Huntington Disease Bioinformatics Research

Xinyi Zhang CID:02020037

Overview

Huntington Disease (HD) is a neurodegenerative disease that can cause movement, cognition, and behaviors abnormalities. Symptoms include chorea, dystonia, cognitive decline, and incoordination of the body (Dayalu & Albin, 2014). HD patients cannot live independently; instead, they need extra full-time care from others until death.

HD is an autosomal dominant disease, meaning it has a 50% chance of inheriting the disease to the next generation. Meanwhile, the age of onset is normally around 30 to 40 years old. Those facts make HD relatively dangerous because people around 30 to 40 years old may already have offspring. Thus, their children are very likely to inherit this fatal disease.

Genetic information

The gene that causes HD is the Huntingtin gene (*HTT*). The gene position is 4p16.3, which is on chromosome 4. It starts from base 3074681 and ends at base 3243960. Therefore, it is a 169,279 bp-long gene (https://www.ncbi.nlm.nih.gov). *HTT* gene codes for the huntingtin protein, which can bind to many transcription factors to regulate DNA transcription.

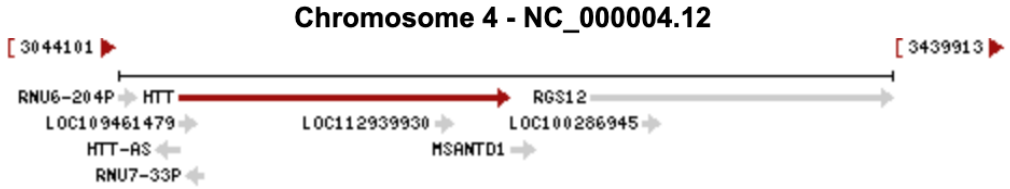
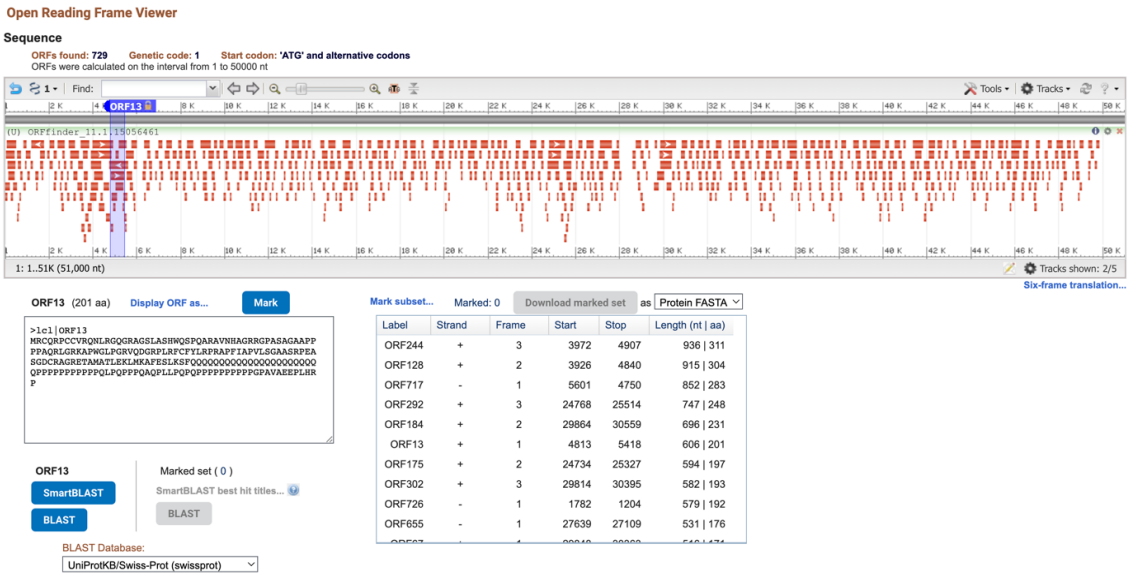


Figure 1: The position of *the HTT* gene on chromosome 4. The location is at 4p16.3.

The disease is caused by more than 36 repeats of CAG bases at 4p16.3, while a healthy individual without HD normally has 10 to 35 repeats (Lee et al., 2018). Generally, the longer the repeats, the earlier the disease will be onset. However, when the number of CAG repeats is equal to, or more than 40 times, it definitely becomes fatal (Lee et al., 2018). Any number between 35 to 40 has the chance of being healthy. When the repeats reach 55 times, Juvenile Huntington Disease (JHD) is more likely to happen (https://www.orpha.net), which has an early age of onset at or under 20-year-old.

CAG codes for glutamine (Q). These triplet nucleotide repeats are positioned at the 5' end of the coding sequence, which will be translated into a polyQ amino acid strand (Finkbeiner, 2011). Thus, HD patients have longer polyQ strands. The BLASTN search function can show the sequence for *the HTT* gene. By downloading the FASTA file and running it in the Open Reading Frame (ORF)Viewer (https://www.ncbi.nlm.nih.gov), the amino acid sequence of the gene with a start and stop codon can be shown. Part of the result is displayed in figure 2.

ORF13 is the ORF that contains polyQ amino acid sequences. In this case, 23 glutamines are being produced which indicates that it is a healthy case. I can assume that there will be more Qs shown if a mutant *HTT* gene is translated.



Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF244	+	3	3972	4907	936 | 311
ORF128	+	2	3926	4840	915 | 304
ORF717	-	1	5601	4750	852 | 283
ORF292	+	3	24768	25514	747 | 248
ORF184	+	2	29864	30559	696 | 231
ORF13	+	1	4813	5418	606 | 201
ORF175	+	2	24734	25327	594 | 197
ORF302	+	3	29814	30395	582 | 193
ORF726	-	1	1782	1204	579 | 192
ORF655	-	1	27639	27109	531 | 176

Figure 2: Open Reading Frame search for *HTT* gene. ORF13 has the polyQ amino acid sequence, which makes it the potential site for mutation.

The *HTT* gene is mostly expressed in the brain. By observing figure 3, one can find out that the mRNA is expressed the most frequently in the pre-frontal cortex (http://biogps.org), at 137.05 rpkm.

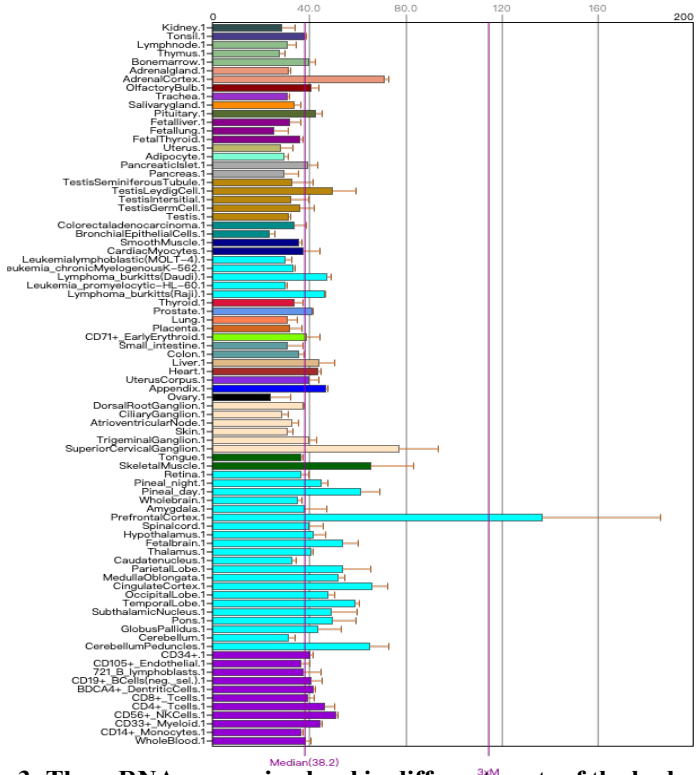


Figure 3: The mRNA expression level in different parts of the body of *H.sapiens*. The brain regions (blue) have higher expression levels than other regions.

Protein

Huntingtin protein has the role of post-translational modification, microtubule-mediated transport, and other functions inside the cell (Yalinca et al., 2019). Similar to mRNA, shown in figure 4, huntingtin is mostly expressed inside the brain (https://www.proteinatlas.org). Therefore, if huntingtin is mutant, there will be problems regarding basic activities one needs in daily life, especially considering that the autonomic nervous system is also being affected.

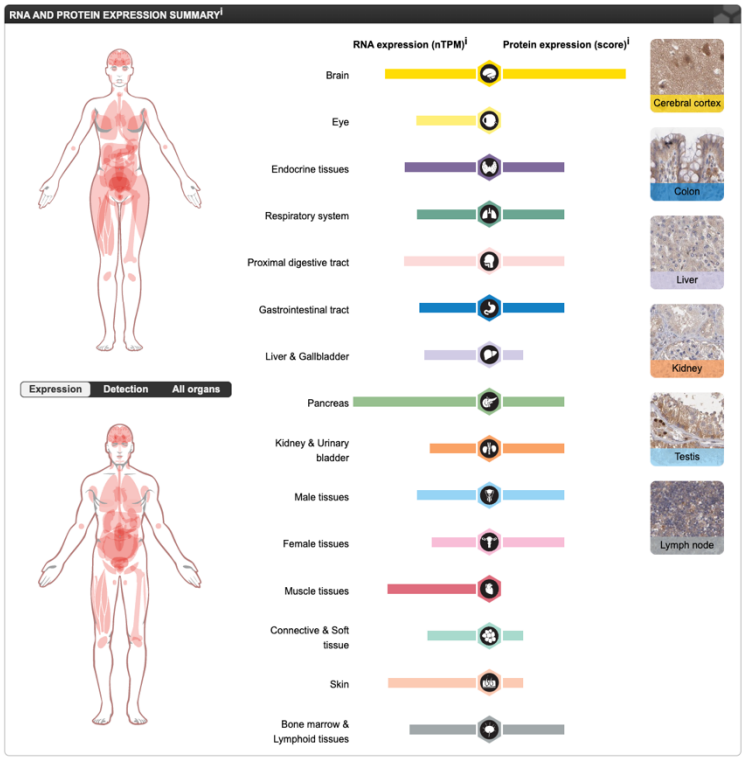


Figure 4: The RNA and protein expression summary of huntingtin in the human body. The brain has both high RNA and protein expression but the protein expression is more obvious compared with the other organs.

The tract of polyQ repeats in exon 1 of huntingtin protein is essential in HD. This sequence is 17-residue long (Yalinca et al., 2019). Figure 5 shows its protein structure (https://www.uniprot.org). The structure is at the N-terminus of the huntingtin protein. The polyQ region can increase the random coil length which may help with negative interaction with other proteins, leading to a pathogenic effect (Kim et al., 2009).

Figure 5: Secondary structure of exon 1 of the huntingtin protein. It consists of an amino-terminal α helix, a poly17Q region, and a polyproline helix.

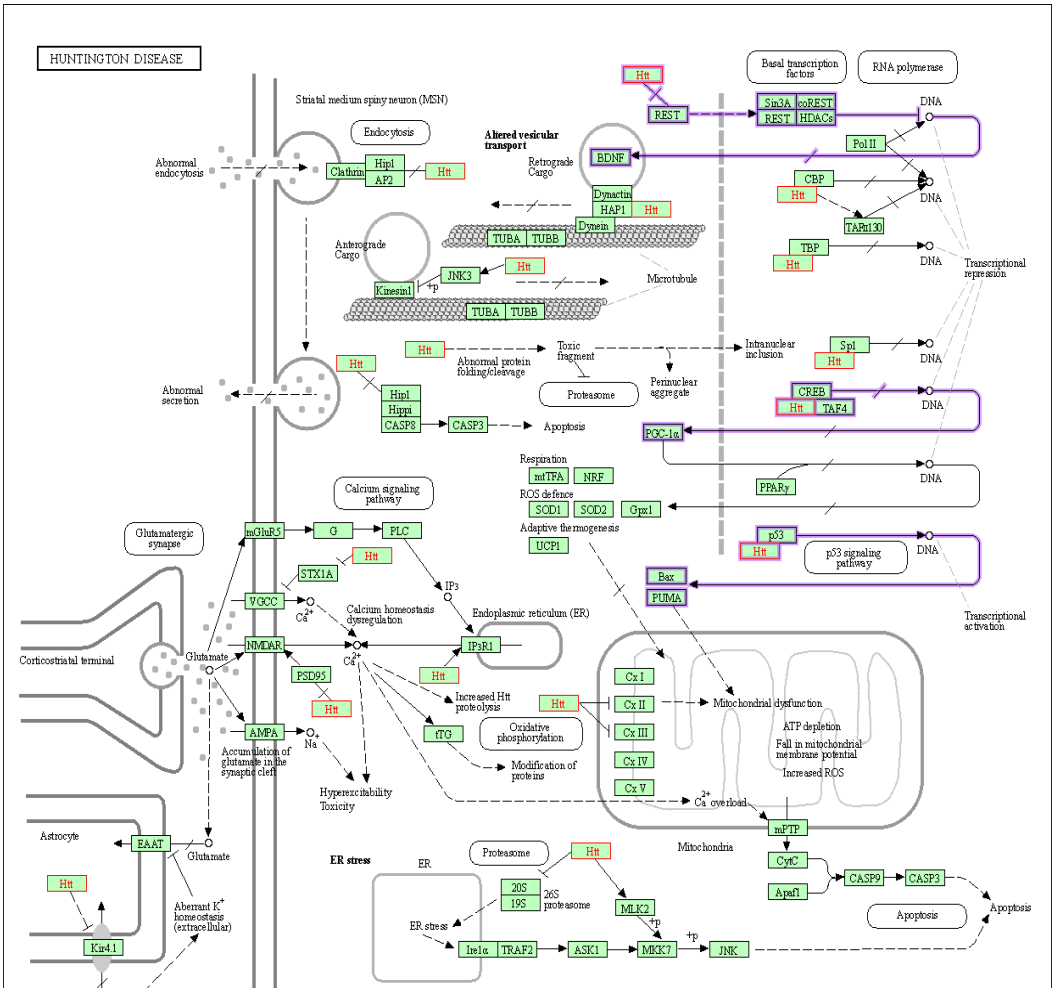
Cellular Pathway

The mutant *HTT* gene also has a negative impact on the cellular pathway, including proteasomal dysfunction, ER stress, impairment of autophagy, and many more (https://www.genome.jp).

One of the most important pathways that should pay extra attention is the "transcription network," which is shown in purple in figure 6. This is because the polyQ strand will translocate to the nucleus, leading to transcription impairment and cause neuron cell death (https://www.genome.jp).

Some evidence shows that BDNF (Brain-Derived Neurotrophic Factor) reduction causes HD pathogenesis (Xie et al., 2010). As we can see in figure 6, the transcription for BDNF is being disrupted and thus negatively affect the retrograde transport of BDNF on the microtubule. Therefore, the diagram proves the evidence.

In addition, the p53 tumor suppressor gene also affects cell function. Mutant huntingtin protein binds to p53 and activates the p53 signaling pathway. As figure 6 shows, this leads to mitochondrial dysfunction, which finally causes apoptosis of neuron cells.



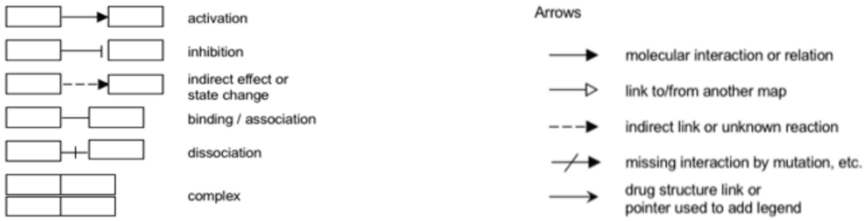


Figure 6: Part of the cellular pathway of the *HTT* gene. The purple lines highlight the transcription network. There are multiple interactions that are blocked because of the mutation.

Variants

One of the variants of the *HTT* gene can cause another disease called Lopes-Maciel-Rodan syndrome (LOMARS) which is being discovered in 2016. This syndrome has similar symptoms as HD which include dystonia and swallowing problems, but some different symptoms include small and cold hands and feet (Lopes et al., 2016). Different from HD, LOMARS is autosomal recessive.

A Single Nucleotide Variation (dsSNV) named rs34315806 occurs at position 3160307 on *the HTT* gene and causes LOMARS to be onset (https://www.ncbi.nlm.nih.gov). SNV means a single nucleotide change. This missense mutation causes allele change from C to T, which will lead to amino acid change. The UCSC website indicates that the original amino acid being produced is Threonine and the missense variation turns the amino acid into Methionine (https://genome.ucsc.edu).

This protein transition is called thr1260-to-met (T1260M) substitution, which can be found on ClinVar website (https://www.ncbi.nlm.nih.gov).

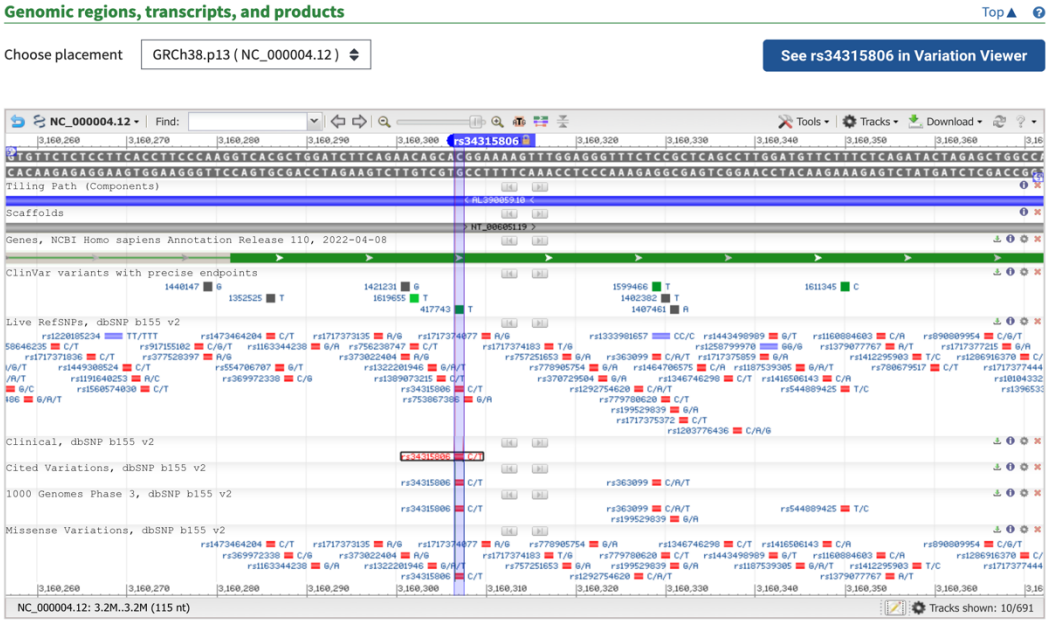


Figure 7: SNV rs34315806 variant on HTT gene leads to LOMARS. It is a missense mutation.

Some other variants are listed in table 1. Four of them are related to LOMARS, in which all of them have Single-Nucleotide Polymorphism (SNP), which has a similar meaning to SNV. Number 0002 variant has the same missense mutation as rs34315806 (Number 0003), which is allele change from C to T, but the amino acid changes from Proline to Leucine (https://genome.ucsc.edu). Number 0004 is special because it is a splice donor variant meaning that there is genetic alteration happening at the boundary of intron and exon. Thus, more severe changes to the potential amino acid sequence may be produced.

Table 1: The 5 variants related to *the HTT* gene

Number	Phenotype	Mutation	SNP	gnomAD	ClinVar
.0001	HUNTINGTON DISEASE	HTT, (CAG)n REPEAT EXPANSION	-	-	RCV000030659
.0002	LOPES-MACIEL-RODAN SYNDROME	HTT, PRO703LEU	rs768047421	rs768047421	RCV000477706...
.0003	LOPES-MACIEL-RODAN SYNDROME	HTT, THR1260MET (rs34315806)	rs34315806	rs34315806	RCV000477735...
.0004	LOPES-MACIEL-RODAN SYNDROME	HTT, IVS34DS, G-A, +1	rs1060505027	-	RCV000477676...
.0005	LOPES-MACIEL-RODAN SYNDROME	HTT, PHE2719LEU	rs1060505028	-	RCV000477714

Animal models

Animal models help us to understand HD better. *HTT* gene also exists in many different animals. Just as figure 8 shows, there are homologs existed in primates, rodents, birds, and many more (https://www.ensembl.org).

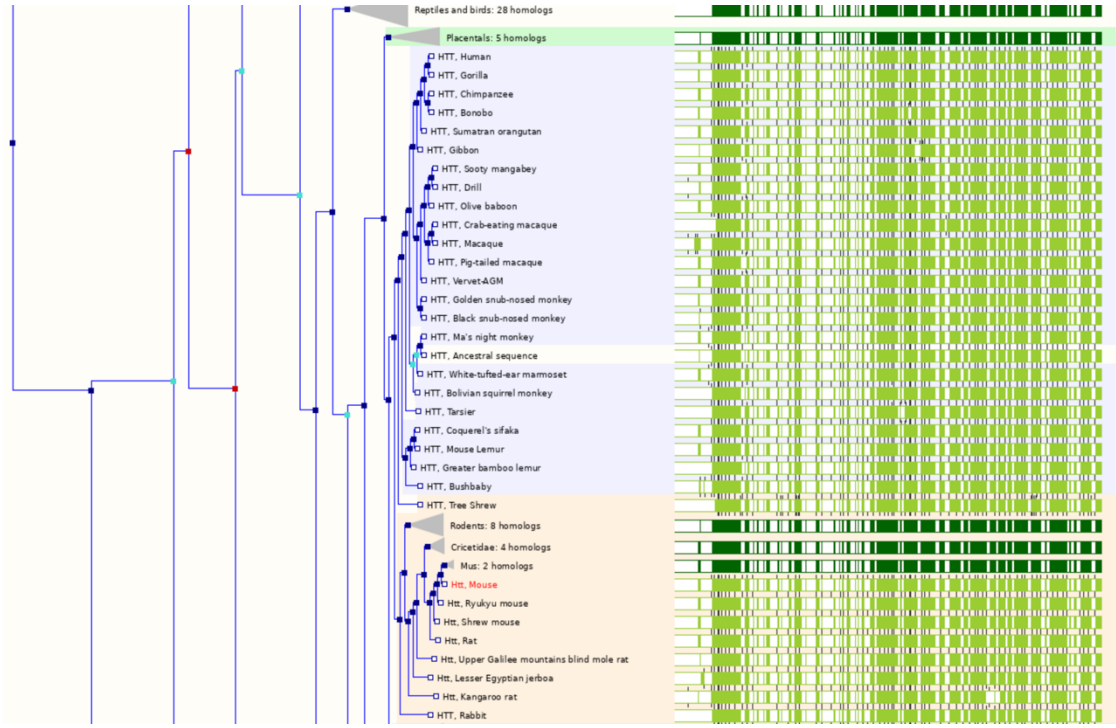


Figure 8: Gene tree showing homologs of the *HTT* gene in different animals.

One of the most common types of experimental models of HD is done on *Mus musculus* (mice). Mice *HTT* gene is also mostly expressed in the brain region (Figure

9), but mice do not normally develop HD. However, gene knock-in technology can insert the mutants into the DNA of the mice to help with the investigation.

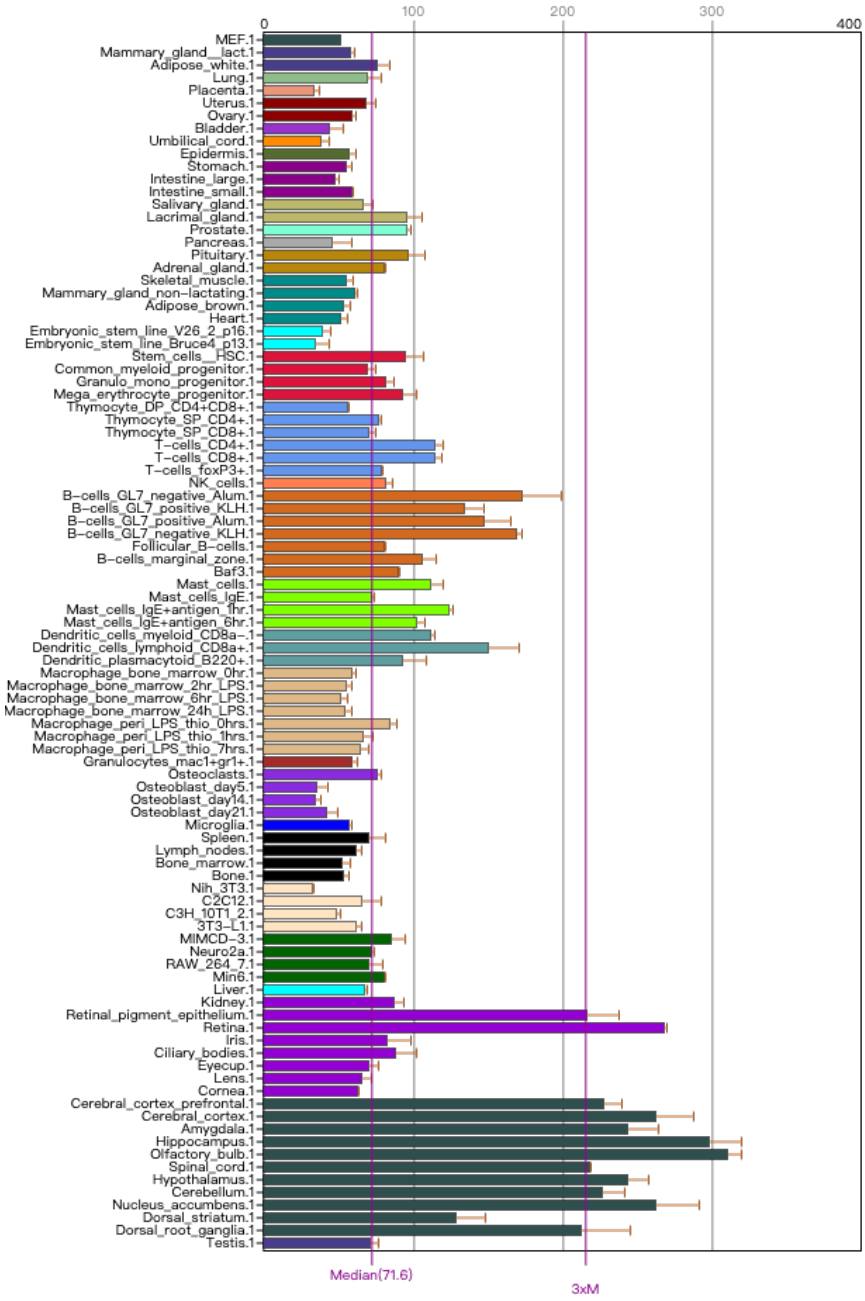


Figure 9: The mRNA expression level of the HTT gene in *M.musculus*. The dark green part is about the brain regions with the highest frequency.

Different from humans, the location of the HTT gene in mice is on chromosome 5 (http://www.informatics.jax.org). On the MGI website, there are 60+ mice models so far related to HD. Part of them is shown in figure 10.

Term Browser | Genes (27) | Models (64)

Excel File | Text File | All mouse models of Huntington's disease with phenotypic similarity to the human disease

Disease Term	Allelic Composition	Genetic Background	Reference	Phenotypes
Huntington's disease	Tg(HDexon1)62Gpb/0	B6CBA-Tg(HDexon1)62Gpb/1J	J:99425, J:111237	View
Huntington's disease	$Htt^{tm2Detl}$/Htt^{+}	B6J.129P2-$Htt^{tm2Detl}$	J:220868	View
Huntington's disease	Tg(Ppp1r1b-HTT*)1Meeh/0	C57BL/6J-Tg(Ppp1r1b-HTT*)1Meeh	J:139807	View
Huntington's disease	Tg(GFAP-HTT*160Q)1Xjl/0	either: (involves: FVB/NCrl) or (involves: C3H * C57BL/6 * FVB/NCrl)	J:156460	View
Huntington's disease	Tg(GFAP-HTT*160Q)31Xjl/0	either: (involves: FVB/NCrl) or (involves: C3H * C57BL/6 * FVB/NCrl)	J:156460	View
Huntington's disease	Tg(HTT*97Q)IXwy/0	FVB-Tg(HTT*97Q)IXwy	J:137345	View
Huntington's disease	Tg(HTT*)1Xwy/0	FVB-Tg(HTT*)1Xwy	J:157723	View
Huntington's disease	Htt^{tm1Hay}/Htt^{tm1Hay} Tg(HTT*97Q)IXwy/0 Tg(YAC18)18Hay/Tg(YAC18)18Hay	FVB.Cg-Htt^{tm1Hay} Tg(HTT*97Q)IXwy Tg(YAC18)18Hay	J:191147, J:215223	View
Huntington's disease	Tg(YAC72)2511Hay/Tg(YAC72)2511Hay	FVB/N-Tg(YAC72)2511Hay	J:105728	View
Huntington's disease	Tg(YAC128)53Hay/0	FVB/N-Tg(YAC128)53Hay	J:84453, J:105723, J:111237	View
Huntington's disease	Tg(YAC128)55Hay/0	FVB/N-Tg(YAC128)55Hay	J:105723	View
Huntington's disease	Tg(YAC128)55Hay/Tg(YAC128)55Hay	FVB/N-Tg(YAC128)55Hay	J:105723, J:105728	View
Huntington's disease	Tg(YAC128)#Hay/0	FVB/N-Tg(YAC128)#Hay	J:98736	View
Huntington's disease	$Htt^{tm1Detl}$/$Htt^{tm1Detl}$	involves: 129P2/OlaHsd * C57BL/6J	J:67074	View
Huntington's disease	$Htt^{tm1Detl}$/Htt^{+}	involves: 129P2/OlaHsd * C57BL/6J	J:67074	View
Huntington's disease	$Htt^{tm2Detl}$/$Htt^{tm2Detl}$	involves: 129P2/OlaHsd * C57BL/6J	J:123681	View
Huntington's disease	$Htt^{tm1.1Tna}$/$Htt^{tm1.1Tna}$	involves: 129P2/OlaHsd * C57BL/6J	J:76017	View
Huntington's disease	$Htt^{tm1.1Tna}$/Htt^{+}	involves: 129P2/OlaHsd * C57BL/6J	J:76017	View
Huntington's disease	$Cnr1^{tm1Map}$/$Cnr1^{tm1Map}$ Tg(HD82Gln)81Gschi/0	involves: 129S1/Sv * 129X1/SvJ * C3H * C57BL/6 * CD-1	J:172874	View
Huntington's disease	Tg(HD82Gln)81Gschi/0	involves: 129S1/Sv * 129X1/SvJ * C3H * C57BL/6 * CD-1	J:172874	View
Huntington's disease	Htt^{tm1Mem}/Htt^{tm6Mem} Tg(CAG-cre/Esr1*)5Amc/0	involves: 129S1/Sv * 129X1/SvJ * C57BL/6 * CBA * Swiss Webster	J:260244	View
Huntington's disease	Gt(ROSA)26Sor$^{tm1(HD*103Q)Xwy}$/? Tg(Nes-cre)1Kln/?	involves: 129S1/Sv * 129X1/SvJ * C57BL/6 * SJL	J:99759	View

Figure 10: Mouse models related to HD. Each model represents different experiments that different groups have done. By clicking the "phenotypes," more information regarding the phenotype coded by *the HTT* gene in the mice can be shown.

In order to make HD symptoms onset, CAG repeats need to be inserted into the *HTT* gene of the mice. The database provides many aspects of HD studies. For example, an experiment done on YAC128 mice revealed that homozygosity of the alleles leads to a more severe phenotype (Graham et al., 2005). Alternatively, by comparing different phenotypes of those models, it can be observed that mice show similar symptoms as humans with HD, including abnormal activity in movement, reduction in brain regions, weight loss, and many others.

Case study

An unique case study that I found at the Coriell Institute about HD is about a family in Venezuela. It is a Barranquitas village that has the highest concentration of HD patients with the genetic disease influence of more than 10 generations. In the cell line section of Coriell Institute, there is a subcollection called “Venezuelan Huntington Disease."

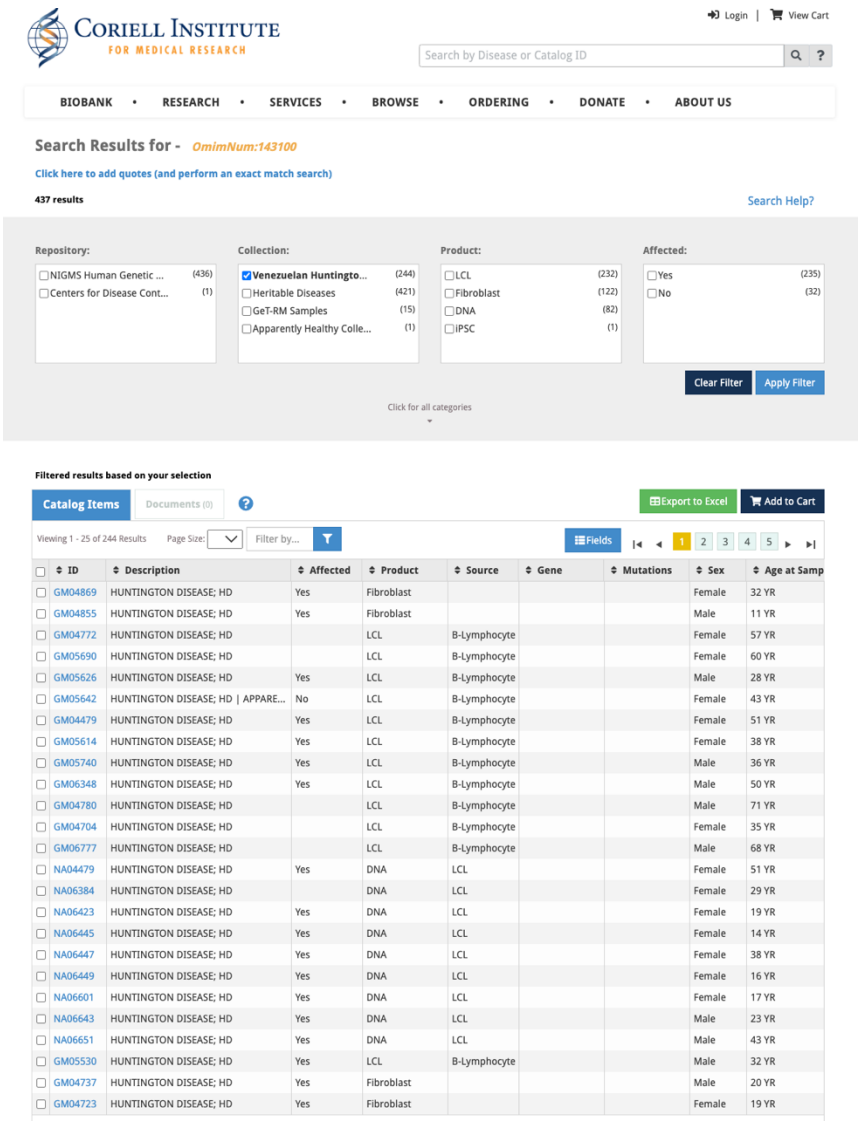


Figure 11: Database for the Venezuelan Huntington Disease. Of the 244 data collected, 235 are affected by HD.

I then searched for more information about the case study. The study was done by Dr. Nancy Wexler for more than 40 years. In one of her papers, among the 4,000 individuals she tested, she found out the distribution of the CAG repeats among people in that village. As shown in figure 12, there are two peaks that are representative. One is 17 repeats which show that this is the number most common for healthy individuals. While the other peak length is at 44, which is above the full penetration threshold. Theoretically, people can have an infinite number of CAG repeats, while in this study, the maximum length is 86 (Wexler, 2004).

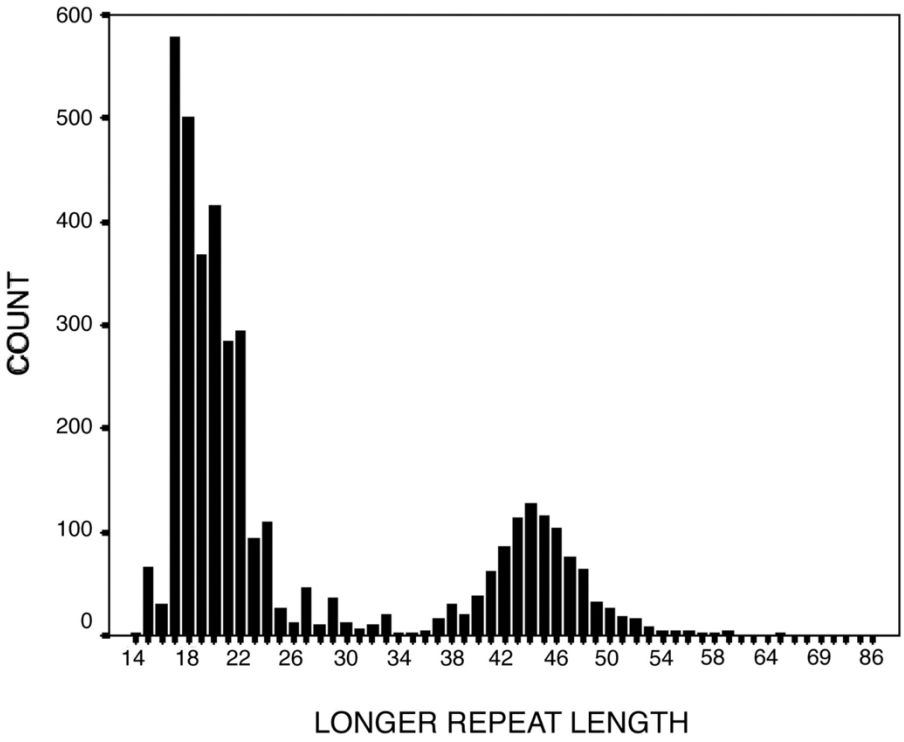


Figure 12: Distribution of the CAG repeat of the Venezuelan kindreds. There are two normal distributions; one peak represents the normal allele, while the other peak means the mutant allele.

In addition, it is being found that the mean age of onset of the Venezuelan kindreds (34.35 ± 10.07)) is earlier than American (37.47 ± 13.28) and Canadian kindreds (40.36 ± 12.97). One explanation for this is poverty. Many of them live near highly polluted water, and they have a problematic diet. This also verifies the correlation between age of onset and environment. It has been found that the environment is responsible for 63% of the variance in age of onset (Wexler, 2004). In the database shown in figure 11, most individuals have a large family, and records indicate that some people have 10+ children. Sadly, many of the children are also affected. Undoubtedly, it reveals the poverty of the area things need to do to educate people there about HD.

Word count: 1488

Reference List

BioGPS - your gene portal system. *BioGPS Blog*. http://biogps.org/#goto=genereport&id=3064 [Accessed: November 2, 2022].

Dayalu, P. & Albin, R.L. (2014) *Huntington disease: Pathogenesis and treatment*, *Neurologic Clinics*. Elsevier. https://www.sciencedirect.com/science/article/abs/pii/S0733861914000711?via%3Dihub

Ensembl Genome Browser 108.(n.d.) *Gene: Htt ENSMUSG00000029104* (no date) *Gene tree - mus_musculus - ensembl genome browser 108*. https://www.ensembl.org/Mus_musculus/Gene/Compara_Tree?g=ENSMUSG00000029104%3Br [Accessed: November 3, 2022].

Finkbeiner, S. (2011) "Huntington's disease," *Cold Spring Harbor Perspectives in Biology*, 3(6). https://doi.org/10.1101/cshperspect.a007476 .

Graham, R.K. *et al.* (2005) "Levels of mutant Huntingtin influence the phenotypic severity of Huntington disease in Yac128 Mouse models," *Neurobiology of Disease*, 21(2), 444–455. https://doi.org/10.1016/j.nbd.2005.08.007 .

Kim, M.W. *et al.* (2009) "Secondary structure of Huntingtin Amino-terminal region," *Structure*, 17(9), 1205–1212. https://doi.org/10.1016/j.str.2009.08.002 .

Kyoto Encyclopedia of Genes and Genomes. (n.d.) *Huntington disease - Homo Sapiens (human) KEGG PATHWAY: Huntington disease - Homo sapiens (human)*. https://www.genome.jp/pathway/hsa05016+3064 [Accessed: November 1, 2022].

Lee, J.K. *et al.* (2018) "Effect of trinucleotide repeats in the Huntington's gene on Intelligence," *EBioMedicine*, 31, 47–53. https://doi.org/10.1016/j.ebiom.2018.03.031 .

Lopes, F. *et al.* (2016) "Identification of novel genetic causes of Rett syndrome-*like*phenotypes," *Journal of Medical Genetics*, 53(3), 190–199. https://doi.org/10.1136/jmedgenet-2015-103568 .

Mouse Genome Informatics. (n.d.) *HTT MGI Mouse gene detail - MGI:96067 - huntingtin*. http://www.informatics.jax.org/marker/MGI:96067 [Accessed: November 1, 2022].

Orphanet. (n.d.) *Juvenile huntington disease*. https://www.orpha.net/consor/cgibin/OC_Exp.php?lng=EN&Expert=248111 [Accessed: October 31, 2022].

UCSC Genome Browser. (n.d.) *Common (1000 genomes phase 3 maf >= 1%) short genetic variants from dbSNP release 153 (RS34315806)*. https://genome.ucsc.edu/cgi-bin/hgc?hgsid=1488472387_josAs7S4AAxAayta58AxkmJlow7M&db=hg38&c=chr4&l=3160206&r=3160407&o=3160306&t=3160307&g=dbSnp153Common&i=rs34315806 [Accessed: November 1, 2022].

UCSC Genome Brower.(n.d.) *All short genetic variants from dbsnp release 153 (RS768047421)*. https://genome.ucsc.edu/cgi-bin/hgc?hgsid=1488477691_YgAKpfAlqrz4suKga8pyiTXPELWs&db=hg38&c=chr4&l=3131546&r=3131747&o=3131646&t=3131647&g=dbSnp153&i=rs768047421 [Accessed: November 1, 2022].

UniProt.(n.d.)*P42858 · HD_HUMAN* https://www.uniprot.org/uniprotkb/P42858/entry#structure [Accessed: November 2, 2022].

U.S. National Library of Medicine. (n.d.) *Loc112939930 SHARPR-mpra regulatory region 10599 [homo sapiens (human)] - gene - NCBI National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/gene/112939930 [Accessed: November 1, 2022].

U.S. National Library of Medicine. (n.d.) *Orffinder home - NCBI National Center for Biotechnology Information*. Available at: https://www.ncbi.nlm.nih.gov/orffinder/ [Accessed: November 1, 2022].

U.S. National Library of Medicine. (n.d.) *RS34315806 RefSNP report - dbSNP - NCBI National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/snp/rs34315806#history [Accessed: November 1, 2022].

U.S. National Library of Medicine. (n.d.) *VCV000417743.9 - Clinvar - NCBI National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/clinvar/variation/417743/?oq=rs34315806%5BExternal%2Ballele%2BID%5D&m=NM_001388492.1%28HTT%29%3Ac.3779C%3ET+%28p.Thr1260Met%29 [Accessed: November 1, 2022].

The Human Protein Atlas (n.d.) *Tissue expression of HTT - summary - The human protein atlas*. https://www.proteinatlas.org/ENSG00000197386-HTT/tissue [Accessed: November 2, 2022].

The U.S.–Venezuela Collaborative Research Project and Wexler, N.S. (2004) "Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset," *Proceedings of the National Academy of Sciences*, 101(10), 3498–3503. https://doi.org/10.1073/pnas.0308679101 .

Xie, Y., Hayden, M.R. and Xu, B. (2010) “BDNF overexpression in the forebrain rescues Huntington's disease phenotypes in Yac128 Mice,” *Journal of Neuroscience*, 30(44), 14708–14718. https://doi.org/10.1523/jneurosci.1637-10.2010 .

Yalinca, H. *et al.* (2019) “The role of post-translational modifications on the energy landscape of Huntingtin N-terminus,” *Frontiers in Molecular Biosciences*, 6. https://doi.org/10.3389/fmolb.2019.00095 .