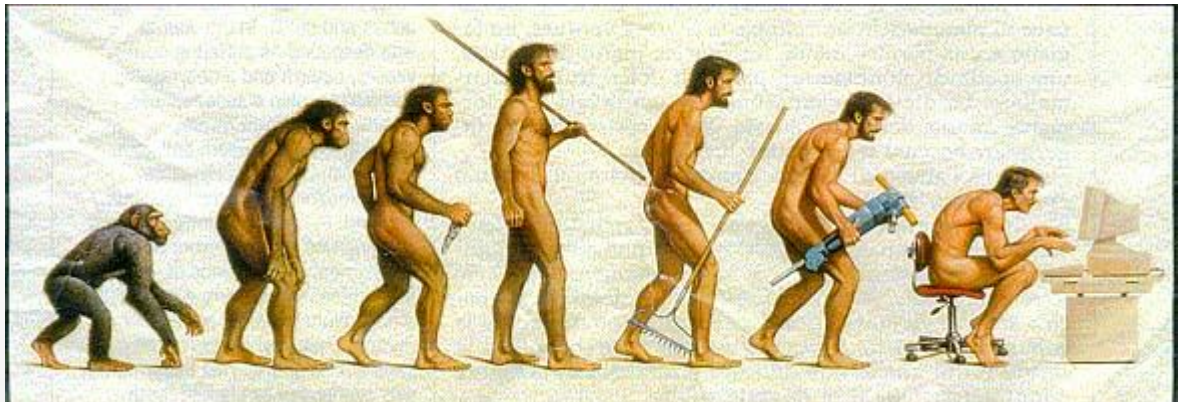
Current single-step national beef cattle evaluation models used by the American Hereford Association and International Genetic Solutions, computational aspects, and implications of marker selection

B. L. Golden[1], M. L. Spangler[2], W. M. Snelling[3], D. J. Garrick [1,4]

[1]Theta Solutions, LLC; [2]University of Nebraska, Lincoln; [3]USDA Meat Animal Research Center; [4]Massey University, Palmerston North, NZ

Topics

- Marker Effects Model
- Traits
  - Using Marker Subsets
  - MCMC Sampling for Accuracy



Using technology built for computer gaming

Single Step Super Hybrid Marker Effects Model

- Both Genotyped and Non-genotyped Animals
- Multiple Traits
- Multiple Components
- Extra Polygenic Effects

Super Hybrid Marker Effects Model

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = \begin{bmatrix} X_n \\ X_g \end{bmatrix} \beta + Z u_{1-c} + \begin{bmatrix} Z_n^* & 0 \\ 0 & Z_g^* M_g \end{bmatrix} \begin{bmatrix} u_c \\ \alpha \end{bmatrix} + e$$

$$Z = \begin{bmatrix} Z_n & Z_g \end{bmatrix} = \begin{bmatrix} Z_n^* & 0 \\ 0 & Z_g^* \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} A^{nn} & A^{ng} \\ A^{gn} & A^{gg} \end{bmatrix}$$

$$\begin{bmatrix} X'X & X'Z & X'Z_n & Z'Z_g M_g \\ S & Z'Z + A^{-1}\frac{\sigma_e^2}{(1-c)\sigma_g^2} & Z'Z_n & Z'Z_g M_g \\ S & S & Z_n'Z_n + A^{nn}\frac{\sigma_e^2}{c\sigma_g^2} & A^{ng} M_g \frac{\sigma_e^2}{c\sigma_g^2} \\ S & S & S & Q \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u}_{1-c} \\ \hat{u}_c \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z_n'y \\ M_g'Z_g'y \end{bmatrix}$$

$$Q = M_g' Z_g' Z_g M_g + I\frac{\sigma_e^2}{\sigma_\alpha^2} + M_n' A^{nn} M_n \frac{\sigma_e^2}{c\sigma_g^2}$$

$$Q = M_g^{'} Z_g^{'} Z_g M_g + I \frac{\sigma_e^2}{\sigma_\alpha^2} + M_n^{'} A^{nn} M_n \frac{\sigma_e^2}{c\sigma_g^2}$$

$$A^{-1} = \begin{bmatrix} A^{nn} & A^{ng} \\ A^{gn} & A^{gg} \end{bmatrix}$$

$$M_n = A_{ng} A_{gg}^{-1} M_g$$

$$A^{nn} m_{i\_n} = -A^{ng} m_{i\_g}$$

$$-M_g^{'} A^{gn} M_n = M_n^{'} A^{nn} M_n$$

EBV = Genomic Effect + Extra Polygenic Effect

Genotyped animals:

$$\hat{u}_g = M_g\hat{\alpha} + \hat{u}_{c-1,\_g}$$

Non-genotyped animals:

$$\hat{u}_n = \hat{u}_c + \hat{u}_{c-1,\_n}$$

Other Effects

- Breed-Year (cluster) additive genetic groups
- Heterosis: two-breed type out-crossing equations
- Contemporary group
- J equation – genetic merit of genotyped animals
- K equation – centering across all loci

AHA Traits Produced

- Birth wt.
- Weaning wt.
- Yearling wt.
- Milk
- Total maternal
- Calving ease
- Calving ease total maternal

Carcass wt.

REA

Marbling score

Fat thickness

Sus Cow Fert

DMI

Udder & Teat

IGS Traits Produced

- Birth wt.
- Weaning wt.
- Yearling wt.
- Milk
- Total maternal
- Calving ease
- Calving ease total maternal

Carcass wt.

REA

Marbling score

Fat thickness

[Yield grade]

Stayability

Docility

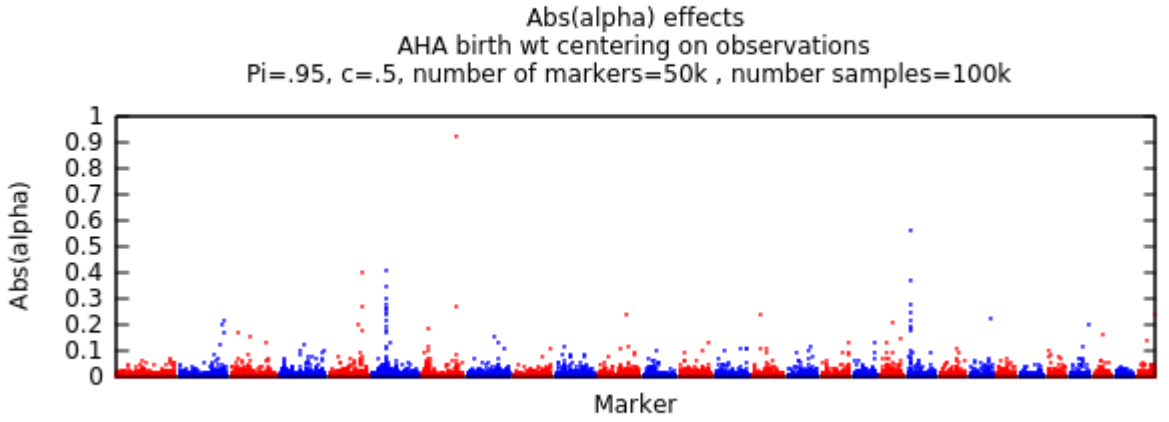
Marker Subsets Illustration

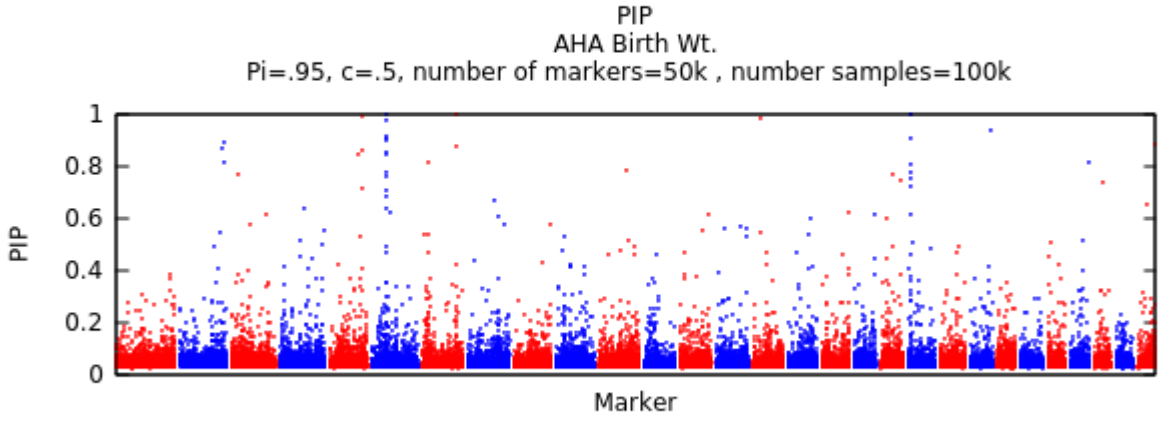
- AHA Birth weight data
  - Training data (born before 1/1/2018):
    - 236,463 birth weight phenotypes
    - 3 generation pedigree
    - 35,386 were genotyped – imputed to 50k
  - Test data (born after 1/1/2018)
    - 9,538 genotyped animals with phenotype

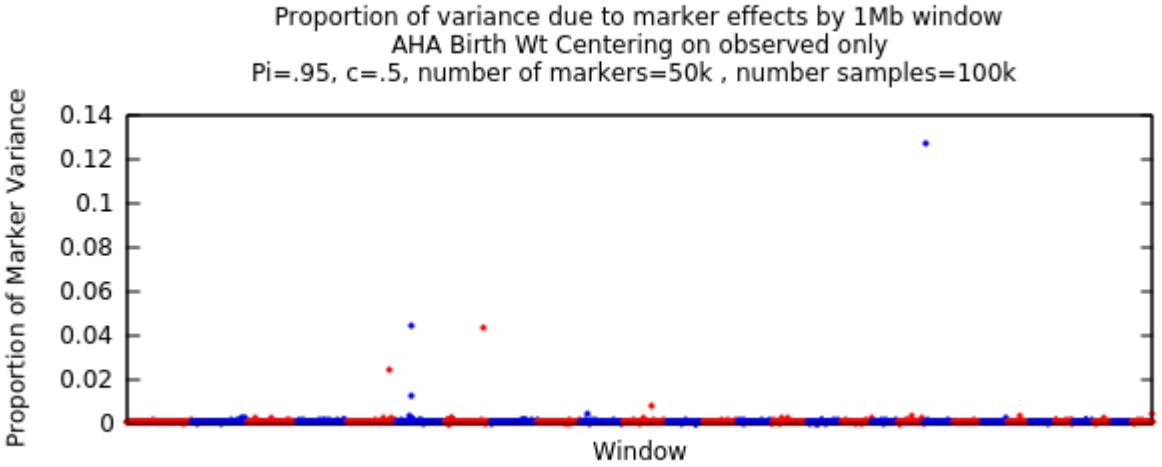
Step 1

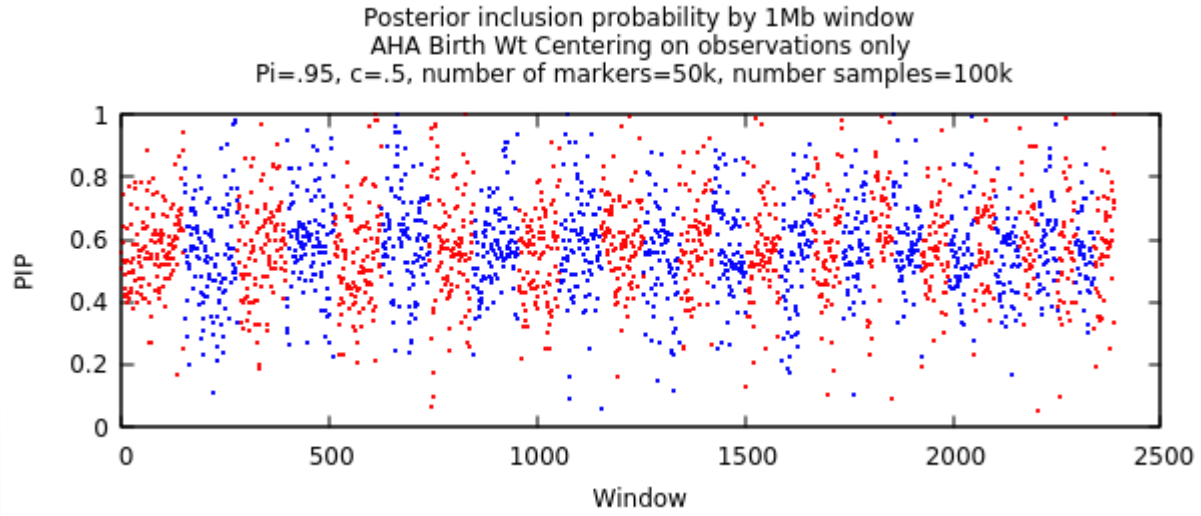
- Bayes C π=.95 on the **training data**
  - c = .5
  - MCMC Gibbs Sampler
    - 4 parallel chains
    - Seeded with PCG solutions
    - Burn in 5,000 samples
    - Sampled 25,000 x 4 = 100,000 samples
  - Posterior distributions and posterior inclusion probabilities (PIP)

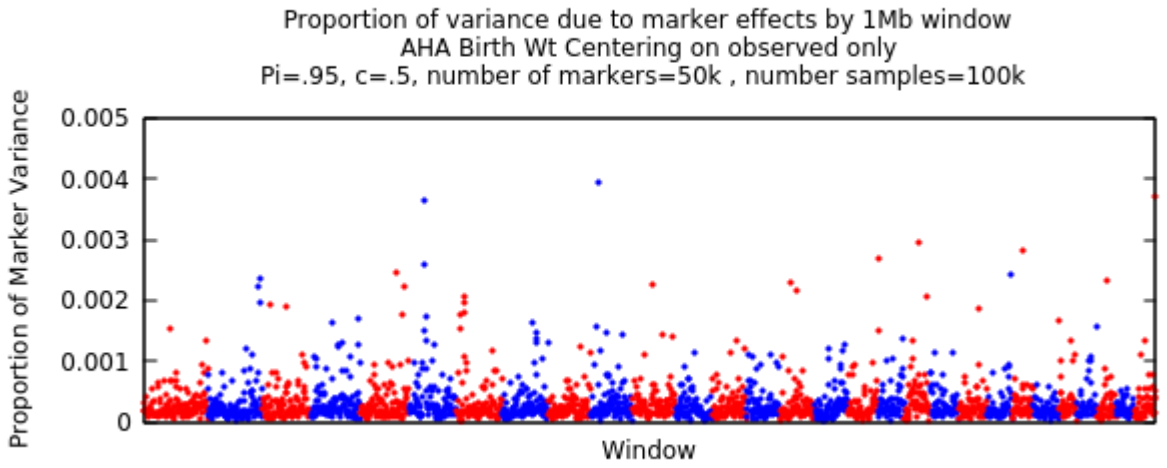
Marker Posterior Means and PIP

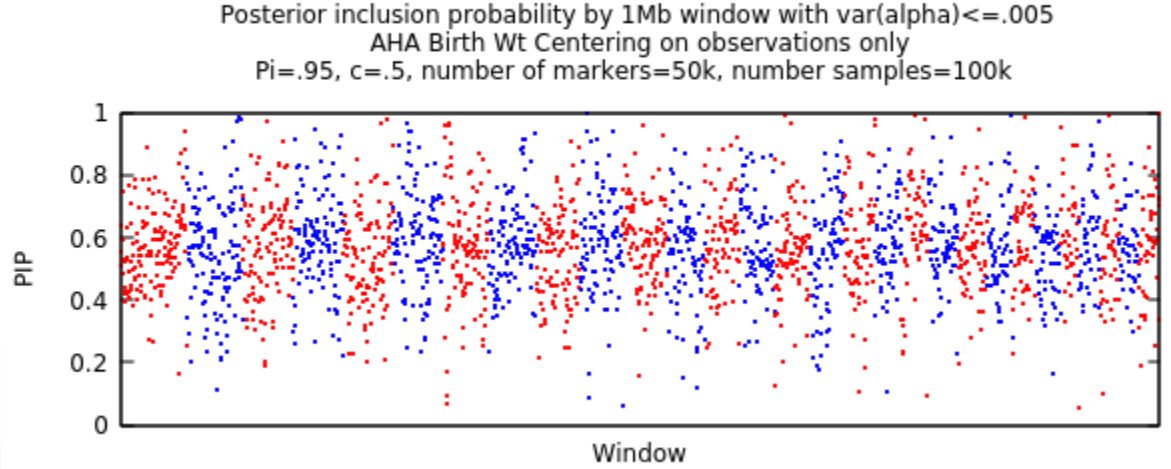





Proportion of variance due to marker effects by 1Mb window
AHA Birth Wt Centering on observed only
Pi=.95, c=.5, number of markers=50k , number samples=100k
Proportion of Marker Variance
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Window


Posterior inclusion probability by 1Mb window
AHA Birth Wt Centering on observations only
Pi=.95, c=.5, number of markers=50k, number samples=100k
PIP
1
0.8
0.6
0.4
0.2
0
0
500
1000
1500
2000
2500
Window

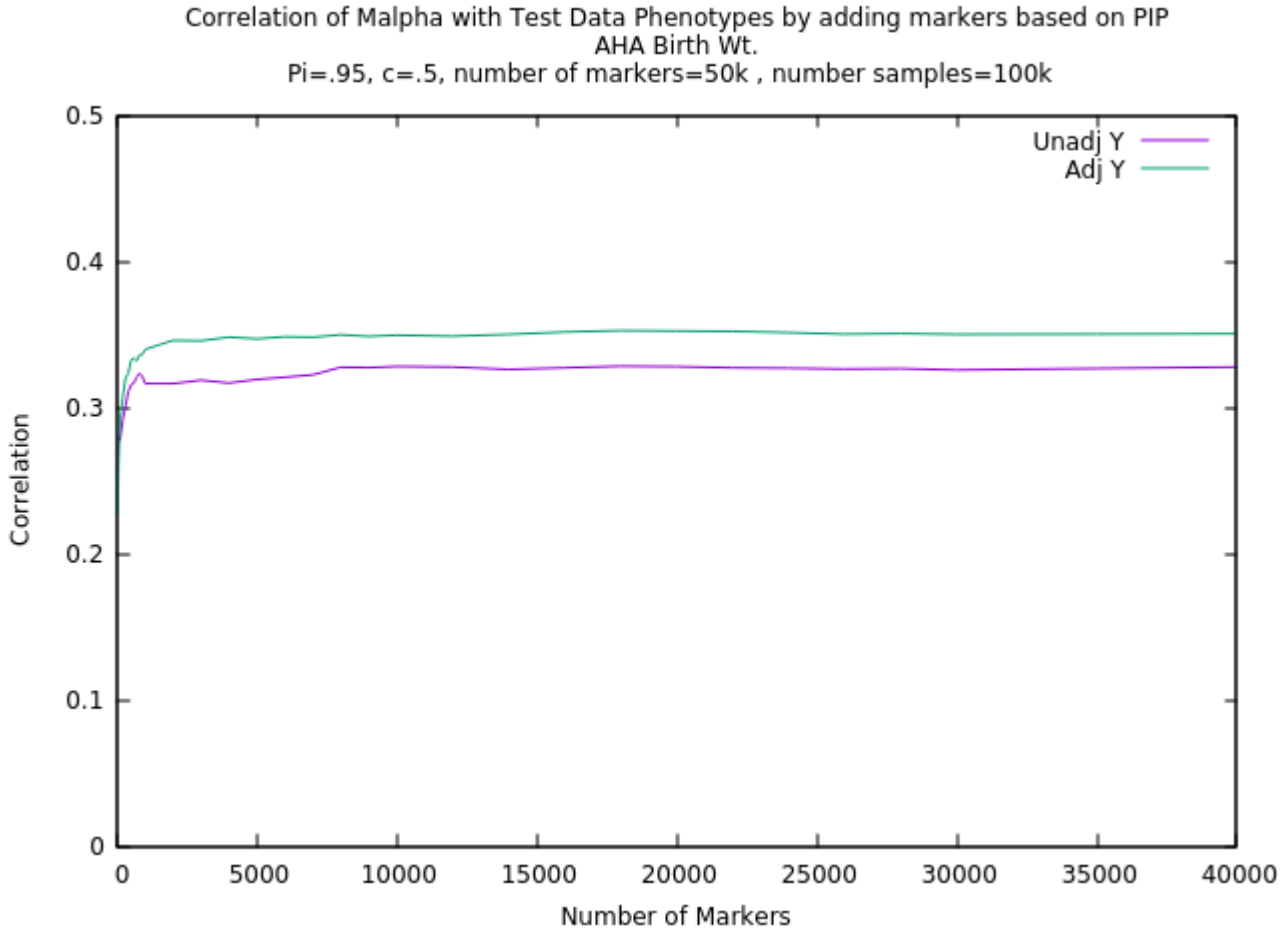

Proportion of variance due to marker effects by 1Mb window
AHA Birth Wt Centering on observed only
Pi=.95, c=.5, number of markers=50k , number samples=100k
Proportion of Marker Variance
0.005
0.004
0.003
0.002
0.001
0
Window

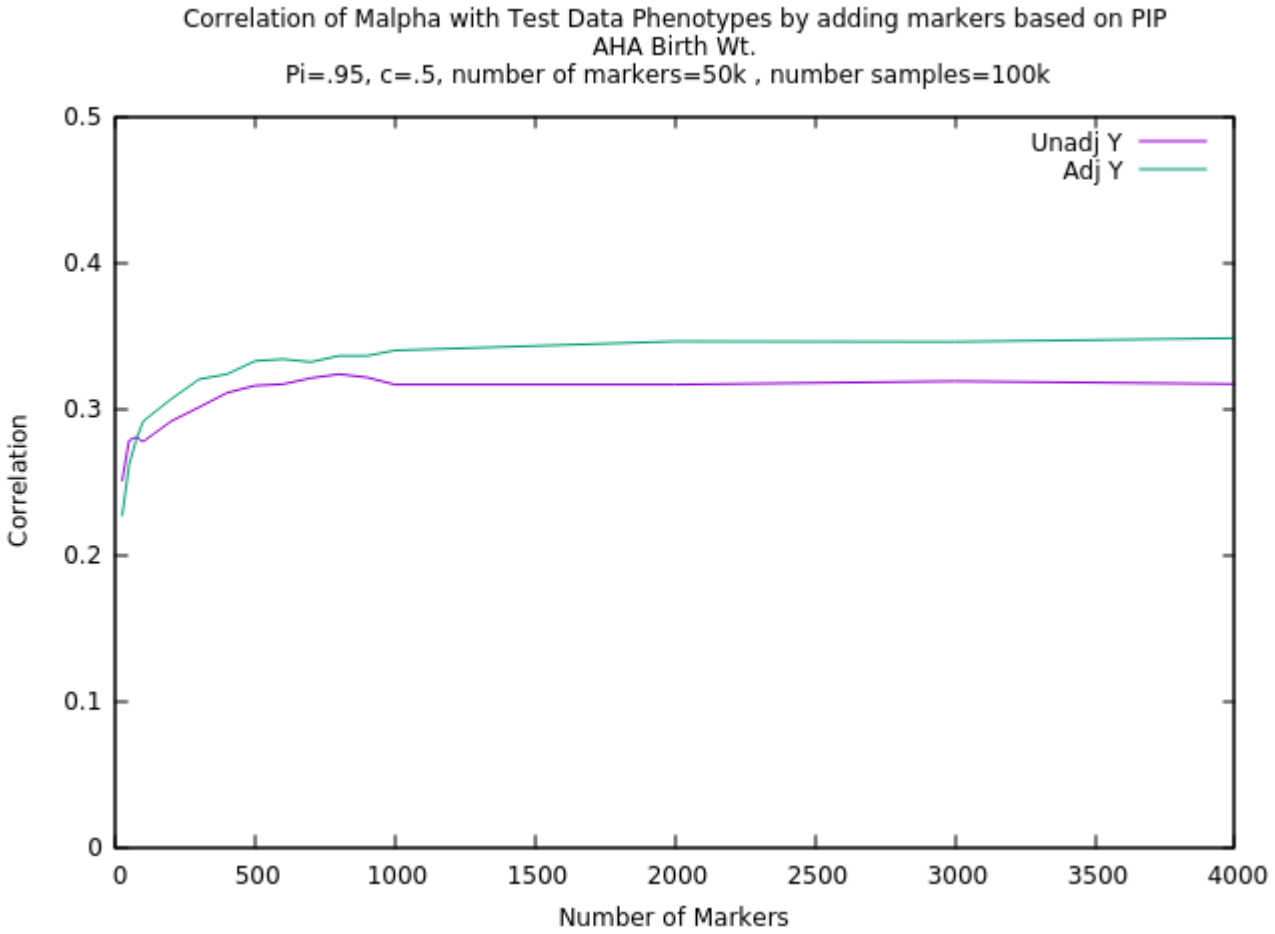

Posterior inclusion probability by 1Mb window with var(alpha)<=.005
AHA Birth Wt Centering on observations only
Pi=.95, c=.5, number of markers=50k, number samples=100k
PIP
1
0.8
0.6
0.4
0.2
0
Window

Step 2

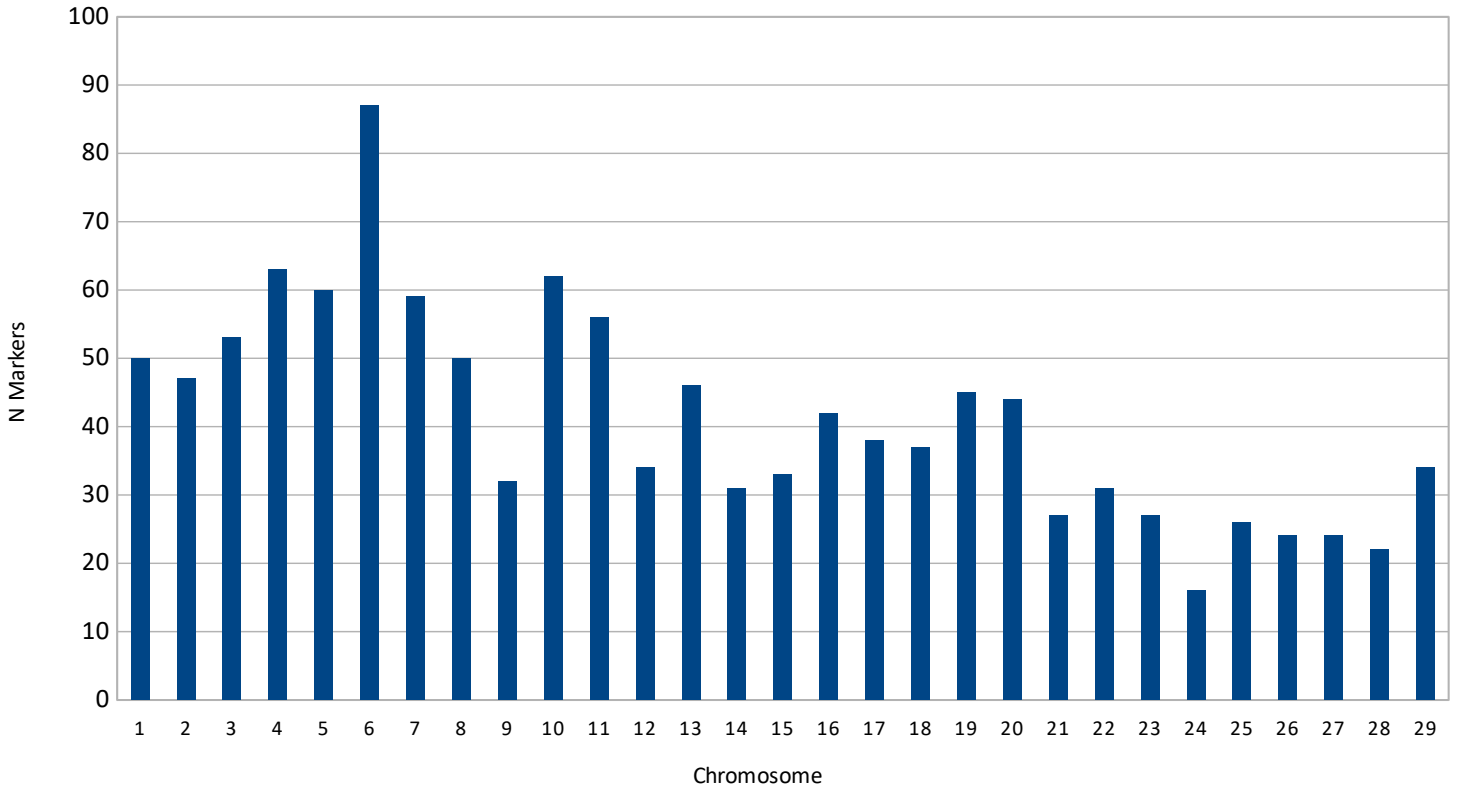
- Successive Bayes C π=0 on the **training data**
  - Adding markers by next highest PIP
- Compute prediction accuracy of subsets in **test data** (t)
  - Using marker subset posterior means from **training data**

$$r_{p,i} = Correlation(M_{t,i}\hat{\alpha}_i, P_t)$$

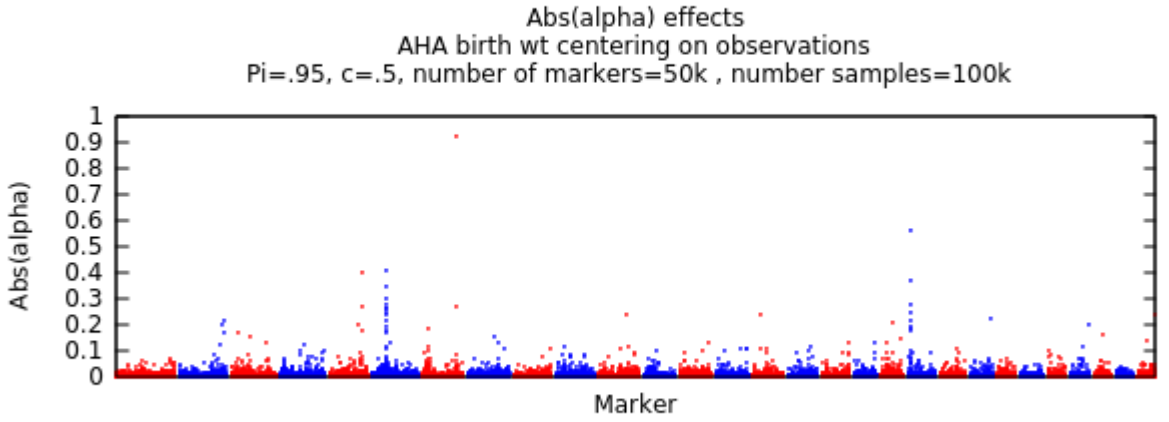

Correlation of Malpha with Test Data Phenotypes by adding markers based on PIP
AHA Birth Wt.
Pi=.95, c=.5, number of markers=50k , number samples=100k
Unadj Y
Adj Y
Correlation
Number of Markers
0
0.1
0.2
0.3
0.4
0.5
0
5000
10000
15000
20000
25000
30000
35000
40000

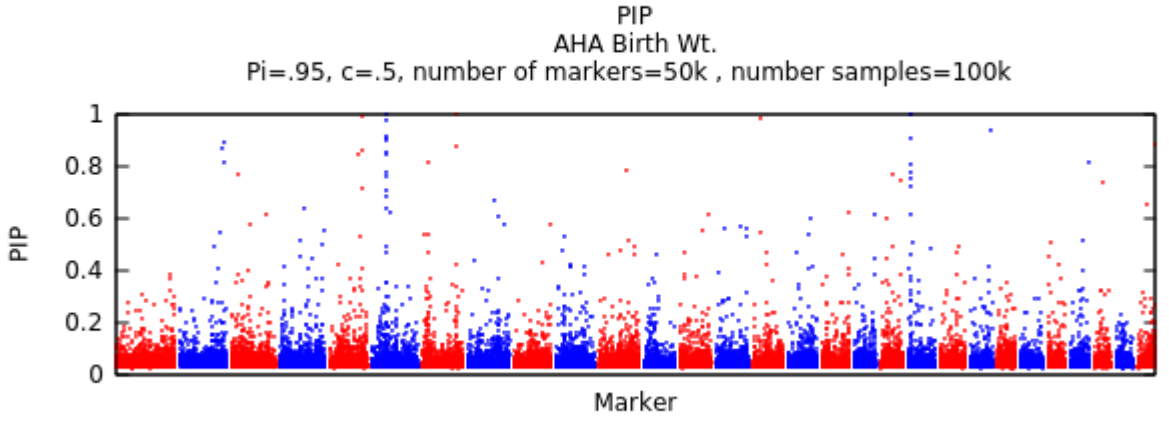

Correlation of Malpha with Test Data Phenotypes by adding markers based on PIP
AHA Birth Wt.
Pi=.95, c=.5, number of markers=50k , number samples=100k
Unadj Y
Adj Y
Correlation
Number of Markers
0
0.1
0.2
0.3
0.4
0.5
0
500
1000
1500
2000
2500
3000
3500
4000

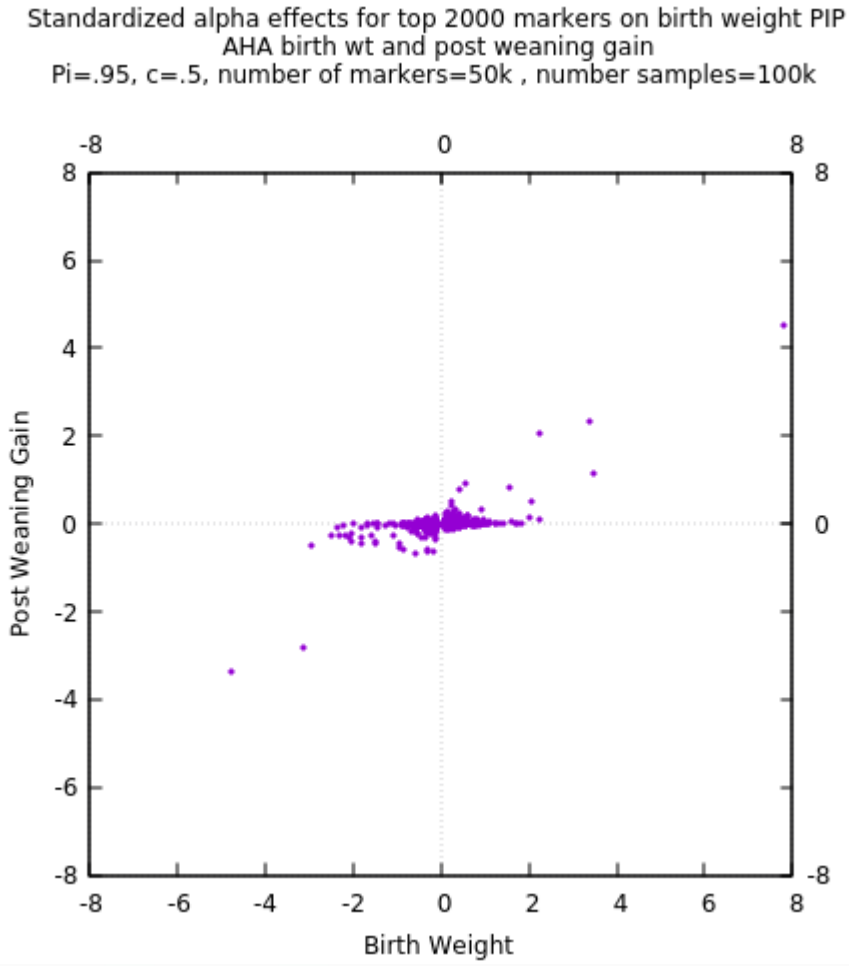
1,200 Highest PIP Markers by Autosome

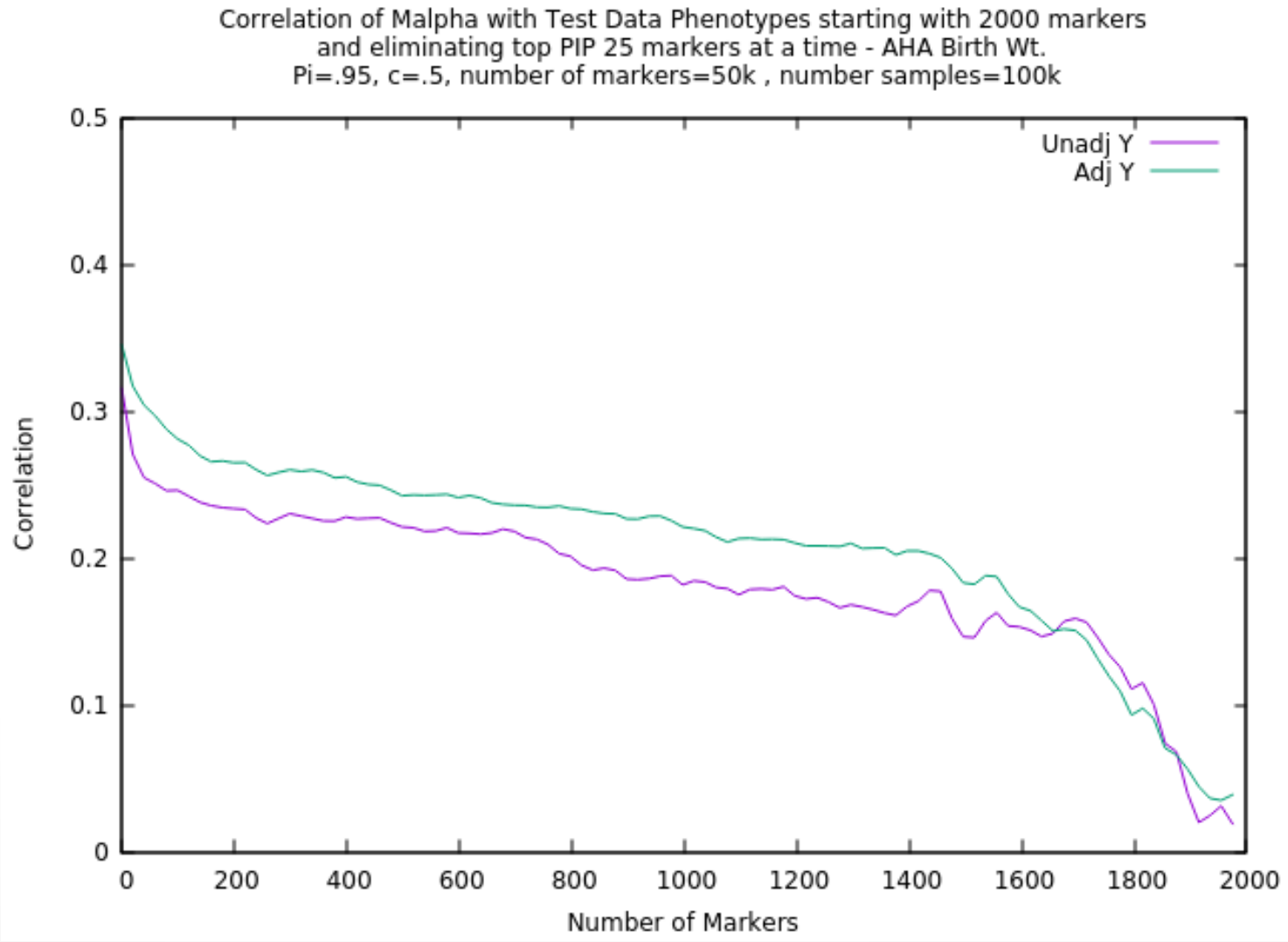


Marker Posterior Means and PIP






Standardized alpha effects for top 2000 markers on birth weight PIP
AHA birth wt and post weaning gain
Pi=.95, c=.5, number of markers=50k , number samples=100k
Post Weaning Gain
Birth Weight
-8
-6
-4
-2
0
2
4
6
8


Correlation of Malpha with Test Data Phenotypes starting with 2000 markers
and eliminating top PIP 25 markers at a time - AHA Birth Wt.
Pi=.95, c=.5, number of markers=50k , number samples=100k
Unadj Y
Adj Y
Correlation
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Number of Markers

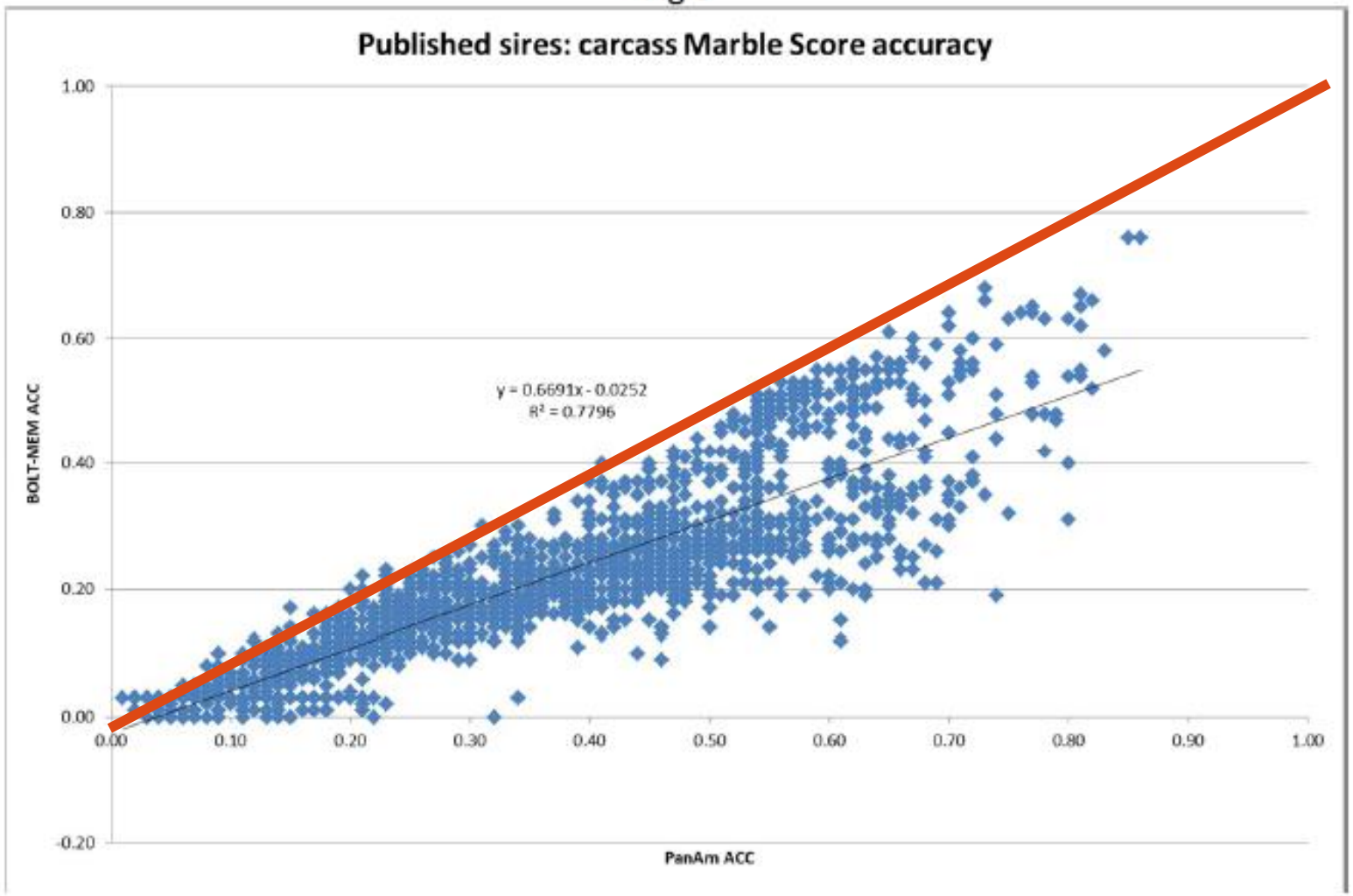
Accuracy Calculation

- No “Approximation Bias”
- “Direct” method to estimate Prediction Error Variance (PEV)

$$BIFAccuracy = 1 - \sqrt{\frac{PEV}{(1+f) * GeneticVariance}}$$

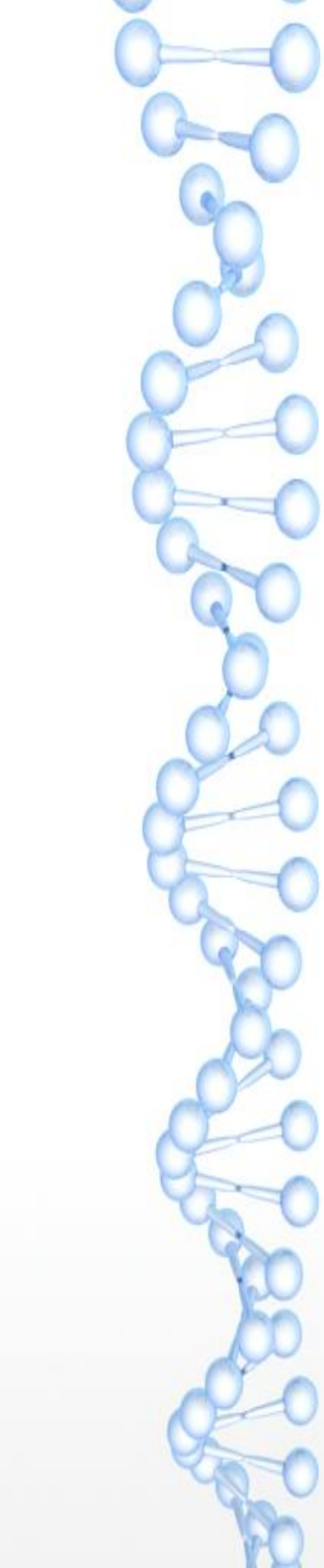
What is the Effect on Accuracy Values?

Fig 6



Conclusions

- Marker effects models are feasible
- MCMC Gibbs Sampling is feasible
- Our biggest opportunity is improved markers



Questions?

Completely Reimagined Models for All Traits

- Stayability
- Carcass
- Calving Ease

Predictions of Cow Productivity

- Days to Calving
- Calving Interval
- Cow Longevity – proportional hazard model
- Stayability - MAP
  - Snelling, et al., 1995
  - Brigham, et al., 2007
- Random Regression
  - Jamrozik, et al., 2014

New Stayability Model

- ERT with the most impact in maternal selection index
  - Relatively small changes can make a big difference
    - More calves to sell
    - Fewer replacement females – even more calves to sell
    - Heavier weaning weights of sale calves from older cows
    - Less calving difficulty

Random Regression Stayability

- The random (e.g., genetic) effects are described as a curve (polynomial) on age of cow at record
- Observations:
  - 0 – Cow did not have calf at a given age
  - 1 cow had a calf at a given age
  - Missing – unknown if the cow had a calf at a given age

Advantages of RR Method

- Uses observations in a more sensible way
  - Simultaneous solution to all ages
  - Censoring from culling is “missing” value
- Handles missing values in a more sensible
  - e.g., donor cow
- Easy to implement an animal model
- Easy to implement genomic information
  - Marker Effects Model
- Faster to obtain answers
  - PCG solver
  - Gibbs sampler

Our Marker Effects Stayability Model

- Year of birth
- Age at first calving
- Random contemporary group (intercept and slope)
- Random permanent environment due to the dam (intercept and slope)
- Genetic marker random effects (intercept and slope)
  - Genotyped
  - Non-genotyped