Beyond GWAS

the application of SNP-chips in genomic selection

Joanna Szyda

Wrocław University of Life and Environmental Sciences, Institute of Genetics

The Biostatistics Group

Outline

Big data in livestock research

1. ~~GWAS~~
2. Cattle breeding industry → SNP
3. Research applications → SNP + WGS

Big data for breeding industry

SNP-chip for breeding industry

Routine genomic evaluation

- SNP effect estimation 3 times a year

  IV, VIII, XII

- Animals' additive genetic merit

Routine genomic evaluation

- Animals' additive genetic merit
  - sum of additive effects of all genes
  - breeding value
  - based on SNP information → direct genomic value
  - based on SNP and pedigree information → genomically enhanced breeding value
  - „old" bulls, „young" bulls, cows

Genomic evaluation – data

(pseudo)phenotypes

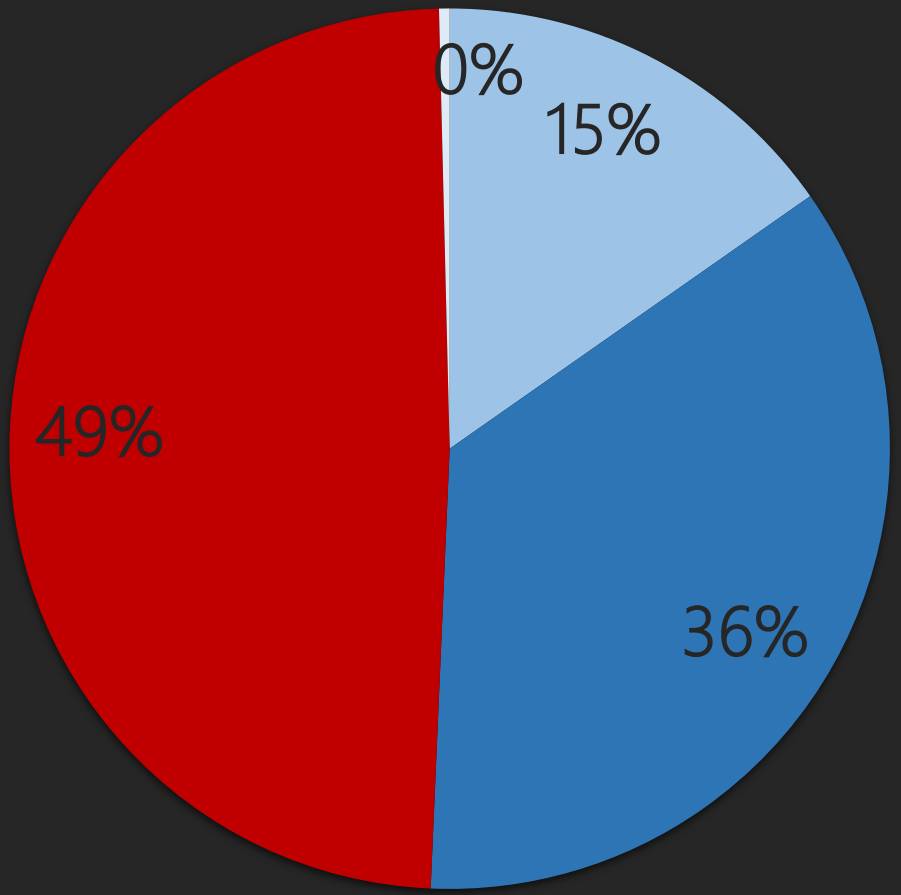
- cows→ phenotype→ bulls→ pseudophenotype

• Production	e.g. milk yield
• Udder health	e.g. SCS
• Fertility	e.g. cow conception rate
• Longevity	= functional herd life
• Type	e.g. udder depth

Genomic evaluation – data

SNP genotypes

- Euro Genomics 10K
- Illumina Bovine 50Kv1
- Illumina Bovine 50Kv2
- Other LD



Imputation

08.2017: 71 743 genotyped animals (53 492♂ / 18 251♀)

Genomic evaluation – data

Animals

Training data set „old" bulls	Test data set „young" bulls & cows
SNP genotypes	SNP genotypes
(pseudo)phenotypes	~~(pseudo)phenotypes~~

Genomic evaluation – training data set

SNP effect estimation

- Based on the training data set

$$\boldsymbol{y = 1\mu + Z_1 q + Z_2 a + e}$$

  - y deregressed breeding value
  - μ general mean
  - q SNP additive effect — q ~ N(0,G)
  - a additive polygenic effect — **a** ~ N(0,**A**)
  - **e** residual — **e** ~ N(0,**R**)
  - $Z_1 \in \{ -1, 0, 1 \}$ — dense
  - $Z_2 \in \{ 0, 1 \}$ — sparse

Genomic evaluation – SNP effect estimation

Covariance structure

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

q ~ N(0,G)

$$\begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & \dots \\ 0 & \dots & 1 \end{bmatrix} x\hat{\sigma}_a^2$$

a ~ N(0,A)

$$\begin{bmatrix} \frac{1}{edc_1} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \frac{1}{edc_n} \end{bmatrix} \hat{\sigma}_e^2$$

e ~ N(0,R)

Genomic evaluation – dimensions in 08.2017

SNP effect estimation

- Polish genomic evaluation run: 08.2017

$$\boldsymbol{y = 1\mu + Z_1 q + Z_2 a + e}$$

- y [1 × 30 548]
- μ [1 × 30 548]
- q [1 × 46 267]
- a [1 × 138 684]
- e [1 × 30 548]
- Z_1 [30 548 × 46 267]
- Z_2 [30 548 × 138 684]

Genomic evaluation – SNP effect estimation

Solutions

- Mixed model equations

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z_1 & X^T R^{-1} Z_2 \\ & Z_1^T R^{-1} Z_1 + G^{-1} & Z_1^T R^{-1} Z_2 \\ & & Z_2^T R^{-1} Z_2 + A^{-1} \end{bmatrix} \begin{bmatrix} \mu \\ q \\ a \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z_1^T R^{-1} y \\ Z_2^T R^{-1} y \end{bmatrix}$$

- No direct inverse possible

→ Iteration on data with residual update

(Legarra & Misztal 2009)

Genomic evaluation – results

Genomically Enhanced Breeding Value

- DGV of bull „i"

$$DGV_i = X_i \hat{b}_i + Z_{1i} \hat{q}_i$$

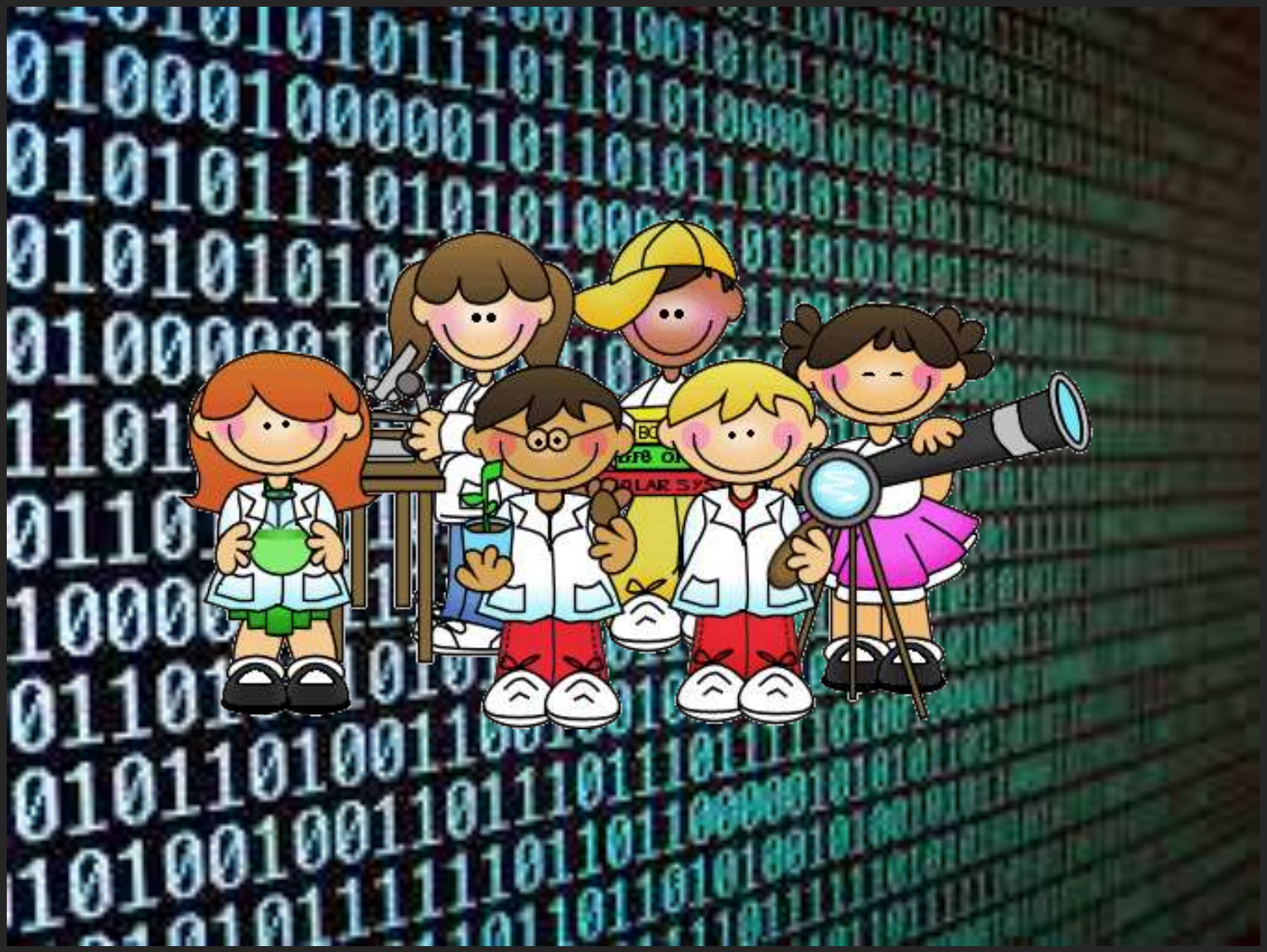
- GEBV for the training data set

$$GEBV = \begin{bmatrix} r_{DGV} & r_{EBV} \end{bmatrix} \begin{bmatrix} r_{DGV} & r_{DGV} r_{EBV} \\ r_{DGV} r_{EBV} & r_{EBV} \end{bmatrix}^{-1} \begin{bmatrix} DGV \\ EBV \end{bmatrix}$$

- GEBV for the test data set

$$GEBV = \begin{bmatrix} r_{DGV} & r_{PI} \end{bmatrix} \begin{bmatrix} r_{DGV} & r_{DGV} r_{PI} \\ r_{DGV} r_{PI} & r_{PI} \end{bmatrix}^{-1} \begin{bmatrix} DGV \\ PI \end{bmatrix}$$

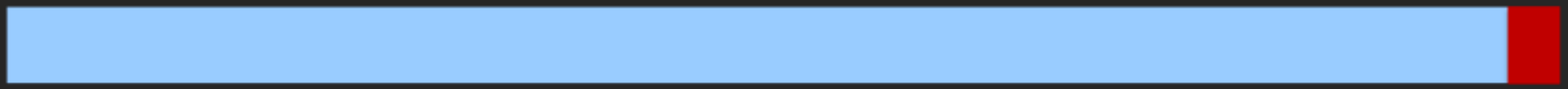
Big data for fun



WGS-SNPs validation by chip-SNPs

SNP calling → Illumina HiSeq2000

SNP calling → Illumina Bovine HD 777 000



correct SNPs (2 219 387)

wrong SNPs (74 764)

$$\text{logit}(\text{P}) = \sum_{i=1}^{5} QUAL^{i} + \sum_{i=1}^{5} DP^{i} + \sum_{i=1}^{5} GQ^{i} + SEQ_{D} + SEQ_{U}$$

… nevertheless GWAS for # of hoof disorders

3 650 cows → multiple records → 74 000 SNPs

GWAS step 1: 2 significant SNPs → BTA7 & BTA14

$$y = X_1\beta + X_2 SNP_a + Z_1 a + Z_2 p + Z_3 k + e$$

1000 Bull Genomes → region imputation → exon

GWAS step 2: → RGMB (BTA7) & STK3 (BTA14)

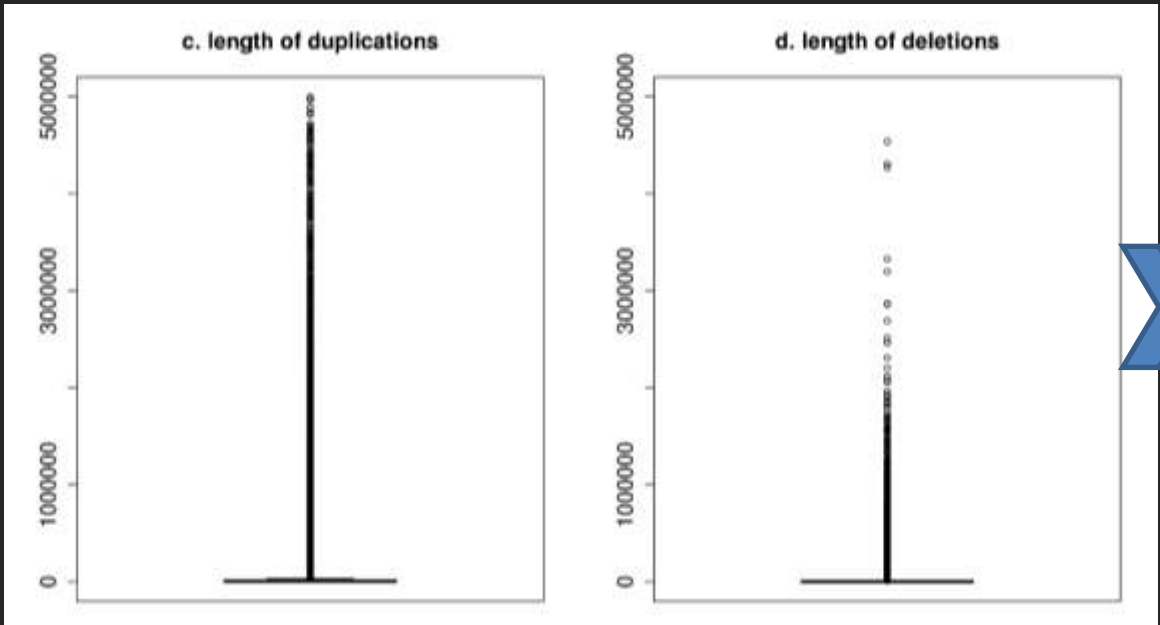
$$y = X_1\beta + \sum_{i=1}^{N} X_{i2} SNP_{ia} + \sum_{i=1}^{N} X_{i3} SNP_{id} + \sum_{i=1}^{N}\sum_{j>i}^{N} X_{ij4} SNP_{ije} + Za + e$$

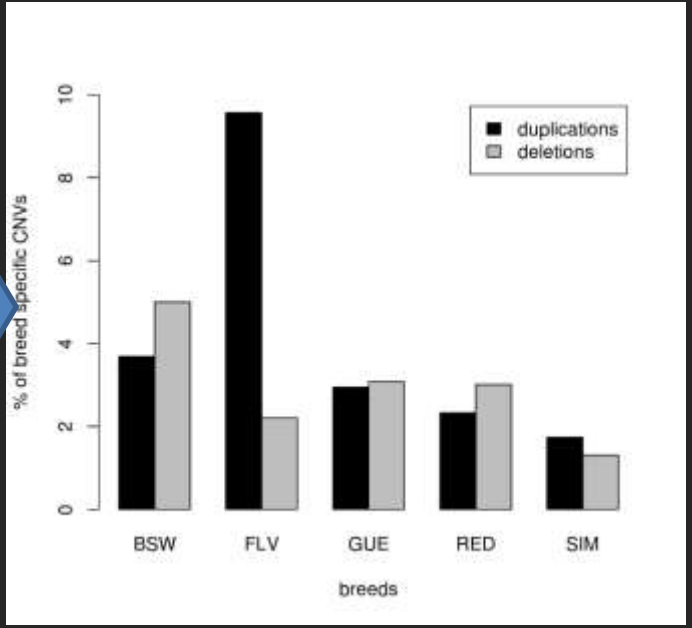
… pure WGS → Inter-breed genetic differences

Whole genome DNA sequence:

152 bulls → Brown Swiss, Fleckvieh, Guernsey, Simmental, Norwegian Red

Copy Number Variation → deletions / duplications





- Magdalena Frąszczak
- Magda Mielczarek
- Tomasz Suchocki
- Andrzej Żarnecki
- Kacper Żukowski
- Bernt Guldbrandsen

- Genomika Polska
- 1000 Bull Genomes Consortium
- Zuchtdata AG 

... Thank you for attention