



ArrayExpress database doubles in size to 100,000 hybridisations

Hinxton, 11 December 2007 – ArrayExpress, the publicly available database of transcriptomics data at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), has doubled in size in 2007, reaching the 100,000-hybridisation milestone. The database now holds snapshots of gene expression (identifying which genes are specifically expressed in a particular tissue or in response to a drug, for example) for more than 180 species under thousands of experimental conditions.

The latest acceleration in growth reflects not only the increased numbers of direct submissions, but also the mass import of data from the Gene Expression Omnibus (GEO), which is produced by the US National Institute of Biotechnology Information. The import of data from GEO is the first step towards the regular exchange of data among public repositories for transcriptomic data. Similar data exchange agreements among biological data providers are widely recognised as being the most effective way of maintaining and quality-controlling the public record of biological information.

GEO data entering ArrayExpress are curated and scored for compliance to MIAME, the microarray community's mini-

um information standard. Users can therefore search for experiments that have been submitted to either database, annotated with common terms, and can download them in MAGETAB – a user-friendly tab-delimited format that simplifies meta-analyses of experiments from different labs. MAGETAB was developed under the direction of the MGED society, which works to simplify data sharing for microarray researchers.

The imported GEO data, like all the data in ArrayExpress, have also been integrated with other EBI resources. For example, users can now hop from an ArrayExpress entry straight to the relevant genes in Ensembl or proteins in UniProt, simplifying the process of data analysis and download for biomedical researchers.

The rate of growth of ArrayExpress seems set to increase further in the future, as new high-throughput sequencing-based transcriptomics applications are already resulting in the generation of huge amounts of data. Dealing with this barrage of new data will be the next challenge for ArrayExpress and its collaborators. ●

Contact:

Louisa Wright PhD, EMBL-EBI Scientific Outreach Officer, Hinxton, UK, Tel: +44 1223 494665, www.ebi.ac.uk, louisa@ebi.ac.uk
Anna-Lynn Wegener, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 452, www.embl.org, wegener@embl.de

Notes for Editors:

About EBI

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL) and is located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK). The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (EMBL-Bank), protein sequences (UniProt), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database), data from microarray experiments (ArrayExpress), protein-protein interactions (IntAct) and pathway information (Reactome). The EBI hosts several research groups and its scientists continually develop new tools for the biocomputing community.

www.ebi.ac.uk

About EMBL

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 20 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom). Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. EMBL's International PhD Programme has a student body of about 170. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome.

www.embl.de

About ArrayExpress

ArrayExpress is an international public repository for MIAME-compliant transcriptomics and related data. MIAME describes the Minimum Information About a Microarray Experiment needed to clearly interpret experimental results. ArrayExpress contains data from gene expression, comparative genomic hybridisation, genotyping and chromatin-immunoprecipitation (ChIP-chip) experiments.

www.ebi.ac.uk/arrayexpress

Policy regarding use

EMBL press releases may be freely reprinted and distributed via print and electronic media. Text, photographs & graphics are copyrighted by EMBL. They may be freely reprinted and distributed in conjunction with this news story, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: www.embl.org