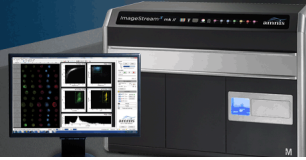


New ImageStream<sup>x</sup> Mark II

Cytometry without compromise



amnis<sup>®</sup>  
part of EMD Millipore



## Systematic Classification of Primary Immunodeficiencies Based on Clinical, Pathological, and Laboratory Parameters

This information is current as of July 13, 2012.

Crina Samarghitean, Csaba Ortutay and Mauno Vihinen

*J Immunol* 2009; 183:7569-7575; Prepublished online 16 November 2009;

doi: 10.4049/jimmunol.0901837

<http://www.jimmunol.org/content/183/11/7569>

---

**Supplementary Material** <http://www.jimmunol.org/content/suppl/2009/11/16/jimmunol.0901837.DC1.html>

**References** This article **cites 71 articles**, 15 of which you can access for free at: <http://www.jimmunol.org/content/183/11/7569.full#ref-list-1>

**Subscriptions** Information about subscribing to *The Journal of Immunology* is online at: <http://jimmunol.org/subscriptions>

**Permissions** Submit copyright permission requests at: <http://www.aai.org/ji/copyright.html>

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at: <http://jimmunol.org/cgi/alerts/etoc>

---

*The Journal of Immunology* is published twice each month by The American Association of Immunologists, Inc., 9650 Rockville Pike, Bethesda, MD 20814-3994. Copyright © 2009 by The American Association of Immunologists, Inc. All rights reserved. Print ISSN: 0022-1767 Online ISSN: 1550-6606.



# Systematic Classification of Primary Immunodeficiencies Based on Clinical, Pathological, and Laboratory Parameters<sup>1</sup>

Crina Samarghitean,<sup>\*†</sup> Csaba Ortutay,<sup>\*</sup> and Mauno Vihinen<sup>2\*†</sup>

**The classification of diseases has several important applications ranging from diagnosis and choice of treatment to demographics. To date, classifications have been successfully created manually, often within international consortia. Some groups of diseases, such as primary immunodeficiencies (PIDs), are especially hard to nosologically cluster due, on one hand, to the presence of a wide variety of disorders and, in contrast, because of overlapping characteristics. More than 200 PIDs affecting components of the innate and adaptive immune systems have been described. Clinical, pathological, and laboratory characteristics were collected and used to group PIDs. A consensus of at least five independent methods provided a novel classification of 11 groups, which revealed previously unknown features and relationships of PIDs. Comparison of the classification to independent features, including the severity and therapy of the diseases, functional classification of proteins, and network vulnerability, indicated a strong statistical support. The method can be applied to any group of diseases. *The Journal of Immunology*, 2009, 183: 7569–7575.**

**P** primary immunodeficiencies (PIDs)<sup>3</sup> are a large and heterogeneous group of disorders that have been organized manually into different categories (1–8) sometimes without a consensus. PIDs are mainly rare hereditary disorders of the immune system that often have serious consequences (1, 2). These diseases represent a challenge in their diagnosis and treatment due to overlapping symptoms and similarities between diseases. Infections are the hallmarks of PIDs (1, 9–11). A diagnosis will often be considered when infections are frequent or severe, resistant to standard therapies, or caused by unusual (opportunistic) organisms. Other manifestations include autoimmunity (12, 13) and cancer diseases (14), granulomatosis (15), hemophagocytic syndrome (16, 17), angioedema (18), autoinflammation (19–21), thrombotic microangiopathy, or predisposition to allergy.

Clinical descriptions have already been made for more than 200 PIDs (4, 7, 22), for which 167 genetic etiologies have been described. PIDs have historically been defined and classified according to immunological phenotypes (3). Before molecular analyses were widely available, PIDs were classified according to the affected immune function, as follows: Ab production (B cells), cellular immunity (T cells) or both (combined immunodeficiencies), phagocyte function (neutrophils, monocytes), and complement activation, etc. This classification has been useful for certain practical purposes, but not from the mechanistic point of view, because many PIDs do not easily fit into the scheme. On clinical grounds, immunodeficiencies can be classified into two broad groups ac-

ording to whether all features are the result of the immune defect (immunodeficiency syndromes) or whether many, even prominent ones, cannot be explained by the immune defect (syndromes with immunodeficiency).

The behavior of even the most complex of systems is based on the interaction of their components. These components can be reduced to a series of nodes that are connected to each other by links, which together form a network (23). Most real networks in technological, social, and biological systems share common designs that are simple and quantifiable. In medicine, network analysis has been used to characterize, e.g., the spread of epidemics (24), to determine ways to control them (25), and to identify novel target genes for prostate cancer (26).

When networks formed from diseases, the involved genes, and their phenotypes have been investigated (27–29), only a few PIDs have been included. Information about a protein interaction network for the immunome (30) has been used together with gene ontology terms (31) to predict novel PID candidate genes (32).

In this study, our goal was to develop a systematic, mathematical classification of PIDs. Previous PID classifications have been derived from observational correlations between pathological and clinical features. The foundation of the method is the description of the diseases based on 87 clinical and laboratory parameters.

Altogether, six methods belonging to two categories were used to organize the diseases based on the characteristics. Three clustering methods were applied to the multivariate problem to form disease groups in which members are most similar to each other (33).

The other three methods are from the emerging field of community analysis of networks. A community is a set of nodes with many edges (connections) inside the community and few edges outside it. Community analysis is a powerful tool for finding groups in interconnected entities (34–36); therefore, PIDs interpreted as a network can be analyzed this way, and diseases strongly associated with each other can be identified. Our systematic approach can be applied for classifying any other disease group.

## Materials and Methods

To obtain a systematic classification for PIDs, a novel approach was developed. The method applies advanced computational tools for clustering and network analysis. We used altogether six methods to group PIDs based on characteristics that they share. Three of the methods were for clustering, and three for network community analysis.

\*Institute of Medical Technology, University of Tampere, Tampere, Finland; and  
†Tampere University Hospital, Tampere, Finland

Received for publication June 9, 2009. Accepted for publication September 23, 2009.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> This work was supported by the Finnish Academy, the Sigrid Juselius Foundation, the Medical Research Fund of Tampere University Hospital, and Tampere City Hall.

<sup>2</sup> Address correspondence and reprint requests to Dr. Mauno Vihinen, Institute of Medical Technology, FI-33014 University of Tampere, Finland. E-mail address: mauno.vihinen@uta.fi

<sup>3</sup> Abbreviations used in this paper: PID, primary immunodeficiency; DC, dendritic cell; ICD, International Classification of Diseases; IDR, ImmunoDeficiency Resource.

Copyright © 2009 by The American Association of Immunologists, Inc. 0022-1767/09/\$2.00

### Immunodeficiencies and selection of the parameters

Data for PIDs were collected from the ImmunoDeficiency Resource (IDR) (4), IDdiagnostics (37), IDbases (38), and literature (1, 2, 5–7, 39–43). Only detailed reports with statistical information including clinical symptoms and measured laboratory values characterizing PIDs were included. When diseases without specific symptoms were excluded, there were altogether 194 PIDs left.

For each disease, all signs, symptoms, and laboratory values mentioned in the literature were collected. The initial list contained 420 parameters. Parameters characterizing only one to four diseases were omitted or merged to others.

For example, IgM lymphoma, which is according to the literature associated only with hyper-IgM syndrome type 2, was merged to the more general term of other malignancies. Similarly, encephalitis, meningoencephalitis, meningitis, conjunctivitis, iritis, episcleritis, and brain abscess were grouped to a more general term of CNS infections. Cerebellar ataxia, pathognomonic for ataxia-telangiectasia, and ataxia-telangiectasia-like disease were merged to neurological or CNS abnormalities together with other signs, such as peripheral neuropathy, speech delay, and convulsions. Finally, after iterative process, we had 87 informative parameters (supplemental Table S1).<sup>4</sup> All of the parameters had an equal weight in the analysis.

### Cluster and network community analysis

Cluster and network analyses were performed in the R statistical environment (44) using the *igraph* (45) and *cluster* program libraries. Three different variations of *K*-means clustering were used to analyze the dataset. The Clustering Large Applications (*clara*) method computes a list representing the clustering of the data into *k* clusters. Partitioning Around Medoids (*pam*) partitions (clusters) the data into *k* clusters around medoids, which are representative objects of a dataset from which the distances to the other points in the cluster are computed. The Fuzzy Analysis Clustering (*fanny*) method computes a partition grouping of the data into *k* clusters. The number of clusters was chosen by maximizing the average width of the clusters. In fuzzy clustering, data elements can belong to more than one cluster, and thus, each disease has a set of membership levels, which indicate the strength of association to each cluster.

Three methods were applied to find highly interconnected parts of the network. Community structure via short random walks is a walktrap community analysis, which searches for densely connected subgraphs, i.e., communities (34). When moving from one node to a connected one, short random walks tend to stay in the same community. The second method uses community structure detection based on the leading eigenvector of the community matrix (35). The method looks for densely connected subgraphs by calculating the leading nonnegative eigenvector of the modularity matrix of the graph. The third method tries to find communities in graphs via a spin-glass model and simulated annealing (36).

### Combination of clustering and network results

The data for PIDs are incomplete because in many diseases just a few, even a single, patient was known, and therefore, the most prominent signs were difficult to define. In rare diseases, some symptoms may occur frequently just by chance. Therefore, to obtain the most reliable and robust grouping and a consistent and robust view of the disease grouping patterns, a consensus classification based on the co-occurrence of the diseases in four, five, or six methods was generated. Using this approach, the results are expected to be independent from the biases of the individual methods.

### Statistics

To evaluate which of the binary parameters significantly supported the identified disease clusters, we tested whether a parameter had significantly different distribution in the individual clusters compared with the entire dataset by calculating *p* values using the hypergeometric distribution. The threshold for significance used was a *p* value 0.05. A similar evaluation was performed for the binary clinical and functional properties of the diseases and respective proteins.

### Correlation of disease clusters to clinical, genetic, functional, and network properties

The consensus network graph was used to visualize a number of properties for the diseases, and involved genes and proteins. Information about the prevalence of the PIDs was obtained from the IDR (4) and other sources (1, 2, 40, 46). When the prevalence was not known and only a few cases were

reported in literature, it was assumed to be  $<1/10^7$ . Data on inheritance were retrieved from the IDR and literature (1, 2, 6, 40). For treatment modalities, the most common treatments were listed for each disease (1, 2, 39). Functional classifications of the proteins in the immunome (47) were obtained from the Immunome Knowledge Base (48).

## Results

### PID network

Novel PID grouping was obtained by applying altogether six methods for clustering and network analysis. Our approach revealed associations that were not previously obvious and led to the identification of distinct novel groups. Fig. 1A shows the results when at least five methods agreed on the clustering. The diseases are indicated by the affected genes, when known. If all six methods are required to agree, the only difference is that the 11 major groups are divided further to smaller subgroups. There is one giant cluster that contains the majority of the PIDs and some separate clusters and singleton PIDs. Some details of the classification may change in the future when more information becomes available, yet still the major features will remain. In the consensus graph for the PIDs (Fig. 1A), 1,285 pairs of diseases are grouped together by at least five methods of a possible 18,721, and of the 12,721 that are grouped by at least one method. The results are also available in an interactive web page at [http://bioinf.uta.fi/PID\\_classification](http://bioinf.uta.fi/PID_classification).

### Disease clusters

The analytical approach combining six different and independent methods provides a highly robust classification when at least five of the methods were required to agree on the grouping details. The dendrogram reveals 11 well-defined disease clusters (DCs) of at least 4 diseases (Fig. 1A). First, we analyzed the nature of the PIDs in the clusters and then investigated the properties of the diseases and the corresponding genes and proteins in the clusters.

Most of the clusters are very homogeneous and contain related diseases. An overall view of the DCs shows that clusters III and VII contain (almost) exclusively SCIDs, whereas in group IX the diseases are related to the complement system and in DCs I and XI to phagocyte functions; in DC VIII are fever syndromes, and in DC X Fanconi anemias. All of the MHC II genes are in cluster III, and all of the known MHC I diseases are in cluster V. The classical complement pathway diseases are in clusters V and IX. Diseases in cluster II are related to DNA instability and DNA damage repair, except for G6PC3. DCs IV, V, and VI are the most heterogeneous. Cluster IV contains numerous receptor and signaling molecule-related diseases. Some of the proteins behind these disorders form transmembrane channels. DC V contains mainly Ab and complement deficiencies together with some SCIDs. The majority of the group VI diseases are related to phagocytosis and apoptosis. All of the groups have strong statistical support (supplemental Table S2).

These groups, which have been exclusively generated based on disease characteristics, indicate the power of the method. Other information for the PID genes, proteins, and their functions further support the classification (Fig. 1, B–E). Also, these results are statistically significant (supplemental Table S2).

The previous PID classifications have relied heavily on the cell types in which the disease-related genes are normally expressed. Thus, one of the major differences to these is that Ab deficiencies, combined PIDs, and diseases related to phagocytosis are widely scattered in our graph (Fig. 1). These diseases are very heterogeneous in their symptoms and signs, and affect numerous parts of the immune system. The highest concentration of SCIDs is in DCs II, IV, and VII, whereas Ab deficiencies mainly appear in clusters IV and V. Phagocyte diseases are exclusive to DCs I and XI, but also appear in DCs IV, V, and VI.

<sup>4</sup> The online version of this article contains supplemental material.

To test whether the etiologically rather homogeneous DCs shared any similarities, the distribution of characteristics describing independent clinical, functional, genetic, and network properties of the diseases and the corresponding genes and proteins were investigated. These features were not used for the original clustering of the PIDs.

The PIDs were divided into four groups according to their severity. Severity was not among the symptoms used in our classification because it is not routinely used in the clinical description of the diseases. SCID is a pediatric emergency situation because the condition is life threatening, whereas some of the other PIDs are just mild. In fact, the vast majority of PID cases have been thought to remain undiagnosed due to mild symptoms. The severity shows a very homogeneous pattern in some of the clusters (Fig. 1B). Diseases in DCs II, VII, and IX belong to two categories, whereas in almost all of the remaining clusters there are almost exclusively moderate and severe or severe and life-threatening diseases.

PIDs are treated in a number of ways that can be grouped to a small number of categories, including Ig treatment, the use of antibiotics, antifungals or antivirals, immunomodulators, and the reconstitution of the immune system by (haploidentical) bone marrow transplantation or hematopoietic stem cell transplantation. The data for the treatment of the diseases in Fig. 1C indicate that the majority of the clusters are very homogeneous in regard to treatments and there are clear differences between DCs. There are DCs in which all of the diseases are treated with the same battery of therapeutic modalities, and in almost all the clusters some of the therapeutic options can be used for all of the coclustered PIDs. Also, based on these results, the DCs reliably reflect the properties of diseases, and the therapy applied to diseases within DCs is usually similar.

The cellular functions have been determined for all the genes and proteins required in the immunome (entirety of immune system) (49). There are functional groups for, e.g., the surface receptors in clusters of differentiation classification, chemokines, and their receptors, humoral immunity proteins, and those involved in cellular immunity, Ag processing, and transcription factors. The distribution of the functional properties in the PID classification is shown in Fig. 1D.

The majority of the group I proteins are involved in inflammation and cellular immunity. Group II and VII proteins have functions as transcription factors involved in humoral and cellular immunity. Humoral immunity is most prevalent in DC IV, complement proteins in DC IX, and both humoral and complement functions in DC V. Inflammation is the function involved in DC VIII. Many of the proteins have several classifications because they are typically overlapping. Also, at this level, the grouped proteins and diseases share many common properties because the functions are very homogeneous within DCs and differ between them.

Vulnerability is a systems biology measure that indicates how crucial a certain node is for the network (30). Vulnerabilities of proteins in the immunome protein interaction network were color coded compared with the average vulnerability in the entire network (Fig. 1E). There are some clusters in which the vulnerabilities are related, especially those in DCs III, VI, and VII. The most vulnerable diseases are widely scattered throughout the network. Only some of the SCID proteins that are related with the most severe PIDs are highly vulnerable. Also, previously, the vast majority of disease genes were shown to be peripheral in the network (27).

Inheritance pattern (supplemental Fig. S1) and prevalence (supplemental Fig. S2) did not show any correlation within DCs at all.

This was expected because these characteristics are not likely to affect the etiology of diseases.

## Discussion

We developed a novel approach for nosology and produced a systematic classification of PIDs based on parameters across several clinical, pathological, and physiological dimensions. Our approach combines existing clustering and network partition methods to classify these diseases. The new classification shares certain features with previous groupings, yet is different in a number of details. For example, the new classification indicates that cell-type expression, which has previously been one of the major classification criteria, cannot be very reliably used for classification of PIDs.

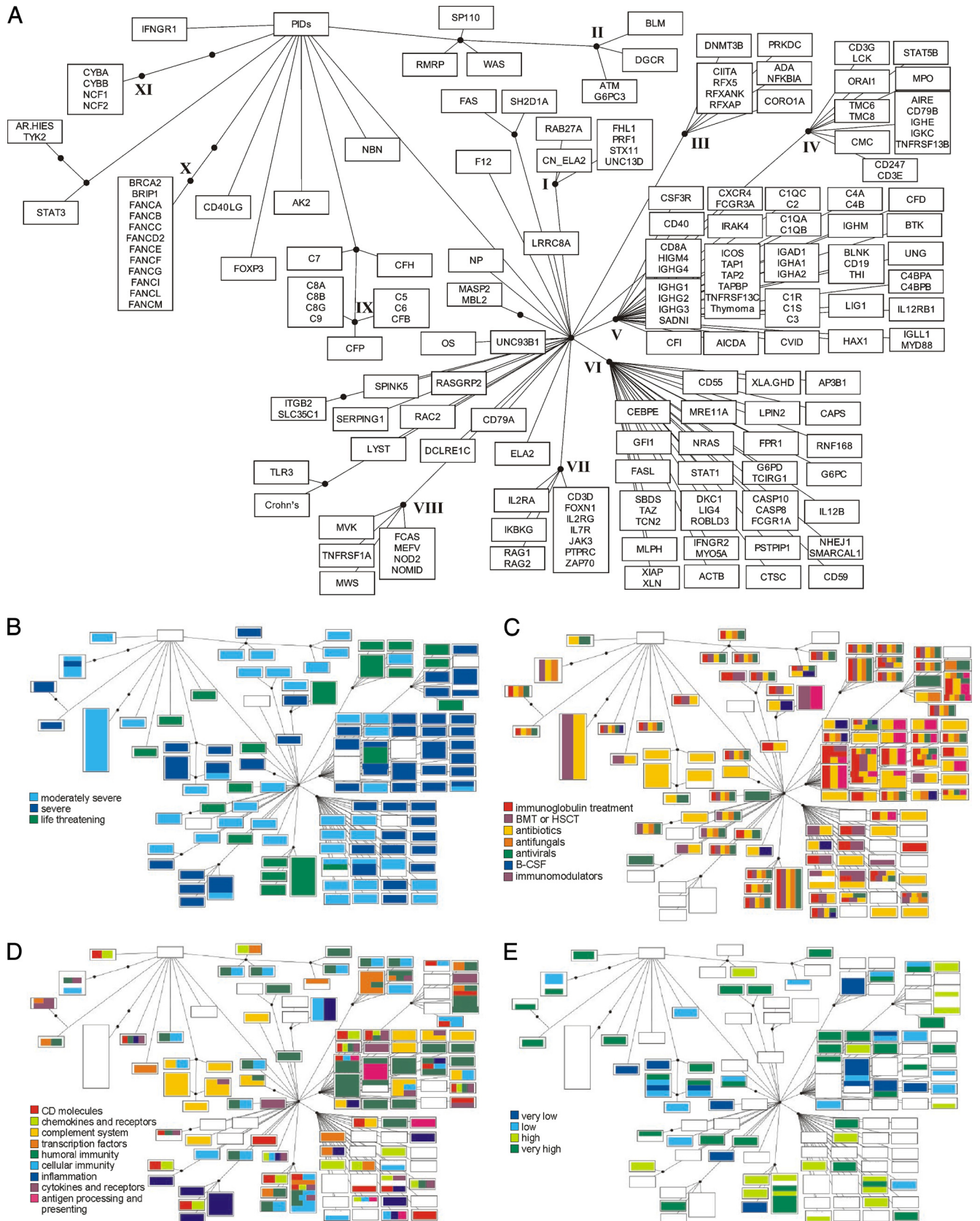
Clustering methods are widely used in many fields, such as in microarray data analysis. In medicine, cluster analysis has been used in the nosological splitting of different phenotypes, for example in Marshall and Stickler syndromes (50, 51), and more recently to classify patients with chronic pain (52) and to develop a new taxonomy for airway diseases (53, 54). Because the disease data were not complete due to many of the PIDs being extremely rare, we combined both network and clustering methods and used their consensus to obtain a robust grouping for PIDs. Considering the consensus of four, five, or six independent methods makes the results robust and independent from the individual methods.

### *Comparison with previous classifications*

Previous *in silico* disease classification methods have been based on shared genes in disorders (27), protein interactions (55), protein complexes (28), or tissue-specific gene expression (29). For example, Human Phenotype Ontology terms describe clinical features and can be used for disease grouping (56). However, because only a few PIDs were included in these classification studies, we unfortunately cannot compare our results with them.

The new PID classification differs in details from those published earlier. The most comprehensive classification with the largest number of PIDs is from the European Society for Immunodeficiencies (ESID) registry (7) based on the International Union of Immunological Societies classification of 150 diseases (6). This scheme contains seven defined disease groups. The IDR uses and expands the classification (from Ref. 2) in 11 classes. The American Academy of Allergy, Asthma, and Immunology, the American College of Allergy, Asthma, and Immunology, and the Joint Council of Allergy, Asthma, and Immunology have classified 97 PIDs in 5 groups (39). The International Classification of Diseases (ICD10) contains 100 PIDs in 10 categories (8). These classifications have been useful, although they have disagreed on a number of disorders. The task of classifying the widely variable PIDs is hardly any more possible to do manually due to the very high dimensionality of the data. A systematic, mathematical classification can be used as an alternative or complement approach to the existing methods.

The earlier classifications were color coded and visualized in supplemental Fig. 3. The previous studies contained only subsections of the PIDs. The color codes were also chosen so that interclassification comparisons are possible because related diseases in the different groupings have the same colors. The more homogeneous the color is within a cluster, the more similar the classifications are. The IDR and ESID classifications agree very well in DCs I, II, IV, V, VII, VIII, IX, X, and XI. ICD10 agrees well in DCs III, V, VII, IX, X, and XI. There are, however, only 100 PIDs in the ICD system. The American Academy of Allergy, Asthma, and Immunology grouping behaves similarly to ICD, except for DC X, diseases that are not included at all. In conclusion, the novel



**FIGURE 1.** Consensus for the six methods used to group the PIDs. **A**, Relationship of the diseases. The white rectangles indicate the grouping of diseases by five (those attached to the first black bullet from the PIDs root) and six (the second bullet) methods. Diseases are indicated by the systematic names of affected genes, when known. Otherwise, the following names were used when no genes are identified in relation to the disease: AR-HIES, autosomal recessive hyperimmunoglobulin E recurrent infection syndrome; CMC, chronic mucocutaneous candidiasis; CVID, common variable immunodeficiency; FHL1, familial hemophagocytic lymphohistiocytosis type 1; HIGM4, hyper-IgM syndrome type 4; SADNI, specific Ab deficiency with normal Ig concentrations; THI, transient hypogammaglobulinemia of infancy; thymoma, thymoma with immunodeficiency (Good's syndrome); and XLA/GHD, X-linked hypogammaglobulinemia with growth hormone deficiency. When one gene is involved in more than one disease, the following abbreviations were used:

classification agrees with numerous features in the earlier groupings. All of the previous classifications have a smaller number of groups, which are typically divided in the new nosology. There are also a number of diseases that are not consistent with the novel classification, or with the other previous groupings. The systematic mathematical approach is capable of solving these cases and places the diseases in groups using solely the given parameters.

#### *Parameters characterizing primary immunodeficiencies*

In PIDs, the signs and symptoms of infections may be repetitive, severe, or refractory to therapy and caused by organisms with low virulence (39). We grouped the parameters for infections according to microbial taxonomy and site of infection.

Autoimmune diseases and malignancies are complications of many immunodeficiencies. Based on the frequency of associations with autoimmune diseases, PIDs can be grouped in three groups, as follows: systematic (>80% of the patients with the disorder have autoimmune disease symptoms), strong (20–80%), and mild (<20% of the patients) and absent (12, 57). Some PID patients appear susceptible also to atopy and lupus-like syndromes (58).

Malignancies occur with great frequency in certain immunodeficiencies. The types of malignancies depend on the PID, the age of the patient, and any possible viral infection(s). B cell malignancies, especially non-Hodgkin's lymphomas, are predominant. Other types of malignancies encountered in PIDs are T cell malignancies and leukemia (14).

Problems in lymphoproliferation (hepatomegaly, splenomegaly, lymphadenopathy) are typical for some PIDs. EBV infection is associated with many lymphoproliferation-linked immunodeficiencies (59–61).

Some PID patients have chronic respiratory problems, such as asthma, chronic obstructive disease, chronic inflammatory lung disease, emphysema/lung cysts, or interstitial pneumonia (62–65). Cardiovascular diseases such as congenital cardiac anomalies, cardiomyopathy, and hematologic abnormalities are found in certain PIDs (66, 67).

Some of the gastrointestinal diseases are associated with PIDs, including esophageal atresia, Crohn's disease, chronic inflammatory bowel disease, ulcerative colitis, granulomatous colitis, celiac disease, malabsorption, Hirschsprung disease, and anal stenosis (15, 68–70). Hepatobiliary tract diseases in PID patients include storage liver disease, hepatic vascular occlusion disease, and sclerosing cholangitis (71, 72). Kidney diseases associated with PIDs include renal anomalies, renal dysfunction, amyloidosis, renal failure, IgA nephropathy, and glomerulonephritis (73).

Although physical findings are often absent, may be nonspecific, or very discreet, many PIDs have characteristic features. A common feature in PID patients is failure to thrive (child) or wasting (adult). Facial abnormalities, such as microcephaly or dysmorphism, are characteristic in some PIDs. Neurological abnormalities can include ataxia, peripheral neuropathy, speech delay, mental retardation, retinal lesions, photophobia, convulsions, or psychomotor retardation, whereas gastrointestinal abnormalities appear as severe gingivostomatitis, recurrent aphthae, periodontitis, delayed shedding of primary teeth, palatal weakness/cleft, gastric outlet obstruction, or diarrhea (1, 2, 39–41).

Ligamentous laxity/hyperextensive joints, limited extension of elbows, costochondral junction abnormality, rib abnormalities, metaphyseal chondrodysplasia/dysostosis, pectus carinatum, spondyloepiphyseal dysplasia, hip degeneration, or short limb dwarfism are among the skeletal abnormalities found in PIDs (74). The skin is frequently affected in immunodeficiencies. Erythroderma, eczema/atopic dermatitis, pyoderma, or granuloma is common. The presence of petechiae or bruises suggests a bleeding problem, as in phagocyte disorders (75–79).

The definitive diagnosis of PIDs depends on laboratory evidence, including assessments of humoral and cellular immunity and molecular analysis. The immunologic phenotype is based on laboratory tests of immune function, such as serum Ig levels, specific Ab titers, peripheral blood lymphocyte subpopulations, measures of T cell function, assays of phagocytes, and complement function or serum component level (1, 2, 37, 39–41).

Most of the parameters have binary values (yes, no), whereas, for example, the laboratory parameters are quantitative. The parameters were chosen so that they represent different important features of PIDs.

#### *The new PID classification*

The obtained 11 disease clusters are very robust due to them being the consensus of at least five methods. The *p* values show the significance of the observations. More detailed subgrouping is available by using the consensus of all the six methods. In Fig. 1A, diseases coclustered by all the six methods are within the boxes, whereas in the DCs at least five methods agree on the placement of PIDs within the graph. In addition to the actual classification, the PID parameters could offer guidelines for medical descriptions of PIDs. The classification allows a novel and fresh look at the relationships of PIDs, the genes behind them, and the encoded proteins. The network is far more complex than the previous mainly cell-type-based groupings might have led to imagine. Based on the classification and the parameters, it might be possible to develop novel diagnostic schemes for PIDs.

The correlation with other independent information not used for the original classification implies that the classification reliably reflects numerous properties of the diseases and the genes and proteins mutated in them. Data for the disease parameters were not complete, especially for the ultra-rare PIDs, and thus the homogeneity of the groups could even increase in the future when more reliable statistical information for the symptoms and laboratory characteristics will be available.

Diseases affecting genes involved in the same pathway do often cocluster, and they share a similar etiology, for example, JAK3 and IL2RG, and proteins in IFN-mediated immunity, including STAT1 and IFNGR2; MHC II diseases in DC III are all parts of the same protein complex as well as the MHC I components in DC V; complement components are in DC V and IX, and components of membrane bound oxidase in DC XI; BLNK, BTK, and MyD88, which have been suggested to be downstream of CD19 signaling, are all in DC V.

The new PID classification resulting from our approach can have several applications. Because it was generated independently from the existing classifications using solely mathematical analysis of clinical

CAPS, cryopyrin-associated periodic syndromes; Crohn's, Crohn's disease; CN\_ELA2, cyclic neutropenia; ELA2, severe congenital neutropenia; FCAS, familial cold autoinflammatory syndrome; MWS, Muckle-Wells syndrome; NOMID, neonatal-onset multisystem inflammatory disease; WAS, Wiskott-Aldrich syndrome; and XLN, X-linked neutropenia. OS denotes Omenn syndrome, in which multiple disease genes were identified. Independent properties of the diseases and their respective proteins have a rather homogeneous distribution with the disease clusters for *B*, disease severity; *C*, therapeutic options; *D*, functional properties of proteins; and *E*, vulnerability of the nodes in the PID protein interaction network. For supporting statistical parameters, see supplemental Table S2. The color-coded properties are shown in each subgraph.

and laboratory parameters, it can be used in evaluation and development of other classifications. Disease groups defined by experts and found also by our independent approach have a strong indication of their existence. The new classification will be used in IDR (4) for organizing diseases and information about them. Another possible application is detailed demographics and mortality records.

The classification and the dataset also serve the development of diagnostic expert systems, which requires objective criteria for diagnosis. Expert systems are useful, especially in case of rare diseases like PIDs. The computer-based disease classification applied in this study can also identify the key symptoms and laboratory parameters that can help the experts to diagnose correctly these diseases.

The approach can be applied also to other groups of diseases.

## Acknowledgments

We thank Jukka Lehtiniemi for help with the figures.

## Disclosures

The authors have no financial conflict of interest.

## References

- Stiehm, E. R., H. D. Ochs, J. A. Winklerstein, and E. Rich. 2004. *Immunologic Disorders in Infants and Children*. Elsevier, New York.
- Ochs, H. D., C. I. E. Smith, and J. M. Puck. 2007. *Primary Immunodeficiency Diseases: A Molecular and Genetic Approach*. Oxford University Press, Oxford.
- Notarangelo, L., J. L. Casanova, M. E. Conley, H. Chapel, A. Fischer, J. Puck, C. Roifman, R. Seger, and R. S. Geha. 2006. Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee Meeting in Budapest, 2005. *J. Allergy Clin. Immunol.* 117: 883–896.
- Samarghitean, C., J. Väliäho, and M. Vihinen. 2007. IDR knowledge base for primary immunodeficiencies. *Immunome Res.* 3: 6.
- Shearer, W. T., R. H. Buckley, R. J. Engler, A. F. Finn, Jr., T. A. Fleisher, T. M. Freeman, H. G. Herrod III, A. I. Levinson, M. Lopez, R. R. Rich, et al. 1996. Practice parameters for the diagnosis and management of immunodeficiency: The Clinical and Laboratory Immunology Committee of the American Academy of Allergy, Asthma, and Immunology (CLIC-AAAAI). *Ann. Allergy Asthma Immunol.* 76: 282–294.
- Geha, R. S., L. D. Notarangelo, J. L. Casanova, H. Chapel, M. E. Conley, A. Fischer, L. Hammarstrom, S. Nonoyama, H. D. Ochs, J. M. Puck, et al. 2007. Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee. *J. Allergy Clin. Immunol.* 120: 776–794.
- Guzman, D., D. Veit, V. Knerr, G. Kindle, B. Gathmann, A. M. Eades-Perner, and B. Grimbacher. 2007. The ESID Online Database network. *Bioinformatics* 23: 654–655.
- WHO. 2007. *International Statistical Classification of Diseases and Related Health Problems*. World Health Organization, Geneva.
- Riches, P. G. 1992. Viral infections complicating primary immunodeficiencies. *Clin. Ter.* 140: 123–129.
- Antachopoulos, C., T. J. Walsh, and E. Roilides. 2007. Fungal infections in primary immunodeficiencies. *Eur. J. Pediatr.* 166: 1099–1117.
- Bustamante, J., S. Boisson-Dupuis, E. Jouanguy, C. Picard, A. Puel, L. Abel, and J. L. Casanova. 2008. Novel primary immunodeficiencies revealed by the investigation of pediatric infectious diseases. *Curr. Opin. Immunol.* 20: 39–48.
- Carneiro-Sampaio, M., and A. Coutinho. 2007. Tolerance and autoimmunity: lessons at the bedside of primary immunodeficiencies. *Adv. Immunol.* 95: 51–82.
- Arkwright, P. D., M. Abinun, and A. J. Cant. 2002. Autoimmunity in human primary immunodeficiency diseases. *Blood* 99: 2694–2702.
- Salavoura, K., A. Kolialexi, G. Tsangaris, and A. Mavrou. 2008. Development of cancer in patients with primary immunodeficiencies. *Anticancer Res.* 28: 1263–1269.
- Assari, T. 2006. Chronic granulomatous disease; fundamental stages in our understanding of CGD. *Med. Immunol.* 5: 4.
- Janka, G. E. 2007. Hemophagocytic syndromes. *Blood Rev.* 21: 245–253.
- Filipovich, A. H. 2008. Hemophagocytic lymphohistiocytosis and other hemophagocytic disorders. *Immunol. Allergy Clin. North Am.* 28: 293–313.
- Frank, M. M. 2008. Hereditary angioedema. *J. Allergy Clin. Immunol.* 121: S398–S401.
- Shinkai, K., T. H. McCalmont, and K. S. Leslie. 2008. Cryopyrin-associated periodic syndromes and autoinflammation. *Clin. Exp. Dermatol.* 33: 1–9.
- Stojanov, S., and M. F. McDermott. 2005. The tumor necrosis factor receptor-associated periodic syndrome: current concepts. *Expert Rev. Mol. Med.* 7: 1–18.
- Brydges, S., and D. L. Kastner. 2006. The systemic autoinflammatory diseases: inborn errors of the innate immune system. *Curr. Top. Microbiol. Immunol.* 305: 127–160.
- Marodi, L., and L. D. Notarangelo. 2007. Immunological and genetic bases of new primary immunodeficiencies. *Nat. Rev. Immunol.* 7: 851–861.
- Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5: 101–113.
- Pastor-Satorras, R., and A. Vespignani. 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86: 3200–3203.
- Pastor-Satorras, R., and A. Vespignani. 2002. Immunization of complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65: 036104.
- Savli, H., A. Szendroi, I. Romics, and B. Nagy. 2008. Gene network and canonical pathway analysis in prostate cancer: a microarray study. *Exp. Mol. Med.* 40: 176–185.
- Goh, K. I., M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi. 2007. The human disease network. *Proc. Natl. Acad. Sci. USA* 104: 8685–8690.
- Lage, K., E. O. Karlberg, Z. M. Storling, P. I. Olason, A. G. Pedersen, O. Rigina, A. M. Hinsby, Z. Tumer, F. Pociot, N. Tommerup, et al. 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* 25: 309–316.
- Lage, K., N. T. Hansen, E. O. Karlberg, A. C. Eklund, F. S. Roque, P. K. Donahoe, Z. Szallasi, T. S. Jensen, and S. Brunak. 2008. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. USA* 105: 20870–20875.
- Ortutay, C., and M. Vihinen. 2008. Efficiency of the immunome protein interaction network increases during evolution. *Immunome Res.* 4: 4.
- Barrell, D., E. Dimmer, R. P. Huntley, D. Binns, C. O'Donovan, and R. Apweiler. 2009. The GOA database in 2009: an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.* 37: D396–D403.
- Ortutay, C., and M. Vihinen. 2009. Identification of candidate disease genes by integrating gene ontologies and protein-interaction networks: case study of primary immunodeficiencies. *Nucleic Acids Res.* 37: 622–628.
- Kaufman, L., and P. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Pons, P., and M. Latapy. 2005. Computing communities in large networks using random walks. *arXiv:physics/0512106*.
- Newman, M. E. J., and M. Girvan. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* 69: 026113.
- Newman, M. E. J. 2006. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74: 036104.
- Samarghitean, C., J. Väliäho, and M. Vihinen. 2004. Online registry of genetic and clinical immunodeficiency diagnostic laboratories, IDiagnostics. *J. Clin. Immunol.* 24: 53–61.
- Piirilä, H., J. Väliäho, and M. Vihinen. 2006. Immunodeficiency mutation databases (IDbases). *Hum. Mutat.* 27: 1200–1208.
- Bonilla, F. A., I. L. Bernstein, D. A. Khan, Z. K. Ballas, J. Chinen, M. M. Frank, L. J. Kobrynski, A. I. Levinson, B. Mazer, R. P. Nelson, Jr., et al. 2005. Practice parameter for the diagnosis and management of primary immunodeficiency. *Ann. Allergy Asthma Immunol.* 94: S1–S63.
- Spickett, G. 2006. *Oxford Handbook of Clinical Immunology and Allergy*. Oxford University Press, Oxford.
- De Vries, E. 2006. Patient-centered screening for primary immunodeficiency: a multi-stage diagnostic protocol designed for non-immunologists. *Clin. Exp. Immunol.* 145: 204–214.
- Eades-Perner, A. M., B. Gathmann, V. Knerr, D. Guzman, D. Veit, G. Kindle, and B. Grimbacher. 2007. The European internet-based patient and research database for primary immunodeficiencies: results 2004–06. *Clin. Exp. Immunol.* 147: 306–312.
- Wehr, C., T. Kivioja, C. Schmitt, B. Ferry, T. Witte, E. Eren, M. Vlkova, M. Hernandez, D. Detkova, P. R. Bos, et al. 2008. The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* 111: 77–85.
- Teo, Y. Y. 2008. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr. Opin. Lipidol.* 19: 133–143.
- Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* 1695.
- Prevalence of rare diseases: bibliographic data, Orphanet Report Series, May 2009:1. [http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence\\_of\\_rare\\_diseases\\_by\\_alphabetical\\_list.pdf](http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf)
- Ortutay, C., M. Siermala, and M. Vihinen. 2007. Molecular characterization of the immune system: emergence of proteins, processes, and domains. *Immunogenetics* 59: 333–348.
- Ortutay, C., and M. Vihinen. 2009. Immunome Knowledge Base (IKB): an integrated service for immunome research. *BMC Immunol.* 10: 3.
- Ortutay, C., and M. Vihinen. 2006. Immunome: a reference set of genes and proteins for systems biology of the human immune system. *Cell. Immunol.* 244: 87–89.
- Ayme, S., and M. Preus. 1984. The Marshall and Stickler syndromes: objective rejection of lumping. *J. Med. Genet.* 21: 34–38.
- Verloes, A. 1995. Numerical syndromology: a mathematical approach to the nosology of complex phenotypes. *Am. J. Med. Genet.* 55: 433–443.
- Sheffer, C. E., J. A. Deisinger, J. E. Cassisi, and K. Lofland. 2007. A revised taxonomy of patients with chronic pain. *Pain Med.* 8: 312–325.
- Haldar, P., I. D. Pavord, D. E. Shaw, M. A. Berry, M. Thomas, C. E. Brightling, A. J. Wardlaw, and R. H. Green. 2008. Cluster analysis and clinical asthma phenotypes. *Am. J. Respir. Crit. Care Med.* 178: 218–224.
- Wardlaw, A. J., M. Silverman, R. Siva, I. D. Pavord, and R. Green. 2005. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin. Exp. Allergy* 35: 1254–1262.
- Feldman, I., A. Rzhetsky, and D. Vitkup. 2008. Network properties of genes harboring inherited disease mutations. *Proc. Natl. Acad. Sci. USA* 105: 4323–4328.

56. Robinson, P. N., S. Kohler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* 83: 610–615.
57. Carneiro-Sampaio, M., and A. Coutinho. 2007. Immunity to microbes: lessons from primary immunodeficiencies. *Infect. Immun.* 75: 1545–1555.
58. Karim, M. Y. 2006. Immunodeficiency in the lupus clinic. *Lupus* 15: 127–131.
59. Tran, H., J. Nourse, S. Hall, M. Green, L. Griffiths, and M. K. Gandhi. 2008. Immunodeficiency-associated lymphomas. *Blood Rev.* 22: 261–281.
60. Wallet-Faber, N., C. Bodemer, S. Blanche, E. Delabesse, C. Eschard, N. Brousse, and S. Fraitag. 2008. Primary cutaneous Epstein-Barr virus-related lymphoproliferative disorders in 4 immunosuppressed children. *J. Am. Acad. Dermatol.* 58: 74–80.
61. Elenitoba-Johnson, K. S., and E. S. Jaffe. 1997. Lymphoproliferative disorders associated with congenital immunodeficiencies. *Semin. Diagn. Pathol.* 14: 35–47.
62. Garcia-Laorden, M. I., J. Sole-Violan, F. Rodriguez de Castro, J. Aspa, M. L. Briones, A. Garcia-Saavedra, O. Rajas, J. Blanquer, A. Caballero-Hidalgo, J. A. Marcos-Ramos, et al. 2008. Mannose-binding lectin and mannose-binding lectin-associated serine protease 2 in susceptibility, severity, and outcome of pneumonia in adults. *J. Allergy Clin. Immunol.* 122: 368–374, 374.e1–2.
63. Basile, N., S. Danielian, M. Oleastro, S. Rosenzweig, E. Prieto, J. Rossi, A. Roy, and M. Zelazko. 2009. Clinical and molecular analysis of 49 patients with X-linked agammaglobulinemia from a single center in Argentina. *J. Clin. Immunol.* 29: 123–129.
64. Bott, L., J. Lebreton, C. Thumerelle, J. Cuvellier, A. Deschildre, and A. Sardet. 2007. Lung disease in ataxia-telangiectasia. *Acta Paediatr.* 96: 1021–1024.
65. Quinti, I., A. Soresina, G. Spadaro, S. Martino, S. Donnanno, C. Agostini, P. Claudio, D. Franco, A. Maria Pesce, F. Borghese, et al. 2007. Long-term follow-up and outcome of a large cohort of patients with common variable immunodeficiency. *J. Clin. Immunol.* 27: 308–316.
66. Shprintzen, R. J. 2008. Velo-cardio-facial syndrome: 30 years of study. *Dev. Disabil. Res. Rev.* 14: 3–10.
67. Sweeney, R. T., G. J. Davis, and J. A. Noonan. 2008. Cardiomyopathy of unknown etiology: Barth syndrome unrecognized. *Congenit. Heart Dis.* 3: 443–448.
68. Marks, D. J., K. Miyagi, F. Z. Rahman, M. Novelli, S. L. Bloom, and A. W. Segal. 2009. Inflammatory bowel disease in CGD reproduces the clinicopathological features of Crohn's disease. *Am. J. Gastroenterol.* 104: 117–124.
69. Barton, L. L., S. L. Moussa, R. G. Villar, and R. L. Hulet. 1998. Gastrointestinal complications of chronic granulomatous disease: case report and literature review. *Clin. Pediatr.* 37: 231–236.
70. Daniels, J. A., H. M. Lederman, A. Maitra, and E. A. Montgomery. 2007. Gastrointestinal tract pathology in patients with common variable immunodeficiency (CVID): a clinicopathologic study and review. *Am. J. Surg. Pathol.* 31: 1800–1812.
71. Cliffe, S. T., M. Wong, P. J. Taylor, E. Ruga, B. Wilcken, R. Lindeman, M. F. Buckley, and T. Roscioli. 2007. The first prenatal diagnosis for veno-occlusive disease and immunodeficiency syndrome, an autosomal recessive condition associated with mutations in SP110. *Prenat. Diagn.* 27: 674–676.
72. Hayward, A. R., J. Levy, F. Facchetti, L. Notarangelo, H. D. Ochs, A. Etzioni, J. Y. Bonnefoy, M. Cosyns, and A. Weinberg. 1997. Cholangiopathy and tumors of the pancreas, liver, and biliary tree in boys with X-linked immunodeficiency with hyper-IgM. *J. Immunol.* 158: 977–983.
73. De Heer, E., and D. J. Peters. 2008. Innate immunity as a driving force in renal disease. *Kidney Int.* 73: 7–8.
74. Sordet, C., A. Cantagrel, T. Schaevebeke, and J. Sibilia. 2005. Bone and joint disease associated with primary immune deficiencies. *Joint Bone Spine* 72: 503–514.
75. Heyworth, P. G., A. R. Cross, and J. T. Curnutte. 2003. Chronic granulomatous disease. *Curr. Opin. Immunol.* 15: 578–584.
76. De Raeve, L., M. Song, J. Levy, and F. Mascart-Lemone. 1992. Cutaneous lesions as a clue to severe combined immunodeficiency. *Pediatr. Dermatol.* 9: 49–51.
77. Saurat, J. H. 1985. Eczema in primary immune-deficiencies: clues to the pathogenesis of atopic dermatitis with special reference to the Wiskott-Aldrich syndrome. *Acta Derm. Venereol. Suppl.* 114: 125–128.
78. Moin, A., A. Farhoudi, M. Moin, Z. Pourpak, and N. Bazargan. 2006. Cutaneous manifestations of primary immunodeficiency diseases in children. *Iran J. Allergy Asthma Immunol.* 5: 121–126.
79. Chowdhury, M. M., A. Anstey, and C. N. Matthews. 2000. The dermatosis of chronic granulomatous disease. *Clin. Exp. Dermatol.* 25: 190–194.