

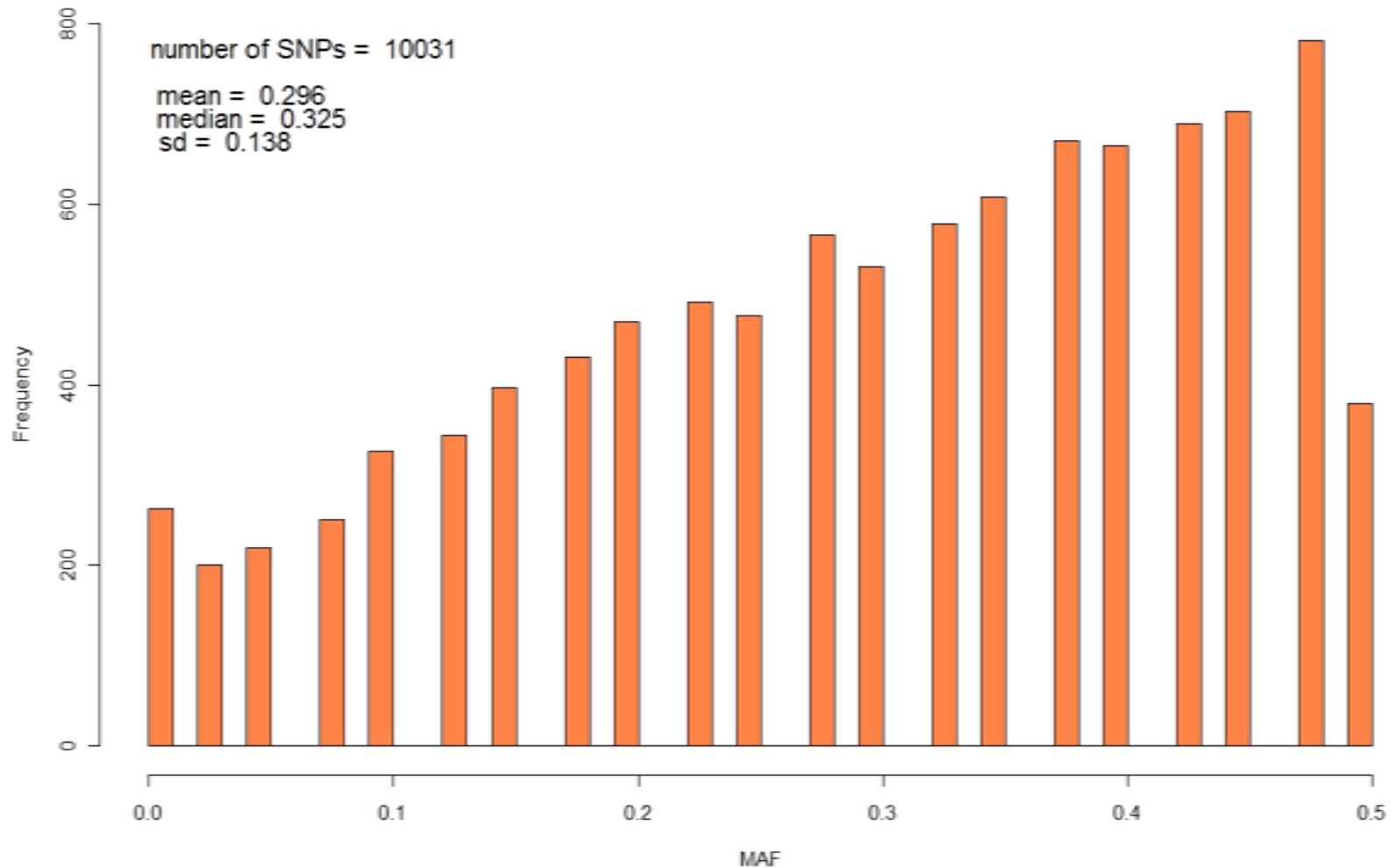
# THE ESTIMATION OF SNP EFFECTS ON A BINARY AND A QUANTITATIVE TRAIT

Kacper Żukowski, Joanna Szyda

# The aim of the work

- comparison effects of Single Nucleotide Polymorphisms (SNP) on a binary and a quantitative trait
- choose most significant SNPs

# Material - MAF



# SNP selection

	10031
MAF > 1%	-264

---

**9767 SNPs in training**

# gBLUP

$$y = \mu + Za + e$$

- ▣  $y$  EBVs vector estimated by standard animal model
- ▣  $\mu$  overall mean
- ▣  $a$  random SNP effects assuming  $a \sim N\left(0, I \frac{\sigma_a^2}{n}\right)$
- ▣  $Z$  design matrix for random SNPs effects the elements given by -1, 0 and 1 for SNP genotype 11, 12(21) and 22
- ▣  $e$  random errors assuming
- ▣  $n$  number of SNPs  $e \sim N(0, I\sigma_e^2)$

# gBLUP

- ▣ All animals (2326)

binary trait as classifier:

- ▣ affected animals (687)
- ▣ unaffected animals (1639)

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$G = I \frac{\hat{\sigma}_\alpha^2}{n}$$

$$R = I\sigma_e^2$$

# Modification of gBLUP (CgBLUP)

$$y = \mu + Z_{unaff} a_{unaff} + Z_{aff} a_{aff} + e$$

- $a$  random SNP effects assuming for unaffected animals  $a_{unaff} \sim N\left(0, I \frac{\hat{\sigma}_\alpha^2}{n}\right)$
- $a$  random SNP effects assuming for affected animals  $a_{aff} \sim N\left(0, I \frac{\hat{\sigma}_\alpha^2}{n}\right)$
- $Z_{unaff}$  and  $Z_{aff}$  design matrix for random SNPs effects
- $\hat{\sigma}_\alpha^2$  additive variance for quantitative trait

# Modification of gBLUP

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_{unaff} & X'R^{-1}Z_{aff} \\ Z'_{unaff}R^{-1}X & Z'_{unaff}R^{-1}Z_{unaff} + G^{-1} & Z'_{unaff}R^{-1}Z_{aff} \\ Z'_{aff}R^{-1}X & Z'_{aff}R^{-1}Z_{unaff} & Z'_{aff}R^{-1}Z_{aff} + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ a_{unaff} \\ a_{aff} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'_{unaff}R^{-1}y \\ Z'_{aff}R^{-1}y \end{bmatrix}$$

$$G = I \frac{\hat{\sigma}_\alpha^2}{n}$$

$$R = I\sigma_e^2$$

# Logistic regression model

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 Z_i$$

- $y$  a probability of being affected,
- $\beta_0$  intercept
- $\beta_1$  additive effect of  $i^{\text{th}}$  SNP
- $Z_i$   $i^{\text{th}}$  column for SNP in incidence matrix  $Z_i$

# Significance testing

- for gBLUP and CgBLUP models
  - p-value was calculated from normal distribution

- for Logistic regression model
  - Likelihood ratio test is given by:

$$LRT = 2(\ln L_1 - \ln L_2)$$

- p-value was calculated from chi-squared distribution

# Significant SNPs

for  $\alpha=0.05$

$$H_0 : \hat{\beta}_1 = 0 \quad H_1 : \hat{\beta}_1 \neq 0$$

Chromosome	gBLUP			CgBLUP		Logistic regression model*
	all	affected	unaffected	affected	unaffected	
1	275	296	162	169	192	657
2	177	170	167	184	172	533
3	195	235	181	222	219	557
4	126	119	156	222	207	528
5	89	80	181	167	159	553

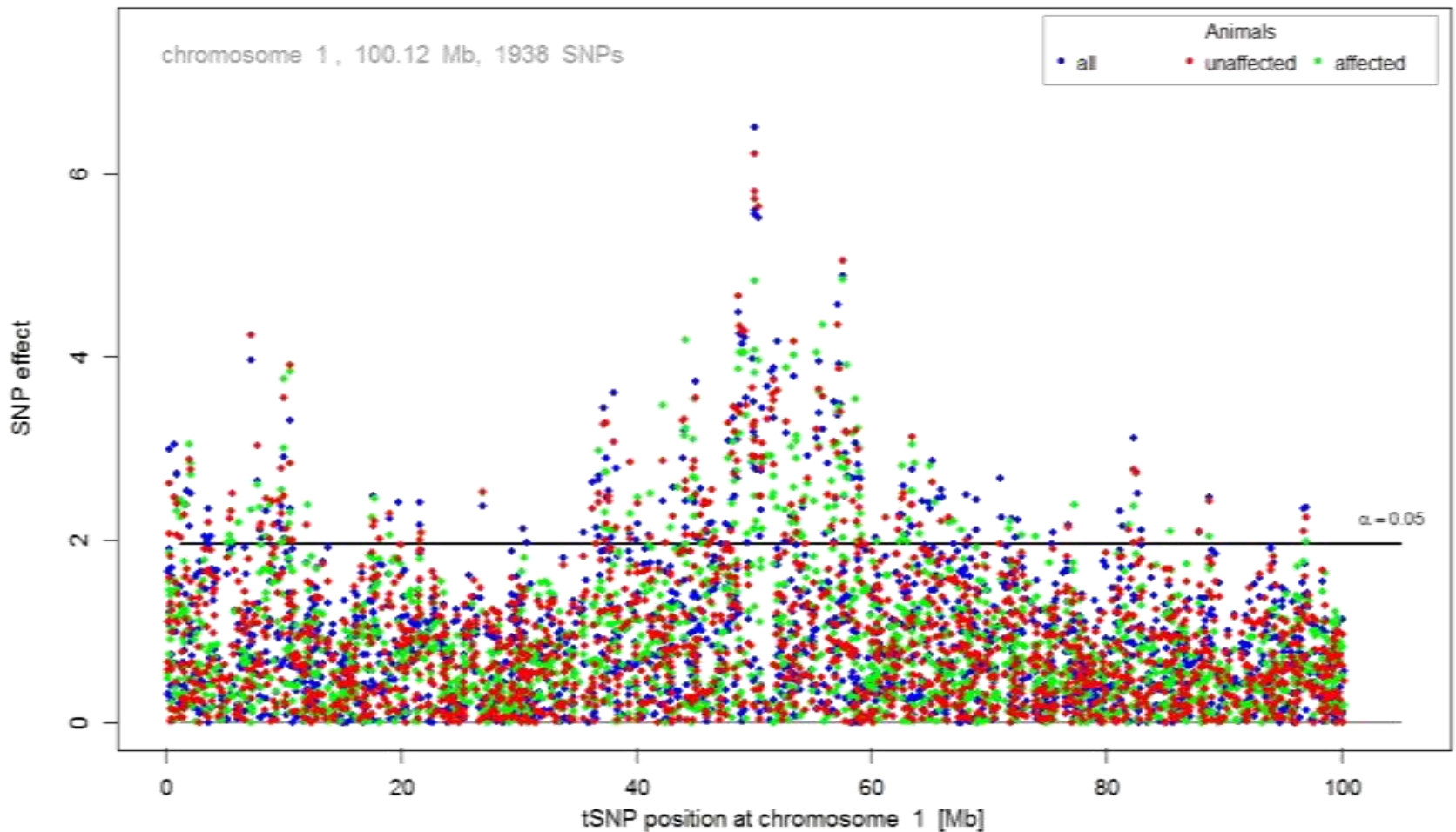
\*after Bonferroni correction

# Correlation between SNPs effects

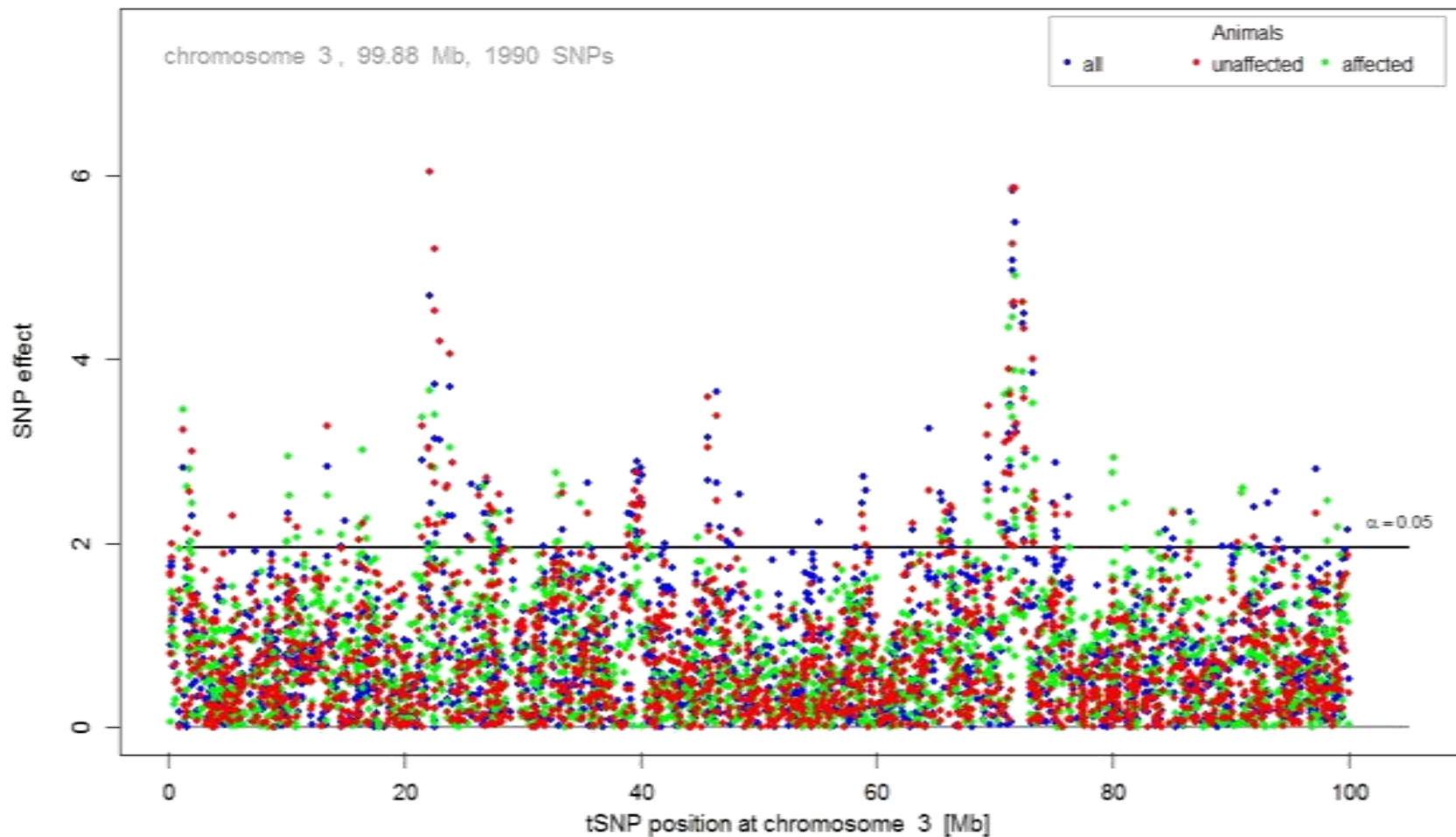
gBLUP all	gBLUP affected	gBLUP unaffected	CgBLUP affected	CgBLUP unaffected	Logistic regression model	
1,00	<b>0,80</b>	<b>0,95</b>	-0,02	0,07	0,31	<b>gBLUP all</b>
	1,00	<b>0,66</b>	-0,01	0,08	0,30	<b>gBLUP affected</b>
		1,00	-0,02	0,06	0,30	<b>gBLUP unaffected</b>
			1,00	<b>0,83</b>	0,01	<b>CgBLUP affected</b>
				1,00	0,03	<b>CgBLUP unaffected</b>
					1,00	<b>Logistic regression model</b>



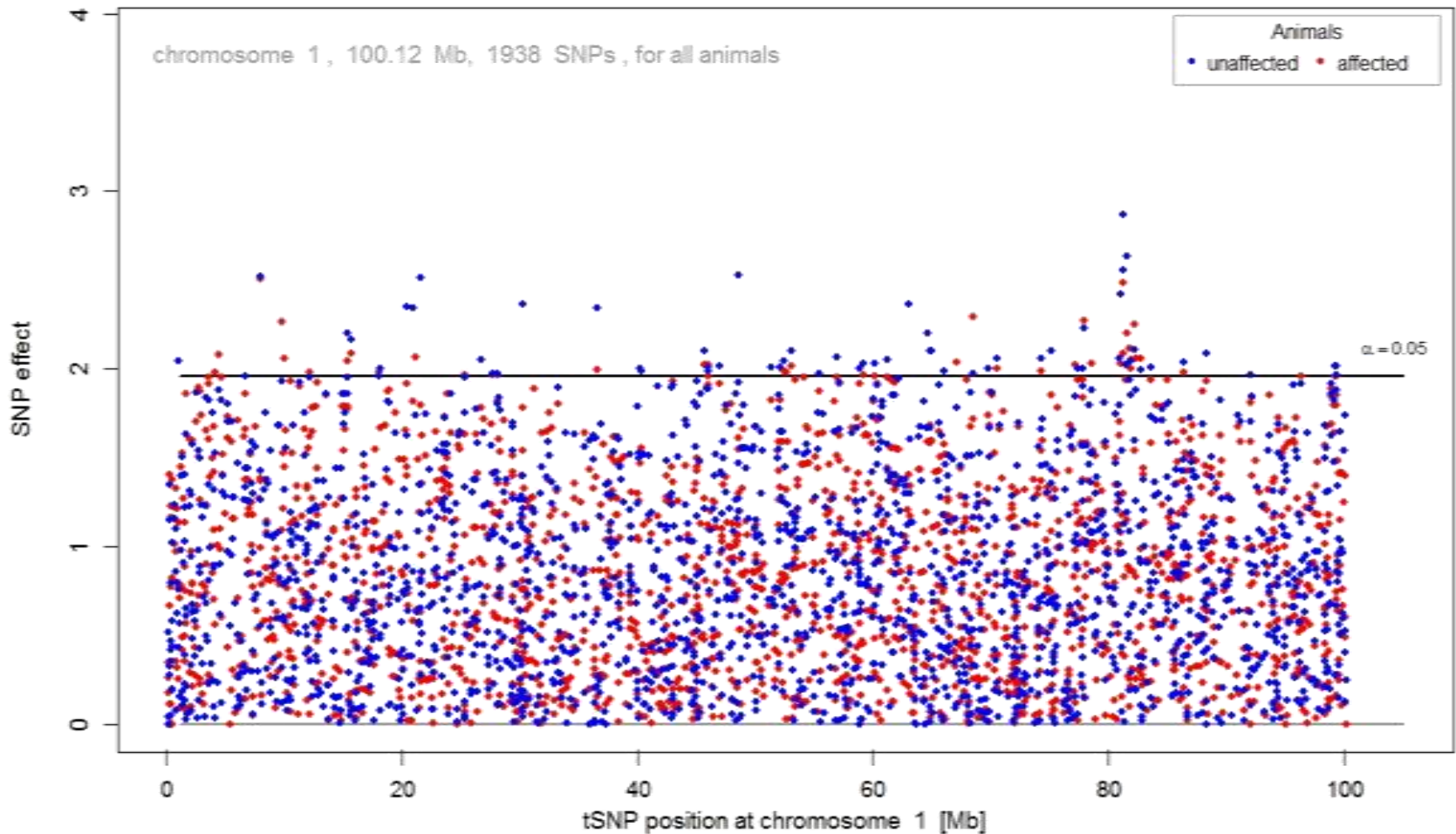
# Results, gBLUP



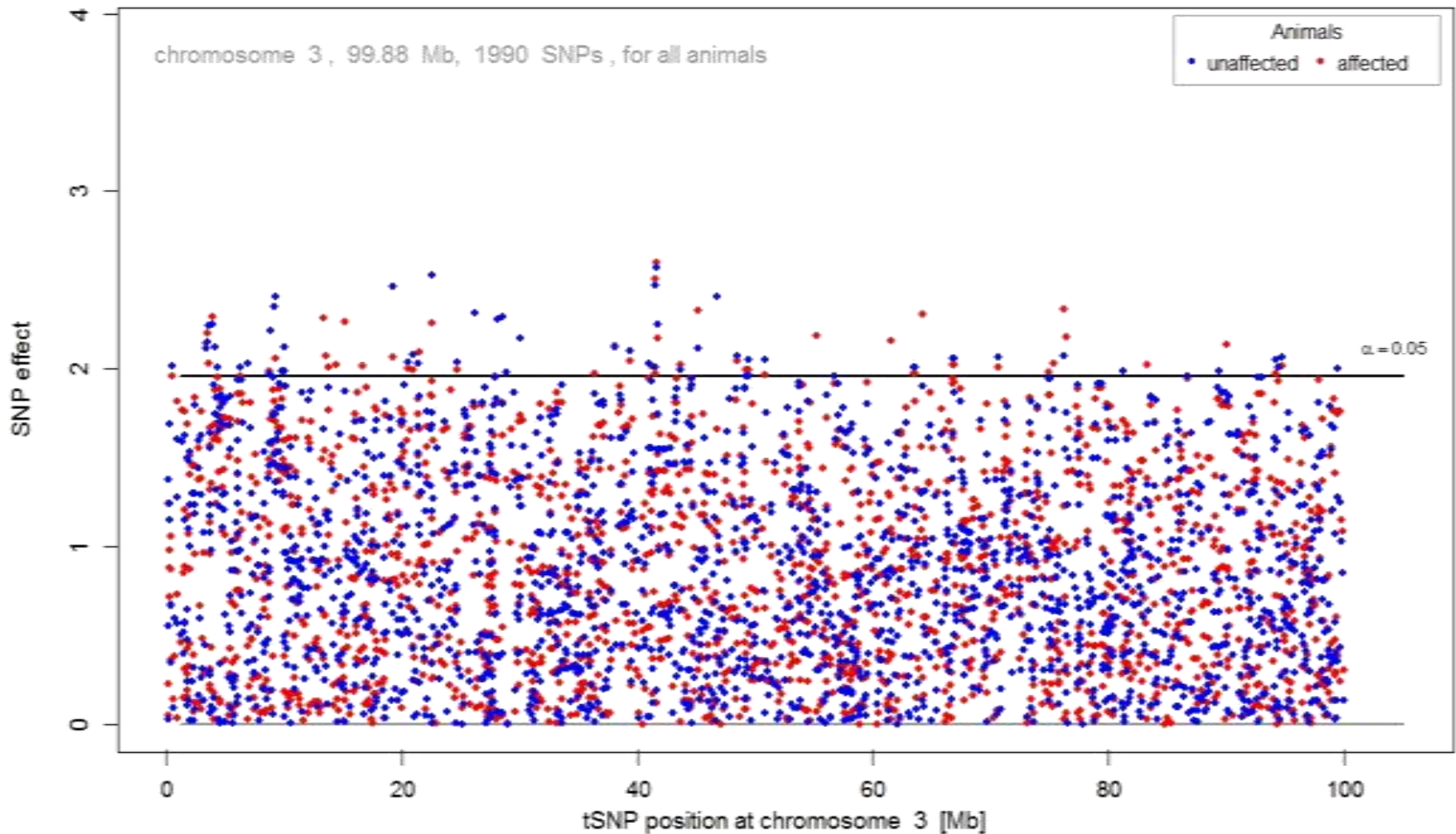
# Results, gBLUP



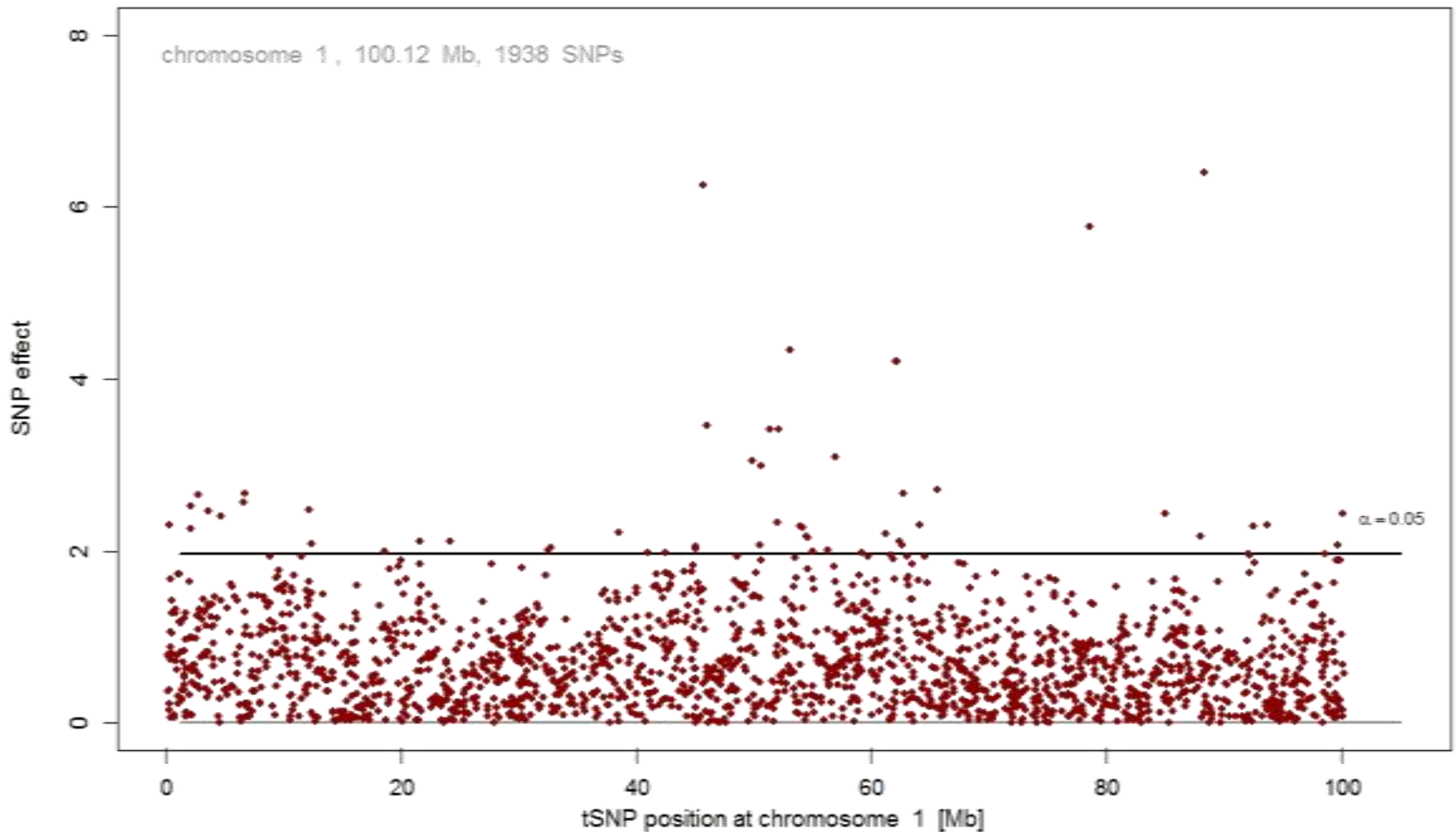
# Results, CgBLUP



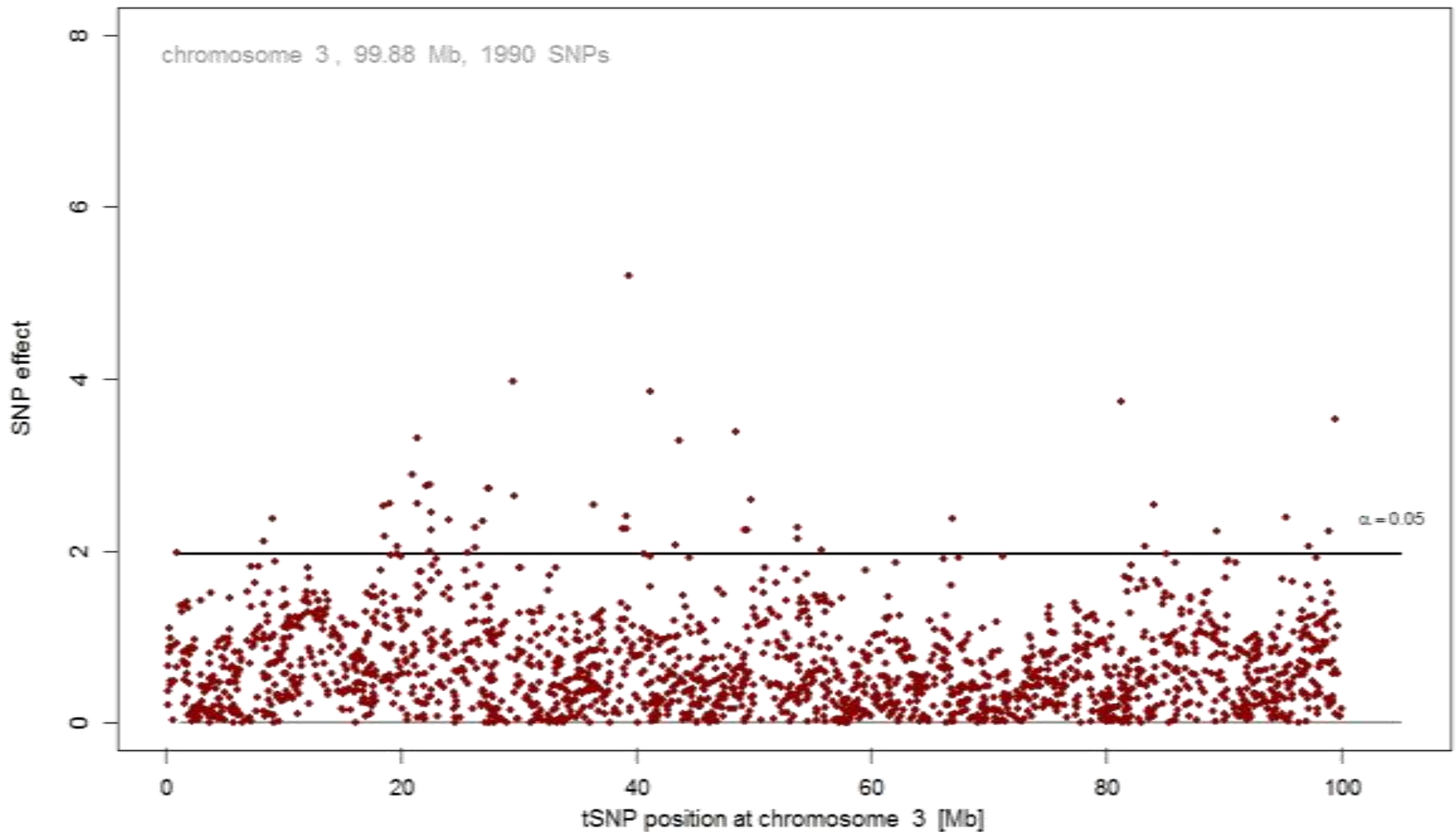
# Results, CgBLUP



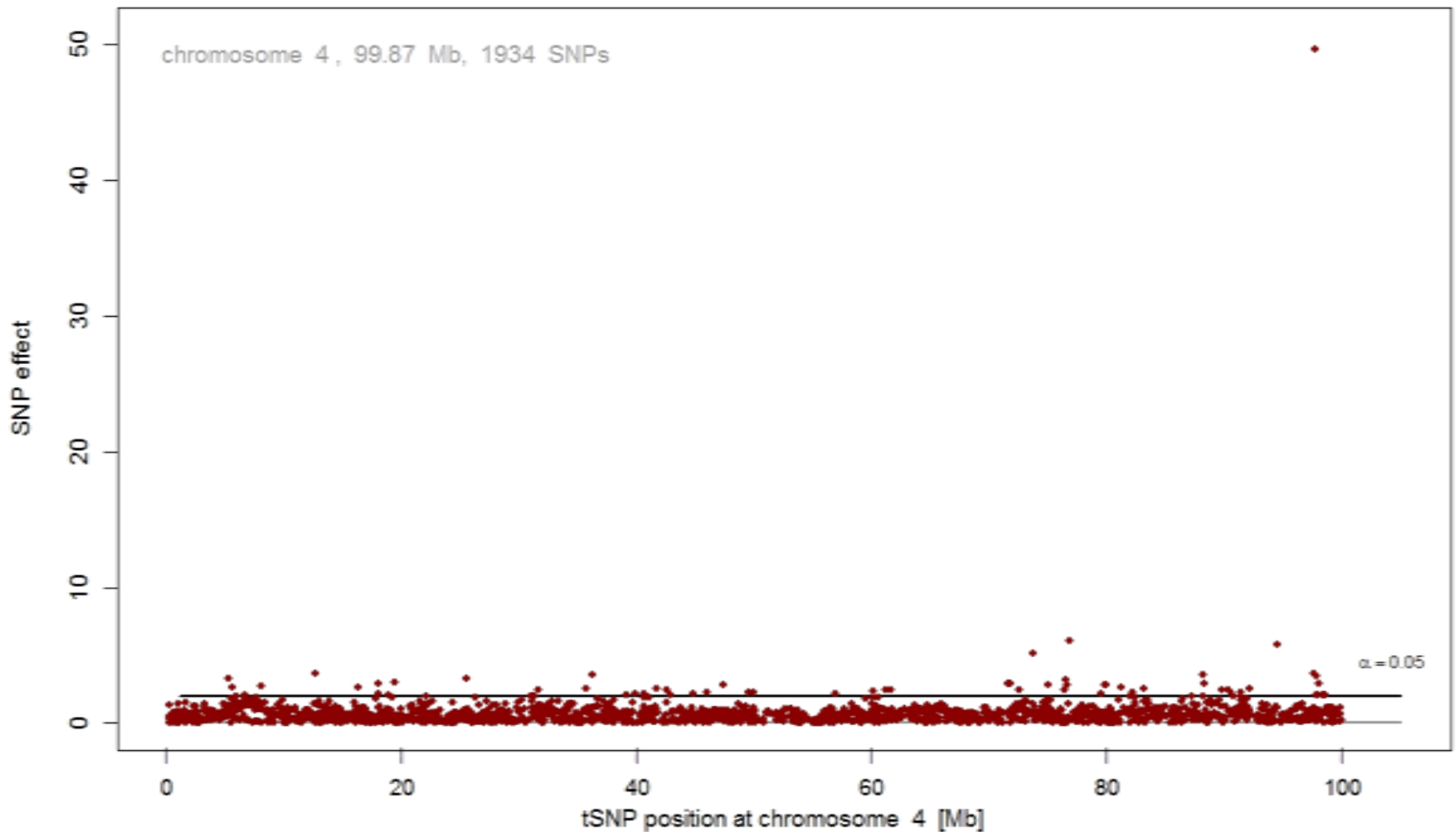
# Results, LOGIT



# Results, LOGIT



# Results, LOGIT



# Conclusions

- Gene located on 22,0-22,5 Mb on chromosome 3 could have influence on a quantitative and a binary trait simultaneously
- High correlation within gBLUP and CgBLUP models in contrast to between models

# Thank you for attention

