Leibniz Universität Hannover

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

Causal Relationship over Knowledge Graphs

Care KG

Hao Huang
supervised by Prof. Dr. Maria-Esther Vidal

Leibniz University of Hannover, Germany
TIB Leibniz Information Centre for Science and Technology

TrustKG

funded by

Leibniz Association

TIB

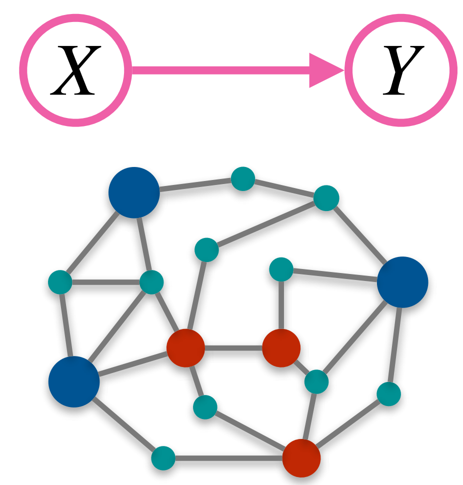
Ph.D. student at LUH
starting at 2022.01.01

Leibniz
Universität
Hannover

Research Interests:

Knowledge Graphs (KGs);

Causal inference & representation over KGs;

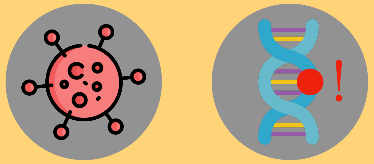


Outline

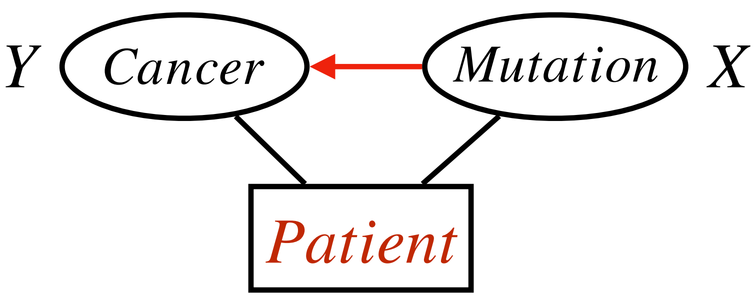
- Motivating Example
- Related Work

- Problem Statement
- CareKG-Proposed Solution
- Experimental Study

- Lessons Learned & Timeline
- Conclusions & Future Work



0	0
0	0
1	0
1	0
1	1
1	1
0	1
0	1



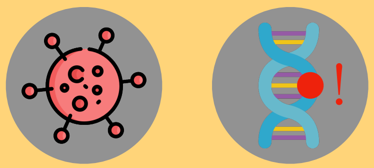
Homogenous Units U_{XY}

Contingency Matrix

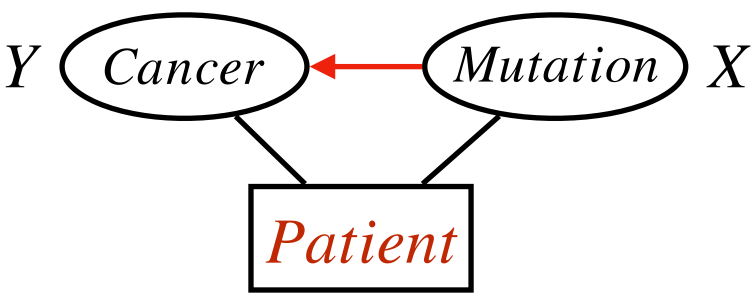
	Cancer	No Cancer
Mutation	2	2
No Mutation	2	2

Mutation $\not\rightarrow$ Cancer

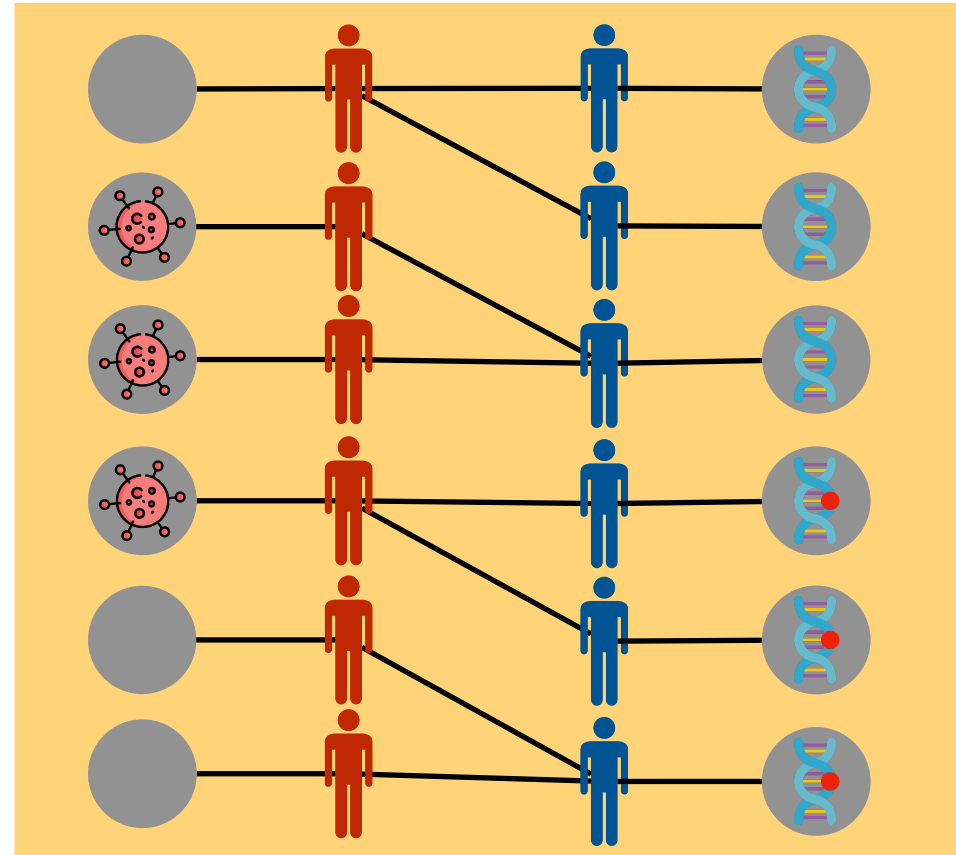
Inherited mutation

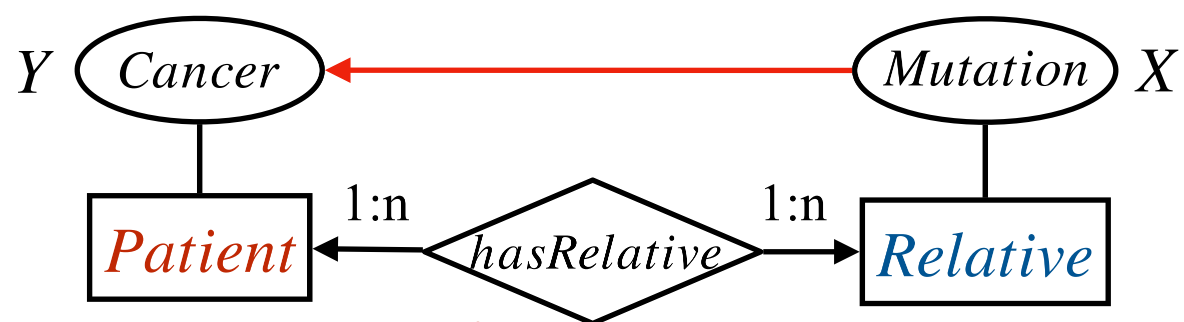


0	0
0	0
1	0
1	0
1	1
1	1
0	1
0	1



Homogenous Units U_{XY}





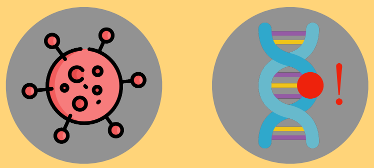
Response Units U_Y

Treated Units U_X

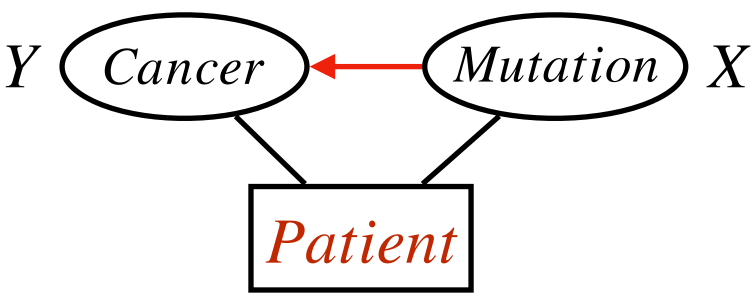
Heterogeneity

	Cancer	No Cancer
Mutation	2	2
No Mutation	2	2

Mutation $\not\rightarrow$ Cancer

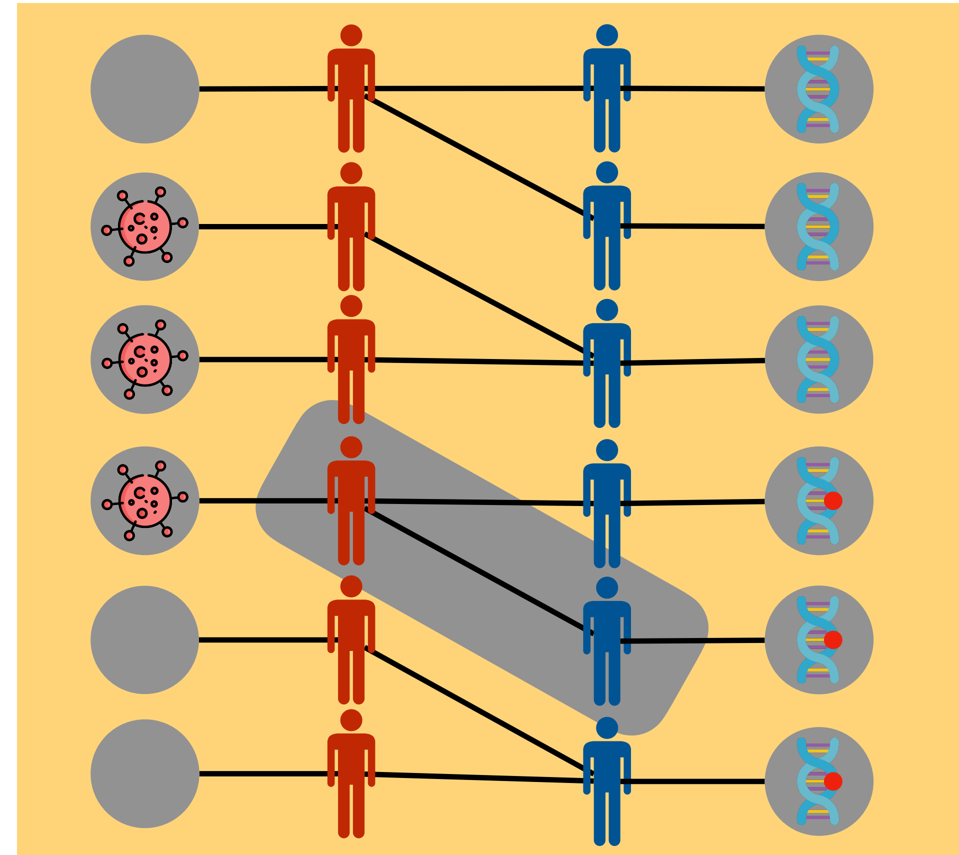


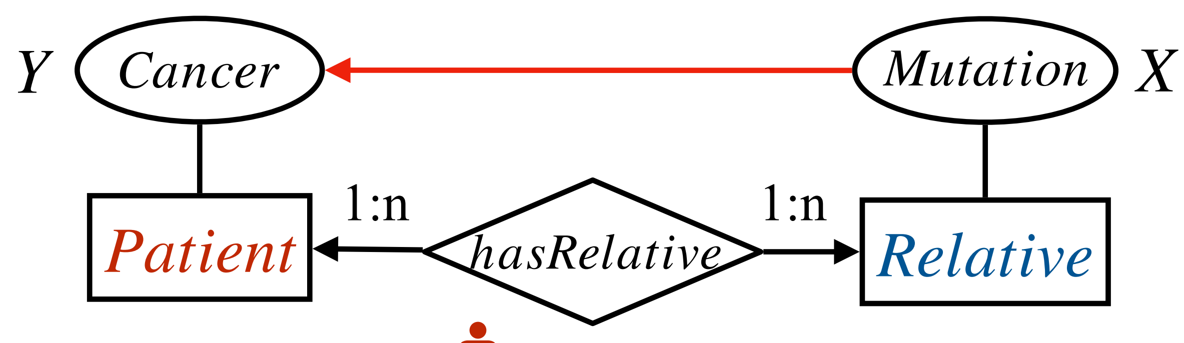
0	0
0	0
1	0
1	0
1	1
1	1
0	1
0	1



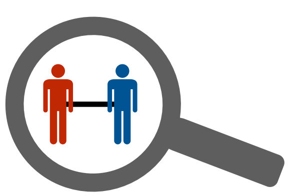
Homogenous Units U_{XY}

Inherited mutation





Response Units U_Y Treated Units U_X

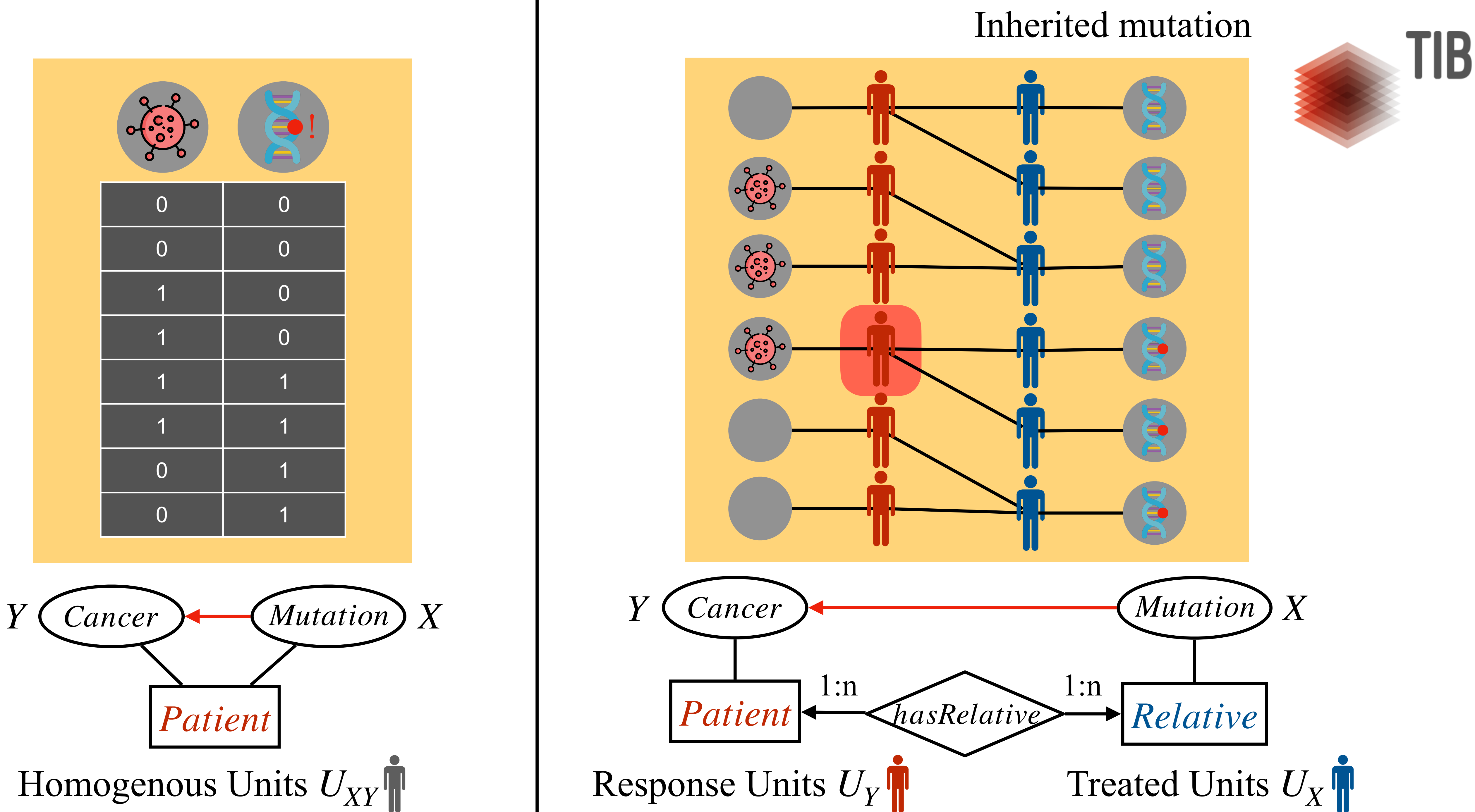


	Cancer	No Cancer
Mutation	2	2
No Mutation	2	2

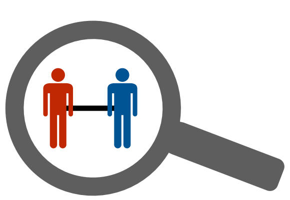
Mutation $\not\rightarrow$ Cancer

Perspective: a relational path $\mathscr{P}$; a sequence of classes / relations.

Unit: an instance of the relational path



0	0
0	0
1	0
1	0
1	1
1	1
0	1
0	1

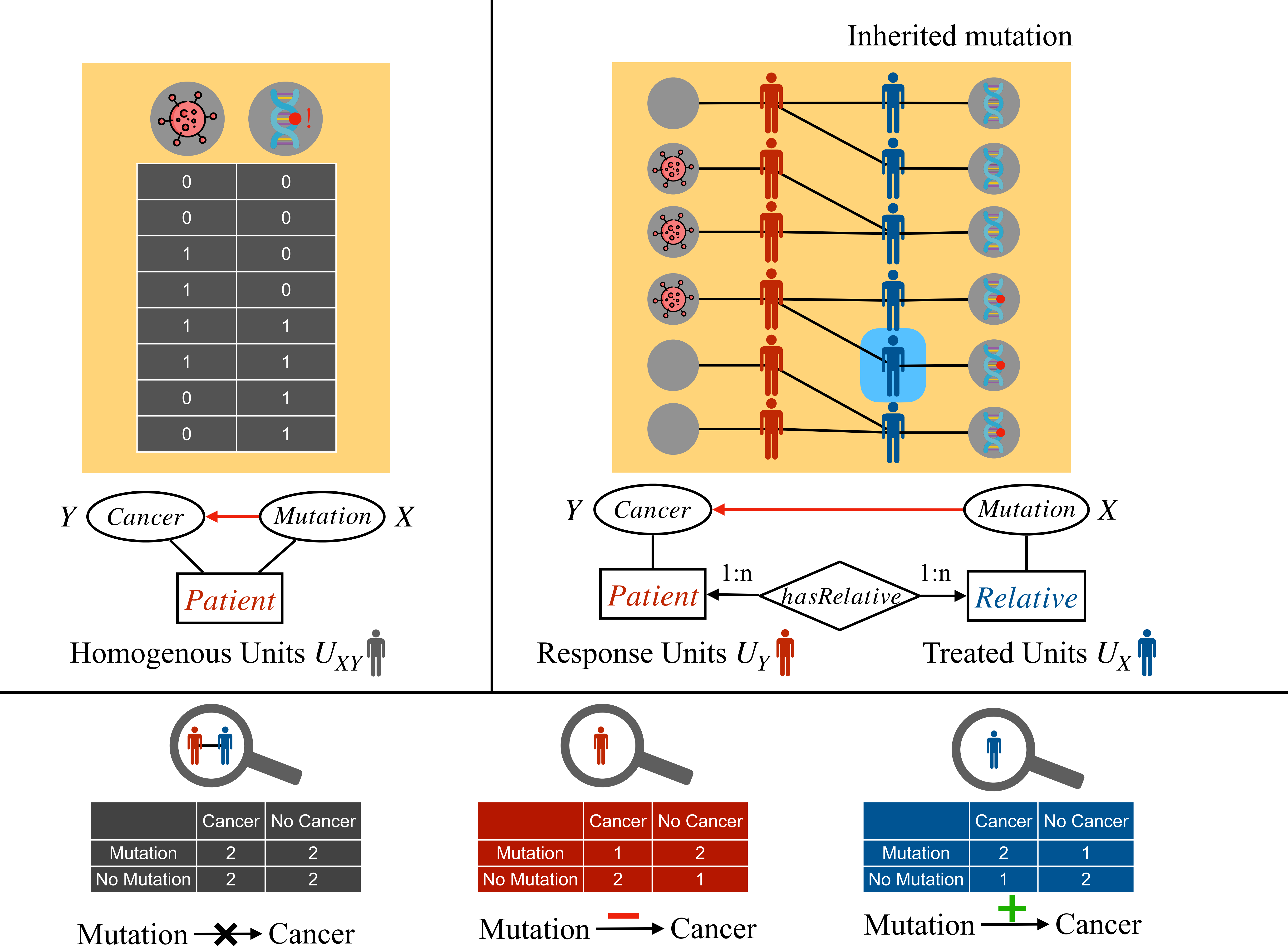


	Cancer	No Cancer
Mutation	2	2
No Mutation	2	2

Mutation $\not\rightarrow$ Cancer

	Cancer	No Cancer
Mutation	1	2
No Mutation	2	1

Mutation $\xrightarrow{-}$ Cancer

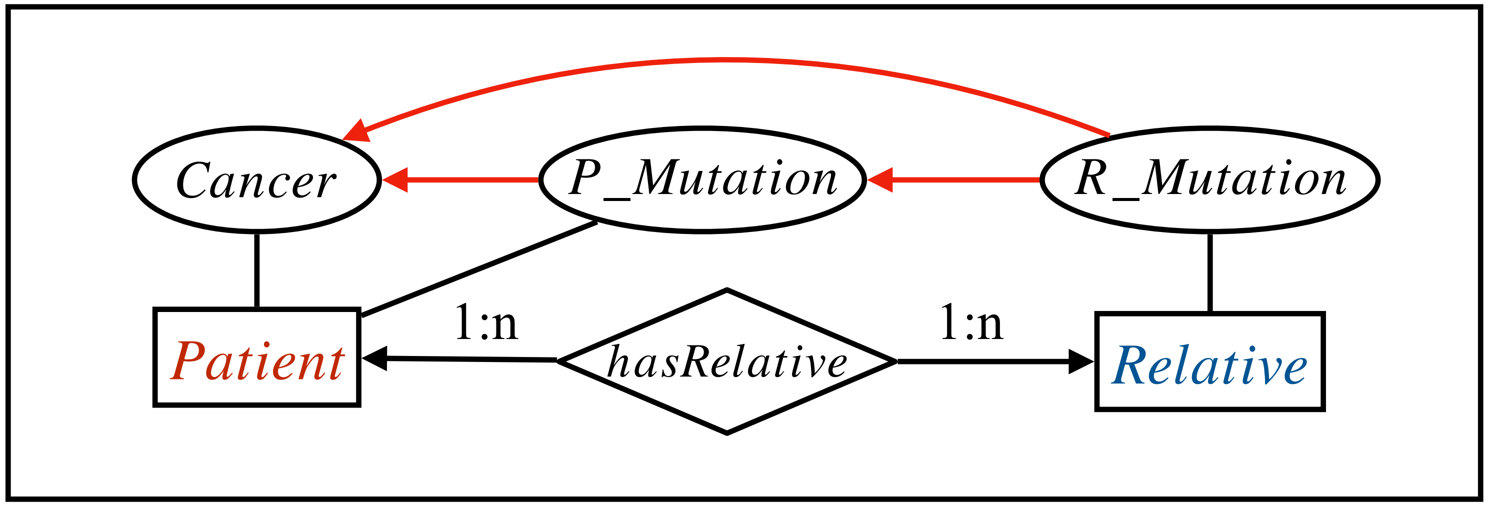


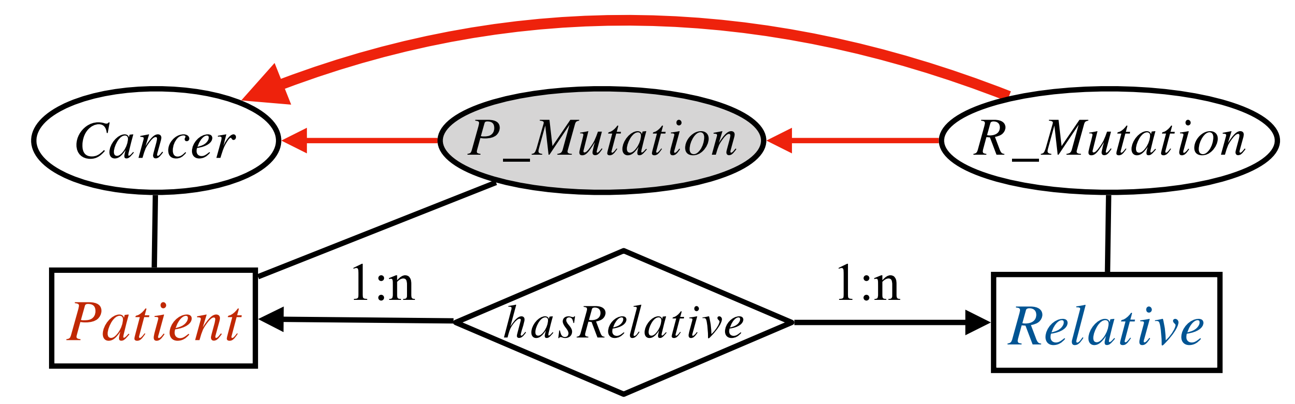
0	0
0	0
1	0
1	0
1	1
1	1
0	1
0	1

	Cancer	No Cancer
Mutation	2	2
No Mutation	2	2

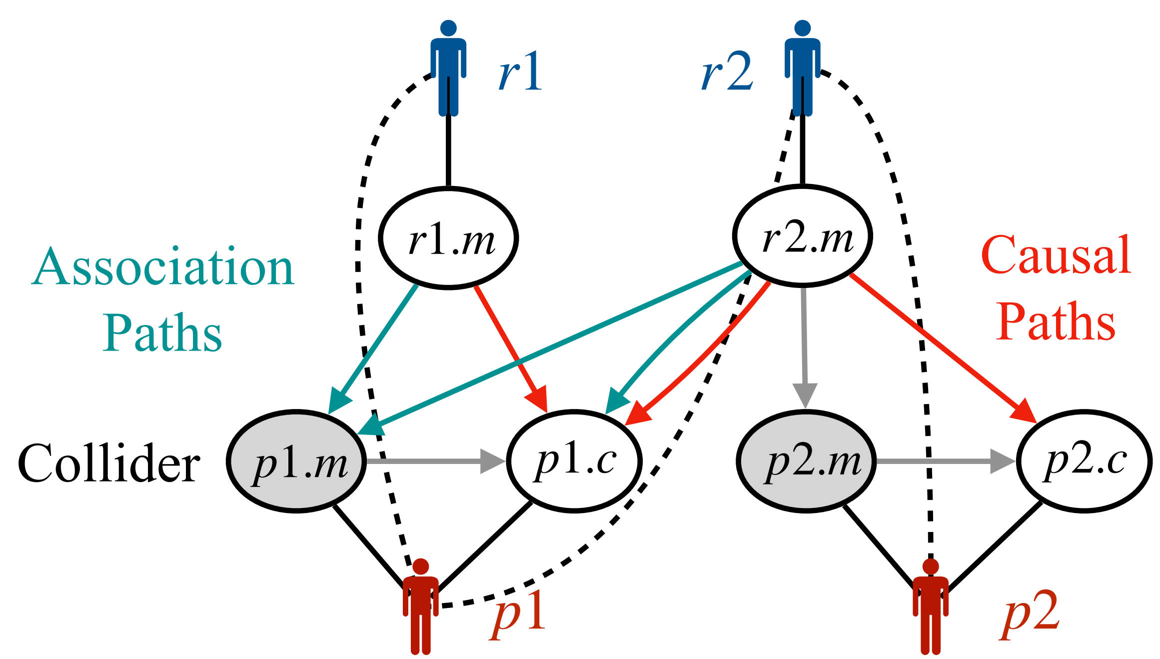
	Cancer	No Cancer
Mutation	1	2
No Mutation	2	1

	Cancer	No Cancer
Mutation	2	1
No Mutation	1	2

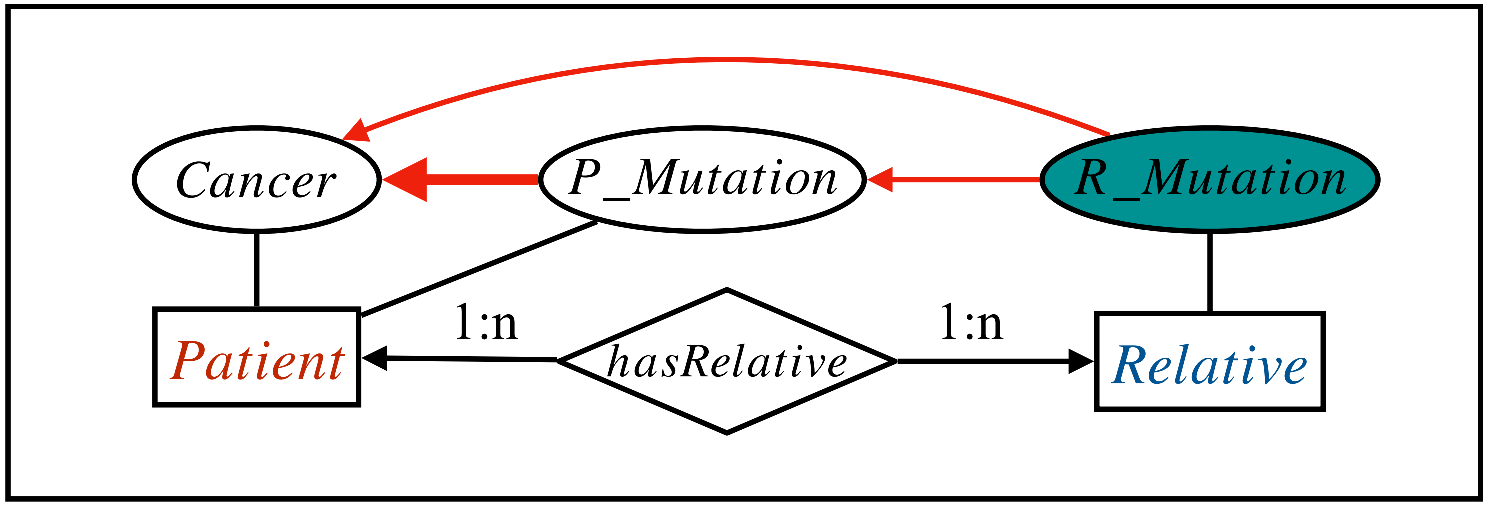

Cancer
P_Mutation
R_Mutation
Patient
1:n
hasRelative
1:n
Relative



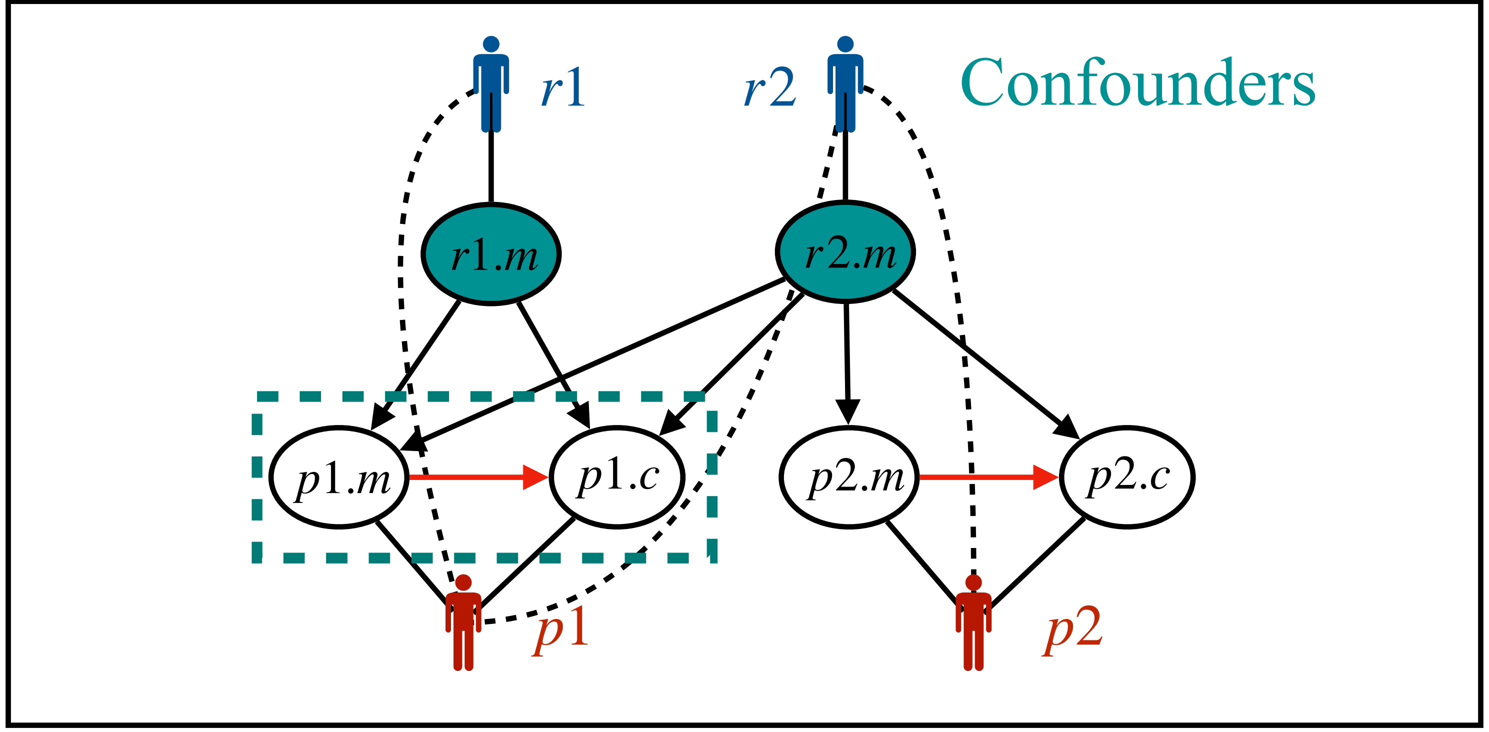
Cardinality (class : relation 1:1 vs 1:n)



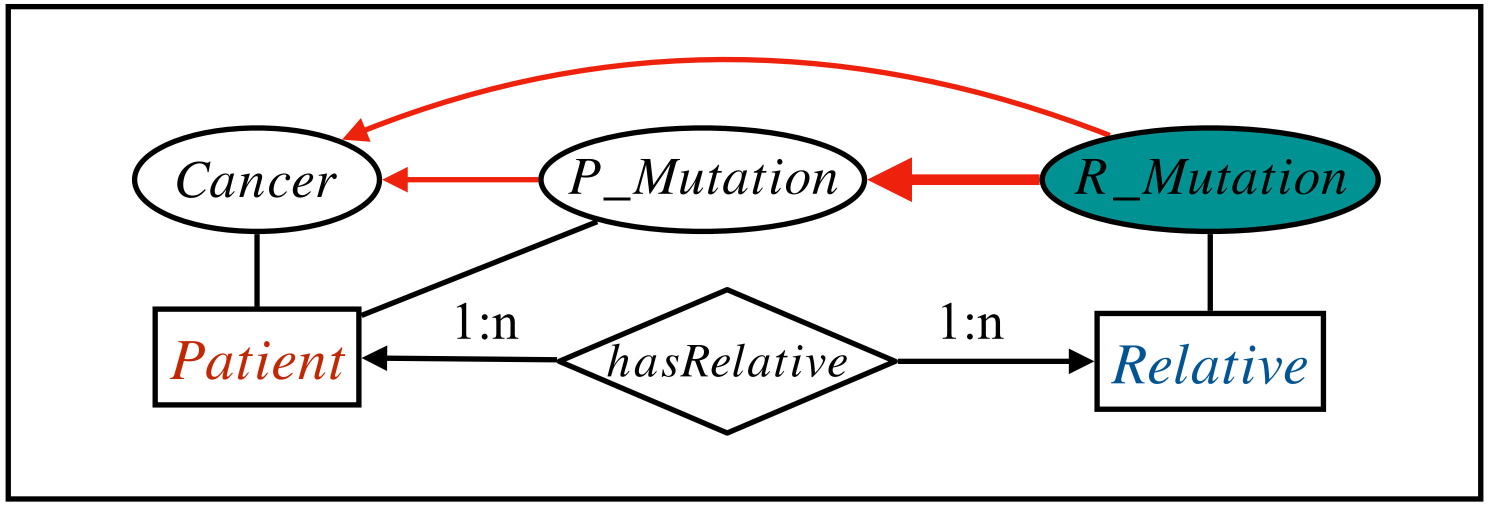
Ignoring cardinality prevents differentiating association paths from causal paths



Cardinality (class : relation or class : attribute = 1:n)



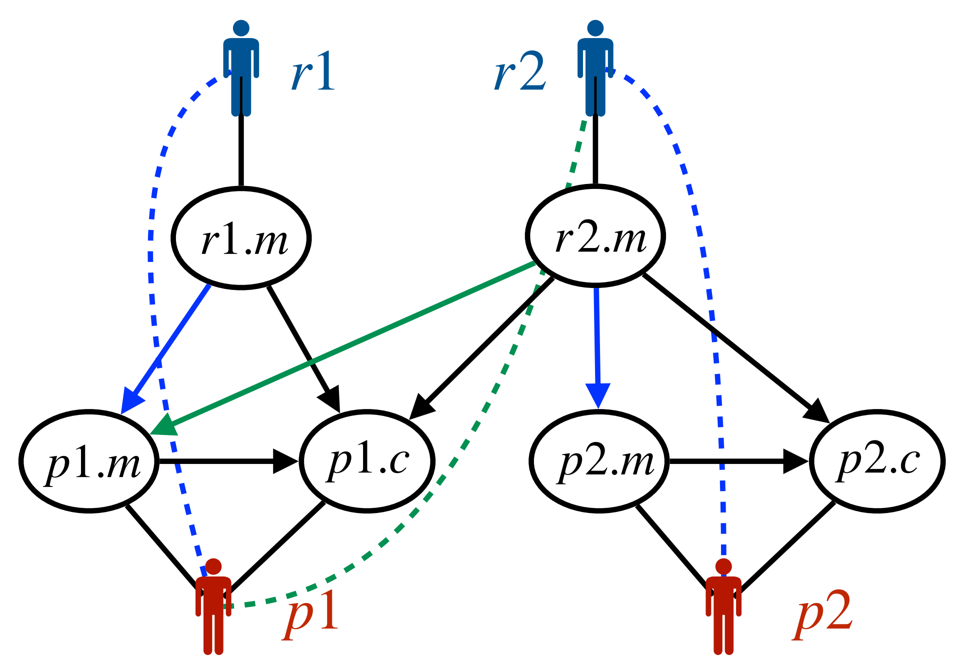
Ignoring cardinality prevents identifying multiple confounders, or causes and effects



$hasRelative = hasDRelative \cup hasIDRelative$

$hasDRelative \cap hasIDRelative = \varnothing$

$CE_{hasRelative} = f(CE_{hasDRative}, CE_{hasIDRelative})$



Ignoring subsumption, disjoint among concepts will miss the chance of sufficient inference

Related Work

- Causality over Relational Data

  Causal Relational Learning (CaRL) [1]

  D-seperation in Relational Data [2]

- Causality over Knowledge Graph

  Causal KG [3]

  Differential Causal Rule Mining [4]

[1] Babak Salimi, Harsh Parikh, Moe Kayali, Lise Getoor, Sudeepa Roy, and Dan Suciu. 2020. Causal relational learning. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 241–256.

[2] Maier, Marc, Katerina Marazopoulou, and David Jensen. "Reasoning about independence in probabilistic models of relational data." *arXiv preprint arXiv:1302.4381* (2013).

[3] Jaimini, Utkarshani, and Amit Sheth. "CausalKG: Causal Knowledge Graph Explainability using interventional and counterfactual reasoning." *IEEE Internet Computing* 26.1 (2022): 43-50.

[4] Simonne, Lucas, et al. "Differential Causal Rules Mining in Knowledge Graphs." *Proceedings of the 11th on Knowledge Capture Conference*. 2021.

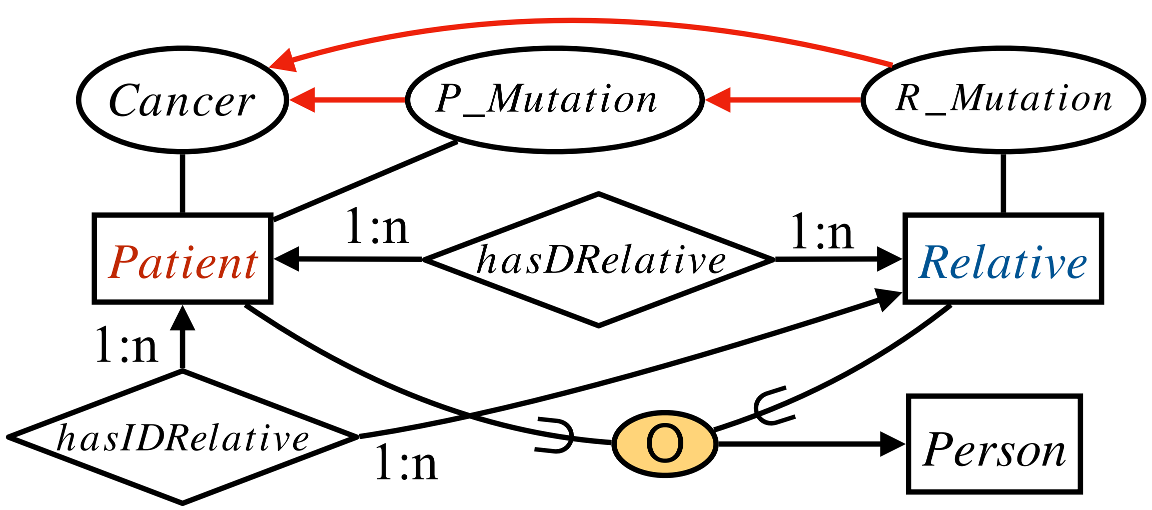
Related Work

- Causality over Relational Data
  Causal Relational Learning (CaRL) [1]

Relational Causal Rule of CaRL

$A[\mathbf{Y}] \Leftarrow A_1[\mathbf{X}_1], \ldots, A_k[\mathbf{X}_k]\ WHERE\ Q(\mathbf{W})\ (Y, X_k \in \mathbf{W})$

$Cancer[P] \Leftarrow R\_Mutation[R]\ WHERE\ hasDRelative(P, R)$



Lacks of :

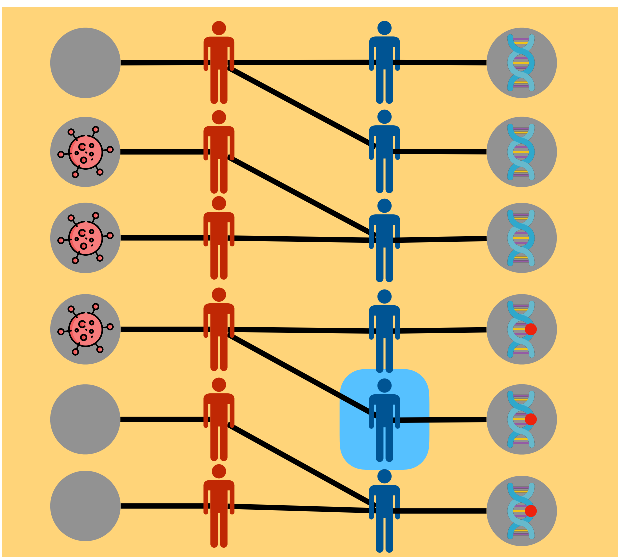
Class Disjoint	Patient - Relative
Cardinality Constraint	Patient: hasDRelative = 1:1 or 1:n ?
Integrity Constraint	hasDRelative ∩ hasIDRelative = ∅

[1] Babak Salimi, Harsh Parikh, Moe Kayali, Lise Getoor, Sudeepa Roy, and Dan Suciu. 2020. Causal relational learning. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 241–256.

Related Work

Causal Query of CaRL

$Agg(A[\mathbf{Y}]) \Leftarrow A[\mathbf{X}]?\ (WHEN\ \ < cnd >\ \ PEERS\ TREATED)$



Interference between units

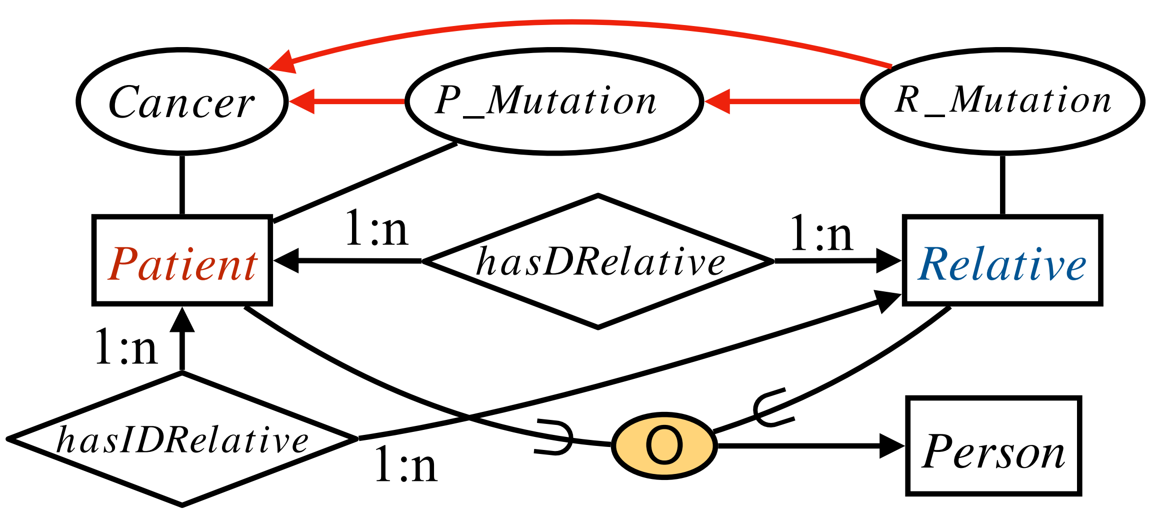
peer cause effect

isolated cause effect

Aggregation Function: AVG, MAX, MIN, …

Condition: Less / More than k | At (Least / Most) k | ALL | None

$AVG\_Cancer[P] \Leftarrow R\_Mutation[R]?$



Lacks of :

Perspective	
The Relational Path	[Patient, hasDRelative, Relative] or [Patient, hasIDRelative, Relative]
Constraint	Relative - Patient
Peer Definition	Relative who share same patients or same Mutation?

Related Work

- Causality over Knowledge Graph

  Differential Causal Rule Mining [4]

Target class & Strata: $C(X_1) \wedge C(X_2)$ $[\wedge ST_i(X_1) \wedge ST_i(X_2)]$

Treatment: $\bigwedge_k p_k(X_1, T_1) \wedge p_k(X_2, T_2) \wedge compare(T_1, T_2)$

Outcome: $p(X_1, O_1) \wedge p(X_2, O_2) \Rightarrow \underbrace{compare(O_1, O_2)}_{\text{head}}$

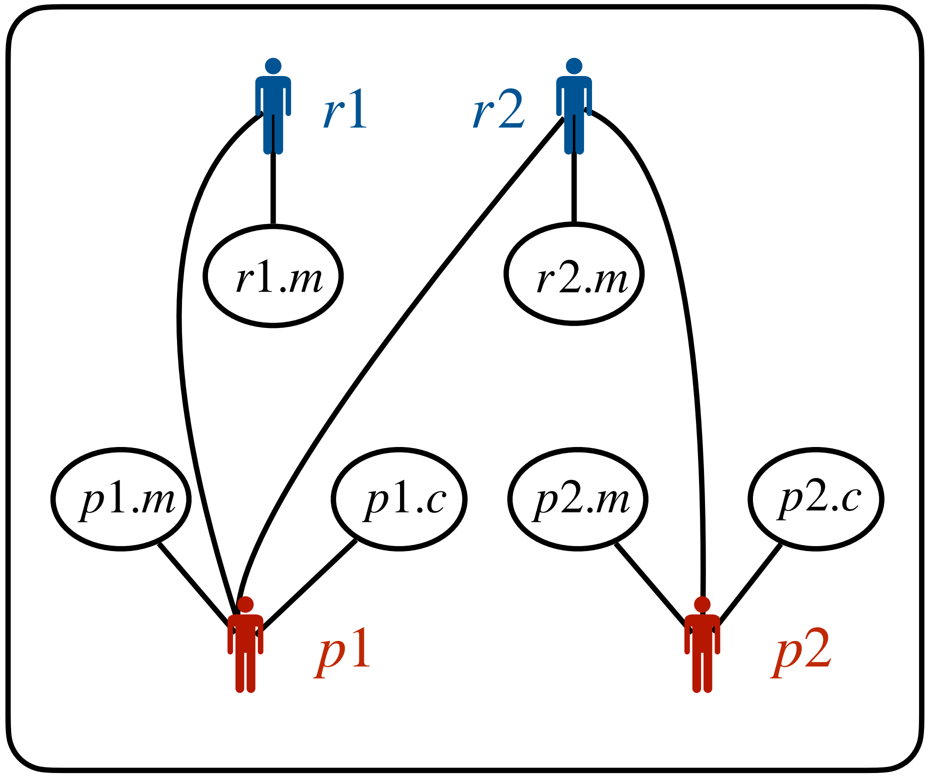
$Patient(P_1) \wedge Patient(P_2) \wedge Mutation(P_1, M_1) \wedge Mutation(P_2, M_2) \wedge greatThan(M_1, M_2)$
$\wedge\ Cancer(P_1, C_1) \wedge Cancer(P_2, C_2) \Rightarrow greatThan(C_1, C_2)$

+: offers context (strata) where causal rule hold.
+: allows multiple attributes as treatment.

- : lacks of consideration of semantics of KGs, e.g. cardinality, integrity constraint, etc.
- : offer fixed perspective from target class.

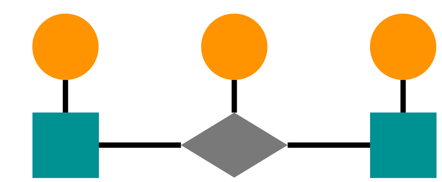
[4] Simonne, Lucas, et al. "Differential Causal Rules Mining in Knowledge Graphs." *Proceedings of the 11th on Knowledge Capture Conference*. 2021.

Problem Definition



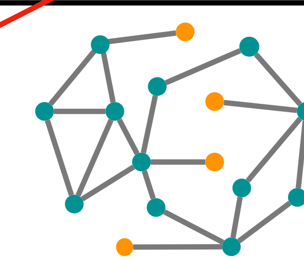
Classes Relations Attributes

Ontology: $\mathcal{O} =< \mathcal{C}, \mathcal{R}, \mathcal{A} >$



Entities & Literals Properties Triples

Knowledge Graph: $KG = <V, E, T>$



A surgery is only applicable to Patients in stage I

Axioms: ζ

Problems:

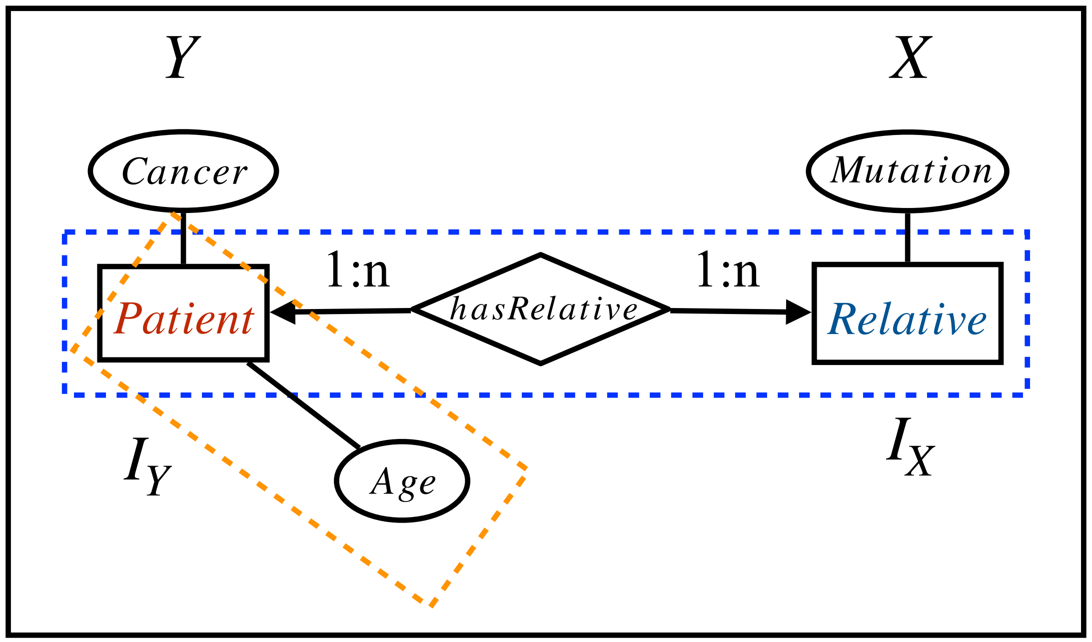
1. How to represent causal relations in KGs?
2. How to formulate causal queries using semantics?

Causal Relation Representation

Ontological Causal Rule (CareKG)

$$Y[I_Y] \Leftarrow X[I_X] \; WHERE \; \underbrace{P(I_X, I_Y)}_{\text{Relational Path}}, \underbrace{CTX(I_X, I_Y)}_{\text{Context}}$$

Relational Path Context



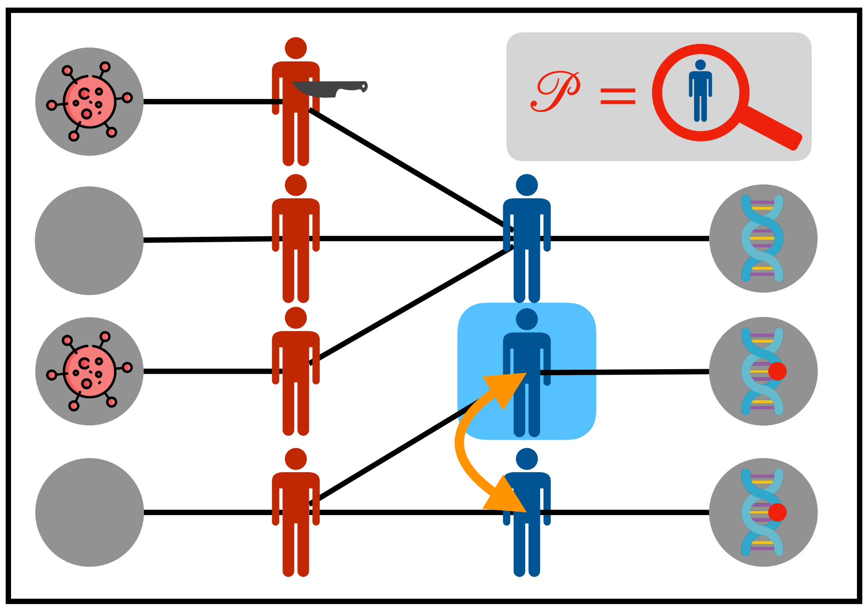
I_X is the concept (class or relation) of cause X, I_Y is the concept of effect Y

Property Path Patient.hasAge.Age=Young $\in$ Context $CTX(I_X, I_Y)$

Causal Query

$\text{FUN}(Y[I_Y]) \Leftarrow \text{FUN}(X[I_X])?$ FROM PATH P, UNDER PERSPECTIVE $\mathscr{P}$, SUBJECT TO $\langle axiom \rangle$, WHEN $\langle cnd \rangle$

FUN: (1) AVG, MAX, SUM, COUNT, COMB…; (2) DISC, REPLC

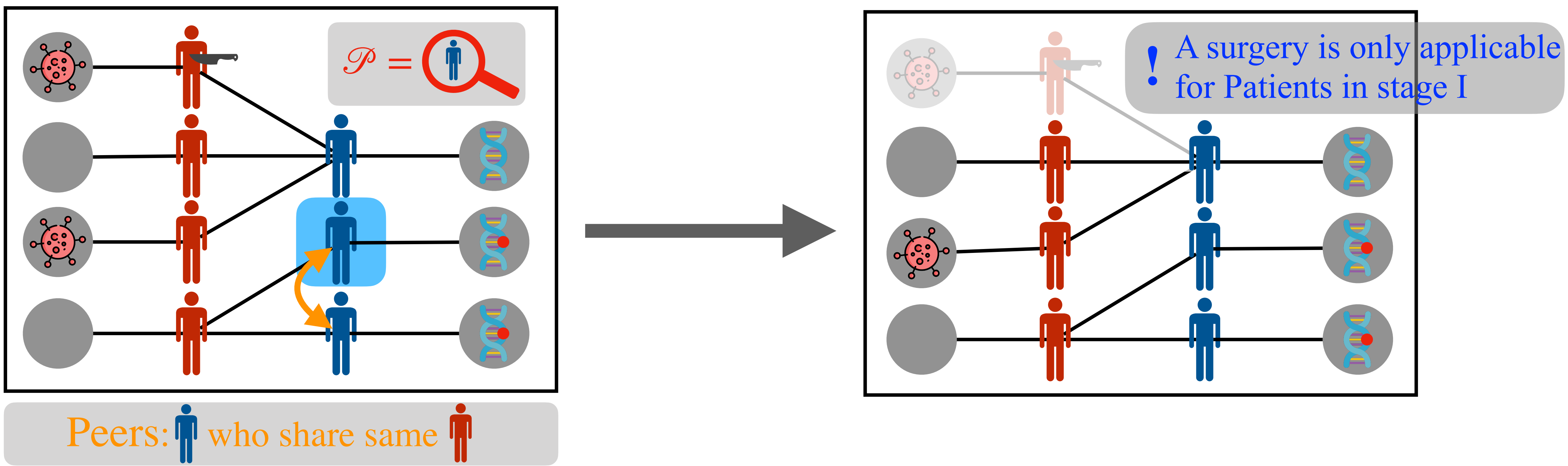


Peers: who share same

Causal Query

$$\mathrm{FUN}(Y[I_Y]) \Leftarrow \mathrm{FUN}(X[I_X])?\ \mathrm{FROM\ PATH}\ P,\ \mathrm{UNDER\ PERSPECTIVE}\ \mathscr{P},$$
$$\mathrm{SUBJECT\ TO}\ \langle axiom \rangle,\ \mathrm{WHEN}\ \langle cnd \rangle$$

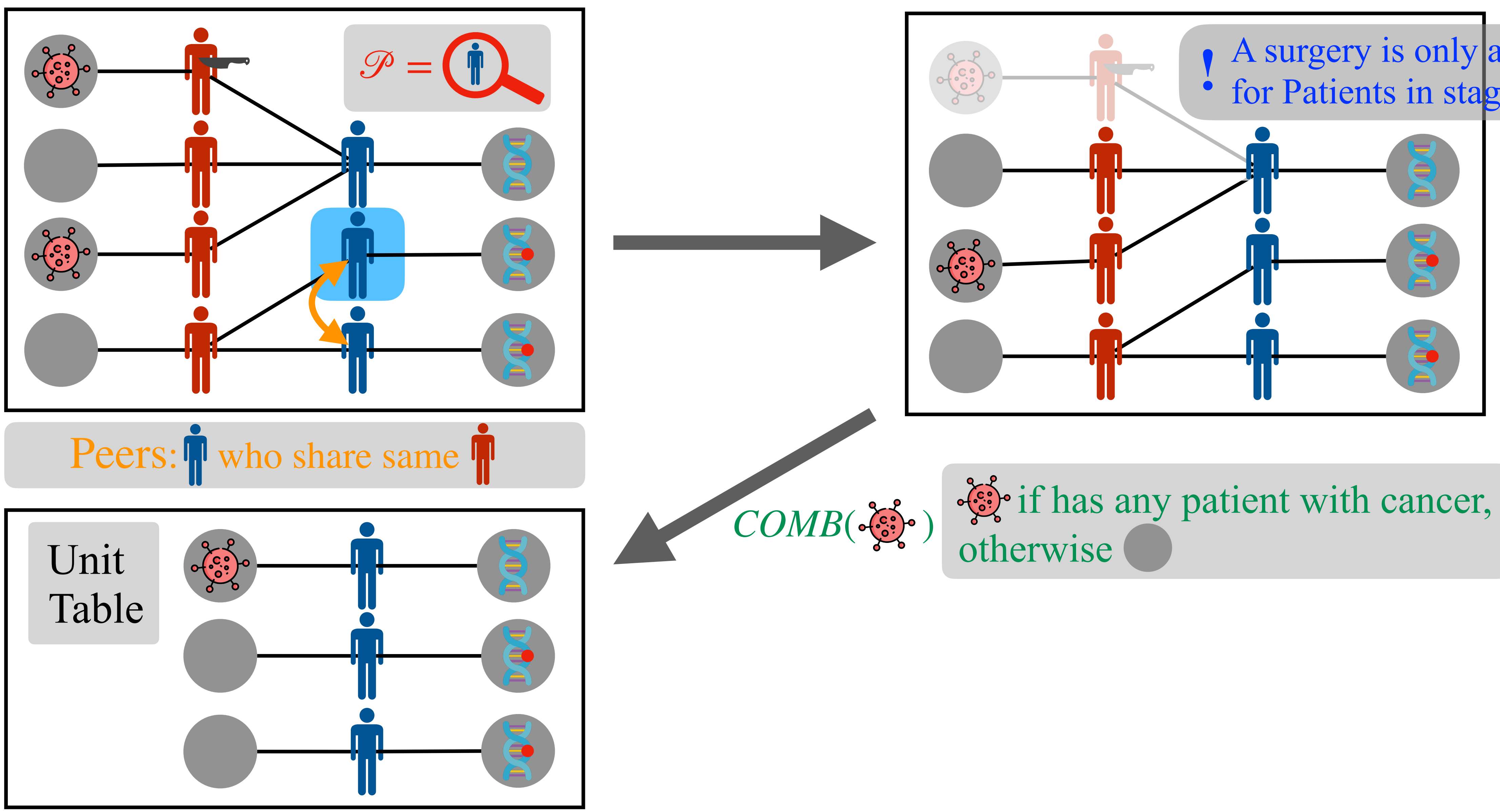
FUN: (1) AVG, MAX, SUM, COUNT, COMB…; (2) DISC, REPLC



Causal Query

$\mathrm{FUN}(Y[I_Y]) \Leftarrow \mathrm{FUN}(X[I_X])?$ FROM PATH P, UNDER PERSPECTIVE $\mathscr{P}$, SUBJECT TO $\langle axiom \rangle$, WHEN $\langle cnd \rangle$

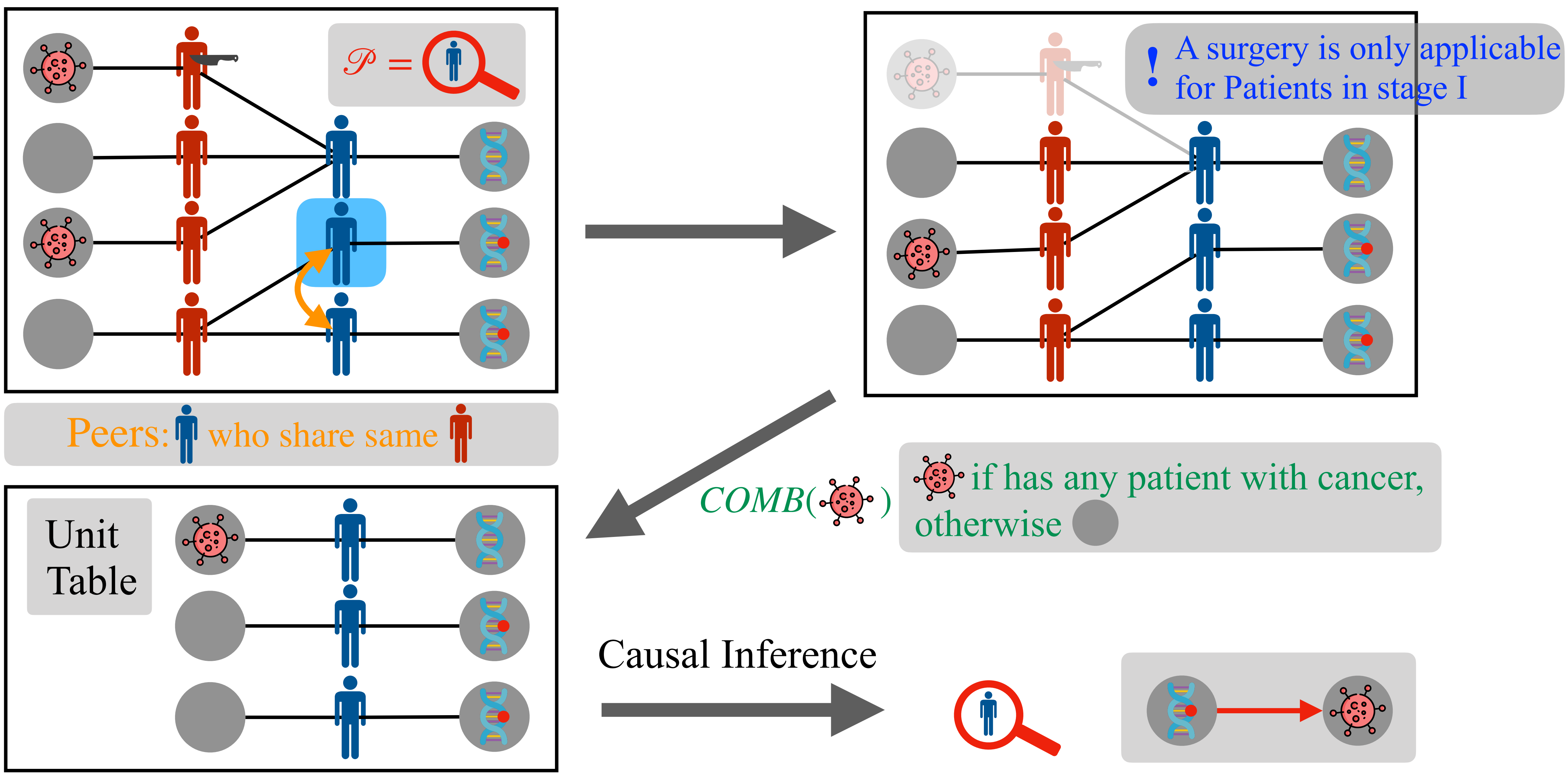
FUN: (1) AVG, MAX, SUM, COUNT, COMB…; (2) DISC, REPLC



Causal Query

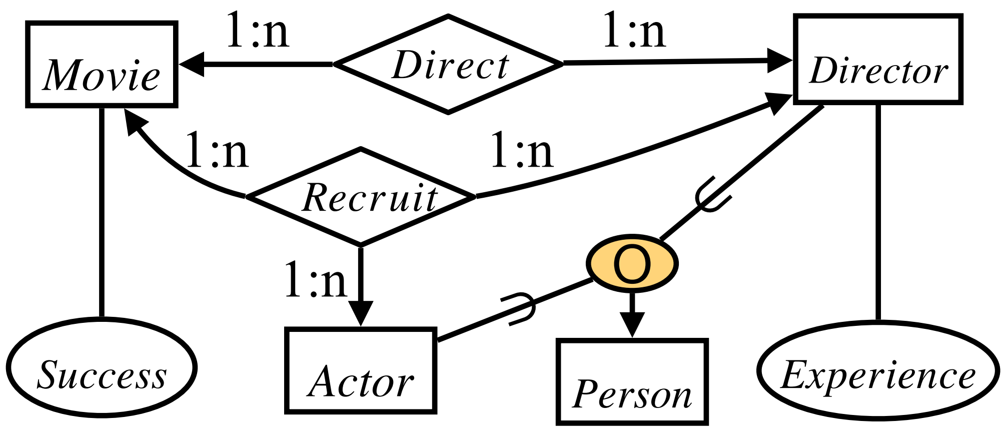
$\mathrm{FUN}(Y[I_Y]) \Leftarrow \mathrm{FUN}(X[I_X])?$ FROM PATH P, UNDER PERSPECTIVE $\mathscr{P}$,
SUBJECT TO $\langle axiom \rangle$, WHEN $\langle cnd \rangle$

FUN: (1) AVG, MAX, SUM, COUNT, COMB…; (2) DISC, REPLC



Setting of Experiment 1

Experimental KG:
A synthetic KG "Ex - Movie"



Director	300
Actor	3043
Movie	582

$$M.Suc = 0.5 \times AVG(D_1.Exp) + 0.5 \times AVG(D_2.Exp)$$
$$(D_1 \in Director, D_2 \in Director \cap Actor)$$

Objective:
Proof the expressiveness of CareKG is better than CaRL[1]

Causal Rules

Relational Causal Rule (CaRL)

$$Success[M] \Leftarrow Experience[D]\ WHERE\ Direct(D, M)$$
$$Success[M] \Leftarrow Experience[D]\ WHERE\ Recruit(D, A, M)$$

Ontological Causal Rule (CareKG)

$$Success[M] \Leftarrow Experience[D]\ WHERE\ [Director(D), Direct(D, M), Movie(M)]$$
$$Success[M] \Leftarrow Experience[D]\ WHERE\ [Director(D), Recruit(D, A, M), Movie(M)]$$

[1] Babak Salimi, Harsh Parikh, Moe Kayali, Lise Getoor, Sudeepa Roy, and Dan Suciu. 2020. Causal relational learning. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 241–256.

Results of Experiment 1

Causal Queries (CaRL)

$Success[M] \Leftarrow Experience[D]?$

$AVG\_Success[M] \Leftarrow Experience[D]?$

Paths

$P_1 = [Director(D), Direct(D, M), Movie(M)]$

$P_2 = [Director(D), Recruit(D, A, M), Movie(M)]$

Causal Queries (CareKG)

$Success[M] \Leftarrow Experience[D]?\ FROM\ PATH\ P, UNDER\ PERSPECTIVE\ \boxed{Direct/Recruit}$

$Success[M] \Leftarrow AVG(Experience[D])?\ FROM\ PATH\ P, UNDER\ PERSPECTIVE\ \boxed{Movie}$

$AVG(Success[M]) \Leftarrow Experience[D]?\ FROM\ PATH\ P, UNDER\ PERSPECTIVE\ \boxed{Director,}$
$SUBJECT\ TO\ \ < axiom >$

(value %)	Path	Perspective		
		Director	Movie	Direct / Recruit
CareKG	P_1	44.12 (± 5.47)	78.70 (± 4.77)	41.81 (± 4.22)
	P_2	46.97 (± 5.55)	71.42 (± 5.82)	45.02 (± 0.68)
CaRL	P_1	44.07 (± 5.47)	-	-
	P_2	47.25 (± 5.56)	-	-

Table 1. perspective & path

$P = P_1, \mathscr{P} = Director$

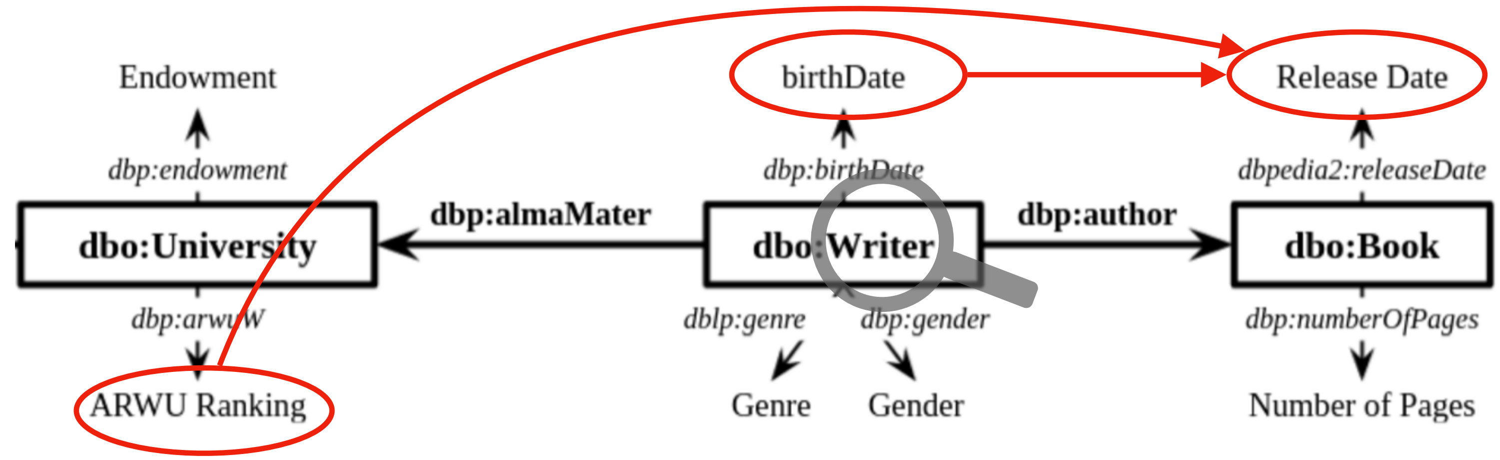
(value %)	SUBJECT TO $\langle axiom \rangle$: I_X is		
	Director (default)	Director ∩ Actor	Director - Actor
CareKG	44.12 (± 5.47)	67.75 (± 10.15)	39.73 (± 6.11)
CaRL	44.07 (± 5.47)	-	-

Table 2. axiom

Conclusion: CareKG has better expressiveness than CaRL

Setting of Experiment 2

Experimental KG:
A real KG "DBpediaW"



Target:
Show how CareKG works different from Differential Causal Rule Mining [4].

Ontological Causal Rule (CareKG)

$$ReleaseDate[B] \Leftarrow birthDate[W]\ WHERE\ [Writer(W), author(W,B), Book(B)]$$

$$ReleaseDate[B] \Leftarrow Ranking[U]\ WHERE\ [University(U), almaMater(W,U), Writer(W), author(W,B), Book(B)]$$

Causal Queries (CareKG)

$$MIN(ReleaseDate[B]) \Leftarrow birthDate[W]?\ FROM\ PATH\ P,\ UNDER\ PERSPECTIVE\ Writer(W)$$

$$MIN(ReleaseDate[B]) \Leftarrow AVG(Ranking[U])?\ FROM\ PATH\ P,\ UNDER\ PERSPECTIVE\ Writer(W)$$

[4] Simonne, Lucas, et al. "Differential Causal Rules Mining in Knowledge Graphs." *Proceedings of the 11th on Knowledge Capture Conference*. 2021.

Results of Experiment 2

Result from CareKG:

If a writer is 1 year younger than other, he / she publish first book 0.86 year earlier.

If a writer's university is 1 rank higher than another one's university, he / she publish book first book 0.017 year earlier.

Results from Differential Causal Rule [1]

Strata	Treatment	Effect
$genre(Fiction) \wedge gender(male) \wedge country(US)$	$univ.arwuW(i_1) \geq univ.arwuW(i_2)$	$ageFirst(i_1) \geq ageFirst(i_2)$
$genre(Fiction) \wedge country(US)$	$birthYear(i_1) \leq birthYear(i_2)$	

Rule 1: For male writers who writing fiction, and from US, the writer from higher ranked university publish first book earlier.

Rule 2: For writers who writing Fiction, and from US, the younger writer publish first book earlier.

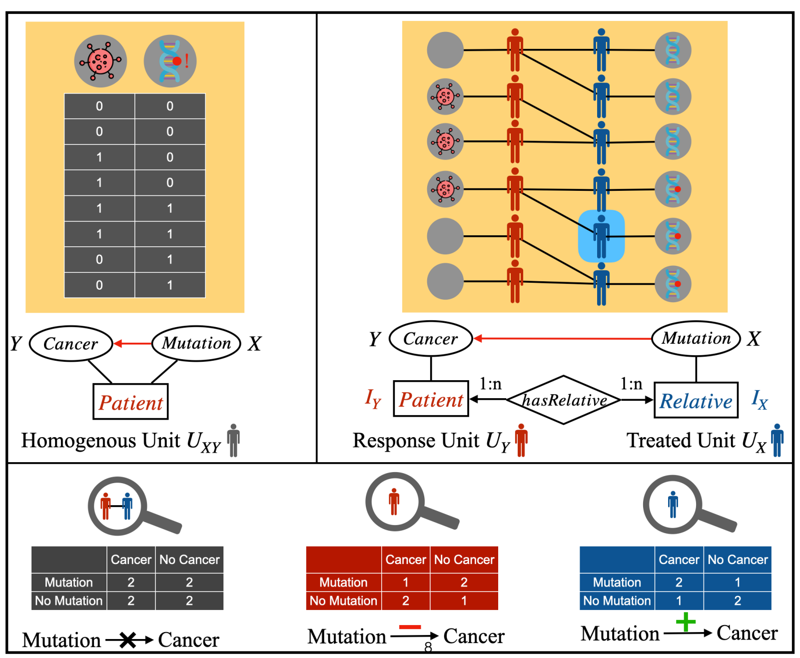
[4] Simonne, Lucas, et al. "Differential Causal Rules Mining in Knowledge Graphs." *Proceedings of the 11th on Knowledge Capture Conference*. 2021.

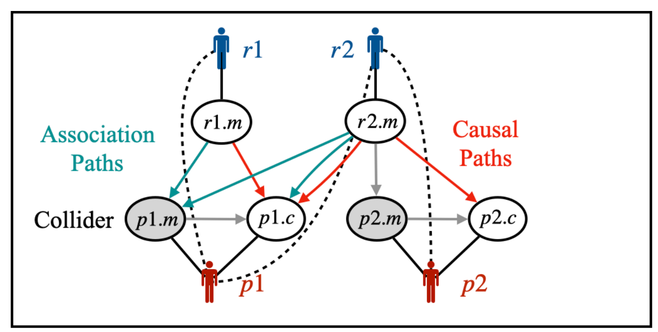
Lessons Learned

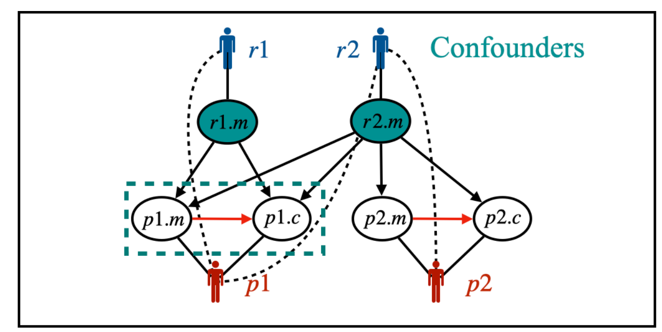
1. Different **perspectives** allow different **unit table** definitions.

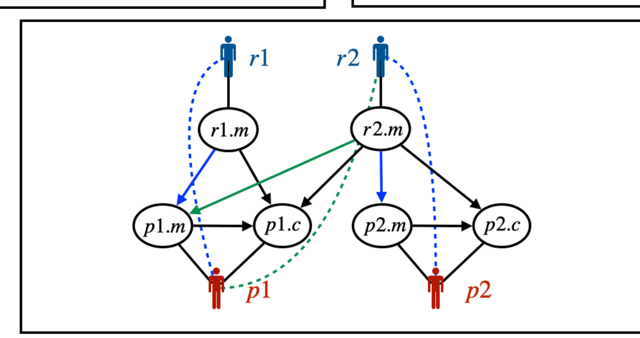
2. The classes, relations, and **metadata** upon them between them make causal inference complicated.

3. Be careful of **multi-value covariates** (causes, effects, confounders, etc)









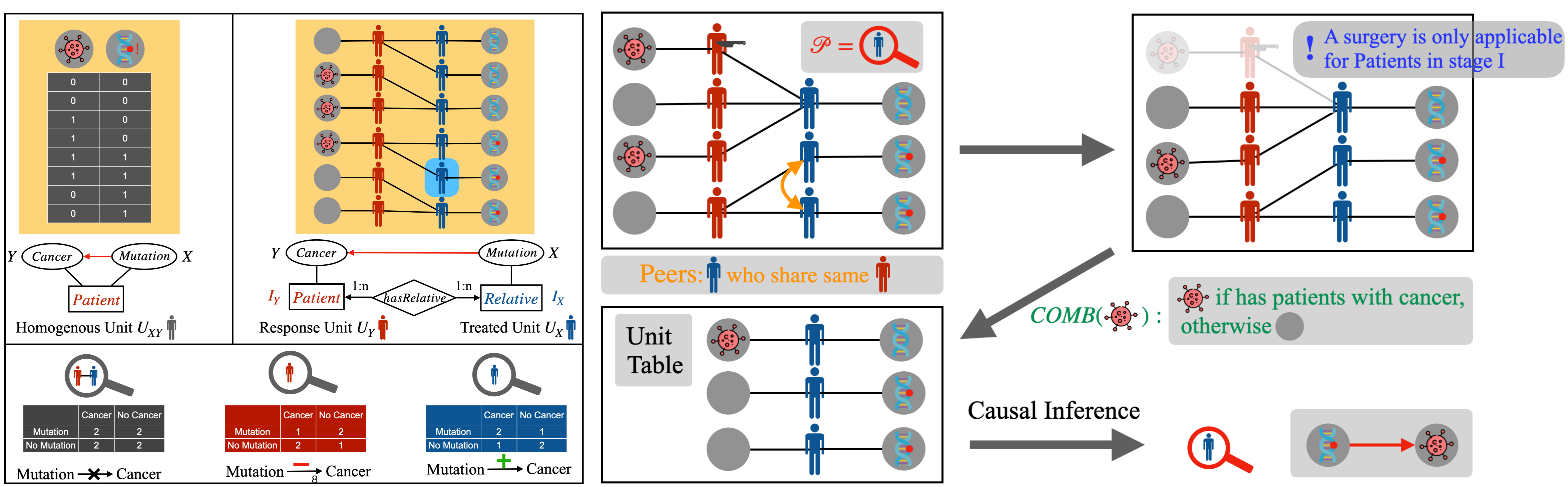
Conclusions & Future Work

Conclusions

1. Offering a **new formalism** for causal query over KGs
2. Extending **causal inference** over KGs
3. Using **semantics** when doing causal queries over KGs

Future Work

1. Making causal inference **efficient** in large scale KGs
2. Studying causal inference **theory** over data structure like KGs.



Time Line

12.2022

12.2023

Causal Inference Theory Over KGs

12.2024

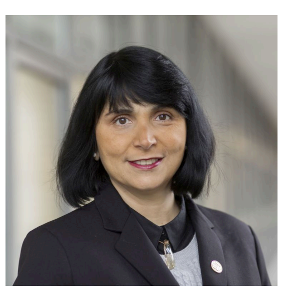
Applications over Medical KGs

Thanks & Questions

Thanks for the **funding** from Leibniz Association
& the **travel grant** from SIGIR!
Thanks for supports of my colleagues from TrustKG project!

Hao.Huang@tib.eu

CareKG: https://github.com/SDM-TIB/CareKG



Maria.vidal@tib.eu

TrustKG Project:
KG building tools: https://github.com/SDM-TIB