Stiftung Tierärztliche Hochschule Hannover
University of Veterinary Medicine Hannover, Foundation

EAAP 2018
Dubrovnik, Croatia
Session 66

# Multi-locus random-SNP-effect genome-wide association studies for health traits in horses

O. Distl, J. Metzger, B. Ohnesorge
Institute for Animal Breeding and Genetics

## General analysis model for GWAS

$$\text{Phenotype} = X\beta + S\alpha + Qv + Zu + e$$

$X\beta$: fixed systematic effects

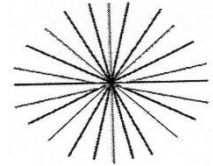$S\alpha$: allelic SNP-effects

$Qv$: Population structure effects

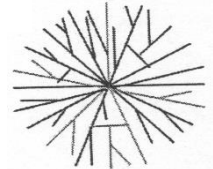$Zu$: random additive quantitative effects

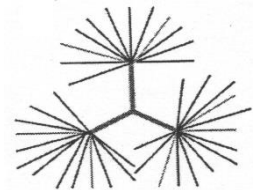$e$:   random unknown residuals

# Stratification types
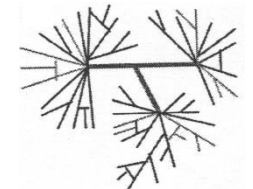
(I) **Randomly and equally distributed**

(II) **Familial relationships**

(III) **Population structure**

(IV) **Population and family structure**

(V) **Subpopulation, population and family structure**

Stiftung Tierärztliche Hochschule Hannover
University of Veterinary Medicine Hannover, Foundation

**MLM-GRM**
**Traits under analysis: drEBVs, Phenotypes**
**SNP by SNP analysis**

**Trait =**

      **SNP-allele$_i$ +**

      **Breed contributions (Han-Tb-Hol-Trak)$_j$ +**

      **Cryptic structure (PrinComp)$_k$ +**

      **Fixed effects$_l$ +**

      **Family structure-Quantitative effects $(A*\sigma^2_a)_m$ +**

      **Residual**

# Multi-locus Models

**Multi-locus models**
>   **problem of multicollinearity**
>   **number of variables**
>   **higher power to detect QTN and lower MSE**

**Multi-locus approaches**

>   **BLASSO** (Bayesian least absolute shrinkage and selection operator)
>   **ISIS-EM-BLASSO** (iterative modified-sure independence screening EM BLASSO)
>   **SCAD** (smoothly clipped absolute deviation**)**
>   **FarmCPU** (fixed and random model circulating probability unification)
>   **BayesR**
>   **mrMLM** (multi-locus random-SNP-effect mixed linear model)
>   **pLARmEB** (polygenic-background-control-based least angle regression plus
>   empirical Bayes)

Comparison of statistical power for QTN detection in simulations using ISIS EM-BLASSO, SCAD, EMMA, FarmCPU and mrMLM



Source: Tamba et al. 2017 PLoS Comput Biol 13(1): 1005357

Comparison of mean squared error of QTN estimates in simulations using ISIS EM-BLASSO, SCAD, EMMA, FarmCPU and mrMLM
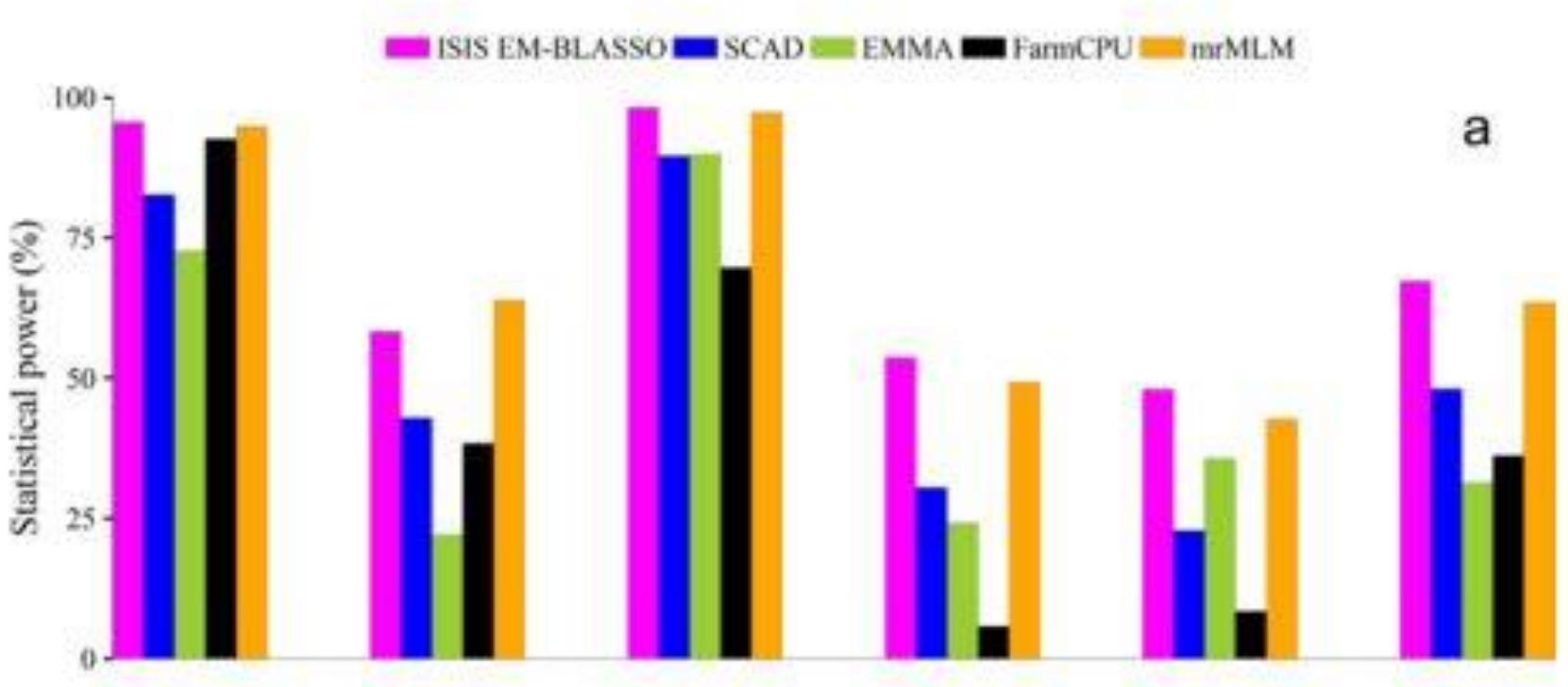


Source: Tamba et al. 2017 PLoS Comput Biol 13(1): 1005357

# Multi-locus Model - mrMLM
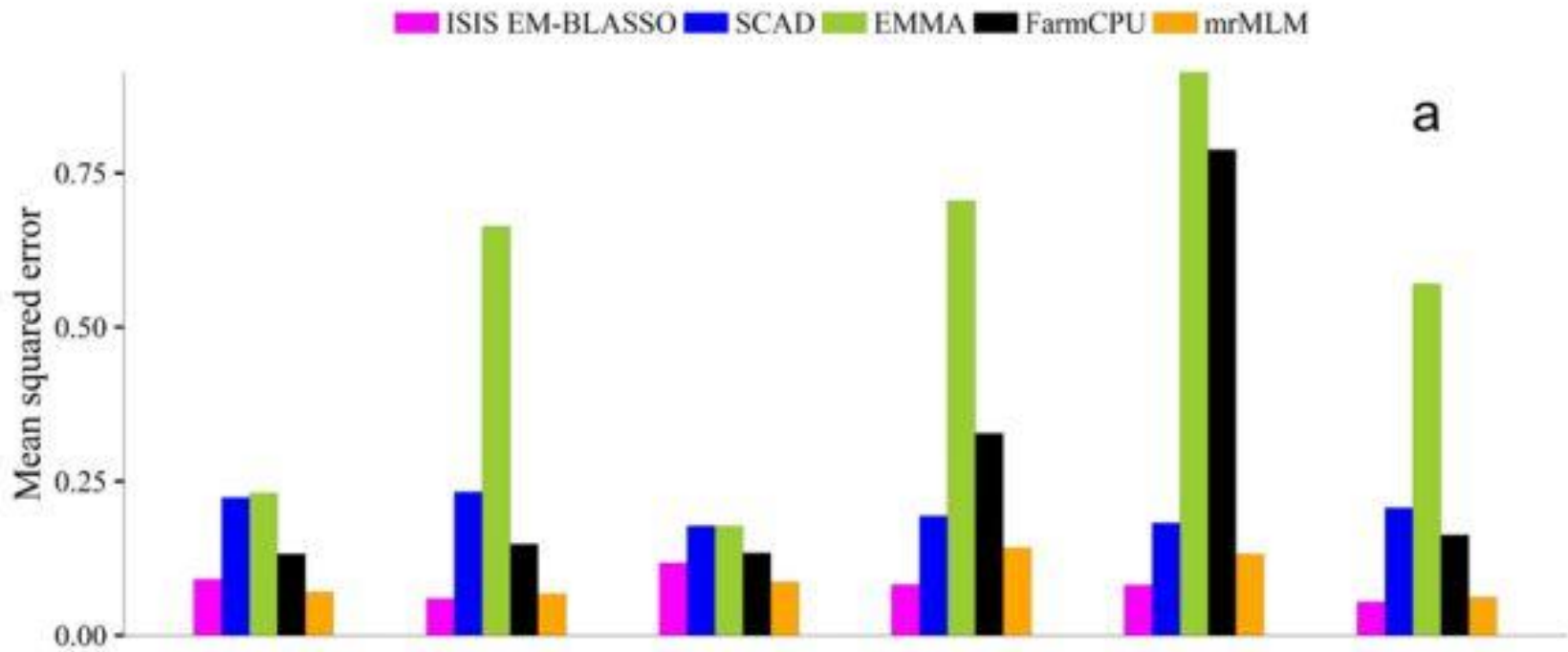
Trait

    Equine recurrent uveitis (ERU 0/1-trait)

    Recurrent-remitting episodes of inflammations of both eyes or one eye

    Complex disease of the eyes with aberrant autoimmune response

    Most common cause of blindness in warmblood and draught horses (3-15%)

Genotyping

    Equine SNP50 beadchip for cases and controls (n=250)

    Next generation sequencing data from 80 horses

    Imputed SNP data set with ~ 2 mio SNPs

    Genomic relationship matrix

ERU: inflammatory products in the eye

# Multi-locus Model - mrMLM

mrMLM (2-step procedure)

    (1) Single SNP random linear mixed model (RLMLM)

    (2) Multi-locus random linear mixed model


Selection of SNPs for the second step

    P-value < 0.01

    Elimination of all SNPs within 200 kb (collinearity) around a selected SNP
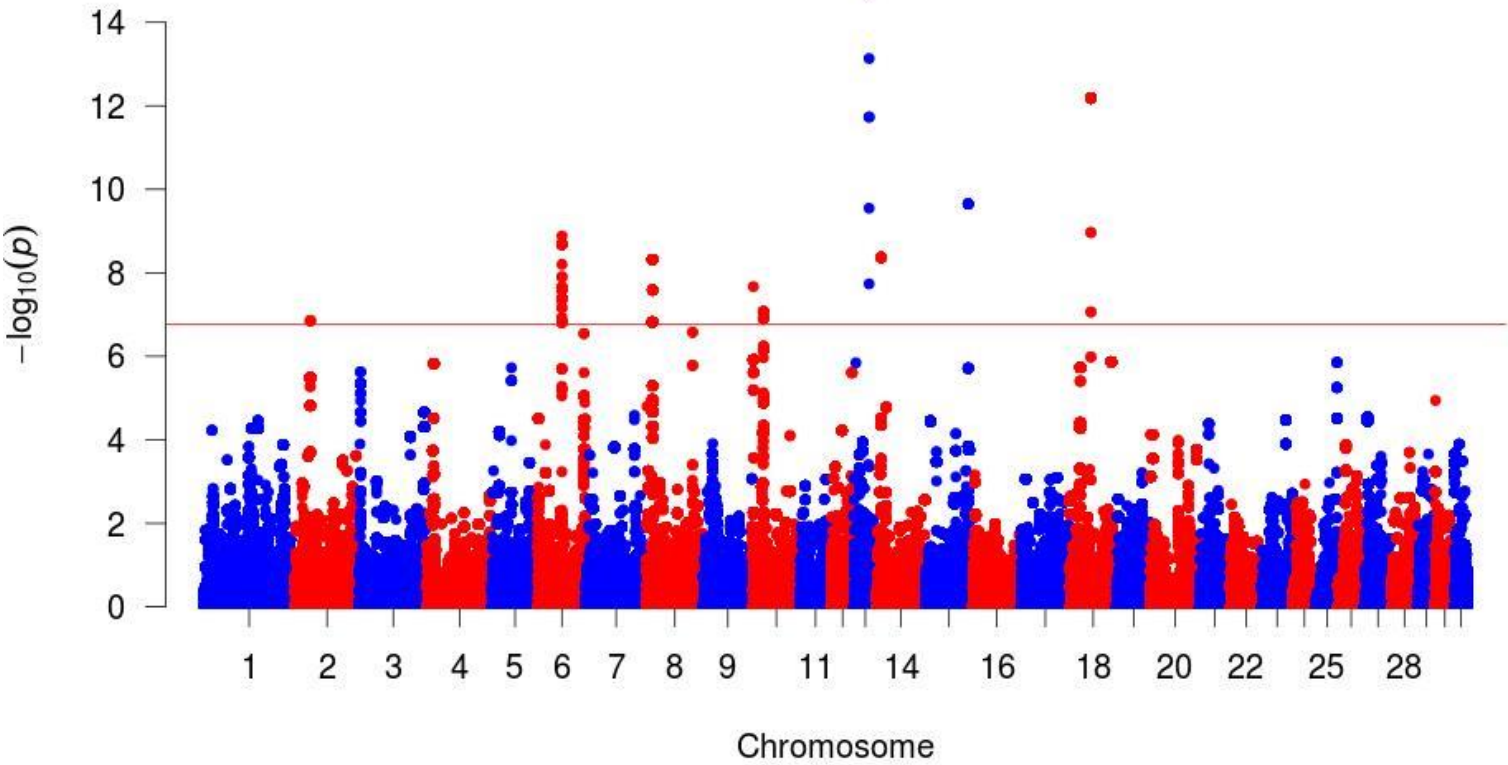
Second step

    Empirical Bayes

    Thresholds for P-values < 0.0002

## Manhattan plot using RLMLM – Trait: ERU

Stiftung Tierärztliche Hochschule Hannover
University of Veterinary Medicine Hannover, Foundation

# Multi-locus Model Estimates – mrMLM - ERU

| ECA | SL | ML | Position (Mb) | QTN-effect | LOD score | $R^2$(%) |
|---|---|---|---|---|---|---|
| 1 | | | 62 | -0.14 | 5.9 | 4.2 |
| 2 | * | * | 80 | 0.21 | 11.6 | 10.7 |
| 2 | * | | 106 | -0.11 | 3.8 | 2.3 |
| 3 | | * | 46 | -0.21 | 7.2 | 10.2 |
| 3 | | * | 115 | -0.22 | 12.4 | 12.6 |
| 4 | | | 103 | -0.12 | 3.7 | 3.6 |
| 5 | | | 35 | 0.12 | 3.6 | 3.5 |
| 9 | | * | 20 | -0.20 | 8.5 | 7.0 |
| 12 | | * | 2 | -0.18 | 6.8 | 5.4 |
| 12 | | * | 20 | -0.16 | 6.5 | 6.3 |
| 14 | * | * | 80 | 0.19 | 10.0 | 7.9 |
| 19 | | | 46 | 0.13 | 4.4 | 3.6 |
| 22 | | | 13 | -0.13 | 3.4 | 3.9 |
| 29 | | * | 28 | 0.20 | 6.2 | 7.9 |

# Multi-locus Model – ISIS-EM-BLASSO

Model

ISIS-EM-BLASSO (iterating procedure)

    (1) Reducing of SNPs using SCAD via correlation learning (SIS)

    (2) SCAD for moderate-scale variable selection

    (3) Expectation-Maximization (EM)-Bayesian LASSO algorithm

Selection of SNPs for the final step

    Critical P-value < 0.01

# Multi-locus Model Estimates– ISIS-EM-BLASSO

| ECA | SL | mrMLM | Position (Mb) | QTN-effect | LOD score | R² (%) | MAF |
|---|---|---|---|---|---|---|---|
| 1 | | | 17 | -0.09 | 3.1 | 2.7 | 0.01 |
| 3 | | # | 115 | 0.08 | 3.8 | 2.8 | 0.12 |
| 8 | * | | 85 | -0.08 | 3.4 | 2.6 | 0.15 |
| 13 | *** | | 28 | -0.18 | 5.5 | 11.8 | 0.38 |
| 15 | *** | | 58 | 0.07 | 3.0 | 1.8 | 0.13 |
| 15 | *** | | 85 | -0.11 | 3.0 | 4.9 | 0.16 |
| 16 | | | 5 | -0.12 | 3.8 | 5.7 | 0.05 |
| 18 | *** | | 14 | -0.14 | 5.0 | 7.8 | 0.07 |
| 19 | | | 14 | -0.12 | 4.7 | 5.8 | 0.10 |
| 21 | | | 31 | 0.06 | 2.6 | 1.3 | 0.17 |
| 24 | | | 10 | -0.09 | 3.4 | 2.8 | 0.15 |
| 25 | | | 38 | 0.12 | 2.9 | 5.0 | 0.06 |
| 28 | | | 6 | -0.07 | 2.0 | 1.6 | 0.06 |

# Conclusions

- Multi-locus random estimation improves detection accuracy and lowers MSE for QTN

- Complementarity among models

- Multi-locus methods with high empirical power in QTN detection accelerate search for causative major variants for complex traits

- Real data examples

  - Validation of results necessary through genotyping

  - Animals to be selected for NGS based on multi-locus results

- Whole genome sequencing data and multi-locus models

  - ISIS-EM-BLASSO most efficient for QTN detection

  - Further models under development (pKWmEB, e.g.)

# Thank you for your attention!