Identification of causal mutations underlying feet and leg disorders in cattle

T. Suchocki[1,2], C.H. Egger-Danner[3], H. Schwarzenbacher[3], M. Mielczarek[1,2] and J. Szyda[1,2]

[1] Wroclaw University of Environmental and Life Sciences, Biostatistics Group, Poland
[2] National Research Institute of Animal Production, Cracow, Poland
[3] ZuchtData EDV Dienstleistungen GmbH, Vienna, Austia

Aim

Combination of SNP chip and whole genome sequence information to analyze feet and legs phenotypes directly measured on cows, in order to be able to consider also non-additive effects of variants.

Material

Animals:

- 1,998 Austrian Fleckvieh
- 979 Austrian Braunvieh

Traits collected until 100^{th} day-in-milk:

- hoof health status defined by a vet (HSV)
- total number of hoof disorders (NHD)
- hoof health status defined by a farrier (HSF)

Genotypes:

- 74,762 SNPs from GeneSeek® Genomic Profiler™ HD panel
- SNP selection criteria comprised the minor allele frequency (MAF) of at least 0.01 and the technical quality of genotyping expressed by a minimum call rate of 99%.

number of records per cow ranged from 1 to 8

Methods - Estimation (co)variance components

Model for NHD:

$$y = X\beta + Z_\alpha \alpha + Z_d d + Z_p p + \epsilon,$$

where

- β - fixed effects comprising: a general mean, breed, parity (from 1 to 4 and greater or equal to 5), calving year-season (four years between 2012 and 2015 and season 1 – between October and March and 2 – between April and September); percent of non Holstein-Friesian genes and hoof status
- α - random additive polygenic effect $\left(\alpha \sim \mathcal{N}(0, A\sigma_\alpha^2)\right)$
- d - random vet effect $\left(d \sim \mathcal{N}(0, I\sigma_d^2)\right)$
- p - random pernament environmental effect $\left(p \sim \mathcal{N}(0, I\sigma_p^2)\right)$

Methods - Estimation (co)variance components

Model for HSV and HSF:

$$logit(p) = X\beta + Z_\alpha\alpha + Z_d d + Z_p p,$$

where

- β, α, d and p - the same form as before

Methods - GWAS - Single SNP

Adding the $X_g g$ to the model for calculation (co)variance parameters, where

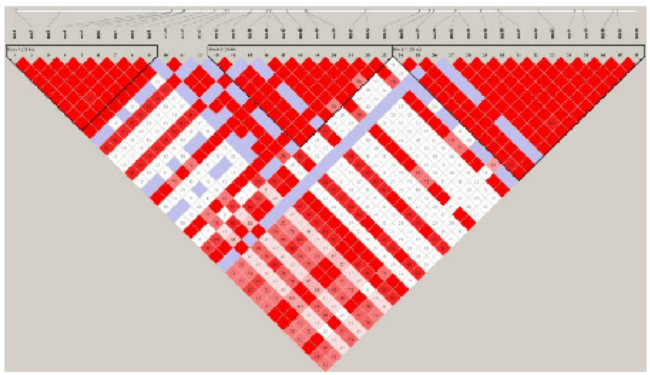
- $X_g = \{-1, 0, 1\}$
- g - vector of an additive SNP effect

Test of significance:

$$W = \frac{\widehat{g}}{SE(\widehat{g})} \sim \mathcal{N}(0, 1)$$

Methods - Multiple testing correction

1. Calculate LD coefficient between all pairs of SNPs and build LD matrix



2. Calculate eigenvalues for LD matrix

$$A = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 7 & 1 \\ 4 & 1 & 4 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 7 & 1 \\ 4 & 1 & 4 \end{bmatrix} + \begin{bmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{bmatrix}$$

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 1 & 4 \\ 1 & 7-\lambda & 1 \\ 4 & 1 & 4-\lambda \end{vmatrix} = 0$$

$$-54\lambda + 15\lambda^2 - \lambda^3 = 0$$

$$-\lambda(\lambda - 6)(\lambda - 9) = 0$$

$$\lambda = 0, 6, 9$$

3. Calculate m_e^1 and m_e^2.

Figure 1: Scheme of calculating m_e^1 and m_e^2

Methods - Multiple testing correction

Methods for calculating effective number of tests (Li and Ji (2005)):

1. $m_e^1 = \sum_{i=1}^{m} (I(\lambda_i \geq 1) + (\lambda_i - \lfloor \lambda \rfloor))$,

where

- λ_i - eigenvalues for pairwise linkage disequilibrium (r^2) matrix between SNPs
- $I(\cdot)$ - indicator variable
- $\lfloor \cdot \rfloor$ - floor function

Li et al. (2012) proposed second version of effective number of tests:

2. $m_e^2 = m - \sum_{i=1}^{m} I(\lambda_i > 1)(\lambda_i - 1)$

Methods - Using WGS

48 whole genome DNA sequences (WGS) of Braunvieh and 30 of Fleckvieh individuals, available through the 1000 Bulls Genome project where used to creation genomic regions, which cover:

- the significant SNP from the HD panel
- selected SNPs from the HD panel flanking the significant SNP
- SNPs from the whole genome sequence inbetween flanking SNPs

Based on significant regions we add $X_{a_1} a_1$, $X_{d_1} d_1$ and $X_{e_1} e_1$ to the model for calculation (co)variance parameters, where

- a_1 - additivie effect of SNP
- d_1 - dominance effect of SNP
- e_1 - additive-by-additive epistatic effect of pair of SNPs

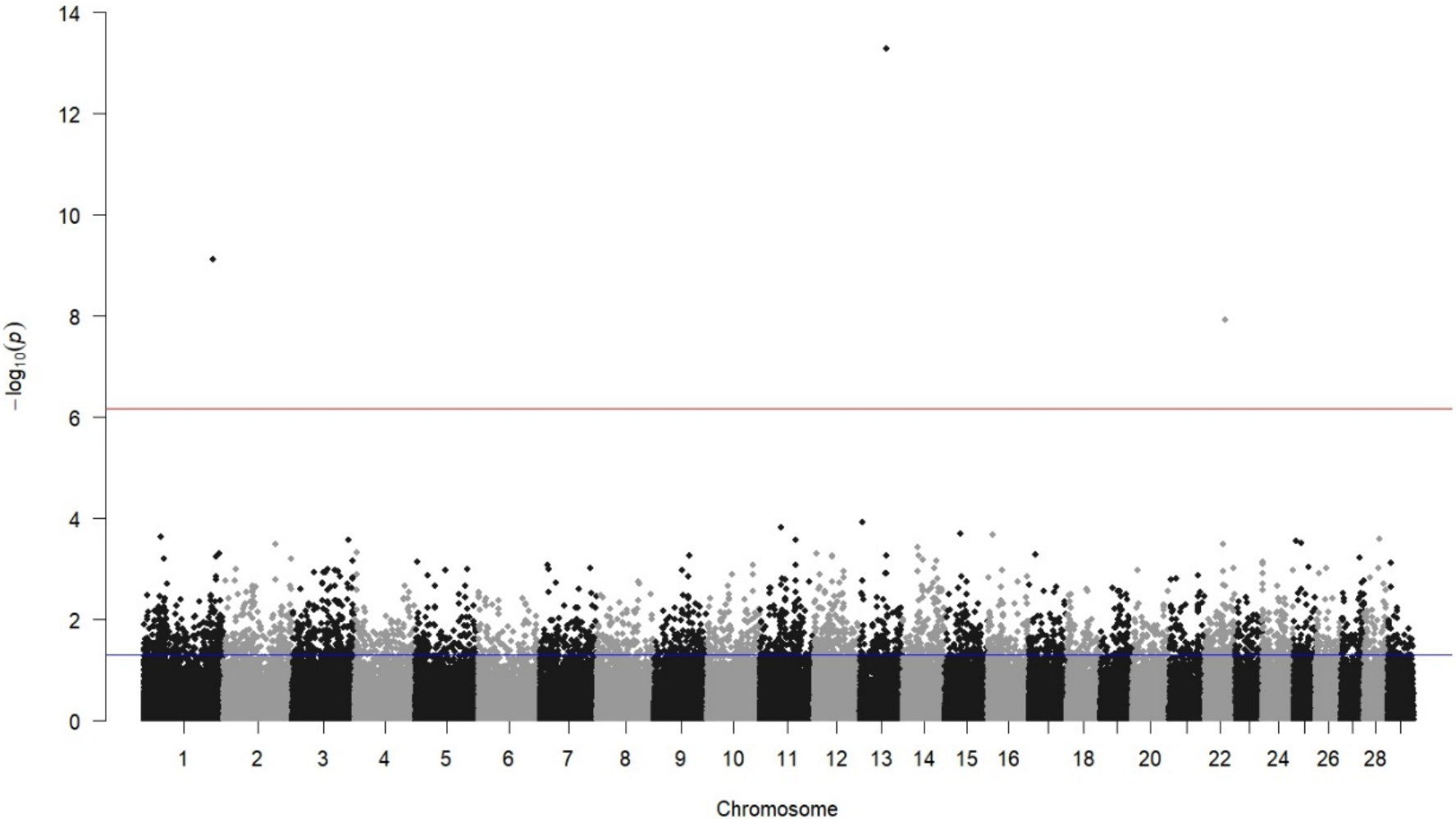
Results - (co)variance components

Trait	Additive polygenic	Permanent environmental	Vet	Residual	Heritability
HSV	0.0017	0.0016	1.0000E-06	0.0437	0.0350
NHD	0.2436	1.0000E-07	0.1681	0.4602	0.2790
HSF	0.0752	0.0028	0.0609	0.1625	0.2490

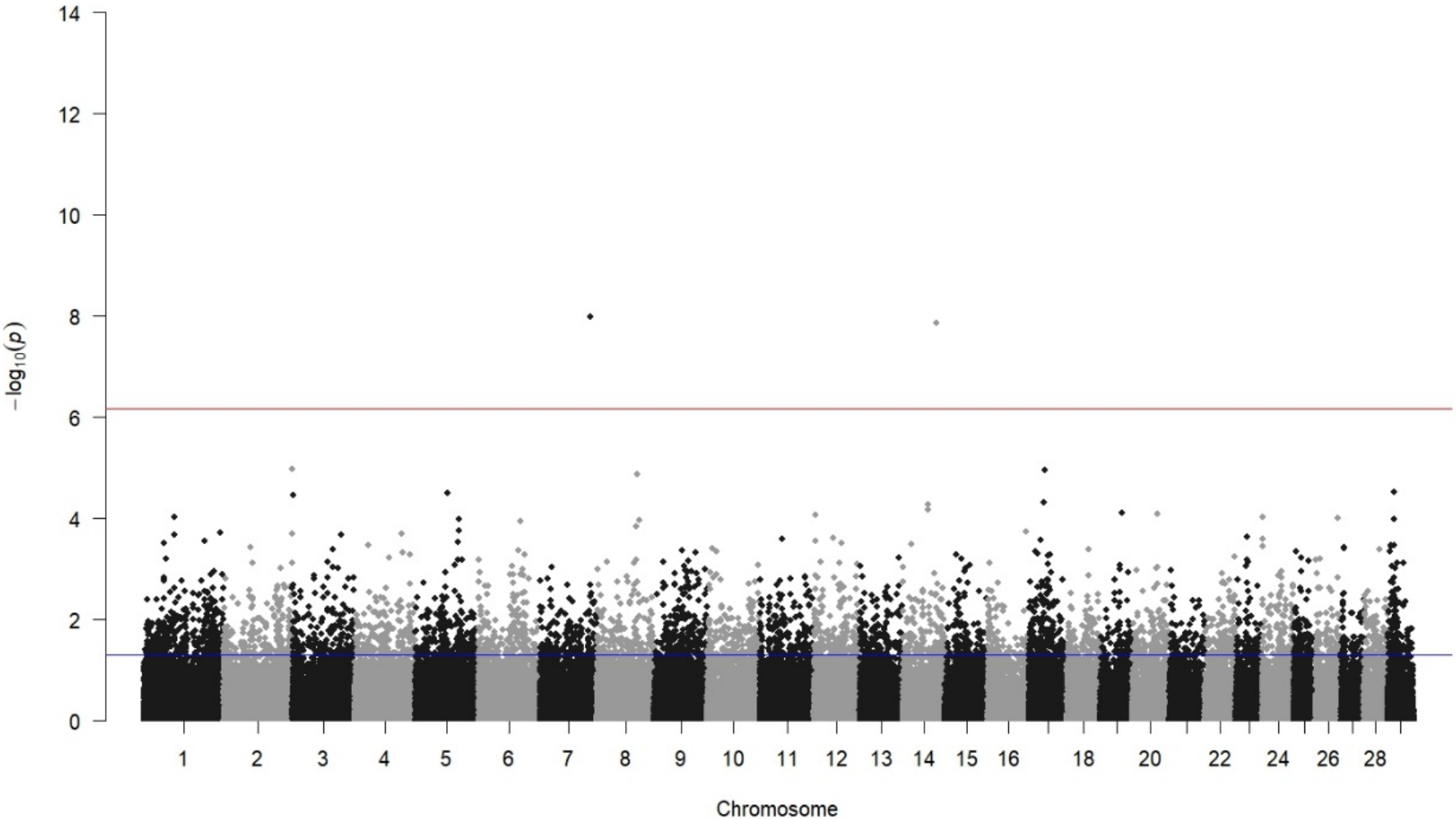
Results - Multiple testing correction

Method	Effective number of tests
Bonferroni	74,762
m_e^1	70,284
m_e^2	69,945

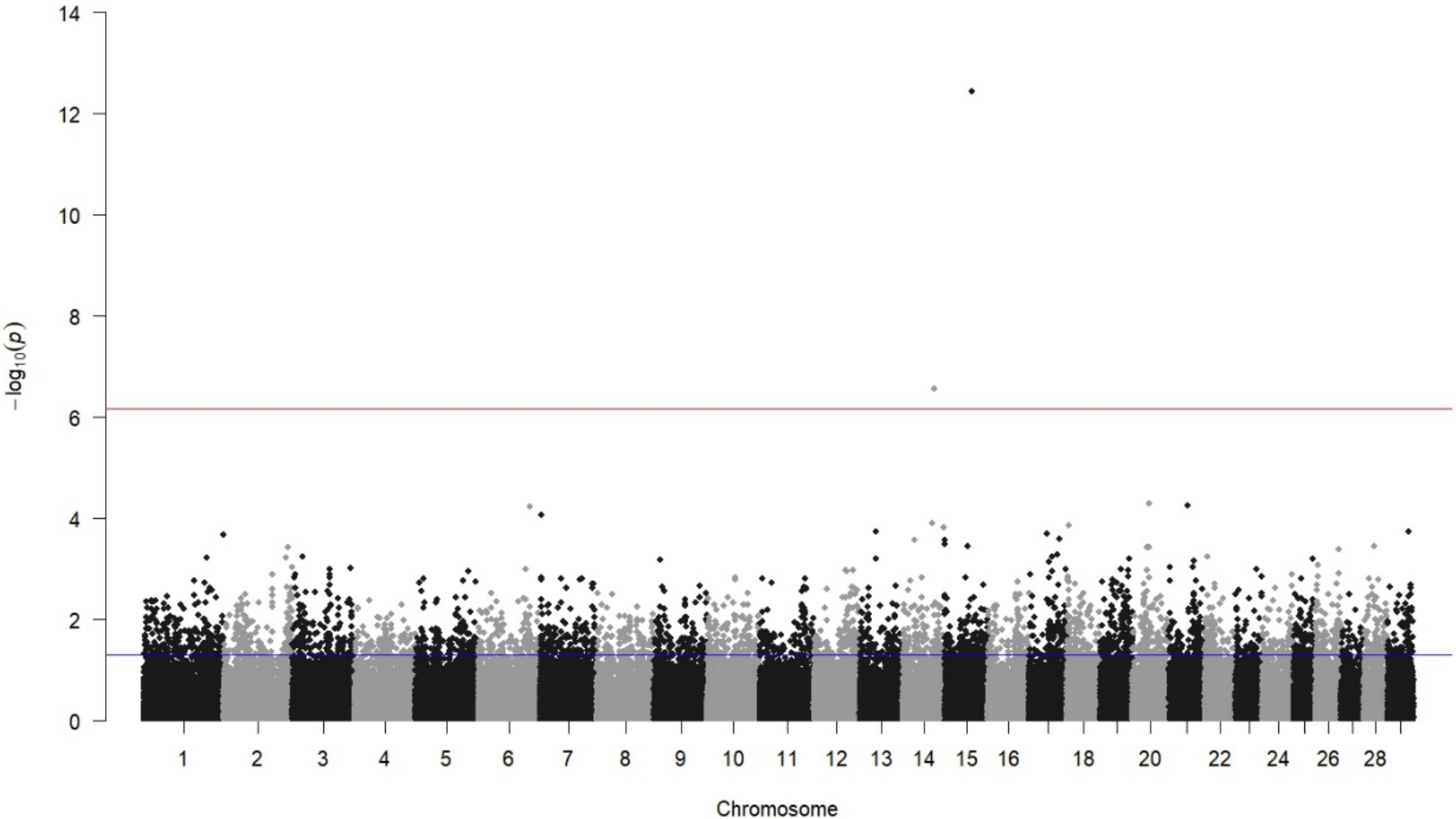
Results - GWAS for HSF



Results - GWAS for NHD



Results - GWAS for HSV



Results - non-additive effects

SNP	Chr	Effect	Allele increasing the risk	P	Effect type A or D
			HSF		
rs43247868	1	0.3459	A	<0.0001	A
rs134142607	13	0.2158	-	0.0267	D
rs111006940	13	0.3912	C	<0.0001	A
rs379823522	22	0.1467	-	0.0458	D
rs110147660	22	0.3206	G	0.0003	A
			NHD		
rs109798552	7	0.1375	A	0.0005	A
rs110532594	14	0.1347	A	0.0005	A
		0.0892	-	0.0336	D
			HSV		
rs207680520	14	0.2764	-	0.0396	D
rs109154693	14	0.5793	C	0.0101	A
rs136200469	15	0.8224	A	<0.0001	A

Results - non-additive effects

SNP	Chr	Effect	Allele increasing the risk	P	Effect type A or D
			HSF		
rs134142607	13	0.2158	-	0.0267	D
⇧ Exon 20 of TOPBP1					
rs379823522	22	0.1467	-	0.0458	D
⇧ Exon 2 of PTPRG					
			HSV		
rs207680520	14	0.2764	-	0.0396	D
⇧ Exon 6 of RRM2B					

Conclusions

- Methods for calculation effective number of tests based on LD could be more accurate then based on Bonferroni method.
- No additive-by-additive epistasis was found.
- Four significant dominance effect were detected.
- We have found three possible causal mutation associated with feet and legs disorders ie. gene TOPBP1 located on BTA1, gene RRM2B on BTA14 and gene PTPRG located on BTA22.

Thank you for your attention!