blupf90+ MME solver and ssGWAS

Zuleica Trujano

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences
Animal Breeding and Genetics Group

blupf90+

MME solver

blupf90+

- `blupf90`: MME solver
- `airemlf90`: variance components using Average Information REML
- `remlf90`: variance components using Expectation Maximization REML

Mixed Model Equations Solver

Variance Components Estimation

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W}+\mathbf{A}^{-1}\otimes\mathbf{G}_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

blupf90+

MME Solver

Default

VC Estimation

- AI-REML:

```
OPTION method VCE
```

- EM-REML:

```
OPTION method VCE
OPTION EM-REML xx
```

xx: # of EM rounds

xx > 0 : switch to aireml
xx < 0 : does not switch if convergence is reached

blupf90+

- Supports virtually any model used in AB&G:
  - animal model
  - models with maternal effect
  - MPE
  - PE
  - Random Regression
  - Social interaction
  - Multiple traits
    - up to 70 if no correlated effects
    - up to [70/number of correlated effects]

blupf90+

- Computes generalized solutions by several methods:
  - Preconditioner Conjugate Gradient (PCG)
    - Default Iterative method (fast)
  - Successive over-relaxation (SOR)
    - an iterative method based on Gauss-Seidel
  - Direct solution using sparse Cholesky factorization
    - FSPAK or YAMS (greater memory requirements)

- Solutions change among methods, but estimable functions should be the same

- Prediction error variances can be obtained using sparse inverse (FSPAK or YAMS)

blupf90+ with PCG

Animal Breeding and Genetics Local Wiki

Iteration on data with preconditioned conjugate gradient (PCG)

Algorithm

Table of Contents –

Preconditioned conjugate gradient (PCG) is an iterative method to solve the linear equations. This method is easily harmonized with the iteration of data technique. Intermediate status is kept in only 4 vectors and the one iteration will be done updating the vectors. BLUP90IOD2 is a program implementing the algorithms. Here we will introduce a basic idea needed to understand what the program does. See Stranden and Lidauer (2000) and Tsuruta et al. (2001) for detailed algorithm.

The mixed model equations can be written as

$$\mathbf{Cx} = \mathbf{b}$$

where $\mathbf{C}$ is the left-hand side matrix, $\mathbf{x}$ is the solution vector and $\mathbf{b}$ is the right-hand side vector. If we have a matrix $\mathbf{M}$ which is an approximation of $\mathbf{C}$, above equations are equivalent to

$$\mathbf{M}^{-1}\mathbf{Cx} = \mathbf{M}^{-1}\mathbf{b}.$$

This matrix $\mathbf{M}$ is called preconditioner. If $\mathbf{M} = \mathbf{C}$, the equations are immediately solved. BLUPF90 uses $\mathbf{M} = \mathrm{diag}(\mathbf{C})$ so its inverse is easily calculated.

The residual is expressed as

$$\mathbf{r} = \mathbf{b} - \mathbf{Cx}$$

and the algorithm tries to reduce with a statistics containing the residual. The convergence criterion is

$$\varepsilon = \frac{||\mathbf{b} - \mathbf{Cx}||^2}{||\mathbf{b}||^2}$$

where $||\cdot||$ means the norm.

If $\mathbf{M}^{-1}\mathbf{C}$ has a better condition than **C**, the convergence is reached is faster

Parameter file for blupf90+

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4      40593 cross
  5  5          2 cross
  6  0          4 cross
  7  0          8 cross
  8  8     918111 cross
RANDOM_RESIDUAL VALUES
   2.5300        1.3425
   1.3425        29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600        2.2391
   2.2391        30.609
```

Unlimited number of traits and effects

Parameter file for blupf90+

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3   4      40593 cross
  5   5          2 cross
  6   0          4 cross
  7   0          8 cross
  8   8     918111 cross
RANDOM_RESIDUAL VALUES
   2.5300        1.3425
   1.3425        29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600        2.2391
   2.2391        30.609
```

As many columns as the number of traits

Number of levels

Type of effect

- As many rows as the NUMBER_OF_EFFECTS
- Model definition for each trait
- Different models per trait are supported
- If an effect is missing for one trait use 0

Parameter file for blupf90+

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4      40593 cross
  5  5          2 cross
  6  0          4 cross
  7  0          8 cross
  8  8     918111 cross
RANDOM_RESIDUAL VALUES
   2.5300        1.3425
   1.3425        29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600        2.2391
   2.2391        30.609
```

Should be a square matrix with dimension equal to the number of traits

- Use zero (0.0) to indicate uncorrelated residual effects between traits
- e.g. For a 3-trait model

  43.1 0.0 0.0
  0.0 5.1 3.2
  0.0 3.2 10.3

Parameter file for blupf90+

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 ../renf90.dat
NUMBER_OF_TRAITS
           2
NUMBER_OF_EFFECTS
           5
OBSERVATION(S)
    1    2
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
  3  4      40593 cross
  5  5          2 cross
  6  0          4 cross
  7  0          8 cross
  8  8     918111 cross
RANDOM_RESIDUAL VALUES
   2.5300        1.3425
   1.3425        29.714
RANDOM_GROUP
     5
RANDOM_TYPE
 add_an_upginb
FILE
../renadd05.ped
(CO)VARIANCES
   0.7600        2.2391
   2.2391        30.609
```

Definition of random effects

RANDOM_GROUP
RANDOM_TYPE
FILE
(CO)VARIANCES

Definition of random effects

- RANDOM_GROUP
  - Number of the effect(s) from list of effects
  - Correlated effects should be consecutive e.g. Maternal effects, Random Regression
- RANDOM_TYPE
  - diagonal, add_animal, add_sire, add_an_upg, add_an_upginb, add_an_self, user_file, user_file_i, or par_domin
- FILE
  - Pedigree file, parental dominance, or user file
- (CO)VARIANCES
  - Square matrix with dimension equal to the number_of_traits*number_of_correlated_effects

(CO)VARIANCES

- Assuming a 3 trait (T1-T3) and 2 correlated effects (E1-E2)

		E1			E2		
		T1	T2	T3	T1	T2	T3
E1	T1						
	T2						
	T3						
E2	T1						
	T2						
	T3						

RANDOM_TYPE

- *Diagonal*
  - for permanent environment effects
  - assumes no correlation between levels of the effect

- *add_sire*
  - To create a relationship matrix using sire and maternal grandsire
  - Pedigre file:
    - `individual number, sire number, maternal grandsire number`

- *add_animal*
  - To create a relationship matrix using sire and dam information
  - Pedigre file:
    - `animal number, sire number, dam number`

RANDOM_TYPE

- *add_an_upg*
  - As before but using rules for unknown parent group
  - Pedigre file:
    - `animal number, sire number, dam number, parent code`
    - missing sire/dam can be replaced by upg number, usually greater than maximum number of animals
    - Parent code = 3 – # of known parents
      - 1 both parents known
      - 2 one parent known
      - 3 both parents unknown

- *add_an_upginb*
  - As before but using rules for unknown parent group and inbreeding
  - Pedigre file:
    - `animal number, sire number, dam number, inb/upg code`
    - missing sire/dam can be replaced by upg number, usually greater than maximum number of animals
    - inb/upg code = 4000 / [(1+ms)(1-Fs) + (1+md)(1-Fd)]
    - ms (md) is 0 if sire (dam) is known and 1 otherwise
    - Fs(Fd) inbreeding coefficient of the sire (dam)

RANDOM_TYPE

- *Add_an_self*
  - *To create a relationship matrix when there is selfing*
  - Pedigre file:
    - `individual number, parent 1 number, parent 2, number of selfing generations`

- *user_file*
  - An inverted matrix is read from file
  - Matrix is stored only upper- or lower-triangular
  - Matrix file:
    - `row, col, value`

- *user_file_i*
  - As before but the matrix will be inverted by the program

- *par_domin*
  - A parental dominance file created by program RENDOM

OPTIONS for blupf90+

- Program behavior can be modified by adding extra options at the end of the par file
- `OPTION option_name x1 x2 …`
- option_name: each program has its definition of options
- The number of optional parameters (x1, x2, …) to control the behavior depends on the option

Options for blupf90+

Options

```
OPTION conv_crit 1e-12
```

Set convergence criteria (deault 1e-12).

```
OPTION maxrounds 10000
```

Set maximum number of rounds (default 5000).

```
OPTION solv_method FSPAK
```

Selection solutions by FSPAK, SOR or PCG (default PCG).

```
OPTION r_factor 1.6
```

Set relaxation factor for SOR (default 1.4).

```
OPTION sol se
```

Store solutions and standard errors.

```
OPTION store_pev_pec 6
```

Store triangular matrices of standard errors and its covariances for correlated random effects such as direct-maternal effects and random-regression effects in “pev_pec_bf90”.

Options for blupf90+

Missing data
Not pedigree!

```
OPTION missing -999
```

Specify missing observations (default 0) in integer.

```
OPTION residual
```

y-hat and residual will be included in “yhat_residual”.

```
OPTION blksize 3
```

Set block size for preconditioner (default 1).

```
OPTION use_yams
```

Run the program with YAMS (modified FSPAK).

```
OPTION SNP_file snp
```

Specify the SNP file name to use genotype data.

New options for blupf90+

- Storing reliabilities based on PEV

  `OPTION store_accuracy X`

  Number of animal effect

$$Rel = 1 - \frac{PEV}{\sigma_u^2(1+f)}$$

  - Adjusts for f (inbreeding) from **A**, **G**, or **H**
  - Turn inbreeding adjustment off
  - `OPTION correct_accuracy_by_inbreeding_direct 0`

- Storing solutions with original ID if renumf90 was used to renumber the data

  `OPTION origID`

  - Only *solutions.original* is created

New options for blupf90+

- Storing reliabilities with original ID

```
OPTION store_accuracy X orig
```

Number of animal effect

$$Rel = 1 - \frac{PEV}{\sigma_u^2(1+f)}$$

- Storing solutions with original ID if renumf90 was used to renumber the data
  - The option will save acc_bf90 with renumbered and original ID
  - If want to have solutions with original ID as well, combine with OPTION origID

Common parameter file for blupf90+

```
# BLUPF90 parameter file created by RENUMF90
DATAFILE
 renf90.dat
NUMBER_OF_TRAITS
           1
NUMBER_OF_EFFECTS
           2
OBSERVATION(S)
    1
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]
 2          2 cross
 3      12010 cross
RANDOM_RESIDUAL VALUES
  0.60000
 RANDOM_GROUP
     2
 RANDOM_TYPE
 add_an_upginb
 FILE
renadd02.ped
(CO)VARIANCES
  0.40000
OPTION SNP_file genotypes.txt
OPTION map_file gen_map.txt
```

ssGBLUP SNP effects and ssGWAS

Equivalence between GBLUP and SNP-BLUP

GBLUP

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W}+\mathbf{G}^{-1}\boldsymbol{\lambda_1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

↓

GEBV

$\mathrm{Var}(\mathbf{u}) = ?$

$\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$

SNP-BLUP (Ridge Regression)

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{I}\boldsymbol{\lambda_2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

↓

SNP effects

$\mathbf{u} = \mathbf{Za}$

$\mathrm{Var}(\mathbf{u}) = ?$

$\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$

Are GBLUP and SNP-BLUP equivalent?

- Assumption of GBLUP: $\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$
- In SNP-BLUP: $\mathbf{u} = \mathbf{Za}$

$$\mathbf{u} = \mathbf{Za}$$

$$\mathrm{Var}(\mathbf{u}) = \mathrm{Var}(\mathbf{Za})$$

$$\mathrm{Var}(\mathbf{u}) = \mathbf{Z}\,\mathrm{Var}(\mathbf{a})\,\mathbf{Z}'$$

$$\mathrm{Var}(\mathbf{u}) = \mathbf{ZZ}'\sigma_a^2$$

$$\sigma_a^2 = \frac{\sigma_u^2}{2\sum_{i=1}^{SNP} p_i(1-p_i)}$$

$$\mathrm{Var}(\mathbf{u}) = \mathbf{ZZ}'\frac{\sigma_u^2}{2\sum_{i=1}^{SNP} p_i(1-p_i)}$$

$$\mathrm{Var}(\mathbf{u}) = \frac{\mathbf{ZZ}'}{2\sum_{i=1}^{SNP} p_i(1-p_i)}\sigma_u^2$$

Genomic relationship matrix VanRaden (2008)

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2\sum_{i=1}^{SNP} p_i(1-p_i)}$$

$$\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$$

GBLUP assumption!!!

ssGBLUP and ssSNP-BLUP are also equivalent!

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W}+\mathbf{H}^{-1}\frac{\sigma_e^2}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

ssGBLUP

Misztal et al. (2009)
Legarra et al. (2009)
Aguilar et al. (2010)
Christensen & Lund (2010)

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\mathbf{M} & \mathbf{X}_n'\mathbf{Z}_n \\ \mathbf{M}'\mathbf{Z}'\mathbf{X} & \mathbf{M}'\mathbf{Z}'\mathbf{Z}\mathbf{M}+\mathbf{I}\frac{\sigma_e^2}{\sigma_\alpha^2} & \mathbf{M}_n'\mathbf{Z}_n'\mathbf{Z}_n \\ \mathbf{Z}_n'\mathbf{X}_n & \mathbf{Z}_n'\mathbf{Z}_n\mathbf{M}_n & \mathbf{Z}_n'\mathbf{Z}_n+\mathbf{A}^{nn}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\epsilon}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{M}'\mathbf{Z}'\mathbf{y} \\ \mathbf{Z}_n'\mathbf{y}_n \end{bmatrix}$$

ssSNPBLUP or **ssBR**

Fernando et al. (2014)
Liu et al. (2014)
Mantysaari & Stranden (2016)

J. Dairy Sci. 101:10082–10088
https://doi.org/10.3168/jds.2018-14913

Short communication: **Genomic prediction using different single-step methods in the Finnish red dairy cattle population**

H. Gao,*†[1] M. Koivula,‡ J. Jensen,* I. Strandén,‡ P. Madsen,* T. Pitkänen,‡ G. P. Aamand,† and E. A. Mäntysaari‡
*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark
†Nordic Cattle Genetic Evaluation, DK-8200 Aarhus, Denmark
‡Natural Resources Institute Finland (Luke), FIN-31600 Jokioinen, Finland

We confirmed that regular ssGBLUP and ssBR with an extra polygenic effect led to the same predictions.

SNP effects in ssGBLUP

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W}+\mathbf{H}^{-1}\boldsymbol{\lambda}_{\mathbf{1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix}$$

$$\hat{\mathbf{a}} = \alpha b \frac{1}{2\sum p_i(1-p_i)} \mathbf{Z'G}^{-1}\hat{\mathbf{u}}$$

Matrix of SNP content

Genomic relationship matrix

α = blending parameter for $\mathbf{G}$

$$b = 1 - \frac{\lambda}{2}$$

$$\lambda = \frac{1}{n^2}\left(\sum_i \sum_j \mathbf{A}_{22_{ij}} - \sum_i \sum_j \mathbf{G}_{ij}\right)$$

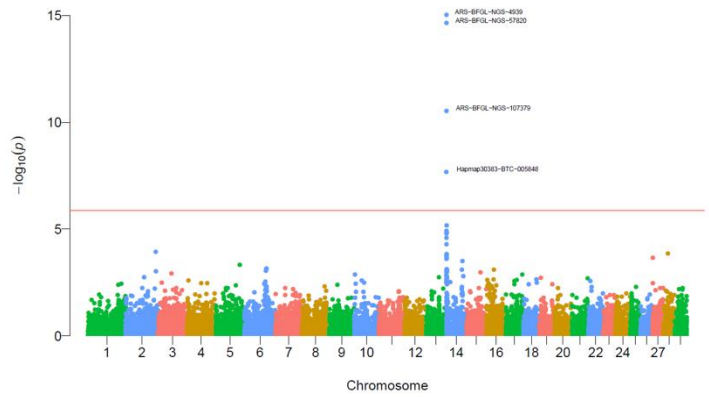
What can we do with SNP effects?

1) Predictions for animals not included in the evaluation

   Indirect Predictions

   Indirect Genomic Predictions

2) Genome-Wide Association Studies (GWAS)



1) Indirect Predictions

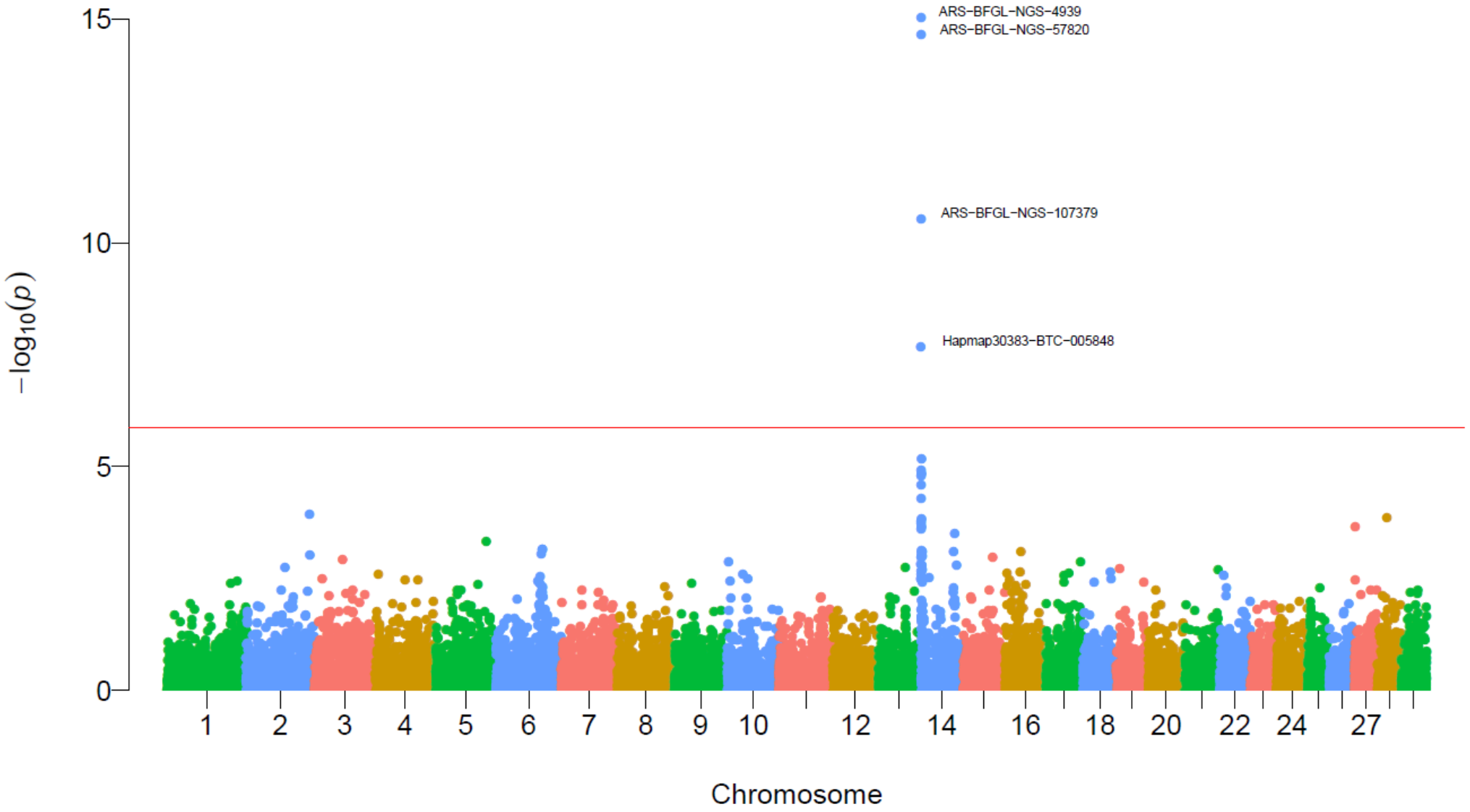
- Interim evaluations
  - Between official runs
- Not all genotyped animals are in the evaluations
  - Animals with incomplete pedigree increase bias and lower R^2
- Commercial products
  - e.g., GeneMax -> genomic testing for non-registered animals

1) Indirect Predictions

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'W} \\ \mathbf{W'X} & \mathbf{W'W+H^{-1}\lambda_1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{bmatrix} \quad \Longrightarrow \quad \hat{\mathbf{a}} = \alpha b \frac{1}{2\sum p_i(1-p_i)} \mathbf{Z'G^{-1}\hat{u}}$$

Indirect Prediction: $\mathbf{IP} = \boldsymbol{u}_{\boldsymbol{m}}^{*} = \mathbf{Z\hat{a}}$

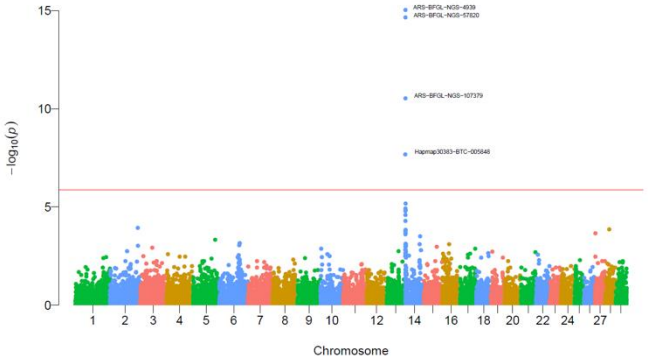
2) Genome-wide Association Studies



Current standard for GWAS

- Single marker regression with **G** to compensate for relationships
  - $\mathbf{y} = \mathbf{Xb} + \mathbf{z}_i\mathbf{a}_i + \mathbf{u} + \mathbf{e}$
    - **z:** gene content {0,1,2}
    - **a:** SNP effect

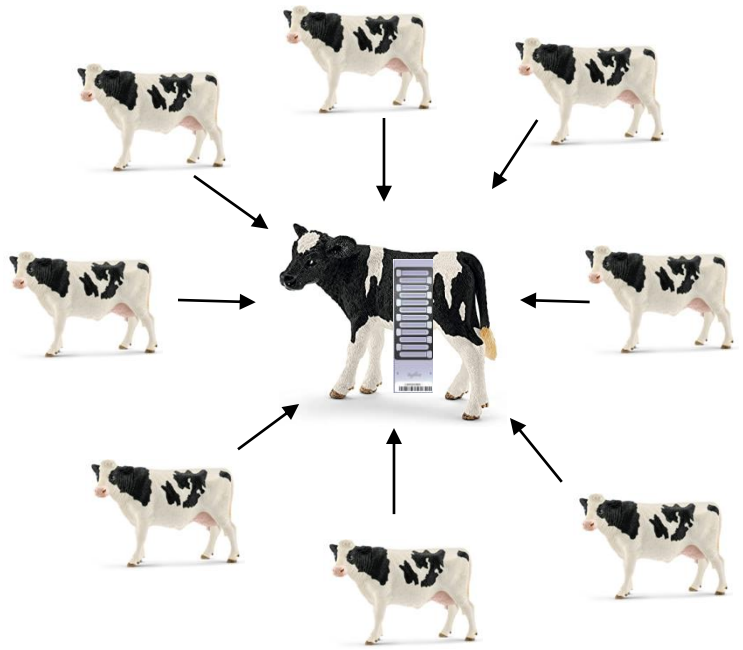
- Estimate SNP effects
- Get p-values as $pval_i = 2\left(1 - \Phi\left(\left|\frac{\hat{a}_i}{sd(\hat{a}_i)}\right|\right)\right)$
- Apply Bonferroni to correct for multiple testing



- **Assumption: Genotyped individuals have phenotypes**

GWAS in livestock populations

- Most animals are non-genotyped
- Animals may not have phenotypes
- Some traits are sex-limited
  - milk, fat, protein



- Single marker regression
  - Only genotyped animals with phenotypes
  - Deregressed EBV

- Need a method that fits the livestock data
  - ssGWAS

Single-step GWAS (historical)

SNP effects

GEBVs

$$\hat{\mathbf{a}} = \alpha b \frac{1}{2\sum p_i(1-p_i)} \mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

VanRaden 2008
Stranden and Garrick 2009
Wang et al. 2012

a) Quadratic SNP variance (Falconer & Mackay, 1996)

$$d_i = \hat{a}_i^2 2p_i(1-p_i)$$

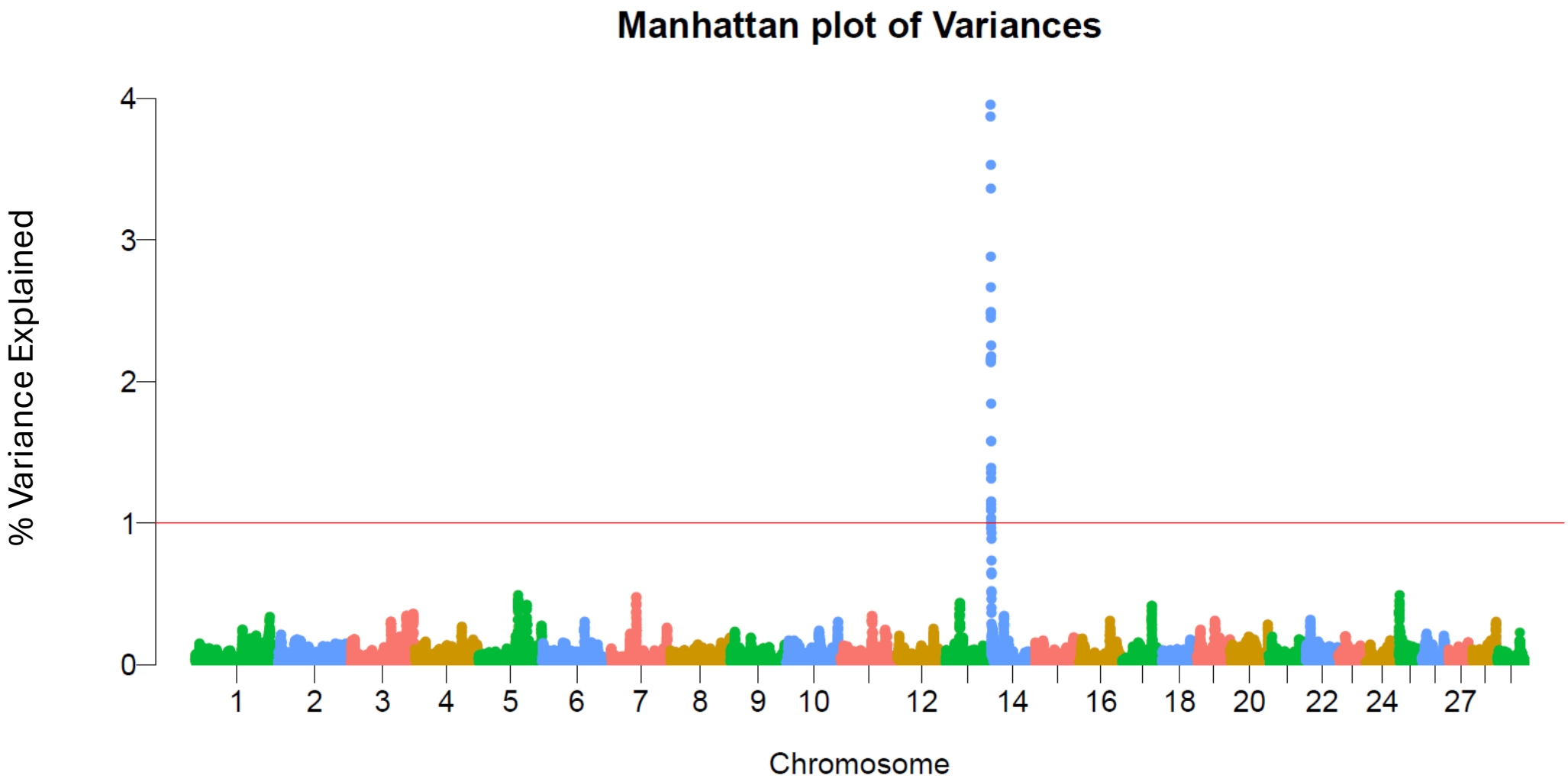
b) NonlinearA SNP variance (VanRaden, 2008)

$$d_i = 1.125^{\frac{|\hat{a}_i|}{sd(\hat{a})}-2}$$

Single-step GWAS (historical)

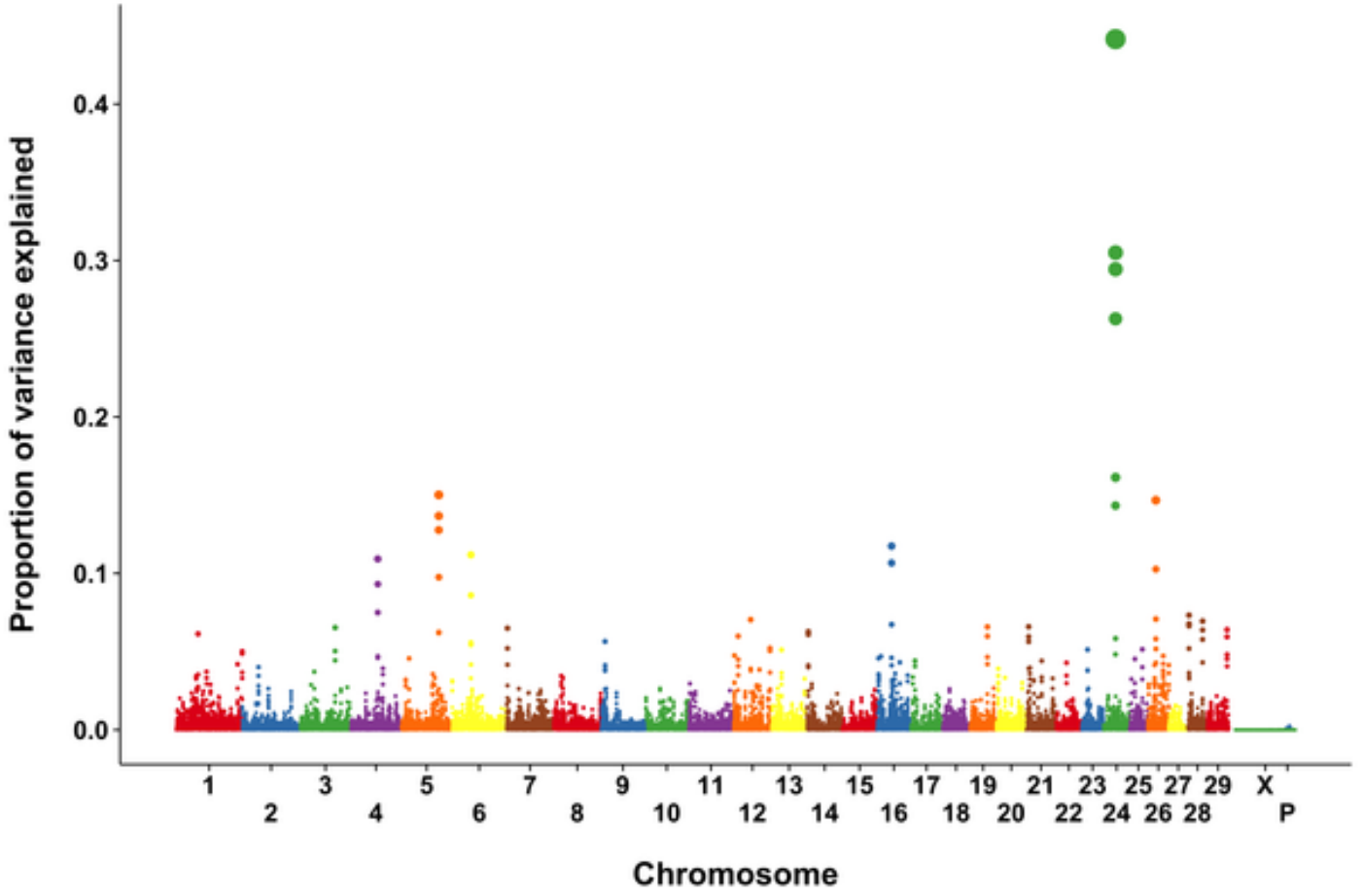
Fat – US Holsteins

No P-value!!!



Single-step GWAS (historical)

Figure 2. Proportion of SNP variance explained by 5-SNP moving windows for rectal temperature from a single-step GBLUP analysis



No P-value!!!

PLOS ONE

Dikmen S, Cole JB, Null DJ, Hansen PJ (2013) Genome-Wide Association Mapping for Identification of Quantitative Trait Loci for Rectal Temperature during Heat Stress in Holstein Cattle. PLOS ONE 8(7): e69202. https://doi.org/10.1371/journal.pone.0069202
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069202

p-values in ssGWAS

1) Factorize and Invert LHS of ssGBLUP with YAMS (Masuda et al., 2014)

blupf90+

2) Solve the MME for $\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}\end{bmatrix}$ using the sparse Cholesky factor

3) Extract coefficients for genotyped animals ($\mathbf{C}^{u_2u_2}$) from LHS^{-1}

4) Obtain individual prediction error variance of SNP effects:

$$Var(\hat{a}_i) = \alpha b \frac{1}{2\sum p_i(1-p_i)} \mathbf{z}_i' \mathbf{G}^{-1}(\mathbf{G}\sigma_u^2 - \mathbf{C}^{u_2u_2})\mathbf{G}^{-1}\mathbf{z}_i \frac{1}{2\sum p_i(1-p_i)} \alpha b$$

postGSf90

(Gualdron-Duarte et al., 2014)

5) Backsolve GEBV to SNP effects ($\hat{a}$): $\hat{a} = \alpha b \frac{1}{2\sum p_i q_i} \mathbf{Z}'\mathbf{G}^{-1}\hat{u}$

6) p-value$_i$ $= 2\left(1 - \Phi\left(\left|\frac{\hat{a}_i}{sd(\hat{a}_i)}\right|\right)\right)$

Φ is the cumulative standard normal function

How to run ssGWAS with p-values in BLUPF90

- blupf90+ to estimate GEBV
  - `OPTION SNP_file snp.dat_clean`
  - `OPTION map_file mrkmap.txt_clean`
  - `OPTION saveGInverse`
  - `OPTION saveA22Inverse`
  - `OPTION snp_p_value`
  - `OPTION no_quality_control`

- postGSf90 to backsolve GEBV to SNP effect and compute p-values
  - `OPTION SNP_file snp.dat_clean`
  - `OPTION map_file mrkmap.txt_clean`
  - `OPTION readGInverse`
  - `OPTION readA22Inverse`
  - `OPTION snp_p_value`
  - `OPTION no_quality_control`

Output from postGSf90

`chrsnp_pval`

contains data to create plot by GNUPLOT

- 1: trait
- 2: effect
- 3: -log10(p-value)
- 4: SNP
- 5: Chromosome
- 6: Position in bp

`chrsnp`

contains data to create plot by GNUPLOT

- 1: trait
- 2: effect
- 3: values of SNP effects to use in Manhattan plots → [abs(SNP_i)/var(SNP)]
- 4: SNP
- 5: Chromosome
- 6: Position

Pft1e2.gnuplot

Pft1e2.R

Sft1e2.gnuplot

Sft1e2.R

Output from postGSf90

```
chrsnpvar
```

contains data to create plot by GNUPLOT

- 1: trait
- 2: effect
- 3: variance explained by *n* adjacents SNP
- 4: SNP
- 5: Chromosome
- 6: Position

Vft1e2.gnuplot

Vft1e2.R

Output from postGSf90

```
snp_sol
```

contains solutions of SNP and weights

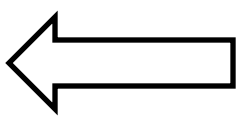
- 1: trait
- 2: effect
- 3: SNP
- 4: Chromosome
- 5: Position
- 6: SNP solution
- 7: weight

if `OPTION windows_variance` is used

- 8: variance explained by *n* adjacents SNP.

if `OPTION snp_p_value` is used

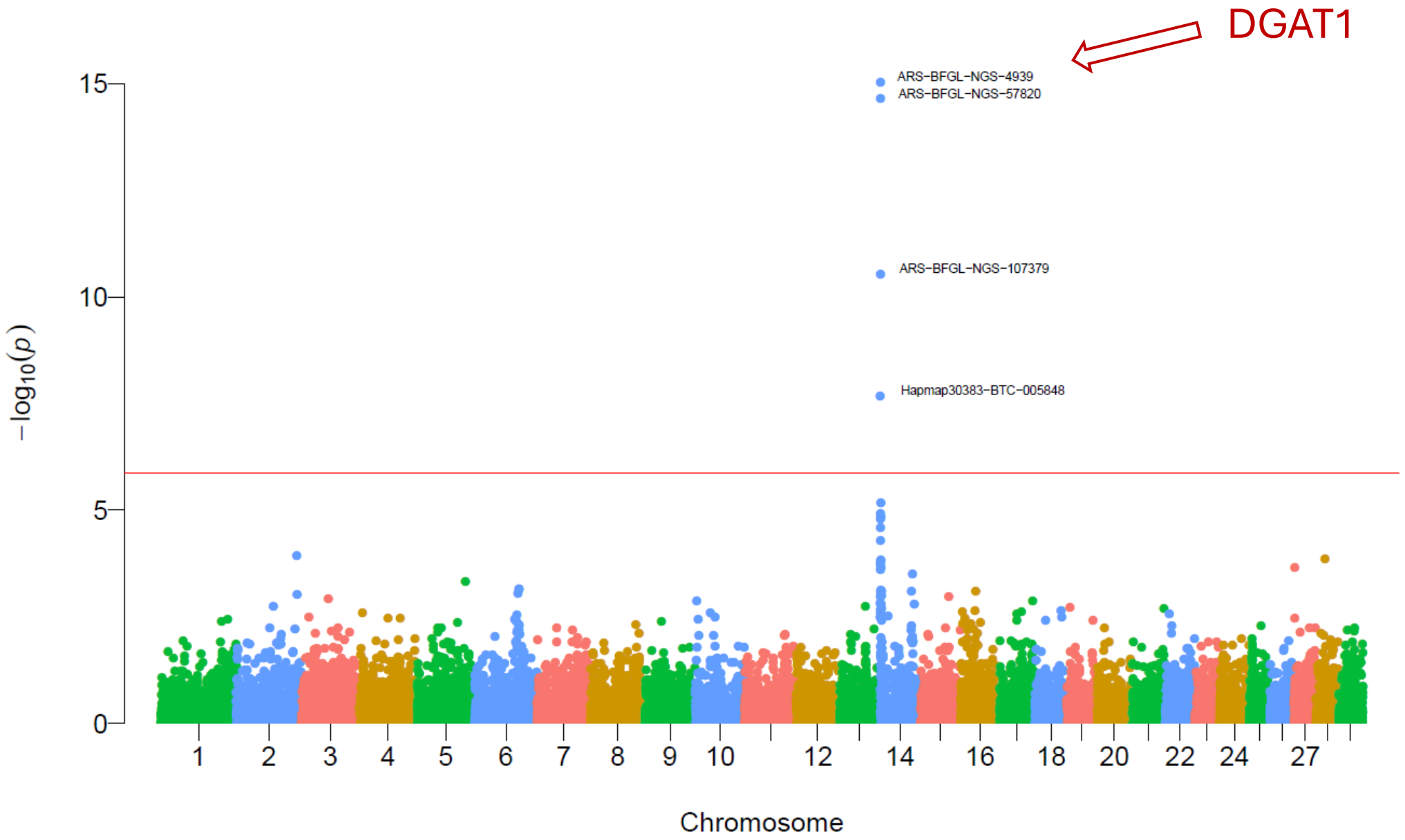
- 9: variance of the SNP solution (used to compute the p-value)



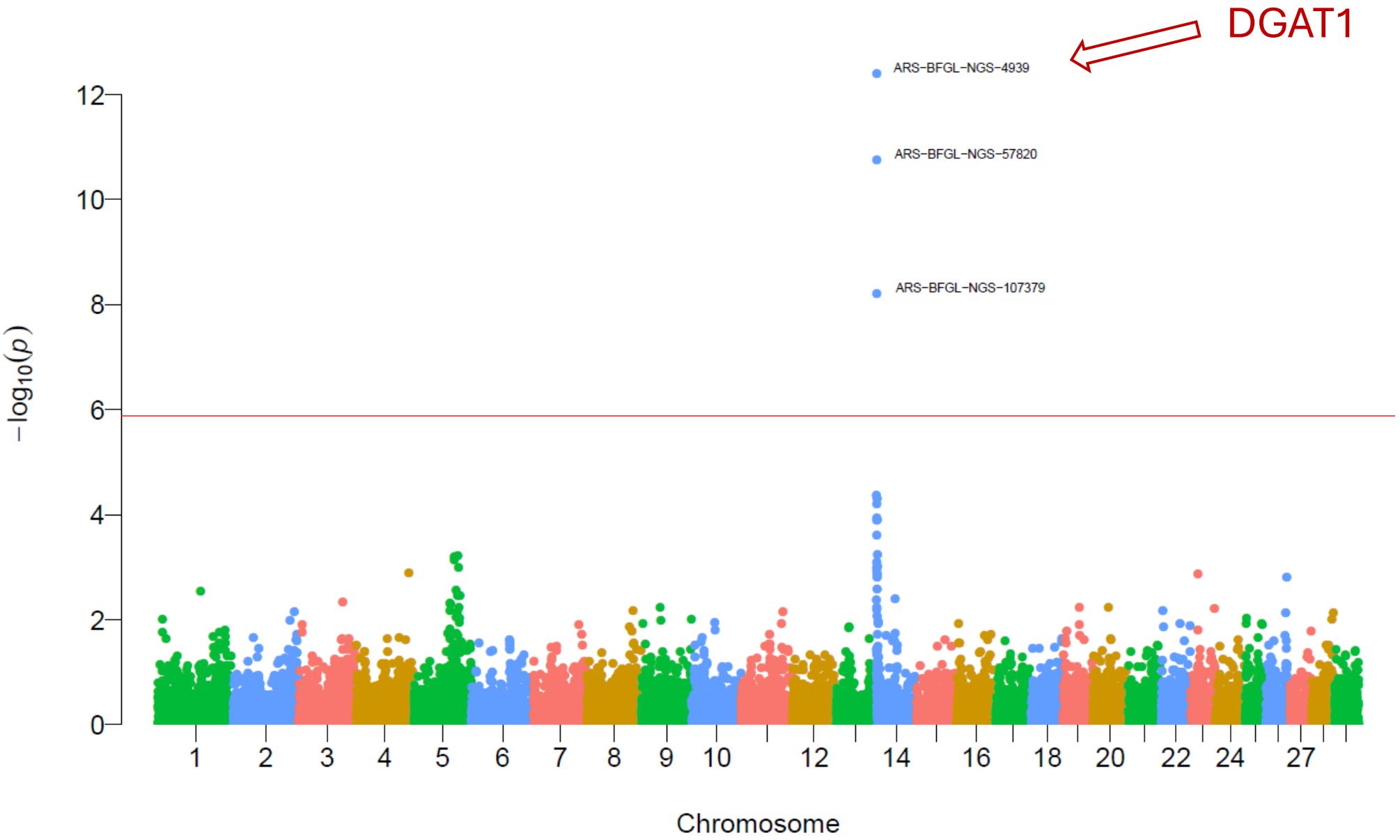
p-values in ssGWAS for US Holsteins

- US HOL 2009 data: milk, fat, protein
- Single-trait models
  - 10k genotyped bulls
  - 752k records for 100k daughters
  - 303k animals in ped

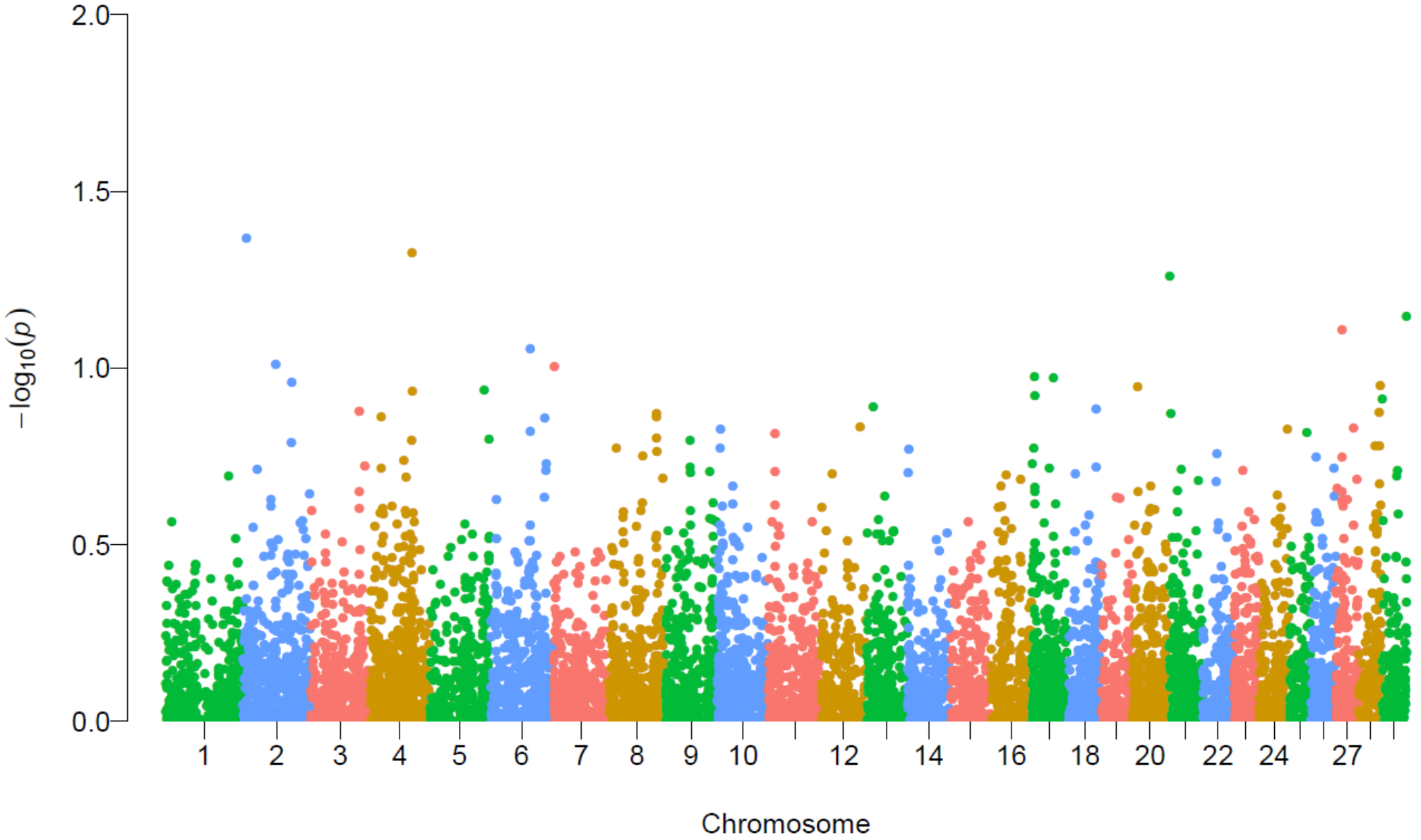
p-values in ssGWAS - Milk



p-values in ssGWAS - Fat



p-values in ssGWAS - Protein

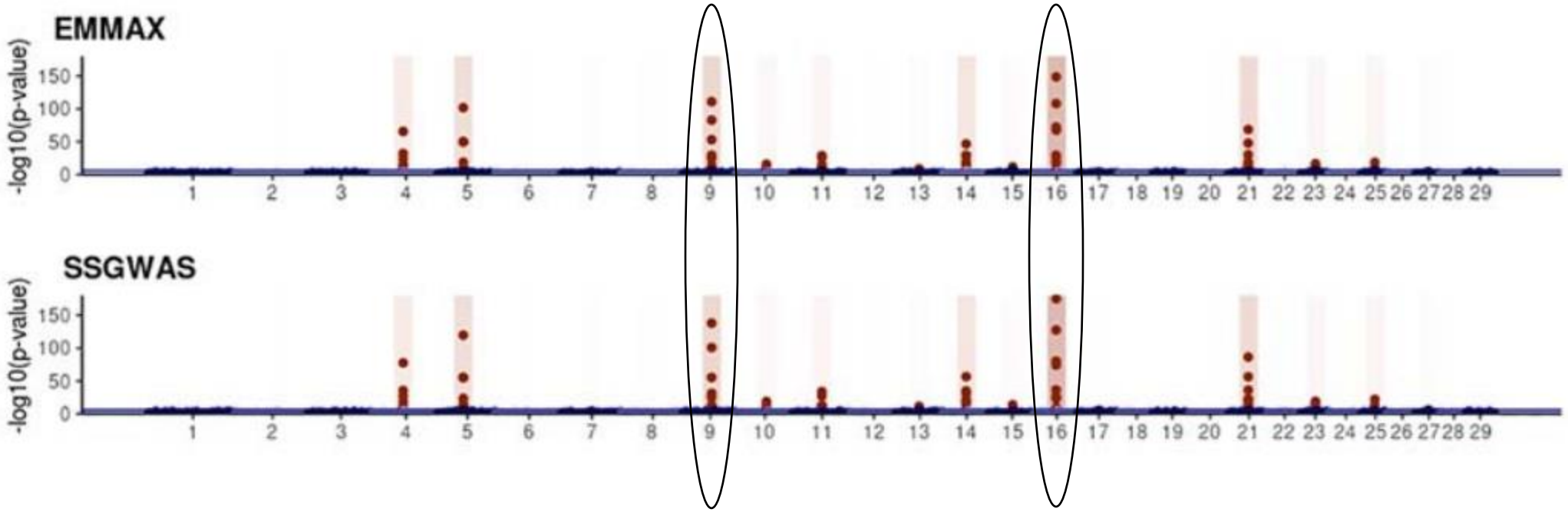


ssGWAS vs. EMMAX

- Simulated population (1 QTN per CHR)

14k genotyped sires
Deregressed EBV
(10 daughters)

14k genotyped sires
500k Pedigree
250k phenotypes



Association	EMMAX (Khang et al., 2010)	ssGWAS (Aguilar et al., 2019)
True Positive	55.2[a] (3.7)	61.6[a] (8.7)
False Positive	0.0	0.0

Mancin et al. (2021)

Thank you