



MICHIGAN

IMPUTATIONSERVER



Christian
Fuchsberger



Sebastian
Schönherr



Cassandra
Spracklen



Lukas Forer



Eleftheria
Zeggini



Albert Smith

Disclosure Slide

Financial Disclosure for:

Christian Fuchsberger

Sebastian Schönherr

Lukas Forer

Cassandra Spracklen

Eleftheria Zeggini

Albert Smith

We have nothing to disclose

Setup

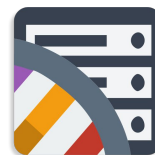
- 6 Sessions:
 - (1) Genotype Imputation
 - (2) Use the server and the Imputation Bot,
 - (3) GWAS,chrX, HLA (4) GWAS pipeline and PGS Server,
 - (5) Helmholtz Munich Imputation Server, (6) TOPMed
- Lectures
- Demos
- Interaction
 - PollEv.com/ashg
- Question & Answer session at the end

Section 1

Imputation and the Server



Christian Fuchsberger
Eurac Research
cfuchsberger@eurac.edu



MICHIGAN
IMPUTATIONSERVER

Learning objectives

Participants will

1. Understand the principles of genotype imputation and the Michigan Imputation Server

Genotype imputation

Key method used in GWAS to

- Increase the number of tested variants
- Fine-mapping becomes more complete
- Meta-analysis using different arrays

0. Imputation setting

GWAS Haplotypes

. . . . **A** **A** **A** . . .
. . . . **G** **C** **A** . . .

Reference Haplotypes (e.g. TOPMed)

C	G	A	G	A	T	C	T	C	C	T	T	C	T	T	C	T	G	T	G	C
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	C	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C
C	G	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C
C	G	A	G	A	C	T	C	T	C	C	G	A	C	C	T	T	A	T	G	C
T	G	G	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	T	G	T	G	C
C	G	A	G	A	C	T	C	T	T	T	T	C	T	T	T	T	G	T	A	C
C	G	A	G	A	C	T	C	T	C	C	G	A	C	C	T	C	G	T	G	C
C	G	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C

1. Identify match among reference

GWAS Haplotypes

. . . . A A A
. . . . G C A

Reference Haplotypes (e.g. TOPMed)

C	G	A	G	A	T	C	T	C	C	T	T	C	T	T	C	T	G	T	G	C
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	C	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C
C	G	A	A	G	C	T	C	T	C	C	G	A	C	C	T	T	A	T	G	C
T	G	G	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	T	G	T	G	C
C	G	A	G	A	C	T	C	T	T	T	T	C	T	T	T	T	G	T	A	C
C	G	A	G	A	C	T	C	T	C	C	G	A	C	C	T	C	G	T	G	C
C	G	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C

2. Impute

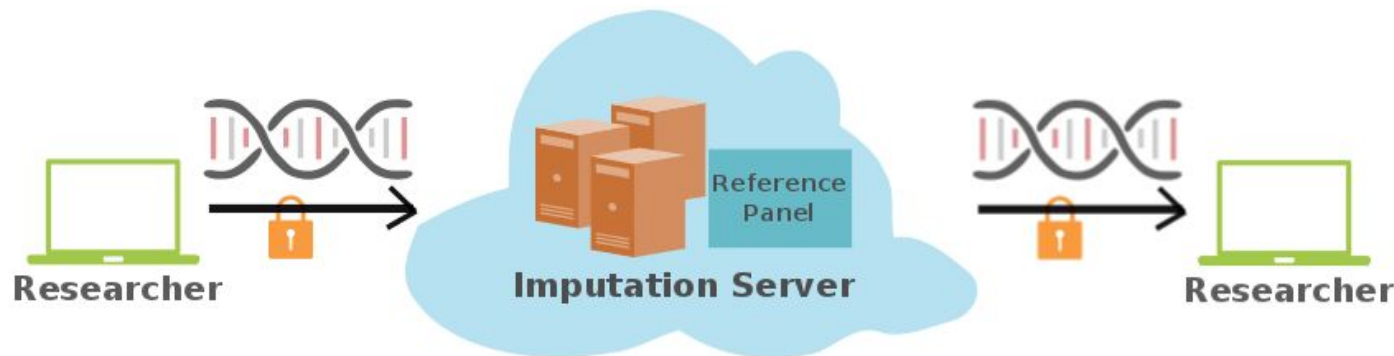
GWAS Haplotypes

c	g	a	g	A	t	c	t	c	c	c	g	A	c	c	t	c	A	t	g	g
c	g	a	a	G	c	t	c	t	t	t	t	C	t	t	t	c	A	t	g	g

Reference Haplotypes (e.g. TOPMed)

C	G	A	G	A	T	C	T	C	C	T	T	C	T	T	C	T	G	T	G	C
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	C	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C
C	G	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C
C	G	A	G	A	C	T	C	T	C	C	G	A	C	C	T	T	A	T	G	C
T	G	G	G	A	T	C	T	C	C	C	G	A	C	C	T	C	A	T	G	G
C	G	A	G	A	T	C	T	C	C	C	G	A	C	C	T	T	G	T	G	C
C	G	A	G	A	C	T	C	T	T	T	T	C	T	T	T	T	G	T	A	C
C	G	A	G	A	C	T	C	T	C	C	G	A	C	C	T	C	G	T	G	C
C	G	A	A	G	C	T	C	T	T	T	T	C	T	T	C	T	G	T	G	C

ASHG 2014: imputation web service



1.

Upload GWAS data

2.

Server performs

- Quality checks
- Pre-phasing
- Imputation
- Encryption

3.

Download results



MICHIGAN
IMPUTATIONSERVER



Interactive polls

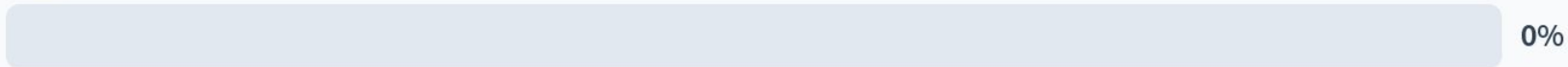


<https://pollev.com/ashg>

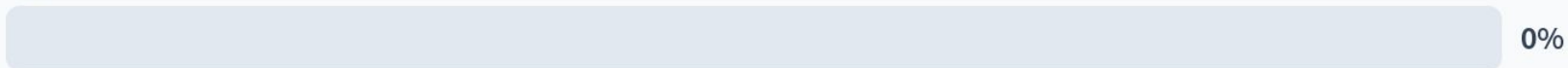


Have you ever used the Michigan Imputation Server

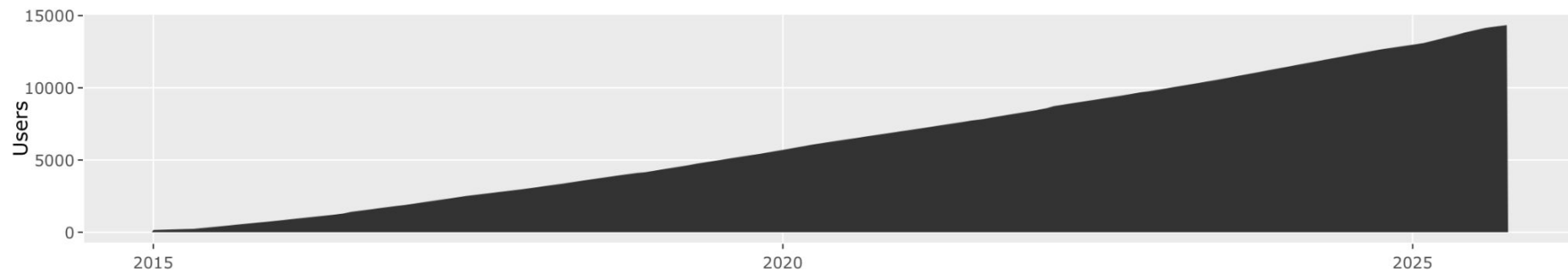
Yes



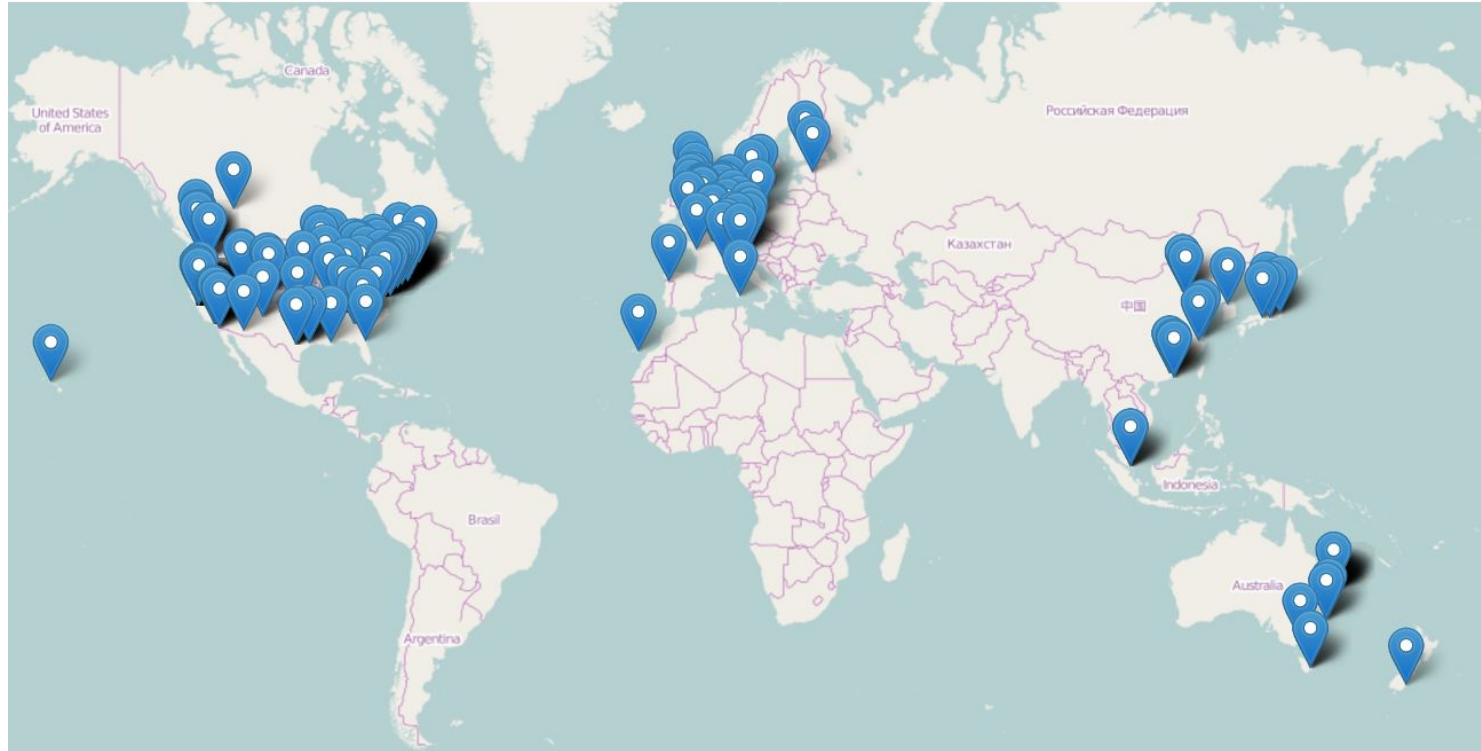
No



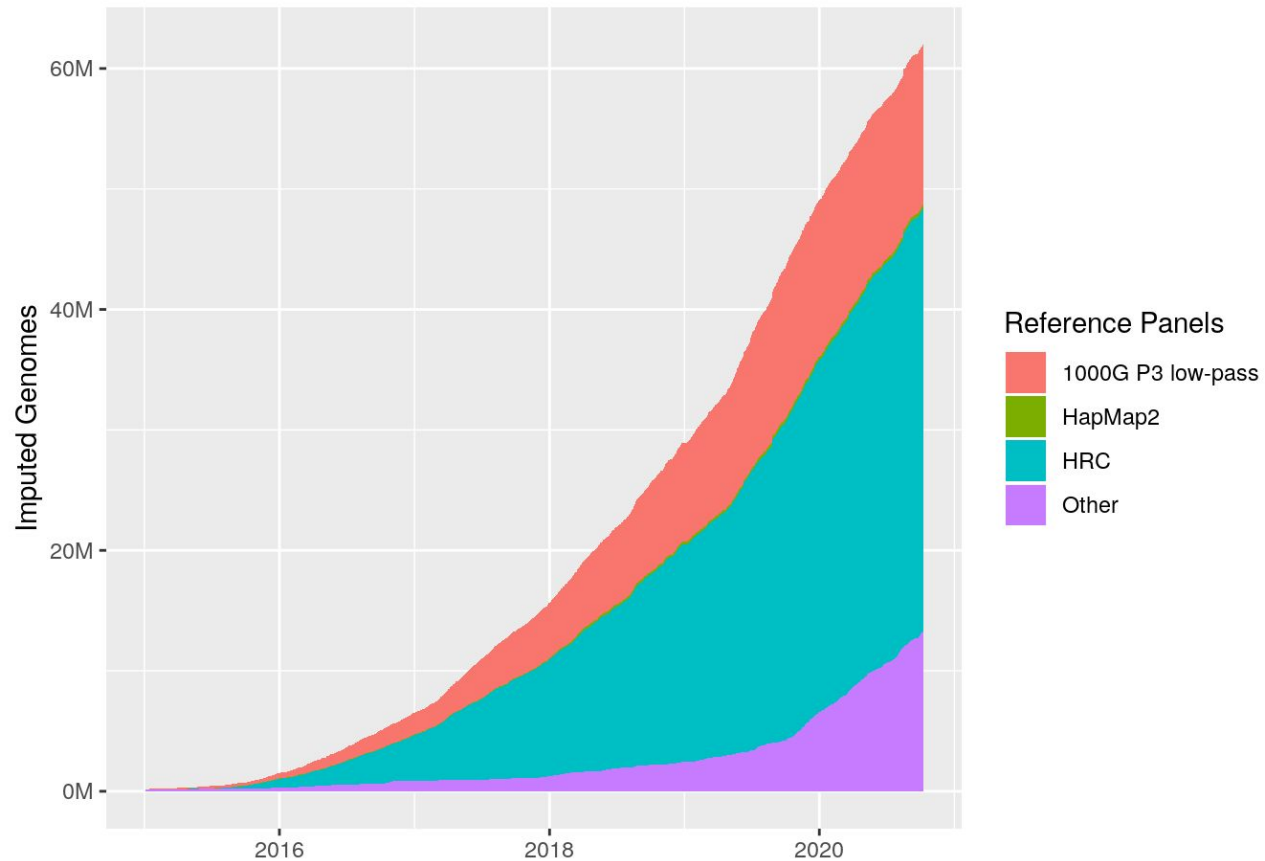
>14,000 users



Used by researcher world-wide



>120M imputed genomes



Summary

Genotype imputation key method in GWAS

Michigan imputation server is easy to use and ensures high quality imputation

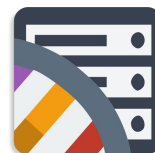
Cloud-services will accelerate genetic research so we can devote our time to more interesting tasks

More info and FAQ can be found here:

<https://genepi.github.io/michigan-imputationserver/>

Section 2

Run a job, Data Preparation and Server Interaction



MICHIGAN
IMPUTATIONSERVER



Sebastian Schönherr
Medical University of Innsbruck
sebastian.schoenherr@i-med.ac.at

Learning objectives

Participants will learn

1. How to submit a genotype imputation job
2. How to prepare your input data
3. Different ways to interact with imputation servers

To run an imputation / PGS jobs, several imputation servers are available (1)

- Michigan Imputation Server (2015)
<https://imputationserver.sph.umich.edu>
- TOPMed Imputation Server (2020)
<https://imputation.biodatacatalyst.nhlbi.nih.gov>
- Helmholtz Munich Imputation Server (2024)
<https://imputationserver.helmholtz-munich.de>
- Mexico City Prospective Study Imputation Server (2025)
<https://imputationserver-reg.sph.umich.edu/>

To run an imputation / PGS jobs, several imputation servers are available (2)

- All servers are built on the same software stack
 - version 1: Hadoop based
 - version 2: Nextflow based (since 2024)
- Each server offers different reference panels and applications
- For this workshop, we provide a hands-on tutorial using the Michigan Imputation Server (MIS) instance

Which genotype imputation service have you utilized to date?

Michigan Imputation Server

0%

TOPMed Imputation Server

0%

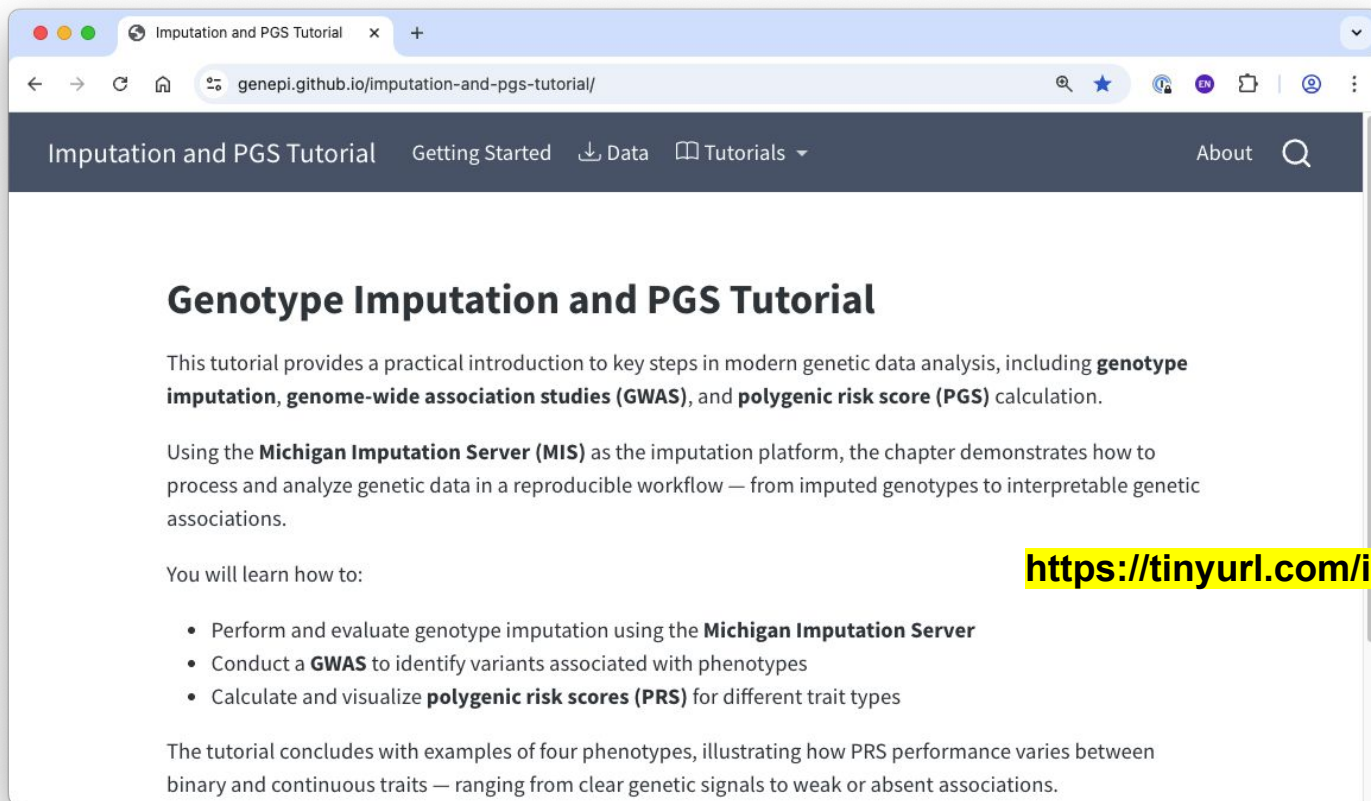
Munich Imputation Server

0%

Other Imputation Service

0%

Hands-On Tutorial



The screenshot shows a web browser window with the address bar displaying 'genepi.github.io/imputation-and-pgs-tutorial/'. The page title is 'Imputation and PGS Tutorial'. The navigation bar includes links for 'Getting Started', 'Data', 'Tutorials', and 'About'. The main content area features the title 'Genotype Imputation and PGS Tutorial' and a description of the tutorial's scope, which includes genotype imputation, GWAS, and PGS calculation. It mentions the Michigan Imputation Server (MIS) as the platform used. A list of learning objectives is provided, and the tutorial concludes with examples of four phenotypes. A QR code is visible on the right side of the page, and a URL is highlighted in a yellow box.

Imputation and PGS Tutorial | Getting Started | Data | Tutorials | About

Genotype Imputation and PGS Tutorial

This tutorial provides a practical introduction to key steps in modern genetic data analysis, including **genotype imputation**, **genome-wide association studies (GWAS)**, and **polygenic risk score (PGS)** calculation.

Using the **Michigan Imputation Server (MIS)** as the imputation platform, the chapter demonstrates how to process and analyze genetic data in a reproducible workflow — from imputed genotypes to interpretable genetic associations.

You will learn how to:

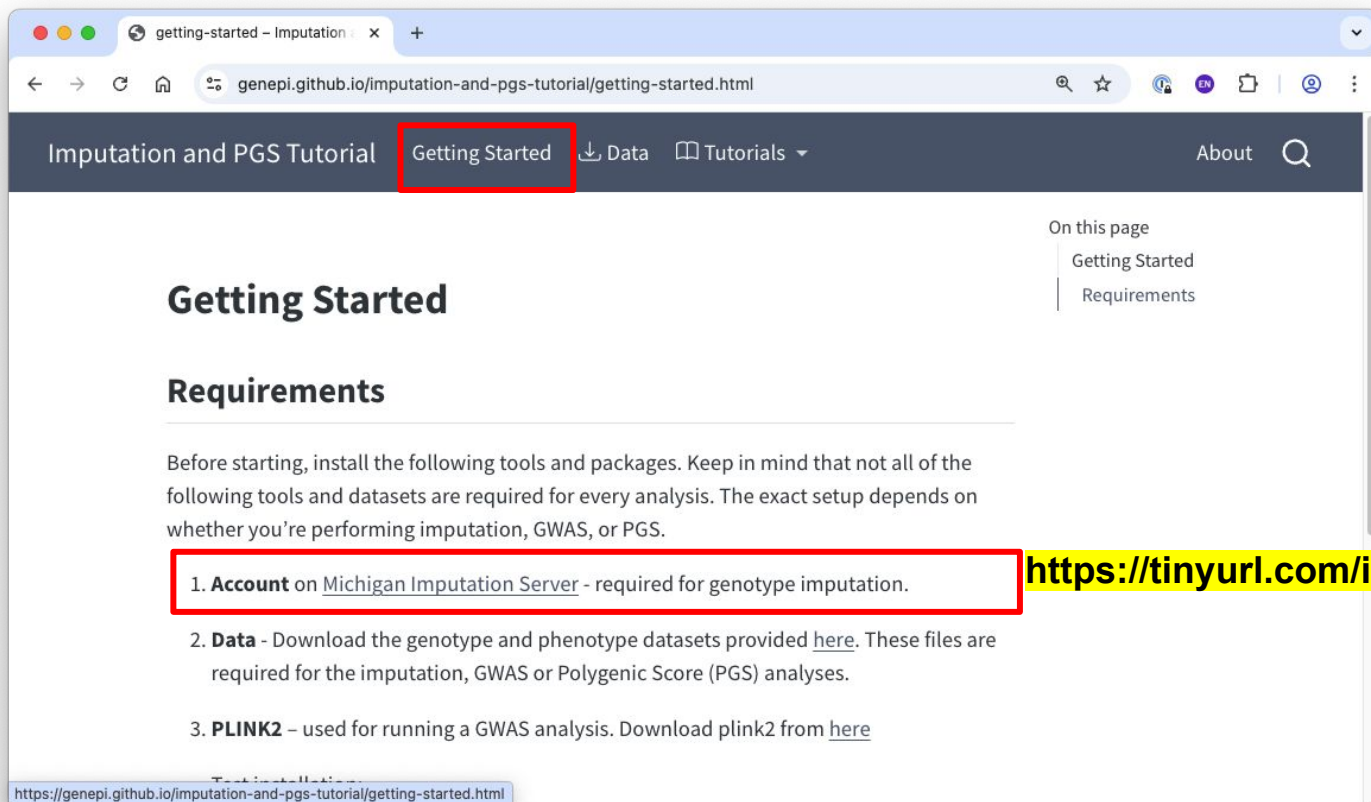
- Perform and evaluate genotype imputation using the **Michigan Imputation Server**
- Conduct a **GWAS** to identify variants associated with phenotypes
- Calculate and visualize **polygenic risk scores (PRS)** for different trait types

The tutorial concludes with examples of four phenotypes, illustrating how PRS performance varies between binary and continuous traits — ranging from clear genetic signals to weak or absent associations.

<https://tinyurl.com/imputationserver-2025>



Running Imputation: Register an Account



Imputation and PGS Tutorial

Getting Started

On this page

- Getting Started
- Requirements

Getting Started

Requirements

Before starting, install the following tools and packages. Keep in mind that not all of the following tools and datasets are required for every analysis. The exact setup depends on whether you're performing imputation, GWAS, or PGS.

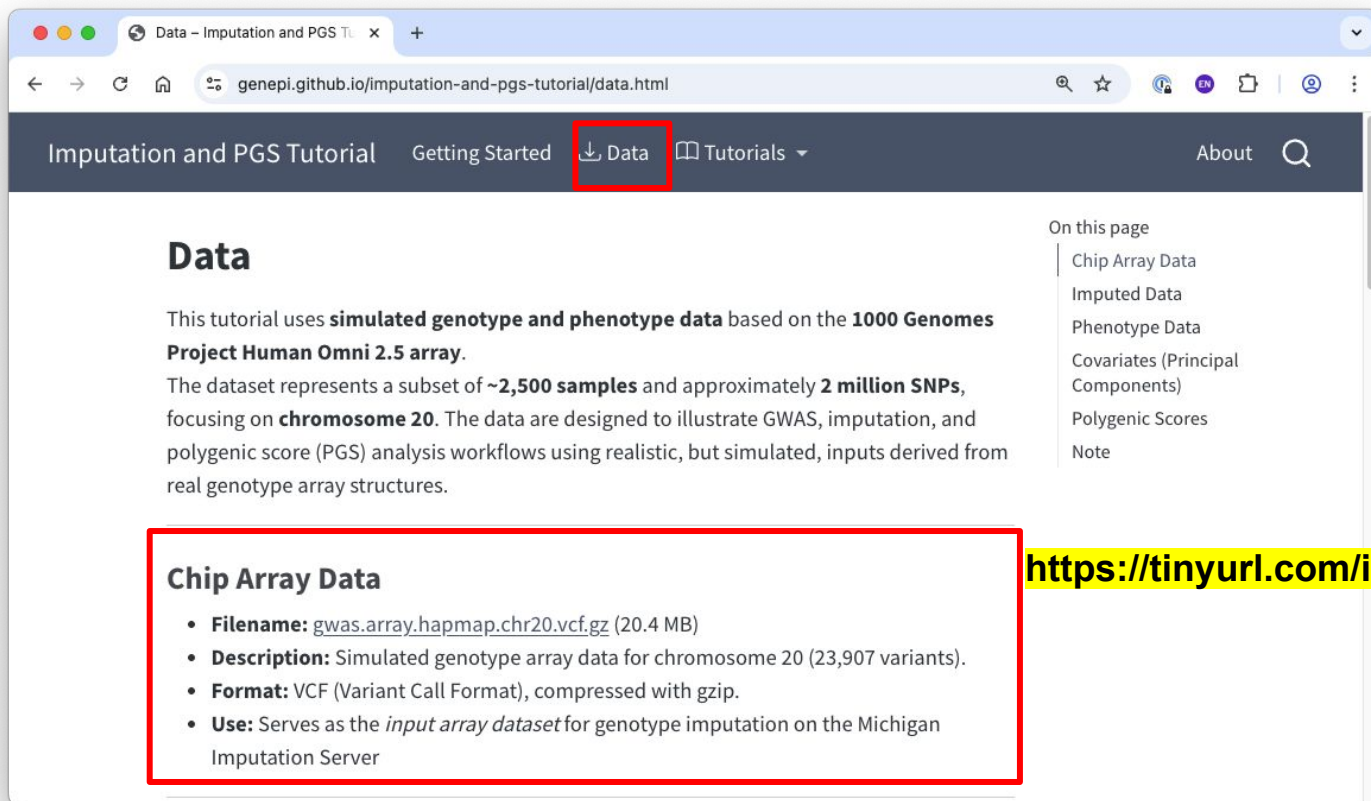
1. **Account** on [Michigan Imputation Server](#) - required for genotype imputation.
2. **Data** - Download the genotype and phenotype datasets provided [here](#). These files are required for the imputation, GWAS or Polygenic Score (PGS) analyses.
3. **PLINK2** - used for running a GWAS analysis. Download plink2 from [here](#)

<https://genepi.github.io/imputation-and-pgs-tutorial/getting-started.html>



<https://tinyurl.com/imputationserver-2025>

Running Imputation: Download Array Data



The screenshot shows a web browser window with the URL `genepi.github.io/imputation-and-pgs-tutorial/data.html`. The navigation bar at the top has links for "Imputation and PGS Tutorial", "Getting Started", "Data" (highlighted with a red box), and "Tutorials". The main content area is titled "Data" and contains the following text:

This tutorial uses **simulated genotype and phenotype data** based on the **1000 Genomes Project Human Omni 2.5 array**. The dataset represents a subset of **~2,500 samples** and approximately **2 million SNPs**, focusing on **chromosome 20**. The data are designed to illustrate GWAS, imputation, and polygenic score (PGS) analysis workflows using realistic, but simulated, inputs derived from real genotype array structures.

On the right side, there is a sidebar titled "On this page" with links to "Chip Array Data", "Imputed Data", "Phenotype Data", "Covariates (Principal Components)", "Polygenic Scores", and "Note".

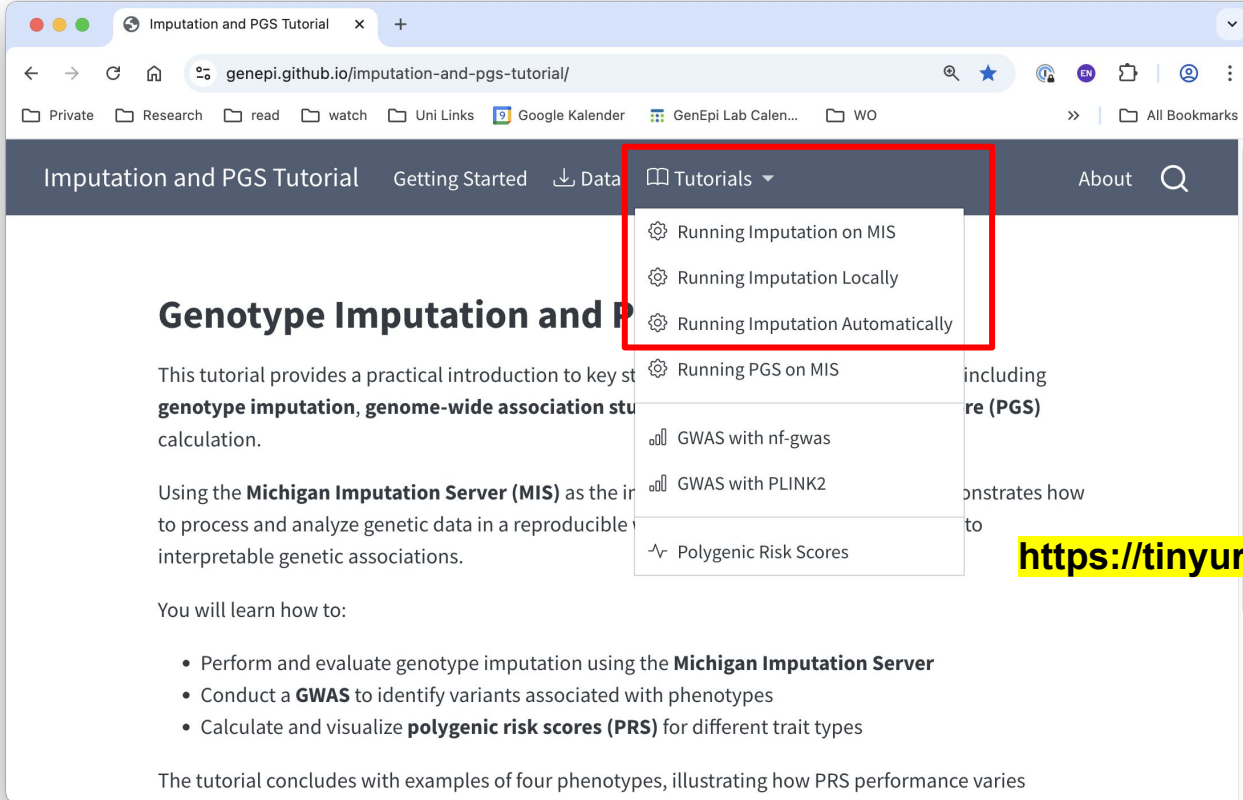
A red box highlights the "Chip Array Data" section, which contains the following information:

- **Filename:** `gwas.array.hapmap.chr20.vcf.gz` (20.4 MB)
- **Description:** Simulated genotype array data for chromosome 20 (23,907 variants).
- **Format:** VCF (Variant Call Format), compressed with gzip.
- **Use:** Serves as the *input array dataset* for genotype imputation on the Michigan Imputation Server



<https://tinyurl.com/imputationserver-2025>

Running Imputation on MIS



The screenshot shows a web browser window with the URL `genepi.github.io/imputation-and-pgs-tutorial/`. The page title is "Imputation and PGS Tutorial". A dark blue navigation bar contains links for "Getting Started", "Data", "Tutorials", and "About". The "Tutorials" dropdown menu is open, showing a list of options: "Running Imputation on MIS", "Running Imputation Locally", "Running Imputation Automatically", "Running PGS on MIS", "GWAS with nf-gwas", "GWAS with PLINK2", and "Polygenic Risk Scores". The "Running Imputation on MIS" option is highlighted with a red box. The main content area has the heading "Genotype Imputation and PGS" and a paragraph: "This tutorial provides a practical introduction to key steps in **genotype imputation**, **genome-wide association studies**, and **polygenic risk scores** calculation. Using the **Michigan Imputation Server (MIS)** as the imputation server, this tutorial demonstrates how to process and analyze genetic data in a reproducible and interpretable genetic associations. You will learn how to:

- Perform and evaluate genotype imputation using the **Michigan Imputation Server**
- Conduct a **GWAS** to identify variants associated with phenotypes
- Calculate and visualize **polygenic risk scores (PRS)** for different trait types

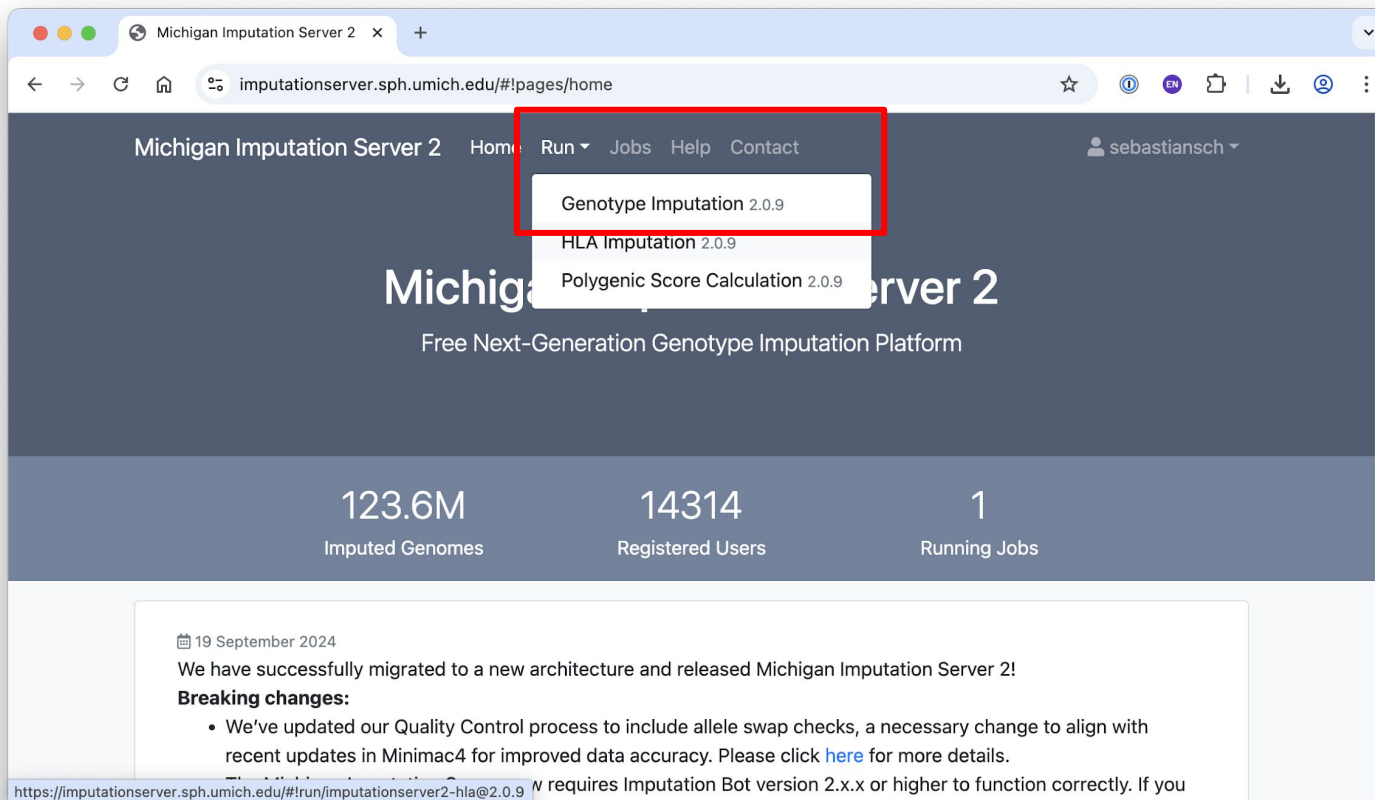
The tutorial concludes with examples of four phenotypes, illustrating how PRS performance varies



<https://tinyurl.com/imputationserver-2025>

Running Imputation on MIS

Running Imputation: Submit a first job



The screenshot shows the Michigan Imputation Server 2 website. The browser's address bar displays `imputationserver.sph.umich.edu/#!/pages/home`. The navigation bar includes links for Home, Run, Jobs, Help, and Contact, along with a user profile for 'sebastiansch'. The 'Run' dropdown menu is open, showing three options: 'Genotype Imputation 2.0.9', 'HLA Imputation 2.0.9', and 'Polygenic Score Calculation 2.0.9'. The main content area features the text 'Michigan Imputation Server 2' and 'Free Next-Generation Genotype Imputation Platform'. Below this, three statistics are displayed: '123.6M Imputed Genomes', '14314 Registered Users', and '1 Running Jobs'. A news section dated '19 September 2024' announces the migration to the new architecture and the release of Michigan Imputation Server 2. It lists 'Breaking changes' and mentions updates to the Quality Control process and Minimac4. A footer note states that the server requires Imputation Bot version 2.x.x or higher to function correctly.

Michigan Imputation Server 2 Home Run Jobs Help Contact sebastiansch

Genotype Imputation 2.0.9
HLA Imputation 2.0.9
Polygenic Score Calculation 2.0.9

Michigan Imputation Server 2
Free Next-Generation Genotype Imputation Platform

123.6M Imputed Genomes 14314 Registered Users 1 Running Jobs

19 September 2024
We have successfully migrated to a new architecture and released Michigan Imputation Server 2!
Breaking changes:

- We've updated our Quality Control process to include allele swap checks, a necessary change to align with recent updates in Minimac4 for improved data accuracy. Please click [here](#) for more details.

<https://imputationserver.sph.umich.edu/#!/run/imputationserver2-hla@2.0.9> requires Imputation Bot version 2.x.x or higher to function correctly. If you

Upload Data and Select Reference Panel

Michigan Imputation Server 2 x +

imputationserver.sph.umich.edu/#!run/imputationserver2%402.0.9

Michigan Imputation Server 2 Home Run Jobs Help Contact sebastiansch

Genotype Imputation 2.0.9

This is the new Michigan Imputation Server Pipeline using Minimac4. Documentation can be found [here](#).

If your input data is **GRCh37/hg19** please ensure chromosomes are encoded with chr1-22,X,Y

If your input data is **GRCh38/hg38** please ensure chromosomes are encoded with chr1-22,X,Y

<https://imputationserver.readthedocs.io>

Run

Name test-job

1 Reference Panel (Details) HapMap 2 (GRCh37/hg19)

2 Input Files (VCF) gwas.array.hapmap.chr20.vcf.gz

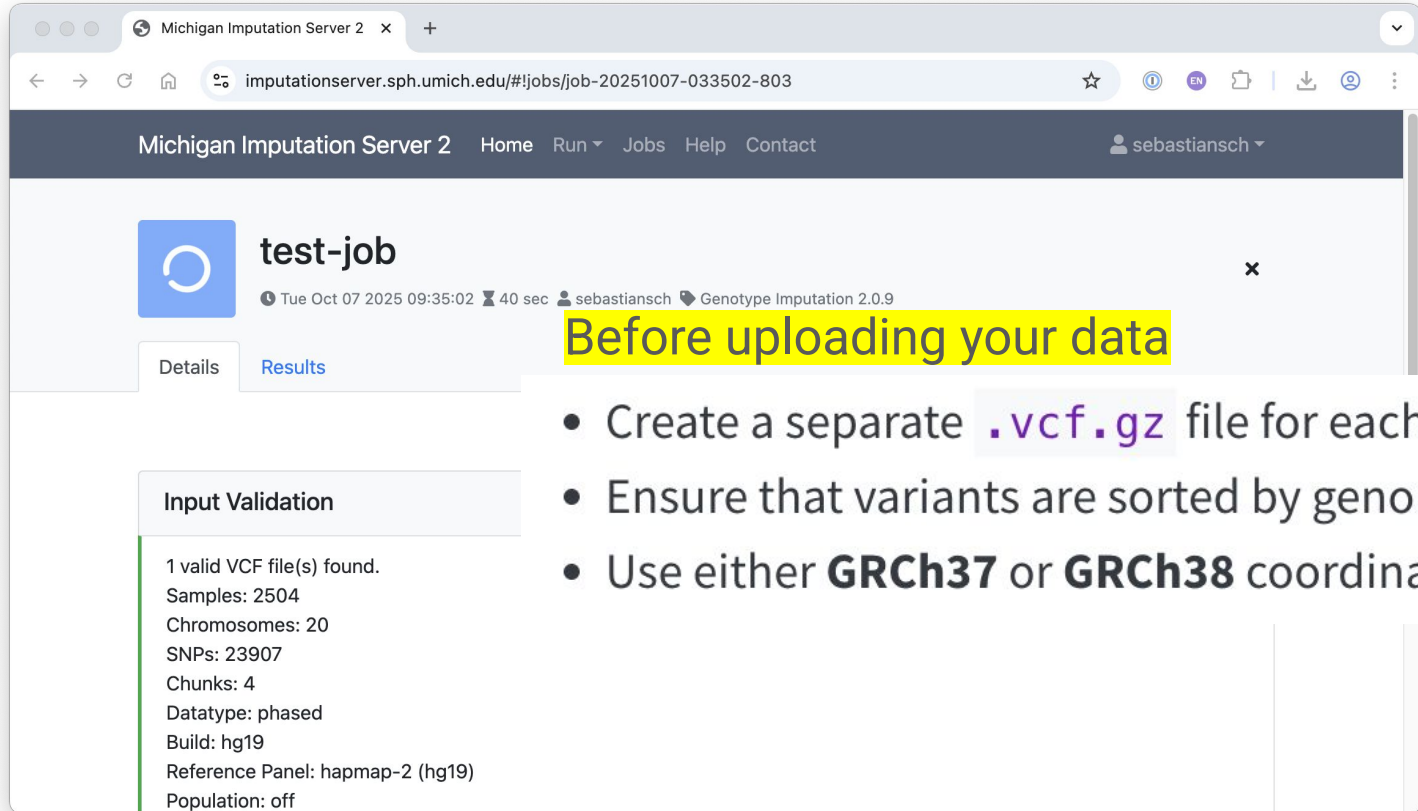
3 Array Build GRCh37/hg19

Please note that the final SNP coordinates always match the reference build.

1000G Phase 1 v3 Shapeit2 (no singletons) (GRCh37/hg19)
1000G Phase 3 (GRCh38/hg38) [BETA]
1000G Phase 3 30x (GRCh38/hg38) [BETA]
1000G Phase 3 v5 (GRCh37/hg19)
CAAPA African American Panel (GRCh37/hg19)
Genome Asia Pilot - GAsP (GRCh37/hg19)
HRC r1.1 2016 (GRCh37/hg19)
HapMap 2 (GRCh37/hg19)
Samoan (GRCh37/hg19)

PANELS

Step 1: Input Validation

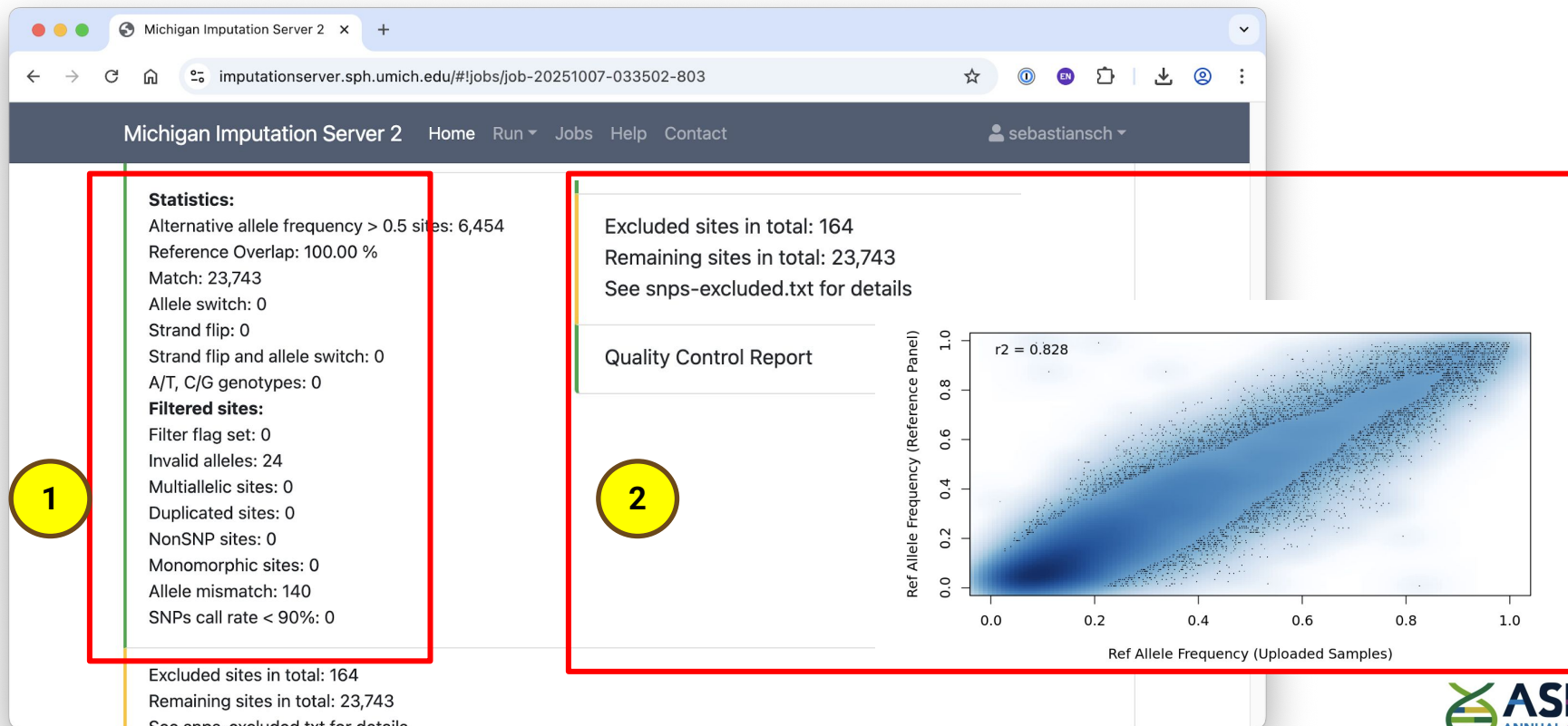


The screenshot shows the Michigan Imputation Server 2 web interface. The browser address bar displays the URL: `imputationserver.sph.umich.edu/#!jobs/job-20251007-033502-803`. The page header includes navigation links: Home, Run, Jobs, Help, and Contact, along with a user profile for 'sebastiansch'. The main content area shows a job titled 'test-job' with a status of 'Tue Oct 07 2025 09:35:02' and a duration of '40 sec'. The job is associated with 'sebastiansch' and 'Genotype Imputation 2.0.9'. Below the job title, there are tabs for 'Details' and 'Results'. The 'Results' tab is active, showing the 'Input Validation' section. This section reports: '1 valid VCF file(s) found.', 'Samples: 2504', 'Chromosomes: 20', 'SNPs: 23907', 'Chunks: 4', 'Datatype: phased', 'Build: hg19', 'Reference Panel: hapmap-2 (hg19)', and 'Population: off'.

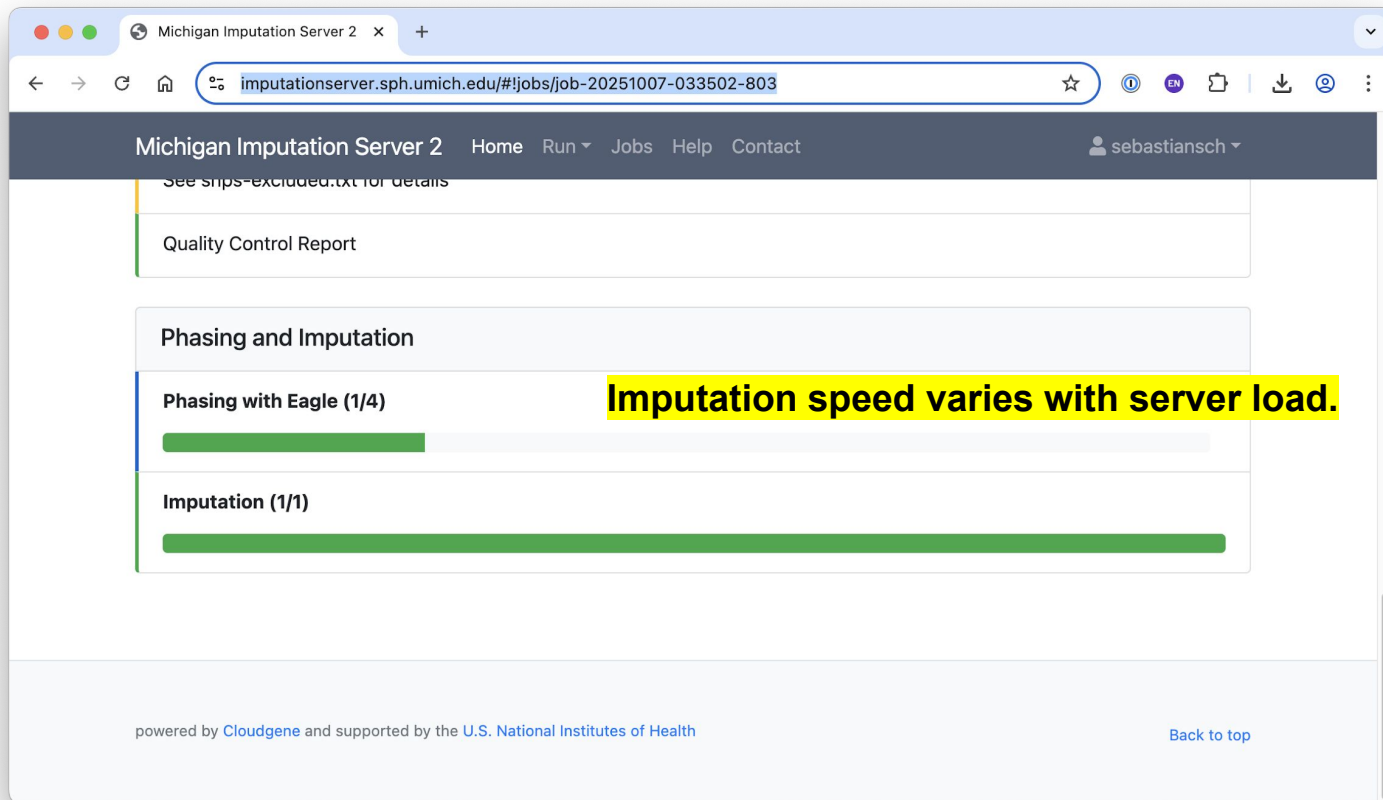
Before uploading your data

- Create a separate `.vcf.gz` file for each chromosome.
- Ensure that variants are sorted by genomic position.
- Use either **GRCh37** or **GRCh38** coordinates.

Step 2: Quality Control



Step 3 + 4: Phasing, Imputation, Encryption



Download Results

The screenshot shows a web browser window with the URL `imputationserver.sph.umich.edu/#ijobs/job-20251007-033502-803/results`. The page title is "Michigan Imputation Server 2" and the user is logged in as "sebastiansch". The job is named "test-job" and is marked as successful with a green checkmark. The job details show it was run on Tue Oct 07 2025 at 09:35:02, took 33 min 18 sec, and used Genotype Imputation 2.0.9. The "Results" tab is selected, showing a "Downloads" section with a list of files: `chr_20.zip` (694 MB), `qc_report.txt` (674 bytes), and `quality-control.html` (1 MB). A red box highlights this list. A yellow banner with the text "Decrypt chr_20.zip using the password sent via email." is overlaid on the right side of the download list.

Michigan Imputation Server 2 Home Run Jobs Help Contact sebastiansch

test-job

Tue Oct 07 2025 09:35:02 33 min 18 sec sebastiansch Genotype Imputation 2.0.9

Details Results

Downloads wget

- > statistics
 - chr_20.zip (694 MB)
 - qc_report.txt (674 bytes)
 - quality-control.html (1 MB)

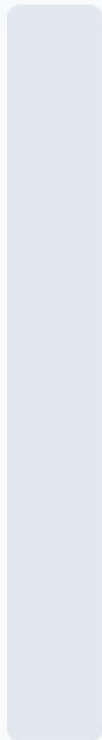
Decrypt chr_20.zip using the password sent via email.

How many jobs are failing?

- 40% in 2015; 20% in 2019; 7% 2020-2024; >10% 2025
- Reason for job failures: Something wrong with your input data **or** phasing/imputation issue on our side

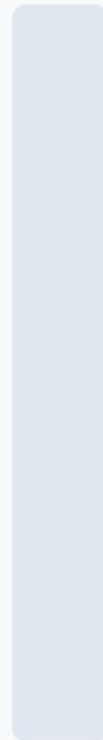
For users: Have you ever encountered a failed imputation job?

0%



True

0%



False

Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app

Input Validation

Input Validation

The provided VCF file is malformed. Error during index creation: [tabix] was bgzip used to compress this file? (see [Help](#)).

Input Validation

The provided VCF file contains more than one chromosome. Please split your input VCF file by chromosome (see [Help](#)).

Input Validation

Unable to parse header with error: Your input file has a malformed header: We never saw the required CHROM header line (starting with one #) for the input VCF file (see [Help](#)).

QC Step

1

Excluded sites in total: 695
Remaining sites in total: 185,791
See [snps-excluded.txt](#) for details
Typed only sites: 397
See [typed-only.txt](#) for details

Warning: 2 Chunk(s) excluded: reference c
Remaining chunk(s): 40

Error: More than 100 obvious strand flips have been detected. Please check strand. Imputation cannot be sta

2

Excluded sites in total: 11,088
Remaining sites in total: 1,968
See snps-excluded.txt for details
Typed only sites: 848,376
See [typed-only.txt](#) for details

Warning: 153 Chunk(s) excluded: reference overlap < 50.0% (see chunks-excluded.txt for details).

Remaining chunk(s): 0

Error: No chunks passed the QC step. Imputation cannot be started!

3

Excluded sites in total: 165,335
Remaining sites in total: 161,526
See snps-excluded.txt for details
Warning: 2 Chunk(s) excluded: < 20 SNPs (see chunks-excluded.txt for details).
Remaining chunk(s): 152

Error: More than 100 allele switches have been detected. Imputation cannot be started!

How to fix input files?

Imputation Preparation Tool

TUTORIAL AVAILABLE ON MIS

- Developed by W. Rayner
- Works for all major reference panels (HRC, TOPMed, Asia, CAAPA, 1000G)
- Checks for consistency between input data and a reference panel
- Updates/removes SNPs, Updates strand, position and ref/alt assignment
- Input Data in PLINK Binary Format (bim, bed, fam)

<https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.3.0.zip>

Uploaded my data 5 mins ago and my job is still waiting...

There is a problem with MIS - email MIS team to let them know

0%

There is a problem with my data

0%

MIS is very busy - check again later

98%

Don't know

2%

Error: No chunks passed the QC step. Imputation cannot be started!

Email MIS team



4%

Data are on the wrong build



35%

Too few variants overlap with the reference panel



52%

Must be a MIS problem, since imputation runs locally



2%

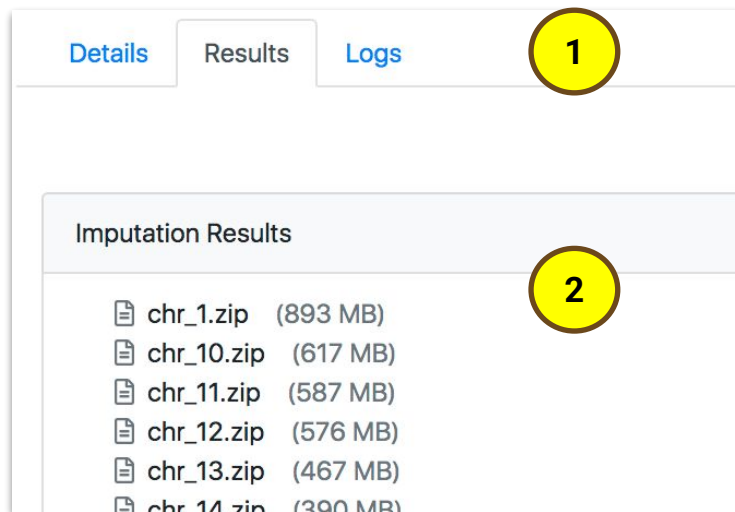
Don't know



8%

How to download
the imputed genotypes?

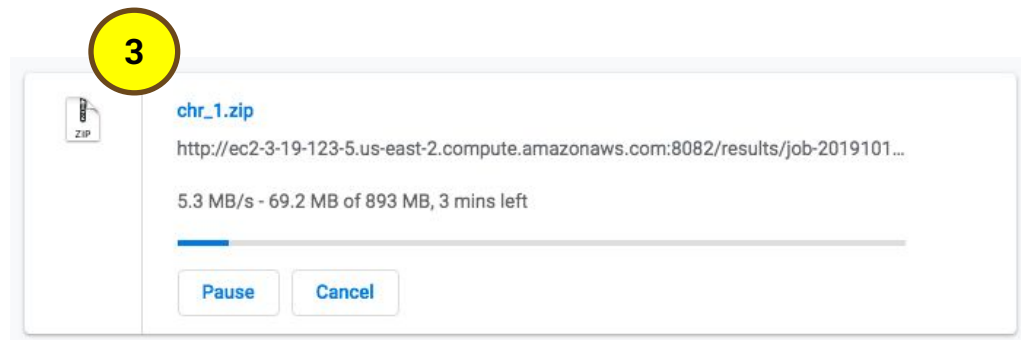
Option 1: Web-Interface



Details Results **1** Logs

Imputation Results **2**

chr_1.zip	(893 MB)
chr_10.zip	(617 MB)
chr_11.zip	(587 MB)
chr_12.zip	(576 MB)
chr_13.zip	(467 MB)
chr_14.zip	(390 MB)



3

chr_1.zip

<http://ec2-3-19-123-5.us-east-2.compute.amazonaws.com:8082/results/job-2019101...>

5.3 MB/s - 69.2 MB of 893 MB, 3 mins left

Pause Cancel

Option 2: Batch Download

Imputation Results  wget

- chr_1.zip (469 MB)
- chr_10.zip (287 MB)
- chr_11.zip (281 MB)
- chr_12.zip (269 MB)
- chr_13.zip (195 MB)
- chr_14.zip (192 MB)
- chr_15.zip (181 MB)
- chr_16.zip (203 MB)
- chr_17.zip (187 MB)
- chr_18.zip (160 MB)
- chr_19.zip (170 MB)
- chr_2.zip (471 MB)
- chr_20.zip (129 MB)

Download data

wget (22) [URLs \(22\)](#)

```
wget https://imputationserver.sph.umich.edu/share/results/1fc3d1b  
wget https://imputationserver.sph.umich.edu/share/results/3d9f5f0  
wget https://imputationserver.sph.umich.edu/share/results/528e41f  
wget https://imputationserver.sph.umich.edu/share/results/ed598ab  
wget https://imputationserver.sph.umich.edu/share/results/7c818b4  
wget https://imputationserver.sph.umich.edu/share/results/1c1e65d
```

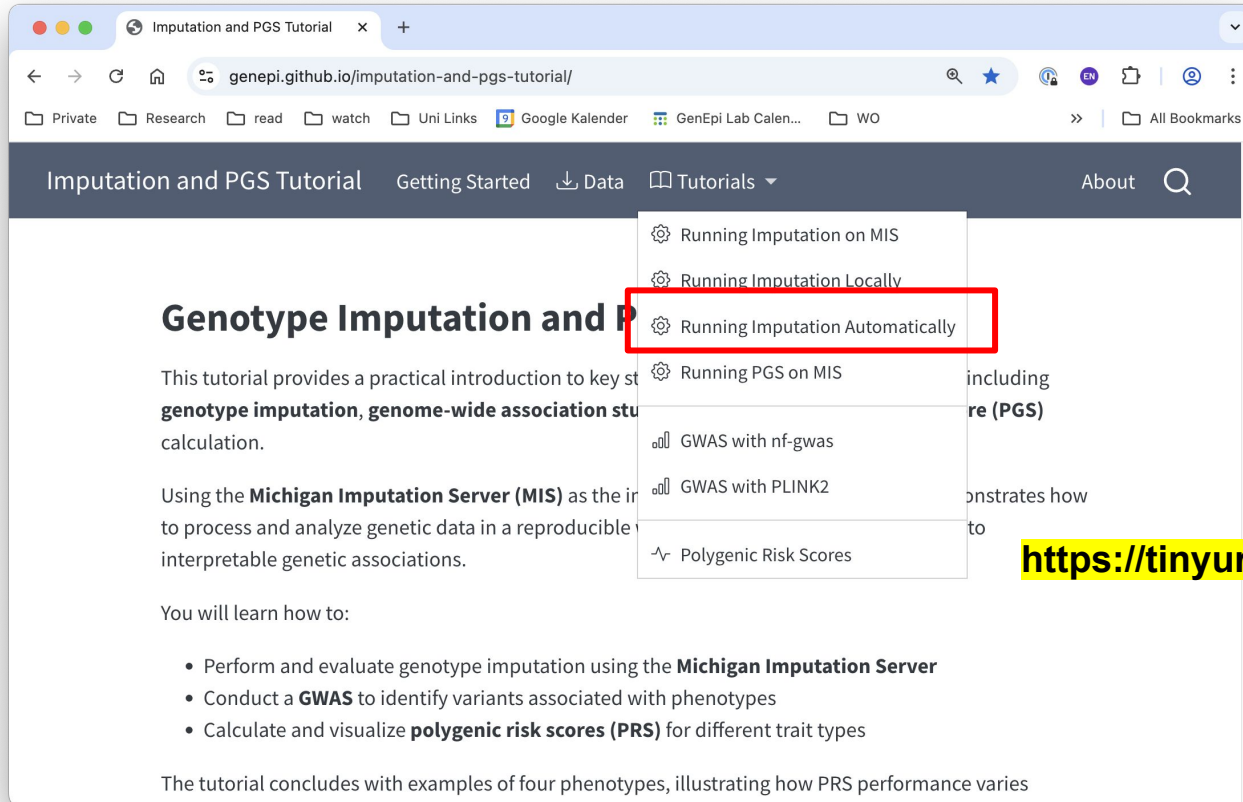
Use the following command to download all results at once:

```
curl -sL https://imputationserver.sph.umich.edu/get/1584555,
```



OK

Option 3: Use Imputationbot



The screenshot shows a web browser window with the URL `genepi.github.io/imputation-and-pgs-tutorial/`. The page title is "Imputation and PGS Tutorial". A dropdown menu is open from the "Tutorials" link in the navigation bar, showing several options. The option "Running Imputation Automatically" is highlighted with a red rectangle. The main content area has the heading "Genotype Imputation and PGS" and a paragraph: "This tutorial provides a practical introduction to key steps in **genotype imputation**, **genome-wide association studies**, and **polygenic risk score** calculation." Below this, it says "Using the **Michigan Imputation Server (MIS)** as the imputation server, this tutorial demonstrates how to process and analyze genetic data in a reproducible and interpretable genetic associations." A list of topics to learn is provided: "Perform and evaluate genotype imputation using the **Michigan Imputation Server**", "Conduct a **GWAS** to identify variants associated with phenotypes", and "Calculate and visualize **polygenic risk scores (PRS)** for different trait types". The page concludes with "The tutorial concludes with examples of four phenotypes, illustrating how PRS performance varies".

Imputation and PGS Tutorial Getting Started Data Tutorials About

Genotype Imputation and PGS

This tutorial provides a practical introduction to key steps in **genotype imputation**, **genome-wide association studies**, and **polygenic risk score** calculation.

Using the **Michigan Imputation Server (MIS)** as the imputation server, this tutorial demonstrates how to process and analyze genetic data in a reproducible and interpretable genetic associations.

You will learn how to:

- Perform and evaluate genotype imputation using the **Michigan Imputation Server**
- Conduct a **GWAS** to identify variants associated with phenotypes
- Calculate and visualize **polygenic risk scores (PRS)** for different trait types

The tutorial concludes with examples of four phenotypes, illustrating how PRS performance varies

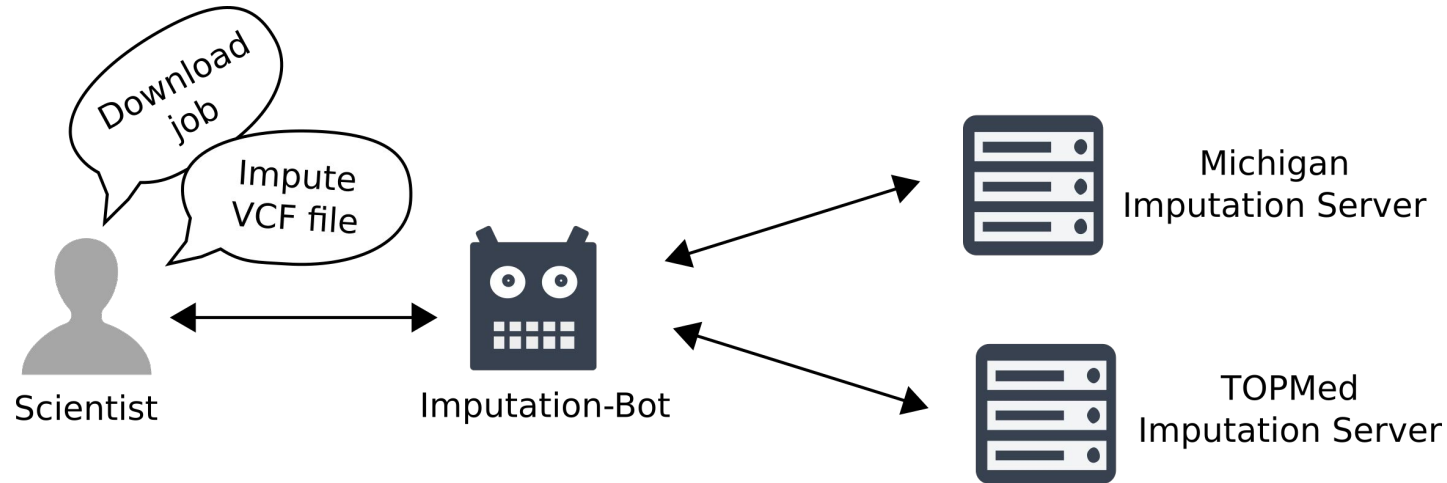
Running Imputation on MIS
Running Imputation Locally
Running Imputation Automatically
Running PGS on MIS
GWAS with nf-gwas
GWAS with PLINK2
Polygenic Risk Scores

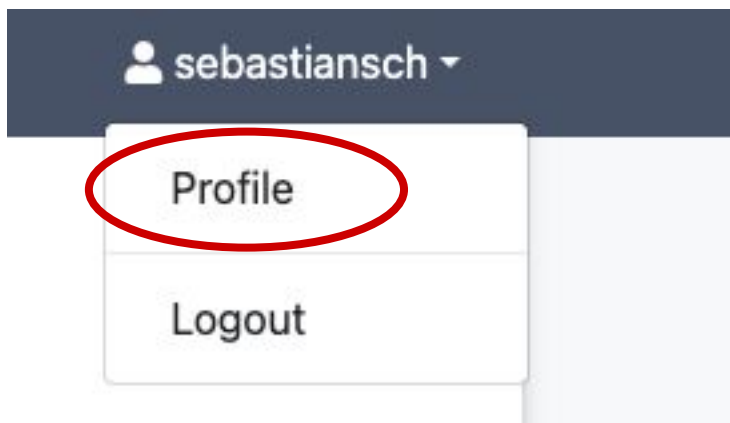


<https://tinyurl.com/imputationserver-2025>

Imputation Bot

- Automate remote imputation
- Submit and monitor jobs from the command line
- Different commands can easily be combined





API Access

This service provides a rich RestAPI to submit, monitor and download jobs.

You need a access token to use the API. [Learn more.](#)

Create API Token

Expires in

30 days



API Token

Make sure to copy your personal access token now. **You won't be able to see it again:**

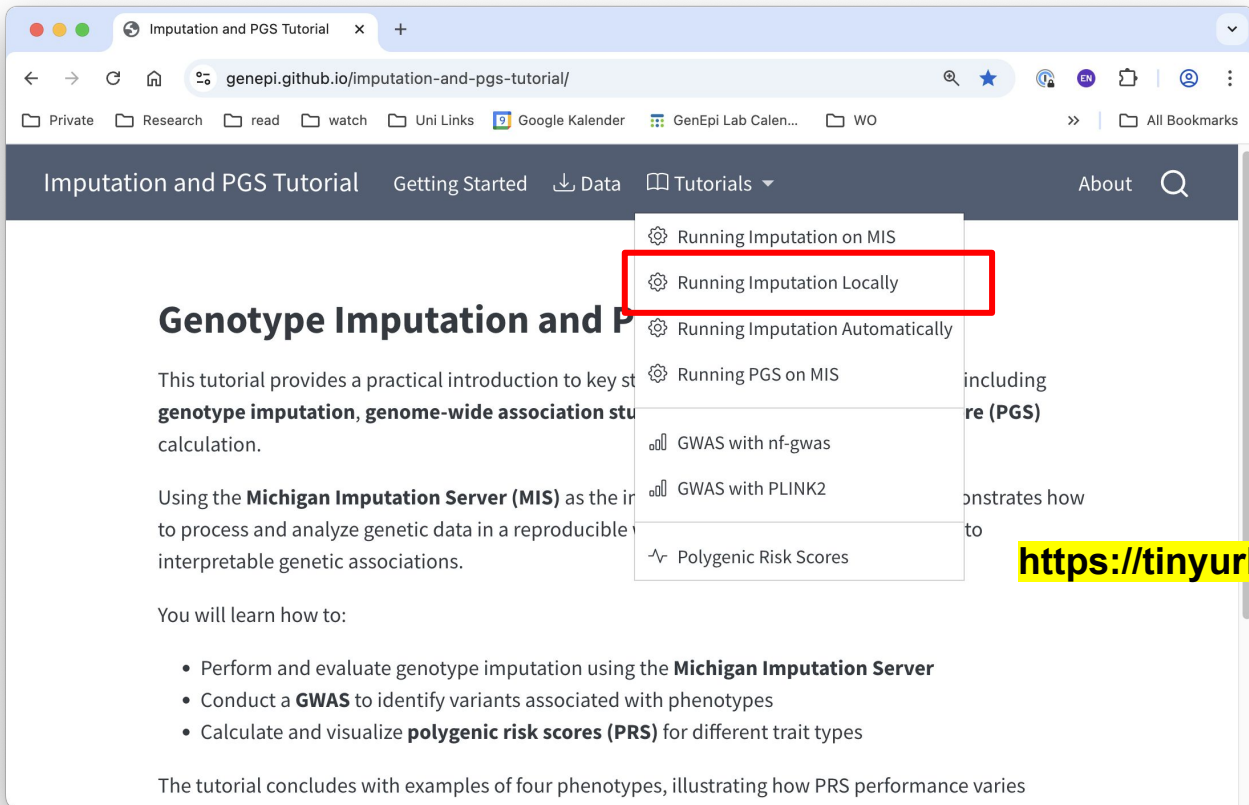
```
eyJhbGciOiJIUzI1NiJ9.eyJzdWUiOiJzZWJhc3RpYW5zY2giLCJ1YmYiOiJlE3NTk4NDczMTgslm1haWwiOiJzZWluc2Nob2VuaGVyckBnbWFpbC5jb20iLCJhcGlfaGFzaCI6Im44SXFHQ2lwN1lQamlwS0pkSjJManU5eVRwazhxZSIsInJvbGVzIjpbXSwiaXNzI
```

OK

Workflow

- `curl -sL imputationbot.now.sh/v2 | bash`
- `./imputationbot add-instance`
- `./imputationbot impute --files gwas.array.hapmap.chr20.vcf.gz --refpanel hapmap-2 --population eur`
- `./imputationbot download job-20251007-104433-914 --password ABCD`

Running Imputation Locally



The screenshot shows a web browser window with the URL `geneipi.github.io/imputation-and-pgs-tutorial/`. The page title is "Imputation and PGS Tutorial". A dropdown menu is open from the "Tutorials" link in the navigation bar, showing several options. The option "Running Imputation Locally" is highlighted with a red rectangular box. The main content area of the page is titled "Genotype Imputation and PGS" and contains introductory text about the tutorial's purpose and a list of topics to be covered.

Imputation and PGS Tutorial Getting Started Data Tutorials About

Genotype Imputation and PGS

This tutorial provides a practical introduction to key steps in **genotype imputation**, **genome-wide association studies (GWAS)**, and **polygenic risk scores (PRS)** calculation.

Using the **Michigan Imputation Server (MIS)** as the imputation server, this tutorial demonstrates how to process and analyze genetic data in a reproducible and interpretable manner.

You will learn how to:

- Perform and evaluate genotype imputation using the **Michigan Imputation Server**
- Conduct a **GWAS** to identify variants associated with phenotypes
- Calculate and visualize **polygenic risk scores (PRS)** for different trait types

The tutorial concludes with examples of four phenotypes, illustrating how PRS performance varies

Running Imputation on MIS

Running Imputation Locally

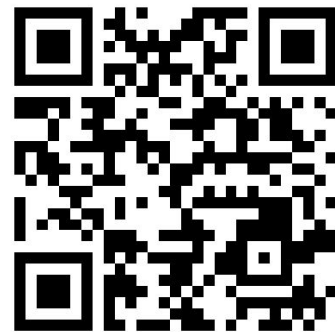
Running Imputation Automatically

Running PGS on MIS

GWAS with nf-gwas

GWAS with PLINK2

Polygenic Risk Scores



<https://tinyurl.com/imputationserver-2025>

Running Imputation Locally



- Since 2024 MIS Workflow is based on Nextflow
- Nextflow is a workflow management system for scalable and reproducible bioinformatics pipelines
- Requirements
 - Install Nextflow
 - Use your own/publicly available reference panel
 - Run Imputation on the commandline

Executors

- AWS Batch
- Azure Batch
- Bridge
- Flux Executor
- Google Cloud Batch
- HTCondor
- HyperQueue
- Kubernetes

Running Imputation Locally



```
# Create temp directory
mkdir imputationserver
cd imputationserver
# Download Input Data
wget https://genepi.i-med.ac.at/downloads/imputation/gwas.array.hapmap.chr20.vcf.gz
# Download and unzip Reference Panel
wget https://imputationserver.sph.umich.edu/resources/ref-panels/imputationserver2-l
unzip imputationserver2-hapmap2.zip -d hapmap2
```

1

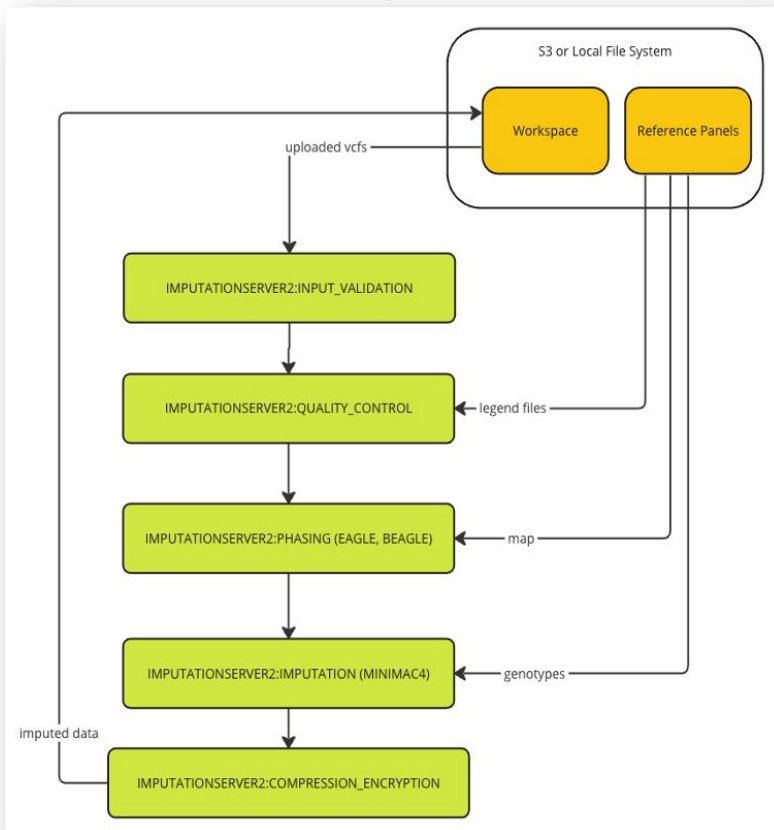
```
params {
  project          = "my-test-project"
  build            = "hg19"
  files            = "gwas.array.hapmap.chr20.vcf.gz"
  allele_frequency_population = "eur"
  mode            = "imputation"
  refpanel_yaml    = "hapmap2/imputation-hapmap2.yaml"
  output          = "results"
  encryption.enabled = false
}
```

2

```
nextflow run genepi/imputationserver2 -r v2.0.9 -c imputation.config -profile docker
```

3

Running Imputation Locally



```
NEXTFLOW ~ version 24.04.2

Launching `main.nf` [fabulous_mendel] DSL2 - revision: 16c60661b1

Loading reference panel from file /Users/seb/repositories/nf-imputationserver/tests/data
Welcome to Imputation Server 2 (v2.0.6)
executor > local (5)
[0a/d99e6b] INPUT_VALIDATION:INPUT_VALIDATION_VCF [100%] 1 of 1 ✓
[91/a9fe51] QUALITY_CONTROL:QUALITY_CONTROL_VCF [100%] 1 of 1 ✓
[cf/093d23] QUALITY_CONTROL:QUALITY_CONTROL_REPORT [100%] 1 of 1 ✓
[6c/fd767c] PHASING:EAGLE (chunk_20_0040000001_0060000000.vcf.gz) [25%] 1 of 4
[-] IMPUTATION:MINIMAC4 [0%] 0 of 1
[-] ENCRYPTION:COMPRESSION_ENCRYPTION_VCF -
```

Summary

- MIS Web Interface provides a fast and reliable way to impute data
- We provide tutorials to run imputation on MIS, locally and via the imputationbot
- MIS applies a strict Quality Control with the goal to return high quality imputation data

Section 3

Performing GWAS using imputed data



Cassandra Spracklen
University of Massachusetts
cspracklen@umass.edu

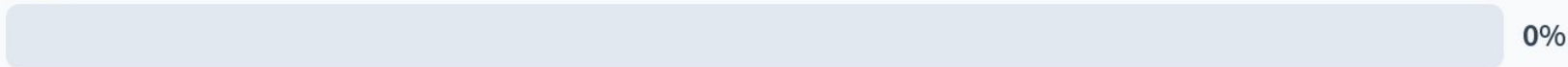
Learning objectives

Participants will learn to:

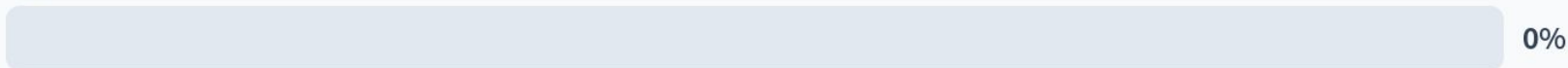
- Identify and understand the use of variant imputation quality information following imputation in the MIS
- Distinguish between some of the available options for GWAS
- Troubleshoot common GWAS errors

Have you ever performed a GWAS?

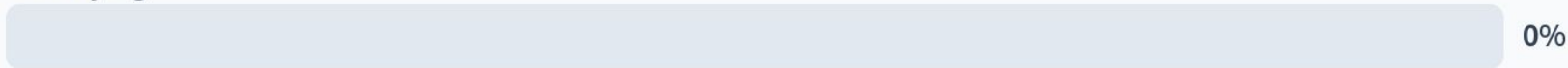
Yes



No



I'm trying



Imputation Quality

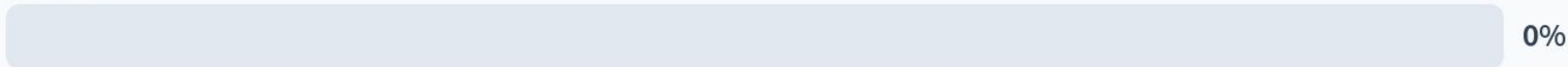
- For each variant, how confident can we be that the imputation dosages are sufficiently “accurate” for association analyses?
- Measure of confidence in imputed dosages: “Rsq” column [range 0-1]

SNP	REF(0)	ALT(1)	ALT_Frq	MAF	AvgCall	Rsq	Genotyped	...
20:61795:G:T	G	T	0.26318	0.26318	0.88455	0.54658	Imputed	...
20:63231:T:G	T	G	0.03843	0.03843	0.98342	0.67736	Imputed	...
20:63244:A:C	A	C	0.16132	0.16132	0.91761	0.49907	Imputed	...

From a chr20.info.gz file

Minimally accepted R_{sq} value for common ($MAF \geq 5\%$) variants?

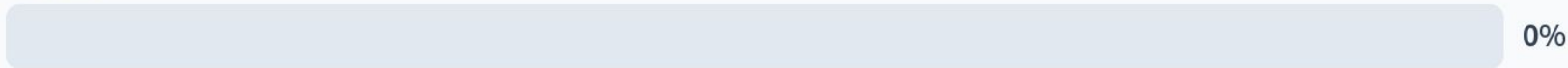
≥ 0.10



≥ 0.30



≥ 0.50

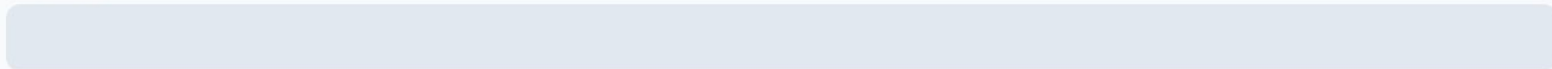


≥ 0.70



Minimally accepted Rsq value for low frequency (MAF<5%) and rare variants?

≥ 0.10



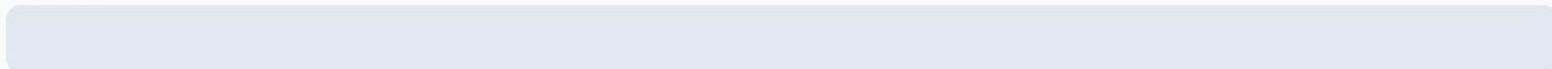
0%

≥ 0.30



0%

≥ 0.50



0%

≥ 0.70



0%



Imputation Quality

- Minimal R^2 value for common variants
 - ≥ 0.30
- Minimal R^2 value for low frequency/rare variants
 - ≥ 0.50
- Before performing GWAS, remove variants that do not meet these thresholds
 - Suggested program: VCFtools
 - Saves computational time when performing GWAS

Which GWAS program(s) have you used?

Performing the GWAS

- Each program has its own input, output formats, and options
- Typical input files
 - Genotype file (.vcf; .bgen; .bed/.bim/.fam)
 - Phenotype/covariate file (.txt; .ped)
 - Some programs use separate phenotype and covariate files
 - Kinship/relationship matrix (EPACTS, SAIGE)

Available GWAS Programs

No File Reformatting (VCF from MIS)

- EPACTS
- Rvtests
- SNPTTEST
- SAIGE

File Formatting Required

- BOLT-LMM
- BGENIE
- regenie
- PLINK

**May not be all-inclusive*

Each GWAS Program Has Strengths, Limitations

EPACTS/Rvtests

- + Many model options - single variant, gene-based
- + Chr X analyses
- + Phenotypic transformation (e.g inverse normal; Rvtests only)
- + Linear mixed model for sample relatedness (quantitative traits only)
- + Generate covariance matrices for downstream analyses (e.g conditional analyses; Rvtests only)
- Memory intensive
- Sample size $\sim \leq 20,000$ (better $\leq 10,000$)

EPACTS: <https://genome.sph.umich.edu/wiki/EPACTS>

Rvtests: <https://genome.sph.umich.edu/wiki/Rvtests>

Each GWAS Program Has Strengths, Limitations

SAIGE

- + Similar to Rvtests, but for very large sample sizes (e.g. biobanks)
- + Able to account for sample relatedness for binary traits
- + Designed to handle unbalanced number of cases and controls
- + Chr X analyses
- Should not be used to examine heritability (biased variance estimates)
- Computational time can vary widely between phenotypes and sample sizes
- Can be conservative for extremely unbalanced case and control ratio
- Odds ratios estimated to conserve computational time

SAIGE: <https://github.com/weizhouUMICH/SAIGE>

Each GWAS Program Has Strengths, Limitations

SNPTEST

- + Frequentist and bayesian methods supported
- + Chr X analyses
- Limited to unrelated individuals
- Computationally intensive

SNPTEST: <https://www.well.ox.ac.uk/~gav/snptest/#introduction>

Each GWAS Program Has Strengths, Limitations

BOLT-LMM/BGENIE/Regenie

- + Great for very large sample sizes (e.g. biobanks)
- + Chr X analyses
- + Computationally efficient (Regenie)
- Requires files to be in BGEN or PLINK format
- Nextflow pipeline for regenie using VCF: <https://github.com/genepi/nf-gwas>
- Not optimal for extremely unbalanced case control ratio (especially with rare variants)

BOLT-LMM: <https://data.broadinstitute.org/alkesgroup/BOLT-LMM/#x1-5600011>

BGENIE: <https://jmarchini.org/bgenie/>

Regenie: <https://github.com/rgcgithub/regenie>

Each GWAS Program Has Strengths, Limitations

PLINK

- + Quick
- + Multiple versions; often as intermediary tool to the other programs
- + Can run on the command line (unix not required)
- + Chr X analyses
- Requires files to be in PLINK format (.bed/.bim/.fam)
- Limited model options

PLINK: <https://www.cog-genomics.org/plink/2.0/>

Summary of common GWAS analysis tools

	EPACTS	Rvtests	SNPTEST	SAIGE	BOLT-LMM	Bgenie	Regenie
Input VCF	Y	Y	Y				
Sample relatedness (Quantitative outcome)	Y	Y		Y	Y	Y	Y
Sample relatedness (Binary outcome)				Y		Y	Y
Case control imbalance				Y			Y
Large sample size (>20,000)				Y	Y	Y	Y

Which program(s) would be best?

A researcher new to genetic analyses and unfamiliar to the UNIX environment wants to perform a GWAS on total cholesterol using a cohort of 5,500 unrelated individuals.

EPACTS/Rvtests

0%

SAIGE

0%

BOLT-LMM

0%

PLINK

0%



Which program(s) would be best?

Researchers want to perform a GWAS using a cohort of 10,000 individuals with household based recruitment (i.e. includes related individuals).

EPACTS/Rvtests

0%

SAIGE

0%

BOLT-LMM

0%

PLINK

0%



Which program(s) would be best?

Researchers want to perform a GWAS using data from BioBank Japan (>200,000 individuals)

(A) EPACTS/Rvtests

0%

(B) SAIGE

0%

(C) BOLT-LMM

0%

(D) PLINK

0%

(E) Regenie/Bgenie

0%



Common Errors When Running a GWAS

- Wording of error messages vary by program, but the same issues will cause errors throughout all of the program
- [Unix] Errors independent of GWAS program
 - File permissions
 - Correct by changing file permissions
 - Directory/file not found
 - Correct by making sure all of the file locations and names are accurate
 - Not enough memory/time
 - Correct by restarting job with adequate memory/time allocation

Common Errors When Running a GWAS

- Common errors
 - IDs don't match
 - Correct by ensuring that the ID in the phenotype, genotype, covariance, kinship matrix are consistent format in all files

Common Errors When Running a GWAS

- Common errors
 - IDs don't match
 - Correct by ensuring that the ID in the phenotype, genotype, covariance, kinship matrix are consistent format in all files
 - File format(s) incorrect
 - Correct by making sure the format of all files are as the program is expecting (e.g. columns, delimiters, headers, file extension)

Common Errors When Running a GWAS

- Common errors
 - IDs don't match
 - Correct by ensuring that the ID in the phenotype, genotype, covariance, kinship matrix are consistent format in all files
 - File format(s) incorrect
 - Correct by making sure the format of all files are as the program is expecting (e.g. columns, delimiters, headers, file extension)
 - Improperly specified options/command
 - Correct by checking all needed options are specified, correct order, no typos

Common Errors When Running a GWAS

- Common errors
 - IDs don't match
 - Correct by ensuring that the ID in the phenotype, genotype, covariance, kinship matrix are consistent format in all files
 - File format(s) incorrect
 - Correct by making sure the format of all files are as the program is expecting (e.g. columns, delimiters, headers, file extension)
 - Improperly specified options/command
 - Correct by checking all needed options are specified, correct order, no typos
 - Peripheral programs not available (e.g. R with EPACTS, SAIGE)
 - Correct by installing other peripheral programs

Common Errors When Running a GWAS

- Common errors
 - IDs don't match
 - Correct by ensuring that the ID in the phenotype, genotype, covariance, kinship matrix are consistent format in all files
 - File format(s) incorrect
 - Correct by making sure the format of all files are as the program is expecting (e.g. columns, delimiters, headers, file extension)
 - Improperly specified options/command
 - Correct by checking all needed options are specified, correct order, no typos
 - Peripheral programs not available (e.g. R with EPACTS, SAIGE)
 - Correct by installing other peripheral programs
 - Invalid estimate (e.g. heritability in BOLT-LMM)
 - Sample too related and/or sample size too small
 - Correct by using a different program

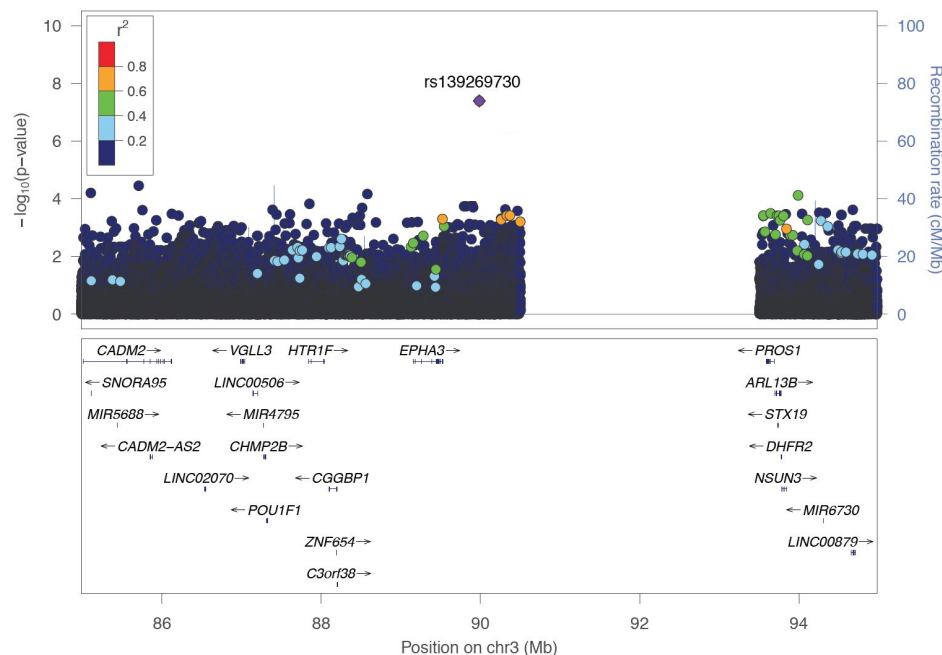
Interpreting GWAS Results

- GWAS results must be carefully reviewed for:

- Imputation quality
- Genomic inflation
- False positives

- Replication datasets

- PheWas



Sex chromosomes

- Humans have two sex chromosomes, X and Y, that in combination determine the sex of an individual.
 - To ensure balanced expression, one of the female's X-chromosome is randomly selected to undergo inactivation (either paternal or maternal determined at early embryonic development).
 - Pseudoautosomal regions (PAR1 and PAR2) are short regions on the X- and Y-chromosomes that are inherited in an autosomal rather than a sex-linked manner

X-chromosome - genotype calls and QC

- Genotype calls on X-chromosomes
 - One copy in males, two copies in females
- Clarity of genotype calls on X-Chromosome
 - 0/1/2 allele coding in females; 0/1 or 0/2 allele coding in males
 - 0/0.5/1 allele coding in females; 0/0.5 allele coding in males
 - Comparing standard error with autosomal data
- Analysis plan to be specific on allelic coding
- Develop QC checks before association analyses, e.g., similar allele frequencies between males and females (differential calls bias)

X-chromosome - imputation

Chromosome X Pipeline

Additionally to the standard QC, the following per-sample checks are executed for chrX:

- Ploidy Check: Verifies if all variants in the nonPAR region are either haploid or diploid.
- Mixed Genotypes Check: Verifies if the amount of mixed genotypes (e.g. 1/.) is < 10 %.

For phasing and imputation, chrX is split into three independent chunks (PAR1, nonPAR, PAR2). These splits are then automatically merged by Michigan Imputation Server and are returned as one complete chromosome X file. Only Eagle is supported.

Interpretation of association

- Biological considerations:
 - Assumption of total X-inactivation but there is evidence of escape and compensation mechanisms
 - Different inheritance/activation in different tissues?
 - Sex-biased gene expression? Unclear if it is real differences in gene expression or influence of hormones.
- Methodology development
 - Statistical methods to estimate the inactivation when estimating the effects; adjusting for sex

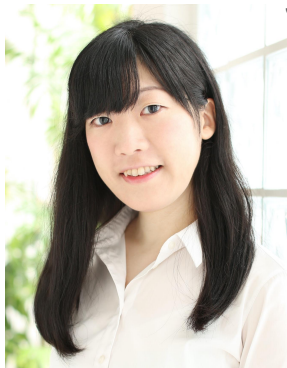
Summary

- Variants must be filtered post-imputation to remove those with poor imputation quality
- There are many GWAS programs available, each with their own strengths and limitations - so be sure to pick one that fits your analyses needs
- As these GWAS programs are widely used or adopted by consortia, there are tutorials and help-pages available
- X-chr imputation and analyses are also possible!

More info and FAQ can be found here:
<https://imputationserver.readthedocs.io>

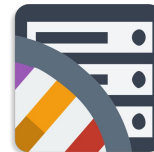
HLA Imputation

Slides Created By:



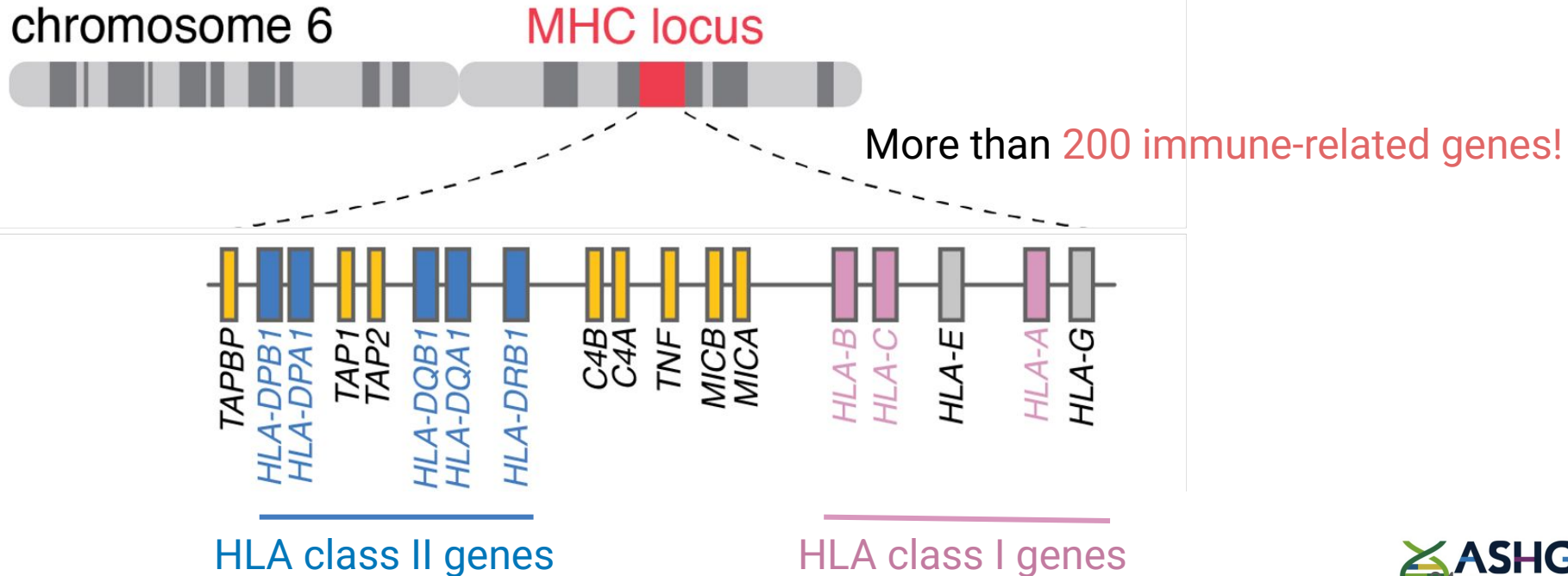
Saori Sakaue
Broad Institute

ssakaue@broadinstitute.org

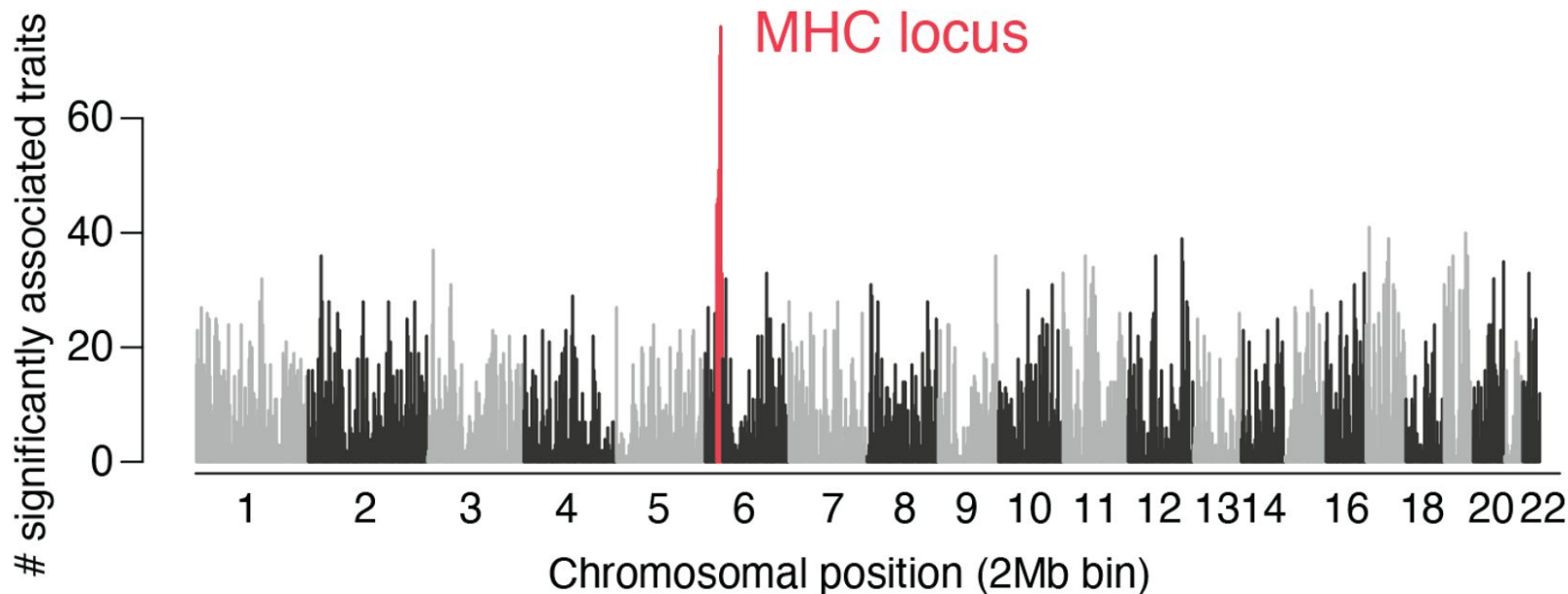


MICHIGAN
IMPUTATIONSERVER

HLA genes are within MHC (major histocompatibility complex) locus on chromosome 6



MHC locus confers the largest number of associations of any locus genome-wide



In UK Biobank, Sakaue et al. Nat Genet 2021, Sakaue et al. Nature Protocols 2023

Genotype Imputation of HLA Region is great, but not perfect

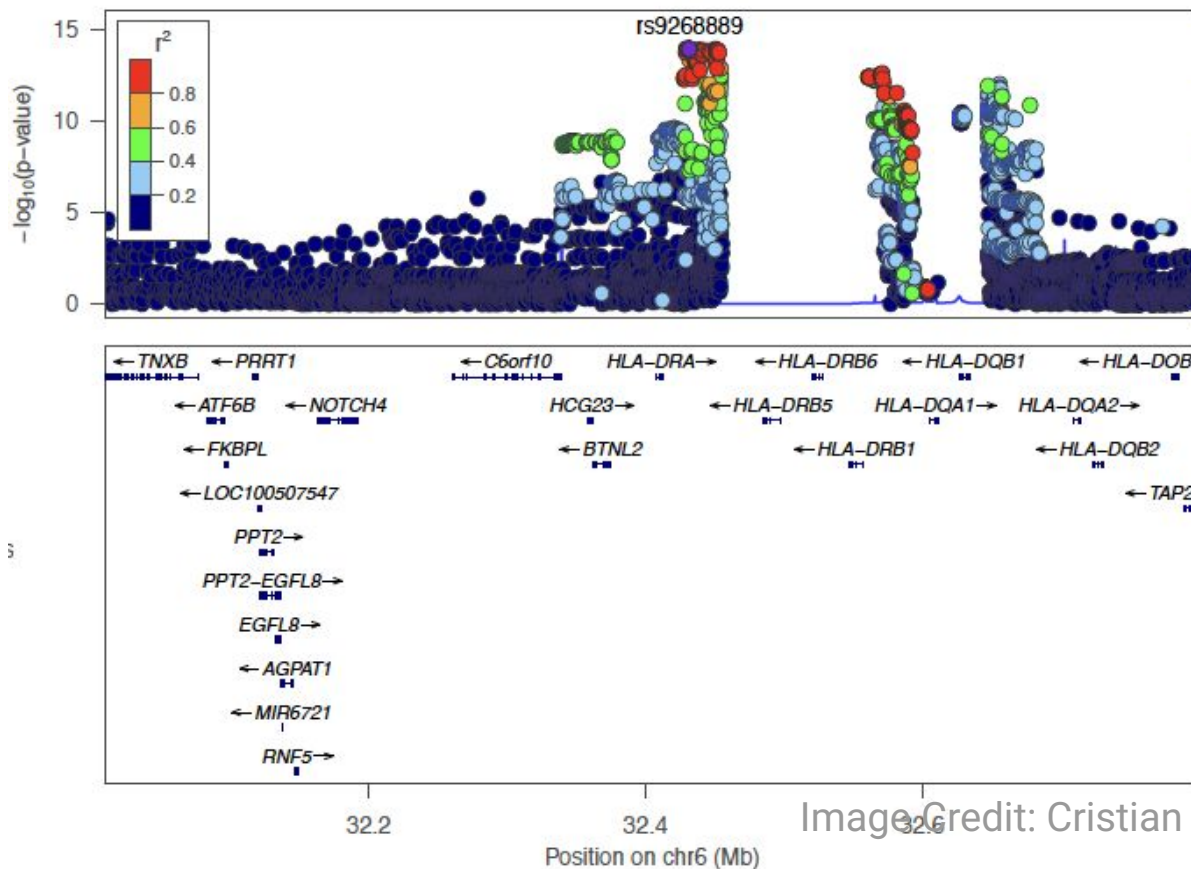


Image Credit: Cristian Valencia

Are you planning to impute the HLA region?

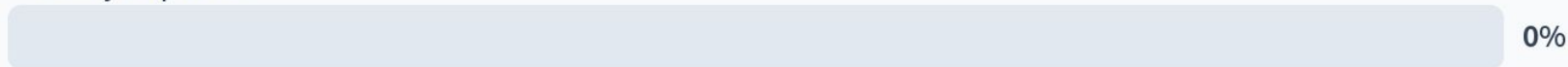
No



Yes



Already imputed!



Genotype imputation in a nutshell

Genotype imputation

Known Genotyped SNPs

Given Reference haplotype with
whole-genome SNPs

Unknown
(Impute!) Untyped SNPs

HLA imputation in a nutshell

Genotype imputation

Known Genotyped SNPs

Given Reference haplotype with whole-genome SNPs

Unknown (Impute!) Untyped SNPs

HLA imputation

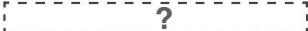
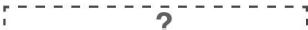
Genotyped SNPs

Reference haplotype with HLA alleles and amino acid sequences

Untyped HLA alleles and amino acid sequences

HLA imputation in a nutshell

Known SNP genotype of the target cohort

Genotype in the MHC region	HLA alleles	HLA amino acids	HLA intragenic SNPs
 CGA.ATCT..GTCTTCTGT.CTAA			
 CAA.ATCT..GTCCT.TGT.CTAA			
 CAA.ATTT..TGCTTCAGT.CTAA			

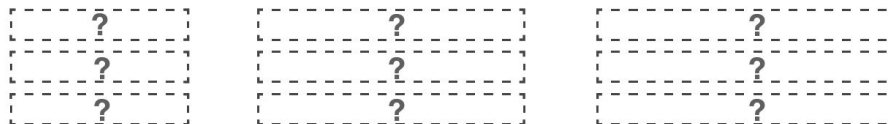
QCed Plink *.{bed,bim,fam}

HLA imputation in a nutshell

Known SNP genotype of the target cohort

Genotype in the MHC region

CGA.ATCT..GTCTTCTGT.CTAA
CAA.ATCT..GTCCT.TGT.CTAA
CAA.ATTT..TGCTTCAGT.CTAA



+

Given HLA imputation reference panel

Scaffold genotype in the MHC	HLA alleles	HLA amino acids	HLA intragenic SNPs
CGAGATCTCAGTCTTCTGTTCTAA	DRB1*04:01	GGSCMAALTVTLMVL	GGAGGACCTGTGAACCA
CAAGATTTCCTTCATCTGTTCTAA	DRB1*01:01	GGSCMTALTVTLMVL	GGAAGACCTGCGAACCA
CGAGATCTCCTGCTTCAGTTCTAA	DRB1*01:02	GGSCMTALTVTLMVL	GGAGGACCTGCGAACCA
CAAGATCTCCGTCCTCTGTTCTAA	DRB1*15:01	GGSCMTALTVTLMVL	GGAAGACCTGTGAACCG



Multi-ancestry panel
(EUR, EAS, SAS, AFR, LAT)

Luo et al. Nat Genet 2021, Sakaue et al. Nat Protocols 2023

HLA imputation in a nutshell

Known SNP genotype of the target cohort

Genotype in the MHC region

CGA.ATCT..GTCTTCTGT.CTAA
CAA.ATCT..GTCCT.TGT.CTAA
CAA.ATTT..TGCTTCAGT.CTAA

?	?	?
?	?	?
?	?	?

+

Given HLA imputation reference panel

Scaffold genotype in the MHC	HLA alleles	HLA amino acids	HLA intragenic SNPs
CGAGATCTCAGTCTTCTGTTCTAA	DRB1*04:01	GGSCMAALTVTLMVL	GGAGGACCTGTGAACCA
CAAGATTTCCTTCATCTGTTCTAA	DRB1*01:01	GGSCMTALTVTLMVL	GGAAGACCTGCGAACCA
CGAGATCTCCTGCTTCAGTTCTAA	DRB1*01:02	GGSCMTALTVTLMVL	GGAGGACCTGCGAACCA
CAAGATCTCCGTCCTCTGTTCTAA	DRB1*15:01	GGSCMTALTVTLMVL	GGAAGACCTGTGAACCG

↓ Haplotype phasing + Imputation

Want Estimation of HLA imputation for the target cohort

CGAGATCTCAGTCTTCTGTTCTAA	DRB1*04:01	GGSCMAALTVTLMVL	GGAGGACCTGTGAACCA
CAAGATCTCCGTCCTCTGTTCTAA	DRB1*15:01	GGSCMTALTVTLMVL	GGAAGACCTGTGAACCG
CAAGATTTCCTGCTTCAGTTCTAA	DRB1*01:01	GGSCMTALTVTLMVL	GGAAGACCTGCGAACCA

HLA imputation in a nutshell

Known SNP genotype of the target cohort

Genotype in the MHC region

CGA.ATCT..GTCTTCTGT.CTAA
CAA.ATCT..GTCCT.TGT.CTAA
CAA.ATTT..TGCTTCAGT.CTAA

?

?

?

+

Given HLA imputation reference panel

Scaffold genotype in the MHC

HLA alleles

HLA amino acids

HLA intragenic SNPs

CGAGATCTCAGTCTTCTGTTCTAA → DRB1*04:01 → GGSCMAALTVTLMVL → GGAGGACCTGTGAACCA
CAAGATTTCCTTCATCTGTTCTAA → DRB1*01:01 → GGSCMTALTVTLMVL → GGAAGACCTGCGAACCA
CGAGATCTCCTGCTTCAGTTCTAA → DRB1*01:02 → GGSCMTALTVTLMVL → GGAGGACCTGCGAACCA
CAAGATCTCCGTCCTCTGTTCTAA → DRB1*15:01 → GGSCMTALTVTLMVL → GGAAGACCTGTGAACCG

↓ Haplotype phasing + Imputation

Want Estimation of HLA imputation for the target cohort

CGAGATCTCAGTCTTCTGTTCTAA → DRB1*04:01 → GGSCMAALTVTLMVL → GGAGGACCTGTGAACCA
CAAGATCTCCGTCCTCTGTTCTAA → DRB1*15:01 → GGSCMTALTVTLMVL → GGAAGACCTGTGAACCG
CAAGATTTCCTGCTTCAGTTCTAA → DRB1*01:01 → GGSCMTALTVTLMVL → GGAAGACCTGCGAACCA

- Beagle
- SHAPEIT
- Eagle
- Beagle
- Minimac

HLA imputation in a nutshell

Known SNP genotype of the target cohort

Genotype in the MHC region

CGA.ATCT..GTCTTCTGT.CTAA
CAA.ATCT..GTCCT.TGT.CTAA
CAA.ATTT..TGCTTCAGT.CTAA

?	?	?
?	?	?
?	?	?

+

Given HLA imputation reference panel

Scaffold genotype in the MHC	HLA alleles	HLA amino acids	HLA intragenic SNPs
CGAGATCTCAGTCTTCTGTTCTAA	DRB1*04:01	GGSCMAALTVTLML	GGAGGACCTGTGAACCA
CAAGATTTCCTTCATCTGTTCTAA	DRB1*01:01	GGSCMTALTVTLML	GGAAGACCTGCGAACCA
CGAGATCTCCTGCTTCAGTTCTAA	DRB1*01:02	GGSCMTALTVTLML	GGAGGACCTGCGAACCA
CAAGATCTCCGTCCTCTGTTCTAA	DRB1*15:01	GGSCMTALTVTLML	GGAAGACCTGTGAACCG

↓ Haplotype phasing + Imputation

Want Estimation of HLA imputation for the target cohort

CGAGATCTCAGTCTTCTGTTCTAA	DRB1*04:01	GGSCMAALTVTLML	GGAGGACCTGTGAACCA
CAAGATCTCCGTCCTCTGTTCTAA	DRB1*15:01	GGSCMTALTVTLML	GGAAGACCTGCGAACCA
CAAGATTTCCTGCTTCAGTTCTAA	DRB1*01:01	GGSCMTALTVTLML	GGAAGACCTGCGAACCA

- Beagle
- SHAPEIT
- Eagle
- Beagle
- Minimac



HLA imputation in MIS

Michigan Imputation Server

Home

Run ▾

Jobs

Help

Contact

saorisakaue ▾

Genotype Imputation HLA

Thank you for using our multi-ethnic HLA imputation server. This server is designed to impute HLA alleles from low-resolution genotyping arrays using a reference panel of HLA sequences from five global populations.

Please cite this manuscript if you would like to use this server: Luo, Y., Kanai, M., Choi, W., Li, X., Yamamoto, T., Haas, D. W., Guo, X., Palmer, N. D., Chen, Y.-D. I., Rotter, J. I., Taylor, K. D., Rich, S., ... Raychaudhuri, S. (2020). **A high-resolution HLA reference panel capturing global population diversity enables multi-ethnic fine-mapping in HIV host response**. <https://doi.org/10.1101/2020.07.16.20155606>

If your input data is **GRCh37/hg19** please ensure chromosomes are encoded without prefix (e.g. **20**).
If your input data is **GRCh38hg38** please ensure chromosomes are encoded with prefix 'chr' (e.g. **chr20**).

<https://imputationserver.readthedocs.io>

▶ Run

Genotype Imputation (Minimac4)

Genotype Imputation and Polygenic Scores (Beta Version)

Genotype Imputation HLA (Minimac4)

Deprecated

Genotype Imputation (Minimac3)

Name

optional job name

Reference Panel

(Details)

-- select an option -- ▾

Input Files (VCF)

File Upload ▾

HLA imputation in MIS

Michigan Imputation Server Home Run Jobs Help Contact saorisakaue

Run

Name optional job name

Reference Panel
(Details)

Input Files (VCF)

✓ -- select an option --
Four-digit Multi-ethnic HLA reference panel (GRCh37/hg19)
Four-digit Multi-ethnic HLA reference panel v2 (GRCh37/hg19)
Multi-ethnic HLA reference panel (GRCh37/hg19)

Select Files

Multiple files can be selected by using the **ctrl** / **cmd** or **shift** keys.

Array Build GRCh37/hg19

Please note that the final SNP coordinates always match the reference build.

Phasing Eagle v2.4 (phased output)

Latest HLA reference panel

HLA imputation in MIS

Input Files (VCF)



File Upload

QCed genotype data on chromosome 6 in VCF format

Select Files

Multiple files can be selected by using the **ctrl** / **cmd** or **shift** keys.

Array Build

GRCh37/hg19

Please note that the final SNP coordinates always match the reference build.

Phasing

Eagle v2.4 (phased output)

Mode

Quality Control & Imputation

☐ AES 256 encryption

Imputation Server encrypts all zip files by default. Please note that AES encryption does not work with standard unzip programs. Use 7z instead.

HLA imputation in MIS

Michigan Imputation Server

HomeRunJobsHelpContact

saorisakaue

Input Files (VCF)

File Upload

Select Files

Multiple files can be selected by using the **ctrl** / **cmd** or **shift** keys.

Array Build

GRCh37/hg19

Please note that the final SNP coordinates always match the reference build.

Phasing

Eagle v2.4 (phased output)

Mode

Quality Control & Imputation

☐ AES 256 encryption

Imputation Server encrypts all zip files by default. Please note that AES encryption does not work with standard unzip programs. Use 7z instead.

Can be prephased (recommended in small sample size) or phasing by Eagle at MIS

Submit Job

ASHG
2018 ANNUAL MEETING 2025
BOSTON, MA • OCTOBER 14-18

Let's look at the output from MIS!

```
ssakaue@wmbed-37d:~/Downloads/chr_6 (2)$ zcat chr6.dose.vcf.gz | less -S
```

Let's look at the output from MIS!

```
##fileformat=VCFv4.1
##filedate=2022.11.14
##contig=<ID=6>
##INFO=<ID=AF,Number=1,Type=Float,Description="Estimated Alternate Allele Frequency">
##INFO=<ID=MAF,Number=1,Type=Float,Description="Estimated Minor Allele Frequency">
##INFO=<ID=R2,Number=1,Type=Float,Description="Estimated Imputation Accuracy (R-square)">
##INFO=<ID=ER2,Number=1,Type=Float,Description="Empirical (Leave-One-Out) R-square (available only for genotyped variants)">
##INFO=<ID=IMPUTED,Number=0,Type=Flag,Description="Marker was imputed but NOT genotyped">
##INFO=<ID=TYPED,Number=0,Type=Flag,Description="Marker was genotyped AND imputed">
##INFO=<ID=TYPED_ONLY,Number=0,Type=Flag,Description="Marker was genotyped but NOT imputed">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DS,Number=1,Type=Float,Description="Estimated Alternate Allele Dosage : [P(0/1)+2*P(1/1)]">
##FORMAT=<ID=HDS,Number=2,Type=Float,Description="Estimated Haploid Alternate Allele Dosage ">
##FORMAT=<ID=GP,Number=3,Type=Float,Description="Estimated Posterior Probabilities for Genotypes 0/0, 0/1 and 1/1 ">
##pipeline=michigan-imputationserver-1.5.8
##imputation=minimac4-1.0.2
##phasing=eagle-2.4
##panel=apps@multiethnics-hla-panel-4digit-v2@1.0.0
##r2Filter=0.0
```

Imputed SNPs within MHC

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	010061321010_R01C01_10007854	010061321010_R02C01_10020793	010061321010_
6	27970031		rs149946	G	T	.	PASS	AF=0.22479;MAF=0.22479;R2=0.99214;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27976200		rs9380032	G	T	.	PASS	AF=0.02975;MAF=0.02975;R2=0.97885;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27979188		rs4141691	A	G	.	PASS	AF=0.11754;MAF=0.11754;R2=0.94642;IMPUTED	GT:DS:HDS:GP	0 0:0.003:0.0	
6	27979625		rs10484402	A	G	.	PASS	AF=0.04041;MAF=0.04041;R2=0.92706;IMPUTED	GT:DS:HDS:GP	0 0:0.003:0.0	
6	27981673		rs9368540	G	A	.	PASS	AF=0.03706;MAF=0.03706;R2=0.98634;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27984726		rs74505854	A	C	.	PASS	AF=0.00719;MAF=0.00719;R2=0.94391;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27984907		rs17765055	T	C	.	PASS	AF=0.04632;MAF=0.04632;R2=0.99897;ER2=0.97293;TYPED	GT:DS:HDS:GP	0 0:0	
6	27986199		rs72848791	C	T	.	PASS	AF=0.04361;MAF=0.04361;R2=0.99732;ER2=0.97465;TYPED	GT:DS:HDS:GP	0 0:0	
6	27986529		rs9368544	A	C	.	PASS	AF=0.04631;MAF=0.04631;R2=0.99885;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27998258		rs149990	G	A	.	PASS	AF=0.11782;MAF=0.11782;R2=0.99974;ER2=0.99765;TYPED	GT:DS:HDS:GP	0 0:0	
6	27999044		rs9368545	A	T	.	PASS	AF=0.04627;MAF=0.04627;R2=0.99825;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	27999421		rs16893573	C	T	.	PASS	AF=0.02852;MAF=0.02852;R2=0.98601;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	28001003		rs17708949	A	C	.	PASS	AF=0.03124;MAF=0.03124;R2=0.98465;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0	
6	28001610		rs149942	T	C	.	PASS	AF=0.26963;MAF=0.26963;R2=0.99868;ER2=0.98731;TYPED	GT:DS:HDS:GP	0 0:0	
6	28002388		rs149943	G	A	.	PASS	AF=0.11781;MAF=0.11781;R2=0.99984;ER2=0.99889;TYPED	GT:DS:HDS:GP	0 0:0	
6	28003271		rs183926	T	A	.	PASS	AF=0.00996;MAF=0.00996;R2=0.99551;ER2=0.95540;TYPED	GT:DS:HDS:GP	0 0:0	

Let's look at the output from MIS!

```
ssakaue@wmbed-37d:~/Downloads/chr_6 (2)$ zcat chr6.dose.vcf.gz | less -S  
ssakaue@wmbed-37d:~/Downloads/chr_6 (2)$ zcat chr6.dose.vcf.gz | grep HLA | less -S
```

Imputed HLA alleles

6	29910247	HLA_A*01	A	T	.	PASS	AF=0.15456;MAF=0.15456;R2=0.99719;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910248	HLA_A*01:01	A	T	.	PASS	AF=0.15234;MAF=0.15234;R2=0.99517;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910249	HLA_A*01:02	A	T	.	PASS	AF=0.00110;MAF=0.00110;R2=0.85198;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910250	HLA_A*01:136	A	T	.	PASS	AF=0.00002;MAF=0.00002;R2=0.04644;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910251	HLA_A*02	A	T	.	PASS	AF=0.26026;MAF=0.26026;R2=0.99741;IMPUTED	GT:DS:HDS:GP	0 1:0.998:0.0
6	29910252	HLA_A*02:01	A	T	.	PASS	AF=0.22914;MAF=0.22914;R2=0.98881;IMPUTED	GT:DS:HDS:GP	0 1:0.996:0.0
6	29910253	HLA_A*02:02	A	T	.	PASS	AF=0.00471;MAF=0.00471;R2=0.99686;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910254	HLA_A*02:03	A	T	.	PASS	AF=0.00094;MAF=0.00094;R2=0.88473;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910255	HLA_A*02:04	A	T	.	PASS	AF=0.00017;MAF=0.00017;R2=0.89691;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910256	HLA_A*02:05	A	T	.	PASS	AF=0.01556;MAF=0.01556;R2=0.99839;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910257	HLA_A*02:06	A	T	.	PASS	AF=0.00364;MAF=0.00364;R2=0.98705;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910258	HLA_A*02:07	A	T	.	PASS	AF=0.00133;MAF=0.00133;R2=0.94265;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910259	HLA_A*02:10	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.06774;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910260	HLA_A*02:11	A	T	.	PASS	AF=0.00072;MAF=0.00072;R2=0.79302;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910261	HLA_A*02:135	A	T	.	PASS	AF=0.00005;MAF=0.00005;R2=0.22275;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910262	HLA_A*02:17	A	T	.	PASS	AF=0.00108;MAF=0.00108;R2=0.91078;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910263	HLA_A*02:195	A	T	.	PASS	AF=0.00003;MAF=0.00003;R2=0.00997;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910264	HLA_A*02:20	A	T	.	PASS	AF=0.00019;MAF=0.00019;R2=0.41177;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910265	HLA_A*02:22	A	T	.	PASS	AF=0.00026;MAF=0.00026;R2=0.84742;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910266	HLA_A*02:279	A	T	.	PASS	AF=0.00008;MAF=0.00008;R2=0.24552;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910267	HLA_A*02:55	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.02045;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910268	HLA_A*02:56	A	T	.	PASS	AF=0.00007;MAF=0.00007;R2=0.08360;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910269	HLA_A*02:60	A	T	.	PASS	AF=0.00004;MAF=0.00004;R2=0.49123;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910270	HLA_A*02:76	A	T	.	PASS	AF=0.00030;MAF=0.00030;R2=0.40308;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910271	HLA_A*02:87	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.00984;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910272	HLA_A*03	A	T	.	PASS	AF=0.12657;MAF=0.12657;R2=0.99727;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910273	HLA_A*03:01	A	T	.	PASS	AF=0.12061;MAF=0.12061;R2=0.99073;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910274	HLA_A*03:02	A	T	.	PASS	AF=0.00441;MAF=0.00441;R2=0.97833;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910275	HLA_A*03:36N	A	T	.	PASS	AF=0.00063;MAF=0.00063;R2=0.40664;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910276	HLA_A*03:89	A	T	.	PASS	AF=0.00006;MAF=0.00006;R2=0.05589;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910277	HLA_A*11	A	T	.	PASS	AF=0.06332;MAF=0.06332;R2=0.99759;IMPUTED	GT:DS:HDS:GP	1 0:1.000:1.0
6	29910278	HLA_A*11:01	A	T	.	PASS	AF=0.05966;MAF=0.05966;R2=0.96821;IMPUTED	GT:DS:HDS:GP	1 0:0.958:0.9
6	29910279	HLA_A*11:02	A	T	.	PASS	AF=0.00059;MAF=0.00059;R2=0.84655;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910280	HLA_A*11:32	A	T	.	PASS	AF=0.00021;MAF=0.00021;R2=0.04209;IMPUTED	GT:DS:HDS:GP	0 0:0.024:0.0
6	29910281	HLA_A*11:50Q	A	T	.	PASS	AF=0.00250;MAF=0.00250;R2=0.41268;IMPUTED	GT:DS:HDS:GP	0 0:0.018:0.0
6	29910282	HLA_A*23	A	T	.	PASS	AF=0.02822;MAF=0.02822;R2=0.99569;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910283	HLA_A*23:01	A	T	.	PASS	AF=0.02797;MAF=0.02797;R2=0.99150;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910284	HLA_A*23:15	A	T	.	PASS	AF=0.00000;MAF=0.00000;R2=0.00005;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0

REF ALT

"T": Presence of the allele
"A": Absence of the allele

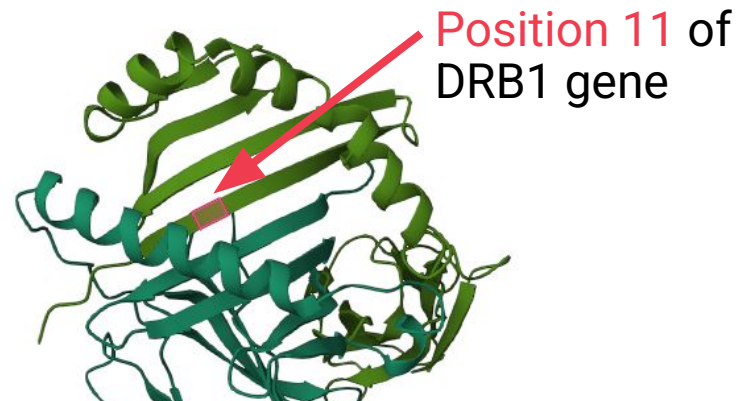
Imputed HLA alleles

6	29910247	HLA_A*01	A	T	.	PASS	AF=0.15456;MAF=0.15456;R2=0.99719;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910248	HLA_A*01:01	A	T	.	PASS	AF=0.15234;MAF=0.15234;R2=0.99517;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910249	HLA_A*01:02	A	T	.	PASS	AF=0.00110;MAF=0.00110;R2=0.85198;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910250	HLA_A*01:136	A	T	.	PASS	AF=0.00002;MAF=0.00002;R2=0.04644;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910251	HLA_A*02	A	T	.	PASS	AF=0.26026;MAF=0.26026;R2=0.99741;IMPUTED	GT:DS:HDS:GP	0 1:0.998:0.0
6	29910252	HLA_A*02:01	A	T	.	PASS	AF=0.22914;MAF=0.22914;R2=0.98881;IMPUTED	GT:DS:HDS:GP	0 1:0.996:0.0
6	29910253	HLA_A*02:02	A	T	.	PASS	AF=0.00471;MAF=0.00471;R2=0.99686;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910254	HLA_A*02:03	A	T	.	PASS	AF=0.00094;MAF=0.00094;R2=0.88473;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910255	HLA_A*02:04	A	T	.	PASS	AF=0.00017;MAF=0.00017;R2=0.89691;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910256	HLA_A*02:05	A	T	.	PASS	AF=0.01556;MAF=0.01556;R2=0.99839;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910257	HLA_A*02:06	A	T	.	PASS	AF=0.00364;MAF=0.00364;R2=0.98705;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910258	HLA_A*02:07	A	T	.	PASS	AF=0.00133;MAF=0.00133;R2=0.94265;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910259	HLA_A*02:10	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.06774;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910260	HLA_A*02:11	A	T	.	PASS	AF=0.00072;MAF=0.00072;R2=0.79302;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910261	HLA_A*02:135	A	T	.	PASS	AF=0.00005;MAF=0.00005;R2=0.22275;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910262	HLA_A*02:17	A	T	.	PASS	AF=0.00108;MAF=0.00108;R2=0.91078;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910263	HLA_A*02:195	A	T	.	PASS	AF=0.00003;MAF=0.00003;R2=0.00997;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910264	HLA_A*02:20	A	T	.	PASS	AF=0.00019;MAF=0.00019;R2=0.41177;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910265	HLA_A*02:22	A	T	.	PASS	AF=0.00026;MAF=0.00026;R2=0.84742;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910266	HLA_A*02:279	A	T	.	PASS	AF=0.00008;MAF=0.00008;R2=0.24552;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910267	HLA_A*02:55	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.02045;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910268	HLA_A*02:56	A	T	.	PASS	AF=0.00007;MAF=0.00007;R2=0.08360;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910269	HLA_A*02:60	A	T	.	PASS	AF=0.00004;MAF=0.00004;R2=0.49123;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910270	HLA_A*02:76	A	T	.	PASS	AF=0.00030;MAF=0.00030;R2=0.40308;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910271	HLA_A*02:87	A	T	.	PASS	AF=0.00001;MAF=0.00001;R2=0.00984;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910272	HLA_A*03	A	T	.	PASS	AF=0.12657;MAF=0.12657;R2=0.99727;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910273	HLA_A*03:01	A	T	.	PASS	AF=0.12061;MAF=0.12061;R2=0.99073;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910274	HLA_A*03:02	A	T	.	PASS	AF=0.00441;MAF=0.00441;R2=0.97833;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910275	HLA_A*03:36N	A	T	.	PASS	AF=0.00063;MAF=0.00063;R2=0.40664;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910276	HLA_A*03:89	A	T	.	PASS	AF=0.00006;MAF=0.00006;R2=0.05589;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910277	HLA_A*11	A	T	.	PASS	AF=0.06332;MAF=0.06332;R2=0.99759;IMPUTED	GT:DS:HDS:GP	1 0:1.000:1.0
6	29910278	HLA_A*11:01	A	T	.	PASS	AF=0.05966;MAF=0.05966;R2=0.96821;IMPUTED	GT:DS:HDS:GP	1 0:0.958:0.9
6	29910279	HLA_A*11:02	A	T	.	PASS	AF=0.00059;MAF=0.00059;R2=0.84655;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910280	HLA_A*11:32	A	T	.	PASS	AF=0.00021;MAF=0.00021;R2=0.04209;IMPUTED	GT:DS:HDS:GP	0 0:0.024:0.0
6	29910281	HLA_A*11:50Q	A	T	.	PASS	AF=0.00250;MAF=0.00250;R2=0.41268;IMPUTED	GT:DS:HDS:GP	0 0:0.018:0.0
6	29910282	HLA_A*23	A	T	.	PASS	AF=0.02822;MAF=0.02822;R2=0.99569;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910283	HLA_A*23:01	A	T	.	PASS	AF=0.02797;MAF=0.02797;R2=0.99150;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0
6	29910284	HLA_A*23:15	A	T	.	PASS	AF=0.00000;MAF=0.00000;R2=0.00005;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0

REF ALT

"T": Presence of the allele
"A": Absence of the allele

How are imputed HLA variants associated with disease?



Disease ↔ ?

1 HLA alleles

HLA-DRB1*01:01 → GGSCMAALTVTLMVLSSP
HLA-DRB1*04:01 → GGSCMTALTVTLMVLSSP
HLA-DRB1*15:01 → GGSCMTALTVTLMVLSSP

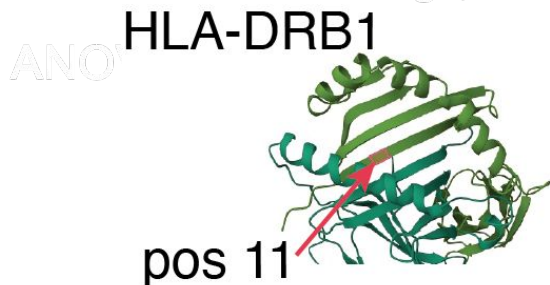
2 HLA amino acid positions (Omnibus test)

3 HLA haplotypes (Conditional haplotype test)

Omnibus test to ask **which HLA amino acid position** affects the disease

Full model: $\log(odds_i) = \beta_0 + \sum_k \beta_k \text{covariate}_{k,i} + \sum_{m=1}^{M-1} \beta_m \text{AM}_{m,i}$

Reduced model: $\log(odds_i) = \beta_0 + \sum_k \beta_k \text{covariate}_{k,i}$



$M = 6$ possible amino acid residues

	$M - 1$					
	L	S	V	G	D	(P)
⋈ RFLWQ L KFECH	0.1	0.0	0.9	0.0	0.0	1.0
⋈ RFLEY S TSECH	0.9	0.0	0.0	0.0	1.0	0.1
⋈ RFLEQ V KHECH	0.0	1.0	0.0	0.9	0.0	0.1
⋈ RFLWQ G KYKCH	0.0	0.0	1.9	0.0	0.1	0.0
⋈ RFLKQ D KFECH	2.0	0.0	0.0	0.0	0.0	0.0
⋈ RFLWQ P KRECH						

Omnibus test to ask **which HLA amino acid position** affects the disease

Full model: $\log(odds_i) = \beta_0 + \sum_k \beta_k \text{covariate}_{k,i} + \sum_{m=1}^{M-1} \beta_m \text{AM}_{m,i}$

Reduced model: $\log(odds_i) = \beta_0 + \sum_k \beta_k \text{covariate}_{k,i}$

ANOVA(Full model, Reduced model)

How much does **this amino acid position**'s polymorphism increase the explained variance for the trait?

→ Determine the single most significant amino acid position

Summary

1. HLA amino acid sequences and alleles characterize antigen presentation and disease risk within HLA.
2. HLA amino acid sequences and alleles can be accurately imputed from genotyped SNPs by MIS.
3. Imputed HLA alleles can be used to fine-map causal disease mechanisms.

Summary

1. HLA amino acid sequences and alleles characterize antigen presentation and disease risk within HLA.
2. HLA amino acid sequences and alleles can be accurately imputed from genotyped SNPs by MIS.
3. Imputed HLA alleles can be used to fine-map causal disease mechanisms.

nature protocols

Review article

<https://doi.org/10.1038/s41596-023-00853-4>

Tutorial: a statistical genetics guide to identifying HLA alleles driving complex disease

Sakaue et al. *Nature Protocols* 2023

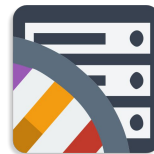


Section 4

nf-gwas and PGS Server



Lukas Forer
Medical University of Innsbruck
lukas.forer@i-med.ac.at
@lukfor



MICHIGAN
IMPUTATIONSERVER

Learning objectives

Participants will

1. Learn how to run a GWAS pipeline
2. Learn how to use the PGS extension
3. Learn how to analyze PGS calculated by the server

Summary of common GWAS analysis tools

	EPACTS	Rvtests	SNPTEST	SAIGE	BLOT-LMM	Bgenie	regenie
Input VCF	Y	Y	Y				
Sample relatedness (Quantitative outcome)	Y	Y		Y	Y	Y	Y
Sample relatedness (Binary outcome)				Y		Y	Y
Case control imbalance				Y			Y
Large sample size (>20,000)				Y	Y	Y	Y

nf-gwas

- A GWAS pipeline based on REGENIE and works with VCF input
- Includes several pre- and post-processing steps
- Based on Nextflow
 - Allows to build a portable, reproducible, scalable pipeline
 - Runs on clusters (e.g. SLURM) or in the cloud (e.g. AWS Batch)

JOURNAL ARTICLE

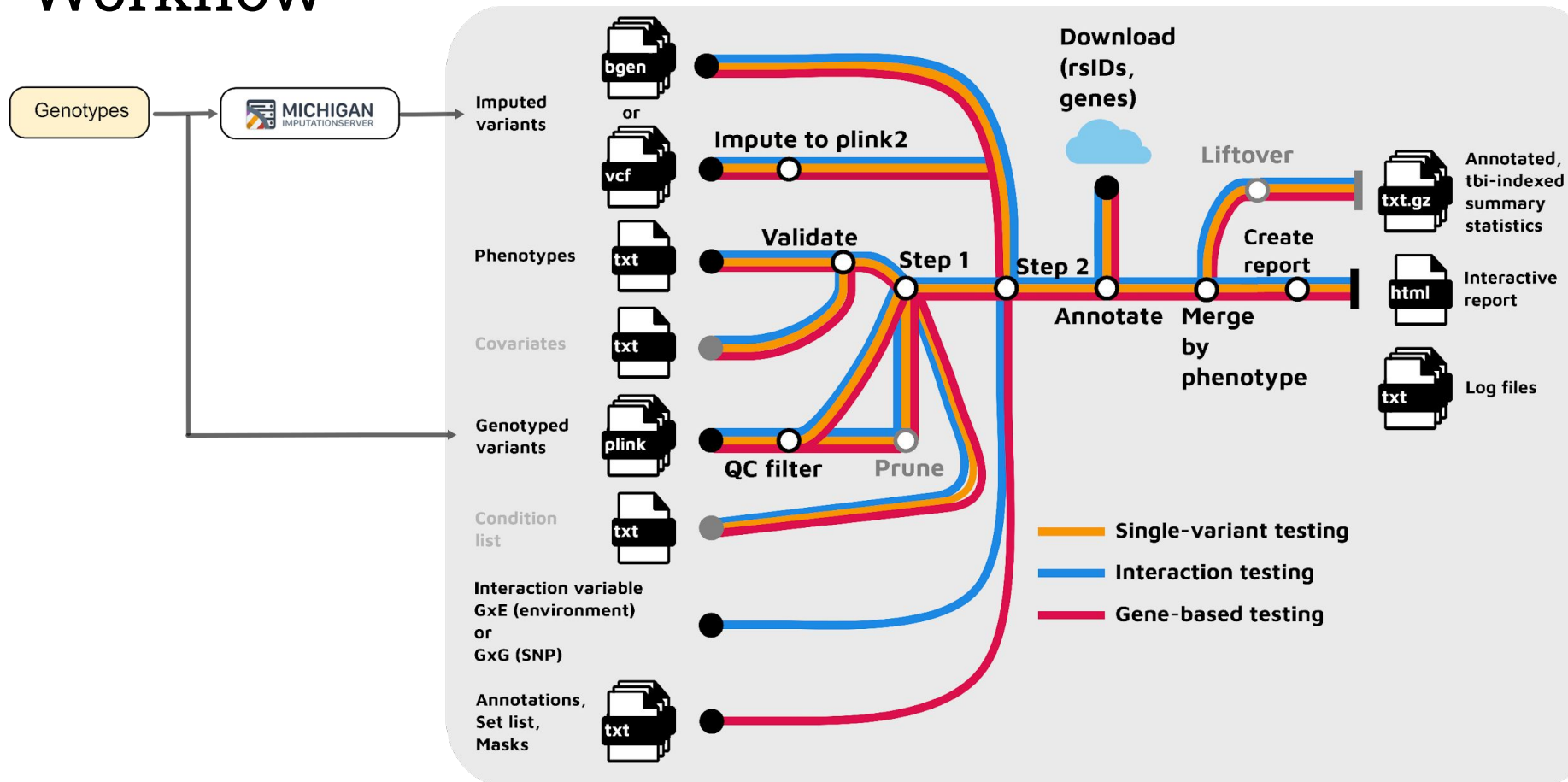
Performing highly parallelized and reproducible GWAS analysis on biobank-scale data

Sebastian Schönherr, Johanna F Schachtl-Riess, Silvia Di Maio, Michele Filosi,
Marvin Mark, Claudia Lamina, Christian Fuchsberger, Florian Kronenberg, Lukas Forer ✉

[Author Notes](#)

NAR Genomics and Bioinformatics, Volume 6, Issue 1, March 2024, lqae015,
<https://doi.org/10.1093/nargab/lqae015>

Workflow



Download Phenotypes

Phenotype Data

- **Filename:** [phenotypes.txt](#) (131.6 KB)
- **Description:** Simulated phenotype file containing four traits.
- **Format:** Plain text (tab-delimited).
- **Columns:**
 - [sample_id](#) — individual sample identifiers
 - [pheno_1](#), [pheno_2](#), [pheno_3](#), [pheno_4](#) — four simulated phenotypes
- **Use:** Provides phenotype data for GWAS and PGS evaluation.

Covariates (Principal Components)

- **Filename:** [covariates.txt](#) (292.3 KB)
- **Description:** Covariate file containing the first 10 genetic principal components (PCs)
- **Format:** Tab-delimited, compressed with gzip.
- **Columns:**
 - [sample](#) — individual identifiers matching the genotype data
 - [PC1](#) – [PC10](#) — first ten principal components representing population structure
- **Use:** Used as covariates in association analyses to correct for population stratification

Simulated Phenotypes

Phenotype	Type	Highly Genetic?
pheno_1	Binary (e.g., case/control)	✓ Yes
pheno_2	Continuous (e.g., height)	✓ Yes
pheno_3	Binary or categorical	✗ No
pheno_4	Continuous	✗ No



<https://tinyurl.com/imputationserver-2025>

You can also download the pre-imputed data

Easy to configure

my-gwas.config

```
params {  
  project = 'test-gwas-with-pcs'  
  genotypes_association = 'gwas.imputed.chr20.dose.vcf.gz'  
  genotypes_association_format = 'vcf'  
  association_build = 'hg19'  
  phenotypes_filename = 'phenotypes.txt'  
  phenotypes_columns = 'pheno_2,pheno_4'  
  covariates_filename = 'covariates.txt'  
  covariates_columns = 'PC1,PC2,PC3,PC4,PC5,PC6,PC7,PC8,PC9,PC10'  
  regenie_test = 'additive'  
  regenie_min_imputation_score = 0.3  
  regenie_skip_predictions = true  
  rsids_filename = "rsids-v154-hg19-chr20.index.gz"  
  binning_size = 50000  
}
```

} Imputed genotypes
from Imputation Server

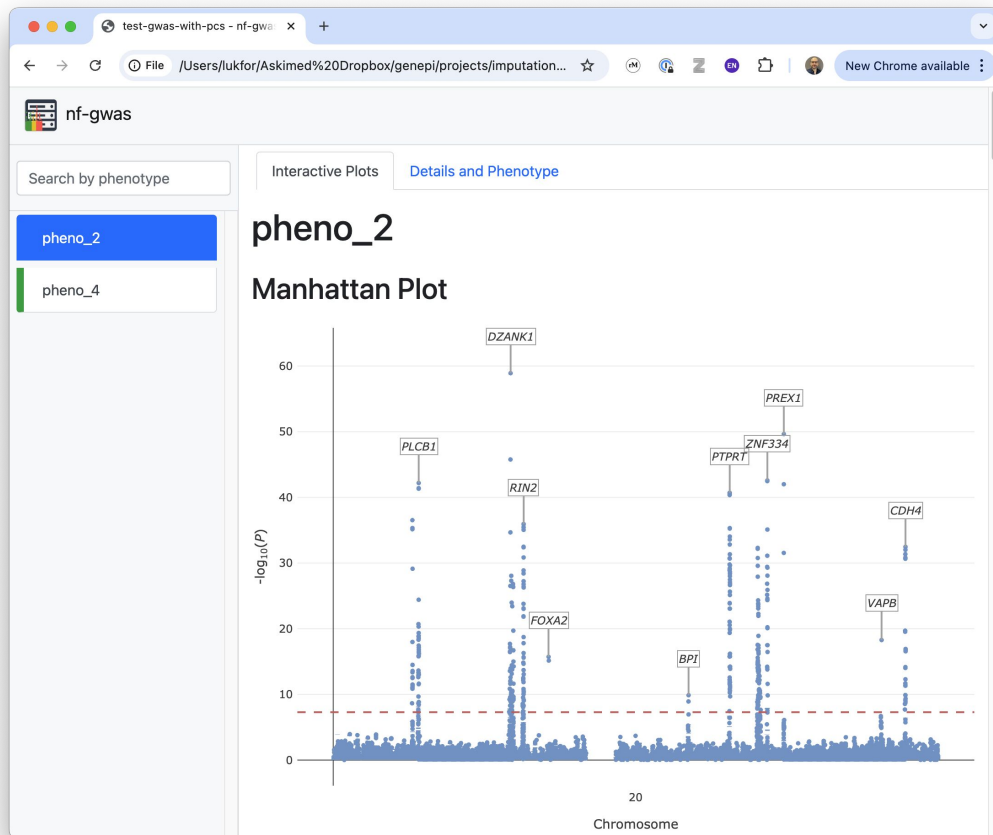
} phenotypes

} covariates

} Annotation and
visualization

```
nextflow run genepi/nf-gwas -r v1.0.11 -profile docker -c my-gwas.config
```

Results Phenotype 2

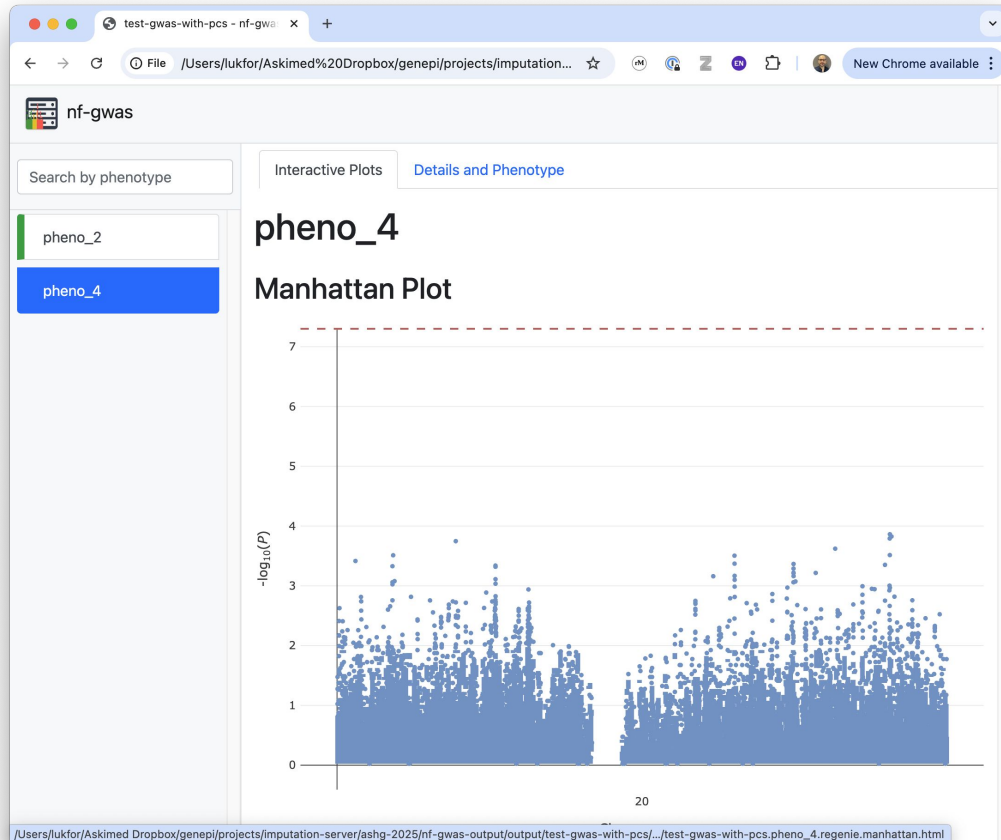


Top Loci

Variant	ID	Ref Allele	Effect Allele	Nearest Gene	$-\log_{10}(P)$	Beta
20:18445266	rs6111973	C	T	DZANK1	58.9235	-0.522676
20:46858654	rs7269193	A	G	PREX1	49.648	0.626932
20:45142619	rs847055	A	T	ZNF334	42.5918	0.493058
20:8887891	rs6056246	T	C	PLCB1	42.2005	0.437334
20:41227725	rs6102924	C	T	PTPRT	40.7084	-0.627067
20:19800969	rs1884767	A	G	RIN2	35.9524	-0.385179
20:59507557	rs6071441	C	T	CDH4	32.4673	0.480876
20:57020044	rs6070466	C	G	VAPB	18.2868	0.925181
20:22392054	rs13038903	A	G	FOXA2	15.7283	-3.3011
20:36950990	rs6024863	A	T	BPI	9.8814	0.225461
20:1728212	rs605739	T	G	SIRPG	3.95055	0.124383
20:4453463	rs297660	C	T	PRNP	3.89483	-0.14028
20:2513877	rs6083567	A	G	TMC2	3.79769	-0.217792
20:21387531	rs13040764	A	G	NKX2-4	3.7689	-0.763363
20:31440180	rs6088022	T	G	MAPRE1	3.69196	-0.442159
20:25957535	rs2252934	G	C	ZNF337	3.57123	-0.184881
20:29859763	rs6089068	C	T	DEFB115	3.51671	0.189548
20:53303956	rs6023507	C	T	DOK5	3.47376	0.357591
20:25221547	rs6050477	T	C	PYGB	3.47251	-0.120575

Loci are defined ± 200 kb around lead SNPs

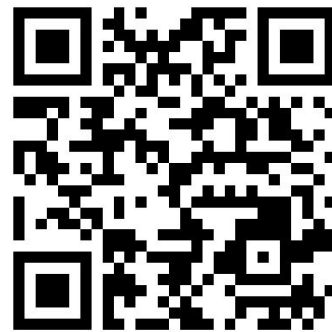
Results Phenotype 4



No significant loci.

Tutorial: Setup, Parameters, and Example Analyses

- On the workshop website, you'll find a step-by-step tutorial on how to:
 - Setup and install the pipeline
 - Explore all available parameters
 - Run association tests for binary traits (e.g. pheno_1 and pheno_2)
 - Generate and interpret QQ plots
- Also available for PLINK2



<https://tinyurl.com/imputationserver-2025>

JOURNAL ARTICLE

Imputation Server PGS: an automated approach to calculate polygenic risk scores on imputation servers

Lukas Forer, Daniel Taliun, Jonathon LeFaive, Albert V Smith, Andrew P Boughton, Stefan Coassin, Claudia Lamina, Florian Kronenberg, Christian Fuchsberger, Sebastian Schönherr 

Nucleic Acids Research, Volume 52, Issue W1, 5 July 2024, Pages W70–W77,
<https://doi.org/10.1093/nar/gkae331>

Imputation Server PGS

- Makes it easier to use polygenic scores (PGS) with imputed genotypes by calculating them directly on the server
- PGS: aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for a phenotype
- Available since April 2022
- More than 5,000 submitted jobs
- Supports more than 3,000 scores

Have you previously used the PGS tool on the Michigan Imputation Server?

Yes



2%

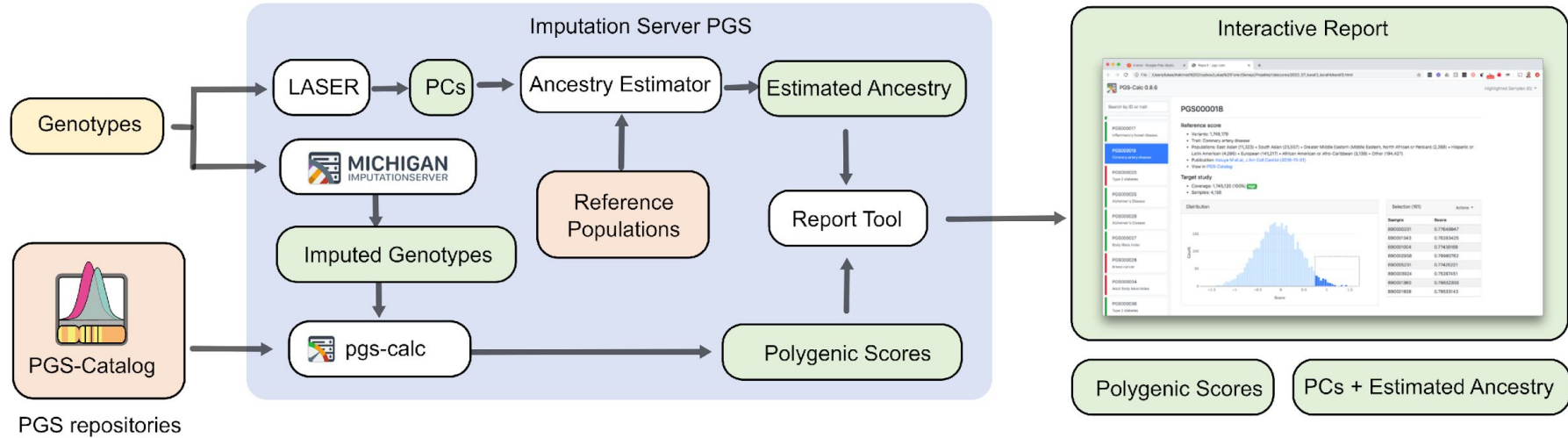
No



98%



Workflow



 Select Files

Multiple files can be selected

Genotype Imputation 2.0.9

HLA Imputation 2.0.9

Polygenic Score Calculation 2.0.9

Array Build

GRCh37/hg19 ▾

Please note that the final SNP
coordinates always match the
reference build.

rsq Filter

0.3 ▾

PGS Calculation

Please select a collection of polygenic scores to enable on the fly PGS calculation.

Scores

✓ -- select an option --

PGS-Catalog v20240318

Trait Category

Tutorial (Simulated Scores!)

Ancestry Estimation

Disabled ▾

Results

[Details](#)[Results](#)[Logs](#)

Downloads

> statistics

- qc_report.txt (674 bytes)
- quality-control.html (1 MB)
- scores.coverage (222 bytes)
- scores.html (32 KB)
- scores.info (1 KB)
- scores.report.zip (175 KB)
- scores.txt (228 KB)**

Logs

- step1-nextflow.log (59 KB)
- step1-report.html (2 MB)
- step1-timeline.html (251 KB)
- step1-trace.csv (3 KB)



PGS-Server 1.6.1

Samples

score_1

score_2

score_3

score_4

score_1

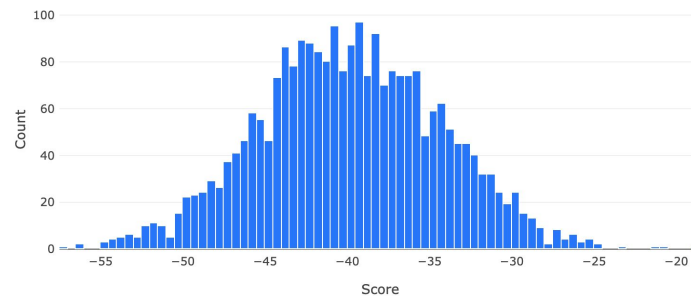
Reference score

- Variants: 32
- Variants ignored: 0

Target study

- Coverage: 32 (100%) **high**
- Samples: 0 / 2.504 (0%)

Distribution



What can we do with these scores now?

Example: compare how well each score explains the simulated phenotypes

- Merge scores with phenotypes

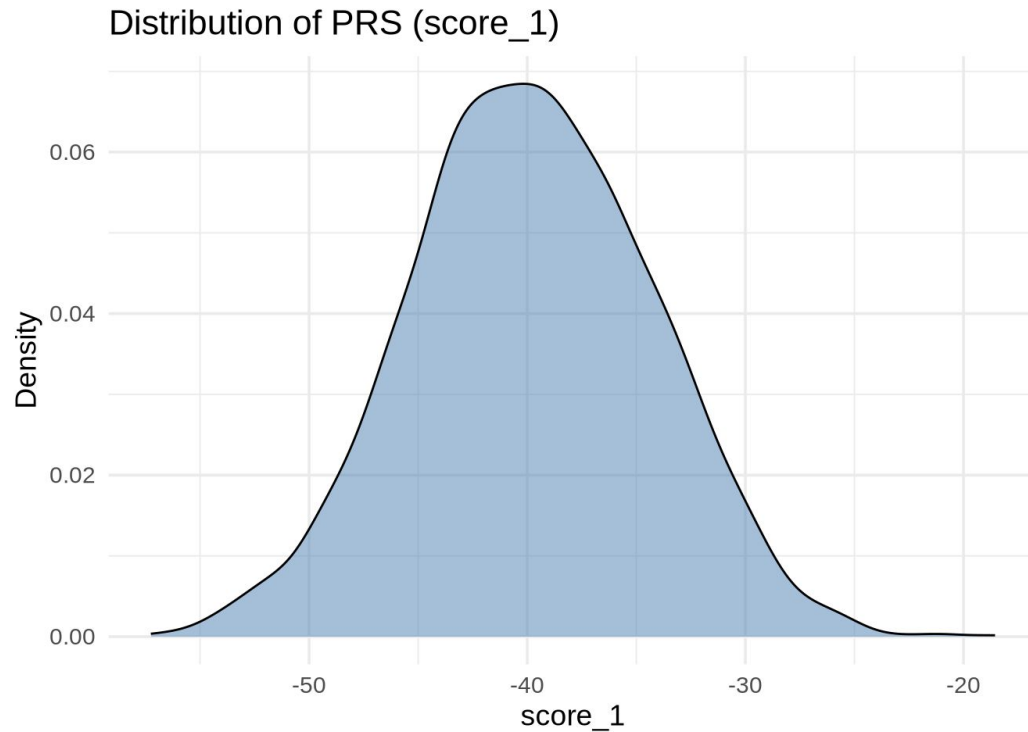
```
# Read PRS scores and phenotypes
scores <- read_csv("data/scores.txt")
phenos <- read_table("data/phenotypes.txt")

# Merge datasets by sample ID
merged <- inner_join(scores, phenos, by = c("sample" = "IID"))
```

- You can also download the calculated polygenic scores from the website

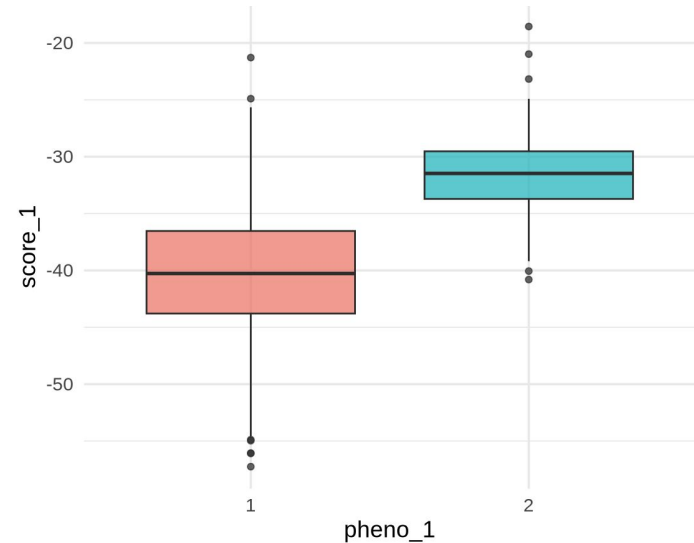
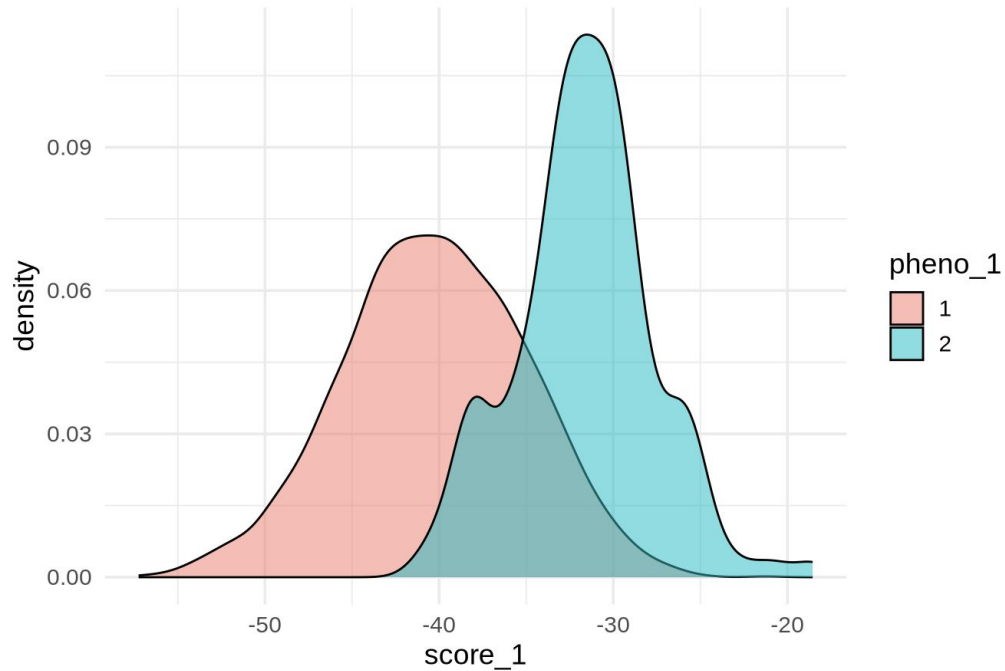
<https://tinyurl.com/imputationserver-2025>

```
ggplot(merged, aes(x = score_1)) +  
  geom_density(alpha = 0.5, fill = "steelblue") +  
  labs(  
    title = "Distribution of PRS (score_1)",  
    x = "score_1",  
    y = "Density"  
  )
```



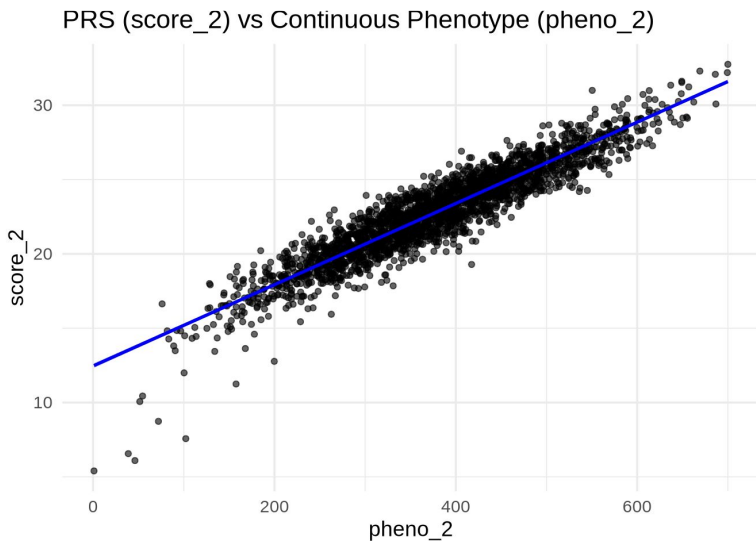
```
ggplot(merged, aes(x = score_1, fill = as.factor(pheno_1))) +
  geom_density(alpha = 0.5) +
  labs(
    title = "Distribution of PRS (score_1) by Binary Phenotype (pheno_1)",
    x = "score_1",
    fill = "pheno_1"
  )
)
```

Distribution of PRS (score_1) by Binary Phenotype (pheno_1)



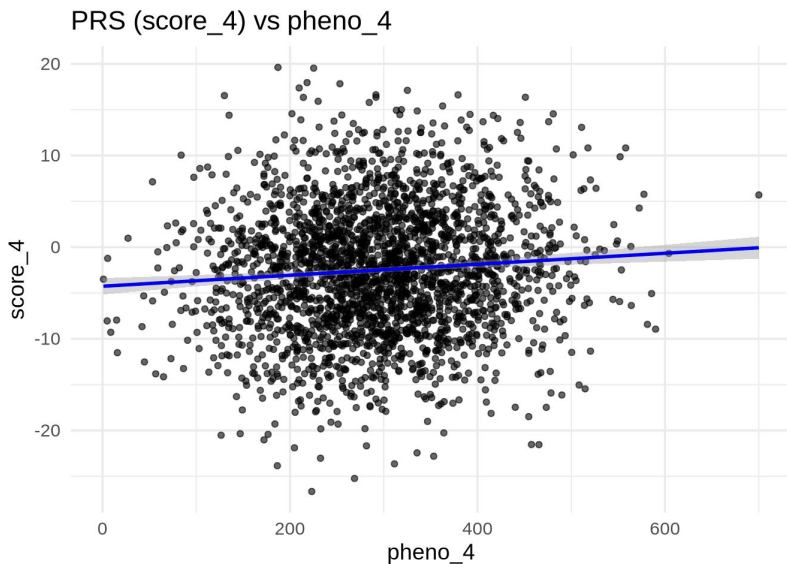
```
ggplot(merged, aes(x = pheno_2, y = score_2)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "PRS (score_2) vs Continuous Phenotype (pheno_2)",
    x = "pheno_2",
    y = "score_2"
  )
```

geom_smooth() under formula method



```
ggplot(merged, aes(x = pheno_4, y = score_4)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "PRS (score_4) vs pheno_4",
    x = "pheno_4",
    y = "score_4"
  )
```

geom_smooth() under formula method



<https://tinyurl.com/imputationserver-2025>

Summary

- Imputation Servers are easy to use and ensure high-quality imputation
- GWAS and PGS play a key role in modern genetic research
- Several pipelines and extensions of the Imputation Server

Aim: Taking the burden out of genetic analysis

Section 5

Helmholtz Munich Imputation Server



Eleftheria Zeggini

Director, Institute of Translational
Genomics

HELMHOLTZ
MUNICH

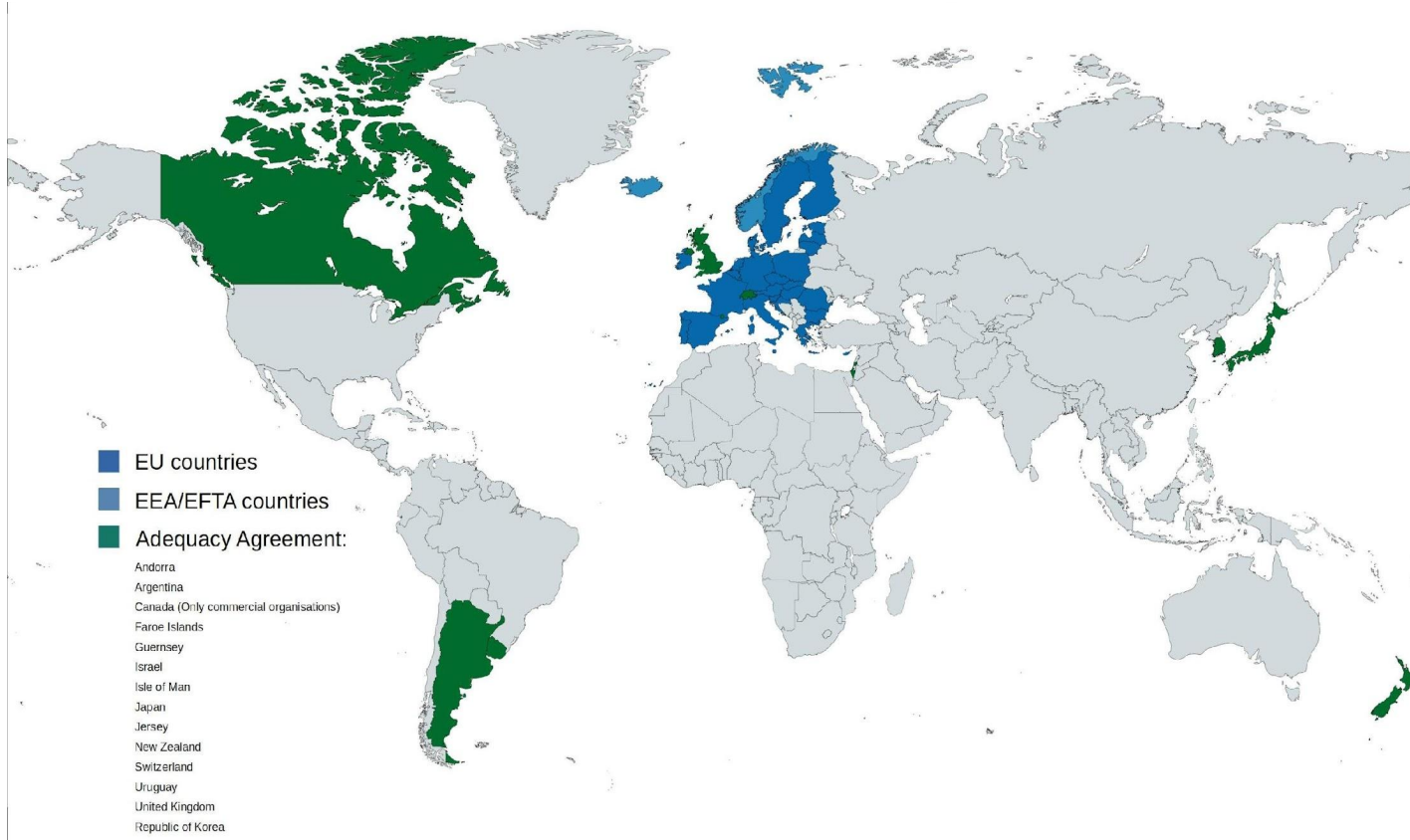
Why do we need another imputation server?



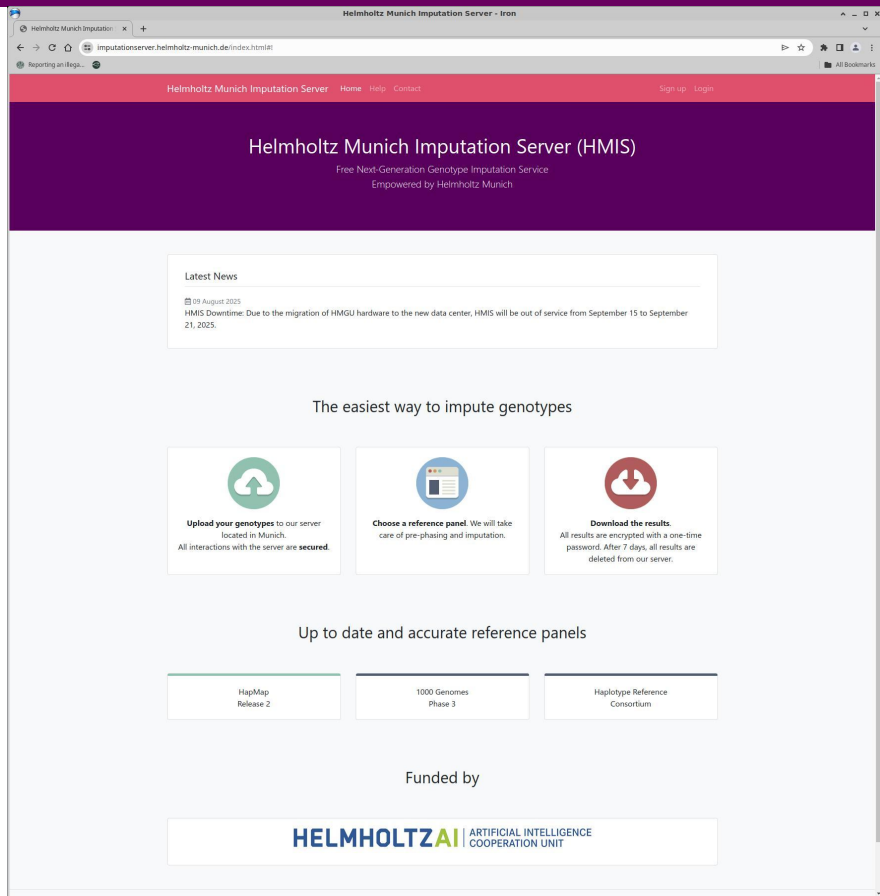
- The transfer of personal data to countries outside the European Economic Area (EEA) is generally prohibited by default
- Therefore, all personal data on EU citizens must reside and be processed on servers physically located within the EU
- Human genetic data are considered a special category of personal data subject to stricter requirements
- There are reasons when personal data can be moved outside the EEA:
 - Consent is given by the individual for the transfer
 - A Standard Contractual Clause (SCC) has been agreed between the two parties
 - There is adequacy agreement with country



EU Adequacy Agreements



Helmholtz Munich Imputation Server



The server is designed to assist users in the EU to comply with the 2016 General Data Protection Regulation (GDPR) law

nature genetics

Correspondence | Published: 15 November 2024

Toward GDPR compliance with the Helmholtz Munich genotype imputation server

[N. William Rayner](#) , [Young-Chan Park](#), [Christian Fuchsberger](#), [Andrei Barysenka](#) & [Eleftheria Zeggini](#) 

[Nature Genetics](#) 56, 2580–2581 (2024) | [Cite this article](#)



HELMHOLTZ MUNICH 



- Based on the freely available Michigan software
- Currently version 2 based on Cloudfuge and Hadoop
- Moving to version 3, currently integrating the changes we have made from the standard Michigan software

Notable Differences vs Michigan



- More information is required on sign up to comply with GDPR

HMIS sign up



Institute name and address



email of legal representative
and country are required so
as to meet our obligations
under GDPR

The screenshot shows the registration page of the Helmholtz Munich Imputation Server. The page is titled "Sign up" and is divided into two main sections: "Personal details" and "Affiliation details".

Personal details:

- Username:
- Full Name:
- E-Mail:
- Password:
- Confirm password:

Affiliation details:

- Supervisor or Legal-representative E-Mail:
- Institute name:
- Address:
- City:
- State:
- Postal Code:
- Country:

After alignment with the responsible administrative personnel of my institution I confirm that I am fully authorized to accept these [Terms of Service](#) including the Data Processing Agreement on behalf of my institution.

I herewith accept these Terms of Service including the Data Processing Agreement on behalf of my institution: ☐

[Register](#)

powered by Cloudgene and supported by the Institute of Translational Genomics (ITG) | [Privacy Policy](#) | [Legal notice](#)

Notable Differences vs Michigan



- More information is required on sign up to comply with GDPR
- Short and long (>25,000 samples) processing queues introduced to ensure fair access to the resource
- Longer more complex passwords are now required, >15 characters

Current Reference Panels

- 1000G GRCh37
- 1000G GRCh38
- HRC
- HapMap2



Upcoming Reference Panels

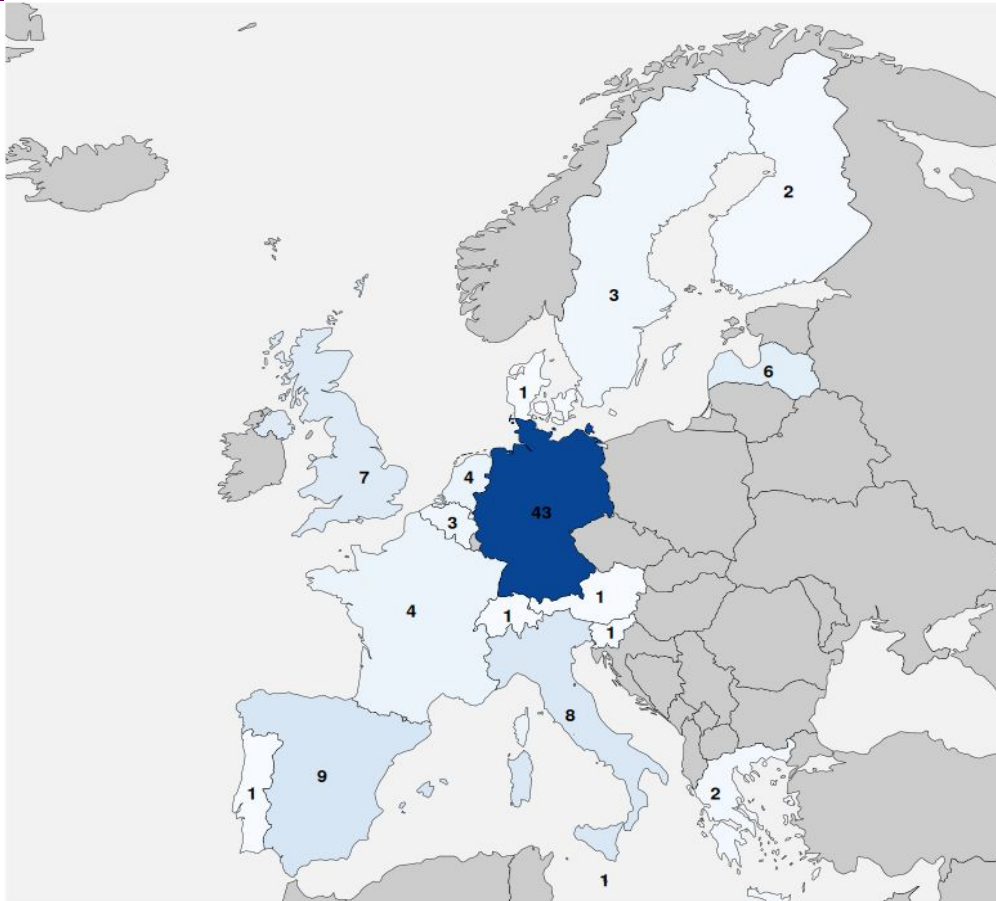
- German National Cohort (NAKO)
~15,500
Will be expanded to 30,000
- Genome of Europe



Registered Users



We have 101 registered users, largely from across the EU



Number of genomes

- Over 900,000 genomes imputed to date



0 surveys completed



0 surveys underway



Section 6

The TOPMed Imputation Server



Albert Smith
University of Michigan

albertvs@umich.edu

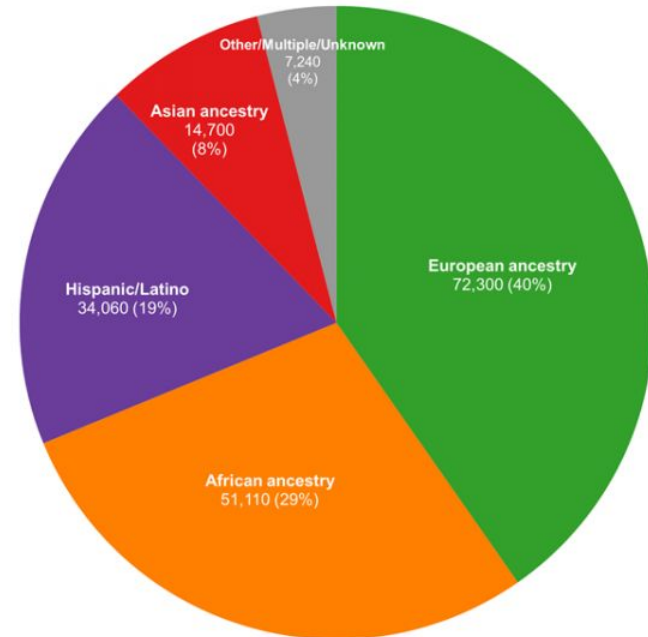
 @avsmith

TOPMed Program

Ancestry & Ethnicity

Phases 1-7 (~180K Participants)

- Trans-Omics for Precision Medicine (TOPMed) Program
- A Precision Medicine Initiative sponsored by National Heart, Lung and Blood Institute
- Integrating whole-genome sequencing and other omics data
- >180k participants from >90 studies



TOPMed Imputation

- Current reference panel based on TOPMed Freeze 8 Calls
- Michigan Imputation Server ported to Amazon Web Services
- R2 panel released April 2020
- Updated R3 panel released December 2023
- <https://imputation.biodatacatalyst.nhlbi.nih.gov>
- Registration as before, open access to TOPMed panel

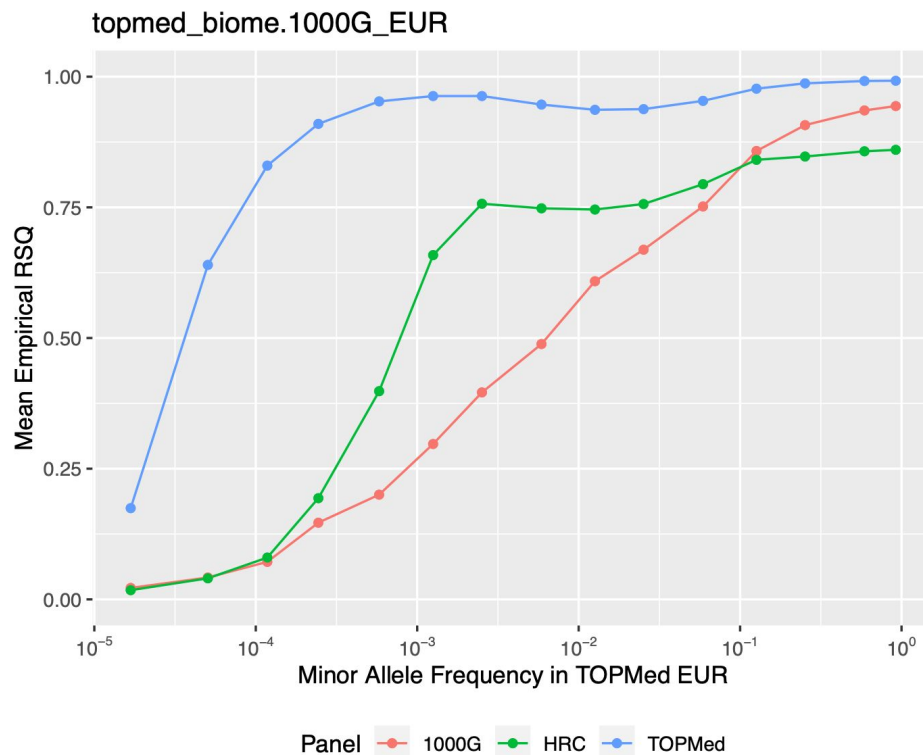
TOPMed Panel Compared

	TOPMed_r2	HRC	1000G Genomes
N samples	97K	39K	2,500
Ancestry	Multiethnic	European	Multiethnic
N variants	308M	39M	88M
Avg. depth	38X	8X	4X
Genome build Position	b38	b37	b37

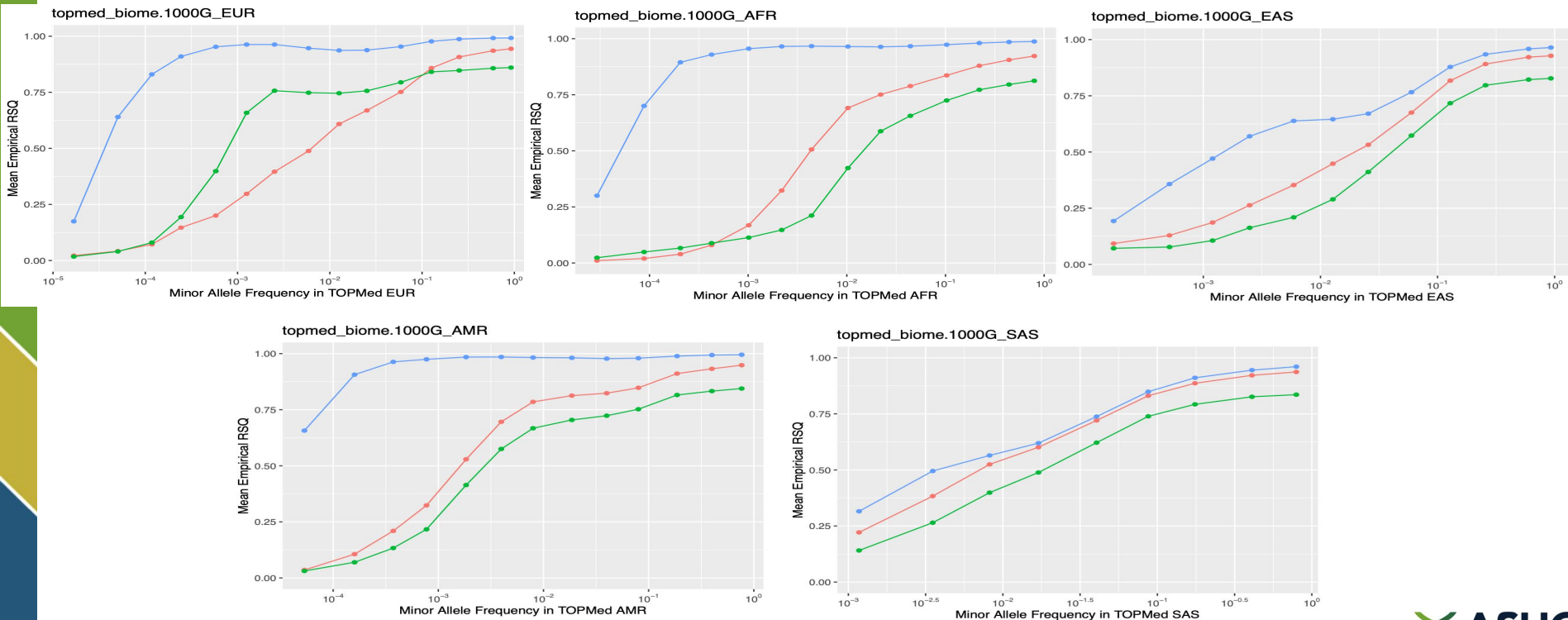
TOPMed Panel Freq Distribution

Variation type	Non-reference allele frequency bins				Totals
	(0, 0.005]	(0.005, 0.01]	(0.01, 0.05]	(0.05, 1)	
SNVs	270,352,495	3,365,284	5,330,340	7,020,861	286,068,980
Insertions	5,462,262	74,150	130,506	148,595	5,815,513
Deletions	15,406,052	185,606	297,186	333,748	16,222,592
Totals	291,220,809	3,625,040	5,758,032	7,503,204	308,107,085

Imputation Panel Quality

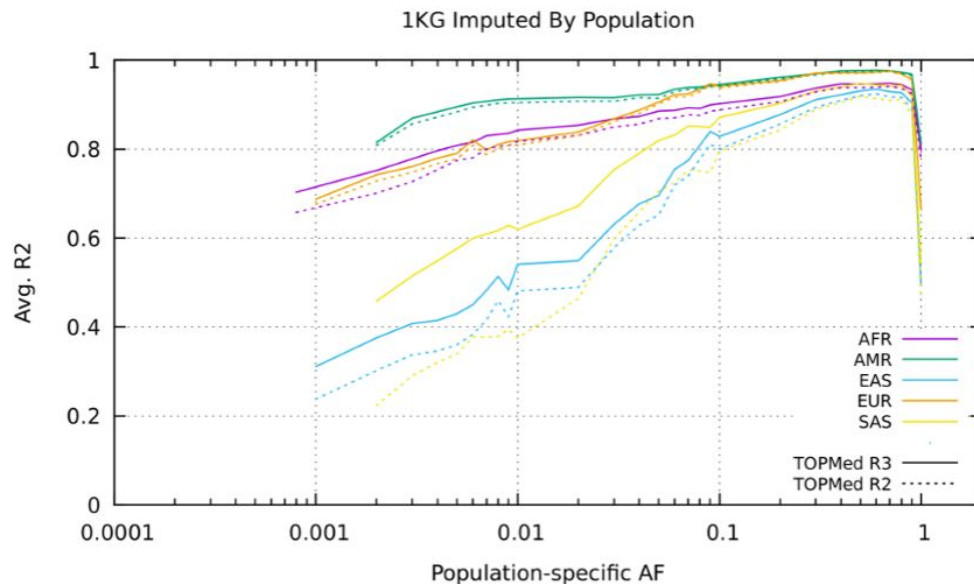


Imputation Panel Quality



Updated TOPMed R3 Panel

- Existing panel has better coverage for EUR, AFR and AMR samples
- Targeted improvement for SAS and EAS samples
- Expanded panel (97k => 143k)
- Released Dec 2023 (R2 deprecated)



TOPMed Panel Compared

	TOPMed_r3	TOPMed_r2	HRC	1000G Genomes
N samples	143K	97K	39K	2,500
Ancestry	Multiethnic	Multiethnic	European	Multiethnic
N variants	414M	308M	39M	88M
Avg. depth	38X	38X	8X	4X
Genome build	b38	b38	b37	b37
Position				

TOPMed_r2: Deprecated Dec 2025

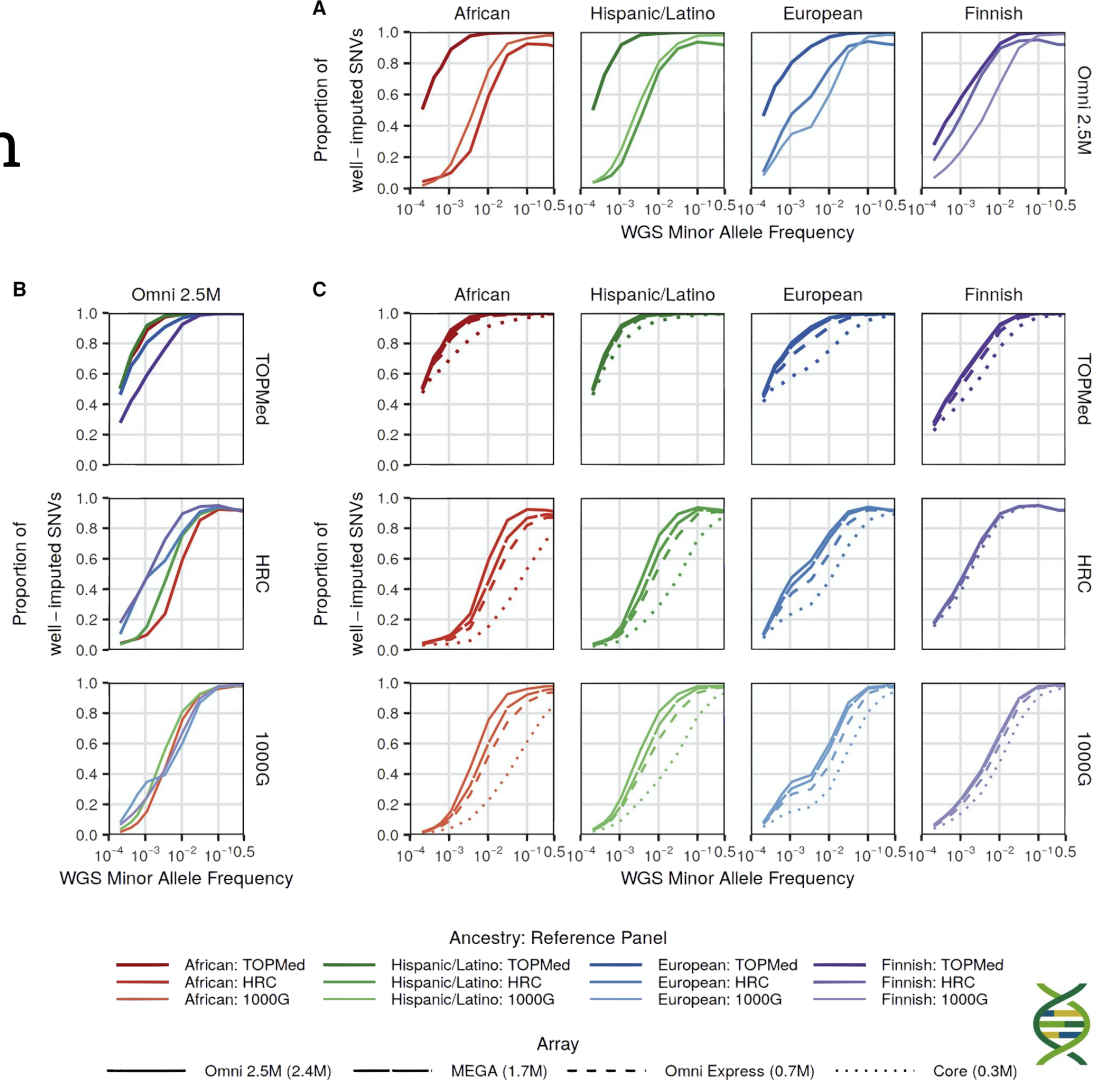
TOPMed_r3: Contains HRC and HGDP

HRC: Haplotype Reference Consortium Panel

1000G: 1000 Genomes Project Reference Panel

TOPMed Imputation Compared to WGS

- Proportion well imputed ($r^2 > 0.8$) down to MAF:
 - 0.14% in African
 - 0.11% in Hispanic/Latino
 - 0.35% in European
 - 0.85% in Finnish
- Similar performance for arrays with >700k variants
- Source: Hanks et al.
<https://doi.org/10.1016/j.ajhg.2022.07.012>



imputation.biodatacatalyst.nhlbi.nih.gov

1

National Heart, Lung, and Blood Institute

BioData CATALYST
TOPMed Imputation Server

[Home](#) [About](#) [Help](#) [Contact](#)

[Sign up](#) [Login](#)

TOPMed Imputation Server

Free Next-Generation Genotype Imputation Service

[Sign up now](#) [Login](#)

83.7M
Imputed Genomes

6353
Registered Users

1
Active Jobs

The easiest way to impute genotypes



Upload your genotypes to our secured service.



Choose a reference panel. We will take care of pre-phasing and imputation.



Download the results.
All results are encrypted with a one-time password. After 7 days, all results are deleted from our server.

The TOPMed Imputation Server is powered by software invented and developed by the [University of Michigan](#) and driven by data provided by the investigators of the [TOPMed Program](#).



ASHG
ANNUAL MEETING 2025
BOSTON, MA • OCTOBER 14-18

imputation.biobacatalyst.nih.gov

NIH National Heart, Lung, and Blood Institute BioData CATALYST TOPMed Imputation Server

Home Run Jobs About Help Contact albert

Genotype Imputation (Minimac4) 1.7.4

This is the new Michigan Imputation Server Pipeline using Minimac4. Documentation can be found [here](#).

If your input data is **GRCh37/hg19** please ensure chromosomes are encoded without prefix (e.g. **20**).
If your input data is **GRCh38hg38** please ensure chromosomes are encoded with prefix 'chr' (e.g. **chr20**). <https://imputationserver.readthedocs.io>

Run

Name

Reference Panel

[\(Details\)](#)

Input Files [\(VCF\)](#)

Select Files

Multiple files can be selected by using the **ctrl** / **cmd** or **shift** keys.

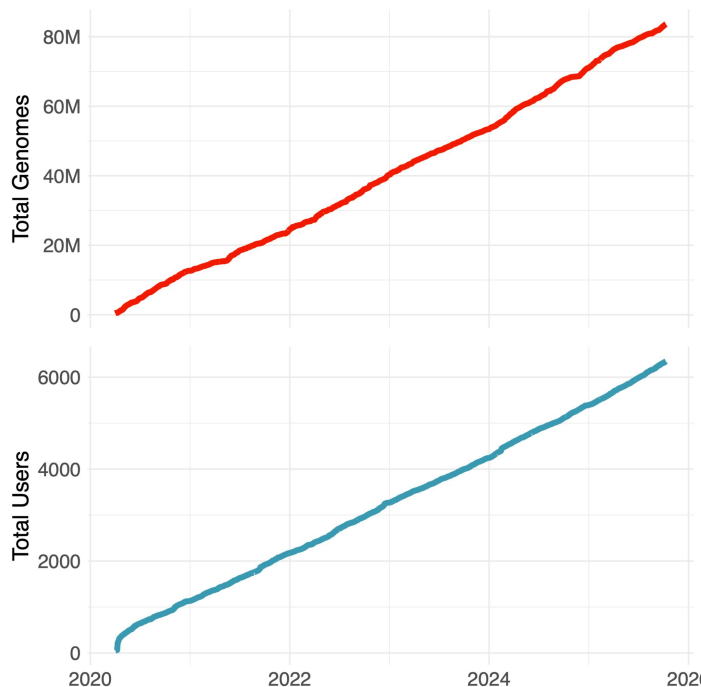
Array Build

Please note that the final SNP coordinates always match the reference build.

rsq Filter

TOPMed Imputation

- 84M genomes imputed
- Supplanted 1000g & HRC imputation for many studies
- Particularly benefits ethnically diverse cohorts
- Satisfying GDPR-related concerns of European users remains a challenge (Better served by Munich Imputation Server)



TOPMed Imputation

- TOPMed Server migrating to Nextflow based architecture in near future
- Beta testers needed
- If you have imputation ready samples, contact me <albertvs@umich.edu> or our support address <imputationserver@umich.edu>
- NOTE: *This will be the same R3 panel. Testers should have new samples.*

Imputation Resources

- Michigan Imputation Server
<https://imputation.sph.umich.edu/>
- Helmholtz Munich Imputation Server
<https://imputationserver.helmholtz-munich.de/>
- TOPMed Imputation Server
<https://imputation.biodatacatalyst.nhlbi.nih.gov/>
- MCPS Imputation Server
<https://imputationserver-reg.sph.umich.edu/>

Your questions:

Nobody has responded yet.

Hang tight! Responses are coming in.



Thank you!

Your questions

If we run out of time, please send your questions to

Christian Fuchsberger - cfuchsb@umich.edu