

# Working Group on Genomic Reliability

February 2015

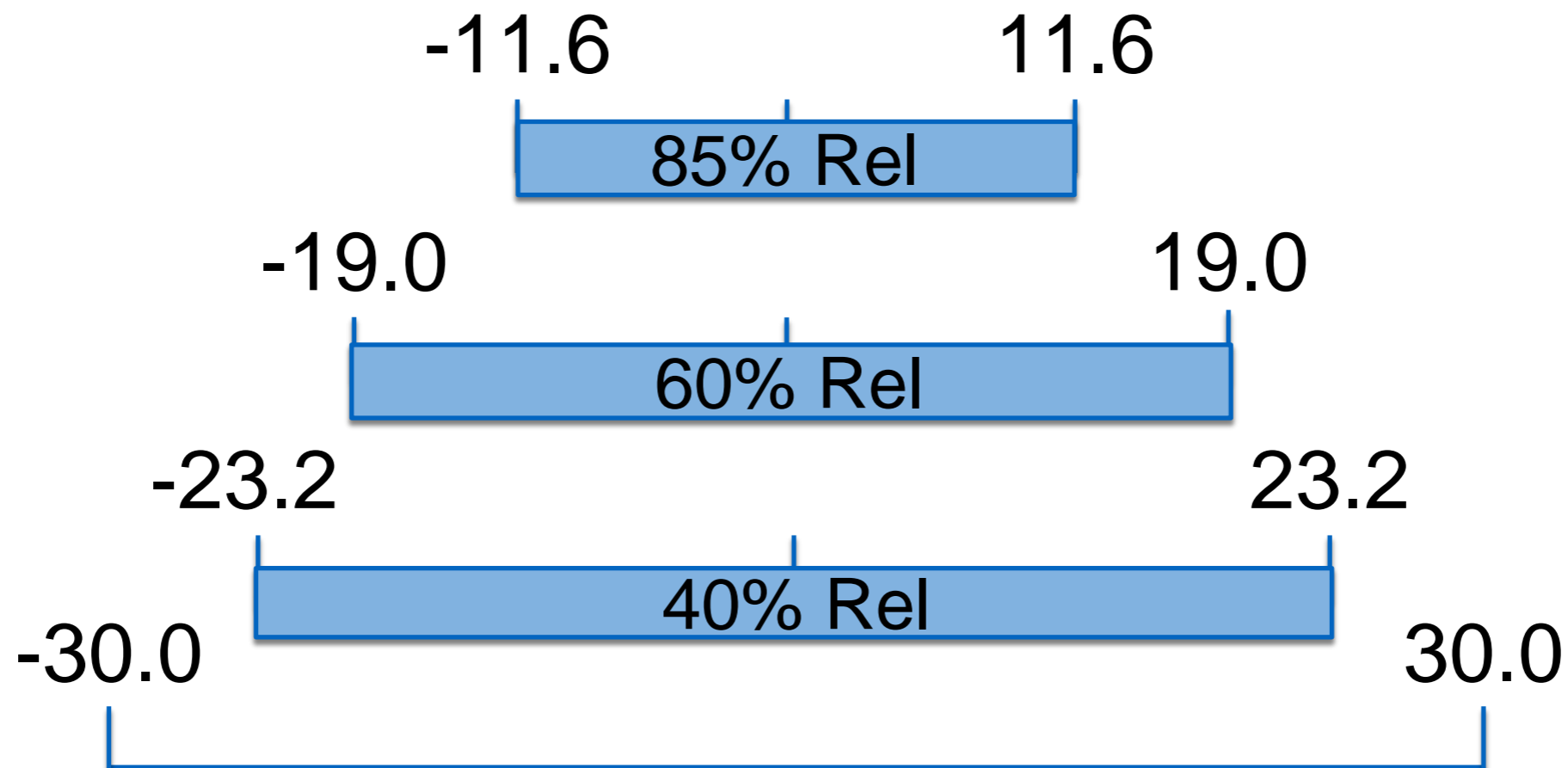
Mario Calus, Paul VanRaden, Martin Lidauer, Vincent Ducrocq,  
Zengting Liu and Bevin Harris

# Background

- Genomic reliability is a measure of “precision” of the genomic breeding value
- Measure of accuracy of prediction of young sires without daughter information
- It is reasonable to expect the validation  $R^2$  and average genomic reliability of a cohort of young sires to be of a similar magnitude

# Background

- Genomic reliability should be proportional to the expected change in the genomic breeding value as new information becomes available
- Example: 99% interval, genetic standard deviation=10



# ITB Berlin 2014

- Need to under take a simulation model
- Naive method had been developed
- Undergone limited testing on national data sets
- Showed some promise but difficult to assess without simulated data

# Simulation

- Mimicking a dairy cattle population using QMSim (Sargolzaei and Schenkel, 2009)
- Recent generations:
  - 45,000 females and 5,000 males per generation
  - 1,000 males selected for mating, based on 64% reliable EBV or at random
  - Heritability = 0.3; Bull reliability is ~0.8 (with 45 daughters)
  - ~9,000 QTL with effects from normal distribution & ~42,000 SNPs

# Simulation Data for analysis

- 2,000 training bulls from 2 generations
- 6,000 validation bulls from the next 3 generations
- TBV, genotypes and pedigree are provided for all 8,000 bulls and DYDs on 2,000 training bulls
- Heritability = 0.3; Bull reliability is  $\sim 0.8$  (with 45 daughters)

# Analysis: simulation data

- Very Preliminary Results: no replication
- Undertaken to ensure the simulated data behaves as expected
- Comparing traditional BLUP results with methods for genomic evaluation

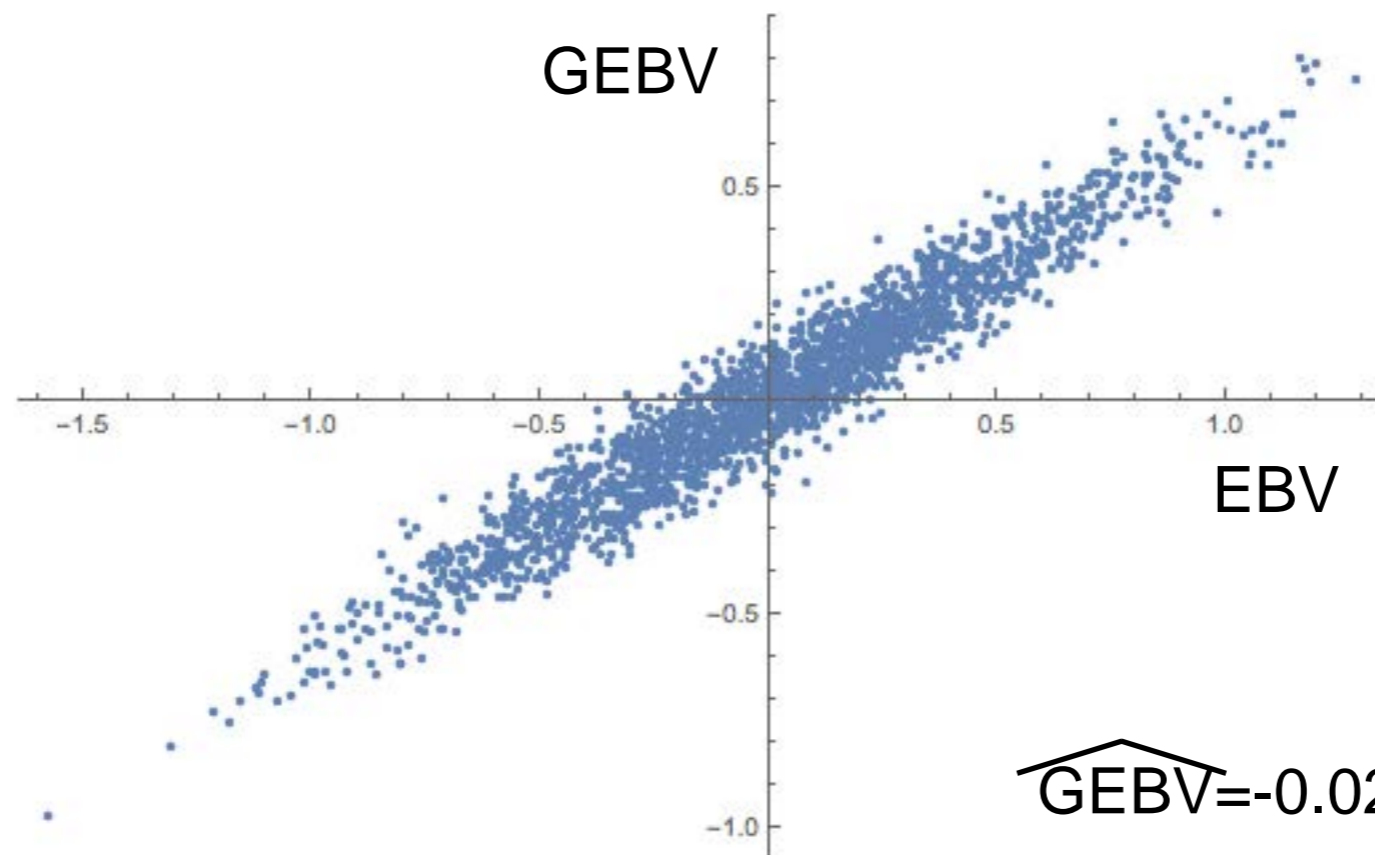
# Analysis: simulation data

- The data was checked for any big differences between G and A.
- For 8000 by 8000 bulls, only 1 pair of bulls (806434 and 880612) had a difference bigger 0.2 (related by 0.34 in G vs 0.13 in A)
- Genomic selection did not distort the genomic vs. pedigree relationships and shows that QMSim worked properly
- This is consistent with given the allele frequencies changed little, which is reasonable given many QTLs with small effects



# Simulation

- No selection dataset
- 2000 training sires BLUP EBVs versus GEBVs using GBLUP



$$\widehat{\text{GEBV}} = -0.02 + 0.98 \text{ EBV} \quad R^2 = 0.95$$

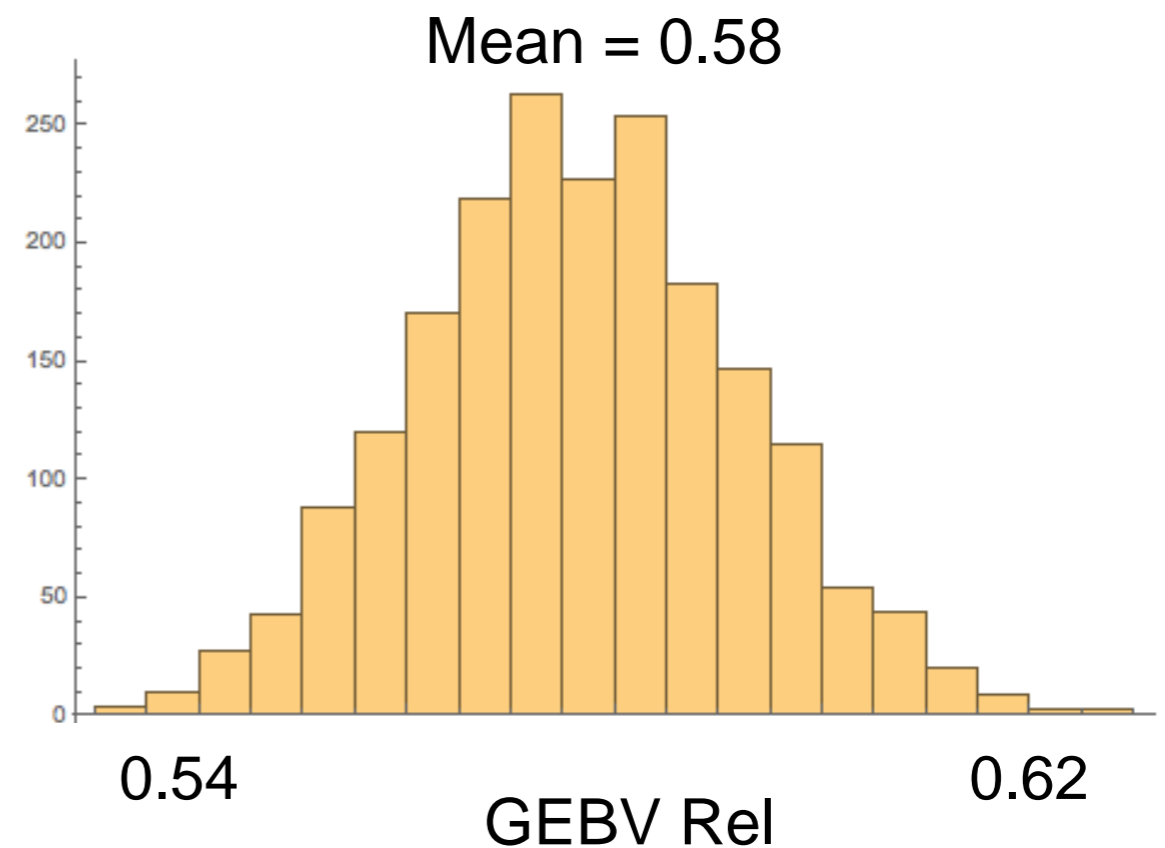
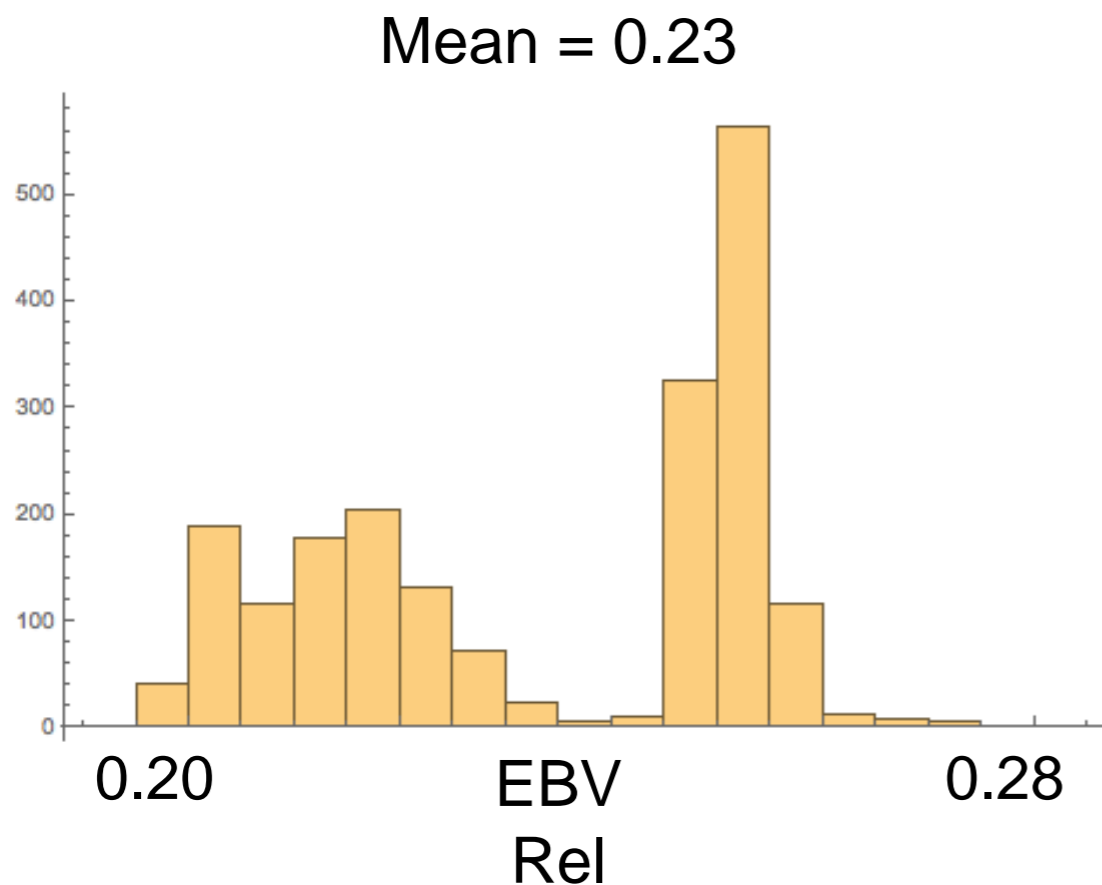
# Simulation

- No selection dataset
- 3 Generations of test sires (1000 sires per generation)
- $R^2$  between TBV and both the BLUP EBVs and GEBVs

Generation	EBV	GEBV
1	0.19	0.51
2	0.07	0.43
3	0.04	0.40

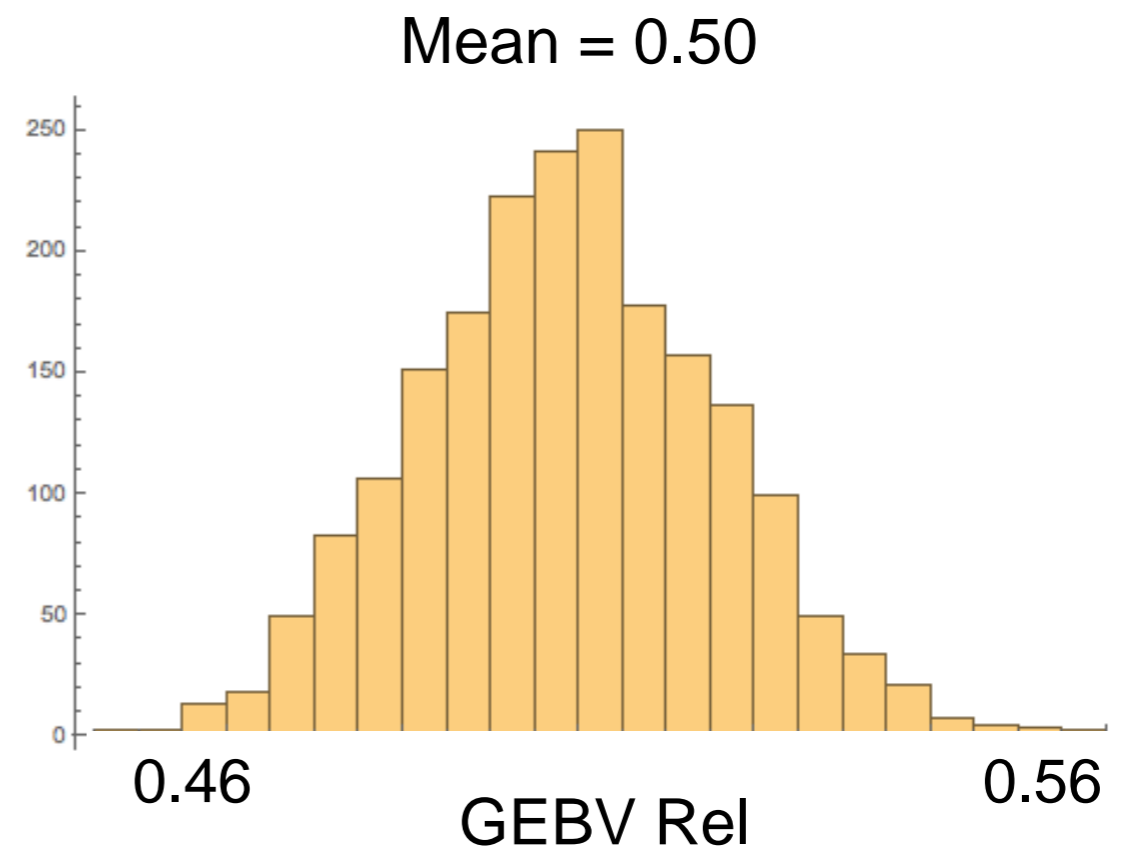
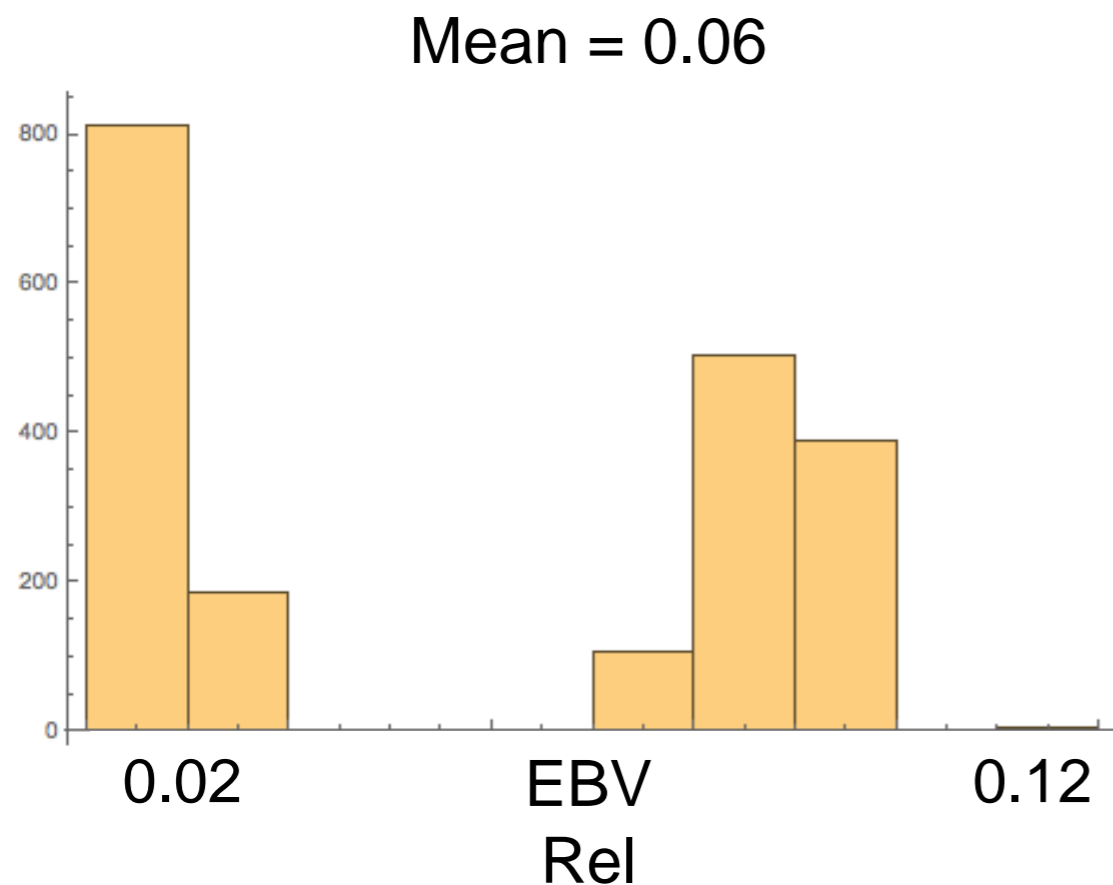
# Simulation

- No selection dataset
- Generation 1 of test sires
- Comparison of direct reliabilities from BLUP EBVs and GEBVs



# Simulation

- No selection dataset
- Generation 3 of test sires
- Comparison of direct reliabilities from BLUP EBVs and GEBVs



# Simulation

- No selection dataset
- 3 Generations of test sires (1000 sires per generation)
- $R^2$  between TBV and both the BLUP EBVs and GEBVs

Generation	$R^2$ EBV	Mean Rel EBV	$R^2$ GEBV	Mean Rel GEBV
1	0.19	0.23	0.51	0.58
2	0.07	0.11	0.43	0.52
3	0.04	0.06	0.40	0.50

# Simulation

- Selection dataset  
Validation Regression

Generation	Bayes A 20%A	GBLUP 20%A	GBLUP 0%A	EBV
1	0.95	0.96	0.87	0.68
2	0.93	0.94	0.82	0.67
3	0.90	0.92	0.79	-0.13

Validation  $R^2$

Generation	Bayes A 20%A	GBLUP 20%A	GBLUP 0%A	EBV
1	0.30	0.30	0.32	0.07
2	0.25	0.26	0.27	0.03
3	0.21	0.22	0.23	0.00

# Simulation

- Selection dataset
  - The highest R-square was with no polygenic variance, but this analysis fails the regression test
  - The lack of daughter proofs for low GEBV bulls can cause low squared correlations
  - An adjustment may be required from the reduced R-square to get the correct PEV
  - The Interbull genomic validation group may have to derive formulas to adjust reliabilities for genomic pre-selection

# Simulation

- **With Selection**

ValidationI = R<sup>2</sup> between TBV and (G)EBV

Model = ASREML PEV using assumed variance

	Reliabilities			
	ValidationI		Model	
	BLUP	GBLUP	BLUP	GBLUP
Animals				
Tr. gen. 1	0.857	0.861	0.949	0.953
Tr. gen. 2	0.841	0.849	0.949	0.953
Val. gen. 1	0.126	0.472	0.280	0.633
Val. gen. 2	0.048	0.398	0.138	0.565
Val. gen. 3	0.000	0.349	0.066	0.536



# Simulation

- **With Selection**

ValidationI = R<sup>2</sup> between TBV and (G)EBV

Model = ASREML PEV using estimated variances from the data

	Reliabilities			
	ValidationI		Model	
	REML	GREML	REML	GREML
Animals				
Tr. gen. 1	0.818	0.814	0.538	0.758
Tr. gen. 2	0.788	0.806	0.535	0.756
Val. gen. 1	0.110	0.431	0.158	0.481
Val. gen. 2	0.045	0.362	0.078	0.423
Val. gen. 3	0.000	0.307	0.038	0.398

# Future Work

- Finish validating simulation model
  - Produce multiple replicates from the simulation model
  - Extend simulation model to produce data sets useful single step models
- Provide data-sets to members to test their own genomic reliability software
- Provide recommendations on calculation methods
- Provide recommendations on detecting genomic reliability over-estimation – relative validation  $R^2$