Genomic selection in Polish Holstein

Joanna Szyda

Wrocław University of Life Sciences

Andrzej Żarnecki

National Research Institute for Animal Production

Stanisław Kamiński

University of Warmia and Mazury

1. Material

- markers
- traits
- animals

2. Methods

- DGV estimation model
- GEBV
- reliability

3. Results

- GEBV
- candidate genes

4. Future

MARKERS

Illumina bovine snp50 beadchip

- 54 001 SNPs

46 267 tSNPs

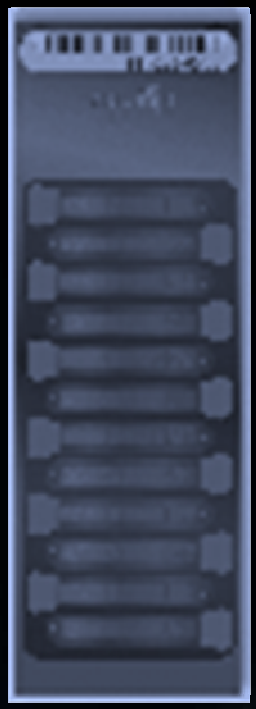
- MAF ≥ 0.01
- call rate ≥ 90%

average call rate

- 99.66% (all SNPs)
- 99.75% (tSNPs)

average MAF

- 0.23 (all SNPs)
- 0.26 (tSNPs)



All traits subjected to the national evaluation

production

- 3

udder health

- 1

conformation

- 21

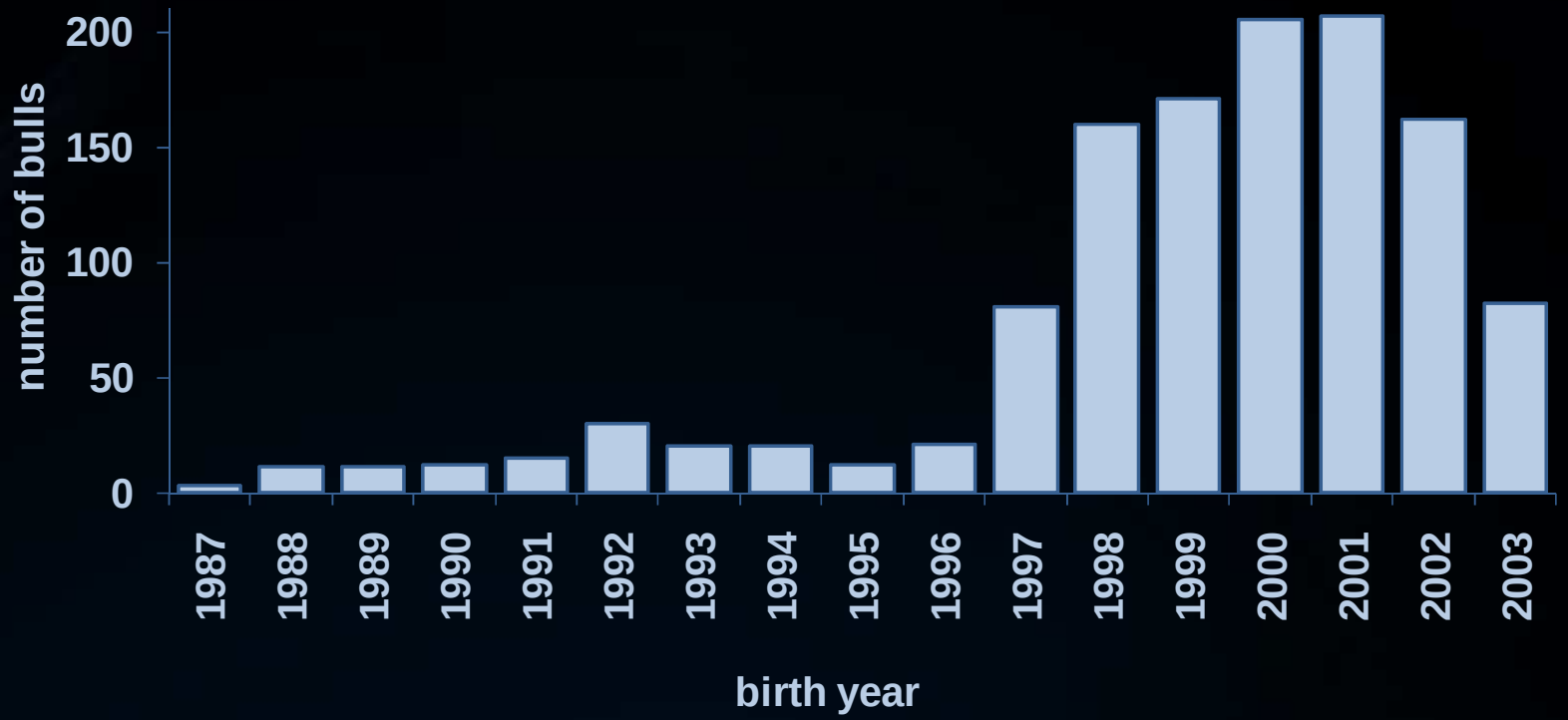
fertility

- 4

... except functional longevity

ANIMALS

ANIMALS



- 2008 training data set 1 227 bulls
- 2009 prediction data sets 55 + 80 bulls
- 2010 training data set 2 000 bulls

SNP effect estimation

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zq} + \mathbf{e}$$

- **y** deregressed EBV
- **μ** general mean
- **q** SNP
- **Z** ∈{ -1, 0, 1 }
- **e** residual

covariance structure

$$\mathbf{q} \sim N\left(0, I\frac{\hat{\sigma}_a^2}{46267}\right) \quad 46267 \times 46267$$

$$\begin{bmatrix} 1 & & "0" \\ & \cdots & \\ "0" & & 1 \end{bmatrix} \frac{\hat{\sigma}_a^2}{46267}$$

$$\mathbf{e} \sim N(0, D\sigma_e^2) \quad 1227 \times 1227$$

$$\begin{bmatrix} \frac{1}{EDC_1} & & "0" \\ & \cdots & \\ "0" & & \frac{1}{EDC_{1227}} \end{bmatrix} \sigma_e^2$$

DGV and GEBV

DGV → $\hat{Zq}$

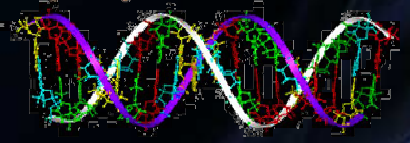
GEBV → $[r_{DGV} \quad r_{PI}] \begin{bmatrix} r_{DGV} & r_{DGV}r_{PI} \\ r_{PI} & \end{bmatrix}^{-1} \begin{bmatrix} DGV \\ PI \end{bmatrix}$

~ genomic relationship matrix Q

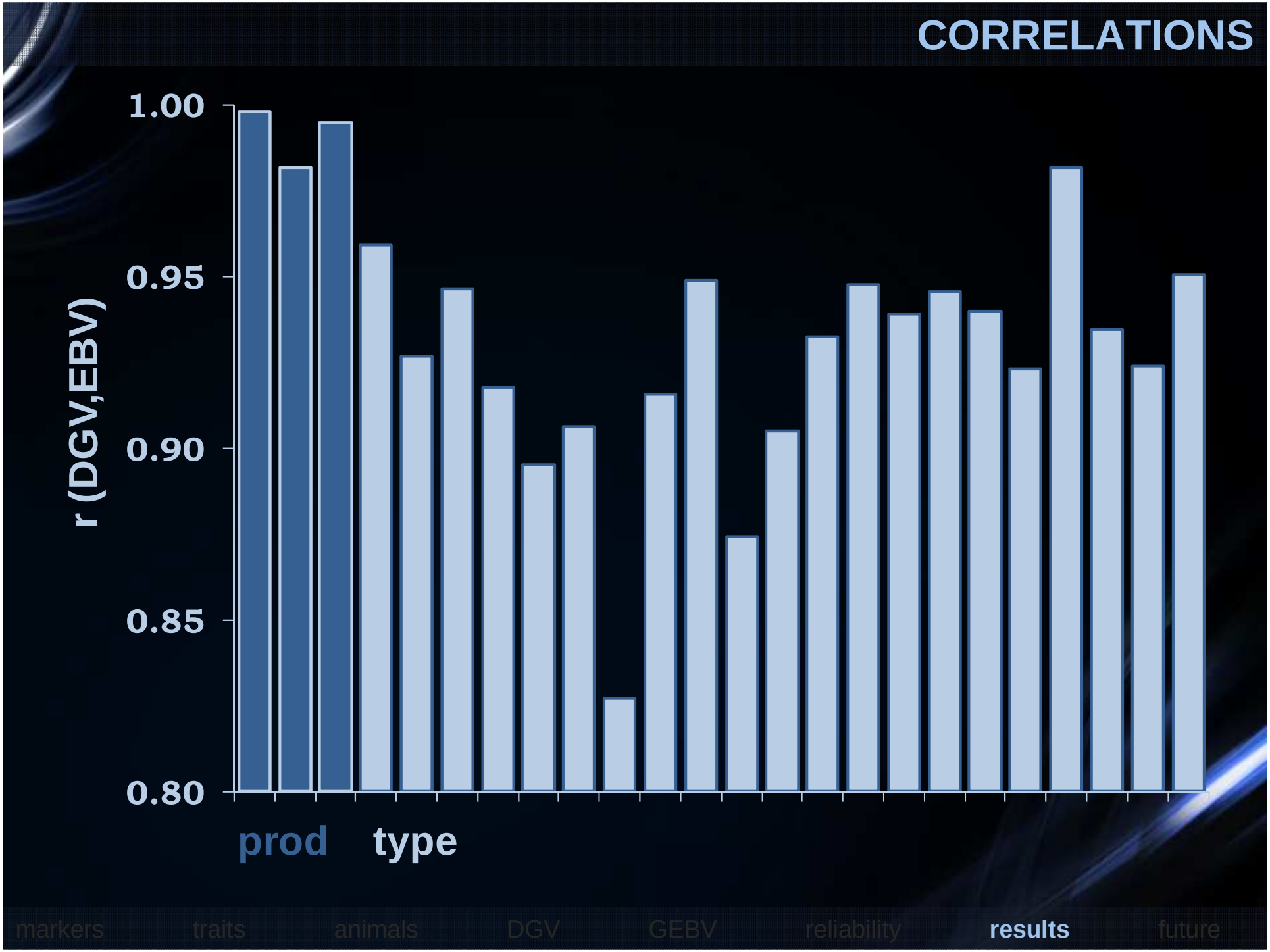
$DGV \sim N(0, Q\hat{\sigma}_a^2)$

ancestors (PI)

SNP genotypes (DGV)



CORRELATIONS



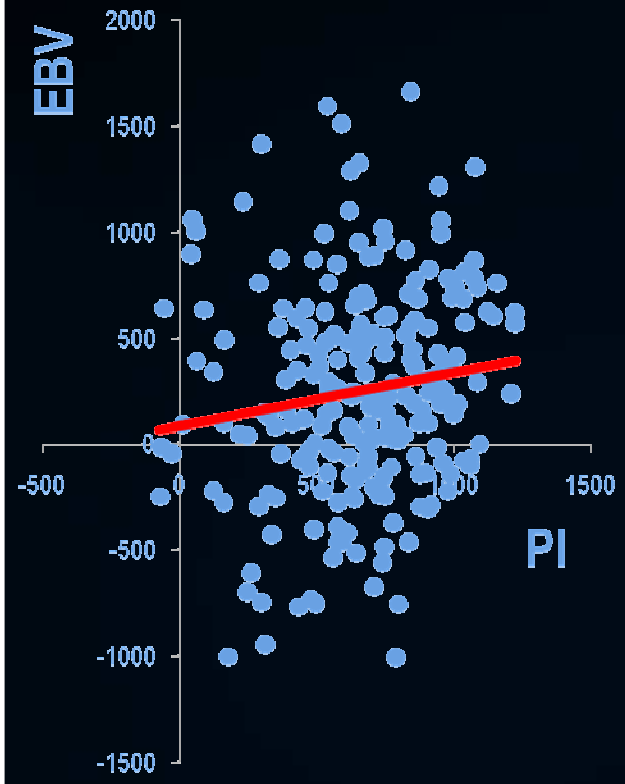
PREDICTIVE ABILITY

1227 genotyped bulls → validation for milk yield, May 2010

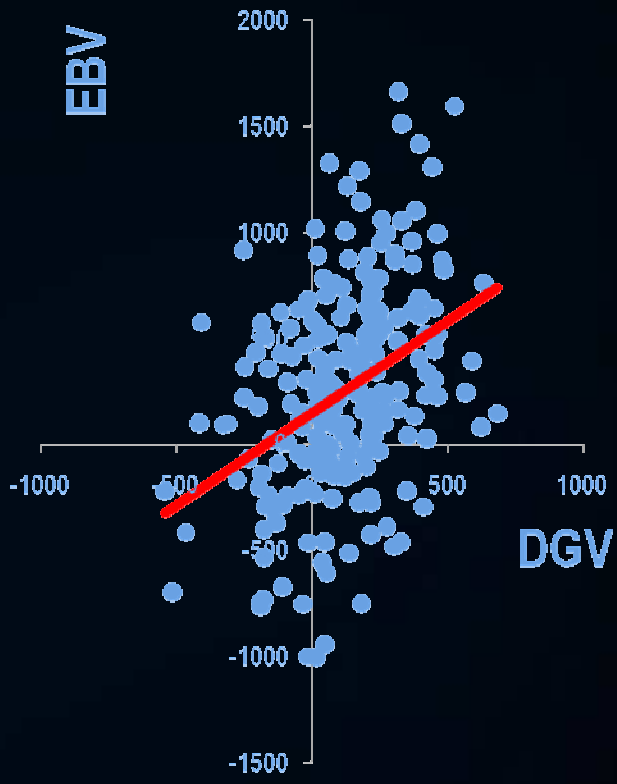
training: 1897-2001 → 996 bulls

prediction: >2001

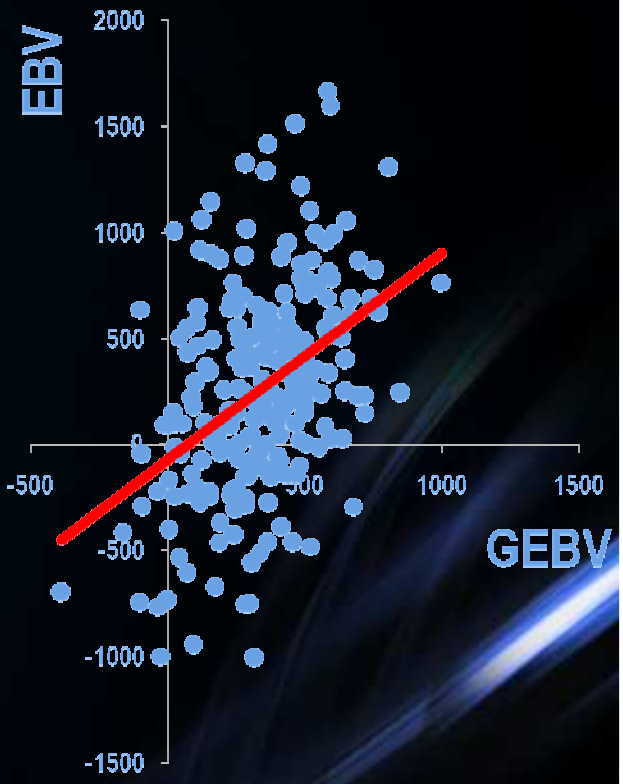
y=85+0.25x:



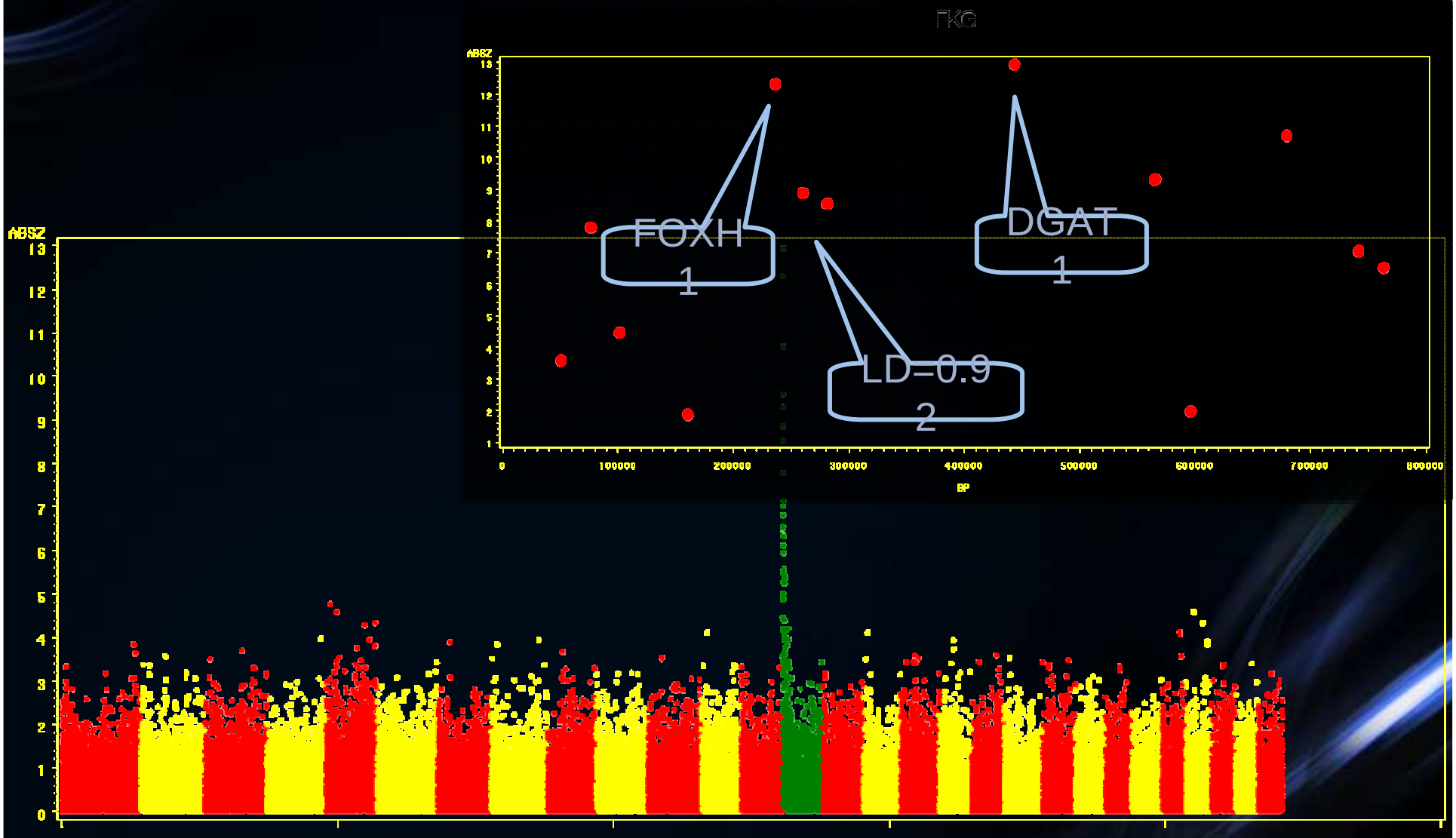
y=144+0.87x:



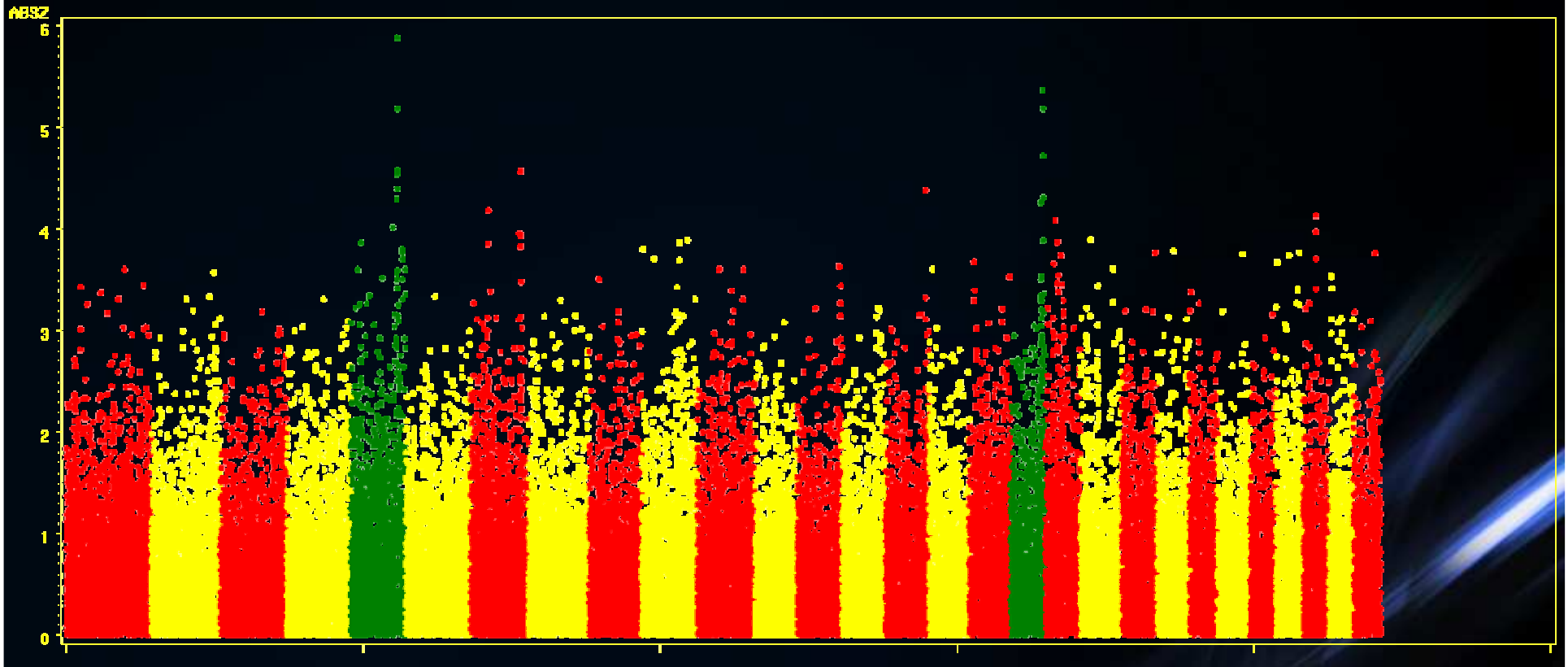
y=72+0.98x:



SNP effects → fat yield



SNP effects → stature



FUTURE PROJECTS

Dairy industry

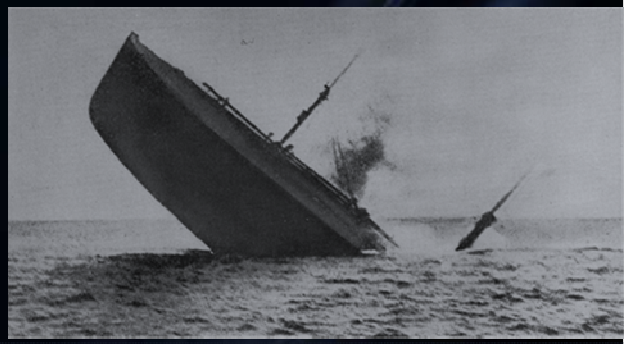
- small chip
- larger training data set
- MACE

Genetics

- Candidate gene hunting (SCS, semen quality)
- Interactions between genes

Methodology

- Covariance between SNPs



Thank you for your attention ...

ACCURACY of DGV

$$DGV \sim N(0, Q\hat{\sigma}_a^2) \qquad Q = Z_{ANIMxSNP} Z_{ANIMxSNP}^T \frac{1}{p_{het}^*}$$

$$y = X\beta + Zu + e$$

- y - deregressed EBV, b - fixed effects, u - DGV, e - error

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + Q^{-1}\hat{\sigma}_g^{-2} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

$$\text{Rel} = diag\left\{ Q\hat{\sigma}_a^2 - \left(Z^T R^{-1} Z + Q^{-1}\hat{\sigma}_a^{-2} \right)^{-} \right\} \frac{1}{\hat{\sigma}_a^2}$$

ACCURACY of DGV

$$y = X\beta + Zu + e$$

- y deregressed EBV
- b fixed effects
- u DGV
- e error

$$u \sim N(0, Q\hat{\sigma}_g^2) \qquad Q = Z_{SNP} Z_{SNP}^{T} \frac{1}{p_{het}^{*}}$$