Translational Bioinformatics for Immunogenomics

Matt Krantz

6/22/2024

Table of contents

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'hUigin 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
SNP2HLA	Command line interface	PLINK binary format	HLA class I and II alleles	Jia, 2013
HIBAG	R	Plink binary format	HLA class I and II alleles	Zheng, 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.

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

B*46:01 B*73:01 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

A*3 A*11

: 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 cytoxic 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 Ligand Motif	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–3DS1*013 and* 014–are observed with the greatest frequency in any population.

2.8 KIR Allele Imputation Programs

Name	Programming Language	Input Data	Output	Reference
PONG	R	PLINK binary format	KIR3DL1/S1 alleles (Global Model includes 51 alleles)	Harrison, 2022
KIR*IMP	Online portal	HAPS/SAMPLE format	19 KIR types: 17 loci (presence/absence and copy number) plus 2 extended haplotype classifications (A and B haplotypes)	Vukcevic, 2015

3 ERAP

ERAP is located on chromsome 5.

4 Epistatic Interactions

4.1 KIR-HLA

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

Bonus

5 Genotype Imputation

5.1 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.

5.2 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.

5.3 Reference Panels

Reference Panel	Genome Assembly	No. of Samples	Sites (chr1-22)	Chr.	Imputation Server
1000 Genomes Phase 3 (Version 5)	GRCh37/hg19	2,504	49,143,605	1-22, X	Michigan Imputation Server
TOPMed (Version 3)	GRCh38/hg38	133,597	445,600,184	1-22, X	TOPMed Imputation Server

5.4 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

5.5

6 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).

6.1 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/`

6.2 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
```

6.3 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"

6.4 Application Containers with Docker

7 Presenting Your Medical Research

7.1 Font and Font Size

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.

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

7.2 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.

7.3 Timing

You should estimate approximately 1 minute per slide.

7.4 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).

Tools that you can use to get started creating SVG include Microsoft PowerPoint, draw.io (free), and Inkscape (free). Draw.io is best for diagrams and flowcharts. Inkscapeis 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.”

7.5 References

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

ℹ Format

Last Name. *Journal Abbreviation.* Year.

8 On Being a Physician-Scientist

8.1 Academic Tenure-Track Offer Letters

The Burroughs Wellcome Fund provides a comprehensive list of offer letter components in their article, "Academic Tenure-Track Offer Letters."