

# INFOBIOMED

## NEWSLETTER

N.3 SEPTEMBER 2005



```
TCCTCCTAGCAACGCGGCCCTCTCGGCTCTTAAGGCGTGACTTGCCATGGATGTTCTGCC  
TTTGGAAACCAAGTGGGTGGTCACAGCACAGAGGC GTTGGTTTCATGGCTGTCACATGCAC  
AGAGACAGCCCTGTCCACTGAGCCTGTTCTGTGGCAGAAGGAACCCCTAGCCTGGACC  
TCCTTGGGTGGAAACACCAGCCCTCAGCTCTGACTGTCATTTTCAGATGTGAAGAGAA  
ACACAGTGTGCTGATGTTGTATTGCCGTGAACAACGAGAGTTTCTTTGACTTC  
TAACTGCCTTAAAGCTAGAGCCCGGCATTAGGATGGCAGTGGTGGTAAATGAGGTATG
```

**BIOMEDICAL INFORMATICS  
TO SUPPORT INDIVIDUALISED HEALTHCARE**

**SUPPORT INDIVIDUALISED HEALTHCARE**



## SUMMARY

### NEWS

PG. 2

Harvard Project to scan millions of medical files	pg. 2
PSI will determine the shapes of 3,00-4,00 Proteins in the Second Phase of Project	pg. 2
IFPMA improves Biomedical Data Transparency with Launch of First Worldwide Clinical Trials Portal	pg. 3
Perlegen and RIKEN Partner on 47-Disease Genomic Association Study	pg. 4
Public collections of DNA and RNA sequence reach 100 gigabases	pg. 4

### INFOBIOMED PROFILES

PG. 5

Leicester	pg. 5
HHUD	pg. 5

### RESOURCES

PG. 7

Bioinformatics Links Directory	pg. 7
Infobiomed Brokerage Service	pg. 7
Infobiomed Course Database (ICD) Publication	pg. 7
Pride, an Open-Source Proteomics Database, launched	pg. 8
Ballview: a tool for research and education in molecular modeling and a visualization application	pg. 8

### EVENTS

PG. 9

### COMMENT

PG. 9

## NEWS

### Harvard Project to scan millions of medical files

**Source:** Boston Globe, Gareth Cook, Globe Staff.  
July 3rd, 2005

[http://www.boston.com/news/science/articles/2005/07/03/harvard\\_project\\_to\\_scan\\_millions\\_of\\_medical\\_files/](http://www.boston.com/news/science/articles/2005/07/03/harvard_project_to_scan_millions_of_medical_files/)

Harvard Scientists are building a software tool that by using artificial intelligence will be able to scan medical files looking for links between patients' DNA and illnesses. It will draw out key information that with the traditional method would take researchers months or years. It will speed the process of information retrieval from medical records thus providing researchers with more information in a shorter span of time. As for any concerns about patient privacy that may arise, the program called "I2B2" (Informatics for Integrating Biology and the Bedside") will comply with the guidelines set out in federal and state laws.

This project that started in 2004 will last five years and is funded with 20 million dollars will undertake as its first study an effort to understand the genetics roots of asthma. A disease that affects a large population that usually does not respond well to traditional treatments and suffers from repeated asthma attacks. In this study the computer will be programmed to search out for key phrases that may prove useful such as the smoking habits of the patients. One of the main challenges is to get the computer to extract accurate useful information from the medical records that can be used in research.

When dealing with genetics information the main problem is the absence of this kind of data in medical files so the project team is looking into the possibility of setting a DNA database. If this were to be established a fairly complete Biobank with genetic and clinical data would be available for researchers to study and retrieve information on several diseases. However for this possibility to become a reality there are some considerations to be taken into account such as ethical and logistic issues.

### PSI will determine the shapes of 3,000-4,000 Proteins in the Second Phase of Project.

**Source:** GenomeWeb News Bulletin.  
July 1st, 2005

[www.nigms.nih.gov/psi/](http://www.nigms.nih.gov/psi/)

GenomeWeb News published in its July 1st bulletin the announcement made by the NIH of the launch of the second phase of the Protein Structure Initiative (PSI). The initial pilot phase of this project run from September 2000 to June 2005. During this first phase the objective had been to develop new approaches and tools for the determination and generation of 3D protein structure models.

The long-range goal of the Protein Structure Initiative as stated in its mission statement is "to make the three-dimensional atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences" (<http://nigms.nih.gov/psi/mission.html>). The determination of the 3D structure of proteins will help to better understand and study their functions, the roles that they play both in health and disease and will allow to advance in the design of new drugs among other things.

During the first phase of this project the focus was set on the development of tools that would make the process of determining the structure of proteins a highly automated one and to incorporate new methods that turn DNA sequence information into protein structures. During this phase the structure more than 1000 proteins was determined including some unique structures solved (structures sharing less than 30 percent of their sequence with other known proteins): More than 700. The aim of this second phase, also called the rapid production phase, is to determine the shapes of 2000-3000 proteins shifting focus from the development of tools to the use of the already developed tools to rapidly determine the structure of proteins. The structures determined will also be submitted to the Protein Data Bank (<http://www.rcsb.org/pdb/>). A repository where researchers can access PSI-generated information.

The project is mostly funded by the National Institute of General Medical Sciences (NIGMS), part of the National Institutes of Health.

This new production phase includes two types of centers. Four large-scale centers, established during the pilot phase and six specialized centers that will develop novel methods for determining the structure of those proteins that have been up to now hard to study; this includes small protein complexes.

The four large-scale centers are:

- 1:: Joint Center for Structural Genomics:  
<http://www.jcsg.org>
- 2:: Midwest Center for Structural Genomics:  
<http://www.mcsg.anl.gov>
- 3:: New York Structural GenomiX Research Consortium: <http://www.nysgrc.org>
- 4:: Northeast Structural Genomics Consortium:  
<http://www.nesg.org>

and the six specialized centers are:

- 1:: Accelerated Technologies Center for Gene to 3D Structure: <http://www.atcg3d.org>
- 2:: Center for Eukaryotic Structural Genomics:  
<http://www.uwstructuralgenomics.org>
- 3:: Center for High-Throughput Structural Biology
- 4:: Center for Structures of Membrane Proteins
- 5:: Integrated Center for Structure and Function Innovation
- 6:: New York Consortium on Membrane Protein Structure

## IFPMA improves Biomedical Data Transparency with Launch of First Worldwide Clinical Trials Portal

In Geneva September 21 2005, the International Federation of Pharmaceutical Manufactures & Associations (IFPMA) announced the launch of IFPMA Clinical Trials Portal, the first Internet search engine to link to on-line information about clinical trials worldwide that was developed with IBM.

IFPMA Clinical Portal is located at <http://www.ifpma.org/clinicaltrials.html> and contains more than 250.000 links.

According to Dr. Daniel Vasella, Chairman and CEO of Novartis and President of the IFPMA, the web portal will *"provide doctors, patients and their families with simple access to the most complete information on clinical trials of drugs and vaccines"*.

An important part of this Portal is the search engine that includes the access of more relevant clinical trial information through pharmaceutical company sites or pharmaceutical industry association resources such as the US pharmaceutical association –PhRMA– or government sites such as the US National Library of Medicine's. The European Union's will participate in the future too.

The portal has been designed as a single entry allowing you to search for comprehensive information on on-going clinical trials (registry) or results of completed trials (database) conducted by the innovative pharmaceutical industry.

In the first stage of the development of this portal the priority established that the patient could also have access to this information. The pharmaceutical industry recognizes the requirement of further efforts to increase its ease of use, include more languages, facility the comprehension of medical terminology to patients, etc., that will be improved in the next stages.

IFPMA, the European Federation of Pharmaceutical Industries and Associations (EFPIA), the Japan Pharmaceutical Manufacturers Association

(JPMA), the Pharmaceutical Research and Manufacturers of America (PhRMA) and Canada's Research-Based Pharmaceutical Companies (Rx&D) decided in its "Joint Position on the Disclosure of Clinical Trial Information via Clinical Trial Registries and Databases", issued in January 2005 to develop a blueprint for improving clinical trial transparency, an objective that IFPMA Clinical Trials Portal will contribute to achieve.

### **Perlegen and RIKEN Partner on 47-Disease Genomic Association Study**

<http://www.bionity.com/news/e/47796/?sort=2>

As part of a collaboration agreement, Perlegen subsidiary Perlegen Sciences Japan and the Japanese Institute of Physical and Chemical Research, or RIKEN, will conduct genomic association studies to find the causes of 47 common diseases.

The statement provided by the organizations said that the collaboration will involve the analysis of more than 250,000 SNPs in samples collected as part of the BioBank Japan initiative coordinated by the University of Tokyo's Institute of Medical Science, which includes 66 hospitals across the country. At present, the collection includes DNA and clinical data from about 150,000 patients.

Perlegen will use "next-generation" Affymetrix technology in first-pass genome scanning, while RIKEN's SNP Research Center will conduct fine mapping in candidate regions associated with disease, and will conduct follow-up research on promising genomic regions, the statement said. Initial studies will focus on cancer and heart disease, but the project will eventually also include atherosclerosis, rheumatoid arthritis, epilepsy, heart failure, chronic obstructive pulmonary disease, asthma, arrhythmia, cerebral infarction, and several cancers, the partners said.

The results of these studies may help "identify... people who are at high risk of developing the disease, which may lead to better prevention, earlier detection and improved treatments."

### **Public collections of DNA and RNA sequence reach 100 gigabases**

**Source:** Medical News Today. 23 August 2005

<http://www.medicalnewstoday.com/medical-news.php?newsid=29489&nfid=rssfeeds>

*"The world's three leading public repositories for DNA and RNA sequence information have reached 100 gigabases [100,000,000,000 bases; the 'letters' of the genetic code] of sequence. Thanks to their data exchange policy, which has paved the way for the global exchange of many types of biological information, the three members of the International Nucleotide Sequence Database Collaboration [INSDC, <http://www.insdc.org>] - EMBL Bank [Hinxton, UK], GenBank [Bethesda, USA] and the DNA Data Bank of Japan [Mishima, Japan] all reached this milestone together".*

## INFOBIOMED PROFILES

### Leicester



## University of Leicester

The University of Leicester is one of the UK's leading research and teaching universities. It was founded as a University College in 1921, was granted a Royal Charter to act as an independent university in 1957, and currently has over 18,568 students including 10,327 at postgraduate level.

The University of Leicester's Research and Business Development Office (RBDO) provides central support for the financial administration of the project. The RBDO has a full time staff of 22, is responsible for all central administrative and managerial functions relating to research and in 2001, on behalf of around 800 academic staff and over 10,000 post-graduate students, administered an income of around 52 Million Euro. The eight-strong Research Finance Section of the University Research and Business Development Office provides full financial support for all EU projects, including provision for fully audited accounting.

The University Research and Business Development Office employs a full time European Projects Manager who has several years experience in industry and in academia on all aspects of FP5 management, including proposal writing and submission, contract negotiation at Brussels with the Commission, as well as financial and technical co-ordination of projects (Management Reports, Technological Implementation Plans, Cost Statements etc.). In recent years the Office has been involved in management and co-ordination of several EU projects, not least Framework Five (81 projects).

The INFOBIOMED project is hosted in the Department of Genetics, which has an excellent and formidable international reputation for its pioneering bio-medical research. The department's excellence was recently recognized by the award of 5\*/5 in the 2001 National University's Research

Assessment Exercise, thereby characterizing it as the best genetics department in the UK. Suitable facilities and computing resources for the project are available in the Genetics Department, and supplementary needs for high-capacity computing may be met via our collaboration with EBI, and/or by a Linux cluster recently established locally. Department staff provides support for all IT matters and email/network security.

### INFOBIOMED research group in Leicester

The group has recently established itself in the UK, and is enlarging as part of this move. The group is led by Professor Anthony J Brookes, who has been an international leader in the SNP research field since its early days, and who created and still runs the Human Genome Variation database (HGVbase). Senior researchers in the group include programmers with relevant experience, who organize and check day-to-day project activities. Most programming is performed also in conjunction with computer science graduates / PhD students and various international collaborators.

### HHUD



<http://www.uniklinik-duesseldorf.de/kks>

In the midst of the largest accumulation of universities and research facilities in Europe the Heinrich-Heine University at Düsseldorf (HHUD) is the prominent regional center of science and research in various regards. Particularly the university clinic plays a leading role in many fields of medical research for years. The increasing need of high-quality studies for the improvement of established treatments or development of new innovative therapies - in the sense of a long persisting health and improved quality of life of the patient - has to be taken into account. This demand culmi-

nated into the establishment of the Coordination Center for Clinical Trials (German: Koordinierungszentrum für Klinische Studien Düsseldorf, KKSD). The coordination center at the university clinic has evolved from the Institute for Theoretical Surgery and is until today at home in the surgical clinic's building.

As a central service unit of the medical faculty the KKSD is considerably involved in the planning, conduction and evaluation of multicentric clinical trials. In addition, non-university projects concerning clinical studies are accompanied by the KKSD as a partner.

At present, the team consists of 18 coworkers from as different backgrounds as medicine, computer science and biology. By the end of November 2005 the KKSD is involved in 30 clinical trials which – besides several innovative pharmacological studies - encompass exciting studies as the first national study on the effectiveness of acupuncture and a trial on the evaluation of preventive colorectal surgery in patients with HNPCC syndrome.

An important task is fulfilled by the WG public relations and training which provides training in the field of clinical trials in form of courses for students of the medical sciences and health professionals of all disciplines. Furthermore, the KKS offers physicians and scientists the opportunity to work for a limited duration as a so-called "rotation physician" to be introduced to the field.

The spatial like contentwise proximity let the KKSD grow to a junction between classical medicine, medical informatics and bioinformatics and a center of an active experience and information exchange between providers and users of medical technology. Over the past years the KKSD has earned major experience in the development, implementation and evaluation of clinical information and medical decision-support systems. There is proven expertise in the field of validation and evaluation of the usability and user-friendliness of medical IT-systems, including clinical information and decision-support systems, and in the evaluation of the clinical impact of these systems.

The head of the unit has coordinated the working group "evaluation" of the German-MEDWIS programme (1995 - 2000), was involved in the VATAM project about evaluation (1998) and is participating in the European Network on Evaluation of IT in Healthcare. It also coordinated an EU INCO COPERNICUS Concerted Action about "Information Technology for the Quality Assurance in Acute Abdominal Pain" (EURO-AAP, 1995 - 1997). The unit was a major partner for validation and evaluation of medical information and knowledge-based systems in the following EU projects TELEGASTRO (Telematics Technology for Harmonizing and Promoting Quality in Gastroenterology, AIM Programme), ELCANO (European and Latin-American Countries Associated in a Networked database of Outstanding Guidelines in unusual clinical cases, INCO/DC Programme), SMARTIE (Smart Medical Applications Repository of Tools for Informed Expert Decision, IST-2000-25429) and LIQUID (Language Independent Querying for Information Discovery, IST-2000-25324). In addition, the unit is involved to a great extent in quality assessment (standardized quality management involving standard operating procedures, member of the Steering Committee on Quality Management of the university clinic).

## RESOURCES

### Bioinformatics Links Directory

[http://bioinformatics.ubc.ca/resources/links\\_directory/](http://bioinformatics.ubc.ca/resources/links_directory/)

The description of directory provided in its web page is: *"The Bioinformatics Links Directory features curated links to molecular resources, tools and databases. All of the resources are free or available for a nominal fee. The links listed in this directory are selected on the basis of recommendations from bioinformatics experts in the field. We make every effort to list useful, well tested resources instead listing all tools"*.

The website also has a mailbox for suggestions from the community of bioinformatics to improve their service.

They included in the directory in 2004 the list of the links contained in the Nucleic Acids Research Webserver issue.

### Infobiomed Brokerage Service

<http://www.infobiomed.org>

The Network of Excellence INFOBIOMED's Mobility Brokerage Service (MBS) is envisaged as a secure web-based marketplace addressed to create and to reinforce a Biomedical Informatics (BMI) European Community.

In this web site, host organizations will be able to publish job or mobility offers and candidates will be able to post their mobility or job demands, including their personal data, interest and subjects. The MBS is intended to automatically match offers and demands, apart from allowing access to the information posted.

The MBS will be open to the whole BMI community from the beginning.

### Infobiomed Course Database (ICD) Publication

The INFOBIOMED project has as one of the goals of its Training and Mobility program to set up a comprehensive database of academic courses on BMI and related MI and BI courses.

The ICD will be accessed through the INFOBIOMED website and will be open to any organisation willing to introduce its training activities. Initially, the focus will be on integrative BMI courses and programs, however, BI and MI courses can be included as well. This database will at the beginning include information on the current academic programs of the network partners, however, it has also the future goal of becoming the one-stop repository for European BMI training information.

To reach this goal a state of the art Information Retrieval (IR) methodology will be used to easily access and administer the course database. The ICD will be a useful tool first of all by the organisations to entry their courses, and then by the users to perform searches through it.

Other activities performed in the field of training by INFOBIOMED have been the study of the current BMI education situation in Europe and comparing it to the ideal situation (represented by an integrative ideal training curriculum in BMI).

This has allowed the Training and Mobility Committee of the Network of Excellence INFOBIOMED to identify the current existing gaps in training in BMI, and to subsequently address those gaps with an ad hoc action plan.

Comparison of the "ideal" situation versus the "reality" and gap Identification will lead to Issue recommendations on how to address these gaps, which will derive on the organisation of newly developed or adapted training activities.

## PRIDE, An Open-Source Proteomics Database, launched

**Source:** GenomeWeb Daily News.  
30 August 2005 and PRIDE web page

<http://www.genomeweb.com/articles/view.asp?Article=2005830121628> and  
<http://www.ebi.ac.uk/pride/init.do>

The Proteomics Identifications Database (PRIDE, <http://www.ebi.ac.uk/pride/init.do>), an open-source proteomics database was launched last week of August 2005. PRIDE was developed by a consortium of research institutes and universities in the UK and Belgium, that includes the European Bioinformatics Institute in Hinxton, UK; and Flanders Interuniversity Institute for Biotechnology and Ghent University, both in Belgium.

The database *“is a centralized standards compliant, public data repository of proteomics data”* that includes data from the results of the Human Proteome Organization's Plasma Proteome Project, and of the human platelet proteome set published by Ghent University. Future plans include the addition of the results of other proteomic research initiatives such as HUPO's Liver Proteome Project. Submission of data is opened to any proteomics laboratory generating *“protein and peptide identification data to be published in peer-reviewed publications”*.

The PRIDE database, source code, data, and support tools are freely available for web access or download.

## BALLView: a tool for research and education in molecular modeling and a visualization application.

**Source:** Bioinformatics. 6 December 2005

<http://bioinformatics.oxfordjournals.org/cgi/content/full/22/3/365>

The molecular modeling framework BALL (Biochemical Algorithms Library) has developed the BALLView, a free molecular modeling and molecular graphics tool. It provides fast OpenGL-based visualization of molecular structures, molecular mechanics methods (minimization, MD simulation using the AMBER and CHARMM force fields), calculation and visualization of electrostatic properties (FDPB). BALLView is a new tool for research in structural bioinformatics and molecular modeling. Its interface, that uses the scripting language Python, offers novel users and teachers an easy way to access to the full functionality.

BALLView is Open Source software available under GNU Public License (GPL). This viewer project page is linked at <http://www.ballview.org> and runs on all major platforms (Windows, MacOS X, Linux, and most Unix flavours).

## EVENTS

**HIMSS 06 Annual Conference and Exhibition**  
San Diego, USA. February 12-16, 2006  
<http://www.himss06.org/>

**EFMI Special Topic Conference 2006 & ROMEDINF**  
Integrating Biomedical Information: from e-cell to e-Patient. Timisoara, Romania. 6-8 April 2006.  
<http://medinfo.umft.ro/stc2006/>

**AMIA 2006 Spring Congress**  
Phoenix, AZ, USA. May 16-18, 2006  
<http://www.amia.org/meetings/s06/>

**APIII 2006: Frontiers in Oncology and Pathology Informatics**  
Vancouver, British Columbia | August 16 - 18, 2006  
<http://apiii.upmc.edu/>

**MIE 2006**  
Maastricht, NL. 27 – 30 August 2006.  
<http://www.mie2006.org>

**6th Nordic Conference on eHealth and Telemedicine “From Tools to Services”**  
Helsinki, Finland. 31 August – 1 September 2006.  
<http://www.congex.fi/nceht2006>

## COMMENT

Coinciding with the beginning of the third and last year of the NoE, we will initiate in the newsletter 4 several articles that confer significance to the title “Structuring European Biomedical Informatics to Support Individualised Healthcare”. One of the objectives is *to create a European BMI community that extends beyond the proposed core network to serve as an open forum for dialogue between the actors involved, to widely spread the knowledge acquired and developed in the framework of the network to the scientific community, healthcare professionals, citizens, industry, authorities and other stakeholders and to create a long-lasting, self-sustainable structure in the European BMI field.* To better achieve these objectives, newsletter four and sequent pretends to be an open forum for the dissemination of all BMI related initiatives in Europe. Providing a comprehensive and as complete as possible a view of this discipline, we will start a new section in our quarterly newsletter called “BMI in Europe” that will serve as a vehicle and showcase for all current relevant initiatives.

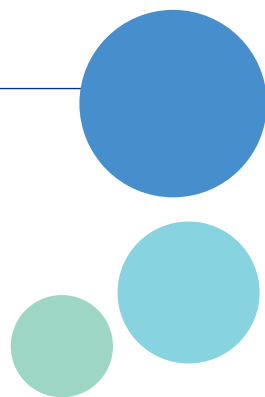
We will emphasize the dissemination of the different work currently carried out from the newly approved EC funded projects and the Networks of Excellence to the technological platforms among others. We will also include interviews with different experts both European and from the USA allowing us to provide the readers with a valid comparison between the development of this discipline in both areas of the world.

Our main aim is to give an overview of BMI in Europe giving the readers a general idea of each project and providing them with information to incite their interest for these initiatives. We will encourage them to look for more information when possible including web pages of reference for these initiatives.

In the next number we will give the big picture of BMI related initiatives in Europe while in later numbers we will briefly describe each of them. We, the dissemination and communication workpackage of the INFOBIOMED Network of Excellence, hope this new focus of our newsletters will be of interest to the scientific community and all other stakeholders interested in BMI.

*Infobiomed Editorial Board*

BIOMEDICAL INFORMATICS  
TO SUPPORT INDIVIDUALISED HEALTHCARE





BIOMEDICAL INFORMATICS TO SUPP

<http://www.infobiomed.org>

