Version 2.3 (Mar, 2019)

User guide of software LOHAS

Hsin-Chou Yang, Lun-Ching Chang, Han-Wei Li, Yu-Ting Lin, and Chia-Wei Chen

Institute of Statistical Science, Academia Sinica

Correspondence: Hsin-Chou Yang (hsinchou@stat.sinica.edu.tw), Institute of Statistical Science, Academia Sinica, 128, Academia Road, Section 2 Nankang, Taipei 115, Taiwan.

Table of Contents:

1. LOHAS LICENSE

All copyright are reserved by authors of **LOHAS**. We welcome any noncommercial use of **LOHAS** for your own research. Please do NOT modify or distribute the program of **LOHAS** in any form without the permission of authors of **LOHAS**. Commercial use of **LOHAS** should be directed to hsinchou@stat.sinica.edu.tw. For free software **LOHAS**, we assume no warranty and no responsibility for the results of analyses. If the work is published based on the results from the use of **LOHAS**, please cite the following references:

Yang, H.-C., Chang, L.-C., Huggins, R. M., Chen, C.-H. and Mullighan, C. G. (2011/05). LOHAS: Loss-of-heterozygosity analysis suite. *Genetic Epidemiology* **35**, 247-260.

Yang, H.-C. and Li, H.-W. (2014/06). Analysis of homozygosity disequilibrium using whole-genome sequencing data. *BMC Proceedings* **8**, S15.

Yang, H.-C. and Lin, Y.-T. (2016/11). Homozygosity disequilibrium and its gene regulation. *BMC Proceedings* **10**, 27. (http://rdcu.be/mHtp)

Yang, H.-C. and Chen, C.-W. (2018/09). Homozygosity disequilibrium associated with treatment response and its methylation regulation. *BMC Proceedings* **12**, 45. (https://rdcu.be/6Z4m)

2. INTRODUCTION

LOHAS (Loss-Of-Heterozygosity Analysis Suite), written in R and R GUI, was developed for studies of homozygosity disequilibrium (HD). HD describes a phenomenon in which a non-random pattern is observed for a sizable run of homozygosity (ROH) in the human genome. HD can result from autozygosity, natural selection, and chromosomal aberrations. **LOHAS** can be used to identify ROHs associated with complex disorders, detect loss of heterozygosity (LOH) in cancer research, and characterize long contiguous stretches of homozygosity (LCSH) in population genetics studies using genotype data. LOHAS can estimate homozygosity intensity, identify samples with unusual genomes, cluster samples with close HD structures, and map the genomic

segments bearing HD by analyzing whole-genome SNP data and sequencing data.

LOHAS has been successfully applied to analyze genome-wide genotype data from the following large-scale genomic studies:

- A study of general populations (Yang et al., *Genetic Epidemiology*, 2011) – Data from the HapMap Project and the 1000 Genomes Project;
- A study of cancer research (Yang et al., *Genetic Epidemiology*, 2011) – Data from an acute lymphoblastic leukemia study;
- A study of a complex disease (Yang et al., *PLoS One*, 2012) – Data from the Welcome Trust Case-Control Consortium and the North American Rheumatoid Arthritis Consortium;
- A next-generation sequencing and gene regulation study (Yang and Li, *BMC Proceedings*, 2014; Yang and Lin, *BMC Proceedings*, 2016) – Data from the T2D-GENES Project;
- A pharmacogenomics and pharmacoepigenomics study (Yang and Chen, *BMC Proceedings*, 2018) – Data from the GOLDN Project.

3. SOFTWARE DOWNLOAD AND INSTALLATION

Execution of **LOHAS** requires the installation of **LOHAS** program and R program. For calculation acceleration, Windows users should install Rtools (https://cran.r-project.org/bin/windows/Rtools/). Procedures for downloading and installing the two programs are described as follows:

1. **LOHAS**:

   LOHAS program is available at the **LOHAS** website at http://www.stat.sinica.edu.tw/hsinchou/genetics/loh/LOHAS.htm. The zipped file “LOHAS_20190307.rar” can be downloaded and then unzipped to obtain a directory “LOHAS” containing the program codes of **LOHAS** and five illustrative examples.

2. R:

   Users can download R program “R-3.1.2-win.exe” from the **LOHAS** website. Or users can download latest R from the website of “The R Project for

Statistical Computing" at http://www.r-project.org/. Users click "CRAN" (Comprehensive R Archive Network) in the left of the page and then select a suitable mirror site to download R. Select a platform (Linux, MacOS X, or Windows) for R execution in your end. Click the hyperlink "base" and select "R-3.1.2-win.exe". Then execute the file to install R to "C:\Program Files\R\R-3.1.2". After finishing the installation of R, doubly click the icon "R i386 3.1.2" or "R x64 3.1.2" to initialize R in a 32-bit or 64-bit system, respectively. A window "RGui" with a sub-window "R Console" jumps up await for the subsequent analysis action. Users are suggested to update packages in R. They can select "Packages" in the tool bar, click "Update packages" and then select a suitable mirror site to update packages. A window "CRAN mirror" jumps up and the icon "OK" is clicked to update packages. Note that the analyses provided by **LOHAS** require a number of additional R packages, e.g., locfit, tcltk, tcltk2, and geeM. These packages will be automatically downloaded if users use a latest version of R. **Note that users are suggested to use program R-3.1.2 or a version of R program newer than program R-2.15.0 for execution of LOHAS.**

3. Rtools:

   Windows users can download Rtools from the website of "Building R for Windows" at https://cran.r-project.org/bin/windows/Rtools/. According to the R version installed, users should choose the corresponding Rtools version to download (e.g., Rtools32.exe is required for R-3.1.0 and later). Then execute the file to install Rtools to "C:\Rtools".

4. LOHAS INITIALIZATION

Once the directory "LOHAS" mentioned in the previous section has been copied to a working directory (e.g., "C:\" is set as a working directory in this user guide), **LOHAS** can be initialized by the following procedures.

1. Initialize software R by doubly clicking the icon "R-3.1.2".
2. Key in the command, **LOHAS.gui=paste("C:/LOHAS/Program/LOHAS_interface.r",sep="")**, in the command line in the window "R Console" and press the Enter key.

3. Type the command, **source(LOHAS.gui)**, in the command line and press the Enter key to initialize **LOHAS**. The **LOHAS** interface (see **Figure 4-1**) jumps up and waits for the data entry after pressing the Enter key. The above-mentioned commands are also provided in file “LOHAS_path.txt” in the directory “C:/LOHAS/Program/”.

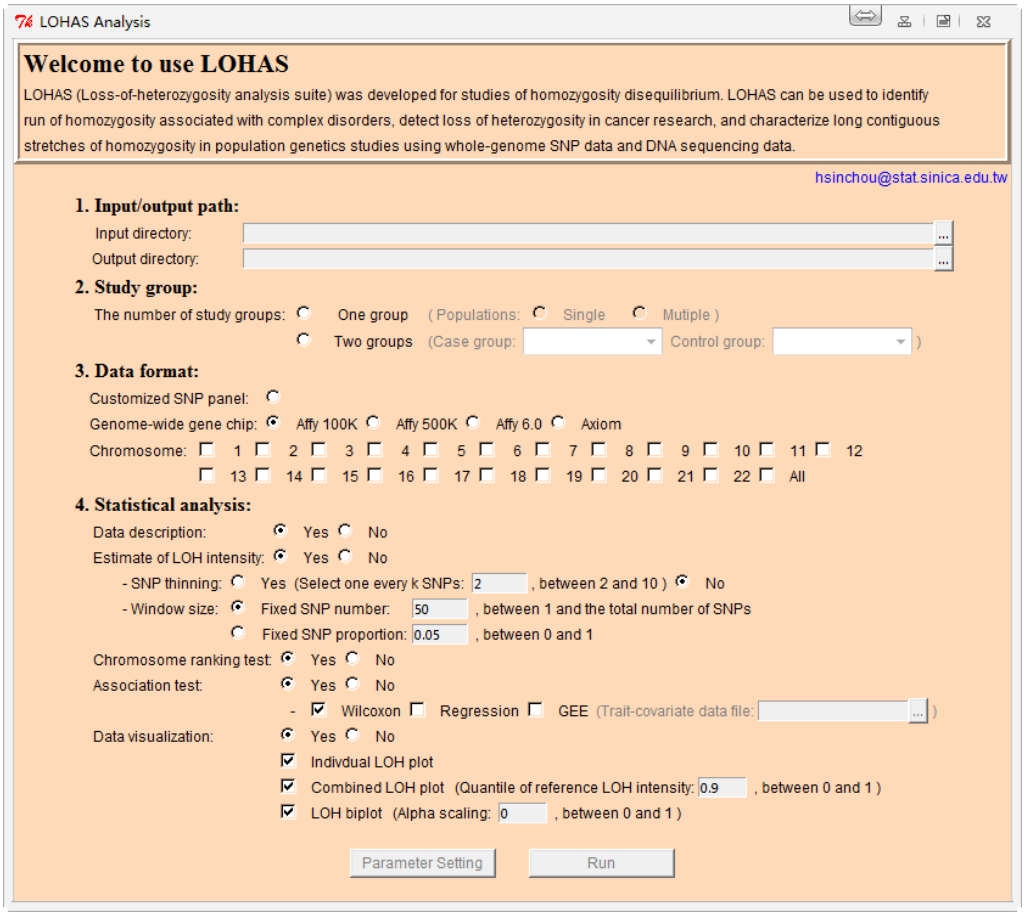


Figure 4-1. Interface of LOHAS.

5. LOHAS INTERFACE AND FUNCTIONS

LOHAS has a user-friendly interface developed by R GUI (**Figure 4-1**). The interface contains a preface for a short introduction of **LOHAS**. Directory structure of **LOHAS** is shown (**Figure 5-1**).

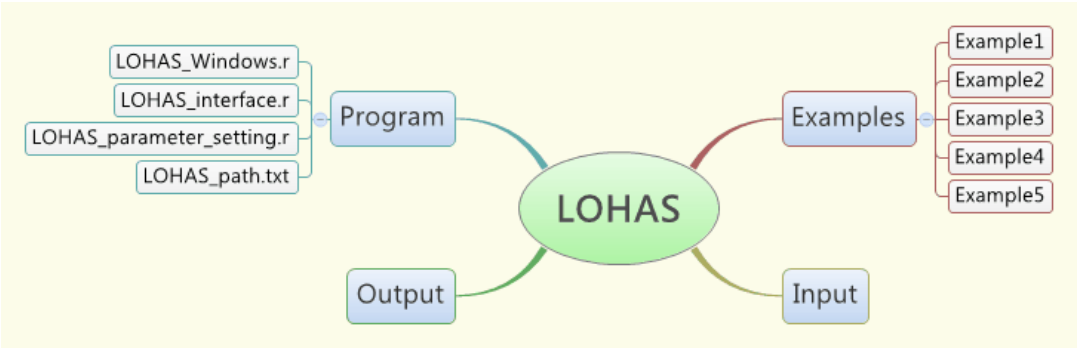


Figure 5-1. Directory structure of LOHAS (Windows System).

Five main item questions are designed for providing required/optional information for **LOHAS** data analysis.

Item 1: Input/output path:

- Input directory: Users should provide the working directory path where the populations/groups data folder exists.
- Output directory: Users should provide the working directory path where their output files will be saved.

Item 2: Study group:

- The number of study groups: Users should select “one group” for a one-population or multiple-population analysis or select “two groups” for a case-control analysis. Moreover, users should specify the first or second group as a case group.

Item 3: Parameter setting:

After the previous two settings, users can click icon “Parameter Setting” in the bottom of the **LOHAS** interface to input parameters (**Figure 5-2**). The format (columns) of genotype data files includes ID of SNP, ID of chromosome where the SNP is located, physical position of SNP (unit: base pair), genotype AB call, RS number, a threshold of minor allele frequency (MAF) for defining a variant as a rare variant, use of a locus weight in a local polynomial model, and the starting and ending characters used for a nomenclature of output files of study individuals:

- “Column index of SNP ID” (**required**): The column index of SNP ID in a genotype data file. The information must be provided.
- “Column index of chromosome” (**required**): The column index of chromosome ID in a genotype data file. The information must be

provided.

● “Column index of physical position” (**required**): The column index of physical position of SNP in a genotype data file. The information must be provided.

● “Column index of ABcall” (**required**): The column index of genotype AB call in a genotype data file. The information must be provided.

● “Column index of RS number” (**optional**): The column index of RS number of SNP in a genotype data file. The information could be provided.

● “MAF threshold” (**optional**): The threshold of MAF for defining a variant as a rare variant in a genotype data file. If MAF of a variant is smaller than the threshold (e.g., 0.05) then the variant is treated as a rare variant. MAF threshold is required if “Yes” is selected in “MAF weight” (see the next setting “MAF weight”). **Note that this setting is designed for an analysis of DNA sequencing data and therefore it is only relevant for the analysis of data from “customized SNP panel” (see Item 3 in Section 5). In an analysis of SNP gene chips, this threshold will be ignored.**

● “MAF weight” (**optional**): There are two choices, “Yes” or “No”. “Yes” indicates that a locus weight that is generated by **LOHAS** will be incorporated into a local polynomial model (Yang and Lin, *BMC Proceedings*, 2016). “No” indicates that no locus weight will be considered in the analysis. It is validated after “MAF threshold” is set. **Similar to “MAF_threshold”, this setting is designed for an analysis of DNA sequencing data and therefore it is only relevant for the analysis of data from “customized SNP panel” (see Item 3 in Section 5). In an analysis of SNP gene chips, this setting will be ignored.**

● “Sample ID start field” (**optional**): The inputted number indicates that **LOHAS** will automatically start to extract character for a LOHAS-output filename nomenclature of study individuals from that character position of the genotype filenames. The information could be provided.

● “Sample ID end field” (**optional**): The inputted number indicates that **LOHAS** will stop to extract character for a LOHAS-output filename

nomenclature of study individuals from that character position of the genotype filenames. The information could be provided

After setting up the abovementioned parameters in the first panel in **Figure 5-2**, users should select populations/groups listed in the second panel (i.e., the "Population/Group" panel that provides a list of folder names under the input directory. Then **click >> to confirm the current settings for each population/group in the third panel (i.e., the "Group Setting Summary" panel).** A window jumps up to confirm the setting or reminds a need for a further setting. Finally, click the SAVE bottom to save the settings.

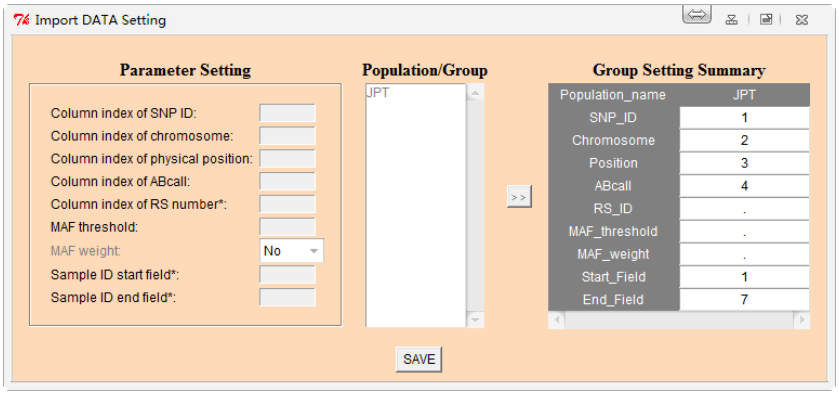


Figure 5-2. Interface for parameter setting.

Item 4: Data format:

- Customized SNP panel – Users should choose this format if their genotype data were NOT generated from Affymetrix 100K, 500K, and Array6.0.
- Genome-wide gene chip – When users choose this format they should specify the SNP gene chips of their data: Affymetrix 100K, 500K or Array6.0. If Affymetrix 100K or 500K is chosen, since version 2.3, **LOHAS** will run a merge-chip analysis if data of two chips are provided.
- Chromosome: Users should specify the chromosome(s) of interest in their analysis. Users can check "All" for an analysis of all chromosomes. It includes an analysis of sex chromosome if data for sex chromosome is provided.

Item 5: Statistical analysis:

- Data description – Check “Yes” for requesting an output of data description in the analysis.
- Estimate of LOH/LCSH intensity – Check “Yes” for requesting an estimation of homozygosity intensity.
  - ✧SNP thinning: Check “Yes” for requesting a SNP thinning procedure. A number (between 2 and 10) should be inputted. For example, if a number “2” is inputted then one SNP is chosen per two consecutive SNPs in the analysis. This option can save computational time if the number of variants is too large.
  - ✧Window size: Users can consider a fixed number of SNPs (between 1 and the total number of SNPs) or a fixed proportion of SNPs on a chromosome (between 0 and 1) for each sliding window. **Note that a rule of thumb for the number of SNPs in a window is at least 30 for an estimation of regression coefficients in a local polynomial regression.** An over-small number of SNPs in a window could result in an acute oscillation of the estimated homozygosity intensities.
- Chromosome ranking test – Check “Yes” to perform a chromosome ranking test.
- Association test – Check “Yes” for performing a homozygosity association test. Three methods are provided: (1) Wilcoxon non-parametric test, (2) regression model, and (3) generalized estimating equation (GEE) model. If users choose to run a regression or GEE model, they **MUST** provide a trait-covariate file (see **Section 6.4**), else an error message will be shown.
- Data visualization – Check “Yes” for executing a data visualization analysis. Three types of graphs are provided: (1) individual homozygosity intensity plot, (2) combined homozygosity intensity plot (users should provide a quantile threshold of reference homozygosity intensity; the value ranges between 0 and 1), and (3) homozygosity intensity biplot (users should provide an alpha scaling coefficient; the value ranges between 0 and 1).

Once all item questions are answered, icon “Run” can be clicked to submit

computational job. Then, **LOHAS** will check the inputted data information and data files. If the inputted information is invalid or the data files are ill-format, **LOHAS** will show warning message(s) or error message(s), which provides users to make corrections. If the inputted data pass the examination, **LOHAS** starts to perform analysis and a message "Please wait a while, **LOHAS** 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 **LOHAS** is finished." jumps up to acknowledge users the completion of **LOHAS** computation. **Note that users can interrupt the execution of LOHAS anytime by clicking ESC in the window "R Console".** Once the execution of **LOHAS** is finished, the numerical results and graphical outputs will be automatically saved in the output directory that users pre-specify. We suggest that users should remove figure files from a previous analysis before a new analysis in case of the confusion of multiple figure files from old and new analyses.

6. DATA INPUT FORMAT

This section introduces the input format of two data files (genotype file and trait-covariate file).

6.1 Genotype directories and files

Genotype files of all samples in the same population or group should be saved in the respective genotype data directories according to the pre-specified input path (see Item 1 in **Section 5.1**). Genotype data of each sample should be saved in a single genotype file (Except for the cases of Affymetrix 100K and Affymetrix 500K, genotype data of each sample should be saved in two files by chip; see below). Each genotype file contains at least four columns with header "SNP_ID", "Chromosome", "Physical_position", and "ABcall". An additional column "RS_number" is optional. These columns should be arranged in compliance with the setting the parameter setting (see Item 3 in **Section 5** and **Figure 5-2**). **Note that the genotype file must be a tab-delimited format and SNPs have to sort according to physical**

position of SNPs by chromosome in each genotype file.

- *Customized SNP panel:*

  Please refer to Example 4 for illustration (see **Section 7.4**).

- *Affymetrix 100K, 500K and Array6.0:*

  Affymetrix 100K and 500K chips contain two chips. Users **MUST** add chip types in file names in order to specify data files for different chips. For example, for an individual Abnorm_01 who is a female and genotyped with Affymetrix 100K, then the filenames of genotype data are "Abnorm_01_F_Hind_Genotype.txt" and "Abnorm_01_F_Xba_Genotype.txt" for Hind and Xba chips, respectively. The filename nomenclature rule is applied to Nsp and Sty chips for Affymetrix 500K similarly. Please refer to Example 3 for illustration (see **Section 7.3**). For Array6.0, there is only one chip in this platform and it is not necessary to make additional subdirectories. Please refer to Example 1 and Example 2 for illustration (see **Section 7.1** and **Section 7.2**).

6.2 Trait-covariate data file

LOHAS provides three types of homozygosity association tests (Wilcoxon, linear regression, and GEE). Wilcoxon test examines an equality of medians of homozygosity intensity of two independent groups. If users choose a homozygosity association analysis of a regression model or GEE model, a quantitative trait or binary trait can be analyzed with/without an adjustment for continuous and/or discrete covariates. For independent samples, a regression model tests the association between a trait and homozygosity intensity; a linear regression is used to examine the association between a quantitative trait and homozygosity intensity and a logistic regression is used to examine the association between a binary trait and homozygosity intensity. For correlated samples in a family study, GEE examines the association between a quantitative (binary) trait and homozygosity intensity by using an identity (logit) link.

Estimated homozygosity intensity is obtained from the output of **LOHAS**. Trait and covariate data are provided by users and saved in a trait-covariate file. In the trait-covariate data file, the first two columns are family ID (FID) and

individual ID (IID). **Note that this file MUST be a tab-delimited format.** Then trait data and covariate data are listed after the first two columns. Header names of trait variables and covariate variables should be given. The header names will be used to inform **LOHAS** a trait or covariate variable is quantitative or binary as follows: A prefix Q in header name indicates a **q**uantitative trait, B indicates a **b**inary trait, C indicates a **c**ontinuous covariate, and D indicates a **d**iscrete covariate. Analyses will be run independently if multiple phenotypes are provided in the trait-covariate data file, and all covariates in the data file will be included in models simultaneously. Please refer to Example 4 for illustration (see **Section 7.4**).

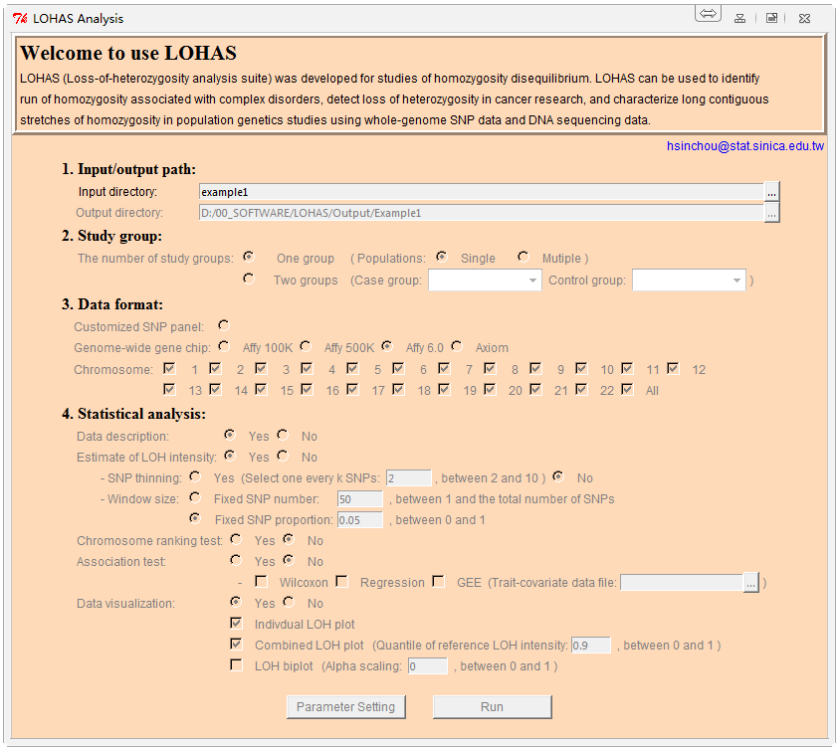
7. FIVE ILLUSTRATIVE EXAMPLES

LOHAS provides five illustrative examples. The first example demonstrates a LCSH analysis of “one group and single population” based on whole-genome SNP array data (see **Section 7.1**). The second example demonstrates a LCSH analysis of “one group and multiple populations” based on whole-genome SNP array data (see **Section 7.2**). The third example demonstrates a LOH analysis of “two groups” based on whole-genome SNP array data (see **Section 7.3**). The fourth and final examples demonstrate a homozygosity association study of “two groups” based on DNA sequencing data without incorporating a locus weight (see **Section 7.4**) and having a locus weight (see **Section 7.5**), respectively. Data of the examples are provided in directory “C:\LOHAS\Examples”. No folder for Example 5 is created because the same dataset has been provided in Example 4.

7.1 Example 1: A LCSH analysis of “one group and single population”

This example provides a LCSH analysis based on the Affymetrix Array6.0 data of two Japanese samples (JPT) from the International HapMap Project II. Genotype data files “NA18998.txt” and “NA19012.txt” are provided in the directory “C:\LOHAS\Examples\Example1”. Each of the two files contains four columns: they are probe set (SNP ID), chromosome, physical position, and genotype of SNP.

This example can be run easily by keying in “Example1” in the input directory (see **Figure 7-1-1**) and pressing the “Run” button. **LOHAS** starts to perform analysis and a message “Please wait a while, **LOHAS** is running…” will be shown in the command line in the R Console window. When the computation is finished, a message “Computation of **LOHAS** is finished.” is shown to acknowledge users the completion of **LOHAS** computation. The computational procedure will take about 8.5 minutes using a machine with an Intel® Core™ i5-2500K CPU @ 3.30GHz and RAM of 16.0 GB. Results of the analysis will be automatically saved in the output directory “C:\LOHAS\Output\Example1\” including two subdirectories (“Graphical results” and “Numerical results”) and two output files (“Data description.txt” and “Log.txt”). File “Data description.txt” describes the data attributes and parameter settings in the analysis (see **Figure 7-1-2**). File “Log.txt” provides a running progress and warning/error messages occurred in the **LOHAS** execution. In addition, file “ParSetting_forGroups.csv” provides the information about parameter setting in the analysis. Folder “TMP” contains temporary files generated by R in this analysis and can be used to continue a previous unfinished analysis due to an unexpected termination.



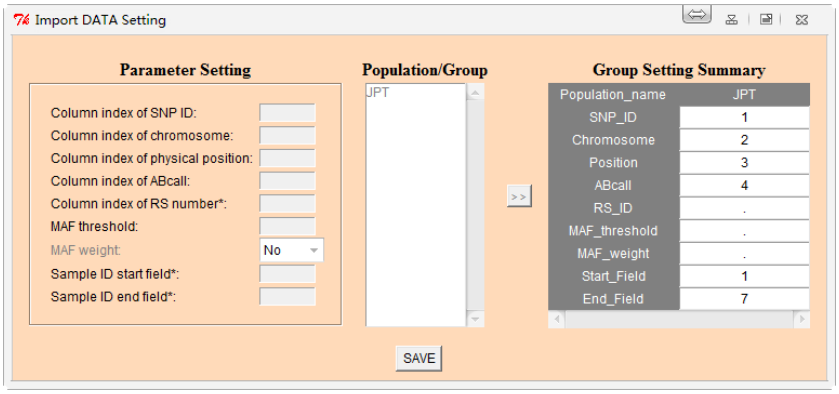


Figure 7-1-1. Execution and parameter setting of Example 1.

```
================================
===          Data description          ===
================================
1. Input/output path
    1) Input directory                                  :: D:/00_SOFTWARE/LOHAS/Examples/Example1
    2) Output directory                                 :: D:/00_SOFTWARE/LOHAS/Output/Example1
2. Study group
    1) The number of study groups                       :: 1 (1 population)
    2) The number of individuals                        :: 2 in 'JPT'
3. Data format
    1) Gene chip                                        :: Affy 6.0
    2) The chromosome(s) studied in LOH analysis        :: 1-22
        2-1)  The number of total markers               :: 869224 in 'JPT'
        2-2)  The number of markers across all individuals  :: 869224
        2-3)  The number of annotated markers (with locations)  :: 869224
4. Statistical analysis
    1) Data description                                 :: Yes
    2) Estimate of LOH intensity                        :: Yes
        2-1) SNP thinning                               :: No
        2-2) Window size                                :: Fixed SNP proportion 0.050000
    3) Chromosome ranking test                          :: No
    4) Association test                                 :: No
        4-1) Wilcoxon test                              :: No
        4-2) Regression test                            :: No
        4-3) GEE test                                   :: No
        4-4) trait-covariate file                       :: None
    5) Data visualization                               :: Yes
        1) Individual LOH plot                          :: Yes
        2) Combined LOH plot                            :: Yes. Quantile of reference LOH intensity is 0.9
        3) LOH biplot                                   :: No. Alpha scaling is 0
```

Figure 7-1-2. Data description in Example 1.

All graphical results are saved in the directory “Graphical results” which contains two subdirectories (“Combined plot” and “Individual plot”). Folder “Combined plot” contains all the combined homozygosity intensity plots. Here, “combined” means that a reference line (red color) is plotted combined with a homozygosity intensity curve of a study sample (blue color). The reference line is drawn based on a pre-specified threshold of homozygosity intensity (e.g., a value of 0.9) in a one-group analysis or an x% quantile of homozygosity intensities calculated based on normal samples in a two-group analysis (e.g., 90% quantile). Folder “Individual plot” contains all the individual homozygosity

intensity plots of all samples. Here we illustrate the results of combined homozygosity intensity plots. The combined homozygosity plots of sample NA18998 show that most of homozygosity intensities are lower than a threshold of 0.9 (see **Figure 7-1-3**). In contrast, the combined homozygosity plots of sample NA19012 show that this sample carries multiple regions of LCSH (chromosomes 2, 5, 6, 11, 17 and 21) (see **Figure 7-1-4**).

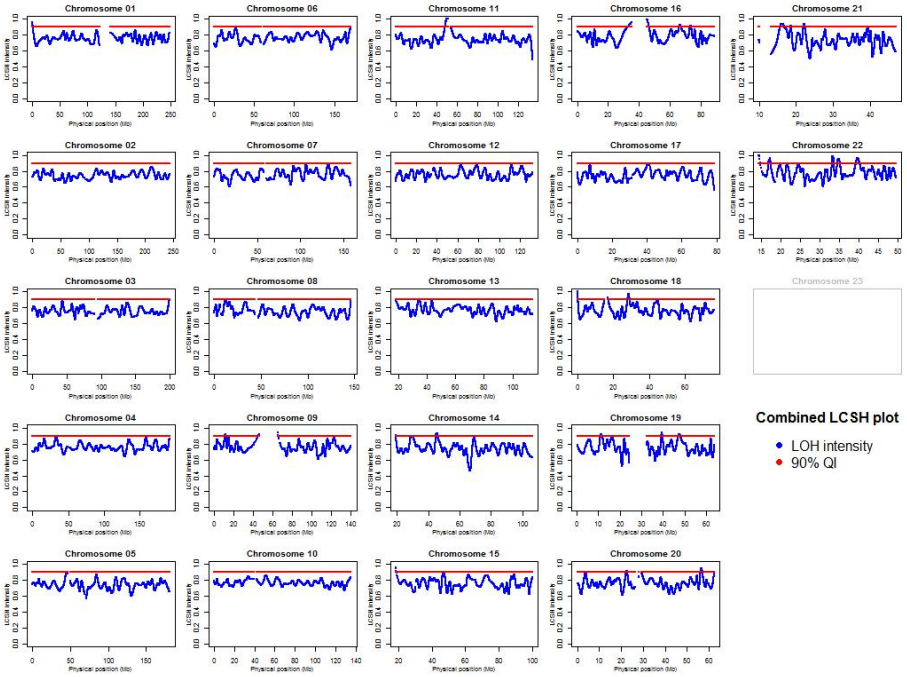


Figure 7-1-3. Combined homozygosity intensity plots of sample NA18998 in Example 1.

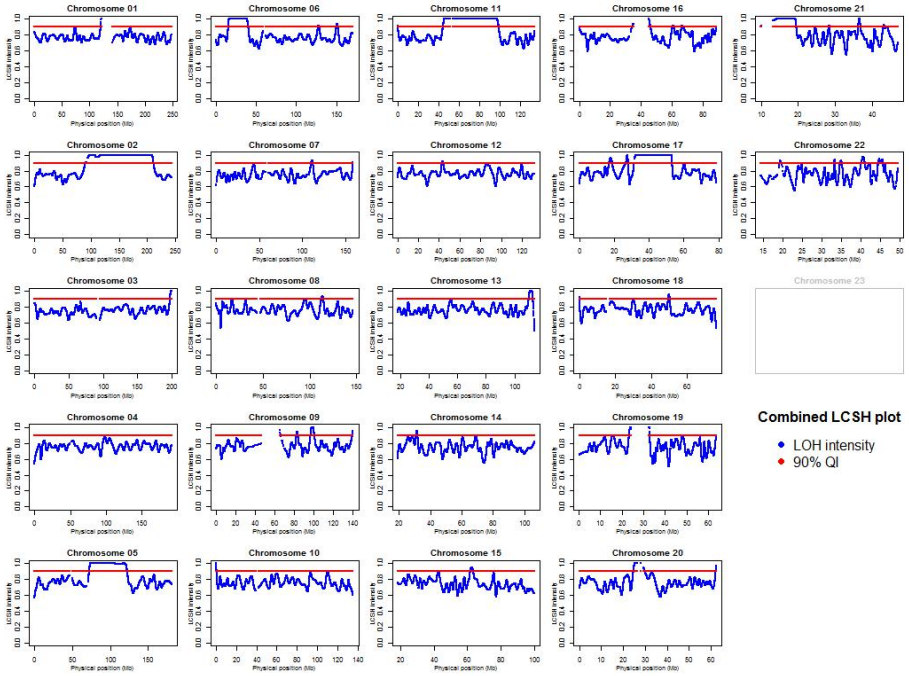


Figure 7-1-4. Combined homozygosity intensity plots of sample NA19012 in Example 1.

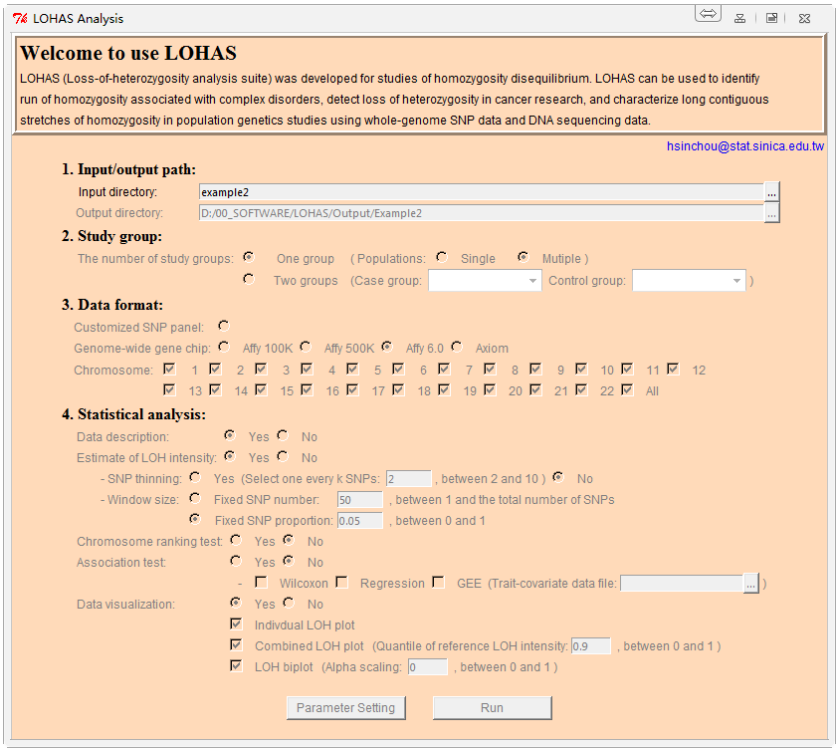
All numerical results are saved in directory “Numerical results”. Because only one population was analyzed in this example, only one folder “JPT” was created. The numerical results of homozygosity intensity are saved in Excel files by chromosome in the directory “C:\LOHAS\Output\Example1\LCSH intensity\JPT”. Each of the Excel files provides information, by chromosome, including SNP ID (probe set and/or RS number), physical position, and homozygosity intensities by samples. In addition, file “summary.csv” provides detailed summary information, including chromosome, physical position, SNP ID, RS number, MAF, major homozygote genotype, locus weight, and the summary results of homozygosity intensities including, (a) overall median and mean of homozygosity intensity for all samples, and (b) population/group median, mean, and standard deviation of homozygosity intensity for individuals in each population and group.

7.2 Example 2: A LCSH analysis of “one group and multiple populations”

This example illustrates a LCSH analysis based on the Affymetrix Array 6.0 data of two Caucasian samples (CEU), two Chinese samples (CHB), and two African samples (YRI) from the International HapMap Project II. Genotype data

files “NA06985.txt” and “NA06991.txt” are provided in the sub-directories in the directory “C:\LOHAS\Examples\Example2\CEU”. Genotype data files “NA18524.txt” and “NA18526.txt” are provided in the sub-directories in the directory “C:\LOHAS\Examples\Example2\CHB”. Genotype data files “NA18500.txt” and “NA18501.txt” are provided in the sub-directories in the directory “C:\LOHAS\Examples\Example2\YRI”. Each data file contains four columns: they are probe set (SNP ID), chromosome, physical position, and genotype of SNP.

This example can be run easily by keying in “Example2” in the input directory (see **Figure 7-2-1**) and pressing the “Run” button. **LOHAS** starts to perform analysis and a message “Please wait a while, **LOHAS** is running…” will be shown in the command line. When the computation is finished, a message “Computation of **LOHAS** is finished.” is shown to acknowledge users the completion of **LOHAS** computation. The computational procedure will take about 30.5 minutes using a machine with an Intel® Core™ i5-2500K CPU @ 3.30GHz and RAM of 16.0 GB. File “Data description.txt” describes the data attributes and parameter settings in the analysis (see **Figure 7-2-2**).



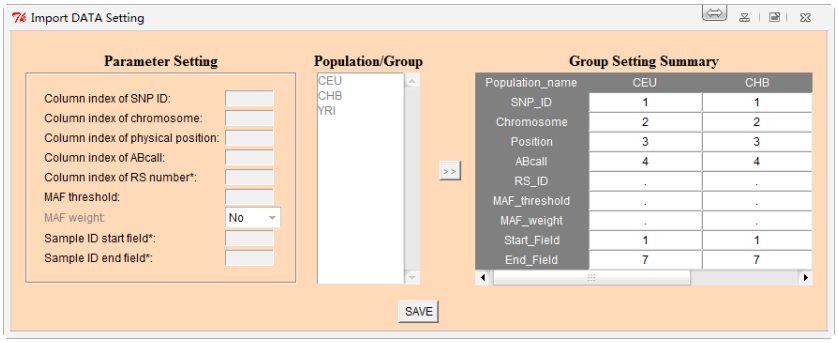


Figure 7-2-1. Execution and parameter setting of Example 2.

```
=====================================
===          Data description       ===
=====================================
1. Input/output path
    1) Input directory                                  :: D:/00_SOFTWARE/LOHAS/Examples/Example2
    2) Output directory                                 :: D:/00_SOFTWARE/LOHAS/Output/Example2
2. Study group
    1) The number of study groups                       :: 1 (3 populations)
    2) The number of individuals                        :: 2 in 'CEU', 2 in 'CHB', 2 in 'YRI'
3. Data format
    1) Gene chip                                        :: Affy 6.0
    2) The chromosome(s) studied in LOH analysis        :: 1-22
        2-1)  The number of total markers               :: 869224 in 'CEU', 869224 in 'CHB', 869224 in 'YRI'
        2-2)  The number of markers across all individuals  :: 869224
        2-3)  The number of annotated markers (with locations)  :: 869224
4. Statistical analysis
    1) Data description                                 :: Yes
    2) Estimate of LOH intensity                        :: Yes
        2-1) SNP thinning                               :: No
        2-2) Window size                                :: Fixed SNP proportion 0.050000
    3) Chromosome ranking test                          :: No
    4) Association test                                 :: No
        4-1) Wilcoxon test                              :: No
        4-2) Regression test                            :: No
        4-3) GEE test                                   :: No
        4-4) trait-covariate file                       :: None
    5) Data visualization                               :: Yes
        1) Individual LOH plot                          :: Yes
        2) Combined LOH plot                            :: Yes. Quantile of reference LOH intensity is 0.9
        3) LOH biplot                                   :: No. Alpha scaling is 0
```

Figure 7-2-2. Data description in Example 2.

Graphical results are saved in the directory “Graphical results” which contains three subdirectories (“Biplot”, “Combined plot”, and “Individual plot”). First, all chromosome-wise biplots of 22 autosomes are saved in folder “Biplot”. For example, a biplot drawn based on SNPs on chromosome 1 shows that two CEU samples (blue), two CHB samples (red), and two YRI samples (purple) are separated (see **Figure 7-2-3**). Markers are presented with a green-to-pink color spectrum according to their ordered physical positions. Moreover, the biplot legend provides the location of the centromeric gap (symbol: circle), and the number of SNPs on the p-arm and q-arm of a study chromosome and the physical positions of the initial and last SNPs interrogated on the study SNP

chip. Second, all chromosome-wise combined homozygosity intensity plots are saved in the folder “Combined plot” by population. The plots of CEU, CHB and YRI populations are saved in folders “CEU”, “CHB” and “YRI”, respectively. Finally, the results of individual homozygosity intensity plots are arranged in the folder “Individual plot”.

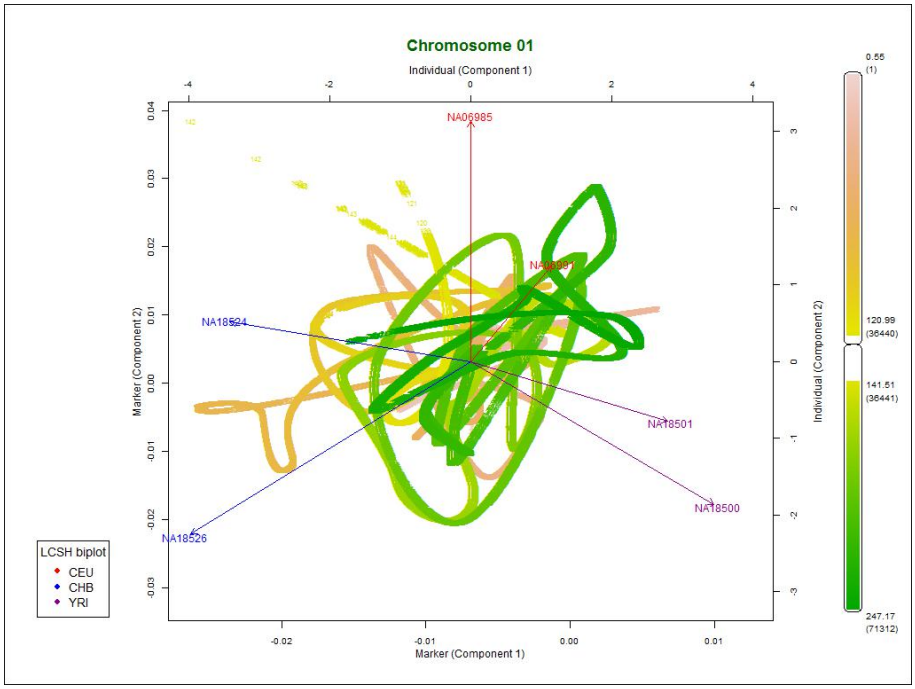


Figure 7-2-3. Biplot drawn based on SNPs on chromosome 1 for six samples from three populations (CEU, CHB and YRI) in Example 2.

Numerical results are saved in the directory “Numerical results” which contains two subdirectories (“LCSH biplot” and “LCSH intensity”). Numerical results of LCSH biplot are saved in Excel files by chromosome by population in the folder “LCSH biplot”. The output consists of three parts. First, a label of singular value, the proportion of explained variation, and the cumulative proportion of variation explained by principal component are listed. Second, population name and individual ID are listed and followed by scores of first two important “individual” components for each sample. Finally, marker ID and physical position are listed and followed by scores of first two important “marker” components for each SNP. In addition, the numerical results of homozygosity intensity are saved in Excel files by population, by sample, and by chromosome in the folder “C:\LOHAS\Output\Example2\LCSH intensity\”. Each of the Excel files provides information including SNP ID (probe set and/or

RS number), physical position, homozygosity intensities of each samples, and a pre-specified threshold of 0.9.

7.3 Example 3: A LOH analysis of “two groups”

This example illustrates a LOH analysis based on the Affymetrix 100K data of two cases and ten normal controls. The data analyzed in this example were generated from a simulation study and provided by ALOHA software (Yang et al., *BMC Genomics*, 2010). All normal controls do not carry any regions of LOH. The first cancer patient is a female and carries regions of LOH on chromosomes 2, 7, 8, 9, 10, 11 and 12. The second cancer patient is a male and carries regions of LOH on chromosomes 7, 8, 9, 10, 11 and 12.

Genotype data of the two cancer patients are saved in the directory “C:\LOHAS\Examples\Example3\Case”. For the first sample (a female), genotype data of SNPs on Hind and Xba are saved in files “Abnorm_01_F_Hind_Genotype.txt” and “Abnorm_01_F_Xba_Genotype.txt”, respectively. For the second sample (a male), genotype data of SNPs on Hind and Xba are saved in files “Abnorm_02_M_Hind_Genotype.txt” and “Abnorm_02_M_Xba_Genotype.txt”, respectively. The similar data structure and nomenclature rule are also applied to the data of control samples. Genotype data of ten normal controls are saved in the directory “C:\LOHAS\Examples\Example3\Control”.

This example can be run easily by keying in “Example3” in the input directory (see **Figure 7-3-1**) and pressing the “Run” button. Then **LOHAS** starts to perform analysis and a message “Please wait a while, **LOHAS** is running…” will be shown in the command line in the R Console window. When the computation is finished, a message “Computation of **LOHAS** is finished.” is shown to acknowledge users the completion of **LOHAS** computation. The computational procedure will take about 56 minutes using a machine with an Intel® Core™ i5-2500K CPU @ 3.30GHz and RAM of 16.0 GB. Results of the analysis will be automatically saved in the output directory “C:\LOHAS\Output\Example3\”, including two subdirectories (“Graphical results” and “Numerical results”) and two output files (“Data description.txt” and “Log.txt”).

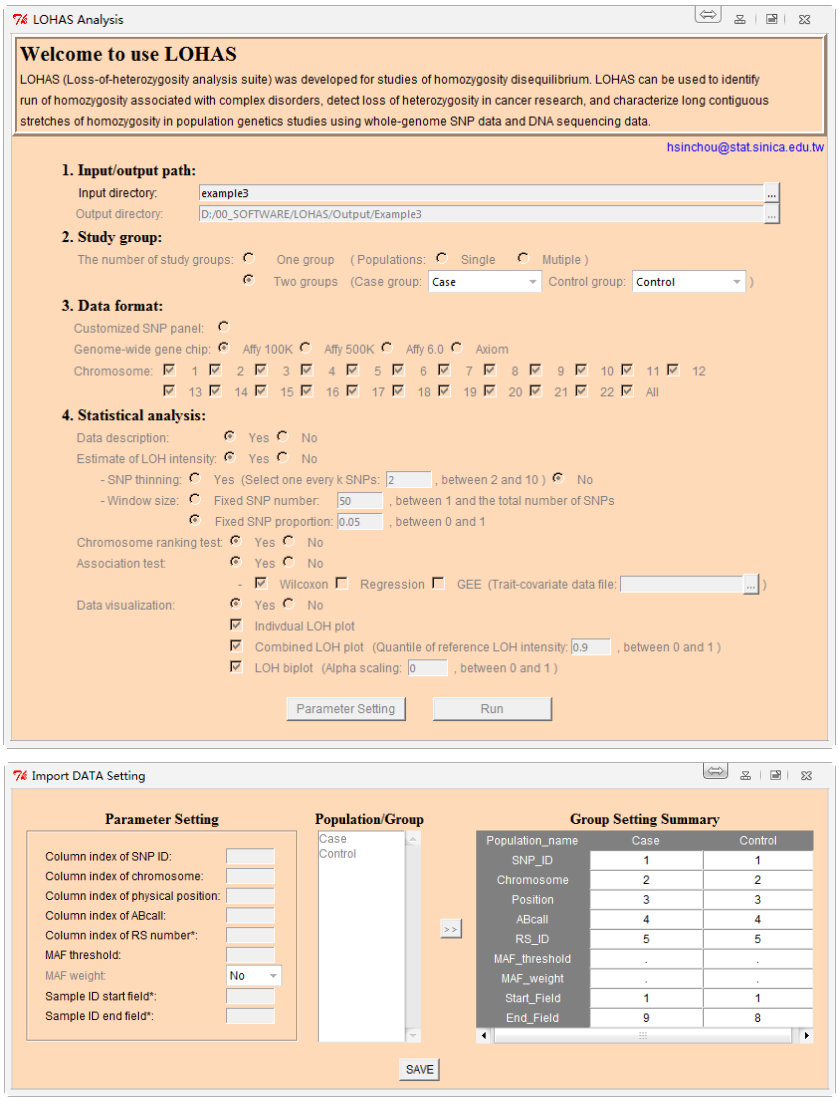


Figure 7-3-1. Execution and parameter setting of Example 3.

File “Data description.txt” describes the data attributes and parameter settings in the analysis (see **Figure 7-3-2**). File “Log.txt” provides running progress and warning/error messages related to **LOHAS** execution in the analysis. In this example, all Hind-only analysis (Chip 1), Xba-only analysis (Chip 2), and merge-chip analysis were carried out. Their results were saved in the same folder but different filenames. **Note that the results from an analysis of Chip 1, Chip 2 and merged chip are saved in files with a label “Chip1”, “Chip2” and “Chip12” in filenames, respectively.**

```
====================================
===          Data description      ===
====================================
1. Input/output path
   1) Input directory                                  :: D:/00_SOFTWARE/LOHAS/Examples/Example3
   2) Output directory                                 :: D:/00_SOFTWARE/LOHAS/Output/Example3
2. Study group
   1) The number of study groups                       :: 2 (Patient group: Case)
   2) The number of individuals                        :: 2 in 'Case', 10 in 'Control'
3. Data format
   1) Gene chip                                        :: Affy 100K
   2) The chromosome(s) studied in LOH analysis        :: 1-23
      2-1)  The number of total markers                :: 115560 in 'Case', 115560 in 'Control'
      2-2)  The number of markers across all individuals :: 115560
      2-3)  The number of annotated markers (with locations) :: 115560
4. Statistical analysis
   1) Data description                                 :: Yes
   2) Estimate of LOH intensity                        :: Yes
      2-1) SNP thinning                                :: No
      2-2) Window size                                 :: Fixed SNP proportion 0.050000
   3) Chromosome ranking test                          :: Yes
   4) Association test                                 :: Yes
      4-1) Wilcoxon test                               :: Yes
      4-2) Regression test                             :: No
      4-3) GEE test                                    :: No
      4-4) trait-covariate file                        :: None
   5) Data visualization                               :: Yes
      1) Individual LOH plot                           :: Yes
      2) Combined LOH plot                             :: Yes. Quantile of reference LOH intensity is 0.9
      3) LOH biplot                                    :: No. Alpha scaling is 0
```

Figure 7-3-2. Data description in Example 3.

Graphical results are saved in the directory “Graphical results” which contains four subdirectories (“Combined plot”, “Individual plot”, “Biplot”, and “Association test”). First, all chromosome-wise homozygosity intensity plots are saved in the folder “Combined plot” by case-control groups. The plots of cases (Group 1) and controls (Group 2) are saved in directories “Group 1” and “Group 2”, respectively. “Group 1” or “Group 2” may be a case or control group depending upon the users’ assignment in LOHAS interface. Homozygosity intensity plots of the first and second patients in the analysis of merged chip were saved in file “IP_Popu1_Abnorm_01_Chip12_AllChr.jpeg” (see **Figure 7-3-3**) and “IP_Popu1_Abnorm_02_Chip12_AllChr.jpeg” (see **Figure 7-3-4**), respectively. As expected, the first cancer is a female (refer to the non-homozygous pattern in sex chromosome) and shows multiple regions of LOH on chromosomes 2, 7, 8, 9, 10, 11, and 12. The second cancer patient is a male (refer to the homozygous pattern in sex chromosome) and carries regions of LOH on chromosomes 7, 8, 9, 10, 11 and 12. Second, the results of individual homozygosity intensity plots are arranged in the folder “Individual plot”. Third, all chromosome-wise biplots of 22 autosomes are saved in folder “Biplot”. For example, the biplot from the analysis of merged chip was saved in

file “BP_Chip12_AllChr.jpeg” (see **Figure 7-3-5**). As expected, one or two patient samples are pulled out the majority of samples in biplots of chromosomes 2, 7, 8, 9, 10, 11, and 12. Finally, a nonparametric Wilcoxon association test was performed to identify regions of LOH that a patient group had a higher median of homozygosity intensity than a control group. All chromosome-wise p-values of 22 autosomes are saved in folder “Association test”. For example, the p-value plot of Wilcoxon association test in the analysis of merged chip was saved in file “Wilcoxon_test_Chip12_AlChr.jpeg” (see **Figure 7-3-6**). Because sample size was small no significant regions of LOH were found in this example.

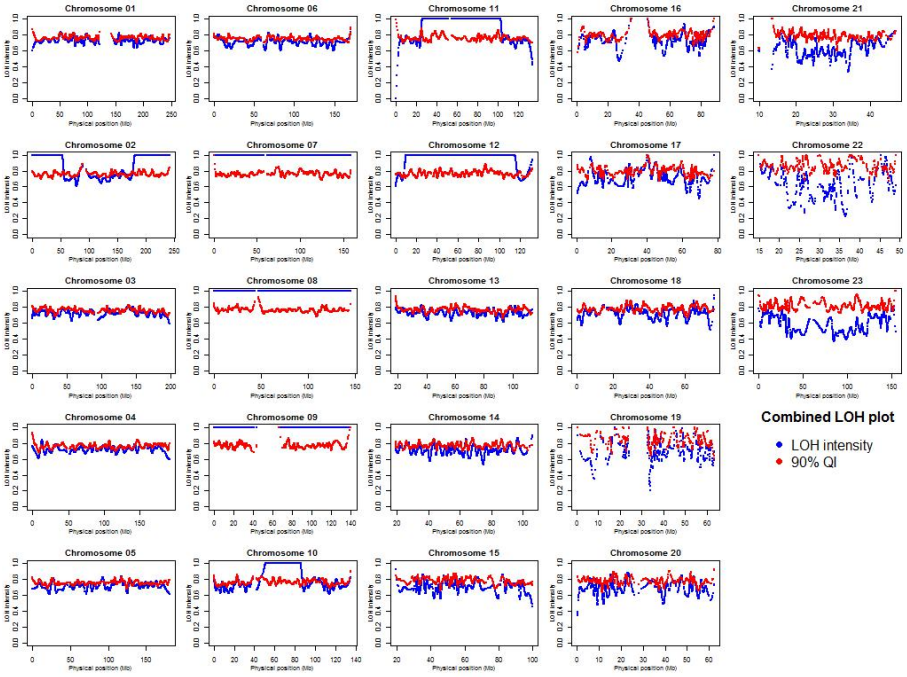


Figure 7-3-3. Combined homozygosity intensity plots of the first cancer patient.

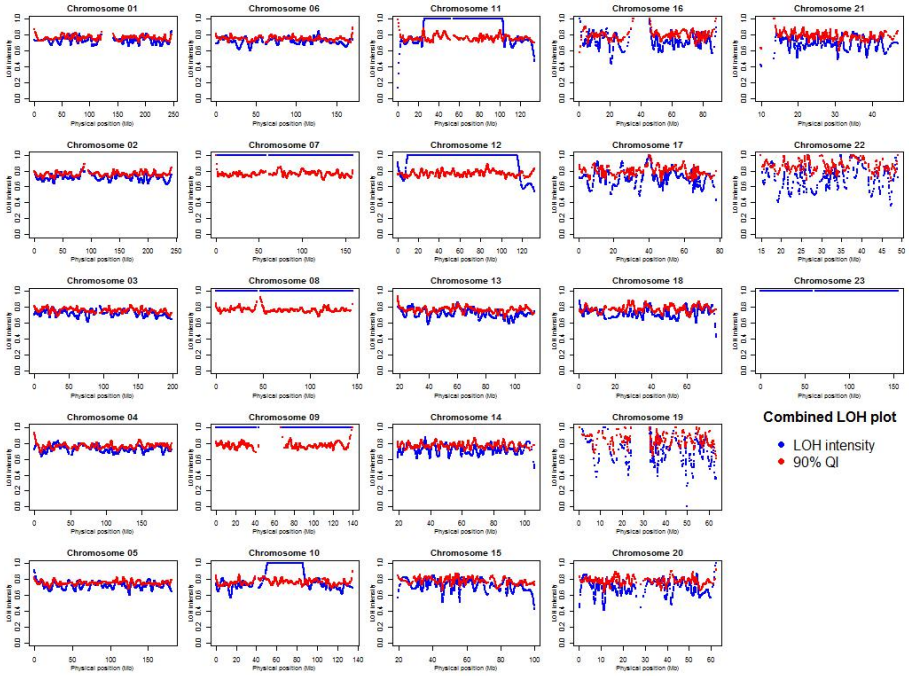


Figure 7-3-4. Combined homozygosity intensity plots of the second cancer patient.

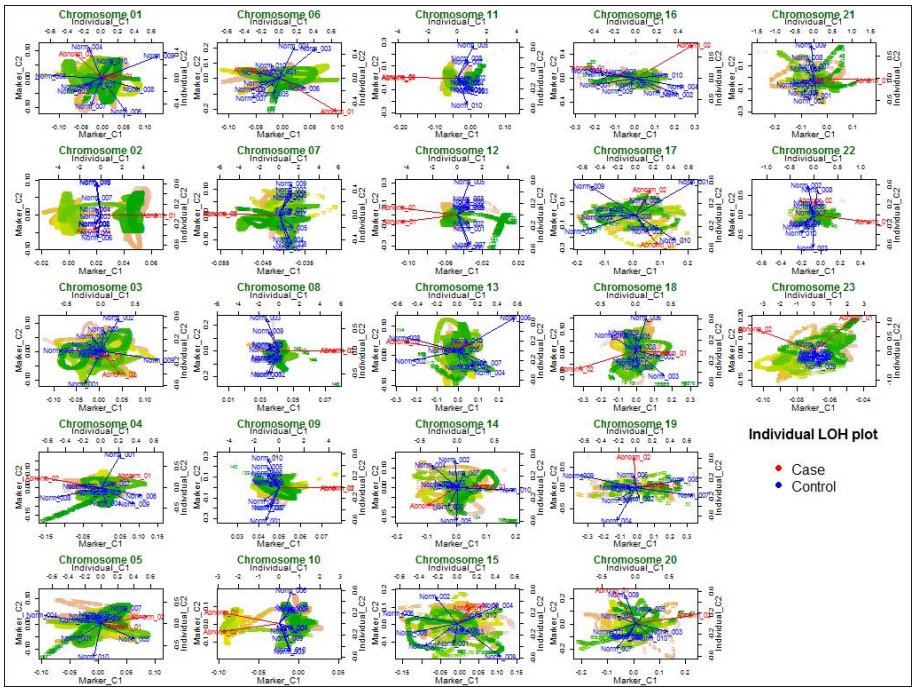


Figure 7-3-5. Chromosome-wise biplots drawn for 2 cancer patients (red color) and 10 normal controls (blue color) in Example 3.

Numerical results are saved in the directory “Numerical results” which contains four subdirectories (“LOH intensity”, “LOH biplot”, “Chromosome

ranking test", and "Association test"). First, numerical results of homozygosity intensity are saved in Excel files by group and by chromosome in the folder "C:\LOHAS\Output\Example3\LOH intensity\". Each of the Excel files provides information including SNP ID (probe set and/or RS number), physical position, homozygosity intensities of samples, and a 90% quantile of homozygosity intensity in a control group. Second, numerical results of LOH biplot are saved in Excel files by chromosome in the folder "LOH biplot". The output consists of three parts as illustrated in **Section 7.2**. In this example, the first two components explain 31% of total variation for chromosome 1 in a merge chip analysis. The scores of "individual" components and "marker" components are provided too. Third, numerical results of p-values of chromosome ranking tests are saved in Excel files by analysis of Chip1, Chip2 and merged chip in the folder "C:\LOHAS\Output\Example3\Chromosome ranking test".

Finally, numerical results of p-values in Wilcoxon association tests are provided in the last column in Excel files by chromosome in the folder "Association test". In addition, the Excel file also contains the following information for each anchor SNP of sliding windows: SNP ID (probe set), RS number, physical position and some statistics including mean, median, standard deviation of homozygosity intensity, and proportions of samples which have higher homozygosity intensities than a reference threshold in case group and in control group. In this example, the reference threshold indicates the red curve in the combined homozygosity intensity plot, i.e., the 90% quantile of homozygosity intensity in control group.

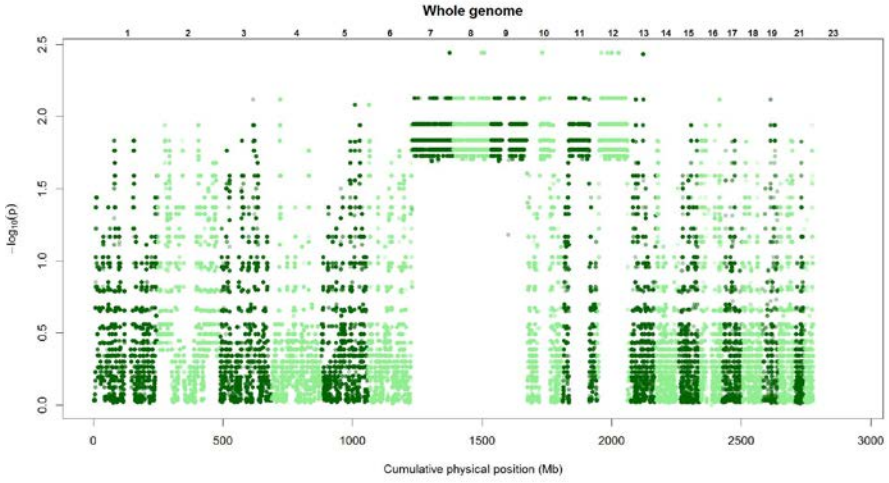


Figure 7-3-6. P-value plot of Wilcoxon test in Example 3.

7.4 Example 4: A homozygosity association analysis of “two groups” based on DNA sequencing data - an analysis which does not consider a locus weight

This example illustrates a homozygosity association analysis based on the DNA sequencing data. The data analyzed in this example were generated based on the DNA sequencing data of chromosome 22 from 14 IBS samples (Iberian Populations in Spain) in the 1000 Genomes Project. Seven samples were randomly assigned as cases (3 males and 4 females) and seven samples were assigned as normal controls (4 males and 3 females). To generate a disease-associated ROH, we replaced the original genotypes of all seven cases with homozygous genotypes in the region from 38.946 Mb to 40.889 Mb but remained the genotypes of all seven controls unchanged. Wilcoxon test was performed to detect the pre-specified disease-associated ROH.

Genotype data files of the seven cases are provided in the sub-directories in the directory “C:\LOHAS\Examples\Example4\Case” and genotype data files of the seven controls are provided in the sub-directories in the directory “C:\LOHAS\Examples\Example4\Control”. Their covariate data are saved in file “covraiates.txt”. Actually, this trait-covariate file was not necessary for association analysis of Wilcoxon test.

This example can be run easily by keying in “Example4” in the input directory (see **Figure 7-4-1**) and pressing the “Run” button. Then **LOHAS** starts to perform analysis and a message “Please wait a while, **LOHAS** is running…” will be shown in the command line in the R Console window. When the computation is finished, a message “Computation of **LOHAS** is finished.” is shown to acknowledge users the completion of **LOHAS** computation. The computational procedure will take about 22.5 hours using a machine with an Intel® Core™ i5-2500K CPU @ 3.30GHz and RAM of 16.0 GB. Results of the analysis will be automatically saved in the output directory “C:\LOHAS\Output\Example4\”, including two subdirectories (“Graphical results” and “Numerical results”) and two output files (“Data description.txt” and “Log.txt”). File “Data description.txt” describes the data attributes and parameter settings in the analysis (see **Figure 7-4-2**). File “Log.txt” provides running progress and warning/error messages related to **LOHAS** execution in

the analysis.

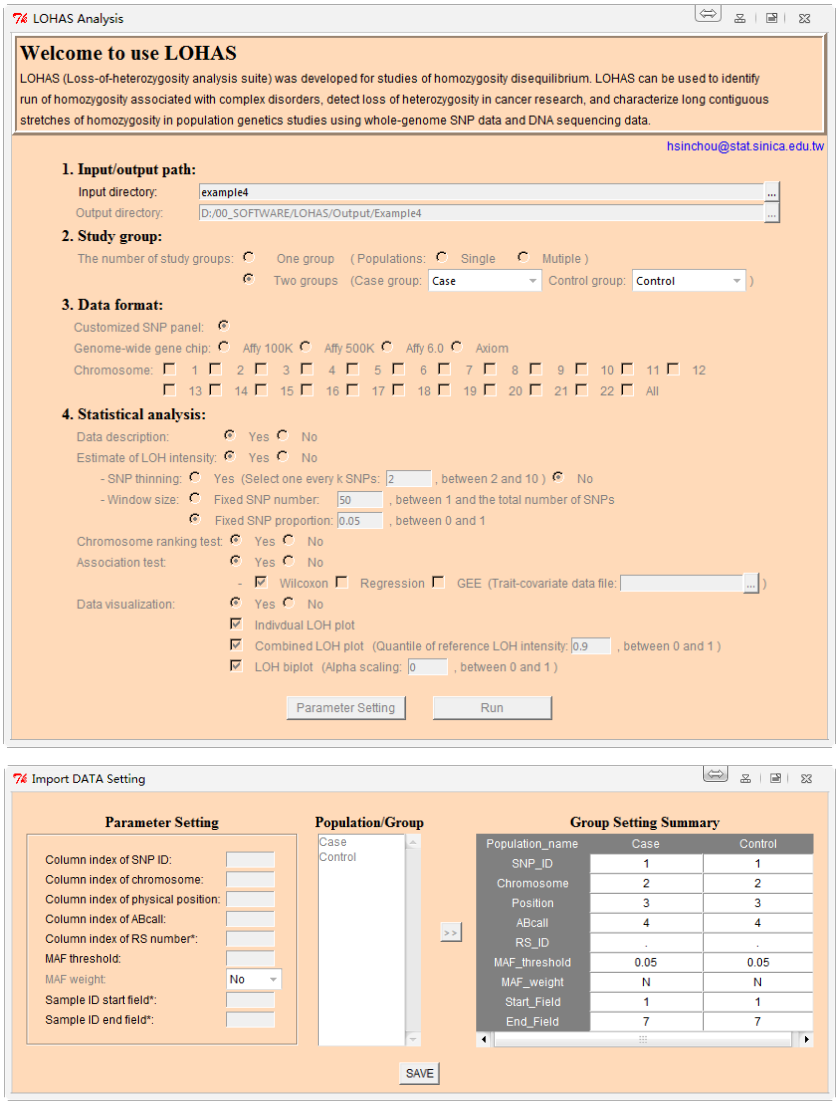


Figure 7-4-1. The execution and parameter setting of Example 4.

```
=====================================
===          Data description       ===
=====================================
1. Input/output path
    1) Input directory                                  :: D:/00_SOFTWARE/LOHAS/Examples/Example4
    2) Output directory                                 :: D:/00_SOFTWARE/LOHAS/Output/Example4
2. Study group
    1) The number of study groups                       :: 2 (Patient group: Case)
    2) The number of individuals                        :: 7 in 'Case', 7 in 'Control'
3. Data format
    1) Gene chip                                        :: Customized SNP panel
    2) The chromosome(s) studied in LOH analysis        :: 22
        2-1)  The number of total markers               :: 509283 in 'Case', 509283 in 'Control'
        2-2)  The number of markers across all individuals  :: 509283
        2-3)  The number of annotated markers (with locations)  :: 509283
4. Statistical analysis
    1) Data description                                 :: Yes
    2) Estimate of LOH intensity                        :: Yes
        2-1) SNP thinning                               :: No
        2-2) Window size                                :: Fixed SNP proportion 0.050000
```

3) Chromosome ranking test	:: Yes
4) Association test	:: Yes
4-1) Wilcoxon test	:: Yes
4-2) Regression test	:: No
4-3) GEE test	:: No
4-4) trait-covariate file	:: None
5) Data visualization	:: Yes
1) Individual LOH plot	:: Yes
2) Combined LOH plot	:: Yes. Quantile of reference LOH intensity is 0.9
3) LOH biplot	:: No. Alpha scaling is 0

Figure 7-4-2. The data description in Example 4.

Graphical results are saved in the directory “Graphical results” which contains four subdirectories (“Combined plot”, “Individual plot”, “Biplot”, and “Association test”). First, a combined homozygosity intensity plot is saved in the folder “Combined plot” by case-control groups. The intensity plots of cases (Group 1) and controls (Group 2) are saved in directories “Group 1” and “Group 2”, respectively. “Group 1” or “Group 2” may be a case or control group depending upon the users’ assignment in **LOHAS** interface. For example, combined homozygosity intensity plot of a case is saved in file “CP_Popu1_HG01518_Chr22.jpeg” (see **Figure 7-4-3**). Homozygosity intensity plot of a control is saved in file “CP_Popu2_HG01617_Chr22.jpeg” (see **Figure 7-4-4**). As expected, a region of ROH is observed in the region around 40 Mb for the case but not for the control. Second, the results of individual homozygosity intensity plots are arranged in the folder “Individual plot”. Third, a biplot of the 14 samples is saved in file “BP_Chr12.jpeg” in folder “Biplot” (see **Figure 7-4-5**). We observe that cases (left-hand side; blue) and controls (right-hand side; red) are separated. Finally, a nonparametric Wilcoxon association test was performed to identify regions of LOH that a case group had a higher median of homozygosity intensity than a control group. P-values of association tests are saved in file “Wilcoxon_test_Chr22.jpeg” in folder “Association test” (see **Figure 7-4-6**). Although sample size is so small, association signal of the pre-specified region of ROH around 40 Mb is still obvious.

Numerical results are saved in the directory “Numerical results” which contains four subdirectories (“LOH intensity”, “LOH biplot”, “Chromosome ranking test”, and “Association test”). First, numerical results of homozygosity intensity are saved in Excel files by group and by chromosome in the folder “C:\LOHAS\Output\Example4\LOH intensity\”. Each of the Excel files provides

information including SNP ID (probe set and/or RS number), physical position, homozygosity intensities of samples, and a 90% quantile of homozygosity intensity in a control group. Second, numerical results of LOH biplot are saved in Excel file in the folder “LOH biplot”. The output consists of three parts as illustrated in **Section 7.2**. In this example, the first two components explain 34% of total variation. The scores of “individual” components and “marker” components are provided too. Third, numerical results of p-values of chromosome ranking tests are saved in Excel files in the folder “C:\LOHAS\Output\Example4\Chromosome ranking test”.

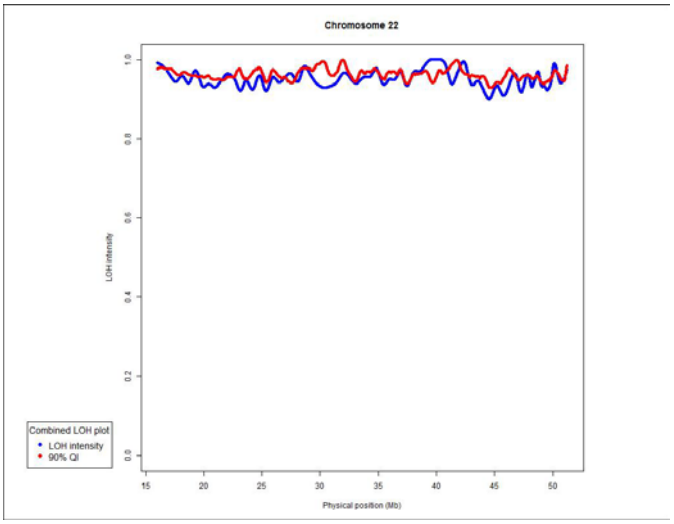


Figure 7-4-3. Combined homozygosity intensity plots of a case (HG01518).

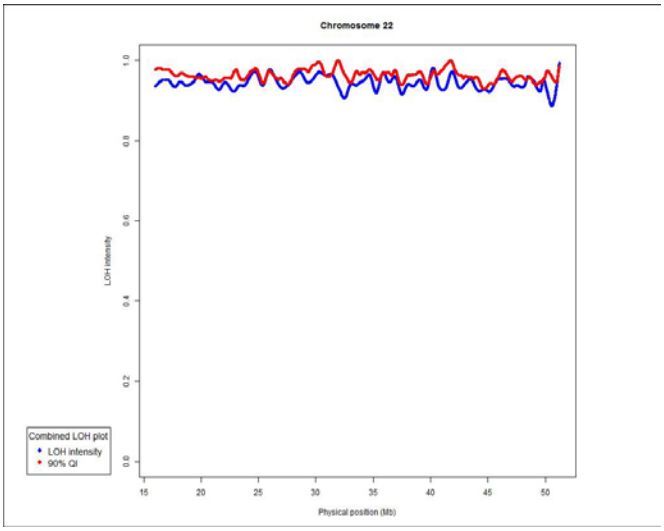


Figure 7-4-4. Combined homozygosity intensity plots of a control (HG01517).

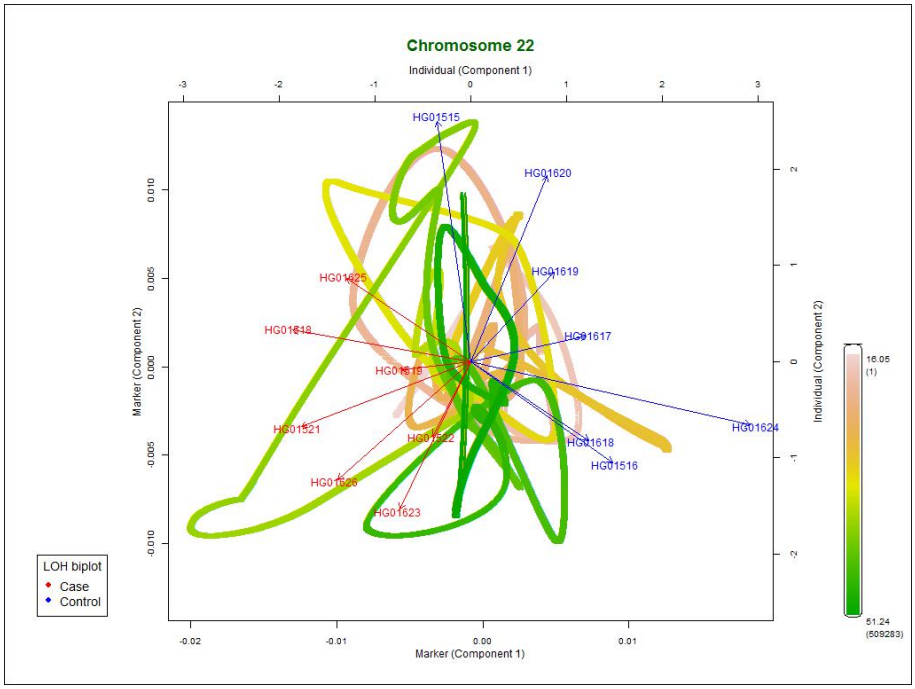


Figure 7-4-5. A biplot drawn for 7 cases (red color) and 7 controls (blue color) in Example 4.

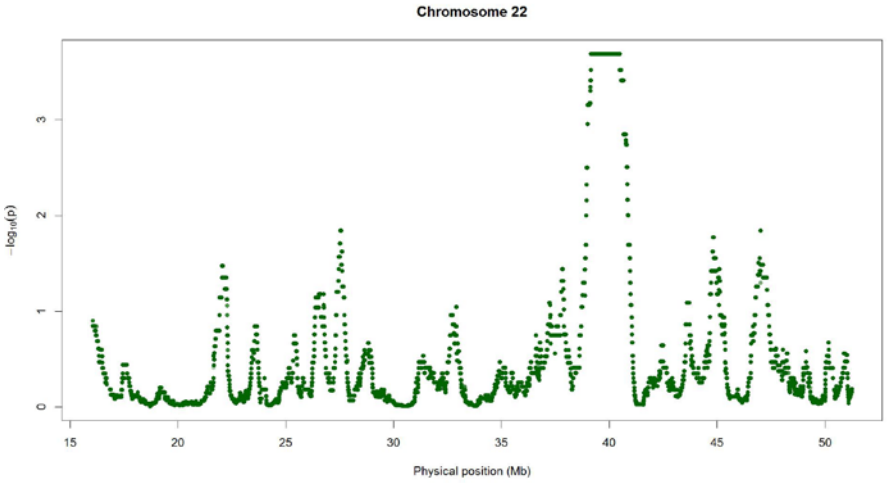


Figure 7-4-6. P-value plot of Wilcoxon homozygosity association test in Example 4.

Finally, numerical results of p-values in Wilcoxon association tests are provided in the last column in Excel files by chromosome in the folder “Association test”. In addition, the Excel file also contains the following information for each anchor SNP of sliding windows: SNP ID (probe set), RS number, physical position and some statistics including mean, median,

standard deviation of homozygosity intensity, and proportions of samples which have higher homozygosity intensities than a reference threshold in case group and in control group. In this example, the reference threshold indicates the red curve in the combined homozygosity intensity plot, i.e., the 90% quantile of homozygosity intensity in control group.

7.5 Example 5: A homozygosity association analysis of “two groups” based on DNA sequencing data - an analysis which considers a locus weight

This example analyzes the same dataset of Example 4 but this analysis considers an additional locus weight which was not considered in Example 4. Lower locus weights are assigned to the common homozygotes of rare variants that have a lower minor allele frequency; a locus weight is the minor allele frequency divided by 0.05.

This example can be run easily by keying in “Example5” in the input directory (see **Figure 7-5-1**) and pressing the “Run” button. Then **LOHAS** starts to perform analysis and a message “Please wait a while, **LOHAS** is running…” will be shown in the command line in the R Console window. When the computation is finished, a message “Computation of **LOHAS** is finished.” is shown to acknowledge users the completion of **LOHAS** computation. The computational procedure will take about 20.5 hours using a machine with an Intel® Core™ i7-960 CPU @ 3.2GHz and RAM of 18.0 GB. Results of the analysis will be automatically saved in the output directory “C:\LOHAS\Output\Example5\”, including two subdirectories (“Graphical results” and “Numerical results”) and two output files (“Data description.txt” and “Log.txt”). File “Data description.txt” describes the data attributes and parameter settings in the analysis (see **Figure 7-5-2**). File “Log.txt” provides running progress and warning/error messages related to **LOHAS** execution in the analysis.

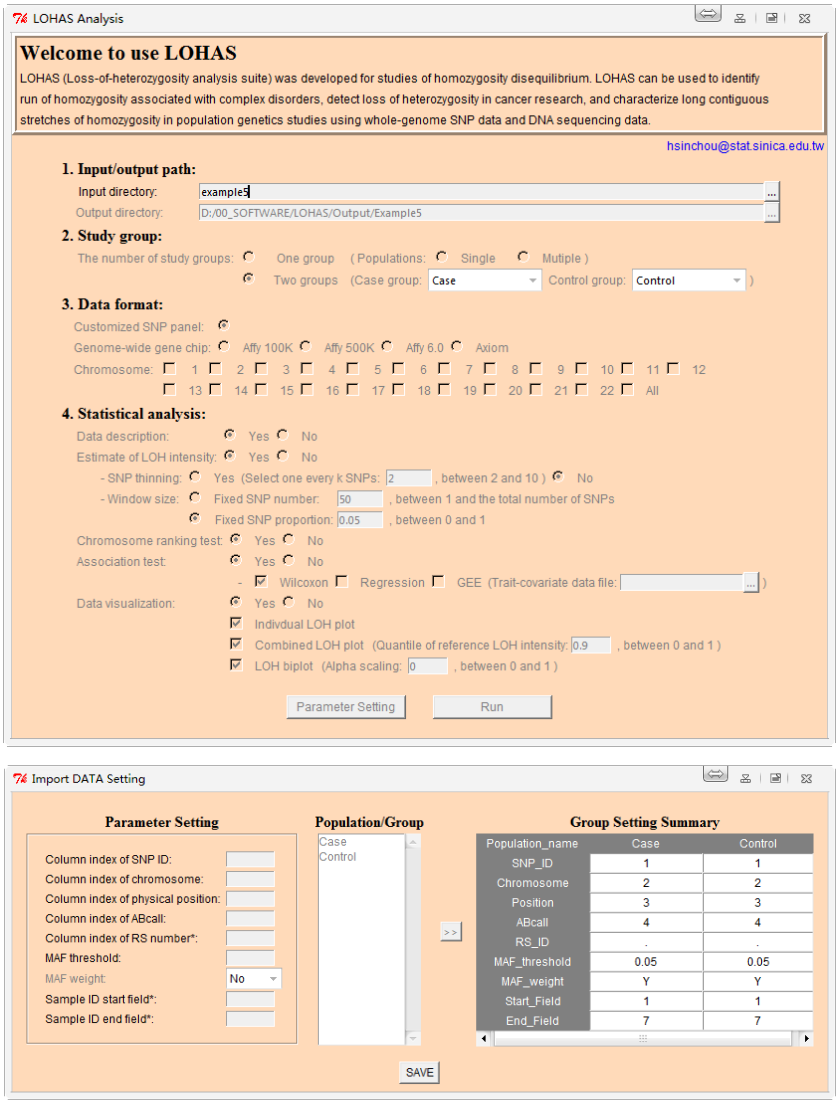


Figure 7-5-1. The execution and parameter setting of Example 5.

```
=====================================
===          Data description       ===
=====================================
1. Input/output path
    1) Input directory                                  :: D:/00_SOFTWARE/LOHAS/Examples/Example4
    2) Output directory                                 :: D:/00_SOFTWARE/LOHAS/Output/Example5
2. Study group
    1) The number of study groups                       :: 2 (Patient group: Case)
    2) The number of individuals                        :: 7 in 'Case', 7 in 'Control'
3. Data format
    1) Gene chip                                        :: Customized SNP panel
    2) The chromosome(s) studied in LOH analysis        :: 22
        2-1)  The number of total markers               :: 509283 in 'Case', 509283 in 'Control'
        2-2)  The number of markers across all individuals  :: 509283
        2-3)  The number of annotated markers (with locations)  :: 509283
4. Statistical analysis
    1) Data description                                 :: Yes
    2) Estimate of LOH intensity                        :: Yes
        2-1) SNP thinning                               :: No
        2-2) Window size                                :: Fixed SNP proportion 0.050000
    3) Chromosome ranking test                          :: Yes
    4) Association test                                 :: Yes
        4-1) Wilcoxon test                              :: Yes
        4-2) Regression test                            :: No
        4-3) GEE test                                   :: No
```

4-4) trait-covariate file :: None
5) Data visualization :: Yes
1) Individual LOH plot :: Yes
2) Combined LOH plot :: Yes. Quantile of reference LOH intensity is 0.9
3) LOH biplot :: No. Alpha scaling is 0

Figure 7-5-2. The data description in Example 5.

The results of this example are similar to the results of **Example 4**. Graphical results are saved in the directory “Graphical results” which contains four subdirectories (“Combined plot”, “Individual plot”, “Biplot”, and “Association test”). First, a combined homozygosity intensity plot is saved in the folder “Combined plot” by case-control groups. The intensity plots of cases (Group 1) and controls (Group 2) are saved in directories “Group 1” and “Group 2”, respectively. “Group 1” or “Group 2” may be a case or control group depending upon the users’ assignment in **LOHAS** interface. For example, combined homozygosity intensity plot of a case is saved in file “CP_Popu1_HG01518_Chr22.jpeg” (see **Figure 7-5-3**). Homozygosity intensity plot of a control is saved in file “CP_Popu2_HG01617_Chr22.jpeg” (see **Figure 7-5-4**). As expected, a region of ROH is observed in the region around 40 Mb for the case but not for the control. Second, the results of individual homozygosity intensity plots are arranged in the folder “Individual plot”. Third, a biplot of the 14 samples is saved in file “BP_Chr12.jpeg” in folder “Biplot” (see **Figure 7-5-5**). We observe that cases (right-hand side; blue) and controls (left-hand side; red) are separately. Finally, a nonparametric Wilcoxon association test was performed to identify regions of LOH that a case group had a higher median of homozygosity intensity than a control group. P-values of association tests are saved in file “Wilcoxon_test_Chr22.jpeg” in folder “Association test” (see **Figure 7-5-6**). Although sample size is so small, association signal of the pre-specified region of ROH around 40 Mb is still obvious.

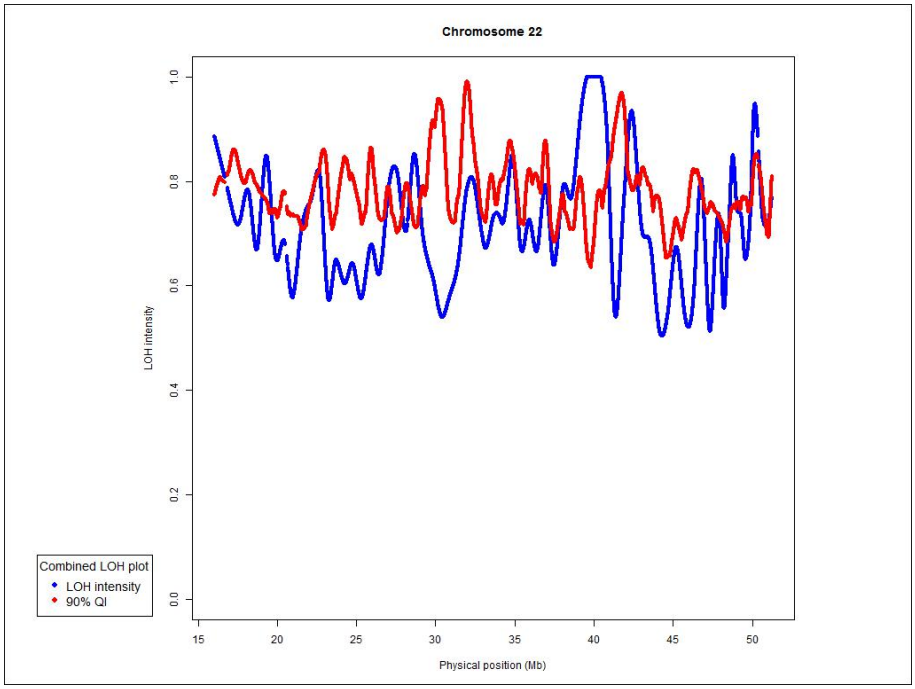


Figure 7-5-3. Combined homozygosity intensity plots of a case (HG01518).

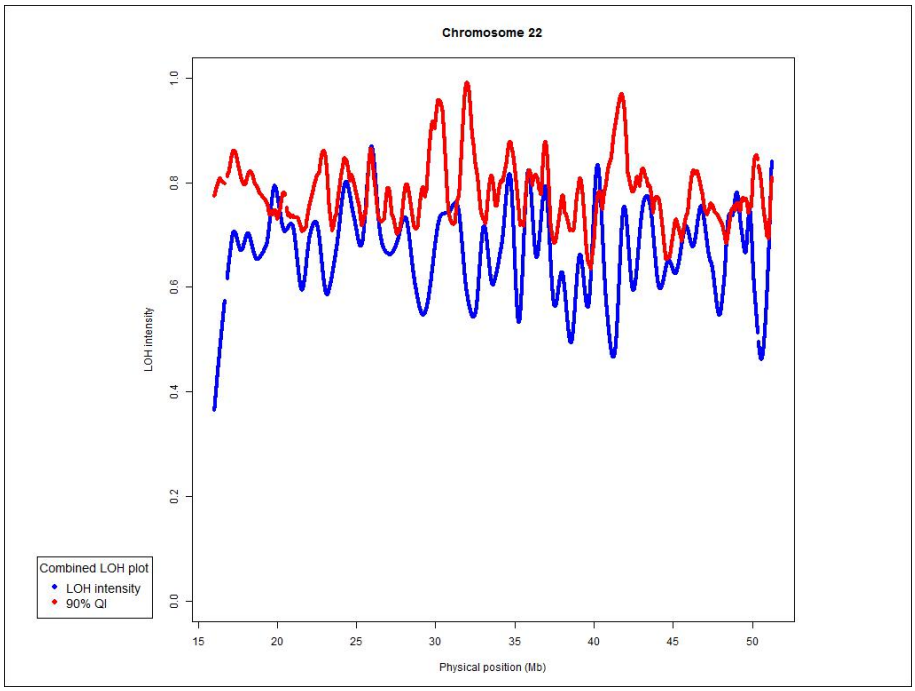


Figure 7-5-4. Combined homozygosity intensity plots of a control (HG01517).

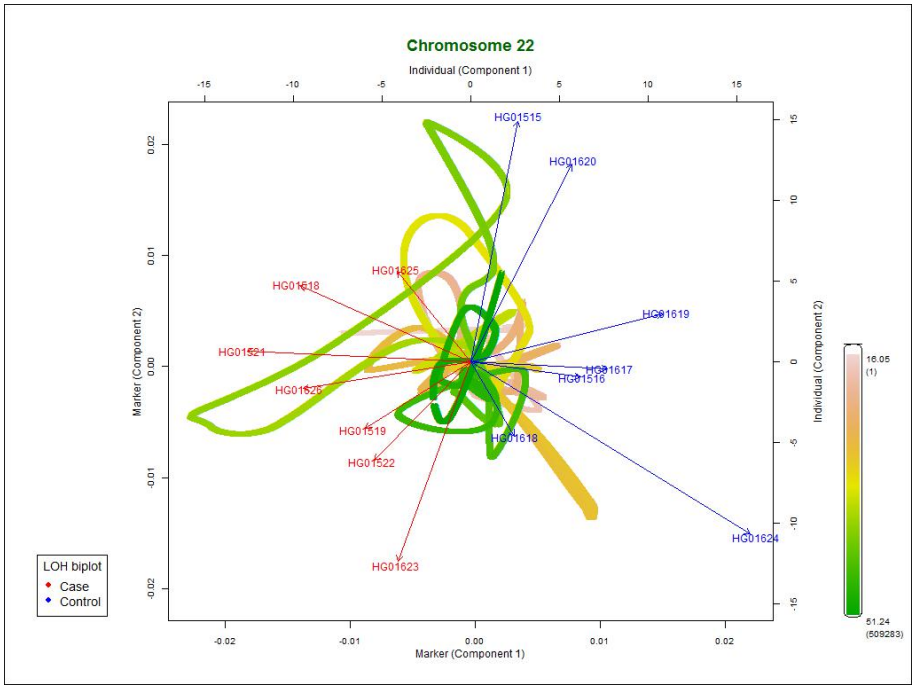


Figure 7-5-5. A biplot drawn for 7 cases (red color) and 7 controls (blue color) in Example 5.

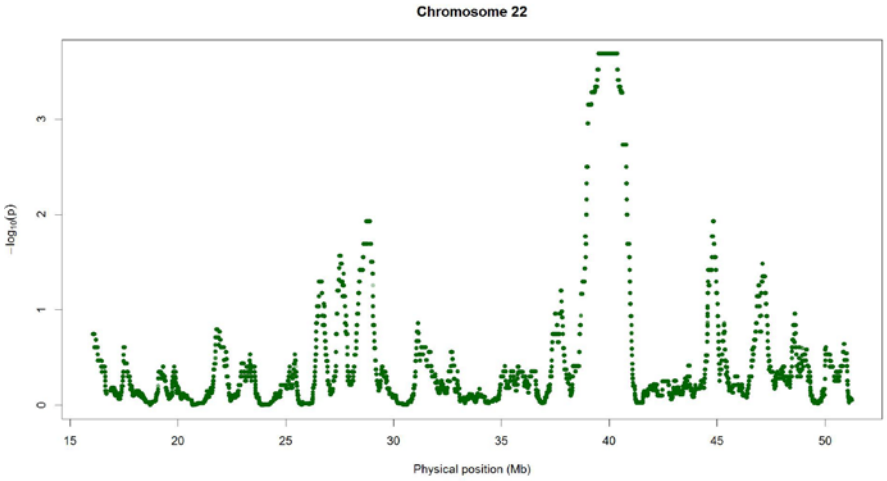


Figure 7-5-6. P-value plot of Wilcoxon homozygosity association test in Example 5.

Numerical results are saved in the directory “Numerical results” which contains four subdirectories (“LOH intensity”, “LOH biplot”, “Chromosome ranking test”, and “Association test”). First, numerical results of homozygosity intensity are saved in Excel files by group and by chromosome in the folder “C:\LOHAS\Output\Example5\LOH intensity\”. Each of the Excel files provides

information including SNP ID (probe set and/or RS number), physical position, homozygosity intensities of samples, and a 90% quantile of homozygosity intensity in a control group. Second, numerical results of LOH biplot are saved in Excel file in the folder “LOH biplot”. The output consists of three parts as illustrated in **Section 7.2**. In this example, the first two components explain 35.46% of total variation. The scores of “individual” components and “marker” components are provided too. Third, numerical results of p-values of chromosome ranking tests are saved in Excel files in the folder “C:\LOHAS\Output\Example5\Chromosome ranking test”.

Finally, numerical results of p-values in Wilcoxon association tests are provided in the last column in Excel files by chromosome in the folder “Association test”. In addition, the Excel file also contains the following information for each anchor SNP of sliding windows: SNP ID (probe set), RS number, physical position and some statistics including mean, median, standard deviation of homozygosity intensity, and proportions of samples which have higher homozygosity intensities than a reference threshold in case group and in control group. In this example, the reference threshold indicates the red curve in the combined homozygosity intensity plot, i.e., the 90% quantile of homozygosity intensity in control group.

8. LOHAS VERSION UPGRADE

Versions:
LOHAS Version 1.0: Jan. 2010
LOHAS Version 1.1: Nov. 2011
LOHAS Version 2.0: Feb. 2013
LOHAS Version 2.1: Nov. 2014
LOHAS Version 2.2: Oct. 2016
LOHAS Version 2.3: Mar. 2017

What are the new features in LOHAS?
In comparison with **LOHAS** v2.2, **LOHAS** v2.3 adds some new analysis function as follows:

1. **LOHAS** adds a parameter setting interface so that users don't need to prepare the file 'Datasetting.txt' by themselves (see Item 3 in **Section 5**).
2. For chromosome ranking test, **LOHAS** calculates KL-distance of individuals to estimated median of HI instead of mean of HI.
3. **LOHAS** provides a new function for users to provide a trait-covariate file to run regression or GEE.
4. **LOHAS** uses R-packages foreach and bigmemory to accelerate the computation and data reading.
5. If an analysis is interrupted by some external reasons (e.g., power down), **LOHAS** can continue the analysis if users DO NOT remove the 'TMP' folder under the output folder.
6. **LOHAS** adds or changes some numerical and graphic results. For example, adding HI summary table, providing Manhattan plots of association results, and making some modifications in biplot.

9. REFERENCES

Yang, H.-C., Lin, H.-C., Huang, M.-C., Li, L.-H., Pan, W.-H., Wu, J.-Y. and Chen, Y.-T. (2010). A new analysis tool for individual-level allele frequency for genomic studies. *BMC Genomics* **11**, 415.

Yang, H.-C., Chang, L.-C., Huggins, R. M., Chen, C.-H. and Mullighan, C. G. (2011). LOHAS: Loss-of-heterozygosity analysis suite. *Genetic Epidemiology* **35**, 247-260.

Yang, H.-C. and Li, H.-W. (2014). Analysis of homozygosity disequilibrium using whole-genome sequencing data. *BMC Proceedings* **8**, S15.