Immune Epitope Database
NEWSLETTER

La Jolla Institute Awarded IEDB Renewal Contract

The La Jolla Institute for Allergy and Immunology and its subcontractor team of Science Applications International Corporation, Technical University of Denmark, and the San Diego Supercomputer Center at UC San Diego were pleased to learn on 3 November 2011 that they had been awarded the IEDB renewal contract from the National Institute for Allergy and Infectious Diseases. The contract period started 15 December 2011 and is expected to continue for seven years. The team will continue to maintain and improve the IEDB as a valuable resource to the research community. Efforts will include enhancing the usability of the website and tools, and expanding the user base.

IEDB v 2.8 Released in December 2011

Several noteworthy new features in the IEDB website were introduced in the latest release this past December. Two of the major new features are the Assay Finder on the advanced MHC Binding and MHC Ligand Elution Search pages, which can be accessed from the "Search" pull-down menu. Assays are now displayed in a tree structure, as seen in Figures 1 and 2. A PubMed export link was also added at the bottom of the Reference List pages. When a user performs a home page or advanced search and clicks on the Reference link at the bottom of the Search Result Summary page, the system displays a Reference List. With this new feature, clicking "PubMed Export" (Figure 3) generates a PubMed query with all the PubMed IDs of the references and opens a new page on the PubMed website.

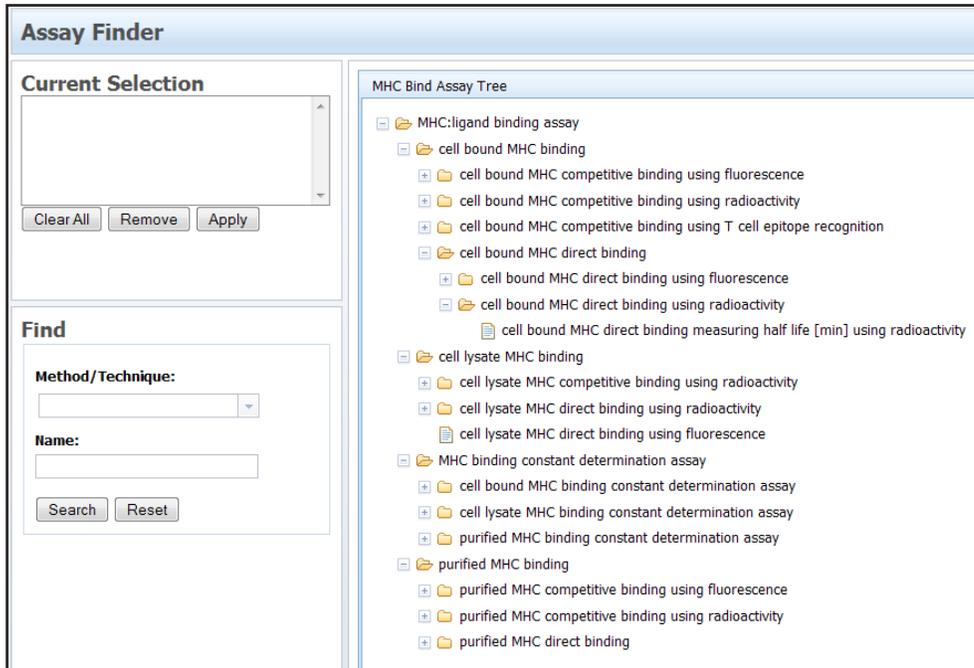### Inside this Issue

Continued from Page 1

Figure 1. The latest release of the IEDB website features a new Assay Finder for the advanced MHC Binding Search, located on the Search pull-down menu. Assays are displayed in a tree structure.
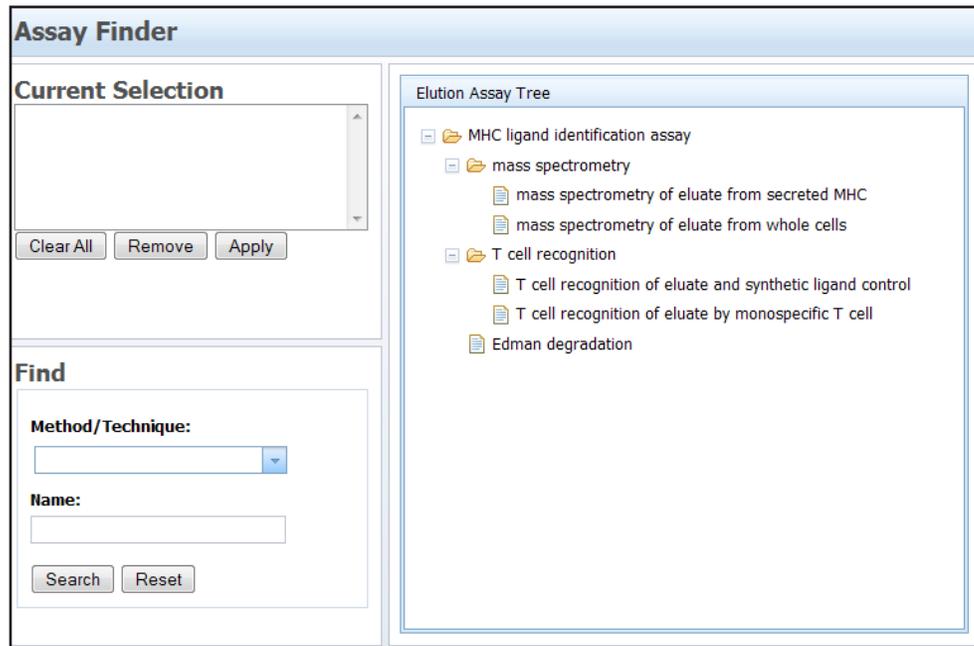

Figure 2 . The latest release of the IEDB website features a new Assay Finder for the advanced MHC Ligand Elution Search, located on the Search pull-down menu. Assays are displayed in a tree structure.

Figure 3. A new PubMed Export feature has been added to the Reference List page.

Finally, select compilations of data are now available for download from the Database Export web page (http://iedb.org/database_export.php) in Comma-Separated Value (CSV) format in the section titled CSV Metric Exports, as seen in Figure 4. Users can download these files to conveniently obtain all the information in the IEDB for specific categories. For example, reference_compact.zip and reference_full.zip contain the list of all references (literature and submissions) in the database. The compact version contains the reference ID, reference type, PubMed ID, author(s), title, abstract, journal, and year. The full version contains the same information plus author affiliations, journal volume, article pages, MeSH terms, and all other reference related information captured in the database. Files are available for all peptidic epitopes, non-peptidic epitopes, T cell assays, B cell assays, MHC binding assays, MHC ligand elution assays, and source organisms.

Figure 4. Comma-Separated Value files for all peptidic and non-peptidic epitopes, all four categories of assays, source organisms, and references are now available to download from the IEDB Database Export page (http://iedb.org/database_export.php).

Analysis Resource v 2.6 Released in January

A new version of the IEDB Analysis Resource was deployed in January. For the release, NetCTLpan version 1.1 was incorporated into the proteasomal cleavage prediction tool, allowing for prediction of over 3,500 MHC alleles. The other major new feature was the addition to the Analysis Resource landing page of a tab called 'IEDB-API' describing the RESTful (REpresentational State Transfer) interface to the MHC class I and II binding prediction tools. Several IEDB Analysis tools can now be accessed via the RESTful Web Services. This service is currently available for MHC-I and MHC-II binding predictions for T-cell epitopes. The service sends a POST request to the tools server and relies on user supplied parameters. Below are some examples for accessing the service using the 'curl' program. Any program, including a web browser, that sends POST requests to the server will work just as well.The full list of parameters and their values for MHC class I and II can be found on the 'IEDB-API' web page (http://tools.immuneepitope.org/main/html/iedb_api.html).

Example for Class-I binding prediction
To run with the default (IEDB recommended) method, a sequence and allele-length combination:

```
$ curl --data "sequence_text=ADFADSFADSFDF&allele=HLA-A*01:01&length=9&species=human"
http://tools.immuneepitope.org/analyze/cgi-bin/mhc_I_binding.py
```

Example for Class-II binding prediction
To run with the default (IEDB recommended) method, a sequence and an allele:

```
$ curl --data "sequence_text=ADFADSFADSFDASFAS&allele=HLA-DRB1*01:01" http://tools.immuneepitope.org/analyze/cgi-bin/mhc_II_binding.py
```

Curation Status

Curation of data relating to peptidic epitopes for all infectious diseases, allergens, and autoimmune diseases, and non-peptidic epitopes for allergens and autoimmune diseases is current for references appearing in PubMed. A query for new potentially relevant epitope references is run biweekly since all reference categories are in maintenance mode for both peptidic and non-peptidic epitopes. As of April 2012, data from approximately 13,600 literature references and direct data submissions have been incorporated into the IEDB. The IEDB contains data for over 92,000 epitopes, 2,912 epitope source organisms, and 611 restricting MHC alleles.

The IEDB at AAAAI 2012

The IEDB exhibit booth was present and staffed at the American Academy of Allergy, Asthma, and Immunology (AAAAI) Annual Meeting held March 2 – 6 in Orlando, Florida. The booth was staffed for three days by Drs. Randi Vita and Yohan Kim of LIAI (Figure 5). Approximately 100 people stopped to visit with them and learn more about the IEDB. The IEDB was well received, with visitors being quite impressed by the content of the database. Research areas of interest included food allergens, dust mites, parasites, pollens,

chemical exposures, autoimmune disorders, fungi, and mycoplasma. In addition, IEDB Principal Investigator Dr. Alessandro Sette presented a talk on "The Immune Epitope Database: Navigation and Applications" on the afternoon of March 5. His talk highlighted the allergen epitope content of the IEDB.

Figure 5. Senior Curator Randi Vita, MD and Bioinformatics postdoc Yohan Kim, PhD staff the IEDB exhibit booth at AAAAI 2012.

The IEDB at Upcoming Conferences

The IEDB exhibit booth will return to the Annual Meeting of the American Association of Immunologists this May 4 -8 in Boston. The exhibit booth will be staffed on all three days, Saturday through Monday, from 9:30 to 4:30. Two senior curators and a bioinformatician will be available to demonstrate the features of the website and the Analysis Resource, answer questions from users, and introduce the IEDB resource to potential new users. Nima Salimi will present on Sunday morning, May 6 at 11:00, a one hour vendor workshop on the IEDB that will cover the basic features of the database website, the available tools, and new features introduced over the past year. This will be another opportunity for meeting attendees to learn more about the IEDB and ask questions. Dr. Kerrie Vaughan will also present a poster titled "A meta-analysis of the existing knowledge of immunoreactivity against hepatitis C virus (HCV)."

The IEDB will be present at FOCIS 2012, the annual meeting of the Federation of Clinical Immunology Societies. The meeting will take place at the Vancouver Convention Centre in Vancouver, British Columbia, Canada, June 20 – 23. The exhibit hall will be open Thursday evening and Friday afternoon and

evening. Two IEDB senior curators will staff the booth. Special emphasis will be placed on the large quantity of autoimmune epitopes available in the database. In addition, Dr. Randi Vita will present a poster titled "The IEDB: Making Immune Epitope Data Freely Accessible".

The fourth meeting at which the IEDB booth is scheduled to be present for 2012 is the 20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB), which will be held July 15 – 17 at the Long Beach Convention Center in Long Beach, CA. This is one of the largest bioinformatics conferences and is an official conference of the International Society for Computational Biology. This will be the first time that the IEDB has exhibited at this meeting. It will be a wonderful opportunity to introduce a whole new community to the IEDB. The booth will be staffed by two senior curators and a bioinformatics postdoc for the three days of the conference. A 25-minute workshop on the IEDB will also be presented.

Recent Publications

Cost sensitive hierarchical document classification to triage PubMed abstracts for manual curation.

Seymour E, Damle R, Sette A, Peters B.

BMC Bioinformatics. 2011 Dec 19;12:482.

PMCID: PMC3314711; PMID: 22182279 [PubMed - in process]

ABSTRACT: The Immune Epitope Database (IEDB) project manually curates information from published journal articles that describe immune epitopes derived from a wide variety of organisms and associated with different diseases. In the past, abstracts of scientific articles were retrieved by broad keyword queries of PubMed, and were classified as relevant (curatable) or irrelevant (not curatable) to the scope of the database by a Naïve Bayes classifier. The curatable abstracts were subsequently manually classified into categories corresponding to different disease domains. Over the past four years, we have examined how to further improve this approach in order to enhance classification performance and to reduce the need for manual intervention.

RESULTS: Utilizing 89,884 abstracts classified by a domain expert as curatable or uncuratable, we found that a SVM classifier outperformed the previously used Naïve Bayes classifier for curatability predictions with an AUC of 0.899 and 0.854, respectively. Next, using a non-hierarchical and a hierarchical application of SVM classifiers trained on 22,833 curatable abstracts manually classified into three levels of disease specific categories we demonstrated that a hierarchical application of SVM classifiers outperformed non-hierarchical SVM classifiers for categorization. Finally, to optimize the hierarchical SVM classifiers' error profile for the curation process, cost sensitivity functions were developed to avoid serious misclassifications. We tested our design on a benchmark dataset of 1,388 references and achieved an overall category prediction accuracy of 94.4%, 93.9%, and 82.1% at the three levels of categorization, respectively.

CONCLUSIONS: A hierarchical application of SVM algorithms with cost sensitive output weighting enabled high quality reference classification with few serious misclassifications. This enabled us to significantly reduce the manual component of abstract categorization. Our findings are relevant to other databases that are developing their own document classifier schema and the datasets we make available provide large scale real-life benchmark sets for method developers.

Contact Information

The Immune Epitope Database and Analysis Resource is suported by a contract from the National Institute of Allergy and Infectious Disease, NIH, DHHS (Contract HHSN272201200010C). 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.

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

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

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

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

Production:
Emily Seymour
Ward Fleri, Ph.D.