Análisis Genómico en Entornos Sanitarios

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
Instituto de Salud Carlos III

IMPaCT

Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología

Program	IMPacT: Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología		
Project Name	IMPaCT-Data: Programa de Ciencia de Datos de IMPaCT		
Expedient	IMP/00019		
Duration	January 2021 – December 2023		
Website	impact-data.bsc.es/		
Work Package	WP3 – Genomics		
Task	T3.2 Adaptación, instalación y uso de software de código abierto para el análisis e integración de distintas fuentes de datos		
Deliverable	E3.4 Análisis Genómico en Entornos Sanitarios		
Version	1.1.1		
Due Date	31/10/2022	***Approval Date***	17/05/2023
Responsible	CRG		
Dissemination Level	X	PU	Public
		CO-IMP	Confidential, only IMPaCT pillars members, including the evaluation commission from IMPaCT.
		CO-DATA	Confidential, only IMPaCT-Data members, including the evaluation commission from IMPaCT.

Authors		
Organization	**Name**	**Role**
Acronym	**Name and Surname**	**Coordination / Author / Reviewer**
BSC-CNS	Lidia López	Coordination
EGA-CRG	Jordi Rambla	Author
EGA-CRG	Amy Curwin	Author
EGA-CRG	Teresa D'Altri	Author
FPGMX-IDIS	Jorge Amigo	Reviewer
SNS-O	Javier Garricho	Reviewer

Versions History			
N.	**Date**	**Description**	**Author**
v 0	04/03/2022	Translated from Spanish version	L. López (BSC-CNS)
v 0.0	04/05/2022	Created	A. Curwin (EGA-CRG)
v 0.1	7/10/2022	Outline sent to reviewers	EGA-CRG
v 0.2	17/11/2022	Content added, sent to participating partners and coordination for review	EGA-CRG
v 0.3	29/11/2022	Sent to reviewers for formal review	A. Curwin (EGA-CRG)
v 0.4	6/12/2022	Minor revisions based on reviewer comments	A. Curwin (EGA-CRG)
v 1.0	13/12/2022	Final version accepted	A. Curwin (EGA-CRG)
v 1.1	17/05/2023	Visibility changed to public and approved	Comité Directivo
v 1.1.1	14/06/2023	Cambio de formato para publicar en la Web de IMPaCT	David Velasco (ISCIII)

Content

Tables

Executive Summary

Deliverable E3.4 “Genomic Analysis in Healthcare Environments" gathers information across partner institutes regarding what genomic processes are performed, what diseases are most studied, what pipelines are used and their portability. This document, produced through surveys, meetings, and discussion with relevant partner institutes, provides a picture of the overall landscape of these aspects within the IMPaCT-Data project. The aim of this deliverable is to understand the current situation in order to provide or suggest the necessary adaptations for integration of the different data sources within the federated ecosystem of IMPaCT-Data.

Introduction

Audience

Deliverable E3.4 “Genomic Analysis in Healthcare Environments” is envisioned and written for those institutes working with health-care data who would like to join a federation where the data is shared in accordance with the legal frame. This is the mission of IMPaCT-Data project that is set the basis for the successful implementation of the Spanish personalized medicine program.

Topic

This deliverable is related to task 3.2 of WP3 “Adaptation, installation and use of open-source software for the analysis and integration of different data sources”. As such, we have met, discussed with and surveyed various partners, including hospitals and clinical research centres, to understand the current landscape of genomic analysis in order to determine what adaptations might be necessary for the potential use of the data analysis system being developed in IMPaCT-Data.

Relation to other Deliverables

This is the third deliverable of WP3, although not directly related to previous deliverables in this WP. E3.4 is related to E5.1 developed by WP5, where they describe the technologies available for the integration of biomedical data.

Deliverable Structure

This deliverable is structured as follows:

Landscape of different data sources and analyses used among the partners.

- Here we describe the study methodology and results of 2 surveys carried out among the project partners. Sections include:
    - Types of processes
    - Diseases
    - Pipelines and general details of the processes

Evaluation of the landscape and proposed workflow.

- Here we elaborate an analysis of the results exposed above, providing some conclusions and proposed future steps.

1 Landscape of different data sources used among partners

1.1 Study methodology

The content of this deliverable is derived from the work of a working group within WP3 that is focussed on understanding the landscape of genomic analysis being performed across healthcare centres within IMPaCT-Data. The working group met regularly, learning different use-cases through presentations by the partners, as well as performing and analyzing 2 surveys. The second survey was a refined version of the first, produced in collaboration with all partners of WP3 and reviewed by WP leaders, ensuring alignment within the respective WPs, as well as with the IMPaCT-Genomica pillar of the consortium. Therefore, analysis of the second survey is the main content of this deliverable, but at times we draw on things learned in the first survey or from discussion within the working group meetings themselves. The surveys were provided in Spanish, however, we have provided English translations within the text of the deliverable. The surveys are included in their original formats in Annex A (2nd survey) and Annex B (1st survey). The target audience was any health care and/or research centre that routinely collects health genomic data for the purposes of diagnosis and/or research.

The 2nd, refined survey was organized in three main sections:

1. Types of processes: in-house versus commercial, volumes performed annually.
2. Diseases: what diseases are being studied by the different processes
3. Pipelines and general details: details of quality controls, types of files, pipelines used, and details of internally developed pipelines and their public availability/portability.

1.2 Results of the survey

The above described survey has been analysed in order to collect the useful information for proposing a workflow toward an integrated and federated ecosystem. 12 partners responded to the survey, thus their answers have produced the results exposed in this section. It is worth noting, the results of the survey were discussed within a working group of WP3. We are aware that, in some cases, respondents answered on behalf of the whole institute, while in others this information was not readily available. We provide here a summary of the survey results, realizing the amount of heterogeneity across institutes is very large. Further investigation will also be needed during the hand-on phase of the implementation.

1.2.1 Types of processes

We have asked the partners to select which sequencing processes they perform in their premises and in what volume. The list provided in the survey and complemented with users' answers is the following. 14 processes were given to select from, as well as “other” to respond in free text. We asked respondents to distinguish between in-house and commercial services, however after analysing and discussing the results with the respondents we realized these should be considered together. Reasons for using in-house versus commercial services varied including lack of resources, available technologies, volume of processes required, time required, etc. In some cases, all diagnostic processes are performed commercially, while techniques performed in-house were more for research purposes. There is a lot of diversity in this regard across the partners and even within a given institute. Therefore, it was decided it did not add anything to the big picture distinguishing between in-house versus commercially performed processes. Table 1 represents the total of in-house and commercial services being performed and used within the different institutes, individual labs or projects.

Table 1. Genomic analysis processes performed across partners

Process	Number / % of reported users	Number of processes reported yearly			
		<100	100-1000	1000-10000	10000
Exomes	**12 / 100%**	4 / 33%	6 / 50%	2 / 17%	–
Gene panels	**11 / 92%**	2 / 18%	6 / 55%	3 / 27%	–
Transcriptomics	**9 / 75%**	3 / 33%	6 / 67%	–	–
Sanger sequencing	**8 / 67%**	1 / 12%	3 / 38%	3 / 38%	1 / 12%
MLPA*	**8 / 67%**	3 / 38%	4 / 50%	1 / 12%	–
Whole Genome	**7 / 58%**	6 / 86%	1 / 14%	–	–
CNV microarrays	**6 / 50%**	1 / 17%	3 / 50%	2 / 33%	–
Genotype microarrays	**6 / 50%**	2 / 33%	3 / 50%	1 / 17%	–
Methylation microarrays	**5 / 42%**	3 / 60%	1 / 20%	1/ 20%	–
Single cell	**5 / 42%**	3 / 60%	1 / 20%	1 / 20%	–
ATAC-seq	**4 / 33%**	3 / 75%	1 / 25%	–	–
Metagenomes	**3 / 25%**	2 / 67%	1 / 33%	–	–
Microbiomes	**3 / 25%**	2 / 67%	1 / 33%	–	–
Expression microarrays	**2 / 17%**	1 / 50%	1 / 50%	–	–
Other processes	**5 / 42%**				

** MLPA (Multiplex Ligation-dependent Probe Amplification)*

Among “Others” the following free text responses were given: Detección de repeticiones expandidas (X-Frágil, SCAs), Detección de inversión F8, CHIP-Seq, Targeted deepSeq,

Paneles metilación, Mapeo óptico del genoma (Bionano), High throughput qPCR (Biomark HD)

The summary above shows that every partner is already doing exomes and virtually everyone is doing gene panels. The number of applied technologies is wider than expected showing that some techniques are being common among the IMPaCT-Data partners.

1.2.2 Diseases

The survey asked which diseases were being analysed by the given processes, among the following 9 choices: Oncology, rare diseases, cardiovascular diseases, neurological diseases, autoimmune diseases, infectious diseases (host), infectious diseases (pathogen), pharmacogenetics and prenatal diagnosis. In “others” we received the following free text responses: hereditary renal, metabolic and ophthalmological disease types.

Table 2. Most used processes with respect to most studied diseases

Process vs Disease studied	Oncology	Rare diseases	Cardio-vascular	Neurological	Auto-immune	Other
Exomes	2	9	4	4	2	3 Prenatal diagnostic
Gene panels	9	6	4	2	2	1 Prenatal diagnostic
Transcript-omics	4	5	2	1	2	2 Infectious (host)
Sanger sequencing	6	7	3	4	2	1 Pharmaco-genetics 1 Prenatal diagnostic
MLPA	2	7	3	2	1	1 Pharmaco-genetics 1 Prenatal diagnostic
Whole genomes	1	4	0	1	0	1 Infectious (pathogen)
Microarray CNVs	2	4	1	1	1	4 Prenatal diagnostic
Microarray genotype	3	2	2	1	1	1 Pharmaco-genetics
Single cell	2	0	1	1	1	NA

Includes processes used by at least 5 of 12 respondents and most commonly studies diseases. Number of respondents is shown.The "Other" column points out the next most studied disease type(s) for the given process.

Generally, the partner institutions have a focus on a specific disease field (group of diseases). Some have a narrow focus, while others span a broader spectrum of expertise. We observe 3 respondents working exclusively with either oncology (2) or cardiovascular disease (1). The other 9 display a variety of at least three pathology types in their study range.

Oncology and rare diseases fields are studied by 9 of the institutes, cardiovascular diseases by 5. Others, like Neurological, autoimmune and infectious diseases, among others, are studied only by a subset of institutes. 2 institutes stand out as working on 6-9 different disease types, while most focussed on 1-3 disease types only.

The following are all the numbers of institutes out of 12 studying a giving disease type, regardless of process: oncology (9), rare diseases (9), cardiovascular diseases (5), neurological diseases (5), autoimmune diseases (2), infectious diseases (host) (3), infectious diseases (pathogen) (2), pharmacogenetics (2) and prenatal diagnosis (4). From "others": hereditary renal diseases (2), metabolic diseases (1) and ophthalmological diseases (1).

1.2.3 Pipelines and general details of the processes

Standard pipelines exist for the analysis of the above mentioned sequencing technologies. Researchers can use them directly, or modify them to make them more suitable for their specific needs. Commercial pipelines are also used for most of the sequencing data, but the customer has no option to customize to their needs. More equipped teams can develop their own *ad hoc* pipelines.

Among the 12 surveyed entities there is a variability in terms of the types of pipelines in use. One of the entities declared to only use standard pipelines for the developed sequencings. Another only uses commercial, while a third only uses in-house developed pipelines. All the other 9 entities use a combination of the described possibilities depending on the sequencing data to be run and the specific case. 9 have developed their own pipelines while 6 entities have modified standard ones for several data types.

Notably, little variability is observed among the different data types in this regard: all of them are approached with a mix of the above described options; none is analysed with the same type of pipeline from all the surveyed performing institutes.

All surveyed produce and use FASTQ and VCF files, and almost all (11 out of 12) use BAM files as well. Only a minority uses CRAM or gVCF files (3 each).

Consistently, all the surveyed entities perform quality controls on the raw sequencing data and the final Variant Calling they produce, and 9 out of 12 do it as well on the alignment. As a reference genome, 5 institutes use exclusively GRCh37 (hg19). One institute uses exclusively GRCh38 (hg38), and 6 use both. A relative homogeneity on the file formats and quality control used by the partners is going to be instrumental for the workflow toward an integrated ecosystem.

All respondents perform variant calling annotation (3 with an external tool, 7 with various tools from which a consensus is made and 2 respond “other”.

We asked about storage and access to the files in a series of questions in the first survey (Annex B). All responded they have access to the files (eg. FASTQ, BAMs, VCF etc.) during the analysis, either through an internal server (9 of 17) or when specifically requested (7 of 17, 1 did not respond). When asked where they store the files in the medium/long-term they were given the following options:

- They are not saved, they are destroyed shortly after the end of the analysis
- On backup tapes or removable/portable hard drives or similar, available upon request or manual upload,
- On hard drives, external storage, or cloud accessible at any time, without any necessary intervention

1 selected all 3, 1 selected options 2 and 3, 4 selected only the second option and 8 selected only the 3rd option (1 did not respond). Therefore, most respondents keep the data for the medium/long-term in some form.

With respect to how long they store the file, respondents could answer in free text. Reviewing the responses (12 of 17 answered), essentially 10 answered indefinitely, with conditions in some cases (eg. every 2 years moved to an external unit, or depending on the type of data; diagnostic or research, etc), 1 answered 10 years, 1 answered not yet determined.

With respect to management of the pipelines, we asked if those developed in-house were publicly available in a repository and if they were containerized for easy deployment (this only applied to 11 of the 12 institutes). In the case of the former, 6 responded no, they are not available in a public repository, 3 responded yes, 1 responded yes and no, and 1 did not know. Of the 3 that responded yes, 2 have them containerized and one not. Apart from this, the

responses to whether the pipelines were containerized were the same as if they were publicly available (6 no, 1 yes and no, and 1 did not know).

We asked whether those who are using their own pipelines (10 of 12) are using a workflow manager such as Nextflow, Galaxy or Snakemake or whether they only use a script (bash, python, etc.). 5 answered they only use a script, 2 indicated the same in addition to Snakemake or Snakemake, Galaxy and “others”. 1 answered both Nextflow and Galaxy, and 2 answered only “others”. Those that answered “others” wrote WDL/cromwell, their own design and internal queue manager. 1 answered “I don’t know”. Also of note, based on discussions we learned most of the partners run their pipelines in HPC (high performance cluster) environments instead of cloud environments. Furthermore, we learned a sharing scheme for pipelines would be highly valued among the partners and we discussed which pipelines would be of most interest to share first. This is discussed further in the conclusions below.

2 Evaluation of the landscape and proposed workflow

The results described above, produced from surveys and discussion among the partners, portrays a two-sided landscape: 1) a great variability among the partners in terms of types of processes and diseases covered but 2) a homogeneity in how they approach the bioinformatics analyses. Being conscious that the 12 respondents are likely not completely representative of the whole group of involved partners, the amount of different settings will hardly disappear if we increase the number of partners interviewed.

Wide ranging scenario has nonetheless different outcomes. Variability in performed technologies and studied diseases provides depth and enriches the overall collection of data, adding value at the point in which the data will be sharable in a federated network. Dissimilarities in the used pipelines could represent an obstacle towards the sharing and reuse of the whole set of scripts and containers, as they would not be of interest for other partners. In some cases, the in-house developed pipelines were containerized and publicly available, but this represented a minority of cases. The used file formats are quite homogeneous among the respondents, and this is a favourable point that represents an explicit opportunity in leveraging common steps like quality control.

3 Conclusions

The survey and discussions indicated that virtually all institutes are using pipelines themselves, many developed in-house or modified from standard versions. Not all have their pipelines containerized and publicly available, but many say they would benefit from the work of others and most would be willing to share knowledge. We discussed with the partners what they would most like to get from such a sharing scheme and a few relevant topics came out from this. For example, pipelines for RNAseq, CNVs and quality control.

Therefore, we would suggest the next steps in this regard, and within the scope of the IMPaCT-Data project, would be to select 2 or 3 existing pipelines currently in production and do a proof of concept of porting (adapting and moving) these pipelines with interested partners.

Acronyms and Abbreviators

In the following table there are some acronyms and abbreviators used in the deliverable

IMPaCT	Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología (Spanish initials)
IMPaCT-Data	Data program for IMPaCT
WP	Work Package
MLPA	Multiplex Ligation-dependent Probe Amplification
CNVs	Copy number variants

Annex A

In this annex we include the complete list of questions from the 2nd, refined survey.

Preguntas para GdT03-1: Análisis genómico

Sección 1: Tipos de procesos

1) Tu institución/proyecto/consorcio institución realiza los siguientes tipos de procesos INTERNAMENTE (en una pregunta posterior se realiza la misma pregunta para servicios comerciales contratados externamente), tanto para investigación como para asistencia médica. Por favor, incluye una indicación aproximada del número que se realizan al año:

Proceso	N/A (cero)	<100	100-1000	1000-10000	>10000
Paneles de genes					
Exomas					
Transcriptomas					
Microarray de genotipado					
Microarray de expresión					
Microarrays de CNVs					
Microarrays de metilación					
Genomas completos					
Metagenómica					
Microbiomas					
Single cell					

MLPA					
Sanger sequencing					
ATAC-Seq					
Otros procesos					

1b) En el caso de responder “Otros procesos”, especifica por favor: (free text)

2) Tu institución/proyecto/consorcio institución realiza los siguientes tipos de procesos COMERCIALES (con anterioridad se ha realizado la misma pregunta para procesos realizados internamente en su organización), tanto para investigación como para asistencia médica. Por favor, incluye una indicación aproximada del número que se realizan al año:

Proceso	N/A (cero)	<100	100-1000	1000-10000	>10000
Paneles de genes					
Exomas					
Transcriptomas					
Microarray de genotipado					
Microarray de expresión					
Microarrays de CNVs					
Microarrays de metilación					
Genomas completos					
Metagenómica					
Microbiomas					

Single cell					
MLPA					
Sanger sequencing					
ATAC-Seq					
Otros procesos					

2) ¿A qué tipos de enfermedad/situación se aplica cada tipo de proceso?

Proceso	N/A	Oncología	Enfermedades raras	E. cardio-vasculares	E. neurológicas	E. autoinmunes	E. infecciosas (huésped)	E. infecciosas (patógeno)	Farmacogenética	Diagnóstico Prenatal	Otros tipos
Paneles de genes											
Exomas											
Transcriptomas											
Microarray de genotipado											
Microarray de expresión											
Microarrays de CNVs											
Microarrays de											

metilación											
Genomas completos											
Metagenómica											
Microbiomas											
Single cell											
MLPA											
Sanger sequencing											
ATAC-Seq											
Otros procesos											

2b) En el caso de responder “Otros procesos”, especifica por favor: (free text)

2c) En el caso de responder “Otros tipos”, especifica por favor: (free text)

3) Tipos de pipelines utilizados. Es indistinto si la pipeline está en producción, en fase de pruebas o en desarrollo (escoger todo lo que aplique)

Proceso	N/A (cero)	Estándar	Estándar modificada	Desarrolladas internamente	Comerciales

Paneles de genes					
Exomas					
Transcriptomas					
Microarray de genotipado					
Microarray de expresión					
Microarrays de CNVs					
Microarrays de metilación					
Genomas completos					
Metagenómica					
Microbiomas					
Single cell					
Otros pipelines					

3b) En el caso de responder “Otras pipelines”, específica por favor: (free text)

Sección 2: Detalles generales de los procesos

4) Puntos en los que se realiza un control de calidad (escoger todo lo que aplique)

1. Microarray
2. Resultado de la secuenciación (FASTQ)
3. Alineamiento (BAM/CRAM)
4. Variant calling (VCF)
5. Otros (especificar)

5) Tipos de ficheros generados (escoger todo lo que aplique)
1. FASTQ
2. BAM
3. CRAM
4. VCF
5. gVCF
6. Otros (especificar)

6) ¿Qué genomas de referencia se utilizan? (escoger todo lo que aplique)
1. GRCh37 (hg19)
2. GRCh38 (hg38)
3. Otros (especificar)

7) ¿Se anotan las variantes encontradas? (escoger una opción)
1. Sí, con una herramienta externa
2. Sí, con varias herramientas de las que se hace un consenso
3. No, se reciben anotadas
4. Otros (especificar)

Sección 3: Gestión de las pipelines

8) Las pipelines propias ¿están disponibles en algún repositorio público? (escoger todo lo que aplique)
1. Sí
2. No
3. No lo sé

9) Las pipelines propias ¿están containerizadas para hacerlas fáciles de desplegar? (escoger todo lo que aplique)
1. Sí
2. No
3. No lo sé

10) Las pipelines propias ¿utilizan algún gestor de workflows? (escoger todo lo que aplique)

1. Nextflow
2. Galaxy
3. Snakemake
4. No, sólo un script (bash, python…)
5. Sí, otros (especificar)
6. No lo sé

Annex B

In this annex we include the complete list of questions from the 1st survey.

Preguntas para GdT03-1

Flujo desde el lab al almacenamiento genómico

1) Tu institución/proyecto/consorcio institución recoge los siguientes tipos de muestras… (escoger todo lo que aplique)

1. Biopsias de tejido fresco
2. Biopsias de tejido parafinado
3. Plasma (DNA libre circulante)
4. Sangre completa
5. Cultivos celulares
6. Otros (especificar)

2) Tu institución/proyecto/consorcio secuencia… (escoger todo lo que aplique)

1. Paneles targeted - variantes germinales
2. Paneles targeted - variantes somáticas
3. Exomas clínicos
4. Exomas completos
5. Genomas completos
6. Transcriptomas completos
7. Microarrays de genotipado
8. Microarrays para la detección de CNVs
9. Otros (especificar)

3) Están integrados circuitos de identificación de cada una de las muestras:

1. Si
2. No

4) ¿Quién realiza las librerías?

1. Laboratorios de mi institución
2. Laboratorios externos (públicos) del servicio de salud/CCAA

3. Laboratorios externos privados nacionales
4. Laboratorios externos privados fuera de España

5) ¿Quién realiza la secuenciación? (escoger todo lo que aplique)
1. Laboratorios de mi institución
2. Laboratorios externos (públicos) del servicio de salud/CCAA
3. Laboratorios externos privados nacionales
4. Laboratorios externos privados fuera de España

6) ¿Quién analiza el resultado directo de la secuenciación (e.g. los FASTQ)? (escoger todo lo que aplique)
1. Mi unidad u otra unidad de mi institución
2. Una unidad especializada/centralizada del servicio de salud
3. Laboratorios externos privados nacionales
4. Laboratorios externos privados fuera de España

7) En caso que el análisis lo realice un tercero, el resultado directo del análisis (e.g. el variant calling) … (escoger todo lo que aplique)
1. No se recibe, ni está accesible
2. No se recibe, pero está accesible vía “web”, “FTP”, carpeta compartida…
3. Se recibe en forma de VCF
4. Se recibe en un formato distinto al VCF
5. (No aplica en mi caso)

8) En caso que el análisis lo realice un tercero, el control de calidad se realiza:
1. En mi institución
2. No se realiza y se reporta la información recibida por un tercero
3. (No aplica en mi caso)

9) La información facilitada a los genetistas moleculares, se entrega:
1. Íntegra, sin filtrar ni priorizar
2. Se filtra a veces, o facilita el filtrado, y se prioriza un subconjunto de variantes
3. Siempre se filtra y se proporciona sólo un subconjunto de variantes

10) Para gestionar los hallazgos incidentales, el contenido del consentimiento informado es conocido en cada muestra analizada:

1. SI
2. No
3. A veces

11) ¿Se realiza confirmación de los resultados de NGS con otras técnicas moleculares?

1. Si, siempre
2. A veces, dependiendo del contexto
3. Nunca

12) ¿Qué se guarda en la HCE? (escoger todo lo que aplique)

1. Sólo las notas que el clínico considera relevantes
2. El informe (PDF) del laboratorio genético (a modo de adjunto)
3. La información del informe molecular en campos estructurados (i.e. Tipo de muestra, gen, tipo de herencia, recurrencia, tabla de calidad etc…)
4. Un adjunto con las variantes en algún formato digital (VCF, Excel…)
5. Ninguno de los anteriores

13) ¿Se tiene acceso a los archivos generados durante el análisis (e,g, FASTQ, BAMs, VCF…)?

1. No, nunca
2. Sí, a petición expresa
3. Sí, disponibles en un servidor del laboratorio externo
4. Sí, disponibles en un servidor del laboratorio interno

14) ¿Dónde se guardan esos archivos a medio/largo plazo? (escoger todo lo que aplique)

1. No se guardan, se destruyen poco después de acabar el análisis
2. En cintas de backup o discos duros extraíbles/portátiles o similares, disponibles bajo petición o carga manual
3. En discos duros, almacenamiento externo, o cloud accesibles en cualquier momento, sin que ninguna intervención sea necesaria

15) ¿Por cuánto tiempo se guardan estos archivos? (texto libre)

16) ¿Quién realiza la custodia de la información genómica?

1. El servicio de Informática de la institución
2. El servicio/laboratorio de Bioinformática
3. Otros