

Joint Likelihood Mapping (JLIM)

JLIM is a cross-trait test of shared causal effect, which is described in [Chun et al. 2017](#). JLIM tests whether two traits – main and secondary – are driven by shared causal effect or not. Typically, the main trait is a large GWAS study, and secondary trait can be an expression Quantitative Trait Loci (eQTL) association study. For main trait, JLIM takes only summary-level association statistics, but for secondary trait, it requires genotype-level data to generate permutation-based null distribution. JLIM is simultaneously released at [Cotsapas lab github](#) and Sunyaev lab website.

Download

JLIM release v1.0.3

How to install

The core JLIM module is implemented as an R extension (**jlimR**). **jlimR** depends on **getopt** module. If it is not installed, **getopt** can be installed from CRAN by:

```
Rscript -e 'install.packages("getopt", repos="http://cran.r-project.org")'
```

After **getopt** has been installed, **jlimR** (included in the distribution file) can be installed by:

```
tar -zxvf jlim-1.0.2.tar.gz
cd jlim-1.0.2
R CMD INSTALL jlimR_1.0.2.tar.gz
```

In case that it is preferred to install R extensions in your home directory (e.g. ~/R) instead of the default system path, please do the following instead:

```
Rscript -e 'install.packages("getopt", "~/R",
repos="http://cran.r-project.org")'

cd jlim-1.0.2

R CMD INSTALL -l ~/R jlimR_1.0.2.tar.gz
```

And then, add your local R library path to **R_LIBS** environment variable in .bashrc or .profile as:

```
export R_LIBS=~/:$R_LIBS
```

How to run JLIM on provided example

Example data

In jlim-1.0.2/examples, we provide the following dataset: - **MS/MS-indexSNP.tsv** Master file containing the list of index SNPs of known GWAS hits to run JLIM. As an example, we include two GWAS hits and ImmunoChip intervals covering them (chr1:160697074-160933065 and chr12:57626582-58488667). The GWAS hits are originally from IMSGC. Nature Genetics 2013. - **MS/** Summary statistics of association to Multiple Sclerosis as primary trait. A separate summary statistic file is provided for each interval listed in the indexSNP file. The data are originally from IMSGC. Nature Genetics 2013. - **LCL/** eQTLs in Lymphoblastoid cell lines (LCL) as secondary trait. A separate directory is provided for each interval listed in the indexSNP file. For each interval, two files are provided for each gene to be analyzed: summary association file (.assoc.linear.gz) and permutation file (.mperm.dump.all.gz), which are both generated by plink. In addition, plink ped file (LCL.ped.gz) is provided to calculate in-sample LD matrix. The data are originally from Geuvadis project (Lappalainen et al. Nature 2013). - **ld0/** Reference LD panel for each interval listed in the indexSNP file. Data is from the 1000 genomes project phase 3 (release 20130502).

Running JLIM on

Using the indexSNP file, jlim_gencfg.sh scans provided eQTL folder to generate a config file. Then, run_jlim.sh can run JLIM test on the config file as the following (note that the r2 resolution limit is set to default, 0.8):

```
cd jlim-1.0.2

bin/jlim_gencfg.sh --tr1-name MS --tr1-dir examples/MS --tr2-dir examples/LCL
--idxSNP-file examples/MS/MS-indexSNP.tsv --refld-dir examples/ld0/ --out
examples/jlim.cfg.tsv

bin/run_jlim.sh examples/jlim.cfg.tsv 0.8 examples/jlim.out.tsv >
examples/jlim.out.log
```

File formats

IndexSNP file

The indexSNP file is a tab-delimited file with five columns: - CHR: chromosome - SNP: SNP ID of index SNP - BP: bp position of index SNP - STARTBP: start of interval around index SNP (bp) - ENDBP: end of interval around index SNP (bp)

CHR.STARTBP.ENDBP combination will be used as an interval identifier to locate files of primary association statistics, secondary association statistics, and reference LD. In each interval, the most

associated SNP will be automatically picked based on primary association p-values, and then the analysis window will be set up to +/- 100kb around the most associated SNP. The JLIM analysis window will not extend over the original interval specified in the indexSNP file. The exact bp position of index SNP does not matter for JLIM as it will pick the most associated SNP within the specified interval.

Example:

CHR	SNP	BP	STARTBP	ENDBP
1	rs35967351	160711804	160697074	160933065
12	rs10783847	58196447	57626582	58488667

Primary trait summary statistics file

Primary trait file is named by *TraitName.CHR.STARTBP.ENDBP.txt*. It is space-delimited and has CHR, BP, and SNP. It also has to carry one of STAT, T, or P. If P is specified, the two-sided P-value will be transformed into Z-statistic. STAT or T will be approximated as Z-statistic.

Example:

CHR	BP	SNP	P
1	160698276	rs564141	9.317999999999999E-3
1	160700126	rs116424351	0.83330000000000004
1	160702475	rs576627	8.5309999999999997E-2
1	160702791	rs579272	1.388E-2
1	160703965	rs983494	8.0780000000000002E-7

Secondary trait files

Three types of files are required for secondary trait: summary statistics file, permutation file, and genotype file. Summary statistics file is a plink linear assoc file. Permutation file is a plink permutation file. Genotype file is a plink ped-format file for all individuals in the cohort. **Currently, JLIM does not allow missing genotypes for the secondary trait genotype file.** All three files should have an identical set of markers. In case of eQTL, summary statistics and plink permutation files are generated for each genes, but a genotype file is shared for all genes in the interval. The gene-specific files are named as *TraitName.SubID.assoc.linear.gz* and *TraitName.SubID.mperm.dump.all.gz*. The SubID can be a gene identifier in case of eQTL. The genotype file is fixed to *TraitName.ped.gz*.

Reference LD file

Reference LD files are provided one for each interval. It is a tab-delimited file without header. The file name is specified as *locus.CHR.STARTBP.ENDBP.txt.gz*. Each row is a marker, and it contains the following columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, and is followed by two alleles for each individual.

Config file

The config file is generated by `jlilim_gencfg.sh`, and contains the following column. JLIM will be executed on each row separately. The config file contains the following columns: - `maintrID`: name of main trait. Specified by `-tr1-name`. - `chrom`: chromosome - `idxSNP`: index SNP name (SNP in indexSNP file) - `idxBP`: index SNP pos (BP in indexSNP file) - `idxP`: P-value of association to primary trait at `idxSNP` - `idx2BP`: SNP that is most associated to primary trait (automatically selected by JLIM) - `idx2P`: P-value of association to primary trait at `idx2BP` - `start`: start of JLIM analysis window - `end`: end of JLIM analysis window - `sectrID`: name of secondary trait ("LCL" in the above example) - `sectrsubID`: sub-identifier of tested association in secondary trait folder (gene names in the above example) - `minP2`: smallest p-value of association to secondary trait within JLIM analysis window. By default, `jlilim_gencfg.sh` excludes secondary traits with `minP2 >= 0.01` from config file. The cut-off can be changed by `-p-tr2-cutoff` option.

- `maintr`: file path to main trait association file - `sectr`: file path to secondary trait association file - `refld`: file path to reference ld file - `mainld`: file path to in-sample ld of main trait cohort if specified (use `refld` if set to ".") - `seclld`: file path to in-sample ld of secondary trait cohort - `perm`: permutation file of secondary trait association

Example:

```
maintrID|chrom|idxSNP|idxBP|idxP|idx2BP|start|end|sectrID|sectrsubID|mainld|maintr|sectr|refld|mainld|seclld|perm
MS|1|155267352|160711046|4.79e-08|160711046|700-08|160838276|160831804|LCL|SLAMP7_ENSG00000026751.1|474-08|examples/MS/MS-1.500897074.160933065.LCL.SLAMP7_ENSG00000026751.1.assoc.linear.gz|examples/ld0/ld0.ms.1.500897074.160933065.dat.gz|examples/LCL/ld0.ms.1.500897074.160933065.LCL.ped.gz|examples/LCL/ld0.ms.1.500897074.160933065.LCL.ped.gz|examples/LCL/ld0.ms.1.500897074.160933065.LCL.SLAMP7_ENSG00000026751.1.mperm.dump.gz
MS|2|110783847|88196447|3.005e-10|88196447|3.005e-10|88196447|88296447|LCL|NHBC_ENSG00000175189.2|3.0054739|examples/MS/MS.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz|examples/LCL/ld0.ms.2.57626582.58488667.dat.gz
```

Output file

The JLIM out contains two columns: - `STAT`: JLIM statistic - `p`: p-value by permutation. The p-value of 0 means that the JLIM statistic is more extreme than permutation.

Example:

STAT	p
3.75229222129719	0
-3.6281538792694	0.86

How to run JLIM on your data

- Specify the indexSNP file.
- Provide primary trait association file.
- Generate the reference LD. We provide a sample script to extract LD info for non-Finnish Europeans from downloaded 1000 genomes project vcf files. For example, if the vcf files are present in `/data/1000genomes/ftp/release/20130502`:

```
bin/fetch.refld0.EUR.pl /data/1000genomes/ftp/release/20130502/
examples/MS/MS-indexSNP.tsv examples/ld0/
```

- Generate secondary trait genotypes (.ped.gz files).
- Generate secondary trait association and mperm files. For example,

```
plink --bfile region1 --pheno gene1tx.pheno --linear --out LCL.gene1 --covar  
yourcov --mperm 1000 --merpm-save-all --chr yourCHR --from-bp yourSTARTBP --  
end-bp yourENDBP
```

```
gzip LCL.gene1.assoc.linear
```

```
gzip LCL.gene1.mperm.dump.all
```

- Run `jlim_gencfg.sh` and `run_jlim.sh` similarly as the above.

From:

<https://sunyaevlab.hms.harvard.edu/wiki/!web/> - **Sunyaev Lab**

Permanent link:

<https://sunyaevlab.hms.harvard.edu/wiki/!web/jlim>

Last update: **2022/10/15 23:04**

