

Need help?

www.ebi.ac.uk/ega/
ega-helpdesk@ebi.ac.uk
 +44 (0) 1223 492630

The European Genome-phenome Archive
 EMBL-EBI
 Wellcome Trust Genome Campus
 Cambridge
 CB10 1SD, UK

The European Genome-phenome Archive (EGA)

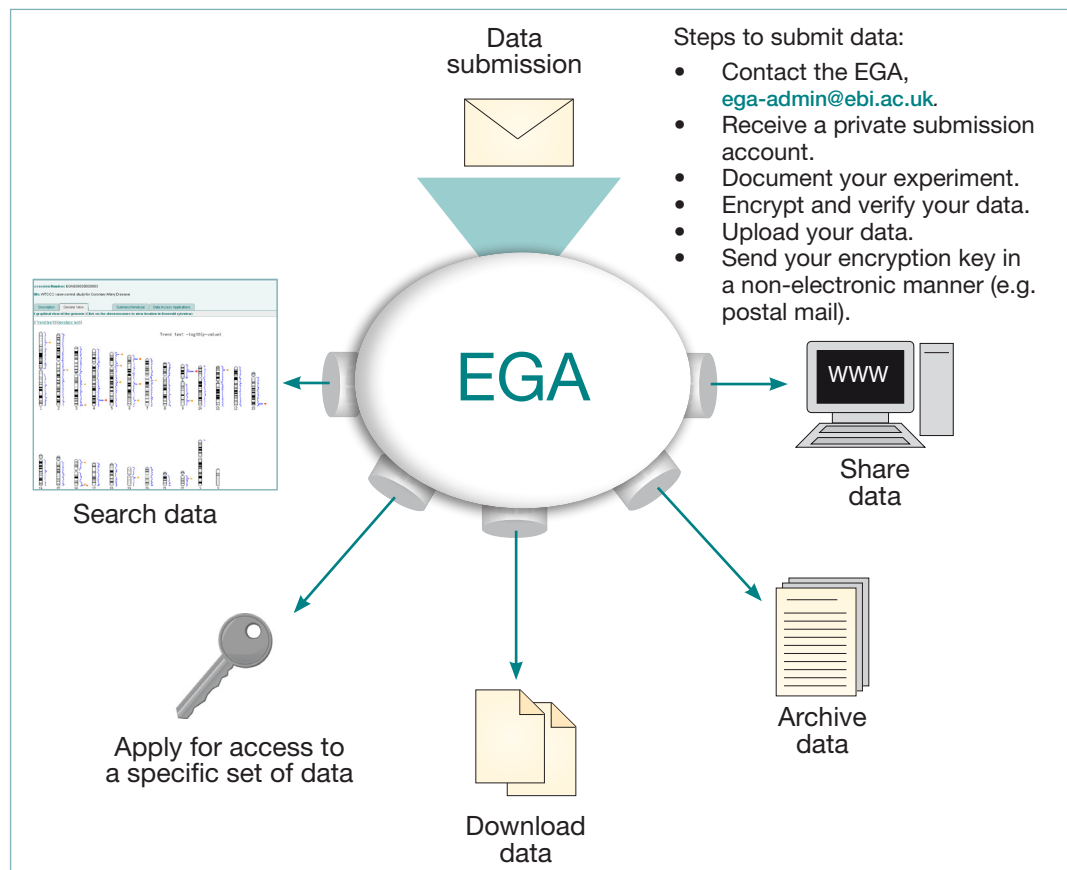
www.ebi.ac.uk/ega/

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 European Genome-phenome Archive (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.

What is the EGA?

The European Genome-phenome Archive (EGA) is a permanent repository for all types of potentially identifiable genetic and phenotypic data from biomedical research projects. The EGA contains data collected from individuals who have given consent for its use in research, but not for open public distribution.

The resource accepts raw data from sequencing, genotyping, transcriptome or epigenetics experiments using next-generation sequencing platforms or array-based technologies. The service can also be used to archive any processed data, 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 data that have been consented for use in research. The EGA supports pre-publication data release in accordance with the Toronto agreement (see Further Reading, side panel).



The steps for submitting data to the EGA and an illustration of how the data can be used to explore genetic and phenotypic information.

The EGA provides an analysis infrastructure to add value to the data submitted into our system. Our quality control applies to both samples and experimental data, without altering the original data. It allows us to merge data collected using different technologies, phase submitted data or determine unobserved genotypes by imputation (a method for substituting the missing data with statistically plausible values) from public resources, such as the 1000 Genomes project (www.1000genomes.org).

The EGA can integrate the data with other available EMBL-EBI resources, for example by providing full genomic annotation via Ensembl for those variants that show significant association with the studied phenotype, or links to ArrayExpress 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), are made available to users in the most widely used formats.

As of May 2011, the EGA contains data from over 50 studies, representing more than 130 000 individuals. These studies include summary statistics, dense genotype experiments between cases and controls, population based studies or resequencing and RNA-seq data from various cancer genomes.

What can I do with the EGA?

- Search EGA or dbGaP experiments by keywords, publication, the platforms used or phenotypes.
- 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 set of data.
- Download genetic or phenotypic data for which permission has been granted.
- Use tools to integrate, download or visualise data from other resources at the EBI.
- Archive data that support a publication, in accordance with the relevant release policy.
- Share data in accordance with the Toronto Agreement.

Submitting data to the EGA

See diagram for the steps to submit your data to the EGA. It is also possible to automate submissions by linking your local Laboratory Information Management System (LIMS) into our submission interface. For further information, examples, and compatible data formats, see www.ebi.ac.uk/ega/data-submission or contact us (see side panel, page 1).

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 and (3) a responsible individual or organisation must be identified for making data access decisions.

Accessing the data in the EGA

The EGA provides secure access to restricted data for authorised researchers and clinicians. In all cases, data access decisions are made by the appropriate DAC and not by the EGA. Links to contact the relevant DAC are provided on each EGA study page. 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. ●

Further reading

Toronto International Data Release Workshop. Pre-publication Data Release. *Nature*, 461, 168-70 (2009)

Flicek, P. *et al.* Ensembl's 10th Year. *Nucleic Acids Res.* 38, D557-62 (2010)

Kaye, J. *et al.* Ethical implications of the use of whole genome methods in medical research. *Eur. J. Hum. Genet.* 18, 398-403 (2010)

About the EGA

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

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