



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

CREATE CHANGE

# Best Linear Unbiased Prediction (BLUP)

Jian Zeng

[j.zeng@uq.edu.au](mailto:j.zeng@uq.edu.au)



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

Institute for Molecular Bioscience



Program in Complex  
Trait Genomics

Slides credit: Ben Hayes

A weighted sum of the count of risk alleles

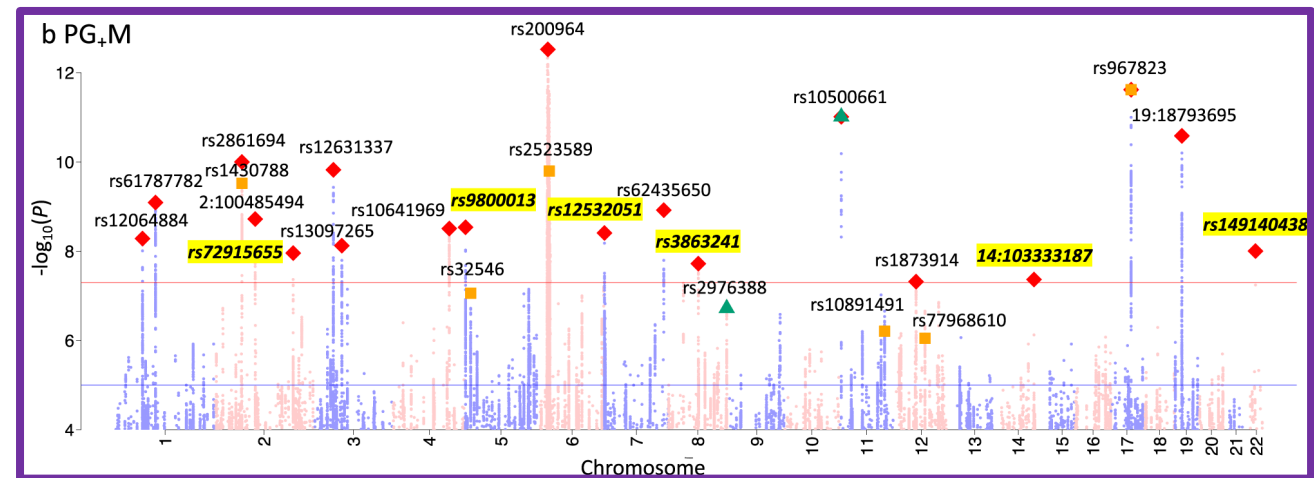
$$PRS = \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} + \dots = \sum_{j=1}^{n_{SNP}} \widehat{\beta}_j x_{ij}$$

How many SNPs?  
Which SNPs?  
What weights?

## Basic method:

Clumping & P-value thresholding (C+PT):

- Select most associated SNP in tower – LD-based clumping
- Select on a p-value threshold



A weighted sum of the count of risk alleles

$$PRS = \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} + \dots = \sum_{j=1}^{n_{SNP}} \widehat{\beta}_j x_{ij}$$

How many SNPs?  
Which SNPs?  
What weights?

**Can we simultaneously use all SNPs?**

Yes! But ...

cannot aggregate GWAS effects

due to linkage disequilibrium (double counting)

A weighted sum of the count of risk alleles

$$PRS = \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} + \dots = \sum_{j=1}^{n_{SNP}} \widehat{\beta}_j x_{ij}$$

How many SNPs?  
Which SNPs?  
What weights?

**Estimate SNP effects with a multiple regression?**

Yes!

But ...

## Linear model

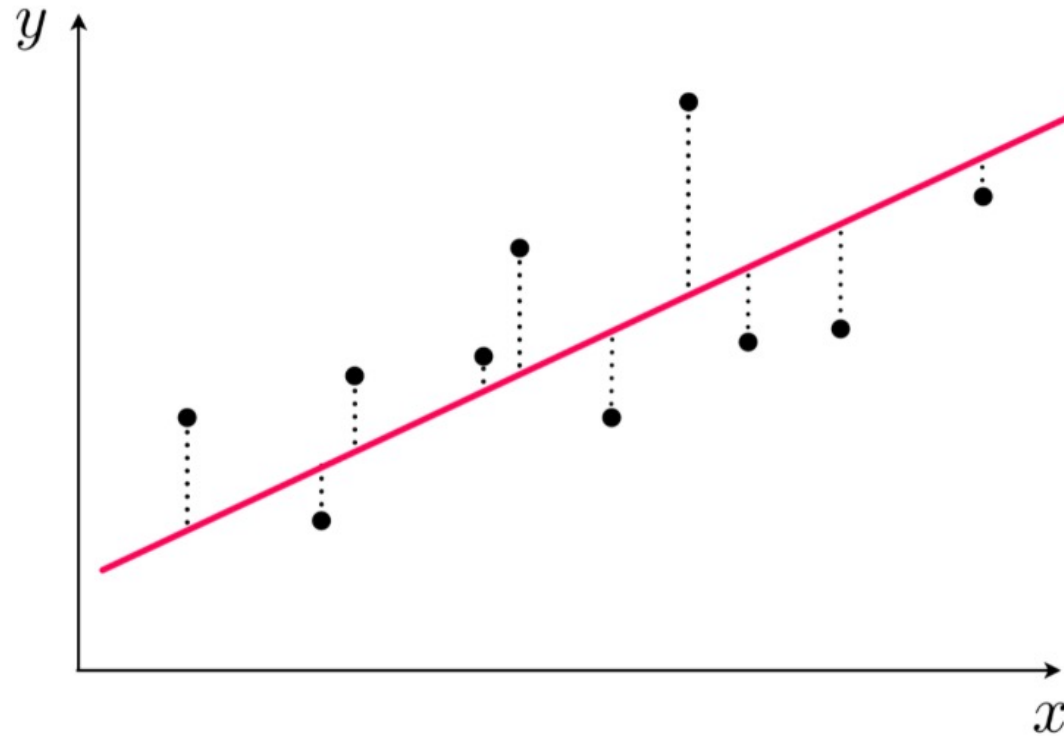
$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where

- $\mathbf{y}$  is a vector of  $n$  phenotypes,
- $\mu$  is the mean,
- $\mathbf{X}$  is an incidence matrix of individuals' genotypes for all SNPs,
- $\boldsymbol{\beta}$  are the **fixed** effects of the  $m$  SNPs,
- $\mathbf{e}$  is a vector of random residuals,  $\mathbf{e} \sim N(0, \sigma_e^2)$

# Least squares method

Least squares (LS): minimising the sum of squares of the residuals.



## Linear model

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

LS solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

No unique solutions when #SNPs > #individuals  
( $p > n$  problem)

# Linear mixed model

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where

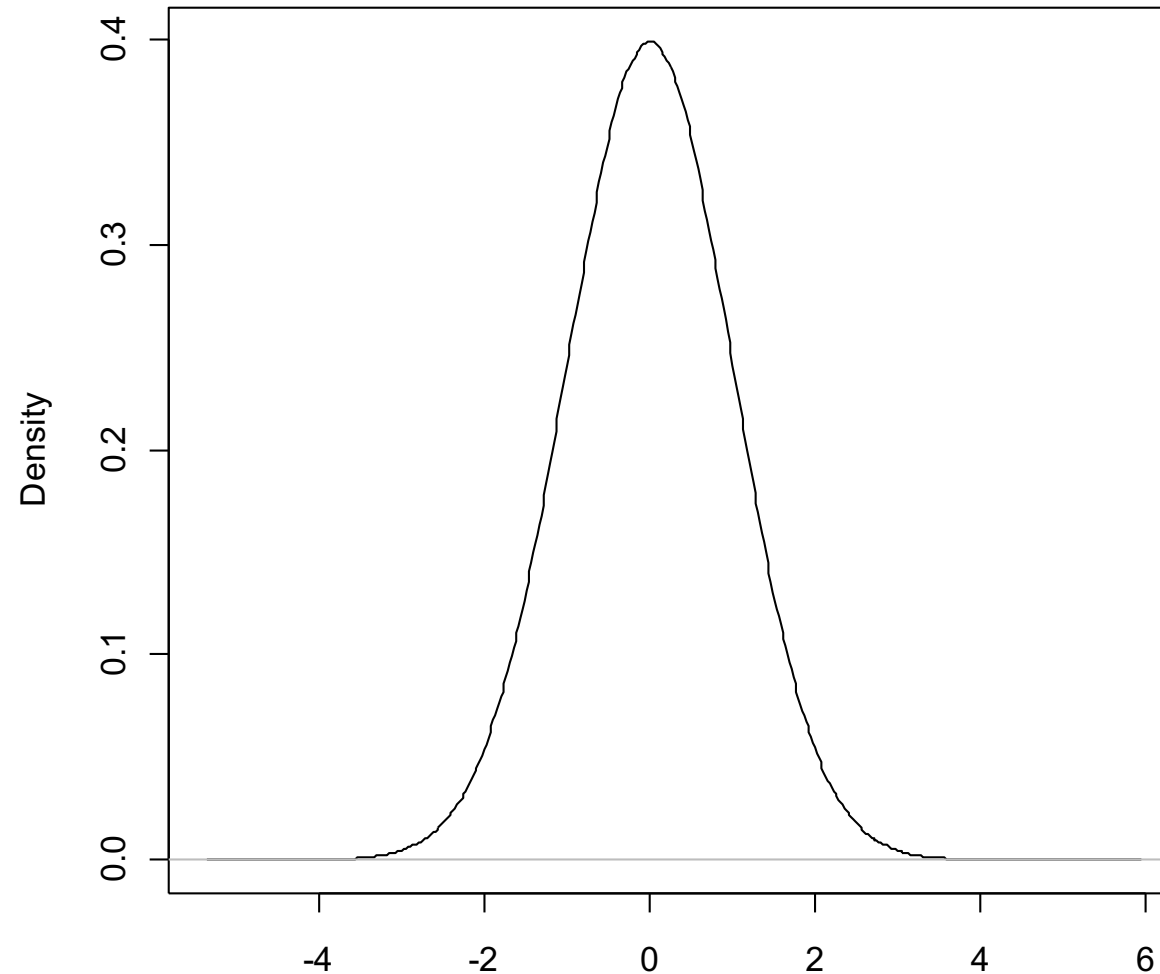
- $\mathbf{y}$  is a vector of  $n$  phenotypes,
- $\mu$  is the mean,
- $\mathbf{X}$  is an incidence matrix of individuals' genotypes for all SNPs,
- $\boldsymbol{\beta}$  are the **random** effects of the  $m$  SNPs,
- $\mathbf{e}$  is a vector of random residuals,  $\mathbf{e} \sim N(0, \sigma_e^2)$

Assume SNP effects come from normal distribution with same variance  $\boldsymbol{\beta} \sim N(0, \sigma_\beta^2)$



# Assumed distribution of SNP effects

$$N(0, \sigma_{\beta}^2)$$



# Best linear unbiased prediction

To estimate random effects (Henderson 1975 & Robinson 1991).

**Best:** minimum mean square error within class of linear predictors

**Linear:** random variables  $\beta$  are linear functions of the data  $\mathbf{y}$

**Unbiased:** the average value of the estimate of  $\beta$  is equal to the average value of the quantity being estimated

**Predictor:** to distinguish random effects from fixed effect estimates

# Best linear unbiased prediction (BLUP)

Linear mixed model

$$\mathbf{y} = \mathbf{1}_n \mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

$\mathbf{I}$  = identity matrix (dimensions  $m \times m$ )

$$\lambda = \sigma_e^2 / \sigma_\beta^2$$

## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

## LS solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

- 10 SNPs
- Only 5 phenotypes

		X									
Individual	y	1	2	3	4	5	6	7	8	9	10
1	0.19	0	0	0	0	0	0	1	2	0	2
2	1.23	1	0	0	1	1	1	2	1	0	1
3	0.86	1	0	0	1	0	0	1	1	1	1
4	1.23	1	1	1	1	0	1	2	1	1	1
5	0.45	0	1	1	1	1	1	2	1	0	1

# Example

Let  $\mathbf{1}_n' = [1 \ 1 \ 1 \ 1 \ 1]$

Assume value of 1 for  $\lambda$

BLUP solutions

		X									
Individual	y	1	2	3	4	5	6	7	8	9	10
1	0.19	0	0	0	0	0	0	1	2	0	2
2	1.23	1	0	0	1	1	1	2	1	0	1
3	0.86	1	0	0	1	0	0	1	1	1	1
4	1.23	1	1	1	1	0	1	2	1	1	1
5	0.45	0	1	1	1	1	1	2	1	0	1

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

5	3	2	2	4	2	3	8	6	2	6		3.96
3	4	1	1	3	1	2	5	3	2	3		3.32
2	1	3	2	2	1	2	4	2	1	2		1.68
2	1	2	3	2	1	2	4	2	1	2		1.68
4	3	2	2	5	2	3	7	4	2	4		3.77
2	1	1	1	2	3	2	4	2	0	2		1.68
3	2	2	2	3	2	4	6	3	1	3		2.91
8	5	4	4	7	4	6	15	9	3	9		6.87
6	3	2	2	4	2	3	9	9	2	8		4.15
2	2	1	1	2	0	1	3	2	3	2		2.09
6	3	2	2	4	2	3	9	8	2	9		4.15

## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

5	3	2	2	4	2	3	8	6	2	6	3.96
3	4	1	1	3	1	2	5	3	2	3	3.32
2	1	3	2	2	1	2	4	2	1	2	1.68
2	1	2	3	2	1	2	4	2	1	2	1.68
4	3	2	2	5	2	3	7	4	2	4	3.77
2	1	1	1	2	3	2	4	2	0	2	1.68
3	2	2	2	3	2	4	6	3	1	3	2.91
8	5	4	4	7	4	6	15	9	3	9	6.87
6	3	2	2	4	2	3	9	9	2	8	4.15
2	2	1	1	2	0	1	3	2	3	2	2.09
6	3	2	2	4	2	3	9	8	2	9	4.15



## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

5	3	2	2	4	2	3	8	6	2	6	3.96
3	4	1	1	3	1	2	5	3	2	3	3.32
2	1	3	2	2	1	2	4	2	1	2	1.68
2	1	2	3	2	1	2	4	2	1	2	1.68
4	3	2	2	5	2	3	7	4	2	4	3.77
2	1	1	1	2	3	2	4	2	0	2	1.68
3	2	2	2	3	2	4	6	3	1	3	2.91
8	5	4	4	7	4	6	15	9	3	9	6.87
6	3	2	2	4	2	3	9	9	2	8	4.15
2	2	1	1	2	0	1	3	2	3	2	2.09
6	3	2	2	4	2	3	9	8	2	9	4.15

## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

5	3	2	2	4	2	3	8	6	2	6	3.96
3	4	1	1	3	1	2	5	3	2	3	3.32
2	1	3	2	2	1	2	4	2	1	2	1.68
2	1	2	3	2	1	2	4	2	1	2	1.68
4	3	2	2	5	2	3	7	4	2	4	3.77
2	1	1	1	2	3	2	4	2	0	2	1.68
3	2	2	2	3	2	4	6	3	1	3	2.91
8	5	4	4	7	4	6	15	9	3	9	6.87
6	3	2	2	4	2	3	9	9	2	8	4.15
2	2	1	1	2	0	1	3	2	3	2	2.09
6	3	2	2	4	2	3	9	8	2	9	4.15

## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

5.96	-0.46	-0.04	-0.04	-0.81	-0.31	-0.01	-1.01	-1.19	-0.50	-1.19		3.96
-0.46	0.65	0.11	0.11	-0.11	0.08	-0.06	-0.06	0.11	-0.18	0.11		3.32
-0.04	0.11	0.72	-0.28	-0.03	0.04	-0.11	-0.11	0.03	-0.07	0.03		1.68
-0.04	0.11	-0.28	0.72	-0.03	0.04	-0.11	-0.11	0.03	-0.07	0.03		1.68
-0.81	-0.11	-0.03	-0.03	0.83	-0.09	-0.05	-0.05	0.17	-0.09	0.17		3.77
-0.31	0.08	0.04	0.04	-0.09	0.68	-0.12	-0.12	0.09	0.24	0.09		1.68
-0.01	-0.06	-0.11	-0.11	-0.05	-0.12	0.76	-0.24	0.05	0.07	0.05		2.91
-1.01	-0.06	-0.11	-0.11	-0.05	-0.12	-0.24	0.76	0.05	0.07	0.05		6.87
-1.19	0.11	0.03	0.03	0.17	0.09	0.05	0.05	0.83	0.09	-0.17		4.15
-0.50	-0.18	-0.07	-0.07	-0.09	0.24	0.07	0.07	0.09	0.68	0.09		2.09
-1.19	0.11	0.03	0.03	0.17	0.09	0.05	0.05	-0.17	0.09	0.83		4.15

## BLUP solutions

Mean	0.47
SNP1	0.29
SNP2	-0.05
SNP3	-0.05
SNP4	0.08
SNP5	-0.02
SNP6	0.13
SNP7	0.13
SNP8	-0.08
SNP9	0.11
SNP10	-0.08

“Smear” the effect  
over SNPs in LD

Now we want to predict PGS of a group of young individuals without phenotypes

$$\text{PGS} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

We have the  $\hat{\boldsymbol{\beta}}$ , and we can get  $\mathbf{X}$  from their genotypes (after genotyping).....

Young individuals	X										
	1	1	1	1	1	1	1	2	1	0	1
	2	1	0	0	1	1	1	1	1	0	1
	3	1	0	0	1	1	1	2	1	0	1
	4	1	0	0	1	1	2	2	1	0	1
	5	0	0	0	0	0	0	1	2	0	2

$$\text{PGS} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

$\mathbf{X}$										$\hat{\boldsymbol{\beta}}$	PGS
1	1	1	1	1	1	2	1	0	1	0.29	0.48
1	0	0	1	1	1	1	1	0	1	-0.05	0.45
1	0	0	1	1	1	2	1	0	1	-0.05	0.58
1	0	0	1	1	2	2	1	0	1	0.08	0.71
0	0	0	0	0	0	1	2	0	2	-0.02	-0.19
										0.13	
										0.13	
										-0.08	
										0.11	
										-0.08	

BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

$\lambda = \sigma_e^2 / \sigma_\beta^2$  is known as the shrinkage parameter

It shrinks LS estimates toward zero to an extent depending on the noise-signal ratio.

BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

$\lambda = \sigma_e^2 / \sigma_\beta^2$  is known as the shrinkage parameter

Ignoring mean and other SNP

$$\begin{aligned} \hat{\alpha}_1 &= \frac{x_1' y}{x_1' x_1 + \lambda} \\ &= (0 * 0.19 + 1 * 1.23 + 1 * 0.86 + 1 * 1.23 + 0 * 0.45) / (3 + 1) \end{aligned}$$

		X	
Individual	Y		1
1	0.19		0
2	1.23		1
3	0.86		1
4	1.23		1
5	0.45		0

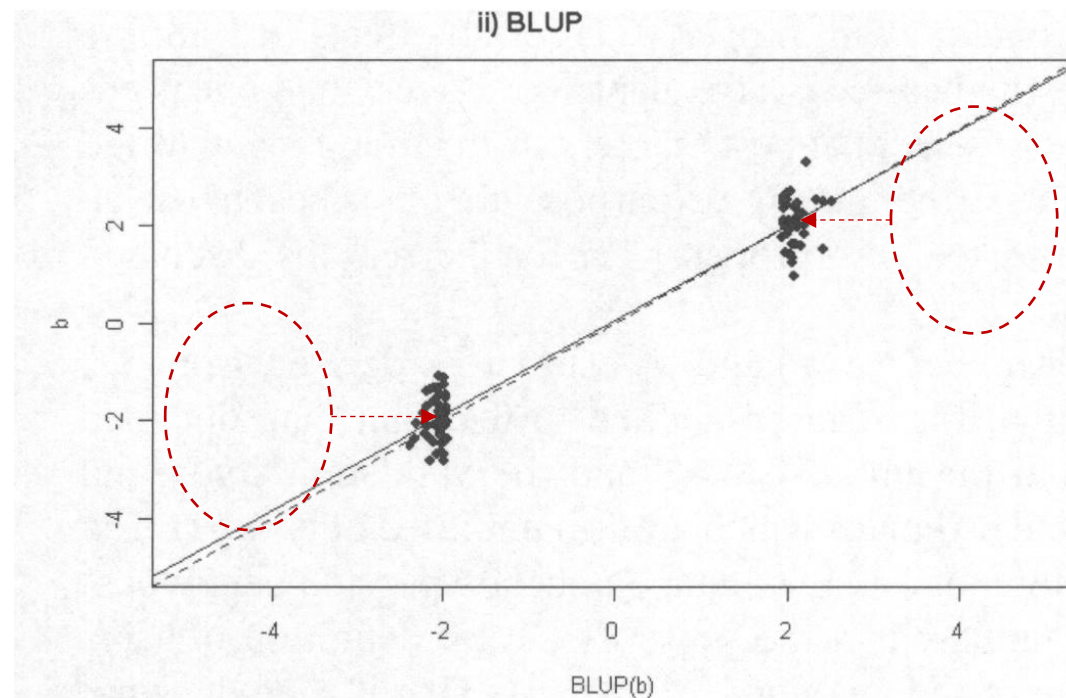
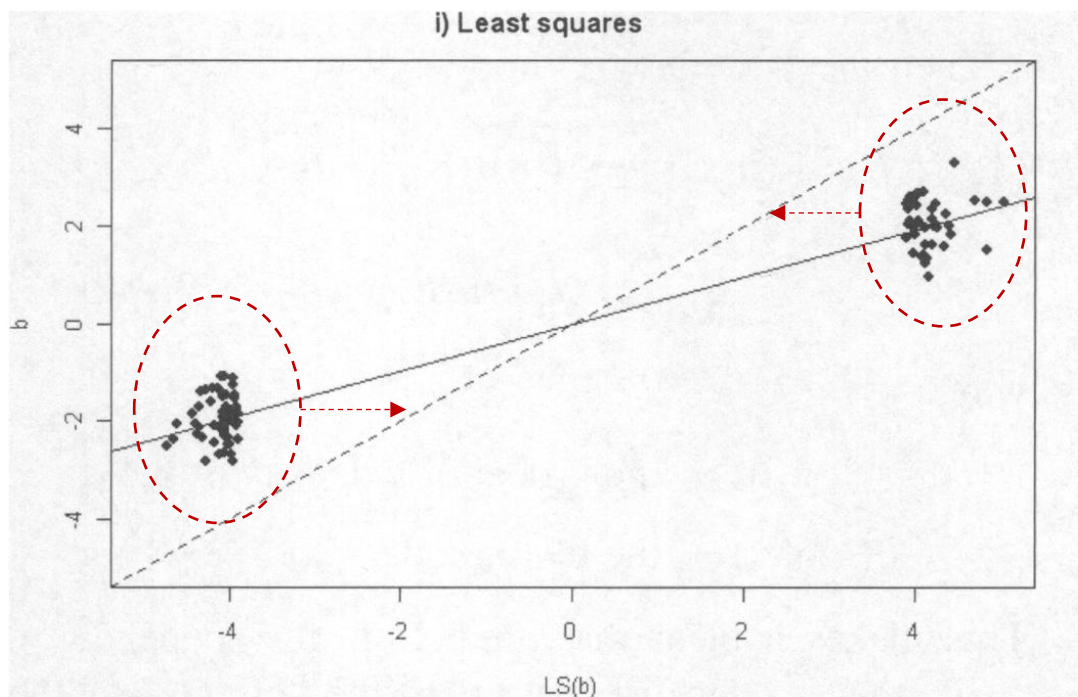


Shrinks LS estimates toward zero

*Statistical Science*  
2009, Vol. 24, No. 4, 517-529  
DOI: 10.1214/09-STS306  
© Institute of Mathematical Statistics, 2009

## Estimating Effects and Making Predictions from Genome-Wide Marker Data

Michael E. Goddard, Naomi R. Wray, Klara Verbyla and Peter M. Visscher



**BLUP avoids selection bias!**

*Statistical Science*  
2009, Vol. 24, No. 4, 517–529  
DOI: 10.1214/09-STS306  
© Institute of Mathematical Statistics, 2009

## Estimating Effects and Making Predictions from Genome-Wide Marker Data

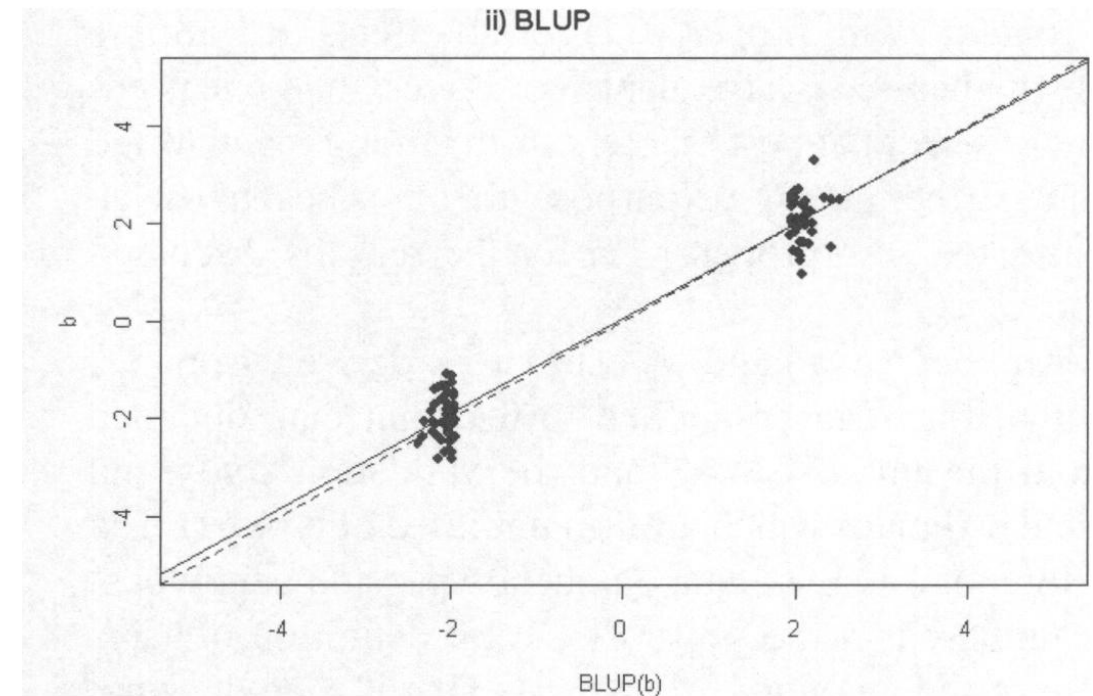
Michael E. Goddard, Naomi R. Wray, Klara Verbyla and Peter M. Visscher

**Unbiased:**  $E[\beta \mid \hat{\beta}_{\text{BLUP}}] = \hat{\beta}_{\text{BLUP}}$

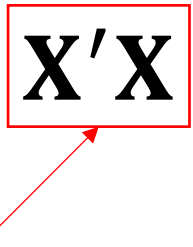
In contrast, for LS estimator:  $E[\hat{\beta}_{\text{LS}} \mid \beta] = \beta$

Desirable property of a genetic predictor:

The regression of  $y$  on the predictor has an intercept of zero and a slope of one.



## BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \mathbf{I} \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$


What's the dimension of this matrix?

**Think about fitting 20 million SNPs!**

## An equivalent model to SNP-BLUP

- If there are many causal variants whose effects are normally distributed with constant variance,
- Then it is equivalent to use a genomic relationship matrix (**GRM**) estimated from SNP markers in normal BLUP equations.
  - $GRM_{ij}$  = proportion of genome that is shared between individuals  $i$  and  $j$

## How to calculate GRM?

Rescale  $\mathbf{X}$  to account for allele frequencies

$$w_{ij} = \frac{(x_{ij} - 2p_i)}{\sqrt{2p_i(1 - p_i)}}$$

Then, the genetic values are  $\mathbf{g} = \mathbf{W}\boldsymbol{\beta}$

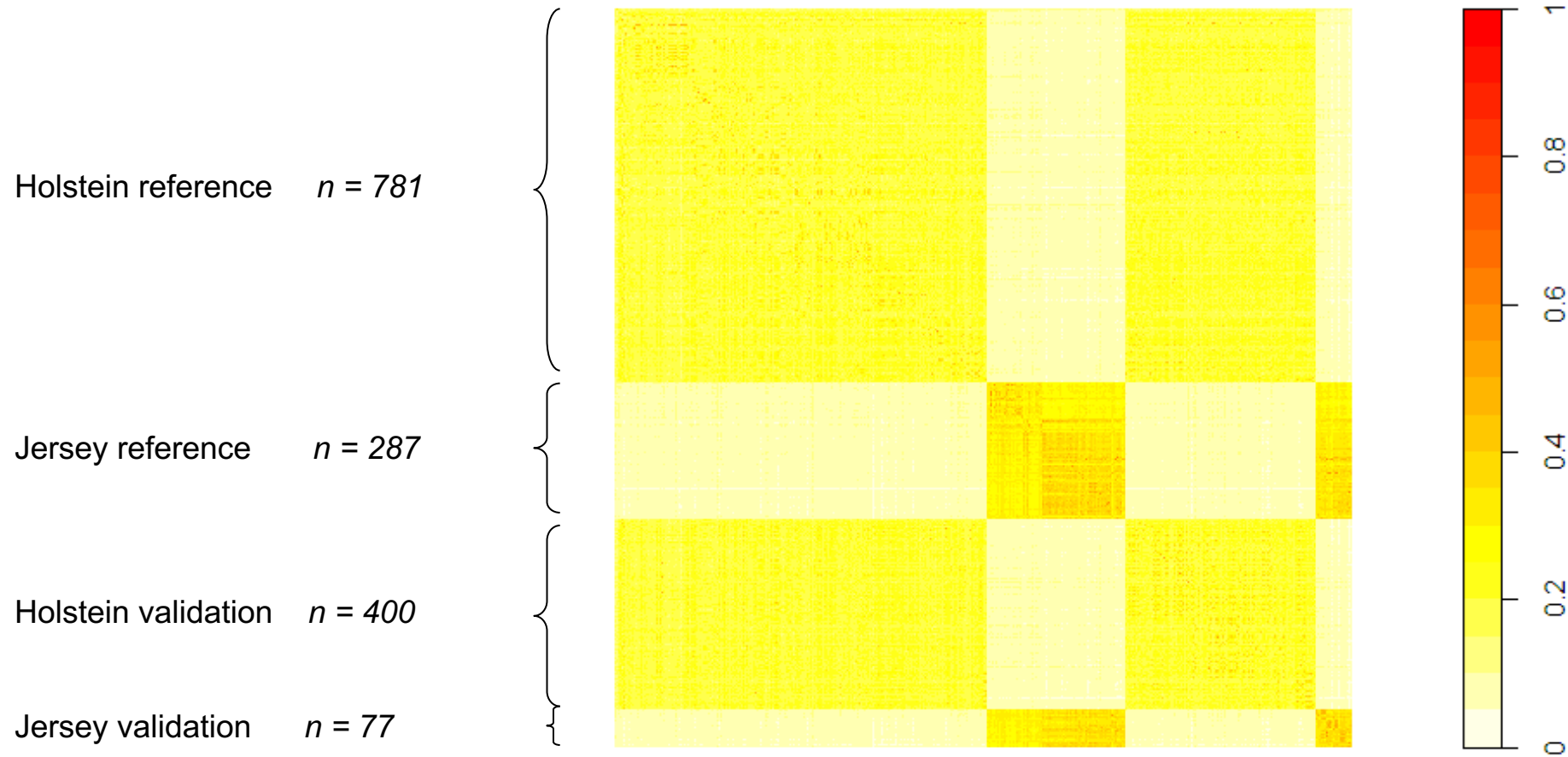
$$\text{Var}(\mathbf{g}) = \mathbf{W} \mathbf{W}' \sigma_{\beta}^2 = \frac{\mathbf{W}\mathbf{W}'}{m} m\sigma_{\beta}^2 = \mathbf{G}\sigma_g^2$$

Hence

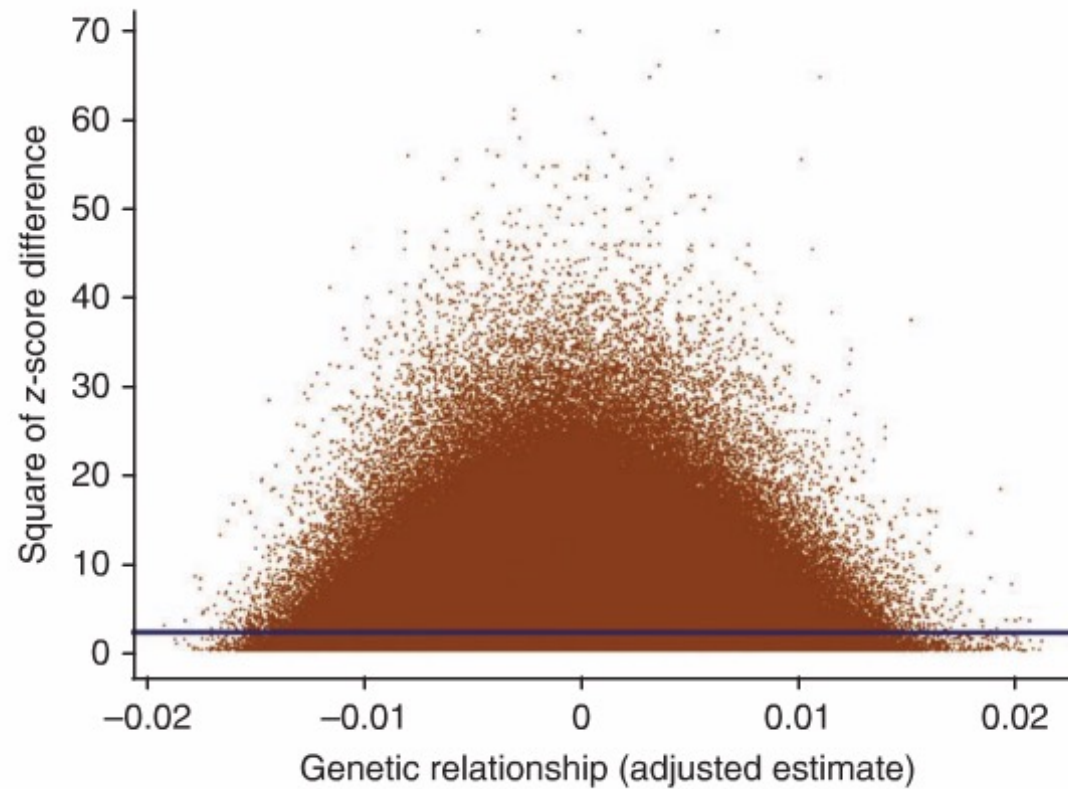
$\mathbf{G} = \frac{\mathbf{W}\mathbf{W}'}{m}$  is the GRM, and  $\sigma_g^2 = m\sigma_{\beta}^2$  is the total genetic variance

# Genomic relationship matrix (GRM)

## In cattle



In humans (unrelated individuals)



Yang et al (2010 Nat Genet)

An equivalent model

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

where

$$\text{Var}(\mathbf{g}) = \mathbf{G} \sigma_g^2$$

BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{Z} \\ \mathbf{Z}' \mathbf{1}_n & \mathbf{Z}' \mathbf{Z} + \mathbf{G}^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{Z}' \mathbf{y} \end{bmatrix}$$



Z matrix maps the phenotypic records onto the genetic values

e.g. 5 individuals with the first 3 having records

$$\mathbf{y} = \mathbf{Z} \times \mathbf{g}$$

<b>y</b>	=	<b>Z</b>	×	<b>g</b>	
$y_1$		1 0 0 0 0		$g_1$	
$y_2$		0 1 0 0 0		$g_2$	
$y_3$		0 0 1 0 0		$g_3$	
				$g_4$	} To be predicted
				$g_5$	

## Model 1 - SNP-BLUP

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n'\mathbf{1}_n & \mathbf{1}_n'\mathbf{X} \\ \mathbf{X}'\mathbf{1}_n & \mathbf{X}'\mathbf{X} + \mathbf{I}\frac{\sigma_e^2}{\sigma_\beta^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix}$$

## Model 2 - GBLUP

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n'\mathbf{1}_n & \mathbf{1}_n'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1}_n & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\mathbf{PGS} = \mathbf{X}\hat{\boldsymbol{\alpha}}$$

- If number of SNPs  $\gg \gg$  large than number of individuals, GBLUP is more computationally efficient
- Calculate prediction accuracy for each individual from inverse coefficient matrix (amount of data in estimate!)

- Prediction error variance  $PEV_i = C^{ii} \sigma_e^2$
- Accuracy  $r_i^2 = 1 - PEV_i / \sigma_g^2$

$$\mathbf{C} = \begin{bmatrix} \mathbf{1}'_n \mathbf{1}_n & \mathbf{1}'_n \mathbf{Z} \\ \mathbf{Z}' \mathbf{1}_n & \mathbf{Z}' \mathbf{Z} + \mathbf{G}^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}^{-1}$$

- Very useful – can calculate how well we predict for individuals without their own phenotype (e.g., young calves, people)

## Polygenic scoring accuracy varies across the genetic ancestry continuum

[Yi Ding](#) , [Kangcheng Hou](#), [Ziqi Xu](#), [Aditya Pimplaskar](#), [Ella Petter](#), [Kristin Boulier](#), [Florian Privé](#), [Bjarni J. Vilhjálmsson](#), [Loes M. Olde Loohuis](#) & [Bogdan Pasaniuc](#) 

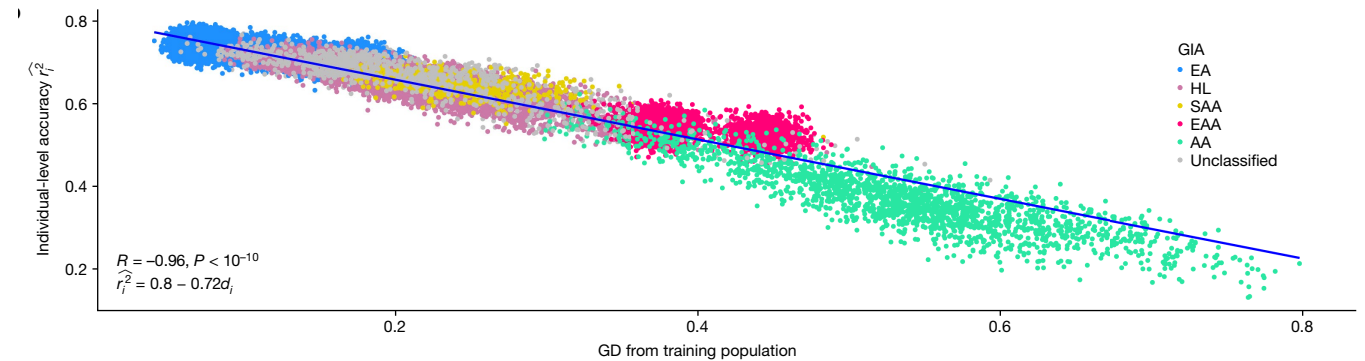
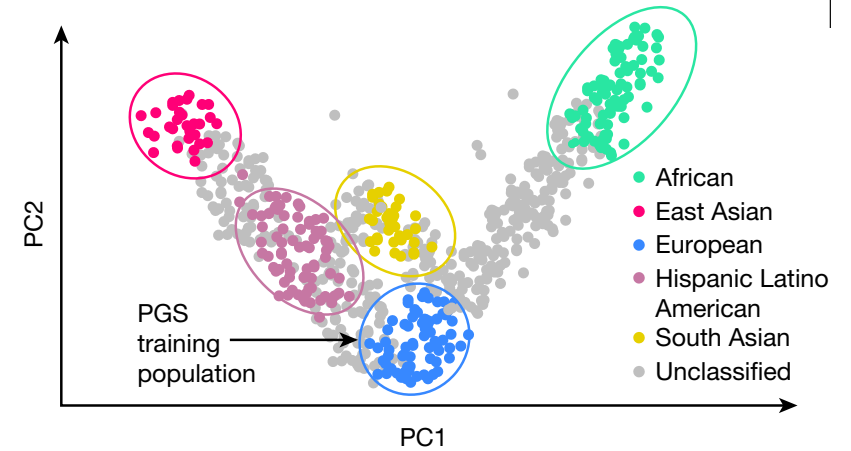
[Nature](#) (2023) | [Cite this article](#)

11k Accesses | 200 Altmetric | [Metrics](#)

### Individual PGS accuracy

We define individual PGS accuracy as the squared correlation between an individual's genetic liability,  $g_i$ , and its PGS estimate,  $\hat{g}_i$ , following the general form in ref. 28:

$$r_i^2 = 1 - \frac{E_D(\text{var}_{\beta|D}(x_i^T \beta))}{\text{var}_{\beta}(x_i^T \beta)}$$



- Moving from GBLUP to SNP-BLUP
- called Backsolving for SNP effects

$$\hat{\beta} = \mathbf{X}'\mathbf{G}^{-1} \hat{\mathbf{g}} / m$$

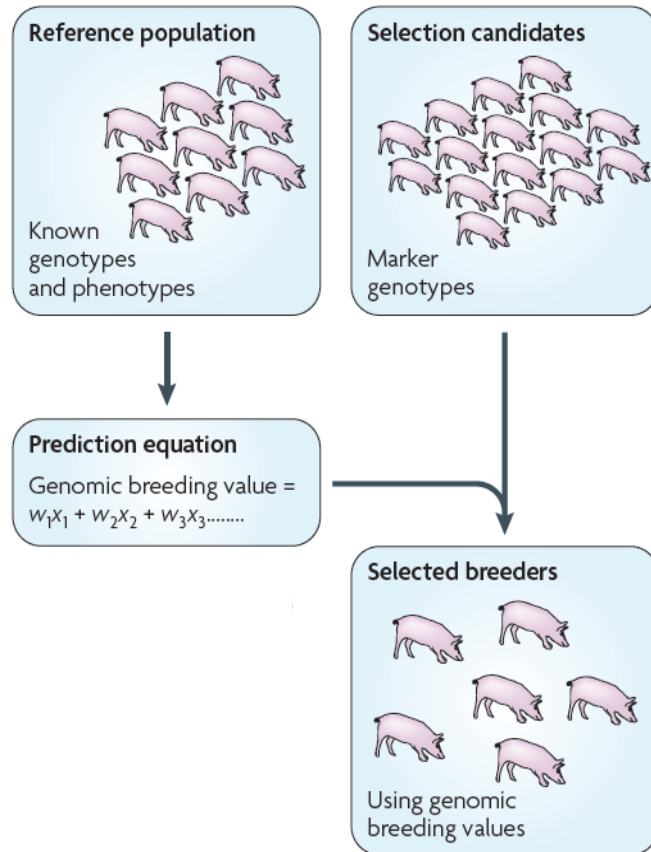
- Can use in alternative form of GWAS

## Examples of BLUP applications

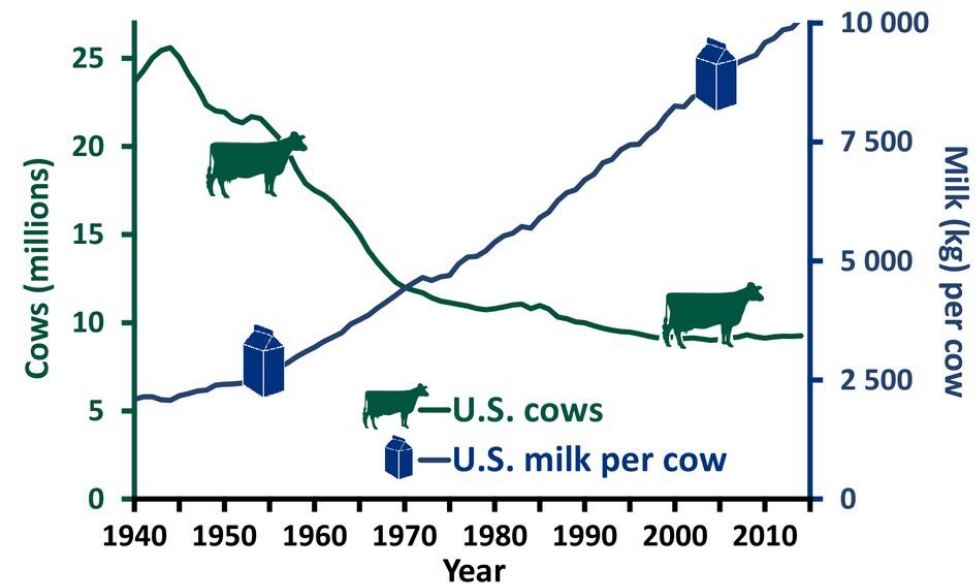
- Genomic selection in livestock
- Disease risk prediction in humans

Use genome-wide SNPs to estimate the breeding value of selection candidates.

“Genomic selection” = “precision medicine” for animals



## U.S. dairy population & milk yield



## Humans – Crohn's disease

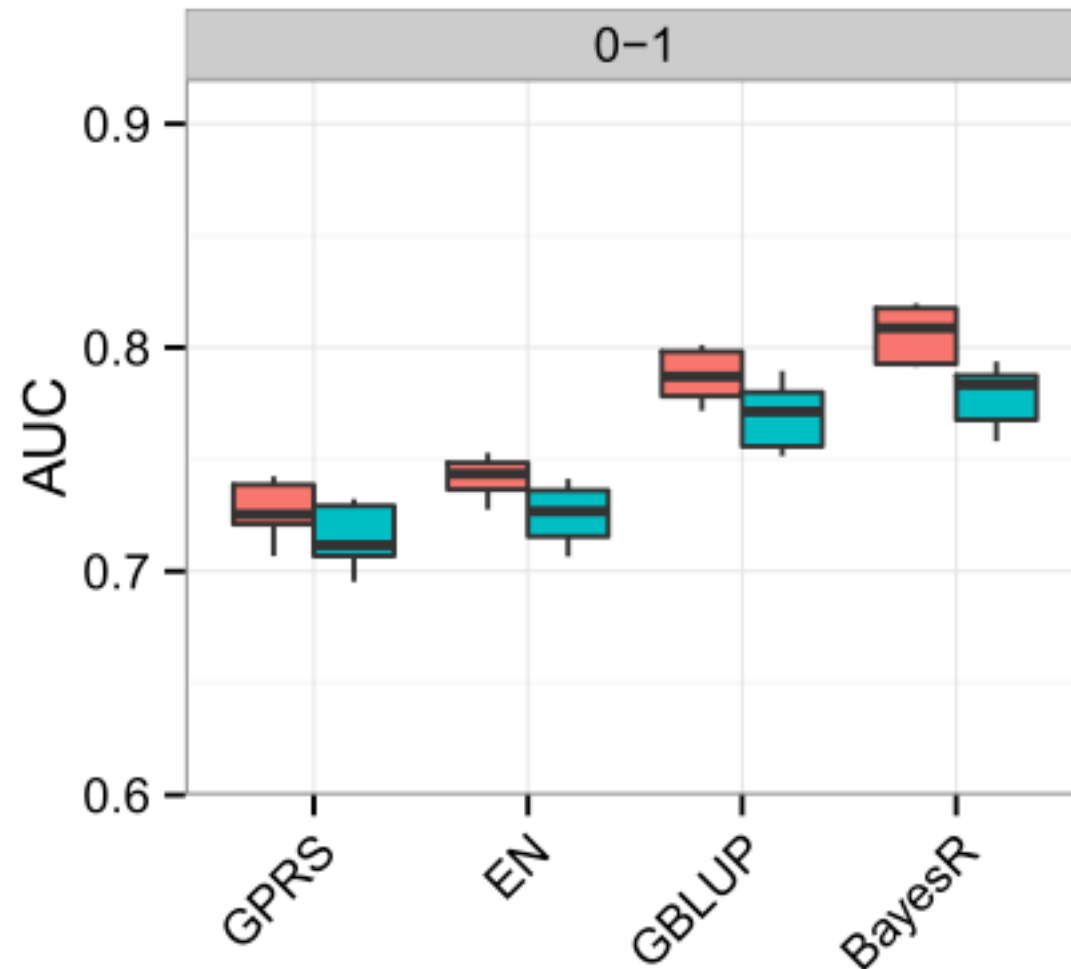
Chen et al. 2017. BMC Medicine.

- Inflammatory Bowel Disease
- Affects 2 in every 1000 people (approx.)
- 68,000 IBD patients and 29,000 healthy controls from 15 cohorts, European descent
- 909,763 GWAS SNPs or 123,437 SNPs on the custom designed ImmunoChip
- Prediction methods:
  - Genetic profile risk scores (GPRS) constructed using effects of all SNPs from GWAS
  - GBLUP
  - Elastic net (EN)
  - BayesR - Bayesian method that models SNP effects as a mixture of 4 normal distributions.

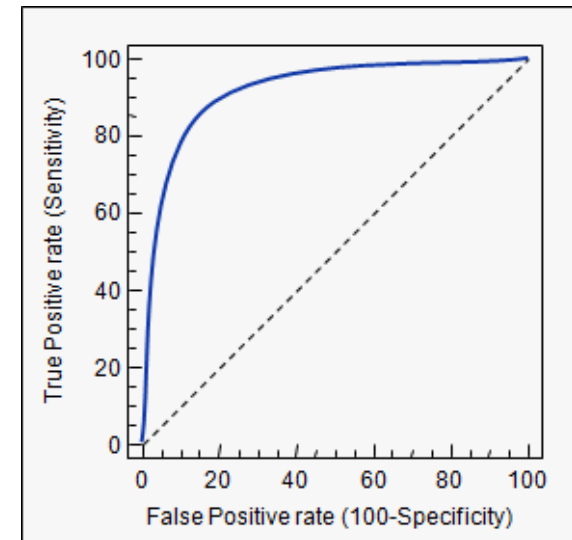


## Humans – Crohn's disease

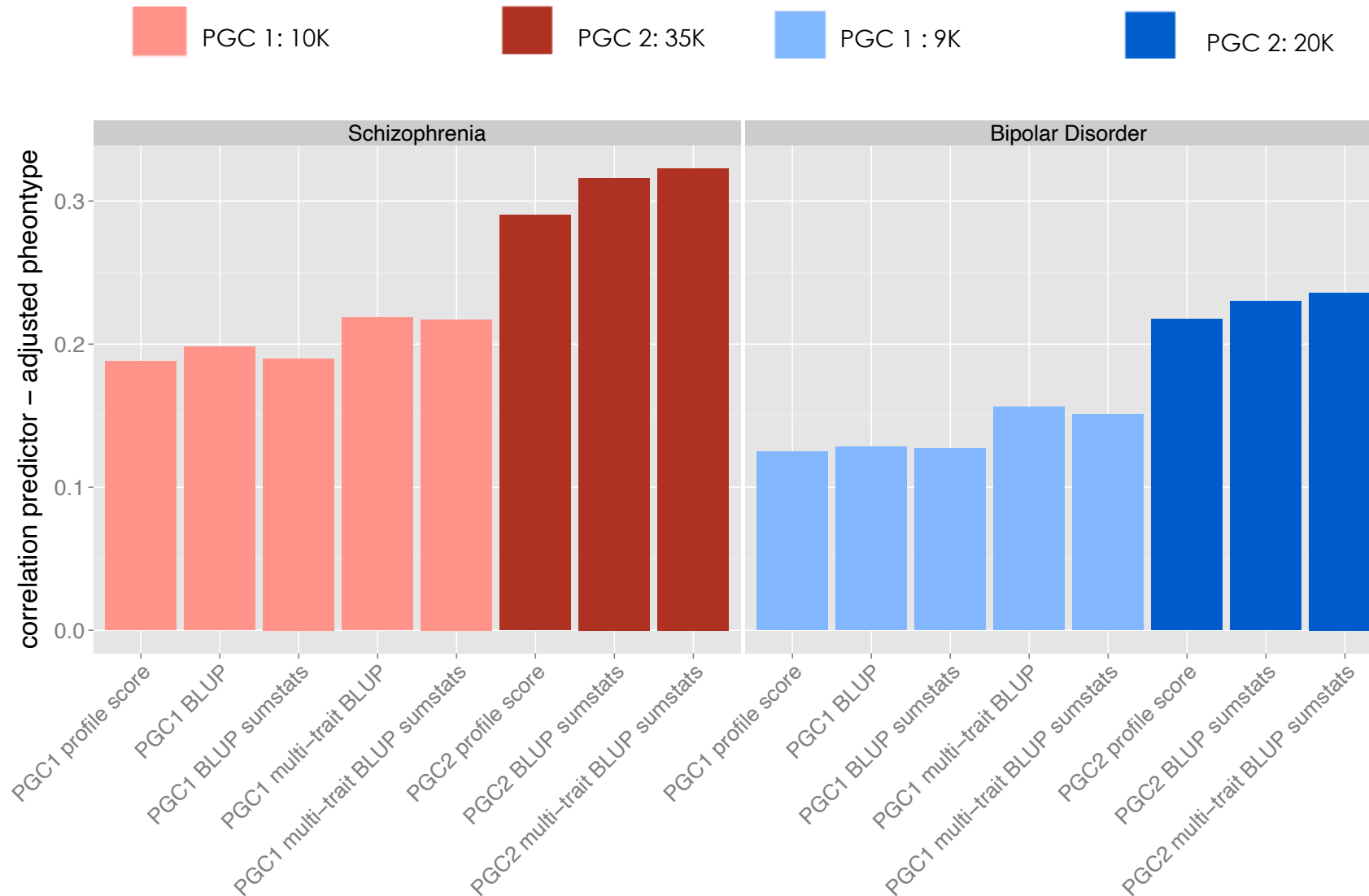
Chen et al. 2017. BMC Medicine.



Assess value of predictions as “Area Under Curve” (AUC) from 5-fold cross-validation



# Predict risk of psychiatric disorders



Maier et al (2015) Joint Analysis of Psychiatric Disorders Increases Accuracy of Risk Prediction for Schizophrenia, Bipolar Disorder, and Major Depressive Disorder. AJHG. (Not summary statistics) CRICOS code 00025B

## BLUP

- Simultaneously estimate all SNP effects as random
  - No need to prune on LD or select p-value threshold
  - No need to know causal variants or biological function
- Assumes normal distribution on SNP effects with equal variance
- Unbiased estimates of SNP effects
- Equivalent models between SNP-BLUP and GBLUP
- Provide per-individual prediction accuracy
- Improved prediction accuracy in practice

# Practical 3: BLUP

[https://cnsgenomics.com/data/teaching/GNGWS24/module5/Practical3\\_BLUP.html](https://cnsgenomics.com/data/teaching/GNGWS24/module5/Practical3_BLUP.html)

To log into your server, type command below in **Terminal** for Mac/Linux users or in **Command Prompt** or **PowerShell** for Windows users.

```
ssh username@hostname
```

And then key in the provided password.