International Journal of Sciences

Published online 01 15, 2022
ISSN 2763-5392

In silico analysis of the impact of non-synonymous Single Nucleotide Polymorphisms (nsSNPs) in the human Il-6 gene related to autoimmune diseases

Arthur Felipe Ferreira de Freitas[1]; Nara Suzy Aguiar de Freitas[2]; Martín Alejandro Montes[3]; Maria de Mascena Diniz Maia[4]*

1-4 Department of Biology, Genetics, Federal Rural University of Pernambuco, Recife, Brazil.

E-mail adresses: arthurffreitas16@gmail.com (Arthur Felipe Ferreira de Freitas), nara.safreitas@ufrpe.br (Nara Suzy Freitas Aguiar), martin.montes@ufrpe.br (Martín Alejandro Montes), mascenadiniz@hotmail.com (Maria de Mascena Diniz Maia)
*Corresponding author

To cite this article:

Freitas, A.F.F.; Freitas, N.S.A.; Montes, M.A.; Maia, M.M.D. ***In silico* analysis of the impact of non-synonymous Single Nucleotide Polymorphisms (nsSNPs) in the human Il-6 gene related to autoimmune diseases.** *International Journal of Sciences*. Vol. 3, No. 1, 2022, pp.01-05. ISSN 2763-5392.

Received: 11 24, 2021; **Accepted:** 11 25, 2021; **Published**: 01 15, 2022

Abstract: Non-synonymous single nucleotide polymorphisms (nsSNPs) represent a common genetic variation that alters amino acids encoded in proteins. All nsSNPs can potentially affect the structure or function of expressed proteins and therefore can have an impact on complex diseases. Interleukin 6 is a cytokine that acts on the immune system, with anti-inflammatory or pro-inflammatory actions depending on the cell in which it was produced. Overproduction of this cytokine results in complications in various autoimmune diseases, such as ankylosing spondylitis and rheumatoid arthritis. From genetic sequence research in the *Ensembl* database and software for protein analysis, it was possible to locate five missenses' mutations (rs1282201857, rs1583431936, rs1257406129, rs756681741 and rs2069849) that affect the expression of the IL6 gene, whose impact leads to the emergence of complex diseases.

Keywords: IL-6. Genetic polymorphisms. Autoimmune diseases. Bioinformatics. *In silico*. nsSNPs.

1. Introduction

Autoimmune diseases are those in which the response given to autoantigens, by the immune system, causes tissue damage or dysfunction. These diseases can act in a systemic way or may directly affect specific organs or systems of the body (Mackay and Burnet, 1963 *apud* NGO ST, *et al*., 2014). Sand develops in individuals with some genetic predisposition and who are exposed to environmental conditions favorable to the onset of the disease (WANG, L *et al*., 2015). Advances in scientific studies suggest that autoimmune diseases are induced through the interaction of genetic factors and epigenetic alterations that arise from the influence of the environment (HEWAGAMA; RICHARDSON, 2009). However, there are still not enough genetic tools to predict the risk of these diseases (WANG, L *et al*., 2015).

IL-6 is a pleiotropic cytokine that is pro-inflammatory because it induces the synthesis of acute phase proteins, besides stimulating the differentiation and clonal expansion of B lymphocytes, contributing to the production of immunoglobulins (Hunter CA, Jones AS, 2015). It is usually produced at low levels; however, it is related to numerous infectious, inflammatory, autoimmune diseases and in some types of cancer when its production is elevated to serum levels (TRIKHA, *et al*., 2003). Its synthesis is performed by several cell types, including endothelial cells, fibroblasts, monocytes and others (ROSSI, *et al*., 2015). The use of anti-IL6 therapy is an alternative for reducing inflammation caused by overproduction of this cytokine. Its blockade showed positive results in the therapeutic efficacy of the records of medicines used in Castleman's disease and inflammatory diseases such as rheumatoid arthritis and Crohn's disease (ROSSI; *et al*., 2015).

Single nucleotide polymorphisms (SNPs) are the main and most frequent forms of variations that occur in the human

genome. The estimate for it to appear in the genome is between 10 and 11 million, that is, it is common for SNPs to appear, on average, every 275 base pairs (bp) (PONTES; *et al*., 2017). SNPs are generally biallelic and have an average mutation rate that occurs in a given location of base pairs being significantly low (10–7 to 10–9). Thus, SNPs are mostly used for demographic analyses of the human population. (PONTES; *et al*., 2017).

The IL6 coding gene is located on the short arm of human chromosome 7 (7p22). This gene is also known as BSF2, HGF, HSF, IFNB2. The SNPs found in the interleukin 6 gene can directly affect the formation of amino acids which, depending on the mutation, can cause damage to the structure and/or function of the resulting protein. SNPs of this gene play a relevant role in the predisposition and progression to lung diseases, such as asthma, bronchitis, chronic obstructive pulmonary disease (COPD) and lung cancer (FERREIRA, *et al*., 2015). IL-6 polymorphisms are strongly related to the risk of developing Kaposi's sarcoma (KS), which is an inflammatory condition angio-proliferative, due to patients previously infected with the HIV virus. (FOSTER, *et al*., 2000). Interleukin-6 (IL 6) has been implicated in the pathogenesis of Ankylosing Spondylitis disease (ASD). The IL6 variant (rs1800796) seems to play an important role in the pathogenesis of Rheumatoid Arthritis (SILVA, 2018). Thus, this study aimed to provide information on the effects of amino acid substitutions on the protein structure of the IL6 gene, function and association with diseases.

2. Methodology

To perform the *in silico* analyses of the 21 missenses mutations(nsSNPs), some software was used to assist in the studies of the effects caused by changes in protein.

Obtaining the data

The *Ensembl* database was used to locate the interleukin 6 gene and isolate the missenses mutations of interest. On the main page of the site, the search frame was populated with Human and IL6 to locate the interleukin 6 gene in humans. After the location of the gene, several important information about IL6 was obtained, such as gene synonyms, chromosome location and transcripts available in the database. For this study, the transcript IL6-201 (ENST00000258743.10) was selected, which contains 1127bp in length and 212aa in protein. In the *bases* screens of the transcription, the item cDNA (complementary DNA) was accessed, being possible to visualize the changes of the gene sequences for and selection of the rs variants of interest for the study.

Analysis and review of the impact missenses mutations

The Software PolyPhen-2 were used to obtain data on possible damage caused by the protein mutation. We use the WHESS.DB tool for quick access to forecast sets deposited in the PolyPhen-2 database. To perform the research and locate the mutation, the protein code obtained in the UniProtKB (P05231) + the position of the mutation in the protein + the wild amino acid + the altered amino acid was inserted in the field (e.g.: P05231 2 N I). Also, in PolyPhen-2, we observed the following results: HumDiv, a rare allele evaluation model, and scores that can range from 0 to 1, and the closer to 1, the mutation is likely to cause harm to the individual.

The PROVEAN software was used as a tool that predicts whether an amino acid or indel substitution has an impact on the biological function of a protein, for this, the list of variants selected for this study was inserted. Thus, it was possible to verify whether the result of the mutation was deleterious (score $\geqslant$ -2.5) or neutral (score < -2.5). SIFT predictions were also verified in the same software, in which it was possible to classify mutations as tolerated (score $\geqslant$ 0.05) or harmful (score > 0.05).

The SNP&GO and PHD-SNP software's scans were used to predict whether the mutation would present disease or be neutral, taking into account the reliability index (IR), which can range from 0 (unreliable) to 10 (reliable), in both software. To make use of SNP&GO, simply fill the tables with the UniProt code of the protein, the position of the mutation, the wild residue and substitute. In PHD-SNP, it is necessary to fill with the original sequence of the protein, enter the code UniProt, the position of the mutation and the substitute residue.

The UniProt database was used to obtain the necessary information on the protein sequences of the IL6 gene in humans (UniProtKB P05231), as well as the graphic representation of the three-dimensional structure of this cytokine.

To evaluate the structure of the protein after the missense mutations in the fast sequence, the I-TASSER software was used, which is a tool capable of generating and determining the structure and function of the molecules of a protein using molds for simulation of assembly of the structures available in the RCSB PDB (Protein Data Bank) database. The system uses the TM-align structural alignment program to combine the first I-TASSER model with all structures in the PDB library.

Software's

ENSEMBL

It is a genome navigator for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation, and transcription regulation. *Ensembl* annotates genes, calculates multiple alignments, predicts regulatory function, and collects disease data. *Ensembl* tools include BLAST, BLAT, BioMart and variant effect predictor (VEP) for all supported species.

UNIPROT

Universal Protein Resource is a comprehensive resource for protein sequence and annotation data.

POLYPHEN-2

Polymorphism Phenotyping v2 is a tool that predicts the possible impact of the alteration of an amino acid on the structure and function of a human protein, using simple physical and comparative considerations.

PROVEAN

Protein Variation Effect Analyzer is a software tool that predicts whether an amino acid or indel substitution (nucleotide insertion or deletion mutation) has an impact on the biological function of a protein. It is useful for filtering sequence variants to identify non-synonymous or indels variants that are predicted to be functionally important.

SNPs & GO

SNPs & GO is a method for predicting unique amino acid polymorphisms (SNPs) deleterious using functional protein annotation probably involved in the insurgency of diseases in humans.

PHD-SNP

It is a software that predicts the unique nucleotide polymorphisms in humans.

I-TASSER

This software develops computational methods to predict the three-dimensional structure of protein molecules from the amino acid sequence and deduce biological functions based on the sequence-structure-function paradigm.

3. Results and Discussion

In the *Ensembl* platform, the presence of 170 SNPs variants was identified in the IL6-201 transcript (ENST00000258743.10) of the interleukin coding gene 6. For the present work, 21 non-synonymous variants (Table 1) were selected in order to observe the possible damages that these mutations could cause in protein structure and function.

The 21 variants were analyzed using different bioinformatics tools to obtain the results. PolyPhen-2 software was used to verify whether the selected variants of the IL6 gene would be classified as benign, possibly harmful, or likely to cause damage to the body. To verify whether the mutation was neutral or deleterious, the PROVEAN software was used and the SIFT was applied to observe whether the alteration would be harmful or tolerated by the organism. (Table 2)

In POLYPHEN-2, PROVEAN and SIFT software (Table 2), rs200700194, rs1282201857, rs1583431936, rs934459268, rs199990564, rs755968353, rs125847874, rs1257406129, rs756681741 and rs2069849 present scores that indicate possibilities of risk in protein formation, which may result in diseases. However, from the analysis of these results in the SNP&GO and PHD-SNP, it was possible to associate only five polymorphisms in the gene with significant results for disease, rs1282201857, rs1583431936, rs1257406129, rs756681741 and rs2069849 (Table 3). Due to the high reliability index (IR), it is possible to deduce that amino acid substitutions with different chemical composition and/or sizes have a greater impact on protein function and structure. Recent studies reveal that the C allele of rs2069849 in interaction with abdominal obesity were directly associated with a higher risk of osteoporosis (Ji YF, *et al.*, 2019).

Analyses in 3D structures corroborate the results of the algorithms, due to the presence and impact of such mutations in conserved regions, important in the function and structure of the protein.

Figure 1 shows the three-dimensional graphic representation of the position of altered amino acids in the protein. RI represents the Reliability Index.

Table 1. Selection of 21 nsSNPs variants of the human IL6 gene for *in silico* analysis

dbSNP rs	Consequences	Gene	Code UniProt	Position in protein	Amino acid
RS200700194	Missense variant	IL6	P05231 2 N I	2	N/I
RS1224052258	Missense variant	IL6	P05231 5 S F	5	S/F
RS774003579	Missense variant	IL6	P05231 6 T I	6	T/I
RS1212373109	Missense variant	IL6	P05231 7 S G	7	S/G
RS1383764310	Missense variant	IL6	P05231 9 F S	9	F/S
RS762283178	Missense variant	IL6	P05231 10 G A	10	G/A
RS767924065	Missense variant	IL6	P05231 13 A S	13	A/S
RS760962452	Missense variant	IL6	P05231 14 F L	14	F/L
RS1282201857	Missense variant	IL6	P05231 17 G V	17	G/V
RS1583431936	Missense variant	IL6	P05231 21 V G	21	V/G
RS1335707674	Missense variant	IL6	P05231 23 P S	23	P/O
RS746848366	Missense variant	IL6	P05231 24 A P	24	A/P
RS1485376559	Missense variant	IL6	P05231 30 V I	30	V/I
RS142759801	Missense variant	IL6	P05231 31 P T	31	P/T
RS934459268	Missense variant	IL6	P05231 83 E G	83	I/O
RS199990564	Missense variant	IL6	P05231 87 E G	87	I/O
RS755968353	Missense variant	IL6	P05231 91 N Y	91	N/Y
RS1258447874	Missense variant	IL6	P05231 95 M T	95	M/T
RS1257406129	Missense variant	IL6	P05231 100 G R	100	G/R
RS756681741	Missense variant	IL6	P05231	112	L/V

			112 L V		
RS2069849	Missense variant	IL6	P05231 201 F L	201	F/L

Table 2. Comparison of *in silico* analysis of 21 snSNPs mutations in the human IL6 gene

dbSNPrs	Aminoacid Subs.	PolyPhen-2	SCORE	PROVEAN	SCORE	SIFT	SCORE
RS200700194	N2I	Possibly harmful	0.912	Deleterious	-3,53	Harmful	0,003
RS1224052258	S5F	Benign	0,229	Neutral	-2,13	Allowable	0,136
RS774003579	T6I	Probably cause damage	0,997	Neutral	-2,33	Harmful	0,002
RS1212373109	S7G	Possibly harmful	0,816	Neutral	-1,01	Allowable	0,088
RS1383764310	F9S	Benign	0,416	Deleterious	-3,55	Harmful	0,001
RS762283178	G10A	Benign	0,003	Neutral	0,04	Allowable	0,365
RS767924065	A13S	Probably cause damage	0.999	Neutral	-1,69	Harmful	0,009
RS760962452	F14L	Possibly harmful	0,894	Neutral	-2,23	Allowable	0,173
RS1282201857	G17V	Probably cause damage	1	Deleterious	-4,71	Harmful	0,002
RS1583431936	V21G	Probably cause damage	0.999	Deleterious	-3,78	Harmful	0
RS1335707674	P23S	Benign	0.176	Neutral	0,48	Harmful	0,032
RS746848366	A24P	Possibly harmful	0.828	Neutral	-1,15	Harmful	0,003
RS1485376559	V30I	Possibly harmful	0.665	Neutral	-0,47	Allowable	0,161
RS142759801	P31T	Benign	0.009	Neutral	-2,18	Allowable	0,767
RS934459268	E83G	Probably cause damage	0.999	Deleterious	-3,5	Harmful	0,018
RS199990564	E87G	Possibly harmful	0.553	Deleterious	-2,95	Harmful	0,029
RS755968353	N91Y	Probably cause damage	0.997	Deleterious	-3,96	Harmful	0,007
RS1258447874	M95T	Probably cause damage	0.994	Deleterious	-3,03	Harmful	0,002
RS1257406129	G100R	Probably cause damage	0.994	Deleterious	-5,37	Harmful	0,02
RS756681741	L112V	Probably	0.996	Deleterious	-2,59	Harmful	0,006
RS2069849	F201L	Probably cause damage	0.997	Deleterious	-4,10	Harmful	0

Table 3. Prediction of missense nsSNPs, human IL-6 gene, disease-causing using the web-based tools, SNPs & GO and PHD-SNP.

dbSNP rs	Amino acid Subs	SNPs & GO	LAUGH	PHD-SNP	LAUGH
RS1282201857	G17V	Sickness	8	Sickness	3
RS1583431936	V21G	Sickness	4	Sickness	6
RS1257406129	G100R	Sickness	6	Sickness	1
RS756681741	L112V	Sickness	5	Sickness	3
RS2069849	F201L	Sickness	4	Sickness	4

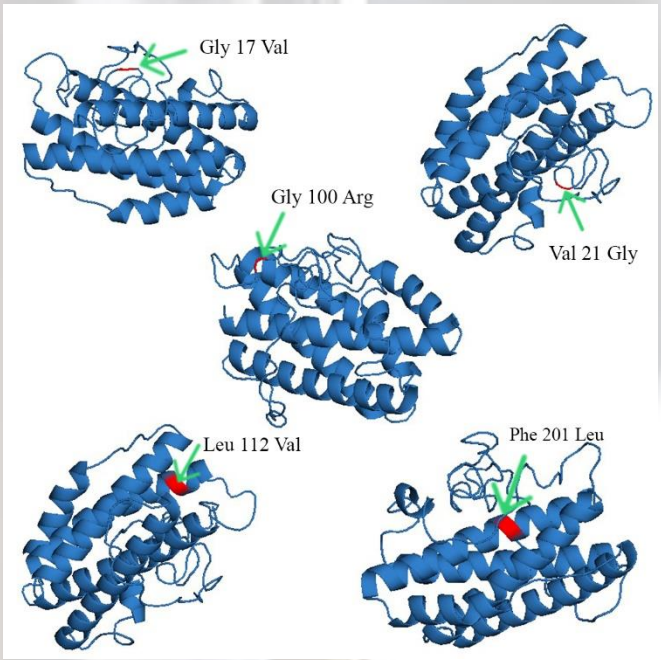

Figure 1. Localization of the position of altered amino acids in the protein due to the missense mutations of the IL6 gene considered in this study that cause diseases.

4. Conclusions

We conclude that computational algorithms carefully validated in the context of other evidence can be an important tool for the classification of missense variants. We can also observe that gene polymorphisms may be correlated with various human diseases and can be considered good therapeutic markers.

5. Acknowledgments

To the Federal Rural University of Pernambuco for its support.

References

[1] NGO, S.T.; STEYN, F.J.; MCCOMBE, P.A. Gender differences in autoimmune disease, Frontiers in Neuroendocrinology, v.35, n.3, p. 347-369, ago. 2014. Disponível em: https://www.sciencedirect.com/science/article/pii/S0091302214000466?via%3Dihub. Acesso em: 05 out. 2021.

[2] 5WANG, Lifeng; WANG, Fu?Sheng; GERSHWIN, M. Eric. Human autoimmune diseases: a comprehensive update. Journal of internal medicine, v.278, n.4, p.369-395, 2015. Disponível em: https://onlinelibrary.wiley.com/doi/10.1111/joim.12395. Acesso em: 05 out. 2021.

[3] HEWAGAMA, Anura; RICHARDSON, Bruce. The genetics and epigenetics of autoimmune diseases. Jounal of Autoimmunity, v.33, n.1, p.3-11. ago. 2009. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0896841109000523?via%3Dihub. Acesso em: 05 out. 2021.

[4] 6HUNTER, C. A.; JONES, S. A. IL-6 as a keystone cytokine in health and disease. Nature Immunology. v.16, n.5, p. 448-457, 2015. Disponível em: https://www.nature.com/articles/ni.3153#:~:text=IL%2D6%20represents%20a%20keystone,location%2C%20location%2C%20location'. Acesso em: 20 set. 2021

[5] TRIKHA, Mohit. *et al*. Targeted Anti-Interleukin-6 Monoclonal Antibody Therapy for Cancer. Clinical cancer research, v.9, n.13, p. 4653-4665, 15 out. 2003. Disponível em: https://clincancerres.aacrjournals.org/content/9/13/4653.long. Acesso em: 05 out. 2021.

[6] ROSSI, Jean-François. *et al*. Interleukin-6 as a therapeutic target. Clinical cancer research, v.21, n.6, p. 1248-1257. mar. 2015. Disponível em: https://clincancerres.aacrjournals.org/content/21/6/1248.long. Acesso em: 05 out. 2021.

[7] Pontes, L., Sousa, J., & Medeiros, R. SNPs and STRs in forensic medicine. A strategy for kinship evaluation. Archives of Forensic Medicine and Criminology, v.67, n.3, p.226-240, mar. 2017. Disponível em: https://www.termedia.pl/SNPs-and-STRs-in-forensic-medicine-A-strategy-for-kinship-evaluation,82,31659,1,1.html. Acesso em: 05 out. 2021.

[8] FERREIRA, Michele Berger. *et al*. Polimorfismo no gene da interleucina-6 e sua relação com suscetibilidade às doenças pulmonares em crianças e adolescentes. Revista Jovens Pesquisadores, Santa Cruz do Sul, v. 5, n. 1, maio 2015. p.2237-048. Disponível em: <https://online.unisc.br/seer/index.php/jovenspesquisadores/article/view/5774>. Acesso em: 15 set. 2021.

[9] FOSTER, Charles B., *et al*. An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency vírus. Blood, v.96, n.7, p.2562-2567, out. 2000. Disponível em: https://www.sciencedirect.com/science/article/pii/S0006497120543693. Acesso em: 05 out. 2021.

[10] SILVA, Marília Lins e. avaliação dos níveis séricos e polimorfismos genéticos de citocinas envolvidas com as respostas th 1, th 2 e th 17 em indivíduos com artrite reumatoide. ATTENA Repositório Digital da UFRPE, Recife, 2018. Disponível em: https://repositorio.ufpe.br/handle/123456789/32903. Acesso em: 20 set. 2021.

[11] Ji YF, Jiang X, Li W, Ge X. Impact of interleukin-6 gene polymorphisms and its interaction with obesity on osteoporosis risk in Chinese postmenopausal women. Environmental Health and Preventive Medicine. 24, 48, 13 de julho de 2019. Disponível em: https://environhealthprevmed.biomedcentral.com/articles/10.1186/s12199-019-0803-y#citeas. Acesso em: 20 set. 2021.