

## Open PHACTS Computational Protocols for *in silico* Target Validation of Cellular Phenotypic Screens: Knowing the Knowns

Edgar Jacoby, Jean-Marc Neefs, Herman Van Vlijmen, Daniela Digles, Barbara Zdrazil, Gerhard Ecker, Christian Herhaus, Andrei Caracoti

*Linking Life Science Data: Design to Implementation, and Beyond*, Vienna, 18-19.02.2016



- **Objectives**
- **6 protocols**
- **Application to a phenotypic Prelamin A/C splicing assay**
- **Conclusions**



## Objectives

- Knowing the knowns in phenotypic screening is a must
- The molecular mechanism needs to be elucidated and validated
- In silico assessment of the hit lists constitutes a key step
- IMI Open PHACTS suite of software and databases provide a seamless integration to achieve this goal
- Enables insightful interpretation of the phenotypic screening results to sustain target validation based on hitherto established drug discovery knowledge



## **6 Protocols to Know the Knowns**

**Protocol 1 : ChEBI/ChEMBL Annotation and Classification**

**Protocol 2 : GO Annotation**

**Protocol 3 : Wikipathways Annotation**

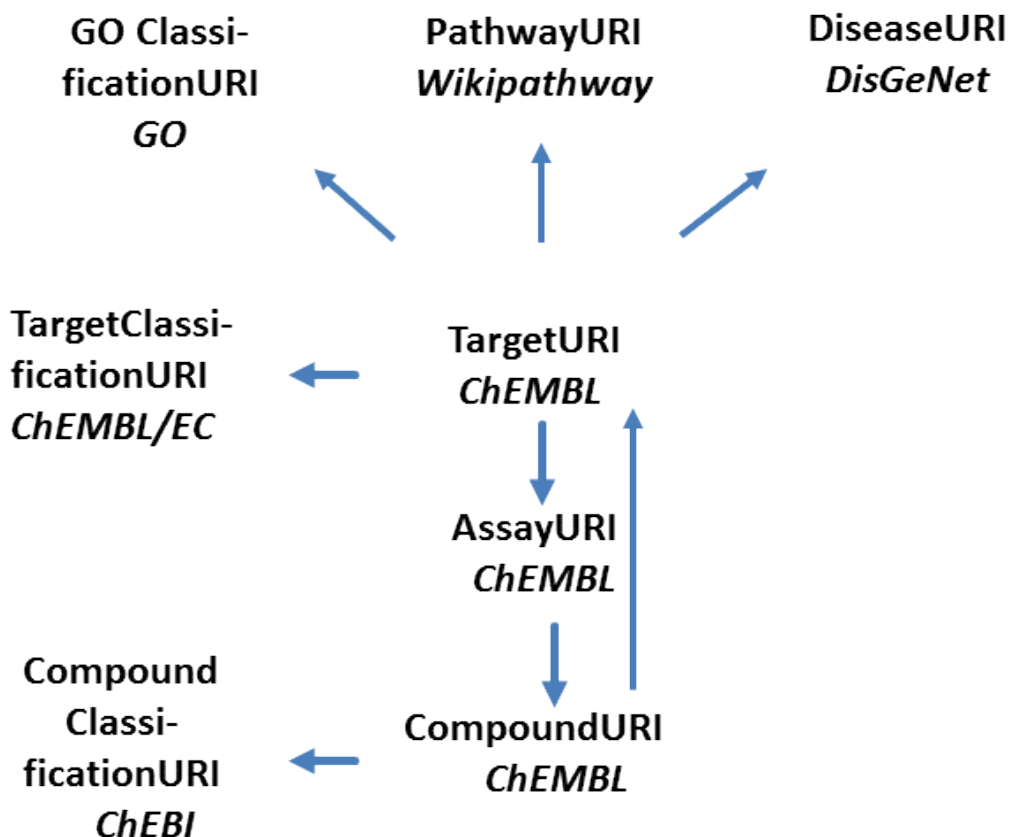
**Protocol 4 : Links to diseases and possible side effects – DisGeNET annotation**

**Protocol 5 : Correlation of the phenotypic and bio-chemical screening data**

**Protocol 6 : Compound tool box to validate/devalidate identified potential targets of protocol 1 based on IUPHAR database**

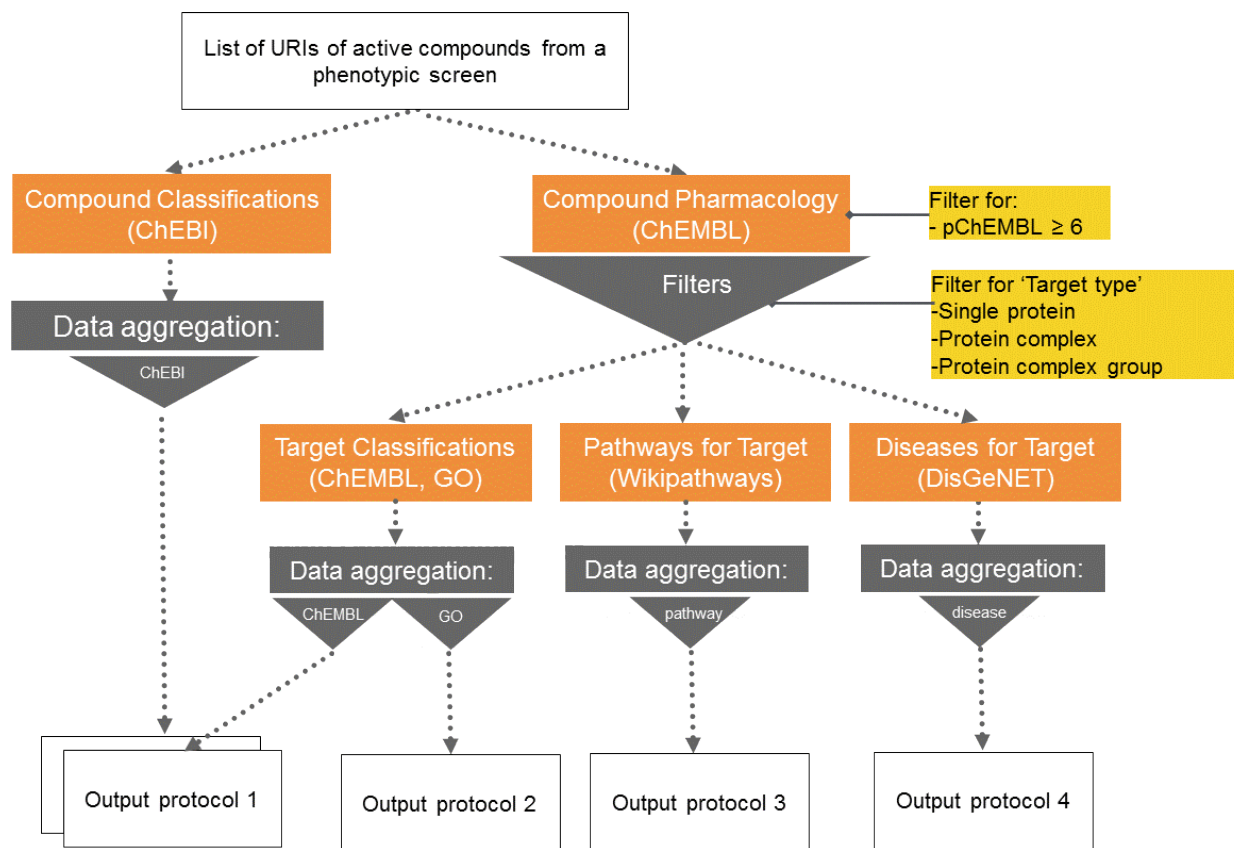


## Open PHACTS Provides the Required Semantic Integration



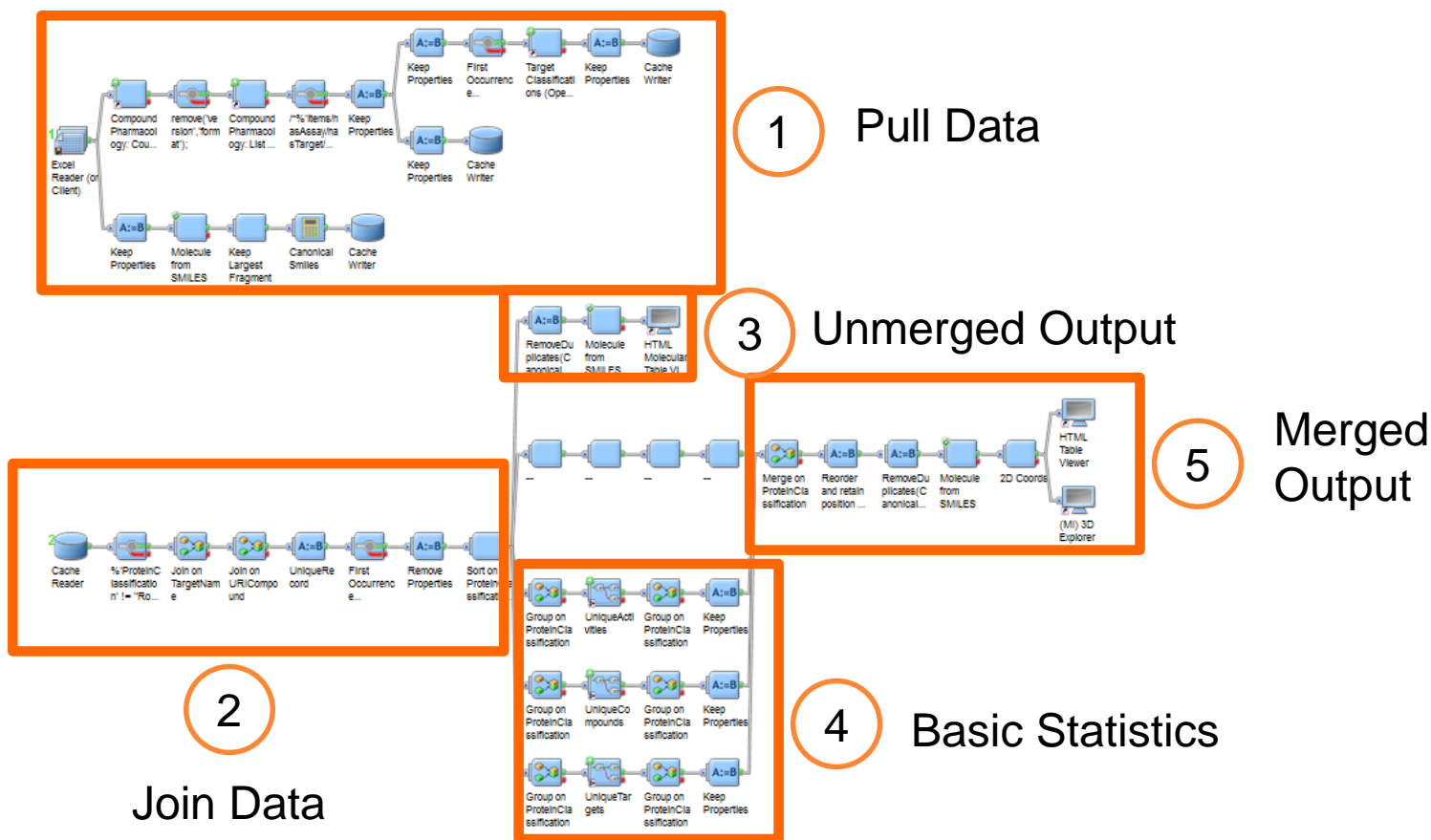


## Protocols 1-4 follow a generic pipelining flowchart



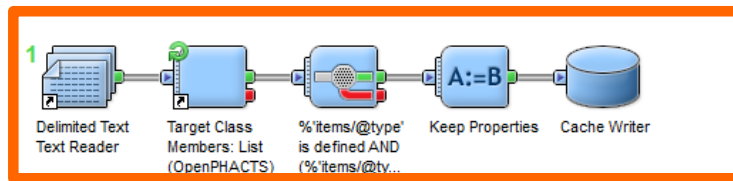


# Pipeline Pilot Implementation of Protocol 1

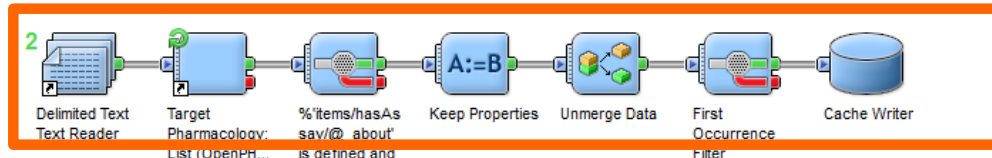




## Pipeline Pilot Implementation of Protocol 5 : Correlation Robot

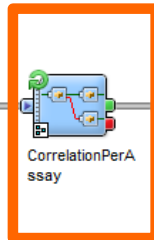
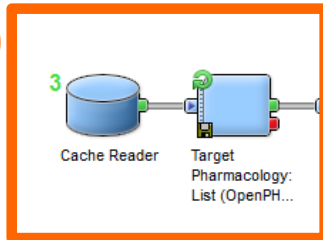


1 Pull all Kinases from Tree



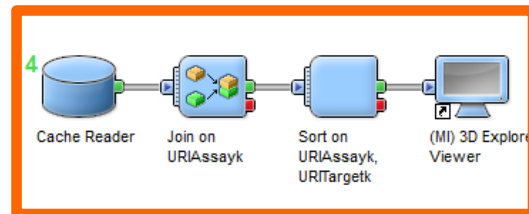
2 Pull Data for Phenotypic Assay

3



5

Correlation Analysis



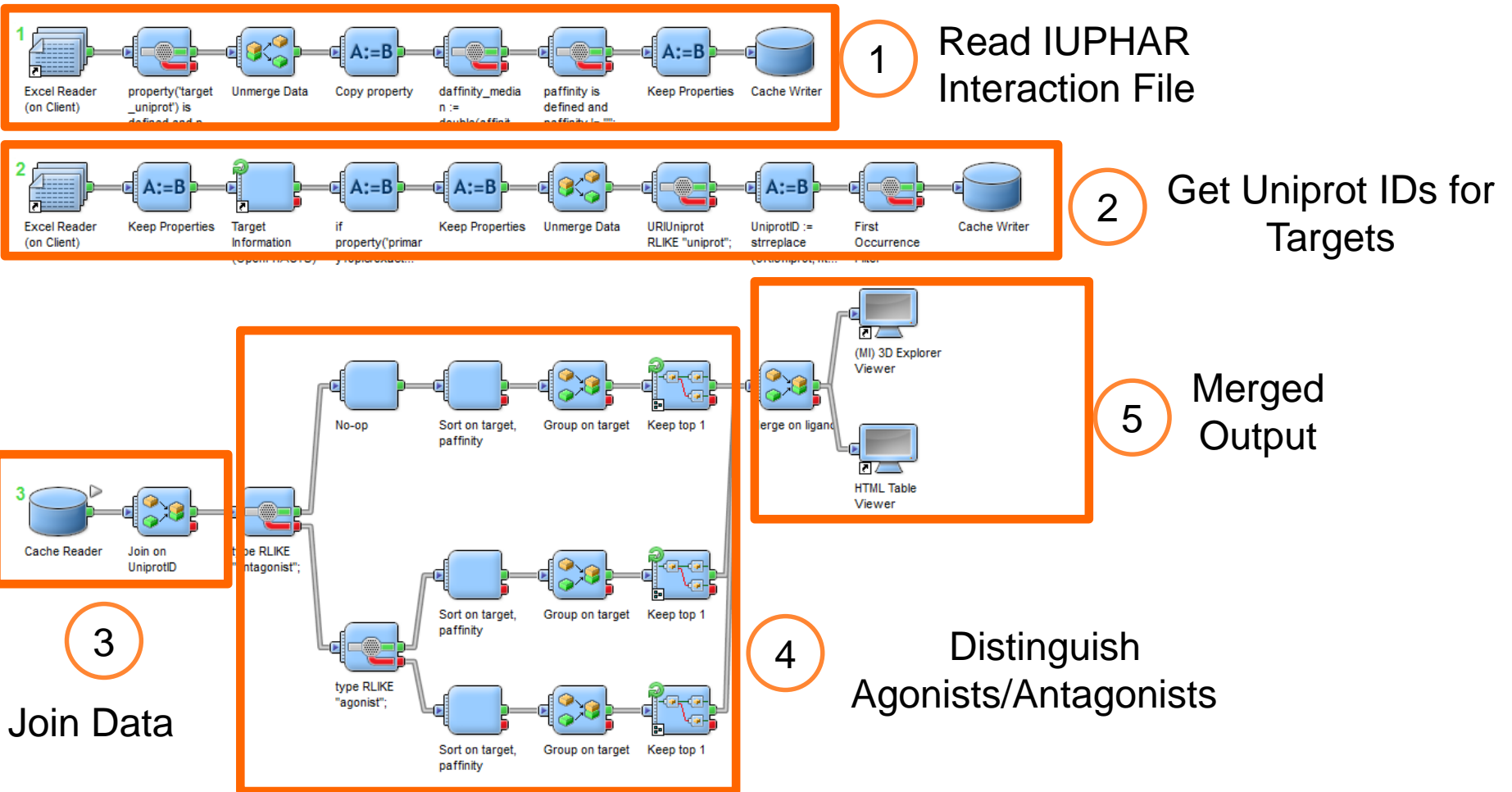
4 Output Table

Pull Data for Each Kinase





## Pipeline Pilot Implementation of Protocol 6 : IUPHARDB





## Application to Lamin A/C Splicing Assay

- Aim : Identify splicing correctors against the Hutchinson-Gilford Progeria Syndrome (HGPS)
- HGPS is a pediatric premature aging disease caused by a spontaneous mutation in the lamin A/C (LMNA) gene
- The mutation activates a cryptic splice site in the LMNA pre-mRNA which results in production of a pre-lamin A protein that cannot be processed properly. The mutant protein accumulates in the nucleus and negatively affects numerous cellular functions
- PUBCHEM\_BIOASSAY: Validation of Assay for Modulators of Lamin A Splicing lists 280 bioactives for which 92 chemically diverse compounds have pChEMBL values  $\geq 6$ .
- The assay measures expression of correctly spliced protein



## Insights to Lamin A/C Splicing Assay

### ChEMBL Annotations:

73 kinase activities are observed based on 10 compounds on 40 targets. Prominent are the CGMC kinases CDK1 and 5, DYRK1A and GSK3B and the MAPKinases p38  $\alpha$  and  $\beta$ , c-Jun1-3 and ERK2. DYRK1A inhibitors are reported in the literature to modulate alternative pre-mRNA splicing of model gene transcripts in cells with submicromolar potencies.

22 Nuclear NR1 activities are observed based on 7 compounds on 2 targets: ROR $\gamma$  and the thyroid hormone receptor 1B.

86 Family A GPCR activities are observed based on 19 compounds on 30 targets. Most prominent are the monoamine receptor activities.

30 epigenetic regulator activities are observed based on 19 compounds on 6 targets.

### GO Annotations :

241 GO Component terms are found; 14 compounds are linked to the spliceosomal complex via the heterogeneous nuclear ribonucleoprotein A1 and the survival motor neuron protein. 1628 GO Process terms are found. Multiple compounds are linked to various DNA related processes via the Bloom syndrome protein. 14 compounds are linked to spliceosomal complex assembly. 426 GO Function terms are found.



## Insights to Lamin A/C Splicing Assay

**Kinase assay correlation robot** : Points to the MAP kinase ERK2 assay ChEMBL1613808 which has 8 compounds in common. The underlying pathway is the MAPK signaling pathway.

**DisGeNet Annotation** : links to 3890 diseases and side-effects; 134 of them have more than 20 potential efficacy targets links. Various neoplasms and cancers are prominent given the link via kinases. Spinal muscular atrophy is linked by 14 hits via the survival motor protein link. It will require further disease biology expertise to recognize relevant links to the observed phenotype.

**ChEBI** terms with  $\geq 5$  compounds include 5 metabolites and 9 antineoplastic agents among which : Fluorouracil, camptothecin and rotenone. Rotenone is discussed in the literature to modulate splicing of several genes.

**IUPHAR BOX** analysis suggests to test 110 compounds to test 94 of 165 potential targets. Very prominent are monoamine receptor ligands and kinase inhibitors.



## Conclusions

- IMI Open PHACTS provides seamless integration for analysis of phenotypic screening data
- Interplay of OPS API calls with data pipelining and analysis tools allows great flexibility
- Knowing the Knowns is possible for everyone !