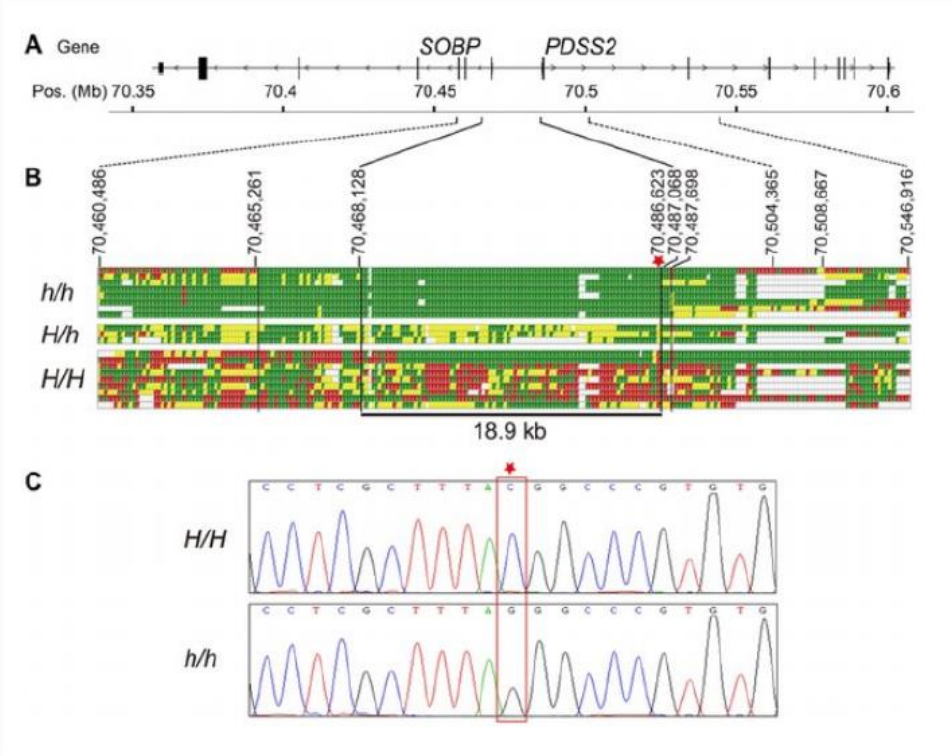
SNP prioritisation for immune responses traits in chicken

T. Suchocki[1,2], B. Czech[4], M. Siwek[3] 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] Roche Poland, Domaniewska 39b, 02 – 672 Warsaw, Poland
[4]UTP University of Science and Technology, Bydgoszcz, Poland

Motivation

Very often significant SNP detected for quantitative traits are not represented by genes and thus is difficult to indicate the causative mutation influenced the trait.



source: researchgate.net

Aim

The second stage of the study.

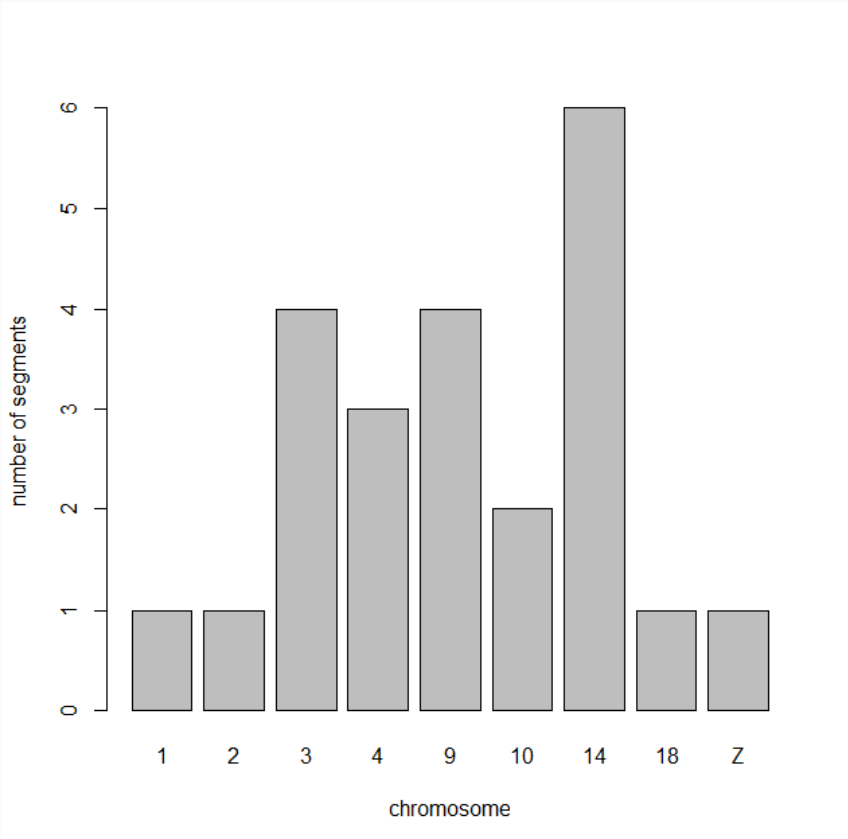
The aim is to create the list of causal snps or at least causal genes influenced immune responses traits of hens using prioritisation method.

Material

1. Animals (288):
   - Green-legged Partridgelike and White Leghorn cross
2. Traits (4):
   - Immune responses are the category or quantitative traits and they are controlled by multiple genes with different magnitudes of phenotypic effects.
   - KLHd0 - specific immune response for a Keyhole LympetHemocyanin antigen at day 0
   - KLHd7 - specific immune response for a Keyhole LympetHemocyanin antigen at day 7
   - LTAd0 - non specific,innate immune responses for Lipoteichoic acid at day 0
   - LPSd0 - non specific,innate immune responses for Lipopolysaccharide at day 0

Material

3. Genotypes(194):
   - sequences of amplicons for *Gallus gallus* sequenced on Illumina MiSeq platform
   - 262 SNP after quality control of sequencing (no amplicon border, low sequencing depth and InDels)
   - 194 SNP after quality control of MAF$>$ 1% and CallRate$>$ 95%



Methods - estimation of BV

Model used for estimation BV and domination deviation:

$$\mathbf{y} = \mu + \mathbf{X}\beta + \mathbf{Z}_a\mathbf{g}_a + \mathbf{Z}_d\mathbf{g}_d + \epsilon,$$

where

- $\mathbf{y} = \{KLHd0, KLHd7, LTAd0, LTSd0\}$
- β - fixed effects of sex and batch
- $\mathbf{g}_a$ - random additive effect
- $\mathbf{g}_d$ - random dominance deviation effect

Methods - estimation of BV

- $\mathbf{g}_a \sim \mathcal{N}\left(\mathbf{0}, \mathbf{G}_a\sigma_a^2\right)$,

  where

$$\mathbf{G}_a = \frac{\mathbf{M}\mathbf{M}'}{2\sum_i p_i q_i}$$

$M_{ij} = \{2-2p_i, 1-2p_i, -2p_i\}$, for AA, AB and BB of SNP_i and individual j

- $\mathbf{g}_d \sim \mathcal{N}\left(\mathbf{0}, \mathbf{G}_d\sigma_d^2\right)$,

  where

$$\mathbf{G}_d = \frac{\mathbf{W}\mathbf{W}'}{2\sum_i p_i^2 q_i^2}$$

$W_{ij} = \{-2q_i^2, -2p_i q_i, -2p_i^2\}$, for AA, AB and BB of SNP_i and individual j

- $\sigma_\epsilon^2 = 0.5\sigma_y^2$, $\sigma_a^2 = 0.3\sigma_y^2$ and $\sigma_d^2 = 0.2\sigma_y^2$

Methods - back solve of SNP effects

SNP effects:

$$\widehat{\mathbf{snp}}_a = \left(\left(\frac{1-k}{n_{snp}} \right)^{-1} + \frac{1}{k}\mathbf{M}\mathbf{A}^{-1}\mathbf{M}^T \right)^{-1} \frac{1}{k}\mathbf{M}\mathbf{A}^{-1}\widehat{\mathbf{g}}_a$$

and

$$\widehat{\mathbf{snp}}_d = \left(\left(\frac{1-k}{n_{snp}} \right)^{-1} + \frac{1}{k}\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^T \right)^{-1} \frac{1}{k}\mathbf{W}\mathbf{A}^{-1}\widehat{\mathbf{g}}_d$$

- $k = 0.2$

Methods - SNP prioritization

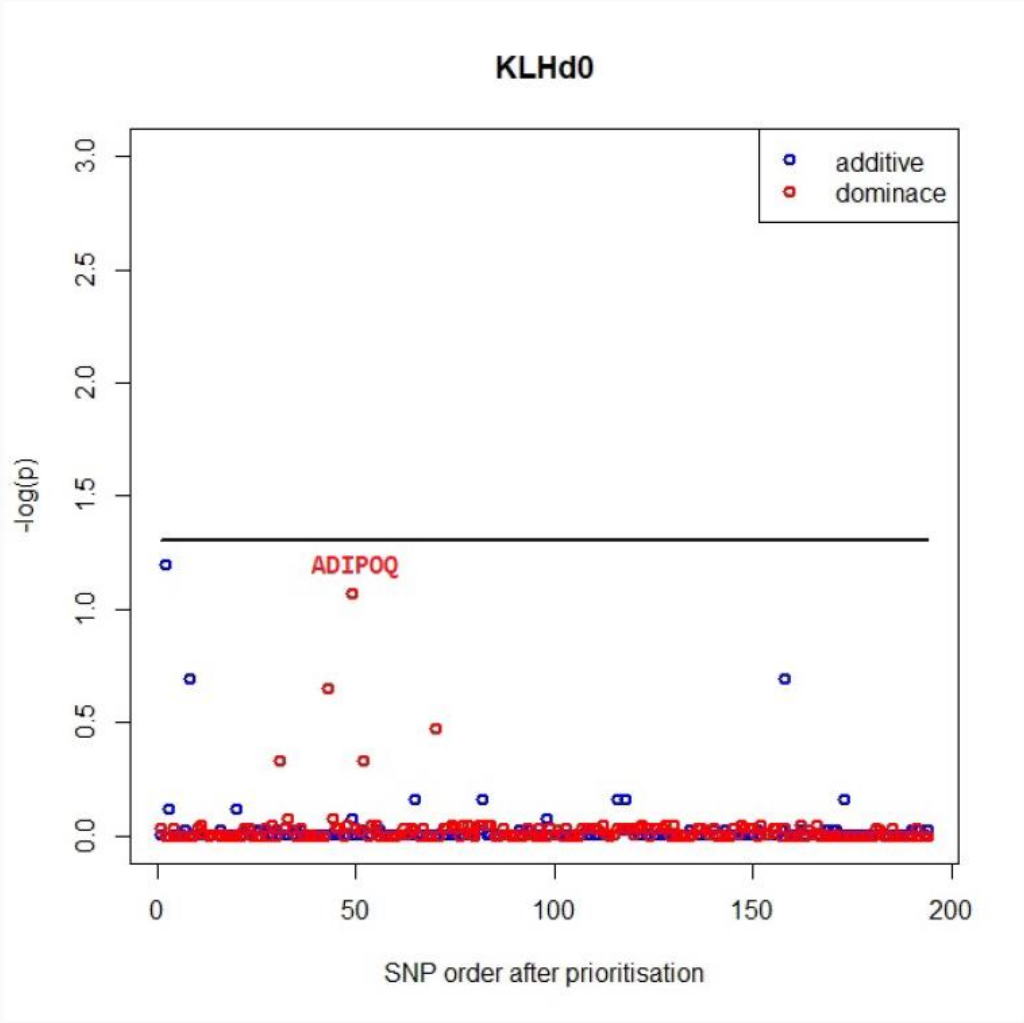
Model used for SNP prioritization:

$$\log(\mathbf{p}) = \mu + log(\mathbf{AF}) + \frac{1}{\mathbf{BaseQRankSum}} + \mathbf{MQ} + \frac{1}{\mathbf{SOR}} + \epsilon$$

where

- **AF** - allele frequency for each ALT allele
- **BaseQRankSum** - Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities
- **MQ** - RMS mapping quality
- **SOR** - symmetric odds ratio to detect strand bias

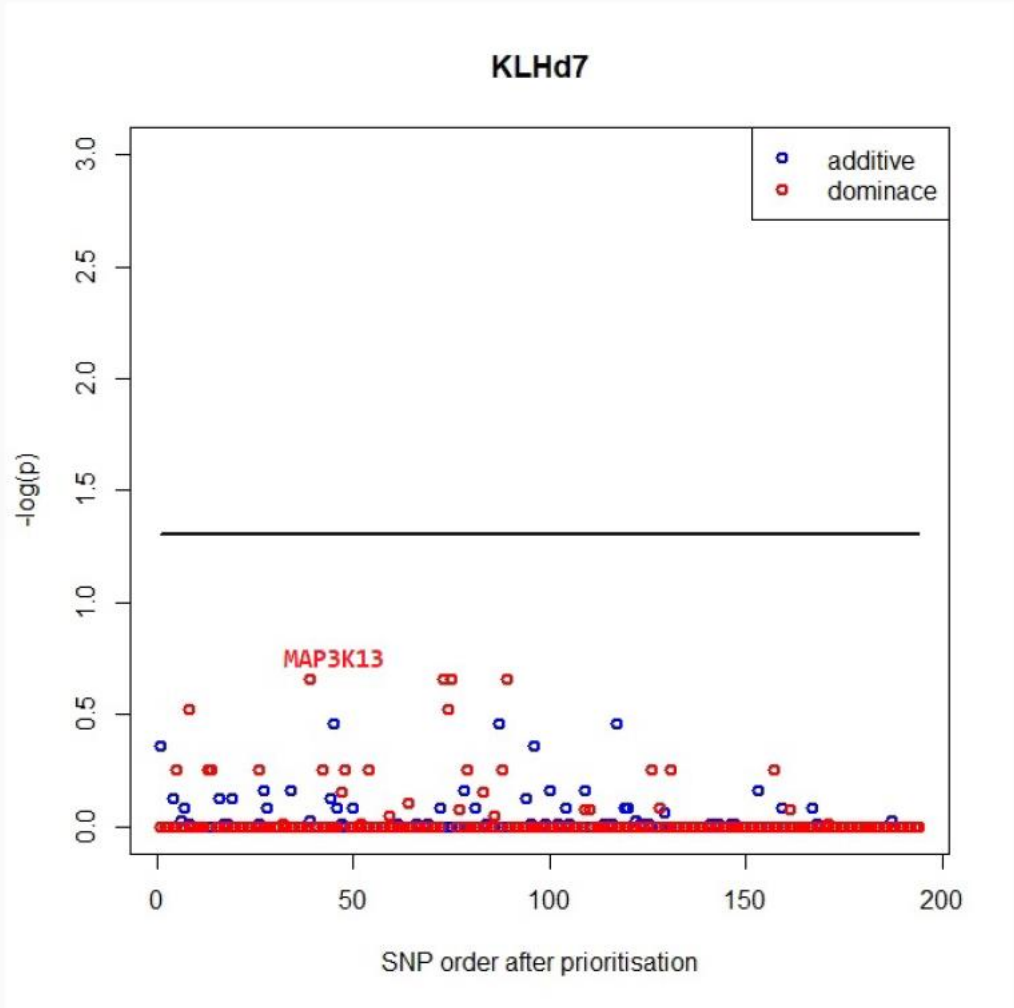
Results - SNP effects KLHd0



Add: chr2 - 2,chr9 - 5, chr10 - 12, chr14 - 1

Dom: chr2 - 3, chr3 - 5, chr4 - 5, chr9 - 1, chr10 - 5, chr14 - 1

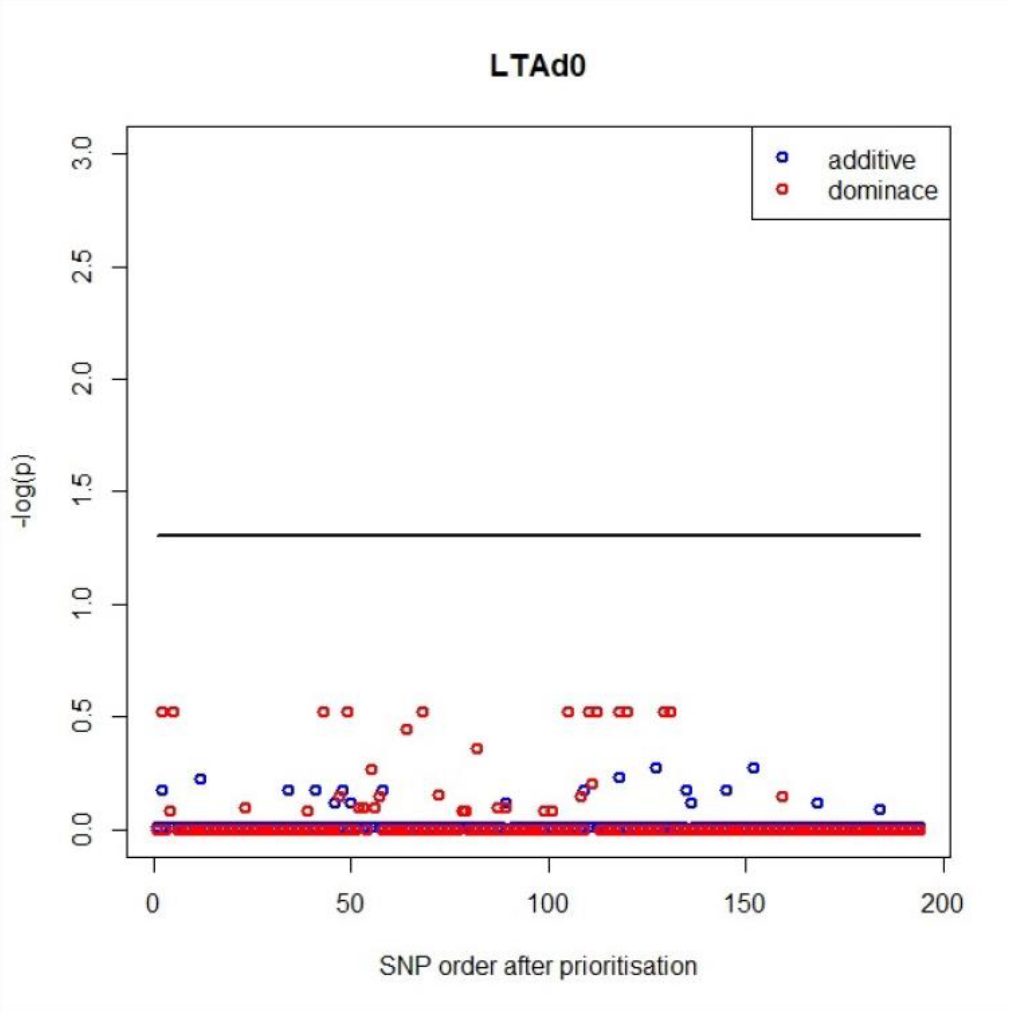
Results - SNP effects KLHd7



Add: chr1 - 1,chr3 - 14, chr4 - 3, chr9 - 2

Dom: chr3 - 1, chr4 - 2, chr9 - 10, chr14 - 5, chrZ - 2

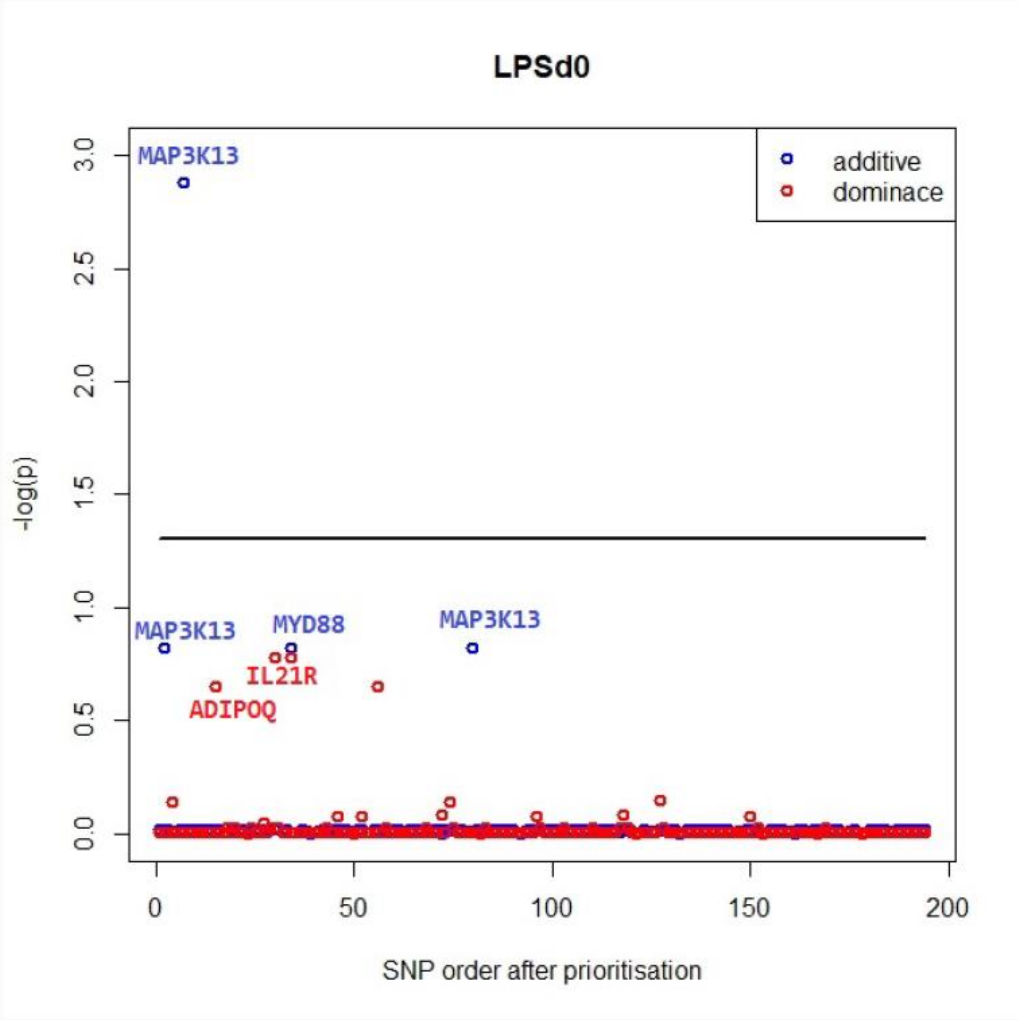
Results - SNP effects LTAd0



Add: chr1 - 1,chr3 - 3, chr4 - 3, chr9 - 8, chr14 -4, chr18 - 1

Dom: chr3 - 8, chr4 - 5, chr9 - 4, chr10 - 2, chr14 - 1

Results - SNP effects LPSd0



Add: chr3 - 13, chr4 - 4, chr9 - 2, chr10 -1

Dom: chr1 - 1, chr4 - 15, chr9 - 2, chr10 - 2

Results - frequency of genes for first 20 prioritizationing SNP

gene or not a gene	frequency
not a gene	35.63%
NOCT	14.38%
MYD88	13.13%
MAP3K13	10.63%
MED12	8.13%
PDGFA	2.50%
ADIPOQ	2.50%
IL21R	2.50%
SNX8	1.88%
PROCR	1.88%
CD36	1.88%
TLR5	1.88%
MAP2K1	1.25%
MSL2	0.63%
E4F1	0.63%
AP3B1	0.63%

Conclusion

- There is no correlation between prioritization and p-value after FDR correction, $\rho \in (0.08, 0.22)$.
- Genes: NOCTA, MYD88, MAP3K13 and MED12 covered almost 46% of first 20 prioritizationing SNP for all four traits.
- 36% of first 20 prioritizationing SNP are not represented by any gene.
- Prioritizationing is more efficient than p-values (Only 30% of SNP with FDR$<$ 0.3 are represented by some genes).

acknowledgement

NATIONAL SCIENCE CENTRE
POLAND

PROJECT NUMBER: NCN 02123

Thank you for your attention!