User Documentation

Samuel González Martín

March 2025

1 Software and Hardware Requirements to Run the Project

To run this software, you must first have R installed along with any IDE that allows handling R files. In this case, RStudio has been used, but any other IDE is viable (Visual Studio Code, TinnR, Eclipse, Rbase, etc.). The project has been developed on Windows, so the installation instructions are specifically for this operating system.

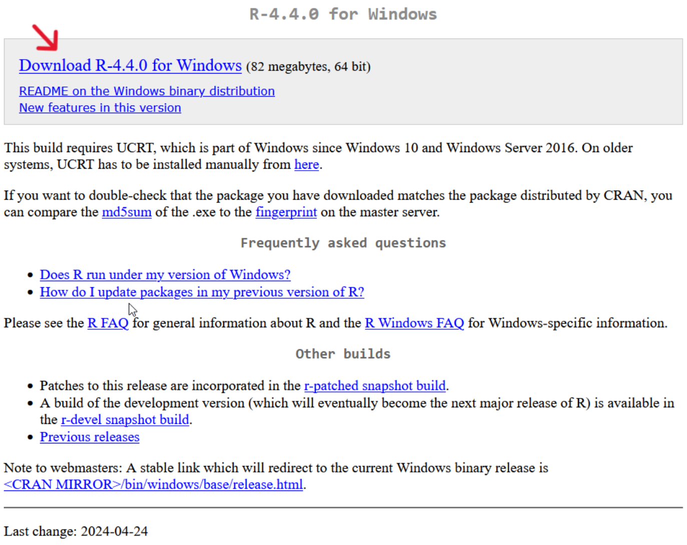
Separately, we must also install the MongoDB driver, MongoDB Compass, so the system can create patient records locally. How to install all these programs is detailed in the next section.

2 Installation / Setup

2.1 Installing R

To install R on Windows, you can download it from the CRAN website [CRAN, 2024]. Once on the website, click on the "*Download*" section, as shown in Figure 1. This will download an executable (.exe) file that will be found in the Downloads folder.

When executing it, you must select the installation language 2. Once selected, the license agreements will appear 3, and you will be given the option to choose the installation directory 4; in this case, the default folder will be used. Then, you will select the components to install and specify if you want to use configuration options, as shown in Figures 5 and 6. Finally, before starting the installation, you will choose a folder to store the program's shortcuts 7 and define additional tasks to be performed during installation 8.



R-4.4.0 for Windows

Download R-4.4.0 for Windows (82 megabytes, 64 bit)

README on the Windows binary distribution

New features in this version

This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from here.

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server.

Frequently asked questions

- Does R run under my version of Windows?
- How do I update packages in my previous version of R?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

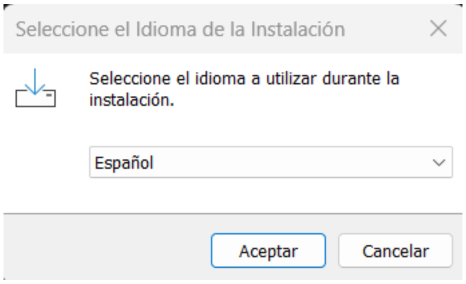
Other builds

- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.html.

Last change: 2024-04-24

Figure 1: Download R

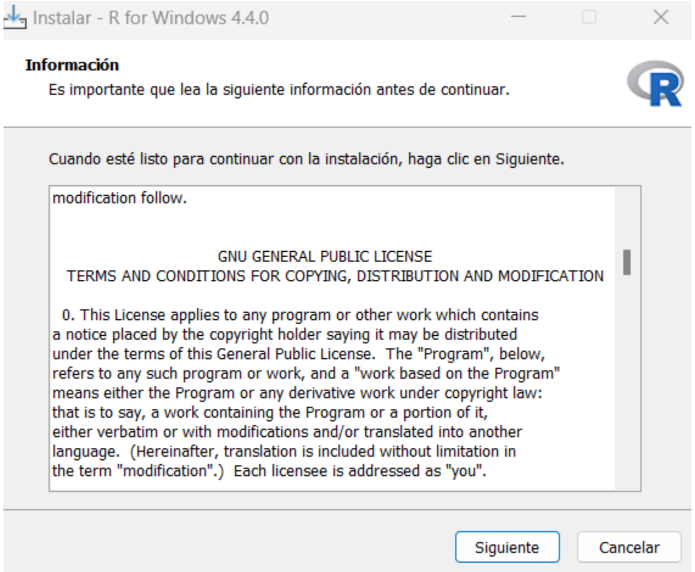


Seleccione el Idioma de la Instalación

Seleccione el idioma a utilizar durante la instalación.

Español

Figure 2: Language Selection



Información

Es importante que lea la siguiente información antes de continuar.

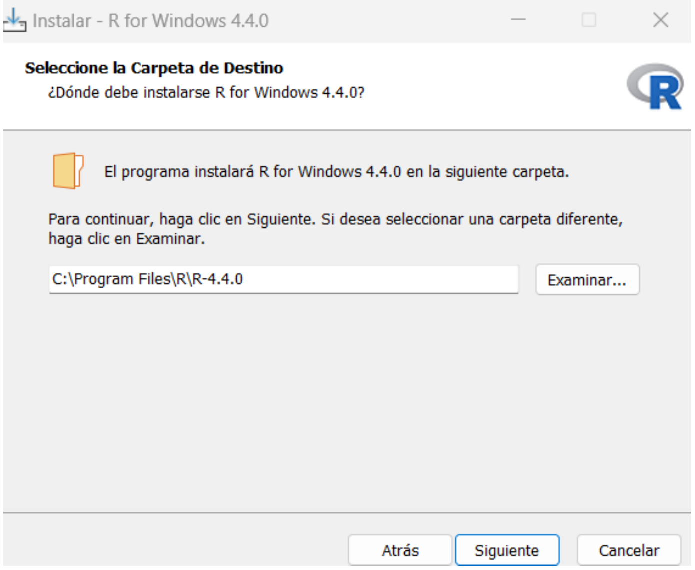
Cuando esté listo para continuar con la instalación, haga clic en Siguiente.

```
modification follow.


                    GNU GENERAL PUBLIC LICENSE
   TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

  0. This License applies to any program or other work which contains
a notice placed by the copyright holder saying it may be distributed
under the terms of this General Public License.  The "Program", below,
refers to any such program or work, and a "work based on the Program"
means either the Program or any derivative work under copyright law:
that is to say, a work containing the Program or a portion of it,
either verbatim or with modifications and/or translated into another
language.  (Hereinafter, translation is included without limitation in
the term "modification".)  Each licensee is addressed as "you".
```

Figure 3: Licenses



Seleccione la Carpeta de Destino

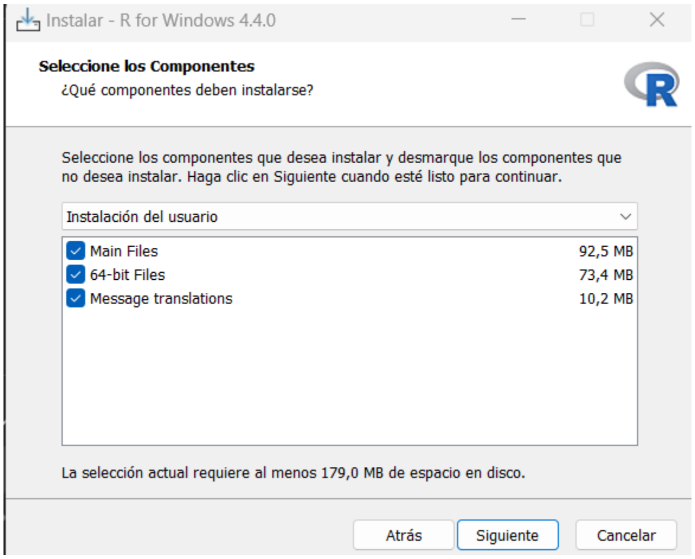
¿Dónde debe instalarse R for Windows 4.4.0?

El programa instalará R for Windows 4.4.0 en la siguiente carpeta.

Para continuar, haga clic en Siguiente. Si desea seleccionar una carpeta diferente, haga clic en Examinar.

C:\Program Files\R\R-4.4.0

Figure 4: Destination Folder Selection



Seleccione los Componentes

¿Qué componentes deben instalarse?

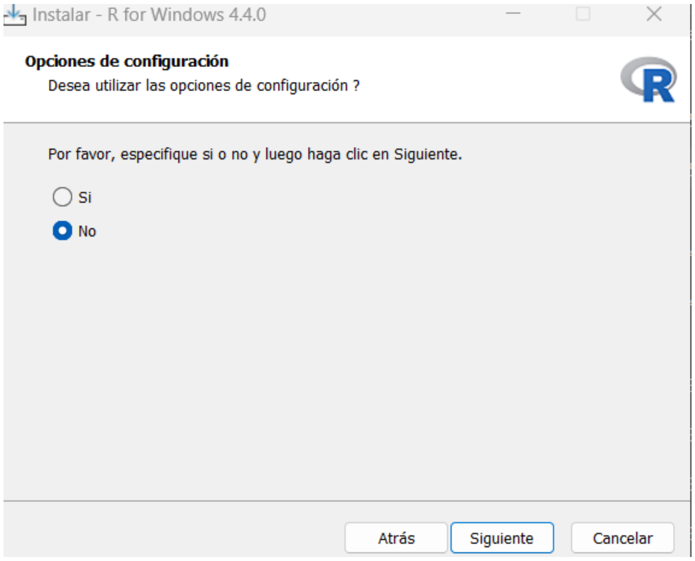
Seleccione los componentes que desea instalar y desmarque los componentes que no desea instalar. Haga clic en Siguiente cuando esté listo para continuar.

Instalación del usuario

☑ Main Files	92,5 MB
☑ 64-bit Files	73,4 MB
☑ Message translations	10,2 MB

La selección actual requiere al menos 179,0 MB de espacio en disco.

Figure 5: Component Selection



Opciones de configuración

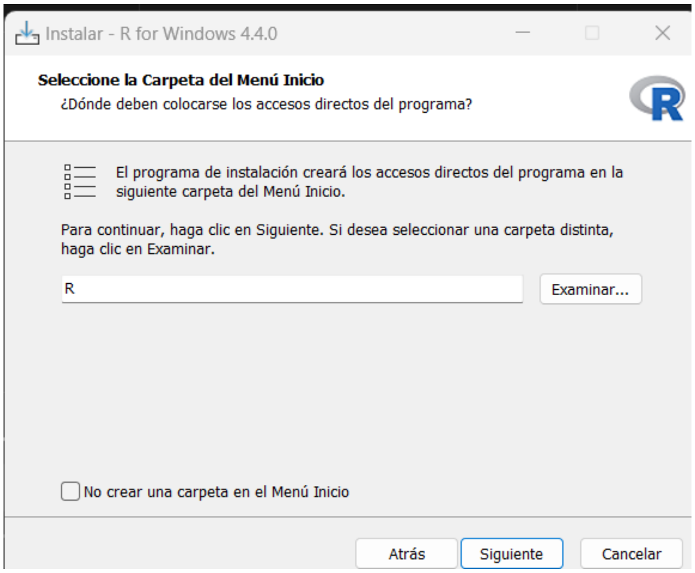
Desea utilizar las opciones de configuración ?

Por favor, especifique si o no y luego haga clic en Siguiente.

○ Si

● No

Figure 6: Configuration Options



Instalar - R for Windows 4.4.0

Seleccione la Carpeta del Menú Inicio

¿Dónde deben colocarse los accesos directos del programa?

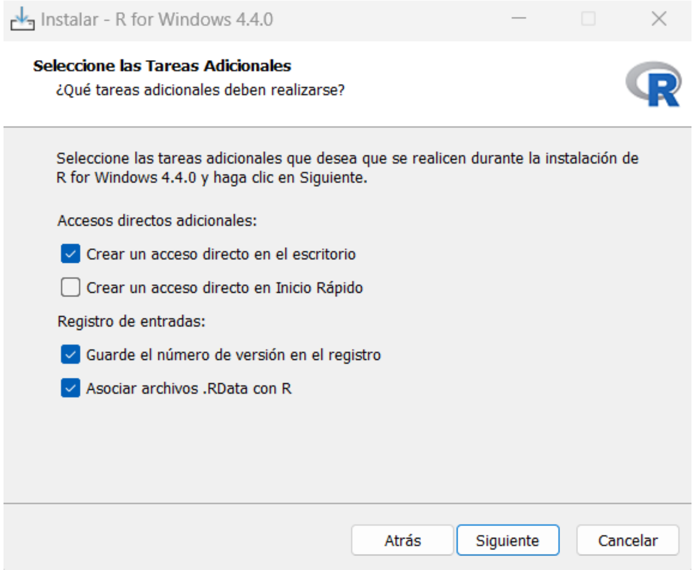
El programa de instalación creará los accesos directos del programa en la siguiente carpeta del Menú Inicio.

Para continuar, haga clic en Siguiente. Si desea seleccionar una carpeta distinta, haga clic en Examinar.

R

☐ No crear una carpeta en el Menú Inicio

Figure 7: Shortcut Folder Selection



Instalar - R for Windows 4.4.0

Seleccione las Tareas Adicionales

¿Qué tareas adicionales deben realizarse?

Seleccione las tareas adicionales que desea que se realicen durante la instalación de R for Windows 4.4.0 y haga clic en Siguiente.

Accesos directos adicionales:

☑ Crear un acceso directo en el escritorio

☐ Crear un acceso directo en Inicio Rápido

Registro de entradas:

☑ Guarde el número de versión en el registro

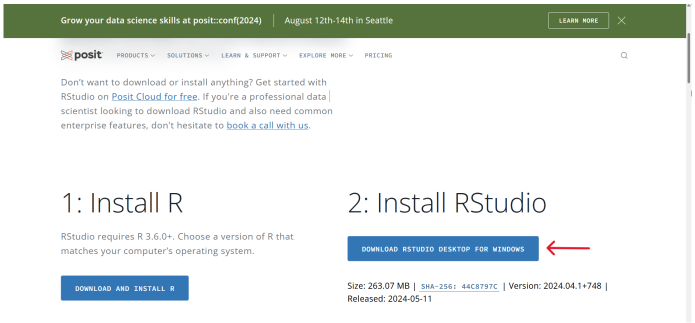
☑ Asociar archivos .RData con R

Figure 8: Additional Tasks

2.2 Installing RStudio

The installation of RStudio is similar to that of R. To download the installation file 9, go to the Posit website [Posit, 2024].

When executing the file, you will be asked to indicate the destination directory 4 and the folder for RStudio shortcuts 11. Once this information is completed, the installation will begin.


Grow your data science skills at posit::conf(2024)
August 12th-14th in Seattle
LEARN MORE
posit
PRODUCTS
SOLUTIONS
LEARN & SUPPORT
EXPLORE MORE
PRICING
DOWNLOAD AND INSTALL R
DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Don't want to download or install anything? Get started with RStudio on Posit Cloud for free. If you're a professional data scientist looking to download RStudio and also need common enterprise features, don't hesitate to book a call with us.

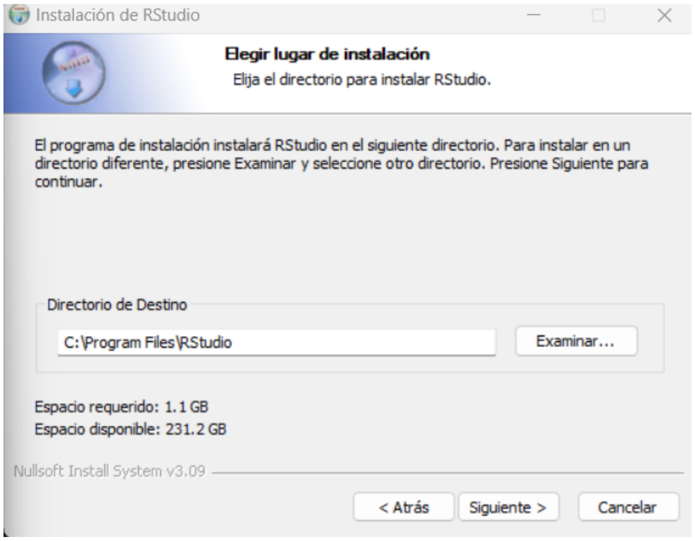
1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

2: Install RStudio

Size: 263.07 MB | SHA-256: 44C8797C | Version: 2024.04.1+748 | Released: 2024-05-11

Figure 9: Download Installation File



Instalación de RStudio

Elegir lugar de instalación

Elija el directorio para instalar RStudio.

El programa de instalación instalará RStudio en el siguiente directorio. Para instalar en un directorio diferente, presione Examinar y seleccione otro directorio. Presione Siguiente para continuar.

Directorio de Destino

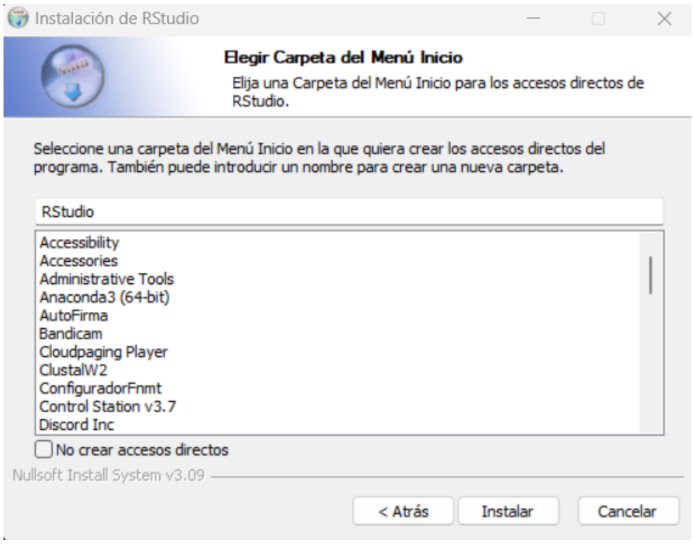
C:\Program Files\RStudio

Espacio requerido: 1.1 GB

Espacio disponible: 231.2 GB

Nullsoft Install System v3.09

Figure 10: Destination Folder Selection



Instalación de RStudio

Elegir Carpeta del Menú Inicio

Elija una Carpeta del Menú Inicio para los accesos directos de RStudio.

Seleccione una carpeta del Menú Inicio en la que quiera crear los accesos directos del programa. También puede introducir un nombre para crear una nueva carpeta.

RStudio

Accessibility
Accessories
Administrative Tools
Anaconda3 (64-bit)
AutoFirma
Bandicam
Cloudpaging Player
ClustalW2
ConfiguradorFnmt
Control Station v3.7
Discord Inc

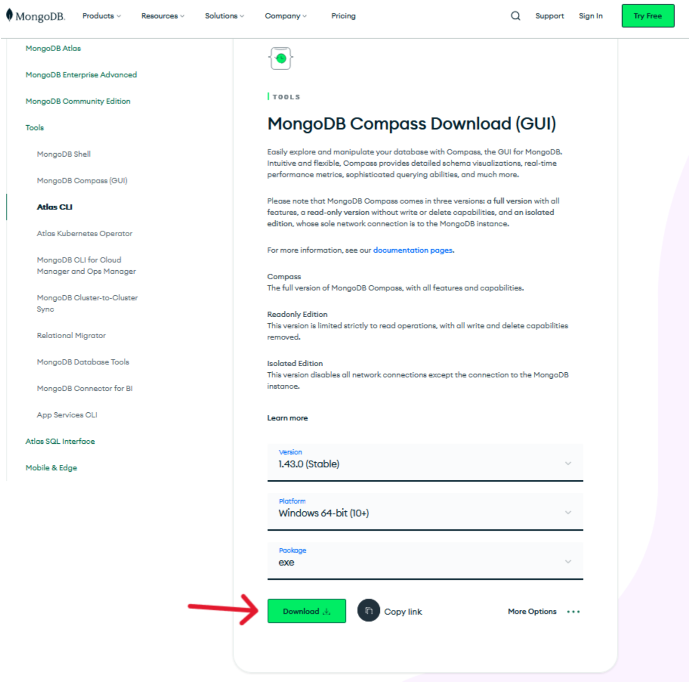
☐ No crear accesos directos

Nullsoft Install System v3.09

Figure 11: Shortcut Folder Selections

2.3 Installing MongoDB Compass

On the official MongoDB website [MongoDB, 2024], you can find the MongoDB Compass installer 12. In this case, no configuration is required; the application will install automatically upon executing the downloaded file.



MongoDB Compass Download (GUI)

Easily explore and manipulate your database with Compass, the GUI for MongoDB. Intuitive and flexible, Compass provides detailed schema visualizations, real-time performance metrics, sophisticated querying abilities, and much more.

Please note that MongoDB Compass comes in three versions: a full version with all features, a read-only version without write or delete capabilities, and an isolated edition, whose sole network connection is to the MongoDB instance.

For more information, see our documentation pages.

Compass
The full version of MongoDB Compass, with all features and capabilities.

Readonly Edition
This version is limited strictly to read operations, with all write and delete capabilities removed.

Isolated Edition
This version disables all network connections except the connection to the MongoDB instance.

Version: 1.43.0 (Stable)

Platform: Windows 64-bit (10+)

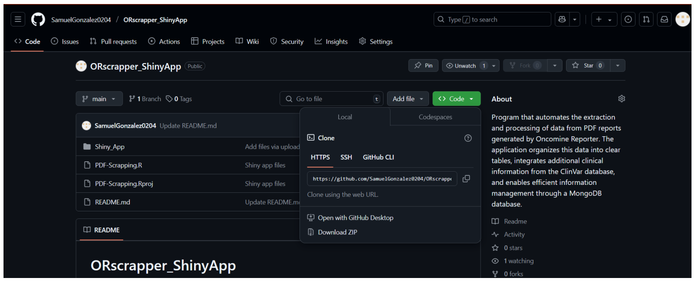
Package: exe

Figure 12: Download MongoDB Compass Installer

Finally, to launch the Shiny application, you only need to clone the GitHub repository, open the PDF-Scrapping.R file, and execute it.

The instructions to clone a repository [GitHub, 2024] are:

- Go to the GitHub repository page.
- Click the green button labeled "¡¿ Code" 13.
- Select the cloning method (HTTPS, SSH, or GitHub CLI).
- Copy the displayed URL.
- Open Git Bash and navigate to the directory where you want to clone the repository.
- Type "git clone" followed by the copied URL.



SamuelGonzalez0204 Update README.md

Shiny_App	Add files via upload
PDF-Scrapping.R	Shiny app files
PDF-Scrapping.Rproj	Shiny app files
README.md	Update README.md

https://github.com/SamuelGonzalez0204/ORscrapp

Clone using the web URL.

About

Program that automates the extraction and processing of data from PDF reports generated by Oncomine Reporter. The application organizes this data into clear tables, integrates additional clinical information from the ClinVar database, and enables efficient information management through a MongoDB database.

ORscrapper_ShinyApp

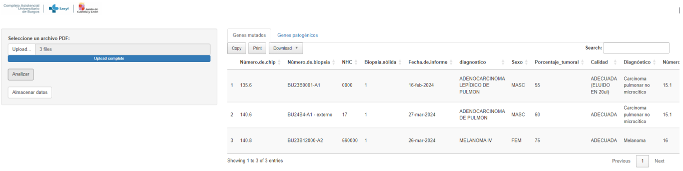
Figure 13: Repository Cloning

3 Manuals and/or Practical Demonstrations

Following the use case diagram in Annex F, demonstrations for each use will be shown.

3.1 Processing PDFs

Figure 14 shows how data has been extracted from three uploaded files.



Seleccione un archivo PDF: 3 files

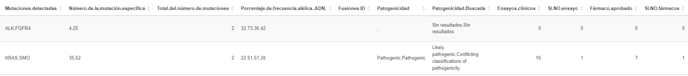
	Número.de.chip	Número.de.biopsia	NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Númer
1	135.6	BU23B0001-A1	0000	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1
2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
3	140.8	BU23B12000-A2	590000	1	26-mar-2024	MELANOMA IV	FEM	75	ADECUADA	Melanoma	16

Showing 1 to 3 of 3 entries

Figure 14: Demonstration of PDF Processing

3.2 Querying Searches in Clinical Databases

The column Patogenicidad.Buscada in the table of Figure 15 corresponds to the searches for mutations in the ClinVar database.

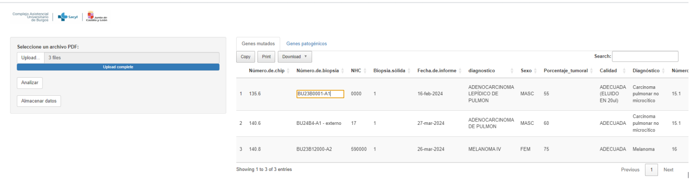


Mutaciones.detectadas	Número.de.la.mutación.especifica	Total.del.número.de.mutaciones	Porcentaje.de.frecuencia.alélica..ADN.	Fusiones.ID	Patogenicidad	Patogenicidad.Buscada	Ensayos.clinicos	SI.NO.ensayo	Fármaco.aprobado	SI.NO.fármacos
ALK,FGFR4	4,25	2	32.73,30.42		,	Sin resultados,Sin resultados	0	0	0	0
KRAS,SMO	35,52	2	22.51,51.20		Pathogenic,Pathogenic	Likely pathogenic,Conflicting classifications of pathogenicity	16	1	7	1

Figure 15: Demonstration of querying clinical databases

3.3 Modifying Tables

Double-clicking on any cell in the tables will open an editor where we can modify the value of that cell.



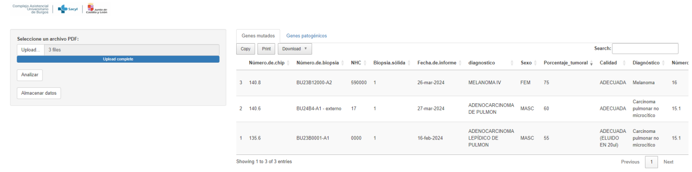
	Número.de.chip	Número.de.biopsia	NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Númerc
1	135.6	BU23B0001-A1	0000	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1
2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
3	140.8	BU23B12000-A2	590000	1	26-mar-2024	MELANOMA IV	FEM	75	ADECUADA	Melanoma	16

Showing 1 to 3 of 3 entries

Figure 16: Demonstration of table modification

3.3.1 Sorting Values

In this example 17, the column Porcentaje_tumoral has been sorted in descending order.



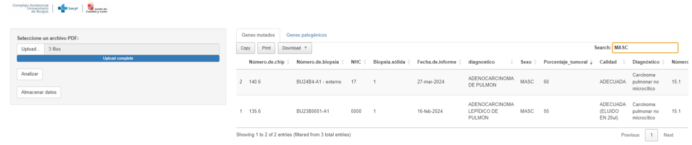
	Número.de.chip	Número.de.biopsia	NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Númerc
3	140.8	BU23B12000-A2	590000	1	26-mar-2024	MELANOMA IV	FEM	75	ADECUADA	Melanoma	16
2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
1	135.6	BU23B0001-A1	0000	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1

Showing 1 to 3 of 3 entries

Figure 17: Demonstration of sorting values

3.3.2 Filtering Rows

In Figure 18, only records corresponding to male patients have been filtered.



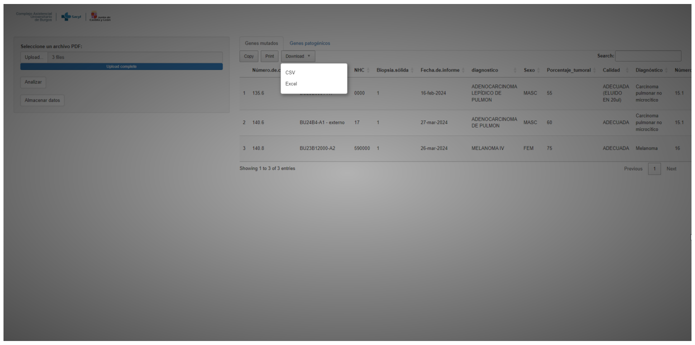
	Número.de.chip	Número.de.biopsia	NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Númerc
2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
1	135.6	BU23B0001-A1	0000	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1

Showing 1 to 2 of 2 entries (filtered from 3 total entries)

Figure 18: Demonstration of filtering rows

3.4 Exporting Data

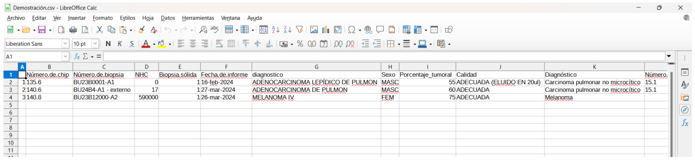
In this example, the data has been extracted into a CSV document 19, and the visualization of the downloaded file is shown in 20.



	Número.de.c		NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Númerc
1	135.6	[illegible]	0000	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1
2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
3	140.8	BU23B12000-A2	590000	1	26-mar-2024	MELANOMA IV	FEM	75	ADECUADA	Melanoma	16

Showing 1 to 3 of 3 entries

Figure 19: Demonstration of exporting data

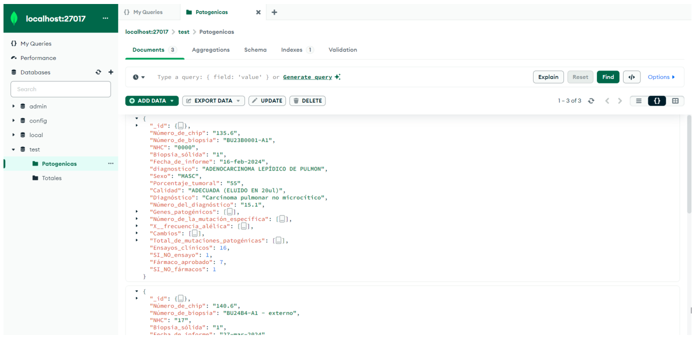

Demostración.csv - LibreOffice Calc
Archivo Editar Ver Insertar Formato Estilos Hoja Datos Herramientas Ventana Ayuda
Liberation Sans
10 pt
A1

	A	B	C	D	E	F	G	H	I	J	K	
1		Número.de.chip	Número.de.biopsia	NHC	Biopsia.sólida	Fecha.de.informe	diagnostico	Sexo	Porcentaje_tumoral	Calidad	Diagnóstico	Número.
2	1	135.6	BU23B0001-A1	0	1	16-feb-2024	ADENOCARCINOMA LEPÍDICO DE PULMON	MASC	55	ADECUADA (ELUIDO EN 20ul)	Carcinoma pulmonar no microcítico	15.1
3	2	140.6	BU24B4-A1 - externo	17	1	27-mar-2024	ADENOCARCINOMA DE PULMON	MASC	60	ADECUADA	Carcinoma pulmonar no microcítico	15.1
4	3	140.8	BU23B12000-A2	590000	1	26-mar-2024	MELANOMA IV	FEM	75	ADECUADA	Melanoma	

Figure 20: Demonstration of exporting data

3.5 Creating Patient History

In Figure 21, the records of the analyzed documents are shown in the MongoDB Compass application.



```
{
  "_id": {…},
  "Número_de_chip": "135.6",
  "Número_de_biopsia": "BU23B0001-A1",
  "NHC": "0000",
  "Biopsia_sólida": "1",
  "Fecha_de_informe": "16-feb-2024",
  "diagnostico": "ADENOCARCINOMA LEPÍDICO DE PULMON",
  "Sexo": "MASC",
  "Porcentaje_tumoral": "55",
  "Calidad": "ADECUADA (ELUIDO EN 20ul)",
  "Diagnóstico": "Carcinoma pulmonar no microcítico",
  "Número_del_diagnóstico": "15.1",
  "Genes_patogénicos": […],
  "Número_de_la_mutación_específica": […],
  "X__frecuencia_alélica": […],
  "Cambios": […],
  "Total_de_mutaciones_patogénicas": […],
  "Ensayos_clínicos": 16,
  "SI_NO_ensayo": 1,
  "Fármaco_aprobado": 7,
  "SI_NO_fármacos": 1
}

{
  "_id": {…},
  "Número_de_chip": "140.6",
  "Número_de_biopsia": "BU24B4-A1 - externo",
  "NHC": "17",
  "Biopsia_sólida": "1",
  "Fecha_de_informe": "27-mar-2024"
```

Figure 21: Demonstration of creating patient history

References

[CRAN, 2024] CRAN (2024). Download r-4.4.0 for windows. the r-project for statistical computing. https://cran.r-project.org/bin/windows/base/. Acceso realizado el 5 de junio de 2024.

[GitHub, 2024] GitHub (2024). Clonar un repositorio - documentación de github. https://docs.github.com/es/repositories/creating-and-managing-repositories/cloning-a-repository. Acceso realizado el 11 de junio de 2024.

[MongoDB, 2024] MongoDB (2024). Download mongodb atlas cli — mongodb. https://www.mongodb.com/try/download/atlascli. Acceso realizado el 5 de junio de 2024.

[Posit, 2024] Posit (2024). Rstudio desktop - posit. https://posit.co/download/rstudio-desktop/. Acceso realizado el 5 de junio de 2024.