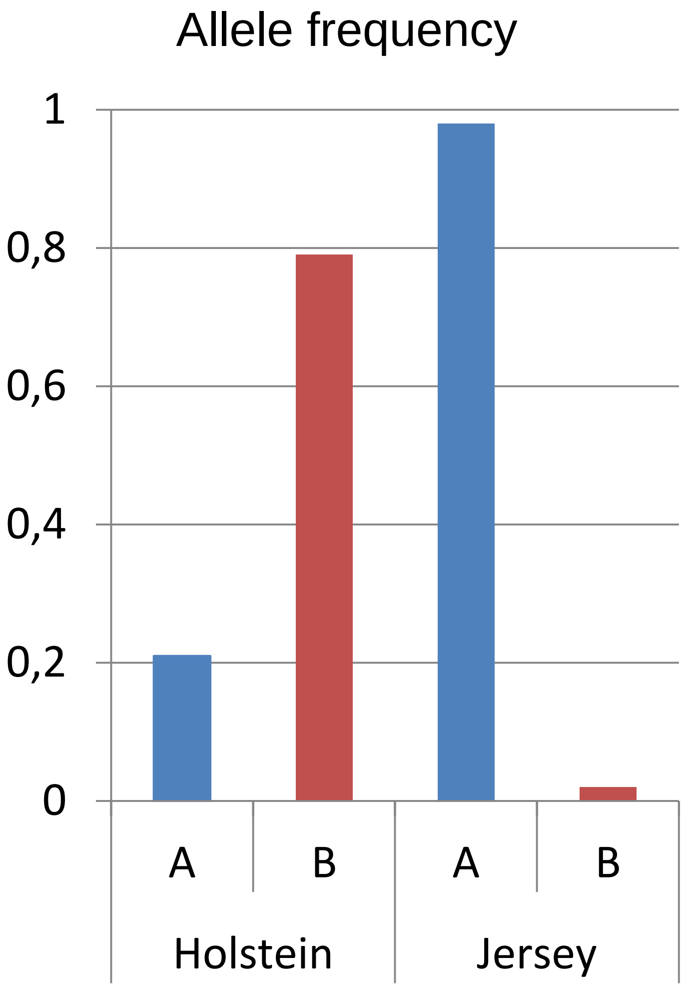
sQTN in *CSN3* region (Ruindong Xiang)

- SNPs affecting inclusion of the 5th exon in the milk cell transcriptome are most significant



Details of sQTN Chr6:87392580





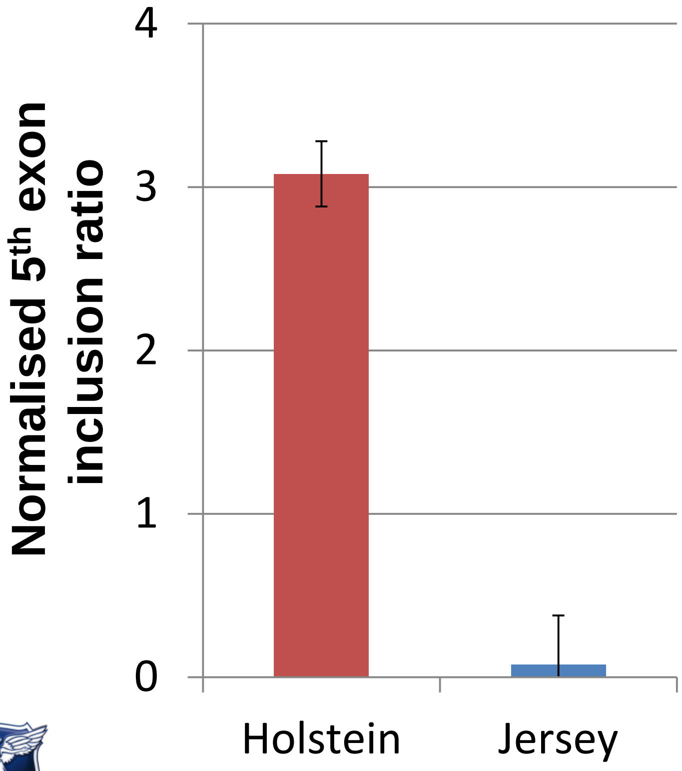
THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA

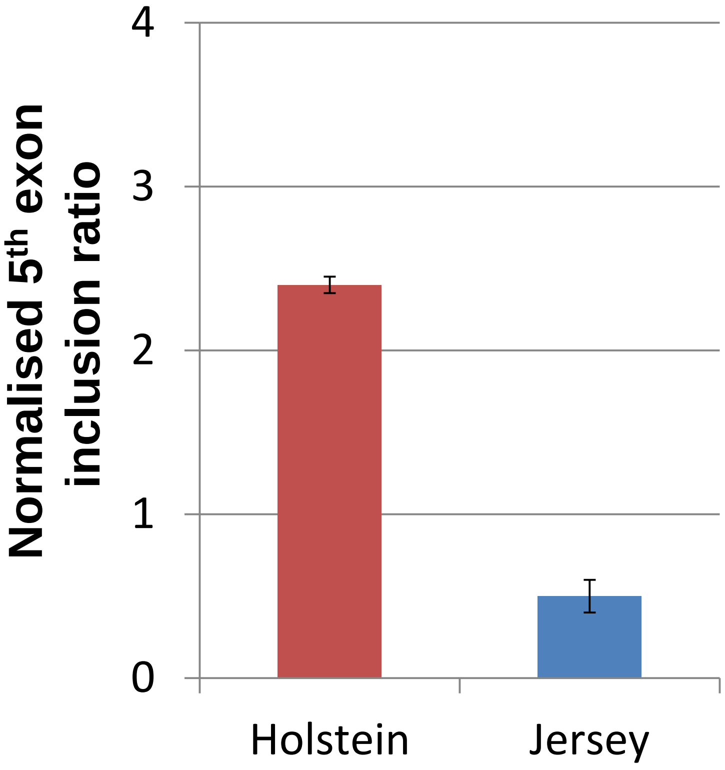
The splicing of 5th exon of *CSN3*

The agreement between the effects of sQTN Chr6:87392580 and the breed splicing of the 5th exon of *CSN3*

Predicted using allele effects



Observed

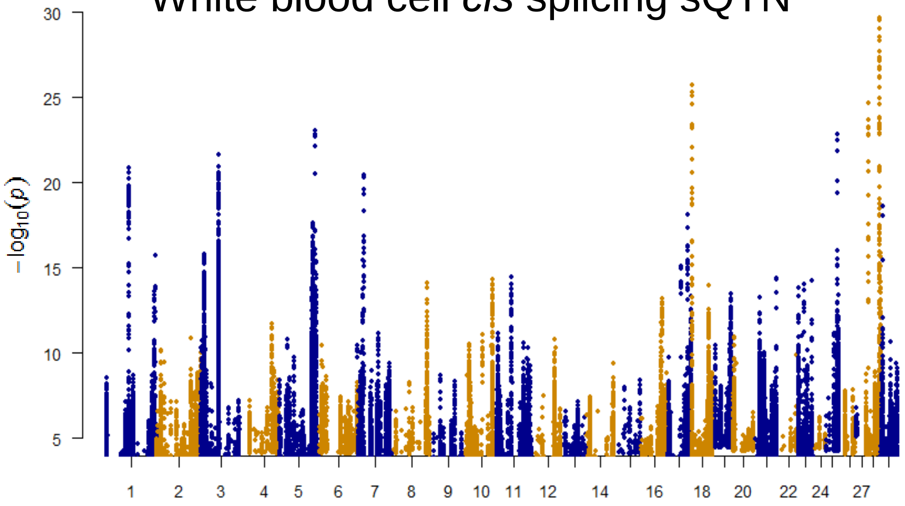


THE UNIVERSITY OF MELBOURNE

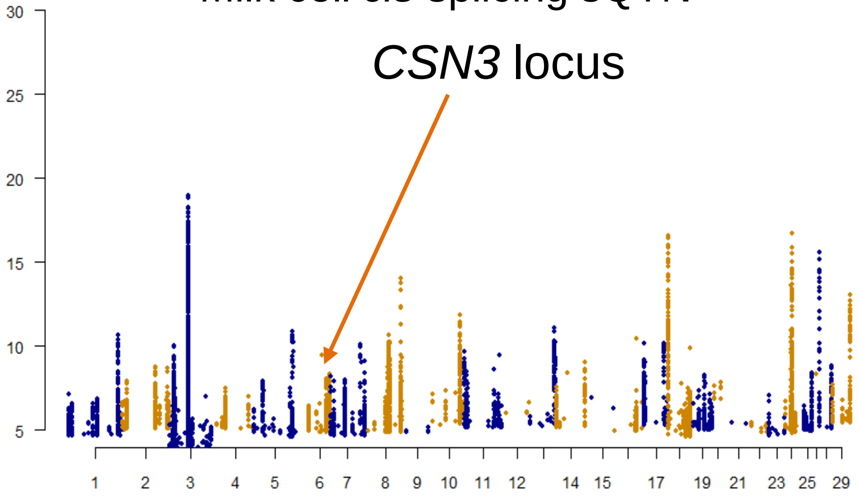
AGRICULTURE VICTORIA

Splicing between genotypes (sQTN)

White blood cell *cis* splicing sQTN



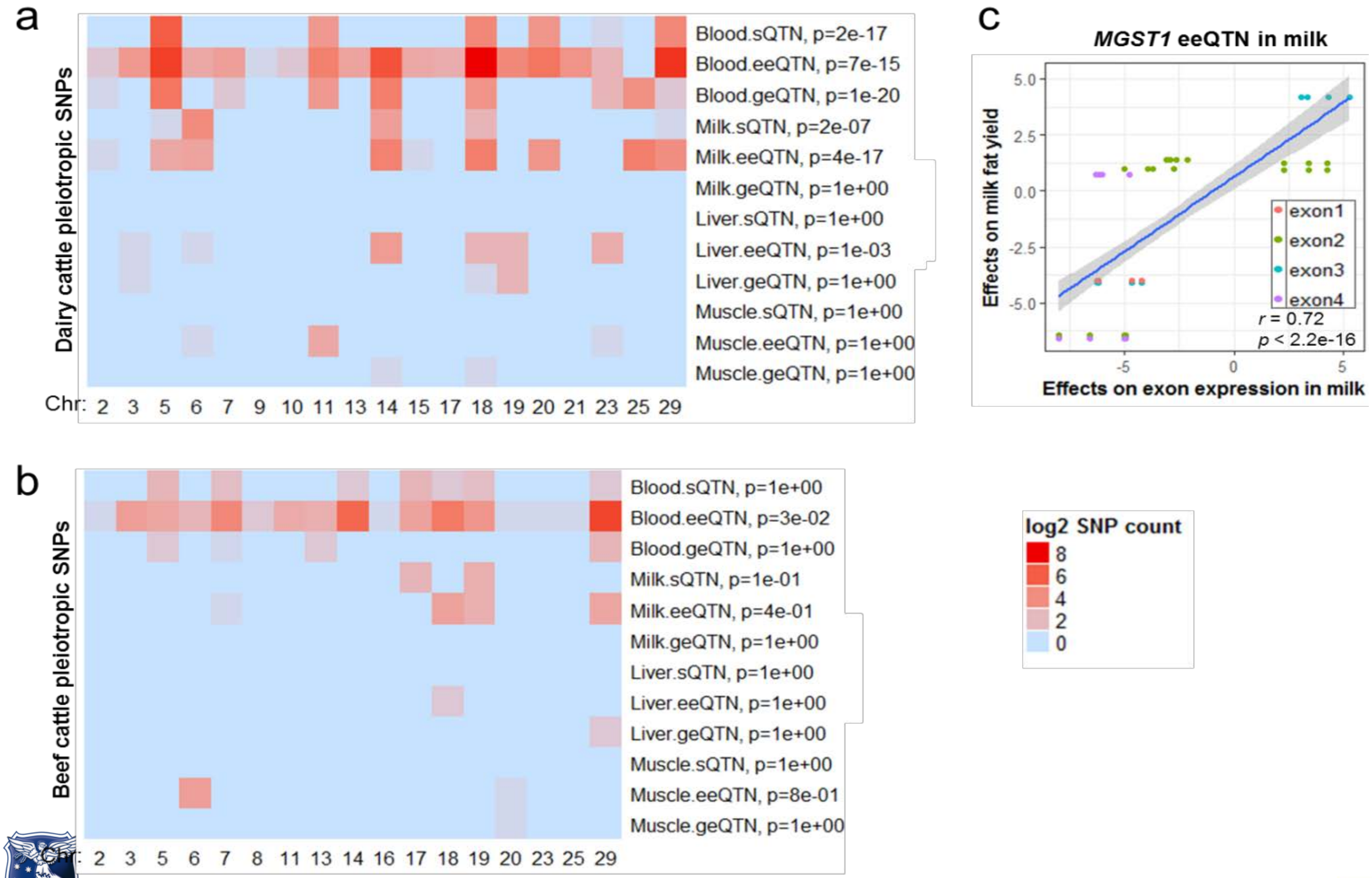
Milk cell *cis* splicing sQTN


CSN3 locus

Tissue	Individuals no.	RNA seq depth	sQTN no.	Gene no.
white blood cells	105	150 PE	138,796	929
milk cells	131	150 PE	28,907	283
Liver	35	100 PE	11,544	49
Muscle	41	100 PE	5,783	76

- Sample size and sequence precision are associated with the amount of sQTN detected

Overlap between eQTL and milk QTL



Solutions

Include target breed in training population

Use denser SNP panels or sequence

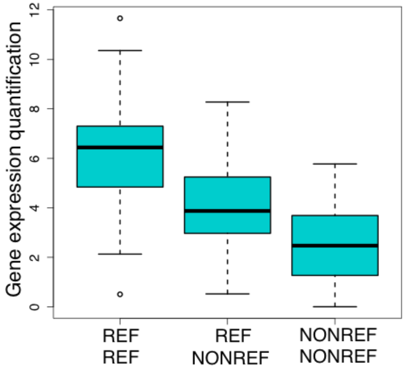
Use Bayesian statistical method not GBLUP

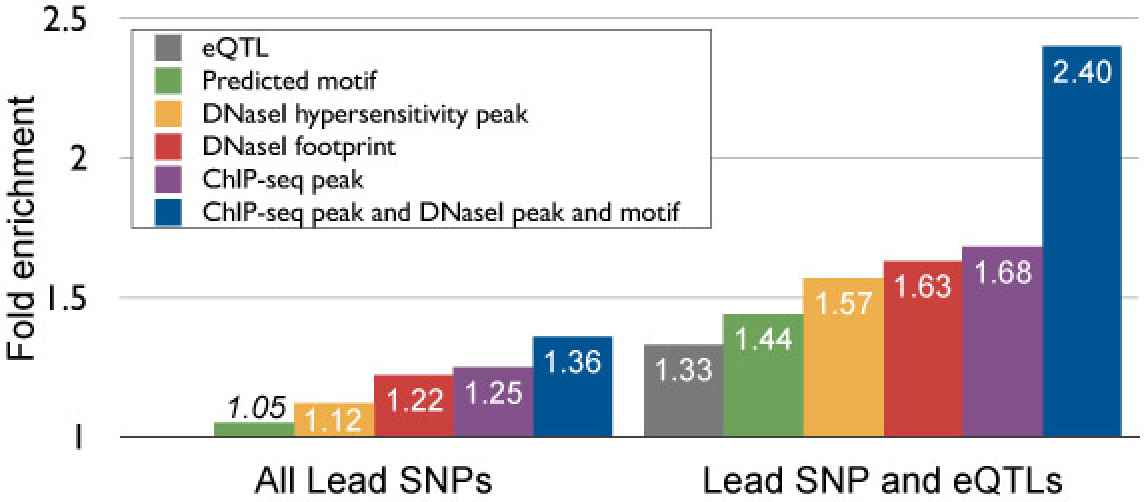
Multi-trait analysis e.g. gene expression data

Use functional annotation of genome

SNP effects at cellular level

- Quantify the impact of a mutation on gene expression levels





Schaub et al 2012

THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA

The bad news

Accuracy only improves a little

You need to capture a high proportion of total variance

Conclusion

Data from the target breed is the most useful

But, training data from other breeds helps

Advantage to use sequence data and Bayesian method

Sequence imputation loses accuracy

Identify near perfect markers and genotype them directly

Expression data and functional annotation helps select best variants

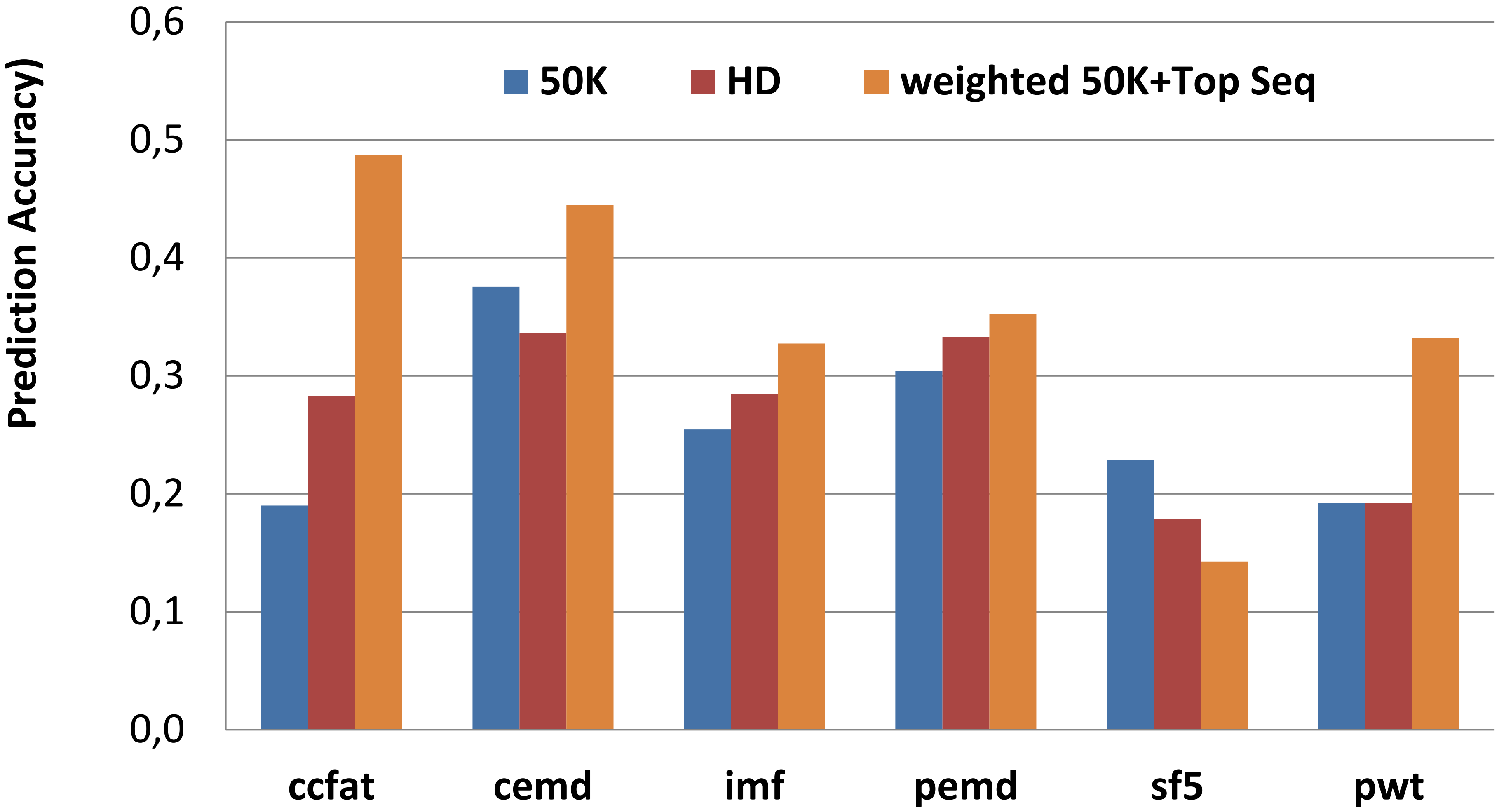
Genomic Prediction

Trait (unit)	Discovery	Reference	MER Validation	BL×MER Validation
CCFAT (mm)	4,452	7,635	912	536
CEMD (mm)	4,473	7,714	904	519
PEMD (mm)	7,114	9,715	1,766	586
IMF (%)	3,905	6,353	843	474
SF5 (Newtons)	4,344	7,392	868	531
PWT (Kg)	8,937	11,067	3,118	543

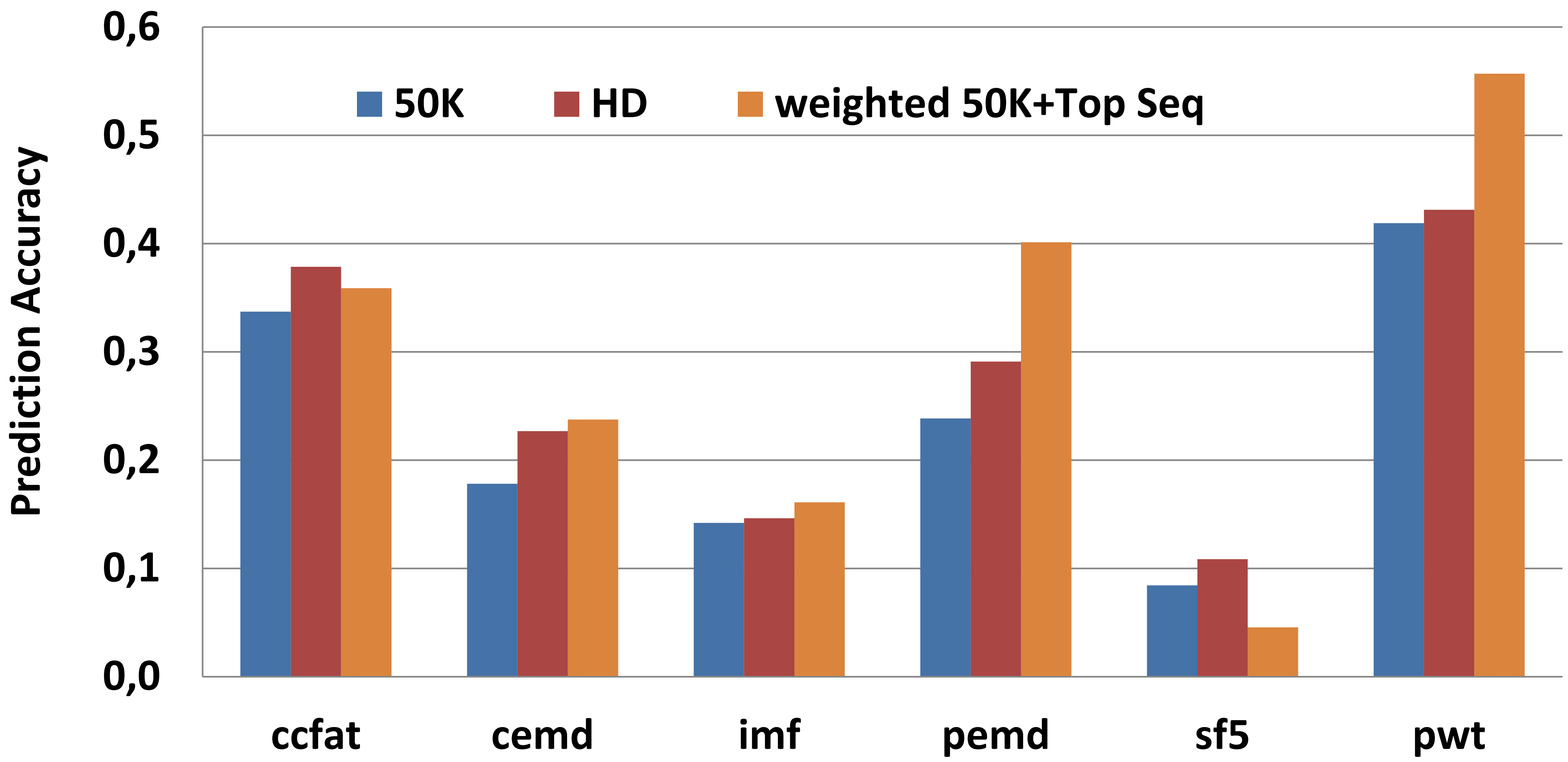
THE UNIVERSITY OF MELBOURNE

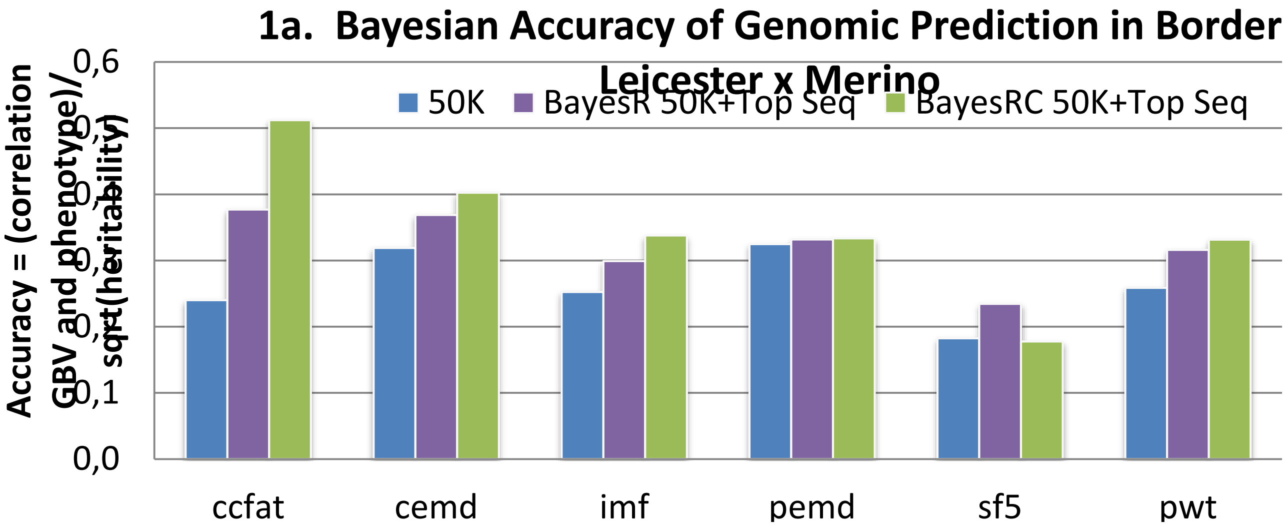
AGRICULTURE VICTORIA

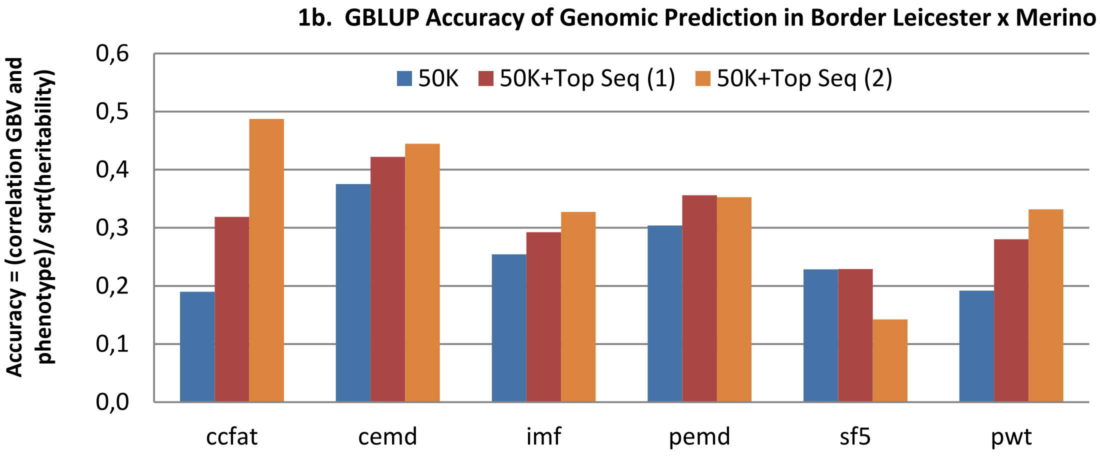
Accuracy of Prediction: Border Leicester x Merino



Accuracy of Prediction: Merino

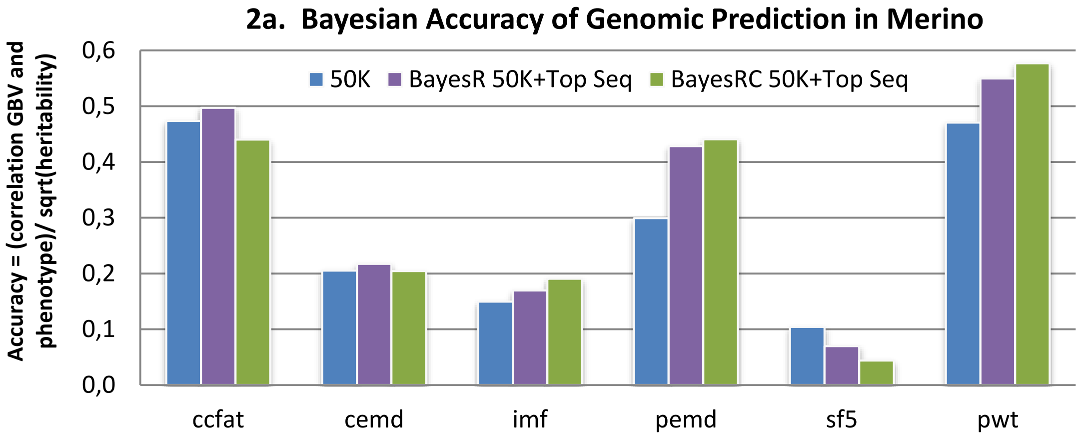



1a. Bayesian Accuracy of Genomic Prediction in Border Leicester x Merino

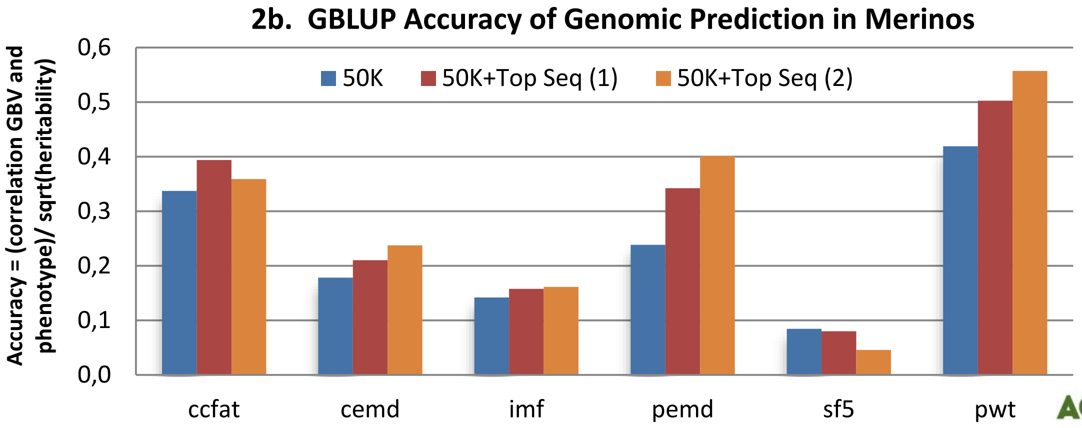

1b. GBLUP Accuracy of Genomic Prediction in Border Leicester x Merino

THE UNIVERSITY OF MELBOURNE

AGRICULTURE VICTORIA



2a. Bayesian Accuracy of Genomic Prediction in Merino

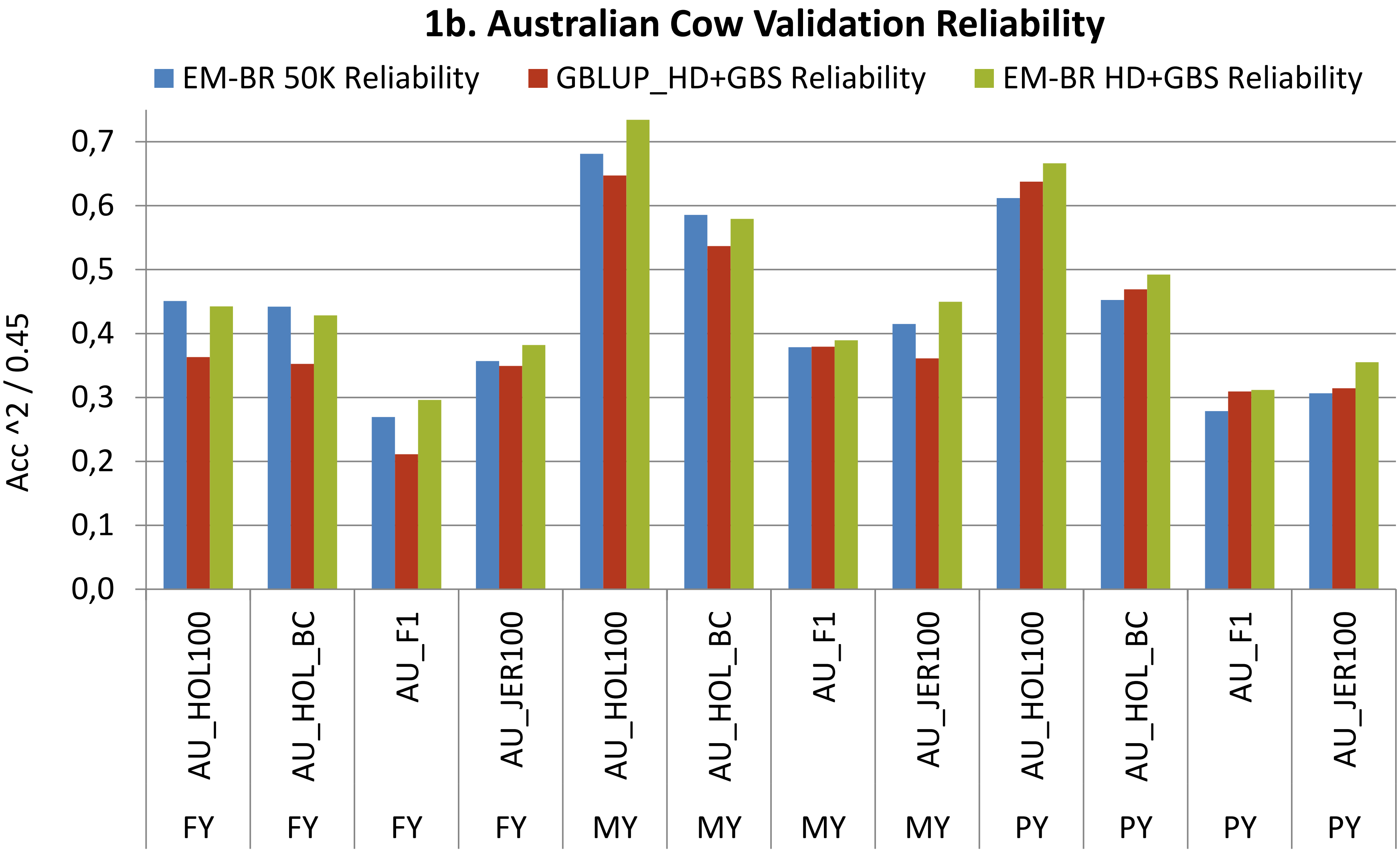


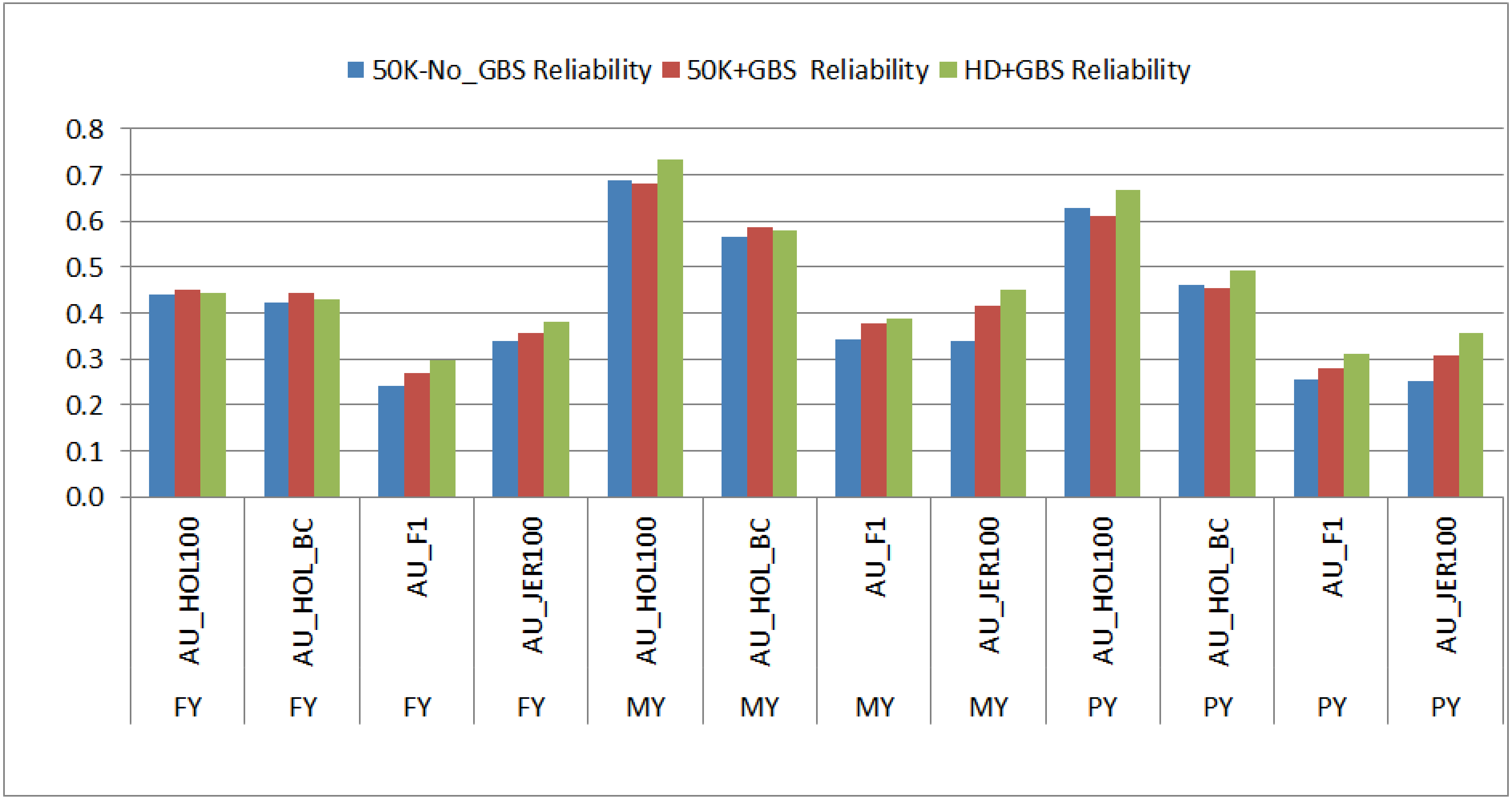
2b. GBLUP Accuracy of Genomic Prediction in Merinos

THE UNIVERSITY OF
MELBOURNE

AGRICULTURE VICTORIA

Dairy Cattle: Validation in Cows of different breed/cross (Reference was mixed bull & cow: inc. Hol, Jer





50K-No_GBS Reliability
50K+GBS Reliability
HD+GBS Reliability
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
AU_HOL100
AU_HOL_BC
AU_F1
AU_JER100
FY
MY
PY