Bioinformatics services related to diagnosis of primary immunodeficiencies

Crina Samarghitean[a,b] and Mauno Vihinen[a,b]

[a]Institute of Medical Technology, University of Tampere and [b]Tampere University Hospital, Tampere, Finland

Correspondence to Mauno Vihinen, Institute of Medical Technology, FI-33014 University of Tampere, Finland
E-mail: mauno.vihinen@uta.fi

Current Opinion in Allergy and Clinical Immunology 2009, 9:531–536

Purpose of review

Most primary immunodeficiencies (PIDs) have overlapping signs and symptoms – presenting a challenge for diagnosis. The information available from the Internet for over 200 PIDs is scattered between numerous services and databases. Patient information has been collected in different patient registries. Several software tools have been developed in order to build the databases, expert systems and other information systems useful in diagnosis or prediction.

Recent findings

Previously released services have been significantly improved and some new bioinformatics tools have been developed to help in diagnosis, prediction, mutation analysis and classification of PIDs. Several national initiatives have been launched for centralized PID information services. The very latest additions are tools and approaches for PID candidate gene prioritization, systematic classification and a medical expert system to help in diagnosis.

Summary

Many bioinformatics tools for PIDs are already freely available over the Internet. We expect bioinformatics tools to further help healthcare professionals in diagnosis, analysis and prediction. Currently, most of the resources are stand-alone and thus their integration will be a challenge for the future. Another challenge is to develop terminologies, ontologies and standards to achieve semantic interoperability.

Keywords

diagnosis, disease classification, expert systems, mutation databases, patient registries, primary immunodeficiencies

Curr Opin Allergy Clin Immunol 9:531–536

1528-4050

Introduction

Primary immunodeficiencies (PIDs) may manifest a wide range of clinical symptoms [1]. Currently, more than 200 PIDs and some 170 PID-related genes are known, many of them reported during the last few years. Diagnosis of PIDs may be demanding due to symptoms and signs often being discreet, overlapping and variable in many PIDs. Also, the large number of PIDs makes the field difficult for nonexperts. Several databases, registries, knowledge bases, prediction tools and expert systems are available and can be helpful for the diagnosis of PIDs.

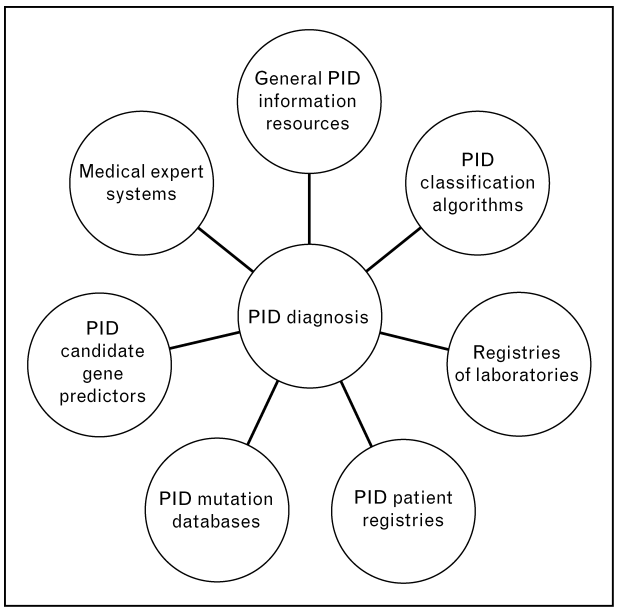
Tools helpful for PID diagnosis can be divided into seven categories (Fig. 1). First, general PID resources provide lots of information about the disorders on many levels ranging from genetics to protein structures, from disease models to interest groups, and so on. Second, classifications of PIDs are helpful for diagnosis as related diseases share many clinical features. Third, some resources list laboratories making genetic and clinical tests for PIDs. Fourth, there are national and international PID patient registries, which are complemented by mutation databases (fifth category), from which information can be used for comparing a case under study to those previously reported. Sixth, bioinformatics tools are also available for prediction or prioritization of novel PID candidate genes and decision-making (seventh category) in PID diagnosis.

General knowledge services for primary immunodeficiencies

Dedicated PID information resources include the ImmunoDeficiency Resource (IDR) and INFO4PI (Table 1) [2••]. These contain tailored information for different user groups such as patients and their families, researchers and healthcare professionals.

The IDR knowledge base is the most comprehensive PID information service available [3]. It integrates a wide spectrum of information, including clinical, biochemical, genetic, genomic, proteomic, structural and computational data. The IDR provides structured, systematic

Figure 1 Schematic grouping of bioinformatics resources and tools providing information about primary immunodeficiencies



PID, primary immunodeficiency.

and validated information for about 170 PIDs for doctors, researchers, students, nurses and patients. The IDR has been integrated with several internal and external databases and services. The service has a user-friendly and versatile search function. There are several introductory texts, including published classifications and diagnostic criteria, protocols and guidelines for PIDs.

Fact files form the core of the IDR – they integrate biomedical knowledge from several heterogeneous sources. Distributed information includes disorders, genes, mutations, protein sequences and structures, model organisms, online resources and patient, nurse and professional organizations and associations. Inherited Disease Markup Language (IDML) was developed for the fact files to provide a standardized method for exchanging genetic and clinical data, and links to other related resources [4].

INFO4PI is designed for diverse user groups. It is an official webpage of the Jeffrey Modell Foundation (JMF). The medical advisory group of the foundation has developed '10 warning signs for PIDs' and 'four stages for immunologic testing', which have been adopted in many

Table 1 Resources for primary immunodeficiency-related information

Bioinformatics services	URL
General PID resources	
IDR	http://bioinf.uta.fi/idr
INFO4PI	http://www.info4pi.org
JMF	http://www.jmfworld.com
ORPHANET	http://www.orpha.net
PID classifications	
AAAAI	http://www.aaaai.org/professionals/resources/pdf/immunodeficiency2005.pdf
ESID/IUIS	http://www.esid.org/downloads/ESID_Diseases_2009_0.pdf
IDR	http://bioinf.uta.fi/xml/idr/classification.xml
Novel mathematical classification	http://bioinf.uta.fi/PID_classification
WHO (ICD10)	http://apps.who.int/classifications/apps/icd/icd10online/
Registries of diagnostic laboratories	
EDDNAL	http://www.eddnal.com
GeneTests	http://www.ncbi.nlm.nih.gov/sites/GeneTests
IDdiagnostics	http://bioinf.uta.fi/IDdiagnostics
Patient registries	
ASCIA	http://www.immunodeficiency.org.au
CEREDIH	http://www.ceredih.fr/
ESID	http://www.esid.org/
IPINET	http://www.aieop.org
USIDnet	http://www.usidnet.org
Mutation databases	
IDbases	http://bioinf.uta.fi/IDbases
Other PID mutation databases	http://bioinf.uta.fi/base_root/mutation_databases_list2.php
RAPID	http://rapid.rcai.riken.jp/RAPID
Novel PID candidate genes	
PID candidates	[2••]
RAPID	http://rapid.rcai.riken.jp/RAPID/SVM
PID diagnosis tools	
PIDexpert	http://bioinf.uta.fi/PIDexpert
UKPIN	http://www.ukpin.org.uk/ESID/index.htm

EDDNAL, European Directory of DNA Diagnostic Laboratories; ESID, European Society for Immunodeficiencies; ICD, International Classification of Diseases; IDR, ImmunoDeficiency Resource; IPINET, Italian Primary Immunodeficiency Network; IUIS, International Union of Immunological Societies; JMF, Jeffrey Modell Foundation; PID, primary immunodeficiency; RAPID, Resource of Asian Primary Immunodeficiency Diseases; UKPIN, United Kingdom Primary Immunodeficiency Network.

countries. The service also includes information for over 50 diagnostic and research centers worldwide and a registry of experts worldwide.

ORPHANET provides information on rare diseases for healthcare professionals, patients and their relatives. ORPHANET includes expert-authored and peer-reviewed information for rare, mostly genetic, diseases. There is a directory of clinics, clinical laboratories, research activities and patient organizations. ORPHANET provides up-to-date information about rare diseases in many languages.

Classification of primary immunodeficiencies

Traditionally, PIDs have been classified into wide categories. The PID Classification Committee of the International Union of Immunological Societies (IUIS) has provided classifications for many years. The essential textbook in the field, Primary Immunodeficiency Diseases [5], contains another grouping of PIDs as well as the International Classification of Diseases (ICD10). The classification in the European Society for Immunodeficiencies (ESID) online patient database further expands the IUIS classification.

As PIDs are highly variable and there are already some 200 diseases, it would be beneficial to use computational approaches to organize the highly multidimensional space of signs, symptoms and other parameters. Consensus of at least five clustering and network analysis methods provided a statistically supported novel classification with novel features and relationships of PIDs (Samarghitean C, Ortutay C, Vihinen M, unpublished data). Many independent properties of the diseases and affected proteins showed good agreement with the classification, including severity and therapy of the diseases, functional classification of the proteins and protein interaction network properties. The results can also be used in developing medical expert systems (MESs).

Registries of diagnostic laboratories

The diagnosis of immunodeficiencies can be difficult because several disorders may have similar symptoms. In some cases, early and reliable diagnosis is crucial for efficient treatment because delayed diagnosis and management can lead to severe and irreversible complications.

For many PIDs, the definitive diagnosis can be obtained only based on both genetic and clinical tests. The physical signs may be nonspecific, very discreet or absent. Due to the rareness of PIDs, there may not be many laboratories analyzing a particular disease. IDdiagnostics has two registries, for PID gene testing and for clinical test laboratories [6]. There are contact addresses for laboratories along with the information for assay method(s) used. For genetic testing, there are details on the turnaround time, how often the samples are run and how many samples are studied annually. No samples should be sent without prior contact with the laboratory.

Some PID laboratory information can be found from general services for diagnostic laboratories. The GeneTests service provides an international directory of genetic testing laboratories and genetic and prenatal diagnosis clinics. There are also expert-authored, peer-reviewed disease descriptions called GeneReviews. The European Directory of DNA Diagnostic Laboratories (EDDNAL) has information about DNA-based diagnostic services. IDR fact files and IDdiagnostics both have cross-references to GeneTests and EDDNAL.

Patient registries

Several PID patient registries have been released in different countries. The information in them can help in making a diagnosis for a patient. The ESID patient registry is the largest [7••]. In July 2009, it contained 10,003 patient entries from 63 centers for 154 PIDs. For each PID, there is an individual database; however, they all share a common core dataset with information about diagnosis, therapy, quality of life and some laboratory data. There are also disease-specific data models for some of the most prevalent PIDs. Submission of mutation data to the patient registry is combined with Immunodeficiency mutation databases (IDbases) [8••]. Access to the data is allowed only to the registered users who have obtained permission.

The same software is used by the United States Immunodeficiency Network (USIDnet) for registry of US patients. Currently, the USIDnet data are collected for eight PIDs, but soon there will be information for over 30 disorders. Another regional registry using the ESID online database system is the Latin American Society for Primary Immunodeficiency Diseases (LASID), which so far includes information for 3321 patients in 14 countries.

As the national and international PID patient registries have been recently reviewed [9], we just briefly mention some in here. The Australian Society of Clinical Immunology and Allergy (ASCIA) PID register of Australia and New Zealand has already described the prevalence of PIDs in Australia for 1209 patients in 88 centers [10]. The Italian primary immunodeficiency network (IPINET) collects patient information, including the pedigree, date of diagnosis, immunological data and clinical manifestations, laboratory data and information about

replacement therapy [11]. The Spanish registry for primary immunodeficiencies (REDIP) has quite similar contents [12]. In 2004, the Spanish registry counted 2607 cases from 82 centers. The Iranian PID registry has followed 930 patients over a period of 30 years [13].

Other important PID patient-related and treatment-related databases include those by the Center for International Blood and Marrow Transplant Research (CIBMTR) and the European Blood and Marrow Transplant Group (EBMT). These facilitate research into allogeneic hematopoietic stem cell transplantation outcomes [14]. Stem cell transplantation for immunodeficiency in Europe registry (SCETIDE) is a specific stem cell transplantation database for PIDs in Europe, which has data going back to 1985. These databases allow addressing questions difficult to answer through clinical trials and may also aid the development of optimal designs for prospective clinical trials [15].

Mutation databases

Although many identified genetic variations in PIDs are novel, it is important to analyze whether the variation in a patient has been previously described. Information about PID-related genetic variations is available from different mutation databases. The majority of these registries are in the *IDbases* maintained at the IMT Bioinformatics group. Currently, there are 122 freely available IDbases, which contain mutation information for 5359 patients from 4489 families having altogether 2322 different mutational events. In many IDbases, there is plenty of information in addition to the description of the actual mutation. The mutation entries are linked to sequence databanks, literature and Online Mendelian Inheritance in Man (OMIM) [16]. The format of IDbases is standardized and uniform throughout.

Currently, some important changes are being made to IDbases. Reference sequences at three levels (DNA, RNA and protein) have been developed in collaboration with RefSeqGene (http://www.ncbi.nlm.nih.gov/RefSeq/) and LRG (Locus, Reference, Genomic) projects (http://www.lrg-sequence.org). Thereby, IDbases will have a stable genomic framework for reporting mutations that allows easy integration with other services. The IDbases can provide new insights into both genotype–phenotype correlations in patients and protein structure–function relationships for the encoded proteins. The IDbases are linked to the University of California Santa Cruz (UCSC) genome browser [17] from where the mutation data can be easily viewed with PhenCode [18], along with other genetic and variation information. Another change ongoing in IDbases is conversion to a relational database model, which has been jointly developed with the GEN2-PHEN consortium (http://www.gen2phen.org/).

Several PID mutation databases are maintained also in other laboratories (Table 1). The coverage and depth of details vary in these registries. Currently, there exists a locus-specific mutation database for almost all PIDs in which the gene defects are known.

The mutation information in the Resource of Asian Primary Immunodeficiency Diseases (RAPID database) is mainly a copy of data in IDbases and IDR. For PID genes and proteins, there is information about mRNA and protein expression as well as on protein–protein interactions and mouse studies. A tool can visualize mutations on protein three-dimensional structures [19].

Although identification of gene defects and variations has become easy and fast to perform, the interpretation of the effects and elucidation of the detailed molecular mechanisms of genetic diseases is much more difficult. Disease-related alterations may have diverse effects on the structure and function at DNA, RNA and protein levels [20•]. Numerous methods can be used for predicting the effects of amino acid substitutions and are collected in the recently developed Pathogenic-Or-Not pipeline (PON-P), which is freely available at http://bioinf.uta.fi/PON-P [20•].

Novel primary immunodeficiency candidate genes

A bioinformatics approach has been applied to predict novel PID candidate genes [2••]. A total of 26 putative PID genes were prioritized. The method combines information about protein interaction network properties and Gene Ontology terms. The analysis was based on a dataset for the immunome, the entirety of genes and proteins essential for mounting immune responses [21]. The approach utilizes the protein interaction network information available in the Immunome Knowledge Base (IKB) [22]. The identified disease gene candidates are mainly involved in cellular signaling, including receptors, protein kinases and adaptors and binding proteins as well as enzymes [2••].

Another PID candidate list of altogether 1442 genes is available from RAPID. Currently, no details for the support vector machine approach are available. The error rate is stated to be 2.14%.

Primary immunodeficiency diagnosis tools

Some services for PID diagnosis have been implemented on the web. The multistage diagnostic protocol designed for nonimmunologists [23] has been converted to linked webpages on the United Kingdom Primary Immunodeficiency Network (UKPIN) site. Recommendations for diagnosis and treatment for some PIDs, both in

English and in Italian, are provided by IPINET. In the Centre de Référence Déficits Immunitaires Héréditaires (CEREDIH), the French national immunodeficiency center service, there is information about French PID diagnosis centers and laboratories. There are also '10 warning signs' list for children and '12 warning signs' list for adults. They have also a decision tree-like schema for diagnostic protocol. All the information in this service is written in French.

More advanced tools for diagnosis are called MESs. These are computer software systems that are based on a set of rules applied to knowledge originally extracted from human experts or generated by computational analyses. In addition to helping in diagnosis and report generation, MESs can improve consistency in decisions, as well as timeliness in decision-making and productivity [24•]. MESs can be integrated with other healthcare applications, such as electronic patient records, and systems for prescribing and dispensing medicines.

PIDexpert is, to our knowledge, the first MES for the diagnosis of PIDs (Samarghitean C, Iltanen K, Varpa K, *et al.*, unpublished results). PIDexpert generates a differential diagnosis from clinical symptoms and suggests potentially useful further clinical and laboratory information required for definitive diagnosis. The main components of the decision support system shell are a knowledge base, a query base, an inference engine and a graphical user interface. The query base and the knowledge base include data and facts from many sources, including the IDR, IDdiagnostics, IDbases, the ESID registry, national registries and literature. Medical experts were consulted for validation of the query and the knowledge base. The data are mainly informal and heuristic. Previously released diagnostic guidelines [25] and practice parameters [26] were coded into the system. These include guidelines for possible/probable/definitive diagnosis for some of the most common PIDs. The inference engine searches for the best fitting patterns among the PIDs. The system also identifies other conditions that might be associated with the disorder and suggests how the diagnosis could be confirmed. PIDexpert will soon be publicly available.

Future trends

Computers and sophisticated software tools are integral components of modern healthcare. Integration of heterogeneous information sources will further increase the interoperability and allow a more holistic view of the data. Standards, recommendations and ontologies will also contribute to the seamless data flow between applications. Many resources will be publicly available, except for patient databases, which cannot be made public due to confidentiality and security issues. It is important for the community to provide information and details to the bioinformatics services to guarantee their coverage, value and completeness. Curation and quality of the data should be secured by permanent infrastructure funding for the most central resources.

Conclusion

A large number of services are already available for the diverse data and information of PIDs. Online knowledge bases collect and distribute many kinds of information and keep users up-to-date with the deluge of data. Patient registries are vital components of any public healthcare program with an important role in decision-making. Databases for patients and mutations are essential for diagnosis. Resources for diagnosis laboratories are frequently used by those wanting to verify a suspected PID case. Classification of PIDs and knowledge of candidate genes have many applications. Expert systems can help in the diagnosis of PIDs and training of healthcare personnel. New methods, algorithms and ontologies are needed to fully exploit all the available data for PIDs.

Acknowledgements

We are grateful to the Finnish Academy, the Medical Research Fund of Tampere University Hospital, the Sigrid Juselius Foundation and Tampere City Hall for financial support.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (pp. 575–576).

1 Marodi L, Notarangelo LD. Immunological and genetic bases of new primary immunodeficiencies. Nat Rev Immunol 2007; 7:851–861.

2 •• Ortutay C, Vihinen M. Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies. Nucleic Acids Res 2009; 37:622–628.
The paper describes a bioinformatics approach applied to predict novel PID candidate genes.

3 Samarghitean C, Väliaho J, Vihinen M. IDR knowledge base for primary immunodeficiencies. Immunome Res 2007; 3:6.

4 Väliaho J, Riikonen P, Vihinen M. Distribution of immunodeficiency fact files with XML – from Web to WAP. BMC Med Inform Decis Mak 2005; 5:21.

5 Ochs HD, Smith CIE, Puck JM. Primary immunodeficiency diseases, a molecular and genetic approach. 2nd ed. Oxford: Oxford University Press; 2007.

6 Samarghitean C, Väliaho J, Vihinen M. Online registry of genetic and clinical immunodeficiency diagnostic laboratories, IDdiagnostics. J Clin Immunol 2004; 24:53–61.

7 •• Eades-Perner AM, Gathman B, Knerr V, *et al.* The European internet-based patient and research database for primary immunodeficiencies: results 2004–06. Clin Exp Immunol 2007; 147:306–312.
This publication presents the first results from the ESID online database.

8 •• Piirilä H, Väliaho J, Vihinen M. Immunodeficiency mutation databases (IDbases). Hum Mutat 2006; 27:1200–1208.
This publication introduces IDbases, their content and applications.

9 Knerr V, Grimbacher B. Primary immunodeficiency registries. Curr Opin Allergy Clin Immunol 2007; 7:475–480.

10 Kirkpatrick P, Riminton S. Primary immunodeficiency diseases in Australia and New Zealand. J Clin Immunol 2007; 27:517–524.

11 Plebani A, Soresina A, Notarangelo LD, *et al.* The Italian network of primary immunodeficiencies. Iran J Allergy Asthma Immunol 2004; 3:165–168.

12 Matamoros Flori N, Mila Llambi J, Espanol Boren T, *et al.* Primary immunodeficiency syndrome in Spain: first report of the national registry in children and adults. J Clin Immunol 1997; 17:333–339.

13 Rezaei N, Pourpak Z, Aghamohammadi A, *et al.* Frequency and clinical manifestations of patients with primary immunodeficiency disorders in Iran: update from the Iranian primary immunodeficiency registry. J Clin Immunol 2006; 26:519–532.

14 Horowitz M. The role of registries in facilitating clinical research in BMT: examples from the Center for International Blood and Marrow Transplant Research. Bone Marrow Transplant 2008; 42:S1–S2.

15 Griffith LM, Cowan MJ, Kohn DB, *et al.* Allogeneic hematopoietic cell transplantation for primary immune deficiency diseases: current status and critical needs. J Allergy Clin Immunol 2008; 122:1087–1096.

16 Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 2009; 37:D793–D796.

17 Karolchik D, Kuhn RM, Baertsch R, *et al.* The UCSC Genome Browser Database: 2008 update. Nucleic Acids Res 2008; 36:D773–D779.

18 Giardine B, Riemer C, Hefferon T, *et al.* PhenCode: connecting ENCODE data with mutations and phenotype. Hum Mutat 2007; 28:554–562.

19 Keerthikumar S, Raju R, Kandasamy K, *et al.* RAPID: Resource of Asian Primary Immunodeficiency Diseases. Nucleic Acids Res 2009; 37:D863–D867.

20 Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the • effects of missense mutations using bioinformatics methods. Hum Mutat 2009; 30:703–714.
The paper presents methods used for predicting the effects of genetic variations and introduces a new service, the PON-P.

21 Ortutay C, Siermala M, Vihinen M. Molecular characterization of the immune system: emergence of proteins, processes, and domains. Immunogenetics 2007; 59:333–348.

22 Ortutay C, Vihinen M. Immunome knowledge base (IKB): an integrated service for immunome research. BMC Immunol 2009; 10:3.

23 de Vries E, Clinical Working Party of the European Society for Immunodeficiencies (ESID). Patient-centred screening for primary immunodeficiency: a multistage diagnostic protocol designed for nonimmunologists. Clin Exp Immunol 2006; 145:204–214.

24 Samarghitean C, Vihinen M. Medical expert systems. Curr Bioinform 2008; • 3:56–65.
This review presents MESs, their benefits and limitations and introduces the PIDexpert model.

25 Conley ME, Notarangelo LD, Etzioni A. Diagnostic criteria for primary immunodeficiencies. Clin Immunol 1999; 93:190–197.

26 Bonilla FA, Bernstein IL, Khan DA, *et al.*, American Academy of Allergy, Asthma and Immunology; Joint Council of Allergy, Asthma and Immunology. Practice parameter for the diagnosis and management of primary immunodeficiency. Ann Allergy Asthma Immunol 2005; 94:S1–S63.