

Methodology Report

IMMUNOCAT—A Data Management System for Epitope Mapping Studies

Jo L. Chung, Jian Sun, John Sidney, Alessandro Sette, and Bjoern Peters

Division of Vaccine Discovery, La Jolla Institute for Allergy & Immunology, La Jolla, CA 92037, USA

Correspondence should be addressed to Bjoern Peters, bpeters@liai.org

Received 30 November 2009; Accepted 7 March 2010

Academic Editor: Anne S. De Groot

Copyright © 2010 Jo L. Chung et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To enable rationale vaccine design, studies of molecular and cellular mechanisms of immune recognition need to be linked with clinical studies in humans. A major challenge in conducting such translational research studies lies in the management and integration of large amounts and various types of data collected from multiple sources. For this purpose, we have established “IMMUNOCAT”, an interactive data management system for the epitope discovery research projects conducted by our group. The system provides functions to store, query, and analyze clinical and experimental data, enabling efficient, systematic, and integrative data management. We demonstrate how IMMUNOCAT is utilized in a large-scale research contract that aims to identify epitopes in common allergens recognized by T cells from human donors, in order to facilitate the rational design of allergy vaccines. At clinical sites, demographic information and disease history of each enrolled donor are captured, followed by results of an allergen skin test and blood draw. At the laboratory site, T cells derived from blood samples are tested for reactivity against a panel of peptides derived from common human allergens. IMMUNOCAT stores results from these T cell assays along with MHC:peptide binding data, results from RAST tests for antibody titers in donor serum, and the respective donor HLA typing results. Through this system, we are able to perform queries and integrated analyses of the various types of data. This provides a case study for the use of bioinformatics and information management techniques to track and analyze data produced in a translational research study aimed at epitope identification.

1. Introduction

A crucial step for rational subunit vaccine design is the selection of antigens to include. For vaccines against infectious agents, antigens capable of inducing protective immune responses are desired. Several strategies based on genomic and proteomic approaches are being used to identify subsets of antigens that are highly expressed in general [1], on the surface [2], or during infection [3]. Antigens from these priority subsets are then followed up individually to test for their capacity to induce protective immunity. An alternative strategy that identifies protective antigens directly is to map targets of immune responses in previously infected hosts that successfully cleared the infection. This strategy is applicable whenever past infectious are known to provide protective immunity. In those cases, the capacity of antigens to induce protective immunity in a vaccine setting has been shown to correlate with the magnitude of the response against that

antigen post infection [4]. Therefore, knowledge of targets of immune responses in infected hosts has high value for vaccine design against infectious diseases. Knowing immune response targets is also crucial for the development of allergy vaccines, whose goal is to modulate the pathologic immune responses of allergic individuals towards those found in non-allergics [5, 6]. Similarly, for cancer vaccines to be successful, it is necessary to identify antigens targeted by immune responses associated with tumor regression [7, 8]. In summary, identifying the targets and characteristics of immune responses in well characterized host populations enables the rational design of vaccines.

One established approach for the identification and characterization of T cell immune response is the use of peptide based epitope mapping strategies. These are especially efficient when used in combination with bioinformatics predictions of candidate peptides [9]. The identification of epitopes, the exact molecular unit of recognition within an

antigen, also provides a mechanistic understanding of cross-reactivity of immune responses for different pathogens. This has recently been applied to study T cell immunity to swine flu [10, 11], and is important when designing cross-protective vaccines.

We have participated in two recently completed large-scale T cell epitope mapping projects, one to characterize epitopes responsible for the protective immunity conveyed by the smallpox vaccine [12–15], another to characterize epitopes in Arenaviruses [16–18], which has led to the generation of a candidate for a cross protective vaccine (M. Kotturi et al., PLoS Pathogens, in press). One lesson learned from these studies is that their data management is challenging, as the epitope response patterns discovered are typically complicated [19]. Also, these studies require the integration of large amounts and various types of data collected from multiple clinical and laboratory sites. Like many other groups, we have managed these data in a collection of spreadsheets, lab notebooks, and database systems designed for a single type of experiment. While each of these provides a sufficient mechanism to capture a specific type of information, the integrated analysis of these data often becomes labor intensive. Worse, problems due to inconsistencies in nomenclature and incompleteness of datasets are often only discovered at the time of analysis rather than at the time of data entry, which can make it hard or impossible to rectify them.

One way to address these issues is to collect data, from the start, in an integrated database system which immediately connects data from different sources. While this requires more work upon data entry, compared to less formal means of data capture such as spreadsheets, it greatly reduces the effort for data analysis. A deciding factor for our group to move toward an integrated database system was the award of a large scale NIH contract that aims to identify T cell epitopes from common allergens with the ultimate goal to facilitate the development of improved allergy vaccines. This study involves tracking donors enrolled at two clinical sites, capturing complex clinical history information, and storing results from multiple experiments performed in-house and through external providers (see Figure 1 for a breakdown of the overall study). Our experience with previous studies which were less complex, suggested that the maintenance and analysis of these data would be challenging. We therefore decided to establish IMMUNOCAT, which aims to store all data relevant for an integrated analysis of epitope mapping experiments, with the ultimate goal to enable rational vaccine design.

IMMUNOCAT integrates and enhances existing information management systems in our laboratory. Two stand-alone web-based relational databases were previously developed. ELICAT stores and analyzes results from ELISPOT [20] experiments used to test T cells for their cytokine secretion profile in response to peptide antigens. TOPCAT was designed to store the binding affinities determined in competitive peptide:MHC binding assays [21] utilizing a TopCount instrument. The databases are linked by the use of common peptide identifiers, which are assigned to every peptide that is synthesized in our studies. For the allergen

epitope mapping study, we needed to link the results of T cell reactivities stored in ELICAT with information about the donor from whom the T cells were derived. We first built a new web-based database called DONORDB to store answers from a clinical questionnaire, skin test results, and records of blood samples. Subsequently, this donor database was integrated with the existing ELICAT and TOPCAT databases (Figure 2) through the use of shared MHC allele and donation identifiers. The result is IMMUNOCAT, an integration of three web-based databases.

In the following, we describe the features and components of IMMUNOCAT, which has been in stable use in our laboratory for over a year. With this system, we are able to manage all data relevant to our epitope mapping studies in a central location and ensure their quality and integrity. Answers to integrated queries that would previously have taken significant effort can now be determined instantly, such as how many blood samples are available from donors that showed T cell reactivity to a certain peptide. In addition to the allergy epitope mapping study described here, IMMUNOCAT has been modified for use in two recently initiated epitope mapping studies for Dengue Virus and Mycobacterium Tuberculosis. This demonstrates the benefits of applying bioinformatics and information management techniques to epitope mapping studies in order to facilitate rational vaccine development.

2. User Administration

IMMUNOCAT provides different functionality to different groups of users. Each user has an individual account created by a database administrator. In order to access the database through the web, users have to identify themselves by supplying a user name and password. Different functionalities are provided based on the user group assignments. For example, there are two primary groups of users for DONORDB. The first group represents staff members at the clinical sites, who enter information from the clinical questionnaire and results from the skin test and blood draw. The second group is staff members at the laboratory site, who track the use of blood samples for each donor. Different users within each group are assigned different levels of access. For example, lab scientist supervisors have the ability to audit, add, delete, or modify the information in the system as necessary. Other lab scientists can only enter raw data and must request lab scientist supervisors to correct, for example, data entry errors. Separate projects using the DONORDB, such as the epitope mapping studies for Dengue Virus, use different web addresses for user access.

3. System Components and Features

3.1. DONORDB at Clinical Site. DONORDB is used at the clinical sites to capture information as donors are enrolled, interviewed, and undergo clinical procedures. For our ongoing allergen epitope mapping study, demographic information of the donor, such as gender, birthdate, ethnicity and parents' birthplaces (as an additional way to identify

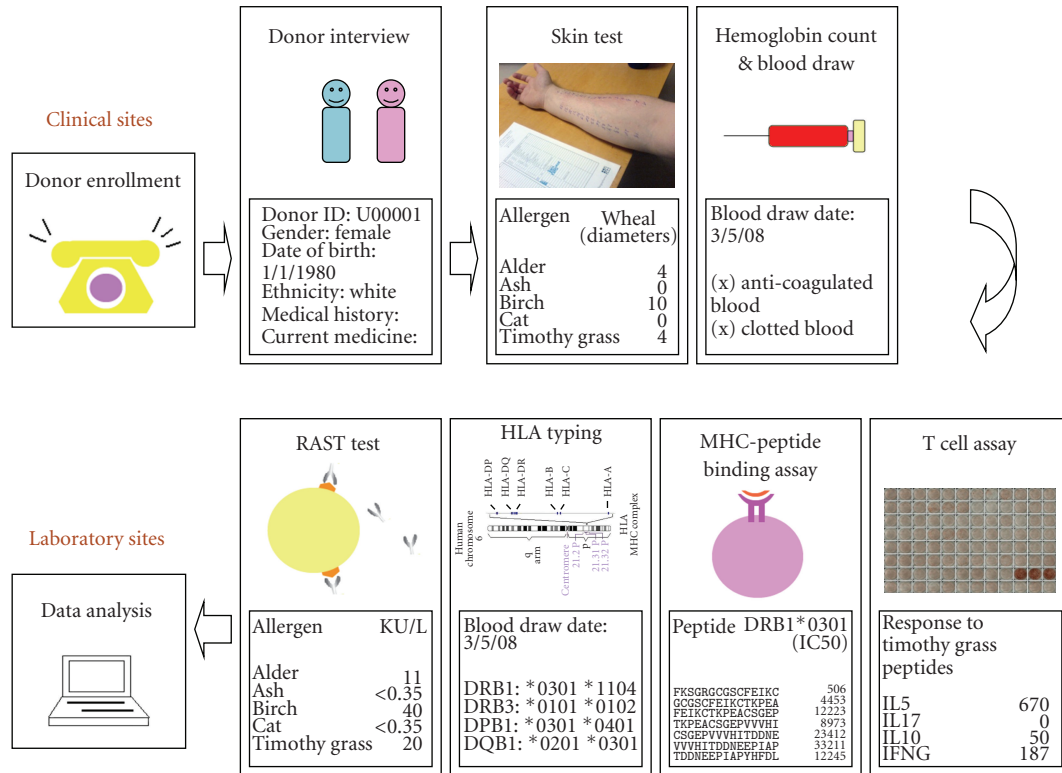


FIGURE 1: The allergy epitope mapping research workflow. In this study, allergic and normal donors were enrolled at two clinical sites (University of California, San Diego and National Jewish Medical Research Center). During the clinical interview, donor demographic information and medical history were recorded. Following that, allergen skin tests, hemoglobin test, and blood sample collection were performed. All the data and blood samples were sent to LIAI for further experiments and analyses. At LIAI, blood samples were tested for their immunogenicity (T cell activities) against a selected panel of peptides derived from common human allergens by ELISPOT assays [20] (T cell assays). Concurrently, the selected peptides were tested for their binding capacities to a set of common MHC molecules [21]. Peptide binding to MHC molecules is a necessary, but not sufficient, requirement for T cell recognitions [22]. Blood samples were also sent to an external company for radioallergosorbent (RAST) and human leukocyte antigens (HLA) typing test. The RAST test detects the amount of the IgE antibody in the blood that reacts with a specific allergen [23]. HLA typing determines the specific HLA types expressed by the donor, which is an indication of MHC restriction [24, 25]. The figure of skin test was adapted from the Wikimedia Commons file “File:Allergy_skin_testing.jpg”. The figure of HLA typing was adapted from the Wikimedia Commons file “File:HLA.jpg”.

ethnicity), is first entered. Next, a questionnaire is completed based on donor interviews (Figure 3). The questionnaire is tied to a scoring scheme that is used to classify donors into different disease categories, namely, three classifications of “allergic rhinitis” (none, possible, and probable) and “allergic asthma”. In terms of clinical procedures, the results of a prick skin test for allergic reactions against 32 common allergens are recorded in terms of flare and wheal diameters. Also, the hemoglobin count of a donor is measured and recorded to determine if he/she is safely capable of donating blood. If so, a separate blood donation visit is scheduled, and the volume of blood drawn is recorded in the database before samples are shipped for further analysis to the laboratory site.

At every step, the system ensures that donors meet all enrollment criteria, by notifying clinical staff if a donor falls under one of the exclusion criteria based on the entered information. During initial enrollment, exclusion would occur if, for example, the patient states that he is currently receiving allergen immunotherapy without yet being in maintenance. After initial enrollment, exclusion

would occur, for example, when the hemoglobin count is too low to allow a blood draw, or when the donor is still taking antihistamine medication right before the skin test. Integrating these questions and criteria into the database makes it easier for the clinical sites to keep track of all aspects of the enrollment criteria, and promotes consistency between sites.

Only anonymized information about the donors is entered at the clinical sites. As donors are enrolled, they get assigned a donation identifier by the system. This identifier is used to match blood samples shipped to the laboratory site with the information collected at the clinical sites. The clinical sites store the donation identifiers in their general patient records, which allow them to, for example, match repeat donations of a patient pre- and post immunotherapy, or match existing patient identifiers established at the individual sites to the DONORDB donation identifiers. No personally identifiable information (such as names or social security numbers) is stored in the database, which ensures that all anonymization requirements for data analysis are met.

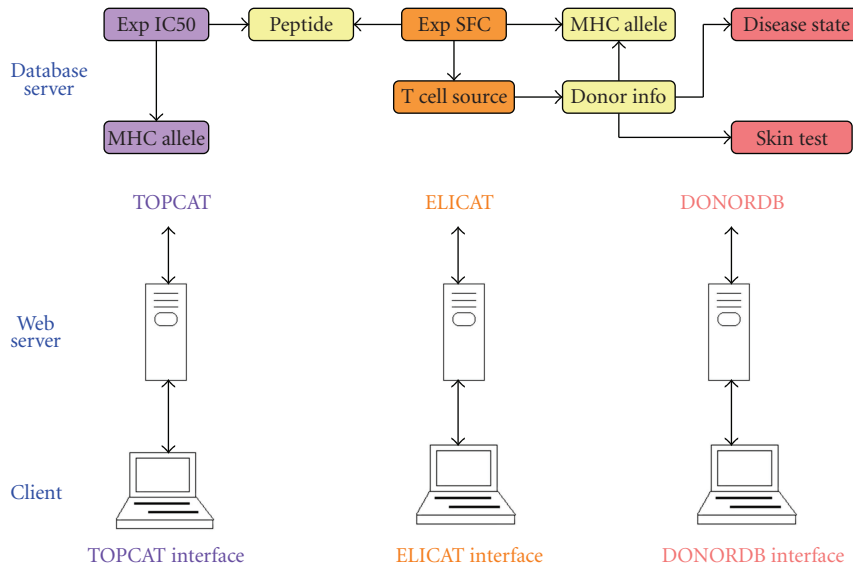


FIGURE 2: The architecture of IMMUNOCAT system. IMMUNOCAT is an integration of three web accessible relational databases. DONORDB is designed for data collected during donor enrollment and clinical tests. ELICAT stores the data from T cell assays performed using donor blood samples stimulated with a set of allergen derived peptides. TOPCAT stores the MHC-peptide binding data of the same set of peptides. Simplified database structures are presented at the top of the figure. Arrows represent the links between database tables. Tables shared by two databases are colored yellow. In DONORDB, donor information table is linked to skin test, disease state, and MHC allele tables. In ELICAT, table of experimental spot forming cell (SFC) is linked to peptide, T cell source, and MHC allele tables. In TOPCAT, table of experimental IC50 is linked to peptide and MHC allele tables. DONORDB is connected with ELICAT using donor IDs and MHC allele IDs. ELICAT and TOPCAT are connected with each other using shared peptide IDs.

By default, the data entry process follows the order in which the study was envisioned to be conducted. However, this order can be specifically overruled by users. At any time point, users can save completed data entry forms, and log out. They can log back in, identify the donor they were working on based on the donation ID, and continue data entry. If necessary, users can jump to different entry parts at their convenience. This was identified as critical functionality, as in clinical practice the idealized workflow outlined above is often interrupted due to changes in patients' schedules, and availability of time for data entry. Similarly, data entry requires internet access which for various reasons may not always be available. We found that enrollment information at clinical sites is therefore often recorded on paper before transferred into the database.

3.2. DONORDB at Laboratory Site. The lab scientists receive blood samples from the clinical sites, and process them to extract peripheral blood mononuclear cells (PBMCs). They use the database to track the availability of PBMCs from each donation, determine the number of vials that could initially be made from the blood shipment, record the use of vials in subsequent experiments, and track the location of the remaining vials in the freezer.

For each donor, data from two experiments performed at outside commercial laboratories are stored directly in DONORDB. The first set of data consists of IgE antibody titers against a panel of allergens determined in a RAST assay. The second set of data consists of the results from

HLA typing assays, which determine the specific set of MHC molecules the donor expresses.

Simple queries can be run against DONORDB in a web-based form (Figure 4). At this point, the system provides query functions to retrieve results from skin tests and the receipt of blood samples, as well as to monitor the number of donors in the two clinical sites. For example, lab scientists can query the database and see whether the blood samples of specific donors have been received at LIAI. They can also identify the donors who were skin test positive for a particular allergen and recruited at a specific clinical site. These queries are used routinely by laboratory scientists to select suitable blood samples for experiments.

3.3. ELICAT Database. ELICAT was designed for the management of data from ELISPOT assays. In these assays [13, 18], T cells are stimulated with peptide antigens, and those that respond by producing cytokines are visualized as individual spots. These recognized peptides are potential candidates for inclusion in an allergy vaccine. The experiments are performed on 96-well plates, and the numbers of spots are counted by an automated ELISPOT reader. The raw data generated is exported to a text file containing a matrix whose elements denote the number of spots detected in each well. This raw data is automatically imported into ELICAT, where it is connected with information about experimental design previously entered through the interface shown in Figure 5. An experiment is defined as one or more ELISPOT plates run with the same layout and cells from the same source. The plate layout specifies which wells are used to

File Edit View History Bookmarks Tools Help

Donor Information

Donor ID number:

Donor ID on the first visit (returning donor only):

1. Gender: Male Female

2. Date of birth:

3. Ethnicity:

Native American Asian/Pacific Islander Black

Hispanic White Other

4. Mother's birth country:

5. Father's birth country:

6. Comments:

Done

FIGURE 3: DONORDB: demographic information. The Figure displays a screen shot of the form used during initial enrollment of a donor.

File Edit View History Bookmarks Tools Help

[Enter Lab Data](#) [Upload Data](#) [Search Database](#) [Log Out](#)

Lab Data **Rast & Skin Test** **HLA Typing**

157 hit(s)

Donor ID	SITE	GENDER	ETHNICITY	YEAR OF BIRTH	BLOOD DRAWN DATE	BLOOD RECEIVED DATE	BLOOD VOLUME	PBMC	SERUM	LOCATION (RACK)	LOCATION (BOX)	FRC VL
D00001	Denver	F	Hispanic	1975	06-18-2008	06-19-2008	450	900	0	R25	B10	34
D00002	Denver	M	White	1978	07-02-2008	07-01-2008	300	410	0	R25	B11	20
D00003	Denver	F	White	1949	10-31-2008	07-01-2008	350	660	0	R25	B11	33
D00004	Denver	F	White	1983	07-01-2008	07-03-2008	450	1300	0	R25	B12	48
D00005	Denver	M	White	1978	07-02-2008	07-03-2008	465	500	0	R25	B12	25
D00006	Denver	F	White	1980	07-08-2008							
D00007	Denver	M	White	1977	07-08-2008	07-09-2008	400	680	0	R25	B13	34
D00008	Denver	M	White	1980	07-08-2008	07-09-2008	340	585	0	R25	B13	29
D00009	Denver	F	White	1950	07-09-2008	07-10-2008	435	710	0	R26	B1	35
D00010	Denver	M	White	1958	07-09-2008	07-10-2008	450	685	0	R26	B1	34
D00011	Denver	M	White	1960	07-14-2008	07-15-2008	500	865	0	R26	B2	43
D00012	Denver	F	White	1963	07-14-2008	07-15-2008	450	705	0	R26	B2	35
D00013	Denver	M	White	1979	07-15-2008	07-16-2008	430	880	0	R25	B3	44

Done

FIGURE 4: DONORDB: search functionality. The Figure shows a screen shot of the results generated when querying for all enrolled donors.

hold the tested peptides and which wells are used to hold the controls. The peptides used are specified in a separate file containing peptide identifiers.

Based on these inputs, ELICAT calculates summary metrics for the ELISPOT results for each peptide or peptide pool. The fraction of spot forming cells per million is calculated based on the number of spots detected in a well

and the number of effector T cells added. Based on replicate spot counts, a *P* value is calculated that evaluates if the detected spots for a peptide are significantly higher than those detected in negative control wells that contained no peptides. Finally, a stimulation index, which divides the average spot count for a peptide by the average background value, is calculated.

Import Experiments

Enter the information below about your experiment: (*=Required)

*Layout	Triplicates
*Starting Plate Number	205
*# of Effector Cells per well	1x10 ⁵
*Peptide File	C:\Documents and Settings\C Browse...

Next

Immunization - Enter the information below: (*=Required)

Please select to use a Predefined or New Immunization

Predefined
 New

*New

Select from existing predefined to build a new Immunization

D00004

Immunized Species

*Species	Human
Human Donor	D00004
Sex	Female
Age	26
*MHC Types Present	HLA-DPA1*0401 HLA-DPB1*0101 HLA-DPB1*0102 HLA-DPB1*0201
Disease State	Allergic
*Immunization Category	Occurrence of allergy
Comments	

FIGURE 5: ELICAT database: experimental design functionality. The figure shows two screens used to capture the experimental design of an ELISPOT experiment. By specifying the donor from whom T cells used in the experiment were derived, the experiment becomes linked to the information for that donor in DONORDB, and some fields such as the HLA alleles expressed by that donor become prepopulated. ELICAT also allows the users to specify the type of cytokines and cells used in the experiments (not shown here).

In the allergen epitope mapping study, two sets of ELISPOT experiments are performed to identify peptide epitopes. First, pools of peptides are screened for reactivity with PBMCs from allergic donors. Secondly, the positive pools are deconvoluted to identify which peptides caused the response. As a cutoff for the positive response >100 spot forming cells per million, $P < .05$ and stimulation index >2 were used. The ELICAT user interface allows description of each assay, including identification of each individual peptide tested or definition of each peptide component of specific pools.

3.4. TOPCAT Database. TOPCAT was the first database established in the laboratory, and its design was largely replicated in ELICAT. TOPCAT is used to store results from assays that evaluate MHC:peptide binding affinity, a necessary, but not sufficient, requirement for T cell

recognition [21]. In our assays [26, 27], the ability of the tested peptides at different concentrations to completely inhibit the binding of a radiolabeled high affinity ligand is determined. The concentration of the tested peptide at which the number of labeled ligands bound is reduced by 50% is the IC50 value. Under the condition utilized, measured IC50 values approximate the Kd value of the binding interaction [28]. The IC50 values are calculated based on radioactivity detection in a 96-well plate measured by TOPCOUNT NXT microscintillation reader. Information on the tested peptides, MHC alleles, and corresponding MHC-peptide binding affinities are stored and analyzed in the database.

All peptides used at the laboratory site are assigned a peptide identifier in the TOPCAT database, and the tubes containing the peptide are labeled accordingly. This identifies the sequence of the peptide and the protein and organism from which the peptide was derived, as well as the purity

of the synthesis. If the same peptide is synthesized multiple times, a new identifier is assigned to each unique synthesis. The ELICAT database uses the same peptide identifiers, and uses them to retrieve information about peptides from the TOPCAT system.

For the allergen epitope mapping study, peptides derived from protein sequences of common human allergens were synthesized. Each is being tested for binding affinity against a panel of 23 human MHC class II alleles. Through the integration of the three databases shown in Figure 2, it now becomes possible to directly link the allergic status of a donor, the peptides derived from the corresponding allergens, and their binding affinity to the MHC alleles present in the donor and the T cell reactivities detected towards these peptides.

4. System Design Process and Implementation

ELICAT and TOPCAT have been in routine use in the lab for more than 5 years. Regarding DONORDB, requirements were gathered in terms of what information needs to be captured for the process of donor enrollment, donor interviews, clinical procedures, blood sample shipments, and lab tests. For each step, the information desired to track progress was identified. Based on this, prototype web pages were created, and discussed with the clinical collaborators and LIAI lab scientists. Test of these prototypes was performed by both sites that analyzed use-scenarios, which led to the identification of additional requirements. These were, primarily, giving the users more flexibility in the order in which data was entered, and the capacity to store additional information. After three iterations, the prototypes were considered complete, and a functional system was implemented. The addition of more fields for new types of studies will require going through the same design process and modifications of the database and web application itself. As we gather experience in the kinds of modifications to expect, we plan to create tools that will ease making standard extensions.

All three databases are relational databases with web-browser-based interfaces and are implemented with SQL server 2005. The user interface was implemented with ASP for TOPCAT, and ASP.NET 2.0 for DONORDB and ELICAT. All three databases are hosted on a Dual-Processor Quad-Core Intel Xeon Rackmount Server. The database information is stored on a RAID 5 drive which ensures against single hard-disk failure, and is backed up daily through the LIAI network.

5. Usage

At this point, clinical information for 86 donors at the UCSD sites and 71 donors at NJMRC was deposited into DONORDB. Immunogenicity and MHC binding data for more than 60,000 peptides are stored in ELICAT and TOPCAT. We have not, so far, encountered data loss during operation. The source code of IMMUNOCAT is available for

download at http://donor.liai.org/Donor_Source.zip. Users will need to have an installation of SQL Server 2005 DBMS and ASP.NET 2.0 framework, and the ability to customize and configure the database for their own purposes. The code will be maintained for at least the next four years, and updates will be made available as new versions of the systems are completed. Prospective users are highly encouraged to contact the authors with any installation problems. For studies that require extensive modifications of the code and for laboratories operating incompatible IT environments, it will be preferable to redesign the application from scratch. In those cases, the present manuscript should still be useful in identifying requirements and reusing appropriate design patterns.

6. Future Prospects

We are planning to implement more search and analysis interfaces that take advantage of the integrated data in IMMUNOCAT. Currently, such analyses are being done by manually running SQL queries, which cannot be expected of laboratory technicians.

We are in the process of extending DONORDB to handle donors from three additional epitope-mapping studies, one dealing with donors from Dengue fever endemic regions, another dealing with donors that have tuberculosis, and a third with donors recently vaccinated against smallpox. This will require modifying some of the information captured for each donor, such as the serological typing of patients for the type of dengue viruses they have been exposed to. Other aspects, such as MHC typing and blood sample collection, will remain the same. In the longer term, we are aiming to make the customization of DONORDB for individual studies easier by allowing the addition of fields from predefined modules. This could, for example, mean that different lab tests performed could be chosen from an ontology such as the Ontology for Biomedical Investigations [29], and that vaccine terms could be taken from the Vaccine Ontology [30].

Finally, one of the goals of IMMUNOCAT is to submit data from completed studies into the Immune Epitope Database [31, 32]. Currently, both TOPCAT and ELICAT have export mechanisms that provide XML formatted data that can be imported into the IEDB. These mechanisms need to be integrated and further updated to integrate the data from DONORDB, and provide comprehensive export functionality of an entire study.

Acknowledgments

This work was supported by NIH contracts HHSN27220070-0048C, HHSN272200900042C, and HHSN272200900044C, and Grant no.5 T32 AI00749-14. The authors want to thank Carla Oseroff, Ravi Kolla, Carrie Moore, David Broide, Debbie Broide, Rafeul Alam, Linda Bannister, Susanna Burr, and Howard Grey for helpful discussions.

References

- [1] G. Grandi, "Antibacterial vaccine design using genomics and proteomics," *Trends in Biotechnology*, vol. 19, no. 5, pp. 181–188, 2001.
- [2] P. A. Cullen and C. E. Cameron, "Progress towards an effective syphilis vaccine: the past, present and future," *Expert Review of Vaccines*, vol. 5, no. 1, pp. 67–80, 2006.
- [3] C. Rollenhagen, M. Sørensen, K. Rizos, R. Hurvitz, and D. Bumann, "Antigen selection based on expression levels during infection facilitates vaccine development for an intracellular pathogen," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 23, pp. 8739–8744, 2004.
- [4] M. Moutaftsi, S. Salek-Ardakani, M. Croft, et al., "Correlates of protection efficacy induced by vaccinia virus-specific CD8⁺ T-cell epitopes in the murine intranasal challenge model," *European Journal of Immunology*, vol. 39, no. 3, pp. 717–722, 2009.
- [5] D. Campbell, R. H. DeKruyff, and D. T. Umetsu, "Allergen immunotherapy: novel approaches in the management of allergic diseases and asthma," *Clinical Immunology*, vol. 97, no. 3, pp. 193–202, 2000.
- [6] R. Valenta and D. Kraft, "From allergen structure to new forms of allergen-specific immunotherapy," *Current Opinion in Immunology*, vol. 14, no. 6, pp. 718–727, 2002.
- [7] J. A. Berzofsky, M. Terabe, S. Oh, et al., "Progress on new vaccine strategies for the immunotherapy and prevention of cancer," *Journal of Clinical Investigation*, vol. 113, no. 11, pp. 1515–1525, 2004.
- [8] R. Offringa, S. H. Van Der Burg, F. Ossendorp, R. E. Toes, and C. J. Melief, "Design and evaluation of antigen-specific vaccination strategies against cancer," *Current Opinion in Immunology*, vol. 12, no. 5, pp. 576–582, 2000.
- [9] A. S. De Groot, H. Sbai, C. S. Aubin, J. McMurry, and W. Martin, "Immuno-informatics: mining genomes for vaccine components," *Immunology and Cell Biology*, vol. 80, no. 3, pp. 255–269, 2002.
- [10] A. S. De Groot, M. Ardito, E. M. McClaine, L. Moise, and W. D. Martin, "Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008–2009 conventional influenza vaccine," *Vaccine*, vol. 27, no. 42, pp. 5740–5747, 2009.
- [11] J. A. Greenbaum, M. F. Kotturi, Y. Kim, et al., "Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 48, pp. 20365–20370, 2010.
- [12] M. Moutaftsi, H.-H. Bui, B. Peters, et al., "Vaccinia virus-specific CD4⁺ T cell responses target a set of antigens largely distinct from those targeted by CD8⁺ T cell responses," *Journal of Immunology*, vol. 178, no. 11, pp. 6814–6820, 2007.
- [13] M. Moutaftsi, B. Peters, V. Pasquetto, et al., "A consensus epitope prediction approach identifies the breadth of murine TCD8⁺-cell responses to vaccinia virus," *Nature Biotechnology*, vol. 24, no. 7, pp. 817–819, 2006.
- [14] V. Pasquetto, H.-H. Bui, R. Giannino, et al., "HLA-A*0201, HLA-A*1101, and HLA-B*0702 transgenic mice recognize numerous poxvirus determinants from a wide variety of viral gene products," *Journal of Immunology*, vol. 175, no. 8, pp. 5504–5515, 2005.
- [15] C. Oseroff, F. Kos, H.-H. Bui, et al., "HLA class I-restricted responses to vaccinia recognize a broad array of proteins mainly involved in virulence and viral gene regulation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 39, pp. 13980–13985, 2005.
- [16] J. Botten, J. Alexander, V. Pasquetto, et al., "Identification of protective Lassa virus epitopes that are restricted by HLA-A2," *Journal of Virology*, vol. 80, no. 17, pp. 8351–8361, 2006.
- [17] J. Botten, J. L. Whitton, P. Barrowman, et al., "HLA-A2-restricted protection against lethal lymphocytic choriomeningitis," *Journal of Virology*, vol. 81, no. 5, pp. 2307–2317, 2007.
- [18] M. F. Kotturi, B. Peters, F. Buendia-Laysa Jr., et al., "The CD8⁺T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus," *Journal of Virology*, vol. 81, no. 10, pp. 4928–4940, 2007.
- [19] A. Sette and B. Peters, "Immune epitope mapping in the post-genomic era: lessons for vaccine development," *Current Opinion in Immunology*, vol. 19, no. 1, pp. 106–110, 2007.
- [20] C. C. Czerkinsky, L. A. Nilsson, and H. Nygren, "A solid-phase enzyme-linked immunospot (ELISPOT) assay for enumeration of specific antibody-secreting cells," *Journal of Immunological Methods*, vol. 65, no. 1–2, pp. 109–121, 1983.
- [21] A. Sette, A. Vitiello, B. Rehman, et al., "The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes," *Journal of Immunology*, vol. 153, no. 12, pp. 5586–5592, 1994.
- [22] J. Klein, *Natural History of the Major Histocompatibility Complex*, John Wiley & Sons, New York, NY, USA, 1986.
- [23] G. Virella, *Medical Immunology*, Marcel Dekker, New York, NY, USA, 5th edition, 2001.
- [24] J. B. Henry, *Clinical Diagnosis and Management by Laboratory Methods*, W.B. Saunders, Philadelphia, Pa, USA, 20th edition, 2001.
- [25] V. Vengelen-Tyler and American Association of Blood Banks, *Technical Manual*, American Association of Blood Banks, Bethesda, Md, USA, 12th edition, 1996.
- [26] J. Sidney, S. Southwood, C. Oseroff, M. F. del Guercio, A. Sette, and H. M. Grey, "Measurement of MHC/peptide interactions by gel filtration," *Current Protocols in Immunology*, chapter 18, unit 18.3, 2001.
- [27] J. Sidney, E. Assarsson, C. Moore, et al., "Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries," *Immunome Research*, vol. 4, no. 1, article 2, 2008.
- [28] Y. Cheng and W. H. Prusoff, "Relationship between the inhibition constant (K_1) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction," *Biochemical Pharmacology*, vol. 22, no. 23, pp. 3099–3108, 1973.
- [29] T. O. Consortium, "Modeling biomedical experimental processes with OBI," in *Proceedings of the the 12th Annual Bio-Ontologies Meeting (ISMB '09)*, p. 41, 2009.
- [30] Y. He, "VO: vaccine ontology," in *Nature Precedings*, 2009.
- [31] B. Peters and A. Sette, "Integrating epitope data into the emerging web of biomedical knowledge resources," *Nature Reviews Immunology*, vol. 7, no. 6, pp. 485–490, 2007.
- [32] B. Peters, J. Sidney, P. Bourne, et al., "The immune epitope database and analysis resource: from vision to blueprint," *PLoS Biology*, vol. 3, no. 3, article e91, 2005.