



# EMBL-EBI

## Press Release

4 February 2003

For immediate release



Ewan Birney

### Scientific contacts:

#### *Ewan Birney*

European Bioinformatics Institute  
Wellcome Trust Genome Campus  
Hinxton

Cambridge, CB10 1SD  
United Kingdom

Tel: +44 (0)1223 494 420

Fax: +44 (0)1223 494 468

Email: [birney@ebi.ac.uk](mailto:birney@ebi.ac.uk)

#### *Lincoln Stein*

Cold Spring Harbor Laboratory  
P.O. Box 100

1 Bungtown Rd.

Cold Spring Harbor

USA - NY 11724

Tel: +1516 367-8380

Fax: +1 516 367-8389

Email: [steinl@cshl.edu](mailto:steinl@cshl.edu)

### EMBL Press Contact:

Office of Information and Public  
Affairs

Meyerhofstr. 1

D-69117 Heidelberg

Germany

Tel: +49 6221 387 252/443

Fax: +49 6221 387 525

email: [info@embl.de](mailto:info@embl.de)

<http://www.embl.de>

## Linking the levels of life

### *From genes to cellular processes with the Genome Knowledge Base*

The best place to learn about a small town you plan to visit would be a travel brochure or a history book. If none exists, you might have to spend days combing libraries and archives for information before your trip. A similar problem confronts scientists when they search genome databases for information about genes. "What they want to know may be there," says Ewan Birney, of the European Bioinformatics Institute (EBI) in the UK, "but figuring out what it means, and whether it's important, may require days in the library."

Birney has teamed up with Lincoln Stein of Cold Spring Harbor Laboratory (New York) to launch the "Genome Knowledge Base" (GKB), a sort of "travelogue" of the human genome. It goes online for the first time today at [www.genomeknowledge.org](http://www.genomeknowledge.org). The information is directly linked to *Ensembl*, the public database of human, mice, and other genomes run by Birney's team and the Sanger Centre.

Scientists are rapidly learning new things about how individual genes participate in processes like cell division, cell differentiation, or cancer. In *Ensembl* and other databases, these discoveries appear as a sort of exhaustive series of footnotes. This often makes it hard to see the forest for the trees. The Genome Knowledge Base aims to make things much simpler.

"When scientists talk about the cell," Birney says, "they say things like, 'Once this molecule recruits its partner kinase, the resulting complex switches on a number of other target genes.' It is a process view, in contrast to the gene-centric or protein-centric view of most databases." Representing biological processes in a database is regarded as one of the biggest current challenges for bioinformatics.

In the past, discoveries came at a slower pace, so researchers had the time to spend a day or two in the library. And the current system generally works fine, Birney says, if a scientist is working on a specific process and comes to a database for information about a gene.

**Genome KnowledgeBase**

Assembly of the pre-replicative complex

DNA replication initiation in eukaryotic cells begins with the formation of the pre-replicative complex (pre-RC) during the late M and G1 phases of the mitotic cell cycle, a process also called DNA replication licensing. The association of initiation proteins (ORC, Cdc6, Cdt1, Mcm2-7) with the origin of replication in both *S. cerevisiae* and humans has been demonstrated by chromatin immunoprecipitation experiments [9335335]. In *S. cerevisiae*, pre-replicative complexes are assembled from late M to G1 [9407030, 9288745]. In mammalian cells as well, pre-replicative complexes are assembled from late M to G1, as shown by biochemical fractionation and immunostaining [11046155, 11801723]. There are significant sequence similarities among some of the proteins in the pre-replicative complex. The ORC subunits Orc1, Orc4 and Orc5 are homologous to one another and to Cdc6 [9829972]. The six subunits of the Mcm2-7 complex are homologous to one another. In addition, Orc1, Orc4, Orc5, Cdc6, and the Mcm2-7 subunits, are members of the AAA+ superfamily of ATPases [9927482]. Since the initial identification of these pre-RC components other factors that participate in this complex have been found, including Cdt1 in human, *Xenopus*, *S. pombe*, and *S. cerevisiae* cells.

**Replication Initiation**

The diagram illustrates the assembly of the pre-replicative complex and the pre-initiation complex. It shows the origin DNA with Mcm10, Cdc6, and the Mcm2-7 complex. The pre-replicative complex is formed by the association of Mcm10, Cdc6, and the Mcm2-7 complex. The pre-initiation complex is formed by the addition of Pol 4/primase, Cdc45, and RPA to the pre-replicative complex.

One of the processes covered in the screenshot is the means by which cells replicate their DNA.

But new types of experiments have changed things. For example, work at the EBI's parent institute, the European Molecular Biology Laboratory in Heidelberg (EMBL), has recently turned up over two hundred genes that might be involved in how mosquitoes transmit malaria. Looking them all up would be a gargantuan task that the GKB aims to make easier.

"You might find that twenty of those genes are part of a very common biological process," Birney says. "But if you can't spot that easily in the database, you might overlook them as unconnected single genes."

While this is not the first effort to tackle the problem, three things make the resource special. First, it will be entirely in the public domain, and Birney and Stein invite scientists everywhere to use it, link into it, and add to it. Secondly, researchers can move smoothly back and forth between the GKB and *Ensembl*, which many scientists regard as the world's best public version of human and other genomes.

Finally, the researchers are taking painstaking steps to ensure a very high quality of information. "Everything in the Knowledge Base points back to original experimental sources," Birney says. "Some of what we know about basic processes like metabolism goes back fifty years or more, and you have to crack old textbooks to find it. New or old, all the information is reviewed by a top expert, and we invite help from scientists in each field."

The initial version of the GKB presents four major processes in human cells. Several more will be covered in a second release planned for March. With active support from the research community, the resource should quickly grow by leaps and bounds.

- Russ Hodge

## **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, 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, 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.

### **Contact:**

Office of Information and Public Affairs  
European Molecular Biology Laboratory (EMBL)  
Meyerhofstr. 1- D-69117 Heidelberg - Germany  
Tel: +49 (0)6221 387252/443  
Fax: +49 (0)6221 387525  
email: [info@embl.de](mailto:info@embl.de)  
<http://www.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](mailto:cath@ebi.ac.uk)  
[www.ebi.ac.uk](http://www.ebi.ac.uk)

## **EMBL member states:**

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

### **Policy regarding use:**

*EMBL press 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](http://www.embl.de).*