Genomics & Proteomics & Significance Tests

Question 1

Firstly, the single nucleotide variations (SNVs) present in the table (Table 1.) have been obtained from the gnomAD database. The information pertaining to the SNVs is reported below in sequence.

Table 1.

4-105236371-A-G
17-746331-T-C
4-1395373-A-G
19-48628944-T-C

Variant: 4-105236371-A-G

Link: https://gnomad.broadinstitute.org/variant/4-105236371-A-G?dataset=gnomad_r4

a) Total population frequency is 0.0006202 for the SNV. The lowest frequency is 0.00004392 for South Asians. The highest frquency is 0.007793 for African/African American. We can check the all population frequency, count and number of homozygotes informations using the table in Figure 1.

Genetic Ancestry Group Frequencies

gnomAD | HGDP | 1KG | Local Ancestry

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African/African American	585	75066	3	0.007793
Admixed American	106	59998	1	0.001767
Remaining	62	62510	1	0.0009918
Middle Eastern	6	6062	0	0.0009898
European (non-Finnish)	233	1180006	1	0.0001975
East Asian	5	44862	0	0.0001115
South Asian	4	91082	0	0.00004392
European (Finnish)	0	63928	0	0.000
Ashkenazi Jewish	0	29602	0	0.000
Amish	0	912	0	0.000
XX	531	812402	1	0.0006536
XY	470	801626	5	0.0005863
Total	**1001**	**1614028**	**6**	**0.0006202**

Figure 1. Population Frequency Table for 4-105236371-A-G.

Gene position is 4-105145875-105279816 (region) for 4-105236371-A-G. (https://gnomad.broadinstitute.org/region/4-105145875-105279816?dataset=gnomad_r4) The gene position is displayed in Figure 2. The variant in question is observed in 6 transcripts of the ***tet methylcytosine dioxygenase (TET2)*** gene.

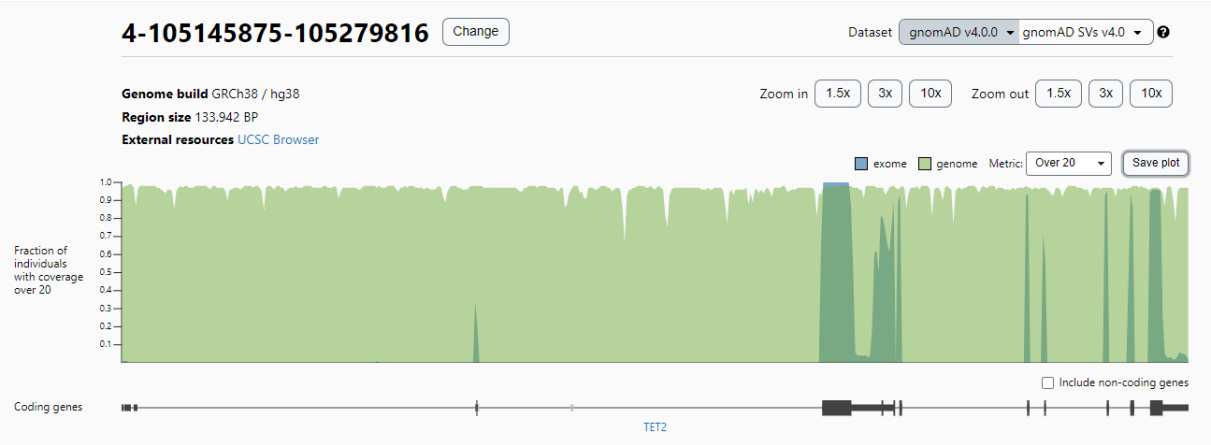


Figure 2. 4-105236371-A-G's gene position.

b) REVEL and PolyPhen Scores

While **REVEL** score is **0.0780**, **Polyphen** (max) score is **0.0100**. The REVEL score of 0.0780 generally indicates a lower probability of pathogenicity. The PolyPhen score, which attempts to predict the effects of a variant on the function of a protein, is 0.0100, typically suggesting a lower probability of pathogenicity as well. Both scores ultimately demonstrate a lower pathogenicity characteristic. The low PolyPhen score can be "non-synonymous".

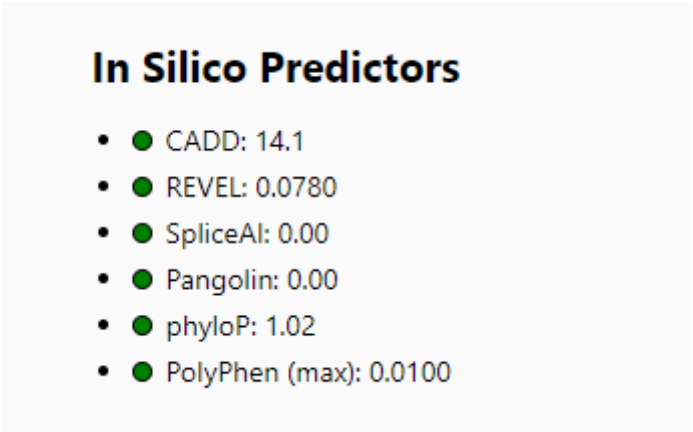
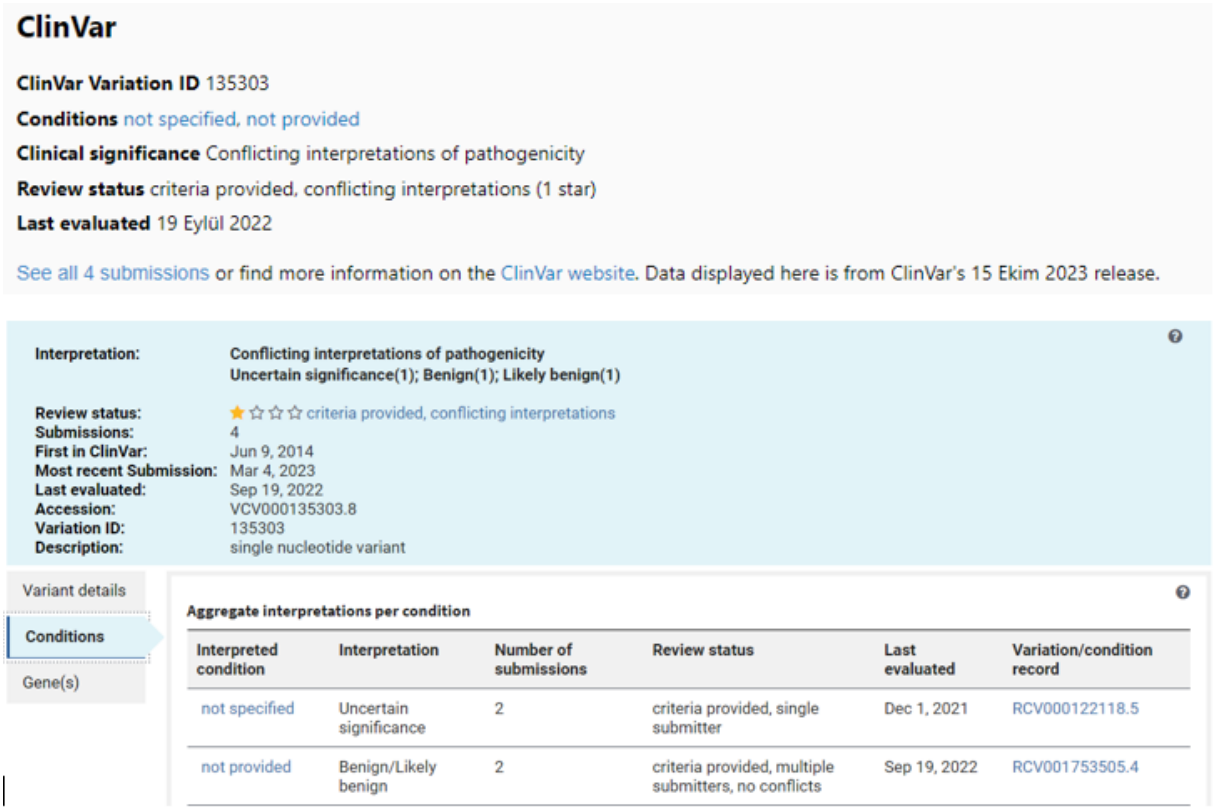


Figure 3. ***In Silico*** **Predictors.**

c) ClinVar Informations

According to ClinVar data, the variant displaying uncertain significance and benign/likely benign characteristics is assessed to have a lower tendency for disease manifestation. These findings from ClinVar are consistent with our REVEL and PolyPhen scores.



ClinVar

ClinVar Variation ID 135303
Conditions not specified, not provided
Clinical significance Conflicting interpretations of pathogenicity
Review status criteria provided, conflicting interpretations (1 star)
Last evaluated 19 Eylül 2022

See all 4 submissions or find more information on the ClinVar website. Data displayed here is from ClinVar's 15 Ekim 2023 release.

Interpretation:	**Conflicting interpretations of pathogenicity** **Uncertain significance(1); Benign(1); Likely benign(1)**
Review status:	★☆☆☆ criteria provided, conflicting interpretations
Submissions:	4
First in ClinVar:	Jun 9, 2014
Most recent Submission:	Mar 4, 2023
Last evaluated:	Sep 19, 2022
Accession:	VCV000135303.8
Variation ID:	135303
Description:	single nucleotide variant

Variant details
Conditions
Gene(s)

Aggregate interpretations per condition

Interpreted condition	Interpretation	Number of submissions	Review status	Last evaluated	Variation/condition record
not specified	Uncertain significance	2	criteria provided, single submitter	Dec 1, 2021	RCV000122118.5
not provided	Benign/Likely benign	2	criteria provided, multiple submitters, no conflicts	Sep 19, 2022	RCV001753505.4

Figure 4. ClinVar Report.

d) Variant Filter/Prioritize

The variant in question exhibits missense characteristics; however, based on the scores and information obtained from gnomAD and ClinVar, the likelihood of it being pathogenic is low. In addition to there are conflicting comments regarding pathogenicity.

Variant: 17-746331-T-C

Link: https://gnomad.broadinstitute.org/variant/17-746331-T-C?dataset=gnomad_r4

a) Total population frequency is **0.005788** for the SNV. The lowest frequency is 0.0003565 for East Asians. The highest frquency is 0.01171for Middle Eastern. We can check the all population frequency, count and number of homozygotes informations using the table in Figure 5.

Genetic Ancestry Group Frequencies

gnomAD | HGDP | 1KG | Local Ancestry

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Middle Eastern	71	6062	1	0.01171
Amish	8	912	0	0.008772
South Asian	622	91086	5	0.006829
Remaining	416	62484	0	0.006658
European (non-Finnish)	7683	1179872	30	0.006512
Admixed American	248	60020	0	0.004132
Ashkenazi Jewish	59	29608	0	0.001993
African/African American	143	75008	1	0.001906
European (Finnish)	74	63818	0	0.001160
East Asian	16	44880	0	0.0003565
XX	4649	812292	17	0.005723
XY	4691	801458	20	0.005853
Total	**9340**	**1613750**	**37**	**0.005788**

Figure 5. Population Frequency Table for 17-746331-T-C

Gene position is 17-744421-753999 for **17-746331-T-C**; displayed the Figure 7. The variant in question is observed in 6 transcripts of the ***gem nuclear organelle associated protein 4 (GEMIN4)*** gene.

https://gnomad.broadinstitute.org/region/17-744421-753999?dataset=gnomad_r4

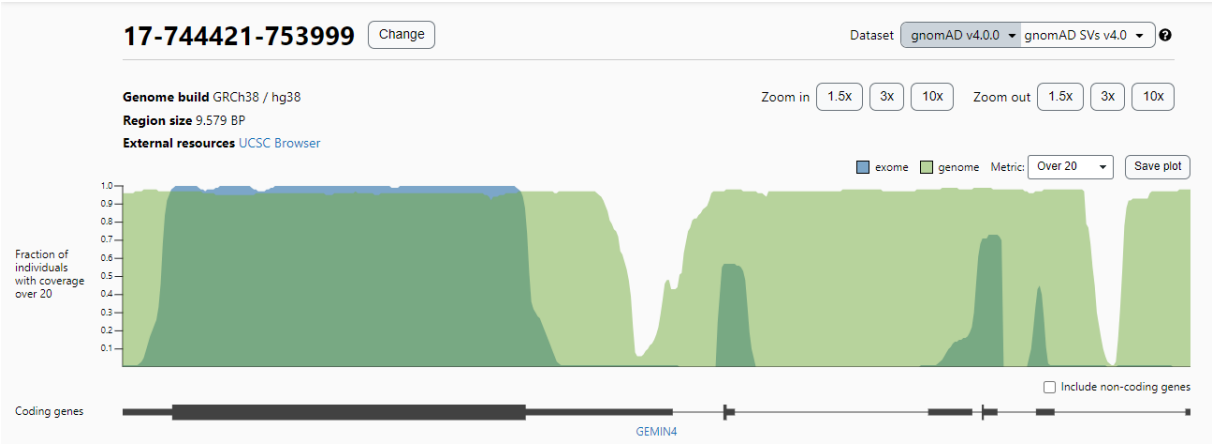


Figure 6. Gene Position for 17-746331-T-C

b) REVEL and PolyPhen Scores

While **REVEL** score is **0.295, Polyphen (max**) score is **0.996.** The REVEL score of 0.295 for the genetic variant indicates a low likelihood of a pathogenic effect, while the PolyPhen (max) score of 0.996 is considerably high, suggesting the potential for a pathogenic impact. However, it is important to note that although a high PolyPhen score implies the potential harmfulness of the genetic variant, in specific contexts and diseases, there is also a possibility that no harmful effect may be observed.

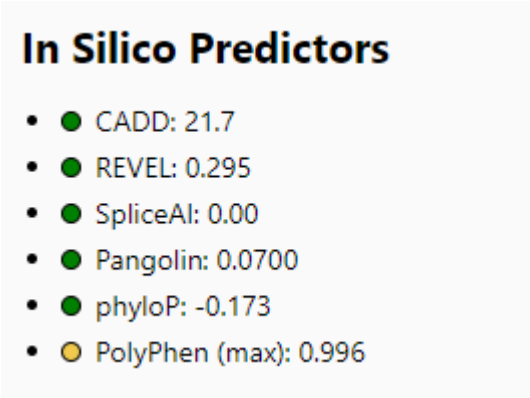


Figure 7. *In Silico* Predictors.

c) ClinVar Informations

According to data obtained from ClinVar, the variant exhibits characteristics indicative of being likely benign and does not display any features of pathogenicity. However, the notable high PolyPhen score of 0.996 is a point of attention. The elevated PolyPhen score of 0.996 suggests a high likelihood of pathogenicity.

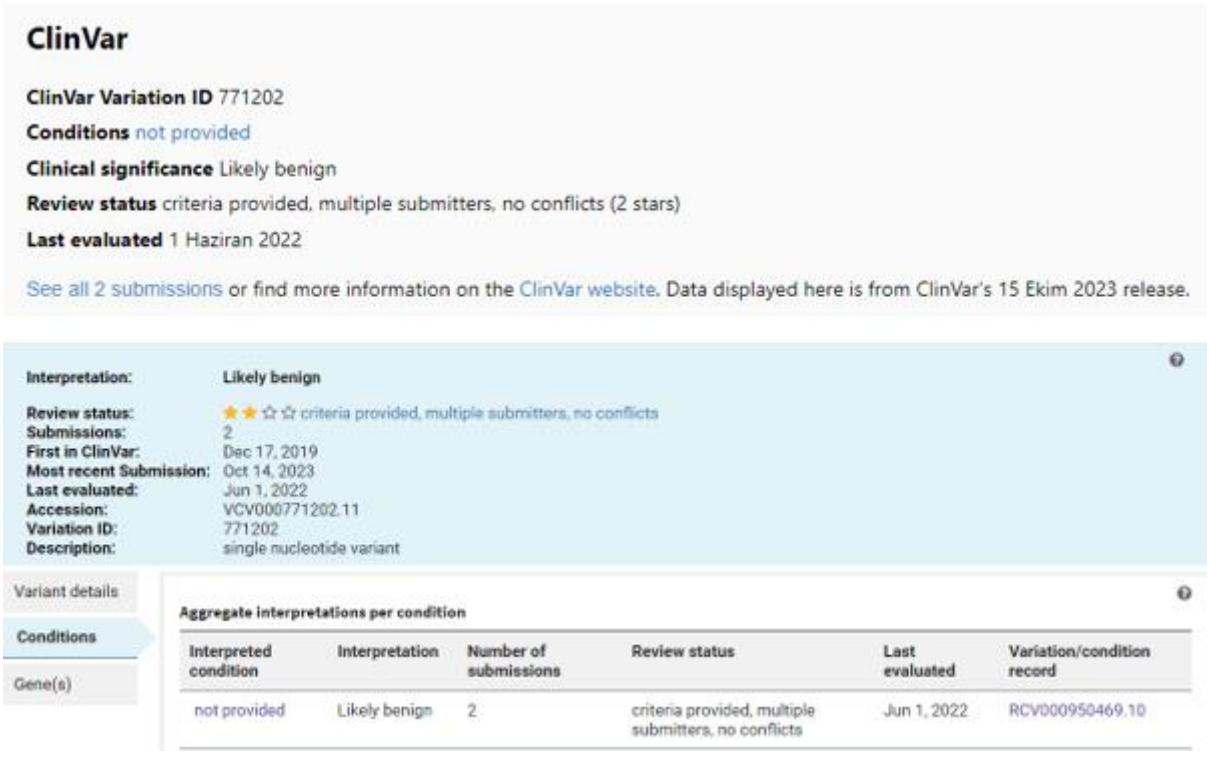


Figure 8. ClinVar Report.

d) Variant Filter/Prioritize

The variant is characterized as missense, and studies conducted have not indicated any harmful effects. However, this does not necessarily imply the absence of direct pathogenicity.

Variant: 4-1395373-A-G

Link: https://gnomad.broadinstitute.org/variant/4-1395373-A-G?dataset=gnomad_r4

a) Total population frequency is **0.3576** for the SNV. The lowest frequency is 0.05896 for East Asians. The highest frquency is 0.3944 for European(non-Finnish). We can check the all population frequency, count and number of homozygotes informations using the table in Figure 9.

Genetic Ancestry Group Frequencies

gnomAD | HGDP | 1KG | Local Ancestry

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (non-Finnish)	411130	1042334	91674	0.3944
European (Finnish)	18017	45780	4329	0.3936
Remaining	18322	54598	3654	0.3356
Ashkenazi Jewish	7922	24574	1546	0.3224
Admixed American	13466	45772	2527	0.2942
South Asian	20648	84166	2997	0.2453
Middle Eastern	1060	4732	176	0.2240
Amish	33	238	11	0.1387
African/African American	4565	49442	497	0.09233
East Asian	2320	39346	98	0.05896
XX	251476	697746	54789	0.3604
XY	246007	693236	52720	0.3549
Total	**497483**	**1390982**	**107509**	**0.3576**

Figure 9. Population Frequency for 4-1395373-A-G

Gene position is 4-1345691-1395989 for **4-1395373-A-G;** displayed the gene position on Figure 10. The variant in question is observed in 6 transcripts of the ***UV stimulated scaffold protein A***

(UVSSA) gene.

https://gnomad.broadinstitute.org/gene/ENSG00000163945?dataset=gnomad_r4

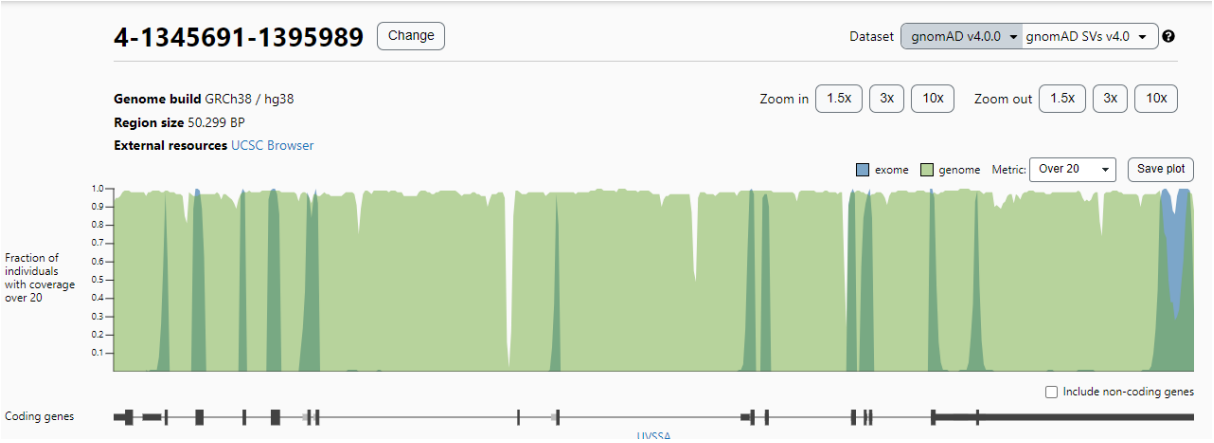


Figure 10. Gene Position for 4-1395373-A-G

b) REVEL and PolyPhen Scores

The variant hasn't got REVEL and PolyPhen scores; however it has got CADD, SpliceAl, Pangolin, and phyloP scores.

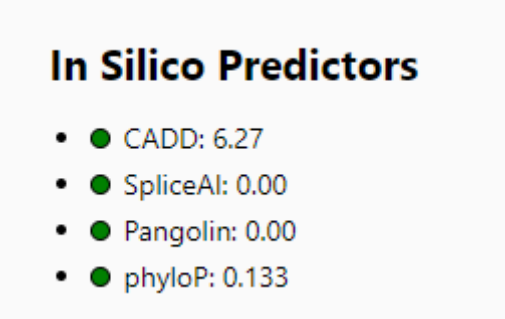


Figure 11. *In Silico* Predictors.

c) ClinVar Informations

Not reported in ClinVar.

d) Variant Filter/Prioritize

The ClinVar data, as well as REVEL and PolyPhen scores specific to the relevant variant, are not available. Therefore, we cannot make a direct inference about the likelihood of the variant causing the disease and its pathogenicity.

Variant: 19-48628944-T-C

Link: https://gnomad.broadinstitute.org/variant/19-48628944-T-C?dataset=gnomad_r4

Total population frequency is **0.00003204** for the SNV. The lowest frequency is 0.000008608 for European(non-Finnish). The highest frquency is 0.001795 for Middle Eastern. We can check the all population frequency, count and number of homozygotes informations using the table in Figure 1.

Genetic Ancestry Group Frequencies

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Middle Eastern	7	3900	0	0.001795
Remaining	3	32902	0	0.00009118
African/African American	1	17572	0	0.00005691
East Asian	2	35784	0	0.00005589
South Asian	3	69598	0	0.00004310
Admixed American	1	43342	0	0.00002307
European (non-Finnish)	3	348500	0	0.000008608
European (Finnish)	0	51760	0	0.000
Ashkenazi Jewish	0	20890	0	0.000
XX	5	284162	0	0.00001760
XY	15	340086	0	0.00004411
Total	**20**	**624248**	**0**	**0.00003204**

Figure 12. Population Frequency for 19-48628944-T-C.

Gene position is 19-48619291-48630717 for **19-48628944-T-C;** displayed the gene position on Figure 10. The variant in question is observed in 6 transcripts of the ***sphingosine kinase 2***

(SPHK2) gene.

https://gnomad.broadinstitute.org/region/19-48619291-48630717?dataset=gnomad_r4

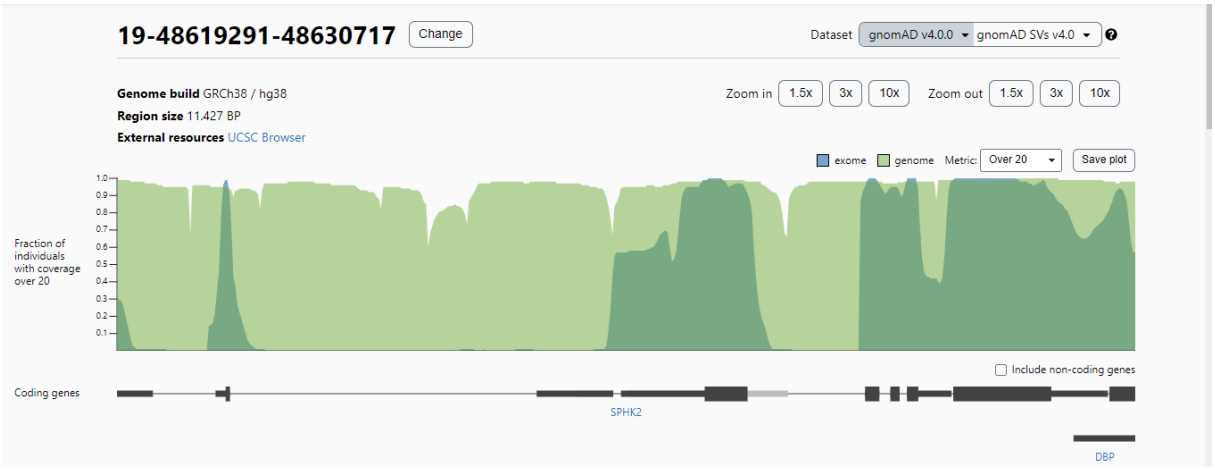


Figure 13. Gene Position for 19-48628944-T-C

b) REVEL and PolyPhen Scores

While REVEL score is 0.420, Polyphen (max) score is 0.998. While a REVEL score of 0.420 and a PolyPhen (Max) score of 0.998, this variant is associated with a lower likelihood of a disease according to REVEL. However, according to PolyPhen, it may signify a more impactful alteration in the protein, indicating a higher probability of pathogenicity.

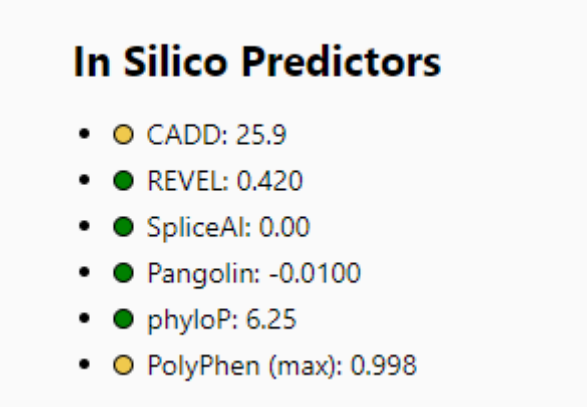


Figure 14. *In Silico* Predictors.

c) ClinVar Informations

Not reported in Clinvar.

d) Variant Filter/Prioritize

The ClinVar data for the relevant variant is not available. However, by examining the REVEL and PolyPhen scores, we can suggest that there is a high likelihood of the variant being pathogenic.

Question 2. Protein Identification Through PMF (Peptide Mass Fingerprinting)

Firstly, we utilized the .txt file provided to us during the search stage on the Mascot page.

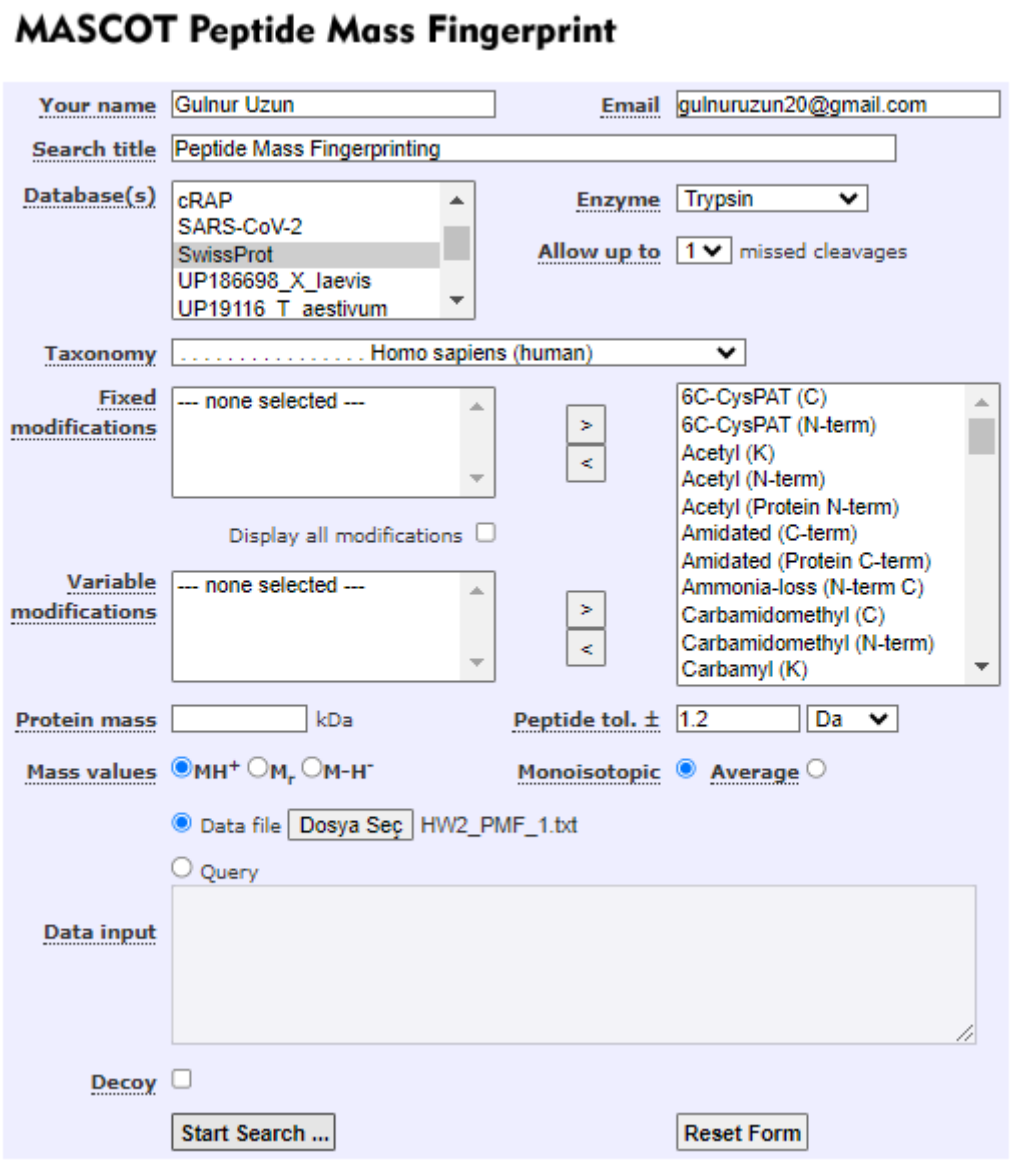


Figure 15. First Step on Mascot.

```
User            : Gulnur Uzun
Email           : gulnuruzun20@gmail.com
Search title    : Peptide Mass Fingerprinting
MS data file    : HW2_PMF_1.txt
Database        : SwissProt 2023_05 (570420 sequences; 206321560 residues)
Taxonomy        : Homo sapiens (human) (20429 sequences)
Timestamp       : 17 Nov 2023 at 14:32:41 GMT
Top Score       : 102 for HYEP_HUMAN, Epoxide hydrolase 1 OS=Homo sapiens OX=9606 GN=EPHX1 PE=1 SV=1
```

Figure 16. Displayed the name of protein.

a) The name of protein is Epoxide hydrolase 1 (Figure 16). e-value is **1.3e-06** for Epoxide hydrolase 1 protein (Figure 17). 1.3e-06 equal to 0.0000013, the e-value of the corresponding protein is a very small number, indicating that the likelihood of the match being random is quite low. The lower the e-value, the more significant the score.

Database:	SwissProt
Score:	102
Expect:	1.3e-06
Monoisotopic mass (M_r):	52915
Calculated pI:	6.77
Taxonomy:	Homo sapiens

Figure 17. Displayed the e-value.

b) In the relevant analysis, the number of peptides significantly matched to the protein is 35.

MS data file: MM2_PMF_1.txt
Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Mass values searched: 135
Mass values matched: 35

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	M	Peptide
22 - 39	2085.9900	2084.9827	2084.9803	0.0024	1	R.DKEETLPLEDGWWGPGTR.S
24 - 39	1842.8500	1841.8427	1841.8584	-0.0157	0	K.EETLPLEDGWWGPGTR.S
44 - 52	1106.5700	1105.5627	1105.5404	0.0223	0	R.EDDSIRPFK.V
44 - 66	2795.4000	2794.3927	2794.3158	0.0769	1	R.EDDSIRPFKVETSDEEIHDLHQR.I
67 - 71	678.3000	677.2927	677.3860	-0.0933	1	R.IDKFR.F
67 - 71	679.4100	678.4027	677.3860	1.0167	1	R.IDKFR.F
92 - 98	951.5300	950.5227	950.5338	-0.0110	1	K.KVISYWR.N
93 - 98	823.4300	822.4227	822.4388	-0.0161	0	K.VISYWR.N
99 - 104	838.3700	837.3627	837.3657	-0.0030	0	R.NEFDWK.K
99 - 105	966.4800	965.4727	965.4607	0.0121	1	R.NEFDWKK.Q
106 - 112	871.4900	870.4827	870.4923	-0.0096	0	K.QVEILNR.Y
113 - 117	691.2500	690.2427	690.3489	-0.1062	0	R.YPHFK.T
160 - 168	1009.6100	1008.6027	1008.6219	-0.0192	0	K.IIPLLTDPK.N
197 - 206	1050.5900	1049.5827	1049.5618	0.0209	1	K.KGFNSVATAR.I
198 - 206	922.4900	921.4827	921.4668	0.0159	0	K.GFNSVATAR.I
198 - 210	1474.7400	1473.7327	1472.7776	0.9551	1	K.GFNSVATARIFYK.L
207 - 210	569.2900	568.2827	569.3213	-1.0386	0	R.IFYK.L
215 - 244	3337.9700	3336.9627	3337.6264	-0.6637	0	R.LGFQEFYIQGGDWGSLICTNMAQLVPSHVK.G
271 - 277	835.4600	834.4527	834.4599	-0.0072	0	R.FLGLTER.D
271 - 286	1892.0100	1891.0027	1891.0455	-0.0428	1	R.FLGLTERDVELLYPVK.E
278 - 286	1075.6000	1074.5927	1074.5961	-0.0034	0	R.DVELLYPVK.E
289 - 295	915.4700	914.4627	914.4684	-0.0057	0	K.VFYSLMR.E
329 - 338	1288.6100	1287.6027	1287.5884	0.0143	0	K.FSTWTNTEFR.Y
339 - 347	1051.5300	1050.5227	1050.4982	0.0246	0	R.YLEDGGLER.K
348 - 372	2902.5500	2901.5427	2901.4946	0.0481	1	R.KFSLDDLLTNVMLYWTTGTIISSQR.F
349 - 372	2773.6200	2772.6127	2773.3997	-0.7869	0	K.FSLDDLLTNVMLYWTTGTIISSQR.F
373 - 386	1729.7900	1728.7827	1728.8294	-0.0466	1	R.FYKENLGQGWMTQK.H
376 - 386	1292.6000	1291.5927	1290.6027	0.9901	0	K.ENLGQGWMTQK.H
392 - 411	2279.1900	2278.1827	2278.1674	0.0153	0	K.VYVPTGFSAFPFELLHTPEK.W
415 - 419	682.2900	681.2827	681.3850	-0.1023	1	R.FKYPK.L
420 - 428	1131.6000	1130.5927	1130.5794	0.0133	0	K.LISYSYMVR.G
429 - 446	1999.9900	1998.9827	1998.9799	0.0028	0	R.GGHFAAFEEPELLAQDIR.K
429 - 447	2128.0800	2127.0727	2127.0749	-0.0022	1	R.GGHFAAFEEPELLAQDIRK.F
448 - 454	863.4900	862.4827	862.4912	-0.0085	0	K.FLSVLER.Q
448 - 455	991.5700	990.5627	990.5498	0.0129	1	K.FLSVLERQ.-

Figure 18. Displayed the number of peptides significantly matched to the protein.

c) We displayed the amino acid sequence of the protein along with the matched peptides on the analysis page, highlighted in red. Protein sequence coverage is 58% for this sequence.

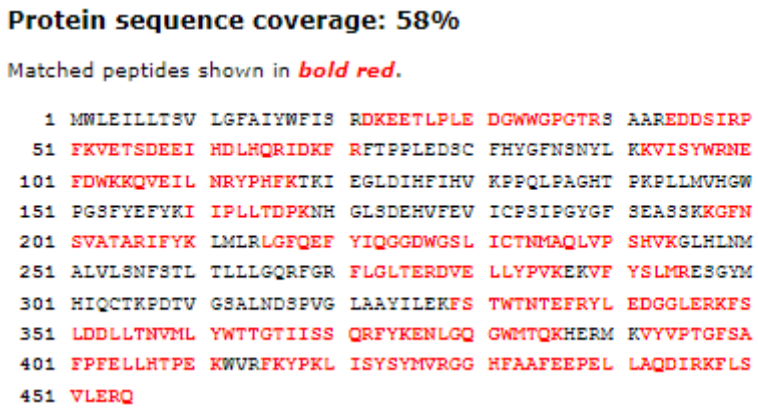


Figure 19. Protein Sequence Coverage.

Question 3: Transcriptomic Data Analysis (Details in Excel file.)

Firstly, the t-test and fold change calculations have been performed in Excel using formulas.

For example IDP2 gene:

T.TEST(C6:J6; L6:S6; 2; 2)

#Explain the calculatio;

C6:J6 is aerobik range; L6:S6 is anaerobik range; 2: Specifies the use of an independent two-sample t-test; 2: Indicates that the variances of both samples are assumed to be equal.

For Fold Change (FC), averages for aerobic and anaerobic conditions were separately obtained, and then the FC was defined as the ratio of the aerobic mean to the anaerobic mean.

Typically, p-values less than 0.05 are considered statistically significant. The p-values and fold-change values have been obtained from the table provided to us. Below are the genes with the smallest p-values:

- For the IDP2 gene, the p-value is 1.01E-11, which is equal to 0.00000000000101246.
- For the PFK1 gene, the p-value is 2.03E-07, which is equal to 0.000000202763.
- For the CIT3 gene, the p-value is 2.02E-06, which is equal to 0.00000201516.
- The top 3 genes with the highest fold-change (FC) values are as follows:
- For the IDP2 gene, the FC value is 40.3443.
- For the CIT3 gene, the FC value is 22.36661.
- For the KGD2 gene, the FC value is 2.341349.

Based on these results, the data analysis and interpretation can be summarized as follows:

1. **IDP2 Gene:**
    - P-value: 1.01E-11 (very small)
    - FC: 40.3443 (high)

    The p-value for this gene is very small, indicating statistical significance. Additionally, the high FC value suggests a substantial increase in gene expression. IDP2 has a higher FC compared to other genes, indicating a potentially significant role in biological processes.

2. **PFK1 Gene:**
    - P-value: 2.03E-07 (very small)
    - FC: 2.03E-07 (low)

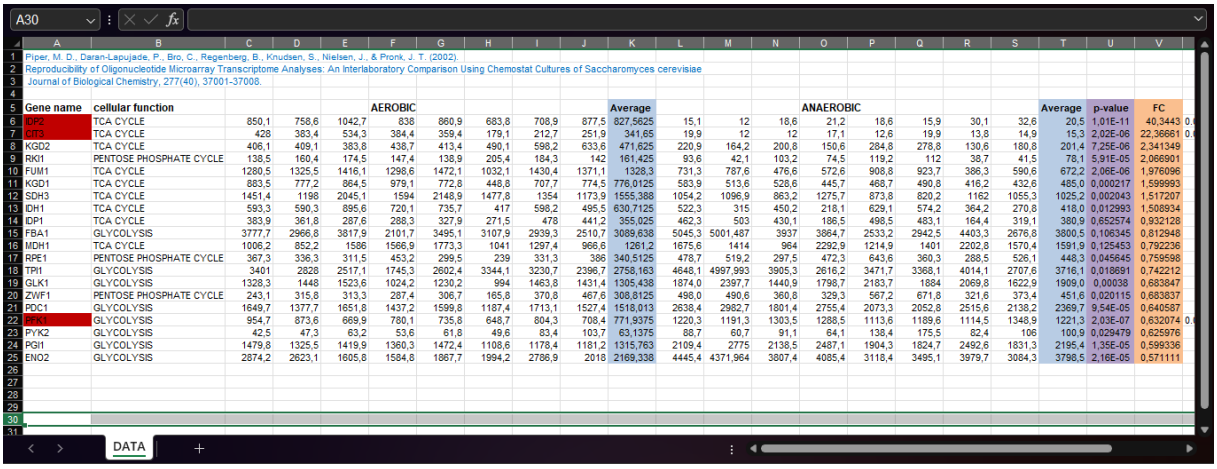
The p-value for PFK1 is statistically significant. However, the low FC value suggests a more limited change in gene expression, potentially resulting in a lower biological impact compared to other genes.

3. **CIT3 Gene:**
    - P-value: 2.02E-06 (very small)
    - FC: 22.36661 (high)

The small p-value and high FC value indicate a significant increase in gene expression for CIT3. This gene may play a crucial role in biological processes due to the noteworthy increase in expression.

4. **KGD2 Gene:**
    - P-value: Not provided
    - FC: 2.341349 (moderate)

Although the p-value is not specified, the moderate FC value suggests a noticeable change in gene expression, albeit to a lesser extent compared to other genes.



Piper, M. D., Daran-Lapujade, P., Bro, C., Regenberg, B., Knudsen, S., Nielsen, J., & Pronk, J. T. (2002).
Reproducibility of Oligonucleotide Microarray Transcriptome Analyses: An Interlaboratory Comparison Using Chemostat Cultures of Saccharomyces cerevisiae
Journal of Biological Chemistry, 277(40), 37001-37008.

Gene name	cellular function	AEROBIC								Average	ANAEROBIC							Average	p-value	FC
IDP2	TCA CYCLE	850,1	758,6	1042,7	838	860,9	683,8	708,9	877,5	827,5625	15,1	12	18,6	21,2	18,6	15,9	30,1	32,6	20,5	1,01E-11	40,3443
CIT3	TCA CYCLE	428	383,4	534,3	384,4	359,4	179,1	212,7	251,9	341,65	19,9	12	12	17,1	12,6	19,9	13,8	14,9	15,3	2,02E-06	22,36661
KGD2	TCA CYCLE	406,1	409,1	383,8	438,7	413,4	490,1	598,2	633,6	471,625	220,9	164,2	200,8	150,6	284,8	278,8	130,6	180,8	201,4	7,25E-06	2,341349
RKI1	PENTOSE PHOSPHATE CYCLE	138,5	160,4	174,5	147,4	138,9	205,4	184,3	142	161,425	93,6	42,1	103,2	74,5	119,2	112	38,7	41,5	78,1	5,91E-05	2,066901
FUM1	TCA CYCLE	1280,5	1325,5	1416,1	1298,6	1472,1	1032,1	1430,4	1371,1	1328,3	731,3	787,6	476,6	572,6	908,8	923,7	386,3	590,6	672,2	2,06E-06	1,976096
KGD1	TCA CYCLE	883,5	777,2	864,5	979,1	772,8	448,8	707,7	774,5	776,0125	583,9	513,6	528,6	445,7	468,7	490,8	416,2	432,6	485,0	0,000217	1,599993
SDH3	TCA CYCLE	1451,4	1198	2045,1	1594	2148,9	1477,8	1354	1173,9	1555,388	1054,2	1096,9	863,2	1275,7	873,8	820,2	1162	1055,3	1025,2	0,002043	1,517207
IDH1	TCA CYCLE	593,3	590,3	895,6	720,1	735,7	417	598,2	495,5	630,7125	522,3	315	450,2	218,1	629,1	574,2	364,2	270,8	418,0	0,012993	1,508934
IDP1	TCA CYCLE	383,9	361,8	287,6	288,3	327,9	271,5	478	441,2	355,025	462,3	503	430,1	186,5	498,5	483,1	164,4	319,1	380,9	0,652574	0,932128
FBA1	GLYCOLYSIS	3777,7	2966,8	3817,9	2101,7	3495,1	3107,9	2939,3	2510,7	3089,638	5045,3	5001,487	3937	3864,7	2533,2	2942,5	4403,3	2676,8	3800,5	0,106345	0,812948
MDH1	TCA CYCLE	1006,2	852,2	1586	1566,9	1773,3	1041	1297,4	966,6	1261,2	1675,6	1414	964	2292,9	1214,9	1401	2202,8	1570,4	1591,9	0,125453	0,792236
RPE1	PENTOSE PHOSPHATE CYCLE	367,3	336,3	311,5	453,2	299,5	239	331,3	386	340,5125	478,7	519,2	297,5	472,3	643,6	360,3	288,5	526,1	448,3	0,045645	0,759598
TPI1	GLYCOLYSIS	3401	2828	2517,1	1745,3	2602,4	3344,1	3230,7	2396,7	2758,163	4648,1	4997,993	3905,3	2616,2	3471,7	3368,1	4014,1	2707,6	3716,1	0,018691	0,742212
GLK1	GLYCOLYSIS	1328,3	1448	1523,6	1024,2	1230,2	994	1463,8	1431,4	1305,438	1874,0	2397,7	1440,9	1798,7	2183,7	1884	2069,8	1622,9	1909,0	0,00038	0,683847
ZWF1	PENTOSE PHOSPHATE CYCLE	243,1	315,8	313,3	287,4	306,7	165,8	370,8	467,6	308,8125	498,0	490,6	360,8	329,3	567,2	671,8	321,6	373,4	451,6	0,020115	0,683837
PDC1	GLYCOLYSIS	1649,7	1377,7	1651,8	1437,2	1599,8	1187,4	1713,1	1527,4	1518,013	2638,4	2982,7	1801,4	2755,4	2073,3	2052,8	2515,6	2138,2	2369,7	9,54E-05	0,640587
PFK1	GLYCOLYSIS	954,7	873,6	669,9	780,1	735,8	648,7	804,3	708,4	771,9375	1220,3	1191,3	1303,5	1288,5	1113,6	1189,6	1114,5	1348,9	1221,3	2,03E-07	0,632074
PYK2	GLYCOLYSIS	42,5	47,3	63,2	53,6	61,8	49,6	83,4	103,7	63,1375	88,7	60,7	91,1	64,1	138,4	175,5	82,4	106	100,9	0,029479	0,625976
PGI1	GLYCOLYSIS	1479,8	1325,5	1419,9	1360,3	1472,4	1108,6	1178,4	1181,2	1315,763	2109,4	2775	2138,5	2487,1	1904,3	1824,7	2492,6	1831,3	2195,4	1,35E-05	0,599336
ENO2	GLYCOLYSIS	2874,2	2623,1	1605,8	1584,8	1867,7	1994,2	2786,9	2018	2169,338	4445,4	4371,964	3807,4	4085,4	3118,4	3495,1	3979,7	3084,3	3798,5	2,16E-05	0,571111

Figure 20. Displayed Excel File.

Question 4: The answer to this question is found in the .R file.

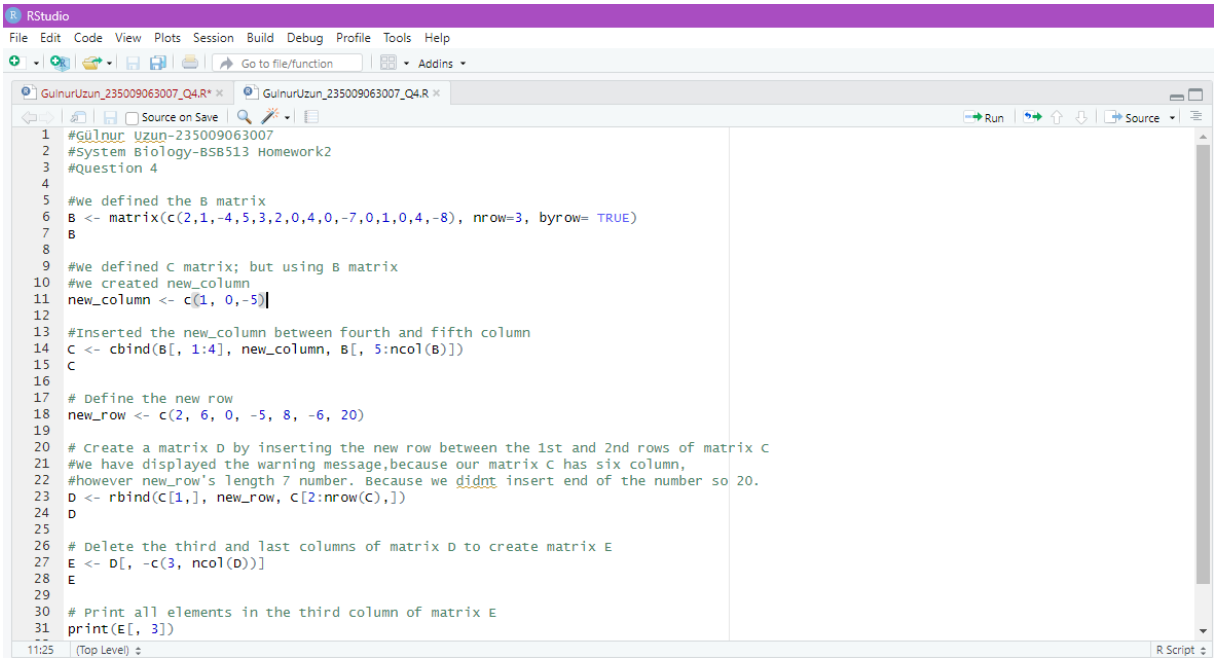
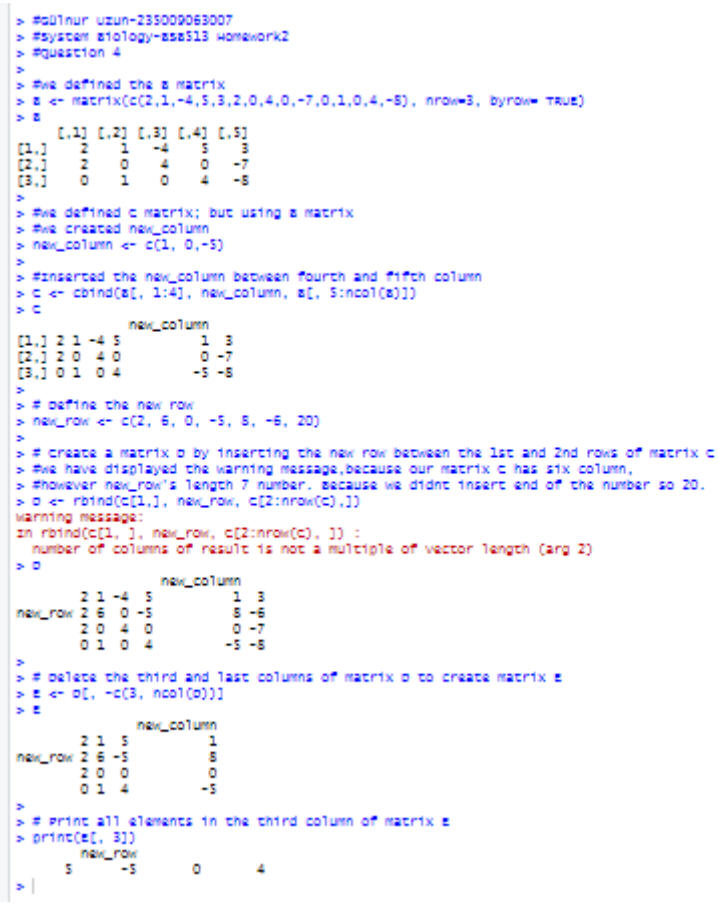


Figure 21. R Script for Question 4.



```
> #gulnur uzun-235009063007
> #system biology-bbb513 Homework2
> #question 4
>
> #we defined the a matrix
> a <- matrix(c(2,1,-4,5,3,2,0,4,0,-7,0,1,0,4,-8), nrow=3, byrow= TRUE)
> a
     [,1] [,2] [,3] [,4] [,5]
[1,]    2    1   -4    5    3
[2,]    2    0    4    0   -7
[3,]    0    1    0    4   -8
>
> #we defined c matrix; but using a matrix
> #we created new_column
> new_column <- c(1, 0,-5)
>
> #inserted the new_column between fourth and fifth column
> c <- cbind(a[, 1:4], new_column, a[, 5:ncol(a)])
> c
             new_column
[1,] 2 1 -4 5          1  3
[2,] 2 0  4 0          0 -7
[3,] 0 1  0 4         -5 -8
>
> # define the new row
> new_row <- c(2, 6, 0, -5, 8, -6, 20)
>
> # create a matrix d by inserting the new row between the 1st and 2nd rows of matrix c
> #we have displayed the warning message,because our matrix c has six column,
> #however new_row's length 7 number. because we didnt insert end of the number so 20.
> d <- rbind(c[1,], new_row, c[2:nrow(c),])
Warning message:
In rbind(c[1, ], new_row, c[2:nrow(c), ]) :
  number of columns of result is not a multiple of vector length (arg 2)
> d
                new_column
       2 1 -4  5          1  3
new_row 2 6  0 -5          8 -6
       2 0  4  0          0 -7
       0 1  0  4         -5 -8
>
> # delete the third and last columns of matrix d to create matrix e
> e <- d[, -c(3, ncol(d))]
> e
              new_column
       2 1  5          1
new_row 2 6 -5          8
       2 0  0          0
       0 1  4         -5
>
> # Print all elements in the third column of matrix e
> print(e[, 3])
       new_row
     5      -5       0       4
>
```

Figure 22. Outputs for R Script (Question 4).