Meta-analysis of rare variant association test

Clara Tang

2015

Rare variants

- Functional rare variants are expected to have larger effect size, single variant association test can still lack power due to few copies of rare alleles

- Several ways to leverage power
  - Increasing sample size
  - **Meta-analysis**
  - **Gene-based/Set-based association test** to group rare variants likely to be functional in gene or pathway

Rare variants

- Functional rare variants are expected to have larger effect size, single variant association test can still lack power due to few copies of rare alleles

- Several ways to leverage power
  - Increasing sample size
  - **Meta-analysis**
  - **Gene-based/Set-based association test** to group rare variants likely to be functional in gene or pathway
  - **Meta-analysis of gene-based association test**

Meta-analysis of gene-based rare variants test

- Gene-based association test statistics can be reconstructed from single variant score statistics
- Distribution of test statistics can be computed using linkage disequilibrium information, i.e. variance-covariance matrices
- Highly comparable to joint association analysis without using raw data

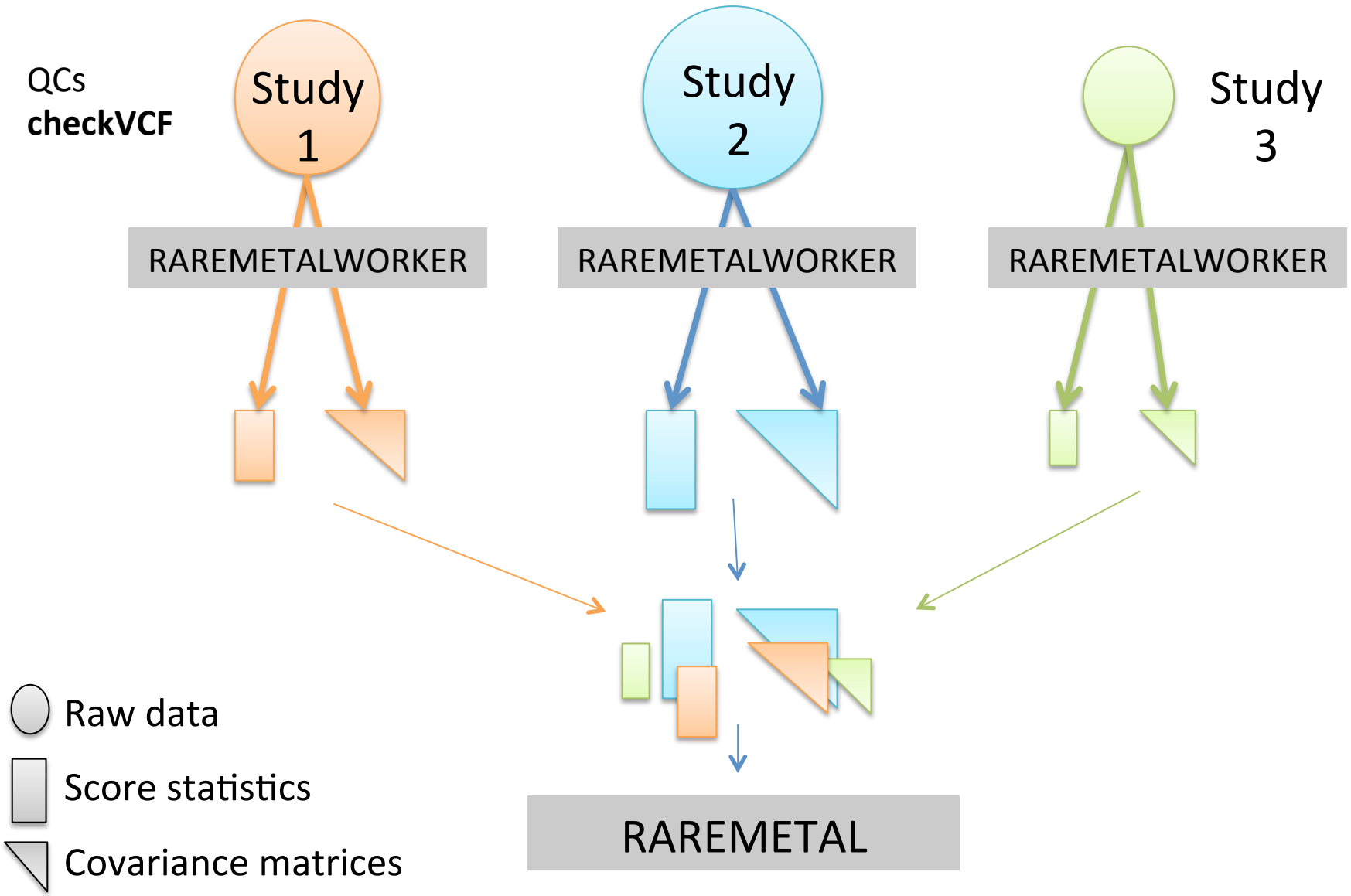
Rare variants meta-analysis

- Meta-SKAT (Lee *et al.* 2013)
- seqMeta (Voorman *et al.* 2013).
- MASS (Tang and Lin, 2013; Hu *et al.* 2013)
- **RAREMETAL** (Liu *et al.* 2014; Feng *et al.* 2014)

RAREMETAL

- Meta-analysis involves two steps
  - Takes summary statistics and LD matrices computed by **RAREMETALWORKER** or rvtests
  - Combine results across studies using **RAREMETAL**
- Both single variant and **burden/set-based** meta-analysis are supported
- It was originally designed for meta-analyzing unrelated individuals but has been extended to cover related samples with the use of linear mixed model
- It can take into account of study-specific covariates and cryptic relatedness
- Most suitable for exome array and sequencing studies

RAREMETAL workflow



What individual group shares?

- For each of the *k* studies for trait *y*, we need to share
  - Single variant score statistics u_k

$$\mathbf{u}_\mathbf{k} = \left(X_k - \bar{X}_k \right)^T \mathbf{y}_\mathbf{k}$$

    computed by linear regression (without kinship matrix) or linear mixed model

  - Covariance matrix V_k

$$V_k = \hat{\sigma}_k^2 N_k \operatorname{cov}\left(X_k \right) = \hat{\sigma}_k^2 \left(X_k - \bar{X}_k \right)^T \left(X_k - \bar{X}_k \right)$$

  - Estimated alternative allele frequencies
  - Genotype call rate and HWE *p*-values
  - Mean and variance for the trait

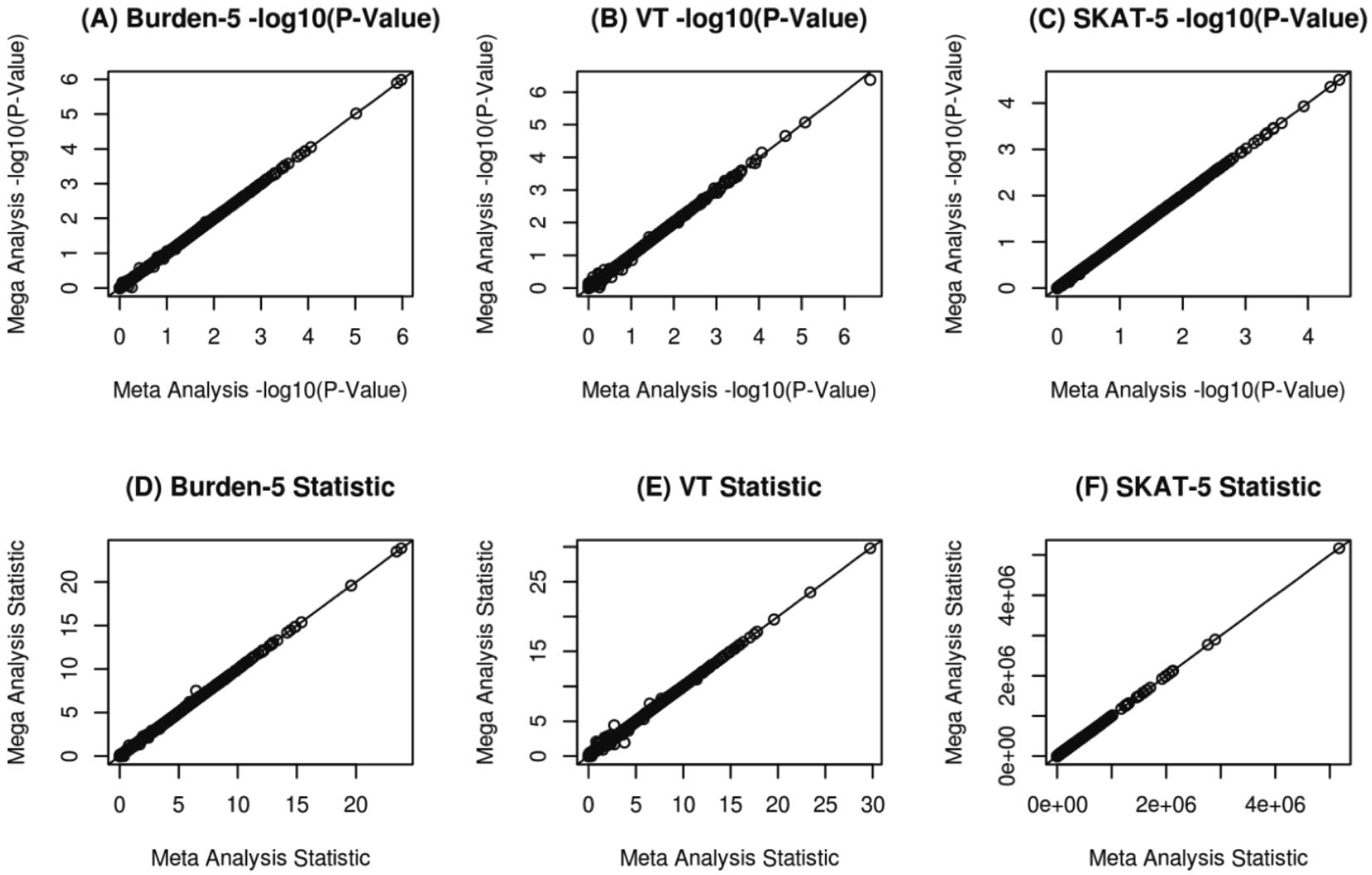
RAREMETAL

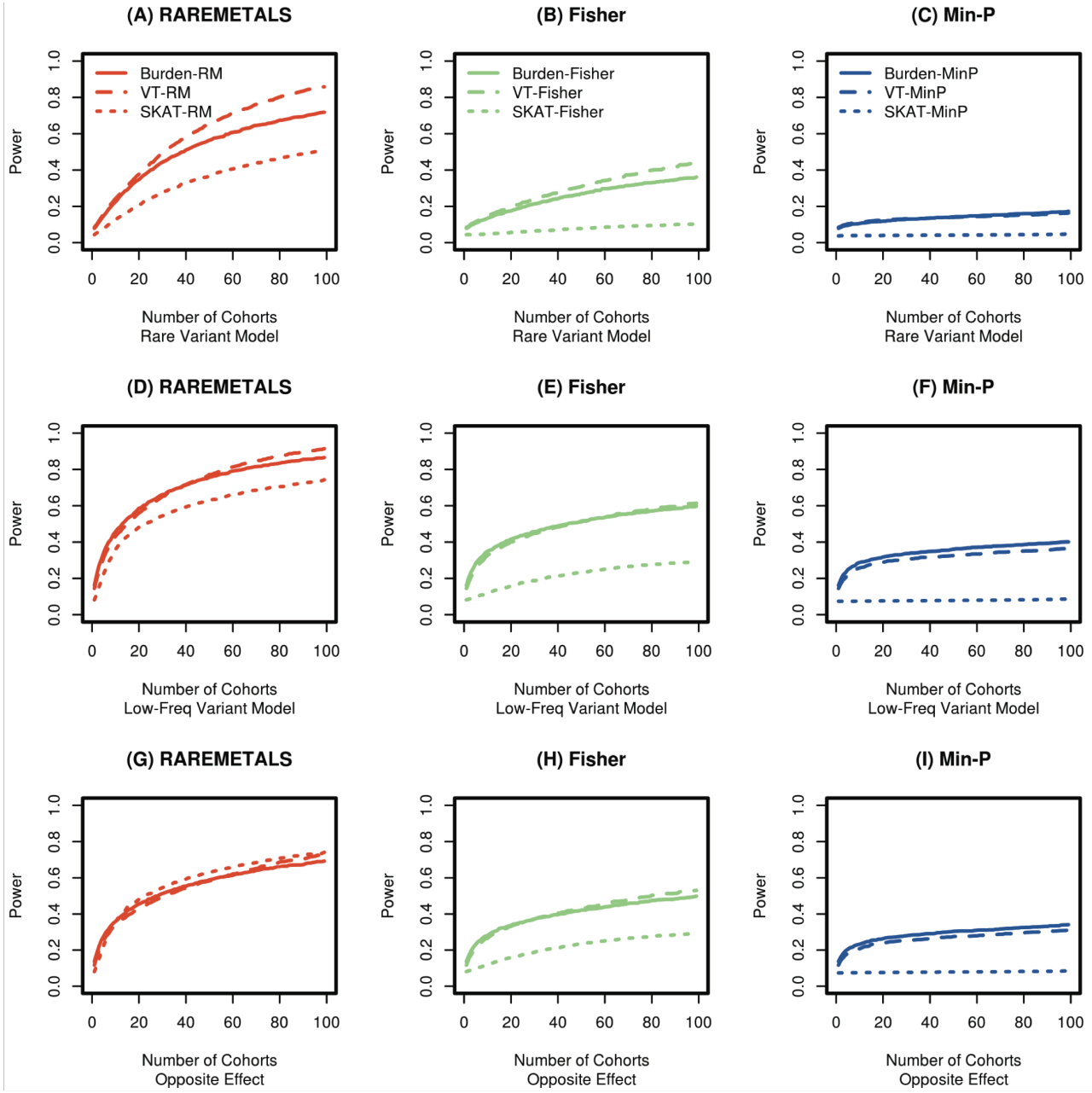
- Single variant association test statistics are first combined across studies using the Cochran-Mantel-Haenszel method.

$$T_{j,\bullet} = U_{j,\bullet} \Big/ \sqrt{V_{j,j,\bullet}}$$

- Perform gene-based or region-based meta analysis
  - Burden
  - Madsen-Browning (MB)
  - Variable Threshold (VT)
  - SKAT

Meta versus joint mega-analysis




(A) RAREMETALS
Burden-RM
VT-RM
SKAT-RM
(B) Fisher
Burden-Fisher
VT-Fisher
SKAT-Fisher
(C) Min-P
Burden-MinP
VT-MinP
SKAT-MinP
(D) RAREMETALS
(E) Fisher
(F) Min-P
(G) RAREMETALS
(H) Fisher
(I) Min-P
Power
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Number of Cohorts
Rare Variant Model
Low-Freq Variant Model
Opposite Effect

Practical

RAREMETALWORKER

+

RAREMETAL

Practical

- Make a new directory
  - `mkdir ~/raremetal`
- Copy the data folder to your home raremetal directory
  - `cp /faculty/clara/2015/raremetal/* ~/raremetal`
- Go to your raremetal directory
  - `cd ~/raremetal`

RAREMETALWORKER

- MERLIN format **PED** and **DAT**

```
--ped input.ped --dat input.dat
```

  - .ped

```
 1 1 0 0 1 1.5 1 23 A A A A A A A A A A
 2 1 0 0 1 1.0 1 34 A C A C A C A C A C
 3 1 0 0 2 0.4 1 43 A A A A A A A A A A
 4 1 0 0 2 0.9 1 13 A C A C A C A C A C
```

  - .dat

```
T YourTraitName
C SEX
C AGE
M 1:123456
M 1:234567
M 2:111111
M 2:222222
M X:12345
```

phenotypes

genotypes

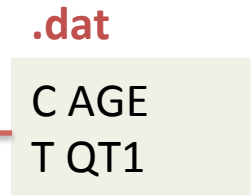
Markers in PED and DAT file must be sorted by chromosome and position.

- and/or **VCF** for genotypes

```
--ped input.ped --dat input.dat --vcf input.vcf.gz
```

RAREMETALWORKER

- Zip and index the vcf file of example 1
  - bgzip example1.anno.vcf
  - tabix -p vcf -f example1.anno.vcf.gz
- Obtain score test statistics and covariance matrix for each variant
  - raremetalworker --ped example1.ped \
    --dat example1.dat \
    --vcf example1.anno.vcf.gz \
    --traitName QT1 \
    --inverseNormal \
    --makeResiduals \
    --prefix example1



- Repeat for example 2

.QT1.singlevar.score.txt

example1.QT1.singlevar.score.txt

```
##ProgramName=RareMetalWorker
##Version=4.13.5
##Samples=1999
##AnalyzedSamples=1680
##Families=1718
##AnalyzedFamilies=1680
##Founders=1718
##AnalyzedFounders=1680
##Covariates=AGE
##CovariateSummaries    min     25th    median  75th    max     mean    variance
##AGE   35      46      49      53      67      49.4708 23.8538
##InverseNormal=ON
##TraitSummaries        min     25th    median  75th    max     mean    variance
##QT1   -2.599  -0.657  0.019   0.646   3.708   0.0335833       0.92307
## - NullModelEstimates
## - Name       BetaHat SE(BetaHat)
## - Intercept  -2.55589e-09    0.024388
##AnalyzedTrait -3.43377        -0.673553       0.00074602      0.675427        3.43377 -2.55589e-09    0.999818
##Sigma_e2_Hat  0.999223
#CHROM  POS     REF     ALT     N_INFORMATIVE   FOUNDER_AF      ALL_AF  INFORMATIVE_ALT_AC      CALL_RATE       HW
```

.QT1.singlevar.score.txt

example1.QT1.singlevar.score.txt

```
##ProgramName=RareMetalWorker
##Version=4.13.5
##Samples=1999
##AnalyzedSamples=1680
##Families=1718
##AnalyzedFamilies=1680
##Founders=1718
##AnalyzedFounders=1680
##Covariates=AGE
##CovariateSummaries    min     25th    median  75th    max     mean    variance
##AGE   35      46      49      53      67      49.4708 23.8538
##InverseNormal=ON
##TraitSummaries        min     25th    median  75th    max     mean    variance
##QT1   -2.599  -0.657  0.019   0.646   3.708   0.0335833       0.92307
## - NullModelEstimates
## - Name       BetaHat SE(BetaHat)
## - Intercept  -2.55589e-09    0.024388
##AnalyzedTrait -3.43377        -0.673553       0.00074602      0.675427        3.43377 -2.55589e-09    0.999818
##Sigma_e2_Hat  0.999223
#CHROM  POS     REF     ALT     N_INFORMATIVE   FOUNDER_AF      ALL_AF  INFORMATIVE_ALT_AC      CALL_RATE       HW
```

example2.QT1.singlevar.score.txt

```
##Covariates=AGE
##CovariateSummaries    min     25th    median  75th    max     mean    variance
##AGE   31      46      50      53      66      49.5781 26.0694
##InverseNormal=ON
##TraitSummaries        min     25th    median  75th    max     mean    variance
##QT1   -3.244  -0.71   -0.085  0.616   4.11    -0.01691        1.03818
## - NullModelEstimates
## - Name       BetaHat SE(BetaHat)
## - Intercept  7.99546e-09     0.0173128
##AnalyzedTrait -3.61543        -0.674726       -0.000751613    0.672368        3.61543 7.99546e-09     0.999907
```

Top associated variant in example 1?

Top associated variant in example 2?

Top associated variant in example 1?

```
sort --key 17 -g example1.QT1.singlevar.score.txt | less -S
```

Top associated variant in example 2?

```
sort --key 17 -g example2.QT1.singlevar.score.txt | less -S
```

Top associated variant in example 1?

```
sort --key 17 -g example1.QT1.singlevar.score.txt | less -S
```

Top associated variant in example 2?

```
sort --key 17 -g example2.QT1.singlevar.score.txt | less -S
```

Example	#CHR	POS	REF	ALT	N	ALL_AF	ALT_AC	ALT_EFFSIZE	PVALUE
1	10	13214753	G	C	1680	6.85E-3	23	-0.7126	6.85E-4
2	10	13214753	G	C	3335	9.00E-3	60	-0.5691	1.24E-05

RAREMETAL

- Zip and index both the score statistics and covariance files
  - bgzip example1.QT1.singlevar.score.txt
  - bgzip example1.QT1.singlevar.cov.txt
  - tabix -s 1 -b 2 -e 2 -c "#" example1.QT1.singlevar.score.txt.gz
  - tabix -s 1 -b 2 -e 2 -c "#" example1.QT1.singlevar.cov.txt.gz
- Repeat for example 2
- Record the file names of score statistics and covariance files into summaryfiles and covfiles
  - ls example1.QT1.singlevar.score.txt.gz \
    example2.QT1.singlevar.score.txt.gz > summaryfiles
  - ls example1.QT1.singlevar.cov.txt.gz \
    example2.QT1.singlevar.cov.txt.gz > covfiles

RAREMETAL

- Run RAREMETAL
  - raremetal --summaryFiles summaryfiles \
    --covFiles covfiles \
    --groupFile groupfile \
    --SKAT --burden --MB --VT \
    --hwe 1.0e-05 \
    --callRate 0.95 \
    --longOutput \
    --tabulateHits \
    --hitsCutoff 1e-04 \
    --prefix COMBINED.QT1 \
    --maf 0.05

meta-analysis results

##Method=SinglevarScore
##STUDY_NUM=2
##TotalSampleSize=5015

Single variant results

#CHROM	POS	REF	ALT	N	POOLED_ALT_AF	DIRECTION	EFFECT_SIZE	EFFECT_SIZE_SD	H2	PVALUE	Gene
10	13214753	G	C	5015	8.28E-03	--	-0.609	0.111	6.04E-03	3.72E-08	*MCM10*
6	44145063	T	C	5015	3.19E-03	++	0.828	0.177	4.34E-03	3.04E-06	*CAPN11*
6	44137663	A	G	5015	0.255	++	0.089	0.023	2.98E-03	1.09E-04	*CAPN11*
6	44151691	A	T	5013	0.267	--	-0.080	0.023	2.50E-03	4.03E-04	*CAPN11*
6	44145170	C	T	5009	0.192	++	0.089	0.025	2.50E-03	4.05E-04	*CAPN11*
10	7772060	G	C	5015	3.99E-04	++	1.573	0.500	1.97E-03	1.66E-03	*ITIH2*
10	7769777	G	A	5015	0.043	++	0.147	0.049	1.82E-03	2.49E-03	*ITIH2*
10	7788586	T	A	5015	1.79E-03	++	0.612	0.236	1.34E-03	9.48E-03	*ITIH2*
6	44137312	C	T	5015	0.012	--	-0.223	0.090	1.21E-03	0.0136	
6	44137284	G	C	5012	0.012	--	-0.218	0.090	1.17E-03	0.0154	

Gene based results

Gene-based *P*	*ASPG*	*CAPN11*	*CDHR2*	*ITIH2*	*MCM10*	*NKX2-2*
MB	0.152	0.030	0.072	3.91E-06	0.021	0.846
SKAT	0.550	3.62E-06	0.393	0.001	3.60E-08	1.000
VT	0.320	0.002	0.212	0.001	4.52E-07	0.972
burden	0.344	0.001	0.043	1.53E-04	1.69E-07	0.947