DECISION

DEcompensated CIrrhoSIs: identification of new cOmbiNatorial therapies based on systems approaches

H2020 – 847949

D1.8 Deposit of Transcriptomic data to European Genome-phenome Archive (EMBL)

WP Leader:	Joan Clària (01 FCRB)
Authors:	Cristina Sanchez (01 EFCLIF) Ferran Aguilar (01 EFCLIF) Pierre de la Grange (Genosplice)
Version	V1
Due date of deliverable	30/09/2022
Actual submission date	31/01/2023
Dissemination level	PU

Abbreviations

ENA	European Nucleotide Archive
DECISION	Decompensated cirrhosis: Identification of new combinatorial therapies based on systems approaches
DMC	Data Management Centre
DMP	Data Management Plan
ID	Identifier
SC	Steering Committee
WP	Work Package
EGA	European Genome-phenome Archive

Contributors

Contributor	Contribution	Estimate of person-months
FCRB, coordination	• *Prepared and processed data from INSERM.* • *Standardized nomenclature to identify the patients.* • *Preparation of databases with phenotypes for the analyses.* • *Curation and submission of the data to ENA.*	*0.25 PM*
EFCLIF	• *Writing of the deliverable report*	*0.25 PM*
Genosplice and EFCLIF	• *Prepared and processed data from INSERM.* • *Standardized nomenclature to identify the patients.* • *Preparation of databases with phenotypes for the analyses.* • *Curation and submission of the data to ENA.*	*0.50 PM*
APHP	• *Deliverable review, correspondence and involvement in discussions*	*0.25 PM*

Executive Summary

The tasks included in the Deliverable have been carried out by the INSERM in collaboration with WP1 leader and other partners providing multi-omics data.

Transcriptomic data of 1.261 individuals from Canonic, Predict and Aclara cohorts have been submitted in the ENA's sub platform (Figure 1). The submission was performed by uploading a spreadsheet template available in the platform.

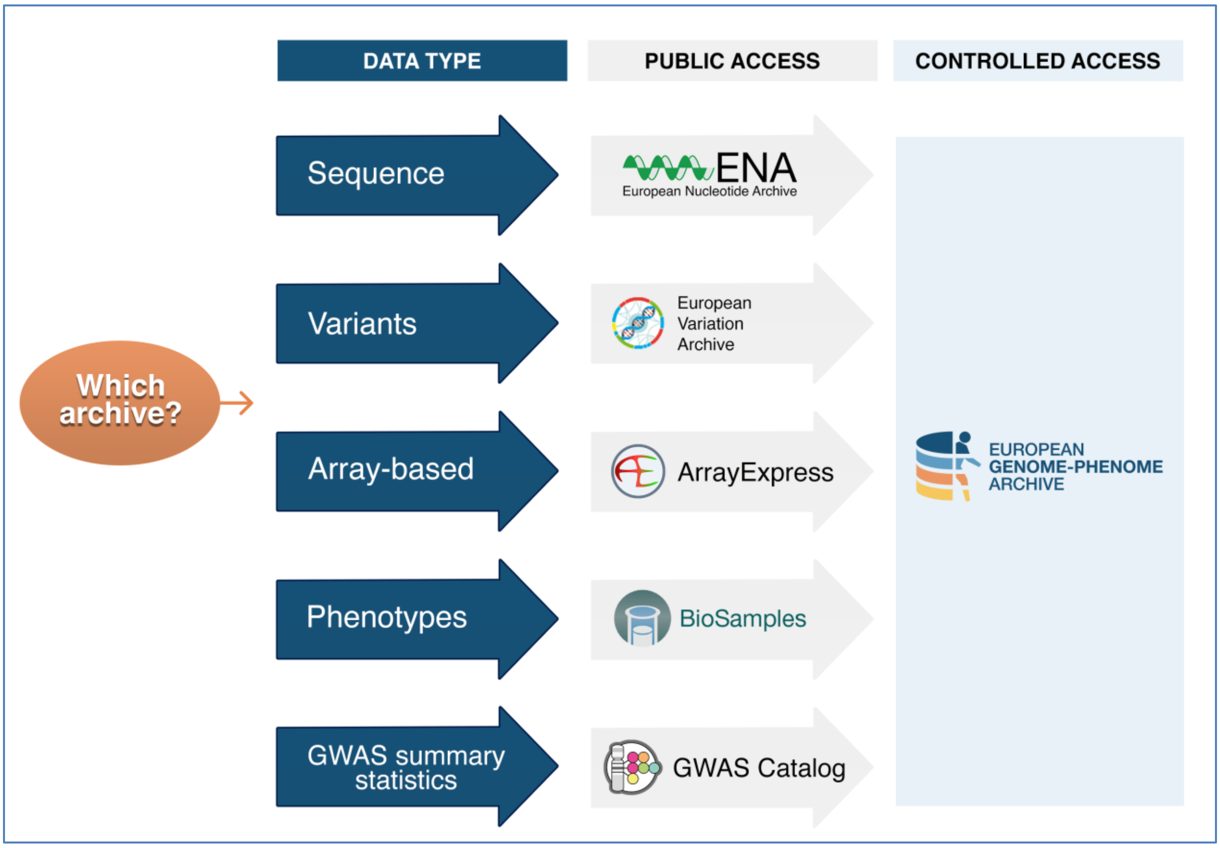


Figure 1. Structure of EGA Sub platform

Table of Contents

1 PREPARING THE DATABASE

1.1 Getting ready the data for the submission to ENA

Transcriptomics data of the DECISION project has been obtained by extracting and sequencing whole blood tempus tube samples.

Sequence of the samples was performed using Illumina NovaSeq 6000 system and the library NuGEN Universal Plus mRNA-seq kit woGlobin via RNA fragmentation, first and second strand synthesis, end repair, adaptor ligation, strand selection and library amplification.

The EGA archive is a controlled access database that allows the submission of many types of data (genomics, transcriptomics etc.)

Data types accepted by EGA can be split into 3 types as shown in Figure 2.

- Sequence data
- Array-based data
- Phenotype data

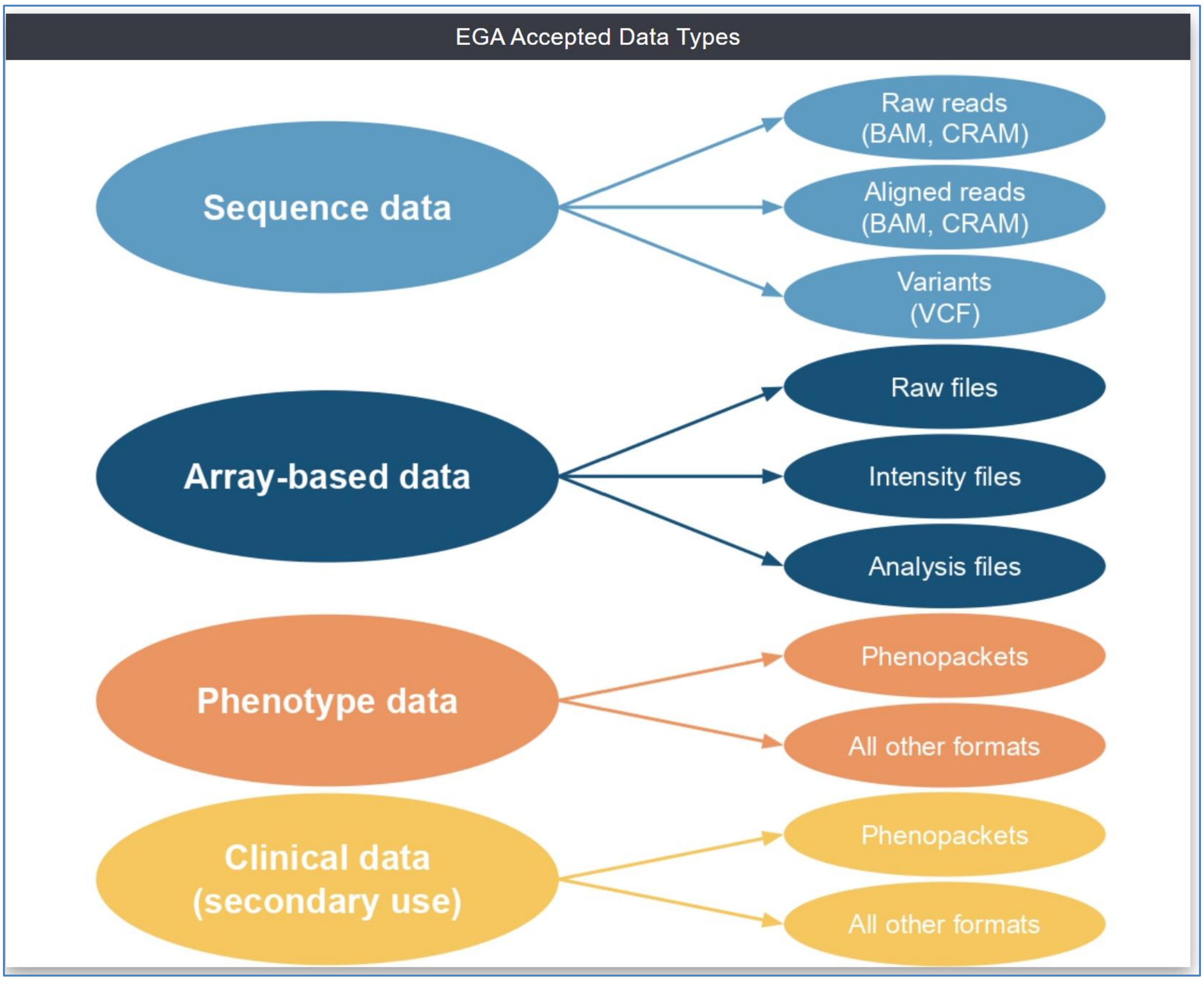


Figure 2. Data types accepted by EGA

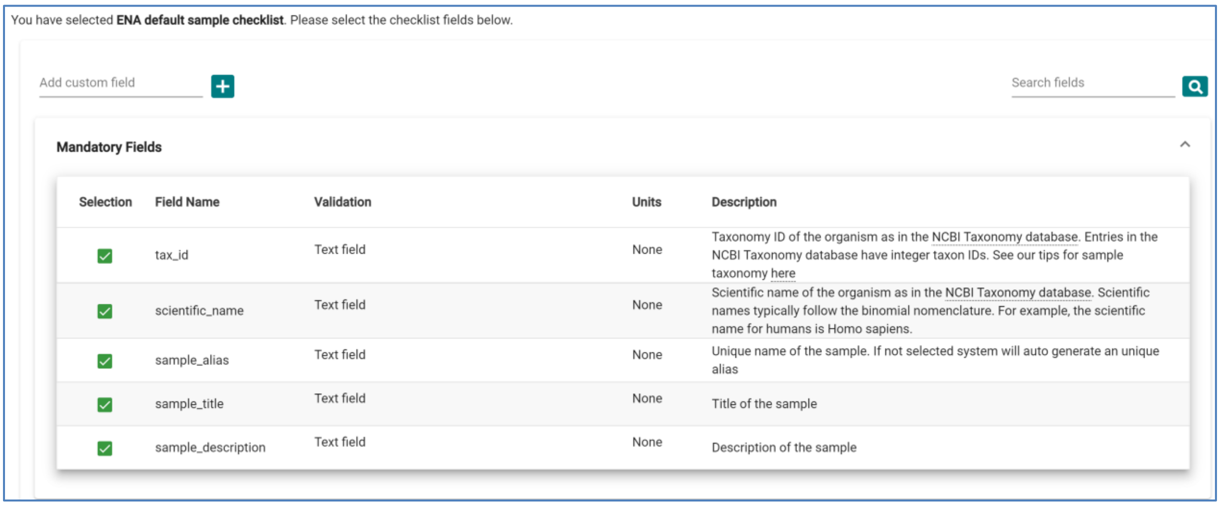
1.2 Submitting data to EGA's platform

The sequence reads were submitted uploading the CRAM files and an md5 code was generated and assigned to each sample.

Submission of information about the sequence data was made through 2 spreadsheet templates available in the platform: one for the samples and the second one for the reads corresponding to each sample.

To perform the submission, detailed information about the samples and the corresponding raw data files was uploaded to the website and validated by the Institution (Figures 3 and 4)

- Tax ID
- Scientific Name
- Sample Alias
- Sample title
- Sample Description

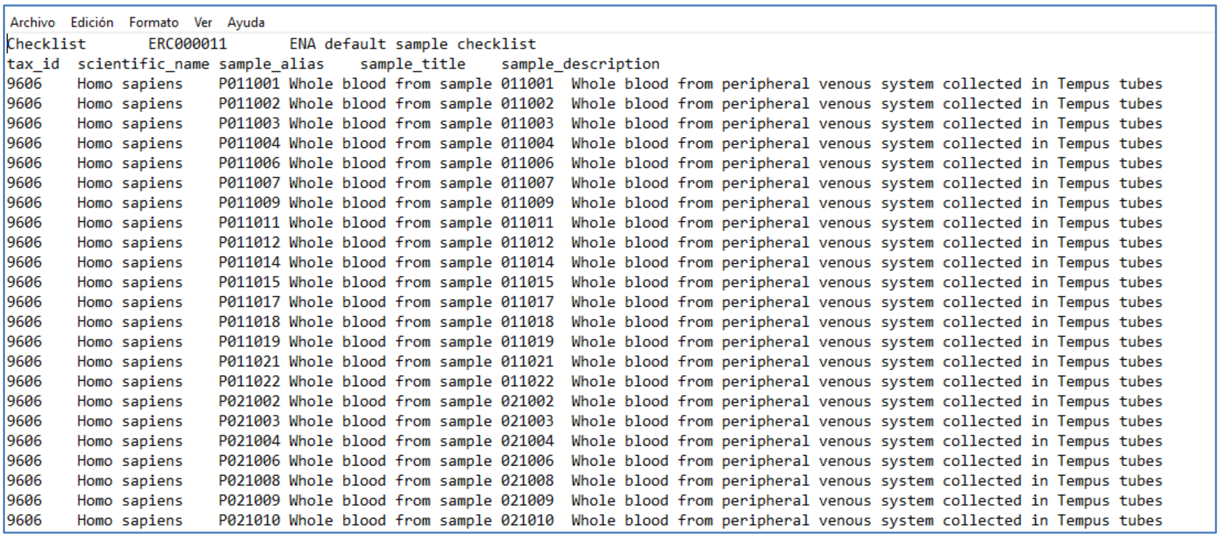


You have selected **ENA default sample checklist**. Please select the checklist fields below.

Mandatory Fields

Selection	Field Name	Validation	Units	Description
☑	tax_id	Text field	None	Taxonomy ID of the organism as in the NCBI Taxonomy database. Entries in the NCBI Taxonomy database have integer taxon IDs. See our tips for sample taxonomy here
☑	scientific_name	Text field	None	Scientific name of the organism as in the NCBI Taxonomy database. Scientific names typically follow the binomial nomenclature. For example, the scientific name for humans is Homo sapiens.
☑	sample_alias	Text field	None	Unique name of the sample. If not selected system will auto generate an unique alias
☑	sample_title	Text field	None	Title of the sample
☑	sample_description	Text field	None	Description of the sample

Figure 3. Required information about the samples



```
Checklist       ERC000011       ENA default sample checklist
tax_id  scientific_name sample_alias    sample_title    sample_description
9606    Homo sapiens    P011001 Whole blood from sample 011001  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011002 Whole blood from sample 011002  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011003 Whole blood from sample 011003  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011004 Whole blood from sample 011004  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011006 Whole blood from sample 011006  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011007 Whole blood from sample 011007  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011009 Whole blood from sample 011009  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011011 Whole blood from sample 011011  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011012 Whole blood from sample 011012  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011014 Whole blood from sample 011014  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011015 Whole blood from sample 011015  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011017 Whole blood from sample 011017  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011018 Whole blood from sample 011018  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011019 Whole blood from sample 011019  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011021 Whole blood from sample 011021  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P011022 Whole blood from sample 011022  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021002 Whole blood from sample 021002  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021003 Whole blood from sample 021003  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021004 Whole blood from sample 021004  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021006 Whole blood from sample 021006  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021008 Whole blood from sample 021008  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021009 Whole blood from sample 021009  Whole blood from peripheral venous system collected in Tempus tubes
9606    Homo sapiens    P021010 Whole blood from sample 021010  Whole blood from peripheral venous system collected in Tempus tubes
```

Figure 4. Detailed information of raw data

After all this information was provided, the samples were uploaded to the platform and linked to each ID.

Once the data had been submitted, the data could be accessed through an accession number. The BioProject accession (primary) is PRJEB56149 and the study accession (secondary) is ERP141065.

The date of public release of the data was set to 20/09/2024 and it can be modified if needed according to the dissemination rules of DECISION project.

2 Acknowledgement and Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847949.