Biodigitalvalley experience:
integrating wet-lab with in-silico biomedical research
BioDigitalValley

External service
- Industrial
- Academic

Research projects
- ParlIS
- Image

Product
- ProteinQuest™

Core know-how
- Java
- Image analysis
- Biostatistics
Data sources:

a) PubMed/Biomed Central English abstracts
b) Figures/captions from biomedical papers: 4,525,203
c) USP patents: 1,461,000 from relevant classes

Dimension:

<table>
<thead>
<tr>
<th>Category</th>
<th>Objects</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteins</td>
<td>22597</td>
<td>269933</td>
</tr>
<tr>
<td>Substances</td>
<td>41379</td>
<td>99488</td>
</tr>
<tr>
<td>Bioprocesses</td>
<td>19798</td>
<td>56355</td>
</tr>
<tr>
<td>Anatomy</td>
<td>1476</td>
<td>9841</td>
</tr>
<tr>
<td>Disease</td>
<td>4408</td>
<td>41818</td>
</tr>
<tr>
<td>Drugs</td>
<td>1795</td>
<td>15704</td>
</tr>
<tr>
<td><strong>TOTAL (BIO-OBJECTS)</strong></td>
<td><strong>91453</strong></td>
<td><strong>493139</strong></td>
</tr>
<tr>
<td>Authors (papers + patents)</td>
<td>4740735 + 1660704</td>
<td></td>
</tr>
<tr>
<td>Institutions</td>
<td></td>
<td>9224</td>
</tr>
</tbody>
</table>

12.5 billions of binary relationships between objects
Data mining

- Prioritize genes/proteins according to your needs
- Find out which protein is involved in a disease and what are the available drugs
- Separate clinical results according to development phase
- Look for experimental support to your data
Data analysis

- Get your own protein networks
- Build heat-maps and find out protein/drug/pathology relationships
- Examine scientific collaboration networks
Today we will discuss the following use cases:

• **Aligning different experiments and getting a meta-analysis**
• **Network analysis**
Integrating wet-lab biology with in-silico research: aligning different experiments and get a meta-analysis
PQ query: DJ1 AND 2D gel

CROPPING

Gel images from literature
Image scaling and registration on a reference pl/MW coordinate system

Background subtraction * and binarization **

Production of a metagel summing up all the images

Spot segmentation on the metagel

Matching of the metagel spots on every original image

Classification of every original image on the basis of the presence/absence of metagel spots
DJ-1 MetaGel (161 experiments considered)
<table>
<thead>
<tr>
<th>d5</th>
<th>spot6</th>
<th>spot7</th>
<th>spot8</th>
<th>spot9</th>
<th>spot10</th>
<th>spot11</th>
<th>spot12</th>
<th>spot13</th>
<th>spot14</th>
<th>spot15</th>
<th>spot16</th>
<th>Paper</th>
<th>Tissue</th>
<th>condition</th>
<th>technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Massimo Natale, Dario Bonino, Paolo Consoli, Tiziana Alberio, Rivka G Ravid, Mauro Fasano, and Enrico M Bucci, Bioinformatics, advance access published on February 19, 2010
Spot co-occurrence analysis defines two main groups of spots, which are different for MW and average pI.
Similarity-based cluster analysis on the 161 considered 2D-gels
Analysis of DJ-1 forms in human brain.

Panels A and B represent metagels obtained for control and patients as documented in the considered literature.

Panels C and D represent the metagels obtained from experiments on 3 control individuals and 3 PD patients.
After validating the method: the ParIS collaboration

- 50 Parkinsonians, 50 neurological patients as controls
- 4 research groups (Aosta Neurology Unit, Insubria University, Mondino Foundation, Biodigitalvalley)
- 2 years of activity
- More than 150 bidimensional gels (among the largest proteomic study of Parkinson’s disease)
- **Final result expected: validation/rejection of all the peripheral biomarkers proposed so far for Parkinson’s disease**
The platform (http://paris.biodigitalvalley.com)

Clinical (1), biochemical (2) and proteomic data (3)

Flexible biostatistics
There is more than 2D gels ...

Drec northern blot

DJ-1 northern blot

Reference slice
Final result: a virtual double-staining experiment
MRI of the head
Reconstructed model
Brain from a dementia patient (FTG imaging)
Orthogonal views

3D projection
Integrating wet-lab biology with in-silico research:
network analysis
An archetypal story: p53

• 1979: discovery of the gene by Lane and Levine
• Late 1980s: Vogelstein recognized that p53 is a tumor suppressor gene, which is frequently mutated in cancer
• 1993: p53 is the “molecule of the year” (Science)
• 1999: more than 15,000 papers mentioning p53
• Lane, Levine and Vogelstein, Nature, 2000: “An appreciation of the existence and complexity of cellular networks should enable more rational design and interpretation of experiments in the future, and should allow more realistic approaches to treatment. After all, the most important question in p53 research is: how do we attack a cellular network that is already compromised by inactivation of one of its most highly connected nodes?”
What caused this change of perspective?

• In 1998, Oltvai, a cell biologist, and Barabasi, a physicist which was studying the structure of internet, were home neighbors in Chicago

• At the time, Barabasi had already shown that internet is a non-random network, and that its connectivity structure influences its function

• One year later, in 1999, they proved that the metabolic pathways of yeast define a network whose structure is very similar to that of internet.
The cell metabolic network vs internet

Each cell looks like a tiny web, extremely uneven, with a few molecules involved in the majority of reactions – like the internet main servers and routers are involved in the majority of the web transactions – while most of the molecules are involved only in few reactions – like the client machines in internet.
Why p53 is so important? Co-occurrence analysis

A network made of proteins which are each involved in at least 20 different cancer types.
p53 is connected to many nodes of the network. It is a network hub (like internet servers)
Can we target hubs in cancer protein networks?


Hubs hierarchy is mirrored by clinical development stage
Biases from Big Pharma on data are very evident!

Bias in centrality measures
(more publications = more experiments = higher degree = higher centrality)

Integrating wet-lab biology with in-silico research: final remarks
• Mining the literature and integrating with other data sources in a creative and flexible way can provide insights which are valuable in terms of money spared from experiments – there is more than PubMed!

• Figures published in the scientific literature provide very relevant information.

• Network analysis seems to be quite appreciated from our customers.
Aknowledgements

BdV research team:
D.Bonino          M.Giordano          L.Montagnoli
A.Ruffino         M.Gullusci         M.Cornaz
A.Poli            M.Natale           J.Guidone

Academic collaborations:
BiodigitalValley srl
Pèpinière d'Entreprises, via Carlo Viola 78
11026 Pont Saint Martin (AO) - Italy
info@biodigitalvalley.com
www.biodigitalvalley.com