

# Testing Association of Genotypes with Discrete and Continuous Traits

Kacper Żukowski  
Anna Macierzyńska  
Heliodor Wierzbicki

Department of Genetics and Animal Breeding  
Wrocław University of Environmental and Life Sciences



# The aim of the work:

- prediction of breeding values of continuous and binary trait,
- investigation of associations between genetic markers and both traits,
- comparison of three different approaches of estimating particular effects

# SNPs selection

- for gBLUP

	<b>10031</b>
MAF > 1%	<b>-264</b>

---

**9767 SNPs in training**

- for gFouley

	<b>10031</b>
MAF > 10%	<b>-934</b>
$r^2 > 0.2$	<b>-3263</b>

---

**5834 SNPs in training**

# Models

- BLUP

y – continuous trait		y – binary trait	
$\sigma_\alpha^2 = 59.73$	$\sigma_e^2 = 46.08$	$\sigma_\alpha^2 = 0.25$	$\sigma_e^2 = 1$
$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$ $\alpha \sim N(0, A\hat{\sigma}_\alpha^2) \qquad e \sim N(0, I\sigma_e^2)$			

- gBLUP

y – continuous trait		y – binary trait	
$\sigma_\alpha^2 = 59.73$	$\sigma_e^2 = 46.08$	$\sigma_\alpha^2 = 0.25$	$\sigma_e^2 = 1$
$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}$ $a \sim N\left(0, I \frac{\hat{\sigma}_\alpha^2}{n_{SNP}}\right) \qquad e \sim N(0, I\sigma_e^2) \qquad Z \in \{-1, 0, 1\}$			

# Foulley's model

- joint analysis of quantitative and binary traits, model proposed by *Foulley et al., 1983*

- model 1 for quantitative trait:  $y_1 = X_1\beta_1 + Z_1u_1 + e_1$

- model 2 for binary trait:  $y_2 = X_2\beta_2 + Z_2u_2 + e_2$

where:

- $y_1$  is vector of quantitative trait
- $y_2$  is vector of binary trait
- $\beta_1$  and  $u_1$  are fixed effects and random effects for quantitative trait
- $X_1$  and  $Z_1$  are design matrices for quantitative trait
- $Z_2 = Z_1$  and  $X_2 = X_1H$  where H is an identity matrix if all factors affecting the quantitative trait also affect the binary trait

# Foulley's model

$$\text{var} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \quad \text{var} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = A \otimes G$$

where:

- $G$  is the genetic covariance matrix for both traits
- $A$  is the **numerator relationship matrix or kinship matrix** (Loiselle *et al.* 1995)
- $\theta$  is vector of location parameters,  $\theta' = [\beta_1, \tau, u_1, v]$

where:

- $\tau = \beta_2 - bH\beta_1$  and  $v = u_2 - bu_1$
- $b$  is the regression coefficient,  $b = r_{12} \left( \frac{1}{\sigma_{e1}} \right) \frac{1}{\sqrt{1 - r_{12}^2}}$

and  $r_{12}$  is the residual correlation coefficient

# Fouley's model - equations

$$\begin{bmatrix}
 X_1^T R_1^{-1} X_1 & X_1^T R_1^{-1} Z_1 & 0 & 0 \\
 Z_1^T R_1^{-1} X_1 & Z_1^T R_1^{-1} Z_1 + A^{-1} g_c^{11} & 0 & A^{-1} g_c^{12} \\
 0 & 0 & X_2^T W^{[i-1]} X_2 & X_2^T W^{[i-1]} Z_2 \\
 0 & A^{-1} g_c^{21} & Z_2^T W^{[i-1]} X_2 & Z_2^T W^{[i-1]} Z_2 + A^{-1} g_c^{22}
 \end{bmatrix}
 \begin{bmatrix}
 \hat{\beta}^{[i]} \\
 \hat{u}^{[i]} \\
 \Delta \hat{\tau}^{[i]} \\
 \Delta \hat{\nu}^{[i]}
 \end{bmatrix}
 =
 \begin{bmatrix}
 X_1^T R_1^{-1} y_1 \\
 Z_1^T R_1^{-1} X_1 \\
 X_2^T q^{[i-1]} \\
 Z_2^T q^{[i-1]}
 \end{bmatrix}
 -
 \begin{bmatrix}
 0 \\
 A^{-1} g_c^{12} \nu^{[i-1]} \\
 0 \\
 A^{-1} g_c^{22} \nu^{[i-1]}
 \end{bmatrix}$$

# gFoulley's model

- joint analysis of quantitative and binary traits combined genomic information

- Model 1 for quantitative trait  $y_1 = \mu_1 + Z_1 u_1 + e_1$

- Model 2 for binary trait  $y_2 = \mu_2 + Z_2 u_2 + e_2$

- $y_1$  vector of quantitative trait
- $y_2$  vector of binary trait
- $\mu_1$  and  $\mu_2$  mean for quantitative and binary trait
- $u_1$  and  $u_2$  vectors for random SNP effects for quantitative and binary trait
- $Z_1$  design matrices random SNP effects  $Z_1 \in \{-1,0,1\}$
- $e_1$  and  $e_2$  vector of random errors
- $Z_2 = Z_1$  and  $X_2 = X_1 H$  where H is an identity matrix if all factors affecting the quantitative trait also affect the binary trait

# gFouley's model

- We assumed:  $u_1 \sim N\left(0, \frac{\hat{\sigma}_{\alpha 1}^2}{n}\right)$   
 $u_2 \sim N\left(0, \frac{\hat{\sigma}_{\alpha 2}^2}{n}\right)$   
 $e_1 \sim N(0, I\sigma_{\varepsilon 1}^2)$

where:

- $\sigma_{\alpha 1}^2$  variance for quantitative trait
- $\sigma_{\alpha 2}^2$  variance for binary trait
- $n_{SNP}$  number of SNPs
- $\sigma_{\varepsilon 1}^2$  error variance

# Results

## Correlation EBV-GEBV

All animals		Quatitative trait				
		EBV	EBV Foulley	EBV Foulley kinship	GEBV	GEBV Foulley
Binary trait	EBV	1	0,15	0,14	0,67	0,19
	EBV Foulley	0,04	1	0,40	0,58	0,20
	EBV Foulley kinship	0,03	0,77	1	0,49	0,12
	GEBV	0,78	0,29	0,29	1	0,34
	GEBV Foulley	-0,05	0,13	0,11	-0,12	1

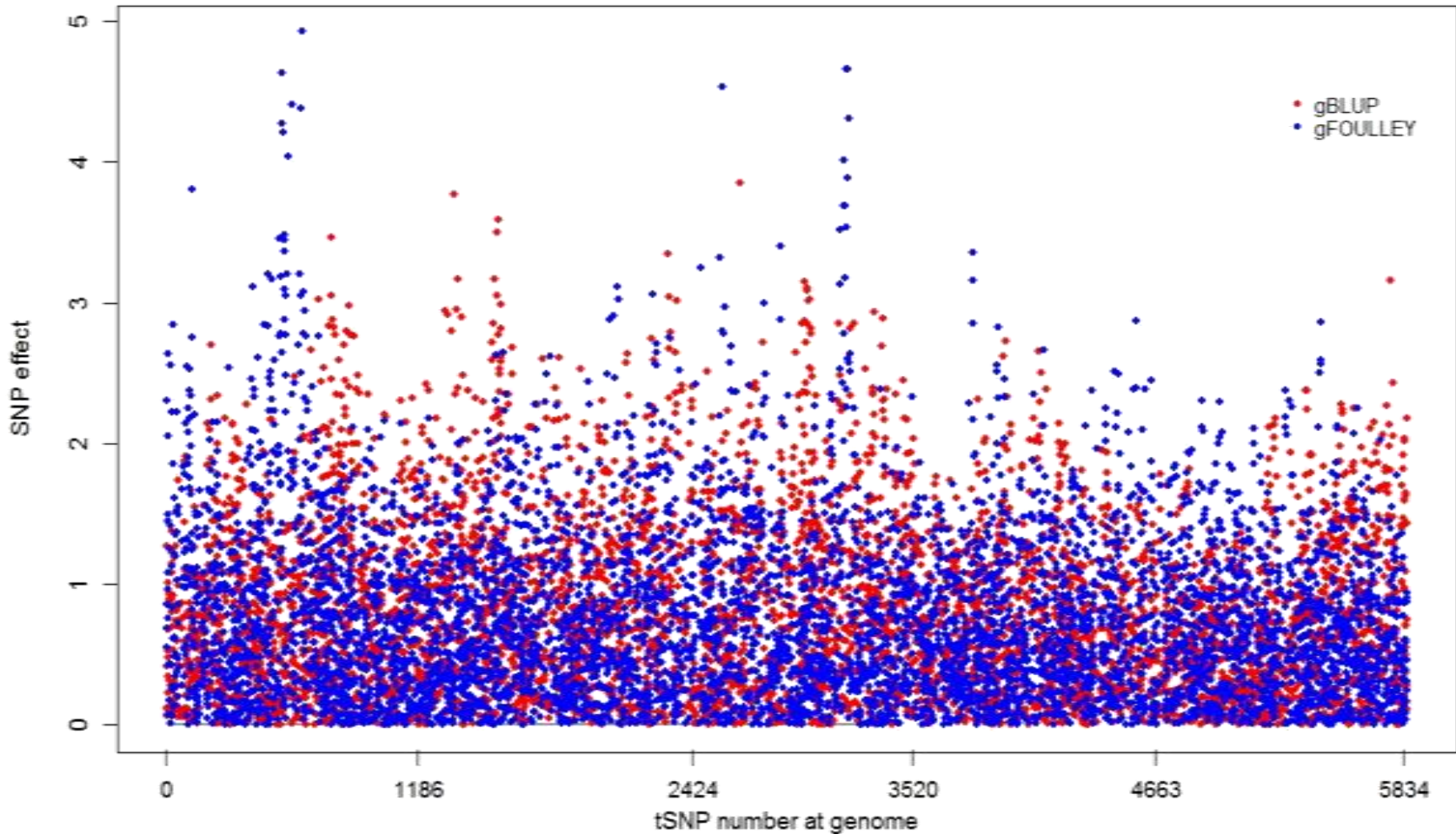
# Results

## Correlation EBV-GEBV

Phenotyped animals		Quatitative trait				
		EBV	EBV Foulley	EBV Foulley kinship	GEBV	GEBV Foulley
Binary trait	EBV	1	0,42	0,36	0,65	0,20
	EBV Foulley	0,25	1	0,45	0,70	0,26
	EBV Foulley kinship	0,24	0,84	1	0,58	0,16
	GEBV	0,84	0,37	0,37	1	0,35
	GEBV Foulley	-0,14	0,20	0,16	-0,19	1

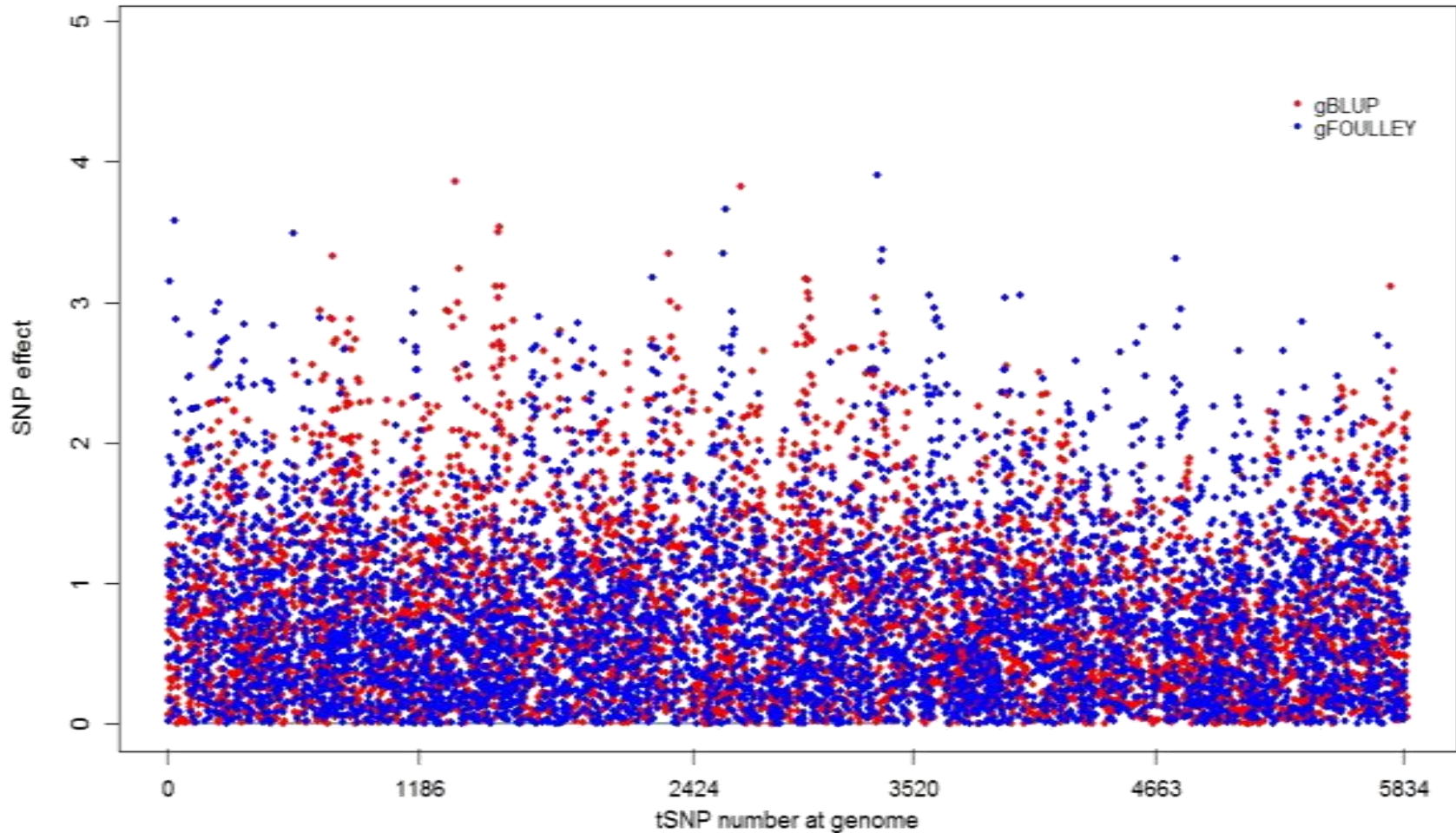
# SNP effects - continuous trait

$$r_{gBLUP, gFOULLEY} = \mathbf{0.13}$$



# SNP effects - binary trait

$$r_{g\text{BLUP},g\text{FOULLEY}} = \mathbf{-0.18}$$



# Conclusions

- additive effects of SNPs in genome were not the same for the continuous and the binary trait
- higher correlations between EBV-GEBV for binary trait
- gFouley overestimates SNP effects
- very low correlation for binary trait and gFouley model
- gFouley model is promising estimation method for joint analysis of continuous and binary trait - needs more investigation
- gFouley computationally demanding



Thank you for attention