MLSCN @ CU & OME

Molecular Library Screening Center Network at Columbia University and the Open Microscopy Environment

Bernd Jagla, 4/22/2006
MLSCN

- This is a nationwide consortium of small molecule screening centers that has been recently funded to produce innovative chemical tools for use in biological research.
- The MLSCN performs HTS on assays provided by the research community, against a large library of small molecules maintained in a central molecule repository.
- The MLSCN has established a collection of 100,000 chemically diverse small molecules.
- All of the results from the MLSCN’s activities will be placed into a public database called PubChem.
Columbia MLSC

• The Columbia Center in the MLSCN focuses on cell biology, high content/high resolution automated cellular imaging and image analysis, and phenotypic assay design and implementation.

• Primary screening will be done with high content assays at HTS performed in intact cells.

• Assay development will initially focus on establishing a repertoire of >50 assays providing broad coverage of signaling pathways, and associated bioinformatics tools.

• Profiling of hits and leads against this repertoire of biology will provide important information on specificity at the biological level to complement information on the compound's selectivity at the protein/target level.
Who is involved?

- James E. Rothman (PI)
- Lars Branden (Project manager)
- Thomas Mayer (assay dev.)
- M. Beard (assay dev.)
- Deby Smith (assay dev.)
- Effie Tzilianos (assay dev.)
- Feng-Li Zhang (assay dev.)
- Mike Wyler (assay dev.)
- Nathalie Aulner (HTS)
- Udo Többen (HTS)
- Bernd Jagla (IT)
- Geoff Barger (automation)
- Martine Lecorps (secretary)
- Collaborator: Ai Yamamoto
- OME:
  - Ilya Goldberg
  - Harry Hochheiser
  - Josiah Johnston
  - Jason Swedlow
- Partek
  - DJ Meyer
  - Michael J. Venezia
- GE
  - Rick Maguire
  - Binayak Roy
  - Sarang Parnaik
  - Marcin Swiatek
Biological profiling

Assays

Cell types

Metabolism

Gene expression levels

Response to stimuli

compound
"Three Dimensional" Biological Profiling
HT screening work-flow

1. Primary HT-compound screen
2. Confirmatory screen
3. Biological Profiling
   - Biological Specificity
   - Toxicology
4. Hit Characterization and Selection
   - Time course
   - Dose response
   - Chemistry
5. Knowledge Base

NIH Compound Repositories

Implemented Assay

Assay information/results Acquisition

Scientific Community

PubChem

Referring Investigator

Columbia Center Scientist

Sister MLSCN Centers

FTP
Cell-based “high content” functional assays

- Automated confocal microscopy
- Up to 30,000 wells per day
- Simultaneous 3 color detection
- Quantitative data analysis
- Three years’ experience using prototype of GE INCell 3000 Analyzer
Biological Profiling: Specificity and Mechanism

### Protein Binding Profiling

<table>
<thead>
<tr>
<th>Protein Binding Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRBK1</td>
</tr>
<tr>
<td>Akt1</td>
</tr>
<tr>
<td>Akt2</td>
</tr>
<tr>
<td>Akt3</td>
</tr>
<tr>
<td>CaMKI d</td>
</tr>
<tr>
<td>CaMKII d</td>
</tr>
<tr>
<td>CaMKII</td>
</tr>
<tr>
<td>CaMKIV</td>
</tr>
<tr>
<td>CDK1/cyclin B</td>
</tr>
<tr>
<td>CDK2/cyclin A</td>
</tr>
<tr>
<td>CDK5/p35</td>
</tr>
<tr>
<td>CHEK1</td>
</tr>
<tr>
<td>CHEK2</td>
</tr>
<tr>
<td>CLK1</td>
</tr>
<tr>
<td>CSNK1D</td>
</tr>
<tr>
<td>CSNK1G2</td>
</tr>
<tr>
<td>CSNK2A1</td>
</tr>
<tr>
<td>CSNK2A2</td>
</tr>
<tr>
<td>DAPK2</td>
</tr>
<tr>
<td>IRAK4</td>
</tr>
<tr>
<td>MAP2K1</td>
</tr>
<tr>
<td>MAP2K2</td>
</tr>
<tr>
<td>MAPK3</td>
</tr>
<tr>
<td>MAPKAP-K5</td>
</tr>
<tr>
<td>PKA</td>
</tr>
<tr>
<td>PIM2</td>
</tr>
<tr>
<td>PRKD2</td>
</tr>
<tr>
<td>ROCK1</td>
</tr>
<tr>
<td>RPS6KA1</td>
</tr>
<tr>
<td>TBK1</td>
</tr>
</tbody>
</table>

### Biological Process/Pathway Profiling

#### Biological Processes (20)
- Apoptosis
- Inflammation
- Cell cycle
- DNA repair
- Secretion
- Transcription
- Etc etc

#### Pathways (40)
- Receptor activation
- Kinase activation
- Transcriptional activation
- Intracellular signalling
- Etc etc

---

Compound X →
Gene-Plus collection (960 compounds) screen

Control

TNFα stimulation (positive control)

Image Analysis
Dimensions of the profiling matrix

- Genome/Proteome
- IC₅₀
- Assays
- Cell type

Known compound vs. unknown compound → New Knowledge

Compound ID/fingerprint → Insight into mechanism of action
Tasks modularized

Automation equipment

INCA

PubChem

Public access

Internal Knowledge Base (OME)

 QC & Statistical analysis
 Partek, Pipeline Pilot, ...

 Off-line Image analysis
 Matlab Developer ImageJ

 Cluster analysis
 Knowledge discovery
 Partek Matlab, TBD

External DBs

SRS
PubChem
BIND
Results

• Run files from INCA can be read
  – Original format can be read in and exported files (tiff, xml)
  – Plate – image relationship is transferred to OME

• Frm (image) files from INCA can be
  – Decompressed
  – Read into OME

• Analysis files
  – Mapped to corresponding runs/plates/images
  – Values are imported into OME
Image view
shoola
Statistical analysis of screening results

• Partek – Screeners solution
  – Statistics software for analyzing one screen

• Pipeline Pilot
  – Workflow management system for
    • statistical analysis
    • integration with chemical compound analysis products
    • Standardized analysis workflows
How is the link established

• set url
Partek Screeners solution

- Check for edge effects
- Z’
- Normalization tools
- Cluster of results
- Principal component analysis
- Clustering tools
- Chemical compound visualisation
PCA view
Compound view
Edge effects

<table>
<thead>
<tr>
<th>Row</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Std. Err.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4032</td>
<td>17008.385855**</td>
<td>952.721384</td>
<td>15.003951</td>
<td>13508.908780</td>
<td>18859.982579</td>
</tr>
<tr>
<td>B</td>
<td>4032</td>
<td>17002.861979</td>
<td>950.744700</td>
<td>14.972822</td>
<td>13453.703540</td>
<td>18910.109857</td>
</tr>
<tr>
<td>C</td>
<td>4032</td>
<td>16976.129457</td>
<td>949.139649</td>
<td>14.947544</td>
<td>13351.936082</td>
<td>18659.802287</td>
</tr>
<tr>
<td>D</td>
<td>4032</td>
<td>16982.392165</td>
<td>947.812605</td>
<td>14.926645</td>
<td>13425.857000</td>
<td>19164.577546</td>
</tr>
<tr>
<td>E</td>
<td>4032</td>
<td>16969.129733</td>
<td>951.581159</td>
<td>14.985995</td>
<td>13382.473094</td>
<td>18671.552071</td>
</tr>
<tr>
<td>F</td>
<td>4032</td>
<td>16978.732403</td>
<td>952.962357</td>
<td>15.007179</td>
<td>13316.972142</td>
<td>18762.027165</td>
</tr>
<tr>
<td>G</td>
<td>4032</td>
<td>16966.322348</td>
<td>967.088109</td>
<td>15.230206</td>
<td>13491.011643</td>
<td>18812.174705</td>
</tr>
<tr>
<td>H</td>
<td>4032</td>
<td>16973.755606</td>
<td>950.754805</td>
<td>14.972981</td>
<td>13224.462682</td>
<td>18781.418374</td>
</tr>
<tr>
<td>I</td>
<td>4032</td>
<td>16980.114425</td>
<td>1266.709488</td>
<td>19.948799</td>
<td>12335.452461</td>
<td>19452.331299</td>
</tr>
<tr>
<td>J</td>
<td>4032</td>
<td>16958.295326</td>
<td>1283.762089</td>
<td>20.217353</td>
<td>12426.390767</td>
<td>19155.462848</td>
</tr>
<tr>
<td>K</td>
<td>4032</td>
<td>16975.580536</td>
<td>1273.223120</td>
<td>20.051379</td>
<td>12309.357032</td>
<td>19358.999191</td>
</tr>
<tr>
<td>L</td>
<td>4032</td>
<td>16950.061808</td>
<td>1270.761199</td>
<td>20.012608</td>
<td>12357.286023</td>
<td>19427.048768</td>
</tr>
<tr>
<td>M</td>
<td>4032</td>
<td>16970.115544</td>
<td>1274.358901</td>
<td>20.069266</td>
<td>12074.821169</td>
<td>19144.761504</td>
</tr>
<tr>
<td>N</td>
<td>4032</td>
<td>16948.031374</td>
<td>1273.829003</td>
<td>20.069921</td>
<td>12347.809253</td>
<td>19276.677213</td>
</tr>
<tr>
<td>O</td>
<td>4032</td>
<td>16931.949904*</td>
<td>1276.129969</td>
<td>20.097158</td>
<td>12386.094982</td>
<td>19211.724266</td>
</tr>
<tr>
<td>P</td>
<td>4032</td>
<td>16963.965335</td>
<td>1262.880281</td>
<td>19.888495</td>
<td>12420.256675</td>
<td>19146.690616</td>
</tr>
</tbody>
</table>

Summary:
* Row O has the smallest mean (16931.949904)
** Row A has the largest mean (17008.385855)
Row A is 0.45% larger than Row O
p-value: 0.214377 — No significant edge effect has been detected
INCell Information
Instrument Annotations

ExcitationRecord
- Instr_id
- index
- excitationID
- pass
- exitationName
- NDName

ImageFormatRecord
- Instr_id
- binning
- horizontalPixels
- verticalPixels
- scanWidth
- scanLength

CameraRecord
- Instr_id
- index
- status
- acquisitionFFMode
- pass
- channel
- filterName

ScanRecord
- Instr_id
- firstFrameChannel
- nextFrameChannel
- numberPassesFirstFrame
- numberPassesNextFrame
- integrationTime
- frameType
- maximumFramesPerWell
- cellCountCheck
- plateType
- columnsPerGroup
- interWellDelay
- cellCountThreshold
- interFrameDelay

bufferToCameraMap
- scanRec_id
- index
- value

EnvironmentalRecord
- Instr_id
- temperature
- humidity
- cO2

DispenserRecord
- Instr_id
- status
- spitOrder
- volume
- flowRate
- atFrame
- delay

TransmissionRecord
- Inst_id
- slitWidth
- cameraIndex

ConfocalRecord
- Instr_id
- Objective

AutofocusRecord
- Instr_id
- autofocusOffset
- stackZ0
- stackDeltaZ