

RAPID: Resource of Asian Primary Immunodeficiency Diseases

Shivakumar Keerthikumar^{1,2,3}, Rajesh Raju^{1,2,3}, Kumaran Kandasamy^{1,3,4},
Atsushi Hijikata⁵, Subhashri Ramabadrani^{1,2}, Lavanya Balakrishnan^{1,2},
Mukhtar Ahmed¹, Sandhya Rani¹, Lakshmi Dhevi N. Selvan¹, Devi S. Somanathan¹,
Somak Ray¹, Mitali Bhattacharjee¹, Sashikanth Gollapudi¹, Y. L. Ramachandra³,
Sahely Bhadra⁶, Chiranjib Bhattacharyya⁶, Kohsuke Imai⁷, Shigeaki Nonoyama⁷,
Hirokazu Kanegane⁸, Toshio Miyawaki⁸, Akhilesh Pandey^{1,4},
Osamu Ohara^{5,9,*} and Sujatha Mohan^{1,2}

¹Institute of Bioinformatics, International Technology Park, Bangalore 560 066, India, ²Research Unit for Immunoinformatics, Research Center for Allergy and Immunology, RIKEN Yokohama Institute, Kanagawa 230-0045, Japan, ³Department of Life Science, Kuvempu University, Jnanasahyadri, Shimoga 577 451, India, ⁴McKusick-Nathans Institute of Genetic Medicine and Departments of Biological Chemistry, Pathology and Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA, ⁵Laboratory for Immunogenomics, Research Center for Allergy and Immunology, RIKEN, Yokohama Institute, Kanagawa 230-0045, Japan, ⁶Department of Computer Science and Automation, Indian Institute of Science, Bangalore-12, India, ⁷Department of Pediatrics, National Defense Medical College, Saitama 359-8513, Japan, ⁸Department of Pediatrics, Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, Toyama 930-0194 and ⁹Department of Human Genome Technology, Kazusa DNA Research Institute, Chiba 292-0818, Japan

Received August 15, 2008; Revised September 21, 2008; Accepted September 23, 2008

ABSTRACT

Availability of a freely accessible, dynamic and integrated database for primary immunodeficiency diseases (PID) is important both for researchers as well as clinicians. To build a PID informational platform and also as a part of action to initiate a network of PID research in Asia, we have constructed a web-based compendium of molecular alterations in PID, named Resource of Asian Primary Immunodeficiency Diseases (RAPID), which is available as a worldwide web resource at <http://rapid.rcai.riken.jp/>. It hosts information on sequence variations and expression at the mRNA and protein levels of all genes reported to be involved in PID patients. The main objective of this database is to provide detailed information pertaining to genes and proteins involved in primary immunodeficiency diseases along with other relevant information about protein-protein interactions, mouse studies and microarray gene-expression profiles in various organs and cells of the immune system. RAPID also hosts a tool, mutation viewer, to predict deleterious

and novel mutations and also to obtain mutation-based 3D structures for PID genes. Thus, information contained in this database should help physicians and other biomedical investigators to further investigate the role of these molecules in PID.

INTRODUCTION

Primary immunodeficiency diseases (PIDs) are a class of disorders resulting from intrinsic defects in genes involved in the development and maintenance of the immune system. More than 150 primary immunodeficiency genes are reported thus far, which are classified into eight different categories by the International Union of Immunological Societies (1). Patients with these intrinsic defects have increased susceptibility to recurrent and persistent infections, and they may also have autoimmune and cancer-related symptoms. Most PIDs are rare and the diagnosed patients for a given condition are often randomly spread out around the world (2). The genetic defects that cause PIDs can affect the expression and function of proteins that are involved in a range of biological process such as

*To whom correspondence should be addressed. Tel: +81 45 503 9695; Fax: +81 45 503 9694; Email: oosamu@rcai.riken.jp
Correspondence may also be addressed to Sujatha Mohan. Tel: +81 45 503 7034; Fax: +81 45 503 9694; Email: sujatha@rcai.riken.jp

immune development, effector-cell functions, signaling cascades and maintenance of immune homeostasis (3). Recent advances in both diagnosis and therapeutic modalities have allowed these defects to be identified earlier and to be more precisely defined, and they have also resulted in more promising long-term outcomes (4). Development of a freely accessible, dynamic and integrated database with inclusion of genomics, transcriptomics and proteomics data of all genes that are involved in PID has the potential to further accelerate research into PIDs as well as provide physicians with easy access to pertinent clinical and molecular data that is otherwise spread throughout the literature.

Resource of Asian Primary Immunodeficiency Diseases (RAPID) is a web-based compendium of molecular alterations in PIDs that is freely available to the academic community at <http://rapid.rcai.riken.jp>. It hosts information on the sequence variations mapped to the mRNA and protein sequences for all the genes reported from PID patients worldwide. Besides molecular alterations, RAPID has the protein-protein interaction network from Human Protein Reference Database (HPRD) (5), a database with protein-centric information for all human proteins, along with a graphical representation of the expression of PID genes from microarray profiling of organs and cells of the immune system from Gene Expression Omnibus (GEO) (6) and Reference Database of Immune Cells (RefDIC) (7). In addition, it incorporates mouse studies from Mouse Genome Informatics (MGI) (8) database for the representation of allele-based phenotypes and anatomical systems affected due to either mouse knockouts, knockins and spontaneous mutations for the available PID genes. With inclusion of this variety of data for the PID disease genes, RAPID can serve as a connecting link between the genotype and the phenotype.

RAPID ARCHITECTURE

RAPID is an object-oriented database. We used Zope (<http://www.zope.org>) for the development of RAPID. Zope is a leading open source web application server and is built using the programming language Python (<http://www.python.org>). MySQL is used as a backend data storage system.

RAPID allows users access to gene-specific PID information either by using the query page or by browsing. RAPID can be queried by various search options including gene symbol, protein name, mouse phenotypes, chromosome number and PID category, and accession numbers of entries from several database resources. The query system includes an autocomplete option that facilitates quick access to the list of items in the database. Users can browse this resource by PID genes, mutation features such as mutation types and effects. This is the first of its kind database to have these user-friendly features for search and display options.

Primary information page is the default main page of every PID gene in the RAPID. It summarizes the external links available in the public domain such as Entrez (9,10), HPRD, IDR, RefDIC, NetPath, OMIM (11),

HGNC (12), PDB (13), Ensembl (14) and UniProt/Swiss-Prot (15). It also includes the disease phenotype linked to the given gene along with the mode of inheritance, alternative names of the gene function and the associated features (Figure 1A).

ANNOTATION OF MUTATION DATA

The sequence variations in PID genes reported from the patients are manually curated by expert biologists from the published literature and mapped to the NCBI RefSeq (16,17) genomic, cDNA and protein sequences to aid the scientific community as per the recommendations for the description of sequence variants by Human Genome Variation Society (HGVS) whose main objectives remain to ensure documentation, collection and free distribution of all variation information (18,19). The main criterion for inclusion of mutation data in the RAPID from the literature is that the PID causing genes mutation analysis has to be performed in patient samples who have already presented with a set of PID clinical presentations or associated features or from cell lines derived from such patients. Carriers and asymptomatic individuals are not included in our annotations.

MUTATION VIEWER

Each entry on the mutation has a link to the mutation viewer, a web-based graphical user interface (GUI) enabled tool named Mutation@A Glance, which can be accessed at <http://rapid.rcai.riken.jp/mutBrowse.cgi> (Figure 1B). It allows users to visualize the mutation position both at the level of DNA and protein sequences as well as homology based 3D structures with various types of information such as SNP, protein domains and functional sites (Hijikata, *et al.*, manuscript in preparation).

GENE-EXPRESSION PROFILES

Gene-expression profiles of PID genes in various organs and cells of the immune systems from a number of microarray experiments available in public repositories are represented graphically for the available PID gene in the form of histograms with immune cell types along with the corresponding average gene-expression intensity values. This shows the expression pattern of these genes in immune cells, which is of relevance to immunologists (Figure 2A).

INTERACTION NETWORKS

All the primary immunodeficiency disease genes have been mapped to their direct physical interactors and, in turn, these interactors are represented differently based on whether an interactor is already known to be PID or not, any of these interactors have their site of expression in immune cells, or any lethality and/or immune system/hematopoietic phenotypes affected due to either mouse knockout, or spontaneous mutation in mice. This information is graphically represented in the form of nodes and

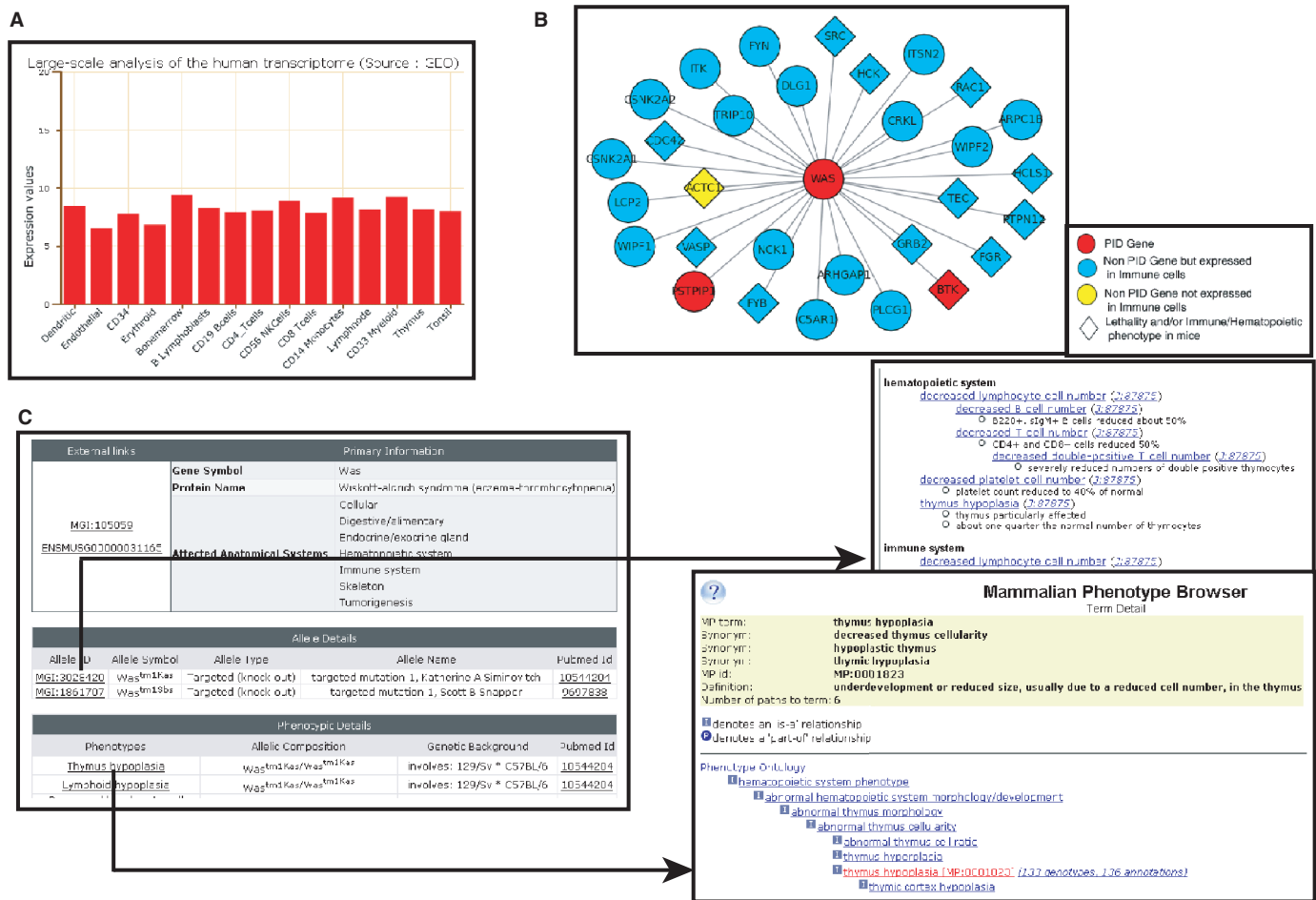


Figure 2. A screenshot of the expression profile, interaction network and mouse studies of *WAS* gene. (A) Microarray gene expression profile of *WAS* gene in a particular human microarray data set from GEO. (B) Interaction networks of *WAS* from HPRD is shown along with the legends, which explain various parameters of interactors in different shapes and colors. (C) Primary information of mouse gene *WAS* is shown along with anatomical systems affected due to mutations and its allelic composition and phenotypes from MGI database.

Table 1. Overall statistics

Number of PID genes	161
Number of PID genes with mutation data	143
Number of non-redundant mutation entries	2455
Number of unique PubMed citations	762

for PID patients. With community participation of interested groups, we anticipate that RAPID will become a primary resource of PIDs in Asia.

ACKNOWLEDGEMENTS

The authors thank Yasuaki Murahashi and Miho Izawa for their help in setting up the web servers, Advanced Center of Computing and Communication (ACCC), RIKEN for providing necessary computing resources, Drs Reiko Kikuno and Fumihiko Ishikawa for their valuable discussion, and all PID physicians involved in the PID Japan project for their valuable input and suggestions.

FUNDING

The Special Coordination Funds for promoting Science and Technology, from the Ministry of Education, Culture, Sports, and Science of Japan. Funding for open access charge: RIKEN Research Center for Allergy and Immunology.

Conflict of interest statement. None declared.

REFERENCES

- Geha, R.S., Notarangelo, L.D., Casanova, J.L., Chapel, H., Conley, M.E., Fischer, A., Hammarström, L., Nonoyama, S., Ochs, H.D., Puck, J.M. *et al.* (2007) International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee. Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee. *J. Allergy Clin. Immunol.*, **120**, 776–794.
- Samarghitean, C., Valiaho, J. and Vihinen, M. (2007) IDR knowledge base for primary immunodeficiencies. *Immunome Res.*, **3**, 6.
- Marodi, L. and Notarangelo, L.D. (2007) Immunological and genetic bases of new primary immunodeficiencies. *Nat. Rev. Immunol.*, **7**, 851–861.

4. Cunningham-Rundles,C. and Ponda,P.P. (2005) Molecular defects in T- and B-cell primary immunodeficiency diseases. *Nat. Rev. Immunol.*, **5**, 880–892.
5. Mishra,G., Suresh,M., Kumaran,K., Kannabiran,N., Suresh,S., Bala,P., Shivakumar,K., Anuradha,N., Raghunath,R., Madhan Raghavan,T. *et al.* (2006) Human Protein Reference Database - 2006 update. *Nucleic Acids Res.*, **34**, D411–D414.
6. Barrett,T. and Edgar,R. (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol.*, **411**, 352–369.
7. Hijikata,A., Kitamura,H., Kimura,Y., Yokoyama,R., Aiba,Y., Bao,Y., Fujita,S., Hase,K., Hori,S. *et al.* (2007) Construction of an open-access database that integrates cross-reference information from the transcriptome and proteome of immune cells. *Bioinformatics*, **23**, 2934–2941.
8. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E. and the members of the Mouse Genome Database Group (2007) The Mouse Genome Database (MGD): new features facilitating a model system. *Nucleic Acids Res.*, **35**, D630–D637.
9. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
10. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
11. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
12. Bruford,E.A., Lush,M.J., Wright,M.W., Sneddon,T.P., Povey,S. and Birney,E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, **36**, D445–D448.
13. Dutta,S., Burkhardt,K., Swaminathan,G.J., Kosada,T., Henrick,K., Nakamura,H. and Berman,H.M. (2008) Data deposition and annotation at the worldwide protein data bank. *Methods Mol. Biol.*, **426**, 81–101.
14. Flicek,P., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F., Cutts,T. *et al.* (2008) Ensembl 2008. *Nucleic Acids Res.*, **36**, D707–D714.
15. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
16. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
17. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
18. Antonarakis,S.E. (1998) Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum. Mutat.*, **11**, 1–3.
19. den Dunnen,J.T. and Antonarakis,S. E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.*, **15**, 7–12.