AMLS Archives of Medical Laboratory Sciences

Original Article

Distribution of Disease-Causing Mutations through Different Protein Domains in Patients with Severe Combined Immunodeficiency

Zahra Shahbazi[1], Shirin Shahbazi[2], Hamzeh Rahimi[1], Reza Mahdian[1*]

1. Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran.
2. Department of Medical Genetics, School of Medical Sciences, Tarbiat Modares University, Tehran, Iran.

Received: July 4, 2019; Accepted: September 20, 2019

Abstract

Background and Aim: Severe combined immunodeficiency (SCID) has been described as the most severe form of primary immunodeficiency disorders (PID). The disease can be caused by mutations in more than 20 different genes with prevalence of 1 in 50000 to 100000 live births. In the present study, we described the protein domain position of variants in 14 main genes in patients with SCID. We also aimed to investigate the correlation between the variant distribution of protein domains and its pathogenicity and clinical outcome of the variant. **Materials and Methods:** Molecular genetic analysis including Sanger sequencing, targeted gene panel and whole exome sequencing were performed on 50 patients with SCID. Moreover, protein domains characteristics were extracted from different databases such as Uniprot and PDB and the reported mutations were obtained from HGMD and ENSEMBL databases. **Results:** Our results showed that the mortality rate had a significant correlation with severity of clinical manifestations in the patients (p-value=0.000). There was also a significant relationship between the protein type and mutation severity (p-value=0.001) and severity of clinical manifestations (p-value=0.025). However, there was no significant relationship between the mortality rate and occurrence of mutations in different domains of proteins (p-value=0.304) and the severity of mutations (p-value= 0.586). **Conclusion:** In severe genetic diseases such as SCID, mutations in related genes have affected the structure of the protein enough to cause severe symptoms. However, there are differences in the pathogenicity of the mutations based on their location on the protein domains. In order to determine these variations and predict the outcome of mutations, it is necessary to use in silico and laboratory methods along with statistical and data mining tools to track these minor differences.

Keywords: Primary Immunodeficiency; Severe Combined Immunodeficiency (SCID); Protein Domain; Variant Interpretation.

***Corresponding Author:** Reza Mahdian; Email: dr.reza.mahdian@gmail.com

Please cite this article as: Shahbazi Z, Shahbazi S, Rahimi H, Mahdian R. Distribution of Disease-Causing Mutations through Different Protein Domains in Patients with SCID. Arch Med Lab Sci. 2019;5(4):1-7. https://doi.org/10.22037/amls.v5i4.30891

Introduction

Severe combined immunodeficiency (SCID) comprises a heterogeneous group of primary immunodeficiency disorders (PIDs) associated with severe decline in T and/or B lymphocytes. The overall prevalence of SCID varies between 1 in 50000-100000 live births worldwide [1]. However, the actual number of PIDs cases is higher in the populations with a high rate of consanguineous marriage like Iran [2]. SCID disease commonly cause severe and repeated infections by opportunistic microorganisms, early onset skin rashes, cutaneous complications, persistent diarrhea, pneumonitis, oral candidiasis and failure to thrive (FTT) within the first year of life [3, 4]. Without immune reconstitution,

patients with SCID rarely survive beyond 6–12 months [5]. However, they usually show a successful response to allogeneic hematopoietic stem cell transplantation (HSCT) [6]. Several contributing genes to SCID have been described so far [7]. The most common genes are CD3ε/δ/ζ,, IL2RG, JAK3, DCLRE1C, RAG1/ RAG2, ADA, PNP [8]. In Iran, the disease shows an autosomal recessive hereditary pattern in most of the cases. However, in populations with low consanguineous marriage rates, the pattern is mainly x-linked and most of the SCID cases are boys with mutations in IL2RG gene [9].

Genetic variants can range from benign to severely pathogenic. The prediction of novel variants severity depends on various factors such as gene conservation, known pathogenic mutations, and protein-level annotations. The best known in silico assessment tools include SIFT (sorting intolerant from tolerant) [10], PolyPhen [11] and CADD (combined annotation dependent depletion), though the implication of recently introduced analysis methods can provide more reliable results [12]. It can be difficult to identify crucial residues for preserving the domain's stability and function for some proteins [13]. However, by evaluating the frequency of the previously reported variants in particular domains, one can realize which domains are commonly correlated with disease severity. This information can be of use in determining whether a particular variant is pathogenic or benign [14].

In the present study, we evaluated the protein domain distribution of SCID-causing mutations in 14 causal genes found in our patients' cohort. Our main goal was to study the correlation between mutations in different protein domains of SCID genes and severity of the disease.

Methods

Patients: The present study included 50 patients with SCID whose diagnosis was established according to the updated diagnostic criteria provided by the European Society for Immunodeficiencies (https://esid.org/Working-Parties/Registry/Diagnosis-criteria) based on physical examination findings, survey on infection history, thoracic radiology for thymus gland detection, complete blood cell count (CBC), lymphocyte transformation test (LTT) for analysis of the lymphocyte function, detection of CD3,4,8,16,19,56 marker levels in blood and determination of serum immunoglobulin level (15, 16). The clinical severity phenotype of patients with SCID was defined by having 2 of the following criteria: early-age onset of the symptoms (< 1 month), mortality (< 1 year), absent CD3+ or CD4+ or CD8 T+ cells, development of opportunistic infections, and development of severe infectious complications during the course of the disease (sepsis, central nervous system infections, osteomyelitis, and invasive bacterial infection) [17]. This study has been approved by the Ethics Committee of Pasteur Institute of Iran.

Genetic diagnosis: The genetic examination was carried out for the patients leading to molecular diagnosis. Defects were explored in different SCID-causing genes by Sanger sequencing, targeted gene panel (TGP) and whole exome sequencing (WES) [15, 18]. The pathogenicity of disease variants was re-assessed based on the updated guideline for interpretation of molecular sequencing presented by the American College of Medical Genetics and Genomics (ACMG) considering the allele frequency in the population database, immunological/functional data, familial segregation and parental genotype (https://www.acmg.net/). Prediction tools like SIFT, Polyphen and CADD were also used for prediction of mutation pathogenicity. Finally, the frame shift, nonsense, splice site and start losing mutations were considered as severe and missense variants as mild mutations. In order to find mutation distribution template in the previously reported variants, mutation histories were extracted from HGMD, ClinVar and Ensemble as genomic databases. The mutation severity was determined based on the effect of mutation on sequence and structure of proteins. Frame shift, nonsense, splicing and start loss mutations were considered as severe while non-frame shift and missense mutations as mild variants.

Distribution of founded mutations in protein domains: In the present study, Uniprot (https://www.uniprot.org/) and PDB (https://www.rcsb.org/) database were used to find protein structures and domains. Proteins were first categorized into enzyme and receptor groups. Enzymatic protein domains were further divided into catalytic and non-catalytic groups. The catalytic

domain contains the part of the protein that has the active site of the enzyme. Receptor protein domains were also divided into three categories: extracellular, transmembrane and intracellular.

Statistical analysis: Statistical analysis was performed using a commercially available software package (SPSS Statistics 22.0.0, SPSS, Chicago, Illinois). A p-value of <0.05 was considered to be statistically significant for all tests.

Results

Patients' characteristics and genetic diagnosis: Demographic, clinical and laboratory data related to all 50 patients are presented in Tables 1, 2 and 3, respectively. The clinical manifestations were severe in 26 and mild in 24 patients. The characteristics of the patients' genetic mutations have been described previously[15]. There were 44 different variants in the patients, 17 variants were previously reported in genomics databases (http://www.hgmd.cf.ac.uk, https://asia.ensembl.org) and 27 variants were novel variations. The pathogenic effect of the variants on corresponding proteins was severe in 27 and mild in 23 patients.

Distribution of the mutations in protein domains: As shown in figures 1 and 2, the mutations found in the present study are distributed in different domains of corresponding proteins. The gene mutations reported in this study including novel mutations are displayed alongside to previously reported mutations in each domain of the corresponding proteins. The frequency and distribution of variants in the protein domains were compared between previous reports and current study as displayed in Figure 3.

Correlation between pathogenicity of the variants and their protein domain location: The results of statistical analysis show that there is no significant relationship between the occurrence of mutations in different domains of the proteins and mortality rate ($\chi2$=4.845, df=1, and p-value=0.304(. No significant relationship was also found between the severity of mutations and the mortality rate ($\chi2$=0.297, df=1, and p-value=0.586) while mortality has a significant correlation with severity of clinical manifestations; people with mild clinical manifestations had higher mortality rates ($\chi2$=29.095, df=1, and p-value=0.000).

The relationship between the mutation severity and protein type was statistically significant. Mutations in proteins with enzymatic activity were often more severe than in receptor proteins ($\chi2$=12.013, df=1, and p-value=0.001). The relationship between protein type and the severity of clinical manifestations was also found to be significant in a way that people with mutations in genes with enzymatic activity displayed more severe clinical symptoms than those with mutations in receptor proteins ($\chi2$=5.024, df=1, and p-value=0.025).

Based on domains of enzymatic proteins, there is no significant relationship between mutation severity in different domains of enzymatic mutation ($\chi2$=3.601, df=1, and p-value=0.058) and receptor proteins ($\chi2$=1.702, df=1, and p-value=0.762) meaning that the occurrence of mutations in these domains does not cause a change in the severity of the mutation. This lack of relationship is also true about the severity of clinical manifestations as well; no significant correlation exists between the severity of clinical manifestations and the enzymatic protein domains ($\chi2$=1.298, df=1, and p-value=0.255) and receptor protein domains ($\chi2$=1.664, df=1, and p-value=0.197).

Table 1. Demographic data of the patients with SCID

Demographic Features		Number
Gender	**Male, number (%)**	29 (50)
	Female, number (%)	21 (50)
Age of Onset (Month)		4.5.00 (0.00-94)
Age of Diagnosis (Month)		5.00 (1.00-105)
Diagnostic Delay (Month)		3.50 (0.00-10.00)
Dead/Alive		38/50

Table 2. Clinical characteristics of the patients with SCID

Clinical Characteristics	Number (%)
Pneumonia, Number	17 (34%)
BCG-Osis, Number	14 (28%)
Oral candidiasis, Number	12 (24%)
FTT (Failure to thrive), Number	11 (22%)
Diarrhea, Number	19 (38%)
Skin infection, Number	5 (10%)
Hives, Number	2 (4)
Rash, Number	1 (2%)
Otitis, Number	2 (4%)
Urinary Tract Infections, Number	2 (4%)
Fever, Number	23 (46%)
LAP (Lymphadenopathy), Number	11 (22%)
Hepatomegaly, Number	12 (24%)
Splenomegaly, Number	7 (14%)

Table 3. Laboratory features of the patients with SCID

Title	Mean (Range)
CBC Test Results	
WBC(cell/ul)	5754.00 (2560-22305)
Neut %	42.00 (17.50-68.75)
Neut count	4145.00 (499.00-8778.0000)
Lymph %	27.05 (3.27-48.17)
Lymphocyte Count (%)	2197.62 (235.00-6850.88)
CD-markers	
CD3%	21.00 (0.00-62.75)
CD3 count	162.19 (0.00-3469.99)
CD4%	8.00 (0.00-38.71)
CD4 count	57.71 (0.00-1951.94)
CD8%	3.35 (0.00-38.90)
CD8 count	57.02 (0.00-2657.56)
CD16-56%	17.34 (0.31-45.50)
CD16 – 56 count	261.78 (2.90-1893.47)
CD19%	1.35 (0.14-78.80)
CD19 count	787.11 (1.15-4180.80)
Serum levels of Immunoglobulins	
IgG(mg/dl)	316.00 (36.50-800.00)
IgM(mg/dl)	56.00 (1.50-205.50)
IgA(mg/dl)	5.50 (0.50-237.00)
IgE(IU/ml)	23.70 (0.02-231.55)

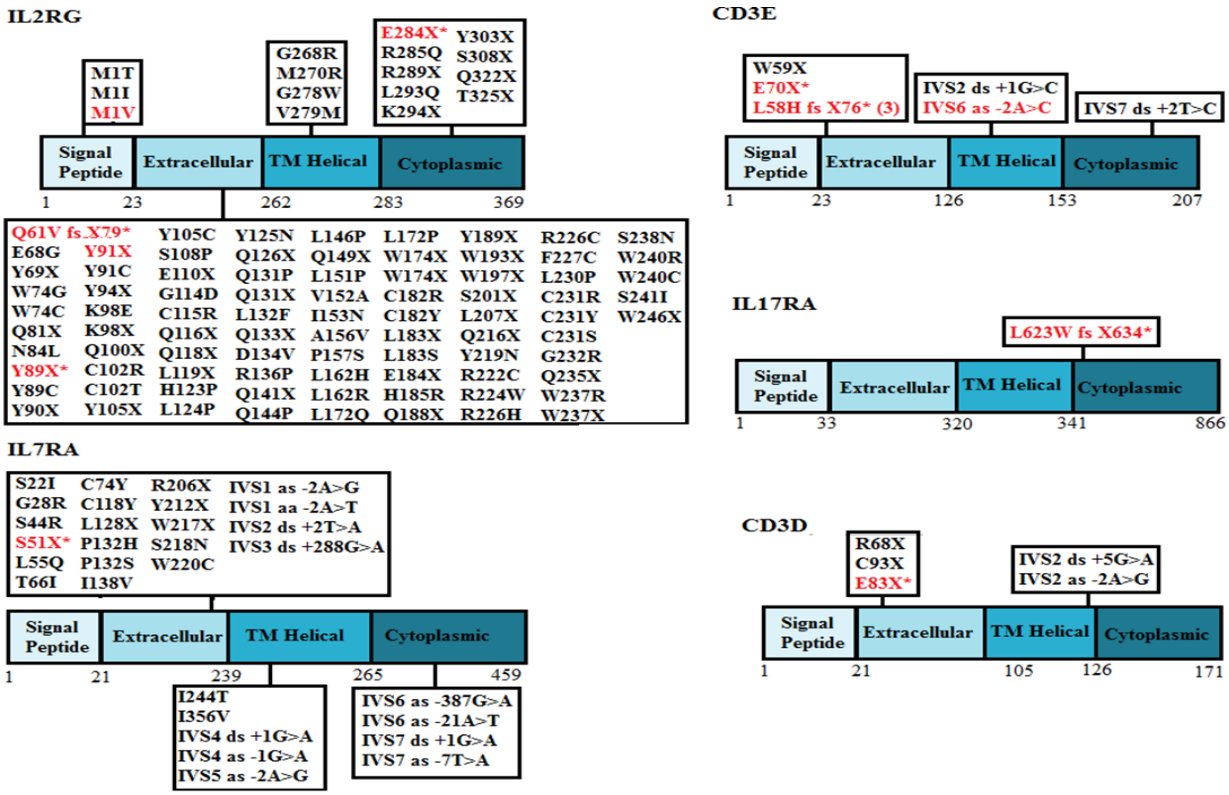

Figure 1. Mutation distribution in different domains of receptor proteins.
Black font: Previously reported mutations; Red font: Mutations found in the present study; *: novel mutations found in the present study.

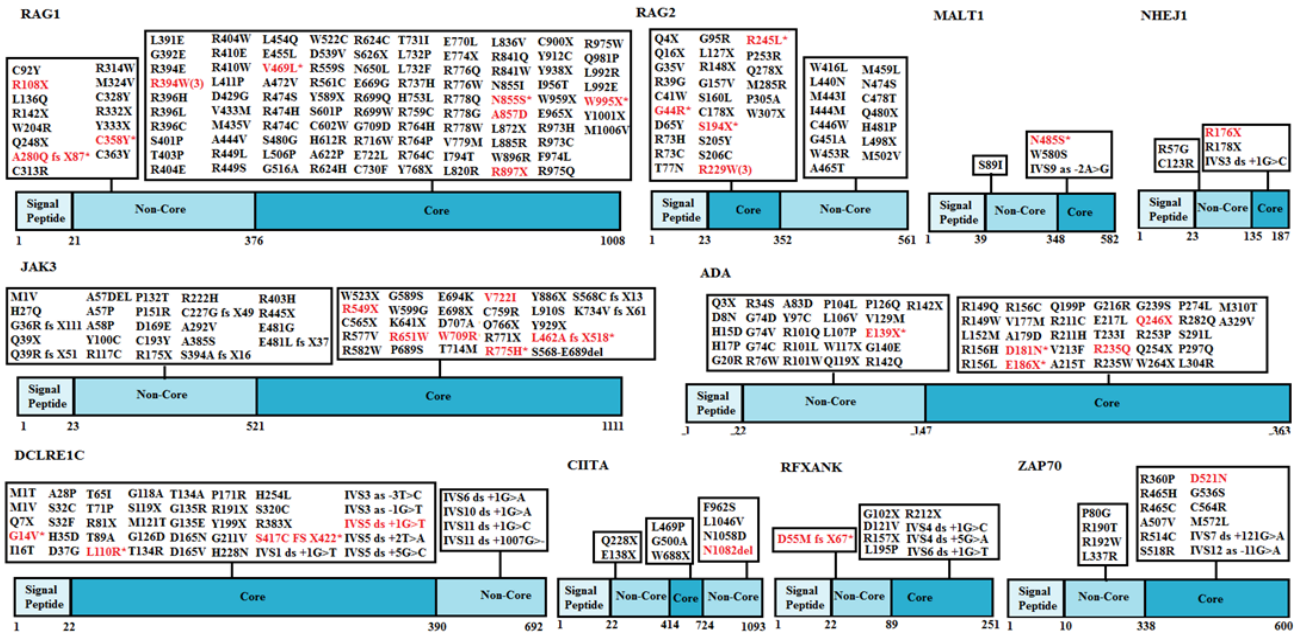

Figure 2. Mutation distributions in different domains of enzyme proteins.
Black font: Previously reported mutations; Red font: Mutations found in the present study; *novel mutations found in the present study.

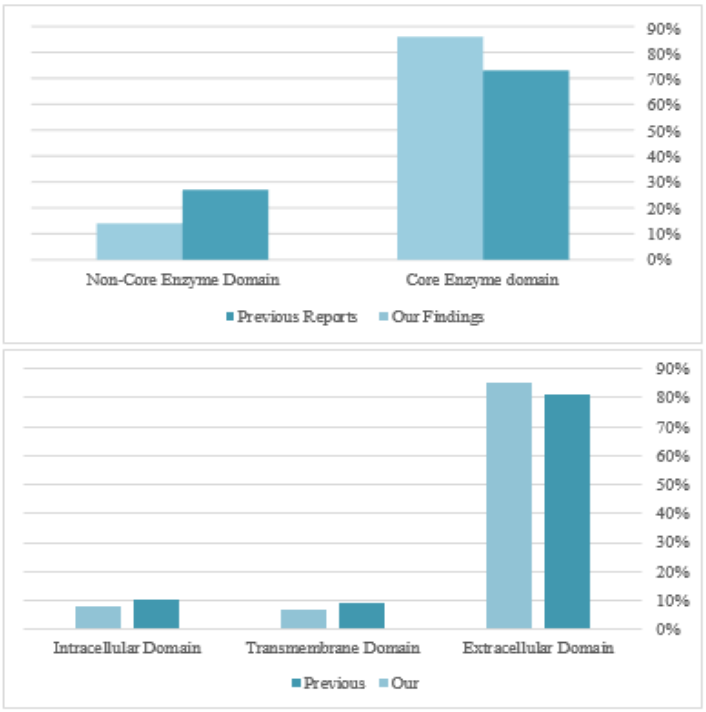

Figure 3. Frequency of mutations in different protein domains in the present study compared with previous reports

Discussion

In the present study, an attempt was made to investigate the mutation distribution in different protein domains and their consequences in clinical manifestations and survival prognosis of our patients with SCID.

Regarding the protein domains affected by gene mutations, genes with enzymatic roles such as ADA, DCLRE1C, JAK3, RAG1, RAG2 and ZAP70 are often mutated in catalytic or core enzyme domains, while genes acting as receptors such as CD3D, CD3E, IL7RA and IL2RG are often mutated in ligand binding or extra cellular domains (figures 1 and 2). These findings were in line with our expectations according to the HGMD (http://www.hgmd.cf.ac.uk) and Atlas-Genetics-Oncology (http://atlasgeneticsoncology.org) databases as well as the previous studies [19-22].

Investigation of mutations distribution in different protein domains can help in pathogenicity interpretation of the changes. Finding a mutation in a domain with several previously reported pathogenic mutations could be an indication of the mutation pathogenicity, as most pathogenic mutations occur in essential domains of proteins. On the other hand, one way to find important protein domains is to look at the domains with previously reported pathogenic mutations [19-22].

It may be inferred that proteins with enzymatic activity have a higher level of tolerance to missense variants. Some of the missense variants may be classified as SNPs and do not cause traceable clinical complications. This may be due to the maintenance of protein function despite conformational changes caused by missense variants. On the other hand, receptor proteins may be vulnerable to missense variants, which might justify why missense variants in receptor genes cause noticeable clinical complications. Although these variants may cause minor changes in the structure of the receptor protein, they may affect the protein function by changing their specific functional conformation.

The same explanation may be applied in interpreting the cause of increased severity of clinical manifestations in patients with enzymatic genes mutations. Due to the higher frequency of missense mutations in the receptor genes in the present study, there may be some residual activity in these proteins. This reduces the severity of manifestations in these patients compared to the patients with mutations in enzymatic genes because most of the mutations in the enzymatic genes were severe and led to a complete loss of function.

However, due to the small number of patients studied in the present study, no definitive conclusion can be drawn. In order to accurately investigate the effect of different variants on the structure of proteins, in silico studies are necessary considering the biological conditions of the living environment, the position of amino acids relative to each other and the characteristics of amino acids around the mutation.

Conclusion

Since prediction of mutation pathogenicity is a critical step in genetic counseling, carrier detection, and prenatal diagnosis, using appropriate and different tools in interpreting these variants can increase the accuracy and certainty of the conclusion. One of the best tools in this field is to investigate in which vital domain of the protein the mutation occurred. This tool is also used in the ACMG guidelines. However, in the present study, we concluded that in addition to statistical studies, it is necessary to conduct in silico studies to determine the effect of mutation location on the disease severity and outcome.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. Erman B, Bilic I, Hirschmugl T, Salzer E, Boztug H, Sanal Ö, et al. Investigation of genetic defects in severe combined immunodeficiency patients from Turkey by targeted sequencing. Scand J Immunol. 2017;85(3):227-34.

2. Puck JM, Group SNSW. Population-based newborn screening for severe combined immunodeficiency: steps toward implementation. J Allergy Clin Immunol. 2007;120(4):760-8.

3. Parvaneh N, Casanova J-L, Notarangelo LD, Conley ME. Primary immunodeficiencies: a rapidly evolving story. J Allergy Clin Immunol. 2013;131(2):314-23.

4. Sponzilli I, Notarangelo LD. Severe Combined Immunodeficiency (SCID): from molecular basisto clinical management. Acta Biomed. 2011;82(1):5-13.

5. Fischer A, Le Deist F, Hacein-Bey-Abina S, Andre-Schmutz I, Basile Gde S, de Villartay JP, et al. Severe combined immunodeficiency. A model disease for molecular immunology and therapy. Immunol Rev. 2005;203:98-109.

6. McGhee SA, Stiehm ER, McCabe ER. Potential costs and benefits of newborn screening for severe combined immunodeficiency. J Pediatr. 2005;147(5):603-8.

7. Kuo CY, Kohn DB. Gene therapy for the treatment of primary immune deficiencies. Curr Allergy Asthma Rep. 2016;16(5):39.

8. Kumrah R, Vignesh P, Patra P, Singh A, Anjani G, Saini P, et al. Genetics of severe combined immunodeficiency. Genes & Diseases. 2020;7(1):52-61.

9. Fazlollahi M, Pourpak Z, Hamidieh A, Movahedi M, Houshmand M, Badalzadeh M, et al. Clinical, Laboratory and Molecular Findings of 63 Patients with Severe Combined Immunodeficiency. A Decade s Experience. J Investig Allergol Clin Immunol. 2017;27(5):299-304.

10. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009;4(7):1073.

11. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics. 2013;76(1):7.20. 1-7.. 41.

12. Mather C, Salipante S, Scroggins S, Turner E, Shirts B, Wu D, et al., editors. Performance characteristics of Combined Annotation-Dependent Depletion (CADD) scores in clinical next-generation sequencing cancer panels. JOURNAL OF MOLECULAR DIAGNOSTICS; 2014: ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA.

13. Liu J, Rost B. Sequence-based prediction of protein domains. Nucleic acids research. 2004;32(12):3522-30.

14. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic acids research. 2017;45(D1):D190-D9.

15. Shahbazi Z, Yazdani R, Shahkarami S, Shahbazi S, Hamid M, Sadeghi-Shabestari M, et al. Genetic mutations and immunological features of severe combined immunodeficiency patients in Iran. Immunology letters. 2019;216:70-8.

16. Aghamohammadi A, Abolhassani H, Kutukculer N, Wassilak SG, Pallansch MA, Kluglein S, et al. Patients with primary immunodeficiencies are a reservoir of poliovirus and a risk to polio eradication. Frontiers in immunology. 2017;8:685.

17. Shahbazi Z, Parvaneh N, Shahbazi S, Rahimi H, Hamid M, Shahbazi D, et al. Graft versus host disease and microchimerism in a JAK3 deficient patient. Allergy, Asthma & Clinical Immunology. 2019;15(1):1-9.

18. Fang M, Abolhassani H, Lim CK, Zhang J, Hammarström L. Next generation sequencing data analysis in primary immunodeficiency disorders–future directions. Journal of clinical immunology. 2016;36(1):68-75.

19. Cossu F. Genetics of SCID. Ital J Pediatr. 2010;36(1):76.

20. Puck JM, Pepper AE, Henthorn PS, Candotti F, Isakov J, Whitwam T, et al. Mutation analysis of IL2RG in human X-linked severe combined immunodeficiency. Blood. 1997;89(6):1968-77.

21. Tasher D, Dalal I. The genetic basis of severe combined immunodeficiency and its variants. Appl Clin Genet. 2012;5:67.

22. Vihinen M, Villa A, Mella P, Schumacher RF, Savoldi G, O'Shea JJ, et al. Molecular modeling of the Jak3 kinase domains and structural basis for severe combined immunodeficiency. Clin Immunol. 2000;96(2):108-18.