Supplement 1:

MultiGWAS report for Full GWAS model with filters

Contents

1 Input Parameters
2 Best-ranked SNPs
2.1 Table of best-ranked SNPs
2.2 Venn diagram of best-ranked SNPs
3 Significative SNPs
3.1 Table of significative SNPs
3.2 Venn diagram of significative SNPs
4 Manhattan and QQ plots
5 Chord diagrams for SNPs by chromosome
6 Profiles for common significative SNPs

MultiGWAS is a tool to do GWAS analysis in tetraploid organisms by executing in parallel and integrating the results from four existing GWAS software: two available for polyploids (GWASpoly and SHEsis) and two frequently used for diploids (PLINK and TASSEL).

MultiGWAS executes two types of GWAS models: Full and Naive. The Full model with control for population structure and individual relatedness and the Naive model without any control. In both models, users can apply different control quality filters for the genomic data.

This report, created by the MultiGWAS tool, is a summary of the input parameters used to run the tool and its main outputs, including (1) Score tables with detailed information on the associations for each tool. (2) Venn diagrams of shared SNPs among the four tools. (3) Heatmaps of significative SNP profiles among the four tools. (4) Manhattan and QQ plots for the association found by each tool. And (5) Chord diagrams for the chromosomes vs. SNP by each tool.

1 Input Parameters

MultiGWAS uses as input a configuration file where the user specifies the input genomic data (genotype and phenotype files) along with different values for parameters used by the MultiGWAS tool and by the other four GWAS software (GWASpoly, SHEsis, PLINK, and TASSEL).

The following table shows the current input parameters specified by the user in the configuration file:

PARAMETER	VALUE
Ploidy (4 or 2)	4
Genotype filename	example-genotype-tetra.csv
Phenotype filename	example-phenotype.csv
Significance level (Genome-wide significance level)	0.05
Correction method (Bonferroni or FDR)	Bonferroni
GWAS model (Full or Naive)	Full
nBest (Number of best-ranked SNPs to be reported)	10
Filtering (TRUE or FALSE)	TRUE
MIND Filter (Individual with missing genotype)	0.1
GENO Filter (SNPs with missing genotype)	0.1
MAF Filter (Minor allele frequency)	0.01
HWE Filter (Hardy-Weinberg test)	1e-10
GWAS Tools	GWASpoly SHEsis PLINK TASSEL

2 Best-ranked SNPs

This section shows a table and a Venn diagram for the best ranked SNPs (For this report, N=10). The configuration file allows defining the parameter N.

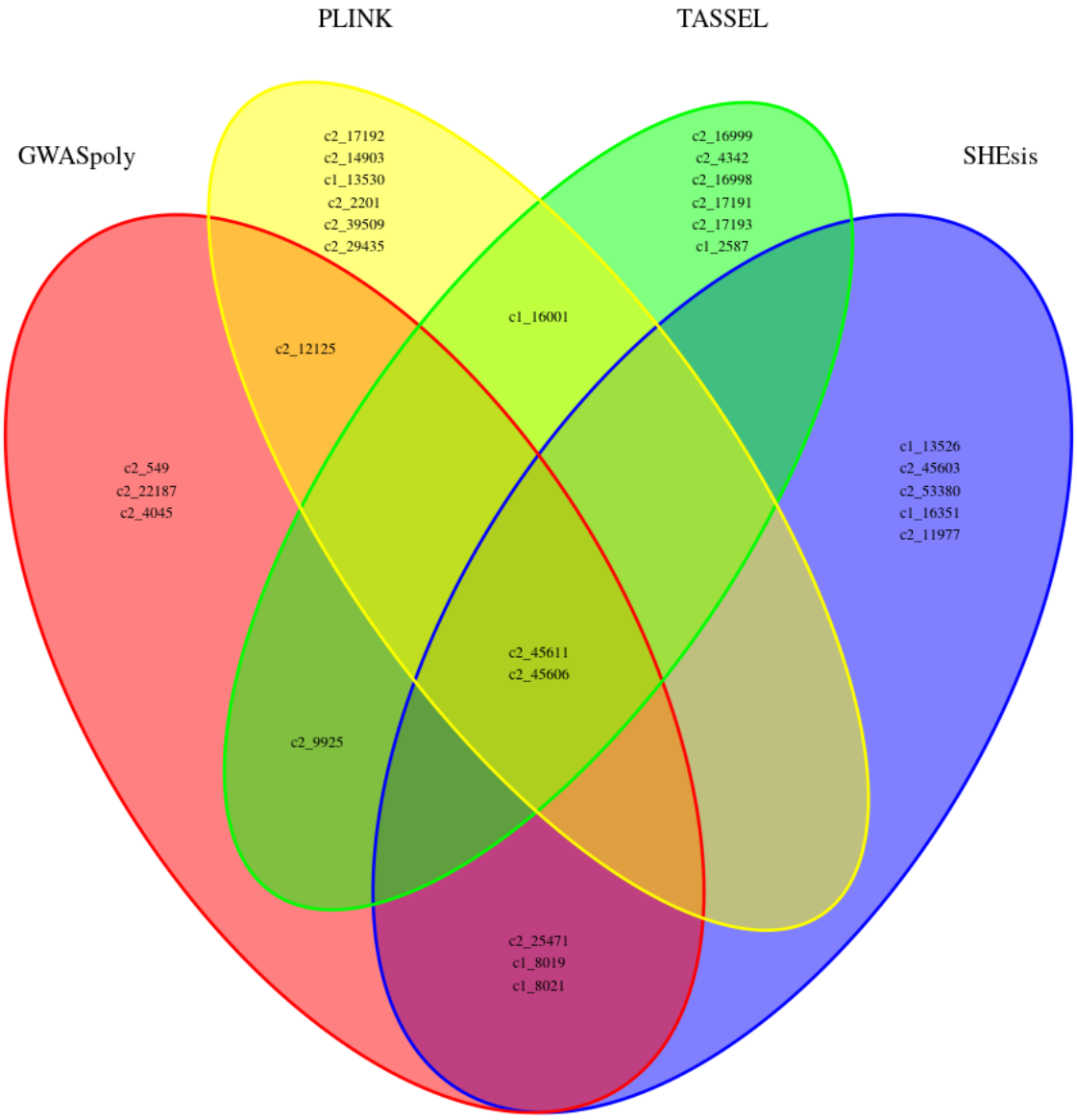
2.1 Table of best-ranked SNPs

The following table shows the associations between SNPs and the trait as transformed p-values to scores as -log 10 (p-value) sorted from the highest to the lowest scores based on each tool analysis and the model selected (Full or Naive). Each SNP has its associated information, such as the chromosome, position in the genome, and name. Also, based on the threshold calculated by a multiple testing correction (i.e., Bonferroni or FDR), the table marks the significance of each SNP as a TRUE or FALSE.

TOOL	MODEL	GC	SNP	CHROM	POSITION	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	additive	0.994	c2_25471	10	48808404	0.000005	5.2800	4.4800	TRUE
GWASpoly	additive	0.994	c2_45611	10	48203431	0.000009	5.0700	4.4800	TRUE
GWASpoly	additive	0.994	c1_8019	10	48863165	0.000012	4.9300	4.4800	TRUE
GWASpoly	additive	0.994	c2_45606	10	48218826	0.000048	4.3200	4.4800	FALSE
GWASpoly	additive	0.994	c2_549	9	16527499	0.000562	3.2500	4.4800	FALSE
GWASpoly	additive	0.994	c2_9925	1	81410256	0.001698	2.7700	4.4800	FALSE
GWASpoly	additive	0.994	c1_8021	10	48862950	0.002188	2.6600	4.4800	FALSE
GWASpoly	additive	0.994	c2_12125	1	71450400	0.002291	2.6400	4.4800	FALSE
GWASpoly	additive	0.994	c2_22187	11	40777537	0.002344	2.6300	4.4800	FALSE
GWASpoly	additive	0.994	c2_4045	9	3710627	0.002754	2.5600	4.4800	FALSE
PLINK	additive	1.279	c1_16001	10	47539878	0.000115	3.9378	4.5248	FALSE
PLINK	additive	1.279	c2_17192	1	70472766	0.001373	2.8623	4.5248	FALSE
PLINK	additive	1.279	c2_12125	1	71450400	0.001766	2.7530	4.5248	FALSE
PLINK	additive	1.279	c2_45606	10	48218826	0.001915	2.7178	4.5248	FALSE
PLINK	additive	1.279	c2_45611	10	48203431	0.002313	2.6358	4.5248	FALSE
PLINK	additive	1.279	c2_14903	1	87322718	0.003177	2.4980	4.5248	FALSE
PLINK	additive	1.279	c1_13530	10	48149399	0.003183	2.4972	4.5248	FALSE
PLINK	additive	1.279	c2_2201	1	77738822	0.003268	2.4857	4.5248	FALSE
PLINK	additive	1.279	c2_39509	9	50174507	0.003291	2.4827	4.5248	FALSE
PLINK	additive	1.279	c2_29435	11	32948117	0.004548	2.3422	4.5248	FALSE
SHEsis	general	3.562	c1_8019	10	48863165	0.000000	10.9914	4.5248	TRUE
SHEsis	general	3.562	c1_13526	10	48020996	0.000000	10.0477	4.5248	TRUE
SHEsis	general	3.562	c2_45603	10	48073593	0.000000	9.8928	4.5248	TRUE
SHEsis	general	3.562	c2_25471	10	48808404	0.000000	9.6498	4.5248	TRUE
SHEsis	general	3.562	c2_53380	1	70371898	0.000000	8.9747	4.5248	TRUE
SHEsis	general	3.562	c2_45606	10	48218826	0.000000	8.1662	4.5248	TRUE
SHEsis	general	3.562	c1_16351	10	48761642	0.000000	8.0013	4.5248	TRUE
SHEsis	general	3.562	c2_45611	10	48203431	0.000000	7.7305	4.5248	TRUE
SHEsis	general	3.562	c1_8021	10	48862950	0.000000	7.5100	4.5248	TRUE
SHEsis	general	3.562	c2_11977	5	3515956	0.000000	6.7570	4.5248	TRUE
TASSEL	general	0.995	c2_16999	8	54838024	0.001094	2.9610	4.5248	FALSE
TASSEL	general	0.995	c2_4342	7	14924207	0.001212	2.9165	4.5248	FALSE
TASSEL	general	0.995	c2_16998	8	54838005	0.001389	2.8573	4.5248	FALSE
TASSEL	general	0.995	c2_17191	1	70474651	0.002149	2.6677	4.5248	FALSE
TASSEL	general	0.995	c2_9925	1	81410256	0.002246	2.6485	4.5248	FALSE
TASSEL	general	0.995	c1_16001	10	47539878	0.002354	2.6282	4.5248	FALSE
TASSEL	general	0.995	c2_45606	10	48218826	0.004562	2.3408	4.5248	FALSE
TASSEL	general	0.995	c2_45611	10	48203431	0.004952	2.3052	4.5248	FALSE
TASSEL	general	0.995	c2_17193	1	70472380	0.005492	2.2602	4.5248	FALSE
TASSEL	general	0.995	c1_2587	2	42179952	0.006113	2.2137	4.5248	FALSE

2.2 Venn diagram of best-ranked SNPs

The Venn diagram shows the best-ranked SNPs with high scores that are either unique or shared by the four software. Shared SNPs appear in the intersection areas of the diagram.



3 Significative SNPs

This section shows a table and a Venn diagram for the signifcative SNPs (score is above the significance threshold for each tool).

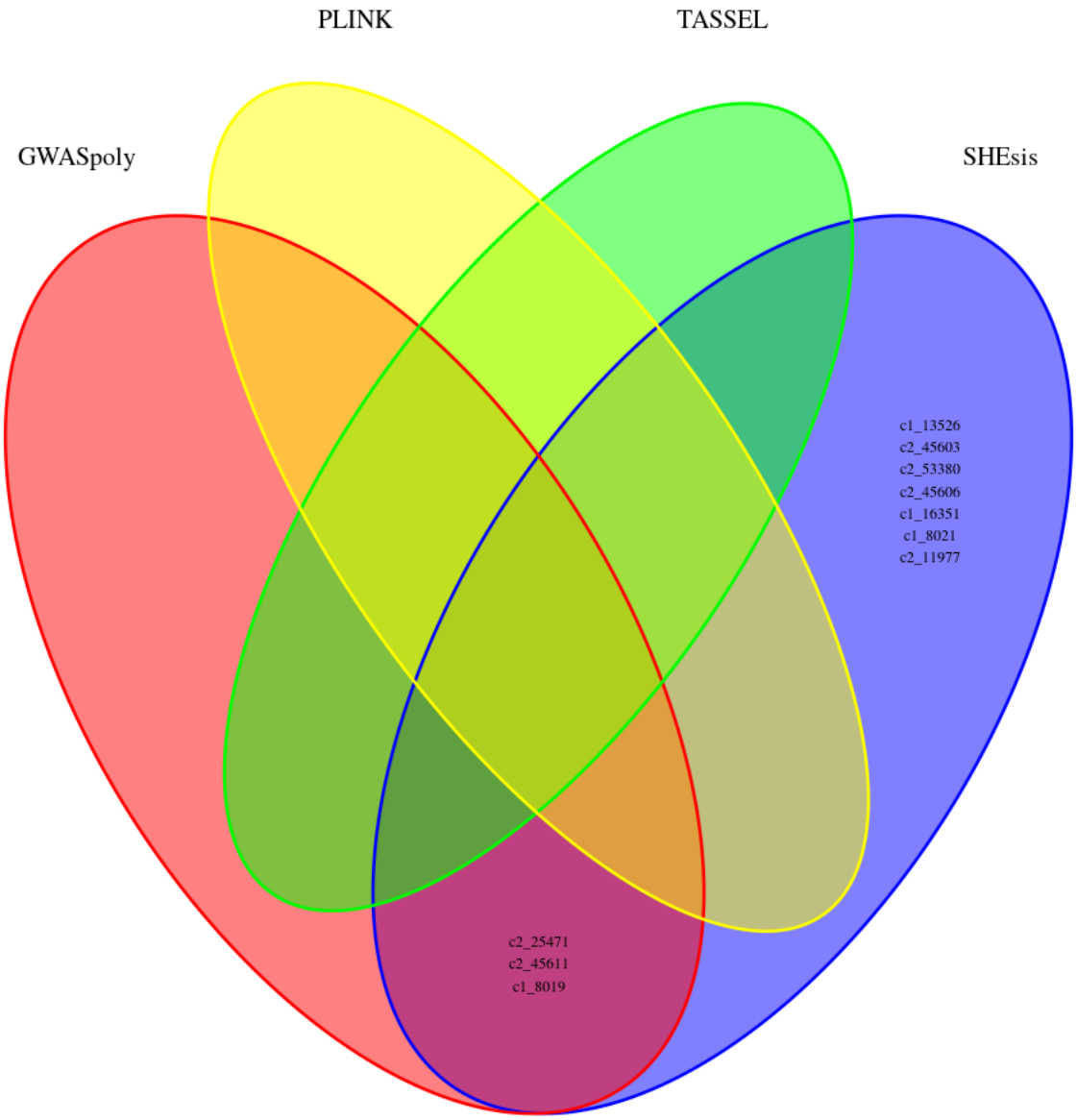
3.1 Table of significative SNPs

The following table shows the associations between SNPs and the trait as transformed p-values to scores as -log 10 (p-value) sorted from the highest to the lowest scores based on each tool analysis and the model selected (Full or Naive). Each SNP has its associated information, such as the chromosome, position in the genome, and name.

TOOL	MODEL	GC	SNP	CHROM	POSITION	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	additive	0.994	c2_25471	10	48808404	5.0e-06	5.2800	4.4800	TRUE
GWASpoly	additive	0.994	c2_45611	10	48203431	9.0e-06	5.0700	4.4800	TRUE
GWASpoly	additive	0.994	c1_8019	10	48863165	1.2e-05	4.9300	4.4800	TRUE
SHEsis	general	3.562	c1_8019	10	48863165	0.0e+00	10.9914	4.5248	TRUE
SHEsis	general	3.562	c1_13526	10	48020996	0.0e+00	10.0477	4.5248	TRUE
SHEsis	general	3.562	c2_45603	10	48073593	0.0e+00	9.8928	4.5248	TRUE
SHEsis	general	3.562	c2_25471	10	48808404	0.0e+00	9.6498	4.5248	TRUE
SHEsis	general	3.562	c2_53380	1	70371898	0.0e+00	8.9747	4.5248	TRUE
SHEsis	general	3.562	c2_45606	10	48218826	0.0e+00	8.1662	4.5248	TRUE
SHEsis	general	3.562	c1_16351	10	48761642	0.0e+00	8.0013	4.5248	TRUE
SHEsis	general	3.562	c2_45611	10	48203431	0.0e+00	7.7305	4.5248	TRUE
SHEsis	general	3.562	c1_8021	10	48862950	0.0e+00	7.5100	4.5248	TRUE
SHEsis	general	3.562	c2_11977	5	3515956	0.0e+00	6.7570	4.5248	TRUE

3.2 Venn diagram of significative SNPs

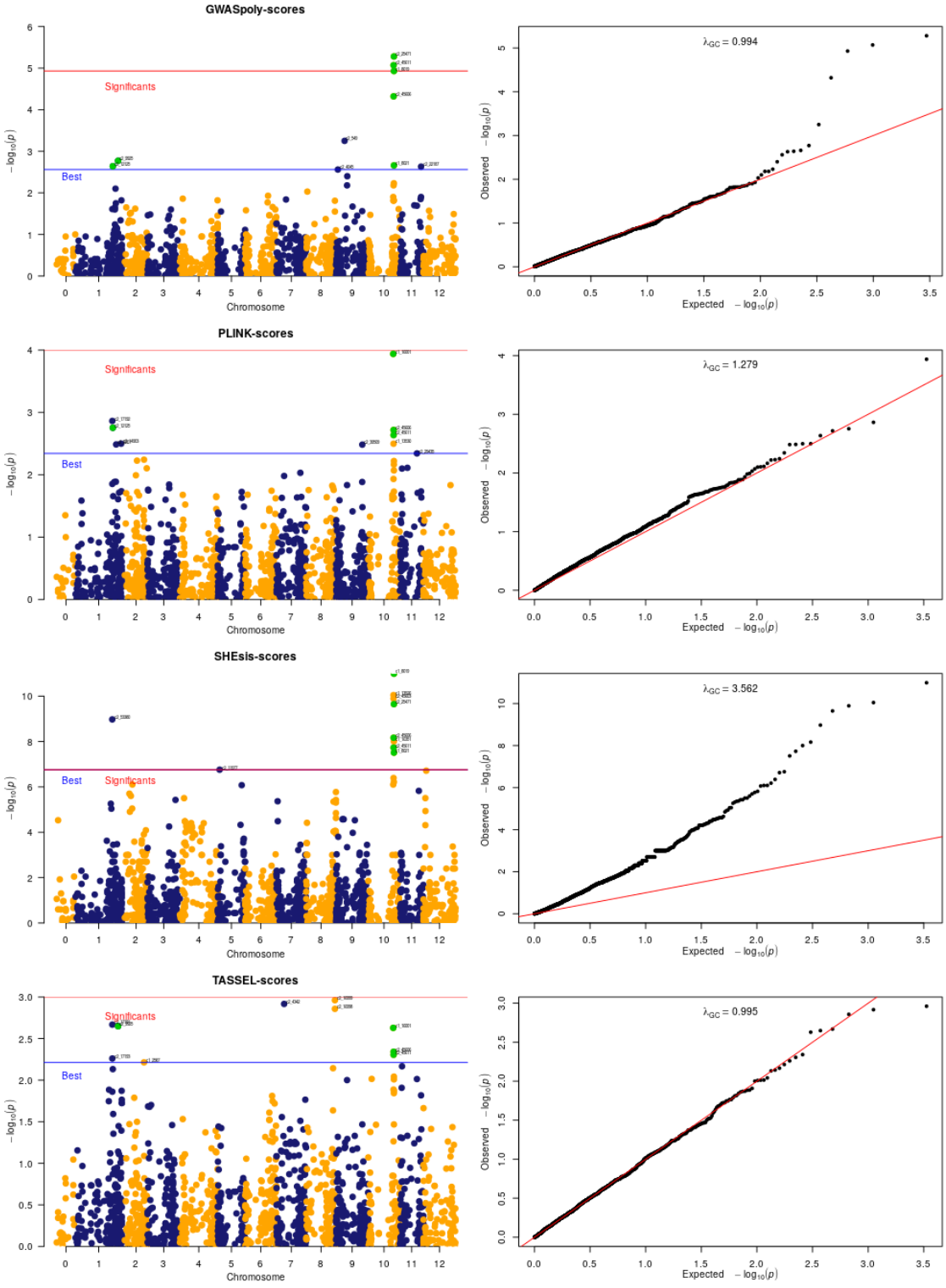
The Venn diagram shows the significative SNPs that are either unique or shared by the four software. Shared SNPs appear in the intersection areas of the diagram.



4 Manhattan and QQ plots

MultiGWAS uses classical Manhattan and Quantile–Quantile (QQ) plots to visually summarize GWAS results and identify both best-ranked and significative SNPs for each GWAS software. The Manhattan plot shows each SNPs distributed in the genetic location (x-axis) vs. the p-value (y-axis). The best-ranked SNPs are above the blue line and the significative SNPs above the blue line. In green are the SNPs that got high scores in more than one software.

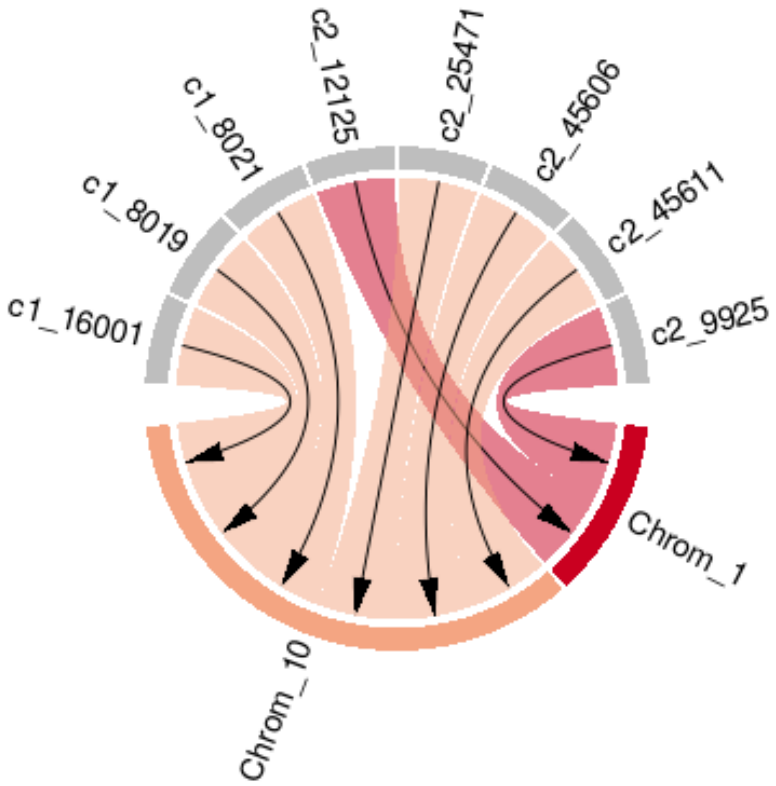
The QQ plot plots the observed (black dot lines) vs. expected -log 10 (p-value) (red line).



5 Chord diagrams for SNPs by chromosome

Chord diagrams are used in MultiGWAS to highlight nearby SNPs found on the same chromosome. These SNPs may have been identified simultaneously by multiple GWAS packets, which can be interpreted as a possible strong association, as in the Manhattan charts, where these associations form clean peaks with multiple SNPs displaying the same signal.

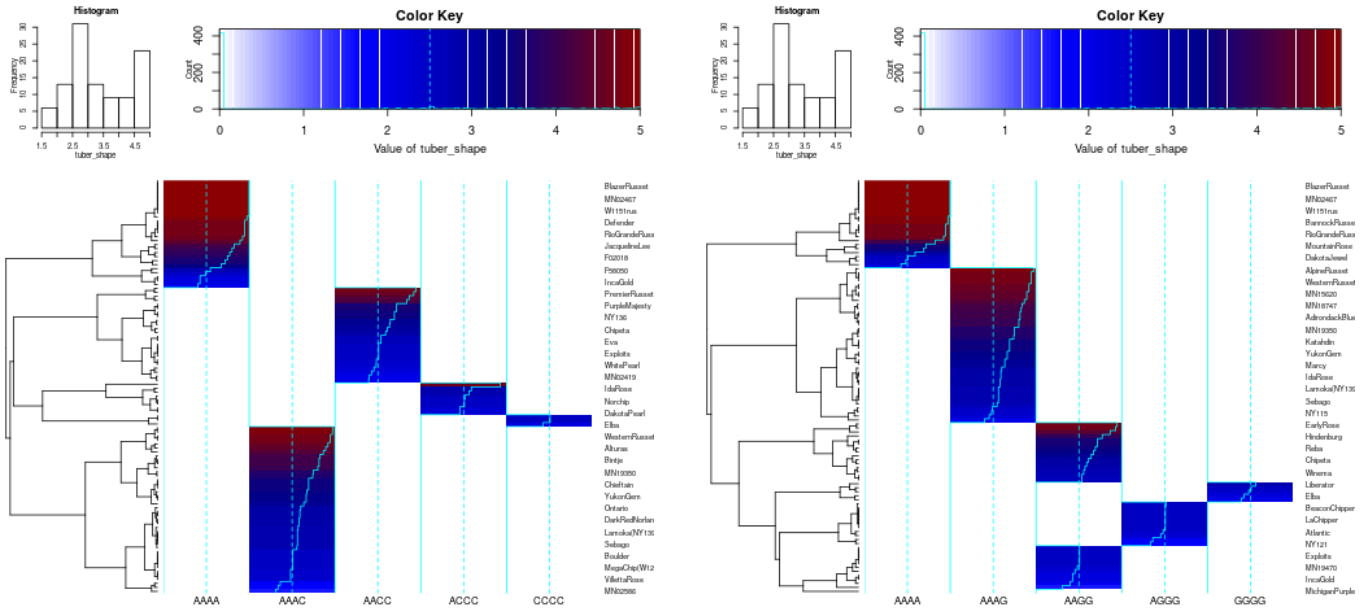
Next, we show the chord diagram for the shared SNPs shown in the Venn diagram of the best-ranked SNPs (Section 2.2).



6 Profiles for common significative SNPs

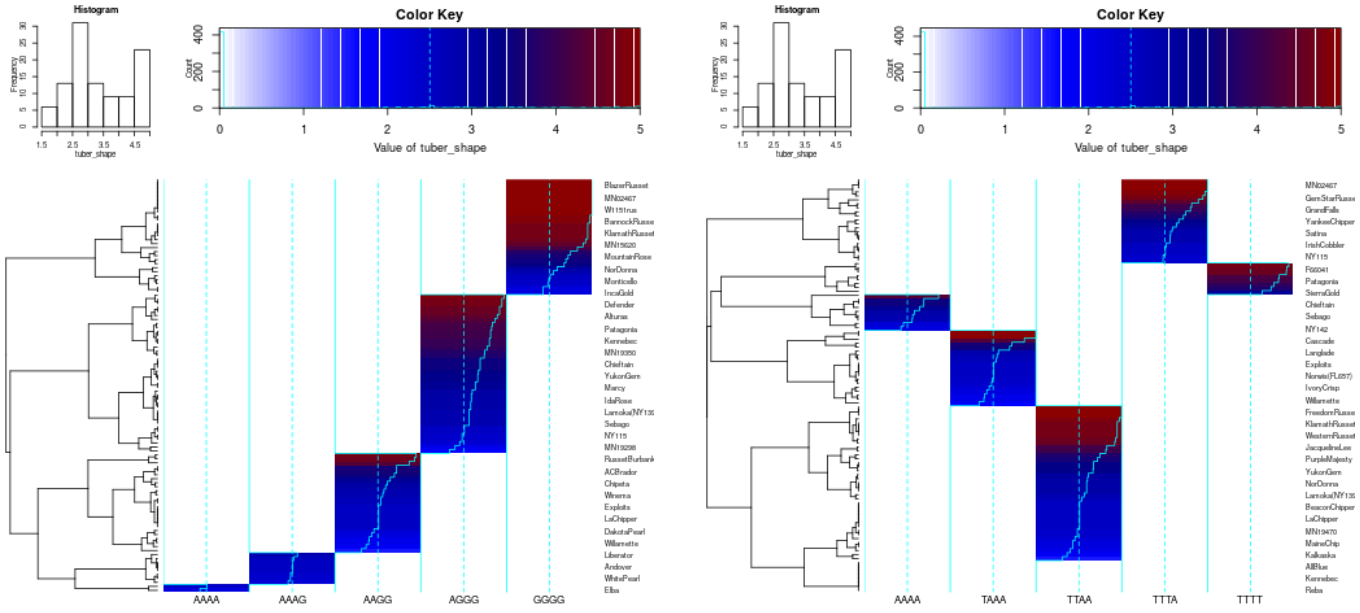
For the common significative SNPs, MultiGWAS provides a figure to visualize each trait by individuals (rows) and genotypes (columns). At the left, the individuals are grouped in a dendrogram by their genotype. At the right, there is the name or ID of each individual. At the bottom, the genotypes are ordered from left to right, starting from the major to the minor allele (i.e., AAAA, AAAB, AABB, ABBB, BBBB). At the top, there is a description of the trait based on a histogram of frequency (top left) and by an assigned color for each numerical phenotype value using a grayscale (top right). Thus, each individual appears as a colored line by its phenotype value on its genotype column. For each column, there is a solid cyan line with the mean of each column and a broken cyan line that indicates how far the cell deviates from the mean.

Because each multiGWAS report shows one specific trait at a time, the histogram and color key will remain the same for all the best-ranked SNPs.



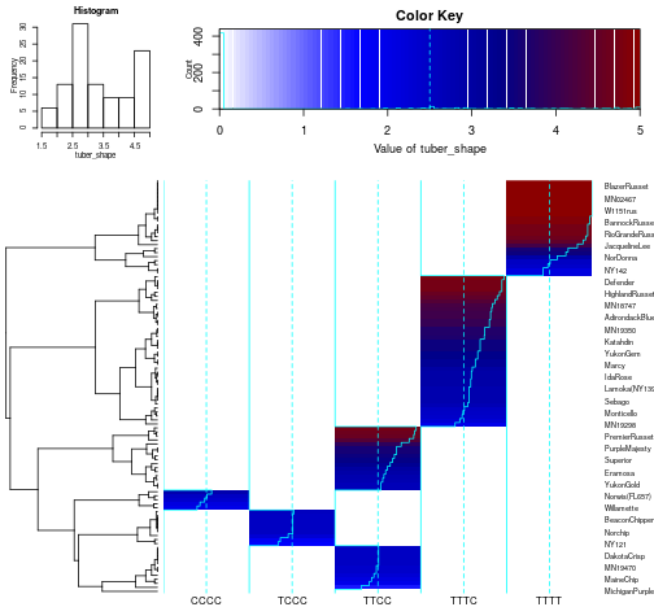
c1_16001

c1_8019

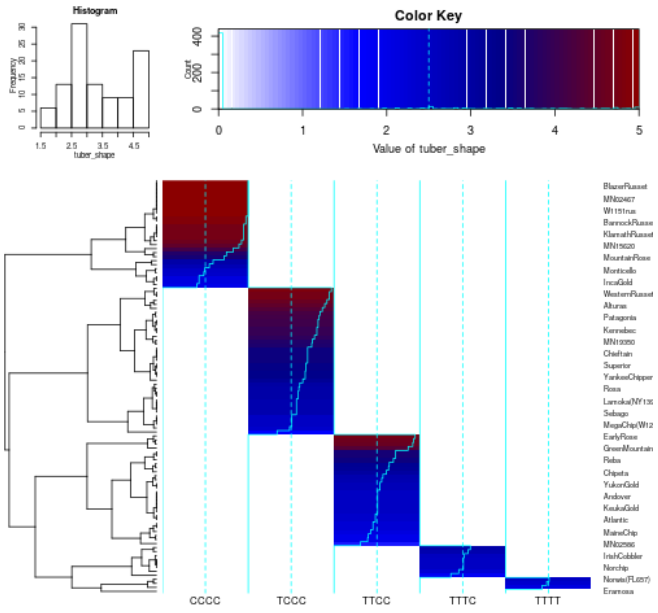


c1_8021

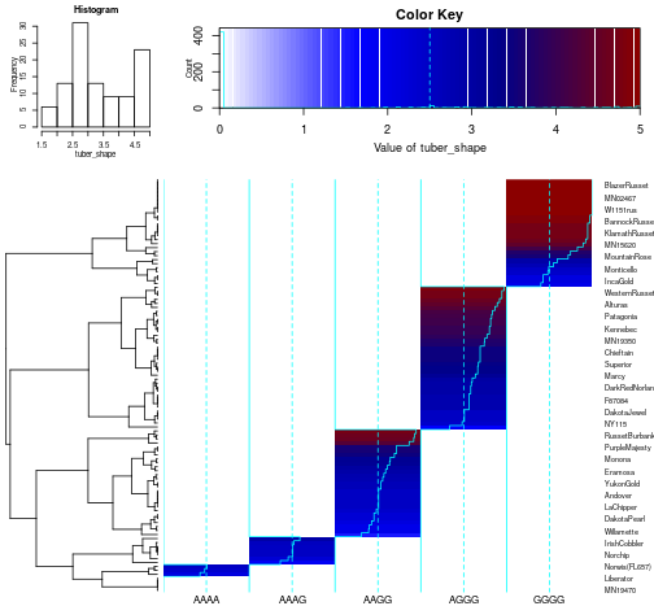
c2_12125



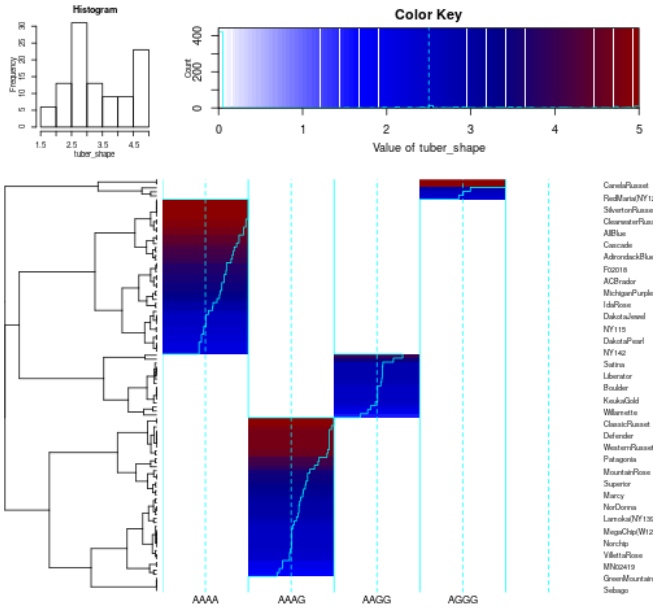
c2_25471



c2_45606



c2_45611



c2_9925