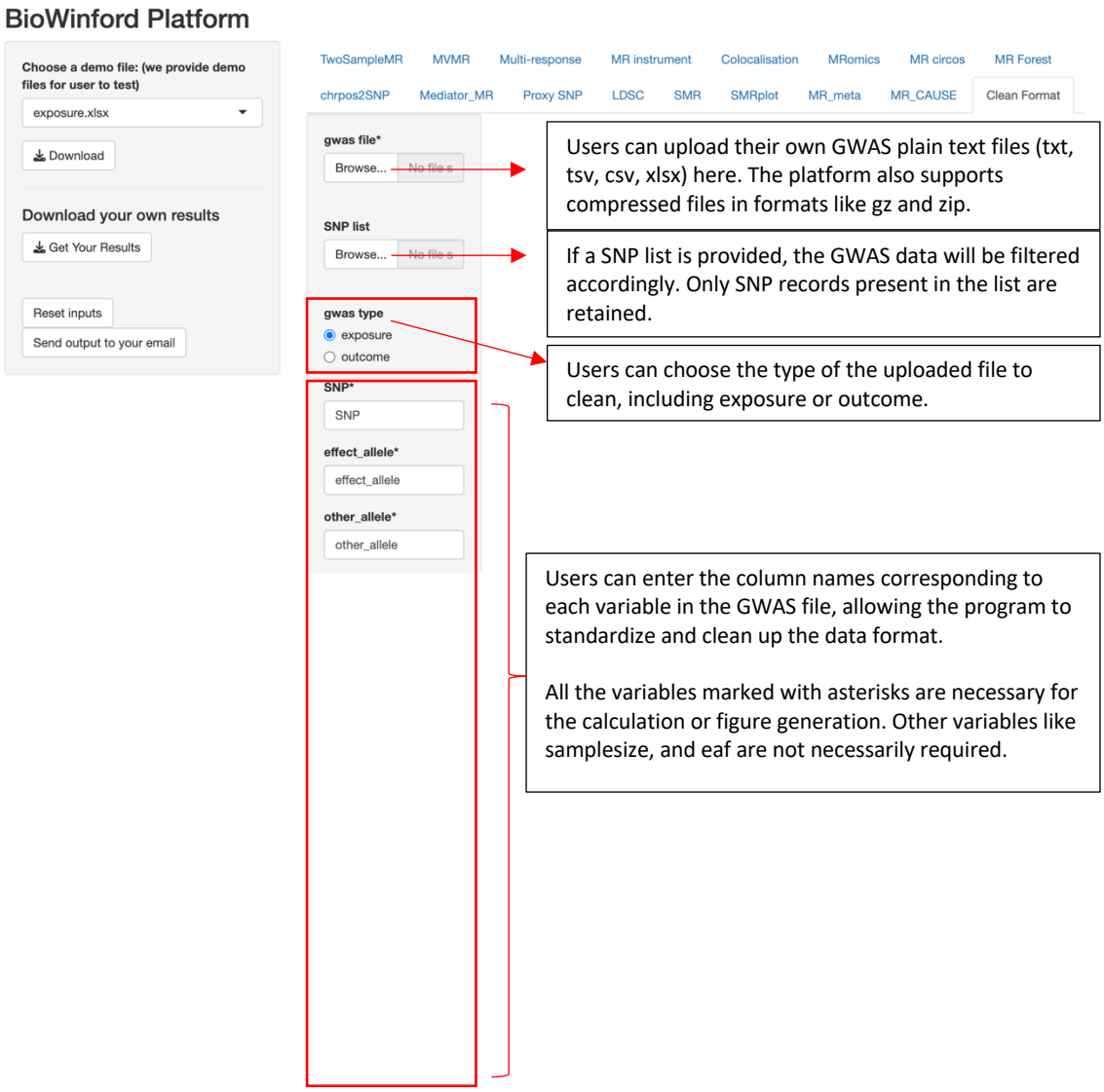
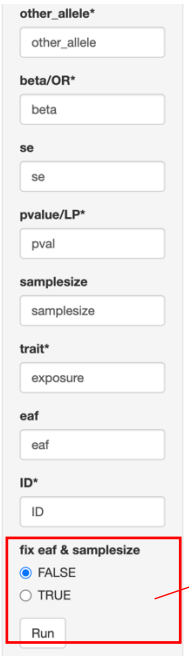
User guide

BioWinfordMR is a platform dedicated to performing Mendelian randomization causal inference. The platform integrates various Mendelian-related analytical functions and GWAS data. Users can learn how to utilize these functions for their analyses by reading the user guide documentation. Next, we will introduce each module individually.

Clean_format

This module can help users easily process GWAS data in different formats into a standard format. The purpose of doing this is to make it easier for other functional modules to read the same standardized input file to complete the analysis. Generally, GWAS data contains information such as SNP, effect allele, other allele, beta, standard error, p-value, effect allele frequency, trait, ID, etc. However, GWAS data from different sources do not have a standardized naming or arrangement order. Therefore, users can clean up this data using the clean format module.

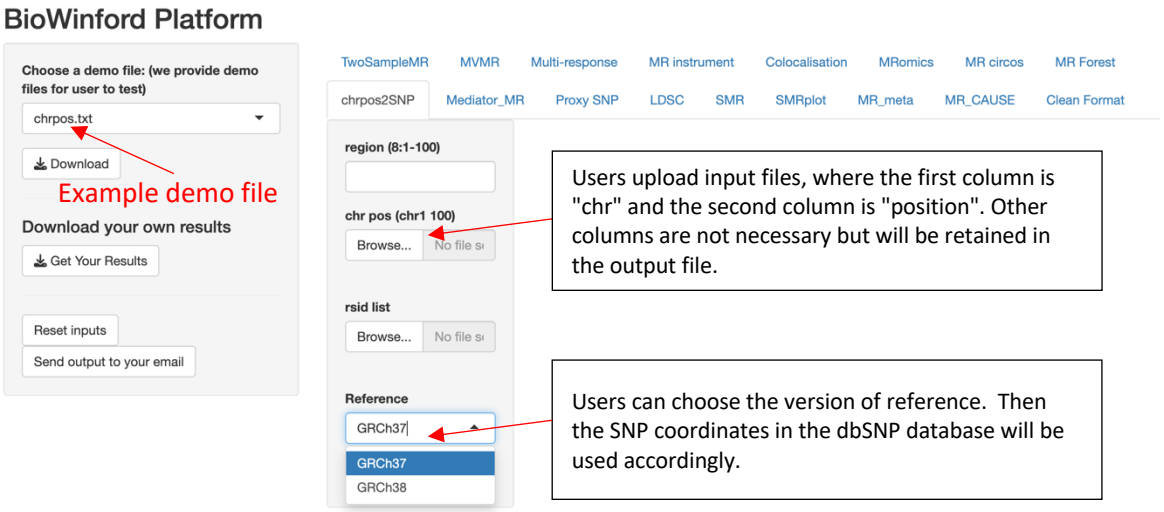




Although samplesize and eaf are not necessary for most MR analyses, it's still recommended to provide these two variables in certain situations, such as when conducting coloc analysis. Therefore, users can choose to supplement the samplesize and eaf information of SNPs through reference data from the 1000 Genome Project when these variables are missing. This may not fully restore the true original information, but it can somewhat compensate for incomplete data.

chrpos2snp

In some special cases, users may have obtained GWAS data that is missing SNP information (rsid). For MR analysis, the rsid for each SNP is necessary. In such situations, users can use the "chrpos2snp" module on the Biowinfordmr platform to map the SNP's chromosome and position coordinates to the dbSNP database and convert them to the corresponding rsid.

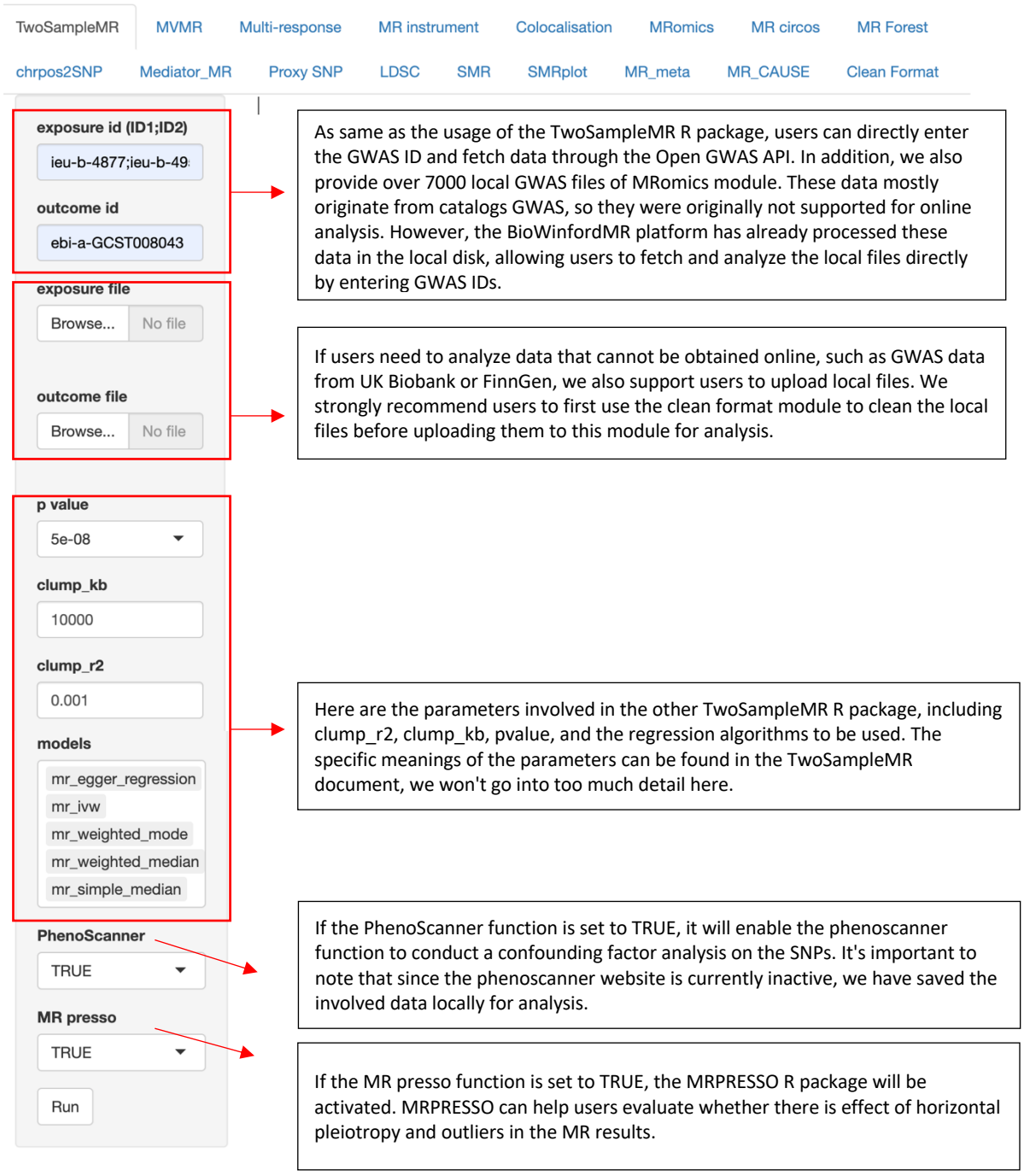


Users upload input files, where the first column is "chr" and the second column is "position". Other columns are not necessary but will be retained in the output file.

Users can choose the version of reference. Then the SNP coordinates in the dbSNP database will be used accordingly.

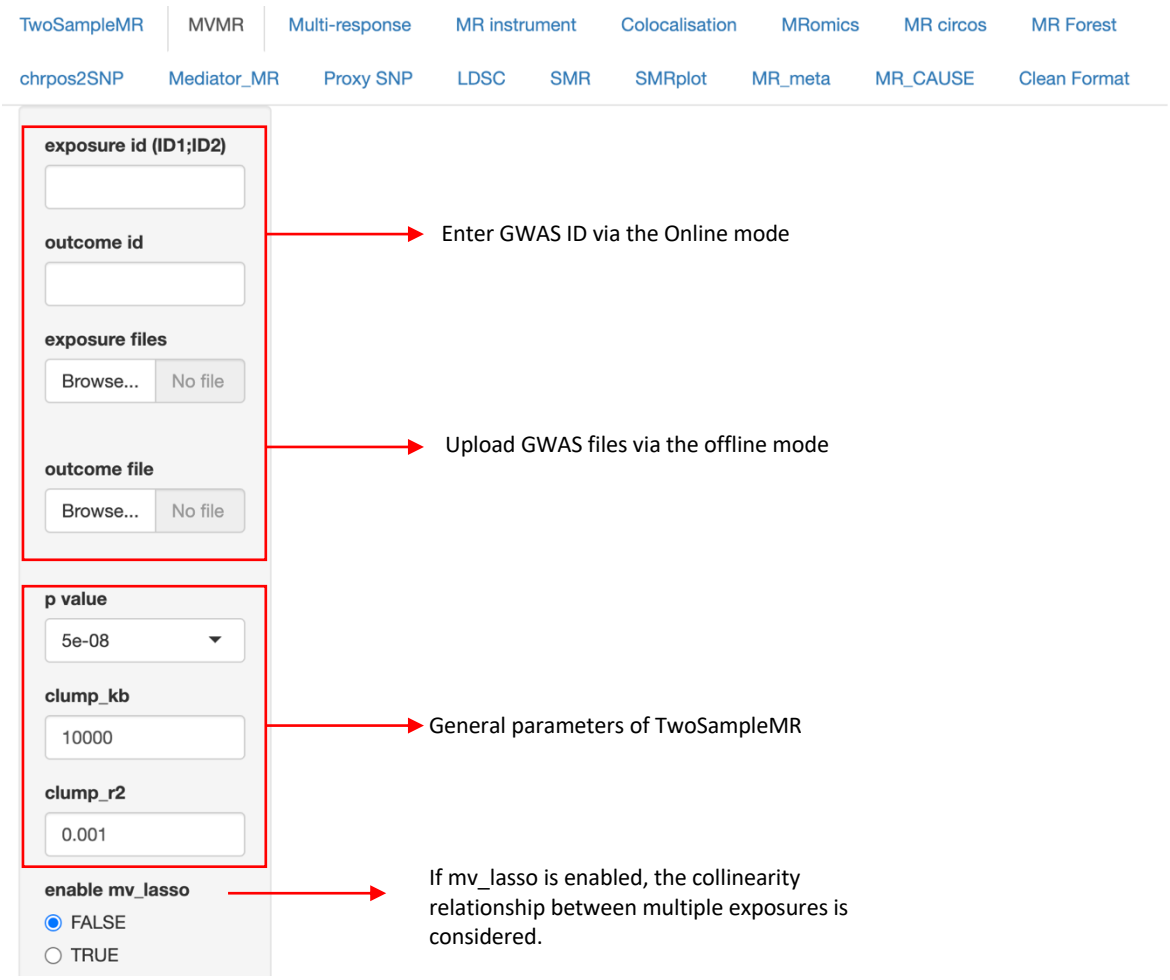
TwoSampleMR

In this module, we mainly implemented the interactive interface for TwoSampleMR R package. We also supplemented other functions, including mr presso and phenoscanner. Mr presso can call the MRPRESSO R package to evaluate horizontal pleiotropy. The phenoscanner function is used to evaluate if there are confounding factors in the exposure. Other parameters are consistent with the twosamplemr R package.



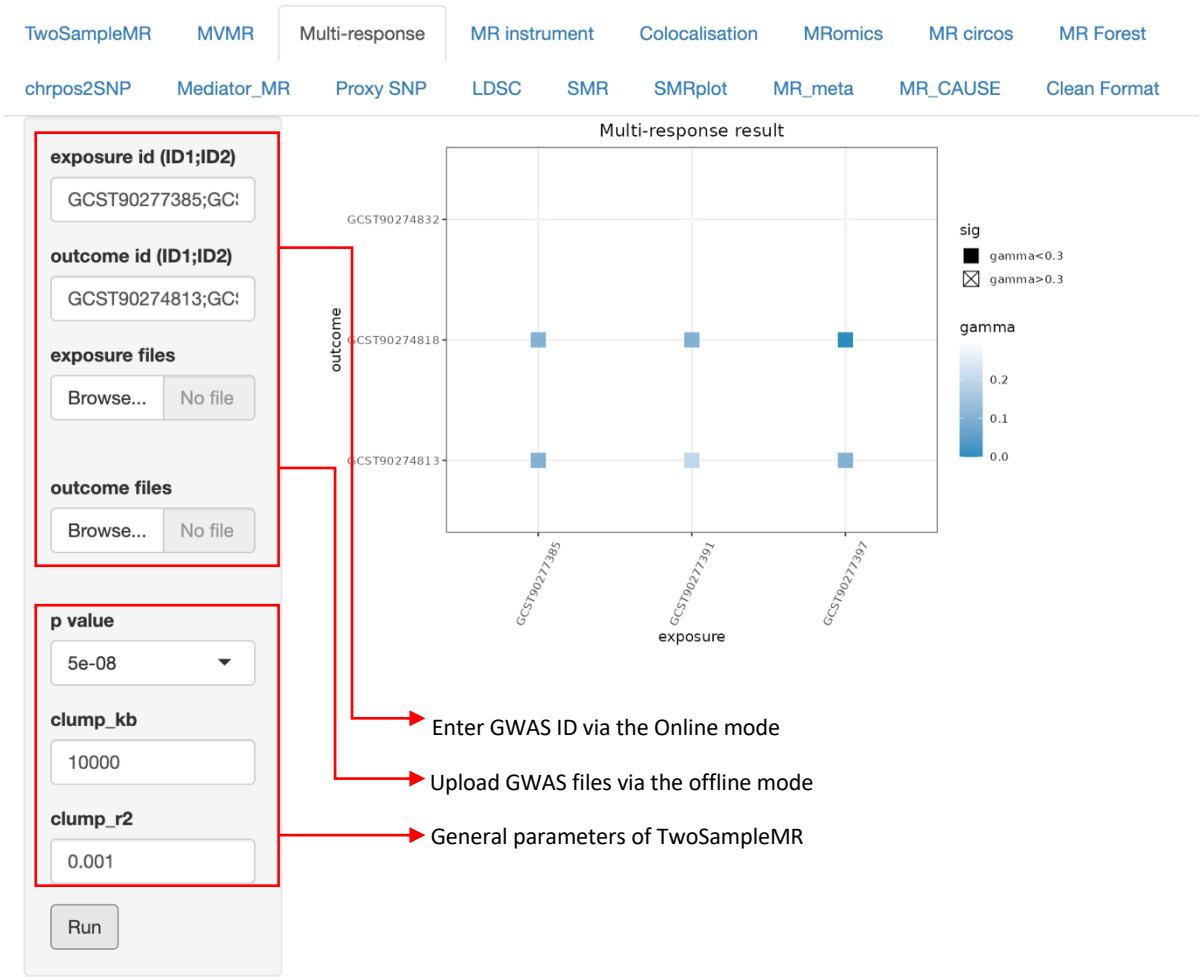
MVMR

In some cases, we need to conduct the causal inference between multiple exposures and outcome. These exposures may be genetically correlated. To address this issue, we need to use the mv_multiple function in the TwoSampleMR pacakge (MVMR). The parameter interface is similar to TwoSampleMR. Users can analyze in online or local file mode. For MVMR module, we also provide the mv_lasso function to eliminate collinearity between multiple exposure factors.



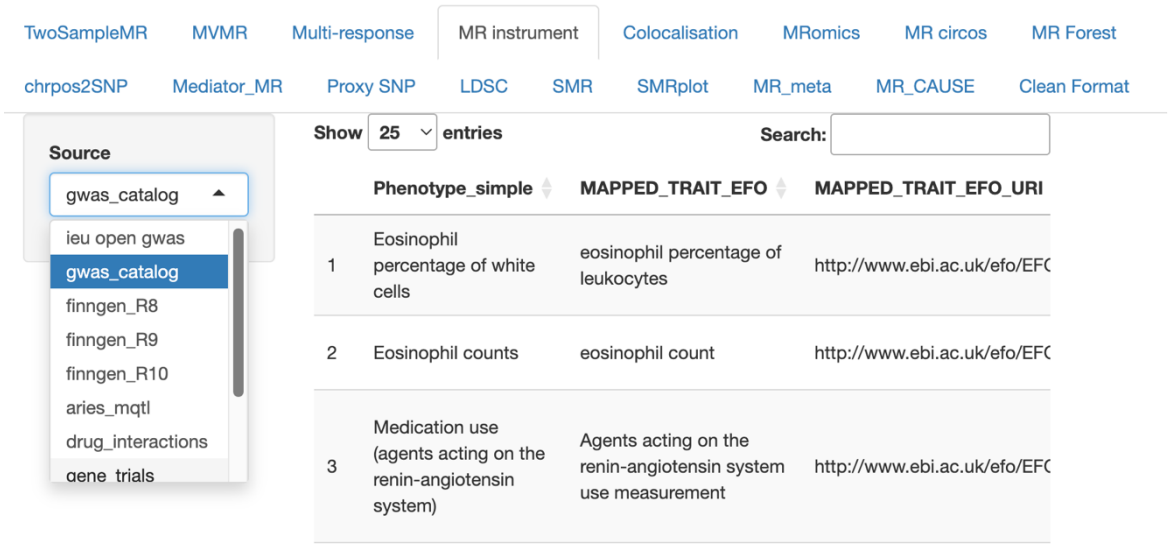
Multi-response

MVMR can only conduct the causal inference between multiple exposures and a single outcome. If we want to consider the relationship between multiple exposures and multiple outcomes, we can use the Multi-response module. This module invokes the MR2 algorithm (https://github.com/lb664/MR2/). The MR2 R package uses a Bayesian model to estimate the correlation between multiple exposures and multiple outcomes. Users can easily use this function via the interface on BiowinfordMR platform.



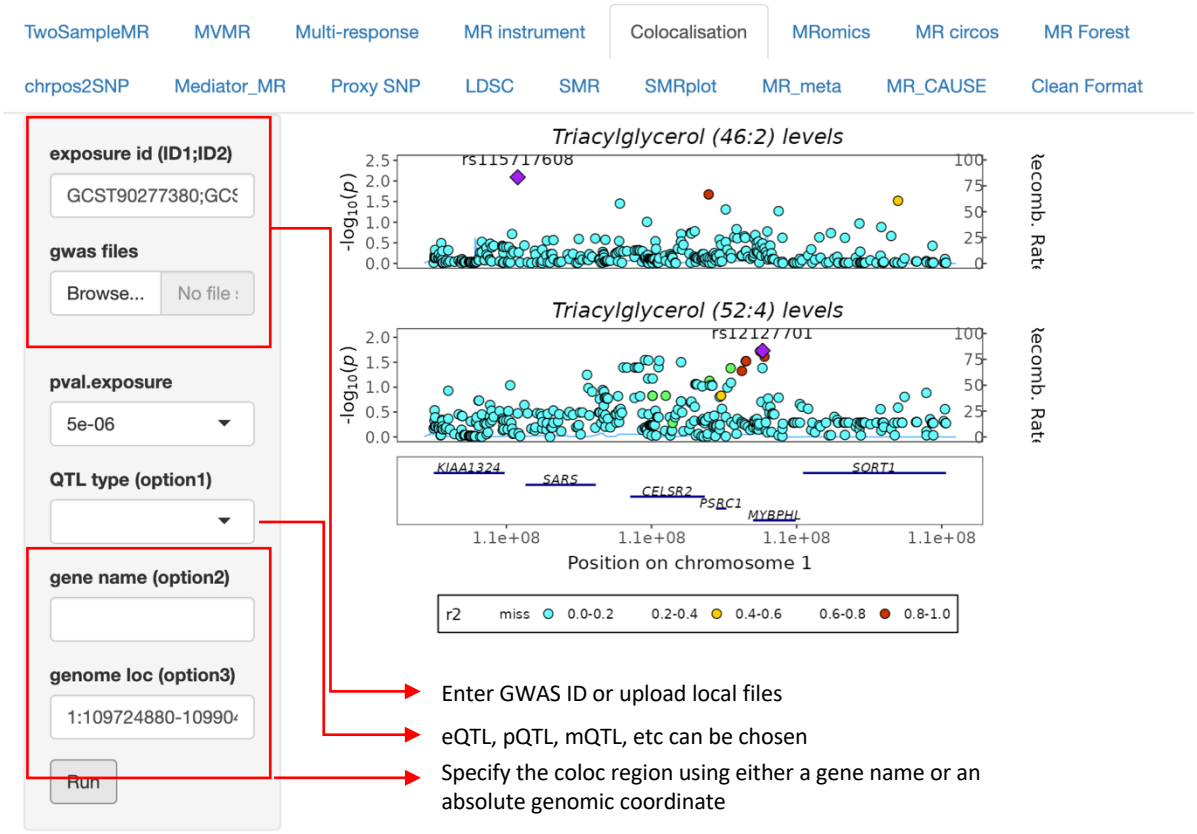
MR instrument

This is an R package that contains a number of data files from various sources to provide instruments in two sample MR (https://github.com/MRCIEU/MRInstruments). We have further supplemented the GWAS data from Finngen, including the versions from R8 to R10. Users can easily search for matching GWAS data in this module via entering keywords.



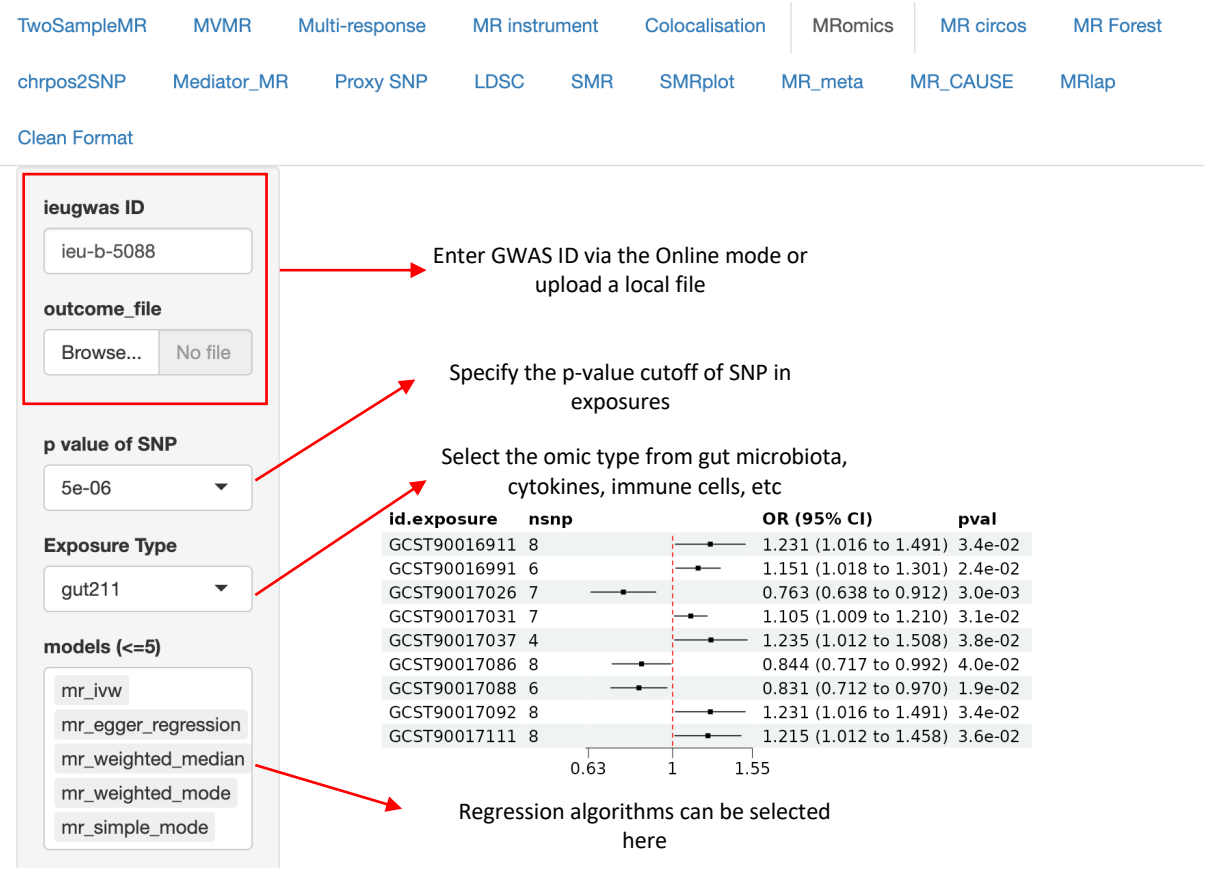
Colocalization

In this module, users can perform colocalization analysis by entering GWAS ID(s) or uploading local file(s). We provide various types of QTL data for users to choose from, including eQTL, pQTL, and mQTL. To facilitate users to specify the colocalization region, in addition to directly entering an absolute coordinate, users can also specify a genomic region by entering a gene symbol. If a user enters a gene symbol, the region used for colocalization will include the whole gene's region from start to end position, as well as the surrounding 2500kb regions.



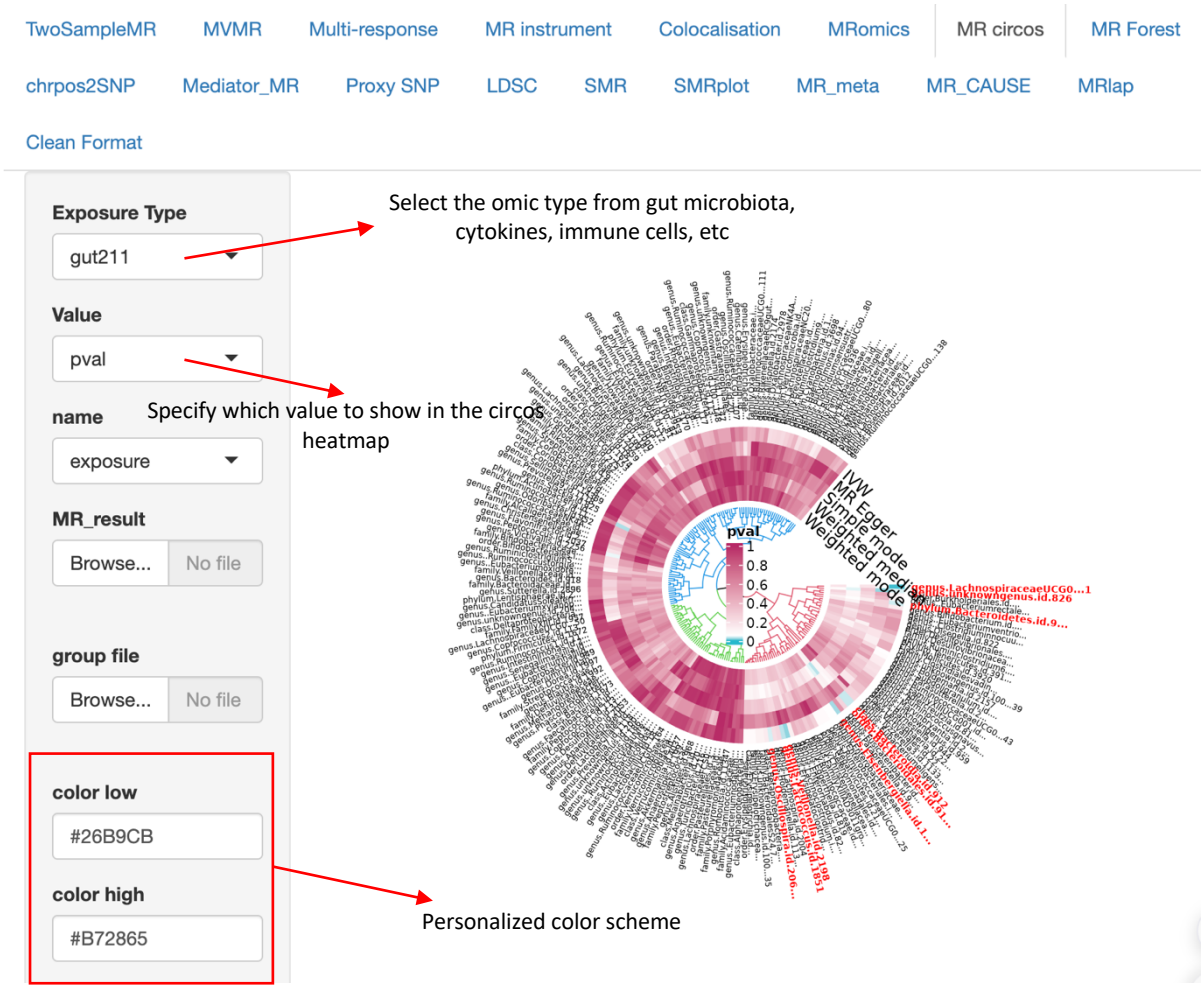
MRomics

The MRomics module currently contains more than 15 types of omics data involving over 7000 GWAS datasets. Such number keeps growing as the BioWinfordMR platform develops. These omics data consist of gut microbiome, oral microbiome, skin microbiome, cytokines factors, immune cells, lipids, serum metabolites, etc. Users only need to provide one GWAS ID or local file to conduct MR analysis in batches with the aforementioned omics data. To mitigate false positives from multiple testing, the original p-values will be adjusted using the Benjamini-Hochberg correction to generate adjusted p-values.



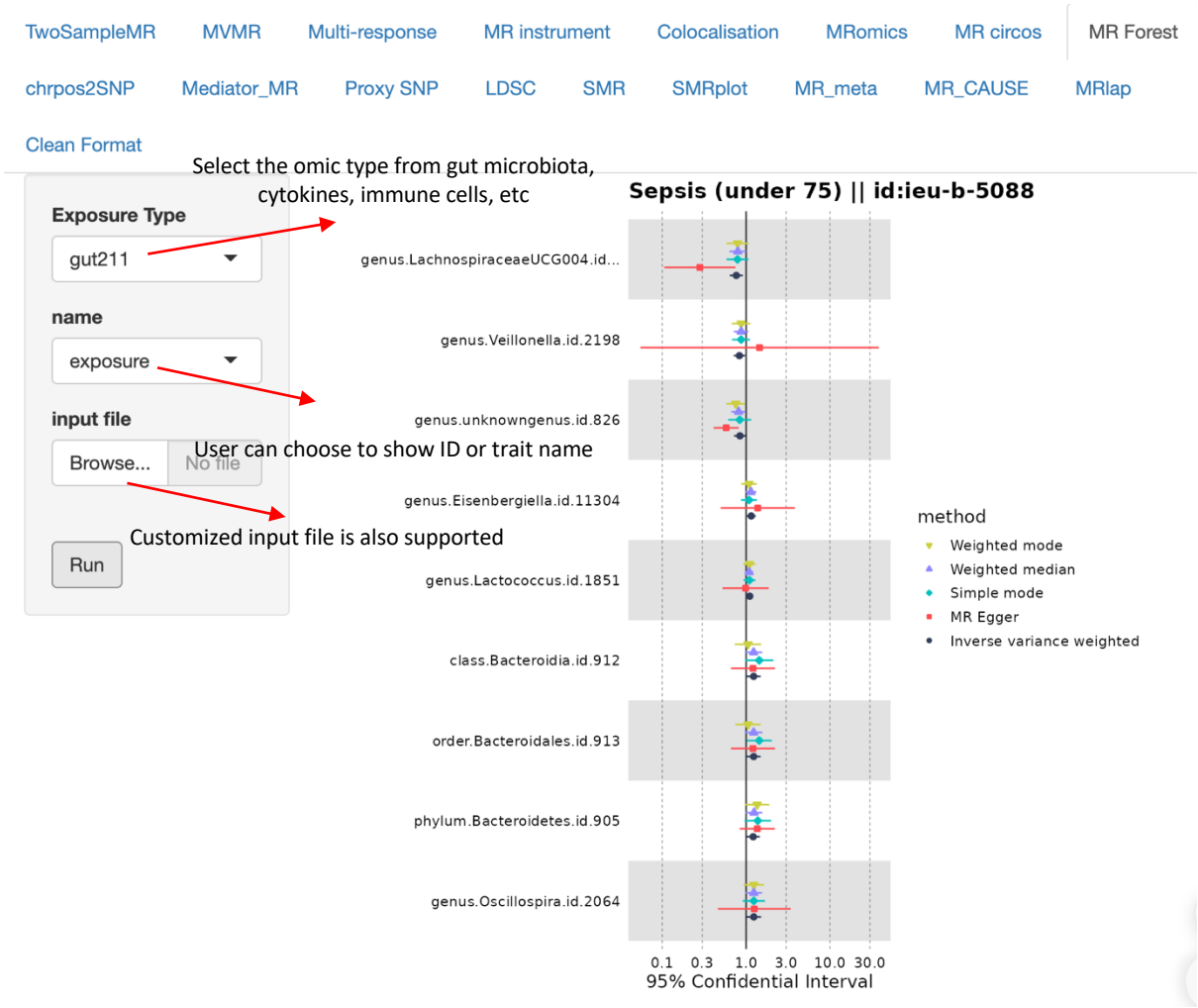
MR circos

This module can directly load the output results from MRomics for visualization analysis. Users can choose the indicators to display (beta, se, OR, p-value) and the type of labels (ID or exposure name). Additionally, this module supports user-defined color schemes.



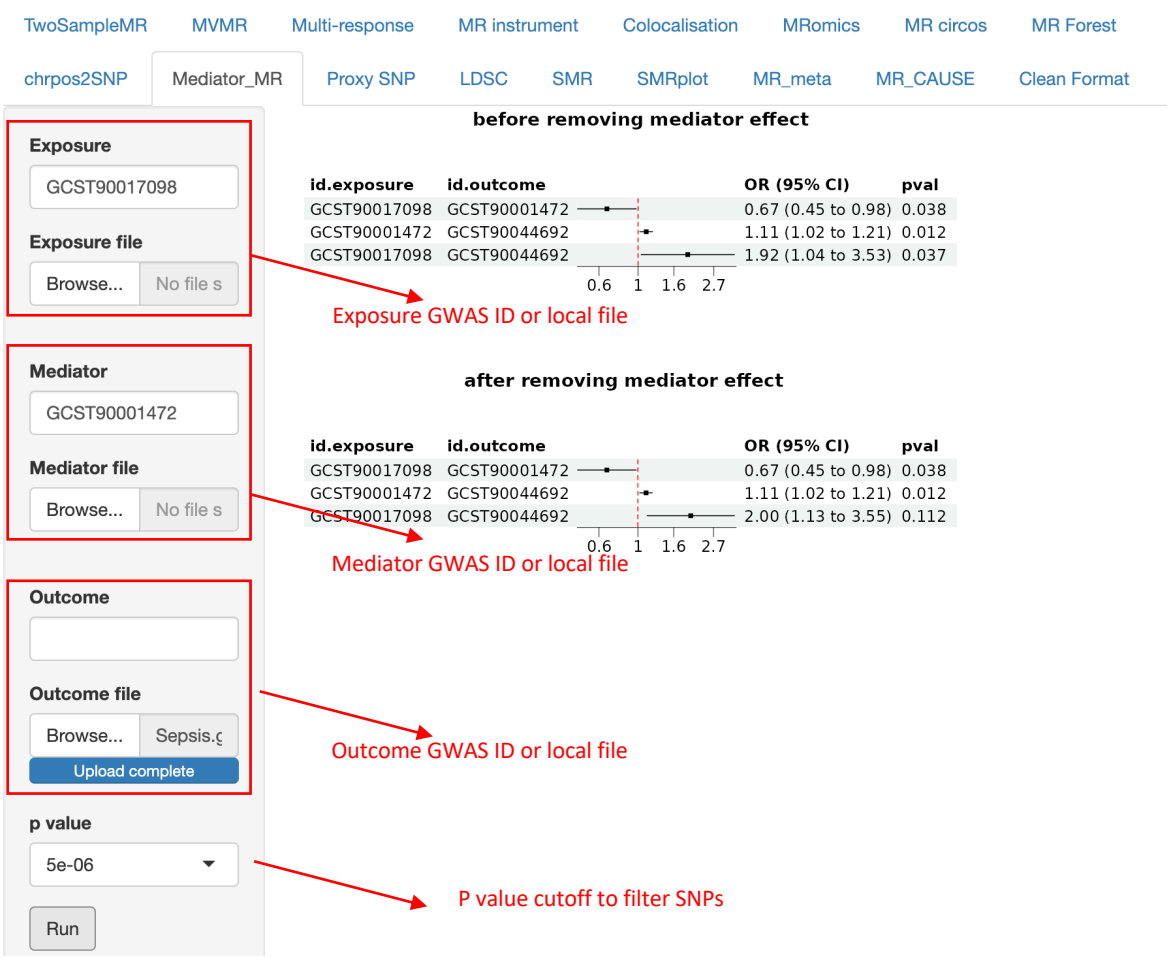
MR forest

This module can directly load the output results from MRomics for visualization analysis. Users can also upload the customized file to show forest graph. The file content and format should be consistent with the mr_result.txt file exported by MRomics.



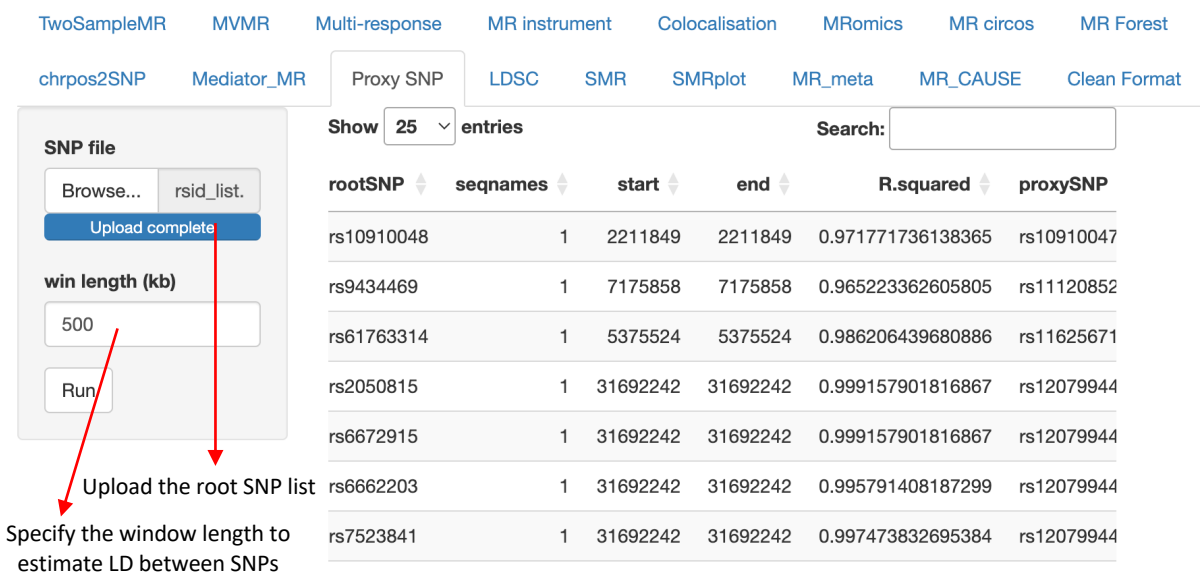
Mediator_MR

Although MR can infer causal relationships between two phenotypes, in most cases, the exposure does not directly act on the outcome but rather through certain mediators. The exposure first affects the mediator, which then influences the outcome. This mediator pathway is known as a complete mediator effect. To identify such mediator pathways like exposure-mediator-outcome, we need to utilize the Mediator_MR module. In the Mediator_MR module, we employ the two-step algorithm. The first step involves estimating the effects of exposure-mediator, mediator-outcome, and exposure-outcome. In the second step, after removing the mediator effect, the genetic relationship from exposure to outcome is recalculated.



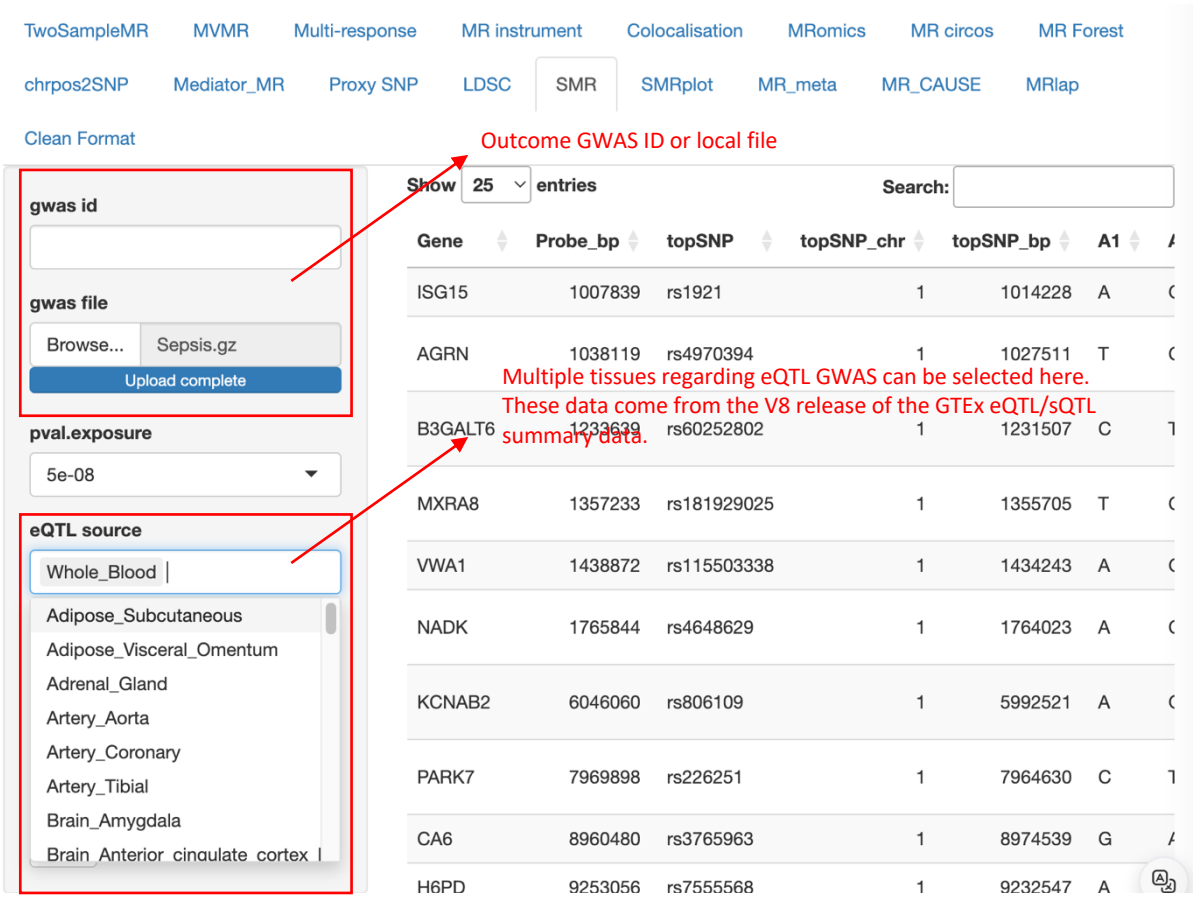
Proxy SNP

In the MR analysis process, sometimes we are limited by a small number of SNPs. In such cases, we can consider using proxy SNPs. In the Proxy SNP module, BioWinfordMR allow users to identify proxy SNPs in a specified window range that are in linkage disequilibrium with the root SNP. The code was retrieved from the Github repository of Kamil Slowikowski (https://gist.github.com/slowkow/3d13aa44cf4f65ca9ad2a0570346ba05?permalink_comment_id=2605197)



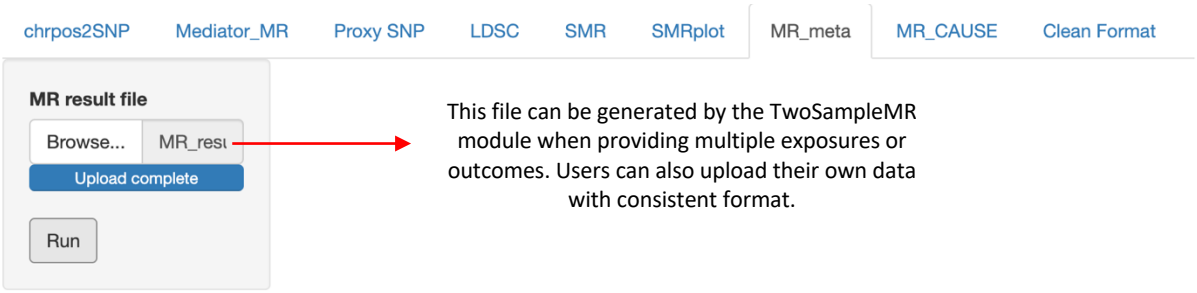
SMR

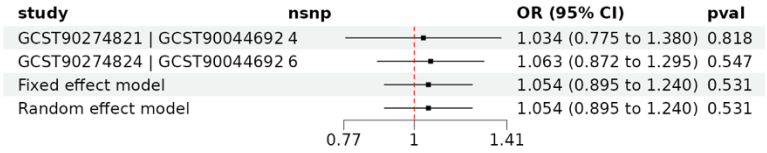
The SMR module allows users to estimate association between the expression level of a gene and a complex trait of interest using summary-level data from GWAS and expression quantitative trait loci (eQTL) studies (Zhu et al. 2016 Nature Genetics). The SMR & HEIDI methodology can be interpreted as an analysis to test if the effect size of a SNP on the phenotype is mediated by gene expression. This tool can therefore be used to prioritize genes underlying GWAS hits for follow-up functional studies. The methods are applicable to all kinds of molecular QTL (xQTL) data, including DNA methylation QTL (mQTL) and protein abundance QTL (pQTL).



MR_meta

For analyzing multiple exposures or multiple outcomes, in addition to considering the MVMR and multi-response modules, users can also consider conducting meta-analysis on multiple MR results via the meta R package. First, we utilize TwoSampleMR to individually analyze each pair of exposure and outcome. Subsequently, the results from multiple analyses are aggregated for comprehensive evaluation using both fixed-effect and random-effect models.






```
Number of studies: k = 2

                                        95%-CI    z p-value
Common effect model  0.0522 [-0.1110; 0.2154] 0.63  0.5309
Random effects model 0.0522 [-0.1110; 0.2154] 0.63  0.5309

Quantifying heterogeneity:
 tau^2 = 0; tau = 0; I^2 = 0.0%; H = 1.00

Test of heterogeneity:
    Q d.f. p-value
 0.02    1  0.8799

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
```

MR_cause

CAUSE is a Mendelian Randomization method using genome-wide summary statistics. CAUSE models correlated and uncorrelated horizontal pleiotropy in order to avoid false positives that can occur using other methods. You can find the tutorial here (https://jean997.github.io/cause/ldl_cad.html).

TwoSampleMR | MVMR | Multi-response | MR instrument | Colocalisation | MRomics | MR circos | MR Forest

chrpos2SNP | Mediator_MR | Proxy SNP | LDSC | SMR | SMRplot | MR_meta | MR_CAUSE | Clean Format

exposure file

Browse... diabetes

Upload complete

outcome file

Browse... Sepsis.g

Upload complete

p value

5e-06

clump_kb

10000

clump_r2

0.001

Run