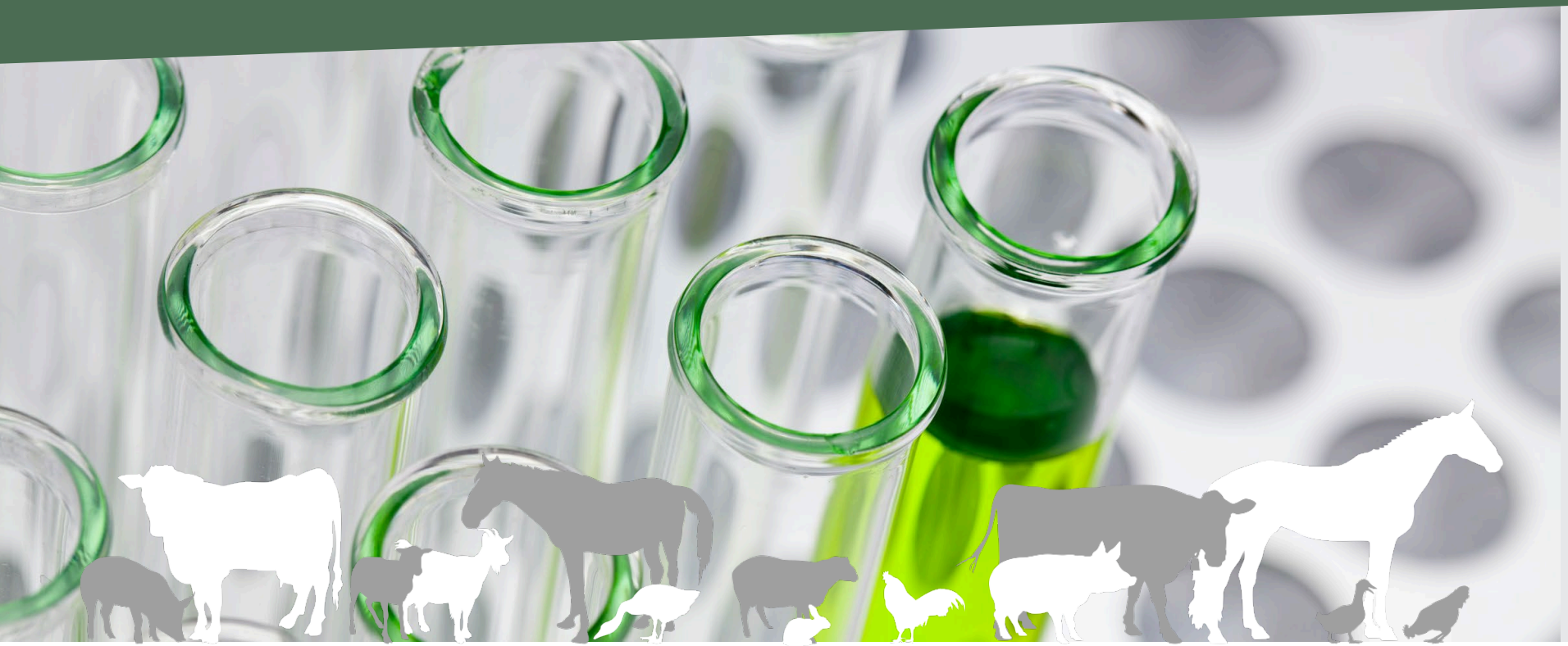


Genome-wide association study for stature for Holstein cattle based on whole genome sequence, using a mixed linear model

Dawid Słomian, Kacper Żukowski, Monika Skarwecka, Krzysztof Kotlarz, Joanna Szyda

Introduction

- 528 Holstein bulls
- Phenotype – Estimated Breeding Value for Stature
- Genotype – 1000 Bull Genomes Project

Purpose of the study

- Reducing dimensionality of data for WGS based on GWAS
- Define significant genomic regions
- Compare computing time and iterations

Data pre-processing (Plink 1.9)

- Filtering data
  - Call rates 0.95
  - Minor allele frequency 0.05

- Dimensionality reduction

Window size	100 000 bp
Step size	1 000 bp
LD treshold	(0.9, 0.75, 0.5) R^2

Data set	Number of SNPs	Size (Gigabytes)
Unpruned	10,771,020	11.4
Tag 0.90	5,874,889	5.8
Tag 0.75	1,343,767	1.4
Tag 0.50	448,468	0.5

Model

$$\boldsymbol{y = 1 + Vu + e}$$

y – bull's EBV for Stature

u – SNP additive genetic effects

V – incidence matrix

e – residuals

Computing server

- Linux Red Hat
- 260GB RAM
- 16 x Intel CPUs with 2.20GHZ

- MiXBLUP 3.0

Wald test for significant SNPs

$$Wald_{unpruned_n} = SNP_{solution_n} / sd(Solutions_{unpruned})$$

$$p\_value_n = 2\ x\ P(Z > |Wald_{unpruned_n}|)$$

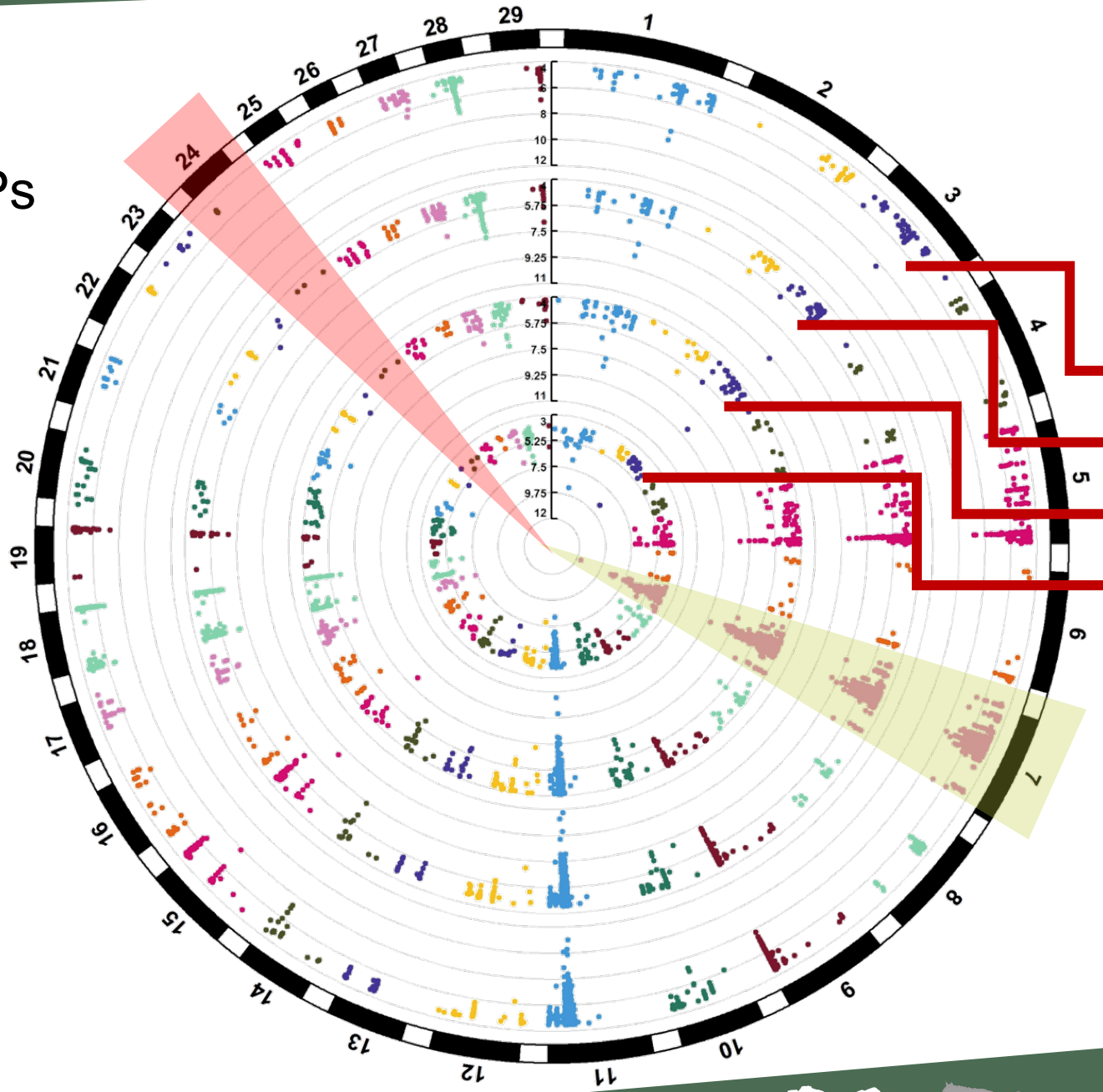
$$p\_value_{corrected_n} = adjust(p\_value_n, method = FDR, n = N)$$

$$Significant_{SNP_{unpruned}} = p\_value_{corrected_n} < 0.1$$

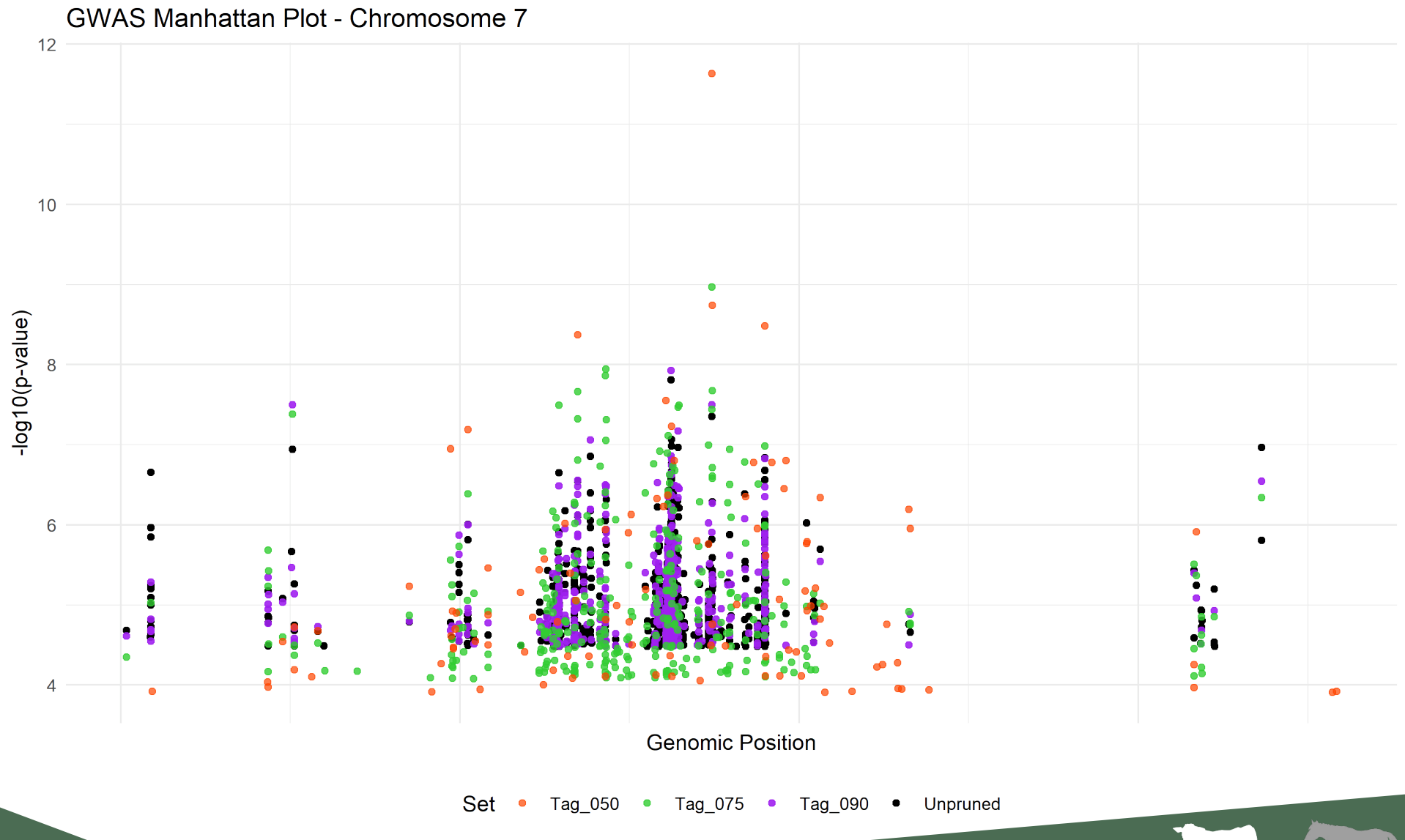
Results

Number of significant SNPs

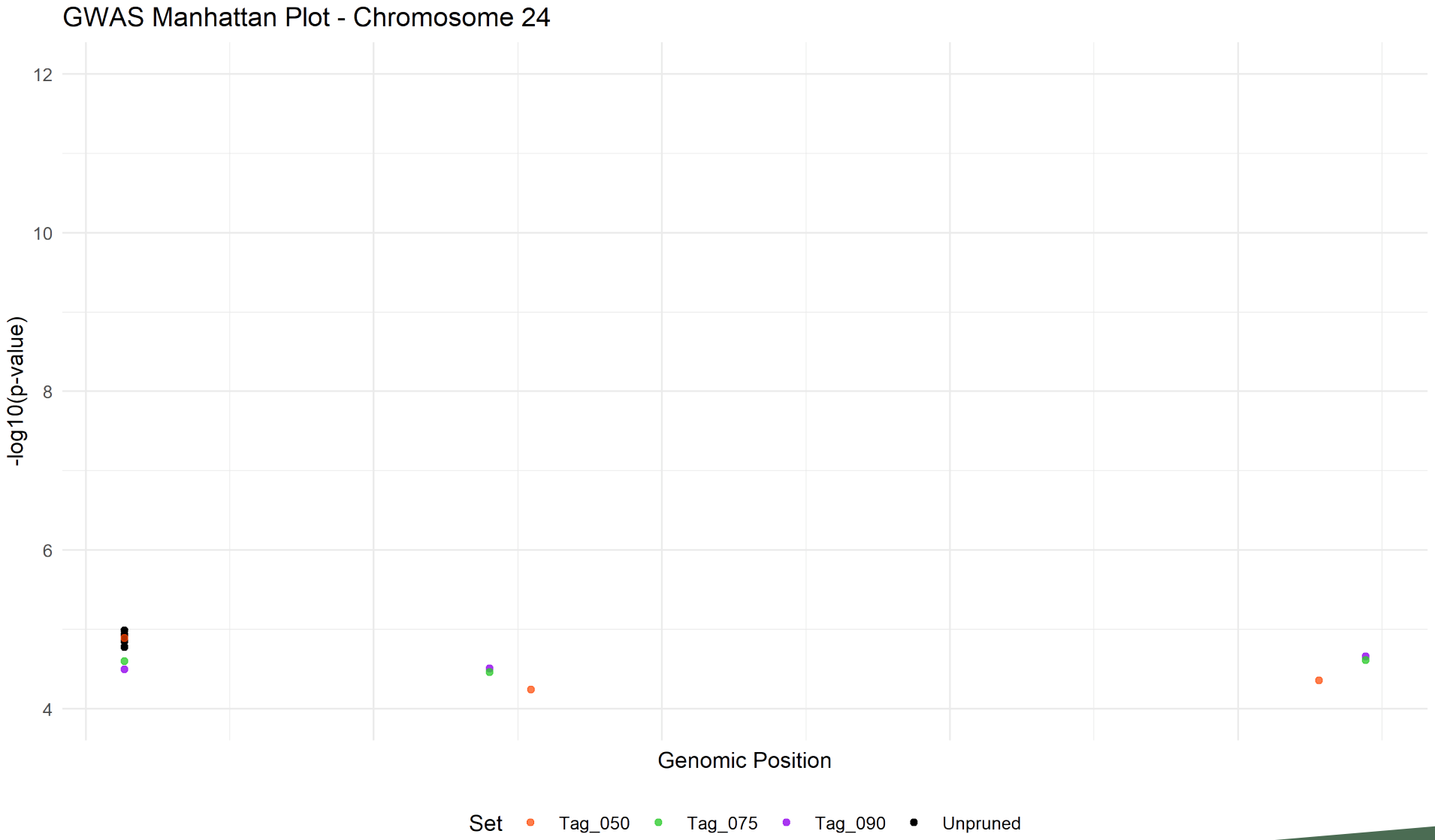
- Unpruned 3,582
- Tag 0.90 1,923
- Tag 0.75 1,118
- Tag 0.50 559



Results


GWAS Manhattan Plot - Chromosome 7
12
10
8
6
4
-log10(p-value)
Genomic Position
Set
Tag_050
Tag_075
Tag_090
Unpruned

Results


GWAS Manhattan Plot - Chromosome 24
12
10
8
6
4
-log10(p-value)
Genomic Position
Set
Tag_050
Tag_075
Tag_090
Unpruned

Time and iteration

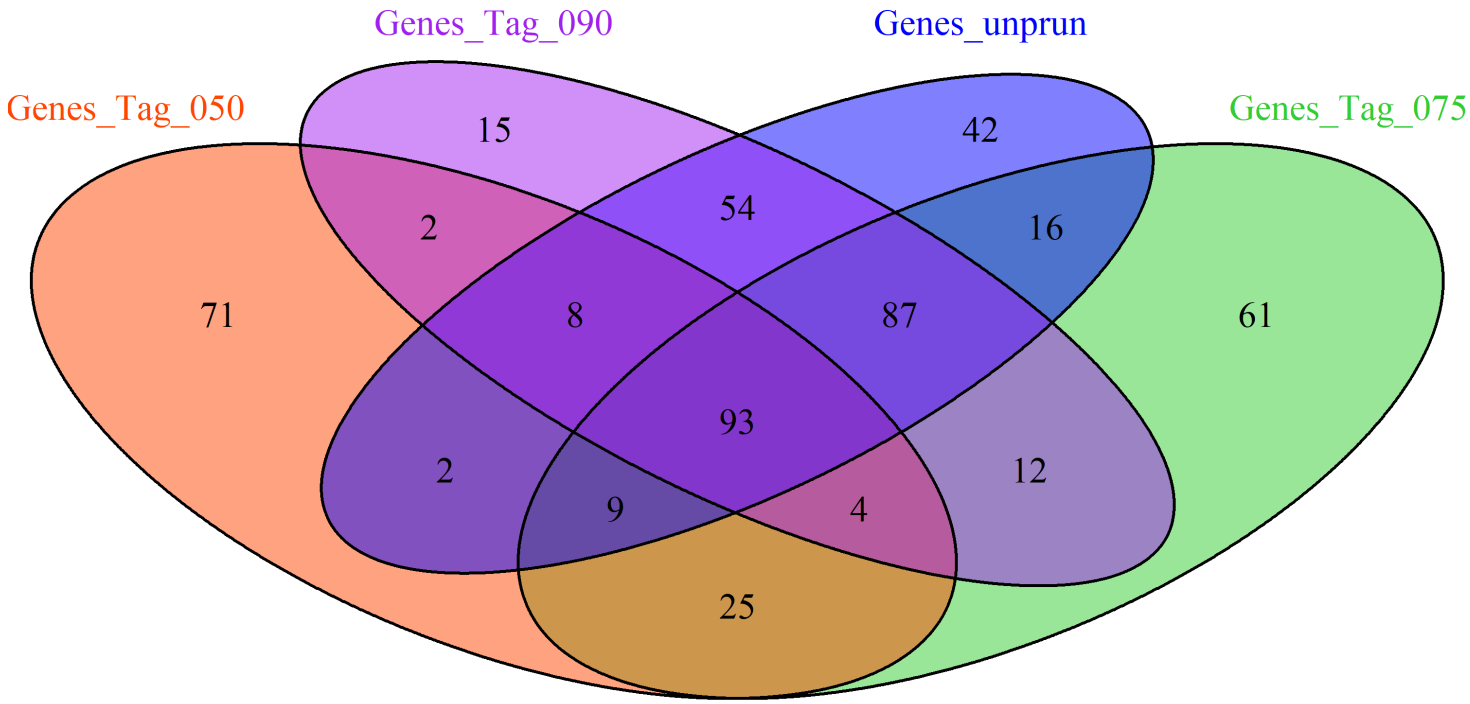
Scenarios	Wall clock time (min)			Number of iterations
	PLINK 1.9	MiXBLUP 3.0	Sum	
Unpruned	19	55	74	42
Tag 0.90	997	42	1 039	33
Tag 0.75	55	9	64	33
Tag 0.50	44	3	47	32

Conclusions

- Tag 0.50 data set provides better defined peaks
  - Exceptions exist
- Dimensionality reduction: wall clock time reduction from 74 mins to 47 mins
- SNP tagging allows for genomic noise reduction

Next steps

- Check variants of significant SNPs
- Validate the tag concept by moving from BPS to genes
- Define the optimal R^2 treshold



Thank You!

dawid.slomian@iz.edu.pl