NCIBI Education and Outreach
## NCIBI Graduate Students

<table>
<thead>
<tr>
<th>Student</th>
<th>Mentor</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jing Zhang</td>
<td>H.V. Jagadish</td>
<td>Provenance of data in MiMI</td>
</tr>
<tr>
<td>Anna Shevardian</td>
<td>H.V. Jagadish</td>
<td>Interaction graphs to visualize and study protein interaction networks</td>
</tr>
<tr>
<td>Abhik Shah</td>
<td>Peter Woolf</td>
<td>Integrating existing knowledge and experimental data using Bayesian Networks</td>
</tr>
<tr>
<td>Sira Sarntivija</td>
<td>Brian Athey</td>
<td>Ontology Mapping; Cell Line Ontology</td>
</tr>
<tr>
<td>Arzucan Ozgur</td>
<td>Drago Radev</td>
<td>NLP Methods for information extraction</td>
</tr>
<tr>
<td>Nerit Glazer</td>
<td>Barbara Mirel</td>
<td>Evaluation and Assessment of User Heuristics</td>
</tr>
<tr>
<td>Gang Su</td>
<td>Fang Meng</td>
<td>Visualization and Data integration for protein-protein and metabolite interactions</td>
</tr>
</tbody>
</table>
NCBC Collaboration

• Big “P” in program
• 3 working groups across NCBCs
  – Ontologies
  – Yellow pages → BIOsitemaps
  – DBP Impact  http://www.ncbcs.org/biositemaps/
• www.ncbcs.org
Shared Events Calendar

NIH Roadmap
National Centers for Biomedical Computing

Home | NCBC Summary | Calendar | 2008 All Hands Meeting | 2006 AHM | SDIWG | Tools and Applications | Ontology Working Group | DBP Interactions

Today | Thursday, April 16 | Print | Week | Month | Agenda

Thursday, April 16
12:00pm  NCBI Tools & Technology Seminar (Webcast Live)

Wednesday, April 22
4:00pm  MAGNet Seminar: "Oncogenic Transcriptional Networks Controlled by NOTCH1 in Human Leukemia"

6:30pm  Simbios Science Seminar: Greg Bowman -- MSMBuilder: Open Source Software for Describing Dynamics with Maps

Tuesday, April 28
NCBI: Fourth Annual NCBI Research Meeting

Wednesday, April 29
NCBI: Fourth Annual NCBI Research Meeting

Thursday, April 30
12:00pm  NCBI Tools & Technology Seminar - Yang Zhang (Webcast Live)

Tuesday, May 5
i2b2: Academic Users' Group Meeting

Events shown in time zone: Eastern Time

NIH Roadmap Initiative
For questions/comments about the NCBC Software and Data Integration Efforts contact:
CCB | i2b2 | NA-MIC | Simbios | NCBO | NCBI | MAGNet

RSS feed and iCal files for individual NCBCs
NCIBI 2008-2009 Pilot Projects

• Open call for pilot projects, but with preference for comorbidity of substance abuse and bipolar disorder

• Selected collaborative project looking at Bipolar Disorder with co-morbid substance abuse among African American women

• Three subprojects:
  – *Genes for Bipolar Disorder among African Americans*, Sharon Lewis, Langston University
  – *Estrogen and Substance Abuse*, Tonya Gerald, North Carolina Central University
  – *Text mining*, Raphael Isokpehi, Wellington Ayensu, Jackson State University
NCIBI/RCMI Workshop on Translational Bioinformatics
July 29-30, 2009

• Provide RCMI investigators with training in the use of NCIBI tools and resources for translational research
• Build research relationships with the goal of expanding informatics resources for the study of health disparities
• Introductory hands-on tutorial with the NCIBI tools, followed by sessions tentatively focused on RTRN translational research clusters that overlap with projects in cancer, diabetes and nephropathy, and bipolar disorder
• Estimated number of attendees ~20
• Identify projects for coming year to expand outreach for NCIBI
Other Current and Future Collaborations

• Building Bridges Post-doc
  – Rich McEachin – Bridge with Laura Beirut Lab at WUSTL
  – Zuofeng Li (Hong Yu, UW Milwaukee) –

• I2B2/NCIBI – Tool integration with HIVE
• GenePattern (Broad Institute)
• Institute for Systems Biology (Seattle)

• Kay Robbins (UT San Antonio)
• Youping Deng (U. Southern MS), and Jack Yang (Harvard)
Tools and Technologies Seminars

- Regular weekly seminars presenting novel development, algorithms and technologies.
- Broadcast over Adobe Connect for all to attend
- Video archived and available for review

http://portal.ncibi.org/gateway/tt.html
Meetings and Demonstrations

- ISMB 2008 – Toronto – Technology Track demonstration
- AMIA Translational Bioinformatics Summit
- Jackson State University - Demonstration
- Yesterday’s NCIBI Workshop!
- Upcoming Tutorials with UM HSL
- Upcoming ISMB meeting (July 2009)
Dissemination of Software Tools
5320 total visits from 63 Countries
41% new visits
59% returning
~ 5.5 pages per visit
## Tool Usage Statistics

<table>
<thead>
<tr>
<th>Tool</th>
<th>total hits</th>
<th>unique</th>
<th>returning</th>
<th>% unique</th>
<th>% returning</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiMI</td>
<td>6238</td>
<td>2521</td>
<td>3717</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Gene2Mesh</td>
<td>2591</td>
<td>1468</td>
<td>1123</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>Metab2Mesh</td>
<td>4501</td>
<td>4080</td>
<td>421</td>
<td>91</td>
<td>9</td>
</tr>
<tr>
<td>MiSearch</td>
<td>1572</td>
<td>872</td>
<td>700</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>Biosearch2D</td>
<td>376</td>
<td>230</td>
<td>146</td>
<td>61</td>
<td>39</td>
</tr>
</tbody>
</table>

### % Total tool hits

- biosearch2D
- Gene2Mesh
- Metab2Mesh
- MiMI
- MiSearch

### % Returning tool hits

- biosearch2D
- Gene2Mesh
- Metab2Mesh
- MiMI
- MiSearch
Cytoscape Plugin Downloads

MiMI Plugin

MetScape Plugin

8/31/07 10/8/07 4/14/08 7/1/08 8/12/08

520 828 890 458 1017

2.0 2.1 2.2 3.0 3.01

0.8

3/15/09

159

4/15/09

present
MiMI Web and Plugin Compared

- Two main routes of entry to get at MiMI data
  - Plug-in via Cytoscape
  - MiMI Website

<table>
<thead>
<tr>
<th></th>
<th>Total Visits</th>
<th>Unique Visits</th>
<th>Returning Visits per IP address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
<td>2009</td>
<td>fold change</td>
</tr>
<tr>
<td>MiMI Web</td>
<td>3990</td>
<td>6238</td>
<td>2</td>
</tr>
<tr>
<td>MiMI Plugin</td>
<td>4144</td>
<td>8515</td>
<td>2</td>
</tr>
</tbody>
</table>
NCIBI Core 4

• Role of Core 4 is twofold
  – Provide reliable, consistent and high quality IT infrastructure for all NCIBI projects and collaborations
  – Explore and implement cutting edge computational capacity and technology for research

• Computer Infrastructure
  – Compute intense (RAM and CPU) servers
  – Virtual Servers
  – Storage
  – Relational Databases
  – Backup / Recovery
Network Overview
Compute Servers

• Server / Compute cycles
  – Servers with large RAM provide the bulk of CPU resources for Memory intensive NLP projects
  – Compute servers for databases
    • Standard server configuration: dual quad-core + 64 GB RAM per server (we started with 8, then 32)

• DC-NCIBI Cluster
  – 152 Opteron cores on 32 computing nodes
  – 40T Unified cluster storage
  – 12T Oracle Grid with 12 Opteron cores and 64 G
  – Time machine backup of critical files
  – Green computing: computing nodes turned off when not used

• GPU servers
  – Using Graphics Processing Units for algorithms amenable to Vector calculation
  – Enormous speed up from parallel pipe architecture of the GPU.
Virtual Servers and Storage

• Virtual Servers
  – Servers to support less intense needs; Web servers
  – Reduces energy / cooling costs by sharing hardware needs
  – Easily reconfigured and adapted to changing needs without additional purchase costs

• Storage
  – Current storage is built on ‘fast’ and ‘slow’ disks
  – Expansion of storage is a regular and essential part of NCIBI
  – Current capacity is ~ 120 TB (raw) (we started with 24)
  – Configured for 2 site mirroring of critical data (first level backup)
  – Tape storage for offline backup of selected data
Databases and Backup

• Relational Databases
  – Databases are essential core function for NCIBI
  – SQLServer and Oracle are primary RDBMS
    • Prototype, Test and Production
  – Tuning and security assessment and performance testing

• Backup Recovery Planning
  – Servers setup for snapshot recovery (immediate retrieval of deleted files)
  – RAID 5, 6, 10 disk arrays (protects against HD failure)
  – Battery backup, monitoring for status, temperature ** (power loss, AC loss control)
  – Offsite Mirror of critical data (protects against single site disaster (fire, sprinkler, etc)
  – Tape Backup for selected data files (long term recovery / restoration)
Data Currency

• Data Warehouse scripts automatically download over 20 biology/bioinformatics data sources totaling over 1.5 terabytes.
  – available to all of the compute servers and therefore all of the NCIBI researchers.

• Nightly Download of PubMed
  – Automatic replication of the nightly downloaded and weekly parsed PubMed data.
Thank you!

- Jerome Kinlaw
- LihShwu Ke
- Vasile Negrea
- Bruce Taylor
- Kurt Seyfried
- Aaron Bookvich
- Zach Wright
- Jim Rees
- Alex Terzian