PreMeta

GENERAL INFORMATION

PreMeta is a software program written in C++ that is designed to facilitate the exchange of information between four software packages for meta-analysis of rare-variant associations: MASS, RAREMETAL, MetaSKAT, and seqMeta. PreMeta has two related purposes: one is to allow the use of different software packages within the same consortium; and the second is to eliminate the need to recalculate summary statistics when investigators join a new consortium that has adopted a different software package.

Each meta-analysis pipeline conducts two separate steps: (1) calculation of summary statistics for each sequencing study; and (2) combination of summary statistics to perform gene-level association tests (i.e., actual meta-analysis). The output files of summary statistics from the four software packages have different formats. Specifically, MASS uses one text file to report all summary statistics. RAREMETAL uses two text files: one contains score statistics and SNP-level information; and the second contains between-SNP covariances by sliding windows. MetaSKAT uses .MSSD and .MInfo files: .MSSD is a binary file with between-SNP covariances; and .MInfo is a text file with information on studies and SNP sets. seqMeta uses an R object to report all summary statistics. PreMeta converts the format of any software output file to the format of any other software output file and thus allows the summary statistics from any one package to be used by any other package for meta-analysis.

The summary statistics pertain to score statistics. The older version of RAREMETAL (e.g., v0.4.0) does not normalize the score statistics by residual variances. Thus, the summary statistics for RAREMETAL (v0.4.0) cannot be directly combined with the summary statistics for the other three packages. PreMeta normalizes the score statistics from RAREMETALWORKER (v0.4.1) (the stand-alone study-level software used by RAREMETAL (v0.4.0)) by the estimated residual variances. The resulting score statistics can then be combined with the score statistics from the other three packages to perform meta-analysis.

The RAREMETAL pipeline is uniquely designed to estimate the covariances for SNPs within sliding windows in study-level analysis via RAREMETALWORKER. The sliding-window covariance estimates cannot be derived from gene-based covariance matrices if SNPs lie in different genes. Thus, gene-based summary statistics generated by operators other than RAREMETALWORKER are not informative enough to recover the sliding-window summary statistics required by RAREMETAL. When PreMeta reformats gene-based summary statistics from other operators for the RAREMETAL pipeline, the between-SNP covariance is set to 0 if the two SNPs do not belong to the same gene. This workaround will produce the correct covariance information in the meta-analysis if the same gene annotation is used to generate gene-based summary statistics and perform meta-analysis (since the covariances between different genes are not used at the end).

SYNOPSIS

PreMeta is a C++ executable program that is run from the command-line:

```
premeta --script SCRIPT_FILE --software META_SOFTWARE
[--version NUM] [--output_dir OUTPUT_DIRECTORY]
```

The **script** argument specifies the script file to use. The script file is described in detail below; it contains a list of the output files from the study-level analyses, as well as the name and the version number of the pipeline used for those files. The **software** argument specifies the meta-analysis software, and should be one of: {MASS, METASKAT, RAREMETAL, SEQMETA}. The **version** argument is optional, and is ignored if software is METASKAT or SEQMETA (in these cases, output will be consistent with version 0.40 for MetaSKAT or 1.5 for seqMeta). Mass is only supported for versions 7.0 and higher. For RareMetal, PreMeta will work on input versions 0.4.1 and 4.13.5, but will only *output* files consistent with the format of 4.13.5 (i.e. PreMeta will not generate output consistent with RareMetal versions older than 4.13.5). The **output_dir** is optional, and specifies the directory to place PreMeta's generated output files.

SCRIPT FILE

Figure 1 shows an example of the PreMeta script file. Lines starting with # are treated as a comment and ignored. The keyword **SOFTWARE** indicates the software that was used to generate the file(s) for the study. **The SOFTWARE keyword should appear before all other keywords for each output study.** The keyword **VERSION** indicates the version of the software; as with the version command-line argument, the version number is primarily ignored (except to distinguish between version 0.4.0 of RareMetal from version 4.13.5). The rest of the keywords describe the files that were generated by each output study. The specific keywords used depends on the software that was used to generate the output files (see Figure 1 below). For each file, the name of the file is mandatory, and then the next two keywords (specifying the *delimiter* used for that file and *comment character* for the file) are optional (default values for these, specified below, will be used if not overridden by the appropriate keyord). Additionally, for each MASS or SEQMETA study that is used as input for PreMeta, if the target software is RAREMETAL or METASKAT, we require the user provide an annotation file that specifies the Reference/Alternative alleles for all SNP positions in the study. This file should have one SNP per line, and contain only three columns:

CHR:POS   REF   ALT

(more columns are permissible, but those three columns must be present, and if there are extra columns, there must be a header line that starts with a comment character and then lists all of the column names in order, among which "CHR:POS", "REF", and "ALT" must be present).

One needs to make sure that the format of the SNP IDs (rs# or chr:pos) are consistent across studies. The SNP ID in the RAREMETAL pipeline takes the form chr:pos. Therefore, if any of those files are generated from the RAREMETAL pipeline, then the SNP ID format should be chr:pos across all studies.

The names of the output files can also be specified in the script file. The keyword X_OUT_FILE shoud be used for each output file that you wish to give a special name to, where X is one of: {MASS, SCORE, COV, MINFO, MSSD, RDATA}. *Note that the choice of which keyword* X_OUT_FILE *to use depends on the* **target software**, *not the study's input software.* Figure 1 below shows an example script file to be used if the target software is RareMetal (hence the keywords SCORE_OUT_FILE and COV_OUT_FILE appear in *every* study).

Figure 1 below shows a sample script file with four studies. The one text file for the first study was generated by the MASS pipeline; the two text files for the second study were generated by the RAREMETAL pipeline; the .MSSD and .MInfo files for the third study were generated by the MetaSKAT pipeline; and the .Rdata file for the fourth study was generated by the seqMeta pipeline. Bold lines are required, normal-font lines are optional (the ALLELE_FILE lines are required if the target software is RAREMETAL or METASKAT). The RESCALE lines are used to specify factors to rescale the summary statistics under inverse normal transformation.

Script File: MASS Inputs

If "SOFTWARE = MASS", then MASS's output file should be specified via the keyword FILE. Additional lines for the comment character (keyword FILE_COMMENT) and delimiter (keyword FILE_SEP) are optional (default values are "#" and "\t", resp.). For a detailed description of the MASS file, refer to SCORE-Seq: http://dlin.web.unc.edu/software/score-seq/ or SCORE-SeqTDS: http://dlin.web.unc.edu/software/score-seqtds/. As mentioned above, if the target software is RAREMETAL or METASKAT, an annotation file must be specified via keyword ALLELE_FILE; and then two lines describing the annotation file's delimiter (keyword ALLELE_SEP)

```
## === THE FIRST STUDY: SUMMARY STAT === ##
SOFTWARE = MASS
VERSION = 7.0
RESCALE = 1.0
FILE = path/study1.txt
FILE_SEP = “\t”
FILE_COMMENT = “#”
ALLELE_FILE = path/allele_info.txt
ALLELE_SEP = “\t”
ALLELE_COMMENT = “#”
SCORE_OUT_FILE = path/raremetal_scores.txt
COV_OUT_FILE = path/raremetal_covariances.txt

## === THE SECOND STUDY: SUMMARY STAT === ##
SOFTWARE = RAREMETAL
VERSION = 0.4.0
RESCALE = 1.0
FILE_SCORE = path/study2_score.txt
FILE_SCORE_SEP = “\t”
FILE_SCORE_COMMENT = “#”
FILE_COV = path/study2_cov.txt
FILE_COV_SEP = “\t”
FILE_COV_COMMENT = “#”
FILE_GROUP = path/group.txt
FILE_GROUP_SEP = “\t”
FILE_GROUP_COMMENT = “#”
SCORE_OUT_FILE = path/raremetal_scores.txt
COV_OUT_FILE = path/raremetal_covariances.txt

## === THE THIRD STUDY: SUMMARY STAT === ##
SOFTWARE = MetaSKAT
VERSION = 0.40
RESCALE = 1.0
FILE_MSSD = path/study3.MSSD
FILE_MInfo = path/study3.MInfo
SCORE_OUT_FILE = path/raremetal_scores.txt
COV_OUT_FILE = path/raremetal_covariances.txt

## === THE FOURTH STUDY: SUMMARY STAT === ##
SOFTWARE = seqMeta
VERSION = 1.5
RESCALE = 1.0
FILE_RDATA = path/study4.Rdata
ALLELE_FILE = path/allele_info.txt
ALLELE_SEP = “\t”
ALLELE_COMMENT = “#”
SCORE_OUT_FILE = path/raremetal_scores.txt
COV_OUT_FILE = path/raremetal_covariances.txt
```

Figure 1: Example script file for PreMeta input. Lines in bold are required, other lines are optional; ALLELE_FILE lines are required if target software is RAREMETAL or METASKAT. Output files are specified via keywords SCORE_OUT_FILE and COV_OUT_FILE, so that this script file is appropriate to use for using PreMeta to convert to RareMetal (use keyword MASS_OUT_FILE if target software is MASS, keywords MINFO_OUT_FILE and MSSD_OUT_FILE if target software is MetaSKAT, or keyword RDATA_OUT_FILE if target software is SeqMeta). If no X_OUT_FILE lines appear, premeta will use default names, and will print out the names used for each study's output files.

and comment character (keyword ALLELE_COMMENT) are optional (default values are “\t” and “#”, respectively).

Script File: RareMetal Inputs

If “SOFTWARE = RAREMETAL”, then three files must be provided: RareMetalWorker's covariance file should be specified via keyword FILE_COV; RareMetalWorker's score file should be specified via keyword FILE_SCORE, and a gene grouping file must be specified via keyword FILE_GROUP. This third file is necessary to convert the sliding-window approach of RareMetalWorker to the gene-based grouping of SNPs for the other three softwares. The format of the gene grouping file must have one gene per line, with format (see documentation for RAREMETAL at http://genome.sph.umich.edu/wiki/RAREMETAL_Documentation#From_a_Group_File):

GENE_NAME   CHR:POS:REF:ALT   CHR:POS:REF:ALT   CHR:POS:REF:ALT   ...

You can optionally specify delimiter and/or comment characters for each file via the FILE_X_SEP or FILE_X_COMMENT keywords, where X is {COV, SCORE, or GROUP}; if not specified, default values of "\t" for delimiter and "#" for comment character will be used. For a detailed description of the Score and Covariance files, refer to the documentation of RAREMETALWORKER (http://genome.sph.umich.edu/wiki/RAREMETALWORKER).

Script File: MetaSKAT Inputs

If "SOFTWARE = METASKAT", then two files must be provided: MetaSKAT's .MSSD file should be specified via keyword FILE_MSSD, and MetaSKAT's .MInfo file should be specified via keyword FILE_MINFO.

Script File: seqMeta Inputs

If "SOFTWARE = SEQMETA", then seqMeta's .Rdata file should be specified via keyword FILE_RDATA.

OUTPUT FILES FOR MASS PIPELINE

For each study, PreMeta generates one text file that can be read by MASS for meta-analysis. PreMeta also prepares the MASS script file "mass_script.txt" that will be used as MASS input. Refer to MASS documentation (http://dlin.web.unc.edu/software/mass/) for detailed description of the MASS script file. The following is an example of the MASS script file.

```
## === STUDY INFORMATION === ##
FILE = path/MASS_STUDY1.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9

## === STUDY INFORMATION === ##
FILE = path/MASS_STUDY2.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9

## === STUDY INFORMATION === ##
FILE = path/MASS_STUDY3.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9
```

Figure 2: Example MASS script file "mass_script.txt" that is output by PreMeta when target software is MASS

OUTPUT FILES FOR MetaSKAT PIPELINE

For each study, PreMeta generates .MSSD and .MInfo files that can be read by MetaSKAT for meta-analysis. For detailed description of the files, refer to the MetaSKAT manual (http://cran.r-project.org/web/packages/MetaSKAT/index.html).

OUTPUT FILES FOR seqMeta PIPELINE

For each study, PreMeta generates a .Rdata file containing a R object that can be read by seqMeta for meta-analysis. Note that the object name loaded in R should be the same as the .Rdata file name. See http://cran.r-project.org/web/packages/seqMeta/index.html for the seqMeta manual for a detailed description of the files.

OUTPUT FILES FOR RAREMETAL PIPELINE

For each study, PreMeta generates two text files (Score and Covariance) that can be read by RAREMETAL for meta-analysis. PreMeta also prepares two lists “score_files.txt’ and “cov_files.txt” that summarize the two sets of these text files across studies. The two list files can be directly used by RAREMETAL. For a detailed description of the lists refer to RAREMETAL documentation at http://genome.sph.umich.edu/wiki/RAREMETAL. The following is an example.

```
## === STUDY INFORMATION === ##
STUDY1_score.txt
## === STUDY INFORMATION === ##
STUDY2_score.txt
## === STUDY INFORMATION === ##
STUDY3_score.txt
## === STUDY INFORMATION === ##
STUDY4_score.txt
```

Figure 3: Example RareMetal “score_files.txt” file that is output by PreMeta when target software is RAREMETAL. This file lists all of the score files prepared by PreMeta.

```
## COV FILES
## === STUDY INFORMATION === ##
STUDY1_cov.txt
## === STUDY INFORMATION === ##
STUDY2_cov.txt
## === STUDY INFORMATION === ##
STUDY3_cov.txt
## === STUDY INFORMATION === ##
STUDY4_cov.txt
```

Figure 4: Example RareMetal “cov_files.txt” file that is output by PreMeta when target software is RAREMETAL. This file lists all of the covariance files prepared by PreMeta.

Learn By Example

This section provides a demonstration of the meta-analysis pipeline starting from the study-level analysis. Suppose we want to meta-analyze three studies. We have 2 genes in this example (4 SNPs in “gene1” and 8 SNPs in “gene2”). For studies 1, 2, and 3, we perform study-level analyses within the RAREMETAL, MetaSKAT, and seqMeta pipelines respectively. Then, we unify the output files into MASS input format using preMeta. At the end, we combine all of the summary statistics from across the three studies and perform the T5 burden test using MASS. We intentionally use the same data set for the three studies, such that the output from different software programs can be compared.

STEP 1: Study-Level Analysis

STUDY 1: RAREMETAL

For study 1, we perform the analysis using RAREMETALWORKER. The score file in Figure 5 and covariance file in Figure 6 were generated by running the command below.

```
raremetalworker –ped study1.ped –dat study1.dat –traitName Trait1 –prefix study1
```

```
##ProgramName=RareMetalWorker
##Version=0.4.1
##Samples=500
##AnalyzedSamples=500
##Families=500
##AnalyzedFamilies=500
##Founders=500
##AnalyzedFounders=500
##Covariates=
##InverseNormal=OFF
##TraitSummaries min 25th median 75th max mean variance
##Trait1 -6.13849 -1.47316 -0.125055 1.36358 7.07711 -0.0263837 4.04374
##AnalyzedTrait -6.13849 -1.47316 -0.125055 1.36358 7.07711 -0.0263837 4.04374
##Heritability=0.000%
#CHROM POS REF ALT N_INFORMATIVE FOUNDER_AF ALL_AF INFORMATIVE_ALT_AC CALL_RATE
  HWE_PVALUE  N_REF  N_HET  N_ALT  U_STAT  SQRT_V_STAT  ALT_EFFSIZE  PVALUE  SE
1   901922 A  T  500  0.107  0.107  107  1  0.486391    400  93    7 60.2106 19.8974 0.613756 0.00247764   0.10
1   902176 A  T  500  0.019  0.019  19   1  1           481  19    0 0.729671 8.58858 0.0399207 0.932295   0.23
1   934735 A  T  500  0.001  0.001  1    1  1           499  1     0 -1.57341 2.00688 -1.57656 0.433037   1.00
1   949422 A  T  500  0.208  0.208  208  1  0.0417426   306  180   14 -1.25632 24.5605 -0.00840504 0.959204   0.08
1   949832 A  T  500  0.074  0.074  74   1  0.744903    429  68    3 11.4939 16.6929 0.166462 0.491107   0.12
1   970687 A  T  500  0.046  0.046  46   1  0.615586    454  46    0 -7.14266 12.9831 -0.171008 0.582217   0.15
1   978628 A  T  500  0.083  0.083  83   1  1           420  77    3 -4.60639 17.4233 -0.0612373 0.791486   0.12
1   1908628A  T  500  0.205  0.205  205  1  0.336366    312  171   17 2.36819 25.0065 0.0152836 0.924551   0.08
1   1968628A  T  500  0.009  0.009  9    1  1           491  9     0 12.4105 5.9722 1.40423 0.0377045   0.34
1   1988634A  T  500  0.003  0.003  3    1  1           497  3     0 10.2907 3.46905 3.45095 0.00301268   0.58
1   2718548A  T  500  0.019  0.019  19   1  1           481  19    0 -8.73673 8.58858 -0.477992 0.309035   0.23
1   2807288A  T  500  0.05   0.05   50   1  0.625924    450  50    0 -6.74858 13.4761 -0.149968 0.616524   0.15
#Genomic control for additive is: 1.03969
```

Figure 5: Score file study1.Trait1.singlevar.score.txt

```
##ProgramName=RareMetalWorker
##Version=0.4.1
#CHROME  CURRENT_POS  MARKERS_IN_WINDOW  COV_MATRICES
1  901922  901922,902176,934735,949422,949832,970687,978628,  0.0486176,-0.00102387,0.000389528,
-0.000253738,0.00355035,0.00255522,0.000613531,
1  902176  902176,934735,949422,949832,970687,978628,  0.00905826,-1.88321e-05,4.75759e-05,9.31695e-05,
-0.000370696,-7.63197e-05,
1  934735  934735,949422,949832,970687,978628,1908628,  0.000494591,-0.000206162,-7.33462e-05,
-4.55936e-05,-8.22667e-05,-0.000203189,
1  949422  949422,949832,970687,978628,1908628,  0.0740757,0.00308054,0.00191493,
-0.00125283,0.00432148,
1  949832  949832,970687,978628,1908628,  0.034219,0.000590734,0.000354837,0.000822667,
1  970687  970687,978628,1908628,1968628,  0.0206995,0.000180392,0.00106055,8.52402e-05,
1  978628  978628,1908628,1968628,  0.0372787,0.00394979,-0.0007404,
1  1908628  1908628,1968628,1988634,2718548,2807288,  0.0767905,-0.000837534,0.000381598,
0.000104072,0.00223012,
1  1968628  1968628,1988634,2718548,2807288,  0.00437996,-2.67615e-05,-0.000169489,0.000545141,
1  1988634  1988634,2718548,2807288,  0.00147783,0.000439086,-0.000148675,
1  2718548  2718548,2807288,  0.00905826,-0.000446024,
1  2807288  2807288,  0.0223012,
```

Figure 6: Covariance file study1.Trait1.singlevar.cov.txt

STUDY 2: MetaSKAT

For study 2, we perform the analysis using the function Generate_Meta_Files in MetaSKAT. The two files study2.MSSD and study2.MInfo were generated by the *R* code in Figures 7 - 9 below.

```
> File.SetID = paste(dir.metaskat.input, “mapping.txt”, sep = “”)
> File.Bed = paste(dir.metaskat.input, “study2.bed”, sep = “”)
> File.Bim = paste(dir.metaskat.input, “study2.bim”, sep = “”)
> File.Fam = paste(dir.metaskat.input, “study2.fam”, sep = “”)
> File.Mat = paste(dir.metaskat.output, “study2.MSSD”, sep = “”)
> File.SetInfo = paste(dir.metaskat.output, “study2.MInfo”, sep = “”)
> FAM <- read.table(File.Fam, header = FALSE)
> y <- FAM[, 6]
> library(SKAT)
> library(MetaSKAT)
> N.Sample <- length(y)
> obj <- SKAT_Null_Model(y ~ 1)
> Generate_Meta_Files(obj, File.Bed, File.Bim, File.SetID, File.Mat, File.SetInfo,N.Sample)
Read SetID file
SetID file has 2 sets
Read Bim file
Bim file has 12 markers
Merge datasets and get set info
Save was done successfully!
```

Figure 7: MetaSKAT: Example *R* code to generate .MSSD and .MInfo files

The information in the two files study2.MSSD and study2.MInfo are retrieved as follows:

```
> Cohort.Info <- Open_MSSD_File_2Read(File.Mat, File.SetInfo)
Number of cohorts = 1
500 samples, 2 sets, 12 SNPs and 12 unique SNPs
> SetID = “gene1”
> t <- MetaSKAT:::Get_META_Data_OneSet(Cohort.Info, SetID)
> gene1.info<-MetaSKAT:::Get_META_Data_OneSet_Align(t$SMat.list, t$Info.list, t$IsExistSNV, 1)
$SMat.list
$SMat.list[[1]]
           [,1]       [,2]      [,3]       [,4]
[1, ]  24.2602  -0.510913  0.194374  -0.12662
[2, ]  -0.5109   4.520070 -0.009397   0.02374
[3, ]   0.1944  -0.009397  0.246801  -0.10287
[4, ]  -0.1266   0.023740 -0.102875  36.96377


$Info.list
$Info.list[[1]]
   SNPID IDX SetID SetID_numeric Score MAF MissingRate Allele1 Allele2 MinorAllele PASS StartPOS
IDX1
1 1:901922 1 gene1 1 14.8898 0.107 0 A TRUE TRUE PASS 1 1
2 1:902176 2 gene1 1   0.1804 0.019 0 A TRUE TRUE PASS 1 2
3 1:934735 3 gene1 1  -0.3891 0.001 0 A TRUE TRUE PASS 1 3
4 1:949422 4 gene1 1  -0.3107 0.208 0 A TRUE TRUE PASS 1 4
```

Figure 8: MetaSKAT: Example *R* code to view summary statistics for gene1

```
> SetID = “gene2”
> t<-MetaSKAT:::Get_META_Data_OneSet(Cohort.Info, SetID)
> gene2.info<-MetaSKAT:::Get_META_Data_OneSet_Align(t$SMat.list, t$Info.list, t$IsExistSNV, 1)

$SMat.list
$SMat.list[[1]]
            [,1]        [,2]        [,3]        [,4]        [,5]      [,6]        [,7]        [,8]
[1,]  38.31846  -0.41793   0.19042   0.05193   1.11283   0.4105    0.52921   1.97095
[2,]  -0.41793   2.18560  -0.01335  -0.08458   0.27203   0.1652   0.04253  -0.36946
[3,]   0.19042  -0.01335   0.73744   0.21910  -0.07419  -0.1098  -0.06825  -0.12315
[4,]   0.05193  -0.08458   0.21910   4.52007  -0.22257  -0.4481  -0.18498  -0.53267
[5,]   1.11283   0.27203  -0.07419  -0.22257  11.12830   0.3957  -0.39567   0.91499
[6,]   0.41051   0.16519  -0.10980  -0.44810   0.39567  17.0753   0.29478   0.17706
[7,]   0.52921   0.04253  -0.06825  -0.18498  -0.39567   0.2948  10.32904   0.09002
[8,]   1.97095  -0.36946  -0.12315  -0.53267   0.91499   0.1771   0.09002  18.60207


$Info.list
$Info.list[[1]]
 SNPID IDX SetID SetID_numeric Score MAF MissingRate Allele1 Allele2 MinorAllele PASS StartPOS IDX1
1 1:1908628 4 gene2 2  0.5856 0.205 0 A TRUE TRUE PASS 45 4
2 1:1968628 5 gene2 2  3.0691 0.009 0 A TRUE TRUE PASS 45 5
3 1:1988634 6 gene2 2  2.5449 0.003 0 A TRUE TRUE PASS 45 6
4 1:2718548 7 gene2 2 -2.1606 0.019 0 A TRUE TRUE PASS 45 7
5 1:2807288 8 gene2 2 -1.6689 0.050 0 A TRUE TRUE PASS 45 8
6 1:949832  1 gene2 2  2.8424 0.074 0 A TRUE TRUE PASS 45 1
7 1:970687  2 gene2 2 -1.7663 0.046 0 A TRUE TRUE PASS 45 2
8 1:978628  3 gene2 2 -1.1391 0.083 0 A TRUE TRUE PASS 45 3
```

Figure 9: MetaSKAT: Example *R* code to view summary statistics for gene2

STUDY 3: SeqMeta

For study 3, we perform the analysis using the function prepScores in the seqMeta package. The .Rdata set study3.Rdata was generated by running the code below.

```
# Generate the summary statistics using geno.Rdata, pheno.Rdata, and SNPInfo.Rdata
> library(seqMeta)
> load(file = paste(dir.seqmeta.input, "geno.Rdata", sep = ""))
> load(file = paste(dir.seqmeta.input, "pheno.Rdata", sep = ""))
> load(file = paste(dir.seqmeta.input, "SNPInfo.Rdata", sep = ""))
> study3 = prepScores(geno, pheno ~ 1, SNPInfo = SNPInfo)
> save(study3, file = paste(dir.seqmeta.output, "study3.Rdata", sep = ""))

$gene1
$gene1$scores
1:901922  1:902176  1:934735  1:949422
60.2106   0.7297    -1.5734   -1.2563

$gene1$cov
4 x 4 sparse Matrix of class "dsCMatrix"
          1:901922  1:902176  1:934735  1:949422
1:901922    98.102    -2.066     0.786    -0.512
1:902176    -2.066    18.278    -0.038     0.096
1:934735     0.786    -0.038     0.998    -0.416
1:949422    -0.512     0.096    -0.416   149.472

$gene1$n
[1] 500

$gene1$maf
1:901922  1:902176  1:934735  1:949422
   0.107     0.019     0.001     0.208

$gene1$sey
[1] 2.011


$gene2
$gene2$scores
 1:949832  1:970687  1:978628  1:1908628  1:1968628  1:1988634  1:2718548
   11.494    -7.143    -4.606      2.368     12.411     10.291     -8.737
1:2807288
-6.749

$gene2$cov
8 x 8 sparse Matrix of class "dsCMatrix"
           1:949832  1:970687  1:978628  1:1908628  1:1968628  1:1988634  1:2718548  1:2807288
1:949832     69.048     1.192     0.716       1.66      0.668      0.444      1.812        1.6
1:970687      1.192    41.768     0.364       2.14      0.172     -0.276     -0.748       -1.6
1:978628      0.716     0.364    75.222       7.97     -1.494     -0.498     -2.154        3.7
1:1908628     1.660     2.140     7.970     154.95     -1.690      0.770      0.210        4.5
1:1968628     0.668     0.172    -1.494      -1.69      8.838     -0.054     -0.342        1.1
1:1988634    -0.444    -0.276    -0.498       0.77     -0.054      2.982      0.886       -0.3
1:2718548    -1.812    -0.748    -2.154       0.21     -0.342      0.886     18.278       -0.9
1:2807288     1.600    -1.600     3.700       4.50      1.100     -0.300     -0.900       45.0

$gene2$n
[1] 500

$gene2$maf
1:949832  1:970687  1:978628  1:1908628  1:1968628  1:1988634  1:2718548  1:2807288
   0.074     0.046     0.083      0.205      0.009      0.003      0.019      0.050

$gene2$sey
[1] 2.011


attr(,"family")
[1] "gaussian"
attr(,"class")
[1] "seqMeta"
```

Figure 10: seqMeta: Example *R* code to generate summary statistics

STEP 2: Reformat

Before we reformat the output files from the three studies, we prepare a group file "gfile.txt" and a PreMeta script file "premeta_script.txt".

```
gene1  1:901922:A:T  1:902176:A:T  1:934735:A:T  1:949422:A:T
gene2  1:949832:A:T  1:970687:A:T  1:978628:A:T  1:1908628:A:T  1:1968628:A:T  1:1988634:A:T
  1:2718548:A:T  1:2807288:A:T
```

Figure 11: Group file gfile.txt

```
#== STUDY 1 from RAREMETALWORKER ==#
SOFTWARE = RAREMETAL
VERSION = 0.4.0
FILE_SCORE = PATH\study1.Trait1.singlevar.score.txt
FILE_COV = PATH\study1.Trait1.singlevar.cov.txt
FILE_GROUP = PATH\group.txt

#== STUDY 2 from MetaSKAT ==#
SOFTWARE = MetaSKAT
VERSION = 0.40
FILE_MSSD = PATH\study2.MSSD
FILE_MInfo = PATH\study2.MInfo

#== STUDY 3 from seqMeta ==#
SOFTWARE = seqMeta
VERSION = 1.5
FILE_RDATA = PATH\study3.Rdata
```

Figure 12: PreMeta script file premeta_script.txt

We then run PreMeta as follows.

```
premeta --script premeta_script.txt  --software MASS --version 7.0
```

PreMeta will generate three files: mass_score_file_1.txt, mass_score_file_2.txt, mass_score_file_3.txt; as well as a MASS script file mass_script.txt.

The contents of the three files for MASS are printed below (remember that they are from one data set).

```
gene1 1:901922 107 500 14.8898302238126 24.2601954046609 -0.510912761269184 0.194374361257298
-0.126615360004763
gene1 1:902176 19 500 0.180444544486391 -0.510912761269184 4.52006943391972
-0.00939723375035285 0.0237403800008899
gene1 1:934735 1 500 -0.389096683350855 0.194374361257298 -0.00939723375035285 0.246801033759267
-0.102874980003863
gene1 1:949422 208 500 -0.310682063941474 -0.126615360004763 0.0237403800008899
-0.102874980003863 36.9637716613879
gene2 1:949832 74 500 2.84239154603675 17.0752683156411 0.294776385011067 0.177063667506646
0.410510737515409 0.165193477506203 -0.109799257504123 -0.448099672516826 0.395673000014855
gene2 1:970687 46 500 -1.76634807571524 0.294776385011067 10.3290436653878 0.0900156075033785
0.529212637519867 0.042534847501597 -0.0682535925025628 -0.184977127506946 -0.395673000014857
gene2 1:978628 83 500 -1.13914059157506 0.177063667506646 0.0900156075033785 18.6020715044485
1.970946131324 -0.369459663763872 -0.123153221254624 -0.532674776270001 0.914993812534354
gene2 1:1908628 205 500 0.585643419881772 0.410510737515409 0.529212637519867 1.970946131324
38.3184570951888 -0.417929606265693 0.19041763125715 0.0519320812519488 1.11283031254178
gene2 1:1968628 9 500 3.0690726173488 0.165193477506203 0.042534847501597 -0.369459663763872
-0.417929606265693 2.18559873383206 -0.0133539637505014 -0.0845751037531756 0.272025187510214
gene2 1:1988634 3 500 2.54485510634904 -0.109799257504123 -0.0682535925025628
-0.123153221254624 0.19041763125715 -0.0133539637505014 0.737435553777689
0.219103923758227 -0.0741886875027857
gene2 1:2718548 19 500 -2.160555107968 -0.448099672516826 -0.184977127506946
-0.532674776270001 0.0519320812519488 -0.0845751037531756 0.219103923758227
4.52006943391972 -0.222566062508357
gene2 1:2807288 50 500 -1.66889439329688 0.395673000014855 -0.395673000014857 0.914993812534354
1.11283031254178 0.272025187510214 -0.0741886875027857 -0.222566062508357 11.1283031254178
```

Figure 13: Example: Output of PreMeta script (target software = MASS)

The values can differ a bit over three studies because of rounding and the different ways to calculate the residual variance (the residual variance estimates from different software programs can differ by a factor of $n/(n-p)$, where n is the sample size and p is the number of covariates).

```
## === THE FIRST INPUT FILE AND COLUMN SPECIFICATION === ##
FILE = PATH\MASS_study1.Trait1.singlevar.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9

## === THE SECOND INPUT FILE AND COLUMN SPECIFICATION === ##
FILE = PATH\MASS_study2.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9

## === THE THIRD INPUT FILE AND COLUMN SPECIFICATION === ##
FILE = PATH\MASS_study3.txt
SKIP = 1
GENE_ID_COLUMN = 1
GVAR_ID_COLUMN = 2
MAC_COLUMN = 4
N_OBS_COLUMN = 5
SCORE_COLUMN = 9
```

Figure 14: Example: MASS script file mass_script.txt, used to perform meta-analysis across 3 studies

STEP 3: Meta-Analysis

Finally, we run the meta-analysis (e.g., T5 burden test) using MASS.

```
    MASS -test T5 -sfile mass_script.txt.out -ofile output_MASS.txt
```

The results are contained in the output file output_MASS.txt.