Immune Epitope Database NEWSLETTER

Volume 8, Issue 1 http://www.iedb.org February 2011

IEDB Version 2.5 Features a Capabilty to Cluster Results

The major new feature of IEDB 2.5, released at the end of October, is the ability to cluster query results by the sequence identity of the epitope structure. Clustering has been implemented to simplify query results, especially since the amount of data returned in many queries has grown as the amount of data in the IEDB has grown. As an example, if the user would query for all Vaccinia virus epitopes they would get a list of over 10,000 records. If they then hit the 'Cluster' button, the epitopes would be grouped according to their sequence identity and the user would end up with a 'cleaned up' view of the results.

An algorithm for grouping the sequences has been developed and tested with several IEDB data sets. It groups related sequences based on sequence identity and allows the user to specify the sequence identity threshold from 70%-100%. As an example, at 85% identity, 7,512 Vaccinia virus epitopes form 2,528 clusters. This drastically reduces the number of results displayed to the user and should result in more efficient navigation. In terms of sorting the clusters of epitopes so that the most relevant ones are displayed first, bioinformaticians at LIAI have developed a metric called the 'evidence score' (EScore). This metric takes into account the number of references, number of positive assays, and total number of assays for each epitope. It is used initially to select the epitopes around which to form clusters. The sum of all of the evidence scores in a cluster is taken to be the 'cluster score' (CScore). Clusters are then sorted by the CScore in a descending order by default.

Figure 1 shows a sample results page with the Cluster feature highlighted. The initial results page is displayed without any clustering. The user can select from four levels of sequence identity for the clustering, 70, 80, 90, or 100%, as indicated with the boxed '2' in the figure. Clustering is initiated when the user clicks the *Cluster* button (indicated as '1' in the figure).

Continued on Page 2

Inside this Issue

Continued from Page 1

A description of the epitope clustering algorithm and additional screen shots of the clustered results can be found in the IEDB Solutions Center under the *Tutorials and Reference Materials* folder (http://iedb.zendesk.com/entries/306422-iedb-epitope-clustering-help). Help can also be reached by clicking on the '?' icon to the right of the *% Identity* button.

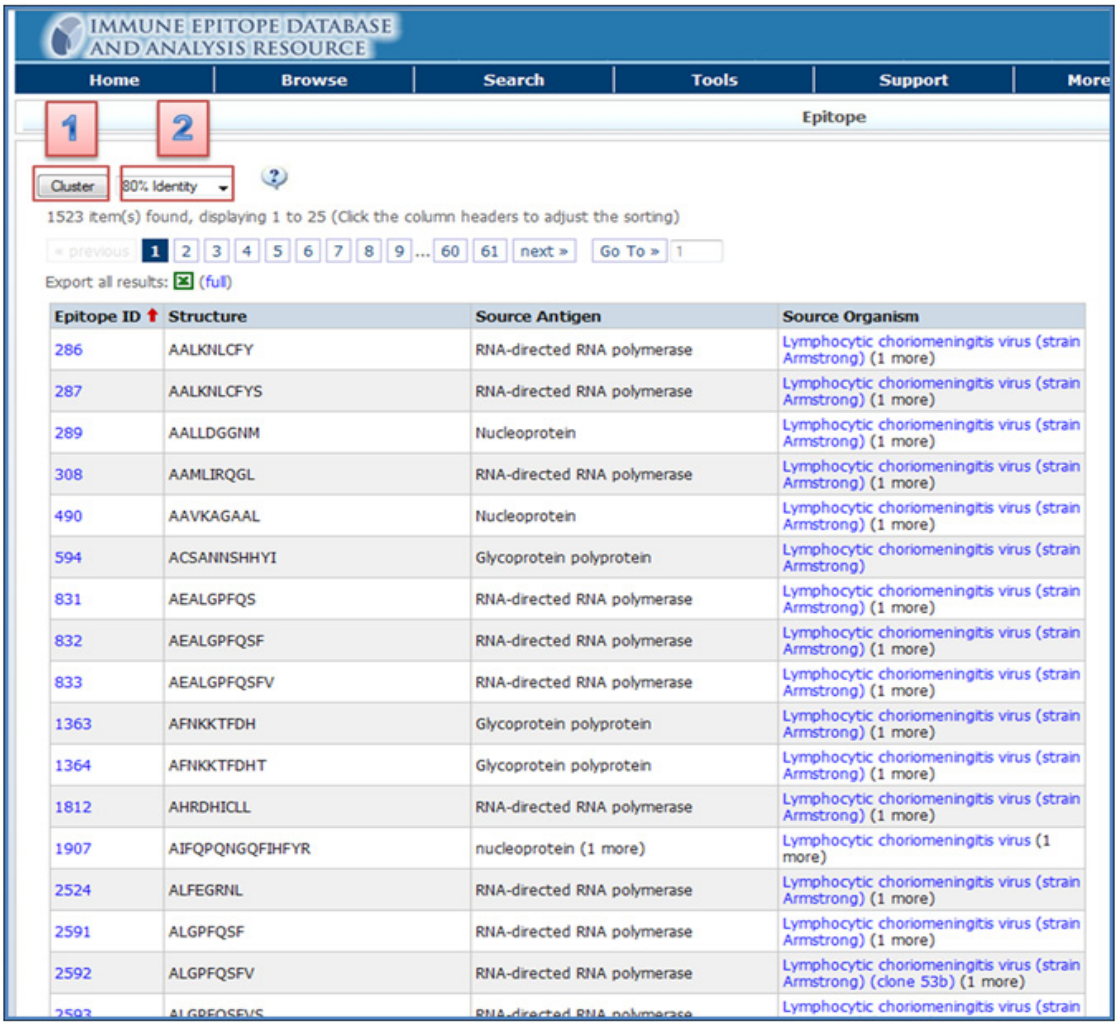

Figure 1. Sample results page with Cluster feature highlighted.

New HLA Nomenclature to be Implemented in the IEDB

The World Health Organization Nomenclature Committee for Factors of the HLA System has adopted changes to the HLA nomenclature system effective April 1, 2010. A detailed description of these changes can be viewed at http://hla.alleles.org/announcement.html. The IEDB will switch to the new nomenclature in all of the interfaces including query reporting, tools, and data submissions by end of March 2011. Users currently in the process of preparing a data submission to the IEDB that includes HLA restriction data should plan to complete submissions using the old nomenclature before March 2011 and the new nomenclature after March 1, 2011. The old names will be retained as synonyms. The nomenclature changes will also be implemented in the epitope prediction tools in the Analysis Resource.

New Version of the Analysis Resource Now Available

A new version of the IEDB Analysis Resource was released in November 2010. Version 2.4 includes major enhancements for MHC class I and II epitope prediction tools. These tools, developed by the Technical University of Denmark (DTU) under separate funding, have been incorporated in the Analysis Resource under the IEDB contract. The first tool, NetMHCpan, was integrated and made available as a selectable method in the MHC class I binding prediction tool. NetMHCpan predicts binding of peptides to a MHC class I molecule using artificial neural networks (ANN). It predicts binding for over 1,650 alleles, including HLA-A, B, C, E, G; non-human primates; mouse; pig; and user-supplied MHC sequence. Predictions can be made for peptide sequences of 8 to 11 residues in length. The method has been trained on over 110,000 peptide/MHC interactions. The second new method, NetMHCIIpan, has been integrated into the MHC class II binding prediction tool as one of the user selections. NetMHCIIpan predicts binding of peptides for over 500 HLA-DR alleles using artificial neural networks.

In addition to the two new prediction methods, the updated Analysis Resource features new user interfaces for the MHC class I and class II binding prediction tools. Simultaneous prediction for multiple alleles is now possible, and interfaces have been reorganized to simplify use. Finally, the consensus processing predictor has been updated to reflect changes in the MHC class I prediction tool.

Six New IEDB "How-To" Videos Available

The series of "How-To" video tutorials available on the IEDB Solutions Center has grown by six since the last Newsletter. There are now a total of 11 videos. Five were added to the *Searching the IEDB* category and one was added to the *Understanding Query Results* category. Their titles are *Using the Organism Finder*, *Peptide/Protein Homology Search*, *Linear Sequence Queries*, *How do I search for human epitope data and specify geographic region?*, *How do I generate a list of all proteins from which epitopes have been identified for an organism of interest?*, and *Results: Epitope Listing, Details of Individual Entries and Using the Export Function*. They can be found in the *Knowledgebase and Forums* menu under *Tutorials and Reference Materials*. The short videos feature actual screen animations and voice-overs from IEDB curators. The collection of videos will grow and evolve as the interfaces and capabilities of the IEDB continually improve. They can be accessed directly at http://iedb.zendesk.com/entries/140865-how-to-videos.

Curation Update

Curation of data relating to peptidic epitopes for all infectious diseases and peptidic and non-peptidic epitopes for allergens is current for references appearing in PubMed as of the end of September 2010, with many newer references recently added. A query for new potentially relevant epitope references is run quarterly to update the database. Curation of peptidic epitopes for all autoimmune diseases is essentially complete. Curation of non-peptidic epitopes for all autoimmune and infectious diseases is in progress and will be completed in early 2011. As of January 2011, data from approximately 11,700 references have been incorporated into the IEDB. The IEDB contains data for over 79,000 epitopes, 2,631 epitope source organisms, and 579 restricting MHC alleles. Users are invited to bring references to our attention that are potentially relevant to the IEDB but do not appear in the database. References that are deemed to meet the IEDB criteria for curation will be queued for processing in accordance to our NIAID-directed priorities (Category A-C priority pathogens, emerging and re-emerging infectious diseases, other infectious diseases, allergies, autoimmune diseases, and transplantation). The IEDB does not curate cancer and HIV references. Citations should be sent to help@iedb.org.

Recent Publications

Peptide binding predictions for HLA DR, DP and DQ molecules.

Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B.

BMC Bioinformatics. 2010 Nov 22;11:568.

PMCID: PMC2998531; PMID: 21092157

BACKGROUND: MHC class II binding predictions are widely used to identify epitope candidates in infectious agents, allergens, cancer and autoantigens. The vast majority of prediction algorithms for human MHC class II to date have targeted HLA molecules encoded in the DR locus. This reflects a significant gap in knowledge as HLA DP and DQ molecules are presumably equally important, and have only been studied less because they are more difficult to handle experimentally.

RESULTS: In this study, we aimed to narrow this gap by providing a large scale dataset of over 17,000 HLA-peptide binding affinities for a set of 11 HLA DP and DQ alleles. We also expanded our dataset for HLA DR alleles resulting in a total of 40,000 MHC class II binding affinities covering 26 allelic variants. Utilizing this dataset, we generated prediction tools utilizing several machine learning algorithms and evaluated their performance.

CONCLUSION: We found that 1) prediction methodologies developed for HLA DR molecules perform equally well for DP or DQ molecules. 2) Prediction performances were significantly increased compared to previous reports due to the larger amounts of training data available. 3) The presence of homologous peptides between training and testing datasets should be avoided to give real-world estimates of prediction performance metrics, but the relative ranking of different predictors is largely unaffected by the presence of homologous peptides, and predictors intended for end-user applications should include all training data for maximum performance. 4) The recently developed NN-align prediction method significantly outperformed all other algorithms, including a naïve consensus based on all prediction methods. A new consensus method dropping the comparably weak ARB prediction method could outperform the NN-align method, but further research into how to best combine MHC class II binding predictions is required.

Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource.

Kim Y, Sette A, Peters B.

J Immunol Methods. 2010 Oct 31.

PMID: 21047510

The Immune Epitope Database and Analysis Resource (IEDB, http://www.iedb.org) hosts a continuously growing set of immune epitope data curated from the literature, as well as data submitted directly by experimental scientists. In addition, the IEDB hosts a collection of prediction tools for both MHC class I and II restricted T-cell epitopes that are regularly updated. In this review, we provide an overview of T-cell epitope data and prediction tools provided by the IEDB. We then illustrate effective use of these resources to support experimental studies.

Continued on Page 5

Continued from Page 4

We focus on two applications, namely identification of conserved epitopes in novel strains of a previously studied pathogen, and prediction of novel T-cell epitopes to facilitate vaccine design. We address common questions and concerns faced by users, and identify patterns of usage that have proven successful.

IEDB-3D: structural data within the immune epitope database.

Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, Bourne PE.

Nucleic Acids Res. 2011 Jan;39 (Database issue):D1164-70. Epub 2010 Oct 28.

PMID: 21030437 [PubMed - in process]

IEDB-3D is the 3D structural component of the Immune Epitope Database (IEDB) available via the 'Browse by 3D Structure' page at http://www.iedb.org. IEDB-3D catalogs B- and T-cell epitopes and Major Histocompatibility Complex (MHC) ligands for which 3D structures of complexes with antibodies, T-cell receptors or MHC molecules are available in the Protein Data Bank (PDB). Journal articles that are primary citations of PDB structures and that define immune epitopes are curated within IEDB as any other reference along with accompanying functional assays and immunologically relevant information. For each curated structure, IEDB-3D provides calculated data on intermolecular contacts and interface areas and includes an application, EpitopeViewer, to visualize the structures. IEDB-3D is fully embedded within IEDB, thus allowing structural data, both curated and calculated, and all accompanying information to be queried using multiple search interfaces. These include queries for epitopes recognized in different pathogens, eliciting different functional immune responses, and recognized by different components of the immune system. The query results can be downloaded in Microsoft Excel format, or the entire database, together with structural data both curated and calculated, can be downloaded in either XML or MySQL formats.

Contact Information

The Immune Epitope Database and Analysis Resource is supported by a contract from the National Institute of Allergy and Infectious Disease, NIH, DHHS (Contract HHSN266200400006C). The newsletter is distributed four times a year. We welcome communication from the users of the IEDB database and invite suggestions for articles in future issues. To subscribe to the IEDB newsletter or to contact project staff, send your email information to the email address below.

Principal Investigator:
Alessandro Sette, Ph.D.
alex@liai.org

Co-Principal Investigator:
Bjoern Peters, Ph.D.
bpeters@liai.org

Email: contact@iedb.org
Web: http://www.iedb.org

Immune Epitope Database and Analysis Resource
c/o La Jolla Institute for Allergy & Immunology
9420 Athena Circle
La Jolla, CA 92037
(858) 752-6500

Project Director:
Stephen Wilson, Ph.D.
swilson@liai.org

Production:
Emily Seymour
Ward Fleri, Ph.D.