

# SNP prioritisation in GWAS with dense marker sets

B. Czech<sup>1</sup>, T. Suchocki<sup>1,2</sup>, J. Szyda<sup>1,2</sup>

<sup>1</sup>Biostatistics Group, Department of Genetics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland; <sup>2</sup>National Research Institute of Animal Production, Balice, Poland



## Conclusions

- The empirical density distribution of GWAS based *P*-values from **whole-genome sequencing** does not significantly differ from the density distribution of GWAS based *P*-values from **targeted amplicon sequencing**
- The empirical density distribution of GWAS based *P*-values from **whole-genome sequencing** is significantly different from the density distribution of GWAS based *P*-values from **SNP microarray**
- **Intercorrelation between SNPs affects the empirical *P*-value density distribution in GWAS**

## Objectives

The comparison of **GWAS based *P*-values** resulting from different real data structures and SNP densities.

## P-value sets

- GWAS based on whole-genome sequencing (WGS)
- targeted amplicon sequencing (TAS)
- SNP microarray

## Methods

- Two sample Kolmogorov-Smirnov test for comparison of empirical density distributions of *P*-values
- Estimation of the proportion of true  $H_0$ :  
 LFDR (Phipson, averaging local FDR)                      mean (averaging *P*-values)  
 histogram method (Mosig and Nettleton)                      convest (Langaas, convex decreasing density)
- Multiple testing correction:  
 Holm                      Hochberg                      Hommel                      Bonferroni                      Benjamini & Hochberg

## Results

### Different multiple testing correction methods

The numbers of significant *P*-values (<0.005) and its proportion to all SNPs

Method of genotyping	Number of SNPs	Holm	Hommel	Hochberg	Bonferroni	Benjamini & Hochberg
WGS	75590	2 (0.002%)	2 (0.002%)	2 (0.002%)	2 (0.002%)	2 (0.002%)
Microarray	73013	9 (0.012%)	9 (0.012%)	9 (0.012%)	9 (0.012%)	14 (0.019%)
TAS	388	0 (0.000%)	0 (0.000%)	0 (0.000%)	0 (0.000%)	0 (0.000%)

### Different methods of estimating the proportion of true $H_0$

Method of genotyping	LFDR	Mean	Histogram	Convest
WGS	0.9999973	0.9999975	1.0000000	1.0000000
Microarray	0.9997295	0.9997317	0.9985244	0.9983977
TAS	0.9970516	0.9974642	0.9820944	0.9811121

### Comparison of *P*-value empirical density distributions → KS test

Microarray vs WGS  $P = 0.001$       Microarray vs TAS  $P = 0.630$       WGS vs TAS  $P = 0.621$

