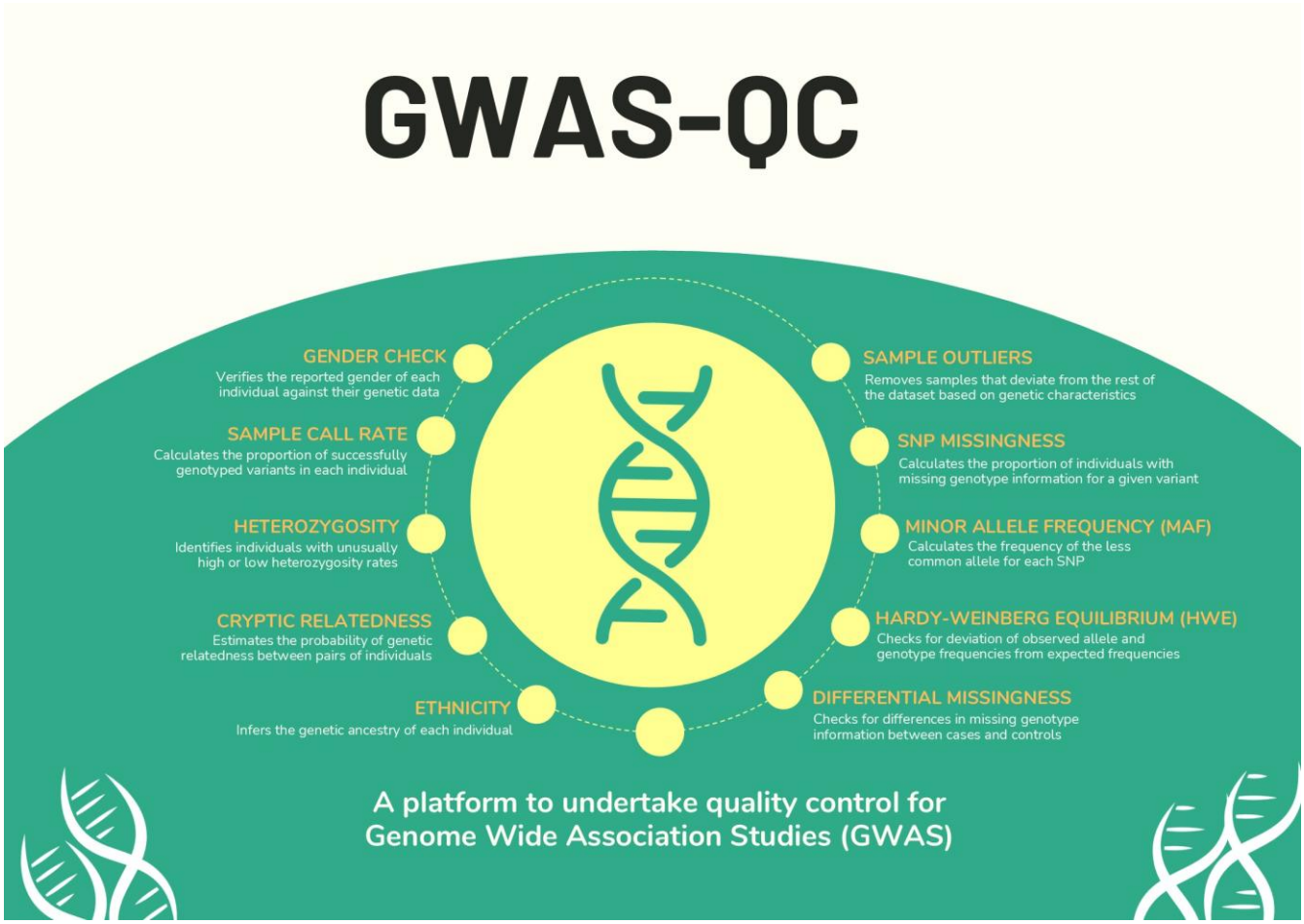


USER MANUAL

Authors
Alexandar V[†], J. Eunice Zhang[†], Prof. SIR Munir Pirmohamed*

University of Liverpool

Table of Contents

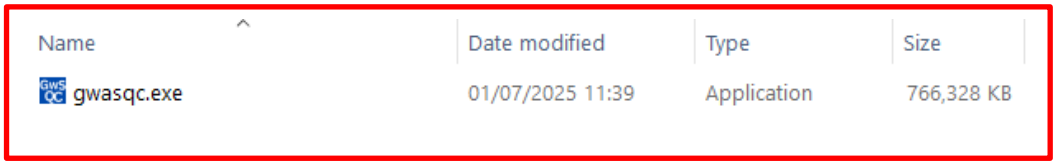
GETTING STARTED

Installation

Compatible with Windows (PC/laptop) or Mac (virtual machine)

***Do NOT install on network drives**

1. Download the GWASQC software.
2. Double click the downloaded “gwasqc.exe” file for installation

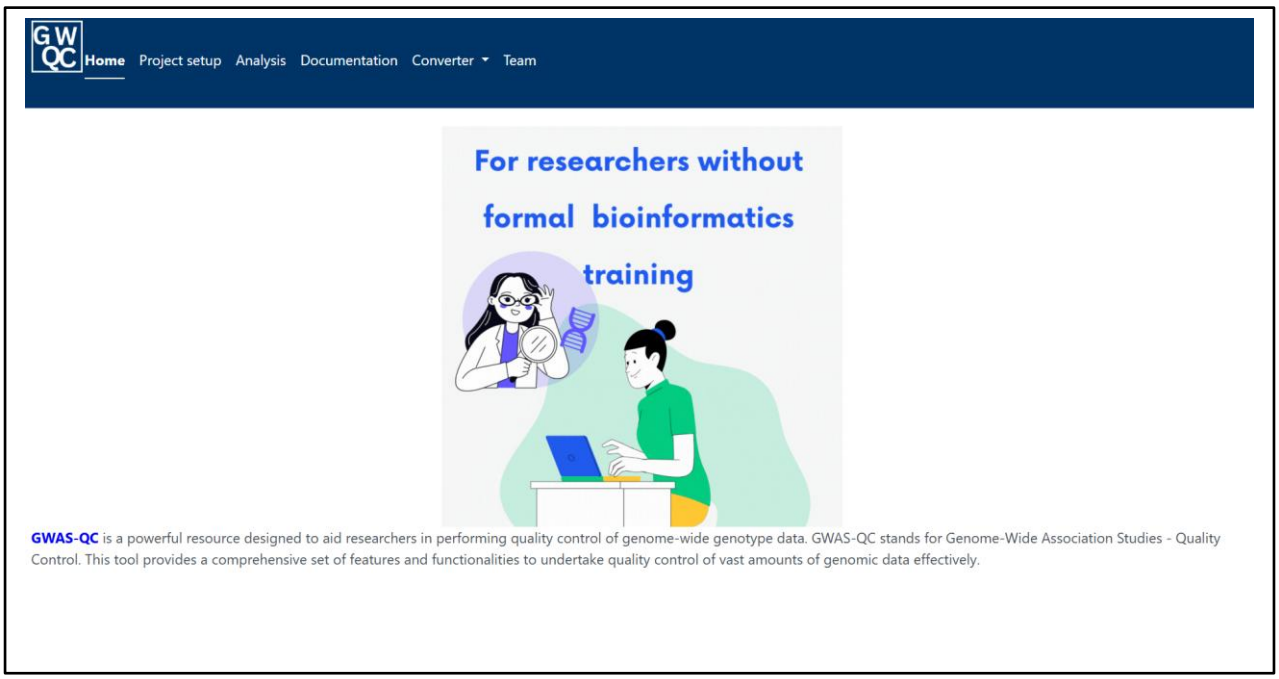


Watch the installation video for detailed instructions.

Run software

After installation, double click the “GWASQC” **desktop** icon.

The software will open in a new window.



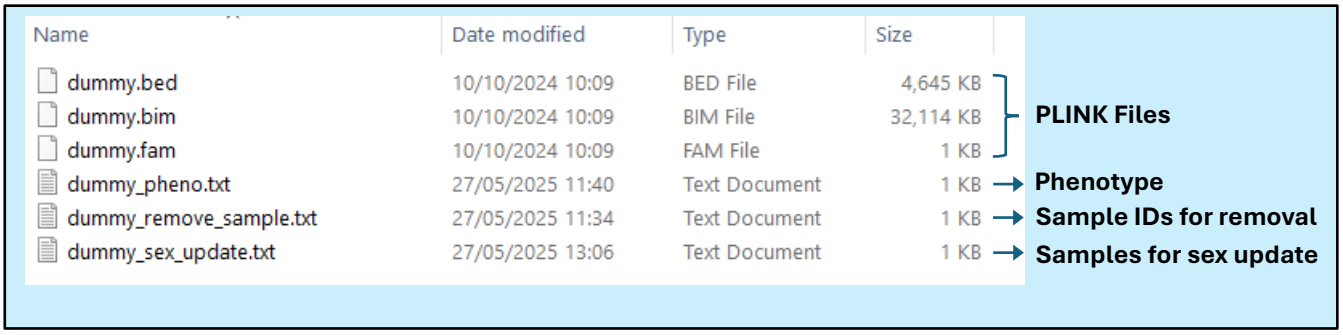
Input genotype file

The software only accepts genotype data in binary PLINK format: BED, BIM and FAM files.

MAP/PED or VCF files can be converted to binary PLINK format using the “Converter” function (see page 7).

Dummy data

A dummy dataset is supplied which users can download and use to explore the application.



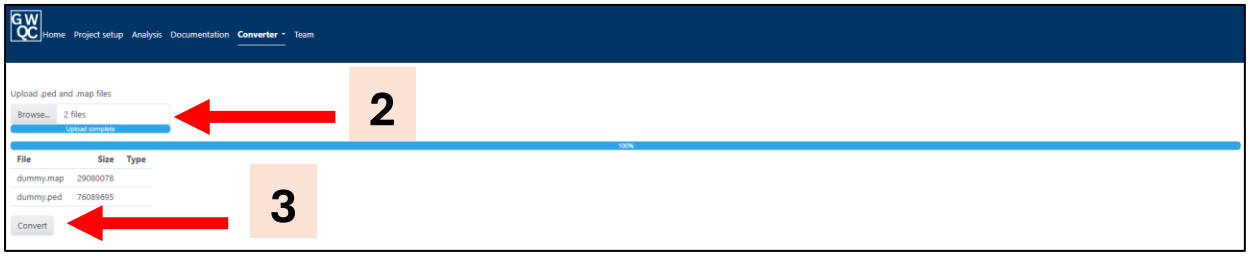
FILE CONVERSION

Converter

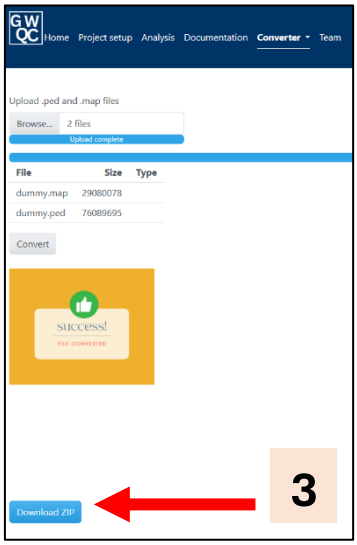
1. Use the “Converter” to convert MAP/PED or VCF files to binary PLINK format.

2. Browse to upload files.

3. Click “Convert”.



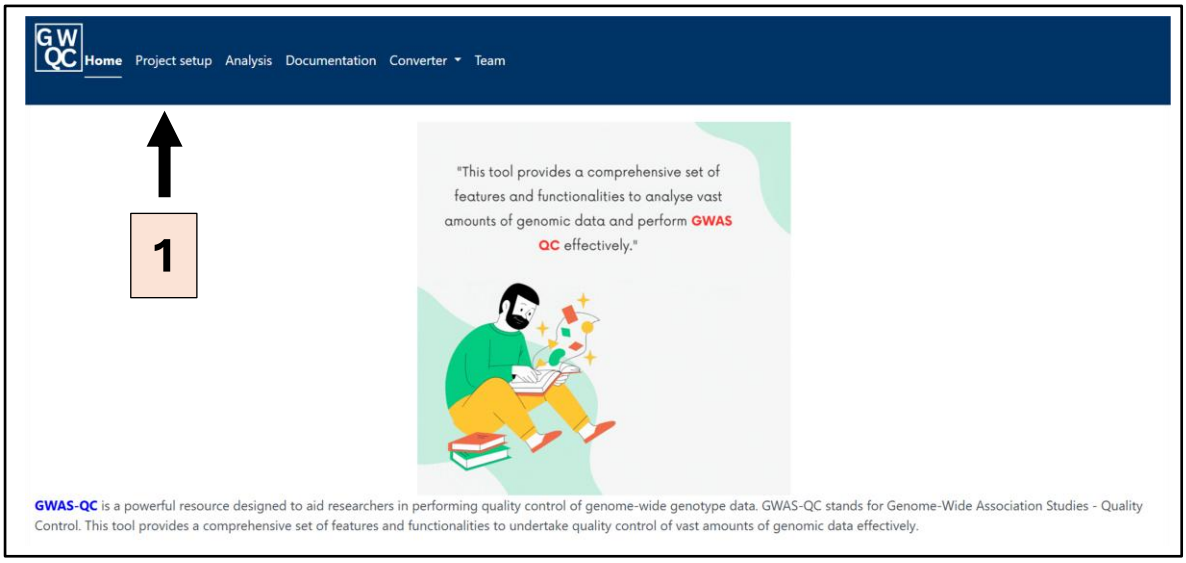
4. Upon successful file conversion, the binary PLINK files will be available for download.



SAMPLE-QC (SQC)

Project setup

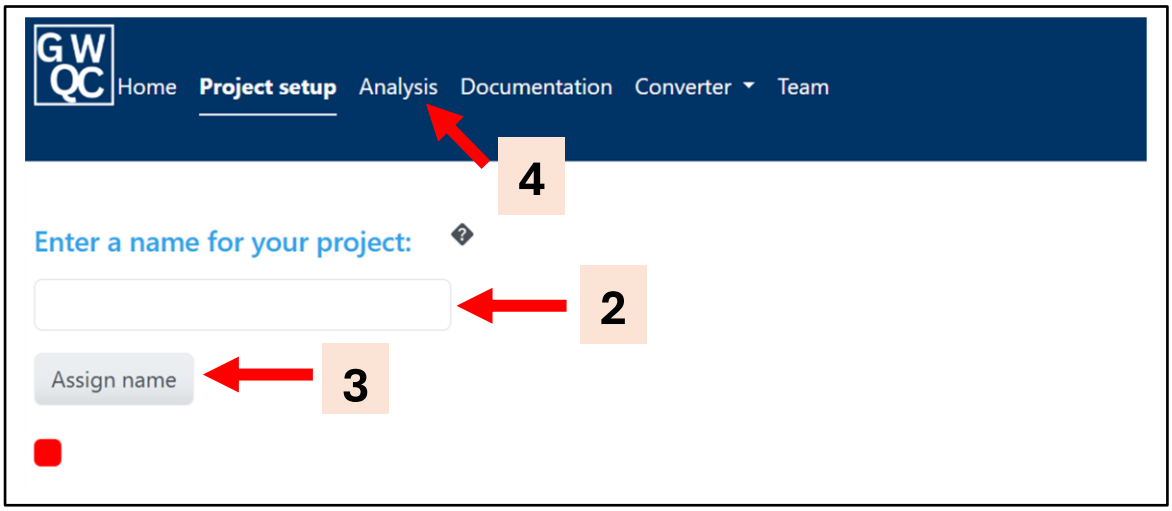
1. Click “Project setup”



2. Enter a name for your project, *preferably less than 5 characters consisting of alphabets and numeric only.* **Note: No special characters and/or whitespace allowed.**

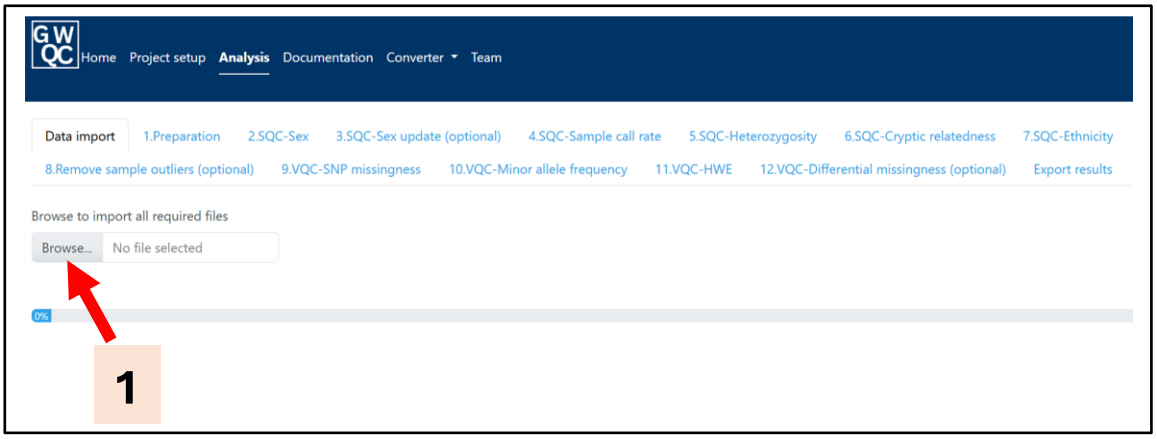
3. Click “Assign name” button.

4. Go to “Analysis” tab when you are prompted to proceed.



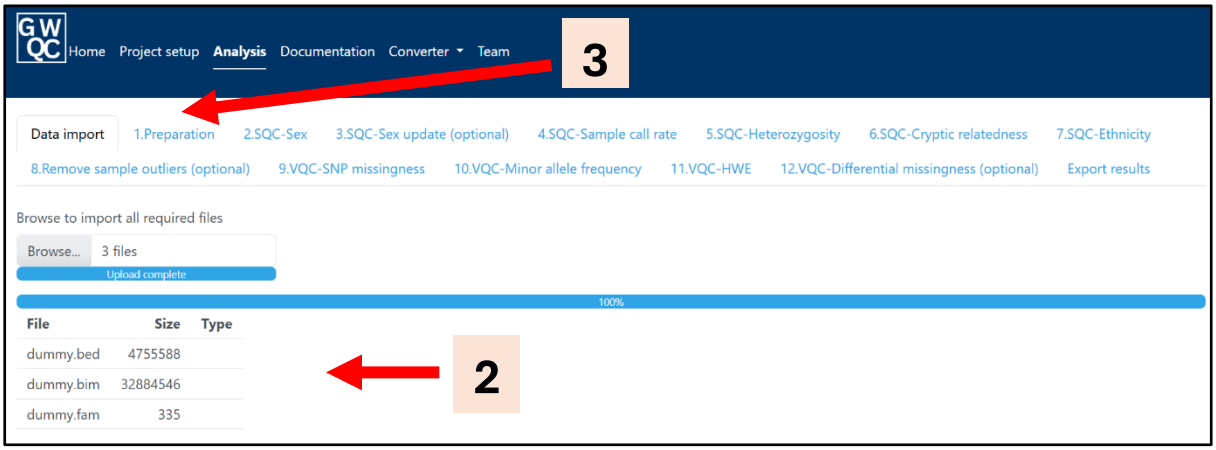
Data import

1. Click “Browse” button to import your files to current project. Multiple files can be imported at once.



2. Upon completion, details of imported files will be displayed.

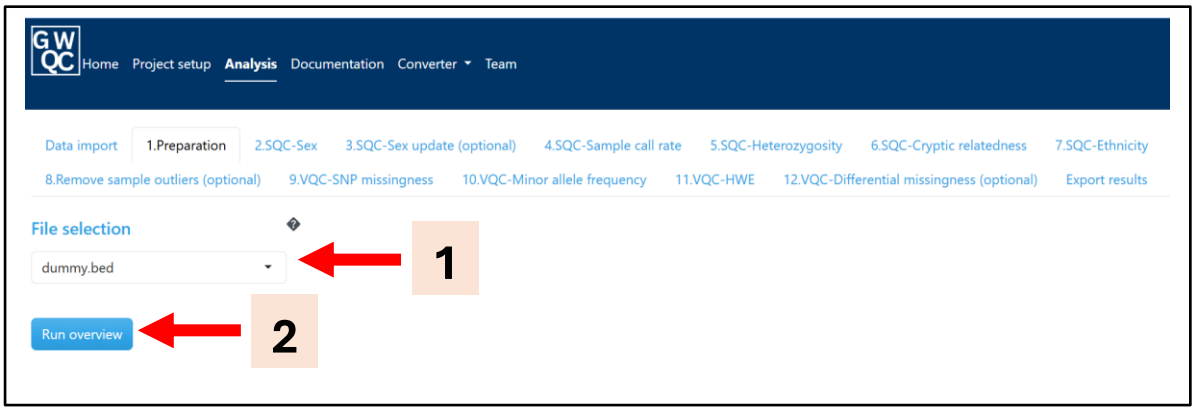
3. Click the “1.Preparation” tab when all the required files have been imported successfully.



1. Preparation

1. Select the project BED file you have uploaded.

2. Click the “Run overview” button.

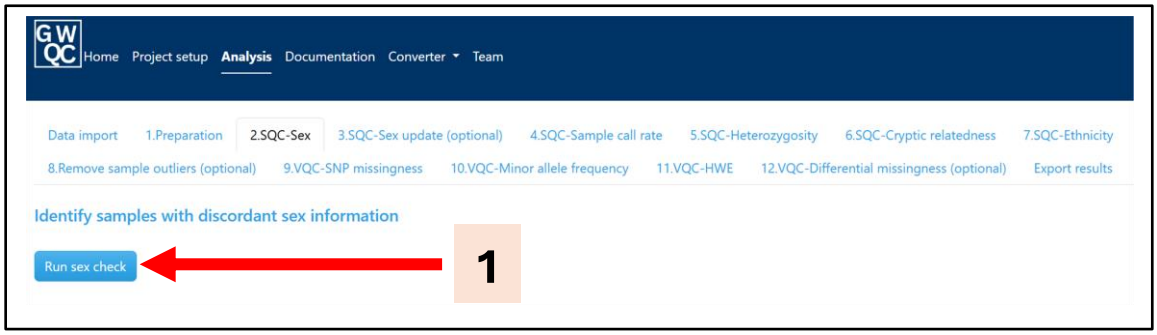


3. Wait for completion and the total number of variants and samples loaded will be displayed.

4. Proceed to “2.SQC-Sex” tab.

2. SQC-Sex

1. Click "Run sex check" button to check for sex concordance between reported and genetic sex.



2. The overall number of individuals with discordant sex information in your sample set will be displayed in a histogram.

3. Information of individuals with discordant sex will be displayed in a table.

   F is the F statistic of sex estimated from X-chromosome inbreeding coefficients where F > 0.8 is Male and F < 0.2 is Female. Male is coded as 1; Female is coded as 2.

4. List of individuals with discordant sex information can be downloaded using the Copy, CSV, Excel, PDF, or Print function.

5. Proceed to either "3.SQC-Update sex (optional)" tab or "4.SQC-Sample call rate" tab.

3.SQC-Update sex (optional)

1. Upload a tab-delimited text file containing Indexes, IDs and updated sex information (**NO** headers required). A dummy sample file is provided in the dummy data folder (“*dummy_sex_update.txt*”). Create your file in the format as follows:

Download the list of individuals with discordant sex information in CSV format (see step 4 in 2.SQC-Sex on page 12).

	A	B	C	D	E	F	G	H
1		INDEX	ID	SEX	GENETIC SEX	STATUS	F	
2	1	126	N573	2	1	PROBLEM	0.9896	
3								
4								

IMPORTANT

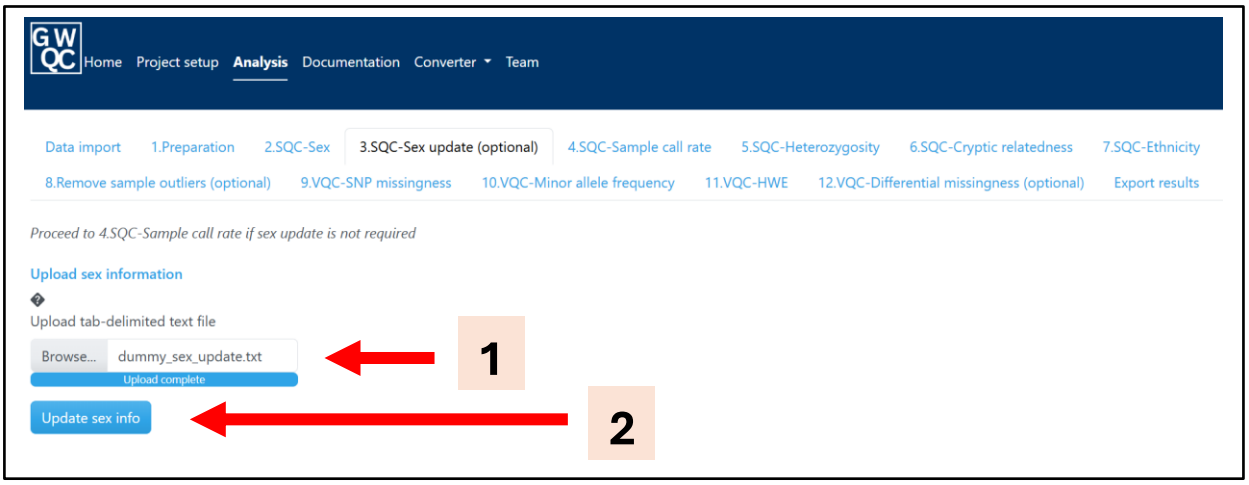
Copy columns B, C, and D into a new excel file and save file as a tab-delimited text file with the following structure with no headers

File format

	A	B	C
1	126	N573	1
2			
3			

INDEX **ID** **Updated Sex (1=Male, 2=Female)**

2. Click “Update sex info” button after file upload is complete.

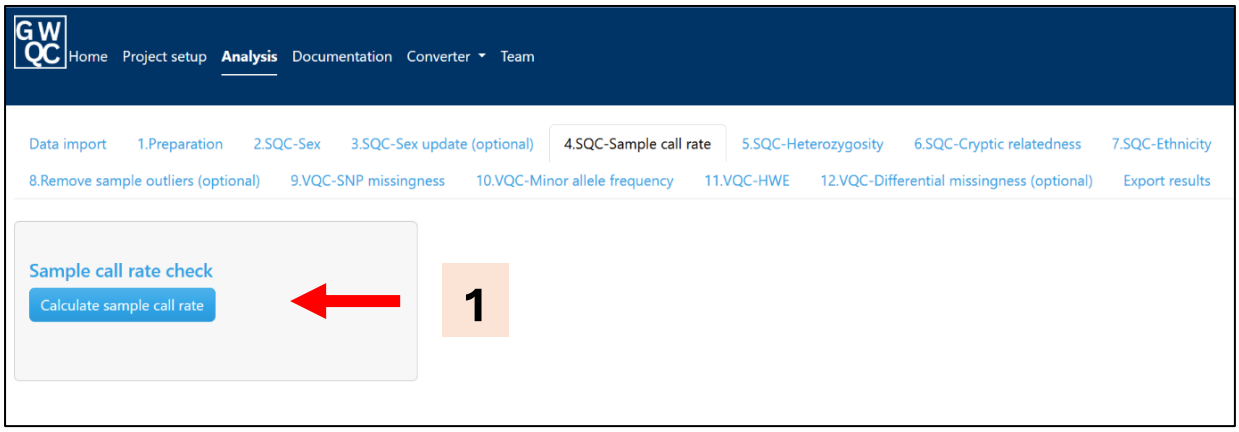


To continue sample QC using files containing updated sex information:

A. Proceed to “Export results” tab to download binary PLINK files with updated sex information (see page 27).
B. Start from “Project setup” to set up a new project (see page 9).
C. Proceed to “Analysis” to upload the updated binary PLINK files with prefix “Sex_uptd” under “Data Import” (see page 10).
D. Proceed to “1.Preparation” tab, select “Sex_uptd.bed”, and click “Run overview” (see page 11).

4.SQC-Sample call rate

1. Click “Calculate sample call rate” button and wait for results.

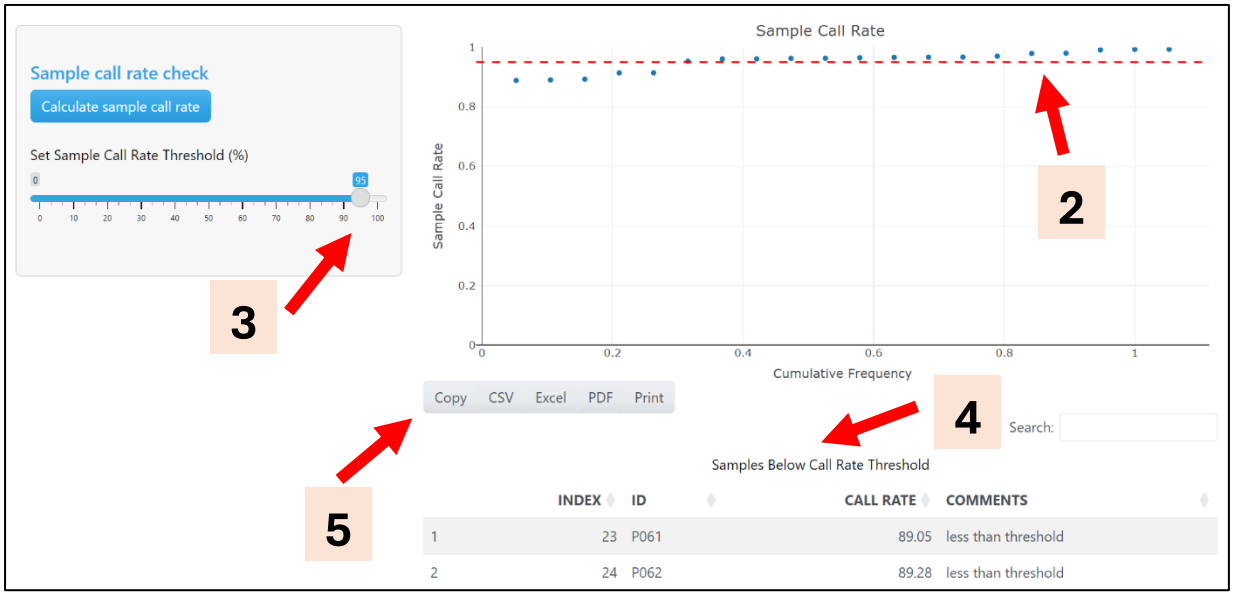


2. An interactive plot displays the sample call rate across individuals, with a red dashed line indicating 95% call rate by default.

3. Move the slider to change the sample call rate threshold and the interactive plot will be updated accordingly.

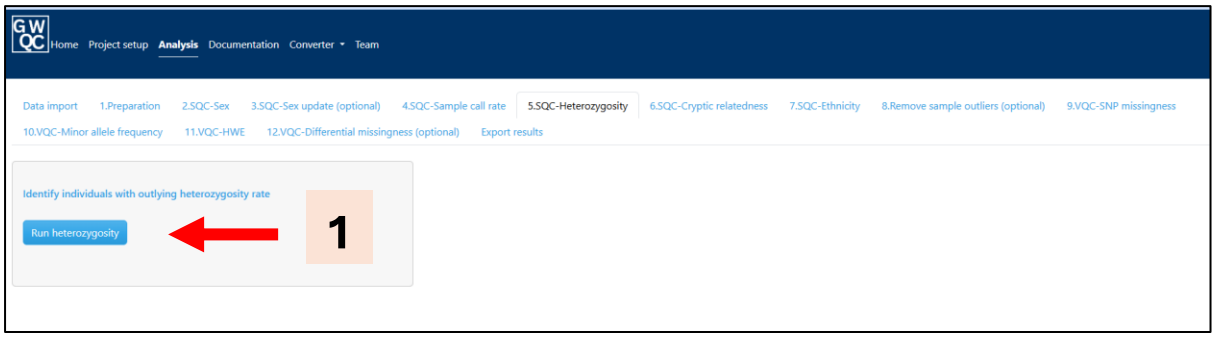
4. IDs of samples below the set call rate threshold are displayed in the table.

5. This list of outliers can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



5.SQC-Heterozygosity

1. Click “Run heterozygosity” button and wait for results.

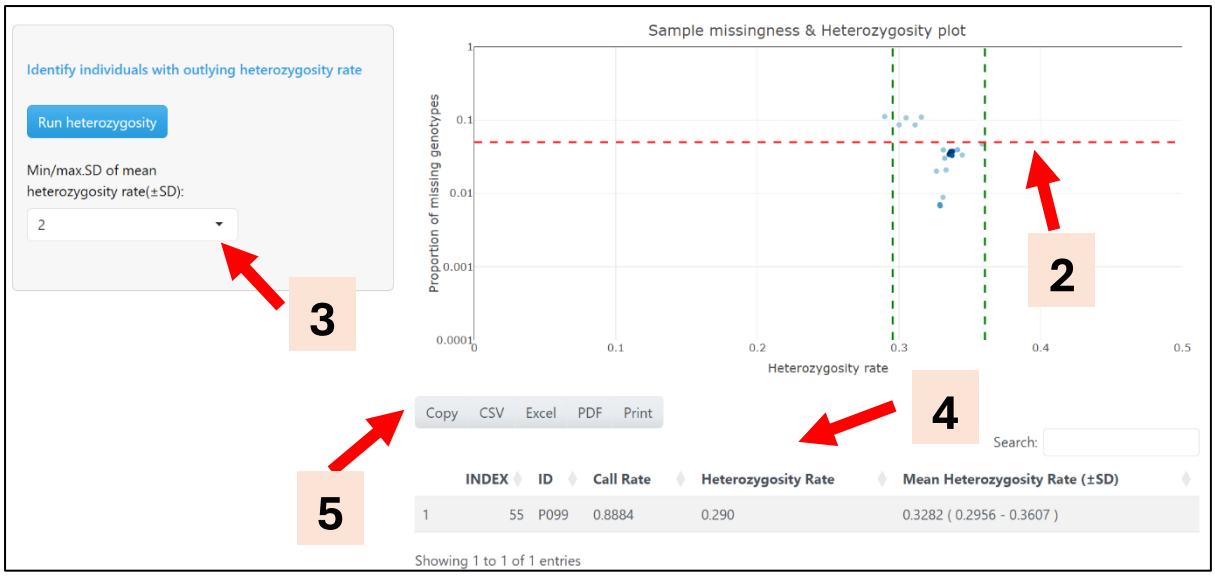


2. By default, samples with ± 2 standard deviation (SD) of mean heterozygosity rate and the chosen call rate from “4.SQC-Sample call rate” are displayed in an interactive plot.

3. Click on the drop down to change the ±SD of mean heterozygosity rate and the interactive plot will be updated accordingly.

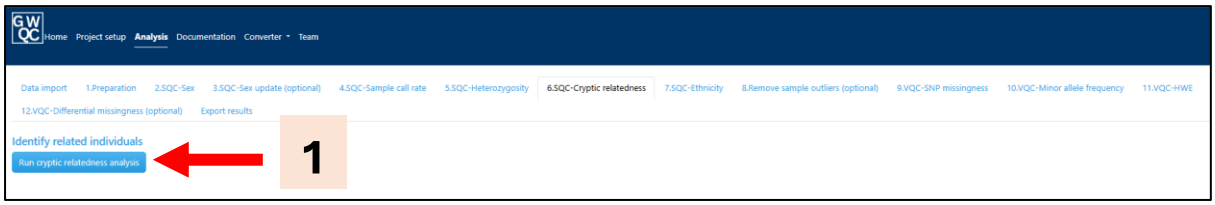
4. IDs of samples beyond the selected heterozygosity rate and call rate are displayed in the table.

5. This list of outliers can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



6.SQC-Cryptic relatedness

1. Click “Run cryptic relatedness analysis” button and wait for results.



2. A histogram showing cryptic relatedness across all samples will be displayed.

3. The recommended IBD threshold.....explain what is IBD.

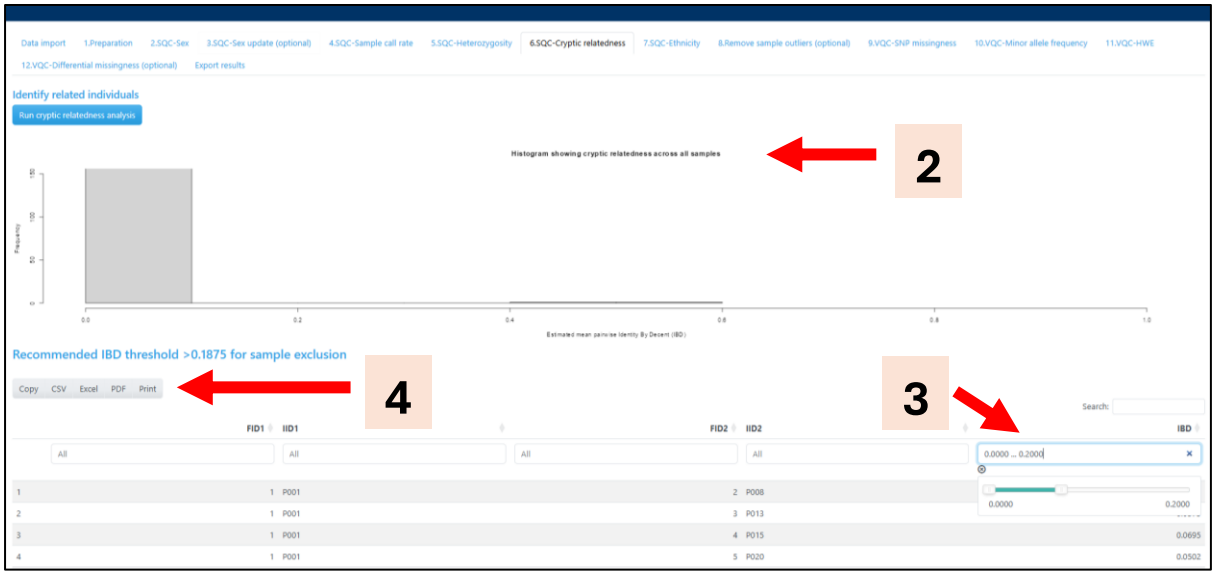
   IBD = 1 duplicates or monozygotic twins,

   IBD = 0.5 first-degree relatives,

   IBD = 0.25 for second-degree relatives

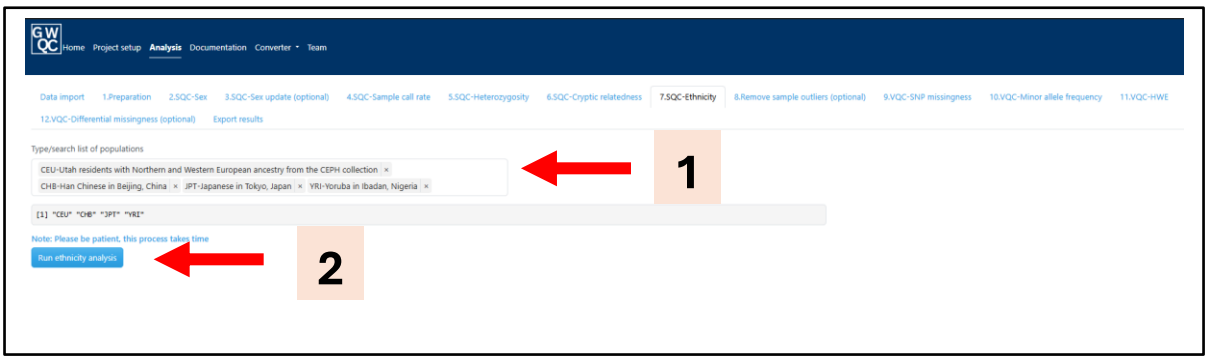
   IBD = 0.125 for third-degree relatives

4. List of individuals beyond the set heterozygosity rate and call rate can be downloaded using the Copy, CSV, Excel, PDF, or Print function.

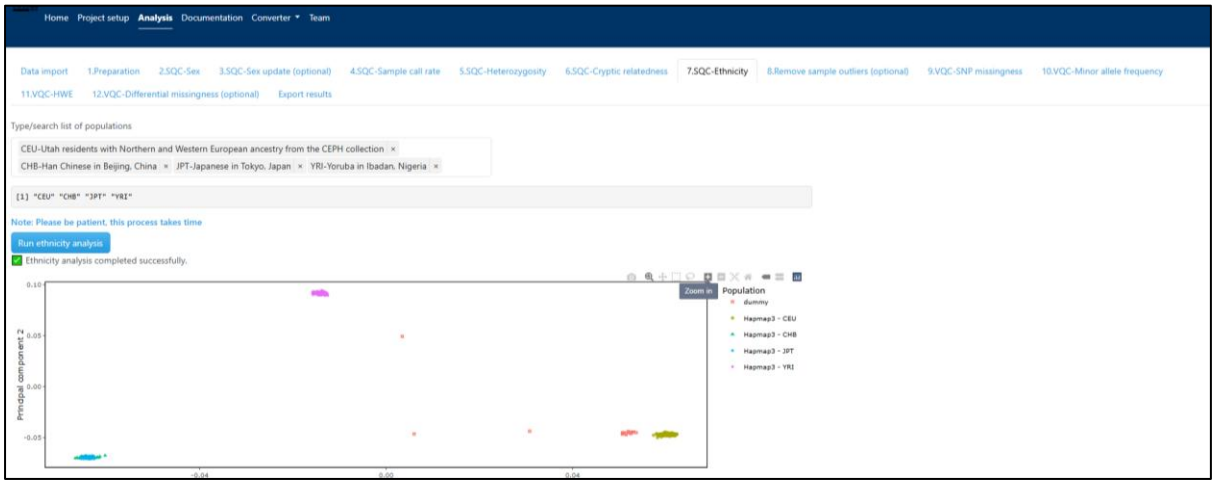


7.SQC-Ethnicity

1. Search and select sub-population of interest in the text box. ***Minimum three populations required***.



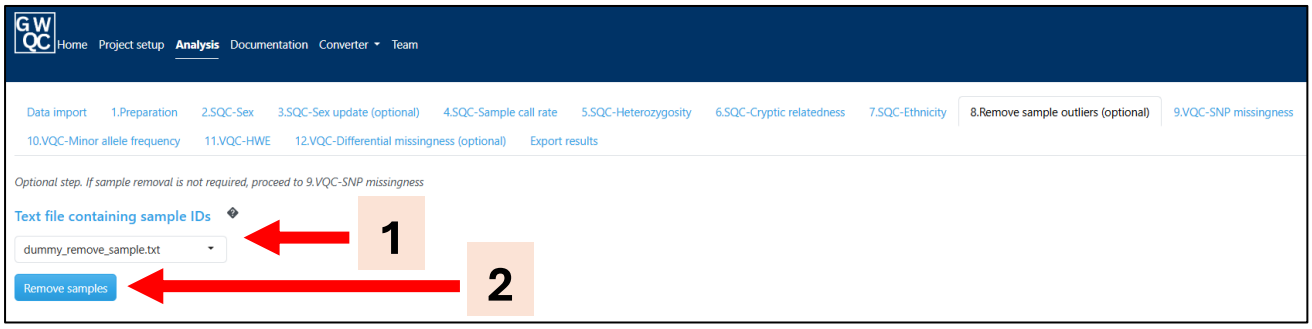
2. Click "Run Ethnicity Analysis" button. This process takes much longer, wait until results appear.



This is an interactive plot. Save, zoom in/out.

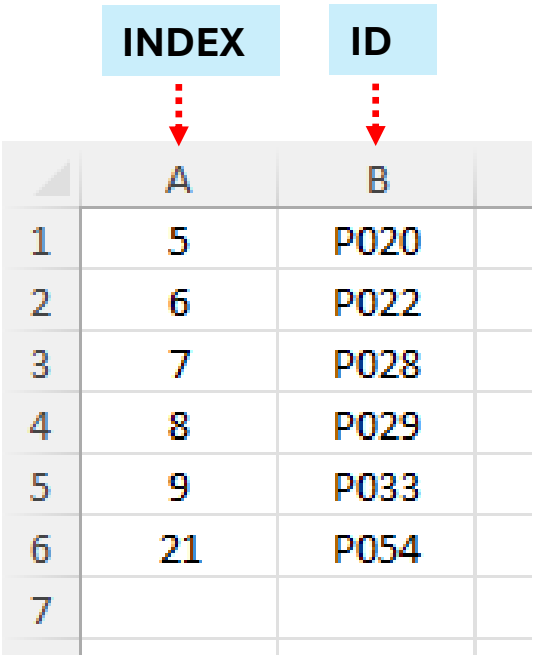
SAMPLE REMOVAL

8.Remove sample outliers (optional)



1. Upload a tab-delimited text file containing Indexes and IDs (**NO** headers required). A dummy sample file is provided in the dummy data folder (“*dummy_remove_sample.txt*”). Create your list in the format given below:

File format

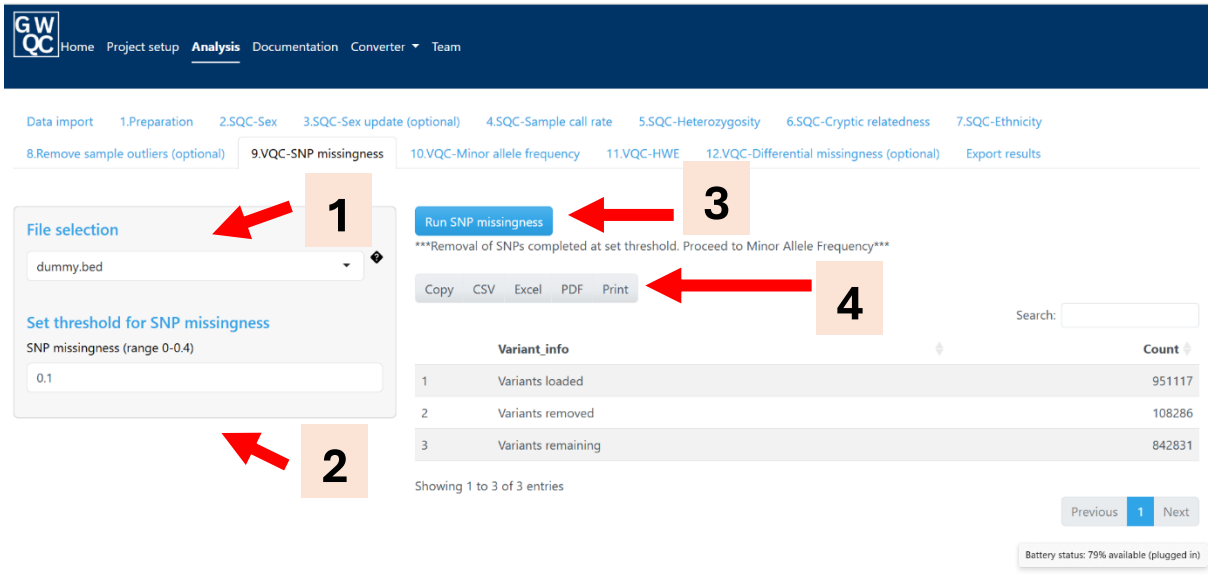


2. Click the “Remove samples” button and wait for the process to complete. A summary table will be displayed detailing number of samples removed. The output files will be named with prefix “***updated_sampleQC***”.

VARIANT-QC (VQC)

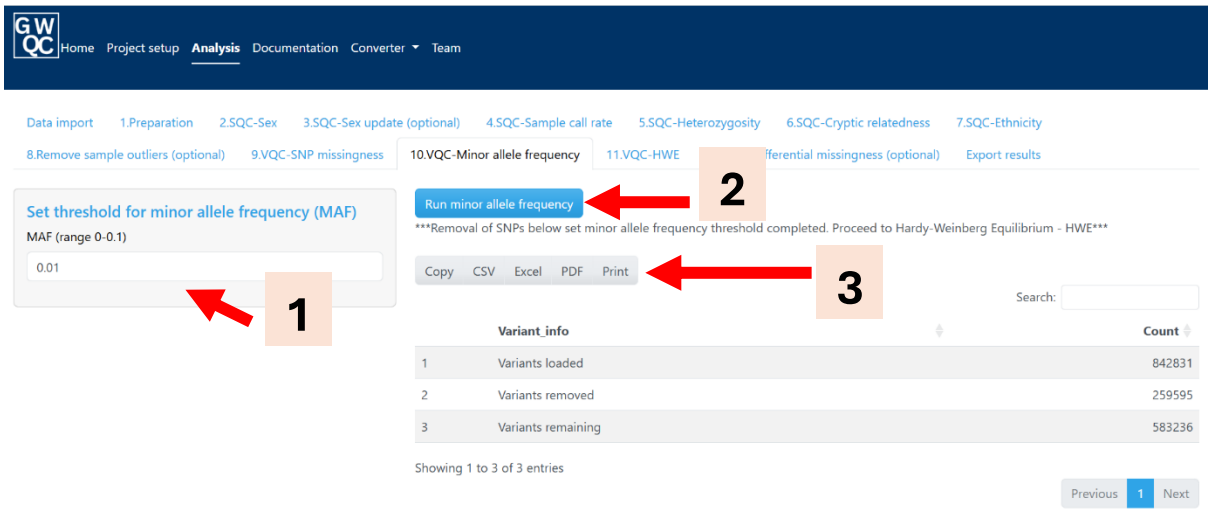
9.VQC-SNP missingness

1. Select required BED file. If you have removed sample outliers, select BED file with prefix "***updated_sampleQC***".
2. Set SNP missingness threshold.
3. Click "Run SNP Missingness".
4. Summary list of variants removed can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



10.SQC-Minor allele frequency

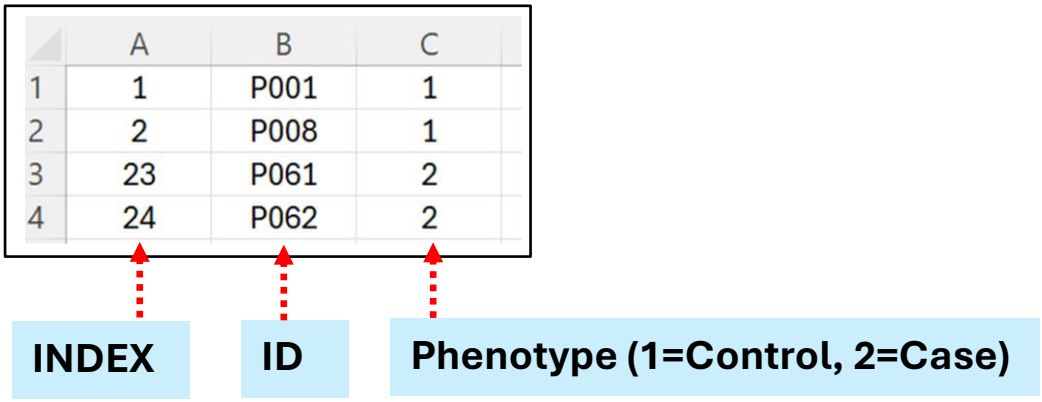
1. Set minor allele frequency (maf) threshold.
2. Click "Run minor allele frequency”.
3. Summary list of variants removed can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



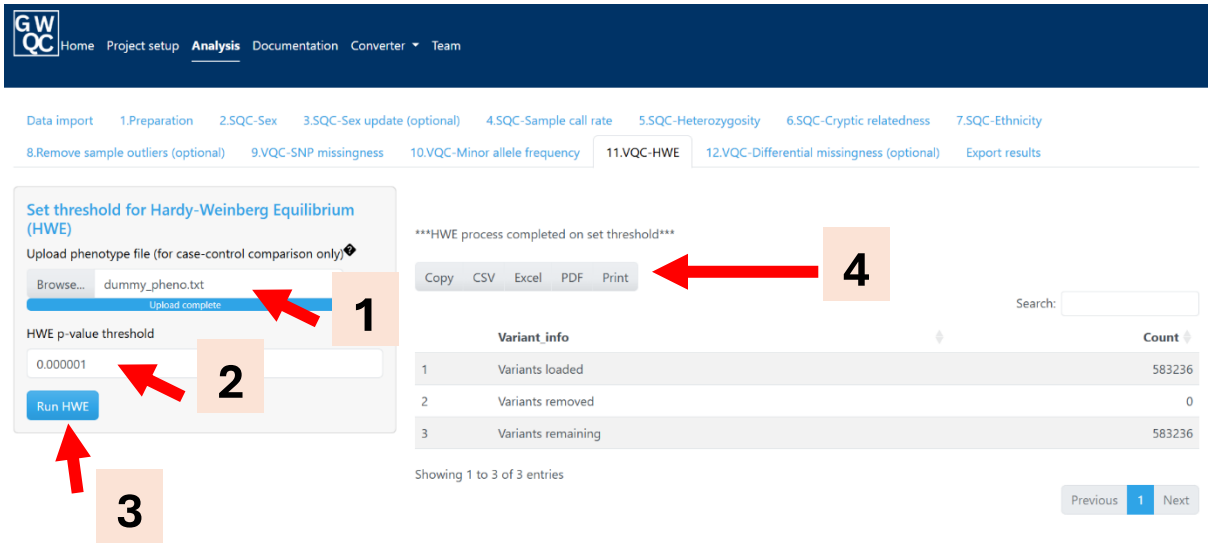
11.SQC-HWE

1. If undertaking a case-control comparison, upload a phenotype file.

Phenotype file should be a tab-delimited text file containing Indexes, IDs and phenotype information (**NO** headers required). A dummy sample file is provided in the dummy data folder (“*dummy_pheno.txt”*). Create your file in the format as follows:



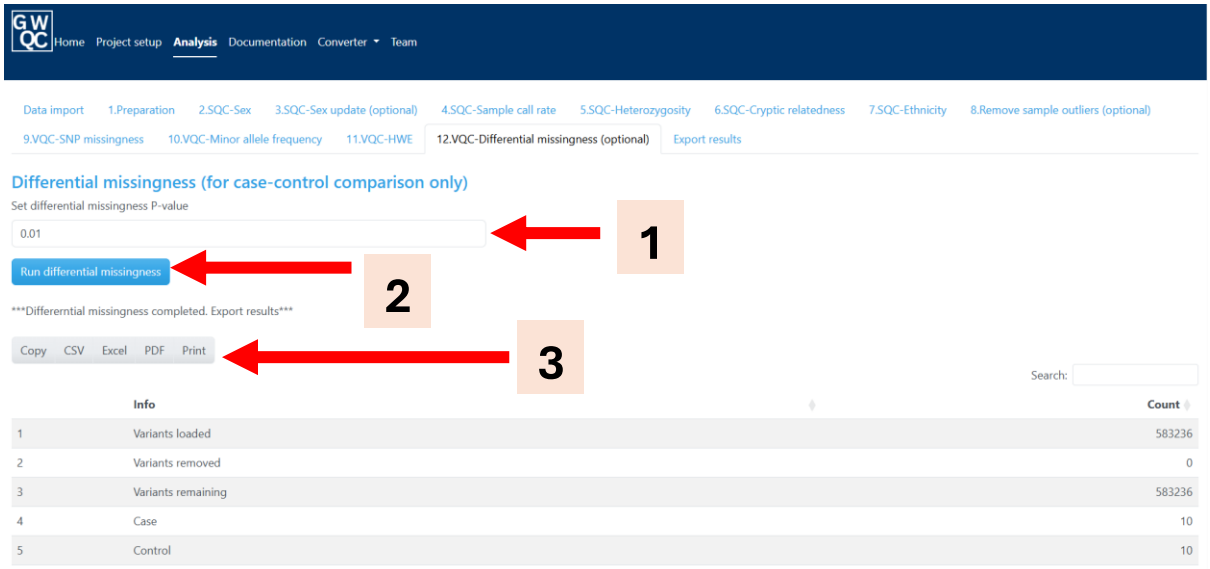
2. Set Hardy-Weinberg Equilibrium (HWE) p-value threshold.
3. Click "Run HWE”.
4. Summary list of variants removed can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



12.SQC-Differential missingness (optional)

This step is only required if you are conducting a case-control comparison

1. Set differential missingness p-value threshold.
2. Click "Run differential missingness".
3. Summary list of variants removed can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



DATA EXPORT

Export results

1. To download updated binary PLINK files, proceed to “Export results” tab.

2. Right click on the output folder you have created on your computer and select “Copy as path”, then paste it here. Make sure you change the backslash to forward slash.

3. Under “Select Category”, click on the drop down and select file required.

4. File status shows you details of the files associated with the category you have selected in step 3.

5. Click “Save ZIP Here”. All files will be saved in the folder you have specified in step 2.