Introduction to the data

- High throughput screening campaigns

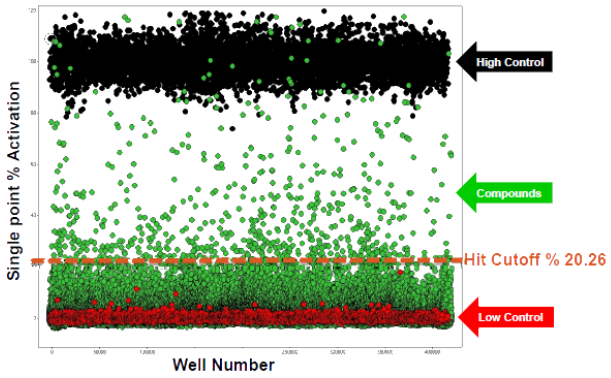
Primary Screen / Counter-screen
>= 315K compounds,
single point, single conc.

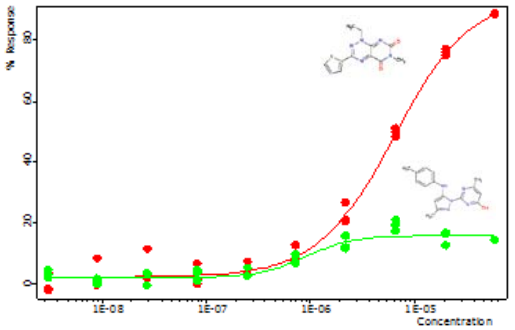
Confirmation / Counter-screen
1000's compounds,
triplicate, single conc.

Dose response
1000 compounds,
triplicate, titration series.

Synthetic chemistry,
Cheminformatics
Late stage assays.

Probe





Introduction to the community

- Molecular Libraries Probe Center Network (MLPCN)
- PubChem is the repository for all MLPCN assays.
  - > 69M chemical substances from 126 vendors
  - > 434K bioassays from 41 institutions (http://pubchem.ncbi.nlm.nih.gov/sources/sources.cgi)
    - Mostly small molecule assays with a small number of nucleotide assays
  - Data is available via PubChem FTP site and via NCBI web services.
  - Analysis tools available from PubChem Website
- Probe reports are available from the MLI website and soon from the NIH Bookshelf.
  - http://mli.nih.gov/mli/mlp-probes/

Engaging the community

- PubChem
  - All assays and structures uploaded to PubChem in a timely manner.
  - Can be non-public (“Embargoed”) for a period of time but must be submitted.
- MLPCN Center
  - Outreach efforts
  - Funding opportunities
  - Publications
  - http://mlpcn.florida.scripps.edu
- Assay provider
  - “Campaign” level SAR analysis
  - Follow up chemistry
  - Publications

Relating to the community

- Each BioAssay has defined fields
  - Description, Protocol, related assays, Xrefs, result definitions etc.
- Each BioAssay result defines
  - Activity score (0 – 100)
  - Activity outcome (Active, Inactive, Unknown, Unspecified)
- PubChem website tools
  - Bioassay Activity Analysis
    - Two dimensional clustering of Compounds and Assays
  - Structure Searching

Some future challenges

- Make it easier for researchers outside MLPCN to enter assays into PubChem.
- Query data with context
  - distinguish between different types of assay
    - e.g. functional and toxicity assays
  - determine the importance of assay endpoints.
- Map / integrate ontologies
- Size of datasets.
  - Millions of files on PubChem FTP site.
  - Improved (computational) access to datasets
  - Local databases and resources to speed processing times.
  - Image based assays?
- Improved handling of dose response
  - Curve fits and/or curve statistics.
- Pathway analysis
- ADMET