# User guide of software KBAT

**Hsin-Chou Yang, Hsin-Yi Hsieh and Cathy S. J. Fann**

Correspondence to hsinchou@stat.sinica.edu.tw

**Table of Contents:**

# 1. KBAT LICENSE

All copyrights are reserved by the authors of **KBAT**. We welcome any noncommercial use of **KBAT** for your own research. Please do NOT modify or distribute **KBAT** in any form without the permission of the authors. Questions regarding commercial use of **KBAT** should be directed to Hsin-Chou Yang at hsinchou@stat.sinica.edu.tw or Cathy S. J. Fann at csjfann@ibms.sinica.edu.tw. Because **KBAT** is open-access/free software, the authors assume no warranty and no responsibility for the results of analyses. If publications use results from **KBAT**, please cite the following reference: Hsin-Chou Yang, Hsin-Yi Hsieh and Cathy SJ Fann. (2007) KBAT: Kernel-based association test.

# 2. INTRODUCTION

**KBAT** (<u>K</u>ernel-<u>b</u>ased <u>a</u>ssociation <u>t</u>est) is a convenient analysis tool for disease gene association mapping. Several powerful association tests in **KBAT** have been developed based on the integrated concept of kernel smoothing, p-value weighting, p-value truncation, and p-value combination. The methods provide systematic genome-wide and candidate-region searches for disease susceptibility genes. Numerical/graphic results are shown together to provide insight into the disease-marker association in study regions. The detailed formulae of the test statistics are provided in the Appendix.

# 3. SOFTWARE DOWNLOAD AND INSTALLATION

**KBAT** was written in the R and R-GUI language, which are publicly available. Execution of **KBAT** requires installation of four programs: (1) **KBAT**, (2) R computing software, (3) Genetics module, and (4) ActiveTcl.

1. Download program **KBAT**:

    **KBAT** is available at http://www.stat.sinica.edu.tw/hsinchou/genetics/association/KBAT.htm. The

zipped file, "KBAT.zip," can be downloaded. Unzip the file to a folder containing **KBAT** and several illustrated data examples.

2. Download program R:

Users should download the R language from "The R Project for Statistical Computing" at http://www.r-project.org/. Once the R language is installed, users are suggested to download update packages for R. Users can select "<u>P</u>ackages" in the tool bar, click "<u>U</u>pdate packages" and then select a suitable mirror site to update packages.

3. Download program "Genetics":

Open "R-2.6.0". Check "Packages" in the tool bar, select "<u>I</u>nstall package(s)", and then select "Genetics" to install.

4. Download program "ActiveTcl":

ActiveTcl can be downloaded from the ActiveState website at http://www.activestate.com/store/productdetail.aspx?prdGuid=f0cd6399-fefb-466e-ba17-220dcd6f4078. Users must register before gaining access to the program.

# 4. KBAT INITIALIZATION

Once the software is installed, **KBAT** can be initialized as follows. Here, we illustrate that the **KBAT** programs are saved in the destination directory "C:\KBAT".

1.  Open "R-2.6.0".

2.  Key in "KBAT<-c("C:\\KBAT\\Real\\KBAT_R.r")" in the command line of the "R Console" window.

3.  Type "source(KBAT)" in the command line to initialize **KBAT**.

# 5. DESCRIPTION OF WORKING DIRECTORIES

The "KBAT" directory contains two sub-directories, "Real" and "Sim". "Real" is designed for real data analysis, and "Sim" is designed for simulated data analysis. As of Oct 2007, the functions for simulation data analysis were under development. The directory "Real" consists of four directories and a program file. The file "KBAT_R.r"

is the main **KBAT** program. The four directories are "Example", "Program", "Input", and "Output". The directory "Example" consists of four subdirectories for four real examples. The directory "Program" consists of several programs within **KBAT**. The directory "Input" is the default data input directory in **KBAT** (as shown in the **KBAT** interface in **Figure 1**). The working directory can be changed by keying the target directory name in the **KBAT** interface. The directory, "Output", is the default result output directory. Users can change the output directory by keying the target directory name in the **KBAT** interface.

# 6. MPDA INTERFACE AND FUNCTIONS

**KBAT** has a user-friendly interface developed using R-GUI (See **Figure 1**). The interface contains a short introduction for **KBAT**. Thirteen points are designed to provide required/optional information for **KBAT** data analysis.

1. Directory of data input: Users should provide the working directory where their data are saved.

2. Directory of results output: Users should provide the working directory where their output should be saved. Note that the output directory must exist before executing **KBAT**.

3. Total number of SNPs: Users should provide the total number of SNPs in the analysis.

4. The first marker of study region: Users should provide the first marker in the study.

5. The last marker of study region: Users should provide the last marker in the study.

6. Weighting procedure: Users should determine which type of weighting procedure will be used, including "Distance", "LD", and "LD and/or distance".

7. Data format of LD information: Users should determine which type of data format of LD information will be provided, including "Not available", "LD measure", and "Genotype data".

8. Determination of bandwidth/window size: Users should determine either bandwidth or window size to be used for window construction.

9. Bandwidth or m (window size = 2m+1): Bandwidth or window size should be

inputted. Input of multiple bandwidths or window sizes in an analysis is allowed.

10. **The truncation threshold (theta):** The truncation threshold (0 to 1) should be provided. A threshold of 1 signifies that no truncation is applied. Input of multiple truncation thresholds in an analysis is allowed.

11. Statistic**s**: Users should choose statistics to be calculated in the analysis. Selection of more than one statistic in a run is allowed.

12. Number of Monte Carlo replications: The number of Monte Carlo replications (10 to 10,000) should be provided.

13. Horizontal axis label: Users can provide a label for the horizontal axis.

Once all of these 13 points have been addressed, the icon "RUN" will submit the job for computation. **KBAT** will check the supplied data information and data files. If the inputted information is invalid or if the data files are formatted incorrectly, **KBAT** shows a warning message or an error message. If the inputted data pass the examination, **KBAT** starts the analysis, and a message "Please wait a while, KBAT is running…" will be shown in the command line. A prompt sign will appear immediately, but the computation is proceeding. Please wait until a new window with the message "Computation of KBAT is finished" acknowledges the completion of the **KBAT** computation. Note that users can interrupt **KBAT** at anytime by clicking the button  in the tool bar of the R-GUI window.

Once **KBAT** has completed its calculations, the numerical results and graphic outputs will be saved in the output directory stipulated by the user. The numerical results will be saved as a filename "output.txt", and the file will be automatically replaced if the next analysis is performed with the same output directory. The graphic results will be saved with a filename respective to the inputted conditions (bandwidth/m and threshold). We suggest that users remove figure files from previous analyses before a new analysis to eliminate confusion between figure files.

**Figure 1.   KBAT** interface
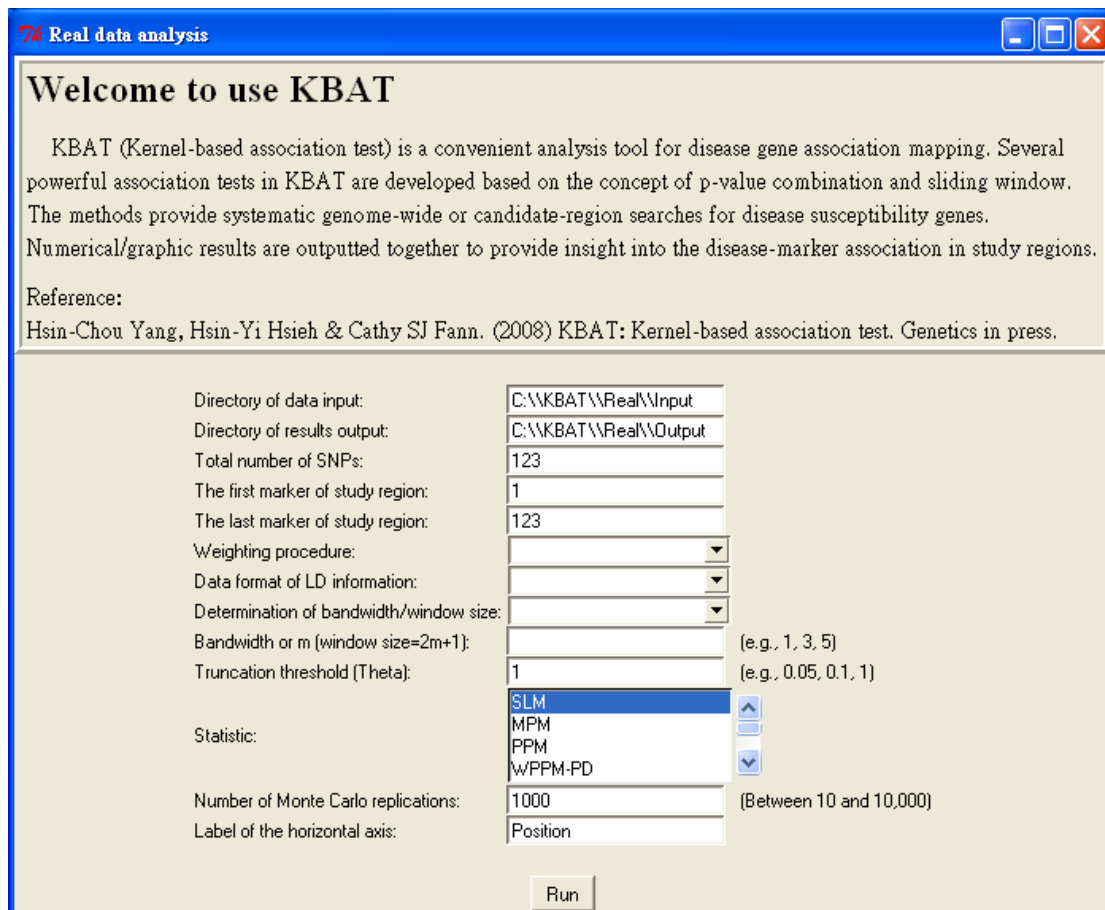
# 7. DATA INPUT FORMAT

Several data files should be provided.

1. The p-value data file with filename "pv.txt": This file contains only one column. p-values from single-locus association tests are arranged in the order of marker position.

2. The map data file with filename "map.txt": This file contains only one column. Physical or genetic positions of markers are recorded. Note that the order of marker positions in this file must match the order of the p-value data.

3. The LD data file with filename "ld.txt": This file contains three columns, which provide intermarker LD coefficients of any two markers. The first two columns are the labels of two markers. The third column records the pair-wise LD coefficient. This file is optional and only should be supplied when users calculate LD-based weights (the other type of data format for LD calculation is shown below).

4. The genotype data file with filename "geno.txt": This file contains n rows and 2p columns. The n rows represent the number of study individuals. The 2p columns

are used to denote p SNP markers, where two columns are used to present a pair of SNP marker alleles. **KBAT** only analyzes SNP markers that are diallelic. Therefore, each column contains at most two numerical values. Missing data can be handled by inputting "NA". This file is optional and only should be provided when users calculate LD-based weights.


# 8. EXAMPLES


In this section, we illustrate **KBAT** with two examples.

**Example 1: Psoriasis data analysis**

This example examines the disease association between psoriasis and 123 SNP markers. Single-locus association tests were performed using TDT-AE (REFERNCE). In total, 123 SNP markers on 17q25 (Helms et al., 2003) were analyzed. The p-values and physical positions for the 123 SNPs were saved in "C:\KBAT\Real\Example\Psoriasis". Here, we reanalyze this data with **KBAT**.

We copy files "pv.txt" and "map.txt" to the working directory "C:\KBAT\Real\Input". In the analysis, statistics SLM, PPM, WPPM-PD, and KBAT-PD were calculated under the two window sizes of 3 (m=1) and 7 (m=3). Formulae of the statistics can refer to the Appendix. Truncation was not considered in the analysis. Because LD information was not available in this example, only distance-based weight was used. The operating procedures are listed below and are also shown in **Figure 2**:

(1) Directory of data input: "C:\\KBAT\\Real\\Input".

(2) Directory of results output: "C:\\KBAT\\Real\\Output".

(3) Total number of SNPs: "123".

(4) The first marker of study region: "1".

(5) The last marker of study region: "123".

(6) Weighting procedure: "Distance".

(7) Data format of LD information: "Not available".

(8) Determination of bandwidth/window size: "Window".

(9) Bandwidth or m (window size = 2m+1): "1,3".

(10) Truncation threshold (Theta): "1".

(11) Statistic: "SLM", "PPM", "WPPM-PD" and "KBAT-PD".

(12) Number of Monte Carlo replications: "1000".

(13) Label of the horizontal axis: "Position".

(14) Press "RUN" to execute **KBAT**.

**Figure 2.** Interface for the example of psoriasis data analysis
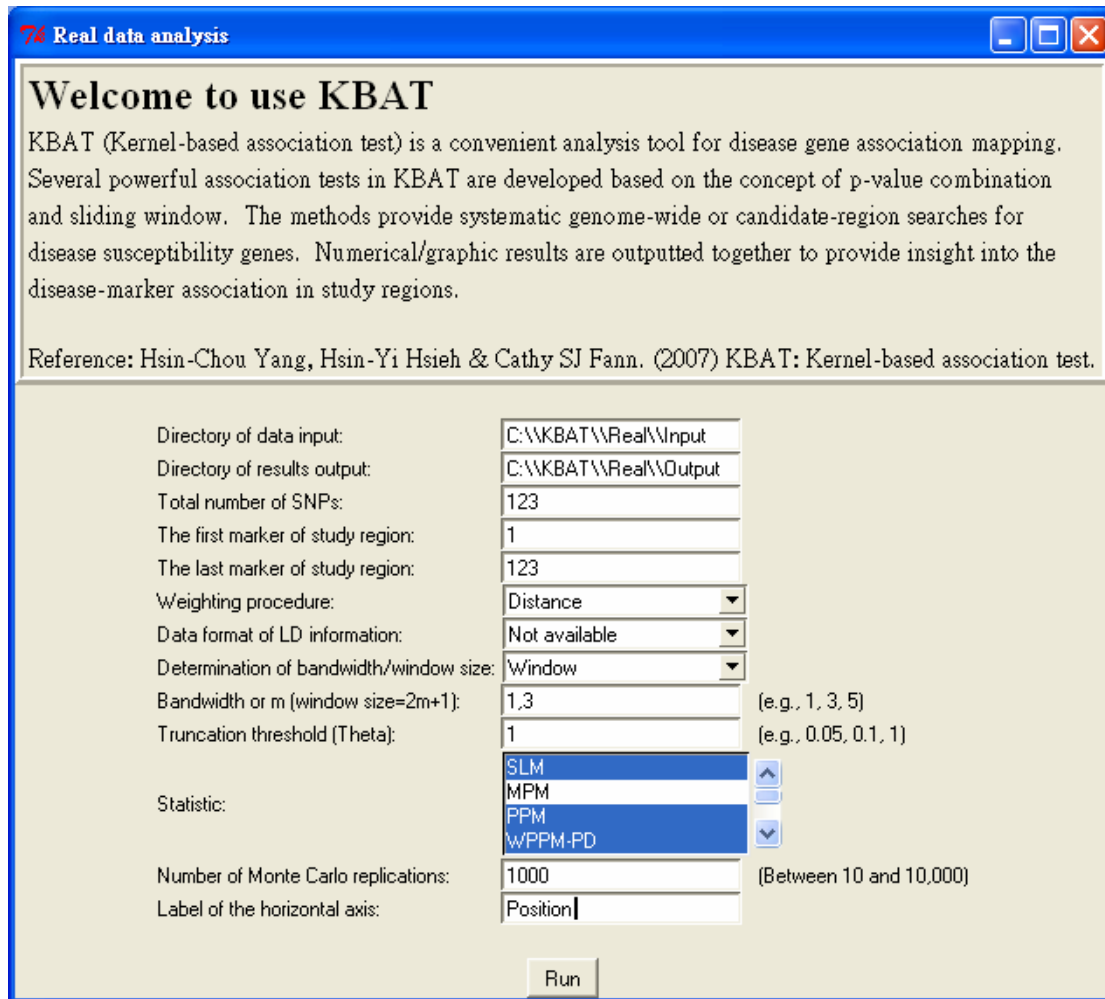


In this example, computation took ~2 minutes on a PC with an Intel P4 3-GHz processor and 2 GB RAM. Once finished, **KBAT** saves the numerical output in "Output.txt" (See **Table 1**) and the figure files (See **Figure 3**) in "C:\KBAT\Real\Output". In **Table 1**, empirical p-values of 123 SNP markers based on the four test statistics are shown as m=1 and m=3, in order. The table title is shown first, followed by the names of variables. The first column is the index of the SNP marker. The second to the fifth columns are empirical p-values of the four test statistics, SLM, PPM, WPPM-PD and KBAT-PD, respectively.

In **Figure 3**, empirical p-values are drawn. The vertical axis is the empirical p-values in a $-\log_{10}$ scale, and the horizontal axis represents the physical position. The titles of the subfigures indicate the test statistic and window size used.

**Table 1.** Numerical output of the psoriasis data analysis

Table: P-values of all statistics for each marker when m is 1 and the truncation threshold is 1

| Marker | SLM | PPM | WPPM-PD | KBAT-PD |
|--------|-----|-----|---------|---------|
| 1 | 1.000000e+00 | 8.030000e-01 | 1.000000e+00 | 8.370000e-01 |
| 2 | 5.153000e-01 | 8.230000e-01 | 5.400000e-01 | 8.130000e-01 |
| 3 | 6.685000e-01 | 6.940000e-01 | 6.440000e-01 | 7.150000e-01 |
| 4 | 6.510000e-01 | 4.970000e-01 | 6.470000e-01 | 4.810000e-01 |
| 5 | 1.809000e-01 | 1.640000e-01 | 2.050000e-01 | 1.470000e-01 |
| 6 | 5.780000e-02 | 2.060000e-01 | 6.500000e-02 | 1.330000e-01 |
| 7 | 1.000000e+00 | 3.740000e-01 | 1.000000e+00 | 6.190000e-01 |
| 8 | 7.457000e-01 | 6.890000e-01 | 7.640000e-01 | 6.520000e-01 |
| 9 | 2.511000e-01 | 4.520000e-01 | 2.430000e-01 | 4.440000e-01 |
| 10 | 3.488000e-01 | 4.740000e-01 | 3.400000e-01 | 3.970000e-01 |
| 11 | 7.760000e-01 | 4.210000e-01 | 7.540000e-01 | 4.690000e-01 |
| 12 | 2.110000e-01 | 2.560000e-01 | 2.240000e-01 | 3.640000e-01 |
| 13 | 9.740000e-02 | 1.980000e-01 | 1.050000e-01 | 2.010000e-01 |
| 14 | 4.564000e-01 | 2.990000e-01 | 4.260000e-01 | 2.560000e-01 |
| 15 | 5.561000e-01 | 7.600000e-01 | 5.480000e-01 | 7.900000e-01 |
| 16 | 1.000000e+00 | 8.450000e-01 | 1.000000e+00 | 8.600000e-01 |
| 17 | 6.723000e-01 | 6.650000e-01 | 6.700000e-01 | 7.250000e-01 |
| 18 | 2.607000e-01 | 4.350000e-01 | 2.570000e-01 | 3.860000e-01 |
| 19 | 3.438000e-01 | 4.880000e-01 | 3.310000e-01 | 4.940000e-01 |
| 20 | 9.557000e-01 | 7.970000e-01 | 9.450000e-01 | 8.520000e-01 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 116 | 6.339000e-01 | 7.270000e-01 | 6.420000e-01 | 7.840000e-01 |
| 117 | 9.656000e-01 | 7.800000e-01 | 9.700000e-01 | 8.570000e-01 |
| 118 | 4.732000e-01 | 8.800000e-01 | 4.610000e-01 | 7.580000e-01 |
| 119 | 1.000000e+00 | 7.520000e-01 | 1.000000e+00 | 8.150000e-01 |
| 120 | 5.403000e-01 | 7.700000e-01 | 5.350000e-01 | 6.800000e-01 |
| 121 | 5.220000e-01 | 5.930000e-01 | 5.270000e-01 | 5.920000e-01 |
| 122 | 4.597000e-01 | 6.870000e-01 | 5.000000e-01 | 5.860000e-01 |
| 123 | 7.069000e-01 | 6.180000e-01 | 6.700000e-01 | 6.790000e-01 |

Table: P-values of all statistics for each marker when m is 3 and the truncation threshold is 1

| Marker | SLM | PPM | WPPM-PD | KBAT-PD |
|--------|-----|-----|---------|---------|
| 1 | 1.000000e+00 | 8.370000e-01 | 1.000000e+00 | 8.450000e-01 |
| 2 | 5.153000e-01 | 6.930000e-01 | 5.280000e-01 | 8.420000e-01 |
| 3 | 6.685000e-01 | 4.160000e-01 | 6.600000e-01 | 3.000000e-01 |
| 4 | 6.510000e-01 | 4.790000e-01 | 6.490000e-01 | 3.640000e-01 |
| 5 | 1.809000e-01 | 4.790000e-01 | 1.820000e-01 | 4.520000e-01 |
| 6 | 5.780000e-02 | 4.320000e-01 | 7.300000e-02 | 4.230000e-01 |
| 7 | 1.000000e+00 | 3.410000e-01 | 1.000000e+00 | 3.240000e-01 |
| 8 | 7.457000e-01 | 3.510000e-01 | 7.710000e-01 | 3.760000e-01 |
| 9 | 2.511000e-01 | 4.110000e-01 | 2.600000e-01 | 5.510000e-01 |
| 10 | 3.488000e-01 | 4.390000e-01 | 3.300000e-01 | 4.190000e-01 |
| 11 | 7.760000e-01 | 3.380000e-01 | 7.460000e-01 | 3.550000e-01 |
| 12 | 2.110000e-01 | 3.350000e-01 | 2.120000e-01 | 3.090000e-01 |
| 13 | 9.740000e-02 | 4.460000e-01 | 1.050000e-01 | 3.240000e-01 |
| 14 | 4.564000e-01 | 5.390000e-01 | 4.560000e-01 | 4.400000e-01 |
| 15 | 5.561000e-01 | 4.040000e-01 | 5.560000e-01 | 5.060000e-01 |

| 16 | 1.000000e+00 | 4.880000e-01 | 1.000000e+00 | 5.700000e-01 |
|---|---|---|---|---|
| 17 | 6.723000e-01 | 7.060000e-01 | 6.940000e-01 | 7.100000e-01 |
| 18 | 2.607000e-01 | 8.060000e-01 | 2.590000e-01 | 7.790000e-01 |
| 19 | 3.438000e-01 | 8.430000e-01 | 3.600000e-01 | 7.320000e-01 |
| 20 | 9.557000e-01 | 7.780000e-01 | 9.560000e-01 | 7.610000e-01 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 116 | 6.339000e-01 | 8.930000e-01 | 6.230000e-01 | 8.760000e-01 |
| 117 | 9.656000e-01 | 8.630000e-01 | 9.670000e-01 | 8.520000e-01 |
| 118 | 4.732000e-01 | 7.950000e-01 | 4.670000e-01 | 7.800000e-01 |
| 119 | 1.000000e+00 | 8.200000e-01 | 1.000000e+00 | 7.750000e-01 |
| 120 | 5.403000e-01 | 8.430000e-01 | 5.610000e-01 | 7.760000e-01 |
| 121 | 5.220000e-01 | 7.600000e-01 | 5.190000e-01 | 7.520000e-01 |
| 122 | 4.597000e-01 | 7.950000e-01 | 4.470000e-01 | 7.060000e-01 |
| 123 | 7.069000e-01 | 6.670000e-01 | 7.020000e-01 | 6.900000e-01 |

------------------------------------------------------------------

**Figure 3.** Graphic output in the psoriasis data analysis. Results will be saved as PDF files based on bandwidths. (A) Results for m=1. (B) Results for m=3.

(A)
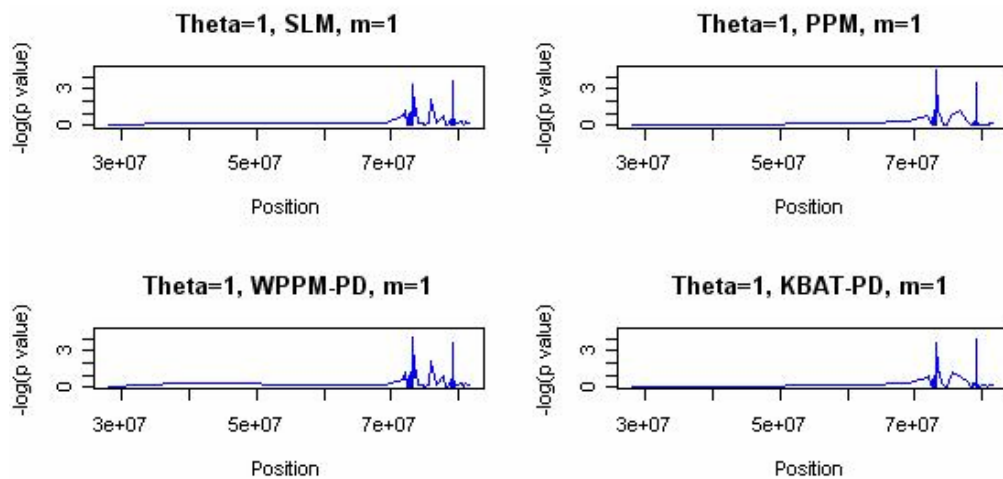


(B)
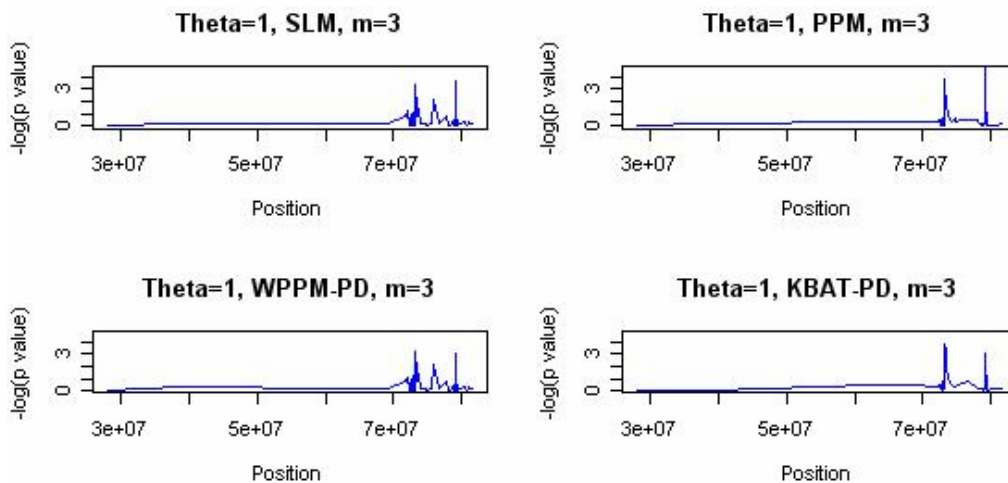


**Example 2: Simulation data analysis**

10

To illustrate the different types of **KBAT** data formats, we generated a data set that contained genotyped data of 31 SNP markers for 500 cases and 500 controls. Genotype data of 31 SNP markers were generated based on a disease model with penetrance PV=(0.1, 0.3, 0.4) for genotype (*dd*, *dD* and *DD*), where *D* was the disease allele. The intermarker recombination was set following a flat U recombination function (Yang et al., 2006). The true disease locus was arranged close to the $16^{th}$ SNP. All files of genotype data, map data, LD data, and p-value data are provided in the directory "C:\KBAT\Real\Example\Sim". Genotype data for the 31 SNP markers are in file "geno.txt". Map data for the 31 SNPs are file "map.txt". Intermarker LD data for the 31 SNPs are in file "ld.txt". Finally, p-value data for single-locus association tests are in file "pv.txt". Here, we reanalyze these data with **KBAT**.

We copied all of the files to the working directory "C:\KBAT\Real\Input". In the analysis, statistics WPPM-PD, WPPM-PDLD, KBAT-PD, and KBAT-PDLD were calculated under the two window sizes of 3 (m=1) and 11 (m=5). Truncation was not considered in the analysis. Calculation of the four statistics requires both LD and position information. Therefore, in addition to files "pv.txt" and "map.txt", LD information should be provided. Users should provide the LD file "ld.txt", and then **KBAT** can directly use the information to calculate p-value weights. In this case, "LD measure" in the item "Data format of LD information" should be selected. Or, users can provide the genotype file "geno.txt", and **KBAT** can help calculate LD. In this case, "Genotype data" in the item "Data format of LD information" should be selected. We illustrate the former situation using the following parameters (See **Figure 4**).

(1)  Directory of data input: "C:\\KBAT\\Real\\Input".

(2)  Directory of results output: "C:\\KBAT\\Real\\Output".

(3)  Total number of SNPs: "31".

(4)  The first marker of study region: "1".

(5)  The last marker of study region: "31".

(6)  Weighting procedure: "LD and/or distance".

(7)  Data format of LD information: "LD measure".

(8)  Determination of bandwidth/window size: "Window".

(9)  Bandwidth or m (window size = 2m+1): "1,5".

(10) Truncation threshold (Theta): "1".

(11) Statistic: "WPPM-PD", "WPPM-PDLD", "KBAT-PD", and "KBAT-PDLD".

(12) Number of Monte Carlo replications: "1000".

(13) Label of the horizontal axis: "Marker".

(14) Press "RUN" to execute **KBAT**.

**Figure 4.** Interface for the example of simulation data analysis
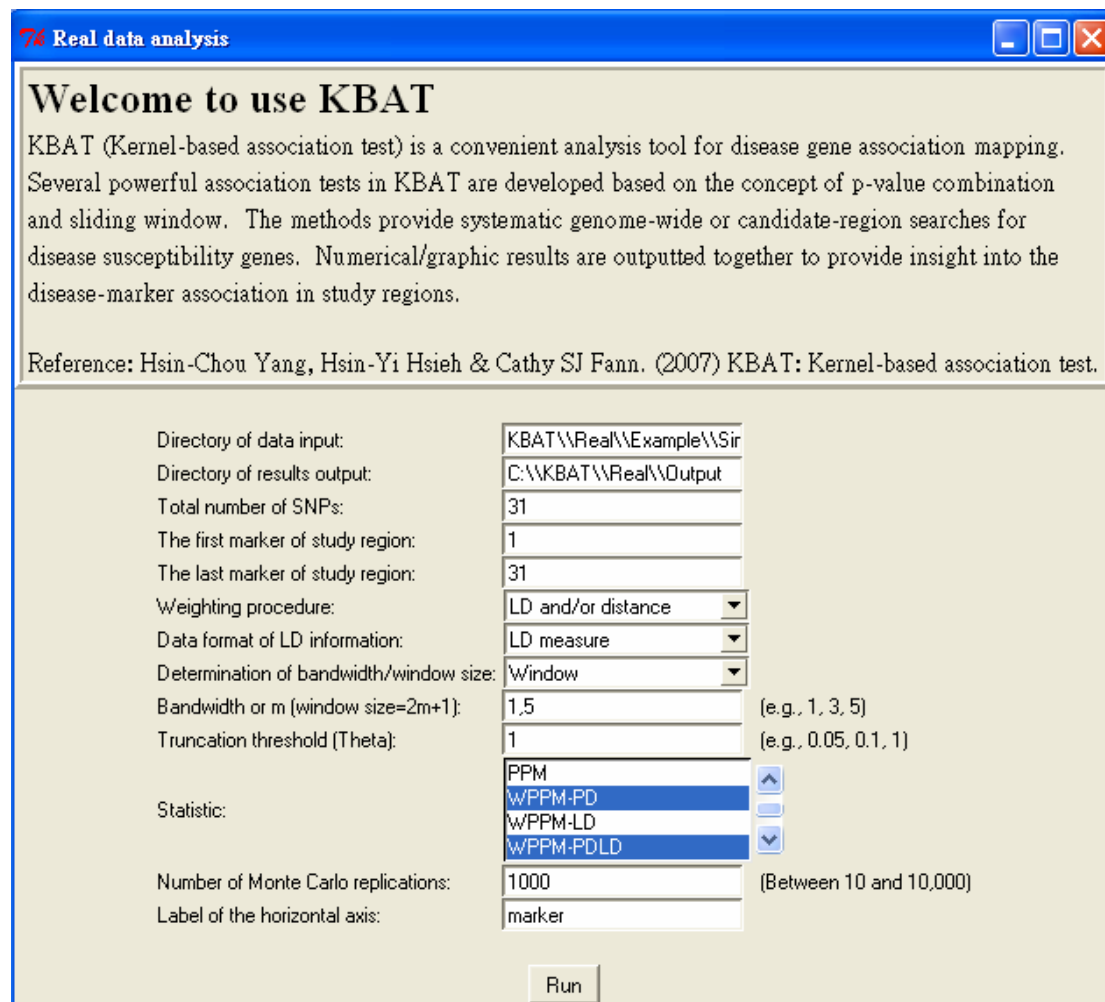


In this example, computation took ~2 minutes on a PC with an Intel P4 3-GHz processor and 2 GB RAM. Once finished, **KBAT** saves the numerical output "Output.txt" (See **Table 2**) and the figure file (See **Figure 5**) in "C:\KBAT\Real\Output". In **Table 2**, empirical p-values of 31 SNP markers based on the four test statistics are shown by different window sizes of m=1 and m=5, in order. The table title is shown first, followed by the names of variables. The first column is the index of the SNP marker. The second to the fifth columns are empirical p-values of the four test statistics, WPPM-PD, WPPM-PDLD, KBAT-PD, and KBAT-PDLD respectively.

In **Figure 5**, empirical p-values are drawn. The vertical axis is the empirical

p-values, and the horizontal axis represents the physical position. The titles of the subfigures indicate the test statistic and window size used.

**Table 2.** Numerical output in the simulation data analysis

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Table: P-values of all statistics for each marker when m is 1 and the truncation threshold is 1

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

| Marker | WPPM-PD | WPPM-PDLD | KBAT-PD | KBAT-PDLD |
|--------|---------|-----------|---------|-----------|
| 1 | 7.240000e-01 | 7.210000e-01 | 6.560000e-01 | 6.520000e-01 |
| 2 | 4.250000e-01 | 7.270000e-01 | 4.250000e-01 | 7.270000e-01 |
| 3 | 4.870000e-01 | 2.360000e-01 | 3.140000e-01 | 1.690000e-01 |
| 4 | 5.360000e-01 | 6.210000e-01 | 5.940000e-01 | 6.630000e-01 |
| 5 | 9.650000e-01 | 9.650000e-01 | 9.790000e-01 | 9.790000e-01 |
| 6 | 9.990000e-01 | 9.870000e-01 | 9.950000e-01 | 9.760000e-01 |
| 7 | 9.430000e-01 | 9.940000e-01 | 9.580000e-01 | 9.950000e-01 |
| 8 | 5.840000e-01 | 6.030000e-01 | 6.080000e-01 | 5.880000e-01 |
| 9 | 1.470000e-01 | 1.480000e-01 | 1.450000e-01 | 1.460000e-01 |
| 10 | 1.570000e-01 | 1.050000e-01 | 1.220000e-01 | 1.020000e-01 |
| 11 | 2.640000e-01 | 5.120000e-01 | 3.210000e-01 | 5.130000e-01 |
| 12 | 6.350000e-01 | 5.850000e-01 | 5.850000e-01 | 5.350000e-01 |
| 13 | 7.810000e-01 | 7.790000e-01 | 7.480000e-01 | 7.440000e-01 |
| 14 | 8.450000e-01 | 8.470000e-01 | 8.450000e-01 | 8.470000e-01 |
| 15 | 1.900000e-02 | 7.380000e-01 | 1.900000e-02 | 7.380000e-01 |
| 16 | 1.000000e-03 | 2.381095e-04 | 1.000000e-03 | 2.348265e-04 |
| 17 | 1.945080e-04 | 4.316122e-04 | 3.251567e-04 | 7.780396e-05 |
| 18 | 4.540000e-01 | 5.470000e-01 | 4.540000e-01 | 5.470000e-01 |
| 19 | 2.170000e-01 | 1.250000e-01 | 2.390000e-01 | 1.760000e-01 |
| 20 | 1.900000e-01 | 1.900000e-01 | 1.360000e-01 | 1.350000e-01 |
| 21 | 3.200000e-02 | 3.200000e-02 | 5.900000e-02 | 6.000000e-02 |
| 22 | 1.070000e-01 | 1.060000e-01 | 3.900000e-02 | 3.800000e-02 |
| 23 | 6.200000e-02 | 5.800000e-02 | 1.350000e-01 | 1.410000e-01 |
| 24 | 1.470000e-01 | 1.310000e-01 | 1.420000e-01 | 1.340000e-01 |
| 25 | 1.670000e-01 | 1.640000e-01 | 1.110000e-01 | 1.090000e-01 |
| 26 | 2.590000e-01 | 2.630000e-01 | 3.840000e-01 | 3.900000e-01 |
| 27 | 9.100000e-01 | 8.990000e-01 | 8.130000e-01 | 7.910000e-01 |
| 28 | 8.290000e-01 | 7.780000e-01 | 8.830000e-01 | 8.450000e-01 |
| 29 | 9.080000e-01 | 7.660000e-01 | 8.880000e-01 | 7.270000e-01 |
| 30 | 7.450000e-01 | 7.040000e-01 | 7.450000e-01 | 7.040000e-01 |
| 31 | 8.650000e-01 | 8.650000e-01 | 8.650000e-01 | 8.650000e-01 |

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Table: P-values of all statistics for each marker when m is 5 and the truncation threshold is 1

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

| Marker | WPPM-PD | WPPM-PDLD | KBAT-PD | KBAT-PDLD |
|--------|---------|-----------|---------|-----------|
| 1 | 0.8220000000 | 0.8300000000 | 0.6680000000 | 0.6840000000 |
| 2 | 0.9140000000 | 0.9880000000 | 0.7640000000 | 0.9580000000 |
| 3 | 0.9310000000 | 0.5260000000 | 0.8650000000 | 0.4970000000 |
| 4 | 0.8190000000 | 0.9360000000 | 0.8950000000 | 0.9460000000 |
| 5 | 0.6400000000 | 0.6800000000 | 0.8580000000 | 0.8810000000 |
| 6 | 0.6860000000 | 0.9570000000 | 0.6860000000 | 0.9570000000 |
| 7 | 0.6950000000 | 0.9850000000 | 0.7220000000 | 0.9860000000 |
| 8 | 0.6310000000 | 0.7600000000 | 0.6670000000 | 0.7000000000 |
| 9 | 0.8510000000 | 0.7360000000 | 0.6260000000 | 0.4930000000 |
| 10 | 0.8030000000 | 0.5110000000 | 0.8030000000 | 0.5110000000 |
| 11 | 0.7000000000 | 0.8450000000 | 0.7000000000 | 0.8450000000 |
| 12 | 0.6140000000 | 0.7340000000 | 0.1270000000 | 0.4180000000 |
| 13 | 0.4530000000 | 0.4530000000 | 0.0380000000 | 0.2580000000 |
| 14 | 0.4050000000 | 0.8320000000 | 0.0320000000 | 0.4920000000 |
| 15 | 0.0050000000 | 0.0610000000 | 0.0050000000 | 0.0610000000 |
| 16 | 0.0010000000 | 0.0008103583 | 0.0010000000 | 0.0005896656 |
| 17 | 0.0010000000 | 0.0020000000 | 0.0010000000 | 0.0020000000 |
| 18 | 0.0830000000 | 0.0810000000 | 0.0070000000 | 0.0440000000 |
| 19 | 0.0690000000 | 0.0190000000 | 0.0020000000 | 0.0040000000 |
| 20 | 0.0350000000 | 0.0320000000 | 0.0010000000 | 0.0250000000 |
| 21 | 0.0720000000 | 0.0690000000 | 0.0720000000 | 0.0690000000 |
| 22 | 0.0690000000 | 0.0680000000 | 0.0690000000 | 0.0680000000 |
| 23 | 0.1050000000 | 0.0820000000 | 0.0410000000 | 0.0410000000 |
| 24 | 0.1370000000 | 0.1040000000 | 0.0790000000 | 0.0640000000 |
| 25 | 0.1590000000 | 0.0640000000 | 0.1780000000 | 0.0990000000 |
| 26 | 0.2500000000 | 0.2100000000 | 0.2500000000 | 0.2100000000 |
| 27 | 0.2410000000 | 0.2330000000 | 0.2410000000 | 0.2330000000 |
| 28 | 0.6340000000 | 0.8010000000 | 0.6340000000 | 0.8010000000 |
| 29 | 0.5080000000 | 0.3450000000 | 0.5080000000 | 0.3450000000 |
| 30 | 0.6450000000 | 0.7590000000 | 0.6450000000 | 0.7590000000 |
| 31 | 0.8770000000 | 0.8770000000 | 0.8770000000 | 0.8770000000 |

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**Figure 5.** Graphic output of the simulation data analysis. The results will be saved as PDF files based on the bandwidths. (A) Result for m=1. (B) Results for m=5.

# 9. REFERENCES

1. Helms C, Cao L, Krueger JG, Wijsman EM, Chamian F, Gordon D, Heffernan M, Daw JAW, Robarge J, Ott J, Kwok PY, Menter A, Bowcock AM. 2003. A putative RUNX1 binding site variant between *SLC9A3R1* and *NAT9* is associated with susceptibility to psoriasis. *Nature Genetics* **35**: 349-256.

2. Yang HC, Lin CY, Fann CSJ. 2006. A sliding-window weighted linkage disequilibrium test. *Genetic Epidemiology* **30**: 531-545.

# 10. APPENDIX — TEST STATISTICS

- **Single locus method (SLM):**

$$Q_{i,m} = p_i, \forall i = 1, \cdots, N.$$

- **Minimum p-value method (MPM):**

$$Q_{i,m} = \min_{j \in \Im(i,m)} \{p_j\}, \forall i = 1, \cdots, N.$$

- **Product p-value method (PPM):**

$$Q_{i,m} = \sum_{j \in \Im(i,m)} \ln(p_j) I[p_j < \tau], \ \forall i = 1, \cdots, N.$$

- **Distance-weight product p-value method (WPPM-PD):**

$$Q_{i,m} = \sum_{j \in \Im(i,m)} w_j(i,m) \ln(p_j) I[p_j < \tau], \ \forall i = 1, \cdots, N, \ \ w_j(i,m) = h_{i,j}^*, \text{ where}$$

$$h_{i,j}^* = h_{i,j} \Big/ \sum_{k \in \Im(i,m)} h_{i,k} \ \text{ and } \ h_{i,j} = 1/(1 + d_{i,j}).$$

- **LD-weight product p-value method (WPPM-LD):**

$$Q_{i,m} = \sum_{j \in \Im(i,m)} w_j(i,m) \ln(p_j) I[p_j < \tau], \ \forall i = 1, \cdots, N, \ \ w_j(i,m) = \hat{\rho}_{i,j}^*, \text{ where}$$

$$\hat{\rho}_{i,j}^* = \hat{\rho}_{i,j} \Big/ \sum_{k \in \Im(i,m)} \hat{\rho}_{i,k} \ \text{ and } \ \hat{\rho}_{i,j} = \left[\hat{\lambda}_{11}(i,j)\hat{\lambda}_{22}(i,j) - \hat{\lambda}_{12}(i,j)\hat{\lambda}_{21}(i,j)\right] \Big/ \left[\hat{\lambda}_{1+}(i,j)\hat{\lambda}_{+2}(i,j)\right].$$

- **Distance-LD-weight product p-value method (WPPM-PDLD):**

$$Q_{i,m} = \sum_{j \in \Im(i,m)} w_j(i,m) \ln(p_j) I[p_j < \tau], \ \forall i = 1, \cdots, N, \ \ w_j(i,m) = \frac{h_{i,j}^* \times \hat{\rho}_{i,j}^*}{\sum_{k \in \Im(i,m)} h_{i,k}^* \times \hat{\rho}_{i,k}^*},$$

$$\text{where } \ h_{i,j}^* = h_{i,j} \Big/ \sum_{k \in \Im(i,m)} h_{i,k} \ \text{ and } \ \hat{\rho}_{i,j}^* = \hat{\rho}_{i,j} \Big/ \sum_{k \in \Im(i,m)} \hat{\rho}_{i,k}.$$

- **Kernel-based association test – physical distance (KBAT-PD):**

$$G_{t,h} = \sum_{i \in W(t,h)} \left( a_i \times \ln(p_i) \right), \text{ where } a_i = \frac{\left( \mathrm{K}((t - t_i)/h) \right)}{\sum_{j \in W(t,h)} \left( \mathrm{K}((t - t_j)/h) \right)}.$$

- **Kernel-based association test – distance-LD (KBAT-PDLD):**

$$G_{t,h} = \sum_{i \in W(t,h)} \left( a_{i,j} \times \ln(p_i) \right), \text{ where } a_{i,j} = \frac{\hat{h}_i^*(t) \times \hat{\rho}_{i,j}^*}{\sum_{j \in W(t,h)} \hat{h}_i^*(t) \times \hat{\rho}_{i,j}^*},$$

$$\hat{h}_i^*(t) = \frac{\left( \mathrm{K}((t_j - t_i)/h) \right)}{\sum_{k \in W(t,h)} \left( \mathrm{K}((t_k - t_i)/h) \right)} \text{ and } \hat{\rho}_{i,j}^* = \hat{\rho}_{i,j} \Big/ \sum_{k \in \mathfrak{I}(i,m)} \hat{\rho}_{i,k}.$$