

Immune Epitope Database NEWSLETTER

Volume 5, Issue 1

<http://www.immuneEPITOPE.org>

January 2008

Inside this Issue

- 1.) Curation Update
Up next in the queue
- 2.) The Norovirus
Norovirus Caliciviridae
- 3.) Recent Publications
*Authored/Co-authored
by IEDB team members*
- 4.) IEDB & Discovery Workshop
4th Annual Meeting
- 5.) Upcoming Events
*Conferences, Exhibits
& Workshops*
- 6.) International Biocurator Meeting
2nd Annual Meeting

Curation Update

Curation of data relating to NIAID Category A, B, and C priority pathogens, NIAID emerging and re-emerging infectious diseases (<http://www3.niaid.nih.gov/research/topics/emerging/list.htm>), Malaria, Hepatitis B, Clostridium tetani, Leishmania, and Candida albicans is current for articles appearing in PubMed as of the end of September 2007. These will be updated in January 2008 to cover newer published articles and pertinent references recently brought to our attention. The IEDB curation team has curated most of the relevant herpesvirus references and has started on allergen references and other infectious diseases. All reference categories will continue to be updated quarterly. Curation of autoimmune diseases will start later this year upon completion of the allergy category.

The Norovirus

genus: *Norovirus* family: *Caliciviridae*

Winter time is often a time where people fall ill. One of the culprits involved is the Norovirus. Noroviruses are a group of related, single-stranded RNA, that causes acute gastroenteritis in humans. This includes vomiting, diarrhea with abdominal cramps, nausea and sometimes a slight fever. These viruses are resistant to many environmental challenges such as freezing temperatures or those as high as 60°C (120°F). They can survive after being steamed in shellfish or in up to 10ppm of chlorine, which is well in excess of levels in public water systems. Despite these resistant characteristics, prevention is possible. Hand washing after using the bathroom and before handling food is important in preventing spread of this virus. Also, care in handling cold food and the use of paid sick leave will help to prevent the spread to others at the work place. To find more about this virus, please visit the Centers for Disease Control and Prevention website at: <http://www.cdc.gov/ncidod/dvrd/revb/gastro/norovirus-factsheet.htm>

Recent Publications

Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines.

Bui HH, Sidney J, Li W, Fusseder N, Sette A.
BMC Bioinformatics. 2007 Sep 26;8(1):361
PMID: 17897458

In an epitope-based vaccine setting, the use of conserved epitopes would be expected to provide broader protection across multiple strains, or even species, than epitopes derived from highly variable genome regions. Conversely, in a diagnostic and disease monitoring setting, epitopes that are specific to a given pathogen strain, for example, can be used to monitor responses to that particular infectious strain. In both cases, concrete information pertaining to the degree of conservancy of the epitope(s) considered is crucial. **RESULTS:** To assist in the selection of epitopes with the desired degree of conservation, we have developed a new tool to determine the variability of epitopes within a given set of protein sequences. The tool was implemented as a component of the Immune Epitope Database and Analysis Resources (IEDB), and is directly accessible at <http://tools.immuneepitope.org/tools/conservancy>. **CONCLUSIONS:** An epitope conservancy analysis tool was developed to analyze the variability or conservation of epitopes. The tool is user friendly, and is expected to aid in the design of epitope-based vaccines and diagnostics.

An analysis of the epitope knowledge related to Mycobacteria.

Blythe MJ, Zhang Q, Vaughan K, de Castro R Jr, Salimi N, Bui HH, Lewinsohn DM, Ernst JD, Peters B, Sette A.
Immunome Res. 2007 Dec 14;3(1):10
PMID: 18081934

Tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, remains a leading cause of infectious disease morbidity and mortality, and is responsible for more than 2 million deaths a year. Reports about extremely drug resistant (XDR) strains have further heightened the sense of urgency for the development of novel strategies to prevent and treat TB. Detailed knowledge of the epitopes recognized by immune responses can aid in vaccine and diagnostics development, and provides important tools for basic research. The analysis of epitope data corresponding to *M. tuberculosis* can also identify gaps in our knowledge, and suggest potential areas for further research and discovery. The Immune Epitope Database (IEDB) is compiled mainly from literature sources, and describes a broad array of source organisms, including *M. tuberculosis* and other Mycobacterial species. **Description:** A comprehensive analysis of IEDB data regarding the genus Mycobacteria was performed. The distribution of antibody/B cell and T cell epitopes was analyzed in terms of their associated recognition cell type effector function and chemical properties. The various species, strains and proteins which the epitope were derived, were also examined. Additional variables considered were the host in which the epitopes were defined, the specific TB disease state associated with epitope recognition, and the HLA associated with disease susceptibility and endemic regions were also scrutinized. Finally, based on these results, standardized reference datasets of mycobacterial epitopes were generated. **Conclusions:** All current TB-related epitope data was cataloged for the first time from the published literature. The resulting inventory of more than a thousand different epitopes should prove a useful tool for the broad scientific community. Knowledge gaps specific to TB epitope data were also identified. In summary, few non-peptidic or post-translationally modified epitopes have been defined. Most importantly epitopes have apparently been defined from only 7% of all ORFs, and the top 30 most frequently studied protein antigens

continue on page 4

Fourth Annual Immune Epitope Database & Discovery Workshop

November 14-15, 2007
La Jolla, California

The Fourth Annual Immune Epitope Database and Discovery Workshop was held November 14 and 15, 2007 at the Sheraton Hotel in La Jolla, California. The meeting provided an opportunity for the contractors of the Immune Epitope Database and Analysis Resource (IEDB) and the Large Scale Antibody and T Cell Epitope Discovery programs to present their project status and plans and to discuss common interests.

The two-day meeting started with a presentation of the status of the IEDB. In the past year, the IEDB has focused on adding epitope data from the literature and generated by the Epitope Discovery groups, increasing the user base and usability of the resource, enhancing the Analysis Resource with new and improved tools, and enhancing data consistence and quality. A significant quantity of data has been added to the IEDB in the past year from literature curation and data submissions. The second order analyses of tuberculosis¹, anthrax², and botulinum² conducted this year have been key to augmenting data quality and utility. Several enhancements have been made to the Analysis Resource, including addition of the antibody epitope prediction tools BepiPred and DiscoTope and the T cell epitope processing predictors NetChop and NetCTL. The 3D visualization and mapping capabilities have been expanded. The coming year will see the deployment of IEDB 2.0 and its many usability enhancements, the curation of allergy and most autoimmune epitope references, enhanced epitope prediction capabilities, MHC class II and antibody prediction tool evaluations, and additional second order analyses. Increasing the IEDB's usability and user base within the scientific community will continue to be a priority.

The IEDB presentation was followed by fourteen 40 minute presentations by the Large-Scale Antibody and T Cell Epitope Discovery contractor teams. The project title and presenters, in the order they presented, are listed in the Executive Summary that will be posted on the IEDB website.

The Fifth Annual Immune Epitope Database and Discovery Workshop will be held on March 10 and 11, 2009 in the Washington D.C. metro area. This will be the last workshop in the series since the five-year Large-Scale Epitope Discovery contracts will be ending in 2009. In

continue on page 5

1.] Blythe MJ, Zhang Q, Vaughan K, de Castro R Jr, Salimi N, Bui HH, Lewinsohn DM, Ernst JD, Peters B, Sette A. An analysis of the epitope knowledge related to Mycobacteria. *Immunome Res.* 2007 Dec 14;3(1):10

2.] Zarebski L, Vaughan K, Sidney J, Peters B, Grey H, Janda KD, Casadevall A, Sette A. Analysis of epitope information related to Bacillus anthracis and Clostridium botulinum. *Expert Rev. Vaccines* 7 (1) (in press)

Upcoming Events

Conferences, Exhibits, Workshops

3rd MASIR Conference
Measuring Antigen-Specific Immune
Responses
La Plagne, France
January 30-February 3, 2008

<http://www.masir.org>

Experimental Biology 2008
San Diego Convention Center
San Diego, California
April 5-9, 2008

<http://eb2008.org/>

**11th Annual Conference on
Vaccine Research**
Baltimore Marriott Waterfront Hotel
Baltimore, Maryland
May 5-7, 2008

<http://nfid.org/conferences/vaccine08/>

American Society for Microbiology
Boston Convention & Exhibit Center
Boston, Massachusetts
June 2-4, 2008

<http://gm.asm.org/>

contain 65% of the epitopes, leaving the majority of *M. tuberculosis* genome unexplored. A lack of information related to the specific strains from which epitopes are derived is also evident. Finally, the generation of reference lists of mycobacterial epitopes should also facilitate future vaccine and diagnostic research.

Integrating epitope data into the emerging web of biomedical knowledge resources.

Peters B, Sette A.
Nat Rev Immunol. 2007 Jun;7(6):485-90. Review.
PMID: 17479127

The recognition of immune epitopes is an important molecular mechanism of the vertebrate immune system to discriminate between self and non-self. Increasing amounts of data on immune epitopes are becoming available due to technological advances in epitope-mapping techniques and the availability of genomic information for pathogens. Organizing this data poses a challenge that is similar to the successful effort that was required to organize genomic data, which needed the establishment of centralized databases that complement the primary literature to make the data readily accessible and searchable by researchers. As described in this Innovation article, the Immune Epitope Database and Analysis Resource aims to achieve the same for the more complex and context-dependent information on immune epitopes, and to integrate this data with existing and emerging knowledge resources.

Meeting Report: NIH Workshop on the Tuberculosis Immune Epitope Database.

Ernst JD, Lewinsohn DM, Behar S, Blythe M, Schlesinger LS, Kornfeld H, Sette A.
Tuberculosis (Edinb). 2007 Dec 6
PMID: 18068490

The Immune Epitope Database (IEDB), an online resource available at <http://immuneepitope.org/>, contains data on T cell and B cells epitopes of multiple pathogens, including *M. tuberculosis*. A workshop held in June, 2007 reviewed the existing database, discussed the utility of reference sets of epitopes, and identified knowledge gaps pertaining to epitopes and immune responses in tuberculosis.

Automating document classification for the Immune Epitope Database.

Wang P, Morgan AA, Zhang Q, Sette A, Peters B.
BMC Bioinformatics. 2007 Jul 26;8:269.
PMID: 17655769

The Immune Epitope Database contains information on immune epitopes curated manually from the scientific literature. Like similar projects in other knowledge domains, significant effort is spent on identifying which articles are relevant for this purpose. **RESULTS:** We here report our experience in automating this process using Naïve Bayes classifiers trained on 20,910 abstracts classified by domain experts. Improvements on the basic classifier performance were made by a) utilizing information stored in PubMed beyond the abstract itself b) applying standard feature selection criteria and c) extracting domain specific feature patterns that e.g. identify peptides sequences. We have implemented the classifier into the curation process determining if abstracts are clearly relevant, clearly irrelevant, or if no certain classification can be made, in which case the abstracts are manually classified. Testing this classification scheme on an independent dataset, we achieve 95% sensitivity and specificity in the 51.1% of abstracts that were automatically classified. **CONCLUSION:** By implementing text classification, we have sped up the reference selection process without sacrificing sensitivity or specificity of the human expert classification. This study provides both practical recommendations for users of text classification tools, as well as a large dataset which can serve as a benchmark for tool developers.

2nd International Biocuration Meeting San Jose, California

October 25-28, 2007



Photo of the meeting attendees at the Dolce Hayes Mansion. IEDB's own, Debbie Shackelford and Leora Zalman are up front and center.

The 2nd International Biocurator Meeting convened 125 or more curators and developers representing 70 of the over 900 currently active biological databases. The target areas of the databases are diverse and include genome and protein sequence annotation, functional genomics, model organisms, protein structure, and cell signaling pathways. The IEDB was represented by Deborah Shackelford and Leora Zalman, who presented a poster entitled "The Immune Epitope Database and Analysis Resource: A searchable open access repository of infectious and noninfectious disease-related B and T cell epitope data". Now that most of the databases represented have established their basic practices of curation and review and are available online, discussions focused on how best to maintain consistency of curation within a team, develop ontology, and how to involve the scientific community in providing critical information and direct curation of their own research reports. A lively workshop discussed how to implement collaborations between journal publishers and database curators to expedite future curation of the literature. The biennial meeting provides one of the few venues for database scientists to discuss their work, initiate collaborations, and foster a sense of community in this diverse and growing area of research.

continued from page 3

September 2007, the NIAID Advisory Council approved two initiative concepts related to epitope discovery. These concepts are available at <http://www.niaid.nih.gov/ncn/budget/concepts/c-ait0907.htm>

Contact Information

The Immune Epitope Database & Analysis Resource is supported by a contract from the National Institute of Allergy & 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.

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

Email: contact@immuneEPITOPE.org
Web: <http://www.immuneEPITOPE.org>

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

Acknowledgements:

Deborah Shackelford, Ph.D.
Leora Zalman, Ph.D.

Production:

Jody Dang
Ward Fleri, Ph.D.