Identification and analysis of pathogenic nsSNPs in human LSP1 gene

Hani Mohammed Ali

1Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia; Email: hmohammedali@kau.edu.sa; haniolfat@hotmail.com; *Corresponding author

Received September 11, 2019; Revised September 20, 2019; Accepted September 29, 2019; Published September 30, 2019

Abstract:
LSP1 (Lymphocyte-specific protein 1) protein plays an important role in neutrophil motility, fibrinogen matrix proteins adhesion, and trans-endothelial migration. Variation in the LSP1 gene is associated with leukemia and lymphomas in tumor cells of Hodgkin’s disease and breast cancer. Despite extensive study on the human LSP1, a comprehensive analysis on the Single Nucleotide Polymorphism (SNPs) of the gene is not available. Therefore, it is of interest to identify, collect, store and analyze the SNPs of the LSP1 gene in relation to several known diseases. Hence, the SNP data (398 rsids) from dbSNP database was downloaded and mapped to the genomic coordinate of “NM_002339.2” transcript expressed by LSP1 (P33241). There were 300 nsSNPs with missense mutation in the dataset. Tools such as SIFT, PROVEAN, Condel, and PolyPhen-2 were further used to identify 29 highly deleterious or damaging on synonymous SNP (nsSNPs) for LSP1. These high confident damaging nsSNPs were further analyzed for disease association using SNPs & GO tool. SNPs of the gene such as nsSNPs C283R, G234R, Y328D and H325P showed disease association with high prevalence.

Keywords: SNP; Lymphocyte-specific protein; computational analysis; F-actin binding protein; neutrophil actin dysfunction

Background:
Human LSP1 (lymphocyte specific protein 1) gene encodes an intracellular F-actin binding protein, recently renamed as leukocyte specific protein. The protein is expressed in lymphocytes, macrophages, neutrophils, and endothelium and regulates adhesion to fibrinogen matrix proteins, neutrophil motility, and transendothelial migration. Due to alternative splicing there are multiple transcript variants which encodes different isoforms. Highest expression of this gene in spleen (RPKM 60.6), appendix (RPKM 43.3) and other tissues [1, 2] is known. LSP1 is found in plasma membrane internal surface of the, the cytoplasm, and is thought to mediate cytoskeleton-driven responses in activated leukocytes that involve receptor capping, cell-cell interactions and cell motility [3]. Lymphocyte specific protein 1 modulates leukocyte populations in resting and inflamed peritoneum [2]. The LSP1 protein is detected in leukemia and lymphomas in tumor cells of Hodgkin’s disease and breast cancer [4]. The motility of melanoma cell is inhibited even at low level of LSP1 expression [5]. Many research showed identifying the deleterious effectiveness and disease associated mutations, thus predicting the pathogenic nsSNPs in correlation to their functional and structural damaging properties [6-9]. Computational studies provide an efficient platform for analysis of genetic mutations for their pathological consequences and in determining their underlying molecular mechanism [10-11]. Single nucleotide polymorphism (SNPs) is a common genetic variations contributing greatly towards the phenotypic variations in the populations. SNPs can alter the functional consequences of proteins. In the coding region of gene, SNPs may be synonymous, non-synonymous (nsSNPs) or nonsense. Synonymous SNPs changes the nucleotide base residue but does not change the amino acid residue in protein sequence due to degeneracy of genetic code. The nsSNPs also called missense
variants, alter amino acid residue in protein sequence and thus change the function of protein through altering protein activity, solubility and protein structure. Nonsense SNPs introduce premature termination in the protein sequence. SNPs have been emerged as the genetic markers for diseases and there are many SNP’s markers available in the public databases. With recent advances in high-throughput sequencing technology, many new SNPs have been mapped to human LSP1 genes. However, not all SNPs are functionally important. Despite extensive studies of LSP1 proteins in human and effect of their polymorphism in diseases, no attempts was made to comprehensively and systematically analyze to establish the functional consequences of SNPs of LSP1 gene. The aim of this study is to identify the high confident pathogenic SNPs of LSP1 gene and determine their functional consequences using computational methods.

Figure 1: Flow chart depicting overall work methodology adopted in this study.

Materials and Methods

SNPs dataset
The SNPs of the LSP1 (Lymphocyte-specific protein 1) protein were retrieved from the dbSNP database [12]. I used “LSP1” as our search term and filter SNPs. Furthermore, I mapped these SNPs on the genomic coordinate of “NM_002339.2” transcript expresses LSP1 protein (P33241) for computation analysis of the effect of missense variant. The protein sequences of genes, LSP1 (P33241) was retrieved from the UniProt database [18]. I employed various computational approaches to identify the pathogenic SNPs and their effect on structural and functional consequences of LSP1 (Figure 1)

Tools used for the prediction of SNPs effects

Predicting deleterious and damaging nsSNPs

SIFT: The algorithm predicted that the tolerant and intolerant coding base substitution based upon properties of amino acids and homology of sequence [13]. The tool considered that vital positions in the protein sequence have been conserved throughout evolution and therefore substitutions at conserved alignment position is expected to be less tolerated and affect protein function than those at diverse positions. I used SIFT version 2.0 [19], which predicted the amino acid substitution score from zero to one. SIFT predicted substituted amino acid as damaging at default threshold score <0.05, while score ≥ 0.05 is predicted as tolerated.

PROVEAN: The online tool uses an alignment-based scoring method for predicting the functional consequences of single and multiple amino acid substitutions, and in-frame deletions and insertions [14]. The tool has a default threshold score, i.e. -2.5, below which a protein variant is predicted as deleterious, and above that threshold, a protein variant is neutral.

Condel (CONsensus DEleteriousness): This tool evaluates the probability of missense single nucleotide variants (SNVs) deleterious. it computes a weighted average of the scores of SIFT, PolyPhen2, Mutation Assessor and FatHMM [15].

PolyPhen-2: This tool is predicting the structural and functional consequences of a particular amino acid substitution in human protein [16]. Prediction of PolyPhen-2 server [20] is based on a number of features including information of structural and sequence comparison. The PolyPhen-2 score varies between 0.0 (benign) to 1.0 (damaging). The PolyPhen-2 prediction output categorizes the SNPs into three basic categories, benign (score < 0.2), possibly damaging, (score between 0.2 and 0.96), or probably damaging (score >0.96).

Predicting disease associated nsSNPs

SNPs & GO: A web server predicting whether an amino acid substitution is associated to a disease or not [17]. It is a SVM (Support Vector
Preprocessing and analyzing SNPs

In order to predict the damaging or deleterious nsSNPs multiple consensus tools were employed. Initially, online tool VEP was used [21]. VEP advantages include: it uses latest human genome assembly GRCh38.p10, and can predict thousands of SNPs from multiple tools including SIFT, Condel, and PolyPhen-2, at a time. 398 nsSNP accession numbers were uploaded to VEP tool and the prediction results were taken for further analysis.

Table 1: List of 40 deleterious missense SNPs on the LSP1 gene identified using prediction tools such as SIFT (score > 0.9), Condel (score > 0.9), Polyphen (score > 0.96) and PROVEAN (score > 2.5).

| SNP id  | AA Change | SIFT (score) | Polyphen (score) | Condel (score) | PROVEAN (score) |
|---------|-----------|--------------|------------------|----------------|-----------------|
| n507274538 | P146Q | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n517796603 | P146Q | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n513181405 | P146Q | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n513181405 | P146Q | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n529973498 | P146Q | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n542628204 | P136V | deleterious (0.98) | probably_damaging (0.92) | deleterious (0.80) | Deleterious |
| n542628204 | P136V | deleterious (0.98) | probably_damaging (0.92) | deleterious (0.80) | Deleterious |
| n543765306 | S135R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n543765306 | S135R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Neutral |
| n5172211080 | S214R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.82) | Deleterious |
| n5172211080 | S214R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.82) | Deleterious |
| n535260247 | S224R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.82) | Deleterious |
| n535260247 | S224R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.82) | Deleterious |
| n542233844 | P255R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Deleterious |
| n542233844 | P255R | deleterious (0.99) | probably_damaging (0.92) | deleterious (0.81) | Deleterious |
| n530673461 | S257F | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n530673461 | S257F | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n568865993 | D277T | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n568865993 | D277T | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n568865993 | D277T | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n568865993 | D277T | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |
| n568865993 | D277T | deleterious (0.99) | probably_damaging (0.91) | deleterious (0.81) | Neutral |

Identifying disease associated nsSNPs

Furthermore, 29 selected amino acid substitutions in LSP1 protein were used to analyze for disease association. LSP1 Protein ID “P33241” isofrom-land its amino acid mutations were submitted to “SNPs & GO” tool [22] and the predicted disease association from three different tools were analyzed. The output of (a) SNPs & GO predicted 4 SNPs C283R, G324R, Y328D and H332P are associated with disease and (b) PhD-SNP predicted 14 SNPs R207P, I2277, Q233R, Q233K, T235I, T235P, E239K, C283R, W297S, Y328D, Y318C, K319T, G324R, H332P are associated with diseases, while (c) PANTHER predicted 4 SNPs C283R, L296H, S276C and G301R as disease associated (Table 2).

Predicting deleterious and damaging nsSNPs

In order to predict the damaging or deleterious nsSNPs multiple consensus tools were employed. Initially, online tool VEP was used [21]. VEP advantages include: it uses latest human genome assembly GRCh38.p10, and can predict thousands of SNPs from multiple tools including SIFT, Condel, and PolyPhen-2, at a time. 398 nsSNP accession numbers were uploaded to VEP tool and the prediction results were taken for further analysis.

300 missense SNPs was mapped to NM_002339.2 on default scores of consensus tools based on sequence and structure homology methods: (a) SIFT (score < 0.5) and (b) PROVEAN (score > 2.5) and Condel (score > 0.522). In order to get a very high confident nsSNPs impacting function and structure of LSP1, I considered high stringent scores across different consensus tools. At parameters of SIFT (score = 0), Polyphen (score >0.96) and Condel (score >0.9), I got 40 nsSNPs (Table 1). These 40nsSNPs were further analyzed by PROVEAN, which gave 29 nsSNP at default cutoff at -2.5 score fall in the predicted category of deleterious and have damaging effect on protein structure and function (Table 1).
Table 2: Prediction of disease associated amino acid substitution using SNPs & GO, PhD-SNP and PANTHER on29 deleterious or damaging missense SNP using tools such as SIFT, Condel, PolyPhen and PROVEAN

| SNP | AA Change | PANTHER | PROVEAN |
|-----|-----------|---------|---------|
| rs1429150883 | H75N | Neutral | Neutral |
| rs1429150928 | H75N | Neutral | Neutral |
| rs1429150882 | H75N | Neutral | Neutral |
| rs1429150929 | H75N | Neutral | Neutral |
| rs1429150883 | H75N | Neutral | Neutral |
| rs1429150928 | H75N | Neutral | Neutral |
| rs1429150882 | H75N | Neutral | Neutral |
| rs1429150929 | H75N | Neutral | Neutral |

Conclusion
A comprehensive analysis of SNPs of the human LSP1 protein with known disease-associated mutations is reported for the first time. The study identified 29 nsSNPs as highly damaging nsSNPs of the human LSP1 protein. These high-confident damaging nsSNPs were further analyzed for disease association by manual data mapping. Prediction analysis shows that SNPs C283R, G324Rand H325P and Y328D have high prevalence for disease association. Data implies that the reported nsSNPs could potentially alter structure and hence the function of LSP1 protein resulting in pathogenicity with abnormal symptoms describing the disease states. These nsSNPs were associated with significant pathogenicity pending experiment verification to link disease prevalence.

Acknowledgements:
This project was not supported by any grants agency. I acknowledge with thanks Deanship of Scientific Research (DSR), at King Abdulaziz University, Jeddah for providing their support. I am also thankful to Dr. Ahmad Firoz, Assistant Professor, Department of Biological Science, King Abdulaziz University, Jeddah, KSA, for his critical reviewing of the manuscript.
References:
[1] https://www.ncbi.nlm.nih.gov/gene/4046
[2] Jongstra-Bilen J et al. Blood 2000 96:1827 [PMID 10961883].
[3] Palker TI et al. Hybridoma 1998 17:497 [PMID 9890705].
[4] Pullford K et al. Immunology 1999 96:262 [PMID 10233704].
[5] Howard TH et al. Blood 1998 91:4786 [PMID 9616178].
[6] Tang Z et al. Nucleic Acids Res. 2017 45:W98 [28407145].
[7] Bhattacharya A et al. Nucleic Acids Res. 2014 42:D86 [PMID 24163105].
[8] Wang Z & Moult J. Hum Mutat. 2001 17:263 [PMID11295823].
[9] Bromberg Y & Rost B. BMC Bioinformatics 2009 10:58 [PMID 19758472].
[10] Calabrese R et al. Hum Mutat. 2009 30:1237 [PMID 19514061].
[11] Bao L et al. Nucleic Acids Res 2005 33:W480 [PMID 15980516].
[12] Sherry ST et al. Nucleic Acids Res 2001 29:308 [PMID 11125122].
[13] Sim NL et al. Nucleic Acids Res 2012 40 W452 [PMID 22689647].
[14] Choi Y & Chan AP. Bioinformatics 2015 31(16) 2745 [PMID 25851949].
[15] Adzhubei IA et al. Nat Methods 2010 7:248 [PMID 20354512].
[16] Hecht M et al. BMC Genomics 2015 16: S1 [PMID 26110438 ].
[17] Calabrese R et al. 2009 Human Mutation 30:1237 [PMID 19514061].
[18] https://www.uniprot.org
[19] http://sift.jcvi.org/
[20] http://genetics.bwh.harvard.edu/pph2/
[21] http://www.ensembl.org/Tools/VEP
[22] http://snps.biocold.org/snps-and-go/snps-and-go.html

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Edited by P Kanguene

Citation: Ali, Bioinformation 15(9): 621-626 (2019)
