Entrega 2 - Grupo 1

Grupo 1:

- Ana Sofia Villa Benavides 201923361
- Ronald Yesid Diaz Pardo 202111309
- Gabriela Moreno Prieto 201713956

Correcciones Primera Entrega:

Enfermedad: Enfermedad: Síndrome de Bardet-Biedl [SBB]

Pregunta: ¿Cómo interactúan los diferentes genes asociados al Síndrome de Bardet-Biedl [SBB] en las vías de señalización celular y cómo estas interacciones contribuyen a la variabilidad fenotípica observada en los pacientes?

¿Podría la variabilidad fenotípica observada en pacientes con el Síndrome de Bardet-Biedl (SBB) ser explicada por las interacciones entre los diferentes genes asociados, especialmente en las vías de señalización celular?

**(se modificó la pregunta para que fuera más clara, pero sigue preguntando lo mismo)

Justificación:

El síndrome de Bardet-Biedl [SBB] es una rara enfermedad hetero genética asociado a la hiperfagia (obesidad central), retrasos mentales, disfunción renal, hipogonadismo y distrofia de conos rojos, sin embargo, se ha encontrado nuevas condiciones clínicas que afectan a los pacientes como como disfunciones neurológicas, de comunicación y de lenguaje, dismorfia facial y anomalías dentales (Beales et al., 1999)

Adicionalmente, esta enfermedad causa una ciliopatia expresada en múltiples sistemas del cuerpo como el muscular, visual y funcionamiento y los principales genes causantes de este síndrome (BBS del 1 al 10) también tienen incidencia en la aparición de otros síndromes como el de joubert que también es multisistémico por lo que su variabilidad multisistémica y su incidencia en otros genes podría ayudar hallar mejores tratamientos para diferentes personas afectadas por estos síndromes, así mismo, al ser una enfermedad con progresión lenta sus síntomas suelen verse a una edad alrededor de los 9 años en promedio y la enfermedad ya habrá avanzado mucho para este punto (Beales et al., 1999), por lo que conocer las maneras en que diversos patrones contribuyen a esta variabilidad fenotípica podría ayudar a mejorar diagnósticos.

En particular, la pregunta de investigación puede abordarse utilizando las diferentes herramientas del EBI. Inicialmente, es necesario realizar una revisión de la literatura sobre el Síndrome de Bardet-Biedl en general y sobre los genes específicos asociados con esta enfermedad; para ello, se utiliza la herramienta *Europe PMC*. En esta etapa, también es importante revisar los antecedentes para identificar los aspectos previamente explorados y cuáles podrían ser causas de la variabilidad fenotípica observada según la literatura.

Además, en esta etapa inicial, *Ensembl* es útil para buscar información sobre los fenotipos asociados con este síndrome y explorar detalles sobre los genes relacionados, como su ubicación genómica, estructura y las variantes genéticas conocidas. Más a fondo, *Ensembl* también resulta útil a través de su sección de *Pathways* y el *Variant Effect Predictor* (VEP), ya que estas herramientas permiten examinar la interacción de los genes en vías específicas asociadas con la enfermedad y predecir los posibles efectos de las variantes identificadas.

Por otro lado, *ENA* complementa la búsqueda realizada previamente de los genes, permitiendo descargar o acceder a las secuencias completas de los genes, lo cual es útil para análisis posteriores con otras herramientas, como *AlphaFold*. Además, *ENA* permite explorar las secuencias disponibles para las diferentes variantes de cada gen, lo que es útil para identificar variaciones en exones u otras regiones que podrían influir en la función del gen.

Herramientas Segunda Entrega:

FAIR

Dada la importancia del buen manejo de datos, en nuestro proyecto estableceremos parámetros que guiarán todo el proceso, alineándonos con los principios FAIR. Para empezar, mantendremos los datos y archivos organizados mediante una estructura o convención de nombres consistente que todos utilizaremos (por ejemplo, 20242_ET1_DescripcionContenido.ext). En caso de realizar cambios o establecer convenciones adicionales, estos serán detallados en el archivo README final del proyecto.

Además, utilizaremos un repositorio público en GitHub donde, al concluir el proyecto, se publicarán nuestros datos, junto con los archivos o scripts utilizados para aplicar las diferentes herramientas. También llevaremos un registro detallado de cada paso de procesamiento y preprocesamiento de nuestros datos, especificando la herramienta, los parámetros y la versión empleada.

Al utilizar datos provenientes de bases de datos o plataformas del EBI, que se rigen por los principios FAIR, garantizaremos que los datos mantengan su integridad ética y sean accesibles. No obstante, también realizaremos una verificación para revisar los artículos de origen de los datos. De este modo, aseguraremos que nuestro proyecto y la información asociada sean fáciles de buscar, accesibles, interoperables y reutilizables.

ENSEMBL

Para la determinación de los genes humanos involucrados en el síndrome BBS se ha realizado la búsqueda del fenotipo como "Bardet-Biedl syndrome" en el buscador de Ensembl para human donde se identificó el fenotipo/enfermedad asociada con esto, la cual tiene un identificador MONDO:0015229.

Posteriormente con el uso de la herramienta BioMart de Ensembl se realizó una búsqueda más especializada de este fenotipo. Como filtro principal se usó el fenotipo y dentro del fenotipo se aplicaron aquellos que tuvieran el nombre de la enfermedad.

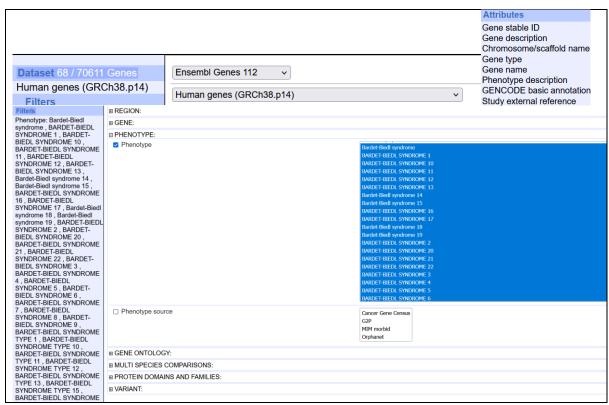


Figura 1. Filtros BioMart. Filtros usados para la búsqueda en BioMart.

De esta búsqueda se obtuvieron 1267 resultados, así que se exportaron los datos obtenidos a un documento CSV. Posteriormente, este archivo se pasó a un documento Excel para una mejor visualización de los resultados disponible aquí.

A partir de estos resultados se puede decir que la enfermedad está asociada a 28 genes presente en 16 cromosomas.

Tabla 1. Resultados BioMart. Resumen de genes obtenidos en la búsqueda realizada en BioMart y cromosoma en el que está ubicado cada gen.

Nombre de Gen	Cromosoma
ARL6	3
BBIP1	10
BBS1	11
BBS10	12
BBS12	4
BBS2	16
BBS4	15
BBS5	2
BBS7	4
BBS9	7
CCDC28B	1
CEP19	3
CEP290	12
CFAP418	8
IFT172	2
IFT27	22

IFT74	9
LZTFL1	3
MKKS	20
MKS1	17
NPHP1	2
SCAPER	15
SCLT1	4
SDCCAG8	1
TMEM67	8
TRIM32	9
TTC8	14
WDPCP	2

PROGGRAMATIC ACCESS

Para nuestra investigación consideramos muy beneficioso el uso de acceso programático para acceder a información actualizada a medida que realizamos nuestro proyecto, esto debido a que la enfermedad BBS es una enfermedad multi genética donde la cantidad de genes que tienen un rol importante en la enfermedad es alta, por lo que realizar estas tareas de búsqueda manuales no es ventajoso y no nos permiten aprovechar el potencial de acceso a APIs con resultados obtenidos de diferentes herramientas a partir de otras de EBI.

Un ejemplo de ciclo de uso de herramientas de acceso programático seria en el caso de usar accesos a API y obtener datos que luego podemos abrir usando pandas en Python, como los diferentes genes asociados a la enfermedad por medio de Ensembl, y luego usar estos genes para buscar sus secuencias completas o variaciones en ENA y también las secuencias de proteínas que codifican y variaciones/mutaciones de estas codificaciones de la proteína, para posteriormente hacer un llamado de API en Expression atlas, donde evaluamos como estas variaciones en la codificación de las proteínas que consideremos más relevantes para nuestro estudio afectan la función del gen respectivo y por lo tanto su expresión. Así mismo, se resalta la importancia de estar actualizados respecto a estudios más recientes de la enfermedad y genes involucrados y cuáles de estos causan los mayores impactos negativos en el fenotipo y con qué frecuencia dentro de las personas que tengan estas proteínas, por lo que el uso de API de European PMC nos serviría mucho para esto y por medio de una línea de código de Python en Jupyter podemos descargar los artículos o estudios más relevantes para nosotros de acuerdo a ciertos filtros.

Cabe resaltar que se ha empezado a hacer uso de algunas de estas herramientas. En el caso de Ensembl se ha hecho uso de la REST API de ensembl donde se hace una búsqueda de un fenotipo y con los datos provistos allí se puede acceder a los genes asociados a esta enfermedad.

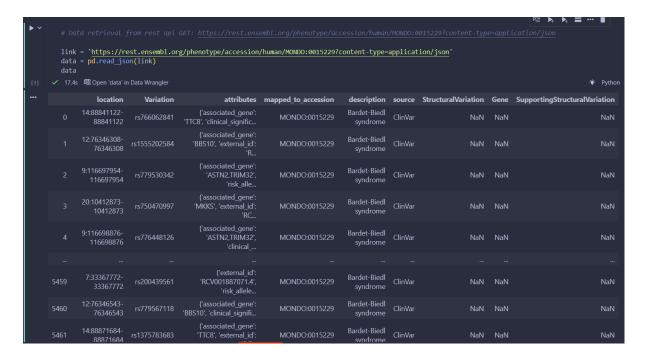
Ejemplo de llamado de API con fenotipo de Bardet-Biedl syndrome (MONDO:0015229):

https://rest.ensembl.org/phenotype/accession/human/MONDO:0015229?content-type=application/json

Donde se obtuvo el siguiente resultado:

```
"description": "Bardet-Biedl syndrome",
  "attributes": {
    "clinical_significance": "likely benign",
    "risk_allele": "C",
"associated_gene": "WDPCP",
"external_id": "RCV002080877.4"
  "mapped_to_accession": "MONDO:0015229",
  "Variation": "rs780109239",
  "source": "ClinVar",
  "location": "2:63433726-63433726"
},
  "source": "dbVar",
  "location": "11:66521250-66526212",
  "StructuralVariation": "nsv5672558"
  "description": "Bardet-Biedl syndrome",
"mapped_to_accession": "MONDO:0015229"
},
{
  "Variation": "rs140537219",
  "description": "Bardet-Biedl syndrome",
  "attributes": {
    "clinical_significance": "uncertain significance",
    "risk_allele": "T",
"associated_gene": "BBS12",
    "external_id": "RCV001941506.4"
  "mapped_to_accession": "MONDO:0015229",
  "location": "4:122742611-122742611",
"source": "ClinVar"
},
  "attributes": {
    "associated_gene": "WDPCP",
    "risk_allele": "C",
    "clinical_significance": "uncertain significance",
    "external_id": "RCV001245531.6"
  "mapped_to_accession": "MONDO:0015229",
  "description": "Bardet-Biedl syndrome",
  "Variation": "rs759281211",
  "source": "ClinVar",
  "location": "2:63174803-63174803"
```

Y se puede realizar en Python por medio del siguiente código donde se usa la librería de pandas como "pd":



Finalmente, estos códigos de acceso programático estarán disponibles en nuestro repositorio de GitHub, el cual se puede visualizar por medio de <u>este link</u>.

ENA

Como se indicó inicialmente, ENA complementa nuestra búsqueda previa sobre los fenotipos asociados a esta enfermedad y nos permitirá conocer secuencias completas de los genes y sus variantes en otras regiones u exones y así ver como esto afectaría la función del gen en cuestión.

Para usar ENA se decide que la búsqueda se hará para las siguientes categorías:

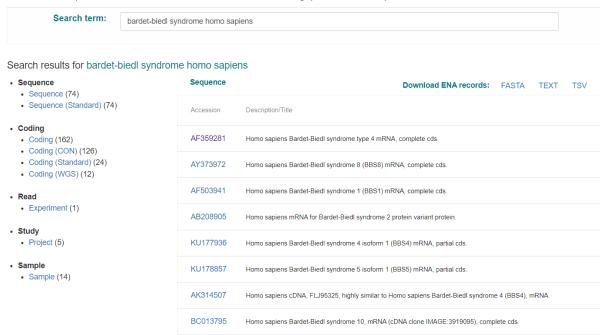
- 1. Coding
- 2. Sequence

Inicialmente se realizó una búsqueda de free-text con ebi-search (funcionalidad que provee ENA por defecto en la página de inicio) para la enfermedad de "bardet-biedl syndrome" y para la cual definimos que esta sea para la especie homo sapiens, de manera que la búsqueda en free-text queda como: "bardet-biedl syndrome homo sapiens".

En lo que respecta a sequence se identifican 74 secuencias, 162 de coding por defecto.

Text Search

Uses EBI Search to perform a free text search across ENA data. For more detailed usage please refer to the help & documentation section.



Cabe resaltar que por facilidad de movimiento de archivos a otras herramientas de EBI como Expression atlas se usaran formatos de archivos FASTA, debido a que estos tienen más soporte en las diferentes herramientas y la facilidad de lectura que este tiene al no incluir otra información como los autores y descripciones de estudios hace que también sea una buena opción para un posible análisis de datos con Pandas Python o, si se requiere, se puede hacer uso del formato txt en pandas debido a que contiene información extra que fasta omite.

	DOWINOOU ENATECOIUS: FASTA TEXT TSV
Accession	Description/Title
CAG5073462	Cotesia congregata Similar to BBS7: Bardet-Biedl syndrome 7 protein (Homo sapiens)
KAI4072508	Homo sapiens (human) Bardet-Biedl syndrome 1
KAI4013421	Homo sapiens (human) Bardet-Biedl syndrome 9
KAI2575011	Homo sapiens (human) Bardet-Biedl syndrome 4
KAI4013424	Homo sapiens (human) Bardet-Biedl syndrome 9
KAI4036821	Homo sapiens (human) Bardet-Biedl syndrome 5
KAI2575010	Homo sapiens (human) Bardet-Biedl syndrome 4
KAI2561191	Homo sapiens (human) Bardet-Biedl syndrome 1
KAI2545488	Homo sapiens (human) Bardet-Biedl syndrome 9
KAI2545489	Homo sapiens (human) Bardet-Biedl syndrome 9
KAI2545483	Homo sapiens (human) Bardet-Biedl syndrome 9
KAI4058613	Homo sapiens (human) Bardet-Biedl syndrome 4
KAI2578586	Homo sapiens (human) Bardet-Biedl syndrome 2

Download ENA records: FASTA TEXT TSV

Coding

En lo que respecta a datos de coding identificamos muchas descripciones similares para los mismos genes. La diversidad de secuencias en los resultados de búsqueda es beneficiosa para la investigación, ya que permite explorar una amplia gama de variantes genéticas de los genes asociados con el síndrome de Bardet-Biedl. Analizando diferentes secuencias, se pueden identificar variantes alélicas, transcritos alternativos y posibles mutaciones, proporcionando una visión más completa de cómo estas diferencias afectan la función del gen y contribuyen a la enfermedad.