Peer reviewed / ISSN: 2663-1040 *AJ Life Sci. 2020, 3 (2): 10-24* **Research Article**

Abasyn Journal of Life Sciences

Open Access

DOI: 10.34091/AJLS.3.2.2

In Silico Identification of Novel Acute Myeloid Leukemia Associated Missense SNPs in Human *CEBPA* Gene

Mudassir Khan[1], Mehran Akhtar[2], Maharij Haroon Jadoon[3], Dilawar Khan[1*]

[1]*Atta-Ur-Rahman School of applied Biosciences (ASAB), National University of Science and Technology (NUST), Islamabad, Pakistan*
[2]*Department of Biotechnology, COMSATS University Islamabad, Abbottabad Campus, Pakistan*
[3]*Research Centre FOR Modelling and Simulation (RCMS), National University of Science and Technology (NUST), Islamabad, Pakistan*

Abstract

Single nucleotide polymorphisms (SNPs) in *CEBPA* gene have been found to be associated with cancer especially Acute Myeloid Leukemia (AML). Therefore, the identification of functional and structural polymorphisms in *CEBPA* is important to study and discover therapeutics targets and potential malfunctioning. For this purpose, several bioinformatics tools were used for the identification of disease-associated nsSNPs, which might be vital for the structure and function of *CEBPA*, making them extremely important. *In silico* tools used in this study included SIFT, PROVEAN, PolyPhen2, SNP&GO and PhD-SNP, followed by ConSurf and I-Mutant. Protein 3D modelling was carried out using I-TASSER and MODELLER v9.22, while GeneMANIA and string were used for the prediction of gene-gene interaction in this regard. From our study, we found that the L345P, R333C, R339Q, V328G, R327W, L317Q, N292S, E284A, R156W, Y108N and F82L mutations were the most crucial SNPs. Additionally, the gene-gene interaction showed the genes having correlation with *CEBPA's* co-expressions and importance in several pathways. In future, these 11 mutations should be investigated while studying diseases related to *CEBPA*, especially for AML. Being the first of its kind, future perspectives are proposed in this study, which will help in precision medicine. Animal models are of great significance in finding out *CEBPA* effects in disease.

Keywords: *CEBPA*; *in silico*; AML; nsSNPs; mutations

Article Info:
Received:
September 3, 2020
Received Revised:
October 17, 2020
Accepted:
October 18, 2020
Available online:
December 31, 2020

**Corresponding Author:*
dilawar_qau@yahoo.com

How to cite:
Khan M, Akhtar M, Jadoon MH, Khan D. In silico identification of novel Acute Myeloid Leukemia associated missense SNPs in human CEBPA gene. *Abasyn Journal of Life Sciences 2020; 3(2): 10-24.*

© 2020 Abasyn Journal of Life Sciences.

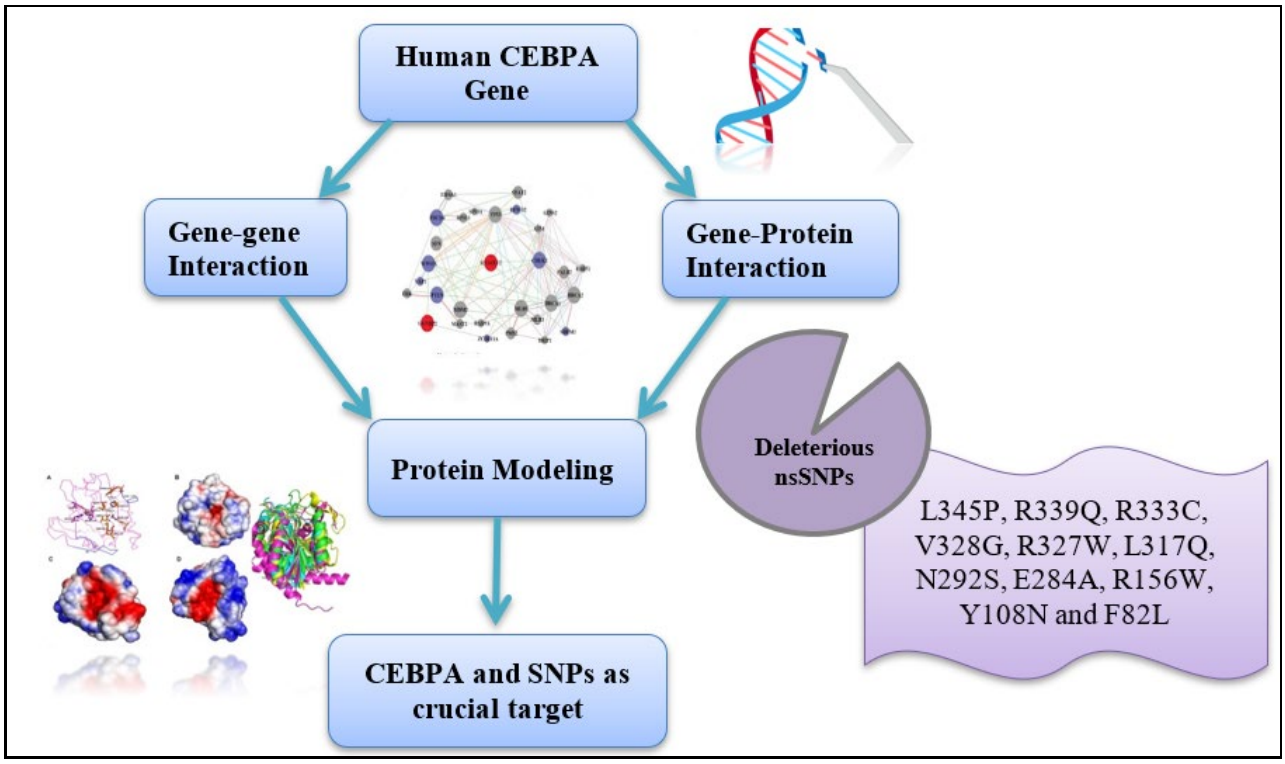

Fig. 1. Graphical Abstract

1. INTRODUCTION

In human's genome, the common genetic variations are SNPs, which are widely used in association-studies with quantitative traits and complex diseases[1-5]. Along human genome, SNPs occur in every 100-300 bases, which represents 90% of genetic variations. They are found in human genome in different densities in both coding and non-coding regions[6]. However, SNPs are abundant in the non-coding regions of human genome, including untranslated and regulatory regions as well as introns. Additionally, the phenotypic functions can also be affected by single nucleotide variation, contributing towards the development of disease. SNPs can influence transcription factor binding or gene expression, while transcriptional activity may be modified by SNPs of UTR regions[7], ribosomal translation of mRNA and RNA stability[8].

Humans *CEBPA* gene encodes CCAAT/enhancer-binding protein alpha[9]. It is involved in blood cells differentiation as a transcription factor[10]. Alteration of specific genes including CBP complex can lead to cell differentiation arrest in AML[11], bZIP transcription factor protein is encoded by intron less gene, which can bind to certain gene enhancers and promoters as a homodimer. It can form heterodimer with CEBP-gamma and CEBP-beta, also c-Jun as distinct transcription factors. However, *CEBPA* is required for development of abnormal AML and for normal mature granulocyte formation because it is essential for myeloid lineage commitment[12].

Various studies have reported that 50% of genetic disorders are due to mutated nsSNPs[13]. Genetic variations in respect of deleterious effect has been reported in a study in *ABCA1* gene, which could possibly lead towards the development of hypoalphalipoproteinemia disease. Additionally, in *STEAP2* gene, nsSNPs can cause prostate cancer by upregulating the mentioned gene, identified in similar studies[14].

We have analyzed *CEBPA* gene nsSNPs to find out the most deleterious ones, in order to highlight its potential role in AML. *CEBPA* can interact with Cyclin-dependent kinase 4 and Cyclin-dependent kinase 2 [15]. However, pediatric and adult patient of AML are linked to good outcome, shown mutation of *CEBPA* gene[16]. Genetic abnormalities, hematopoietic progenitors, characterize AML including blocking of hematopoiesis granulocytes and blasts excessive proliferation. It has been shown that differentiation of myeloid progenitors stops by blocking of CCAAT/enhancer-binding protein alpha and suppression of *CEBPA* expression, during granulocyte differentiation role of CCAAT/enhancer-binding protein alpha and as tumor

© 2020 Abasyn Journal of Life Sciences.

suppressor gene role of *CEBPA* is important in AML prognosis[17]. Therefore, structural and functional variants of *CEBPA* needed to be sort out and studied. Hence, in this study several *in silico* tools were used for finding nsSNPs, which could possibly cause damage to *CEBPA* protein. We proposed 3D model of possible deleterious nsSNPs protein of *CEBPA* and its wild type. This study covers its protein analysis using *in silico* which can be very helpful in future studies of disease treatment associated with *CEBPA*, caused by nsSNPs.

2. METHODOLOGY

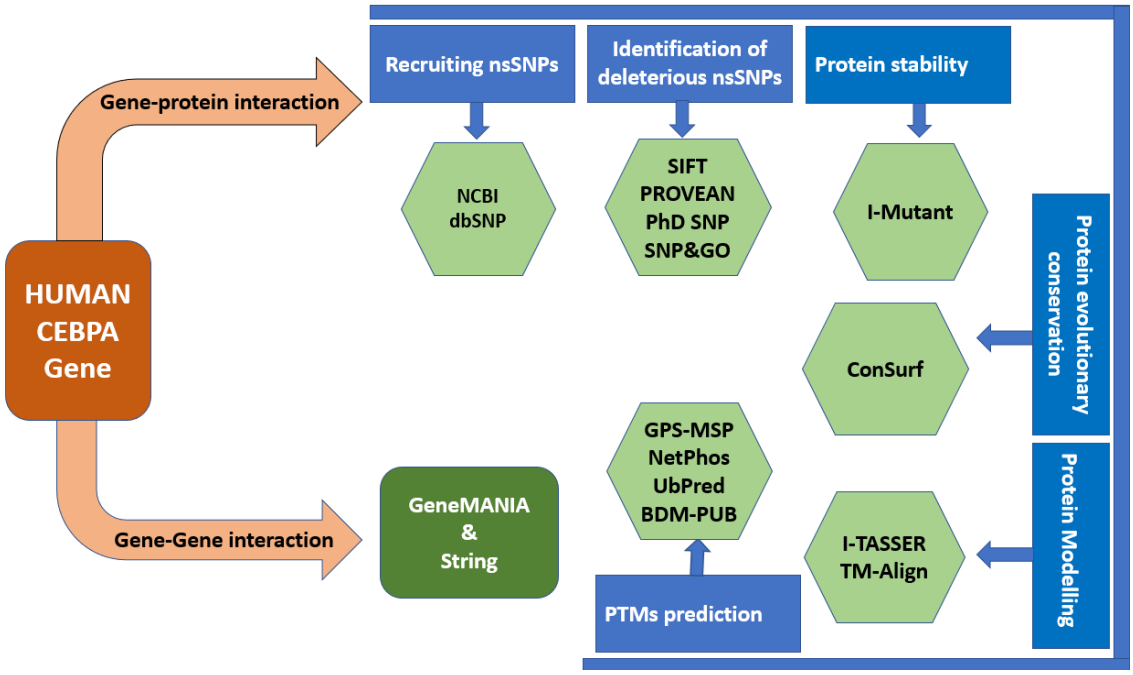

Fig. 2. The graphical illustration of methodology

This research work (Fig. 2) has been completed in several steps including web servers, tools and databases. However, GRCh38 has been used as reference human genome. The detailed description of methodology is as follow.

2.1. Recruiting nsSNPs

National Centre for Biotechnology Information (NCBI) dbSNP (Accessed: 17 July 2020) was used in recruitment of all the SNPs of *CEBPA*. Missense SNPs were selected from *CEBPA* gene window in NCBI where gene view was selected.

2.2. Identification of deleterious nsSNPs

The effect of nsSNPs on protein was identified using four bioinformatical tools. These tools were SIFT (Sorting Intolerant From Tolerant) (https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html), PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org/seq_submit.php), SNP&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html) and PhD-SNP (Predictor of human Deleterious SNP) (https://snps.biofold.org/phd-snp/phd-snp.html). All those SNPs were selected which predicted to be deleterious or intolerant. Further screening of selected SNPs were carried out using PolyPhen2 (Polymorphism Phenotyping 2) (http://genetics.bwh.harvard.edu/pph2/).

© 2020 Abasyn Journal of Life Sciences.

2.3. Stability analysis of protein

I-Mutant 2.0 was used for the scrutiny of target protein stability upon substitution due to nsSNP (https://folding.biofold.org/i-mutant/). Change in mutated protein stability is predicted using this web server, based support vector machine. It also provides predictions with R1 (Reliability index) which ranges from 0 to 10, 0 and 10 shows lowest and highest reliability respectively. *CEBPA* protein fasta sequence was submitted to find the possible effect of deleterious nsSNPs on protein of *CEBPA*, the condition was set to 7.0 pH and 25°C temperature (default parameters).

2.4. Prediction of protein evolutionary conservation

Evolutionary conservation in protein sequence was predicted using ConSurf (https://consurf.tau.ac.il/), which analysed the bases of phylogenetic relations between homologous sequences[18]. For this purpose, to predict conservation degree of amino acid residue, total 50 homologous sequences were used. Additionally, the residues which were highly conserved and aligned with deleterious nsSNPs were analyzed further. After the completion of SNPs relevant information, the next step is to perform structural modelling of relevant protein.

2.5. Prediction of 3D protein structure

I-TASSER (Iterative Threading ASSEmbly Refinement) is a stratified approach used to predict function and structure of protein. I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used to predict 3D protein model, which is 3D homology modelling tool. *CEBPA* wild type 3D proteins models were generated. Furthermore, all the mutant structures were generated by MODELLER v9.22 by using wild type *CEBPA* protein generated by I-TASSER as template. TM-align (https://zhanglab.ccmb.med.umich.edu/TM-align/) was used for the comparison of selected mutants and wild type *CEBPA*, which also predicted the RMSD (Root Mean Square Deviation). Those which had greater RMSD values as compare to *CEBPA* wild type were selected for further study[19]. However, Interactive visualization and molecular features of the resultant protein structure was studied by Chimera v1.11.

2.6. Prediction of PTM sites

Protein function can be predicted by studying post transcriptional modification in protein. GPS-MSP v3.0 (http://msp.biocuckoo.org/online.php) was used to predict methylation sites in *CEBPA*. At tyrosine, threonine and serine positions of *CEBPA* protein sequence, phosphorylation sites predictions were done NetPhos 3.1 (http://www.cbs.dtu.dk/services/NetPhos/). Neural network ensembles were used by NetPhos 3.1, and a threshold of 0.5 was set. Those residues were predicted as phosphorylated which had high score than our selected threshold. In addition to that, **UbPred** (http://www.ubpred.org/) and **BDM-PUB** (http://bdmpub.biocuckoo.org/prediction.php) were used for prediction of ubiquitylation sites. Balanced cut-off was selected for UbPred[20]. Lysine residues were predicted by UbPred which showed equal or higher score then threshold.

2.7. Gene-Gene interaction of *CEBPA*

GeneMANIA (http://genemania.org/) can be used hypothesis generation about function of gene, analyzation of gene lists and genes prioritization for functional assays. GeneMANIA was used to study the interaction of *CEBPA* gene. However, STRING (https://string-db.org/cgi) (Accessed: 17 August 2020 using manual search for *CEBPA* in search box) was used for the prediction of effect of nsSNPs of *CEBPA* on rest of related genes. It is also used for the observation of association with other genes[21]. It predicts gene-gene interaction based on pathways, co-expression, protein domain similarity, genetics, co-localization and protein interaction.

© 2020 Abasyn Journal of Life Sciences.

3. RESULTS

Understanding of SNPs functions will help us to understand the phenotypic variation of human genetics, especially of complex human diseases. However, functional SNPs identification from pool containing both neutral and functional SNPs can lead towards the development of disease by truncated protein formation.

3.1. Recruited nsSNPs

World's largest database for variations of nucleotide is dbSNP which house data from Genome wide association studies (GWAS). Therefore, around 1616 SNPs were recruited from it, in which 269 were nonsynonymous SNPs, 369 in 3'UTR while 190 were located in 5'UTR, 152 were coding synonymous and 540 were others, represented in figure 2. However, from further analysis of SNPs, seven resulted in stop codons having direct effect on protein structure. In addition to that, truncated protein is also the result of SNPs that directly lead towards the diseases.

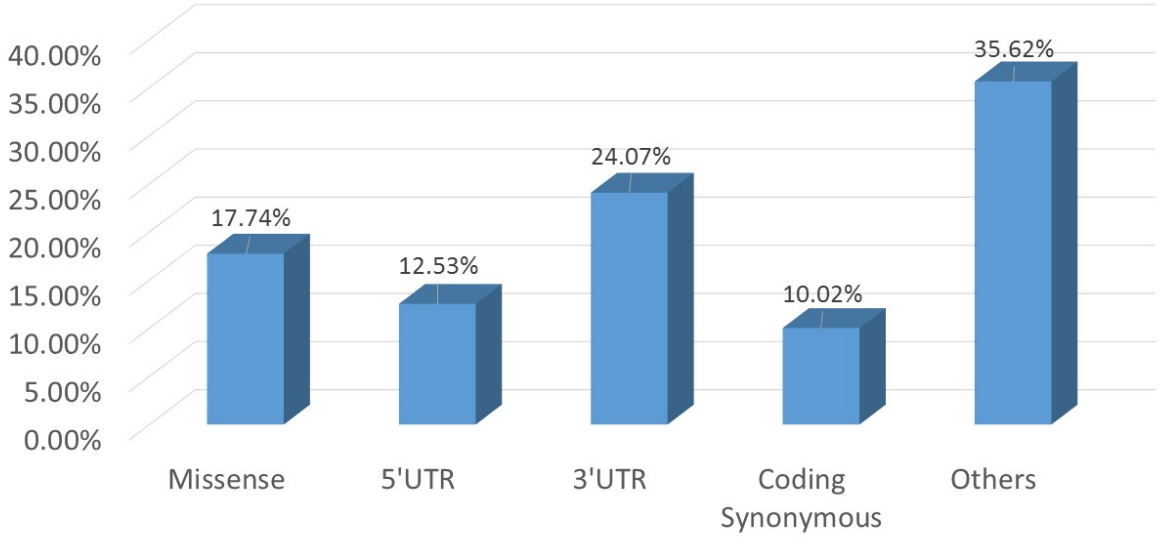

Fig. 3. Percentage of different types of SNPs in Human *CEBPA* gene.

3.2. Deleterious nsSNPs identification

All recruited nsSNPs were subjected to different four bioinformatical tools which includes PhD-SNP, SNPGO, and PROVEAN and SIFT, for further predictions of its effect on structure and function of *CEBPA* protein.

Threshold value of -2.5 was set in **PROVEAN** and all those variants were considered deleterious which were below this threshold. 61 nsSNPs had deleterious effect according to results of PROVEAN.

SIFT shows us predicted values which in turn shows that whether substitution of an amino acid affects function of protein based on sequence homology and also amino acids physical properties. Therefore, SIFT is selected for this purpose to sort out our results further. In **SIFT,** 0.5 TI (Tolerance Index) was considered as threshold value and below this value all the results were considered intolerant or effected.

95 SNPs were predicted by SIFT which were intolerant by SIFT and 41 nsSNPs were found diseased predicted by **PhD-SNP.** ***SNPs***&GO is a server for the prediction of single point protein mutations likely to be involved in the insurgence of diseases in humans. 21 SNPs were labelled as diseased by **SNPs and GO** result.

© 2020 Abasyn Journal of Life Sciences.

Fig. 3 shows results of all tools. We selected 20 nsSNPs, which were predicted as deleterious in all of four tools and provided in table 1. These nsSNPs were then submitted to PolyPhen2, which predicted results as possibly damaging and benign. The most confident predictions were considered as probably damaging. It also gives 0 to 1 count score. 15 nsSNPs (16 amino acid residual changes), which were predicted as probably damaging were considered for further analysis.

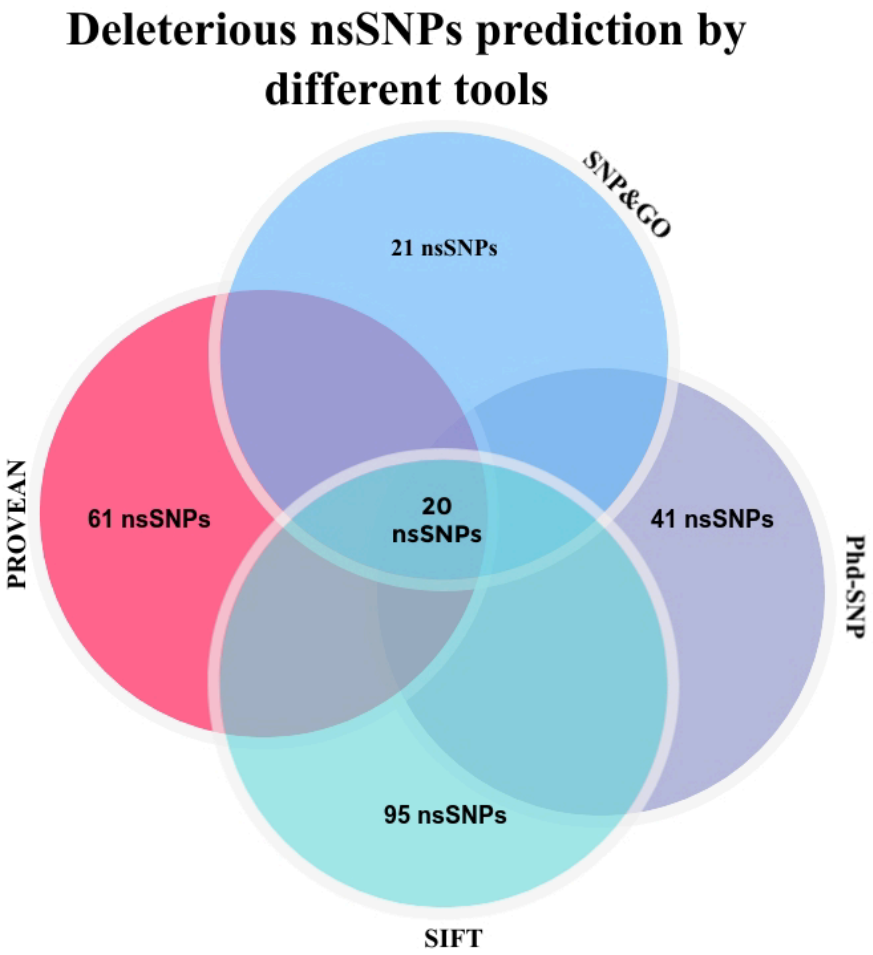

Fig. 4. Representation of predicted deleterious nsSNPs by four *in silico* tools

Table 1. Most damaging 6 nsSNPs predicted by tools

		PROVEAN	Polyphen2 (HumDiv)	PhD-SNP		SNPs & GO		SIFT	
rs ID	Amino acid change	Score	Score	RI	Prob.	RI	Prob.	Prediction	Score
rs1387611667	C357Y	-8.920	0.561	1	0.444	0	0.521	Affected	0.00
rs761752002	L345P	-6.390	0.999	6	0.785	3	0.637	Affected	0.01
rs1304445759	R339Q	-3.899	0.999	0	0.493	1	0.466	Affected	0.00
rs369632687	E334K	-3.899	1.00	5	0.765	6	0.793	Affected	0.00
rs369632687	E334Q	-2.924	1.00	2	0.599	3	0.649	Affected	0.01
rs758726582	R333C	-7.409	1.00	2	0.577	0	0.493	Affected	0.00
rs1422138876	V328G	-6.823	1.00	3	0.627	2	0.392	Affected	0.00
rs1306818311	R327W	-6.468	1.00	1	0.536	2	0.383	Affected	0.00
rs756436149	L317Q	-5.298	1.00	5	0.772	2	0.620	Affected	0.01

© 2020 Abasyn Journal of Life Sciences.

rs1392203731	K298Q	-3.899	1.00	2	0.584	2	0.386	Affected	0.00
rs776590829	N292S	-4.907	1.00	5	0.767	5	0.751	Affected	0.00
rs1196766447	E284A	-5.588	1.00	1	0.574	4	0.311	Affected	0.00
rs1379379731	L220P	-2.663	0.331	2	0.586	1	0.465	Affected	0.00
rs1267025311	R156W	-2.983	1.00	6	0.798	4	0.722	Affected	0.00
rs1257791760	C133Y	-2.887	0.480	6	0.819	4	0.716	Affected	0.01
rs1197023470	C133R	-3.334	0.012	6	0.806	5	0.739	Tolerated	0.10
rs1038352346	G132C	-3.287	0.998	5	0.736	2	0.582	Affected	0.02
rs1245991358	Y108N	-2.900	0.998	1	0.525	1	0.569	Tolerated	0.31
rs917977456	F82L	-3.057	0.990	3	0.649	4	0.280	Affected	0.00
rs1452063514	D63N	-2.992	1.00	4	0.697	4	0.717	Affected	0.00

3.3. Prediction of *CEBPA* stability

Stability of protein is predicted by I-Mutant of *CEBPA* gene for selected nsSNPs and substitution of its amino acid. 15 nsSNPs were submitted to I-Mutant and predicted that 14 of these decrease protein stabilities and one increase stability, All the nsSNPs were individually submitted which were selected and its result of stability was obtained to be increased/decreased with RI ranging from 0 to 10, given in table 2.

Substitution of G132C showed increase in the stability while rest of all showed decrease in stability. This result showed us that all these 14 nsSNPs might cause a greater damage by decreasing stability of *CEBPA* protein. In further analysis, G132C was skipped.

Table 2. I-Mutant prediction for stability of *CEBPA* protein upon selected mutations.

Amino acid change	Stability	RI	Amino acid change	Stability	RI
L345P	Decrease	6	K298Q	Decrease	1
R339Q	Decrease	6	N292S	Decrease	6
E334K	Decrease	8	E284A	Decrease	3
E334Q	Decrease	6	R156W	Decrease	8
R333C	Decrease	6	G132C	Increase	0
V328G	Decrease	7	Y108N	Decrease	0
R327W	Decrease	6	F82L	Decrease	5
L317Q	Decrease	8	D63N	Decrease	4

3.4. *CEBPA* protein evolutionary conservation

It is necessary to know about the evolution to study the mutations which leads to health problems in humans[22]. To know about possible effects of the selected nsSNPs, ConSurf was used for conservation profile study of *CEBPA* amino acid residues. Results obtained provided us structural representation of *CEBPA* protein. Results of all the residue amino acid of *CEBPA* were given but our prime interest was in the location of identified nsSNPs. According to ConSurf L345, L317, L220, C133 and G132 were predicted to be buried, C357, R339, E334, R333, R327, K298, N292, E284, R156 and D63 were Exposed and functionally important while V328 and F82 were Buried and structurally important. Selected nsSNPs conservation score is given in Table 3. Results showed that nsSNP, which were located at highly conserved regions, were most damaging to *CEBPA* protein structure and function.

© 2020 Abasyn Journal of Life Sciences.

Table 3. Conservation profile for the selected residues predicted by ConSurf.

Amino acid change	Conservation score	Prediction
C357	9	Exposed and functionally important
L345	8	Buried
R339	9	Exposed and functionally important
E334	9	Exposed and functionally important
R333	9	Exposed and functionally important
V328	9	Buried and Structurally important
R327	9	Exposed and functionally important
L317	7	Buried
K298	9	Exposed and functionally important
N292	9	Exposed and functionally important
E284	9	Exposed and functionally important
L220	7	Buried
R156	8	Exposed and functionally important
C133	4	Buried
G132	7	Buried
Y108	2	Exposed
F82	9	Buried and Structurally important
D63	9	Exposed and functionally important

F 3.5. 3D modelling of *CEBPA* and its mutants

For generation of 3D structure of wild type, we used I-TASSER. It used 2nbiA and 5jcsS templates for 3D modelling. It generated five 3D structures for the wild type *CEBPA* protein in which structure having lowest C-score (1.43) was selected. The selected mutant structures were generated by MODELLER v9.22 using wild type protein generated by I-TASSER as template. For each mutant structure, RMSD values were calculated. RMSD value shows average distance between α-carbon backbones of mutant and wild type models. Higher RMSD values predict greater deviation between mutant and wild type structure. R339Q had the highest RMSD value of 3.74 Å while L298Q had the lowest RMSD value of 0.43 Å. Mutant structures with RMSD value greater than 2.0Å is considered effected. 11 mutant structures were having greater RMSD values than 2.0Å. Details of all the structures are provided in table 4. Four mutant structures (L298Q 0.43 Å RMSD), E334K (0.49 Å RMSD), D63N (1.41 Å RMSD) and E334Q (1.47 Å RMSD) were skipped for final modelling because their RMSD values were less than 2.0Å. Wild type structure and final selected mutant residues are presented in Figure 5 and 6 respectively.

Fig. 5. 3D structure of wild type *CEBPA* protein modelled by I-TASSER

© 2020 Abasyn Journal of Life Sciences.

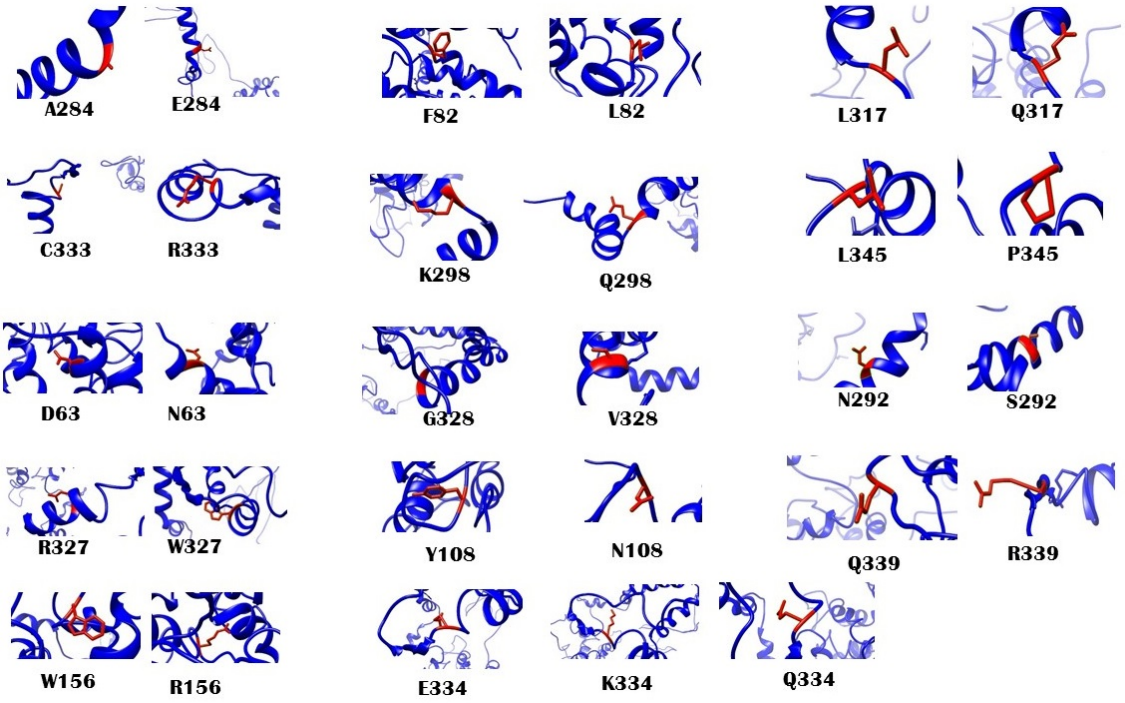

Fig. 6. 3D structures of the modelled mutants of *CEBPA* protein modelled by MODELLER v9.22.

Table 4. RMSD values for selected mutant *CEBPA* proteins compared with *CEBPA* wild protein.

Residual Change	RMSD	Residual Change	RMSD
L345P	3.54 Å	K298Q	0.43 Å
R339Q	3.74 Å	N292S	3.38 Å
E334K	0.49 Å	E284A	3.20 Å
E334Q	1.47 Å	R156W	3.49 Å
R333C	3.55 Å	Y108N	3.53 Å
V328G	3.41 Å	F82L	3.38 Å
R327W	3.60 Å	D63N	1.41 Å
L317Q	3.58 Å		

3.6. Predicted PTMs (Post Translational Modification)

Protein function can be predicted through an extensive study on post-transcriptional modification (PTM) in protein. GPS-MSP 3.0 was used for methylation prediction, which predicted no sites in *CEBPA* to be methylated. It means that nsSNPs might have no role in affecting methylation site. NetPhos3.1 was used for prediction of possible phosphorylation sites. It predicted 28 residues to have potential phosphorylation sites in which 16 were serine specific, 6 were threonine specific and 6 were tyrosine specific. Detailed results are provided in table 5. For prediction of potential ubiquitylation sites, we used BDM-PUB and UbPred. Of the 15 total lysine residues, BDM-PUB predicted 6 sites to be ubiquitinated while UbPred predicted 12 sites to be ubiquitinated. Details of the results of these two tools are given in table 6.

Table 5. Prediction of Phosphorylation Sites by NetPhos 3.1 in *CEBPA* protein

	Position	Score (Threshold 0.5)	Kinase
	3	0.538	CKII
	17	0.563	PKC
	21	0.748	cdk5, p38MAPK, GSK3
	27	0.508	PKG
	61	0.516	CKII, cdc2

© 2020 Abasyn Journal of Life Sciences.

Serine (S)	65	0.534	PKA
	190	0.514	GSK3
	234	0.640	cdk5, PKG, p38MAPK
	266	0.655	PKA
	269	0.505	CKI
	277	0.856	PKC, RSK
	282	0.656	Unsp
	299	0.579	PKC, RSK
	318	0.519	CKII
	332	0.537	CKII
	349	0.501	cdc2
Threonine (T)	216	0.597	PKC
	226	0.521	GSK3, cdk5
	230	0.557	p38MAPK, cdk5
	310	0.567	DNAPK
	318	0.599	unsp
	337	0.842	Unsp
Tyrosine (y)	7	0.919	Unsp
	67	0.875	Unsp
	108	0.520	INSR
	131	0.826	Unsp
	147	0.981	unsp
	147	0.522	SRC

Table 6. *CEBPA* Ubiquitination Prediction Results by UbPred and BDM-PUB

Residue	**UbPred**		**BDM-PUB**	
	Score	Ubiquitinated (Threshold 0.62)	Position	Score (Threshold 0.3)
90	0.66	Yes Low confidence	92	1.84
92	0.70	Yes Medium confidence	161	1.01
161	0.84	Yes High confidence	254	3.11
171	0.92	Yes High confidence	273	4.23
254	0.73	Yes Medium confidence	275	3.23
273	0.27	No	276	5.12
275	0.50	No	280	1.24
276	0.26	No	298	1.87
280	0.34	No	302	2.08
298	0.30	No	304	0.61
302	0.23	No	326	2.46
304	0.24	No	352	0.95
313	0.78	Yes Medium confidence		
326	0.58	No		
352	0.54	No		

3.7. Gene-gene interaction of *CEBPA*

GeneMANIA and STRING were to predict interaction of *CEBPA* gene with other genes. GeneMANIA results showed that *CEBPA* has physical interaction with (*AFP, TGFB1, UHRF1, TOP2A, TK1, TRIM26, NCOA3, PREB, EBF1, ADH7, RUNX1T1, SLC2A4, MMP11, ONECUT1, TNF, DEFA3, UBP1, FDPS, LYZ, PCBP2*). GeneMANIA and STRING predictions of gene interactions are given in figure 5 and 6.

© 2020 Abasyn Journal of Life Sciences.

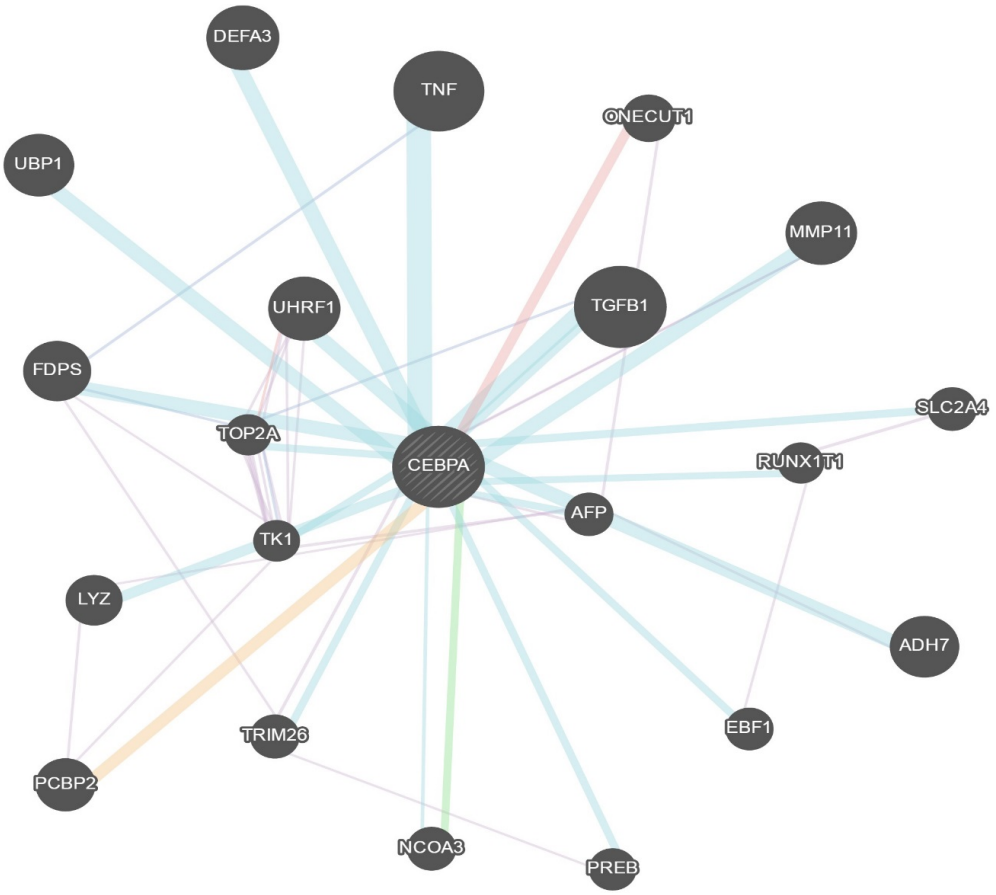

Fig. 7. Prediction of GeneMANIA for gene-gene interaction of Human *CEBPA* gene with other genes.

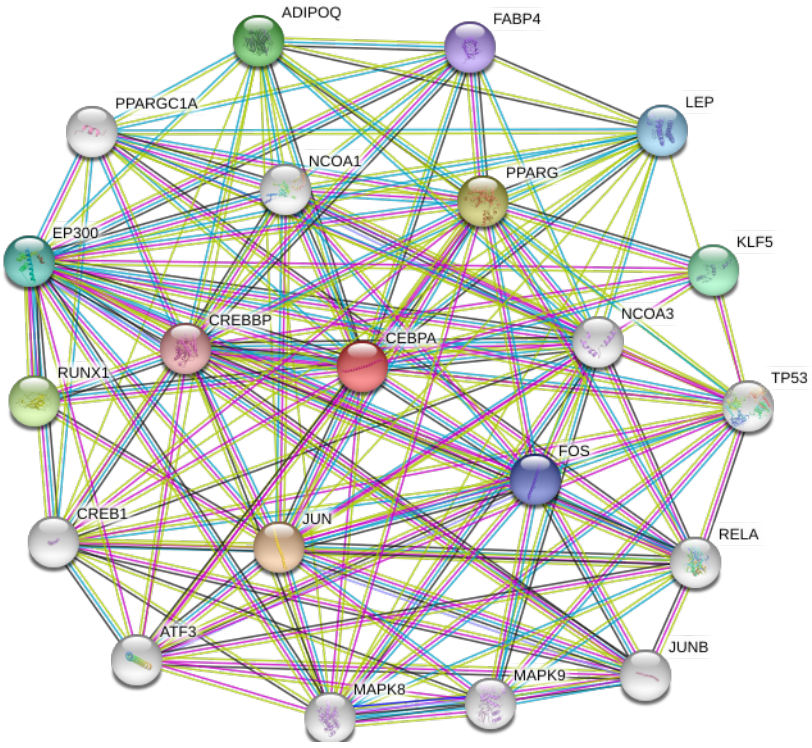

Fig. 8. Prediction of STRING for possible interaction of Human *CEBPA* gene with others.

© 2020 Abasyn Journal of Life Sciences.

4. DISCUSSION

In AML, homozygous *CEBPA* mutation is described in several studies. In three cases of AML, homozygous *CEBPA* mutation involve *CEBPA* locus with mitotic recombination[23]. In a recent study, mutations were found out in *CEBPA* gene, we used different approaches to find out nsSNPs, mentioned in methodology[24]. *CEBPA* loss of function mutation remained in association with deletion-9q within a noncomplex karyotype in AML[25]. Many studies have reported different types of SNPs, which have significance association with different types of cancers. As for AML, this is the first study of its type which we conducted in detail to highlight the disease associated SNPs for AML by investigating *CEBPA* gene. Largest database for SNPs is dbSNP, 1616 SNPs were recruited from it, which consisted of 269 nonsynonymous SNPs, 190 located in 5'UTR, 365 in 3'UTR, 152 coding synonymous and 540 others. The nsSNPs were further analysed in which 7 SNPs resulted in stop codon which means it has direct effect on structure of protein and can possibly lead to disease. Of the 269 nsSNPs, not enough data were available for 31 SNPs and were not included in our analysis.

Four bioinformatics tools were used to know the effect of nsSNPs on the function and structure of *CEBPA* protein. In all these four *in silico* tools PhD-SNP predicted 17.23% SNP&GO 8.82%, SIFT 39.92% and PROVEAN 25.63% to be deleterious or intolerant. 20 nsSNPs were predicted as deleterious by all the four tools, which were then submitted to PolyPhen2. PolyPhen2 predicted results as benign, possibly damaging and probably damaging, 15 nsSNPs were probably damaging was considered the most confident predictions with score of approximately 1 on the scale of 0 to 1 count score and were considered for further analysis. 15 nsSNPs were submitted to I-Mutant, which is used to predict protein stability of *CEBPA* gene for the selected nsSNPs and substitution of its amino acid, it predicted that 14 of these nsSNPs decrease protein stability and while substitution of G132C showed increase in the stability. It showed us that all these 14 nsSNPs might cause greater damage by decreasing *CEBPA* protein stability while G132C was skipped as it increases the stability. After protein modelling, we crosschecked our protein stability predictions in CUPSAT server, which predicts protein stability upon mutation based on protein structure. We found that our I-Mutant predictions were 81.25% in agreement with CUPSAT predictions, which shows that our predictive results are more reliable. To predict conservation profile of *CEBPA* protein, ConSurf was used, which uses combination of evolutionary conservation data and prediction of solvent accessibility. All those residues, which are highly conserved, are predicted to be functionally or structurally important based on their position on protein surface and core[26]. In protein-protein interaction vital amino acids are involved, they are supposed to be more conserved. All nsSNPs, which are present at conserved regions, are most damaging[27]. ConSurf showed us possible effects of nsSNPs in *CEBPA* profile, which showed us *CEBPA* protein structural representation. Location of identified nsSNPs were our priority although all results were given of amino acid residue of *CEBPA*. According to ConSurf L345, L317, L220, C133 and G132 were predicted to be buried, C357, R339, E334, R333, R327, K298, N292, E284, R156 and D63 were Exposed and functionally important while V328 and F82 were Buried and structurally important. Our results showed that, those nsSNPs were most damaging to *CEBPA* protein function and structure, which were located at highly conserved region.

For prediction of Post Translational Modifications (PTMs) sites, several different tools were used. No sites of methylation in *CEBPA* were predicted using GPS-MSP 3.0, which means that nsSNPs might have no role in affecting methylation site. NetPhos3.1 predicted 28 residues to have potential phosphorylation sites in which 16 were serine specific, 6 were threonine specific and 6 were tyrosine specific. BDM-PUB and UbPred were used for prediction of ubiquitination sites. Of the 15 total lysine residues, UbPred predicted 12 sites to be ubiquitinated while BDM-PUB predicted 6 sites to be ubiquitinated.

For prediction of gene-gene interactions, we used STRING and GeneMANIA. From STRING predictions, results showed combined score for each of the genes and found *PPARG, NCOA1, NCOA3, CREBBP, JUN, FOS, CREB1, ATF3, MAPK8, MAPK9, JUNB, RELA, TP53, RUNX1, EP300, PPARGC1A, ADIPOQ, FABP4, LEP* and *KLF5*. It is evident from GeneMANIA that *CEBPA* has physical interaction with *AFP, TGFB1, UHRF1, TOP2A, TK1, TRIM26, NCOA3, PREB, EBF1, ADH7, RUNX1T1, SLC2A4, MMP11, ONECUT1, TNF, DEFA3, UBP1, FDPS, LYZ* and *PCBP2*.

© 2020 Abasyn Journal of Life Sciences.

Our study provides all the analysis and information in detail, which is needed for damaging nsSNPs identification. There are certain limitations in every study and hence in our too. Our study is based on web servers and computer tools which are mainly based on statistical and mathematical algorithms. Therefore, experimental investigation is necessary to confirm these results. Our study provides an insight about 3D protein structure of *CEBPA* protein, its nsSNPs, its gene-gene interaction and potential PTM sites which might be helpful in future studies of *CEBPA* in order to understand its role in AML and all related diseases as well.

5. CONCLUSIONS

Our study concluded 11 nsSNPs to be the most deleterious ones. These nsSNPs included L345P, R339Q, R333C, V328G, R327W, L317Q, N292S, E284A, R156W, Y108N and F82L. These 11 nsSNPs are considered to be very important and can have active role in diseases associated with *CEBPA* gene, in AML. Our study also concluded that the *CEBPA* gene is correlated with other genes in many pathways. Any change in the function or structure of *CEBPA* protein will ultimately affect other pathways, thereby depicting its importance. These nsSNPs are significant for therapeutic targets and personalized medicines as well., and thyese SNPs can be used as diagnostic markers for AML. Although our study was in complete detail, there is still need of confirmation of our predicted results using *in vitro* strategies like modelling in mouse models.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Mudassir Khan https://orcid.org/0000-0002-8184-3557
Mehran Akhtar https://orcid.org/0000-0001-5783-0628
Maharij Haroon Jadoon[3] https://orcid.org/0000-0002-2469-1745
Dilawar Khan https://orcid.org/0000-0002-4543-4641

REFERENCES

1. Akhtar M, Jamal T, Jamal H, Din JU, Jamal M, Arif M, Arshad M, Jalil F. Identification of most damaging nsSNPs in human CCR6 gene: In silico analyses. International journal of immunogenetics. 2019 Dec;46(6):459-71.
2. Akhtar M, Jamal T, ud Din J, Hayat C, Rauf M, ul Haq SM, Khan RS, Shah AA, Jamal M, Jalil F. An in silico approach to characterize nonsynonymous SNPs and regulatory SNPs in human TOX3 gene. Journal of Genetics. 2019 Dec 1;98(5):104.
3. Akhtar M, Khan S, Ali Y, Haider S, ud Din J, Islam ZU, Jalil F. Association study of CCR6 rs3093024 with Rheumatoid Arthritis in a Pakistani cohort. Saudi Journal of Biological Sciences. 2020 Dec 1;27(12):3354-8.
4. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nature genetics. 2007 Nov;39(11):1329-37.
5. Heidema AG, Boer JM, Nagelkerke N, Mariman EC, Feskens EJ. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. BMC genetics. 2006 Dec 1;7(1):23.
6. Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2005 Jun 3;573(1-2):195-204.
7. Milanese M, Segat L, Crovella S. Transcriptional effect of DEFB1 gene 5' untranslated region polymorphisms. Cancer research. 2007 Jun 15;67(12):5997-.
8. Boffa MB, Maret D, Hamill JD, Bastajian N, Crainich P, Jenny NS, Tang Z, Macy EM, Tracy RP, Franco RF, Nesheim ME. Effect of single nucleotide polymorphisms on expression of the gene encoding thrombin-

© 2020 Abasyn Journal of Life Sciences.

activatable fibrinolysis inhibitor: a functional analysis. Blood, The Journal of the American Society of Hematology. 2008 Jan 1;111(1):183-9.
9. Szpirer, C., Riviere, M., Cortese, R., Nakamura, T., Islam, M.Q., Levan, G. and Szpirer, J., 1992. Chromosomal localization in man and rat of the genes encoding the liver-enriched transcription factors CEBP, DBP, and HNF1LFB-1>(CEBP, DBP, and transcription factor 1, TCF1, respectively) and of the hepatocyte growth factor/scatter factor gene (HGF). Genomics, 13(2), pp.293-300.
10. "CEBPA". (Genetics Home Reference. April 20, 2016). Retrieved April 25, 2016.
11. Roumier C, Fenaux P, Lafage M, Imbert M, Eclache V, Preudhomme C. New mechanisms of AML1 gene alteration in hematological malignancies. Leukemia. 2003 Jan;17(1):9-16.
12. Ohlsson E, Schuster MB, Hasemann M, Porse BT. The multifaceted functions of C/EBPα in normal and malignant haematopoiesis. Leukemia. 2016 Apr;30(4):767-75.
13. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC. A catalog of neutral and deleterious polymorphism in yeast. PLoS genet. 2008 Aug 29;4(8):e1000183.
14. Naveed M, Anwar F, Kazmi SK, Tariq F, Tehreem S, Abbas G, Irshad H, Anwar P, Ali A, Mehboob M. In Silico Screening and Pathway Analysis of Disease-Associated nsSNPs of MITF Gene: A study on Melanoma. International Journal of Computer Science and Information Security. 2017 Feb 1;15(2):31.
15. Wang H, Iakova P, Wilde M, Welm A, Goode T, Roesler WJ, Timchenko NA. C/EBPα arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. Molecular cell. 2001 Oct 26;8(4):817-28.
16. Ho PA, Alonzo TA, Gerbing RB, Pollard J, Stirewalt DL, Hurwitz C, Heerema NA, Hirsch B, Raimondi SC, Lange B, Franklin JL. Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. Blood. 2009 Jun 25;113(26):6558-66.
17. Lin TC, Hou HA, Chou WC, Ou DL, Yu SL, Tien HF, Lin LI. CEBPA methylation as a prognostic biomarker in patients with de novo acute myeloid leukemia. Leukemia. 2011 Jan;25(1):32-40.
18. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic acids research. 2016 Jul 8;44(W1):W344-50.
19. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics. 2016 Jun;54(1):5-6.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.
21. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research. 2010 Jul 1;38(suppl_2):W214-20.
22. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic acids research. 2002 Sep 1;30(17):3894-900.
23. Wouters BJ, Sanders MA, Lugthart S, Geertsma-Kleinekoort WM, van Drunen E, Beverloo HB, Löwenberg B, Valk PJ, Delwel R. Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. Leukemia. 2007 Nov;21(11):2382-4.
24. Mustafa MI, Mohammed ZO, Murshed NS, Elfadol NM, Abdelmoneim AH, Hassan MA. In Silico Genetics Revealing 5 Mutations in CEBPA Gene Associated With Acute Myeloid Leukemia. Cancer informatics. 2019 Aug;18:1176935119870817.
25. Fröhling S, Schlenk RF, Krauter J, Thiede C, Ehninger G, Haase D, Harder L, Kreitmeier S, Scholl C, Caligiuri MA, Bloomfield CD. Acute myeloid leukemia with deletion 9q within a noncomplex karyotype is associated with CEBPA loss-of-function mutations. Genes, Chromosomes and Cancer. 2005 Apr;42(4):427-32.
26. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics. 2004 May 22;20(8):1322-4.
27. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. Human molecular genetics. 2001 Oct 2;10(21):2319-28.

© 2020 Abasyn Journal of Life Sciences.

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License. To read the copy of this license please visit: https://creativecommons.org/licenses/by-nc/4.0/

© 2020 Abasyn Journal of Life Sciences.