Performance of genomic selection for traits in mice using Bayesian multi-marker models

D N R G Kapell, D Sorensen, G Su,
L L G Janss, C J Ashworth and R Roehe

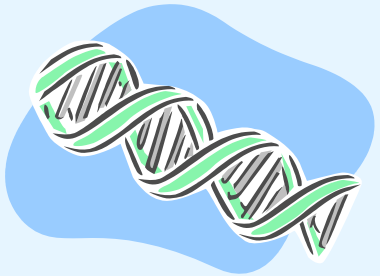
EAAP 26/08/09

Outline

- Aims & Introduction
  - Genomic Selection & Mixtures
- Materials
  - Mice & SNP
- Methods
  - Models & Mixtures
- Results
  - Variances
  - Predictive Ability
  - Accuracy
- Conclusions

Introduction

- Polygenic selection
  - Uses phenotypic records and pedigree
- Genomic selection
  - Uses genome wide dense markers to estimate genomic breeding values
  - Uses several thousand markers, with varying size of effect
  - Differences between methods of genomic selection are often in the prior assumptions regarding the size of the marker effects
  - The marker effect can be estimated by using a mixture of two distributions, one where the marker is associated with no effect and one where the marker is associated with an effect



Aims

- To compare the predictive ability of models based on:
  - Polygenic effects only
  - Genomic effects only
  - Combined genomic & polygenic effects

- Across a range of traits
- With different mixtures of the proportion of markers with or without effects

Material

- Data available from WTCHG (http://gscan.well.ox.ac.uk/)
- 2281 genotyped mice with 10946 SNPs
- 3 traits, ranging from high to low heritability:
  - 6-Wk weight (h^2=0.74)[1]
  - Total activity (h^2=0.34)[1]
  - Hematocrit percentage (h^2=0.11)[1]

[1] (Valdar et al.,2006)

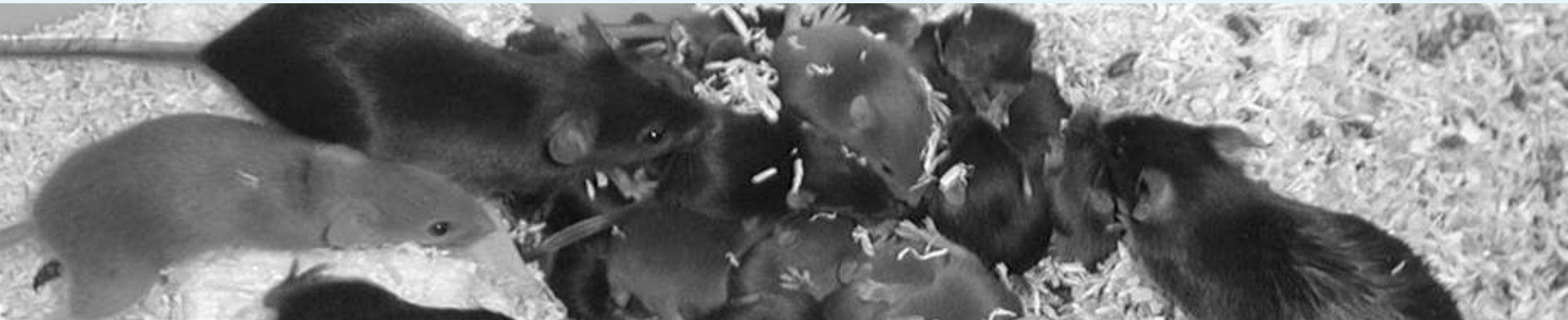
Material

- All traits analysed with full model
  - Fixed effects, covariates and transformations based on Valdar et al.,2006
- Pedigree 2890 animals
- Cages almost completely housed animals from one full sib family (avg 3.1 cages/family)

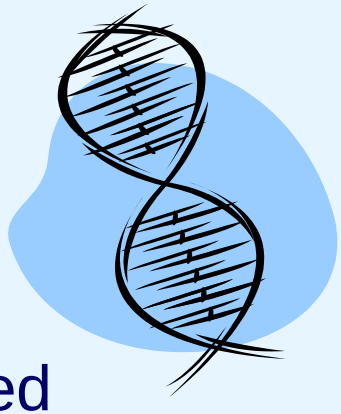
Material

- Descriptive statistics (after transformation)

Trait	n	Mean ± s.e.	Family size
6-Wk weight	1916	272.0 ± 15.0	11.3 ± 8.0
Total activity	1879	44.9 ± 14.8	10.1 ± 7.5
Hematocrit %	1578	22.3 ± 7.9	10.9 ± 7.8



Method: overview



- Model 1: polygenic effects only
- Model 2: genomic effects only
- Model 3: polygenic and genomic effects combined
- Models 2 and 3 with different percentages of markers having an effect
  - Mixture model with 10%, 40% and 70% of markers showing an effect
  - Non mixture model with all markers showing an effect

Method: model

- All models analysed using iBay (L. Janss, 2009)
- Bayesian multi-marker model

$$y = \mu + \sum_i X_{1,i}\beta i + \sum_j X_{2,j}\gamma_j + \sum_k \sigma_k Q_k \alpha_k + Zu + e$$

  - $\sigma_k Q_k \alpha_k$ fits the marker effects, where α_k is a vector with the effects of marker alleles and σ_k is a scaling factor modelling the variance explained by that marker

Method: mixtures

- Prior assigned to the scaling factor σ_k
  - For non mixture models:
    - $\sigma_k \sim TN_{>0}(0,\sigma^2_g)$
    - Whereby σ^2_g can be interpreted approx. as the expected average fitted variance per marker
  - For mixture models:
    - $\sigma_k \sim \begin{cases} N(0,\sigma^2_{g0}) \text{ with probability } \pi_0 \\ TN_{>0}(0,\sigma^2_{g1}) \text{ with probability } \pi_1 = 1-\pi_0 \end{cases}$
    - Whereby the first distribution models the markers with no effect with π_0 the proportion of markers without effect, and the second distribution models with markers that have an effect with proportion π_1

Method: datasets

- Variance components
  - Based on full set of phenotypic data
- Predictive ability
  - Correlation between predicted and observed based on residuals
    - Observed phenotype corrected for fixed effects excl. cage
  - Five subsets of the full set
    - Proportion of training:validation set = 5:1
  - Accuracy
    - Increased correlation of model 2 vs. model 1, depending on variances

Method: sampling

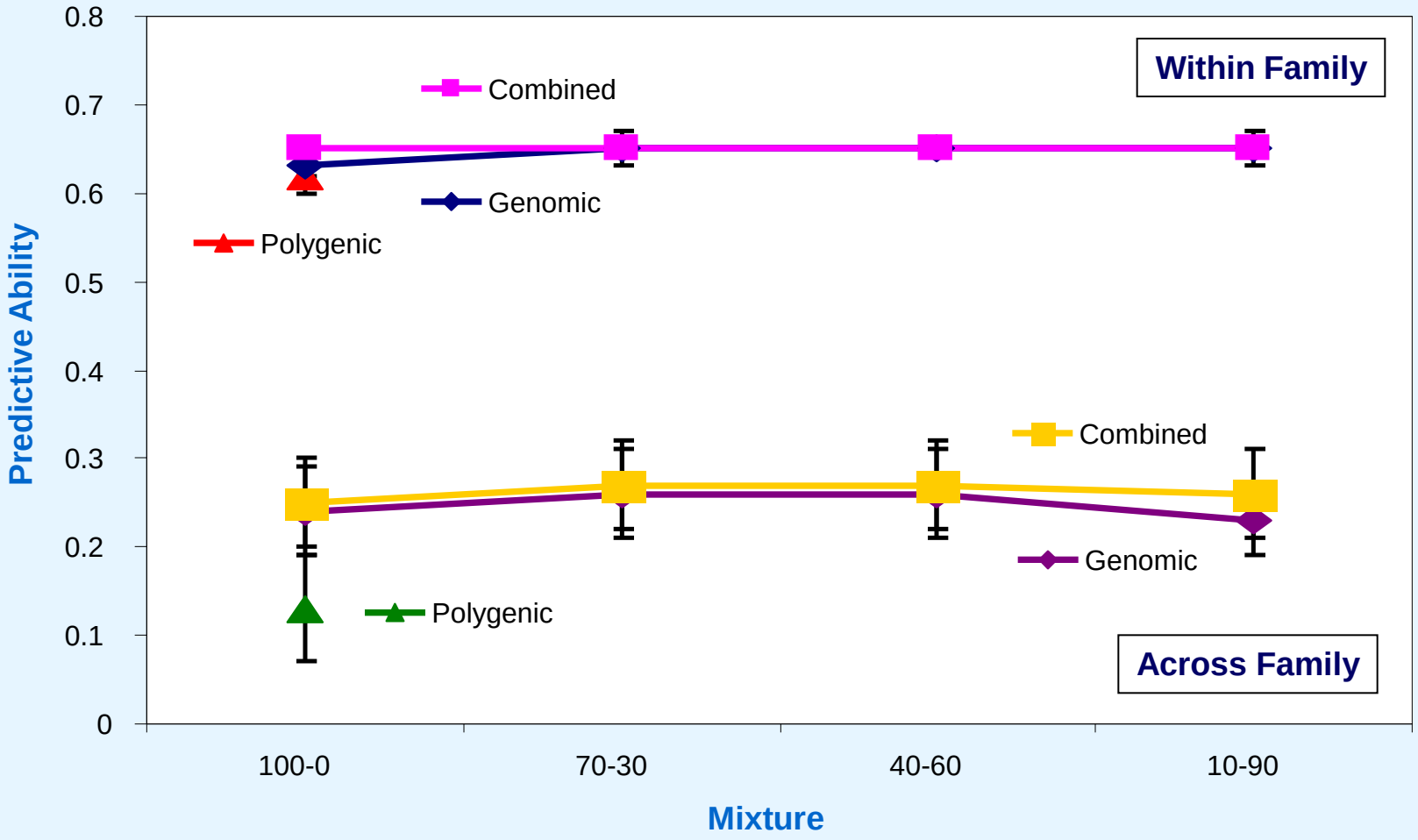
- Selection within family
  - 1/6th of animals per family in validation
  - Splitting full-sib families over datasets
  - Use within family information
- Selection across families
  - 1/6th of all families in validation
  - Select entire families
  - Little or no relationship between families
  - Use across family information

Variances: 6-Wk weight

mix	σ^2_a	σ^2_u	σ^2_c	σ^2_e	h^2_a	h^2_u
			Model 1: Polygenic			
	-	58.4	31.3	21.1	-	0.53
			Model 2: Genomic			
100/0	42.3	-	39.0	35.8	0.36	-
70/30	22.5	-	41.7	42.2	0.21	
40/60	21.9	-	41.8	42.3	0.21	-
10/90	17.2	-	43.8	43.8	0.16	-
			Model 3: Combined			
100/0	34.1	29.3	33.1	22.5	0.29	0.25
70/30	17.0	22.5	34.1	32.3	0.16	0.21
40/60	16.9	22.9	34.2	32.0	0.16	0.22
10/90	13.4	36.1	32.7	25.8	0.12	0.33

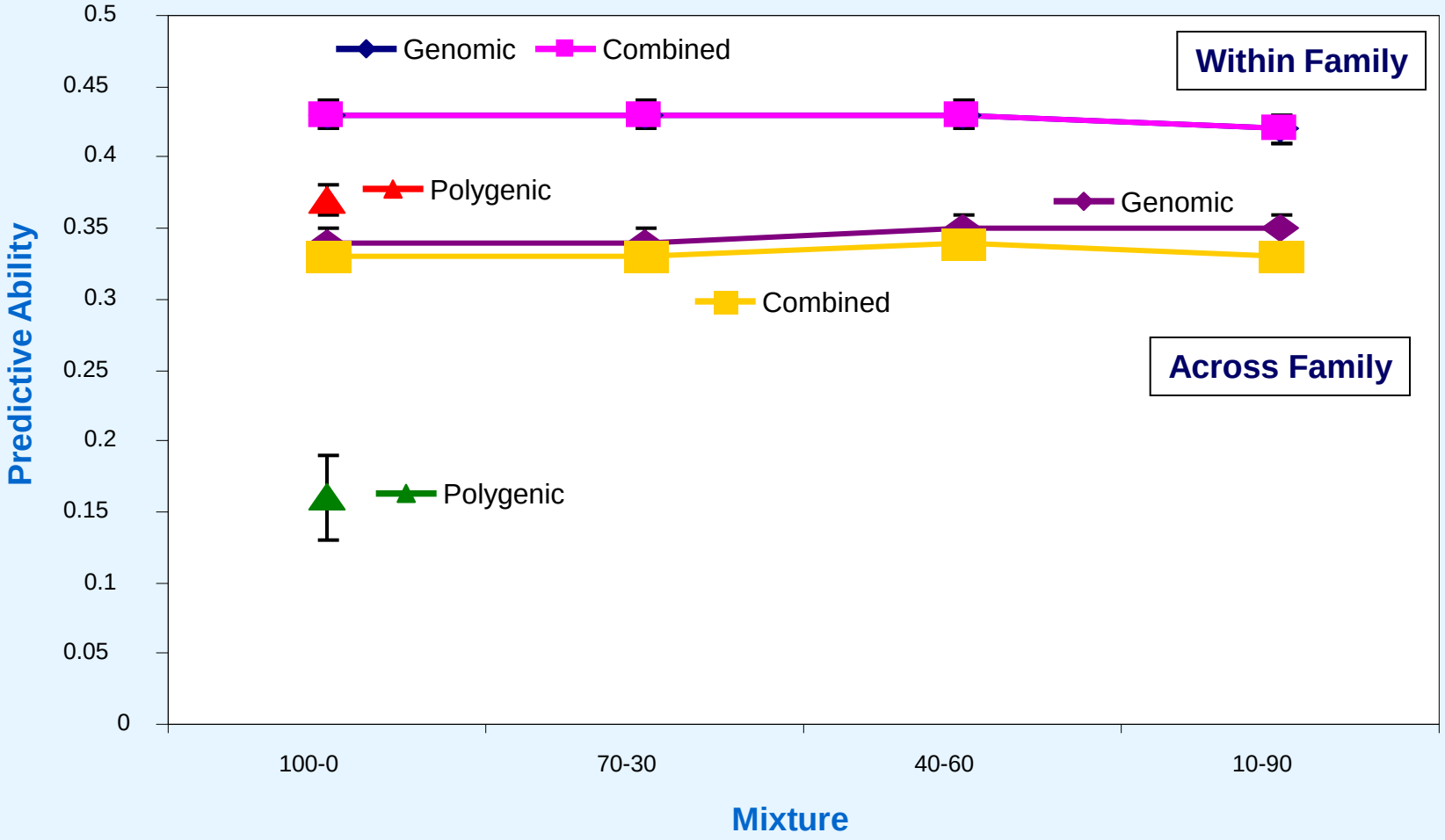
Predictive ability (1)

6-Wk weight (h^2=0.74)



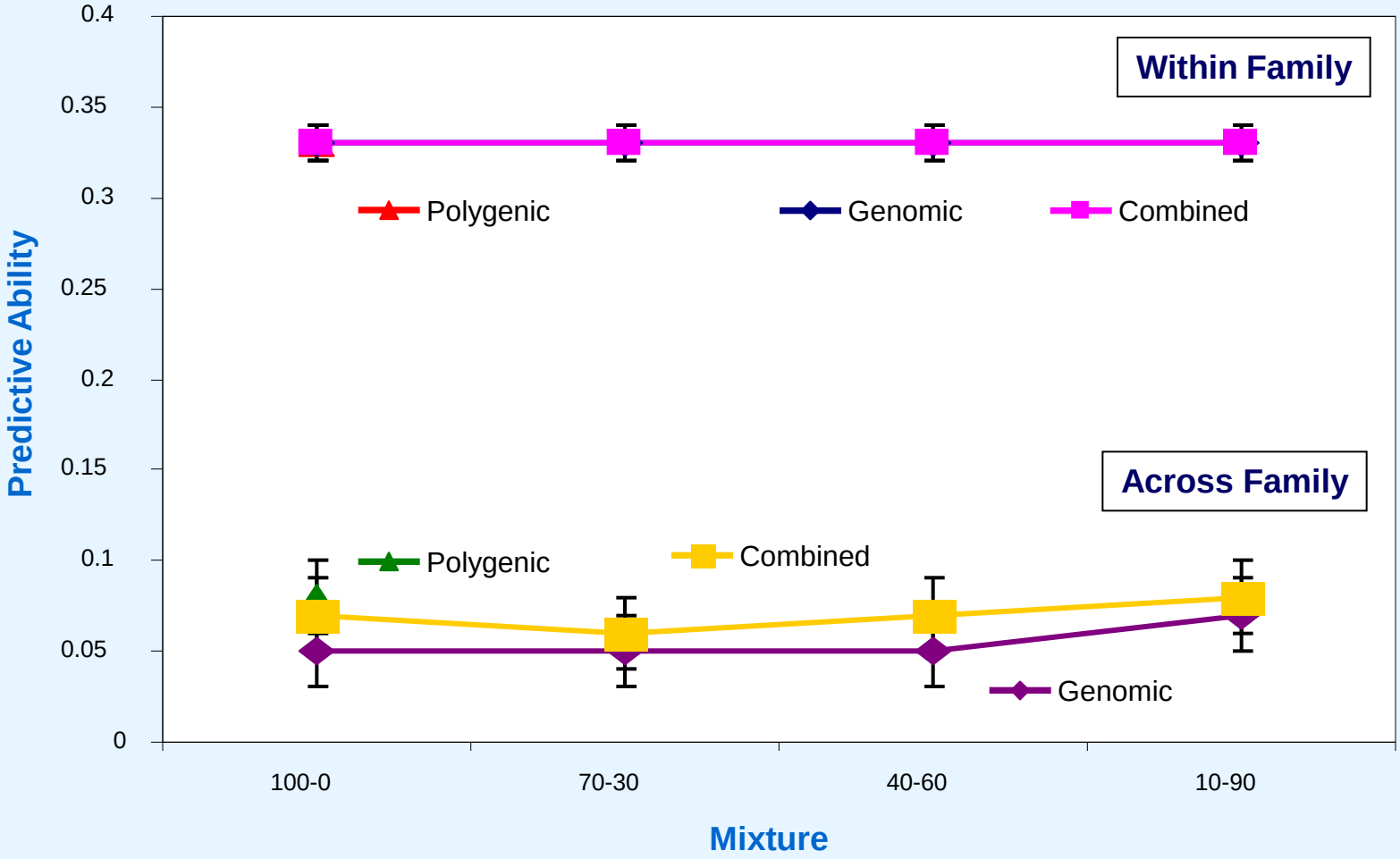
Predictive ability (2)

Total Activity (h^2=0.34)



Predictive ability (3)

Hematocrit % (h^2=0.11)



Accuracy

Increase in accuracy of model 2 (genomic) compared to model 1 (polygenic) with no mixture

Trait	Within family	Across Family
6-Wk weight	0.01	0.18
Total activity	0.10	0.32
Hematocrit %	0.01	-0.16

Conclusions

- Higher predictive ability for selection within family than across family
  - Shown by the higher within family predictability for all traits
- Genomic selection, when compared with polygenic selection, benefits the trait ‘Total activity’ when selection is across families
- Reduction in the proportion markers with an effect has little effect on predictive ability

Conclusions

- Higher increase in accuracy for selection across family than within family, especially with decreased heritability, except for the trait ‘Hematocrit %’
  - Shown by the increased accuracy of the trait ‘Total activity’ compared to the trait ‘6-Wk weight’

- Based on these results it would seem that an increase in the number of markers selected for genomic selection would not automatically lead to an increase in predictive ability

Acknowledgements

- DNRG Kapell is supported by
  - SABRETRAIN Project, funding the Marie Curie Host Fellowships for Early Stage Research Training as part of the 6th Framework Programme of the European Commission
  - Scottish Government, funding the research

SAC

Success through Knowledge