

PREST-plus Documentation

V4.09

Released October 5th, 2010

Input file formats

The genotype data are read from a combination of ped and map files, similar to the ones used by plink. The only difference is that the genotypes must be coded as 1 or 2. If you have an input file for plink you can easily convert it for use with prest with the command `--recode12`.

The pedigree file contains one line per individual with fields separated by space. The first six columns are pedigree id, individual id, father id, mother id, sex and phenotype (not used). Then the genotypes follow, encoded by 1 and 2 and separated by space.

The map file contains one line per marker with 4 fields separated by space: chromosome, marker name, cm distance, bp. The chromosome and bp are not used by prest but are included in the map file for convenience.

Most commonly used command line options

`--file file` : specify pedigree input file

`--map file` : specify input map file

`--wped` : perform within pedigree testing

`--aped` : perform across pedigree testing

`--out prefix`: specify alternative prefix for output files

Note: One of `aped` or `wped` (or both) must be specified, otherwise no IBD calculations are performed.

Additional tests:

`--cm` : specify that a genetic map is provided in the map file, which allows evaluation of the normally approximated statistics

`--mlrt` : Run the maximum-likelihood ratio test. This requires that the `--cm` option is also specified

Quick start

Suppose your input files are in the files `geno.map` and `geno.ped`. If you wish to test only pairs within a pedigree then run command line used to run `prest` is:

```
./prest --file geno.ped --map geno.map --wped
```

To test also pairs across pedigrees, run the following command:

```
./prest --file geno.ped --map geno.map --wped --aped
```

The second option might take a long time to run so beware.

If you have SNP genotype data, you don't need to run `prest` with all the SNPs as this very time consuming and does not improve the accuracy. Generally a random selection of 50k SNPs is more than enough to get accurate IBD estimates.

Always remember to include only the autosomal chromosome 1-22 in the input files, as `prest` currently does not model non-autosomal chromosomes.

Output

The main output file is `prest_results` (unless you have used the command `--out` to redirect it to another file). One line per pair of individuals tested is output with all the statistics computed for them separated by space. The first line is a header lines describing the fields of the output.

Summary of output fields:

FID1,IID1 : id of first individual

FID2, IID2 : id of second individual

reltype : relationship type of pair, among the 11 tested relationships (see below)

commark : number of markers commonly typed in the two individuals

p.IBS0, p.IBS1, p.IBS2 : computed mean IBS sharing of 0/1/2 alleles.

IBS : mean IBS sharing

p.IBD0 , p.IBD1 , p.IBD2 : computed IBD statistics

Relationship types as encoded in prest and expected IBD probabilities

ID	NAME	CODE	KINSHIP	IBD0	IBD1	IBD2
1	Full-Sibling	FS	0.250	0.25	0.50	0.25
2	Half-Sibling	HS	0.125	0.50	0.50	0
3	Grandparent-grandchild	GPC	0.125	0.50	0.50	0
4	Avuncular	AV	0.125	0.50	0.50	0
5	First-Cousin	FC	0.0625	0.75	0.25	0
6	Unrelated	U	0	1	0	0
7	Half-Avuncular	HA	0.0625	0.75	0.25	0
8	Half-First-Cousin	HFC	0.03125	0.875	0.125	0
9	Half-Sib+First-Cousin	HSFC	0.1875	0.375	0.5	0.125
10	Parrent-Offspring	PO	0.25	0	1	0
11	MZ-Twins	MZ	0.5	0	0	1

Changes an improvements since v3.02:

- Prest is now integrated with altest and can test alternative hypothesis for a specified pair, using the command-line option "--pair"
- The EM algorithm for computing the probability of IBD is now up to 10-20x faster thanks to a convergence acceleration procedure based on 2D Newton-Raphson iteration.

- Prest can handle more effectively GWA datasets with millions of SNPs, although we recommend pruning the set of SNPs to around 50k since there are diminishing returns in accuracy of IBD estimated when more SNPs.
- Prest now is more user friendly with a new set of command line options and is flexible with regards to input file. Ped files are no longer restricted in having only numbers as family and individual ids and can have any string. A map file is now required, which provides markers names for reporting mendelian errors. Genetic map can also be included in the file and will be taken into account when performing the MLRT test.