TRABAJO FINAL

Curso: Lenguaje de Programación II.

Semestre: 2021-2

Tema: webscrapeo a UniProt (https://www.uniprot.org/).

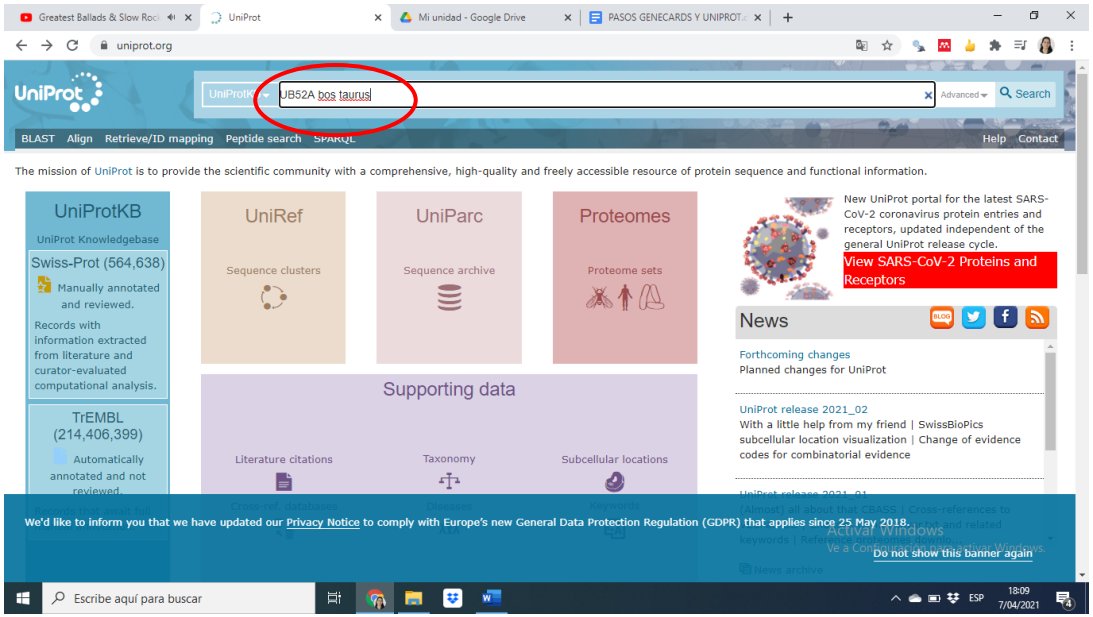
INDICACIONES:

- El trabajo es grupal (integrantes de grupos ya indicados)
- Utilizar *Jupyter notebook*.
- Trabajar en equipo vía *GitHub* entre los miembros de los grupos. (Sugerencia: el jefe de grupo puede crear el repositorio del proyecto e invitar a los demás integrantes a colaborar en el proyecto).

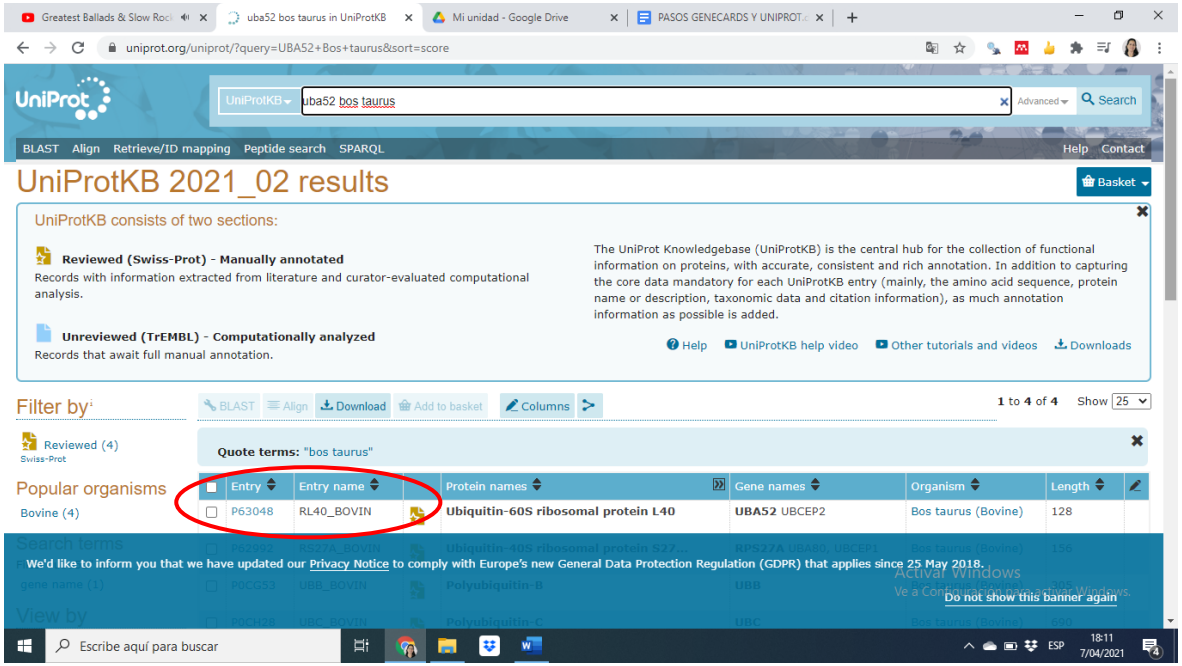
PROCEDIMIENTO REFERENCIAL:

1. Descargar y explorar el archivo ***genes identificados con vecinos.xlsx*** compuesta por varias hojas las cuales tienen una columna con etiqueta: GenAbrev.

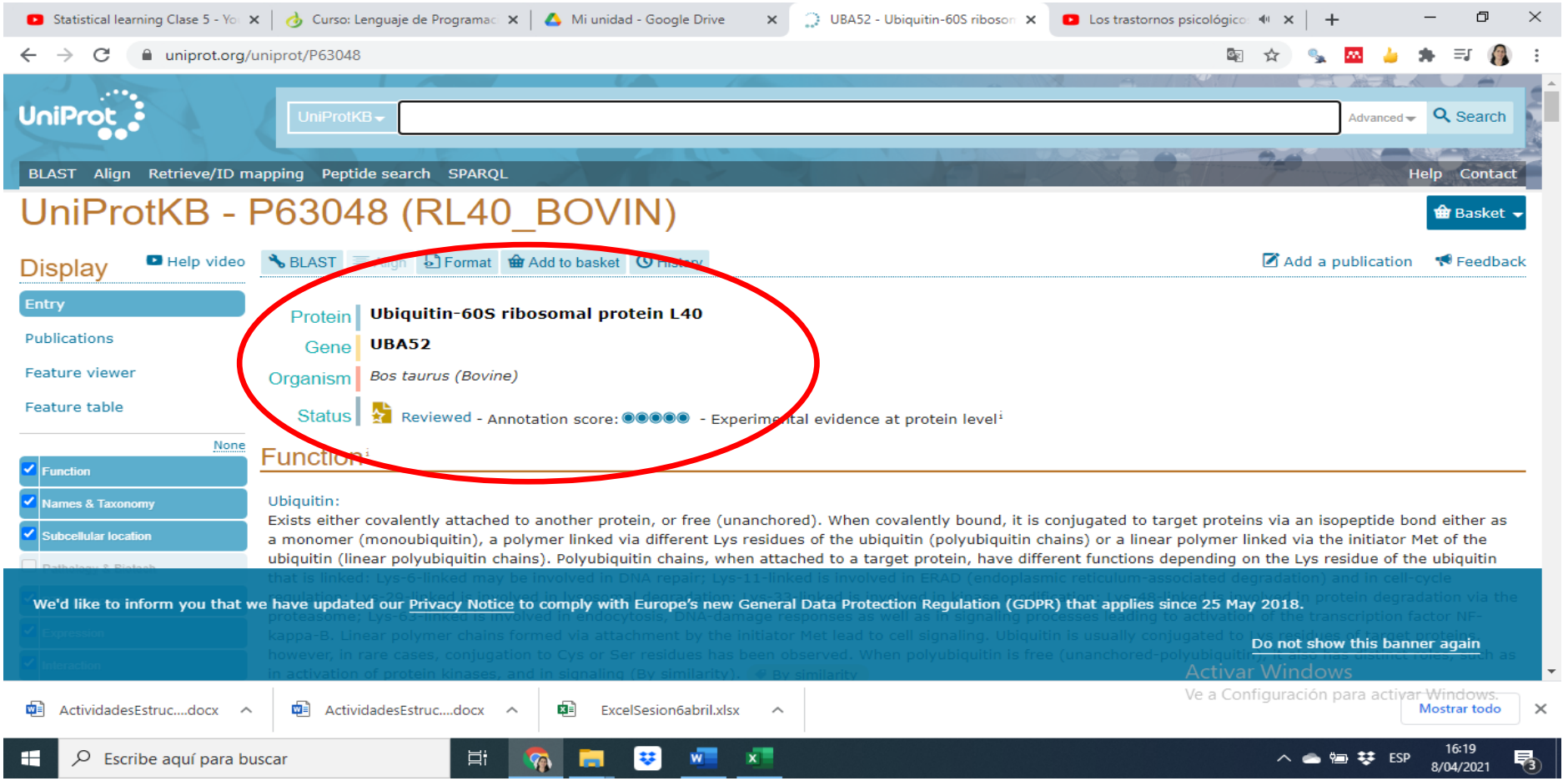
2. Para cada nombre del gen en abreviado (GenAbrev), abrir la página https://www.uniprot.org y en el buscador solicitar la búsqueda de cada una de la forma siguiente: colocar el nombre del gen y la palabra *Bos Taurus.* Abajo se muestra para **UBA52 Bos taurus**



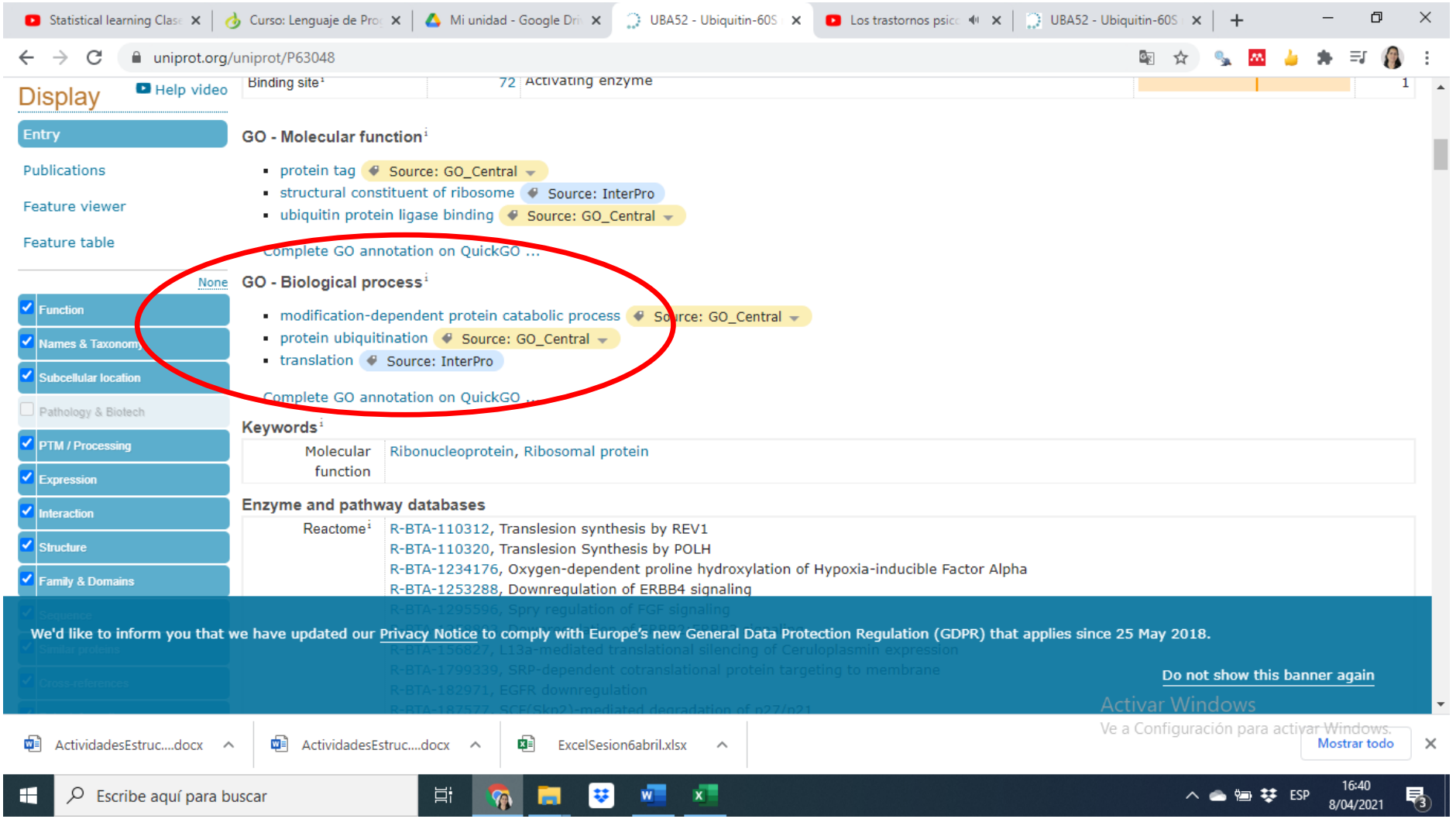
3. Se mostrará una lista de genes con el nombre buscado. Extraer información de ***entry*** y ***entry name***.



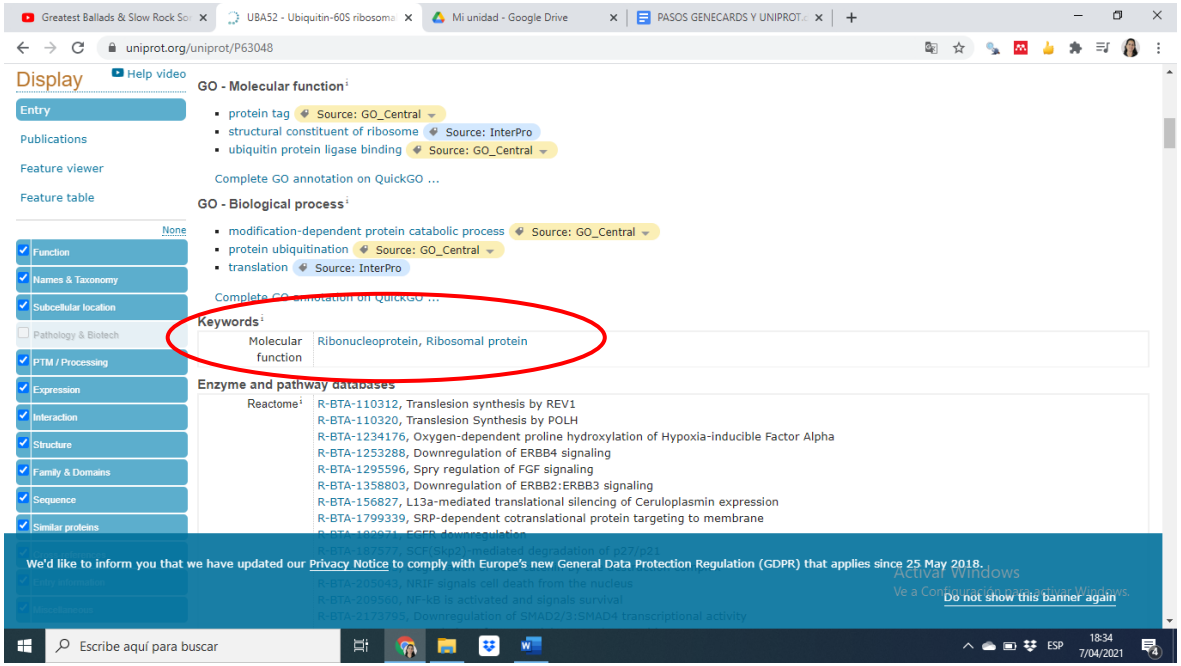
4. Seleccionar la primera opción del gen de la primera columna “entry” (clickear), la cual llevará a la siguiente pantalla y de la cual extraer la información de **Protein**, **Gene**, **Organism** y **Status**.



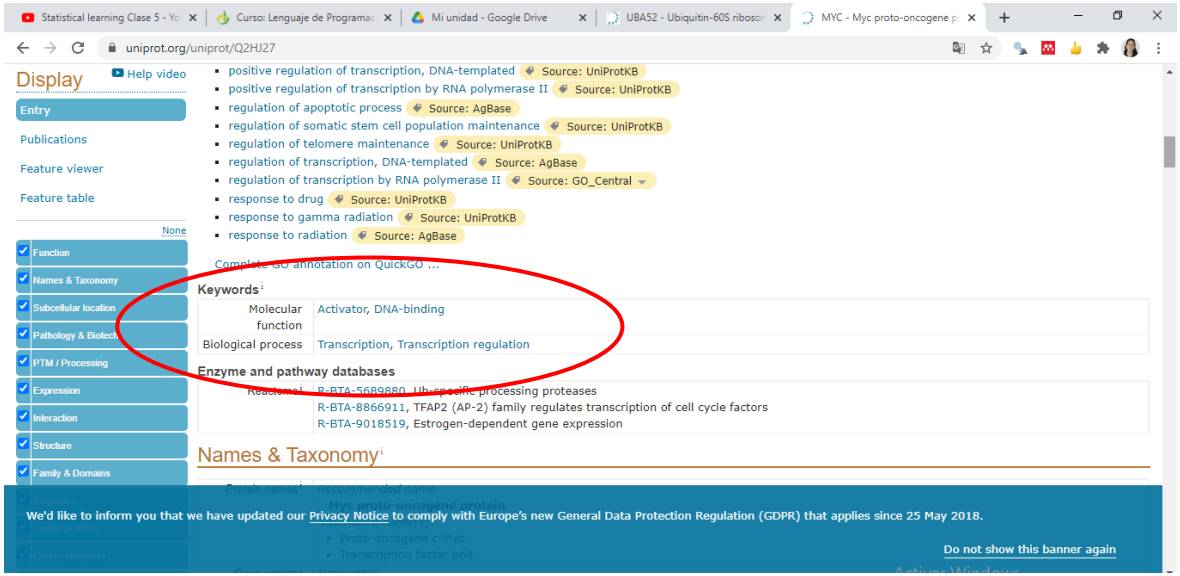
5. Extarer la infomación que corresponde a: **GO - Biological process**


Statistical learning Clas
Curso: Lenguaje de Pro
Mi unidad - Google Dri
UBA52 - Ubiquitin-60S
Los trastornos psic
UBA52 - Ubiquitin-60S
uniprot.org/uniprot/P63048
Display
Help video
Entry
Publications
Feature viewer
Feature table
None
Function
Names & Taxonomy
Subcellular location
Pathology & Biotech
PTM / Processing
Expression
Interaction
Structure
Family & Domains
Binding site
72
Activating enzyme
1
GO - Molecular function
protein tag
Source: GO_Central
structural constituent of ribosome
Source: InterPro
ubiquitin protein ligase binding
Source: GO_Central
Complete GO annotation on QuickGO ...
GO - Biological process
modification-dependent protein catabolic process
Source: GO_Central
protein ubiquitination
Source: GO_Central
translation
Source: InterPro
Complete GO annotation on QuickGO ...
Keywords
Molecular function
Ribonucleoprotein, Ribosomal protein
Enzyme and pathway databases
Reactome
R-BTA-110312, Translesion synthesis by REV1
R-BTA-110320, Translesion Synthesis by POLH
R-BTA-1234176, Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha
R-BTA-1253288, Downregulation of ERBB4 signaling
R-BTA-1295596, Spry regulation of FGF signaling
R-BTA-156827, L13a-mediated translational silencing of Ceruloplasmin expression
R-BTA-1799339, SRP-dependent cotranslational protein targeting to membrane
R-BTA-182971, EGFR downregulation
We'd like to inform you that we have updated our Privacy Notice to comply with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018.
Do not show this banner again
Activar Windows
Ve a Configuración para activar Windows.
ActividadesEstruc....docx
ActividadesEstruc....docx
ExcelSesion6abril.xlsx
Mostrar todo
Escribe aquí para buscar
ESP
16:40
8/04/2021

6. Extraer también la información correspondiente a **Keywords**



La información correspondiente a **keywords** puede variar por ejemplo para: **cip2a bos Taurus**



7. La información debe estructurarse en un archivo **.csv** o **.xlsx** que muestre el código abreviado del gen (GenAbrev) y la información organizada de keywords (diferenciando cada tipo de función).

   Sugerencia de estructura de reporte: **Reporte.xlsx**