HIV Database Immunology Workshop

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

immuno@lanl.gov

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 NATIONAL LABORATORY

Los Alamos HIV-1 Databases

❑ Integrate HIV immunological and sequence data

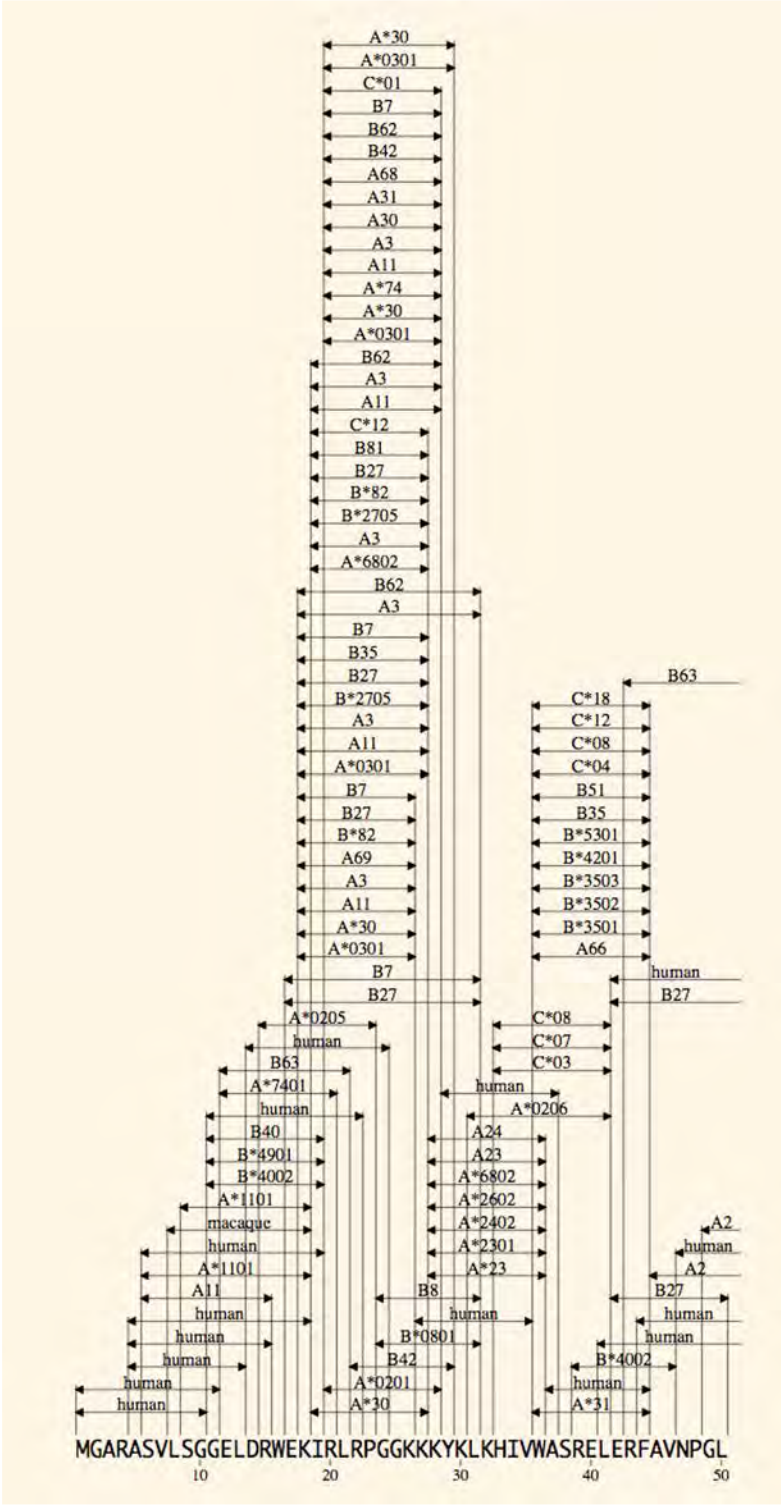
- For the first part of March 2022:
  - 306,644 hits
  - 12,783 visits

- 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/

Los Alamos NATIONAL LABORATORY

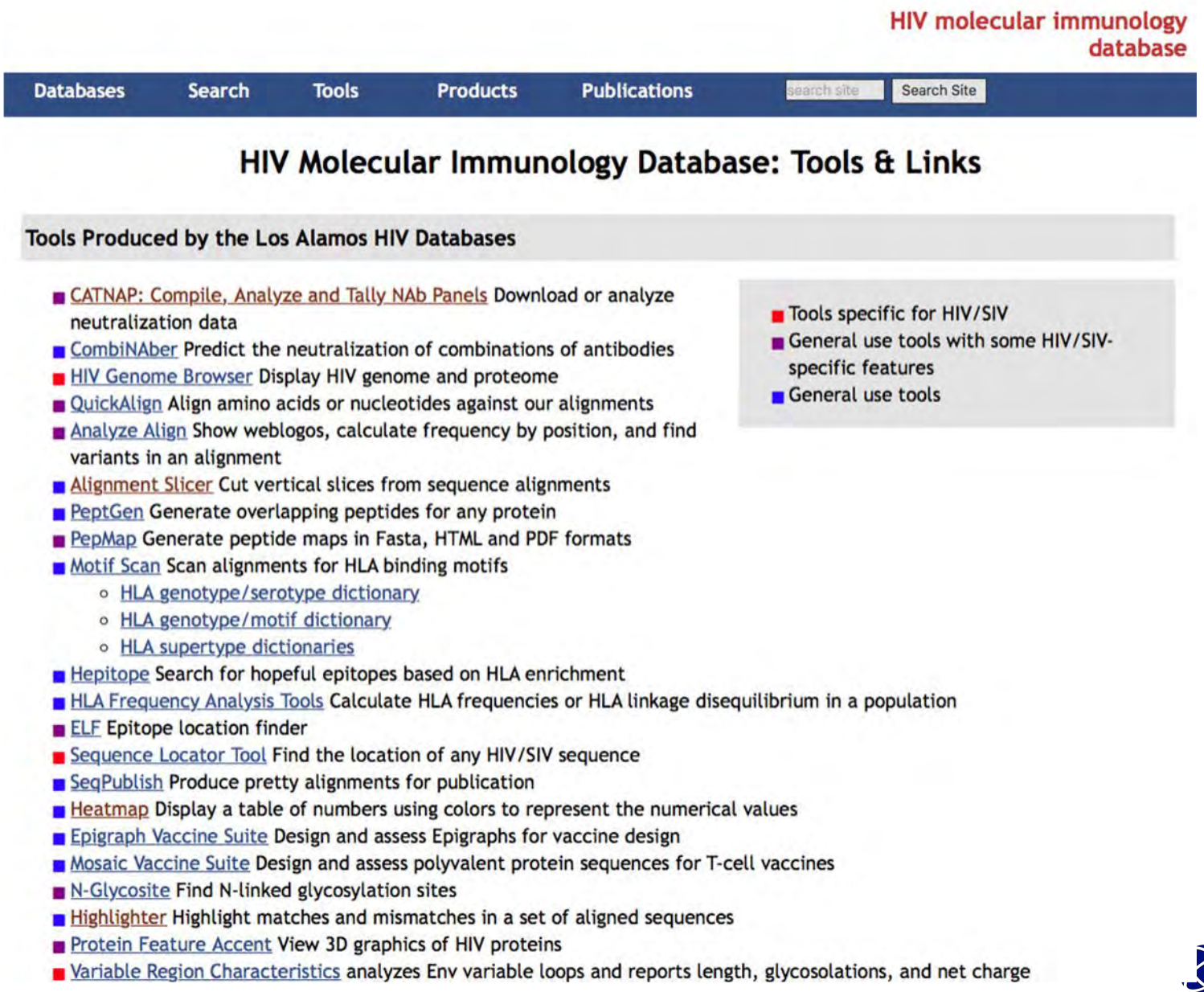
Integration of HIV Sequence and Immunology databases

- 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

Databases | Search | Tools | Products | Publications | search site | Search Site

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

Los Alamos NATIONAL LABORATORY

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)
  - **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

Los Alamos NATIONAL LABORATORY

https://hiv.lanl.gov/

Los Alamos NATIONAL LABORATORY

Databases | Search | Tools | Products | Publications | search site | Search Site

- Sequence DB
- Immunology DB
- Neutralization DB
- Ab contacts DB
- HCV Databases
- HFV Databases

HIV Molecular Immunology Database

The 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

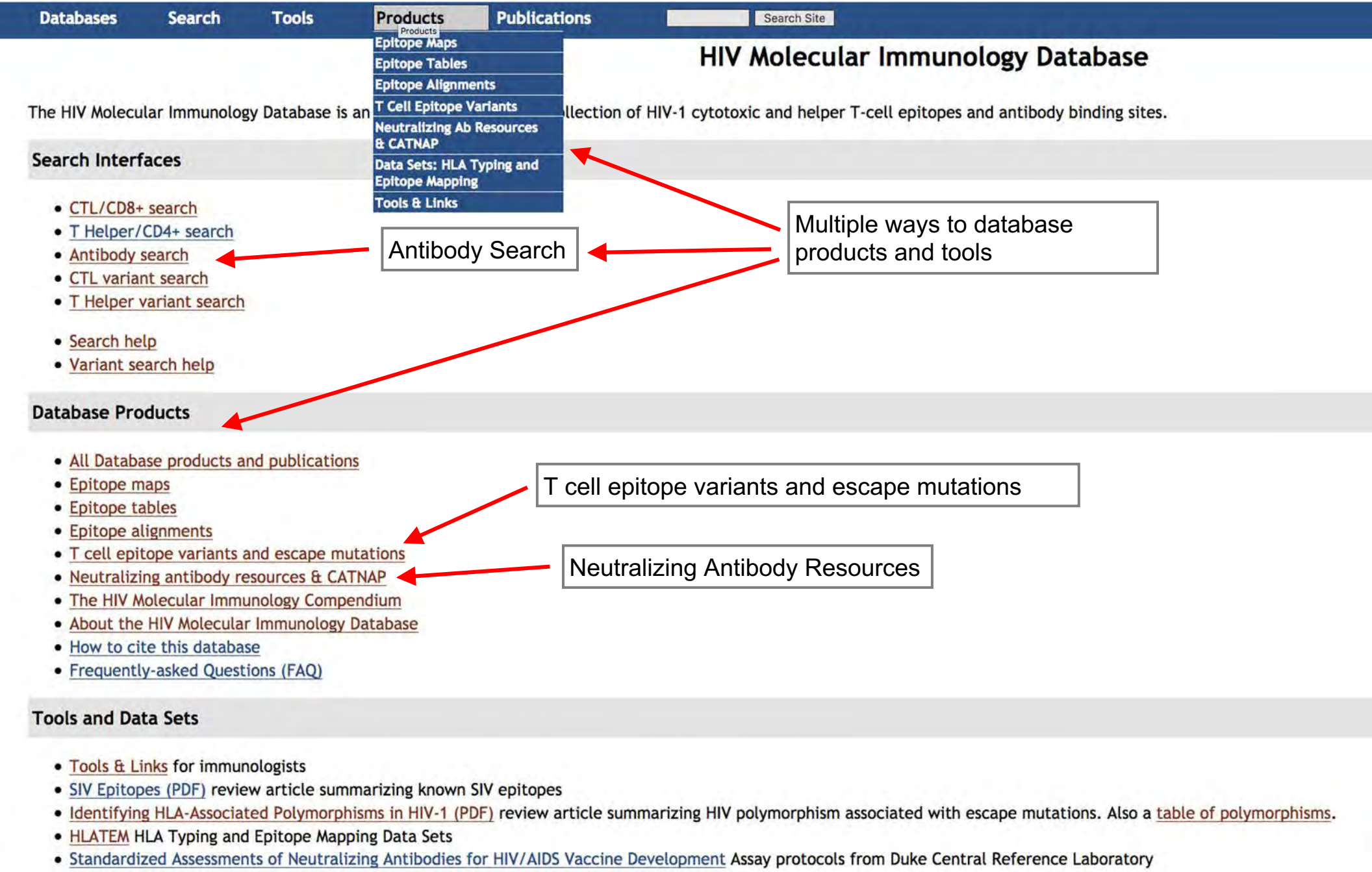
- All Database products and publications
- Epitope maps
- Epitope tables
- Epitope alignments
- Epitope density plots
- T cell epitope variants and escape mutations
- Neutralizing antibody resources & CATNAP
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How 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

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

Los Alamos NATIONAL LABORATORY



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

Los Alamos NATIONAL LABORATORY

Databases Search Tools Products Publications Search Site

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

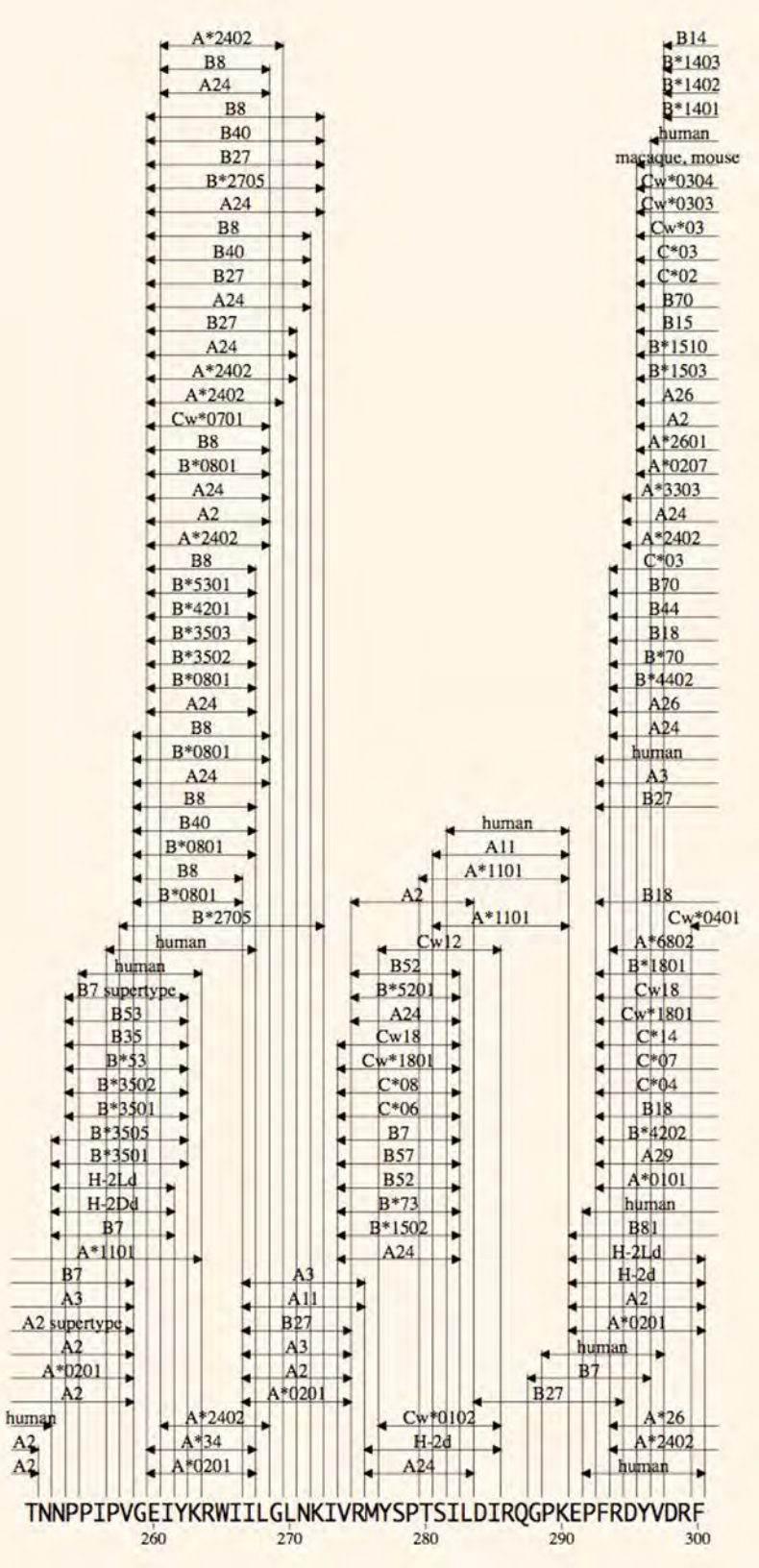
Epitope alignments: epitopes aligned to HIV subtype Reference sequences in Fasta format

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

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

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
MGARASVLSG	p17	1-10	CRF01_AE	human	
ASVLSGGEL	p17	5-13	B	human	
ASILRGGKLDK	p17	5-15	C	human	
SVLSGGQLDR	p17	6-15	B	human	A11
LSGGELDRWEK	p17	8-18		macaque	
GELDRWEKI	p17	11-19	B	human	B*4002, B40
GQLDRWEKI	p17	11-19	B	human	
GKLDSWEKIRLR	p17	11-22	A, CRF01_AE, CRF02_AG	human	

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
GELDRWEKI	p17	11-19		human	B*4002
KIRLRPGGK	p17	18-26		human	A*0301
IRLRPGGKK	p17	19-27	B	human	B*2705
RLRPGGKKK	p17	20-28		human	A*0301
RLRPGGKKKY	p17	20-29	B	human	A*0301
GGKKKYKLK	p17	24-32	B	human	B*0801
KYKLKHIVW	p17	28-36	B	human	A*2402
HLVWASREL	p17	33-41		human	Cw*0804

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

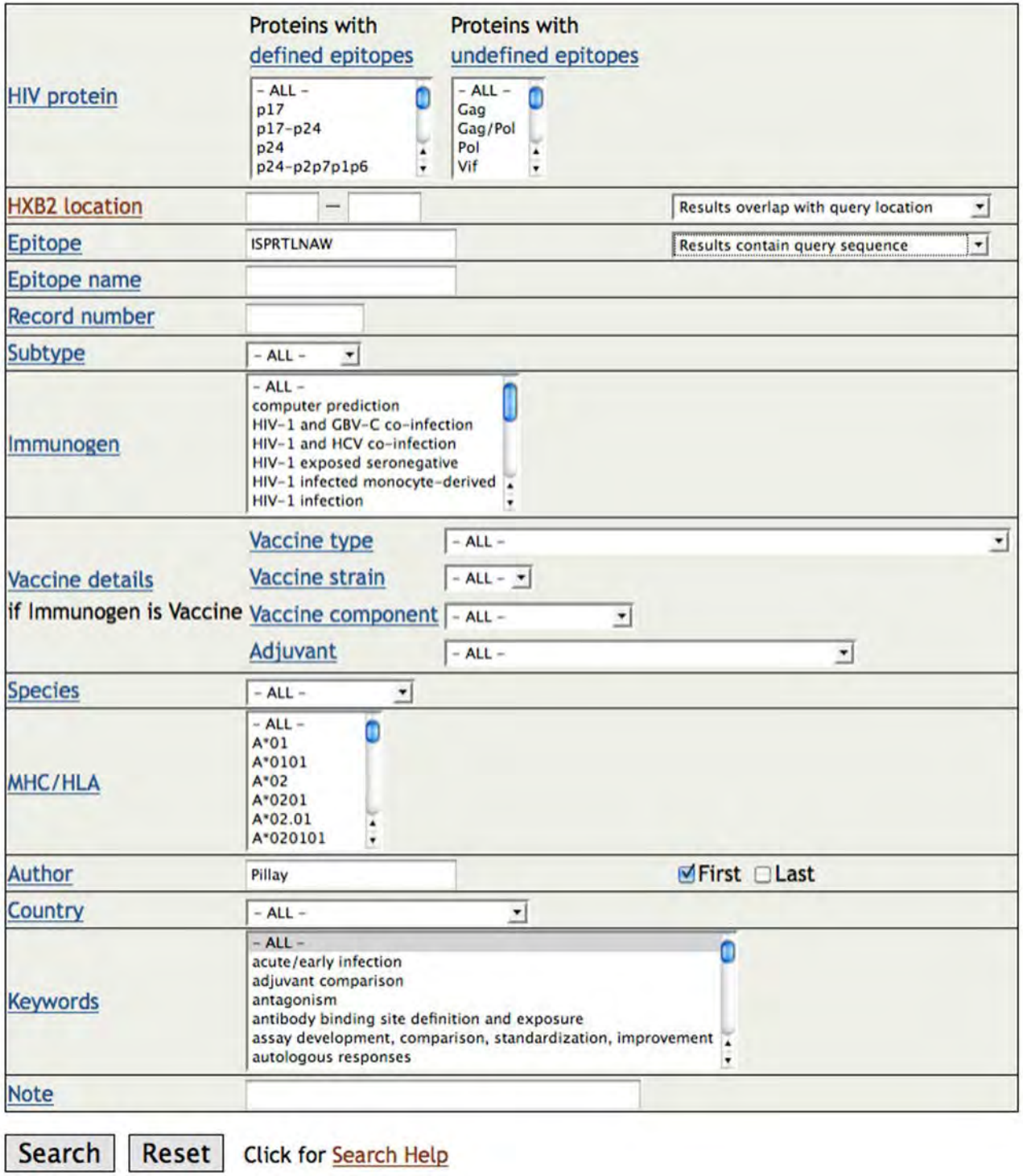
Epitope variants and escape mutations

- 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

Los Alamos NATIONAL LABORATORY

CTL/CD8+ Search (www.hiv.lanl.gov/content/immunology/ctl_search)



- Search by HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords
- Search on epitope location and find fuzzy matches, overlaps and embedded epitopes
- Search examples:
  - *Example:*
    - SLYNTVATL – 285 entries
    - Narrow the search with keyword “escape” – 35 entries

Search for ISPRTLNAW with first author = Pillay

Los Alamos NATIONAL LABORATORY

Search CTL/CD8+ T-Cell Epitope Database

Found 1 matching record:

Displaying record number 53832

Immunological, virological, and epidemiological contexts:

Link to Epitope Maps →

Link to Epitope Alignment →

Variant details with annotator's notes →

HXB2 Location	p24(15-23)	p24 Epitope Map
Author Location	Gag(147-155)	
Epitope	ISPRTLNAW	Epitope Alignment
		Show epitope variants
Subtype	C	
Species (MHC/HLA)	human(B57)	
Immunogen	HIV-1 infection	
Donor MHC/HLA	A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602, Cw*0701	
Country	South Africa	
Experimental methods	CD8 T-cell Elispot - IFNy	
Keywords	epitope processing, responses in children, mother-to-infant transmission, escape, acute/early infection	

Notes

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

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.

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

Los Alamos NATIONAL LABORATORY

Databases Search Tools Products Publications search site Search 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 if Immunogen is Vaccine	Vaccine type: - ALL - Vaccine strain: - ALL - Vaccine component: - ALL - Adjuvant: - 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

Databases Search Tools Products Publications search site Search 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	QTGSEELRSLYNTVATLYCVHQRIE	Epitope Alignment
Species (MHC/HLA)	human(A2)	
Immunogen	HIV-1 infection	
Experimental methods		
Keywords		

Notes

- Gag CTL response studied in three individuals.

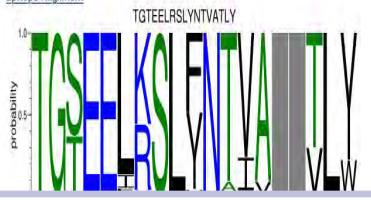
References

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: KIRLRPGGKKKYKLKHIVWASRELE and QTGSEELRSLYNTVATLYCVHQRIE; p24 epitopes: NPPIPVGEIYKRWIILGLNKIV, VHQAISPRTLNAWVKVVEEKAF, NAWVKVVEEKAFSPEVIPMFSA, SALSEGATPQDLNTMLNTVGGH, GHQAAMQMLKETINEEAAEWDR, 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	TGTEELRSLYNTVATLY	Epitope Alignment



https://www.hiv.lanl.gov/content/immunology/ctl_search.html

Los Alamos NATIONAL LABORATORY

Databases | Search | Tools | Products | Publications | search site | Search Site

Search CTL/CD8+ T-Cell Epitope

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 if Immunogen is Vaccine	Vaccine type: - ALL - Vaccine strain: - ALL - Vaccine component: - ALL - Adjuvant: - 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	mother-to-infant transmission mutation acquisition naive T cells Nef-mediated down-regulation neutralization by CTL NK cells non-susceptible form	
Note		

Search | Reset | Click for Search Help

Databases | Search | Tools | Products | Publications | Search Site

Search CTL/CD8+ T-Cell Epitope Database

Found 5 matching records:

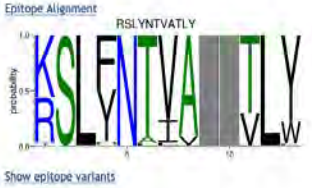
Displaying record number 62248

Download this epitope record as JSON.

HXB2 Location	Gag(76-86) p17(76-86) DNA(805..834)
Author Location	Gag(76-86)
Epitope	RSLYNTVATLY
Epitope Name	Gag-RY11
Subtype	C
Species (MHC/HLA)	human(A*30)
Immunogen	HIV-1 infection
Patient MHC/HLA	517-C: A*30:02, A*68:02, B*15:10, B*42:01, C*03:04, C*17:01; 517-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 - IFNγ, Other
Keywords	responses in children, mother-to-infant transmission, rate of progression, escape

Gag Epitope Map

Epitope Alignment



Show epitope variants

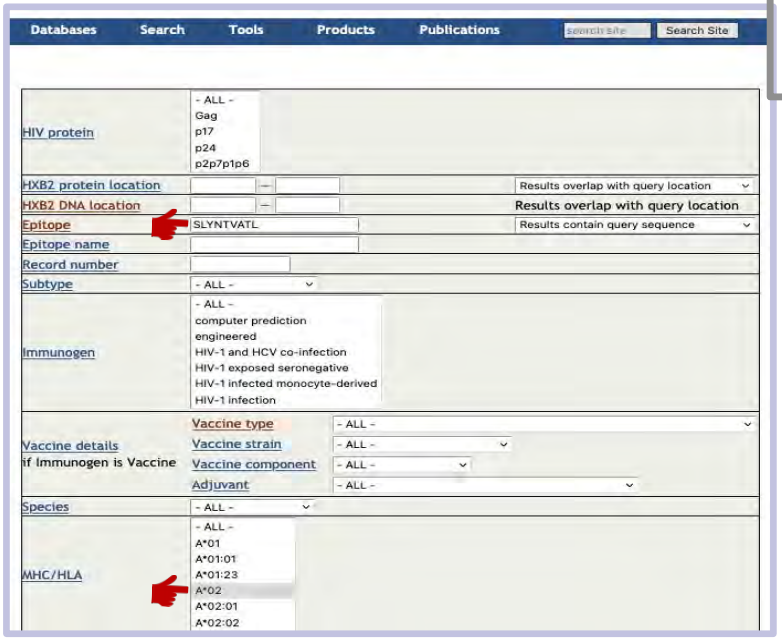
Notes

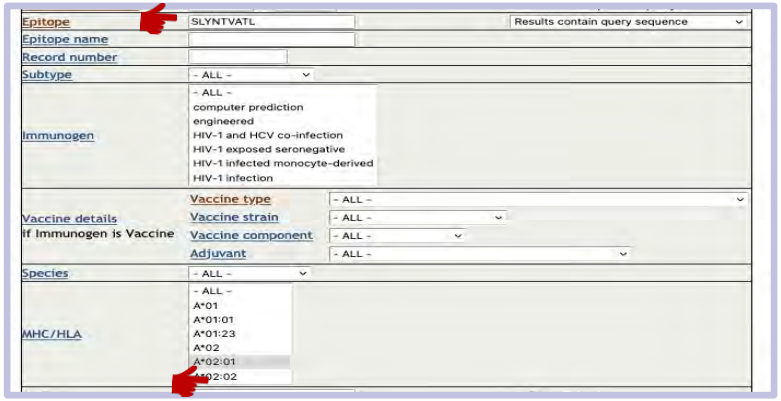
- 11 perinatally infected pediatric slow progressors (PSPs) were followed longitudinally for a decade from birth to examine CTL responses to circulating and autologous HIV. It was found by ultra-deep sequencing that though, t develop variants early in infection; unlike adults, pediatric anti-variant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

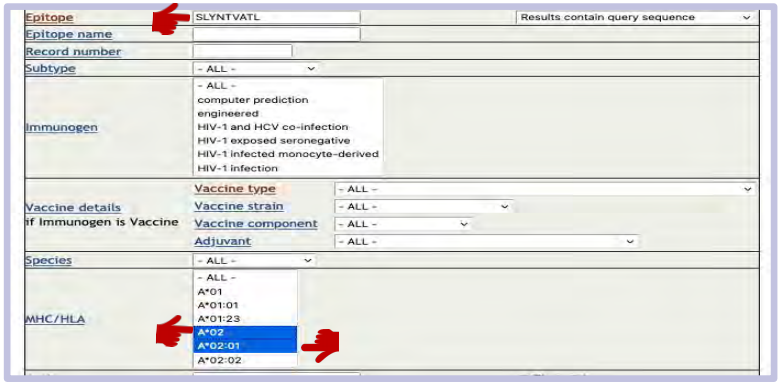
https://www.hiv.lanl.gov/content/immunology/ctl_search.html

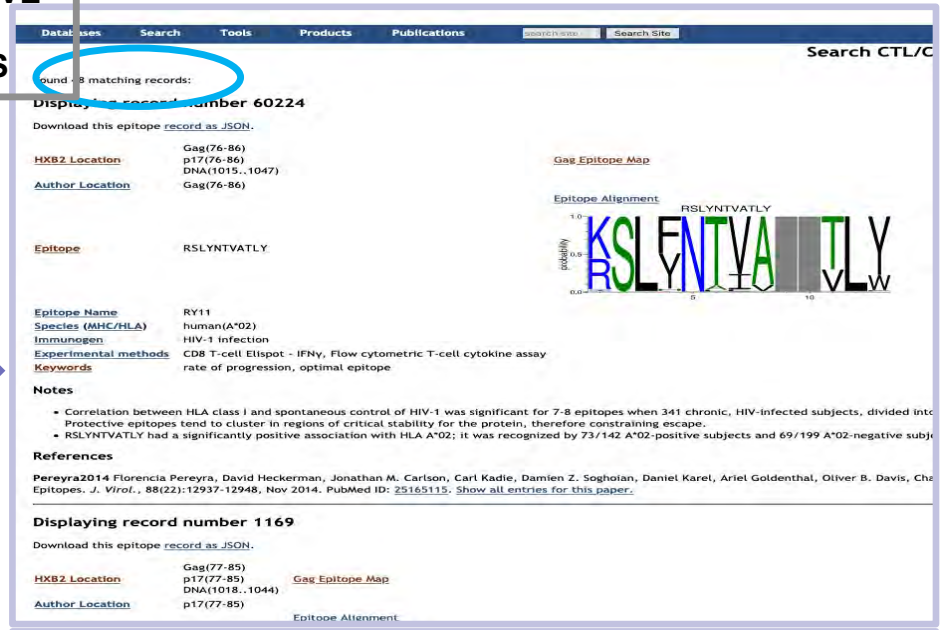
Los Alamos NATIONAL LABORATORY

CUMULATIVE CTL SEARCHES



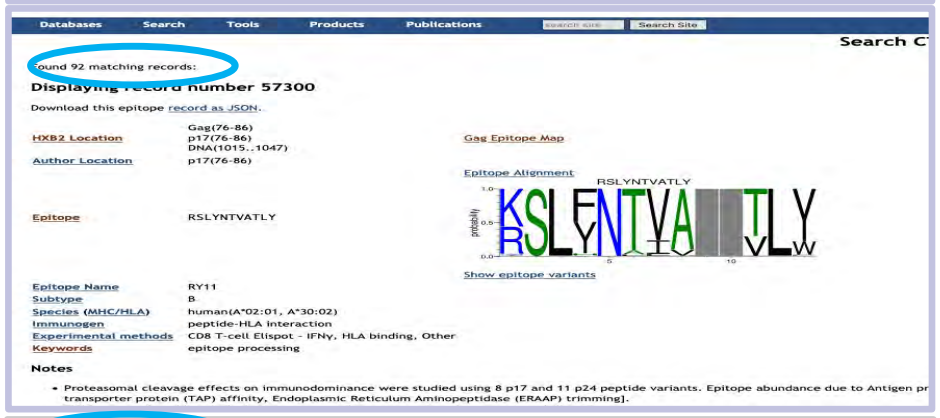






SLYNTVATL
+ A*02

48
records



SLYNTVATL
+ A*02:01

92
records

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

140
records

Los Alamos
NATIONAL LABORATORY

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

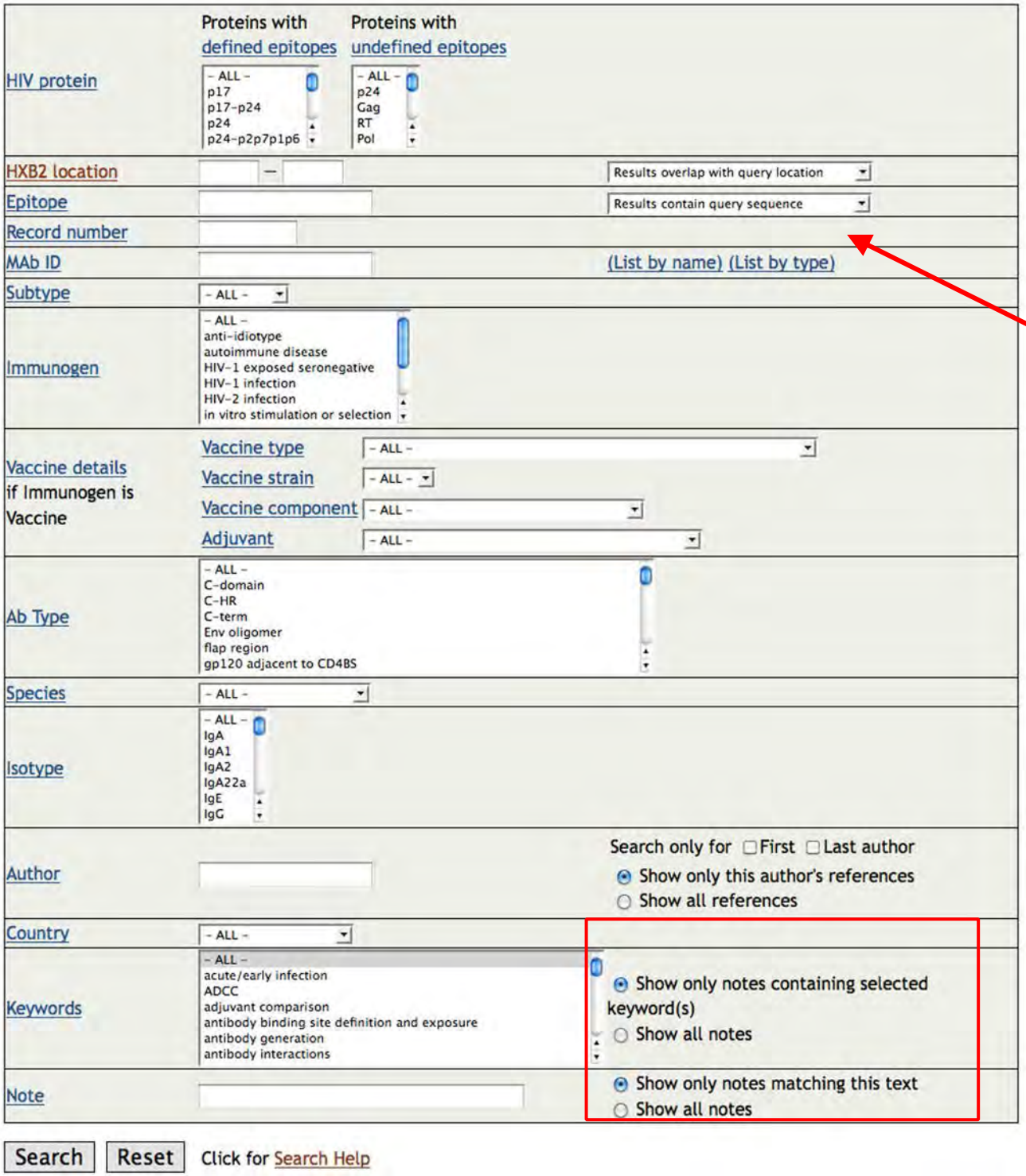
News Archive

No new news.

Questions or comments? Contact us at immuno@lanl.gov

Los Alamos NATIONAL LABORATORY

Antibody Search (https://www.hiv.lanl.gov/content/immunology/ab_search)



Search by

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

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

Los Alamos NATIONAL LABORATORY

Search Antibody Database

Search results for 10E8

Found 30 matching records:

Displaying record number 2708

MAb ID	10E8
HXB2 Location	gp160(671-683) DNA(8235..8273)
Author Location	
Epitope	NWFDISNWLWYIK
Subtype	B
Ab Type	gp41 MPER (membrane proximal external region)
Neutralizing	P (tier 2) View neutralization details
Contexts and Features	Search for contexts and features
Species (Isotype)	human(IgG3)
Patient	Donor N152
Immunogen	HIV-1 infection
Keywords	ADCC, antibody binding site, antibody gene transfer, antibody generation, antibody lineage, antibody sequence, binding affinity, bispecific molecule, broad neutralizer, chimeric antibody, computational epitope prediction, contact residues, glycosylation, immunoprophylaxis, immunotherapy, neutralization, review, structure, subtype comparisons, vaccine antigen design, vaccine-induced immune responses, variant cross-reactivity

Link to Epitope Map

gp160 Epitope Map

Link to Epitope Alignment

Epitope Alignment

Link to CATNAP

Link to Antibody Features Database
(Ab contact positions and related protein features)

Link to patient Donor detail

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_{50}/IC_{80} 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_{50}, IC_{80}, and ID_{50} data, germline genes, and antibody sequences.

Questions or comments? Contact us at immuno@lanl.gov

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS INFO search site Search

Neutralizing Antibody Contacts & Features

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

Submit Reset

Example: Z13e1 contact sites

ID 21

Description Z13e1 contacts

Antibody class MPER

Reference Nelson2007

Type contacts

MAb name Z13 Z13e1 (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 Hxb2 amino acid: S668 L669 W670 N671 W672 F673 N674 I675 T676 N677

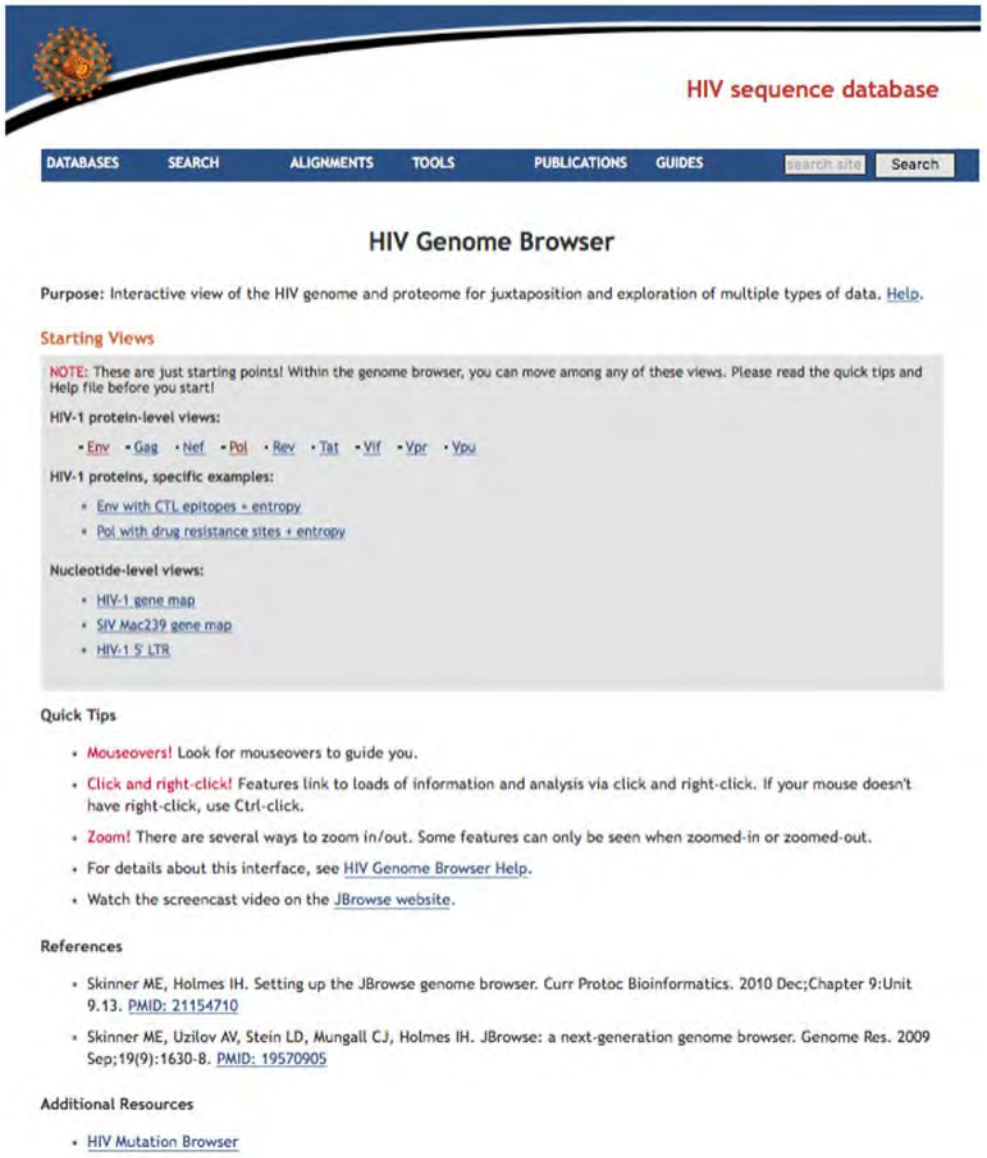
Los Alamos NATIONAL LABORATORY

https://www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp

Genome Browser

A tool at the interface between the sequence and immunology databases

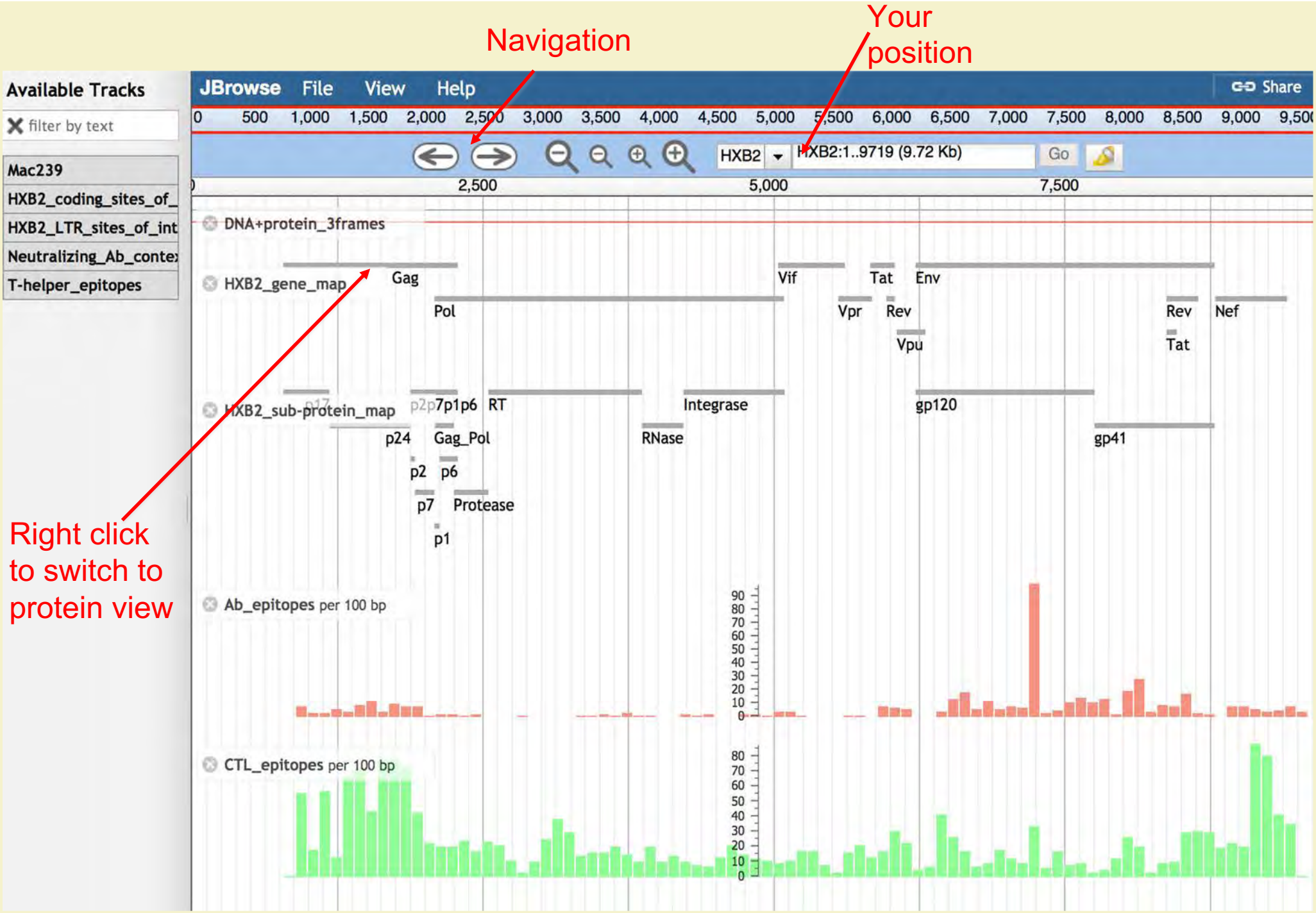
- Tools -> Genome Browser
- Click on Env
- Drag tracks into the map:
  - Entropy M group
  - Ab epitopes
  - Neutralizing Ab contexts
- Zoom in to expand a region of interest
- Click around to explore

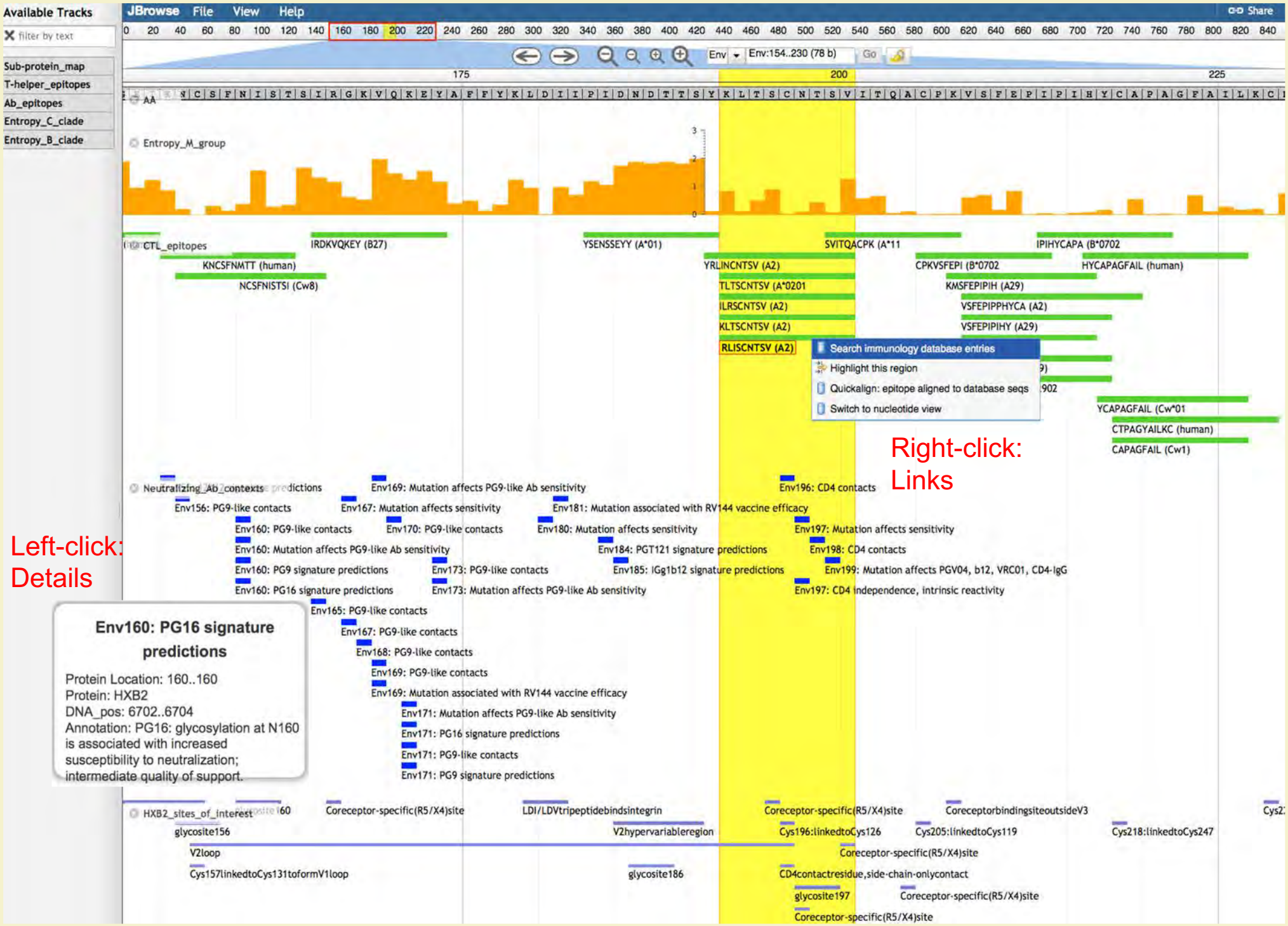


https://www.hiv.lanl.gov/content/sequence/genome_browser/browser.html

Los Alamos NATIONAL LABORATORY

HIV Genome Browser: Nucleotide view




Available Tracks
filter by text
Sub-protein_map
T-helper_epitopes
Ab_epitopes
Entropy_C_clade
Entropy_B_clade
JBrowse File View Help
Share
Env Env:154..230 (78 b) Go
Entropy_M_group
CTL_epitopes
IRDKVQKEY (B27)
YSENSSEYY (A*01)
SVITQACPK (A*11
IPIHYCAPA (B*0702
KNCSFNMTT (human)
NCSFNISTSI (Cw8)
YRLINCNTSV (A2)
CPKVSFEPI (B*0702
HYCAPAGFAIL (human)
TLTSCNTSV (A*0201
KMSFEPIPIH (A29)
ILRSCNTSV (A2)
VSFEPIPPHYCA (A2)
KLTSCNTSV (A2)
VSFEPIPIHY (A29)
RLISCNTSV (A2)
Search immunology database entries
Highlight this region
Quickalign: epitope aligned to database seqs
Switch to nucleotide view
YCAPAGFAIL (Cw*01
CTPAGYAILKC (human)
CAPAGFAIL (Cw1)
Right-click: Links
Neutralizing_Ab_contexts
Env156: PG9-like contacts
Env160: PG9-like contacts
Env160: Mutation affects PG9-like Ab sensitivity
Env160: PG9 signature predictions
Env160: PG16 signature predictions
Env165: PG9-like contacts
Env167: PG9-like contacts
Env168: PG9-like contacts
Env169: PG9-like contacts
Env169: Mutation associated with RV144 vaccine efficacy
Env171: Mutation affects PG9-like Ab sensitivity
Env171: PG16 signature predictions
Env171: PG9-like contacts
Env171: PG9 signature predictions
Env167: Mutation affects sensitivity
Env170: PG9-like contacts
Env173: PG9-like contacts
Env173: Mutation affects PG9-like Ab sensitivity
Env169: Mutation affects PG9-like Ab sensitivity
Env180: Mutation affects sensitivity
Env181: Mutation associated with RV144 vaccine efficacy
Env184: PGT121 signature predictions
Env185: IGg1b12 signature predictions
Env196: CD4 contacts
Env197: Mutation affects sensitivity
Env198: CD4 contacts
Env199: Mutation affects PGV04, b12, VRC01, CD4-IgG
Env197: CD4 independence, intrinsic reactivity
Left-click: Details
Env160: PG16 signature predictions
Protein Location: 160..160
Protein: HXB2
DNA_pos: 6702..6704
Annotation: PG16: glycosylation at N160 is associated with increased susceptibility to neutralization; intermediate quality of support.
HXB2_sites_of_interest
glycosite156
V2loop
Cys157linkedtoCys131toformV1loop
Coreceptor-specific(R5/X4)site
LDI/LDVtripeptidebindsintegrin
V2hypervariableregion
glycosite186
Coreceptor-specific(R5/X4)site
Coreceptorbindingsiteoutside V3
Cys196:linkedtoCys126
Cys205:linkedtoCys119
Cys218:linkedtoCys247
Coreceptor-specific(R5/X4)site
CD4contactresidue,side-chain-onlycontact
glycosite197
Coreceptor-specific(R5/X4)site
Coreceptor-specific(R5/X4)site

CATNAP (**C**ompile, **A**nalyze and **T**ally **NA**b **P**anels)

- CATNAP is both a *database* and a set of *tools*
- Compiles published HIV Ab neutralization data

  >400 Abs, >1000 HIV pseudoviruses, >100,000 $IC_{50/80}$ 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

https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/index.html

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_{50}, IC_{80}, and ID_{50} 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_{50} and IC_{80} 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
- $IC_{50/80}$ data

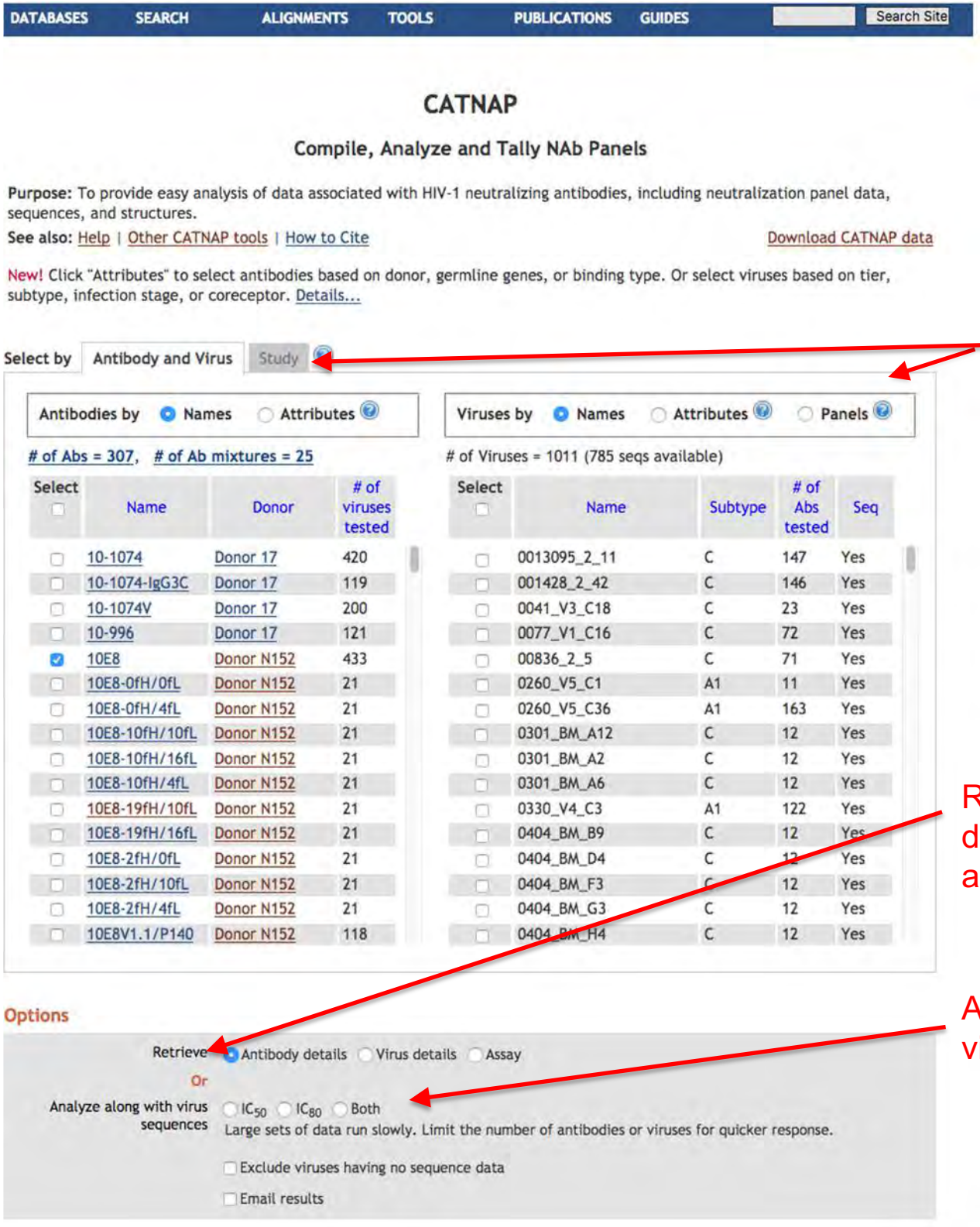
Find Names: quickly convert your list of virus or Ab names to the standardized ones in CATNAP

Custom Input accepts

- Numerical data: IC_{50}, ID_{50}, AUC, or *any* phenotypic data
- Aligned sequences associated with that data

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

Los Alamos NATIONAL LABORATORY



Many ways to select antibodies and viruses

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

Analyze IC_{50}, IC_{80} or both, along with the viral sequences

Los Alamos NATIONAL LABORATORY

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

Go to CATNAP main page

Antibody information

Number of antibodies: 1

Download heavy and light ◉ aa ○ na sequences in Fasta

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
10E8	○ C-term ○ gp41 MPER (membrane proximal external region)	4U6G 5IQ7 5IQ9 4G6F	Donor N152	10E8	Huang2012a	○ 10E8 contacts ○ 10E8 residue prediction ○ 10E8 signature predictions (West2013)	3-15*05	3-3*01	1*01	3-19*01	3*02	L	IC_{50} IC_{80}	

Download heavy/light chain sequences for just the selected set of mAbs.

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 Submit

Number of data: 1551

Download table below with additional virus info

Expand table below to show virus infomation

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

Los Alamos NATIONAL LABORATORY

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 (2648), PGT142 (2649), PGT143 (2650), PGT145 (2651), PGT144 (2652), PGDM1400 (3201), PGDM1401 (3203), PGDM1402 (3204), PGDM1403 (3205), PGDM1404 (3206), PGDM1405 (3207), PGDM1406 (3208), PGDM1407 (3209), PGDM1408 (3210), PGDM1409 (3211), PGDM1410 (3212), PGDM1411 (3213), PGDM1412 (3214), 3BNC117/PGDM1400 (3478), 3BNC117/PGT145 (3479), 8ANC195/PGT145 (3529), 8ANC195/PGDM1400 (3530), VRC01/PGDM1400-10E8v4 (3955), polyclonal Donor 84 (3966), PGDM1400/PRO-140 (3979)
Sequence Database Patient ID Record	85588

Link to patient's HIV sequences

List of all published mAbs derived from this patient

Los Alamos NATIONAL LABORATORY

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

Virus information

Number of viruses: 8

Download ◉aa ○na in Fasta

Download table below

More info in HIV Sequence DB

Download an ENV alignment of just the 8 selected viruses

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

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	EF117267	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	EF117266	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	HM215256		early		0260.5.36, 0260.v5.c36	0260_v5_c36	Yes	
0301_BM_A2	C	MALAWI	2008						HM070482				0301_bmA2	0301bmL_A2	Yes	
0330_V4_C3	A1	TANZANIA	2005				5 or 6	Heterosexual	HM215257	2	early		0330.v4.c3	0330_v4_c3	Yes	The sequence we provide is an upublished 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						HQ595836				0404_bmB9	0404bmL_B9	Yes	
0702_BM_H12	C	MALAWI	2008						HM070525				0702_bmH12	0702bmL_H12	Yes	
0735_V4_C1	A1C	TANZANIA	2005				5 or 6	Heterosexual	HM215259		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

Los Alamos NATIONAL LABORATORY

CATNAP: IC_{50} & IC_{80}/HIV-1 alignment

Collapse or expand details from individual studies

4 antibodies & 1007 viruses selected to search

More virus info in HIV Seq DB

*: Geometric means
(Expand to individual values)

●<0.1 ●<1 ●<10 ●≤50 ○>cutoff or 50 (μg/ml)

Virus name	Tier	10E8 IC50	10E8 IC80	PG9 IC50	PG9 IC80	PG1
001428_2_42	2	1.48106*	6.44387*	[illegible]	[illegible]	[illegible]
0041_V3_C18	2	0.44000	3.70700	[illegible]	[illegible]	1.8
0077_V1_C16	2	1.45938*	9.19287*	[illegible]	0.35301*	UD::
00836_2_5	1B or 2	0.50193*	1.90470*	49.00000*	UD:>50*	31.
0260_V5_C36		9.80180*	30.00417*	1.57709*	13.80908*	[illegible]
0301_BM_A12		-	-	0.66000	-	-
0301_BM_A2		-	-	0.91000	-	-
0301_BM_A6		-	-	0.91000	-	-
0330_V4_C3	2	1.04674*	3.83465*	[illegible]	[illegible]	[illegible]
0404_BM_B9		-	-	[illegible]	-	-
0404_BM_D4		-	-	6.87000	-	-
0404_BM_F3		-	-	[illegible]	-	-
0404_BM_G3		-	-	0.58000	-	-
0404_BM_H4		-	-	[illegible]	-	-
0439_V5_C1	2	1.16920*	4.69871*	UD:>50*	UD:>50*	UD::
0702_BM_B4		-	-	0.83000	-	-
0702_BM_B9		-	-	0.38000	-	-
0702_BM_D1		-	-	0.18000	-	-
0702_BM_H12		-	-	0.37000	-	-
0815_V3_C3	2	0.31354*	1.86820*	UD:>10, >25,...*	UD:>25, >50*	[illegible]
0907_V4_C12		-	-	-	-	-
Geometric mean of detected		0.35352	2.15327	0.18501	0.38293	[illegible]
Geometric mean of all (undetected set to 100)		0.46515	2.43935	0.92140	2.20904	0.9
% detected (detected/total)		95% (411/432)	97% (387/400)	74% (543/729)	69% (285/416)	64%

HXB2

```
MRVKE----KY-QHLW-RWG---WRWGTMLLG---MLMI---CSAT---
----.---------|--------.----|-------.-------|----
-------------10------------20--------------30----
MRVRGILR-NY-QQW--------WMWGVLGFW---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--------LRWGTLILG---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----
MRARAMKR-NC-QNL--------WKWGIMLLG---ILMI---CSAA----
```

Potency and Breadth of neutralization over multiple studies

# of viruses found: 799
# of Abs found: 4
# of studies found:70

Download neutralization data with virus information

Download ◉ aa ○ na in Fasta

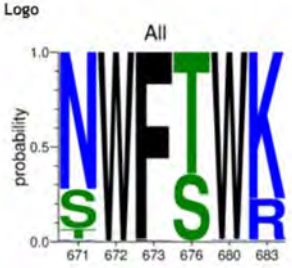
Go to antibody information section

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

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

- 10E8 contacts (LogoByAll LogoBySubtype): N671 W672 F673 T676 W680 K683
- PG9-like contacts (LogoByAll LogoBySubtype): N156 N160 I165 G167 K168 V169 Q170 K171 Y173

Logo



Frequency by position

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)	

Position analysis

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.

Analysis at position 160 for Ab PG9

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				

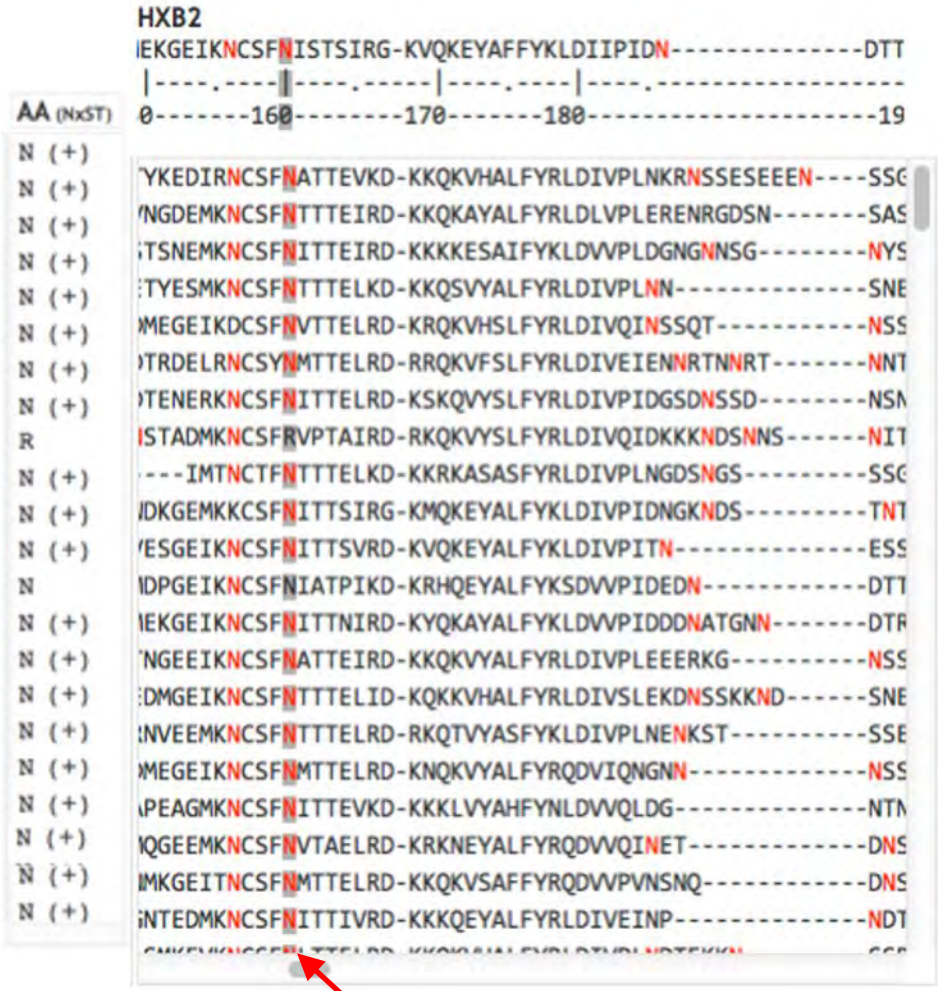
N-linked Glycosylation Motif Counts

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				

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

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



Env position 160 is highlighted

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

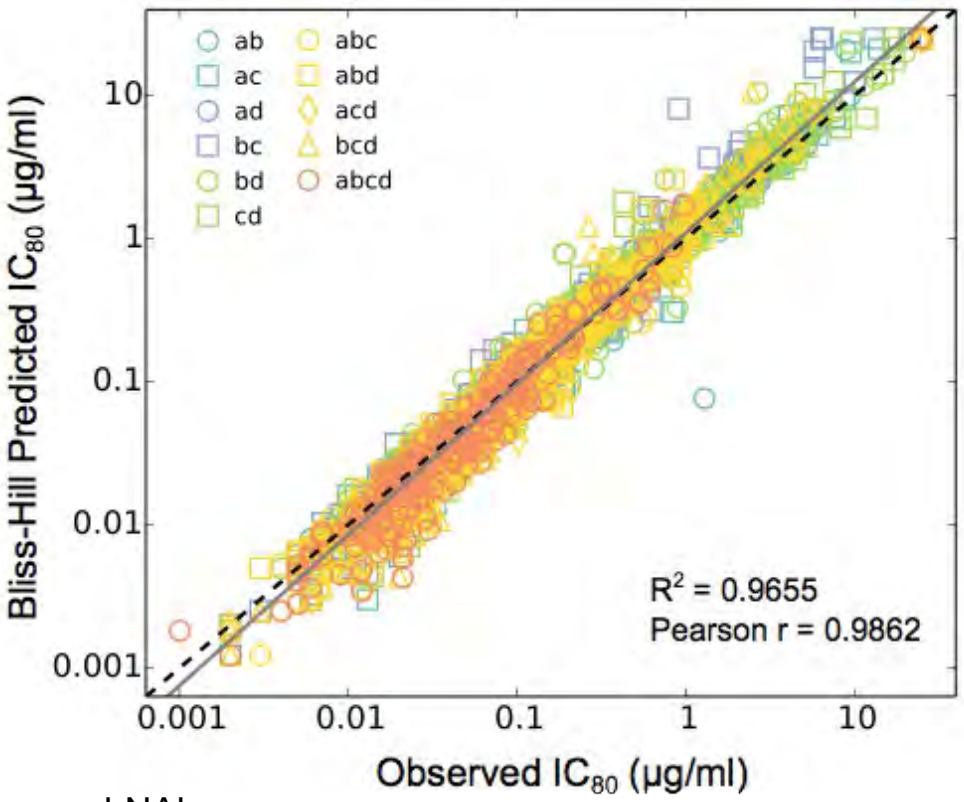
Mutation affects PG9-like Ab sensitivity: Loss of glycan confers resistance; PG9-like class includes PG16, PGT141, 145, CH01-CH04 (V1V2 glycan, Doria-RoseNA2012)
PG16 signature predictions: PG16: glycosylation at N160 is associated with increased susceptibility to neutralization; intermediate quality 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)

Los Alamos NATIONAL LABORATORY

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
  - 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.



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

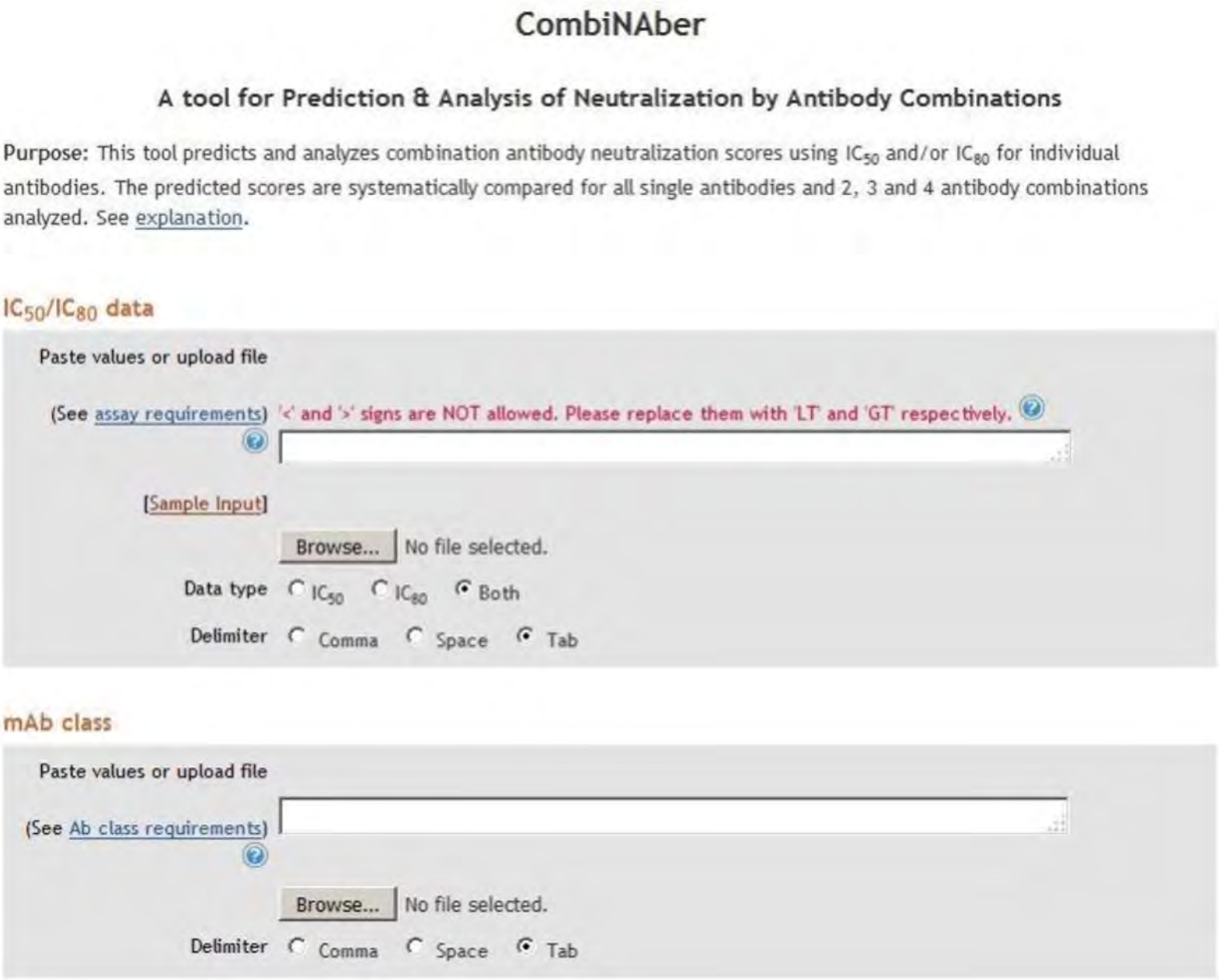
Wagh et al. 2016 PLoS Pathog 12(3): e1005520

- Purpose of CombiNAber:
  - Predict neutralization profiles of bNAb combinations using individual bNAb IC50 & IC80 data as input.
  - Systematically compare bNAb combinations to identify optimal combinations.

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

Los Alamos NATIONAL LABORATORY

CombiNAber Input Page



IC50 / IC80 data:

- Viruses on rows, mAbs on columns
- Giving both IC50 & IC80 data allows prediction using the more accurate Bliss-Hill Model

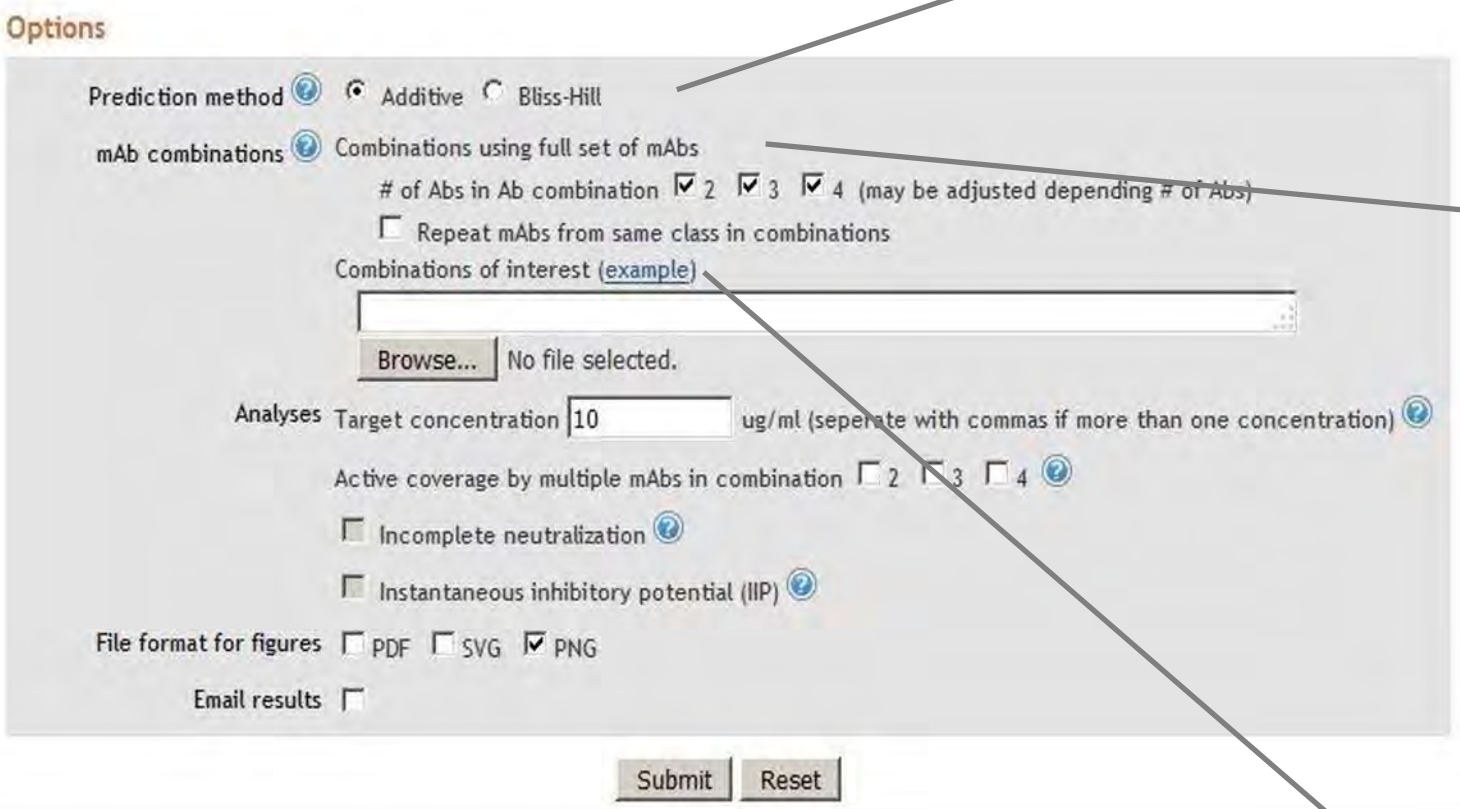
mAb class:

- 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.

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

Los Alamos NATIONAL LABORATORY

CombiNAber Input Page



Prediction method:

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

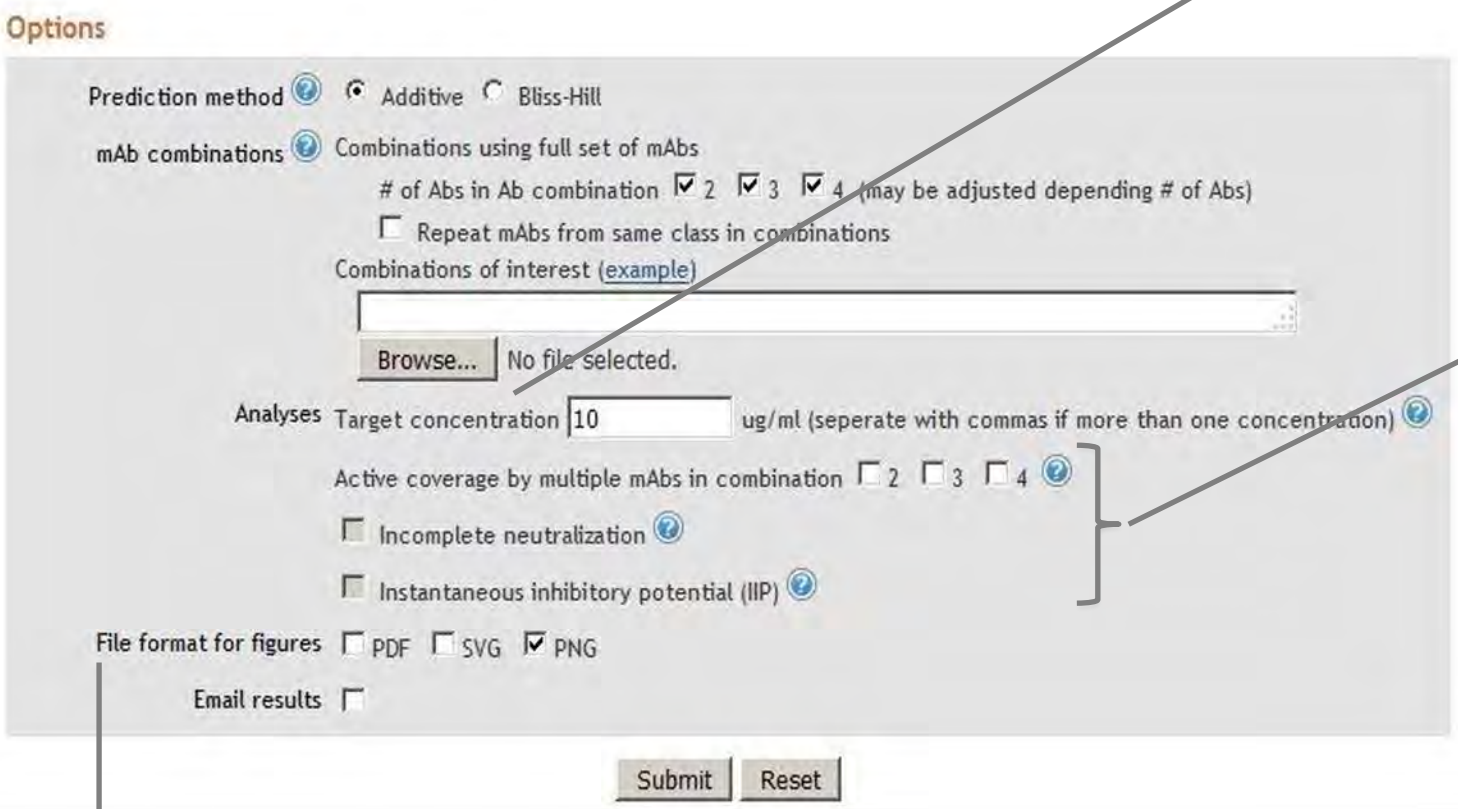
mAb combinations:

- 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

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

Los Alamos NATIONAL LABORATORY

CombiNAber Input Page



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

Los Alamos NATIONAL LABORATORY

CombiNAber Output Page

CombiNAber

Input & options:

IC_{50}/IC_{80} data: 15 ab(s) found, 25 virus(es) found. [See]
Data type: IC50 & IC80
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

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

All results can be downloaded in an archived zip file

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

Los Alamos NATIONAL LABORATORY

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

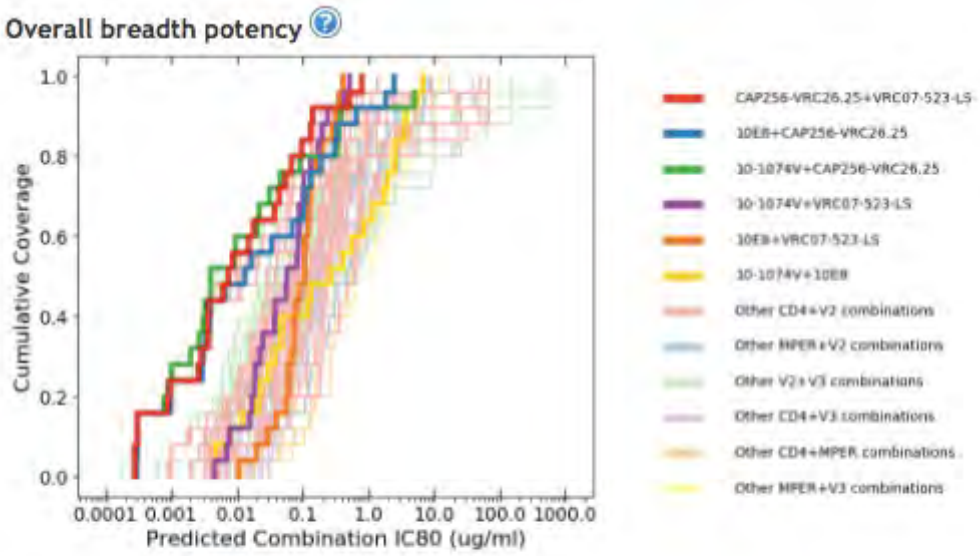
Combination	Class	IC80 Geometric Mean	Coverage at IC80 < 10.0 ug/ml	Percent Viruses with maximum inhibition > 0.95 at 10.0 ug/ml	Median IIP at 10.0 ug/ml	Percent Viruses with IIP > 5 at 10.0 ug/ml	Geometric Mean IC80 with at least 2 active at IC80 < 10.0 ug/ml	Coverage with at least 2 active at IC80 < 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

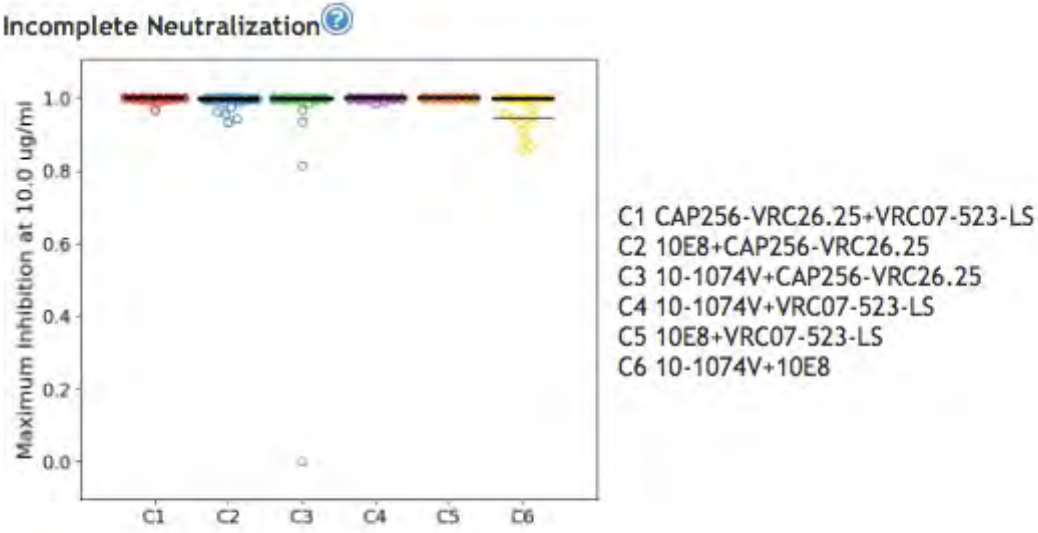
Table shows the best 2 mAb combination of each epitope class (e.g. CAP256-VRC26.25 + VRC07-523-LS is cd4+v2 and 10E8 + CAP256-VRC26.25 is mper+v2) using all the metrics in the table. Overall rank in 2mAb combinations is also shown.

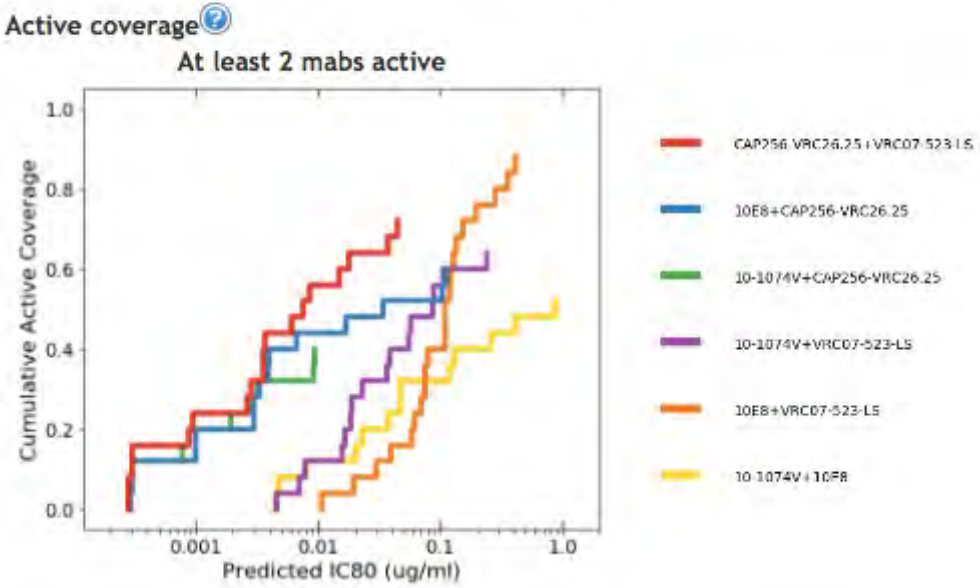
www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

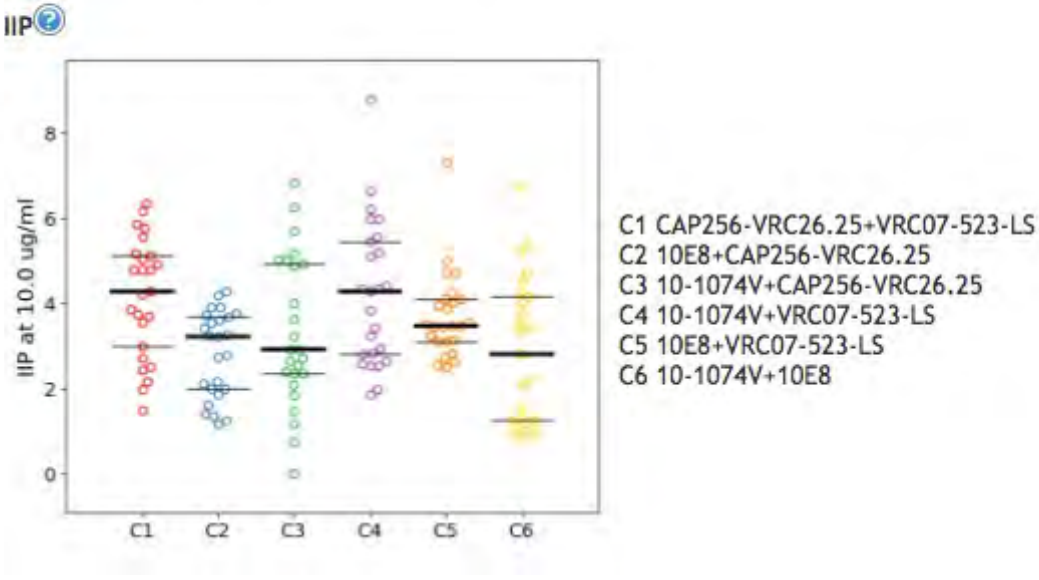
Los Alamos NATIONAL LABORATORY

CombiNAber Output Page









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)

www.hiv.lanl.gov/content/sequence/COMBINABER/combinaber.html

Los Alamos NATIONAL LABORATORY

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; immuno@lanl.gov;
Technical Questions & Suggestions: kshitij@lanl.gov

GenSig

Purpose:

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

Background:

- Bette Korber, Tanmoy Bhattarcharya 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

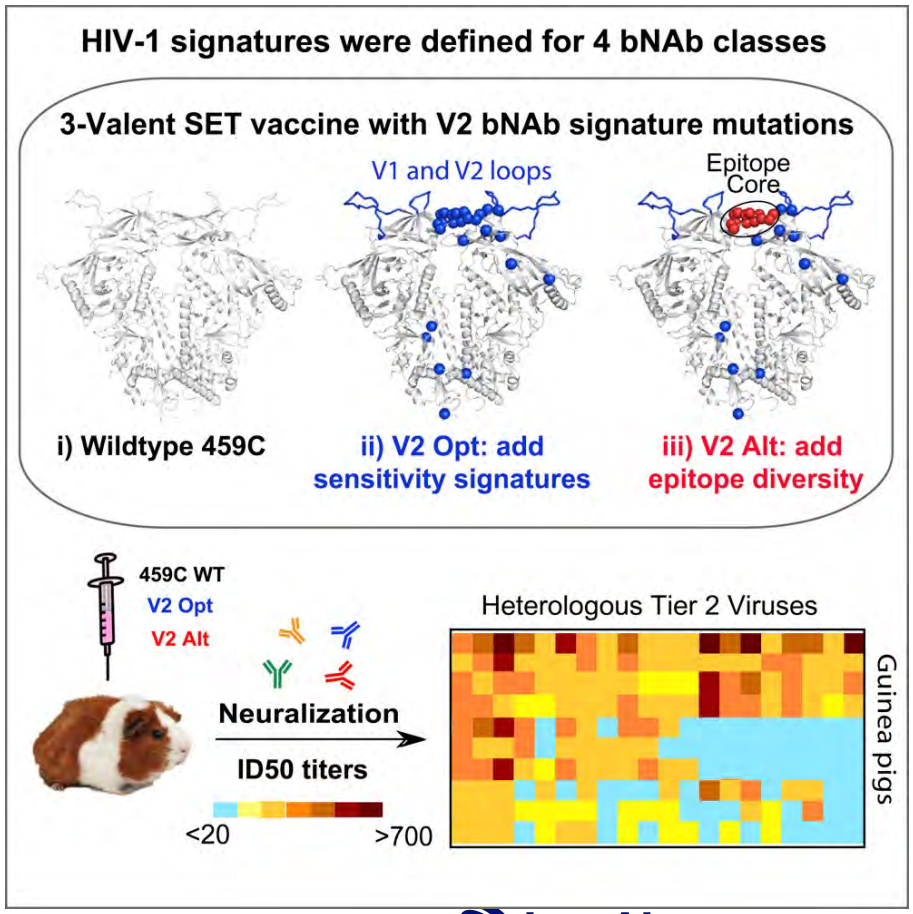
- We found robust amino acid and glycan signatures associated bNAb sensitivity/resistance. These signatures were used to design a trivalent vaccine cocktail that induced broad heterologous tier 2 neutralization in guinea pigs.
  Bricault et al. Cell Host & Microbe 2019 25(1):59-72

Cell Host & Microbe
Volume 25, Issue 1, 9 January 2019, Pages 59-72.e8

Article

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

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



Los Alamos NATIONAL LABORATORY

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

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.

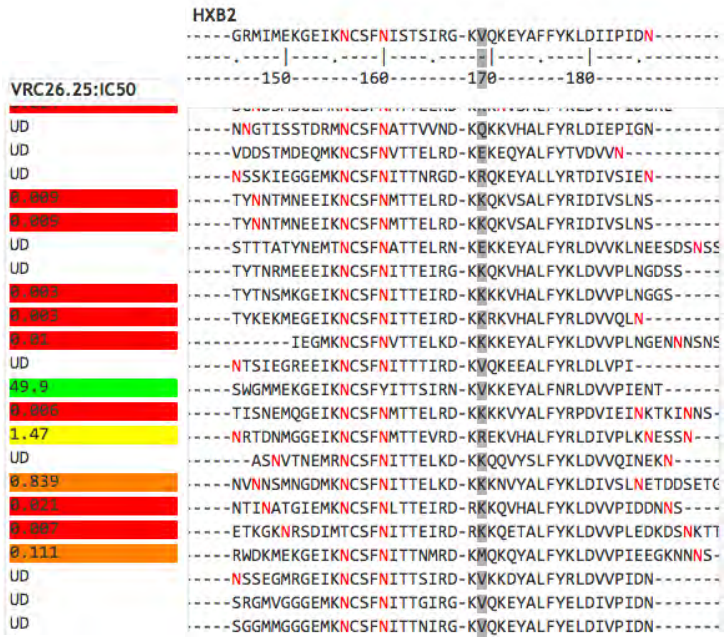
- Option for potential N-linked glycosylation sites (PNGS).

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.

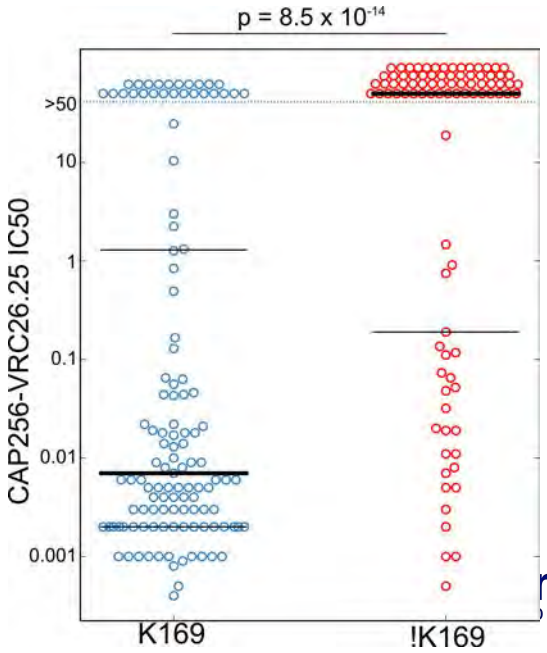
Phylogenetic correction (next slide!)

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



	K169	!K169
Neutralized	94	25
Not-neutralized	27	59

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



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

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

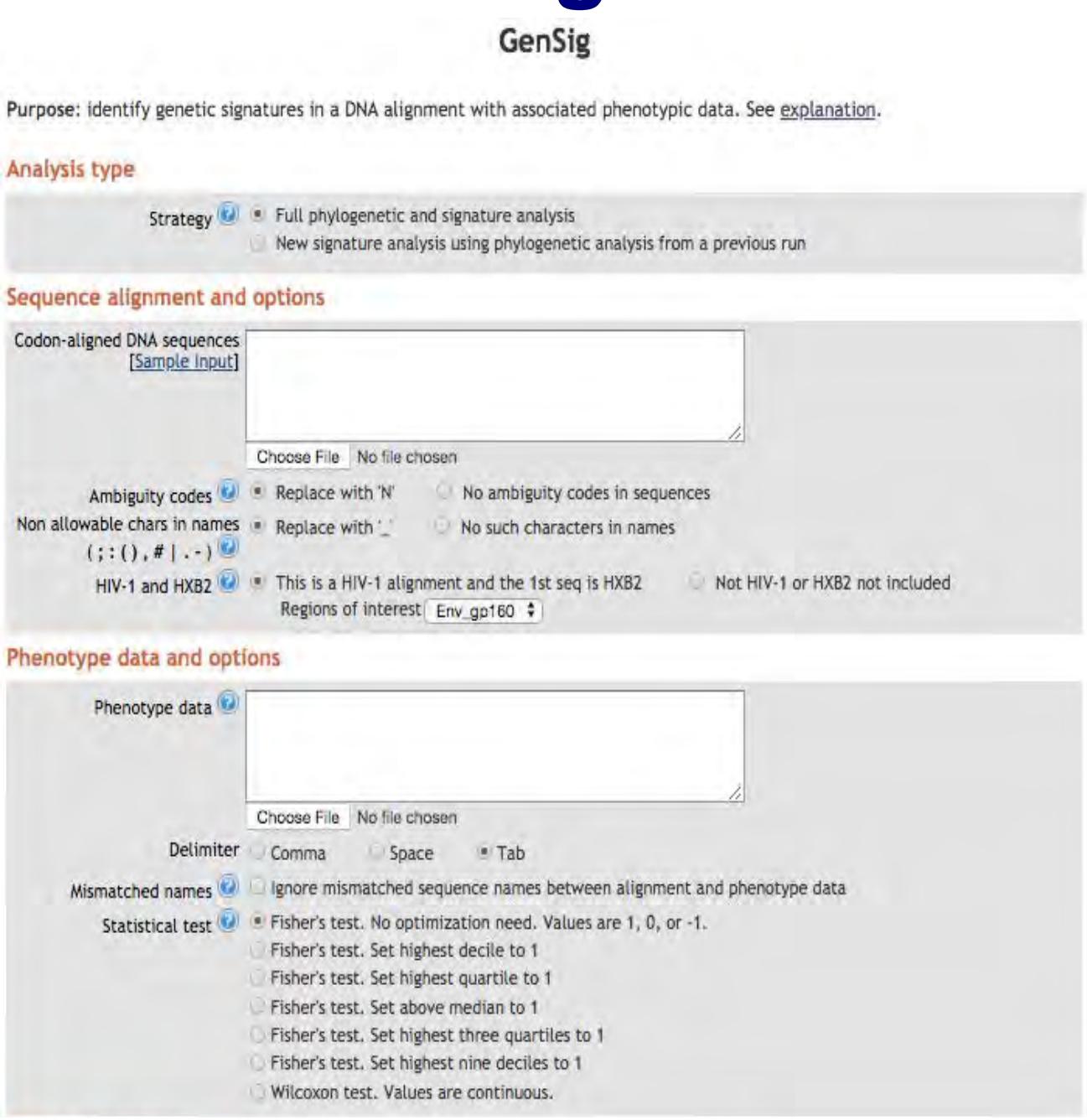
True Signature N-334

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

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

GenSig Input Page



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.

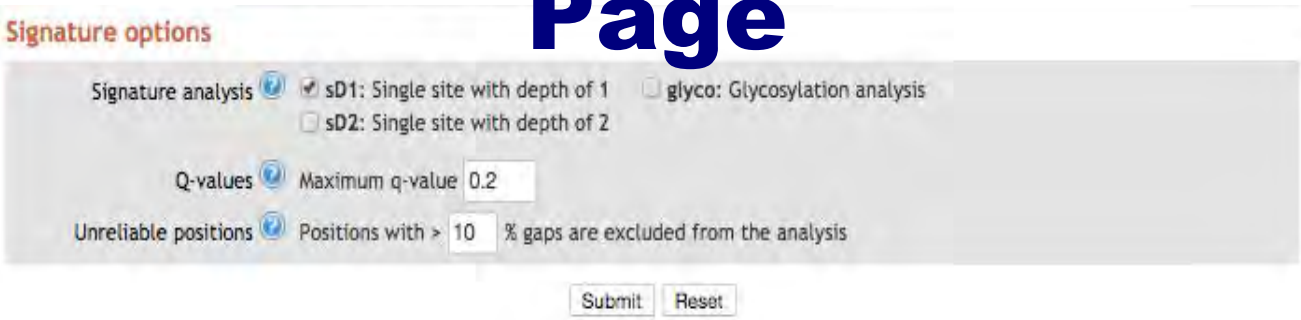
Phenotype data:

- 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.

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

GenSig Input Page



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.

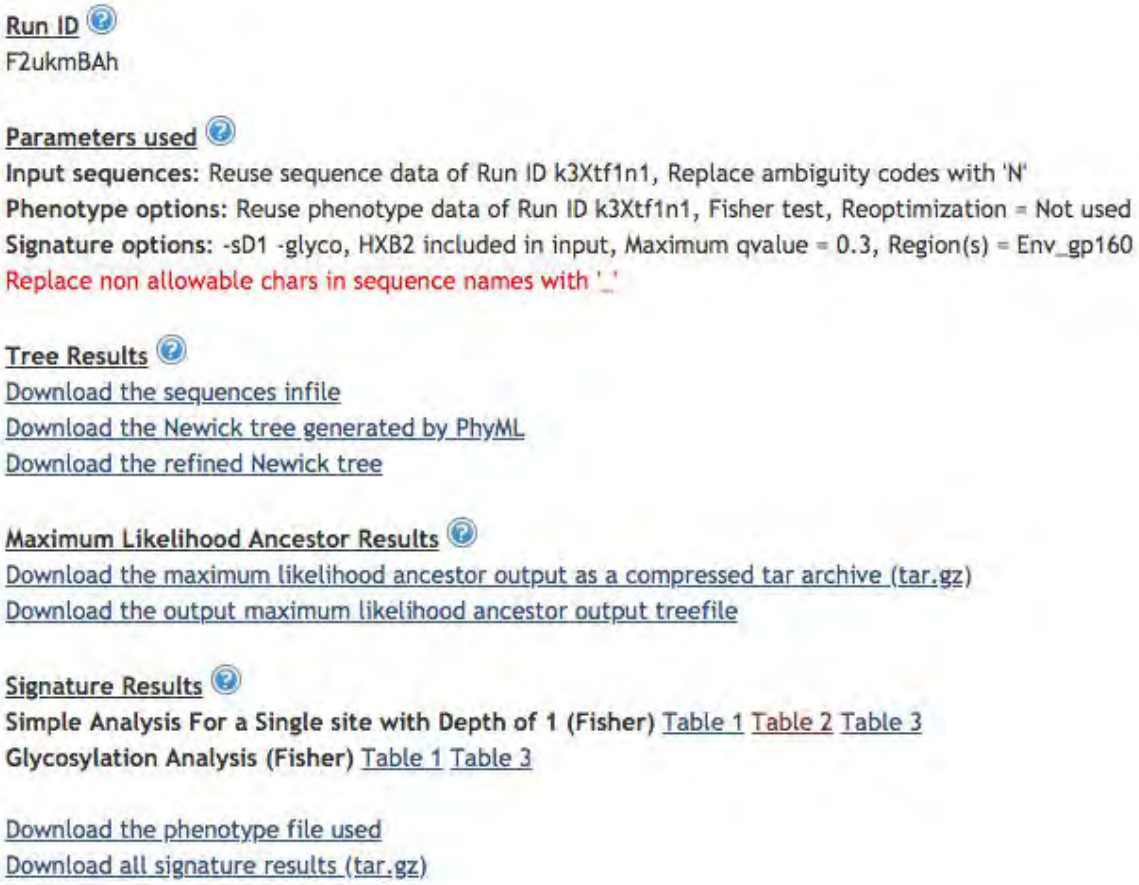
- 0.2 is a reasonable cutoff – 80% chance of true positive, 20% of false positive.

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.

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

GenSig Output Page



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)

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.

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

Los Alamos NATIONAL LABORATORY

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	Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.28	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	Tree
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	Tree
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	Tree

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.

Los Alamos NATIONAL LABORATORY

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	Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.286	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	Tree
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	Tree
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	Tree

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.

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

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	Tree
T1	Fisher:1,0,-1	K4	6234	Env_gp160	4	10	M	ch04	0.000245	50	33	96	18	0.00676	0.286	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	E	ch04	0.0579	70	13	82	32	0.223	2.09	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	5	13	G	ch04	0.0624	14	69	33	81	0.228	0.5	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	Q	ch04	0.00369	38	45	76	38	0.0362	0.424	Tree
T1	Fisher:1,0,-1	E5	6237	Env_gp160	7	19	R	ch04	0.000234	73	10	74	40	0.00643	3.92	Tree
T1	Fisher:1,0,-1	K6	6240	Env_gp160	9	25	N	ch04	0.0396	3	80	14	100	0.183	0.269	Tree
T1	Fisher:1,0,-1	L851	8775	Env_gp160	948	2842	F	ch04	0.0605	34	49	63	51	0.226	0.563	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	E	ch04	0.00277	27	56	16	98	0.0325	2.94	Tree
T1	Fisher:1,0,-1	E106	6540	Env_gp160	118	352	T	ch04	0.00145	61	22	104	10	0.0195	0.268	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	E	ch04	0.0206	79	4	96	18	0.115	3.68	Tree
T1	Fisher:1,0,-1	Q114	6564	Env_gp160	126	376	Q	ch04	0.0206	4	79	18	96	0.115	0.272	Tree

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).

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

Los Alamos NATIONAL LABORATORY

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

Table 3:
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.

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 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
T3	Fisher:1,0,-1	N160	6702	Env_gp160	193	577	N	ch04	0.000761	0	83	13	101	0.113	0	Tree
T3	Fisher:1,0,-1	K171	6735	Env_gp160	205	613	K	ch04	2e-08	4	77	36	54	1.84e-06	0.079	Tree

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 2019https://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

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.

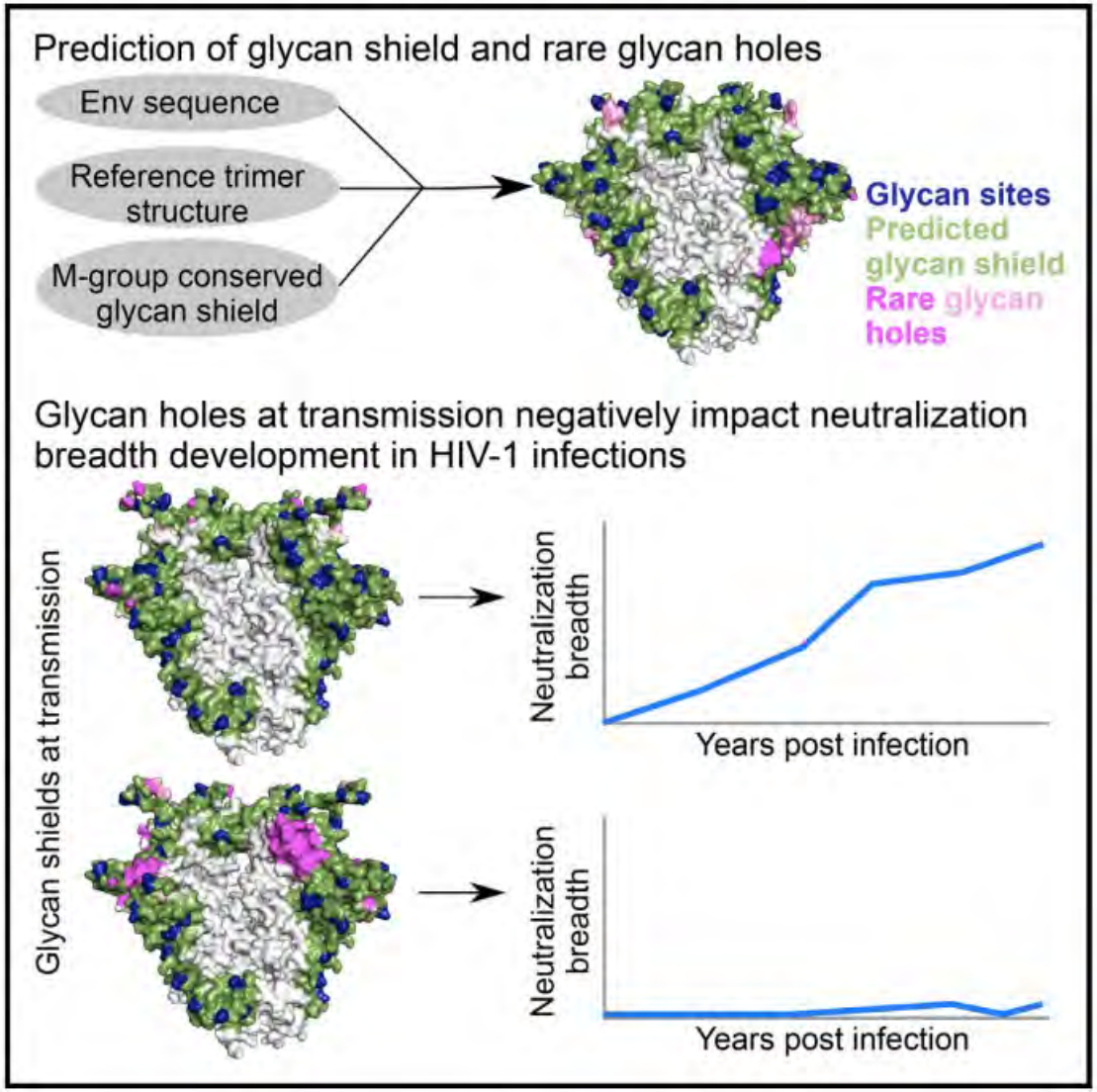
We developed a strategy to accurately predict rare glycan holes given an input Env sequence.
(Wagh et al. Cell Rep 2018 25:893-908)

- 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.



https://www.hiv.lanl.gov/content/sequence/GLYSHIELDMAP/glyshieldmap.html

Los Alamos NATIONAL LABORATORY

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.

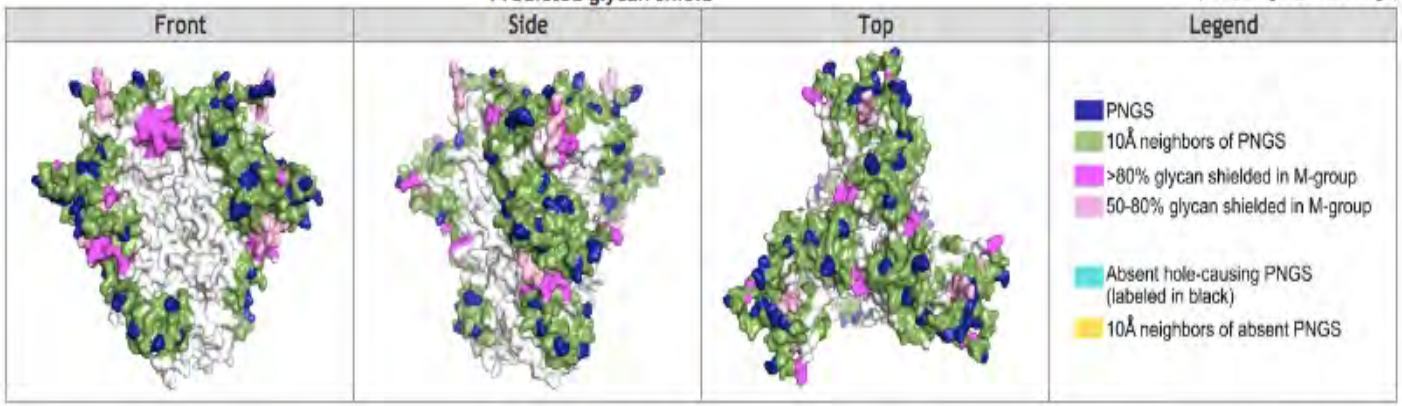
Results:

Download in an archived zip

JRFL : Total glycan hole area (>50% M-group conserved, A^2) = 4721.37

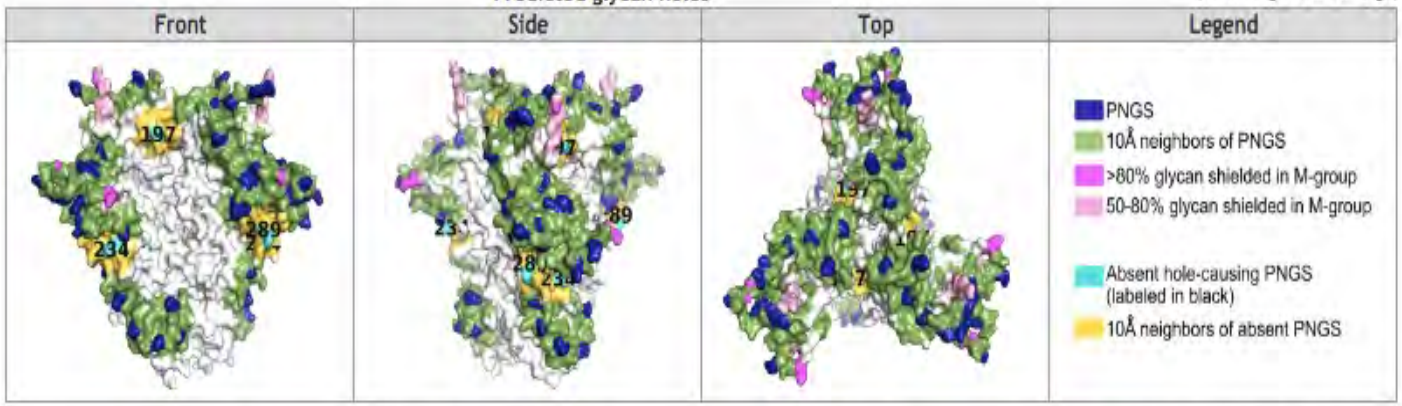
Predicted glycan shield

(click images to view large)



Predicted glycan holes

(click images to view large)



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

Los Alamos NATIONAL LABORATORY

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

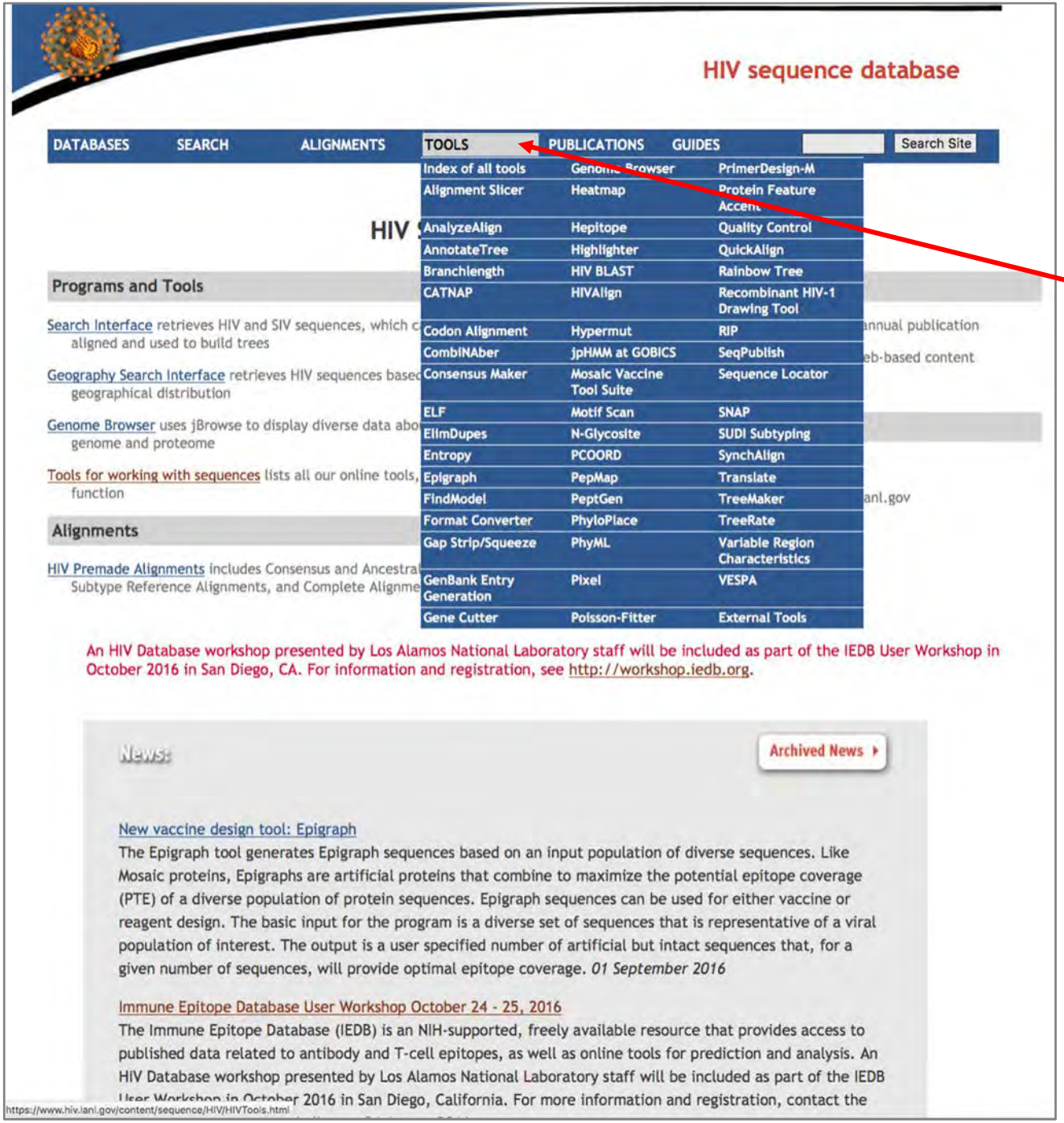
Technical questions & suggestions: kshitij@lanl.gov

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



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.

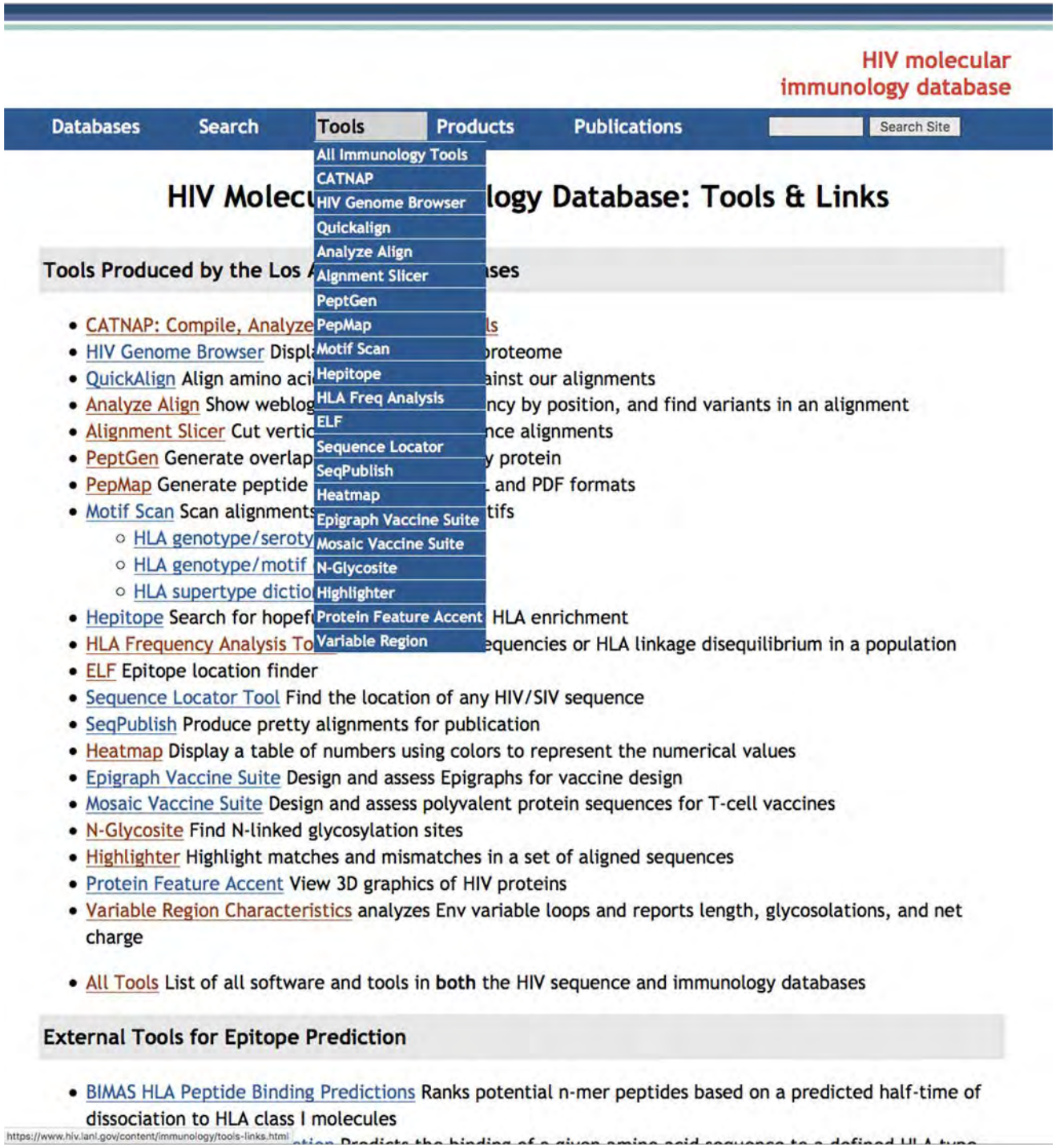
Most tools have explanation pages, and sample data sets.

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

Los Alamos NATIONAL LABORATORY

HIV Immunology Tools are a subset of the HIV Sequence Tools

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



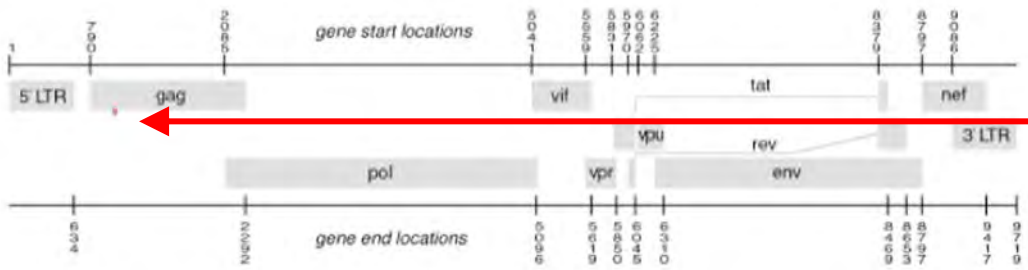
Tools especially useful from immunologists can be accessed from the HIV Immunology "Tools" page

Los Alamos NATIONAL LABORATORY

HIV/SIV Sequence Locator Tool

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

Paste or type a DNA or protein sequence here.



Location in genome mapped in red.

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

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%):

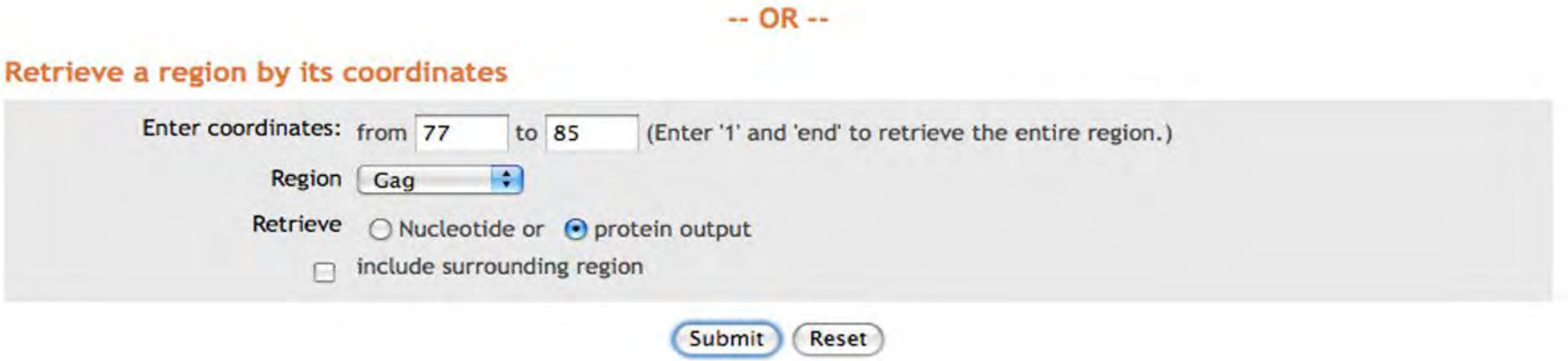
```
Query SLYNTVATL  9
      :::::::::
HXB2  SLYNTVATL
```

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

Los Alamos NATIONAL LABORATORY

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)



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

Los Alamos NATIONAL LABORATORY

PepMap www.hiv.lanl.gov/content/sequence/PepMap/pepmap.html

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

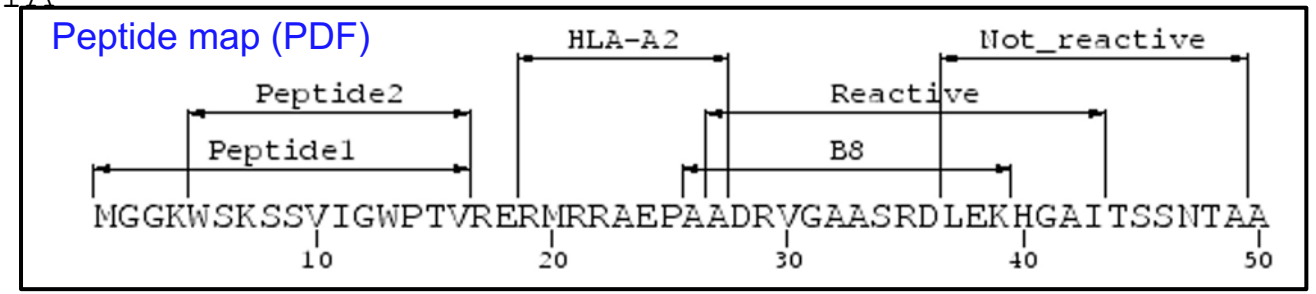
Input:

```
Peptide1      MGGKWSASSVIGGPTV
Peptide2      WSKSSVIGWVTV
HLA-A2        RMRRAEPAV
B8            AADRVGAASRDLEK
Reactive      ADRVGAASRDLEKHGAI
Not_reactive  LEKHGAITSSNTA
```

Peptide map (FASTA)

```
>B.FR.83.HXB2_LAI_IIIB_BRU_K03455
MGGKWSKSSVIGWPTVRERMRRAEPAADRVGAASRDLEKHGAITSSNTAA

>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	AADRVGAASRDLEK	AADRVGAASRDLEK	Nef	26-39	-	-	100.0
Reactive	ADRVGAASRDLEKHGAI	ADRVGAASRDLEKHGAI	Nef	27-43	-	-	100.0
Not_reactive	LEKHGAITSSNTA	LEKHGAITSSNTA	Nef	37-49	-	-	100.0

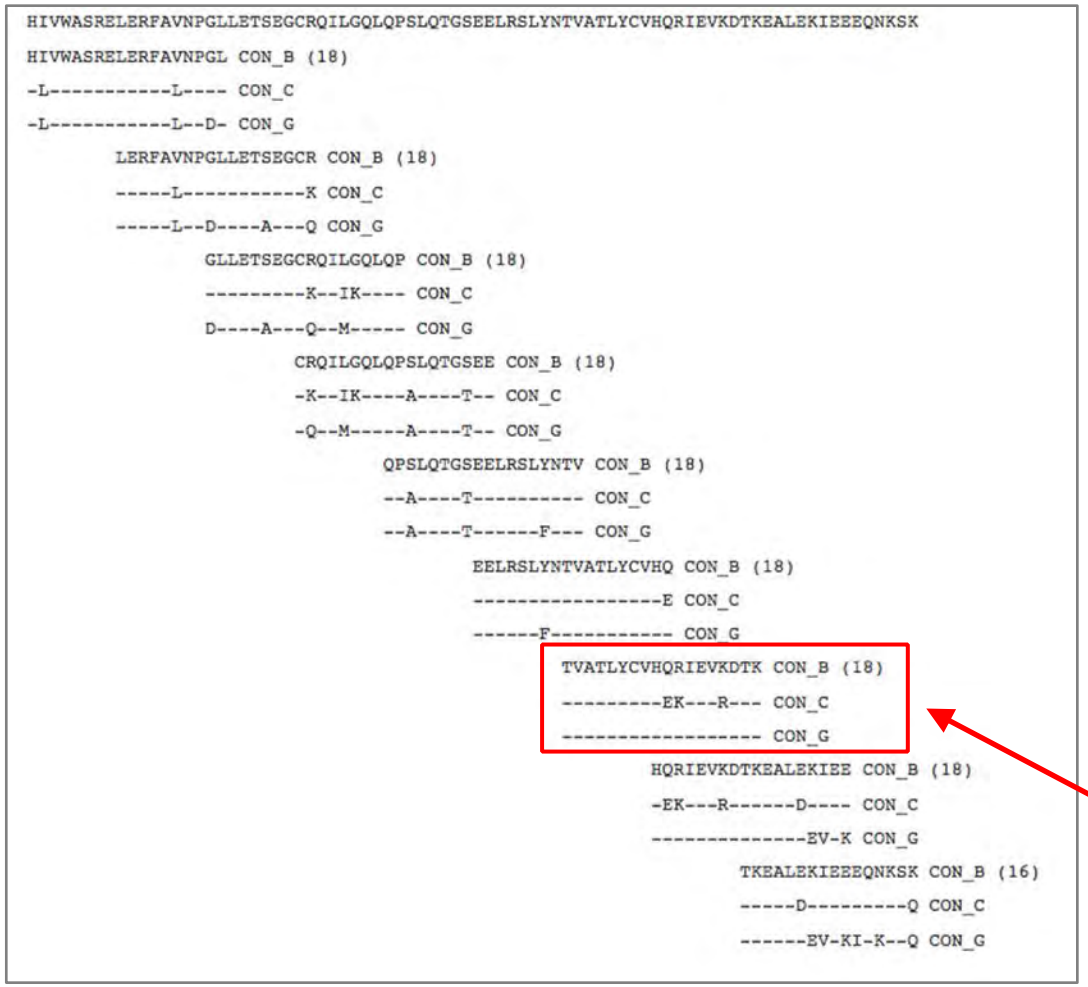
Los Alamos NATIONAL LABORATORY

PeptGen

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

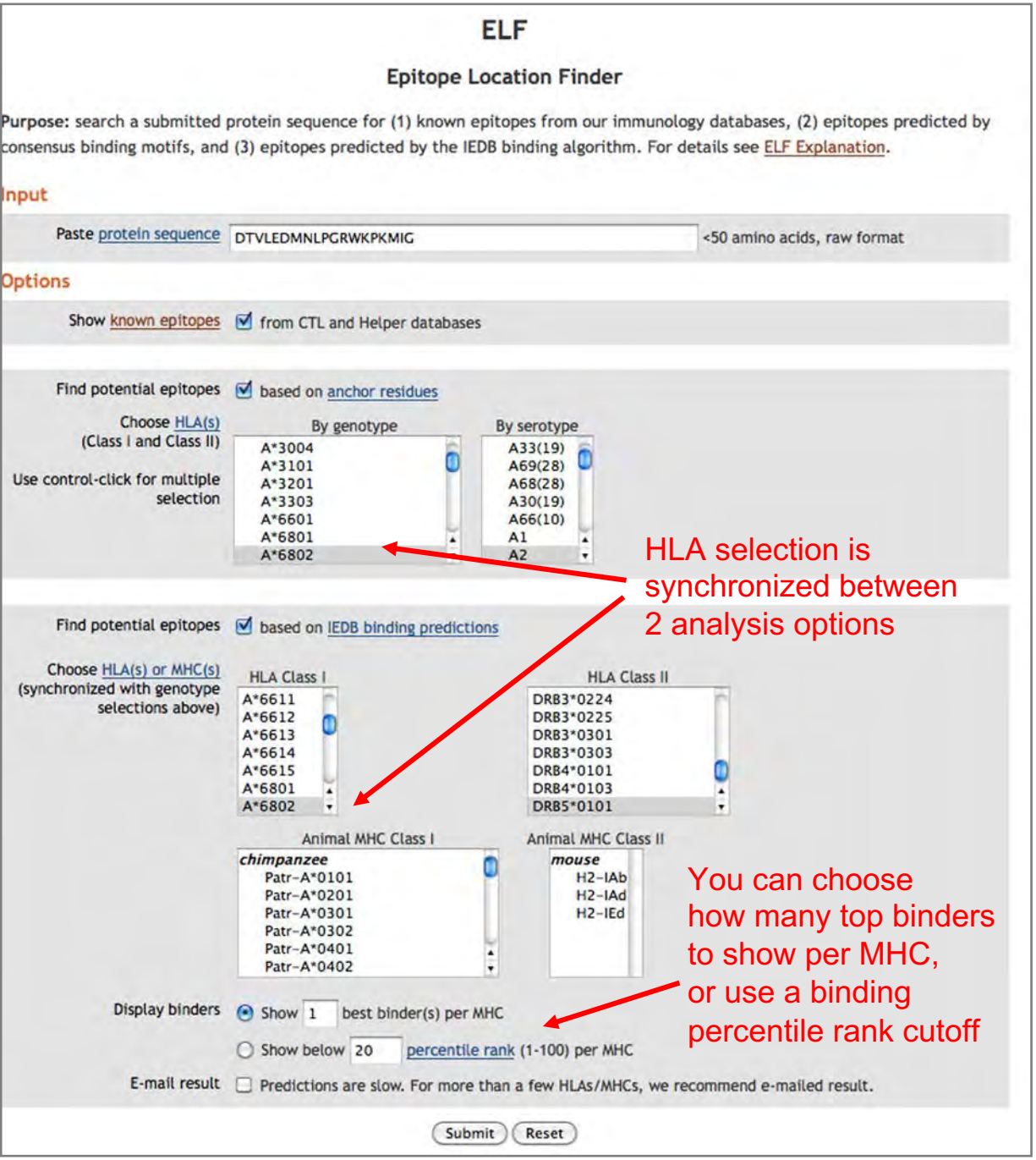
- Generates overlapping peptides for any protein sequence
- Takes alignment as an input and removes duplicate peptides

```
Seq1        HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
Seq2        HLVWASRELERFALNPGLLETSEGCKQIIKQLQPALQTGTEELRSLYNTVATLYCVHEKIEVRDTKEALDKIEEEQNKSQ
Seq3        HLVWASRELERFALNPDLLETAEGCQQIMGQLQPALQTGTEELRSLFNTVATLYCVHQRIEVKDTKEALEEVEKIQKKSQ
```




```
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 GLLETSEGCRQILGQLQP 3 s1 1  s1 - -
8 GLLETSEGCKQIIKQLQP 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 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

https://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html

Los Alamos NATIONAL LABORATORY

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 table

DTVLEDMNLPGRWKPKMIG

DTVLEEMNL A*6802 align ◀ Epitopes
DTVLEDINL A*6802 align ◀ matching
DTVLEEWNL A*6802 align ◀ requested HLAs
DTVLEEMNL A68 align
DTVLEEMNL A28 align
DTVLEDMNL align
EEMNLPGRW B44 align
EEINLPGKW B44 align
EEMNLPGRW B*4402 align
EEMNLPGRW B*4403 align
EEMNLPGRW B18,B40,B44 align
EDMNLPGRW align
EEMNLPGRW B*44 align
EEINLPGKW B*4403 align
EEMNLPGRW 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.

Download this alignment in format table

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

Motifscan

https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan

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

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
     DMNLPGRW      B*1501 (26)
      MNLPGRWK     A*6802 (3.0)
```

Class II

Selected allele(s): DRB5*0101

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
 TVLEDMNLPGRWKPK    DRB5*0101 (17.17)
```

IEDB binding predictions

Clicking on MHC links to the full list of IEDB predictions for that MHC (see next table)

Los Alamos NATIONAL LABORATORY

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	DTVLEDMNLPGRWKPKMIG

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	VLEDMNLPGRW	*NetMHCpan*	27
HLA-B*15:01	1	3	11	9	VLEDMNLPG	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	TVLEDMNL	*NetMHCpan*	36
HLA-B*15:01	1	2	11	10	TVLEDMNLPG	*NetMHCpan*	47
HLA-B*15:01	1	4	11	8	LEDMNLPG	*NetMHCpan*	48

Los Alamos NATIONAL LABORATORY

QuickAlign

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.97.97TZ03      --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

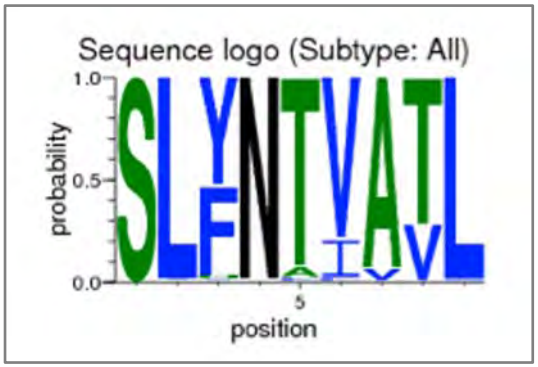
Total sequences = 23
Number of variants = 7

Variant frequency summary

Frequency by position — Go to top

See full raw counts — cutoff: 95%

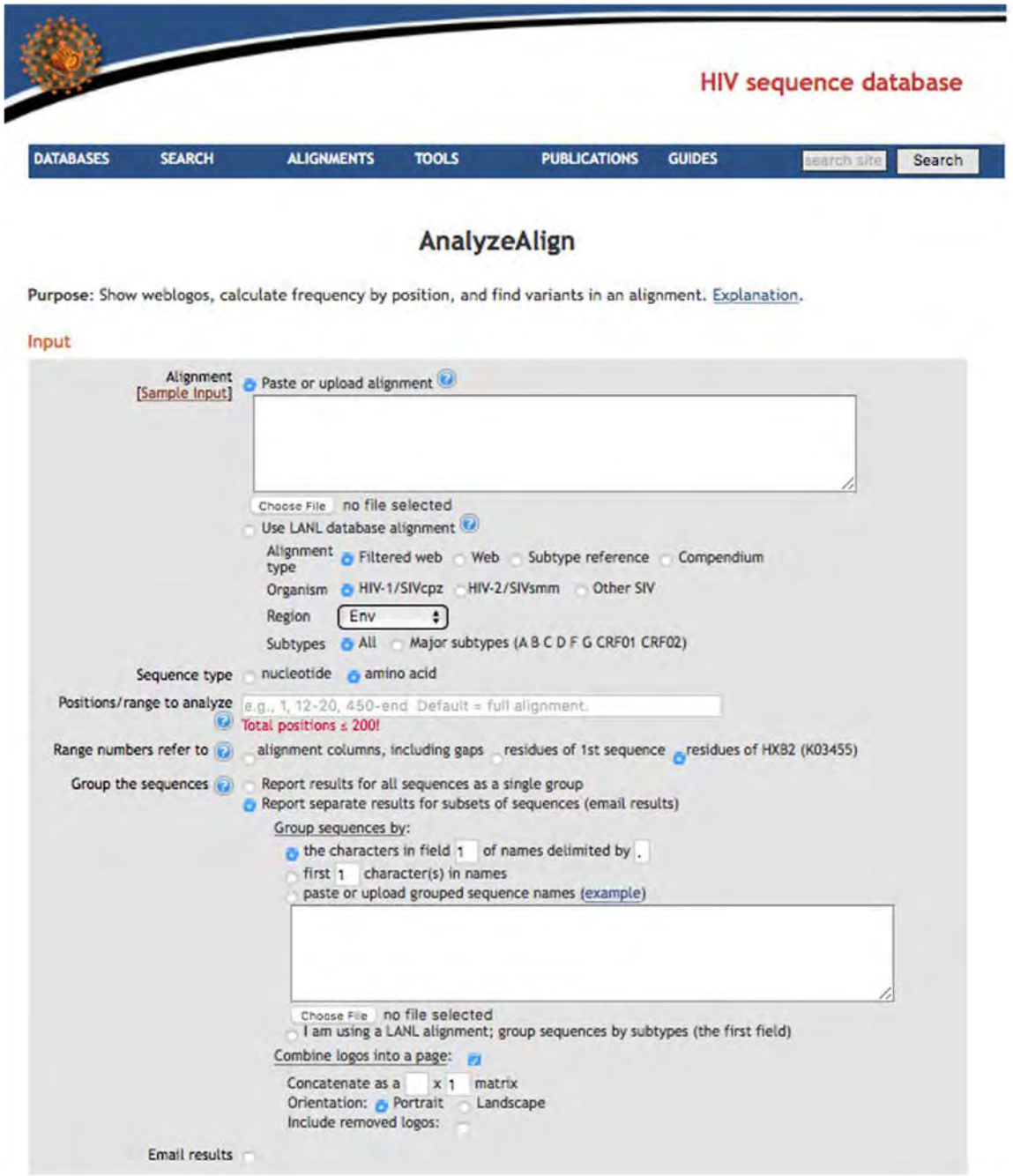
Position	Percentage and raw count of non-gap	Non-gap/total (percentage)
1	S: 99.90% (3113) other: 0.10% (3)	3116/3119 (100.00%)
2	L: 98.90% (3068) other: 1.10% (34)	3102/3119 (99.55%)
3	Y: 52.71% (1633) F: 43.77% (1356) other: 3.52% (109)	3098/3119 (99.42%)
4	N: 99.68% (3104) other: 0.32% (10)	3114/3119 (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%)



Los Alamos NATIONAL LABORATORY

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



https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html

Los Alamos NATIONAL LABORATORY

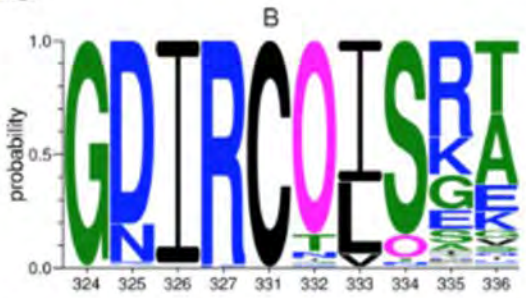
AnalyzeAlign Output

Groups
[Download combined logs PDF EPS]
B A1 A2 C D F1 F2 G 01_AE 02_AG

Group B

Go to top

Logo



Download: PNG PDF EPS

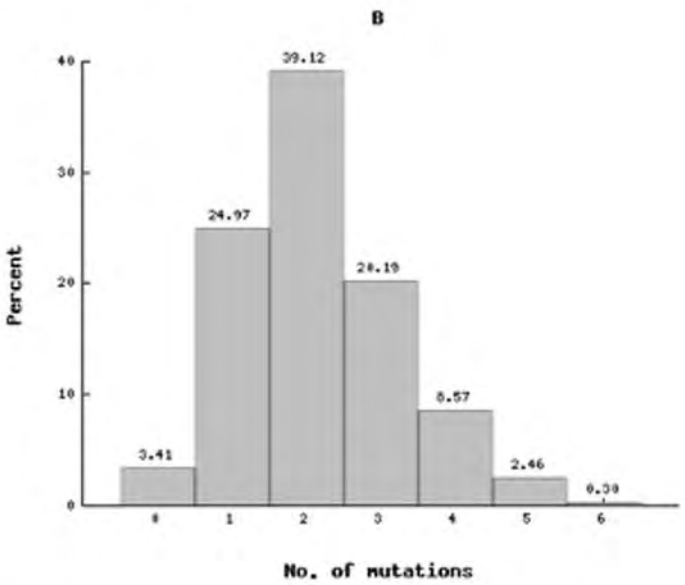
Frequency by position

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%)

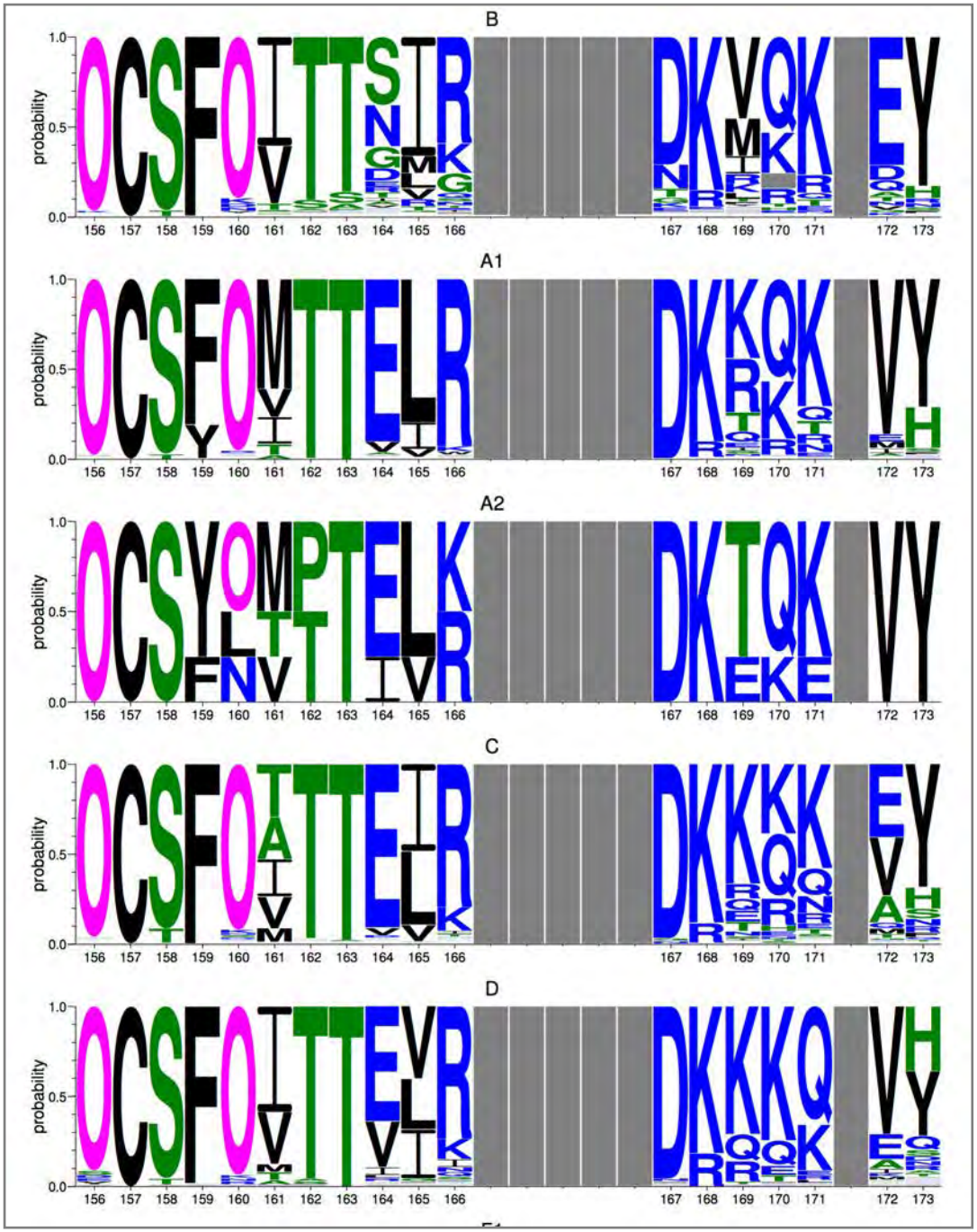
Sequence variants



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--------A	28	1.45	2
--------GA	27	1.39	2
--------GE	27	1.39	2
--------X-	26	1.34	1
--------XE	24	1.24	2
------L--E	20	1.03	2
-----TLOG-	19	0.98	4

Los Alamos NATIONAL LABORATORY

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 0)

Los Alamos
NATIONAL LABORATORY

AnalyzeAlign

Transmitted HIV virus

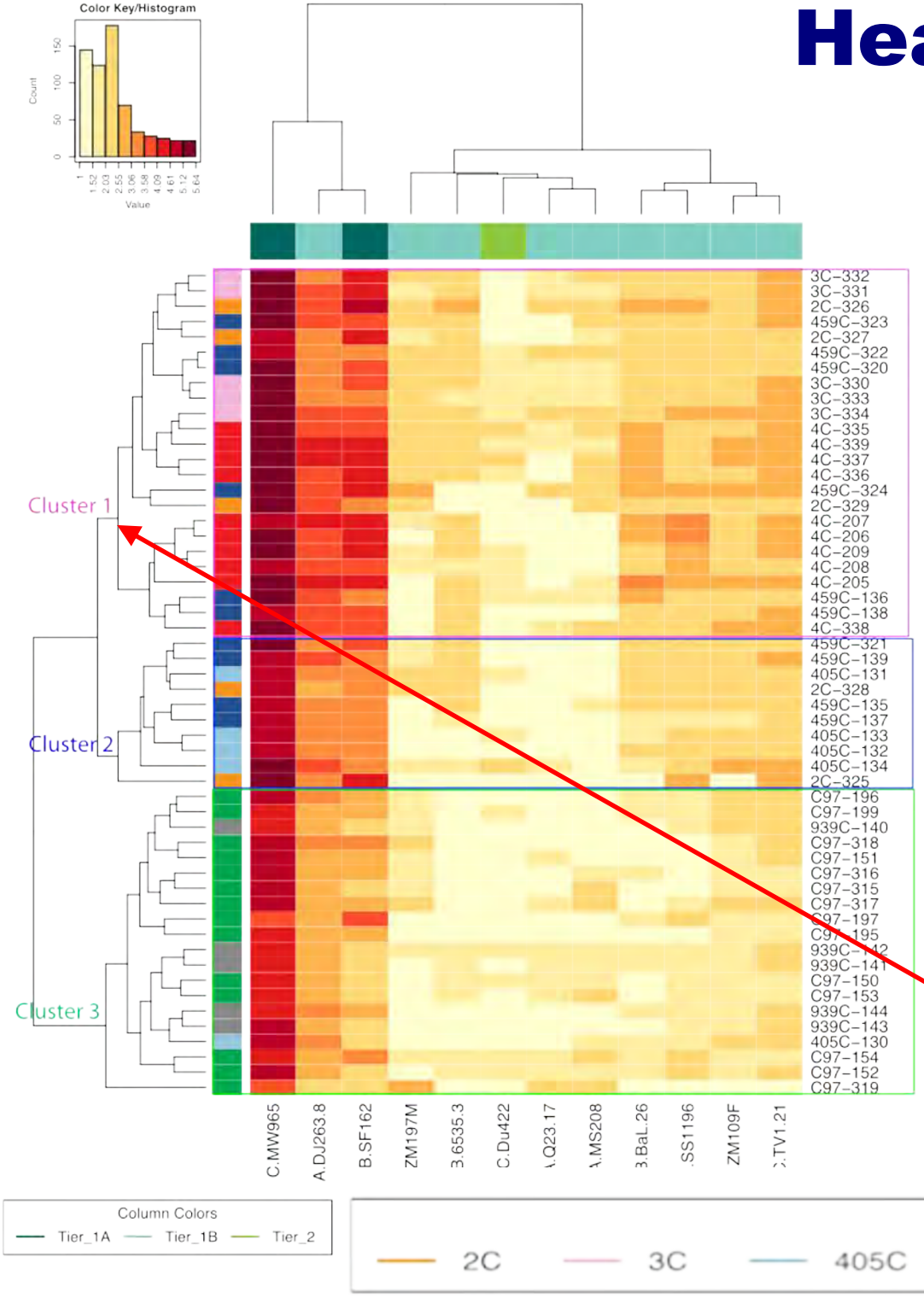
- ❑ Longitudinal samples over time
- ❑ Discontinuous positions under apparent immune selection
- ❑ Only differences from transmitted virus are shown
- ❑ Colors indicate amino acid charge categories

Virus 3 years post-infection

Figure from **Hraber et al. *Viruses* 2015**

Los Alamos NATIONAL LABORATORY

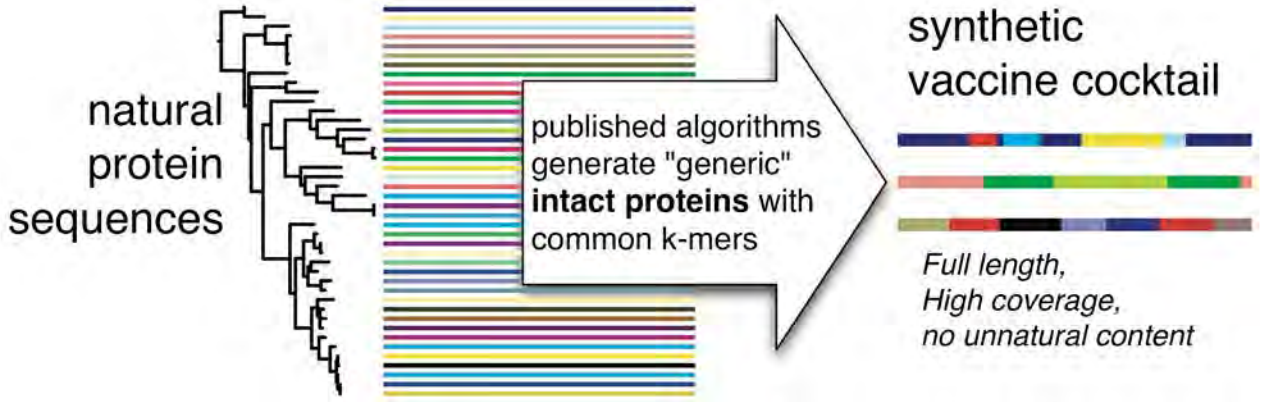
Heatmap



- 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

https://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap_mainpage.html

Vaccine Design Tools (Mosaic/Epigraph)



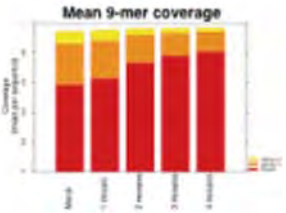
Design Tools

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)

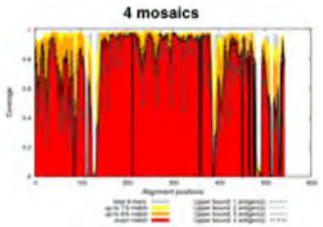
Epigraph — graph theoretic approach (Theiler et al. 2016)

Evaluation tools



Epitope Coverage Assessment (EPICOVER)

Alignment-independent "k-mer" coverage by vaccines or peptides.



Positional Epitope Coverage Assessment (POSICOVER)

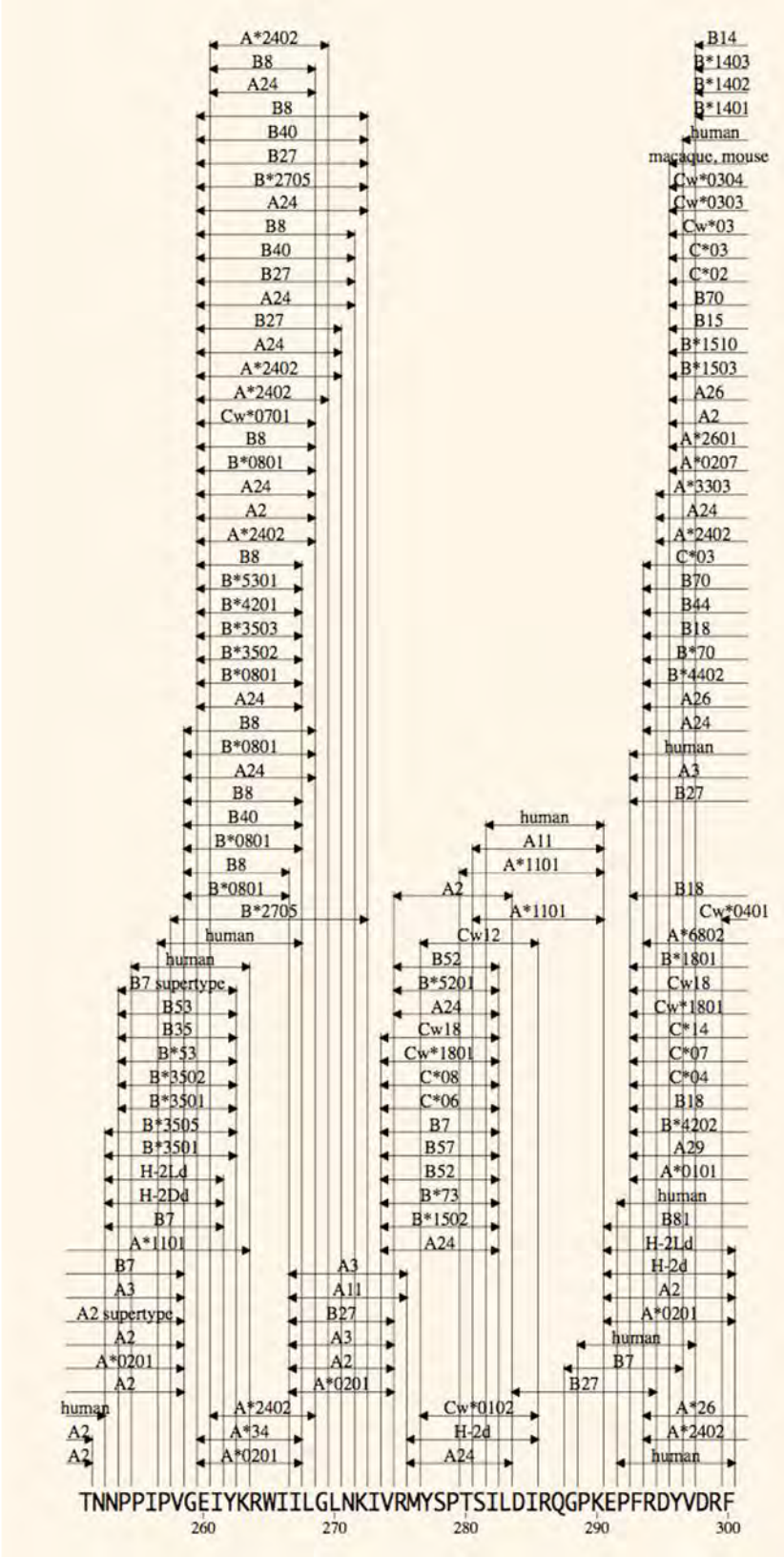
Alignment-based coverage by vaccines or peptides

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

Los Alamos NATIONAL LABORATORY

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

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

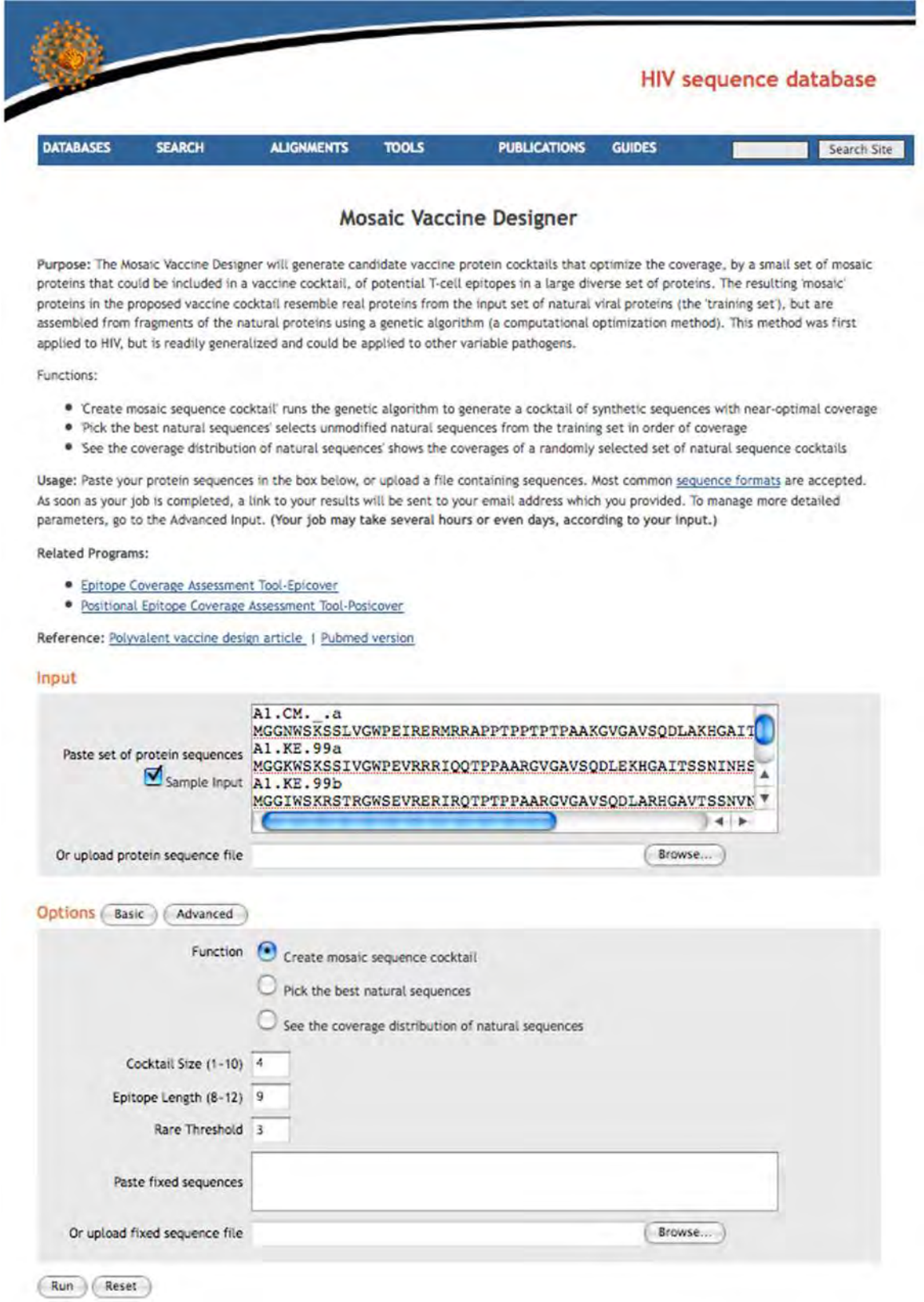
HIV, SIV, HCV, *Chlamydia*

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.

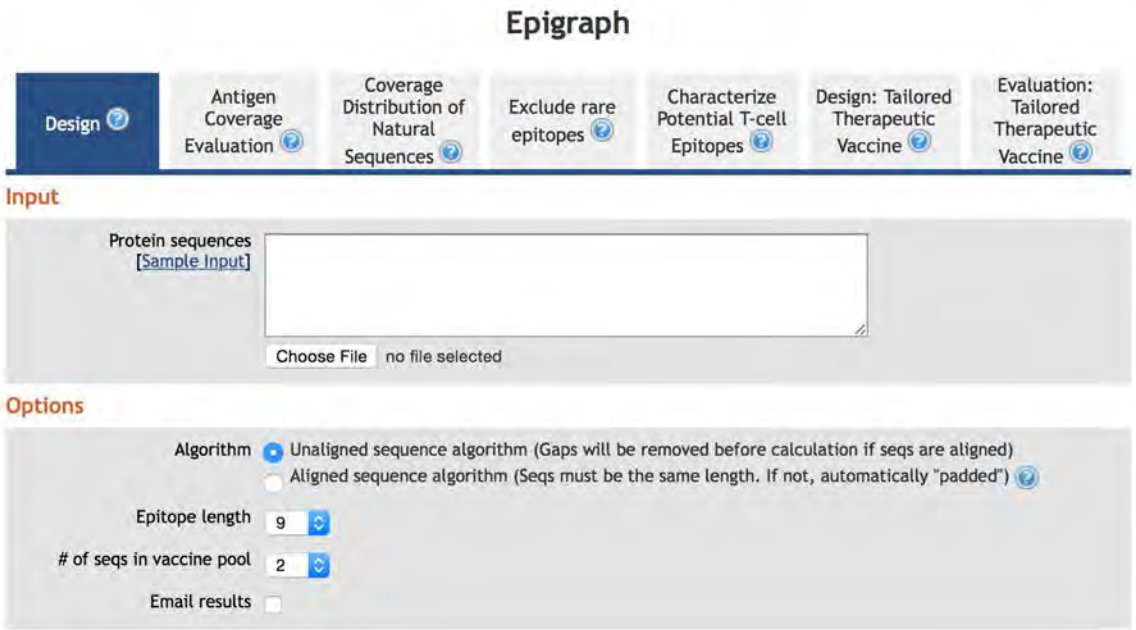
Epitope length: default is 9 amino-acids.

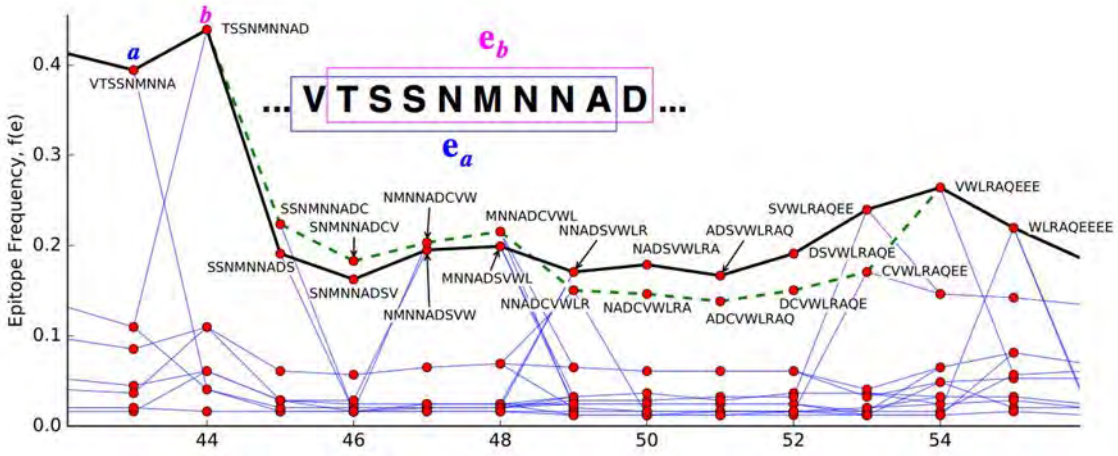
Method: evaluation of acyclic graph

Advantages over mosaic

Essentially optimal (fractionally better coverage)

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



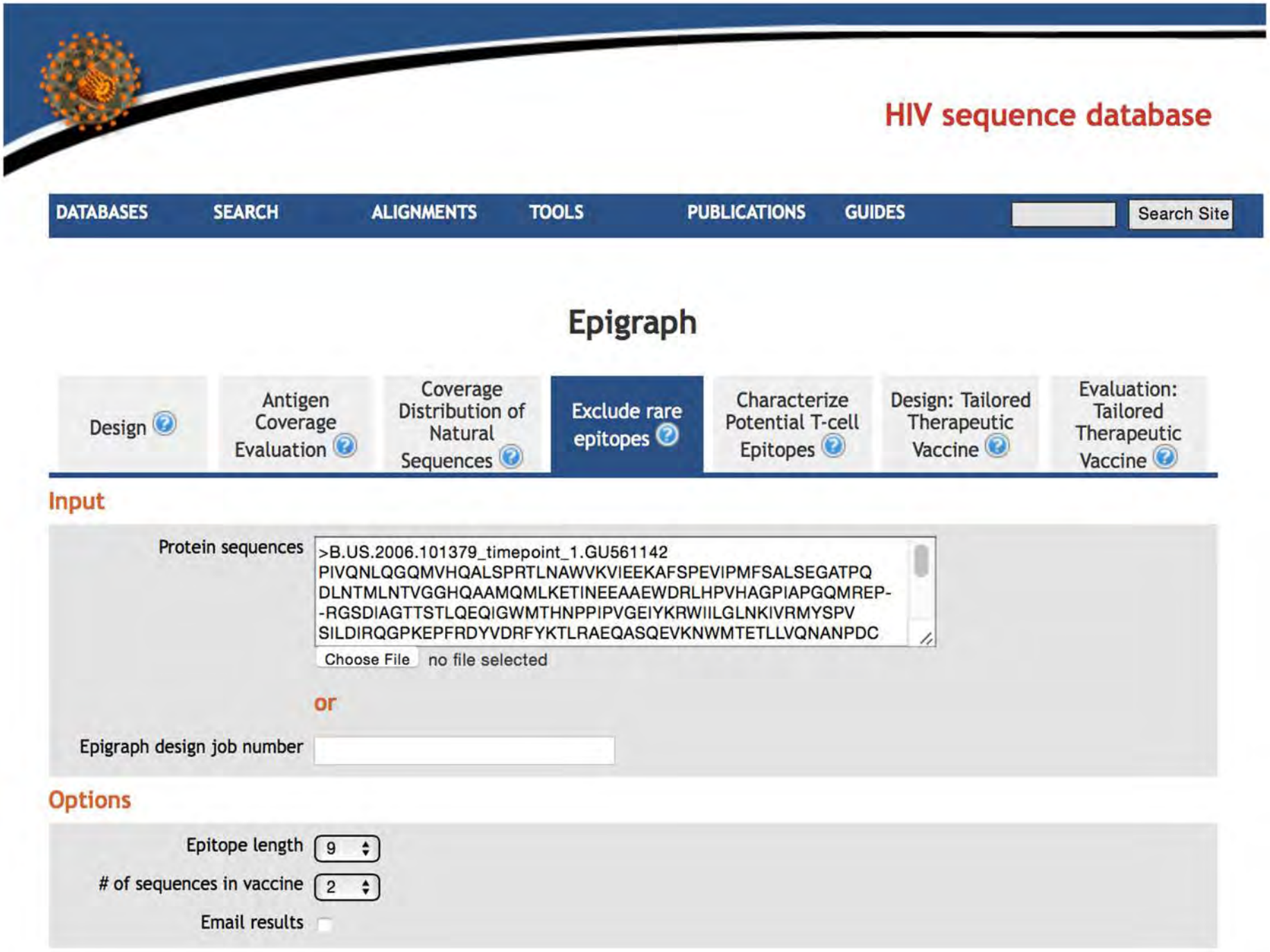


Reference: Theiler, J., Yoon, H., Yusim, K., Picker, L. J., Fruh, K., and Korber, B. (2016). Epigraph: A vaccine design tool applied to an HIV therapeutic vaccine and a pan-filovirus vaccine. *Sci Rep*, 6:33987.

Los Alamos NATIONAL LABORATORY

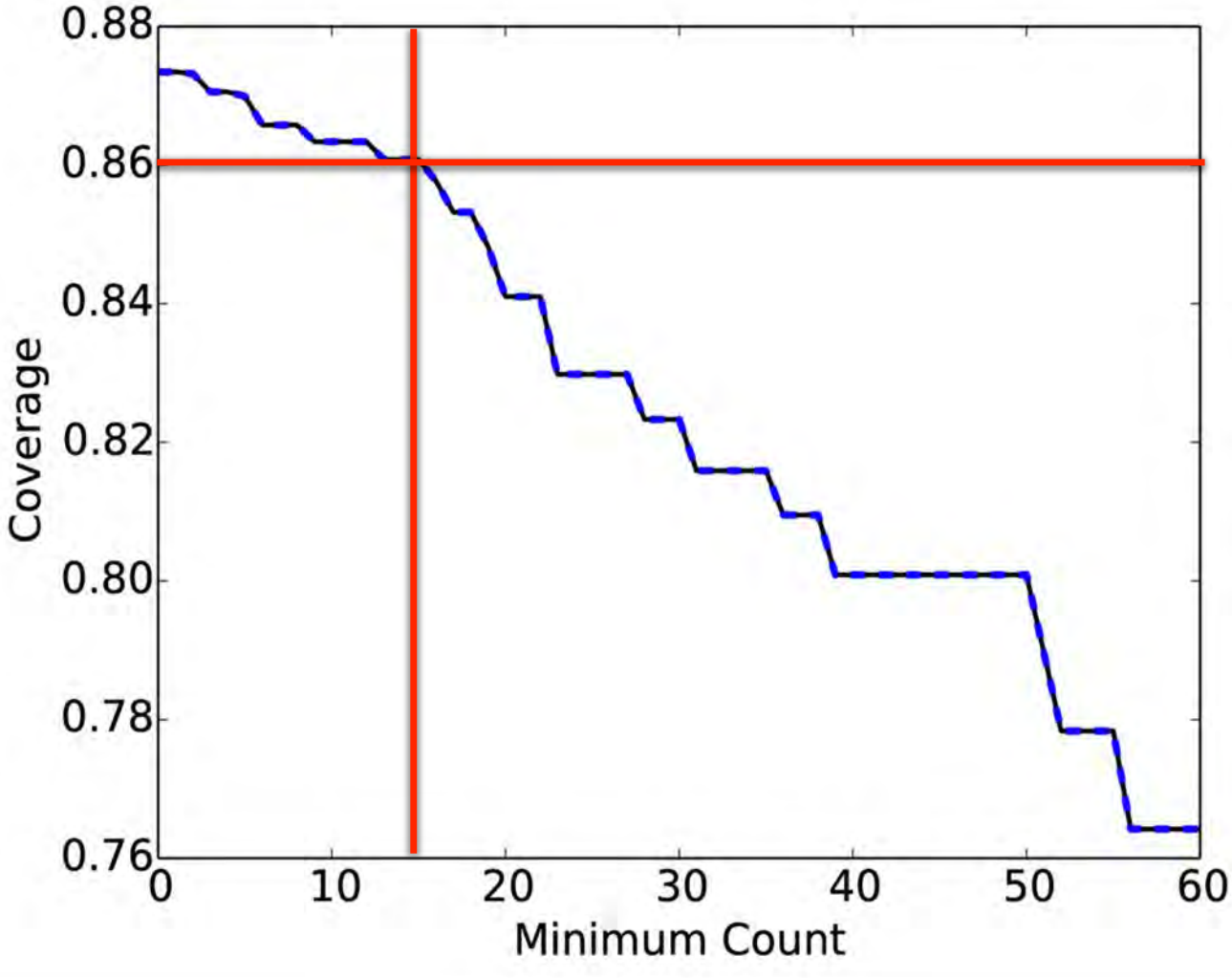
https://www.hiv.lanl.gov/content/sequence/EPIGRAPH/epigraph.html

EPIGRAPH — exclude rarities



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.

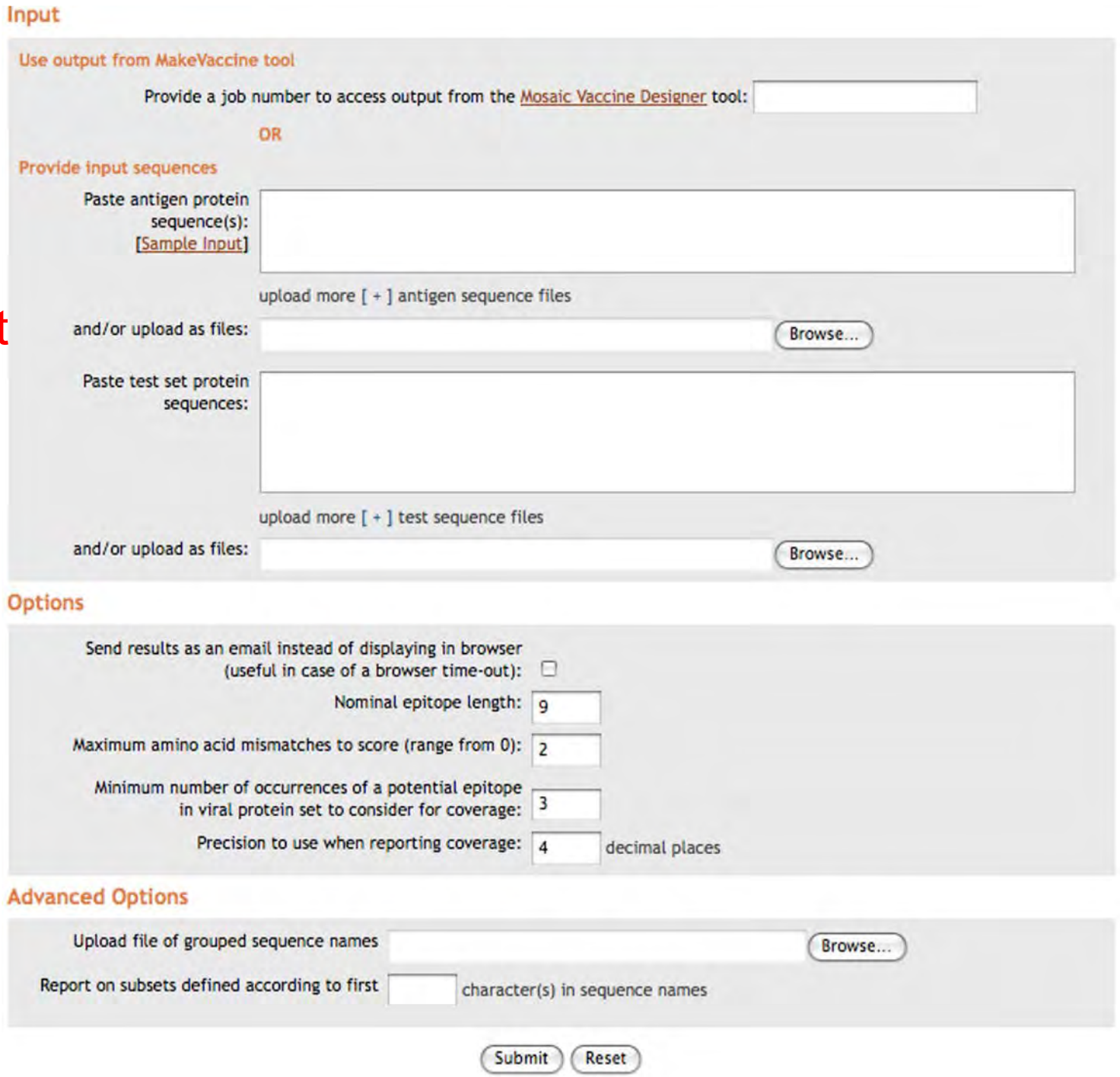
Epitope Coverage Assessment - Epicover

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



Los Alamos NATIONAL LABORATORY

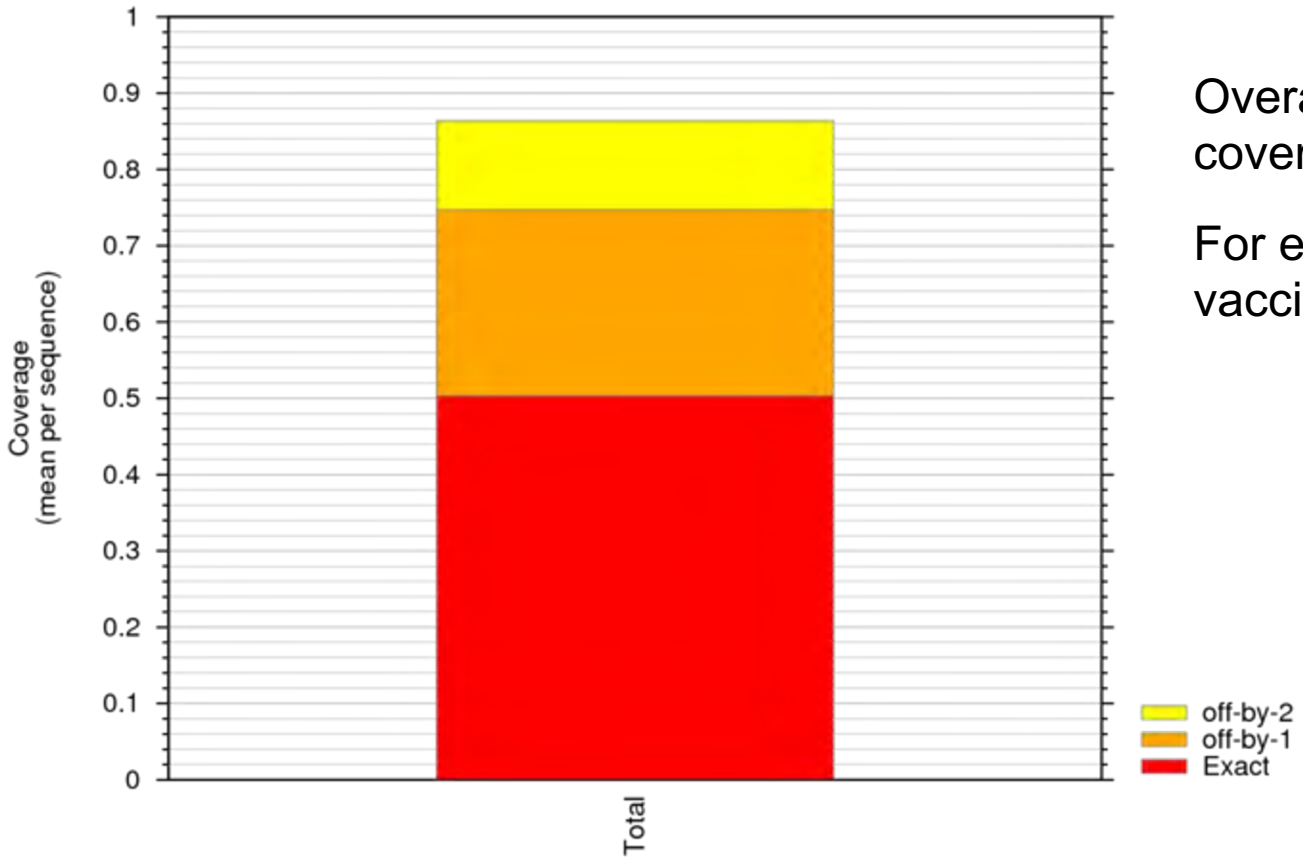
Epicover output (mean coverage per sequence)

vaccine set	subset	subset count	Off-by-0	Off-by-1	Off-by-2	rare(<3,>1)	unique	absent
vaccine_set_from_user	**Total**	**39**	**0.5037**	**0.7474**	**0.8636**	**91**	**61**	**38**
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
vaccine_set_from_user	**Total**	**39**	**0.5037**	**0.7474**	**0.8636**	**91**	**61**	**38**
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
…	…	…	…	…	…	…	…	…



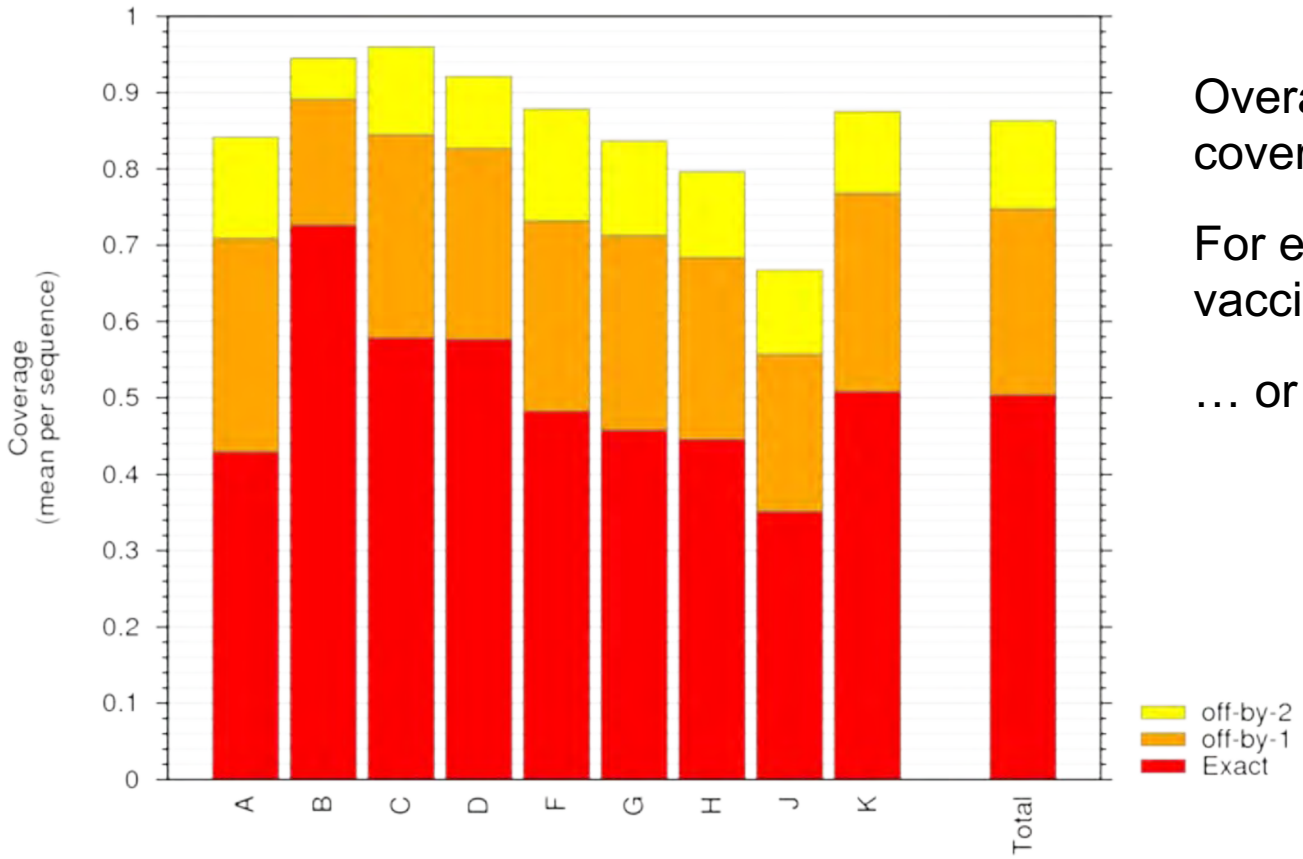
Overall summaries of *k*-mer coverage

For entire set (to compare with other vaccine candidates)

Los Alamos NATIONAL LABORATORY

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
vaccine_set_from_user	**Total**	**39**	**0.5037**	**0.7474**	**0.8636**	**91**	**61**	**38**
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
…	…	…	…	…	…	…	…	…



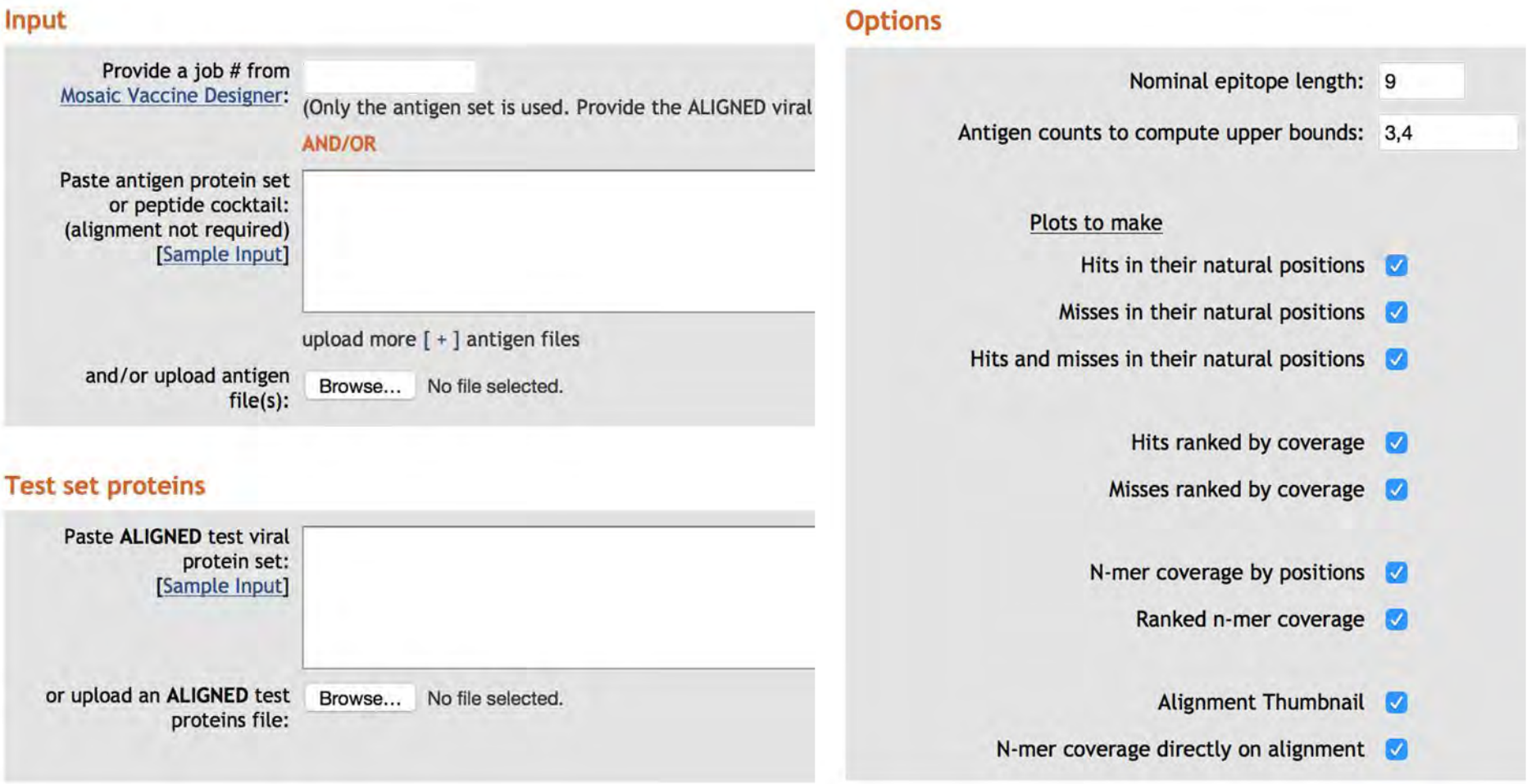
Overall summaries of *k*-mer coverage

For entire set (to compare with other vaccine candidates)

… or by pathogen subset

Los Alamos NATIONAL LABORATORY

Positional Epitope Coverage Assessment - Posicover



- INPUTS
  1. Vaccine/peptide sequences
  2. ALIGNED target set
- OUTPUTS

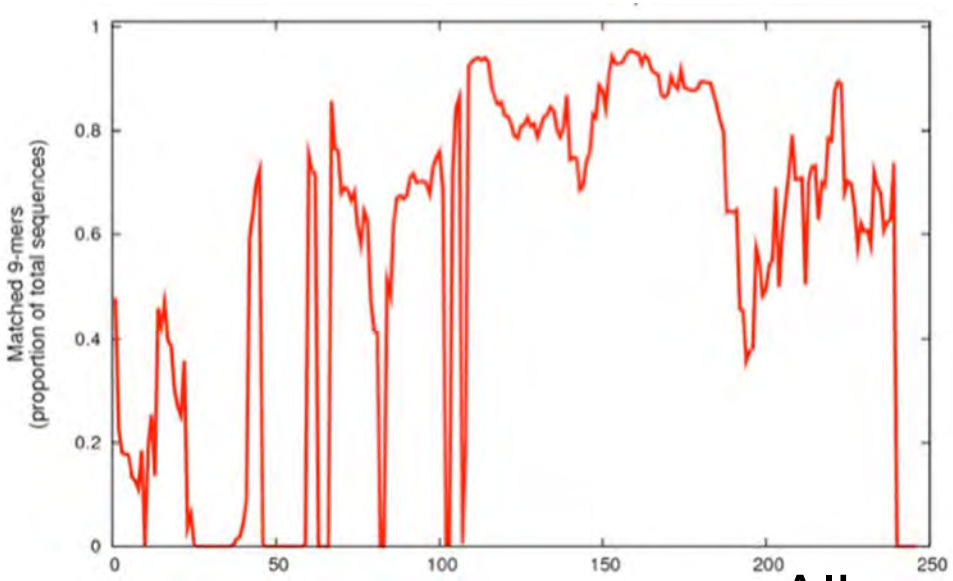
  1-dimensional (by alignment column)

  2-dimensional (by sequence and alignment column)

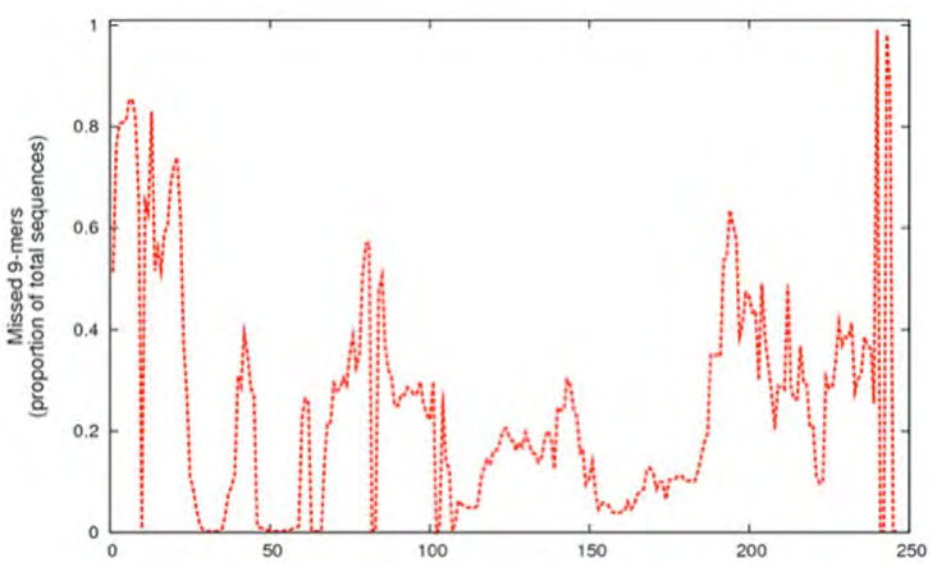
Los Alamos NATIONAL LABORATORY

Posicover output (1-dimensional summaries)

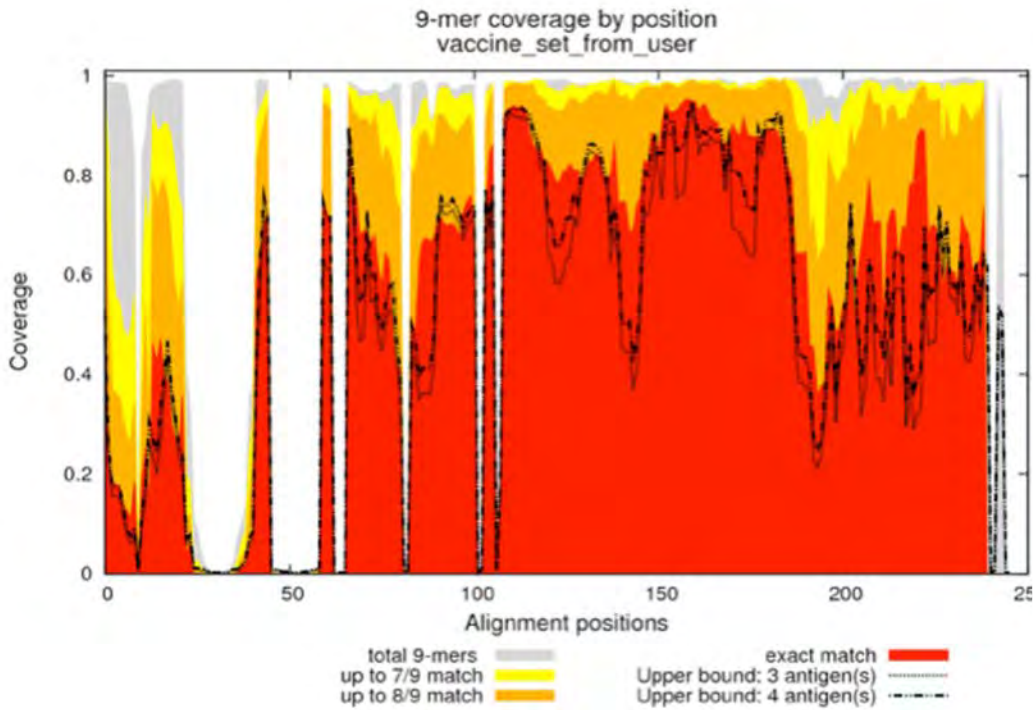
Matched 9-mers

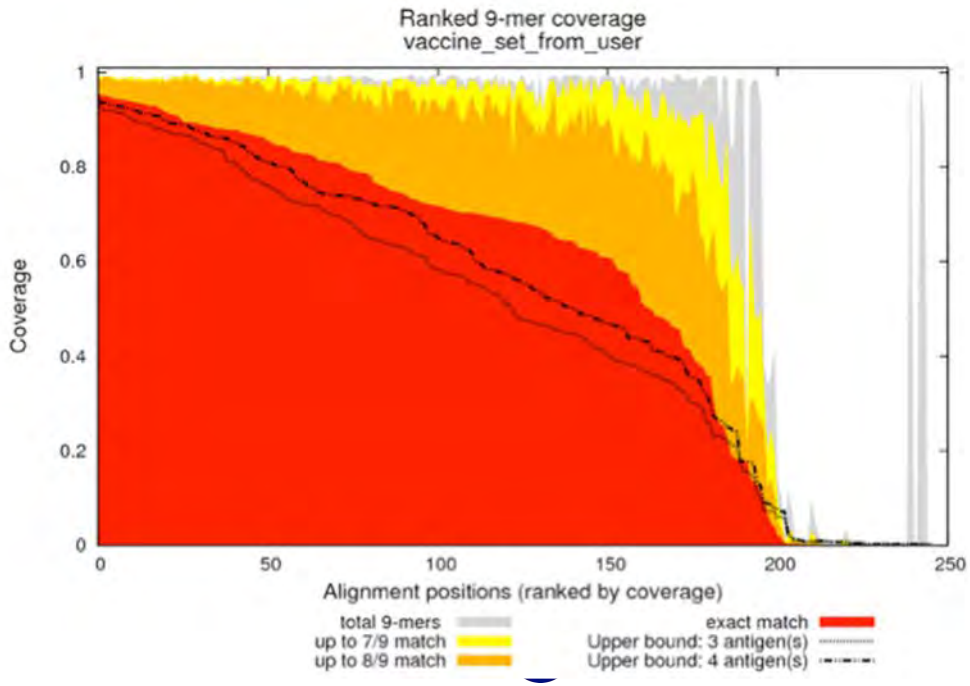


Missed 9-mers

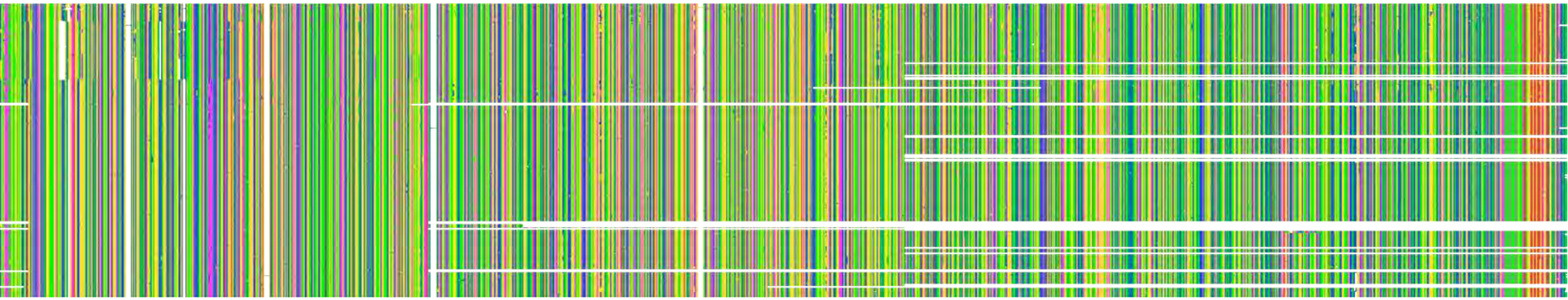


Alignment positions



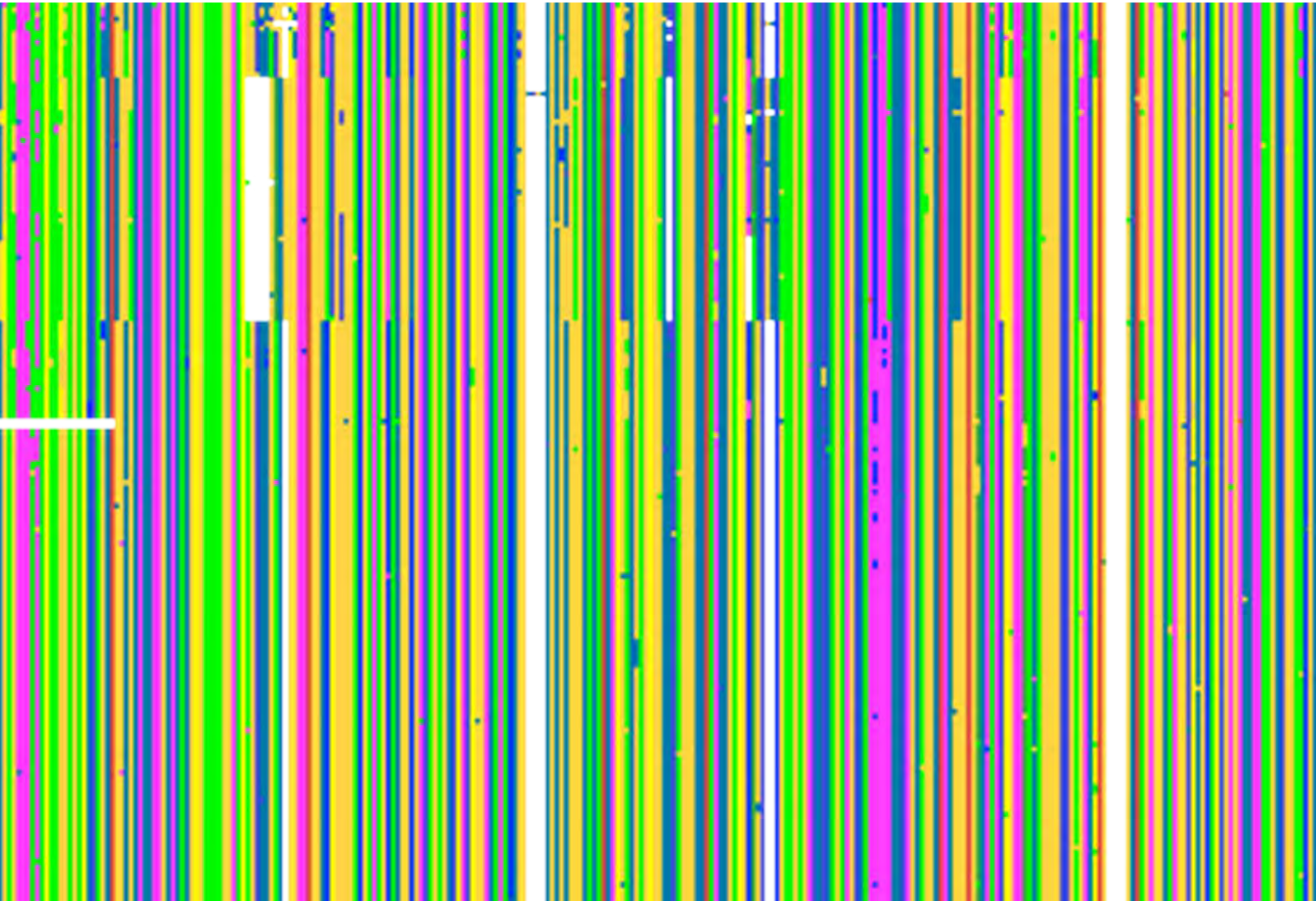


Posicover output (2 dimensional)

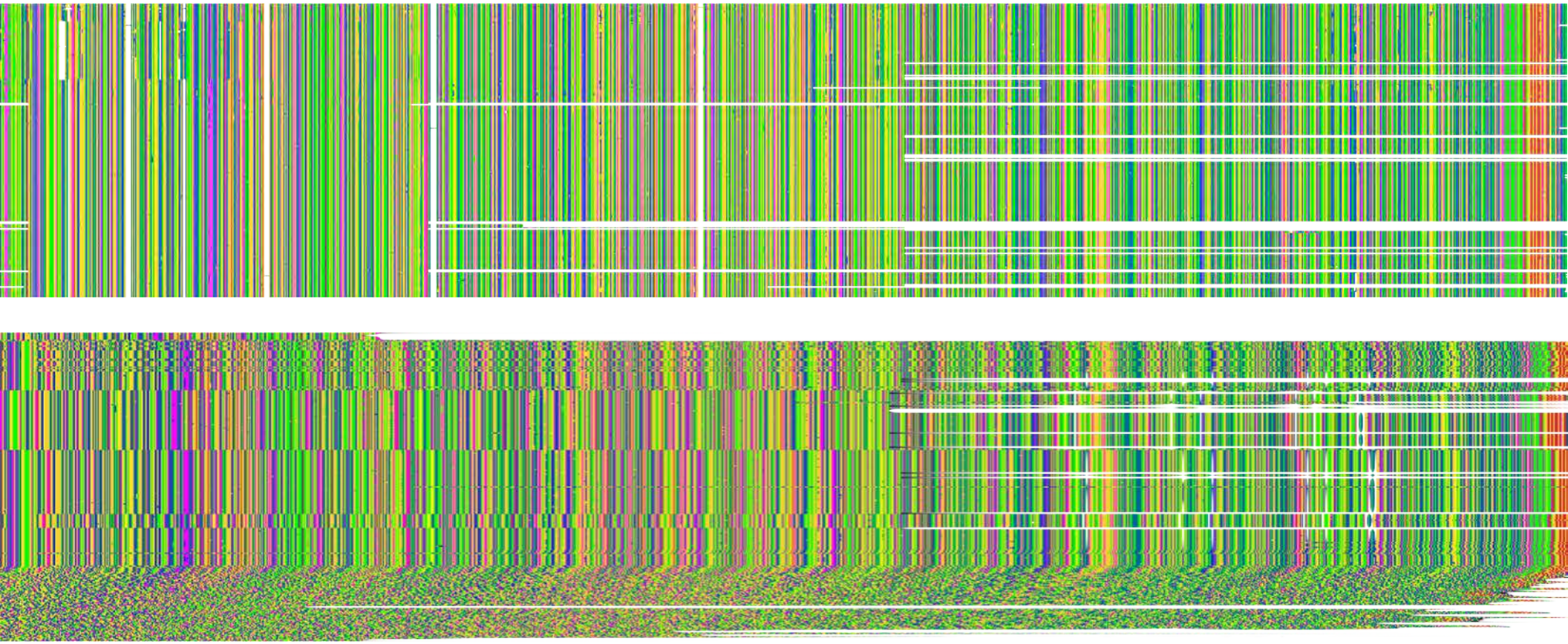


Pixel-based Alignment view
Each amino-acid represented as a single-colored square
Allows quick detection of gross errors in alignment

Posicover output (2 dimensional)

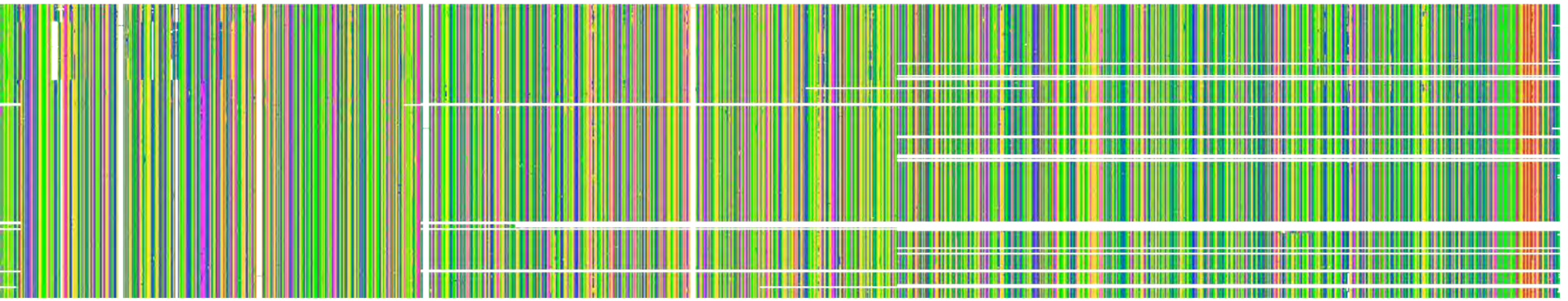


Posicover output (2 dimensional)

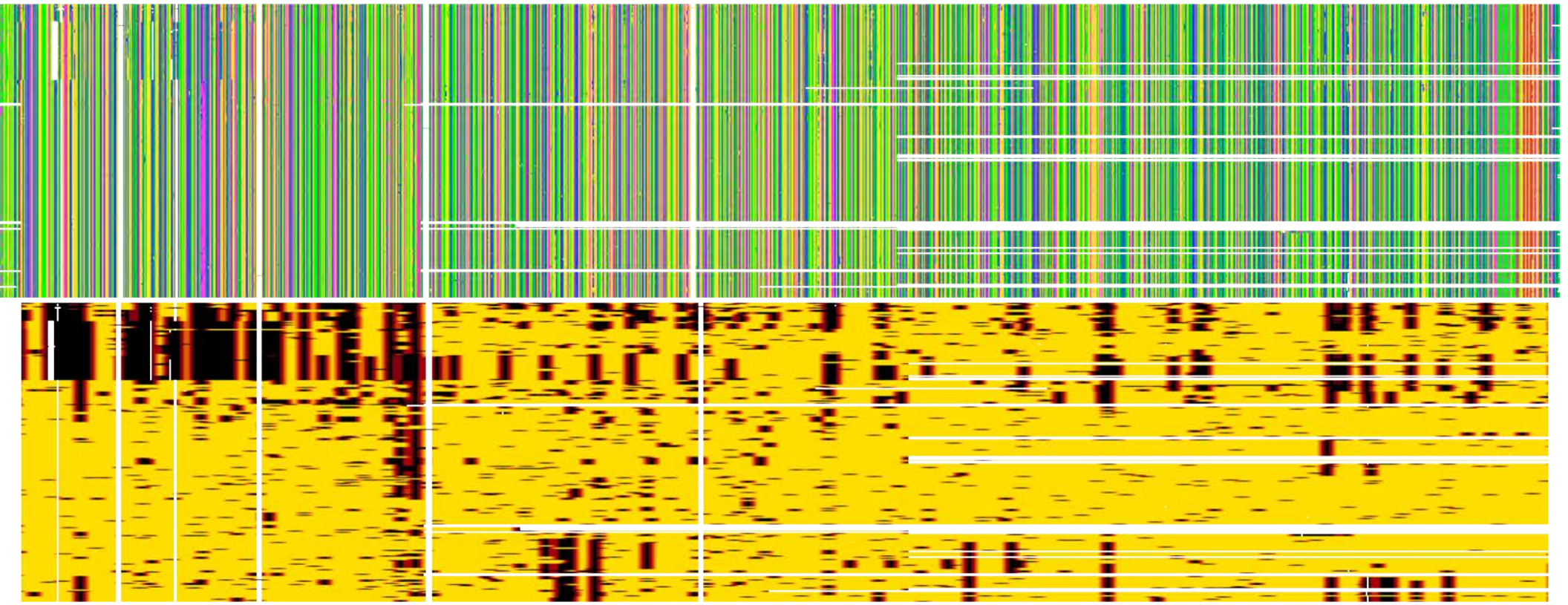


Pixel-based Alignment view
Each amino-acid represented as a single-colored square
Allows quick detection of gross errors in alignment

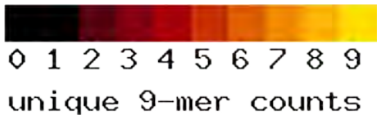
Posicover output (2 dimensional)



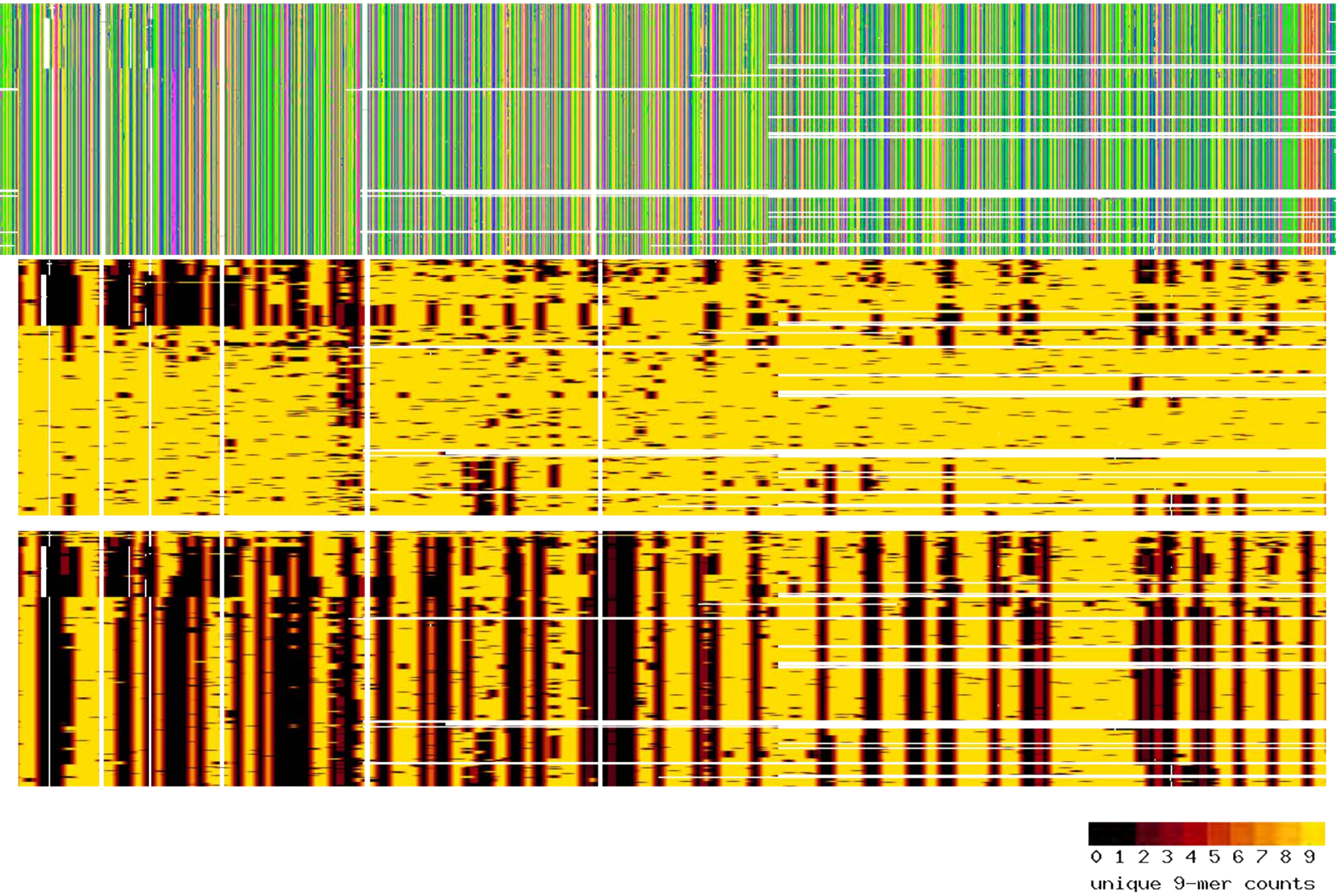
Posicover output (2 dimensional)



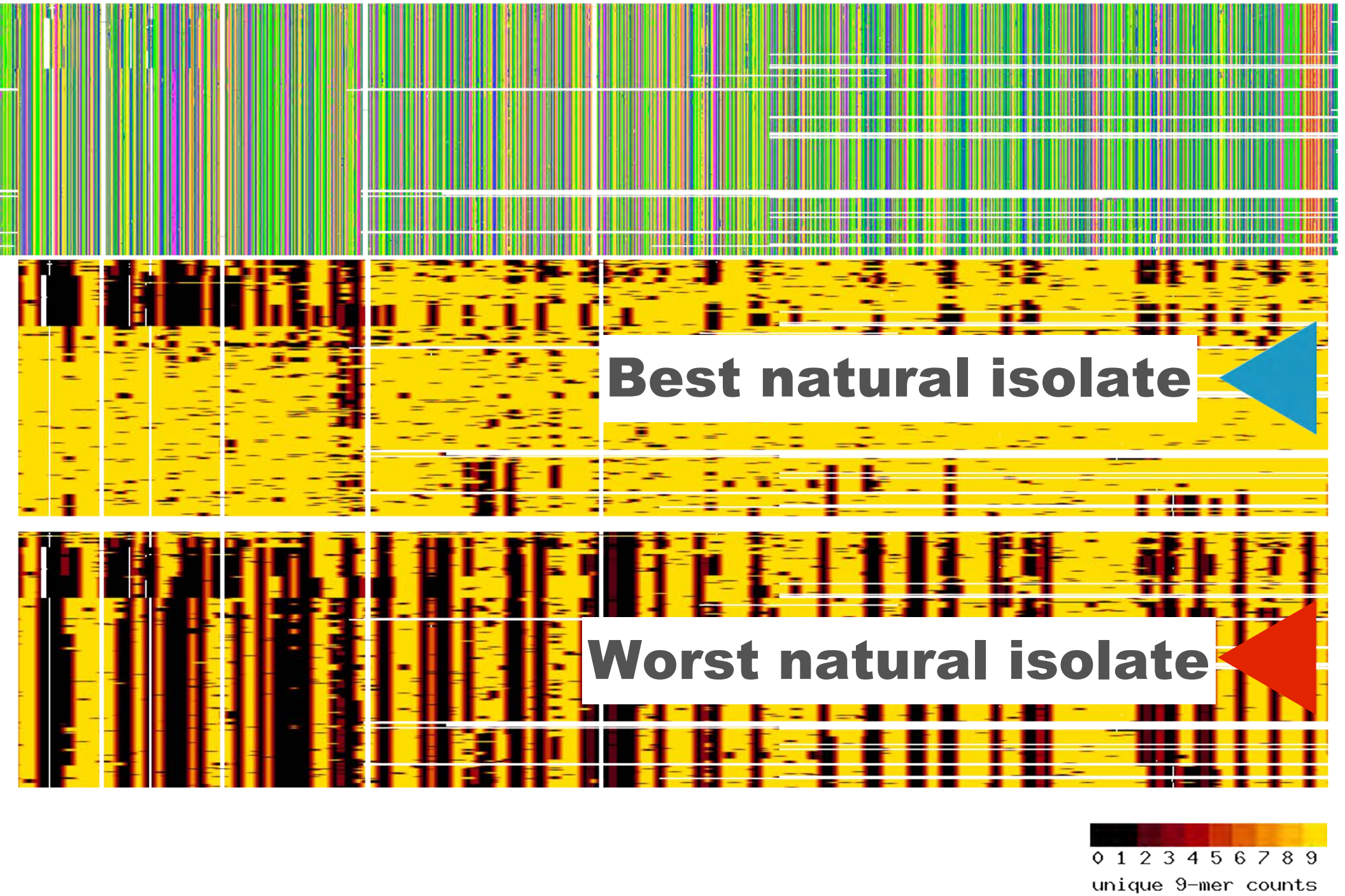
POSICOVER *K*-MER COVERAGE
(YELLOW-BLACK GRADIENT SHOWS HOW MANY OF EACH RESIDUE'S *K*-MERS APPEAR IN VACCINE)



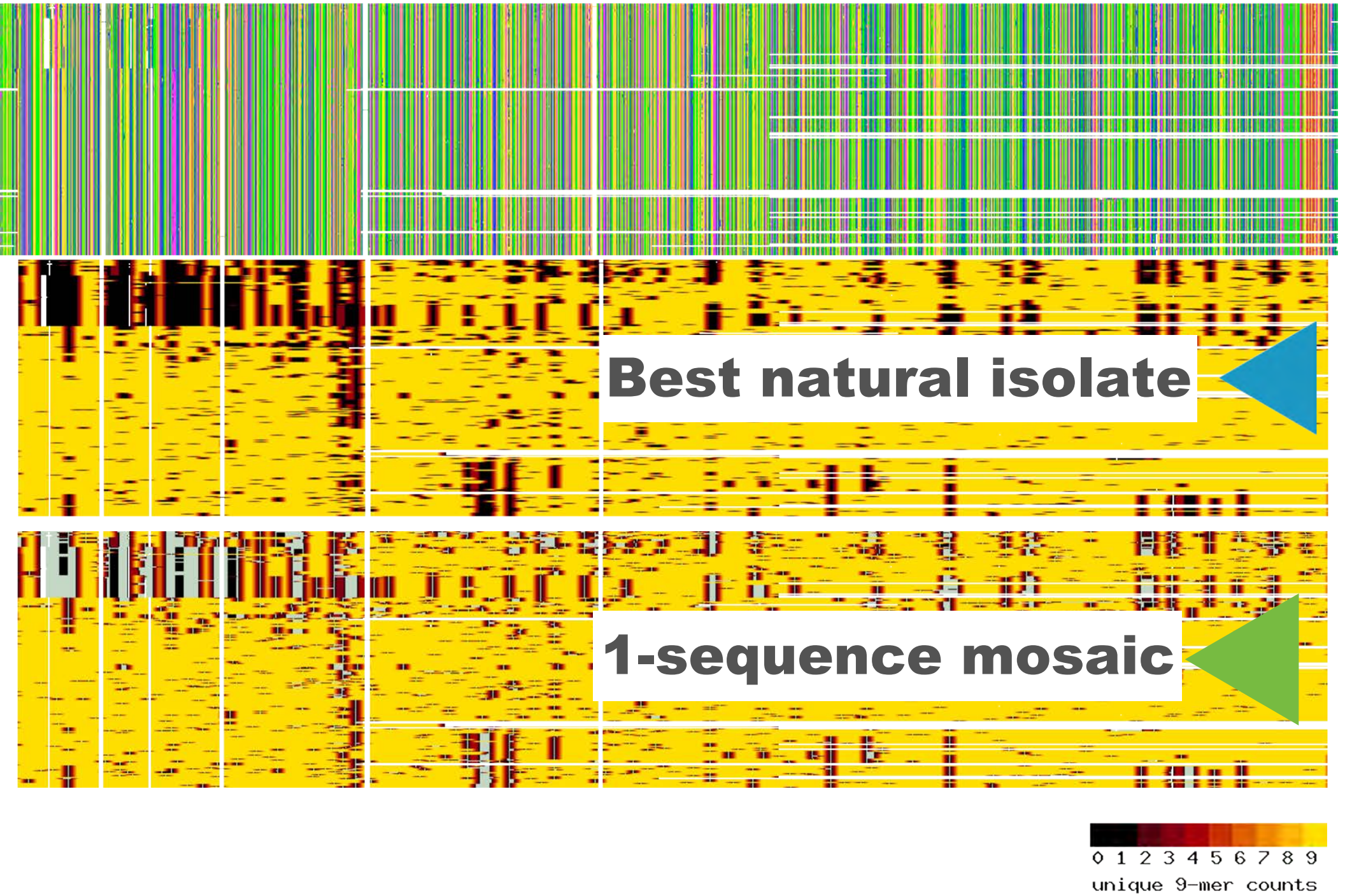
Posicover output (2 dimensional)



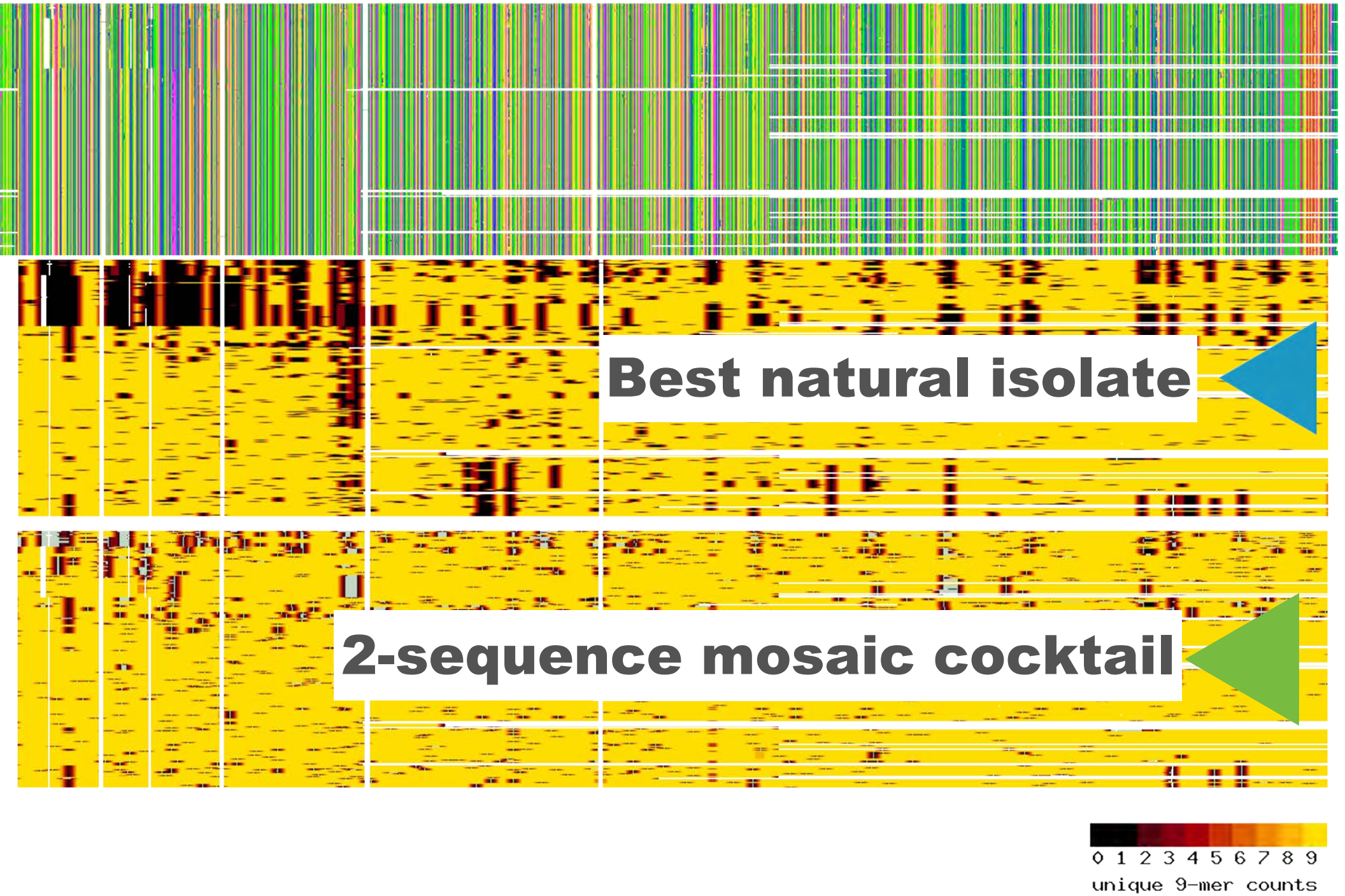
Posicover output (2 dimensional)



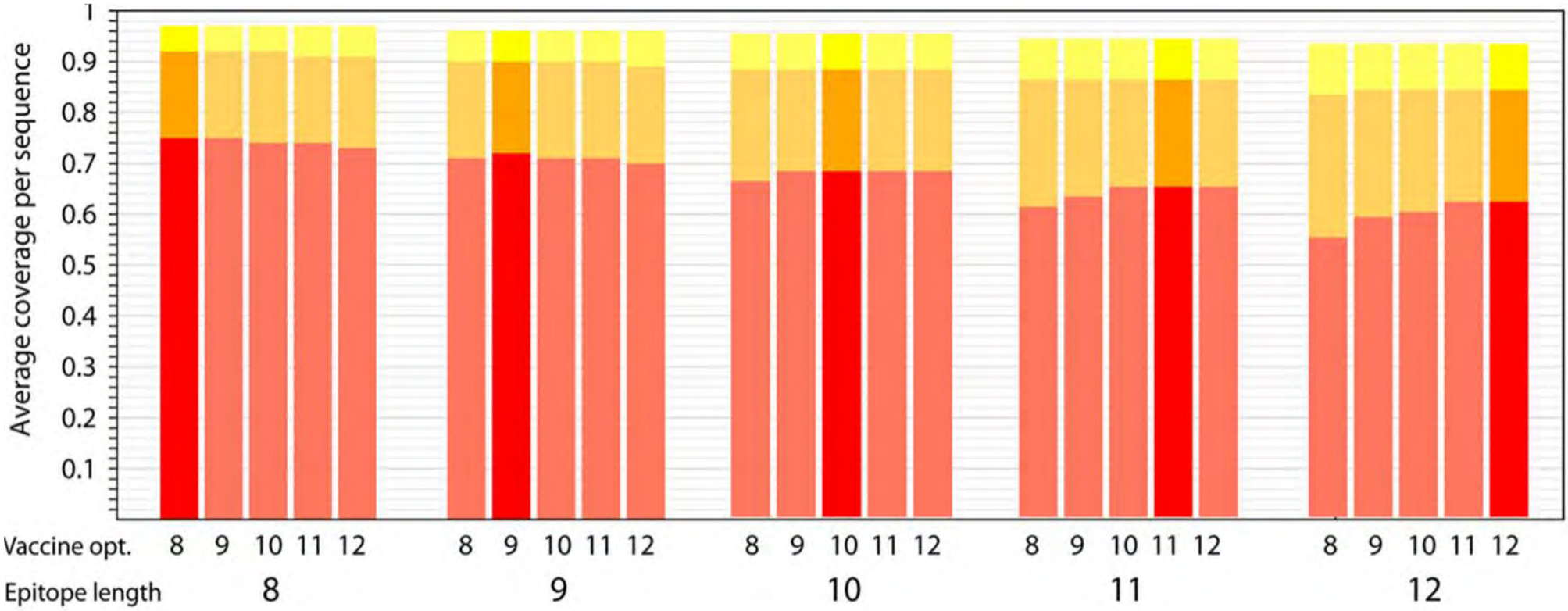
Posicover output (2 dimensional)



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.

Los Alamos NATIONAL LABORATORY

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 or seq-info@lanl.gov

Los Alamos NATIONAL LABORATORY