

# Meta-analysis of GWAS to estimate SNP effects and breeding values

Michael Goddard  
University of Melbourne

# Introduction

What is a meta-analysis?

Why do we need it?

How good is it?

Examples

Proposal for Interbull

# What is a meta-analysis

Combining the results from  $>1$  analysis rather than combining the raw data

## Why do we need it?

To increase power

To increase robustness

Cant combine raw data

# How good is it?

Very widely used to combine information from medical trials

Some information is lost

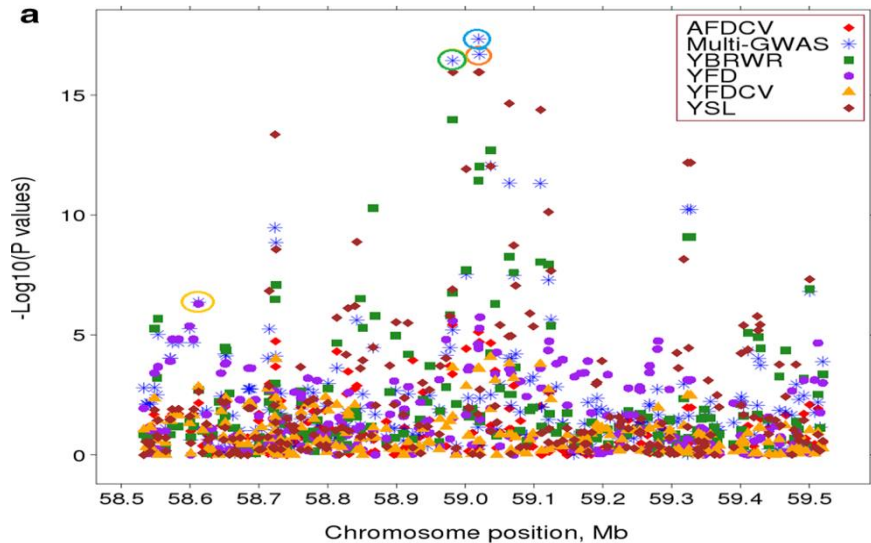
Covariances between estimates

OLS vs GLS

# Example 1- Multi-trait GWAS

Bolormaa et al(2017)

Combine single trait GWASs on correlated traits



**c**  
Variance of local GEBV or Eigenvalue of PC1

# Example 2- Multi-country GWAS of Stature

Bouwman et al (2018)

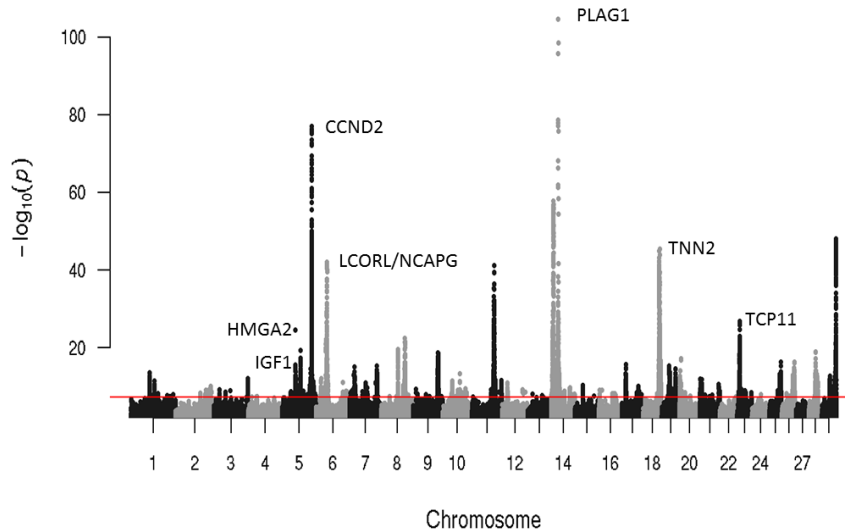
Combine GWASs on stature from different breeds and countries

58,000 bulls from 17 countries-breeds

163 lead variants with  $p < 5 \times 10^{-8}$

Explaining 13% of variance

# Example 2- Multi-country GWAS of Stature



# Example 3- LD score regression in humans

Uses summary data ie chi-square from GWAS and LD  $r^2$  to estimate  $h^2$  (Bulik-Sullivan et al 2015)

Can be used to estimate  $r_g$

Assess the importance of prior information (Finucane et al 2015)  
eg H3K4me3 sites enriched 2 fold for effect on 27 disease traits

Less accurate than raw data?



# Example 4- Joint and conditional GWAS in humans

Yang et al (2012)

Simultaneously estimating the effect of many SNPs on a trait

Least squares equation for multiple regression

$$X'X b = X'y$$

b from published data

$X'X$  from reference sample

$X'Xb \rightarrow X'y$  for sample

Bayesian multiple regression (Zhu and Stephens 2016)

# Example 5- Combining eQTL and GWAS summary statistics

Zhu et al 2017

Regression of expression on phenotype for individual SNP associations

# Opportunities in cattle

Many countries and breeds have genotypes and phenotypes but raw data cannot be combined.

## Opportunities in cattle

Sale of semen increasing based on genomic EBVs

We want them to be as accurate as possible

(for all traits in all breeds and countries and between breeds)

i.e. We want estimated SNP effects to be as accurate as possible

High accuracy  $\leftarrow$  high N, non-linear estimation, one-step, sequence data, functional information

# Opportunities in cattle

## High N

### Human genetics

meta-analysis of 270,000 people for height → more SNPs,  
increased accuracy

in UK 500,000 people with WGS

in USA 1M people with WGS

### Dairy cattle

1,000,000s world wide if we collaborate  
not within-breed, within-country for all traits

# Interbull SNP Mace

Interbull combines progeny test EBVs from different countries

→ more accurate EBVs which are comparable regardless of country of origin

Selection of bulls now largely on genomic EBVs

Lose information if you combine GEBVs from different countries

Better to combine SNP solutions

# Interbull SNP Mace

Lose information if you combine GEBVs from different countries

## Options

Convert GEBVs from country A to country B as for progeny test EBVs

GEBVs are regressed back by  $r_g$

Put genotype from country A through prediction equation of country B

Limited accuracy due to size of reference population in country B

Combine estimates of SNP effects, allowing for  $r_g$ , to get most accurate estimate of SNP effects in country B

# Interbull SNP Mace

Single country equations to estimate SNP effects (g)

$$(Z_1'Z_1 + \lambda I) g_1 = Z_1'y_1$$

Two countries

$$(Z_1'Z_1 + Z_2'Z_2 + \lambda I) g = Z_1'y_1 + Z_2'y_2$$

If individual countries provide  $g_i$  and  $Z_i'Z_i$  we can construct the multi-country BLUP and solve for g

Extensions: include  $r_g < 1$  between countries and weights for records



## Extensions for SNP-MACE model

$$\begin{bmatrix} \mathbf{Z}'_1 \mathbf{Z}_1 + \mathbf{G}^1 & \boldsymbol{\Psi}_{12} + \mathbf{G}^{12} \\ \boldsymbol{\Psi}_{12} + \mathbf{G}^{12} & \mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{G}^2 \end{bmatrix} \times \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1 \mathbf{y}_1 \\ \mathbf{Z}'_2 \mathbf{y}_2 \end{bmatrix}$$

Common daughter information

$$\mathbf{B} = \frac{1}{\sum_j 2p_j(1-p_j)} \mathbf{I} \quad (\text{VanRaden 2008})$$

$$\mathbf{G} = \text{var} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_1^2 \mathbf{B}_1 & r_{12} \sqrt{\sigma_1^2 \mathbf{B}_1 \times \sigma_2^2 \mathbf{B}_2} \\ r_{12} \sqrt{\sigma_1^2 \mathbf{B}_1 \times \sigma_2^2 \mathbf{B}_2} & \sigma_2^2 \mathbf{B}_2 \end{bmatrix}^{-1}$$

## A general SNPMace model

$$\begin{bmatrix} \ddots & & & & & & \\ \vdots & \dots & \dots & \dots & \dots & & \\ \vdots & [\mathbf{Z}'_i \mathbf{Z}_i + \mathbf{G}^{ii}] & \dots & [\boldsymbol{\Psi}_{ii^+} + \mathbf{G}^{ii^+}] & \dots & & \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \\ \vdots & [\boldsymbol{\Psi}_{ii^+} + \mathbf{G}^{ii^+}] & \dots & [\mathbf{Z}'_{i^+} \mathbf{Z}_{i^+} + \mathbf{G}^{i^+i^+}] & \dots & & \\ \vdots & \dots & \dots & \dots & \dots & & \\ \ddots & & & & & & \end{bmatrix} \times \begin{bmatrix} \dots \\ [\hat{\mathbf{g}}_i] \\ \vdots \\ [\hat{\mathbf{g}}_{i^+}] \\ \dots \end{bmatrix} = \begin{bmatrix} \dots \\ [\mathbf{Z}'_i \mathbf{y}_i] \\ \vdots \\ [\mathbf{Z}'_{i^+} \mathbf{y}_{i^+}] \\ \dots \end{bmatrix}$$

# Interbull SNP Mace Project

2018-2019

We will write software to do SNP Mace and deliver it to Interbull

We will test method on Brown Swiss

Interbull have individual records for all countries and can calculate GEBV

We will use SNP Mace to calculate SNP effects and hence GEBVs

Compare the two sets of GEBVs

Based on 50k SNP genotypes

# Interbull Data

- $Z'R^{-1}Z$  matrices were calculated for six countries:

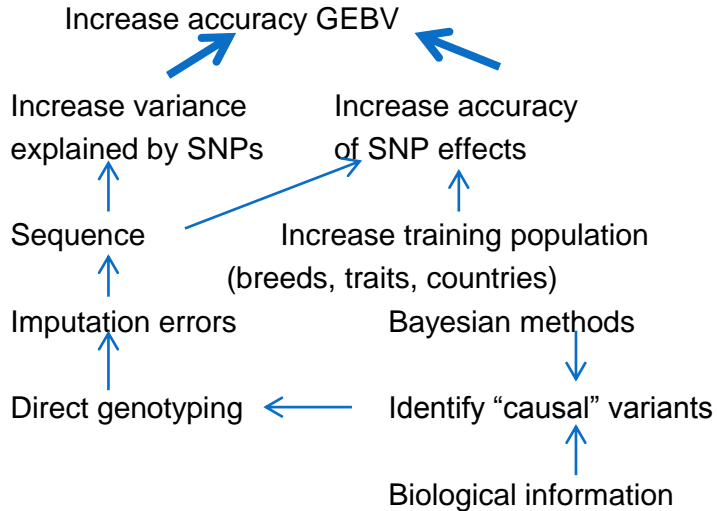
<b>Country</b>	<b>No bulls</b>
CHE	1922
DEA	2578
FRA	171
ITA	1418
SVN	227
USA	796

# Interbull SNP Mace Project

Possible extension of project

Use all sequence variants and Bayesian method instead of BLUP.

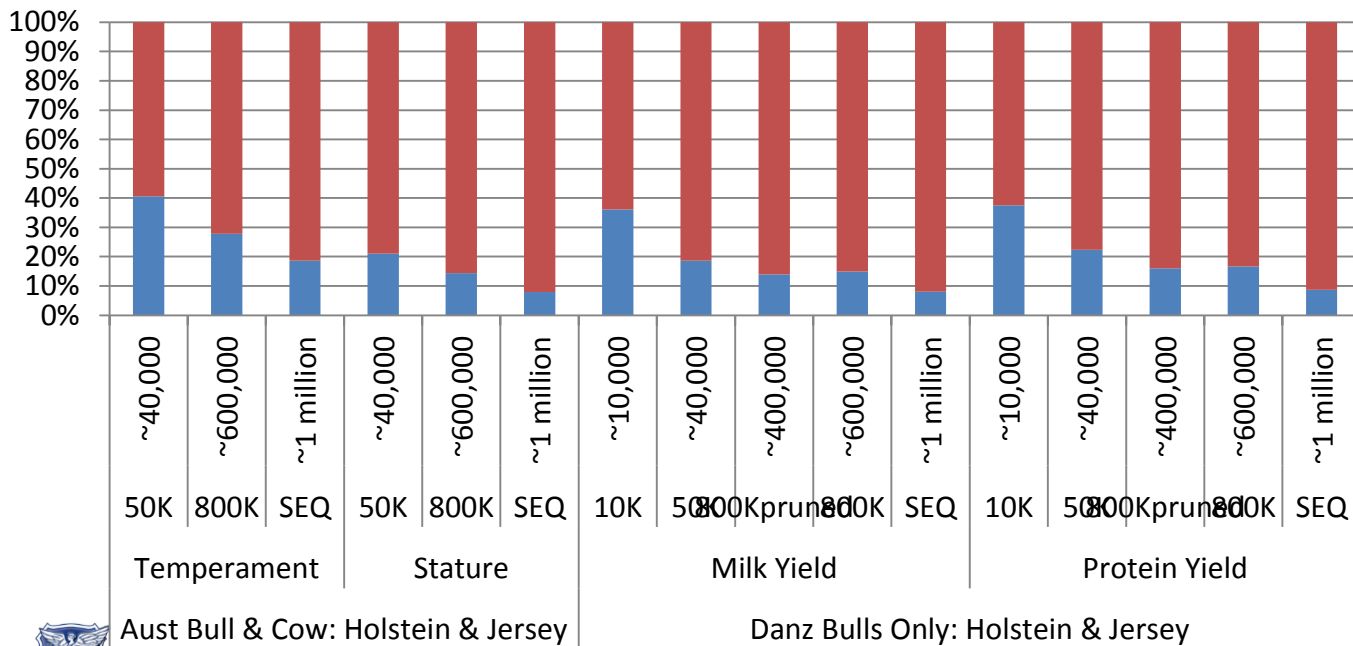
## Medium term



# Variance explained by SNPs and sequence (Iona Macleod)

## Proportion of Total Genetic Variance Explained by SNP and Pedigree: BayesR (Mixed Hol & Jer)

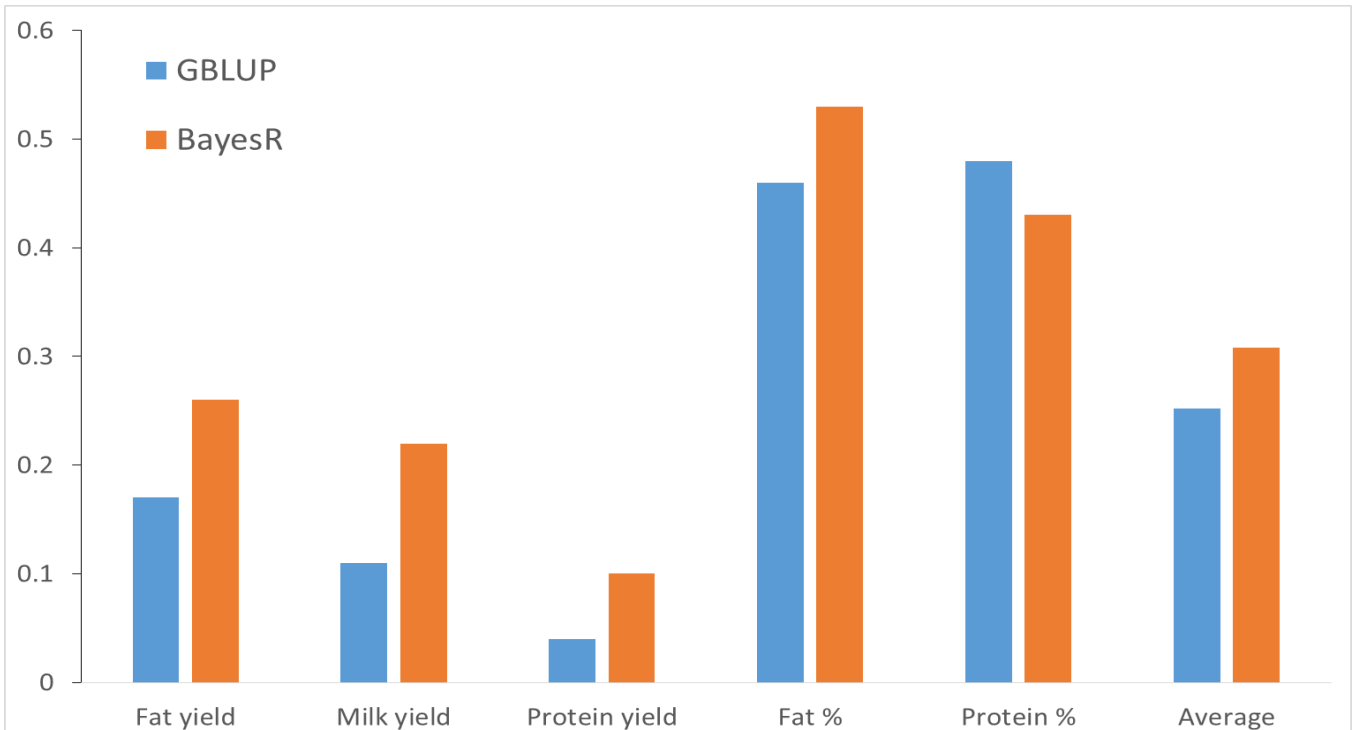
■ % Genetic Var - SNP ■ % Genetic Var - Ped



Aust Bull & Cow: Holstein & Jersey

Danz Bulls Only: Holstein & Jersey

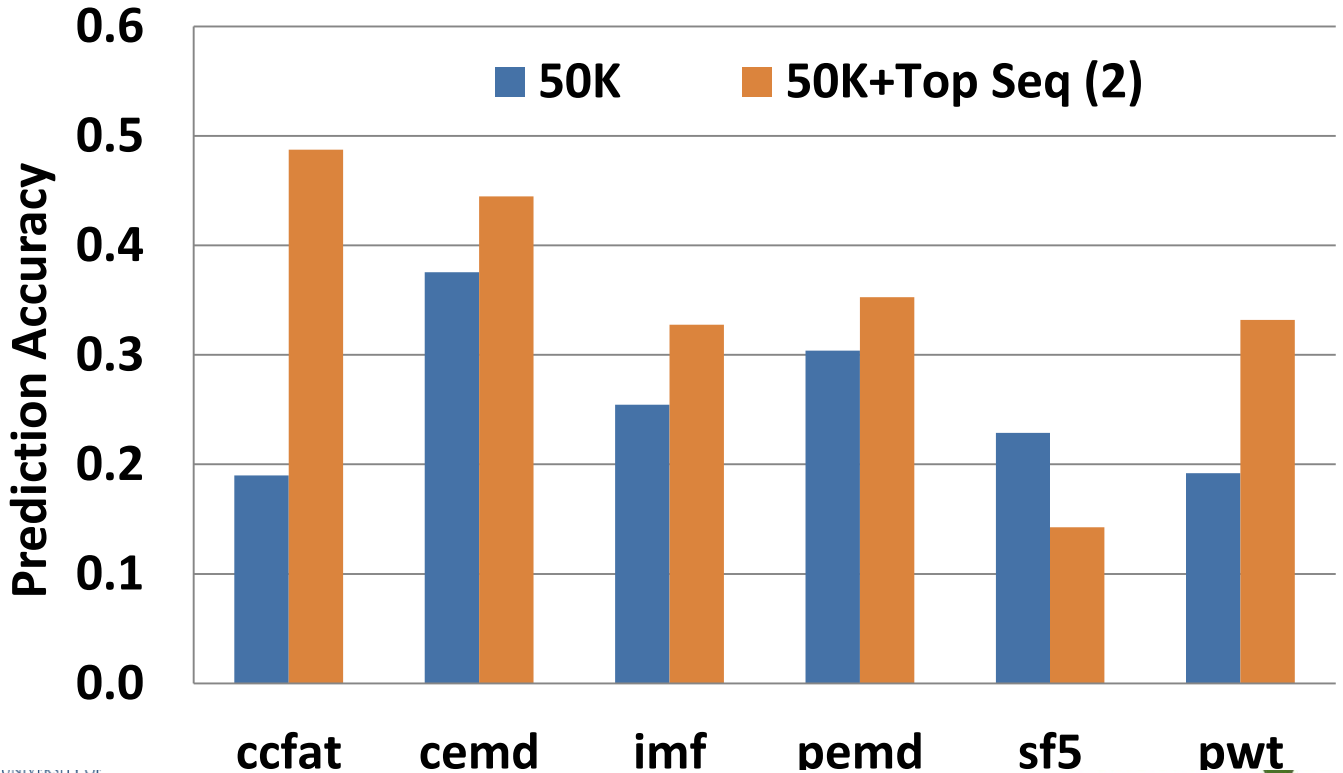
# Accuracy $r(\text{DGV}, \text{DTD})$ in Aussie Red Bulls (Iona MacLeod)





# Meat Traits:

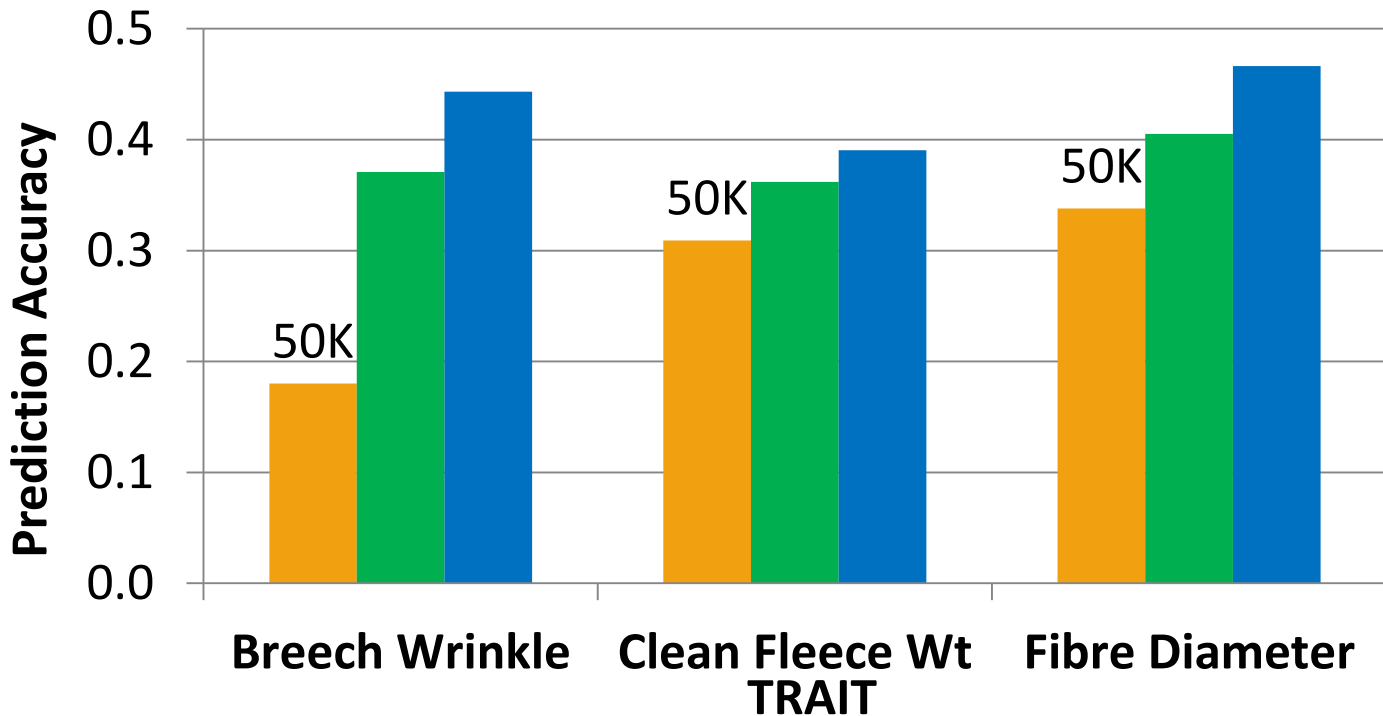
## GBLUP Accuracy - Merino x Border Leicester



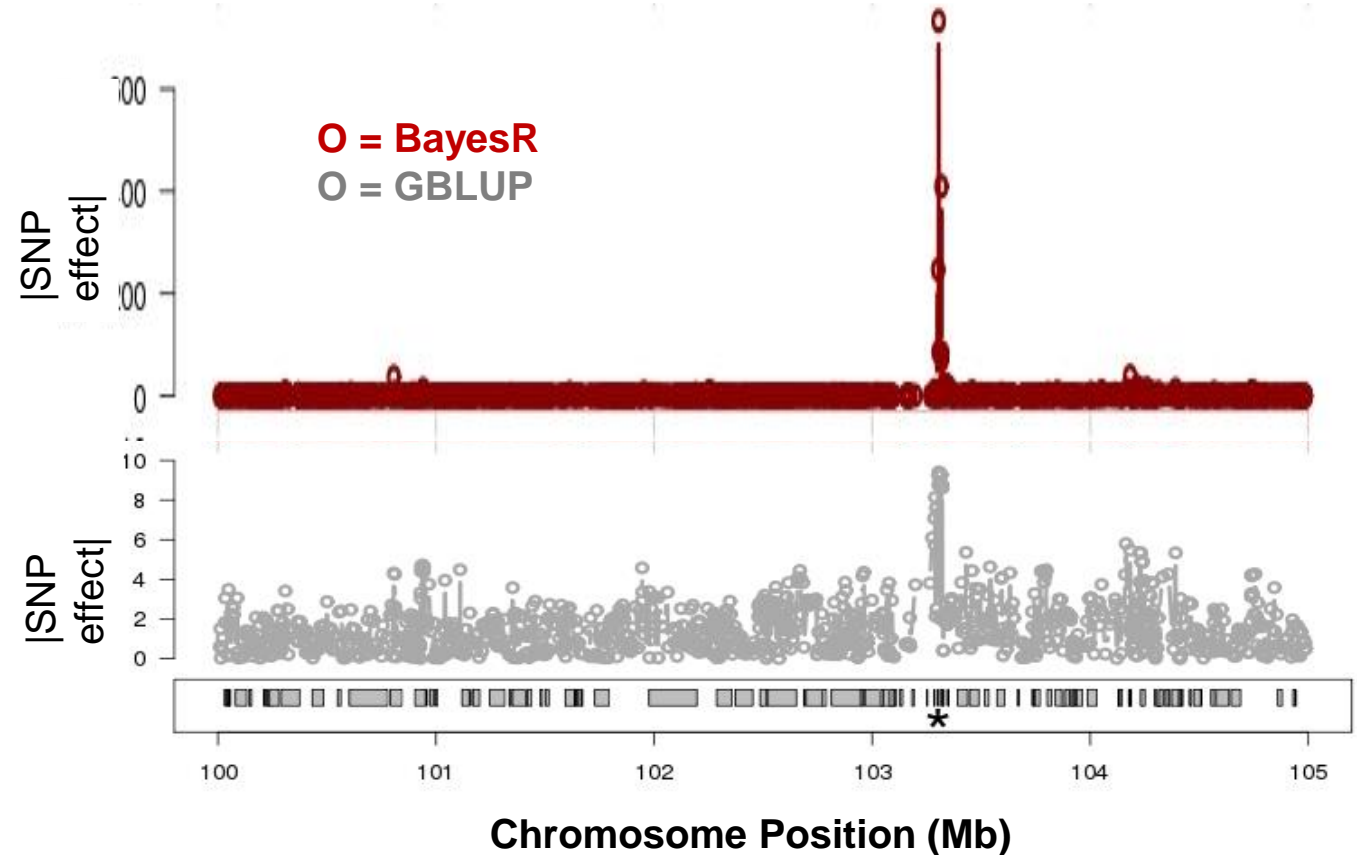
# Wool Traits:

## Prediction Accuracy in Merinos

50K BayesR 50K+Top Seq BayesRC 50K+Top Seq



# BayesR vs BLUP (BTA11)

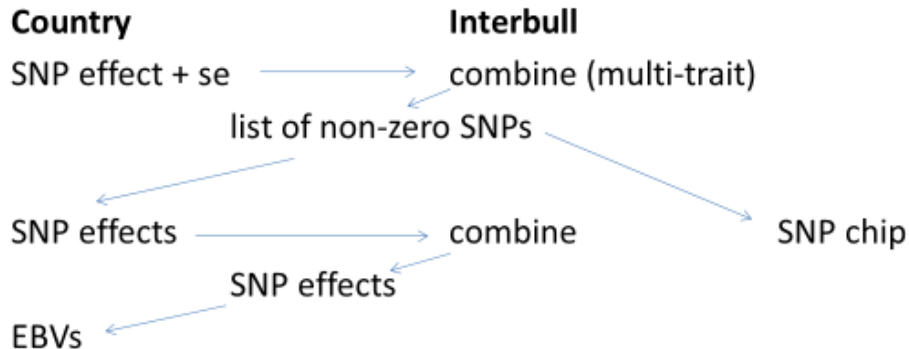


# Cattle stature (Aniek Bouwman, Ben Hayes et al)

Annotation class	Number
intergenic_variant	83
upstream_gene_variant	11
5_prime_UTR_variant	1
intron_variant	55
missense_variant	5
downstream_gene_variant	8
ChiP-SEQ peaks*	8
WBC eQTL	10

# Proposal

Countries could collaborate through Interbull with one or both of these analyses



Breaker

# Conclusions

Meta-analysis could be used more to collaborate between countries

Interbull project is an example

# Acknowledgements

Thank you to the  
Brown Swiss community for access to their data



## Section Breaker