Informatics problems and solutions for phenotypic screenings, an AstraZeneca perspective

Ola Engkvist, Chemistry Innovation Centre, Discovery Sciences

Open Phacts, Santiago de Compostela 16th of February 2014

AstraZeneca

Phenotypic screening at AZ

Important complement to target based approaches

Annotated set (20K) plus diversity set

Investment in in-house chemical proteomics facility

Collaborations for CRISPR technology just announced
Press release

Current data resources

Bioactivity data
- In-house
  - SAR screening
  - HTS
  - Panel screenings
- Commercial
  - GoStar (GVKBio)
  - BioPrint (Cerep)
- Public
  - ChEMBL
  - PubChem

Target Prediction
- Molecular fingerprints
- Bayesian Models (Collaboration with Dr Andreas Bender)

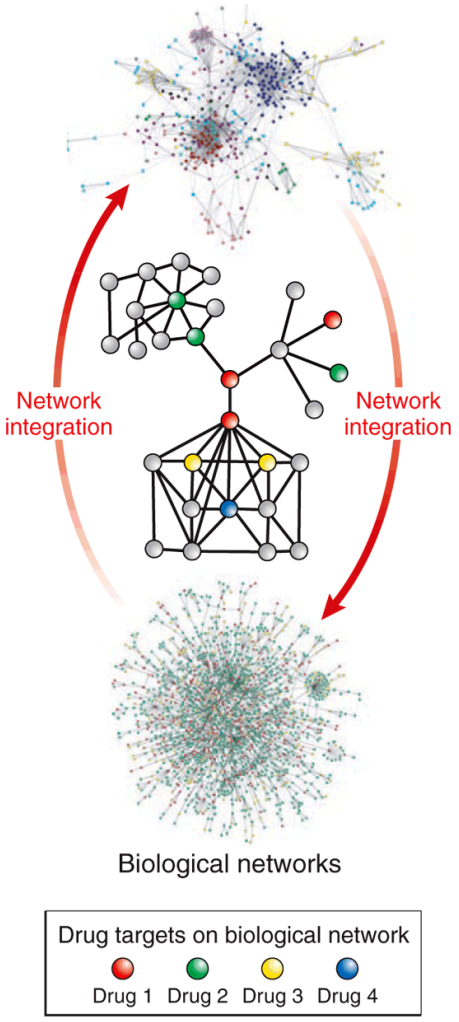
Target Enrichment Analysis
- Published in JBS

Pathway analysis
- GeneGo
- IPA

Textmining
- PubMed

Improvements
- Bio-Assay Ontology (Collaboration with Prof. Stephan Schürer)
- Cortelis Data Fusion

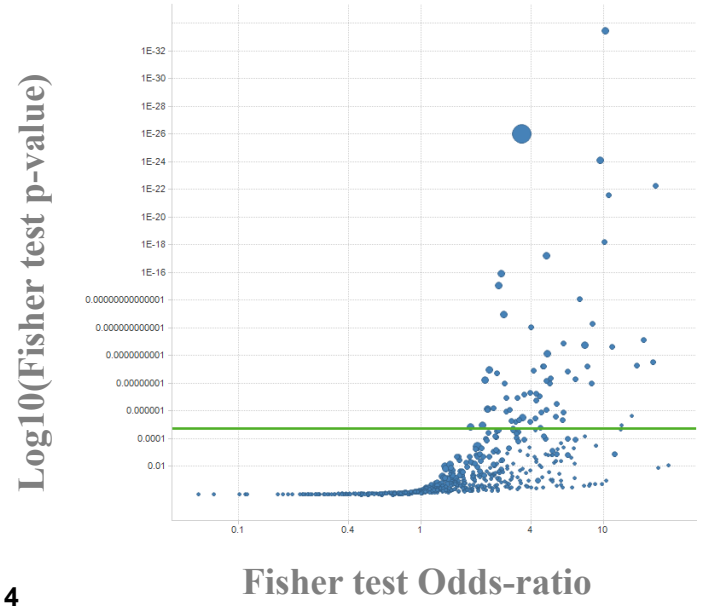
Sustainability
- IMI Open Phacts



Example: Target enrichment analysis

For each target a contingency matrix is generated with the number of compounds (Journal of Biomolecular Screening 2014):

Target X	Ph. Scr. Inactive	Ph. Scr. Active
Target X Inactive	596	17
Target X Active	87	26



Fisher's test:

P-value = 3.24e-12

Odds-ratio = 10.42

Bonferroni Correction for significance cut-off (2060 targets in analysis): 0.05/2060 = 2.43e-05

Enriched targets needs to be experimentally validated:

- **Pharmacological experiments**
- **RNAi**
- **Genome Editing**

Current bottleneck

How can in silico methods best support a target identification hypothesis generated from a chemical proteomics experiment?

- Which are the best known agonist tool compounds for a given target?
- In which cell types and tissues are a target expressed?
- Which is the subcellular location for a protein?
- Which other proteins is an identified protein known experimentally to interact with? Directionality?
- If there is no known small molecule modulators known for a target, which is the closest protein according to a protein classification with a known small molecule?