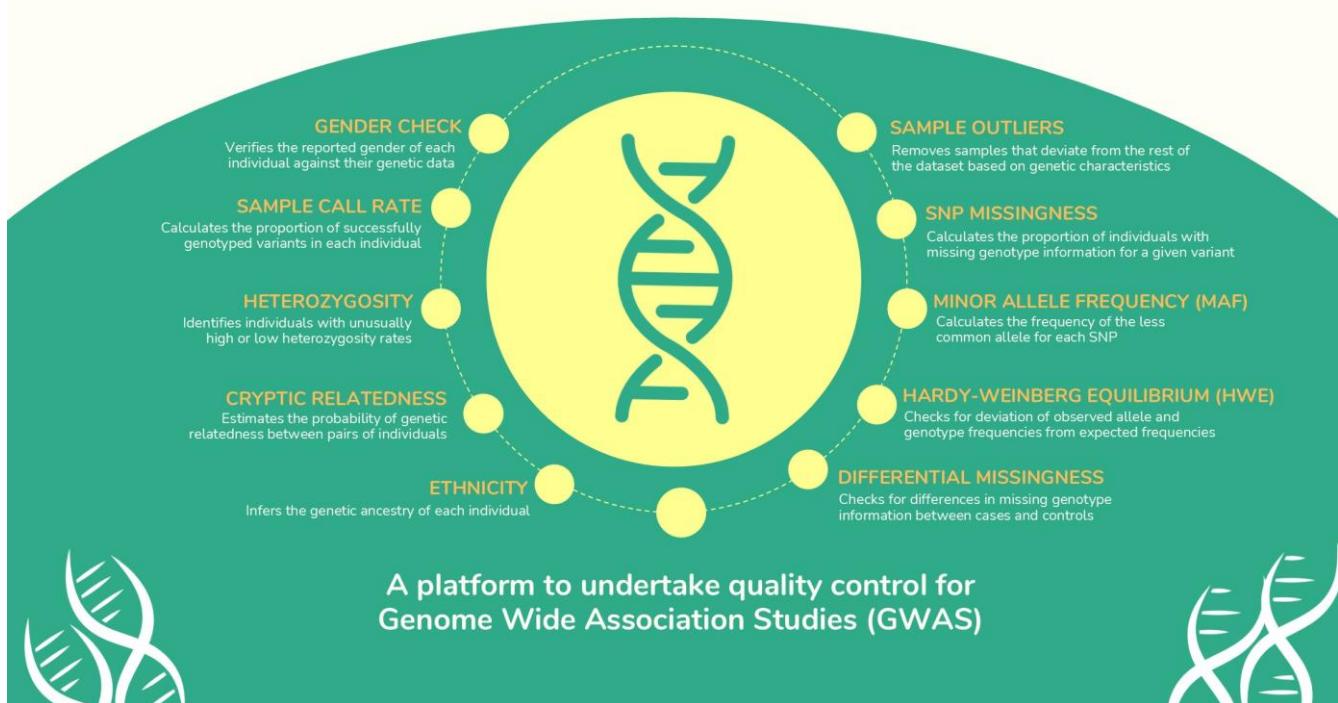


GWAS-QC



USER MANUAL

Authors

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

University of Liverpool

Table of Contents

GETTING STARTED	2
Installation.....	3
Run software.....	4
Input genotype file	5
Dummy data	5
FILE CONVERSION.....	6
Converter.....	7
SAMPLE-QC (SQC).....	8
Project setup.....	9
Data import.....	10
1. Preparation.....	11
2. SQC-Sex.....	12
3.SQC-Update sex (optional).....	13
4.SQC-Sample call rate	15
5.SQC-Heterozygosity	16
6.SQC-Cryptic relatedness	17
7.SQC-Ethnicity.....	18
SAMPLE REMOVAL.....	19
8.Remove sample outliers (optional)	20
VARIANT-QC (VQC)	21
9.VQC- SNP missingness	22
10.SQC-Minor allele frequency	23
11.SQC-HWE	24
12.SQC-Differential missingness (optional)	25
DATA EXPORT.....	26
Export results	27

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

Name	Date modified	Type	Size
 gwasqc.exe	01/07/2025 11:39	Application	766,328 KB

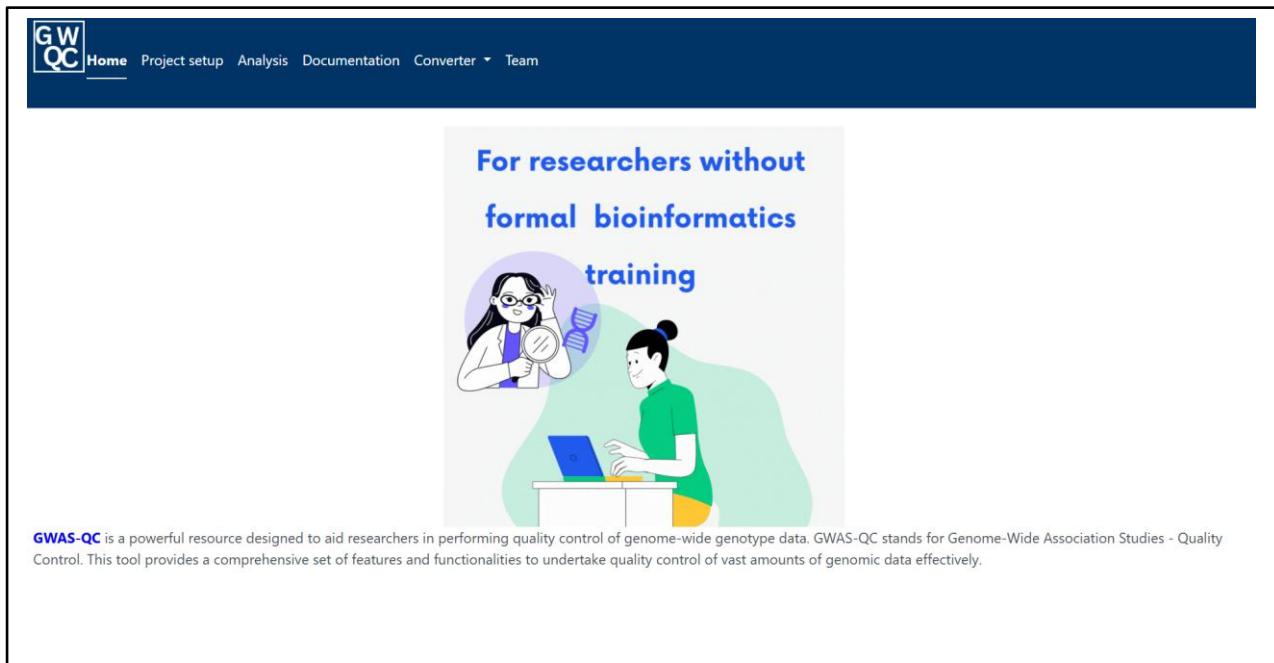
Watch the installation video for detailed instructions.

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.

Name	Date modified	Type	Size
dummy.bed	10/10/2024 10:09	BED File	4,645 KB
dummy.bim	10/10/2024 10:09	BIM File	32,114 KB
dummy.fam	10/10/2024 10:09	FAM File	1 KB
dummy_pheno.txt	27/05/2025 11:40	Text Document	1 KB
dummy_remove_sample.txt	27/05/2025 11:34	Text Document	1 KB
dummy_sex_update.txt	27/05/2025 13:06	Text Document	1 KB

PLINK Files

Phenotype

Sample IDs for removal

Samples for sex update

FILE CONVERSION

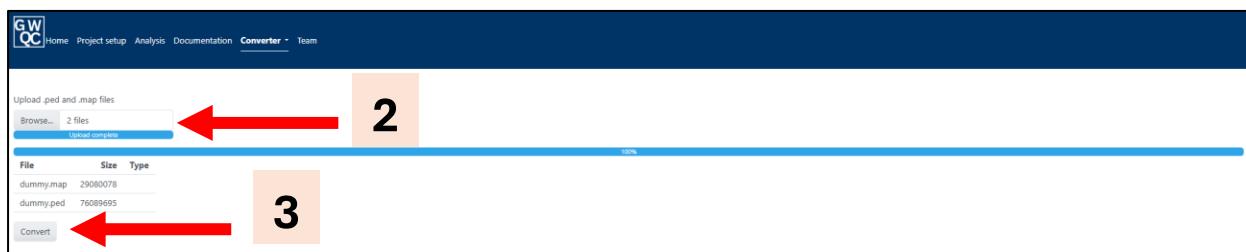
Converter

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

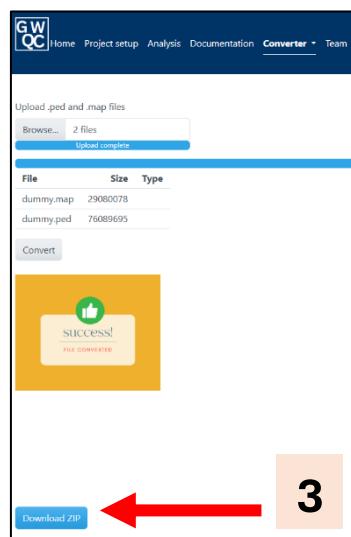


2. Browse to upload files.

3. Click “Convert”.



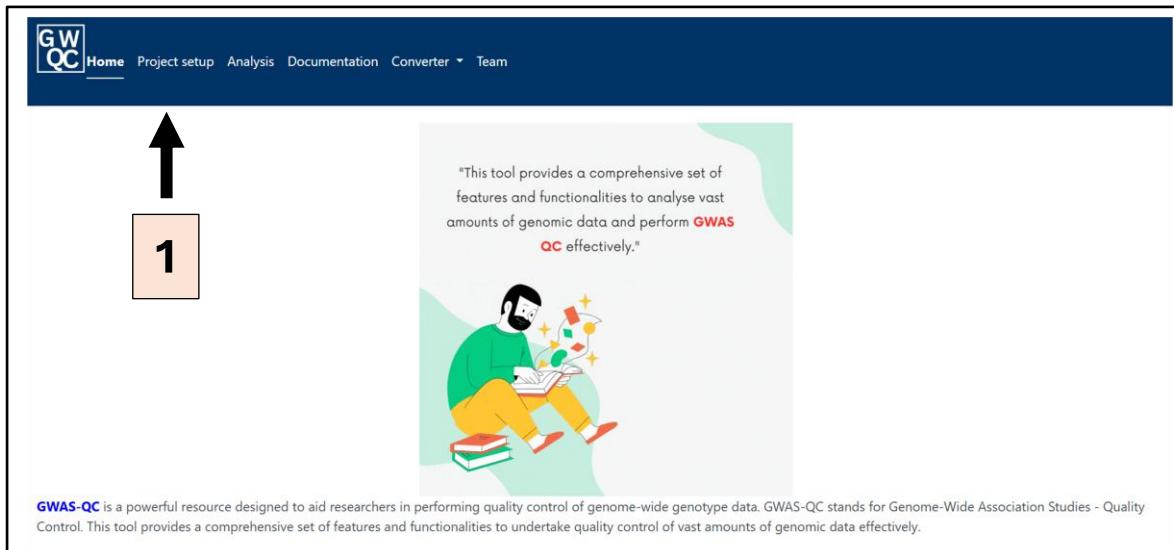
4. Upon successful file conversion, the binary PLINK files will be available for download.



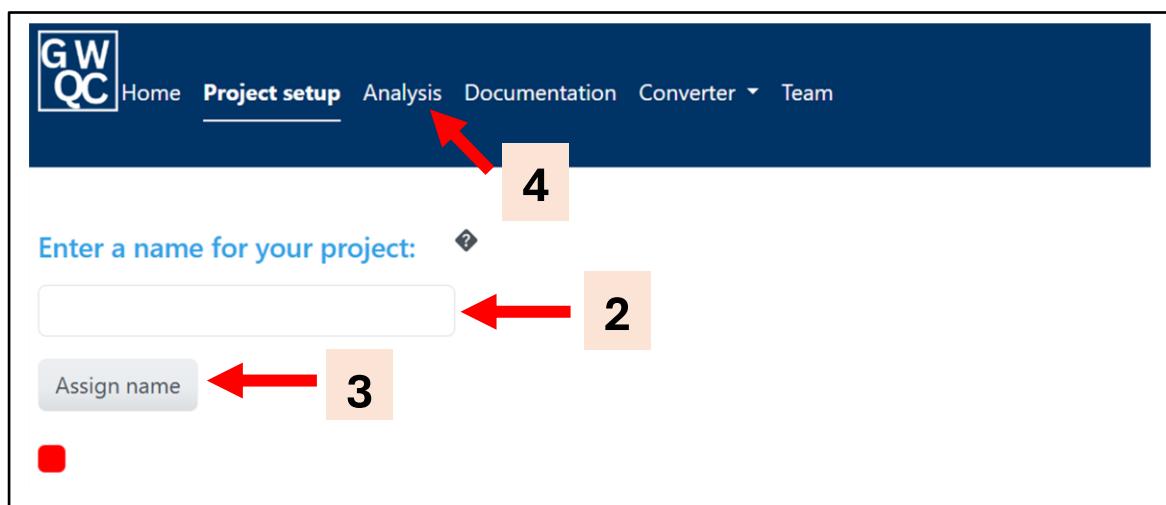
SAMPLE-QC (SQC)

Project setup

1. Click “Project setup”

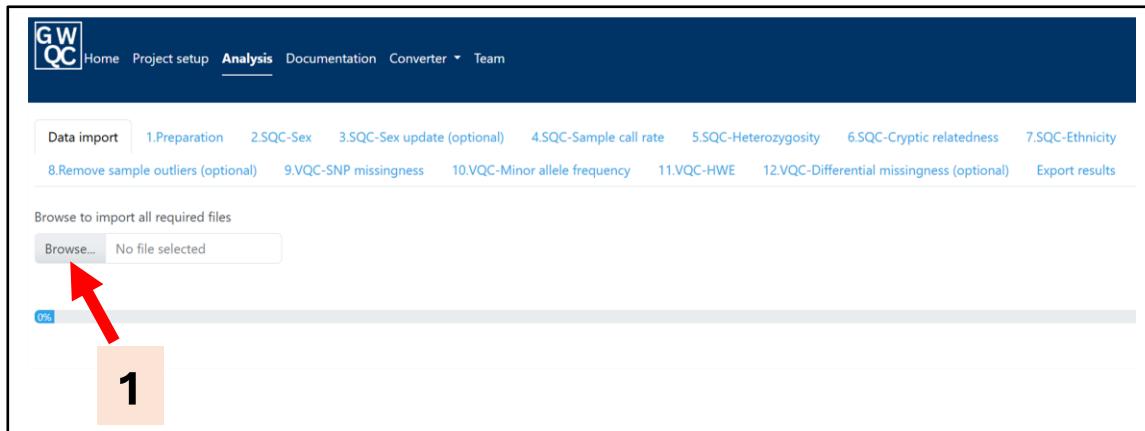


2. Enter a name for your project, *preferably less than 5 characters consisting of alphabets and numeric only*. **Note: No special characters and/or whitespace allowed.**
3. Click “Assign name” button.
4. Go to “Analysis” tab when you are prompted to proceed.

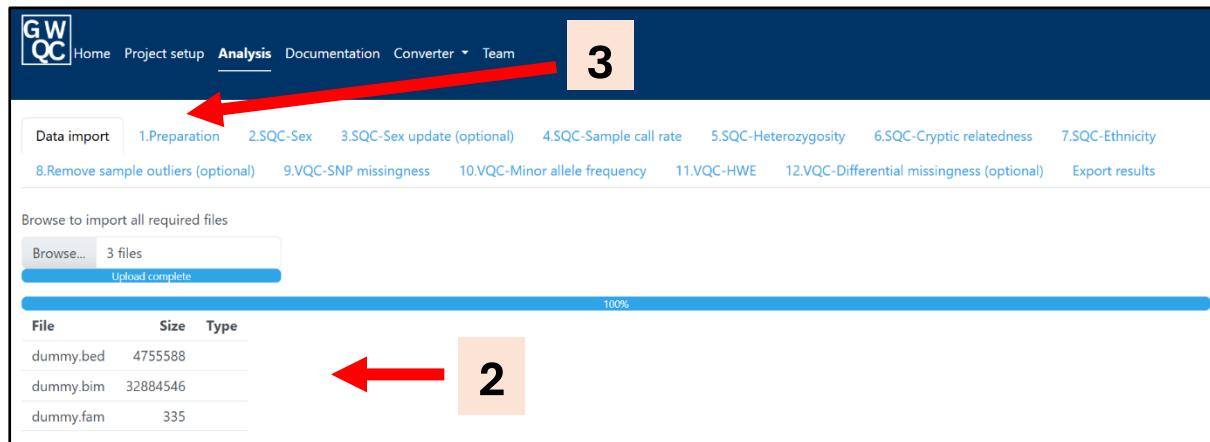


Data import

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

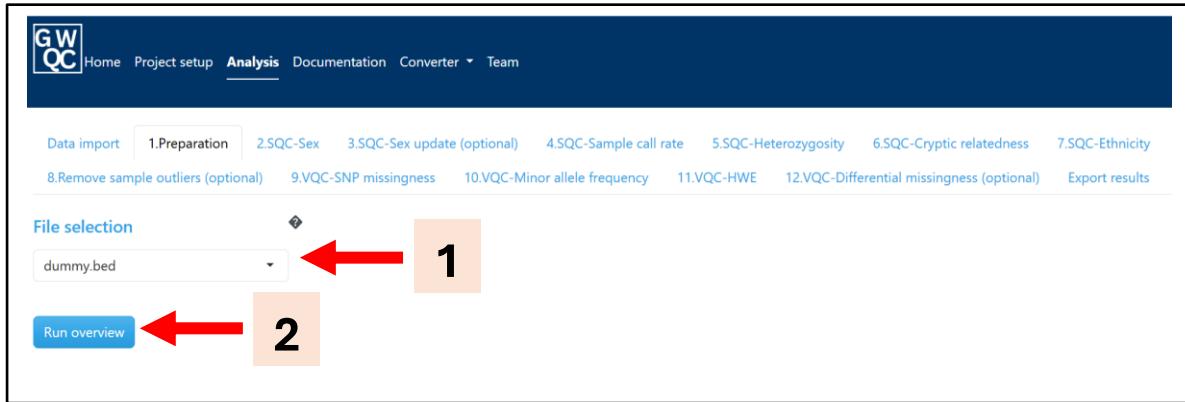


2. Upon completion, details of imported files will be displayed.
3. Click the “1.Preparation” tab when all the required files have been imported successfully.

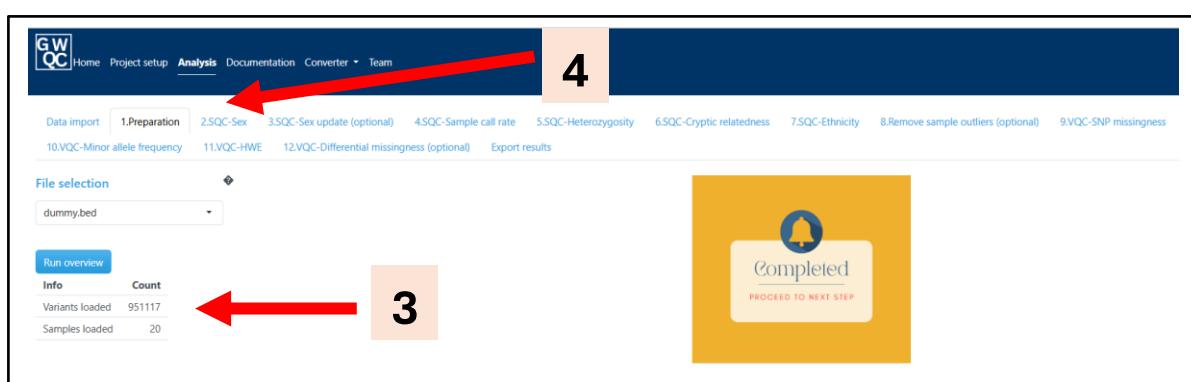


1. Preparation

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

- Click “Run sex check” button to check for sex concordance between reported and genetic sex.

GW QC Home Project setup **Analysis** Documentation Converter ▾ Team

Data import 1.Preparation 2.SQC-Sex 3.SQC-Sex update (optional) 4.SQC-Sample call rate 5.SQC-Heterozygosity 6.SQC-Cryptic relatedness 7.SQC-Ethnicity
8.Remove sample outliers (optional) 9.VQC-SNP missingness 10.VQC-Minor allele frequency 11.VQC-HWE 12.VQC-Differential missingness (optional) Export results

Identify samples with discordant sex information

Run sex check Run sex check

1

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

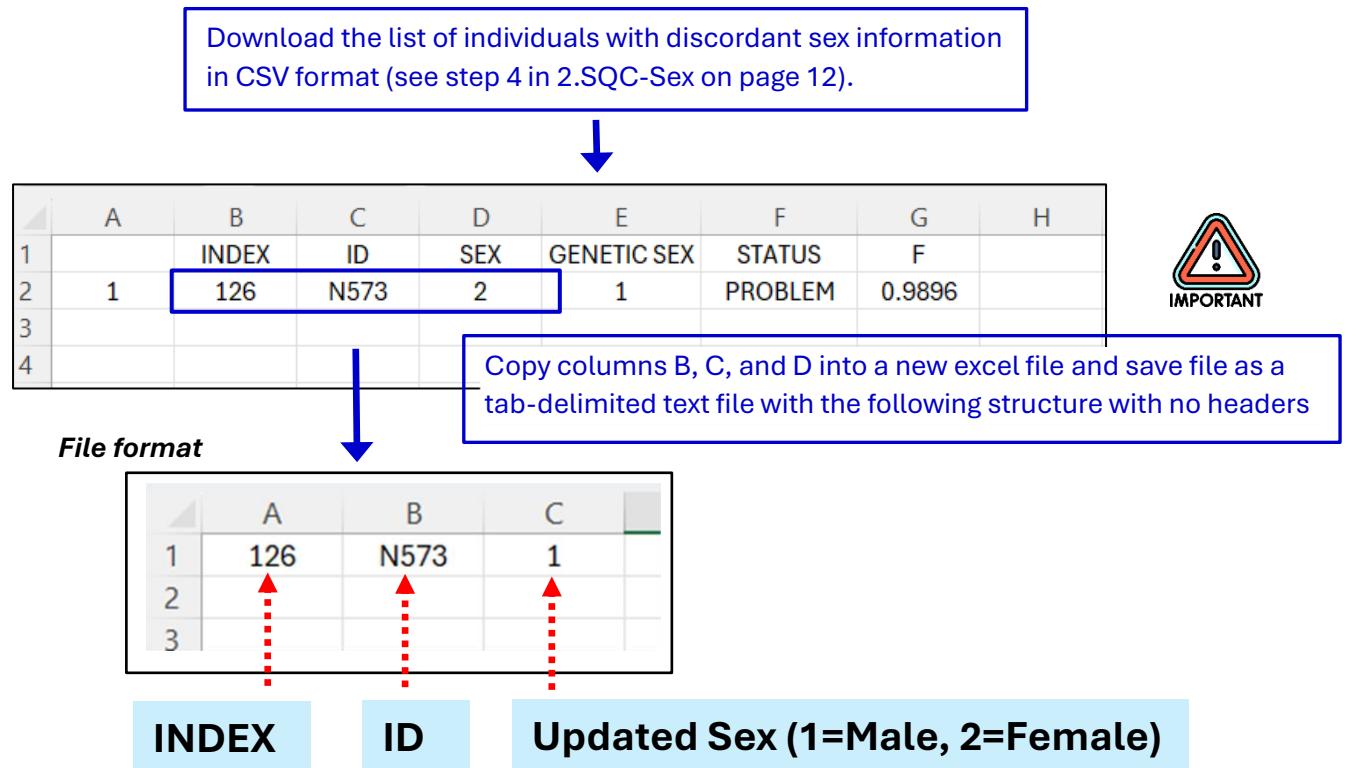
- List of individuals with discordant sex information can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



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

3.SQC-Update sex (optional)

- 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:



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

GW QC Home Project setup Analysis Documentation Converter Team

Data import 1.Preparation 2.SQC-Sex 3.SQC-Sex update (optional) 4.SQC-Sample call rate 5.SQC-Heterozygosity 6.SQC-Cryptic relatedness 7.SQC-Ethnicity
8.Remove sample outliers (optional) 9.VQC-SNP missingness 10.VQC-Minor allele frequency 11.VQC-HWE 12.VQC-Differential missingness (optional) Export results

Proceed to 4.SQC-Sample call rate if sex update is not required

Upload sex information

Upload tab-delimited text file

Browse... dummy_sex_update.txt Upload complete

1

2

Update sex info

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

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

GW QC Home Project setup **Analysis** Documentation Converter ▾ Team

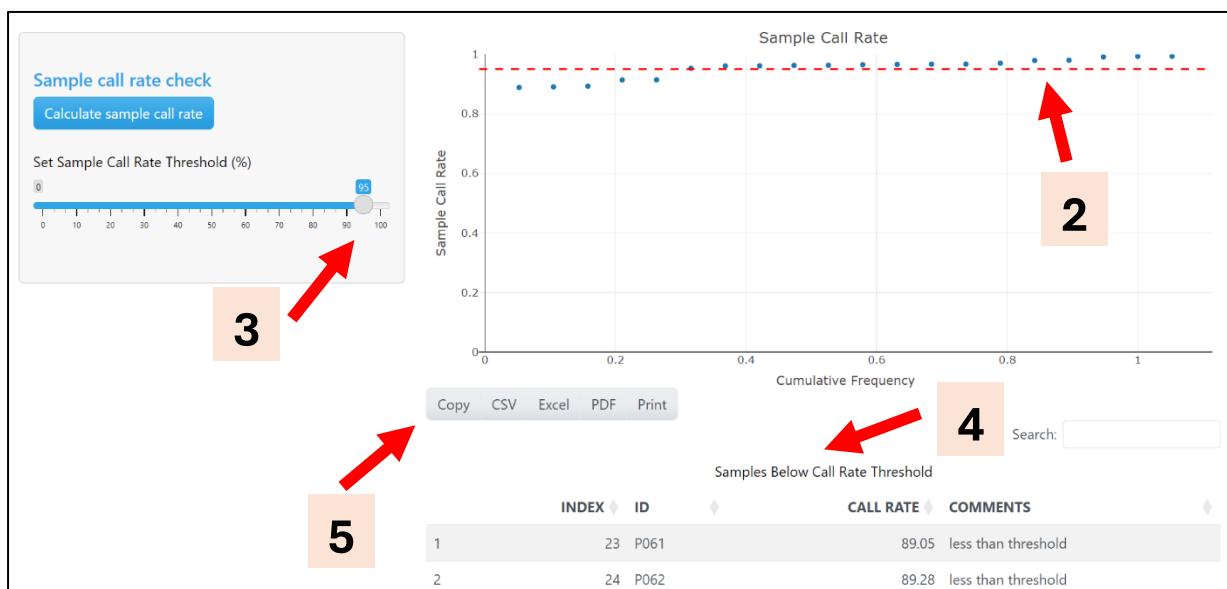
Data import 1.Preparation 2.SQC-Sex 3.SQC-Sex update (optional) 4.SQC-Sample call rate 5.SQC-Heterozygosity 6.SQC-Cryptic relatedness 7.SQC-Ethnicity
8.Remove sample outliers (optional) 9.VQC-SNP missingness 10.VQC-Minor allele frequency 11.VQC-HWE 12.VQC-Differential missingness (optional) Export results

Sample call rate check

1

Calculate sample call rate

- An interactive plot displays the sample call rate across individuals, with a red dashed line indicating 95% call rate by default.
- Move the slider to change the sample call rate threshold and the interactive plot will be updated accordingly.
- IDs of samples below the set call rate threshold are displayed in the table.
- This list of outliers can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



5.SQC-Heterozygosity

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

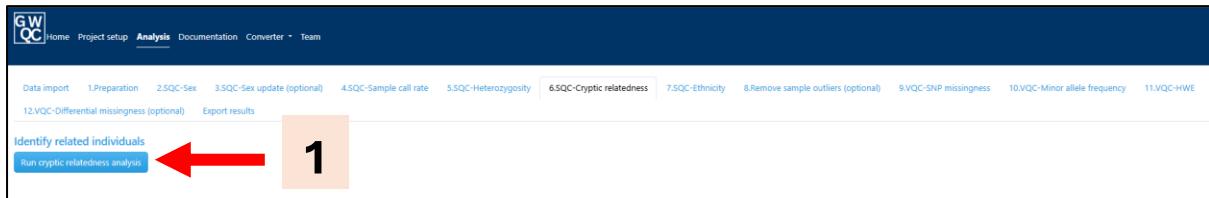
The screenshot shows the GWAS-QC software interface. The top navigation bar has 'GW QC' logo, 'Home', 'Project setup', 'Analysis' (which is underlined), 'Documentation', 'Converter', and 'Team'. Below the navigation is a horizontal menu with various analysis steps: 'Data import', '1.Preparation', '2.SQC-Sex', '3.SQC-Sex update (optional)', '4.SQC-Sample call rate', '5.SQC-Heterozygosity' (which is highlighted in blue), '6.SQC-Cryptic relatedness', '7.SQC-Ethnicity', '8.Remove sample outliers (optional)', and '9.VQC-SNP missingness'. Underneath this menu, there are two rows of numbered options: '10.VQC-Minor allele frequency', '11.VQC-HWE', '12.VQC-Differential missingness (optional)', and 'Export results'. On the left, a sidebar says 'Identify individuals with outlying heterozygosity rate' and contains a 'Run heterozygosity' button. A large red arrow points to this button, and it is highlighted with a red box and labeled '1'.

- 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.
- Click on the drop down to change the $\pm SD$ of mean heterozygosity rate and the interactive plot will be updated accordingly.
- IDs of samples beyond the selected heterozygosity rate and call rate are displayed in the table.
- This list of outliers can be downloaded using the Copy, CSV, Excel, PDF, or Print function.



6.SQC-Cryptic relatedness

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



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

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

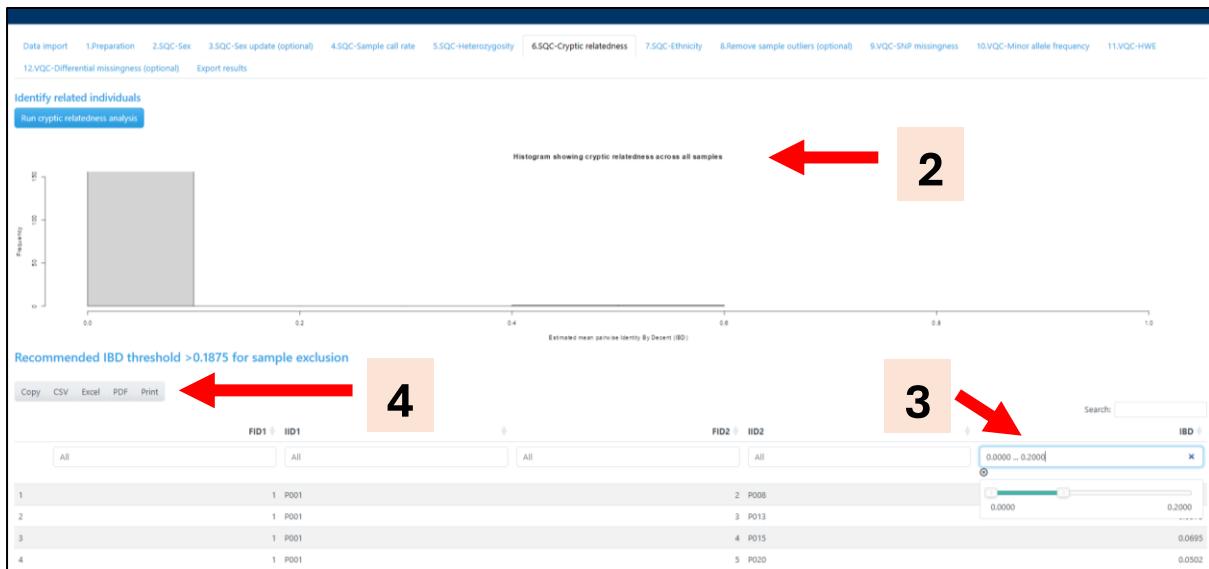
IBD = 1 duplicates or monozygotic twins,

IBD = 0.5 first-degree relatives,

IBD = 0.25 for second-degree relatives

IBD = 0.125 for third-degree relatives

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



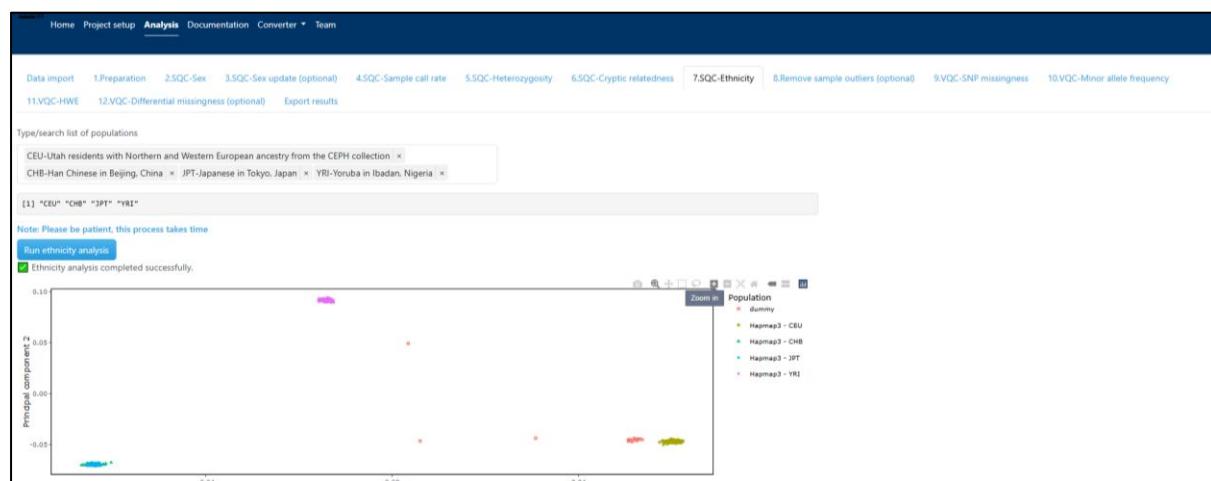
7.SQC-Ethnicity

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



The screenshot shows the GWAS-QC software interface. The 'Analysis' tab is active. In the search bar, the code 'CEU- Utah residents with Northern and Western European ancestry from the CEPH collection' is entered, followed by 'CHB-Han Chinese in Beijing, China', 'JPT-Japanese in Tokyo, Japan', and 'YRI-Yoruba in Ibadan, Nigeria'. Below the search bar, a note says 'Note: Please be patient, this process takes time.' A blue button labeled 'Run ethnicity analysis' is highlighted with a red arrow labeled '2'. Another red arrow labeled '1' points to the search bar.

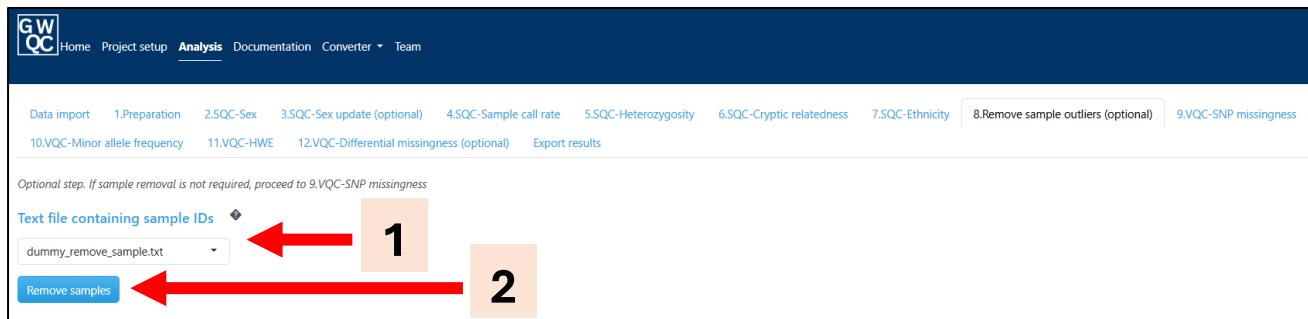
- Click “Run Ethnicity Analysis” button. This process takes much longer, wait until results appear.



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

	INDEX	ID
	A	B
1	5	P020
2	6	P022
3	7	P028
4	8	P029
5	9	P033
6	21	P054
7		

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.

Variant_info	Count
1 Variants loaded	951117
2 Variants removed	108286
3 Variants remaining	842831

Battery status: 79% available (plugged in)

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.

Set threshold for minor allele frequency (MAF)
MAF (range 0-0.1)
0.01

Run minor allele frequency

Removal of SNPs below set minor allele frequency threshold completed. Proceed to Hardy-Weinberg Equilibrium - HWE*

Variant_info

		Count
1	Variants loaded	842831
2	Variants removed	259595
3	Variants remaining	583236

Showing 1 to 3 of 3 entries

Previous 1 Next

11.SQC-HWE

- 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:

	A	B	C
1	1	P001	1
2	2	P008	1
3	23	P061	2
4	24	P062	2

INDEX **ID** **Phenotype (1=Control, 2=Case)**

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

GW QC Home Project setup Analysis Documentation Converter Team

Data import 1.Preparation 2.SQC-Sex 3.SQC-Sex update (optional) 4.SQC-Sample call rate 5.SQC-Heterozygosity 6.SQC-Cryptic relatedness 7.SQC-Ethnicity
8.Remove sample outliers (optional) 9.VQC-SNP missingness 10.VQC-Minor allele frequency 11.VQC-HWE 12.VQC-Differential missingness (optional) Export results

Set threshold for Hardy-Weinberg Equilibrium (HWE)
Upload phenotype file (for case-control comparison only) Upload complete **1**

HWE p-value threshold **2**

Run HWE **3**

HWE process completed on set threshold

Variant_info	Count
1 Variants loaded	583236
2 Variants removed	0
3 Variants remaining	583236

Showing 1 to 3 of 3 entries

Search:

Count **1**

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.

Differential missingness (for case-control comparison only)

Set differential missingness P-value

0.01

Run differential missingness

Differential missingness completed. Export results

Copy CSV Excel PDF Print

	Count
1 Variants loaded	583236
2 Variants removed	0
3 Variants remaining	583236
4 Case	10
5 Control	10

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.

