

METAL Documentation

From Genome Analysis Wiki

Contents

- 1 Useful Wiki Pages
- 2 History
- 3 Brief Description
- 4 Approach
- 5 Basic Usage Instructions
 - 5.1 Help!
 - 5.2 Input File Separators
 - 5.3 Input File Columns
 - 5.4 Specifying Weights in P-value Based Analysis
 - 5.5 Reading Each Input File
 - 5.6 Performing the Final Analysis
- 6 Additional Analysis Options
 - 6.1 Selecting an Analysis Scheme
 - 6.2 Genomic Control Correction
 - 6.3 Sample Overlap Correction
 - 6.4 Strand Information
 - 6.5 Filtering
 - 6.6 Verbose Mode
 - 6.7 Lenient Mode
 - 6.8 Tracking Allele Frequencies
 - 6.9 Custom Variables
 - 6.10 Input File Recommendations
- 7 Example: A METAL Meta-Analysis Script

Useful Wiki Pages

There are a few pages in this Wiki that may be useful to METAL users. Here are links to key pages:

- [The METAL Home Page](#)
- [The METAL Quick Start Tutorial](#)
- [The METAL FAQ](#)
- [The METAL Command Reference](#)

History

METAL was developed by Goncalo Abecasis, Yun Li and Cristen Willer (manuscript available here (<http://www.sph.umich.edu/csg/abecasis/publications/pdf/Bioinformatics.vol.26-pp.2190.pdf>)). The first version was developed in 2007 and was used for the analyses presented in Sanna et al (2008) (<http://www.sph.umich.edu/csg/abecasis/publica>

tions/18193045.html) and Willer et al (2008) (<http://www.sph.umich.edu/csg/abecasis/publications/18193043.html>). Since then, it has become quite a popular tool for the analysis of genomewide association scans.

Brief Description

METAL is a tool for meta-analysis genomewide association scans. METAL can combine either (a) test statistics and standard errors or (b) p-values across studies (taking sample size and direction of effect into account). METAL analysis is a convenient alternative to a direct analysis of merged data from multiple studies. It is especially appropriate when data from the individual studies cannot be analyzed together because of differences in ethnicity, phenotype distribution, gender or constraints in sharing of individual level data imposed. Meta-analysis results in little or no loss of efficiency compared to analysis of a combined dataset including data from all individual studies.

Approach

One of the most common questions we receive is about the approach used by METAL to carry out a meta-analysis using p-values as input. The process is actually quite simple! First, for each marker, a reference allele is selected and a z-statistic characterizing the evidence for association is calculated. The z-statistic summarizes the magnitude and the direction of effect relative to the reference allele and all studies are aligned to the same reference allele. Next, an overall z-statistic and p-value are then calculated from a weighted sum of the individual statistics. Weights are proportional to the square-root of the number of individuals examined in each sample and selected such that the squared weights sum to 1.0. For samples that contain related individuals, a smaller ‘effective’ sample size may be used, but simulations suggest that modest changes in the effective sample size have very little impact on the final p-value.

Basic Usage Instructions

METAL is a command line tool. It is typically run from a Linux, Unix or DOS prompt by invoking the command `metal`. Analyses can be run interactively or a simple script can be provided as input. Interactive analyses are usually convenient when learning how to use METAL, whereas the scripting approach is preferred for production use (as it allows analyses to be conveniently repeated). An example METAL script is included at the bottom of this page.

METAL has lots of options and here we have listed some common ones that, hopefully, will help you get started.

Help!

Issuing the `HELP` command lists all available commands and the current settings for each option. The list of all available commands is also available in the METAL Command Reference.

Input File Separators

METAL expects that each set of results will be summarized in a table. This table must be stored in a text file but otherwise METAL is quite flexible about details such as column separators, column headers and the like. This does mean that an essential bit of information needed before any meta-analysis is a description of each input file.

The first thing you should specify is the column separator. By default, METAL assumes columns are separated by whitespace (which consists of any combination of space and tab characters). You can also specify:

SEPARATOR	WHITESPACE	- the default
SEPARATOR	COMMA	- for comma delimited files that are popular in some platforms
SEPARATOR	TAB	- columns separated by a single tab, so that consecutive tabs indicate an empty column

Input File Columns

Each input file should include the following information:

- A column with marker name, which should be consistent across studies
- A column indicating the tested allele
- A column indicating the other allele

If you are carrying out a sample size weighted analysis (based on p-values), you will also need:

- A column indicating the direction of effect for the tested allele
- A column indicating the corresponding p-value
- An optional column indicating the sample size (if the sample size varies by marker)

If you are carrying out a meta-analysis based on standard errors, you will need:

- A column indicating the estimated effect size for each marker
- A column indicating the standard error of this effect size estimate

The header for each of these columns must be specified so that METAL knows how to interpret the data. As noted below, additional columns including allele frequency information, strand information, and others can also be present.

Here is a typical set of commands that would describe a table where the headers SNP, RefAllele, NonRefAllele, Pvalue and Beta correspond to the MARKER, ALLELE 1 and 2, PVALUE and EFFECT columns:

MARKERLABEL	SNP
ALLELELABELS	RefAllele NonRefAllele
PVALUELABEL	P-value
EFFECTLABEL	Effect

These can be abbreviated as:

MARKER	SNP
ALLELE	RefAllele NonRefAllele
PVALUE	P-value
EFFECT	Effect

Specifying Weights in P-value Based Analysis

The weight for each MARKER can be stored in a column in the table (specified with the WEIGHTLABEL or WEIGHT commands). Most commonly, the weight will be the number of individuals contributing to that particular p-value.

WEIGHTLABEL	N
-------------	---

Alternatively, the same weight can be used for all markers for that inputfile (in which case the fixed weight can be set with the DEFAULTWEIGHT command). The WEIGHTLABEL command takes precedence over the DEFAULTWEIGHT command, so the WEIGHT column label in use must not match any columns in the inputfile.

```
WEIGHTLABEL      DONTUSECOLUMN
DEFAULTWEIGHT    1000
```

Reading Each Input File

Once all appropriate headers have been specified, issuing the `PROCESS` command will read an input file and update summary statistics to take the results it contains into account. Thus:

```
PROCESS          study1-results.tbl
```

Performing the Final Analysis

Once all input files have been processed, simply issue the `ANALYZE` command to execute a meta-analysis. If you'd like to execute interim analysis that include only a subset of the studies, issue the `ANALYZE` command after the corresponding input files have been processed.

```
ANALYZE
```

To allow for heterogeneity, use the `ANALYZE HETEROGENEITY` command. This command will take a little longer to run, because it requires each input file to be examined twice. The METAL heterogeneity analysis requires a second pass of analysis to decide whether observed effect sizes (or test statistics) are homogeneous across samples. The resulting heterogeneity statistic has $n-1$ degrees of freedom for n samples.

```
ANALYZE HETEROGENEITY
```

METAL does not require that all input files report a result for every marker. Any available data is used. To restrict the output to only markers that have at least a specific number of individuals analysed (or weight), use a command like the following:

```
MINWEIGHT 10000
```

For example to restrict the output to show only Markers with a total sample size of at least 10,000 individuals.

Additional Analysis Options

Selecting an Analysis Scheme

```
SCHEME SAMPLESIZE - default approach, uses p-value and direction of effect, weighted according to sample size
SCHEME STDERR     - classical approach, uses effect size estimates and standard errors
STDERR SE         - specify the label for the standard error column.
```

By default, METAL combines p-values across studies taking into account a study specific weight (typically, the sample size) and direction of effect. This behavior can be requested explicitly with the `SCHEME SAMPLESIZE` command. An alternative can be requested with the `SCHEME STDERR` command and weights effect size estimates using the inverse of the corresponding standard errors. To enable this option, you will also need to specify which of your input columns contains standard error information using the `STDERRLABEL` command (or `STDERR` for short). While standard error based weights are more common in the biostatistical literature, if you decide to use this

approach, it is very important to ensure that effect size estimates (*beta* coefficients) and standard errors use the same units in all studies (i.e. make sure that the exact same trait was examined in each study and that the same transformations were applied). Inconsistent use of measurement units across studies is the most common cause of discrepancies between these two analysis strategies.

Genomic Control Correction

```
GENOMICCONTROL OFF      - the default, no adjustment to test statistics
GENOMICCONTROL ON       - automatically correct test statistics to account for small amounts of population stratification
GENOMICCONTROL [value] - correct test statistics using the specified inflation factor
```

METAL has the ability to apply a genomic control correction to all input files. METAL will estimate the inflation of the test statistic by comparing the median test statistic to that expected by chance, and then apply the genomic control correction to the p-values (for SAMPLESIZE weighted meta-analysis) or the standard error (for STDERR weighted meta-analysis). This should only be applied to files with whole genome data (i.e. should not be used for settings where results are only available for a candidate locus or a small number of SNPs selected for follow-up of GWAS results). Genomic control settings can be customized for each input file. We recommend applying genomic control correction to all input files that include genomewide data and, in addition, to the meta-analysis results. To apply genomic control to the meta-analysis results, just perform an initial meta-analysis and then load the initial set of results into METAL to get final, genomic control adjusted results.

Sample Overlap Correction

Correction for sample overlap in sample size weighted meta-analysis (developed by Sebanti Sengupta and implemented by Daniel Taliun).

First, METAL estimates the number of individuals that are common among two or more studies based on Z-statistics from each study. Then, METAL adjusts for sample overlap when calculating overall Z-statistics by correcting the weights with the estimated number of individuals in common.

To enable correction for sample overlap in your sample size weighted meta-analysis, use OVERLAP ON command (valid only with SCHEME SAMPLESIZE). By default, METAL uses Z-statistics <1 for estimating the number of individuals that are common among studies. To change this threshold, use ZCUTOFF [number] command.

More information on the method can be found in:

- Method overview and results
- Full method description (current draft, manuscript est. 2018)

Strand Information

```
USESTRAND ON
STRANDLABEL StrandColumnHeading
```

Input files can contain a column that indicates which strand the alleles are coded on (given as +/-). If this column is present, you should issue the USESTRAND ON command and specify an appropriate header with the STRANDLABEL command. If USESTRAND is off, the strand is assumed to be “+” for all SNPs, although obvious strand problems are identified by METAL and appropriately handled (for example, when one study provides A/G alleles and a different study provides C/T alleles).

Filtering

Custom filters can be used to select SNPs for inclusion in the meta-analysis. This can be used, for example, to select SNPs within a specified minor-allele frequency range for analysis.

Here are some possible filters:

```
ADDFILTER N > 1000
ADDFILTER MAF > 0.01
```

Together, these two filters would only consider entries where the value in the N column is greater than 1000 and the value in the MAF column is also greater than 0.01.

Filters can be defined using the <, >, <=, >=, =, != and IN operators. The IS operator tests membership in a set. For example to restrict analysis to three interesting SNPs, use (*note absence of spaces in list of SNPs*):

```
ADDFILTER MARKER_ID IN (rs1234,rs123456,rs123)
```

To remove all previously defined filters, use the command:

```
REMOVEFILTERS
```

Verbose Mode

```
VERBOSE ON
```

METAL allows for complete output of individual summary statistics for all SNPs in all input files. This can create a very large file and should be used with caution. Typically, one should create custom filters to restrict analyses to interesting SNPs of interest before using this option. This option can be useful for comparing direction of effect across many studies since METAL takes care of all the strand flipping and provides the direction of effect relative to the same allele. This is also a way to double-check that the expected data are being used appropriately by METAL.

Lenient Mode

```
COLUMNCOUNTING STRICT      - requires expected number of columns in every row
COLUMNCOUNTING LENIENT     - tries to interpret rows with fewer columns than expected
```

By default, METAL will skip lines in each input file that don't have the expected number of columns. This is usually a good idea because it avoids producing incorrect results when a column is missing. Sometimes (for example, when there are optional extra columns at the end of each line), the `COLUMNCOUNTING LENIENT` option can be useful.

Tracking Allele Frequencies

```
AVERAGEFREQ ON
MINMAXFREQ ON
```

METAL can optionally track the effect allele frequency across all files and report the mean, minimum and maximum effect allele frequency. These can be quite useful to check that allele frequencies are similar across different cohorts after METAL performs all strand alignment. Large differences in allele frequencies across studies can suggest inconsistent naming of reference alleles across studies. METAL requires all input files to have an allele frequency column when this feature is turned on. To specify the column header for allele frequency information, use the `FREQLABEL` command.

Custom Variables

We allow users to keep cumulative counts of custom variables across input files. An example of this might be to keep track of the sample size when performing standard-error weighted meta-analysis. The name of the custom variable should be defined once, before input files are loaded. The corresponding column label in each input file can be specified using the `LABEL` command. For example, to create a custom variable labeled `TotalSampleSize` that tallies the total of the `N` column across files, one could issue the commands:

```
CUSTOMVARIABLE TotalSampleSize
LABEL TotalSampleSize as N
```

If needed, the `LABEL` command can be used multiple times to customize column headers for each input file.

Input File Recommendations

We strongly recommend that both allele labels, corresponding to the the effect allele and non-effect allele, should be provided for all SNPs. As long as both allele columns are given for each input file, METAL appropriately accounts for situations when different input files use different reference alleles. Alleles can be coded numerically (`A=1,C=2,G=3,T=4`) or alphabetically (`A,C,G,T,a,c,g,t`) and can be on either strand if not an A/T or C/G SNP. For A/T or C/G SNPs, METAL requires SNPs to be on a consistent strand in different input files for the results to be interpretable. For other SNPs, METAL can automatically identify and resolve strand inconsistencies.

P-values that are < 0.0 , > 1.0 or non-numeric will be treated as missing and generate a warning.

The `EFFECT` column can have positive and negative values (beta values from regression, for example), or simply directions of effect relative to the reference allele, listed as “+” and “-“. An `EFFECT` of “+” (or any positive number) with respect to the reference allele A (or effect allele A), for example, represents a case where increasing number of copies of allele A are correlated with increasing trait values. For discrete traits, it is common to report odds ratios, which are always positive. In this case, to calculate the direction of effect, one should look at the log of the odds ratio. METAL can compute the odds ratio for you if you specify `EFFECT log(ODDS_RATIO_COLUMN)`

To perform odds-ratio based meta-analysis, select `SCHEME STDERR` at the beginning of the script. Then, for each file, provide the natural log of the odds ratio as the `EFFECT` column or another appropriate statistic (such as the corresponding regression coefficient from a logistic regression analysis).

Example: A METAL Meta-Analysis Script

```
#THIS SCRIPT EXECUTES AN ANALYSIS OF EIGHT STUDIES
#THE RESULTS FOR EACH STUDY ARE STORED IN FILES Inputfile1.txt THROUGH Inputfile8.txt

#LOAD THE FIRST EIGHT INPUT FILES

# UNCOMMENT THE NEXT LINE TO ENABLE GenomicControl CORRECTION
# GENOMICCONTROL ON
```

```
# === DESCRIBE AND PROCESS THE FIRST INPUT FILE ===
MARKER SNP
ALLELE REF_ALLELE OTHER_ALLELE
EFFECT BETA
PVALUE PVALUE
WEIGHT N
PROCESS inputfile1.txt

# === THE SECOND INPUT FILE HAS THE SAME FORMAT AND CAN BE PROCESSED IMMEDIATELY ===
PROCESS inputfile2.txt

# === DESCRIBE AND PROCESS THE THIRD INPUT FILE ===
MARKER SNP
ALLELE A_REF OTHER_ALLELE
EFFECT BETA
PVALUE pvalue
WEIGHT N
PROCESS inputfile3.txt

# === DESCRIBE AND PROCESS THE FOURTH INPUT FILE ===
MARKER MARKERNAME
ALLELE EFFECTALLELE NON_EFFECT_ALLELE
EFFECT EFFECT1
PVALUE PVALUE
WEIGHT NONMISS
PROCESS inputfile4.txt

# === CARRY OUT AN INTERIM ANALYSIS OF THE FIRST FOUR FILES ===
OUTFILE METAANALYSIS_inputfile1to4_ .tbl
ANALYZE

# LOAD THE NEXT FOUR INPUT FILES

# === DESCRIBE AND PROCESS THE FIFTH INPUT FILE ===
MARKER rsid
ALLELE EFFECT_ALLELE OTHER_ALLELE
EFFECT BETA
PVALUE Add_p
WEIGHT total_N
SEPARATOR COMMAS
PROCESS inputfile5.txt

# === THE SIXTH INPUT FILE HAS THE SAME FORMAT AND CAN BE PROCESSED IMMEDIATELY ===
PROCESS inputfile6.txt

# === DESCRIBE AND PROCESS THE SEVENTH INPUT FILE ===
ALLELE ALLELE OTHER_ALLELE
MARKER SNP
EFFECT BETA
PVALUE PVALUE
WEIGHT N
SEPARATOR WHITESPACE
PROCESS inputfile7.txt

# === DESCRIBE AND PROCESS THE EIGHTH INPUT FILE ===
ALLELE BETA_ALLELE OTHER_ALLELE
MARKER SNP
EFFECT BETA
PVALUE P_VAL
WEIGHT N
PROCESS inputfile8.txt

#for the final meta-analysis of all 8 samples only output results if the
#combined weight is greater than 10000 people

OUTFILE METAANALYSIS_inputfile1-8_ .tbl
MINWEIGHT 10000
ANALYZE

QUIT
```


- This page was last edited on 22 December 2017, at 15:52.
- This page has been accessed 201,375 times.