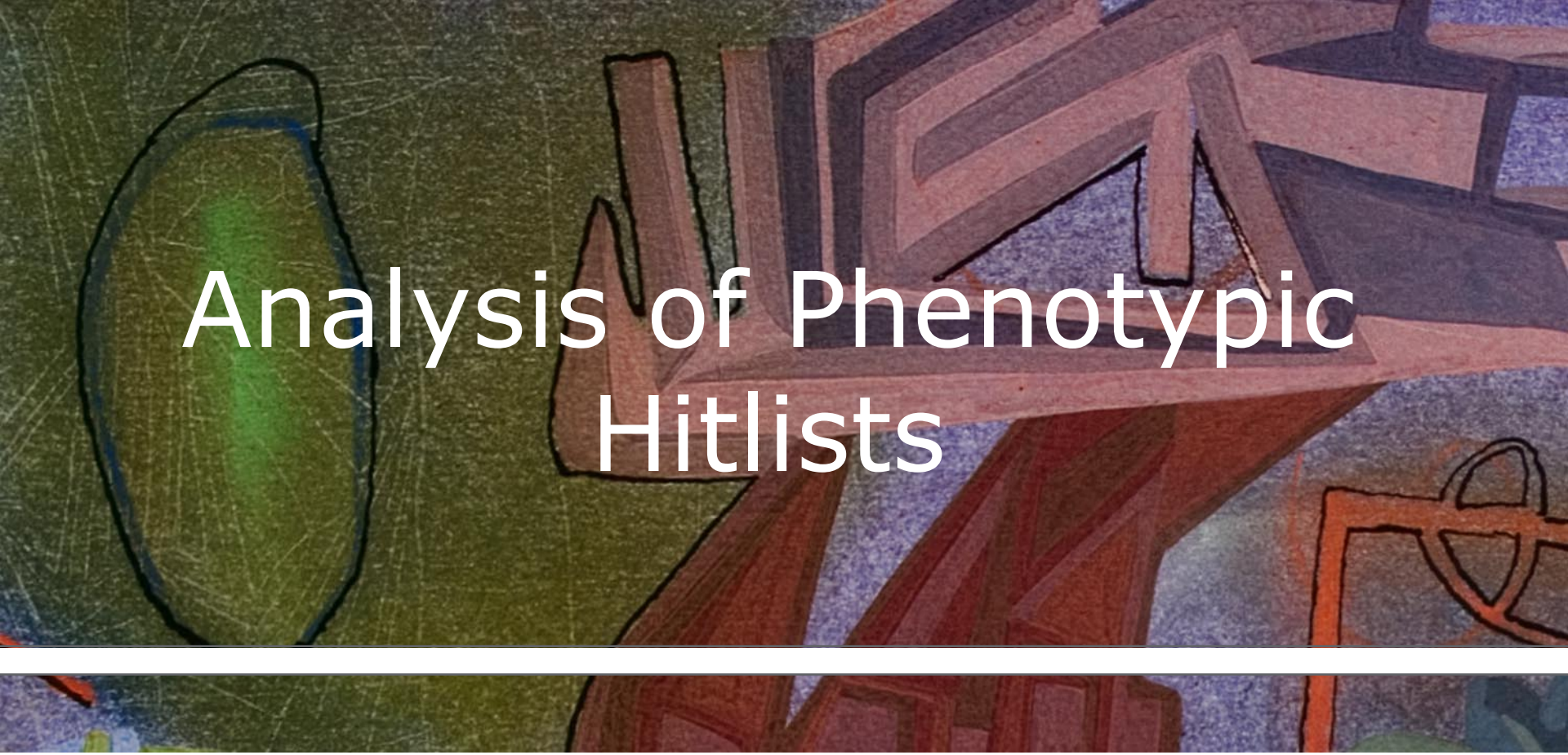


Analysis of Phenotypic Hitlists

Open PHACTS Workshop: Understanding the knowledge management needs of phenotypic screening

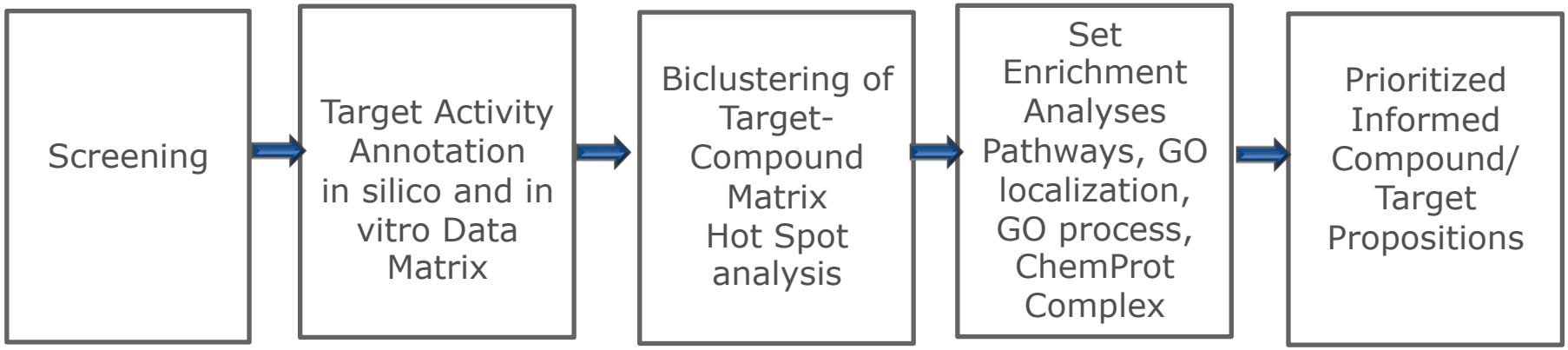
16.02.15

janssen
PHARMACEUTICAL COMPANIES
OF Johnson&Johnson

The Key Challenge of Phenotypic Screening Analysis

- The phenotypic screen is *per definition* target agnostic - This is different than screening in a cell based model with known target.
- Lead optimization is in such assays considered more difficult
- In Silico hit annotation, both based on experimental data or in silico models will support:
    - Experimental target deconvolution
    - Further biological validation
    - Assessment of potential tox liabilities (off-targets)
    - Future assay development and lead optimization

Proposed in-silico Workflow for Phenotypic Data Analysis



Needs for Additional Supporting Data

- Complementary Phenotypic Experiments
  - LINCS
  - HCSPhenoprint
  - Kinetic Cytometry
- Clinical Compound-Target Data
- Compound Data – Derived Cell Permeability Information
- Compound Data Physical Availability Data

Genes and Targets in OPS

- The Issue:
  - Chemical compounds act on targets (proteins) 👍
  - Pathways contain proteins – but WikiPedia refers to genes 👎
  - Tissues express both genes and proteins, NextProt only takes proteins 👍 but we cannot link to gene expression 👎
  - DisGenet connects diseases and genes 👍 but we cannot link compounds to genes 👎
  - One cannot link genetic and mutation information 👎

- A Possible Solution:
  - Use UniProt mapping file to find Entrez Gene IDs (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/idmapping_selected.tab.gz)
  - But the Swiss-Prot curated section only covers 95% of human, 95% of mouse, 90% of rat, 80% of cattle and less for other species
  - Also use NCBI Entrez Gene IDs mapping file to find UniProt AC (ftp://ftp.ncbi.nih.gov/gene/DATA/gene2accession.gz)
  - It's a bit involved, but one can use both files to confirm each other or to get more coverage from each one