Supplementary Information
MultiGWAS: A tool for GWAS analysis on tetraploid organisms by integrating results of four GWAS software

L. Garreta, I. Cerón-Souza, M.R. Palacio, and P.H. Reyes-Herrera

March 20, 2020

Contents

1 Report for Full GWAS model **2**

2 Report for Naive GWAS model **11**

1 Report for Full GWAS model

MultiGWAS report for Full GWAS model

- 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 Profiles for common significative SNPs

This report, created by the MultiGWAS tool, is a summary of the input parameters used to run the tool and its main outputs. The outputs include the best-ranked and significative SNPs found for each tool. For each one, there is a score table, Venn diagram, and an SNP profile.

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
Genotype filename	example-genotype.tbl
Phenotype filename	example-phenotype.tbl
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)	7
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

2 Best-ranked SNPs

This section shows a table and a Venn diagram for the best N ranked SNPs (For this report N=7). The configuration file allows defining the parameter N.

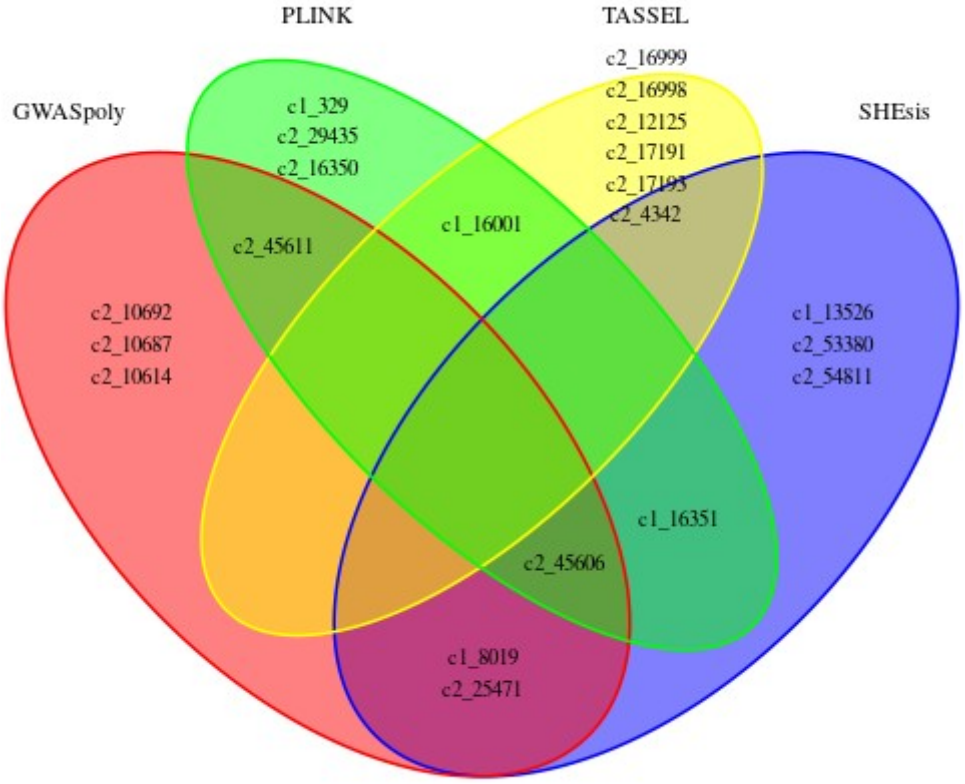
2.1 Table of best-ranked SNPs

The following table shows the associations between SNPs and the traits 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	CHROM	POSITION	SNP	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	Full	10	48863165	c1_8019	0.000017	4.780000	4.250000	TRUE
GWASpoly	Full	10	48808404	c2_25471	0.000027	4.570000	4.270000	TRUE
GWASpoly	Full	10	48203431	c2_45611	0.000044	4.360000	4.270000	TRUE
GWASpoly	Full	10	48218826	c2_45606	0.000021	4.680000	4.500000	TRUE
GWASpoly	Full	4	71591813	c2_10692	0.001175	2.930000	4.250000	FALSE
GWASpoly	Full	4	71592285	c2_10687	0.001175	2.930000	4.250000	FALSE
GWASpoly	Full	4	71827521	c2_10614	0.001175	2.930000	4.250000	FALSE
PLINK	Full	10	67293176	c1_16001	0.000187	1.769349	3.260071	FALSE
PLINK	Full	10	77351069	c1_329	0.000662	1.179470	3.301030	FALSE
PLINK	Full	11	51404231	c2_29435	0.000845	1.118849	3.255273	FALSE
PLINK	Full	10	69323144	c2_45611	0.001054	1.022917	3.255273	FALSE
PLINK	Full	2	41814861	c2_16350	0.001097	0.959793	3.301030	FALSE
PLINK	Full	10	69311500	c2_45606	0.001445	0.848906	3.292256	FALSE
PLINK	Full	10	69809843	c1_16351	0.002539	0.613066	3.283301	FALSE
SHEsis	Full	2	13697423	c1_8019	0.000000	9.471083	3.301030	TRUE
SHEsis	Full	1	30837971	c1_13526	0.000000	8.450065	3.292256	TRUE
SHEsis	Full	5	46046095	c2_53380	0.000000	8.240929	3.260071	TRUE
SHEsis	Full	3	39255236	c2_25471	0.000000	7.824082	3.292256	TRUE
SHEsis	Full	5	49804489	c2_54811	0.000000	6.963331	3.269513	TRUE
SHEsis	Full	1	69809843	c1_16351	0.000000	6.024734	3.283301	TRUE
SHEsis	Full	4	69311500	c2_45606	0.000000	5.955695	3.292256	TRUE
TASSEL	Full	8	54838024	c2_16999	0.000247	3.607621	3.894316	FALSE
TASSEL	Full	8	54838005	c2_16998	0.000329	3.482989	3.894316	FALSE
TASSEL	Full	1	71450400	c2_12125	0.003287	2.483226	3.894316	FALSE
TASSEL	Full	1	70474651	c2_17191	0.003548	2.449995	3.894316	FALSE
TASSEL	Full	1	70472380	c2_17193	0.005137	2.289293	3.894316	FALSE
TASSEL	Full	10	47539878	c1_16001	0.001230	2.910131	4.551206	FALSE
TASSEL	Full	7	14924207	c2_4342	0.001320	2.879298	4.551206	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).

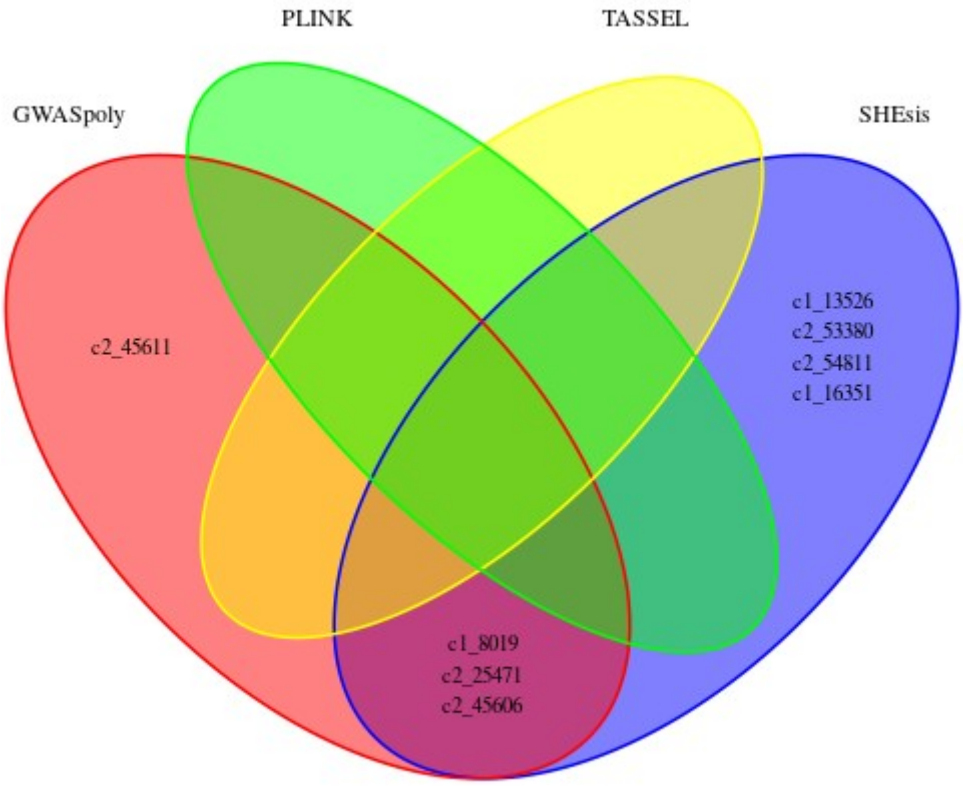
3.1 Table of significative SNPs

The following table shows the associations between SNPs and the traits 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	POSITION	SNP	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	Full	48863165	c1_8019	1.7e-05	4.780000	4.250000	TRUE
GWASpoly	Full	48808404	c2_25471	2.7e-05	4.570000	4.270000	TRUE
GWASpoly	Full	48203431	c2_45611	4.4e-05	4.360000	4.270000	TRUE
GWASpoly	Full	48218826	c2_45606	2.1e-05	4.680000	4.500000	TRUE
SHEsis	Full	13697423	c1_8019	0.0e+00	9.471083	3.301030	TRUE
SHEsis	Full	30837971	c1_13526	0.0e+00	8.450065	3.292256	TRUE
SHEsis	Full	46046095	c2_53380	0.0e+00	8.240929	3.260071	TRUE
SHEsis	Full	39255236	c2_25471	0.0e+00	7.824082	3.292256	TRUE
SHEsis	Full	49804489	c2_54811	0.0e+00	6.963331	3.269513	TRUE
SHEsis	Full	69809843	c1_16351	0.0e+00	6.024734	3.283301	TRUE
SHEsis	Full	69311500	c2_45606	0.0e+00	5.955695	3.292256	TRUE

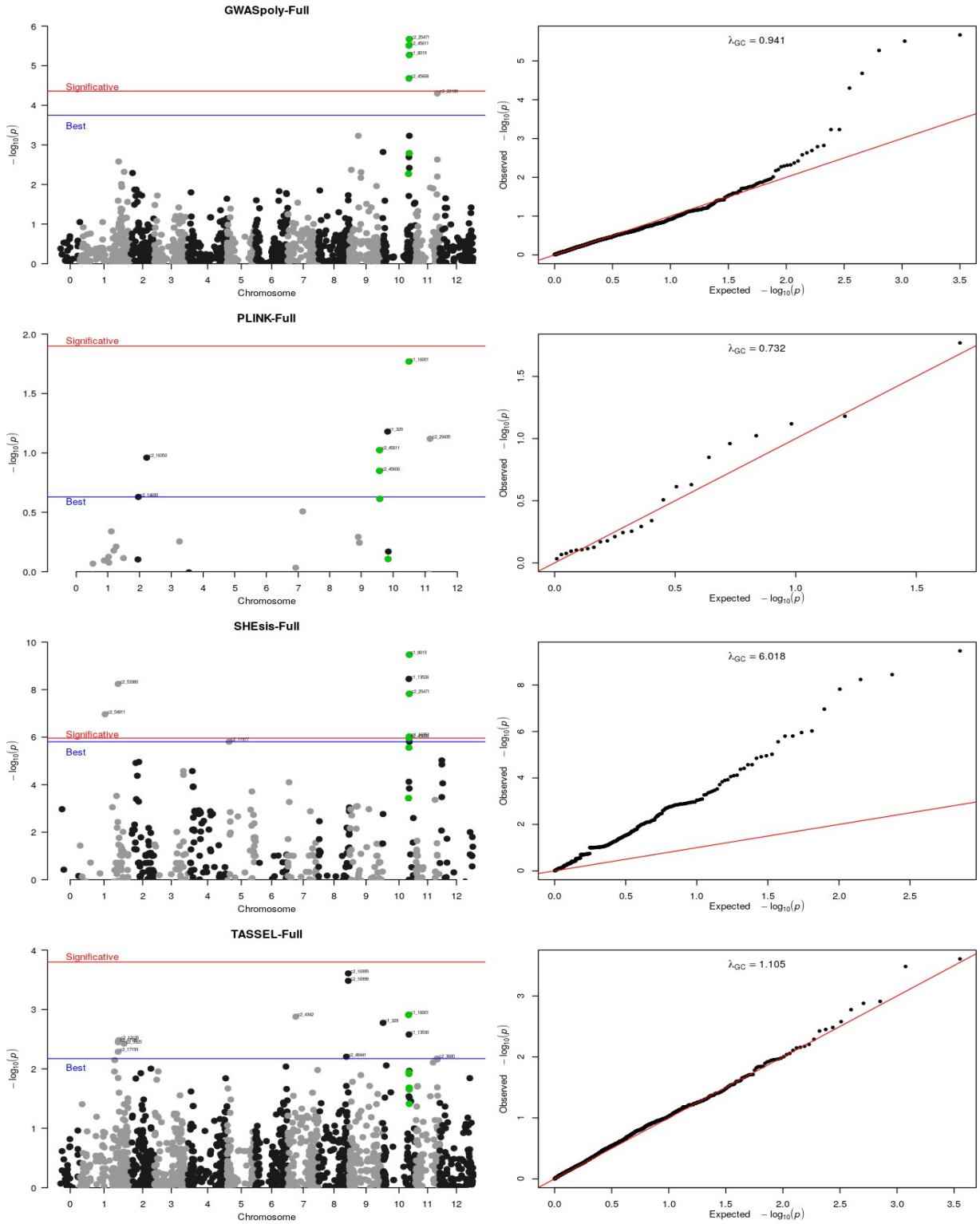
3.2 Venn diagram of significative SNPs

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

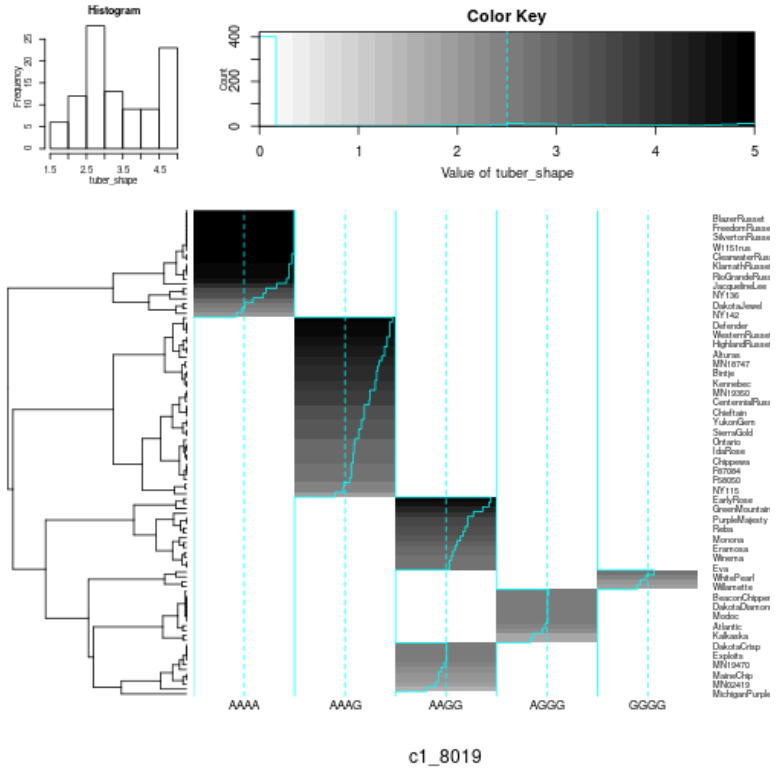
MultiGWAS uses classical Manhattan and Quantile–Quantile (QQ) plots to visually summarize GWAS results and identify both best-ranked and significative SNPs for each GWAS software. The Manhattan plot shows each SNPs distributed in the genetic location (x-axis) vs. the p-value (y-axis). The best-ranked 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).

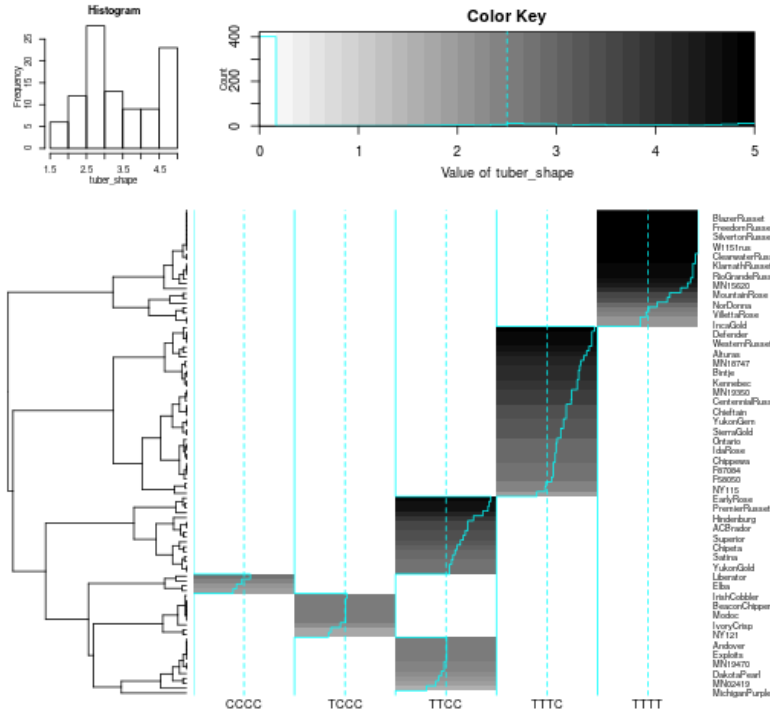


5 Profiles for common significative SNPs

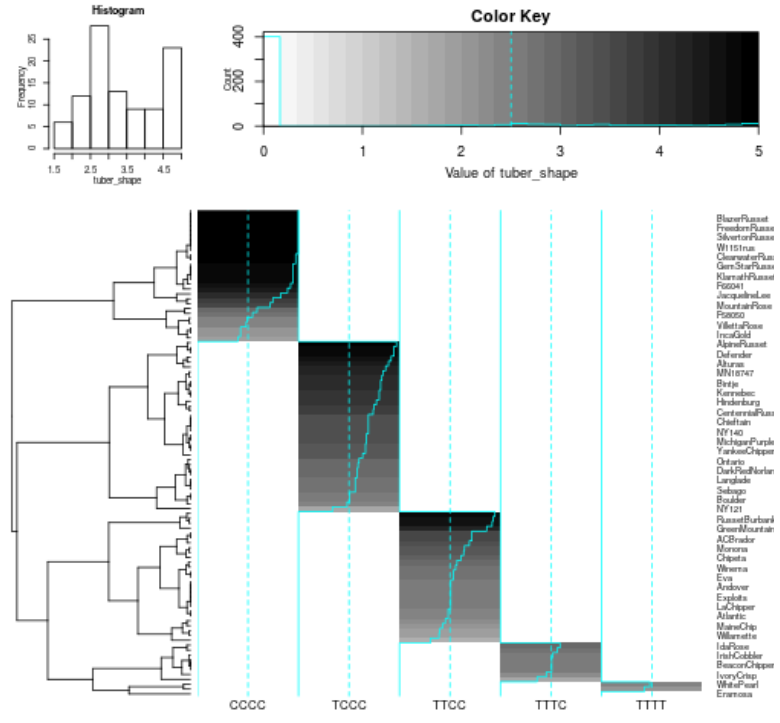
For the common significative SNPs, MultiGWAS provides a figure to visualize each trait by individuals (rows) and genotypes (columns). At the left, the individuals are grouped in a dendrogram by their genotype. At the right, there is the name or ID of each individual. At the bottom, the genotypes are 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.





c2_25471



c2_45606

2 Report for Naive GWAS model

MultiGWAS report for Naive GWAS model

- 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 Profiles for common significative SNPs

This report, created by the MultiGWAS tool, is a summary of the input parameters used to run the tool and its main outputs. The outputs include the best-ranked and significative SNPs found for each tool. For each one, there is a score table, Venn diagram, and an SNP profile.

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
Genotype filename	example-genotype.tbl
Phenotype filename	example-phenotype.tbl
Significance level (Genome-wide significance level)	0.05
Correction method (Bonferroni or FDR)	Bonferroni
GWAS model (Full or Naive)	Naive
NBest (Number of best-ranked SNPs to be reported)	7
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

2 Best-ranked SNPs

This section shows a table and a Venn diagram for the best N ranked SNPs (For this report N=7). The configuration file allows defining the parameter N.

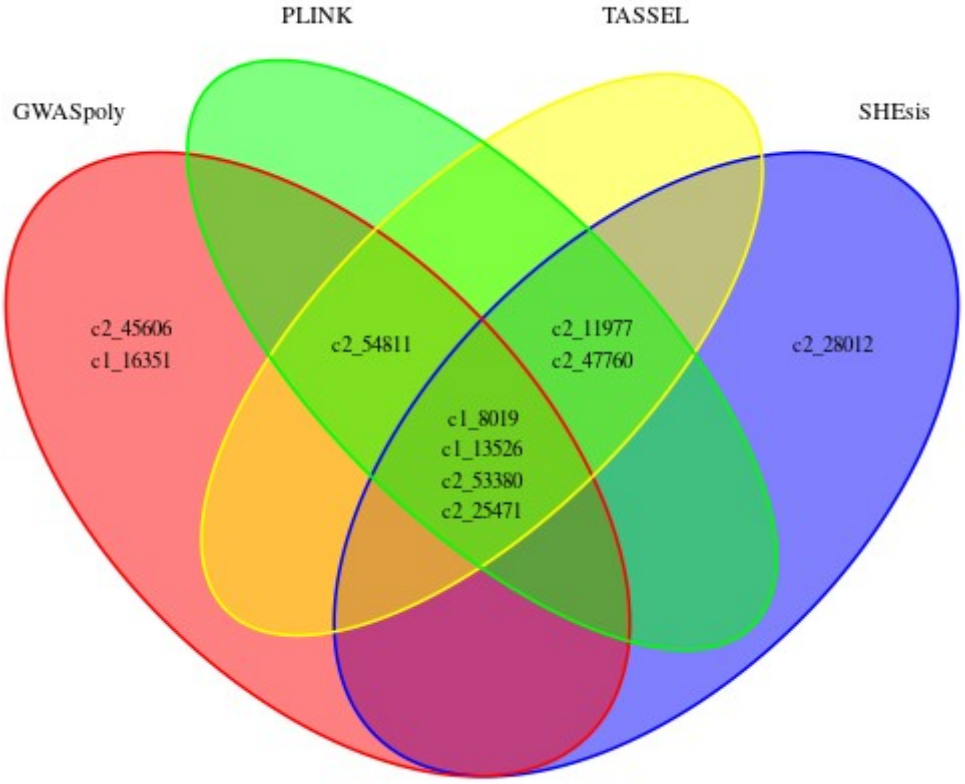
2.1 Table of best-ranked SNPs

The following table shows the associations between SNPs and the traits 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	CHROM	POSITION	SNP	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	Naive	10	48863165	c1_8019	0	11.560000	4.500000	TRUE
GWASpoly	Naive	10	48020996	c1_13526	0	10.610000	4.500000	TRUE
GWASpoly	Naive	1	70371898	c2_53380	0	10.370000	4.500000	TRUE
GWASpoly	Naive	10	48808404	c2_25471	0	9.890000	4.500000	TRUE
GWASpoly	Naive	1	46270954	c2_54811	0	8.540000	4.500000	TRUE
GWASpoly	Naive	10	48218826	c2_45606	0	7.970000	4.500000	TRUE
GWASpoly	Naive	10	48761642	c1_16351	0	7.970000	4.500000	TRUE
PLINK	Naive	5	32820618	c2_11977	0	6.848054	3.283301	TRUE
PLINK	Naive	10	13697423	c1_8019	0	5.040577	3.301030	TRUE
PLINK	Naive	1	46046095	c2_53380	0	4.967240	3.260071	TRUE
PLINK	Naive	2	72026885	c2_47760	0	4.771482	3.292256	TRUE
PLINK	Naive	1	49804489	c2_54811	0	4.691867	3.269513	TRUE
PLINK	Naive	10	30837971	c1_13526	0	4.560996	3.292256	TRUE
PLINK	Naive	10	39255236	c2_25471	0	4.536164	3.292256	TRUE
SHEsis	Naive	5	46046095	c2_53380	0	9.497204	3.313867	TRUE
SHEsis	Naive	2	13697423	c1_8019	0	8.031405	3.357935	TRUE
SHEsis	Naive	1	30837971	c1_13526	0	7.933190	3.330414	TRUE
SHEsis	Naive	2	32820618	c2_11977	0	6.565176	3.334454	TRUE
SHEsis	Naive	3	39255236	c2_25471	0	6.264880	3.342423	TRUE
SHEsis	Naive	4	72026885	c2_47760	0	5.941574	3.342423	TRUE
SHEsis	Naive	3	46475259	c2_28012	0	5.781570	3.318063	TRUE
TASSEL	Naive	5	3515956	c2_11977	0	8.830326	4.549984	TRUE
TASSEL	Naive	10	48863165	c1_8019	0	7.040577	4.549984	TRUE
TASSEL	Naive	1	70371898	c2_53380	0	6.926245	4.549984	TRUE
TASSEL	Naive	1	46270954	c2_54811	0	6.859964	4.549984	TRUE
TASSEL	Naive	2	20034673	c2_47760	0	6.762783	4.549984	TRUE
TASSEL	Naive	10	48020996	c1_13526	0	6.552175	4.549984	TRUE
TASSEL	Naive	10	48808404	c2_25471	0	6.527434	4.549984	TRUE

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).

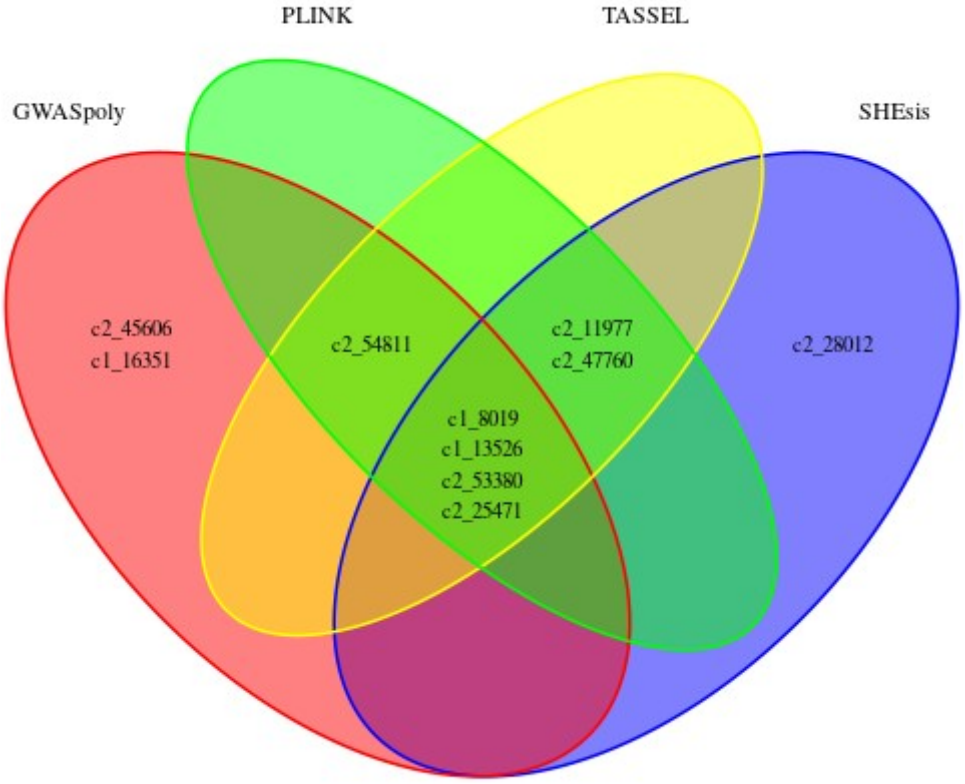
3.1 Table of significative SNPs

The following table shows the associations between SNPs and the traits 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	CHROM	POSITION	SNP	PVALUE	SCORE	THRESHOLD	SIGNIFICANCE
GWASpoly	Naive	10	48863165	c1_8019	0	11.560000	4.500000	TRUE
GWASpoly	Naive	10	48020996	c1_13526	0	10.610000	4.500000	TRUE
GWASpoly	Naive	1	70371898	c2_53380	0	10.370000	4.500000	TRUE
GWASpoly	Naive	10	48808404	c2_25471	0	9.890000	4.500000	TRUE
GWASpoly	Naive	1	46270954	c2_54811	0	8.540000	4.500000	TRUE
GWASpoly	Naive	10	48218826	c2_45606	0	7.970000	4.500000	TRUE
GWASpoly	Naive	10	48761642	c1_16351	0	7.970000	4.500000	TRUE
PLINK	Naive	5	32820618	c2_11977	0	6.848054	3.283301	TRUE
PLINK	Naive	10	13697423	c1_8019	0	5.040577	3.301030	TRUE
PLINK	Naive	1	46046095	c2_53380	0	4.967240	3.260071	TRUE
PLINK	Naive	2	72026885	c2_47760	0	4.771482	3.292256	TRUE
PLINK	Naive	1	49804489	c2_54811	0	4.691867	3.269513	TRUE
PLINK	Naive	10	30837971	c1_13526	0	4.560996	3.292256	TRUE
PLINK	Naive	10	39255236	c2_25471	0	4.536164	3.292256	TRUE
SHEsis	Naive	5	46046095	c2_53380	0	9.497204	3.313867	TRUE
SHEsis	Naive	2	13697423	c1_8019	0	8.031405	3.357935	TRUE
SHEsis	Naive	1	30837971	c1_13526	0	7.933190	3.330414	TRUE
SHEsis	Naive	2	32820618	c2_11977	0	6.565176	3.334454	TRUE
SHEsis	Naive	3	39255236	c2_25471	0	6.264880	3.342423	TRUE
SHEsis	Naive	4	72026885	c2_47760	0	5.941574	3.342423	TRUE
SHEsis	Naive	3	46475259	c2_28012	0	5.781570	3.318063	TRUE
TASSEL	Naive	5	3515956	c2_11977	0	8.830326	4.549984	TRUE
TASSEL	Naive	10	48863165	c1_8019	0	7.040577	4.549984	TRUE
TASSEL	Naive	1	70371898	c2_53380	0	6.926245	4.549984	TRUE
TASSEL	Naive	1	46270954	c2_54811	0	6.859964	4.549984	TRUE
TASSEL	Naive	2	20034673	c2_47760	0	6.762783	4.549984	TRUE
TASSEL	Naive	10	48020996	c1_13526	0	6.552175	4.549984	TRUE
TASSEL	Naive	10	48808404	c2_25471	0	6.527434	4.549984	TRUE

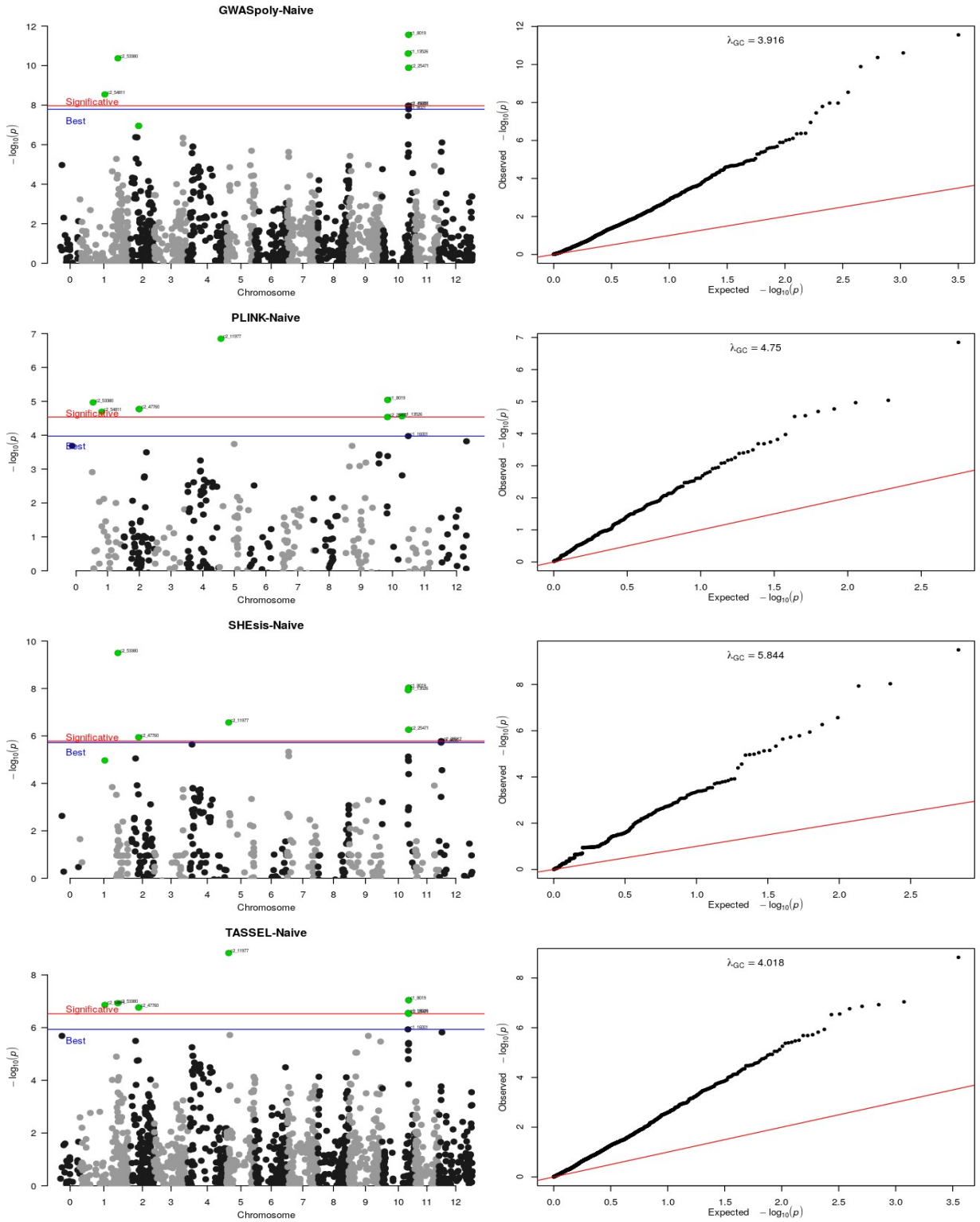
3.2 Venn diagram of significative SNPs

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

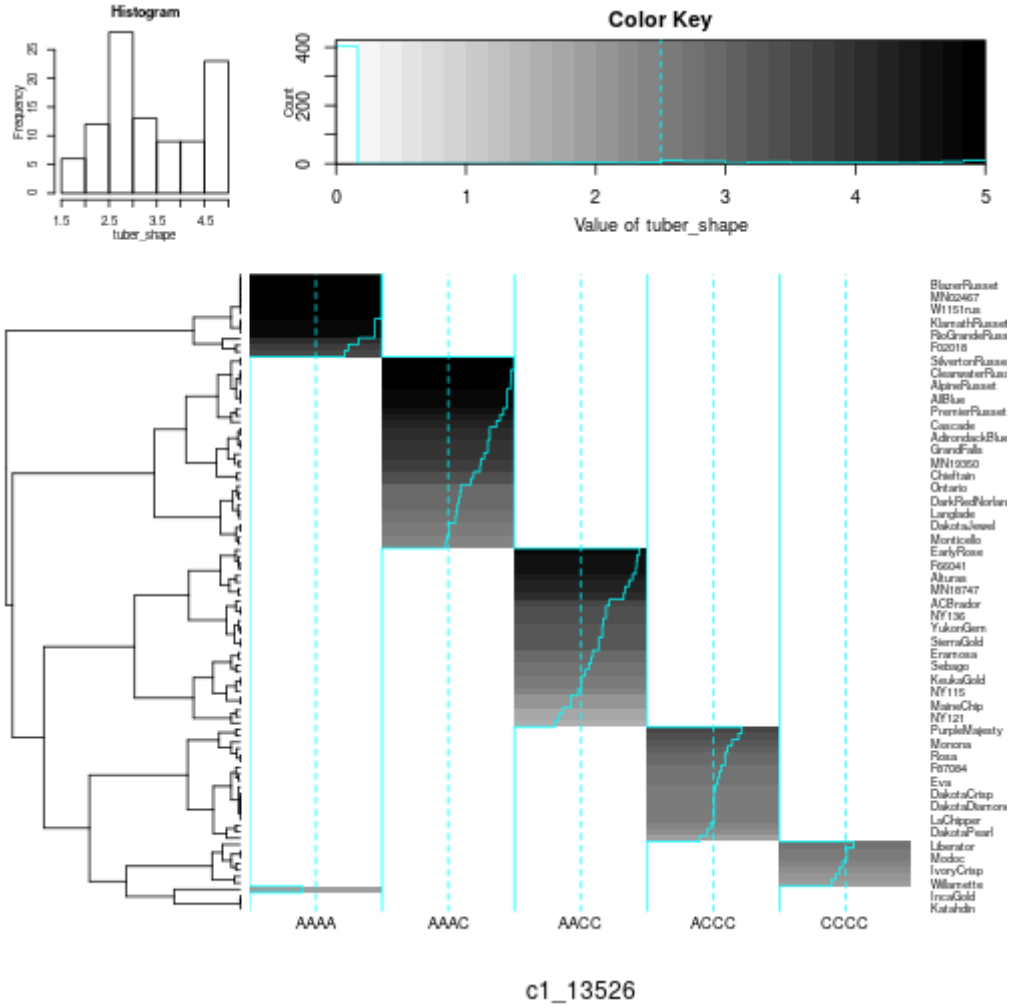
MultiGWAS uses classical Manhattan and Quantile–Quantile (QQ) plots to visually summarize GWAS results and identify both best-ranked and significative SNPs for each GWAS software. The Manhattan plot shows each SNPs distributed in the genetic location (x-axis) vs. the p-value (y-axis). The best-ranked 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).

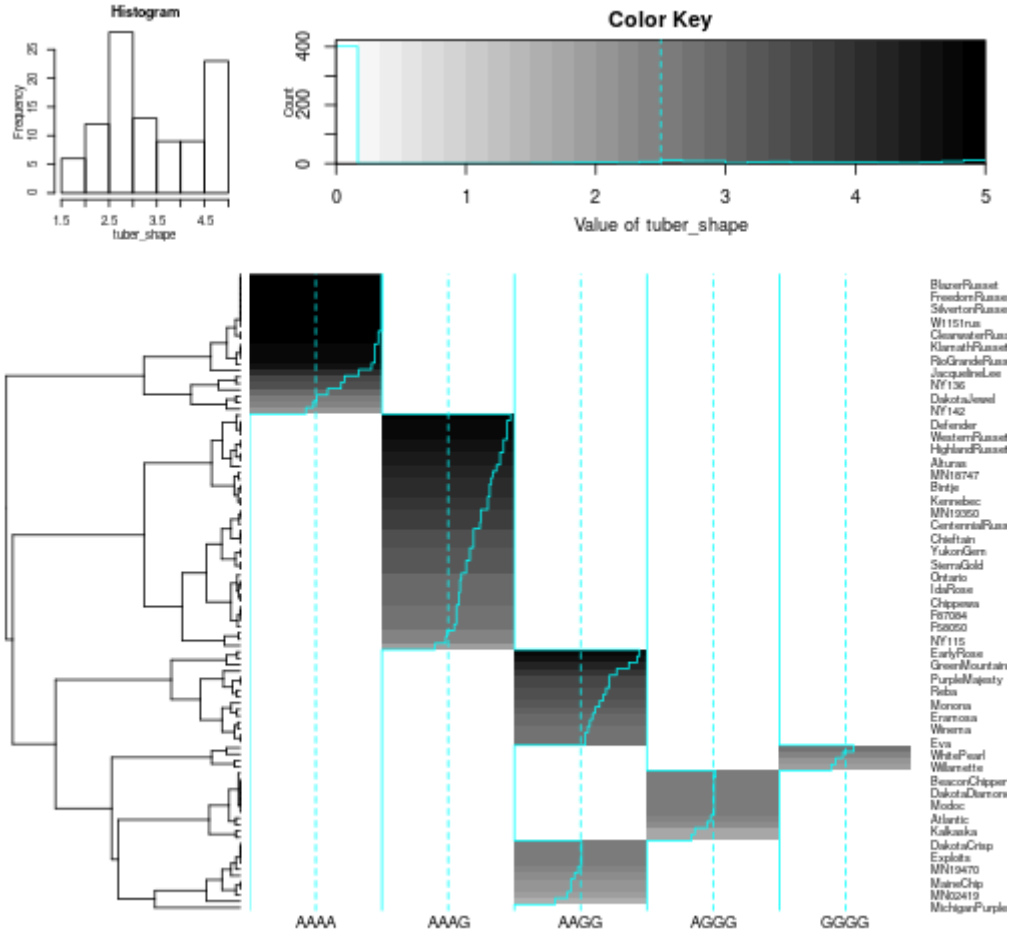


5 Profiles for common significative SNPs

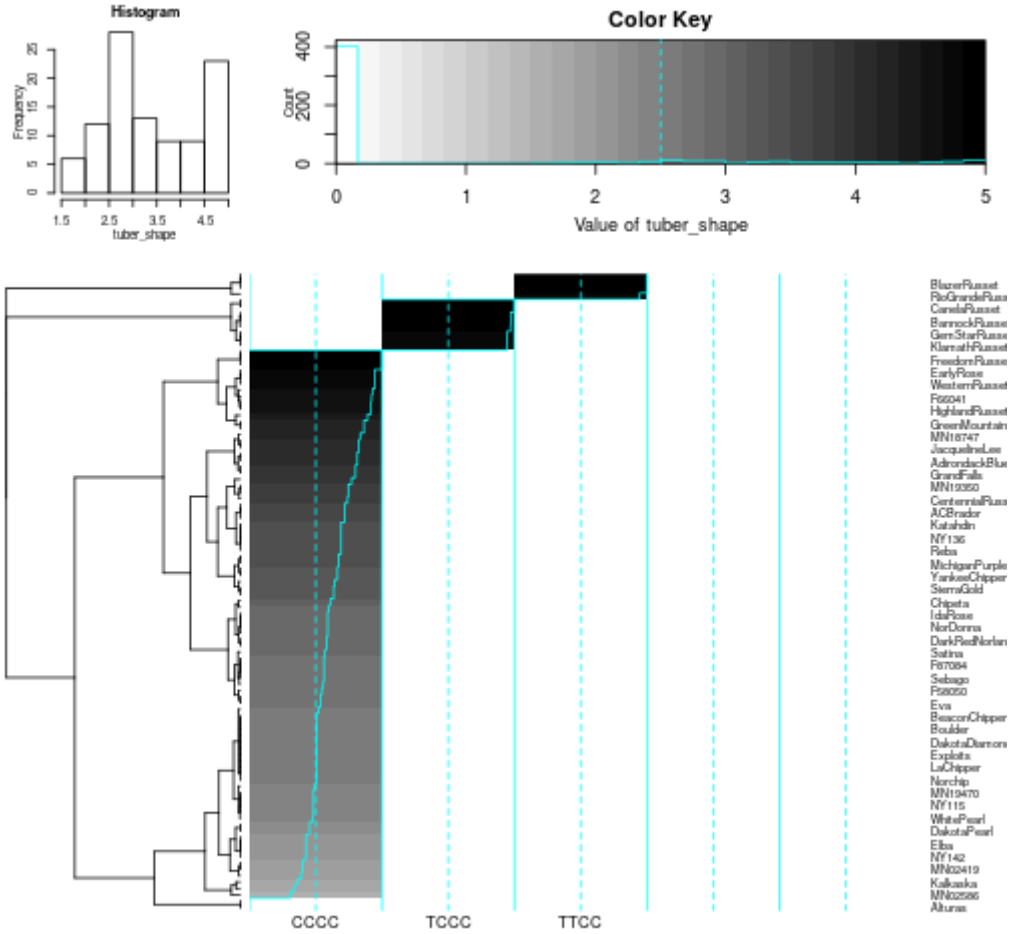
For the common significative SNPs, MultiGWAS provides a figure to visualize each trait by individuals (rows) and genotypes (columns). At the left, the individuals are grouped in a dendrogram by their genotype. At the right, there is the name or ID of each individual. At the bottom, the genotypes are 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.

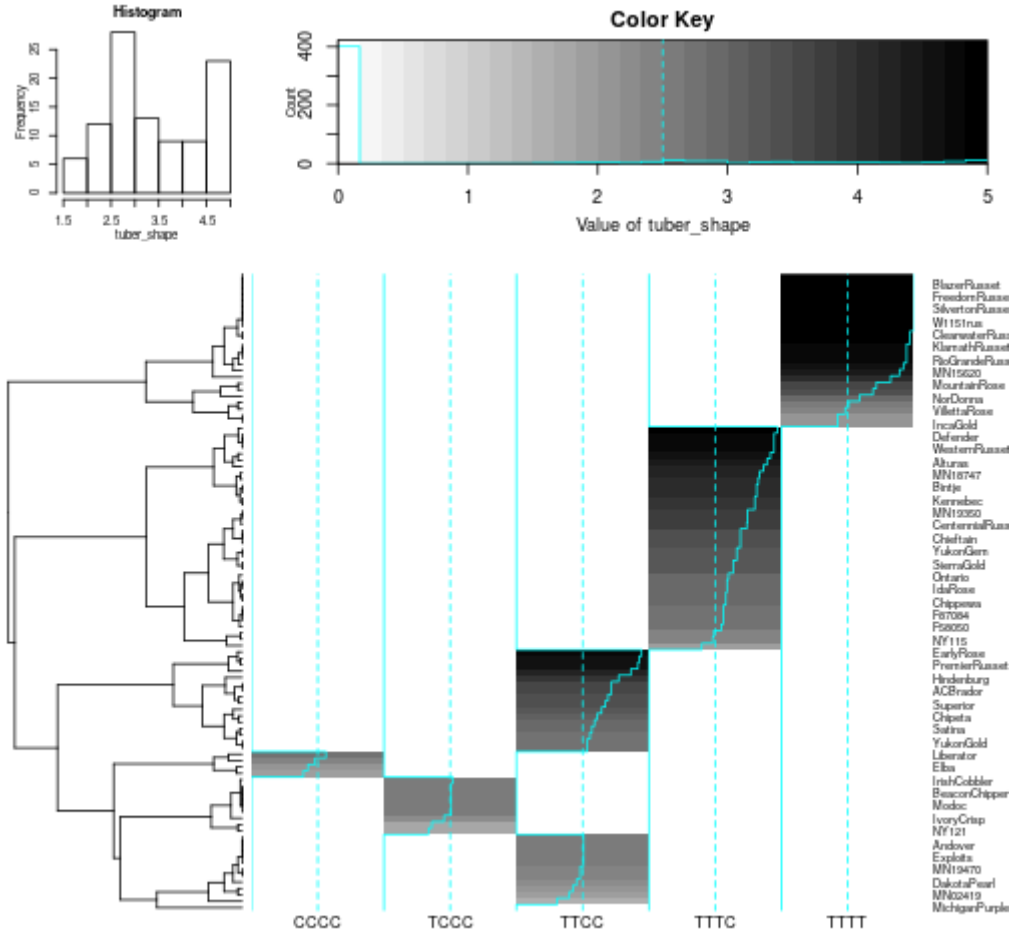



Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
AAAA
AAAG
AAGG
AGGG
GGGG

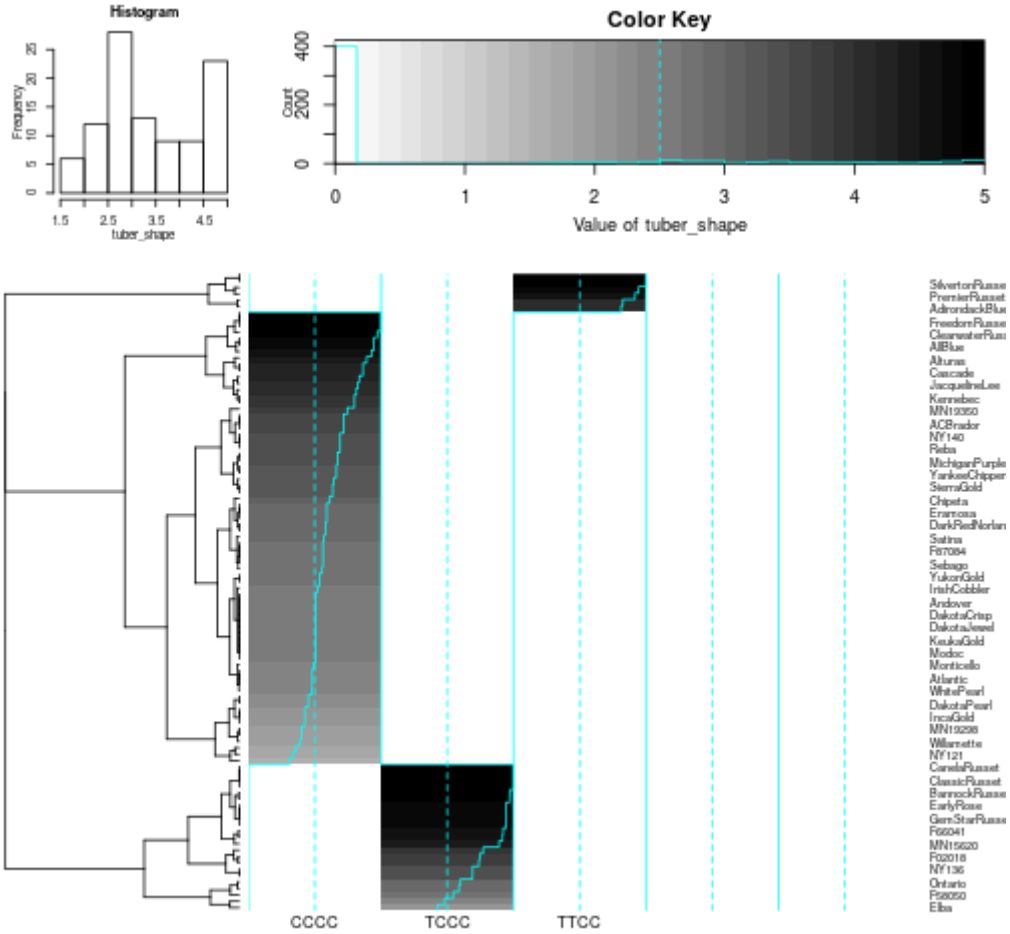
c1_8019


Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
CCCC
TCCC
TTCC

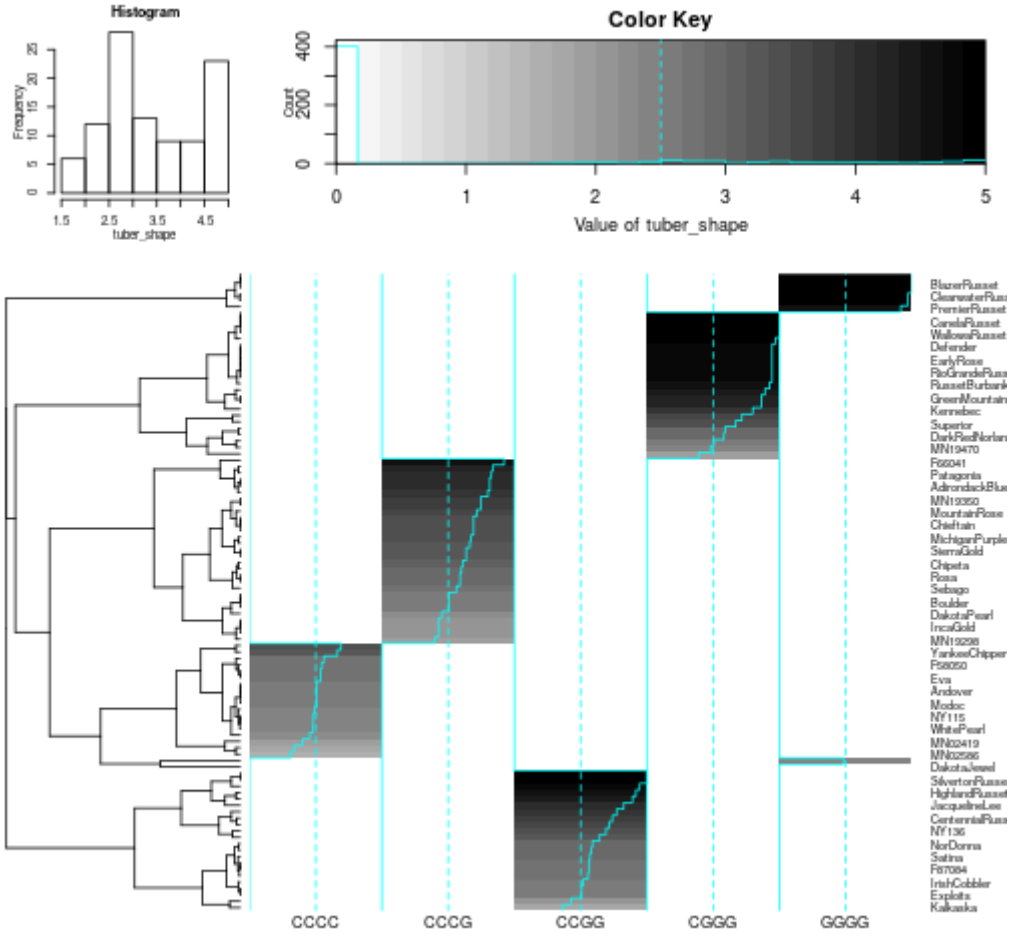
c2_11977


Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
BlazerRusset
FreedomRusset
SilvertonRusset
Wi151rus
ClearwaterRus
KlamathRusset
RioGrandeRuss
MN15620
MountainRose
NorDonna
YllettaRose
IncaGold
Defender
WesternRusset
Alturas
MN18747
Bintje
Kennebec
MN13350
CentennialRuss
Chieftain
YukonGem
SierraGold
Ontario
IdaRose
Chippewa
F87084
F58050
NY115
EarlyRose
PremierRusset
Hindenburg
ACBrador
Superior
Chipeta
Snowden
YukonGold
Liberator
Elba
IrishCobbler
BeaconChipper
Modoc
IvoryCrisp
NY121
Andover
Exploits
MN19470
DakotaPearl
MN02419
MichiganPurple
CCCC
TCCC
TTCC
TTTC
TTTT

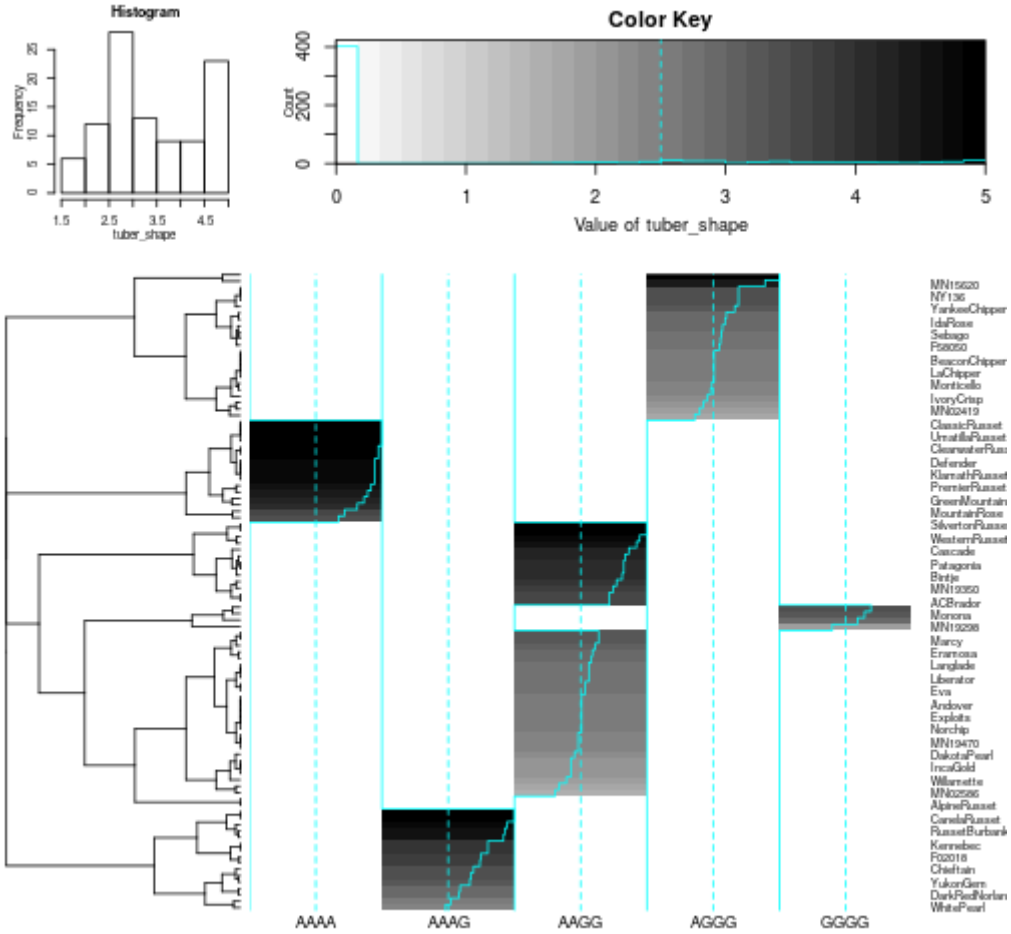
c2_25471


Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
CCCC
TCCC
TTCC

c2_47760


Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
CCCC
CCCG
CCGG
CGGG
GGGG

c2_53380


Histogram
Frequency
tuber_shape
Color Key
Count
Value of tuber_shape
AAAA
AAAG
AAGG
AGGG
GGGG

c2_54811