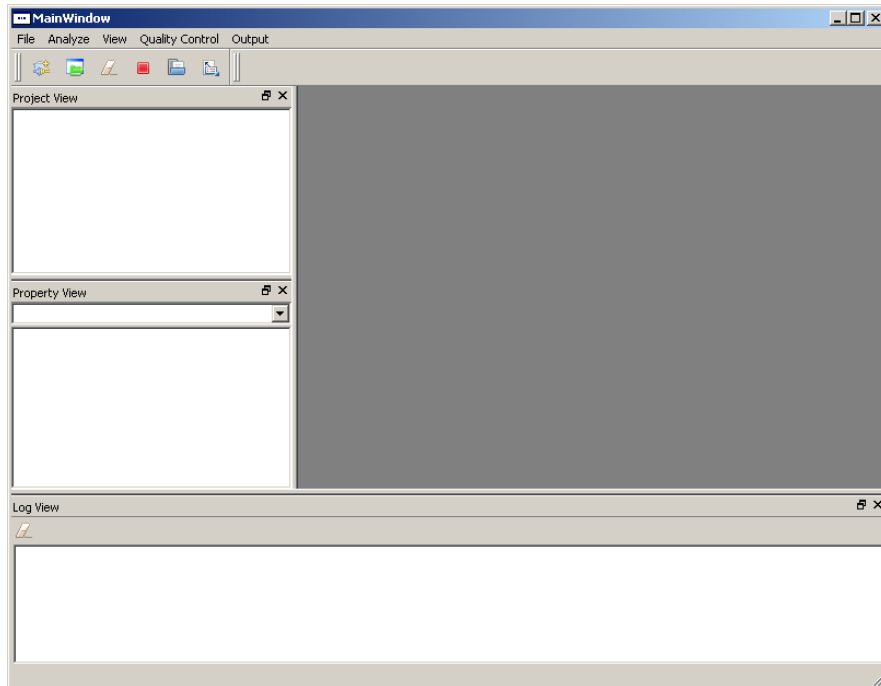


README

After downloaded the program, unzip it to a folder. Double click **GenomeMatrix.exe** the application will start. You will see the window like below.



Genome Matrix may require some runtime library to run. Usually it comes with the operator system. But if the system said you missing some runtime library, you can download it from <http://www.microsoft.com/downloads/details.aspx?familyid=a5c84275-3b97-4ab7-a40d-3802b2af5fc2&displaylang=en>.

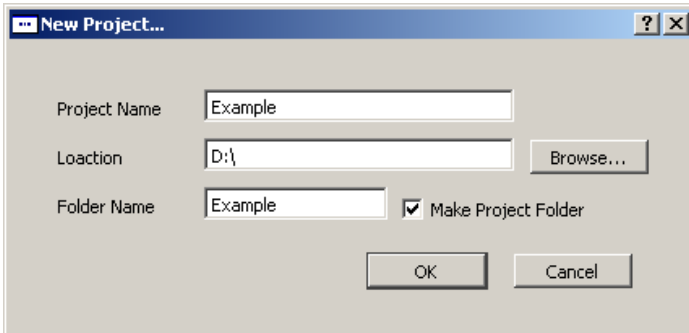


If windows popup an security warning about Unknown Publisher, click run.

1. Project:

Genome Matrix stores every dataset and its results based on project. If you have an exist project, select **File->Open Project** and located to your project file for opening it. To create a new project, select **File->New Project...** and a popup window will show up.

Input your project name, and specify where to put the project. If you want, Genome Matrix can make a folder for the project.



After create a new project, the Data Import Wizard will appear.

2. Data

Genome Matrix supports most of plain text data. For most cases you will need to provide the Case-Control information and SNP RS number. We use RS number to search in database which Gene or Pathway the SNP is belonging to.

Normally you will have two files to import. One main data file and the description of the data file. And there should be a header line for each file.

You must choose the type of your data file. In the Data Import Wizard, Row for individual means each row in your main data presents an individual. It looks like:

ID	CaseControl	OtherInformations	SNPS
IND1	1	...	AA AT TT AA
IND2	2	...	AA TT AT TT
...			

And the description file should contain the SNP rs numbers.

SNPID	OtherInformations
rs1	...
rs2	...
...	

The column for individual means each row in your main data file presents an SNP and each 2 columns is an individual. It looks like:

SNPID	SNPS
rs1	AA AT TT AA
rs2	AA TT AT TT
...	

And the description file should contain the Individual Case-Control Information.

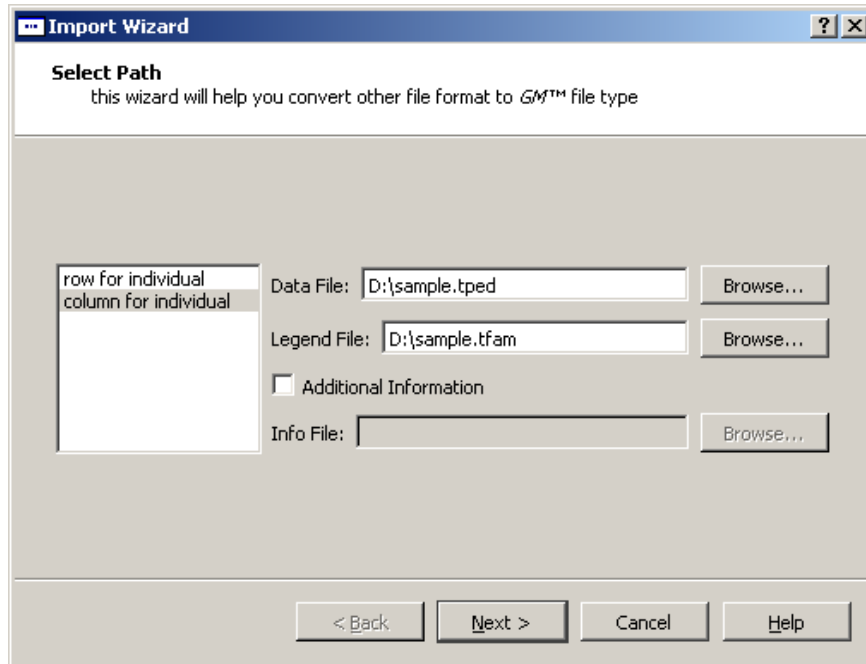
ID	CaseControl	OtherInformations
IND1	1	...

IND2 2

...

...

You can choose the method according to your data type. And use browse or input your data file names manually. If the data files exist and a method have been choose, select **next**.



In Wizard Step 2, you can define the meaning of each field.

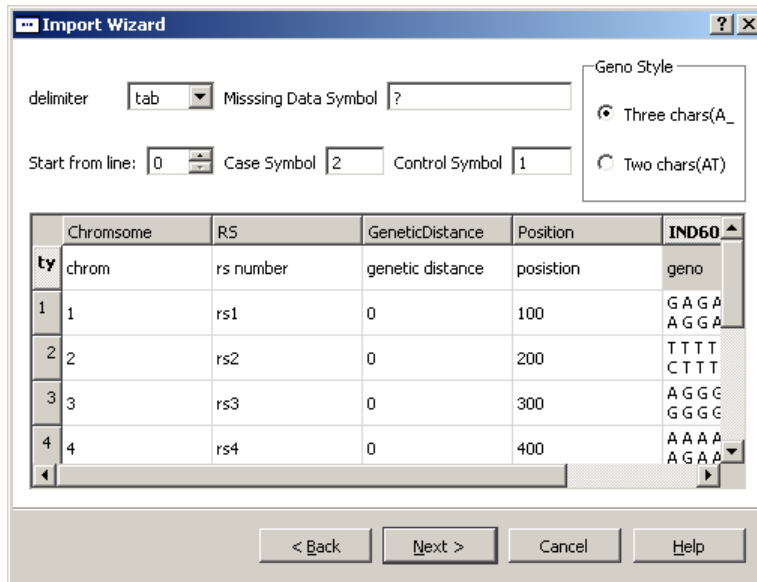
First choose your delimiter of your file. When you have chosen the right delimiter for the file, import wizard will automatically generate the preview of file. The **Missing Data Symbol** field you can specify the missing symbol of the main data file, the same as **Case Symbol** and **Control Symbol**.

If your genotype data is presented by three characters (like A A, A_T, A,A), select three chars in **Geno Style** box. Otherwise (like AA), choose two chars.

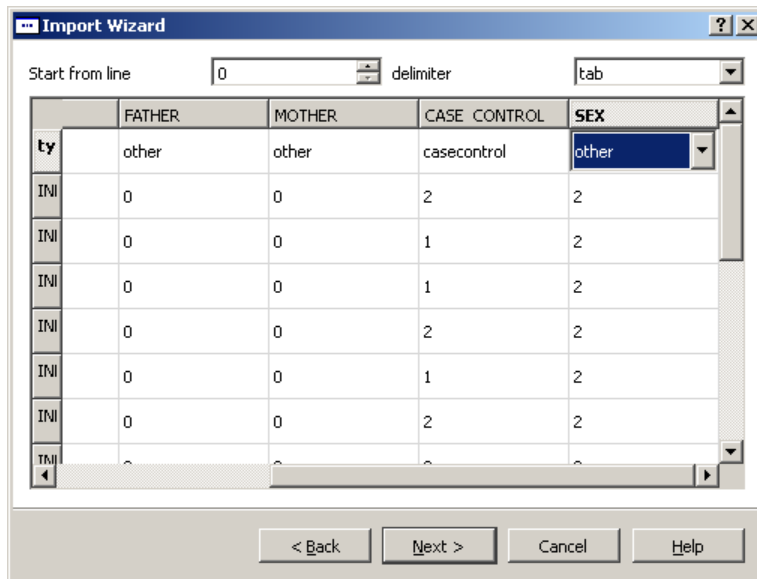
You can set the number of lines you want skip in **Start from line** field if there are several comment lines before the data.

After that, double click on each field in the first row. The types of row are different according to your file type: including genotype, chromosome, rs number, position, genetic distance, case-control and other. The values in chromosome type should be an integer indicates which chromosome the SNP is in. If not provided, the SNP chromosome information will automatically load from database. It is same for the position and genetic distance types. The rs number type accepts inputs with RS and an integer, like rs1000 or RS1000. If value in case-control type matches Case Symbol or Control Symbol you have defined on last step, the individual will be consider as a case or control.

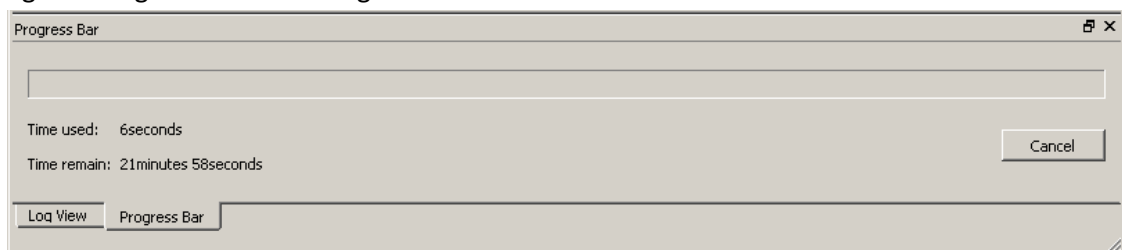
Except genotype, each column's type must be specified. If its type isn't in the list, choose other. The program will make all the left columns' type as genotype by default. Genome Matrix do require rs number and case-control information, be sure you have import them correctly.

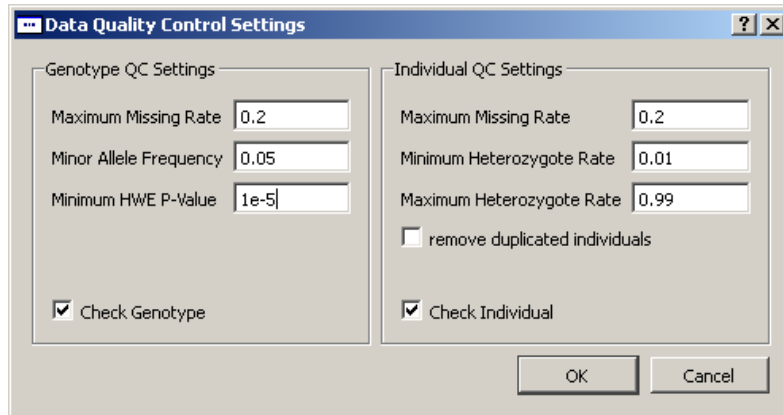


After you finished import the first file, do the same process on the second file.



Press finish to start the converting process. By click Progress Bar at bottom, you can view the progress. Large file will take long time to convert.

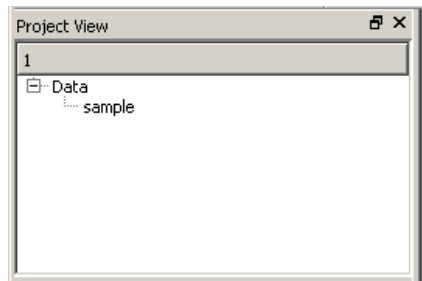




When the converting is finished, you can define the data quality control parameters. If a SNP's missing proportion is more than **Maximum Missing Rate**, its minor allele frequency is below the **Minor Allele Frequency** or its Hardy-Weinberg Equation P-Value is less than **Minimum HWE P-Value**, it will not be used in analyze. To take all the SNPs into computation, do not mark **Check Genotype**.

For individual, if its total missing SNPs proportion is more than **Maximum Missing Rate** or its heterozygote rate is not between **Minimum Heterozygote Rate** and **Maximum Heterozygote Rate**, it will not take count into analyze. To make all the individuals into computation, do not mark **Check Individual**.

Now you have a working data copy. Next time you want to re-open and analyze this data, just click Open Project and double click the data file.



3. Pathway Analyze

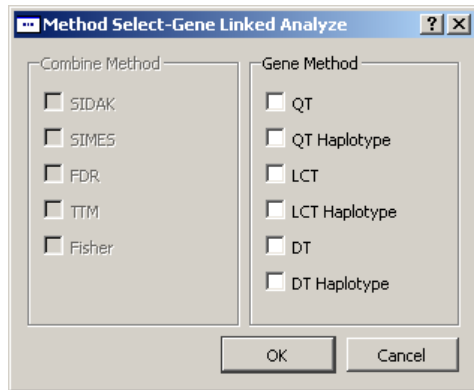
After converted or opened a data, you can do the pathway test. Click **Analyze->Pathway Test**.

We implements Gene Unlinked analyze and Pathway Unlinked analyze according to the description in our paper.

Gene Linked:

There are three major method in Gene Linked analyze. The first suggested statistic is to take a linear combination of P-values for all SNPs within the gene, which is referred to as the LCT. QT test is based on the quadratic form according to our paper. Another way to combine dependent P-values is that we first transform dependent variables into independent variables and then combine independent variables. All the methods for combining independent P-values can be applied to the transformed variables, such as Fisher's combination test or Sidak, Simes, false discovery rate (FDR) method.

To do Gene Linked analyze in Genome Matrix, click **Analyze->Pathway Test->Gene Linked** and a method select window will popup.



Choose by which method you want to analyze the dataset. Details about the methods you can refer to our paper.