ONLINE ISSN :3007-309X
PRINT ISSN :3007-3081

Journal of Medical & Health Science Review

ANALYSIS OF NON-SYNONYMOUS SNPS IN THE SET ONCOGENE AND THEIR IMPACT ON LEUKEMIA

Bushra Akbar[1], Fouzia Tanvir[1*], Atika Irfan[1], Asif Bilal[2], Arooba Nawaz[3], Huma Minhas[4], and Mariya Basharat[1]

[1]Department of Zoology, University of Okara, Pakistan

[2]Department Biological Sciences, The Superior University Lahore, Pakistan

Email: asif.bilal.sgd@superior.edu.pk

[3]Department Allied Health Sciences, The Superior University Lahore, Pakistan

[4]Bashir Family Clinic Main Chowk Chak 101 N.B Sargodha, Pakistan

ARTICLE INFO

Keywords:
SET, Leukemia, *In-silico* analysis, nsSNPs, Oncogene

Corresponding Author:

Fouzia Tanvir,
Department of Zoology,
University of Okara, Pakistan
Email: fouzia.tanvir@uo.edu.pk

ABSTRACT

Suppressor of variegation 3-9, Enhancer of Zeste, Trithorax (SET) is a family of protein. SET is a leukemia protein consisting up of Su (var)3-9, Ezh2, Trithorax (SET) domain in the protein family. In mammals SET has three pairs of family members including SETd1A/SETd1B, MLL1/MLL2, and MLL3/MLL4. SET links directly to various proteins involved in cell cycle regulation and DNA damage, physiological processes regulation, cell differentiation. It is expressed in various tissues such as brain, lung, heart, liver, the gonadal system, kidney, spleen and mainly detected in the nucleus in which it cooperates in the gene transcription regulation, stimulating and suppressing the various gene expression. SET protein, also called I2PP2A (Inhibitor 2 of PP2A), SET is a powerful inhibitor of protein phosphatase 2A (PP2A), separated from a chromosomal rearrangement at 9q34 in a case having acute undifferentiated leukemia. It is overexpressed in chronic myelogenous leukemia (CML). Sometimes it's also overexpressed in CLL relative to normal B cells. The objective of this study was to identify non-synonymous SNPs of SET and their role in causing leukemia utilizing computational analysis. For this purpose, different online tools like MAPP, SNAP, SNAP2 (to identify pathogenic SNPs), SIFT, Polyphen, Polyphen-2, fuNTRp, PhD-SNP, PredictSNP, PANTHER (disease associated SNPs), MetaSNP, SNP&GO, CONSURF (to check protein stability) I-Mutant, and Mu-Pro were used. Protein to protein interaction was detected by STRING, Post-translational modifications (PTMs) by Musitedeep, and protein secondary structure was detected by SOPMA. All SNPs in the SIFT and Polyphen-2 predicted as deleterious. In Polyphen, there were 113 SNPs predicted as possibly damaging, 21 showed probably damaging and 15 SNPs were predicted as benign. In SNAP2, 111 SNPs were predicted as diseased or effected SNPs, respectively. The PTMs found were phosphorylation at amino acid number 2 and 128; palmitoylation at number 13; pyrrolidone carboxylic acid at number 33 and glycosylation at number 41 and 46. The secondary structure of SET consists of Alpha helix (Hh) 121 is (43.68%), extended standard (Ee) 28 is (10.11%), beta turn (Tt) 3 is (1.08%) and random coil (Cc) 125 is (45.13%). SET has been shown strong protein to protein interaction with its own family as well as Granzyme A, Nucleoside diphosphate kinase A. It is concluded that these SNPs found in SET are involved in leukemia.

INTRODUCTION

Leukemia is a severe condition associated with white blood cells (WBC) that impacts both the bone marrow and the bloodstream in the human body. This illness has the potential to compromise the immune system significantly. Leukemia is primarily categorized into two types: acute and chronic, which are distinguished by the rate of progression. In acute leukemia, the affected WBC fail to function as normal cells, whereas in chronic leukemia, they may still exhibit some normal characteristics. Consequently, chronic leukemia can be particularly challenging to identify, as it can closely resemble normal WBC. Additionally, each type of leukemia can be further divided into two subtypes based on the morphology of the WBC: lymphoid and myeloid. Overall, there are four recognized subtypes of leukemia: Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myeloid Leukemia (CML). Accurate identification of leukemia and its specific types is crucial for hematologists to mitigate medical risks and determine appropriate treatment strategies (Ahmed *et al.,* 2019).

The conserved Su (var)3-9, Ezh2, Trithorax (SET) domain in this family of proteins is the same to that of S. cerevisiae SET1. SET1 is an H3K4 methyltransferase and is accountable for all three methylation conditions (mono-, di-, and tri-) of H3K4. The SET/MLL has three pairs of family members in mammals: SETd1A/SETd1B, MLL1/MLL2, and MLL3/MLL4. While the nomenclature has been confused between MLL2 and MLL4, we use here MLL2 for the protein encoded by the gene Kmt2b and gene Wbp7. MLL5, in spite of harboring an SET domain, is also inactive enzymatically (Yang *et al.,* 2017).

The MLL1 (mixed lineage leukemia) gene is situated on the long arm of chromosome 11, specifically at locus 11q23, and is linked to aggressive forms of leukemia. Abnormalities in the MLL1 gene are observed in approximately 70% of infant leukemia cases and about 10% of adult acute myeloid leukemia (AML) cases. While the overall cure rate for children diagnosed with leukemia is relatively high (78-85% five years post-therapy), the prognosis for leukemia associated with MLL1 rearrangements is considerably poorer, with an estimated cure rate of only 20% (Wiersma *et al.,* 2015; Christensen *et al.,* 2011)

A study shows that the participations of the Win motif-WDR5 interaction to complex meeting vary among the human SET1 family members also that such variations might be exploited to modify the enzymatic activities of SET1 family core complexes subset (Alicea-Velázquez *et al.,* 2016; Yang *et al.,* 2017; Separovich *et al.,* 2021).

It can be found in the cytoplasm, nucleus, and associated with the endoplasmic reticulum and plasma membrane, depending on the specific cellular environment (Dacol *et al.,* 2021). The

suppression of PP2A activity, or through specific interactions with various protein partners. This section aims to summarize our understanding of the physiological functions of SET (Bilal *et al.*, 2024).

The mechanisms that lead to its rearrangements are similarly not well understood. In this review, we present a summary of the relevant facts and hypotheses that we believe are essential for comprehending the breakage and rearrangements of KMT2A/MLL, with particular emphasis on therapy-related cases (Gole *et al.,* 2015).

Acute promyelocytic leukemia (APL) is recognized as a unique disease, primarily defined by the fusion of the promyelocytic leukemia (PML) gene with the retinoic acid receptor alpha (RARA). This fusion typically arises from the balanced translocation t (15;17) (q24.1; q21.2). APL accounts for roughly 5–10% of acute myeloid leukemia (AML) cases in the pediatric population, with its incidence increasing with age, peaking between 30 and 40 years (Conneely *et al.,* 2020; Bilal, 2021a,b).

Acute myeloid leukemia (AML) is a diverse hematological clonal disorder characterized by the buildup of immature myeloid cells in the bone marrow, referred to as leukemic blasts (L-blasts). Pediatric acute myeloid leukemia (pedAML) represents 20% of all leukemia cases in children. In the United States, the incidence of childhood AML is estimated to be between 7 and 8 cases per million among individuals aged 0 to 14 years. In the Netherlands and Belgium, around 30 to 35 children are diagnosed with AML each year, resulting in an average incidence of slightly more than one case per million (Depreter & B *et al.*, 2020).

The human MLL gene encodes a protein consisting of nearly 4000 amino acids, which is subsequently processed by the endopeptidase Taspase1. This processing yields two protein fragments that are then assembled into a high molecular weight protein complex. This complex is responsible for both reading and, more significantly, writing H3K4me3 chromatin signatures, which play a vital role in identifying active promoter regions within the genome. The MLL complex performs this epigenetic function in a manner that is specific to different cell types. These characteristics are essential for the maintenance of stem cells, developmental processes, and the functionality of fully differentiated cells, as the MLL complex preserves "transcriptional memory," thereby ensuring "tissue identity" (Marschalek & R *et al.*, 2015).

Main objectives of the study are to identify the nsSNPs of SET leukemia protein, their role in causing leukemia utilizing computational analysis as well as check the effect of nsSNPs on protein stability.

Materials and Methods

Dataset of SET (*Homo sapiens*) Protein

The Human SET protein gene sequence (Accession number: **NC_**AAA60318) and SET protein sequence (**N**P_ M93651.1) in FASTA format were obtained from NCBI. Relevant information from dbSNP and Protein ID from UniProt were also retrieved. OMIM provided additional gene and protein data for SET. Drug Bank databases were used for virtual screening of SET-related compounds. These datasets were collected for subsequent computational analysis.

Identification of disease associated SNPs

Seven computational tools, were utilized, like (Bilal *et al.,* 2022a,b), to see the harmful consequences of nsSNPs including SNPNEXUS from (SIFT and Polyphen). Other tools were CADD, PolyPhen2, PROVEAN, fuNTRp, SNAP2, PMut, SNP&GO, PHD-SNP, MetaSNP, MAPP, PANTHER, SNAP, and PredictSNP. The input was the amino acids sequences in FASTA format.

Effect of SNPs on protein stability and amino acid conservation

To forecast the influence of SNPs on protein stability, three tools were employed for the well-founded outcomes. MuPro, I-Mutant 2.0, and Consurf were used for this purpose.

Identification of PTM, secondary structure, and PPI

Musitedeep server, SOPMA, and STRING server ware used identify the post translational modification, secondary structure prediction, and protein to protein interaction.

Structural validation of native and mutant SET protein

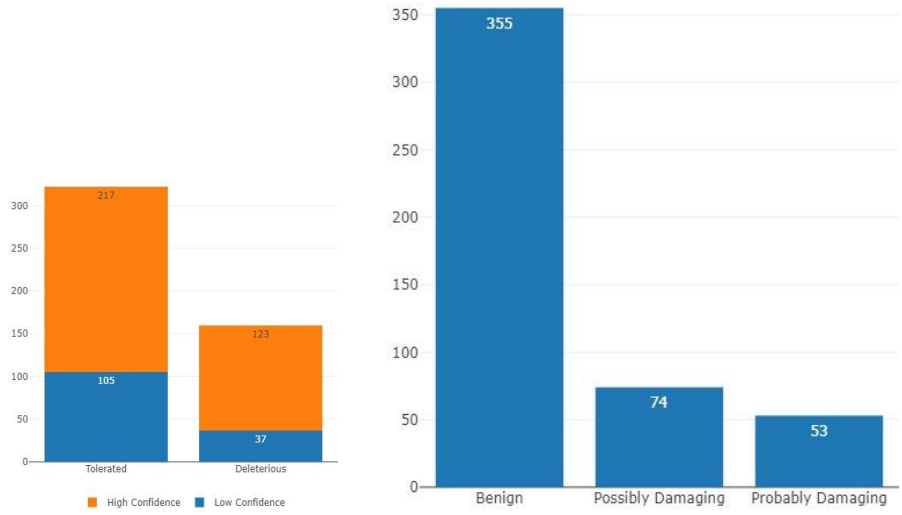
The wild-type and mutant protein structures were refined using ModRefiner, which improves structure quality through a mechanism of minimization. The quality of the refined structures was evaluated using RAMPAGE and ERRAT. RAMPAGE uses a Ramachandran plot analysis to assess amino acid distribution and stability in structure. The TM-align algorithm was utilized to compare wild-type and mutant structures, providing a measure of structural similarity through RMSD values and TM-score. A higher RMSD value predicts greater variations between the native and mutant structures, while a TM-score of 1 signifies a perfect match between the superimposed structures.

RESULTS

Identification of Harmful SNPs in SET

SNPNEXUS was used to observe a total of 7577 SNPs. Here two servers were utilized by SNPNEXUS to observe SNPs, SIFT and Polyphen. By SIFT server it was examined that 123 SNPs were deleterious with a score of less than 0.5. While 217 SNPs from SIFT were tolerable. SIFT identifies harmful SNPs on the basis of conservation of residue in matched

sequence to closely linked sequence. The score in SIFT vary from 0 to 1. The algorithm predicts alterations in amino acids with score less than 0.05 as deleterious and a score above 0.05 will be declared as tolerated (Figure 2A). In Polyphen server, the alterations in amino acids have adverse effects on protein function. Its result was: probably damaging, possibly damaging, and or benign. In our results, 53 SNPs were showing probably damaging effects, 74 possibly damaging, and 355 SNPs were benign (Figure 2B).



SIFT **Polyphen**

Figure 2: SIFT (A) and Polyphen (B) barplot of deleterious SNP

Identification of pathogenic SNPs in SET

After analyzing through SIFT and Polyphen, Polyphen-2 and SNAP2 were used further for the identification of pathogenic SNPs in SET protein. All SNPs in the SIFT and Polyphen-2 predicted as deleterious. In Polyphen, there were 113 SNPs predicted as possibly damaging, 21 showed probably damaging and 15 SNPs were predicted as benign. In SNAP2, 111 SNPs were predicted as diseased or effected SNPs, respectively (Table 1).

Identification of pathogenic nsSNPs										
		SIFT		**Polyphen**		**PPH2**		**fuNTRp**	**SNAP2**	
rsIDs	**A.A**	**PREDICTION**	**SCORE>0.5**	**prediction**	**Sc or**	**prediction**	**PPH2 SCORE=1**	**Score**	**PREDICTION**	**SCORE**

					e					
rs12289 85010	M1I	Deleteri ous - Low Confide nce	0	Possi bly Dama ging	0.3 8	Delete rious	60	0.78	effect	28
rs12289 85010	M1I	Deleteri ous - Low Confide nce	0.03	Possi bly Dama ging	0.3 8	Delete rious	55	0.79	effect	91
rs12879 05948	P3R	Deleteri ous - Low Confide nce	0.01	Possi bly Dama ging	0.4 07	Delete rious	75	0.81	effect	51
rs12879 05948	P3R	Deleteri ous - Low Confide nce	0	Possi bly Dama ging	0.4 07	Delete rious	60	0.9	effect	50
rs11807 24092	L9P	Deleteri ous - Low Confide nce	0	Possi bly Dama ging	0.0 48	Delete rious	87	0.76	effect	78
rs11807 24092	L9P	Deleteri ous - Low Confide nce	0	Possi bly Dama ging	0.0 48	Delete rious	55	0.59	effect	72
rs99206	P10	Deleteri	0.01	Possi	0	Delete	76	0.61	effect	46

7058	L	ous - Low Confidence		bly Damaging		rious				
rs971609222	Q12P	Deleterious - Low Confidence	0	Possibly Damaging	0	Deleterious	79	0.68	effect	88
rs971609222	Q12P	Deleterious - Low Confidence	0.04	Possibly Damaging	0	Deleterious	83	0.83	effect	89
rs923629041	K15N	Deleterious - Low Confidence	0.04	Possibly Damaging	0	Deleterious	86	0.71	neutral	-55
rs923629041	K15N	Deleterious - Low Confidence	0.01	Possibly Damaging	0.058	Deleterious	60	0.88	effect	63
rs1361492173	P16R	Deleterious - Low Confidence	0.01	Possibly Damaging	0.058	Deleterious	87	0.37	effect	48
rs1361492173	P16L	Deleterious - Low Confide	0.01	Possibly Damaging	0	Deleterious	76	0.76	effect	63

		nce								
rs900458307	P19L	Deleterious - Low Confidence	0.01	Possibly Damaging	0.001	Deleterious	79	0.67	neutral	-85
rs138894709	L22P	Deleterious - Low Confidence	0.01	Possibly Damaging	0	Deleterious	61	0.55	effect	16
rs1208440934	E26K	Deleterious - Low Confidence	0	Possibly Damaging	0.001	Deleterious	51	0.69	effect	53
rs1442641412	S28L	Deleterious - Low Confidence	0.03	Possibly Damaging	0	Deleterious	86	0.71	effect	21
rs1181003970	A31V	Deleterious - Low Confidence	0.02	Possibly Damaging	0	Deleterious	60	0.88	effect	9
rs1268249403	L6R	Deleterious - Low Confidence	0.03	Possibly Damaging	0	Deleterious	87	0.37	effect	17
rs1319521099	L7R	Deleterious -	0.01	Possibly	0	Deleterious	76	0.61	effect	19

		Low Confidence		Damaging						
rs12189 33656	P8 H	Deleterious - Low Confidence	0.01	Possibly Damaging	0.052	Deleterious	79	0.67	neutral	-51
rs12189 33656	P8L	Deleterious - Low Confidence	0.02	Possibly Damaging	0.052	Deleterious	60	0.65	effect	56
rs11411 38	P4L	Deleterious	0	Possibly Damaging	0	Deleterious	87	0.46	effect	76
rs11542 565	L13 F	Deleterious	0.05	Probably Damaging	0.063	Deleterious	87	0.63	effect	78
rs77569 3359	M1 R	Deleterious - Low Confidence	0	Probably Damaging	0.063	Deleterious	60	0.6	effect	78
rs10248 56362	E13 K	Deleterious - Low Confidence	0.03	Probably Damaging	0.066	Deleterious	55	0.6	effect	38
rs10248 56362	E13 K	Deleterious -	0.03	Probably	0.066	Deleterious	65	0.52	effect	71

		Low Confidence		Damaging						
rs1276443619	Q29R	Deleterious	0	Probably Damaging	0	Deleterious	51	0.54	effect	86
rs1276443619	Q17R	Deleterious	0	Probably Damaging	0	Deleterious	61	0.67	effect	83
rs1276443619	Q18R	Deleterious	0	Probably Damaging	0.07	Deleterious	61	0.65	effect	84
rs1276443619	Q42R	Deleterious	0.04	Probably Damaging	0.011	Deleterious	72	0.46	effect	48
rs1276443619	Q20R	Deleterious	0.03	Probably Damaging	0.07	Deleterious	55	0.81	effect	2
rs866872429	H89Y	Deleterious	0	Probably Damaging	0	Deleterious	54	0.6	effect	71
rs866872429	H77N	Deleterious	0.03	Probably Damaging	0	Deleterious	44	0.6	effect	74
rs866872429	H102N	Deleterious	0.02	Probably	0.005	Deleterious	7	0.9	effect	1

				Damaging						
rs866872429	H102Y	Deleterious	0.01	Probably Damaging	0.084	Deleterious	59	0.76	effect	34
rs866872429	H80N	Deleterious	0.02	Probably Damaging	0.078	Deleterious	35	0.59	effect	52
rs866872429	H80Y	Deleterious	0.01	Probably Damaging	0.113	Deleterious	24	0.61	neutral	-77
rs483352707	A94P	Deleterious	0.05	Probably Damaging	0.084	Deleterious	13	0.68	effect	57
rs483352707	A82P	Deleterious	0.05	Probably Damaging	0.078	Deleterious	29	0.83	neutral	-90
rs483352707	A83P	Deleterious	0.04	Possibly Damaging	0.113	Deleterious	50	0.71	effect	49
rs483352707	A107P	Deleterious	0.05	Possibly Damaging	0.357	Deleterious	48	0.88	neutral	-53
rs483352707	A85P	Deleterious	0.04	Possibly Dama	0	Deleterious	44	0.37	effect	67

				ging						
rs14087 42550	A10 3S	Deleteri ous	0.04	Possi bly Dama ging	0.3 57	Delete rious	41	0.76	effect	30
rs14087 42550	A91 S	Deleteri ous	0.04	Possi bly Dama ging	0	Delete rious	60	0.67	effect	81
rs14087 42550	A92 S	Deleteri ous	0.04	Possi bly Dama ging	0.0 07	Delete rious	46	0.55	effect	83
rs14087 42550	A11 6S	Deleteri ous	0.04	Possi bly Dama ging	0.0 07	Delete rious	48	0.69	effect	32
rs14087 42550	A94 S	Deleteri ous	0.04	Possi bly Dama ging	0.0 06	Delete rious	6	0.71	effect	26
rs91329 9138	T10 8A	Deleteri ous	0.05	Possi bly Dama ging	0.0 06	Delete rious	47	0.88	effect	29
rs91329 9138	T96 A	Deleteri ous	0.05	Possi bly Dama ging	0.0 03	Delete rious	46	0.37	effect	39
rs91329 9138	T97 A	Deleteri ous	0.05	Possi bly Dama ging	0.0 03	Delete rious	27	0.61	effect	81

rs913299138	T121A	Deleterious	0.05	Possibly Damaging	0.175	Deleterious	88	0.67	neutral	-19
rs913299138	T99A	Deleterious	0.05	Possibly Damaging	0.102	Deleterious	60	0.65	effect	53
rs374932364	Y121H	Deleterious	0.02	Possibly Damaging	0.014	Deleterious	76	0.63	effect	37
rs374932364	Y122H	Deleterious	0.02	Possibly Damaging	0.014	Deleterious	79	0.6	neutral	-19
rs374932364	Y146H	Deleterious	0.02	Possibly Damaging	0.005	Deleterious	44	0.6	neutral	-62
rs374932364	Y124H	Deleterious	0.02	Possibly Damaging	0.003	Deleterious	35	0.52	effect	47
rs765603755	E123D	Deleterious	0.05	Possibly Damaging	0	Deleterious	40	0.54	effect	35
rs765603755	E124D	Deleterious	0.05	Possibly Damaging	0.003	Deleterious	51	0.67	neutral	-6
rs76560	E14	Deleteri	0.05	Possi	0.0	Delete	7	0.78	effect	73

3755	8D	ous		bly Damaging	39	rious				
rs765603755	E126D	Deleterious	0.05	Possibly Damaging	0	Deleterious	59	0.79	effect	70
rs1429567543	K141R	Deleterious	0.02	Possibly Damaging	0.443	Deleterious	35	0.81	effect	17
rs1429567543	K132R	Deleterious	0.02	Possibly Damaging	0.027	Deleterious	50	0.61	effect	23
rs1013985707	F143L	Deleterious	0.03	Possibly Damaging	0.009	Deleterious	48	0.68	effect	41
rs1013985707	F131L	Deleterious	0.03	Possibly Damaging	0.003	Deleterious	44	0.83	effect	50
rs1013985707	F132L	Deleterious	0.03	Possibly Damaging	0.017	Deleterious	41	0.71	effect	32
rs1013985707	F156L	Deleterious	0.03	Possibly Damaging	0.001	Deleterious	60	0.88	effect	7
rs1367036404	S148R	Deleterious	0.01	Possibly	0	Deleterious	48	0.76	effect	51

				Damaging						
rs1367036404	S136R	Deleterious	0.01	Benign	0.003	Deleterious	6	0.67	effect	81
rs1367036404	S137R	Deleterious	0.01	Benign	0.003	Deleterious	47	0.55	effect	88
rs1367036404	S161R	Deleterious	0.01	Benign	0.25	Deleterious	46	0.69	effect	68
rs1367036404	S139R	Deleterious	0.01	Benign	0.25	Deleterious	27	0.71	effect	44
rs1429859567	D150G	Deleterious	0.01	Benign	0.097	Deleterious	88	0.88	effect	64
rs1429859567	D138G	Deleterious	0.01	Possibly Damaging	0.097	Deleterious	35	0.37	effect	88
rs1429859567	D139G	Deleterious	0.01	Possibly Damaging	0.227	Deleterious	40	0.61	effect	80
rs1429859567	D163G	Deleterious	0.01	Possibly Damaging	0.227	Deleterious	51	0.67	effect	57
rs1429859567	D141G	Deleterious	0.01	Possibly Damaging	0.227	Deleterious	7	0.65	neutral	-1
rs764645296	S153L	Deleterious	0.05	Possibly Damaging	0.227	Deleterious	59	0.46	effect	82
rs76464	S14	Deleteri	0.04	Possi	0.0	Delete	24	0.63	effect	14

5296	2L	ous		bly Damaging	42	rious				
rs764645296	S166L	Deleterious	0.04	Possibly Damaging	0.069	Deleterious	13	0.6	effect	3
rs764645296	S144L	Deleterious	0.05	Possibly Damaging	0.069	Deleterious	29	0.6	effect	22
rs1168200235	I158M	Deleterious	0	Possibly Damaging	0.216	Deleterious	50	0.52	effect	5
rs1168200235	I146M	Deleterious	0	Possibly Damaging	0.216	Deleterious	48	0.54	effect	89
rs1168200235	I147M	Deleterious	0	Possibly Damaging	0.12	Deleterious	44	0.67	effect	70
rs1168200235	I171M	Deleterious	0	Possibly Damaging	0.12	Deleterious	41	0.65	effect	89
rs1168200235	I149M	Deleterious	0	Possibly Damaging	0.641	Deleterious	60	0.46	effect	89
rs980437151	K149R	Deleterious	0.03	Possibly	0	Deleterious	48	0.63	effect	88

				Damaging						
rs980437151	K150R	Deleterious	0.03	Possibly Damaging	0	Deleterious	6	0.6	effect	89
rs980437151	K174R	Deleterious	0.03	Possibly Damaging	0.025	Deleterious	47	0.6	effect	70
rs980437151	K152R	Deleterious	0.03	Possibly Damaging	0.025	Deleterious	46	0.52	effect	89
rs142648600	S170W	Deleterious	0.05	Possibly Damaging	0.02	Deleterious	27	0.54	effect	89
rs142648600	S159W	Deleterious	0.04	Possibly Damaging	0	Deleterious	60	0.65	effect	88
rs142648600	S183W	Deleterious	0.03	Possibly Damaging	0	Deleterious	55	0.46	effect	89
rs142648600	S161W	Deleterious	0.03	Possibly Damaging	0	Deleterious	75	0.81	effect	70
rs1268994417	T194I	Deleterious	0.02	Possibly Dama	0	Deleterious	60	0.63	effect	89

				ging						
rs12689 94417	T18 2I	Deleteri ous	0.02	Benig n	0.2 92	Delete rious	87	0.6	effect	17
rs12689 94417	T18 3I	Deleteri ous	0.02	Benig n	0.2 92	Delete rious	55	0.6	effect	31
rs12689 94417	T18 5I	Deleteri ous	0.02	Benig n	0.2 92	Delete rious	79	0.54	effect	10
rs14774 48265	H19 6Y	Deleteri ous	0	Benig n	0.5 91	Delete rious	83	0.67	effect	80
rs14774 48265	H18 4Y	Deleteri ous	0	Possi bly Dama ging	0	Delete rious	86	0.65	effect	28
rs14774 48265	H18 5Y	Deleteri ous	0	Possi bly Dama ging	0	Delete rious	60	0.46	effect	91
rs14774 48265	H20 9Y	Deleteri ous	0	Possi bly Dama ging	0.2 32	Delete rious	87	0.81	effect	51
rs13236 58124	D21 0V	Deleteri ous	0.03	Possi bly Dama ging	0.2 63	Delete rious	79	0.6	effect	78
rs13236 58124	D19 8V	Deleteri ous	0.03	Possi bly Dama ging	0.0 85	Delete rious	61	0.52	neutral	-71
rs13236 58124	D19 9V	Deleteri ous	0.03	Possi bly Dama ging	0.0 85	Delete rious	51	0.54	effect	46

rs1323658124	D223V	Deleterious	0.03	Possibly Damaging	0.597	Deleterious	86	0.67	effect	88
rs1323658124	D201V	Deleterious	0.03	Possibly Damaging	0.381	Deleterious	60	0.65	effect	89
rs756885899	M225T	Deleterious	0.03	Possibly Damaging	0.06	Deleterious	87	0.46	effect	70
rs756885899	M238T	Deleterious	0.04	Possibly Damaging	0.023	Deleterious	76	0.81	effect	89
rs1294476257	E255V	Deleterious	0.01	Possibly Damaging	0.187	Deleterious	71	0.52	effect	66
rs1294476257	E233V	Deleterious	0.01	Possibly Damaging	0.271	Deleterious	87	0.54	neutral	-42
rs753565206	D276E	Deleterious - Low Confidence	0	Possibly Damaging	0.271	Deleterious	60	0.67	effect	76
rs753565206	D264E	Deleterious - Low Confide	0	Possibly Damaging	0.378	Deleterious	55	0.65	neutral	-42

		nce								
rs753565206	D265E	Deleterious - Low Confidence	0	Possibly Damaging	0.378	Deleterious	65	0.46	effect	88
rs753565206	D289E	Deleterious - Low Confidence	0	Possibly Damaging	0.067	Deleterious	51	0.81	effect	89
rs753565206	D267E	Deleterious - Low Confidence	0	Possibly Damaging	0.072	Deleterious	61	0.63	effect	70
rs754819669	D277V	Deleterious - Low Confidence	0	Possibly Damaging	0.072	Deleterious	61	0.6	effect	89
rs754819669	D265V	Deleterious - Low Confidence	0	Possibly Damaging	0.072	Deleterious	72	0.6	effect	89
rs754819669	D266V	Deleterious - Low Confidence	0	Possibly Damaging	0.104	Deleterious	55	0.52	effect	89
rs754819669	D290V	Deleterious -	0	Possibly	0.095	Deleterious	74	0.54	neutral	-68

		Low Confidence		Damaging						
rs754819669	D268V	Deleterious - Low Confidence	0	Possibly Damaging	0.153	Deleterious	54	0.6	effect	88

4.3. Identification of disease associated nsSNPs in SET

All the SNPs were examined for the diseased association. There were 99, 108, 115 SNPs predicted as diseased by SNP&GO, PhD-SNP, and Meta-SNP tools, respectively. There were 94, 81 and 69 SNPs predicted as deleterious by PredictSNP, MAPP, and SNAP, respectively. 93 SNPs predicted as disordered by PredictProtein and 109 SNPs were predicted by PANTHER (Table 2).

Diseases associated nsSNPs														
		SNP &GO	**PhD-SNP**		**Predict SNP**		**MAPP**		**SNAP**		**MetaSNP**		**PANTHER**	
rsIDs	**A.A**	**PREDICTION**	**PREDICTION**	**R1**	**Prediction**	**%age expected accuracy**	**Prediction**	**%age expected accuracy**	**Prediction**	**%age expected accuracy**	**Prediction**	**scoreRI**	**Prediction**	**Pdel**
rs122898501 0	M1I	Disease	Disease	41	Neutral	41	Neutral	56	Neutral	86	Disease	0.32	Neutral	63
rs122898 5	M1I	Disease	Disease	63	Deleterio	77	Deleteri	82	Deleteri	78	Disease	0.8	Disease	82

010					us		ous		ous			3		
rs1287905948	P3R	Disease	Disease	63	Deleterious	92	Neutral	77	Neutral	81	Disease	0.18	Disease	84
rs1287905948	P3R	Disease	Disease	82	Neutral	81	Neutral	86	Neutral	66	Disease	0.45	Disease	92
rs1180724092	L9P	Disease	Disease	84	Neutral	41	Neutral	87	Neutral	76	Disease	0.27	Disease	66
rs1180724092	L9P	Disease	Disease	92	Deleterious	63	Deleterious	45	Deleterious	46	Disease	0.27	Disease	56
rs992067058	P10L	Disease	Disease	66	Deleterious	63	Deleterious	78	Neutral	77	Disease	0.29	Disease	82
rs971609222	Q12P	Disease	Disease	56	Deleterious	82	Deleterious	41	Neutral	87	Disease	0.69	Disease	77
rs971609222	Q12P	Disease	Disease	82	Neutral	84	Deleterious	63	Neutral	45	Disease	0.93	Disease	86
rs923629041	K15N	Disease	Disease	77	Deleterious	92	Deleterious	63	Deleterious	78	Disease	0.67	Disease	77
rs923629041	K15N	Disease	Disease	86	Deleterious	66	Deleterious	82	Deleterious	77	Disease	0.03	Disease	63
rs1361492173	P16R	Disease	Disease	77	Deleterious	56	Deleterious	84	Deleterious	98	Disease	0.16	Disease	98
rs1361492	P16L	Neutral	Disease	63	Deleterio	82	Deleteri	92	Deleteri	41	Disease	0.2	Disease	45

173					us		ous		ous					
rs90045833 07	P19L	Disease	Disease	98	Deleterious	77	Deleterious	66	Deleterious	77	Disease	0.121	Disease	62
rs138894709	L22P	Disease	Disease	45	Deleterious	86	Deleterious	56	Deleterious	92	Disease	0.16	Disease	43
rs1208440934	E26K	Neutral	Neutral	62	Neutral	77	Neutral	82	Neutral	81	Neutral	0.09	Neutral	88
rs1442641412	S28L	Disease	Disease	43	Deleterious	63	Deleterious	77	Deleterious	41	Disease	0.9	Disease	65
rs1181003970	A31V	Disease	Disease	88	Deleterious	98	Neutral	86	Deleterious	63	Disease	0.88	Disease	77
rs1268249403	L6R	Disease	Disease	65	Deleterious	45	Deleterious	77	Deleterious	63	Disease	0.12	Disease	87
rs1319521099	L7R	Neutral	Neutral	77	Neutral	62	Deleterious	63	Neutral	82	Neutral	0.59	Neutral	77
rs1218933656	P8H	Disease	Disease	87	Neutral	43	Deleterious	98	Neutral	84	Disease	0.27	Disease	41
rs1218933656	P8L	Disease	Disease	77	Deleterious	88	Deleterious	45	Deleterious	92	Disease	0.03	Disease	77
rs1141138	P4L	Disease	Disease	77	Deleterious	65	Deleterious	62	Neutral	66	Disease	0.2	Disease	92
rs143	A	Neutr	Neutr	5	Neut	77	Del	43	Neu	56	Neu	0.	Neut	8

7350212	6D	al	al	6	ral		eterious		tral		tral	07	ral	1
rs1154256 5	L13F	Neutral	Neutral	81	Deleterious	87	Deleterious	88	Deleterious	82	Disease	0.06	Neutral	41
rs7756933 59	M1R	Disease	Disease	75	Neutral	77	Neutral	65	Neutral	77	Neutral	0.05	Disease	63
rs1024856 362	E13K	Neutral	Disease	77	Neutral	77	Neutral	77	Neutral	86	Neutral	0.15	Neutral	45
rs1024856 362	E13K	Disease	Disease	84	Deleterious	56	Deleterious	87	Deleterious	87	Disease	0.28	Disease	78
rs1276443 619	Q29R	Disease	Disease	86	Deleterious	81	Deleterious	77	Deleterious	98	Disease	0.09	Disease	41
rs1276443 619	Q17R	Neutral	Disease	86	Neutral	75	Neutral	77	Neutral	41	Neutral	0.19	Neutral	63
rs1276443 619	Q18R	Disease	Disease	78	Deleterious	77	Deleterious	56	Deleterious	77	Disease	0.48	Disease	63
rs1276443 619	Q42R	Disease	Disease	81	Deleterious	84	Deleterious	81	Neutral	92	Disease	0.34	Disease	82
rs1276443 619	Q20R	Disease	Disease	66	Deleterious	86	Deleterious	75	Deleterious	81	Disease	0.1	Disease	84
rs8668724 29	H89N	Disease	Disease	76	Deleterious	86	Deleterious	77	Deleterious	41	Disease	0.07	Disease	92
rs866	H	Disea	Disea	4	Dele	78	Del	84	Del	63	Dis	0.	Dise	6

872429	89Y	se	se	6	terious		eterious		eterious		ease	32	ase	6
rs8668724 29	H77N	Disease	Neutral	77	Deleterious	81	Deleterious	86	Deleterious	63	Disease	0.83	Disease	56
rs8668724 29	H77Y	Disease	Disease	87	Deleterious	66	Deleterious	86	Deleterious	82	Disease	0.18	Disease	82
rs8668724 29	H78N	Neutral	Neutral	45	Deleterious	76	Neutral	78	Deleterious	84	Disease	0.45	Disease	77
rs8668724 29	H78Y	Neutral	Neutral	78	Neutral	46	Neutral	81	Neutral	92	Neutral	0.27	Neutral	86
rs8668724 29	H102N	Neutral	Neutral	77	Neutral	77	Deleterious	66	Neutral	66	Disease	0.27	Disease	77
rs8668724 29	H102Y	Neutral	Disease	98	Deleterious	87	Deleterious	76	Neutral	56	Disease	0.29	Neutral	63
rs8668724 29	H80N	Disease	Disease	41	Deleterious	45	Neutral	46	Deleterious	82	Disease	0.69	Disease	98
rs8668724 29	H80Y	Neutral	Neutral	77	Deleterious	78	Deleterious	77	Deleterious	77	Disease	0.93	Neutral	45
rs4833527 07	A94P	Neutral	Neutral	92	Neutral	77	Neutral	87	Neutral	86	Neutral	0.67	Neutral	62
rs4833527	A82	Neutral	Neutral	81	Neutral	98	Neutral	45	Neutral	77	Neutral	0.0	Neutral	43

07	P											3		
rs4833527 07	A 83 P	Disease	Disease	41	Deleterious	41	Deleterious	78	Deleterious	63	Disease	0.16	Disease	88
rs4833527 07	A 107P	Disease	Disease	63	Deleterious	77	Neutral	77	Neutral	98	Disease	0.2	Disease	65
rs4833527 07	A 85 P	Disease	Disease	63	Deleterious	92	Deleterious	98	Neutral	45	Disease	0.121	Disease	77
rs1408742550	A 103S	Neutral	Neutral	82	Neutral	81	Neutral	41	Neutral	62	Disease	0.16	Neutral	87
rs1408742550	A 91 S	Disease	Disease	84	Deleterious	41	Neutral	77	Neutral	43	Neutral	0.09	Disease	77
rs1408742550	A 92 S	Disease	Neutral	92	Neutral	63	Neutral	92	Neutral	88	Disease	0.9	Disease	77
rs1408742550	A 116S	Disease	Disease	66	Deleterious	63	Deleterious	81	Deleterious	65	Disease	0.88	Disease	56
rs1408742550	A 94 S	Neutral	Neutral	56	Neutral	82	Neutral	41	Neutral	77	Neutral	0.12	Neutral	81
rs9132991 38	T108 A	Disease	Disease	82	Deleterious	84	Deleterious	63	Deleterious	87	Disease	0.59	Disease	75
rs9132991 38	T96 A	Disease	Disease	77	Neutral	92	Neutral	63	Neutral	77	Disease	0.27	Disease	77
rs913	T9	Disea	Disea	8	Dele	66	Del	82	Del	77	Dis	0.	Dise	8

299138	7A	se	se	6	terious		eterious		eterious		ease	03	ase	4
rs9132991138	T121A	Disease	Disease	77	Deleterious	56	Deleterious	84	Deleterious	56	Disease	0.16	Disease	86
rs9132991138	T99A	Disease	Disease	63	Deleterious	82	Deleterious	92	Deleterious	81	Disease	0.2	Disease	86
rs1196624947	D125G	Neutral	Neutral	98	Neutral	77	Neutral	66	Neutral	75	Disease	0.121	Disease	78
rs1196624947	D113G	Disease	Disease	45	Neutral	86	Neutral	56	Neutral	77	Disease	0.16	Disease	81
rs1196624947	D114G	Neutral	Neutral	62	Neutral	77	Neutral	82	Neutral	84	Neutral	0.09	Neutral	66
rs1196624947	D138G	Disease	Disease	43	Deleterious	63	Neutral	77	Neutral	86	Disease	0.9	Disease	76
rs1196624947	D116G	Neutral	Neutral	88	Neutral	98	Neutral	86	Neutral	86	Neutral	0.88	Neutral	46
rs1222712623	E130G	Neutral	Neutral	65	Neutral	45	Neutral	77	Deleterious	78	Disease	0.12	Disease	77
rs1222712	E118	Neutral	Neutral	77	Deleterio	62	Deleteri	63	Neutral	81	Disease	0.5	Disease	87

623	G				us		ous					9		
rs1222712623	E119G	Neutral	Neutral	87	Neutral	43	Deleterious	98	Neutral	66	Disease	0.27	Disease	45
rs1222712623	E143G	Neutral	Neutral	77	Neutral	88	Neutral	45	Neutral	76	Neutral	0.03	Neutral	78
rs1222712623	E121G	Disease	Disease	77	Deleterious	65	Deleterious	62	Neutral	46	Disease	0.16	Disease	77
rs374932364	Y133H	Disease	Disease	56	Deleterious	77	Deleterious	43	Neutral	77	Disease	0.559	Disease	98
rs374932364	Y121H	Neutral	Neutral	81	Deleterious	87	Deleterious	88	Deleterious	87	Disease	0.03	Disease	41
rs374932364	Y122H	Disease	Disease	75	Deleterious	77	Deleterious	65	Deleterious	45	Disease	0.16	Neutral	77
rs374932364	Y146H	Disease	Disease	77	Deleterious	77	Deleterious	77	Deleterious	78	Disease	0.2	Neutral	92
rs374932364	Y124H	Neutral	Disease	81	Deleterious	86	Deleterious	77	Neutral	45	Neutral	0.03	Disease	65
rs765603755	E123D	Neutral	Disease	41	Deleterious	87	Deleterious	87	Neutral	62	Disease	0.2	Disease	77

rs1013985707	F143L	Disease	Disease	86	Deleterious	41	Deleterious	86	Neutral	81	Disease	0.34	Disease	65
rs1013985707	F131L	Disease	Disease	87	Deleterious	63	Neutral	78	Deleterious	75	Disease	0.1	Disease	77
rs1013985707	F132L	Neutral	Disease	45	Deleterious	63	Neutral	81	Neutral	77	Neutral	0.07	Neutral	87
rs1013985707	F156L	Neutral	Disease	78	Neutral	82	Neutral	66	Neutral	84	Neutral	0.32	Neutral	77
rs1013985707	F134L	Disease	Disease	77	Deleterious	84	Neutral	76	Deleterious	86	Disease	0.83	Disease	77
rs1367036404	S148R	Neutral	Neutral	98	Deleterious	92	Neutral	46	Neutral	86	Neutral	0.18	Neutral	56
rs1367036404	S136R	Disease	Disease	41	Neutral	66	Neutral	77	Neutral	78	Disease	0.45	Disease	81
rs1367036404	S137R	Neutral	Neutral	77	Neutral	56	Neutral	87	Neutral	81	Neutral	0.27	Neutral	75
rs1367036404	S161R	Disease	Disease	92	Deleterious	82	Neutral	45	Deleterious	66	Disease	0.27	Disease	77
rs1367036404	S139R	Disease	Disease	81	Deleterious	77	Neutral	78	Deleterious	76	Disease	0.29	Disease	84
rs1429859567	D150	Disease	Neutral	41	Neutral	86	Neutral	77	Deleterious	46	Disease	0.69	Disease	86

	G													
rs1429859567	D138G	Disease	Neutral	63	Deleterious	77	Neutral	98	Neutral	77	Disease	0.93	Disease	86
rs1429859567	D139G	Neutral	Neutral	63	Neutral	63	Neutral	41	Neutral	87	Neutral	0.67	Neutral	78
rs1429859567	D163G	Disease	Disease	82	Deleterious	98	Neutral	77	Deleterious	46	Disease	0.03	Disease	81
rs1429859567	D141G	Disease	Disease	84	Deleterious	45	Deleterious	92	Deleterious	77	Disease	0.16	Disease	63
rs76464529 6	S153L	Disease	Disease	92	Deleterious	62	Neutral	81	Neutral	87	Disease	0.2	Disease	98
rs76464529 6	S141L	Disease	Disease	66	Deleterious	43	Deleterious	41	Deleterious	45	Disease	0.121	Disease	45
rs76464529 6	S142L	Disease	Disease	56	Deleterious	88	Deleterious	63	Deleterious	78	Disease	0.16	Disease	62
rs76464529 6	S166L	Neutral	Neutral	82	Neutral	65	Deleterious	63	Neutral	77	Neutral	0.09	Neutral	43
rs76464529 6	S144L	Disease	Disease	77	Deleterious	77	Deleterious	82	Deleterious	98	Disease	0.9	Disease	88

rs1168200235	I1158M	Disease	Disease	86	Deleterious	87	Deleterious	84	Deleterious	41	Disease	0.88	Disease	65
rs1168200235	I1146M	Disease	Disease	77	Deleterious	77	Deleterious	92	Deleterious	77	Disease	0.12	Disease	77
rs1168200235	I1147M	Disease	Disease	63	Deleterious	77	Deleterious	66	Deleterious	92	Disease	0.59	Disease	82
rs1168200235	I1171M	Neutral	Disease	62	Neutral	77	Neutral	87	Deleterious	84	Neutral	0.88	Neutral	77
rs1168200235	I1149M	Neutral	Disease	43	Neutral	63	Deleterious	77	Neutral	86	Disease	0.12	Neutral	87
rs9804371051	K161R	Neutral	Neutral	88	Neutral	98	Neutral	77	Deleterious	86	Disease	0.59	Disease	77
rs9804371051	K149R	Neutral	Neutral	65	Neutral	45	Neutral	56	Neutral	78	Neutral	0.27	Neutral	77
rs9804371051	K150R	Neutral	Neutral	77	Neutral	62	Neutral	81	Neutral	81	Disease	0.03	Disease	56
rs9804371051	K174R	Neutral	Neutral	82	Neutral	43	Neutral	75	Neutral	66	Neutral	0.16	Neutral	81
rs9804371051	K152R	Disease	Disease	77	Neutral	88	Neutral	77	Neutral	76	Disease	0.59	Disease	75
rs14264860600	S170W	Neutral	Neutral	86	Neutral	65	Neutral	84	Neutral	46	Disease	0.03	Disease	77

rs142648600	S158W	Disease	Disease	7	Deleterious	77	Deleterious	86	Deleterious	77	Disease	0.16	Disease	84
rs142648600	S159W	Disease	Disease	63	Deleterious	87	Deleterious	86	Neutral	87	Disease	0.2	Disease	86
rs142648600	S183W	Disease	Disease	98	Deleterious	77	Deleterious	78	Neutral	45	Disease	0.121	Disease	86
rs142648600	S161W	Disease	Disease	45	Deleterious	77	Deleterious	81	Deleterious	78	Disease	0.16	Disease	78
rs1268994417	T194I	Neutral	Neutral	62	Neutral	56	Neutral	66	Neutral	77	Disease	0.09	Disease	81
rs1268994417	T182I	Disease	Disease	43	Deleterious	81	Deleterious	76	Neutral	98	Disease	0.9	Disease	66
rs1268994417	T183I	Disease	Disease	88	Deleterious	75	Deleterious	46	Neutral	41	Disease	0.88	Disease	76
rs1268994417	T207I	Disease	Disease	65	Deleterious	77	Neutral	77	Deleterious	77	Disease	0.12	Disease	46
rs1268994417	T185I	Neutral	Neutral	77	Deleterious	84	Neutral	87	Deleterious	92	Disease	0.59	Neutral	77
rs1477448265	H196Y	Neutral	Disease	87	Neutral	86	Deleterious	45	Neutral	81	Disease	0.27	Neutral	87
rs147	H	Disea	Disea	7	Dele	86	Del	78	Neu	41	Neu	0.	Dise	4

7448265	184Y	se	se	7	terious		eterious		tral		tral	19	ase	5
rs1477448265	H185Y	Neutral	Disease	77	Neutral	78	Neutral	77	Neutral	63	Disease	0.48	Neutral	78
rs1477448265	H209Y	Disease	Disease	56	Neutral	81	Neutral	98	Neutral	63	Disease	0.34	Neutral	77
rs1477448265	H187Y	Disease	Disease	81	Neutral	66	Neutral	41	Neutral	82	Neutral	0.1	Disease	98
rs1323658124	D210V	Disease	Neutral	75	Neutral	76	Neutral	77	Neutral	84	Disease	0.07	Disease	41
rs1323658124	D198V	Neutral	Neutral	77	Neutral	46	Neutral	92	Neutral	92	Neutral	0.32	Neutral	77
rs1323658124	D199V	Neutral	Disease	84	Deleterious	77	Neutral	81	Deleterious	66	Disease	0.83	Neutral	92
rs1323658124	D223V	Neutral	Neutral	86	Neutral	87	Neutral	41	Neutral	56	Neutral	0.18	Neutral	81
rs1323658	D20	Neutral	Disease	86	Deleterio	45	Deleteri	63	Neutral	82	Disease	0.4	Neutral	41

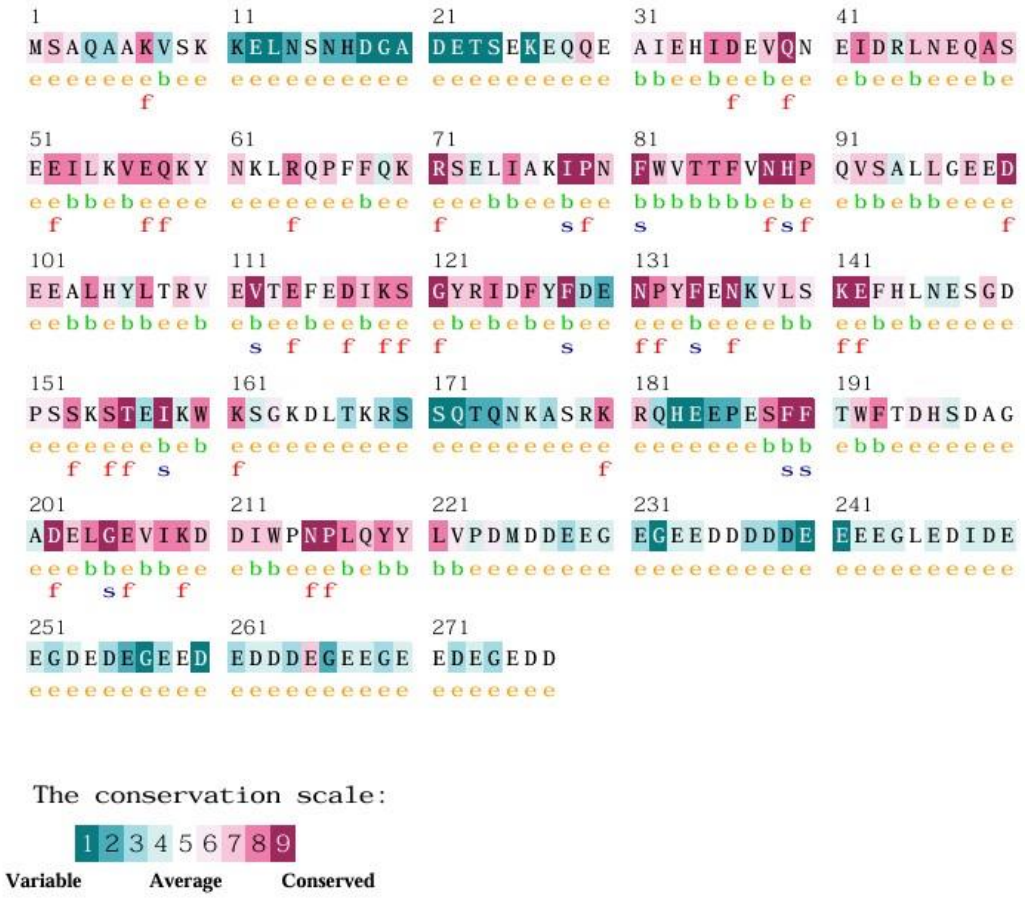
124	1V				us		ous					5		
rs75688585899	M225T	Neutral	Neutral	78	Neutral	78	Neutral	63	Neutral	77	Neutral	0.27	Neutral	63
rs75688585899	M238T	Disease	Disease	81	Neutral	77	Deleterious	82	Neutral	86	Neutral	0.27	Disease	63
rs129447625 7	E242V	Disease	Disease	66	Neutral	98	Neutral	84	Neutral	77	Neutral	0.29	Disease	82
rs129447625 7	E230V	Disease	Disease	76	Deleterious	41	Deleterious	92	Deleterious	63	Disease	0.69	Disease	84
rs129447625 7	E231V	Disease	Disease	46	Deleterious	77	Deleterious	66	Neutral	98	Disease	0.93	Disease	92
rs129447625 7	E255V	Disease	Disease	77	Deleterious	92	Deleterious	56	Deleterious	45	Disease	0.67	Disease	66
rs129447625 7	E233V	Disease	Disease	87	Deleterious	81	Deleterious	82	Deleterious	62	Disease	0.03	Disease	56
rs75356520 06	D276E	Disease	Disease	46	Deleterious	41	Deleterious	77	Deleterious	43	Disease	0.16	Disease	82
rs75356520 06	D264E	Disease	Disease	77	Neutral	63	Neutral	86	Neutral	88	Disease	0.2	Disease	77
rs75356520 06	D265E	Disease	Disease	87	Deleterious	63	Deleterious	77	Deleterious	65	Disease	0.121	Disease	86

rs753565206	D289E	Neutral	Disease	45	Neutral	82	Neutral	63	Neutral	77	Neutral	0.16	Neutral	77
rs753565206	D267E	Disease	Disease	78	Deleterious	84	Deleterious	98	Deleterious	82	Neutral	0.09	Disease	63
rs754819669	D277V	Disease	Disease	77	Deleterious	92	Deleterious	45	Deleterious	77	Disease	0.9	Disease	98
rs754819669	D265V	Disease	Disease	98	Deleterious	66	Deleterious	62	Deleterious	86	Disease	0.88	Disease	45
rs754819669	D266V	Disease	Disease	41	Deleterious	56	Deleterious	43	Neutral	77	Disease	0.12	Disease	62
rs754819669	D290V	Disease	Disease	77	Deleterious	82	Deleterious	88	Deleterious	63	Disease	0.59	Disease	43
rs754819669	D268V	Disease	Disease	92	Deleterious	77	Deleterious	65	Deleterious	98	Disease	0.27	Disease	88

4.4. Effect of nsSNPs on protein stability and conservation of amino acids

Protein stability of nsSNPs was determined by utilizing Mu-Pro and I-mutant online servers. The results showed that most of nsSNPs has decreased protein stability in all tested servers (Table 3). Mu-Pro analysis revealed that all SNPs had decreased protein stability. While I-mutant analysis showed 2 SNPs LP9 and Q12P having increase protein stability.

From Consurf, SET protein has a variety of residues, some exposed, some buried, and even some predicted to be functional or structural. The conservation scale ranges from 1 (variable) to 9 (conserved). The residues marked as 'e' are exposed, 'b' are buried, 'f' are predicted functional, and 's' are predicted as structural. Based on this information, it seems like the SET protein has a mix of conserved and variable regions, with some residues playing important functional or structural roles.The Amino acids having numbers 2, 7, 39, 52, 57-58, 64, 66, 70, 72, 79, 80, 88, 90, 98, 100 to 102, 109, 111, 114, 117, 119 to 121, 123, 131 to 133, 136, 141 to 142, 149, 153, 155-156, 161, 180, 181, 202, 206, 209, 215-216, 218, 265 are highly conserved and exposed, at 8, 31, 32, 74, 83, 87, 95, 103, 106, 110, 115, 140, 222, are neutral, while amino acids at numbers 1, 3 to 6, 40, 44, 51, 55, 60, 62 to 63, 65, 67, 69, 73, 76 to 77, 91, 94, 97, 99, 105, 108, 113, 116, 125, 127, 129 to 130, 135, 137 to 138, 144, 146 to 148, 150 to 152, 154, 157, 159, 162 to 179, 182 to 187, 191, 194 to 201, 214, 266 to 277 are exposed neutral.



e - An exposed residue according to the neural network algorithm.

b - A buried residue according to the neural network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

Figure 3. Prediction of evolutionary conservation of amino acids.

Table 3. Prediction of effect of nsSNPs on protein stability, amino acid conservation.

		protein stability				Sequence Conservation
		Mu Pro		**I Mutant**		**CONSURF**
rsIDs	**A.A**	**PREDICTION**	**DETAL DELTA**	**STABILITY**	**RI**	**Conservation Score**
rs1228985010	M1I	DECREASE stability	-0.845	Decrease	3	6,b
rs1228985010	M1I	DECREASE stability	-1.336	Decrease	6	6,b
rs1287905948	P3R	DECREASE stability	-0.9107	Decrease	1	4,b
rs1287905948	P3R	DECREASE stability	-0.331	Decrease	3	4,b
rs1180724092	L9P	DECREASE stability	-1.722	Increase	4	4,b
rs1180724092	L9P	DECREASE stability	-0.522	Decrease	9	9,e,f
rs992067058	P10L	DECREASE stability	-0.8003	Decrease	10	9,e,f
rs971609222	Q12P	DECREASE stability	-0.341	Increase	1	7,e
rs971609222	Q12P	DECREASE stability	-0.267	Decrease	1	7,e
rs923629041	K15N	DECREASE stability	-0.868	Decrease	9	4,e
rs923629041	K15N	DECREASE stability	-0.611	Decrease	9	9,b,s
rs1361492173	P16R	DECREASE stability	-0.614	Decrease	9	9,b,s
rs1361492173	P16L	DECREASE	-1.892	Decrease	7	5,b

		stability				
rs900458307	P19L	DECREASE stability	-0.487	Decrease	8	9,b,s
rs138894709	L22P	DECREASE stability	-0.568	Decrease	9	9,b,s
rs1208440934	E26K	DECREASE stability	-0.884	Decrease	9	7,e
rs1442641412	S28L	DECREASE stability	-1.703	Decrease	8	8,e,f
rs1181003970	A31V	DECREASE stability	-1.097	Decrease	9	8,e
rs1268249403	L6R	DECREASE stability	-0.498	Decrease	6	8,e,f
rs1319521099	L7R	DECREASE stability	-1.065	Decrease	2	4,b
rs1218933656	P8H	DECREASE stability	-0.909	Decrease	5	5,b
rs1218933656	P8L	DECREASE stability	-1.714	Decrease	8	7,b
rs1141138	P4L	DECREASE stability	-1.58	Decrease	6	3,b
rs1437350212	A6D	DECREASE stability	-0.436	Decrease	3	7,e
rs11542565	L13F	DECREASE stability	-0.739	Decrease	9	7,e
rs775693359	M1R	DECREASE stability	-0.763	Decrease	3	7,e
rs1024856362	E13K	DECREASE stability	-1.2	Decrease	1	3,e
rs1024856362	E13K	DECREASE stability	-0.874	Decrease	7	8,e,f
rs1276443619	Q29R	DECREASE stability	-0.991	Decrease	5	9,e,f

rs1276443619	Q17R	DECREASE stability	-0.812	Decrease	2	5,e
rs1276443619	Q18R	DECREASE stability	-0.988	Decrease	4	7,b
rs1276443619	Q42R	DECREASE stability	-0.988	Decrease	6	7,b
rs1276443619	Q20R	DECREASE stability	-0.817	Decrease	8	7,b
rs866872429	H89N	DECREASE stability	-0.49536	Decrease	7	6,b
rs866872429	H89Y	DECREASE stability	-2.0804	Decrease	5	7,b
rs866872429	H77N	DECREASE stability	-1.627	Decrease	9	8,b
rs866872429	H77Y	DECREASE stability	-0.262	Decrease	3	9,e,f
rs866872429	H78N	DECREASE stability	-0.657	Decrease	6	5,e
rs866872429	H78Y	DECREASE stability	-0.845	Decrease	7	3,e
rs866872429	H102N	DECREASE stability	-1.336	Decrease	2	8,e
rs866872429	H102Y	DECREASE stability	-0.9107	Decrease	4	9,b,s
rs866872429	H80N	DECREASE stability	-0.331	Decrease	5	8,b
rs866872429	H80Y	DECREASE stability	-1.722	Decrease	8	8,b
rs483352707	A94P	DECREASE stability	-0.522	Decrease	6	4,e
rs483352707	A82P	DECREASE stability	-0.8003	Decrease	9	6,e
rs483352707	A83P	DECREASE	-0.341	Decrease	5	6,e

		stability				
rs483352707	A107P	DECREASE stability	-0.267	Decrease	2	6,e
rs483352707	A85P	DECREASE stability	-0.868	Decrease	7	6,e
rs1408742550	A103S	DECREASE stability	-0.611	Decrease	2	7,e
rs1408742550	A91S	DECREASE stability	-0.614	Decrease	6	7,e
rs1408742550	A92S	DECREASE stability	-1.892	Decrease	3	7,e
rs1408742550	A116S	DECREASE stability	-0.487	Decrease	6	5,e
rs1408742550	A94S	DECREASE stability	-0.568	Decrease	1	5,e
rs913299138	T108A	DECREASE stability	-0.884	Decrease	3	9,b,s
rs913299138	T96A	DECREASE stability	-1.703	Decrease	4	9,e,f
rs913299138	T97A	DECREASE stability	-1.097	Decrease	9	9,e,f
rs913299138	T121A	DECREASE stability	-0.498	Decrease	10	9,e,f
rs913299138	T99A	DECREASE stability	-1.065	Decrease	1	9,e,f
rs1196624947	D125G	DECREASE stability	-0.909	Decrease	1	5,e
rs1196624947	D113G	DECREASE stability	-1.714	Decrease	9	5,e
rs1196624947	D114G	DECREASE stability	-1.58	Decrease	9	8,e
rs1196624947	D138G	DECREASE stability	-0.436	Decrease	9	8,b,s

rs1196624947	D116G	DECREASE stability	-0.739	Decrease	7	6,e
rs1222712623	E130G	DECREASE stability	-0.763	Decrease	8	6,e
rs1222712623	E118G	DECREASE stability	-1.2	Decrease	9	4,e
rs1222712623	E119G	DECREASE stability	-0.874	Decrease	9	5,e
rs1222712623	E143G	DECREASE stability	-0.991	Decrease	8	5,e
rs1222712623	E121G	DECREASE stability	-0.812	Decrease	9	8,e
rs374932364	Y133H	DECREASE stability	-0.988	Decrease	6	4,b
rs374932364	Y121H	DECREASE stability	-0.988	Decrease	2	8,b
rs374932364	Y122H	DECREASE stability	-0.817	Decrease	5	9,e,f
rs374932364	Y146H	DECREASE stability	-0.49536	Decrease	8	9,e,f
rs374932364	Y124H	DECREASE stability	-0.991	Decrease	6	9,e,f
rs765603755	E123D	DECREASE stability	-0.812	Decrease	3	3,e
rs765603755	E124D	DECREASE stability	-0.988	Decrease	6	5,e
rs765603755	E148D	DECREASE stability	-0.988	Decrease	1	5,e
rs765603755	E126D	DECREASE stability	-0.817	Decrease	3	4,e
rs1429567543	K141R	DECREASE stability	-0.49536	Decrease	4	5,e
rs1429567543	K129R	DECREASE	-2.0804	Decrease	9	5,e

		stability				
rs1429567543	K130R	DECREASE stability	-1.627	Decrease	10	9,b,s
rs1429567543	K154R	DECREASE stability	-0.262	Decrease	1	8,b
rs1429567543	K132R	DECREASE stability	-0.657	Decrease	1	8,b
rs1013985707	F143L	DECREASE stability	-0.845	Decrease	9	3,b
rs1013985707	F131L	DECREASE stability	-1.336	Decrease	9	9,e,f
rs1013985707	F132L	DECREASE stability	-0.9107	Decrease	9	3,b
rs1013985707	F156L	DECREASE stability	-0.331	Decrease	7	7,b
rs1013985707	F134L	DECREASE stability	-1.722	Decrease	8	9,e,f
rs1367036404	S148R	DECREASE stability	-0.522	Decrease	9	8,e,f
rs1367036404	S136R	DECREASE stability	-0.8003	Decrease	9	8,e,f
rs1367036404	S137R	DECREASE stability	-0.341	Decrease	8	3,b
rs1367036404	S161R	DECREASE stability	-0.267	Decrease	9	9,b,s
rs1367036404	S139R	DECREASE stability	-0.868	Decrease	6	9,b,s
rs1429859567	D150G	DECREASE stability	-0.611	Decrease	2	9,e,f
rs1429859567	D138G	DECREASE stability	-0.614	Decrease	5	9,e,f
rs1429859567	D139G	DECREASE stability	-1.892	Decrease	8	9,e,f

rs1429859567	D163G	DECREASE stability	-0.487	Decrease	6	9,e,f
rs1429859567	D141G	DECREASE stability	-0.568	Decrease	3	5,b
rs764645296	S153L	DECREASE stability	-0.884	Decrease	9	5,b
rs764645296	S141L	DECREASE stability	-1.703	Decrease	3	9,e,f
rs764645296	S142L	DECREASE stability	-1.097	Decrease	1	9,e,f
rs764645296	S166L	DECREASE stability	-0.498	Decrease	7	5,e
rs764645296	S144L	DECREASE stability	-1.065	Decrease	5	9,b,s
rs1168200235	I158M	DECREASE stability	-0.909	Decrease	2	8,e
rs1168200235	I146M	DECREASE stability	-1.714	Decrease	4	9,e,f
rs1168200235	I147M	DECREASE stability	-1.58	Decrease	6	9,b,s
rs1168200235	I171M	DECREASE stability	-0.812	Decrease	9	7,e
rs1168200235	I149M	DECREASE stability	-0.988	Decrease	10	7,e
rs980437151	K161R	DECREASE stability	-0.988	Decrease	1	8,e,f
rs980437151	K149R	DECREASE stability	-0.817	Decrease	1	8,b
rs980437151	K150R	DECREASE stability	-0.49536	Decrease	9	3,b
rs980437151	K174R	DECREASE stability	-2.0804	Decrease	9	5,e
rs980437151	K152R	DECREASE	-1.627	Decrease	9	8,b

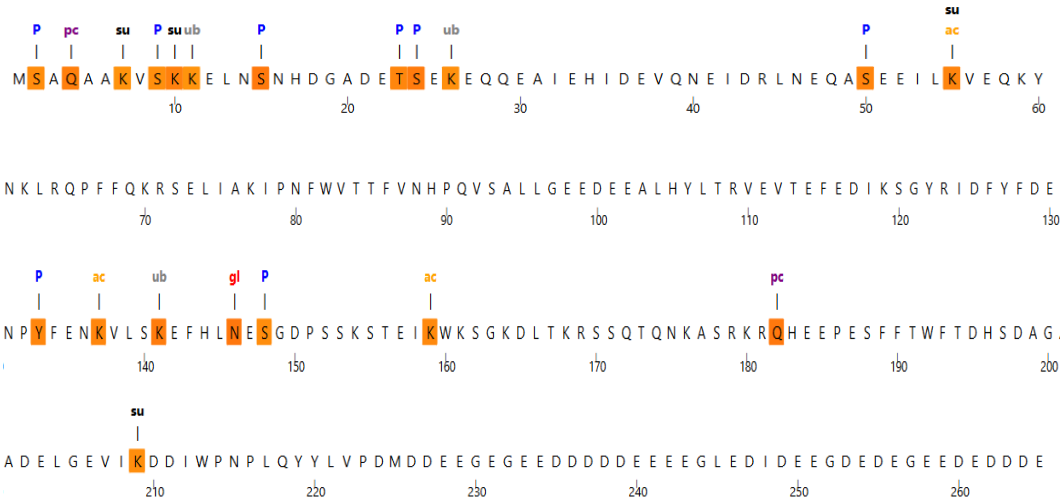
		stability				
rs142648600	S170W	DECREASE stability	-0.262	Decrease	7	8,b
rs142648600	S158W	DECREASE stability	-0.657	Decrease	8	9,e,f
rs142648600	S159W	DECREASE stability	-0.845	Decrease	9	9,e,f
rs142648600	S183W	DECREASE stability	-1.336	Decrease	9	8,b
rs142648600	S161W	DECREASE stability	-0.9107	Decrease	8	8,b,s
rs1268994417	T194I	DECREASE stability	-0.331	Decrease	9	5,e
rs1268994417	T182I	DECREASE stability	-1.722	Decrease	6	5,e
rs1268994417	T183I	DECREASE stability	-0.522	Decrease	2	8,b
rs1268994417	T207I	DECREASE stability	-0.8003	Decrease	5	7,e
rs1268994417	T185I	DECREASE stability	-0.341	Decrease	8	5,e
rs1477448265	H196Y	DECREASE stability	-0.267	Decrease	6	5,e
rs1477448265	H184Y	DECREASE stability	-0.868	Decrease	3	3,b
rs1477448265	H185Y	DECREASE stability	-0.611	Decrease	9	6,b
rs1477448265	H209Y	DECREASE stability	-0.614	Decrease	3	5,b
rs1477448265	H187Y	DECREASE stability	-1.892	Decrease	1	6,e
rs1323658124	D210V	DECREASE stability	-0.487	Decrease	7	3,e

rs1323658124	D198V	DECREASE stability	-0.568	Decrease	5	7,b
rs1323658124	D199V	DECREASE stability	-0.884	Decrease	2	9,e,f
rs1323658124	D223V	DECREASE stability	-1.703	Decrease	4	7,b
rs1323658124	D201V	DECREASE stability	-1.097	Decrease	6	7,b
rs756885899	M225T	DECREASE stability	-0.498	Decrease	8	3,e
rs756885899	M238T	DECREASE stability	-1.065	Decrease	7	5,e
rs1294476257	E242V	DECREASE stability	-0.909	Decrease	5	6,b
rs1294476257	E230V	DECREASE stability	-1.714	Decrease	9	9,b,s
rs1294476257	E231V	DECREASE stability	-1.58	Decrease	3	5,b
rs1294476257	E255V	DECREASE stability	-0.436	Decrease	6	5,b
rs1294476257	E233V	DECREASE stability	-0.739	Decrease	7	8,b,s
rs753565206	D276E	DECREASE stability	-0.763	Decrease	2	8,e,f
rs753565206	D264E	DECREASE stability	-1.2	Decrease	4	5,b
rs753565206	D265E	DECREASE stability	-0.874	Decrease	5	5,b
rs753565206	D289E	DECREASE stability	-0.991	Decrease	8	7,b
rs753565206	D267E	DECREASE stability	-0.812	Decrease	6	3,b
rs754819669	D277V	DECREASE	-0.988	Decrease	9	3,b

		stability				
rs754819669	D265V	DECREASE stability	-0.988	Decrease	5	9,e,f
rs754819669	D266V	DECREASE stability	-0.817	Decrease	2	8,e,f
rs754819669	D290V	DECREASE stability	-1.2	Decrease	7	9,e,f
rs754819669	D268V	DECREASE stability	-0.874	Decrease	2	9,e,f

4.5. Post translational modifications (PTMs)

According to MustiDeep, there is Phosphorylation at amino acid number 2, 9, 15, 23, 24, 50, 133 and 148; Pyrrolidone carboxylic acid at number 4 and 182 and Glycosylation at number 146; Ubiquitination at number 11, 26 and 141; SUMOylation at number 7, 10, 55 and 209; Acetylation at number 55, 137 and159 (Figure 4).



- Phosphorylation: **P**
- Glycosylation: **gl**
- Ubiquitination: **ub**
- SUMOylation: **su**
- Acetylation: **ac**
- Methylation: **me**
- Pyrrolidone carboxylic acid: **pc**
- Palmitoylation: **pa**
- Hydroxylation: **Hy**

- Zinc: **z**
- Copper: **c**
- Ferrous: **fe**

Figure 4: Post translational modifications (PTMs) in SET.

4.6. Protein's secondary structure

According to SOPMA, the secondary structure of SET consists of Alpha helix (Hh) 121 is (43.68%), extended standard (Ee) 28 is (10.11%), beta turn (Tt) 3 is (1.08%) and random coil (Cc) 125 is (45.13%) (Figure 5).

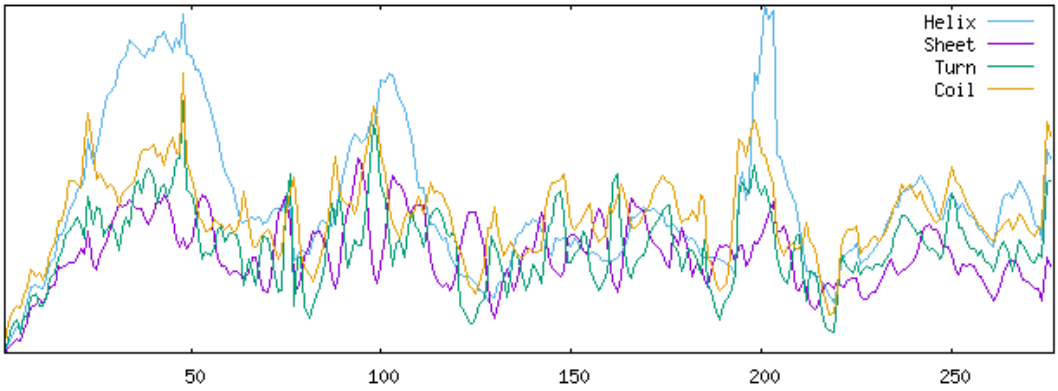


Figure 5: Secondary structure of SET protein.

Protein to protein interaction (PPI)

SET has been shown strong protein to protein interaction with its own family as well as Granzyme A, Nucleoside diphosphate kinase A, DNA-(apurinic or apyrimidinic site) lyase, Acidic leucine-rich nuclear phosphoprotein 32 family member A, Shugoshin 2, Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform, Exportin-1 and with Nuclear pore complex protein Nup214 having score almost 0.9. (Fig No.1).

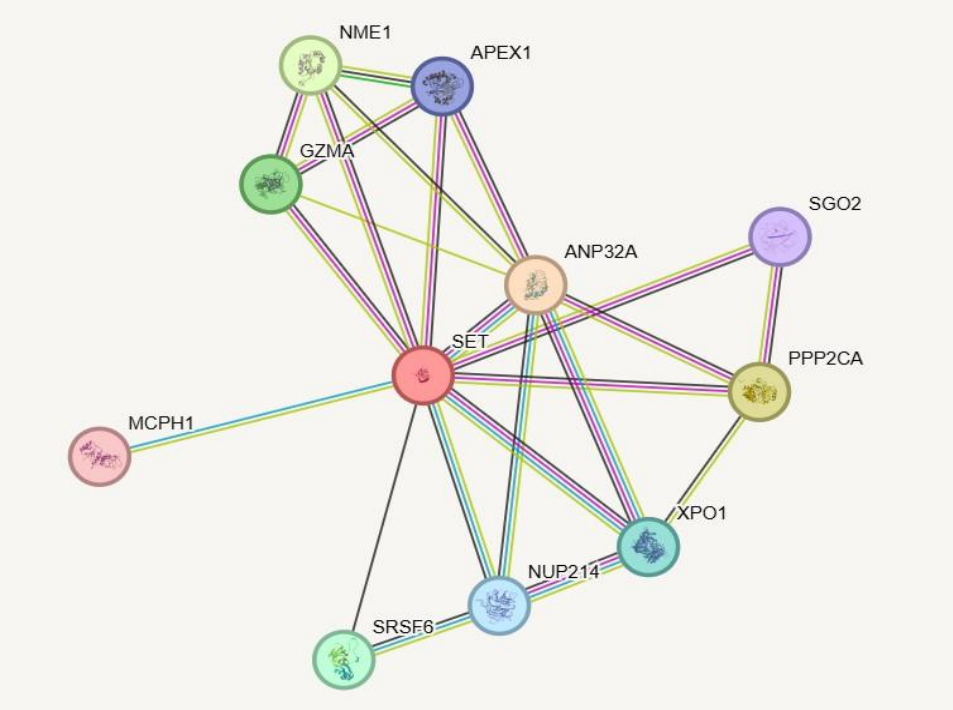


Figure 6: Protein to protein interaction (PPI) of SET.

DISCUSSION

The study shows that the conserved Su (var)3-9, Ezh2, Trithorax (SET) domain in this family of proteins is the same to that of S. cerevisiae SET1. SET1 is an H3K4 methyltransferase and is accountable for all three methylation conditions (mono-, di-, and tri-) of H3K4. The SET/MLL has three pairs of family members in mammals: SETd1A/SETd1B, MLL1/MLL2, and MLL3/MLL4. The Cosgrove group checked all bacterial core SET/MLL complex (with WRAD) in vitro and detected that MLL1/2 are mainly mono-and di-methyltransferases and MLL3/4 act majorly as mono-methyltransferases, although SETd1A/B may catalyze each three methylation conditions the same to the yeast ortholog. SET protein causes human leukemia. Objective of this study was to identify non-synonymous SNPs of SET and their role in causing leukemia utilizing computational analysis. We've used different tools to identify pathogenic diseased synonymous and non-synonymous nsSNPs in SET. This study shows that there were total 7577 SNPs in the protein SET. In which by SIFT server it was examined that 123 SNPs were deleterious with a score of less than 0.5. While 217 SNPs from SIFT were tolerable. And 53 SNPs are probably damaging, 74 possibly damaging and 355 are benign in the polyphen server.

In SNAP2, 111 SNPs were predicted as diseased or effected SNPs. There were 99, 108, 115 SNPs predicted as diseased by SNP&GO, PhD-SNP, and Meta-SNP tools, respectively.

There were 94, 81 and 69 SNPs predicted as deleterious by PredictSNP, MAPP, and SNAP, respectively. 93 SNPs predicted as disordered by Predict Protein and 109 SNPs were predicted by PANTHER. The results showed that almost all the nsSNPs has decreased protein stability in all tested servers (Table 3) Mu-Pro and I-mutant. While I-mutant analysis showed 2 SNPs LP9 and Q12P having increase protein stability. According to MustiDeep, there is Phosphorylation at amino acid number 2, 9, 15, 23, 24, 50, 133 and 148; Pyrrolidone carboxylic acid at number 4 and 182 and Glycosylation at number 146; Ubiquitination at number 11, 26 and 141; SUMOylation at number 7, 10, 55 and 209; Acetylation at number 55, 137 and159.

Our study showed that SET in the protein protein interaction (PPI) has been shown strong protein to protein interaction with its own family as well as Granzyme A, Nucleoside diphosphate kinase A, DNA-(apurinic or apyrimidinic site) lyase, Acidic leucine-rich nuclear phosphoprotein 32 family member A, Shugoshin 2, Serine/threonineprotein phosphatase 2A catalytic subunit alpha isoform, Exportin-1 and with Nuclear pore complex protein Nup214 having score almost 0.9.

The SET1/MLL (mixed lineage leukemia) family of methyltransferases is evolutionarily conserved from yeast to mammals. These enzymes facilitate the mono-, di-, or tri-methylation of histone 3 at lysine 4 (H3K4) on chromatin through their Suppressor of variegation 3–9, Enhancer of Zeste, Trithorax (SET) domain. In humans, six H3K4 methyltransferases (HMTs) are essential: MLL1 (MLL/KMT2A), MLL2 (KMT2B), MLL3 (KMT2C), MLL4 (KMT2D), SETD1A (KMT2F), and SETD1B (KMT2G). In contrast, yeast possesses only one HMT, the Set1/COMPASS complex, which is capable of catalyzing all three methylation states (Sugeedha *et al.,* 2021). The Mixed-Lineage Leukemia gene (MLL/MLL1/KMT2A) encodes a member of the KMT2 family of methyltransferase enzymes, characterized by several structural domains, including a C-terminal SET domain responsible for the methylation of histone H3 at lysine 4 (H3K4). Rearrangements and translocations of the MLL gene, which typically impact one allele, account for approximately 70% of infant leukemias and 5–10% of acute myeloid leukemia (AML) cases in children and adults. Frequently, the MLL gene rearrangement associated with leukemia results in the formation of the MLL fusion oncoprotein, which loses the C-terminal SET domain of MLL and acquires a partial sequence from its fusion partner, such as AF4, AF9, AF10, or ENL, thereby recruiting DOT1L-associated transcription elongation complexes. (Lu *et al.,* 2017). The SET1/MLL complexes play a significant role in the regulation of cell survival, proliferation, and apoptosis, and their subunits are critically implicated in cancer

development. Genetic alterations, including mutations and deletions of the catalytic subunits, are frequently linked to various human cancers; however, their functional implications can be intricate and context-dependent. Chromosomal translocations involving MLL1 are known to drive mixed lineage leukemias, while the wild-type forms of MLL1, MLL2, and SET1A contribute to the maintenance of certain cancer types. MLL3 and MLL4 are recognized as established tumor suppressors, although MLL4 is also essential for the support of MLL-rearranged leukemia. Recurrent mutations in SET1B have been identified in primary hepatic neuroendocrine tumors, with one specific mutation enhancing cell proliferation, migration, and invasion. Conversely, mutations or deletions of the core subunits are infrequently observed in cancer. Instead, there is a notable prevalence of amplification and increased expression of these core subunits in human cancers, with all four core subunits demonstrating significant roles in promoting tumorigenesis, as will be elaborated in the subsequent sections (Jiang & H *et al.*,2020). SET-domain-containing methyltransferases serve as the catalytic subunits within the MLL/SET/COMPASS complex (Mozzetta et al. 2015). This domain exhibits a high degree of conservation from yeast to humans (Figure 1.12A). The SET domain derives its name from the Drosophila proteins Suppressor of variegation 3-9 (Su(var)3-9), Enhancer of zester (E(z)), and Trithorax (Trx). It is characterized by a length of approximately 130 amino acids and possesses catalytic activity directed towards the Ɛ-amino group of lysine residues (Herz et al. 2013). This section will provide an overview of the structure of the SET domain, its role in facilitating methyltransferase activity, and additional motifs that may be located adjacent to the SET domain (Samsudin *et al.,* 2017). Chronic myelogenous leukemia (CML) ranks among the most prevalent types of leukemia. According to the American Cancer Society, there were 1,280 reported deaths and an estimated 9,280 new cases of CML expected to be diagnosed in the United States in 2024. CML, like other forms of leukemia, primarily impacts hematopoietic stem cells (HSCs). The disease is characterized by specific translocation breakpoints that lead to the fusion of two genes: the breakpoint cluster region (BCR) located on chromosome 22 and the Abelson murine leukemia viral oncogene homolog 1 (ABL1) found on chromosome 9. The Philadelphia chromosome plays a crucial role in the pathophysiology of CML, as it activates tyrosine kinase (TK) through the BCR-ABL1 oncoprotein. The advent of tyrosine kinase inhibitors (TKIs) has revolutionized the treatment landscape for CML, earning widespread recognition. (Das *et al.,* 2024). Chronic myeloid leukemia (CML) is a form of leukemia that begins with a chronic phase (CP), marked by a significant proliferation of myeloid precursors and mature cells. This is followed by a later acute phase known as "blast crisis" (BC), which bears

similarities to acute myeloid leukemia (AML). The pathophysiology of CML is defined by the presence of the reciprocal translocation t(9;22)_(q34;q11), which results in the formation of a BCR-ABL fusion protein that exhibits abnormal and unregulated activity. The introduction of imatinib, a BCR/ABL kinase inhibitor, represented a substantial advancement in the treatment of CML. Nevertheless, the efficacy of imatinib, along with second and third-generation tyrosine kinase inhibitors, is frequently insufficient, as resistant CML stem cells and residual disease remain in numerous patients. Genomic sequencing of individuals with CML has revealed additional genetic alterations, including mutations in tumor suppressor genes such as RB1, TP53, and CDKN2A. Furthermore, it has been noted that BCR-ABL kinase mutations are detected in only about 50% of patients who exhibit poor responses and disease progression. Consequently, further advancements in our comprehension of CML biology are necessary (Pippa *et al., 2020).* Artificial intelligence has transformed the fields of disease prognosis, diagnosis, and management, particularly in the context of Chronic Myeloid Leukemia (CML). This transformation is largely attributed to the emergence of guideline-based clinical systems, known as expert systems, as well as the application of machine learning (ML) and deep learning (DL) techniques in data analysis and clinical imaging. Machine learning algorithms play a crucial role in the early detection of CML by analyzing clinical and laboratory data, while deep learning methods, including Convolutional Neural Networks (CNNs), enhance the automation of CML classification and diagnosis through the interpretation of medical images. These technological advancements promote timely detection, swift intervention, and better patient outcomes. In contrast to conventional statistical and experimental prediction techniques, artificial intelligence provides significant, practical, and non-invasive analytical capabilities, particularly in complex and uncertain scenarios such as forecasting cancer prognosis and survival rates (Ram *et al.,* 2024).

The SET domain comprises of 12 β strands that organized into 5 partly interwoven sheets. I–IV Sheets are antiparallel, while sheet V is parallel. I and V Sheets are interwoven, in that β12 contributes in both sheets. Sheet I comprises strands β1, β2, and β12; sheet II strands β3 and β11; sheet III strands β4, β10, and β9; sheet IV strands β5, β7, and β6; also sheet V strands β8 and β12. A knot-like structure is present where the C terminus of the SET domain occurs beneath the β8-β9 connection. The all-β SET domain suffers no similarity to the classical α/β AdoMet binding fold of other methyltransferase (Trievel *et al.,* 2002).

The SET domain is interrupted among β5 and β6 via a Rubisco LSMT-specific domain (we mention to this as the SET-inserted domain, or iSET domain), comprising of 115 remnants also six α helices. The inserted region makes a single folding unit at once with the N- and C-terminal helices of the N-terminal lobes, α1 and α8, which we will mention to as the “nSET” and “cSET” regions, appropriately. The SET domain and cSET iSET and nSET also regions of the Nterminal lobe significantly link with each other. Inclusively, these regions make a domain which is shaped like a catcher's mitt also grasps the SET domain on three sides, covering a lot of SET domain sheets III and IV (Trievel *et al.*, 2002).

The designation "chronic" indicates a more gradual development of cancer compared to the acute varieties of leukemia, particularly in the context of chronic myelogenous leukemia. The term "myelogenous" pertains to the specific type of cells that are impacted. Chronic myelogenous leukemia (CML) is a rare form of bone marrow cancer, which is the spongy tissue found within bones responsible for the production of blood cells. In CML, there is an increase in the number of white blood cells present in the bloodstream. This condition may also be known as chronic, granulocytic, or myeloid leukemia. It predominantly affects older adults, although it can occur at any age, including in younger individuals.Chronic myeloid leukaemia (CML) is a slow-progressing malignancy of the bone marrow. The normal function of bone marrow includes the production of red blood cells (erythrocytes) that transport oxygen, white blood cells (leucocytes) that defend against infections, and platelets (thrombocytes) that facilitate blood clotting. In CML, there is an overproduction of white blood cells, which initially function adequately. However, as the disease advances, the bone marrow accumulates immature white blood cells known as myeloblasts (Pandey *et al.*, 2022). The proliferation of myeloblasts disrupts the production of other blood cells, resulting in a deficiency of red blood cells (anaemia) and platelets. This form of leukaemia typically manifests in individuals over the age of 60. Common symptoms include significant fatigue, fever, and weight loss. Patients may also experience splenomegaly, leading to a sensation of fullness in the abdomen and a decreased appetite. Approximately half of those diagnosed with chronic leukaemia are identified incidentally during blood tests conducted for other reasons, often presenting without initial symptoms. The disease progresses through three distinct phases: the chronic phase, the accelerated phase, and the blast crisis phase. In the chronic phase, there is a predominance of mature white blood cells, with myeloblasts constituting less than 10 percent of the total blood cells. Symptoms during this phase are typically mild or absent and may gradually worsen over a period of months to years. The

accelerated phase is characterized by an increase in myeloblasts, which make up 10 to 29 percent of blood cells, and symptoms become more pronounced. This phase usually develops over a span of 4 to 6 months, although some patients may progress more rapidly. In the blast crisis phase, myeloblasts account for 30 percent or more of blood or bone marrow cells, and symptoms such as severe splenomegaly, bone pain, and significant weight loss become particularly acute. This phase poses a heightened risk of life-threatening infections (Pandey *et al.*, 2022).

Human SETD1A is a nuclear-localized protein composed of 1707 amino acids, characterized by four evolutionarily conserved domains that it shares with yeast Set1: the RRM, NSET, SET, and post-SET domains. The RRM (RNA recognition motif) domain is situated in the N-terminal region, while the NSET, SET, and post-SET domains are located in the C-terminal region. The methyltransferase activity of SETD1A is dependent on the presence of the NSET, SET, and post-SET domains. The post-SET domain, which is a conserved cysteine-rich motif that follows the SET domain, has an unclear function in the context of SETD1A. The SET domain is enzymatically active and is a feature common to all six SET1/MLL family proteins, as well as other histone methyltransferases such as SUV39 and EZH1/2. The NSET and SET domains engage with a conserved complex known as COMPASS (complex of proteins associated with Set1), demonstrating adequate enzymatic activity. In mammals, the WRAD complex, consisting of four subunits—WDR5, RbBPP5, ASH2L, and DPY30—is present in all SET1/MLL family proteins (Kikuchi *et al.*, 2023).

Allogeneic hematopoietic cell transplantation (HCT) represents the most effective post-remission treatment for individuals diagnosed with acute myeloid leukemia (AML). It is particularly prevalent among younger patients exhibiting intermediate-risk or adverse-risk cytogenetic profiles. The decision to proceed with transplantation primarily hinges on factors such as cytogenetic and molecular risk stratification, patient age, comorbid conditions, therapeutic response, and the availability of an appropriate donor. Notably, more than 25% of AML cases are classified as secondary (s-AML), which can occur following prior chemotherapy and/or radiotherapy (referred to as therapy-related AML or t-AML) or may develop from a preceding myeloid disorder (AHD-AML), including myelodysplastic syndrome (MDS) or myeloproliferative neoplasm (MPN). s-AML has been recognized as an independent factor associated with unfavorable outcomes, particularly in cases of non-MDS s-AML. However, it is not currently incorporated into the existing risk classification systems that inform HCT decisions, such as the European LeukemiaNet (ELN) criteria. Despite this,

patients with s-AML can still be evaluated for HCT, which has demonstrated efficacy in treating this population. The frequency of HCT utilization among these patients and its impact on their survival remains to be thoroughly investigated (Nilsson *et al.*, 2019).

Nuclear oncogene SET has been recognized as inhibitor 2 of PP2A (I2PP2A) and is significantly associated with leukaemogenesis. In patients with chronic myeloid leukaemia (CML), the expression of the PP2A inhibitor SET leads to the inhibition of PP2A's phosphatase activity. SET is found to be overexpressed in various solid tumors, CML, and possibly other forms of leukaemia. Studies have indicated that the induced downregulation of SET can restore PP2A activity. The reactivation of PP2A in these cells has been demonstrated to facilitate the dephosphorylation of critical regulators that govern cell survival and proliferation, while also reducing the activity and degradation of the tyrosine kinase inhibitor BCR-ABL. Consequently, the activation of PP2A results in growth suppression, increased apoptosis, restored differentiation, and a decrease in in vivo leukaemogenesis of BCR-ABL positive cells. SETBP1 is associated with a decrease in PP2A activity, which provides a growth advantage to hematopoietic progenitors. This protein is overexpressed in cases of acute myeloid leukemia (AML) due to the (12;18) translocation involving the ETV6 gene. Additionally, mutated forms of SETBP1 have been identified in colorectal carcinomas, while it is often lost in the prevalent chromosomal abnormalities at 18q21. Notably, over 60% of the aberrations at 18q21 are deletions, which are linked to various malignancies, including malignant lymphoma, acute leukemia, pancreatic carcinoma, and colorectal cancer (Grech *et al.*, 2016; Bilal *et al.*, 2024a,b,c).

The SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1) is classified within the family of Suppressor of Variegation 3–9 (SUV39) proteins, and it is a member of the group of SET domain-containing protein lysine methyltransferases (PKMTs) that play a significant role in epigenetic regulation. SETDB1 is distinguished by its highly conserved bifurcated SET domain, which features an intervening sequence of approximately 150 amino acids. The SET domain was initially identified in the Suppressor of Variegation 3–9 (SUV39), Enhancer of Zeste (EZ), and Trithorax genes of Drosophila species. Further investigations have revealed the presence of SET domains in over 40 species, including Saccharomyces cerevisiae (SET1 gene), Schizosaccharomyces pombe (Clr4+ gene), and humans (SETDB1 along with other SET domain-containing HKMT genes). Moreover, SETDB1 orthologs have been examined in various other species, including Mus musculus, Rattus norvegicus, Danio rerio, Bos taurus, and Macaca mulatta, among others (Markouli *et al.*, 2021; Afzal *et al.*, 2024).

MLL1 possesses a conserved trans-activation domain (TAD) that engages with the CREB-binding protein (CBP). A structural analysis of a ternary complex involving the activation domain of the transcription factor c-Myb, the MLL1 TAD, and the kinase-inducible domain-interacting domain (KIX) of CBP has been documented. The binding of the MLL1 TAD enhances the stability of the interaction between c-Myb and CBP by inducing conformational alterations in the disordered regions of the KIX domain. Furthermore, the association of MLL1 TAD promotes the interaction between phosphorylated CREB and CBP. The transactivation mediated by MLL1 TAD is significantly inhibited by the coexpression of adenovirus E1A12S, a competitive antagonist of CBP, or by MLL1 TAD mutants that lack the ability to bind CBP. Notably, CBP appears to influence the recruitment of MLL1 to either E2F1-mediated early pro-survival genes or late pro-apoptotic genes in a mouse model of hepatocellular carcinoma. The interaction between MLL1 and CBP is evolutionarily conserved; in Drosophila, TRX forms a stable complex with dCBP, which collaborates with TRX in the regulation of homeotic genes. Likewise, p300/CBP interacts with the mammalian SET1 complex. (Sha *et al.,* 2020; Umar *et al.,* 2025). It is concluded that the SET protein is exhibited in divergent tissues like as lung, heart, spleen, gonadal system, liver, brain, kidney, and in the nucleus where it actively participates in repressing or stimulating many genes expression, the gene transcription regulation. The SET is upregulated in both hematological also in solid tumors, involving breast cancer or colorectal cancer, also in most instance its consequences as an oncogene are because of the PP2A consequent inactivation. The role of SET has been researched in deepness in chronic myeloid leukemia (CML). SET is overexpressed in CML via the BCR-ABL1. After analyzing through SIFT and Polyphen, Polyphen-2 and SNAP2 were used further for the identification of pathogenic SNPs in SET protein. We have determined that All SNPs in the SIFT and Polyphen-2 predicted as deleterious. In Polyphen, there were 113 SNPs predicted as possibly damaging, 21 showed probably damaging and 15 SNPs were predicted as benign. In SNAP2, 111 SNPs were predicted as diseased or effected SNPs, respectively (Table 1). All the SNPs were examined for the diseased association. There were 99, 108, 115 SNPs predicted as diseased by SNP&GO, PhD-SNP, and Meta-SNP tools, respectively. There were 94, 81 and 69 SNPs predicted as deleterious by PredictSNP, MAPP, and SNAP, respectively. 93 SNPs predicted as disordered by PredictProtein and 109 SNPs were predicted by PANTHER. Mu-Pro analysis revealed that all SNPs had decreased protein stability. While I-mutant analysis showed 2 SNPs LP9 and Q12P having increase protein stability.

References

Wiersma, M. (2015). Epigenetic regulation at MLL1 target genes (Doctoral dissertation, University of Birmingham).

Yang, W., & Ernst, P. (2017). Distinct functions of histone H3, lysine 4 methyltransferases in normal and malignant hematopoiesis. Current opinion in hematology, 24(4), 322-328.

Separovich, R. J., Wong, M. W., Chapman, T. R., Slavich, E., Hamey, J. J., & Wilkins, M. R. (2021). Post-translational modification analysis of Saccharomyces cerevisiae histone methylation enzymes reveals phosphorylation sites of regulatory potential. Journal of Biological Chemistry, 296.

Afzal, M., Ali, U., Riaz, A., Tanvir, F., Bilal, A., & Ahmad, S. (2024). In-silico analysis of deleterious single nucleotide polymorphisms (SNPs) of leukemia inhibitory factor (LIF), and their conformational predictions. Journal of Population Therapeutics and Clinical Pharmacology, 31(1), 2792-2811.

Dacol, E. C., Wang, S., Chen, Y., & Lepique, A. P. (2021). The interaction of SET and protein phosphatase 2A as target for cancer therapy. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 1876(1), 188578.

Gole, B., & Wiesmüller, L. (2015). Leukemogenic rearrangements at the mixed lineage leukemia gene (MLL)—multiple rather than a single mechanism. Frontiers in cell and developmental biology, 3, 41.

Bilal, A., Ahmad, S., Tanvir, F., Tariq, M., Ramzan, K., Saleem, M., & Saleem, H. G. M. (2022a). Predictive modeling of N-acetyl transferase 2 single nucleotide polymorphisms and breast cancer risk using in-silco approaches. The Journal of Microbiology and Molecular Genetics, 3(2), 105-121.

Bilal, A., Ahmad, S., Nisa, F. U., Ali, F., Ramzan, K., Tariq, M., & Saleem, M. (2022b). ROLE OF TUMOR NECROSIS FACTOR-? ETA (TNF-?) IN GASTRIC CANCER: SINGLE NUCLEOTIDE POLYMORPHISMS ANALYSIS-AN IN-SILICO STUDY. THE JOURNAL OF MICROBIOLOGY AND MOLECULAR GENETICS, 3(3), 191-206.

Eldeeb, M. (2023). Pediatric Leukemia: Mechanisms of Initiation and Suppression.

Conneely, S. E., & Rau, R. E. (2020). The genomics of acute myeloid leukemia in children. Cancer and Metastasis Reviews, 39(1), 189-209.

Depreter, B. (2020). Exploring the molecular and phenotypical heterogeneity between patients with pediatric acute myeloid leukemia (Doctoral dissertation, Ghent University).

Marschalek, R. (2015). MLL leukemia and future treatment strategies. Archiv der Pharmazie, 348(4), 221-228.

Bilal, A., Tanvir, F., Ahmad, S., Shah, S. H. A., Ahmad, H. A., & Kanwal, N. (2024a). Pre-clinical study of the bioactive compound Asiaticoside against the proteins inducing human mammary carcinoma using molecular docking and ADME analysis. Remittances Review, 9(2), 3543-3576.

Christensen, D. J., Chen, Y., Oddo, J., Matta, K. M., Neil, J., Davis, E. D., ... & Weinberg, J. B. (2011). SET oncoprotein overexpression in B-cell chronic lymphocytic leukemia and non-Hodgkin lymphoma: a predictor of aggressive disease and a new treatment target. Blood, The Journal of the American Society of Hematology, 118(15), 4150-4158.

Alicea-Velázquez, N. L., Shinsky, S. A., Loh, D. M., Lee, J. H., Skalnik, D. G., & Cosgrove, M. S. (2016). Targeted disruption of the interaction between WD-40 repeat protein 5 (WDR5) and mixed lineage leukemia (MLL)/SET1 family proteins specifically inhibits MLL1 and SETd1A methyltransferase complexes. Journal of Biological Chemistry, 291(43), 22357-22372.

Bilal, A., Tanvir, F., Ahmad, S., Kanwal, N., Zulfiqar, H., & Ishaq, R. (2024b). Pharmacokinetic Properties of Bioactive Compounds of Aloe vera against Pregnancy-Associated Plasma Protein A (PAPP-A) inducing Triple-Negative Breast Cancer. Kurdish Studies, 12(5), 157-168.

Ahmed, N., Yigit, A., Isik, Z., & Alpkocak, A. (2019). Identification of leukemia subtypes from microscopic images using convolutional neural network. Diagnostics, 9(3), 104.

Sugeedha, J., Gautam, J., & Tyagi, S. (2021). SET1/MLL family of proteins: functions beyond histone methylation. Epigenetics, 16(5), 469-487.

Lu, R., & Wang, G. G. (2017). Pharmacologic targeting of chromatin modulators as therapeutics of acute myeloid leukemia. Frontiers in oncology, 7, 241.

Jiang, H. (2020). The complex activities of the SET1/MLL complex core subunits in development and disease. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1863(7), 194560.

Samsudin, N. B. (2017). H3K4 methyl marks landscape during Caenorhabditis elegans embryogenesis and their implication in lifespan regulation. The University of Manchester (United Kingdom).

Das, S., Mondal, S., Patel, T., Himaja, A., Adhikari, N., Banerjee, S., ... & Jha, T. (2024). Derivatives of D (-) glutamine-based MMP-2 inhibitors as an effective remedy for the management of chronic myeloid leukemia-Part-I: Synthesis, biological screening and in silico binding interaction analysis. European Journal of Medicinal Chemistry, 116563.

Pippa, R., & Odero, M. D. (2020). The role of MYC and PP2A in the initiation and progression of myeloid leukemias. Cells, 9(3), 544.

Ram, M., Afrash, M. R., Moulaei, K., Parvin, M., Esmaeeli, E., Karbasi, Z., ... & Sabahi, A. (2024). Application of artificial intelligence in chronic myeloid leukemia (CML) disease prediction and management: a scoping review. BMC cancer, 24(1), 1026.

Pandey, M. K., & Pal, S. (2022). Evaluation of chronic myelogenous leukemia (CML) as the chronic phase of disease using machine learning techniques. International Journal of Mechanical Engineering Education, 7, 1-12.

Bilal, A., Tanvir, F., Ahmad, S., Azam, A. R., Qasim, M., Zafar, H., & Tanvir, F. (2024c). THE RAPEUTICAL EVALUATION OF BIOACTIVE COMPOUNDS OF NIGELLA SATIVA FOR HER2-POSITIVE BREAST CANCER TREATMENT. Journal of Population Therapeutics & Clinical Pharmacology. 31 (9), 3149-3164

Kikuchi, S., Kaneda, A., & Hoshii, T. (2023). Catalytic and non-catalytic functions of histone methyltransferase SETD1A in diseases. Chiba Medical Journal, 99(2), 17-25.

Nilsson, C., Hulegårdh, E., Garelius, H., Möllgård, L., Brune, M., Wahlin, A., ... & Lehmann, S. (2019). Secondary acute myeloid leukemia and the role of allogeneic stem cell transplantation in a population-based setting. Biology of blood and marrow transplantation, 25(9), 1770-1778.

Grech, G., Baldacchino, S., Saliba, C., Grixti, M. P., Gauci, R., Petroni, V., ... & Scerri, C. (2016). Deregulation of the protein phosphatase 2A, PP2A in cancer: complexity and therapeutic options. Tumor Biology, 37, 11691-11700.

Markouli, M., Strepkos, D., & Piperi, C. (2021). Structure, activity and function of the SETDB1 protein methyltransferase. Life, 11(8), 817.

Bilal, A. (2021a). Clinical diagnosis and treatment of absence seizures: Case study. MAR Ophthalmology, 2(1).

Bilal, A. (2021b). Impacts of depression on pregnancy: A review. Occup Med Health Aff, 9(2).

Christensen, D. J., Chen, Y., Oddo, J., Matta, K. M., Neil, J., Davis, E. D., ... & Weinberg, J. B. (2011). SET oncoprotein overexpression in B-cell chronic lymphocytic leukemia and non-Hodgkin lymphoma: a predictor of aggressive disease and a new treatment target. *Blood, The Journal of the American Society of Hematology*, *118*(15), 4150-4158.

Trievel, R. C., Beach, B. M., Dirk, L. M., Houtz, R. L., & Hurley, J. H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell*, *111*(1), 91-103.

Umar, M., Bilal, H. M., Riaz, A., Tariq, H. M. A., Mehmood, M., Munib, H. M., ... & Bilal, A. (2025). ELEVATED SERUM URIC ACID LEVELS AS AN ANALYTICAL BIOMARKER FOR HYPERTENSION RISK AMONG YOUTH: A CROSS-SECTIONAL STUDY. Insights-Journal of Health and Rehabilitation, 3(2 (Health & Rehab)), 232-240.