

Genome Reporting for Healthy Populations—Pipeline for Genomic Screening from the GENCOV COVID-19 Study

Erika Frangione,^{1,2} Monica Chung,^{1,2} Selina Casalino,^{1,2} Georgia MacDonald,^{1,2} Sunakshi Chowdhary,^{1,2} Chloe Mighton,^{1,2,3,4} Hanna Faghfoury,⁵ Yvonne Bombard,^{3,4} Lisa Strug,⁶ Trevor Pugh,^{5,7} Jared Simpson,⁷ Limin Hao,⁸ Matthew Lebo,^{8,9} William J. Lane,⁹ Jennifer Taher,^{1,3} Jordan Lerner-Ellis,^{1,2,3,10} and GENCOV Study Workgroup^{1,*}

¹Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

²Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada

³University of Toronto, Toronto, Ontario, Canada

⁴Unity Health Toronto, Toronto, Ontario, Canada

⁵University Health Network, Toronto, Ontario, Canada

⁶The Hospital for Sick Children, Toronto, Ontario, Canada

⁷Ontario Institute for Cancer Research, Toronto, Ontario, Canada

⁸Laboratory of Molecular Medicine, Partners Personalized Medicine, Boston, Massachusetts

⁹Harvard Medical School & Brigham and Women's Hospital, Boston, Massachusetts

¹⁰Corresponding author. *Jordan.Lerner-Ellis@sinaihealth.ca*

*GENCOV Study Workgroup: Saranya Arnoldo^{1,2}, Navneet Aujla¹, Erin Bearss³, Alexandra Binnie², Bjug Borgundvaag^{1,3}, Laurent Briollais⁸, Howard Chertkow⁴, Marc Clausen⁵, Marc Dagher^{1,6}, Luke Devine^{1,3}, David Di Iorio¹, Steven Marc Friedman^{3,7}, Chun Yiu Jordan Fung^{3,8}, Anne-Claude Gingras^{1,3,8}, Lee W. Goneau⁹, Deepanjali Kaushik², Zeeshan Khan¹⁰, Elisa Lapadula^{3,8}, Tiffany Lu³, Tony Mazzulli^{1,3}, Allison McGeer^{1,3,8}, Shelley L. McLeod^{1,3}, Gregory Morgan^{1,3,8}, David Richardson², Seth Stern¹⁰, Ahmed Taher^{1,7,10}, Iris Wong¹⁰, and Natasha Zarei¹⁰

¹University of Toronto, Toronto, Ontario, Canada

²William Osler Health System, Toronto, Ontario, Canada

³Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

⁴Baycrest Health Sciences, Toronto, Ontario, Canada

⁵Unity Health Toronto, Toronto, Ontario, Canada

⁶Women's College Hospital, Toronto, Ontario, Canada

⁷University Health Network, Toronto, Ontario, Canada

⁸Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario, Canada

⁹Dynacare Medical Laboratories, Toronto, Ontario, Canada

¹⁰Mackenzie Health, Richmond Hill, Ontario, Canada

Published in the Human Genetics section

Genome sequencing holds the promise for great public health benefits. It is currently being used in the context of rare disease diagnosis and novel gene identification, but also has the potential to identify genetic disease risk factors in healthy individuals. Genome sequencing technologies are currently being used to identify genetic factors that may influence variability in symptom severity and immune response among patients infected by SARS-CoV-2. The GENCOV study aims to look at the relationship between genetic, serological, and biochemical factors and variability of SARS-CoV-2 symptom severity, and to evaluate the utility of returning genome screening results to study participants. Study participants select which results they wish to receive with a decision aid. Medically actionable information for diagnosis, disease risk estimation, disease prevention, and patient management are provided in a comprehensive genome

report. Using a combination of bioinformatics software and custom tools, this article describes a pipeline for the analysis and reporting of genetic results to individuals with COVID-19, including HLA genotyping, large-scale continental ancestry estimation, and pharmacogenomic analysis to determine metabolizer status and drug response. In addition, this pipeline includes reporting of medically actionable conditions from comprehensive gene panels for Cardiology, Neurology, Metabolism, Hereditary Cancer, and Hereditary Kidney, and carrier screening for reproductive planning. Incorporated into the genome report are polygenic risk scores for six diseases—coronary artery disease; atrial fibrillation; type-2 diabetes; and breast, prostate, and colon cancer—as well as blood group genotyping analysis for ABO and Rh blood types and genotyping for other antigens of clinical relevance. The genome report summarizes the findings of these analyses in a way that extensively communicates clinically relevant results to patients and their physicians. © 2022 Wiley Periodicals LLC.

Basic Protocol 1: HLA genotyping and disease association

Basic Protocol 2: Large-scale continental ancestry estimation

Basic Protocol 3: Dosage recommendations for pharmacogenomic gene variants associated with drug response

Support Protocol: System setup

Keywords: bioinformatics • COVID-19 • genome reporting • genomics • next generation sequencing (NGS)

How to cite this article:

Frangione, E., Chung, M., Casalino, S., MacDonald, G., Chowdhary, S., Mighton, C., Faghfoury, H., Bombard, Y., Strug, L., Pugh, T., Simpson, J., Hao, L., Lebo, M., Lane, W. J., Taher, J., Lerner-Ellis, J., & GENCOV Study Workgroup. (2022). Genome reporting for healthy populations—pipeline for genomic screening from the GENCOV COVID-19 study. *Current Protocols*, 2, e534. doi: 10.1002/cpz1.534

INTRODUCTION

The implementation of serological and molecular tools to inform COVID-19 patient management—the GENCOV study (Taher et al., 2021)—incorporates genome sequencing (GS) technologies to identify and characterize genetic factors that may have an impact on symptom severity and immune response in SARS-CoV-2-infected patients. Sequencing of host genomes may provide useful insights for clinical diagnosis, disease prevention, and patient management strategies. In the GENCOV study, medically relevant genetic results and other genetic findings from the analysis of host genome sequencing data are communicated to participants in a comprehensive genome report. The GENCOV study enrolled approximately 1500 patients with a PCR-positive COVID-19 nasopharyngeal swab. Blood samples were collected at baseline, 1 month, 6 months, and 1 year after diagnosis. Antibody isotype (IgG, IgA, and IgM), titers, and viral neutralization were analyzed and returned to study participants. DNA used for host genome sequencing was isolated from blood lymphocytes. Viral genome lineages were also sequenced and returned. Pre- and post-test counseling is provided by a certified genetic counsellor to manage clinical referrals by a medical geneticist.

This article describes a bioinformatic pipeline for the purpose of returning genetic results in a genome report. The genome report includes HLA genotyping predictions, an estimation of individual large-scale continental ancestry, and pharmacogenomic

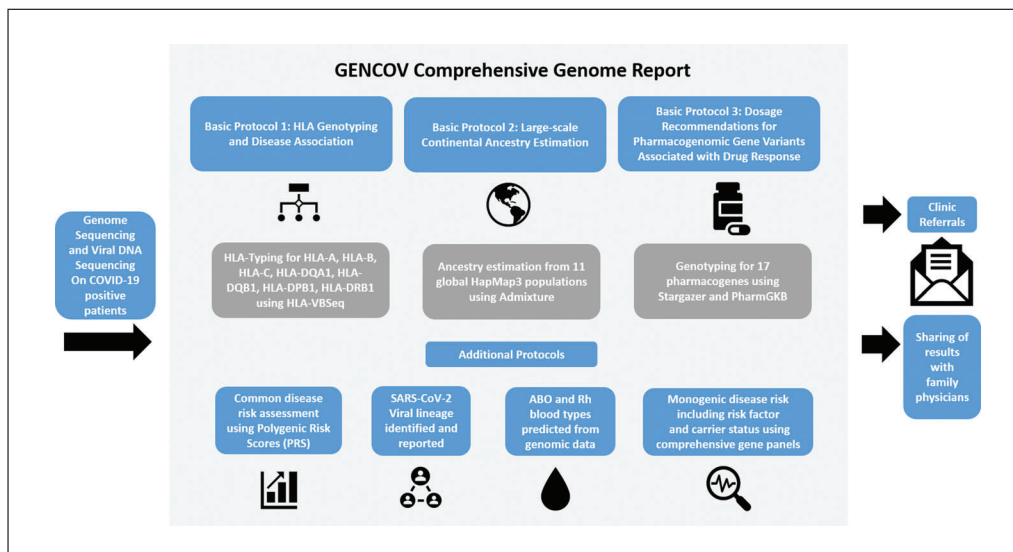


Figure 1 Components of the GENCOV Comprehensive Genome Report to be shared with participants and family physicians. Elements of the genomic report include basic protocols developed for HLA genotyping and disease association, large-scale continental ancestry estimation, and pharmacogenomic dosage recommendations. Additional protocols incorporated into the report have been adapted from collaborators, which include PRS score assessment, ABO and Rh blood group genotyping, and viral lineage identification. An additional protocol for determining monogenic disease risk was also developed in-house and included on the report.

variation for drug response and dosage recommendations. In addition, this report includes existing algorithms to determine patient ABO, Rh, and other blood group genotypes, and an assessment of common disease risk by calculating polygenic risk score (PRS) for six common diseases (atrial fibrillation, coronary artery disease, type 2 diabetes, prostate cancer, colorectal cancer, and breast cancer). The genomes are examined for clinically significant variation related to genetic conditions for the following categories: blood and immunology, endocrine, metabolic/mitochondrial, musculoskeletal, hearing loss, neurology, cardiology, ophthalmology, renal, skin, gastrointestinal disorders, and hereditary cancer. Any clinically significant gene variants detected are assessed using the American College of Medical Genetics (ACMG/AMP) criteria for interpreting sequence variation (Richards et al., 2015). SARS-CoV-2 viral lineage was also determined and included on the report (see Fig. 1, and File S1 in Supporting Information).

The genomic analyses for this study are being performed in-house provided through the Canadian Genomics COVID-19 Network (CanCOGeN) at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children. Base calling is performed using Illumina HiSeq 2500 to generate standard genome sequencing output in the form of FASTQ read files. After quality control (QC), metrics are applied to validate sequence generation, and the FASTQ reads are aligned to the GRCh37 reference human genome using Burrows-Wheeler Aligner (BWA; Li & Durbin, 2009). The Genome Analysis Toolkit (GATK) is used to determine copy number variants (CNV) and structural variants (SVs; Mckenna et al., 2010). Detected variants are annotated using a custom pipeline based on Annovar, which is subsequently outputted into a .tsv file format for further analysis. In addition to the standard FASTQ output, Variant Call Format (VCF) files for SNV and indel variants, including SV and CNV calls, are generated through secondary analysis by a third-party vendor, the Franklin Genoox Platform (<https://www.genoox.com>). In our current genomic analysis pipelines for the study, FASTQ, VCF, and Binary Alignment Map (BAM) files are used as input.

Basic Protocol 1 explains how to utilize an HLA genotyping software application, HLA-VBSeq, to generate estimated genotypes for 22 HLA loci, including *HLA-A*,

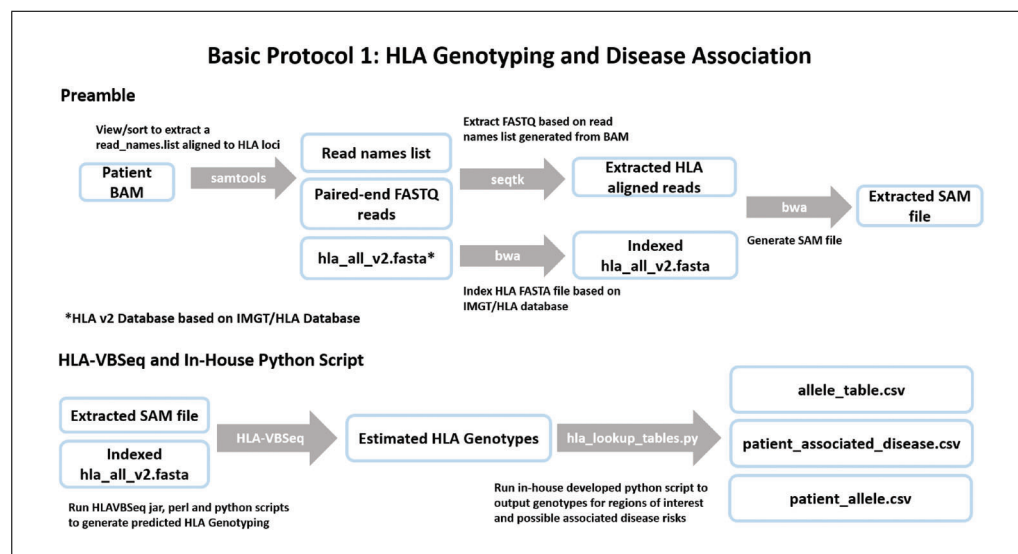


Figure 2 Basic Protocol 1 outlines the pipeline for HLA genotyping and disease association as included on the genome report. This pipeline utilizes the software HLA-VBSeq for HLA genotype calls in combination with an in-house-developed database containing HLA-disease associations as identified from the literature.

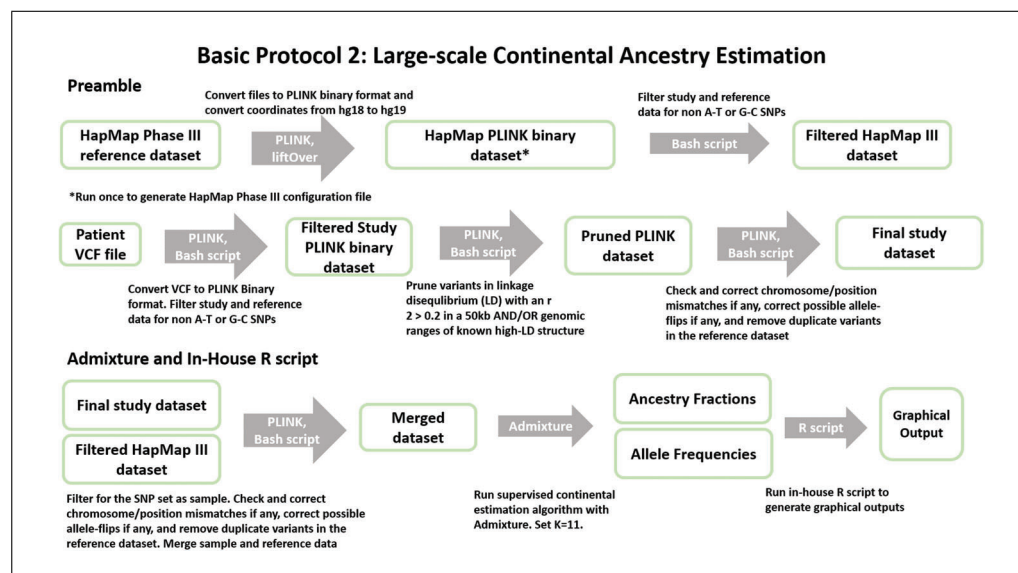


Figure 3 Basic Protocol 2 outlines the pipeline for large-scale continental ancestry estimation as included on the genome report. This pipeline utilizes the HapMap Phase III reference dataset and ADMIXTURE to determine the admixed ancestral populations across GENCOV participants.

-B, and -C. Using the predicted genotypes from this output, custom python scripts were developed to map the genotypes to an in-house database containing relevant HLA-disease associations (Fig. 2). Basic Protocol 2 describes a pipeline for estimating individual large-scale continental ancestry against a reference population dataset, HapMap3, to generate an individual admixed ancestry report in a graphical representation (Fig. 3). Basic Protocol 3 describes the genotyping of 17 relevant pharmacogenomics genes using the software Stargazer, and maps these identified genotypes to current dosage recommendations from the PharmGKB database (Fig. 4). In order to run all three protocols, users will have the option to git clone from our repository or alternatively access Docker images from DockerHub. The protocols outlined in this pipeline have been validated against the NA12878 genome available through Illumina Platform Pedigree (Eberle et al., 2017). Downloads and

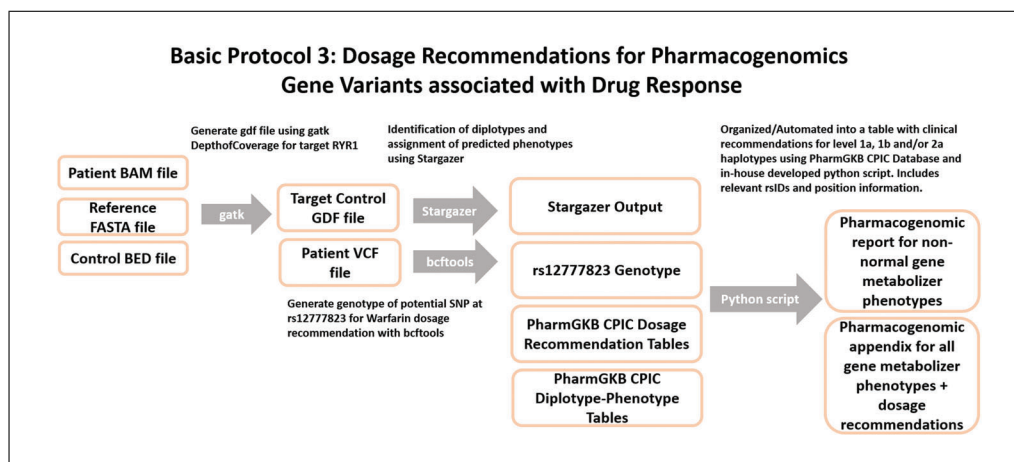


Figure 4 Basic Protocol 3 outlines the pipeline for pharmacogenomics dosage recommendations as included on the genome report. This pipeline utilizes the Stargazer software and formal PharmGKB recommendations to identify pharmacogenomics variants with clinically relevant gene-drug interactions across GENCOV participants.

generation of this sample data are included as script in each of the Basic Protocol github repositories.

HLA GENOTYPING AND DISEASE ASSOCIATION

Basic Protocol 1 explains how to utilize the HLA-VBSeq software to generate estimated HLA genotypes for 22 HLA genes, including *HLA-A*, *HLA-B*, *HLA-C*. Using the output from HLA-VBSeq, python scripts were developed to map predicted genotypes to an in-house developed database containing relevant HLA-disease associations.

Necessary Resources

Hardware

This pipeline has been tested on both Linux and Mac OS

Software

This pipeline is available for distribution through github and dockerhub. In addition, the user must have HLA-VBSeq (<http://nagasakilab.csmil.org/hla/>, RRID:SCR_022284), samtools-1.14 (<http://htslib.org/>, RRID:SCR_002105), bwa-0.7.17 (<http://bio-bwa.sourceforge.net/>, RRID:SCR_010910), seqtk-1.3 (<https://github.com/lh3/seqtk>, RRID:SCR_018927), gatk-4.2.4.1 (<https://software.broadinstitute.org/gatk/>, RRID:SCR_001876), python 2 and 3, Java8, and perl installed on their system.

Files

A sample's BAM, VCF and paired-end FASTQ files. The BAM and VCF files must be generated from the paired-end FASTQ read files being aligned to the GRCh37 (or equivalent) reference human genome. In addition, the VCF file must contain SNV and indel variants.

1. Download the HLA repository from <https://github.com/jlelabs/HLA>. This repository can be downloaded on Github's web interface or cloned using the git command line.

```
git clone https://github.com/jlelabs/HLA
```

2. Download the following files from HLA-VBSeq at <http://nagasakilab.csmil.org/hla/> and transfer them into your Configuration folder within the HLA repository.

```
HLAVBSeq.jar
```

BASIC PROTOCOL 1


```

parse_result.pl
call_hla_digits.py
hla_all_v2.fasta
Allelelist_v2.txt

```

3. Go into the HLA repository using the following command in terminal. The HLA_Repository_Path will be the file path in which the repository was either downloaded or cloned.

```
cd ${HLA_Repository_Path}
```

4. To run the HLA pipeline, input the following into terminal:

```

./HLA.sh -i <SAMPLE_ID> -v <VCF> -b <BAM> --f1 <FASTQ1> --f2 <FASTQ2> -o
<OUTPUT_DIRECTORY>

```

There are six mandatory arguments:

- i: the sample id, or another desired name designated by the user that will be used to name all outputs.
- v: the path to the sample VCF file.
- b: the path to the sample BAM file.
- f1: the path to the sample FASTQ read 1 file.
- f2: the path to the sample FASTQ read 2 file.
- o: the file path of the desired output directory, in which all files will be saved, including final results and script logs.

BASIC PROTOCOL 2

LARGE-SCALE CONTINENTAL ANCESTRY ESTIMATION

Basic Protocol 2 outlines how to estimate individual ancestry against a reference HapMap3 set using PLINK (Purcell et al., 2007) and ADMIXTURE software (Alexander, Novembre, & Lange, 2009). PLINK was used for the filtering and merging, while ADMIXTURE was used to output subject ancestry. In-house R scripts were developed to produce a visual representation of a given individual's results. Expected outputs for this pipeline can be found in the Guidelines for Understanding Results section.

Necessary Resources

Hardware

This pipeline has been tested on both Linux and Mac OS

Software

This pipeline is available for distribution through github and dockerhub. The user must have their specific OS distributions of PLINK and PLINK2 (<http://www.nitrc.org/projects/plink>, RRID:SCR_001757), liftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>, RRID:SCR_018160), ADMIXTURE-1.3.0 (<http://www.genetics.ucla.edu/software/ADMIXTURE/>, RRID:SCR_001263), and R installed in their system.

Files

A sample VCF file containing SNV and indel variants. The VCF file must be generated from the paired end FASTQ read files being aligned to the GRCh37 (or equivalent) reference human genome.

1. Download the Ancestry repository from <https://github.com/jlelals/Ancestry>. This repository can be downloaded on Github's web interface or cloned using the git command line.

```
git clone https://github.com/jlelals/Ancestry
```

2. Go into the Ancestry repository using the following command in terminal. The Ancestry_Repository_Path will be the file path in which the repository was either downloaded or cloned.

```
cd ${Ancestry_Repository_Path}
```

3. Run the following command to download the reference HapMap3 set. This will download and configure the needed reference files, and will only have to be run once.

```
./ancestry_hapmap3.sh
```

4. To run the Ancestry pipeline, input the following into terminal.

```
./ancestry.sh -i <SAMPLE_ID> -v <VCF> -o <OUTPUT_DIRECTORY>
```

There are three mandatory arguments:

- i: the sample id, or another desired name designated by the user that will be used to name all outputs.
- v: the path to the sample VCF file.
- o: the file path of the desired output directory, in which all files will be saved, including final results, and script logs.

DOSAGE RECOMMENDATIONS FOR PHARMACOGENOMIC GENE VARIANTS ASSOCIATED WITH DRUG RESPONSE

BASIC PROTOCOL 3

Basic Protocol 3 was developed to assign individual metabolizer phenotype status and link appropriate recommendations using bcftools, GATK, and Stargazer. The Stargazer software was used to identify the genotype of 17 pharmacogenes of interest, while an in-house python script was developed to generate recommendations. The recommendations and metabolizer phenotypes were adapted from the diplotype-phenotype tables and the gene-medication dosing guidelines from PharmGKB and the Clinical Pharmacogenetics Implementation Consortium (CPIC). Expected outputs for this pipeline can be found in the Guidelines for Understanding Results section.

Necessary Resources

Hardware

This pipeline has been tested on both Linux and Mac OS

Software

This pipeline is available for distribution through github and dockerhub. The user must have Stargazer_v1.0.8 (<https://stargazer.gs.washington.edu/stargazerweb/index.html>), bcftools-1.14 (<http://samtools.sourceforge.net/mpileup.shtml>, RRID:SCR_005227), gatk-4.2.4.1 (<https://software.broadinstitute.org/gatk/>, RRID:SCR_001876), python2 and python3, Java8, and perl installed on their system.

Files

A sample's BAM and VCF files. The BAM and VCF files must be generated from the paired end FASTQ read files being aligned to the GRCh37 (or equivalent) reference human genome. In addition, the VCF file must contain SNV and indel variants.

1. Download the Pharmacogenomics repository from <https://github.com/jlelabs/Pharmacogenomics>. This repository can be downloaded on Github's web interface or cloned using git command line.

```
git clone https://github.com/jlelabs/Pharmacogenomics
```

2. Download the Stargazer software at <https://stargazer.gs.washington.edu/stargazerweb/res/form.html> and move it into the Configuration folder.
3. Go into the Pharmacogenomics repository using the following command in terminal. The Pharmacogenomics_Repository_Path will be the file path in which the repository was either downloaded or cloned.

```
cd ${Pharmacogenomics_Repository_Path}
```

4. To run the Pharmacogenomics pipeline, input the following into terminal.

```
./Pharmacogenomics.sh -i <SAMPLE_ID> -v <VCF> -b <BAM> -o  
<OUTPUT_DIRECTORY>
```

There are four mandatory arguments:

- i: the sample id, or another desired name designated by the user that will be used to name all outputs.
- v: the path to the sample VCF file.
- b: the path to the sample BAM file.
- o: the file path of the desired output directory, in which all files will be saved, including final results and script logs.

SYSTEM SETUP

Necessary Resources

Hardware

This pipeline has been tested on both Linux and Mac OS. Minimum RAM requirements are 16 GB of RAM and 4 CPUs.

Software

This pipeline requires Docker installation

Files

A sample's BAM, VCF, and paired-end FASTQ files. The BAM and VCF files must be generated from the paired end FASTQ read files being aligned to the GRCh37 (or equivalent) reference human genome. In addition, the VCF file must contain SNV and indel variants.

Using Docker and DockerHub

1. Download the desired repository from DockerHub.

Basic Protocol 1:

```
docker pull monicahzchung/hla
```

Basic Protocol 2:

```
docker pull monicahzchung/ancestry
```

Basic Protocol 3:

```
docker pull monicahzchung/pharma
```

2. To run the images, execute the following commands:

Basic Protocol 1:

```
docker run -v INPUT_DIRECTORY:/INPUT/ \  
-v OUTPUT_DIRECTORY:/OUTPUT \  
hla \  
sample_id \  

```



```

/INPUT/VCF \
/INPUT/BAM \
/INPUT/FASTQ1 \
/INPUT/FASTQ2 \
/OUTPUT

```

Basic Protocol 2:

```

docker run -v INPUT_DIRECTORY:/INPUT/ \
-v OUTPUT_DIRECTORY:/OUTPUT \
ancestry \
sample_id \
/INPUT/VCF \
/OUTPUT

```

Basic Protocol 3:

```

docker run -v INPUT_DIRECTORY:/INPUT/ \
-v OUTPUT_DIRECTORY:/OUTPUT \
pharma \
sample_id \
/INPUT/VCF \
/INPUT/BAM \
/OUTPUT

```

Manually installing dependencies

NOTE: It is recommended to use the most up-to-date software installations where possible for the basic protocols outlined above.

1. Python2
2. Python3.85+ with the following packages:
 - numpy
 - pandas
 - xlrd
 - openpyxl
3. Java8
4. Perl
5. bwa-0.7.17 (<http://bio-bwa.sourceforge.net/>, RRID:SCR_010910)
6. bcftools-1.14 (<http://samtools.sourceforge.net/mpileup.shtml>, RRID:SCR_005227)
7. samtools-1.14 (<http://htslib.org/>, RRID:SCR_002105)
8. seqtk-1.3 (<https://github.com/lh3/seqtk>, RRID:SCR_018927)
9. gatk-4.2.4.1 (<https://software.broadinstitute.org/gatk/>, RRID:SCR_001876)
10. Stargazer v1.0.8 (<https://stargazer.gs.washington.edu/stargazerweb/index.html>)
11. R 4.1.2 with the following packages:
 - tidyverse
 - dplyr
 - RColorBrewer
 - reshape2
12. ADMIXTURE-1.3.0 (<http://www.genetics.ucla.edu/software/ADMIXTURE/>, RRID:SCR_001263)

13. PLINK and PLINK2 (<http://www.nitrc.org/projects/plink>, RRID:SCR_001757)
14. HLA-VBSeq (<http://nagasakilab.csml.org/hla/>, RRID:SCR_022284)
15. liftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>, RRID:SCR_018160)
 - a. Chain is included (hg18ToHg19.over.chain) within Basic Protocol 2: Large-scale Continental Ancestry Estimation Configuration folder.

GUIDELINES FOR UNDERSTANDING RESULTS

Outputs

Each of the genomic analysis pipelines requires a mandatory output directory in which the user must specify with the -o flag. In the specified file path, the following folders and files will be generated. The expected results and directory structure for each of the Basic Protocols are explained in detail below and were validated using the NA12878 genome. The table of possible HLA-disease associations can be found in File S2 in Supporting Information.

Basic Protocol 1: HLA Genotyping and Disease Association

Folder structure

```
sample_output_directory/
| - sample_HLA_REPORT/
|   | - sample_allele_table.csv
|   | - sample_patient_allele_table.csv
|   | - sample_patient_associated_disease.csv
|   | - sample_report.d8.txt
|   | - HLA.log
| - sample_HLA_SUPPLEMENTARY/
|   | - sample_Allele_Depth.txt
|   | - sample_results.txt
|   | - sample_Top_2_Results.txt
```

Sample_HLA_REPORT/folder:

The HLA_REPORT folder contains information to be shared in the genome reports, as well as the log file for the protocol. The sample_report.d8.txt file is expected output from the HLA-VBSeq HLA calling software, which estimates most likely HLA alleles to 8-digit resolution. For the reporting of HLA-disease risk association, either the sample_patient_allele_table.csv or sample_allele_table.csv can be used for report generation. The sample_allele_table.csv aggregates the alleles by gene, whereas sample_patient_allele_table.csv does not. Sample results from each of the expected outputs are shown below.

The sample_patient_allele_table.csv file reports on the HLA loci of interest, the top two most likely HLA alleles adapted from the sample_report.d8.txt file, a risk column that indicates whether the specific allele(s) are associated with known cases of disease risk as identified in an in-house developed HLA database, and HLA ID, which is an internal ID that links potential increased risk allele(s) cases to an identified case in the disease risk database. One or more HLA IDs could potentially be assigned to a given HLA locus depending on the number of disease associations identified. If multiple disease associations for a given allele are detected, the internal HLA IDs will be listed together in the HLA ID column. The sample_patient_allele_table.csv file

Table 1 Case 1 of sample_patient_allele_table.csv Output

Gene	Allele	Risk	HLA ID
<i>HLA-A</i>	<i>A*02:01:01:01</i>	None	None
<i>HLA-A</i>	<i>A*33:03:01:01</i>	None	None
<i>HLA-B</i>	<i>B*27:04:01</i>	Risk	HLAC20
<i>HLA-B</i>	<i>B*58:01:01:03</i>	None	None
<i>HLA-C</i>	<i>C*03:02:02:x1</i>	None	None
<i>HLA-C</i>	<i>C*04:01:01:14</i>	None	None
<i>HLA-DPB1</i>	<i>DPB1*02:01:02:29</i>	None	None
<i>HLA-DPB1</i>	<i>DPB1*04:01:34</i>	None	None
<i>HLA-DQA1</i>	<i>DQA1*01:02:01:04</i>	None	None
<i>HLA-DQA1</i>	<i>DQA1*04:01:01:02</i>	None	None
<i>HLA-DQB1</i>	<i>DQB1*04:02:01:04</i>	None	None
<i>HLA-DQB1</i>	<i>DQB1*06:09:01:01</i>	None	None
<i>HLA-DRB1</i>	<i>DRB1*07:01:01:02</i>	None	None
<i>HLA-DRB1</i>	<i>DRB1*08:02:01:01</i>	None	None

Table 2 Case 2 of sample_patient_allele_table.csv Output

Gene	Allele	Risk	HLA ID
<i>HLA-A</i>	<i>A*02:01:01:01</i>	None	None
<i>HLA-A</i>	<i>A*33:03:01:01</i>	None	None
<i>HLA-B</i>	<i>B*40:01:02:04</i>	None	None
<i>HLA-B</i>	<i>B*58:01:01:03</i>	None	None
<i>HLA-C</i>	<i>C*03:02:02:x1</i>	None	None
<i>HLA-C</i>	<i>C*04:01:01:14</i>	None	None
<i>HLA-DPB1</i>	<i>DPB1*02:01:02:29</i>	None	None
<i>HLA-DPB1</i>	<i>DPB1*04:01:34</i>	None	None
<i>HLA-DQA1</i>	<i>DQA1*05:01:01:01</i>	Risk	HLAC7 HLAC10
<i>HLA-DQA1</i>	<i>DQA1*05:01:01:02</i>	Risk	HLAC10
<i>HLA-DQB1</i>	<i>DQB1*02:01:01</i>	Risk	HLAC6 HLAC10
<i>HLA-DQB1</i>	<i>DQB1*02:01:02</i>	Risk	HLAC10
<i>HLA-DRB1</i>	<i>DRB1*07:01:01:02</i>	None	None
<i>HLA-DRB1</i>	<i>DRB1*08:02:01:01</i>	None	None

is primarily used to verify the detection of any internal HLA IDs linked to one or more specific alleles.

Report files

Table 1 shows an example of the sample_patient_allele_table.csv file in which a single locus, *HLA-B*, was detected as being associated with an increased risk to a specific disease. In this case, the internal ID HLAC20 links the particular *HLA-B*27:04* allele to a potential disease risk of Ankylosing Spondylitis (AS).

Table 2 shows an example of the sample_patient_allele_table.csv file, in which multiple loci in combination, in this case *HLA-DQA1* and *HLA-DQB1*, are detected as having alleles associated with an increased risk for a specific disease. In this case,

Table 3 Case 1 of sample_allele_table.csv Output

Gene	Allele 1	Allele 2	Risk
<i>HLA-A</i>	<i>A*02:01</i>	<i>A*33:03</i>	Average risk
<i>HLA-B</i>	<i>B*27:04</i>	<i>B*58:01</i>	Increased risk
<i>HLA-C</i>	<i>C*03:02</i>	<i>C*04:01</i>	Average risk
<i>HLA-DQA1</i>	<i>DQA1*01:02</i>	<i>DQA1*04:01</i>	Average risk
<i>HLA-DQB1</i>	<i>DQB1*04:02</i>	<i>DQB1*06:09</i>	Average risk
<i>HLA-DPB1</i>	<i>DPB1*02:01</i>	<i>DPB1*04:01</i>	Average risk
<i>HLA-DRB1</i>	<i>DRB1*07:01</i>	<i>DRB1*08:02</i>	Average risk

Table 4 Case 2 of sample_allele_table.csv Output

Gene	Allele 1	Allele 2	Risk
<i>HLA-A</i>	<i>A*02:01</i>	<i>A*33:03</i>	Average risk
<i>HLA-B</i>	<i>B*40:01</i>	<i>B*58:01</i>	Average risk
<i>HLA-C</i>	<i>C*03:02</i>	<i>C*04:01</i>	Average risk
<i>HLA-DQA1</i>	<i>DQA1*05:01</i>	<i>DQA1*05:01</i>	Increased risk
<i>HLA-DQB1</i>	<i>DQB1*02:01</i>	<i>DQB1*02:01</i>	Increased risk
<i>HLA-DPB1</i>	<i>DPB1*02:01</i>	<i>DPB1*04:01</i>	Average risk
<i>HLA-DRB1</i>	<i>DRB1*07:01</i>	<i>DRB1*08:02</i>	Average risk

the internal IDs HLAC6, HLAC7, and HLAC10 link the particular *DQA1*05:01* and *DQB1*02:01* alleles in combination to a potential disease risk for Celiac Disease (CD).

The sample_allele_table.csv file reports on the HLA loci of interest, the top two most likely HLA alleles adapted from the sample_report.d8.txt file, and a risk column that indicates whether the specific allele(s) are associated with known cases of disease risk from the internal HLA database. Note that the HLA alleles in this table only report to 4-digit resolution, for purposes of simplifying the reporting of allele status to participants.

Table 3 shows an example of the sample_allele_table.csv file in which a single locus, *HLA-B*, was detected as being associated with an increased risk to a specific disease (Ankylosing Spondylitis). Disease information on the report will be generated in a separate table.

Table 4 shows an example of the sample_allele_table.csv file in which multiple loci in combination, in this case *HLA-DQA1* and *HLA-DQB1*, are detected as having alleles associated with an increased risk to a specific disease (Celiac Disease).

The sample_patient_associated_disease.csv file reports on seven columns indicating the internal HLA ID assigned to one or more of the Increased Risk cases within the HLA-disease risk database, Associated Disease, HLA Locus, Allele, Haplotype or Genotype Assignment, Risk, “OR[95%CI] or RR, *p*-value, references”, and Interpretation. The sixth column indicates the HLA-disease association’s odds ratio at a 95% confidence interval, relative risk, and/or *p*-value as identified from the literature, including the literature reference that identified a disease association to the specific allele(s) detected. The interpretation column provides in-depth information regarding the potential disease risk, including a brief summary of the disease, the general population risk in Canada, and any known information about the identified risk alleles. The sample_patient_associated_disease.csv

will be blank if all seven HLA loci from the previous two tables report an Average Risk status and do not have an internal HLA ID that matches to a disease association case in the HLA database. Otherwise, if an “Increased Risk” is detected, the associated disease risk information from the HLA database will be populated in the `sample_patient_associated_disease.csv` table for all specific allele(s) identified. Multiple disease associations may be reported for a single HLA locus depending on genotype assignment.

Table 5 shows an example of the `sample_patient_associated_disease.csv` file in which a single allele at locus *HLA-B* was detected as being associated with an increased risk of a specific disease (Ankylosing Spondylitis). Any known Odds Ratio, Relative Risk, and associated literature references will be presented in the table, including disease information and any known information about the specific risk allele identified.

Table 6 shows an example of the `sample_patient_associated_disease.csv` file in which alleles in combination at multiple loci, *HLA-DQA1* and *HLA-DQB1*, were detected as being associated with an increased risk of a specific disease (Celiac Disease). Any known Odds Ratio, Relative Risk, and associated literature references will be presented in the table, including disease information and any known information about the specific risk allele identified.

Sample_HLA_Supplementary/folder:

The `HLA_SUPPLEMENTARY` folder contains supplementary information that is not shared in genome reports, including all outputs generated as part of the HLA-VBSeq algorithm. These files include a `sample_results.txt` file, which is a tab-delimited text file containing the following columns: HLA allele ID, genomic locus length, number of reads that the algorithm assigned to the HLA allele, normalized number of reads, and relative abundance. The `sample_Allele_Depth.txt` file is a tab-delimited text file with the following columns: HLA allele name, and an average depth of coverage. The `sample_Top_2_Results.txt` file is a tab-delimited text file that reports the top two average depth of coverage calls for 7 HLA loci of interest: *HLA-A*, *-B*, *-C*, *-DQA1*, *-DQB1*, *-DRB1*, *-DPB1*). These results are adapted from the `sample_Allele_Depth.txt` file.

Basic Protocol 2: Large-Scale Continental Ancestry Estimation

Folder structure

`sample_output_directory/`

```
| - sample_ANCESTRY_REPORT/
|   | - sample_HapMap3.png
|   | - sample_HapMap3.txt
| - sample_ANCESTRY_SUPPLEMENTARY/
|   | - sample_ADMIXTURE_ref_HAPMAP3.11.P
|   | - sample_ADMIXTURE_ref_HAPMAP3.11.Q
|   | - sample.pruned.fam
|   | - log/
|     | - all log files from PLINK
```

Sample_ANCESTRY_REPORT/folder:

The `ANCESTRY_REPORT` folder contains information to be included in the genome reports, similar to the `HLA_REPORT` folder. The `sample_HapMap3.png` file is a graphical visualization output from the in-house-developed R script, taking inputs from `ADMIXTURE` and `PLINK`. The `ADMIXTURE` software estimates large continental ancestry for a given sample, and the `.fam` file generated from `PLINK` is used to map

Table 5 Case 1 of sample_patient_associated_disease.csv Output

HLA ID	Associated disease	Locus	Allele/haplotype/ genotype assignment	Risk	OR[95%CI] or RR, <i>p</i> -value, references	Interpretation
HLAC20	Ankylosing Spondylitis (AS)	<i>B</i>	<i>B</i> *27:04 × <i>I</i>	Increased risk	OR = 1.76 (1.20-2.58), <i>p</i> = .004, study population = meta-analysis including Asian, African, American, and Europeans (Yang_2013_24261772); RR = 1.14 (1.01-1.29), <i>p</i> = .041, study population = meta-analysis (Lin_2017_28526894)	AS is a form of inflammatory arthritis that affects the spine and its surrounding joints and ligaments, causing chronic pain. AS usually manifests between 15 and 30 years of age. As many as 1 in 100 Canadians are estimated to live with AS (approximately 1%). The exact cause of AS is currently unknown but is thought to be caused by an interaction of genetic, immunological, and environmental factors. (Zhu_2019_31666997) (Arthritis Society, 2019 https://arthritis.ca/about-arthritis/arthritis-types-(a-z)/types/ankylosing-spondylitis). As many as 90% of people who develop AS carry an <i>HLA-B</i> *27 allele; however having the <i>HLA-B</i> *27 allele is not diagnostic for AS. A positive <i>HLA-B</i> *27 genetic test result can assist healthcare providers in confirming a diagnosis of AS in the presence of symptoms such as joint pain, stiffness and/or swelling of the spine, neck or chest, as well as inflammation of the urethra and eye (Arthritis Society, 2019; https://arthritis.ca/about-arthritis/arthritis-types-(a-z)/types/ankylosing-spondylitis)

Table 6 Case 2 of sample_patient_associated_disease.csv Output

HLA ID	Associated disease	Locus	Allele/haplotype/genotype assignment	Risk	OR[95% CI] or RR, p-value, references	Interpretation
HLA*06:01:01:01	Celiac Disease (CD)	DQA1	DQA1*02:01×1	Increased risk	OR = 20 (8.75-40.73), <i>p</i> < .001, study population = Slovenian (Schweiger_2016_27138053)	CD is an autoimmune disorder wherein the surface of the small intestine is damaged by gluten (a group of proteins present in wheat, rye, barley, spelt, kamut, and other various cereals), causing gastrointestinal issues like diarrhea, constipation, and bloating. The ingestion of gluten-containing foods is a known trigger of CD symptoms. CD is estimated to affect about 1 in 114 Canadians, and 0.5%-1% of the general population. CD is thought to be caused by a combination of genetic, immunological, and environmental factors (Caio_2019_31331324) (Health Canada, 2018; https://www.canada.ca/en/health-canada/services/food-nutrition/reports-publications/food-safety/ceeliac-disease-gluten-connection-1.html ; Jamnik et al., 2017; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5640059/). Genetic testing for the HLA-DQA1 and HLA-DQB1 alleles may be used to predict an individual's risk of developing CD. Individuals with HLA-DQA1 or HLA-DQB1 are estimated to have a 36%-53% risk of developing CD. These genes are also quite common in the general population and, therefore, are necessary but not sufficient for the development of CD (Martina_2018_30561391).

(Continued)

Table 6 Case 2 of sample_patient_associated_disease.csv Output, continued

HLA ID	Associated disease	Locus	Allele/haplotype/genotype assignment	Risk	OR[95%CI] or RR, <i>p</i> -value, references	Interpretation
HLA C7	Celiac Disease (CD)	DQA1	DQA1*05:01 × 1	Increased risk	OR = 9.43 (4.52-19.81), <i>p</i> < .001, study population = Slovenian (Schweiger_2016_27138053)	CD is an autoimmune disorder wherein the surface of the small intestine is damaged by gluten (a group of proteins present in wheat, rye, barley, spelt, kamut, and other various cereals), causing gastrointestinal issues like diarrhea, constipation, and bloating. The presence of gluten-containing foods is a known trigger of CD symptoms. CD is estimated to affect about 1 in 114 Canadians, and 0.5%-1% of the general population. CD is thought to be caused by a combination of genetic, immunological, and environmental factors (Caio_2019_31331324) (Health Canada, 2018 https://www.canada.ca/en/health-canada/services/food-safety/ceeliac-disease-gluten-connection-1.html ; Jamnik et al. 2017; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5640059/). Genetic testing for the HLA-DQA1 and HLA-DQB1 alleles may be used to predict an individual's risk of developing CD. Individuals with HLA-DQA1 or HLA-DQB1 are estimated to have a 36%-53% risk of developing CD. These genes are also quite common in the general population and, therefore, are necessary but not sufficient for the development of CD (Martina_2018_30561391).

(Continued)

Table 6 Case 2 of sample_patient_associated_disease.csv Output, continued

HLA ID	Associated disease	Locus	Allele/haplotype/genotype assignment	Risk	OR[95% CI] or RR, <i>p</i> -value, references	Interpretation
HLA C10	Celiac Disease (CD)	DQA1-DQB1	DQA1*05:01(or *05:02)×I-DQB1*02:02(or *02:01)×I/DDQA1*02×I-DQB1*02:02×I	Increased risk	OR = 19.68 (2.44-158.91), <i>p</i> < .001, study population = Slovenian (Schweiger_2016_27138053); OR = 15.95 (2.08-121.8), <i>p</i> = .0004, study population = Brazilian (Mixed) (Almeida_2016_28042478); Risk = 1 : 7, study population = Brazilian (Mixed) (Almeida_2016_28042478); OR = 8.32 (3.02-22.93), <i>p</i> < .01, study population = Syrian (Murad_2018_29793442); OR = 9.19 (2.34-20.14), <i>p</i> < 10 ⁻⁷ , study population = Spanish (Martínez-Ojinaga_2018_29699404)	CD is an autoimmune disorder wherein the surface of the small intestine is damaged by gluten (a group of proteins present in wheat, rye, barley, spelt, kamut, and other various cereals), causing gastrointestinal issues like diarrhea, constipation, and bloating. The ingestion of gluten-containing foods is a known trigger of CD symptoms. CD is estimated to affect about 1 in 114 Canadians, and 0.5%-1% of the general population. CD is thought to be caused by a combination of genetic, immunological, and environmental factors (Caio et al., 2019 31331324) (Health Canada, 2018; https://www.canada.ca/en/health-canada/services/food-safety/food-nutrition/reports-publications/food-safety/celiac-disease-gluten-connection-1.html ; Jamnik et al. 2017; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5640059/). Genetic testing for the HLA-DQA1 and HLA-DQB1 alleles may be used to predict an individual's risk of developing CD. Individuals with HLA-DQA1 or HLA-DQB1 are estimated to have a 36%-53% risk of developing CD. These genes are also quite common in the general population and, therefore, are necessary but not sufficient for the development of CD (Martina_2018_30561391).

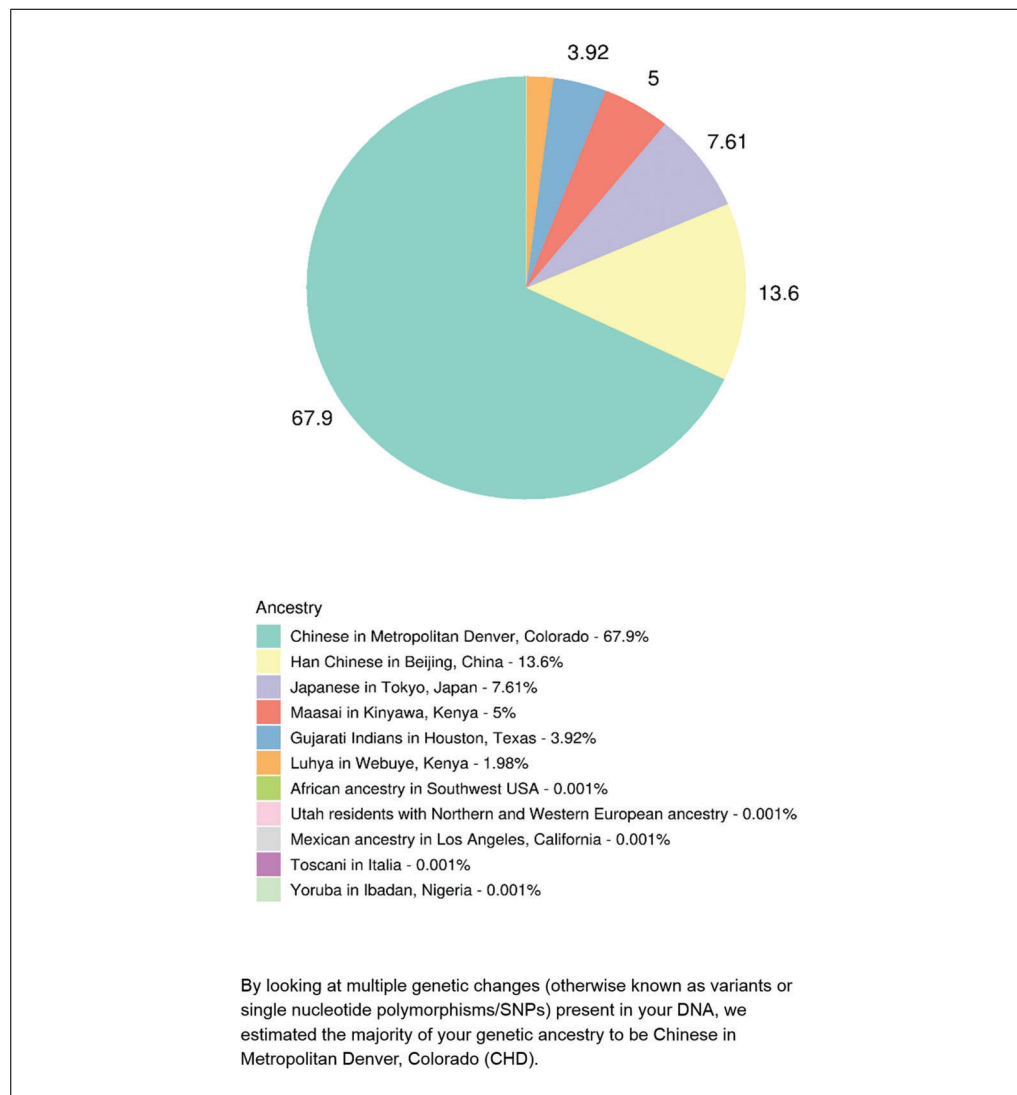


Figure 5 Combined pie chart and associated caption summarizing a given sample's estimated ancestry based on HapMap3 reference data. In the pie chart, estimated ancestries are represented by a percentage, with a threshold of equal to or greater than 3% admixed ancestry. Estimated ancestries of less than 3% are not depicted on the pie chart, but indicated instead in the chart legend.

each of the mapped references to their designated populations from the HapMap3 reference database (The International HapMap3 Consortium et al., 2010) HapMap3. The results are visualized using the R packages tidyverse, dplyr, RColorBrewer, and reshape2. For reporting, `sample_HapMap3.png` and `sample_HapMap3.txt` are used as the figure and caption, respectively.

Sample_ANCESTRY_SUPPLEMENTARY/ folder:

The ANCESTRY_SUPPLEMENTARY folder contains supplementary information not shared in genome reports, similar to the `sample_HLA_SUPPLEMENTARY` folder. The files included in the folder include the `.P` and `.Q` file outputs from the ADMIXTURE software and the `.fam` PLINK binary file generated from the final filtered merged dataset. The `.P` and `.Q` files are the allele frequencies of the inferred ancestral populations and the ancestry fractions respectively, while the `.fam` file is a tab-delimited text file containing the columns family ID, within-family ID, within-family ID of father, within-family ID of mother, sex code, and phenotype value, if available. In addition, the log folder is also located in this supplementary folder, containing all the log files

associated with each of the PLINK operations, as well as the pipeline log file. The `sample_HapMap3.png` shows a graphical representation of the participant's admixed ancestry as a pie chart (Fig. 5).

Basic Protocol 3: Dosage Recommendations for Pharmacogenomic Gene Variants Associated With Drug Response

Folder structure

`sample_output_directory/`

```
| - sample_PHARMACOGENOMICS_REPORT/  
|   | - sample_appendix.xlsx  
|   | - sample_report.xlsx  
|   | - Pharmacogenomics.log  
| - sample_PHARMACOGENOMICS_SUPPLEMENTARY/  
|   | - sample_final_stargazer_output.txt  
|   | - sample_rsID_Warfarin.txt  
|   | - sample_Target_GDF_RYR1.gdf (and associated files)
```

Sample_PHARMACOGENOMICS_REPORT/ folder:

The PHARMACOGENOMICS_REPORT folder contains curated pharmacogenomics information to be included in the genome reports, similarly to the HLA_REPORT and ANCESTRY_REPORT folders. This folder contains the `appendix.xlsx` and `report.xlsx` files, as well as the `Pharmacogenomics.log` pipeline log file. The `report.xlsx` file contains the three columns Gene/Genotype/rsID, Medication(s), and Metabolizer Phenotype. For the Gene/Genotype/rsID column, a gene, its predicted genotype, and associated rsIDs are listed. In the Medication(s) column, the list of medications possibly impacted by the variants present in the gene are provided, while the Metabolizer Phenotype column describes the potential metabolic activity of the gene variants associated with the listed medications for the given individual. The list of all possible pharmacogenes and associated recommendations can be found in File S3 in Supporting Information.

Report files

General metabolizer phenotype classification

Table 7 shows an example report of an individual who has intermediate metabolism for the genes *CYP2C19*, *CYP2C9*, an unknown metabolizer status for *CYP2D6* and *UGT1A1*, and poor metabolism for *CYP3A5*. The specific genotype of *CYP2C9* indicates a poor metabolizer status for the medication warfarin, including a specific variant rsID located in the *CYP2C* chromosome 10 region that is also known to be associated with poor warfarin metabolism (Whirl-Carrillo et al., 2021).

The `appendix.xlsx` file contains a more detailed version of `report.xlsx`, containing the six following columns: Gene/Genotype/rsID/PharmGKB ID, Medication(s), Metabolizer Phenotype, Implications, Recommendation, Strength of Recommendation. The first three columns in `appendix.xlsx` are identical to the columns represented in `report.xlsx`, except that the Gene/Genotype/rsID/PharmGKB ID column also includes the PharmGKB ID for the given clinical and dosage recommendations. Implications for the individual's metabolizer phenotype for the listed medications are detailed, and a dosage recommendation is also assigned. The strength of recommendation is also included as indicated from the PharmGKB/CPIC Database (Whirl-Carrillo et al., 2021).

Table 8 shows an example of the Appendix file for an individual who has an uncertain susceptibility for malignant hyperthermia based on their *CACNA1S* genotype, and has a *CFTR* genotype that is non-responsive to the medication Ivacaftor in cystic fibrosis patients. The implications of these gene-drug interactions, the dosage recommendations,

Table 7 Sample_report.xlsx

Gene genotype rsID	Medication(s)	Metabolizer phenotype
<i>CYP2C19</i> *1/*2 rs12769205	Amitriptyline Citalopram Escitalopram Clomipramine Clopidogrel Dexlansoprazole Doxepin Imipramine Lansoprazole Omeprazole Pantoprazole Sertraline Trimipramine Voriconazole	Intermediate metabolizer
<i>CYP2C9</i> *1/*2 rs1799853	Celecoxib Flurbiprofen Ibuprofen Lornoxicam Fosphenytoin Phenytoin Meloxicam Piroxicam Tenoxicam	Intermediate metabolizer
<i>CYP2D6</i> *3/*68+*4 rs35742686 rs3892097	Amitriptyline Atomoxetine Clomipramine Codeine Desipramine Doxepin Fluvoxamine Hydrocodone Imipramine Nortriptyline Ondansetron Paroxetine Tamoxifen Tramadol Trimipramine Tropisetron	Unknown
<i>CYP3A5</i> *3/*3 rs776746 rs776746	Tacrolimus	Poor metabolizer
<i>UGT1A1</i> *1/*60 rs4124874	Atazanavir	Unknown
<i>CYP2C9</i> *1/*2 rs1799853	Warfarin	Poor function
rs12777823 A/G	Warfarin	Poor function

Table 8 Sample_appendix.xlsx

GeneGenotypers IDPharmGKB ID	Medication(s)	Metabolizer phenotype	Implications	Recommendation	Strength of recommen- dation
<i>CACNA1S</i> Reference/ Reference PA166180457	Desflurane Enflurane Halothane Isoflurane Methoxyflurane Sevoflurane Succinylcholine	Uncertain Susceptibility	These results do not eliminate the chance that this patient is susceptible to Malignant Hyperthermia. The genetic cause of about half of all MH survivors, with MH susceptibility confirmed by contracture test, remains unknown	Clinical findings, family history, further genetic testing and other laboratory data should guide use of halogenated volatile anesthetics or depolarizing muscle relaxants	Strong
<i>CFTR</i> *1/*1 PA166114461	Ivacaftor	Ivacaftor non-responsive in CF patients	Ivacaftor treatment is recommended only in cystic fibrosis (CF) patients that are either homozygous or heterozygous for certain <i>CFTR</i> variants.	Ivacaftor not recommended	Moderate

and the strength of the recommendations are indicated according to current PharmGKB guidelines (Whirl-Carrillo et al., 2021). Additional rows for the Appendix are omitted from this table due to the length. The full example output is available as part of File S4 in Supporting Information.

The metabolizer phenotype structure of a typical gene-drug association is shown in Table 9 for *CYP2B6* and Efavirenz. Generally, a given gene variant will report either a Normal Metabolizer, Intermediate Metabolizer, Poor Metabolizer, Rapid Metabolizer, or Ultrarapid Metabolizer phenotype based on genotype. The assignment of metabolizer phenotypes is adapted from the PharmGKB diplotype-phenotype tables. Each of these possible phenotypes has specific implications, dosage recommendations, and a strength of recommendation for a given gene. This may include a recommendation for increasing or decreasing dosages, or an alternative treatment plan.

Metabolizer Phenotype Classification—Unique Cases

Some genes are associated with specific metabolizer statuses that differ significantly from the typical phenotype classifications, and are assigned based on specific dosage guidelines. In these unique cases, a metabolizer phenotype may be associated with a certain disease or condition susceptibility, such as in the case of *RYR1* and *CACNA1S* (Table 10). The assignment of the metabolizer phenotypes for *RYR1* and *CACNA1S* are based on the malignant hyperthermia causative variant list (Whirl-Carrillo et al., 2021). If a given individual is heterozygous for a malignant hyperthermia variant, they will be assigned the metabolizer phenotype of Malignant Hyperthermia Susceptible, or will otherwise be reported as Uncertain Susceptibility.

Table 9 Adapted Dosage Recommendation Table

Gene	Medication(s)	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2B6</i>	Efavirenz	Ultrarapid metabolizer	Slightly lower dose-adjusted trough concentrations of efavirenz compared with normal metabolizers	Initiate efavirenz with standard dosing (600 mg/day)	Strong	PA166182603
<i>CYP2B6</i>	Efavirenz	Rapid metabolizer	Slightly lower dose-adjusted trough concentrations of efavirenz compared with normal metabolizers	Initiate efavirenz with standard dosing (600 mg/day)	Strong	PA166182603
<i>CYP2B6</i>	Efavirenz	Normal metabolizer	Normal efavirenz metabolism	Initiate efavirenz with standard dosing (600 mg/day)	Strong	PA166182603
<i>CYP2B6</i>	Efavirenz	Intermediate metabolizer	Higher dose-adjusted trough concentrations of efavirenz compared with normal metabolizers; increased risk of CNS adverse events	Consider initiating efavirenz with decreased dose of 400 mg/day	Moderate	PA166182603
<i>CYP2B6</i>	Efavirenz	Poor metabolizer	Higher dose-adjusted trough concentrations of efavirenz compared with normal metabolizers; significantly increased risk of CNS adverse events and treatment discontinuation	Consider initiating efavirenz with decreased dose of 400 or 200 mg/day	Moderate	PA166182603

Table 10 *RYR1* and *CACNA1S* Metabolizer Phenotype Classifications

Gene	Medication(s)	Genotype	Variant	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>RYR1</i> <i>CACNA1S</i>	Desflurane	An individual heterozygous for an <i>RYR1</i> or <i>CACNA1S</i> malignant hyperthermia causative variant as designated by the EMHG	Causative variant	Malignant hyperthermia susceptible	Individuals are at increased risk of developing malignant hyperthermia if administered halogenated volatile anesthetics or the depolarizing muscle relaxant succinylcholine	Halogenated volatile anesthetics or depolarizing muscle relaxants succinylcholine are relatively contraindicated in persons with MHS. They should not be used, except in extraordinary circumstances where the benefits outweigh the risks. In general, alternative anesthetics are widely available and effective in patients with MHS	Strong	PA166180457
	Enflurane							
	Halothane							
	Isoflurane							
	Methoxyflurane							
	Sevoflurane							
	Succinylcholine							

(Continued)

Table 10 *RYR1* and *CACNA1S* Metabolizer Phenotype Classifications, *continued*

Gene	Medication(s)	Genotype	Variant	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>RYR1</i> <i>CACNA1S</i>	Desflurane Enflurane Halothane Isoflurane Methoxyflurane Sevoflurane Succinylcholine	An individual negative for a <i>RYR1</i> or <i>CACNA1S</i> malignant hyperthermia causative variant as designated by the EMHG	Non causative variant	Uncertain susceptibility	These results do not eliminate the chance that this patient is susceptible to Malignant Hyperthermia. The genetic cause of about half of all MH survivors, with MH susceptibility confirmed by contracture test, remains unknown	Clinical findings, family history, further genetic testing and other laboratory data should guide use of halogenated volatile anesthetics or depolarizing muscle relaxants	Strong	PA166180457

Similarly, in *CFTR*, the metabolizer phenotype may indicate non-responsiveness to Ivacaftor in Cystic Fibrosis (CF) patients (Table 11). The metabolizer phenotype for this gene is based on the *CFTR* variant list included in the Ivacaftor label in PharmGKB, as well as the presence of *G551D* and *F508del* variants. Other *CFTR* variants with therapeutic recommendations within the PharmGKB database are also taken into consideration for dosing guidelines. If an individual is heterozygous or homozygous for *G551D* or any variant included in the *CFTR* variant list, then they are considered Ivacaftor responsive in CF patients. However, if an individual is homozygous for the *F508del* variant, they are considered Ivacaftor non-responsive in CF patients (Whirl-Carrillo et al., 2021). Additionally, clinically actionable variants in *CFTR* are also included in the monogenic disease risk and carrier status sections of the genome report.

In the case of rasburicase and *G6PD*, the metabolizer phenotype is determined through consideration of variant classes as defined by WHO classification, and the sex of the patient (Table 12). Variants in class I-III are considered deficient, while class IV is defined as non-deficient. If a male patient is heterozygous or homozygous for a class IV variant, and a female patient is homozygous for a class IV variant, they are reported as having a Normal Metabolizer phenotype. In contrast, if a male patient is heterozygous or homozygous for a class I-III variant, and a female patient is homozygous for a class I-III variant, they are reported as a Deficient metabolizer of rasburicase. Finally, if a female patient has a genotype including one class IV variant and one class I-III variant, their metabolizer phenotype is reported as Variable. Due to X-linked mosaicism, female patients with this genotype may display either the deficient or non-deficient phenotype. However, an enzyme activity test can be used in place to assign *G6PD* phenotypes in these possible cases.

For warfarin dosing (Table 13), the genotypes of *CYP2C9*, *CYP4F2*, *VKORC1*, and the carrier status of rs12777823 contributed to generate formal recommendations for a given patient. For *CYP2C9*, alleles *2 and *3 were reported as Poor Function phenotypes, while alleles *5, *6, *8, and *11 were reported as Intermediate Metabolizers. If a patient has a *CYP2C9* genotype made up of an allele from the Poor Function classification and an allele from the Intermediate Metabolizer classification, both recommendations are included in the main body of the report. In *CYP4F2*, the rs2108622 T variant was reported as an Increased Metabolizer phenotype, while in *VKORC1*, the 1639G>A variant was reported as a Poor Function phenotype. Lastly, if a patient was a carrier of the rs12777823 A variant, they will be assigned a Poor Function phenotype, while a rs12777823 T variant is assigned as Unknown for warfarin.

For each of the four regions affecting warfarin, if a patient was not a carrier of the variants listed with formal recommendations from PharmGKB, they are reported to have a Normal Metabolizer phenotype. The Normal Metabolizer assignments for these four regions are performed in consideration of the CPIC guidelines for pharmacogenetics-guided warfarin dosing (Johnson et al., 2017). Additionally, the self-reported ancestry of the patient also has considerations for warfarin dosing, and is included in the strength of recommendations on the report (Whirl-Carrillo et al., 2021).

For the genes *CYP2C9*, *DPYD*, and *CYP2D6*, the activity score was also taken into consideration for a combination of certain medications and metabolizer phenotypes (Table 14). An activity score is defined as the sum of values assigned to each allele of a given gene, and typically ranges between 0 and 3.0, although some genes such as *CYP2D6* have activity scores that can exceed this range (Crews et al., 2014). For the case of *CYP2D6*, in which the copy number of variants can affect recommendations, a score of 0 indicates a poor metabolizer status, 0.25 to 1.0 indicates an intermediate metabolizer status, 1.25 to 2.25 indicates a normal metabolizer status, and >2.25 indicates an ultrarapid or

Table 11 *CFTR* Metabolizer Phenotype Classifications

Gene	Medication(s)	Genotype	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CFTR</i>	Ivacaftor	<i>G551D/X</i> <i>CFTR</i> variant/ <i>X</i>	Ivacaftor responsive in <i>CF</i> patients	Significant improvement in lung function, weight, risk of pulmonary exacerbation, patient reported outcomes, and reduction in sweat chloride concentrations through enhanced <i>CFTR</i> channel activity (increase probability of open channel)	Use ivacaftor according to the product label	Strong	PA166114461
<i>CFTR</i>	Ivacaftor	<i>F508del/F508del</i>	Ivacaftor non-responsive in <i>CF</i> patients	Ivacaftor treatment is recommended only in cystic fibrosis (<i>CF</i>) patients that are either homozygous or heterozygous for certain <i>CFTR</i> variants.	Ivacaftor not recommended	Moderate	PA166114461

Table 12 G6PD Metabolizer Phenotype Classifications

Gene	Medication(s)	Genotype	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
G6PD	Rasburicase	Male: class IV/X, Female: class IV/class IV	Normal metabolizer	Low or reduced risk of hemolytic anemia	No reason to withhold rasburicase based on G6PD status	Strong	PA166119846
G6PD	Rasburicase	Male: class I-III/X, Female: class I-III/class I-III	Deficient	At risk of acute hemolytic anemia	Rasburicase is contraindicated; alternatives include allopurinol	Strong	PA166119846
G6PD	Rasburicase	Female: class IV/class I-III	Variable	Unknown risk of hemolytic anemia	To ascertain that G6PD status is normal, enzyme activity must be measured; alternatives include allopurinol	Moderate	PA166119846

Table 13 CYP2C9, VKORC1, and CYP4F2 Phenotype Classifications for Warfarin

Gene	Medication(s)	Variant	Metabolizer phenotype	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2C9</i>	Warfarin	*2	Poor function	Based on <i>CYP2C9</i> *2, administer lower dosage through validated published pharmacogenomics algorithms.	Non-African Ancestry: Strong African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin	*3	Poor function	Based on <i>CYP2C9</i> *3, administer lower dosage through validated published pharmacogenomics algorithms.	Non-African Ancestry: Strong African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin	*5	Intermediate metabolizer	Based on <i>CYP2C9</i> 5*, calculate dosage from validated pharmacogenetic algorithms.	Non-African Ancestry: Optional African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin	*6	Intermediate metabolizer	Based on <i>CYP2C9</i> *6, calculate dosage from validated pharmacogenetic algorithms.	Non-African Ancestry: Optional African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin	*8	Intermediate metabolizer	Based on <i>CYP2C9</i> *8, calculate dosage from validated pharmacogenetic algorithms.	Non-African Ancestry: Optional African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin	*11	Intermediate metabolizer	Based on <i>CYP2C9</i> *11, calculate dosage from validated pharmacogenetic algorithms.	Non-African Ancestry: Optional African Ancestry: Moderate	PA166104949
<i>CYP2C9</i>	Warfarin		Normal metabolizer	Based on the variants present in <i>CYP2C9</i> , administer recommended dosage.	N/A	N/A

(Continued)

Table 13 CYP2C9, VKORC1, and CYP4F2 Phenotype Classifications for Warfarin, *continued*

Gene	Medication(s)	Variant	Metabolizer phenotype	Recommendation	Strength of recommendation	PharmGKB ID
rsID	Warfarin	rs12777823 A	Poor function	Based on rs12777823 carrier A, decrease calculated dosage by 10%-25%.	African Ancestry: Moderate	PA166104950
rsID	Warfarin		Normal metabolizer	Based on the absence of rs12777823, administer recommended dosage.	N/A	N/A
rsID	Warfarin	rs12777823 T	Unknown	Based on the rs12777823 carrier T, you have a genetic change that may alter how effective this medication is for you. However, at this time we are not entirely sure what this change means or how/if this alters standard dosing recommendations. Changes to current medication(s) or dosing are not recommended based on this finding.		
CYP4F2	Warfarin	rs2108622 T	Increased function	Based on rs2108622 carrier T in CYP4F2, increase dose by 5%-10%.	Non-African Ancestry: Optional	PA166104949
CYP4F2	Warfarin		Normal metabolizer	Based on rs2108622 in CYP4F2, administer recommended dosage.	N/A	N/A
VKORC1	Warfarin	1639G>A	Poor function	Based on VKORC1 1639G>A, administer lower dosage through validated published pharmacogenomics algorithms.	Non-African Ancestry: Strong African Ancestry: Moderate	PA166104949
VKORC1	Warfarin		Normal metabolizer	Based on the absence of VKORC1 1639G>A, administer recommended dosage.	N/A	N/A

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2C9</i>	Celecoxib Flurbiprofen Ibuprofen Lornoxicam	1.5	Intermediate metabolizer	Mildly reduced metabolism.	Initiate therapy with recommended starting dose. In accordance with the prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals.	Moderate	PA166191841
<i>CYP2C9</i>	Celecoxib Flurbiprofen Ibuprofen Lornoxicam	1	Intermediate metabolizer	Moderately reduced metabolism; higher plasma concentrations may increase probability of toxicities.	Initiate therapy with lowest recommended starting dose. Titrate dose upward to clinical effect or maximum recommended dose with caution. In accordance with the prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals. Carefully monitor adverse events such as blood pressure and kidney function during course of therapy.	Moderate	PA166191841
<i>CYP2C9</i>	Fosphenytoin Phenytoin	1.5	Intermediate metabolizer	Slightly reduced phenytoin metabolism; however, this does not appear to translate into increased side effects.	No adjustments needed from typical dosing strategies. Subsequent doses should be adjusted according to therapeutic drug monitoring, response, and side effects. An <i>HLA-B*15:02</i> negative test does not eliminate the risk of phenytoin-induced SJS/TEN, and patients should be carefully monitored according to standard practice.	Moderate	PA166122806

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, continued

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2C9</i>	Fosphenytoin Phenytoin	1	Intermediate metabolizer	Reduced phenytoin metabolism, higher plasma concentrations will increase probability of toxicities.	For first dose, use typical initial or loading dose. For subsequent doses, use approximately 25% less than typical maintenance dose. Subsequent doses should be adjusted according to therapeutic drug monitoring, response and side effects. An <i>HLA-B*15:02</i> negative test does not eliminate the risk of phenytoin-induced SJS/TEN, and patients should be carefully monitored according to standard practice.	Moderate	PA166122806
<i>CYP2C9</i>	Meloxicam	1.5	Intermediate metabolizer	Mildly reduced metabolism.	Initiate therapy with recommended starting dose. In accordance with the prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals.	Moderate	PA166192301

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, continued

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2C9</i>	Meloxicam	1	Intermediate metabolizer	Moderately reduced metabolism; higher plasma concentrations may increase probability of toxicities.	Initiate therapy with 50% of the lowest recommended starting dose. Titrate dose upward to clinical effect or 50% of the maximum recommended dose with caution. In accordance with the meloxicam prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals. Upward dose titration should not occur until after steady state is reached (at least 7 days). Carefully monitor adverse events such as blood pressure and kidney function during course of therapy. Alternatively, consider alternative therapy. Choose an alternative therapy not metabolized by <i>CYP2C9</i> or not significantly impacted by <i>CYP2C9</i> genetic variants <i>in vivo</i> or choose an NSAID metabolized by <i>CYP2C9</i> but with a shorter half-life.	Moderate	PA166192301
<i>CYP2C9</i>	Piroxicam	1.5	Intermediate metabolizer	Mildly reduced metabolism.	Initiate therapy with recommended starting dose. In accordance with the prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals.	Moderate	PA166192321

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, continued

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2C9</i>	Piroxicam	1	Intermediate metabolizer	Moderately reduced metabolism; higher plasma concentrations may increase probability of toxicities.	Choose an alternative therapy not metabolized by <i>CYP2C9</i> or not significantly impacted by <i>CYP2C9</i> genetic variants <i>in vivo</i> or choose an NSAID metabolized by <i>CYP2C9</i> but with a shorter half-life.	Moderate	PA166192341
<i>CYP2C9</i>	Tenoxicam	1.5	Intermediate metabolizer	Mildly reduced metabolism.	Initiate therapy with recommended starting dose. In accordance with the prescribing information, use the lowest effective dosage for shortest duration consistent with individual patient treatment goals.	Moderate	PA166192341
<i>CYP2C9</i>	Tenoxicam	1	Intermediate metabolizer	Moderately reduced metabolism; higher plasma concentrations may increase probability of toxicities.	Choose an alternative therapy not metabolized by <i>CYP2C9</i> or not significantly impacted by <i>CYP2C9</i> genetic variants <i>in vivo</i> or choose an NSAID metabolized by <i>CYP2C9</i> but with a shorter half-life.	Optional	PA166192341
<i>DPYD</i>	Capecitabine	1.5	Intermediate metabolizer	Decreased DPD activity (leukocyte DPD activity at 30% to 70% that of the normal population) and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Reduce starting dose by 50% followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). Patients with the c.[2846A>T];[2846A>T] genotype may require >50% reduction in starting dose.	Moderate	PA166109594

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, *continued*

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>DPYD</i>	Capecitabine	1	Intermediate metabolizer	Decreased DPD activity (leukocyte DPD activity at 30% to 70% that of the normal population) and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Reduce starting dose by 50% followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). Patients with the c.[2846A>T];[2846A>T] genotype may require >50% reduction in starting dose.	Strong	PA166109594
<i>DPYD</i>	Capecitabine	0.5	Poor metabolizer	Complete DPD deficiency and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Avoid use of 5- fluorouracil or 5-fluorouracil prodrug-based regimens. In the event, based on clinical advice, alternative agents are not considered a suitable therapeutic option, 5-fluorouracil should be administered at a strongly reduced dose with early therapeutic drug monitoring.	Strong	PA166109594
<i>DPYD</i>	Capecitabine	0	Poor metabolizer	Complete DPD deficiency and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Avoid use of 5-fluorouracil or 5-fluorouracil prodrug-based regimens.	Strong	PA166109594

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, *continued*

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>DPYD</i>	Fluorouracil	1.5	Intermediate metabolizer	Decreased DPD activity (leukocyte DPD activity at 30% to 70% that of the normal population) and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Reduce starting dose by 50% followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). Patients with the c.[2846A>T];[2846A>T] genotype may require >50% reduction in starting dose.	Moderate	PA166122686
<i>DPYD</i>	Fluorouracil	1	Intermediate metabolizer	Decreased DPD activity (leukocyte DPD activity at 30% to 70% that of the normal population) and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Reduce starting dose by 50% followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). Patients with the c.[2846A>T];[2846A>T] genotype may require >50% reduction in starting dose.	Strong	PA166122686
<i>DPYD</i>	Fluorouracil	0.5	Poor Metabolizer	Complete DPD deficiency and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Avoid use of 5- fluorouracil or 5-fluorouracil prodrug-based regimens. In the event, based on clinical advice, alternative agents are not considered a suitable therapeutic option, 5-fluorouracil should be administered at a strongly reduced dose with early therapeutic drug monitoring.	Strong	PA166122686

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, *continued*

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>DPYD</i>	Fluorouracil	0	Poor metabolizer	Complete DPD deficiency and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs.	Avoid use of 5-fluorouracil or 5-fluorouracil prodrug-based regimens.	Strong	PA166122686
<i>CYP2D6</i>	Atomoxetine	1	Intermediate metabolizer	Possibly higher atomoxetine concentrations as compared to normal metabolizers but questionable clinical significance. Normal metabolizers may be at an increased risk of increased discontinuation as compared to poor metabolizers.	Initiate with a dose of 0.5 mg/kg and increase to 1.2 mg/kg/day after 3 days If no clinical response and in the absence of adverse events after 2 weeks, consider obtaining a peak plasma concentration (1 to 2 hr after dose administered). If <200 ng/ml, consider a proportional increase in dose to approach 400 ng/ml.	Moderate	PA166181885
<i>CYP2D6</i>	Atomoxetine	1	Intermediate metabolizer	Decreased metabolism of atomoxetine and higher atomoxetine concentrations as compared to normal metabolizers. Individuals with activity score of 1.0 with <i>CYP2D6</i> *10 may be at an increased risk of increased discontinuation as compared to poor metabolizers.	Initiate with a dose of 0.5 mg/kg/day and if no clinical response and in the absence of adverse events after 2 weeks, consider obtaining a plasma concentration 2-4 hr after dosing. If response is inadequate and concentration is <200 ng/ml, consider a proportional dose increase to achieve a concentration to approach 400 ng/ml. If unacceptable side effects are present at any time, consider a reduction in dose.	Moderate	PA166181885

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, continued

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2D6</i>	Atomoxetine	0.5	Intermediate metabolizer	Decreased metabolism of atomoxetine and higher atomoxetine concentrations as compared to normal metabolizers. Intermediate metabolizers may be at an increased risk of discontinuation as compared to poor metabolizers.	Initiate with a dose of 0.5 mg/kg/day and if no clinical response and in the absence of adverse events after 2 weeks, consider obtaining a plasma concentration 2-4 hr after dosing. If response is inadequate and concentration is <200 ng/ml, consider a proportional dose increase to achieve a concentration to approach 400 ng/ml. If unacceptable side effects are present at any time, consider a reduction in dose.	Moderate	PA166181885
<i>CYP2D6</i>	Tamoxifen	1	Intermediate metabolizer	Lower endoxifen concentrations compared to normal metabolizers; higher risk of breast cancer recurrence, event-free and recurrence-free survival compared to normal metabolizers.	Consider hormonal therapy such as an aromatase inhibitor for postmenopausal women or aromatase inhibitor along with ovarian function suppression in premenopausal women, given that these approaches are superior to tamoxifen regardless of <i>CYP2D6</i> genotype. If aromatase inhibitor use is contraindicated, consideration should be given to use a higher but FDA approved tamoxifen dose (40 mg/day). Avoid <i>CYP2D6</i> strong to weak inhibitors.	Optional	PA166176068

(Continued)

Table 14 Pharmacogenomic Activity Scores for *CYP2C9*, *DPYD*, *CYP2D6*, *continued*

Gene	Medication(s)	Activity score	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>CYP2D6</i>	Tamoxifen	1	Intermediate metabolizer	Lower endoxifen concentrations compared to normal metabolizers; higher risk of breast cancer recurrence, event-free and recurrence-free survival compared to normal metabolizers.	Consider hormonal therapy such as an aromatase inhibitor for postmenopausal women or aromatase inhibitor along with ovarian function suppression in premenopausal women, given that these approaches are superior to tamoxifen regardless of <i>CYP2D6</i> genotype. If aromatase inhibitor use is contraindicated, consideration should be given to use a higher but FDA approved tamoxifen dose (40 mg/day). Avoid <i>CYP2D6</i> strong to weak inhibitors.	Moderate	PA166176068
<i>CYP2D6</i>	Tamoxifen	0.5	Intermediate metabolizer	Lower endoxifen concentrations compared to normal metabolizers; higher risk of breast cancer recurrence, event-free and recurrence-free survival compared to normal metabolizers.	Consider hormonal therapy such as an aromatase inhibitor for postmenopausal women or aromatase inhibitor along with ovarian function suppression in premenopausal women, given that these approaches are superior to tamoxifen regardless of <i>CYP2D6</i> genotype. If aromatase inhibitor use is contraindicated, consideration should be given to use a higher but FDA approved tamoxifen dose (40 mg/day). Avoid <i>CYP2D6</i> strong to weak inhibitors.	Moderate	PA166176068

Table 15 HLA-A, HLA-B Metabolizer Phenotype Table

Gene	Medication(s)	Variant	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
HLA-A	Carbamazepine		Normal or reduced risk	Normal or reduced risk of carbamazepine-induced SJS/TEN, DRESS and MPE	Use carbamazepine per standard dosing guidelines.	Strong	PA166105008
HLA-A	Carbamazepine	*31:01	Increased risk	Increased risk of carbamazepine-induced SJS/TEN, DRESS and MPE	If patient is carbamazepine-naïve and alternative agents are available, do not use carbamazepine.	Strong	PA166105008
HLA-B	Abacavir		Normal or reduced risk	Low or reduced risk of abacavir hypersensitivity	Use abacavir per standard dosing guidelines.	Strong	PA166104997
HLA-B	Abacavir	*57:01	Increased risk	Significantly increased risk of abacavir hypersensitivity	Abacavir not recommended.	Strong	PA166104997
HLA-B	Allopurinol		Normal or reduced risk	Low or reduced risk of allopurinol SCAR	Use allopurinol per standard dosing guidelines.	Strong	PA166105003
HLA-B	Allopurinol	*58:01	Increased Risk	Significantly increased risk of allopurinol SCAR	Allopurinol is contraindicated.	Strong	PA166105003
HLA-B	Carbamazepine		Normal or Reduced Risk	Normal or reduced risk of carbamazepine-induced SJS/TEN, DRESS and MPE.	Use carbamazepine per standard dosing guidelines.	Strong	PA166105008
HLA-B	Carbamazepine	*15:02	Increased risk	Increased risk of carbamazepine-induced SJS/TEN	If patient is carbamazepine-naïve, do not use carbamazepine.	Strong	PA166105008

(Continued)

Table 15 HLA-A, HLA-B Metabolizer Phenotype Table, *continued*

Gene	Medication(s)	Variant	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
HLA-B	Fosphenytoin Phenytoin		Normal or reduced risk	Normal phenytoin metabolism	No adjustments needed from typical dosing strategies. Subsequent doses should be adjusted according to therapeutic drug monitoring, response, and side effects. An <i>HLA-B</i> *15:02 negative test does not eliminate the risk of phenytoin-induced SJS/TEN, and patients should be carefully monitored according to standard practice.	Strong	PA166122806
HLA-B	Fosphenytoin Phenytoin	*15:02	Increased risk	Increased risk of phenytoin-induced SJS/TEN	If patient is phenytoin-naïve, do not use phenytoin/fosphenytoin. Avoid carbamazepine and oxcarbazepine. Optional recommendation: If the patient has previously used phenytoin continuously for longer than three months without incidence of cutaneous adverse reactions, cautiously consider use of phenytoin in the future. The latency period for drug-induced SJS/TEN is short with continuous dosing and adherence to therapy (4–28 days), and cases usually occur within three months of dosing.	Strong	PA166122806
HLA-B	Oxcarbazepine		Normal or reduced risk	Normal or reduced risk of oxcarbazepine-induced SJS/TEN	Use oxcarbazepine per standard dosing guidelines.	Strong	PA166176623
HLA-B	Oxcarbazepine	*15:02	Increased risk	Increased risk of oxcarbazepine-induced SJS/TEN	If patient is oxcarbazepine-naïve, do not use oxcarbazepine.	Strong	PA166176623

Table 16 *IFNL3* Metabolizer Phenotype Table

Gene	Medication(s)	Variant	Metabolizer phenotype	Implications	Recommendation	Strength of recommendation	PharmGKB ID
<i>IFNL3</i>	Peginterferon alfa-2a	rs12979860 CC	Favorable response	For PEG-IFN ALPHA and RBV based on the variation rs12979860 Approximately 70% chance for SVR after 48 weeks of treatment. Consider implications before initiating PEG-IFN alpha and RBV containing regimens For protease inhibitor combinations with PEG-IFN ALPHA and RBV based on the variation rs12979860 Approximately 90% chance for SVR after 24-48 weeks of treatment. Approximately 80%-90% of patients are eligible for shortened therapy (24-28 weeks vs. 48 weeks). Weighs in favor of using PEG-IFN alpha and RBV containing regimen		Strong	PA166110235
	Peginterferon alfa-2b Ribavirin						
<i>IFNL3</i>	Peginterferon alfa-2a Peginterferon alfa-2b Ribavirin	rs12979860 CT/TT	Unfavorable Response	For PEG-IFN ALPHA and RBV based on the variation rs12979860 Approximately 30% chance for SVR after 48 weeks of treatment. Consider implications before initiating PEG-IFN alpha and RBV containing regimens For protease inhibitor combinations with PEG-IFN ALPHA and RBV based on the variation rs12979860 Approximately 60% chance for SVR after 24-48 weeks of treatment. Approximately 50% of patients are eligible for shortened therapy (24-28 weeks). Consider implications before initiating PEG-IFN and RBV containing regimens. For PEG-IFN ALPHA and RBV: Approximately 70% chance for SVR c after 48 weeks of treatment. Consider implications before initiating PEG-IFN alpha and RBV containing regimens.		Strong	PA166110235

extensive metabolizer status. In the cases of *CYP2C9* and *DPYD*, 0 to 0.5 indicates a poor metabolizer status, 1.0 to 1.5 indicates an intermediate metabolizer status, and 2.0 indicates a normal metabolizer status (Whirl-Carrillo et al., 2021). The activity scores for these genes are appended from the PharmGKB diplotype-phenotype tables and used to assign formal recommendations that may differ slightly between score assignments within the same metabolizer phenotype classification. For example, in Table 14, the implications and recommendations for *CYP2C9* genotypes at an activity score of 1 and 1.5 vary, despite being both assigned as intermediate metabolizers.

For *HLA-A* and *HLA-B* genotypes determined using Basic Protocol 1, the listed variants for each gene are also used to determine their metabolizer phenotypes (Table 15). For *HLA-B*, a metabolizer phenotype of Increased Risk is assigned with the presence of the *57:01 variant for Abacavir, the presence of *58:01 for Allopurinol, and the presence of the *15:01 variant for Carbamazepine, Fosphenytoin, Phenytoin, and Oxcarbazepine. For *HLA-A*, a metabolizer phenotype of Increased Risk is assigned with the presence of *31:01 for Carbamazepine. Otherwise, in both *HLA-A* and *HLA-B*, they are assigned the metabolizer phenotype of Normal or Reduced Risk.

For *IFNL3*, the genotype at rs12979860 is used to determine the metabolizer phenotype, with the genotype CC indicating a Favorable Response, while CT or TT indicate an Unfavorable Response (Table 16). Otherwise, a metabolizer phenotype of Unknown is assigned for *IFNL3*.

Sample_PHARMACOGENOMICS_SUPPLEMENTARY/ folder:

The PHARMACOGNEOMICS_SUPPLEMENTARY folder contains supplementary data and the intermediary files generated in the Pharmacogenomics pipeline. The final_stargazer_output.txt file contains concatenated results from the Stargazer software for all the pharmacogenes analyzed in this pipeline. The rsID_Warfarin.txt file contains the results from the bcftools view command for rsid12777823 to determine an individual's warfarin metabolizer status. Lastly, the Target_GDF_RYR1.gdf file is also included in this folder, for the purposes of analyzing the possible presence of structural variants in *CYP2D6* with the Stargazer software. The .gdf file is generated using the GATK software, and analyzed using the control gene *RYR1* as recommended from Lee, Wheeler, Thummel, and Nickerson (2019). The *VDR* gene .bed file is also included in the Pharmacogenomics Configuration folder to be used as an alternative control gene, and was also outlined in (Lee et al., 2019).

COMMENTARY

Background Information

Basic Protocol 1: HLA genotyping

HLA stands for "Human Leukocyte Antigen," and is associated with the body's regulation and the function of immune response. An individual's HLA type is grouped into a different class depending on the genes that are present. For example, class I consists of the *HLA-A*, *HLA-B*, and *HLA-C* genes, and class II consists of the *HLA-DRB1* and *HLA-DQB1* genes (Matzaraki, Kumar, Wijmenga, & Zhernakova, 2017). Studies have shown that certain genetic changes within the HLA region may be associated with susceptibility to and severity of SARS-CoV-2 infection. For example, studies have suggested that having a certain genetic variant in class I *HLA-B*, *HLA-B*46:01* may result in an individual being more susceptible to a SARS-CoV-2

infection. Conversely, having a different genetic variant in class I *HLA-B*, *HLA-B*15:03* may result in having greater protection against a SARS-CoV-2 infection (Anastassopoulou, Gkizarioti, Patrinos, & Tsakris, 2020). Certain genetic variants in the HLA region have previously been found to be associated with an increased risk for autoimmune conditions such as type 1 diabetes, ankylosing spondylitis, and rheumatoid arthritis (Gao et al., 2019). Therefore, analysis of HLA status may provide valuable information on potential disease risk in addition to providing novel information on COVID-19 immune response. As we acquire more information about the association between HLA, autoimmune conditions, and COVID-19, knowing HLA status from a genome report may help inform response to SARS-CoV-2, vaccine response, early

candidates for vaccine development, treatment, and clinical decision-making tools for individuals with a past infection.

The HLA typing pipeline was developed to estimate the most probable HLA genotypes for a given individual and indicate potential disease risks associated with certain HLA alleles. The software HLA-VBSeq v2 (Mimori et al., 2019) was used to estimate the most probable HLA types from whole-genome-sequencing patient data. Read names were extracted from BAM files for 22 HLA loci, including *HLA-A*, *-B*, *-C*, *-DQA1*, *-DQB1*, *-DPB1*, and *-DRB1*. The extracted reads were then aligned to the HLA v2 database from the HLA-VBSeq package, which is based on the IMGT/HLA database Release 3.31.0 and the Japanese HLA reference dataset (Mimori et al., 2019). The HLA types were then estimated for paired-end read data using the HLA-VBSeq package along with the average depth of coverage for each HLA allele. The top called allele for each HLA loci was reported to four-digit resolution. The top called alleles for the seven major HLA loci were then compared to an in-house database containing currently known HLA-disease associations as reported from the literature. An increased disease risk was reported if a patient's HLA genotype or genotypes were found to be associated with a relevant disease or condition. Alleles were only reported if they met HLA-VBSeq's default coverage threshold for HLA typing.

HLA alleles associated with increased autoimmune disease risk were identified through a review of literature available on PubMed (<https://pubmed.ncbi.nlm.nih.gov>). Papers reviewing the association between the HLA locus and human autoimmune disease in general were identified using the following search terms: "HLA" OR "MHC" AND "human disease" OR "autoimmune disease". A list of autoimmune diseases consistently reported to be associated with HLA was compiled as follows: Type 1 Diabetes (T1D), Celiac Disease (CD), Rheumatoid Arthritis (RA), Ankylosing Spondylitis (AS), Behçet's Disease (BD), Multiple Sclerosis (MS), and Graves' Disease (GD). Subsequent searches were conducted on PubMed for review articles highlighting the role of HLA in each of the above diseases. The following search terms were used: "Type 1 Diabetes" OR "T1D" OR "Insulin Dependent Diabetes Mellitus" OR "IDDM" OR "Celiac Disease" OR "Coeliac Disease" OR "Rheumatoid Arthritis" OR "Ankylosing Spondylitis" OR "Behçet's Disease" OR "Be-

hçet's Syndrome" OR "Multiple Sclerosis" OR "Graves' Disease" AND "HLA" OR "MHC" AND "association" OR "risk". From this search, a list of HLA alleles consistently reported to be associated with each disease was compiled, followed by an evaluation of the significance of disease risk. Subsequent searches were conducted to identify the disease risks associated with each HLA allele. The following search terms were used. For T1D: "Type 1 Diabetes", "T1D", "IDDM", "Insulin-Dependent Diabetes Mellitus", "HLA", "MHC", "DRB1", "DQA1", "DQB1", "DR", "class II", "allele", "HLA genotype", "risk", "disease risk", "genetic risk", "case-control". For CD: "Celiac Disease", "Coeliac Disease", "HLA", "MHC", "HLA genotype", "disease risk", "risk", "genetic risk", "HLA-DQ", "case-control". For RA: "Rheumatoid Arthritis", "HLA", "MHC", "association", "risk", "genetic risk", "HLA-DRB1", "shared epitope", "susceptibility", "case-control". For AS: "ankylosing spondylitis", "HLA", "HLA-B27", "susceptibility", "association", "case-control". For BD: "Behçet's disease", "Behçet Disease", "HLA", "MHC", "HLA-B*51", "HLA-B51", "association", "class I", "case-control". For MS: "Multiple Sclerosis", "HLA", "genetics", "association", "susceptibility", "HLA-DRB1*15", "HLA-DR15". For GD: "Graves' Disease", "HLA", "class II", "DQA1*0501", "case-control", "association". Papers were included if they were a case-control study, or a meta-analysis of a case-control study, and were published after 1990. Disease risks were included if they were published in a case-control study or a meta-analysis of case-control studies, were reported as an odds ratio (OR), or relative risk (RR), had an OR RR > 1, and were found to be statistically significant by the published study. The full list of references used for the development of the in-house HLA disease association database can be found in File S6 in Supporting Information.

Basic Protocol 2: Ancestry estimation

A person's race typically refers to their physical appearance and characteristics, such as their skin and eye color. A person's ethnicity refers to communality in cultural heritage, language, social practice, traditions, and geopolitical factors. Genetic ancestry is often inferred using ancestry informative markers (AIMs) based on genetic/genomic data, and can thus be considered a biological variable. Genetic ancestry can sometimes disagree with self-reported race and/or ethnicity (Mersha &

Abebe, 2015). Identifying genetic ancestry, currently used primarily in the recreational genetic testing context, may be important for understanding which populations are disproportionately affected by COVID-19, as well as if there are any significant differences in COVID-19 susceptibility and response across individuals of various genetic ancestries. To illustrate current understanding of COVID-19 disease outcomes among various populations, the American Center for Disease Control (CDC) shared statistics in March 2021 that highlight COVID-19 infection rates in the U.S.A. as categorized by race and ethnicity. The CDC identified that American Indian, Non-Hispanic persons had a $1.6\times$ increase in risk of infection compared to White, Non-Hispanic persons, a $3.5\times$ increase in hospitalization, and a $2.4\times$ increase in death. Similarly for Asian, Black, and Hispanic persons, a $0.7\times$, $1.1\times$, and $2.0\times$ increase in risk of infection were reported, respectively, a $1.0\times$, $2.8\times$, and $3.0\times$ increased risk of hospitalization, and a $1.0\times$, $1.9\times$, and a $2.3\times$ risk of death compared to White, Non-Hispanic persons (COVID-NET, accessed April 23rd, 2021; see Internet Resources). These statistics note the highly elevated rates in cases, hospitalizations, and death across various populations compared to White, Non-Hispanic Americans, particularly in American-Indigenous, Black, and Hispanic persons. Alongside these findings, other studies conducted across the U.S.A. and U.K. have also investigated COVID-19 risk among diverse populations, and have concluded that patients of Black, Asian, and Hispanic ancestry were more likely to be infected with COVID-19, as well as being at a higher risk both for admission to intensive care and for death (Sze et al., 2020). In addition, a genetic association study identified a particular gene cluster on chromosome 3 (45,859,651-45,909,024, hg19) from Neanderthals that may indicate risk for respiratory failure after infection with COVID-19, as well as act as a risk factor for other severe symptoms and hospitalization. It is carried by $\sim 50\%$ of people in South Asia and $\sim 16\%$ of people in Europe, suggesting another potential link between COVID-19 risk and genetic ancestry in some infected individuals (Zebberg & Pääbo, 2020). Although there are currently major limitations when using ancestry-based algorithms to predict genetic ancestry, we sought to incorporate this information into a comprehensive report for the purpose of evaluating the utility of genomic results as part of the GENCOV study (Taher et al., 2021).

Large-scale continental individual ancestry was estimated through the combination of patient genotypes with those from the reference dataset, HapMap3 (The International HapMap 3 Consortium et al., 2010). The International HapMap 3 Consortium have genotyped 1.6 million common single nucleotide polymorphisms (SNPs) in 1184 reference individuals from 11 global populations (The International HapMap 3 Consortium et al., 2010, see Table 17). Various filtering strategies were applied to both the patient genotypes and the reference dataset, removing duplicate SNPs and correcting for chromosome and position mismatches using bash scripting and PLINK software. Furthermore, the genetic variants used for estimation were pruned for variants in linkage disequilibrium (LD) with an $r^2 > 0.2$ in a 50 kb window. Variants were also excluded if they fell within genomic ranges of known high-LD structure following the PlinkQC protocol for ancestry estimation using the HapMap3 reference dataset (Meyer, 2021a, Meyer, 2021b). Ancestry estimation can alternatively be performed using reference data available through The 1000 Genomes Project initiative to increase the accuracy of estimation reporting and for validation purposes (1000 Genomes Project Consortium et al., 2015, Meyer, 2021b).

ADMIXTURE was used to output estimated patient ancestry against the reference dataset following filtration. ADMIXTURE is a model-based ancestry estimation software, modeling the probability of the observed genotypes using ancestry proportions and population allele frequencies (Alexander et al., 2009). In-house scripts were then used to generate a visual summary of patient admixed ancestry.

Basic Protocol 3: Pharmacogenomics

Pharmacogenomics is the study of relationships between genetic variation and its influence on how bodies respond to medications. More often than not, prescribed medications are assumed to be equally effective in all patients, regardless of factors such as age, body mass index, sex, or diet, provided that they are given appropriately adjusted doses. However, genetic variation among patients may lead to varying drug responses even when accounting for most of these traits. Genetic variation can also alter drug metabolism by affecting how much of a given medication is absorbed into the body and how efficiently it is able to travel to its intended destination, as well as how quickly it gets broken down.

Table 17 HapMap3 Reference Dataset

Population	Population abbreviation	Number of samples
African ancestry in Southwest USA	ASW	83
Utah residents with Northern and Western European ancestry from the CEPH collection	CEU	165
Han Chinese in Beijing, China	CHB	84
Chinese in Metropolitan Denver, Colorado	CHD	85
Gujarati Indians in Houston, Texas	GIH	88
Japanese in Tokyo, Japan	JPT	86
Luhya in Webuye, Kenya	LWK	90
Maasai in Kinyawa, Kenya	MKK	171
Mexican ancestry in Los Angeles, California	MXL	77
Toscani in Italia	TSI	88
Yoruba in Ibadan, Nigeria	YRI	167
TOTAL		1184

If a given individual's genetic variation alters any of these processes, certain medications may not work as well or cause side effects. Pharmacogenomic testing provides valuable insight into how individuals with particular genetic variants might respond differently to medication(s). Therefore, identifying gene variants with relevant pharmacogenomic variants can allow healthcare providers to give more personalized recommendations related to drug dosage and/or treatment regimens.

The pharmacogenomics reports were based on the recommendations from The Pharmacogenomics Knowledge Base (Whirl-Carrillo et al., 2021, accessed as of October 2021). PharmGKB is a resource containing clinically actionable gene-drug associations and genotype-phenotype relationships (Thorn, Klein, & Altman, 2013). From PharmGKB, dosing guidelines were taken from the Clinical Pharmacogenetics Implementation Consortium (CPIC) for CPIC Level A and PharmGKB Level of Evidence 1A. Stargazer is a software application that detects single nucleotide variants, insertion-deletion variants, and structural variants, including gene deletions, duplications, and conversions, for identifying genotypes in pharmacogenes (Lee et al., 2019). Using Stargazer, the genotypes of 17 pharmacogenes were identified, including structural variant analysis in *CYP2D6*. Additionally, the genotypes of two HLA loci, *HLA-A* and *HLA-B*, and rs12777823 were also

included. The resulting genotypes were then linked to a metabolizer phenotype reference database compiled in-house. Pharmacogenes that were unable to be genotyped for certain participants through the Stargazer software were included as part of the report Appendix with haplotype candidates listed instead. The metabolizer phenotype database was generated from the gene diplotype-phenotype tables and the gene-medication dosing guidelines from PharmGKB and CPIC, accessed as of December 2021. If a genotype was not defined with CPIC Level A and PharmGKB Level of Evidence 1A, a metabolizer phenotype of Unknown and a standardized recommendation were assigned to the given gene medication.

Additional protocols

In addition to the basic protocols outlined above, comprehensive gene panels were developed to aid in variant filtration and assessment. A list of medically actionable gene panels was developed according to specifications of medically actionable genes as reported by the American College of Medical Genetics and Genomics (Miller et al., 2021), The Clinical Genome Resource (ClinGen, Rehm et al., 2015), and Reble et al. (2021) for a comprehensive panel including 208 genes. Additional panels containing genes associated with carrier status and Mendelian, rare genetic diseases were included under the following categories:

audiology and ophthalmology, 1387 genes; biochemical and mitochondrial, 944 genes; blood and immunology, 898 genes; cardiovascular, 1000 genes; carrier screening, 920; dermatological, 582 genes; endocrine, 460 genes; gastrointestinal, 612 genes; musculoskeletal, 1619 genes; neurological, 2639 genes; oncology, 735 genes; renal, 860 genes. These comprehensive disease gene panels were compiled using ClinGen, Clinical Genomics Database (CGD, Solomon, Nguyen, Bear, & Wolfsberg, 2013), Online Mendelian Inheritance in Man (OMIM, Hamosh, Scott, Amberger, Valle, & McKusick, 2000), BabySeq (Holm et al., 2018), and The Gene Curation Coalition (GenCC) (DiStefano et al., 2022) gene databases. Genes from these databases were cross-referenced with panels from reputable genetic testing companies. Additional genes associated with COVID-19 were identified through a literature review. Periodic updates of this database are needed to incorporate new evidence for gene-disease relationships as they are identified. The comprehensive gene panel database can be found in File S5 in Supporting Information.

Following the development of gene panels, participant data are processed and analyzed using the Franklin Genoox Platform (<https://www.genoox.com>) in combination with GATK best practices for variant calling. Reads are aligned using BWA against the GRCh37 human reference, with duplicate reads removed. Variant calling is performed using GATK (version 4.1) and FreeBayes (version 1.1.0). Base calling is performed using bcl2fastq 2.20 (for HiSeq 2500) or HiSeq Analysis Software (for HiSeq X). Reads are mapped to the b37 reference sequence using the bwa-mem algorithm, and duplicate reads are marked using Picard Tools. Local realignment and base quality score recalibration are also performed using GATK. Variants are called using GATK HaplotypeCaller, followed by variant quality score recalibration (VQSR) for filtration. Variants are identified for further analysis using the following filtering parameters: (1) variants classified as pathogenic or likely pathogenic in ClinVar (Landrum et al., 2018); (2) variants with a Genoox classification of pathogenic/likely pathogenic/leaning pathogenic, aggregated allele frequency $\leq 5\%$, \pm identified in ClinVar, with an established gene-disease relationship; (3) variants with a prior classification within Genoox; (4) common risk factor variants (aggregated allele frequency $> 5\%$); and (5) rare risk factor variants (aggregated allele

frequency $\leq 5\%$). Variants determined to be likely pathogenic or pathogenic using ACMG criteria are reported back to participants in a screening context in the report results sections as ‘you are at risk for’ or ‘you are a carrier of’.

Copy number variants (CNVs) are initially detected and annotated using an Annovar-based custom pipeline developed by The Center for Applied Genomics (TCAG), and were further filtered for variant inclusion based on (1) OMIM morbid genes; (2) ISCA triplo- or haplo-sensitivity scores; (3) size ($> 20\text{Kb}$); and, (4) frequency ($< 1\%$). CNVs with $> 90\%$ overlap with known benign, polymorphic CNV regions excluded. Variants are filtered further by quality parameters (e.g., confidence of call), internal and population frequencies, published evidence, ClinVar submissions, validity of gene-disease relationships, and participant preferences. Example variant interpretation reports can be found in File S1 in Supporting Information.

Two other analyses were implemented in our genome reports according to existing protocols developed by collaborators within the GENCOV study. These include polygenic risk score calculation for six diseases (atrial fibrillation, coronary artery disease, type 2 diabetes, prostate cancer, colorectal cancer, breast cancer; Hao et al., 2022) and ABO, Rh, and other blood group genotyping (Halls et al., 2020, Lane et al., 2018).

A polygenic risk score (PRS) takes into account multiple genetic changes to provide a personalized risk for developing a specific disease. A PRS provides a “relative” risk of developing a disease compared to individuals of similar age, ethnicity, and genetic background. This score does not account for factors other than genetics that may influence an individual’s risk for common conditions, including lifestyle factors (e.g., diet and exercise) and family history. Using whole genome sequencing data, polygenic risk scores (PRS) were calculated using existing algorithms developed in Hao et al. (2022). All relevant code for PRS calculations is available and adapted from the following public github repository: <https://github.com/MGB-Personalized-Medicine/PRS-adjustment>. The scores were developed from GWAS and replicated in multiple cohorts. Each individual carries 0, 1, or 2 risk alleles at each locus, which are weighted by their per-allele effect size and summed. Unadjusted, raw PRS are calculated with PLINK software and known SNPs and associated effect size (from the discovery GWAS)

Table 18 Example of Polygenic Risk Score Results for Atrial Fibrillation

Disease	General population lifetime risk	Your result
Atrial fibrillation	1 in 4	Average polygenic risk
Risk interpretation: The individual's calculated polygenic risk score, derived from 6730541 loci, has NOT been associated with an increased risk for atrial fibrillation, defined as a greater than 2-fold risk. Scores that fall among the top 7% of polygenic risk score values are associated with a greater than 2-fold risk of developing atrial fibrillation among 409,258 participants of British ancestry when compared to the average individual (Khera_2018_30104762).		
Disease information: Atrial fibrillation (sometimes called Afib) is a type of irregular heart rhythm (arrhythmia) that affects the upper chambers of the heart (atria) and causes the heart to beat very quickly. This may lead to heart failure and stroke. Atrial fibrillation is the most common type of heart arrhythmia, affecting 1%-2% of the general population and approximately 200,000 Canadians (Andrade_2020_33191198; Heart and Stroke Foundation of Canada_2020_Atrial fibrillation). Data from an American cohort of men and women (N = 8725) obtained between 1968-1999 estimated that the lifetime risk (up to 95 years of age) to develop atrial fibrillation was 1 in 4 for men and women 40 years of age or older and 1 in 6 for men and women 40 years of age or older specifically without a previous diagnosis of congestive heart failure or myocardial infarction (Lloyd-Jones_2004_15313941). While genetics plays a role in the risk for atrial fibrillation, it is not the whole picture. There can be many contributors to the development of atrial fibrillation, some of which are coronary artery disease, high blood pressure, diabetes, unhealthy weight, and heavy alcohol consumption.		

for type 2 diabetes (6,917,436 SNPs), atrial fibrillation (6,730,541 SNPs), coronary artery disease (6,630,150 SNPs; Khera et al., 2018), female breast cancer (3820 SNPs; Mavadat et al., 2015, Michailidou et al., 2017), colorectal (81 SNPs; Huyghe et al., 2019), and prostate cancer (147 SNPs; Schumacher et al., 2018). Raw scores are adjusted for ancestry using ancestry-informative principal components to calculate residualized PRS. Individuals with PRS exceeding a threshold corresponding with an odds ratio (OR) >2 are considered at “increased polygenic risk.” All other PRS results are considered “average polygenic risk.” For example, if a person’s OR for developing colon cancer given that they have a certain genetic change(s)/variant(s) is equal to 2, then the individual has a 2-fold increased risk of developing colorectal cancer compared to individuals without the same genetic change(s)/variant(s). The threshold assignment of OR > 2 used to evaluate polygenic risk during the PRS calculation is discussed in Hao et al. (2022).

An example PRS results table for atrial fibrillation as generated through this protocol is shown in Table 18. Each results table indicates the common disease condition analyzed, the general population lifetime risk of the disease, the simplified polygenic risk result,

a summary of the disease risk interpretation, and general disease information as summarized from the literature. The general lifetime risk of disease reflects the risk of developing the particular disease within a certain time interval, which may vary depending on the research study as well the age of a given participant. The simplified polygenic risk result indicates whether the participant is at an increased or average polygenic risk given the calculated and adjusted PRS. The risk interpretation summary provides more details about this polygenic risk result, including the number of loci used for the PRS calculation and the research study used as reference for determining increased and average risk. The general disease information summary provides a brief overview of the common disease being analyzed, the percentage of affected individuals within the general Canadian population, and average lifetime risk, as well as summarizing other non-genetics related factors that may contribute to the risk of developing the disease. The full list of references included on the PRS tables can be found in File S6 in Supporting Information.

Blood group genotyping is included on the genome report to reveal potential associations between ABO and other blood genotypes and susceptibility to a COVID-19

Table 19 Example of Blood Group Genotyping Results

RBC Antigens	
ABO (RhD) blood type	O (RhD+)
Extended RBC antigen genotyping	M+, N-, Mi(a-), S-, s+, U+, He-, P1+, C+, E+, c+, e+, C(W-), C(X-), V-, VS-, Lu(a-), Lu(b+), K-, k+, Kp(a-), Kp(b+), Js(a-), Js(b+), Kp(c-), Fy(a+), Fy(b+), Jk(a+), Jk(b+), Di(a-), Di(b+), Wr(a-), Wr(b+), Sc1+, Sc2-, Do(a+), Do(b+), Hy+, Jo(a+), Co(a+), Co(b-), LW(a+), LW(b-), Cr(a+), Kn(a+), Kn(b-), McC(a+), Yk(a+), McC(b-), KCAM+, KDAS-, Vel+
Rare RBC antigen(s)	
Platelet antigens	
Extended platelet antigen genotyping	HPA-1a+, HPA-1b-, HPA-2a+, HPA-2b-, HPA-3b+, HPA-3a-, HPA-4a+, HPA-4b-, HPA-5a+, HPA-5b-, HPA-6bw-, HPA-15b+, HPA-15a-
Rare platelet antigen(s)	

DISCUSSION

These red blood cell (RBC) and human platelet antigen (HPA) predictions are based on published genotype to phenotype correlations for the alleles present. Some antigens have also been serologically determined using traditional blood typing methods. During pregnancy or transfusion, alloantibodies to blood group antigens and platelet antigens can form against foreign RBCs that contain immunogenic blood group and platelet antigens that the recipient is missing. These alloantibodies can cause clinically important complications during future transfusions and pregnancy.

Blood product donation

This individual would be a desirable donor for the following uncommon antigen negative change(s) found in $\leq 40\%$ of donors: N-, HPA-3a-, HPA-15a-. This individual does not have a rare antigen negative changes.

infection. For example, studies have shown that individuals with blood group A may be at a higher risk for developing COVID-19 than those with non-A blood groups, while individuals with blood group O may actually be more protected against a COVID-19 infection than those with non-O blood groups (Severe Covid-19 GWAS Group et al., 2020). Individuals who are Rhesus-negative (Rh⁻) may also have a protective advantage against COVID-19 versus those who are Rhesus-positive (Rh⁺; Ray, Schull, Vermeulen, & Park, 2021). Germline short variants are called from the red blood cell and platelet antigen genes and promoter regions using Genomic Analysis Tool Kit (GATK) v3.7-0-gcfedb67 (UnifiedGenotyper and EMIT_ALL_SITES output mode). Sequencing coverage is extracted from the alignment file using BEDTools v2.17.0. Antigen typing using bloodTyper is performed at the relevant allele nt positions corresponding to specific antigens using a 4× calling cutoff and SV detection using a combination of read depth, paired reads, and split read methods. The read depth method evaluates copy number and can detect SV changes such as deletions or replacements. The paired read method in which aligned regions are abnor-

mally spaced is useful for observing breakpoints and evaluating read alignments, deletions, and duplications. The split read method detects when a single read matches alignments from disparate, nonsequential locations, indicating breakpoint transition (Halls et al., 2020, Lane et al., 2018).

Example blood group genotyping results tables are shown in Table 19. The first table summarizes the red blood cell (RBC) antigen predictions, including ABO blood type and RhD status. An extended list of RBC antigens is also predicted, including any rare RBC antigens within a given participant. The second table summarizes human platelet antigen (HPA) predictions, similarly including any extended or rare platelet antigen results for the participant. The third table discussion summarizes the RBC and HPA prediction methodology, including any potential implications of particular antigen predictions that may result in complications during pregnancy or blood transfusions for the participant. Finally, the blood product donation summary indicates potential uncommon or rare antigen blood donation candidacy based on predicted genotypes, as well as potential compatibility for blood donations in participants who are determined to

Table 20 Troubleshooting

Problem	Possible cause	Solution
Installation	System incompatibilities—setup and script configuration differs between operating systems (e.g., PLINK, samtools, bcftools, htlib)	Download the most recent version of each package for the correct OS—Mac and Linux distributions recommended. Use Docker images for the ease of running each of the pipelines without manual installation and configuration of packages. OR Follow the manual instructions step-by-step and refer to each of the package distributors for possible installation issues.
Input files	Incompatibility between BAM/VCF/FASTQ files Truncation or corruption errors BAMs are not aligned to GRCh37/hg19 VCF files generated outside of GATK Best Practices Chromosome notation (e.g., “1” vs. “chr1”)	Ensure all files are generated from the same FASTQ input, and aligned to the same reference genome. Verify the integrity of your FASTQ, VCF, BAM files. BAMs must be aligned to GRCh37 (hg19) for HLA genotyping using HLA-VBSeq. VCF generation using GATK Best Practices pipeline is recommended. Ensure chromosome prefixes are compatible between reference fasta and input files.
Computing resources	When your local system does not meet the hardware requirements needed to run each of the pipelines and Docker images	Access pipelines through an HPC resource. OR If using Docker images, opt to download the Github repository and run script on non-containerized environment.
Package does not exist	Package directory is not included in the PATH variable Configuration has not been run successfully	Follow instructions for given OS system on adding package directory in PATH variable. Look into your system C compiler, and package distributors for successful configuration.

have rare blood type results (Halls et al., 2020, Lane et al., 2018).

Viral lineage was also determined for each participant and included on the final genome report. Viral consensus sequences are determined using the CanCOGeN-VirusSeq standardized analysis pipeline based on freebayes v1.7 (<https://github.com/freebayes/freebayes>, RRID:SCR_010761). Quality control is performed using ncov-tools (<https://github.com/jts/ncov-tools>). The viral lineages reported are predictions and not definitive. They are thus representative of the current knowledge of the lineage at the time of issuing the report to participants.

Critical Parameters

HLA calls made by WGS may not be as accurate as PCR-based sequencing methods; therefore, clinical HLA testing is recommended to confirm HLA assignments. Phase of HLA genotypes may not be determined

through this protocol. Reported HLA-disease associations are correlative, not causative. Having a specific HLA genotype is not sufficient for developing the associated disease. Further, the in-house-developed HLA database does not contain an exhaustive list of all possible disease associations published in the literature. Reported pharmacogenomic associations are risks as determined by current CPIC guidelines. The presence of a specific genotype-medication metabolizer relationship does not necessarily mean that an adverse reaction to the medication will occur. Reported metabolizer phenotypes and dosing recommendations are meant to be used as guidelines for certified health professionals. Medication prescriptions and dosage should ultimately be left to the discretion of the appropriate health-care professional/provider. Reported ancestry estimates are generated using datasets containing genomic data from populations of primarily European ancestry; therefore ancestry

Table 21 Time and Compute Considerations

Basic protocol	Runtime ^a	System resources required
HLA	2 hr	N/A
Ancestry	12 hr	Minimum 8 GB RAM
Pharmacogenomics	1 hr	N/A
HLA (Docker image)	N/A	Minimum 16 GB RAM, 4 CPUs
Ancestry (Docker image)	N/A	Minimum 16 GB RAM, 4 CPUs
Pharmacogenomics (Docker)	N/A	Minimum 16 GB RAM, 4 CPUs

^a Runtime is based on a 64 GB RAM M1 MAX system

estimates may not be as accurate for individuals of Non-White/Non-European ancestry. The predictive value of a PRS may also not be as accurate for individuals of Non-White/Non-European ancestry due to similar dataset limitations. In some cases, reported ancestry may not accurately represent self-reported ancestry.

Troubleshooting

Possible problems, causes, and suggested solutions are provided in Table 20. In general, troubleshoot possible input errors through log files in output folders.

Time Considerations

Time and compute considerations can be found in Table 21.

Acknowledgments

This work was supported by funding from the Canadian Institutes of Health Research (Funding Reference Number VR4-172753, 461170, 461304). We would also like to acknowledge CanCOGen for Host and Viral genome sequencing (<https://genomecanada.ca/challenge-areas/cancogen/>).

HostSeq data storage and sharing sub-Committee

Steven Jones (Chair, BC Cancer Agency)
 Inanc Birol (University of British Columbia)
 Guillaume Bourque (McGill University)
 Lisa Strug (Hospital for Sick Children, Toronto)
 Joe Whitney (Hospital for Sick Children)
 Bhooma Thiruv (Hospital for Sick Children)

Author Contributions

Erika Frangione: conceptualization, formal analysis, methodology, software, writing original draft, writing review and editing; **Monica Chung:** conceptualization, formal analysis, methodology, software, writing original draft, writing review and editing; **Selina Casalino:** conceptualization, formal analysis, methodology, writing original draft, writing review and editing; **Georgia MacDonald:** conceptualization, methodology, writing original draft, writing review and editing; **Sunakshi Chowdhary:** project administration, writing review and editing; **Chloe Mighton:** project administration, writing review and editing; **Hanna Faghfoury:** project administration; **Yvonne Bombard:** project administration; **Lisa Strug:** data curation; **Trevor Pugh:** data curation; **Jared Simpson:** data curation; **Limin Hao:** methodology, software, writing review and editing; **Matthew Lebo:** methodology, software, writing review and editing; **William J. Lane:** formal analysis, methodology, software, writing review and editing; **Jennifer Taher:** conceptualization, methodology, project administration, supervision, writing review and editing; **Jordan Lerner-Ellis:** conceptualization, methodology, project administration, supervision, writing original draft, writing review and editing; **Saranya Arnoldo:** project administration; **Navneet Aujla:** formal analysis; **Erin Bearss:** project administration; **Alexandra Binnie:** project administration; **Bjug Borgundvaag:** project administration; **Laurent Briollais:** project administration; **Howard Chertkow:** project administration; **Marc Clausen:** project administration; **Marc Dagher:** project administration;

Luke Devine: project administration; **David Di Iorio:** Formal analysis; **Steven Marc Friedman:** project administration; **Chun Yiu Jordan Fung:** project administration; **Anne-Claude Gingras:** project administration; **Lee W. Goneau:** project administration; **Deepanjali Kaushik:** project administration; **Zeeshan Khan:** project administration; **Elisa Lapadula:** project administration; **Tiffany Lu:** project administration; **Tony Mazzulli:** project administration; **Allison McGeer:** project administration; **Shelley L McLeod:** project administration; **Gregory Morgan:** Formal analysis; **David Richardson:** project administration; **Seth Stern:** project administration; **Ahmed Taher:** project administration; **Iris Wong:** project administration; **Natasha Zarei:** project administration.

Conflict of Interest

Matthew S. Lebo and Limin Hao are employed by a non-for-profit molecular diagnostics lab (Laboratory for Molecular Medicine at Mass General Brigham) that offers genome screening and polygenic risk score assessments. William J. Lane is a member of the Scientific Advisory Board of CareDx, Inc., member of the Next Generation Sequencing Steering Committee of One Lambda, Inc., and his institution is a founding member of the Blood Transfusion Genomics Consortium (BGC) that has received fees from Thermo Fisher Scientific Inc. to help co-develop a high density DNA genotyping array.

Data Availability Statement

Data openly available in a public repository that issues datasets with DOIs.

Literature Cited

- Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., ... Abecasis, G. R., 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. doi: 10.1038/nature15393
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1664–1664. doi: 10.1101/gr.094052.109
- Anastassopoulou, C., Gkizarioti, Z., Patrinos, G. P., & Tsakris, A. (2020). Human genetic factors associated with susceptibility to SARS-CoV-2 infection and Covid-19 disease severity. *Human Genomics*, 14(1), 40. doi: 10.1186/s40246-020-00290-4
- Crews, K. R., Gaedigk, A., Dunnenberger, H. M., Leeder, J. S., Klein, T. E., Caudle, K. E., ... Skaar, T. C. (2014). Clinical pharmacogenetics implementation consortium guidelines for cytochrome p450 2d6 genotype and codeine therapy: 2014 update. *Clinical Pharmacology & Therapeutics*, 95(4), 376–382.
- DiStefano, M. T., Goehringer, S., Babb, L., Alkuraya, F. S., Amberger, J., Amin, M., ... Rehm, H. L. (2022). The gene curation coalition: A global effort to harmonize gene–disease evidence resources. *Genetics in Medicine*, 24(8), 1732–1742. doi: 10.1016/j.gim.2022.04.017
- Eberle, M. A., Fritzilas, E., Krusche, P., Källberg, M., Moore, B. L., Bekritsky, M. A., ... Bentley, D. R. (2017). A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Research*, 27(1), 157–164. doi: 10.1101/gr.210500.116
- Gao, J., Zhu, C., Zhu, Z., Tang, L., Liu, L., Wen, L., & Sun, L. (2019). The human leukocyte antigen and genetic susceptibility in human diseases. *Journal of Bio-X Research*, 2(3), 112–120. doi: 10.1097/JBR.0000000000000044
- Halls, J., Vege, S., Simmons, D. P., Aeschlimann, J., Bujiriri, B., Mah, H. H., ... Lane, W. J. (2020). Overcoming the challenges of interpreting complex and uncommon RH alleles from whole genomes. *Vox Sanguinis*, 115(8), 790–801. doi: 10.1111/vox.12963
- Hamosh, A., Scott, A. F., Amberger, J., Valle, D., & McKusick, V. A. (2000). Online mendelian inheritance in man (OMIM). *Human Mutation*, 15(1), 57–61. doi: 10.1002/(SICI)1098-1004(200001)15:13.0.CO;2-G
- Hao, L., Kraft, P., Berriz, G. F., Hynes, E. D., Koch, C., Kumar, P. K. V., ... Lebo, M. S. (2022). Development of a clinical polygenic risk score assay and reporting workflow. *Nature Medicine*, 28, 1006–1013. doi: 10.1038/s41591-022-01767-6
- Holm, I. A., Agrawal, P. B., Ceyhan-Birsoy, O., Christensen, K. D., Fayer, S., Frankel, L. A., ... Beggs, A. H. (2018). The babyseq project: Implementing genomic sequencing in newborns. *BMC Pediatrics*, 18(1), 225. doi: 10.1186/s12887-018-1200-1
- Huyghe, J. R., Bien, S. A., Harrison, T. A., Kang, H. M., Chen, S., Schmit, S. L., ... Peters, U. (2019). Discovery of common and rare genetic risk variants for colorectal cancer. *Nature Genetics*, 51(1), 76–87. doi: 10.1038/s41588-018-0286-6
- Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., ... Wadelius, M. (2017). Clinical pharmacogenetics implementation consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clinical Pharmacology and Therapeutics*, 102(3), 397–404. doi: 10.1002/cpt.668
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219–1224. doi: 10.1038/s41588-018-0183-z

- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Maglott, D. R. (2018). Clinvar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. doi: 10.1093/nar/gkx1153
- Lane, W. J., Westhoff, C. M., Gleadall, N. S., Aguad, M., Smeland-Wagman, R., Vege, S., ... MedSeq Project. (2018). Automated typing of red blood cell and platelet antigens: A whole-genome sequencing study. *The Lancet Haematology*, 5(6), e241–e251. doi: 10.1016/S2352-3026(18)30053-X
- Lee, S. B., Wheeler, M. M., Thummel, K. E., & Nickerson, D. A. (2019). Calling star alleles with Stargazer in 28 pharmacogenes with whole genome sequences. *Clinical Pharmacology & Therapeutics*, 106(6), 1328–1337. doi: 10.1002/cpt.1552
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Matzaraki, V., Kumar, V., Wijmenga, C., & Zhernakova, A. (2017). The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology*, 18(1), 76. doi: 10.1186/s13059-017-1207-1
- Mavaddat, N., Pharoah, P. D. P., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., ... Garcia-Closas, M. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*, 107(5), djv036. doi: 10.1093/jnci/djv036
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. doi: 10.1101/gr.107524.110
- Mersha, T. B., & Abebe, T. (2015). Self-reported race/ethnicity in the age of genomic research: Its potential impact on understanding health disparities. *Human Genomics*, 9, 1. doi: 10.1186/s40246-014-0023-x
- Meyer, H. (2021a). Processing Hapmap III reference data for ancestry estimation. Available at <https://meyer-lab-cshl.github.io/plinkQC/articles/HapMap.html>
- Meyer, H. (2021b). Ancestry estimation based on reference samples of known ethnicities. Available at <https://meyer-lab-cshl.github.io/plinkQC/articles/AncestryCheck.html>
- Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., ... Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678), 92–94. doi: 10.1038/nature24284
- Miller, D. T., Lee, K., Chung, W. K., Gordon, A. S., Herman, G. E., Klein, T. E., ... Martin, C. L. (2021). ACMG sf v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 23(8), 1381–1390. doi: 10.1038/s41436-021-01172-3
- Mimori, T., Yasuda, J., Kuroki, Y., Shibata, T. F., Katsuoka, F., Saito, S., ... Yamamoto, M. (2019). Construction of full-length Japanese reference panel of class I HLA genes with single-molecule, real-time sequencing. *Pharmacogenomics Journal*, 19, 136–146. doi: 10.1038/s41397-017-0010-4
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81, 559–575. doi: 10.1086/519795
- Ray, J. G., Schull, M. J., Vermeulen, M. J., & Park, A. L. (2021). Association between ABO and RH blood groups and SARS-CoV-2 infection or severe Covid-19 illness. *Annals of Internal Medicine*, 174(3), 308–315. doi: 10.7326/M20-4511
- Reble, E., Salazar, M. G., Zakoar, K.-R., Khalouei, S., Clausen, M., Kodida, R., ... Bombard, Y. (2021). Beyond medically actionable results: An analytical pipeline for decreasing the burden of returning all clinically significant secondary findings. *Human Genetics*, 140(3), 493–504. doi: 10.1007/s00439-020-02220-9
- Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., ... Watson, M. S. (2015). ClinGen—the clinical genome resource. *New England Journal of Medicine*, 372(23), 2235–2242. doi: 10.1056/NEJMSr1406261
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. doi: 10.1038/gim.2015.30
- Schumacher, F. R., Olama, A. A. A., Berndt, S. I., Benlloch, S., Ahmed, M., Saunders, E. J., ... Eeles, R. A. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics*, 50(7), 928–936. doi: 10.1038/s41588-018-0142-8
- Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., ... Karlsen, T. H., Severe Covid-19 GWAS Group. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *The New England Journal of Medicine*, 383(16), 1522–1534. doi: 10.1056/NEJMoa2020283
- Solomon, B. D., Nguyen, A.-D., Bear, K. A., & Wolfsberg, T. G. (2013). Clinical genomic database. *Proceedings National Academy of Science USA*, 110(24), 9851–9855. doi: 10.1073/pnas.1302575110
- Sze, S., Pan, D., Nevill, C. R., Gray, L. J., Martin, C. A., Nazareth, J., ... Pareek, M. (2020). Ethnicity and clinical outcomes in Covid-19: A systematic review and meta-analysis. *EClinicalMedicine*,

- 29, 100630. doi: 10.1016/j.eclinnm.2020.100630
- Taher, J., Mighton, C., Chowdhary, S., Casalino, S., Frangione, E., Arnoldo, S., ... Lerner-Ellis, J. (2021). Implementation of serological and molecular tools to inform Covid-19 patient management: Protocol for the GENECOV prospective cohort study. *BMJ Open*, 11(9), e052842. doi: 10.1136/BMJOPEN-2021-052842
- Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., ... McEwen, J. E., The International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), 52–58. doi: 10.1038/nature09298
- Thorn, C. F., Klein, T. E., & Altman, R. B. (2013). PharmGKB: The pharmacogenomics knowledge base. *Methods in Molecular Biology*, 1015, 311–320. doi: 10.1007/978-1-62703-435-7_20
- Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C. F., Whaley, R., & Klein, T. E. (2021). An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics*, 110(3), 563–572.
- Zeberg, H., & Pääbo, S. (2020). The major genetic risk factor for severe Covid-19 is inherited from Neanderthals. *Nature*, 587(7835), 610–612. doi: 10.1038/s41586-020-2818-3

Internet Resources

- https://gis.cdc.gov/grasp/covidnet/covid19_3.html
- COVID-NET: Covid-19-associated hospitalization surveillance network, centers for disease control and prevention (2021).