

EBI (EMBL)



European Bioinformatics Institute of the
European Molecular Biology Laboratory

PRESS RELEASE

December 12, 2002

Talking genes: *New database shares DNA chip results*

Contacts:

European Bioinformatics Institute, EMBL

Cath Brooksbank, PhD
Scientific Outreach Officer
Hinxton Hall, Hinxton
Cambridge CB10 1SD
Great Britain
Tel: +44 (0)1223 492525
Fax: +44 (0)1223 494468
email: cath@ebi.ac.uk
www.ebi.ac.uk

EMBL Press Office:

Office of Information and Public Affairs
Meyerhofstr. 1
D-69117 Heidelberg, Germany
Tel: +49-6221-387252/452
Fax: +49-6221-387525
email: info@embl.de
www.embl.de

Scientific contacts:

Alvis Brazma, brazma@ebi.ac.uk
Helen Parkinson, parkinson@ebi.ac.uk
Ugis Sarkans, ugis@ebi.ac.uk
Mohammadreza Shojatalab, shoja@ebi.ac.uk

Policy regarding use:

EMBLpress releases may be freely reprinted and distributed via print and electronic media. Text, photographs and graphics are copyrighted by the 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.de



Talking genes:

New database shares DNA chip results

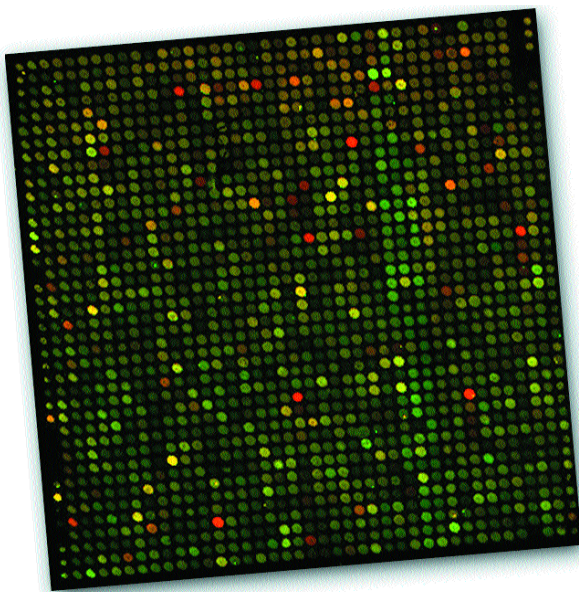
Hinxton (UK) and Heidelberg (Germany)

The European Bioinformatics Institute (EBI), one of the world's most important providers of biological data, has launched a new database for a type of information that is changing the way biologists and medical researchers work. *ArrayExpress* gives scientists all over the world free and direct access to data from experiments using "microarrays" – commonly called DNA chips. The service, which can be found on the Internet at www.ebi.ac.uk/arrayexpress, has now opened its doors for direct submission of data from scientists.

Microarray experiments give biologists a look at the complete genetic activity of cells. They can provide a fast-track to identifying disease-related genes and to answering questions like why patients respond to medications in different ways. The new database will allow scientists to compare results from a wide range of experiments and to ask completely new types of questions about living organisms.

Setting up the database and getting it ready for submissions has been an enormous challenge. "Microarrays generate huge amounts of data," says Alvis Brazma, who heads the project at the EBI. "A single experiment – for example, comparing the gene activity of normal breast tissue versus breast cancer tissue – can yield five million data points. Making the most use of that information will require storing, accessing and comparing such experiments from labs across the world."

DNA microarrays consist of rows of molecules mounted on a glass slide. These act as probes to reveal which genes are active in specific types of cells. All cells taken from the same person draw on an identical "archive" of genes, but each uses only about 20% of this collection to produce protein molecules. A cell's appearance and behavior depend on which subset of genes is active at a given time. Neurons use a different subset than skin or muscle cells. And if something changes – for example, if the cell is injured or suffers an infection – new genes will switch on and others will shut down.



A complete overview of this activity can help scientists identify genes that play key roles in disease and other biological processes. But before microarrays were developed, it was difficult for scientists to monitor the activity of more than a few genes at a time. Microarrays overcome this limitation because they can take snapshots of tens of thousands of genes – even a whole genome – at a time.

DNA chips can compare the activity of thousands of genes in different types of cells. Results from such experiments can give researchers hints about key molecules involved in diseases and other biological processes.



Jaak Vilo and
Alvis Brazma, EBI

DNA chips are already in use in major laboratories throughout the world, including at the European Molecular Biology Laboratory (EMBL) in Heidelberg, the parent institute of the EBI. Recently, Wilhelm Ansorge's research group, in close collaboration with the group of Bernhard Korn of the RZPD Ressourcenzentrum für Genomenforschung GmbH (Berlin-Heidelberg), has assembled a massive chip that can probe nearly the entire human genome. "The chip contains 52,145 samples representing known and predicted human genes," Ansorge says. "Nearly all human genes are represented here. This represents the largest and best-characterized collection of cDNA clones in the world, to date."

Each experiment with such a chip will produce an immense amount of data – potentially very useful if researchers can access the information and combine it with results from other experiments. Ansorge and researchers from all over the world have been active collaborators in helping the ArrayExpress team overcome a series of scientific and computational hurdles to achieve this goal.

Yet less than a year ago there was no common language for the exchange and comparison of microarray data. "Developing such a language was a key step in building a public repository that can handle different types of microarrays and the wide variety of protocols used in doing these experiments," says Ugis Sarkans, ArrayExpress's software development coordinator.

Many researchers from academia and industry put their heads together to create the language and new standards. The result was the *Microarray Gene Expression Markup Language (MAGE-ML)* and a protocol called *MIAME (Minimal Information About a Microarray Experiment)* that has to be provided so that researchers can correctly interpret results.

ArrayExpress has three goals. First, it will serve as an archive where microarray data can be stored. (While this data is essential to publications in scientific journals, there is too much of it to be printed.) Secondly, standardization will enable researchers to compare results from diverse experiments. Finally, it will promote a wider use of microarrays because it will encourage researchers to exchange chip designs and experimental methods.

What will a researcher get out of submitting data to ArrayExpress? The whole is far greater than the sum of its parts, according to Helen Parkinson, ArrayExpress's curation coordinator. "High-throughput microarray experiments produce datasets that no single biologist can analyze. These data have the potential to answer questions about biology beyond the scope of the



original experiment. By making them publicly available, everyone will have the opportunity to mine them."

The scientific press has also recognized the importance of ArrayExpress. Two journals – *Nature* and *EMBO Journal* – already require that microarray data in support of a paper be submitted to a publicly accessible database. Several more journals, including *Cell* and *The Lancet*, are encouraging authors to adhere to MIAME standards.

Other publications will probably follow suit – in the past, such requirements have been a key step in ensuring the success of gene, protein, and other databases.



DNA chips are made by mounting molecular probes on ordinary glass slides.

The ArrayExpress team is making it as easy as possible for researchers to submit their data, especially as many chip users aren't experts in the use of databases. A data submission and annotation tool called MIAMExpress is available on the web at the address www.ebi.ac.uk/miamexpress. "Putting one's data into the standard MAGE-ML format may be quite a daunting task for a small laboratory without good bioinformatics support," says Mohammadreza Shojatalab, the main developer of the tool. "MIAMExpress will do it for you."

Several commercial companies have also jumped on board to make the process of submission easy. Many users rely on standardized DNA chips produced by industry. ArrayExpress will hold descriptions of popular microarrays. "We are working with Affymetrix and Agilent to make that happen," Sarkans says, "and we're hoping that other chip manufacturers will follow suit. Researchers will be able to use MIAMExpress to provide necessary information about their experiments without having to describe the chips themselves."

- by Cath Brooksbank & Russ Hodge

Scientific contacts:

Alvis Brazma, brazma@ebi.ac.uk
Helen Parkinson, parkinson@ebi.ac.uk
Ugis Sarkans, ugis@ebi.ac.uk
Mohammadreza Shojatalab, shoja@ebi.ac.uk

The MIAMExpress support team, which includes both data management and software development experts, can also be reached by e-mail (miamexpress@ebi.ac.uk).

The Microarray Gene Expression Society (MGED), which developed the standards used by ArrayExpress, will hold its next meeting in February 2003, as part of the AAAS Annual meeting. Please see the AAAS website for details.

**About EMBL**

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 16 member states, including most of the EU, Switzerland and Israel. 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, Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and an external research programme in Mouse Biology in 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, and to develop new instruments and methods in the life sciences. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome. For more information see the EMBL website at: www.embl.de

or contact:

the Office of Information and Public Affairs
European Molecular Biology Laboratory (EMBL)
Meyerhofstr. 1
D-69117 Heidelberg, Germany
Tel: +49 (0)6221 387252
Fax: +49 (0)6221 387525
email: info@embl.de

About the EBI

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL), and is located on the Wellcome Trust Human 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 (SWISS-PROT and TrEMBL), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database) and a new resource to hold data from microarray experiments (ArrayExpress). The EBI hosts several research groups and its scientists continually develop new tools for the biocomputing community.

contact:

Cath Brooksbank PhD
Scientific Outreach Officer
European Bioinformatics Institute, EMBL
Hinxton Hall, Hinxton
Cambridge CB10 1SD
Great Britain
Tel: +44 (0)1223 492525
Fax: +44 (0)1223 494468
email: cath@ebi.ac.uk
www.ebi.ac.uk

The EMBL member states are:

Austria, Belgium, Denmark, Finland, France, Germany, Greece, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom