EGA: Quick tour

This Quick tour provides a brief introduction to the European Genome-phenome archive (EGA).

Learning objectives:

- A basic understanding of the EGA and how to use it
- Where to go for more information about the EGA

What is the EGA?

The European Genome–phenome Archive (EGA) [2] is a permanent repository for all types of potentially identifiable genetic and phenotypic data from biomedical research projects.

Why do we need the EGA?

Participants in medical or genetic research projects have typically provided consent for their data to be used in research but not for open public distribution. These data require a secure archiving, processing and disseminating service that respects the original informed consent agreements. The EGA was created as a service to make sure that all such data can be made available for the researchers who have been granted access to the data in a secure and controlled way.

About the EGA

The resource accepts raw data [3] from sequencing, genotyping, transcriptome [4] or epigenetics experiments using next-generation sequencing platforms or array-based technologies. The EGA can also be used to archive any processed data [5], such as the locations of individual variations (e.g. SNPs) from the raw data or summary statistics from a particular project. The samples can be associated with phenotype [6] data that have been consented for use in research. The EGA supports pre-publication data release in accordance with the Toronto agreement [7].
The EGA can integrate the data with other available EMBL-EBI resources, for example by providing full genomic annotation [8] via Ensembl [9] for those variants that show significant association with the studied phenotype, or links to ArrayExpress [10] for accessing expression data deposited from the same cohort members. The data can also be integrated across individual studies. EGA data, together with data from our partner Data Access Committees (DACs [11]), are made available to users in the most widely used formats.

The studies in the EGA include summary statistics, dense genotype [12] experiments between cases and controls, population based studies or resequencing [13] and RNA-seq [14] data from various cancer genomes.

What can I do with the EGA?

- Search EGA [15] and dbGaP [16] experiments by keyword(s), publication(s), the platform(s) used or phenotype [6][s).
- Use your personal EGA account to receive alerts on new data that may be of interest to you.
- Find out how to apply for access to a particular dataset.
- Download genetic or phenotypic data (for which you have been granted permission).
- Use tools to integrate, download or visualise data from other resources at the EBI.
- Archive data, in accordance with the relevant release policy.
- Share data in accordance with the Toronto Agreement.

Searching for and gaining access to data in the EGA

Search for your study of interest
Choose your search result

Select your preferred dataset(s).
Figure 3 Identify the dataset(s) you wish to access.

**Contact the Data Access Committee (DAC).**
Figure 4 Apply to the appropriate Data Access Committee for access to the dataset(s) you desire.

**Accessing the data in the EGA**

The EGA [15] provides secure access to restricted data for authorised researchers and clinicians. In all cases, data access decisions are made by the appropriate Data Access Committee (DAC) and not by the EGA. Links to contact the relevant DAC are provided on each EGA study page, and you can also view a list of all DACs [11].

Approved users are issued with a personal EGA account associated with the relevant permissions. A data access agreement made between the DAC and the applicant dictates how data can be used, stored or transferred once it is downloaded from our system. There is more information about the DACs on our submission pages [17].

**Submitting data to the EGA**

The EGA [15] accepts a variety of genetic and phenotypic human data (see Figure 5).
Figure 5  The EGA provides a service for archiving, processing and disseminating all types of potentially identifiable genetic and phenotypic human data - on subjects who have consented for the data to be used in biomedical research, but not for unlimited public data release. Alternative data repositories that would allow public access to data are shown here.

You can also automate submissions by linking your local Laboratory Information Management System (LIMS) into our submission interface. For further information, examples, and compatible data formats, see the EGA submission page [18] or email the ega-helpdesk [at] ebi.ac.uk (EGA helpdesk).

Each submission should include statements that verify:

1. Submitted data to be consistent with the initial informed consent and with the applicable laws and regulations.
2. Submitter must have authorisation on behalf of the organisation.
3. A responsible individual or organisation must be identified for making data access decisions.

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.
Get help and support on the EGA

Support

- For all enquiries, please contact the ega-helpdesk [at] ebi.ac.uk (EGA helpdesk).

References


Collaborators

The EGA works closely with the Wellcome Trust Case Control Consortium and also coordinates with NCBI's dbGAP.

Funding

The EGA is supported by EMBL, the European Union, the Wellcome Trust and the UK Medical Research Council AP.

Contributors

Gary Saunders [1]

EMBL-EBI
Genetic Variation Scientific Curator - Paschall team: Variation

Gary Saunders is an EBI curator of the European Variation Archive and related resources: the Database of Genomics Variants archive and the European Genome-phenome Archive. It is Gary’s responsibility to manage the data within these resource(s) to ensure accuracy, clarity and discoverability. Previous to this position, Gary was a curator of the GENCODE project, which provides the gene set for the Ensembl genome browser.

Gary moved into curation following the completion of his PhD at the University of Glasgow, where he employed a variety of phylogenetic and bioinformatic methods to investigate drug resistance in nematode parasites of human and livestock importance. ORCID ID: 0000-0002-7468-0008