

Documentation

Multi-ethnic Imputation System (MI-System)

[\(https://misystem.cgm.ntu.edu.tw/\)](https://misystem.cgm.ntu.edu.tw/)

Index	Page #
I. Log in to the system	2
II. Universally unique identifier (UUID)	3
III. Data preparation	4
IV. Main Functions	
1. Imptation Pipelines	4
(a) Upload dataset	5
(b) Select quality control (QC) parameters	6
(c) Reference panel selection	6
(i) Choose from reference panels offered by the system	6
(ii) Custom reference panel	7
(iii) Combine reference panels	7
(d) Imputation results download	10
2. Utility Functions	11
(a) Create reference panel	11
(b) Split chrommosome	13
(c) Liftover	14
V. Download Results	16

I. Log in to the system

To get started with Multi-ethnic Imputation System (MI-System) , users are required to provide their email address by clicking on the start button at the top left corner of the welcome page (Figure T1). Press “Start new project” once the email address is entered.

MI-System HomePage **Start** Tutorial

Multi-ethnic Imputation System

Enter your email as user id

Email* Type email

※Email will only use to identify the user and create the user folder

Start new project

Access to previous project:

Uuid

※UUid will use to identify the project and create the project folder

Access

Figure T1. Logging in with email ID in to the system

II. UUID

MI-System will randomly assign a universally unique identifier (UUID) after logging in. The UUID is sent to the user's email as well (**Figure T2**). This UUID can be used to access all results in the future **Figure T3**.

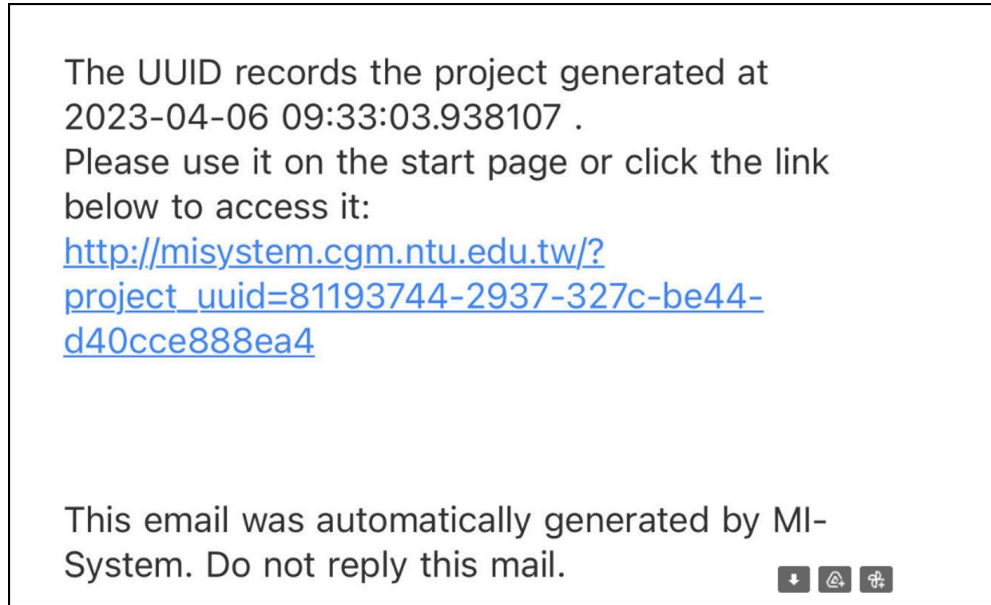


Figure T2. Email notification to users about universally unique identifier (UUID) after logging in



Figure T3. Screenshot of UUID after logging in

III. Data preparation

MI-System accepts both PLINK binary format files (.bed, .bim, .fam) and VCF (vcf.gz) files. Create separate files for each chromosome. Sort the variants by genomic position. All data needs to be in build GRCh37. Plink files can be created by PLINK software available freely (<https://www.cog-genomics.org/plink/2.0/>). Use VCFtools or GATK to create genotype VCF files. Plink command can be used to convert vcf to plink binary files or vice versa using the following commands

```
plink2 --vcf example.vcf.gz --make-bed --out output_example
```

```
plink2 --bfile example_data --keep-allele-order --recode vcf --out output_vcf
```

IV. Main Functions

1. Imputation pipelines

To start an imputation job, after logging into the system, select the function button from the options at the top left of the homepage. Choose from any of the three imputation pipelines(**Figure T4**):

- (i) **Imputation:** SHAPEIT2 (pre-phasing) and IMPUTE2 (imputation)
- (ii) **Imputation_IMPUTE5 alpha test:** SHAPEIT4 (prephasing) and IMPUTE5 (imputation)
- (iii) **Imputation_Beagle5 alpha test:** Beagle 5.1 (prephasing and imputation)

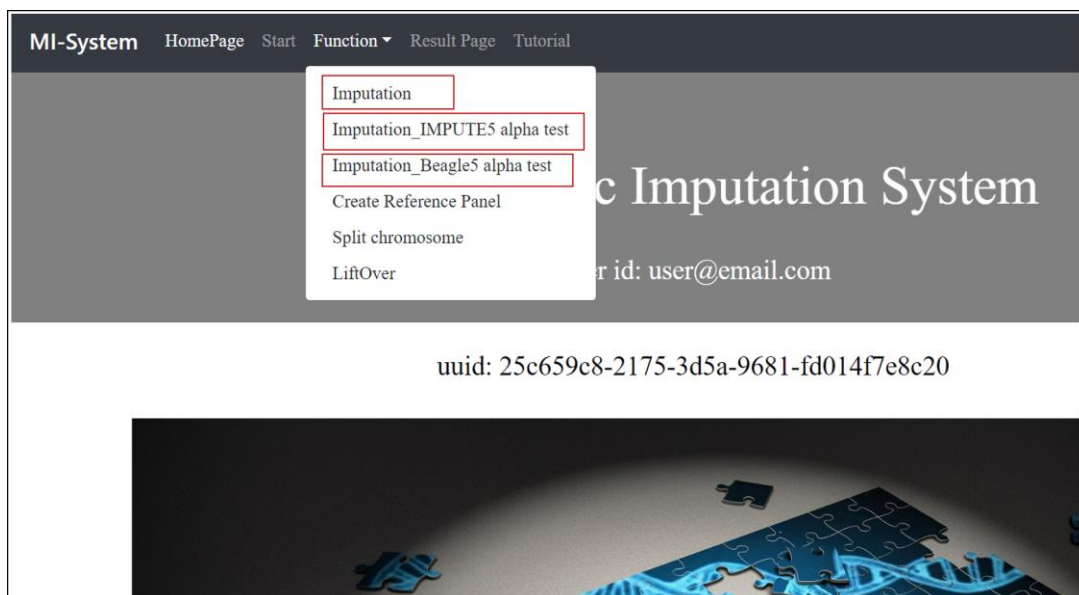


Figure T4: Select imputation pipeline for starting the imputation job

(a) Upload dataset:

Start by assigning a project name. Upload either **PLINK binary files** (*example.bed*, *example.bim*, *example.fam*) (**Figure T5**) or *.vcf* format files directly from your source by pressing the “**Choose files**” button. Alternatively, files can be uploaded by pasting the url of the data by checking the “**Use URL to upload**” option (**Figure T6**). Choose the **Chromosome number** from the drop down options. If the imputation requires to be done on the full chromosome, check the “**impute the full chromosome**” button, otherwise mention the **start position** and the **end position**, to indicate the region of interest that needs imputation. All data needs to be uploaded for each chromosome at a time in human genome build HG19 (GRCh37). If you have whole genome data, split your data into chromosomes using the utility function **Split chromosome** (see page 13) If your data is in GRCh38 build then use the utility function **Liftover** (see page 14) to convert your data to hg19.

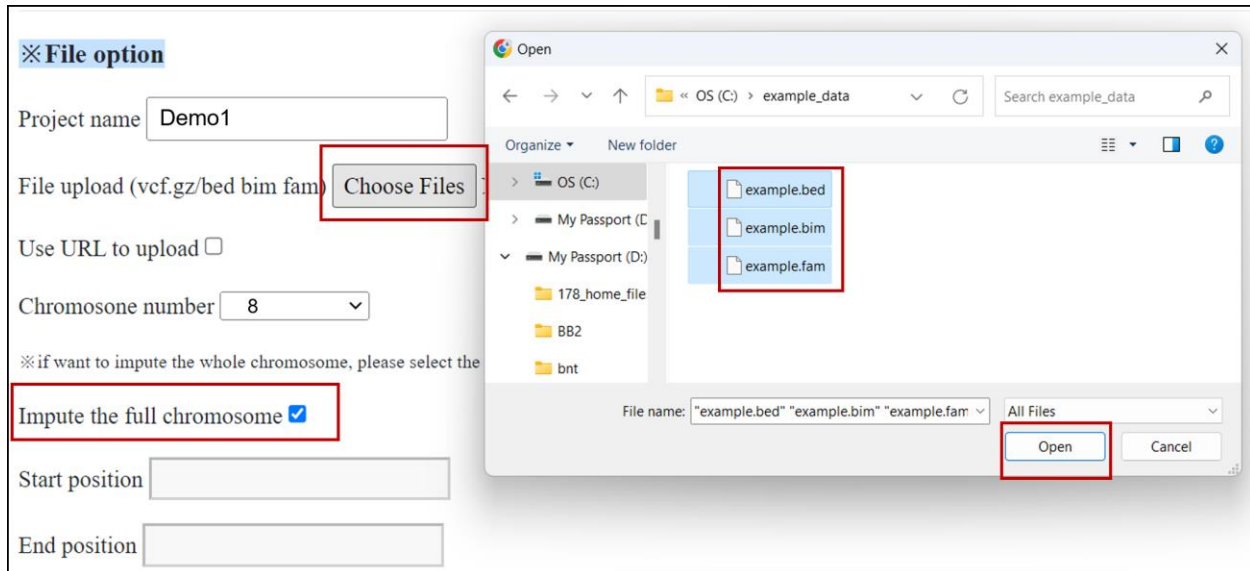


Figure T5. Upload files, each chromosome at a time, from user source directly

Use URL to upload ☒

Allow google drive link

※Notice that the link need to open the authority of the google drive

File upload url (vcf.gz)

File upload url (bed)

File upload url (bim)

File upload url (fam)

Chromosome number

※if want to impute the whole chromosome, please select the function

Figure T6. Upload files, each chromosome at a time, by pasting the url of the source file location

(b) Select quality control (QC) parameters

Use drop down menus to select the thresholds for minor allele frequencies (MAF), Single SNP missing rate option (call rate), individual SNP missing rate option (missing genotype) and Hardy-Weinberg equilibrium (**Figure T7**).

(c) Reference panel selection

MI-System has three different options for reference panel.

(i) ***Choose from reference panels offered by the system***

(a) 1000 Genomes Phase III

Populations :

- European (EUR)
- South Asians (SAS)
- East-Asian (EAS)
- Ad-mixed Americans (AMR)
- African (AFR)
- All samples

(b) Hapmap3

(c) Taiwan Biobank

※Quality control option

Minor allele frequency option 0.05

Single SNP missing rate option 0.05

Individual SNP missing rate option 0.05

Hardy-Weinberg Equilibrium option 0.000001

※Reference option (if need to use custom reference, please select 'Custom')

Reference group select

-----Combine reference

※if need to use the merge re

Improved reference pan

Merge reference panels

-----Custom reference

※Reference option (if need

File upload (.sample, .k

1000G_phases3

European(EUR)

South Asian(SAS)

East Asian(EAS)

Ad Mixed American(AMR)

African(AFR)

1000G_phase3 total

HapMap 3

HapMap 3 -- hg18

Taiwan biobank

Taiwan biobank

Custom reference

Custom

Choose Files No file chosen

Notice

MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

☐ have read the Privacy Statement of MI-System and accept MI-System to process my upload

Submit

Figure T7. Select quality control thresholds and Reference panel to be used while conducting imputation.

Once the reference panel is selected from the drop down menu (**Figure T7**), the users are required to confirm our privacy statement, after which the imputation process can be started immediately by clicking on *submit* button.

(ii) *Custom reference panel*

Users can upload their own customized data that they may want to use as a reference panel for conducting imputation. Choose the **Custom** option from the drop down menu of **Reference group select** (**Figure T8**), then upload your reference panel files for each chromosome at a time (reffile.sample, reffile.legend.gz, reffile.hap.gz, genetic_map.txt). Using custom panels is only available for pipeline SHAPEIT2-IMPUTE2. To create appropriate formats for your customized panels use the utility function Create Reference Panel (see page 11)

※Reference option (if need to use custom reference, please select 'Custom')

Reference group select Custom ▼

-----Combine reference panels-----

※if need to use the merge reference or two phased reference function, please select the second reference that different with the first reference.

Improved reference panels None ▼

Merge reference panels Taiwan biobank ▼

-----Custom reference-----

※Reference option (if need to use custom reference, please upload the reference files)

File upload (.sample, .legend.gz, .hap.gz, genetic_map.txt) Choose Files No file chosen

↓

Notice

MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

☐ I have read the Privacy Statement of MI-System and accept MI-System to process my upload

Submit

Figure T8. Use custom reference panel to conduct imputation

Once done, confirm our privacy statement, and click on *submit* button to start imputation.

(iii) Combine reference panels

This can be done by selecting two reference panels and selecting either *Merge reference panels* option (Figure T9) or *Improved reference panels* option (Figure T10). You may choose both using the panels provided by the system, or you may upload your own reference panel, and combine it with one of the panels provided by the system. Combining panels option is only available for pipeline SHAPEIT2-IMPUTE2. To create appropriate formats for your customized panels use the utility function Create Reference Panel (see page 11)

※Reference option (if need to use custom reference, please select 'Custom')

Reference group select East Asian(EAS) ▼

-----Combine reference panels-----

※if need to use the merge reference or two phased reference function, please select the second reference that different with the first reference.

Improved reference panels None ▼

Merge reference panels Taiwan biobank ▼

↓

Notice

MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

☐ I have read the Privacy Statement of MI-System and accept MI-System to process my upload

Submit

Figure T9. Combine reference panels : Merge reference panels

※Reference option (if need to use custom reference, please select 'Custom')

Reference group select Custom ▼

-----Combine reference panels-----

※if need to use the merge reference or two phased reference function, please select the second reference that different with the first reference.

Improved reference panels Taiwan biobank ▼

Merge reference panels None ▼

↓

Notice

MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

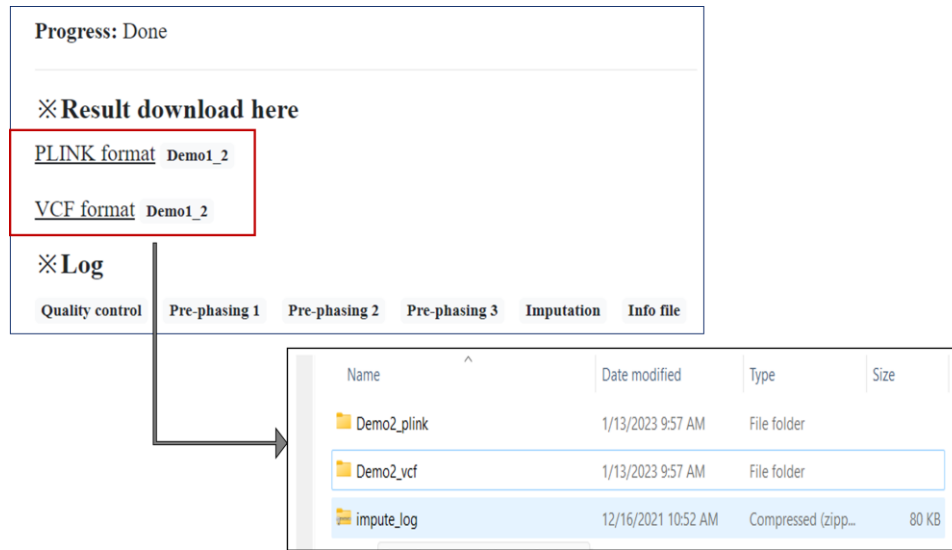
☐ I have read the Privacy Statement of MI-System and accept MI-System to process my upload

Submit

Figure T10. Combine reference panels : Improved reference panels

(d) Imputation results download

Go to the results page from the menu in the home page, choose Imputation and select your project name to go to the results page (see page 16). Click to download and save the imputation results. You can also download all the log files (Figure T11). High quality plots can also be downloaded from the results page (Figure T12).



Progress: Done

※Result download here

[PLINK format Demo1_2](#)

[VCF format Demo1_2](#)

※Log

Quality control Pre-phasing 1 Pre-phasing 2 Pre-phasing 3 Imputation Info file

Name	Date modified	Type	Size
Demo2_plink	1/13/2023 9:57 AM	File folder	
Demo2_vcf	1/13/2023 9:57 AM	File folder	
impute_log	12/16/2021 10:52 AM	Compressed (zip) file	80 KB

Figure T11. Download results and log files for imputation.

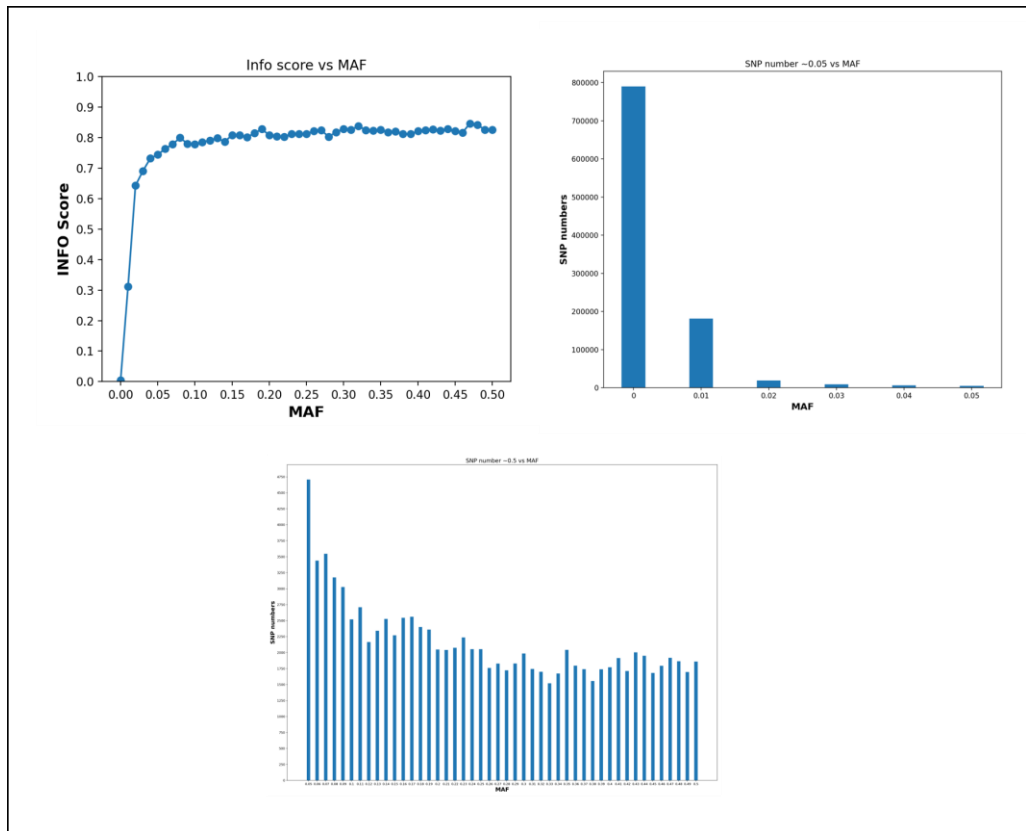


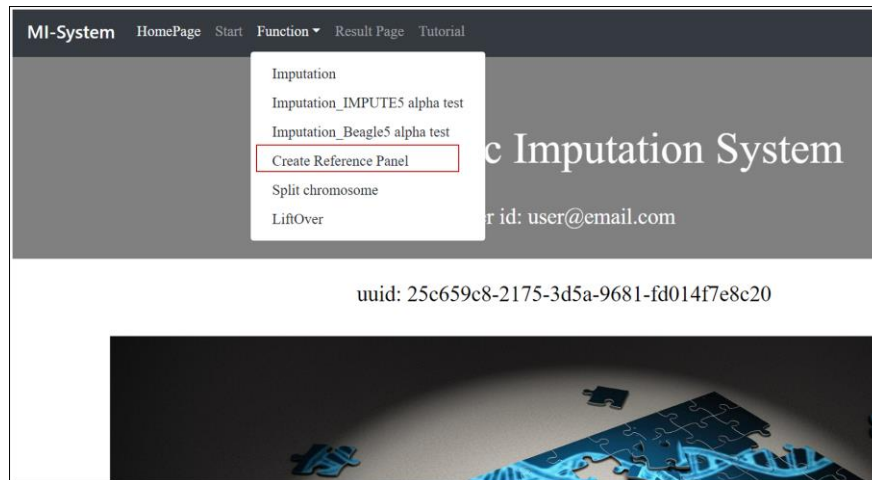
Figure T12. Downloadable plots showing quality of the imputed data

2. Utility Functions

The utility functions are provided to help the users to preprocess and customize their datasets to be used in the MI-System for conducting imputation. Three utility functions are provided

- (i) **Create reference panel:** Users can use this function to create suitable reference file formats by uploading their data to the system. The reference panel formatted files can then be downloaded (reffile.sample, reffile.legend.gz, reffile.hap.gz, genetic_map.txt), which can be used as **Custom reference** by uploading it for function **Imputation** (page)

Upload each chromosome data either by browsing from your computer or by pasting the url of the data source, in either PLINK binary format or VCF format. Once done the users are required to confirm our privacy statement, after which the process can be started immediately by clicking on submit button (**Figure T13**). Go to the result page to download the reference format files **Figure T14**.



The screenshot shows the 'Create Reference Panel' form. The form has a section titled 'File option' with the following fields and options:

- Project name:
- File upload (vcf.gz/bed bim fam): (highlighted with a red box)
- Chromosome number:
- Use URL to upload: ☒
- Allow google drive link: ☐
- File upload url (vcf.gz):
- File upload url (bed):
- File upload url (bim):
- File upload url (fam):

A file upload dialog is open, showing the contents of a folder named 'example_data'. The files listed are 'example.bed', 'example.bim', and 'example.fam'. The 'File name' field is empty, and the 'File type' is set to 'All Files'. The 'Open' button is highlighted.

Below the form, there is a 'Notice' section with the following text:

MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

Below the notice, there is a checkbox labeled 'I have read the Privacy Statement of MI-System and accept MI-System to process my upload' (highlighted with a red box). At the bottom, there is a 'Submit' button (highlighted with a red box).

Figure T13. Upload data to run the process of creating reference panel format files

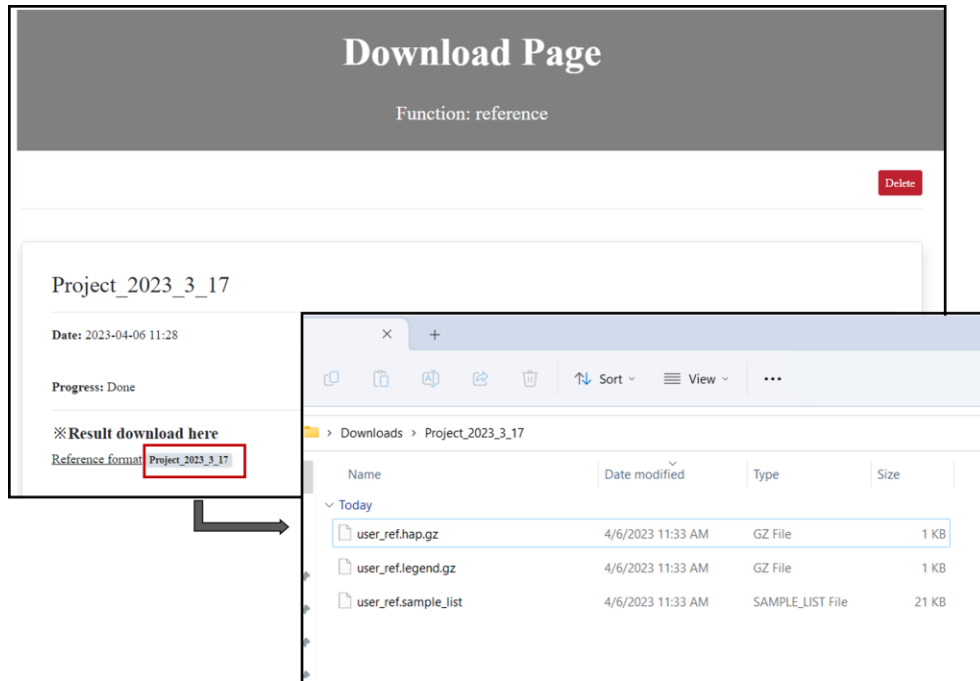
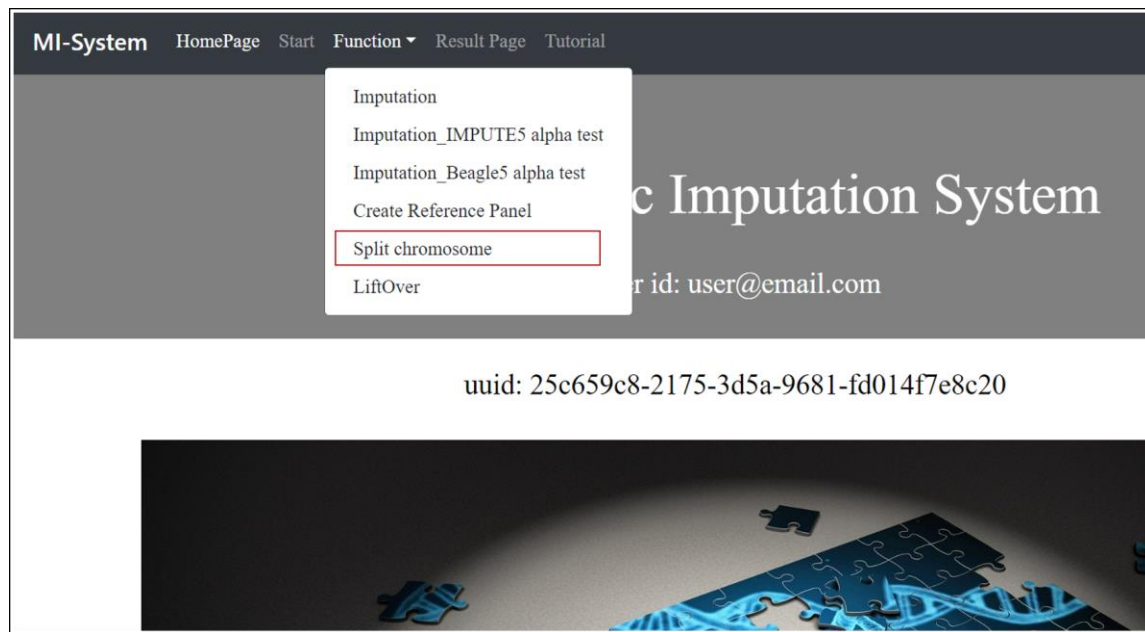


Figure T14. Result page for create reference files

- (ii) **Split chromosome:** This function is provided to help users to break their whole genome data into chromosomes (**Figure T15**). Imputation in MI-System can be done for one chromosome at a time. To obtain data for each chromosome users can upload their whole genome data and obtain split data for all chromosomes in Plink binary format (chr#.bed, chr#.bim, chr#.fam) by clicking **Result page** tab from the welcome page and selecting the **Project name (assigned by users)** from the **Split_chromosome** menu.



Split chromosome

✖ **File option**

Project name

Data name

File upload (vcf.gz/bed bin fam) 3 files

Use URL to upload ☐

Notice

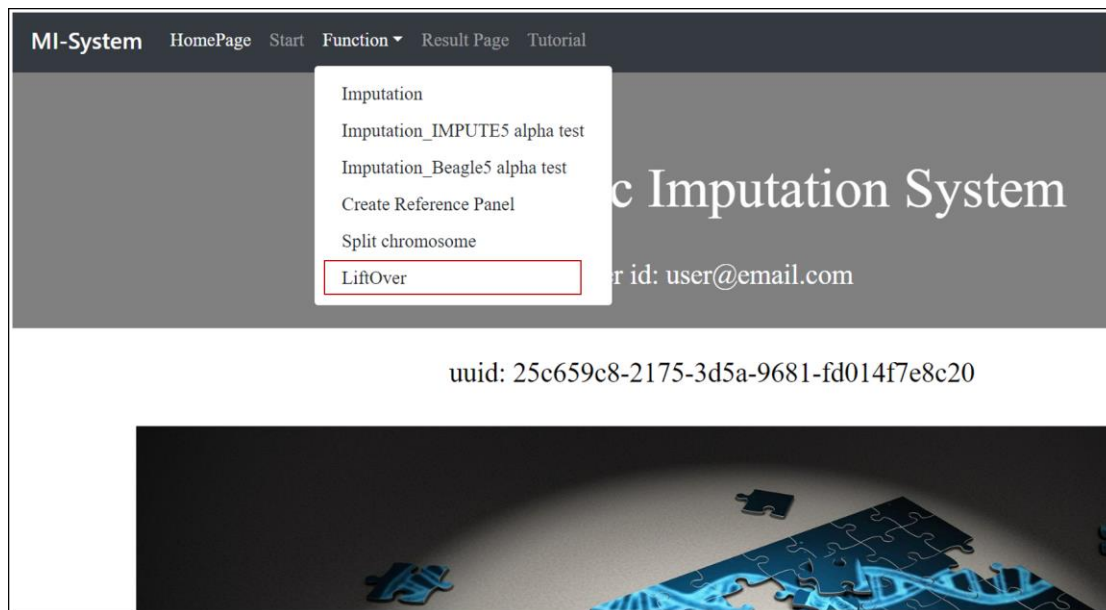
MI-System NEVER collects and shares user uploaded data and results with others. All data and metadata of one task will be removed automatically from the server without any copies after 7 days of the report generated.

☐ I have read the Privacy Statement of MI-System and accept MI-System to process my upload

Figure T15. Submission of job for Splitting whole genome data into chromosomes

- (iii) **Liftover:** MI-System requires user to upload data in GRCh37 build. Therefore, if your data is in GRCh38 build, you are required to convert it to GRCh37. Liftover function can be used to convert the genome builds of user uploaded data (**Figure T16**) by selecting from the drop down menu. Users can upload PLINK binary format files and once the function is successfully run, can download split data in

PLINK binary format again. These files can be used into all other functions of MI-System.



The screenshot shows the 'LiftOver' submission page. The title 'LiftOver' is at the top. Below it, the section '※File option' contains several fields: 'Project name' (Demo), 'Data name' (Try_data), 'Convert function' (Hg38 to Hg19), and 'File upload (vcf.gz/bed bim fam)' (Choose Files 3 files). The 'Convert function' dropdown and the 'Choose Files' button are highlighted with red boxes. Below these fields is a checkbox for 'Use URL to upload'. A light blue 'Notice' box contains text about data collection and removal. Below the notice is a checkbox for 'I have read the Privacy Statement of MI-System and accept MI-System to process my upload'. At the bottom, a 'Submit' button is highlighted with a red box.

Figure T15. Submission of job for Splitting whole genome data into chromosomes

V. Download results

For all jobs run by the system, use the UUID that was emailed to you to access the results from the system. Once you enter the system using the UUID, then select the **Result page** from the top left of the welcome page. It will direct you to the result page with four different drop down menus, for imputation function and each of the utility functions, respectively (**Figure T16**). For the results that you desire to download, select the project name from the drop down menu and select it. It will take you to the results page and where all files can be downloaded using simple clicks. Once you have downloaded all results you can delete your results from each of the result page by clicking on the delete button at the right top of the page. Alternately the results will be automatically deleted after 7 days.

MI-System HomePage Start Function Result Page tutorial

Multi-ethnic Imputation System

Please select your project and download here

Imputation
Choose a Project:
Submit

Reference
Choose a Project:
Submit

Split_chromosome
Choose a Project:
Submit

LiftOver
Choose a Project:
Submit

Figure T16. Result page showing the drop down menus for each function