In-Silico Evaluation of the Missense Mutations of *KMT2C* Gene in Breast Cancer

[1]Ankitha V, [2]Sailakshmi Iyer

Department of Biomedical Sciences,
School of Bio Sciences and Technology (SBST),
Vellore Institute of Technology, Vellore, India

***Abstract*: Breast cancer is the second most diagnosed cancer in women globally. More than 5 to 10% of the breast cancer cases are due to inherited gene mutations in genes such as *BRCA1, BRCA2, KMT2C, PTEN*, etc. Various studies have reported the involvement of the *KMT2C* gene in breast cancer but is still unclear. Early detection and proper medical care intervention are the keys to a longer survival rate. However, several studies and clinical trials are underway in identifying a potential cure for breast cancer and its subtypes. The focus of the study was to evaluate missense mutations of the *KMT2C* gene specific for breast cancer.**

***Methodology*— The evaluation of the missense mutations of the *KMT2C* gene-specific for breast cancer was done using various openly accessible bioinformatics tools such as COSMIC, UniProt PredictSNP, I-Mutant Suite, ConSurf, fathmm, and Align GVGD.**

***Result*— The study was successful in identifying the top 3 missense mutations out of the list of mutations retrieved that interfere with the function of the protein encoded by the *KMT2C* gene.**

Index Terms*—Missense mutations, Breast Cancer, COSMIC, *KMT2C

I. INTRODUCTION

Breast cancer is the second most commonly diagnosed cancer and the leading cause of death among women worldwide. Breast cancer develops in the cells of the breast and is a multistep process involving different cell types. Various risk factors play a vital role in the development of breast cancer such as an increase in age (Women aged 45 or above are at a higher risk of developing breast cancer), late pregnancy, inherited gene mutations, etc. Early diagnosis and timely medical care are the best approaches in the prevention of breast cancer as many developed countries have reported to have a 5-year survival rate in about 80% of breast cancer patients [1]. According to the GLOBOCAN 2020 cancer statistics, 19.3 million new cases and 10 million deaths were reported in the year 2020 globally with an exception to basal cell carcinoma and NMSC. An incidence rate of 11. 7% and a mortality rate of 6.9% was reported of the total cases accounted for female breast cancer [2].

Mutations in several genes such as *BRCA1, BRCA2, PALB2, KMT2C, TP53*, etc. have been reported to be associated with breast cancer. The study takes into account the *KMT2C* gene which tops the list of 20 elite genes mutated in breast cancer as reported by the COSMIC database [3] [4]. In recent years it's been very evidently said that epigenetic alterations can cause alterations in the genome and lead to cancer development. Epigenetic mechanisms like DNA modifications, histone modifications, and chromatin remodeling have started to gain more attention as they could hide potential clues for cancer therapies. KMTs or otherwise known as histone lysine methyltransferases are considered key players in epigenetics. They methylate different amino acids on histone thereby affecting the structure of the chromatin and affecting gene transcription of target genes. The *KMT2C* gene located on chromosome 7 (cytogenetic location 7q36.1) encodes Histone lysine N-methyltransferase 2C protein, which plays a crucial role in histone methylation specifically methylating H3 'Lys-4' and also acts as a transcriptional coactivator. The frequency of *KMT2C* gene mutation is about 8% and is shown to be commonly mutated in breast cancer and multiple other cancers [5] [7].

Studies have also reported that KMT2C affects the function of DNA damage repair genes and that low expression of *KMT2C* adversely affects the overall survival of breast cancer patients contributing to chemotherapy resistance thus suggesting that *KMT2C* could serve as a biomarker of prognosis [6]. However, research studies are underway in deciphering the role of *KMT2C* in breast cancer as it's still unclear. Our study aims to evaluate the missense mutations of the *KMT2C* gene specific for breast cancer using in-silico approaches and successfully identify the most commonly occurring *KMT2C* gene mutations in breast cancer.

II. MATERIALS AND METHODS

A. Data Retrieval

The canonical FASTA sequence of the protein histone-lysine N-methyltransferase 2C with (UniProt ID: Q8NEZ4) and encoded by the *KMT2C* gene were retrieved from the UniProt database. The list of missense mutations reported in the case of breast cancer was obtained from the COSMIC database. A total of 284 missense mutations were procured *(Appendix).* The UniProt database harbors complete information related to a protein derived from literature studies. The COSMIC (Catalogue of Somatic Mutations in Cancer) database helps in exploring the somatic mutations associated with cancers by searching for a gene, cancer type, or mutation as input [5] [7].

B. Pathogenicity prediction

The pathogenicity prediction of the missense mutations was carried out using the PredictSNP tool. The PredictSNP tool employs eight prediction tools namely, MAPP, nsSNP Analyzer, PANTHER, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP, and categorizes the mutations as neutral or deleterious. The FASTA sequence of the protein followed by the list of mutations was given as input for prediction [8].

C. Protein Stability prediction

The stability of protein was predicted using the I-Mutant 3.0 Suite server wherein the FASTA sequence of the protein without the headers and the mutations were given as input. The I-Mutant tool effectively predicts the stability of a protein as "increasing " or "decreasing" in stability upon single point mutation in the protein sequence. The tool also enables the user to predict the stability of the protein from a protein structure if available [9].

D. Evolutionary Conservational analysis

The in-silico tool ConSurf was used to identify the functional regions of a protein. The tool identifies the functional regions by estimating the amino acid sites that make up a protein and are assigned different conservation grades ranging from a scale of 1-9, 1 being highly variable and 9 being highly conserved. The FASTA sequence of the protein was given as input [10].

E. Functional analysis of cancer-promoting mutations

To identify cancer-promoting mutations, an online bioinformatics tool 'Fathmm' (Functional Analysis through Hidden Markov Models) was used. The online tool is very useful in predicting cancer-promoting mutations, germline polymorphism, in both the coding and non-coding variants comprising the human genome. The Uniprot ID of the protein followed by comma-separated mutations was given as input Example: (Q8NEZ4 G292E, C438Y....) [11].

F. Missense mutation analysis

The analysis of missense mutations in the *KMT2C* gene was carried out via the Align GVGD bioinformatics tool. The tool effectively categorizes the biochemical properties of the amino acids such as the composition, volume, and polarity. The values obtained are in the form of Grantham Variation score (GV) and Grantham Difference score (GD). The scores range from a spectrum of classes (C0, C15, C25, C35, C45, C55, C65) with C65 being most likely and C0 least likely to interfere with the function. The FASTA sequence of the protein followed by the mutations was given as input [12].

III. RESULTS

The results of the missense mutations of the *KMT2C* gene obtained after analysis through various bioinformatics tools are plotted into a bar graph and tabular column.

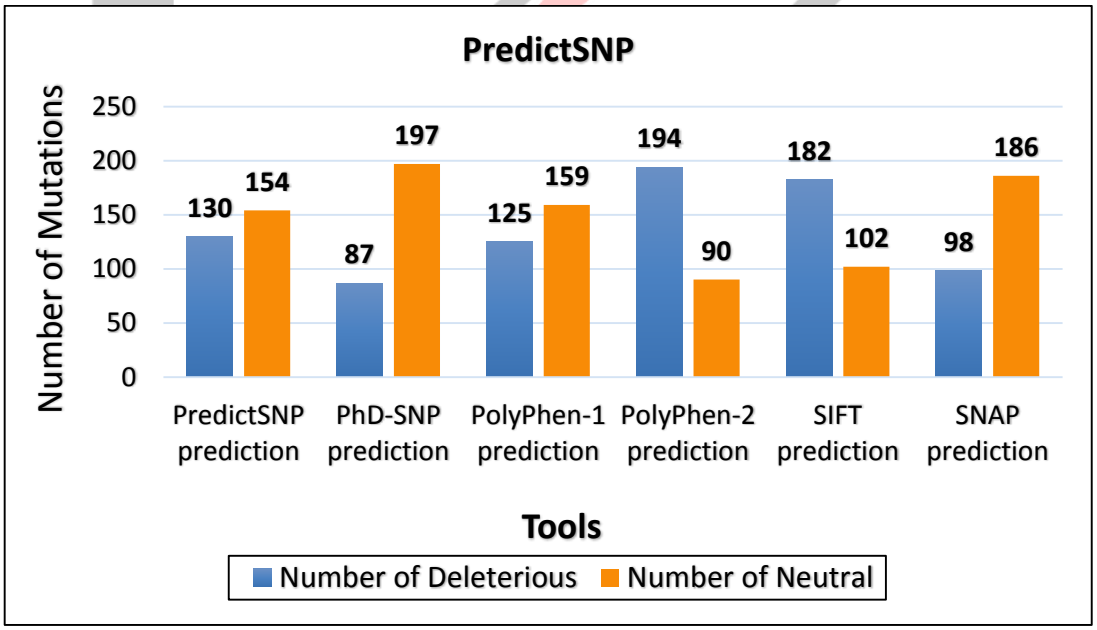

Figure 1: Graph representing the results for pathogenicity prediction using 8 tools of PredictSNP software.

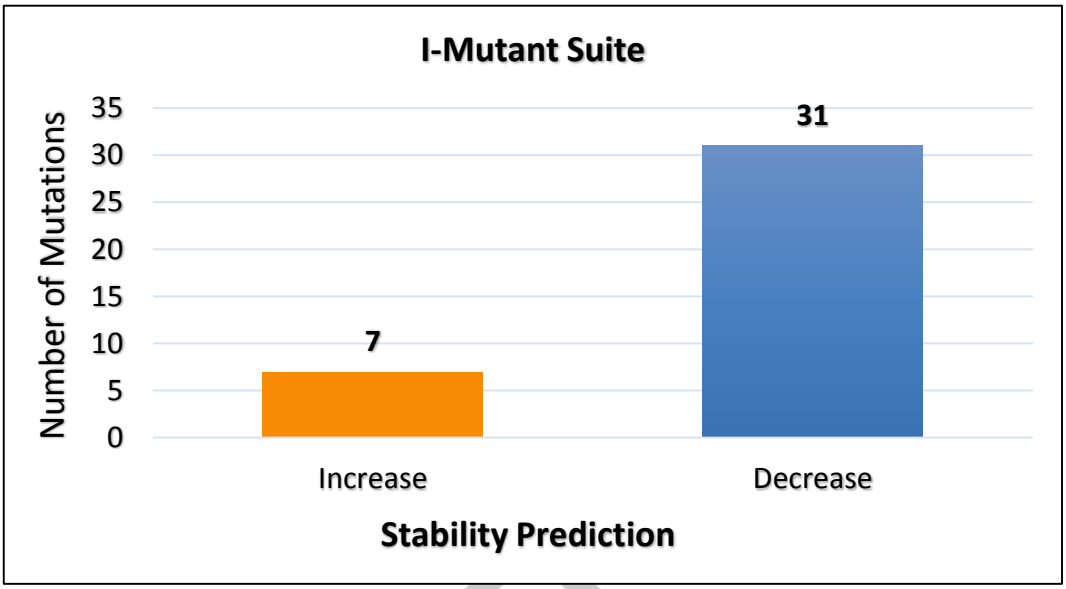

Figure 2: Graph representing the results of protein stability prediction using iMutant software.

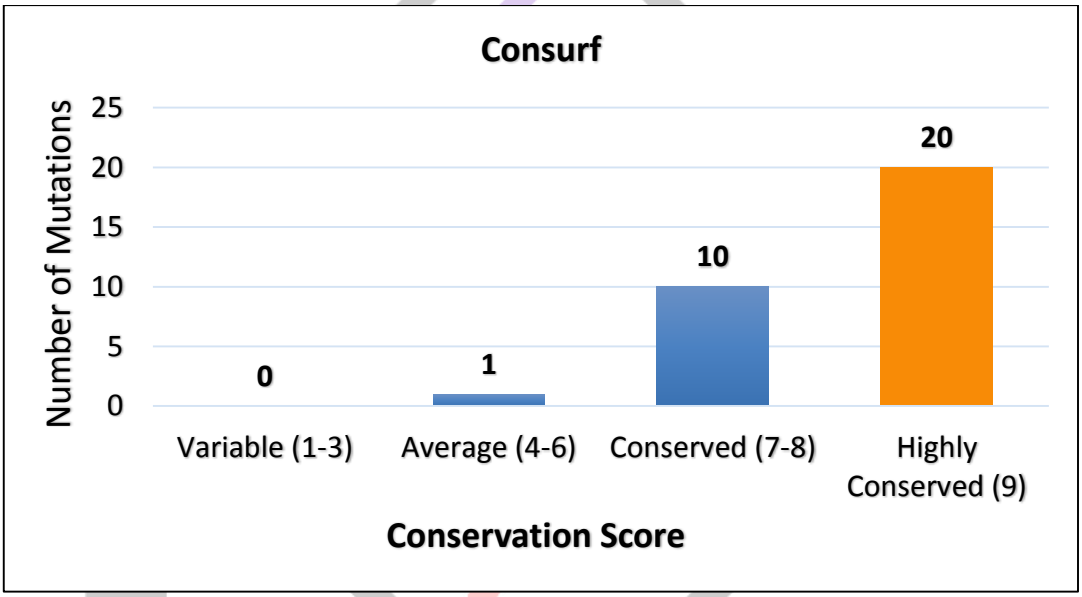

Figure 3: Graph representing the results of conservation analysis by ConSurf.

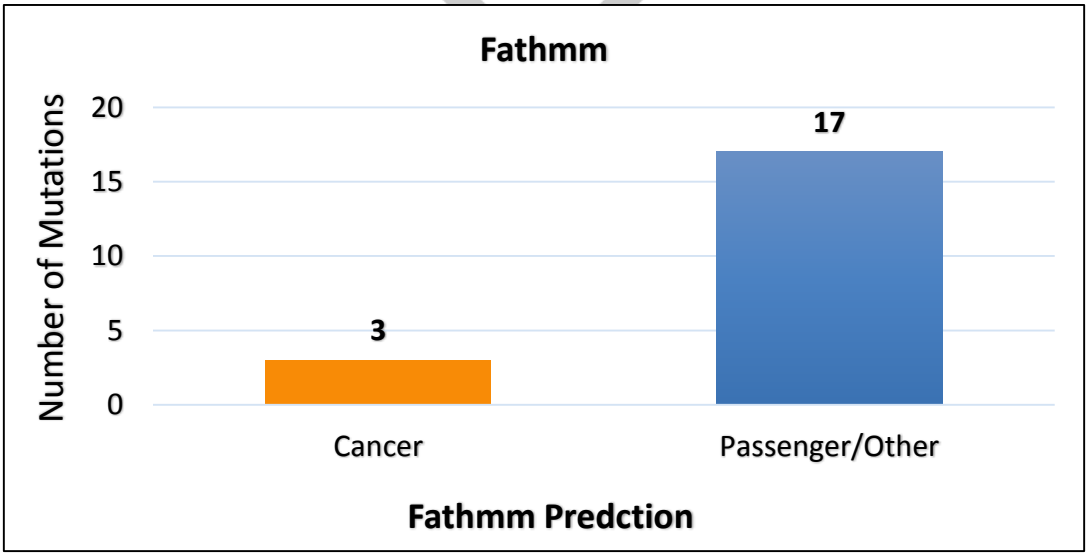

Figure 4: Graph representing the results of Fathmm Prediction

Table 1: Table representing the Align GVGD scores for 3 mutations.

Mutation	Align GVGD Scores	GV	GD
H367D	Class C65	0	81.24
W430C	Class C65	0	214.36
G3106D	Class C65	0	93.77

Table 2: Table representing the top 3 mutations shortlisted out of the 284 missense mutations.

Mutation	Pathogenicity	Stability	Conservation	Fathmm	Align GVGD
H367D	DELETERIOUS	Decrease	Highly Conserved	CANCER	Class C65
W430C	DELETERIOUS	Decrease	Highly Conserved	CANCER	Class C65
G3106D	DELETERIOUS	Decrease	Highly Conserved	CANCER	Class C65

IV. DISCUSSION

A. Pathogenicity Prediction

PredictSNP, an online web tool was used to predict the effect of the amino acid substitutions. PredictSNP classified the 284 missense mutations using 8 different tools as represented in *Figure 1*. Among the 284 mutations, PredictSNP classified 130 mutations as deleterious and 154 mutations as neutral. PhD-SNP classified 87 mutations as deleterious and 197 mutations as neutral. PolyPhen-1 tool classified 125 mutations as deleterious and 159 mutations as neutral. PolyPhen-2 classified 194 mutations as deleterious and 90 mutations as neutral. SIFT tool classified 182 mutations as deleterious and 102 mutations as neutral. SNAP classified 98 mutations as deleterious and 186 mutations as neutral. MAPP and PANTHER tools were unable to predict/classify the mutations into deleterious and neutral. Overall, 38 missense mutations were classified as deleterious by all the prediction tools.

B. Protein Stability Prediction

The 38 missense mutations classified by PredictSNP were subjected to stability prediction to analyze the stability of the protein upon mutation by using an online web tool- I-Mutant 3.0 Suite. Among 284 mutations, I-Mutant Suite predicted 31 mutations to have "Decrease" in stability and 7 mutations to have "Increase" in stability as represented in *Figure 2*.

C. Evolutionary Conservation Analysis

ConSurf server was used to estimate the evolutionary rates of amino acids. The 31 mutations with decreased stability were subjected to conservation analysis and among the 31 mutations, 1 mutation was said to have average conservation with a ConSurf score of 5, 10 mutations were said to be conserved with a ConSurf score of 7 and 8, and 20 mutations were said to be highly conserved with the ConSurf score of 9 as represented in *Figure 3*.

D. Functional analysis of cancer-promoting mutations

The highly conserved 20 mutations with ConSurf score 9 were then exposed to Fathmm, an online web tool used to distinguish cancer-promoting mutation from other mutations. Among the 20 mutations, 3 mutations were predicted as cancer-promoting and 17 mutations were predicted as Passenger/other types of mutation as represented in *Figure 4*.

E. Missense mutation analysis

The 3 cancer-promoting mutations obtained were finally exposed to the Align GVGD webtool, which characterizes the biochemical properties of amino acids at each position. All the 3 mutations were predicted to fall under the spectrum of Class C65 which is said to "most likely" interfere with the function of the protein as represented in *Table 1*.

V. CONCLUSION

Globally breast cancer is the second most predominant cancer among women and is responsible for millions of deaths worldwide annually. Various genes are altered in the case of Breast Cancer, but the primary focus of this study is *KMT2C* (Lysine Methyltransferase 2C) gene which possesses histone methylation activity and is shown to be mutated in breast cancer. However, various research studies are being carried out in trying to elucidate the role of KMT2C in breast cancer. Extensive computational pipelines were used in this study to evaluate the missense mutations of the *KMT2C* gene in breast cancer. Web-based servers that are openly accessible such as PredictSNP, I-Mutant, ConSurf, Fathmm, and Align GVGD were used for the evaluation of pathogenicity, protein stability, evolutionary conservation, analysis of cancer-promoting mutations of the 284 missense mutations of the *KMT2C* gene. The study was successful in shortlisting the top 3 mutations (H367D, W430C, and G3106D) as they were

found to be deleterious, decrease in stability, highly conserved, cancer-promoting, and most likely to interfere with the function of the protein as depicted in *Table 2*.

VI. REFERENCES

[1] Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., Shi, W., Jiang, J., Yao, P. P., & Zhu, H. P. (2017). Risk Factors and Preventions of Breast Cancer. *International journal of biological sciences*, *13*(11), 1387–1397. https://doi.org/10.7150/ijbs.21635

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

[3] Breast Cancer Association Consortium, Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K. A., Parsons, M. T., Fortuno, C., Wang, Q., Bolla, M. K., Dennis, J., Keeman, R., Alonso, M. R., Álvarez, N., Herraez, B., Fernandez, V., Núñez-Torres, R., Osorio, A., … Easton, D. F. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*, *384*(5), 428–439. https://doi.org/10.1056/NEJMoa1913948

[4] Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., Speedy, H. E., … Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

[5] Chen, X., Zhang, G., Chen, B., Wang, Y., Guo, L., Cao, L., Ren, C., Wen, L., & Liao, N. (2019). Association between histone lysine methyltransferase KMT2C mutation and clinicopathological factors in breast cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, *116*, 108997. https://doi.org/10.1016/j.biopha.2019.108997

[6] Liu, X., Qiu, R., Xu, M., Meng, M., Zhao, S., Ji, J., & Yang, Y. (2021). KMT2C is a potential biomarker of prognosis and chemotherapy sensitivity in breast cancer. *Breast cancer research and treatment*, *189*(2), 347–361. https://doi.org/10.1007/s10549-021-06325-1

[7] The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, https://doi.org/10.1093/nar/gkaa1100

[8] Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology*, *10*(1), e1003440. https://doi.org/10.1371/journal.pcbi.1003440

[9] Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, *33*(Web Server issue), W306–W310. https://doi.org/10.1093/nar/gki375

[10] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, *44*(W1), W344–W350. https://doi.org/10.1093/nar/gkw408

[11] Shihab, H. A., Gough, J., Mort, M., Cooper, D. N., Day, I. N., & Gaunt, T. R. (2014). Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Human genomics*, *8*(1), 11. https://doi.org/10.1186/1479-7364-8-11

[12] Tavtigian, S. V., Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., de Silva, D., Zharkikh, A., & Thomas, A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of medical genetics*, *43*(4), 295–305. https://doi.org/10.1136/jmg.2005.033878

VII. APPENDIX

List of missense mutations

Mutation	COSMIC ID	Mutation	COSMIC ID	Mutation	COSMIC ID	Mutation	COSMIC ID
P12S	COSM6936867	S401N	COSM6986053	G917R	COSM5075578	R1718S	COSM9179993
T79M	COSM9277533	H414Y	COSM397339	S939C	COSM9360088	S1733L	COSM7344680
E97Q	COSM1488416	L418V	COSM5783371	G964R	COSM5025006	F1736S	COSM9179381
C134Y	COSM452745	M422I	COSM5597096	L974F	COSM9277527	E1748K	COSM6906215
N169D	COSM9122409	G429D	COSM9122421	C977Y	COSM452723	Q1775H	COSM9360052
K170T	COSM6904604	W430C	COSM745836	G981D	COSM4774834	V1776G	COSM3832359
Q203H	COSM6354030	C438Y	COSM6906214	Q982K	COSM7677293	L1793F	COSM1488392
I237V	COSM7344623	D457G	COSM6193626	C1004F	COSM9146000	K1832E	COSM6956463
A241S	COSM3879611	P468A	COSM7344615	D1029N	COSM5026861	T1837P	COSM3832355
A241G	COSM7344621	E477Q	COSM7344698	L1067V	COSM162510	S1840C	COSM7344585
H252Y	COSM4391355	E523K	COSM6945125	E1096V	COSM7685622	V1846M	COSM6931558
D274E	COSM9180071	M524R	COSM7344613	E1097D	COSM7344684	S1860C	COSM7344583
V278I	COSM7344619	R526L	COSM7344611	R1108T	COSM6921653	P1863A	COSM9277024
R284Q	COSM1179106	E547K	COSM7673555	C1114F	COSM6904618	P1877T	COSM7693978
G292E	COSM4381097	V576F	COSM7344696	E1122Q	COSM5767522	R1890G	COSM7344678
C302Y	COSM7674122	S593L	COSM6945663	E1152G	COSM6337227	P1891T	COSM9360066
P309S	COSM1179670	L596V	COSM6945661	E1165Q	COSM9179419	R1917T	COSM7238094
C310S	COSM1087662	D599V	COSM5777507	T1171S	COSM452719	T1957M	COSM3879574
A311G	COSM7697559	Q666K	COSM9122413	Q1186E	COSM9180064	Y1963H	COSM9180196
G315C	COSM4381930	R679K	COSM6910102	L1187S	COSM9122427	T1969I	COSM7344676
T316N	COSM3304643	E681K	COSM7344639	T1191K	COSM162504	E1992Q	COSM452707
Q318E	COSM7675995	L687F	COSM9122429	I1223F	COSM9122425	S2025L	COSM6969506
E329K	COSM7699065	M689I	COSM9122433	S1334C	COSM7670974	P2029H	COSM7344674
K339N	COSM4587276	S691C	COSM3396809	S1336F	COSM5966019	P2042L	COSM9122437
E340K	COSM9277531	P711A	COSM7344692	E1356V	COSM7344682	R2066Q	COSM6941895
D348E	COSM7344617	E718A	COSM4774832	T1358A	COSM9277023	Y2174H	COSM9277525
D348N	COSM7335280	R719M	COSM7677800	I1381V	COSM452715	F2190V	COSM9122435
S349R	COSM9267409	G722E	COSM452729	S1416L	COSM3304239	T2195A	COSM7344670
C359S	COSM3832375	L732F	COSM6109527	A1432T	COSM7344601	H2205P	COSM9360068
T360N	COSM3304593	E765G	COSM227525	S1469F	COSM4771399	A2223V	COSM7344668
T361I	COSM3304589	V781M	COSM7688685	P1472L	COSM7344597	E2231K	COSM7344666
H365Q	COSM4361025	P787R	COSM7688131	T1536A	COSM6904313	G2232D	COSM9277523
H367D	COSM3675545	M812I	COSM6908440	P1544L	COSM7344595	P2268S	COSM7344664
H367Y	COSM452741	K822R	COSM5788541	T1545I	COSM212884	G2282R	COSM7344577
C370Y	COSM7693063	S854R	COSM7344605	D1554G	COSM7343415	R2296H	COSM6906126
R380L	COSM225885	I868T	COSM7344635	P1606L	COSM4384337	R2388H	COSM3698285
G382V	COSM9180118	S888T	COSM7344686	N1612D	COSM7344593	P2467S	COSM9112748
C385F	COSM9180259	G892E	COSM6915123	R1634S	COSM9360074	R2515K	COSM1673707
S397L	COSM9360064	R894Q	COSM4595027	A1679S	COSM6961809	H2540L	COSM6911050
G398E	COSM5775847	S914R	COSM3832371	S1686N	COSM7344633	D2567Y	COSM162512

R2571W	COSM7344569	R3224L	COSM6904620	I4055M	COSM7344563	R4418G	COSM3785043
G2579R	COSM7449877	D3264N	COSM3832339	S4066L	COSM9179415	L4420R	COSM5964207
H2604R	COSM7343414	V3284F	COSM7665385	A4100T	COSM9360060	D4425H	COSM6916529
R2610Q	COSM1581241	P3289L	COSM7344658	L4136V	COSM7344561	D4425Y	COSM452671
Q2632E	COSM9360072	T3303P	COSM3832335	P4149L	COSM6981257	D4425A	COSM7344555
E2687D	COSM9180069	L3346V	COSM213176	S4151F	COSM6986032	N4431D	COSM9360084
E2705Q	COSM6927226	S3394N	COSM1488390	V4158M	COSM6920052	V4439F	COSM6970961
E2718K	COSM6942212	R3400C	COSM7254176	S4190N	COSM9277515	T4468M	COSM7344645
E2718Q	COSM6977876	G3438A	COSM9277519	L4193H	COSM7344559	C4474F	COSM9358125
P2794Q	COSM9122439	P3496L	COSM9122441	R4194T	COSM7697588	H4484L	COSM9122423
E2838K	COSM6938843	V3524A	COSM452695	C4198W	COSM3396683	I4489M	COSM7450038
E2885K	COSM6938843	P3555T	COSM7344654	R4211W	COSM6697327	C4493Y	COSM7344643
Q2903E	COSM6916531	G3584R	COSM452693	I4244L	COSM7449780	K4499R	COSM7684495
S2935L	COSM7344631	I3590L	COSM1673700	A4252V	COSM6854690	H4512R	COSM9277513
S2937T	COSM6906831	R3606T	COSM7667922	S4255P	COSM9122445	E4515K	COSM9360070
R2963C	COSM485080	K3609N	COSM6933662	E4256K	COSM6965172	D4530H	COSM6005475
N2967T	COSM162506	E3717K	COSM9361231	S4260C	COSM6930681	G4542A	COSM9180019
E3050K	COSM9122415	E3724K	COSM3262541	F4277S	COSM7344649	R4549C	COSM245709
N3086K	COSM6471771	E3732V	COSM6947858	H4290Y	COSM7344647	Y4580S	COSM5791320
I3092L	COSM9277521	P3794R	COSM9277517	P4296R	COSM5960672	N4593S	COSM9360082
K3100R	COSM7344567	M3827I	COSM6193622	S4300A	COSM2153780	R4595G	COSM6906127
A3103T	COSM6969504	G3834R	COSM9122419	R4334W	COSM1087522	R4595P	COSM9277511
A3103S	COSM7678298	E3845Q	COSM7669531	R4334Q	COSM6981121	R4597C	COSM6916678
L3104F	COSM7408660	R3853Q	COSM6932630	E4343K	COSM6966994	V4719D	COSM9360080
G3106D	COSM452697	K3866N	COSM6177139	D4344H	COSM3832333	R4779W	COSM7211286
G3135V	COSM6920789	R3868M	COSM7344652	R4353T	COSM6931506	N4811K	COSM7672384
Q3142R	COSM7344660	E3873Q	COSM162508	E4378K	COSM4943274	G4842R	COSM5788832
A3148S	COSM3832343	V3938I	COSM4333154	E4381K	COSM6910062	E4880K	COSM9122417
P3169S	COSM5798402	D3996N	COSM5935298	R4400G	COSM9179324	D4892G	COSM6904609
S3213L	COSM1312866	E4008K	COSM6958684	R4400Q	COSM5453040	H4895Q	COSM7344545
A3222G	COSM6924026	P4035T	COSM9306553	G4415E	COSM162514	G4902A	COSM6919858