Identification of VUS Across the ATP7B Gene

Deniz Saltık 29381
Durmuş Erdem Kertmen 28071
Göktuğ Gençkaya 27888
Metin Alperen Balta 28231

Contents

>> Wilson’s Disease

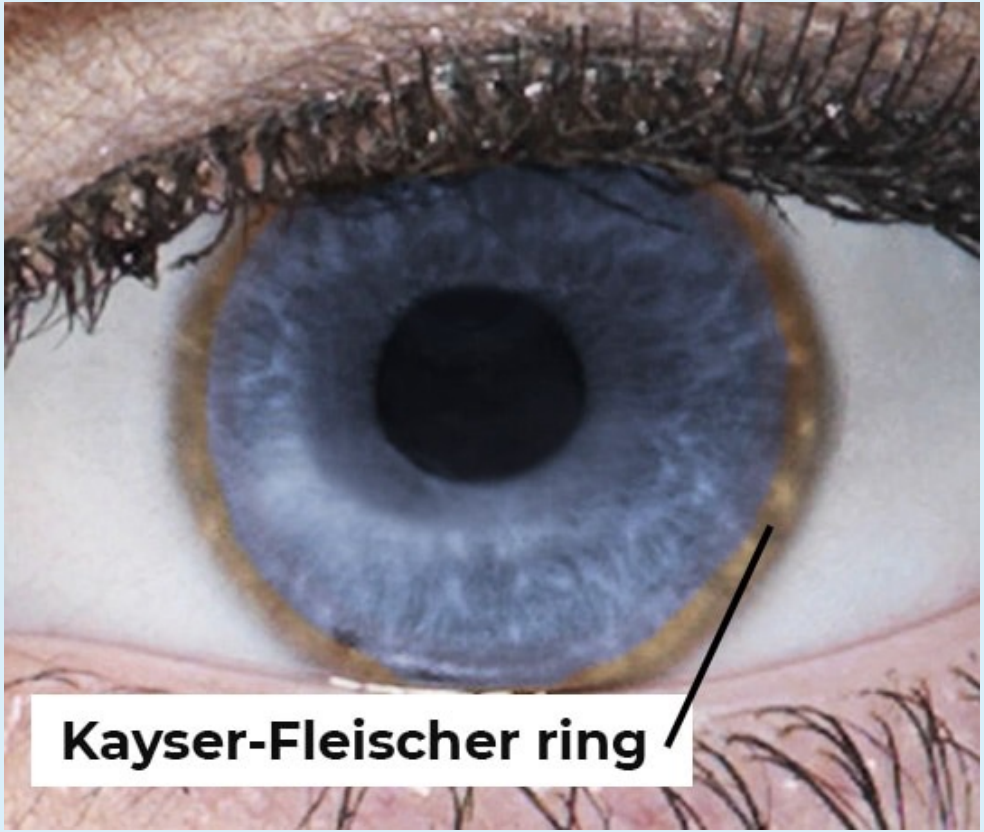
>> Copper-transporting p-type ATPase 2, ATP7B

>> Methods

>> Results

- Conservation score graph
- Identification of Vus

>> Discussion



Wilson’s Disease

>> Wilson's disease is an autosomal recessive inherited disorder of hepatic copper metabolism resulting in the accumulation of copper in many organs and tissues. (Ferenci P., 2004)

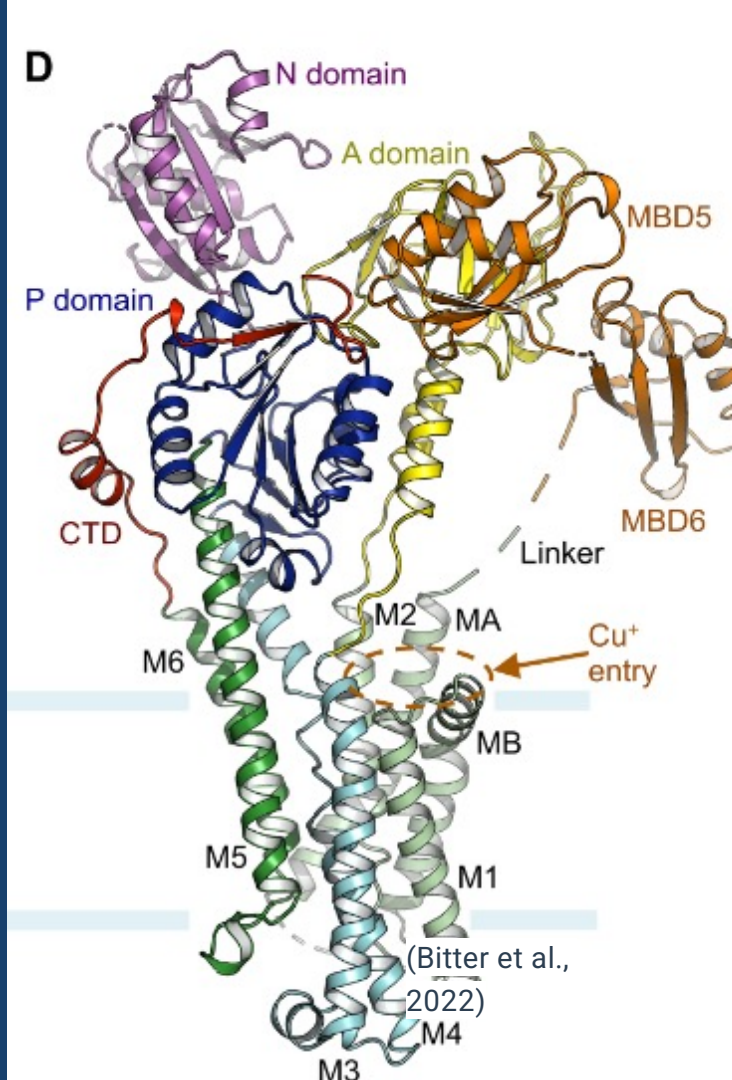
- Neurological symptoms
- Liver malfunctions
- Kayser-Fleischer rings

Copper-transporting p-type ATPase 2, ATP7B

>> Copper-transporting p-type ATPase 2 is mostly found in liver.

>> It's main purpose is to transport copper from the liver to other parts of the body and to remove excess copper from the cells.

>> It's deficiency causes Wilson's disease.



(Bitter et al., 2022)

Methods

1- Research

2- Receiving the protein sequence from UniProt

3- Finding Homologous sequences using BlastP

4- Alignment of the homologous sequences using MEGA11

5- Constructing the Phylogenetic Tree using MEGA11

6- Rerooting the Phylogenetic Tree using FigTree

7- Calculating conservation scores by Python

8- Retrieving mutations from ClinVar

9- Identifying Vus by Python

Research

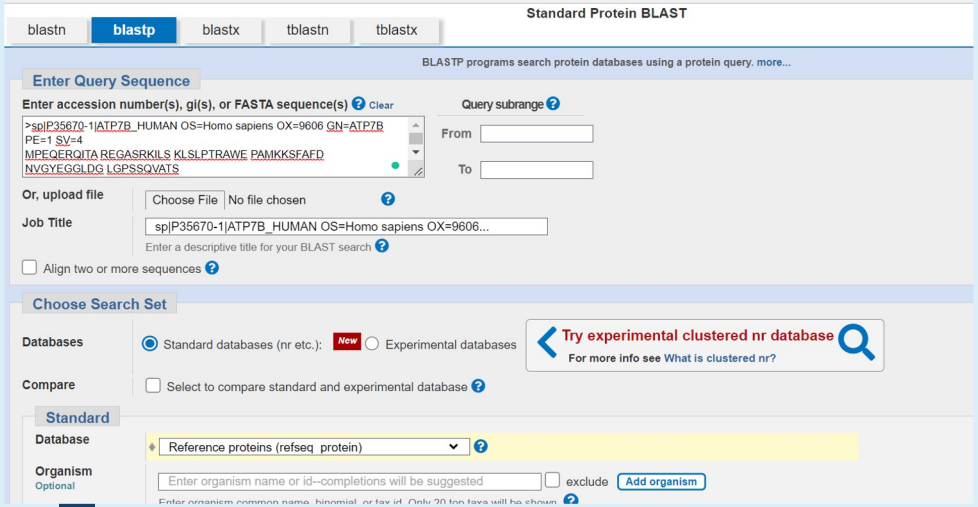
>> From Google Scholar, Wilson's disease is chosen

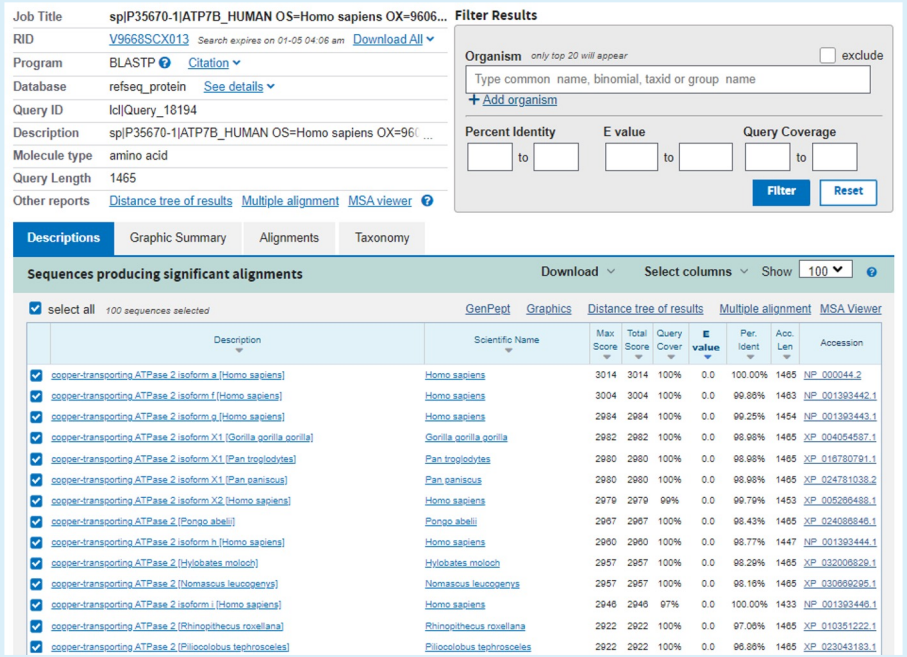
>> The protein (Copper-Transporting p-type ATPase 2,ATP7B) which causes Wilson's disease to occur is identified.

>> The protein sequence for Copper-Transporting p-type ATPase 2 is retrieved from UniProt.

Finding Homologous Sequences

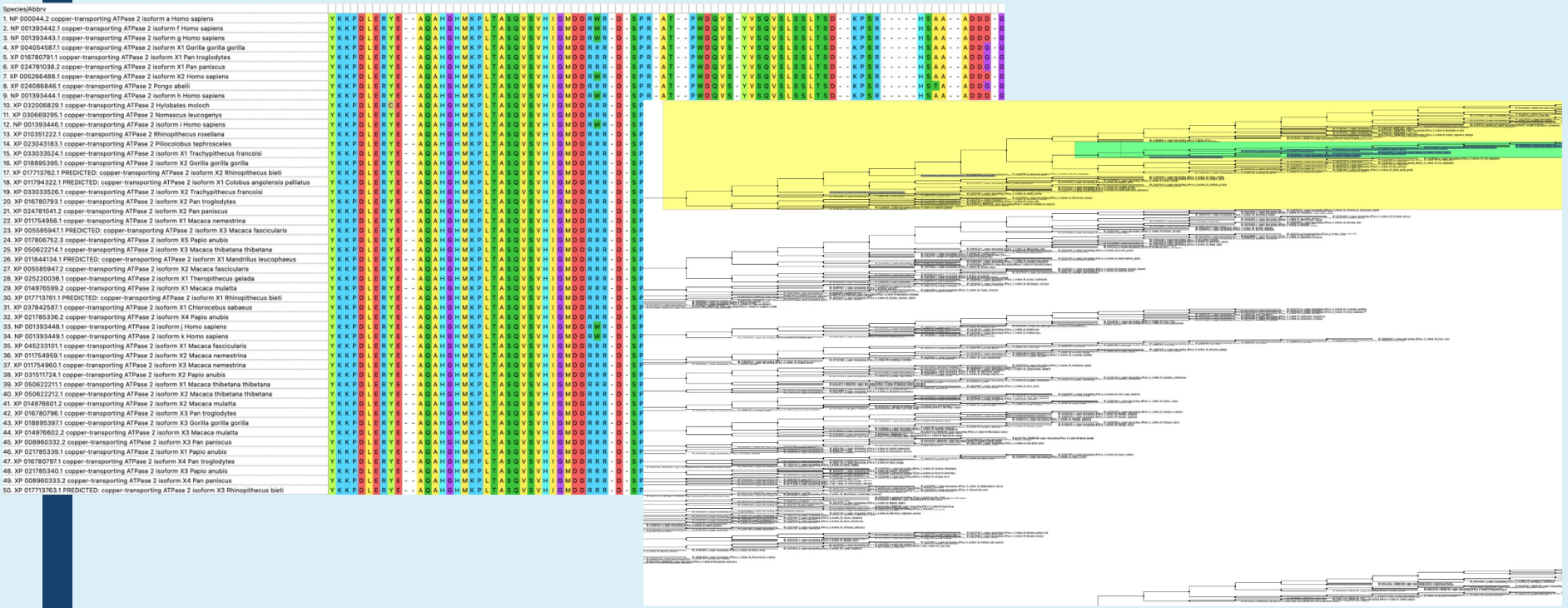
>> By using BlastP





Alignment and Phylogenetic Tree

>> By using MEGA11 and FigTree

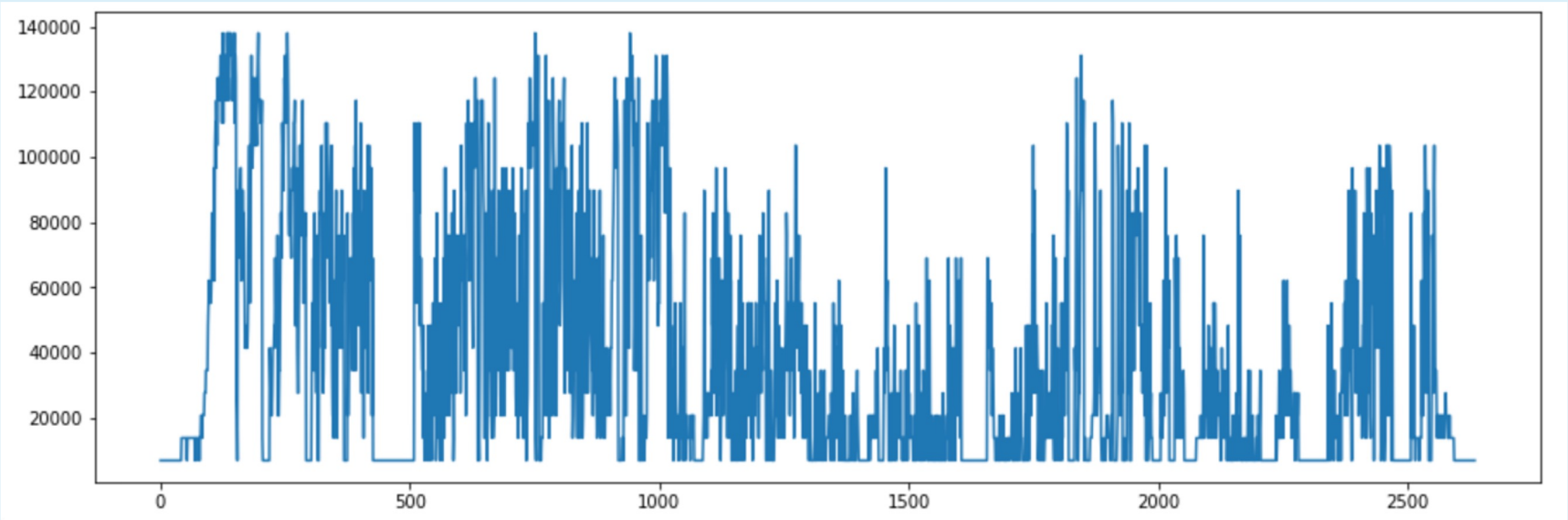


Calculating the Conservation Scores

>> By using the Python Code

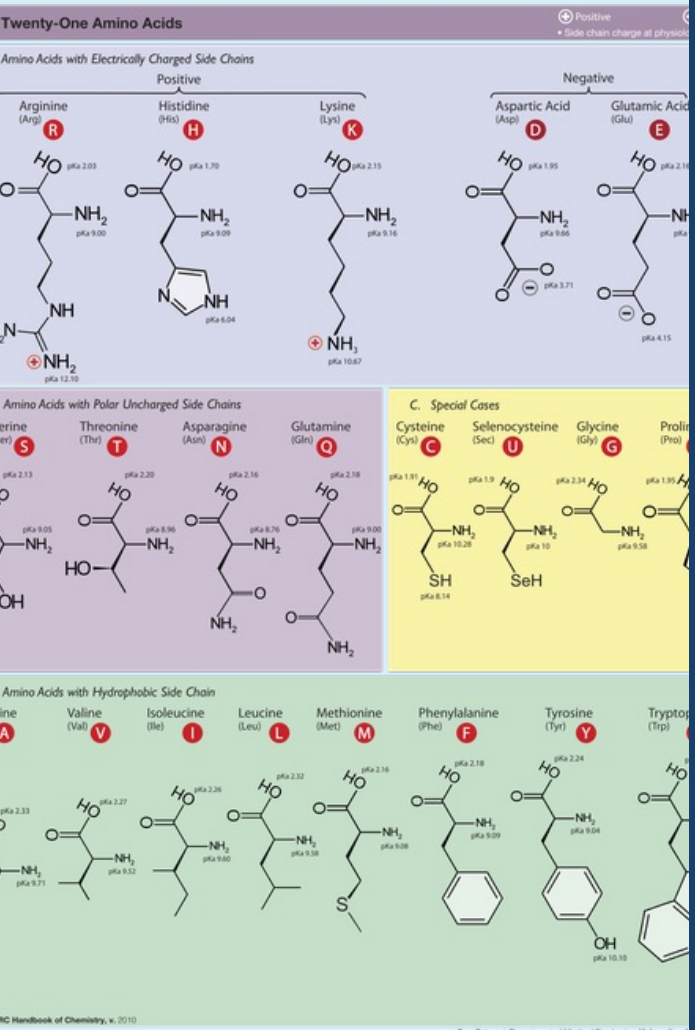
```
matrixOfFrequencies = [] #there will be 20 columns (for amino acids) and position many rows

lengthOfMSA = len((list(MSA_dict.values()))[0])
dictLen = len(MSA_dict)

eMSA_dict = {}
count = 0
for key,value in MSA_dict.items():
    eMSA_dict[count]= value
    count+=1

numberOfiad = []
for pos in range(lengthOfMSA):
    for key, value in aa_dict.items():
        aa_dict[key] = 0
    aaDictForPos = aa_dict
    totalBases = 0
    iadForPos = 0
    frequencyArray = []
    for i in range(dictLen):
        if eMSA_dict[i][pos] != "-":
            aa = eMSA_dict[i][pos]
            if aa == "X":
                aaDictForPos["A"]+=1
            else:
                aaDictForPos[aa]+=1
            totalBases+=1
        else:
            iadForPos +=1
```

Conservation score graph



Identification of VUS

Name	Mutation	Amino acid #	Protein Change	Significance	Polarity/Charge	with gaps	nucleotide
Tyr1464Ser	Y-S	1464	Y1386S, Y1353S	No	P_P	2555	4391A>C
Gln1463Arg	Q-R	1463	Q1379R, Q1256	No	P_+	2554	4388A>G
Gln1463Ter		1463	Q1463*, Q1256*	No		2554	4387C>T
Asp1460Gly	D-G	1460	D1253G, D1376	No	-_N	2543	4379A>G
Leu1454Pro	L-P	1454	L1454P, L1343P	No	N_N	2534	4361T>C
Asp1450Asn	D-N	1450	D1020N, D1169N	No	-_P	2471	4383G>A
Ala1443Ser	A-S	1443	A1236S, A1332S	No	N_P	2461	4327G>T
Arg1440Gln	R-Q	1440	R1233Q, R1362	No	\+_P	2453	4319G>A
Arg1440Trp	R-W	1440	R1356W, R1362	No	\+_N	2453	4318C>T
Ser1439Pro	S-P	1439	S1328P, S1355P	No	P_N	2452	4315T>C
Ser1432Phe	S-F	1432	S1225F, S1321F	No	P_N	2443	4295C>T
Leu1430Met	L-M	1430	L1223M, L1346M	No	N_N	2441	4288C>A
Ser1429Leu	S-L	1429	S1429L, S1222L	No	P_N	2440	4286C>T
Tyr1424Cys	Y-C	1424	Y1340C, Y1313C	No	P_N	2435	4271A>G
Ser1423Asn	S-N	1423	S1216N, S1345N	No	P_P	2433	4286G>A
Arg1411Gln	R-Q	1411	R1300Q, R1411Q	No	\+_P	2416	4232G>A
Ser1398Thr	S-T	1398	S1191T, S1287T	Yes	P_P	2403	4192T>A
Pro1394Leu	P-L	1394	P1187L, P1283L	No	N_N	2399	4181C>T
Met1392Thr	M-T	1392	M1185T, M1281T	No	N_P	2397	4175T>C
His1389Arg	H_R	1389	H1182R, H1311R	No	\+_+	2394	4166A>G
Ala1388Val	A-V	1388	A1388V, A1304V	No	N_N	2393	4163C>T
Leu1381Val	L-V	1381	L1100V, L1154V,	No	N_N	2384	4141C>G
Tyr1376Ser	Y-S	1376	Y1376S, Y1169S	Yes	P_P	2379	4127A>C
Cys1375Ser	C-S	1375	C1375S, C1168S	No	N_P	2378	4124G>C
Leu1373Pro	L-P	1373	L1373P, L1289P	No	N_N	2376	4118T>C
Gln1372Lys	Q-K	1372	Q1091K, Q1145K	No	P_+	2375	4114C>A
Gln1372Glu	Q-E	1372	Q1261E, Q1288E	No	P_-	2375	4114C>G
Leu1371Arg	L-R	1371	L1164R, L1371R	Yes	N_+	2374	4112T>G
Ser1369Leu	S-L	1369	S1369L, S1258L	Yes	P_N	2372	4106C>T
Leu1368Val	L-V	1368	L1368V, L1161V	No	N_N	2371	4102C>G
Val1364Leu	V-L	1364	V1364L, V1157L	Yes	N_N	2367	4090G>C
Met1359Ile	M-I	1359	M1152I, M1275I,	Yes	N_N	2362	4077G>T
Ala1357Val	A-V	1357	A1246V, A1273V	Yes	N_N	2360	4070C>T
Ser1356Pro	S-P	1356	S1356P, S1149P	Yes	P_N	2359	4066T>C
Gly1355Asp	G-D	1355	G1355D, G1271	Yes	N_-	2358	4064G>A
Gly1355Val	G-V	1355	G1355V, G1244V	Yes	N_N	2358	4064G>T
Gly1355Ser	G-S	1355	G1355S, G1148S	Yes	N_P	2358	4064G>A
Trp1353Arg	W-R	1353	W1146R, W1275	Yes	N_+	2356	4057T>C
Val1349Leu	V-L	1349	V1349L, V1271L	No	N_N	2352	4045G>C
Met1344Leu	M-L	1344	M1137L, M1233L	No	N_N	2347	4030A>C



DISCUSSION

- Changes between amino acids with same polarities may not change the protein structure/function, even though the conservation score is high.
- Number and relevance of sequences may cause overfitting/underfitting for classifying the VUS.
- Change in the threshold for conservation score will result in differences.

References

- Ferenci, P. (2004). diagnosis and current therapy of Wilson's disease. *Alimentary pharmacology & therapeutics*, *19*(2), 157-165. https://doi.org/10.1046/j.1365-2036.2003.01813.x
- *Kayser-Fleischer rings*. (2018, October). National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/liver-disease/wilson-disease/symptoms-causes
- Bitter, R. M., Oh, S., Deng, Z., Rahman, S., Hite, R. K., & Yuan, P. (2022). Structure of the Wilson disease copper transporter ATP7B. *Science Advances*, *8*(9). https://doi.org/10.1126/sciadv.abl5508
- *Amino Acid Table*. (n.d.). Rosalind. https://rosalind.info/glossary/amino-acid/

THANK YOU FOR LISTENING !!!