

# **Translational Bioinformatics for Immunogenomics**

# Table of contents

<b>Welcome</b>	<b>5</b>
<b>Introduction</b>	<b>6</b>
<b>1 Human Leukocyte Antigens</b>	<b>7</b>
1.1 Background . . . . .	7
1.2 HLA Class I . . . . .	7
1.3 HLA Class II . . . . .	7
1.4 HLA Nomenclature . . . . .	7
1.5 Functional Divergence . . . . .	7
1.6 HLA Imputation Programs . . . . .	8
<b>2 Killer Cell Immunoglobulin-like Receptors</b>	<b>9</b>
2.1 Background . . . . .	9
2.2 KIR Locus . . . . .	9
2.3 KIR Diversity . . . . .	12
2.4 NK Cell Education . . . . .	12
2.5 KIR Nomenclature . . . . .	13
2.5.1 Inhibitory KIR . . . . .	13
2.5.2 Activating KIR . . . . .	13
2.5.3 Broad KIR Haplotypes . . . . .	13
2.6 KIR Ligand Motifs . . . . .	13
2.7 KIR3DL1 and KIR3DS1 . . . . .	14
2.8 KIR Allele Imputation Programs . . . . .	14
<b>3 ERAP</b>	<b>16</b>
<b>4 Epistatic Interactions</b>	<b>17</b>
4.1 KIR-HLA . . . . .	17
<b>Bonus</b>	<b>18</b>
<b>5 Drug Allergy</b>	<b>19</b>
5.1 Cephalosporins . . . . .	19
5.2 Bupropion . . . . .	19
5.3 Fluoroquinolones . . . . .	19

5.4	Iron . . . . .	19
5.5	Local Anesthetics . . . . .	19
5.6	Penicillins . . . . .	19
5.7	Radiocontrast . . . . .	19
5.8	Sulfa Antibiotics . . . . .	19
<b>6</b>	<b>Genotype Imputation</b>	<b>20</b>
6.1	Michigan Imputation Server . . . . .	20
6.2	TOPMed Imputation Server . . . . .	20
6.3	Reference Panels . . . . .	20
6.4	Genome Assemblies . . . . .	21
<b>7</b>	<b>Bioinformatic Best Practices</b>	<b>22</b>
7.1	Project Organization . . . . .	22
7.2	Version Control with Git . . . . .	22
7.3	File Naming Conventions . . . . .	23
7.4	Application Containers with Docker . . . . .	24
<b>8</b>	<b>Presenting Your Medical Research</b>	<b>25</b>
8.1	Font . . . . .	25
8.2	Font Size . . . . .	25
8.3	Word Count . . . . .	26
8.4	Timing . . . . .	26
8.5	Figures . . . . .	26
8.6	References . . . . .	26
8.7	Equipment . . . . .	27
8.7.1	Laptop . . . . .	27
8.7.2	Hub . . . . .	27
8.7.3	USB Drive . . . . .	27
8.7.4	Presentation Remote . . . . .	27
<b>9</b>	<b>On Being a Physician-Scientist</b>	<b>28</b>
9.1	Academic Medicine Jobs . . . . .	28
9.1.1	AAMC Faculty Salary Report . . . . .	28
9.1.2	Tenure-Track Offer Letters . . . . .	28
9.2	NIH Loan Repayment Program . . . . .	28
9.3	Online Resources . . . . .	28
9.3.1	Edge for Scholars . . . . .	28
9.3.2	NCFDD . . . . .	28
9.4	Professional Organizations . . . . .	29
9.5	Suggested Readings . . . . .	29
9.5.1	Not Discussed . . . . .	29
9.5.2	Publishing Your Medical Research . . . . .	29



# Welcome

A few key web sources include the [Immuno Polymorphism Database](#) and [Allele Frequency Net Database](#).

# Introduction

# 1 Human Leukocyte Antigens

## 1.1 Background

HLA is located on chromosome 6 in the Major Histocompatibility Complex (MHC).

## 1.2 HLA Class I

HLA class I molecules are expressed by healthy nucleated cells.

## 1.3 HLA Class II

HLA class II molecules are expressed by professional antigen-presenting cells (APC)—dendritic cells, macrophages, and B cells.

## 1.4 HLA Nomenclature

## 1.5 Functional Divergence

Heterozygosity of HLA class I genes is associated with better outcomes after HIV infection. This is thought to be due to a greater repertoire of HIV peptides presented and cytotoxic T cell response. However, looking at HLA class I allotype alone does not take into account differences in actual peptide repertoire. Viard and O’Uigin developed a metric to measure this difference, termed “functional divergence.” Functional divergence predicts the peptide repertoire as a continuum. They showed that greater functional divergence was associated with better HIV outcomes. Functional divergence may be relevant to other diseases where HLA heterozygosity confers advantage, such as infection, vaccination, and immunotherapy.

You can download functional divergence estimates for pairwise combinations of HLA-A, HLA-B, and HLA-C alleles from their article’s Supplementary Materials. The functional divergence measure ranges from 0 (i.e., smallest functional divergence) to 1 (i.e., greatest functional divergence).

## 1.6 HLA Imputation Programs

Name	Programming Language	Input Data	Output	Reference
<a href="#">SNP2HLA</a>	Command line interface	PLINK binary format	HLA class I and II alleles	(Jia et al. 2013)
<a href="#">HIBAG</a>	R	Plink binary format	HLA class I and II alleles	(Zheng et al. 2014)



## 2 Killer Cell Immunoglobulin-like Receptors

### 2.1 Background

KIR is located on chromosome 19 (19q13.4) in the Leukocyte Receptor Complex (LRC). KIR is expressed on the surface of Natural Killer (NK) cells and some T cells. KIR do not undergo somatic rearrangement—a key difference from T-cell receptors. KIR interacts with HLA class I—their cognate ligand—to recognize and destroy unhealthy tissue cells while preventing the same from occurring to healthy cells. Therefore, NK cells play a role in fighting infections, resisting some cancers, pregnancy, and preventing autoimmunity. For further reading and references, I highly recommend the review article by Pollock, Harrison, and Norman on the immunogenetics and co-evolution of KIR and HLA class I.

### 2.2 KIR Locus

Adapted from Pollock, Harrison, and Norman. JACI: In Practice. 2022.

Gene

3DL3

2DS2

2DL2/3

2DL5B

2DS3

2DL1

2DL4

3DL1

3DS1

2DL5A

2DS5

2DS4  
2DS1  
3DL2  
Function  
Inhibit.  
Activ.  
Inhibit.  
Inhibit.  
Activ.  
Inhibit.  
Activ.  
Inhibit.  
Activ.  
Activ.  
Activ.  
Inhibit.  
Alleles  
228  
65  
98  
47  
71  
173  
112  
184  
39  
44

88

39

33

166

Allotypes

112

22

50

21

23

65

58

92

22

19

38

20

12

115

HLA Class I Ligand Motifs

B7H7

A\*11 C1

*B46:01 B73:01* C1 C2

PVR

?

C2

HLA-G

Bw4+ HLA-A and Bw4+ HLA-B

Bw4+ HLA-B and HLA-F

PVR

C2

A\*11 HLA-C

C2

A3 A11

: Adapted from *Pollock, Harrison, and Norman. JACI: In Practice. 2022.*

## 2.3 KIR Diversity

KIR diversity is influenced by gene content variation and sequence variation. Distinct DNA sequences of KIR genes are called “alleles.” Distinct polypeptide sequences of KIR genes are called “allotypes.” Because different DNA sequences of KIR gene can lead to the same polypeptide, there are more alleles than allotypes for a given KIR gene.

KIR Diversity Concept	Definition
Gene Content Variation	Presence/absence, fusion, duplication
Sequence Variation	May alter ligand affinity or specificity, signal transduction ability, or surface expression (e.g., promoter activity, translation, intracellular trafficking)
Allele	Distinct DNA sequence
Allotype	Distinct polypeptide sequence

## 2.4 NK Cell Education

NK Cell Education (i.e., Arming, Licensing)	Corresponding Pairs of KIR and HLA Class I Ligands	Cytotoxicity and other Effector Abilities
Strong	Many	More
Weak	Few	Less

## 2.5 KIR Nomenclature

### 2.5.1 Inhibitory KIR

The main role of inhibitory KIR is to prevent cytotoxic NK and T cells from killing tissue cells—unless their HLA class I expression is lost or altered by infection or mutagenesis.

### 2.5.2 Activating KIR

Activating KIR help identify diseased cells for destruction by cytotoxic NK and T cells. Binding of foreign peptides by HLA class I molecules retained by infected cells may be most critical for activating KIR.

### 2.5.3 Broad KIR Haplotypes

Broad KIR Haplotype	KIR Copy Number Variation	KIR Gene Organization	Activating KIR
A	Relatively stable	Generally non-variable	Less
B	Extensive	Highly variable	More

## 2.6 KIR Ligand Motifs

Table 2.4: Adapted from *Pollock, Harrison, and Norman. JACI: In Practice. 2022.*

KIR Ligand Motif	HLA-A Allotypes	HLA-B Allotypes	HLA-C Allotypes
A3/A11	A*03, A*11		

KIR Lig- and Mo- tif			
	HLA-A Allotypes	HLA-B Allotypes	HLA-C Allotypes
Bw4	A*23, A*24, A*32	B*07:27, B*08:02, B*08:03, (B13), B*15:13, B*15:16, B*15:17, B*15:23, B*15:24, B*15:36, B*15:43, B*15:67, B*27:01, B*27:02, B*27:03, B*27:04, B*27:05, B*27:07, B*37, B*38, B*40:13, B*40:19, B*44, B*47, B*49, B*51, B*52, B*53, B*56:07, B*57, B*58, B*59	_____
C1	C*01, C*03, C*07, C*08, C*12:02, C*12:03, C*12:06, C*12:08, C*13, C*14, C*16	B*46, B*73	_____
C2	C*02, C*03:07, C*04, C*05, C*06, C*12:04, C*12:05, C*12:07, C*14:04, C*15, C*16:02, C*17, C*18	_____	_____

## 2.7 KIR3DL1 and KIR3DS1

Because of significant non-allelic recombination in the KIR region, the distinction between KIR genes and alleles can be confusing. Specifically, KIR3DL1 and KIR3DS1 are alleles of the same gene. Of the KIR3DS1 allotypes—3DS1013 and 014—are observed with the greatest frequency in any population.

## 2.8 KIR Allele Imputation Programs

Name	Programming			Reference
	Lan- guage	Input Data	Output	
<a href="#">PONG</a>	R	PLINK bi- nary for- mat	KIR3DL1/S1 alleles (Global Model includes 51 alleles)	<a href="#">Harrison, 2022</a>
<a href="#">KIR*IMP</a>	Online portal	HAPS/SAMPLE for- mat	AMPLEtypes: 17 loci (presence/absence and copy number) plus 2 extended haplotype classifications ( A and B haplotypes)	<a href="#">Vukcevic, 2015</a>

## 3 ERAP

ERAP is located on chromosome 5.



## 4 Epistatic Interactions

### 4.1 KIR-HLA

Epistatic interactions between KIR and HLA are associated with ankylosing spondylitis ([Hanson, 2020](#))

## Bonus

## **5 Drug Allergy**

**5.1 Cephalosporins**

**5.2 Bupropion**

**5.3 Fluoroquinolones**

**5.4 Iron**

**5.5 Local Anesthetics**

**5.6 Penicillins**

**5.7 Radiocontrast**

**5.8 Sulfa Antibiotics**

## 6 Genotype Imputation

### 6.1 Michigan Imputation Server

The [Michigan Imputation Server](#) is a free next-generation genotype imputation platform. You can learn more about the Michigan Imputation Server by visiting their [Getting Started](#) documentation. The [1000 Genomes Phase 3 \(Version 5\)](#) Reference Panel is available on the Michigan Imputation Server.

### 6.2 TOPMed Imputation Server

The [TOPMed Imputation Server](#) is another free next-generation genotype imputation platform developed by the [University of Michigan](#) and powered by data from the [TOPMed Program](#) investigators. You can learn more about the TOPMed Imputation Server by visiting their [Getting Started](#) documentation. The [TOPMed Version 3](#) Reference Panel was released in December 2023.

### 6.3 Reference Panels

Reference Panel	Genome			Chr.	Imputation Server
	As- sem- bly	No. of Sam- ples	Sites (chr1- 22)		
<a href="#">1000 Genomes Phase 3 (Version 5)</a>	GRCh37/2.504	2,504	49,143,605	1-22, X	<a href="#">Michigan Imputation Server</a>
<a href="#">TOPMed (Version 3)</a>	GRCh38/138.38	138,389	445,600,184	1-22, X	<a href="#">TOPMed Imputation Server</a>

## 6.4 Genome Assemblies

The [Genome Reference Consortium](#) (GRC) is the main source of human genome assembly data. The most recent human genome assembly version is GRCh38, released in 2013. The “h” in “GRCh” stands for “human.” The GRC also maintains genome assembly data for rat (r), mouse (m), zebrafish (z), and chicken (g for gallus). Major updates, called “versions”, are released every few years. Minor updates are called “patches” and are released more frequently.

GRCh38 is referred to as “hg38” in the [University of California Santa Cruz \(UCSC\) Genome Browser](#). The “hg” stands for “human genome.” Before the GRCh38 genome assembly, the version numbers of the GRC and UCSC Genome Browser genome assemblies did not match. For example, when the GRCh37 genome assembly was released in 2009, the UCSC Genome Browser version was “hg19.” Therefore, to minimize confusion, starting with the GRCh38 genome assembly, the [UCSC Genome Browser](#) version number was matched as “hg38.”

GRC Version	UCSC Version	Year Released	Genome Coverage	Alternate Haplotypes
GRCh37	hg19	2009	~92.5%	3 regions with 9 alternate loci
GRCh38	hg38	2013	95%	178 regions with 261 alternate loci

## 6.5

## 7 Bioinformatic Best Practices

I recommend the tutorial, “A Reproducible Data Analysis Workflow With R Markdown, Git, Make, and Docker” as a starting point for R-based data analyses ([Peikert & Brandmaier, 2021](#)).

### 7.1 Project Organization

#### Bash Commands to Create Folder Directory Structure for Your R Project

```
cd </path/to/parent/directory>
mkdir <your-r-project-folder>
cd <your-r-project-folder>
touch README.md
mkdir data doc src bin outputs
```

Once you have downloaded your raw data to your data folder, you should make the contents of the data folder read-only (non-editable) with the following command: `chmod u-w -R data/`

### 7.2 Version Control with Git

I recommend the [Using Git and GitHub with RStudio Cheatsheet](#) for additional helpful commands.

#### Verify Git Installation and Version

```
which git # request path to your Git executable
git --version # check your Git version
```

#### Introduce Yourself to Git

```
git config --global user.name "<username>"
git config --global user.email "<email>"
```

## Create a New Repository on GitHub

Go to GitHub to create your new repository, then initialize your repository from the command line.

```
cd </path/to/your-r-project-folder>
echo "# your-r-project-folder" >> README.md
git init
git add README.md
git commit -m "first commit"
git branch -M main
git remote add origin https://github.com/<user.name>/<your-repository>.git
git push -u origin main
```

## Add, Commit, and Push Files to Remote Repository

```
git add <file-name>
git commit -m "description"
git push
```

## 7.3 File Naming Conventions

In your README.md, you should define naming conventions for your project files. The main elements for a file naming convention are metadata, separator, and version tracking. I recommend the [File Naming Conventions Worksheet](#) (Briney, 2020) to develop your file naming conventions.

Metadata	Separator	Version Tracking
3 to 5 pieces max (e.g. sample ID, date in ISO 8601 format such as YYYY-MM- DD)	Dashes (-), underscore (_), or camel case (i.e., capitalize each word without spaces)	Numeric (e.g., v01) or Status (e.g., raw, processed)

#### Example

My naming convention for R Markdown analysis files is: “analysis-YYYY-MM-DD-version.Rmd” where version starts with “v01.” This is my first analysis file, “analysis-YYYY-MM-DD-v01.Rmd”

## 7.4 Application Containers with Docker






## 8 Presenting Your Medical Research

### 8.1 Font

You should use a [sans-serif](#) font like Arial to maximize readability. “Serifs” are extending features at the end of letters. Times New Roman is a serif font.

Table 8.1: Sans-serif versus serif fonts<sup>1</sup>

Font	Illustration	Examples
Sans-serif font		Arial, Calibri, Helvetica
Serif font		Times New Roman, Georgia, Garamond
Serifs (colored in red)		

### 8.2 Font Size

<sup>1</sup>Font images are recreated by User:Stannered, original by en>User:Chmod007 - en:Image:Serif and sans-serif 01.png, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=2058303>

Slide Section	Font Size
Title	36 – 44
Text (e.g., Bullets, Figures, Tables)	24 – 28
References	20 – 24

### 8.3 Word Count

The fewer words, the better. A rule to follow is the **7×7 rule**: no more than 7 lines and no more than 7 words per line.

### 8.4 Timing

You should estimate approximately 1 minute per slide.

### 8.5 Figures

I recommend creating your figures as [Scalable Vector Graphics](#) (SVG). The main advantages of the SVG format include always maintaining its resolution and smaller file size than pixel-based image formats (e.g., JPEG).

Some tools that you can use to get started creating SVG include [Microsoft PowerPoint](#) (subscription), [Adobe Illustrator](#) (subscription), [draw.io](#) (free), and [Inkscape](#) (free). [Draw.io](#) is best for diagrams and flowcharts. [Inkscape](#) is better for flexible drawings. Both [draw.io](#) and [Inkscape](#) are integrated with [Bioicons](#), an open-source extension which includes >1700 icons for scientific illustrations.

In Microsoft PowerPoint, you can create an SVG file by selecting all shapes, right-clicking, choosing “Save as Picture”, and then picking “SVG” as the “Save as Type.”

### 8.6 References

Cite references at the bottom of your slides as you present information.

**i** Format

Last Name. *Journal Abbreviation*. Year.

## 8.7 Equipment

### 8.7.1 Laptop

Bring your own laptop to presentations in case there isn't a desktop computer for you to use, or it is not functioning, reliable or frustratingly slow.

### 8.7.2 Hub

What is worse than not being able to connect your laptop to the correct cable? While a good host for a presentation should have a hub (or dongle if that's your preferred terminology), you can come prepared with your own too—particularly important if you have a laptop with only USB-C ports and no HDMI port.

There are lots of options for hubs. If you are looking for a recommendation, I've found that [Anker](#) usually has a selection of high-quality and affordable hubs.

### 8.7.3 USB Drive

Do you want the entire audience to see your most recent emails when you login to download the PowerPoint you emailed yourself? No. Me either. To avoid this, bring your presentation loaded onto a USB drive, which should ideally have both USB-A and USB-C ports. Or, you can also avoid this by using your own laptop—where the presentation should already be downloaded.

### 8.7.4 Presentation Remote

I don't feel as strongly about bringing your own presentation remote as your own laptop, hub, and USB drive—but I think it is another piece of equipment to consider. This helps keep you from being tethered to and white knuckling the podium during your talk.

One option to consider is the [Logitech Spotlight Presentation Remote](#)—which includes features such as magnification, vibration alerts for time management (e.g., 5 minutes left), 3 hours use from 1 minute of charging, and connection by USB receiver or Bluetooth—in addition to slide advancement.

## 9 On Being a Physician-Scientist

### 9.1 Academic Medicine Jobs

#### 9.1.1 AAMC Faculty Salary Report

Looking to get an idea of academic faculty salaries? The annual [AAMC Faculty Salary Report](#) compiles academic faculty salaries by rank, degree, department/specialty, medical school type, region, and more. This is often available for free through your university library. Get to know your librarian!

#### 9.1.2 Tenure-Track Offer Letters

What goes into a tenure-track offer letter? The [Burroughs Wellcome Fund](#) provides a comprehensive list of offer letter components in their article, “[Academic Tenure-Track Offer Letters](#).”

### 9.2 NIH Loan Repayment Program

[NIH Loan Repayment Program](#)

### 9.3 Online Resources

#### 9.3.1 Edge for Scholars

[Edge for Scholars](#)

#### 9.3.2 NCFDD

[NCFDD](#)

## **9.4 Professional Organizations**

[American Physician Scientists Association](#)

## **9.5 Suggested Readings**

### **9.5.1 Not Discussed**

[Not Discussed](#) by Michael Stein

### **9.5.2 Publishing Your Medical Research**

[Publishing Your Medical Research](#) by Daniel W. Byrne

## References

- Jia, Xiaoming, Buhm Han, Suna Onengut-Gumuscu, Wei-Min Chen, Patrick J. Concannon, Stephen S. Rich, Soumya Raychaudhuri, and Paul I. W. De Bakker. 2013. “Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens.” Edited by Jianming Tang. *PLoS ONE* 8 (6): e64683. <https://doi.org/10.1371/journal.pone.0064683>.
- Zheng, X, J Shen, C Cox, J C Wakefield, M G Ehm, M R Nelson, and B S Weir. 2014. “HIBAG—HLA Genotype Imputation with Attribute Bagging.” *The Pharmacogenomics Journal* 14 (2): 192–200. <https://doi.org/10.1038/tpj.2013.18>.