

# 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)

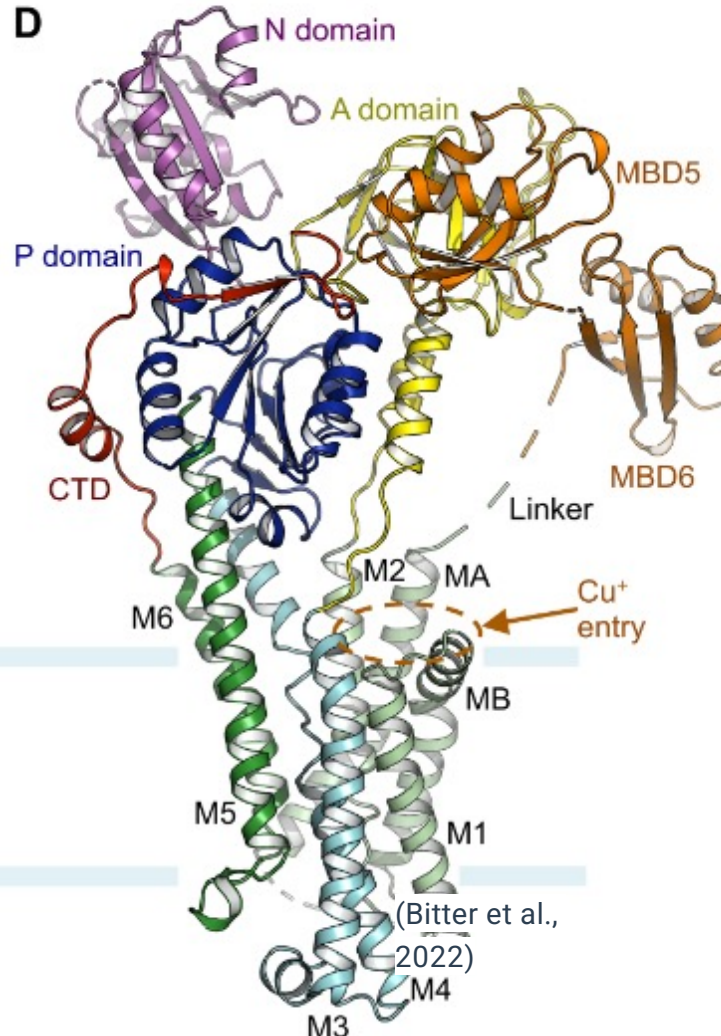
- Neurological symptoms
- Liver malfunctions
- Kayser-Fleischer rings



**Kayser-Fleischer ring**

# 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.



# 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

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

Enter Query Sequence

BLASTP programs search protein databases using a protein query. more...

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) Clear

Query subrange [?](#)

>sp|P35670-1|ATP7B\_HUMAN OS=Homo sapiens OX=9606 GN=ATP7B  
PE=1 SV=4  
MPGEQRIQTA REGASRKILS KLSLPTRAWE PAMKKSFAFD  
NVGYEGGLDG LGPSSQVATS

From

To

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Databases ☒ Standard databases (nr etc.): [New](#) ☐ Experimental databases [Try experimental clustered nr database](#) [?](#)  
For more info see [What is clustered nr?](#)

Compare ☐ Select to compare standard and experimental database [?](#)

Standard

Database  [?](#)

Organism  ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 taxa will be chosen [?](#)

Job Title sp|P35670-1|ATP7B\_HUMAN OS=Homo sapiens OX=9606...  
RID Y9668SCX013 Search expires on 01-05 04:06 am [Download All](#) [?](#)  
Program BLASTP [?](#) [Citation](#) [?](#)  
Database refseq\_protein [See details](#) [?](#)  
Query ID IclQuery\_18194  
Description sp|P35670-1|ATP7B\_HUMAN OS=Homo sapiens OX=9606...  
Molecule type amino acid  
Query Length 1465  
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude  
Type common name, binomial, taxid or group name  
[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to   
[Filter](#) [Reset](#)

**Descriptions** [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments Download Select columns Show 100 [?](#)

☒ select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform a (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	3014	3014	100%	0.0	100.00%	1465	<a href="#">NP_000044.2</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform f (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	3004	3004	100%	0.0	99.98%	1463	<a href="#">NP_001393442.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform g (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	2984	2984	100%	0.0	99.25%	1454	<a href="#">NP_001393443.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform X1 (Gorilla gorilla gorilla)</a>	<a href="#">Gorilla gorilla gorilla</a>	2982	2982	100%	0.0	99.98%	1465	<a href="#">XP_004054587.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform X1 (Pan troglodytes)</a>	<a href="#">Pan troglodytes</a>	2980	2980	100%	0.0	99.98%	1465	<a href="#">XP_016780791.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform X1 (Pan caninus)</a>	<a href="#">Pan caninus</a>	2980	2980	100%	0.0	99.98%	1465	<a href="#">XP_024781038.2</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform X2 (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	2976	2976	99%	0.0	99.79%	1453	<a href="#">XP_005266488.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 (Pongo abelii)</a>	<a href="#">Pongo abelii</a>	2957	2957	100%	0.0	98.43%	1465	<a href="#">XP_02408846.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform h (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	2960	2960	100%	0.0	99.77%	1447	<a href="#">NP_001393444.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 (Hylobates moloch)</a>	<a href="#">Hylobates moloch</a>	2957	2957	100%	0.0	98.29%	1465	<a href="#">XP_032008829.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 (Nomascus leucogenus)</a>	<a href="#">Nomascus leucogenus</a>	2957	2957	100%	0.0	99.16%	1465	<a href="#">XP_030692095.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 isoform l (Homo sapiens)</a>	<a href="#">Homo sapiens</a>	2946	2946	97%	0.0	100.00%	1433	<a href="#">NP_001393446.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 (Rhinopithecus roosei)</a>	<a href="#">Rhinopithecus roosei</a>	2922	2922	100%	0.0	97.06%	1465	<a href="#">XP_010361222.1</a>
<input checked="" type="checkbox"/>	<a href="#">cgoecr-transferring ATPase 2 (Ptilinopus techroceles)</a>	<a href="#">Ptilinopus techroceles</a>	2922	2922	100%	0.0	99.98%	1465	<a href="#">XP_023043183.1</a>

# Alignment and Phylogenetic Tree

>> By using MEGA11 and FigTree

Species/Abbrev	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
1. NP 000044.2 copper-transporting ATPase 2 isoform a Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
2. NP 001393442.1 copper-transporting ATPase 2 isoform 1 Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
3. NP 001393443.1 copper-transporting ATPase 2 isoform g Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
4. XP 004045487.1 copper-transporting ATPase 2 isoform X1 Gorilla gorilla gorilla	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
5. XP 016780791.1 copper-transporting ATPase 2 isoform X1 Pan troglodytes	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
6. XP 024781038.2 copper-transporting ATPase 2 isoform X1 Pan paniscus	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
7. XP 005266488.1 copper-transporting ATPase 2 isoform X2 Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
8. XP 024068446.1 copper-transporting ATPase 2 Pongo abelii	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
9. NP 001393444.1 copper-transporting ATPase 2 isoform 1 Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
10. XP 032006829.1 copper-transporting ATPase 2 Hylobates moloch	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
11. XP 03068295.1 copper-transporting ATPase 2 Nomascus leucogenys	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
12. NP 001393446.1 copper-transporting ATPase 2 isoform 1 Homo sapiens	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
13. XP 010331221.1 copper-transporting ATPase 2 Rhinopithecus roosei	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
14. XP 023043183.1 copper-transporting ATPase 2 Ptilocotylus lephrosceles	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
15. XP 033033024.1 copper-transporting ATPase 2 isoform X1 Trachypitecus francoisi	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
16. XP 018895395.1 copper-transporting ATPase 2 isoform X2 Gorilla gorilla gorilla	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
17. XP 017713762.1 PREDICTED: copper-transporting ATPase 2 isoform X2 Rhinopithecus bieti	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
18. XP 011784322.1 PREDICTED: copper-transporting ATPase 2 isoform X1 Colobus angolensis palliatus	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
19. XP 033033261.1 copper-transporting ATPase 2 isoform X2 Trachypitecus francoisi	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
20. XP 016780793.1 copper-transporting ATPase 2 isoform X2 Pan troglodytes	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
21. XP 024781041.2 copper-transporting ATPase 2 isoform X2 Pan paniscus	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
22. XP 011754956.1 copper-transporting ATPase 2 isoform X1 Macaca nemestrina	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
23. XP 00585947.1 PREDICTED: copper-transporting ATPase 2 isoform X3 Macaca fascicularis	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
24. XP 017067062.1 copper-transporting ATPase 2 isoform X5 Papio anubis	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
25. XP 050622214.1 copper-transporting ATPase 2 isoform X3 Macaca thibetana thibetana	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
26. XP 018444134.1 PREDICTED: copper-transporting ATPase 2 isoform X1 Mandrillus leucophaeus	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
27. XP 005859472.1 copper-transporting ATPase 2 isoform X2 Macaca fascicularis	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
28. XP 032520038.1 copper-transporting ATPase 2 isoform X1 Theropithecus gelada	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
29. XP 014876599.2 copper-transporting ATPase 2 isoform X1 Macaca mulatta	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
30. XP 017713761.1 PREDICTED: copper-transporting ATPase 2 isoform X1 Rhinopithecus bieti	Y	K	K	P	D	L	R	E	Y	E	-	A	A	H	G	M	K	P	L	A	S	V	V	H	G	M	D	R	R	-	D	P	R	-	A	-	P	W	Q	V <td>-</td> <td>Y</td> <td>V</td> <td>S</td> <td>Q</td> <td>V</td> <td>L</td> <td>S</td> <td>L</td> <td>S</td> <td>D</td> <td>-</td> <td>K</td> <td>P</td> <td>S</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>H</td> <td>A</td> <td>A</td> <td>-</td> <td>A</td> <td>D</td> <td>D</td> <td>-</td> <td>D</td>	-	Y	V	S	Q	V	L	S	L	S	D	-	K	P	S	-	-	-	-	-	H	A	A	-	A	D	D	-	D
31. XP 037842457.1 copper-transporting ATPase 2 isoform X1 Chlorocebus sabaeus	Y	K	K	P	D	L	R	E	Y	E	-	A	A</																																																								

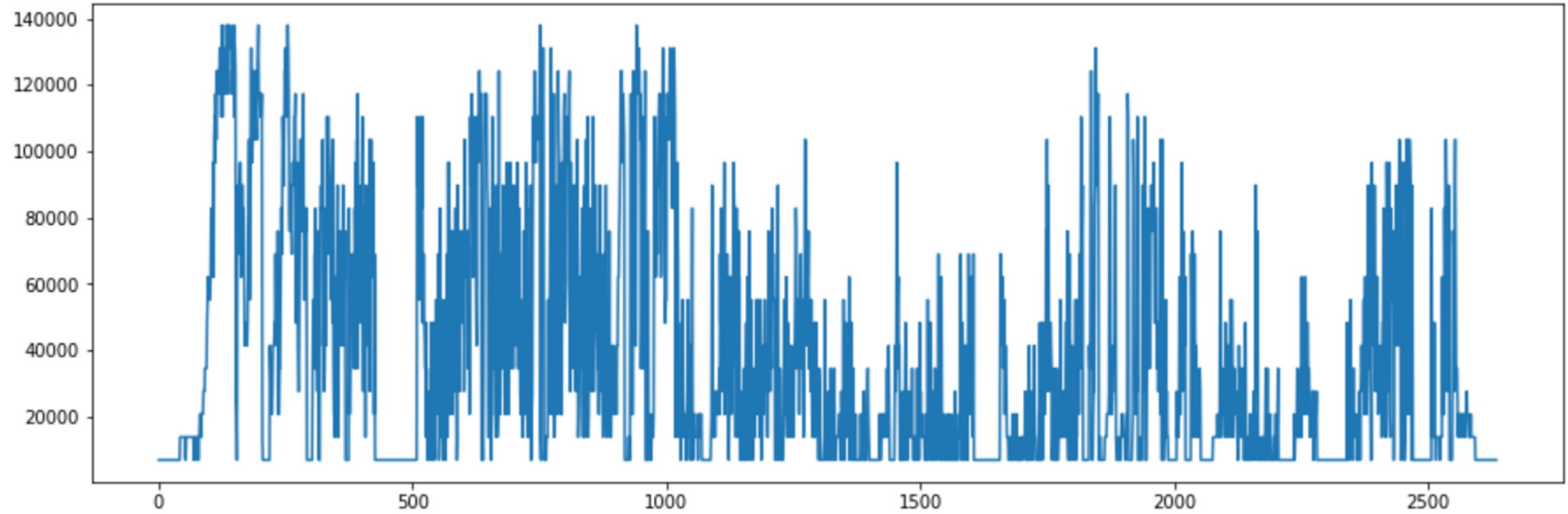


# Calculating the Conservation Scores

>> By using the Python Code

```
61 matrixOfFrequencies = [] #there will be 20 columns (for amino acids) and position many rows
62
63 lengthOfMSA = len((list(MSA_dict.values()))[0])
64 dictlen = len(MSA_dict)
65
66 eMSA_dict = {}
67 count = 0
68 for key,value in MSA_dict.items():
69     eMSA_dict[count]= value
70     count+=1
71
72 numberOfIad = []
73 for pos in range(lengthOfMSA):
74     for key, value in aa_dict.items():
75         aa_dict[key] = 0
76     aaDictForPos = aa_dict
77     totalBases = 0
78     iadForPos = 0
79     frequencyArray = []
80     for i in range(dictlen):
81         if eMSA_dict[i][pos] != "-":
82             aa = eMSA_dict[i][pos]
83             if aa == "X":
84                 aaDictForPos["A"]+=1
85             else:
86                 aaDictForPos[aa]+=1
87             totalBases+=1
88         else:
89             iadForPos +=1
```

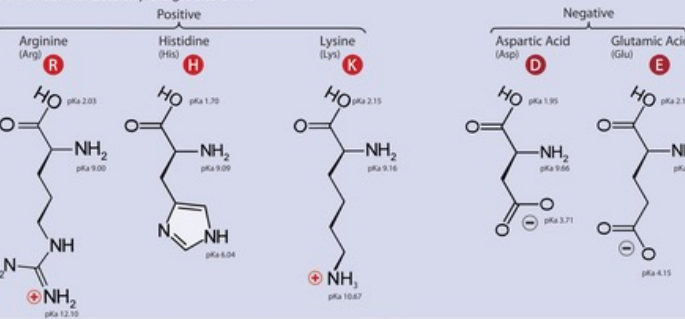
# Conservation score graph



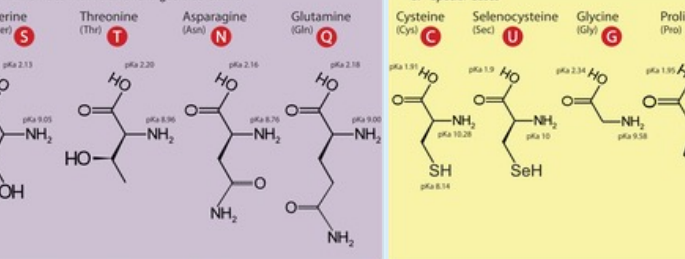
# Identification of VUS

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	P_-	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, D1169 No	-_P	2471	4383G>A
Ala1443Ser	A-S	1443	A1236S, A1332S No	N_P	2461	4327G>T
Arg1440Gln	R-Q	1440	R1233Q, R1362 No	W_P	2453	4319G>A
Arg1440Trp	R-W	1440	R1356W, R1362 No	W_N	2453	4318C>T
Ser1439Pro	S-P	1439	S1328P, S1355F 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, S1345I No	P_P	2433	4286G>A
Arg1411Gln	R-Q	1411	R1300Q, R1411I No	W_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, M1281 No	N_P	2397	4175T>C
His1389Arg	H_R	1389	H1182R, H1311I No	W_+	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, C1168I No	N_P	2378	4124G>C
Leu1373Pro	L-P	1373	L1373P, L1289P No	N_N	2376	4118T>C
Gln1372Lys	Q-K	1372	Q1091K, Q1145I No	P_+	2375	4114C>A
Gln1372Glu	Q-E	1372	Q1261E, Q1288 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, S1149F Yes	P_N	2359	4066T>C
Gly1355Asp	G-D	1355	G1355D, G1271 Yes	N_-	2358	4064G>A
Gly1355Val	G-V	1355	G1355V, G1244I Yes	N_N	2358	4064G>T
Gly1355Ser	G-S	1355	G1355S, G1148I Yes	N_P	2358	4064G>A
Trp1353Arg	W-R	1353	W1146R, W127I Yes	N_+	2356	4057T>C
Val1349Leu	V-L	1349	V1349L, V1271L No	N_N	2352	4045G>C
Met1344Leu	M-L	1344	M1137L, M1233I No	N_N	2347	4030A>C

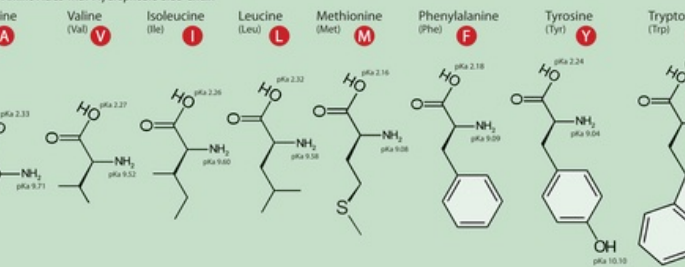
## Amino Acids with Electrically Charged Side Chains



## Amino Acids with Polar Uncharged Side Chains



## Amino Acids with Hydrophobic Side Chain



# 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 !!!