



Implementing single step GBLUP in pigs

Andreas Hofer
SUISAG

SABRE-TP
12.06.2014 , Zug



Outline



- ▶ What is single step GBLUP?
- ▶ Plan of implementation by SUISAG
- ▶ Validation of genetic evaluations
- ▶ First results
- ▶ Discussion

What is single step GBLUP?

► Traditional several steps:

- Estimation of SNP-effects based on training data
- Estimate genomic breeding values for genotyped selection candidates (dGBV)
- Blending with traditional BVs to obtain GoBV

► Single step:

- Combine phenotypic, pedigree and SNP-data in one single analysis to directly obtain GoBV
- Several studies show same or improved accuracy
- Simpler to implement in established process of BV estimation

BLUP vs ssGBLUP

(Legarra, Christensen, Aguilar, Misztal: Single step, a general approach for genomic selection, Livest Sci 2014)

► MME BLUP
$$\begin{pmatrix} X'X & X'W \\ W'X & W'W + A^{-1}\lambda \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ W'y \end{pmatrix}$$

► MME ssGBLUP
$$\begin{pmatrix} X'X & X'W \\ W'X & W'W + H^{-1}\lambda \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ W'y \end{pmatrix}$$

$$H^{-1} = A^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{pmatrix}$$

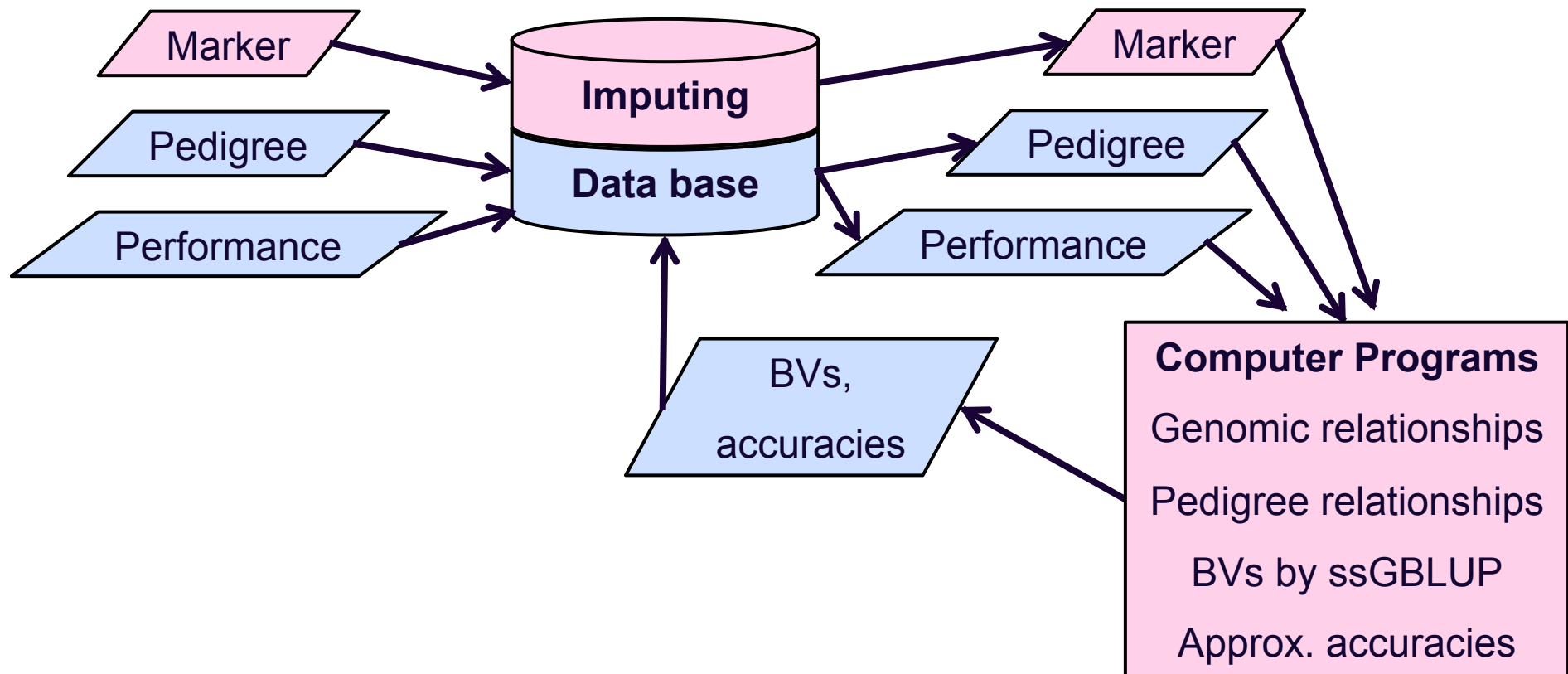
$$H = \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{12}A_{22}^{-1}G \\ GA_{22}^{-1}A_{21} & G \end{pmatrix}$$

$$\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$$



- ▶ Several approaches to construct $\mathbf{G}$ (Van Raden, JDS 2008)
- ▶ $\mathbf{G}$ can be singular $\rightarrow \mathbf{G}_{\alpha} = (\mathbf{1}-\alpha)\mathbf{G} + \alpha\mathbf{A}_{22}$ (usually $0.05 \leq \alpha \leq 0.2$)
 - α can be viewed as proportion of genetic variance not explained by the markers.
- ▶ $\mathbf{G}$ should be „on the same base“ as $\mathbf{A}$
 - Base population allele freq. are not known $\rightarrow \text{tr}(\mathbf{G}) = \text{tr}(\mathbf{A}_{22})$
 - Genotyped animals are not a random sample of base animals (genetic trend, variance) $\rightarrow \mathbf{G}_n = \mathbf{a} + \mathbf{b}\mathbf{G}$
 $\mathbf{a}$ and $\mathbf{b}$ = function of $\text{tr}(\mathbf{G})$, $\text{tr}(\mathbf{A}_{22})$, mean of $\mathbf{G}$ and $\mathbf{A}_{22}$
- ▶ Scale contributions $\rightarrow \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1}$ (usually $\tau \leq 1$, $\omega \leq 1$)
- ▶ Combined: $\tau[\mathbf{a} + \mathbf{b}((1-\alpha)\mathbf{G} + \alpha\mathbf{A}_{22})]^{-1} - \omega\mathbf{A}_{22}^{-1}$

From traditional BLUP to Single Step GBLUP



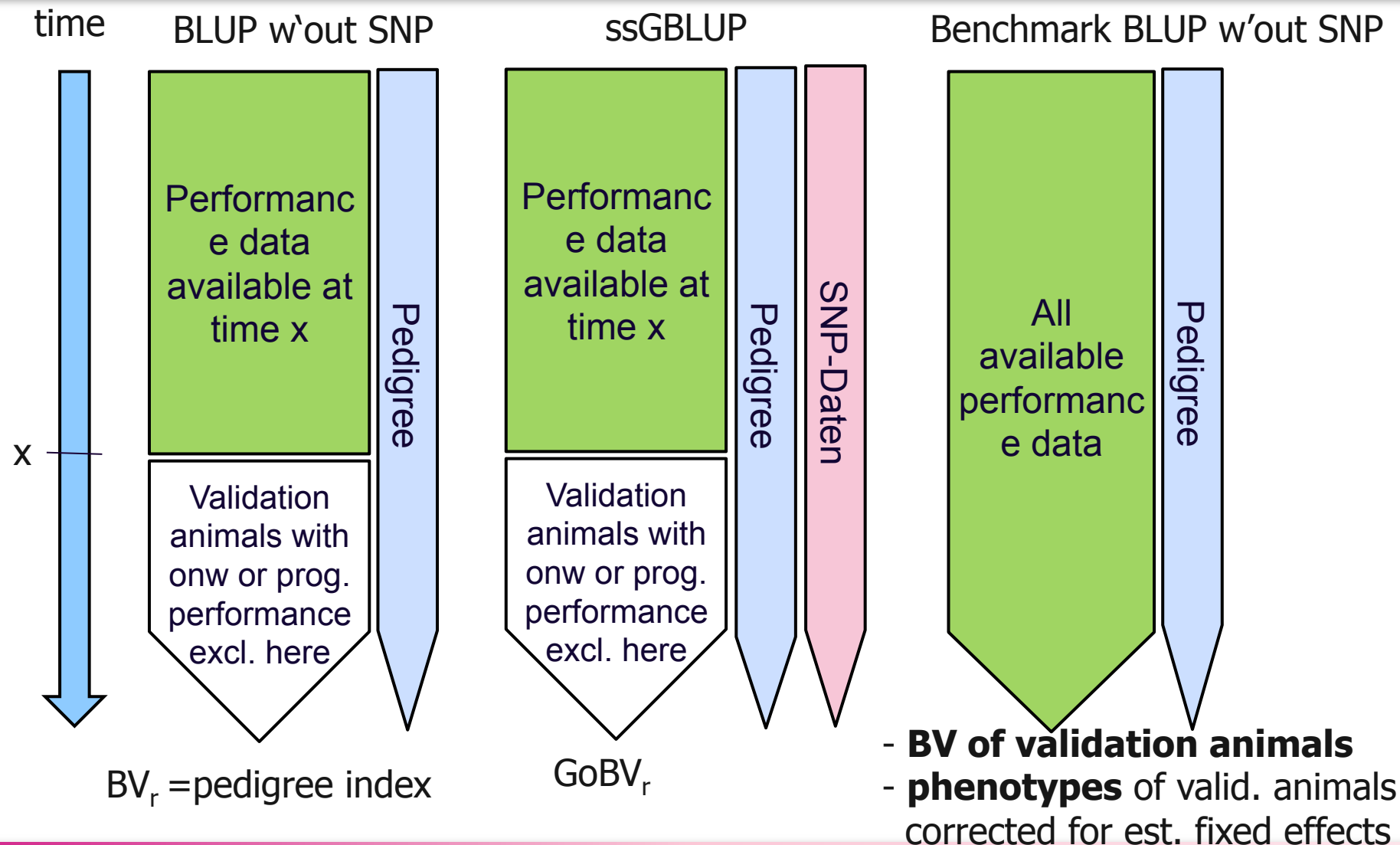


Software



- ▶ Handling SNP-data: ?? so far own programs
- ▶ Imputing: Fimpute (Sargolzaei, licence fee)
- ▶ $G^{-1} - A_{22}^{-1}$: HGinv (Stranden, MTT, with Mix99)
 - Various options available: $\tau[a + b((1-\alpha)G + \alpha A_{22})]^{-1} - \omega A_{22}^{-1}$
- ▶ Solve MME: MiX99 (Lidauer, MTT, licence fee)
 - Variance components not yet implemented for ssGBLUP
 - But estimates are often very similar → little impact ??
- ▶ Approx. accuracies: ??
 - Modify own programs to add genomic contributions according to Misztal et al. 2013 JDS.

Validation of predicting ability of estimated BVs

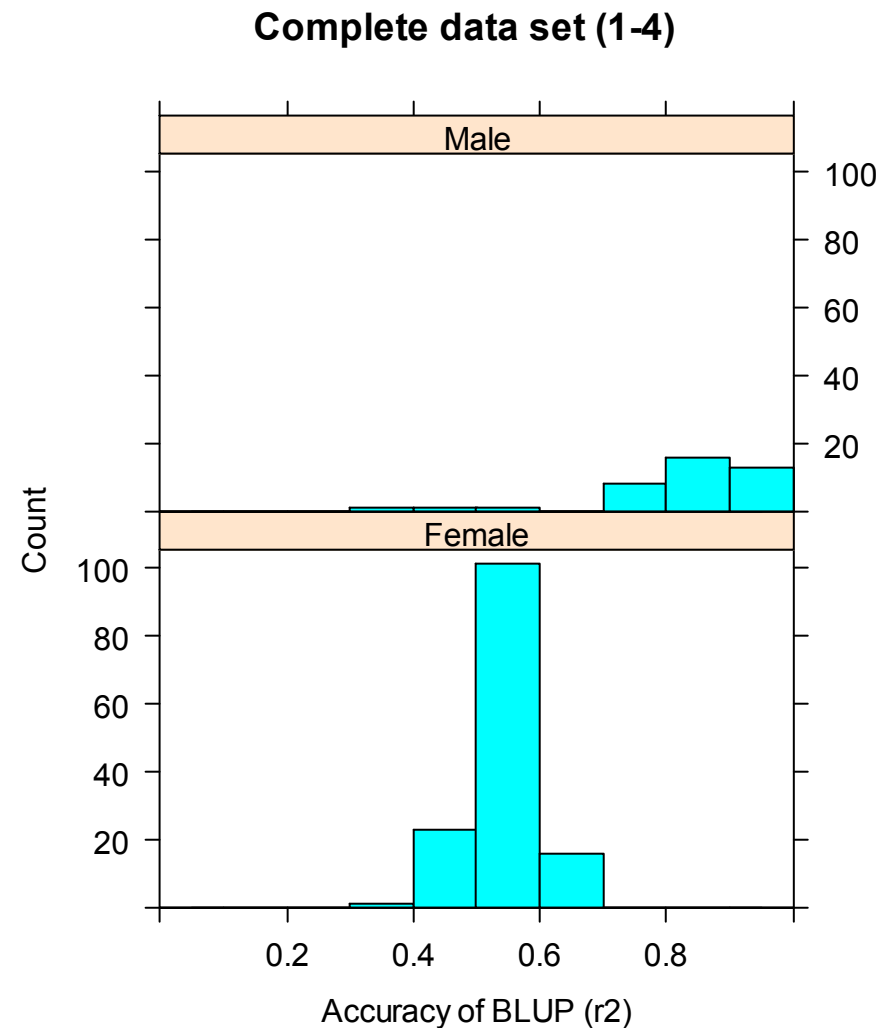
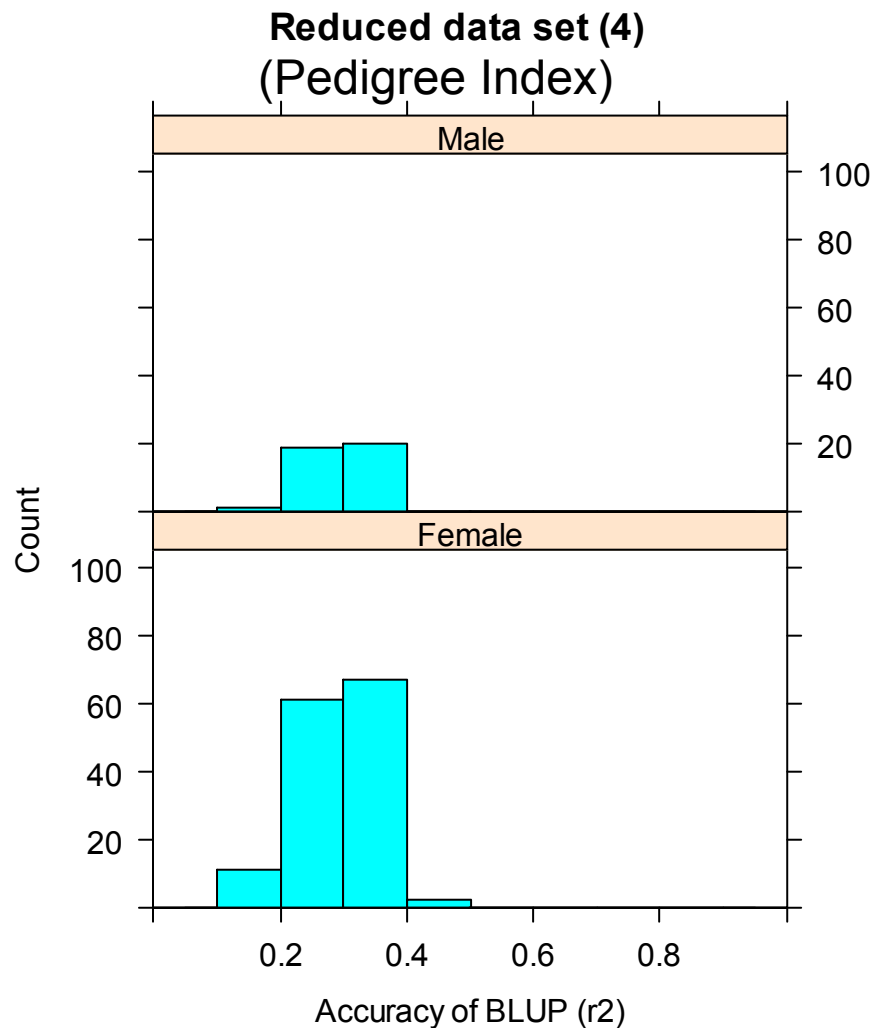


First results: Data sets



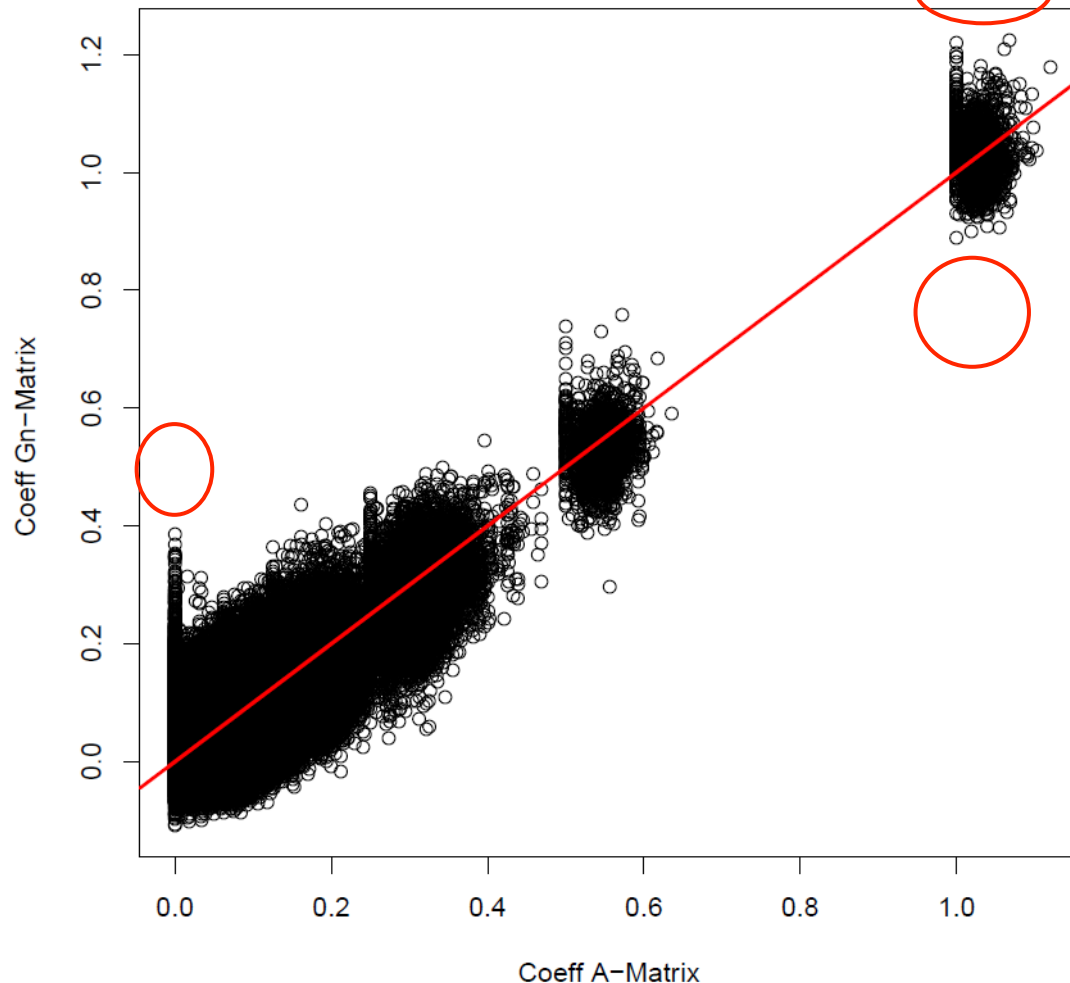
- ▶ Single trait analysis for litter size (NBA)
- ▶ Split data set by birth date of litter at 1.10.2011
- ▶ Validation animals have genotypes (60K) and own or prog. performance only $\geq$ 1.10.2011
 - 40 boars:
 - ∅ each with 116 daughters with 2.9 litters/daughter
 - 141 sows:
 - ∅ each with 5 own litters, 81 also with daughters

Accuracy (r^2) of EBV of validation animals in reduced and complete data



G_n vs A_{22} after exclusion of outliers

Comparison A- with Gn-Matrix (Nanimals= 1596)



Outliers =

- High/low g_{ij}
- Off-diagonals:
 $|a_{ij} - g_{ij}| > 0.4$

5 Pairs of duplicates

Total 59 animals eliminated

Correlations A-G

- Diagonals = **-0.09 ??**
- Off-diagonals = 0.64



Single trait runs



	1	2ab	3ab $\tau\omega$	4ab $\tau\omega$	5old	6cut
BYear ped	1994	1994	1994	1994	1988	1996
BYear SNP	1996	1996	1996	1996	1992	2000
Phen from	2000	2000	2000	2000	1996	2004
α	0.1	0.1	0.1	0.1	0.1	0.1
a/b	0/1	0.07/0.94	0.07/0.94	0.07/0.94	0.08/0.93	0.07/0.94
τ/ω	1/1	1/1	1.5/0.6	1/0.7	1/0.7	1/0.7

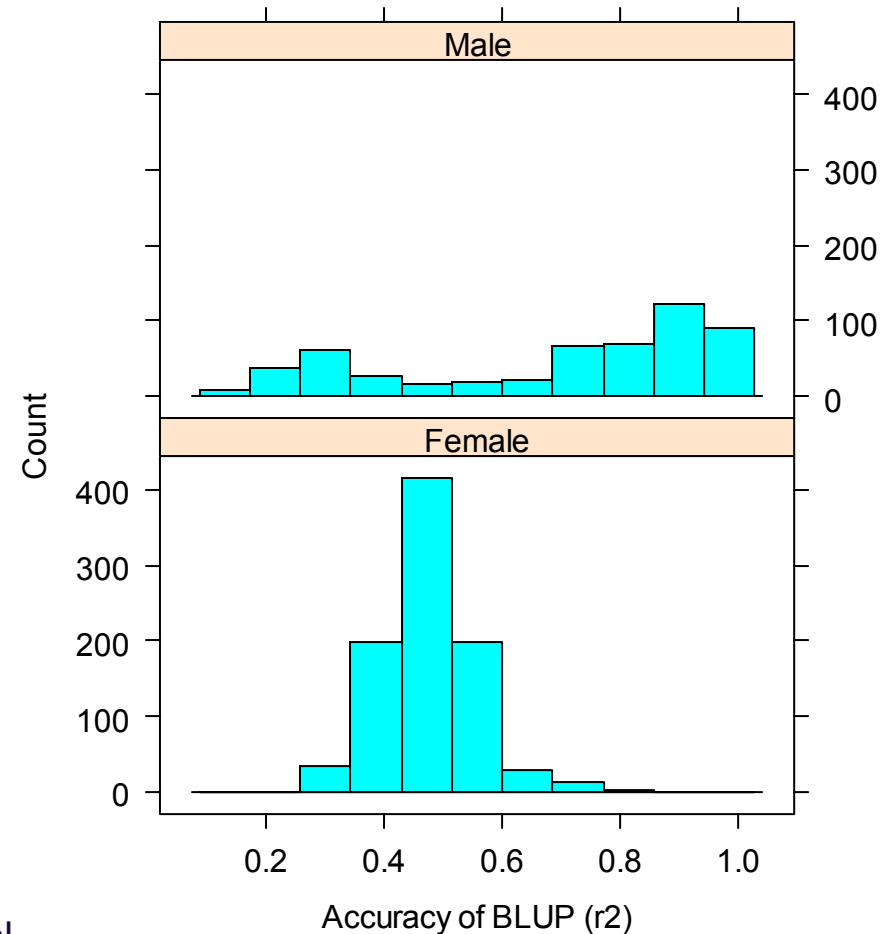
Reference animals in runs 1-4

► 533 males
 $r^2_{\hat{a},a} = 0.69$
 (only 254 with $r^2_{\hat{a},a} > 0.8$)

► 885 females
 $r^2_{\hat{a},a} = 0.47$

► Limited information from SNPs, because gain in accuracy for animal $i = f(\sum (g_{ij} - a_{ij})^2 * r^2_{\hat{a},a_j})$ (Misztal et al. 2013)

Reference animals red data set (4) (N= 1418)



Validation criteria = forward prediction of BV or phenotypes

► Breeding values of validation animals (40m,141f)

- Relative increase of correlation with BVc (complete data, no SNP)

$$r(\text{GoBVr}, \text{BVc}) / r(\text{BVr}, \text{BVc})$$

- Regression of BVc on GoBVr or BVr should be 1

$$\text{BVc} = b_0 + b_1 \times \text{GoBVr} + e \quad \text{vs}$$

$$\text{BVc} = b_0 + b_1 \times \text{BVr} + e$$

► Same for phenotypes corrected for fixed effects (with complete data, no SNP) averaged per sow

- $r(\text{GoBVr}, y_{\text{cor}}) / r(\text{BVr}, y_{\text{cor}})$

- $y_{\text{cor}} = b_0 + b_1 \times \text{GoBVr} + e \quad \text{vs}$

$$y_{\text{cor}} = b_0 + b_1 \times \text{BVr} + e$$



Forward prediction of BVs

	1	2ab	3ab $\tau\omega$	4ab $\tau\omega$	5old	6cut
BY2010-2004 c	1.29	1.29	1.29	1.29	1.38	1.18
BY2010-2004 r	1.18	1.18	1.18	1.18	1.26	1.09
BY2010-2004 rS	0.80	1.14	1.02	1.09	1.17	1.01
r(BVr, BVc) m	0.61	0.61	0.61	0.61	0.62	0.60
r(GoBVr, BVc) m	0.72	0.72	0.72	0.72	0.72	0.72
rel. increase	1.17	1.17	1.18	1.17	1.16	1.20
b1(BVr) m	0.86	0.86	0.86	0.86	0.88	0.85
b1(GoBVr) m	0.90	0.88	1.14	1.04	1.03	1.05
r(BVr, BVc) f	0.65	0.65	0.65	0.65	0.66	0.65
r(GoBVr, BVc) f	0.62	0.62	0.67	0.68	0.67	0.66
rel. increase	0.95	0.96	1.04	1.04	1.03	1.02
b1(BVr) f	0.93	0.93	0.93	0.93	0.95	0.94
b1(GoBVr) f	0.65	0.67	1.03	0.93	0.93	0.94



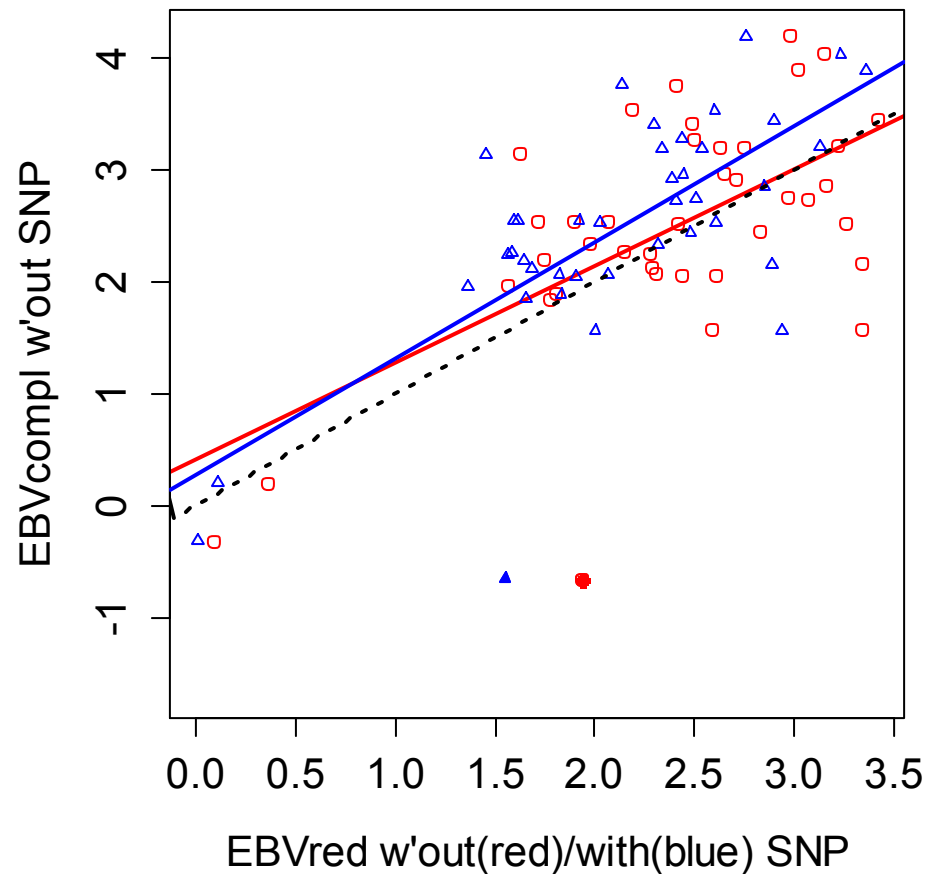
Forward prediction of averaged phenotypes of 141 valid. females

	1	2ab	3ab $\tau\omega$	4ab $\tau\omega$	5old	6cut
r(BVr, y_cor)	0.21	0.21	0.21	0.21	0.22	0.22
r(GoBVr, y_cor)	0.28	0.25	0.25	0.26	0.26	0.25
rel. Increase	1.34	1.21	1.22	1.22	1.19	1.13
b1(BVr)	0.84	0.84	0.84	0.84	0.89	0.90
b1(GoBVr)	0.84	0.77	1.09	0.98	1.01	1.01

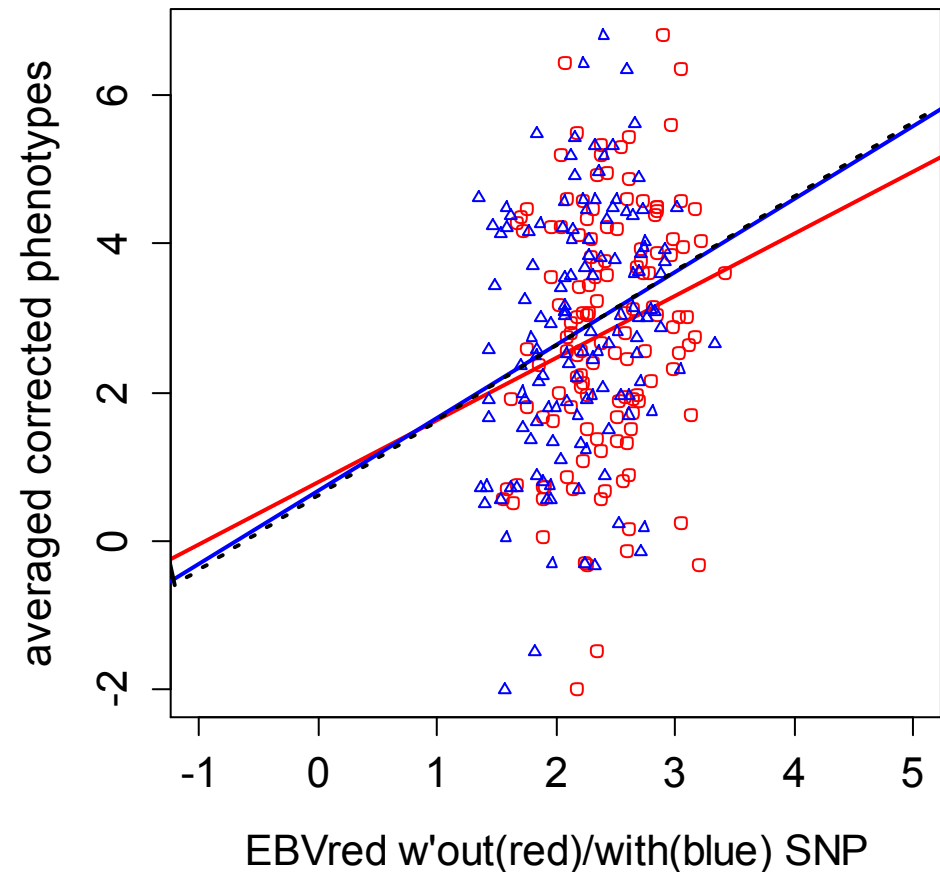
- Larger increase of accuracy if sire genotyped (not shown)
- Conclusion: 4 seems to provide „best“ predictions
 - Interaction among parameters $\alpha, a/b, \tau/\omega$?
 - Repeatable for other validation animals?
 - Best option for litter size also best for other traits?
- **additional validation runs needed!**

So far only moderate improv. of predictions (plots for run 4)

Validation boars



Validation phenotypes





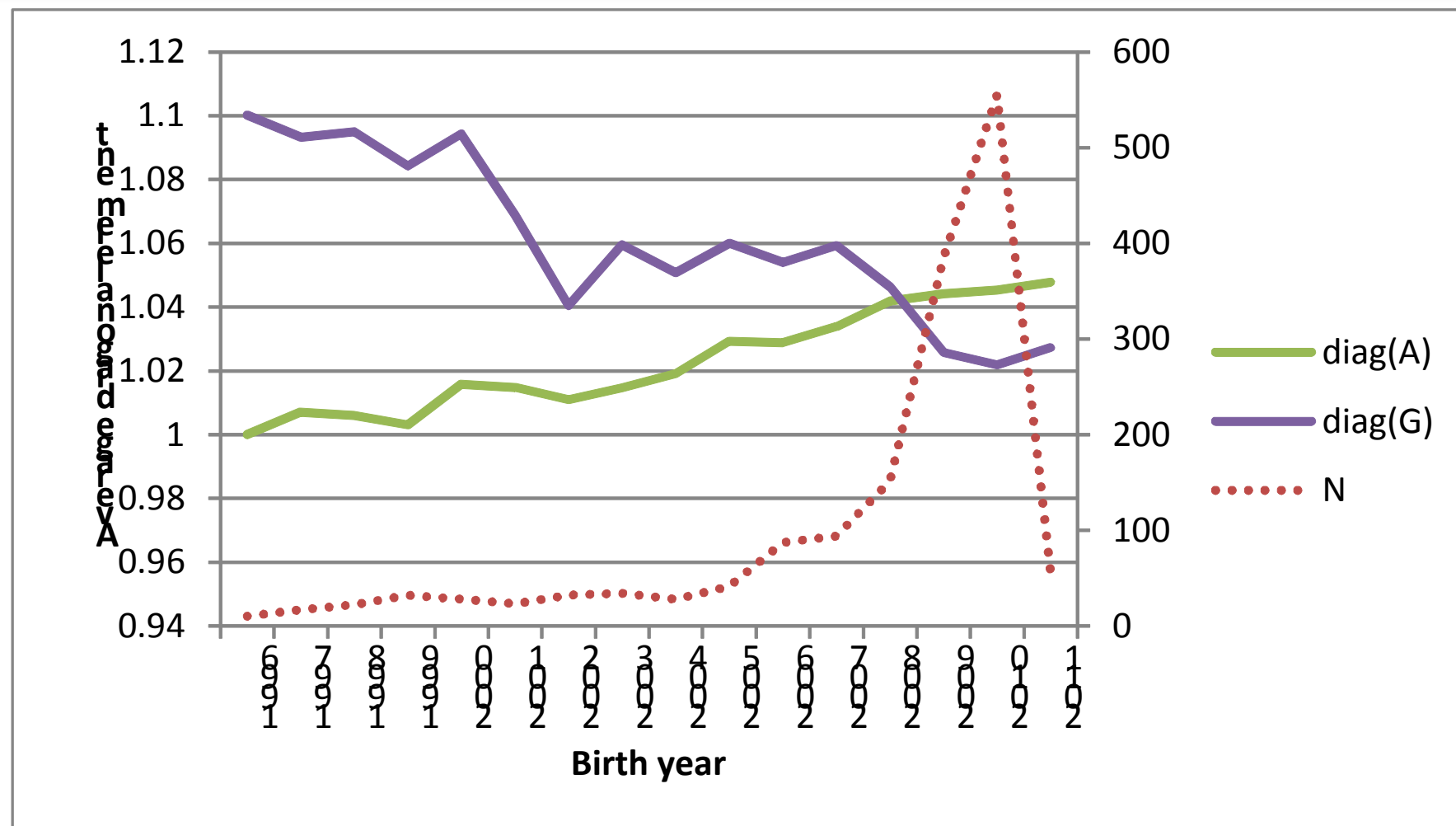
► Method of validation

- Aim of genetic evaluation is prediction → forward prediction is method of choice
- Used BVc or y_{cor} as surrogate for the unknown true breeding value of validation animals
 - Others use DYDs or deregressed proofs ?
 - The variable used should be highly correlated to true breeding value and independent of prediction method (at least not favouring SNP)

► Lack of correlation bw diagonals of A and G

- Diag(A) reflect inbreeding → pedigree depth?
 - Adding 2 years of pedigree in relation to SNP-animals did not help
- Diag(G) reflect homozygosity, but $E(G) = A$
- What could be potential problems?

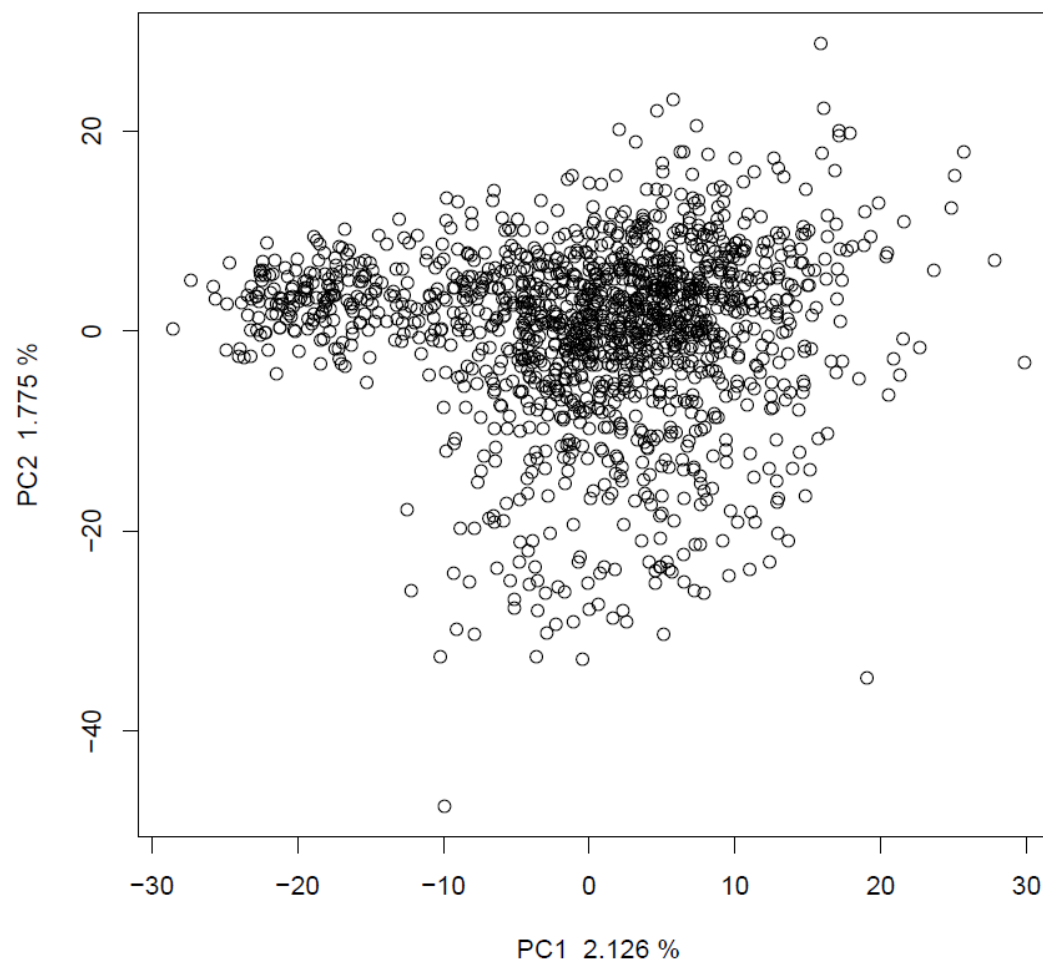
Diagonal elements of A and G vs birth year (run 5)



PCA SNPs before Imputation

(only SNPs with no missing genotypes, run 6)

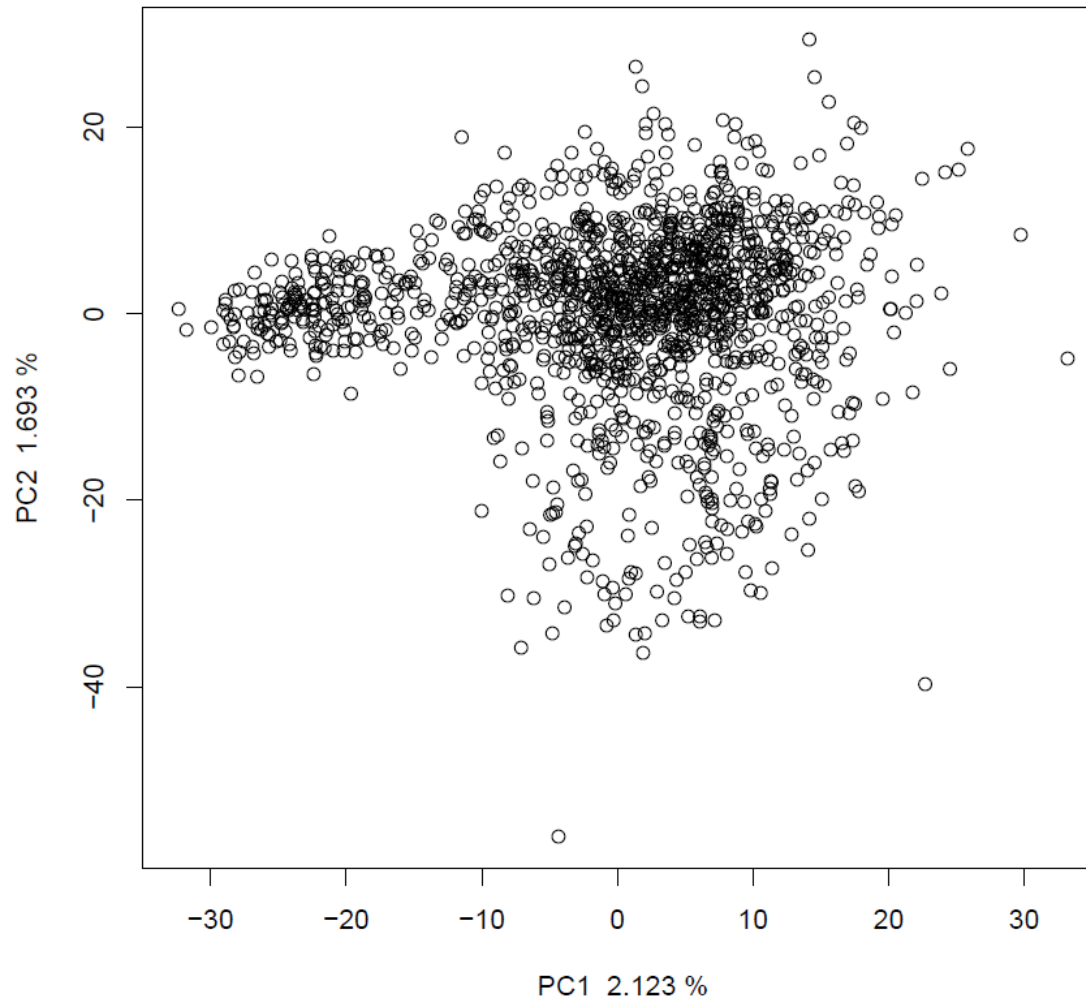
PCA Ntiere= 1471 Nsnp= 17560



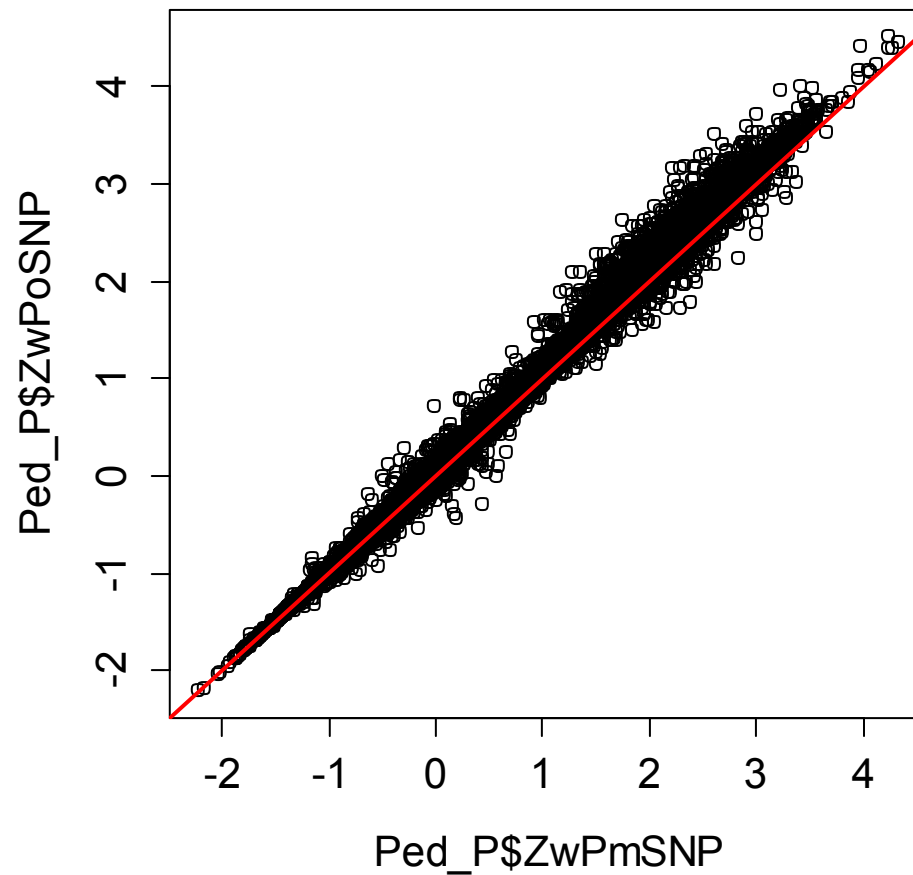
PCA SNPs after Imputation

(every 2.5nd SNP, run 6)

PCA Ntiere= 1471 Nsnp= 18564

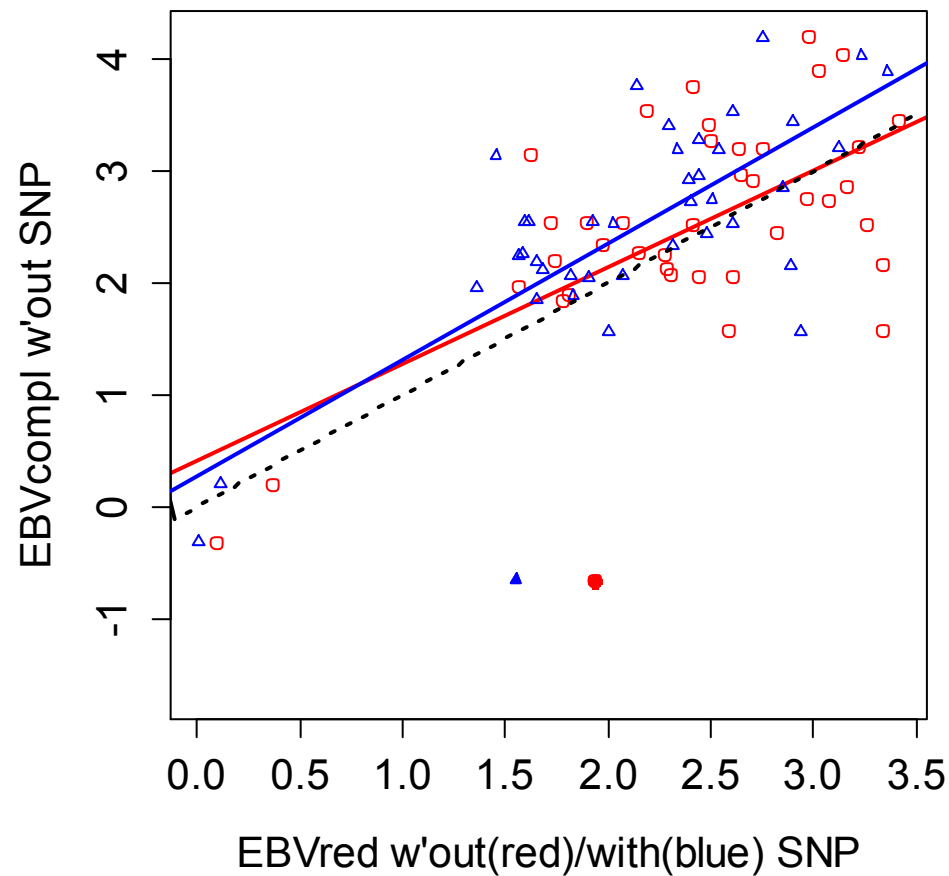


BVr vs GoBVr (run 4)



GoBVR or BVR vs BVC for 40 boars (run 4)

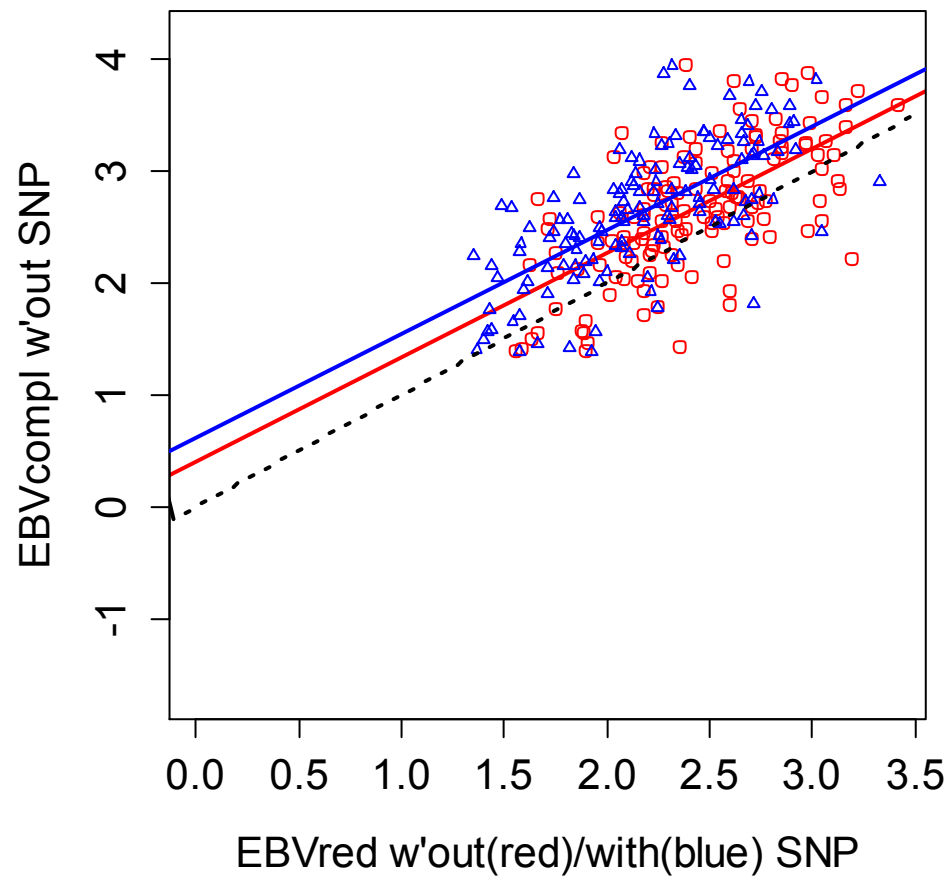
Validation boars



BVR:
 $r = 0.61$, $b1 = 0.86$
 GoBVR:
 $r = 0.72$, $b1 = 1.04$

GoBVR or BVR vs BVC for 141 females (run 4)

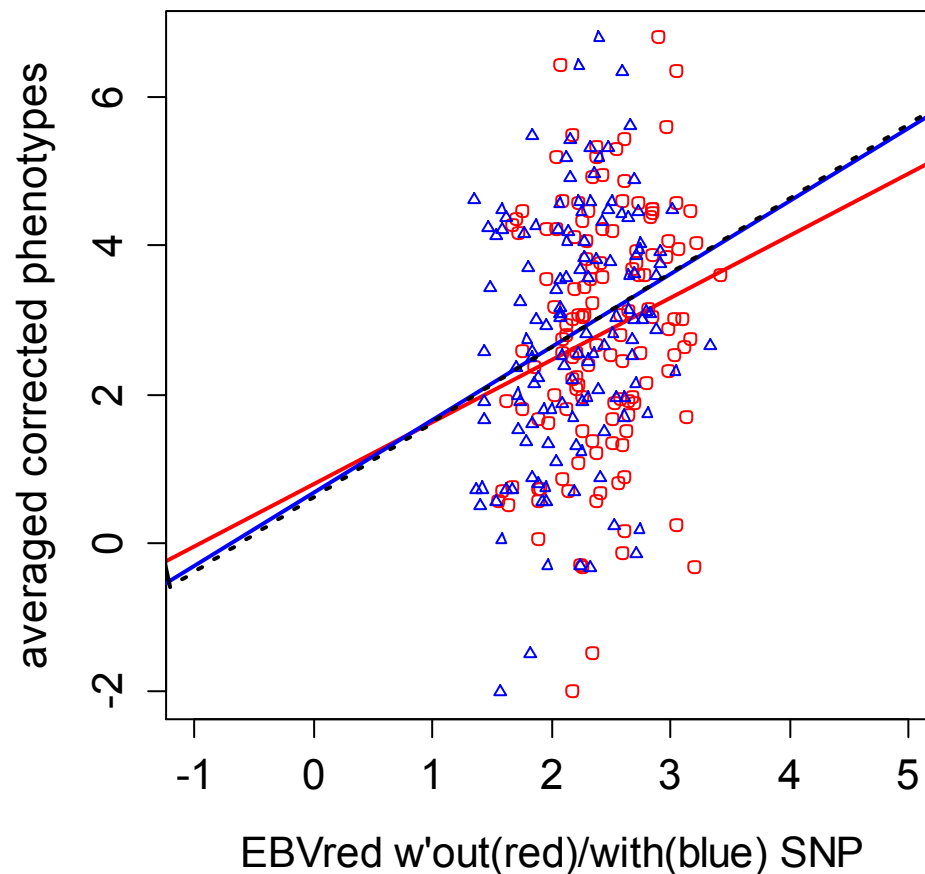
Validation sows



BVR:
 $r = 0.65$, $b1 = 0.93$
 GoBVR:
 $r = 0.68$, $b1 = 0.93$

GoBvR or BVr vs average of y_cor for 141 females (run 4)

Validation phenotypes



BVr:
 $r = 0.21$, $b1 = 0.84$
 GoBvR:
 $r = 0.26$, $b1 = 0.98$