

## 1. Introducción

En esta actividad se ha diseñado una base de datos relacional orientada al ámbito de la genómica, cuyo objetivo es representar de forma estructurada información sobre genes, sus secuencias, variantes genéticas y anotaciones clínicas. Cada gen incluye una descripción, una localización cromosómica y varias secuencias (exones, intrones, etc.). Además, se registran variantes con sus alelos y tipo, y se asocian anotaciones que aportan contexto biológico o médico. La base de datos también integra estudios científicos relacionados, asegurando integridad mediante restricciones específicas y buenas prácticas en el diseño e implementación del mismo.

## 2. Diagrama Entidad-Relación

El diagrama refleja cómo se organiza la información genética en la base de datos, destacando las siguientes entidades y relaciones:

### Entidades principales:

#### 3. Gen

- a. Identificado por **ID gen**.
- b. Atributos: nombre, descripción y localización cromosómica.
- c. Tiene una relación **1:N** con **Secuencia** (un gen contiene múltiples secuencias).
- d. Se relaciona con **Variantes genéticas** (presenta variantes múltiples) y puede estar anotado.

#### 4. Secuencia

- a. Atributos: **ID secuencia**, tipo (exón, intrón, etc.), representación de ADN, posición relativa al gen.
- b. Está vinculada a un único gen mediante la relación **Contiene**.

#### 5. Variantes genéticas

- a. Atributos: **ID variante**, tipo de variante (SNP, inserción, delección), alelo mutado, alelo referencia y posición relativa.
- b. Relacionada con **Gen** por medio de **Presenta** (un gen puede tener muchas variantes).
- c. Puede ser objeto de anotaciones y de estudios científicos.

## 6. Estudios

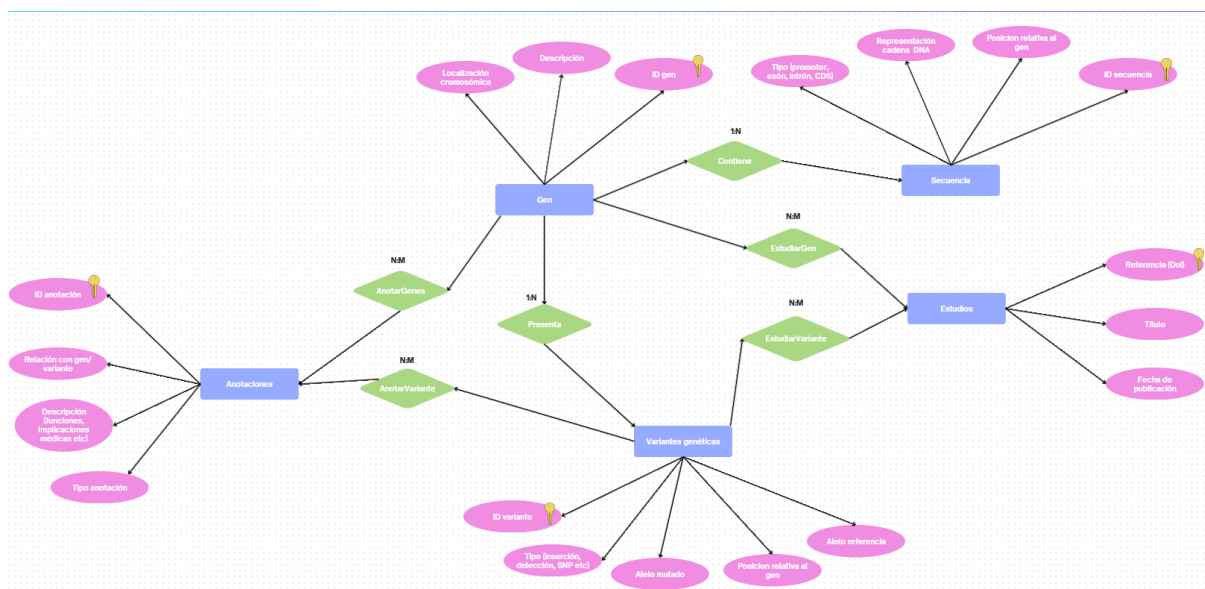
- Atributos: **Referencia (DOI)** como clave primaria, título y fecha de publicación.
- Se vincula con genes y variantes por relaciones N:M (**EstudiarGen** y **EstudiarVariante**).

## 7. Anotaciones

- Atributos: **ID anotación**, tipo (clínica, funcional), descripción y relación con gen o variante.
- Puede aplicarse tanto a genes como a variantes mediante relaciones N:M (**AnotarGenes** y **AnotarVariante**).

## Relaciones entre entidades:

- Contiene**: entre **Gen** y **Secuencia**, 1:N.
- Presenta**: entre **Gen** y **Variantes genéticas**, 1:N.
- EstudiarGen** y **EstudiarVariante**: relaciones N:M entre **Estudios** y **Gen/Variantes**.
- AnotarGenes** y **AnotarVariante**: relaciones N:M que permiten asociar **Anotaciones** tanto a genes como a variantes.



### 3. Esquema Modelo relacional

#### 1. Tabla: **gene**

Contiene la información básica de los genes identificados.

- **Clave primaria:** **GENE\_ID**: nombre del gen (obligatorio, no nulo)
- **Atributos:**
  - **SUMMARY**: descripción del gen (obligatoria)
  - **CHROMOSOMIC\_POSITION**: posición en el cromosoma

#### 2. Tabla: **sequence**

Almacena las distintas secuencias asociadas a cada gen.

- **Clave primaria:** **SEQUENCE\_ID** (valor no nulo)
- **Clave foránea:** **GENE\_ID** → **gene**
- **Atributos:**
  - **POSITION\_RELATIVE\_TO\_GENE**: posición entera positiva (no nula)
  - **DNA\_CHAIN**: cadena de ADN (longitud entre 10 y 1000 caracteres)
  - **SEQ\_TYPE**: tipo de secuencia (exón, intrón, CDS, promotor, etc.)

#### 3. Tabla: **variants**

Representa las variantes genéticas asociadas a los genes.

- **Clave primaria:** **VAR\_ID**
- **Clave foránea:** **GENE\_ID** → **gene**
- **Atributos:**
  - **VAR\_TYPE**: tipo de variante (SNP, inserción, delección...)

- **REFERENCE\_ALLELE**: alelo de referencia (valor por defecto: "-")
- **MUTATED\_ALLELE**: alelo mutado (valor por defecto: "-")
- **POSITION\_RELATIVE\_TO\_GENE**: posición relativa al gen (entero positivo)

#### 4. Tabla: **annotations**

Incluye información adicional sobre genes o variantes, como funciones biológicas o implicaciones clínicas.

- **Clave primaria**: **ANNOTATION\_ID**
- **Atributos**:
  - **ANNOT\_TYPE**: tipo de anotación (funcional, clínica, etc.)
  - **SUMMARY**: resumen o descripción de la anotación
  - **RELATION\_TO\_GENE**: relación con el gen o la variante

#### 5. Relación: **geneannotations**

Tabla intermedia que relaciona genes con anotaciones.

- **Claves foráneas**:
  - **ANNOTATION\_ID** → annotations
  - **GENE\_ID** → gene
- **Cardinalidad**: N:M

#### 6. Relación: **annotationvariant**

Relaciona variantes genéticas con sus anotaciones correspondientes.

- **Claves foráneas**:
  - **ANNOTATION\_ID** → annotations

- VAR\_ID → variants
- **Cardinalidad:** N:M

## 7. Tabla: studies

Almacena estudios científicos que analizan genes o variantes.

- **Clave primaria:** DOI
- **Atributos:**
  - TITLE: título del estudio
  - PUBLICATION\_DATE: fecha de publicación
  - DOI: referencia única (se espera el formato `aaaa/111`)

## 8. Relación: variantstudy

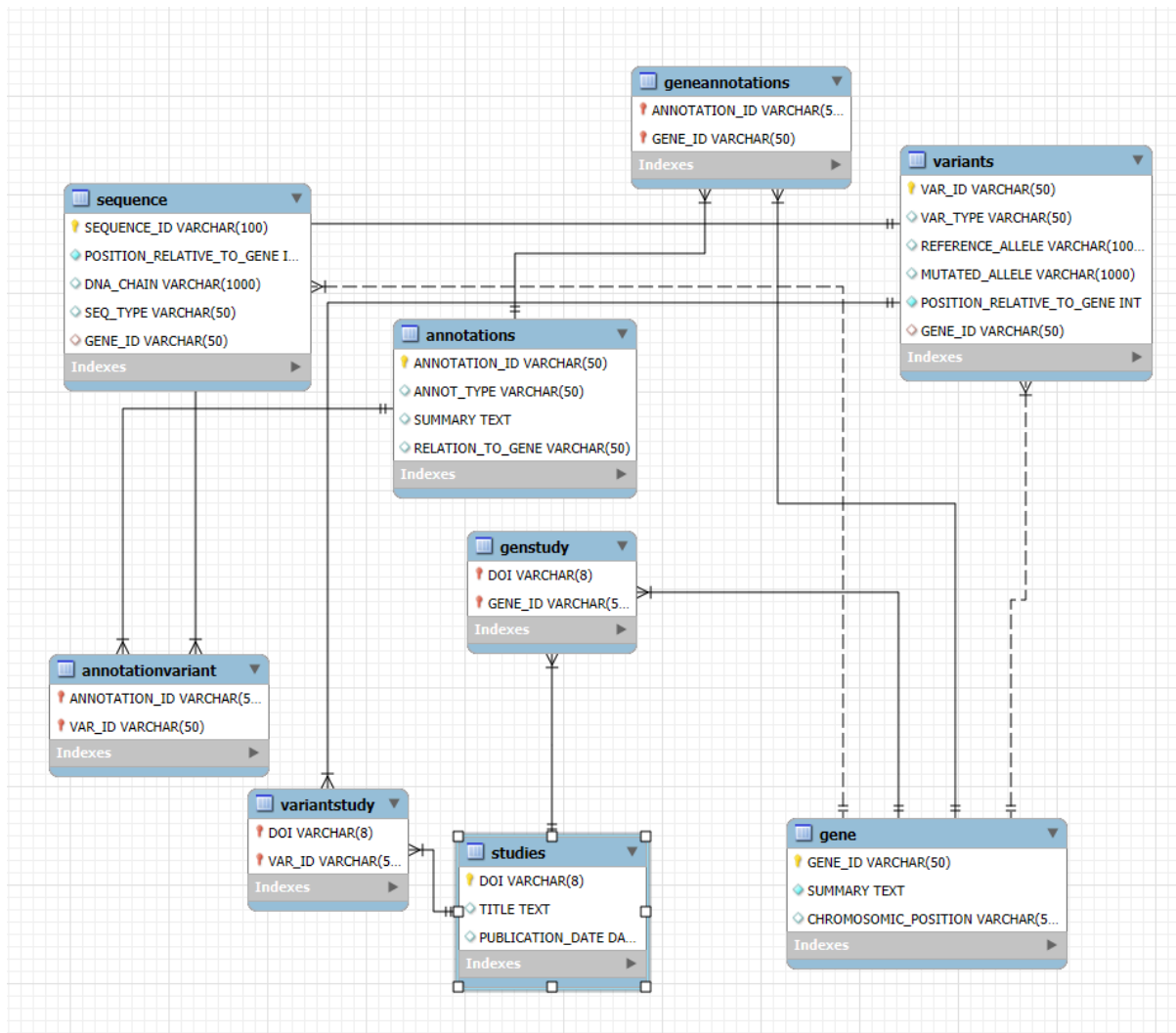
Asocia variantes genéticas con estudios que las han analizado.

- **Claves foráneas:**
  - VAR\_ID → variants
  - DOI → studies
- **Cardinalidad:** 1:N

## 9. Relación: genestudy

Asocia genes con estudios científicos.

- **Claves foráneas:**
  - GENE\_ID → gene
  - DOI → studies
- **Cardinalidad:** N:M



## 4. Script

**Nota:** Este script también se encuentra en el siguiente repositorio de GitHub:

[https://github.com/pedrofortesgonzalez/BBDD\\_equipo2\\_lote7.git](https://github.com/pedrofortesgonzalez/BBDD_equipo2_lote7.git)

-- 0. Crear la base de datos

```
CREATE DATABASE IF NOT EXISTS tallergrupal_equipo2_lote7; -- Crea la base de datos si no existe
USE tallergrupal_equipo2_lote7; -- Selecciona la base de datos
```

-- 1. Tabla GENE

```
CREATE TABLE IF NOT EXISTS GENE ( -- Crea la tabla GENE
```

```
    GENE_ID VARCHAR(50) PRIMARY KEY NOT NULL, -- ID del gen, clave primaria y no nula
```

```
    SUMMARY TEXT NOT NULL, -- Resumen, no nulo
```

```
    CHROMOSOMIC_POSITION VARCHAR(50) -- Posición cromosómica del gen
```

```
) ENGINE=InnoDB; -- Usamos el motor InnoDB
```

-- 2. Tabla SEQUENCE

## Creación de una base de datos genómica

Pedro Fortes González, Jon Corral López, Xose Manuel Tome Castro y Cristina Lázaro Ruiz

```
CREATE TABLE IF NOT EXISTS SEQUENCE ( -- Crea la tabla SEQUENCE
    SEQUENCE_ID VARCHAR(100) PRIMARY KEY NOT NULL, -- ID de la secuencia, clave primaria y
no nula
    POSITION_RELATIVE_TO_GENE INT NOT NULL, -- Posición relativa al gen, no nula
    DNA_CHAIN VARCHAR(1000) CHECK (CHAR_LENGTH(DNA_CHAIN) >= 10), -- Cadena de ADN,
longitud mínima de 10 caracteres
    SEQ_TYPE VARCHAR(50) -- Tipo de secuencia (cDNA, etc.)
) ENGINE=InnoDB;
```

### -- 3. Tabla VARIANTS

```
CREATE TABLE IF NOT EXISTS VARIANTS ( -- Crea la tabla VARIANTS
    VAR_ID VARCHAR(50) PRIMARY KEY, -- ID de la variante, clave primaria
    VAR_TYPE VARCHAR(50), -- Tipo de variante (SNP, indel, etc.)
    REFERENCE_ALLELE VARCHAR(1000) DEFAULT "«-»", -- Alelo de referencia, valor por defecto
    MUTATED_ALLELE VARCHAR(1000) DEFAULT "«-»", -- Alelo mutado, valor por defecto
    POSITION_RELATIVE_TO_GENE INT NOT NULL CHECK (POSITION_RELATIVE_TO_GENE > 0) --
Posición de la variante, debe ser positiva
) ENGINE=InnoDB;
```

### -- 4. Tabla ANNOTATIONS

```
CREATE TABLE IF NOT EXISTS ANNOTATIONS ( -- Crea la tabla ANNOTATIONS
    ANNOTATION_ID VARCHAR(50) PRIMARY KEY, -- ID de la anotación, clave primaria
    ANNOT_TYPE VARCHAR(50), -- Tipo de anotación (funcional, reguladora, etc.)
    SUMMARY TEXT, -- Resumen de la anotación
    RELATION_TO_GENE VARCHAR(50) -- Relación de la anotación con el gen
) ENGINE=InnoDB;
```

### -- 5. Tabla STUDIES

```
CREATE TABLE IF NOT EXISTS STUDIES ( -- Crea la tabla STUDIES
    DOI VARCHAR(8) PRIMARY KEY CHECK(DOI REGEXP '^[A-Za-z]{4}/[0-9]{3}$'), -- DOI del
estudio, formato específico
    TITLE TEXT, -- Título del estudio
    PUBLICATION_DATE DATE -- Fecha de publicación del estudio
) ENGINE=InnoDB;
```

### -- 6. Tablas intermedias y claves foráneas

-- Para "sequence" y "variants" introduzco clave foránea

```
ALTER TABLE SEQUENCE
ADD COLUMN GENE_ID VARCHAR(50), -- Añade columna para relacionar con GENE
ADD CONSTRAINT fk_sequence_gene FOREIGN KEY (GENE_ID) REFERENCES GENE(GENE_ID); -- Clave
foránea que hace referencia a GENE
```

```
ALTER TABLE VARIANTS
```

```
ADD COLUMN GENE_ID VARCHAR(50), -- Añade columna para relacionar con GENE
```

## Creación de una base de datos genómica

Pedro Fortes González, Jon Corral López, Xose Manuel Tome Castro y Cristina Lázaro Ruiz

```
ADD CONSTRAINT fk_variant_gene FOREIGN KEY (GENE_ID) REFERENCES GENE(GENE_ID); -- Clave foránea que hace referencia a GENE
```

```
-- Tablas intermedias
```

```
CREATE TABLE IF NOT EXISTS geneannotations ( -- Relaciona ANNOTATIONS con GENE
    ANNOTATION_ID VARCHAR(50),
    GENE_ID VARCHAR(50),
    PRIMARY KEY (ANNOTATION_ID, GENE_ID), -- Clave primaria compuesta
    FOREIGN KEY (ANNOTATION_ID) REFERENCES ANNOTATIONS(ANNOTATION_ID), -- Relación con ANNOTATIONS
    FOREIGN KEY (GENE_ID) REFERENCES GENE(GENE_ID) -- Relación con GENE
);
```

```
CREATE TABLE IF NOT EXISTS annotationvariant ( -- Relaciona ANNOTATIONS con VARIANTS
    ANNOTATION_ID VARCHAR(50),
    VAR_ID VARCHAR(50),
    PRIMARY KEY (ANNOTATION_ID, VAR_ID), -- Clave primaria compuesta
    FOREIGN KEY (ANNOTATION_ID) REFERENCES ANNOTATIONS(ANNOTATION_ID), -- Relación con ANNOTATIONS
    FOREIGN KEY (VAR_ID) REFERENCES VARIANTS(VAR_ID) -- Relación con VARIANTS
);
```

```
CREATE TABLE IF NOT EXISTS genstudy ( -- Relaciona GENE con STUDIES
    DOI VARCHAR(8),
    GENE_ID VARCHAR(50),
    PRIMARY KEY (DOI, GENE_ID), -- Clave primaria compuesta
    FOREIGN KEY (DOI) REFERENCES STUDIES(DOI), -- Relación con STUDIES
    FOREIGN KEY (GENE_ID) REFERENCES GENE(GENE_ID) -- Relación con GENE
);
```

```
CREATE TABLE IF NOT EXISTS variantstudy ( -- Relaciona VARIANTS con STUDIES
    DOI VARCHAR(8),
    VAR_ID VARCHAR(50),
    PRIMARY KEY (DOI, VAR_ID), -- Clave primaria compuesta
    FOREIGN KEY (DOI) REFERENCES STUDIES(DOI), -- Relación con STUDIES
    FOREIGN KEY (VAR_ID) REFERENCES VARIANTS(VAR_ID) -- Relación con VARIANTS
);
```

```
-- 7. Introducción de datos ficticios
```

```
-- - Tablas principales
```

```
INSERT INTO GENE (GENE_ID, SUMMARY, CHROMOSOMIC_POSITION)
VALUES ('GENE001', 'Este gen está involucrado en la regulación del ciclo celular.', '1q23.3'); -- Inserta un gen
```



## Creación de una base de datos genómica

Pedro Fortes González, Jon Corral López, Xose Manuel Tome Castro y Cristina Lázaro Ruiz

```
INSERT INTO SEQUENCE (SEQUENCE_ID, POSITION_RELATIVE_TO_GENE, DNA_CHAIN, SEQ_TYPE, GENE_ID)
```

```
VALUES ('SEQ001', 123, 'ATGCGTACGTGACTGAGTCGATCGA', 'cDNA', 'GENE001'); -- Inserta una secuencia para el gen
```

```
INSERT INTO VARIANTS (VAR_ID, VAR_TYPE, REFERENCE_ALLELE, MUTATED_ALLELE, POSITION_RELATIVE_TO_GENE, GENE_ID)
```

```
VALUES ('VAR001', 'SNP', 'A', 'G', 150, 'GENE001'); -- Inserta una variante para el gen
```

```
INSERT INTO ANNOTATIONS (ANNOTATION_ID, ANNOT_TYPE, SUMMARY, RELATION_TO_GENE)
```

```
VALUES ('ANNOT001', 'Functional', 'Esta variante afecta la función de la proteína', 'GENE001'); -- Inserta una anotación
```

```
INSERT INTO STUDIES (DOI, TITLE, PUBLICATION_DATE)
```

```
VALUES ('GENE/123', 'Estudio sobre la función de GENE001', '2023-06-15'); -- Inserta un estudio
```

-- - Tablas intermedias

-- Relaciona variantes con anotaciones

```
INSERT INTO annotationvariant (ANNOTATION_ID, VAR_ID)
```

```
VALUES ('ANNOT001', 'VAR001');
```

-- Relaciona genes con anotaciones

```
INSERT INTO geneannotations (ANNOTATION_ID, GENE_ID)
```

```
VALUES ('ANNOT001', 'GENE001');
```

-- Relaciona genes con estudios

```
INSERT INTO genstudy (DOI, GENE_ID)
```

```
VALUES ('GENE/123', 'GENE001');
```

-- Relaciona variantes con estudios

```
INSERT INTO variantstudy (DOI, VAR_ID)
```

```
VALUES ('GENE/123', 'VAR001');
```