5th Annual Bioinformatics Forum for SMEs

Piemonte Bioindustry Park, 6-7 October 2011
Welcome

13.30 Fabrizio Conicella
Workshop Overview

14.00 Dominic Clark
Purpose 1

• The EMBL-EBI is the outstation of EMBL with the primary responsibility to provide bioinformatics services to European researchers.

• Part of our Mission is:
  • To help disseminate cutting-edge technologies to industry

• we have established an annual series of information workshops designed to allow SMEs to hear about molecular and genomic databases and services that are provided freely, to understand the international data sharing agreements underlying these data resources and the road map for future developments.
Data from the 2010/2011 EBI Survey

- Total no. of respondents = 3,557
- “The EMBL-EBI User survey”
- Industry respondents made up 8% of this total:
  - SMEs = 156
  - Large commercial = 131
Which country?

Large

SME
Sector and nature of work

- SME
- Health
- Comm
- Acad

Dry Science
Wet Science
What is your role?
Which of the following EMBL-EBI resources is THE MOST IMPORTANT for your work?
What do our industry users do?

- retrieve specific database records (e.g. data that were mentioned in a paper)
- try to find data similar to my own (e.g. by sequence search)
- explore a scientific topic such as a disease (e.g. by keyword searching)
- derive information by analysing whole databases (e.g. gene regulatory networks)
- browse to see what data is new (e.g. look at the latest additions to Ensembl)
- other
Do you have bioinformatics support?

- **SME**
  - No: 0%
  - Yes: 40%
  - Yes, me: 60%

- **Health**
  - No: 20%
  - Yes: 80%
  - Yes, me: 0%

- **Commercial**
  - No: 40%
  - Yes: 60%
  - Yes, me: 0%

- **Academic**
  - No: 80%
  - Yes: 20%
  - Yes, me: 0%
How do you use the data?

- **SME**
  - Download: 20%
  - Programmatic: 10%
  - Website: 70%

- **Health**
  - Download: 20%
  - Programmatic: 10%
  - Website: 70%

- **Comm**
  - Download: 20%
  - Programmatic: 10%
  - Website: 70%

- **Acad**
  - Download: 20%
  - Programmatic: 10%
  - Website: 70%
SME Meeting have been held in a number of locations within Europe

• 2007: Cambridge
• 2008: Berlin.
• 2009: Vienna.
• 2010: Munich
• 2011: Piemonte
Purpose 2

• The workshop will provide the opportunity to introduce the tools and data resources that are available and the opportunities for SMEs to use and integrate these resources in order to derive business benefit from the existing and developing infrastructure.

• The topics of the workshop have been specifically selected on the basis of previous interactions with SMEs, priorities specified by SMEs as part of a recent user survey and discussions with a regional focus group.
Purpose 3

• The data resources and services include:
  • Genomics resources and Ensembl Genome Browser
  • Proteomics resources and UniProt
  • Chemical resources (with a focus on the ChEMBL bioactivity database)
  • EMBL-EBI Patent services and the non-redundant patent sequence database
  • Transcriptomics (ArrayExpress and the expression Atlas)
  • Integration services and EBI Search
  • Web services
Overview of the EBI data resources and services

14.10 Dominic Clark
What is EMBL-EBI?

- Europe’s main centre for bioinformatics services and research
- Non-profit organisation
- Based on the Wellcome Trust Genome Campus
- Part of the European Molecular Biology Laboratory
EMBL 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

Associate member state:
Australia
The five branches of EMBL

EMBL is a basic research institute funded by public research monies from 20 member states.

EMBL-EBI is the Bioinformatics Services Institute

Heidelberg

Basic research in molecular biology
Administration
EMBO

Hamburg

Structural biology

EBI Hinxton

Bioinformatics

Grenoble

Structural biology

Monterotondo

Mouse biology
EMBL-EBI’s mission statement

- To provide **freely available** data and bioinformatics **services** to all facets of the scientific community in ways that promote scientific progress
- To contribute to the advancement of biology through basic investigator-driven **research** in bioinformatics
- To provide advanced bioinformatics **training** to scientists at all levels, from PhD students to independent investigators
- To help disseminate cutting-edge technologies to **industry**
- To **coordinate** biological data provision across Europe
TODAY: EMBL-EBI’s Mission

Services

Disseminate new technologies to industry

Basic Research

Advanced Training
Supporting Industry

- Cosmetics
- Forestry
- Biotech
- Fisheries
- Big Pharma
- Computing
- Medical devices
- Agri-food
- Environmental protection
- Biofuels

EMBL-EBI
Key facts about services

• The world’s most comprehensive collection of molecular databases
• European node for globally coordinated data collection and dissemination projects
• Core databases produced in collaboration with other world leaders, including
  • Wellcome Trust Sanger Institute (UK)
  • NCBI and Cold Spring Harbor Laboratory (US)
  • National Institute of Genetics (Japan)
  • Swiss Institute of Bioinformatics (Switzerland)
What services do we provide?

Labs around the world send us their data and we...

...and provide user-friendly tools to enable the scientific community to benefit from it

A virtuous circle

Archive it

Classify it

Share it with other data providers

Analyse it
Bioinformatics underpins life-science research

1. Genomes contain genes
2. Genes are transcribed
3. Transcripts translate to protein sequences
4. Proteins form 3D structures
5. Proteins interact with each other and with small molecules to form pathways
6. Pathways combine to build systems
Databases: molecules to systems

Genomes
- Ensembl
- Ensembl Genomes
- EGA

Nucleotide sequence
- ENA

Functional genomics
- ArrayExpress
- Expression Atlas

Structures
- PDBe

Chemical entities
- ChEBI

Interactions & pathways
- IntAct
- Reactome

Protein sequences
- UniProt

Protein families + motifs
- InterPro

Protein expression
- PRIDE

Chemogenomics
- ChEMBL

Literature and ontologies
- CiteXplore
- UKPMC
- GO

Systems
- BioModels

EMBL-EBI
Principles of service provision

- **Accessibility** – all data and tools freely available without restriction
- **Compatibility** – we develop and promote the use of standards in bioinformatics
- **Comprehensive** – agreements with other data providers ensure that our resources contain comprehensive and up-to-date data; agreements with publishers ensure that published data are made public
- **Portability** – resources can be downloaded and installed locally
- **Quality** – Our databases are enhanced through annotation and cross-referencing
User training  www.ebi.ac.uk/training

Learn to make the most of Europe’s most widely used databases for life scientists

Train at EMBL-EBI
Gain hands-on experience in our state-of-the-art training room

Train at your place
Choose the training modules relevant to you and your colleagues and let us bring the Bioinformatics Roadshow* to you

Train online
Beta release launched: train wherever and whenever you like and at your own pace

* The Bioinformatics Roadshow is funded by the European Commission under SLING, grant agreement number 226073 (Integrating Activity) within the Research Infrastructures of the FP7 Capacities Specific Programme.
Training a diverse research community

The databases are useless unless our users know how to use them
Staff growth and nationalities

(a) Staff growth from 1998 to June 2010, and projected staff growth.
(b) Nationalities of EMBL-EBI members of personnel as of June 2010
EMBL-EBI staff, March 2011
EBI users on an ordinary day

IN 2010, EMBL-EBI’s website was visited by ~3.4 million unique IP
Standards development

- Genome annotation: www.geneontology.org
- Nucleotide sequence: www.insdc.org
- Functional Genomics Data Society: www.fged.org
- HUPO-Proteomics Standards Initiative (PSI): www.psidev.info/
- Protein sequence: www.uniprot.org
- Protein structure: www.wwpdb.org
- Cheminformatics: www.ebi.ac.uk/chebi
- Pathways: www.reactome.org, www.biopax.org
- Metabolomics Standards Initiative (MSI): www.metabolomicsociety.org
- Systems modelling standards: www.sbml.org

Genomics Standards Consortium (GSC): http://gensc.org

EMBL-EBI
<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.00</td>
<td>Registration and refreshments</td>
</tr>
<tr>
<td>13.30</td>
<td>Welcome from Bioindustry Park and BioPmed Cluster</td>
</tr>
<tr>
<td></td>
<td><strong>(Fabrizio Conicella, BioPmed)</strong></td>
</tr>
<tr>
<td>14.00</td>
<td>Workshop Overview <strong>(Dominic Clark, Industry Programme Manager, EMBL-EBI)</strong></td>
</tr>
<tr>
<td>14.10</td>
<td>Overview of EMBL-EBI data resources and services <strong>(Dominic Clark, EMBL-EBI)</strong></td>
</tr>
<tr>
<td>14.30</td>
<td>EMBL-EBI Integration Services: EBI Search <strong>(Jenny Cham, EMBL-EBI)</strong></td>
</tr>
<tr>
<td>15.00</td>
<td>ENSEMBL &amp; Genomics: High level overview and use cases <strong>(Jana Vandrovcova, ENSEMBL team, EMBL-EBI)</strong></td>
</tr>
<tr>
<td>15.30</td>
<td>COFFEE BREAK</td>
</tr>
<tr>
<td>Time</td>
<td>Session</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>16.00</td>
<td>UniProt &amp; Proteomics Services: High level overview and use cases (Michele Magrane, EMBL-EBI)</td>
</tr>
<tr>
<td>16.30</td>
<td>Transcriptomics and High Throughput Sequencing: High level overview and use cases (Gabriella Rustici, EBI)</td>
</tr>
<tr>
<td>17.00</td>
<td>Chemical Resources (Structure, Bioactivity and Metabolomics) Anne Hersey, EMBL-EBI: High level overview and use cases</td>
</tr>
<tr>
<td>17.30</td>
<td>Presentations from Regional Organisation: Introduced by Fabrizio Conicella, BioPmed.</td>
</tr>
<tr>
<td>17.35</td>
<td>Enrico Bucci, BioDigital Valley.</td>
</tr>
<tr>
<td>18.05</td>
<td>Francois Rechenmann, Genostar.</td>
</tr>
<tr>
<td>18.35</td>
<td>Closing comments, Fabrizio Conicella &amp; Dominic Clark</td>
</tr>
<tr>
<td>18.45</td>
<td>La Merenda Sinoira (“Wine and Cheese”) Networking Session, sponsored by BioPmed</td>
</tr>
</tbody>
</table>
## Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>09.00</td>
<td>Welcome Coffee</td>
</tr>
<tr>
<td>09.30</td>
<td>Introduction to the second day <em>(Dominic Clark, EMBL-EBI and Fabrizio Conicella, BioPmed)</em></td>
</tr>
<tr>
<td>09.40</td>
<td>EMBL-EBI Web Services <em>(Rodrigo Lopez, EMBL-EBI)</em> and worked examples</td>
</tr>
<tr>
<td>10.10</td>
<td>EMBL-EBI Patent services <em>(Jennifer McDowall, EMBL-EBI)</em> and worked examples</td>
</tr>
<tr>
<td>10.40</td>
<td>COFFEE</td>
</tr>
</tbody>
</table>
## Agenda

<table>
<thead>
<tr>
<th>11.10</th>
<th>EMBL-EBI Services Roadmap and strategic alliances <em>(Dominic Clark, EMBL-EBI)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>11.30</td>
<td>ELIXIR &amp; European Research Infrastructure Projects Introduction, <em>(Dominic Clark, EMBL-EBI)</em></td>
</tr>
<tr>
<td>11.40</td>
<td>Industry challenges and opportunities <em>(Bryn Williams-Jones, Connected Discovery representing the Pistoia Alliance)</em></td>
</tr>
<tr>
<td>12.10</td>
<td>Public tools for analysis &amp; Systems microscopy <em>(Bernd Fischer, EMBL)</em></td>
</tr>
<tr>
<td>12.40</td>
<td>The EuroBioImaging European Research Infrastructure Project <em>(Pamela Zolda, EIBIR – European Institute for Biomedical Imaging Research, Vienna)</em></td>
</tr>
<tr>
<td>13.10</td>
<td>Closing comments</td>
</tr>
<tr>
<td>13.10</td>
<td>Buffet Lunch</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>14.00</td>
<td><strong>Hands-on tutorials (bring your own laptop!) two parallel sessions:</strong></td>
</tr>
<tr>
<td></td>
<td>Proteomics resources and UniProt (Michele Magrane, EMBL-EBI) - Eridano Room</td>
</tr>
<tr>
<td></td>
<td>Transcriptomics (ArrayExpress &amp; the expression Atlas) (Gabriella Rustici, EMBL-EBI) - Salette Room</td>
</tr>
<tr>
<td>15.30</td>
<td>Final Departures</td>
</tr>
</tbody>
</table>
Introducing our people profiles for today…

**Francesca**
Target Validation
“My job involves cloning & expression analysis”

**Ola**
Discovery
“Our approach is to look into an interactome”

**Marc**
Discovery
“I’m interested in biomarkers”

**John**
Medicinal Chemistry
“I’m constantly looking for new information”
My job is **cloning and expression analysis**. The information I am interested in is published **cDNA sequences** and information on **orthologues**. I also need to know whether my targets of interest are over-expressed in a model organism for the disease I’m looking at; for example, in an inflammation model.

On a day-to-day basis, having a simple **way to easily search** across biological information would be useful. I’d like to be able to quickly find out what is known about a target, without having to forage around in lots of different websites.

**EBI resources useful for Francesca:**
- Ensembl
- Gene Expression Atlas
- EBI Search
EMBL-EBI Integration Services: EBI Search
Jenny Cham
Acknowledgements

Organisers
• Fabrizio Conicella
• Alberto Baldi
• Samantha Balma

EMBL-EBI
• Gabriella Rustici
• Holly Foster
• Mary Todd-Bergman

• All EMBL(-EBI) speakers: Jenny, Jennifer, Gabriella, Jana, Michele, Rodrigo, Anne, Bernd

• Regional Speakers: Enrico and Francois
• Pamela Zolda
• Bryn Williams-Jones