

Rentrez. Ejercicio guiado 3

Mario Pascual González

Enunciado: Estudio de la Fibrosis Quística

La Fibrosis Quística es una enfermedad genética hereditaria que afecta principalmente a los pulmones y al sistema digestivo. Es causada por mutaciones en el gen CFTR (Cystic Fibrosis Transmembrane Conductance Regulator). Usando los recursos aprendidos de **rentrez**, resuelve las siguientes cuestiones relacionadas con el gen CFTR y la enfermedad. Asegúrate de incluir el código utilizado para obtener cada respuesta.

Primer ejercicio

1. Identificación del Gen CFTR

- Encuentra el ID de Entrez para el gen CFTR en la especie Homo Sapiens.

Resolución Primero debemos encontrar el parámetro de la base de datos **gene** que hace referencia al tipo de gen

```
library(rentrez)
entrez_db_searchable("gene")
```

```
## Searchable fields for database 'gene'
## ALL      All terms from all searchable fields
## UID      Unique number assigned to a gene record
## FILT     Limits the records
## TITL     gene or protein name
## WORD     Free text associated with record
## ORGN     scientific and common names of organism
## MDAT     The last date on which the record was updated
## CHR      Chromosome number or numbers; also 'mitochondrial', 'unknown' properties
## MV       Chromosomal map location as displayed in MapViewer
## GENE      Symbol or symbols of the gene
## ECNO     EC number for enzyme or CAS registry number
## MIM      MIM number from OMIM
## DIS      Name(s) of diseases associated with this gene. When available, OMIM name will be used
## ACCN     Nucleotide or protein accession(s) associated with this gene
## UGEN     UniGene cluster number for this gene
## PROP     Properties of Gene record
## CDAT     The date on which this record first appeared
## NCAC     nucleotide accessions of sequences
## NUID     nucleotide uids of sequences
## PACC     protein accessions
## PUID     protein uids
## PMID     PubMed ids of accessions linked to the record
```

```
## TID      taxonomy id
## GO       Gene Ontology
## DOM      Domain Name
## DDAT     The date on which the record was discontinued
## CPOS     Chromosome base position
## GFN      Gene full name
## PFN      Protein full name
## GL       Gene length
## XC       Exon count
## GRP      Relationships for this gene
## PREF     Preferred symbol of the gene
## AACCC    Assembly accession
## ASM      Assembly name
## EXPR     Gene expression
```

Se utilizará TITL para el nombre del gen, mientras que se usará ORGN para especificar la especie.

```
search <- entrez_search(db = "gene",
                        term = "CFTR[TITL] AND Homo Sapiens[ORGN]",
                        retmax = 100)
summary(search)
```

```
##              Length Class              Mode
## ids          75      -none-             character
## count        1      -none-             numeric
## retmax       1      -none-             numeric
## QueryTranslation 1      -none-             character
## file         1      XMLInternalDocument externalptr
```

```
print(paste("Tenemos un total de", search$count, "ids para el gen CFTR", sep = " "))
```

```
## [1] "Tenemos un total de 75 ids para el gen CFTR"
```

Ahora debemos encontrar el que se llama exactamente CFTR.

```
ids_gen_correcto <- search$ids[sapply(search$ids, function(id) {"CFTR" == entrez_summary(db = "gene", id = id)$TITL})]
print(paste("ID del gen que buscábamos: ", ids_gen_correcto))
```

```
## [1] "ID del gen que buscábamos: 1080"
```

Ejercicio 2

2. Localización Cromosómica y Coordenadas Genéticas

- Determina en qué cromosoma se encuentra el gen CFTR y sus coordenadas genéticas exactas.

```
summ_gen <- entrez_summary(db = "gene",
                           id = ids_gen_correcto)
print(paste("Cromosoma:", summ_gen$genomicinfo$chrloc))
```

Resolución

```
## [1] "Cromosoma: 7"

print(paste("Coordenadas de inicio dentro del cromosoma:", summ_gen$genomicinfo$chrstart))

## [1] "Coordenadas de inicio dentro del cromosoma: 117480024"

print(paste("Coordenadas de fin dentro del cromosoma:", summ_gen$genomicinfo$chrstop))

## [1] "Coordenadas de fin dentro del cromosoma: 117668664"
```

Ejercicio 3

3. Variantes Clínicas Asociadas

- Calcula cuántas variantes clínicas relacionadas con la Fibrosis Quística están registradas en ClinVar para el gen CFTR.

```
links <- entrez_link(dbfrom = "gene",
                    db = "clinvar",
                    id = ids_gen_correcto)
print(paste("Hay asociadas", length(links$links$gene_clinvar), "variantes clínicas registradas en clinvar"))
```

Resolución

```
## [1] "Hay asociadas 4832 variantes clínicas registradas en clinvar del gen CFTR"
```

Ejercicio 4

4. Investigación de Transcritos

- Identifica cuántos transcritos del gen CFTR están disponibles en la base de datos de NCBI.

```
busq <- entrez_link(dbfrom = "gene",
                  db = "nuccore",
                  id = ids_gen_correcto)
print(paste("Cantidad de transcritos:", length(busq$links$gene_nuccore), sep = " "))
```

Resolución

```
## [1] "Cantidad de transcritos: 230"
```

Ejercicio 5

5. Secuenciación y Análisis

- Recupera la secuencia FASTA de los transcritos del gen CFTR.

```
link_fasta <- busq$links$gene_nucore_refseqrna  
  
fasta <-entrez_fetch(db = "nucore",  
                    id = link_fasta,  
                    rettype = "FASTA")  
write(fasta, file = "/home/mario/Universidad/Proyectos-R/Rentrez-R/FastaR-1080.fasta")
```

Resolución Podemos ver los contenidos de este fichero fasta tal que:

```
cd /home/mario/Universidad/Proyectos-R/Rentrez-R/FastaR-1080.fasta cat FastaR-1080.fasta
```

Y contar la cantidad de lecturas del gen tal que:

```
cat FastaR-1080.fasta | grep -c ">"
```

Ejercicio 6

6. Exploración de Publicaciones

- Encuentra los cinco artículos más recientes sobre la Fibrosis Quística relacionados con el gen CFTR en PubMed.

Resolución Primero tendremos que encontrar la opción que tiene pubmed para filtrar por los artículos que contengan información sobre el gen CFTR.

```
entrez_db_searchable("pubmed")
```

```
## Searchable fields for database 'pubmed'  
## ALL      All terms from all searchable fields  
## UID      Unique number assigned to publication  
## FILT     Limits the records  
## TITL     Words in title of publication  
## MESH      Medical Subject Headings assigned to publication  
## MAJR     MeSH terms of major importance to publication  
## JOUR      Journal abbreviation of publication  
## AFFL     Author's institutional affiliation and address  
## ECNO     EC number for enzyme or CAS registry number  
## SUBS     CAS chemical name or MEDLINE Substance Name  
## PDAT     Date of publication  
## EDAT     Date publication first accessible through Entrez  
## VOL      Volume number of publication  
## PAGE     Page number(s) of publication  
## PTYP     Type of publication (e.g., review)  
## LANG     Language of publication  
## ISS      Issue number of publication
```

```

## SUBH Additional specificity for MeSH term
## SI Cross-reference from publication to other databases
## MHDA Date publication was indexed with MeSH terms
## TIAB Free text associated with Abstract/Title
## OTRM Other terms associated with publication
## COLN Corporate Author of publication
## CNTY Country of publication
## PAPX MeSH pharmacological action pre-explosions
## GRNT NIH Grant Numbers
## MDAT Date of last modification
## CDAT Date of completion
## PID Publisher ID
## FAUT First Author of publication
## FULL Full Author Name(s) of publication
## FINV Full name of investigator
## TT Words in transliterated title of publication
## LAUT Last Author of publication
## PPDT Date of print publication
## EPDT Date of Electronic publication
## LID ELocation ID
## CRDT Date publication first accessible through Entrez
## BOOK ID of the book that contains the document
## ED Section's Editor
## ISBN ISBN
## PUBN Publisher's name
## AUCL Author Cluster ID
## EID Extended PMID
## DSO Additional text from the summary
## AUID Author Identifier
## PS Personal Name as Subject
## COIS Conflict of Interest Statements
## WORD Free text associated with publication
## P1DAT Date publication first accessible through Solr

```

Ahora podemos realizar la búsqueda:

```

search_pubmed <- entrez_search(db = "pubmed",
                               term = "CFTR[TIAB] AND (2024[PDAT])",
                               retmax = 5)

summary(search_pubmed)

```

```

##          Length Class          Mode
## ids          5    -none-      character
## count        1    -none-      numeric
## retmax       1    -none-      numeric
## QueryTranslation 1    -none-      character
## file         1    XMLInternalDocument externalptr

```

```

print(paste("Cantidad de artículos del 2024 sobre el gen CFTR:", search_pubmed$count))

```

```

## [1] "Cantidad de artículos del 2024 sobre el gen CFTR: 150"

```

```

if(length(search_pubmed$ids) > 0) {
  summaries <- entrez_summary(db="pubmed", id=search_pubmed$ids)

  # Imprimir títulos y autores (o cualquier información relevante) de los artículos
  lapply(summaries, function(s) {
    cat("Título:", s$title, "\n")
    cat("Autores:", paste(s$authors, collapse=" "), "\n\n")
  })
}

```

```

## Título: One year effect of tezacaftor and ivacaftor on functional exercise capacity and muscle strength
## Autores: c("Anne-Sophie A", "Penelle M", "Clémence G", "Berardis S", "Goubau C", "Reychler G", "Gohy")
##
## Título: The Rectal Gland of the Shark: The Road to Understanding the Mechanism and Regulation of Transcription
## Autores: c("Silva P", "Evans DH"), c("Author", "Author"), c("", "")
##
## Título: Year in review 2023 - Back to the future.
## Autores: c("Cristiani L", "Fernandes FF"), c("Author", "Author"), c("", "")
##
## Título: Cystic fibrosis and fat malabsorption: Pathophysiology of the cystic fibrosis gastrointestinal
## Autores: c("McDonald CM", "Reid EK", "Pohl JF", "Yuzyuk TK", "Padula LM", "Vavrina K", "Altman K"), c("", "")
##
## Título: A uniquely efficacious type of CFTR corrector with complementary mode of action.
## Autores: c("Marchesin V", "Monnier L", "Blattmann P", "Chevallard F", "Kuntz C", "Forny C", "Kamperling J")
##
## $'38434346'
## NULL
##
## $'38433340'
## NULL
##
## $'38431442'
## NULL
##
## $'38429959'
## NULL
##
## $'38427726'
## NULL

```