

Imputation Demo

dhriti.sengupta@wits.ac.za

Pre-Imputation QC

```
[dhriti@n07 Day3_Impute_input]$ ls  
chr21.bed  chr21.bim  chr21.fam
```

```
[dhriti@n07 Day3_Impute_input]$ wc -l chr21.bim  
32131  chr21.bim
```

```
[dhriti@n07 Day3_Impute_input]$ wc -l chr21.fam  
2018  chr21.fam
```

Based on a simulated dataset from African genomes

Filtering out SNPs with high missingness

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21 --geno 0.05 --allow-no-sex --make-bed --out chr21.gen  
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/  
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3  
Logging to chr21.gen.log.  
Options in effect:  
  --allow-no-sex  
  --bfile chr21  
  --geno 0.05  
  --make-bed  
  --out chr21.gen  
  
451213 MB RAM detected; reserving 225606 MB for main workspace.  
32131 variants loaded from .bim file.  
2018 people (916 males, 1093 females, 9 ambiguous) loaded from .fam.  
Ambiguous sex IDs written to chr21.gen.nosex .  
2018 phenotype values loaded from .fam.  
Using 1 thread (no multithreaded calculations invoked).  
Before main variant filters, 2018 founders and 0 nonfounders present.  
Calculating allele frequencies... done.  
Total genotyping rate is 0.983142.  
930 variants removed due to missing genotype data (--geno).  
31201 variants and 2018 people pass filters and QC.  
Among remaining phenotypes, 1009 are cases and 1009 are controls.  
--make-bed to chr21.gen.bed + chr21.gen.bim + chr21.gen.fam ... done.  
[dhriti@n07 Day3_Impute_input]$
```

Filtering out SNPs with high deviation for HWE

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21.gen0 --hwe 0.0001 --make-bed --allow-no-sex --out chr21.gen0.hwe
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.gen0.hwe.log.
Options in effect:
--allow-no-sex
--bfile chr21.gen0
--hwe 0.0001
--make-bed
--out chr21.gen0.hwe

451213 MB RAM detected; reserving 225606 MB for main workspace.
31201 variants loaded from .bim file.
2018 people (916 males, 1093 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.gen0.hwe.nosex .
2018 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2018 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.997914.
--hwe: 157 variants removed due to Hardy-Weinberg exact test.
31044 variants and 2018 people pass filters and QC.
Among remaining phenotypes, 1009 are cases and 1009 are controls.
--make-bed to chr21.gen0.hwe.bed + chr21.gen0.hwe.bim + chr21.gen0.hwe.fam ...
'
```

Filtering out low MAF SNPs

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21.genotype.hwe --maf 0.01 --allow-no-sex --make-bed --out chr21.genotype.hwe.maf
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.genotype.hwe.maf.log.
Options in effect:
--allow-no-sex
--bfile chr21.genotype.hwe
--maf 0.01
--make-bed
--out chr21.genotype.hwe.maf

451213 MB RAM detected; reserving 225606 MB for main workspace.
31044 variants loaded from .bim file.
2018 people (916 males, 1093 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.genotype.hwe.maf.nosex .
2018 phenotype values loaded from .fam.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2018 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.997983.
4896 variants removed due to minor allele threshold(s)
(--maf/--max-maf/--mac/--max-mac).
26148 variants and 2018 people pass filters and QC.
Among remaining phenotypes, 1009 are cases and 1009 are controls.
--make-bed to chr21.genotype.hwe.maf.bed + chr21.genotype.hwe.maf.bim +
chr21.genotype.hwe.maf.fam ... done.
[dhriti@n07 Day3_Impute_input]$ █
```

Removing individuals with high missingness

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21.genotype.hwe.maf --mind 0.05 --make-bed --out chr21.genotype.hwe.maf.mind
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.genotype.hwe.maf.mind.log.
Options in effect:
  --bfile chr21.genotype.hwe.maf
  --make-bed
  --mind 0.05
  --out chr21.genotype.hwe.maf.mind

451213 MB RAM detected; reserving 225606 MB for main workspace.
26148 variants loaded from .bim file.
2018 people (916 males, 1093 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.genotype.hwe.maf.mind.nosex .
2018 phenotype values loaded from .fam.
Warning: Ignoring phenotypes of missing-sex samples. If you don't want those
phenotypes to be ignored, use the --allow-no-sex flag.
3 people removed due to missing genotype data (--mind).
IDs written to chr21.genotype.hwe.maf.mind.irem .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2015 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate in remaining samples is 0.998313.
26148 variants and 2015 people pass filters and QC.
Among remaining phenotypes, 1008 are cases and 1007 are controls.
--make-bed to chr21.genotype.hwe.maf.mind.bed + chr21.genotype.hwe.maf.mind.bim +
chr21.genotype.hwe.maf.mind.fam ... done.
```

Identifying duplicate SNPs

```
[dhriti@n07 Day3_Impute_input]$ cp chr21.bim chr21.bim.backup      # For safety!
```

```
[dhriti@n07 Day3_Impute_input]$ mv chr21.bim chr21.bim.original
```

```
[dhriti@n07 Day3_Impute_input]$ head chr21.bim.original
```

21	h3a_37_21_10461773_C_T	0	10461773	T	C
21	h3a_37_21_10467257_T_C	0	10467257	C	T
21	seq-h3a_37_21_10467257_T_C	0	10467257	C	T
21	seq-h3a_37_21_10470269_C_T	0	10470269	C	T
21	seq-h3a_37_21_10475112_A_G	0	10475112	G	A

```
[dhriti@n07 Day3_Impute_input]$ awk '{$2=$1":'$4; print}' chr21.bim.original > chr21.bim
```

```
[dhriti@n07 Day3_Impute_input]$ awk '{print $2}' chr21.bim > chr21.pos.txt
```

```
[dhriti@n07 Day3_Impute_input]$ sort chr21.pos.txt | uniq -d > chr21.dup.txt
```

```
[dhriti@n07 Day3_Impute_input]$ wc -l chr21.dup.txt
```

```
313 chr21.dup.txt
```

QC in a single step

```
[dhrithi@n07 Day3_Impute_input]$ plink --bfile chr21 --geno 0.05 --hwe 0.0001 --maf 0.01 --mind 0.05 --exclude chr21.dup.txt --snps-only 'just-acgt' --allow-no-sex --make-bed --out chr21.qc
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.qc.log.
Options in effect:
--allow-no-sex
--bfile chr21
--exclude chr21.dup.txt
--geno 0.05
--hwe 0.0001
--maf 0.01
--make-bed
--mind 0.05
--out chr21.qc
--snps-only just-acgt

451213 MB RAM detected; reserving 225606 MB for main workspace.
32123 out of 32131 variants loaded from .bim file.
2018 people (916 males, 1093 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.qc.nosex .
2018 phenotype values loaded from .fam.

--exclude: 31494 variants remaining.
3 people removed due to missing genotype data (--mind).
IDs written to chr21.qc.irem .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2015 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate in remaining samples is 0.984057.
877 variants removed due to missing genotype data (--geno).
--hwe: 152 variants removed due to Hardy-Weinberg exact test.
4675 variants removed due to minor allele threshold(s)
(--maf/--max-maf/--mac/--max-mac).
25790 variants and 2015 people pass filters and QC.
Among remaining phenotypes, 1008 are cases and 1007 are controls.
--make-bed to chr21.qc.bed + chr21.qc.bim + chr21.qc.fam ... done.
```

Strand and reference allele check!

Strand Home Source Strand Files Ilmn Strand Files AB to TOP Ref/Alt <https://www.well.ox.ac.uk/~wrayner/strand/RefAlt.html>

Ref/Alt allele mapping

These files are for use in plink with the --reference-allele command, allowing Allele1 to be set to match the reference genome.
NOTE: at the present time the files are not 100% correct for indels, whilst the allele assignment is correct, the allele listed may be truncated.

BDCHP-1X10-HUMANHAP240S_11216501_A

[BDCHP-1X10-HUMANHAP240S_11216501_A-b36.strand.RefAlt.zip](#)

[BDCHP-1X10-HUMANHAP240S_11216501_A-b37.strand.RefAlt.zip](#)

[BDCHP-1X10-HUMANHAP240S_11216501_A-b38.strand.RefAlt.zip](#)

BDCHP-1x10-HUMANHAP300v1-1_11219278_C

[BDCHP-1x10-HUMANHAP300v1-1_11219278_C-b36.strand.RefAlt.zip](#)

[BDCHP-1x10-HUMANHAP300v1-1_11219278_C-b37.strand.RefAlt.zip](#)

[BDCHP-1x10-HUMANHAP300v1-1_11219278_C-b38.strand.RefAlt.zip](#)

BDCHP-1X10-HUMANHAP550_11218540_C

[BDCHP-1X10-HUMANHAP550_11218540_C-b36.strand.RefAlt.zip](#)

[BDCHP-1X10-HUMANHAP550_11218540_C-b37.strand.RefAlt.zip](#)

[BDCHP-1X10-HUMANHAP550_11218540_C-b38.strand.RefAlt.zip](#)

Cardio-Metabo_Chip_11395247_A

[Cardio-Metabo_Chip_11395247_A-b36.strand.RefAlt.zip](#)

[Cardio-Metabo_Chip_11395247_A-b37.strand.RefAlt.zip](#)

[Cardio-Metabo_Chip_11395247_A-b38.strand.RefAlt.zip](#)

cardio-metabo_chip_11395247_c

[cardio-metabo_chip_11395247_c-b36.strand.RefAlt.zip](#)

[cardio-metabo_chip_11395247_c-b37.strand.RefAlt.zip](#)

[cardio-metabo_chip_11395247_c-b38.strand.RefAlt.zip](#)

Preparing input files -Converting file to VCF format

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21.qc --recode vcf-iid --out chr21.qc.final
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.qc.final.log.
Options in effect:
  --bfile chr21.qc
  --out chr21.qc.final
  --recode vcf-iid

451213 MB RAM detected; reserving 225606 MB for main workspace.
25790 variants loaded from .bim file.
2015 people (914 males, 1092 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.qc.final.nosex .
2015 phenotype values loaded from .fam.
Warning: Ignoring phenotypes of missing-sex samples. If you don't want those
phenotypes to be ignored, use the --allow-no-sex flag.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2015 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998394.
25790 variants and 2015 people pass filters and QC.
Among remaining phenotypes, 1008 are cases and 1007 are controls.
--recode vcf-iid to chr21.qc.final.vcf ... done.
[dhriti@n07 Day3_Impute_input]$ bgzip chr21.qc.final.vcf
[dhriti@n07 Day3_Impute_input]$ tabix -p vcf chr21.qc.final.vcf.gz
[dhriti@n07 Day3_Impute_input]$
```

Zip and index the VCF file

Why use “snps-only just-acgt”

```
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21 --geno 0.05 --hwe 0.0001 --maf 0.01 --mind 0.05 --exclude chr21.dup.txt --allow-no-sex --make-bed --out chr21.qc.temp
[dhriti@n07 Day3_Impute_input]$ plink --bfile chr21.qc.temp --recode vcf-iid --out chr21.qc.final.temp
PLINK v1.90b6.18 64-bit (16 Jun 2020)          www.cog-genomics.org/plink/1.9/
(C) 2005-2020 Shaun Purcell, Christopher Chang   GNU General Public License v3
Logging to chr21.qc.final.temp.log.
Options in effect:
--bfile chr21.qc.temp
--out chr21.qc.final.temp
--recode vcf-iid

451213 MB RAM detected; reserving 225606 MB for main workspace.
25792 variants loaded from .bim file.
2015 people (914 males, 1092 females, 9 ambiguous) loaded from .fam.
Ambiguous sex IDs written to chr21.qc.final.temp.nosex .
2015 phenotype values loaded from .fam.
Warning: Ignoring phenotypes of missing-sex samples. If you don't want those
phenotypes to be ignored, use the --allow-no-sex flag.
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 2015 founders and 0 nonfounders present.
Calculating allele frequencies... done.
Total genotyping rate is 0.998393.
25792 variants and 2015 people pass filters and QC.
Among remaining phenotypes, 1008 are cases and 1007 are controls.
--recode vcf-iid to chr21.ac.final.temp.vcf ... done.
Warning: At least one VCF allele code violates the official specification;
other tools may not accept the file. (Valid codes must either start with a
'<', only contain characters in {A,C,G,T,N,a,c,g,t,n}, be an isolated '*', or
represent a breakend.)
```

Michigan Imputation Server

Free Next-Generation Genotype Imputation Service

63M

Imputed Genomes

6661

Registered Users

13

Running Jobs

Latest News

18 March 2020 Due to coronavirus-related impacts support may be slower than usual. If you haven't heard back from us after a week or so, feel free to e-mail again to check on the status of things. Take care!

07 November 2019 Updated MIS to v1.2.4! Major improvements: Minimac4 for imputation, improved chrX support, QC check right after upload, better documentation. Checkout out our [GitHub repository](#) for further information.

17 October 2019 Michigan Imputation Server @ ASHG19. All information is available [here](#).

27 November 2018

Tweets by [@umimpute](#)



Michigan Imputation Server Retweeted



Cristen Willer, PhD
@cristenw

Hooray! Best multi-ethnic imputation reference panel ever, now available! [@umimpute](#)

<https://twitter.com/umimpute/status/12483657662290698>

27



Apr 10, 2020

Michigan Imputation Server Retweeted



Mike Inouye
@minouye271

- Genotype Imputation (Minimac4)
- Genotype Imputation HLA Playground (Minimac4)
-
- Deprecated
- Genotype Imputation (Minimac3)

63M

Imputed Genomes

6661

Registered Users

13

Running Jobs

Latest News

18 March 2020 Due to coronavirus-related impacts support may be slower than usual. If you haven't heard back from us after a week or so, feel free to e-mail again to check on the status of things. Take care!

07 November 2019
Updated MIS to v1.2.4! Major improvements: Minimac4 for imputation, improved chrX support, QC check right after upload, better documentation. Checkout out our [GitHub repository](#) for further information.

17 October 2019
Michigan Imputation Server @ ASHG19. All information is available [here](#).

27 November 2018

Tweets by [@umimpute](#)

Michigan Imputation Server Retweeted

 **Cristen Willer, PhD** @cristenw
Hooray! Best multi-ethnic imputation reference panel ever, now available! @umimpute
<https://twitter.com/umimpute/status/12483657662290698>
27

Apr 10, 2020

Michigan Imputation Server Retweeted

 **Mike Inouye** @minouye271

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.

rsq Filter

off



Phasing

Eagle v2.4 (phased output)



Population

-- select an option --



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.

 I will not attempt to re-identify or contact research participants. I will report any inadvertent data release, security breach or other data management incident of which I become aware.

Submit Job

Processing of genetic personal data: Processing of genetic personal data of patients submitted by you to the University may fall under the General Data Protection Regulation (EC) 2016/679 (the "GDPR"). If that's the case, the University will be the processor of such personal data and you agree that you are the controller of the data (as such terms are defined in the GDPR). The University will process personal data submitted by you only as directed. Personal data will be processed by the University on your behalf for the purpose of computing in support of your genetic research, but only for so long as necessary to complete the computing. Once complete, the data is encrypted, made available to you, and deleted from our servers once you have retrieved it.

Genotype Imputation (Minimac4) 1.4.1

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

Test1

Reference Panel
([Details](#))

- ✓ -- select an option --
- 1000G Phase 1 v3 Shapeit2 (no singlettons) (GRCh37/hg19)
- 1000G Phase 3 v5 (GRCh37/hg19)
- CAAPA African American Panel (GRCh37/hg19)
- Genome Asia Pilot - GAsP (GRCh37/hg19)
- HapMap 2 (GRCh37/hg19)
- HRC r1.1 2016 (GRCh37/hg19)

Input Files ([VCF](#))

 Select Files

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

Run

Name

Test1

Reference Panel
([Details](#))

1000G Phase 3 v5 (GRCh37/hg38)

Input Files ([VCF](#))

File Upload

chr21.qc.fin.vcf.gz

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.

rsq Filter

off

Phasing

Eagle v2.4 (phased output)

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.

rsq Filter

off



Phasing

Eagle v2.4 (phased output)



Population

AFR



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.

 I will not attempt to re-identify or contact research participants. I will report any inadvertent data release, security breach or other data management incident of which I become aware.

Submit Job

Processing of genetic personal data: Processing of genetic personal data of patients submitted by you to the University may fall under the General Data Protection Regulation (EC) 2016/679 (the "GDPR"). If that's the case, the University will be the processor of such personal data and you agree that you are the controller of the data (as such terms are defined in the GDPR). The University will process personal data submitted by you only as directed. Personal data will be processed by the University on your behalf for the purpose of computing in support of your genetic research, but only for so long as necessary to complete the computing. Once complete, the data is encrypted, made available to you, and deleted from our servers once you have retrieved it.

Uploading Data...

rsq Filter

off

Phasing

Eagle v2.4 (phased output)



Population

AFR



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.

- I will not attempt to re-identify or contact research participants.
- I will report any inadvertent data release, security breach or other data management incident of which I become aware.

Submit Job

Processing of genetic personal data: Processing of genetic personal data of patients submitted by you to the University may fall under the General Data Protection Regulation (EC) 2016/679 (the "GDPR"). If that's the case, the University will be the processor of such personal data and you agree that you are the controller of the data (as such terms are defined in the GDPR). The University will process personal data submitted by you only as directed. Personal data will be processed by the University on your behalf for the purpose of computing in support of your genetic research, but only for so long as necessary to complete the computing. Once complete, the data is encrypted, made available to you, and deleted from our servers once you have retrieved it.

Security: The University of Michigan recognizes the importance of maintaining the security of the information it processes and maintains, and strives to ensure reasonable security measures are in place, including physical, administrative, and technical safeguards to protect personal information. The University will inform you without undue delay if it becomes aware of any unauthorized access or breach of personal data that is processed on your behalf.

GDPR, Consent, and Public Health: The University understands that personal data submitted to the University contains genetic information, which may require consent from the data subject. To the extent that you submit this type of data for processing by the University, you represent to the University that you have obtained such consent or that, pursuant to Article 9(2)(i) of the GDPR, "processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high



Test1



🕒 Mon Oct 12 2020 18:02:38 ✅ 11 sec 🚀 dhriti 🎨 Genotype Imputation (Minimac4) 1.4.1

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

Input Validation

1 valid VCF file(s) found.

Samples: 2015

Chromosomes: 21

SNPs: 25790

Chunks: 3

Datatype: unphased

Build: hg19

Reference Panel: apps@1000g-phase-3-v5 (hg19)

Population: afr

Phasing: eagle

Mode: imputation

Quality Control

Calculating QC Statistics [1/1]



Quality Control

Calculating QC Statistics

Statistics:

Alternative allele frequency > 0.5 sites: 0

Reference Overlap: 98.48 %

Match: 19,498

Allele switch: 5,512

Strand flip: 0

Strand flip and allele switch: 0

A/T, C/G genotypes: 248

Filtered sites:

Filter flag set: 0

Invalid alleles: 0

Multiallelic sites: 0

Duplicated sites: 0

NonSNP sites: 0

Monomorphic sites: 0

Allele mismatch: 140

SNPs call rate < 90%: 0

Excluded sites in total: 140

Remaining sites in total: 25,258

See [snps-excluded.txt](#) for details

Typed only sites: 392

See [typed-only.txt](#) for details

Allele mismatch: 140

SNPs call rate < 90%: 0

Excluded sites in total: 140

Remaining sites in total: 25,258

See [snps-excluded.txt](#) for details

Typed only sites: 392

See [typed-only.txt](#) for details

Quality Control (Report)

Execution successful.

Pre-phasing and Imputation

Chr 21

O

- Waiting
- Running
- Complete

Chr 21

- Waiting
- Running
- Complete

Data Compression and Encryption

Exported data.

We have sent an email to **dhriti.sengupta@wits.ac.za** with the password.



noreply@imputationserver.sph.umich.edu <noreply@imputationserver.sph.umich.edu>

Dhriti Sengupta

Monday, 12 October 2020 at 13:29

[Show Details](#)

Dear Dhriti Sengupta,

the password for the imputation results is: **InAp1BE0iob512**

The results can be downloaded from <https://imputationserver.sph.umich.edu/start.html#!jobs/job-20201012-120238-190/results>



Test1



⌚ Mon Oct 12 2020 18:02:38 ⏸ 2 h 6 min 17 sec 🚙 dhriti 🎨 Genotype Imputation (Minimac4) 1.4.1

Details

Results

Quality-Control Report

wget

qcreport.html (830 KB)

QC Statistics

wget

snps-excluded.txt (5 KB)
 typed-only.txt (6 KB)

Imputation Results

wget

chr_21.zip (1 GB)

Logs

wget

chr_21.log (731 bytes)



test1



⌚ Thu May 26 2022 17:24:26 ⏳ 2 h 32 min 35 sec 🚙 dhriti 🛡 Genotype Imputation (Minimac4) 1.6.6

Your job retires on **Fri Jun 03 2022 01:56:12!** All results will be deleted at that time. Please ensure that you have downloaded all data.

Details

Results

Quality-Control Report

wget

qcreport.html (1 MB)

QC Statistics

wget

snps-excluded.txt (106 bytes)

typed-only.txt (33 KB)

Imputation Results

wget

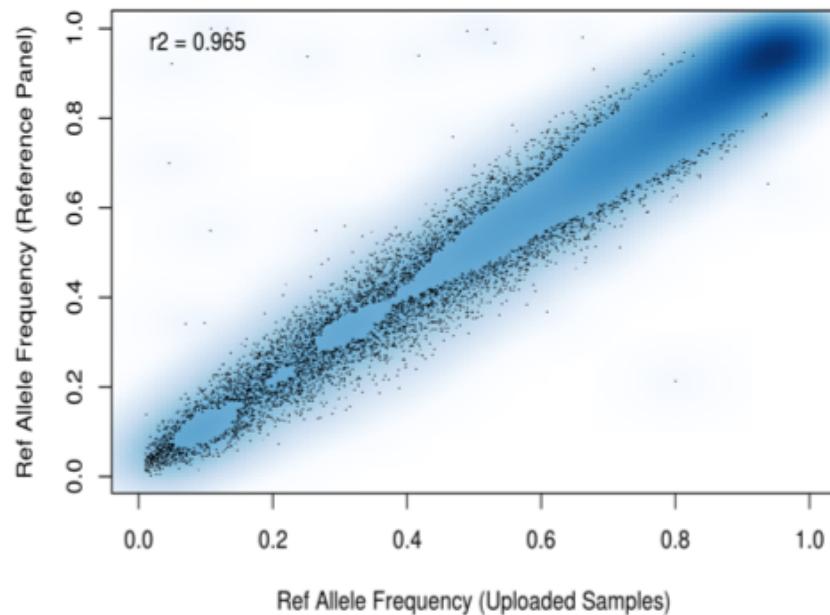
chr_21.zip (877 MB)

results.md5 (43 bytes)

Allele-Frequency Correlation

Uploaded Samples vs. Reference Panel

The plot shows the densities of frequencies falling into each part. The first 5000 points from areas of lowest regional densities will be plotted.



Potential Frequency Mismatches

Markers where chisq is greater than 300.

```
## Total mismatches: 14
```

```
## Mismatched frequencies for '21:14415250' f[A,G] = [0.4178608,0.5821392] vs [0.9402,0.0598], chisq 617.1278
## Mismatched frequencies for '21:15182128' f[A,C] = [0.1322645,0.8677354] vs [0.9992,8e-04], chisq 2241.123
## Mismatched frequencies for '21:18795734' f[T,A] = [0.06972112,0.9302789] vs [0.3411,0.6589], chisq 433.6367
## Mismatched frequencies for '21:19160902' f[C,A] = [0.04529617,0.9547038] vs [0.6989,0.3011], chisq 2043.053
## Mismatched frequencies for '21:19711057' f[A,G] = [0.2517378,0.7482622] vs [0.9372,0.0628], chisq 1175.271
## Mismatched frequencies for '21:21665581' f[C,T] = [0.5310174,0.4689826] vs [0.9682,0.0318], chisq 449.5633
## Mismatched frequencies for '21:26215690' f[C,T] = [0.01143141,0.9885686] vs [0.1392,0.8608], chisq 324.1745
## Mismatched frequencies for '21:27671435' f[C,A] = [0.4900646,0.5099354] vs [0.9947,0.0053], chisq 587.1627
## Mismatched frequencies for '21:28416202' f[A,T] = [0.5188537,0.4811463] vs [0.9977,0.0023], chisq 536.1323
## Mismatched frequencies for '21:40785851' f[A,G] = [0.938152,0.06184799] vs [0.6528,0.3472], chisq 504.34
## Mismatched frequencies for '21:42273055' f[C,A] = [0.8006455,0.1993545] vs [0.2126,0.7874], chisq 969.2262
## Mismatched frequencies for '21:46058931' f[A,G] = [0.04972376,0.9502763] vs [0.9213,0.0787], chisq 2998.947
## Mismatched frequencies for '21:46618408' f[T,C] = [0.107196,0.892804] vs [0.5492,0.4508], chisq 788.1865
## Mismatched frequencies for '21:46931004' f[T,C] = [0.1077993,0.8922007] vs [1,0], chisq 2524.323
```

#Position		#Position	FilterType	Info
21:9414787:A:G		21:14779379:C:A	Allele mismatch	Ref:C/G
21:9419380:A:G		21:14779935:A:C	Allele mismatch	Ref:A/T
21:9420811:A:G		21:15224795:C:A	Allele mismatch	Ref:C/T
21:9495801:G:T		21:15510522:T:C	Allele mismatch	Ref:T/A
21:9503254:T:C		21:16470416:G:T	Allele mismatch	Ref:G/G
21:9504140:T:C		21:16790950:A:T	Allele mismatch	Ref:A/A
21:9511443:G:A		21:16871095:A:C	Allele mismatch	Ref:A/G
21:9519678:A:G		21:17351007:C:T	Allele mismatch	Ref:T/G
21:9520544:A:G		21:17416393:A:C	Allele mismatch	Ref:A/T
21:9521794:T:C		21:17629281:T:A	Allele mismatch	Ref:T/T
21:9522805:T:C		21:18050165:T:C	Allele mismatch	Ref:T/T
21:9527292:T:C		21:18143428:G:A	Allele mismatch	Ref:G/T
21:9581058:C:T		21:18153667:T:A	Allele mismatch	Ref:T/T
21:9581965:A:G		21:18238562:C:A	Allele mismatch	Ref:C/C
21:9651350:C:T		21:18461843:C:A	Allele mismatch	Ref:C/T
21:9654378:C:G		21:18580135:A:C	Allele mismatch	Ref:C/C
21:9827728:G:C		21:18947445:A:C	Allele mismatch	Ref:C/C
21:9829715:A:G		21:19044603:G:A	Allele mismatch	Ref:G/C
21:9829819:G:A		21:19153541:C:T	Allele mismatch	Ref:T/G
21:9830700:T:G		21:19154417:A:C	Allele mismatch	Ref:A/T
21:9834527:T:C		21:19972599:T:A	Allele mismatch	Ref:A/A
21:9858714:A:G		21:20298246:A:C	Allele mismatch	Ref:A/G
21:9858773:T:C		21:20991988:G:A	Allele mismatch	Ref:A/A
21:9891936:T:C		21:21576652:C:A	Allele mismatch	Ref:C/C
21:9918900:A:T		21:21812926:G:A	Allele mismatch	Ref:G/C
21:9923066:A:G		21:22211075:C:A	Allele mismatch	Ref:C/G
21:9962070:T:C		21:22227731:T:C	Allele mismatch	Ref:T/T
21:10008665:T:A		21:22457518:G:T	Allele mismatch	Ref:G/G
		21:22829511:A:G	Allele mismatch	Ref:A/A
		21:23258485:G:T	Allele mismatch	Ref:G/G
		21:23400405:A:T	Allele mismatch	Ref:T/T

Download data



wget (1)

URLs (1)

```
wget https://imputationserver.sph.umich.edu/share/results,
```

Use the following command to download all results at once:

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



OK

Quality-Control Report

 qcreport.html (830 KB) wget

QC Statistics

 snps-excluded.txt (5 KB) wget typed-only.txt (6 KB)

Imputation Results

 chr_21.zip (1 GB) wget

Logs

 chr_21.log (731 bytes) wget

Downloading the imputed data

```
[dhriti@n07 Day3_PostImpute]$ wget https://imputationserver.sph.umich.edu/share/results/650443cf894bd6e9d19cc859b6cea873f8611f1fed304afa41dc974d0f5b35ec/chr_21.zip
```

Google Gmail Layout Copy Reset Bookmarks

```
[dhriti@n07 Day3_PostImpute]$ wget https://imputationserver.sph.umich.edu/share/results/650443cf894bd6e9d19cc859b6cea873f8611f1fed304afa41dc974d0f5b35ec/chr_21.zip--2020-10-13 02:14:54-- https://imputationserver.sph.umich.edu/share/results/650443cf894bd6e9d19cc859b6cea873f8611f1fed304afa41dc974d0f5b35ec/chr_21.zipResolving imputationserver.sph.umich.edu (imputationserver.sph.umich.edu)... 141.211.29.100Connecting to imputationserver.sph.umich.edu (imputationserver.sph.umich.edu)|141.211.29.100|:443... connected.HTTP request sent, awaiting response... 200 OKLength: 1569046086 (1.5G) [application/zip]Saving to: 'chr_21.zip' + [REDACTED]
```

Try Premium Log in Sign Up

100%[=====>] 1,569,046,086 11.1MB/s in 2m 26s

EMILY'S POV
2020-10-13 02:17:25 (10.2 MB/s) - 'chr_21.zip' saved [1569046086/1569046086]
My pale hands are shaking as I nervously step out of the yellow cab. My

YOU'LL ALSO LIKE

```
[dhriti@n07 Day3_PostImpute]$ unzip chr_21.zip
```

Archive: chr_21.zip

```
[chr_21.zip] chr21.dose.vcf.gz password:  
inflating: chr21.dose.vcf.gz
```

noreply Dhriti Sengupta Monday

```
inflating: chr21.info.gz
```



Dear Dhriti Sengupta,
the password for the imputation results is: InApI BE0iob5l2

The results can be downloaded from https://imputationserver.sph.umich.edu/share/results/650443cf894bd6e9d19cc859b6cea873f8611f1fed304afa41dc974d0f5b35ec/chr_21.zip

Know your imputed data

```
[dhriti@n07 Day3_PostImpute]$ bcftools view -h chr21.dose.vcf.gz > header.txt
[dhriti@n07 Day3_PostImpute]$ head -n 20 header.txt
##fileformat=VCFv4.1
##FILTER=<ID=PASS,Description="All filters passed">
##filedate=2020.10.12
##contig=<ID=21>
##pipeline=michigan-imputationserver-1.4.1
##imputation=minimac4-1.0.2
##phasing=eagle-2.4
##r2Filter=0.0
##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">
##bcftools_viewVersion=1.10.2-91-g365d117+htslib-1.10.2-106-g9c35744
[dhriti@n07 Day3_PostImpute]$ █
```

Know your imputed data.. contd

Alt Allele
frequency

Imputation
accuracy

bcftools view -H chr21.dose.vcf.gz head -n 20 cut -f 1-10 > vcf_info.txt									
[[dhriti@n07 Day3_PostImpute]\$ bcftools view -H chr21.dose.vcf.gz head -n 20 cut -f 1-10 > vcf_info.txt									
21	9411245	21:9411245:C:A	C	A	.	PASS	AF=0.00023;MAF=0.00023;R2=0.00099;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411302	21:9411302:G:T	G	T	.	PASS	AF=0.00294;MAF=0.00294;R2=0.05542;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411313	21:9411313:G:A	G	A	.	PASS	AF=0.00548;MAF=0.00548;R2=0.0639;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411332	21:9411332:G:T	G	T	.	PASS	AF=0.00016;MAF=0.00016;R2=0.00208;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411347	21:9411347:G:C	G	C	.	PASS	AF=0.00305;MAF=0.00305;R2=0.00582;IMPUTED	GT:DS:HDS:GP	0 0:0.002:0.002,0:0.998,0.00
,	,	,	,	,	,	,	,	,	,
21	9411358	21:9411358:C:T	C	T	.	PASS	AF=0.00123;MAF=0.00123;R2=0.00115;IMPUTED	GT:DS:HDS:GP	0 0:0.001:0.001,0:0.999,0.00
,	,	,	,	,	,	,	,	,	,
21	9411381	21:9411381:G:T	G	T	.	PASS	AF=0.01138;MAF=0.01138;R2=0.21607;IMPUTED	GT:DS:HDS:GP	0 0:0.022:0,0.021:0.978,0.02
,	,	,	,	,	,	,	,	,	,
21	9411409	21:9411409:T:C	T	C	.	PASS	AF=0.00049;MAF=0.00049;R2=0.0009;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411410	21:9411410:C:T	C	T	.	PASS	AF=0.40383;MAF=0.40383;R2=0.07737;IMPUTED	GT:DS:HDS:GP	1 0:0.972:0.502,0.47:0.264,0
5,0.236									
21	9411417	21:9411417:C:T	C	T	.	PASS	AF=0.00047;MAF=0.00047;R2=0.01729;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411441	21:9411441:T:G	T	G	.	PASS	AF=0.00085;MAF=0.00085;R2=0.07319;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411446	21:9411446:C:T	C	T	.	PASS	AF=0.01232;MAF=0.01232;R2=0.24179;IMPUTED	GT:DS:HDS:GP	0 0:0.043:0.001,0.041:0.957,
.043,0									
21	9411449	21:9411449:G:T	G	T	.	PASS	AF=0.00054;MAF=0.00054;R2=0.00025;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411476	21:9411476:A:T	A	T	.	PASS	AF=0.00068;MAF=0.00068;R2=0.00065;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411485	21:9411485:C:A	C	A	.	PASS	AF=0.00045;MAF=0.00045;R2=0.00033;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411486	21:9411486:C:A	C	A	.	PASS	AF=0.00041;MAF=0.00041;R2=0.0003;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411497	21:9411497:A:G	A	G	.	PASS	AF=0.007;MAF=0.007;R2=0.06293;IMPUTED	GT:DS:HDS:GP	0 0:0.021:0.001,0.02:0.979,0.021,0
21	9411500	21:9411500:G:T	G	T	.	PASS	AF=0.37488;MAF=0.37488;R2=0.07131;IMPUTED	GT:DS:HDS:GP	0 0:0.581:0.15,0.431:0.484,0
451,0.065									
21	9411542	21:9411542:T:A	T	A	.	PASS	AF=0.00027;MAF=0.00027;R2=6e-05;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0
21	9411601	21:9411601:T:C	T	C	.	PASS	AF=0.00043;MAF=0.00043;R2=0.08606;IMPUTED	GT:DS:HDS:GP	0 0:0:0,0:1,0,0

Filtering by genotyping probability and other parameters

The first step can take 10-20 minutes depending on your computational resources !!!

```
[dhrifti@n07 Day3_PostImpute]$ bcftools filter chr21.dose.vcf.gz -e "F_MISSING > 0.05 || MAF < 0.01 || R2<0.6" | bcftools convert -Oz -o chr21.filter.vcf.gz  
[dhrifti@n07 Day3_PostImpute]$ bcftools view -H chr21.filter.vcf.gz | head -n 20 | cut -f 1-10 > vcf_info_filt.txt
```

```
[dhriti@n07 Day3_PostImpute]$ vi vcf_info_filt.txt
```

Higher Imputation accuracy

21	14488541	21:14488541:T:A T	A	.	PASS	AF=0.03287;MAF=0.03287;R2=0.72035;IMPUTED	GT:DS:HDS:
GP	0 0:0:0,0:1,0,0						
21	14500720	21:14500720:G:A G	A	.	PASS	AF=0.01089;MAF=0.01089;R2=0.73474;IMPUTED	GT:DS:HDS:
GP	0 0:0:0,0:1,0,0						
21	14501861	21:14501861:C:T C	T	.	PASS	AF=0.01089;MAF=0.01089;R2=0.73474;IMPUTED	GT:DS:HDS:
GP	0 0:0:0,0:1,0,0						
21	14505954	21:14505954:T:C T	C	.	PASS	AF=0.0219;MAF=0.0219;R2=0.68842;IMPUTED	GT:DS:HDS:GP 01
	0:0:0,0:1,0,0						

Index the zipped VCF file and check the number of SNPs

```
[dhriti@n07 Day3_PostImpute]$ tabix -p vcf chr21.filter.vcf.gz  
[dhriti@n07 Day3_PostImpute]$ bcftools index -n chr21.filter.vcf.gz  
234475  
[dhriti@n07 Day3_PostImpute]$
```

Convert back to PLINK

```
[dhriti@n07 Day3_PostImpute]$ plink --vcf chr21.filter.vcf.gz --hwe 1e-5 --keep-allele-order --allow-no-sex --make-bed  
--out chr21.filter.qc  
PLINK v1.90b6.25 64-bit (5 Mar 2022)          www.cog-genomics.org/plink/1.9/  
(C) 2005-2022 Shaun Purcell, Christopher Chang   GNU General Public License v3  
Logging to chr21.filter.qc.log.  
Options in effect:  
  --allow-no-sex  
  --hwe 1e-5 _____ HWE filter  
  --keep-allele-order  
  --make-bed  
  --out chr21.filter.qc  
  --vcf chr21.filter.vcf.gz  
  
451213 MB RAM detected; reserving 225606 MB for main workspace.  
--vcf: chr21.filter.qc-temporary.bed + chr21.filter.qc-temporary.bim +  
chr21.filter.qc-temporary.fam written.  
234475 variants loaded from .bim file.  
2015 people (0 males, 0 females, 2015 ambiguous) loaded from .fam.  
Ambiguous sex IDs written to chr21.filter.qc.nosex .  
Using 1 thread (no multithreaded calculations invoked).  
Before main variant filters, 2015 founders and 0 nonfounders present.  
Calculating allele frequencies... done.  
Total genotyping rate is exactly 1.  
--hwe: 150 variants removed due to Hardy-Weinberg exact test.  
234325 variants and 2015 people pass filters and QC.  
Note: No phenotypes present.  
--make-bed to chr21.filter.qc.bed + chr21.filter.qc.bim + chr21.filter.qc.fam  
... done.
```