



# MetaboAnalyst 6.0

-- a unified platform for metabolomics data processing,  
analysis and interpretation

Causal Analysis using Mendelian randomization

# Module Overview



This module offers functions to estimate the causal relationship between metabolites and phenotypes through Mendelian randomization (MR) analysis.

- ✓ There are many metabolomics-based genome-wide association studies (mGWAS) conducted to understanding the genetic regulations of metabolites in complex phenotype.
- ✓ By leveraging those SNP-tagged metabolites and summary statistics from public GWAS repositories, we can now test potential causal relationships between those genetically influenced metabolites and a disease outcome of interest using the well-established two-sample Mendelian randomization method.
- ✓ MR can estimate whether a relationship between a metabolite and a phenotype is causal, while reduce the impact of confounding factors and reverse causality that often plague observational studies.



# Content

- ❑ 1. Introduction;
- ❑ 2. Choose the Module;
- ❑ 3. Causal Analysis;
  - 3.1 Specify metabolite and phenotype of interest
  - 3.2 SNP filtering and harmonization
  - 3.3 Select statistical methods
  - 3.4 Mendelian randomization results
- ❑ 4. Download Results.

# 1. Introduction

## Background

- It is now possible to estimate causal relationship between metabolites and a phenotype of interest.
- If a metabolite is causal for a given disease, genetic variants which influence the levels of that metabolite, either directly or indirectly, should result in a higher risk of the disease.

## Data Formats

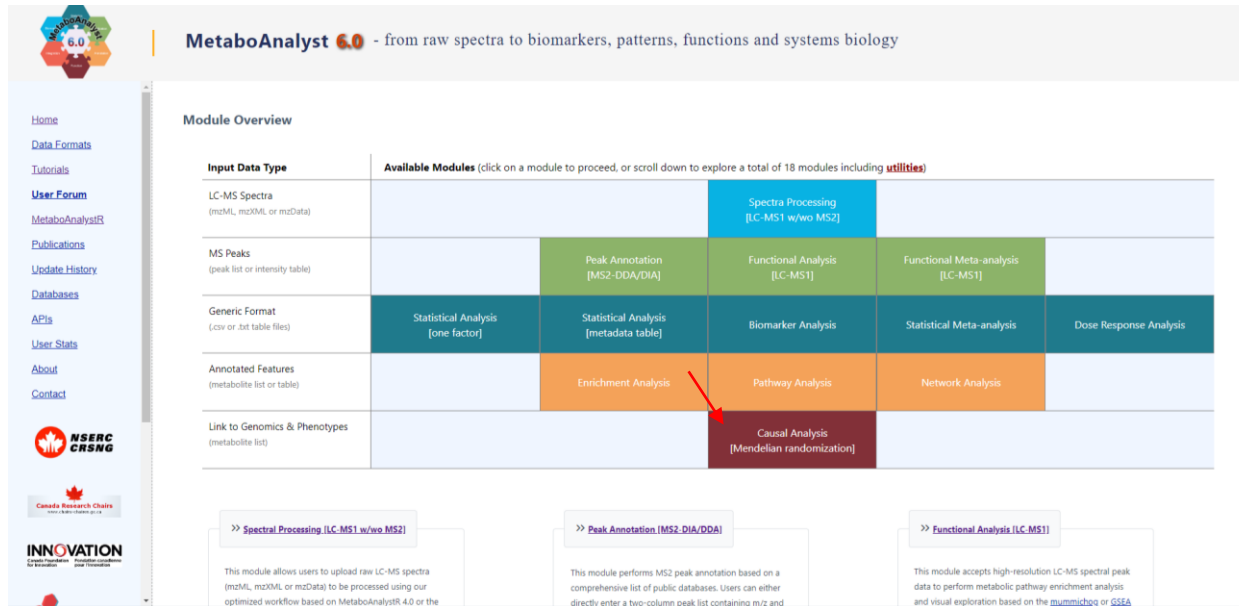
- No data upload required
- Users select an exposure (i.e., metabolites) and an outcome (i.e., diseases) of interest from the available options from our built-in databases

## Expected Results

- This module provides user comprehensive results on potential causal relationships between exposure-outcome based on two-sample MR
- i. Intermediate results from harmonization steps
  - ii. The results from various statistical routines to estimate the causal effects and associated diagnostic and visualization plots

## 2. Choose the Module

Go to MetaboAnalyst (<https://www.metaboanalyst.ca>), and select the module



**MetaboAnalyst 6.0** - from raw spectra to biomarkers, patterns, functions and systems biology

**Module Overview**

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including <a href="#">utilities</a> )				
LC-MS Spectra (mzML, mzXML or mzData)			Spectra Processing [LC-MS1 w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS1]	Functional Meta-analysis [LC-MS1]	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes (metabolite list)			Causal Analysis [Mendelian randomization]		

**>> Spectra Processing [LC-MS1 w/wo MS2]**

This module allows users to upload raw LC-MS spectra (mzML, mzXML or mzData) to be processed using our optimized workflow based on MetaboAnalystR 4.0 or the

**>> Peak Annotation [MS2-DIA/DDA]**

This module performs MS2 peak annotation based on a comprehensive list of public databases. Users can either directly enter a two-column peak list containing m/z and

**>> Functional Analysis [LC-MS1]**

This module accepts high-resolution LC-MS spectral peak data to perform metabolic pathway enrichment analysis and visual exploration based on the [mummichog](#) or [GSEA](#)



### **3. Causal Analysis via two-sample Mendelian randomisation (2SMR)**

# 3.1 Specify metabolite and phenotype of interest

## Please specify metabolites (exposure) and outcome of interest

Causal analysis is based on Mendelian randomization (MR) which leverages genetic variants such as single-nucleotide polymorphisms (SNPs) as instrumental variables (IV) to estimate exposure-outcome associations. The growing number of mGWAS studies and two-sample MR method permit causal analysis between metabolite and outcome of interest as described below:

1. Identify SNPs that are significantly associate with a metabolite of interest from our large collections of the recent [mGWAS studies](#) (covering > 4000 metabolites including their ratios);
2. Obtain the estimates of associations between these same SNPs with an outcome of interest from public repository. We use [Open GWAS Project](#).
3. Perform SNP filtering and harmonize the effect sizes for SNPs on the exposures and the outcomes to be for the same reference allele.
4. Conduct MR analysis, sensitivity analyses, and explore the graphical outputs

Please note you may not be able to perform causal analysis in some cases when no suitable SNPs are found in the two repositories.

### 1. Select a metabolite of interest (exposure):

Due to its complex and computing intensive nature, MR analysis is typically performed with [one metabolite/exposure at a time](#), to make sure that each step is performed properly as well as to avoid performance issue (max 5).

Q cys

- Cysteineglutathione disulfide
- Methylcysteine
- 3-(Cystein-5-yl)acetaminophen
- S-N-Methylcysteine
- 3-Indolepropionic acid/(S-(5-Adenosyl)-L-homocysteine
- 3-Indolepropionic acid/(S-Sulfo-L-cysteine
- L-Cystine/3-Indolepropionic acid

Filter selected metabolites

- L-Cystathionine

Users should first select an exposure (i.e., metabolites) and an outcome (i.e., diseases) of interest.

Search to choose the metabolite of interest (e.g. *Cystathionine*) from the left box. Once you select it, it will be automatically added into the right box.

### 2. Specify an outcome of interest:

Enter a key word to see available options from the public repository.

Type 2 diabetes | finn-b-E4\_DM2

Proceed

Click "Proceed" to continue

For instance, we are interested in Type 2 diabetes. Type the name to see a list of matched studies. Here we choose finn-b-E4\_DM2

## 3.2 SNP filtering and harmonization



- Multiple SNPs could be identified as potential instrumental variables (IV) from the mGWAS and GWAS studies.
- To perform proper 2SMR, the IVs should be
  - Independent (i.e. not correlated with each other)
  - Showing strong effect (i.e. significant p-values)
  - No horizontal pleiotropy (i.e. affect the outcome only through the metabolite).
- Users need to carefully examine SNPs and apply different filtering and harmonization methods for each criterion



# 3.2 SNP Filtering and harmonization

To properly conduct two-sample MR analysis, the instrumental variables (IV) should be **independent** (i.e. not correlated with each other), showing **strong effect** (i.e. significant p-values), and **no horizontal pleiotropy** (i.e. affect the outcome only through the metabolite). The step provides following procedures to facilitate proper MR analysis:

- Acquisition of independent IVs by performing linkage disequilibrium (LD) clumping.
- In cases where the SNP query is absent in the outcome GWAS, a proxy SNP in LD with the input SNP, utilizing the 1000 Genomes Project (phase 3).
- Harmonizing exposure and outcome data to make sure that the effects of the SNPs on exposure and outcome are associated with the same allele. You should also review the table below to perform further harmonization based on other metadata (such as population, study info, etc)
- To control horizontal pleiotropy, you should manually exclude SNPs that are associated with multiple metabolites.

1. LD Clumping	<input checked="" type="radio"/> Do not check for LD between SNPs <input type="radio"/> Use clumping to prune SNPs for LD
2. LD Proxies	<input checked="" type="radio"/> Do not use proxies <input type="radio"/> Use proxies and allow palindrome SNPs (advanced settings)
3. Allele Harmonization	<input type="radio"/> Assume all alleles are presented on the forward strand <input checked="" type="radio"/> Try to infer the forward strand alleles using allele frequency information <input type="radio"/> Correct the strand for non-palindromic SNPs, but drop all palindromic SNPs

Submit

Harmonization steps require intensive computing and also access via remote server.

It could take a long time or time out. Please be patient or try more time

SNP ID ↑↓	Associated Metabolites ↑↓	Nearest Gene ↑↓	P-value ↑↓	Biofluid ↑↓	Population ↑↓	Study ↑↓	Include
▼ <b>L-Cystathionine (7)</b>							
<a href="#">rs117782586</a>	L-Cystathionine	<a href="#">JRKL</a>	4.637e-08	Blood	European	<a href="#">28263315</a>	<input checked="" type="checkbox"/>
<a href="#">rs146276253</a>	L-Cystathionine	<a href="#">ANKRD13C</a>	3.436e-08	Blood	European	<a href="#">28263315</a>	<input checked="" type="checkbox"/>
<a href="#">rs150320192</a>	L-Cystathionine	<a href="#">SRSE11</a>	3.566e-08	Blood	European	<a href="#">28263315</a>	<input checked="" type="checkbox"/>

## 3.3 Select statistical methods

MR analysis methods are based on the *TwoSampleMR* and *MRInstruments* R packages. Among these methods, the *median estimator* and *MR Egger regression* allow for genetic pleiotropy. You can use mouse-over of the corresponding question marks to learn more about each method.

<input checked="" type="checkbox"/> Wald ratio <sup>?</sup>	<input type="checkbox"/> Maximum likelihood <sup>?</sup>	<input checked="" type="checkbox"/> MR Egger <sup>?</sup>
<input type="checkbox"/> Simple median <sup>?</sup>	<input checked="" type="checkbox"/> Weighted median <sup>?</sup>	<input type="checkbox"/> Inverse variance weighted radial <sup>?</sup>
<input type="checkbox"/> Inverse variance weighted (MRE) <sup>?</sup>	<input type="checkbox"/> Inverse variance weighted (FE) <sup>?</sup>	<input checked="" type="checkbox"/> Simple mode <sup>?</sup>
<input checked="" type="checkbox"/> Weighted mode <sup>?</sup>	<input type="checkbox"/> Weighted mode (NOME) <sup>?</sup>	<input type="checkbox"/> Simple mode (NOME) <sup>?</sup>
<input type="checkbox"/> Sign concordance test <sup>?</sup>	<input type="checkbox"/> Unweighted regression <sup>?</sup>	

A total of 14 MR methods are offered currently. Some of them are more robust and can better tolerate violations of the assumptions to certain degree

- Mouse over the question marks for each method to see their main features.
- You can also find more detailed introduction on the forum:

<https://omicsforum.ca/t/what-are-the-differences-between-the-mr-analysis-methods/1045>

## 3.4 Mendelian randomization results

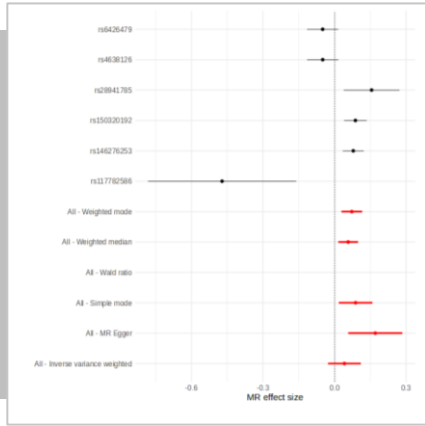
∨ L-Cystathionine

Methods	SNP Count	Causal Effect Estimates			Heterogeneity Tests			Horizontal Pleiotropy		
		Beta	SE	P value	Q	Q_df	Q_pval	Egger Intercept	SE	P value
Inverse variance weighted	6	0.041396	0.034939	0.23609	35.367	5	1.2709e-06	-	-	-
MR Egger	6	0.17071	0.057839	0.041906	14.006	4	0.0072767	-0.1179	0.047732	0.068948
Simple mode	6	0.088203	0.032068	0.040286	-	-	-	-	-	-
Weighted median	6	0.057145	0.021665	0.0083484	-	-	-	-	-	-
Weighted mode	6	0.072379	0.0219	0.021358	-	-	-	-	-	-

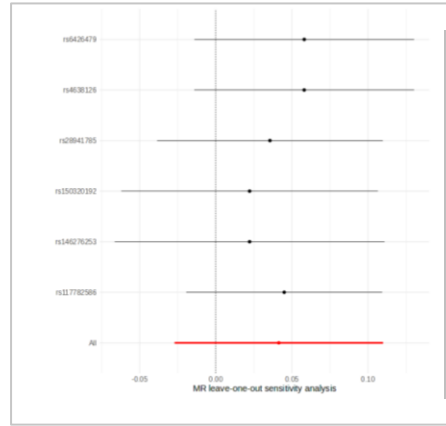
- The MR results are organized per metabolite (exposure).
- For metabolite, it shows the SNPs instrumental variables, along with their corresponding causal effect estimates, standard errors and p-values. Key values such as the MR-Egger regression intercept and its corresponding p-value are presented.
- Not all methods selected from the previous page would yield results depending of the data used.

# 3.4 Graphical outputs from MR analysis

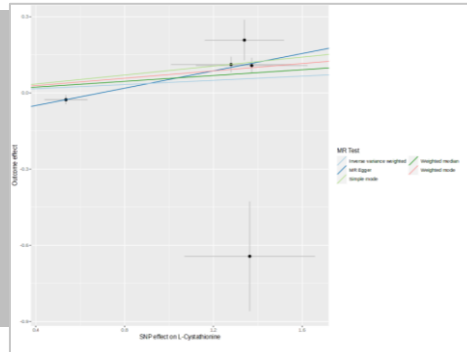
**Forest plot** compares the causal effect calculated using the methods that include all the SNPs to using each SNP separately.



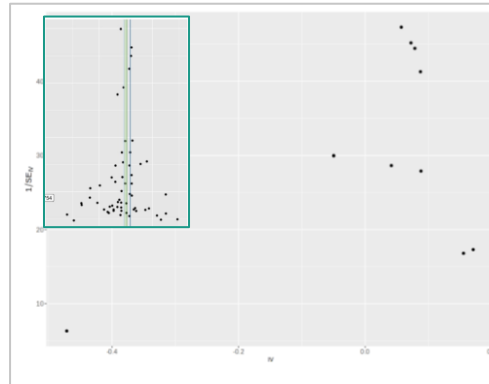
**Leave one out sensitivity analysis:** assesses whether a single SNP is having a disproportionately larger impact on an association. Each dot represents the MR analysis excluding that specific SNP using IVW method.



**Scatter plot** shows the relationships between SNP effects on exposure vs on the outcome. The slopes indicating the causal association



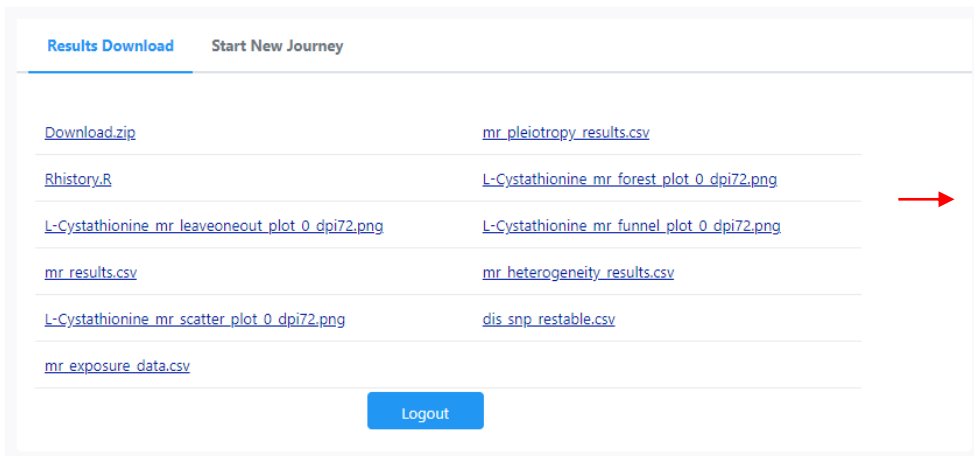
**Funnel plot:** Funnel shape will become more obvious with many SNPs (i.e. green box inset). Its asymmetry and wider spread may suggest horizontal pleiotropy.



# 4. Download Results

## Download Results & Start New Journey

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button. Finally, you can continue to explore other compatible modules using the **Start New Journey** tab.



The screenshot shows a web interface with two tabs: 'Results Download' (active) and 'Start New Journey'. Below the tabs is a list of downloadable files arranged in two columns. At the bottom of the list is a blue 'Logout' button.

Results Download	Start New Journey
<a href="#">Download.zip</a>	<a href="#">mr_pleiotropy_results.csv</a>
<a href="#">Rhistory.B</a>	<a href="#">L-Cystathionine_mr_forest_plot_0_dpi72.png</a>
<a href="#">L-Cystathionine_mr_leaveoneout_plot_0_dpi72.png</a>	<a href="#">L-Cystathionine_mr_funnel_plot_0_dpi72.png</a>
<a href="#">mr_results.csv</a>	<a href="#">mr_heterogeneity_results.csv</a>
<a href="#">L-Cystathionine_mr_scatter_plot_0_dpi72.png</a>	<a href="#">dis_snp_restable.csv</a>
<a href="#">mr_exposure_data.csv</a>	

Logout

All results can be downloaded here.



# In summary

If you have any questions, please read/post into OmicForum ([www.omicsforum.ca](http://www.omicsforum.ca))

Or contact us:

[zhiqiang.pang\[at\]xialab.ca](mailto:zhiqiang.pang@xialab.ca)

[jeff.xia\[at\]xialab.ca](mailto:jeff.xia@xialab.ca)

- Two-sample MR analysis allows researchers to estimate potential causal relationships between a metabolite and a phenotype of interest based on public data (mGWAS and GWAS summary statistics)
- Performing 2MSR requires identification of suitable SNPs (i.e. instrument variables) and performing filtering and harmonization. The process is computing intensive and better start with one metabolite at a time
- The module offers various MR methods with different strengths and limitations. They may give different estimates. Carefully examine the graphical outputs are necessary to reach robust conclusions