

Decreasing computing time of approximated reliabilities of GEBV in ssGBLUP using different core sizes for APY

S. Sanchez-Sierra*, M. Bermann*, N. Vukasinovic†, D. Gonzalez-Peña†, G. Varga†, M. A. Sánchez-Castro†, G. A. Oliveira Junior†, I. Misztal*, D. Lourenco*

*Department of Animal and Dairy Science, University of Georgia, Athens, GA 30605

†Zoetis Genetics, Kalamazoo, MI 49007

Motivation

- Direct calculation of PEV in large-scale evaluations is not feasible
- Several approximations exist
- One of them is based on the algorithm for proven and young - APY
- Although precise, it can have high computing cost

Objective: Decrease the computing cost for approximating GEBV reliabilities in ssGBLUP by reducing the size of the core set in APY without affecting the quality of the approximations

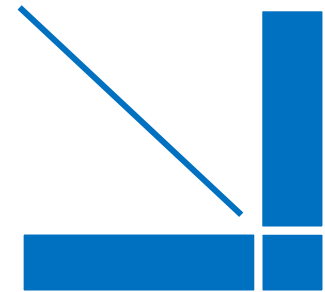


Genomic reliabilities

➤ GBLUP with APY

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha \mathbf{G}_{APY}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{M}_{nn}^{-1} & \mathbf{G}^{nc} \\ \mathbf{G}^{cn} & \mathbf{G}^{cc} \end{bmatrix} =$$

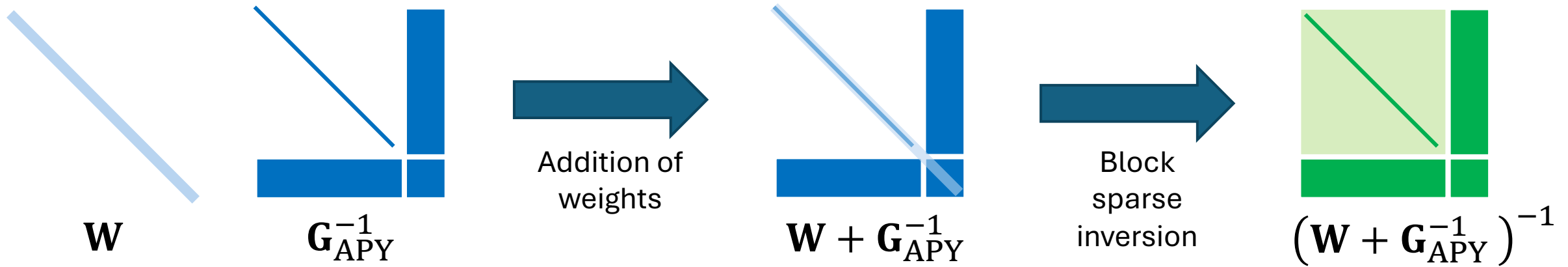


$$\text{diag}(\mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} + \mathbf{G}_{APY}^{-1})^{-1} \approx \text{diag}(\mathbf{W} + \mathbf{G}_{APY}^{-1})^{-1}$$

ACCF90GS2

Bermann et al. (2022)

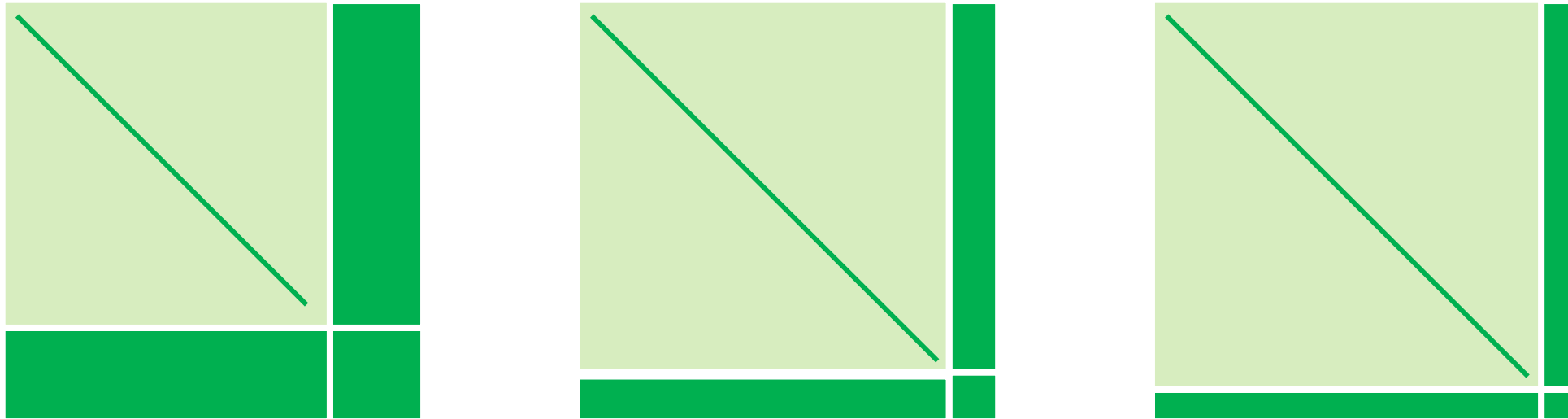
➤ Reliability **approximation** using G-APY



$$\max(O(n_c^3), O(n_n n_c^2))$$

Computing cost reduction

- Smaller core blocks in G-APY lead to faster computations



To what extent will the quality of the reliabilities be affected?

Datasets

Trait	Pedigree	Genotypes	Records	Model	h^2
Calf respiratory disease	4.5 M	1.6 M	1.5 M	Repeatability	0.042
Cystic ovaries	427 K	107 K	127 K	Repeatability	0.054

Methods

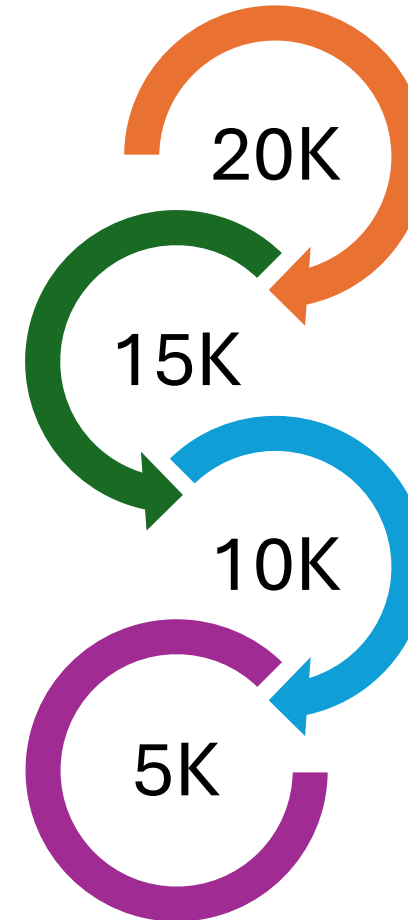
- **Approximated** reliabilities comparison under different core sizes in G-APY



Original core size

Vs

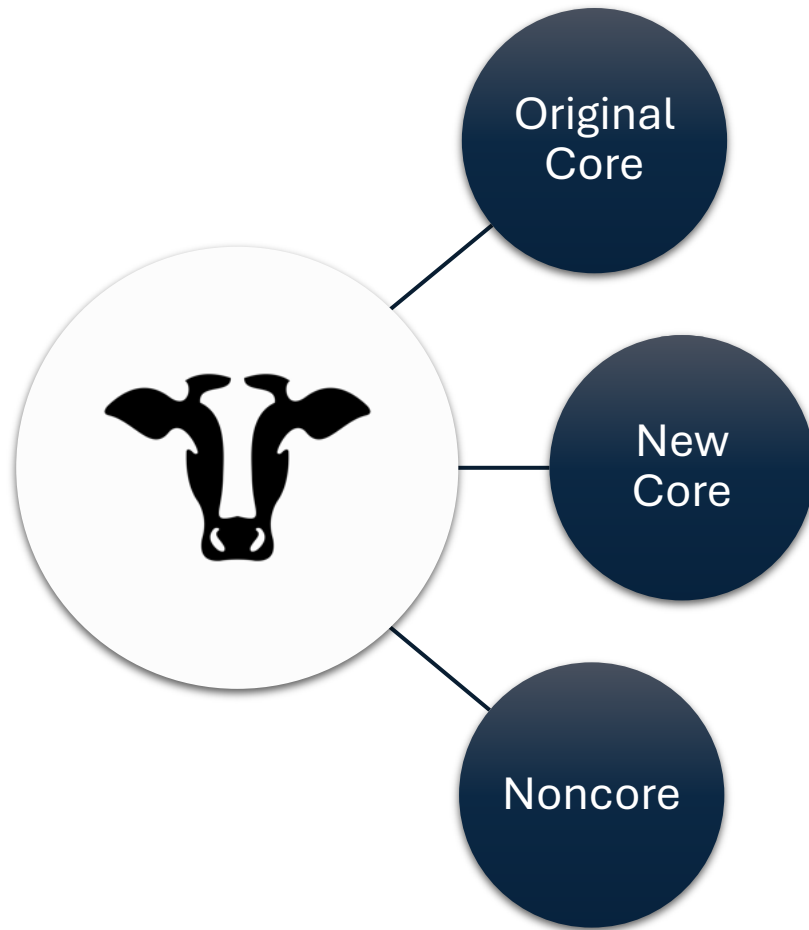
New Core sizes



Benchmark

Assessments

Evaluation criteria



Scatter plots

Regression coefficients

Correlation coefficient

Absolute difference

Computing time



Results

Calf respiratory disease

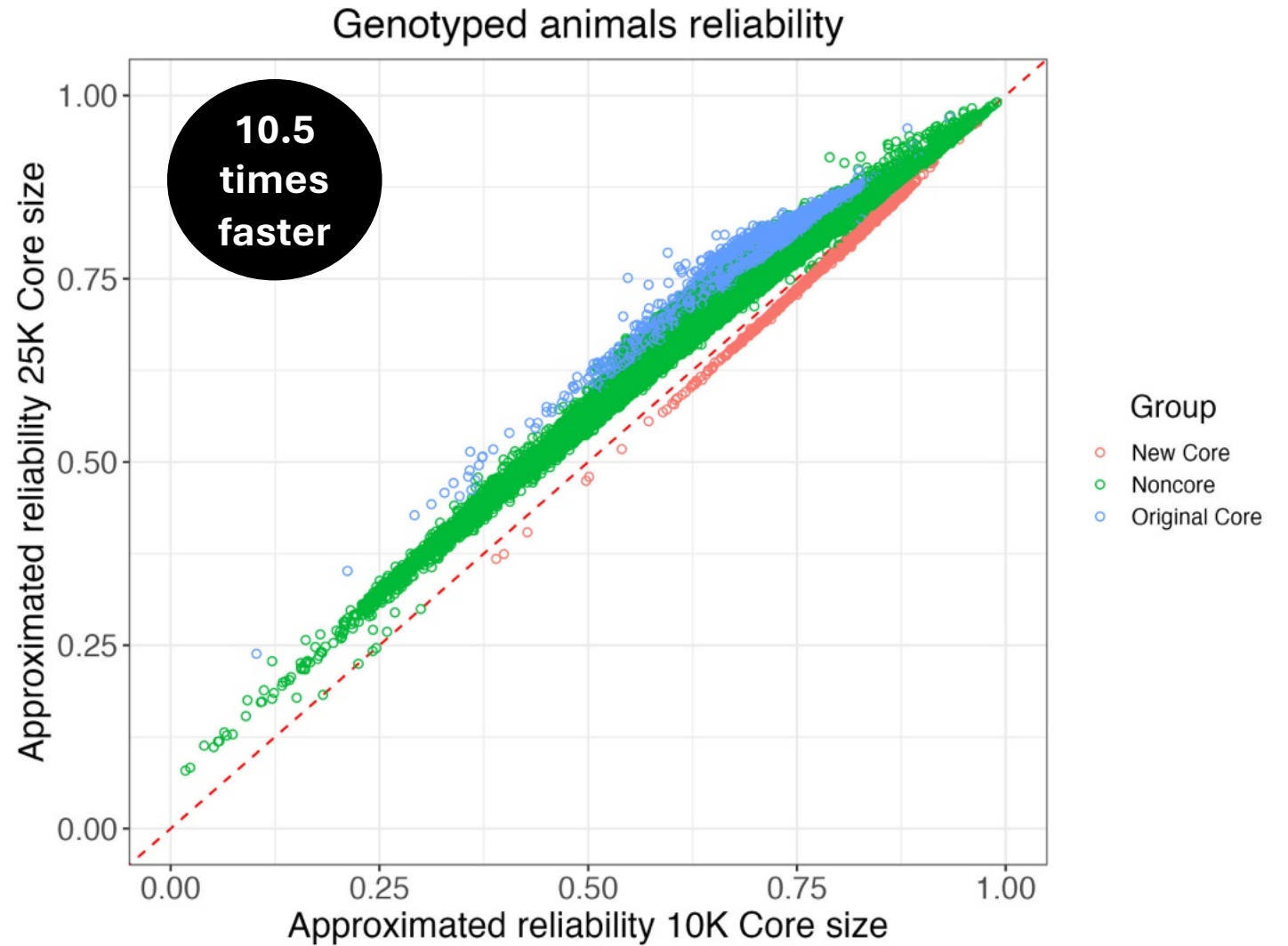
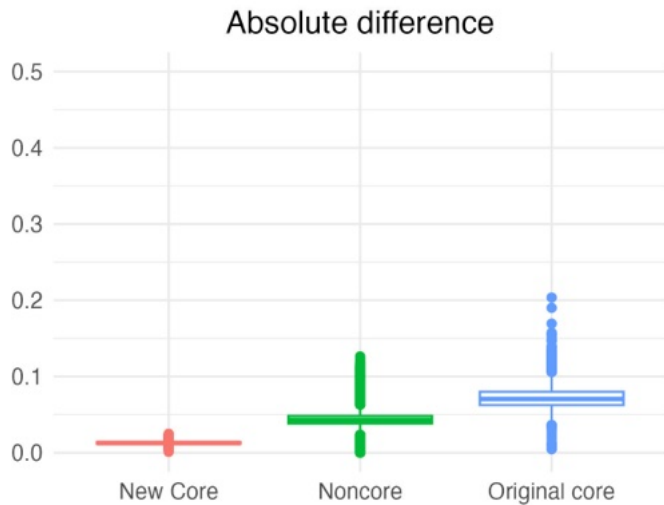
25K core vs 10K core

5K core



25K vs 10K

	Correlation	Intercept	Slope
New Core	1.00	-0.05	1.05
Noncore	1.00	0.14	0.87
Original core	0.98	0.25	0.75
Computing time (min)		25k	456.80
		10k	43.10

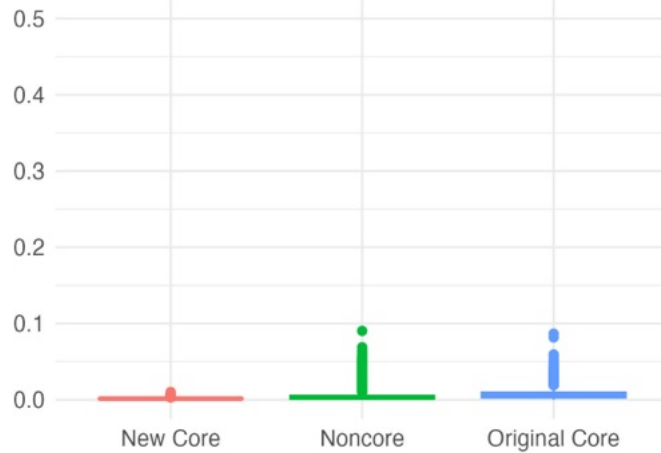


25K vs 10K

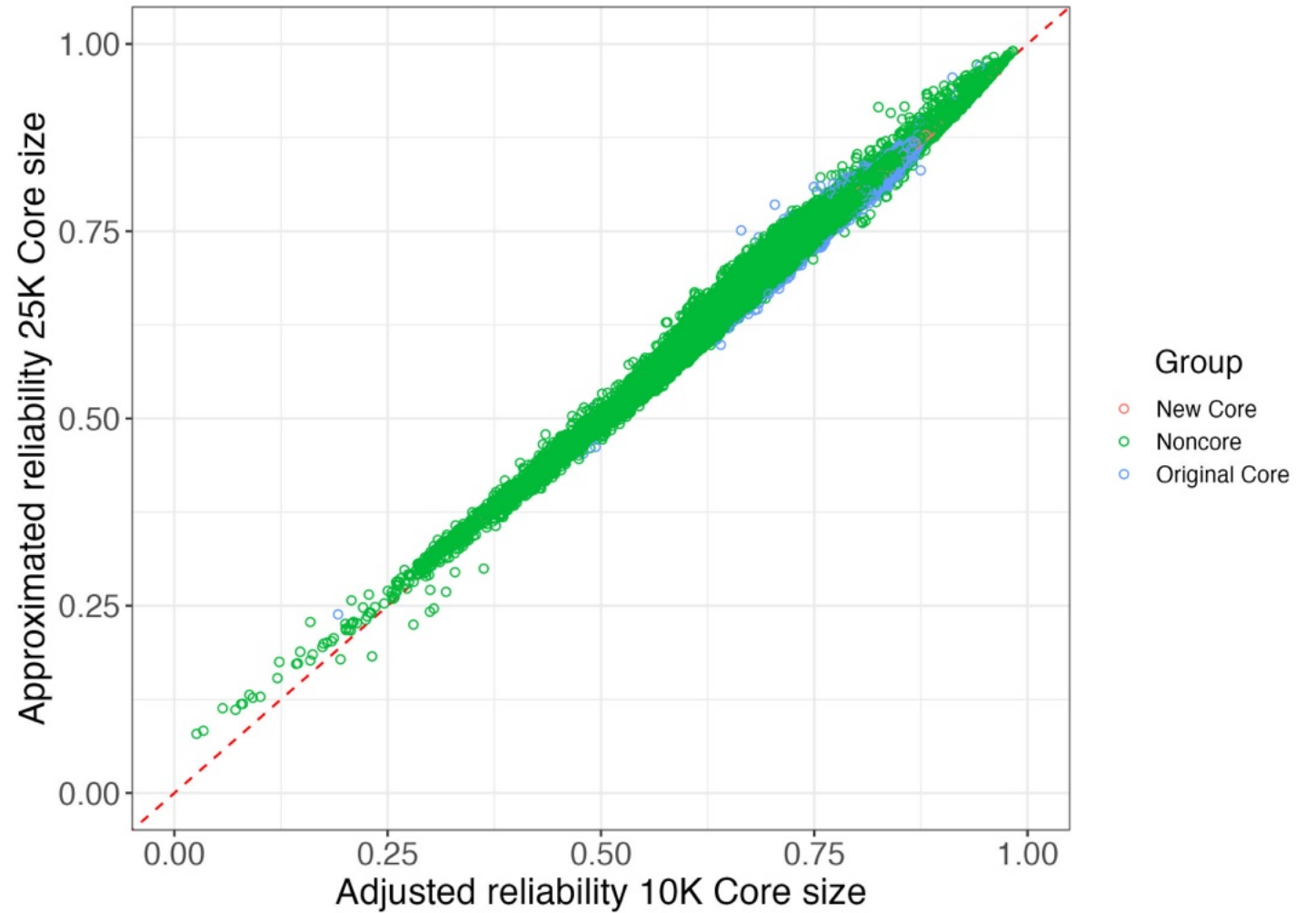
Adjusted Reliability

$$rel_{25k} = a * rel_{10k}^b * e^{-c*rel_{10k}}$$

Absolute difference

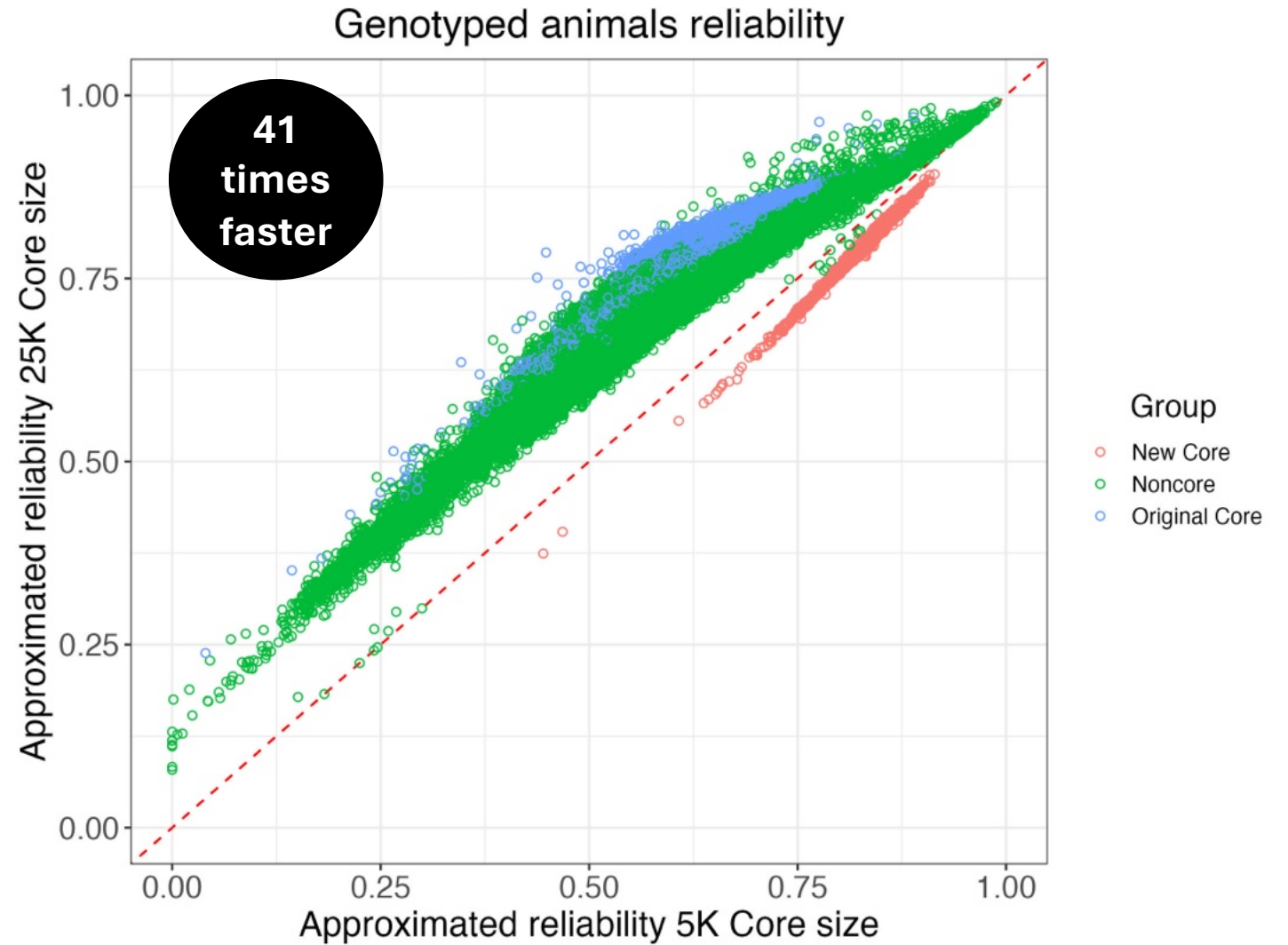
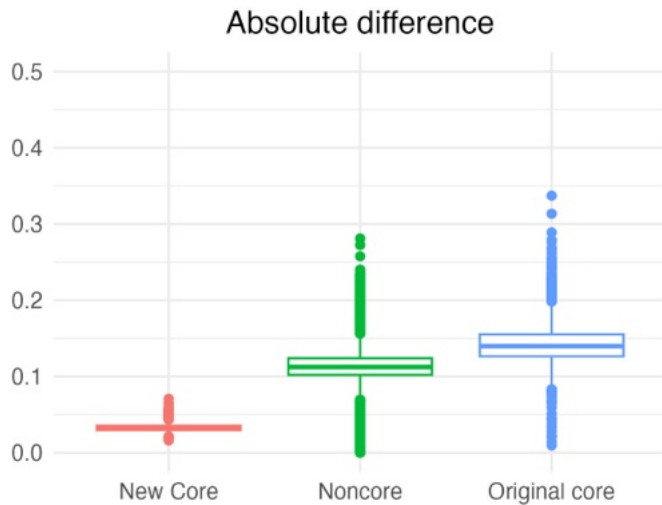


Genotyped animals reliability



25K vs 5K

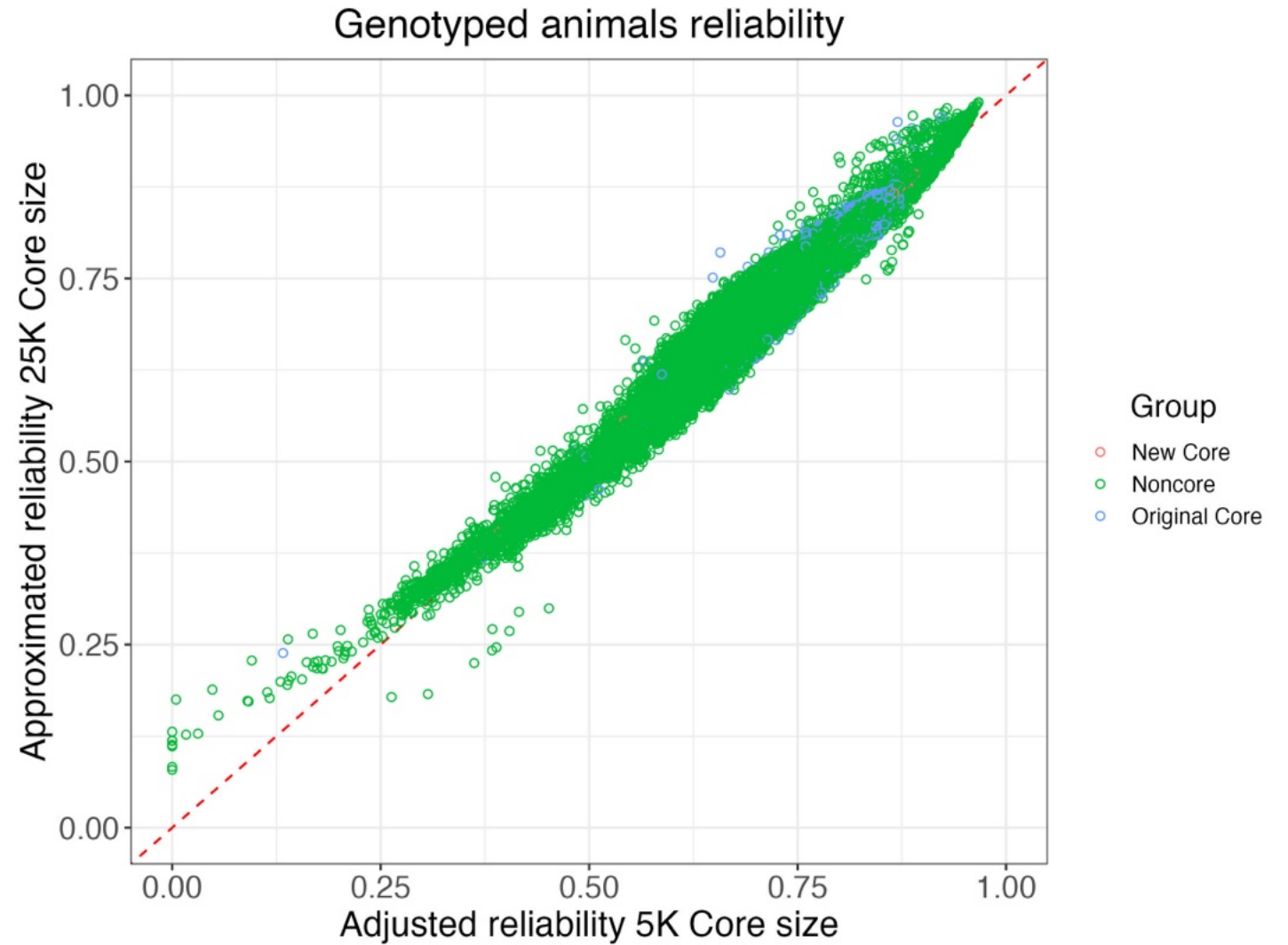
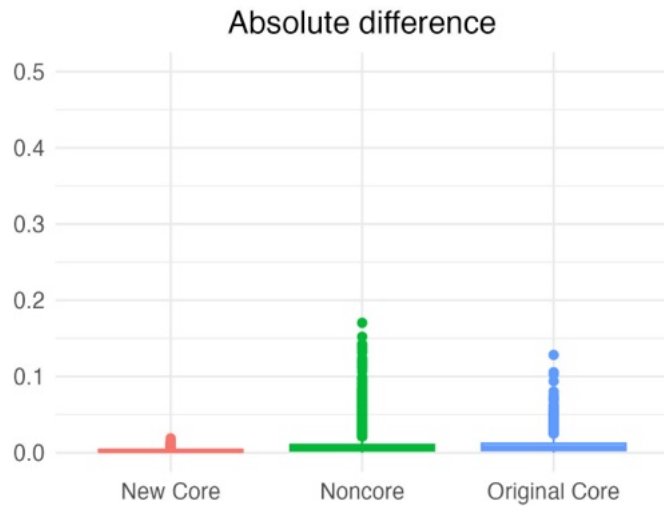
	Correlation	Intercept	Slope
New Core	1.00	-0.16	1.15
Noncore	0.97	0.28	0.74
Original core	0.94	0.38	0.64
Computing time (min)		25k	456.8
		5k	11.1



25K vs 5K

Adjusted Reliability

$$rel_{25k} = a * rel_{5k}^b * e^{-c*rel_{5k}}$$



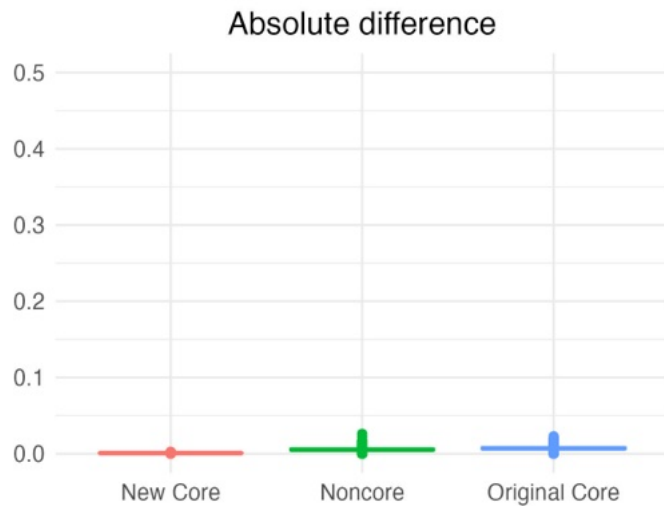
Results

Cystic ovaries

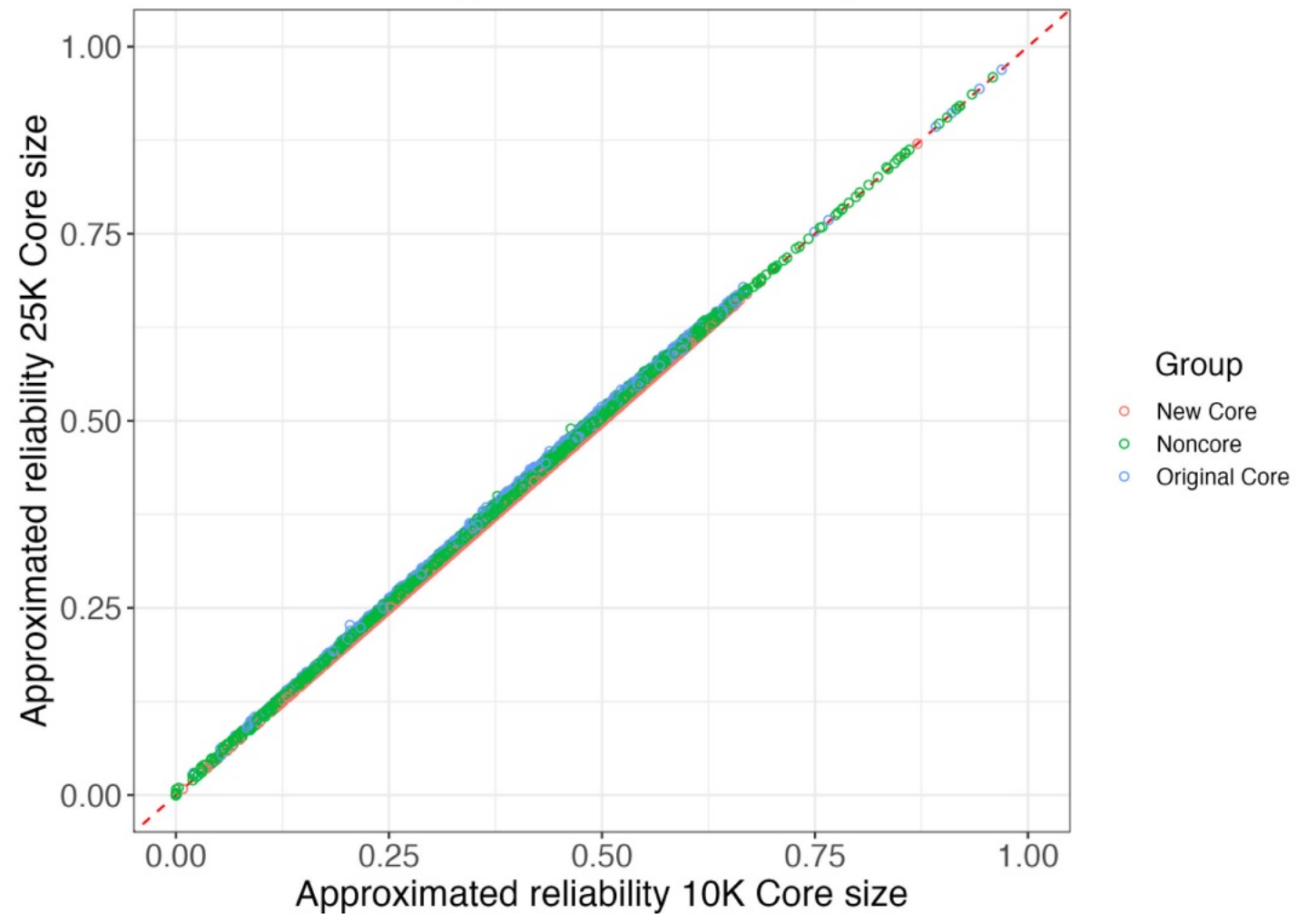
25K core vs 10K core
5K core

25K vs 10K

	Correlation	Intercept	Slope
New Core	1.00	0.00	1.00
Noncore	1.00	0.00	1.00
Original core	1.00	0.00	1.00
Computing time (min)		25k	7.44
		10k	2.54

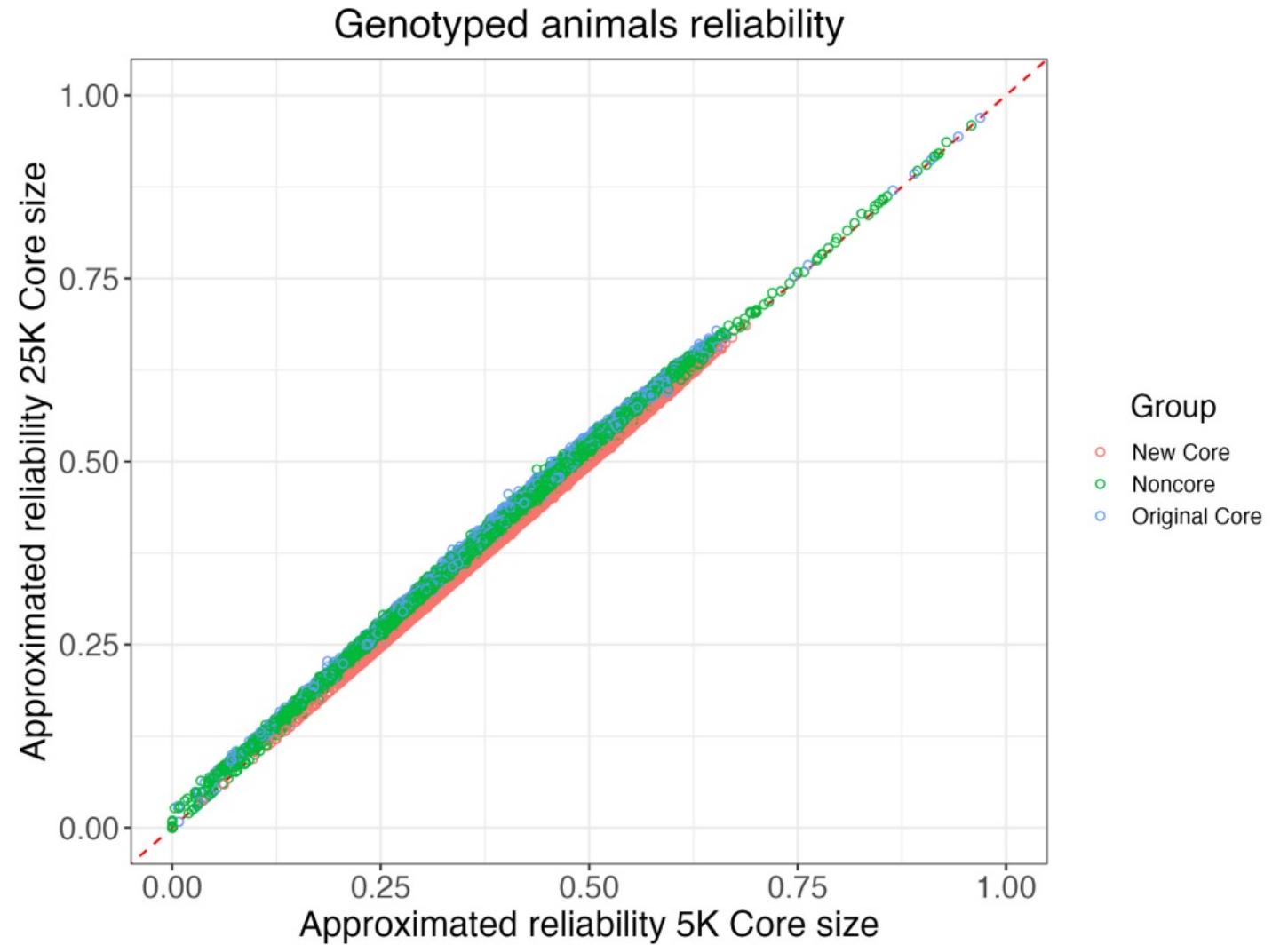
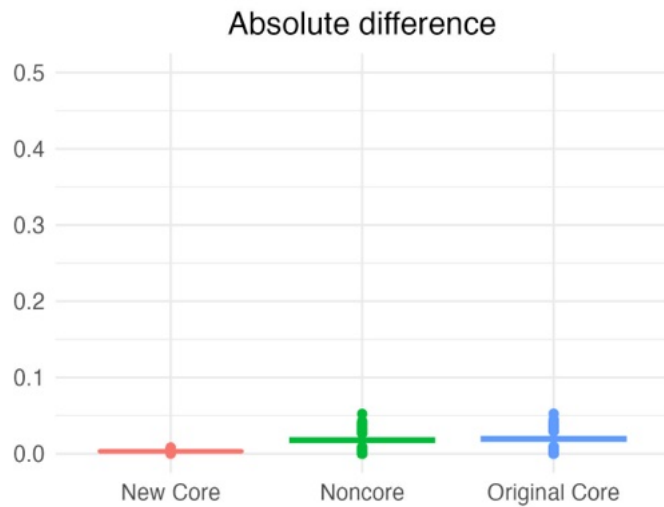


Genotyped animals reliability



25K vs 5K

	Correlation	Intercept	Slope
New Core	0.99	0.00	1.01
Noncore	0.99	0.02	1.01
Original core	0.99	0.02	1.01
Computing time (min)		25k	7.44
		5k	0.76



Conclusions

- Decreasing the core size in \mathbf{G}_{APY}^{-1} significantly reduces the computing time when approximating reliabilities in large datasets
- Changes in the approximated reliabilities occur and depend on the Genotyped/Core animal proportion
- Nonlinear transformation of approximated reliabilities could help to improve the quality of the estimations
- Parameters in the transformation are trait and model-dependent



Acknowledgments

