Instituto Politécnico Nacional

Escuela Superior de Cómputo

Bioinformatics

Practice 7 - GenBank

Group: 3CV9

Student:
Ramos Diaz Enrique

Professor:
Rosas Trigueros Jorge Luis

Development Date:
November 25th 2020

Due Date:
December 2nd 2020

1 Theoretical Framework

1.1 GenBank

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis. The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information [1].

1.2 Gene

A Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide [2].

Human genes examples and their descriptions:

1. **MC1R - Melanocortin 1 Receptor Gene:** This receptor is a major determining factor in sun sensitivity and is a genetic risk factor for melanoma and non-melanoma skin cancer. Over 30 variant alleles have been identified which correlate with skin and hair color, providing evidence that this gene is an important component in determining normal human pigment variation. [3].
2. **SLC12A2 - Solute Carrier Family 12 Member 2:** The protein encoded by this gene mediates sodium and chloride transport and reabsorption. The encoded protein is a membrane protein and is important in maintaining proper ionic balance and cell volume. This protein is phosphorylated in response to DNA damage. Three transcript variants encoding two different isoforms have been found for this gene [4].
3. **TYRP1 - Tyrosinase Related Protein 1:** This gene encodes a melanosomal enzyme that belongs to the tyrosinase family and plays an important role in the melanin biosynthetic pathway. Defects in this gene are the cause of rufous oculocutaneous albinism and oculocutaneous albinism type III [5].

1.3 Database of Short Genetic Variations

The Database of Short Genetic Variations (dbSNP) contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations [6].

2 Material and Equipment

- GenBank web page [1].
- ClinVar Miner web page [7].
- PubMed web page [8].

3 Practice Development

3.1 Searching for genes

In this practice some variations from three different human genes will be reviewed and analyzed in order to determine their respective effects in the body. The selected genes are the following:

1. MC1R - Melanocortin 1 Receptor Gene (Human) [3]
2. SLC12A2 - Solute Carrier Family 12 Member 2 (Human) [4] [8]
3. TYRP1 - Tyrosinase Related Protein 1 (Human) [5] [8]

3.2 MC1R - Melanocortin 1 Receptor Gene (Human)

Go to GenBank [1] and on the search bar type for the **MC1R gene** (be sure to select "Gene" on the list options). The Figure 1 shows the correct result to use from this search. Click on the *Genome Browser* button to open a new tab that looks like Figure 2.

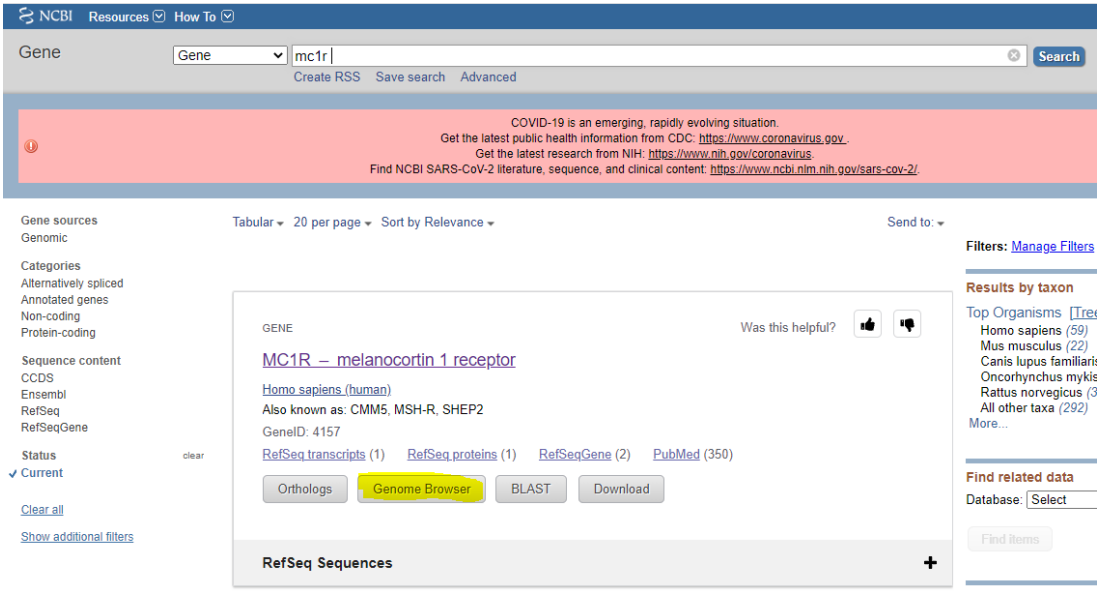

Figure 1: **MC1R** results on GenBank [3].

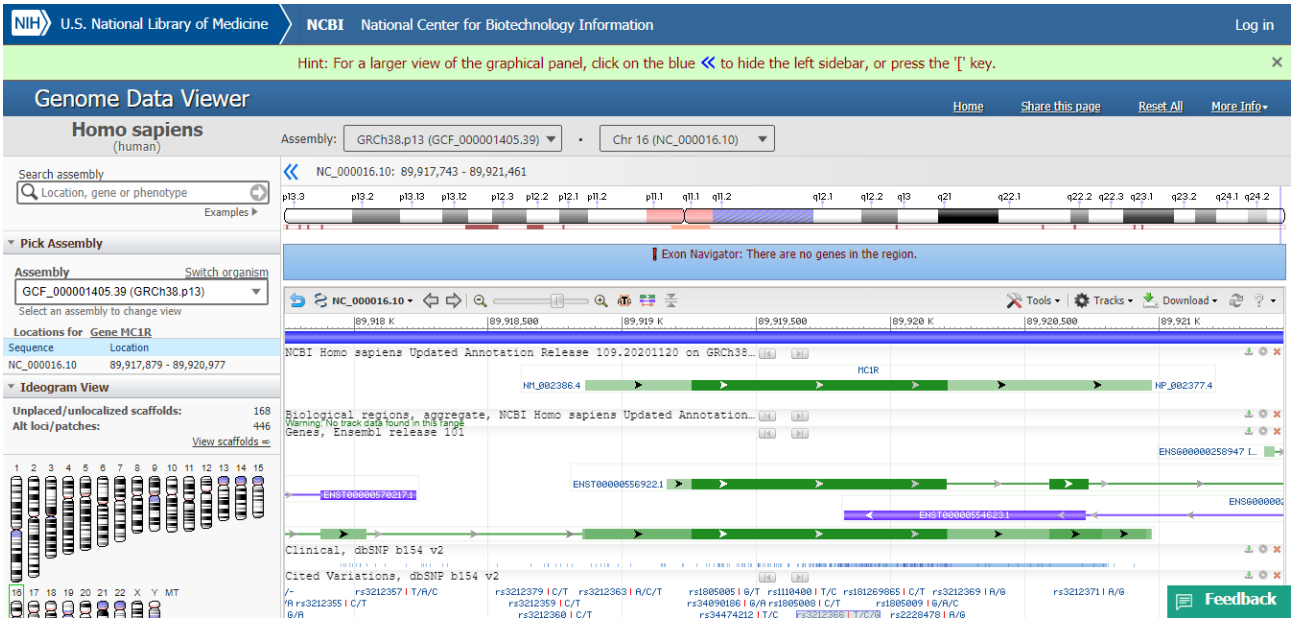

Figure 2: **MC1R** Genome Browser.

On the Genome Browser page, navigate until the *Cited Variations* - *dbSNP* section and search the **rs1805007** variation id. To see its SNP summary with more detail, place the mouse on the variation id link and then click on the *SNP Summary* link (see Figure 3).

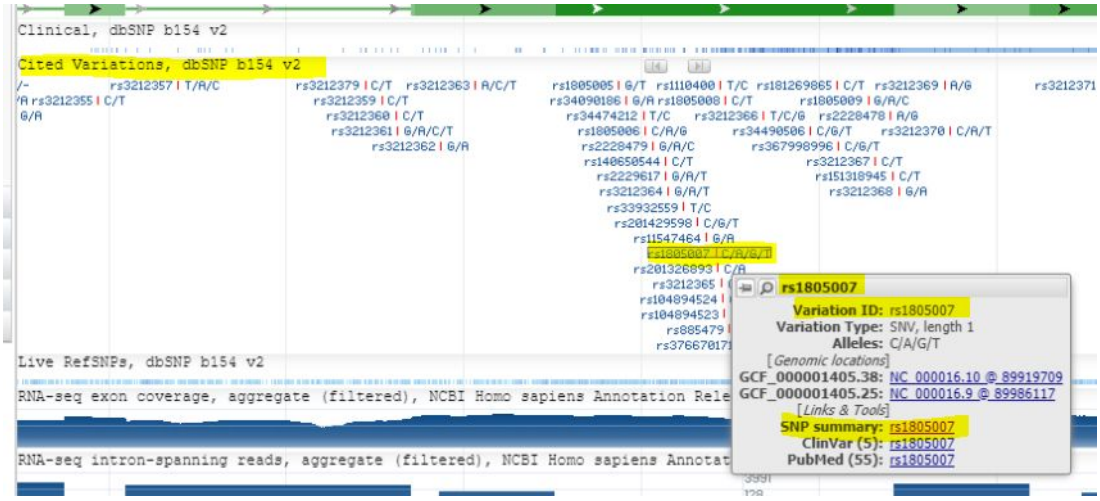

Figure 3: **MC1R rs1805007** Variation.

Again a new tab will be open, showing a table with all the SNP summary associated to this variation on the **MC1R** gene, as can be seen in Figure 4. Figure 5 indicates this alteration **that corresponds to people with red hair and freckles**.

Gene: MC1R, melanocortin 1 receptor (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
MC1R transcript	NM_002386.4:c.451C>A	R [CGC] > S [AGC]	Coding Sequence Variant
MC1R transcript	NM_002386.4:c.451C>G	R [CGC] > G [GGC]	Coding Sequence Variant
MC1R transcript	NM_002386.4:c.451C>T	R [CGC] > C [TGC]	Coding Sequence Variant
melanocyte-stimulating hormone receptor	NP_002377.4:p.Arg151Ser	R (Arg) > S (Ser)	Missense Variant
melanocyte-stimulating hormone receptor	NP_002377.4:p.Arg151Gly	R (Arg) > G (Gly)	Missense Variant
melanocyte-stimulating hormone receptor	NP_002377.4:p.Arg151Cys	R (Arg) > C (Cys)	Missense Variant

Figure 4: **MC1R rs1805007** SNP Summary.

Figure 5: **MC1R rs1805007** Effect: Read Hair and Freckles.

3.3 SLC12A2 - Solute Carrier Family 12 Member 2 (Human)

Go to GenBank [1] and on the search bar type for the **SLC12A2 gene**. The Figure 6 shows the correct result to use from this search. Click on the *Genome Browser* (see Figure 7).

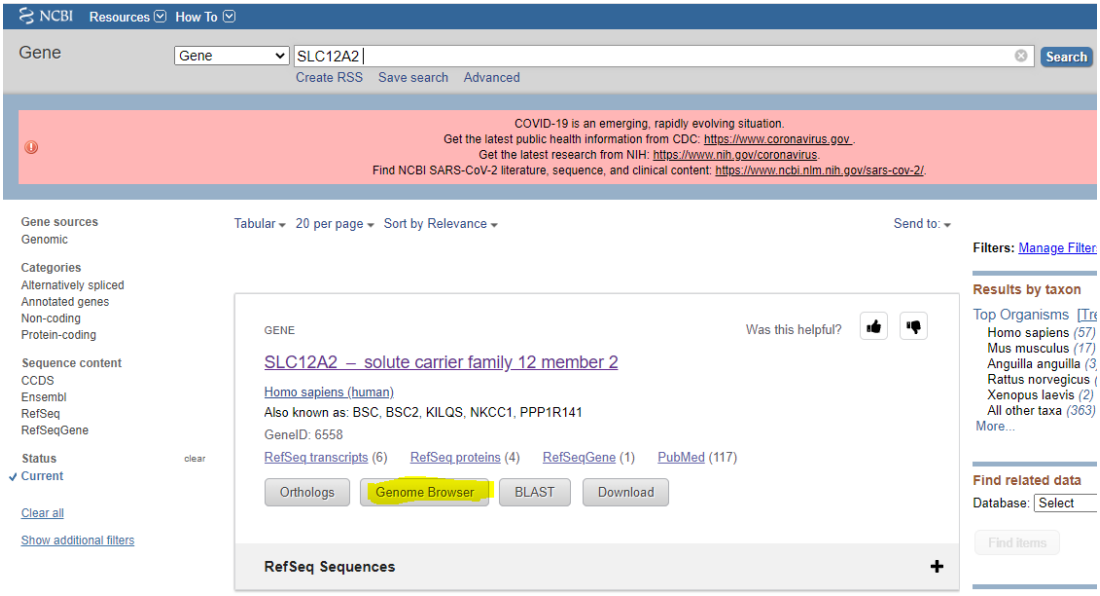

Figure 6: **SLC12A2** results on GenBank [4].

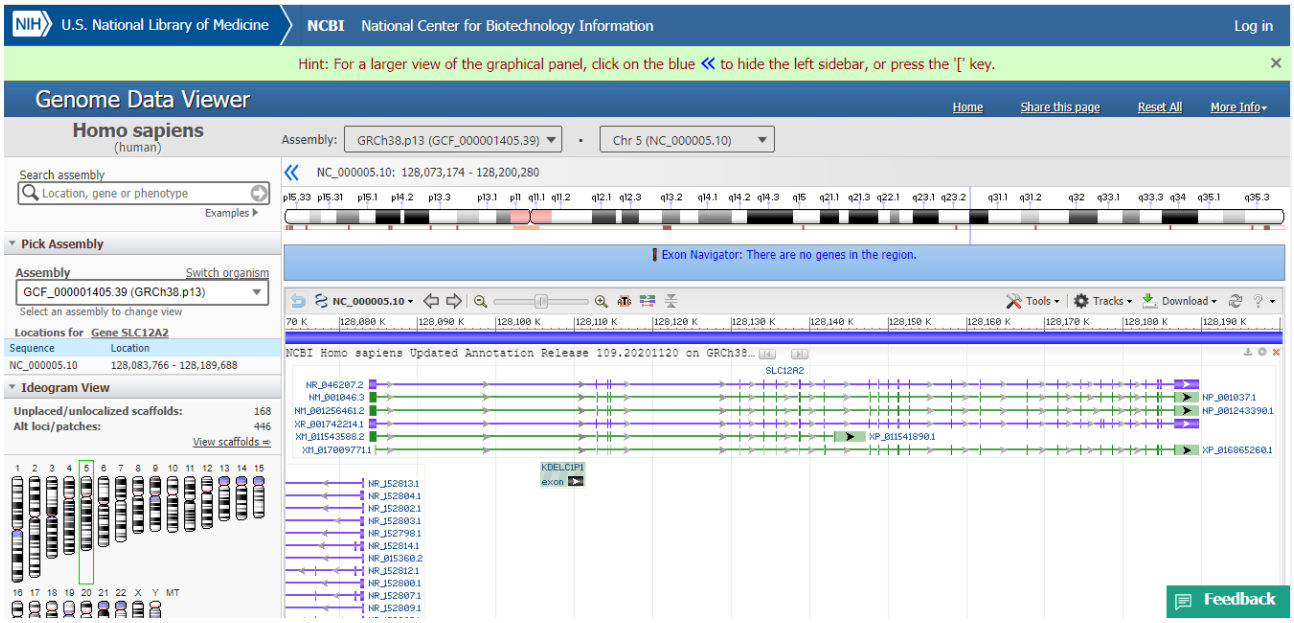

Figure 7: **SLC12A2** Genome Browser.

On the Genome Browser page, navigate until the *Cited Variations - dbSNP* section and search the **rs886040968** variation id. To see its SNP summary, place the mouse on the variation id link and then click on the *SNP Summary* link (see Figure 8).

Figure 8: **SLC12A2 rs886040968** Variation.

The table with all the SNP summary associated to this variation on the **SLC12A2** gene can be seen in Figure 9. Figure 10 indicates the **V[GT] > F[T] alteration, that is likely pathogenic and corresponds to infant onset multiple organ failure** [9].

Gene: SLC12A2, solute carrier family 12 member 2 (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
SLC12A2 transcript variant 1	NM_001046.3:c.3076_3086del	V [GT] > F [T]	Coding Sequence Variant
SLC12A2 transcript variant 2	NM_001256461.2:c.3028_303…	V [GT] > F [T]	Coding Sequence Variant
SLC12A2 transcript variant 3	NR_046207.2:n.3265_3275del	N/A	Non Coding Transcript Variant
SLC12A2 transcript variant X1	XR_001742214.1:n.3300_331…	N/A	Non Coding Transcript Variant
SLC12A2 transcript variant X2	XM_011543588.2:c.	N/A	Genic Downstream Transcript Variant
SLC12A2 transcript variant X3	XM_017009771.1:c.1318_132…	V [GT] > F [T]	Coding Sequence Variant
solute carrier family 12 member 2 isoform 1	NP_001037.1:p.Val1026fs	V (Val) > F (Phe)	Frameshift Variant
solute carrier family 12 member 2 isoform 2	NP_001243390.1:p.Val1010fs	V (Val) > F (Phe)	Frameshift Variant
solute carrier family 12 member 2 isoform X2	XP_016865260.1:p.Val440fs	V (Val) > F (Phe)	Frameshift Variant

Figure 9: **SLC12A2 rs886040968** SNP Summary.

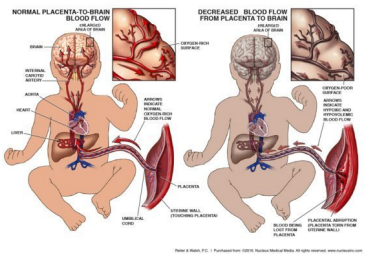

Figure 10: **V[GT]>F[T] variation:** Infant Onset Multiple Organ Failure.

Go back to the **SLC12A2** Genome Browser page, now navigate to the *Clinical - dbSNP* section and search the **rs1581138934** variation id. To see its SNP summary, place the mouse on the variation id link and then click on the *SNP Summary* link (see Figure 11)

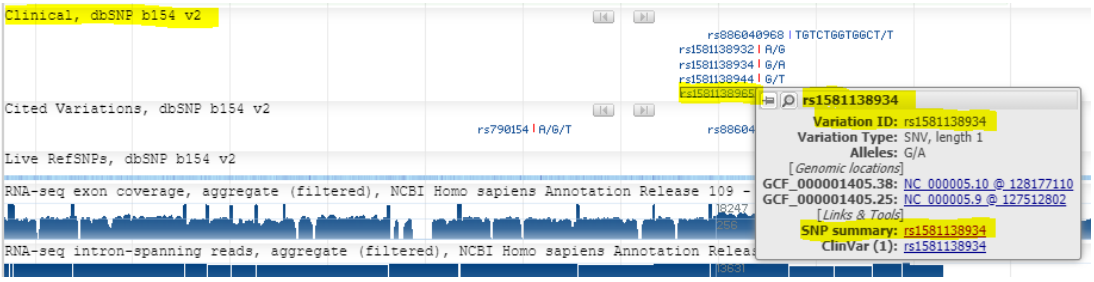

Figure 11: **SLC12A2 rs1581138934** Variation.

The table with all the SNP summary associated to this variation on the **SLC12A2** gene can be seen in Figure 12. Figure 13 indicates the **E[GAA]>K[AAA] alteration, that is associated to the hearing loss** [10].

Gene: SLC12A2, solute carrier family 12 member 2 (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
SLC12A2 transcript variant 1	NM_001046.3:c.2935G>A	E [GAA] > K [AAA]	Coding Sequence Variant
SLC12A2 transcript variant 2	NM_001256461.2:c.2930-145...	N/A	Intron Variant
SLC12A2 transcript variant 3	NR_046207.2:n.3124G>A	N/A	Non Coding Transcript Variant
SLC12A2 transcript variant X1	XR_001742214.1:n.3093G>A	N/A	Non Coding Transcript Variant
SLC12A2 transcript variant X2	XM_011543588.2:c.	N/A	Genic Downstream Transcript Variant
SLC12A2 transcript variant X3	XM_017009771.1:c.1177G>A	E [GAA] > K [AAA]	Coding Sequence Variant
solute carrier family 12 member 2 isoform 1	NP_001037.1:p.Glu979Lys	E (Glu) > K (Lys)	Missense Variant
solute carrier family 12 member 2 isoform X2	XP_016865260.1:p.Glu393Lys	E (Glu) > K (Lys)	Missense Variant

Figure 12: **SLC12A2 rs1581138934** SNP Summary.

Figure 13: **E[GAA]>K[AAA] variation:** Hearing Loss.

3.4 TYRP1 - Tyrosinase Related Protein 1 (Human) (Human)

Go to GenBank [1] and on the search bar type for the **TYRP1 gene**. The Figure 14 shows the correct result to use from this search. Click on the *Genome Browser* (see Figure 15).

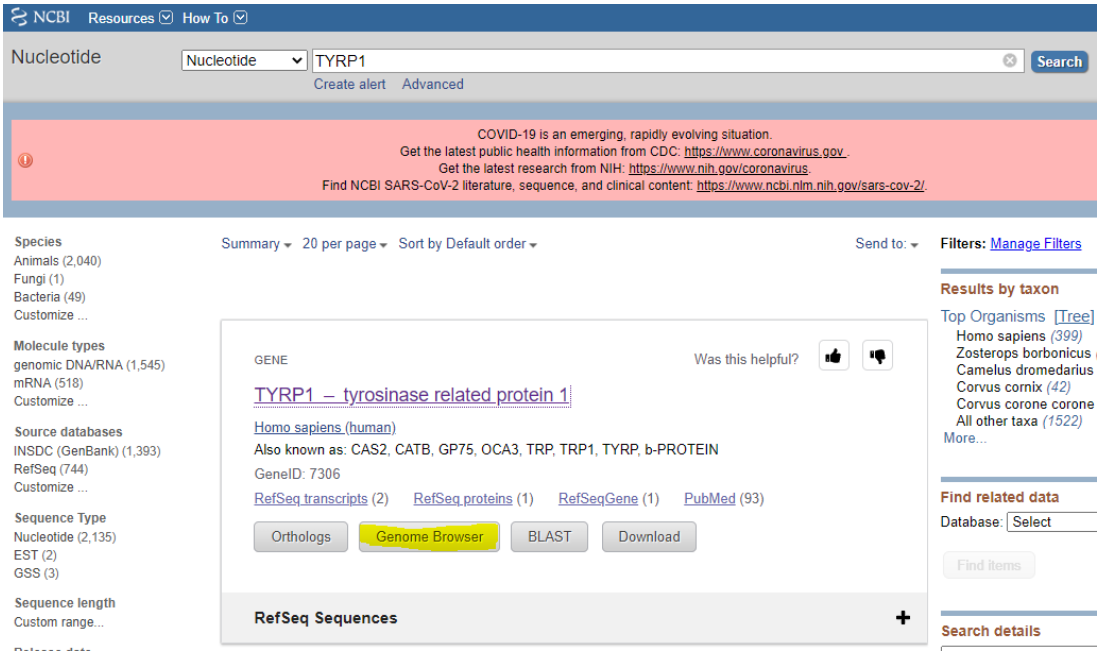

Figure 14: **TYRP1** results on GenBank [5].

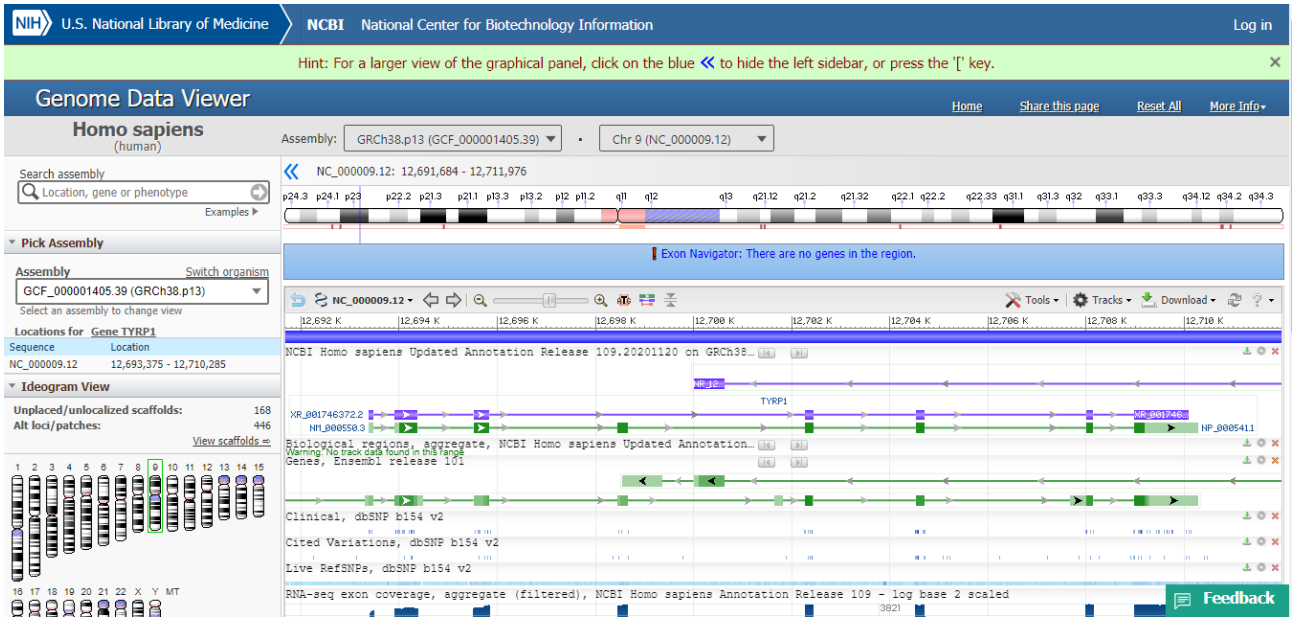

Figure 15: **TYRP1** Genome Browser.

On the Genome Browser page, navigate until the *Cited Variations - dbSNP* section and search

the **rs387907171** variation id. To see its SNP summary, place the mouse on the variation id link and then click on the *SNP Summary* link (see Figure 16).

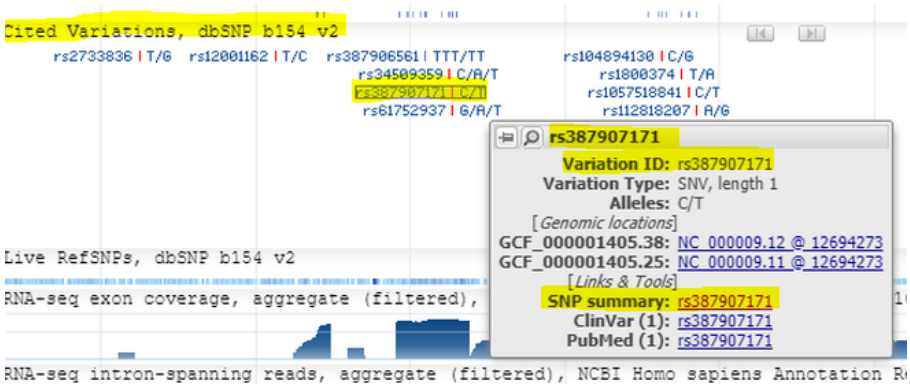

Figure 16: **TYRP1 rs387907171** Variation.

The table with all the SNP summary associated to this variation on the **TYRP1** gene can be seen in Figure 17. Figure 18 indicates the **R[CGC]>C[TGC] alteration, that corresponds to a skin, hair or eye pigmentation variation (code 11), more specifically to blond hair and if someone has blue eyes or not.** [11] [12].

Gene: TYRP1, tyrosinase related protein 1 (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
5,6-dihydroxyindole-2-carboxylic acid oxidase precursor	NP_000541.1:p.Arg93Cys	R (Arg) > C (Cys)	Missense Variant
TYRP1 transcript	NM_000550.3:c.277C>T	R [CGC] > C [TGC]	Coding Sequence Variant
TYRP1 transcript variant X1	XR_001746372.2:n.466C>T	N/A	Non Coding Transcript Variant

Figure 17: **TYRP1 rs387907171** SNP Summary.

Figure 18: **R[CGC]>C[TGC] variation:** Blond Hair and Blue/Non Blue Eyes.

Go back to the **TYRP1** Genome Browser page and on the *Cited Variations - dbSNP* section search the **rs387906561** variation id. To see its SNP summary, place the mouse on the variation id link and then click on the *SNP Summary* link (see Figure 19)

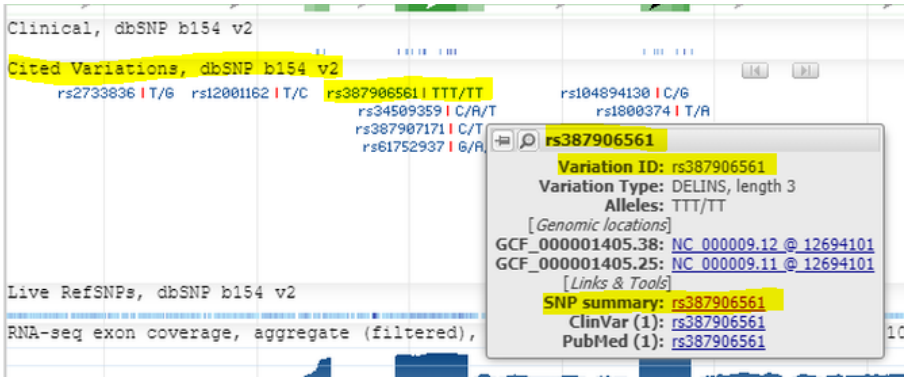

Figure 19: **TYRP1 rs387906561** Variation.

The table with all the SNP summary associated to this variation on the **TYRP1** gene can be seen in Figure 20. Figure 21 indicates the **L[TTG]>*[TG] alteration, that is pathogenic and corresponds to oculocutaneous albinism type 3** [13].

Gene: TYRP1, tyrosinase related protein 1 (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
5,6-dihydroxyindole-2-carboxylic acid oxidase precursor	NP_000541.1:p.Ala35_Leu36...	L (Leu) > * (Ter)	Stop Gained
TYRP1 transcript	NM_000550.3:c.107del	L [TTG] > * [TG]	Coding Sequence Variant
TYRP1 transcript variant X1	XR_001746372.2:n.296del	N/A	Non Coding Transcript Variant

Figure 20: **TYRP1 rs387906561** SNP Summary.

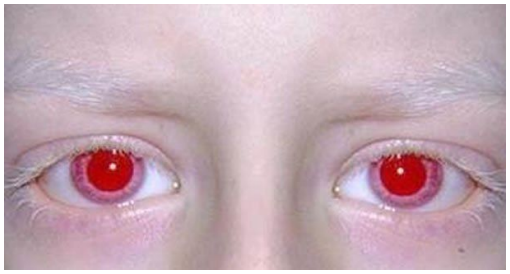

Figure 21: **L[TTG]>*[TG] variation:** Oculocutaneous Albinism Type III.

4 Conclusions and recommendations

Variations or mutations on the genes of an organism can cause simply effects such as the skin or eyes color, and very complex phenoms that could become into pathogenic diseases. It's very important to know the nature of a gene, among the functionality of its in order to have a better

idea and understanding of how its variations would behave or manifest. However, sometimes this isn't enough, because there are many cases where the results of an alteration seems to not be related to the main task of the corresponding gene.

5 References

[1] N. C. of Biotechnology Information, "GenBank," https://www.ncbi.nlm.nih.gov/genbank/, [Online; last access November 30, 2020].

[2] ——, "Gene," https://www.ncbi.nlm.nih.gov/gene/, [Online; last access November 30, 2020].

[3] GenBank, "MC1R melanocortin 1 receptor [Homo sapiens (human)]," https://www.ncbi.nlm.nih.gov/gene/4157, [Online; last access December 1, 2020].

[4] ——, "SLC12A2 solute carrier family 12 member 2 [Homo sapiens (human)]," https://www.ncbi.nlm.nih.gov/gene/6558, [Online; last access December 1, 2020].

[5] ——, "TYRP1 tyrosinase related protein 1 [Homo sapiens (human)]," https://www.ncbi.nlm.nih.gov/gene/7306, [Online; last access December 1, 2020].

[6] N. C. of Biotechnology Information, "dbSNP," https://www.ncbi.nlm.nih.gov/snp/, [Online; last access November 30, 2020].

[7] ——, "ClinVar Miner," https://clinvarminer.genetics.utah.edu/, [Online; last access December 1, 2020].

[8] ——, "PubMed," https://pubmed.ncbi.nlm.nih.gov/, [Online; last access November 30, 2020].

[9] C. Miner, "Submissions for variant NM_001046.3(SLC12A2):c.3076_3086del (p.Val1026fs) (rs886040968)," https://clinvarminer.genetics.utah.edu/submissions-by-variant/NM_001046.3%28SLC12A2%29%3Ac.3076_3086del%20%28p.Val1026fs%29, [Online; last access December 1, 2020].

[10] ——, "Submissions for variant NM_001046.3(SLC12A2):c.2935G>A (p.Glu979Lys) (rs1581138934)," https://clinvarminer.genetics.utah.edu/submissions-by-variant/NM_001046.3%28SLC12A2%29%3Ac.2935G%3EA%20%28p.Glu979Lys%29, [Online; last access December 1, 2020].

[11] ——, "Submissions for variant NM_000550.3(TYRP1):c.277C>T (p.Arg93Cys) (rs387907171)," https://clinvarminer.genetics.utah.edu/submissions-by-variant/NM_000550.3%28TYRP1%29%3Ac.277C%3ET%20%28p.Arg93Cys%29, [Online; last access December 1, 2020].

[12] N. C. of Biotechnology Information, "Skin/hair/eye pigmentation, variation in, 11," https://www.ncbi.nlm.nih.gov/gtr/conditions/C2677086/, [Online; last access December 1, 2020].

[13] C. Miner, "Submissions for variant NM_000550.3(TYRP1):c.107del (p.Ala35_Leu36insTer) (rs387906561)," https://clinvarminer.genetics.utah.edu/submissions-by-variant/NM_000550.3%28TYRP1%29%3Ac.107del%20%28p.Ala35_Leu36insTer%29, [Online; last access December 1, 2020].