INVESTIGATION OF PUTATIVE FUNCTIONAL SNPS OF HUMAN HAT1 PROTEIN: A COMPREHENSIVE *IN SILICO* STUDY

O. AVSAR

Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Hitit University, Corum, 19030, Turkey
E-mail: orcunavsar@hitit.edu.tr

Histone modifications such as acetylation play a fundamental role in DNA packaging and genome regulation and HAT1 protein is involved in gene transcription, DNA repair, and chromatin assembly. Single nucleotide polymorphisms (SNPs) in the human HAT1 gene may be correlated with human diseases such as cancers, inflammatory, and neuropsychiatric diseases. Hence, identification of putative functional SNPs which affect structure and/or function of protein is important for understanding the molecular mechanisms of pathogenesis of diseases and discovery of potential therapeutic agents. In this study, numerous bioinformatics tools were used to determine the most damaging nsSNPs for the function and/or structure of HAT1 protein. In silico analysis was carried out by different algorithmic programs including SIFT, PolyPhen-2, PROVEAN, SNPs&GO, and PhD-SNP. Our study concludes that mutation of Leucine → Arginine at position 416 (rs199575205) is major deleterious mutation which may lead to damage of HAT1 protein. Analyis of HAT1 gene variants by computational tools is a first and comprehensive in silico study. Future in vitro and in vivo studies should include this nsSNP as main target for the development of therapeutics for diseases that are associated with this missense polymorphism.

***Key words**: HAT1, nsSNP, mutation, in silico, histone.*

ВИВЧЕННЯ ІМОВІРНО ФУНКЦІОНАЛЬНИХ ОД НОНУКЛЕОТИДНИХ ПОЛІМОРФІЗМІВ БІЛКА HAT1 ЛЮДИНИ: КОМПЛЕКСНЕ ДОСЛІДЖЕННЯ *IN SILICO*

Модифікації гістонів, зокрема, ацетилювання, відіграють надзвичайно важливу роль у пакуванні ДНК і регуляції геному. Білок HAT1 залучений до транскрипції генів, репарації ДНК і складання хроматину. Однонуклеотидні поліморфізми (SNP) гену *HAT1* людини можуть бути пов'язані з такими хворобами людини, як рак, запалення, неврологічні і психіатричні захворювання. Таким чином, ідентифікація імовірних функціональних SNP, що впливають на структуру і/або функцію білку, є важливою для розуміння молекулярних механізмів патогенезу хвороб і виявлення потенційних терапевтичних засобів. Багато інструментів біоінформатик и було використано у цьому дослідженні з метою визначення nsSNP, що чинять найбільш шкідливий вплив на функцію та/або структуру білку HAT1. *In silico* аналіз проводили за допомогою різних алгоритмічних програм, зокрема, SIFT, PolyPhen-2, PROVEAN, SNPs&GO і PhD-SNP. Ми прийшли до висновку, що мутація, за якої відбувається заміна лейцин-аргінін в положенні 416 (rs199575205), – це основна шкідлива мутація, яка може призвести до пошкодження білку HAT1. Аналіз варіантів гену *HAT1* за допомогою обчислювальної техніки – це перше комплексне дослідження *in silico*. Наступні дослідження *in vitro* та *in vivo* повинні включати цей nsSNP як основний цільовий об'єкт розробки терапевтичних засобів для лікування хвороб, пов'язаних з цим міссенсним поліформізмом.

***Ключові слова:** HAT1, nsSNP, мутація, *in silico*, гістон.*

REFERENCES

Adzhubei I, Jordan DM, Sunyaev SR. (2013) Predicting functional effect of human UNIT 7.20 missense mutations using PolyPhen-2. Curr Protoc Hum Genet 76: 7.20.1-7.20.41. doi: 10.1002/0471142905.hg0720s76

Annan RS, Zappacosta Fю (2005) Protein posttranslational modifications: phosphorylation site analysis using mass spectrometry Methods Biochem Anal 45:85–106.

Arshad M, Bhatti A, John P. (2018) Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study PLoS ONE 13:e0188143. doi: 10.1371/journal.pone.0188143

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules Nucleic Acids Res 8: W344–350. doi: 10.1093/nar/gkw408

Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. (2004) ConSeq: The Identification of Functionally and Structurally Important Residues in Protein Sequences Bioinformatics 20:1322–1324. doi: 10.1093/bioinformatics/bth070

Bhagwat M (2010) Searching NCBI's dbSNP database. Curr Protoc Bioinformatics Chapter: Unit-1.19. doi: 10.1002/0471250953.bi0119s32

Capriotti E, Fariselli P, Casadio R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure Nucleic Acids Res 33: W306-310. doi: 10.1093/nar/gki375

Capriotti E, Calabrase R, Fariselli P, Martelli PL, Altman RB, Casadio R. (2013) WS-SNPs&GO: a

web server for predicting the deleterious effect of human protein variants using functional annotation BMC Genomics 14 (Suppl 3): S6.

Capriotti E, Fariselli P. (2017) PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants Nucleic Acids Res 45: W247–W252. doi: 10.1093/nar/gkx369

Carugo O, Pongor S. (2001) A normalized root-mean-square distance for comparing protein three-dimensional structures Protein Sci 10:1470–1473. doi: 10.1110/ps.690101

Chen R, Davydov EV, Sirota M, Butte AJ. (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association PLoS ONE 5: e13574. doi: 10.1371/journal.pone.0013574

Choi Y, Chan AP. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels Bioinformatics 31:2745–2747. doi: 10.1093/bioinformatics/btv195

Colovos C, Yeates T. (1993) ERRAT: an empirical atom-based method for validating protein structures Protein Sci 2:1511–1519.

Deng W, Wang C, Zhang Y, Xu Y, Zhang S, Liu X et al. (2016) GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences Scientific Reports 6:39787.

Du K, Sharma M, Lukacs GL. (2005) The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. Nat Struct Mol Biol 12:17–25. doi: 10.1038/nsmb882

Gruber JJ, Geller B, Lipchik AM, Chen J, Salahudeen AA, Ram AN. (2019) HAT1 drives a gene-metabolite circuit that links nutrient metabolism to histone production. bioRxiv. doi: https://doi.org/10.1101/664615

Hu Z, Zhou J, Jiang J, Yuan J, Zhang Y, Wei X. (2019) Genomic characterization of genes encoding histone acetylation modulator proteins identifies therapeutic targets for cancer treatment Nature Communications 10:733. doi: https://doi.org/10.1038/s41467-019-08554-x

Jin X, Tian S, Li P. (2017) Histone acetyltransferase 1 promotes cell proliferation and induces cisplatin resistance in hepatocellular carcinoma Oncol Res 25:939–946. doi: 10.3727/096504016X14809827856524

Kaur T, Khakur T, Singh J, Kamboj SS, Kaur M. (2017) Identification of functional SNPs in human LGALS3 gene by in silico analyses Egypt J Med Hum18:321–328. doi: http://dx.doi.org/10.1016/j.ejmhg.2017.02.001

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. (2015) The Phyre2 web portal for protein modeling, prediction and analysis Nat Protoc 10:845–858. doi: https://doi.org/10.1038/nprot.2015.053

Kucukkal TG, Petukh M, Li L, Alexov E. (2015) Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins Curr Opin Struct Biol 32:18–24. doi: 10.1016/j.sbi.2015.01.003

Kumar P, Henikoff S, Ng PC. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm Nature Protocols 4:1073–1081. doi: https://doi.org/10.1038/nprot.2009.86

Lee J, Son MJ, Son CY, Jeong GH, Lee KH, Lee KS. (2020) Genetic variation and autism: A field synopsis and systematic meta-analysis Brain Sci 10:692. doi: 10.3390/brainsci10100692

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions Bioinformatics 25:2744–2750. doi: 10.1093/bioinformatics/btp528

Mayer S, Rüdiger S, Ang HC, Joerger AC, Fersht AR. (2007) Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro J Mol Biol 372: 268–276. doi: 10.1016/j.jmb.2007.06.044

Miller MP, Kumar S. (2001) Understanding human disease mutations through the use of interspecific genetic variation Hum Mol Genet 10:2319–2328. doi: 10.1093/hmg/10.21.2319

Nakamura N. (2018) Ubiquitin system Int J Mol Sci 19:1080. doi: 10.3390/ijms19041080

Osman MM, Khalifa AS, Mutasim AEY, Massaad SO, Gasemelseed MM, Abdagader MA et al. (2016) In silico Analysis of Single Nucleotide Polymorphisms (Snps) in Human FTO Gene JSM Bioinformatics, Genomics and Proteomics 1:1003.

Pascovici D, Wu JX, McKay MJ, Joseph C, Noor Z, Kamath K, et al. (2019) Clinically relevant post-translational modification analyses-maturing workflows and bioinformatics tools Int J Mol Sci 20:16. doi: 10.3390/ijms20010016

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004) UCSF Chimera- a visualization system for exploratory research and analysis J Comput Chem 13:1605–1612. doi: 10.1002/jcc.20084

Prabakaran S, Lippens G, Steen H, Gunawardena J. (2012) Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding WIREs Syst Biol Med 4:565–583. doi: 10.1002/wsbm.1185

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW et al. (2010) Identification, analysis

and prediction of protein ubiquitination sites. Proteins 78:365–380. doi: 10.1002/prot.22555

Ramensky V, Bork P, Sunyaev S. (2002) Human non-synonymous SNPs: server and survey Nucleic Acids Res 30:3894–900. doi: 10.1093/nar/gkf493

Rattan SI, Derventzi A, Clark BF. (1992) Protein synthesis, posttranslational modifications, and aging Ann N Y Acad Sci 663:48–62. doi: 10.1111/j.1749-6632.1992.tb38648.x

Shaw G. (2013) Polymorphism and single nucleotide polymorphisms (SNPs) BJU Inter 112:664–665. doi: 10.1111/bju.12298

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40: W452-W457. doi: 10.1093/nar/gks539

Singh SM, Kongari N, Cabello-Villegas J, Mallela K. (2010) Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-β aggregates PNAS 107:15069–15074. doi: https://doi.org/10.1073/pnas.1008818107

Varga J, Korbai S, Neller A, Zsindely N, Bodai L. (2019) Hat1 acetylates histone H4 and modulates the transcriptional program in Drosophila embryogenesis Scientific Reports 9:17973. doi: https://doi.org/10.1038/s41598-019-54497-0

Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S et al. (2020) MusiteDeep: a deep learning based web-server for protein post-translational modification site prediction and visualization. Nucl Acids Res 48: W140–W146. doi: https://doi.org/10.1093/nar/gkaa275

Wen PP, Shi SP, Xu HD, Wang LN, Qiu JD. (2016) Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization Bioinformatics 32:3107–3115. doi: 10.1093/bioinformatics/btw377

Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy Mol Cell Proteomics 7:1598–1608. doi: 10.1074/mcp.M700574-MCP200

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. (2015) The I-TASSER Suite: Protein structure and function prediction Nat Methods 12:7–8. doi: 10.1038/nmeth.3213

Yang XJ, Seto E. (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications Mol Cell 31:449–461. doi: 10.1016/j.molcel.2008.07.002

Zhang Y, Skolnick J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score Nucl Acids Res 33:2302–2309. doi: 10.1093/nar/gki524

Received December 17, 2019
Received March 31, 2021
Accepted January 18, 2022