Genomic evaluation based on selected variants from imputed whole-genome sequence data in Australian sheep populations

Nasir Moghaddar[1,2], I. MacLeod[1,3], N. Duijvesteijn[1,2], S. Bolormaa[1,3], M. Khansefid[1,3], H. Almamun[1,2], S. Clark[1,2], A. Swan[1,4], H. Daetwyler[1,3], J. van der Werf [1,2]

[1] *CRC for Sheep Industry Innovation*
[2] *School of Environmental and Rural Science, UNE, Armidale, NSW*
[3] *Bioscience Research, Agriculture Victoria, Bundoora, VIC 3083, Australia*
[4] *Animal Genetics and Breeding Unit (AGBU), Armidale, NSW 2351, Australia*

Background

- Genomic prediction accuracy for Australian sheep industry is moderate.

Background

- Genomic prediction accuracy for Australian sheep industry is moderate.

- Multi-breed & Crossbred Population.

- high *Ne,* Esp in Merino *(~850 J. Kijas 2011).*

- Smaller size of shared haplotypes.

Background

- Need stronger LD and denser marker genotypes, esp in a multi-breed population?

Background

- Need stronger LD and denser marker genotypes, esp in a multi-breed population?

- Results of using HD genotypes in sheep population *Moghaddar et al 2017*
  - slightly better prediction within breed (~2.3%)
  - no or very small improvement from across breed information.

Background

- Need stronger LD and denser marker genotypes, esp in a multi-breed population?

- Results of using HD genotypes in sheep population *Moghaddar et al 2017*
  - slightly better prediction within breed (~2.3%)
  - no or very small improvement from across breed information.

- Whole Genome Sequence data provides new opportunities.
  - It potentially covers the causal mutations
  - Marker variants in high LD with causal mutation

Background

Objective

To test whether selected variants from WGS data improve genomic prediction accuracy of Australian sheep populations

Methods

- **Phenotypes:**
  - Combined research and industry datasets
  - 6 traits on growth and eating quality traits (2008 to 2015 drops).

Trait Name	Trait	Reference set	Mean	SD
Post Weaning Weight	PWT (kg)	29,025	46.70	13.11
PW Eye Muscle Depth	PEMD (mm)	24,871	26.34	4.65
Carcass Eye Muscle Depth	CEMD (mm)	16,418	29.56	4.85
Carcass Fat	CFAT (mm)	16,284	4.04	2.27
Intra Muscular Fat	IMF (%)	13,518	4.35	1.14
Shear Force day 5 ageing	SF5 (Newtons)	15,494	25.3	14.05

Methods

- **Phenotypes:**
  - Pre-corrected for environmental and non-direct additive effect.
  - Phenotypes were divided into 3 non-overlapping data subsets:

1) GWAS subset	*4,300 to 4,900 animals Randomly selected*
2) Genomic prediction	*6,353 to 11,067*
3) 2 Validation subsets:	***Purebred Merino*** / ***Crossbred Merino (BLM)*** — *350 to 2036 Lowly related*

Methods

- **Genotypes:**

  - **50k genotypes**: 35,980 — *33% imputed from 12k*
  - **HD genotype**: 2,266 — *key animals*
  - **WGS**: 726 animals — *10x coverage*

  Sheep-CRC and Sheep Genome DB (Daetwyler et al., 2017)

- **Genotype Imputation:**

  - 50k → HD → WGS

  - *MiniMac Imputation R^2 > = 0.4,*
  -
  - *Final WGS set: 31,154,249 variants* *S. Bolormaa et al WCGALP-2018*

Methods

- **Selected Variants:**
- **Based on GWAS on sequence data:**
  - Only GWAS data subset *N. Duijvesteijn et al WCGALP-2018*
  - -Log(P_Value) ≥ 3
  - Pruning: *(LD ≥ 0.95, 100kb windows , ~4,500 variants)*
  - Other (P_Value) threshold tested on one trait.

Methods

- **Genomic prediction**:
  - **GBLUP performed based on**: *MTG2 program (Lee et al 2016).*
    - *1) Routine 50k genotypes.*
    - *2) Whole Genome Sequence data*
    - *3) Selected Sequence Variants*
    - *3) 50k + Sel_Seq*
      - *1) Fitted as one variance component.*
      - *2) Fitted as two variance components jointly.*

        MTG2 program (Lee et al 2016).
- **Prediction accuracy**:
  - *r(GBV, Phen)/h*
- **Bias of prediction**:
  - [1- *RegCoeff* ...]

Results

Heritability estimates based on 50k, WGS and 50k + Selected Variants

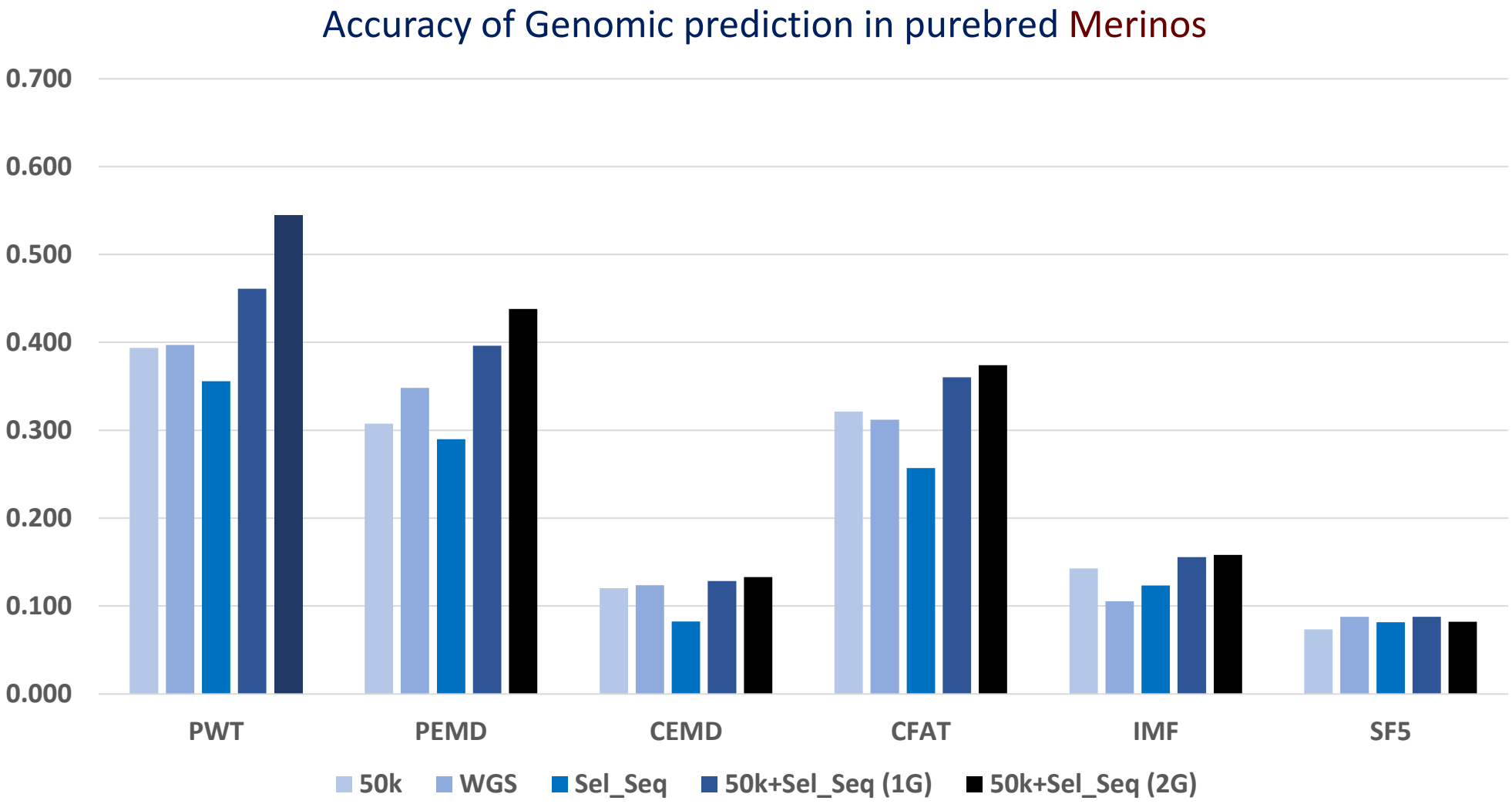
Trait	Size	h^2, 50k	h^2, WGS	h^2 (50k , Sel-Variants)
PWT (*kg*)	11,067	0.21	0.25	0.16 , 0.06
PEMD (*mm*)	9,715	0.23	0.26	0.19 , 0.09
CEMD (*mm*)	7,714	0.16	0.19	0.14 , 0.03
CFAT (*mm*)	7,635	0.19	0.21	0.13 , 0.07
IMF (*%*)	6,353	0.38	0.42	0.33 , 0.07
SF5 (*Newtons*)	7,392	0.24	0.29	0.14 , 0.11

Results

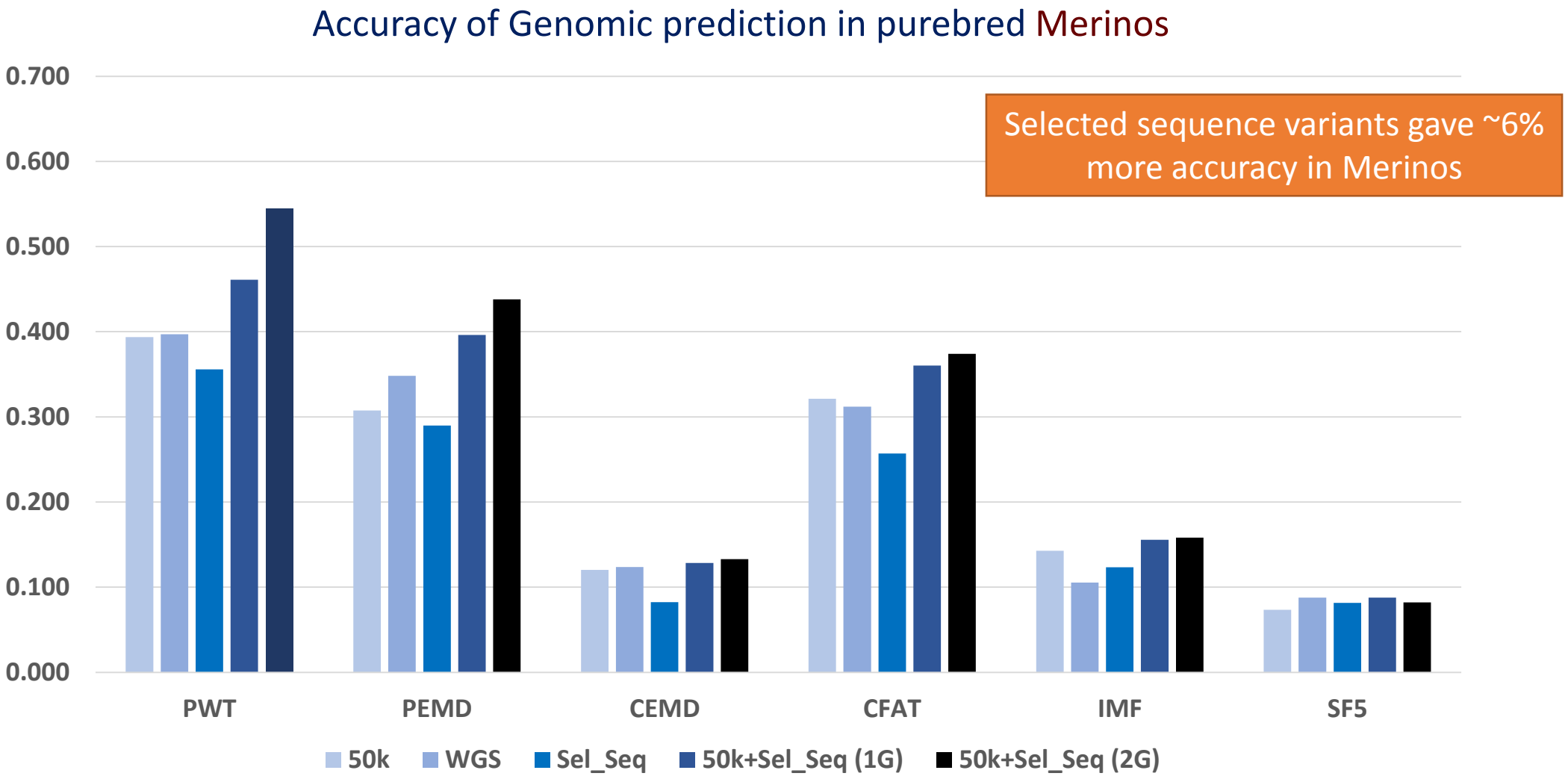
Heritability estimates based on 50k, WGS and 50k + Selected Variants

Trait	Size	h^2, 50k	h^2, WGS	h^2 (50k , Sel-Variants)
PWT (*kg*)	11,067	0.21	0.25	0.16 , 0.06
PEMD (*mm*)	9,715	0.23	0.26	0.19 , 0.09
CEMD (*mm*)	7,714	0.16	0.19	0.14 , 0.03
CFAT (*mm*)	7,635	0.19	0.21	0.13 , 0.07
IMF (*%*)	6,353	0.38	0.42	0.33 , 0.07
SF5 (*Newtons*)	7,392	0.24	0.29	0.14 , 0.11

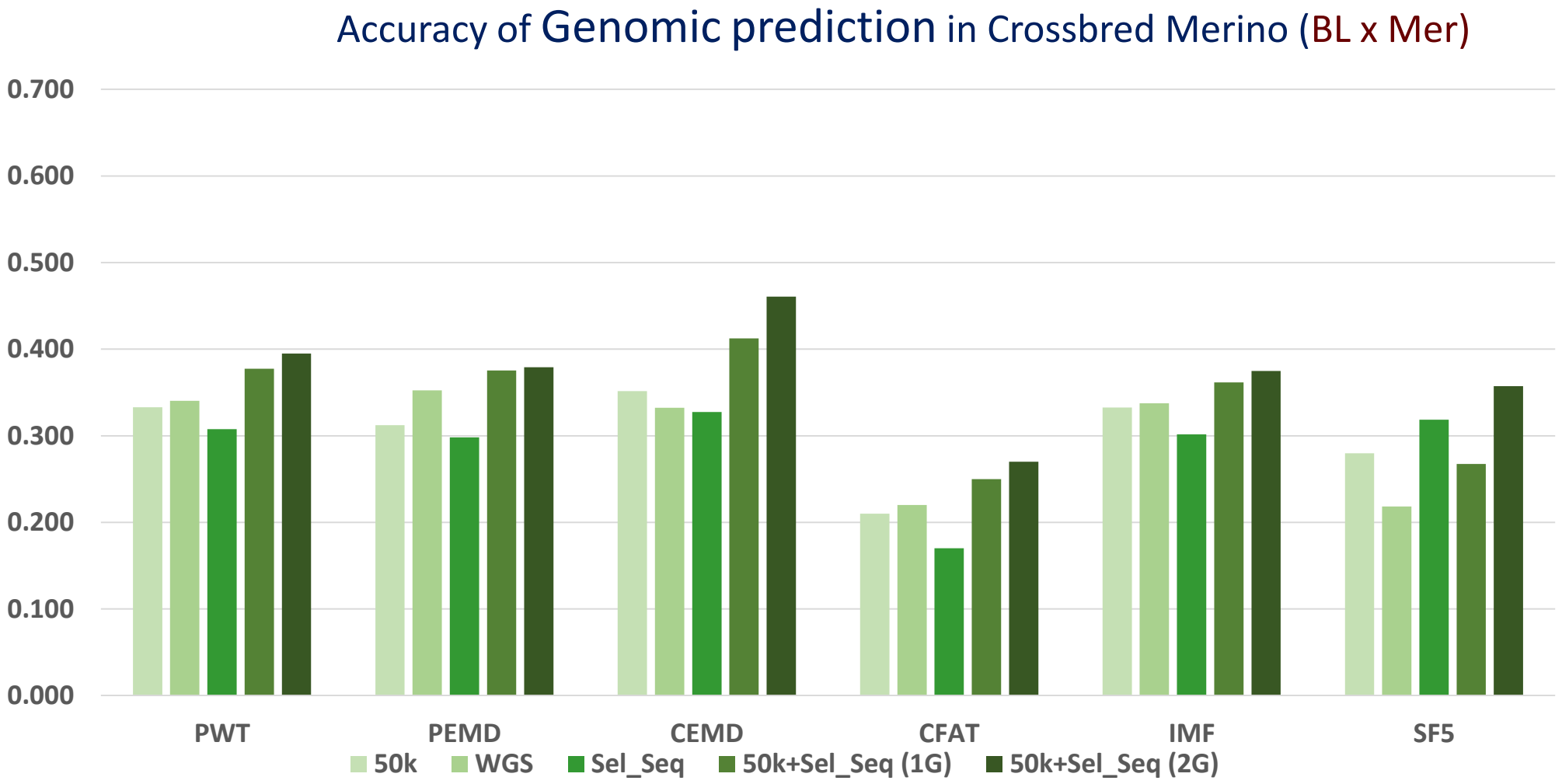
Results


Accuracy of Genomic prediction in purebred Merinos
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
PWT
PEMD
CEMD
CFAT
IMF
SF5
50k
WGS
Sel_Seq
50k+Sel_Seq (1G)
50k+Sel_Seq (2G)

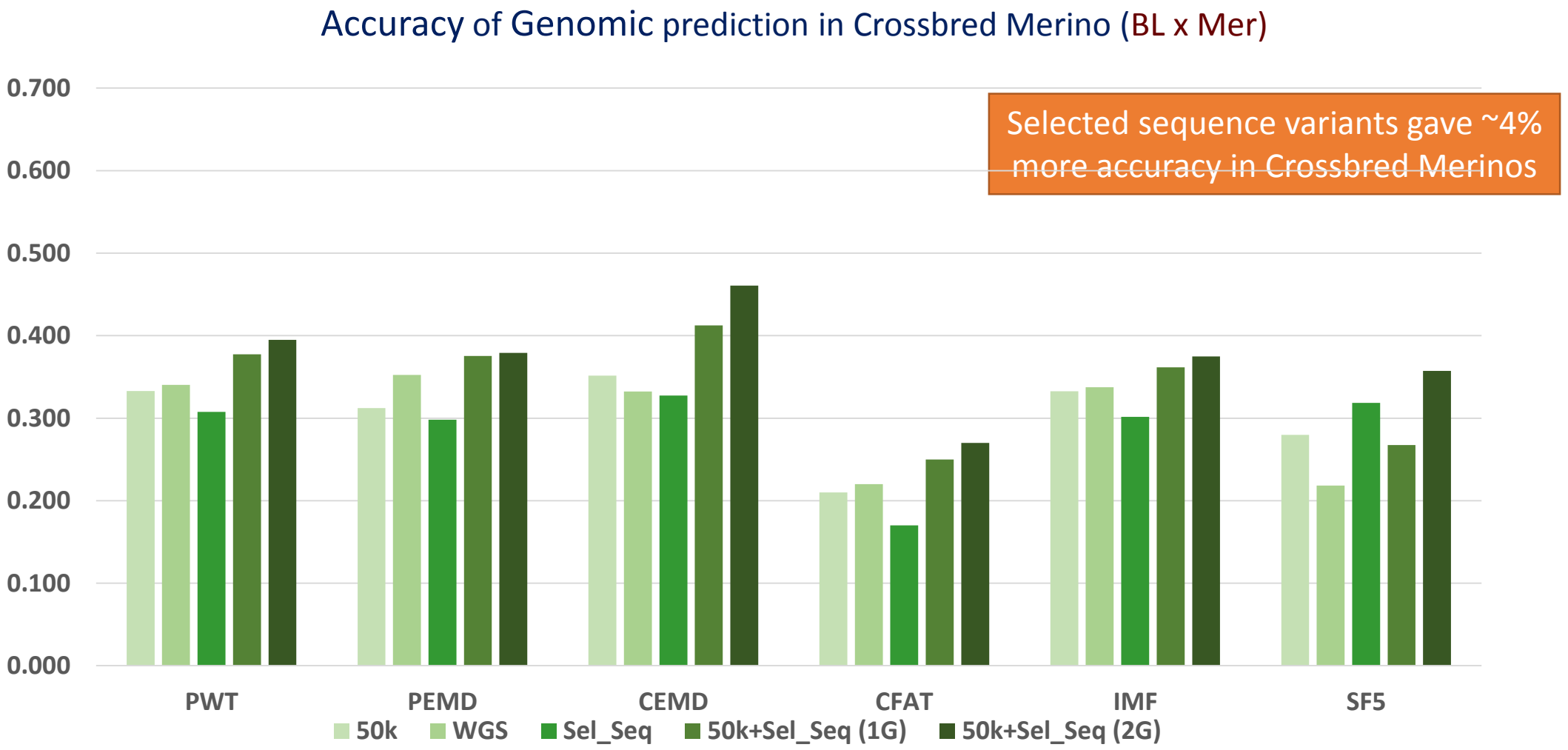
Results


Accuracy of Genomic prediction in purebred Merinos
Selected sequence variants gave ~6% more accuracy in Merinos
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
PWT
PEMD
CEMD
CFAT
IMF
SF5
50k
WGS
Sel_Seq
50k+Sel_Seq (1G)
50k+Sel_Seq (2G)

Results


Accuracy of Genomic prediction in Crossbred Merino (BL x Mer)
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
PWT
PEMD
CEMD
CFAT
IMF
SF5
50k
WGS
Sel_Seq
50k+Sel_Seq (1G)
50k+Sel_Seq (2G)

Results


Accuracy of Genomic prediction in Crossbred Merino (BL x Mer)
Selected sequence variants gave ~4% more accuracy in Crossbred Merinos
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
PWT
PEMD
CEMD
CFAT
IMF
SF5
50k
WGS
Sel_Seq
50k+Sel_Seq (1G)
50k+Sel_Seq (2G)

Results

Regression coefficient of adjusted phenotypes from GBV in purebred and crossbred Merino validation sets

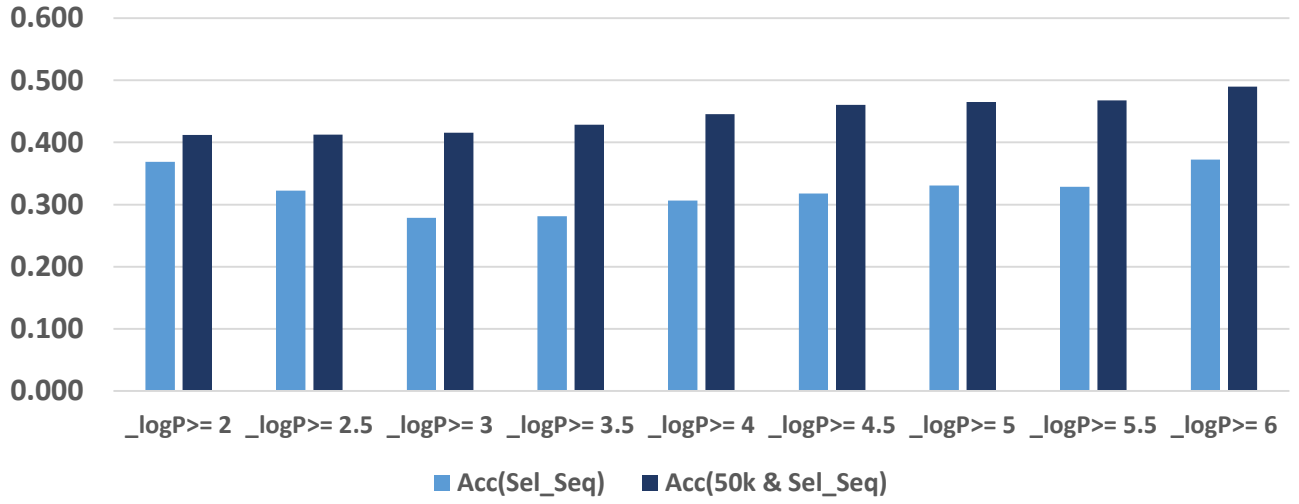
	Purebred Merinos				Crossbred Merinos			
Trait	50k	WGS	Sel-Variants[1]	50k+Sel-Variants	50k	WGS	Sel-Variants	50k+Sel-Variants
PWT	0.92	0.91	1.14	1.06	0.89	0.89	0.80	0.88
PEMD	0.87	0.90	0.74	0.88	0.92	0.95	0.68	0.84
CEMD	0.89	0.88	0.70	0.77	1.14	1.00	0.69	1.62
CFAT	1.06	1.10	0.61	0.91	0.36	0.72	1.44	1.07
IMF	0.51	0.49	0.44	0.50	0.85	0.88	0.84	0.88
SF5	0.34	0.46	0.61	0.35	0.64	0.56	1.44	0.65

[1]: Selected sequence variants, -LogP ≥ 3

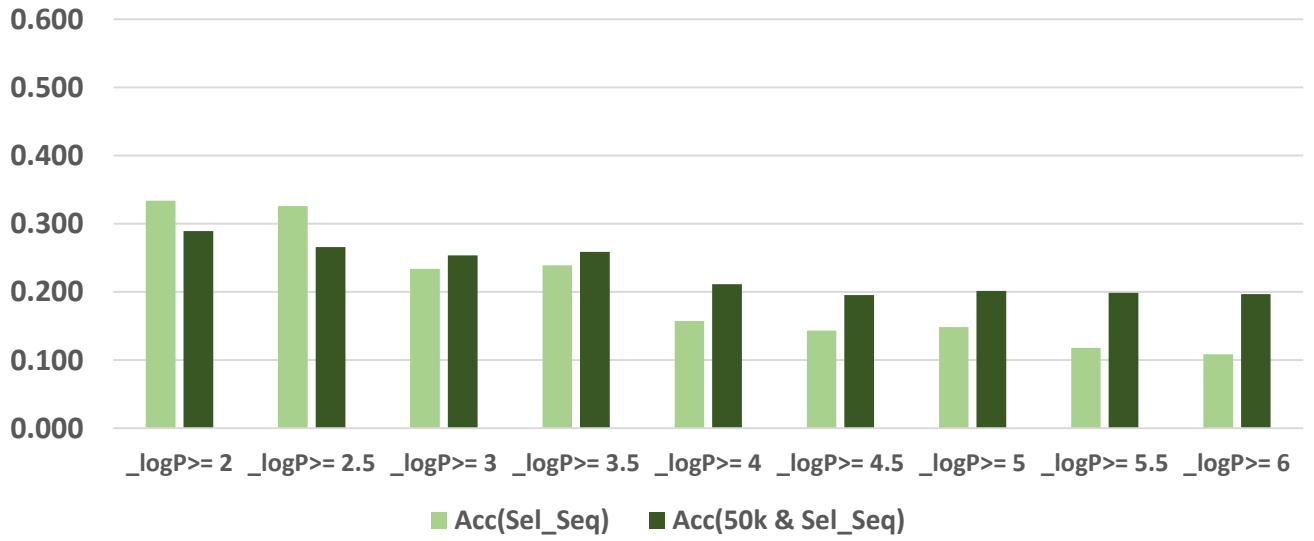
Results

Genomic prediction of CFAT using selected variants based on different GWAS threshold

Purebred Merinos

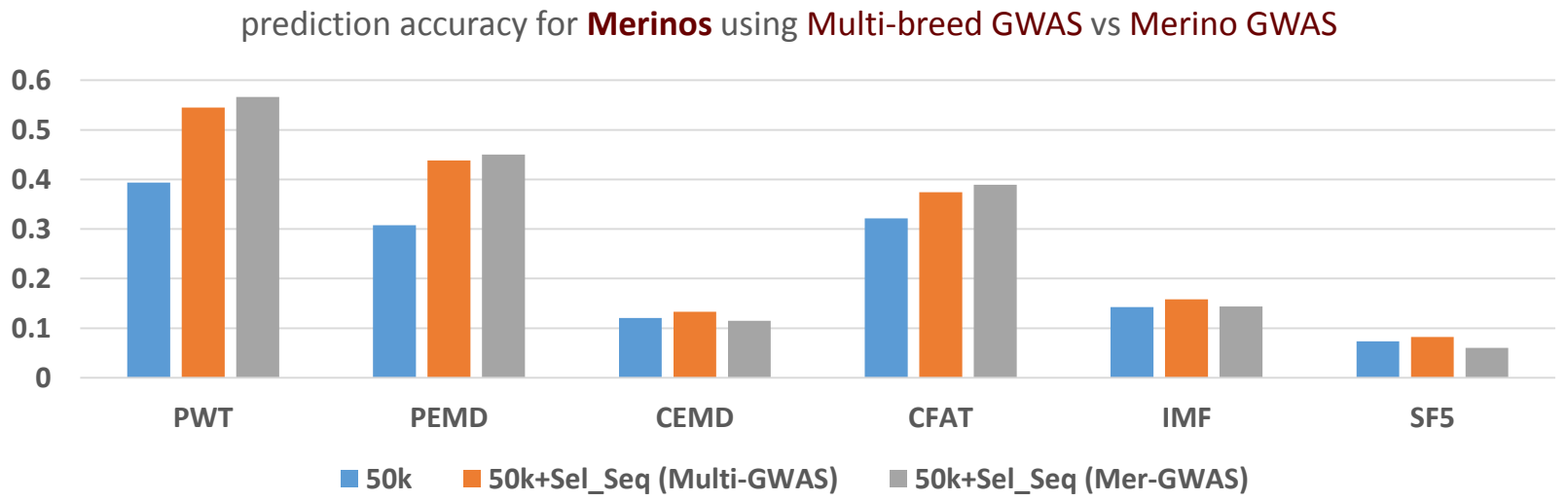


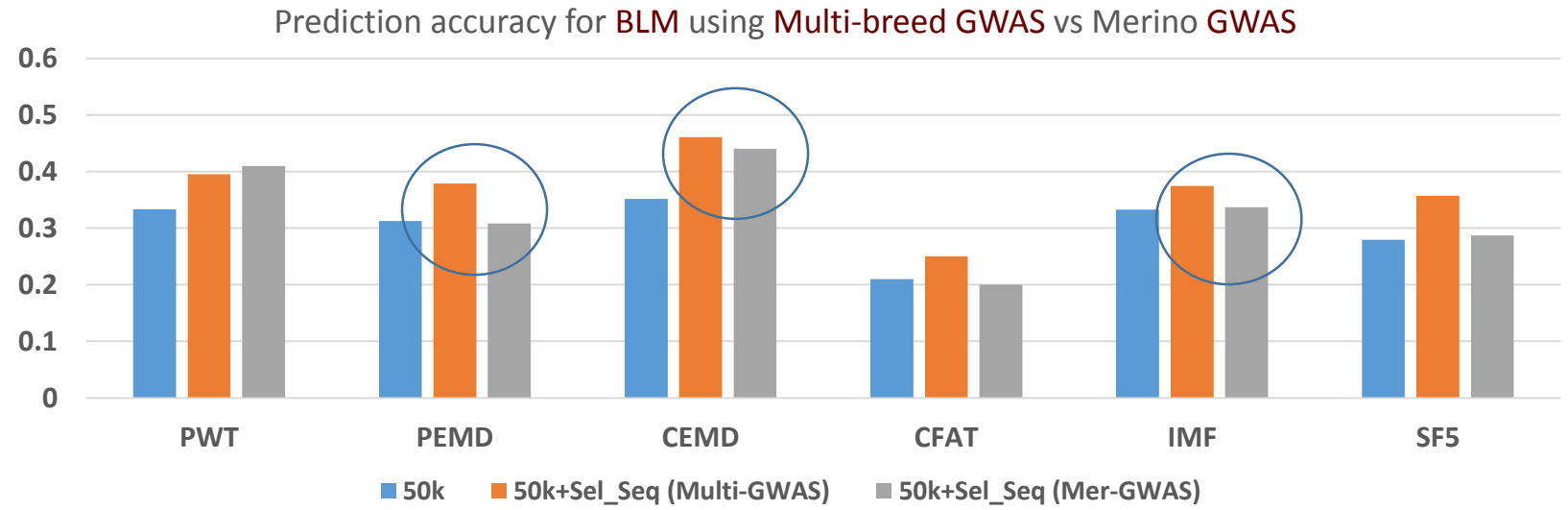
Crossbred Merinos



Results

Multi-breed GWAS vs single breed GWAS





Conclusions

- **Genomic prediction accuracy increased substantially by using selected sequence variants.**

- **Multi-breed GWAS outperformed single breed GWAS.**

- **Stronger threshold on selected variant didn’t persistently improved the accuracy between different traits and purebred or crossbred animals**

- **GBLUP methods can accommodate the selected variants by fitting them as a separate variance component and to avoid double counting of part of additive genetic variance.**

Acknowledgements

- *Sheep CRC*

- *AGBU*

- *UNE*

- *SheepGenome DB*

- *BioSci Research Agriculture, Victoria*

- *Meat Livestock Australia (MLA)*

- *Sheep Genetics Australia (SG)*

-*INF Research Farm Staff*

- *Breeders for contributing phenotypic data*

Thanks