CIIPro: An online cheminformatics portal for large scale chemical data analysis

Daniel P. Russo, Marlene T. Kim, Wenyi Wang, Daniel Pinolini, Hao Zhu
Zhu Research Group
Center for Computational and Integrative Biology
Rutgers University, Camden, NJ
Outline

• Motivation: Current state of toxicology
• High throughput screening
• In vitro in vivo correlation
• Chemical In vitro In vivo profiling
• Weighted Estimated Biological Similiarity
• Demo/tutorial
Current State of Toxicology

• Large amounts of chemicals without toxicity information
• High rate of failure of drugs due to toxicity, adverse drug reactions owing to a lack of understanding of toxicology mechanisms
• Ethical concerns on the use of animals, high cost of testing, and difficulty in species extrapolation has pressured alternative methods to prioritize chemical toxicity assessment
  • *In vitro* assays
  • QSAR, readacross, etc
High throughput screening

• HTS uses robotics in various in vitro cellular assays in a rapid standardized manner
• Several fields such as drug discovery and toxicology have undertook large HTS efforts (PDSP, ToxCast) yielding a wealth of compounds linked to rich, biological data
• Updated daily, various public repositories (i.e., PubChem) have curated, stored, and made this data publicly available.
• Compound response data from PubChem can offer unique insights to a compounds’ in vivo response
In vitro in vivo correlation

• What can we learn from these data and what are its applications?

• 1) Predictive models of complex endpoints
  • QSAR pit falls
    • Compounds linked to adverse toxicity endpoints are chemically diverse and likely influence multiple biomolecules
    • Activity cliffs where compounds are chemically similar but display differing bioactivities
  • Supplementation of QSAR with *in vitro* assays have improved model performance

• 2) Animal toxicity data exists at the highest level of biological organization
  • There currently exists a knowledge gap between toxicants and their relative biological endpoints
  • *In vitro*-* in vivo* correlations and *in silico* tools applied to the growing pool of assay data can offer insight as to what biomolecules are adversely influenced by toxicants
Chemical In vitro In vivo profiling

• The current computational tools linking target compounds to *in vitro* assays and biological endpoints are lacking
  • Introduce CIIPRO – freely accessible through the web at ciipro.rutgers.edu
• Compounds equipped with *in vivo* endpoints can be used to query PubChem
  • CIIPros’ unique algorithm will extract the data based upon the *in vitro in vivo* correlation and output in a format conducive to modeling
• Extracted assays can be ranked by various statistical parameters
Weighted Estimated Biological Similarity

• Within the portal, the extracted biological data can be used to make predictive models using user defined parameters
  • WEBS algorithm overview
  • Users can optimize predictions using user defined parameters
    • Define Bio similarity cutoff, confidence and nearest neighbor parameters
• Models can be evaluated using LOO cross validation (add 5 fold?)

• Go into future features?
  • In house dataset models capable of predictions?
  • Users can create and store models?