HIV Database Workshop
www.hiv.lanl.gov
seq-info@lanl.gov

Presenters:
Will Fischer, Brian Foley, Bette Korber

Database PIs: Bette Korber, Thomas Leitner,
Karina Yusim, Brian Foley

Additional database staff: Werner Abfalterer,
Kumkum Ganguly, Jennifer Macke, James Szinger, and Hyejin Yoon

Contract Officer Representative: Anjali Singh, NIAID, NIH

Theoretical Biology and Biophysics, T-6
Los Alamos National Laboratory
Los Alamos HIV Database

“I think of it as the gift that keeps on giving”, A. Fauci


http://tinyurl.com/HIV-DB-2018
Los Alamos HIV Database

HIV Databases, funded by NIH

- Integrate HIV immunological and viral and host sequence data
- > 60 computational tools, some HIV specific; many applicable to other pathogens
- Tables, summaries, web search interfaces
- Annual Compedia

- HIV Sequence database – founded in 1986, Gerald Myers
  - Sequence data from GenBank with added metadata fields from the literature
  - Metadata and accession numbers incorporated in the sequence names
  - Premade and on the fly alignments – align indels and reduce sequences per person
  - Web searches: subtype, geographic location, patient details, sampling year, etc ~40 fields

- HIV Immunology database – founded in 1995, Bette Korber
  - Comprehensive HIV epitope database,
  - Integrates HIV immunological and sequence data
  - Web searches: epitope, protein, HLA type, immunogen, keywords, patient details, etc

Other pathogen databases

- HCV Database – founded in 2003, Carla Kuiken, initially funded by NIH

- HFV Database – founded in 2009, C. Kuiken, initially funded by DoD, >80 viral species
  - Filovirus portion of the database was updated during and after the 2015 outbreak
  - Premade sequence alignments on genus, species and one-per-outbreak sequence levels
  - Epitope lists and genomic maps, functional domains
  - Ebola Genome browser
HIV Database Workshop Logistics

■ Day 1, Jan 30, Tues
  □ HIVSequence Database

■ Day 2, Jan 31, Wed
  □ HIV Immunology Database
    ■ Part 1:
      □ HIV Immunology Database overview
      □ Antibody searches and entries in HIV database
      □ Neutralizing Antibody Resources
      □ CATNAP, both tailored for HIV and applicable to any pathogen
      □ CombiNaber, applicable for any pathogen
      □ HIV Genome Browser

■ Part 2:
  □ T cell epitopes and searches and entries in HIV database
  □ More computational tools for Immunologists, many applicable for any pathogen
  □ Vaccine design and evaluation tools, applicable for any pathogen
The HIV databases contain data on HIV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials. The website also gives access to a large number of tools that can be used to analyze these data. This project is funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID), a part of the National Institutes of Health (NIH). Click on any of the links below to access a database. Editorial Board

News:

CATNAP: Custom Input
The original CATNAP tool can compile, analyze and tally neutralizing antibody panels from a database of publicly available HIV neutralization data. A new version, CATNAP: Custom Input, is now available. This version allows users to input their own neutralization panel data and perform the same analyses. HIV Env sequences are available as a premade alignment, or can be provided by the user. 12 March 2015

HIV Molecular Immunology 2014
HIV Molecular Immunology 2014 is now available online. The PDF version is hypertext enabled and features clickable table-of-contents, indexes, references and links to external web sites. 04 February 2015

2014 HIV Sequence Compendium
2014 was the last year that we printed and shipped the HIV Sequence Compendium. Printed copies of the 2014 compendium are still available on request. 21 January 2015
All kinds of basic information about HIV and about our database

Previous workshop presentations

Yes! We do respond to this e-mail address!

Questions or comments? Contact us at seq-info@lanl.gov
HIV Sequence Database

Programs and Tools
- **Search Interface** retrieves HIV and SIV sequences, which can then be aligned and used to build trees
- **Geography Search Interface** retrieves HIV sequences based on geographical distribution
- **Genome Browser** uses jBrowse to display diverse data about the HIV-1 genome and proteome
- **Tools for working with sequences** lists all our online tools, organized by function

Alignments
- **HIV Premade Alignments** Includes Consensus and Ancestral Sequences, Subtype Reference Alignments, and Complete Alignments

Information
- **HIV Sequence Compendium** print or order our annual publication
- **Tutorials and other Information** unpublished web-based content
- **Links** to other HIV/AIDS tools and information

About this website
- **FAQ** general information about this website
- **Site Statistics** usage Information for www.hiv.lanl.gov

News:

**IQ-TREE interface**
IQ-tree is a fast and effective stochastic algorithm for finding ML trees. We have developed a convenient web server for building trees with this method. A nice feature of this method is the ability to output a table of site-specific rates of evolution for each position in the alignment. 18 September 2017

**IEDB User Workshop 2017**
The Immune Epitope Database (IEDB) will hold its 2017 User Workshop on October 25-26, 2017 in Rockville, Maryland. Staff from the LANL HIV Databases will be there to talk about our Immunology Database, Sequence Database, and bioinformatics tools. More information is available at [http://workshop.iedb.org/](http://workshop.iedb.org/). 18 July 2017
HIV Immunology Database Overview

- Experimentally characterized immunological and associated viral data

- Key information from each paper on HIV T cell epitopes or mAbs
  - ~10,000 CTL, >1,500 Helper epitopes and >3,000 Antibody records
  - Epitope sequence, location, immunogen, vaccine details, patient details...
  - Epitope Variants (escape, reduced binding, etc.)
  - Host HLA or MHC, Ab isotype, binding region
  - Neutralizing Antibody Resources, contact residues, etc.
  - Notes summarize main findings

- HIV T cell epitopes and Antibody data organization
  - T Cells (CTL and Helper epitopes)
    - One reference per entry, epitope/HLA combinations are often repeated
    - CTL and T-helper database organization is identical
  - B Cells (Antibodies)
    - One entry for each monoclonal antibody
    - Many references per entry (> 800 for some well studied mAbs)
    - Antibody is entered and annotated whether or not epitope is defined

- HIV Immunology Database products
  - Epitope maps, summary tables and yearly compendium
  - Computational tools for immunologists
  - Neutralizing antibody resources
Tools for Immunologists

Most tools are applicable to any organism and some to any numerical data

- **CATNAP**: Compile, Analyze and Tally published and your own NAb Panels
- **CombiNAb**: Predict and analyze neutralization by antibody combinations
- **Sequence Locator**: Find epitope location on the reference genome
- **PepMap**: Map an input set of peptides on the reference sequence (Fasta, PDF and HTML)
- **PeptGen**: Generate sets of overlapping peptides for epitope mapping.
- **QuickAlign** and **AnalyzeAlign**: Align query sequences or discontinuous positions to an alignment, create WebLogos, calculate frequency by position, tally variants in an alignment
- **ELF**: Epitope Location Finder. Search query sequence for
  - Known epitopes from our HIV immunology databases
  - HLA binding motifs
  - Epitopes predicted by the IEDB binding algorithm.
- **N-Glycosite**: Find potential N-linked glycosylation sites in an alignment
- **Mosaic** and **Epigraph**: Generate candidate vaccine protein cocktails with optimized potential epitope coverage, calculate and visualize coverage
- **Heatmap**: Display and organize neutralization or other quantitative data.
- And more …
Neutralizing Antibody Resources

www.hiv.lanl.gov/content/immunology/neutralizing_ab_resources.html

2 new tools useful for Ab analysis

Coming soon:

- Genetic Signature tool
  - Finds phylogenetically corrected genetic signatures in a sequence alignment in conjunction with a phenotype file.

- Filtered Forests
  - Machine learning predictions of bNAb viral sensitivity

Tools

- CATNAP: Compile, Analyze and Tally NAb Panels
  - Analyses of panels of antibody data for identification of potential genetic signatures.
    - Database CATNAP analyzes published IC50/IC90 data for anti-HIV neutralizing antibodies.
    - Custom CATNAP analyzes any numerical data associated with a protein alignment.
    - Hybrid CATNAP analyzes your neutralization data together with published data.

- CombiNAber
  - Predict the neutralization of combinations of antibodies

- HIV Genome Browser
  - A customization of JBrowse displaying genome and proteome features of HIV, including epitopes and neutralizing antibody features.

- External Tools for Germline Antibody Reconstruction
  - A list of external computational tools for modeling antibody evolution and germ line reconstruction from antibody or T-cell receptor sequence data.

Search interface

- Neutralizing antibody contexts and features
  - Search for locations of important neutralizing antibody binding sites and other HIV-1 Env features.

Tables

- Neutralizing antibody contexts and features (.xls)
  - A summary of the information from the search interface above, presented in a single .xls spreadsheet. Each row of the table corresponds to one residue of HIV-1 Env, and each column represents a protein feature or set of known binding residues of broadly neutralizing antibodies. Loops and other features of Env are shown in the first 3 columns on the left. The entropy (sequence variability) of each residue is presented numerically and color coded. Abbreviated references are listed under each column heading, and full references are on the second page of the Excel file.

- Best neutralizing antibodies
  - A table presenting the most broadly-neutralizing HIV-1 antibodies, with links to papers, Ab sequences, structures, notes on breadth of neutralization, locations of Ab contacts or key residues, and heavy and light chain composition.

Protocols

- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

Questions or comments? Contact us at immuno@lanl.gov
Epitope alignments: epitopes aligned to HIV subtype Reference sequences in Fasta format

Reactive peptide maps and tables (with HLA and other patient data) from several large-scale studies scanning HIV proteins.
Antibody Search

An example workflow following from antibody search:

- Search database for a particular antibody record

- For a neutralizing antibody collect comparative neutralization data for that antibody tested against different viruses and in different studies (CATNAP)

- Estimate the effectiveness of multi-antibody cocktails against different viruses (CombiNAber)
Antibody Search (https://www.hiv.lanl.gov/content/immunology/ab_search)

Search by

- HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, Author, Country, Keywords, Isotype
- MAb ID
  - List by Ab name
  - List by Ab type
    - By binding site, for example binding to similar region like V3 or near a common functional domain like CD4 binding site CD4Bs)
- Search examples:
  - 2F5 – 1 record with 815 references
  - Ab type: gp120 CD4BS – 438 records
  - Search for 10E8

Can show only notes and references containing selected keywords or user’s text.
Found 30 matching records:

Displaying record number 2708

**Mab ID**: 10E8

**HXB2 Location**: gp160(671-683)

**DNA**: 8235..8273

**Author Location**

**Epitope**: NWFDISNWLYIK

**Subtype**: B

**Ab Type**: gp41 MPER (membrane proximal external region)

**Neutralizing**: P (tier 2)

**Contexts and Features**: Search for contexts and features

**Species (Isotype)**: human(IgG3)

**Patient**: Donor N152

**Immunogen**: HIV-1 Infection

**Keywords**: ADCC, antibody binding site, antibody gene transfer, antibody generation, antibody lineage, antibody sequence, binding affinity, bispecific molecule, broad neutralizer, chimeric antibody, computational epitope prediction, contact residues, glycosylation, immunophrophylaxis, immunotherapy, neutralization, review, structure, subtype comparisons, vaccine antigen design, vaccine-induced immune responses, variant cross-reactivity

Notes

Showing 44 of 44 notes.

- **10E8**: Next generation of a computational neutralization fingerprinting (NFP) as a way to predict polyclonal Ab responses to HIV infection is presented. A new panel of 20 pseudoviruses, termed f61, was developed to aid in the assessment of experimental neutralization. This panel was used to assess 22 well-characterized bNAbs and mixtures thereof (HJ16, VRC01, BANC195, IgG1b12, PGT121, PGT128, PGT135, PG9, PGT151, 35O22, 10E8, 2F5, 4E10, VRC27, VRC-CH31, VRC-PG20, PG04, VRC23, 12A12, 38NC117, PGT145, CH01). The new algorithms accurately predicted VRC01-like and PG9-like antibody specificities.

  *Doria-Rose 2017* (neutralization, computational epitope prediction)

- **10E8**: The amino acid at gp120 position 375 is embedded in the Phe43 cavity, which affects susceptibility to ADCC. Most M-group strains of HIV-1 have serine at position 375, but CRF01 typically has histidine, which is a bulky residue. Mabs 2G12 and 10E8 were not affected by changes in residue 375, while recognition by CD4i mAbs 17b and A32 was increased by mutations of residue 375 to histidine or tryptophan. Participants in the AIDSVAX vaccine trial were infected by CRF01, and a significant part of the efficacy of this vaccine rested on ADCC responses. The ADCC response of MAbs derived from AIDSVAX participants (CH29, CH38, CH40, CH51, CH52, CH54, CH77, CH80, CH81, CH89, CH91, CH94) was dependent on the presence of 375H and greatly decreased by the presence of 375S.

  *Prevost 2017* (ADCC, vaccine-induced immune responses)
## 10E8 Donor

### Patient Detail

<table>
<thead>
<tr>
<th>Databases</th>
<th>Search</th>
<th>Tools</th>
<th>Products</th>
<th>Publications</th>
</tr>
</thead>
</table>

### Databases

**Patient Code**: Donor N152  
**Patient Sex**: Male  
**Risk Factor**:  
**Infection Country**:  
**Infection City**:  
**Infection Year**:  
**HLA Type**:  
**Patient Ethnicity**:  
**Progression**: Slow progressor (SP)  
**Species**: human  
**Patient Note**: At time of leukapheresis for MAb 10E8, patient had been infected with HIV-1 for 20 years; infected with clade B; selected because his serum neutralizing activity was among the most potent and broad in the cohort.  
**CTL CD8+ Records**:  
**T-Helper CD4+ Records**:  
**Antibody Records**: 2708, 2709, 3072, 3073, 3074, 3075, 3076, 3077, 3078, 3079, 3080, 3081, 3147, 3148, 3149, 3150, 3151, 3152, 3153, 3154, 3471, 3472, 3477, 3518, 3527  
**Sequence Database Patient ID Record**: 75177

[Link to patient’s HIV sequences]
## Neutralizing Antibody Contexts & Features

**Purpose:** To provide exact coordinates of known neutralizing antibody binding sites and other HIV-1 Env features. The data are also summarized in a [spreadsheet (.xls)](https://www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp). For details, see [Help](https://www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp).

### MAb name
- 10-1074
- 10-996
- **10E8**
- 12A12

### Antibody class
- CD4bs
- CD4i
- CH4bs
- glycan

### Env AA position
- 315,323

### Type
- Antibody related feature
- Other Env feature

### Reference
- Andrab2015
- Balli-Jaggoorsingh2013
- Bhiman2015
- Blattner2014

### Database ID
- 1
- 2
- 3
- 4

### View Neutralizing Antibody Contexts & Features

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Antibody class</th>
<th>Reference</th>
<th>MAb name</th>
<th>Env pos.</th>
<th>Feature</th>
<th>HXB2 AA</th>
<th>Entropy Group M</th>
<th>Entropy Subtype B</th>
<th>Entropy Subtype C</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>10E8 contacts</td>
<td>MPER</td>
<td>Huang2012a</td>
<td>10E8</td>
<td>671</td>
<td>gp41</td>
<td>N</td>
<td>0.779</td>
<td>0.669</td>
<td>0.895</td>
<td>10E8 N671 structure and neutralization: key epitope position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>672</td>
<td>gp41</td>
<td>W</td>
<td>0.017</td>
<td>0.023</td>
<td>0.014</td>
<td>10E8 W672 structure and neutralization: key epitope position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>673</td>
<td>gp41</td>
<td>F</td>
<td>0.058</td>
<td>0.065</td>
<td>0.073</td>
<td>10E8 F673 structure and neutralization: key epitope position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>676</td>
<td>gp41</td>
<td>T</td>
<td>0.603</td>
<td>0.610</td>
<td>0.674</td>
<td>10E8 T676 structure and binding: key epitope position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>660</td>
<td>gp41, gp41 transmembrane</td>
<td>W</td>
<td>0.069</td>
<td>0.083</td>
<td>0.081</td>
<td>10E8 W660 structure and neutralization: key epitope position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>663</td>
<td>gp41, gp41 transmembrane</td>
<td>K</td>
<td>0.577</td>
<td>0.499</td>
<td>0.559</td>
<td>10E8 K/R663 structure: key epitope position.</td>
</tr>
</tbody>
</table>

Important position(s) with Hxb2 amino acid: N671, W672, F673, T676, W680, K683

---

[Los Alamos National Laboratory](https://www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp)
10E8 Neutralization information in CATNAP

- Expand the table to show heavy and light chain sequences and sources for germline data.

Link to structure in PDB.
CATNAP
Compile, Analyze and Tally NAb Panels

Purpose:

• To compile published data on HIV NAbS and their neutralization data.

• To integrate and juxtapose on one screen neutralization data (or any numerical data) and viral sequence data.

• To explore potential genetic signatures associated with HIV neutralization based on either published or your own data.

• To find potential genetic signatures in any kind of numerical data associated with sequences.

• With input from Anthony West (West et al, PNAS 2013).

• Designed by Hyejin Yoon, Jennifer Macke, Bette Korber, Karina Yusim

https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/index.html
The CATNAP family of tools has been designed to facilitate the analysis of neutralizing antibodies (NAb) through the identification of potential genetic signatures resulting from a NAb’s interaction with a protein. While interactions between NAb and HIV-1 Env are the emphasis, the Custom Input version can accommodate other types of data, including other proteins and organisms.

**Custom Input requires**
- Numerical data (IC50, ID50, AUC, any phenotypic data)
- Aligned sequences associated with the data

**You can also combine your own HIV data with the published HIV data (Hybrid CATNAP)**

**CATNAP**

Purpose: Analyze our database of HIV-1 IC50 and IC90 neutralization data from publicly-available sources, in conjunction with HIV-1 Envelope sequences. Access our extensive databases of information about neutralizing antibodies and viruses used in published neutralization studies. Alignments of Env sequences for these viruses are also provided.

Help: [CATNAP Help](http://hiv.lanl.gov/catnap).

**CATNAP: Custom Input**

Purpose: Find potential genetic signatures based on your own numerical data in association with protein sequences. In addition to neutralization data, this tool is flexible enough to accommodate almost any kind of data in conjunction with almost any protein sequence.

Help: [Custom CATNAP Help](http://hiv.lanl.gov/catnap).

**CATNAP: Hybrid**

Purpose: Compare and analyze your HIV-1 IC50 and IC90 neutralization data with published data. This tool will display your data side-by-side with data from our database of published HIV-1 neutralization data.

Help: [Hybrid CATNAP Help](http://hiv.lanl.gov/catnap).

**Reference**

Both DATABASE and ANALYSIS:

- Database of all the published IC_{50} and IC_{80} assays we can find (110 currently)

- Data:
  - Antibody data (>300): donor ID, links to Immuno DB, PDB structures, germline, binding type, etc.
  - Aligned virus data (>1000): subtype, accession, neutralization tier, virus name aliases, patient health status, various viral panels, etc.
  - Information about Env positions: entropy, functional domain, Ab contacts and signature predictions

- Analysis:
  - Env sequence data side-by-side with IC_{50}/IC_{80} values
  - AA composition, N-glycosylation sites, basic statistics
  - Antibody potency and breadth summarized over multiple studies
  - Amino acid associations with neutralization.
  - Links to other analysis tools
Select Antibodies and Viruses in Several Ways:

- Individual or all antibody and viruses
Select Antibodies and Viruses in Several Ways:

- Individual or all antibody and viruses
- Select by study
Select Antibodies and Viruses in Several Ways:

- Individual or all antibody and viruses
- Select by study
- Select antibodies by attributes (germline and binding region)
- Select viruses by attributes (Tier, Subtype, Infection stage)
Select Antibodies and Viruses in Several Ways:

- Individual or all antibody and viruses
- Select by study
- Select antibodies by attributes (germline and binding region)
- Select viruses by attributes (Tier, Subtype, Infection stage)
- Select viruses by a virus panel
CATNAP
Compile, Analyze and Tally NAb Panels

Purpose: To provide easy analysis of data associated with HIV-1 neutralizing antibodies, including neutralization panel data, sequences, and structures.

See also: Help | Other CATNAP tools | How to Cite

New! Click "Attributes" to select antibodies based on donor, germline genes, or binding type. Or select viruses based on tier, subtype, infection stage, or coreceptor. Details...

Select by Antibody and Virus Study

Example: 10E8 and PG9

Retrieve Antibody, Virus or Assay details

Analyze IC$_{50}$, IC$_{80}$ or Both along with the viral sequences
### 10E8 Neutralization information in CATNAP

**Antibody information**

Number of antibodies: 1

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Antibody binding type</th>
<th>Structure</th>
<th>Donor</th>
<th>Clonal lineage</th>
<th>Isolation paper</th>
<th>Neutralizing antibody feature</th>
<th>Heavy V (IGHV)</th>
<th>Heavy D (IGHD)</th>
<th>Heavy J (IGHJ)</th>
<th>Light V (IGKV or IGLV)</th>
<th>Light J (IGKJ or IGLJ)</th>
<th>Light chain type</th>
<th>Genetic signature analysis</th>
<th>LANL comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>10E8</td>
<td>o C-term</td>
<td>4U6G</td>
<td>Donor N152</td>
<td>10E8 Huang2012a</td>
<td>10E8 contacts 10E8 residue prediction 10E8 signature predictions (West2013)</td>
<td>3-15<em>05 3-3</em>01 1*01</td>
<td>3-19<em>01 3</em>02 L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o gp41 MPER</td>
<td>5I0Q 5I0Q</td>
<td>4G6F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Assay**

Analyze assay data in CATNAP

Number of data: 1551

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Virus</th>
<th>Reference</th>
<th>IC50</th>
<th>Mean IC50</th>
<th>IC80</th>
<th>Mean IC80</th>
</tr>
</thead>
<tbody>
<tr>
<td>10E8</td>
<td>0G13095_2_11</td>
<td>Asokan et al. J Virol 89:12501 (2015)</td>
<td>0.002</td>
<td>0.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chuang et al. J Virol. 87:10047 (2013)</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Doria-Rose et al. J Virol. 90:76 (2016)</td>
<td>0.00200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huang et al. Immunity 45:1108 (2016a) dataset 1</td>
<td>0.003</td>
<td>0.05800</td>
<td>0.07723</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kong et al. J Virol 89:2659 (2015) dataset 1</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kong et al. J Virol 89:2659 (2015) dataset 2</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Link to structure in PDB
**CATNAP:** Virus info (in addition to the Ab and assay info)

Automatically submit all selected sequences in a batch to the HIV sequence search interface.

### More Info in HIV Sequence DB

<table>
<thead>
<tr>
<th>Virus name</th>
<th>Subtype</th>
<th>Country</th>
<th>Patient health</th>
<th>Days post infection</th>
<th>Fiebig</th>
<th>Risk factor</th>
<th>Accession</th>
<th>Tier</th>
<th>Alias</th>
<th>HIV DB name</th>
<th>Seq data</th>
<th>LANL comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>216_F2_E3_5</td>
<td>A1C</td>
<td>TANZANIA</td>
<td></td>
<td>6</td>
<td>Heterosexual</td>
<td>HM215277</td>
<td>216_F2_E3_5</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>231965_C1</td>
<td>D</td>
<td>UGANDA</td>
<td>acute infection</td>
<td>early</td>
<td>1 or 2</td>
<td>JQ361079</td>
<td>231965</td>
<td>Yes</td>
<td>231965_c01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>231966_C2</td>
<td>D</td>
<td>UGANDA</td>
<td>acute infection</td>
<td>early</td>
<td>1 or 2</td>
<td>JK512899</td>
<td>231966</td>
<td>Yes</td>
<td>231966_c02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234_F1_16_57</td>
<td>C</td>
<td>TANZANIA</td>
<td></td>
<td>5</td>
<td>Heterosexual</td>
<td>HM215278</td>
<td>234_F1_16_57</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>235_47</td>
<td>02_AG</td>
<td>CAMEROON</td>
<td></td>
<td>6</td>
<td>Not Recorded</td>
<td>EUS13195</td>
<td>235_47</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>242_14</td>
<td>02A1</td>
<td>CAMEROON</td>
<td></td>
<td>6</td>
<td></td>
<td>EUS13195</td>
<td>242_14</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>246_F3_C10_2</td>
<td>AC</td>
<td>TANZANIA</td>
<td></td>
<td>6</td>
<td>Heterosexual</td>
<td>HM215279</td>
<td>246_F3_C10_2</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>246F_C1G</td>
<td>C</td>
<td>ZAMBIA</td>
<td>acute infection</td>
<td>early</td>
<td>2</td>
<td>F4J96194</td>
<td>ZM246</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>247_23</td>
<td>DU</td>
<td>CAMEROON</td>
<td></td>
<td>Not Recorded</td>
<td></td>
<td>EU63891</td>
<td>247_23</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>249M_B10</td>
<td>C</td>
<td>ZAMBIA</td>
<td>acute infection</td>
<td>early</td>
<td>Heterosexual</td>
<td>F4J96182</td>
<td>249M_080503_SGA_B10</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25710_2_43</td>
<td>C</td>
<td>INDIA</td>
<td>acute infection</td>
<td>45</td>
<td>5</td>
<td>FF117271</td>
<td>25710_2</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25711_2_4</td>
<td>C</td>
<td>INDIA</td>
<td>acute infection</td>
<td>45</td>
<td>3</td>
<td>FF117272</td>
<td>25711_2</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25925_2_22</td>
<td>C</td>
<td>INDIA</td>
<td>acute infection</td>
<td>45</td>
<td>3</td>
<td>FF117273</td>
<td>25925_2</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26191_2_48</td>
<td>C</td>
<td>INDIA</td>
<td>acute infection</td>
<td>45</td>
<td>3</td>
<td>FF117274</td>
<td>26191_2</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence does not match accession. This sequence/clone was the one used in neutralization studies but it has not yet been deposited in GenBank.

Link to the sequence record in the HIV Sequence DB.
### CATNAP: IC₅₀ & IC₈₀/HIV-1 alignment

<table>
<thead>
<tr>
<th>Virus name</th>
<th>Tier</th>
<th>10E IC₅₀</th>
<th>10E IC₈₀</th>
<th>PGG IC₅₀</th>
<th>PGG IC₈₀</th>
<th>PG1</th>
</tr>
</thead>
<tbody>
<tr>
<td>001428_2_42</td>
<td>2</td>
<td>1.4106*</td>
<td>6.4438*</td>
<td>0.1309*</td>
<td>0.3327*</td>
<td>0.25</td>
</tr>
<tr>
<td>0041_V3_C18</td>
<td>2</td>
<td>1.4593*</td>
<td>8.1928*</td>
<td>0.1204*</td>
<td>0.3501*</td>
<td>1.6</td>
</tr>
<tr>
<td>0077_V1_C16</td>
<td>2</td>
<td>0.5359*</td>
<td>1.0470*</td>
<td>0.0375*</td>
<td>0.1370*</td>
<td>0.1</td>
</tr>
<tr>
<td>00836_2_5</td>
<td>1B or 2</td>
<td>1.8018*</td>
<td>1.9079*</td>
<td>0.0800*</td>
<td>0.1779*</td>
<td>0.8</td>
</tr>
<tr>
<td>0260_V5_C36</td>
<td>2</td>
<td>1.0467*</td>
<td>1.8346*</td>
<td>0.0160*</td>
<td>0.2254*</td>
<td>0.1</td>
</tr>
<tr>
<td>0301_BM_A2</td>
<td>2</td>
<td>0.3135*</td>
<td>1.0682*</td>
<td>0.4206*</td>
<td>0.6392*</td>
<td>0.2</td>
</tr>
<tr>
<td>0301_BM_A6</td>
<td>2</td>
<td>1.1592*</td>
<td>4.6987*</td>
<td>UD: &gt;50*</td>
<td>UD: &gt;50*</td>
<td>UD:</td>
</tr>
<tr>
<td>0330_V4_C3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0404_BM_B9</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0404_BM_D4</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0404_BM_F3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0404_BM_G3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0404_BM_H4</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0439_V5_C1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0702_BM_B4</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0702_BM_B9</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0702_BM_D1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0815_V3_C3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0907_V4_C12</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Geometric mean of detected**

- Geometric mean of all (undetected set to 100)

**% detected (detected/total)**

- 95% (411/432)
- 97% (387/400)
- 74% (543/729)
- 69% (285/416)
- 64%

### Antibody context and feature position(s) (based on HXB2)

**See Spreadsheet of neutralizing antibody contexts and features (xlsx) for more information.**

- **10E Contacts** (LogoOnly, LogoOnlySubtype): N671 W672 F673 T676 W680 K683
- **PGP-like Contacts** (LogoOnly, LogoOnlySubtype): N156 H160 I165 G167 L168 V191 Q170 K171 Y173

### Position analysis

- **Analyze** HXB2 position 100 for Ab PG9

### Run CombiNaber

- **Submit**

**# of viruses tested**

- 10E IC₅₀: 432
- 10E IC₈₀: 400
- PGP IC₅₀: 729
- PGP IC₈₀: 416
- PGT121 IC₅₀: 634
- PGT121 IC₈₀: 393
- VRC01 IC₅₀: 781
- VRC01 IC₈₀: 444

388 virus(es) tested against all antibodies retrieved will be submitted to CombiNaber.
Note: The new Genetic Signature Tool calculating phylogenetically corrected signatures will be linked soon to CATNAP (pending submitted publication).
CombiNABer

Our newest tool, designed by Kshitij Wagh, Hyejin Yoon, Bette Korber

Background
- Kong et al, 2015, J Virol
- Wagh et al, 2016, PLOS Pathogens
- Questions: Kshitij Wagh, kshitij@lanl.gov

Purpose: predict neutralization by antibody combinations (to optimize immunotherapy options)

Input:
- Neutralization data (IC50 and/or IC80) with antibody and virus names
- Antibody type (i.e. binding region)

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html
Single mAbs
**CombiNAbber**

**Single mAbs**

**2-mAb combinations**
CombinaBere

Single mAbs

3-mAb combinations
CombiNABer

Single mAbs

4-mAb combinations
CombiNABer

Single mAbs


Best 4, 3, 2, 1 Combinations
HIV Genome Browser:

• A customization of JBrowse Genome Browser, built to incorporate many sources of information from our Sequence and Immunology databases.

• A one-stop source of information about HIV genome and immunological data. It retrieves the vast and diverse information available at HIV Immunology database and allow to look at the whole HIV genome and zoom in to a region of interest and see all information we have in the database about this region:
  - HXB2 gene map, HXB2 sub-protein map, Mac239 map
  - Overlapping epitopes, antibody binding sites
  - HXB2 coding sites of interest (e.g. functional domains, drug resistance sites, motifs, glycosylation sites, etc.)
  - HXB2 LTR sites of interest (RNA structural elements, primer binding sites, etc.)
  - Neutralizing Ab contact residues, signatures and other NAb-associated features
  - HIV sequence variability (Entropy: M group, B clade, C clade)
  - Links to the database annotation, alignments, tools, PubMed etc.

• DNA- and Protein-level views are available

• Dreamt of by Christian Brander;
• Implemented by Shihai Feng;
• Help from Jennifer Macke, Brian Foley, Jim Szinger, Karina Yusim
HIV Genome Browser: Nucleotide view

Right click to switch to protein view

Navigation

Your position
The HIV database sequence analysis tool set

All tools can be accessed from the HIV sequence database

Click top level to link to full page of tools, where all >60 computational analysis tools are organized in groups by function/purpose.

Most tools have explanation pages, and sample data sets.

Many tools were inspired by user comments — please ask for more!

---

<table>
<thead>
<tr>
<th><strong>Tooltags</strong></th>
<th><strong>Tools</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Index of all tools</td>
<td>Genomic Browser</td>
</tr>
<tr>
<td>Alignment Slicer</td>
<td>Heatmap</td>
</tr>
<tr>
<td>AnalyzeAlign</td>
<td>Hepsite</td>
</tr>
<tr>
<td>AnnotateTree</td>
<td>Highlighter</td>
</tr>
<tr>
<td>BranchingTree</td>
<td>HIV BLAST</td>
</tr>
<tr>
<td>CATMAP</td>
<td>HIVAlign</td>
</tr>
</tbody>
</table>

**Programs and Tools**

- **Search Interface** retrieves HIV and SIV sequences, which can be aligned and used to build trees.
- **Geography Search Interface** retrieves HIV sequences based on geographical distribution.
- **Genome Browser** uses JBrowse to display diverse data about the genome and proteome.
- **Tools for working with sequences** lists all our online tools, grouped by function.

**Alignments**

- **HIV Premade Alignments** includes Consensus and Ancestral subtype reference alignments, and complete alignments.

An HIV Database workshop presented by Los Alamos National Laboratory staff will be included as part of the IEDB User Workshop in October 2016 in San Diego, CA. For information and registration, see [http://workshop.iedb.org](http://workshop.iedb.org).
HIV Immunology Tools are a subset of the HIV Sequence Tools

www.hiv.lanl.gov/content/immunology/tools-links.html

Tools especially useful from immunologists can be accessed from the HIV Immunology “Tools” page
HIV/SIV Sequence Locator Tool

- Calculates DNA or protein fragment location relative to a reference strain
  - Available for HIV-1, SIV, HCV, and similar tools exit in HFV database
  - Such numbers, often included in the literature, are frequently incorrect

Paste or type a DNA or protein sequence here.

Location in genome mapped in red.

Table of protein regions touched by query sequence. AA = amino acid, NA = nucleic acid.

<table>
<thead>
<tr>
<th>CDS</th>
<th>AA position relative to protein start in HXB2</th>
<th>AA position relative to query sequence start</th>
<th>AA position relative to polyprotein start in HXB2</th>
<th>NA position relative to CDS start in HXB2</th>
<th>NA position relative to HXB2 genome start</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gag</td>
<td>77 → 85</td>
<td>1 → 9</td>
<td>NA</td>
<td>229 → 255</td>
<td>1018 → 1044</td>
</tr>
<tr>
<td>p17</td>
<td>77 → 85</td>
<td>1 → 9</td>
<td>NA</td>
<td>229 → 255</td>
<td>1018 → 1044</td>
</tr>
</tbody>
</table>

Alignment of the query sequence to HXB2 (Similarity 100.0%):

Query SLYNTVATL 9

http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html
HIV/SIV Sequence Locator Tool

- Can also retrieve reference sequences
  - by coordinates (range of base or amino-acid positions)
  - by single position (retrieves flanking sequences)

--- OR ---

Retrieve a region by its coordinates

Include surrounding region

50 aa long stretch

http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html
PepMap  www.hiv.lanl.gov/content/sequence/PepMap/pepmap.html

- Maps an input set of peptides on the query sequence
- Can be used to map epitopes, functional domains, or any protein region of interest
- Peptide name can contain any kind of useful information

Input:

Peptide1: MGGKWSASSVIGGPTV
Peptide2: WSKSSVIGWVTY
HLA-A2: RMRRAEPAV
B8: ADRVGAASRDLEK
Reactive: ADRVGAASRDLEKHAI
Not_reactive: LEKHGAIITSSNTA

Peptide map (FASTA)

>Peptide1
MGGKWSASSVIGGPTV-----------------------------------
>Peptide2
WSKSSVIGWVTY-----------------------------------
>HLA-A2
RMRRAEPAV-----------------------------------
>B8
ADRVAASRDLEK-----------------------------------
>Reactive
ADRVAASRDLEKHAI-----------------------------------
>Not_reactive
LEKHGAIITSSNTA-----------------------------------

Peptide map (PDF)

Location table

<table>
<thead>
<tr>
<th>Epitope Name</th>
<th>Query Peptide</th>
<th>Reference Peptide</th>
<th>Protein</th>
<th>AA position in Protein</th>
<th>Polypeptide</th>
<th>AA position in Polypeptide</th>
<th>Similarity%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide1</td>
<td>MGGKWSASSVIGGPTV</td>
<td>MGGKWSASSVIGGPTV</td>
<td>Nef</td>
<td>1-16</td>
<td>-</td>
<td>-</td>
<td>87.5</td>
</tr>
<tr>
<td>Peptide2</td>
<td>WSKSSVIGWVTY</td>
<td>WSKSSVIGWVTY</td>
<td>Nef</td>
<td>5-16</td>
<td>-</td>
<td>-</td>
<td>91.7</td>
</tr>
<tr>
<td>HLA-A2</td>
<td>RMRRAEPAV</td>
<td>RMRRAEPAA</td>
<td>Nef</td>
<td>19-27</td>
<td>-</td>
<td>-</td>
<td>88.9</td>
</tr>
<tr>
<td>B8</td>
<td>ADRVGAASRDLEK</td>
<td>ADRVGAASRDLEK</td>
<td>Nef</td>
<td>26-39</td>
<td>-</td>
<td>-</td>
<td>100.0</td>
</tr>
<tr>
<td>Reactive</td>
<td>ADRVGAASRDLEKHAI</td>
<td>ADRVGAASRDLEKHAI</td>
<td>Nef</td>
<td>27-43</td>
<td>-</td>
<td>-</td>
<td>100.0</td>
</tr>
<tr>
<td>Not_reactive</td>
<td>LEKHGAIITSSNTA</td>
<td>LEKHGAIITSSNTA</td>
<td>Nef</td>
<td>37-49</td>
<td>-</td>
<td>-</td>
<td>100.0</td>
</tr>
</tbody>
</table>
PeptGen

- Generates overlapping peptides for any protein sequence
- Takes alignment as an input and removes duplicate peptides

 Seq1  HIVWASRELFAVNPGGLLETSEGCRQILGQLPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEENQNSK
 Seq2  HLWASRELFAVPNPGGLLETSEGCKIIKQLQPALQTGTEELRSLYNTVATLYCVHEKIEVRDTKEALDKIEEENQNSQ
 Seq3  HLWASRELFAVPNPDLETAEGCQQIMQQLQPALQTGTEELRSLFNTVATLYCVHQRIEVKDTKEALEEEKQKKSQ
**QuickAlign**

- Aligns query sequence to an alignment, creates WebLogos, calculates frequency by position, tallies variants in an alignment
- Can be used to align epitopes, functional domains, or any protein or any region of interest
- Shows results by groupings (subtypes for example) and all groups together

### Variant frequency summary

<table>
<thead>
<tr>
<th>Variant</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>--F-------</td>
<td>11</td>
<td>47.83%</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>30.43%</td>
</tr>
<tr>
<td>--F--I--V--</td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td>--F--V--</td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td>--L-------</td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td>--P--A--V--</td>
<td>1</td>
<td>4.35%</td>
</tr>
</tbody>
</table>

Total sequences = 23
Number of variants = 7

### Frequency by position

<table>
<thead>
<tr>
<th>Position</th>
<th>Percentage and raw count of non-gap (percentage)</th>
<th>Non-gap/total (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S: 99.90% (3113) other: 0.10% (3)</td>
<td>3116/3119 (100.00%)</td>
</tr>
<tr>
<td>2</td>
<td>L: 98.90% (368) other: 1.10% (4)</td>
<td>3102/3119 (99.55%)</td>
</tr>
<tr>
<td>3</td>
<td>Y: 52.71% (1633) other: 47.29% (1536)</td>
<td>3098/3119 (99.42%)</td>
</tr>
<tr>
<td>4</td>
<td>N: 99.68% (3104) other: 0.32% (10)</td>
<td>3114/3119 (99.94%)</td>
</tr>
<tr>
<td>5</td>
<td>T: 92.86% (2887) A: 5.05% (157) other: 2.09% (65)</td>
<td>3109/3119 (99.78%)</td>
</tr>
<tr>
<td>6</td>
<td>V: 79.35% (2448) I: 18.15% (560) other: 2.50% (77)</td>
<td>3085/3119 (99.01%)</td>
</tr>
<tr>
<td>7</td>
<td>A: 92.95% (2889) V: 6.53% (203) other: 0.51% (16)</td>
<td>3108/3119 (99.74%)</td>
</tr>
<tr>
<td>8</td>
<td>T: 72.52% (2254) V: 27.06% (841) other: 0.42% (13)</td>
<td>3083/3119 (99.74%)</td>
</tr>
<tr>
<td>9</td>
<td>L: 99.00% (3078) other: 1.00% (31)</td>
<td>3109/3119 (99.78%)</td>
</tr>
</tbody>
</table>
AnalyzeAlign

- New tool similar to QuickAlign, but takes sequence positions/range (including discontinuous) to analyze in an alignment
- Has many analysing options:
  - WebLogo specifications
  - Frequency cutoffs
  - Choice of the master sequence to find variants
  - User-specified color scheme
  - Combining multiple logos on a page
  - Showing potential N-linked glycosylation sites (Nx[ST], denoted as 0)

MAb PG9 binding regions, Env 156-173, bNAb PG9 contact region
Transmitted HIV virus

- Longitudinal samples over time
- Discontinuous positions under apparent immune selection
- Only differences from transmitted virus are shown
- Colors indicate amino acid charge categories

Virus 3 years post-infection

Figure from Hraber et al. *Viruses* 2015
Heatmap

- Two-dimensional clustering analysis
- A graphical way of displaying a table of numbers by using colors to represent numerical values.
- Strategy borrowed from the gene expression array literature to organize and visualize neutralization data, but is also useful for other complex data

**Example**: (Bricault et al, 2015, J Virol)

- **Rows**: ID50s in guinea pigs vaccinated by 4 different strains and combinations of 2, 3, and 4 strains (2C, 3C, 4C)
- **Columns**: Tier 1A, 1B test Envs
- Higher intensity color – higher ID50s
- Vertical bar – animals colored by vaccine
- Horizontal bar – Envs colored by neutralization tier
- Animals vaccinated by 4 strains in combination cluster together on the top and have highest ID50s
Epitopes up to 14 aa long are mapped on HXB2
- HXB2 sequence may differ
- Epitopes with identical boundaries and HLA fields are included in the maps only once
- The epitope maps are interactive!
  - Clicking on an epitope leads to the epitope entry

www.hiv.lanl.gov/content/immunology/maps/maps.html
**CTL/CD8+ Epitope Summary (B-list)**

- A comprehensive list of all unique epitopes in the database (including with unknown HLA, boundaries not fully defined...)
- Similar lists for Helper epitopes and linear Ab binding sites
- Unlike epitope maps that show epitope locations, each epitope sequence is shown

![Epitope Table](image1)

**Best-defined CTL/CD8+ Epitope Summary (A-list)**

- Experimentally validated optimal epitopes with known HLA presenting molecules
- Defined/curated by Christian Brander and colleagues

![Epitope Table](image2)

**Epitope variants and escape mutations**

- Experimental epitope variants from the literature
  - Search interfaces
  - Summary tables (~3500 CTL epitope variants)
- HLA associated HIV polymorphisms (Zabrina Brumme, Bruce Walker)
  - Database review and a table

![Image](image3)

www.hiv.lanl.gov/content/immunology/tables/ctl_summary.html

www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html

www.hiv.lanl.gov/content/immunology/pdf/2010/escape_article_supplement.html
### CTL/CD8+ Search

[www.hiv.lanl.gov/content/immunology/ctl_search](www.hiv.lanl.gov/content/immunology/ctl_search)

- **Search by HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords**

- **Search on epitope location and find fuzzy matches, overlaps and embedded epitopes**

- **Search examples:**
  - **Example:**
    - SLYNTVATL – 285 entries
    - Narrow the search with keyword “escape” – 35 entries

Search for **ISPRTLNAW**
With the first author Pillay
### Search CTL/CD8+ T-Cell Epitope Database

**Found 1 matching record:**

**Displaying record number 53832**

**HX82 Location**  p24(15-23)

**Author Location**  Gag(147-155)

**Epitope**  ISPRTLNAW

**Subtype**  C

**Species (MHC/HLA)**  human(B57)

**Immunogen**  HIV-1 infection

**Donor MHC/HLA**  A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602, Cw*0701

**Country**  South Africa

**Experimental methods**  CD8 T-cell Elispot - IFNγ

**Keywords**  epitope processing, responses in children, mother-to-infant transmission, escape, acute/early infection

**Notes**

- HIV-specific CTLs in infants were shown to be able to select for viral escape variants early in life, despite a lack of escape with the same CTL specificity in the mother. Infant CTL responses may be compromised by transmission of escape variants that arose in the mother and also those that arose in the father, if the father was the source of the mother's infection.
- ISPRTLNAW is the C consensus form of the epitope and was the autologous form in the mother, and was transmitted to her infant. By 33 weeks a new dominant form of the epitope had emerged in the infant, mSPRRTLNW, and two additional variants had arisen, one with a substitution proximal to the epitope, pISPRTLNAW, and ISPRTLNAW.

**References**

Epitope Alignment

Also available as a separate tool QuickAlign

www.hiv.lanl.gov/content/sequence/QUICK_ALIGNv2/QuickAlign.html
### Variant Details

Displaying record number 53832

<table>
<thead>
<tr>
<th>HXB2 Location</th>
<th>p24(15-23)</th>
<th>p24 Epitope Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epitope</td>
<td>ISPRTLNAW</td>
<td>Epitope Alignment</td>
</tr>
<tr>
<td>Variants</td>
<td>mSPRLNAW</td>
<td>escape documented in this paper</td>
</tr>
<tr>
<td></td>
<td>lSPRLNAW</td>
<td>diminished response</td>
</tr>
<tr>
<td></td>
<td>pI1SPRLNAW</td>
<td>not determined</td>
</tr>
<tr>
<td>Species (MHC/HLA)</td>
<td>human(B57)</td>
<td></td>
</tr>
</tbody>
</table>

### Variant Details

Showing all 3 variants.

<table>
<thead>
<tr>
<th>Variant ID.</th>
<th>1413</th>
<th>1414</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epitope Seq.</td>
<td>ISPRTLNAW</td>
<td>ISPRTLNAW</td>
</tr>
<tr>
<td>Variant Seq.</td>
<td>mSPRLNAW</td>
<td>lSPRLNAW</td>
</tr>
<tr>
<td>Mutations</td>
<td>I/M</td>
<td>I/L</td>
</tr>
<tr>
<td>Epitope Location</td>
<td>I1M</td>
<td>I1L</td>
</tr>
<tr>
<td>HXB2 Location</td>
<td>I15M</td>
<td>I15L</td>
</tr>
<tr>
<td>Mutation Type</td>
<td>E: escape documented in this paper</td>
<td>DR: diminished response</td>
</tr>
<tr>
<td>Method</td>
<td>CD8 T-cell Elispot - IFNy, Sequence</td>
<td>CD8 T-cell Elispot - IFNy, Sequence</td>
</tr>
<tr>
<td>Note</td>
<td>This is de novo variant seen in infant by week 33 of age. The index peptide was recognized, but not the variant.</td>
<td></td>
</tr>
</tbody>
</table>

Mutation type examples:
- **E** escape
- **IE** inferred escape
- **DR** diminished response
- **SF** susceptible form
- **etc...**
**ELF** (Epitope Location Finder)

**ELF** helps identify potential T cell epitopes in a reactive peptide from a person with known HLA type by:

- Highlighting appropriate HLA anchor motifs in the peptide
- Aligning all known epitopes embedded in the peptide from the database to your query sequence, with links to epitope entries

We also have **MotifScan** tool that shows HLA binding and custom motifs on the sequence alignment.

[Image of ELF interface]

- HLA selection is synchronized between 2 analysis options
- You can choose how many top binders to show per MHC, or use a binding percentile rank cutoff.
**ELF (reported epitopes in HIV database)**

Epitopes from our CTL database aligned to your query sequence

Bold red letters indicate residues that differ from the query sequence. The symbol green square means the HLA of the epitope matches one of your submitted HLAs. Click on the epitope to see full database entry. Click on "align" to align the epitope to the sequence database via QuickAlign.

Epitopes shown here are completely within the bounds of your query. Epitopes that overlap the ends of your query are included in the "View database records" links above.

Clicking on an epitope takes you to respective CTL or Helper epitope Database entries

Clicking on the “align” button takes you to “QuickAlign” for that epitope
**ELF (predicted MHC binding)**

### Potential epitopes based on anchor residues

These peptides have C-terminal anchor residues, highlighted in **blue**, and internal anchors highlighted in **magenta**. These anchor residues match one or more motifs associated with the submitted HLA.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>HLA Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTVELDNLPGWKEPNIG</td>
<td>A*0205 .......[L]</td>
</tr>
<tr>
<td>DTVELDNL</td>
<td>A*6802 .......[TV]....[VL]</td>
</tr>
<tr>
<td>TVLEDHNL</td>
<td>A*0206 .......[Q]........</td>
</tr>
<tr>
<td>LEDHNLPCR</td>
<td>DRB5<em>0101,DRB5</em>0101 [FYLM]..[QV2M]....[RX]</td>
</tr>
</tbody>
</table>

**Motifscan**

### Potential epitopes based on IEDB binding predictions

Top binders for each MHC are highlighted in **blue**.

**Prediction method:** IEDB recommended

**Low percentile = good binders**

**Show up to 1 binder(s) per MHC**

#### Class I

Selected allele(s): A*6802, B*1501

<table>
<thead>
<tr>
<th>Peptide</th>
<th>HLA Allele</th>
<th>Binders</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTVELDNLPGWKEPNIG (Click MHC to see full list of IEDB predictions for that MHC)</td>
<td>DRB5*0101 (26)</td>
<td></td>
</tr>
<tr>
<td>MNLPGWKE</td>
<td>A*6802 (3.0)</td>
<td></td>
</tr>
</tbody>
</table>

#### Class II

Selected allele(s): DRB5*0101

<table>
<thead>
<tr>
<th>Peptide</th>
<th>HLA Allele</th>
<th>Binders</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTVELDNLPGWKEPNIG (Click MHC to see full list of IEDB predictions for that MHC)</td>
<td>DRB5*0101 (17.17)</td>
<td></td>
</tr>
</tbody>
</table>

Clicking on MHC links to the full list of IEDB predictions for that MHC (see next table)
Potential epitopes based on IEDB database MHC binding predictions

<table>
<thead>
<tr>
<th>Allele</th>
<th>#</th>
<th>Start</th>
<th>End</th>
<th>Peptide Length</th>
<th>Sequence</th>
<th>Method used</th>
<th>Percentile Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>6</td>
<td>13</td>
<td>8</td>
<td>DNLPGRW</td>
<td>NetMHCspan</td>
<td>26</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>VLEDMLPG</td>
<td>NetMHCspan</td>
<td>27</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>9</td>
<td>VLEDMLPG</td>
<td>Consensus (ANN,SMM,CombLib_Sidney2008)</td>
<td>27.60</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>8</td>
<td>17</td>
<td>10</td>
<td>NLPGRKPKM</td>
<td>NetMHCspan</td>
<td>31</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>7</td>
<td>17</td>
<td>11</td>
<td>NLPGRKPKM</td>
<td>NetMHCspan</td>
<td>35</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>TVLEDML</td>
<td>NetMHCspan</td>
<td>36</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>10</td>
<td>TVLEDMLPG</td>
<td>NetMHCspan</td>
<td>47</td>
</tr>
<tr>
<td>HLA-B*15:01</td>
<td>1</td>
<td>4</td>
<td>11</td>
<td>8</td>
<td>LEDMLPG</td>
<td>NetMHCspan</td>
<td>48</td>
</tr>
</tbody>
</table>
HIV epitopes are densely packed at the population level

- Vaccinating a diverse population with individual epitopes is infeasible
- Escape forms for one HLA are frequently sensitive for a different HLA
- It may not be necessary to predict epitopes — but only to deliver them
- Optimized immunogen cocktails could deliver most epitopes likely to be present in infecting virus
Vaccine Design Tools (Mosaic/Epigraoh)

**Design Tools**

Generate candidate vaccine protein cocktails that optimize coverage of potential T-cell epitopes (as linear $k$-mers) based on frequencies in sets of natural pathogen sequences — “all-natural” throughout, including breakpoints

**Mosaic Vaccine Designer — genetic algorithm** (Fischer et al. 2007)

**Epigraoh — graph theoretic approach** (Theiler et al. 2016)

**Evaluation tools**

**Epitope Coverage Assessment (EPICOVER)**

*Alignment-independent* “$k$-mer” coverage by vaccines or peptides.

**Positional Epitope Coverage Assessment (POSICOVER)**

*Alignment-based* coverage by vaccines or peptides.
Mosaic Vaccine Designer

Inputs

Target set: natural protein sequences from a diverse pathogen population (alignment optional).

Cocktail size: how many mosaic protein sequences to generate.

Epitope length: default is 9 amino-acids.

Method: genetic algorithm

Linear optimization: helpful for both T-cell and linear aspects of B-cell epitopes

Epitope length is transferable...

DEMONSTRATED EFFECTIVENESS

Improved immunogenicity

HIV, SIV, HCV, Chlamydia

Protection from challenge (non-human models):

SHIV, Influenza, FMDV, Ebola

Many human HIV trials in process
*k*-mer coverage is relatively stable for different values of *k* (potential epitope lengths)

In other words, optimizing for potential CD8+ T-cell epitopes (*k*=9) yields good coverage of potential CD4+ T-cell epitopes (*k*=12), too.

**EPIGRAPH**

**Inputs**

**Target set**: natural protein sequences for the pathogen population (alignment optional).

**Cocktail size**: how many mosaic proteins in the output set.

**Epitope length**: default is 9 amino-acids.

**Method**: evaluation of acyclic graph

**Advantages over mosaic**

**Essentially optimal** (fractionally better coverage)

**Much faster**: allows iteration and comparison of multiple input sets and alternate designs

EPIGRAPH — exclude rarities

HIV sequence database

Epigraph

Input

Protein sequences

Choose File: no file selected

Epigraph design job number

Options

Epitope length

# of sequences in vaccine

Email results
**Epitope Coverage Assessment - Epicover**

**Inputs:**

1. Vaccine set
2. Test set (target sequences)

Can report on subsets defined according to the first several characters in sequence names or user-defined subsets.
### Epicover output (mean coverage per sequence)

<table>
<thead>
<tr>
<th>vaccine set</th>
<th>subset</th>
<th>subset count</th>
<th>Off-by-0</th>
<th>Off-by-1</th>
<th>Off-by-2</th>
<th>rare(&lt;3,&gt;1)</th>
<th>unique</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>vaccine_set_from_user</td>
<td>Total</td>
<td>39</td>
<td>0.5037</td>
<td>0.7474</td>
<td>0.8636</td>
<td>91</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>A</td>
<td>6</td>
<td>0.4294</td>
<td>0.7086</td>
<td>0.8417</td>
<td>7</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>B</td>
<td>4</td>
<td>0.7263</td>
<td>0.8911</td>
<td>0.9460</td>
<td>44</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>C</td>
<td>4</td>
<td>0.5786</td>
<td>0.8449</td>
<td>0.9602</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>D</td>
<td>4</td>
<td>0.5764</td>
<td>0.8268</td>
<td>0.9218</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>F</td>
<td>8</td>
<td>0.4821</td>
<td>0.7316</td>
<td>0.8786</td>
<td>2</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>G</td>
<td>4</td>
<td>0.4578</td>
<td>0.7126</td>
<td>0.8367</td>
<td>5</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

Overall summaries of k-mer coverage
### Epicover output (mean coverage per sequence)

<table>
<thead>
<tr>
<th>vaccine set</th>
<th>subset</th>
<th>subset count</th>
<th>Off-by-0</th>
<th>Off-by-1</th>
<th>Off-by-2</th>
<th>rare(&lt;3,&lt;1)</th>
<th>unique</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>vaccine_set_from_user</td>
<td>Total</td>
<td>39</td>
<td>0.5037</td>
<td>0.7474</td>
<td>0.8636</td>
<td>91</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>A</td>
<td>6</td>
<td>0.4294</td>
<td>0.7086</td>
<td>0.8417</td>
<td>7</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>B</td>
<td>4</td>
<td>0.7263</td>
<td>0.8911</td>
<td>0.9460</td>
<td>44</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>C</td>
<td>4</td>
<td>0.5786</td>
<td>0.8449</td>
<td>0.9602</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>D</td>
<td>4</td>
<td>0.5764</td>
<td>0.8268</td>
<td>0.9218</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>F</td>
<td>8</td>
<td>0.4821</td>
<td>0.7316</td>
<td>0.8786</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>G</td>
<td>4</td>
<td>0.4578</td>
<td>0.7126</td>
<td>0.8367</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
</tbody>
</table>

Overall summaries of k-mer coverage

For entire set (to compare with other vaccine candidates)
Epicover output (mean coverage per sequence)

<table>
<thead>
<tr>
<th>vaccine set</th>
<th>subset</th>
<th>subset count</th>
<th>Off-by-0</th>
<th>Off-by-1</th>
<th>Off-by-2</th>
<th>rare(&lt;3,&gt;1)</th>
<th>unique</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>vaccine_set_from_user</td>
<td>Total</td>
<td>39</td>
<td>0.5037</td>
<td>0.7474</td>
<td>0.8636</td>
<td>91</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>A</td>
<td>6</td>
<td>0.4294</td>
<td>0.7086</td>
<td>0.8417</td>
<td>7</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>B</td>
<td>4</td>
<td>0.7263</td>
<td>0.8911</td>
<td>0.9460</td>
<td>44</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>C</td>
<td>4</td>
<td>0.5786</td>
<td>0.8449</td>
<td>0.9602</td>
<td>47</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>D</td>
<td>4</td>
<td>0.5764</td>
<td>0.8268</td>
<td>0.9218</td>
<td>12</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>F</td>
<td>8</td>
<td>0.4821</td>
<td>0.7316</td>
<td>0.8786</td>
<td>2</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>vaccine_set_from_user</td>
<td>G</td>
<td>4</td>
<td>0.4578</td>
<td>0.7126</td>
<td>0.8367</td>
<td>5</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

Overall summaries of k-mer coverage

For entire set (to compare with other vaccine candidates)

... or by pathogen subset
### Positional Epitope Coverage Assessment - Posicover

#### INPUTS
1. Vaccine/peptide sequences
2. ALIGNED target set

#### OUTPUTS
1. 1-dimensional (by alignment column)
2. 2-dimensional (by sequence and alignment column)

#### Options
- Nominal epitope length: 9
- Antigen counts to compute upper bounds: 3, 4

#### Plots to make
- Hits in their natural positions
- Misses in their natural positions
- Hits and misses in their natural positions
- Hits ranked by coverage
- Misses ranked by coverage
- N-mer coverage by positions
- Ranked n-mer coverage
- Alignment Thumbnail
- N-mer coverage directly on alignment
Posicover output (1-dimensional summaries)

**Matched 9-mers**

![Graph showing matched 9-mers](image1)

Alignment positions

**Missed 9-mers**

![Graph showing missed 9-mers](image2)

Alignment positions

![Graph showing 9-mer coverage](image3)

![Graph showing ranked 9-mer coverage](image4)
Pixel-based Alignment view
Each amino-acid represented as a single-colored square
Allows quick detection of gross errors in alignment
Posicover output (2 dimensional)
Posicover output (2 dimensional)

Pixel-based Alignment view
Each amino-acid represented as a single-colored square
Allows quick detection of gross errors in alignment
Posicover output (2 dimensional)
Posicover output (2 dimensional)

Posicover k-mer coverage
(Yellow-black gradient shows how many of each residue’s k-mers appear in vaccine)
Posicover output (2 dimensional)
Posicover output (2 dimensional)
Posicover output (2 dimensional)

Best natural isolate

1-sequence mosaic

unique 9-mer counts
Posicover output (2 dimensional)

Best natural isolate

2-sequence mosaic cocktail
Thank you for attending!

Please fill out our evaluation form!
Your comments will help us provide future training.

Contact us: seq-info@lanl.gov or immuno@lanl.gov
EPIGRAPH — exclude rarities

- Including *only* k-mers above an occurrence threshold drops coverage, but reducing responses to rare epitopes may be helpful.

Here, including only 9-mers that occur at least 14 times drops coverage very little.