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

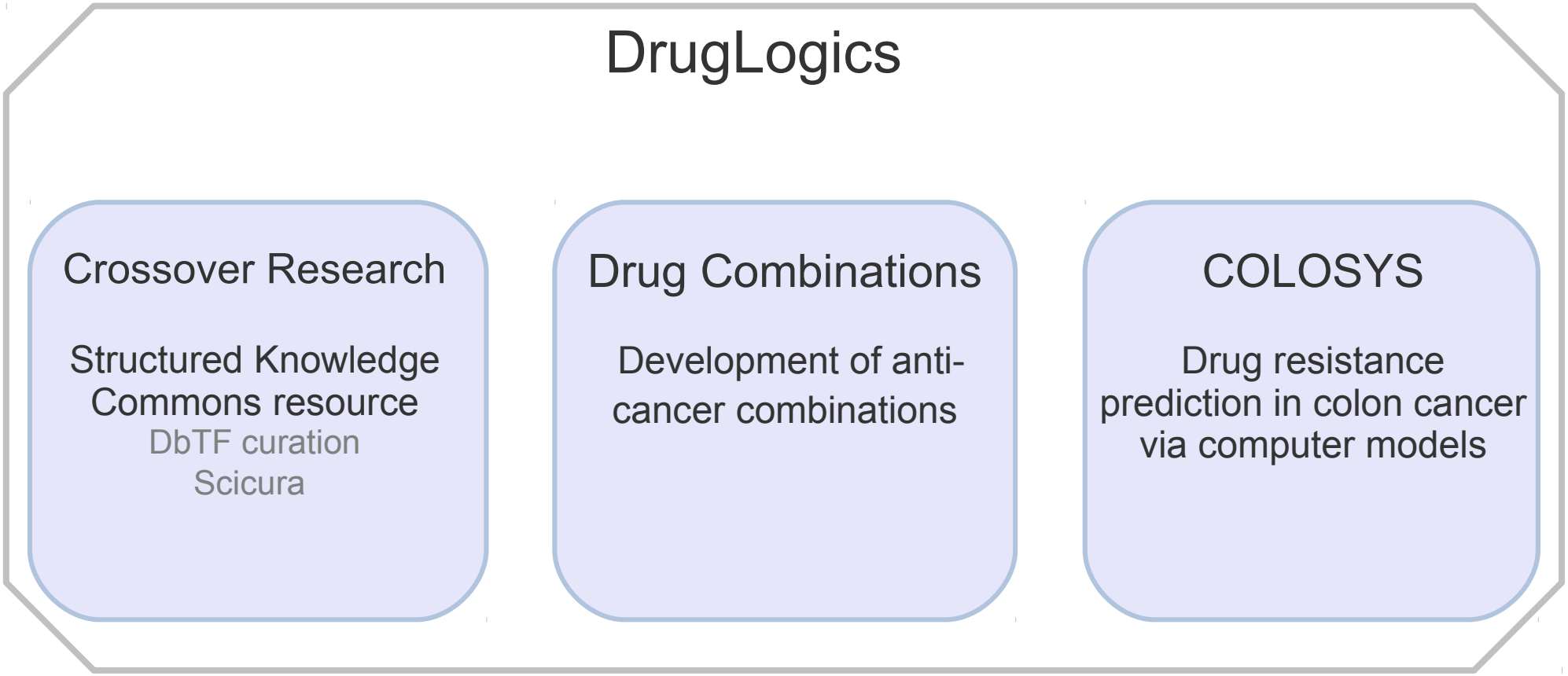
In collaboration with
EBI – Reactome, PSI-MI
Institut Curie
ENS

Norwegian University of Technology and Sciences

vasundra.toure@ntnu.no
@Vasundra_T
@DrugLogics

The DrugLogics project

Towards the development of precision and personalised medicine



My tasks within the DrugLogics/NTNU-Health project

Facilitate the process of building biological models with causal statements

prior knowledge

REACTOME

ACSN

Signor

IntAct

"CausalDB"

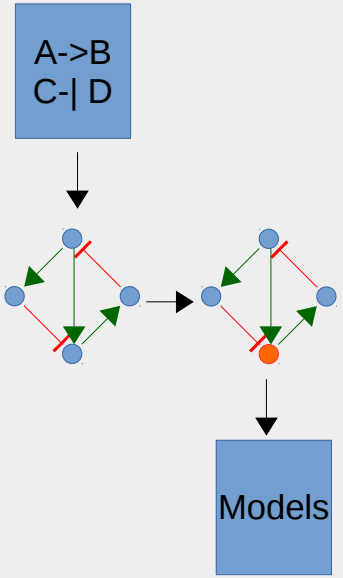
Representation of causal statements

Extraction from prior knowledge

Availability of the data

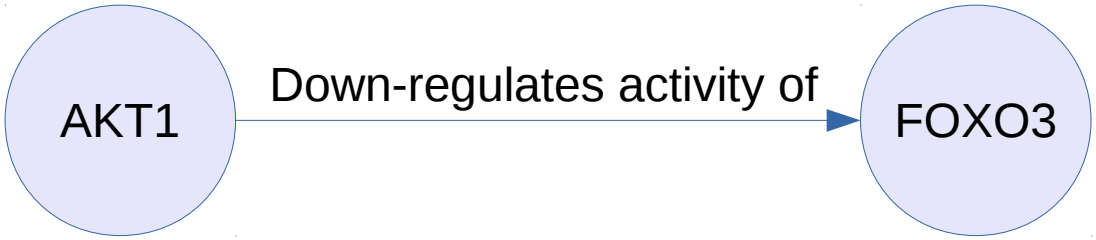
DrugLogics' pipeline

Logical modeling to predict drug targets and synergies



Representation of causal statements

Causal interaction between two biological entities (gene, RNA, protein, complexes, etc…)



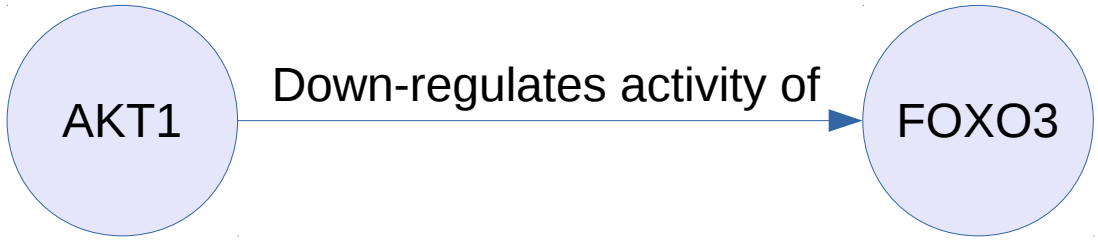
Representation of causal statements

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?

Representation of causal statements

Research ideas on information that should be ideally encoded

Entity – Source (Regulator) / Target (Regulated)

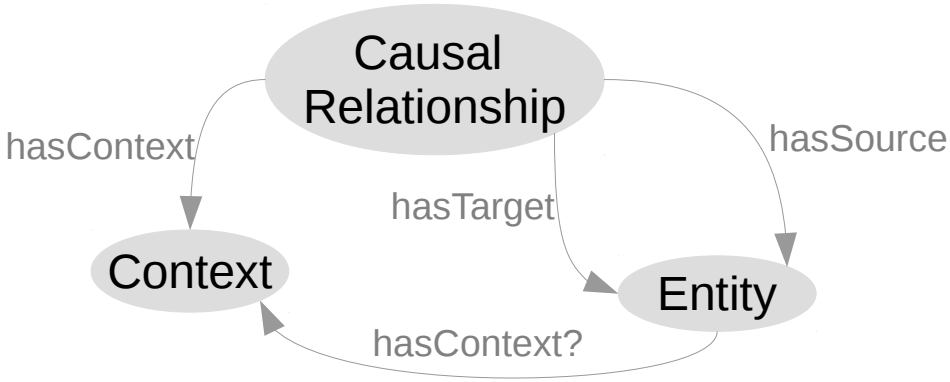
- ID – ex:causalDB:FOXO3
- Reference ID - HGNC; Uniprot; Entrez
  - *For Complex: ComplexPortal ID?*
  - *For Families?*
- Name – ex:FOXO3
- Molecule type - gene, RNA, protein, complex
- Acting entity – Reference ID
- Molecular function – GO:MF
- State – active / inactive

Causal Relationship

- Regulation type – down-regulates
- Mechanism - PSI MOD?
- Modified residue – Tyr@P202
- Interaction depth – 0 (direct); 1; 2; etc…
- Text – Scicura (http://scicura.org/info.html)
- Provenance – ex:Reactome
- Evidence – ECO
- Confidence score?

Context

- Species – TaxID
- Tissue type – Brenda Tissue Ontology (BTO), Uberon?
- Cell type – BTO, Cell Line Ontology (CLO)?
- Experimental conditions – if evidence is experimental
- Tissue / Cell state



Controlled Vocabulary and Ontologies – essential to make data sustainable, shareable and interoperable

Representation of causal statements

Controlled vocabulary / ontologies for representing causal interaction type

Effect	Gene Ontology	PSI-MI causal interaction	Relation Ontology	BEL statement	IntAct	Signalink
Positive regulation	positively regulates	up-regulates	activity directly positively regulates activity of	increases	activates	stimulation
Negative regulation	negatively regulates	down-regulates	activity directly negatively regulates activity of	decreases	inhibits	inhibition

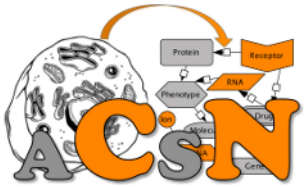
→ Need unification

Extraction from prior knowledge

Aggregation of causal data from several existing resources

REACTOME

Pathways, reactions



Pathways of cancer related signaling networks

Signor

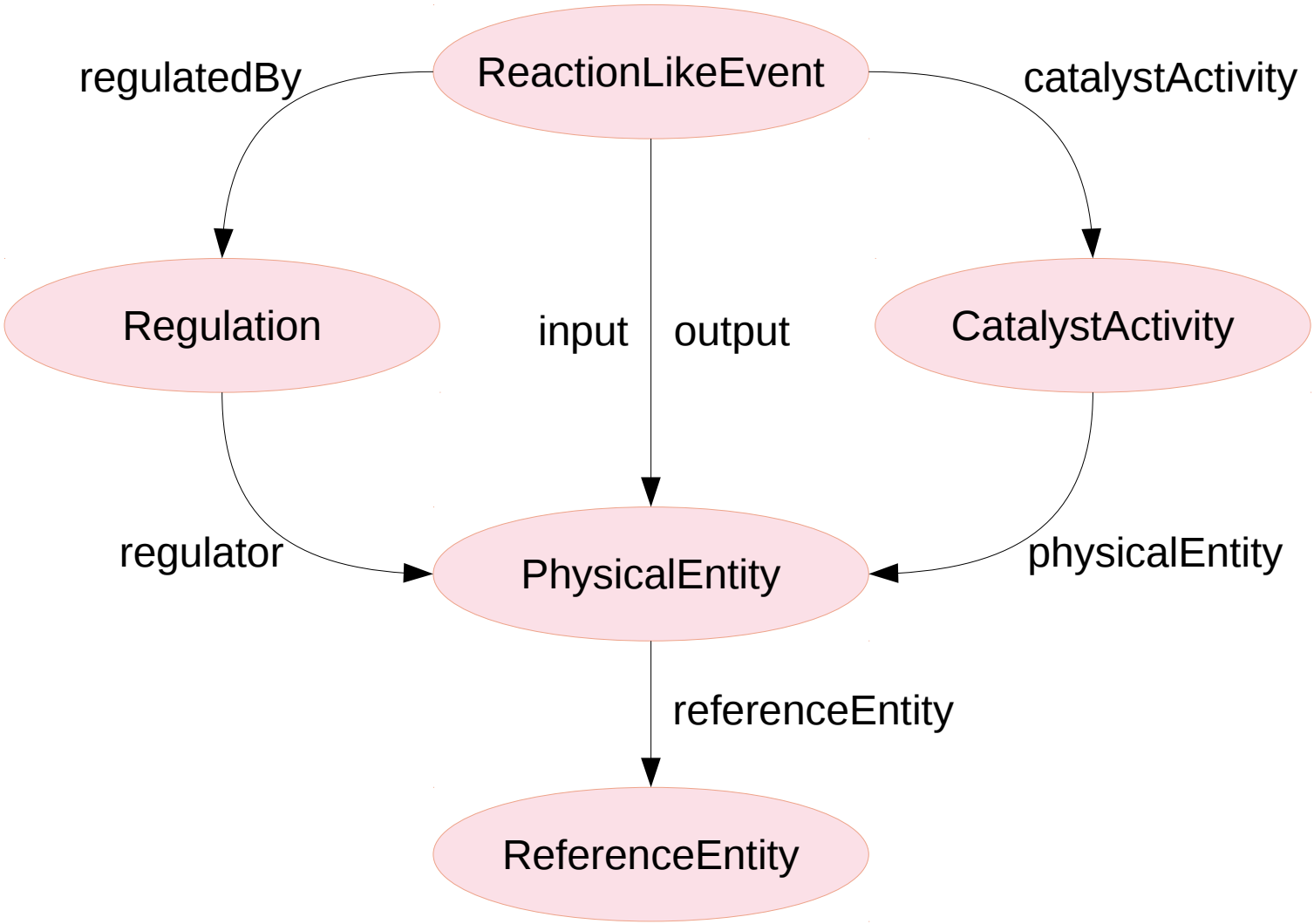
DB of causal interactions

IntAct

DB of molecular interactions

Example: extraction from Reactome

Reactome data model extraction using Neo4j and Cypher Query language



Example: extraction from Reactome

Reactome data model extraction using Neo4j and Cypher Query language

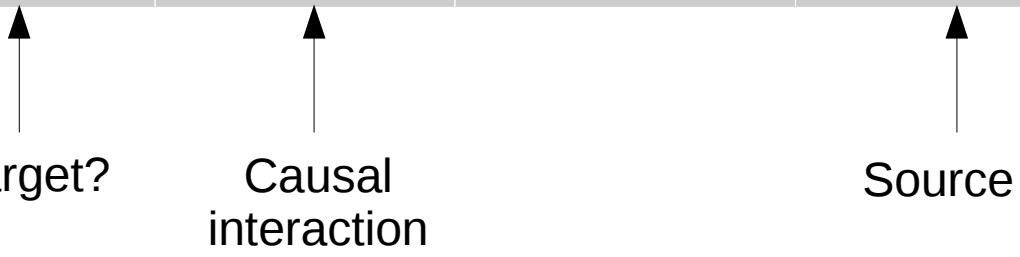
Example: Get all reactions regulated by a physical entity or catalysed by a catalyst activity

```
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

Example: 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]

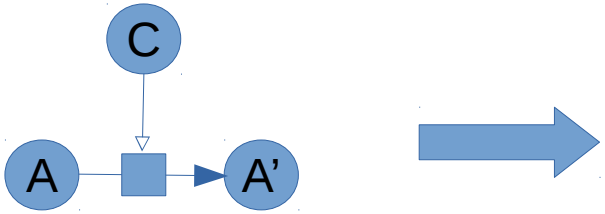


Example: questions / issues raised

- Exclude trivial molecules

malate + NAD+ <=> oxaloacetate + NADH + H+

- Missing IDs for the modified mechanism type
- How to transform a reaction network to causal statements?



C -| A
C -> A’
A -| A’
A’ -| A

- Definition of necessary and sufficient contextual information

→ MICAST: Minimum Information for representing Causality Statements?

Current work

- Collaboration with Curie and ENS - Paris
  - Define the representation of causal statements
  - Extraction of causal statements from ACSN
  - Consensus representations with GO, PSI-MI, COLOMOTO
  - Standardisation of the pipeline → SBML-Qual?

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 Liv Thommesen	Rune Nydal
Steven Vercruysse Wim De Mulder Vladimir Mironov	Åsmund Flobak Marcio Acencio	Ane Møller Gabrielsen
Vasundra Touré	Barbara Niederdorfer Evelina Folkesson	Anamika Chatterjee
Stian Holmås Rafel Riudavets	Kathleen Heck	

NTNU
Norwegian University of Science and Technology

Funding:

NTNU Helse

ERACOSYSMED