Identification of candidate disease genes by integrating Gene Ontologies and protein interaction networks: Case study of primary immunodeficiencies

Supplementary discussion

Csaba Ortutay[1] and Mauno Vihinen[1,2,*]

[1]Institute of Medical Technology, FI-33014 University of Tampere, Finland

[2]Tampere University Hospital, FI-33520 Tampere, Finland

[*]To whom correspondence should be addressed. Tel: +358-3-3551 7735; Fax: +358-3-3551 7710; Email: mauno.vihinen@uta.fi

Biological function of the PID candidate genes

When grouped to general categories, among the candidate genes, there are 8 receptors, 8 protein kinases and 7 regulators. Only two of the candidate proteins are involved in the central processing of invaders. Granzyme B is an endopeptidase in cytolytic T cells and natural killer (NK) cells. The other proteolytic enzyme is cathepsin G, which is expressed in neutrophilic polymorphonuclear leukocytes. Both of these enzymes are essential for killing and digesting pathogens and are thus suspected PID candidates. The reason why they have not yet been identified from patients could be that there are also other proteases and, as very promiscuous enzymes, can therefore easily be replaced. Further, these proteins might be disease related only for certain rare pathogens.

CD4, CD9, ITGB1 (CD29), THY1 (CD29), KIT (CD117) and IL2RB (CD122) are cell surface receptors. CD4 appears on the surface of helper and regulatory T-cells, macrophages and dendritic cells. It is part of the T-cell receptor (TCR) and important for signaling in T-cells. CD9 interacts with several proteins including integrins and others involved in cell adhesion, motility and platelet activation. Interleukin 2 receptor β (IL2RB) forms a central T-cell signaling receptor with the α and γ chains. Mutations in the α–(IL2RA) and so-called common γ–chain (IL2RG) lead to SCID (1), thus it is likely that β–chain mutations would also be PID-related. Integrin β1 subunit (ITGB1) appears in several alternatively spliced forms and has numerous interacting partners when bound to different α subunits to constitute the functional dimer. THY1 (CD90) is a heavily glycosylated GPI-anchored protein that has a wide expression pattern. Its detailed function is not yet known but there is data implying that it is important for cell-cell and cell-ligand interactions in inflammation, apoptosis, and T-cell

activation. KIT is a cytokine receptor protein tyrosine kinase (PTK) that is essential for proliferation, differentiation and survival of several cell types. It is a proto-oncogene and related to several forms of cancers, as listed in KinMutBase (2).

Nuclear factor κB (NF-κB) is known to regulate the expression of hundreds of genes active in immunity, inflammation, cell death and other processes. The specificity of the responses relates to numerous processes and features. There are at least four distinct pathways related to it (for a review see 3). The stimulation of classical and alternative pathway cell surface receptors leads to activation of IKK kinase complex. Once activated, the complex phosphorylates inhibitory I-κB, which releases from the NF-κB dimer, and becomes activated and transported to the nucleus where it recognizes 9-10 bp long binding sites on the genomic DNA and acts as a transcription factor.

Several proteins are involved in these pathways depending on the receptors and constitution of the IKK and NF-κB complexes. Many of these proteins have several family members, and thus the number of different combinations is high when complexes are formed. *RELA*, one of the candidate genes, belongs to the central Rel/NF-κB gene family. Other PID candidates in this pathway are IRAK1, RIPK1, TRAF1 and TRAF2. Signals from the receptors are transmitted further by interactions with adapter proteins and kinases. Interleukin-1 receptor-associated kinase 1 (IRAK1) is a serine/treonine kinase that regulates NF-κB. Several other PID candidate genes also regulate NF-κB, including the receptor (TNFRSF)-interacting serine-threonine kinase 1 (RIPK1), and TNF receptor-associated factors 1 and 2 (TRAF1, TRAF2). TRAFs mediate the connection of tumor necrosis factor receptor family members and several downstream partners in addition to NF-κB along the MAPK8/JNK pathway.

TRAF1 and TRAF2 form a heterodimer.

Since the NF-κB pathway is so essential and there are many proteins involved in these cascades, several diseases have been found in patients related to it (for a review see 4). There are PIDs among these diseases, namely X-linked hyper IgM syndrome and hypohydrotic ectodermal dysplasia due to mutations in NEMO (IKBKG) (5), and autosomal dominant anhidrotic ectodermal dysplasia and T-cell immunodeficiency because of IKKα (NFKBIA) (6) mutations. Both these proteins are subunits of the IKK complex. IRAK4 deficiency arises from impaired IRAK4 function (7). Along with IRAK1, one of our PID candidates, it participates in IL-1 receptor and Toll-like receptor (TLR) pathways. Information for disease causing mutations in these genes and other PID genes is available from IDbases (8).

Other PID candidates with signaling functions include Fyn, a Src family PTK, and MAPK14, a MAP family serine/threonine kinase. The Src and MAP kinase families are both very essential signaling routes that regulate various activities within cells. Knock-out mouse model for Fyn shows several defects in T-cell development and function (9,10). Soluble and receptor kinases are often cancer-causing when mutated. Both inactivating and constitutively activating gain-of-function mutations can be deleterious.

Insulin-like growth factor 1 receptor (IGF1R) is another PTK. It recognizes IGF-1, which is a polypeptide hormone and similar to insulin, which binds to another PID candidate, insulin receptor (INSR), another receptor PTK. INSR is the causative gene for diabetes mellitus, an autoimmune disease. Our method selects it because the patients could also have PID symptoms which, however, are usually hidden because of diabetes. Again, there is a possibility that some patients might also have PID symptoms.

Tyrosine kinase 2 (TYK2) belongs to the same family as Janus kinase 3 (JAK3), which leads to SCID when mutated (11,12), and JAK1 and JAK2, which are also PID candidates. The family members are involved in cytokine receptor signaling pathways. They exert their function by phosphorylating STAT family member proteins, which form homo- and heterodimers. The dimers are then translocated to the nucleus to act as transcription activators. STAT1 and -2 cause PID when mutated (13) and STAT3 is a PID candidate (14). The importance of these pathways is further indicated by suppressor of cytokine signaling 1 (SOCS1), another PID candidate gene. SOCS1 is a negative regulator of the JAK/STAT pathways.

Low density lipoprotein-related protein 1 (LRP1) is a low density lipoprotein (LDL) family member and plays a role in receptor-mediated endocytosis and phagocytosis. LRP1 binds to apoE-containing lipoproteins in a C1q-mediated apoptotic cascade.

X-ray repair complementing defective repair in Chinese hamster cells 5 (XRCC5) is a subunit of ATP-dependent DNA helicase II or Ku heterodimer that works in the repair of DNA double strand breaks by non-homologous end joining in the generation of immunological receptor and antibody genes. This is a highly likely PID gene as the knock-out in mice has a lethal effect (15,16). The effect may also be lethal in humans and therefore patients have not been identified with mutations in this gene. Defective V(D)J recombination of immunological receptors and antibodies severely affects adaptive immunity.

Some of the candidate genes are so essential that it is likely that mutations in them have lethal effects and therefore may not appear in PID patients. For example, deletion of the Fas-associated death domain (FADD) gene causes embryonic lethality in mice (17,18). In platelet-derived growth factor receptor, beta polypeptide (PDGFRB) mutations cause perinatal

lethality (19).

Estimation of the performance of the method

Our bioinformatics method can prioritize possible candidate genes. Accurate estimation of selectivity and specificity of the method cannot be done because of lack of a test set. We can roughly estimate the performance as follows: The high network scores together with the significant gene ontology terms resulted in 39 genes (Fig. 2). 13 of them were known PID genes. When extrapolating based on this ratio we expect at least 8 to 9 genes of the predicted 26 to have a connection to PIDs. This ratio is somewhat higher than the 10% estimated for some earlier methods (20). Recently some methods have been reported to have higher selectivity, around 20-40% (21,22). These methods restrict the search space and require the chromosomal location of the disease gene to be known. Our approach does not need or use this kind of information and is thus a true *de novo* candidate prediction method.

References

1. Sharfe, N., Dadi, H.K., Shahar, M. and Roifman, C.M. (1997) Human immune disorder arising from mutation of the α chain of the interleukin-2 receptor. *Proc Natl Acad Sci U S A*, **94,** 3168-3171.

2. Ortutay, C., Väliaho, J., Stenberg, K. and Vihinen, M. (2005) KinMutBase: a registry of disease-causing mutations in protein kinase domains. *Hum Mutat*, **25,** 435-442.

3. Niederberger, E. and Geisslinger, G. (2008) The IKK-NF-κB pathway: a source for novel molecular drug targets in pain therapy? *FASEB J. doi:10.1096/fj.08-109355*

4. Courtois, G. and Gilmore, T.D. (2006) Mutations in the NF-κB signaling pathway: implications for human disease. *Oncogene*, **25,** 6831-6843.

5. Zonana, J., Elder, M.E., Schneider, L.C., Orlow, S.J., Moss, C., Golabi, M., Shapira, S.K., Farndon, P.A., Wara, D.W., Emmal, S.A. *et al.* (2000) A novel X-linked disorder of immune deficiency and hypohidrotic ectodermal dysplasia is allelic to incontinentia pigmenti and due to mutations in IKK-γ (NEMO). *Am J Hum Genet*, **67,** 1555-1562.

6. Courtois, G., Smahi, A., Reichenbach, J., Doffinger, R., Cancrini, C., Bonnet, M., Puel, A., Chable-Bessia, C., Yamaoka, S., Feinberg, J. *et al.* (2003) A hypermorphic IκBα mutation is associated with autosomal dominant anhidrotic ectodermal dysplasia and T cell immunodeficiency. *J Clin Invest*, **112,** 1108-1115.

7. Kuhns, D.B., Long Priel, D.A. and Gallin, J.I. (1997) Endotoxin and IL-1 hyporesponsiveness in a patient with recurrent bacterial infections. *J Immunol*, **158,** 3959-3964.

8. Piirilä, H., Väliaho, J. and Vihinen, M. (2006) Immunodeficiency mutation databases (IDbases). *Hum Mutat*, **27,** 1200-1208.

9. Stein, P.L., Lee, H.M., Rich, S. and Soriano, P. (1992) pp59^{fyn} mutant mice display differential signaling in thymocytes and peripheral T cells. *Cell*, **70,** 741-750.

10. Pereira, D.B. and Chao, M.V. (2007) The tyrosine kinase Fyn determines the localization of TrkB receptors in lipid rafts. *J Neurosci*, **27,** 4859-4869.

11. Macchi, P., Villa, A., Giliani, S., Sacco, M.G., Frattini, A., Porta, F., Ugazio, A.G.,

Johnston, J.A., Candotti, F., O'Shea, J.J. *et al.* (1995) Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID). *Nature*, **377,** 65-68.

12. Russell, S.M., Tayebi, N., Nakajima, H., Riedy, M.C., Roberts, J.L., Aman, M.J., Migone, T.S., Noguchi, M., Markert, M.L., Buckley, R.H. *et al.* (1995) Mutation of Jak3 in a patient with SCID: essential role of Jak3 in lymphoid development. *Science*, **270,** 797-800.

13. Dupuis, S., Dargemont, C., Fieschi, C., Thomassin, N., Rosenzweig, S., Harris, J., Holland, S.M., Schreiber, R.D. and Casanova, J.L. (2001) Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. *Science*, **293,** 300-303.

14. Minegishi, Y., Saito, M., Tsuchiya, S., Tsuge, I., Takada, H., Hara, T., Kawamura, N., Ariga, T., Pasic, S., Stojkovic, O. *et al.* (2007) Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature*, **448,** 1058-1062.

15. Nussenzweig, A., Chen, C., da Costa Soares, V., Sanchez, M., Sokol, K., Nussenzweig, M.C. and Li, G.C. (1996) Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature*, **382,** 551-555.

16. Zhu, C., Bogue, M.A., Lim, D.S., Hasty, P. and Roth, D.B. (1996) Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell*, **86,** 379-389.

17. Zhang, J., Cado, D., Chen, A., Kabra, N.H. and Winoto, A. (1998) Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1. *Nature*, **392,** 296-300.

18. Yeh, W.C., Itie, A., Elia, A.J., Ng, M., Shu, H.B., Wakeham, A., Mirtsos, C., Suzuki, N., Bonnard, M., Goeddel, D.V. *et al.* (2000) Requirement for Casper (c-FLIP) in regulation of death receptor-induced apoptosis and embryonic development. *Immunity*, **12,** 633-642.

19. Soriano, P. (1994) Abnormal kidney development and hematological disorders in PDGF β-receptor mutant mice. *Genes Dev*, **8,** 1888-1896.

20. Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J. and Pickard, B.S. (2005) Speeding disease gene discovery by sequence based candidate prioritization. *BMC Bioinformatics*, **6,** 55.

21. Kohler, S., Bauer, S., Horn, D. and Robinson, P.N. (2008) Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*, **82,** 949-958.

22. Radivojac, P., Peng, K., Clark, W.T., Peters, B.J., Mohan, A., Boyle, S.M. and Mooney, S.D. (2008) An integrated approach to inferring gene-disease associations in humans. *Proteins: Structure, Function, and Bioinformatics,* **72,** 1030-1037.