METAL

Goncalo Abecasis and Cristen Willer, 2007

METAL is a tool for performing meta-analysis of p-values from two or more individual studies. Metal creates a single summary p-value from studies which could not be analyzed together because of differences in ethnicity, phenotype distribution, gender, inability to share individual-level data, or any other reason.

For each marker, an arbitrary reference allele is selected and a z-statistic characterizing the evidence for association is used as input. The z-statistic summarizes the magnitude and the direction of effect relative to the reference allele. An overall z-statistic and p-value are then calculated from the weighted average 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. If a sample contains related individuals, a smaller 'effective' population size may be used, but simulations suggest that modest changes in the effective sample size have very little impact on the final p-value.

Usage instructions;

METAL is invoked with the command 'metal' and allows for analysis to be performed interactively. A convenient alternative is to save all commands into a single text file which can be provided as input. An example is at the bottom of this document.

METAL allows for a variety of tabular formats in the input files, but the following information must be provided for each marker in each file;

Mandatory input for each marker;
- Marker name
- Reference allele (also known as the 'effect allele') and the non-reference allele
- P-value

- Weight (sample size)
- Direction of effect relative to reference allele

Tables must have column headers that specify where the mandatory input can be found. The default name for the Marker column is 'MARKER', but can be changed to match the relevant input file column with the following command;

> MARKER   SNP

Similarly, the reference allele column, P-value column and effect column can be changed to match the input file;

> ALLELE   myRefAllele   myNonRefAllele
> PVALUE   myPVALUE
> EFFECT   myBETA

We strongly recommend that both allele labels, corresponding to the the effect allele and non-effect allele, respectively, are given for all SNPs. Alleles can be numeric (1,2,3,4) or alphabetical (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 A/C, A/G, C/T, and G/T SNPs, METAL will flip the strand the alleles are on if not consistent between input files and METAL will output results with respect to the lowest numeric reference allele (see Examples 1, 2, and 3, below). If all files are consistent (for example, using the HapMap allele naming conventions), the strand of the alleles is left alone. 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.

P-values of 0 or any other non-numeric value are assumed to be missing. Missing values are tolerated and a meta-analysis p-value will include results from any input file with

non-missing values, even if only one input file has a p-value for this marker (see MINWEIGHT below for exclusion of markers with a small combined N).

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.

The weight for each MARKER can be assigned using a column;
> WEIGHTLABEL mySampleSizeColumn

Or;
> WEIGHT   mySampleSizeColumn

Or the default weight for the entire file can be specified with the following command;
E.g., if you have a sample size of 2000
> DEFAULTWEIGHT   2000

METAL assumes your files are tab or space delimited (or a mixture), but allows for comma-separated files using the following command;

> SEPARATOR   commas

Once the appropriate WEIGHT, MARKER, PVALUE and EFFECT labels are defined, load an input file;
> PROCESS   myresults_bmi.txt

METAL does not require that all input files have a p-value result to calculate a meta-analysis p-value. Any available data is used. To restrict the output to only markers that have at least a specific weight (number of individuals), then use;

> MINWEIGHT 10000

For example to restrict the output to show only Markers with at least 10,000 individuals.

Once all input files have had their column names defined and been loaded, then define your output filename (optional) and analyze!

> OUTPUTFILE   myoutputfilename

> ANALYZE

Example 1; Strand flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	T/G	+	a/c	+
Input file 2	T/G	+	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Example 2; Reference allele flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	C/A	-	a/c	+
Input file 2	C/A	-	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Example 2; Strand flips, numeric flips, and reference allele flips required

	ALLELES	EFFECT	ALLELES Analyzed	EFFECT Analyzed
Input file 1	G/T	-	a/c	+
Input file 2	2/1	-	a/c	+
Input file 3	A/C	+	a/c	+
Output			a/c	+

Example text file to run metal;

```
# THIS FILE EXECUTES AN ANALYSIS OF ALL AVAILABLE INFORMATION

mkdir output-metal

metal << EOT

# loading in the first half of inputfiles
MARKER   SNP
ALLELE   REF_ALLELE OTHER_ALLELE
EFFECT   BETA
WEIGHT   N
PVALUE   PVALUE
PROCESS inputfile1.txt
PROCESS inputfiles2.txt
PVALUE   pvalue
ALLELE   A_REF OTHER_ALLELE
MARKER   SNP
EFFECT   BETA
WEIGHT   N
PROCESS inputfile3.txt
MARKER   MARKERNAME
ALLELE   EFFECTALLELE NON_EFFECT_ALLELE
EFFECT   EFFECT1
WEIGHT   NONMISS
PVALUE   PVALUE
PROCESS inputfile4.txt
# meta-analysis can be performed at any stage
# and will include inputfiles 1-4
OUTFILE METAANALYSIS_inputfile1to4_ .tbl
ANALYZE

# load the second half of inputfiles
MARKER   rsid
ALLELE   EFFECT_ALLELE OTHER_ALLELE
EFFECT   BETA
WEIGHT   total_N
PVALUE   Add_p
SEPARATOR   COMMAS
PROCESS inputfile5.txt
PROCESS inputfile6.txt
ALLELE   ALLELE OTHER_ALLELE
MARKER   SNP
EFFECT   BETA
WEIGHT   N
PVALUE   PVALUE
SEPARATOR WHITESPACE
PROCESS inputfile7.txt
ALLELE   BETA_ALLELE OTHER_ALLELE
PVALUE   P_VAL
MARKER   SNP
EFFECT   BETA
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
EOT
```