



# Trabajo Práctico Especial

*Introducción a la Bioinformática*

---

**Integrantes:**

- Gutierrez, Julián (leg. 51141)
- Alderete, Facundo (leg. 51063)
- Medvedeff, Alexis (leg. 50066)
- Scigliano, Agustín (leg. 51277)

**Fecha de entrega:** 25 de noviembre de 2015

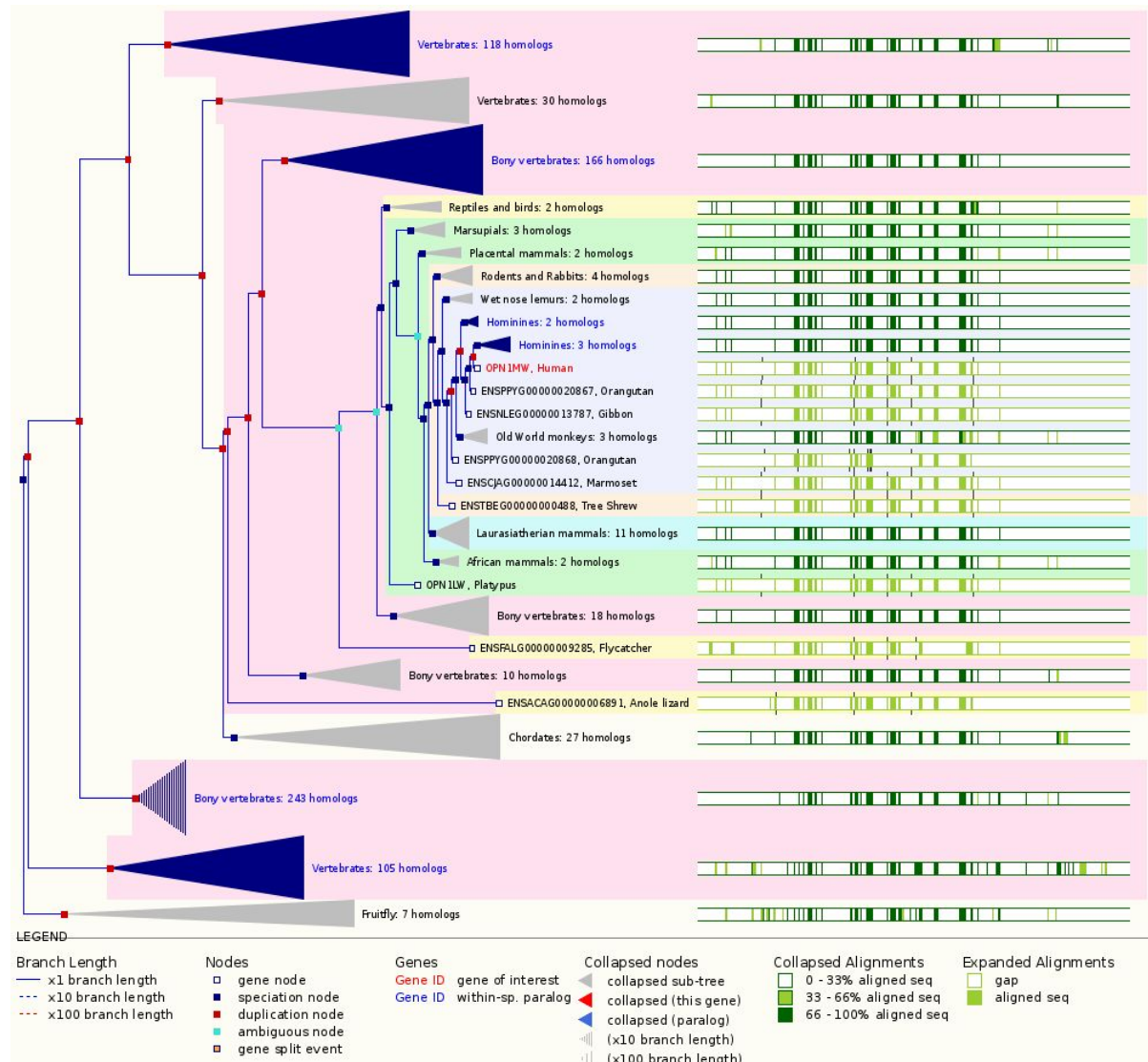
## Ejercicio 5

- a. NPC1 - [http://www.ncbi.nlm.nih.gov/nuccore/NM\\_000513.2](http://www.ncbi.nlm.nih.gov/nuccore/NM_000513.2)

El nombre oficial de este gen es "opsin 1 (cone pigments), medium-wave-sensitive", resumido como OPN1MW, y referenciado en NCBI bajo la secuencia NM\_000513.2. Este gen provee instrucciones para la creación de una proteína que es esencial para la normal percepción del color en la visión. Esta proteína se encuentra en la retina, que es el tejido fotosensible ubicado en el fondo del globo ocular. Existen varios tipos de receptores de la luz, que se diferencian uno del otro gracias al pigmento (opsina) que es sensible a distintas longitudes de onda de luz. OPN1MW provee instrucciones para la fabricación de la opsina que es sensible a las longitudes de onda amarillas y verdes (onda media).

Nos pareció interesante este gen porque es el responsable de la existencia de las personas llamadas "daltónicas", las cuales no son capaces de percibir los colores de manera normal, ya que impiden el correcto funcionamiento de los receptores de la luz presentes en el ojo.




- b. Gene Tree obtenido de la base de datos de ENSEMBL para el gen OPN1MW:



Desde el sitio de HomoloGene podemos pudimos obtener los genes y proteínas homólogos al nuestro:

### Genes

*Genes identified as putative homologs of one another during the construction of HomoloGene.*

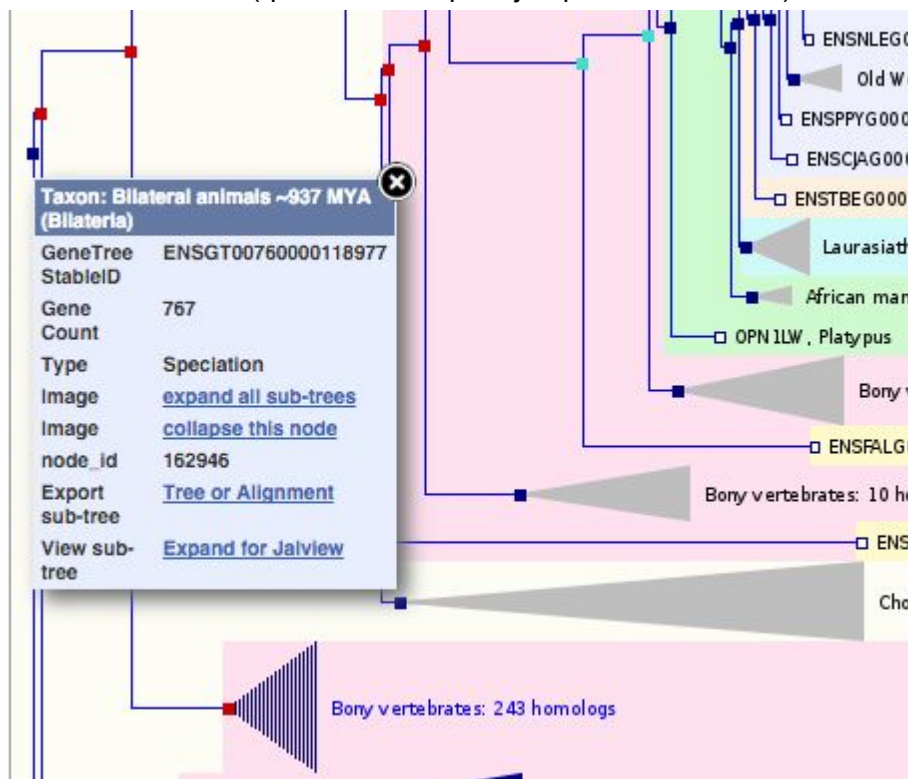
-  [OPN1MW, \*H.sapiens\*](#)  
opsin 1 (cone pigments), medium-wave-sensitive
-  [OPN1MW2, \*H.sapiens\*](#)  
opsin 1 (cone pigments), medium-wave-sensitive 2
-  [LOC101060233, \*H.sapiens\*](#)  
medium-wave-sensitive opsin 1-like

### Proteins

*Proteins used in sequence comparisons and their conserved domain architectures.*

-  [NP\\_000504.1](#)   
364 aa
-  [NP\\_001041646.1](#)   
364 aa
-  [XP\\_003960138.1](#)   
364 aa

Si expandimos la información de la base del árbol de ENSEMBL (haciendo clic sobre la misma) podemos ver que el grupo taxonómico al cual pertenece este gen son los animales con simetría bilateral (quedan fuera, por ejemplo, las medusas).



La principal diferencia que notamos entre las bases de datos HomoloGene y ENSEMBL es que la primera ofrece información muy reducida (sin ningún tipo de gráficos), mientras que la segunda ofrece gráficos interactivos, los cuales despliegan gran cantidad de información cuando se navega por ellos. El gen que elegimos parece estar presente en los animales que poseen una simetría física, más específicamente dos ojos, incluyendo todo tipo de vertebrados (reptiles, mamíferos, marsupiales, humanos) e incluso insectos, como las moscas de la fruta.

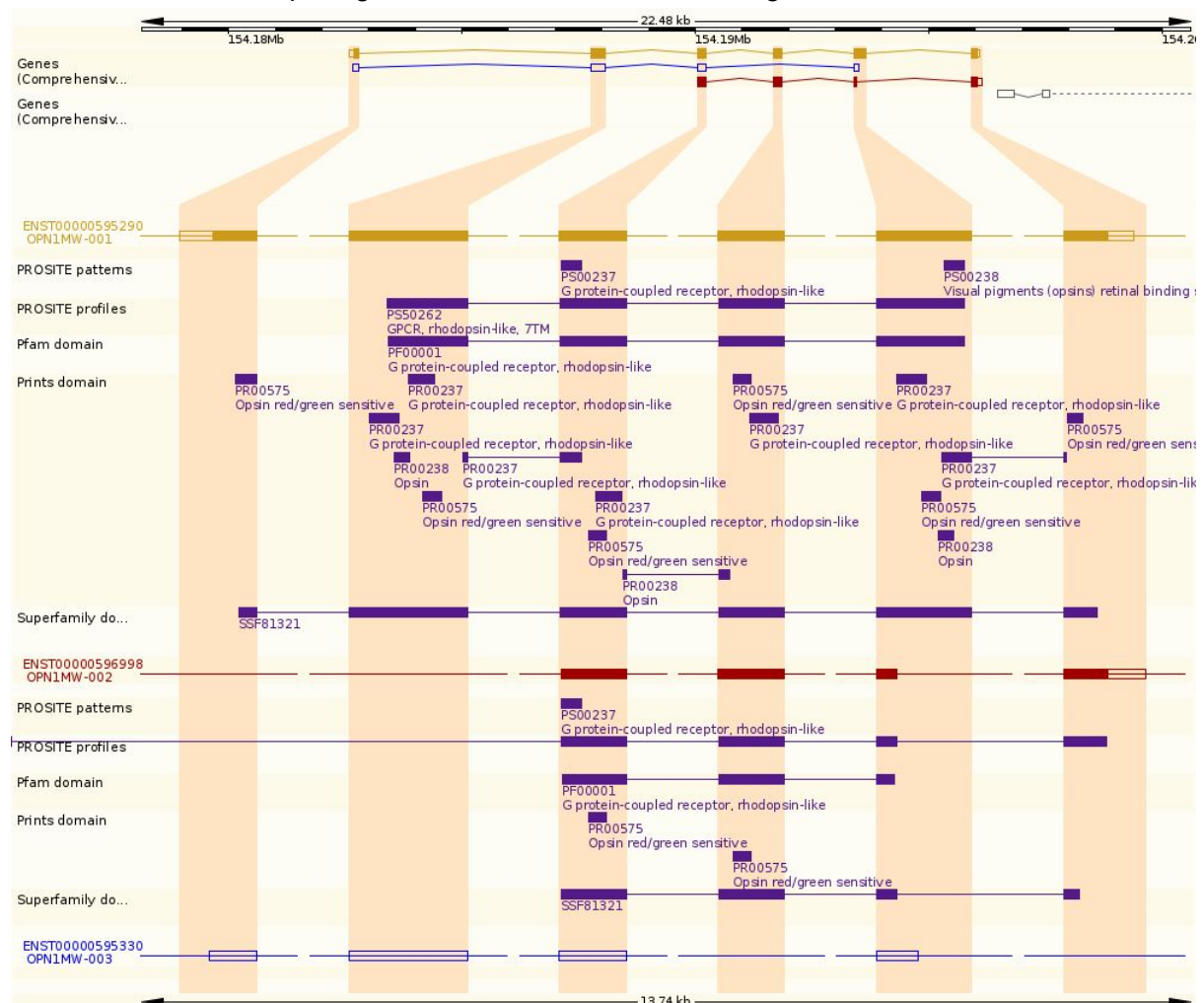
c. Existen 3 transcripts para nuestro gen en la base de datos ENSEMBL:

Transcripts

Hide transcript table

Show/hide columns (1 hidden)									Filter		
Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags			
OPN1MW-001	ENST00000595290	1242	364aa	Protein coding	CCDS14743	P04001	NM_000513 NP_000504	TSL:1 GENCODE basic APPRIS P1			
OPN1MW-002	ENST00000596998	588	164aa	Protein coding	-	H0Y642	-	CDS 5' incomplete TSL:5			
OPN1MW-003	ENST00000595330	692	No protein	Processed transcript	-	-	-	TSL:5			

En cuanto al Splicing, ENSEMBL nos muestra las siguientes variantes:



NCBI nos muestra únicamente dos resultados cuando buscamos las alternativas de splicing, como se ve en la siguiente imagen:

Gene sources

Genomic

Categories

Alternatively spliced

Annotated genes

Protein-coding

Sequence content

CCDS

Ensembl

RefSeq

Status

Current

Chromosome locations

more...

[Clear all](#)

[Show additional filters](#)

Tabular Sort by Relevance

Send to:

clear

Did you mean OPN1MW as a gene symbol?

Search Gene for [OPN1MW](#) as a symbol.

**Search results**

Items: 2

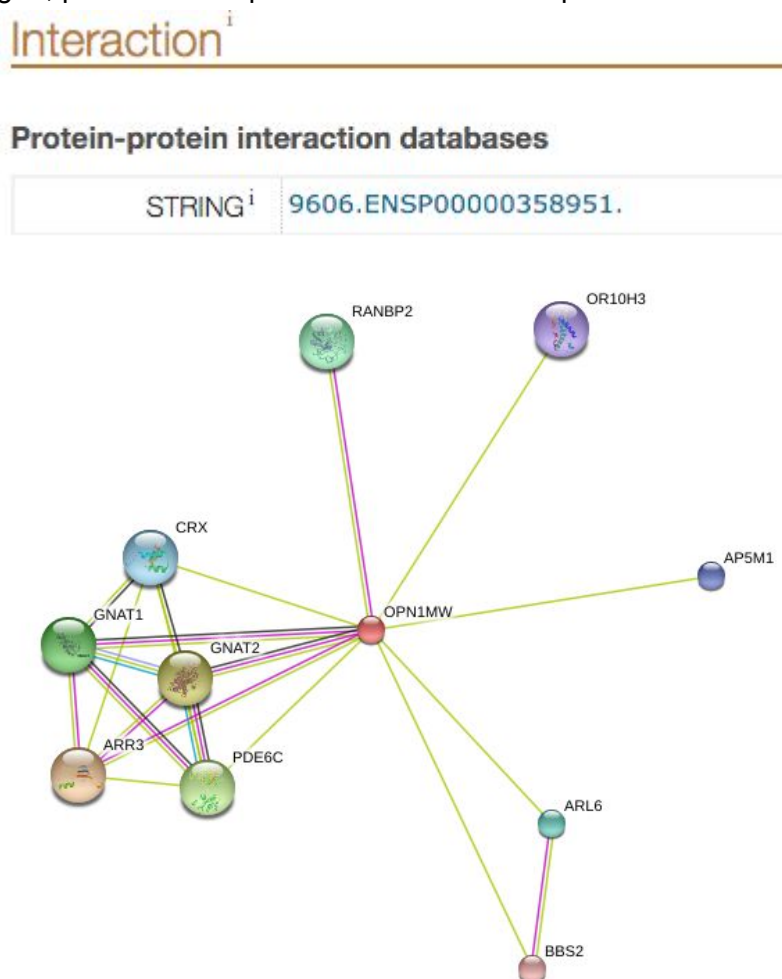
Filters activated: Alternatively spliced. [Clear all](#) to show 18 items.

Showing Current items.

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> <a href="#">Opn1mw</a> ID: 14539	opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan) [ <i>Mus musculus</i> (house mouse)]	Chromosome X, NC_000086.7 (74127466..74150756)	Gcp
<input type="checkbox"/> <a href="#">LOC741998</a> ID: 741998	long-wave-sensitive opsin 1 [ <i>Pan troglodytes</i> (chimpanzee)]	Chromosome X, NC_006491.3 (154826201..154845458)	OPN1LW

La forma en la que el sitio de ENSEMBL muestra la información vuelve a ser la más completa, y la que devuelve un resultado más que NCBI. Esto nos lleva a pensar que es ENSEMBL la base de datos más confiable y precisa.

- d. Desde la base de datos de Uniprot (<http://www.uniprot.org/uniprot/P04001>) para nuestro gen, podemos ver que interactúa con otras proteínas de la siguiente manera:





### Your Input:

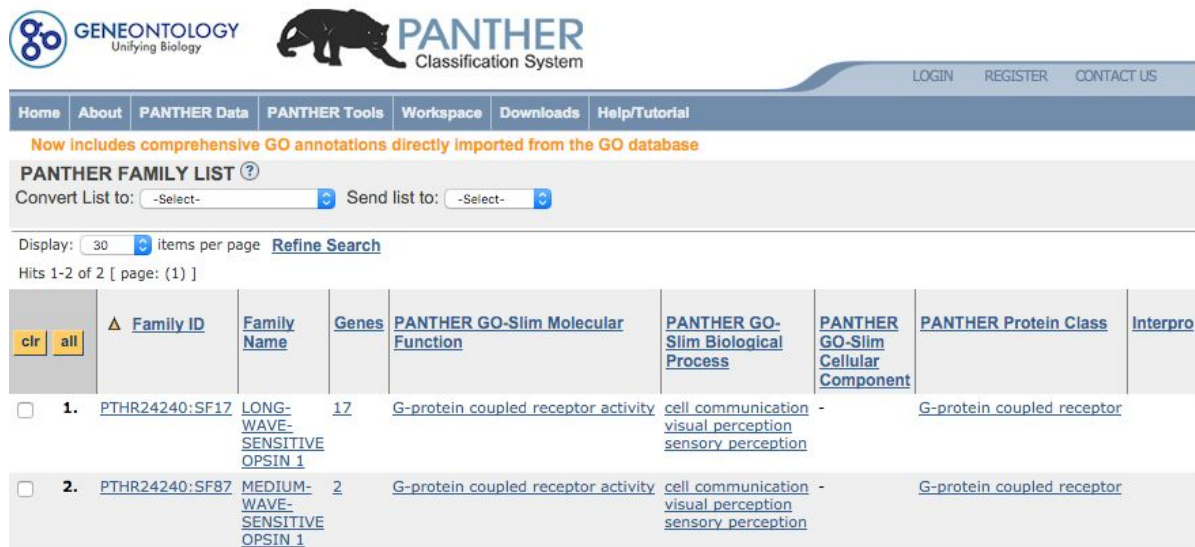
● OPN1MW opsin 1 (cone pigments), medium-wave-sensitive (364 aa)  
(Homo sapiens)

### Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
● ARR3	arrestin 3, retinal (X-arrestin); May play a role in an as yet undefined retina-specific signal [...]									0.850
● GNAT2	guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 2; Guani [...]									0.732
● PDE6C	phosphodiesterase 6C, cGMP-specific, cone, alpha prime (858 aa)									0.674
● GNAT1	guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 1; Guani [...]									0.620
● RANBP2	RAN binding protein 2; E3 SUMO-protein ligase which facilitates SUMO1 and SUMO2 conjugation by [...]									0.603
● ARL6	ADP-ribosylation factor-like 6; Involved in membrane protein trafficking at the base of the cil [...]									0.560
● CRX	cone-rod homeobox; Binds and transactivates the sequence 5'-TAATC[CA]-3' which is found upstrea [...]									0.525
● AP5M1	adaptor-related protein complex 5, mu 1 subunit; As part of AP-5, a probable fifth adapter prot [...]									0.480
● OR10H3	olfactory receptor, family 10, subfamily H, member 3; Odorant receptor (Potential) (316 aa)									0.472
● BBS2	Bardet-Biedl syndrome 2; The BBSome complex is required for ciliogenesis but is dispensable for [...]									0.451

Descubrimos que no se poseen entradas acerca de la interacción con otras proteínas en la base de datos de NCBI, por lo que los resultados se encuentran únicamente en Uniprot. Es aquí donde obtuvimos los gráficos que mostramos arriba, junto con las referencias. Se explica claramente cuáles son las proteínas con las que interactúa nuestro gen, y en qué consiste dicha interacción.

- e. El sistema de clasificación Panther (<http://www.pantherdb.org/>), el cual forma parte del entorno Gene Ontology, arroja la siguiente información al ingresar nuestro gen:



The screenshot shows the PANTHER Classification System interface. At the top, there are logos for GENE ONTOLOGY and PANTHER. Below the logos, there is a navigation bar with links: Home, About, PANTHER Data, PANTHER Tools, Workspace, Downloads, and Help/Tutorial. The main content area displays the PANTHER FAMILY LIST for the input gene OPN1MW. The list shows two hits, both of which are G-protein coupled receptor activity. The first hit is PTHR24240:SF17, and the second is PTHR24240:SF87. Both hits are associated with the same GO term: G-protein coupled receptor activity. The interface also includes a sidebar with links to various PANTHER tools and a search bar.

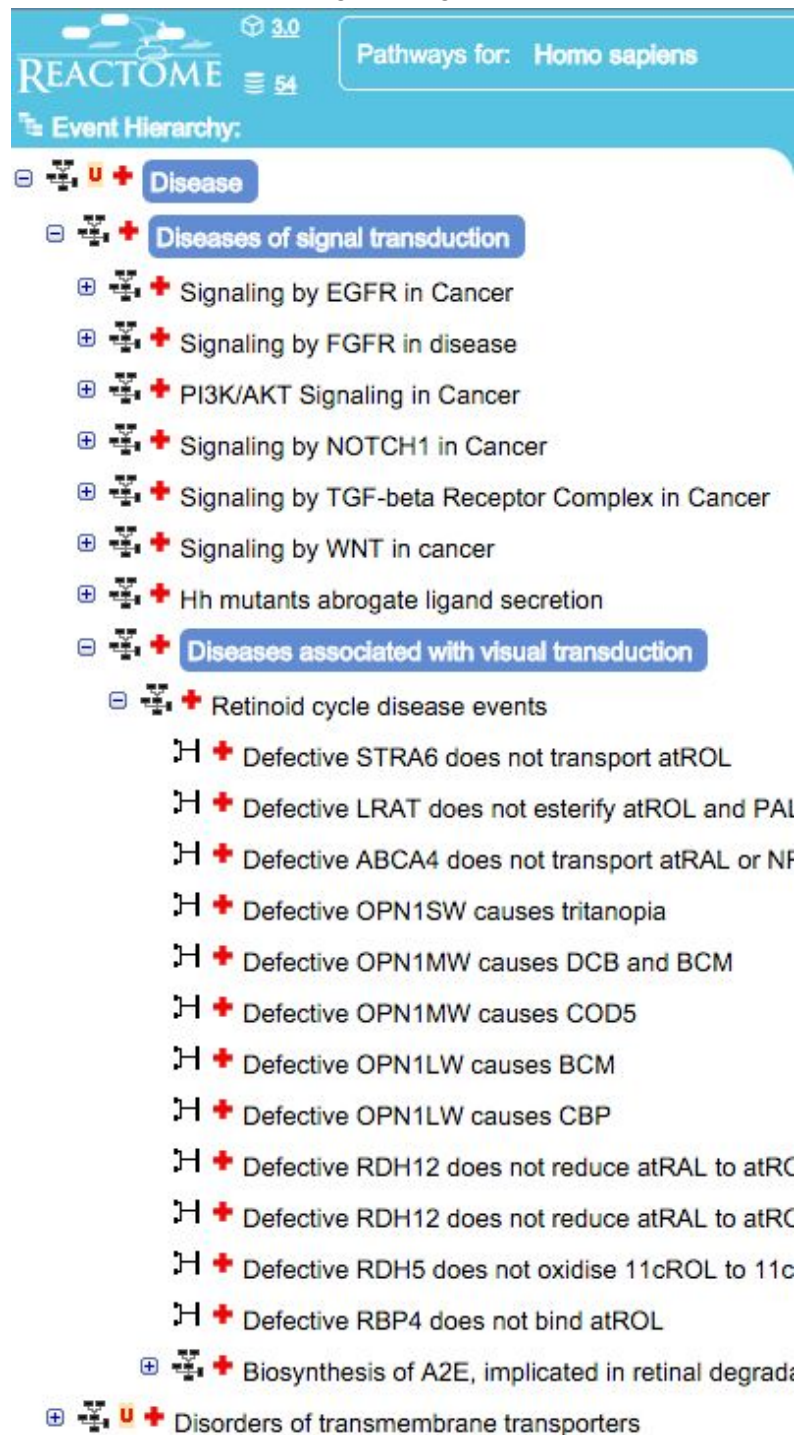
clr	all	Family ID	Family Name	Genes	PANTHER GO-Slim Molecular Function	PANTHER GO-Slim Biological Process	PANTHER GO-Slim Cellular Component	PANTHER Protein Class	Interpro
<input type="checkbox"/>	<input type="checkbox"/>	1. PTHR24240:SF17	LONG-WAVE-SENSITIVE OPSIN 1	17	G-protein coupled receptor activity	cell communication - visual perception sensory perception		G-protein coupled receptor	
<input type="checkbox"/>	<input type="checkbox"/>	2. PTHR24240:SF87	MEDIUM-WAVE-SENSITIVE OPSIN 1	2	G-protein coupled receptor activity	cell communication - visual perception sensory perception		G-protein coupled receptor	

Como podemos ver, no se especifica un componente celular para nuestra proteína. Los procesos biológicos a los que se hace referencia son:

- Comunicación celular: cualquier interacción entre una célula y su entorno. Puede involucrar envío de señales o unión entre dos o más células.
- Percepción visual: serie de eventos requerida para que un organismo reciba un estímulo visual, lo convierta en una señal molecular y reconozca las características de dicha señal. Los estímulos visuales se reciben como fotones y se procesan en el cerebro para formar una imagen.
- Percepción sensorial: serie de eventos requerida para que un organismo reciba un estímulo sensorial, lo convierta en una señal molecular e interprete dicha señal. Es un proceso neurológico.

La función molecular en la que trabaja nuestro gen es "G-protein coupled receptor activity", la cual consiste en combinar una señal extracelular (luz en este caso) y transmitir dicha señal por la membrana activando la proteína G asociada. Permite el intercambio de GDP por GTP en la subunidad alfa de un complejo de proteínas G heterotriméricas.

- f. La base de datos de pathways Reactome (<http://www.reactome.org/PathwayBrowser/>) nos indica que nuestro gen está en la familia de enfermedades que consisten en una deficiencia en la traducción de señales visuales, específicamente las correspondientes a la retina. Esta jerarquía de eventos se ve claramente en el siguiente gráfico:



- g. Buscamos en la base de datos de variantes genéticas una variante para la cual haya información de frecuencia de la misma (existen muchas para nuestro gen que no tienen frecuencia asociada), y nos quedamos con la variación rs183474841 ([http://www.ncbi.nlm.nih.gov/projects/SNP/snp\\_ref.cgi?rs=183474841](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=183474841)). Podemos ver que se trata de una variación de tipo SNV (single nucleotide variation) para una molécula de tipo genómica en humanos. La información sobre la frecuencia que tiene esta variante en la población es la siguiente:

Population Diversity (Alleles in RefSNP orientation) . See additional population frequency from 1000Genome [\[here\]](#)

ss#	Sample Ascertainment			HWP	Alleles	
	Population	Individual Group	Chrom. Sample Cnt.		C	T
ss1556692296	<a href="#">EAS</a>		1008	AF	1.000	
	<a href="#">EUR</a>		1006	AF	1.000	
	<a href="#">AFR</a>		1322	AF	0.922	0.078
	<a href="#">AMR</a>		694	AF	0.997	0.003
	<a href="#">SAS</a>		978	AF	1.000	
Summary	Average Het. +/- std err:	Individual Count	Founders Count	Individual Overlap	Genotype Conflict	
	0.054 +/- 0.155	0	0	0	0	

El grupo étnico más afectado parecen ser las personas del continente africano.