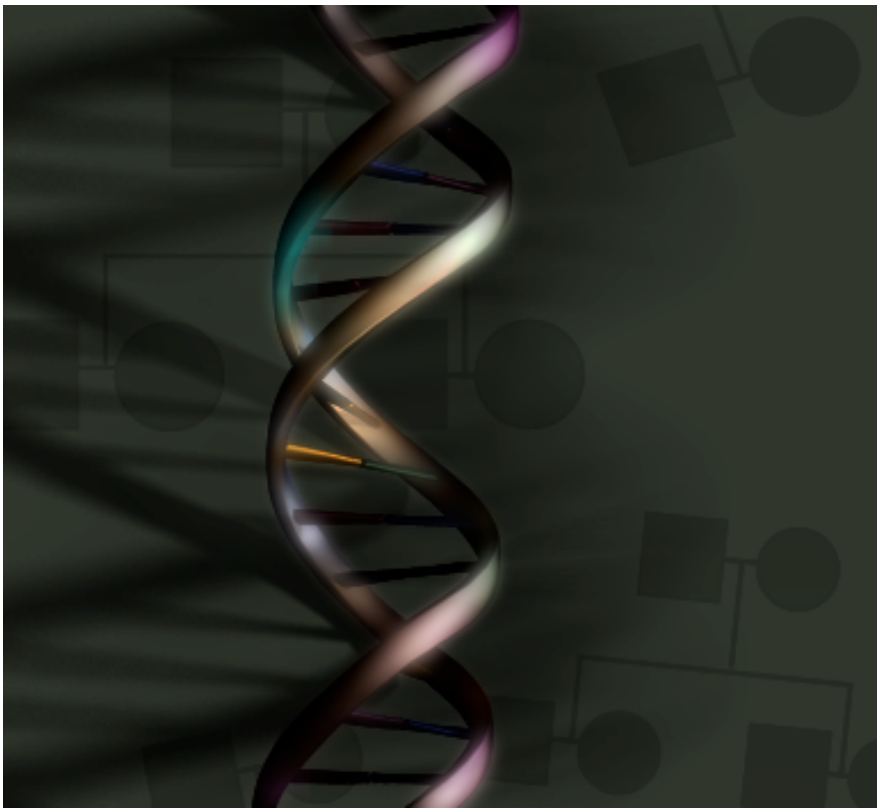
MDR-PDT USER'S GUIDE



MDR-PDT VERSION 2.0.1

Compile Date: Dec. 2008

519 Light Hall • 2215 Garland Ave • Nashville, TN •
http://chgr.mc.vanderbilt.edu/twiki/bin/view/MdrPdtBeta/WebHome

MdrPDT User's Guide - Version 2.0

Table of Contents / Reference

About this document.

This document is an introduction to the use of the MDR-PDT analysis program described in Martin, et al (2006). The analysis program described is currently approaching completion, however as testers give us input, it is possible that things will change. Be sure that the version of this document matches the version of the application being used.

Because the details of the method are described in Martin, et al (2006), this document will be focused on the use of the application rather than how the analysis works.

Datasets Appropriate for MDR-PDT

While MDR (Ritchie, et al., 2001) was designed to discover locus-disease associations in case-control data, MDR-PDT was designed to discover locus-disease associations in pedigree data using the genotype-PDT statistic (Martin et al., 2003). This version of MDR-PDT is capable of analyzing input in standard pedigree format, which allows for a mix of trio data and discordant sibships. In both types of family structures, the informative element is the transfer of genotypes to an affected child as compared with the genotypes transferred to an unaffected child. In the trio data, a virtual unaffected sibling is created who is the genotypic complement of the affected child as derived from the original affected child and the two parents.

Each locus must contain only two alleles and must denote missing data with a zero for each allele that is missing.

Pedigree Format

All data files intended for use with MDR-PDT should be written with one of the following delimiters: space, tab or commas. The application ignores empty fields, so two or more spaces side beside are treated as a single space.

By default, pedigree data is expected to be formatted using MERLIN QTDT format, but an older version, which includes a ten column header, can be used with a simple configuration change. For both formats, Affected status is denoted with a 2 and and Unaffected status is denoted with a 1.

Column Number	Description
1	Pedigree number (must be an integer)
2	Individual ID number (must be an integer)
3	ID of father (0 if this individual is a founder)
4	ID of mother (0 if this individual is a founder)
5*	First offspring ID
6*	Next paternal sibling ID
7*	Next paternal sibling ID
8	Sex (1=male, 2=female, 0=unknown)
9*	Proband status (1=proband, all others have a 0 in this field.)
10	Disease status (2=affected, 1=unaffected, 0=unknown)
11	The 1rst allele pair (must be pairs of integers)
12	
...	...
n-1	The 1rst allele of the last pair
n	

TABLE 1: A DESCRIPTION OF THE DATA IN PEDIGREE FORMAT.

* Used only in non-merlin format. Merlin format is enabled by default

An optional .dat file can be used to label SNPs with rs-numbers to enhance readability of the output. This option is available ONLY with Merlin formatted data and the format is similar to the MERLIN .dat format. This format is a two column structure where the first column denotes the type of field (A - Disease, M - Marker). For MdrPDT, we require the first entry to be the disease. An example of a 3 SNP example might look similar to Table 2.

A	disease_label
M	marker-1-label
M	marker-2-label
M	marker-3-label

TABLE 2: A DESCRIPTION OF THE MERLIN DAT FORMAT.

Application Usage

There are currently 3 modes of operation: configuration generation, analysis and model exploration.

Configuration Generation

To simplify the generation of configuration files, the application provides a way to display an example configuration file. Users can capture this display to a file and use it for analyses. The example configuration produced depends on the Analysis Style selected as the 1rst argument. Currently, there is only one option, PDT.

Optional arguments may follow in order to specify important runtime options. While all other arguments are optional, they must be present in the order on the command line to be interpreted correctly (i.e. If you wish to change the Max Model Size, you have to set Data Filename and Min Model Size as well even if you are happy with the defaults).

ARG #	PURPOSE	POSSIBLE VALUES
1	Analysis Type	PDT
2	Data Filename	Name of your pedigree formatted input file
3	Min. Model Size	1 to Max Model Size
4	Max. Model Size	Min Model Size or more
5	Cross Validation Count	How many folds are to be used in cross validation. (1 means don't use cross validation)
6	P-Test Count	Set the number of PTests to be run
7	Random Seed	Override the default random seed

Distributed with the application is a simulated data set named HighPower-A.ped. In order to create configuration file for running a 1 & 2 SNP search with permutation tests on the example pedigree file, type the following:

```
./mdrpdt2 PDT HighPower-A.ped 1 2 > HighPower-A.mdrpdt
```

The resulting file, `HighPower-A.mdrpdt` should be ready to use. At this point, changes can be made to `HighPower-A.mdrpdt` according to the user's needs.

The following is the result of running the line above with the the beta release version of the application.

```
bash-3.1$ ./mdrpdt2 PDT HighPower-A.ped 1 2 > HighPower-A.mdrpdt
bash-3.1$ more HighPower-A.mdrpdt
########################## Input ############################
# The dataset to be used
INPUTFILE                       HighPower-A.ped
# Input format. Merlin has 6 columns in ped file and an optional .dat file
MERLIN_FORMAT                   NO
# Specify the location of the optional dat file (contains the labels for each locus
MERLIN_DAT
# Set the value associated with affected status
AFFECTED_VALUE                  2
# Set the value associated with the unaffected status
# Individuals with neither Affected nor Unaffected will be ignored by the analysis
UNAFFECTED_VALUE                1
# Ignore 0 or more pedigrees from the data
EXCLUDE_PEDIGREES
# Ignore 0 or more SNPs (1...N) from the data
EXCLUDE_LOCUS
# Maximum amount of missing data before a SNP is ignored from analysis
MISSING_THRESHOLD               0.1


########################## Basic Settings ###################
# Describes the type of models of interest. Models consist of 1 or more SNPs.
# The minimum number of SNPs in a model to be investigate.
# Valid Settings: 1..COMBO_END
COMBO_START                     1
# The maximum number of SNPs to be considered.
# Valid Settings: [COMBO_START..MAX_INT)
COMBO_STOP                      2
# Set the number of cross validation folds are used in analysis
# Recommend settings: 1, 5, 10 (1 is no cross validation)
CROSSVALINTERVAL                5
# Maximum number of models to be reported
REPORTMODELCOUNT                5
# Threshold for reporting models. At most REPORTMODELCOUNT will be reported (those are
sorted according to the T-Statistic
REPORT_THRESHOLD                0
# Write out details regarding the contents of each cross validation fold
VERBOSE_FOLDING                 0


######################### Permutation Tests ##################
# Number of permutation runs to be executed. 1000 is recommended
PTEST_COUNT                     1000
# The seed associated with the tests. Each test gets a new seed
PTEST_SEED                      1397
# Short circuit the permutation tests. See manual for explanation
PTEST_SHORTCIRCUIT              0
# How many simultaneous threads will be run
# Each PTest can theoretically be run in it's own thread (Multiple threads
# will not benefit if no ptests are being performed)
THREAD_COUNT                    1
```

```
# Determine how to handle mendelian errors when encountered.
# Acceptable Values:
#   1 - Report errors, but do nothing
#   2 - Report errors, and zero out loci in families where genotyping error has been
found
#   3 - Report errors and remove pedigrees where the number of genotyping errors exceeds
threshold
MENDELIAN_ERROR_LEVEL              1
# Set the threshold, if level is 3
MENDELIAN_PEDIGREE_THRESHOLD       0



######################### Report Names ########################
# Pedigree report (genotypes and folding details)
EXT_PEDIGREE                       pedigree
# Distribution report (each item from the distribution)
EXT_DISTRIBUTION                   dist
# Write a copy of the 'cleaned' output to file
WRITE_CLEAN_DATAFILE               No
# Extension of the file (above)
EXT_CLEANED_DATA                   ped
```

Because this dataset was created prior the incorporation of MERLIN style input, we have to change the MERLIN_FORMAT setting to NO. Notice the change made in Bold/Red above.

Analysis

To perform analysis, simply create a valid configuration file and execute the application with the configuration file as the sole parameter.

To run the analysis setup with the example data file using the configuration file, HighPower-a.mdrpdt, type the following:

Example:

```
./mdrpdt2 HighPower-a.mdrpdt
```

The following is an excerpt from a run based on the example dataset included with the application.

```
./mdrpdt2 HighPower.mdrpdt
                                Model Size: 1-2
                    Cross Validation Folds: 5
                         Permutation Tests: 1000
                        Random Number Seed: 1397
                    Configuration File Name: HighPower-A.mdrpdt
                         Dataset File Name: HighPower-A.ped
            Total Number of Loci in Dataset: 25
                        Excluded Pedigrees: None
                  Total Genotype Errors: 0
                         Total Individuals: 1500
                Participating Individuals: 500 ( 500A | 0U )
                  Participating Founders: 1000
```

```
                    Participating DSPs: 500
```

This is just an summary of the analyses that have been run.

Application output continued:

Missing Data

Missing data can be very disruptive to an MDR style analysis. To help users identify potential problems, MdrPDT reports the percentage of Missing data at each SNP, as well as the number if each SNP is found in each fold.

```
--------------- Missing Data -----------------
SNP              Fold 1         Fold 2         Fold 3         Fold 4         Fold 5    Total
Missing
 1              100A/100U      100A/100U      100A/100U      100A/100U      100A/100U      0%
 2              100A/100U      100A/100U      100A/100U      100A/100U      100A/100U      0%
 3              100A/100U      100A/100U      100A/100U      100A/100U      100A/100U      0%
 4              100A/100U      100A/100U      100A/100U      100A/100U      100A/100U      0%
 5              100A/100U      100A/100U      100A/100U      100A/100U      100A/100U      0%
                                                  ...
                    Loci For Analysis: 25
```

The first few lines represent details for the specific models that were selected with the highest t-statistic for each model size.

```
Fold 1 Details ( 15  ):

 Genotype       Affected  Unaffected
      15      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
     1/1*       148/37       141/42          289      1.0496
     1/2*       197/47       182/47          379      1.0824
     2/2         55/16        77/11          132      0.7143
  ------------------------------------------------------------------------
                   100          100          200
Missing:          0.00%        0.00%

            T-Statistic (Training) 2.04
             T-Statistic (Testing) -1.00
                Matched Odds Ratio 0.67


Fold 2 Details ( 13  ):

 Genotype       Affected  Unaffected
      13      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
     1/1*       188/36       155/37          343      1.2129
     1/2        155/51       193/54          348      0.8031
     2/2*        57/13         52/9          109      1.0962
  ------------------------------------------------------------------------
```

```
                    100          100          200
Missing:           0.00%        0.00%

            T-Statistic (Training) 2.67
             T-Statistic (Testing) 0.43
                Matched Odds Ratio 1.13
```

(Repeated for each fold for the top 1 SNP models)

```
Fold 1 Details (  5 10  ):

 Genotype       Affected  Unaffected
  5   10      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
 1/1 1/1*       202/52       164/37          366      1.2317
 1/2 1/1         33/12        75/28          108      0.4400
 2/2 1/1*         15/4         11/1           26      1.3636
 1/1 1/2         57/11        88/19          145      0.6477
 1/2 1/2*        71/17         40/9          111      1.7750
 2/2 1/2           0/0          5/1            5      0.0000
 1/1 2/2*        16/1         15/2           31      1.0667
 1/2 2/2*         6/3          2/2            8      3.0000
 2/2 2/2*         0/0          0/1            0         nan
  ------------------------------------------------------------------------
                    100          100          200
Missing:           0.00%        0.00%

            T-Statistic (Training) 5.78
             T-Statistic (Testing) 3.65
                Matched Odds Ratio 3.27


Fold 2 Details (  5 10  ):

 Genotype       Affected  Unaffected
  5   10      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
 1/1 1/1*       205/49       158/43          363      1.2975
 1/2 1/1          36/9        85/18          121      0.4235
 2/2 1/1*         16/3         11/1           27      1.4545
 1/1 1/2         54/14        84/23          138      0.6429
 1/2 1/2*        67/21         41/8          108      1.6341
 2/2 1/2           0/0          5/1            5      0.0000
 1/1 2/2*        15/2         12/5           27      1.2500
 1/2 2/2*         7/2          3/1           10      2.3333
 2/2 2/2           0/0          1/0            1      0.0000
  ------------------------------------------------------------------------
                    100          100          200
Missing:           0.00%        0.00%

            T-Statistic (Training) 6.22
```

```
T-Statistic (Testing) 2.97
   Matched Odds Ratio 2.73
```

(Repeated for each fold for the top 2 SNP models)

As the permutations are performed, a dot will be added to the log. When all tests have completed, the distribution size will be reported.

```
-------------- Running PTests (100) ------------
.....................................................................................
..........

Distribution Size: 100
```

Based on the various model's estimated p-values, the final part of the report contains all "statistically" interesting models. Interesting models are those whose p-values fall below the threshold set by the configuration parameter: REPORT_THRESHOLD. The distribution required for the t-statistics is an omnibus distribution based on the model's MOR value.

In the example below, a REPORT_THRESHOLD of 0 was used. This means that all REPORTMODELCOUNT models will be reported for each order being searched. In the example below, only the top two ranks of the single locus search are shown. The "winning" model for a given rank is selected based first on cross validation consistency followed by the matched odds ratio in the case of a tie. P-Values are determined based on the average MOR over all folds of a given rank.

Model Rank	XV Cons.	Model	MDR-PDT Train. Stat.	MDR-PDT Test Stat.	Matched Odds Ratio	MOR P-Value
		[15]	2.043	-1.000	0.667	
		[13]	2.674	0.429	1.130	
		[13]	2.663	0.525	1.148	
		[13]	2.263	1.260	1.429	
		[13]	2.252	1.286	1.450	
1	4	[13]	2.463	0.875	1.165	0.32
		[12]	1.969	-0.277	0.857	
		[8]	1.938	0.152	1.048	
		[2]	1.829	-0.480	0.857	
		[2]	1.904	-0.267	0.931	
		[8]	1.858	0.309	1.100	
2	2	[8]	1.898	0.231	0.959	0.63

Model Exploration

During regular analyses, only the best model for each model size is described in high detail. However, should users wish to explore other models, mdr-pdt provides a way to interactively explore any model they wish.

Exploration is easy to do, but it requires a few pieces of information. First, the configuration file specifying details about the data (format and location) as well as the analysis style desired. One or more models should be given on the command line separated by space. The models are specified using the SNPs position in the repository separated by "x"s.

Example:

If in a previous analysis two other significant models were reported as 1x10 and 5x10, we could type the following:

```
./mdrpdt2 HighPower-A.mdrpdt 1x10 5x10
```

Below is an example of the exploration of a single model, 2x10. For the sake of space, only the first two folds are shown.

```
bash-3.1$ ./mdrpdt2 HighPower-A.mdrpdt 2x10
                                Model Size: 1-2
                    Cross Validation Folds: 5
                         Permutation Tests: 100
                       Random Number Seed: 1397
                  Configuration File Name: HighPower-A.mdrpdt
                         Dataset File Name: HighPower-A.ped
            Total Number of Loci in Dataset: 25
                       Excluded Pedigrees: None
                  Total Genotype Errors: 0
                         Total Individuals: 1500
                Participating Individuals: 500 ( 500A | 0U )
                  Pariticipating Founders: 1000
                       Participating DSPs: 500


Fold 1 Details (  2 10  ):

 Genotype       Affected  Unaffected
  2   10      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
 1/1 1/1*       131/37       124/30          255      1.0565
 1/2 1/1         97/28       106/27          203      0.9151
```

```
 2/2 1/1*         22/3         20/9           42      1.1000
 1/1 1/2          68/18        72/11          140     0.9444
 1/2 1/2          44/10        50/15          94      0.8800
 2/2 1/2*         16/0         11/3           27      1.4545
 1/1 2/2*         18/2          9/3           27      2.0000
 1/2 2/2           3/2          7/2           10      0.4286
 2/2 2/2*          1/0          1/0            2      1.0000
  ------------------------------------------------------------------------
                   100          100          200
Missing:          0.00%        0.00%

           T-Statistic (Training) 1.68
            T-Statistic (Testing) -0.41
               Matched Odds Ratio 0.89


Fold 2 Details (  2 10  ):

 Genotype       Affected  Unaffected
  2   10      Trn/Test     Trn/Test         Total         Ratio
  ------------------------------------------------------------------------
 1/1 1/1*       134/34       123/31          257      1.0894
 1/2 1/1        101/24       105/28          206      0.9619
 2/2 1/1          22/3         26/3           48      0.8462
 1/1 1/2*        67/19        64/19          131      1.0469
 1/2 1/2         42/12        55/10           97      0.7636
 2/2 1/2*         12/4         11/3           23      1.0909
 1/1 2/2*         16/4         10/2           26      1.6000
 1/2 2/2           5/0          6/3           11      0.8333
 2/2 2/2*          1/0          0/1            1         inf
  ------------------------------------------------------------------------
                   100          100          200
Missing:          0.00%        0.00%

           T-Statistic (Training) 1.51
            T-Statistic (Testing) 0.76
               Matched Odds Ratio 1.26
```

Pedigree Report

During the parsing of a pedigree file, and throughout the generation of virtual sibs and discordant sib pairs, mdr-pdt logs important details.

Pedigree Contributions

Below is the first few lines from the pedigree report generated when analyzing some data with a few errors in one of the pedigrees.

```
    ---------- Mendelian Error Report ----------
```

```
Pedigree: 601  5/25 Affected Loci [ 0 1 3 4 6 ]
 - 1    Missing: P)0 C)0  no mendelian errors found
 - 2    Missing: P)0 C)0  no mendelian errors found
 - 7    Missing: P)5 C)0  Errors: 5   (1) 2  1 x 1  (2) 3  1 x 2   (4) 3  2 x 1   (5) 1
3 x 2   (7) 3  1 x 2
  * Purging 3 members due to questionable genotyping information.
Pedigree: 602  No Errors
Pedigree: 603  No Errors
```

For individual 7 of pedigree 601, 5 genotypes are found to have genotype errors. The genotype is presented for each locus, followed by each of the parent's genotypes.

At the end of the report, it is noted that, due to these errors exceeding a threshold, the pedigree is removed from analysis:

```
-------------------- DSP Contribution ------------------
Pedigree: 601 Contributed 0 DSPs.
Pedigree: 602 Contributed 1 DSPs.
Pedigree: 603 Contributed 1 DSPs.
```

Distribution Report

The contents of the distribution report is simply the values observed from the highest scoring model for each of the N runs where the models are sorted by increasing MOR values.

Index	MOR	Model	R. Index
1	0.7875	14 16	100
2	0.8195	5 16	99
3	0.8496	2 24	98
4	0.8519	8 24	97
5	0.8723	12 15	96
6	0.8775	12 21	95
7	0.8882	13	94
8	0.9062	10	93
9	0.9185	17	92
10	0.9216	7	91
11	0.9241	16 18	90
12	0.9332	3 11	89
13	0.9359	17	88
14	0.9424	24	87
15	0.9534	2 13	86
16	0.955	14	85
17	0.9663	15	84
18	0.9701	2 13	83
19	0.9769	10	82
20	0.9807	4	81

Compiling MDR-PDT

MdrPDT is built using the gcc (gnu c compiler) version 4.1.2 64 bit linux machines running RedHat OS. The make system should be compatible with any system with a modern version of the gcc compiler suite with standard make as long as the prerequisites below are met.

Prerequisites

STL – The Standard Template Library is assumed to be available on all machines with gcc 3.2 or higher installed. Under most circumstances, if it isn't installed, the administrator of the system can add the package on using the package manager associated with the distribution installed on the machine.

Boost 1.33.1 (or later)– Boost is an open source extension to the STL and offers a few classes used heavily by MdrPDT. It is required to have this library available before building the application. It is generally assumed that users can install or have boost installed. Boost can be downloaded at: http://www.boost.org

MPI - To run MdrPDT in parallel on a cluster, users should be have MPI installed and the headers and libraries properly situated in the gcc search paths.

Compilation

If boost is properly installed, compilation should be as easy as running make:

```
make
```

Unless a problem is encountered, the end result should be a functional verison of MdrDPT inside the bin directory.

```
rlab-torstenson:mdrpdt-2.0.1.19$ ls -ltr bin
total 3928
drwxr-xr-x  3 torstees  admin      102 Jun  8 10:22 lib
-rwxr-xr-x  1 torstees  admin   544120 Jun  8 10:25 mdrpdt-OSX
```

The name of the executable will depend on a number of possible details:

- OS - If using MinGW or OSX, the OS will be prepended at the end of the application name
- Users on 64bit platforms using a 64bit compiler might see 64 at the end
- Parallel versions will have a p at the end
- Debug versions will have a lower case d at the very end

There is no “install” script associated with this build. Users are free to move the executable to a place of their own choosing.

Some useful make arguments:

- DEBUG=1 - This will compile the application in debug mode
- USE_MPI=1 - This will compile the application to run in parallel

Multiprocessing and MdrPDT

Threading

It is becoming far more common to find inexpensive hardware sporting 2 or more processors, or cores. MdrPDT can take advantage of these extra processors by using more than 1 thread. When available, it is highly recommended to increase the number of threads to the number of processors available to the user during the time of execution. This can have a dramatic affect on the amount of time required to perform a large number of PTests.

It should be noted that there is overhead associated with threads, and thus, if a system is unable to dedicate as many processors to MdrPDT during execution as there are threads, performance can actually be adversely affected. Users should verify the number of available processors prior to beginning analysis.

Parallel Computing

MdrPDT has been configured to allow parallel execution using MPI. For users with access to a cluster, this can allow significant improvement to the speed at which the final results are available. Because each cluster can be very different, we don't provide a binary for parallel execution, but users can compile their own using a slightly different call to make. Otherwise, everything else is the same.

```
make USE_MPI=1
```

When running MdrPDT on a cluster, users are expected to know the procedures specific to the actual cluster. In General, execution will follow something along the lines of:

```
mpirun mdrpdt -n 10 config.mdrpt
```

It should be noted that multithreaded applications incur far less overhead than parallel. Users will only benefit from parallel processing if the number of nodes available to them from the cluster exceeds the number of processors that can be dedicated to their process on a single machine during execution.

Configuration Parameters

Input Parameters

INPUTFILE string Name of the pedigree dataset

INPUTFILE ../pedigree-data/somefile.ped

MERLIN_FORMAT Yes/No Indicate if data is in MERLIN, 6 col. header, format

MERLIN_FORMAT Yes

MERLIN_DAT string Name of the dat file

MERLIN_DAT ../pedigree-data/somefile.dat

This is only used if MERLIN_FORMAT is Yes

AFFECTED_VALUE integer Override default value to indicate affected status

AFFECTED_VALUE 2

UNAFFECTED_VALUE integer Override default value to indicate unaffected status

UNAFFECTED_VALUE 1

EXCLUDE_PEDIGREES list of pedigree IDs

EXCLUDE_PEDIGREES 100 101 102 110

EXCLUDE_LOCUS List of Integers

Follow the command with 1 or more loci that should not be considered during analyses. These do not shift the columnar id of subsequent snp numbers (i.e. model 10x11 is the same whether or not you excluded locus 9)

MISSING_THRESHOLD float

MISSING_THRESHOLD 0.1

MDR-PDT can ignore SNPs that have too much missing data. In the example above, MdrPDT would not use SNPs whose missing data were greater than 0.1.

Search Parameters

COMBO_START integer Specifies the minimum size of models to be evaluated

COMBO_START 1

COMBO_END integer Specifies the maximum size of models to be evaluated

COMBO_END 2

CROSSVALINTERVAL integer Number of cross validation folds to use (1 means use no cross validation)

CROSSVALINTERVAL 5

REPORTMODELCOUNT integer Number of models to be reported on after analysis.

REPORTMODELCOUNT 10

REPORT_THRESHOLD float P-Value cutoff for a model to be reported

REPORT_THRESHOLD 0.1

Users can use this to avoid reporting on statistically uninteresting models. At most, only REPORTMODELCOUNT models will be reported.

Permutation Tests

PTEST_COUNT integer Number of permutations to be performed

PTEST_COUNT 1000

PTEST_SEED integer Random number seed

PTEST_SEED 1397

This is a bit of a misnomer, but the name is left for backward compatibility. This seed affects, in addition to permutations, the groupings made for cross validation and all other random draws used by the program.

PTEST_SHORTCIRCUIT TBD

THREAD_COUNT integer

THREAD_COUNT 4

Indicate how many separate threads the PTests are to be run.

Mendelian Errors

MENDELIAN_ERROR_LEVEL integer

1) *Report errors, but do nothing*
2) *Report errors, and zero out loci in families where genotyping error has ben found*
3) *Report errors and remove pedigrees where the number of genotyping errors exceeds threshold*

Reporting

Users can change the extension of various reports to suit their needs. All reports are named after the configuration filename.

EXT_DISTRIBUTION string Set the extension used for the distribution(s) used in calculating a model's significance.

EXT_PEDIGREE string Set the extension used for the pedigree report.

WRITE_CLEAN_DATAFILE Y/N

4) *Users can have MdrPDT output a copy of the pedigree data after the cleaning has been performed.*

EXT_CLEANED_DATA string

Related Bibliography

Martin, E.R., Ritchie, M.D., Hahn, L.W., Kang, S., Moore, J.H. (2006) A Novel Method to Identify Gene-Gene Effects in Nuclear Families: The MDR-PDT. Genetic Epidemiology 30:111-123.

Related Web Sites

MDR-PDT at Marylyn Ritchie's lab – http://chgr.mc.vanderbilt.edu/ritchielab/MDRPDT.html

PDT from Eden Martin at Duke – http://www.chg.duke.edu/software/index.html

MDR at Jason Moore's lab – http://www.epistasis.org/mdr.html

MERLIN File Format - http://www.sph.umich.edu/csg/abecasis/merlin/tour/input_files.html