

# Using estimated allele frequency changes to map genomic regions under selection in farm animals

**Henner Simianer<sup>1</sup> and Eduardo Pimentel<sup>2</sup>**

<sup>1</sup> Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Germany

<sup>2</sup> Institut für Tierzucht, Bayerische Landesanstalt für Landwirtschaft, Grub, Germany





# Background

- ⇒ According to theory of natural selection, alleles with a positive effect on fitness should increase in frequency under selection
- ⇒ Alleles that undergo a rapid increase in frequency might have a selective advantage
- ⇒ If we can identify regions carrying such alleles, this might improve our understanding of – natural and anthropogenic - selection

But:

- ⇒ what is a ‚rapid increase in frequency‘?
- ⇒ how can we measure it?

# Some ideas ...



...how to quantify allele frequency change over time in a population:



Have two samples, one from time  $t = 0$  and one from time  $t = 1$ , estimate allele frequencies at the two timepoints and then  $\widehat{\Delta p}_i = \hat{p}_{i1} - \hat{p}_{i0}$



Have one sample from the present population, use the length of the surrounding haplotype as a proxy for the age of an allele  
→ frequent alleles with long surrounding haplotypes have risen rapidly in frequency (idea underlying the EHH statistic, Sabeti et al., 2002)



Have one sample from the present population and pedigree data  
→ estimate allele frequencies in the present and the base population (using the approach of Gengler et al., 2007) and calculate the difference

# Material

- 2'294 progeny tested Holstein Friesian elite bulls (details see Pimentel et al. 2011)
- 21'646 animals in the pedigree.
- Oldest ancestor born in 1906
- average number of discrete generation equivalents: 4.05
- Conventional BLUP breeding values for 12 production, fitness, and fertility traits (positive values = desirable)
- All bulls were genotyped with the Illumina BovineSNP50 BeadChip
  - 39'557 autosomal SNPs after filtering

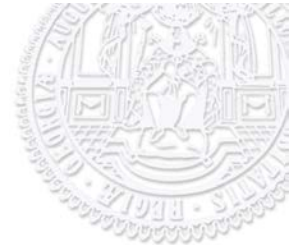


# Processing



## For each of the 12 traits:

- SNP-effects were estimated via random regression BLUP (Meuwissen et al., 2001)
- For each SNP, the allele with a positive effect was chosen and standardized  
→ standardized positive allele effect  $\alpha_i$  for 39'557 SNPs
- At each SNP the present and past frequency of the positive allele was estimated  
→ estimated allele frequency change  $\Delta p_i$  of the positive allele at 39'557 SNPs



According to Wright's (1937) classical formula

$$\Delta p_i = \frac{p_i(1 - p_i)}{2\bar{w}} \frac{\partial \bar{w}}{\partial p_i}$$

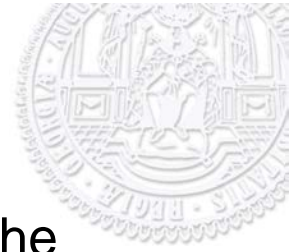
the change of allele frequency  $\Delta p_i$  at locus  $i$  is:

⇒ a quadratic function of the initial allele frequency  $p_i$  and

⇒ the change of average fitness  $\bar{w}$  caused by a change of allele frequency

Estimated SNP effect  $\alpha_i$  can be taken as a proxy for  $\frac{\partial \bar{w}}{\partial p_i}$

Can we find empirical evidence supporting Wright's formula using genomic data?



# Some hypotheses

- 1) For traits under selection, alleles with effects in the direction of the genetic trend increase in frequency over time

$$H_0 : \overline{\Delta p_i} = 0 \quad vs. \quad H_1 : \overline{\Delta p_i} \neq 0$$

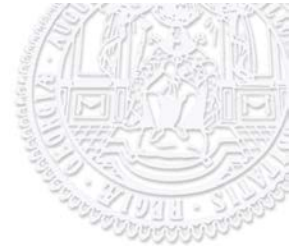
- 2) The change of allele frequency is proportional to effect size

$$H_0 : cor(\alpha_i, \Delta p_i) = 0 \quad vs. \quad H_1 : cor(\alpha_i, \Delta p_i) \neq 0$$

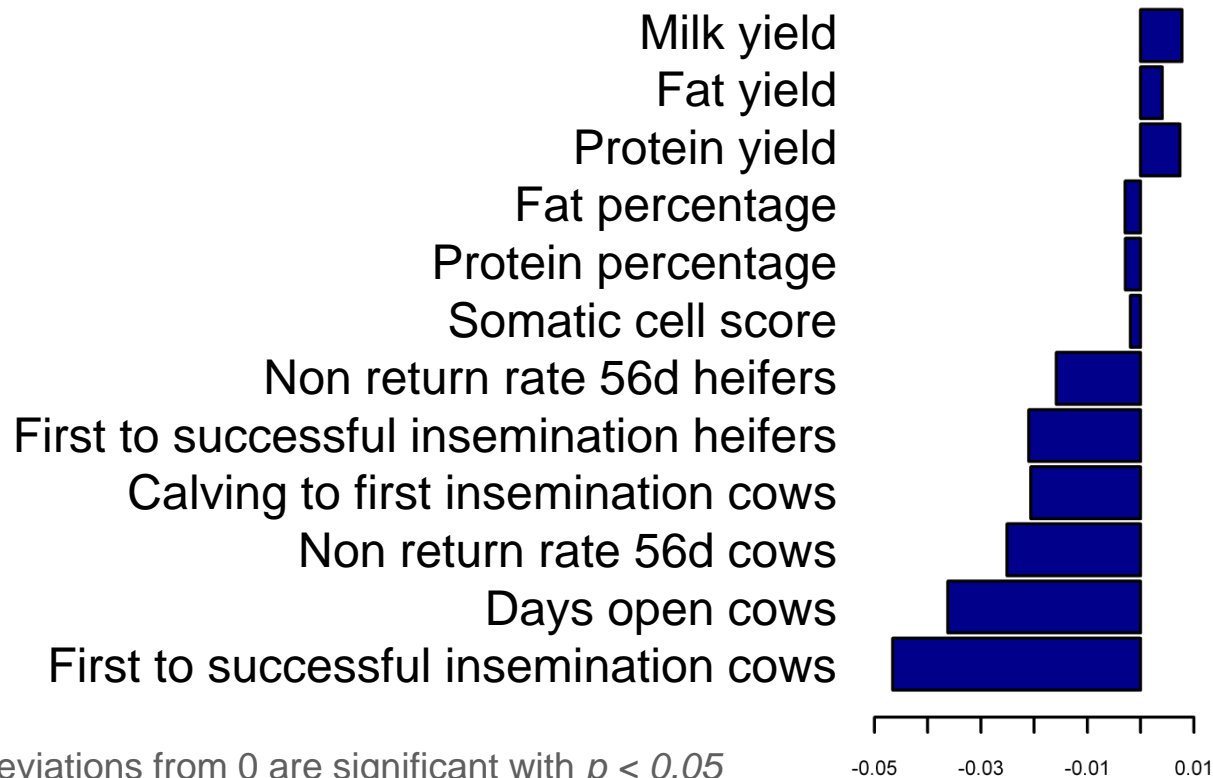
- 3) The change of allele frequency is a quadratic function of the initial allele frequency, thus with  $\Delta p_i = a_0 + a_1 p_{i0} + a_2 p_{i0}^2$

$$H_0 : a_1, a_2 = 0 \quad vs. \quad H_1 : a_1, a_2 \neq 0$$

# Results



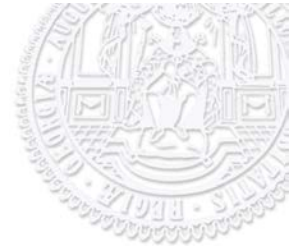
Average  $\overline{\Delta p_i}$  frequency change  
of the positive allele



All deviations from 0 are significant with  $p < 0.05$

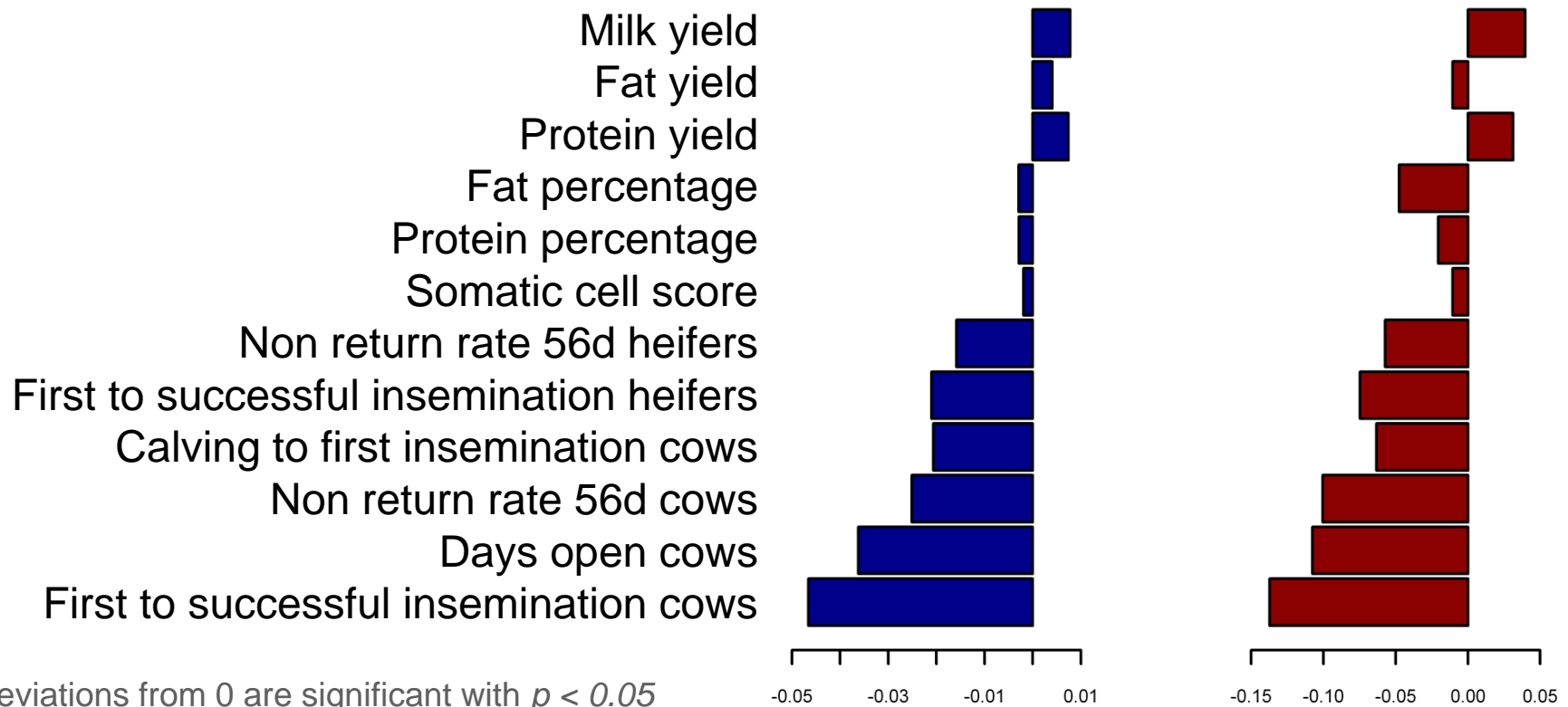


# Results



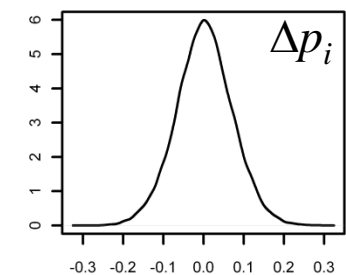
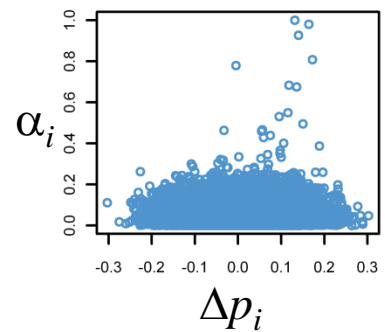
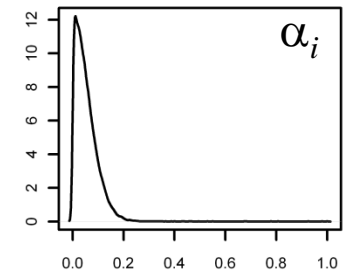
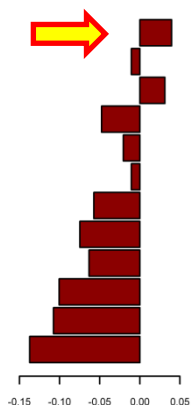
Average correlation of SNP effect size and frequency change of the positive allele  $\text{cor}(\alpha_i, \Delta p_i)$

Average allele frequency change  $\Delta p_i$  of the positive allele

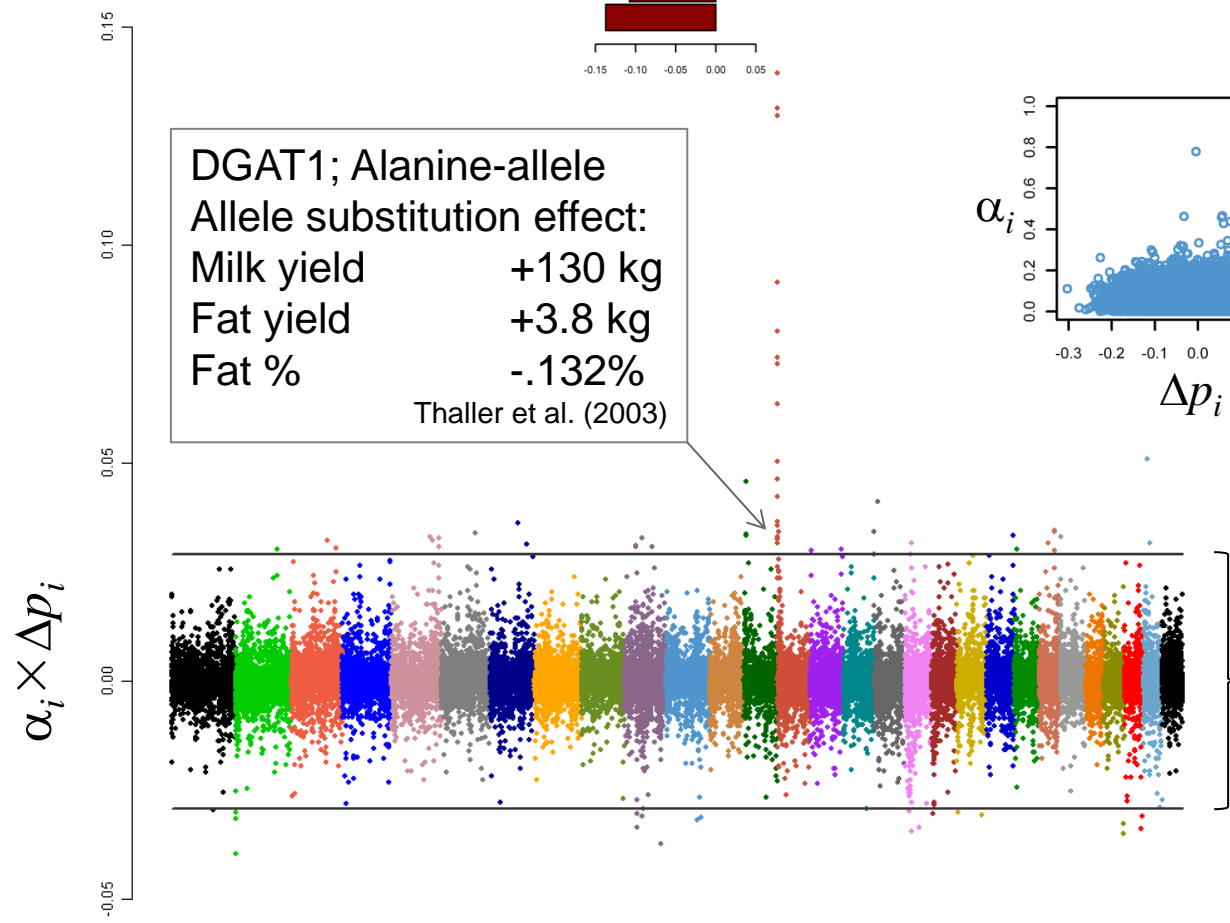


All deviations from 0 are significant with  $p < 0.05$

# Milk yield



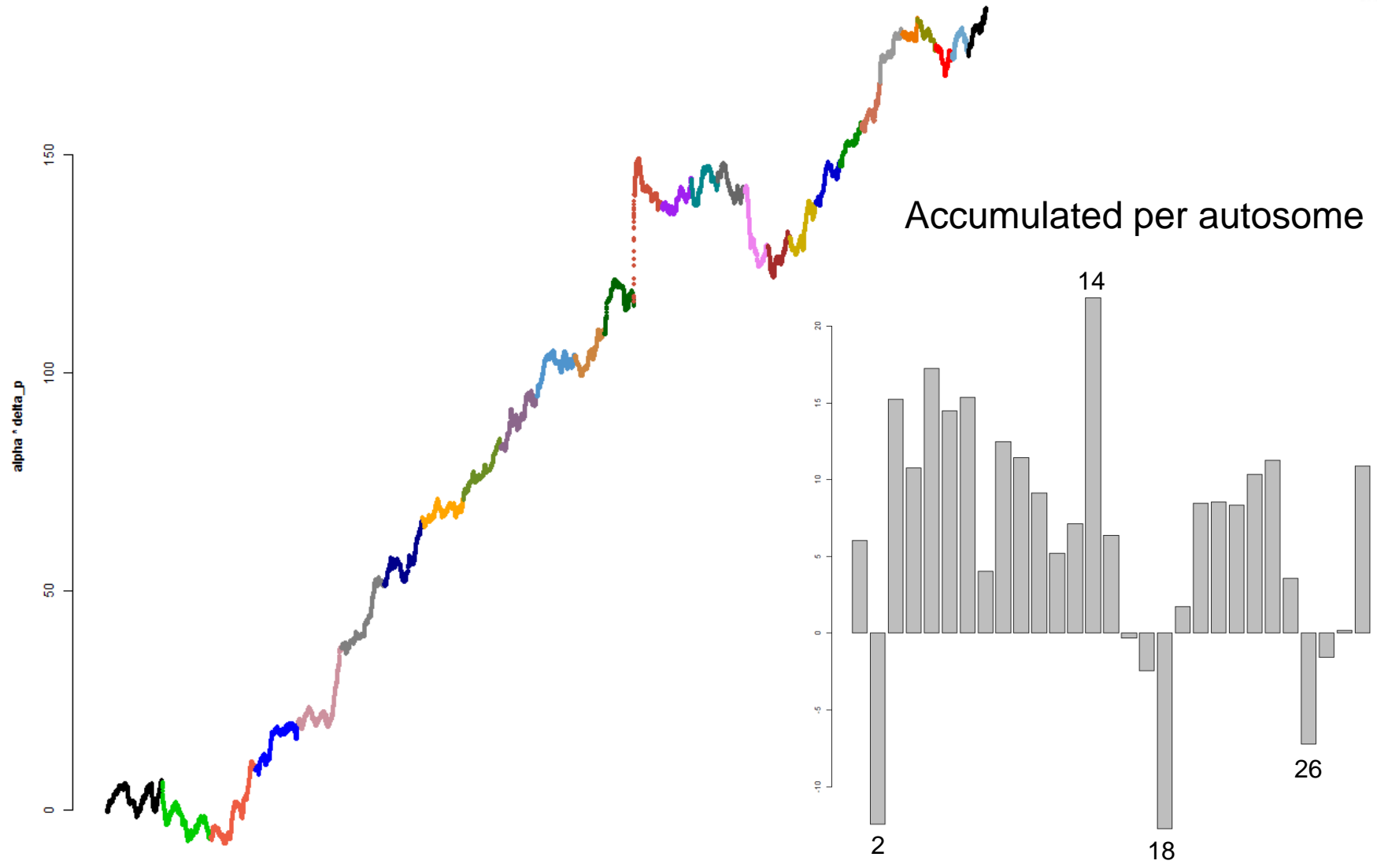
DGAT1; Alanine-allele  
 Allele substitution effect:  
 Milk yield +130 kg  
 Fat yield +3.8 kg  
 Fat % -.132%  
 Thaller et al. (2003)



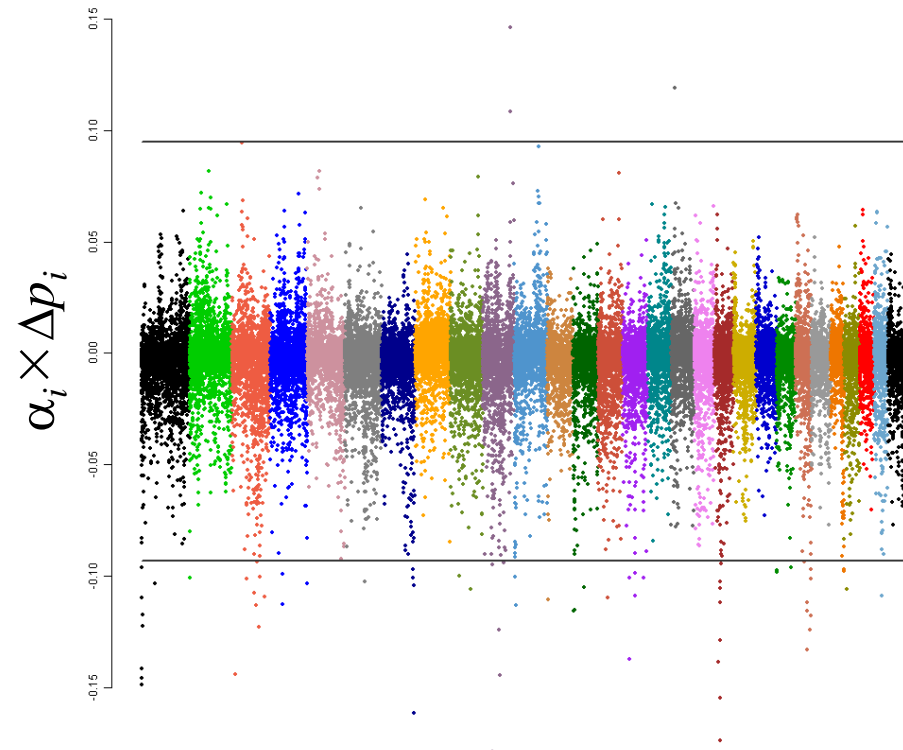
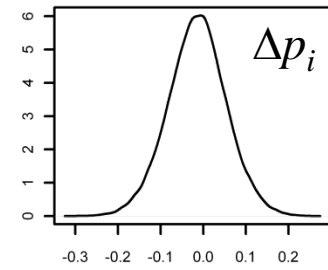
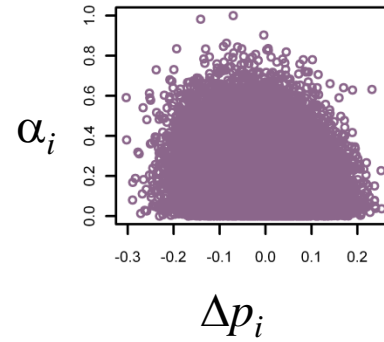
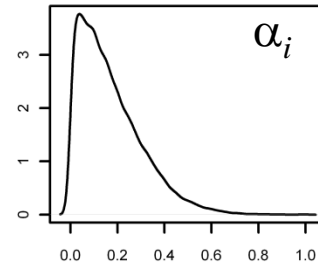
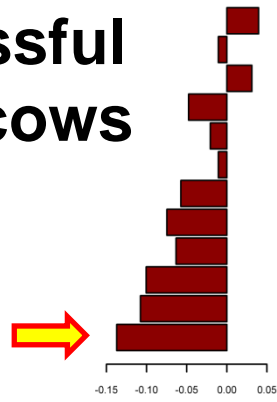
Permutation based genome-wide 95% confidence limits

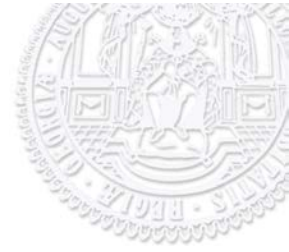


# Frequency change x effect size accumulated across the genome – trait **milk yield**

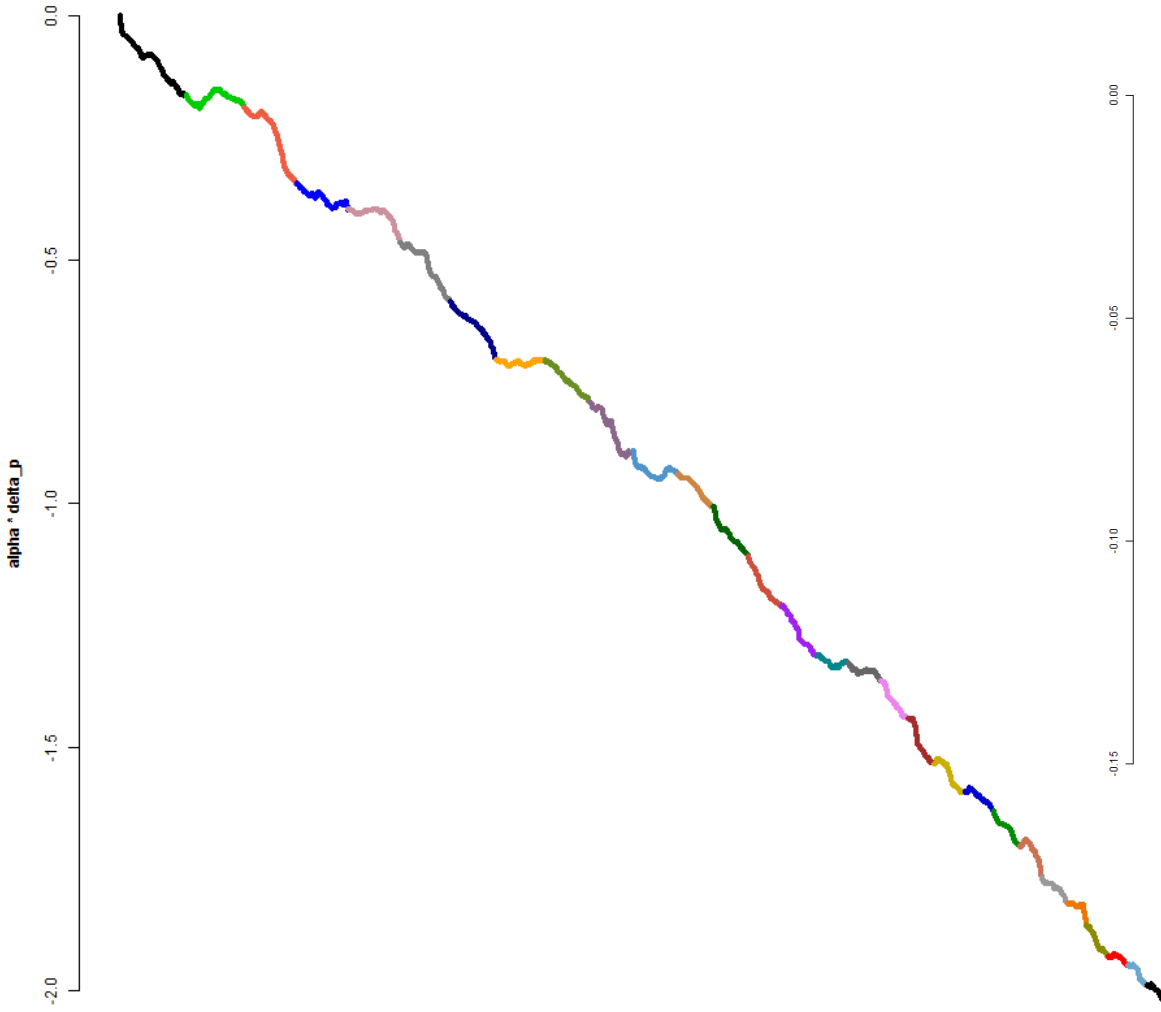


# First to successful insemination cows

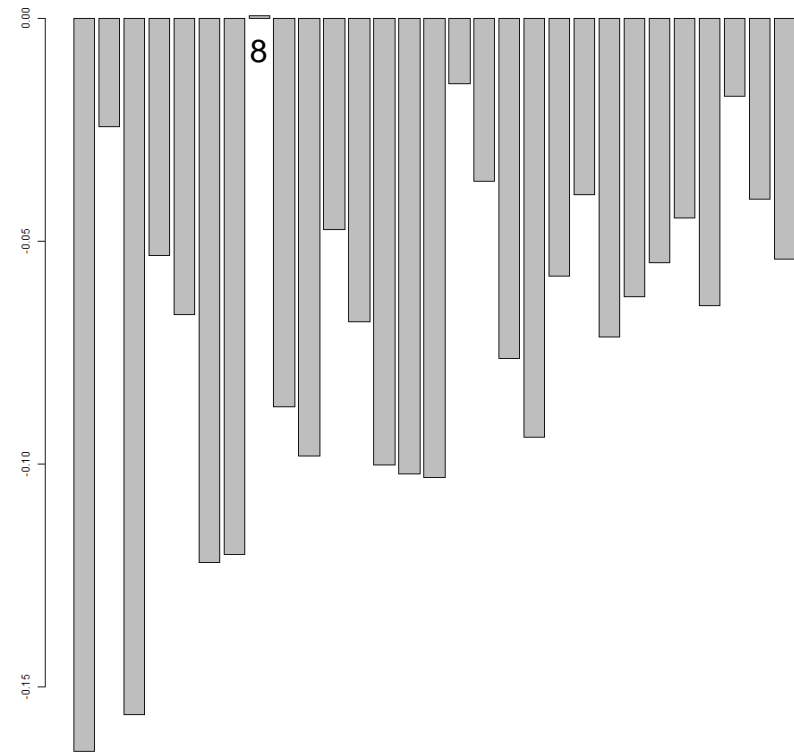


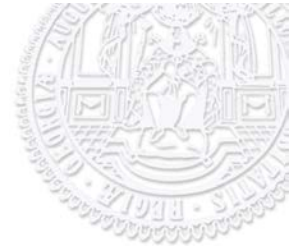


# Frequency change x effect size accumulated across the genome – trait ‘first to successful insemination cows’



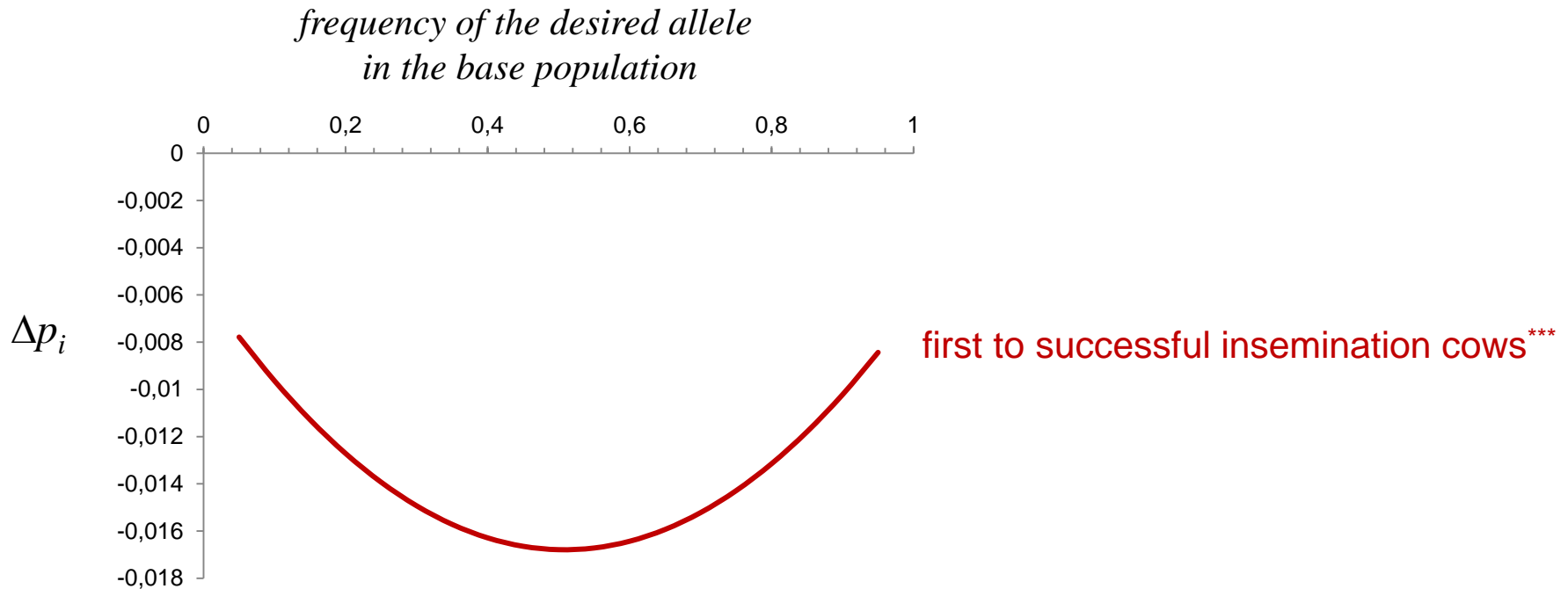
Accumulated per autosome





Is the allele frequency change a quadratic function of the initial allele frequency?

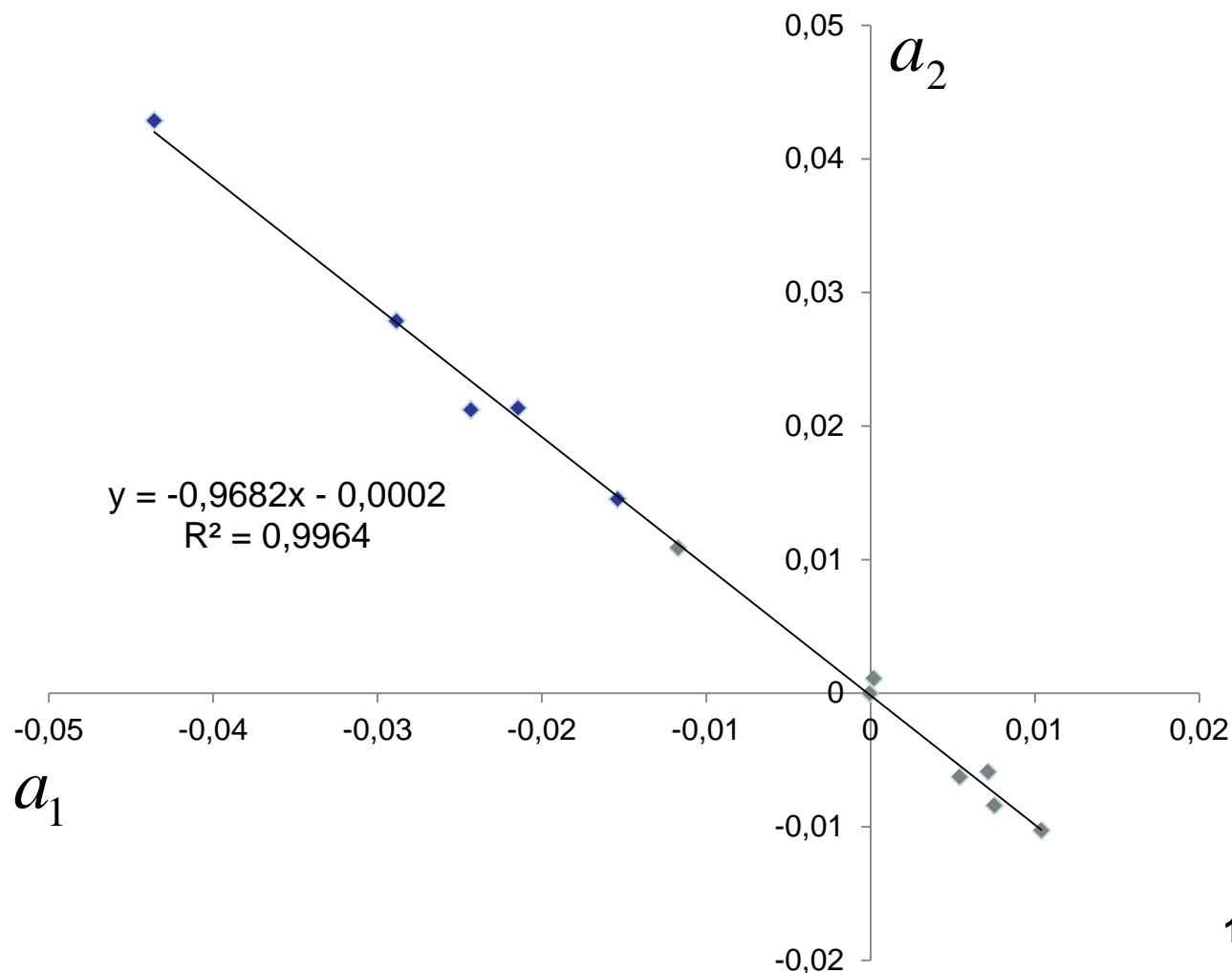
$$\Delta p_i = \frac{p_i(1 - p_i)}{2\bar{w}} \frac{\partial \bar{w}}{\partial p_i}$$

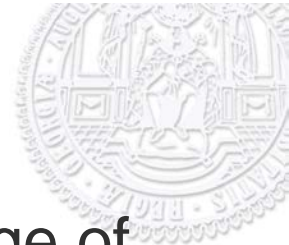




$$\Delta p_i = a_o + a_1 p_{i0} + a_2 p_{i0}^2$$

If  $a_1 = (- a_2)$  the maximum of  $|\Delta p_i|$  will be at  $p_{i0} = 0.5$





# Summary and Conclusions

- The suggested approach quantifies the recent change of allele frequency by combining genotype, phenotype, and pedigree information
- Selection in the Holstein Friesian population over the last ~40 years appears to have focused on milk and protein yield, while fertility and fitness traits deteriorated
- Anthropogenic selection dominates natural selection in modern cattle breeding



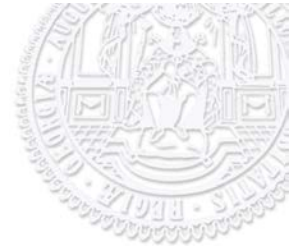


# Summary and Conclusions

- The results confirm Wright's classical formula:
  - For traits under selection, alleles with effects in the direction of the genetic trend increase in frequency over time
  - The change of allele frequency is positively correlated with effect size
  - The change of allele frequency is a quadratic function of the initial allele frequency
- Selection can be mapped to chromosomes and single genes
- Most traits appear to be highly polygenic with a complex pattern of allele effects vs. frequency changes



Eduardo Pimentel



# Acknowledgements

Presented results were obtained within the project **FUGATO-plus GenoTrack** which was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven.

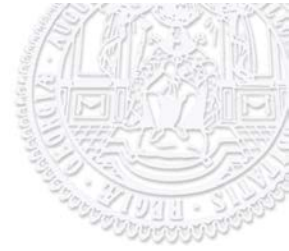


Bundesministerium  
für Bildung  
und Forschung



LOHMANN  
TIERZUCHT

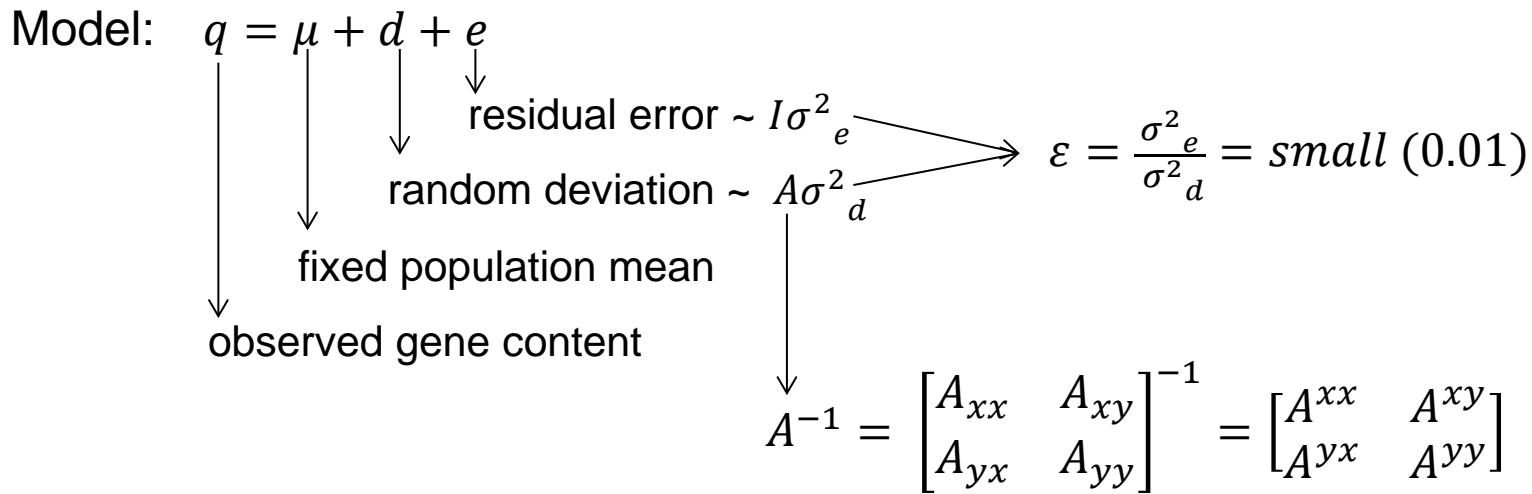
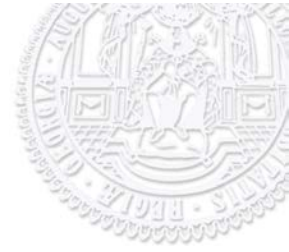
# Estimating allele frequency in the base population



Approach suggested by Gengler et al. (2007)

Basic idea:

- ⇒ Gene content is observed in a genotyped sample
- ⇒ An ungenotyped sample is linked to the genotyped animals via pedigree information
- ⇒ Treat individual gene content (0,  $\frac{1}{2}$  or 1) as a continuous variable and use BLUP to predict the gene content in the ungenotyped individuals



BLUP mixed model equations (MME):

$$\begin{bmatrix} N_y & 0 & 1' \\ 0 & \varepsilon A^{xx} & \varepsilon A^{xy} \\ 1 & \varepsilon A^{yx} & I + \varepsilon A^{yy} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \widehat{d}_x \\ \widehat{d}_y \end{bmatrix} = \begin{bmatrix} \Sigma q_y \\ 0 \\ q_y \end{bmatrix}$$

Solve MME for each SNP locus. Then:

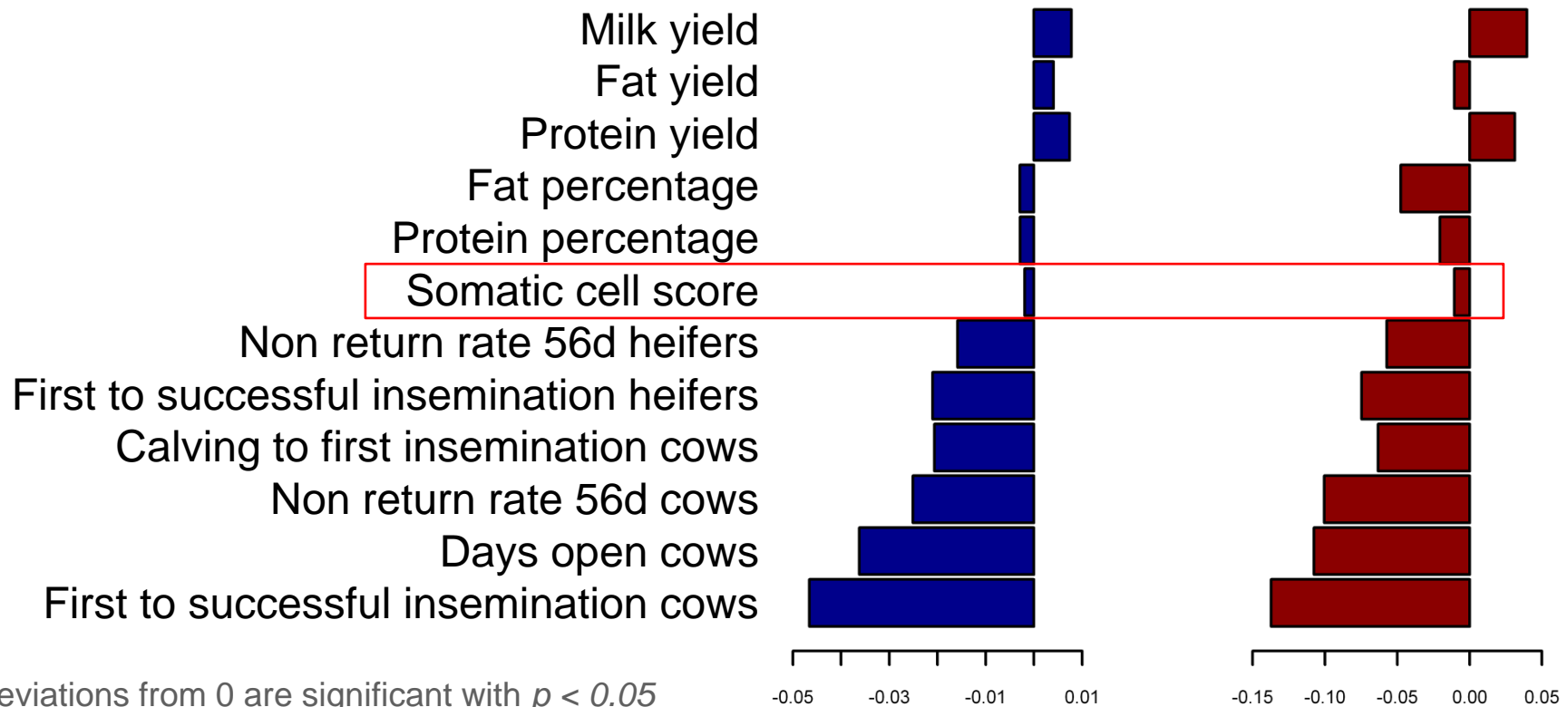
- ⇒ estimate gene content at locus  $i$  in base population  $x$  as  $\hat{p}_{ix} = \frac{\Sigma(\hat{\mu} + \widehat{d}_x)}{N_x}$
- ⇒ estimate gene content at locus  $i$  in the current population  $y$  as  $\hat{p}_{iy} = \frac{\Sigma(\hat{\mu} + \widehat{d}_y)}{N_y}$
- ⇒ calculate the estimated allele frequency change as  $\Delta p_i = \hat{p}_{iy} - \hat{p}_{ix}$

# Results

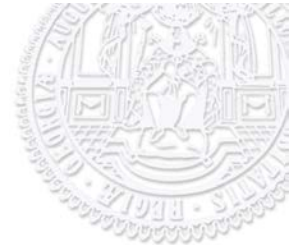


Average correlation of SNP effect size and frequency change of the positive allele  $\text{cor}(\alpha_i, \Delta p_i)$

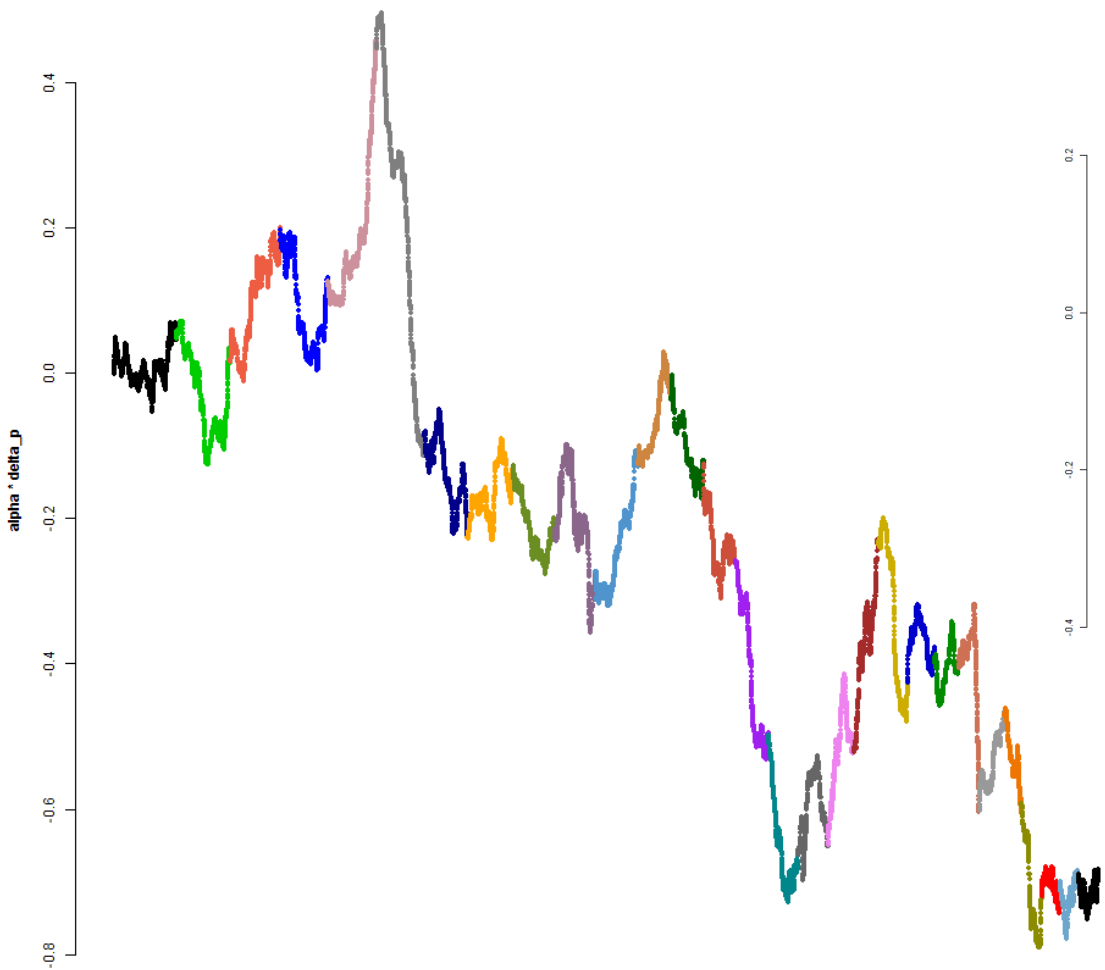
Average allele frequency change  $\Delta p_i$  of the positive allele



All deviations from 0 are significant with  $p < 0.05$



# Frequency change x effect size accumulated across the genome – trait **somatic cell score**



Accumulated per autosome

