Implementing genomic selection in livestock species

Julius van der Werf

CRC for Sheep Industry Innovation
School of Environmental and Rural Science, UNE, Armidale, NSW

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

SHEEP CRC

CRC AUSTRALIA

Outline

1. Potential benefits of genomic selection in breeding programs

2. Can we predict the accuracy of genomic selection?

3. What information is needed for accurate predictions?

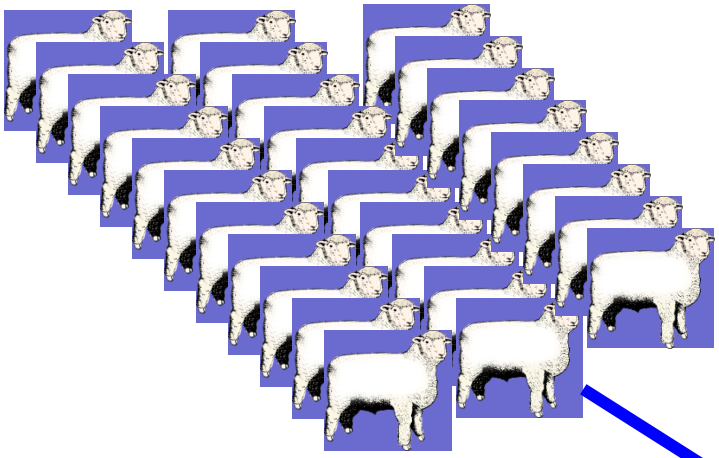
4. Requirements for the reference population

   how large, how related, how long-lasting, multi-breed?

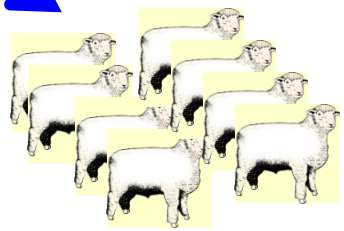
5. Strategies for genotyping

   low density chips, high density chips, sequence data?

Genomic Prediction: basic idea



1) Somebody (else) measures lots of sheep, and their DNA
→ Reference population



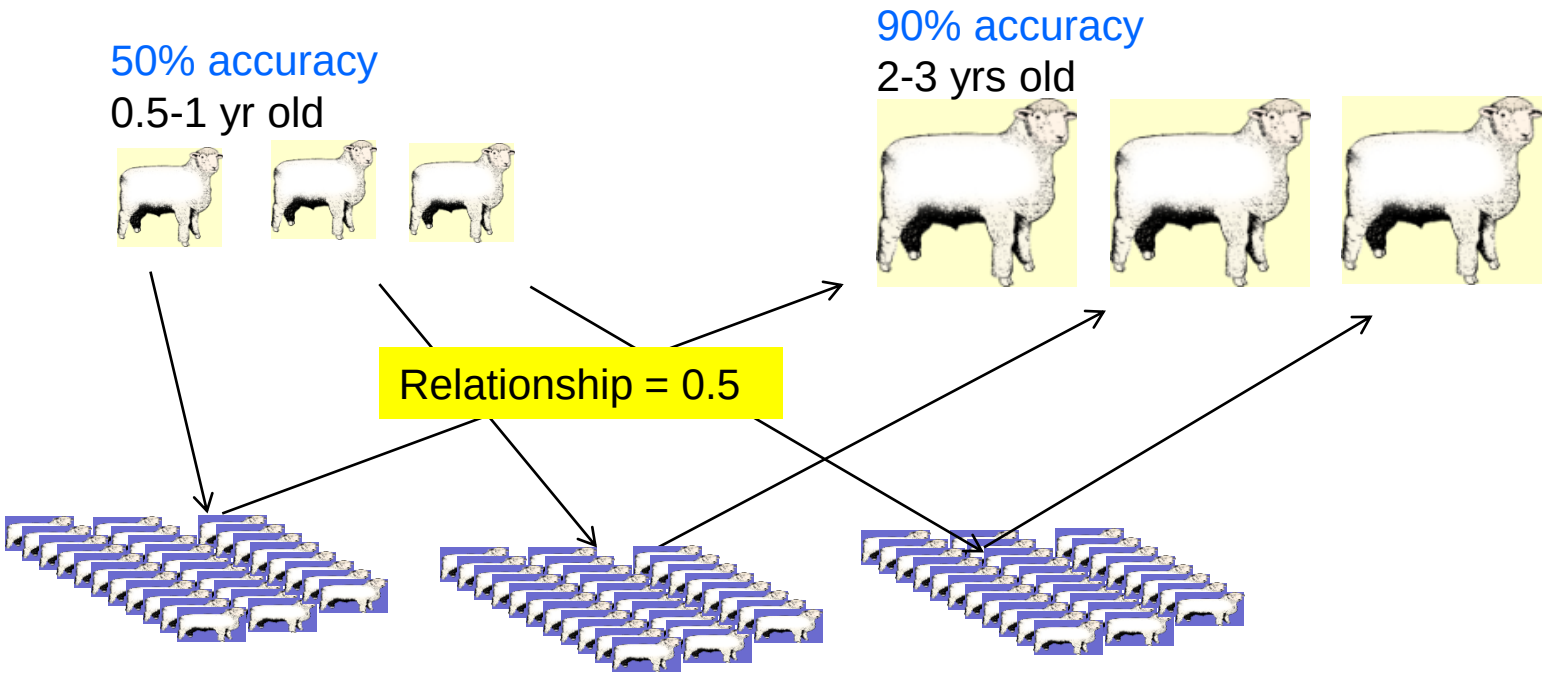
2) A breeder tests DNA on young rams

Prediction from DNA → genomic breeding values - GBV

GBV + Current ASBV → Improved ASBV

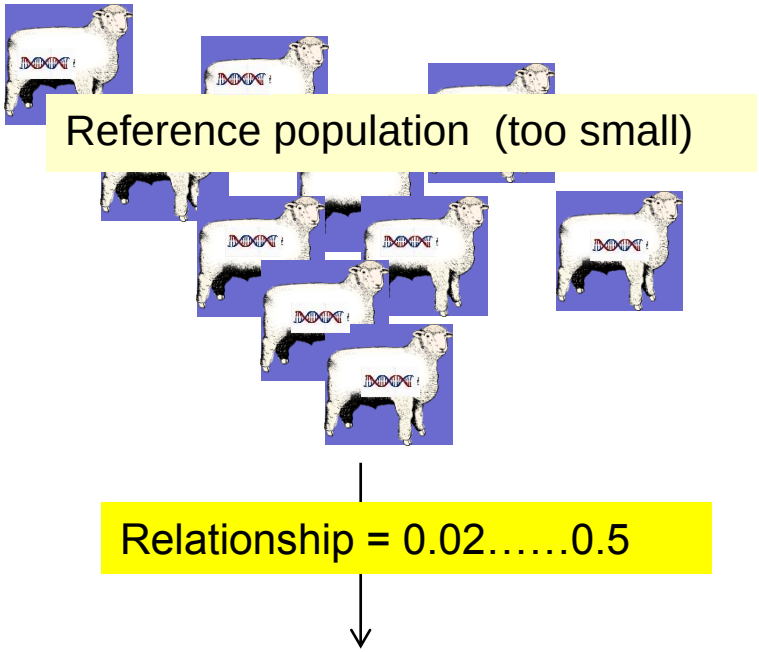
Merit depends on trait measurability

Compare: Progeny Testing



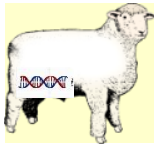
Each progeny group only informs one sire

Genomic Testing

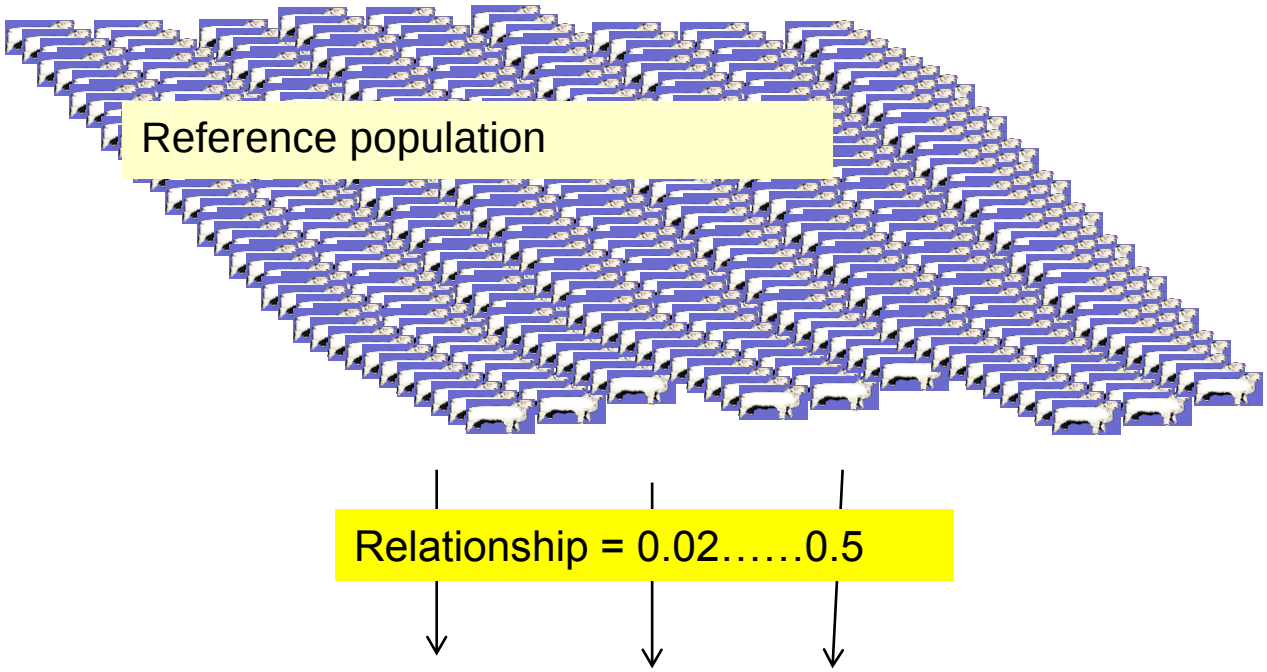


use information on “relatives” while sire is still young

51% accuracy
0.5-1 yrs old

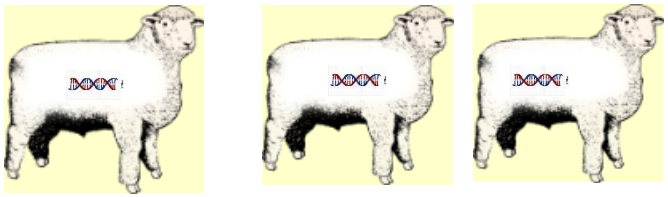


Genomic Testing



70% accuracy

0.5-1 yrs old



Summarizing Genomic Prediction

- What information is used?

- Based on very many small – genomic- relationships
- Does not require ‘direct relatives’ to be tested
- Can be based on distant relatives ‘some generations away’
- …..but the number of small relatives needs to be large (thousands)
- Can not predict across breed

Outline: Sheep Genomic Analysis

- What information is used?
- How useful is this information?
- How to use it?

Genomic Selection: Benefit

Overall:

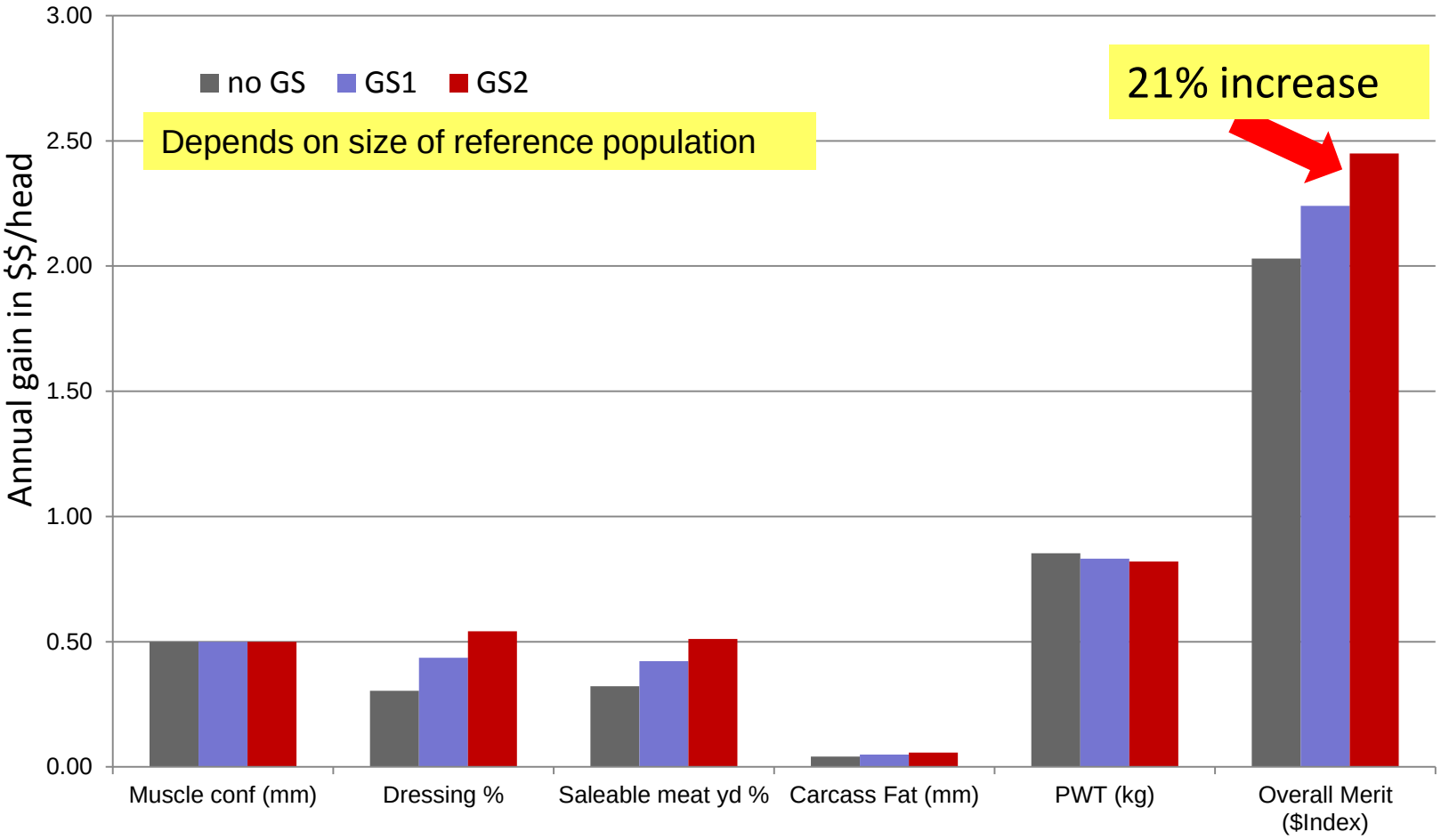
More accurate prediction of genetic merit for breeding objective

Specific:

Traits that are usually difficult to improve

- difficult or expensive to measure
- can not be measured early
- low heritability

Possible Benefits



Modeling genomic selection in breeding programs

1. Selection index approach: multiple information, multiple traits

   Accuracy component

2. Optimizing selection across age classes

   Generation Interval component

3. For specific breeding objectives

Percent increase in rate of genetic gain when using genomic selection

Selection on a single trait

Predicted accuracy of Molecular EBV = 55% *(VQTL=30%)*

Trait Measurability	Heritability	
	0.10	0.50
Measured < 1 year, males and female	37	6
Measured > 1 year, males and females	64	18
Measured >1 year, females only	109	39
Measured on Correlated Trait, Genetic Correlation = 0.9	48	11
Measured on Correlated Trait, Genetic Correlation = 0.5	143	62

Outline

1. Potential benefits of genomic selection in breeding programs

2. Can we predict the accuracy of genomic selection?

3. What information is needed for accurate predictions?

4. Requirements for the reference population

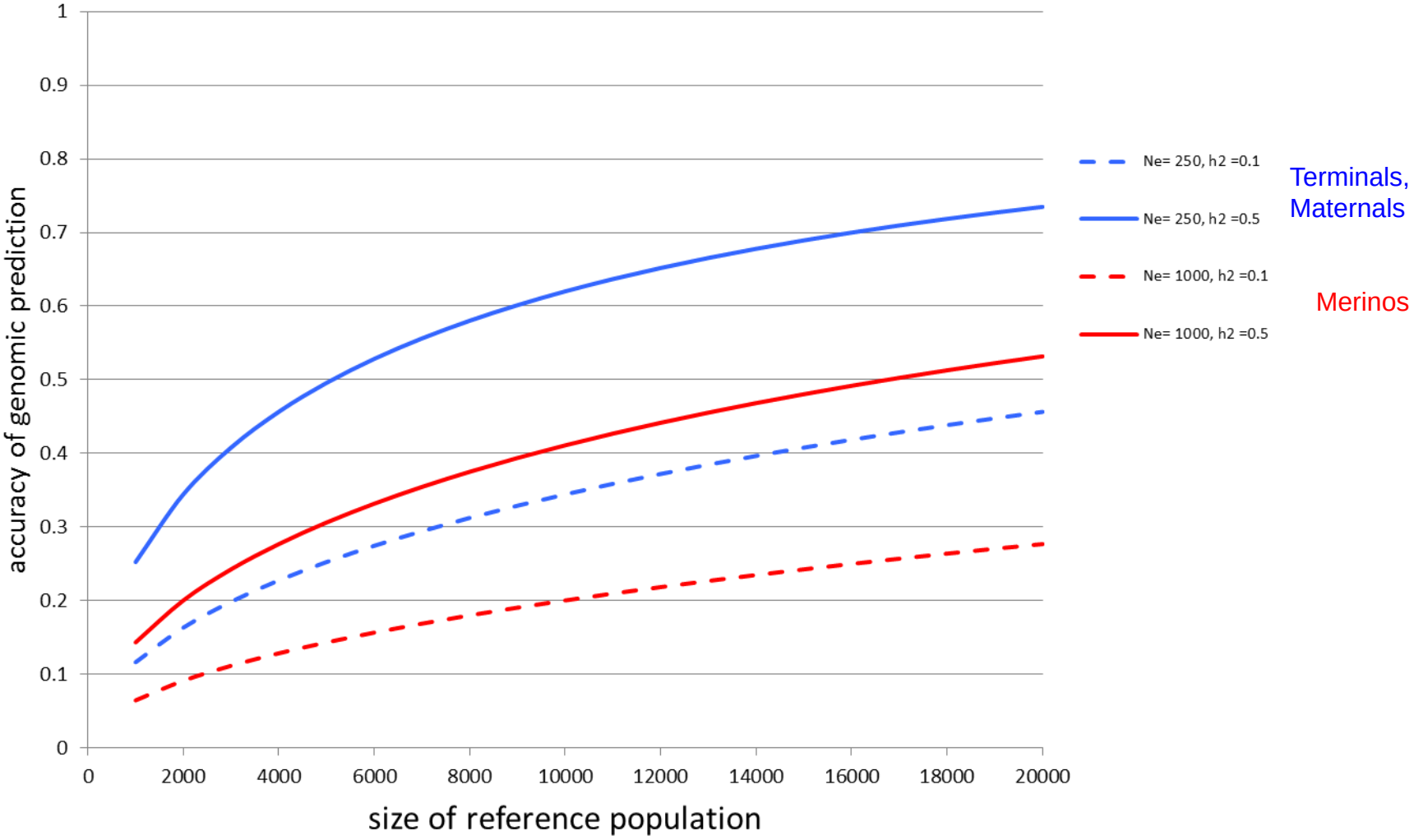
   how large, how related, how long-lasting, multi-breed?

5. Strategies for genotyping

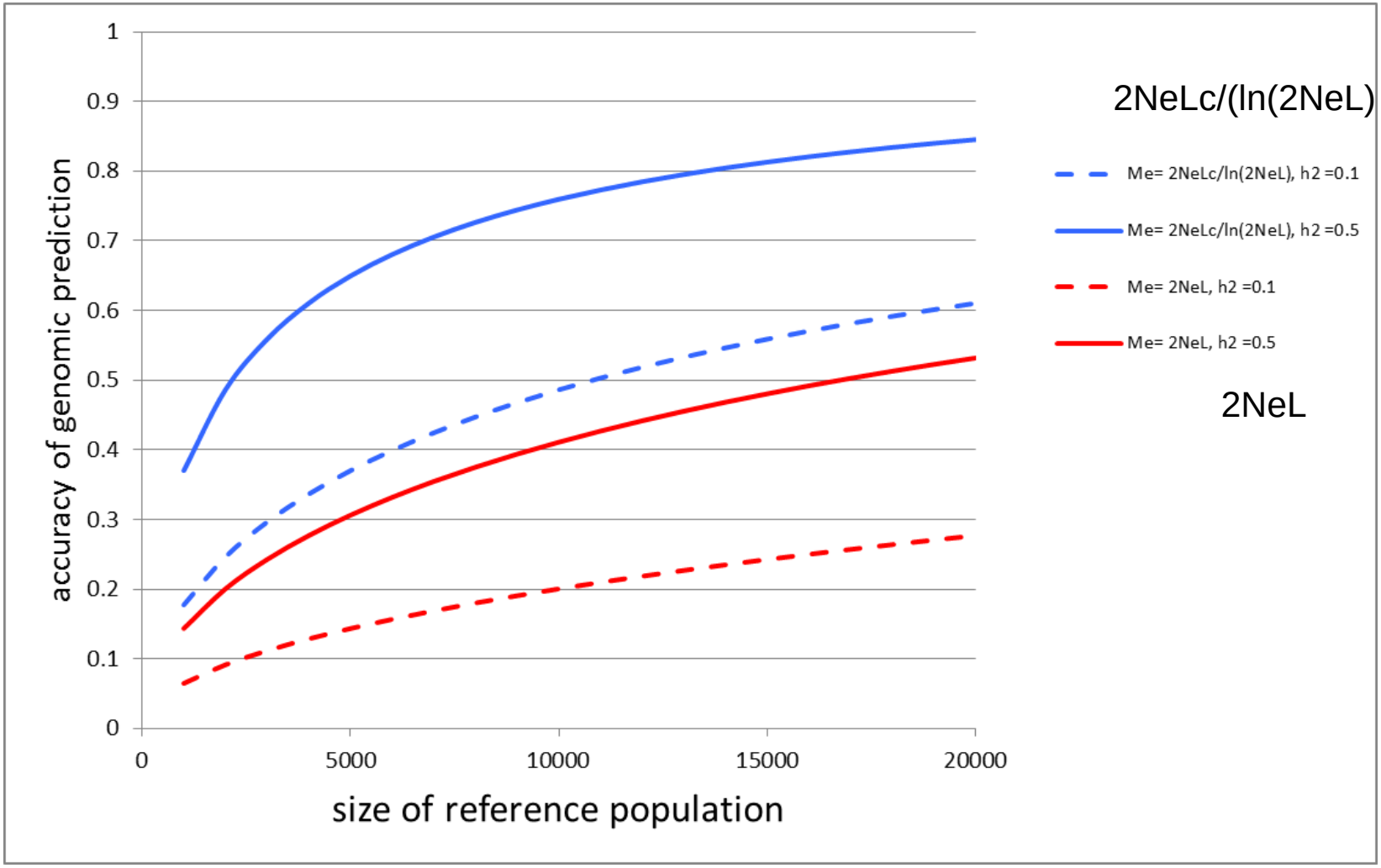
   low density chips, high density chips, sequence data?

Accuracy of genomic prediction depending on size of reference population

Goddard 2009



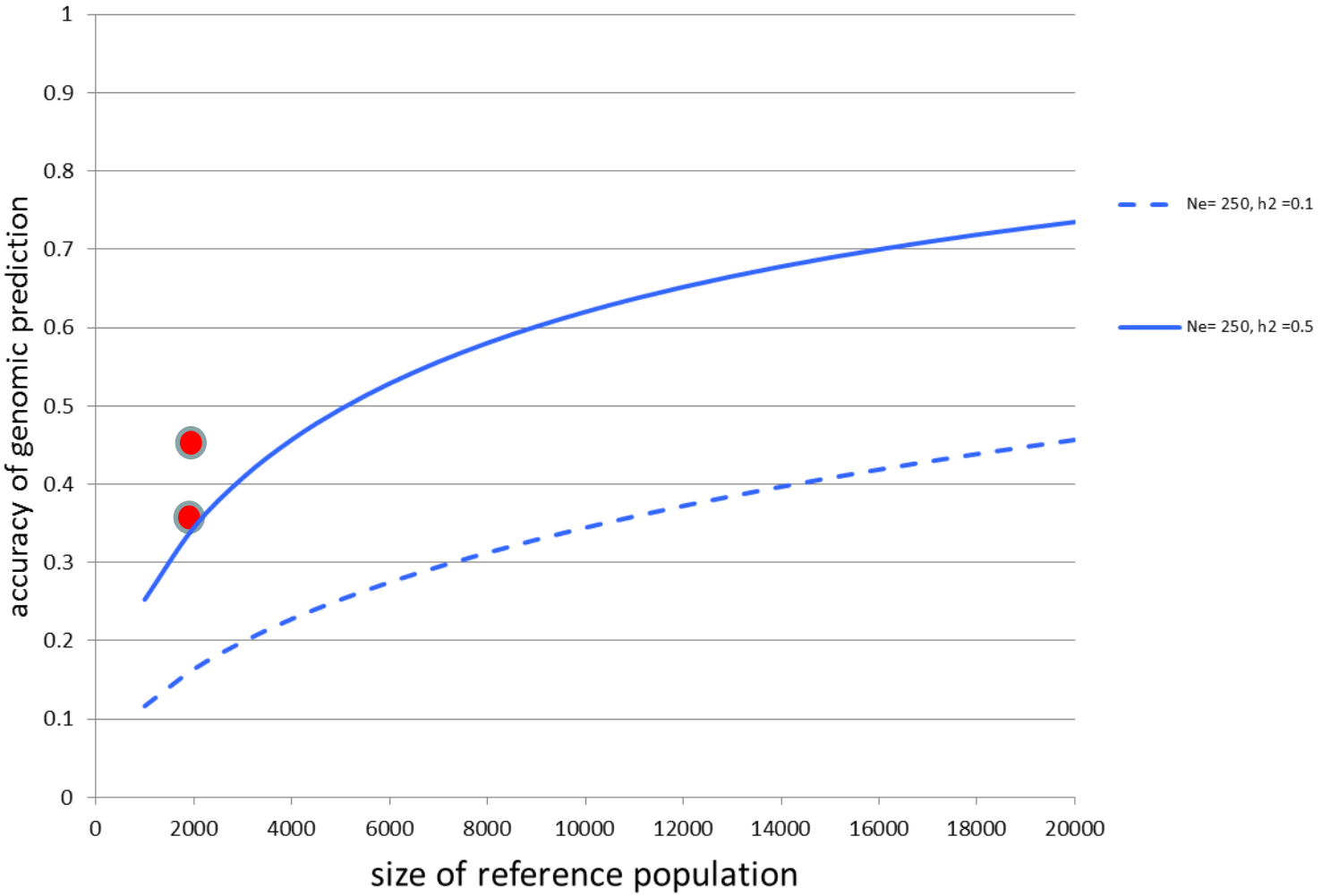
Accuracy, depending on how Me is approximated



design of reference population

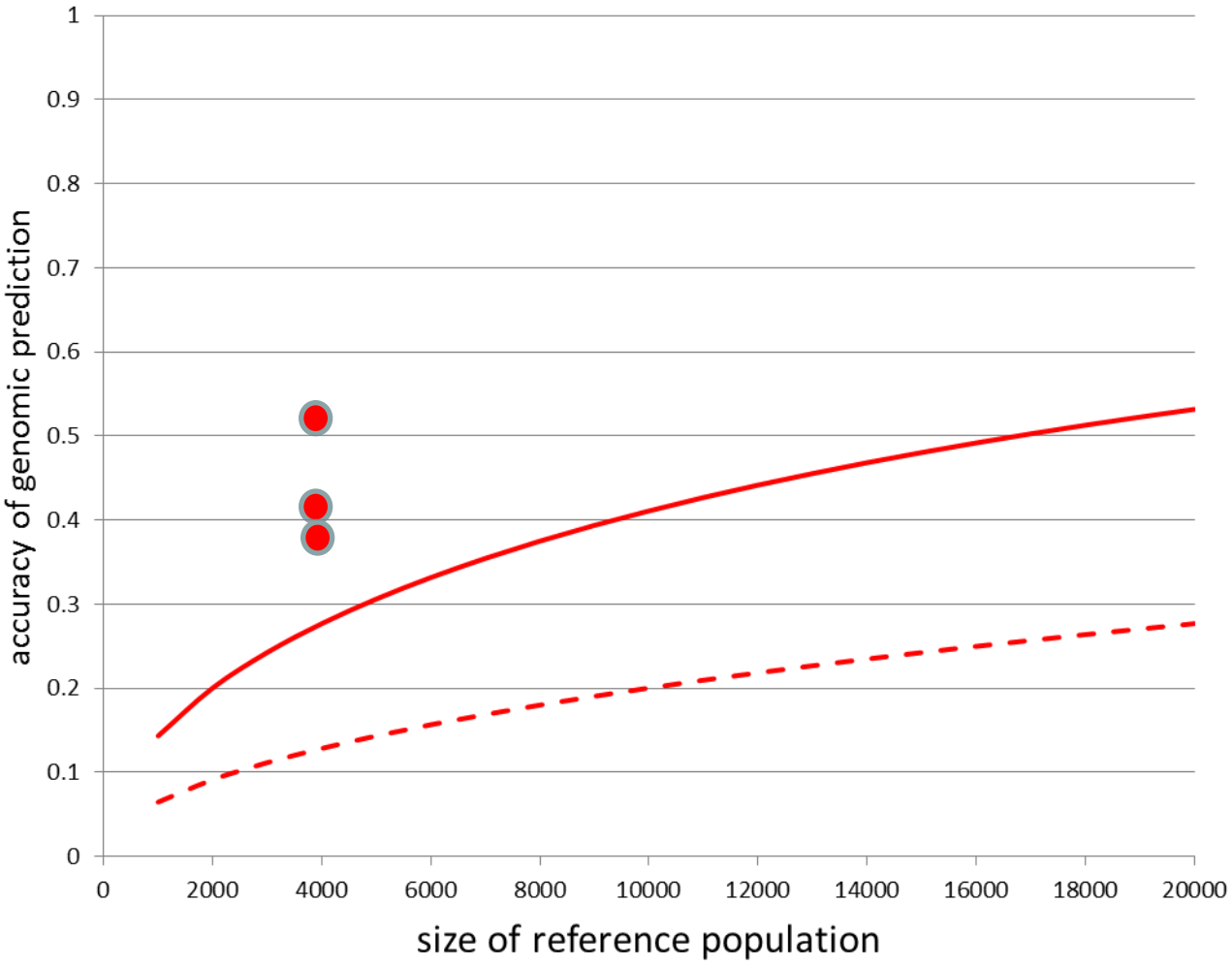
- Relatedness between reference population and selection candidates
- Across breeds or lines?
- Number of sires, nr of progeny per sire, which dams?

Realized accuracy 1



Terminals,
Maternals

Realized accuracy 2



Merino

Accuracy of genomic prediction for Post Weaning Weight from a mixed breed reference population

Reference population		GEBV accuracy	
		G1	
Type	**Size**	**BL**	**Merino**
(1) = Merino	1000	-0.02 [a]	0.53 [b]
(2) = Merino	2000	-0.04 [a]	0.57 [bc]
(3) = Merino	3000	-0.08 [a]	0.59 [c]
BLxMerino	1514	0.49 [c]	0.45 [a]
BLxMerino + (1)	2514	0.42 [bc]	0.56 [bc]
BLxMerino + (2)	3514	0.37 [b]	0.54 [bc]
BLxMerino + (3)	4514	0.36 [b]	0.56 [bc]

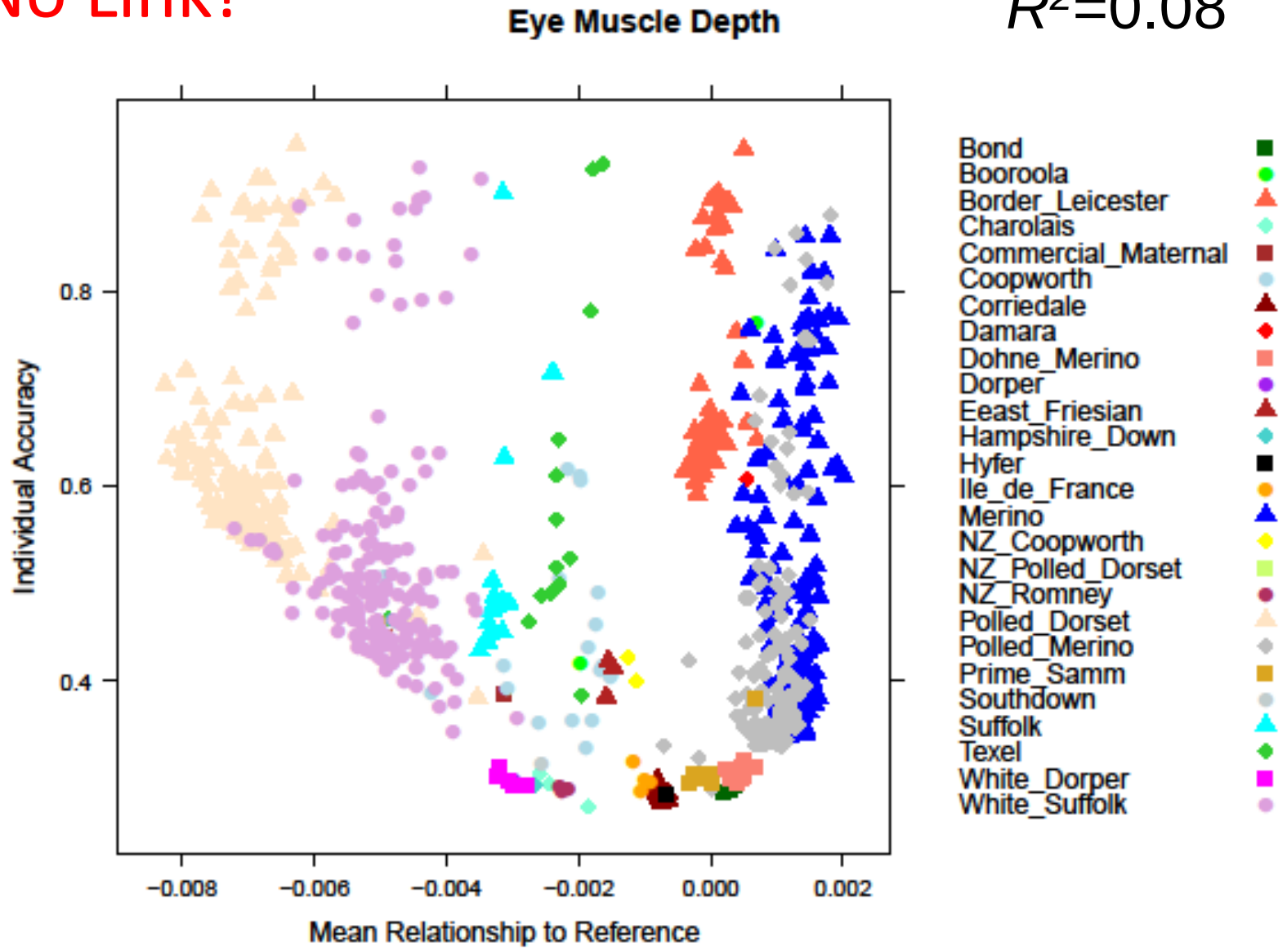
Moghaddar, van der Werf and Swan, 2013 AAABG)

une

Accuracy and Mean Relationship to Ref

→ No Link!

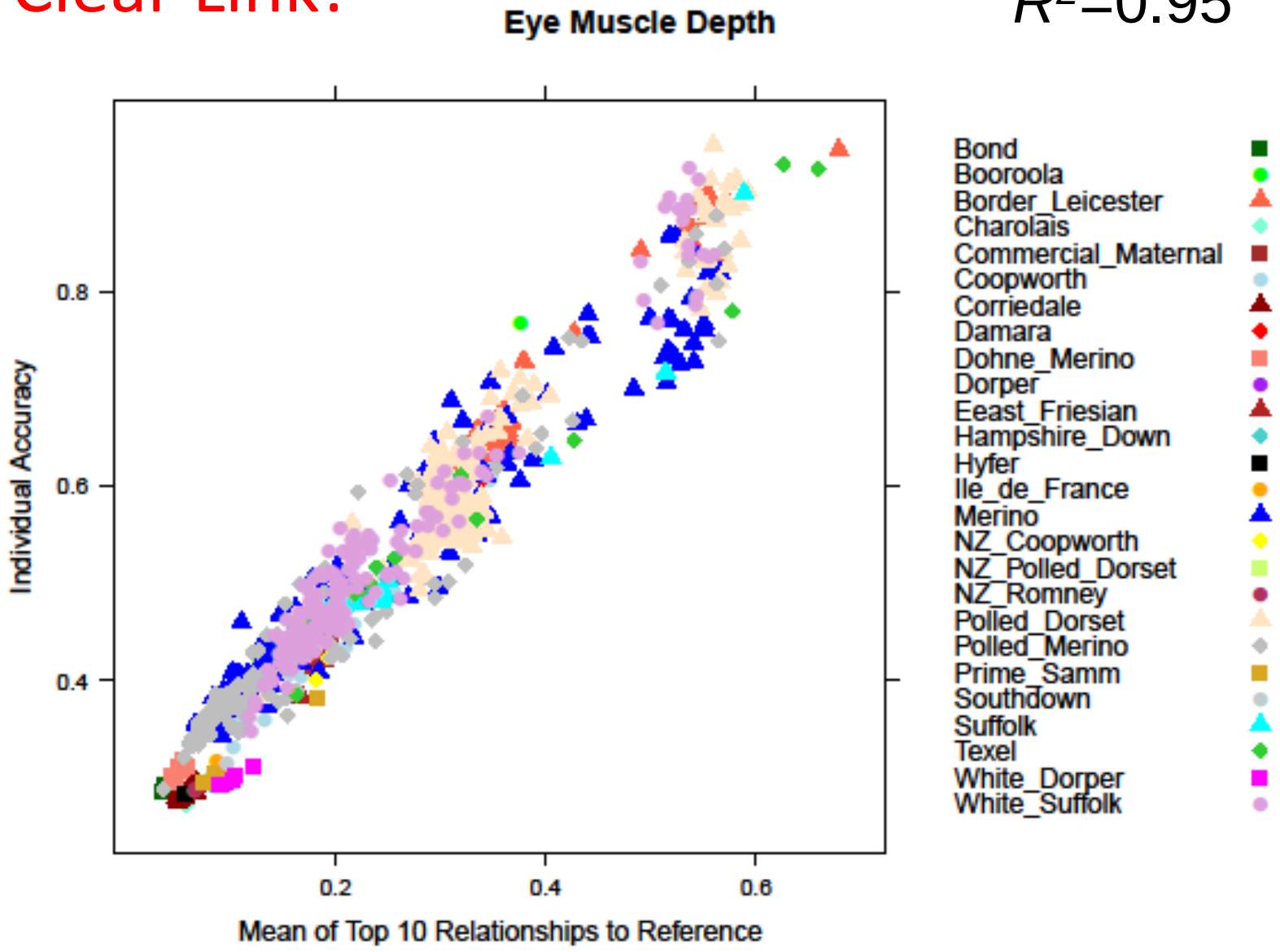
R^2=0.08



Accuracy and Mean of Top 10 Relationships

→ Clear Link!

R^2=0.95



Genomic prediction

$$\begin{bmatrix} X'X & X'X & 0 \\ Z'X & Z'Z + G^{11} & G^{12} \\ 0 & G^{21} & G^{22} \end{bmatrix} \begin{bmatrix} b \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ 0 \end{bmatrix}$$

$\hat{g}_2 = -(G^{22})^{-1} G^{21} \hat{g}_1$

Genomic regression

Example:
Data on sire 1, sons 2 and 3, 4 unrelated, want to predict 5

A-matrix (pedigree-based)

1	0.5	0.5	0	0.5
0.5	1	0.25	0	0.25
0.5	0.25	1	0	0.25
0	0	0	1	0
0.5	0.25	0.25	0	1

G-matrix (DNA-based)

1	0.5	0.5	0.02	0.5
0.5	1	0.20	0.015	0.20
0.5	0.20	1	0.025	0.30
0.02	0.015	0.025	1	0.025
0.5	0.20	0.30	0.025	1

BLUP

$\hat{u}_5 = 0.1136.y_1 + 0.0455.y_2 + 0.0455.y_3$

GBLUP

$\hat{g}_5 = 0.1135.y_1 + 0.0328.y_2 + 0.0591.y_3 + 0.00519.y_4$

Genomic prediction

$$\begin{bmatrix} X'X & X'X & 0 \\ Z'X & Z'Z + G^{11} & G^{12} \\ 0 & G^{21} & G^{22} \end{bmatrix} \begin{bmatrix} b \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ 0 \end{bmatrix}$$

$\hat{g}_2 = -(G^{22})^{-1} G^{21} \hat{g}_1$

Genomic regression

Example:
Data on sire 1, sons 2 and 3, 4 unrelated, want to predict 5

A-matrix (pedigree-based)

1	0.5	0.5	0	0.5
0.5	1	0.25	0	0.25
0.5	0.25	1	0	0.25
0	0	0	1	0
0.5	0.25	0.25	0	1

G-matrix (DNA-based)

1	0.5	0.5	0.02	0.5
0.5	1	0.20	0.015	0.20
0.5	0.20	1	0.025	0.30
0.02	0.015	0.025	1	0.025
0.5	0.20	0.30	0.025	1

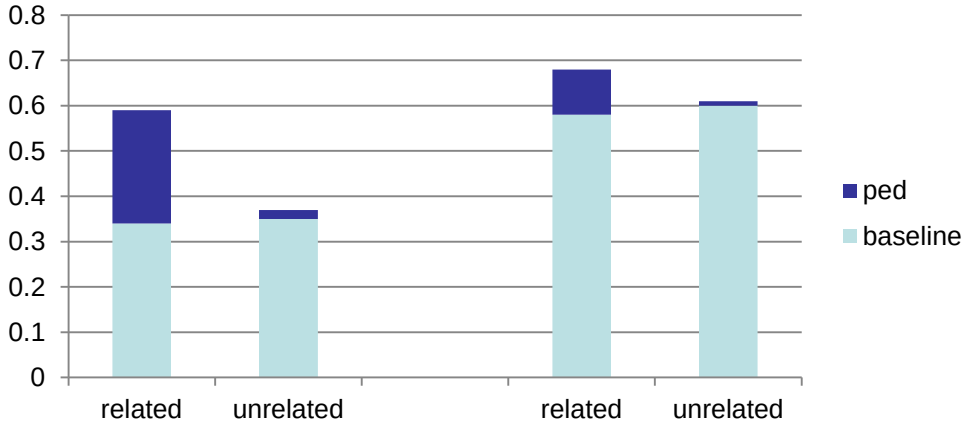
BLUP uses: Family Info

GBLUP uses: Family Info
Segregation within family
Info on 'unrelated'

Sources of information contributing to GBV accuracy

	BLUP	GBLUP	half life
1. Variation between families	++	++	1 gen
2. Variation within families	0	+	1 gen
3. Markers tracking effects of genome segments/LD	0	++.+	several gen's

Info on 'unrelated'



smaller ref pop

larger ref pop

Depending on size of reference population

Results – Simulation

Sam Clark

Method	**Close** Ped 0 - 0.25 Genom 0.08 – 0.35	**Distant** 0 - 0.125 0.08 – 0.26	**Unrelated** 0 - 0.05 0.08 – 0.16
BLUP- Shallow pedigree	0.39	0.00	0.00
BLUP- Deep Pedigree	0.42	0.21	0.04
gBLUP	**0.57**	0.41	0.34

Additional accuracy from family info

'baseline accuracy': graphs predict 0.36 for Ne=100, N=1750, h^2=0.3

Accuracy Real Data (INF) Sam Clark

	Close related sires		Distantly related sires	
Method	Empirical Acc actual correlation with ASBV	Predicted Acc correlation derived from gBLUP	Empirical Acc	Predicted Acc
BLUP-S	?	?	0.00	0.00
BLUP-D	0.62	0.37	0.02	0.05
gBLUP	0.65	0.41	0.27	0.19

Genomic prediction FAQ

- How well can we predict distantly related individuals?
  - Ok if reference population is large enough
  - Can NOT predict across breed *Daetwyler et al., 2011*
- How quick does the genomic prediction erode?
  - Fast if based on relationships, slower if based on 'distant relatives
- Do we need relatives?
  - Relatives give more accuracy, but not everyone can have them
- How large does a reference population need to be?
  - Design based as if prediction is based on 'unrelated'

Reference Pop: How many are needed?

$\%V_A$ explained by GBV

Breed	merino	WS, PD	BL
Ne	1000	250	100
Size of reference pop'n	30,000	10,000	5,000
Progeny measured per year[1]	3750	1250	625
h2=0.1	0.33	0.34	0.35
h2=0.3	0.51	0.53	0.54
h2=0.5	0.60	0.62	0.63
Predicted benefit in dG	40%	20%	?

$\cong h^2$

assuming the reference population is 'refreshed' every 8 years

Reference Pop: How many are needed?

$\%V_A$ explained by GBV

Breed	merino	WS, PD	BL
Size of reference pop'n	12,000	4,000	2,000
Progeny measured per year[1]	1500	500	250
h2=0.1	0.22	0.23	0.23
h2=0.3	0.36	0.37	0.38
h2=0.5	0.44	0.46	0.47
Predicted benefit in dG	20%	10%	?

$\cong \frac{1}{2} h^2$

assuming the reference population is 'refreshed' every 8 years

Outline

1. Potential benefits of genomic selection in breeding programs

2. Can we predict the accuracy of genomic selection?

3. What information is needed for accurate predictions?

4. Requirements for the reference population

   how large, how related, how long-lasting, multi-breed?

5. Strategies for genotyping

   low density chips, high density chips, sequence data?

Implication

- To predict a selection candidate
  - It needs to have relatives in reference populations
  - We can afford a lower degree of relationship than with BLUP
    - » Can predict several generations away
  - Need large reference population

Optimal Genotyping Strategies

- If genotyping is expensive
  - Genotype males only
  - Genotype only ‘best’ males
  - multi-stage selection
  - But enough to be able to select!

Acknowledgements

DPI Vic:	Hans Daetwyler, Ben Hayes
UNE:	Nasir Moghaddar, Sam Clark, John Hickey, Brian Kinghorn, Cedric Gondro
AGBU:	Andrew Swan, Daniel Brown
Sheep CRC:	Ken Geenty, Klint Gore, James Rowe
Sheep Genetics/MLA:	Rob Banks, Alex Ball, Sam Gill

Thanks