

The knowledge-driven exploration of  
integrated biomedical knowledge sources  
facilitates the generation of new hypotheses

Student: Vinh Nguyen, Kno.e.sis Center, Wright State University

Mentor: Dr. Olivier Bodenreider

# Context

- Chagas disease
  - Infectious disease caused by *Trypanosoma cruzi* (*T. cruzi*) parasite
  - Affects 8-10 million people in South America, about 20,000 deaths per year
- Collaborator
  - Center for Tropical and Emerging Global Diseases, University of Georgia
  - Finding drug targets and vaccine

# Context

## PubMed Articles

### **Trypanosoma cruzi calreticulin: a possible role in Chagas' disease autoimmunity.**

Ribeiro CH, López NC, Ramírez GA, Valck CE, Molina MC, Aguilar L, Rodríguez M, Maldonado I, Martínez R, González C, Troncoso R, Lavandero S, Gingras AR, Schwaebler W, Ferreira A.

Programa Disciplinario de Inmunología, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile.

#### **Abstract**

Trypanosoma cruzi (T. cruzi) is the causative agent of Chagas' disease, an endemic and chronic illness that affects 18 million people in Latin America. The mechanisms underlying its pathogenesis are controversial. There is a growing body of evidence supporting the view that T. cruzi infection elicits severe autoimmune responses in the host, which significantly contribute to the pathogenesis of Chagas' disease, and several recent studies have reported the presence of autoantibodies and effector T lymphocytes against parasite and self antigens in infected patients and experimentally infected animals. T. cruzi calreticulin (TcCRT) is a 45kDa protein, immunogenic in humans, rabbits and mice. It has a high degree of homology with human (HuCRT) and mouse calreticulin (MoCRT), which would explain why an immune response to TcCRT could contribute to autoimmune reactions in Chagas' disease. Anti-TcCRT antibodies generated in A/J mice immunized with recombinant TcCRT (rTcCRT) reacted with rHuCRT and bound to neonatal and adult isogenic cardiomyocytes cultured in vitro. Interestingly, histological alterations, such as edema formation and cell infiltrates, which include CD3(+) cells, were detected in heart sections from immunized

## Experimental data

	A	B	C	D
1	Gene Id	SAM result	Assigned Oligo	A Log2 Ratio
2	Tc00.1047053397937.10	NON-SIG		-0.1
3	Tc00.1047053397937.5	SIG	QTC00006_E_2	0.12
4	Tc00.1047053398265.9	NON-SIG	QTC00009_B_1	0.49
5	Tc00.1047053398477.10	SIG	QTC00009_B_7	-1.18
6	Tc00.1047053398751.10	SIG		-0.93
7	Tc00.1047053399033.19	SIG		0.51
8	Tc00.1047053399373.9	SIG	QTC00009_B_9	-1.11
9	Tc00.1047053399389.10	NON-SIG	QTC00009_B_11	-0.64
10	Tc00.1047053400945.10	NON-SIG	QTC00009_B_15	-1.19
11	Tc00.1047053401469.10	SIG	QTC00004_N_23	-1.43
12	Tc00.1047053401473.9	NON-SIG	QTC00009_B_17	-1.22
13	Tc00.1047053401569.10	SIG	QTC00009_B_19	-0.79
14	Tc00.1047053402647.9	NON-SIG	QTC00009_B_21	0.95
15	Tc00.1047053402863.9	NON-SIG	QTC00009_B_23	-0.9
16	Tc00.1047053403875.10	NON-SIG	QTC00009_D_3	-0.27
17	Tc00.1047053403875.19	NON-SIG		-0.23
18	Tc00.1047053403985.10	NON-SIG	QTC00020_J_13	-1.19
19	Tc00.1047053404001.10	SIG		0.62
20	Tc00.1047053404001.20	SIG		-1.19
21	Tc00.1047053404431.10	NON-SIG		-0.71
22	Tc00.1047053404843.10	SIG		-0.73
23	Tc00.1047053404843.20	NON-SIG	QTC00009_D_5	-0.25
24	Tc00.1047053404975.30	NON-SIG	QTC00009_D_9	-0.16
25	Tc00.1047053405025.10	NON-SIG	QTC00009_D_11	-0.21
26	Tc00.1047053405165.10	SIG		1.24
27	Tc00.1047053406999.10	NON-SIG	QTC00022_O_10	-0.67
28	Tc00.1047053407335.18	NON-SIG		1.16
29	Tc00.1047053407335.9	NON-SIG	QTC00003_D_6	0.85

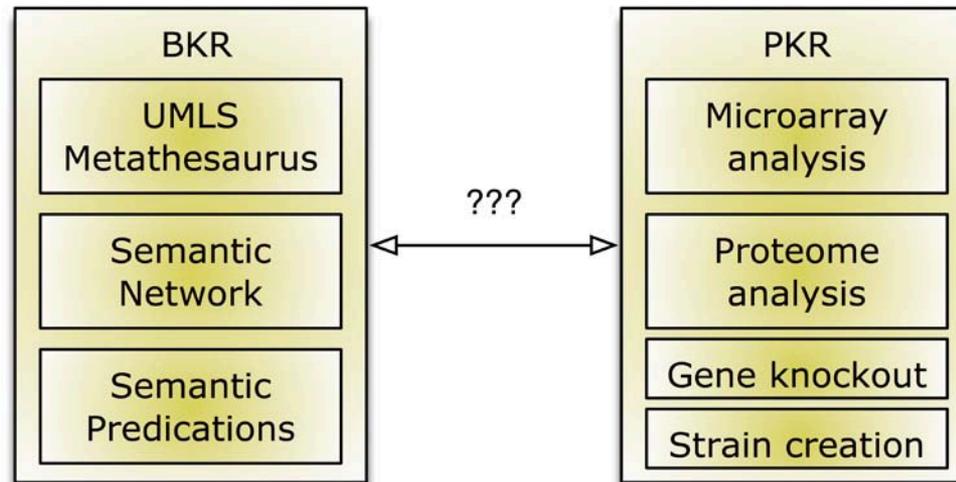
# Motivation

- Data sources
  - Text data from PubMed
  - Experimental data about Chagas disease
- Questions
  - Q1: How to query experimental and text data together?
  - Q2: How to find new insights into Chagas disease?

The knowledge-driven **exploration** of **integrated** biomedical knowledge sources facilitates the generation of new hypotheses

# Q1 - Integration

- Knowledge bases
  - The Biomedical Knowledge Repository (BKR)
  - The Parasite Knowledge Repository for Chagas disease (PKR)



# Knowledge source - BKR

- Created at NLM by Dr. Olivier Bodenreider and Dr. Thomas Rindflesch
- ~ 20 million predications from 6 million articles by SemRep
- Example

**Trypanosoma cruzi calreticulin: a possible role in Chagas' disease autoimmunity.**

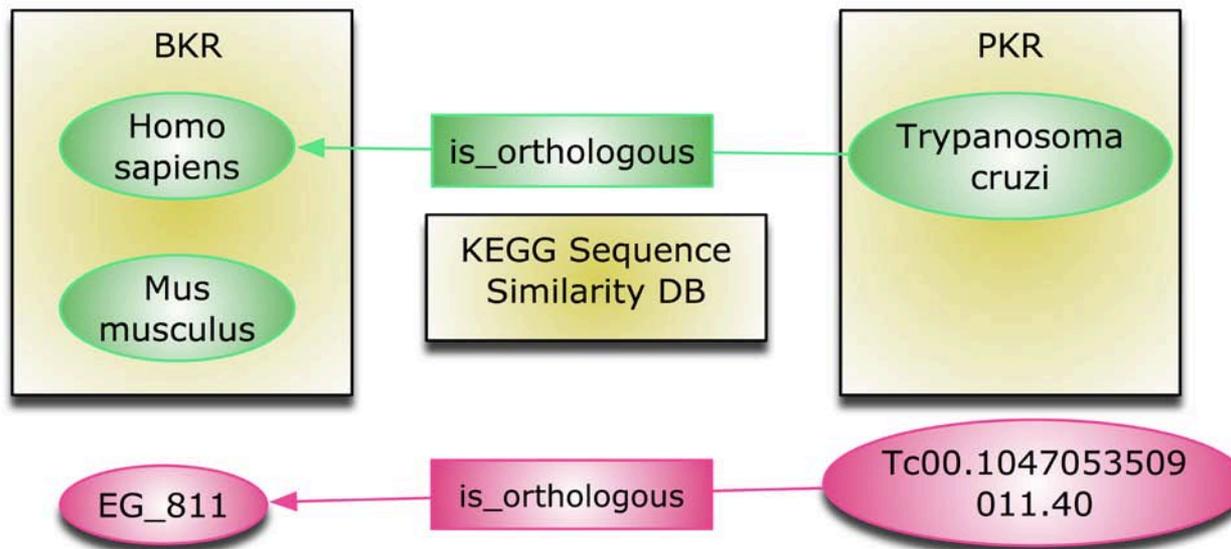


# Knowledge source - PKR

- Created by *T.cruzi* project at Kno.e.sis, WSU
- Microarray analysis
  - Measure the expression level of *T. cruzi* genes
- Proteome analysis
  - Measure the presence of proteins encoded by *T. cruzi* genes
- Ontologies for annotation
  - Parasite Experiment Ontology (PEO)
  - Parasite Lifecycle Ontology (PLO)

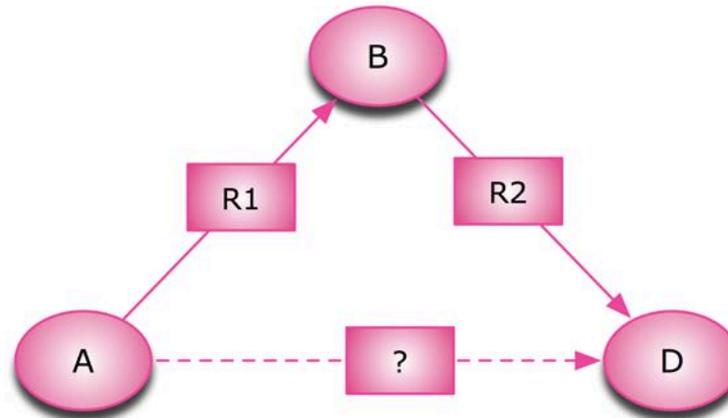
# Linking source

- Ortholog relation
  - Gene function is usually conserved across species



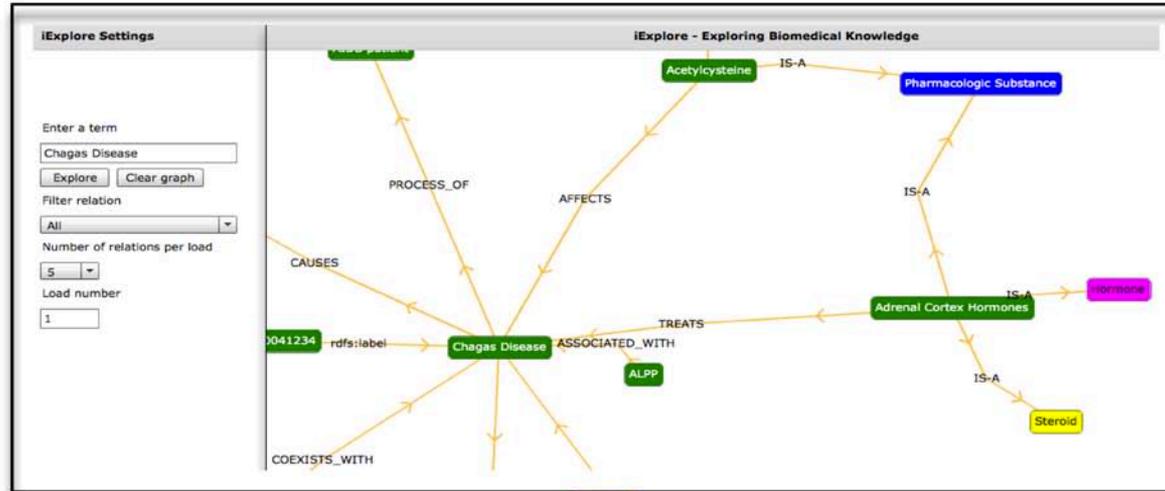
# Q2 – How to explore?

- Knowledge-driven exploration process

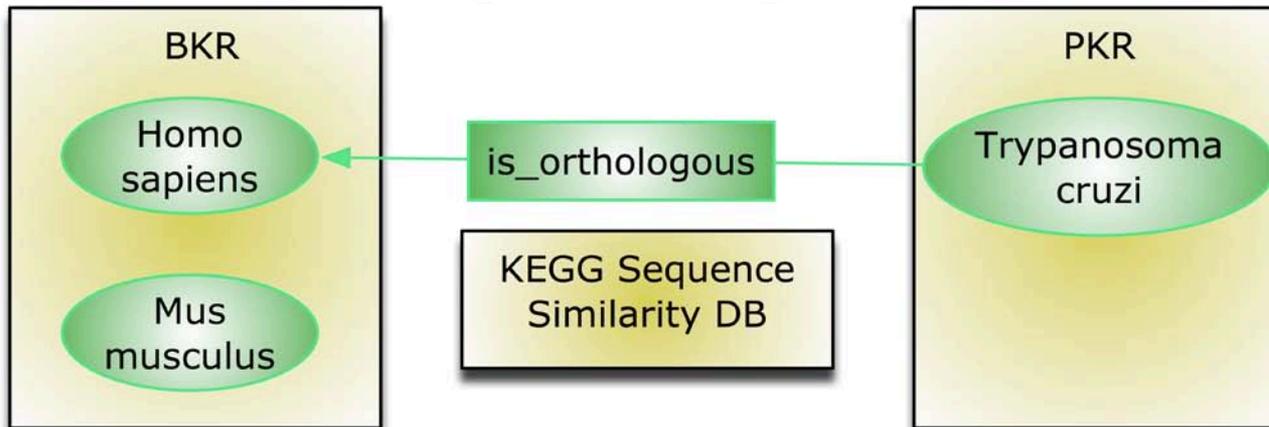


- Two steps
  - *Navigate* the chain of named relationships
    - Graph expansion
    - Graph restriction
  - *Interpret* the chain to infer new hypothesis

# iExplore

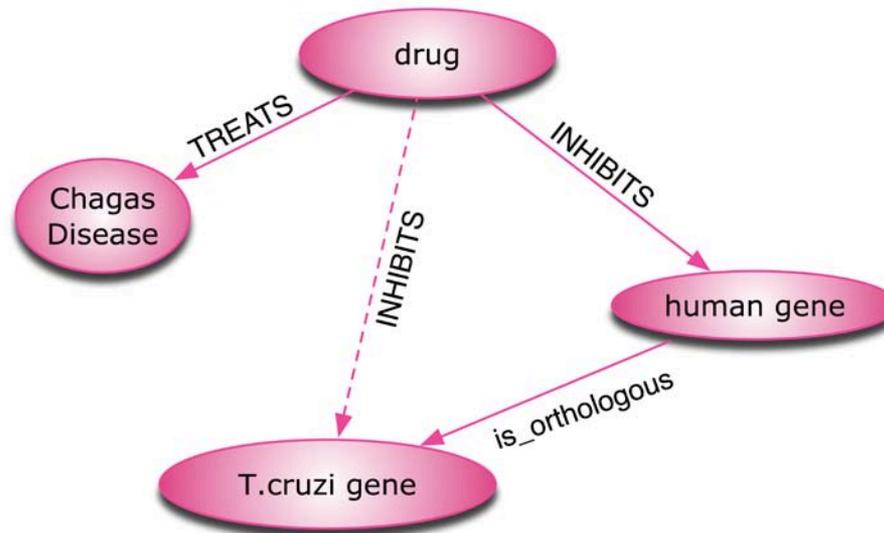


## Integrated knowledge base



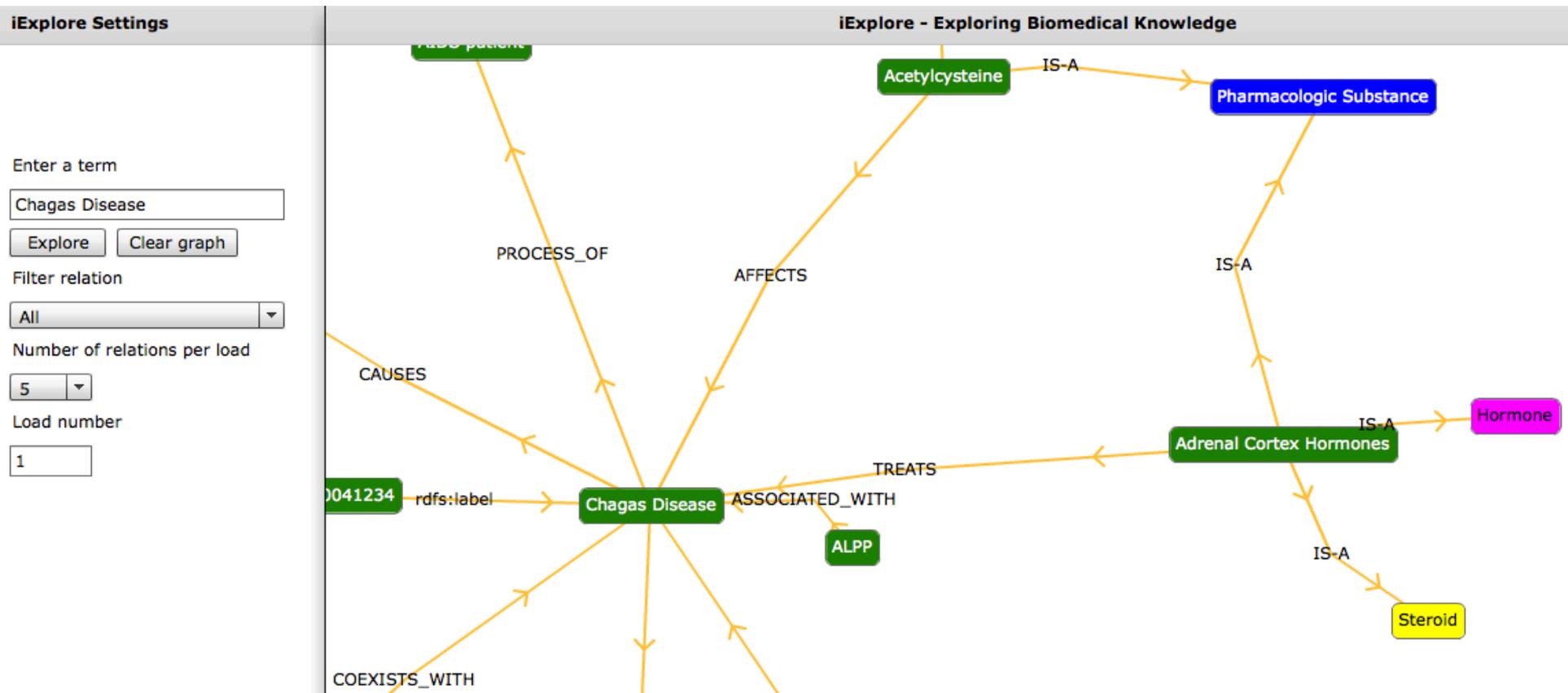
# Exploration process

- Sample chain



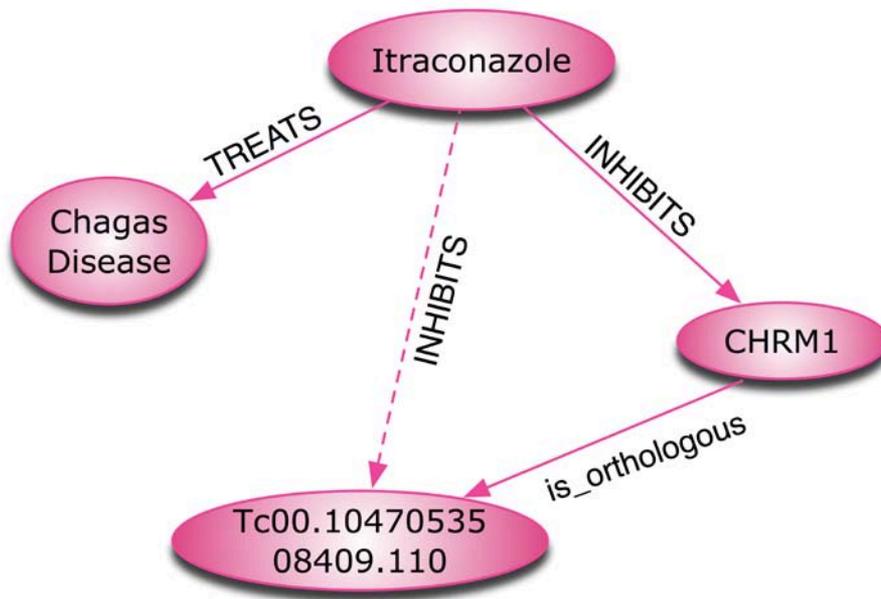
# Exploration process – Live demo

- iExplore



# Exploration process - Interpretation

- Sample chain -> Hypothesis



The *T. cruzi* orthologs of human genes inhibited by itraconazole may also be inhibited by itraconazole and thus would be candidates for further studies into the mechanism(s) of action of the drug itraconazole on *T. cruzi*

# Significance

- Biologically
  - Biologists follow the exploration process to generate novel hypotheses
  - Can easily be generalized to other diseases
- Technically
  - Technical details (e.g., SPARQL queries) hidden from the user
  - Scalable over millions of predications

# Contribution

- Data integration
- Development of *iExplore*
- Paper writing
  - Vinh Nguyen, Olivier Bodenreider, Todd Minning, Amit Sheth. *The knowledge-driven exploration of integrated knowledge sources facilitates the generation of new hypotheses*. Linked Science, 2011 (submitted)

# Technologies

- Semantic Web
  - RDF, SPARQL standards
  - Virtuoso triple store
- iExplore development
  - Client side: Flex application
  - Server side: Java

# Conclusion

- Integrate knowledge bases
  - BKR
  - PKR
- Knowledge-driven exploration
  - Two-step process: navigation and interpretation
  - Assist biologists to explore the integrated knowledge bases

# Future work

- Validation of novel hypotheses
  - Collaborating with UGA researchers in the biomedical aspect.
  - Incorporate the validation into *iExplore*
- Automate the interpretation step
  - Learning the way biologists use their background knowledge to interpret the chain of named relationship

# Acknowledgements

Dr. Olivier Bodenreider

Dr. Thomas Rindflesch

May Cheh

Dr. Clement McDonald

Dr. Bastien Rance

Jonathan Mortensen

John Nguyen

Thang Nguyen

Lister Hill Center

Dr. Cartic Ramakrishnan, Dr. Priti Parikh, Joshua Dotson,  
Sarasi Lalithsena from Kno.e.sis center, WSU.

