postGSf90 - ssGWAS

Single-step GBLUP

- Integrates all available information
  - Phenotypes
  - Genotypes
  - Pedigree
- ssGBLUP vs. BayesX methods
  - infinitesimal model i.e. same variance for all SNPs

ssGWAS

- Combining methods
  - Unequal variances
  - Use all available information like in ssGBLUP

- Improve Accuracy of estimation of GEBV
  - For breeding and selection
  - Accuracy for estimation of SNP effects for GWAS

Equivalent Model

VanRanden et al 2009; Goddard, 2009; Habier el al 2007

Model that estimate SNPs effects

$$y = \mu + Za + e, \qquad \mathrm{var}(a) = D\sigma_a^2 \qquad u = Za$$

Model that estimate Breeding Values

$$y = \mu + u + e, \qquad \mathrm{var}(u) = G\sigma_u^2, \quad G = ZDZ'/k$$

Genomic Information ≈ genomic relationship

Simple conversion between :

Breeding values and SNP effects

$$u = Za$$

$$a = DZ'(ZDZ')^{-1}u$$

Stranden & Garrick, 2009

Equivalent Model

- SNP effects from GEBV's (Henderson, 1973; Strandén and Garrick, 2009):

$$\hat{u} = \frac{\sigma_u^2}{\sigma_a^2} DZ'G^{-1}\hat{a}_g = DZ'[ZDZ']^{-1}\hat{a}_g$$

Differential weight to each SNP

- Also, for each SNP effect (i-th):

$$\hat{\sigma}_{u,i}^2 = \hat{u}_i^2 2p_i(1-p_i)$$

postGSf90 par files

1) Parameter files:

(1) BLUPF90 (and preGSf90 for S1)

(2) postGSf90

2) OPTIONs:

BLUPf90 / PreGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION saveGInverse
OPTION weightedG w # A vector with length = M
```

postGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION ReadGInverse
OPTION chrinfo mapfile #format: snpID chr pos
OPTION weightedG w
# OPTION which_weight 1
# OPTION SNP_moving_average  n
# OPTION Manhattan_plot
```

3) Document:

http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90#gwas_options_postgsf90

Genet. Res., Camb. (2012), **94**, *pp.* 73–83. © Cambridge University Press 2012
doi:10.1017/S0016672312000274

Genome-wide association mapping including phenotypes from relatives without genotypes

H. WANG[1]*, I. MISZTAL[1], I. AGUILAR[2], A. LEGARRA[3] AND W. M. MUIR[4]
[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA*
[2] *Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay*
[3] *INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France*
[4] *Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151, USA*

(*Received 19 September 2011; revised 8 December 2011, and 9 March 2012; accepted 13 March 2012*)

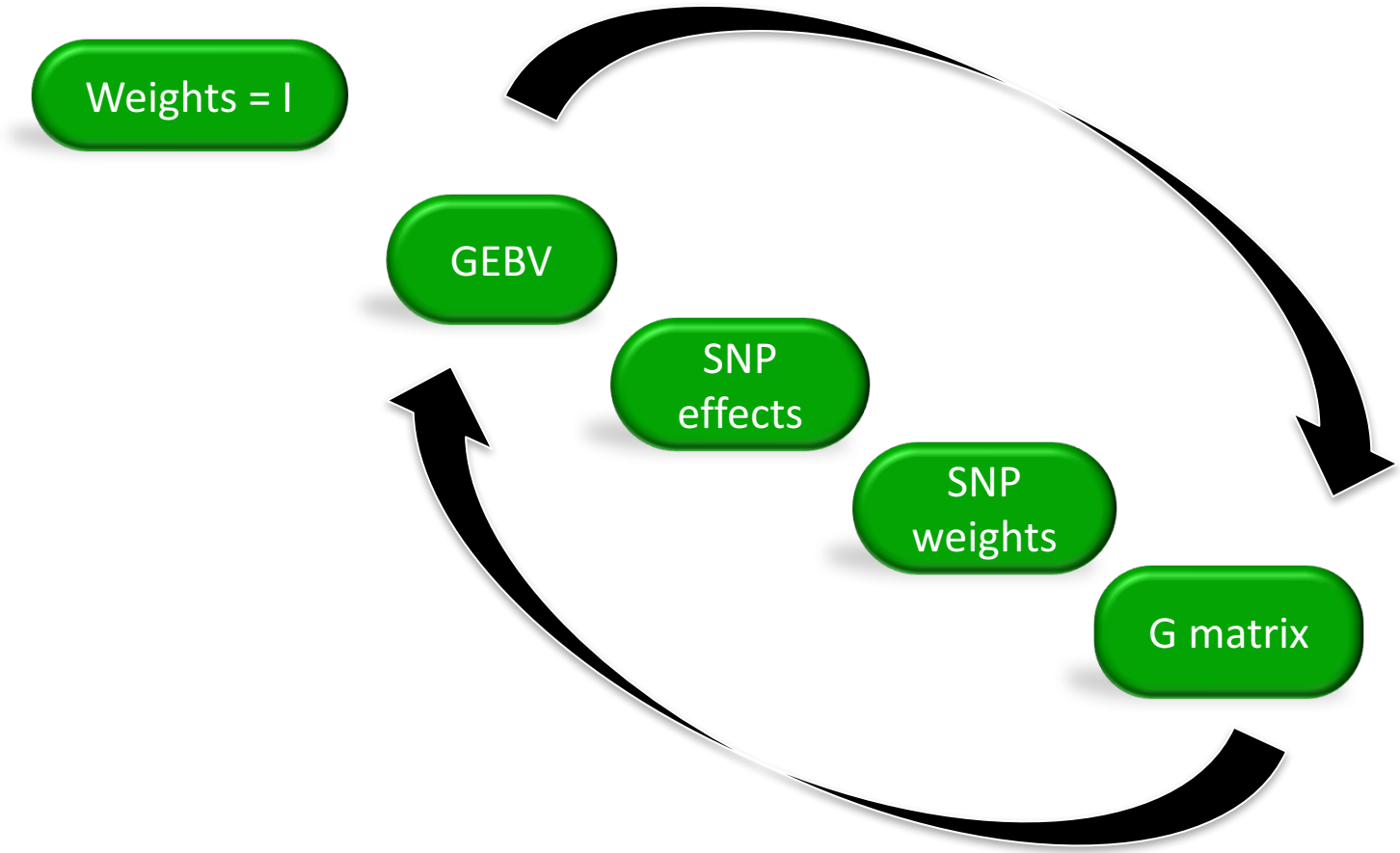
Computing algorithm

- Denote *t* as an iteration number and *i* as the *i*-th SNP

1. t=0, $D_{(t)}=I$, $G_{(t)}=ZD_{(t)}Z'\lambda$
2. Compute $\hat{a}_g$ by ssGBLUP
3. Calculate $\hat{u}_{(t)} = \lambda D_{(t)} Z' G_{(t)}^{-1} \hat{a}_g$
4. Calculate $d^*_{i_{(t+1)}} = \hat{u}^2_{i_{(t)}} 2p_i(1-p_i)$ for all SNPs (Zhang et al., 2010)
5. Normalize $D_{(t+1)} = \frac{tr(D_{(0)})}{tr(D^*_{(t+1)})} D^*_{(t+1)}$
6. Calculate $G_{(t+1)} = ZD_{(t+1)}Z'\lambda$
7. t=t+1
8. Exit , or loop to step 2 or 3

Weighted ssGBLUP

WssGBLUP(Wang et al., 2012)



- Gives more weight to important markers

Simulated data

1. QMSim
2. Simple model: $\mathbf{y} = \mathbf{1}\mu + \mathbf{Z_a}\mathbf{a} + \mathbf{e}$
3. 10 QTLs w. 3000 SNP markers on 2 chromosomes
4. N = 15,800

   N_g= 1500
5. h^2=0.5, all due to QTLs (No Polygen)
6. 10 replications

Different Scenarios

- Scenario 1
  - Run only one BLUP and get GEBV
  - Estimate SNP effects from GEBV using weighted Genomic matrix
  - Multiple trait or random correlated effects
- Scenario 2
  - Get EBVs with weighted genomic relationship matrix
  - Estimate SNP effects from GEBV using updated solutions
  - Single trait analysis - fit one genomic relationship matrix

postGSf90 bash script

- Scenario 1：

```
# run 1 time GBLUP to get GEBVs:
echo par.b90 | blupf90 | tee log.blupf90
# run x times PreGSf90 - postGSf90 to get SNPeff:
for i in 1 2 3 4 5 6 7 8 … … x
    do
     echo par.b90 | preGSf90 | tee log_preGS_$i
     echo postpar.b90  | postGSf90 | tee logpost_$i
     cp snp_sol snp_sol_$i
      #format: tr, eff, snpID, chr, pos, sol, w
     cp chrsnp  chrsnp_$i
     cp w w_$i
     awk '{ print $7 }' snp_sol >  w
done
```

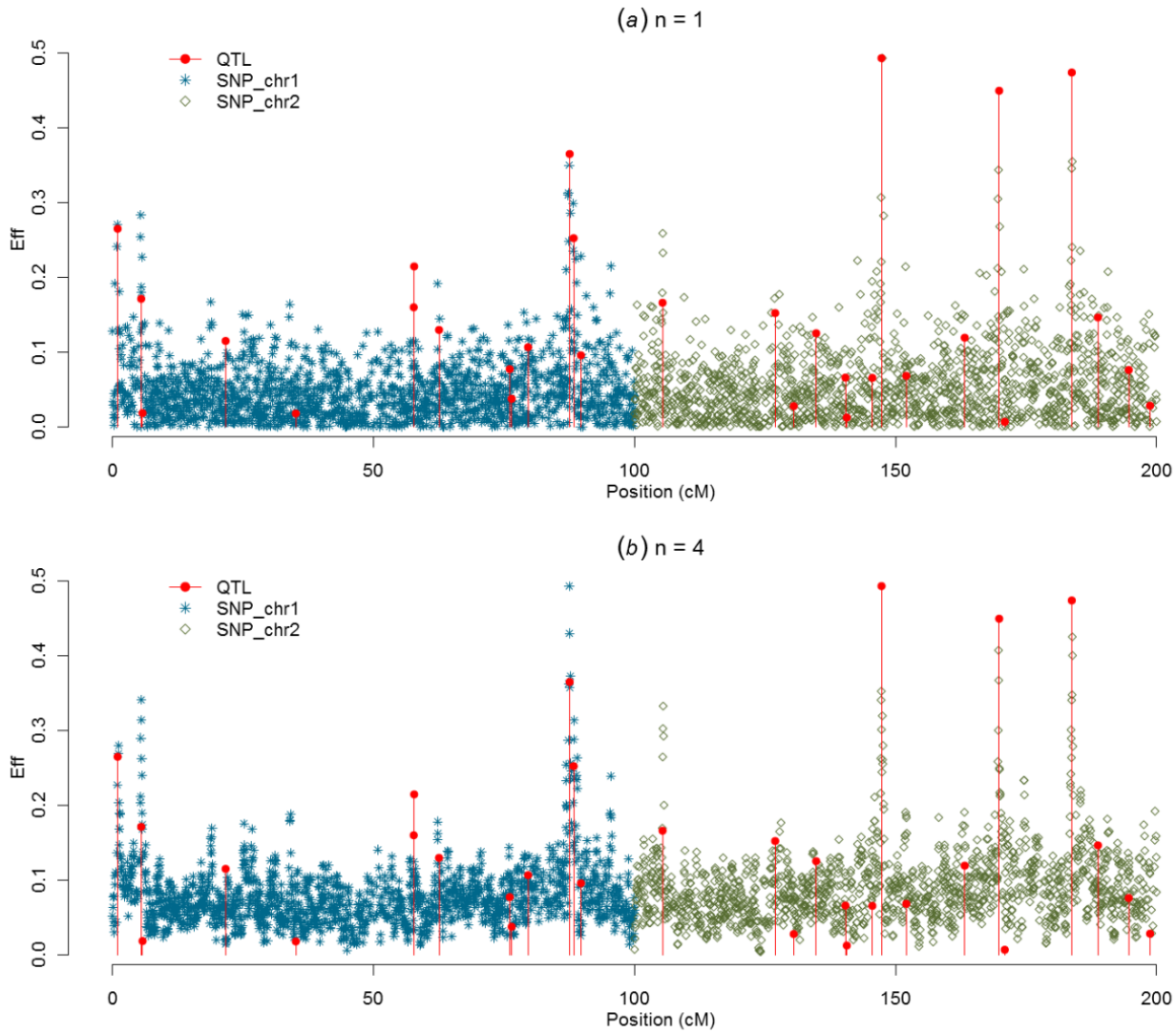
- Scenario 2 :

```
for i in 1 2 3 4 5 6 7 8 … … x
    do
      echo par.b90 | blupf90 | tee logpre_$i
      cp solutions solutions_$i
      echo postpar.b90  | postGSf90 | tee
logpost_$i
      cp snp_sol snp_sol_$i
      cp chrsnp  chrsnp_$i
      cp w w_$i
      awk '{ print $7 }' snp_sol>  w
  done
```

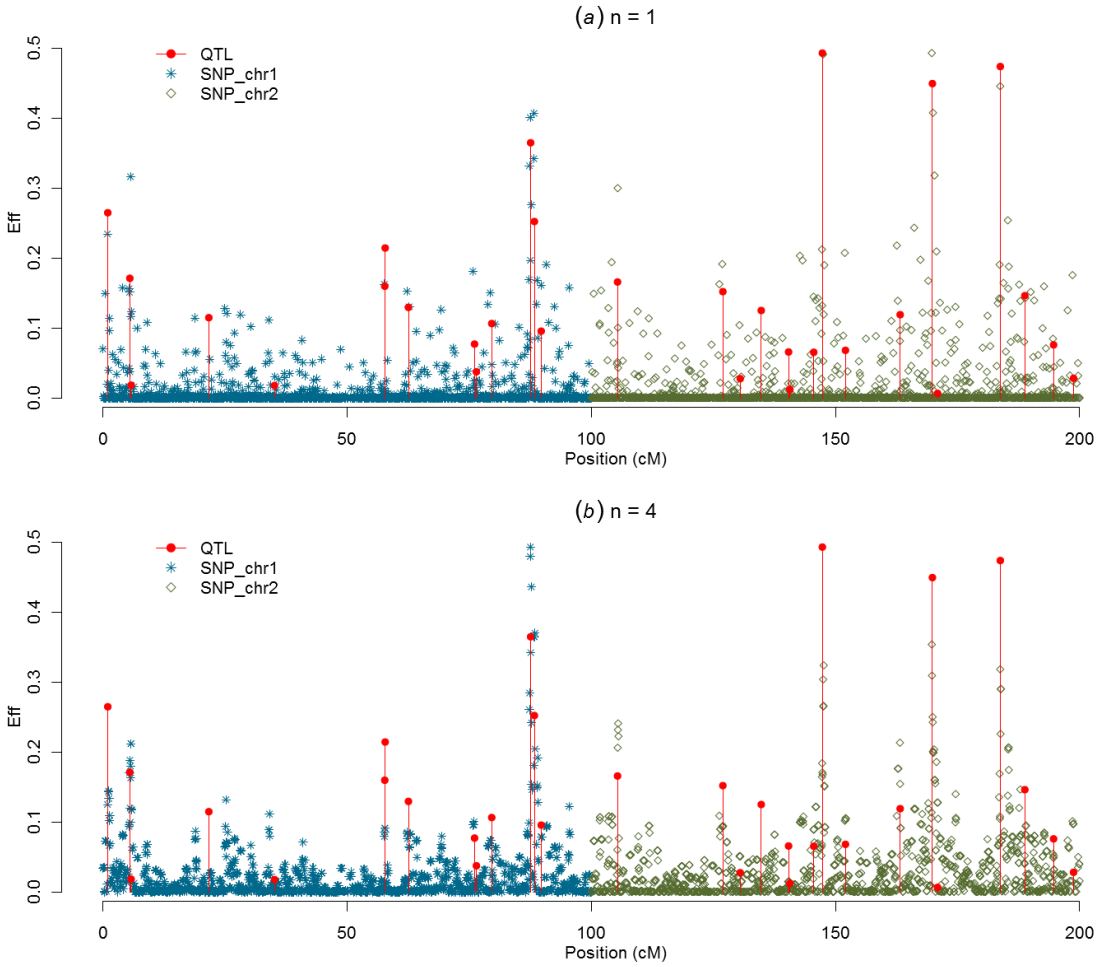
Methods

1. Single marker model: WOMBAT
2. BayesB using de-regressed proofs : GENSEL
3. ssGBLUP: S1 & S2

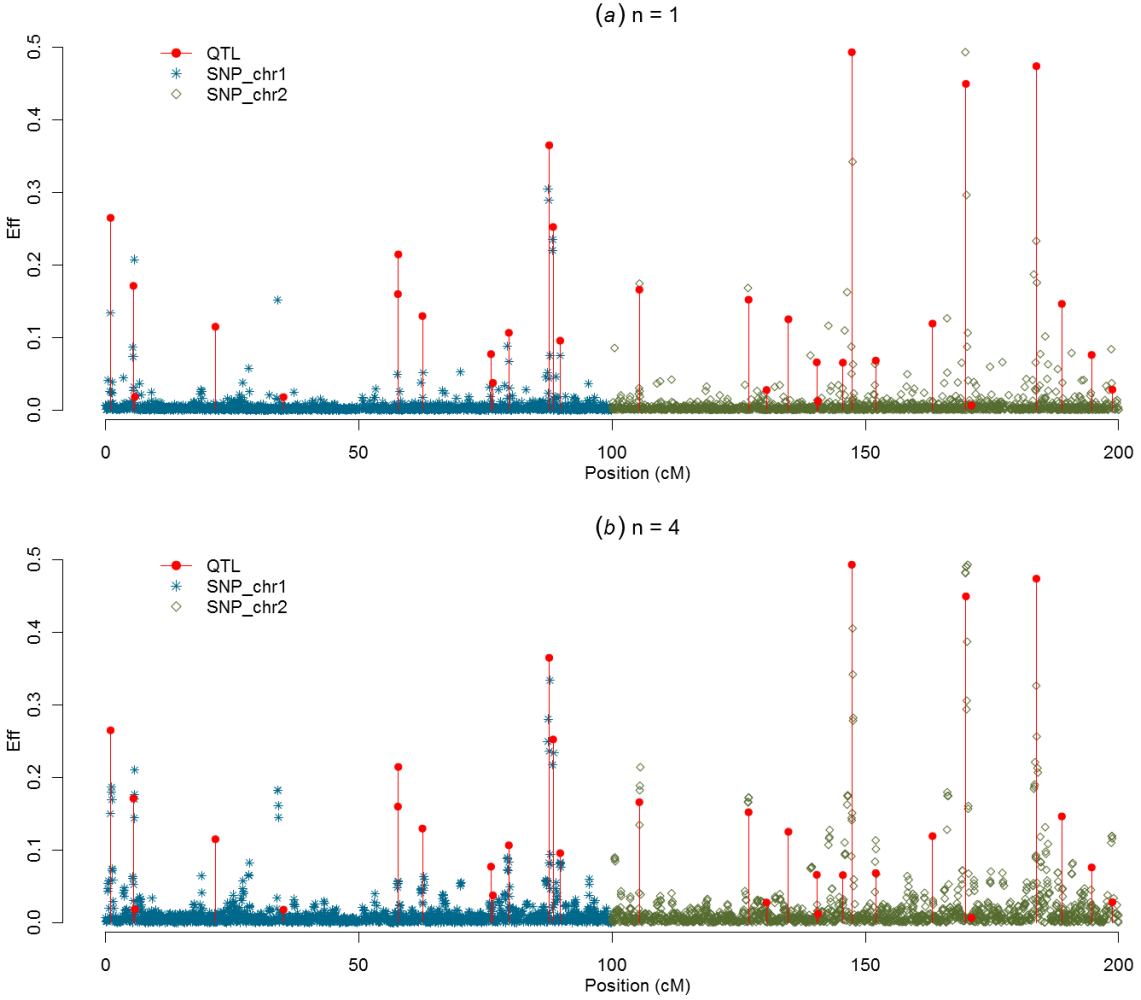
Manhattan plot of S1



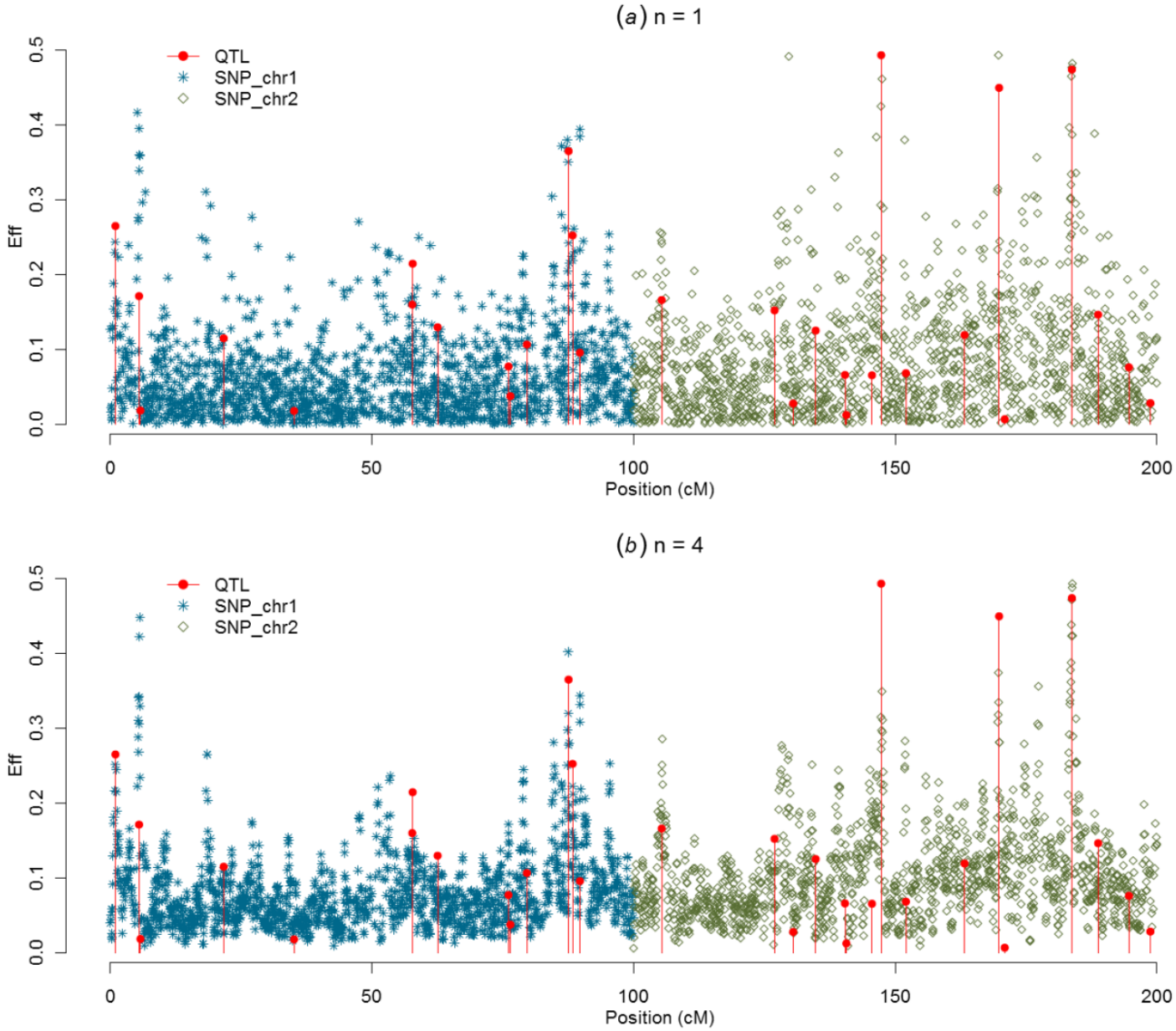
Manhattan plot of S2



Manhattan plot of BayesB



Manhattan plot of WOMBAT



Accuracy of (G)EBVs

	EBVs							
BLUP	0.81 (0.01)							
	it1*	it2	it3	it4	it5	it6	it7	it8
ssGBLUP	0.87 (0.01)	0.89 (0.01)	0.88 (0.01)	0.88 (0.02)	0.88 (0.02)	0.87 (0.02)	0.87 (0.02)	0.87 (0.02)
	NW†	*c*=0.1						
BayesB_DP	0.88 (0.02)	0.88 (0.02)						

Accuracy of SNP effects

Table 3. *Average correlations (standard deviations) between QTL effects and sum of cluster of m SNP effects using ssGBLUP*

S1*	1†	2	4	8	16	40
it1	0.53 (0.07)	0.68 (0.05)	0.79 (0.03)	0.81 (0.02)	0.80 (0.03)	0.62 (0.08)
it2	0.46 (0.07)	0.66 (0.05)	0.78 (0.02)	0.82 (0.02)	0.81 (0.02)	0.63 (0.08)
it3	0.43 (0.07)	0.64 (0.05)	0.77 (0.02)	0.81 (0.02)	0.80 (0.02)	0.62 (0.08)
it4	0.42 (0.07)	0.63 (0.05)	0.77 (0.02)	0.81 (0.02)	0.80 (0.02)	0.62 (0.08)
it5	0.41 (0.07)	0.63 (0.05)	0.76 (0.02)	0.80 (0.02)	0.79 (0.02)	0.61 (0.08)
it6	0.41 (0.07)	0.62 (0.05)	0.75 (0.02)	0.80 (0.02)	0.79 (0.02)	0.61 (0.07)
it7	0.41 (0.07)	0.62 (0.05)	0.75 (0.02)	0.80 (0.02)	0.79 (0.02)	0.61 (0.07)
it8	0.41 (0.07)	0.62 (0.05)	0.75 (0.02)	0.80 (0.02)	0.79 (0.02)	0.60 (0.07)
S2	1	2	4	8	16	40
it1	0.53 (0.07)	0.68 (0.05)	0.79 (0.03)	0.81 (0.02)	0.80 (0.03)	0.62 (0.08)
it2	0.44 (0.09)	0.65 (0.06)	0.77 (0.03)	0.82 (0.03)	0.81 (0.02)	0.63 (0.06)
it3	0.41 (0.08)	0.62 (0.05)	0.75 (0.03)	0.79 (0.03)	0.79 (0.03)	0.65 (0.06)
it4	0.40 (0.07)	0.61 (0.05)	0.73 (0.03)	0.77 (0.03)	0.78 (0.03)	0.64 (0.06)
it5	0.40 (0.07)	0.60 (0.05)	0.72 (0.04)	0.76 (0.04)	0.77 (0.04)	0.64 (0.06)
it6	0.40 (0.07)	0.60 (0.05)	0.72 (0.04)	0.75 (0.04)	0.76 (0.04)	0.63 (0.06)
it7	0.40 (0.07)	0.60 (0.05)	0.72 (0.04)	0.75 (0.04)	0.76 (0.04)	0.63 (0.06)
it8	0.40 (0.07)	0.60 (0.05)	0.71 (0.04)	0.75 (0.04)	0.76 (0.04)	0.63 (0.06)

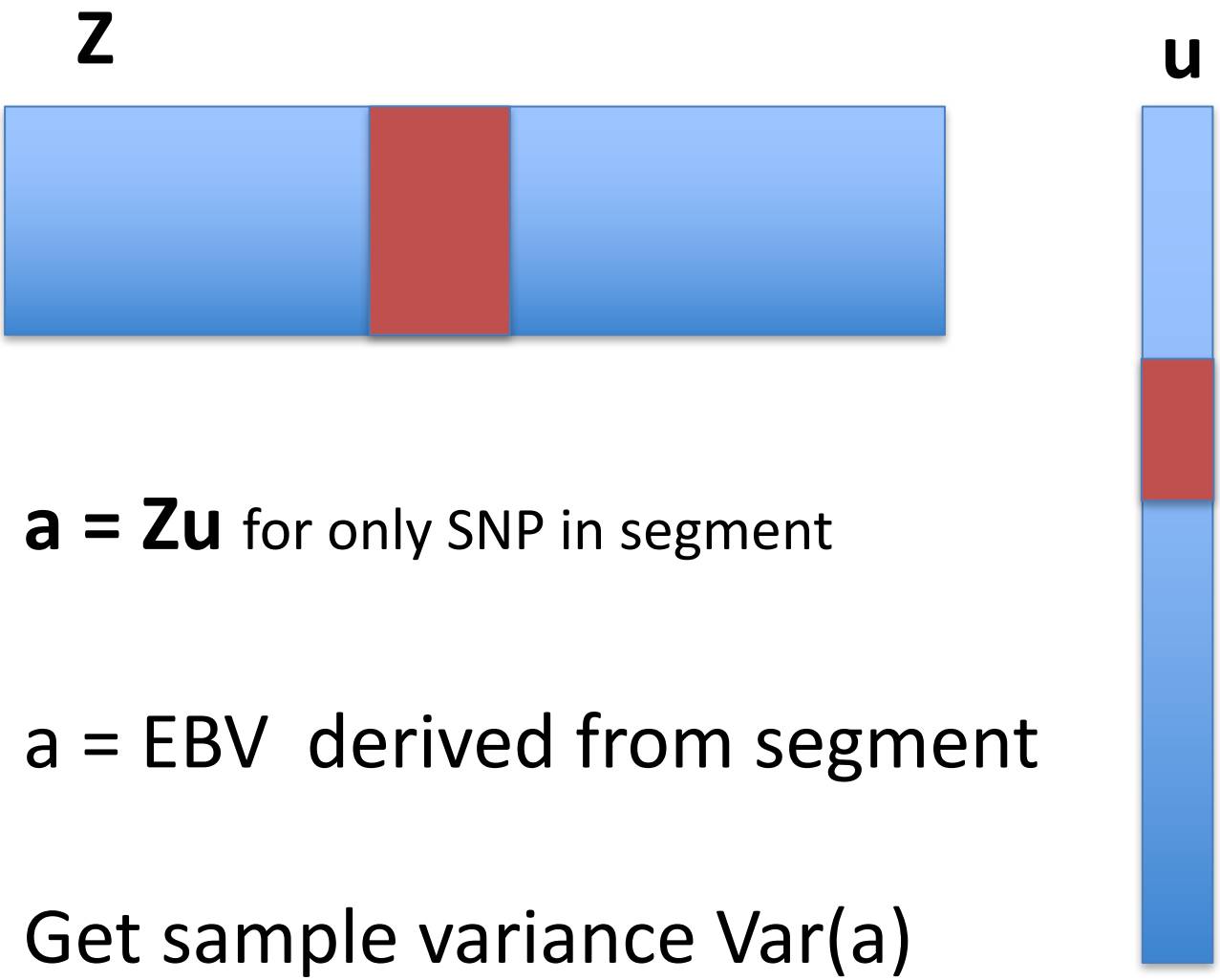
* S1: update weights for SNP effects but not for GEBVs; S2: update weights for both GEBVs and SNP effects in each iteration.

† Number of SNPs (i.e. *m* ranges from 1 to 40) in each cluster.

Variances explained by segments

- ISU propose to present results from GWAS using variance explained by windows of adjacent SNP
- Fan et al 2011, Onteru et al 2011, Peters el al 2012, etc.
- Potentially use of bootstrap to get significance of detected QTL

Windows Variances



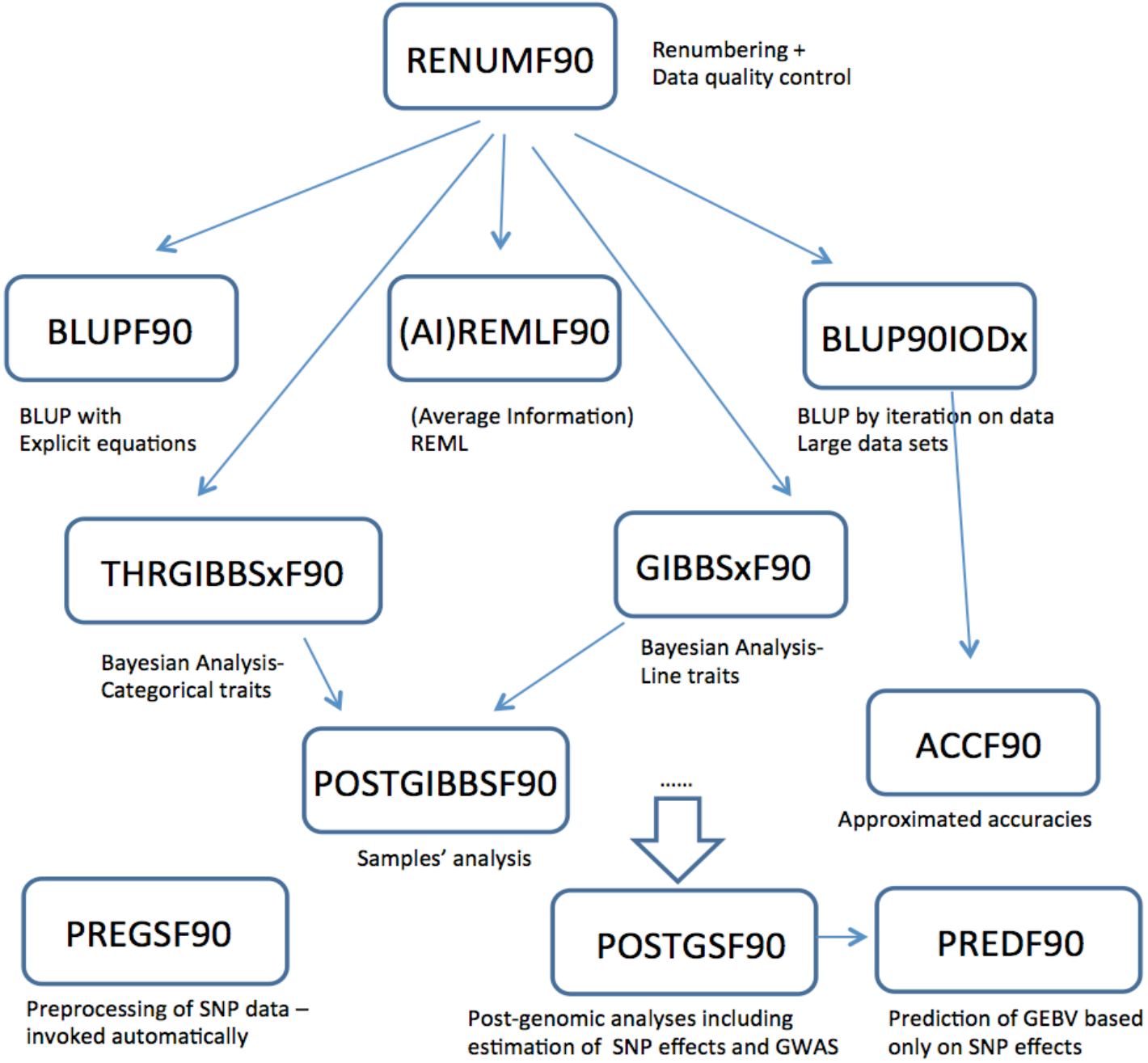
a = Zu for only SNP in segment

a = EBV derived from segment

Get sample variance Var(a) from genotyped individuals

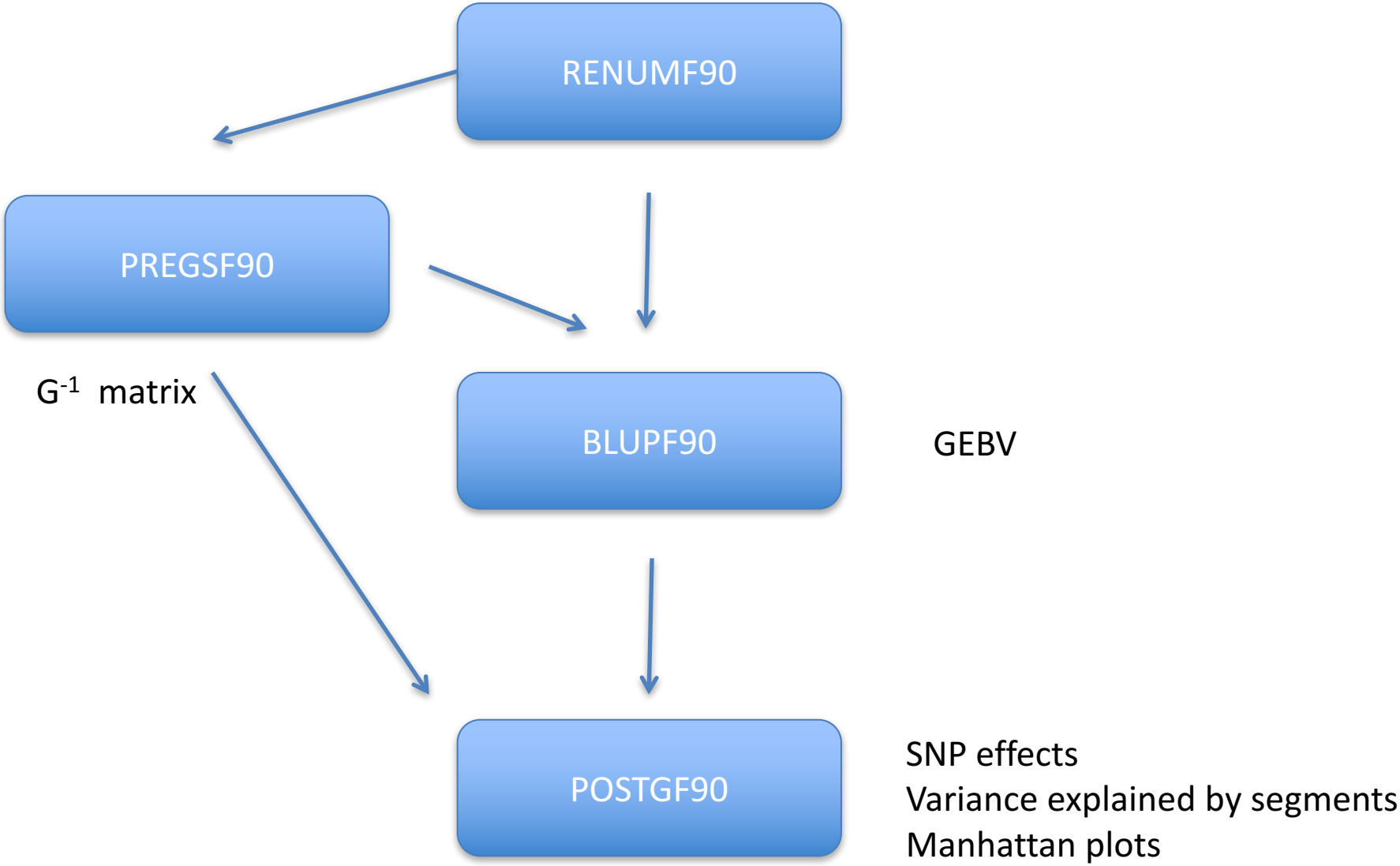
postGSf90

- Genomic POST processing program
- Extract SNP effects from *solutions* after genomic evaluations (GLBUP and ssGLBUP)
- Calculate variance explained by segments



Controlled by the same parameter file!

ssGWAS – postGSf90



postGSf90 par files

1) Parameter files:

(1) BLUPF90 (and preGSf90)

(2) postGSf90

2) OPTIONs:

BLUPf90 / PreGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION saveGInverse
OPTION weightedG w # A vector with length = M
```

postGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION ReadGInverse
OPTION chrinfo mapfile #format: snpID chr pos
OPTION weightedG w
# OPTION Manhattan_plot
```

3) Document:

http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90#gwas_options_postgsf90

POSTGSF90 Options

```
OPTION Manhattan_plot
```

Plot using GNUPLOT the Manhattan plot (SNP effects) for each trait and correlated effect.

```
OPTION Manhattan_plot_R
```

Plot using R the Manhattan plot (SNP effects) for each trait and correlated effect.
Tif images are created: *manplot_St1e2.tif.*
Note: *t1e2* corresponds to trait 1, effect 2.
CAIRO packaged is required.

```
OPTION plotsnp <n>
```

Control the values of SNP effects to use in Manhattan plots

- 1: plot regular SNP effects: abs(val)
- 2: plot standardized SNP effects: abs(val/sd) (default)

POSTGSF90 Options

```
OPTION windows_variance n
```

Calculate the variance explained by *n* adjacents SNPs.

```
OPTION windows_variance_mbp n
```

Calculate the variance explained by *n* Mb window of adjacents SNPs.

```
OPTION windows_variance_type n
```

Set windows type for variances calculations

- 1: moving windows
- 2: exclusive windows

```
OPTION which_weight x
```

Generate a weight variable to be used in the creation of a weighted genomic relationship matrix G=ZDZ'

- 1: w = y^2 * (2(p(1-p)))
- 2: w = y^2

with scaled weight = w * nSnp/sum(w)

Output files 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.

```
chrsnp
```

contains data to create plot by GNUPLOT

- 1: trait
- 2: effect
- 3: values of SNP effects to use in Manhattan plots
- 4: SNP
- 5: Chromosome
- 6: Position

```
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

```
windows_segment
```

contains information of windows segments used to get variance explainded

- 1: label
- 2: window size (number of SNP)
- 3: Start SNP number for the window
- 4: End SNP number for the window
- 5: identification of window: (ChrNumber)'_'(startPositionMBP)
- 6: Start (ChrNumber)'_'(Position) for the window
- 7: End (ChrNumber)'_'(Position) for the window

```
windows_variance
```

contains variance explained for the biggest non-overlapping windows segments

- 1: trait
- 2: effect
- 3: Start SNP number or SNP name for the window
- 4: End SNP number or SNP name for the window
- 5: window size (number of SNP)
- 6: Start (ChrNumber)'_'(Position) for the window
- 7: End (ChrNumber)'_'(Position) for the window
- 8: identification of window: (ChrNumber)'_'(startPositionMBP)
- 9: variance explained by *n* adjacents SNP

```
snp_pred
```

contains allele frequencies + SNP effects

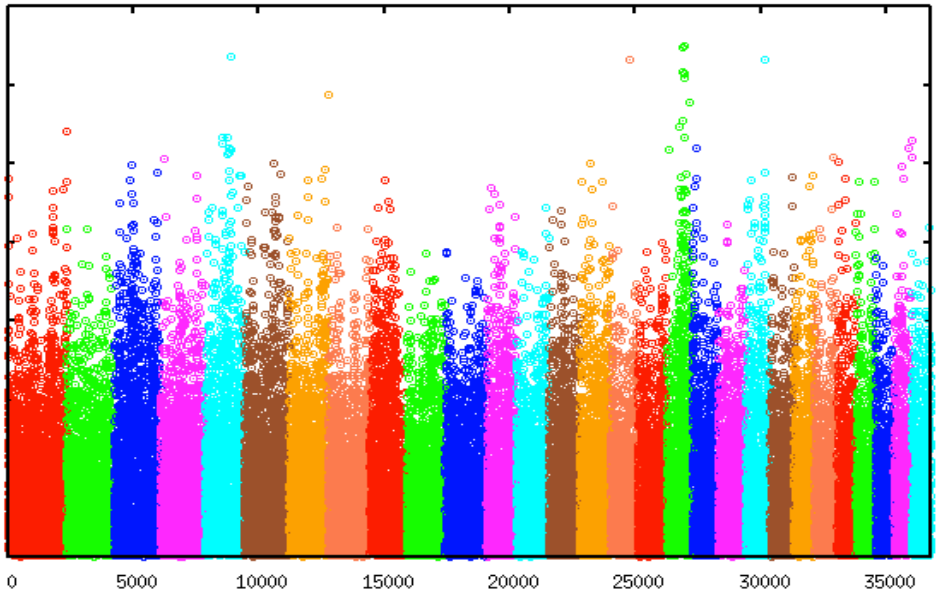
```
Graphic control files
```

Several files are created to generate graphics using either GNUPLOT or R
File names rules:

```
e.g.: 'Sft1e2.R'

first letter indicate
  'S' for solutions of SNP
  'V' for variance explained

t1e2
  indicates that the file is for the trait 1 and the effect 2

filename extension
   .gnuplot
   .R
   .pdf
   .png
   .tif
```

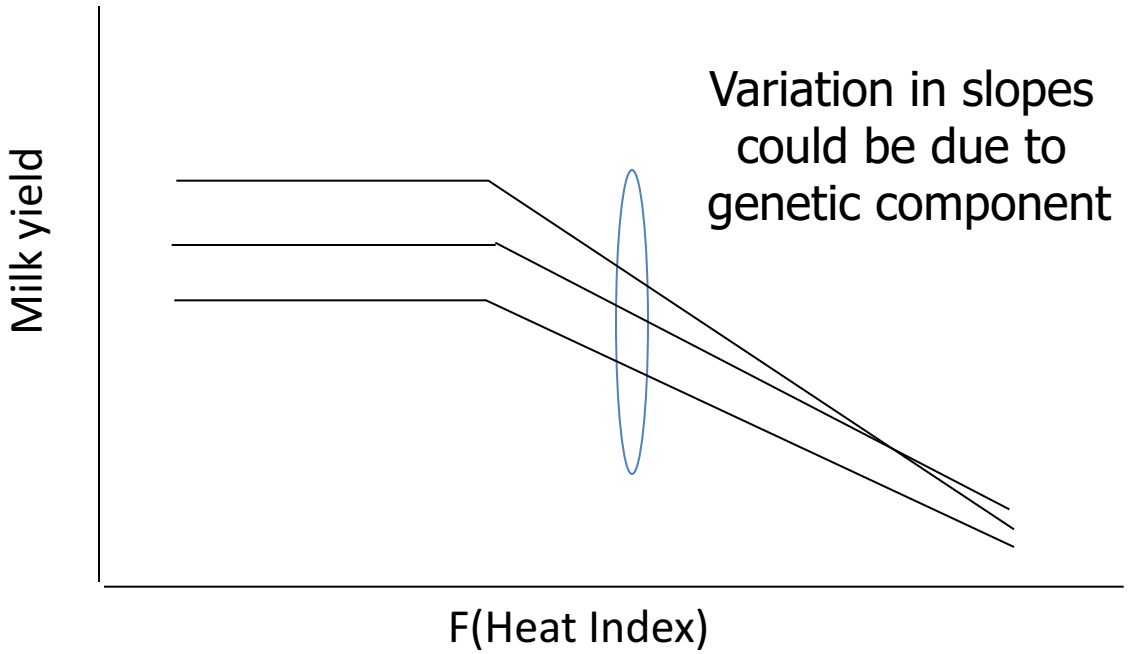
Single-Step GWAS Conception Rate

- Multiple-Trait US Holsteins Service records from AI
  - ~ 5 millions records, ~ 2.5 millions pedigrees
  - ~ 5,600 genotyped bulls
- Computing time
  - Complete evaluation 2 h
  - Estimates of SNPs 2 m



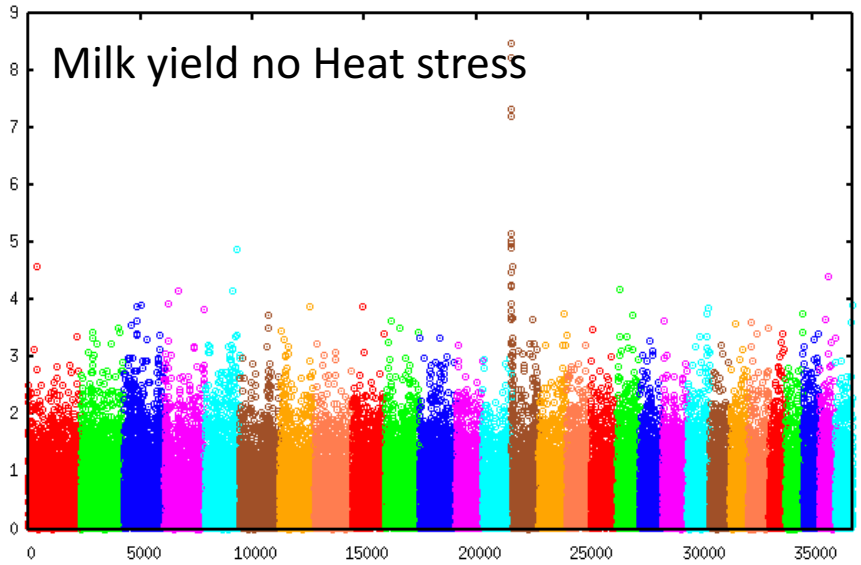
Model to study genetic of heat stress

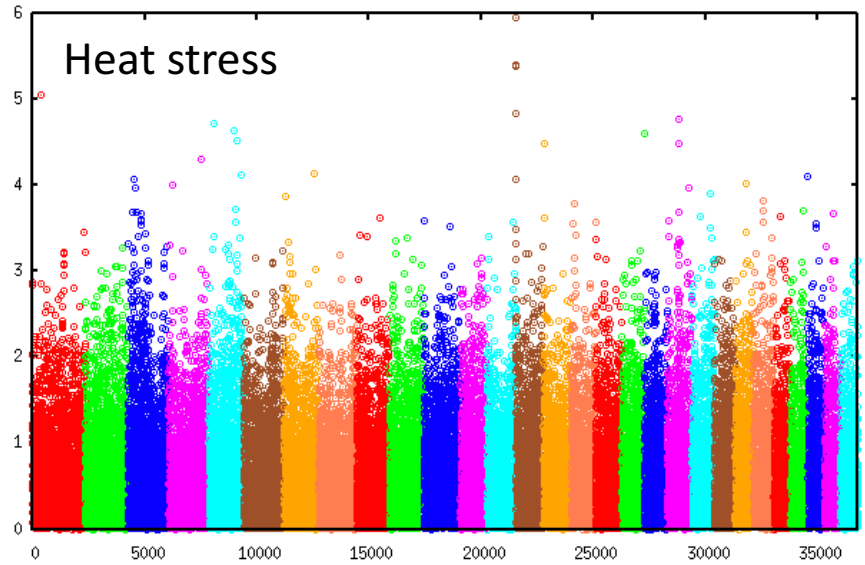
- Performance data + weather data (Ravagnolo & Misztal, 2000)



Single-Step GWAS Heat Stress

- Multiple-Trait Test-Day model heat tolerance
  - ~ 90 millions records, ~ 9 millions pedigrees
  - ~ 3,800 genotyped bulls
- Computing time
  - Complete evaluation ~ 16 h





Variance explained Heat Stress