

# HIV Database Immunology Workshop

<http://www.hiv.lanl.gov/>

[immuno@lanl.gov](mailto:immuno@lanl.gov)

[seq-info@lanl.gov](mailto:seq-info@lanl.gov)

## **Presenters:**

Elizabeth-Sharon Fung, Jennifer Macke, Kshitij Wagh, Will Fischer

**Database PI:** Brian Foley

## **Additional database staff:**

Werner Abfalterer, Katie Belobrajdic, Will Fischer, Elizabeth-Sharon Fung, Kumkum Ganguly, Jennifer Macke, James Szinger, Hyejin Yoon



**Contract Officer Representative:** Anjali Singh, NIAID, NIH

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

LA-UR-22-22582

**HIV DB Workshop slides:**

<https://hiv.lanl.gov/hws>



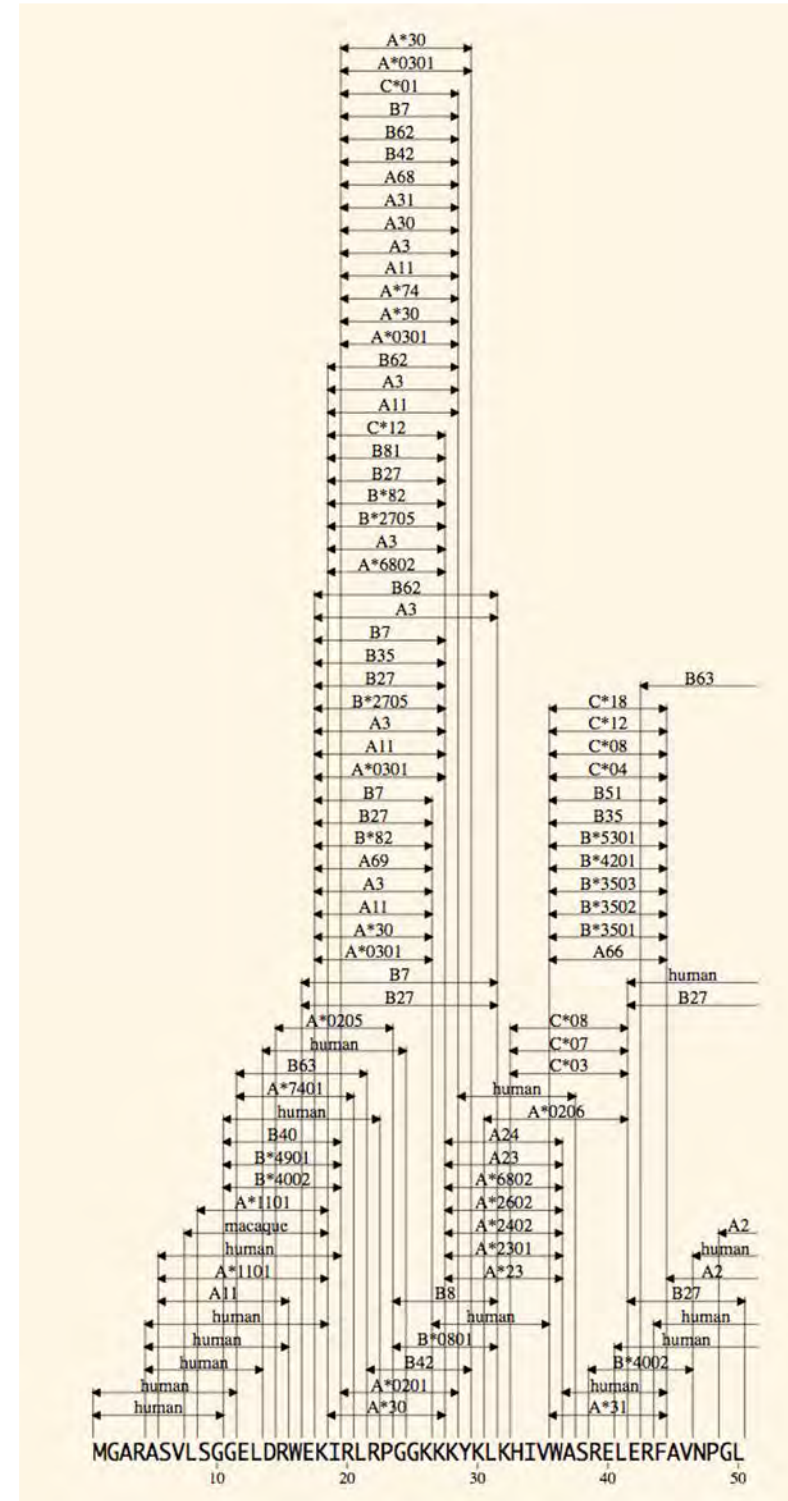
# HIV Immunology Database Workshop

- **Day 2, Keystone 2022**
  - **HIV Immunology Database**
    - Part 1:
      - HIV Immunology Database overview
      - T cell epitopes – entries and searches
      - Antibody Database – entries and searches
      - Neutralizing Antibody Resources
    - Part 2:
      - Antibody Features Database
      - Genome Browser
      - CATNAP, both tailored for HIV and applicable to any pathogen
    - Part 3:
      - CombiNAber, applicable to any pathogen
      - Glycan Shield Tool
      - GenSig
    - Part 4:
      - More computational tools for Immunologists, many applicable for any pathogen
      - Vaccine design and evaluation tools, applicable to any pathogen

# Los Alamos HIV-1 Databases

- Integrate HIV immunological and sequence data

- For the first part of March 2022:
  - 306,644 hits
  - 12,783 visits
- Citations in research articles and patents when searched on 'HIV AND Database AND LANL':
  - 2,045 patents or published applications in the US from 857 patent families (<https://www.uspto.gov/>)
  - 1,421 patents or published applications in any country (<https://www.wipo.int/>)
  - 14,200 Google Scholar citations
- HIV Sequence Database: Over 1,016,500 searchable annotated HIV/SIV sequences available as custom alignments or premade 1-sequence-per-person alignments.
- HIV Immunology Database: Searchable annotated T cell epitopes and Antibody entries.
  - Over 11,200 CD8+ epitope entries
  - Over 1,600 CD4+ epitope entries
  - Over 3,650 Antibody entries
  - Neutralization data for >500 Abs and Ab mixtures and almost 1500 pseudoviruses, most with sequences.
- Over 50 bioinformatics tools with simple web interfaces.
  - Of them many (68%) are general-purpose tools





<https://hiv.lanl.gov/>

The HIV databases contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an [Editorial Board](#).

- SEQUENCE DATABASE ▶
- IMMUNOLOGY DATABASE ▶
- OTHER VIRUSES ▶

News

[Archived News ▶](#)

[HIV Molecular Immunology 2020](#)

*HIV Molecular Immunology 2020* is now available online. The PDF version is hypertext enabled and features clickable table-of-contents, indexes, references and links to external web sites. 27 January 2022

[2020 Alignments](#)

The 2020 Web, Filtered Web, Super Filtered Web, and Consensus Alignments are now available [online](#). The curated web alignments contain a full range of sequences available through the end of 2020. New consensus sequences are available, described by [Linchangco et al. 2022](#). 24 January 2022

Questions or comments? Contact us at [seq-info@lanl.gov](mailto:seq-info@lanl.gov)



# Integration of HIV Sequence and Immunology databases

- ❑ Los Alamos HIV Database: the first pathogen-specific database
  - ❑ HIV Sequence Database – founded in 1986 by G. Myers
  - ❑ HIV Immunology Database - founded in 1994 by B. Korber
- ***What makes our database unique is the integration of HIV sequence and immunological data via multiple tools, for example:***
  - **CATNAP** superimposes Ab neutralization data with the virus data, and links to structures, germline V/D/J genes, Ab sequences, Ab contact residues, viral Env alignments, positions associated with neutralization sensitivity ...
  - **AnalyzeAlign** shows the diversity and HIV variability of epitopes
  - **HIV Genome Browser** provides an interactive detailed view of the HIV genome or proteome with HIV sequence variability, functional domains and antibody and T cell epitopes marked by genome position
  - Multiple **tools** tap into the **Patient database**, containing available donor HIV sequences, Ab sequences, monoclonal and polyclonal Ab data, HLAs, and T-cell epitopes

# Many “HIV Immunology” tools are broadly applicable

- *Tools list is color-coded by range of use*



## HIV Molecular Immunology Database: Tools & Links

### Tools Produced by the Los Alamos HIV Databases

- [CATNAP: Compile, Analyze and Tally NAb Panels](#) Download or analyze neutralization data
- [CombiNAber](#) Predict the neutralization of combinations of antibodies
- [HIV Genome Browser](#) Display HIV genome and proteome
- [QuickAlign](#) Align amino acids or nucleotides against our alignments
- [Analyze Align](#) Show weblogos, calculate frequency by position, and find variants in an alignment
- [Alignment Slicer](#) Cut vertical slices from sequence alignments
- [PeptGen](#) Generate overlapping peptides for any protein
- [PepMap](#) Generate peptide maps in Fasta, HTML and PDF formats
- [Motif Scan](#) Scan alignments for HLA binding motifs
  - [HLA genotype/serotype dictionary](#)
  - [HLA genotype/motif dictionary](#)
  - [HLA supertype dictionaries](#)
- [Hepitope](#) Search for hopeful epitopes based on HLA enrichment
- [HLA Frequency Analysis Tools](#) Calculate HLA frequencies or HLA linkage disequilibrium in a population
- [ELF](#) Epitope location finder
- [Sequence Locator Tool](#) Find the location of any HIV/SIV sequence
- [SeqPublish](#) Produce pretty alignments for publication
- [Heatmap](#) Display a table of numbers using colors to represent the numerical values
- [Epigraph Vaccine Suite](#) Design and assess Epigraphs for vaccine design
- [Mosaic Vaccine Suite](#) Design and assess polyvalent protein sequences for T-cell vaccines
- [N-Glycosite](#) Find N-linked glycosylation sites
- [Highlighter](#) Highlight matches and mismatches in a set of aligned sequences
- [Protein Feature Accent](#) View 3D graphics of HIV proteins
- [Variable Region Characteristics](#) analyzes Env variable loops and reports length, glycosolations, and net charge

- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-specific features
- General use tools

# Beyond HIV

- ❑ **Only 10 of our over 50 computational tools are strictly HIV-specific. The remaining ones either have a useful component or are fully applicable to other organisms**
- ❑ **A striking example of successful use of our tools beyond HIV is Mosaic/Epigraph vaccine design:**
  - ❑ Rabies in bats (Stading *et al*, Plos Negl Trop Dis, 2017)
  - ❑ Filoviruses (Theiler *et al*, Sci Rep. 2016, Fenimore, PLoS One, 2012)
  - ❑ Chlamydia trachomatis (Badamchi-Zadeh *et al*, Front Immunol, 2016)
  - ❑ Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in pigs (Cui *et al*, Vaccine reports, 2016)
  - ❑ Hepatitis C (Yusim *et al*, Clin Vaccine Immunol, 2013)
  - ❑ Foot-and-Mouth Disease in livestock (Devendra *et al*, 2018)
  - ❑ Hepatitis B (Yusim *et al*, in preparation)
- **The whole database structure and tools are transferable to other pathogens.** We successfully modeled several databases using the HIV database as a prototype and translating multiple tools:  
(<https://www.hiv.lanl.gov/content/otherviruses.html>):
  - **HCV Sequence** (Kuiken *et al*, Nucleic Acid Res, 2008) **and Immunology** (Yusim *et al*, Appl Bioinformatics, 2005) Databases
  - **Hemorrhagic Fever Viruses (HFV) Sequence Database** (80 viral species, found in 10 different genera comprising five different families: arena-, bunya-, flavi-, filo- and togaviridae) (Kuiken *et al*, Nucleic Acid Res, 2012)
  - **Filovirus Sequence and Immunology Database** (Yusim *et al*, Database, 2016) ([hfv.lanl.gov](http://hfv.lanl.gov))
  - **SARS COV-2 Database** (<https://cov.lanl.gov/>)
  - Lack of funding, so only the sequence portions of HCV, HFV, Filovirus databases are automatically updated

# HIV Immunology Database Entries and Annotation

- 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)
- Descriptions of HIV T cell epitopes and Antibodies with associated **data are harvested** from regular periodic literature searches:
  - Epitope sequence, location, immunogen, vaccine details, patient details...
  - Epitope Variants (escape, reduced binding, etc.)
  - Host HLA or MHC, binding region, germline genes, etc
  - Neutralizing Antibody Resources, contact residues, positions related to neutralization sensitivity or resistance, etc.
  - Notes summarizing main findings
- Multiple **search interfaces and database products**:
  - 5 search interfaces for T cell epitopes, epitope variants and antibodies
  - Computational tools for immunologists
  - Epitope maps and summary tables that can also serve as search interfaces
  - HLA typing and very large epitope mapping data sets
  - Neutralizing **antibody resources**:
    - Neutralization, germline and antibody sequence data through the CATNAP tool/database
    - Links to Germline Antibody Reconstruction tools
    - Search interface and a table for Ab contact residues, positions related to neutralization sensitivity or resistance, etc.
    - Assay protocols and neutralization serotype discovery data

# HIV Immunology Database - 2021 Additions

## Continuing Efforts

- Curated annotations
- Maintained and updated tools, maps and tables
- Published annual compendium

## Upgrades

- Expanded and searchable patient database
- HLA nomenclature updated
- JSON and CSV download capability





<https://hiv.lanl.gov/>

The HIV databases contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an [Editorial Board](#).

- SEQUENCE DATABASE ▶
- IMMUNOLOGY DATABASE ▶
- OTHER VIRUSES ▶

News

[Archived News ▶](#)

[HIV Molecular Immunology 2020](#)

*HIV Molecular Immunology 2020* is now available online. The PDF version is hypertext enabled and features clickable table-of-contents, indexes, references and links to external web sites. 27 January 2022

[2020 Alignments](#)

The 2020 Web, Filtered Web, Super Filtered Web, and Consensus Alignments are now available [online](#). The curated web alignments contain a full range of sequences available through the end of 2020. New consensus sequences are available, described by [Linchangco et al. 2022](#). 24 January 2022

Questions or comments? Contact us at [seq-info@lanl.gov](mailto:seq-info@lanl.gov)

## HIV Molecular Immunology Database

Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

- [CTL/CD8+ search](#)
- [T Helper/CD4+ search](#)
- [Antibody search](#)
- [CTL variant search](#)
- [T Helper variant search](#)
- [Patient search](#)
- [Search help](#)
- [Variant search help](#)
- [JSON API for search](#)

### Database Products

- [All Database products and publications](#)
- [Epitope maps](#)
- [Epitope tables](#)
- [Epitope alignments](#)
- [Epitope density plots](#)
- [T cell epitope variants and escape mutations](#)
- [Neutralizing antibody resources & CATNAP](#)
- [The HIV Molecular Immunology Compendium](#)
- [About the HIV Molecular Immunology Database](#)
- [How to cite this database](#)
- [Frequently-asked Questions \(FAQ\)](#)

### Tools and Data Sets

- [Tools & Links](#) for immunologists
- [SIV Epitopes \(PDF\)](#) review article summarizing known SIV epitopes
- [Identifying HLA-Associated Polymorphisms in HIV-1 \(PDF\)](#) review article summarizing HIV polymorphism associated with escape mutations. Also a [table of polymorphisms](#).
- [HLATEM](#) HLA Typing and Epitope Mapping Data Sets
- [Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development](#) Assay protocols from Duke Central Reference Laboratory

<https://www.hiv.lanl.gov/content/immunology/index.html>

**Databases**   **Search**   **Tools**   **Products**   **Publications**     

## HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an online collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

### Search Interfaces

- [CTL/CD8+ search](#)
- [T Helper/CD4+ search](#)
- [Antibody search](#)
- [CTL variant search](#)
- [T Helper variant search](#)
- [Search help](#)
- [Variant search help](#)

### Database Products

- [All Database products and publications](#)
- [Epitope maps](#)
- [Epitope tables](#)
- [Epitope alignments](#)
- [T cell epitope variants and escape mutations](#)
- [Neutralizing antibody resources & CATNAP](#)
- [The HIV Molecular Immunology Compendium](#)
- [About the HIV Molecular Immunology Database](#)
- [How to cite this database](#)
- [Frequently-asked Questions \(FAQ\)](#)

### Tools and Data Sets

- [Tools & Links](#) for immunologists
- [SIV Epitopes \(PDF\)](#) review article summarizing known SIV epitopes
- [Identifying HLA-Associated Polymorphisms in HIV-1 \(PDF\)](#) review article summarizing HIV polymorphism associated with escape mutations. Also a [table of polymorphisms](#).
- [HLATEM](#) HLA Typing and Epitope Mapping Data Sets
- [Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development](#) Assay protocols from Duke Central Reference Laboratory

**Products**

- Epitope Maps
- Epitope Tables
- Epitope Alignments
- T Cell Epitope Variants
- Neutralizing Ab Resources & CATNAP
- Data Sets: HLA Typing and Epitope Mapping
- Tools & Links

**Antibody Search**

**Multiple ways to database products and tools**

**T cell epitope variants and escape mutations**

**Neutralizing Antibody Resources**

The diagram illustrates navigation paths from the HIV Molecular Immunology Database homepage. Red arrows originate from a central box labeled 'Multiple ways to database products and tools'. One arrow points to the 'Antibody search' link in the 'Search Interfaces' section. Another arrow points to the 'T cell epitope variants and escape mutations' link in the 'Database Products' section. A third arrow points to the 'Neutralizing antibody resources & CATNAP' link in the 'Database Products' section. A fourth arrow points to the 'Tools & Links' link in the 'Products' dropdown menu. A fifth arrow points to the 'Antibody Search' section header.



# HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

## Search Interfaces

- [CTL/CD8+ search](#)
- [T Helper/CD4+ search](#)
- [Antibody search](#)
- [CTL variant search](#)
- [T Helper variant search](#)
- [Search help](#)
- [Variant search help](#)

## Database Products

- [All Database products and publications](#)
- [Epitope maps](#)
- [Epitope tables](#)
- [Epitope alignments](#)
- [T cell epitope variants and escape mutations](#)
- [Neutralizing antibody resources & CATNAP](#)
- [The HIV Molecular Immunology Compendium](#)
- [About the HIV Molecular Immunology Database](#)
- [How to cite this database](#)
- [Frequently-asked Questions \(FAQ\)](#)

## Tools and Data Sets

- [Tools & Links](#) for immunologists
- [SIV Epitopes \(PDF\)](#) review article summarizing known SIV epitopes
- [Identifying HLA-Associated Polymorphisms in HIV-1 \(PDF\)](#) review article summarizing HIV polymorphism associated with escape mutations. Also a [table of polymorphisms](#).
- [HLATEM](#) HLA Typing and Epitope Mapping Data Sets
- [Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development](#) Assay protocols from Duke Central Reference Laboratory

## Epitope Tables

These tables summarize the epitopes from our database. HIV-1 epitope data may also be obtained in the form of downloadable [maps](#) or [alignments](#).

- [CTL epitopes](#)
- [Best-defined \("A-list"\) CTL epitopes](#)
- [CTL epitope variants and escape mutations](#)
- [T-helper epitopes](#)
- [T Helper epitope variants and escape mutations](#)
- [Antibody epitopes](#)
- [Best Neutralizing Antibodies](#)
- [Antibody-Dependent Cell-Mediated Cytotoxicity \(ADCC\)](#)
- [Antibody index by name](#)
- [Antibody index by binding type](#)
- [SIV epitopes](#)
- [Neutralizing antibody resources](#)

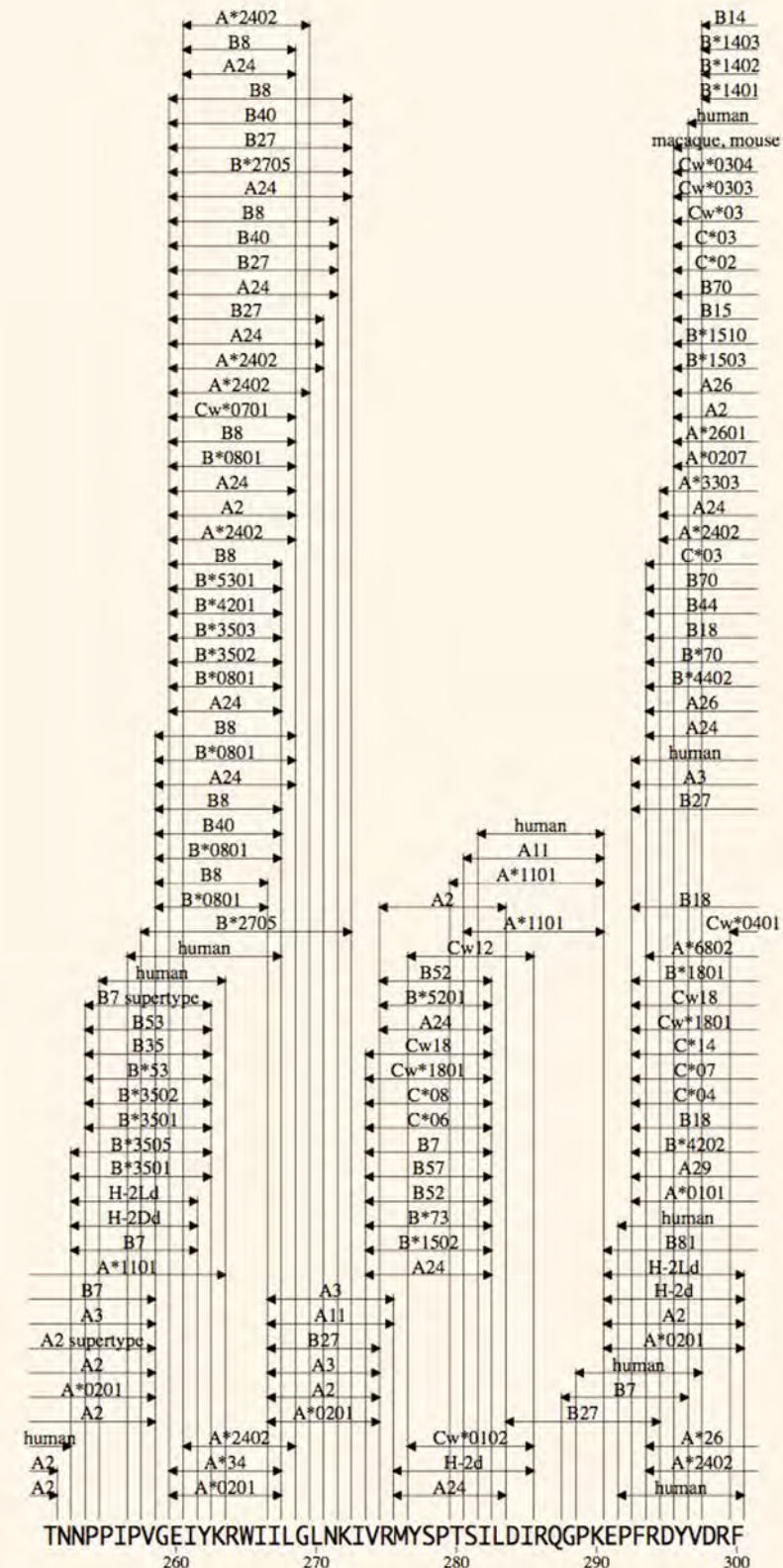
Epitope alignments: epitopes aligned to HIV subtype Reference sequences in Fasta format

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

# p17 CTL/CD8+ Epitope Map

- 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*

<http://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	Protein	HXB2 Location	Subtype	Species	HLA
<a href="#">MGARASVLSG</a>	p17	1-10	CRF01_AE	human	
<a href="#">ASVLSGGEL</a>	p17	5-13	B	human	
<a href="#">ASILRGGKLDK</a>	p17	5-15	C	human	
<a href="#">SVLSGGQLDR</a>	p17	6-15	B	human	A11
<a href="#">LSGGELDRWEK</a>	p17	8-18		macaque	
<a href="#">GELDRWEKI</a>	p17	11-19	B	human	B*4002, B40
<a href="#">GQLDRWEKI</a>	p17	11-19	B	human	
<a href="#">GKLDSWEKIRLR</a>	p17	11-22	A, CRF01_AE, CRF02_AG	human	

[www.hiv.lanl.gov/content/immunology/tables/ctl\\_summary.html](http://www.hiv.lanl.gov/content/immunology/tables/ctl_summary.html)

## 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	Protein	HXB2 Location	Subtype	Species	HLA
<a href="#">GELDRWEKI</a>	p17	11-19		human	B*4002
<a href="#">KIRLRPGGK</a>	p17	18-26		human	A*0301
<a href="#">IRLRPGGKK</a>	p17	19-27	B	human	B*2705
<a href="#">RLRPGGKKK</a>	p17	20-28		human	A*0301
<a href="#">RLRPGGKKKY</a>	p17	20-29	B	human	A*0301
<a href="#">GGKKKYKLLK</a>	p17	24-32	B	human	B*0801
<a href="#">KYKLRHIVW</a>	p17	28-36	B	human	A*2402
<a href="#">HLVWASREL</a>	p17	33-41		human	Cw*0804

[www.hiv.lanl.gov/content/immunology/tables/optimal\\_ctl\\_summary.html](http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html)

## 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

[www.hiv.lanl.gov/content/immunology/pdf/2010/escape\\_article\\_supplement.html](http://www.hiv.lanl.gov/content/immunology/pdf/2010/escape_article_supplement.html)

<a href="#">HIV protein</a>	Proteins with <a href="#">defined epitopes</a> - ALL - p17 p17-p24 p24 p24-p2p7p1p6	Proteins with <a href="#">undefined epitopes</a> - ALL - Gag Gag/Pol Pol Vif
<a href="#">HXB2 location</a>	<input type="text"/> - <input type="text"/>	Results overlap with query location
<a href="#">Epitope</a>	ISPRTLNAW	Results contain query sequence
<a href="#">Epitope name</a>	<input type="text"/>	
<a href="#">Record number</a>	<input type="text"/>	
<a href="#">Subtype</a>	- ALL -	
<a href="#">Immunogen</a>	- ALL - computer prediction HIV-1 and GBV-C co-infection HIV-1 and HCV co-infection HIV-1 exposed seronegative HIV-1 infected monocyte-derived HIV-1 infection	
<a href="#">Vaccine details</a>	<a href="#">Vaccine type</a> <a href="#">Vaccine strain</a> if Immunogen is Vaccine <a href="#">Vaccine component</a> <a href="#">Adjuvant</a>	- ALL - - ALL - - ALL - - ALL -
<a href="#">Species</a>	- ALL -	
<a href="#">MHC/HLA</a>	- ALL - A*01 A*0101 A*02 A*0201 A*02.01 A*020101	
<a href="#">Author</a>	Pillay	<input checked="" type="checkbox"/> First <input type="checkbox"/> Last
<a href="#">Country</a>	- ALL -	
<a href="#">Keywords</a>	- ALL - acute/early infection adjuvant comparison antagonism antibody binding site definition and exposure assay development, comparison, standardization, improvement autologous responses	
<a href="#">Note</a>	<input type="text"/>	

- 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 first author = Pillay**

Search

Reset

Click for [Search Help](#)



## Search CTL/CD8+ T-Cell Epitope Database

Found 1 matching record:

Displaying record number 53832

<a href="#">HXB2 Location</a>	p24(15-23)
<a href="#">Author Location</a>	Gag(147-155)
<a href="#">Epitope</a>	ISPRTLNAW
<a href="#">Subtype</a>	C
<a href="#">Species (MHC/HLA)</a>	human(B57)
<a href="#">Immunogen</a>	HIV-1 infection
<a href="#">Donor MHC/HLA</a>	A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602, Cw*0701
<a href="#">Country</a>	South Africa
<a href="#">Experimental methods</a>	CD8 T-cell Elispot - IFN $\gamma$
<a href="#">Keywords</a>	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, mSPRTLNAW, and two additional variants had arisen, one with a substitution proximal to the epitope, pISPRTLNAW, and ISPRTLNAW.

### References

**Pillay2005** Thillagavathie Pillay, Hua-Tang Zhang, Jan W. Drijfhout, Nicola Robinson, Helen Brown, Munira Khan, Jagadesa Moodley, Miriam Adhikari, Katja Pfafferott, Margaret E. Feeney, Anne St. John, Edward C. Holmes, Hoosen M. Coovadia, Paul Klenerman, Philip J. R. Goulder, and Rodney E. Phillips. Unique Acquisition of Cytotoxic T-Lymphocyte Escape Mutants in Infant Human Immunodeficiency Virus Type 1 Infection. *J. Virol.*, 79(18):12100-12105, Sep 2005. PubMed ID: [16140787](#). [Show all entries for this paper.](#)

Immunological, virological, and epidemiological contexts:

[Link to Epitope Maps](#)

[Link to Epitope Alignment](#)

[Variant details with annotator's notes](#)

[p24 Epitope Map](#)

[Epitope Alignment](#)  
[Show epitope variants](#)

Additional information provided in the entry:

- Location, Donor MHC/HLA, experimental methods, Notes
- Link to all entries for a reference
- PubMed links to papers
- Link to Epitope Maps
- Link to Epitope Alignment (aligned to large set of seq.)
- Epitope variants if studied in the paper

DatabasesSearchToolsProductsPublicationssearch siteSearch Site

Search CTL/CD8+ T-Cell Epitope Database

HIV protein

- ALL -  
Gag  
p17  
p24  
p2p7p1p6

HXB2 protein location

Results contained within query location

HXB2 DNA location

Results overlap with query location

Epitope

SLYNTVATL

Results contain query sequence

Epitope name

Record number

Subtype

- ALL -

Immunogen

- ALL -  
computer prediction  
engineered  
HIV-1 and HCV co-infection  
HIV-1 exposed seronegative  
HIV-1 infected monocyte-derived  
HIV-1 infection

Vaccine details

Vaccine type

Vaccine strain

if Immunogen is Vaccine Vaccine component

Adjuvant

- ALL -  
- ALL -  
- ALL -  
- ALL -

Species

- ALL -

MHC/HLA

- ALL -  
A\*01  
A\*01:01  
A\*01:23  
A\*02  
A\*02:01  
A\*02:02

Author

☐ First ☐ Last

Country

- ALL -

Keywords

early-expressed proteins  
early treatment  
elite controllers  
enhancing activity  
epitope processing  
escape  
genital and mucosal immunity

Note

Search

Reset

Click for Search Help

Search CTL/CD8+ variants

HIV molecular immunology database

DatabasesSearchToolsProductsPublicationssearch siteSearch Site

Search CTL/CD8+ T-Cell Epitope Database

Found 301 matching records:

Displaying record number 57

Download this epitope record as JSON.

HXB2 Location

Gag(69-93)  
p17(69-93)  
DNA(994..1068)

Gag Epitope Map

Author Location

p17(69-93 BH10)

Epitope

QTGSEELSLYNTVATLYCVHQRIE

Epitope Alignment

Species (MHC/HLA)

human(A2)

Immunogen

HIV-1 infection

Experimental methods

Keywords

Notes

- Gag CTL response studied in three individuals.

References

Johnson1991 R. P. Johnson, A. Trocha, L. Yang, G. P. Mazzara, D. L. Panicali, T. M. Buchanan, and B. D. Walker. HIV-1 Gag-Specific Cytotoxic T Lymphocytes Recognize Multiple Highly Conserved Epitopes. Fine Specificity of the Gag-Specific Response Defined by Using Unstimulated Peripheral Blood Mononuclear Cells and Cloned Effector Cells. *J. Immunol.*, 147:1512-1521, 1991. This study presented a detailed study of gag-specific CTL from HIV-1 seropositive individuals. Seven p24 and two p17 epitopes were described, that were recognized by class I-restricted CD3+CD8+ CTL. p17 epitopes: KIRLRPGGKKYKLAHIVWASRELE and QTGSEELSLYNTVATLYCVHQRIE; p24 epitopes: NPPIPVGEIYKRWILLGLNKIV, VHQAISPTLNANWYKVEEKAF, NAWVKYVEEKAFSPVPMFSA, SALSEGATPDILNTMLNTVGGH, GHQAAMQMLKETINEEAEDWR, and RAEQASQEVK. PubMed ID: 1715361. [Show all entries for this paper.](#)

Displaying record number 58923

Download this epitope record as JSON.

HXB2 Location

Gag(70-86)  
p17(70-86)  
DNA(997..1047)

Gag Epitope Map

Author Location

Gag(70-)

Epitope Alignment

GTGTEELSLYNTVATLY

probability

1.0

0.5


0

GTGTEELSLYNTVATLY

Epitope

GTGTEELSLYNTVATLY

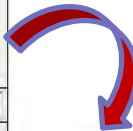
[https://www.hiv.lanl.gov/content/immunology/ctl\\_search.html](https://www.hiv.lanl.gov/content/immunology/ctl_search.html)

 **Los Alamos**  
NATIONAL LABORATORY



## Search CTL/CD8+ T-Cell Epitope

<a href="#">HIV protein</a>	- ALL - Gag p17 p24 p2p7p1p6
<a href="#">HXB2 protein location</a>	<input type="text"/> -- <input type="text"/> <input type="button" value="Results contained within query location"/>
<a href="#">HXB2 DNA location</a>	<input type="text"/> -- <input type="text"/> <input type="button" value="Results overlap with query location"/>
<a href="#">Epitope</a>	<input type="text" value="SLYNTVATL"/> <input type="button" value="Results contain query sequence"/>
<a href="#">Epitope name</a>	<input type="text"/>
<a href="#">Record number</a>	<input type="text"/>
<a href="#">Subtype</a>	- ALL - <input type="button" value="v"/>
<a href="#">Immunogen</a>	- ALL - computer prediction engineered HIV-1 and HCV co-infection HIV-1 exposed seronegative HIV-1 infected monocyte-derived HIV-1 infection
<a href="#">Vaccine details</a>	<a href="#">Vaccine type</a> - ALL - <input type="button" value="v"/> <a href="#">Vaccine strain</a> - ALL - <input type="button" value="v"/> <a href="#">Vaccine component</a> - ALL - <input type="button" value="v"/> <a href="#">Adjuvant</a> - ALL - <input type="button" value="v"/>
<a href="#">Species</a>	- ALL - <input type="button" value="v"/>
<a href="#">MHC/HLA</a>	- ALL - A*01 A*01:01 A*01:23 A*02 A*02:01 A*02:02
<a href="#">Author</a>	<input type="text"/> <input type="button" value="First"/> <input type="button" value="Last"/>
<a href="#">Country</a>	- ALL - <input type="button" value="v"/>
<a href="#">Keywords</a>	<input type="text" value="mimotopes"/> mother-to-infant transmission mutation acquisition naive T cells Nef-mediated down-regulation neutralization by CTL NK cells non-susceptible form
<a href="#">Note</a>	<input type="text"/>



Databases Search Tools Products Publications

### Search CTL/CD8+ T-Cell Epitope Database

Found 5 matching records:

Displaying record number 62248

Download this epitope record as JSON

[HXB2 Location](#) Gag(76-86)  
p17(76-86)  
DNA(805..834)

[Author Location](#) Gag(76-86)

[Gag Epitope Map](#)

[Epitope Alignment](#)

RSLYNTVATLY

Epitope RSLYNTVATLY

[Epitope Name](#) Gag-RY11  
[Subtype](#) C  
[Species \(MHC/HLA\)](#) human(A\*30)  
[Immunogen](#) HIV-1 infection  
[Patient MHC/HLA](#) S17-G: A\*30:02, A\*68:02, B\*15:10, B\*42:01, C\*03:04, C\*17:01; S17-M: A\*30:02, A\*30:04, B\*42:01, B\*58:02, C\*06:02, C\*17:01  
[Country](#) South Africa  
[Experimental methods](#) CD8 T-cell Elispot - IFN-γ, Other  
[Keywords](#) responses in children, mother-to-infant transmission, rate of progression, escape

[Show epitope variants](#)

**Notes**

- 11 perinatally infected pediatric slow progressors (PSPs) were followed longitudinally for a decade from birth to examine CTL responses to circulating and autologous HIV. It was found by ultra-deep sequencing that though, I develop variants early in infection, unlike adults, pediatric anti-variant CTL are generated

[https://www.hiv.lanl.gov/content/immunology/ctl\\_search.html](https://www.hiv.lanl.gov/content/immunology/ctl_search.html)



# CUMULATIVE CTL SEARCHES

Databases	Search	Tools	Products	Publications	Search Site
HIV protein	- ALL -				
HXB2 protein location	Gag				
HXB2 DNA location	p17				
Epitope	p24				
Epitope name	SLYNTVATL				
Record number					
Subtype	- ALL -				
Immunogen	- ALL -				
Vaccine details	computer prediction				
If Immunogen is Vaccine	engineered				
Vaccine type	HIV-1 and HCV co-infection				
Vaccine strain	HIV-1 exposed seronegative				
Vaccine component	HIV-1 infected monocyte-derived				
Adjuvant	HIV-1 infection				
Species	- ALL -				
MHC/HLA	- ALL -				
	A*01				
	A*01:01				
	A*01:23				
	A*02				
	A*02:01				
	A*02:02				

Databases	Search	Tools	Products	Publications	Search Site
Epitope	SLYNTVATL				
Epitope name					
Record number					
Subtype	- ALL -				
Immunogen	- ALL -				
Vaccine details	computer prediction				
If Immunogen is Vaccine	engineered				
Vaccine type	HIV-1 and HCV co-infection				
Vaccine strain	HIV-1 exposed seronegative				
Vaccine component	HIV-1 infected monocyte-derived				
Adjuvant	HIV-1 infection				
Species	- ALL -				
MHC/HLA	- ALL -				
	A*01				
	A*01:01				
	A*01:23				
	A*02				
	A*02:01				
	A*02:02				

Databases	Search	Tools	Products	Publications	Search Site
Epitope	SLYNTVATL				
Epitope name					
Record number					
Subtype	- ALL -				
Immunogen	- ALL -				
Vaccine details	computer prediction				
If Immunogen is Vaccine	engineered				
Vaccine type	HIV-1 and HCV co-infection				
Vaccine strain	HIV-1 exposed seronegative				
Vaccine component	HIV-1 infected monocyte-derived				
Adjuvant	HIV-1 infection				
Species	- ALL -				
MHC/HLA	- ALL -				
	A*01				
	A*01:01				
	A*01:23				
	A*02				
	A*02:01				
	A*02:02				

Databases	Search	Tools	Products	Publications	Search Site
Found 8 matching records:					
Displaying record number 60224					
Download this epitope record as JSON.					
HXB2 Location	Gag(76-86)				
Author Location	p17(76-86)				
	DNA(1015..1047)				
	Gag(76-86)				
Epitope	RSLYNTVATLY				
Epitope Name	RY11				
Species (MHC/HLA)	human(A*02)				
Immunogen	HIV-1 infection				
Experimental methods	CD8 T-cell Elispot - IFN $\gamma$ , Flow cytometric T-cell cytokine assay				
Keywords	rate of progression, optimal epitope				
Notes					
<ul style="list-style-type: none"> <li>Correlation between HLA class I and spontaneous control of HIV-1 was significant for 7-8 epitopes when 341 chronic, HIV-infected subjects, divided into controllers and non-controllers, were compared.</li> <li>Protective epitopes tend to cluster in regions of critical stability for the protein, therefore constraining escape.</li> <li>RSLYNTVATLY had a significantly positive association with HLA A*02; it was recognized by 73/199 A*02-positive subjects and 69/199 A*02-negative subjects.</li> </ul>					
References					
Pereyra2014 Florencia Pereyra, David Heckerman, Jonathan M. Carlson, Carl Kadie, Damien Z. Soghoian, Daniel Karet, Ariel Goldenthal, Oliver B. Davis, Charles E. DeZiel, T. Epitopes. J. Virol., 88(22):12937-12948, Nov 2014. PubMed ID: 25165115. Show all entries for this paper.					
Displaying record number 1169					
Download this epitope record as JSON.					
HXB2 Location	Gag(77-85)				
Author Location	p17(77-85)				
	DNA(1018..1044)				
	p17(77-85)				
Epitope	RSLYNTVATLY				
Epitope Name	RY11				
Subtype	B				
Species (MHC/HLA)	human(A*02:01, A*30:02)				
Immunogen	peptide-HLA interaction				
Experimental methods	CD8 T-cell Elispot - IFN $\gamma$ , HLA binding, Other				
Keywords	epitope processing				
Notes					
<ul style="list-style-type: none"> <li>Proteasomal cleavage effects on immunodominance were studied using 8 p17 and 11 p24 peptide variants. Epitope abundance due to Antigen processing (TAP) affinity, Endoplasmic Reticulum Aminopeptidase (ERAAP) trimming].</li> </ul>					

Databases	Search	Tools	Products	Publications	Search Site
Found 92 matching records:					
Displaying record number 57300					
Download this epitope record as JSON.					
HXB2 Location	Gag(76-86)				
Author Location	p17(76-86)				
	DNA(1015..1047)				
	p17(76-86)				
Epitope	RSLYNTVATLY				
Epitope Name	RY11				
Subtype	B				
Species (MHC/HLA)	human(A*02:01, A*30:02)				
Immunogen	peptide-HLA interaction				
Experimental methods	CD8 T-cell Elispot - IFN $\gamma$ , HLA binding, Other				
Keywords	epitope processing				
Notes					
<ul style="list-style-type: none"> <li>Proteasomal cleavage effects on immunodominance were studied using 8 p17 and 11 p24 peptide variants. Epitope abundance due to Antigen processing (TAP) affinity, Endoplasmic Reticulum Aminopeptidase (ERAAP) trimming].</li> </ul>					

Databases	Search	Tools	Products	Publications	Search Site
Found 140 matching records:					
Displaying record number 60224					
Download this epitope record as JSON.					
HXB2 Location	Gag(76-86)				
Author Location	p17(76-86)				
	DNA(1015..1047)				
	Gag(76-86)				
Epitope	RSLYNTVATLY				
Epitope Name	RY11				
Species (MHC/HLA)	human(A*02)				
Immunogen	HIV-1 infection				
Experimental methods	CD8 T-cell Elispot - IFN $\gamma$ , Flow cytometric T-cell cytokine assay				
Keywords	rate of progression, optimal epitope				
Notes					
<ul style="list-style-type: none"> <li>Correlation between HLA class I and spontaneous control of HIV-1 was significant for 7-8 epitopes when 341 chronic, HIV-infected subjects, divided into controllers and non-controllers, were compared.</li> <li>Protective epitopes tend to cluster in regions of critical stability for the protein, therefore constraining escape.</li> <li>RSLYNTVATLY had a significantly positive association with HLA A*02; it was recognized by 73/142 A*02-positive subjects and 69/199 A*02-negative subjects.</li> </ul>					
References					
Pereyra2014 Florencia Pereyra, David Heckerman, Jonathan M. Carlson, Carl Kadie, Damien Z. Soghoian, Daniel Karet, Ariel Goldenthal, Oliver B. Davis, Charles E. DeZiel, T. Epitopes. J. Virol., 88(22):12937-12948, Nov 2014. PubMed ID: 25165115. Show all entries for this paper.					

SLYNTVATL  
+ A\*02

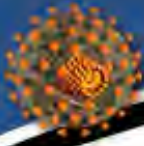
48  
records

SLYNTVATL  
+ A\*02:01

92  
records

SLYNTVATL  
+(A\*02+A\*02:01)

140  
records



## HIV sequence database

[DATABASES](#)[SEARCH](#)[ALIGNMENTS](#)[TOOLS](#)[PUBLICATIONS](#)[GUIDES](#)

### News Archive

**Note:** news releases from the LANL HIV Databases are available as [RSS feeds](#).

#### [Variant Visualizer](#)

Variant Visualizer is a new tool that provides new options and output styles for visualizing variants in an alignment of nucleotide or protein sequences. The tool is similar to [Highlighter](#), but the Variant Visualizer has more options for refining the graphical output. The output can be refined interactively without rerunning the tool. *21 July 2021*

#### [HIV Immunology Database JSON API](#)

A *JSON API* (JavaScript Object Notation - Application Programming Interface) is now available for the HIV Molecular Immunology Database to retrieve curated epitope and related data from the database in JSON format, as an alternative to the existing HTML format. It is fully documented via OpenAPI and allows the contents of the HIV Immunology Database to be queried and extracted. Data extraction may be automated for multiple searches and extracted data may then be manipulated with the user's choice of programming language. *19 March 2021*

#### [HIV Molecular Immunology 2018-19](#)

*HIV Molecular Immunology 2018-19* is now available online. The PDF version is hypertext enabled and features clickable table-of-contents, indexes, references and links to external web sites. *14 September 2020*

## HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

### Search Interfaces

- [CTL/CD8+ search](#)
- [T Helper/CD4+ search](#)
- [Antibody search](#)
- [CTL variant search](#)
- [T Helper variant search](#)
- [Patient search](#)
- [Search help](#)
- [Variant search help](#)
- [JSON API for search](#)

### Database Products

- [All Database products and publications](#)
- [Epitope maps](#)
- [Epitope tables](#)
- [Epitope alignments](#)
- [Epitope density plots](#)
- [T cell epitope variants and escape mutations](#)
- [Neutralizing antibody resources & CATNAP](#)
- [The HIV Molecular Immunology Compendium](#)
- [About the HIV Molecular Immunology Database](#)
- [How to cite this database](#)
- [Frequently-asked Questions \(FAQ\)](#)

### Tools and Data Sets

- [Tools & Links](#) for immunologists
- [SIV Epitopes \(PDF\)](#) review article summarizing known SIV epitopes
- [Identifying HLA-Associated Polymorphisms in HIV-1 \(PDF\)](#) review article summarizing HIV polymorphism associated with escape mutations. Also a [table of polymorphisms](#).
- [HLATEM](#) HLA Typing and Epitope Mapping Data Sets
- [Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development](#) Assay protocols from Duke Central Reference Laboratory

### News

[News Archive](#)

No new news.

Questions or comments? Contact us at [immuno@lanl.gov](mailto:immuno@lanl.gov)



# Antibody Search [\(\[https://www.hiv.lanl.gov/content/immunology/ab\\\_search\]\(https://www.hiv.lanl.gov/content/immunology/ab\_search\)\)](https://www.hiv.lanl.gov/content/immunology/ab_search)

<a href="#">HIV protein</a>	Proteins with <a href="#">defined epitopes</a> - ALL - p17 p17-p24 p24 p24-p2p7p1p6	Proteins with <a href="#">undefined epitopes</a> - ALL - p24 Gag RT Pol
<a href="#">HXB2 location</a>	<input type="text"/> - <input type="text"/>	Results overlap with query location
<a href="#">Epitope</a>	<input type="text"/>	Results contain query sequence
<a href="#">Record number</a>	<input type="text"/>	
<a href="#">MAb ID</a>	<input type="text"/>	( <a href="#">List by name</a> ) ( <a href="#">List by type</a> )
<a href="#">Subtype</a>	- ALL -	
<a href="#">Immunogen</a>	- ALL - anti-idiotypic autoimmune disease HIV-1 exposed seronegative HIV-1 infection HIV-2 infection in vitro stimulation or selection	
<a href="#">Vaccine details if Immunogen is Vaccine</a>	<a href="#">Vaccine type</a> <a href="#">Vaccine strain</a> <a href="#">Vaccine component</a> <a href="#">Adjuvant</a>	- ALL - - ALL - - ALL - - ALL -
<a href="#">Ab Type</a>	- ALL - C-domain C-HR C-term Env oligomer flap region gp120 adjacent to CD4BS	
<a href="#">Species</a>	- ALL -	
<a href="#">Isotype</a>	- ALL - IgA IgA1 IgA2 IgA22a IgE IgG	
<a href="#">Author</a>	<input type="text"/>	Search only for <input type="checkbox"/> First <input type="checkbox"/> Last author <input checked="" type="radio"/> Show only this author's references <input type="radio"/> Show all references
<a href="#">Country</a>	- ALL -	
<a href="#">Keywords</a>	- ALL - acute/early infection ADCC adjuvant comparison antibody binding site definition and exposure antibody generation antibody interactions	<input checked="" type="radio"/> Show only notes containing selected keyword(s) <input type="radio"/> Show all notes
<a href="#">Note</a>	<input type="text"/>	<input checked="" type="radio"/> Show only notes matching this text <input type="radio"/> Show all notes

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  
- notes and references only  
containing selected keywords OR  
- notes containing user's text

# Search Antibody Database

## Search results for 10E8

Found 30 matching records:

Displaying record number 2708

<a href="#">MAb ID</a>	10E8	<a href="#">Link to Epitope Map</a>	<a href="#">gp160 Epitope Map</a>
<a href="#">HXB2 Location</a>	gp160(671-683) DNA(8235..8273)		
<a href="#">Author Location</a>		<a href="#">Link to Epitope Alignment</a>	<a href="#">Epitope Alignment</a>
<a href="#">Epitope</a>	NWFDISNWLWYIK		
<a href="#">Subtype</a>	B		
<a href="#">Ab Type</a>	gp41 MPER (membrane proximal external region)		
<a href="#">Neutralizing</a>	P (tier 2) <a href="#">View neutralization details</a>	<a href="#">Link to CATNAP</a>	
<a href="#">Contexts and Features</a>	<a href="#">Search for contexts and features</a>	<a href="#">Link to Antibody Features Database (Ab contact positions and related protein features)</a>	
<a href="#">Species (Isotype)</a>	human(IgG3)		
<a href="#">Patient</a>	<a href="#">Donor N152</a>	<a href="#">Link to patient Donor detail</a>	
<a href="#">Immunogen</a>	HIV-1 infection		
<a href="#">Keywords</a>	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, immunoprophylaxis, immunotherapy, neutralization, review, structure, subtype comparisons, vaccine antigen design, vaccine-induced immune responses, variant cross-reactivity		

### Notes

Showing 44 of 44 notes.

### Notes from the papers

- 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, 8ANC195, IGg1b12, PGT121, PGT128, PGT135, PG9, PGT151, 35O22, 10E8, 2F5, 4E10, VRC27, VRC-CH31, VRC-PG20, PG04, VRC23, 12A12, 3BNC117, PGT145, CH01). The new algorithms accurately predicted VRC01-like and PG9-like antibody specificities. [Doria-Rose2017](#) (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. [Prevost2017](#) (ADCC, vaccine-induced immune responses)



## Neutralizing Antibody Resources

### Tools

- [CATNAP: Compile, Analyze and Tally NAb Panels](#)

Analysis of panels of antibody data for identification of potential genetic signatures.

- [Database CATNAP](#) analyzes published IC<sub>50</sub>/IC<sub>80</sub> 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.

- [HIV Genome Browser](#)

A customization of jBrowse displaying genome and proteome features of HIV, including epitopes and neutralizing antibody features.

- [Env browser](#): direct link with Ab epitopes and contact features shown.

- [CombiNAber](#)

Predict the neutralization of combinations of antibodies.

- [Neutralization Index](#)

Compute a tier-like score for anti-HIV sera and antibodies.

- [GenSig](#)

Identify genetic signatures from a DNA alignment and associated phenotypic data. Can be used to predict an antibody's signature sites based on Env sequences and neutralization data.

- [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.

### Antibody Contacts

- [Neutralizing antibody contacts and features database](#)

Search for antibody contact locations and other HIV-1 Env features.

Some contacts are also available from a spreadsheet: [Neutralizing antibody features .xlsx](#)

### Protocols and Data

- [Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols](#) from Duke Central Reference Laboratory
- [Neutralization Serotype Discovery Panel](#). A large panel of Env-pseudotyped viruses assayed against plasmas from chronic infection. The panel and plasmas were selected to represent M-group diversity.
- [CATNAP data downloads](#) for HIV-1 antibodies and their IC<sub>50</sub>, IC<sub>80</sub>, and ID<sub>50</sub> data, germline genes, and antibody sequences.

Questions or comments? Contact us at [immuno@lanl.gov](mailto:immuno@lanl.gov)

## Neutralizing Antibody Contacts & Features

Example:  
Z13e1 contact  
sites

**Purpose:** to provide HIV-1 Env coordinates of contacts and other sites associated with neutralizing antibodies. Some of these data are also summarized in a [spreadsheet \(.xlsx\)](#). For details, see [Help](#).

**MAb name** VRC-PG19  
VRC-PG20  
Y498  
Z13  
Z13e1

**Antibody class** CD4BS  
CD4i  
cluster A  
glycan  
gp120/gp41 interface

**Env AA position** 315,323

**Site type** Ab-drug interaction  
ADCC  
binding  
contacts  
Env feature

**Reference** Alam2017  
Andrabi2015  
Balla-Jhagjhoorsingh2013  
Barnes2018  
Bhiman2015

**Database ID** 1  
2  
3  
4  
5

<b>ID</b>	21
<b>Description</b>	Z13e1 contacts
<b>Antibody class</b>	MPER
<b>Reference</b>	<a href="#">Nelson2007</a>
<b>Type</b>	contacts
<b>MAb name</b>	<a href="#">Z13 Z13e1</a> (Click MAb name to get to Immunology DB notes)

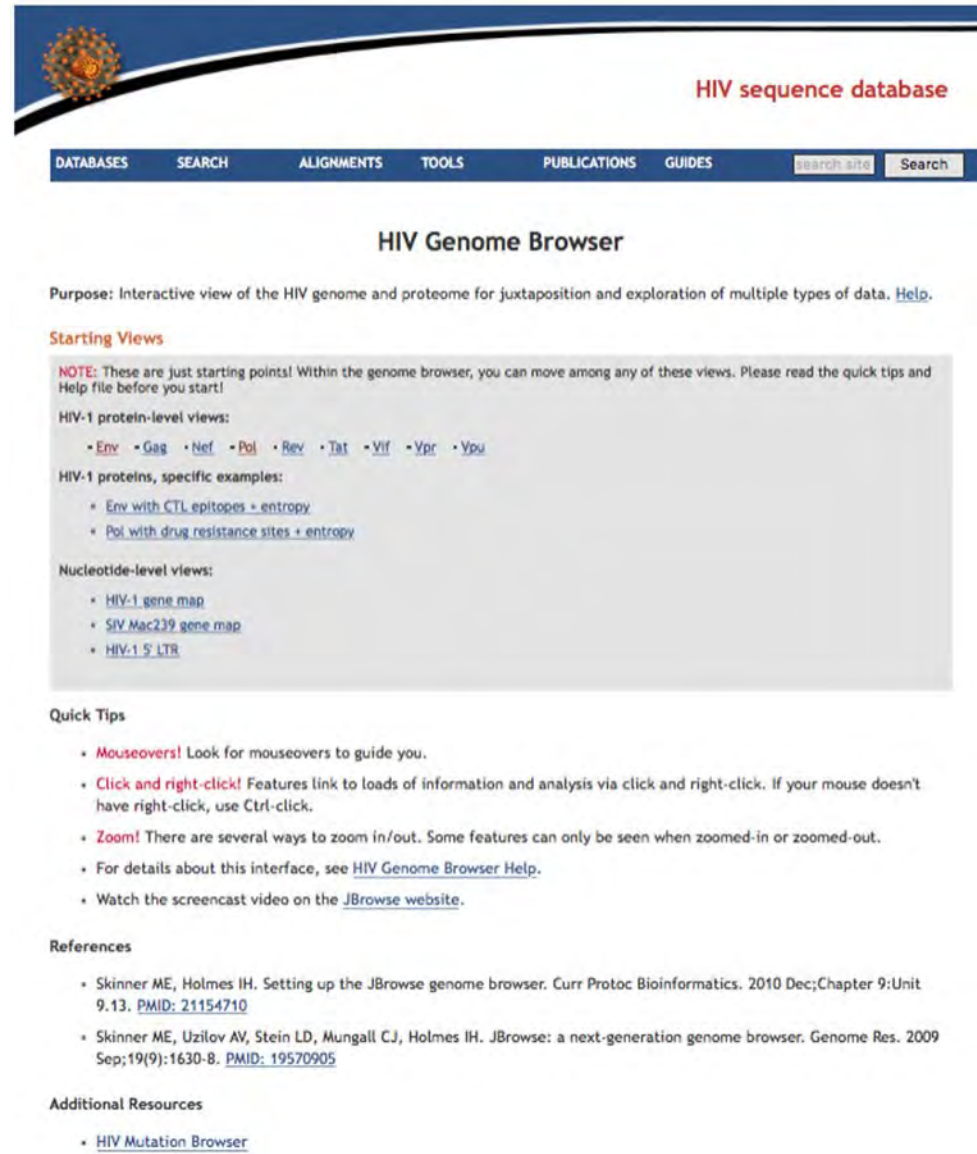
Env pos.	Feature	HXB2 AA	Entropy Group M	Entropy Subtype B	Entropy Subtype C	Annotation
668	gp41	S	0.842	0.596	0.758	Z13e1 contact site
669	gp41	L	0.020	0.023	0.000	Z13e1 contact site
670	gp41	W	0.026	0.028	0.035	Z13e1 contact site
671	gp41	N	0.779	0.669	0.885	Z13e1 contact site; crucial binding residue
672	gp41	W	0.017	0.023	0.014	Z13e1 contact site
673	gp41	F	0.058	0.065	0.073	Z13e1 contact site
674	gp41	N	1.182	1.029	1.344	Z13e1 contact site; crucial binding residue
675	gp41	I	0.050	0.059	0.069	Z13e1 contact site
676	gp41	T	0.683	0.610	0.674	Z13e1 contact site
677	gp41	N	1.187	1.199	1.100	Z13e1 contact site

Important position(s) with HxbZ amino acid: 5668 L669 W670 N671 W672 F673 N674 I675 T676 N677

# Genome Browser

## A tool at the interface between the sequence and immunology databases

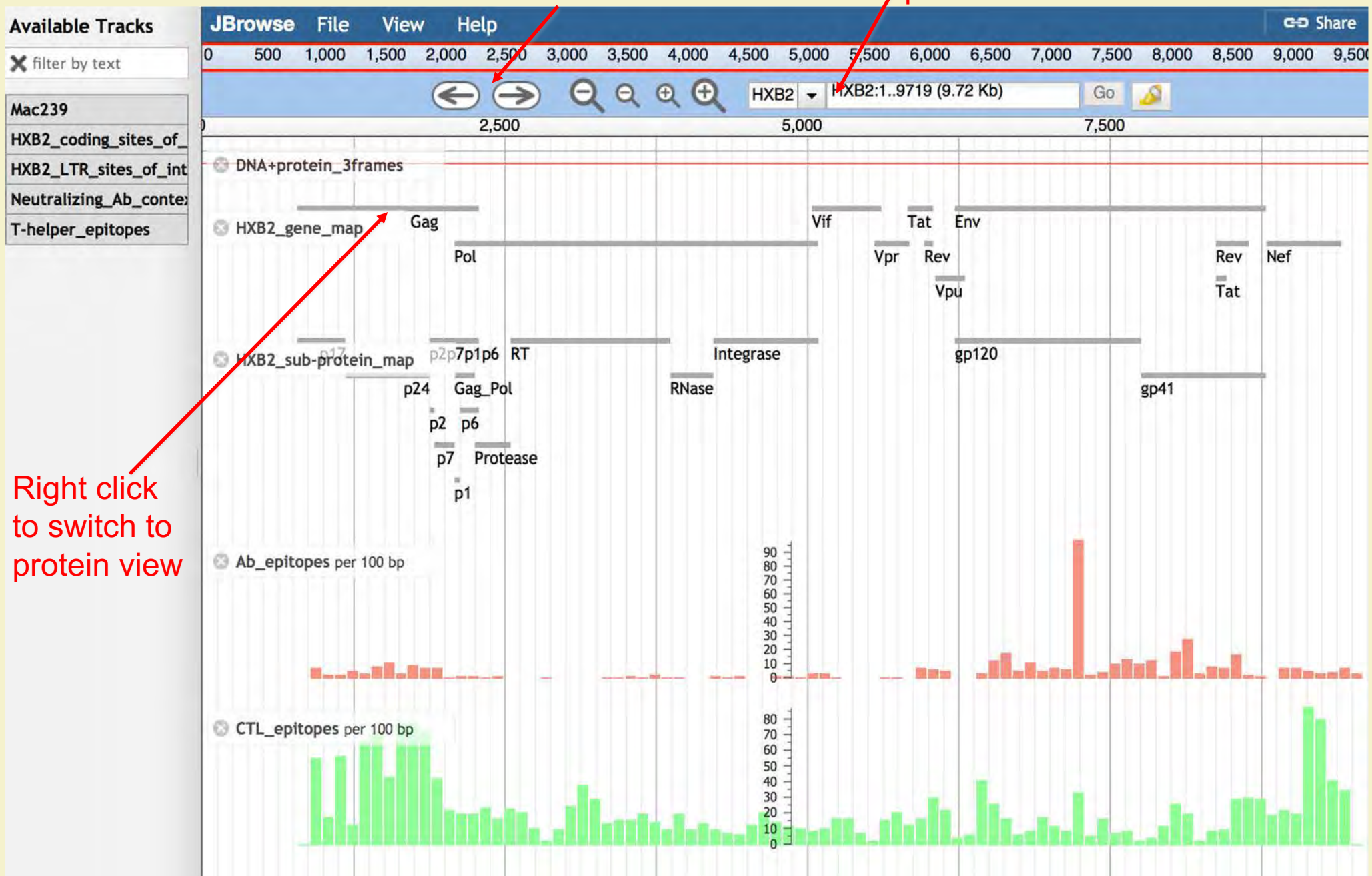
- Tools -> Genome Browser
- Click on Env
- Drag tracks into the map:
  - Entropy M group
  - Ab epitopes
  - Neutralizing Ab contexts
- Zoom in to expand a region of interest
- Click around to explore

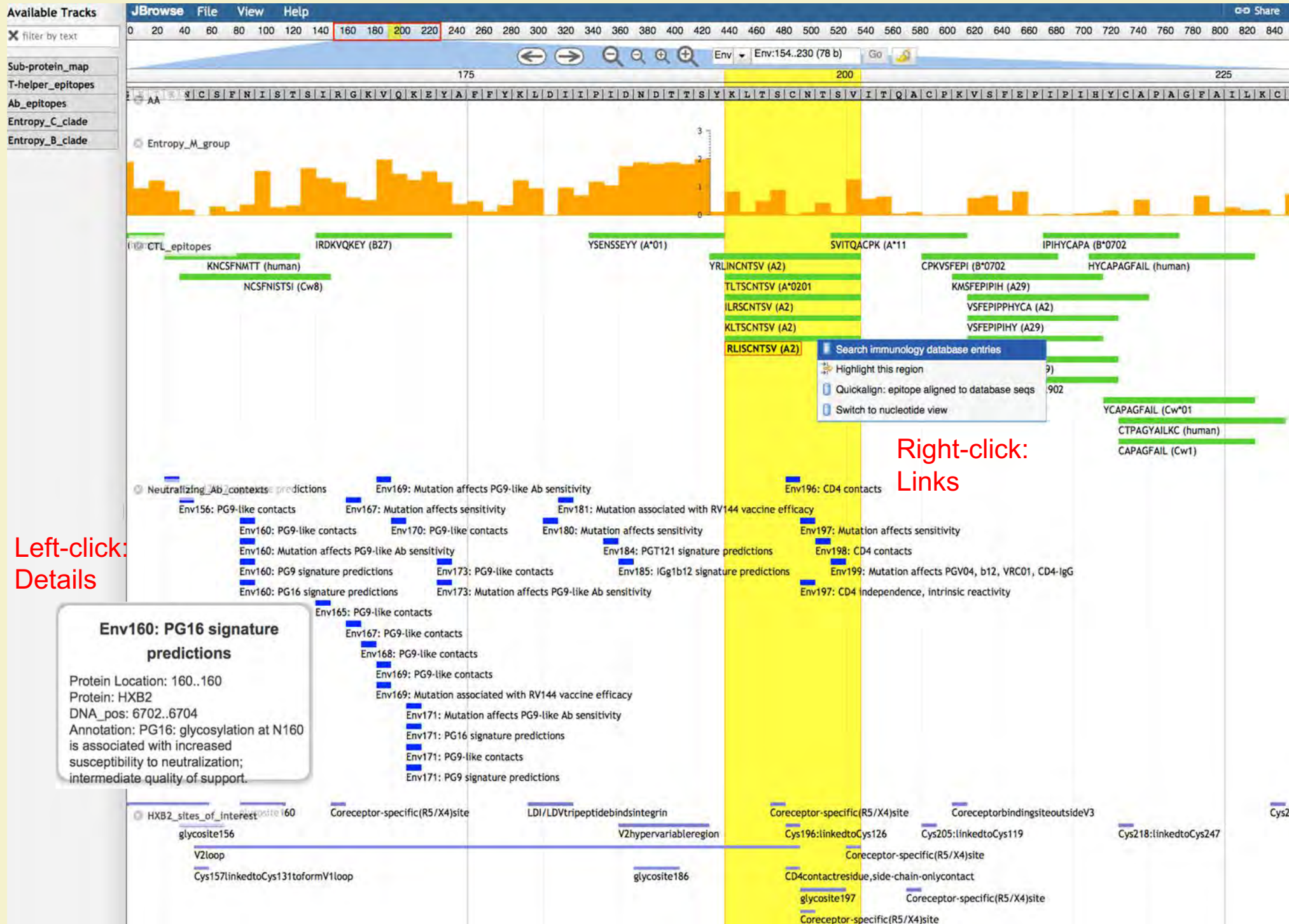


The screenshot shows the HIV Genome Browser interface. At the top, there is a navigation bar with links: DATABASES, SEARCH, ALIGNMENTS, TOOLS, PUBLICATIONS, GUIDES, and a search box. Below the navigation bar, the title "HIV Genome Browser" is displayed. The main content area includes a purpose statement: "Purpose: Interactive view of the HIV genome and proteome for juxtaposition and exploration of multiple types of data. [Help](#)." It also features a "Starting Views" section with a note: "NOTE: These are just starting points! Within the genome browser, you can move among any of these views. Please read the quick tips and Help file before you start!" The "Starting Views" section lists "HIV-1 protein-level views" (Env, Gag, Nef, Pol, Rev, Tat, Vif, Vpr, Vpu) and "HIV-1 proteins, specific examples" (Env with CTL epitopes + entropy, Pol with drug resistance sites + entropy). It also lists "Nucleotide-level views" (HIV-1 gene map, SIV Mac239 gene map, HIV-1 5' LTR). Below this, there is a "Quick Tips" section with four tips: "Mouseovers!", "Click and right-click!", "Zoom!", and "For details about this interface, see HIV Genome Browser Help." The "References" section lists two papers: Skinner ME, Holmes IH. Setting up the JBrowse genome browser. Curr Protoc Bioinformatics. 2010 Dec;Chapter 9:Unit 9.13. PMID: 21154710; and Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. JBrowse: a next-generation genome browser. Genome Res. 2009 Sep;19(9):1630-8. PMID: 19570905. Finally, the "Additional Resources" section lists "HIV Mutation Browser".



# HIV Genome Browser: Nucleotide view







# CATNAP (Compile, Analyze and Tally NAb Panels)

- CATNAP is both a *database* and a set of *tools*
- Compiles published HIV Ab neutralization data
  - >400 Abs, >1000 HIV pseudoviruses, >100,000 IC<sub>50/80</sub> values
- Integrates neutralization data with viral sequences
- Provides important Ab and Virus details:
  - Ab binding region, links to PDB structures, links to the donor info
  - Clonal lineage, germline V/D/J designation, Ab sequences
  - Virus tier, subtype, country, infection stage, aligned sequence
- Select Abs and viruses in multiple ways:
  - Individual or groups of Abs and viruses, or results by study
  - Antibodies by germline V/D/J genes and binding region
  - Viruses by tier, subtype, infection stage, or established viral panels
  - Your favorite list of viruses and antibodies
- Defines genetic neutralization signatures associated with sequences

## CATNAP

### Compile, Analyze and Tally NAb Panels

The CATNAP family of tools has been designed to facilitate the analysis of neutralizing antibodies (NAbs) through the identification of potential genetic signatures resulting from a NAB's interaction with a protein. While interactions between NAbs and HIV-1 Env are the emphasis, the Custom Input version can accommodate many other types of data, including other proteins and organisms.

### CATNAP

**Purpose:** Analyze our database of IC<sub>50</sub>, IC<sub>80</sub>, and ID<sub>50</sub> neutralization data from publicly-available sources, in conjunction with HIV Env sequences. Or download these data for your own analyses.

- [CATNAP Help](#)
- [CATNAP download](#): download all CATNAP neutralization data, Env alignment, antibody sequences, and germline genes
- [Find Names](#): convert your mAb and virus names to CATNAP standard names

### CATNAP: Custom Input

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

- [Custom CATNAP Help](#)

### CATNAP: Hybrid

**Purpose:** Compare and analyze your HIV-1 IC<sub>50</sub> and IC<sub>80</sub> 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.

- [Hybrid CATNAP Help](#)

### Reference

The URL for citation is <https://hiv.lanl.gov/catnap>

When using this tool in a publication, please cite:

Yoon et al. CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. Nucleic Acid Res 2015 Jul 1;43(W1):W213-9.

[PMID 26044712](#).

### Downloads:

- Virus sequences
- Virus alignment
- Ab sequences
- IC<sub>50/80</sub> data

**Find Names:** quickly convert your list of virus or Ab names to the standardized ones in CATNAP

### Custom Input accepts

- Numerical data: IC<sub>50</sub>, ID<sub>50</sub>, AUC, or *any* phenotypic data
- Aligned sequences associated with that data

**Hybrid** combines your HIV neutralization data with published data from CATNAP's database

## 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](#)

[Download CATNAP data](#)

**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**

Many ways to select antibodies and viruses

Antibodies by ☒ Names ☐ Attributes

# of Abs = 307, # of Ab mixtures = 25

Select	Name	Donor	# of viruses tested
<input type="checkbox"/>	10-1074	<a href="#">Donor 17</a>	420
<input type="checkbox"/>	10-1074-IgG3C	<a href="#">Donor 17</a>	119
<input type="checkbox"/>	10-1074V	<a href="#">Donor 17</a>	200
<input type="checkbox"/>	10-996	<a href="#">Donor 17</a>	121
<input checked="" type="checkbox"/>	10E8	<a href="#">Donor N152</a>	433
<input type="checkbox"/>	10E8-OfH/OfL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-OfH/4fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-10fH/10fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-10fH/16fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-10fH/4fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-19fH/10fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-19fH/16fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-2fH/OfL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-2fH/10fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8-2fH/4fL	<a href="#">Donor N152</a>	21
<input type="checkbox"/>	10E8V1.1/P140	<a href="#">Donor N152</a>	118

Viruses by ☒ Names ☐ Attributes ☐ Panels

# of Viruses = 1011 (785 seqs available)

Select	Name	Subtype	# of Abs tested	Seq
<input type="checkbox"/>	0013095_2_11	C	147	Yes
<input type="checkbox"/>	001428_2_42	C	146	Yes
<input type="checkbox"/>	0041_V3_C18	C	23	Yes
<input type="checkbox"/>	0077_V1_C16	C	72	Yes
<input type="checkbox"/>	00836_2_5	C	71	Yes
<input type="checkbox"/>	0260_V5_C1	A1	11	Yes
<input type="checkbox"/>	0260_V5_C36	A1	163	Yes
<input type="checkbox"/>	0301_BM_A12	C	12	Yes
<input type="checkbox"/>	0301_BM_A2	C	12	Yes
<input type="checkbox"/>	0301_BM_A6	C	12	Yes
<input type="checkbox"/>	0330_V4_C3	A1	122	Yes
<input type="checkbox"/>	0404_BM_B9	C	12	Yes
<input type="checkbox"/>	0404_BM_D4	C	12	Yes
<input type="checkbox"/>	0404_BM_F3	C	12	Yes
<input type="checkbox"/>	0404_BM_G3	C	12	Yes
<input type="checkbox"/>	0404_BM_H4	C	12	Yes

Retrieve data: Antibody, Virus or Assay details (then download the data or virus alignment for just the set you selected)

### Options

Retrieve ☒ Antibody details ☐ Virus details ☐ Assay

Or

Analyze along with virus sequences ☐ IC<sub>50</sub> ☐ IC<sub>80</sub> ☐ Both

Large sets of data run slowly. Limit the number of antibodies or viruses for quicker response.

☐ Exclude viruses having no sequence data

☐ Email results

Analyze IC<sub>50</sub>, IC<sub>80</sub> or both, along with the viral sequences



[Go to CATNAP main page](#)

## Antibody information

Number of antibodies: 1

Download heavy and light ☒ aa ☐ na sequences in

Download table below

Expand table below to show heavy and light chain sequences and sources for germline data

Antibody	Antibody binding type	Structure	Donor	Clonal lineage	Isolation paper	Neutralizing antibody feature	Heavy V (IGHV)	Heavy D (IGHD)	Heavy J (IGHJ)	Light V (IGKV or IGLV)	Light J (IGKJ or IGLJ)	Light chain type	Genetic signature analysis	LANL comments
<a href="#">10E8</a>	<ul style="list-style-type: none"> <li>C-term</li> <li>gp41 MPER (membrane proximal external region)</li> </ul>	<a href="#">4U6G</a> <a href="#">5IQ7</a> <a href="#">5IQ9</a> <a href="#">4G6F</a>	Donor <a href="#">N152</a>	10E8	<a href="#">Huang2012a</a>	<ul style="list-style-type: none"> <li><a href="#">10E8 contacts</a></li> <li><a href="#">10E8 residue prediction</a></li> <li><a href="#">10E8 signature predictions (West2013)</a></li> </ul>	3-15*05	3-3*01	1*01	3-19*01	3*02	L	<a href="#">IC<sub>50</sub></a> <a href="#">IC<sub>80</sub></a>	

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

Pre-calculated results from the Genetic Signatures tool

Link to structures in PDB

Link to the patient record for the mAb donor

## Assay

Analyze assay data in CATNAP

Number of data: 1551

Download table below with additional virus info

Expand table below to show virus information

Antibody	Virus	Reference	IC50	Mean IC50	IC80	Mean IC80
10E8	0013095_2_11	<a href="#">Asokan et al. J Virol 89:12501 (2015)</a>	0.002		0.058	
		<a href="#">Chuang et al. J Virol. 87:10047 (2013)</a>	0.013			
		<a href="#">Doria-Rose et al. J Virol. 90:76 (2016)</a>	0.00200		0.05800	
		<a href="#">Huang et al. Immunity 45:1108 (2016a) - dataset 1</a>	0.003	0.00454		0.07723
		<a href="#">Huang et al. Nature 491:406 (2012a)</a>	0.003		0.069	
		<a href="#">Kong et al. J Virol 89:2659 (2015) - dataset 1</a>	0.017		0.194	
		<a href="#">Kong et al. J Virol 89:2659 (2015) - dataset 2</a>	0.005		0.061	

Assay data may include results from multiple studies.

When multiple studies are included, a geometric mean is provided

Example: Retrieve Ab details and assay data for mAb 10E8



## Example: PGT141 mAb Donor

### Patient Detail

Patient Code	Donor 84 (Donor 584)
Patient Sex	
Risk Factor	
Infection Country	RW
Infection City	
Infection Year	
HLA Type	
Patient Ethnicity	
Progression	
Species	human
Patient Note	Donor classified as elite neutralizer; infected for at least 3 years, at least 18 years old and asymptomatic without ARV at time of enrollment; infected with clade C. Andrabi2015 referred to this patient as Donor 584.
CTL CD8+ Records	
T-Helper CD4+ Records	
Antibody Records	PGT141 ( <a href="#">2648</a> ), PGT142 ( <a href="#">2649</a> ), PGT143 ( <a href="#">2650</a> ), PGT145 ( <a href="#">2651</a> ), PGT144 ( <a href="#">2652</a> ), PGDM1400 ( <a href="#">3201</a> ), PGDM1401 ( <a href="#">3203</a> ), PGDM1402 ( <a href="#">3204</a> ), PGDM1403 ( <a href="#">3205</a> ), PGDM1404 ( <a href="#">3206</a> ), PGDM1405 ( <a href="#">3207</a> ), PGDM1406 ( <a href="#">3208</a> ), PGDM1407 ( <a href="#">3209</a> ), PGDM1408 ( <a href="#">3210</a> ), PGDM1409 ( <a href="#">3211</a> ), PGDM1410 ( <a href="#">3212</a> ), PGDM1411 ( <a href="#">3213</a> ), PGDM1412 ( <a href="#">3214</a> ), 3BNC117/PGDM1400 ( <a href="#">3478</a> ), 3BNC117/PGT145 ( <a href="#">3479</a> ), 8ANC195/PGT145 ( <a href="#">3529</a> ), 8ANC195/PGDM1400 ( <a href="#">3530</a> ), VRC01/PGDM1400-10E8v4 ( <a href="#">3955</a> ), polyclonal Donor 84 ( <a href="#">3966</a> ), PGDM1400/PRO-140 ( <a href="#">3979</a> )
Sequence Database Patient ID Record	<a href="#">85588</a>

Link to patient's HIV sequences

List of all published mAbs derived from this patient

# Example: retrieve virus info for a set of 8 commonly-used pseudoviruses

## Virus information

Number of viruses: 8

☒ aa ☐ na in

Download an ENV alignment of just the 8 selected viruses

table below

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

[More info in HIV Sequence DB](#)

Virus name	Subtype	Country	Year	Patient health	Days post infection	Days from seroconversion	Fiebig	Risk factor	Accession	Tier	Infection stage	Coreceptor	Alias	HIV DB name	Seq data	LANL comments
0013095_2_11	C	INDIA	2000	acute infection	45		4	Heterosexual	<a href="#">EF117267</a>	2	intermediate	CCR5	0013095, 0013095-2.11, 0013095.2.11, HIV-0013095-2.11, HIV-0013095.2.11, HIV_0013095_2_11	HIV_0013095_2	Yes	
001428_2_42	C	INDIA	2000	acute infection	45		4	Heterosexual	<a href="#">EF117266</a>	2	intermediate	CCR5	001428, 001428-2.42, HIV-001428-2.42, HIV-001428.2.42, HIV_001428_2_42	HIV_001428_2	Yes	
0260_V5_C36	A1	TANZANIA	2005			early		Heterosexual	<a href="#">HM215256</a>		early		0260.5.36, 0260.v5.c36	0260_v5_c36	Yes	
0301_BM_A2	C	MALAWI	2008						<a href="#">HM070482</a>				0301_bmA2	0301bmL_A2	Yes	
0330_V4_C3	A1	TANZANIA	2005				5 or 6	Heterosexual	<a href="#">HM215257</a>	2	early		0330.v4.c3	0330_v4_c3	Yes	The sequence we provide is an unpublished sequence that is 7 nucleotides longer at the 3' end, but otherwise identical to the GenBank sequence (personal communication David Montefiori group).
0404_BM_B9	C	MALAWI	2009						<a href="#">HQ595836</a>				0404_bmB9	0404bmL_B9	Yes	
0702_BM_H12	C	MALAWI	2008						<a href="#">HM070525</a>				0702_bmH12	0702bmL_H12	Yes	
0735_V4_C1	A1C	TANZANIA	2005				5 or 6	Heterosexual	<a href="#">HM215259</a>		early		0735.v4.c1	0735_v4_c1	Yes	

Link to the sequence record in the HIV Sequence DB

Comments explain when our sequence differs from GenBank sequence



Collapse or expand details from individual studies

[More virus info in HIV Seq DB](#)

```

HXB2
MRVKE---KY-QHLW-RWG---WRWGTMLLG---MLMI---CSAT---
-----|-----|-----|-----
-----10-----20-----30-----
MRVRGILR-NY-QQW-----WMWGVLFWF---MLMI---CNGV---
MRVRGILR-NW-QLW-----WTWGILGFW---MVMN---CNVR---
MRVMGSMR-NC-QRW-----WIWGILGFW---MLMT---CNME---
MRVRGIRR-NY-QHW-----WIWGILGFW---MLMI---CKGGR---
MRVMGIQR-NS-QCF-----LSWGMLVLG---IMMI---CSAV---
MRVRGMMR-NW-QQW-----WIWGILGFW---MLMI---CSVL---
MRVRGMMR-NW-QQW-----WIWGILGFW---MLMM---CSVL---
MRVRGMMR-NW-QQW-----WIWGILGFW---MLMM---CSVL---
MRVMGMQR-NS-RHL-----LLRWGIRILG-MIMI---CRTA---
MRVRGILR-NC-PQW-----WTWGILGFW---MLMI---CSVW---
MRVRGILR-NC-PQW-----WTWGILGFW---MLMI---CSVW---
MRVRGILR-NC-PQW-----WTWGILGFW---MLMI---CSVW---
MRVRGILR-NC-PQW-----WTWGILGFW---MLMI---CSVW---
MRVRGILR-NC-PQW-----WTWGILGFW---MLMI---YSVW---
MRVMGIQR-NC-QHL-----LRWGTLLLG---LIII---CSTA---
MRVRGILR-NW-ELW-----WIWGILGFW---MFMI---CNML---
MRVRGILK-NW-KLW-----WIWGILGFW---MFMI---CNML---
MRVRGILR-NW-ELW-----WIWGILGFW---MFMI---CNML---
MRVRGILR-NW-KLW-----WIWGILGFW---IFMI---CNTL---
MRVMGIQM-NW-QQW-----WIWGILGFW---MLMV---CNGT---
MRAREMKR-NC-QNL-----WKWGIMLLG---ILMI---CSAA---

```

## Potency and Breadth of neutralization over multiple studies

Download aa na in Fasta

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

- 

[See full raw counts](#)

	Percentage and raw count of non-gap			
671	N: 72.14% (3947)	S: 21.09% (1154)	T: 5.63% (308)	other: 1.13% (62)
672	W: 99.80% (5460)	other: 0.20% (11)		
673	F: 98.83% (5407)	other: 1.17% (64)		
676	T: 64.17% (3511)	S: 35.48% (1941)	other: 0.35% (19)	
680	W: 98.92% (5412)	other: 1.08% (59)		
683	K: 77.82% (4257)	R: 21.39% (1170)	other: 0.79% (43)	

Analyze HXB2 position 160 for Ab PG9 :

Pick Ab and click on contact position to analyze, or enter your own position

Run CombiNaber  Submit

# of viruses tested							
10E8 IC50: 432	10E8 IC80: 400	PG9 IC50: 729	PG9 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.



Amino Acid Counts

AA	Count	# for detected	# for undetected	Fisher test p-value	Odds ratio
N	544	425	119	< 2.2e-16	25.55874
D	10	0	10	1.37e-06	0
K	9	0	9	5.403e-06	0
S	5	1	4	0.01897	0.0884202
Y	5	1	4	0.01897	0.0884202
X	4	3	1	1	1.081824
R	3	0	3	0.01834	0
T	1	0	1	0.265	0
V	1	0	1	0.265	0
H	1	0	1	0.265	0
-	1	0	1	0.265	0
I	1	0	1	0.265	0
Total	585	430	155		
no seq	144				
Grand total	729				

**EXAMPLE:** a quick “position analysis” of the neutralization associated with specific amino acids at Env position 160 for mAb PG9

Note: The **Genetic Signature Tool** calculates phylogenetically corrected signatures across all Env positions

N-linked Glycosylation Motif Counts

NxST	Count	# for detected	# for undetected	Fisher test p-value	Odds ratio
g+	531	424	107	< 2.2e-16	31.48806
g-	53	6	47	< 2.2e-16	0.03273309
-	1	0	1	0.265	0
Total	585	430	155		
no seq	144				
Grand total	729				

Odds ratio >1: enriched for detected (neutralized)  
Odds ratio <1: enriched for undetected (not neutralized)

## HXB2

IEKGEIKNCSFNISTSRG-KVQKEYAFFYKLDIIPIDN-----DTT  
|-----||-----|-----|-----|-----|-----|  
0-----160-----170-----180-----19

## AA (NxST)

N (+) YKEDIRNCSFNATTEVKD-KKQKVHALFYRLDIVPLNKRNSSESEEN---SSG  
N (+) NGDEMKNCSFNITTEIRD-KKQKAYALFYRLDLVPLERENRGDSN-----SAS  
N (+) TSNEMKNCSFNITTEIRD-KKKKESAI FYKLDVVPDNGNNSG-----NYS  
N (+) TYESMKNCSFNITTELKD-KKQSVYALFYRLDIVPLNN-----SNE  
N (+) MEGEIKDCSFNVITTEL RD-KRQKVHSLFYRLDIVQINSSQT-----NSS  
N (+) ITRDELNCSYNMTTEL RD-RRQKVSLFYRLDIVEIENNR TNRT-----NNT  
N (+) ITENERKNCSFNITTEL RD-KSKQVYSLFYRLDIVPIDGSDNSSD-----NSN  
R ISTADMKNCSFRVPTAIRD-RKQKVYSLFYRLDIVQIDKKKND SNNS-----NIT  
N (+) ---IMT NCTFNITTELKD-KKRKASAFYRLDIVPLNGDSNGS-----SSG  
N (+) IDKGEMKKCSFNITTSIRG-KMQKEYALFYKLDIVPIDNGKNDS-----TNT  
N (+) ESGEIKNCSFNITTSVRD-KVQKEYALFYKLDIVPITN-----ESS  
N IDPGEIKNCSFNITATPIKD-KRHQYALFYKSDVVPIDEDN-----DTT  
N (+) IEKGEIKNCSFNITTNIRD-KYQKAYALFYKLDVVPIDDDNATGNN-----DTR  
N (+) NGEEIKNCSFNATTEIRD-KKQKVYALFYRLDIVPLEEERKG-----NSS  
N (+) DMGEIKNCSFNITTELID-KQKKVHALFYRLDIVSLEKDNSSKKND-----SNE  
N (+) INVEEMKNCSFNITTEL RD-RKQTVYASFYKLDIVPLNENKST-----SSE  
N (+) MEGEIKNCSFNMTTEL RD-KNQKVYALFYRQDVIQNGNN-----NSS  
N (+) PEAGMKNCSFNITTEVKD-KKKLVYAHFYRLDVVQLDG-----NTN  
N (+) IQGEEMKNCSFNVTAE LR D-KRKNEYALFYRQDVVQINET-----DNS  
N (+) IMKGEITNCSFNMTTEL RD-KKQKVS AFFYRQDVVPVNSNQ-----DNS  
N (+) INTEDMKNCSFNITTI VRD-KKKQYALFYRLDIVEINP-----NDT  
N (+) CMKELNCSFNITTEL RD-KKQKAYALFYRLDIVPLNKRNSSESEEN-----SSG

## About this position

Position: Env 160 (193 in alignment above)

Entropy, M group: 0.401

Functional domain: gp120 ([Kwong2000](#)), V2 ([Leonard1990](#))

## Antibody features of this position

Mutation affects PG9-like Ab sensitivity: Loss of glycan confers resistance; PG9-like class includes PG16, PGT141, 145, CH01-CH04 (V1V2 glycan, [Doria-RoseNA2012](#))

PG16 signature predictions: PG16: glycosylation at N160 is associated with increased susceptibility to neutralization; intermediate quality of support. (V1V2 glycan, [West2013](#))

PG9-like contacts: PG9 glycan contact; PG9-like class includes PG16, PGT141, 145, CH01-CH04 (V1V2 glycan, [McLellan2011](#))

PG9 signature predictions: PG9: N160 is associated with increased susceptibility to neutralization; intermediate quality of support. (V1V2 glycan, [West2013](#))

(For more information, check [Neutralizing Antibody Contexts & Features](#))

Env position 160 is highlighted

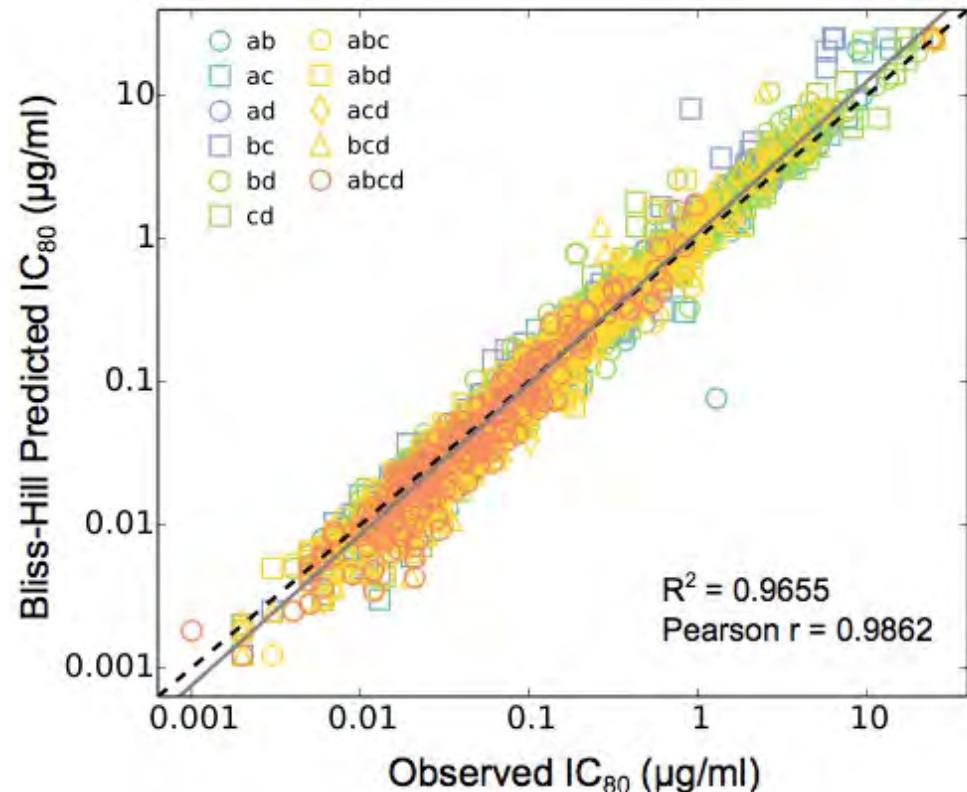
# CombiNAber

- Background

- Papers: Wagh *et al*, 2016, *PLOS Pathogens* 12(3) e1005520; Kong *et al*, 2015, *J Virol* 89(5):2659-71. Questions: [Kshitij Wagh, kshitij@lanl.gov](mailto:kshitij@lanl.gov)
- To counter HIV-1 Env diversity, bNAb combinations that improve breadth and potency over single bNAbs will be needed.
- We have developed “Bliss-Hill” and “Additive” models that accurately predict bNAb combination neutralization given individual bNAb data.
- We used these models to systematically predict and compare all bNAb combinations and identified optimal bNAb combinations.

- Purpose of CombiNAber:

- Predict neutralization profiles of bNAb combinations using individual bNAb IC<sub>50</sub> & IC<sub>80</sub> data as input.
- Systematically compare bNAb combinations to identify optimal combinations.



bNAbs:

- a) CD4bs - VRC07, 3BNC117
- b) V2 – PG9
- c) V3 – PGT128, 10-1074
- d) MPER – 10E8

Wagh *et al*. 2016 PLoS Pathog 12(3): e1005520



# CombiNAber Input Page

## CombiNAber

### A tool for Prediction & Analysis of Neutralization by Antibody Combinations

**Purpose:** This tool predicts and analyzes combination antibody neutralization scores using  $IC_{50}$  and/or  $IC_{80}$  for individual antibodies. The predicted scores are systematically compared for all single antibodies and 2, 3 and 4 antibody combinations analyzed. See [explanation](#).

#### $IC_{50}/IC_{80}$ data

Paste values or upload file

(See [assay requirements](#)) '<' and '>' signs are NOT allowed. Please replace them with 'LT' and 'GT' respectively. 



[Sample Input]

 No file selected.

Data type ☐  $IC_{50}$  ☐  $IC_{80}$  ☒ Both


Delimiter ☐ Comma ☐ Space ☒ Tab


#### $IC_{50}$ / $IC_{80}$ data:

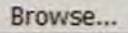
- Viruses on rows, mAbs on columns
- Giving both  $IC_{50}$  &  $IC_{80}$  data allows prediction using the more accurate Bliss-Hill Model

#### mAb class

Paste values or upload file

(See [Ab class requirements](#)) 



 No file selected.

Delimiter ☐ Comma ☐ Space ☒ Tab

#### mAb class:

- Rough epitopes of where each mAb targets. e.g. VRC01 target CD4bs, PGT121 targets V3.
- User can use any nomenclature of epitopes, just have to be consistent.



# CombiNAber Input Page

## Options

The screenshot shows the 'Options' section of the CombiNAber web application. It includes several configuration options: 'Prediction method' with radio buttons for 'Additive' (selected) and 'Bliss-Hill'; 'mAb combinations' with a checked option for 'Combinations using full set of mAbs', a checkbox for '# of Abs in Ab combination' (checked for 2, 3, and 4), and an unchecked checkbox for 'Repeat mAbs from same class in combinations'; a text input for 'Combinations of interest (example)' with a 'Browse...' button; 'Analyses' with a 'Target concentration' input set to '10' and a unit of 'ug/ml'; checkboxes for 'Active coverage by multiple mAbs in combination' (2, 3, 4), 'Incomplete neutralization', and 'Instantaneous inhibitory potential (IIP)'; 'File format for figures' with checkboxes for 'PDF', 'SVG', and 'PNG' (checked); and an 'Email results' checkbox. At the bottom are 'Submit' and 'Reset' buttons. Three arrows originate from the form: one from the 'Additive' radio button points to the 'Prediction method' text block; one from the 'Combinations using full set of mAbs' checkbox points to the 'mAb combinations' text block; and one from the 'Combinations of interest' text input points to the final bullet point in the 'mAb combinations' text block.

### Prediction method:

- Bliss-Hill model is more accurate, but requires both IC50 & IC80 titers.
- If only IC50 or IC80 titers available, choose “Additive” model.

### mAb combinations:

- Choose how many Abs you want in combinations.
- Should mAbs from the same class be repeated in combinations?  
Note: best combinations typically have mAbs from different epitopes.
- All mAb combinations that are consistent with above two options will be predicted and compared.
- Combinations of interest to be highlighted in analyses and figures

# CombiNAber Input Page

## Options

Prediction method ☒ Additive ☐ Bliss-Hill

mAb combinations ☒ Combinations using full set of mAbs

# of Abs in Ab combination ☒ 2 ☒ 3 ☒ 4 (may be adjusted depending # of Abs)

☐ Repeat mAbs from same class in combinations

Combinations of interest ([example](#))

No file selected.

Analyses Target concentration  ug/ml (separate with commas if more than one concentration)

Active coverage by multiple mAbs in combination ☐ 2 ☐ 3 ☐ 4

☐ Incomplete neutralization

☐ Instantaneous inhibitory potential (IIP)

File format for figures ☐ PDF ☐ SVG ☒ PNG

Email results ☐

## Target concentration:

Concentration of each mAb that will correspond to a “physiological” concentration of interest.

e.g. in AMP trials trough concentration ~10µg/ml.

By default combination IC50 and IC80 titers will be predicted.

User can select more stringent metrics that we have shown to be relevant for in vivo success (see Wagh et al. PLoS Pathogens 2016):

- How many viruses are actively neutralized by at least 2/3/4 mAbs in the combination? (higher the number of mAbs, lower the chance of escape of viruses)
- Incomplete neutralization and IIP measure how completely are viruses predicted to be neutralized by mAb combinations

File formats: PNG by default, but can also generate PDF and SVG

For most jobs, email results is preferred to avoid webpage time-out.

[www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html](http://www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html)

# CombiNAber Output Page



## CombiNAber

### Input & options:

IC<sub>50</sub>/IC<sub>80</sub> data: 15 ab(s) found, 25 virus(es) found. [\[See\]](#)

Data type: IC<sub>50</sub> & IC<sub>80</sub>

mAb class: 4 class(es) found

Prediction method: Bliss-Hill

mAb combinations: Combinations using full set of abs, # of mAbs in combination = 2,3. Combinations of interest provided

Repeat mAbs from same class in combinations: No

Target concentration: 10.0 ug/ml

Active coverage by multiple mAbs in combination: 2

Incomplete neutralization: No

Instantaneous inhibitory potential (IIP): No

File format for figures: png

Input parameters

### Results

Target concentration 10.0 ug/ml

[Single mAbs](#) [2 mAbs combinations](#) [3 mAbs combinations](#)

[Comparison of best mAb and best combinations with different number of mAbs](#) [Combinations of interest](#)

[Download in an archived zip](#)

All results can be downloaded in  
an archived zip file

Links to the part of the webpage  
with analyses of single mAbs,  
2/3/4 mAb combinations, and  
best mAbs/combinations with  
different number of mAbs and  
combinations of interest



# CombiNAber Output Page

2. 2 mAbs combinations (target concentration = 10.0 ug/ml)

We will focus on 2mAb combinations

[See summary explanation](#)

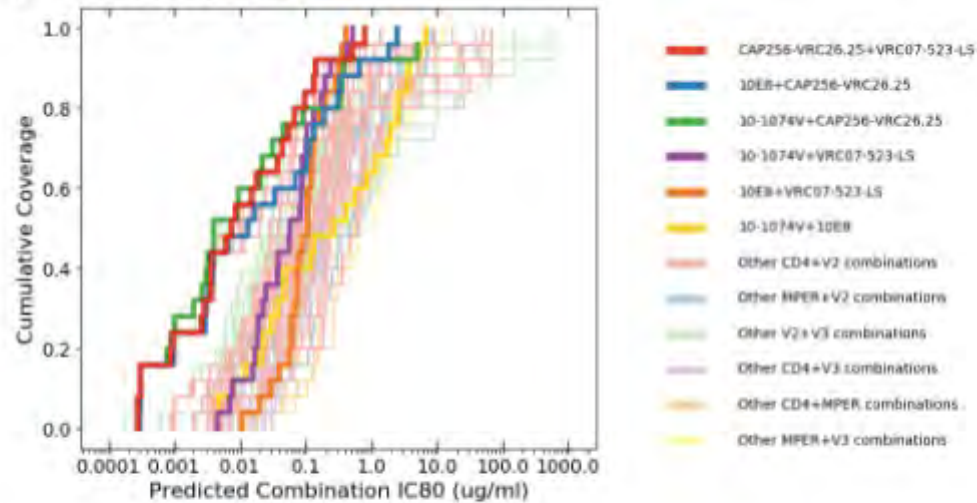
Summary of best-in-class 2mAb combinations using BH model predictions and target concentration = 10.0 ug/ml

Combination	Class	IC80 Geometric Mean	Coverage at IC80 < 10.0 ug/ml	Percent Viruses with maximum inhibition > 0.95 at 10.0 ug/ml	Median IIP at 10.0 ug/ml	Percent Viruses with IIP > 5 at 10.0 ug/ml	Geometric Mean IC80 with at least 2 active at IC80 < 10.0 ug/ml	Coverage with at least 2 active at IC80 < 10.0 ug/ml	Overall Rank
CAP256-VRC26.25+VRC07-523-LS	cd4+v2	0.00907	100.0	100.0	4.27757	32.0	0.00297	72.0	1
10E8+CAP256-VRC26.25	mper+v2	0.01585	100.0	92.0	3.22039	0.0	0.0038	60.0	6
10-1074V+CAP256-VRC26.25	v2+v3	0.01092	96.0	88.0	2.91682	24.0	0.00138	40.0	7
10-1074V+VRC07-523-LS	cd4+v3	0.05175	100.0	100.0	4.27957	36.0	0.02999	64.0	9
10E8+VRC07-523-LS	cd4+mper	0.09344	100.0	100.0	3.48373	4.0	0.09024	88.0	39
10-1074V+10E8	mper+v3	0.215	100.0	72.0	2.81661	12.0	0.04877	52.0	53

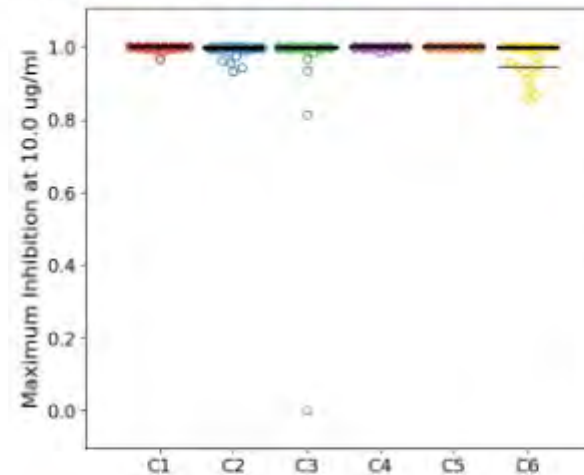
Table shows the best 2 mAb combination of each epitope class (e.g. CAP256-VRC26.25 + VRC07-523-LS is cd4+v2 and 10E8 + CAP256-VRC26.25 is mper+v2) using all the metrics in the table. Overall rank in 2mAb combinations is also shown.

# CombiNAber Output Page

Overall breadth potency ?



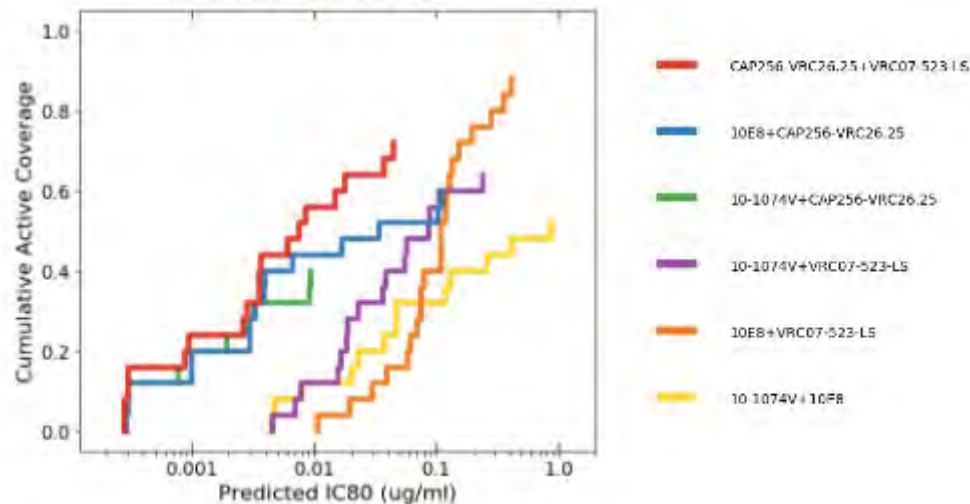
Incomplete Neutralization ?



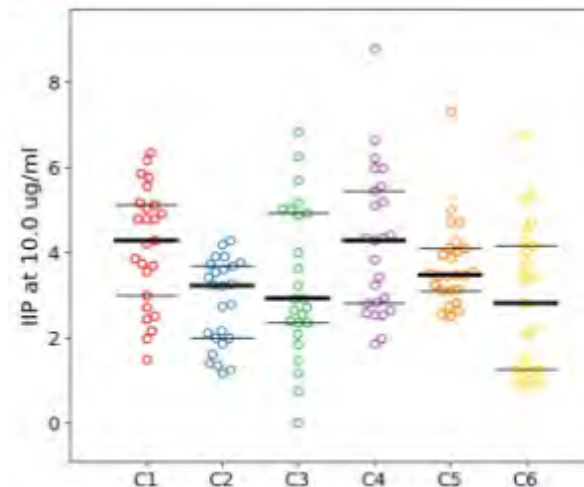
C1 CAP256-VRC26.25+VRC07-523-LS  
C2 10E8+CAP256-VRC26.25  
C3 10-1074V+CAP256-VRC26.25  
C4 10-1074V+VRC07-523-LS  
C5 10E8+VRC07-523-LS  
C6 10-1074V+10E8

Active coverage ?

At least 2 mabs active



IIP ?



C1 CAP256-VRC26.25+VRC07-523-LS  
C2 10E8+CAP256-VRC26.25  
C3 10-1074V+CAP256-VRC26.25  
C4 10-1074V+VRC07-523-LS  
C5 10E8+VRC07-523-LS  
C6 10-1074V+10E8

Next are shown figures measuring each neutralization metric, e.g. IC80 breadth-potency curves (top-left), active coverage (bottom left), incomplete neutralization (top right) and IIP (bottom right)

# CombiNAber Output Page

Similar analyses are shown for single mAbs, 3/4 mAb combinations and combinations of interest (if chosen).

Also, best 2-mAb combination vs best 3-mAb combination vs... are shown.

Detailed information:

<https://www.hiv.lanl.gov/content/sequence/COMBINABER/help.html>

Scientific reading:

Wagh et al. PLoS Pathogens 2016

<https://doi.org/10.1371/journal.ppat.1005520>

General Questions: [seq-info@lanl.gov](mailto:seq-info@lanl.gov); [immuno@lanl.gov](mailto:immuno@lanl.gov);

Technical Questions & Suggestions: [kshitij@lanl.gov](mailto:kshitij@lanl.gov)



## Purpose:

Find “signatures” (sequence features, i.e. amino-acids or glycans) statistically associated with any phenotype.

## Background:

- Bette Korber, Tanmoy Bhattacharya et al. developed a phylogenetically corrected strategy to identify amino acids and glycans significantly associated with any phenotype.

Bhattacharya et al. Science 2007 315(5818):1583-6

- We found robust amino acid and glycan signatures associated bNAb sensitivity/resistance. These signatures were used to design a trivalent vaccine cocktail that induced broad heterologous tier 2 neutralization in guinea pigs.

Bricault et al. Cell Host & Microbe 2019 25(1):59-72

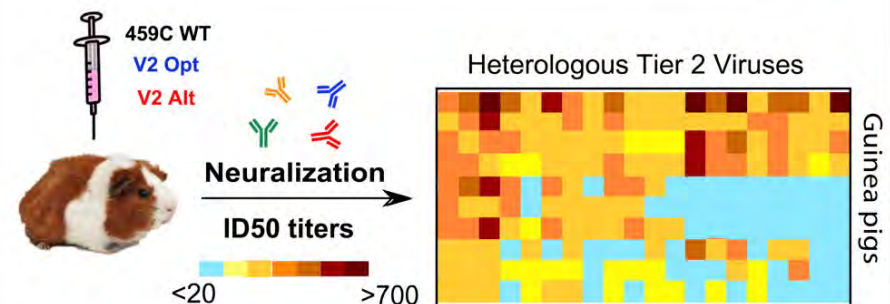
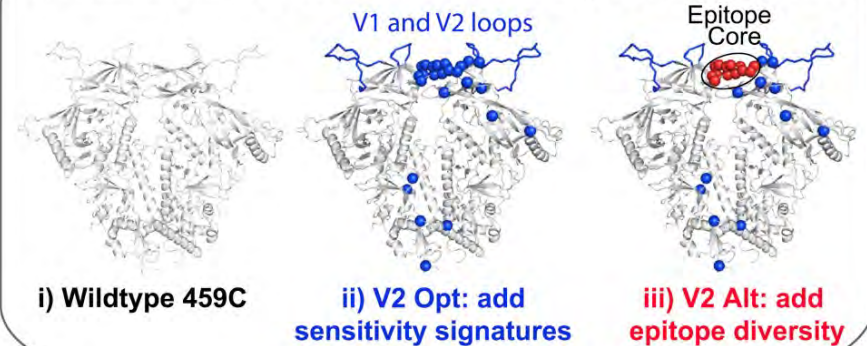
## Article

## HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design

Christine A. Bricault<sup>1, 17</sup>, Karina Yusim<sup>2, 3, 17</sup>, Michael S. Seaman<sup>1</sup>, Hyejin Yoon<sup>2</sup>, James Theiler<sup>2, 3</sup>, Elena E. Giorgi<sup>2, 3</sup>, Kshitij Wagh<sup>2, 3</sup>, Maxwell Theiler<sup>2</sup>, Peter Hraber<sup>2</sup>, Jennifer P. Macke<sup>2</sup>, Edward F. Kreider<sup>4</sup>, Gerald H. Learn<sup>4</sup>, Beatrice H. Hahn<sup>4</sup>, Johannes F. Scheid<sup>5, 6</sup>, James M. Kovacs<sup>7, 8, 9</sup>, Jennifer L. Shields<sup>1</sup>, Christy L. Lavine<sup>1</sup>, Fadi Ghantous<sup>1</sup>, Michael Rist<sup>1</sup>, Madeleine G. Bayne<sup>1</sup>, George H. Neubauer<sup>1</sup>, Katherine McMahan<sup>1</sup>, Hanqin Peng<sup>7, 8</sup>, Coraline Chéneau<sup>1</sup>, Jennifer J. Jones<sup>10</sup>, Jie Zeng<sup>10</sup>, Christina Ochsenbauer<sup>10</sup>, Joseph P. Nkolola<sup>1</sup>, Kathryn E. Stephenson<sup>1, 11</sup>, Bing Chen<sup>7, 8</sup>, S. Gnanakaran<sup>2, 3</sup>, Mattia Bonsignori<sup>12, 13</sup>, LaTonya D. Williams<sup>12</sup>, Barton F. Haynes<sup>12, 13, 14</sup>, Nicole Doria-Rose<sup>15</sup>, John R. Mascola<sup>15</sup>, David C. Montefiori<sup>12, 16</sup>, Dan H. Barouch<sup>1, 11, 18, 19</sup>, Bette Korber<sup>2, 3, 18</sup> ✉

### HIV-1 signatures were defined for 4 bNAb classes

#### 3-Valent SET vaccine with V2 bNAb signature mutations



# GenSig

## Inputs:

Codon-aligned nucleotide sequence alignment and matching phenotype features for the sequences.

- Can be any proteins and any phenotypes!

## Methods:

Each amino acid at each position in the alignment is tested for statistical association with phenotype.

- Option for potential N-linked glycosylation sites (PNGS).

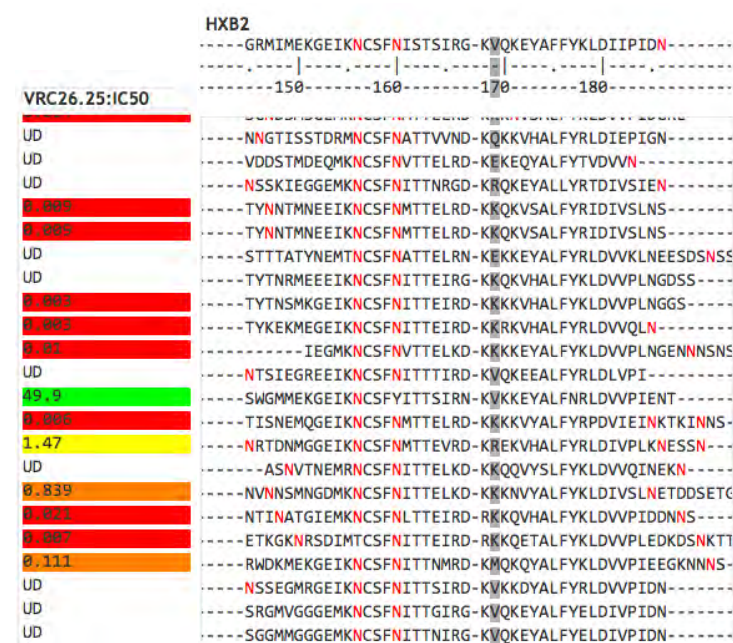
Fisher's exact test: For a binary phenotype (e.g. neutralized or not, above or below median), 2x2 contingency table with and without a given aa.

Wilcoxon test: Compare phenotype scores for sequences with and without aa.

Phylogenetic correction (next slide!)

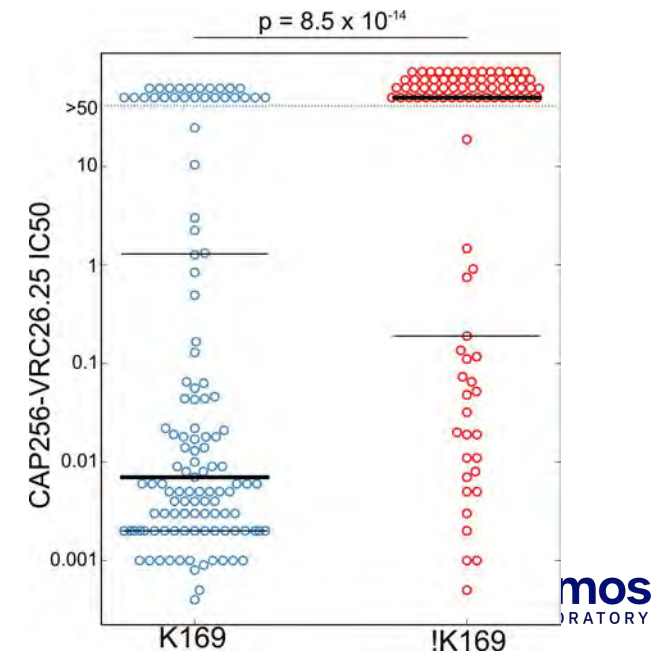
False discovery rate (q-value) correction for multiple tests! (for Env ~4000 tests!)

<https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/gs.html>



	K169	!K169
Neutralized	94	25
Not-neutralized	27	59

$$p = 6.52 \times 10^{-12}$$



mos  
LABORATORY

## Phylogenetic correction:

Simple associations can give strongly biased results due to clade effects.

Example: 10-1074 (and all V3g bNAbs) do not neutralize CRF01 strains due to lack of glycan at 332. Thus, any aa enriched in CRF01 will give false positive, e.g. H-375.

How to correct for this?

- Look for associations between amino acid transitions on a phylogenetic tree and the phenotype  
(Bhattacharya et al Science 2007 315:1583-6).
- If an amino acid is truly associated with the phenotype, then phylogenetic transition to this amino acid will also be associated.
- However, if the association arises due to a clade effect, then phylogenetic transition to this amino acid will not be associated.

### Clade-biased signature H-375

Uncorrected association

Table 1:	!H	H
10-1074 resistant	48	23
10-1074 sensitive	136	0
odds ratio = 0		
p-value = 1.1e-12		
q-value = 2e-10		

### Phylogenetic Signature Strategy

#### Clade-biased signature H-375

Table 2:	!H->H	!H->!H
10-1074 resistant	0	48
10-1074 sensitive	0	136
p-value = 1		

#### True Signature N-334

Table 2:	!N->N	!N->!N
10-1074 resistant	22	15
10-1074 sensitive	1	134
odds ratio = 0.02		
p-value = 6e-17		
q-value = 1.1e-14		




# GenSig Input Page

GenSig


Purpose: identify genetic signatures in a DNA alignment with associated phenotypic data. See [explanation](#).

## Analysis type


- Strategy 
- ☒ Full phylogenetic and signature analysis
  - ☐ New signature analysis using phylogenetic analysis from a previous run

## Sequence alignment and options

Codon-aligned DNA sequences  
[\[Sample Input\]](#)


Ambiguity codes 

- ☒ Replace with 'N'
- ☐ No ambiguity codes in sequences

Non allowable chars in names 

- ☒ Replace with '\_'
- ☐ No such characters in names


(; : ( ) , # | . -)

HIV-1 and HXB2 


- ☒ This is a HIV-1 alignment and the 1st seq is HXB2
- ☐ Not HIV-1 or HXB2 not included

Regions of interest


## Phenotype data and options

Phenotype data 

Delimiter ☐ Comma ☐ Space ☒ Tab

Mismatched names 

- ☐ Ignore mismatched sequence names between alignment and phenotype data

Statistical test 

- ☒ Fisher's test. No optimization need. Values are 1, 0, or -1.
- ☐ Fisher's test. Set highest decile to 1
- ☐ Fisher's test. Set highest quartile to 1
- ☐ Fisher's test. Set above median to 1
- ☐ Fisher's test. Set highest three quartiles to 1
- ☐ Fisher's test. Set highest nine deciles to 1
- ☐ Wilcoxon test. Values are continuous.

**Strategy:** new signature analysis or reuse an phylogenetic analysis from an older run (save time if your sequences are same, but phenotype features are different).

## Sequence alignment:


Can be HIV-1 or not. If HIV-1, then include HXB2 reference in your alignment. Otherwise the tool assumes no reference sequence.


## Phenotype data:


- Mismatch names – if your alignment or phenotype file have sequences that are not in the other, it will ignore them.
- Statistical test – choose Fisher's or Wilcoxon.
- For Fisher's user can supply binary phenotypes (0,1; with -1 being not tested), or supply actual values and use the options to decide how to break up the data (e.g. above or below median).
- For Wilcoxon, phenotype should have actual values.

# GenSig Input Page

**Signature options**

Signature analysis  ☒ sD1: Single site with depth of 1 ☐ glyco: Glycosylation analysis  
☐ sD2: Single site with depth of 2

Q-values  Maximum q-value

Unreliable positions  Positions with >  % gaps are excluded from the analysis

## Signature options:

“Depth”: Click for phylogenetic correction!

- Depth 1 treats all amino acids separately. Most commonly used.
- Depth 2 treats all pairs of 2 amino acids as a single group (e.g. Asp+Glu vs others). Rarely used, only when the user has a particularly strong reason for this.

“Glyco” – whether or not to test glycans (encoded by NxS/T motifs) as signatures.

## Maximum q-value:

The maximum cutoff for displaying signatures on the output page.

- 0.2 is a reasonable cutoff – 80% chance of true positive, 20% of false positive.

## Unreliable positions:

- Positions with lots of gaps indicate evolution by insertion/deletions, e.g. hypervariable loops in HIV-1 Env. Alignment and phylogeny in these regions are unreliable and thus, signatures cannot be reliably calculated. User can choose cutoff for removing such positions – 10% is default reasonable value.

# GenSig Output Page

Run ID: Use it for reusing phylogenetic analyses for a future job.

Parameters used.

Tree Results: Each blue underlined section is a link to download sequences or phylogenetic trees.

Maximum Likelihood Ancestor Results:  
These are the internal node states calculated for phylogenetic corrected signatures.

Signature Results: You can click to go to particular table results. Table 1 is simple Fisher's. Tables 2 & 3 are two flavors of phylogenetic corrected signatures.

The next two links for downloading the phenotype file and all signature results.

Run ID ⓘ

F2ukmBAh

Parameters used ⓘ

**Input sequences:** Reuse sequence data of Run ID k3Xtf1n1, Replace ambiguity codes with 'N'  
**Phenotype options:** Reuse phenotype data of Run ID k3Xtf1n1, Fisher test, Reoptimization = Not used  
**Signature options:** -sD1 -glyco, HXB2 included in input, Maximum qvalue = 0.3, Region(s) = Env\_gp160  
Replace non allowable chars in sequence names with '\_'

Tree Results ⓘ

[Download the sequences infile](#)  
[Download the Newick tree generated by PhyML](#)  
[Download the refined Newick tree](#)

Maximum Likelihood Ancestor Results ⓘ

[Download the maximum likelihood ancestor output as a compressed tar archive \(tar.gz\)](#)  
[Download the output maximum likelihood ancestor output treefile](#)

Signature Results ⓘ

Simple Analysis For a Single site with Depth of 1 (Fisher) [Table 1](#) [Table 2](#) [Table 3](#)  
Glycosylation Analysis (Fisher) [Table 1](#) [Table 3](#)

[Download the phenotype file used](#)  
[Download all signature results \(tar.gz\)](#)



Table 1:  
r1 = row 1 = feature (1)  
r2 = row 2 = feature (0)  
c1 = column 1 = 1A  
c2 = column 2 = A

[Download all unfiltered entries](#)  
[Download filtered entries with qvalue ≤ maximum qvalue \(0.3\)](#)  
[Download this table in a completed format](#)

[Download Env features and interactions with antibodies at signature sites](#)  
(For more information, check [Neutralizing antibody binding sites and other HIV-1 Env features](#))

Generate Tree using following options:

region	localaa	localnuc	test	feature	tree
--------	---------	----------	------	---------	------

Filtered entries with qvalue ≤ maximum qvalue (0.3)

tbl	statTest	HXB2aa	HXB2nuc	region	localaa	localnuc	testaa	feature	pvalue	r1c1	r1c2	r2c1	r2c2	qvalue	oddsRatio	Build Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	K	ch04	0.0249	61	22	66	48	0.133	2.01	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.268	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	<a href="#">Tree</a>
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	<a href="#">Tree</a>

Tbl = Type of test (if T1 simple; T2 and T3 phylogenetic)

statTest = Grouping of the phenotype data (I gave binary phenotype; but if I used above/below median option, then it would show here)

HXB2aa/HXB2nuc = reference amino acid position in HXB2 numbering or in HXB2 genome, respectively

Localaa/ localnuc = position in the alignment

<https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/gs.html>

Table 1 (Non-phylogenetically corrected results):

Table shows results from simple Fisher’s test that had q-value < cutoff provided.

User can download all entries (“unfiltered”), or just the filtered table shown below (both text and table formats).

User can also see which known antibodies have these signature sites below.

Table 1:  
r1 = row 1 = feature (1)  
r2 = row 2 = !feature (0)  
c1 = column 1 = !A  
c2 = column 2 = A

[Download all unfiltered entries](#)  
[Download filtered entries with qvalue ≤ maximum qvalue \(0.3\)](#)  
[Download this table in a completed format](#)  
[Download Env features and interactions with antibodies at signature sites](#)  
(For more information, check [Neutralizing antibody binding sites and other HIV-1 Env features](#))

Generate Tree using following options:  
region  localaa  localnuc  test  feature

Filtered entries with qvalue ≤ maximum qvalue (0.3)

tbl	statTest	HXB2aa	HXB2nuc	region	localaa	localnuc	testaa	feature	pvalue	r1c1	r1c2	r2c1	r2c2	qvalue	oddsRatio	Build Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	K	ch04	0.0249	61	22	66	48	0.133	2.01	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.286	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	<a href="#">Tree</a>
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	<a href="#">Tree</a>

Testaa = amino acid tested.  
This is A in the explanation.

Feature=phenotype

r1c1, r1c2, r2c1, r2c2 are the 2x2 contingency table entries.

	!M4	M4
1 (Neutralized)	50	33
0 (not neutralized)	96	18

oddsRatio = Odds of finding the phenotype with 'testaa' versus not testaa (!testaa).

- Measures association strength – close to 1 is weak association, far from 1 strong.
- Odds ratio (OR) should be interpreted with p-value – a low p-value does not guarantee good OR and vice versa. Best signatures are those with lowest p-values and good OR.



1: Simple Analysis For a Single site with Depth of 1 (Fisher)

Table 1:  
r1 = row 1 = feature (1)  
r2 = row 2 = !feature (0)  
c1 = column 1 = !A  
c2 = column 2 = A

[Download all unfiltered entries](#)  
[Download filtered entries with qvalue ≤ maximum qvalue \(0.3\)](#)  
[Download this table in a completed format](#)

[Download Env features and interactions with antibodies at signature sites](#)  
(For more information, check [Neutralizing antibody binding sites](#) and [other HIV-1 Env features](#))

Go to top

Generate Tree using following options:

region  localaa  localnuc  test  feature  tree

Filtered entries with qvalue ≤ maximum qvalue (0.3)

tbl	statTest	HXB2aa	HXB2nuc	region	localaa	localnuc	testaa	feature	pvalue	r1c1	r1c2	r2c1	r2c2	qvalue	oddsRatio	Build Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	K	ch04	0.0249	61	22	66	48	0.133	2.01	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.286	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	<a href="#">Tree</a>
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	<a href="#">Tree</a>
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	<a href="#">Tree</a>
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	<a href="#">Tree</a>
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	<a href="#">Tree</a>

Clicking "Tree" will generate the tree for your signature of interest:

- Tree shows the sequence at given position and phenotype for visual examination.

For simple signatures (table 1), it is advisable to look at the tree:

- Association could be due to a clade effect (e.g. resistant clade could be enriched in your signature).



# GenSig Output Page

Table 2:

r1 = row 1 = feature (1)  
r2 = row 2 = !feature (0)  
c1 = column 1 = !A->A  
c2 = column 2 = !A->!A

Other tables are similar in format, but they are phylogenetically corrected signatures!

- E.g. Table 2 measures whether phylogenetic transition to given test aa (!A -> A) is significantly associated with phenotype, as compared to !A -> !A.

Table 3:

r1 = row 1 = feature (1)  
r2 = row 2 = !feature (0)  
c1 = column 1 = A->!A  
c2 = column 2 = A->A

Similar outputs for glycan signatures:

Table T1g will be simple glycan associations and T3g will phylogenetically corrected glycan signatures.

[Download all unfiltered entries](#)

[Download filtered entries with qvalue ≤ maximum qvalue \(0.3\)](#)

[Download this table in a completed format](#)

[Download Env features and interactions with antibodies at signature sites](#)

(For more information, check [Neutralizing antibody binding sites and other HIV-1 Env features](#))

[Go to top](#)

Generate Tree using following options:

region  localaa  localnuc  test  feature

Filtered entries with qvalue ≤ maximum qvalue (0.3)

tbl	statTest	HXB2aa	HXB2nuc	region	localaa	localnuc	testaa	feature	pvalue	r1c1	r1c2	r2c1	r2c2	qvalue	oddsRatio	Build Tree
T3	Fisher:1,0,-1	<u>N160</u>	6702	Env_gp160	193	577	N	ch04	0.000761	0	83	13	101	0.113	0	<a href="#">Tree</a>
T3	Fisher:1,0,-1	<u>K171</u>	6735	Env_gp160	205	613	K	ch04	2e-08	4	77	36	54	1.84e-06	0.079	<a href="#">Tree</a>

# GenSig

GenSig is a powerful tool to identify statistically significant and phylogenetically corrected signatures in any protein for any phenotype.

Detailed information:

<https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/help.html>

Scientific reading:

Application to bNAbs: Bricault et al. Cell Host & Microbe 2019 <https://doi.org/10.1016/j.chom.2018.12.001>

Method development: Bhattacharya et al. Science 2017 <https://doi.org/10.1126/science.1131528>

Questions: seq-info@lanl.gov; immuno@lanl.gov

[www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html](https://www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html)

# Glycan Shield Mapping

## Background:

Several studies have shown that rare glycan holes in Env immunogens can induce immunodominant Ab responses targeting glycan holes that lack breadth.

We developed a strategy to accurately predict rare glycan holes given an input Env sequence.

(Wagh et al. Cell Rep 2018 25:893-908)

- Accurately predict known experimental glycan holes.
- TF Envs with complete glycan shields led to improved breadth development in HIV-1 infected individuals.

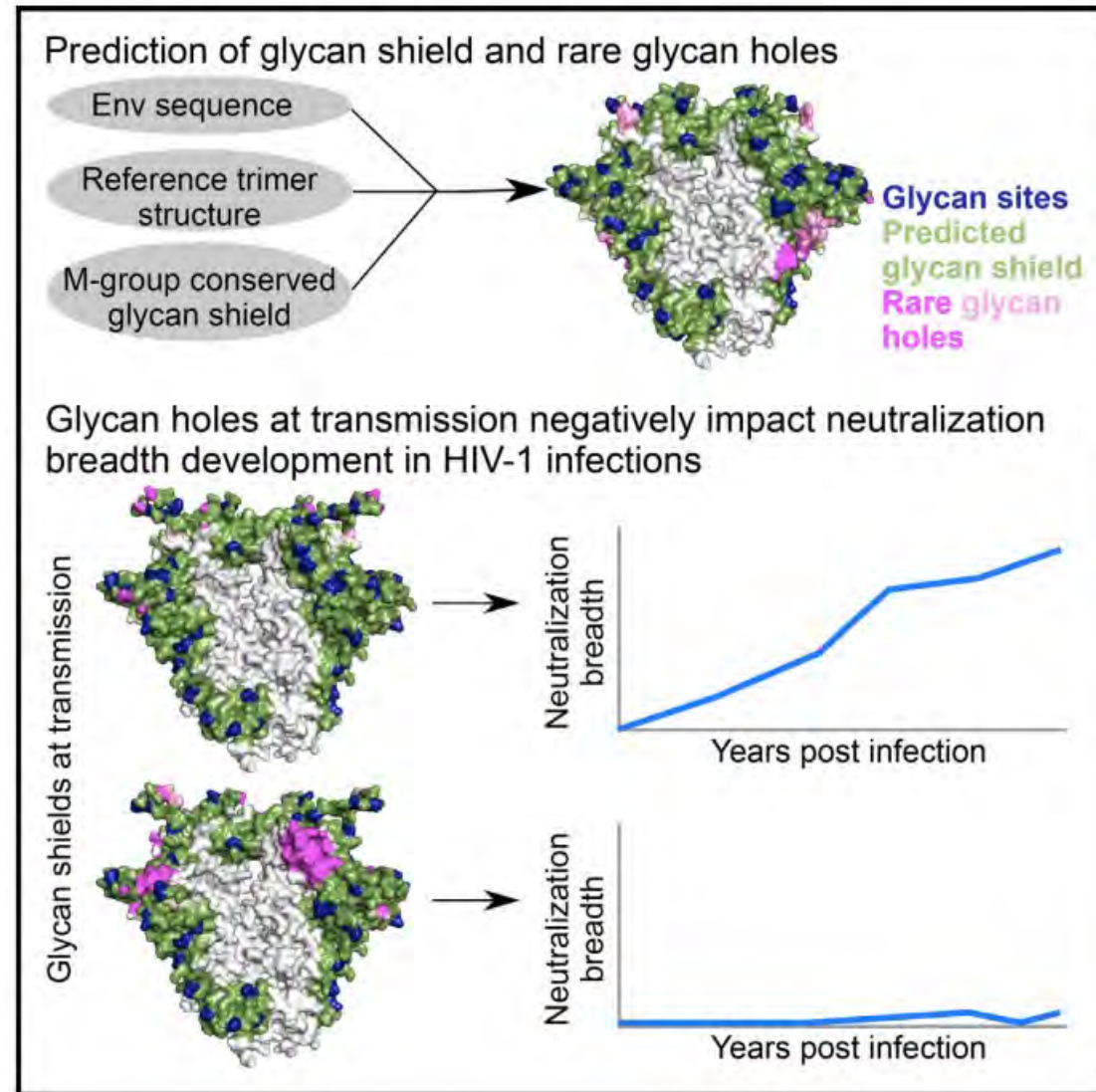
## Purpose:

To characterize glycans shields for Env sequences and identify missing glycans that cause rare glycan holes.

## Input:

Amino acid alignment of user HIV-1 Env sequences with HXB2.

- Sequences should span HXB2 sites 31 to 664 to ensure mapping on to Env trimer.





# Glycan Shield Mapping

## Outputs:

Link to download all results.

For each Env, total glycan hole area.

Graphical representation of glycan shields for each Env.

- Top row: glycan shield from different views. Magenta/pink = glycan holes. Green = predicted glycan shield.
- Bottom row shows missing glycans (cyan) that cause glycan holes.

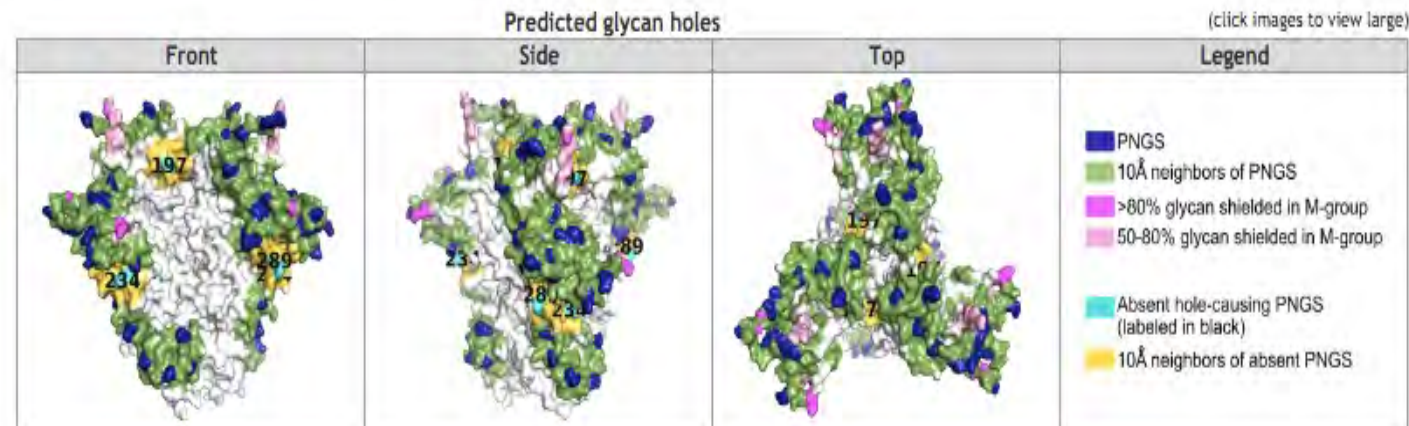
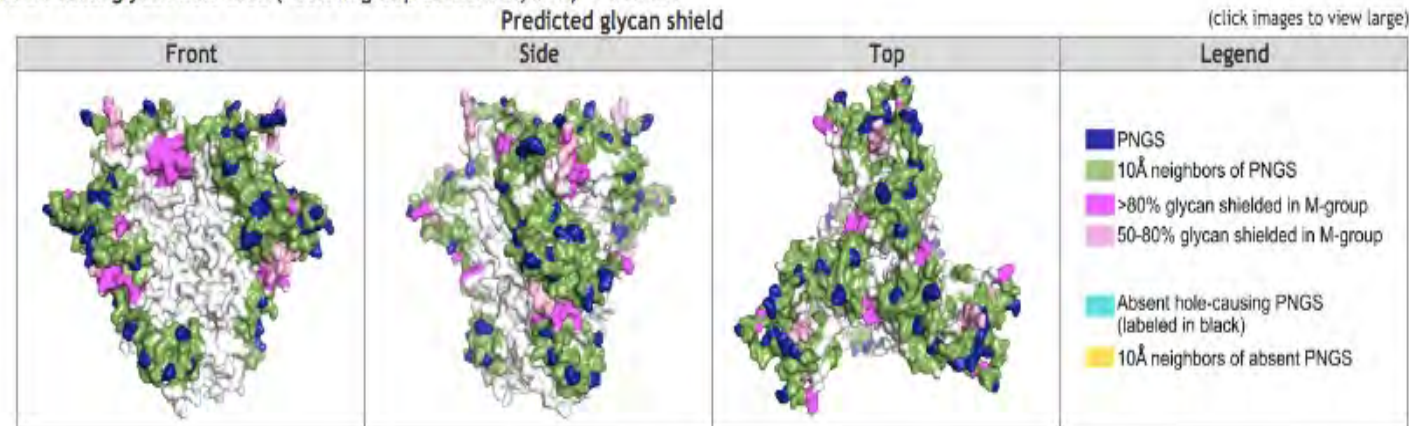
Glycan hole positions – surface-exposed amino acids that fall in glycan holes.

Table indicates missing glycans that lead to glycan holes, with glycan hole area attributed to each glycan and Env sites for each glycan hole.

Results:

[Download in an archived zip](#)

JRFL : Total glycan hole area (>50% M-group conserved, Å<sup>2</sup>) = 4721.37



Glycan hole positions: 93 94 95 126 163 164 165 182 183 184 192 193 194 195 196 197 198 199 200 231 232 233 234 236 237 238 265 266 267 268 269 270 271 272 273 285 286 287 288 289 290 308 309 312 313 314 315 348 411 412 423 433 459 466 484 485

Absent PNGS	Glycan hole area due to absent PNGS	Glycan hole area due to cumulative addition of PNGS	Glycan hole positions covered by absent PNGS
N197	1853.11	1853.11	126 163 164 165 182 183 184 192 193 194 195 196 197 198 199 200 308 309 312 313 314 315 423 433
N289	1641.65	1641.65	231 265 266 267 268 269 270 271 273 286 287 288 289 290 348 484
N234	1145.59	965.01	93 94 95 231 232 233 234 236 237 238 270 271 272 273 285 287 484 485

# Glycan Shield Mapping

Glycan Shield Mapping tool can help vaccine designers by identifying glycan holes in their immunogens and the require glycan mutations that will “fill” these glycan holes.

Detailed information:

<https://www.hiv.lanl.gov/content/sequence/GLYSHIELDMAP/help.html>

Scientific reading:

Wagh et al. Cell Reports 2018

<https://doi.org/10.1016/j.celrep.2018.09.087>

Questions: seq-info@lanl.gov; immuno@lanl.gov

Technical questions & suggestions: kshitij@lanl.gov

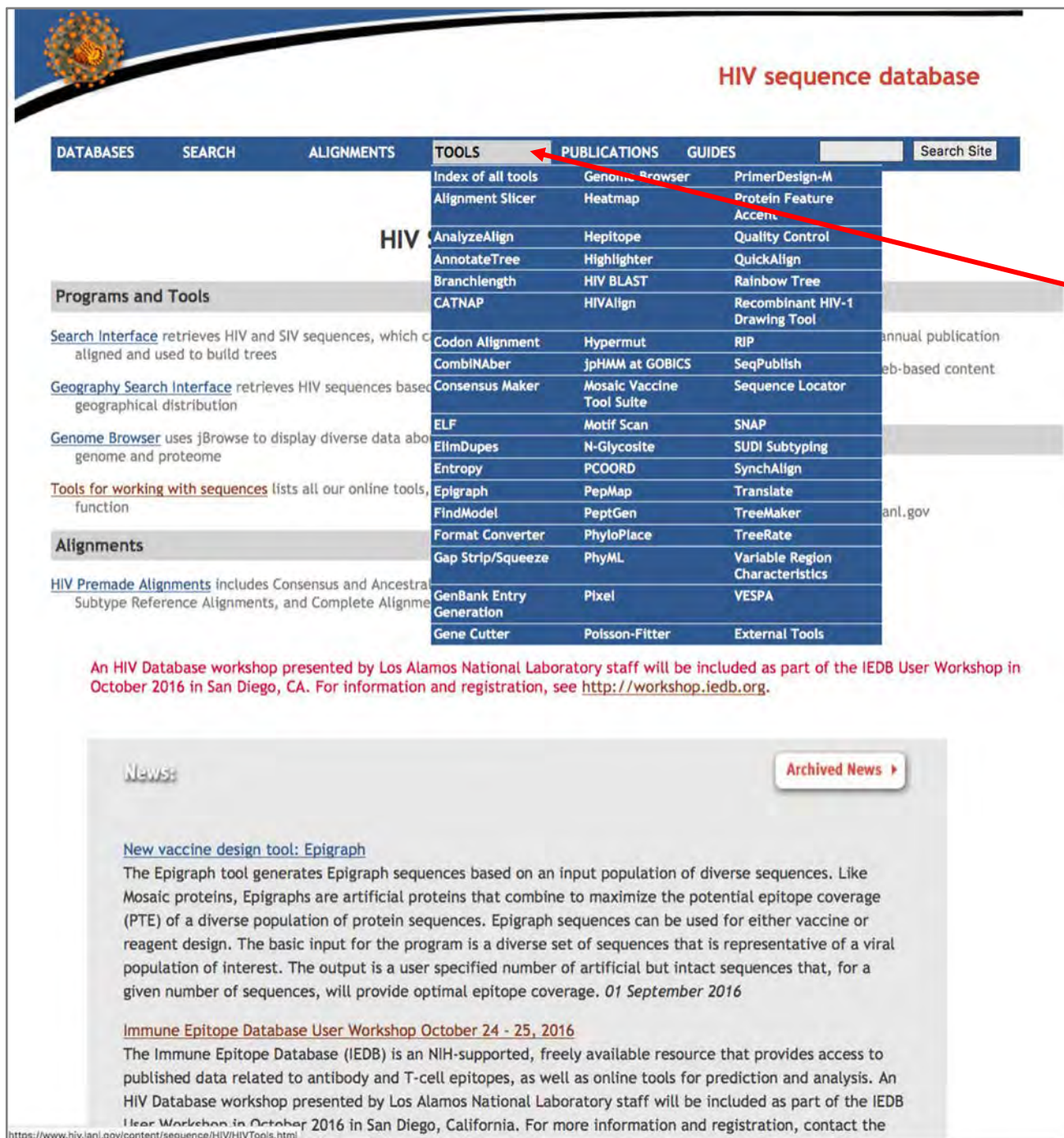
# Selected tools for Immunologists

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

- **CATNAP**: Compile, Analyze and Tally published and your own NAb Panels
- **CombiNAber**: 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 ...



# The HIV database sequence analysis tool set



**HIV sequence database**

**TOOLS**

Index of all tools	Genome Browser	PrimerDesign-M
Alignment Slicer	Heatmap	Protein Feature
AnalyzeAlign	Hepitope	Quality Control
AnnotateTree	Highlighter	QuickAlign
Branchlength	HIV BLAST	Rainbow Tree
CATNAP	HIVAlign	Recombinant HIV-1
Codon Alignment	Hypermut	Drawing Tool
CombinAber	jpHMM at GOBICS	Sequence Locator
Consensus Maker	Mosaic Vaccine	SNAP
ELF	Motif Scan	SUDI Subtyping
ElmDupes	N-Glycosite	SynchAlign
Entropy	PCOORD	Translate
Epigraph	PepMap	TreeMaker
FindModel	PeptGen	TreeRate
Format Converter	PhyloPlace	Variable Region
Gap Strip/Squeeze	PhyML	Characteristics
GenBank Entry	Pixel	VESPA
Generation	Poisson-Fitter	External Tools

**Programs and Tools**

[Search Interface](#) retrieves HIV and SIV sequences, which are 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 genome and proteome

[Tools for working with sequences](#) lists all our online tools, 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>.

**News:**

[New vaccine design tool: Epigraph](#)

The Epigraph tool generates Epigraph sequences based on an input population of diverse sequences. Like Mosaic proteins, Epigraphs are artificial proteins that combine to maximize the potential epitope coverage (PTE) of a diverse population of protein sequences. Epigraph sequences can be used for either vaccine or reagent design. The basic input for the program is a diverse set of sequences that is representative of a viral population of interest. The output is a user specified number of artificial but intact sequences that, for a given number of sequences, will provide optimal epitope coverage. 01 September 2016

[Immune Epitope Database User Workshop October 24 - 25, 2016](#)

The Immune Epitope Database (IEDB) is an NIH-supported, freely available resource that provides access to published data related to antibody and T-cell epitopes, as well as online tools for prediction and analysis. 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, California. For more information and registration, contact the

<https://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html>

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!

# HIV Immunology Tools are a subset of the HIV Sequence Tools

[www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)

HIV molecular immunology database

DatabasesSearchToolsProductsPublicationsSearch Site

HIV Molecular Immunology Database: Tools & Links

Tools Produced by the Los Alamos National Laboratory

- [CATNAP](#): Compile, Analyze, and Display HIV Sequences
- [HIV Genome Browser](#) Display HIV genome sequences
- [QuickAlign](#) Align amino acid sequences
- [Analyze Align](#) Show weblogs of alignments
- [Alignment Slicer](#) Cut vertical slices from alignments
- [PeptGen](#) Generate overlapping peptide sequences
- [PepMap](#) Generate peptide maps
- [Motif Scan](#) Scan alignments for motifs
- [Hepitope](#) Search for hepatitis B surface antigen epitopes
- [HLA Freq Analysis Tool](#) Analyze HLA frequencies by position, and find variants in an alignment
- [ELF](#) Epitope location finder
- [Sequence Locator](#) Find the location of any HIV/SIV sequence
- [SeqPublish](#) Produce pretty alignments for publication
- [Heatmap](#) Display a table of numbers using colors to represent the numerical values
- [Epigraph Vaccine Suite](#) Design and assess Epigraphs for vaccine design
- [Mosaic Vaccine Suite](#) Design and assess polyvalent protein sequences for T-cell vaccines
- [N-Glycosite](#) Find N-linked glycosylation sites
- [Highlighter](#) Highlight matches and mismatches in a set of aligned sequences
- [Protein Feature Accent](#) View 3D graphics of HIV proteins
- [Variable Region Characteristics](#) analyzes Env variable loops and reports length, glycosylations, and net charge
- [All Tools](#) List of all software and tools in both the HIV sequence and immunology databases

Tools especially useful from immunologists can be accessed from the HIV Immunology “Tools” page

## External Tools for Epitope Prediction

- [BIMAS HLA Peptide Binding Predictions](#) Ranks potential n-mer peptides based on a predicted half-time of dissociation to HLA class I molecules

<https://www.hiv.lanl.gov/content/immunology/tools-links.html>



# 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 exist in HFV database
  - Such numbers, often included in the literature, are frequently incorrect

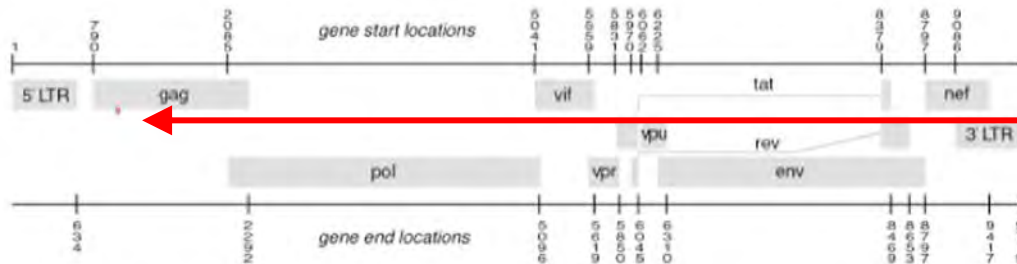
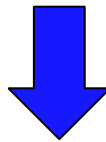
Find the location of a sequence

Sequence type ☒ Let program decide ☐ HIV ☐ SIV

Paste your input here  
[Sample Input]

SLYNTVATL

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.

CDS	AA position relative to protein start in HXB2	AA position relative to query sequence start	AA position relative to polyprotein start in HXB2	NA position relative to CDS start in HXB2	NA position relative to HXB2 genome start
Gag	77 → 85	1 → 9	NA	229 → 255	1018 → 1044
p17	77 → 85	1 → 9	NA	229 → 255	1018 → 1044

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

<https://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html>

Query SLYNTVATL 9  
: : : : : : :  
HXB2 SLYNTVATL



# 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

Enter coordinates: from  to  (Enter '1' and 'end' to retrieve the entire region.)

Region

Retrieve ☐ Nucleotide or ☒ protein output

☐ include surrounding region

Submit

Reset

Include surrounding region

Reference Strain	Type	Region	Start	End
HXB2	pro	complete	77	85
Retrieved Sequence: SLYNTVATL				

Reference Strain	Type	Region	Start	End
HXB2	pro	complete	56	106
Retrieved Sequence: GCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEE				

50 aa long stretch

<https://www.hiv.lanl.gov/content/sequence/LOCATE/locate.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

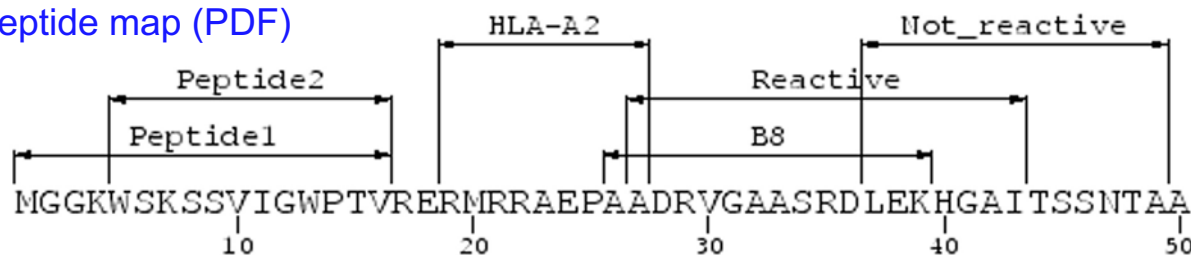
## Input:

Peptide1 MGGKWSASSVIGGPTV  
 Peptide2 WSKSSVIGWVTV  
 HLA-A2 RMRRAEPAV  
 B8 AADRVGAAASRDLEK  
 Reactive ADRVGAAASRDLEKHGAI  
 Not\_reactive LEKHGAITSSNTA

>B.FR.83.HXB2\_LAI\_IIIB\_BRU\_K03455 Peptide map (FASTA)  
 MGGKWSKSSVIGWPTVRERMRAEPAADRVGAAASRDLEKHGAITSSNTAA

>Peptide1  
 MGGKWSASSVIGGPTV-----  
 >Peptide2  
 ----WSKSSVIGWVTV-----  
 >HLA-A2  
 -----RMRRAEPAV-----  
 >B8  
 -----AADRVGAASRDLEK-----  
 >Reactive  
 -----ADRVGAASRDLEKHGAI-----  
 >Not\_reactive  
 -----LEKHGAITSSNTA-----

## Peptide map (PDF)



## Location table

Epitope Name	Query Peptide	Reference Peptide	Protein	AA position In Protein	Polyprotein	AA position In Polyprotein	Similarity%
Peptide1	MGGKWSASSVIGGPTV	MGGKWSKSSVIGWPTV	Nef	1-16	-	-	87.5
Peptide2	WSKSSVIGWVTV	WSKSSVIGWPTV	Nef	5-16	-	-	91.7
HLA-A2	RMRRAEPAV	RMRRAEPAA	Nef	19-27	-	-	88.9
B8	AADRVGAAASRDLEK	AADRVGAAASRDLEK	Nef	26-39	-	-	100.0
Reactive	ADRVGAASRDLEKHGAI	ADRVGAASRDLEKHGAI	Nef	27-43	-	-	100.0
Not_reactive	LEKHGAITSSNTA	LEKHGAITSSNTA	Nef	37-49	-	-	100.0

# PeptGen

<https://www.hiv.lanl.gov/content/sequence/PEPTGEN/peptgen.html>

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

Seq1 HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK  
Seq2 HLVWASRELERFALNPGLLETSEGCKQIIKQLQPALQTGTEELRSLYNTVATLYCVHEKIEVRDTKEALDKIEEEQNKSQ  
Seq3 HLVWASRELERFALNPDLLETAEGCQQIMGQLQPALQTGTEELRSLFNTVATLYCVHQRIEVKDTKEALEEVEKIQKKSQ

```
HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
HIVWASRELERFAVNPGL CON_B (18)
-L-----L---- CON_C
-L-----L--D- CON_G
    LERFAVNPGLLETSEGCR CON_B (18)
    -----L-----K CON_C
    -----L--D---A---Q CON_G
        GLLLETSEGCRQILGQLQP CON_B (18)
        -----K--IK---- CON_C
        D---A---Q--M----- CON_G
            CRQILGQLQPSLQTGSEE CON_B (18)
            -K--IK---A---T-- CON_C
            -Q--M-----A---T-- CON_G
                QPSLQTGSEELRSLYNTV CON_B (18)
                --A---T----- CON_C
                --A---T-----F--- CON_G
                    EELRSLYNTVATLYCVHQ CON_B (18)
                    -----E CON_C
                    -----F----- CON_G
                        TVATLYCVHQRIEVKDTK CON_B (18)
                        -----EK---R--- CON_C
                        ----- CON_G
                            HQRIEVKDTKEALEKIEE CON_B (18)
                            -EK---R-----D--- CON_C
                            -----EV-K CON_G
                                TKEALEKIEEEQNKSK CON_B (16)
                                -----D-----Q CON_C
                                -----EV-KI-K--Q CON_G
```

```
1 HIVWASRELERFAVNPGL 1 s1 1 s1 - -
2 HLVWASRELERFALNPGL 1 s2 1 - s2 -
3 HLVWASRELERFALNPDL 1 s3 1 - - s3

4 LERFAVNPGLLETSEGCR 2 s1 1 s1 - -
5 LERFALNPGLLETSEGCK 2 s2 1 - s2 -
6 LERFALNPDLLETAEGCQ 2 s3 1 - - s3

7 GLLLETSEGCRQILGQLQP 3 s1 1 s1 - -
8 GLLLETSEGCKQIIKQLQP 3 s2 1 - s2 -
9 DLLETAEGCQQIMGQLQP 3 s3 1 - - s3

10 CRQILGQLQPSLQTGSEE 4 s1 1 s1 - -
11 CKQIIKQLQPALQTGTEE 4 s2 1 - s2 -
12 CQQIMGQLQPALQTGTEE 4 s3 1 - - s3

13 QPSLQTGSEELRSLYNTV 5 s1 1 s1 - -
14 QPALQTGTEELRSLYNTV 5 s2 1 - s2 -
15 QPALQTGTEELRSLFNTV 5 s3 1 - - s3

16 EELRSLYNTVATLYCVHQ 6 s1 1 s1 - -
17 EELRSLYNTVATLYCVHE 6 s2 1 - s2 -
18 EELRSLFNTVATLYCVHQ 6 s3 1 - - s3

19 TVATLYCVHQRIEVKDTK 7 s1&s3 2 s1 - s3
20 TVATLYCVHEKIEVRDTK 7 s2 1 - s2 -

21 HQRIEVKDTKEALEKIEE 8 s1 1 s1 - -
22 HEKIEVRDTKEALDKIEE 8 s2 1 - s2 -
23 HQRIEVKDTKEALEEVEK 8 s3 1 - - s3
```



# ELF (Epitope Location Finder)

**ELF**  
Epitope Location Finder

**Purpose:** search a submitted protein sequence for (1) known epitopes from our immunology databases, (2) epitopes predicted by consensus binding motifs, and (3) epitopes predicted by the IEDB binding algorithm. For details see [ELF Explanation](#).

**Input**

Paste [protein sequence](#)  <50 amino acids, raw format

**Options**

Show [known epitopes](#) ☒ from CTL and Helper databases

Find potential epitopes ☒ based on [anchor residues](#)

Choose [HLA\(s\)](#)  
(Class I and Class II)  
Use control-click for multiple selection

By genotype

- A\*3004
- A\*3101
- A\*3201
- A\*3303
- A\*6601
- A\*6801
- A\*6802

By serotype

- A33(19)
- A69(28)
- A68(28)
- A30(19)
- A66(10)
- A1
- A2

Find potential epitopes ☒ based on [IEDB binding predictions](#)

Choose [HLA\(s\) or MHC\(s\)](#)  
(synchronized with genotype selections above)

HLA Class I

- A\*6611
- A\*6612
- A\*6613
- A\*6614
- A\*6615
- A\*6801
- A\*6802

HLA Class II

- DRB3\*0224
- DRB3\*0225
- DRB3\*0301
- DRB3\*0303
- DRB4\*0101
- DRB4\*0103
- DRB5\*0101

Animal MHC Class I

*chimpanzee*

- Patr-A\*0101
- Patr-A\*0201
- Patr-A\*0301
- Patr-A\*0302
- Patr-A\*0401
- Patr-A\*0402

Animal MHC Class II

*mouse*

- H2-IAb
- H2-IAd
- H2-IEd

Display binders ☒ Show  best binder(s) per MHC


☐ Show below  percentile rank (1-100) per MHC

E-mail result ☐ Predictions are slow. For more than a few HLAs/MHCs, we recommend e-mailed result.

- ELF helps identify potential T cell epitopes in a reactive peptide from a person with known HLA type:
  - Highlights appropriate HLA anchor motifs in the peptide
  - Aligns reported epitopes embedded in the peptide from the database to your query sequence, with links to epitope entries
  - Finds potential epitopes based on Immune Epitope Database (IEDB) binding predictions  
<http://www.immuneepitope.org/>
- The **MotifScan** tool shows HLA binding and custom motifs on the sequence alignment

# 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  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.

Download this alignment in format

DTVLEDMNLPGRWKPKMIG

[DTVLE\*\*EM\*\*NL](#) A\*6802

[align](#)

 Epitopes

[DTVLE\*\*I\*\*NL](#) A\*6802

[align](#)

 matching

[DTVLE\*\*EW\*\*NL](#) A\*6802

[align](#)

 requested HLAs

[DTVLE\*\*EM\*\*NL](#) A68

[align](#)

[DTVLE\*\*EM\*\*NL](#) A28

[align](#)

[DTVLEDMNL](#)

[align](#)

[E\*\*E\*\*MNLPGRW](#) B44

[align](#)

[E\*\*E\*\*I\*\*N\*\*LPGKW](#) B44

[align](#)

[E\*\*E\*\*MNLPGRW](#) B\*4402

[align](#)

[E\*\*E\*\*MNLPGRW](#) B\*4403

[align](#)

[E\*\*E\*\*MNLPGRW](#) B18,B40,B44

[align](#)

[EDMNLPGRW](#)

[align](#)

[E\*\*E\*\*MNLPGRW](#) B\*44

[align](#)

[E\*\*E\*\*I\*\*N\*\*LPGKW](#) B\*4403

[align](#)

[E\*\*E\*\*MNLPGRW](#)

[align](#)

[LPGRWKPKMI](#) Cw3

[align](#)

[LPGRWKPKMI](#) B7

[align](#)

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.

this alignment in format

```
DTVLEDMNLPGRWKPKMIG
DTVLEDMNL (A*0205 .....[L])
DTVLEDMNL (A*6802 .[TV].....[VL])
TVLEDMNLP (A*0206 .[VQ].....)
LEDMLNLPGR (DRB5*0101,DRB5*0101 [FYLM]..[QVIM]....[RK])
```

[https://www.hiv.lanl.gov/content/immunology/motif\\_scan/motif\\_scan](https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan)

**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

this alignment in format

DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)

DMNLPGRW	<a href="#">B*1501</a> (26)
MNLPGRWK	<a href="#">A*6802</a> (3.0)

### Class II

Selected allele(s): DRB5\*0101

this alignment in format

DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)

TVLEDMNLPGRWKPK	<a href="#">DRB5*0101</a> (17.17)
-----------------	-----------------------------------

**IEDB binding predictions**

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

## IEDB Analysis Resource

[Home](#)[Help](#)[Example](#)[Reference](#)[Download](#)[Contact](#)

## MHC-I binding predictions - Prediction Results

### Input Sequences

#	Name	Sequence
1	sequence 1	DTVLED MNLPGRWKPKMIG

Prediction method: IEDB recommended | Low percentile = good binders

Check to expanded the result: ☐

Allele	#	Start	End	Peptide Length	Sequence	Method used	Percentile Rank
HLA-B*15:01	1	6	13	8	DMNLPGRW	NetMHCpan	26
HLA-B*15:01	1	3	13	11	VLED MNLPGRW	NetMHCpan	27
HLA-B*15:01	1	3	11	9	VLED MNLPG	Consensus (ANN,SMM,CombLib_Sidney2008)	27.60
HLA-B*15:01	1	8	17	10	NLPGRWKPKM	NetMHCpan	31
HLA-B*15:01	1	7	17	11	MNLPGRWKPKM	NetMHCpan	35
HLA-B*15:01	1	2	9	8	TVLED MNL	NetMHCpan	36
HLA-B*15:01	1	2	11	10	TVLED MNLPG	NetMHCpan	47
HLA-B*15:01	1	4	11	8	LED MNLPG	NetMHCpan	48

# QuickAlign

[https://www.hiv.lanl.gov/content/sequence/QUICK\\_ALIGNv2/QuickAlign.html](https://www.hiv.lanl.gov/content/sequence/QUICK_ALIGNv2/QuickAlign.html)

- 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

Query:	SLYNTVATL
Query Length:	9
HXB2 Location:	Gag 77-85 = p17 77-85
Alignment:	GAG, 458 sequences

Summarize

Query	SLYNTVATL
A1.KE.86.ML170	--F-----
A1.KE.94.Q23	--F-----
A1.SE.94.SE7253	--F----V-
A1.SE.94.SE7535	-----
A1.SE.95.SE8538	-----
A1.SE.95.SE8891	-----
A1.SE.95.UGSE8131	-----
A1.TZ.97.97T203	--F----V-

Summary for subtype A

Variant	Count	Percent
SLYNTVATL		
--F-----	11	47.83
-----	7	30.43
--F--I-V-	1	4.35
--F---V-	1	4.35
-----V-	1	4.35
----L----	1	4.35
--F-A--V-	1	4.35

Total sequences = 23  
Number of variants = 7

Variant  
frequency  
summary

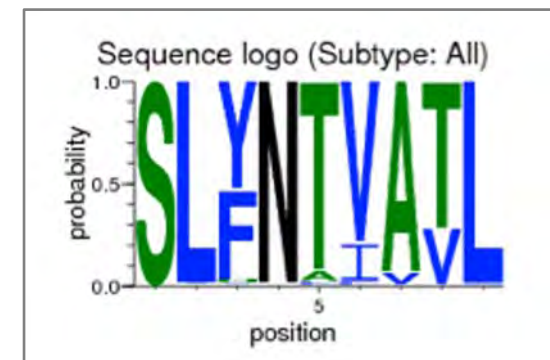
## Frequency by position

[See full raw counts](#)

[Go to top](#)

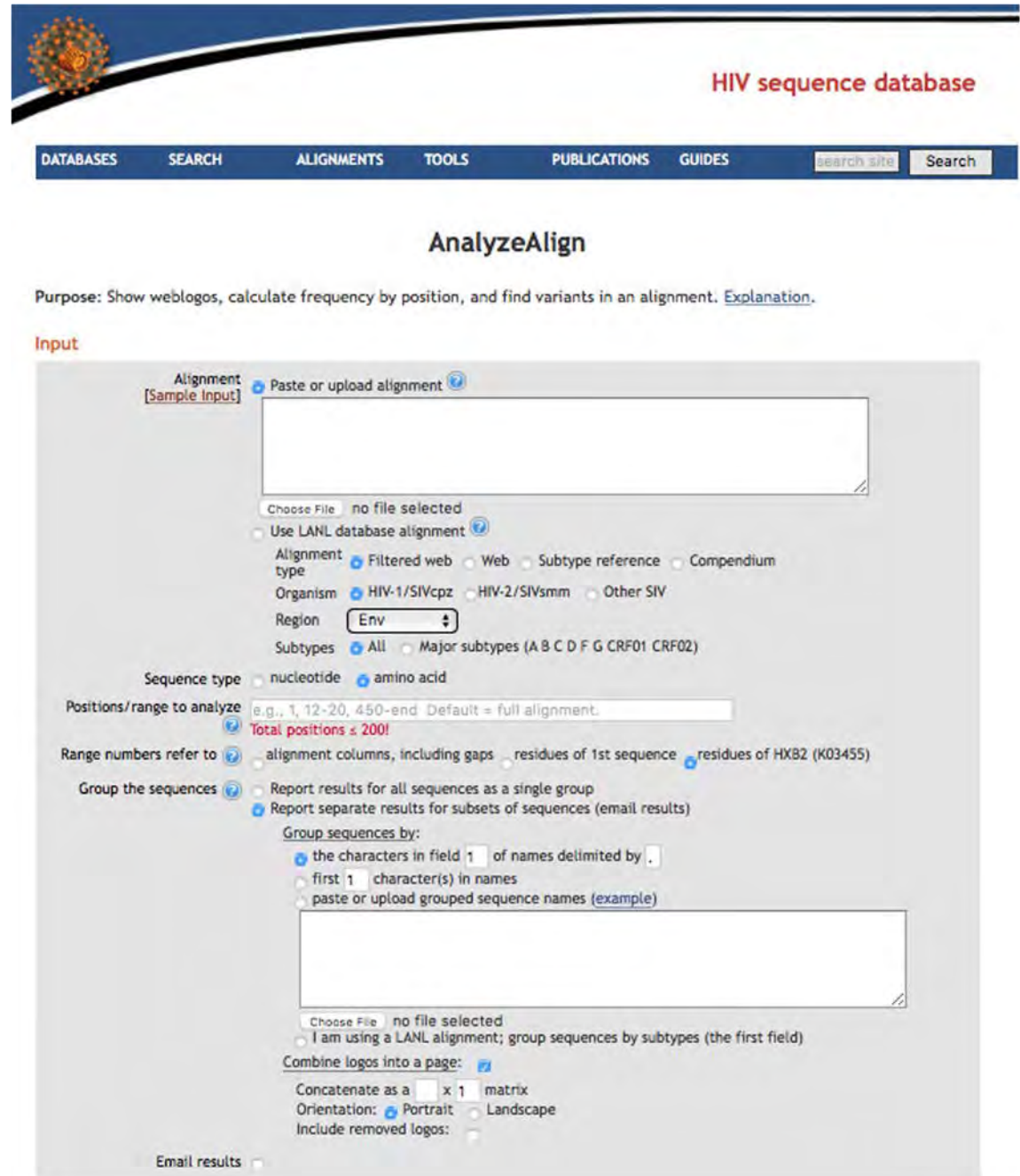
cutoff: 95%

Position	Percentage and raw count of non-gap			Non-gap/total (percentage)
1	S: 99.90% (3113)	other: 0.10% (3)		3116/3119 (100.00%)
2	L: 98.90% (3068)	other: 1.10% (34)		3102/3119 (99.55%)
3	Y: 52.71% (1633)	F: 43.77% (1356)	other: 3.52% (109)	3098/3119 (99.42%)
4	N: 99.68% (3104)	other: 0.32% (10)		3114/3119 (99.94%)
5	T: 92.86% (2887)	A: 5.05% (157)	other: 2.09% (65)	3109/3119 (99.78%)
6	V: 79.35% (2448)	I: 18.15% (560)	other: 2.50% (77)	3085/3119 (99.01%)
7	A: 92.95% (2889)	V: 6.53% (203)	other: 0.51% (16)	3108/3119 (99.74%)
8	T: 72.52% (2254)	V: 27.06% (841)	other: 0.42% (13)	3108/3119 (99.74%)
9	L: 99.00% (3078)	other: 1.00% (31)		3109/3119 (99.78%)



# AnalyzeAlign

- Go to tools drop down menu, select Analyze Align
  - Click on “use LANL database”
    - Could enter your own alignment, it doesn’t need to be HIV
  - Click on “Major subtypes”
  - “Positions/range to analyze”
    - 324-326,327, 331-336
  - Click on “Logo options”
    - At the bottom click on Mark potential N-linked glycosylation sites



The screenshot shows the 'AnalyzeAlign' tool interface on the HIV sequence database website. The top navigation bar includes links for DATABASES, SEARCH, ALIGNMENTS, TOOLS, PUBLICATIONS, and GUIDES, along with a search bar. The 'AnalyzeAlign' section has a purpose statement: 'Purpose: Show weblogs, calculate frequency by position, and find variants in an alignment. [Explanation.](#)'

The 'Input' section contains several options for alignment input and analysis parameters:

- Alignment [Sample Input]:** Includes a text area for pasting or uploading an alignment, a 'Choose File' button, and a 'no file selected' status.
- Use LANL database alignment:** Includes radio buttons for 'Filtered web', 'Web', 'Subtype reference', and 'Compendium'.
- Alignment type:** Includes radio buttons for 'Filtered web', 'Web', 'Subtype reference', and 'Compendium'.
- Organism:** Includes radio buttons for 'HIV-1/SIVcpz', 'HIV-2/SIVsmm', and 'Other SIV'.
- Region:** Includes a dropdown menu set to 'Env'.
- Subtypes:** Includes radio buttons for 'All', 'Major subtypes (A B C D F G CRF01 CRF02)', and 'Other SIV'.
- Sequence type:** Includes radio buttons for 'nucleotide' and 'amino acid'.
- Positions/range to analyze:** Includes a text input field with the value 'e.g., 1, 12-20, 450-end' and a 'Default = full alignment.' label.
- Range numbers refer to:** Includes radio buttons for 'alignment columns, including gaps', 'residues of 1st sequence', and 'residues of HXB2 (K03455)'.
- Group the sequences:** Includes radio buttons for 'Report results for all sequences as a single group' and 'Report separate results for subsets of sequences (email results)'.
- Group sequences by:** Includes radio buttons for 'the characters in field 1 of names delimited by .', 'first 1 character(s) in names', and 'paste or upload grouped sequence names (example)'.
- Combine logos into a page:** Includes a 'Combine logos into a page:' checkbox and a 'Concatenate as a x 1 matrix' label.
- Orientation:** Includes radio buttons for 'Portrait' and 'Landscape'.
- Include removed logos:** Includes a checkbox.
- Email results:** Includes a checkbox.



# AnalyzeAlign Output

## Groups

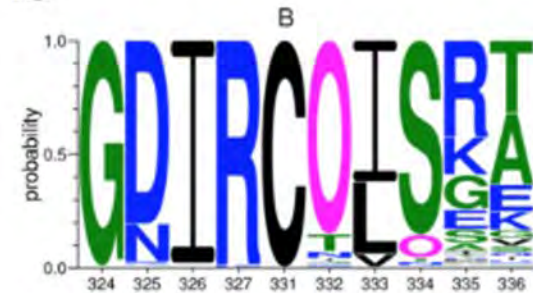
[Download combined logs [PDF](#) [EPS](#)]

B [A1](#) [A2](#) [C](#) [D](#) [F1](#) [F2](#) [G](#) [O1\\_AE](#) [O2\\_AG](#)

## Group B

[Go to top](#)

## Logo



Download: [PNG](#) [PDF](#) [EPS](#)

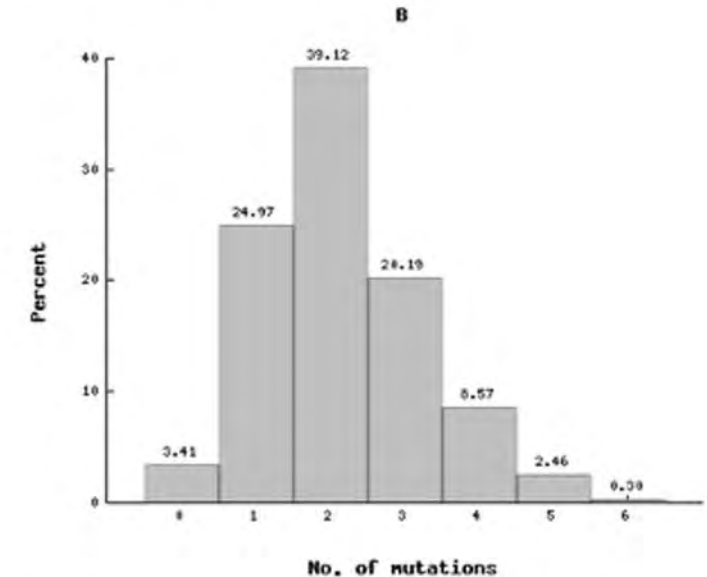
## Frequency by position

[See full raw counts](#)

cutoff: 95%

	Percentage and raw count of non-gap	Non-gap/total (percentage)	Gap/total (percentage)
324	G: 99.43% (1924) other: 0.57% (11)	1935/1937 (99.90%)	2/1937 (0.10%)
325	D: 80.84% (1565) N: 16.63% (322) other: 2.53% (49)	1936/1937 (99.95%)	1/1937 (0.05%)
326	I: 98.24% (1900) other: 1.76% (34)	1934/1937 (99.85%)	3/1937 (0.15%)
327	R: 98.76% (1913) other: 1.24% (24)	1937/1937 (100.00%)	0/1937 (0.00%)
331	C: 99.90% (1935) other: 0.10% (2)	1937/1937 (100.00%)	0/1937 (0.00%)
332	O: 85.23% (1651) T: 7.80% (151) N: 2.89% (56) other: 4.08% (79)	1937/1937 (100.00%)	0/1937 (0.00%)
333	I: 62.26% (1206) L: 31.80% (616) V: 5.73% (111) other: 0.21% (4)	1937/1937 (100.00%)	0/1937 (0.00%)
334	S: 85.54% (1657) O: 10.22% (198) other: 4.23% (82)	1937/1937 (100.00%)	0/1937 (0.00%)
335	R: 42.72% (827) K: 16.48% (319) G: 15.34% (297) E: 8.68% (168) S: 6.30% (122) A: 3.05% (59) I: 1.91% (37) T: 1.03% (20) other: 4.49% (87)	1936/1937 (99.95%)	1/1937 (0.05%)
336	T: 32.27% (625) A: 31.23% (605) E: 11.82% (229) K: 8.21% (159) G: 3.67% (71) V: 3.30% (64) S: 2.74% (53) Q: 1.91% (37) other: 4.85% (94)	1937/1937 (100.00%)	0/1937 (0.00%)

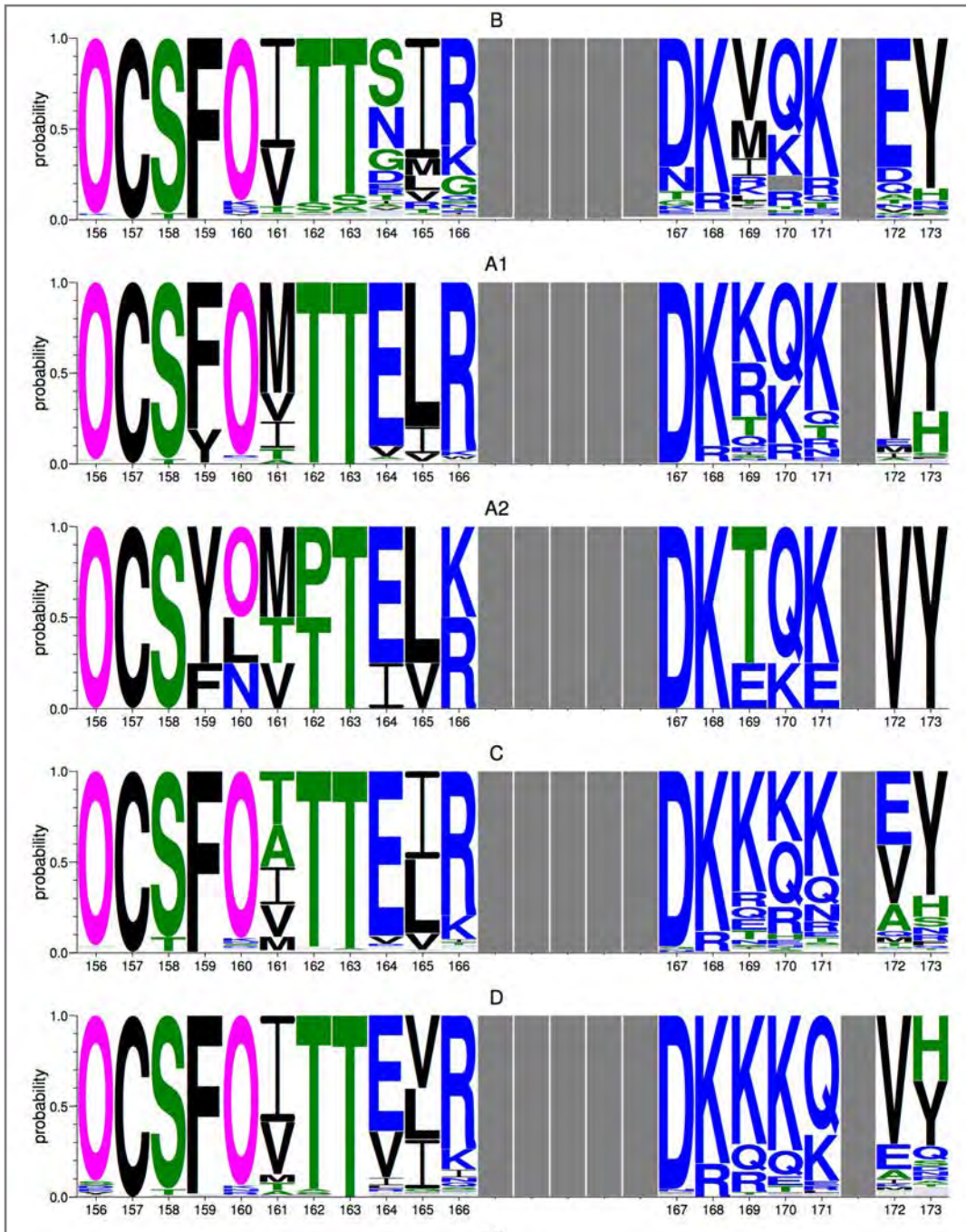
## Sequence variants



Variant	Count	Pct.	No. of mutations
GDIRCOISRT			
---A	134	6.92	1
---L---A	89	4.59	2
---	66	3.41	0
---E	65	3.36	1
---KA	57	2.94	2
---L---	51	2.63	1
---G---	47	2.43	1
---L-S-	33	1.7	2
---K	31	1.6	1
---EA	31	1.6	2
-N---	28	1.45	2
---GA	27	1.39	2
---GE	27	1.39	2
---K-	26	1.34	1
---KE	24	1.24	2
---L---E	20	1.03	2
---TLOG-	19	0.98	4

# MAb PG9 binding regions, Env 156-173, bNAb PG9 contact region

## 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
  - WebLogo color scheme
  - Combining multiple logos on a page
  - Showing potential N-linked glycosylation sites (Nx[ST], denoted as **O**)



# AnalyzeAlign



Transmitted HIV virus

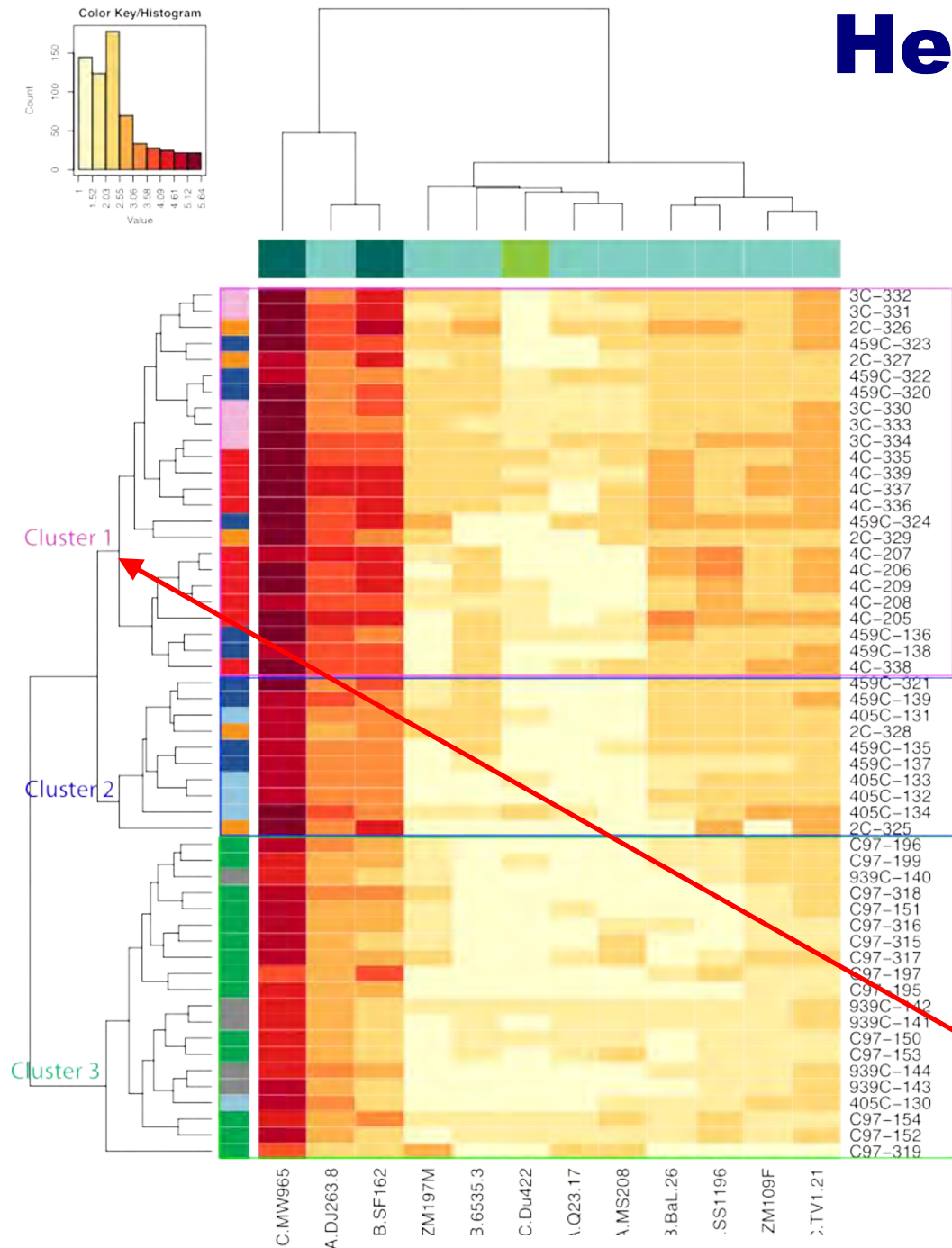
- ☐ Longitudinal samples over time
- ☐ Discontinuous positions under apparent immune selection
- ☐ Only differences from transmitted virus are shown
- ☐ Colors indicate amino acid charge categories

Virus 3 years post-infection

Figure from Hraber et al. *Viruses* 2015

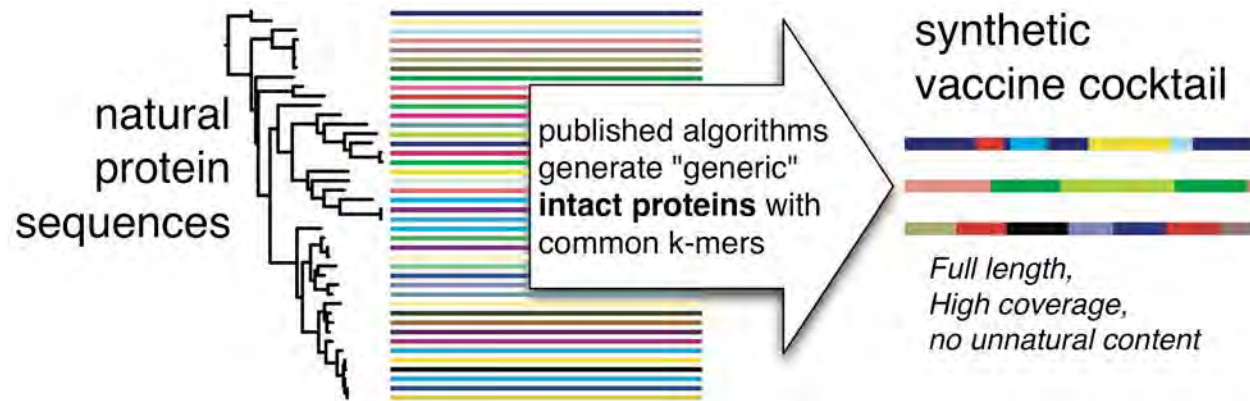


# Heatmap



- 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

# Vaccine Design Tools (Mosaic/Epigraph)



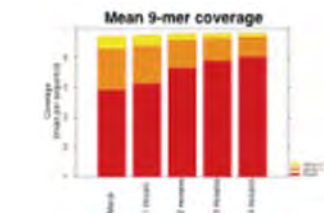
## Design Tools

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)

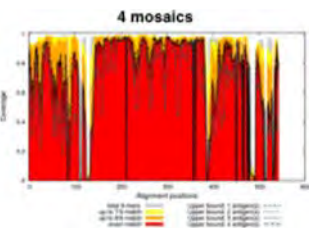
**Epigraph — graph theoretic approach** (Theiler et al. 2016)

## Evaluation tools



### **Epitope Coverage Assessment (EPICOVER)**

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



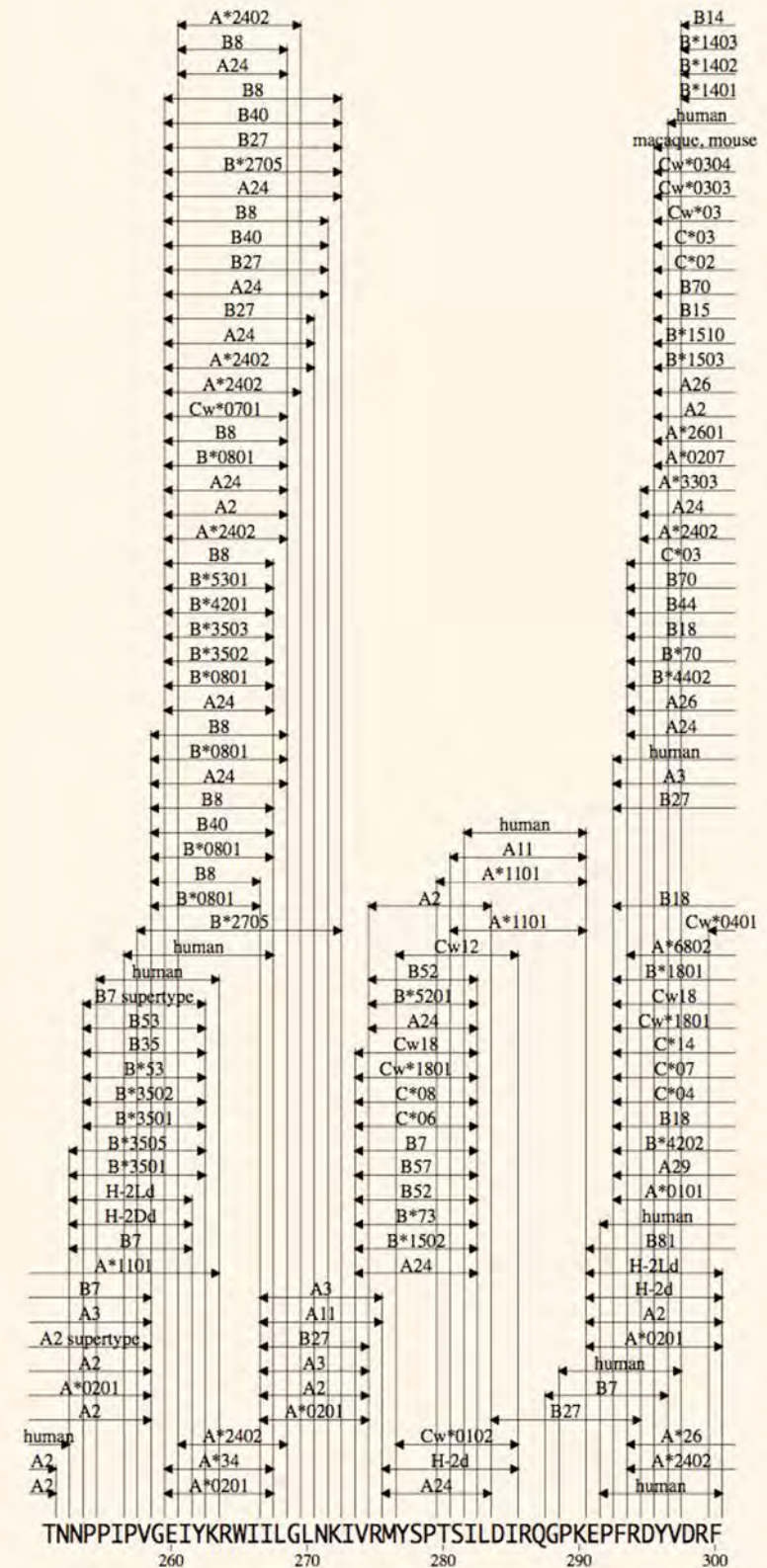
### **Positional Epitope Coverage Assessment (POSICOVER)**

Alignment-based coverage by vaccines or peptides

<https://www.hiv.lanl.gov/content/sequence/MOSAIC/>

# 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





# Mosaic Vaccine Designer

**HIV sequence database**

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

## Mosaic Vaccine Designer

**Purpose:** The Mosaic Vaccine Designer will generate candidate vaccine protein cocktails that optimize the coverage, by a small set of mosaic proteins that could be included in a vaccine cocktail, of potential T-cell epitopes in a large diverse set of proteins. The resulting 'mosaic' proteins in the proposed vaccine cocktail resemble real proteins from the input set of natural viral proteins (the 'training set'), but are assembled from fragments of the natural proteins using a genetic algorithm (a computational optimization method). This method was first applied to HIV, but is readily generalized and could be applied to other variable pathogens.

**Functions:**

- 'Create mosaic sequence cocktail' runs the genetic algorithm to generate a cocktail of synthetic sequences with near-optimal coverage
- 'Pick the best natural sequences' selects unmodified natural sequences from the training set in order of coverage
- 'See the coverage distribution of natural sequences' shows the coverages of a randomly selected set of natural sequence cocktails

**Usage:** Paste your protein sequences in the box below, or upload a file containing sequences. Most common [sequence formats](#) are accepted. As soon as your job is completed, a link to your results will be sent to your email address which you provided. To manage more detailed parameters, go to the Advanced Input. (Your job may take several hours or even days, according to your input.)

**Related Programs:**

- [Epitope Coverage Assessment Tool-Epicover](#)
- [Positional Epitope Coverage Assessment Tool-Posicover](#)

**Reference:** [Polyvalent vaccine design article](#) | [Pubmed version](#)

### Input

Paste set of protein sequences  
☒ Sample Input

```
A1.CM...a
MGGNWSKSSLVGWPEIRERMRRAPPTPTTPPAKGVGAVSQDLAKHGAI
A1.KE.99a
MGGKWSKSSIIVGWPEVRRRIQQTTPAARGVGAVSQDLEKHGAITSSNINHS
A1.KE.99b
MGGIWSKRSTRGWSEVRERIRQTTPPAARGVGAVSQDLARHGAVTSSNVN
```

Or upload protein sequence file

### Options

**Basic** **Advanced**

**Function** ☒ Create mosaic sequence cocktail  
☐ Pick the best natural sequences  
☐ See the coverage distribution of natural sequences

Cocktail Size (1-10)

Epitope Length (8-12)

Rare Threshold

Paste fixed sequences

Or upload fixed sequence file

## 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

## DEMONSTRATED EFFECTIVENESS

Improved immunogenicity

HIV, SIV, HCV, *Chlamydia*

Protection from challenge (non-human models):

SHIV, Influenza, FMDV, Ebola

*Phase IIb Human HIV trial (Imbokodo) recently completed (no efficacy)*

*Phase III HIV trial (Mosaico; HVTN 706/HPX3002) in progress*

# 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

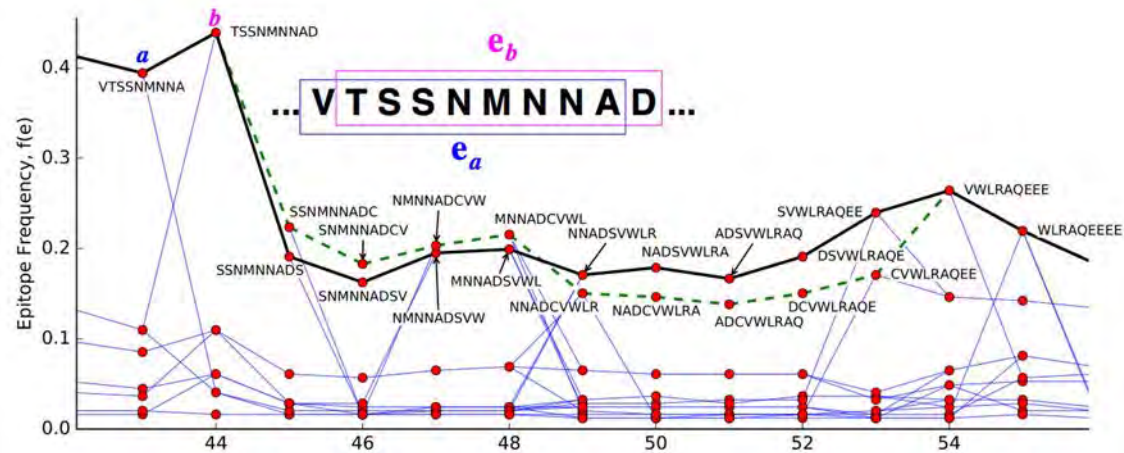
## Advantages over mosaic

**Essentially optimal** (fractionally better coverage)

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


**Reference:** Theiler, J., Yoon, H., Yusim, K., Picker, L. J., Fruh, K., and Korber, B. (2016). Epigraph: A vaccine design tool applied to an HIV therapeutic vaccine and a pan-filovirus vaccine. *Sci Rep*, 6:33987.

<https://www.hiv.lanl.gov/content/sequence/EPIGRAPH/epigraph.html>

The image shows the EPIGRAPH web interface with the "Design" tab selected. The interface includes a navigation bar with links: Design, Antigen Coverage Evaluation, Coverage Distribution of Natural Sequences, Exclude rare epitopes, Characterize Potential T-cell Epitopes, Design: Tailored Therapeutic Vaccine, and Evaluation: Tailored Therapeutic Vaccine. The "Input" section has a text box for "Protein sequences" with a "Sample Input" link and a "Choose File" button. The "Options" section includes a radio button for "Unaligned sequence algorithm" (selected) and a radio button for "Aligned sequence algorithm". It also has dropdown menus for "Epitope length" (set to 9) and "# of seqs in vaccine pool" (set to 2), and a checkbox for "Email results".



# EPIGRAPH — exclude rarities



## HIV sequence database

DATABASES   SEARCH   ALIGNMENTS   TOOLS   PUBLICATIONS   GUIDES

### Epigraph

[Design ?](#)   [Antigen Coverage Evaluation ?](#)   [Coverage Distribution of Natural Sequences ?](#)   **[Exclude rare epitopes ?](#)**   [Characterize Potential T-cell Epitopes ?](#)   [Design: Tailored Therapeutic Vaccine ?](#)   [Evaluation: Tailored Therapeutic Vaccine ?](#)

#### Input

Protein sequences

>B.US.2006.101379\_timepoint\_1.GU561142  
PIVQNLQGQMVMHQALSPRTLNAWVKVIEEKAFSPEVIPMFSAALSEGATPQ  
DLNTMLNTVGGHQAAMQMLKETINEEAAEWDRHLHPVHAGPIAPGQMREP-  
-RGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWILGLNKIVRMYSVP  
SILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNPDC

no file selected

or

Epigraph design job number

#### Options

Epitope length

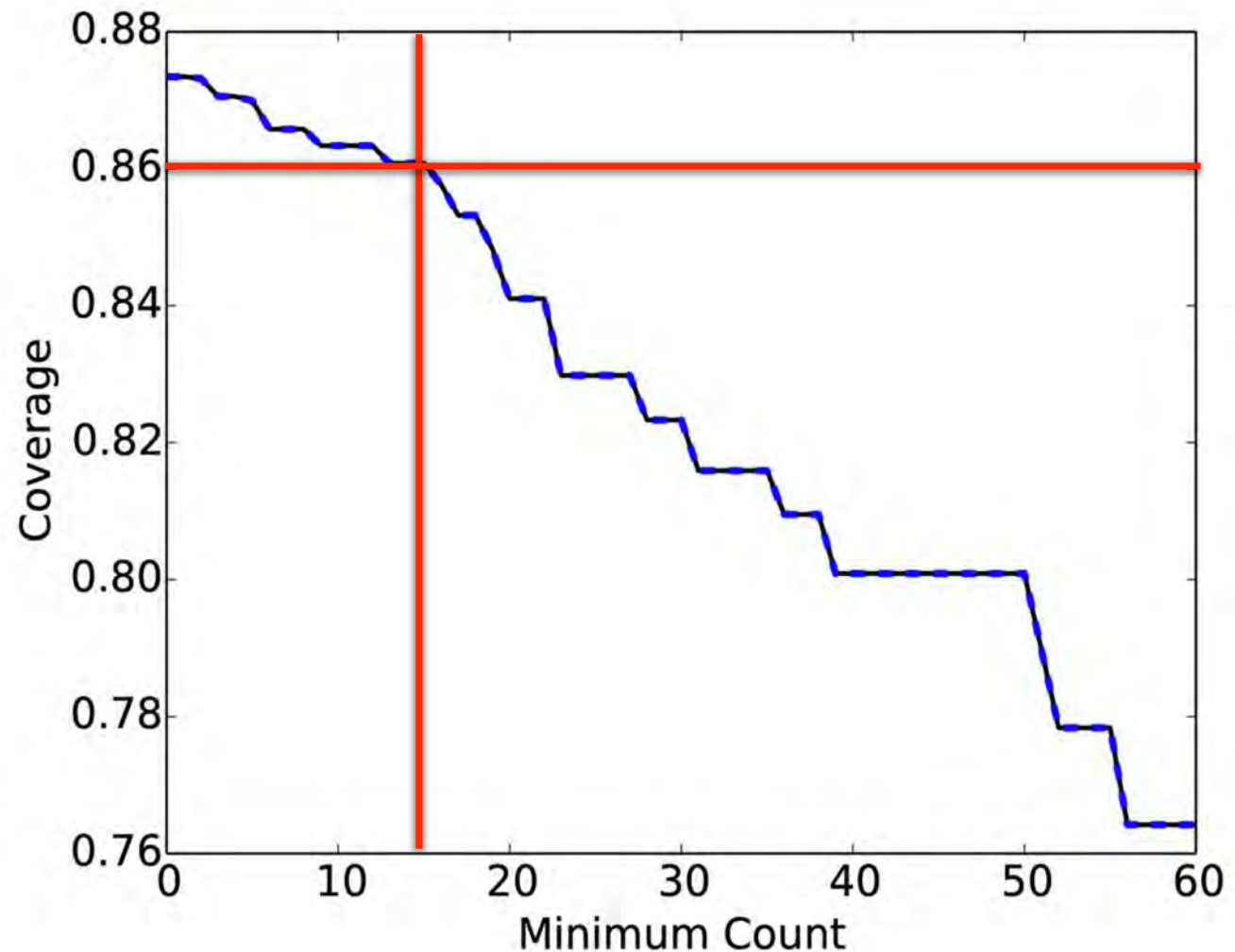
# of sequences in vaccine

Email results ☐



## 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.

# Epitope Coverage Assessment - Epicover

## Input

Use output from MakeVaccine tool

Provide a job number to access output from the [Mosaic Vaccine Designer](#) tool:

OR

Provide input sequences

Paste antigen protein  
sequence(s):  
[\[Sample Input\]](#)

upload more [ + ] antigen sequence files

and/or upload as files:

Browse...

Paste test set protein  
sequences:

upload more [ + ] test sequence files

and/or upload as files:

Browse...

## Options

Send results as an email instead of displaying in browser  
(useful in case of a browser time-out): ☐

Nominal epitope length:

Maximum amino acid mismatches to score (range from 0):

Minimum number of occurrences of a potential epitope  
in viral protein set to consider for coverage:

Precision to use when reporting coverage:

decimal places

## Advanced Options

Upload file of grouped sequence names

Browse...

Report on subsets defined according to first

character(s) in sequence names

Submit

Reset

## 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)

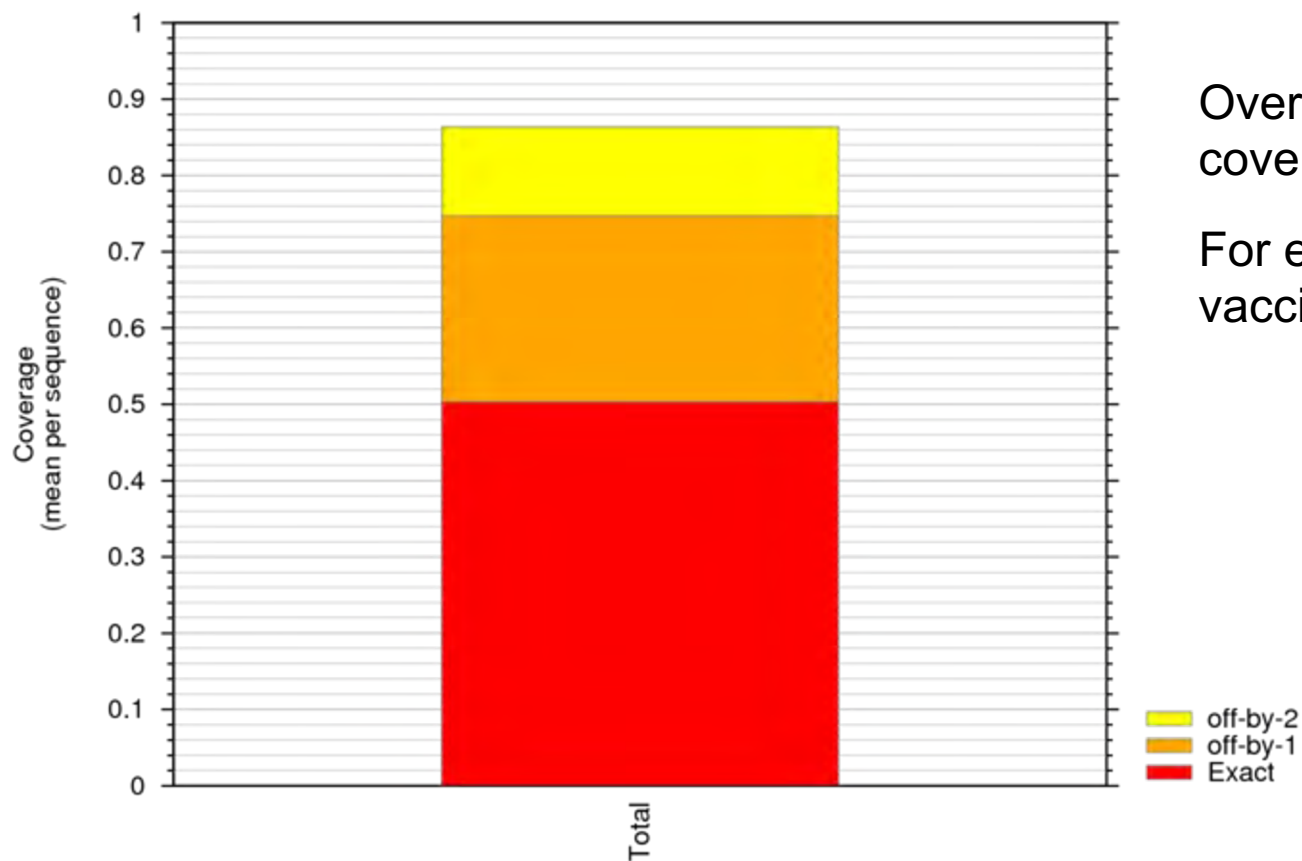
vaccine set	subset	subset count	Off-by-0	Off-by-1	Off-by-2	rare(<3,>1)	unique	absent
<b>vaccine_set_from_user</b>	<b>Total</b>	<b>39</b>	<b>0.5037</b>	<b>0.7474</b>	<b>0.8636</b>	<b>91</b>	<b>61</b>	<b>38</b>
vaccine_set_from_user	A	6	0.4294	0.7086	0.8417	7	1	38
vaccine_set_from_user	B	4	0.7263	0.8911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

Overall summaries of *k*-mer coverage



# Epicover output (mean coverage per sequence)

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

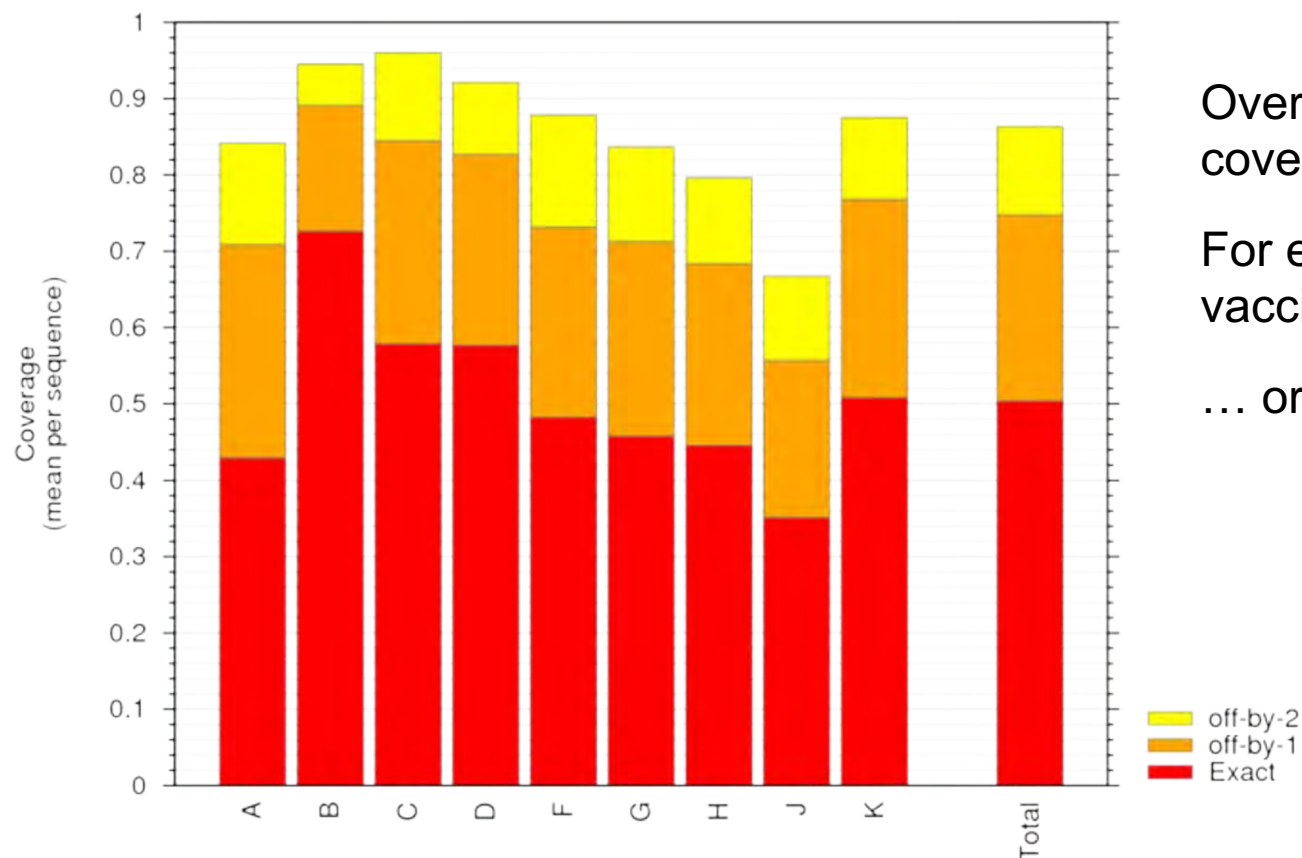


Overall summaries of *k*-mer coverage

For entire set (to compare with other vaccine candidates)

# Epicover output (mean coverage per sequence)

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



Overall summaries of *k*-mer coverage

For entire set (to compare with other vaccine candidates)

... or by pathogen subset

# Positional Epitope Coverage Assessment - Posicover

## Input

Provide a job # from   
[Mosaic Vaccine Designer](#): (Only the antigen set is used. Provide the ALIGNED viral  
**AND/OR**

Paste antigen protein set  
or peptide cocktail:  
(alignment not required)  
[\[Sample Input\]](#)

upload more [ + ] antigen files

and/or upload antigen  
file(s):  No file selected.

## Test set proteins

Paste **ALIGNED** test viral  
protein set:  
[\[Sample Input\]](#)

or upload an **ALIGNED** test  
proteins file:  No file selected.

## Options

Nominal epitope length:

Antigen counts to compute upper bounds:

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 ☒

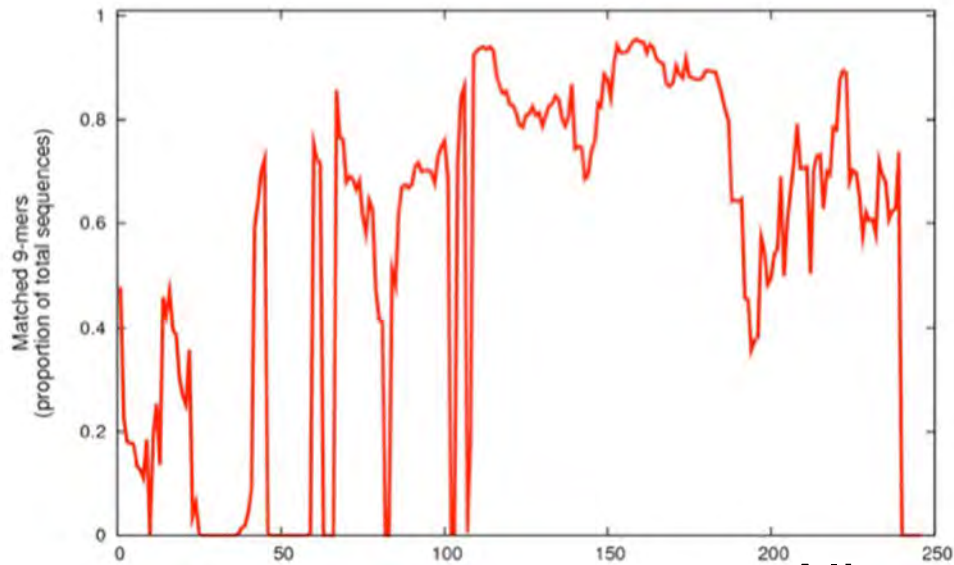
N-mer coverage directly on alignment ☒

- INPUTS
  1. Vaccine/peptide sequences
  2. ALIGNED target set
- OUTPUTS
  - 1-dimensional (by alignment column)
  - 2-dimensional (by sequence and alignment column)

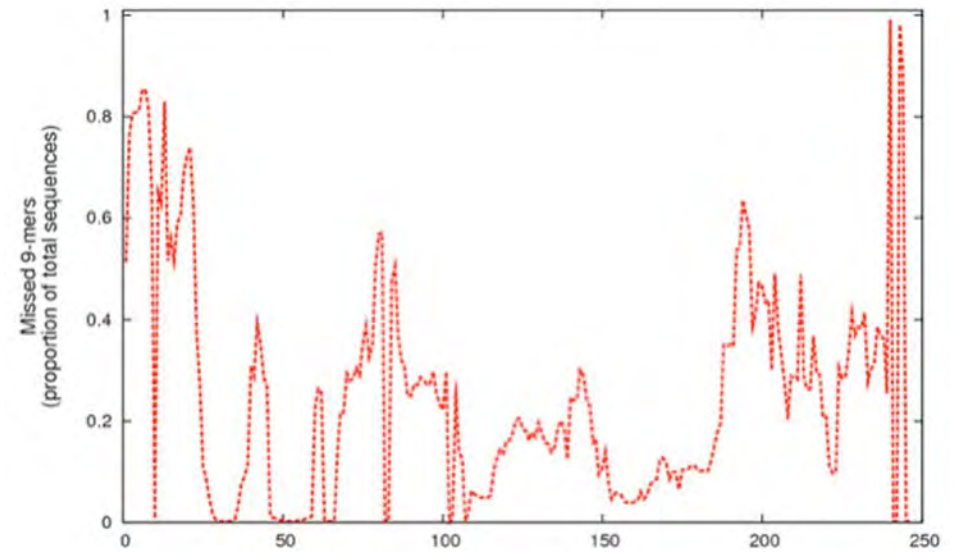


# Posicover output (1-dimensional summaries)

## Matched 9-mers

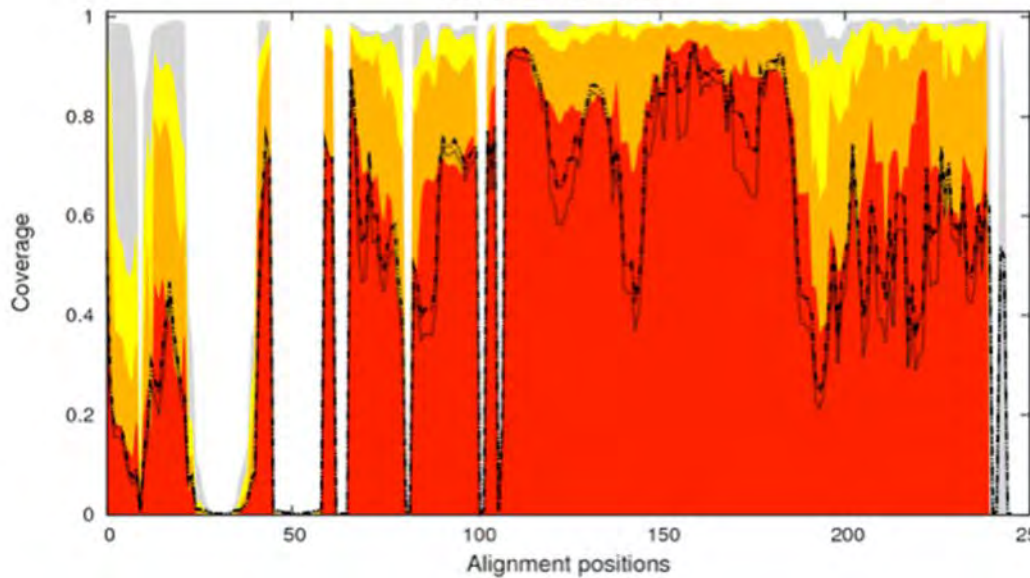


## Missed 9-mers



## Alignment positions

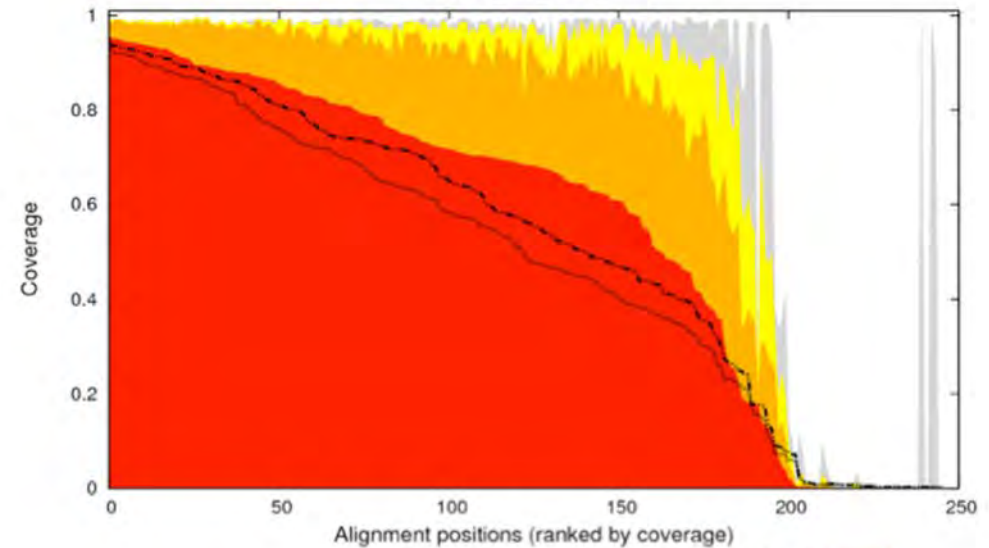
### 9-mer coverage by position vaccine\_set\_from\_user



total 9-mers  
up to 7/9 match  
up to 8/9 match

exact match  
Upper bound: 3 antigen(s)  
Upper bound: 4 antigen(s)

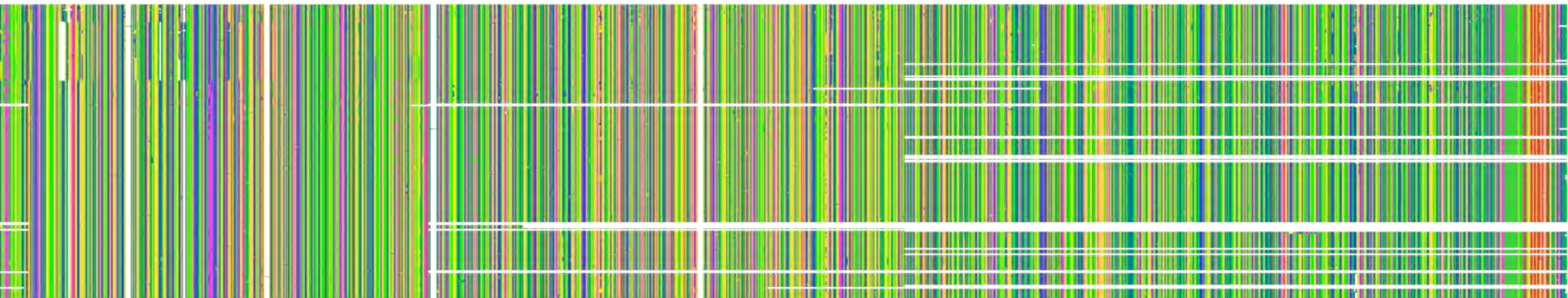
### Ranked 9-mer coverage vaccine\_set\_from\_user



total 9-mers  
up to 7/9 match  
up to 8/9 match

exact match  
Upper bound: 3 antigen(s)  
Upper bound: 4 antigen(s)

# Posicover output (2 dimensional)

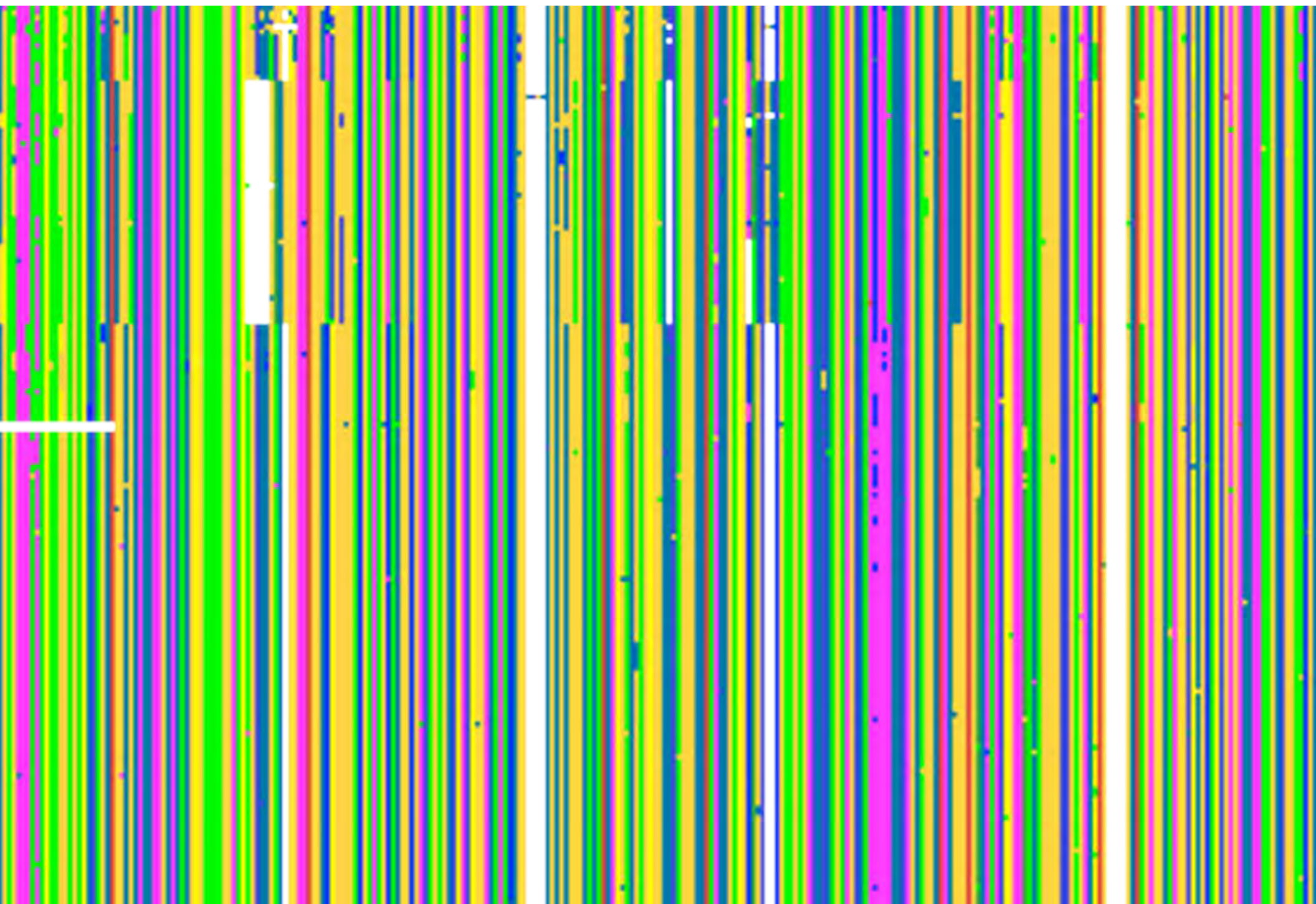


Pixel-based Alignment view

Each amino-acid represented as a single-colored square

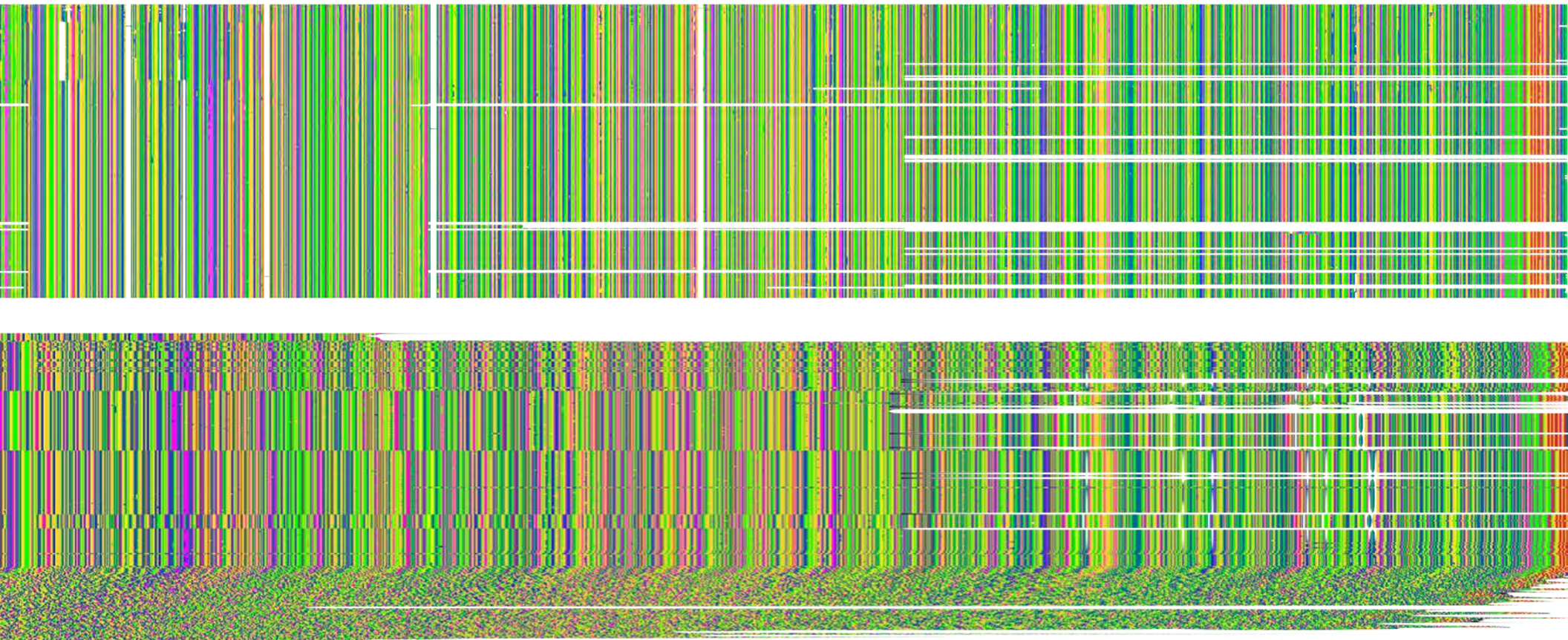
Allows quick detection of gross errors in alignment

# Posicover output (2 dimensional)





# Posicover output (2 dimensional)

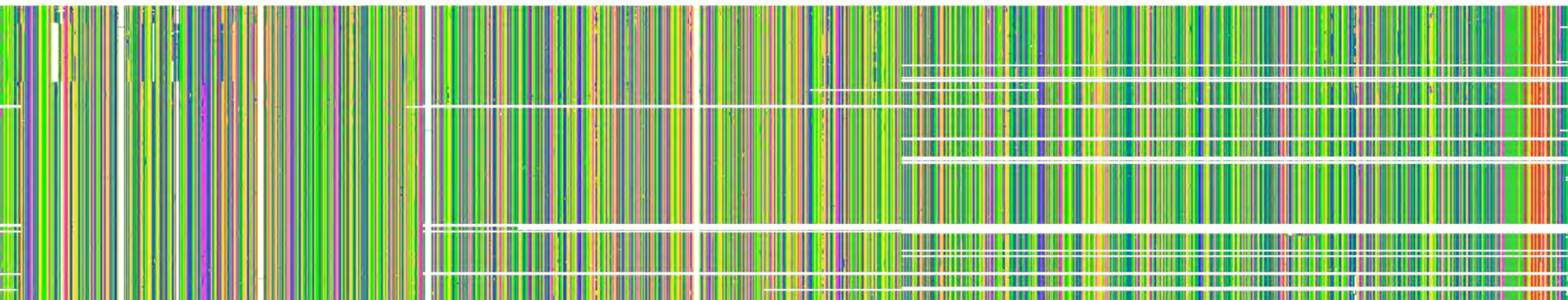


Pixel-based Alignment view

Each amino-acid represented as a single-colored square

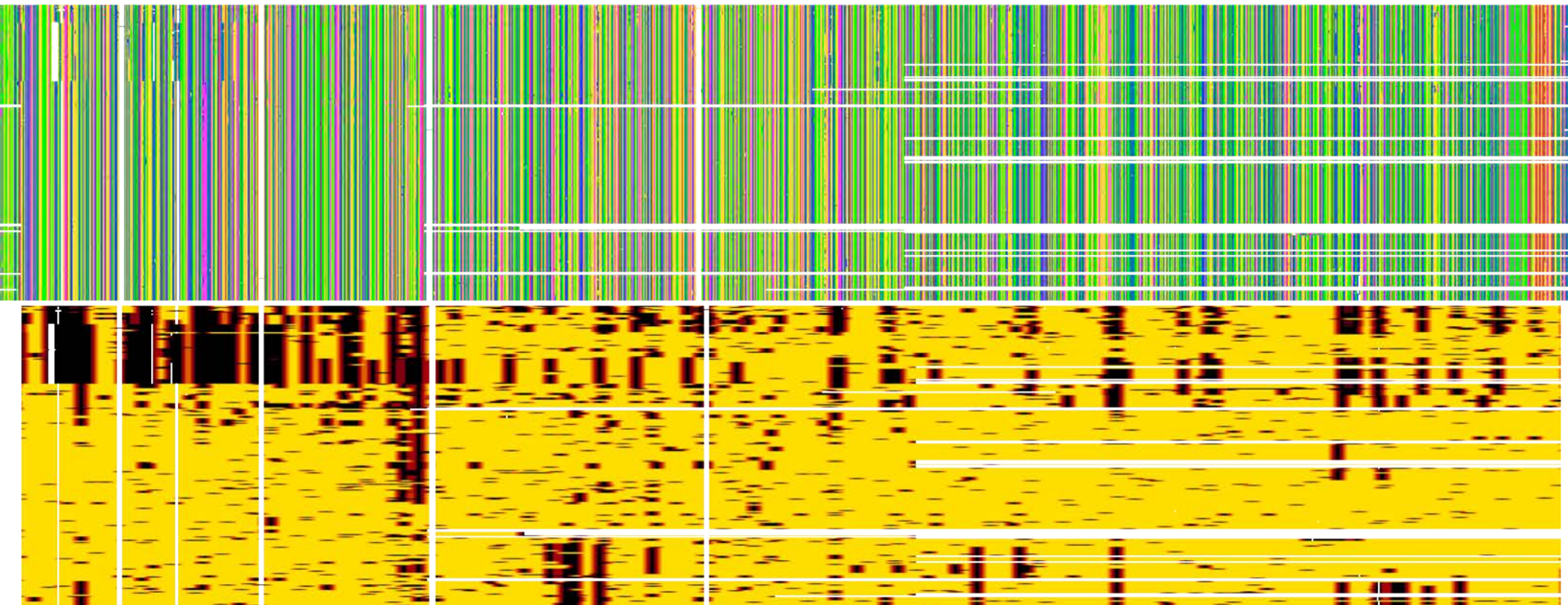
Allows quick detection of gross errors in alignment

# Posicover output (2 dimensional)

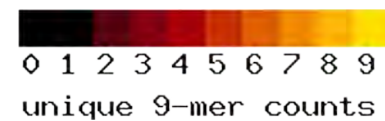




# Posicover output (2 dimensional)

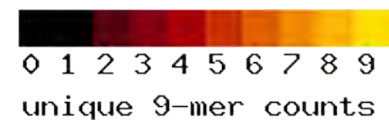
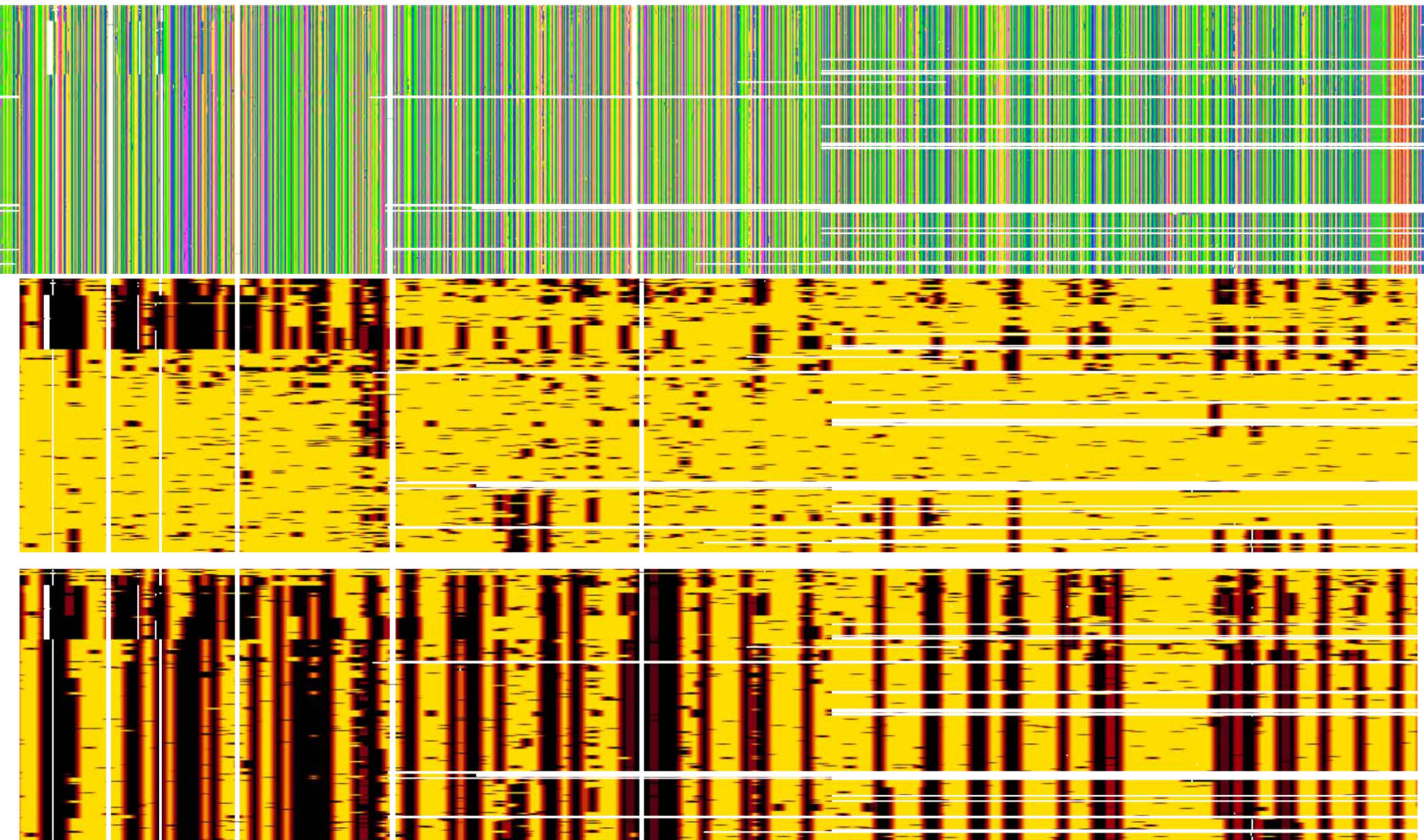


POSICOVER *K*-MER COVERAGE  
(YELLOW-BLACK GRADIENT SHOWS HOW MANY OF EACH  
RESIDUE'S *K*-MERS APPEAR IN VACCINE)



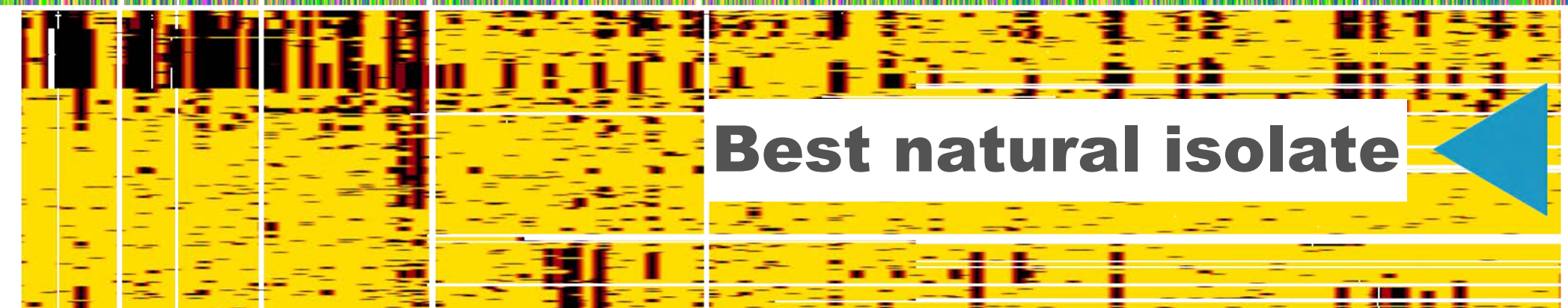
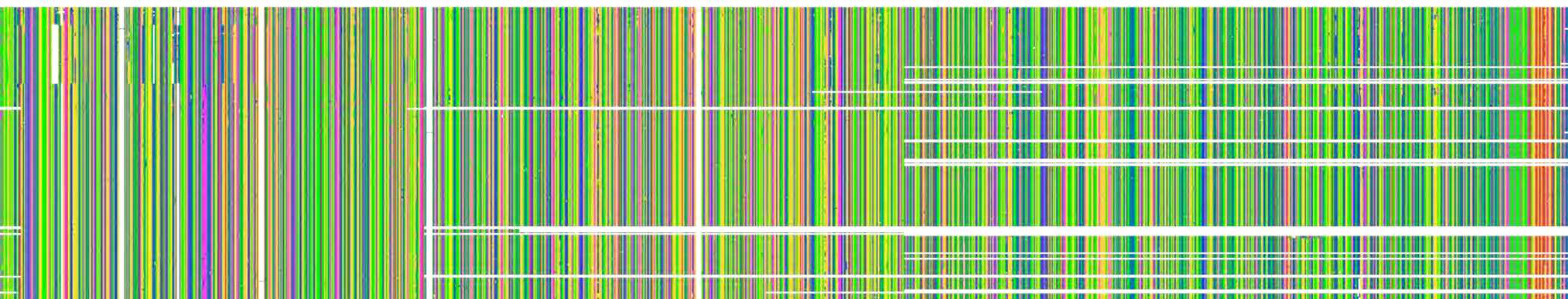


# Posicover output (2 dimensional)

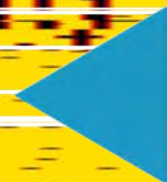




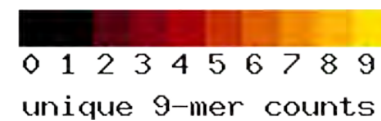
# Posicover output (2 dimensional)



**Best natural isolate**

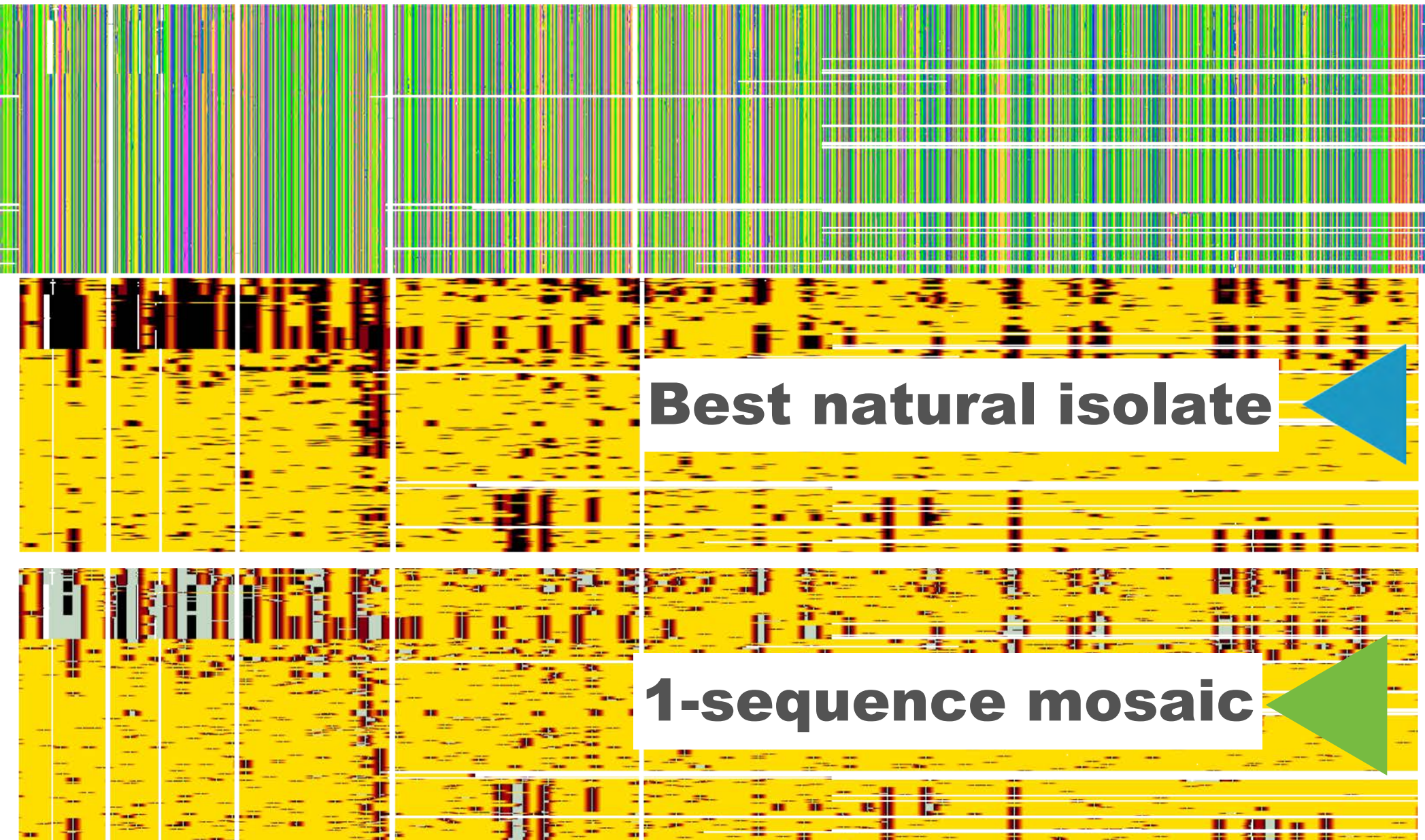


**Worst natural isolate**



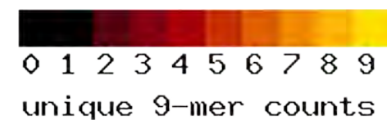
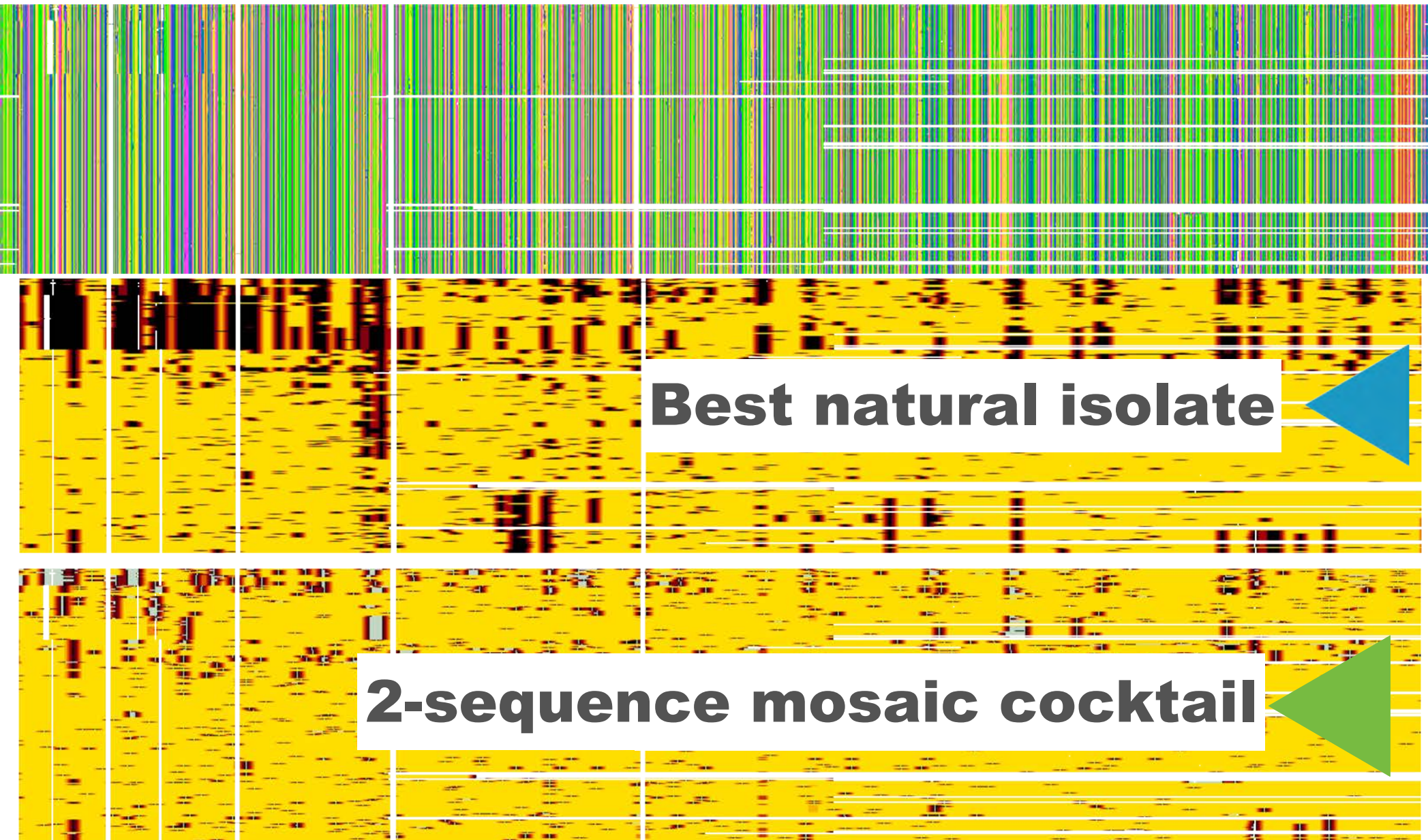


# Posicover output (2 dimensional)

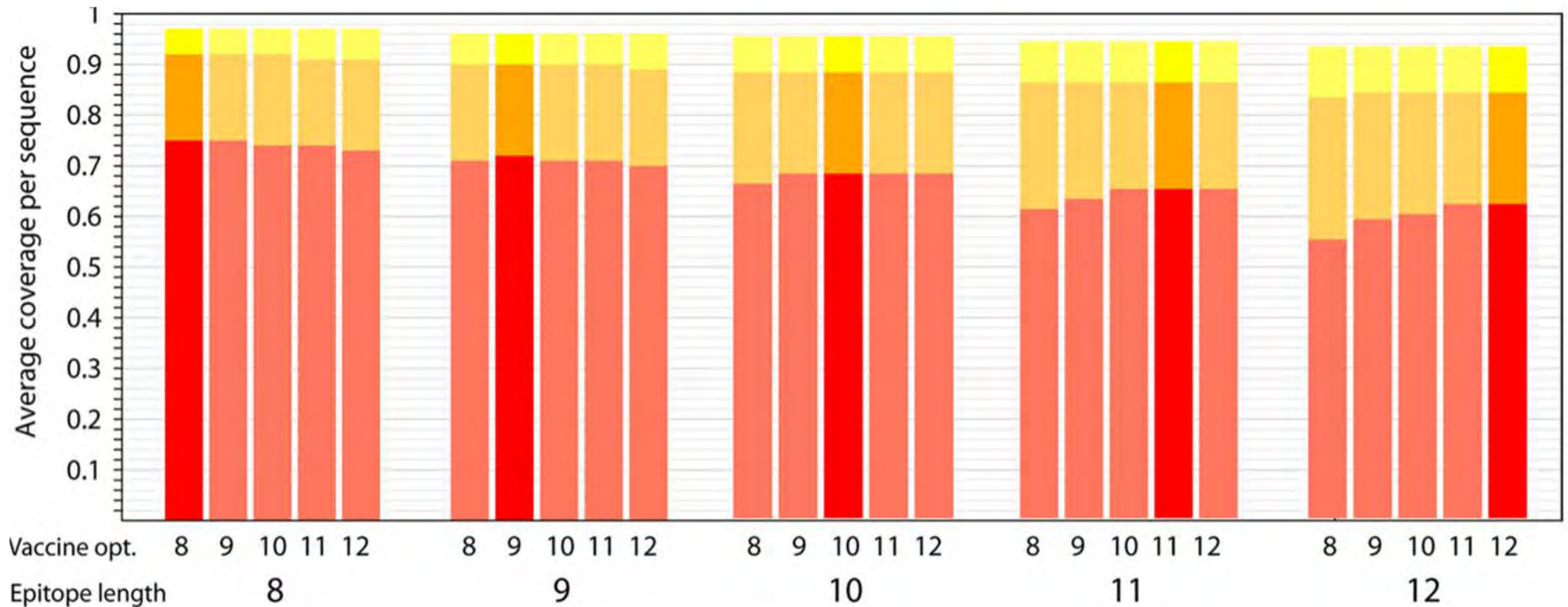




# Posicover output (2 dimensional)



# ***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.

[Korber et al., 2009] T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces. J Virol, 83(17):8300–14.

**Thank you for attending!**

Please send us comments, questions, and suggestions!

Your comments will help us provide future training and better tools.

Slides available at <https://hiv.lanl.gov/hws>

Contact us: [immuno@lanl.gov](mailto:immuno@lanl.gov) or [seq-info@lanl.gov](mailto:seq-info@lanl.gov)