NIH Bioassay Research Database (BARD)

AJIT JADHAV
NIH/NCATS
SEPTEMBER 25, 2013
Presentation Highlights

• Introduction to NCATS
  ➢ ‘Open Source’ Drug Discovery
• Motivations for BARD
• Interacting with BARD
• Extensibility & Community
The Best of Times, The Worst of Times

Fundamental science unprecedentedly advanced, but:

- Poor transition of basic or clinical observations into interventions that tangibly improve human health
- Drug/device/diagnostic development system in crisis
- Clinical trials system in crisis
- Poor adoption of demonstrably useful interventions

People unhealthier and funders of biomedical research enterprise (public and private) impatient
NCATS Mission

To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.
Catalyzing Collaborations Within NIH
Catalyzing Collaborations Outside NIH

- Complements — does not compete — with the work of others
- Revolutionizes the process of translation by promoting innovative research
- Galvanizes and supports new partnerships
- Supports and augments regulatory science and its application
- Expands the precompetitive space
NCATS Programs and Initiatives

Clinical and Translational Science Activities
• Clinical and Translational Science Awards

Rare Diseases Research and Therapeutics
• Therapeutics for Rare and Neglected Diseases
• Bridging Interventional Development Gaps
• Office of Rare Diseases Research

Re-engineering Translational Sciences
• NIH Chemical Genomics Center
• Toxicology in the 21st Century
Creating a Human Genome Translation Toolbox

Transcriptome Reference Sets (MRT Project)

Small molecules (MLP)

cDNA collection (MGC)

siRNAs

siRNA

m7G

AAAAA...

RNA degradation

KO mice genome-wide (KOMP)
The “Non-Druggable” Genome Problem

Drug Target Classes

- GPCRs
- Nuclear Receptors
- Enzymes
- Others
- Ion Channels
- Unknown

46% (n=483)

Human Genome

- Receptor (1543, 5%)
- Kinase (868, 2.8%)
- Select regulatory molecule (988, 3.2%)
- Transferase (610, 2.0%)
- Synthase and synthetase (313, 1.0%)
- Oxidoreductase (656, 2.1%)
- Lyase (117, 0.4%)
- Ligase (56, 0.2%)
- Isomerase (163, 0.5%)
- Hydrolase (1227, 4.0%)
- Chaperone (159, 0.5%)
- Cytoskeletal structural protein (876, 2.8%)
- Extracellular matrix (437, 1.4%)
- Immunoglobulin (264, 0.9%)
- Ion channel (406, 1.3%)
- Motor (376, 1.2%)
- Structural protein of muscle (296, 1.0%)
- Protein sensor (902, 2.9%)
- Select calcium binding protein (34, 0.1%)
- Intracellular transporter (350, 1.1%)
- Transporter (533, 1.7%)
- Molecular function unknown (12809, 41.7%)


Venter et al., (2001) Science 291:1304
Molecular Libraries Program Enabled by Convergent Developments

- Human Genome Project
- Availability of targets
- HTOS
- Availability of compounds
- Compound Libraries
- Availability of screening, ADME
- Robotic Technology
- Assay Technology
- “Open Source” Chemical Biology and Drug Discovery
NIH Chemical Genomics Center

• Obligatory collaboration model
• Currently > 230 collaborations with investigators worldwide
• Assay development, HTS, chemical informatics, medicinal chemistry: “target to lead”
• Focus is unprecedented targets, rare/neglected diseases
• Mission
  ➢ Chemical and siRNA probes/leads
  ➢ New technologies/paradigms to improve efficiency and success rates of target-to-lead stage of drug development
  ➢ Chemical genomics: general principles of siRNA action, small molecule - target interactions
NCATS’ Quantitative HTS (qHTS)

A. Assay concentration ranges over 4 logs (high: ~60 μM)
1536-well plates, inter-plate dilution series
Assay volumes 5 μL

B. Automated concentration-response data collection

C. Automated curve fitting and classification

Class 1
Class 2
Class 3
Class 4

D. Sample ID | R₁ | R₂ | R₃ | n  | Curve Class | EC₅₀ (nM) | Hill Coef
---|---|---|---|----|-------------|---------|---------
NGCC00033801 | H  | H  | H  | 1  | 1.1         | 0.251   | 1.4     
MLSS003223801 | H  | H  | H  | 2  | 1.1         | 0.356   | 1.1     
MLSS003283001 | H  | H  | H  | 1  | 1.1         | 2.51    | 1.0     
MLSS003238001 | H  | H  | H  | 2  | 1.1         | 3.16    | 1.1     
NGCC0002898801 | H  | H  | H  | 1  | 1.2         | 0.050   | 2.2     
MLSS0001206401 | H  | H  | H  | 2  | 2.1         | 10      | 1.8     
NGCC0006940301 | H  | H  | H  | 2  | 2.1         | 12.89   | 1.0     
MLSS0001129601 | H  | H  | H  | 1  | 2.2         | 15.89   | 0.9     
NGCC0004916001 | H  | H  | H  | 2  | 2.2         | 15.89   | 0.9     

P450-cyta2
Comprehensive Drug Repurposing Library

The NCGC Pharmaceutical Collection: A Comprehensive Resource of Clinically Approved Drugs Enabling Repurposing and Chemical Genomics

Ruili Huang,* Noel Southall,* Yuhong Wang, Adam Yasgar, Paul Shinn, Ajit Jadhav, Dac-Trung Nguyen, Christopher P. Austin†


• Chronic Lymphocytic Leukemia – IND filing 2011
• Niemann-Pick C – IND filing 2012
• Other rare disease repurposing projects in pipeline
• Collaboration with Lilly Open Innovation Drug Discovery

~3500 drugs
Drug Repurposing Library - “NPC Browser”

http://tripod.nih.gov/npc/
Example Probes from NCATS

**ML165/RUC2**
- Collaborator: Barry Coller Lab (Rockefeller University)
- Mechanism: potent inhibitor of platelet $\alpha$IIb$\beta$3 receptor
- Indication: myocardial infarction
- Status: IND filing on advanced derivative expected by early 2014

**ML290**
- Collaborator: Alexander Agoulnik (Florida International University)
- Mechanism: potent and selective agonist of relaxin receptor RXFP1
- Indication: chronic heart failure
- Status: Will be developed in collaboration with a pharma company
Example Probes from NCATS

• **ML365 and ML285**
  - Collaborator: Lew Cantley (Harvard) and Matt Vander Heiden (MIT)
  - Mechanism: potent activators of PKM2
  - Indication: various cancers
  - Status: collaborating with public and private organizations

• **ML323**
  - Collaborator: Zhihao Zhuang (University of Delaware)
  - Mechanism: potent and selective inhibitor of USP1/UAF1
  - Indication: various cancers
  - Status: *in vivo* POC stage, to be further developed
  - Publication: *Nature Chemical Biology*, under review.

Structure released
Dec. 2013
**ML323**
The BioAssay Research Database

- Originated from the NIH Molecular Libraries Program
- Motivated to make the bioassay data generated by the MLP more accessible and amenable to exploration and hypothesis generation
- Joint effort between NCATS, UNM, U. Miami, Broad, Vanderbilt, Burnham, Scripps
BARD Collaborators

NIH Molecular Libraries Program – Glenn McFadden, Ajay Pillai

NIH Chemical Genomics Center – Chris Austin (PI), John Braisted, Rajarshi Guha, Ajit Jadhav, Trung Nguyen, Tyler Peryea, Noel Southall, Cordelle Tanega

Broad Institute – Benjamin Alexander, Jacob Asiedu, Kay Aubrey, Joshua Bittker, Steve Brudz, Simon Chatwin, Paul Clemons, Vlado Dancik, Siva Dandapani, Andrea de Souza, Dan Durkin, David Lahr, Jeri Levine, Judy McGloughlin, Phil Montgomery, Jose Perez, Stuart Schreiber (PI), Gil Walzer, Xiaorong Xiang

University of New Mexico – Cristian Bologa, Steve Mathias, Tudor Oprea, Larry Sklar (PI), Oleg Ursu, Anna Waller, Jeremy Yang

University of Miami – Saminda Abeyruwan, Hande Küküc, Vance Lemmon, Ahsan Mir, Magdalena Przydzial, Kunie Sakurai, Stephan Schürer, Uma Vempati, Ubbo Visser

Vanderbilt University – Eric Dawson, Bill Graham, Craig Lindsley (PI), Shaun Stauffer

Sanford-Burnham Medical Research Institute – “T.C.” Chung, Jena Diwan, Michael Hedrick, Gavin Magnuson, Siobhan Malany, Ian Pass, Anthony Pinkerton, Derek Stonich, John Reed (PI)

Scripps Research Institute – Yasel Cruz, Mark Southern, Hugh Rosen (PI)
Goals of BARD

BARD’s mission is to enable novice and expert scientists to effectively utilize MLP and other public data to generate new hypotheses

- Developed as an open-source, industrial-strength platform to support public translational research
- Foster new methods to interpret & analyze chemical biology data
- Develop and adopt an Assay Data Standard
- Enable co-location of data and methods
Components of the BARD Platform

- CAP, Data Dictionary, and Results
- Deposition Data model created & populated
- Dictionary defined as OWL using Protégé
- Warehouse loaded with all PubChem AIDs and results
- Warehouse loaded with GO terms, KEGG terms, and DrugBank annotations
- CAP UI with View and basic editing
Interacting with the BARD
Searching the BARD

- Full text search via Lucene/Solr
  - *Key entry point for new users*
  - Search code runs queries in parallel
  - Fast faceting, auto suggest, filters
- Entities are linked manually via Solr schema
  - Allows us to pick up entities related to the query
  - Supply matching context
- Custom NCATS code for fast structure searches
The BARD API

- A RESTful application programming interface
- Access to individual & collections of entities
- Documented via the wiki
- Versioned
- Hit a URL, get a JSON response
  - Every language supports parsing JSON documents
  - Easy to inspect via the browser or REST clients

http://bard.nih.gov/api/v18/
API Architecture

• Java, read-only, deployed on Glassfish cluster
• Different functionality hosted in different containers
  - Maintenance, security
  - Stability
  - Performance
Open Source as Far as Possible

http://bard.nih.gov/api

Caching Layer

Jersey Webapps deployed on HA Application Server Cluster

ETL Database

Text Search Engine

Structure Search Engine
API Resources

- Covers many data types
- Each resource supports a variety of sub-resources
  - Usually linked to other resources
- Use /_info to see what sub-resources are available

```
{"collection:[
  "/assays",
  "/biology",
  "/cap",
  "/compounds",
  "/documents",
  "/etag",
  "/experiments",
  "/exptdata",
  "/projects",
  "/search",
  "/substances",
  "/targets"
], "link": null}
```

Returns assay information

Available resources:
- GET /assays/_info
- GET /assays/{aid}/annotations
- GET /assays/{aid}
- POST /assays/
- GET /assays/_info
- GET /assays/etag/{etag}/facets
- GET /assays/etag/{etag}
- GET /assays/{aid}/targets
- GET /assays/{aid}/documents
- GET /assays/{aid}/projects
- GET /assays/{aid}/experiments
- GET /assays/{aid}/compounds
- GET /assays/{aid}/substances
- GET /assays/{aid}/experiments/{eid}
- POST /assays/annotations
- GET /assays/_schema
- GET /assays/etag
- POST /assays/etag
- PUT /assays/etag/{etag}
- GET /assays/etag/{etag}/_info
- GET /assays/recent/{n}
- /v1/assays/?filter=query_string[field]
## API Resources

<table>
<thead>
<tr>
<th>Entity</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>/assays</td>
<td>990</td>
</tr>
<tr>
<td>/biology</td>
<td>1080</td>
</tr>
<tr>
<td>/cap</td>
<td>1942</td>
</tr>
<tr>
<td>/compounds</td>
<td>42,572,799</td>
</tr>
<tr>
<td>/documents</td>
<td>10,499</td>
</tr>
<tr>
<td>/experiments</td>
<td>1314</td>
</tr>
<tr>
<td>/exptdata</td>
<td>32,754,242</td>
</tr>
<tr>
<td>/projects</td>
<td>144</td>
</tr>
<tr>
<td>/substances</td>
<td>113,751,456</td>
</tr>
</tbody>
</table>
Data Warning

We’re still in the process of data curation and QC so while you can access all BARD data, keep in mind that much of it will be undergoing review and curation and so may change
Summary Resources

- A number of entities have a /summary sub-resource (Compounds, Projects)
- Aggregates information, suitable for dashboards

```json
{
    assay_count: 12,
    experiments: [...],
    description: "The primary qHTS data of human PKM2 protocol for the PK isoforms M1, liver (L) and renal (R) assay that determined the activity of PK by coupling potency of the synthesized analog. While generally further studied for potential cancer therapeutic aspects.
    cmpd_synthesis_count: -12,
    name: "Discovery of Lead Compounds which Modulate...",
    probe_reports: [...],
    cmpd_purchase_count: -12,
    depositor: null,
    experiment_count: 12,
    targets: [...],
    probes: [...]  
}
```
JSON Responses

- All responses are currently JSON
- Entities can include other entities (recursively)
- JSON Schema is available via /_schema

```json
{
  type: "object",
  properties: {
    bardAssayId: {
      type: "number"
    },
    capAssayId: {
      type: "number"
    },
    category: {
      type: "integer"
    },
    summary: {
      type: "integer"
    },
    assays: {
      type: "integer"
    },
    classification: {
      type: "integer"
    }
  }
}
```
Extending the API

• Concept of plugins
  ➢ Expands the resource hierarchy
  ➢ Has to be written in Java

• What can a plugin accept?
  ➢ Anything

• What can a plugin provide?
  ➢ Anything - plain text, XML, JSON, HTML, Flash

• Plugin manifest - describe available resources, argument types
What Can a Plugin Expect?

- Direct access to the database via JDBC
- Faster access to REST API via co-localization
- No local storage in the BARD warehouse
  - But the plugin can use its own storage (such as an embedded database)
- Plugins have access to system JARs (e.g., XOM) but should bundle their own required dependencies
Plugin Validation

• Run a series of checks on a plugin before deployment
  - Catches manifest/resource errors
  - Doesn’t check for correctness (not our job)
  - Examines plugin Java class
  - Examines final plugin package

• Run as a command line tool or from your code
• Could be made into an Ant/Maven plugin
What Plugins are Available?

• The plugin registry provides a list of deployed plugins
  ➢ Path to the plugin
  ➢ Version
  ➢ Availability

• Long term goal is to have a plugin store

```json
[
  {
    path: "/plugins/smartcyp",
    title: "SMARTCyp",
    version: "1.1",
    available: true
  },
  {
    path: "/plugins/whichcyp",
    title: "WhichCyp",
    version: "1.0",
    available: true
  },
  {
    path: "/plugins/badapple",
    title: "BADAPPLE evidence-based promiscuity scores",
    version: "0.9beta",
    available: true
  },
  {
    path: "/plugins/ssearch",
    title: "Structure Search Plugin",
    version: "1.1",
    available: true
  },
  {
    path: "/plugins/ainfo",
    title: "BARD Assay Information",
    version: "1.1",
    available: true
  },
  {
    path: "/plugins/cs1s",
    title: "Chemical Structure Lookup Service Wrapper",
    version: "1.1",
    available: true
  }
]
```

/plugins/registry/list
Exemplar Plugins

• Look at the plugin repository on Github
• Ranges from
  ➢ trivial calls to external service
  ➢ Incorporate external tools/programs
• Current set of plugins highlight plugin development features
• Plugins let us move non-essential functionality out of the core API
Based on Lars Olsen lab’s work. [http://www.farma.ku.dk/smartcyp/](http://www.farma.ku.dk/smartcyp/)
More Than Just Data

BARD is not just a data store - it’s a platform

• Seamlessly interact with users’ preferred tools
• Allows the community to tailor it to their needs
• Serve as a meeting ground for experimental and computational methods
• Enhance collaboration opportunities
## BARD API Source Repository

[https://github.com/ncats/](https://github.com/ncats/)

### GitHub Repository: ncats/bard

#### Modified result factory to add physical properties as endpoints for SP.

- **joinbraisted** authored 20 hours ago

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>capextract</td>
<td>Modified result factory to add physical properties as endpoints for SP.</td>
</tr>
<tr>
<td>entity</td>
<td>Removed classes (ie target class) from ProteinTarget responses.</td>
</tr>
<tr>
<td>pcparser</td>
<td>rough rsr-manager pkg pcparser updates</td>
</tr>
<tr>
<td>plugin</td>
<td>Updated to support form arguments</td>
</tr>
<tr>
<td>resourcemgr</td>
<td>Commit ExperimentHandler with check for empty context</td>
</tr>
<tr>
<td>rest</td>
<td>updated compound summary to report protein targets rather than target...</td>
</tr>
<tr>
<td>search</td>
<td>Check for NPE when processing filter queries</td>
</tr>
<tr>
<td>service</td>
<td>rough rsr-manager pkg pcparser updates</td>
</tr>
<tr>
<td>tools</td>
<td>fix double close on statement</td>
</tr>
</tbody>
</table>
BARD Plugin Source Code Repository

https://github.com/ncats/bardplugins

<table>
<thead>
<tr>
<th>Branch</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>master</td>
<td>Added example plugin to highlight per-request initialization versus s...</td>
<td>a day ago</td>
</tr>
<tr>
<td></td>
<td>Fixed typo in email spec</td>
<td>7 months ago</td>
</tr>
<tr>
<td></td>
<td>Updated to include a no arg resource, just to test manifest</td>
<td>6 months ago</td>
</tr>
<tr>
<td></td>
<td>Added example plugin to highlight per-request initialization versus s...</td>
<td>a day ago</td>
</tr>
<tr>
<td></td>
<td>Updated to latest bard plugin jar file</td>
<td>4 months ago</td>
</tr>
<tr>
<td></td>
<td>Use an informative 404 rather than a redirect</td>
<td>22 days ago</td>
</tr>
<tr>
<td></td>
<td>Added two more resources (get classes in bulk and get accs for class)</td>
<td>24 days ago</td>
</tr>
<tr>
<td></td>
<td>Update hostnames</td>
<td>6 months ago</td>
</tr>
<tr>
<td></td>
<td>Updated smartcyp plugin to include query params in manifest</td>
<td>6 months ago</td>
</tr>
<tr>
<td></td>
<td>Fixed typo in email spec</td>
<td>7 months ago</td>
</tr>
<tr>
<td></td>
<td>Updated to latest bard plugin jar file</td>
<td>4 months ago</td>
</tr>
</tbody>
</table>
Questions?

ajadhav@mail.nih.gov
**Other NCATS Cheminformatics Resources**

**Hopping for success**

Molecular scaffold is a fundamental concept in medicinal chemistry. Perhaps nowhere this is more apparent than during early stage discovery, where scaffolds are the primary means by which medicinal chemists communicate. Extending this concept further, we recently introduced **Scaffold Hopper** as a tool that utilizes scaffolds to facilitate context “hopping.” Starting with either (i) a set of structures (e.g., hits from an HTS screening campaign), (ii) publication, or (iii) search terms, **Scaffold Hopper** allows the user to quickly “hop” between related contexts (i.e., compounds, documents, targets, MeSH terms) with just a single click. (We’re currently working to add additional contexts such as assays, clinical trials, and patents.) Below is a quick preview of the tool

[Show as slideshow]

Click on the Launch button below to take it out for a spin:

---

**Note for Mac users:** Due to recent Java security updates, Java webstart is no longer launched when clicking on the .jnlp file. You can still launch it manually by selecting “Open With” then “Java Web Start”