Five-Minute SMR Overview

Brian Lee

leebn@sas

27 Jan 2021

https://cnsgenomics.com/software/smr/#SMR&HEIDIanalysis

Hopefully this site will be up; if not, the Wayback Machine saved a few copies as of 27 Jan 2020.

https://web.archive.org/web/20200113104038/http://cnsgenomics.com/software/smr/download/SMR.pdf

Step 1: Data Accumulation/Dependencies

- You will need:
  - Some data-wrangling software/capabilities (R 4.0.X + dplyr, Python 3.x + pandas)
  - A copy of the SMR executable (download from website or Shen Lab members)
  - A PLINK-format binary file set (i.e. *.bed, *.bim, *.fam)
  - Summary GWAS statistics
    - Either use your own GWAS or download from IGAP, Jansen et al 2019 (NatGen), or Kunkle et al 2019.
  - Summary *QTL (i.e. eQTL, mQTL, …) data
    - Commonly calculate your own using MatrixEQTL/FastQTL/PLINK (**not** recommended) or download from either GTEx or Synapse (requires an account)

Step 2a: Storing *QTL Information

*QTL summary data is stored in three files:

- *.esi file (SNP info, similar to PLINK's *.bim, < 10 minutes to create)
- *.epi file (gene/probe information, often requires annotation packages and files)
- *.besd file (links the two and often created *first*; **if you get this wrong, must start over**)

How do we make these files?

Step 2b(i): Make initial *QTL File Set

- Shen Lab members often create the initial eQTL file set using data formatted in the MatrixEQTL format (cols are SNP (*str*), gene (*str)*, beta (*double*), t-stat (*double*), p-value *(double*), FDR (*double*)).
- If you can, use MatrixEQTL to calculate *QTL's for this reason!
- Before moving forward, ensure the following information is available for all SNP's:
  - BP
  - Chr
  - Effect/coded allele + 'other' allele
  - Effect allele frequency

Step 2b(ii): Make initial *QTL File Set

- Ensure the following information is available for all probes:
  - Chr
  - Physical position
- Remove all SNP's and probes that don't have all mandatory information.
- Command: "smr --eqtl-summary matrixEqtlOutputHere.txt --matrix-eQTL-format --make-besd --out myBesdFileset"
- Will create myBesdFileset.besd, *.epi, *.esi but you must update the *.epi and *.esi

Step 2c(i): Update *.esi and *.epi

- You must manually update the *.esi and *.epi files if using the Matrix eQTL method.
- **Do not** change the order of the SNP's/probes in this file.
- *.esi has no header; columns are Chr # (*int*), SNP (*str*), Dummy variable (make all rows '0'), BP (*int*), effect/coded allele (*str*), other allele (*str*), frequency of effect allele (*double*)
- *.epi has no header; columns are Chr # (*int*), probe ID (*str*), Dummy variable (make all rows '0'), physical position (*int*), gene ID (*str*), gene orientation ('str')
- If the probe isn't associated with a specific gene, 'NA' works fine. This information (and likely the probe ID) are likely for your information when producing the SMR output report.
- Gene orientation is only used for the SMR-related visualizations. If this data is unavailable, just default to '+'.

Step 2c(ii): Update *.esi and *.epi

- Run the following:

‘smr --beqtl-summary my_beqtl --update-esi my_updated_esi.esi’

‘smr --beqtl-summary my_beqtl --update-epi my_updated_epi.epi’

- If portions of the information necessary for the *.epi and *.esi files are unavailable (i.e. BP or minor allele frequency), you must exclude that SNP/probe. This is why we checked for information first.
- The *.besd, *.epi, and *.esi files should have the same file name root.

Step 3: Summary GWAS Statistics

- Shen Lab members have already made .ma files for Kunkle, Jansen, and IGAP data.
- This tab-delimited file has a header. Column names are SNP, A1, A2, freq, b, se, p, n
- A1 = coded/effect allele, A2 = ‘other’ allele
- freq = frequency of A1
- b = beta
- se = standard error
- n = Sample size
- File format: .ma

Step 4: SMR

- Run the SMR and HEIDI test as follows:

‘smr --bfile myPLINKbfile --gwas-summary mygwas.ma --beqtl-summary myeqtl --out mysmr --thread-num 10’

- Results will be in mysmr.smr
- Primarily look at column p_SMR
- To perform Bonferroni correction, find number of tests via (number of lines of mysmr.smr – 1).