

# Accuracy of indirect predictions for large datasets

Andre Garcia, I. Aguilar, A. Legarra, S. Miller, S. Tsuruta,  
I. Misztal and D. Lourenco



07/21/2020 – ASAS annual meeting



# Introduction

- ◇ Why Indirect Predictions (IP)?
- ◇ Large number of genotyped animals
- ◇ No phenotypes + Missing pedigrees
- ◇ Non-registered animals
- ◇ Interim evaluations

# Introduction

◆ How to obtain IP from ssGBLUP?

$$\hat{\mathbf{a}} = \alpha\delta \frac{1}{2 \sum pq} \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}_{22} \quad \text{IP} = \mathbf{Z}\hat{\mathbf{a}}$$

◆ How about Accuracy of IP?

$$\text{acc}_i^{\text{IP}} = \sqrt{1 - \mathbf{z}_i \text{LHS}^{aa} \mathbf{z}_i' \sigma_u^{-2}}$$



SNP-BLUP MME  
PEC of SNP

VanRaden 2008  
Stranden & Garrick 2009  
Wang et al. 2012  
Liu et al. 2017

# Introduction

- ◆ Accuracy of IP based on SNP PEC from ssGBLUP

$$\text{PEC} = \text{var}(\hat{a}) = \alpha\delta \frac{1}{2 \sum pq} Z' G^{-1} (G\sigma_u^2 - C^{u22}) G^{-1} Z \frac{1}{2 \sum pq} \alpha\delta$$



ssGBLUP LHS

For genotyped animals

- ◆ Once PEC are available

$$\text{acc}_i^{\text{IP}} = \sqrt{1 - \alpha\delta z_i \text{PEC}(\hat{a}_i) z_i' \alpha\delta \sigma_u^{-2}}$$

Gualdrón Duarte et al. 2014

Legarra et al. 2018

Aguilar et al. 2019

Liu et al. 2017

# Objective

- ◆ Obtain SNP PEC and accuracy of IP from ssGBLUP
- ◆ Investigate how many genotyped animals are needed to compute PEC
  - ◆ Is the accuracy of IP robust?
  - ◆ How does it compare to the accuracy of GEBV?



# Materials and Methods

- ◇ Data from the American Angus Association
  - ◇ 38,000 post weaning gain (PWG) records
  - ◇ ~230,000 animals in pedigree
  - ◇ 60,000 genotyped animals
    - ◇ 5,467 validation animals (born in 2018)
    - ◇ No phenotypes or pedigree in the data



# Materials and Methods

## ◇ Benchmark

- ◇ GEBV from ssGBLUP
- ◇ Accuracy based on PEV
- ◇ Genotypes for validation animals included

## ◇ Testing

- ◇ IP and IP accuracy for validation animals
- ◇ No data or genotypes of validation animals were used to compute SNP effects and SNP PEC

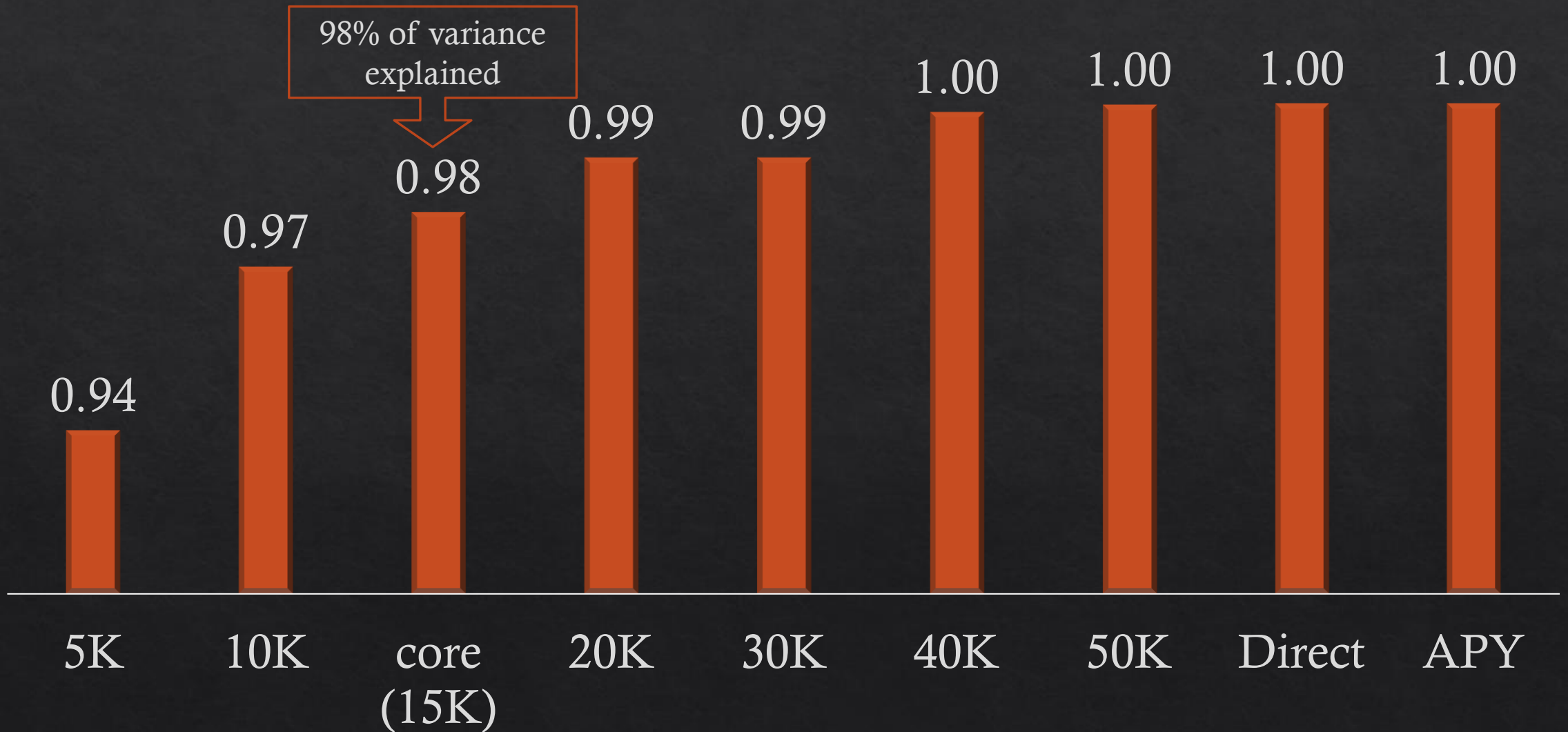
# Materials and Methods

- ◇ Testing scenarios
  - ◇ 54K direct  $G^{-1}$  (all but validation animals)
  - ◇ 54K APY  $G^{-1}$
  - ◇ 5 to 50K
  - ◇ Core animals (15K)
- ◇  $\text{Corr}(\text{GEBV}_{\text{acc}}, \text{IP}_{\text{acc}})$
- ◇  $\text{GEBV}_{\text{acc}} = b_0 + b_1 * \text{IP}_{\text{acc}}$



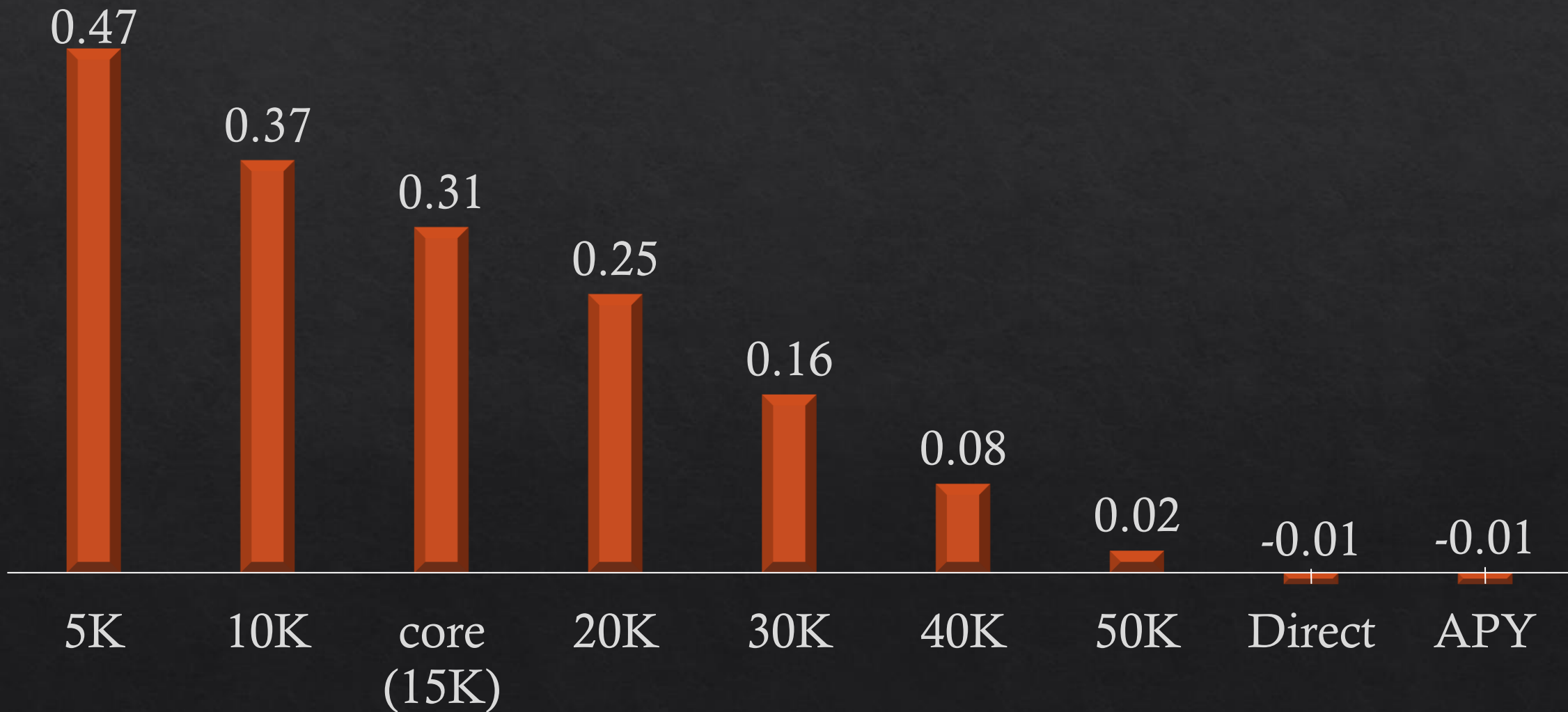
# Results

$$\text{corr}(\text{GEBV}_{\text{acc}}, \text{IP}_{\text{acc}})$$



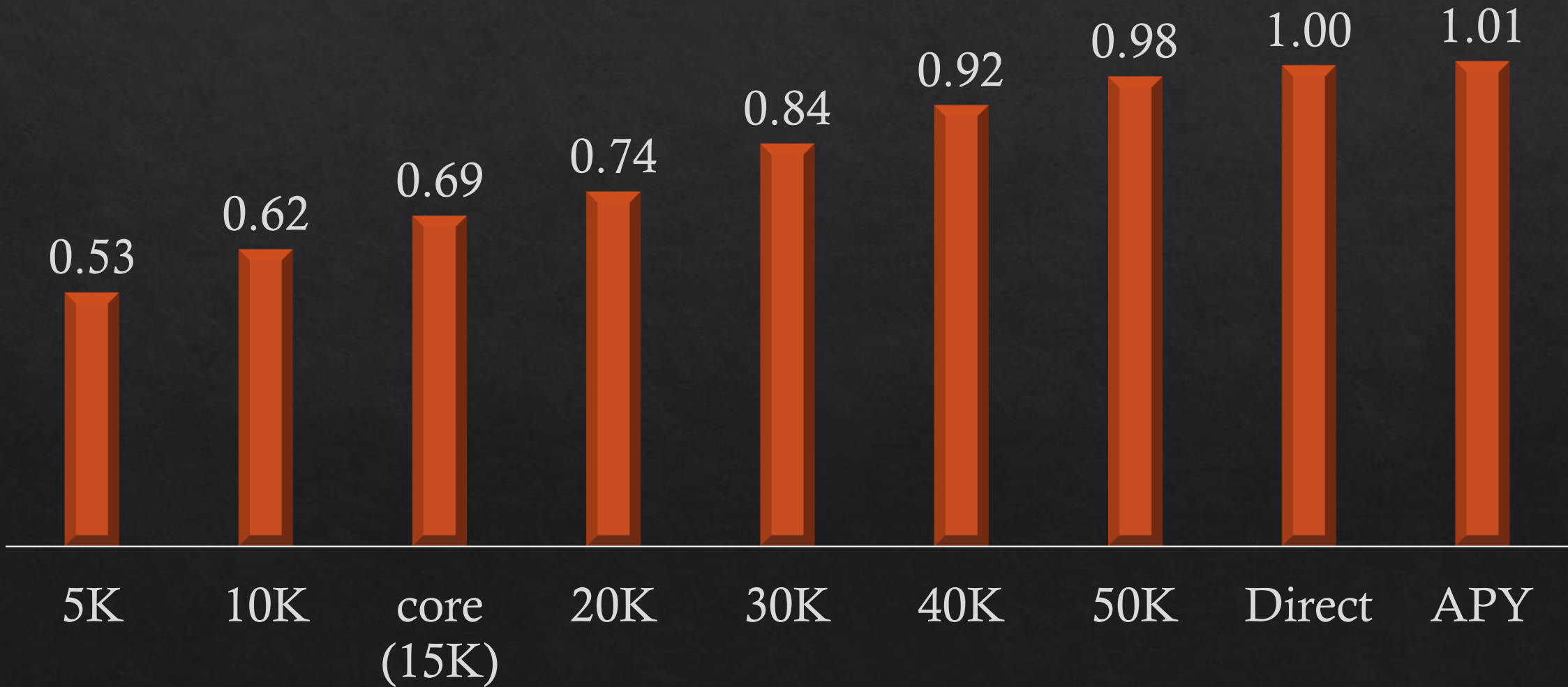
$$\text{GEBV}_{\text{acc}} = b_0 + b_1 * \text{IP}_{\text{acc}}$$

Ideally  $b_0 = 0$



$$\text{GEBV}_{\text{acc}} = b_0 + b_1 * \text{IP}_{\text{acc}}$$

Ideally  $b_1 = 1$



# Final Remarks

- ◇ Accuracy of IP are available from ssGBLUP
- ◇ Correlations with GEBV accuracy are  $\geq 0.99$ 
  - ◇ When enough genotyped animals are used
- ◇ Fine tuning is needed to obtain  $b_0 = 0$  and  $b_1 = 1$
- ◇ Extend computations of accuracy to large scale evaluations
- ◇ Extend ideas to approximate accuracy of GEBV



# Acknowledgements



andre.garcia@uga.edu  
nce.ads.uga.edu

