



# Premios RRI 2022. Acceso Abierto datos *ómicos*. “Caminando hacia la Ciencia Abierta”

Pablo Mínguez

[pablo.minguez@quironosalud.es](mailto:pablo.minguez@quironosalud.es)

# Ciencia Abierta

La Ciencia Abierta es el conjunto de prácticas que permite que los conocimientos científicos estén disponibles, accesibles y sean reutilizables.



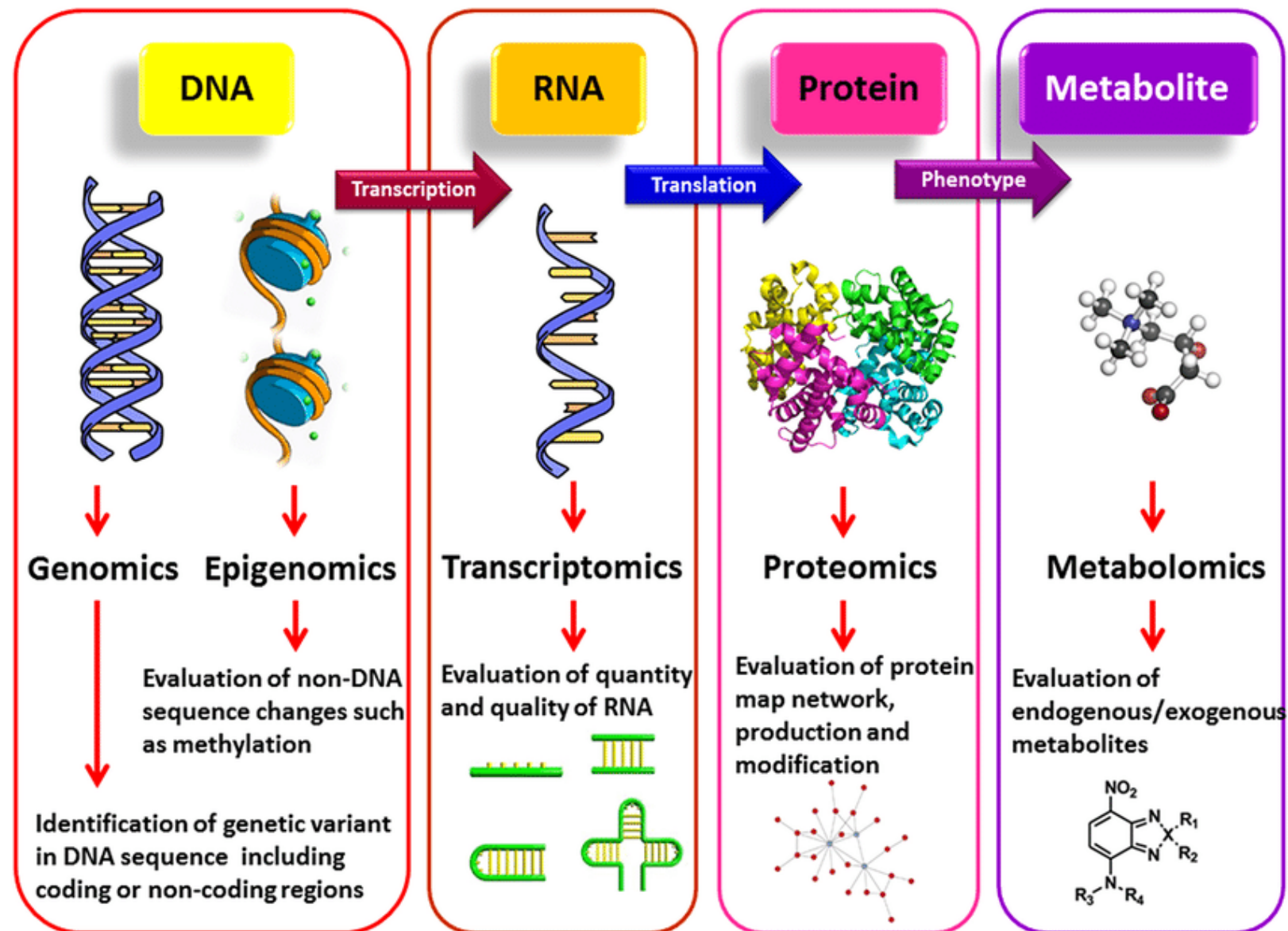
- **UNESCO** (41a reunión, nov 2021): marco internacional para las políticas y prácticas de la Ciencia Abierta.
- **Comisión Europea:** Ciencia Abierta como estándar de trabajo en programas de innovación e investigación (<https://bit.ly/EU-OpenScience-web>).
- **España** Ley de la Ciencia (art 37, septiembre 2022): Ciencia como bien común. Iniciativas para facilitar el libre acceso a los datos generados por la investigación. Transferencia de conocimiento a la sociedad.
- **IIS-FJD:** Política específica de Ciencia Abierta (<https://bit.ly/IIS-FJD-OpenScience>) desde 2019.

# Datos ómicos

Datos moleculares masivos. Producidos por experimentos *ómicos*.  
Y además llevan asociados metadatos.

## Datos:

Diferentes tipos de moléculas o niveles de regulación



## Origen:

- Partes de organismos (tumor, líneas celulares).
- Organismos
  - Humanos: algunos tipos de datos pueden identificar a una persona

## Metadatos:

- Condiciones de laboratorio, drogas suministradas, ...
- Características y condiciones
- Características, enfermedades,

# ¿Porqué se han de compartir los datos *ómicos*?

- **Transparencia** en la revisión del trabajo.
- **Fiabilidad** de los resultados. Los hace replicables.
- Permite un **uso secundario** regulado.
  - **Colaboración** entre productores y usuarios secundarios iniciadas mediante licencias.
  - **Productores:** Asegura el crédito. **Usuarios** secundarios: Asegura la fiabilidad.
  - **Rentabiliza** la inversión por las agencias financiadoras.
  - Aumenta el **impacto** y la visibilidad de la investigación.
- **Revistas y agencias financiadoras lo exigen.**

# Política sobre datos en los proyectos de investigación



[https://www.isciii.es/QueHacemos/Financiacion/Documents/PI\\_2023/PI23\\_PREGUNTAS\\_FRECUENTES.pdf](https://www.isciii.es/QueHacemos/Financiacion/Documents/PI_2023/PI23_PREGUNTAS_FRECUENTES.pdf)

- **¿Qué datos de investigación se deben depositar?**

Se han de depositar en repositorios de acceso abierto todos los datos que subyacen a la investigación, esto es, los datos brutos generados o producidos en el transcurso de la investigación. Así mismo, se deberán publicar junto a los artículos científicos los datos finales que sean necesarios para garantizar la verificación y reproducibilidad de los resultados presentados. Según las disciplinas y los flujos de producción, análisis y utilización de los datos, se pueden publicar los datos en diversos estadios a lo largo del ciclo de vida de la investigación. Asimismo, en el depósito y publicación de los datos se deberá tener en cuenta:

- La protección de los datos personales, que incluye la protección de las libertades y los derechos fundamentales de las personas físicas aplicados a un proyecto de I+D+I, así como su protección ante la posible utilización por terceros no autorizados.
- Los aspectos éticos, que afectan a los datos que pueden mostrarse, el tiempo y el anonimato de las personas implicadas, y respetan la dignidad y la integridad para garantizar su privacidad y confidencialidad.

Los requisitos específicos que las editoriales científicas puedan requerir a los autores, referentes a los datos utilizados específicamente para un artículo, así como los repositorios recomendados o reconocidos para ello.

- **¿Qué repositorios se podrán utilizar para depositar los datos?**

Se deberán depositar los datos de investigación en repositorios de confianza (TDR, Trusted Digital Repositories), institucional, multidisciplinar (ej. Zenodo); o bien en infraestructuras colaborativas de datos, en disciplinas que utilizan datos de forma masiva (DID, Data Intensive Domain), bien dentro del dominio o con carácter general o multidisciplinar (ej. EUDAT).

# Política sobre datos en los proyectos de investigación



[https://www.isciii.es/QueHacemos/Financiacion/Documents/PI\\_2023/PI23\\_PREGUNTAS\\_FRECUENTES.pdf](https://www.isciii.es/QueHacemos/Financiacion/Documents/PI_2023/PI23_PREGUNTAS_FRECUENTES.pdf)

## • ¿Qué tipo de investigación se deben depositar?

Se han de hacer de acceso abierto todos los datos que subyacen a la investigación, esto es, los datos brutos y los datos procesados de la investigación. Así mismo, se deberán publicar junto a los artículos científicos los datos necesarios para garantizar la verificación y reproducibilidad de los resultados presentados. Según las necesidades de acceso y utilización de los datos, se pueden publicar los datos en diversos estadios a lo largo del ciclo de vida de la investigación. Asimismo, en el depósito y publicación de los datos se deberá tener en cuenta:

**Compartición**  
**Aspectos éticos**



**Plan de Gestión**  
**de datos**

El acceso a los datos debe incluir la protección de las libertades y los derechos fundamentales de las personas implicadas en el proceso de I+D+I, así como su protección ante la posible utilización por terceros no autorizados. Debe tenerse en cuenta el tiempo y el anonimato de las personas implicadas, así como el tiempo y el anonimato de las personas implicadas, para garantizar su privacidad y confidencialidad. Los datos científicos pueden requerir a los autores, referentes a los datos utilizados en la investigación, que se depositen en los repositorios recomendados o reconocidos para ello.

## • ¿Dónde se utilizarán para depositar los datos?

Se deberán depositar los datos de investigación en repositorios de confianza (TDR, Trusted Digital Repositories), institucional, multidisciplinarios (ej. Zenodo), o bien en infraestructuras colaborativas de datos, en disciplinas que utilizan datos de forma masiva (DID, Data Intensive Domain), bien dentro del dominio o con carácter general o multidisciplinar (ej. EUDAT).

# Política sobre datos en los proyectos de investigación



[https://www.isciii.es/QueHacemos/Financiacion/Documents/PI\\_2023/PI23\\_PREGUNTAS\\_FRECUENTES.pdf](https://www.isciii.es/QueHacemos/Financiacion/Documents/PI_2023/PI23_PREGUNTAS_FRECUENTES.pdf)

Compartición  
Aspectos éticos



Plan de Gestión  
de datos

- **¿Qué es un Plan de Gestión de Datos, PGD?**

Un Plan de Gestión de Datos (PGD o DMP, Data Management Plan, por sus siglas en inglés) es una declaración formal que describe los datos que se utilizan y producen durante el curso de las actividades de investigación. El PGD establece cómo se van a gestionar los datos en un proyecto de investigación, tanto durante el proyecto como una vez que éste ha finalizado.

El PGD define la metodología de extracción de la información, la unidad de medida, los procesos de creación, limpieza y análisis, el formato, el software de lectura y procesamiento, las licencias y permisos de acceso y (re)utilización, su sistema de control de versiones, su autoría, su difusión, y su lugar de almacenamiento y preservación a largo plazo. Este plan se creará al comienzo del proyecto de investigación financiado y, según la naturaleza y evolución del proyecto, puede que no sea un documento definitivo, sino que variará, se completará y/o modificará a lo largo del ciclo de vida de la investigación.

- **¿Debo incluir en la solicitud de mi proyecto el PGD?**

**NO**, en la memoria de la propuesta solamente se deberá incorporar una descripción inicial que contenga qué datos se van a recoger o generar en el marco del proyecto (tipologías y formatos), cómo será el acceso a los mismos (quién, cómo y cuándo se podrá acceder a ellos), de quién son los datos y en qué repositorio está previsto su depósito, difusión y preservación. Asimismo, se recogerán, en su caso, las condiciones éticas o legales específicas que los regulen (ej. privacidad de los datos y su reglamentación; datos protegidos o protegibles por propiedad intelectual o industrial, etc.) que condicionen su disponibilidad, uso y/o reutilización. Esta declaración inicial en el momento de la solicitud no condicionará la creación de un Plan de Gestión de Datos formal en el caso de que el proyecto resulte finalmente financiado.

- **¿Cuándo debo enviar el Plan de Gestión de Datos?**

El PGD se deberá presentar, si así es requerido, junto con los informes de seguimiento intermedio y final del proyecto financiado. Asimismo se recomienda la publicación en acceso abierto del PGD junto a los datasets utilizados, y en formato legible por máquina.

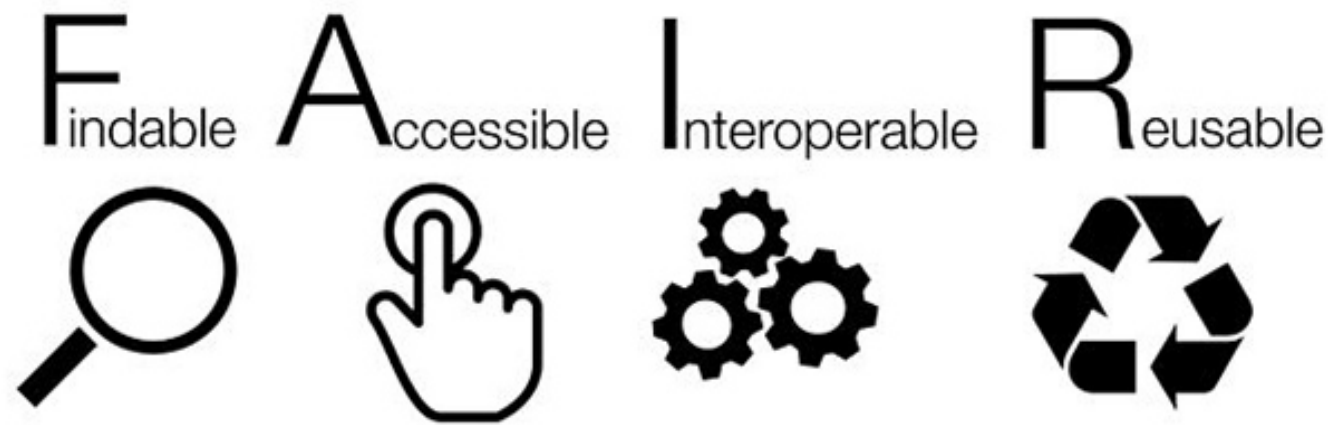
Se deberán utilizar para datos de investigación colaborativas de datos, en disciplinas que utilizan datos de forma masiva (DID, Data Intensive Domain), bien de carácter general o multidisciplinar (ej. EUDAT).

# Plan de Gestión de datos

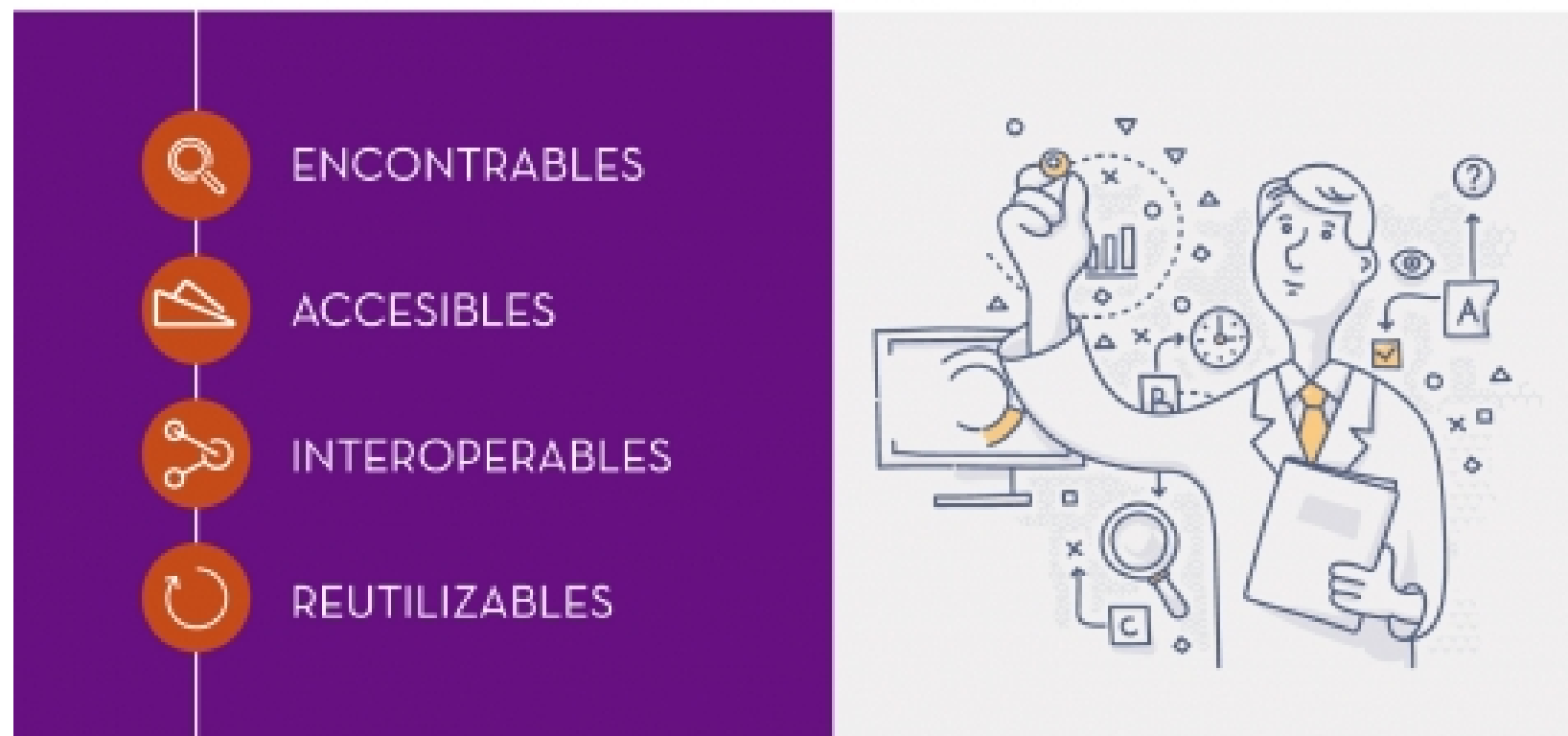
- Política por la cual se gestionan los datos en un proyecto de investigación, durante y después de su ejecución.
- Asigna responsabilidades y recursos: **Gobernanza**.
- Garantiza un adecuado tratamiento, protección y salvaguarda de todos los datos recogidos, generados o procesados.
- Cumplimiento de leyes aplicables: RGPD (Reglamento General de Protección de Datos), LOPDGDD (Protección de Datos Personales y garantía de los derechos digitales).

- En cuanto a **datos ómicos**: Establece como se manejan y comparten estos datos: **Principios FAIR**.

# Los datos *ómicos* han de cumplir los Principios FAIR



## Principios FAIR para los Datos Científicos



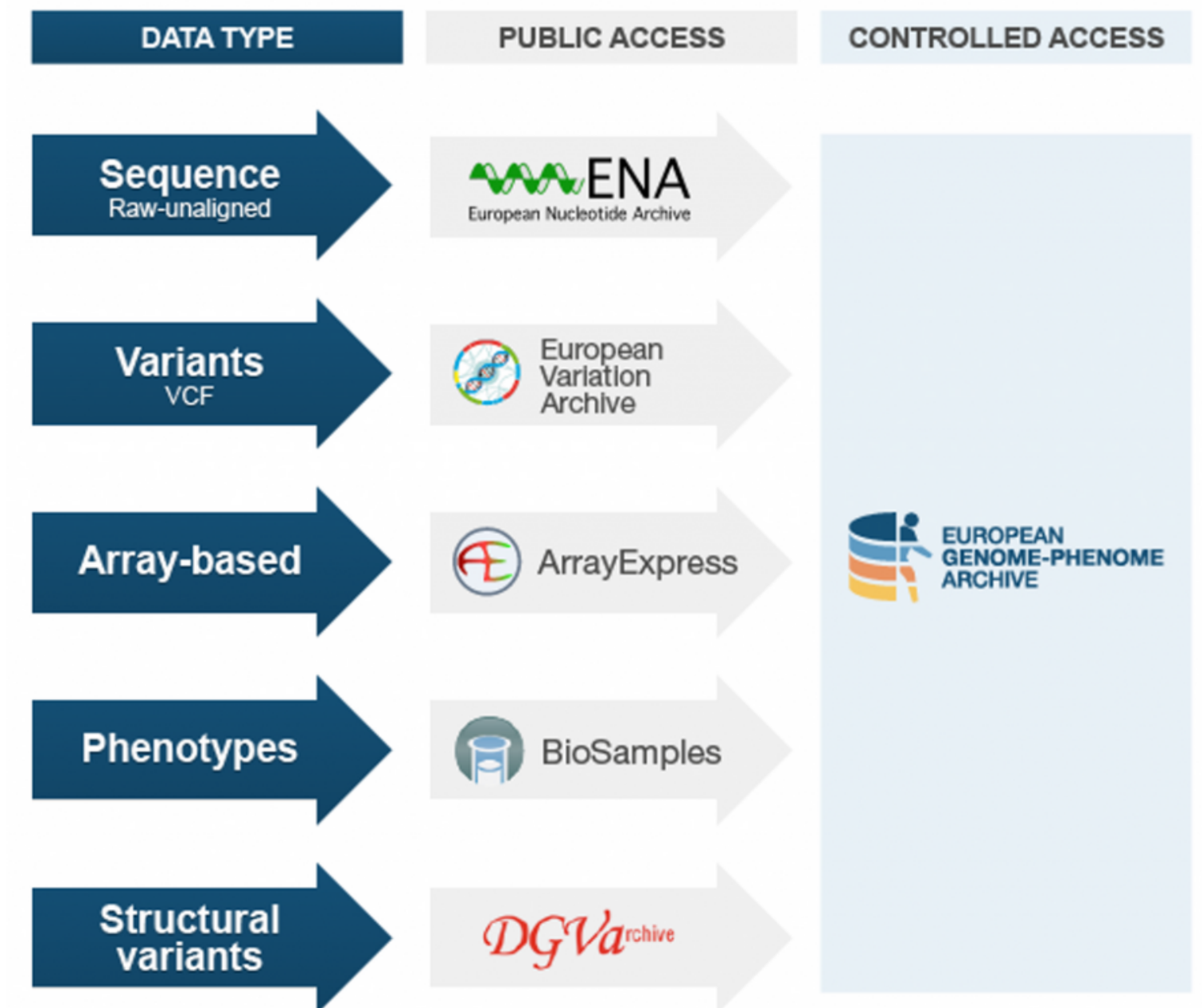
## Datos *ómicos* FAIR:

- **Encontrables:** Datos y metadatos pueden ser encontrados mediante herramientas de búsqueda. ID y repositorio reconocido.
- **Accesibles:** Pueden ser descargados (no necesariamente de forma libre). Su reutilización está regulada.
- **Interoperables:** Tienen formatos estándares para permitir su intercambio y reutilización.
- **Reusables:** Pueden ser usados por terceros bajo unas reglas claras (licencia), conociendo su procedencia y permitiendo su reconocimiento.

# El depósito en repositorios reconocidos permite cumplir los principios FAIR

- **Secuencias de DNA and RNA:** ENA (<https://www.ebi.ac.uk/ena>), datos de libre acceso.
- **Datos de expresión génica (transcriptómica):** Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) o ArrayExpress (<https://www.ebi.ac.uk/biostudies/arrayexpress>), datos bajo demanda/licencia (crudos), libre acceso (conteos).
- **Datos asociación genotipo-fenotipo:** European Genome-phenome Archive (EGA, <https://ega-archive.org>), datos bajo demanda/licencia (crudos + metadatos).
- **Polimorfismos genéticos:** dbSNP, dbVar, clinVAR, o European Variation Archive (EVA), datos libre acceso.

Repositorios reconocidos:  
<https://www.re3data.org>



# Política de Ciencia Abierta del IIS-FJD

Investigación e Innovación Responsable - Responsible Research and Innovation (RRI).

Aspectos clave de la RRI:

2019

<https://bit.ly/IIS-FJD-OpenScience>

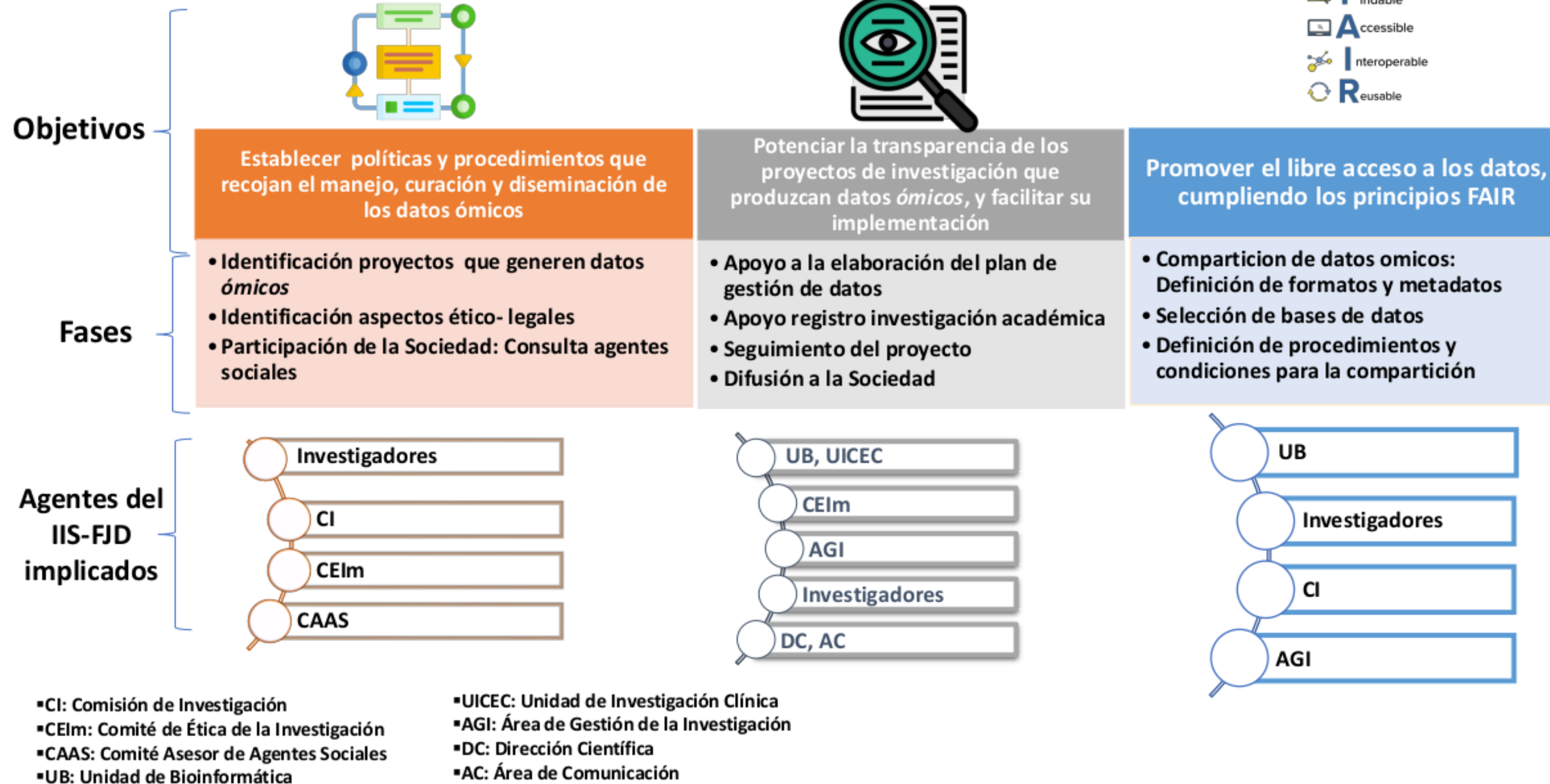


1.- Investigación e Innovación Responsable. A partir de información disponible del proyecto RRI Tools (<https://www.rri-tools.eu/>)

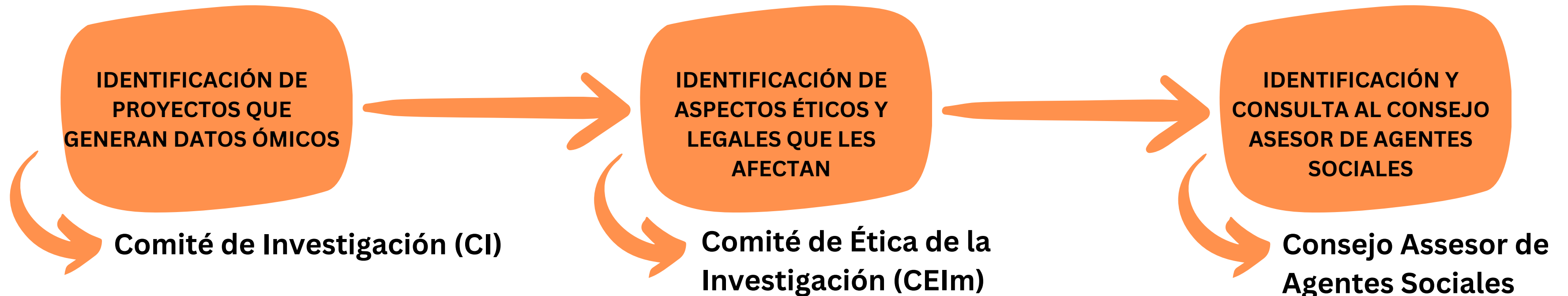
# Caminando hacia la Ciencia Abierta: Implementación del acceso abierto a los datos *ómicos* en el IIS-FJD (RRI22/00005)



Promover la práctica de la Ciencia Abierta en el IIS-FJD, en particular en lo referente a los datos *ómicos*



# Objetivo 1: Establecer políticas y procedimientos que recojan el manejo, curación y diseminación de datos ómicos



## Antes de la solicitud:

- Revisión de proyectos de investigación a presentar en convocatorias competitivas
  - Control de requisitos.
  - Calidad
  - Necesidad de asesoramiento
- **Identificación de contenido ómico y necesidades según tipo:**
  - humanos, otros organismos, líneas celulares.
- **RESULTADO: Visto Bueno y asesoramiento.**

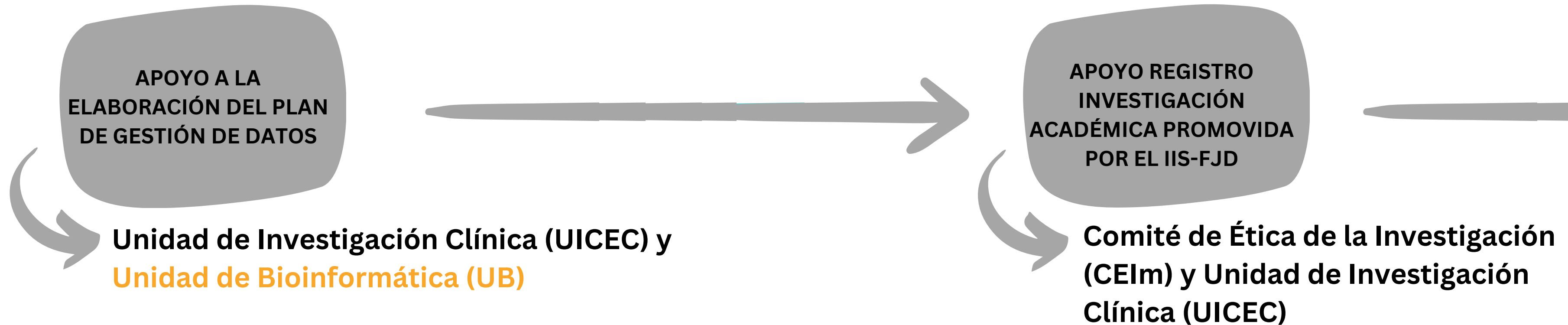
## Antes de la solicitud:

- Revisión de proyectos en seres humanos (datos o muestras).
- ## Una vez financiado:
- Evaluación de aspectos metodológicos y ético-legales: Consentimientos Informados, manejo de muestras biológicas, análisis genéticos.
  - Revisión de aspectos relativos a la compartición de datos, tratamiento, pseudonimización y uso secundario.
  - **RESULTADO: dictamen del CEIm.**

## Antes de la solicitud:

- Asesoramiento sobre aspectos complementarios al desarrollo científico que orienten mejor nuestros resultados hacia la sociedad.
- **RESULTADO: Incorporación de aspectos sociales de relevancia.**

# Objetivo 2: Potenciar la transparencia de los proyectos de investigación que produzcan datos ómicos, y facilitar su implementación



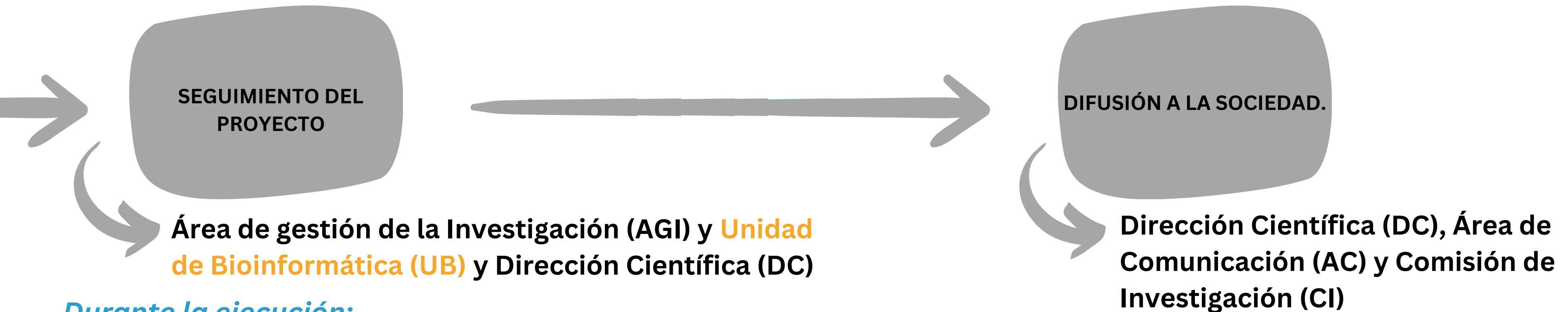
## *Redacción de la solicitud:*

- Asesoramiento en la elaboración del Plan de Gestión de Datos
  - tipología y formato de datos a recoger/generar
  - procedimiento para acceso a los mismos
  - titularidad de los datos
  - repositorio para su depósito
  - garantías de requisitos éticos o legales
- Asesoramiento en el Plan de Difusión de los resultados orientados a la sociedad, entre otros.
- **RESULTADO: Planes de Gestión de Datos y Difusión.**

## *Una vez financiado:*

- Registro de ensayos clínicos en una base de datos de acceso público
- El CEIm emite inclusión expresa de la obligación y la UICEC presta soporte para alta y registro.
- **RESULTADO: Registro de investigaciones clínicas.**

# Objetivo 2: Potenciar la transparencia de los proyectos de investigación que produzcan datos ómicos, y facilitar su implementación



## *Durante la ejecución:*

- Información sobre compromisos del IP con agentes financiadores.
- **Apoyo en el planteamiento y/o análisis de datos ómicos (diseño experimental y análisis)**
- **Depósito de datos.**
  - **Asesoramiento personalizado**
  - **Cursos/sesiones informativas**
- Registro interno de datos *ómicos* generados y compartidos.
- Cálculo de indicadores.
- **RESULTADO: Correcta ejecución del proyecto en relación al tratamiento de datos.**

## *Durante la ejecución:*

- Diseminación del conocimiento
- Participación de la sociedad en el proceso investigador e innovador (CAAS).
- Canales utilizados:
  - Noticias y notas de prensa: <https://www.fjd.es/iis-fjd/es/actualidad/noticias>
  - Redes sociales (Twitter, Facebook).
  - Pantallas internas.
- **RESULTADO: Integración de la sociedad en la investigación y difusión de resultados.**

# Objetivo 3: Promover el acceso abierto a los datos cumpliendo los principios FAIR

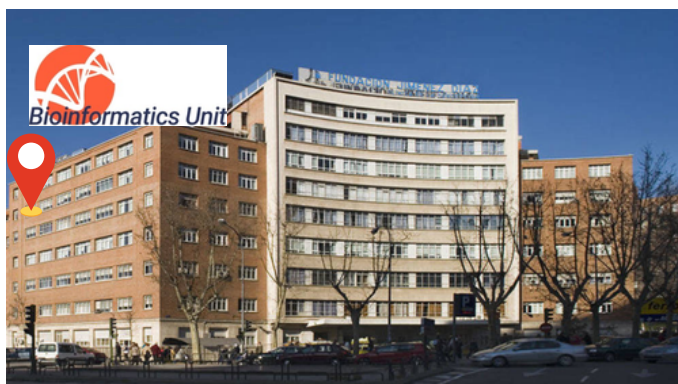
IMPLEMENTACIÓN DE  
PRINCIPIOS FAIR

Unidad de Bioinformática (UB)

## *Durante la ejecución:*

- Aplicación de principios FAIR a los datos *ómicos*:
  - Estandarización de los datos *ómicos*
  - Identificación y estandarización de metadatos
  - Selección de repositorios adecuados
  - Identificación de los procedimientos, condiciones y agentes implicados en la compartición.
- Reutilización interna de datos genómicos.
- Aplicación de políticas de Ciencia Abierta al software de análisis de datos *ómicos*.
- **RESULTADOS:** Integración de la Ciencia Abierta en los datos *ómicos* producidos por el IIS-FJD.

# Objetivos de la Unidad de Bioinformática del IIS-FJD



<https://bit.ly/UnidadBioinformaticallIS-FJD>

desde 2015

- **Pablo Mínguez (coordinador)**
  - Lorena de la Fuente (bioinformática). Técnico Bioinfo AES18 (01/01/2019 - 30/04/2021)
  - Gonzalo Núñez Moreno (bioinformático) Ayud. Inv. CAM. (01/03/2021 - 15/04/2023)
  - **Técnico Bioinformático AES23**
  - **Fortalece 2023**
- **Apoyo al diseño y análisis de experimentos ómicos** (Transcriptómica, Genómica, Proteómica, Epigenómica) a grupos del IIS-FJD (y externos) en régimen de colaboración.
  - **Diseño, implementación y mantenimiento de software bioinformático.**
  - **Asistencia al IIS-FJD en proyectos institucionales** con necesidades bioinformáticas y **de compartición de datos.**
    - Fortalece, IMPaCT-Genómica (IMP/00009), IMPaCT-Data (IMP00019), RRI22/00005, TransBioNet (RED2018-102404-T), unCOVER (H2020), STOP-Coronavirus (COV20\_00181), 2 proyectos integrados ISCIII (PIE13/00051 y PIE15/00048).
  - **Transferencia de conocimiento.**
    - **Definición de estándares** en la comunidad científica (**IMPaCT, CIBERER, TransBioNet, SEBiBC**).
    - **Formación** (cursos, TFM/Gs, clases en Másteres, congresos).
    - Asesoramiento/Apoyo en la **compartición de datos ómicos**.

# Servicios de la Unidad de Bioinformática (Transcriptómica)

## RNASeq (mRNA, miRNA)

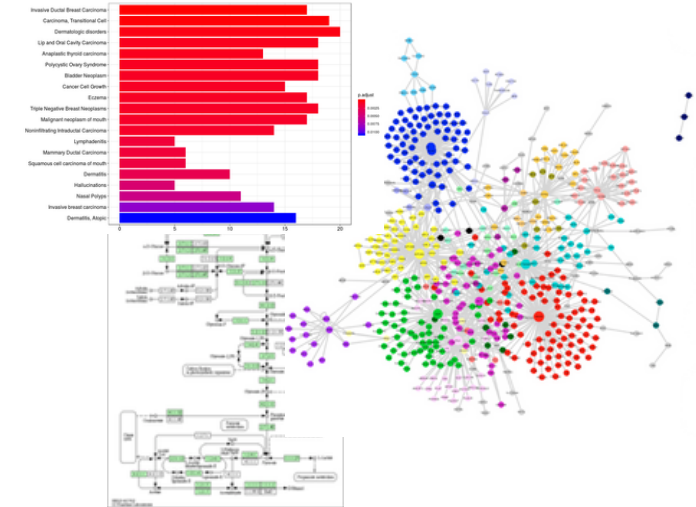
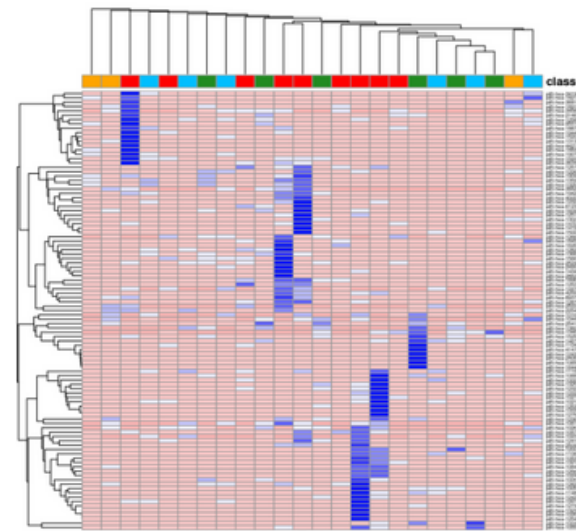
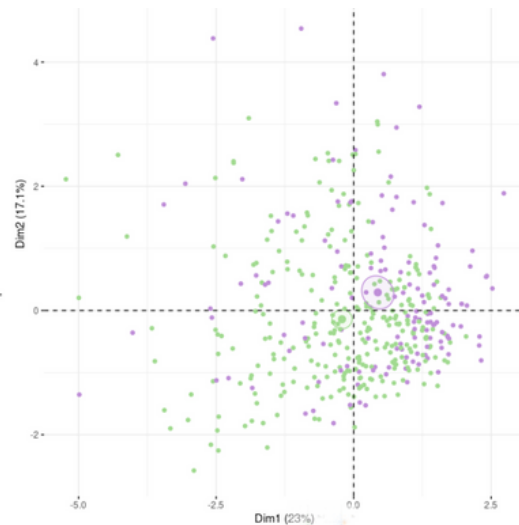
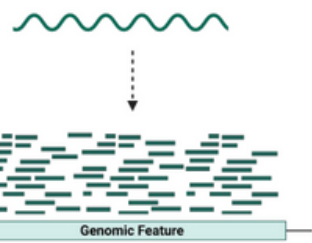
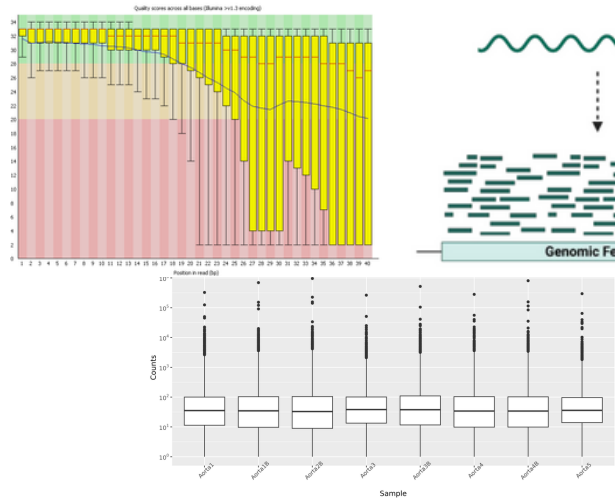
Diseño experimental (batches). Control de calidad. Procesado de muestras (normalización, mapeo, conteo de lecturas)

Descripción de muestras (PCA, clustering, heatmaps)

Expresión diferencial (firmas moleculares) Clasificadores/predictores (aprendizaje automático)

Caracterización funcional (enriquecimientos, redes)

**FAIRificación**  
Producción de estándares y compartición



- Dato crudo (fastq)
- Procesado (conteos)
- Metadatos

**Seguimiento: escritura, propuestas de nuevos análisis y compartición**

# Resultados: Depósito de datos *ómicos* (Transcriptómica)

RNASeq (mRNA, miRNA)



<https://www.ncbi.nlm.nih.gov/geo/>

**142** experimentos accesibles y otros en proceso

## *Proyectos y artículos:*

- PI13/00051. Del Puerto-Nevado *et al.*, Mol Onco. 2019. **GSE115329, GSE115313**
- Mendez-Barbero *et al.*, Ebiomedicine. 2019. **GSE114116**
- Gonzalez-Alonso *et al.*, 2020. **GSE89216**
- Serrano-Lopez *et al.*, Leuk Lymphoma. 2022. **GSE181458**
- Nuñez-Borque *et al.*, Pediatr Allergy Immunol. 2021. Nuñez-Borque *et al.*, Sec. Molecular Innate Immunity. 2023. **GSE245653**

## *En cuarentena, a la espera de publicación:*

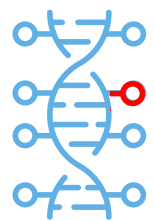
- Laboratorio de Patología Vascular - mRNAs
- Laboratorio de Nefrología - mRNAs
- Microbiología - miRNAs

# Servicios de la Unidad de Bioinformática (Genómica)

## DNaseq (germinales)

**Pipeline**  
nextflow

- CES, WES, WGS
- long, short reads



**SNVs**

GATK  
DRAGEN  
DeepVariant

Control de calidad

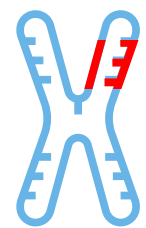
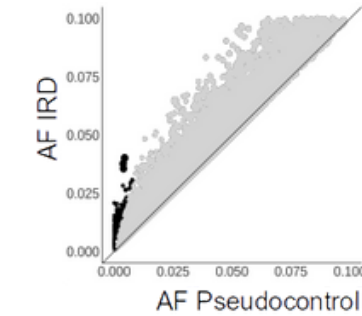
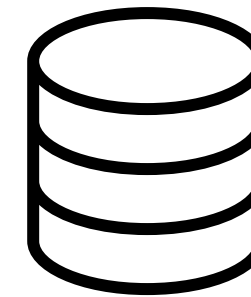
Mapeo

Llamada de variantes  
(3 software)

Anotación de variantes

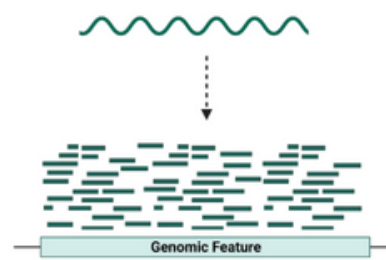
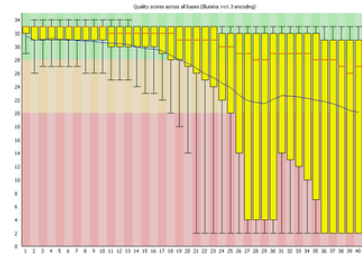
Filtrado y priorización  
(variantes y genes candidatos)

Base de datos de frecuencias poblacionales interna (**Genética**) (6698 casos EERR).



**CNVs, SVs**

connvading  
exomedepth  
gatk  
panelmops

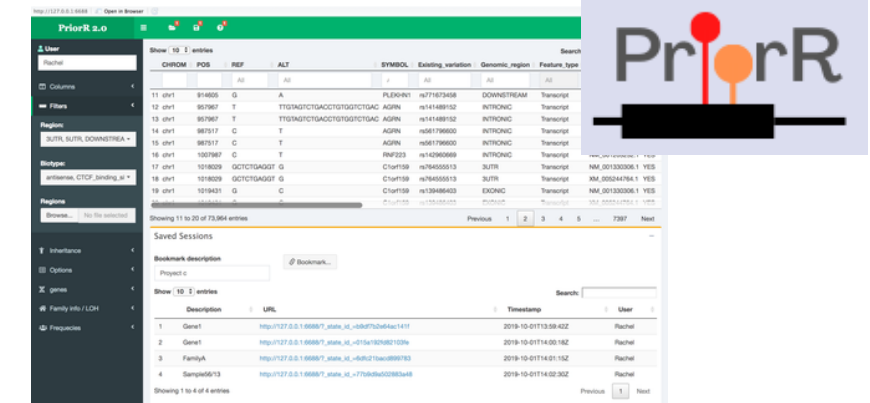


DeepVariant



**Ve!P**

**AnnotSV**



Programa de visualización

## DNaseq (somáticas, mosaico)

# Servicios de la Unidad de Bioinformática (Genómica)

## DNASEq

**FAIRificación:** Generación de estándares y  
Compartición de datos



- Producir los estándares
  - Datos (fastqs)
  - Metadatos
  - Ontologías (HPO)
- Encriptado
- Definir reglas de compartición (Ontología)
- Definir el Data Access Committee (DAC)

## Genetic landscape of inherited retinal dystrophies

In the context of research, this dataset contains 423 IRD samples; 411 of them analyzed with Clinical Exome Sequencing solutions, and 12 with Whole Exome Sequencing.

423 samples DAC: EGAC00001002010 Technology: NextSeq 500

Access Policy 1 Study 846 Files (865.4 GB) Metadata Request Access

DS NPUNCU IRB MOR US PS IS RTN

IIS- Fundacion Jimenez Diaz Data Access Agreement. Please contact any of the members of the IIS- Fundacion Jimenez Diaz DAC to access the whole document.

**DUO:000007 (version: 2019-01-07)**disease specific research This data use permission indicates that use is allowed provided it is related to the specified disease.  
**Modifiers** MONDO:0019117

Classes (308) Properties (134) Individuals (20)

Tree List

- data use modifier (18)
  - clinical care use
  - collaboration required
  - ethics approval required
  - genetic studies only
  - geographical restriction
  - institution specific restriction
  - no general methods research
  - not for profit, non commercial use only (2)
    - non-commercial use only
    - not for profit organisation use only
  - population origins or ancestry research prohibited
  - project specific restriction
  - publication moratorium
  - publication required
  - research specific restrictions
  - return to database or resource
  - time limit on use
  - user specific restriction
- data use permission (5)
  - general research use (2)
    - health or medical or biomedical research (1)
      - disease specific research
    - no restriction
    - population origins or ancestry research only
  - investigation OBI (10)

# Resultados: Depósito de datos ómicos (Genómica)

**DNaseq**



**1238** muestras secuenciadas (CES, WES, WGS) compartidas en EGA y **registradas en el proyecto B1MG**

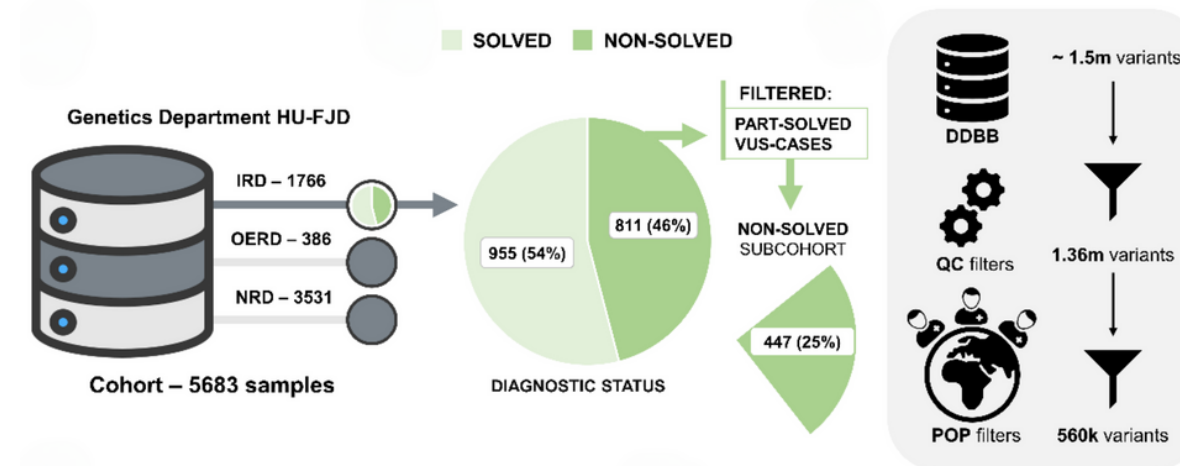
## Proyectos y artículos:

- Proyectos europeos: B1MG, EasiGenomics, unCoVer
- Proyectos nacionales: STOP-Coronavirus, IMPaCT
- Proyectos nacionales intramurales
  
- Perea-Romero *et al.*, Human Genetics 2021
- Perea-Romero *et al.*, Sci Rep, 2021
- Perea-Romero *et al.*, npj genomic medicine 2022
- Lopez-Rodriguez *et al.*, Sci Rep 2022
- Romero *et al.*, npj genome medicine 2022
- Lopez-Rodriguez *et al.*, Andrology 2022.
- Damian-Verde *et al.*, Human Genomics. 2023

Testeadores de la nueva versión de EGA



**Otros resultados:** Base de datos de frecuencias de la cohorte IIS-FJD (Iancu *et al.*, IJMS 2022)



Participación en proyectos COVID

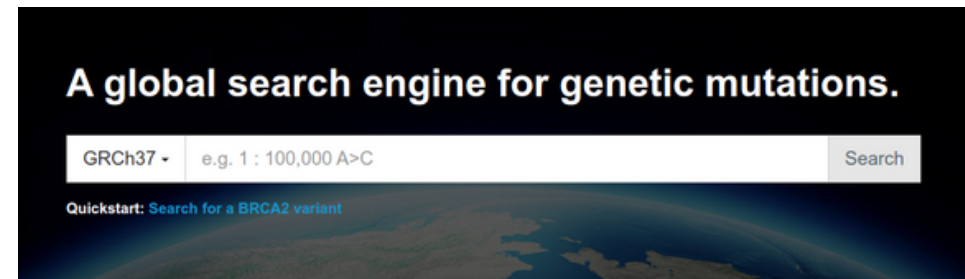


# Resultados: Depósito de datos ómicos (Genómica)

DNaseq

Redes Beacon

<https://beacon-network.org>



Collaborative Spanish Variant Server

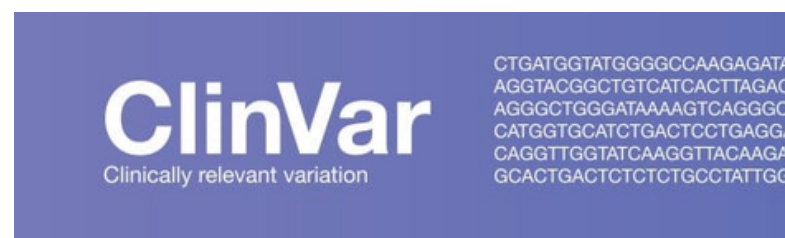
<http://csvs.babelomics.org>

784  
muestras  
en ambos

Chr	Position	Alleles	Gene	Id	Genotypes <sup>Ⓞ</sup>				Freq. <sup>Ⓞ</sup>			1000G AAF (phase 3) <sup>Ⓞ</sup>		ExAC AAF <sup>Ⓞ</sup>	ESP6500 AAF <sup>Ⓞ</sup>		gnomAD AAF <sup>Ⓞ</sup>		SIFT <sup>Ⓞ</sup>	Po
					0/0	0/1	1/1	./.	0 Freq	1 Freq	MAF	ALL	EUR <sup>Ⓞ</sup>	ALL	ALL	EA <sup>Ⓞ</sup>	GENOME	ALLELOME	ALL	
1	215796330	AA>	KCTD3.USH2A	rs60500520	457	6	0	0	0.994	0.006	0.006	0.056	0.003	.	.	.	0.030	.	.	
1	215796350	T>G	KCTD3.USH2A	rs6667509	333	115	15	0	0.843	0.157	0.157	0.103	0.108	.	.	.	0.099	.	.	
1	215796405	>T	KCTD3.USH2A	rs138675447	462	1	0	0	0.999	0.001	0.001	0.024	.	.	.	.	0.019	.	.	
1	215796517	C>A	KCTD3.USH2A	rs74964983	462	1	0	0	0.999	0.001	0.001	0.022	.	.	.	.	0.017	.	.	
1	215796529	T>C	KCTD3.USH2A	rs187725500	462	1	0	0	0.999	0.001	0.001	0.000	0.001	.	.	.	0.000	.	.	
1	215796575	A>G	KCTD3.USH2A	rs60590154	458	5	0	0	0.995	0.005	0.005	0.023	0.003	.	.	.	0.003	.	.	
1	215796618	G>A	KCTD3.USH2A	rs571469756	462	1	0	0	0.999	0.001	0.001	0.000	0.001	.	.	.	.	.	.	
1	215796730	C>T	KCTD3.USH2A	rs79275452	462	1	0	0	0.999	0.001	0.001	0.022	.	.	.	.	0.017	.	.	
1	215796735	G>C	KCTD3.USH2A	rs770103003	462	1	0	0	0.999	0.001	0.001	.	.	.	.	.	.	.	.	
1	215796796	C>T	KCTD3.USH2A	rs79178663	462	1	0	0	0.999	0.001	0.001	0.025	0.001	.	.	.	0.026	.	.	

2 artículos asociados:

- Lopez-Lopez *et al.*, Hum Genomics 2023
- Peña-Chilet *et al.*, NAR 2021



También apoyamos la compartición de información clínica de variantes genómicas (Iancu *et al.*, npj genome medicine 2021)

# Resultados: Depósito de datos ómicos (IMPACT-Genómica)

DNASeq

<https://genomica-impact.es>

WP3: Enfermedades raras  
Grupos de trabajo:

Coordinadores Grupos de Trabajo

MODELO DE INFORME  
DIAGNÓSTICO



Dr. Miguel Ángel Moreno  
Hospital Universitario  
Ramón y Cajal  
CIBERER

CREACIÓN RED  
DE EXPERTOS



Dra. Belén Pérez  
Centro de Diagnóstico de  
Enfermedades Moleculares  
CIBERER

CREACIÓN RED DE  
EXPERTOS



Dr. Víctor Martínez  
Hospital Universitario Parc Taulí

FENOTIPADO HPO



Dr. Jaime Cruz  
Hospital Universitario 12 de  
Octubre

VALIDACIÓN  
FUNCIONAL



Dr. Francesc Palau  
Hospital Sant Joan de Déu  
CIBERER

ALMACENAMIENTO Y  
DATA SHARING



Dr. Pablo Mínguez  
Instituto de Investigación Sanitaria  
Fundación Jiménez Díaz  
CIBERER



## Grupo de Trabajo Data sharing

### Funciones

- > Estudio de la situación actual de la red de hospitales del ámbito de enfermedades raras, en cuestión de datos genómicos acumulados y su compartición.
- > Divulgación de las posibilidades de EGA/*Beacons*.
- > Apoyo para que las secuencias generadas en IMPACT-GENÓMICA estén disponibles en EGA/*Beacon*.
- > Facilitar la implantación de una infraestructura para alojar y compartir datos genómicos.

### Tareas

- > Encuesta a los coordinadores de las CCAA de enfermedades raras para detectar cuál es la situación actual en cuestión de datos genómicos acumulados y su compartición.
- > Taller online abierto sobre EGA/*Beacon* el 27 de enero de 2022, con más de 200 asistentes. La grabación está disponible aquí: <https://youtu.be/RI9wCs7JVKw>.
- > Creación de un grupo de trabajo transversal [GdT-T] con el programa IMPACT-Data para ofrecer soluciones EGA/*Beacon* acordes a las necesidades de IMPACT-GENÓMICA.

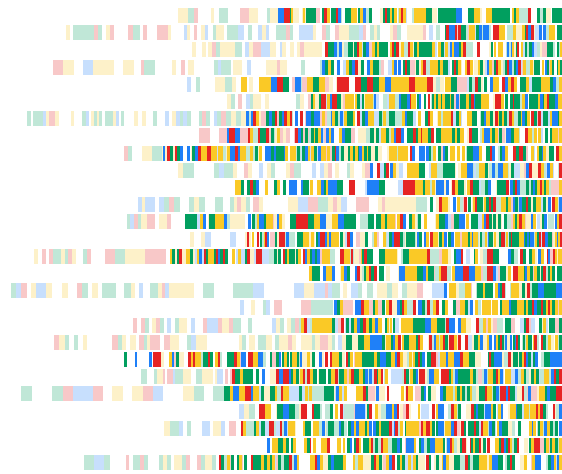
# Servicios de la Unidad de Bioinformática (Genómica de virus y otros)

## DNaseq SARS-CoV-2, Epigenética y otros

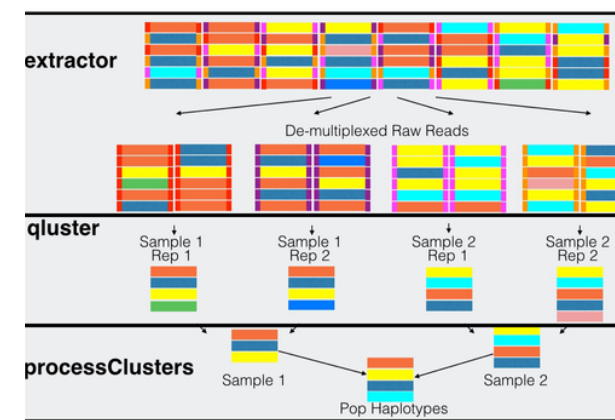
Pre-procesado de las muestras secuenciadas a alta profundidad

Extracción de haplotipos en la muestras

FAIRificación de los datos



<https://seekdeep.brown.edu>



**ENA**  
European Nucleotide Archive

- Preprocesamiento de los datos
- Estandarización de los metadatos
- Depósito

Otro tipo de datos que estamos manejando: Epigenética (Bismark + RnBeads) y Proteómica.

# Servicios de la Unidad de Bioinformática (Genómica de virus)

## DNaseq SARS-CoV-2



**110** secuencias víricas (SARS-CoV2) en colaboración con Microbiología (Celia perales)

### *Proyectos y artículos:*

- Proyectos nacionales: STOP-Coronavirus
- Soria *et al.*, Access Microbiol. 2021
- Martínez-Gonzalez *et al.*, Microbiol Spect. 2022
- Martínez-Gonzalez *et al.*, J Clin Invest. 2022
- Martínez-Gonzalez *et al.*, Pathogens. 2022
- Somovilla *et al.*, Antimicrob Agents Chemother. 2023

# Cursos y Seminarios sobre compartición de datos 2022/2023

## CURSO IIS-FJD: INTRODUCCIÓN AL ANÁLISIS DE DATOS ÓMICOS

Aprende conceptos y recursos básicos para el análisis de experimentos de DNaseSeq, RNASeq y proteómica

**Cuándo:** 10 y 11 de octubre de 2022, de 9 a 12:30 am

**Dónde:** Aula Francisco Grande HU-FJD

**Tipo:** Presencial

**Imparte:** Unidad de Bioinformática IIS-FJD

### PROGRAMA:

**Módulo 1: Análisis Genómico.** Bases de datos, protocolos de detección de variantes y análisis.

Raquel Romero, 2 horas

**Módulo 2: Análisis Transcriptómico.**

Aplicaciones, pasos de un protocolo de cuantificación de expresión, análisis exploratorio de las muestras, normalizaciones y expresión diferencial.

Gonzalo Núñez, 1.5 horas

**Módulo 3: Análisis e interpretación funcional de experimentos ómicos.** Bases de datos, programas de enriquecimiento funcional y de análisis de redes biológicas

Pablo Mínguez, 3 horas

**Módulo 4: Compartición de datos genómicos y transcriptómicos.**

Pablo Mínguez, 0.5 horas

**Programa completo:**

<https://bit.ly/ProgramaCursoBioinfoIISFJD2022>

### INSCRIPCIONES:

Haz tu inscripción [aquí!](#)

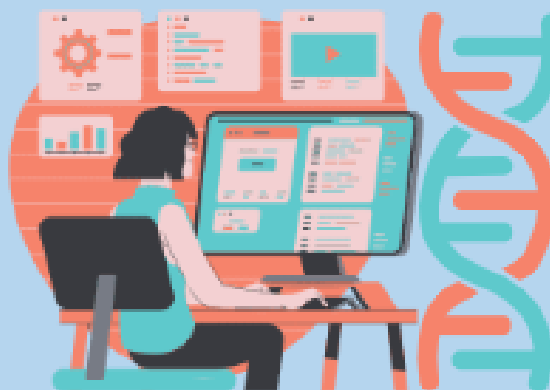
**Antes del 23 de Septiembre**

**Número de plazas:** 30.

Se priorizará que todos los laboratorios tengan acceso al curso con al menos 1 persona, el resto de plazas se asignará por orden de inscripción.

**Prácticas guiadas, trae tu ordenador portátil**

**Información sobre la Unidad de Bioinformática del IIS-FJD:**  
<https://bit.ly/UnidadBioinformaticaIIS-FJD>



**iis FJD** INSTITUTO DE INVESTIGACIÓN SANITARIA FUNDACIÓN JIMÉNEZ DÍAZ

## CURSO IIS-FJD: INTRODUCCIÓN AL ANÁLISIS DE DATOS ÓMICOS

### Módulo 4: Recursos para la compartición de datos genómicos y transcriptómicos.

Pablo Mínguez, 0.5 horas

ÁREA: GENÉTICA Y GENÓMICA

### Plan de Gestión de Datos ómicos de investigación. Aproximación tecnológica y legal

Pablo Mínguez y Guillermo Lazcoz

Genética y Genómica de Enfermedades Raras y Complejas

[pablo.minguez@quironosalud.es](mailto:pablo.minguez@quironosalud.es)  
[guillermo.lazcoz@quironosalud.es](mailto:guillermo.lazcoz@quironosalud.es)

V REUNIÓN ANUAL DEL ÁREA DE GENÉTICA Y GENÓMICA DEL IIS-FJD 05 de OCTUBRE del 2023



### Unidad de Bioinformática IIS-FJD

Plataforma de apoyo al análisis y gestión de datos ómicos

Pablo Mínguez

[pablo.minguez@quironosalud.es](mailto:pablo.minguez@quironosalud.es)

V REUNIÓN ANUAL PLATORMAS DEL IIS-FJD 13 de septiembre del 2023



Sala de Prensa > Agenda > Taller online sobre compartición de datos genómicos

### Taller online sobre compartición de datos genómicos

From Thursday, January 27, 2022 to Thursday, January 27, 2022

El taller, organizado por IMPaCT-GENÓMICA y CIBERER, tiene el objetivo de introducir a la comunidad CIBERER, IMPaCT-Genómica, y otros entornos de investigación genómica, algunas tecnologías existentes para compartir datos genómicos. Por un lado se presentará la base de datos EGA (European Genome-Phenome Archive, <https://ega-archive.org>), un recurso europeo con base en Barcelona, que aloja secuencias genómicas de pacientes, acredita la compartición de estas y permite la petición/cesión de secuencias entre investigadores. Por otro lado se introducirán los sistemas Beacon, un proyecto originario de la GA4GH, que propone un sistema estándar para hacer descubribles los datos genómicos humanos sin comprometer información personal. Durante el taller se presentarán ambos recursos, se discutirá sobre su papel en los proyectos de Medicina Personalizada que producen secuencias genómicas (IMPACT-Genómica), se expondrán las diferentes estrategias disponibles para que estas tecnologías lleguen a los centros de producción de secuencias y permitan generar una red de compartición de datos colaborativa, que asegure el cumplimiento de aspectos legales y éticos y, que contribuya a poner el valor la secuenciación producida para el estudio de las enfermedades de pacientes.

Toda la información de la jornada la podéis encontrar en el siguiente [enlace](#).

Inscripciones [aquí](#).

<https://youtu.be/RI9wCs7JVKw>



Taller sobre compartición de datos genómicos: Bases de datos EGA y sistemas de descubrimiento

Centro de Investigación Cl...  
878 suscriptores

Susíbime

2

Compartir

176 visualizaciones 28 feb 2022

Ponentes: Jordi Rambla (Centre for Genomic Regulation, EGA Team), Teresa D'Altri (Centre for Genomic Regulation, EGA Team), Jose Luis Fernández Rueda (Fundación Progreso y Salud, U715-CIBERER, BiER), Javier Pérez Florido (Fundación Progreso y Salud, U715-CIBERER, BiER), Joaquín Dopazo (Fundación Progreso y Salud, U715-CIBERER, BiER).

CURSO: PRINCIPIOS DE INVESTIGACIÓN E INNOVACIÓN RESPONSABLE (RRI).

# Software propio en repositorios y registros públicos

<https://github.com/TBLabFJD>



Translational  
Bioinformatics Lab - IIS-  
FJD  
TBLabFJD

Pinned Order updated.

<b>NextVariantFJD</b> (Public) Variant calling pipeline implemented in Nextflow ● Perl	<b>RNAseq_scripts</b> (Public) ● R
<b>GLOWgenes</b> (Public) Comprehensive prioritization of gene diseases candidates by disease-aware evaluation of heterogeneous molecular networks ● Python ☆ 1	<b>VisoQLR</b> (Public) Interactive analyzer, viewer and editor for isoform quantification using long-reads sequencing data for single locus analysis ● R ☆ 3
<b>PriorR</b> (Public) PriorR ● R ☆ 2	<b>neuralnetwork</b> (Public) ● R



<https://bio.tools>



Everything: 'Bioinformatics Unit IIS-FJD' x Search bio.tools

Sort by Score Updated Added Name Citation Count Public

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here

**GLOWgenes** |

Prioritization of gene diseases candidates by disease-aware evaluation of heterogeneous evidence networks

Bioinformatics Genomics  
Command-line tool CC-BY-NC-SA-4.0 IMPaCT-Data

**Mini-IsoQLR** |

This pipeline was developed to detect and quantify isoforms from the expression of minigenes, whose cDNA was

Bioinformatics  
Genome alignment RNA-Seq quantification  
Command-line tool CC-BY-NC-SA-4.0 IMPaCT-Data

**PTMCode** |

PTMCode is a resource of known and predicted functional associations between protein post-translational mo which are also propagated through orthologs between 19 different eukaryotic species. A total of 1.6 million sites a

Bioinformatics  
Database portal CC-BY-NC-SA-3.0 IMPaCT-Data

<https://impact-data.bsc.es/outcomes/software>



IMPACT Data INICIO QUIÉNES SOMOS RESULTADOS COLABORACIONES NOTICIAS Y EVENTOS

## HERRAMIENTAS SOFTWARE

Inicio > Resultados > Herramientas software

Las herramientas software para el análisis de datos de IMPACT-Data se pueden encontrar en el dominio IMPACT-Data de acceso público en bio.tools. Bio.tools es un registro de componentes software y bases de datos dirigida a investigadores en el campo de las ciencias biológicas y biomédicas para facilitarles el trabajo de encontrar, entender, utilizar y citar recursos de uso diario.

**PriorR**

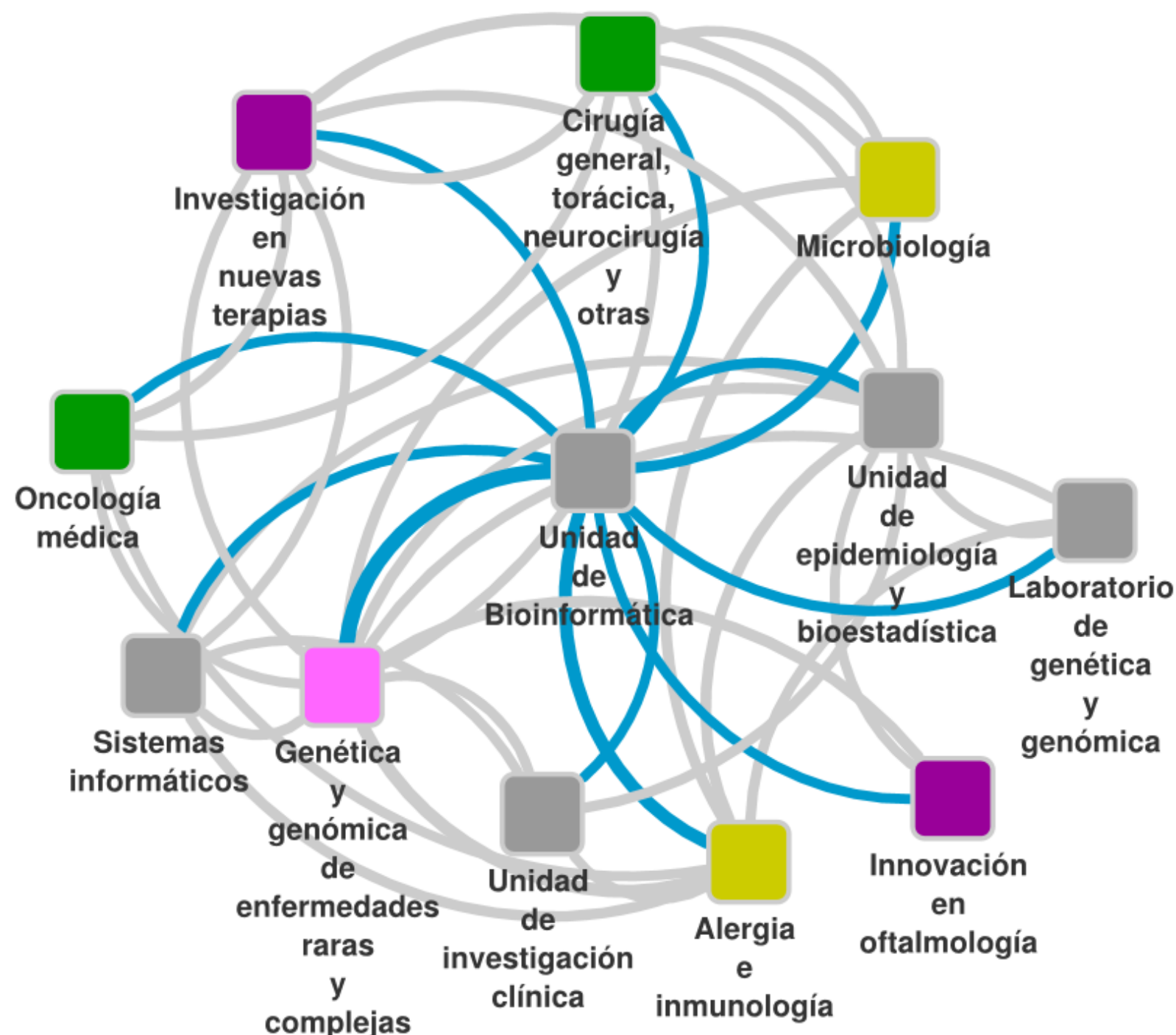
PriorR is a prioritization program of disease-linked genetic variants developed within the Genetics&Genomics Department of La Fundación Jiménez Díaz University Hospital. PriorR is conceived to analyse the output of the FJD-pipeline of SNVs or CNVs. This software program offers a number of useful functionalities for variant analysis such as: filtering by a virtual panel of genes, manual control of different population frequencies or pathogenicity predictors or filtering out variants that have been already found by another protocol.

**FJD-pipeline**

# Colaboraciones de la Unidad de Bioinformática 2022

## Publicaciones 2022 (memoria IIS-FJD)

Total	D1	Q1	Q2	Sin IF	IF medio	IF acumulado
13	1	9	4	0	9.94	90.156



- Genómica (Enfermedades raras, Genética).
- Genómica de virus SARS-CoV2 (CNB).
- Transcriptómica (Alergia, Oncología).
- COVID (STOP-Coronavirus, SCOURGE).
- Herramientas/algoritmos/bases de datos bioinformáticas (Genómica).

## 2023 y más

- Genómica (Enfermedades raras).
- Genómica de virus (SARS-CoV2) y COVID (STOP-Coronavirus, SCOURGE).
- CIBERER, CSVS.
- Transcriptómica (Patología vascular, Alergia).
- Epigenética (Pediatría).
- IA (Hematología).
- Herramientas (Genómica): GLOWgenes, VisoQLR.

# Indicadores del proyecto RRI22/00005

## A nivel de Ciencia:

- Número de publicaciones derivadas de proyectos financiados con fondos públicos en medios Open Access (Objetivo  $\geq 25\%$ ).
- % de los datos de investigación procedentes de los proyectos financiados con fondos públicos, disponibles en abierto en repositorios de datos abiertos normalizados y reconocidos en su disciplina (Objetivo  $\geq 50\%$ ).

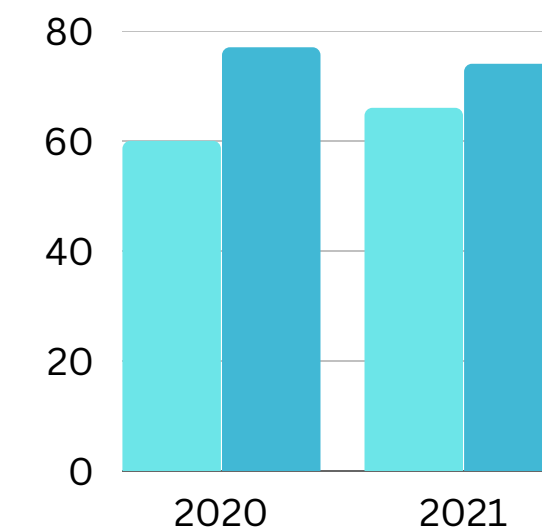
## A nivel de la Sociedad:

- Número y porcentaje de proyectos con participación de los actores sociales en el diseño y desarrollo (Objetivo  $\geq 10\%$  en los dos últimos años).
- Número de actividades de difusión hacia el público general en los últimos cinco años (Objetivo  $\geq 5$ ).

## A nivel del IIS-FJD:

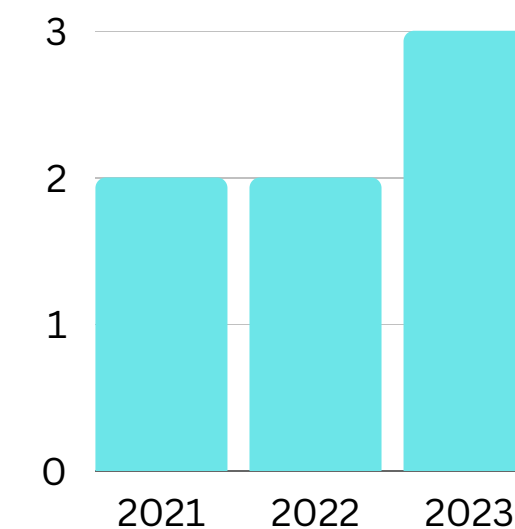
- Número de actividades formativas desarrolladas en materia de Ciencia Abierta (Objetivo  $\geq 1$ /año).
- Valoración media de las actividades formativas desarrolladas en materia de Ciencia Abierta.
- Número de asesoramientos en materia de Open Access proporcionados.
- Número de planes de datos elaborados/socitados (Objetivo = 100% solicitado/año).

## % Publicaciones

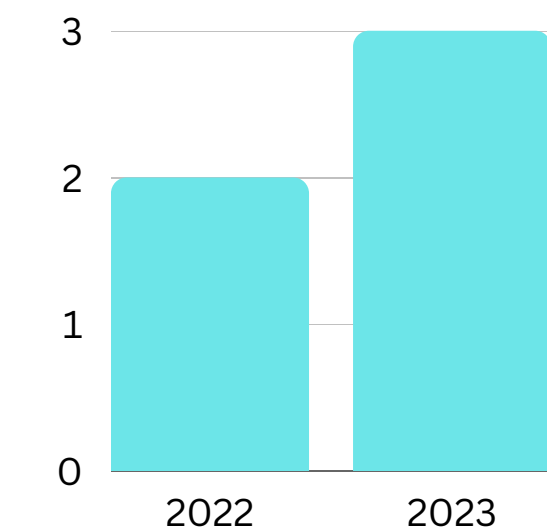


51% datos depositados en el año 2021/Proyectos financiación pública que generen datos

## # Actividades



## # Asesoramientos



# Conclusiones

**La Implementación de la Ciencia Abierta en el tratamiento de datos ómicos ha permitido:**

- Apoyar a los investigadores en el manejo de estos datos.
- Fomentar su utilización.
- Difundir sus resultados a la Sociedad y
- Cumplir objetivos con organismos financiadores, iniciativas y proyectos nacionales y europeos.

# GRACIAS



Contacto: [pablo.minguez@quironsalud.es](mailto:pablo.minguez@quironsalud.es)



Instituto de Salud Carlos III

IMPACT IMPACT

GENÓMICA



DATA



Unión Europea

Fondo Europeo de Desarrollo Regional  
"Una manera de hacer Europa"



ciber | ER

 D.G. de Investigación e Innovación  
C. DE EDUCACIÓN E INVESTIGACIÓN  
Comunidad de Madrid

 rare  
genomics