# RVD27 command line program instruction



Figure 1 The overall Flowchart of RVD2 program



Figure 2 The structure of hypothesis testing functions in RVD2

|  |  |  |
| --- | --- | --- |
|  | one-sided | two-sided |
| one sample | Germline test  (HCC1187 data) | NA |
| two samples (control-case paired) | Paired difference test  (Synthetic data) | Somatic test  (HCC1187 data) |

Figure 3 Overview of three test functions. Chi square test, which can improve specificity, is optional in all the tests. In the paper we performed paired difference test on the synthetic dataset, germline test and somatic test on the clinical dataset.



Figure 4 The flowchart of a demo using computer generated simulation data

**RVD2 command line syntax:**

usage: rvd [-h] [--version] [-v]

{gen,gibbs,one\_sample\_test,germline\_test,paired\_difference\_test,somatic\_test}

RVD is a hierarchical bayesian model for identifying rare variants from short-read sequence data.

Positional arguments:

{gen,gibbs,one\_sample\_test,germline\_test,paired\_difference\_test,somatic\_test}

sub-command help

**gen** Demo: generate simulation sample data from the RVD model

-h, --help show this help message and exit

-N Number of replicates in computer simulation data

-J Number of positions in computer simulation data

-s SEEDINT random process seed.

**gibbs** fit the RVD model using Gibbs sampling

*positional arguments:*

dcfile depth chart file name

*optional arguments:*

-h, --help show this help message and exit

-o OUTPUTFILE output HDF5 file name, default (output)

-p ,--pool POOL number of workers in multithread pool, default None

-g, --ngibbs NGIBBS sampling size, default 4000

-m, --nmh NMH Metropolis-Hastings sampling size, default 10

-b, --burnin BURNIN , default 0.2

-t, --thin THIN thin, default 2

-s SEEDINT random process seed.

**one\_sample\_test** One side Bayesian posterior density test of one single sample

*positional arguments:*

HDF5Name HDF5 sample file

*optional arguments:*

-h, --help show this help message and exit

-i, --intvl INTVL interval of interest in in posterior distribution.

-a, --alpha ALPHA hypothesis test credible level

-o OUTPUTFILE output HDF5 file name, default (output)

**germline\_test** Germline test on a single sample, which includes a one side Bayesian density

test and an optional chi square test.

*positional arguments:*

HDF5Name HDF5 sample file

*optional arguments:*

-h, --help show this help message and exit

-i, --intvl INTVL interval of interest in in posterior distribution.

-a, --alpha ALPHA hypothesis test credible level

-o OUTPUTFILE output HDF5 file name, default (output)

-c, --chi2 Whether to include chi square test in the germline test, default True (Include)

**paired\_difference\_test** One sided posterior density difference test on control-case paired sample, with

an optional chi square test.

*positional arguments:*

controlHDF5Name HDF5 control sample file

caseHDF5Name HDF5 case sample file

*optional arguments:*

-h, --help show this help message and exit

-i, --intvl INTVL interval of interest in in posterior distribution.

-a, --alpha ALPHA hypothesis test credible level

-o OUTPUTFILE output HDF5 file name, default (variants\_paired\_difference)

-c, --chi2 Whether to include chi square test in the paired difference test, default True

-s SEEDINT random process seed.

-n N Posterior difference distribution sampling size.

**somatic\_test** Somatic test, which includes a two sided posterior density difference test and

chi square test on the control-case paired sample.

*positional arguments:*

controlHDF5Name HDF5 control sample file

caseHDF5Name HDF5 case sample file

*optional arguments:*

-h, --help show this help message and exit

-i, --intvl INTVL interval of interest in in posterior distribution.

-a, --alpha ALPHA hypothesis test credible level

-o OUTPUTFILE output HDF5 file name, default (variants\_paired\_difference)

-c, --chi2 Whether to include chi square test in the paired difference test, default True

-s SEEDINT random process seed.

-n N Posterior difference distribution sampling size.

**Optional arguments:**

-h, --help show this help message and exit

--version show program's version number and exit

-v, --verbose increase verbosity (specify multiple times for more)