DECISION

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

H2020 – 847949

D1.7 Deposit of Genotyping data to European Genome-phenome Archive (EMBL)

WP Leader:	Joan Clària (01 FCRB)
Authors:	Cristina Sanchez (01 EFCLIF)
	Ferran Aguilar (01 EFCLIF)

Version	V1
Due date of deliverable	30/09/2022
Actual submission date	31/01/2023
Dissemination level	PU

Abbreviations

EGA	European Genome-phenome 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
GSA	Global Screening Array

Contributors

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

Executive Summary

The tasks included in the Deliverable have been carried out by the EFCLIF DMC in collaboration with WP1 leader and other partners providing multi-omics data.
Genomics data of 1.841 individuals from Canonic, Predict and Aclara cohorts have been submitted in the EGA's sub platform Array Express for genomic data (Figure 1). This sub platform uses the Annotare tool (a user-friendly interface) to upload the data without informatic expertise.

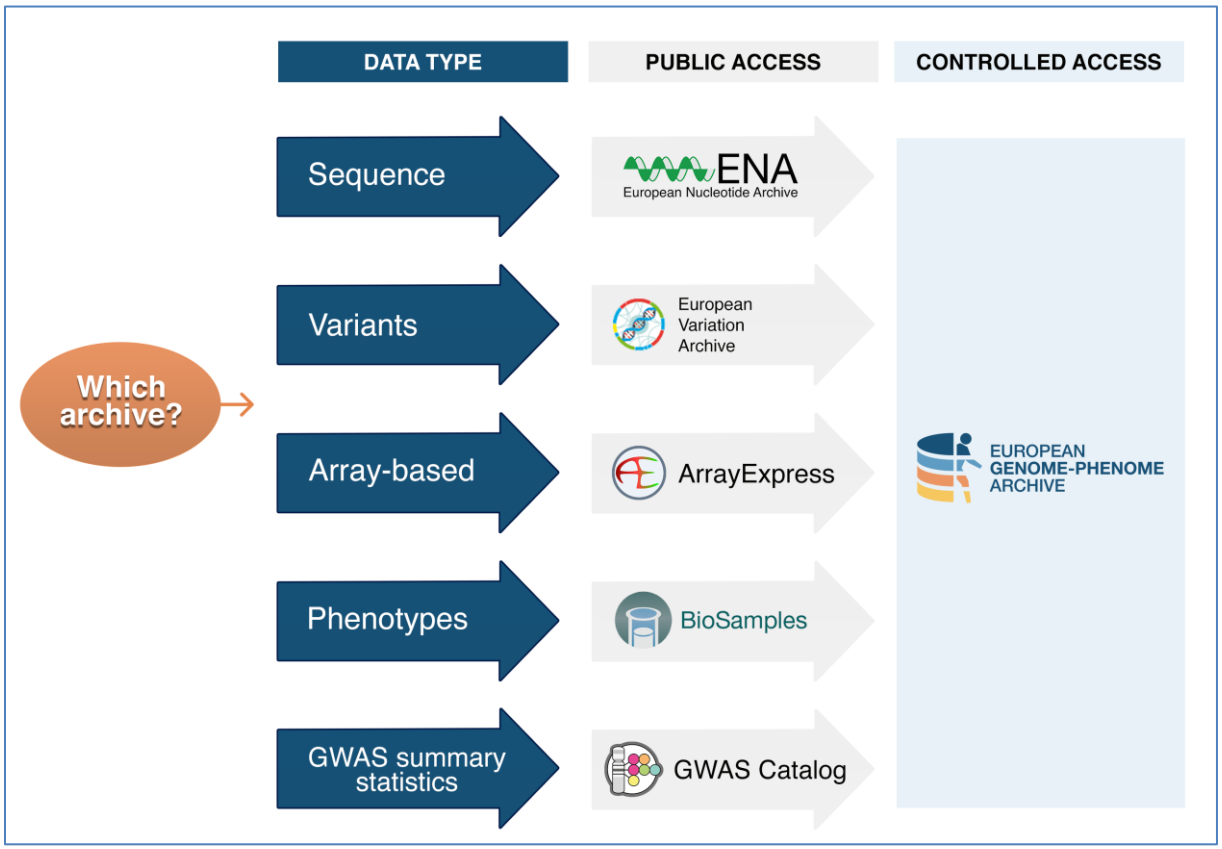

Figure 1. Structure of EGA Sub platform

Table of Contents

1 PREPARING THE DATABASE

1.1 Getting ready the data for the submission to EGA

Genomics data of the DECISION project has been obtained using GSA from EDTA Buffy Coat plasma samples.

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

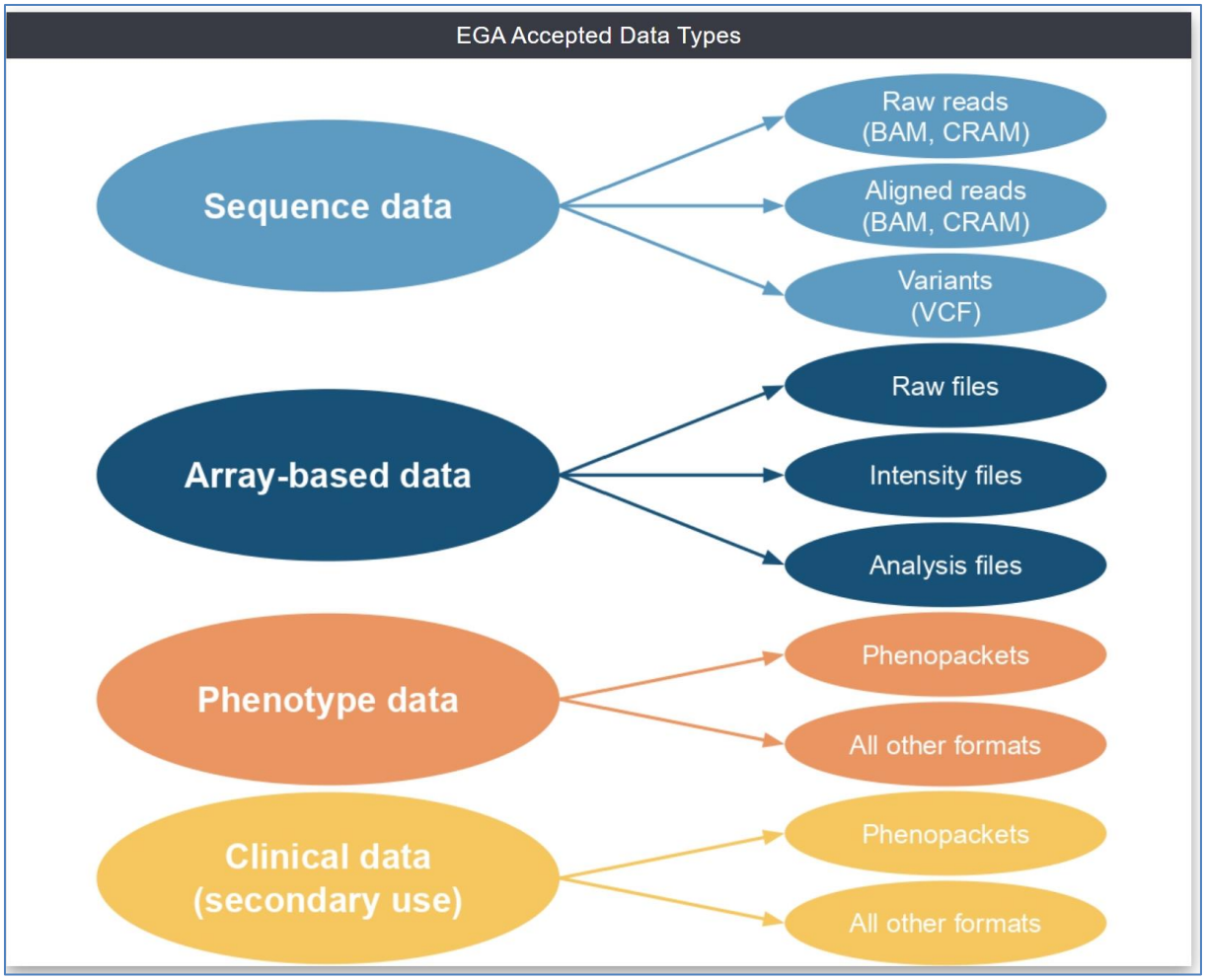

Figure 2. Data types accepted by EGA

1.2 Submitting data to EGA's platform

Submission of Array data is made through ArrayExpress, using the Annotare Platform.

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

- Description of the project
- Experiment type
- Date of the experiment
- Number of samples
- ID of the samples
- Organism
- Age
- Sex
- Disease

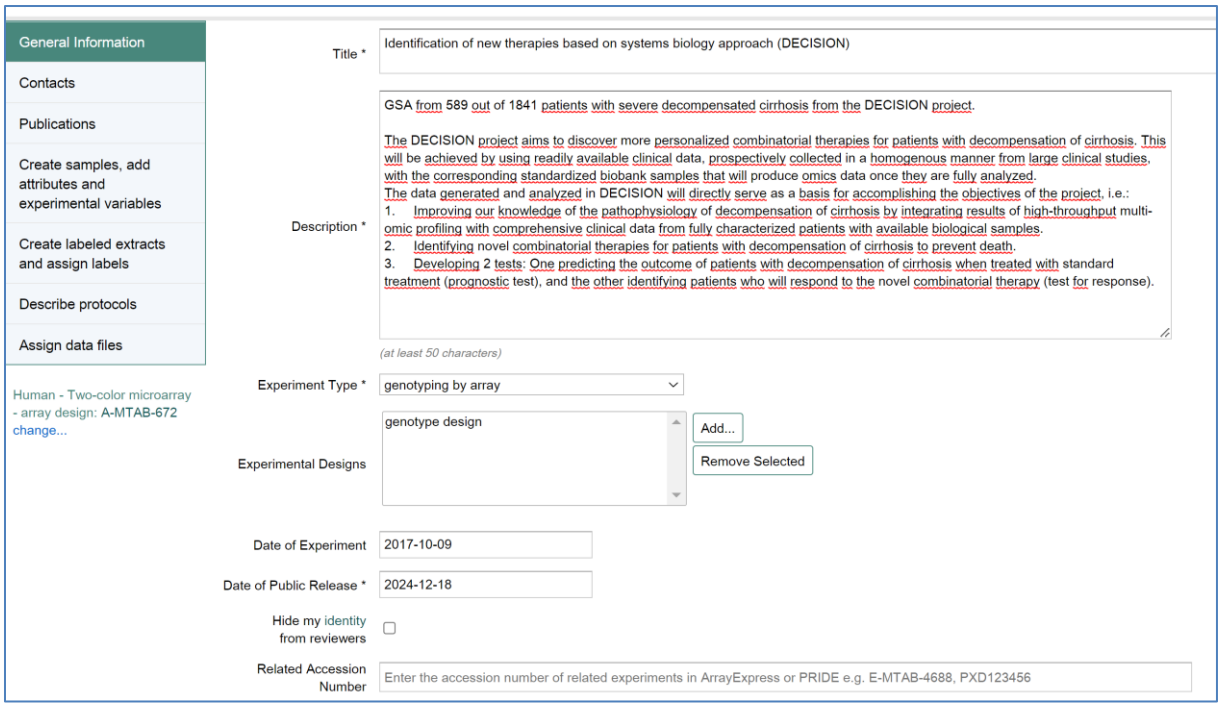

Figure 3. General information of the project and the submission

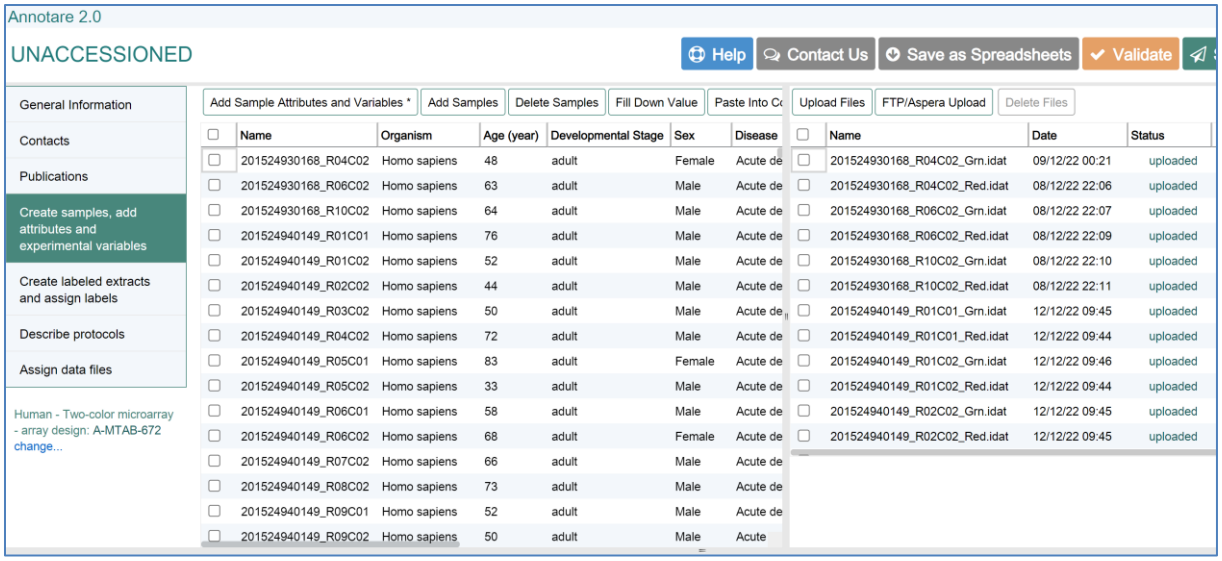

Figure 4. Detailed information of each ID submitted

After all this information was provided, the samples were uploaded to the platform and linked to each ID. Data produced by the DECISION project was obtained using a two-colour microarray. This means that, for each patient and sample, 2 raw files were provided: 1 for each dye which can be green or red.

Currently, the Annotare platform only supports 2:2:1:1 relationship between samples, labelled extracts, hybridisation assays and data files (Figure 5).

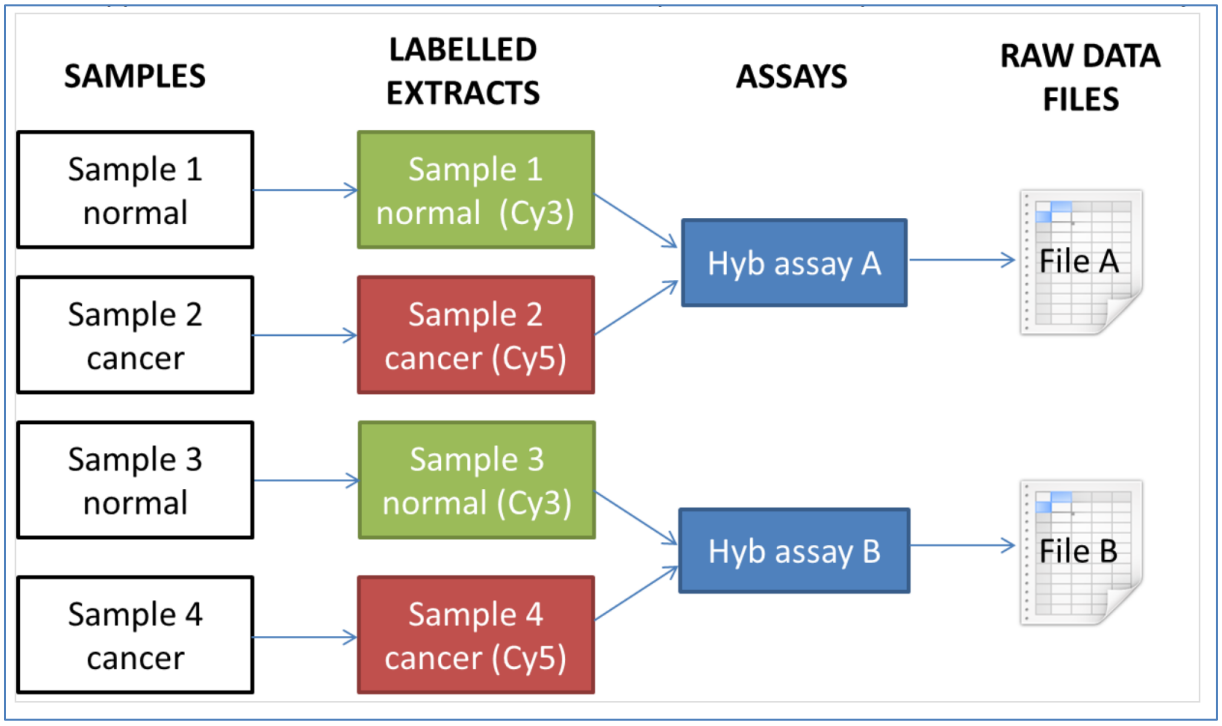

Figure 5. Example of the data structure supported by Annotare platform.

Since our data design is different from the one accepted by the platform, a curator helped with the final submission.

Once the data was submitted, the data could be accessed through an accession number. The date of public release of the data was set to 15/12/2024 and it can be modified if needed according to the dissemination rules of the DECISION project.

Data has been divided into 5 batches (589, 25, 278, 387 and 562 patients respectively) according to the date of sequencing of the data, and different submissions were made for every batch. Nevertheless, all the submissions can be found together with the same accession number.

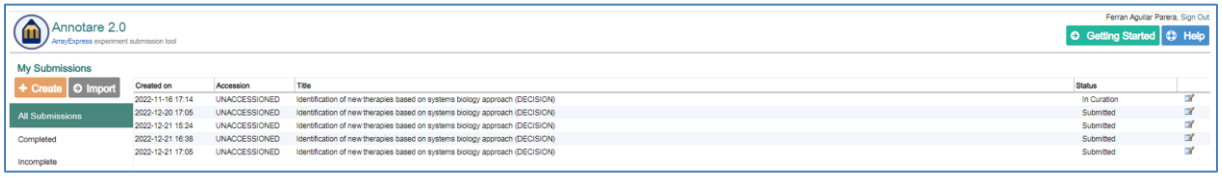

Figure 6. Summary of the 5 batches submitted

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.

This report reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.