



CONTENTS

Immunologic data management

Immunoinformatics & disease

Infectious diseases & computational vaccinology

Allergen informatics

Autoimmunity informatics

Expert opinion & five-year view

Key issues

References

Affiliations

Immunoinformatics and its relevance to understanding human immune disease

Vladimir Brusic[†] and Nikolai Petrovsky

Immunology research is characterized by the production of increasingly vast amounts of data, fuelled by genomics and proteomics projects and large-scale screening of pathogen- and antigen-host interactions. The need to store, manage and analyze this rapidly growing resource of experimental, clinical and epidemiologic data has given rise to the field known as immunoinformatics. Immunoinformatics represents computational methods and resources that are used in the study of immune function. It lies at the intersection of experimental and computational sciences and encompasses domain-specific databases, computational models and strategies drawn from artificial intelligence. For example, computational or artificial intelligence models are increasingly being used to simulate and improve our understanding of immune system behavior, such as antigen processing and presentation, as well as for analysis of host and pathogen genomes. Systemic models focus on simulating the behavior of cells or whole organs and can be used for applications such as tracking the course of infection or optimization of immunization protocols. Immunomics, the large-scale screening of immune processes, which includes powerful immunoinformatic tools, offers great promise for future translation of basic immunology research advances into clinical practice. Immunoinformatics is central to the research fields of immunogenomics, immunoproteomics and computational vaccinology.

Expert Rev. Clin. Immunol. 1(1), 145–157 (2005)

Sequencing of human and other higher organism genomes, as well as the genomes of various human pathogens, has provided large amounts of data relevant to the study of human immune pathology and, in particular, complex host–pathogen interaction. Functional and clinical data are available in specialist databases and in the medical and scientific literature. Additional sources of information include patient records, disease registries, tissue repositories and hospital databases. Collectively, these sources contain a large proportion of accumulated knowledge on human disease. The large quantity of accumulated data are a goldmine for the study of human immune disease and the quest for improved methods of disease prevention or treatment. Unfortunately, these data sources are largely unconnected, and while molecular databases and scientific literature are in the public domain, medical

records and patient samples are subject to privacy constraints and are off limit for many researchers.

Immunotherapy and immunoprophylaxis aim to modify patients' immune responses and thereby alleviate, cure or prevent disease (infection, cancer, allergy, transplant rejection or autoimmunity). These approaches are particularly promising in cases where alternatives such as surgery are limited, ineffective or do not exist. Vaccination is the single most successful immunologic intervention that has provided examples of stellar successes such as eradication of smallpox and the saving of some 170 million life-years [1]. Yet in 200 years of modern vaccine development, no more than 50 successful vaccines have been developed to counter the vast number of pathogens that threaten humans [2]. Immunotherapy and immunoprophylaxis offer high hopes in the fight against emerging infectious diseases such

[†]Author for correspondence
Institute for Infocomm Research,
21 Heng Mui Keng Terrace,
119613 Singapore
Tel.: +65 96 212 415
Fax: +65 67 748 056
vladimir@i2r.a-star.edu.sg

KEYWORDS:
allergen informatics,
autoimmunity, cancer
informatics, computational
vaccinology, immunologic
databases, immunomics

as new [3,4] or drug-resistant [5,6] pathogen strains, cancers [7], allergies [8] and autoimmunity [9]. Experiments have shown that immunotherapy can protect laboratory animals against viral infection [10], cancer [11] and autoimmunity [12]. Nevertheless, translation of successful immunotherapy applications in laboratory animals to safe and effective treatments for humans is a complex undertaking and the results, while promising, remain limited. Adverse effects of immunotherapy, life-threatening anaphylaxis being amongst the extreme, have been observed in animal experiments and concerns of this possibility have put stringent limits on immune experiments involving humans [12]. Human studies are also complicated by the diversity of the human immune system across the population, diversity of pathogens, variation in cancer antigens, individualized responses to allergens and unclear triggers of autoimmunity.

Immunology is a highly combinatorial science because of the diversity of the immune system and molecular variation of targets of the immune response. The presence of more than 10^{11} antibody specificities and greater than 10^{11} T-cell receptors in an individual, and genetic processes of recombination, somatic hypermutation, gene conversion and class switching, collectively contribute to a huge variability of immune system products. In addition, the number of 9-amino acid long linear targets of immune responses is greater than 10^{11} , while the number of conformational epitopes is far greater. Since more than 2000 alleles of human leukocyte antigens (HLA) have been reported and each individual expresses three to six HLA class I and at least that many HLA class II molecules, the theoretic number of HLA haplotypes is more than 10^{12} . Consequently, the number of products of the immune system in humans easily exceeds 10^{10} , which dwarfs the approximate 10^6 nonimmune-system products (including post-translational modifications and alternative splicing) encoded by the human genome. Due to the combinatorial nature of the immune system, the number of experiments required for systematic studies, such as T-cell epitope mapping, can be extraordinarily high. The large number of experimental data generated by such mapping experiments requires appropriate bioinformatic infrastructure spanning general and specialist databases, basic bioinformatics tools (e.g., sequence comparison algorithms) and sophisticated prediction tools that simulate molecular interactions. In addition, tools for analysis of molecular structure as well as for mining literature for functional information are increasingly necessary for immunologic research.

The immune system interacts with self (own organism) and foreign (an array of foreign antigens) agents. To maintain a healthy state in the organism, the immune system must mount adequate responses against pathogenic foreign and mutated self, and tolerate self as well as most environmental antigens. The study of self and foreign antigens requires the use of standard bioinformatics techniques. However, there are properties of the immune system that require specific bioinformatics methods and techniques. A peculiarity of the immune system is that immune recognition is effected through a large number of immune receptors ($>10^{12}$) that have a highly similar structure (primary, secondary and tertiary),

with small differences in sequence often resulting in large functional differences. Another property of immune processes is that they are commonly regulated with minute amounts of bioactive proteins (such as cytokines) and molecular and cellular process responses could be faster than our current ability to measure them. Immunoinformatics, therefore, requires general bioinformatic techniques, as well as those that address the highly combinatorial problems specific to immunology.

Immunologic data management

Immunoinformatics, also referred to as computational immunology, represents computational methods and resources that are used in the study of immune function. It focuses on the management, analysis and modeling of immunologic data [13]. The last few years have seen a resurgence in the field previously known as theoretic immunology [14], for which origins can be traced as far back as the 1910s (mathematical models of malaria transmission by Sir Ronald Ross). The more recent foundations of theoretic immunology were crystalized in the 1970s and 80s with the first international meeting on theoretic immunology held in Santa Fe, New Mexico, in 1987. At that time, theoretic immunology focused on mathematical modeling of the immune processes. However, over the years, the scope of this field has expanded to cover other aspects of immune system informatics including immunologic databases, molecular and systemic modeling, structural immunoinformatics and simulation of wet-laboratory experiments. An exciting aspect of the current immunoinformatics wave is that for the first time, rather than being isolated fields, there is a rapid closing of the gaps between immunoinformatics and both experimental and clinical immunology. Immunoinformatics is being adopted as standard methodology by rapidly increasing numbers of the mainstream immunologists, just as molecular biology, which first arose as a discrete field and has now been incorporated as a mainstream ingredient of immunology research. Immunoinformatics plays a critical role in enabling management and analysis of the growing amounts of immunologic data. Immunoinformatics resources include data sources such as general and specialist molecular databases, patient and disease registries and the scientific literature. Currently, these data are growing both in number and complexity. Analysis of these data requires sophisticated immunoinformatics tools for accurate description of immune processes, identification of new knowledge, and development of new research applications and disease treatments. Current developments in immunoinformatics include improved design and management of data sources, development of novel informatics tools for analysis of immunologic data and creation of predictive models that simulate specific immune interactions at either the molecular, single-cell or system level. Immunoinformatics resources may be used to assist in [15]:

- Understanding the origin, structure and function of immune-related genes, proteins and their products
- Searching for genes, proteins and their variants involved in disease

- Studying gene and protein expression in physiologic and pathologic conditions
- Studying immune system disorders
- Studying infectious disease, cancer, autoimmunity and allergy
- Discovery, design and engineering of immune therapies, prophylaxis and diagnostics
- Building computational models for accurate simulation of laboratory experiments

Data sources

Data sources include scientific literature, molecular databases, clinical records and patient registries. Scientific literature provides a wealth of data including functional and higher-level information not captured in molecular databases. However, this source of data is not structured and information appears as free text. The PubMed database contains more than 15 million abstracts of life science articles and is currently adding 600,000 records a year [101]. Many of the articles are directly linked to full-text articles available online. The astounding quantity of information is exemplified by the number of records related to specific topics added annually. For example, in 2004 alone there were 4200, 6700 and 17,700 newly added abstracts containing the respective terms major histocompatibility complex (MHC) or HLA, allergy or allergen and cytokine. Accessibility to abstracts is now universal and the quality and quantity of extracted information largely depends on the availability of text-mining tools. Due to the large number of articles available, the quality of immunology research will increasingly depend on a researcher's ability to access data sources that contain critical information.

The number of molecular databases is growing steadily. The latest database issue of *Nucleic Acids Research* lists 781 databases, of which 20 have been classified as immunology related [16]. Included in this database listing are at least ten immunology databases classified under other categories (IMGT/3Dstructure-DB, IMGT/PRIMER-DB, HIV Molecular Immunology Database, IPD, ALPSbase, ANTIMIC, APD and SDAP). This listing of databases is incomplete as many more specialist databases are available for access and data analysis. An additional listing of immunology-related databases is available in [15], IMGT Bloc-notes [102]. Allergy-related data sources are also available in [17]. Immunology databases are designed to facilitate both basic and applied research in immunology. Database users need to be aware of the important issues that can affect the extent and accuracy of information extracted from specific databases. Individual databases vary in the depth and breadth of coverage, accuracy, completeness, redundancy, conformance to standards and nomenclature, and the level of detail of information stored therein. Accessibility to the data, ease of use and the availability of online analysis tools also varies between databases. Accurate, complete, nonredundant and detailed information is critical for support of immunology research and practice and, therefore, database users must ensure that data extracted from all these disparate sources are of sufficient quality. The issues of database quality have been extensively reviewed elsewhere [15].

Various laboratory tests, responses to therapies and interventions are captured in clinical records. Traditionally, clinical records have been useful for medical-risk management, improvement of patient care, clinical admission and discharge planning, billing, and access to patient data by physicians. Three major uses of patient records have been identified: recording and use of medical data for co-operative care, decision support using accumulated medical knowledge, and use of patient data in research and reporting [18]. Due to data security issues, restricted information entered in patient records, lack of co-ordination between institutions, lack of willingness to share data and a relatively slow adoption rate of information technology, the use of patient records for research remains limited [19]. Detailed and complete collections of patient records (including data on genotypic and phenotypic measurements, various laboratory tests, patient selection and treatment and outcomes that focus on specific conditions) have clear value in diagnostics, optimization of treatments and prediction of clinical outcomes. Patient registries, including national level registries, capture condition and disease-related patient data. Examples of patient registries include the Australia and New Zealand Dialysis and Transplant Registry of Renal Transplant Patients [20] and the Immune Tolerance Registry of North America [21]. Given the increasingly stringent regulatory environment, combining patient registries and knowledge of molecular processing will grow in importance in designing novel therapies [22,23]. Additional rich sources of clinically relevant data are registries of clinical trials. Until recently, these data were not accessible to researchers. In an unprecedented decision, the International Federation of Pharmaceutical Manufacturers and Associations agreed to create a free, publicly accessible clinical trials registry that will contain a description of the trials and permit access to a clinical trials results database [103]. Combining these data with patient and clinical records would provide a rich source of information for the study of disease. Critical issues that need to be resolved before researchers can get access to these data sources include patient privacy and protection from misuse of patient data and the definition of acceptable uses of patient and clinical records. This will require input from clinicians, hospital administrators, researchers, informaticians, lawyers, patient groups, representatives and regulators. The key issues to be resolved will include acceptable use of patient and clinical data, protection of patients, clinicians and health service providers from potential misuse, and ensuring the safety of data. Immunology provides multiple views to relevant data. These data may describe simple interactions, such as binding of a single peptide to a MHC molecule, or complex systemic data, such as the description of an immune response or the description of complex immune responses (e.g., vaccine responses within a population) (FIGURE 1).

Data analysis tools

Numerous mathematical, statistical or computer simulation models – computational models – have shaped immunology research [24]. The amount of currently available data from multiple sources easily surpasses our ability to manually collect and integrate data and perform various analyses. This has led to

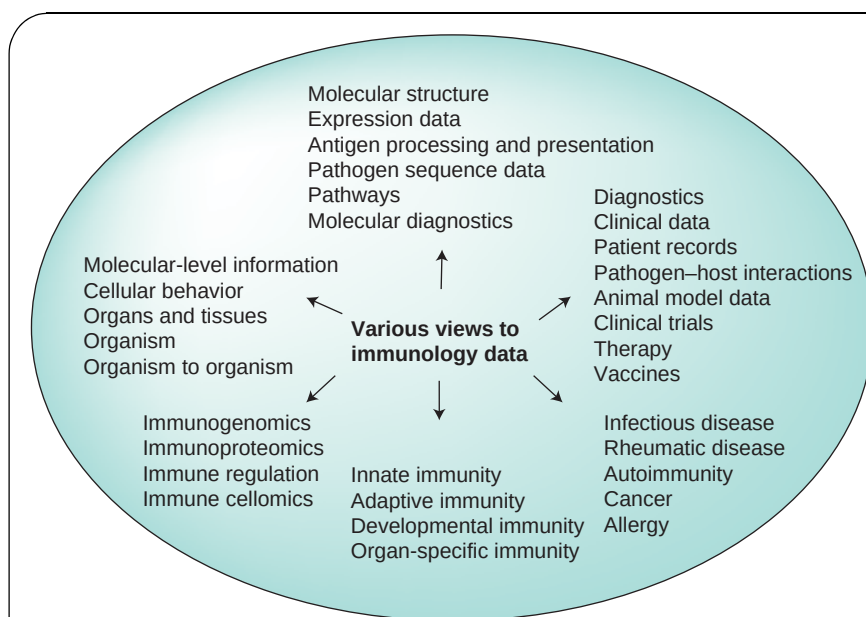


Figure 1. Various views to data generated by immunology-related studies. This list is representative of data of various complexity and extent.

an increased reliance on computational approaches and modeling. The main groups of tools that are used in immunoinformatics or related fields include those for literature mining, statistical and comparative tools, models of molecular interactions and system-level models. Automated extraction of information from free texts is known as text mining. Text mining is at the early stages of development in life sciences, and examples include identification of functional domains in proteins [25], gene grouping and profiling [26], and identification of gene networks [27]. Specific immunologic applications of text mining will appear over time; for example, the authors of this article are involved in research into the automatic identification of allergen crossreactivity information directly from scientific texts.

Statistical or mathematical models have been used extensively in developing immunologic assays [24], including measurement of antibody affinity, agglutination, quantification of antibody or ligand binding, radioimmunoassays, precipitation reactions in gels, estimation of antibody or antigen concentration, immunofluorescence, enzyme immunoassays and enzyme-linked immunosorbent assays (ELISA), among others. Computational modeling has also been applied in quantification of cellular assays including measurement of cytotoxicity, migration, phagocytosis, limiting dilution analysis, basophil degranulation and cell proliferation, among others. Most of these computational models are programmed within instruments and are therefore not visible to the user. Another set of data analysis tools include bioinformatic methods used for database searching and biologic sequence comparison including standard methods such as FASTA, BLAST or Smith-Waterman algorithms [28] for pairwise sequence comparison, or methods for multiple sequence analysis. The analysis of DNA focuses largely on characterization of DNA sequences and genes, while protein sequence analysis focuses on:

- Finding conserved patterns of functional domains in related proteins
- Prediction of protein secondary structure from sequence
- Prediction of 3D structure by homology modeling or *de novo*
- Prediction of function of whole proteins or their domains

Examples of use of standard bioinformatics tools and references can be found in the recent review [28]. Several immunoinformatics tools have been specifically developed to be accessible via the internet, such as a suite of immunogenetics tools [29,30]. These user-friendly, online tools enable the analysis of antibody and T-cell receptor sequences (IMGT/V-QUEST [31], IMGT/JunctionAnalysis [32]), graphic representation of 2D structures (IMGT Colliers de Perles) and the contact analysis in antigen receptors and MHC 3D structures (IMGT/Structural Query) [33].

Newer computational tools that focus on molecular interactions include methods for mapping T-cell epitopes [34]. Computational methods for simulation of antigen processing include prediction of proteasomal cleavage [35,36], and peptide binding to the transporter associated with antigen processing (TAP) [37,38]. The basic methods for prediction of antigen processing and presentation have been extended to more sophisticated computational models for the identification of promiscuous HLA-restricted T-cell epitopes [39,40], MHC supermotifs [41] and T-cell epitope hotspots [42]. Individual prediction systems normally produce different results. An example is illustrated in FIGURE 2, where the 121-amino acid long sequence of the nuclear export protein NS2 from influenza A virus (GenPept accession NP_859033) was searched for 9-mer candidate binders to a mouse MHC molecule H-2K^d using four internet-accessible systems. The prediction results were compared for top ten predictions by each system (TABLE 1). The three best predictions (peptides 98–106, 72–80 and 105–113) were deemed as top ten candidates by all four methods. Five peptides (8–16, 40–48, 30–38, 20–28 and 35–43) were predicted by three methods. Other peptides in the list have lower probability of binding compared with the peptides with overall rank 1–8 (TABLE 2). This example shows that individual predictions need to be taken with care and predictions may be improved by a consensus of multiple methods.

Computational models of cellular or higher level processes or interactions have a longer history than those focusing on molecular processes, but are also more complex. These include models of T-cell responses to viruses [43], analysis of MHC diversity under host–pathogen coevolution [44], maturation of B-cells [45], or even dynamic models of the immune system, which can simulate both cellular and humoral immune

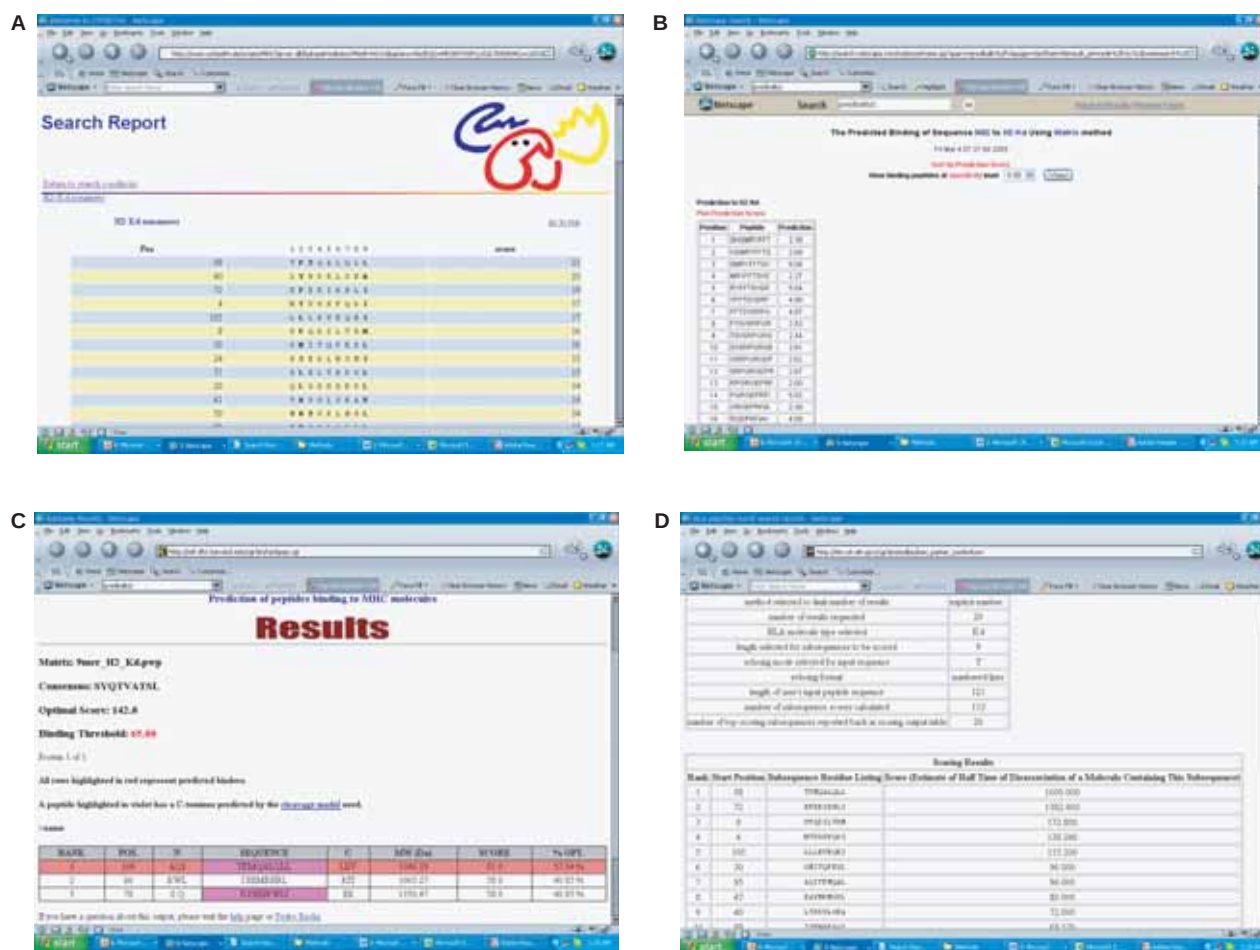


Figure 2. Prediction of 9-mer peptide binders from nuclear export protein NS2 from influenza A virus (GenPept accession NP_859033) to the H-2K^d molecule of BALB/c mouse. Four prediction systems were used and these predictions were compared: (A) SYFPEITHI [104], (B) PRED^{BALB/c} [105], (C) RANKPEP [106] and (D) BIMAS [107].

responses [46]. A brief review describing major issues concerning mathematical modeling can be found in [47]. In the past, such models had limited application because the best computer resources only allowed modeling of immune cells and antigens in numbers that were many orders of magnitude smaller than the size of the real immune system. However, with the ongoing dramatic increase in computational power, particularly grid computing, computational models can now start to match the complexity of the immune system. In turn, this will allow system-level modeling of the immune system using molecular-level models that represent real antigens and receptors, rather than idealizations that use bit-strings. The outcome will be realistic models that should be able to match the complexity of the real immune system and greatly assist our understanding not only of normal functioning of the immune system, but also elucidate how perturbations of the system lead to immune diseases. An example of the application of such models is a numerical simulator of cellular and humoral immune responses to HIV-1 infection [48]. This model uses

predictive parameters of disease progression and found that HIV-1 proteins accumulate single-position mutations primarily in the peptide sequences that are recognition targets for cytotoxic CD8⁺ T-cells. These results indicate that cell-mediated immunity has a major role in control of the viral load and provides a possible explanation for differences between normal and rapid progressors versus long-term nonprogressors. The results also suggested that differences in the early immune response may account for different disease progression. The authors propose that such computational models may help define the predictive parameters of HIV dynamics and disease progression, which has potential applications in vaccine and immunotherapy development.

Immunoinformatics & disease

A major driver for immunology research is to use the understanding of structural and functional elements of the immune system at the cellular, organ, organism and population level, and to develop practical applications in the areas of disease diagnostics,

Table 1. Prediction of 9-mer peptide binders from NS2 protein to H-2K^d.

RANKPEP			PRED ^{BALB/c}			SYFPEITHI			BIMAS		
Pos	Peptide	Pred	Pos	Peptide	Pred	Pos	Peptide	Pred	Pos	Peptide	Pred
98	TFMQALQLL	81	72	KFEEIRWLI	8.21	98	TFMQALQLL	21	98	TFMQALQLL	1600
8	SFQDILTRM	80	30	GMITQFESL	8.14	40	LYRDSLGEA	20	72	KFEEIRWLI	1382
40	LYRDSLGEA	61	98	TFMQALQLL	7.90	72	KFEEIRWLI	19	8	SFQDILTRM	172
105	LLLEVEQEI	59	99	FMQALQLL	7.85	4	NTVSSFQDI	17	4	NTVSSFQDI	138
79	LIEEMRHRL	58	37	SLKLYRDSL	7.67	105	LLLEVEQEI	17	105	LLLEVEQEI	115
72	KFEEIRWLI	55	20	QLGSSSED	7.26	8	SFQDILTRM	16	30	GMITQFESL	96
20	QLGSSSEDL	52	79	LIEEMRHRL	7.25	30	GMITQFESL	16	95	AQITFMQAL	96
95	AQITFMQAL	50	68	QLSQKFEEI	6.91	24	SSEDLNGMI	15	47	EAVMRMGDL	80
68	QLSQKFEEI	50	105	LLLEVEQEI	6.84	37	SLKLYRDSL	15	40	LYRDSLGEA	72
37	SLKLYRDSL	49	112	EIRTSFQFL	6.82	20	QLGSSSEDL	14	89	ITENSFAQI	69

Note: Four prediction systems were compared: SYFPEITHI, PRED^{BALB/c}, RANKPEP and BIMAS. Individual predictions indicating position of the peptide (Pos), peptide and prediction score (Pred) are listed.

prevention and treatment. The complexity of the human immune system (characterized by variability, plasticity, degeneracy and adaptivity), the variability of pathogens and of self and environmental antigens, and the presence of multiple regulatory pathways, ensure that huge amounts of data will be necessary to decipher the workings of the immune system. Since the immune system interacts with and is trained on self-proteins and also undergoes a process of adaptive learning through interaction with a large array of foreign antigens, the amount of data produced by large-scale immunology projects will soon overwhelm the already vast amounts of data produced by genomics projects. This will ensure an increased dependency on immunoinformatics for management and analysis of these data. The current main applications of immunoinformatics in the study of disease include models to understand the pathogenesis of disease and dynamics of the immune response, computational vaccinology and diagnostic tools. Advances in instrumentation, informatics hardware and software and an unprecedented growth of life sciences (both basic and applied) have the potential to deliver significant transformation of medicine and healthcare. Clinical immunology will benefit from access to a vast repository of accumulated information and knowledge. A brief list of immunoinformatics resources and representative applications are listed in TABLE 3. Currently, these sources and tools are mostly separated. Adequate integration of these resources will enhance data analysis and support for immunology research. Data warehousing, which focuses on integration of data and tools for discovery and decision making, is a suitable technology that enables extracting further value from data [49,50].

Medical informatics and bioinformatics are playing an increasingly critical role in research, diagnosis, prevention and treatment of tumors. Large-scale cancer projects already require integration of huge amounts of data, and necessitate

tools that allow access to and analysis of large repositories of cancer-related data. An example of such an initiative is the Cancer Biomedical Informatics Grid (caBIG), which connects a network of 500 individuals and 50 institutions who share data and analyze tools to speed up the development of innovative approaches for the prevention and treatment of cancer [51]. Changes in the cancer field are being supported by an informatics infrastructure that enables patients, physicians and researchers to instantaneously access rapidly growing amounts of information. Large data warehouses are required to host the terabyte-level of experimentally generated data for use by cancer clinicians and researchers [50]. Competent immune responses to cancer are likely to be restricted to the immunome of a specific cancer, including the set of antigens that drive successful immune responses. However, that set of antigens varies between different tumors and is still difficult to uncover in its entirety. Bioinformatics tools used for the discovery of the components of the immunome include sequence analysis, antigen processing and presentation mapping tools, as well as tools for analysis of gene and protein expression, and managing data from high-throughput immunology assays. The databases, analysis tools and bioinformatics methods for support of large-scale screening are rapidly growing in number.

The 2005 database issue of *Nucleic Acids Research* lists 14 cancer-related molecular databases [16]. These databases mainly focus on cancer-related genes and gene expression. Genetic databases are useful for basic research; however, they are not linked to databases that contain functional or clinical data. Intelligent design of high-throughput methods should enable bridging this gap and assist the assigning of functional properties to tumor-related genes [52]. Systems biology approaches target identification of a small number of antigens expressed by

Table 2. Prediction of 9-mer peptide binders from NS2 protein to H-2K^d.

Number	Pos	Peptide	Rank				
			RANKPEP	PRED ^{BALB/c}	SYFPEITHI	BIMAS	Overall
1	4	NTVSSFQDI	-	-	4	4	9
2	8	SFQDILTRM	2	-	6	3	4
3	20	QLGSSSEDL	7	6	10	-	7
4	24	SSEDLNGMI	-	-	8	-	14
5	30	GMITQFESL	-	2	7	6	6
6	37	SLKLYRDSL	10	5	9	-	8
7	40	LYRDSLGEA	3	-	2	9	5
8	47	EAVMRMGDL	-	-	-	8	14
9	68	QLSQKFEEI	9	8	-	-	12
10	72	KFEEIRWLI	6	1	3	2	2
11	79	LIEEMRHRL	5	7	-	-	10
12	89	ITENSFAQI	-	-	-	10	16
13	95	AQITFMQAL	8	-	-	7	11
14	98	TFMQALQLL	1	3	1	1	1
15	99	FMQALQLLL	-	4	-	-	13
16	105	LLLEVEQEI	4	9	5	5	3
17	112	EIRTSFQQL	-	10	-	-	16

Note: Four prediction systems were compared: SYFPEITHI, PRED^{BALB/c}, RANKPEP and BIMAS. Ranked predictions for each method and overall ranking are listed.

cancer cells that are suitable targets of immune responses for natural defense against cancer. A recently reported study describes proteomic mapping of *in vivo* targets for antibodies in lungs, and solid tumors in experimental animals, and defines aminopeptidase-P and annexin A1 as targets of anti-cancer immune responses [53]. Translating such studies into broader clinical applications is complicated by the diversity of the human population as well as the diversity of tumors. Informatic methods have been used for classification of tumors into subtypes, which supports decision making for the selection of therapeutic approaches; however, applications in cancer immunology are yet to come [54].

Bioinformatics methods such as database searching, sequence comparison, motif analysis and structural modeling are useful for assigning tentative functions to uncharacterized genes [55]. Listings of tumor antigens are available [56]. This list includes antigens that have defined T-cell epitopes, but excludes other groups such as analogs or artificially modified epitopes, viral cancer-related antigens or antibody-targeted antigens. Current efforts include consolidation of antigen data and integration with immunoinformatics tools such as antigen processing and presentation predictions [34]. Cancer immunotherapies focus on immunomodulation [22] or target cancer antigens [57]. Animal studies have demonstrated

successful protection against some types of cancer using cells that express cancer antigen and produce interleukin (IL)-12 [11]. Computational models of antigen processing and presentation link molecular-level interaction with the fate of antigens [34–42]. Molecular-level interactions initiate and regulate cellular immune responses. Cellular processes define immune responses and models of these processes are useful for understanding the kinetics of the cells during immune responses. Various models of this type have been developed for immunology and include continuous, discrete and automata-based models [58]. Mathematical models can translate speculative hypotheses into quantitative descriptions. The parameters of mathematical models can be tuned to represent the real behavior of the immune system and in turn can be used to refine hypotheses or offer explanations of processes that are difficult to measure. Practical applications of mathematical models continue to primarily focus on description of immune processes. However, there are developments in which mathematical models have been used for optimization of immunization protocols and experimentally verified in mice [94].

Thus, cancer research is increasingly dependent on informatics resources for access to accumulated data and knowledge, as well as identification of key pointers that support the design of experiments, improved diagnosis and therapeutic approaches.

Table 3. Immunoinformatics resources and selected cancer immunoinformatics applications.

Immunoinformatics	Ref.
<i>Data sources</i>	
Molecular databases	[16]
Specialized lists, nomenclature	[56,108]
Clinical records	[18,19]
Clinical trials	[103]
Patient registries	[20,21]
Literature	[101]
Web resources	[29,51,85]
<i>Tools</i>	
Database searching	[109]
Sequence analysis	Multiple sites
Statistical analysis	[24]
Molecular modeling	[87]
Data warehousing	[49,50]
Data mining	[85]
Text mining	[25–27]
Predictive models	[34–40,76–80]
Systemic models	[43–47]
<i>Examples of applications</i>	
Gene characterization	[52,55]
Antigen characterization	[53]
Antigen receptor repertoire analysis	[31,32]
3D structure analysis of immune receptors	[33]
Assessment of allergenicity	[28,76–80]
Identification of T-cell epitopes	[34–42]
Deimmunization of proteins	[82]
Design of immunotherapy	[7,22,57]
Vaccine design	[2,40,62–64,70,72,75]
Optimization of immunization protocols	[94]
Study of immune system behavior	[43–48]

In the past, the main problem with utilization of cancer data was lack of access capability. Current problems are the opposite – an enormous amount of data are readily available, but it is difficult to identify key data from the huge quantities of data distributed across the world.

Infectious diseases & computational vaccinology

Bioinformatics is growing in importance in the study of infectious disease. It is essential for reporting and tracking infectious disease [59], defining major circulating variants of pathogens [60,61], microbial diagnostics [62], pathogen databases [16] and vaccine development [63,64]. Computational models have been developed with the aim to facilitate understanding of complex pathogen-related processes. Reported computational models of infection and immune responses to infection include those focusing on viral [65,66], bacterial [67], parasitic [68] and fungal pathogens [69].

Traditionally, major applications involving informatics of infectious disease involve epidemiology and vaccine research. With the advancement of various ‘omics’ [61], molecular diagnosis and characterization of pathogens and immune responses is rapidly growing. For example, proteomics enables high-throughput profiling of antibody responses to infection [70]. Deciphering immune responses to various pathogens is a complex task because of the variability of pathogens and variability of the human immune system [2]. The enormous combinatorial space of molecular interactions involved in pathogen interactions with human individuals and populations necessitates the use of computational and mathematical approaches for preliminary screening, interpretation of results and simulations that offer possible explanations of biologic processes.

The recent epidemics of the severe acute respiratory syndrome (SARS) demonstrated the importance of informatics and data management for dealing with emerging epidemics [3]. Genome sequencing, sequence comparison and phylogenetic analysis were crucial for the identification of 11 open reading frames of the SARS virus and its characterization as a distinct member of the coronavirus family [71]. Sequence analysis of the phylogenetic relationship of influenza viral strains circulating in Taiwan revealed the accumulation of changes by antigenic drift in isolates of hemagglutinin (HA) and nonstructural genes. These studies also showed the emergence of a new cluster that evolved from the old lineage (YA88) [4]. The authors analysis of HLA-DR hotspots (peptides that bind multiple HLA-DR molecules) in 1091 non-redundant sequences of influenza HA showed that there are four to six hot spots in each of the four studied subtypes of HA [HEINY ET AL., UNPUBLISHED OBSERVATIONS]. The length of each hotspot is between 15 and 36 amino acids and 20 different peptides and their 168 variants (mainly single amino acid variation) were identified that represent the total antigenic diversity of the 1091 studied HA sequences. This analysis included sequence comparison, elimination of duplicates, analysis of antigenic diversity and computational identification of the HLA-DR hotspot using the computational system, MULTIPRED [42]. An important contribution to the development of vaccines against infectious diseases is screening of complete genomes of pathogens [72]. To date, genomes of more than 100 pathogens have been sequenced [73] and screening these genomes for T-cell epitopes is underway. Large-scale screening of pathogens for T- and B-cell

epitopes from NIAID class A–C pathogens and other pathogens of global importance is also currently underway [74]. These pathogens include *Bacillus anthracis*, *Mycobacterium tuberculosis*, *Clostridium tetani*, *Francisella tularensis*, *Yersinia pestis* and various viruses including smallpox, flaviviruses, arenaviruses, rabies, influenza, SARS and hepatitis A, among others. The importance of these large-scale efforts is that they combine bioinformatics, genomics and proteomics with traditional experimental approaches. Such large-scale efforts will for the first time enable a global view of human immunity to pathogens. In this way, the diversity of pathogen genomes will be mapped together with T- and B-cell epitopes contained in pathogen genomes in the context of population-wide HLA allelic variants. The database component of this project will enable access to data and utilization of specialized tools for further analysis [74]. Vaccine formulations using various constructs identified by computational screening have been studied extensively; vaccinology is set to be a major beneficiary of bioinformatics techniques for epitope research and screening [2,40,62,63,70,72,75].

Allergen informatics

Developments in allergen informatics include formation of improved allergen databases, identification of B- and T-cell epitopes, structural characterization of allergens, assessment of allergenicity and allergic crossreactivity and analysis of genetic factors that contribute to allergic responses, as well as genomics and proteomics studies of allergy [17,28]. Clinical allergology will be a major beneficiary of technologic advances in instrumentation and informatics, which will enable improved diagnosis, prevention and treatment of allergic disease [76]. Two major reviews of the application of bioinformatics to allergy were published recently [17,28]. The vigorous activity of the last 2 years has seen no less than four reports of algorithms for assessment of allergenicity [77–80]. These tools need to be further developed and, in particular, these prediction tools need to be thoroughly tested with known allergens and additional clinical data. In the meantime, users of these algorithms need to exercise care in interpretation of prediction results and test them with experimental validation. Standards for identification of allergenicity are not well established and in the meantime the Food and Agriculture Organization of the United Nations/World Health Organization recommendations are in

place. These recommendations are not perfect and exceptions to the rules are known. By pooling allergy data and tools we can expect the refinement of these recommendations, which will involve more sophisticated informatics methods. Bioinformatics already plays an important role in improved detection and characterization of food allergens [81], as well as improvement of clinical diagnosis and development of novel therapies or prevention measures [82]. However, the full realization of this potential will depend on the development of more advanced allergen-specific databases as well as improved methods for data mining of these databases.

Autoimmunity informatics

Bioinformatics applications in autoimmunity are at an early stage. Factors involved in autoimmune disease include complex genetic, clinical and immunologic traits [83]. In addition, the reporting rate of autoimmunity studies in journals is relatively low [84], and autoimmunity databases are rare and small in size [83]. Nevertheless, bioinformatics infrastructure for complex disease databases is available [85], and patient registries already

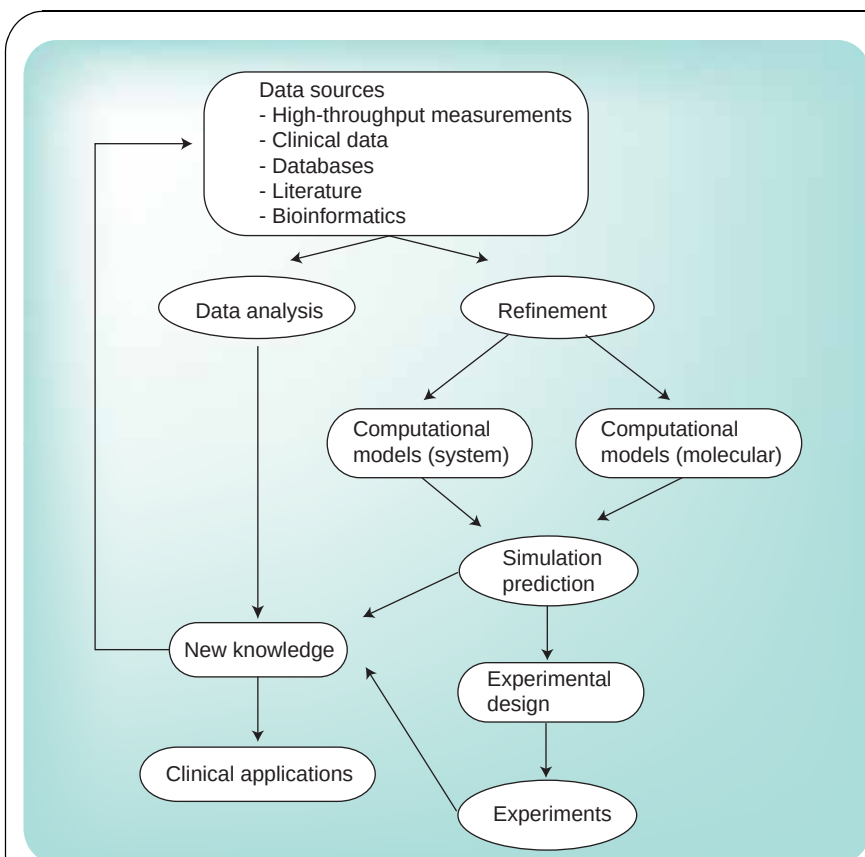


Figure 3. The immunomics paradigm. The existing data sources are analyzed and used to build and subsequently refine computational models for simulation of observed systems or processes. Simulations produce new knowledge or are used for hypothesis generation and subsequent experimental design to test the hypothesis. The experiments can be either large scale measurements or targeted experiments to test a hypothesis. Data analysis, computational simulations and experiments collectively produce new knowledge that is returned to enrich data sources and develop clinical applications in the areas of diagnosis, therapy or prevention.

contain large amounts of clinical data [86]. Combination of this clinical data with modern high-throughput screening technologies offer hope for development of autoimmunity vaccines, an endeavor that will only be possible through application of bioinformatics techniques [12,75]. Specific examples of bioinformatics applications in autoimmunity research include structural design of inhibitors of protein-tyrosine phosphatase- β , a target for diabetes and obesity treatment [87], and computational identification of human thyroglobulin peptides that are pathogenic in an HLA-DR3 transgenic mouse model of autoimmune thyroiditis [88].

Expert opinion & five-year view

Increased application of high-throughput experimental techniques combined with advances in immunoinformatics will result in explosive growth of immunology, similar to the events that previously transformed genetics into genomics. The quest for deciphering the functions of genes and their products with the aim of reducing human disease has led to the creation of large quantities of data. On the other hand, automation of patient records and clinical management systems now provides access to a large quantity of clinical information. The development of computational infrastructure has been revolutionized through computer networking and connectivity of multiple data sources. Huge computational resources are available for sharing data and computing power through Grid computing infrastructure [52,89]. Computational analysis tools are also growing in number, sophistication and quality [90]. In addition to computational infrastructure, instrumentation is rapidly improving through miniaturization and automation, enabling ever-increasing throughput for sample analysis. Examples include immunoprofiling (hundreds of variables) of cells and molecules from small blood samples [91], highly sensitive immuno-PCR assays for detecting pathogens [92] and phage-display libraries combined with immunoassays [93], among others.

The authors' prediction is that the next 5 years will bring increased integration of basic and clinical immunology research, informatics infrastructure and the latest high-throughput

instrumentation advances, thereby transforming immunogenetics into immunomics. This will for the first time allow understanding of immune function at the system rather than just the cellular level. Basic and preclinical research will obviously progress faster than clinical applications, which generally lag by at least 10–15 years due to the extraordinary complexity and time-consuming nature of the clinical trial process. Vast amounts of data will be available for support of molecular immunology research and large numbers of targets for vaccine and immunotherapies will be identified. Large functional studies will start appearing and will be supported by laboratories that already have well-established molecular immunology departments and are ready for introducing large-scale genomics and proteomics techniques. The resulting large amounts of data will in turn require further development of immunoinformatics resources. These developments will permit for rapid growth of the knowledge base in basic immunology and enable the identification of genes and proteins that are involved in immune function, but are not yet characterized. Awareness of the novel technologic possibilities in the clinical environment will emerge. In the next few years we will see the widespread acceptance of the immunomics paradigm (FIGURE 3). Consequently, immunologic research will become increasingly based on integration of basic and clinical immunology foundations with large-scale screening using high-throughput instrumentation, targeted experiments and computational tools.

Clinical immunology will be among the major growth fields in medicine, and immunomics will ultimately deliver benefits to patients from improved diagnosis, prophylaxis and treatment of immune diseases.

Acknowledgements

Vladimir Brusic's work in immunoinformatics is funded in part by the US Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant No. 1 U19 AI56541 and Contract No. HHSN266200400085C.

Key issues

- Immunoinformatics, also known as computational immunology, represents computational methods and resources used in the study of immune function.
- Immunologic data are available in largely disconnected sources including molecular and genomic databases, specialist databases, medical and scientific literature, patient records, disease registries, tissue repositories and hospital databases.
- Data analysis tools used in immunoinformatics include those for database searching, sequence comparison, literature mining, statistical and comparative tools, models of molecular interactions and system-level models of immune processes.
- Online tools enable the analysis of the antibody and T-cell receptor sequences, and prediction of antigen processing and presentation: proteasomal cleavage, peptide binding to major histocompatibility complex or transporter associated with antigen processing, major histocompatibility complex supermotifs, promiscuous human leukocyte antigen-restricted T-cell epitopes and T-cell epitope hotspots.
- The main applications of immunoinformatics in the study of disease focus on pathogenesis of disease, dynamics of immune responses, computational vaccinology and diagnostic tools.
- Immunoinformatics plays a significant role in research, diagnosis, prevention and treatment of tumors and infectious diseases, and in computational vaccinology, while applications in the fields of allergy and autoimmunity are emerging.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest

- 1 Ehreth J. The global value of vaccination. *Vaccine* 21, 596–600 (2003).
- 2 Brusic V, August JT. The changing field of vaccine development in the genomics era. *Pharmacogenomics* 5(6), 597–600 (2004).
- 3 Knudsen TB, Kledal TN, Andersen O, Eugen-Olsen J, Kristiansen TB. Severe acute respiratory syndrome – a new coronavirus from the Chinese dragon's lair. *Scand. J. Immunol.* 58(3), 277–284 (2003).
- 4 Chan CH, Chan Y, Shieh HK *et al.* Phylogenetic analysis of influenza B virus in Taiwan, 1997 to 2001. *J. Microbiol. Immunol. Infect.* 37(3), 135–144 (2004).
- 5 Chopra I. Antibiotic resistance in *Staphylococcus aureus*: concerns, causes and cures. *Expert Rev. Anti-infect. Ther.* 1(1), 45–55 (2003).
- 6 Mansky LM. HIV mutagenesis and the evolution of antiretroviral drug resistance. *Drug Resist. Update* 5(6), 219–223 (2002).
- 7 Berzofsky JA, Terabe M, Oh S *et al.* Progress on new vaccine strategies for the immunotherapy and prevention of cancer. *J. Clin. Invest.* 113(11), 1515–1525 (2004).
- 8 Prescott SL, Jones CA. An update of immunotherapy for specific allergies. *Curr. Drug Targets Inflamm. Allergy* 1(1), 65–75 (2002).
- 9 Cope AP, Feldmann M. Emerging approaches for the therapy of autoimmune and chronic inflammatory disease. *Curr. Opin. Immunol.* 16(6), 780–786 (2004).
- 10 Zhang W, Yang H, Kong X *et al.* Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. *Nature Med.* 11(1), 56–62 (2005).
- 11 De Giovanni C, Nicoletti G, Landuzzi L *et al.* Immunoprevention of HER-2/neu transgenic mammary carcinoma through an interleukin-12-engineered allogeneic cell vaccine. *Cancer Res.* 64(11), 4001–4009 (2004).
- Demonstrates the complete protection of susceptible mice from mammary carcinoma by vaccination.
- 12 McDevitt H. Specific antigen vaccination to treat autoimmune disease. *Proc. Natl Acad. Sci. USA* 101(Suppl. 2), 14627–14630 (2004).
- 13 Brusic V, Petrovsky N. Immunoinformatics – the new kid in town. *Novartis Found. Symp.* 254, 3–13 (2003).
- 14 Rammensee HG. Immunoinformatics: bioinformatic strategies for better understanding of immune function. Introduction. *Novartis Found. Symp.* 254, 1–2 (2003).
- 15 Brusic V, Zeleznikow J, Petrovsky N. Molecular immunology databases and data repositories. *J. Immunol. Methods* 238(1–2), 17–28 (2000).
- 16 Galperin MY. The molecular biology database collection: 2005 update. *Nucleic Acids Res.* 33, D5–D24 (2005).
- 17 Brusic V, Millot M, Petrovsky N, Gendel SM, Gigonzac O, Stelman SJ. Allergen databases. *Allergy* 58(11), 1093–1100 (2003).
- 18 Haux R, Ammenwerth E, Herzog W, Knaup P. Healthcare in the information society. A prognosis for the year 2013. *Int. J. Med. Inform.* 66(1–3), 3–21 (2002).
- Provides vision for the advanced use of patient records.
- 19 Hasman A, Safran C, Takeda H. Quality of healthcare: informatics foundations. *Methods Inf. Med.* 42(5), 509–518 (2003).
- 20 McDonald S. Indigenous transplant outcomes in Australia: what the ANZDATA registry tells us. *Nephrology (Carlton)* 9(Suppl. 4), S138–S143 (2004).
- 21 DiMichele DM, Kroner BL. The North American Immune Tolerance Registry: practices, outcomes, outcome predictors. *Thromb. Haemost.* 87(1), 52–57 (2002).
- 22 Bencsath M, Blaskovits A, Borvendeg J. Biomolecular cytokine therapy. *Pathol. Oncol. Res.* 9(1), 24–29 (2003).
- 23 Girardi M, Knobler R, Edelson R. Selective immunotherapy through extracorporeal photochemotherapy: yesterday, today, and tomorrow. *Hematol. Oncol. Clin. North Am.* 17(6), 1391–1403 (2003).
- 24 Merrill SJ. Computational models in immunological methods: an historical review. *J. Immunol. Methods* 216(1–2), 69–92 (1998).
- 25 Eskin E, Agichtein E. Combining text mining and sequence analysis to discover protein functional regions. *Pac. Symp. Biocomput.* 288–299 (2004).
- 26 Glenisson P, Coessens B, Van Vooren S, Mathys J, Moreau Y, De Moor B. TXTGate: profiling gene groups with text-based information. *Genome Biol.* 5(6), R43 (2004).
- 27 Karopka T, Scheel T, Bansemer S, Glass A. Automatic construction of gene relation networks using text mining and gene expression data. *Med. Inform. Internet Med.* 29(2), 169–183 (2004).
- 28 Brusic V, Petrovsky N, Gendel SM, Millot M, Gigonzac O, Stelman SJ. Computational tools for the study of allergens. *Allergy* 58(11), 1083–1092 (2003).
- 29 Lefranc MP. IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunogenetics and immunoinformatics. *Mol. Immunol.* 40(10), 647–660 (2004).
- 30 Lefranc MP, Giudicelli V, Kaas Q *et al.* IMGT, the international ImmunoGeneTics information system®. *Nucleic Acids Res.* 33, D593–D597 (2005).
- Provides databases, online tools and internet resources for the sequences, genes and structures of the immunoglobulins, T-cell receptors and major histocompatibility complex, and related proteins of the immune system.
- 31 Giudicelli V, Chaume D, Lefranc MP. IMGT/V-QUEST, an integrated software for immunoglobulin and T-cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res.* 32, W435–W440 (2004).
- 32 Monod MY, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T-cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics* 20, I379–I385 (2004).
- 33 Kaas Q, Ruiz M, Lefranc MP. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T-cell receptor and MHC structural data. *Nucleic Acids Res.* 32, D208–D210 (2004).
- 34 Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes – a framework for modelling, testing, and applications. *Methods* 34(4), 436–443 (2004).
- 35 Hakenberg J, Nussbaum AK, Schild H *et al.* MAPPP: MHC class I antigenic peptide processing prediction. *Appl. Bioinformatics* 2(3), 155–158 (2003).
- 36 Saxova P, Buus S, Brunak S, Kesmir C. Predicting proteasomal cleavage sites: a comparison of available methods. *Int. Immunol.* 15(7), 781–787 (2003).
- 37 Doytchinova I, Hemsley S, Flower DR. Transporter associated with antigen processing preselection of peptides binding to the MHC: a bioinformatic evaluation. *J. Immunol.* 173(11), 6813–6819 (2004).
- 38 Petrovsky N, Brusic V. Virtual models of the HLA class I antigen processing pathway. *Methods* 34(4), 429–435 (2004).

- 39 Bian H, Hammer J. Discovery of promiscuous HLA-II-restricted T-cell epitopes with TEPITOPE. *Methods* 34(4), 468–475 (2004).
- 40 De Groot AS, Jesdale B, Martin W *et al.* Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. *Vaccine* 21(27–30), 4486–4504 (2003).
- 41 Doytchinova IA, Guan P, Flower DR. Quantitative structure-activity relationships and the prediction of MHC supermotifs. *Methods* 34(4), 444–453 (2004).
- 42 Srinivasan KN, Zhang GL, Khan AM, August JT, Brusic V. Prediction of class I T-cell epitopes: evidence of presence of immunological hot spots inside antigens. *Bioinformatics* 20(Suppl. 1), I297–I302 (2004).
- Reduces complexity of screening T-cell epitopes by focusing on hotspots – regions of antigens that contain a high concentration of supertype-promiscuous peptides.
- 43 Chao DL, Davenport MP, Forrest S, Perelson AS. A stochastic model of cytotoxic T-cell responses. *J. Theor. Biol.* 228(2), 227–240 (2004).
- 44 Borghans JAM, Beltman JB, De Boer RJ. MHC polymorphism under host–pathogen coevolution. *Immunogenetics* 55, 732–739 (2004).
- 45 Shahaf G, Allman D, Cancro MP, Mehr R. Screening of alternative models for transitional B-cell maturation. *Int. Immunol.* 16(8), 1081–1090 (2004).
- 46 Bernaschi M, Castiglione F. Design and implementation of an immune system simulator. *Comput. Biol. Med.* 31(5), 303–331 (2001).
- Describes a comprehensive simulator of both humoral and cellular immune responses.
- 47 Yates A, Chan CC, Callard RE, George AJ, Stark J. An approach to modelling in immunology. *Brief Bioinform.* 2(3), 245–257 (2001).
- 48 Castiglione F, Poccia F, D’Offizi G, Bernaschi M. Mutation, fitness, viral diversity, and predictive markers of disease progression in a computational model of HIV Type 1 infection. *AIDS Res. Hum. Retroviruses* 20(12), 1314–1323 (2004).
- 49 Schonbach C, Kowalski-Saunders P, Brusic V. Data warehousing in molecular biology. *Brief Bioinform.* 1(2), 190–198 (2000).
- 50 Hu H, Brzeski H, Hutchins J *et al.* Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. *Pharmacogenomics* 5(7), 933–941 (2004).
- 51 Sanchez W, Gilman B, Kher M, Lagou S, Covitz P. *caGRID White Paper (cancer biomedical informatics grid prototype project)*. National Cancer Institute Center for Bioinformatics (NCICB), USA. July 23 (2004).
- 52 Koenig-Hoffmann K, Bonin-Debs AL, Boche I *et al.* High throughput functional genomics: identification of novel genes with tumor suppressor phenotypes. *Int. J. Cancer* 20, 113(3), 434–439 (2005).
- 53 Oh P, Li Y, Yu J *et al.* Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* 429(6992), 629–635 (2004).
- 54 Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin. Cancer Res.* 10(21), 7252–7259 (2004).
- 55 Glynne RJ, Watson SR. The immune system and gene expression microarrays – new answers to old questions. *J. Pathol.* 195(1), 20–30 (2001).
- 56 Novellino L, Castelli C, Parmiani G. A listing of human tumor antigens recognized by T-cells: March 2004 update. *Cancer Immunol. Immunother.* 54(3), 187–207 (2005).
- 57 Dalglish AG. Cancer vaccines as a therapeutic strategy. *Expert Rev. Vaccines* 3(6), 665–668 (2004).
- 58 Motta S, Brusic V. Mathematical modelling of the immune system. In: *Modelling in Molecular Biology*. Ciobanu G, Rozenberg G (Eds), Natural Computing Series, Springer Verlag, Germany, 193–218 (2004).
- 59 Persson L, Bartlett M. Notifiable Diseases Database system: review and development strategy. *NSW Public Health Bull.* 15(1–2), 10–12 (2004).
- 60 Ferdinand S, Valetudie G, Sola C, Rastogi N. Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. *Res. Microbiol.* 155, 647–654 (2004).
- 61 Kiechle FL, Zhang X, Holland-Staley CA. The -omics era and its impact. *Arch. Pathol. Lab. Med.* 128, 1337–1345 (2004).
- 62 Bodrossy L, Sessitsch A. Oligonucleotide microarrays in microbial diagnostics. *Curr. Opin. Microbiol.* 7(3), 245–254 (2004).
- 63 Mora M, Veggi D, Santini L, Pizza M, Rappuoli R. Reverse vaccinology. *Drug Discov. Today* 8(10), 459–464 (2003).
- 64 De Groot AS, Berzofsky JA. From genome to vaccine – new immunoinformatics tools for vaccine design. *Methods* 34(4), 425–428 (2004).
- 65 Kellam P, Holzerlandt R, Gramoustianou E, Jenner R, Kwan A. Viral bioinformatics: computational views of host and pathogen. *Novartis Found. Symp.* 254, 234–247 (2003).
- Comprehensive study utilizing bioinformatics to analyze both virus and host.
- 66 Wodarz D. Mathematical models of HIV and the immune system. *Novartis Found. Symp.* 254, 193–207 (2003).
- 67 Marino S, Kirschner DE. The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *J. Theor. Biol.* 227(4), 463–486 (2004).
- 68 Young K, Frodsham A, Doumbo OK *et al.* Inverse associations of human leukocyte antigen and malaria parasite types in two West African populations. *Infect. Immun.* 73(2), 953–955 (2005).
- 69 Holcombe M. Computational models of cells and tissues: machines, agents and fungal infection. *Brief Bioinform.* 2(3), 271–278 (2001).
- 70 Davies DH, Liang X, Hernandez JE *et al.* Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proc. Natl Acad. Sci. USA* 102(3), 547–552 (2005).
- 71 Rota PA, Oberste MS, Monroe SS *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300(5624), 1394–1399 (2003).
- Bioinformatics enabled rapid characterization of a highly pathogenic strain of coronavirus family.
- 72 Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. *Trends Biotech.* 23(2), 84–91 (2005).
- 73 De Groot AS, Rappuoli R. Genome-derived vaccines. *Expert Rev. Vaccines* 3(1), 59–76 (2004).
- 74 Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH, Wilson S. A roadmap for the immunomics of category A-C pathogens. *Immunity* (2005) (In Press).
- First global effort for systematic mapping of T- and B-cell epitopes including complete genome mapping and organization of the results in a database.
- 75 Locher CP, Soong NW, Whalen RG, Punnonen J. Development of novel vaccines using DNA shuffling and screening strategies. *Curr. Opin. Mol. Ther.* 6(1), 34–39 (2004).

- 76 Petrovsky N, Silva D, Brusic V. The future for computational modelling and prediction systems in clinical immunology. *Novartis Found. Symp.* 254, 23–32 (2003).
- 77 Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J.* 17(9), 1141–1143 (2003).
- 78 Soeria-Atmadja D, Zorzet A, Gustafsson MG, Hammerling U. Statistical evaluation of local alignment features predicting allergenicity using supervised classification algorithms. *Int. Arch. Allergy Immunol.* 133(2), 101–112 (2004).
- 79 Li KB, Issac P, Krishnan A. Predicting allergenic proteins using wavelet transform. *Bioinformatics* 20(16), 2572–2578 (2004).
- 80 Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, van Ham RC. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 5(1), 133 (2004).
- Allergenicity prediction tool compliant with Food and Agriculture Organization of the United Nations/World Health Organization recommendations.
- 81 Gendel SM. Bioinformatics and food allergens. *J. AOAC Int.* 87(6), 1417–1422 (2004).
- 82 Niederberger V, Valenta R. Genetically modified allergens. *Immunol. Allergy Clin. North Am.* 24(4), 727–738 (2004).
- 83 Antonelli A, Tuomi T, Nannipieri M *et al.* Autoimmunity to CD38 and GAD in Type I and Type II diabetes: CD38 and HLA genotypes and clinical phenotypes. *Diabetologia* 45(9), 1298–1306 (2002).
- 84 Royle P, Bain L, Waugh N. Systematic reviews of epidemiology in diabetes: finding the evidence. *BMC Med. Res. Methodol.* 5(1), 2 (2005).
- 85 Burren OS, Healy BC, Lam AC *et al.* Development of an integrated genome informatics, data management and workflow infrastructure: a toolbox for the study of complex disease genetics. *Hum. Genomics* 1(2), 98–109 (2004).
- 86 Weinstock-Guttman B, Jacobs LD, Brownschidle CM *et al.* Multiple sclerosis characteristics in African American patients in the New York State Multiple Sclerosis Consortium. *Mult. Scler.* 9(3), 293–298 (2003).
- 87 Lund IK, Andersen HS, Iversen LF *et al.* Structure-based design of selective and potent inhibitors of protein-tyrosine phosphatase beta. *J. Biol. Chem.* 279(23), 24226–24235 (2004).
- 88 Flynn JC, McCormick DJ, Brusic V *et al.* Pathogenic human thyroglobulin peptides in HLA-DR3 transgenic mouse model of autoimmune thyroiditis. *Cell. Immunol.* 229(2), 79–85 (2004).
- 89 Teo YM, Wang X, Ng YK. GLAD: a system for developing and deploying large-scale bioinformatics grid. *Bioinformatics* (2005) (In Press).
- 90 Srinivasan KN, Brusic V, August JT. New technologies for vaccine development. *Drug Dev. Res.* 62, 383–392 (2004).
- 91 Kantor AB, Alters SE, Cheal K, Dietz LJ. Immune systems biology: immunoprofiling of cells and molecules. *Biotechniques* 36(3), 520–524 (2004).
- Describes enabling technology that permits large amounts of immunologic profiling measurements.
- 92 Liang H, Cordova SE, Kieft TL, Rogelj S. A highly sensitive immuno-PCR assay for detecting Group A Streptococcus. *J. Immunol. Methods* 279(1–2), 101–110 (2003).
- 93 Liu F, Guttikonda S, Suresh MR. Bispecific monoclonal antibodies against a viral and an enzyme: utilities in ultrasensitive virus ELISA and phage display technology. *J. Immunol. Methods* 274(1–2), 115–127 (2003).
- 94 Pappalardo F, Lollini PL, Castiglione F, Motta S. Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics* (2005) (In Press).
- 102 IMGT Bloc-notes
<http://imgt.cines.fr>
(Accessed April 2005)
- 103 International Federation of Pharmaceutical Manufacturers and Associations. Joint position on the disclosure of clinical trial information via clinical trial registries and databases. January 2005
www.efpia.org/4_pos/sci_regu/Clinicaltrials2005.pdf
(Accessed April 2005)
- 104 SYFPEITHI
www.syfpeithi.de
(Accessed April 2005)
- 105 PREDBALB/c
research.i2r.a-star.edu.sg/predBalbc
(Accessed April 2005)
- 106 RANKPEP
mif.dfci.harvard.edu/Tools/rankpep.html
(Accessed April 2005)
- 107 BIMAS
www.bimas.cit.nih.gov/molbio/hla_bind
(Accessed April 2005)
- 108 Allergen Nomenclature
www.allergen.org
(Accessed April 2005)
- 109 ExPASy Proteomics Server
www.expasy.org
(Accessed April 2005)

Affiliations

- Vladimir Brusic, PhD, MBA
Head, Knowledge Discovery Department,
Institute for Infocomm Research, 21 Heng Mui
Keng Terrace, 119613 Singapore;
, Honorary Professor, School of Land and Food
Sciences and Institute for Molecular Bioscience,
University of Queensland, Brisbane, Australia
Tel.: +65 96 212 415
Fax: +65 67 748 056
vladimir@i2r.a-star.edu.sg
- Nikolai Petrovsky, MD, PhD
Professor and Director, Department of
Diabetes and Endocrinology, Flinders Medical
Centre, Flinders Drive, Bedford Park, South
Australia 5042
Tel.: +61 882 044 572
Fax: +61 882 045 987
nikolai.petrovsky@flinders.edu.au

Websites

- 101 PubMed
www.ncbi.nlm.nih.gov/PubMed
(Accessed April 2005)