LUMEN ET VIRTUS
ISSN: 2177-2789

GENOMIC IDENTIFICATION OF NON-SYNONYMOUS VARIANTS OF THE IL1A GENE ASSOCIATED WITH HUMAN DISEASES BY BIOINFORMATICS ANALYSIS

https://doi.org/10.56238/levv15n41-105

Submitted on: 29/09/2024 **Publication date:** 29/10/2024

Arthur Felipe Ferreira de Freitas[1], Nara Suzy Aguiar de Freitas[2], Maria Helena Queiroz de Araújo Mariano[3], Eliezer Rushansky[4] and Maria de Mascena Diniz Maia[5]

ABSTRACT

Interleukin-1 alpha is a cytokine that stands out for its essential pro-inflammatory role that works by activating key cells to fight infections. Some genetic mutations can promote the replacement of amino acids in this protein, leading to consequences that may be associated with diseases. Thus, this work sought to identify Single Nucleotide Polymorphisms (SNPs) of high risk to human health and predict their morphofunctional effects on the protein. The variants were located in the IL1A-201 transcript, available in the ENSEMBL genomic database. After selecting the missense variants, we used the SIFT, PolyPhen-2, MetaLR and Mutation Assessor programs to predict the impacts of amino acid substitutions on protein function and structure. Our analysis revealed the presence of the rs1190431689 variant, which is the change from an Alanine to a Valine, the rs997926639, which corresponds to the change from a Proline to a Serine, and the rs1681201328 variant, which represents the replacement of a Glutamic Acid by a Glycine. All these polymorphisms presented Deleterious (SIFT), Probably Harmful (PolyPhen-2) and Harmful (MetaLR) results in the prediction software, which indicates that these are variants that can be of high risk to human health and can be associated with the development of several pathologies.

Keywords: Cytokine. Genetic polymorphisms. Prediction in Silico.

[1] Undergraduate student in Biological Sciences
Federal Rural University of Pernambuco
E-mail: arthur.ffreitas@ufrpe.br
[2] Dr. in Genetics from the Federal University of Pernambuco
Federal Rural University of Pernambuco
E-mail: nara.safreitas@ufrpe.br
[3] Dr. in Biological Sciences from the Federal University of Pernambuco
University of Pernambuco
E-mail: irpe.diretoria@gmail.com
[4] Post-graduation in Internal Medicine from the University of Pernambuco
University of Pernambuco
E-mail: eliezer.rushansky@upe.pe.gov.br
[5] Dr. in Biological Sciences from the Federal University of Pernambuco
Federal Rural University of Pernambuco
E-mail: maria.dmaia@ufrpe.br

INTRODUCTION

The human immune system is composed of a complex network of cells and molecules that act synergistically to protect the body against exogenous invaders. Within this context, the innate immune system, represented by components such as neutrophils and barrier epithelial cells, plays a key role as the first line of defense against pathogens (Heeb *et al*., 2020). However, the immune response is not limited to innate immunity alone, but is complemented by the adaptive immune system, which ensures a more specific and memory response against recurring threats (Heeb *et al.,* 2020).

From this perspective, cytokines, which are crucial signaling molecules in the immune system, act as intercellular mediators that regulate inflammation, immune response, and various physiological and pathological processes (Varella, 2001). These proteins, produced by a variety of cells, such as macrophages, T cells, and epithelial cells, play a central role in communication between immune cells and in the coordination of both innate and adaptive immune responses (Freitas *et al*., 2022).

Within the class of cytokines, interleukin-1 alpha (IL-1α) stands out for its role as an essential pro-inflammatory cytokine. IL-1α, together with IL-1 beta (IL-1β), is part of the IL-1 family, which also includes the antagonist protein IL-1ra (Moreira *et al.*, 2006). These cytokines are critical in mediating inflammatory responses, being produced primarily by macrophages, epithelial cells, and other cell types in response to signs of stress or infection (Graves *et al*., 1990; Moreira *et al*., 2006). IL-1α, specifically, is a protein that acts both autocrinely, that is, acting on the cells that produce it, and paracrine, acting on the cells neighboring those that produce it, facilitating cell signaling in inflammatory processes and participating in the activation of lymphocytes and neutrophils, key cells in the fight against infections (Becerra *et al*., 2023).

From a molecular point of view, *IL-1α* is encoded by the IL1A gene, located on human chromosome 2, and synthesizes a protein consisting of a chain of 271 amino acids, presenting a highly evolutionarily conserved structure (Moreira *et al.,* 2006). Structural studies, such as those carried out by Graves *et al.* (1990), reveal that IL-1α is formed by a nucleus of antiparallel beta sheets organized in a beta barrel, an essential structural feature for its stability and interaction with cell receptors. This complex molecular organization allows IL-1α to bind to specific receptors on the surface of target cells, triggering a signaling cascade that amplifies the inflammatory response.

Not only does IL-1α play a vital role in the activation and regulation of immune responses, but it is also associated with several pathological conditions, such as autoimmune diseases and chronic inflammation (Kubaski *et al.*, 2013). Therefore, mutations

in genes that encode immune proteins, such as IL1A, is a field of genomic investigation that has revealed the positive correlation between genetic variations and the development of diseases (Kubaski *et al.*, 2013).

The analysis of single nucleotide polymorphisms *(SNPs)* is fundamental for understanding the molecular consequences of amino acid substitutions and their clinical implications. In this context, the present work aimed to perform a comprehensive analysis of the genetic variants of the IL1A gene available in biological databases, with the objective of identifying and predicting their morphofunctional effects on the protein.

MATERIAL AND METHODS

DATASET

Genetic data were collected from the *Ensembl* platform, a publicly accessible repository that integrates reference genomic data (Harrison et al., 2024). The IL1A-201 transcript (ENST00000263339.4) was selected as the basis for the analysis, which allowed us to obtain information on the nucleotide sequence, protein structure, and interleukin-1 alpha (IL-1α) domains. In addition, we used the database to examine Single Nucleotide Polymorphisms (*SNPs)* associated with the *IL1A gene.*

PREDICTION OF THE IMPACT OF VARIANTS

The functional impact of the identified nonsynonymous variants was predicted using a robust set of bioinformatics tools, which evaluate amino acid substitutions based on structural, functional, and evolutionary criteria:

- *SIFT (Sorting Intolerant From Tolerant):* Evaluates the intolerance of substituted amino acids, based on sequence homology and physicochemical characteristics. It classifies variants as "Deleterium" (score 0) or "Tolerated" (score > 0.05).
- *PolyPhen-2 (Polymorphism Phenotyping v2):* Estimates the probability of functional impact of variants, categorizing them as "Probably harmful" (score 0.85–1.0), "Possibly harmful" (score 0.15–0.85), or "Benign" (score < 0.15).
- *MetaLR:* Predicts the degree of pathogenicity of variants based on machine learning models, classifying them as "Harmful" (score close to 1) or "Benign" (score close to 0).
- *Mutation Assessor:* Evaluates the impact of variants taking into account the evolutionary conservation of affected positions on protein homologues. Classifies variants as "Neutral", "Low", "Medium", or "High", based on *the functional impact* score.

RESULTS

The IL1A-203 transcript available at *ENSEMBL* has a total of 250 non-synonymous variants, of which 5 are in the *splicing* region of the gene. In addition, 8 variants that introduce a stop codon (*Stop gained*), 7 variants with a frameshift *variant (frameshift variant),* 2 variants that lose a stop codon (*Stop lost),* 2 variants that lose an initiation codon *(Start lost),* 1 insertion that maintains the read phase (*inframe insertion),* 3 deletions that maintain the read phase (*inframe deletion*), 14 variants in the *splicing region,* 1 variant that maintains the stop codon (*Stop retained),* 91 synonymous variants, and 110 variants in the coding sequence.

The classification of *the Missense variants* is shown in Table 1 according to the computer predictors used. The analysis of the variants in the SIFT resulted in 113 *Missense variants* classified as deleterious, however, two of these predictions have low reliability. In addition, 137 variants were considered to have tolerated effects. In the PolyPhen-2 software, 42 variants were classified as Likely Harmful, 47 as Possibly Harmful, and 36 as benign using MetaLR, 3 variants were considered harmful, while 247 were classified as benign. In the *Mutation Assessor,* 89 variants had a medium functional impact on the protein, 123 had a low impact, and 38 were classified as neutral.

Table 2 shows the non-synonymous variants with the greatest potential impact on the protein, filtered based on the results of the prediction programs, resulting in the identification of three SNPs. The rs1190431689 variant corresponds to the substitution of Alanine (Ala) for Valine (Val) at position 2 of the polypeptide chain. This variant was classified as Deleterious (score: 0) in SIFT, Probably Harmful (score: 0.921) in PolyPhen-2, Harmful (score: 0.547) in MetaLR and Medium Impact (score: 0.553) in Mutation Assessor. The rs997926639 variant refers to the substitution of Proline (Pro) for Serine (Ser) at position 5 of the amino acid sequence, with impacts predicted as Deleterious (score: 0) in SIFT, Probably Harmful (score: 0.961) in PolyPhen-2, Harmful (score: 0.683) in MetaLR, and Medium Impact (score: 0.553) in Mutation Assessor. The rs1681201328 variant, located in the *splicing region,* involves the replacement of Glutamic Acid (Glu) with Glycine (Gly) at position 106 of the protein sequence, with predicted impacts such as Deleterious (score: 0) in SIFT, Probably Harmful (score: 0.917) in PolyPhen-2, Harmful (score: 0.503) in MetaLR and Medium Impact (score: 0.553) in Mutation Assessor (Tables 2 and 3).

Table 3 presents the analysis of the *Missense variants* inserted in the *splicing* region of the *IL1A gene.* The rs200686721 variant, which involves replacing Serine (Ser) with Asparagine (Asn) in position 16, was classified as "Tolerated" by the SIFT, with a score of 0.05. On PolyPhen-2, this variant was rated as "Possibly Harmful" with a score of 0.794.

MetaLR also classified it as "Tolerated", with a score of 0.406. The Mutation Assessor assigned a "Medium" impact, with a score of 0.553. The rs1350937353 variant, where Glutamine (Gln) is replaced by Histidine (His) in position 32, presented divergent classifications among the tools. In the SIFT, it received the classification "Deleterium" with a score of 0, indicating a significant negative impact on the protein's function. PolyPhen-2 rated it as "Likely Harmful," with a score of 0.936, pointing to a substantial adverse effect. In contrast, MetaLR considered it "Tolerated" with a score of 0.306, and Mutation Assessor rated it with an "Average" impact of 0.553. The rs1681201211 variant, in which Glutamic Acid (Glu) is replaced by Lysine (Lys) at position 107, also received varying ratings. In the SIFT, it was considered "Deleterium" with a score of 0.03, indicating a significant negative impact. PolyPhen-2 rated it as "Possibly Harmful" with a score of 0.879. In contrast, MetaLR rated it as "Tolerated" with a score of 0.374, and Mutation Assessor indicated a "Low" impact with a score of 0.553. The rs777688692 variant, which involves the replacement of Lysine (K) with Arginine (R) in position 205, was classified as "Deleterium" by SIFT, with a score of 0.03, but received a "Tolerated" rating from MetaLR with a score of 0.154 and "Low" impact by Mutation Assessor. No PolyPhen-2 rating was provided for this variant.

Table 1. Classification of *Missense variants* of the human *IL1A* gene and their prediction of functional impact using different bioinformatics tools.

SIFT	How much	PolyPhen-2	How much	MetaLR	How much	Mutation Assessor	How much
Deleterious	111	Probably harmful	42	Harmful	3	Medium	89
Deleterious (Low reliability)	2	Possibly harmful	47	Benign	247	Low	123
Allowable	137	Benign	36	-	-	Neutral	38
Total	250		250		250		250

Table 2. Missense variants of the *IL1A* gene predicted to have a high risk for the development of human diseases.

Variant Missense	Residue	Prediction			
		SIFT	PolyPhen-2	MetaLR	Mutation Assessor
rs1190431689	2 Ala/Val	Deleterious 0	Probably harmful 0,921	Harmful 0,547	Medium 0,553
rs997926639	5 Pro/Ser	Deleterious 0	Probably harmful 0,961	Harmful 0,683	Medium 0,553
RS1681201328*	106 Glu/Gly	Deleterious 0	Probably harmful 0,917	Harmful 0,503	Medium 0,553

*Variant located in the *splicing region* of the gene.

Table 3. Prediction of *Missense variants* in the *splicing region.*

Variant	Residue	Prediction			
		SIFT	PolyPhen-2	MetaLR	Mutation Assessor
rs200686721	16 S(Ser)/ N (Asn)	Allowable 0,05	Possibly harmful 0,794	Allowable 0,406	Medium
rs1350937353	32 Q(Gln)/ H(His)	Deleterious 0	Probably harmful 0,936	Allowable 0,306	Medium
rs1681201328	106 E(Glu) / G(Gly)	Deleterious 0	Probably harmful 0,917	Harmful 0,683	Medium
rs1681201211	107 E(Glu)/ K(Lys)	Deleterious 0,03	Possibly harmful 0,879	Allowable 0,374	Low
rs777688692	205 K (Lys)/R(Arg)	Deleterious 0,03	-	Allowable 0,154	Low

DISCUSSION

The three-fold internal three-dimensional symmetry of the structural core is remarkably accurate and with high fidelity (Graves *et al*., 1990). Their structural stability is a major concern because only stable proteins can perform their function efficiently (Kulshreshtha *et al*., 2016). Therefore, changes in amino acids with very different physicochemical conformities can result in the destabilization of the molecule and affect the correct performance of *IL-1α* in the body.

By examining the substituted amino acids in each available variant, we can infer that the predictions of rs1190431689 are highly harmful, since the replacement of the amino acid Alanine (Ala) by a Valine (Val) can have severe impacts on the protein structure (Guzzi, 2018). Although both molecules are hydrophobic and have similar chemical structures, which may not significantly affect the three-dimensional structure of the protein or its biological function, the change of an amino acid can alter the physical properties of the protein and affect its biological activity (Redler et al., 2016.). Valine, for example, is a longer and bulkier side chain than alanine, which can affect the protein's conformation, interaction with other proteins or receptors, and its ability to be processed and secreted by producing cells.

The rs997926639 variant is the result of the substitution of an amino acid Proline (Pro) for Serine (Ser) and can have different impacts, depending on the context in which it occurs. Proline is a non-polar amino acid and its replacement with a serine, which is a polar amino acid, can alter the three-dimensional structure of the protein and affect its biological function. The presence of proline in the protein structure helps stabilize the alpha helix, which is an important secondary structure in protein. Serine substitution can disrupt this stabilization and affect the three-dimensional structure of the protein. In addition, serine has a hydroxyl side chain that can alter the local polarity of the protein and affect its interaction with other molecules. Regarding biological function, studies have shown that replacing proline with serine in *IL-1α* can affect its ability to bind to specific receptors and modulate the inflammatory response. https://www.studocu.com/pt-br/document/instituto-federal-de-educacao-ciencia-e-tecnologia-do-rio-grande-do-norte/biologia/tipos-de-mutacao/17397090. Replacing glutamic acid (Glu), an amino acid with a long side chain, with glycine (Gly), a neutral amino acid with a very short side chain, can have a significant impact. Glutamic acid has a carboxyl group on its side chain, which contributes to the negative charge of the protein and can interact with other molecules in a specific way. Glycine, being very small and neutral, can alter the protein's charge and ability to interact. The change can affect the stability of the protein's three-dimensional structure, as well as its biochemical and functional properties.

With these results, it is evident that prediction *software* can be extremely useful for understanding various pathologies, as they allow the identification of variants associated with diseases and how they can interfere with molecular stability (Freitas *et al*., 2022). In addition, we recommend that future studies be directed to understanding the evolutionary history of the *IL1A gene* to determine the degree of conservation of the gene and its evolutionary importance for human health, since knowing the evolution of interleukins

allows the understanding of the origin and regulatory mechanisms of several pathologies associated with cytokines in humans (Freitas *et al*., 2024).

CONCLUSION

This work sought to study and identify the *Missense mutations* associated with the human *IL1A gene,* highlighting variants with high risk for the development of human diseases and understanding their possible physicochemical impacts on the molecule and its effects on the performance of the cytokine in the body. Thus, serving as a basis to assist new studies in the identification of genes of interest associated with the development of human diseases.

REFERENCES

1. Becerra, C. R., et al. (2023). A phase I study of isunakinra, an IL-1 alfa/beta inhibitor, in combination with nivolumab for patients with solid tumors refractory to standard therapies.

2. Freitas, A. F. F., et al. (2022). In silico analysis of the impact of non-synonymous Single Nucleotide Polymorphisms (nsSNPs) in the human Il-6 gene related to autoimmune diseases. DOI:10.29327/229003.3.1-1.

3. Freitas, A. F. F., de Freitas, N. S. A., Mariano, M. H. Q. de A., Rushansk, E., & Maia, M. de M. D. P. (2024). Padrões evolutivos da interleucina-6 (IL-6) e seu impacto para a saúde humana. Brazilian Journal of Health Review, 7(1), 4624–4637. DOI: 10.34119/bjhrv7n1-374. Disponível em: https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/66998. Acesso em: 14 out. 2024.

4. Graves, B. J., et al. (1990). Structure of interleukin 1. alpha. at 2.7-. ANG. resolution. Biochemistry, 29(11), 2679-2684.

5. Guzzi, A. F. (2018). Caracterização in silico do polimorfismo A122V na proteína receptora de quimiocina do tipo 1 (CXCR1) associado com suscetibilidade à mastite em bovinos. (Thesis). VETTESES. ID: vtt-217109. Biblioteca responsável: BR68.1.

6. Heeb, L. E. M., Egholm, C., & Boyman, O. (2020). Evolution and function of interleukin-4 receptor signaling in adaptive immunity and neutrophils. Genes & Immunity, 21(3), 143-149.

7. Kubaski, F., et al. (2013). Análises in silico para predição do fenótipo na substituição de aminoácidos no gene da GALNS. Revista HCPA. Porto Alegre.

8. Kulshreshtha, S., et al. (2016). Computational approaches for predicting mutant protein stability. Journal of Computer-Aided Molecular Design, 30, 401-412.

9. Moreira, P. R., et al. (2006). The IL1A (−889) gene polymorphism is associated with chronic periodontal disease in a sample of Brazilian individuals. Journal of Periodontal Research, 42(1), 23-30.

10. Redler, R. L., et al. (2016). Protein destabilization as a common factor in diverse inherited disorders. Journal of Molecular Evolution, 82, 11-16.

11. Varella, P. P. V., & Forte, W. C. N. (2001). Citocinas: revisão. Rev. Bras. Alergia Imunopatol, 146-154. Disponível em: https://www.studocu.com/pt-br/document/instituto-federal-de-educacao-ciencia-e-tecnologia-do-rio-grande-do-norte/biologia/tipos-de-mutacao/17397090. Acesso em: 20 out. 2024.