mGWAS-Explorer 2.0 Tutorial

– MR Analysis

Updated on 2023-06-07

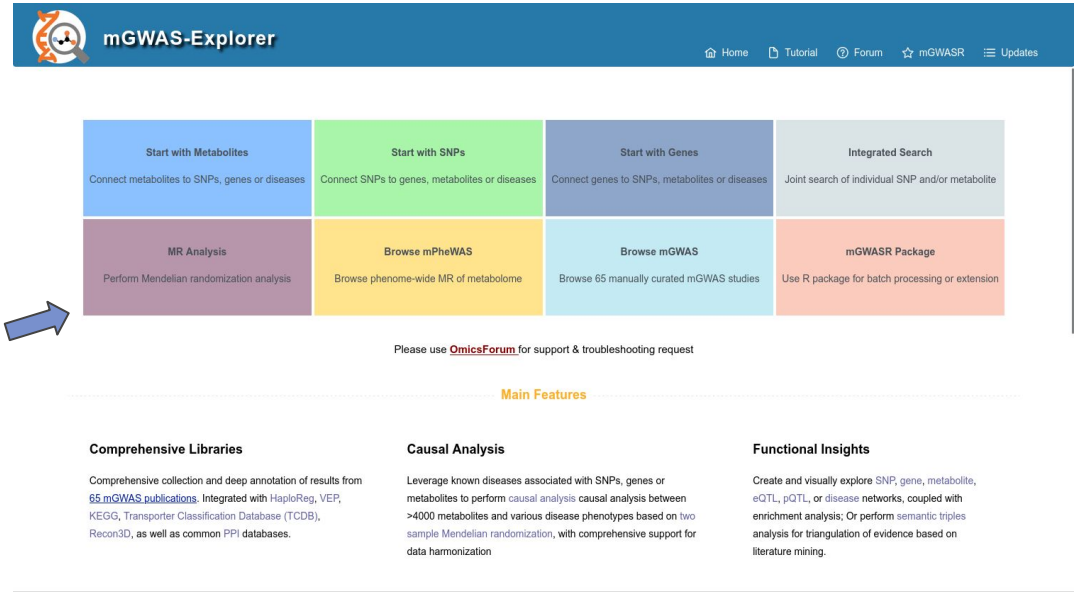
Introduction

- The **MR Analysis module** in mGWAS-Explorer 2.0 is a robust tool designed to facilitate the exploration of potential **causal** relationships between **metabolites** and various **phenotypes**.
- **Metabolites** are small molecules involved in the chemical reactions within cells, and **phenotypes** are observable traits that can range from physical characteristics to disease risk.
- This module leverages the principle of **Mendelian Randomization**, a method used in epidemiology to infer causality from observational data.

Introduction

- The core strength of this tool lies in its ability to support **two-sample Mendelian Randomization (2SMR)** analysis.
- With the assistance of 2SMR, we can analyze data from separate samples for the exposure and outcome, a technique that increases statistical power and reduces bias.
- The **MR Analysis module** allows for causal investigation across over 4000 metabolites, providing a wide range of options for exploration in your research.

Starting up



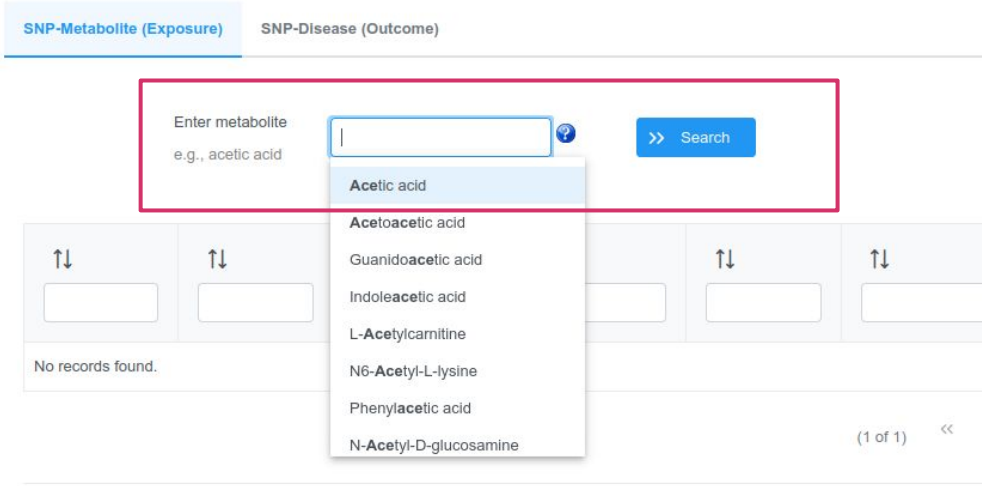
- Go to the mGWAS-Explorer homepage (www.mgwas.ca)
- Click on the “MR Analysis” button to enter the data upload page

Data upload

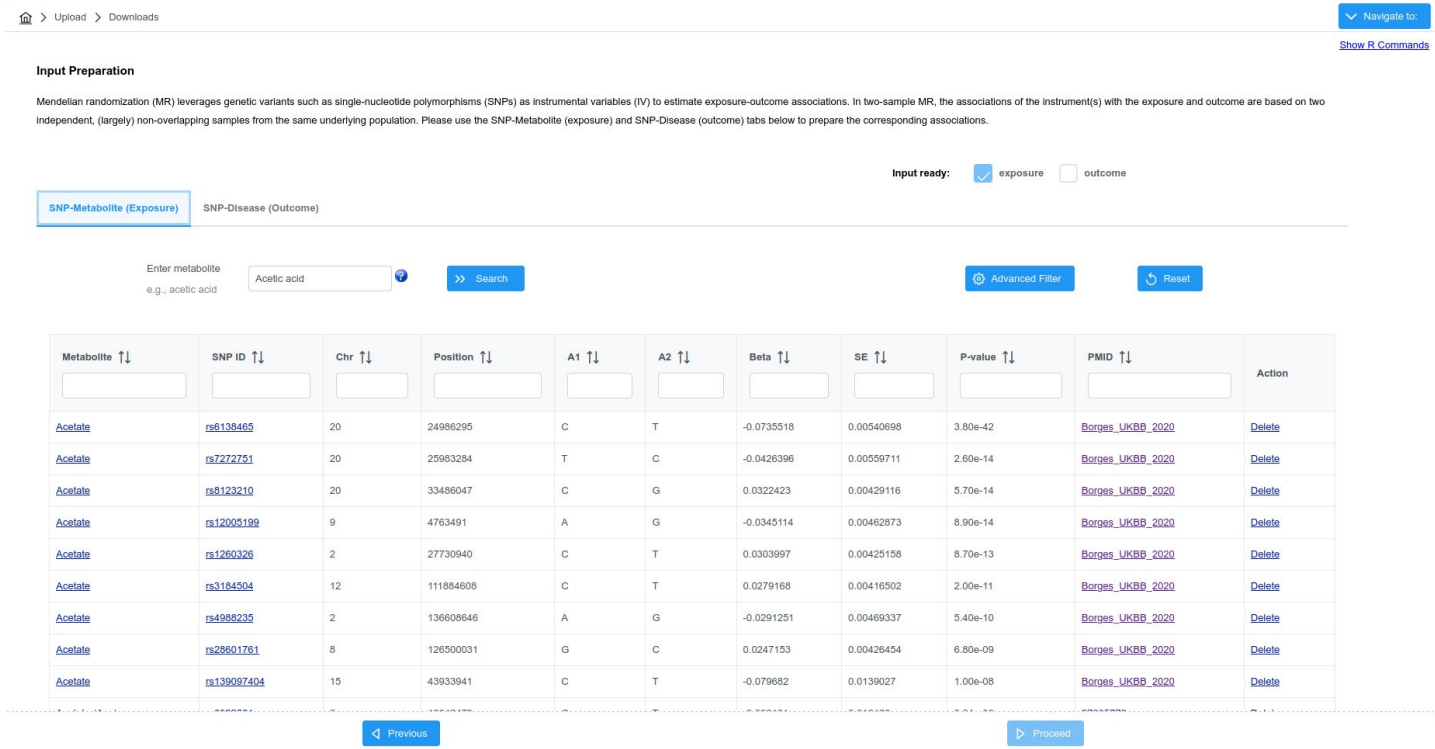
- To initiate your exploration into the potential causal relationships between metabolites and phenotypes, you must first specify your parameters of interest: the metabolites (which represent exposure) and the disease (which represents outcome)
- The process is straightforward. Upon reaching the data upload page of the MR Analysis module, you'll find fields to input your chosen metabolites and disease.
- By inputting the names of the metabolites and the specific disease you're interested in, you effectively set the stage for your investigation.

Data upload - exposure

- A drop-down list will appear when you start typing in the search box
- Please select a metabolite from the drop-down list and then click the "Search" button
- Once the metabolite have been specified, the application springs into action.
- It begins by identifying Single Nucleotide Polymorphisms (SNPs) that are **significantly** associated with the specified metabolites.
- This identification is based on our carefully curated mGWAS data.
- These SNPs serve as **instrumental variables** for your MR analysis.

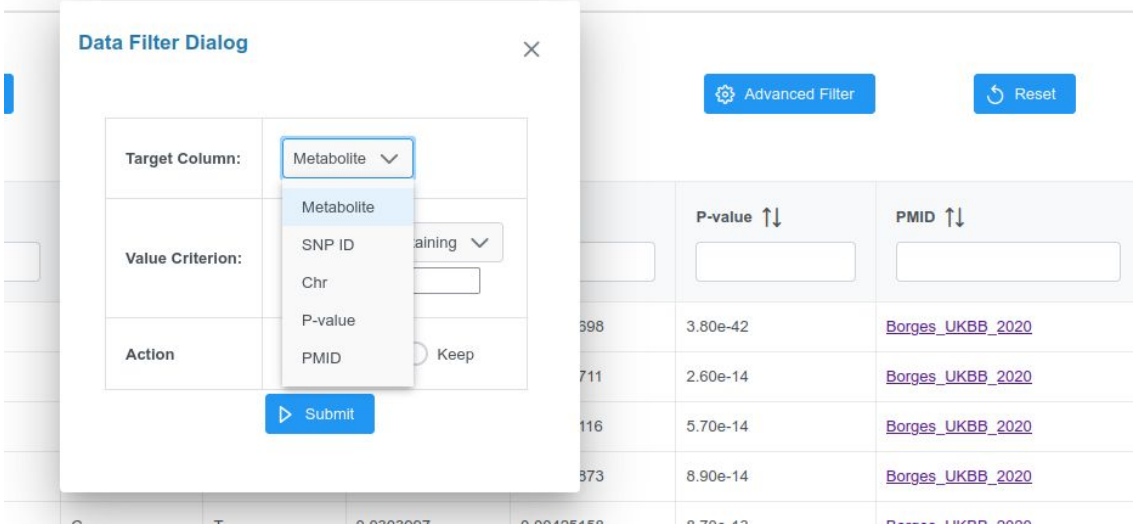


Data upload - exposure

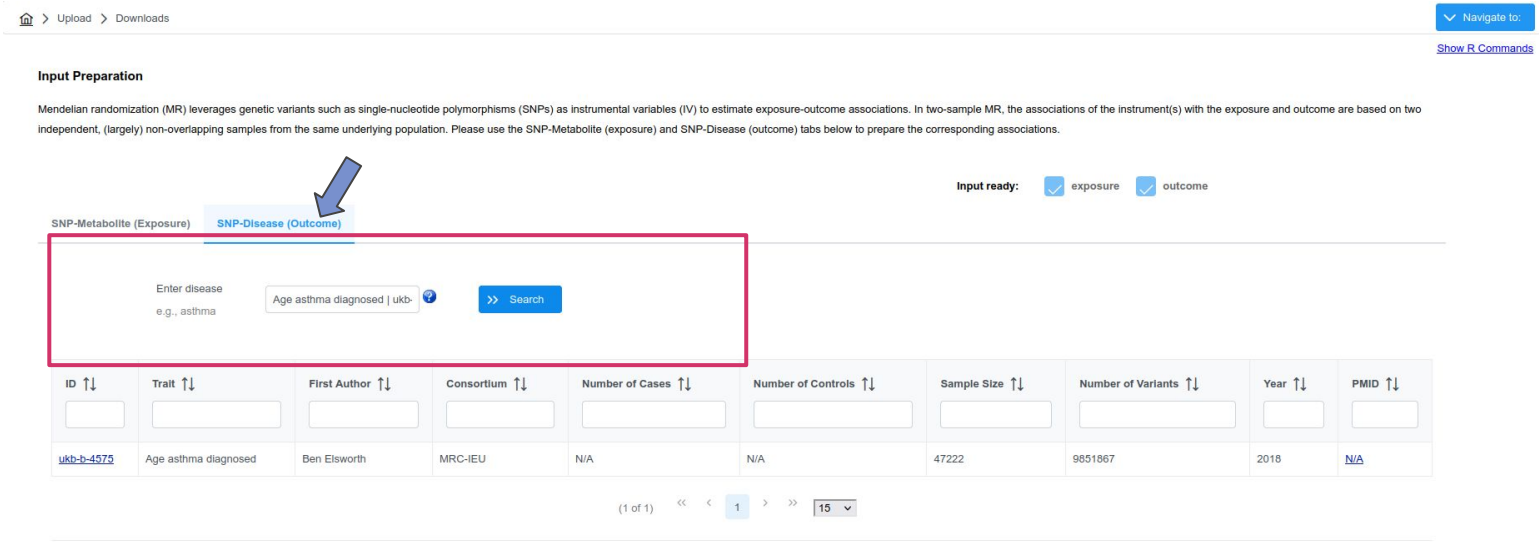


Data upload - exposure

- We also provide the options to filter the instrumental variables by using the "Advanced Filter"

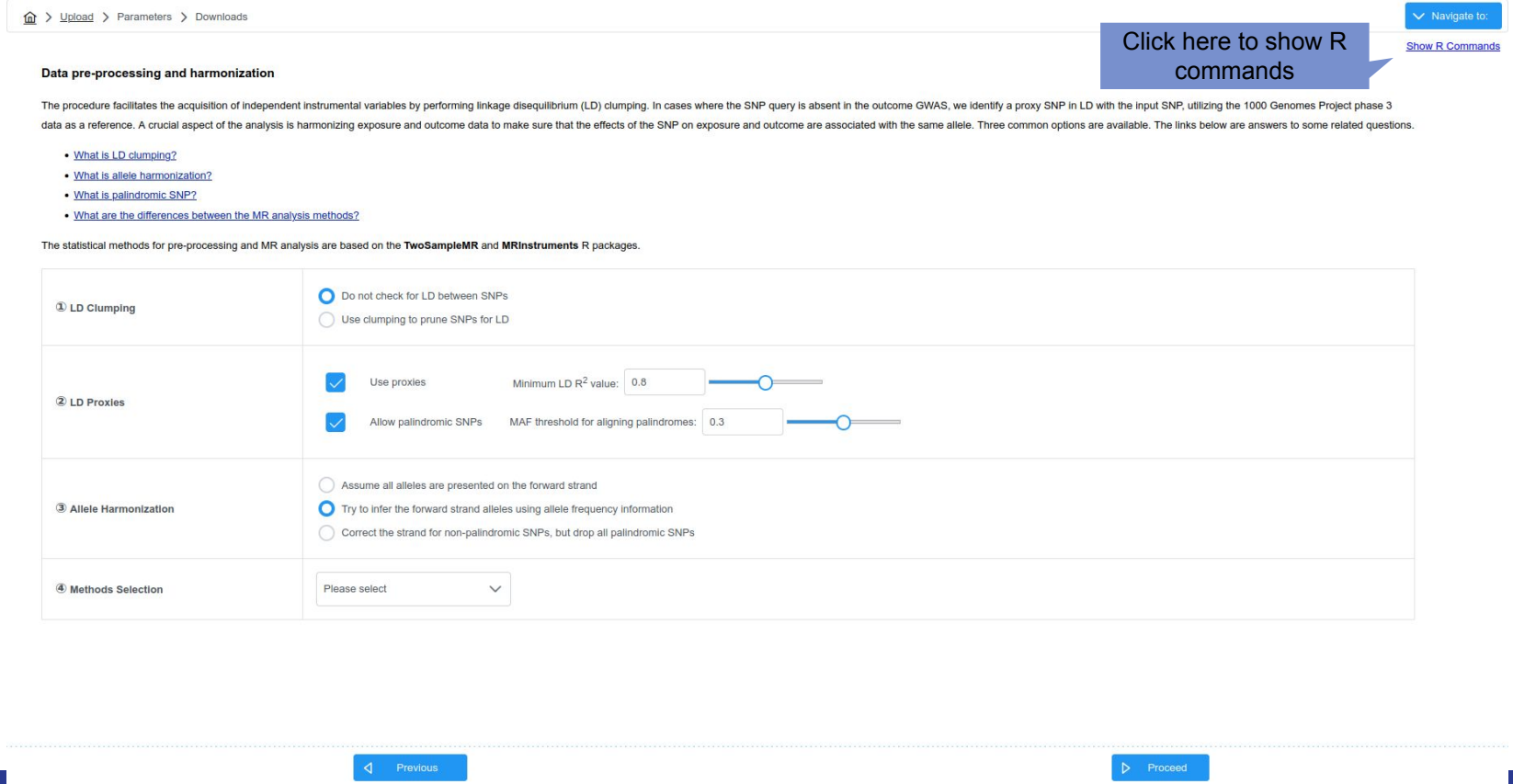


Data upload - outcome



- You can move onto the "Outcome" tab after the "Exposure" data input is ready.
- The application automatically extracts the instrumental SNPs and their association with the disease outcome from a comprehensive resource - the IEU OpenGWAS database.
- Click "Proceed" at the bottom right corner to the "Data pre-processing and harmonization" page

Data pre-processing and harmonization



LD clumping

- Ensuring the robustness and validity of your MR analysis requires careful selection of instrumental variables. The process known as 'clumping', performed by the mGWAS-Explorer 2.0 application, is crucial for this.
- Following the identification of SNPs significantly associated with your chosen metabolites and disease, the application initiates clumping.
- Please refer to our FAQs for an detailed explanation of LD clumping:
- https://omicsforum.ca/t/what-is-ld-clumping/1041

LD proxies

- In instances where the SNP query is absent in the GWAS of the outcome, the mGWAS-Explorer 2.0 application uses an effective solution: the identification of a proxy SNP.
- A **proxy SNP** is a genetic variant that is in strong LD with the input or query SNP. Essentially, this means that the proxy SNP and the query SNP are located close together on the chromosome and are therefore often inherited together.
- As such, the proxy SNP can 'stand in' for the query SNP, providing us with valuable information even in the absence of direct query SNP data in the outcome GWAS.
- The tool identifies these proxy SNPs utilizing the 1000 Genomes Project phase 3 data as a reference.

Allele harmonization

- Data harmonization is the process of ensuring that the SNP data for the exposure (metabolites) and the outcome (disease) are consistent.
- This step is critical for making valid comparisons and ensuring reliable analysis results.
- Given the importance of data harmonization, mGWAS-Explorer 2.0 provides you with options to manage potential inconsistencies:
  - Assume all alleles are presented on the forward strand
  - Infer the forward strand alleles using allele frequency information
  - Correct the strand for non-palindromic SNPs, but drop all palindromic SNPs
- See detailed explanations here:
  - https://omicsforum.ca/t/what-is-allele-harmonization/1043

MR method selection

- Our tool offers 18 distinct MR analysis methods, providing flexibility and versatility in your research.
  - https://omicsforum.ca/t/what-are-the-differences-between-the-mr-analysis-methods/1045
- The choice of method will depend on the specific context of the analysis.
- If you strongly believe that the genetic variants are valid instrumental variables, the **IVW** or **Wald ratio** method would be good choices.
- If you are unsure about the validity of the instrumental variables, the **weighted median** or **MR-Egger** method might be more appropriate.
- However, it's often a good idea to **use several methods** and **compare** the results to assess the robustness of your findings.
- If the different methods give similar results, this can increase your confidence in the causal effect estimate.

Sensitivity assessments

- Sensitivity analyses in Two-Sample MR are essential to validate the main findings and to assess the robustness of the MR estimates against violations of the key assumptions.
- We provide two most common sensitivity analyses:
  - **Heterogeneity test**: Testing for heterogeneity in a 2SMR analysis involves assessing whether the causal effect estimates from different SNPs are consistent with each other.
  - **Pleiotropy test**: This test is based on MR-Egger regression and specifically tests the null hypothesis that the intercept from the MR-Egger regression is equal to zero. If the null hypothesis is rejected, it suggests that there is horizontal pleiotropy, meaning that the SNPs used as instrumental variables affect the outcome through pathways that are not mediated by the exposure.

MR Results - understanding the summary table

Methods	Causal Effect Estimates				Heterogeneity Tests			Horizontal Pleiotropy		
	SNP Count	Beta	SE	P value	Q	Q_df	Q_pval	Egger Intercept	SE	P value
Inverse variance weighted	13	-0.10866	0.14139	0.44217	68.638	12	5.7536e-10	-	-	-
MR Egger	13	0.57832	0.3767	0.15298	51.167	11	3.8573e-07	-0.026583	0.013716	0.078702
Simple mode	13	0.15226	0.14996	0.32996	-	-	-	-	-	-
Weighted median	13	0.13926	0.095877	0.14635	-	-	-	-	-	-
Weighted mode	13	0.17881	0.097133	0.090485	-	-	-	-	-	-

Causal effect estimates

- These are estimates of the effect of the exposure (metabolites) on the outcome (disease), as determined by the MR analysis.
- Each row in the table corresponds to a specific method used and includes the causal effect estimate (**beta**), standard errors (**se**), **p-values,** and **SNP count** for that method.

Methods	Causal Effect Estimates			
	SNP Count	Beta	SE	P value
Inverse variance weighted	13	-0.10866	0.14139	0.44217
MR Egger	13	0.57832	0.3767	0.15298
Simple mode	13	0.15226	0.14996	0.32996
Weighted median	13	0.13926	0.095877	0.14635
Weighted mode	13	0.17881	0.097133	0.090485

Heterogeneity tests

- As we mentioned earlier, heterogeneity refers to variability in the causal estimates from different SNPs.
- The **Cochran's Q statistic** and its corresponding degrees of freedom and p-value are presented for each method.
- A large Q statistic relative to its degrees of freedom, or a small p-value, suggests that there's more heterogeneity among the causal estimates than what would be expected by chance.

Methods	Heterogeneity Tests		
	Q	Q_df	Q_pval
Inverse variance weighted	70.563	13	6.3211e-10
MR Egger	55.181	12	1.6795e-07
Simple mode	-	-	-
Weighted median	-	-	-
Weighted mode	-	-	-

Horizontal pleiotropy tests

- Horizontal pleiotropy occurs when the SNPs have effects on the outcome that are independent of their impact on the exposure.
- In the summary table, **MR-Egger regression intercept** and its corresponding p-value are presented.
- If the regression intercept significantly deviates from zero (typically determined by a p-value less than 0.05), it indicates the presence of horizontal pleiotropy, suggesting that some of the SNPs might be affecting the outcome through pathways other than the metabolite.

Methods	Horizontal Pleiotropy		
	Egger Intercept	SE	P value
Inverse variance weighted	-	-	-
MR Egger	-0.024656	0.013481	0.092352
Simple mode	-	-	-
Weighted median	-	-	-
Weighted mode	-	-	-

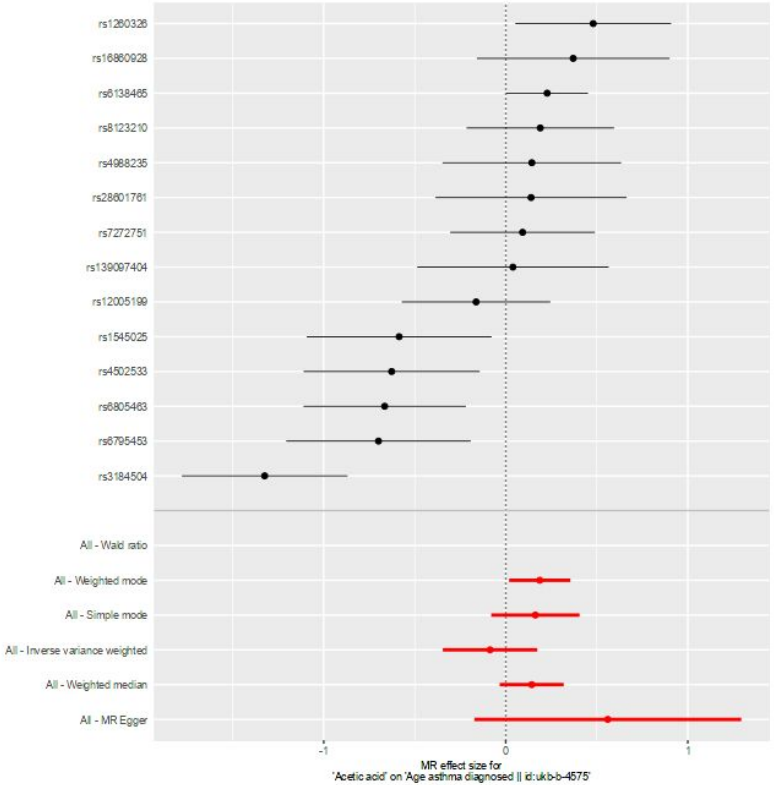
Visualizing your results: understanding the four types of plots

- mGWAS-Explorer 2.0 generates four types of plots (in four tabs) to visually represent your MR results, helping you to understand and interpret the data in an intuitive manner:
  - **Forest plot**: a visual representation of the causal effect estimates for each SNP, as well as the overall causal effect estimate.
  - **Scatter plot**: used to visualize the relationship between genetic associations with the exposure (on the x-axis) and the outcome (on the y-axis). Each point on the plot represents a single SNP.
  - **Funnel plot**: used to assess potential bias in the causal effect estimates, often due to pleiotropy. The plot is a scatter plot of the precision of each SNP's causal effect estimate (on the y-axis) against the estimate itself (on the x-axis).
  - **Leave-one-out plot**: this is a sensitivity analysis used in MR. It assesses the influence of each individual SNP on the overall causal effect estimate. In this analysis, the MR is rerun multiple times, each time leaving out one SNP, to see how much the overall estimate changes.

Forest plot

Here's how you interpret a forest plot in MR analysis:

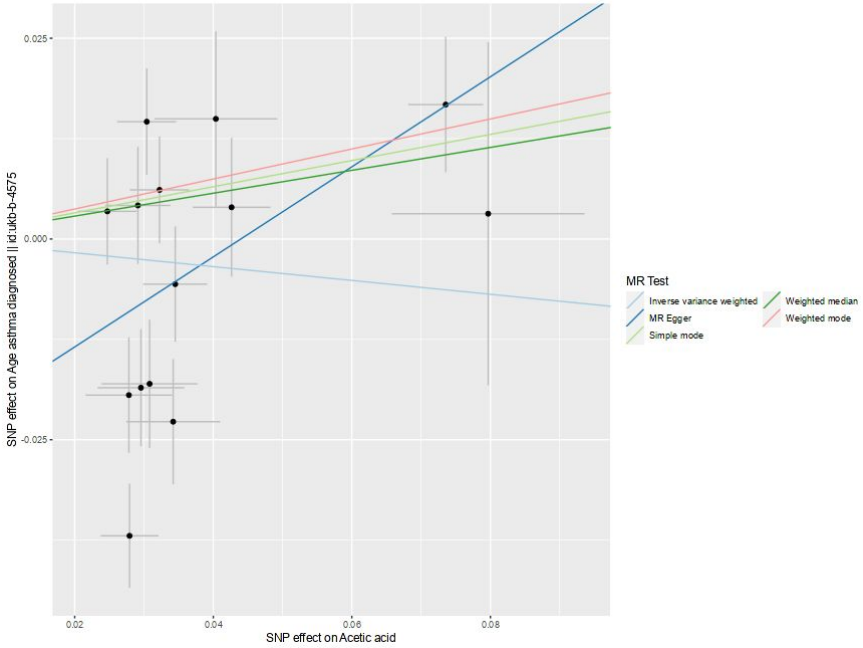
- **Individual SNP estimates:** Each horizontal line in the forest plot represents one SNP. The point on the line indicates the estimated causal effect of the exposure on the outcome for that SNP. The length of the line represents the confidence interval for that estimate.
- **Overall MR estimate**: This is usually represented at the bottom of the plot. The center of the point represents the combined estimate of the causal effect across all SNPs (i.e., the overall MR estimate), and the width of the point represents the confidence interval for the overall estimate.
- **Direction of effect**: If the point estimates are mostly on the right side of the line of no effect (usually represented by a vertical line at 0), it suggests that the exposure increases the risk of the outcome. If they're mostly on the left side, it suggests that the exposure decreases the risk of the outcome.
- **Heterogeneity**: If the individual SNP estimates are widely scattered, it suggests that there is heterogeneity in the causal estimates.
- **Influence of individual SNPs**: If removing one SNP from the analysis changes the overall MR estimate substantially, it suggests that the SNP may be exerting undue influence on the results.



Scatter plot

Here's how you interpret the scatter plot:

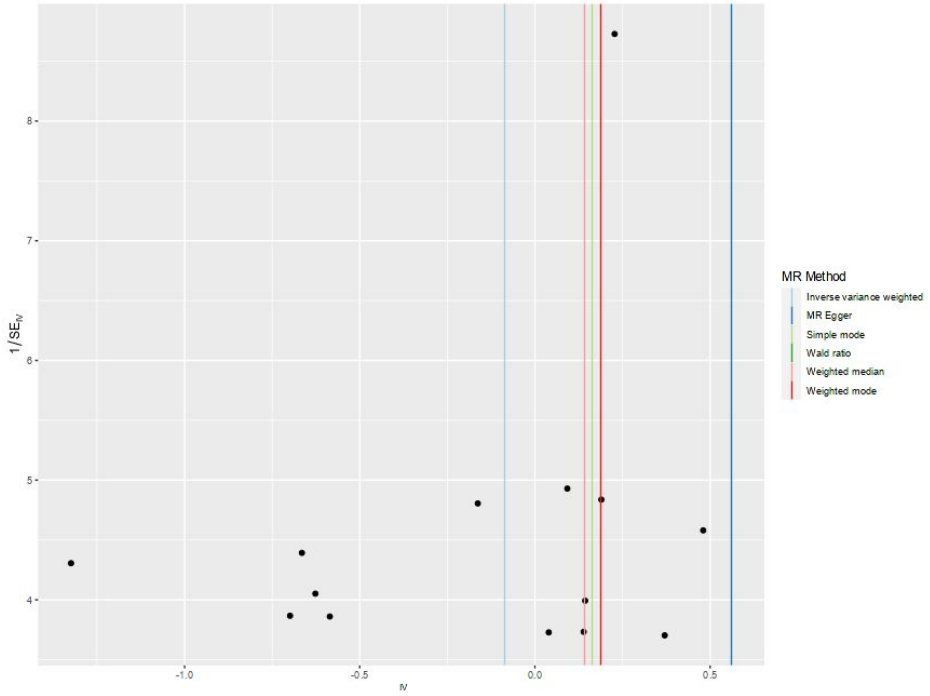
- **Slope and Linearity**: The slope of the line fitted through the points is an estimate of the causal effect of the exposure on the outcome. If the points are scattered around a straight line with a positive slope, it suggests that an increase in exposure is associated with an increase in the outcome, indicating a positive causal effect.
- **Scatter around the line**: If the points are closely clustered around the line, it suggests that the SNPs are good instruments (i.e., they satisfy the assumptions of MR). If the points are widely scattered, it suggests potential violation of the MR assumptions, such as presence of pleiotropy (where a single gene affects more than one trait), or measurement error.
- **Outliers**: Outliers are points that lie far from the line of best fit. These might represent SNPs that have pleiotropic effects, or that are affected by measurement error or linkage disequilibrium (where genetic variants are inherited together more often than would be expected by chance). Outliers can bias the MR estimate, so it's important to consider sensitivity analyses (like MR-Egger regression or leave-one-out analysis) that can help assess the impact of potential outliers on the MR results.



Funnel plot

Here's how you interpret a funnel plot in MR analysis:

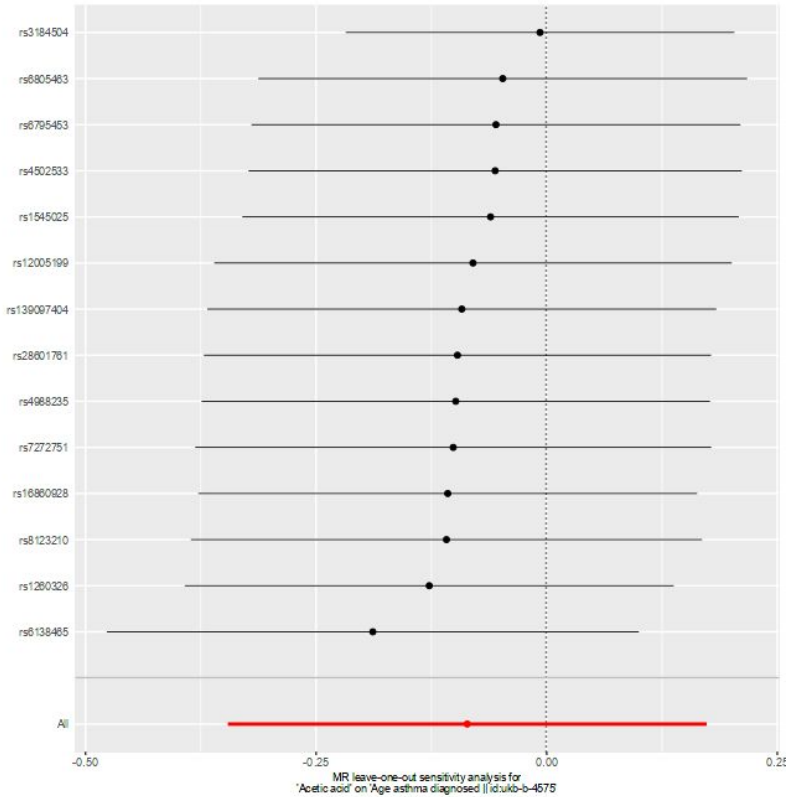
- **Symmetry**: In the absence of bias, the plot should be symmetrical around the vertical line that represents the overall causal effect estimate. This is because each SNP's estimate is expected to vary randomly around the true causal effect, with higher precision estimates (those closer to the top of the plot) varying less than lower precision estimates (those closer to the bottom). If the plot is asymmetrical, it suggests that there may be bias in the MR estimates.
- **Direction of bias**: If the points on the plot are mostly to the right of the overall estimate, it suggests that there is positive bias (i.e., the estimates are skewed towards positive values). If the points are mostly to the left, it suggests that there is negative bias (i.e., the estimates are skewed towards negative values).
- **Pleiotropy**: Asymmetry in the funnel plot can be a sign of pleiotropy, where a SNP influences the outcome through more than one pathway. This violates one of the key assumptions of MR and can bias the results.
- **Heterogeneity and outliers**: A widely scattered plot or the presence of outlier points far from the vertical line can indicate heterogeneity in the causal estimates or the presence of outlier SNPs, which could be due to pleiotropy, linkage disequilibrium, or measurement error. These SNPs might be worth further investigation.
- **Precision**: Points towards the top of the plot are SNPs with higher precision estimates (smaller standard errors), while points towards the bottom are SNPs with lower precision estimates (larger standard errors).



Leave-one-out (LOO) plot

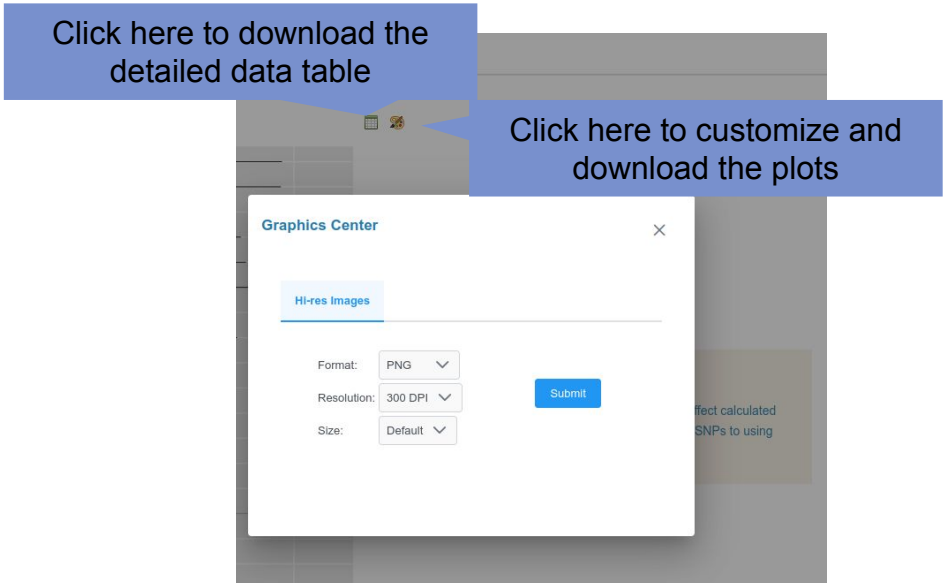
Here's how you interpret a leave-one-out plot in MR analysis:

- **Influence of individual SNPs**: Each point on the LOO plot represents the overall MR estimate obtained when that SNP is excluded from the analysis. If removing a particular SNP changes the overall MR estimate substantially, it suggests that this SNP may be exerting undue influence on the results.
- **Consistency of the MR estimate**: If all the points on the LOO plot are close to the overall MR estimate, it suggests that the MR estimate is consistent and not unduly influenced by any single SNP.
- **Identifying outlier SNPs**: If there are one or more points that are far from the overall MR estimate, these are outlier SNPs that may be violating the assumptions of MR. These SNPs might be worth further investigation to understand why they are outliers.
- **Confidence intervals**: The LOO plot usually also shows the confidence intervals for each leave-one-out MR estimate. If these intervals overlap substantially, it suggests that the MR results are robust to the exclusion of individual SNPs.



Downloading tables and plots

- mGWAS-Explorer 2.0 gives you the ability to download tables and customize/download the plots to suit your specific needs.



Conclusion - putting MR analysis to work

- Throughout this tutorial, we've taken a comprehensive tour of the MR Analysis module in mGWAS-Explorer 2.0, from setting up the exposure and outcome data, selecting and pruning SNPs, to conducting the MR analysis and interpreting the results. We have seen how this powerful tool can facilitate a systematic investigation of potential causal relationships between metabolites and various phenotypes.
- Notably, the MR Analysis module is designed to be versatile and user-friendly. With its intuitive interface, numerous analysis methods, and robust visualization tools, it accommodates researchers of varying levels of expertise and diverse research goals.
- We hope that this overview has given you the confidence to try out the MR Analysis module with your own data. Remember, the true power of a tool lies in its use, so we strongly encourage you to explore this module, experiment with its features, and apply it to your research.
- Your feedback and questions are invaluable to us. If you encounter any issues, have suggestions for improvements, or simply want to share your success stories with mGWAS-Explorer 2.0, please do not hesitate to post on our OmicsForum: https://omicsforum.ca/c/mgwas-explorer/12