IEDB NEWSLETTER

Volume 1, Issue 1 http://www.immuneepitope.org July 2004

Message from the IEDB Team

Introducing the Immune Epitope Database

Stephen Wilson, PhD
Project Director and Deputy-PI

In 2002, the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID) invited the scientific community to respond to a request for building an Immune Epitope Database and Analysis Resource (IEDB). Following a highly competitive process the La Jolla Institute for Allergy and Immunology (LIAI) was awarded the contract on December 15, 2003. The IEDB is intended to serve as the national repository of immune epitope data, facilitating drug, vaccine and immunological response research and discovery, and is scheduled to be first accessible to the public in the fall of 2005.

It is the goal of our team to develop a database that represents both the intrinsic properties of epitopes (eg. MHC binding, antibody affinity, and chemical properties) and extrinsic properties (eg. immunogenicity and antigenicity). A second major component of the IEDB is to host an Analysis Resource, which is essentially a library of data analysis tools. Together with the data in the database, we envision that the scientific community will use the IEDB to search for and access annotated information, and then run sophisticated tools to analyze the data.

continued on page 2

Inside This Issue

Some of the IEDB Team Members at a June 2004 conference. Left to Right; Muthu, Huynh, Steve, Chris, Alex, Tom, Scott, Scott, Rob, Gary, Norma, Jody, Holly, Bjoern.

IEDB in the News

San Diego gains points as biomedical research hub

By Bruce Lieberman
Union Tribune Staff Writer
May 3, 2004

La Jolla will be home to the nation's largest database exploring how the human body's immune system fights disease-causing agents – those occurring naturally and others possibly used as terrorist weapons.

The La Jolla Institute for Allergy and Immunology has received $25 million from the National Institutes of Health to head the 7-year project, scientists at the immunology group said.

By gathering in one place research into SARS, West Nile virus, smallpox, anthrax and other infectious agents, scientists hope to accelerate the development of new and better vaccines. A prototype of the database should be available to scientists around the country within a year.

"Many groups studying one disease or another will gather all their data and submit it to the database," said Alessandro Sette, the institute's lead scientist on the project.

"If many people can discover vaccine targets . . . but they can't share and do discovery work together, that would be a big loss," said Stephen Wilson, an immunologist at the institute and an investigator

continued on page 3

Analysis Resource

The IEDB will be home to a comprehensive set of analysis tools, including

- Epitope binding prediction methods consisting of procedures and algorithms for prediction of MHC binding such as binary ma anchor motif scans, linear coefficient matrices, neural networks and other machine learning prediction methods
- Epitope identification methods including algorithms and other prediction tools for identifying novel antibody and T cell epitopes from genome and protein sequence information; standard methods such as hydrophobicity analysis, solvent accessibility, and thermal mobility tools will also be provided.
- Methods to predict the degree of population coverage afforded by a given epitope or epitope set in different ethnicities, and to assess the degree of conservancy of a given epitope
- The issue of variability (or conservation) of epitopes will be addressed through various bioinformatics tools – these tools will assist in the selection of epitopes with the desired pattern of conservation

The IEDB will host new tools, which will be developed by team members with area expertise, as well as existing tools that will be nominated by the scientific community. All tools will undergo an evaluation, see tool evaluation strategy below.

To recommend a tool to the IEDB, please submit an email to tools@immuneepitope.org

continued from page 1

Finally, through public forums, newsletters, and feedback as a result of data submission directly from the community at large, we hope to engage research professionals across disciplines to help us improve the IEDB and facilitate new research collaborations among users.

In the fall of 2005, the first release of the functional database website will become available at http://www.immuneepitope.org/. Prior to the launch of the production system, we'll use this newsletter and the website as a means to inform the community of our progress. The website will always be the best place for users to provide feedback to the team and NIH about the database and the analysis resource.

In addition to IEDB personnel under contract, the project is guided by an IEDB Working Group (WG). The WG is composed of experts in bioinformatics and immunology who are appointed by the NIAID Program Officer for two-year terms. The WG meets quarterly to review progress and provide critical feedback to program officials at the NIAID.

We recognize that the richness and utility of the IEDB will be a function of how well our team forms a working collaboration with the scientific community. It is with great excitement that we inaugurate the database, and invite you to provide feedback to us at anytime.

Bioinformatic Seminars At La Jolla Institute for Allergy and Immunology

July 16th — 1:00 – 2:00

Immunoinformatics for the study of immune function

Vladimir Brusic, Ph.D., MAppSci, MBA, MEng
Research Manager & Head, Knowledge Discovery Department ;Institute for Infocomm Research (I2R)
Adjunct Associate Professor at John Hopkins Singapore; The University of Canberra, National University of Singapore; Nanyang Technological University, Singapore

September 20th — 2:00 – 3:00

Charles DeLisi, Ph.D.
Metcalf Professor of Science and Engineering and Dean Emeritus, College of Engineering; Boston University

We welcome the scientific community to this series which is held at the LIAI in the Seminar Auditorium. For more information or directions please visit http://www.liai.org

continued from page 1

on the database project. "This is an information-technology solution to a research problem."

Also involved in the project are the San Diego Supercomputer Center, The Scripps Research Institute and Science Applications International Corp – all in La Jolla – and the University of Copenhagen.

The San Diego Supercomputer Center, at UCSD, is already home to the nation's databank of proteins, the molecular building blocks that make up the body's cells and tissues. Having both a protein database and a clearinghouse of immunology research in La Jolla will further make San Diego a hub of biomedical research, Wilson said.

"Many groups studying one disease or another will gather all their data and submit it to the database," said Alessandro Sette, the institute's lead scientist on the project.

The Immune Epitope Database will be used by researchers to measure immune reactions, develop new diagnostic methods, design and evaluate new vaccines, understand how certain microbes can escape the body's immune system and gain insights into what constitutes a successful immune response, Sette said.

Epitopes are small sites on pathogens – or disease-causing agents – that the body's immune system focuses on when it begins an immune response. Many epitopes are small, integral parts of larger proteins that make up an infectious agent, such as the SARS or smallpox viruses.

The human immune system can recognize millions of epitopes, and discriminate between those found in healthy cells and those that signal an infection.

As researchers attempt to develop new vaccines, they need to know which epitopes will illicit the most powerful immune response. Sette believes another huge challenge in immunology will be getting a better understanding of why different human populations respond differently to the same pathogens, or the same vaccine.

"There is a massive effort going on, looking at hundreds of different genes that have variations in different human populations, and correlating those differences to (immune) responses," he said.

Although that genetic analysis is not the subject of the new database, the information will probably be linked to the epitope database to help researchers testing new vaccines.

In all likelihood, immunologist Wilson said, new vaccine discoveries will be made by those working at the intersection between proteomics, the study of proteins, and those who study epitopes.

"It's possible that a protein chemist will solve immunology problems because they'll finally be able to give us resolution of a protein that also has an immunologic function," Wilson said. "When those things come together, proteomics will actually hit the road. . . Someone's protein discovery will transform into a vaccine."

Find this article at:

http://www.signonsandiego.com/news/metro/20040503-9999-1m3immune.html

Coming Soon
in upcoming newsletters

Newsletter #2 — (10/15/04)
Focused Interviews

Newsletter #3 — (1/15/05)
Relevant Published Papers

Newsletter #6 — (10/15/05)
Directions for submission
Links to Updated Links and Tools Tables

Newsletter #7 — (1/5/06)
Tool Evaluation Results
Website Statistics

The IEDB is made possible by funding from:
The National Institutes of Health
The National Institute of Allergy & Infectious Diseases
The Department of Health and Human Services

Tool Evaluation Strategy

The Analysis Resource is designed to be a repository (and listing) of tools developed from scientific developers around the world. A variety of tools are expected to be present, and can be divided roughly into those that are analytical and those which are predictive. Analytical tools are those that extract or interpret data that is contained in the database, whereas predictive tools produce new information from data in the IEDB.

Predictive tools will extrapolate from data held in the database in order to make predictions where experimental results are not available. Many predictive tools aimed at identifying epitopes in protein sequences exist and new ones continue to be developed. As every tool can in some way demonstrate its usefulness, it is necessary to have a strategy in place to quantitatively evaluate tool performance in order to select which tools IEDB resources can support.

It is of great importance that the tool evaluation process is considered fair by the community, and we envision a selection process that is scientifically sound, unbiased and transparent to everyone. We are currently developing two separate evaluation strategies: one for existing tools, and a second to examine new tool development.

The first evaluation strategy, aimed at existing or newly developed tools that are available for predictions on the web or as computer programs, proposes that, just before data is made public in the database, it is given to these tools that will immediately have to return predictions which are stored for later evaluation. The performance of each tool can thus be tracked over time. As this evaluation strategy does not control what training data was accessible to the tool developer, this strategy cannot be used to compare the different methods used for tool development but rather assesses the practical value of accessible tools.

For the second evaluation strategy, designed to compare different methods for tool development, we are considering a "prediction contest." The same training dataset would be handed out to interested scientists, each of which would then generate a tool. Each developer would then submit the predictions generated by their tool for a large test set. At IEDB, a subset of the large test set would then be selected by a computer algorithm (disclosed publicly), allowing one to differentiate between the performance of all tools.

We recognize that the above descriptions only outline the currently considered tool evaluation strategies and both are still being developed. We'll be fully documenting the evaluation strategy in the coming months, and in coordination with the IEDB Working Group. Any input from tool developers or interested users is greatly appreciated, and should be

Newsletter/Contact Info

IEDB news is distributed four times a year. We welcome communication from the users of the IEDB database and invite suggestions for articles in future issues. Upon deployment of the database, we will actively solicit tool and epitope submission. To subscribe to the IEDB Newsletter or contact project staff, send your email information to the address below.

Immune Epitope Database
c/o La Jolla Institute for Allergy and Immunology
10355 Science Center Dr.
La Jolla, CA 92121
Phone: (858) 228-1400
E-mail: contact@immuneepitope.org

Principle Investigator:
Alessandro Sette, PhD (alex@liai.org)
Project Director:
Stephen Wilson, PhD (swilson@liai.org)
Contributors:
Phillip Bourne, PhD
John Sidney
Writers:
Holly Felzer
Production:
Michael Scarpelli
Graphic Design:
Hae-Jean Son

Questions and Answers

Q: I want to submit epitope data to the database. How do I go about it?

A: The database will not be able to accept data until deployment, scheduled for October 2005. There will be several ways to submit your data, depending on the size of your submission. Small data submissions will be submitted electronically, facilitated by a standardized form that will be accessible from the website. The team is currently developing a process to accommodate and facilitate bulk data submissions.

Q: I want to submit a tool to the database. How do I go about it?

A: Submitted tools will be subject to the standard evaluation process, described on page 3. We are actively soliciting tool recommendations from the scientific community. You are encouraged to submit your recommendations and feedback to tools@immuneepitope.org

Q: The website will provide user registration. Why should I register?

A: Database use is free and open to the public. There are benefits to registration however. A registered user may choose receive newsletters, and updates to the database. Also, if you register you can save display/ search preferences,

Q: How can I extract information out of the database?

A: Users will have the ability to perform custom queries on demand. In addition, the result set of popular queries will be available for download on demand.

Q: How will the database be populated?

A: The database will be populated by epitope data from literature, patent literature, and external submissions, which comprise small submissions from individuals, large submissions from other websites and epitope discovery programs. Future newsletters will debut static screens for the data submission and query interfaces.