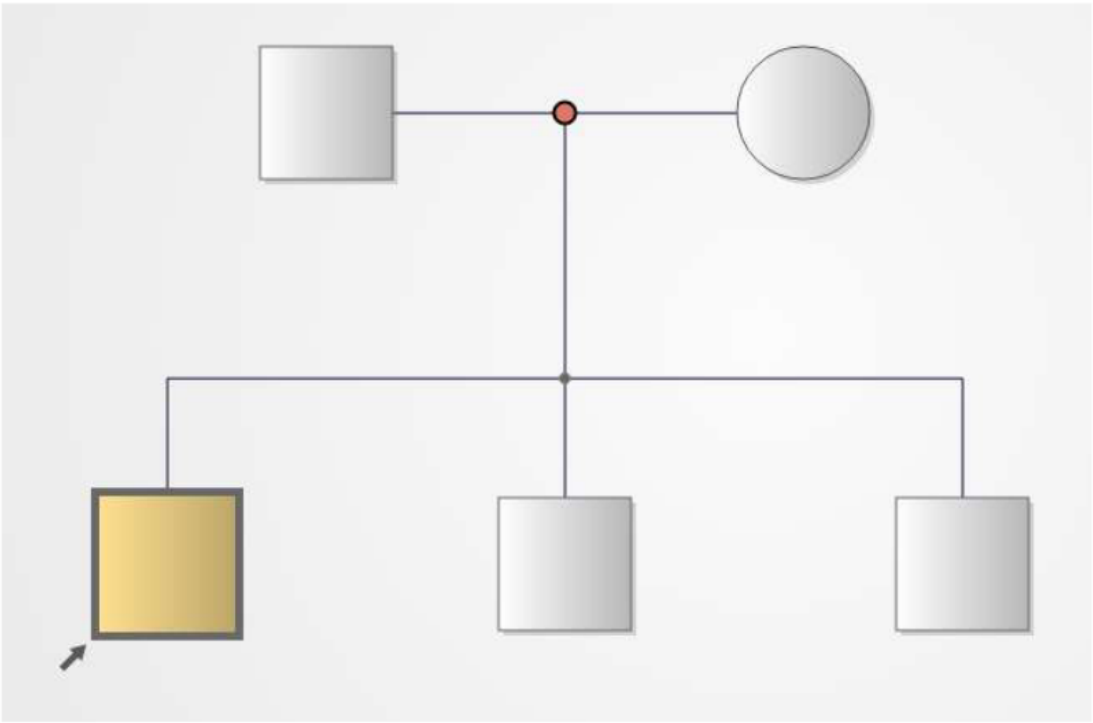
Family_11

Onna Nayyu Leyva Alcantara

2023-10-05



Phenotype: High bile acids, itchiness, hypercholesterolemia

Family:

Affected child with two unaffected, non-carrier siblings, as well as parents.

Symptoms:

High bile acids.

Itching or plurite.

Hypercholesterolemia or high cholesterol.

Data Set:

```
setwd("~/Documentos/a5/Clau/tareaejerc/Tar_2")

Datos = read.csv("FAM11.csv")
# 22,668 variantes con 52 variables.
```

There are 22,668 variants in the exome.

Data Processing:

Coverage, Quality Control (QC) and Allelic Balance.

Tr: Coverage >= 10.

```
Profundidad = subset(Datos, TR >= 10)
  # Vamos a dejar en "profundidad" lo mayor o igual a "10".
# 22,056
```

GQ: Genotype Quality Control >= 30.

```
Quality = subset(Profundidad, GQ >= 30)
# 21,885
```

Ratio: Allelic Balance >= 20 percent.

```
Balancealelico = subset(Quality, RATIO >= 0.20)
# 21,344
```

Filter by annotation:

We will remove by variant type: In this case synonyms.

```
Sinsinonimas = subset(Balancealelico, ANNOTATION != "synonymous SNV")
# 10,793
```

Frequency: Eve_Alt_Freq <= 0.01 or to 1 percent.

```
Filtrado = subset(Sinsinonimas, EVE_ALT_FREQ <= 0.01)
# Frecuencia del alelo alternativo.
# 661
```

Renaming of the columns because "R" does not recognize the originals.

```
##  [1] "X.CHR"          "START"          "END"            "REF"
##  [5] "ALT"            "ZYG"            "FILTER"         "QUAL"
##  [9] "RR"             "VR"             "TR"             "RATIO"
## [13] "GQ"             "PL"             "VCF_INFO"       "BAND"
## [17] "SNP17NF"        "ANNOTATION"     "GENE"           "GENE_NAME"
## [21] "ACCESSION"      "EXON"           "NT_CHANGE"      "AA_CHANGE"
## [25] "ALL_TRX"        "TGP_FREQ"       "ESP_FREQ"       "GRANTHAM_SC"
## [29] "PHASTCONS"      "GERP_ESP"       "GERP"           "PHYLOP_SC"
## [33] "PHYLOP_PRED"    "LRT_SCORE"      "LRT_PRED"       "SIFT_SCORE"
## [37] "SIFT_PRED"      "PPH2_SCORE"     "PPH2_PRED"      "MTT_SCORE"
## [41] "MTT_PRED"       "ESP_PPH2"       "CHIMP"          "EVS_ALLELES"
## [45] "ESP_GENOTYPES" "EVE_ALT_FREQ"  "EVE_HOM"        "Proband.ZYG"
## [49] "Father.ZYG"     "Mother.ZYG"     "UaSib.ZYG"      "UaSib.ZYG.1"
```

Filter by Segregation:

Filtering for a recessive disease:

To clarify the next code is not the better way to codify it, but it's very visual and easy to understand.

```
# Si los cinco integrantes de la familia son homocigotos:
Nohomoall = subset(Filtrado,
                         !(Probandzig == "Proband:hom" &
                           Fatherzig == "Father:hom" &
                           Motherzig == "Mother:hom" &
                           Unafectedzig == "UaSib:hom" &
                           Unafectedzig_1 == "UaSib:hom"))
# 586

# Si los tres hermanos son homocigotos:
Nohomohermas = subset(Nohomoall,
                         !(Probandzig == "Proband:hom" &
                           Unafectedzig == "UaSib:hom" &
                           Unafectedzig_1 == "UaSib:hom"))
# 579

# Si el probando y un hermano son homocigotos:
Nohomoherma = subset(Nohomohermas,
                         !(Probandzig == "Proband:hom" &
                           Unafectedzig == "UaSib:hom"))
# 565

# Si el probando y el otro hermano son homocigotos:
Nohomoherma2 = subset(Nohomoherma,
                         !(Probandzig == "Proband:hom" &
                           Unafectedzig_1 == "UaSib:hom"))
# 553
```

```
# Si el probando y padre son homocigotos:
Nohomopadre = subset(Nohomoherma2,
                          !(Probandzig == "Proband:hom" &
                            Fatherzig == "Father:hom"))
# 546

# Si el probando y madre son homocigotos:
Nohomomadre = subset(Nohomopadre,
                          !(Probandzig == "Proband:hom" &
                            Motherzig == "Mother:hom"))
# 543
```

Only the proband is homozygous:

```
homo = subset(Nohomomadre, Probandzig == "Proband:hom")
# 11
```

Before finishing, in this file I only show the homozygous filtering, although, other types of segregation where searched.

So if we look for an autosomical recesive disease, we obtain 11 related variants, six non-frameshift insertions, two nonsynonymous SNV, two frameshift deletion and one splicing. Although I check everyone, the most striking ones were the last three, and looking for that variants I found that one of the nonsynonymous SNV in the tight junction protein 2 (TJP2) gene cause the symptoms that we were looking for.