Representation and Extraction of Causality Statements

Vasundra Touré[1], Åsmund Flobak[2], Astrid Laegreid[2], Martin Kuiper[1]
[1]Department of Biology
[2]Department of Clinical and Molecular Medicine

In collaboration with
EBI – Reactome, PSI-MI
Institut Curie
ENS

Norwegian University of Technology and Sciences

vasundra.toure@ntnu.no

@Vasundra_T
@DrugLogics

NTNU

The DrugLogics Initiative

Towards the development of precision and personalised medicine

Crossover Research

Structured Knowledge
Commons resource
DbTF curation
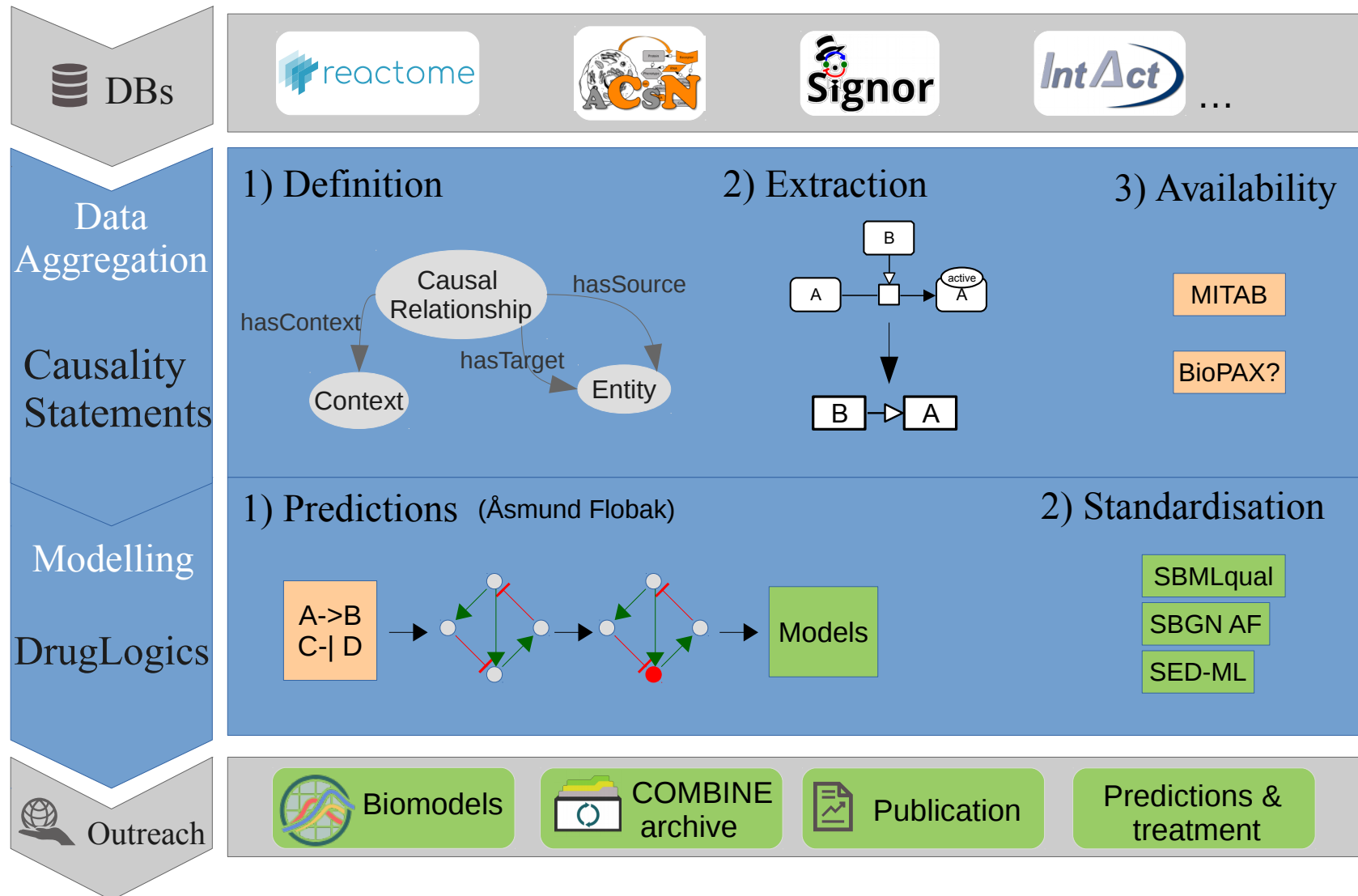Scicura

DrugLogics

Drug development of
anti-cancer combinations

COLOSYS

Drug resistance
prediction in colon cancer
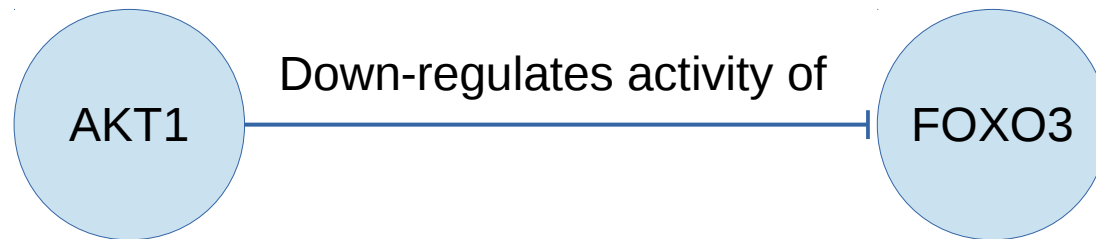via computer models

Personal tasks within the DrugLogics

Facilitate the process of building biological models with causal statements

Representation of causal statements

Causal interaction between two biological entities (gene, RNA, protein, complexes, etc…)
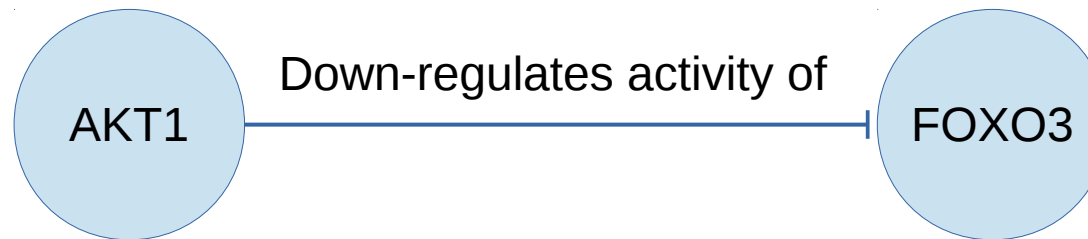
Representation of causal statements

How to represent meaningful causal interactions?

What is FOXO3's state?
(active/inactive)

When and where does
this interaction occurs?

Which molecular function
is down-regulated?

What is the regulation type?
(phosphorylation, acetylation,
dephosphorylation)

Is it a direct or indirect
Interaction?

# Minimum Information for Causality Statements

Entity – Source & Target
- ID – ex: causalDB:FOXO3
- Reference ID - HGNC; Uniprot; Entrez
  - *For Complex: ComplexPortal?*
- Name – ex: FOXO3
- Type - gene, RNA, protein, complex
- Molecular function – GO:MF
- Compartment – GO

Causal Relationship
- Regulation type – down-regulates
- Mechanism - PSI MOD?
- Modified residue – Tyr@P202
- Provenance – ex: Reactome, PMID
- Evidence – ECO
- Confidence – intact-miscore
- *Interaction depth – 0 (direct); 1; 2; etc…*

Context – case specific
- General biological context
  - Species – TaxID
  - Tissue type – Brenda Tissue Ontology (BTO), Uberon?
  - Cell type – BTO, Cell Line Ontology (CLO)?
- Text – Scicura sentence (http://scicura.org/info.html)
- Experimental conditions
  - Tissue / Cell state
- Source context
  - State of chromatin
  - Concentration – could be a number, a range
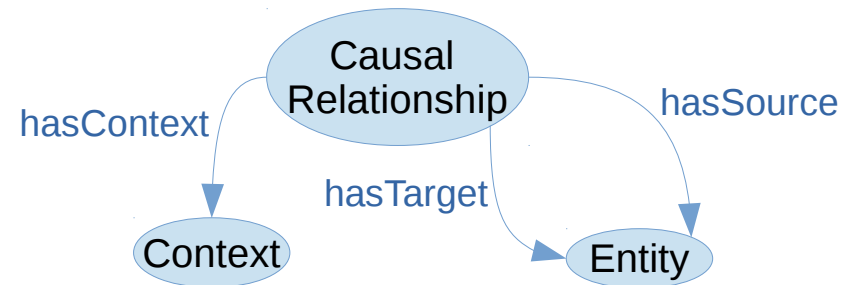
Controlled Vocabulary and Ontologies – essential to make data sustainable, shareable and interoperable
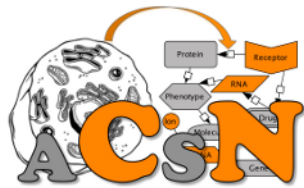
Extraction from prior knowledge

Aggregation of causal data from several existing resources

Pathways, reactions
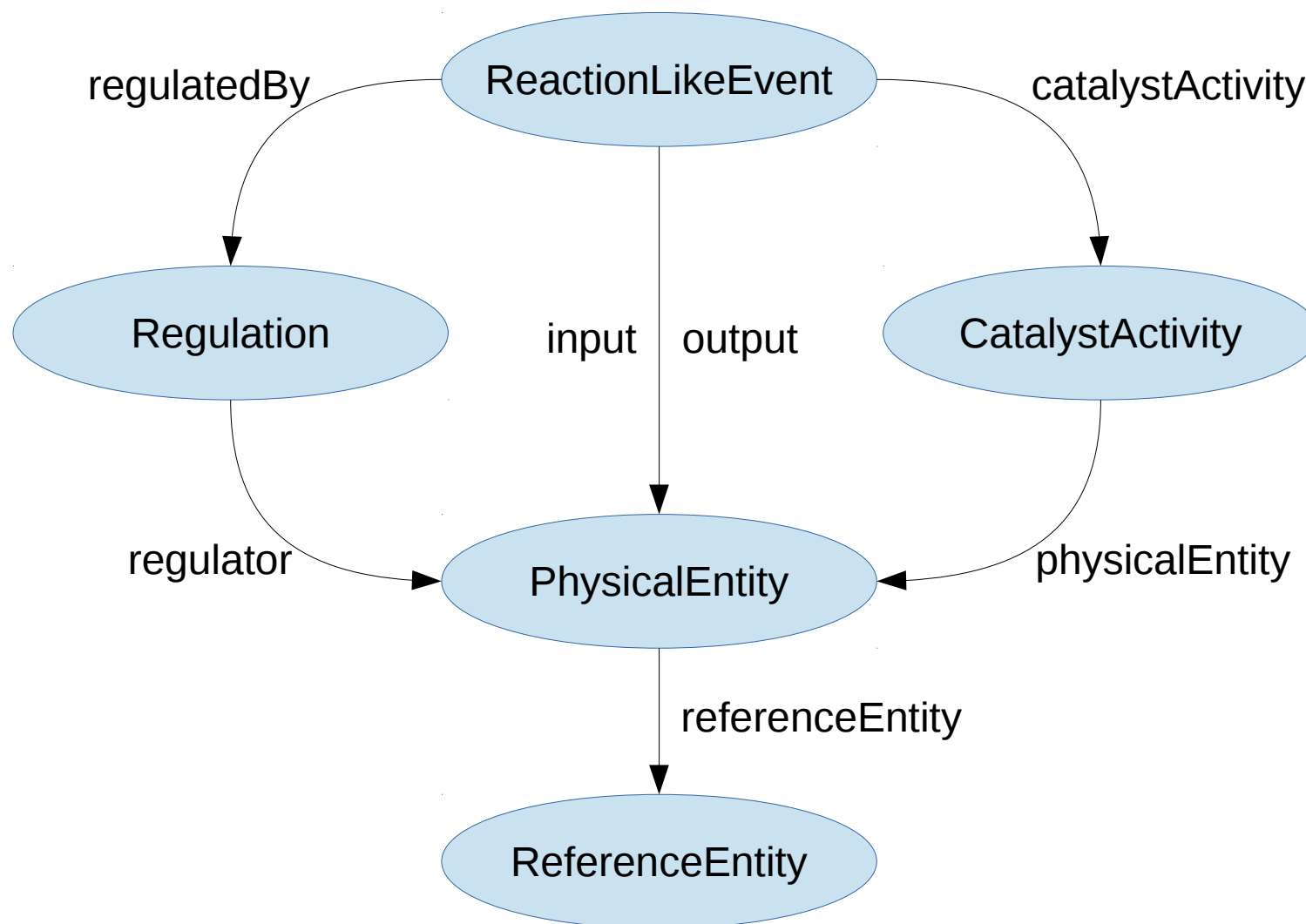

Pathways of cancer related signaling networks

DB of causal interactions

DB of molecular interactions

Case 1: extraction from Reactome

Reactome data model extraction using Neo4j and Cypher Query language

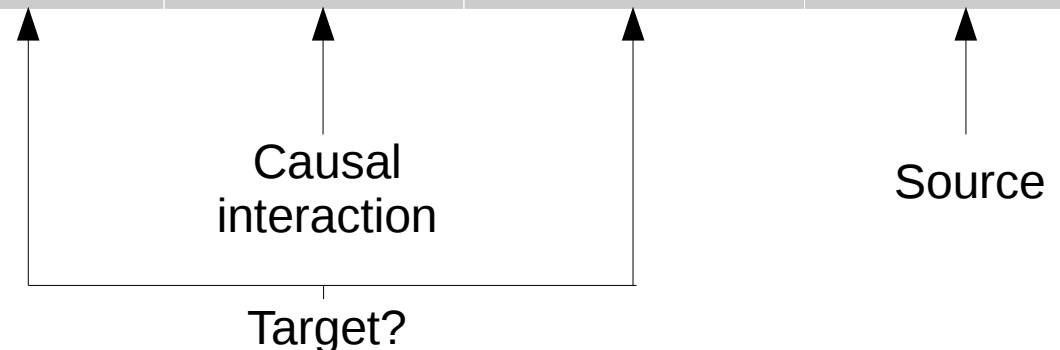**Reactome data model extraction using Neo4j and Cypher Query language**

Example: Get all reactions regulated by a physical entity or catalysed by a catalyst activity

```
MATCH (rle:ReactionLikeEvent)-[:regulatedBy|catalystActivity]->(o)-[:regulator|physicalEntity]->(source:PhysicalEntity)

OPTIONAL MATCH (input:PhysicalEntity)<-[:input]-(rle)-[:output]->(output:PhysicalEntity)

RETURN    rle.stId AS ReactionID,
          rle.displayName AS Reaction,
          COLLECT(input.displayName) AS Inputs,
          COLLECT(output.displayName) AS Outputs,
          o.simpleLabel AS Regulation,
          source.displayName AS Regulator
```

"Cypher is your friend" - A. Fabregat
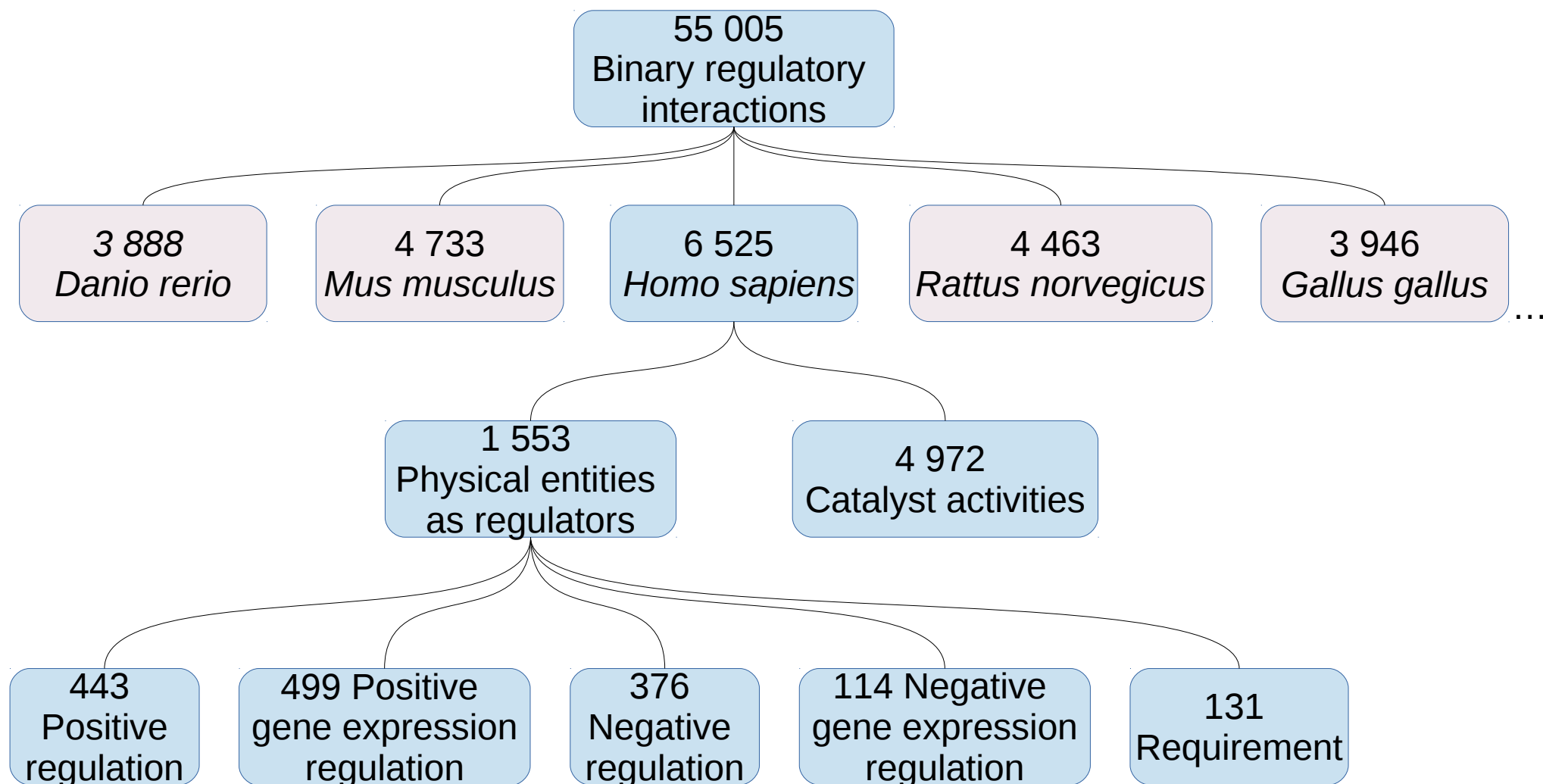
Case 1: resulting outputs

ReactionId	Reaction	Compartment	Inputs	Effect	Outputs	Regulator
R-HSA-452338	Expression of TDGF1 (CRIPTO)	cytosol	["TDGF1 gene [nucleoplasm]"]	NegativeGeneExpressionRegulation	["N-aspartyl-glycosylphosphatidylinositolethanolamine-TDGF1(31-188) [plasma membrane]"]	NR6A1(GCNF):TDGF1 gene [nucleoplasm]
R-HSA-8936628	GP1BA gene transcription is stimulated by the complex containing RUNX1, PRMT1 and GATA1 and inhibited by the complex of RUNX1, SIN3A and PRMT6	plasma membrane	["GP1BA gene [nucleoplasm]"]	NegativeGeneExpressionRegulation	["GP1BA [plasma membrane]"]	RUNX1:CBFB:SIN3A,(SIN3B):PRMT6:HDAC1:GP1BA gene:H3K4me2,H3R2me2a-Nucleosome [nucleoplasm]
R-HSA-8944497	PTEN mRNA translation is negatively regulated by microRNAs	cytosol	["PTEN mRNA [cytosol]"]	NegativeGeneExpressionRegulation	["PTEN [cytosol]"]	miR-20 RISC:PTEN mRNA [cytosol]

Causal interaction

Target?

Source

Case 1: Some numbers – Version 61

Number of pathways: 22 723
Number of reactions: 84 759

55 005
Binary regulatory
interactions

3 888
Danio rerio

4 733
Mus musculus

6 525
Homo sapiens

4 463
Rattus norvegicus

3 946
Gallus gallus

...

1 553
Physical entities
as regulators

4 972
Catalyst activities

443
Positive
regulation

499 Positive
gene expression
regulation

376
Negative
regulation

114 Negative
gene expression
regulation

131
Requirement

Case 1: questions / challenges raised

- Exclude trivial molecules

```
A0  +  NAD+  <=>  A1  +  NADH  +  H+
```
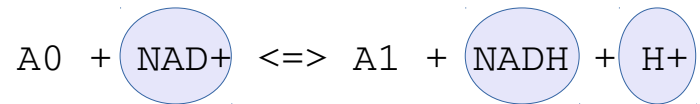
- Missing IDs for the modified mechanism type (event categories)

    Transfers
    Translocates from … to
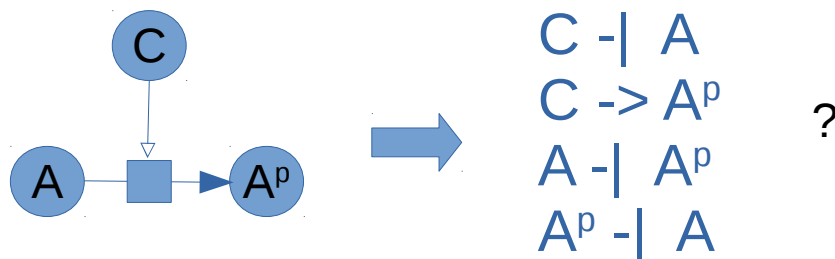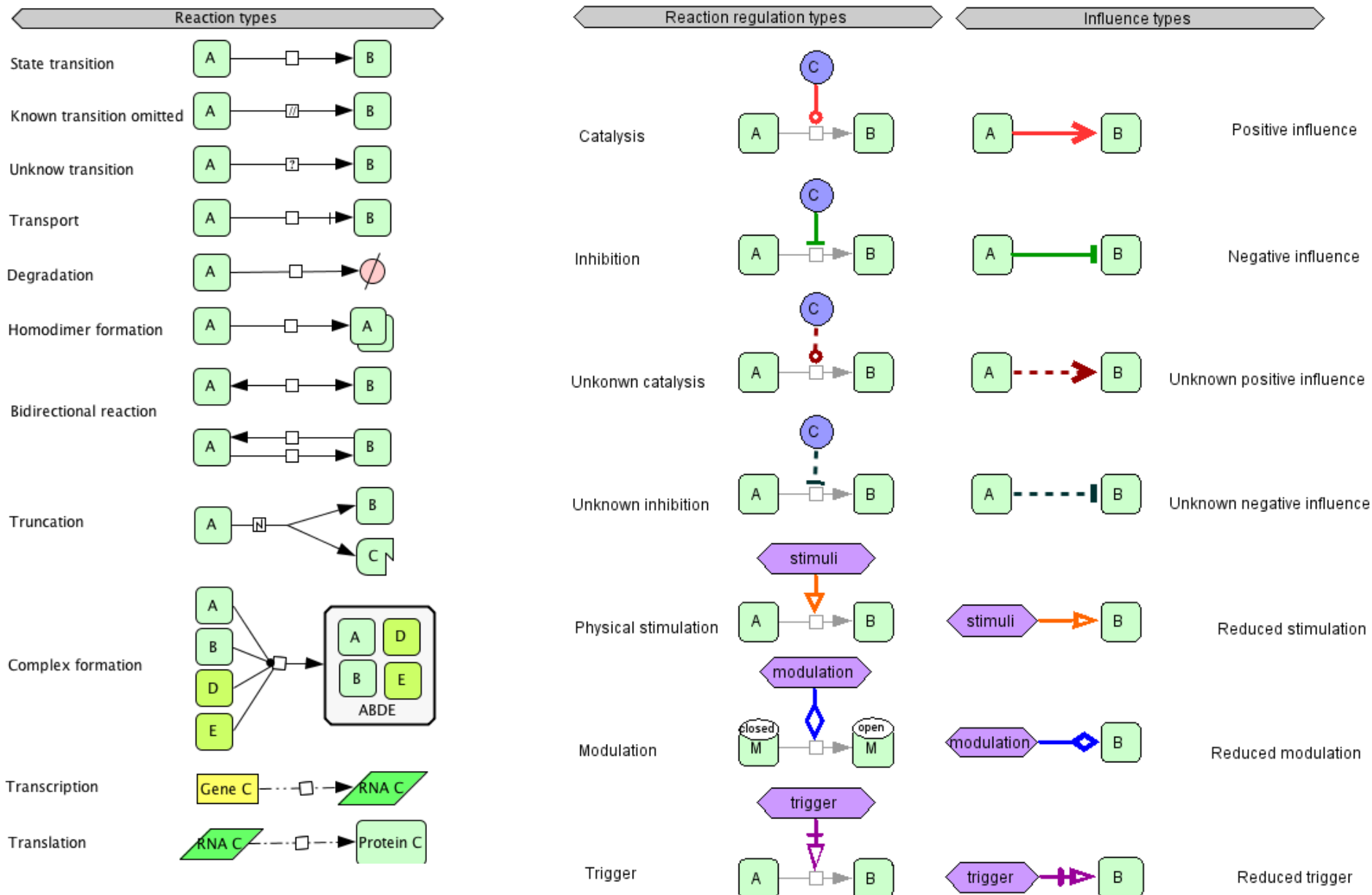   *Transports*
    Exchanges … for …
    Cotransports
    Regulates

- Inference of causal interactions from reaction networks

C -| A
C -> A^p ?
A -| A^p
A^p -| A

Case 2: extraction from ACSN
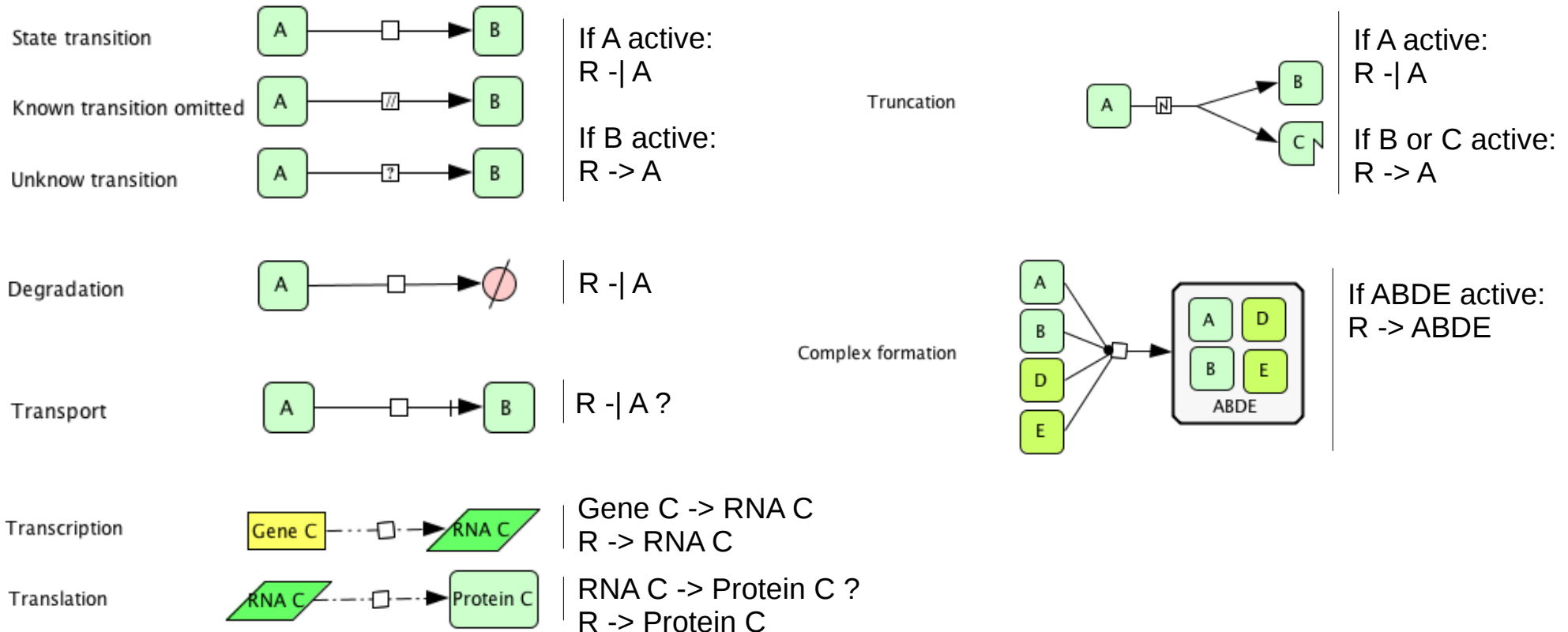

ACSN data model follows SBGN PD schema from CellDesigner tool

How to convert reaction types to causal interactions?

Suppose R, a regulator targeting each reaction below

State transition

If A active:
R -| A

Known transition omitted

If B active:
R -> A

Unknow transition

Truncation

If A active:
R -| A

If B or C active:
R -> A

Degradation

R -| A

Transport

R -| A ?

Complex formation

If ABDE active:
R -> ABDE

Transcription
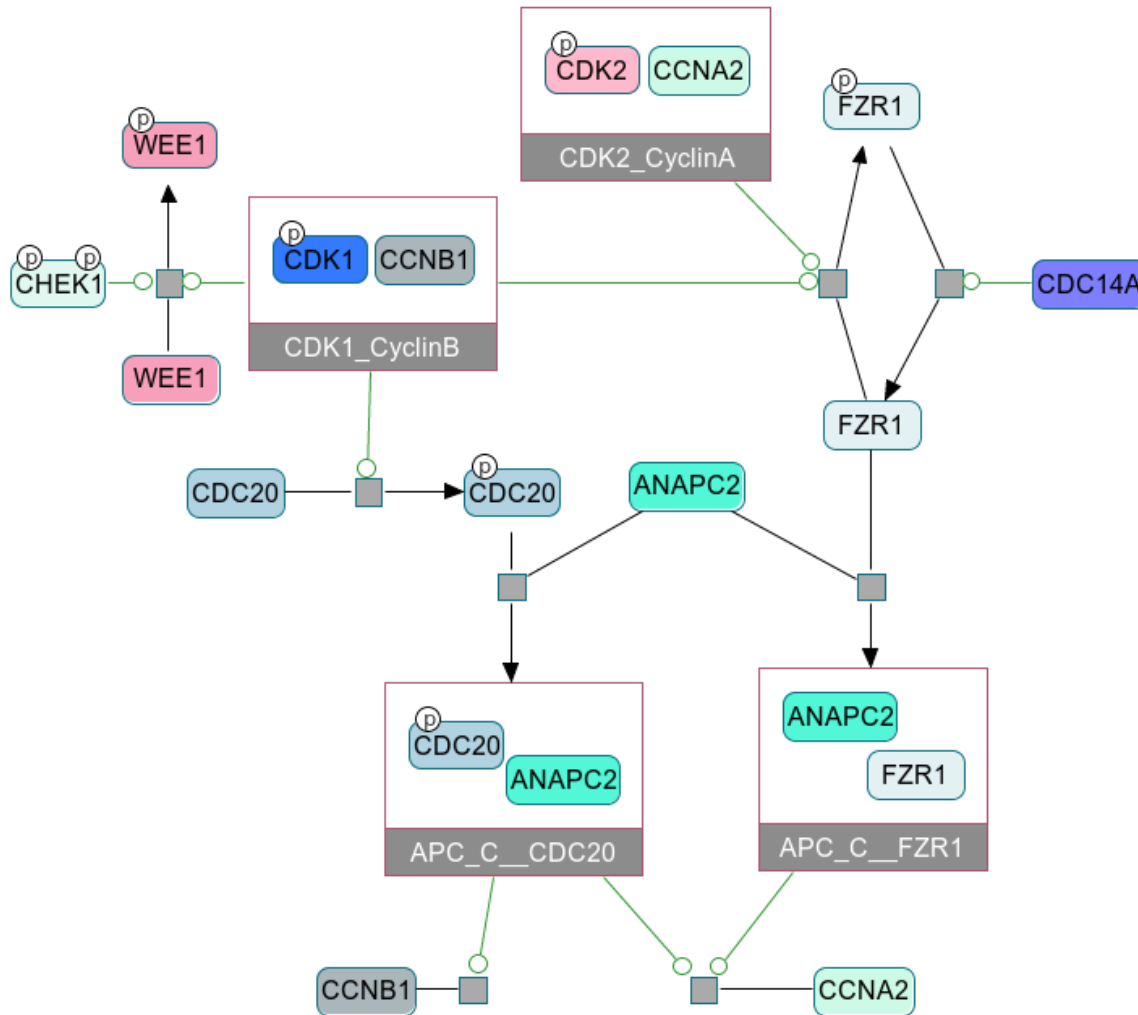
Gene C -> RNA C
R -> RNA C

Translation

RNA C -> Protein C ?
R -> Protein C

Catalysis, unknown catalysis, trigger, physical stimulation → positive influences
Inhibition, unknown inhibition → negative influences
Modulation → not defined influence

Notation
-| inhibition
-> activation

Case 2: extraction from ACSN

Manual inference of causal interactions from the Cell Cycle – APC module

List of causal interactions:

CDK2_CyclinA -| FZR1

CDK1_CyclinB -| FZR1

CDC14A -> FZR1

APC_C_CDC20 -| CCNA2

APC_C_FZR1 -| CCNA2

APC_C_CDC20 -| CCNB1

CDK1_CyclinB -> CDC20

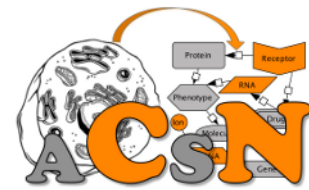Undefined causality:

CDK1_CyclinB regulates Wee1

CHEK1 regulates Wee1

→ How to interpret when both reactant and product are regulators…?

Case 2: extraction from ACSN

Atlas of Cancer Signaling Network – pathways of biochemical interactions

SBML with CellDesigner annotations   BioPAX

Difficult to parse   Easier to parse
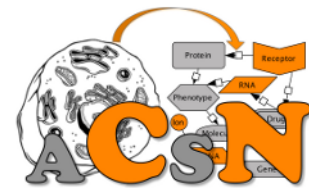   More intuitive

2 659 CATALYSIS:
2 537 – regulatory interactions
122 – boolean gates

Case 2: Possible improvements in ACSN's BioPAX

- BioPAX online validator: No.unification.xref errors

- Family information missing: use 'memberEntityReference' ?

- Complex association / dissociation: better defined with 'spontaneous' in ComplexAssembly?

- Annotations on Entities: PMIDs should be on Reactions

Future work

- Pipeline to extract causal interactions from PD networks with conversion rules defined

- Model checking to validate our models

- Discuss MICAST with the community (curation workshop in Dec)

Thank you for your attention!

The DrugLogics team

Dept of Biology	**Dept of Clinical and Molecular Medicine**	**Dept of Philosophy and Religious Studies**
Martin Kuiper	Astrid Lægreid	Rune Nydal
Steven Vercruysse	Liv Thommesen	Ane Møller Gabrielsen
Wim De Mulder	Åsmund Flobak	Anamika Chatterjee
Vladimir Mironov	Miguel Vazquez	
Vasundra Touré	Marcio Acencio	
Stian Holmås	Barbara Niederdorfer	
Rafel Riudavets	Evelina Folkesson	
	Kathleen Heck	

NTNU
Norwegian University of Science and Technology

Acknowledgments

institut Curie

EMBL-EBI

ENS
ÉCOLE NORMALE
SUPÉRIEURE

Funding:

NTNU Helse

ERA CO SYS MED