

Barcelona is the new home of the European Genome-phenome Archive (EGA), a fundamental resource for biomedical research

- The European Genome-phenome Archive, the EGA, stores genome and phenome data on over 100,000 people, from 200 centres and research groups from all around the world, and is a fundamental resource for the advancement of personalised medicine.
- At the moment, the EGA stores the data generated by over 700 scientific studies on cancer, diabetes, autoimmune diseases, cardiovascular problems and neurological disorders, amongst other illnesses.
- This data, which adds up to around 1,000,000 Gigabytes, will be stored in the Barcelona Supercomputing Center (BSC-CNS) facilities and subsequently analysed by the MareNostrum supercomputer.
- La Obra Social "la Caixa", the Government of Catalonia and the Spanish Ministry of Economy and Competitiveness are behind this co-managed project between the EMBL-European Bioinformatics Institute (EMBL-EBI) and the Centre for Genomic Regulation (CRG), reinforcing the leadership position of Spanish research groups and institutes in genome analysis at a European level.

Barcelona, 14th May, 2014.- In Palau Macaya this morning, **Carmen Vela**, Secretary of State for R&D from the Spanish Ministry of Economy and Competitiveness, **Andreu Mas-Colell**, Minister of Economy and Knowledge from the Government of Catalonia, and Jaime Lanaspá, director general of the "la Caixa" Foundation, along with Luis Serrano, director of the Centre for Genomic Regulation and Arcadi Navarro, affiliate head of the EGA team, ICREA Research Professor and director of the Department of Experimental Sciences and Health at Pompeu Fabra University, have publically launched the European Genome-phenome Archive (EGA) in Barcelona, a project led by the Centre for Genomic Regulation (CRG).

The EGA is the way to guarantee that genome and phenome data, which is notably expensive to obtain, is made available to the international scientific community, in such a way that research can speed up and lead to new discoveries. At the same time, the EGA is the European response to the challenges arising from the pressing need to protect the privacy of human donors who have taken part in genomic studies.

Some time ago, Janet Thornton, director of the EMBL-European Bioinformatics Institute (EBI-EMBL) and advocate of Elixir; Alfonso Valencia, director of the National Institute of Bioinformatics (INB-ISCIII); and Roderic Guigó, programme coordinator of

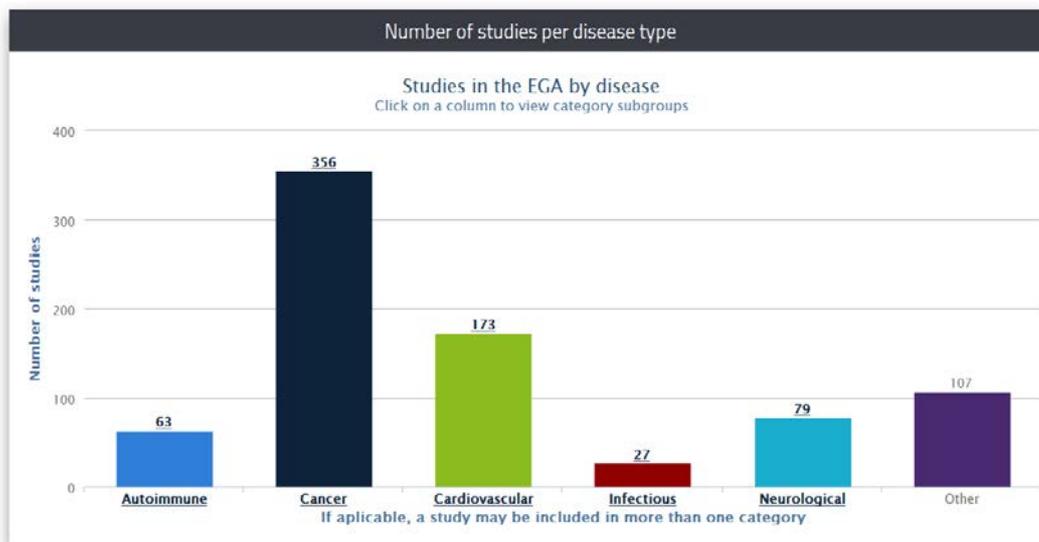
the Bioinformatic and Genomics research group at the Centre for Genomic Regulation (CRG) and lecturer at Pompeu Fabra University (UPF), began exploring the possibility of sharing the EGA and installing a copy of the database at the CRG. All three agreed that within the environment of the European collaboration networks, and specifically within the European bioinformatics infrastructure, ELIXIR, this initiative would be better and stronger. Subsequently, the CRG and the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS) agreed that the data would be physically stored in the facilities of the BSC, a centre that will also collaborate on the analysis of the data using the supercomputer MareNostrum, and through the efforts of its own researchers. Currently, the EGA-CRG team is made up of 6 people and is led by Arcadi Navarro, affiliate head of the team, ICREA Research Professor, and director of the Department of Experimental and Health Sciences at Pompeu Fabra University (UPF), together with a director from the INB-ISCIII.

The implementation of this initiative is the result of the work and joint effort of not only the CRG, but numerous organisations, including: the Obra Social "la Caixa"; the Spanish Ministry of Economy and Competitiveness, through their 'Centre of Excellence Severo Ochoa' programme; the Government of Catalonia; the Barcelona Supercomputing Center (BSC-CNS); the National Institute of Bioinformatics (INB-ISCIII); the ELIXIR Consortium; and the EMBL-European Bioinformatics Institute (EMBL-EBI). The setting up of the EGA was also helped by its being a pilot project within the European ELIXIR infrastructure

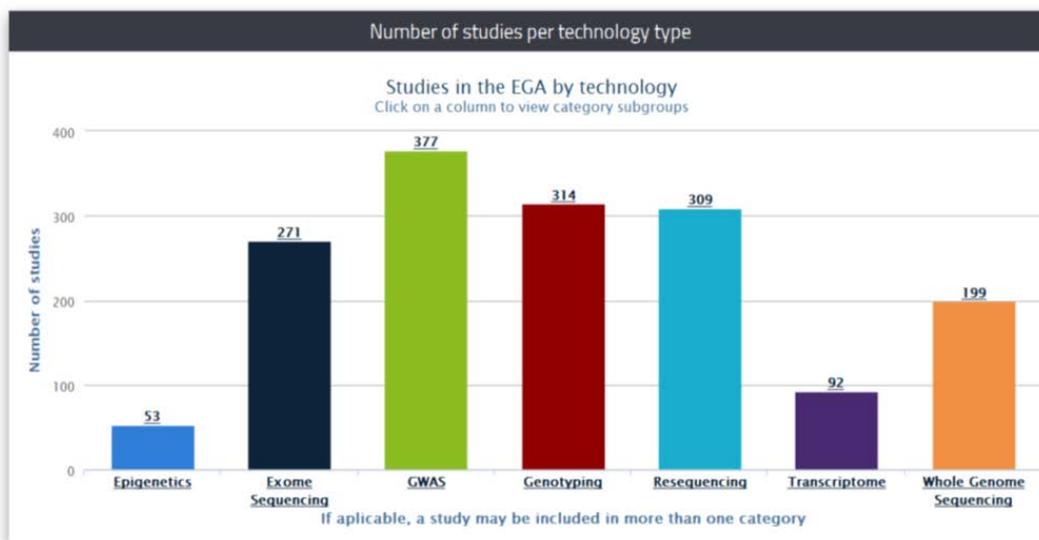
Up until now, the EGA (<http://www.ebi.ac.uk/ega/>) has been one of the EMBL-European Bioinformatics Institute's (EMBL-EBI) services. Its purpose is to permanently and safely archive and share, in a controlled manner, all types of genomic and phenomic data from identifiable people, which comes out of biomedical research projects, particularly studies related to complex diseases.

Basically, the EGA contains confidential data concerning information on genomic variants carried by both patients presenting disease phenotypes and healthy individuals. The information gathered belongs exclusively to people who consent to the publication of their data for use within the framework of scientific research, and to registered scientists.

The data stored in the EGA comes from more than 100,000 people, who, in the majority of cases, have complex diseases. It deals with illnesses that have a huge impact on public health, including many types of cancer (like breast and colon cancer), autoimmune diseases (like multiple sclerosis and diabetes), cardiovascular problems, and psychiatric illnesses, on the list of more than 50 different pathologies.



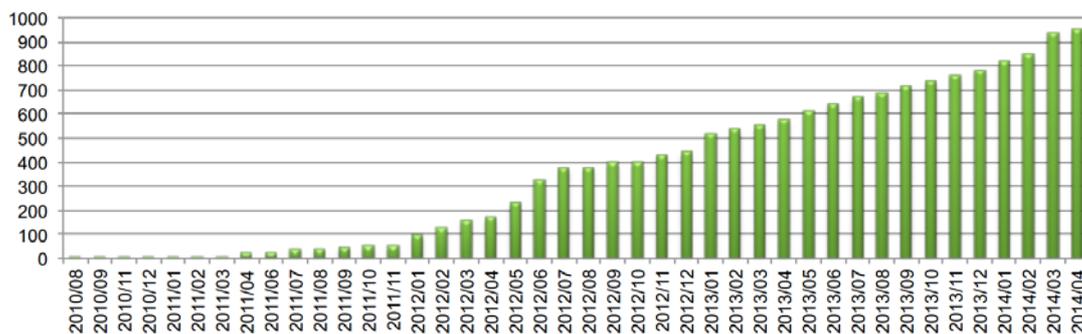
The EGA permanently archives various levels of information obtained using different technologies, including raw sequencing data (which could, for example, be reanalysed in the future with other methods or algorithms), in addition to the final genomic variants provided by the candidates. The EGA is designed to be a repository for all kinds of sequencing, epigenetic, and genotyping experiments, including control cases, population and family studies.



All this data has been generated by research groups, institutes and international consortia, not only in Europe, but across the globe. It involves a total of more than 140 international institutes, which include, by way of example, the Wellcome Trust (United Kingdom), the International Cancer Genome Consortium (ICGC, with Spanish participation), the University of Tokyo (Japan), Beijing University (China), the University of Harvard (USA), the University of Geneva (Switzerland), and the British Columbia Cancer Agency (Canada). They have all entrusted their data to the EGA, to ensure its safety and use for the benefit of human health.

In only the first 4 months of 2014, the data stored in the EGA was transferred more than 20,000 times, to nearly 5,000 users in research groups spread across five continents. In this way, the EGA ensures that the entire scientific community can access the valuable data in order to carry out research that would otherwise be impossible.

Once compressed, the total volume of data stands at approximately 1 PB (1,000,000 GB). Over the past twelve months, the EGA Catalogue has experienced a growth of 50% in the number of studies it contains, and 70% in the number of files. It is expected that in the next 12 months the total volume of files will multiply by 3.



To regulate how the EGA team manages this information, and how it is stored and distributed safely, there are strict protocols that depend on independent data access committees (DACs). At this time, the EGA contains data from almost 800 studies.

"The EGA is, among other things, an infrastructure which is needed in order to ensure that publicly funded omics data is properly stored, distributed quickly, and thoroughly analysed. Its content is essential for maximising the benefits obtained from investment into genomics, which has already become a key and strategic resource for enabling the development of personalised medicine. Only through EGA Europe can maintain its position as a leader in biomedical research", explains Arcadi Navarro. "For this reason, EGA-CRG is not just a backup: it provides more resources (infrastructure and talent) for the entire EGA project, and helps improve and expand its functionality", says Navarro.

"The EGA adds enormous value to Barcelona's already exceptional genomics and health research cluster, which includes leading institutions like the BSC-CNS, CNAG, IRB, CRG and many others, some of which are members of the Global Alliance for Genomics and Health (<http://genomicsandhealth.org/>), a global initiative whose objectives match perfectly with the EGA's mission. In a very important way, the EGA is an excellent example of how a range of different national institutes can join forces to reach a common goal. The innovation does not stop because of the recession", comments Luis Serrano, director of the CRG. "The EGA will boost the Barcelona brand as a reference city for the analysis of the genome and its relationship with disease. In addition, new bioinformatics tools will be developed around the EGA that will allow us to advance in the field of personalised medicine", concludes Serrano.

FOR FURTHER INFORMATION:

Centre for Genomic Regulation

Press office
Gloria Lligadas
gloria.lligadas@crg.eu
Tel. +34 93 316 01 53 – +34 608 550 788

Obra Social "la Caixa"
Communication area
Irene Roch
Iroch@fundaciolacaixa.es
Tel. +34 669 45 70 94

**Department of Economy and Knowledge
Secretariat for Universities and Research**
Laura Nicolás
prensa@gencat.cat
Tel. +34 93 552 67 43 - +34 699 514 029

Secretary of State for R&D
Luis Ordóñez Jimenez
prensaseidi@mineco.es
Tel. +34 91 603 75 09

Carlos III Health Institute
Mila Iglesias Garcia-Zarco
milagrosiglesias@isciii.es
Tel. +34 91 822 24 51

Barcelona Supercomputing Centre - Centro Nacional de Supercomputación
Communication Area
Gemma Ribas
gemma.ribas@bsc.es
Tel. +34 620 429 956