Use of haplotypes based on 50K in Nordic Red

Progress Report December 2012

Esa Mäntysaari;
Ismo Strandén
BGE /BEL
MTT

- D. Boichard A,F, et al. 2012. Genomic selection in French dairy cattle . Animal Production Science, 2012, 52, 115–120
- Reasoning:
  - “it is believed that individual SNPs are unlikely to be systematically close to the QTLs as well as in complete disequilibrium with them.”
  - “Combining neighbouring SNPs into haplotypes is a simple way to increase informativity and is likely to generate a more complete linkage disequilibrium, at the expense of some over-parameterisation”
  - “Because these haplotypes are limited in size compared with conserved segments within breeds, their association with the QTLs is likely to be well conserved in the population and over several generations”

French genomic evaluation

- Around 300-700 haplo-blocks are recognized by trait
  - Few (32-40) haplo-blocks with largest variance are estimated using LD-LA QTL association analysis
  - Rest are selected from Elastic-net algorithm analyses
  - Each block is defined using 4-5 flanking SNPs

$$DRP_i = \mu + \sum_{k=1}^{N} (haplotype_{Pi,k} + haplotype_{Mi,k}) + a_i + e_i$$

Where var(u)=σ^2_u A, and σ^2_u is 30-40% of σ^2_a

Aim in our study:
Compare haplo-block assisted model and GBLUP model

- Data:
  - Nordic Red – bull data:

    1. MTT
    - 3912 reference and 801 candidate bulls
    - 38194 SNPs

    2. NAV (September 2012, Ulrik S Nielsen)
    - 4465 reference and 403 candidate bulls
    - 40070 SNPs
- Method:
  - Haplo-blocks based on 5 SNPs
  - Polygenic pedigree 10% – 50%

- **No Norwegian bulls used**
- **Different genotype data set**
  - **MAF >0.025**
  - **Call rate > 60%**

Implementation steps

1. Fit gaussian SNP model to get general variances
2. Fit all SNP by BayesB
   - Only the reference bulls used
   - Use the a'prior value of σ^2_g from gaussian model
   - Use $\pi = 0.1$ (the number of SNP linked with QTL)
3. Rank the SNP effects w.r.t. SNP estimates Choose appropriate number of SNPs (i.e. 600)
4. Haplotype (phase) the original data
   - We used FindhapV2 (Beagle could be better)

Implementation steps

4. Build 5-snp haplotypes around the peaking SNPs
   - Different rules for beginning of the chromosomes
   - Skip SNP closer than 2 apart from a "better" SNP

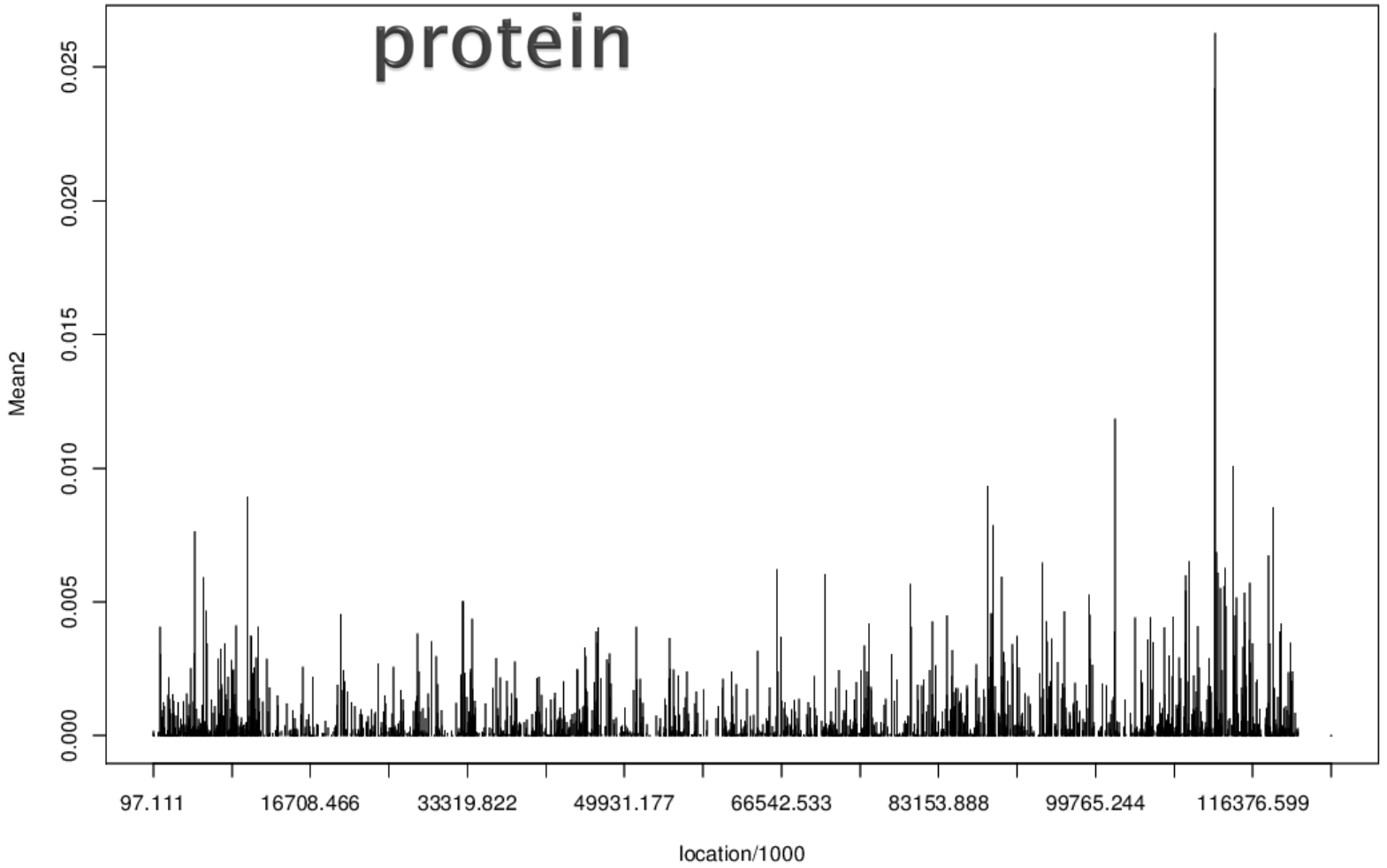
5. Enter the haploblocks into VCE estimation to find the haploblock variances
6. Use the same haploblocks and above variances, and fit "haploblock assisted BV model"
   - Choose appropriate polygenic variance

7. Validate the results

Results

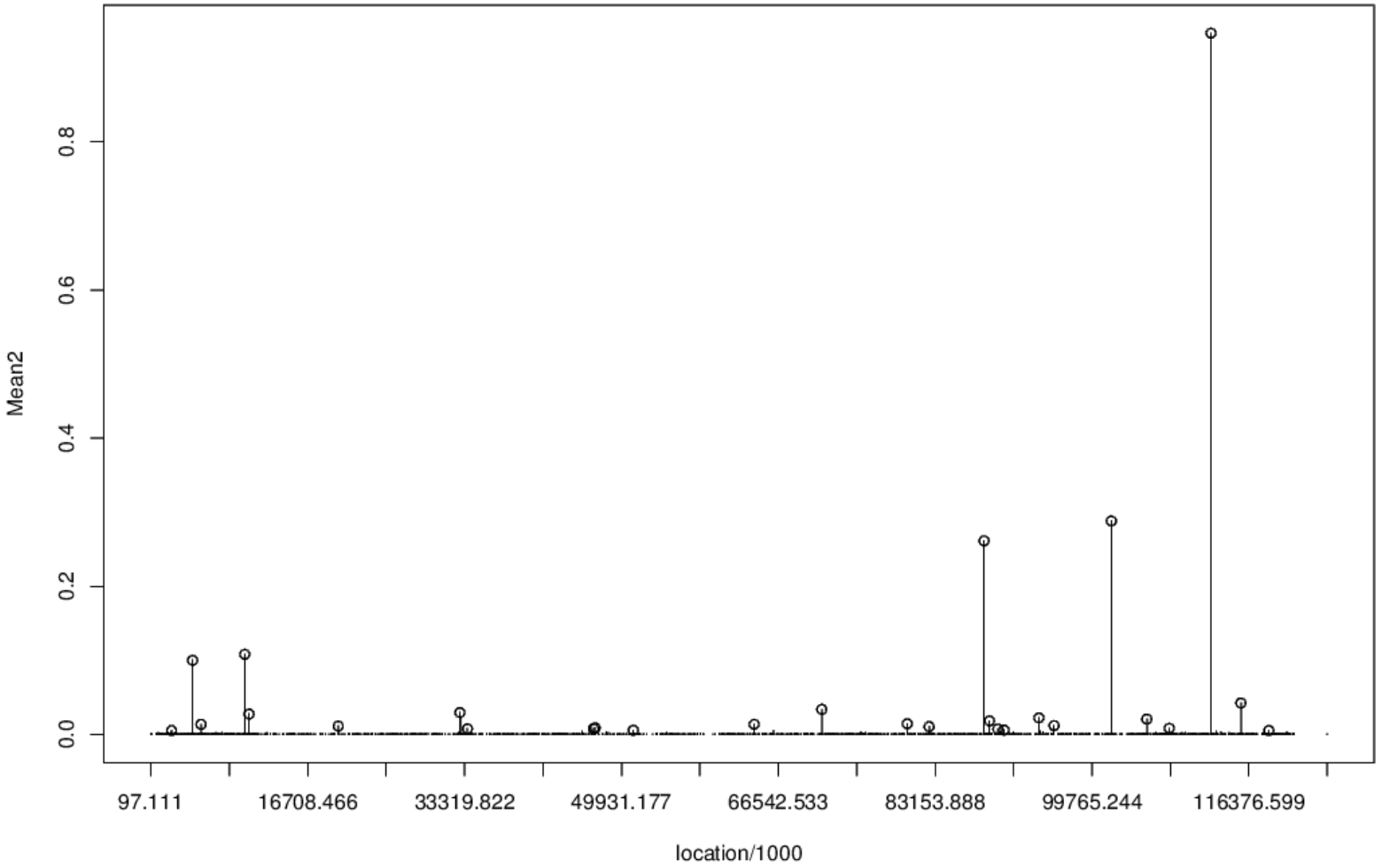
Challenge remains how to find the significant SNP:

- BLUP solutions chr. 5 protein

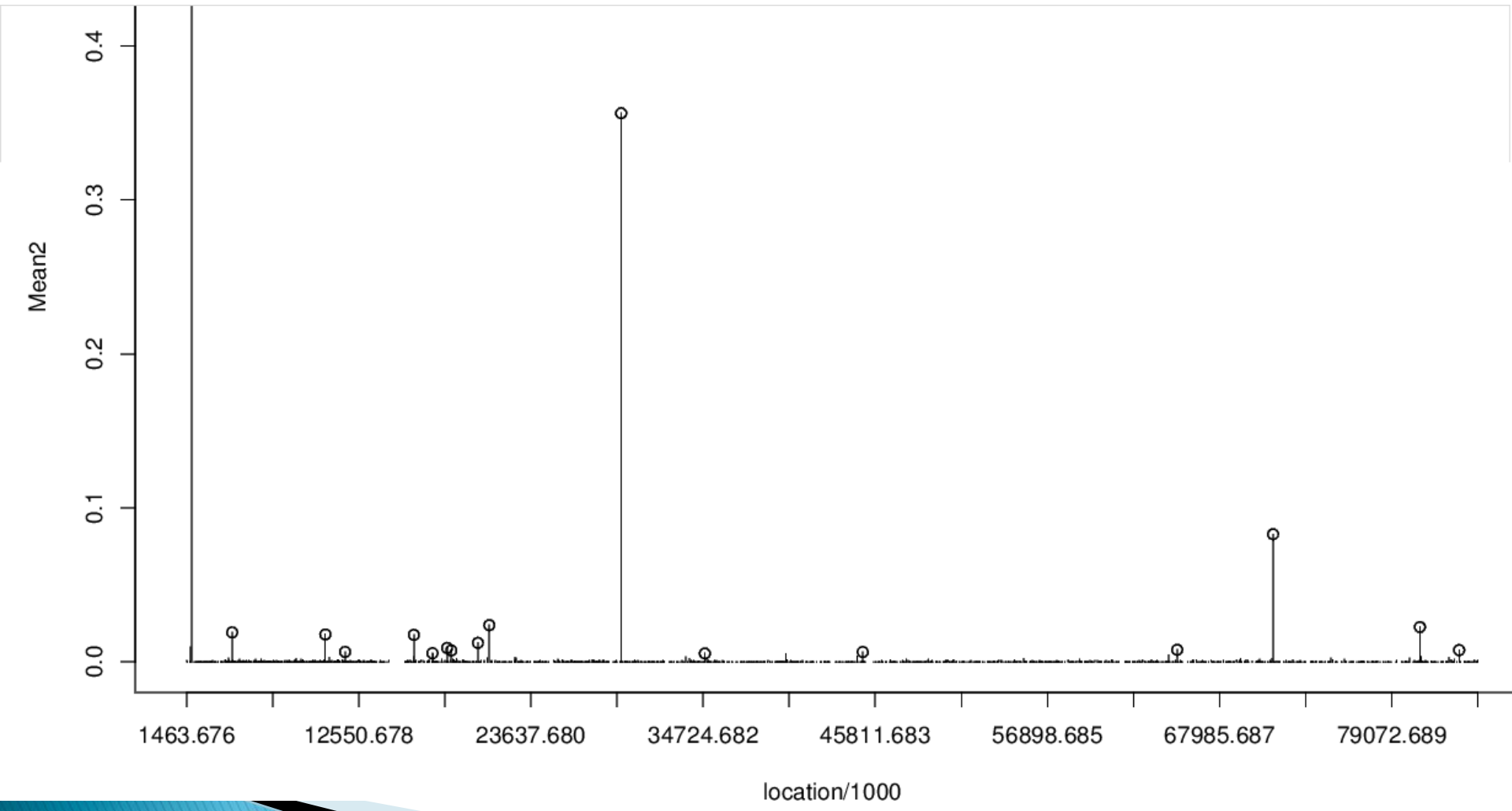


Challenge: how to find the significant SNP:

- Bayes B solutions, Protein chr. 5
- circles ● = selected SNP

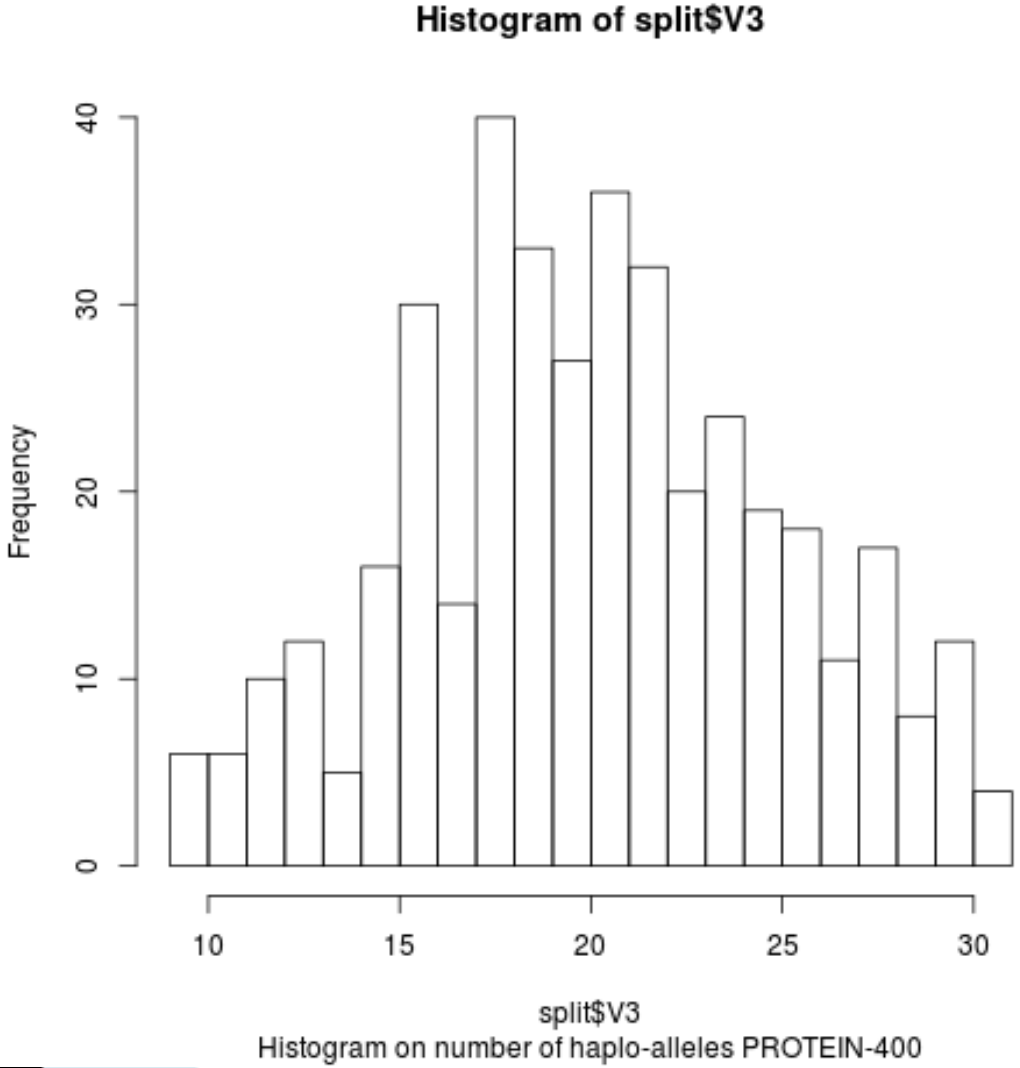


Protein Bayes B SNP solutions: chr. 14



Haplo-blocks - 400

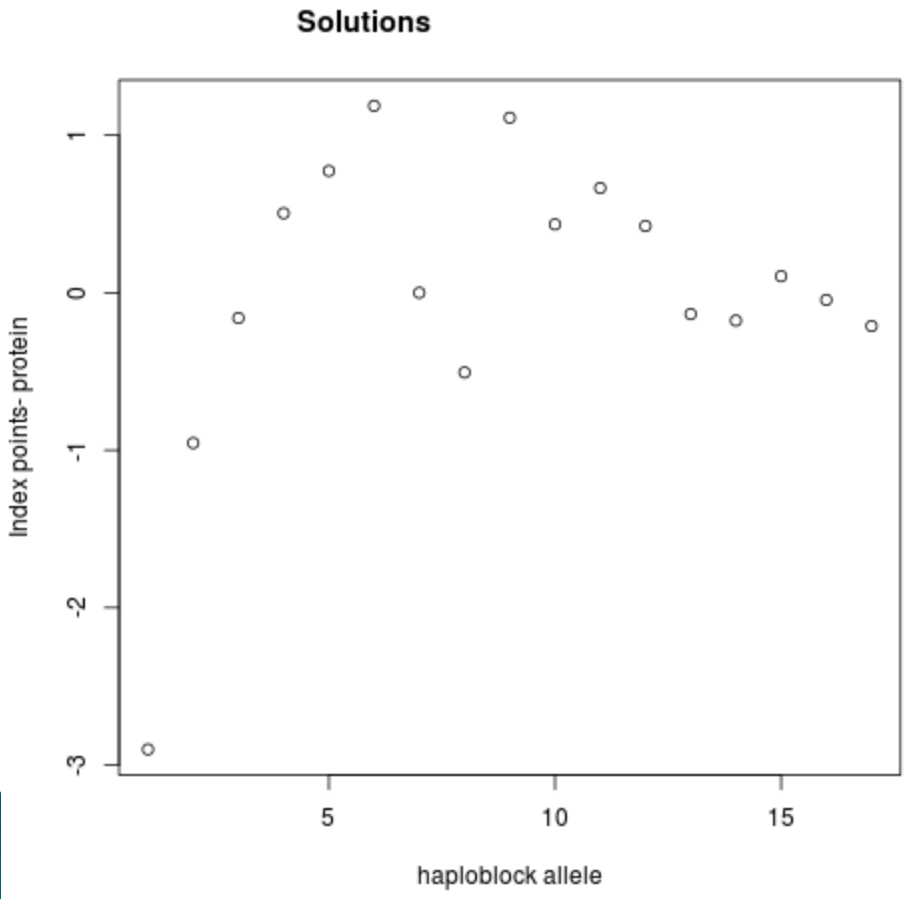
- on average 20 ”alleles” per block

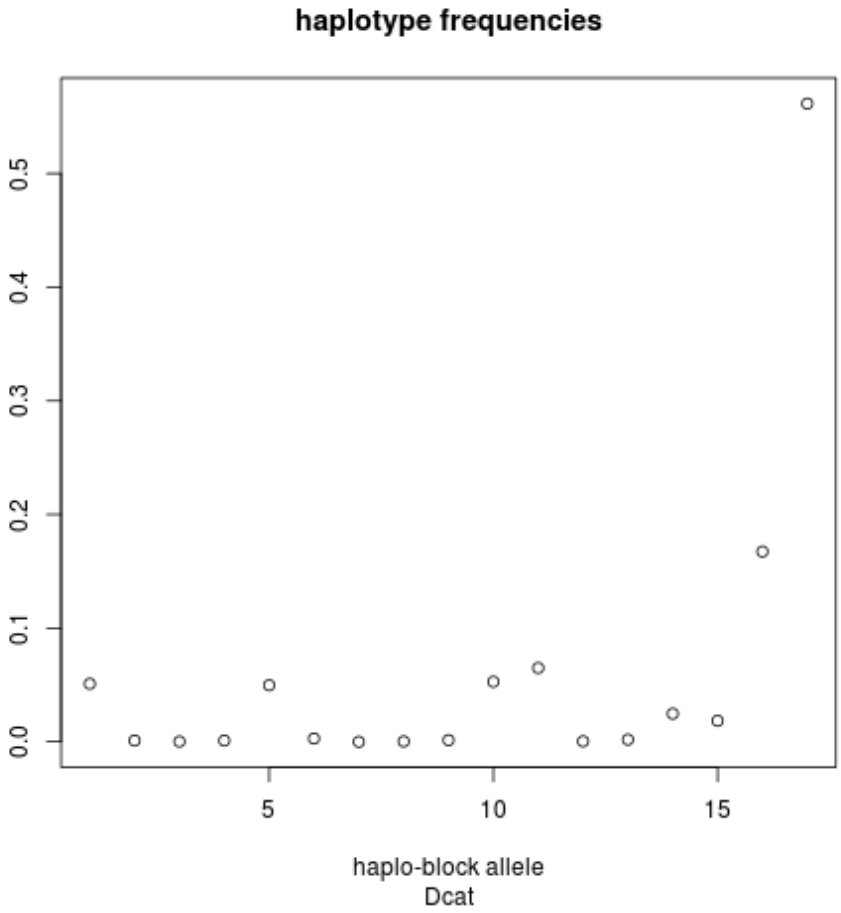


Histogram on number of haplo-alleles PROTEIN-400

"Dcat1 haploblock" has 18 variants

– MTT data





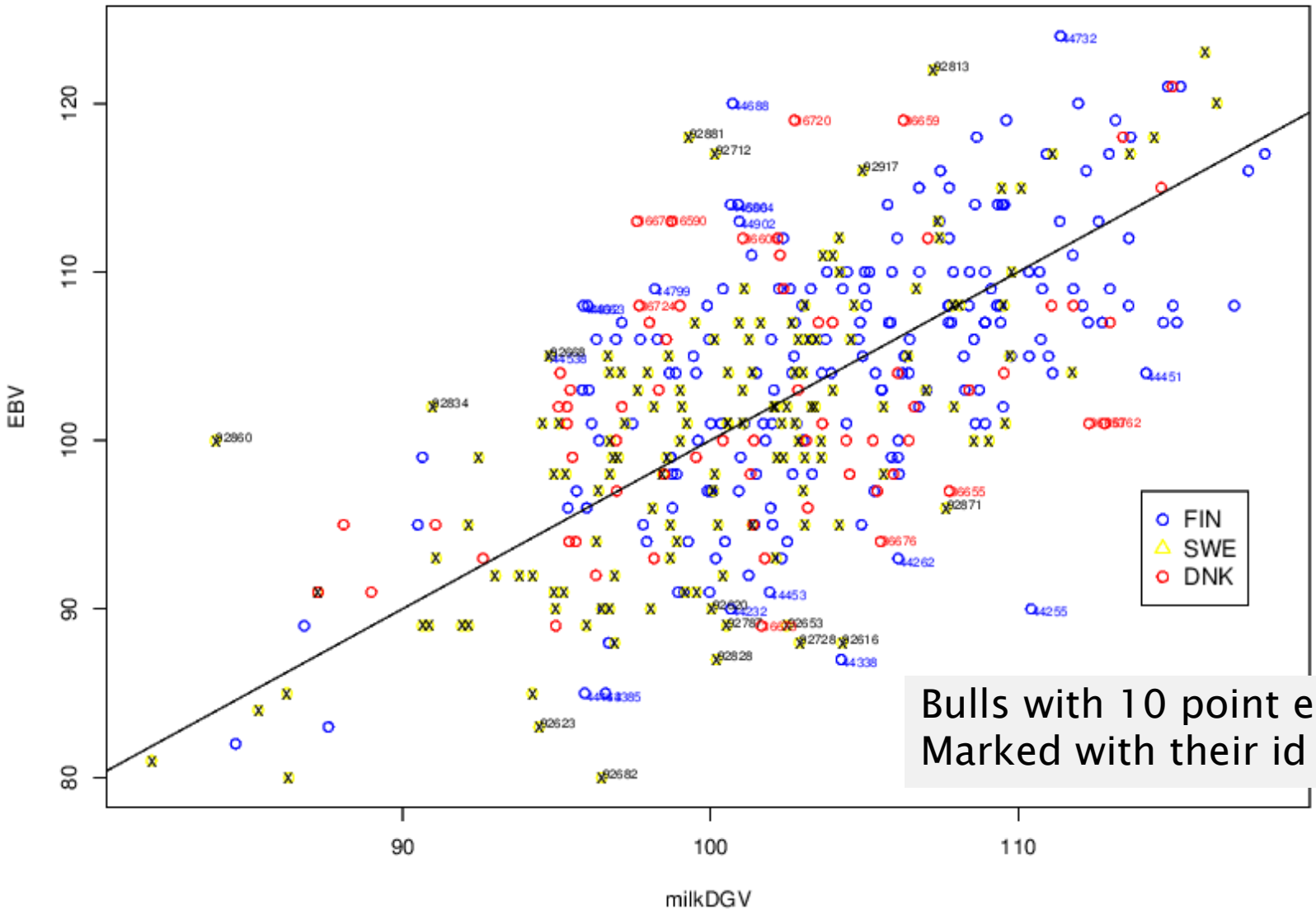
Validation of MA EBV – MTT data 400 haploblocks

	GEBV HB400		GBLUP	
polyg. %	R^2	b_1	R^2	b_1
w=0.10				
Protein	**0.387**	**0.86**	0.332	0.88
Milk	**0.452**	**0.88**	0.334	0.84
Fat	**0.472**	**0.89**	0.462	0.96
	GEBV HB200			
Protein	0.337	0.83		
	GEBV HB600			
Protein	0.392	0.804		

Validation of MA EBV – NAV data 600 haploblocks

	GEBV HB600		GBLUP	
polyg. %	R^2	b_1	R^2 (w=.20)	b_1
variable…				
Protein	**0.260**	**0.80**	0.298 (0.322)	0.70 (0.80)
Milk	**0.377**	**0.91**	0.367 (0.381)	0.82 (0.91)
Fat	**0.310**	**0.74**	0.319 (0.342)	0.69 (0.80)
Fertility	0.139	1.00	0.152 (0.156)	1.03 (1.10)

Example scatter plot: all candidate bulls



Bulls with 10 point error
Marked with their id number

Conclusions

- Haplo-block assisted model worked much better than the GBLUP in MTT data
- For NAV data the first results were not as good
  - Very sensitive to genotype data – missing SNPs etc.
  - Still uncertain how to setup the variances and weight w
  - Optimum number of effects? We used more that French…
- Work continues to clearify why the results are different in MTT and NAV data
  - Bulls are 80% the same
  - Genotypes are from same lab
  - ?