

11.03.2021

# Intermap

An integrative multiomics approach to generating  
therapeutic target hypotheses

Regis A. James, PhD

Bioinformatics Staff Scientist



# This is an optimized way of generating therapeutic target hypotheses.

The screenshot shows a web-based application titled "Intermap Target Detector". The URL is [https://bioinfo-connect.regenon.regn.com/bics/intermap\\_target\\_detector/?GOtermIDs=GO:0005788](https://bioinfo-connect.regenon.regn.com/bics/intermap_target_detector/?GOtermIDs=GO:0005788). The interface includes a sidebar with "Inclusion criteria:" and a main panel displaying a table of search results. A "Shiny" logo is visible in the top right corner.

**Inclusion criteria:**

1. Mouse genes are known to exhibit ANY of the properties selected in the GO term menu.
2. Mouse genes generate non-embryonic-lethal knockout mice (IMPC).
3. Mouse genes have human gene orthologs that are known to host variants that cause genetic (OMIM) disease.
4. OMIM diseases are monogenic (known to be caused by variants in only 1 gene).
5. OMIM diseases are known to have an autosomal recessive mode of inheritance (HPO terms).

**Gene Ontology term(s)**  
endoplasmic reticulum lumen

**Share session** (*copy & paste URL*) or [open in new window](#)

**Click the entities (genes, mice, and diseases) below for details on their annotations (Gene Ontology, Mammalian Phenotype Ontology, and Human Phenotype Ontology, respectively), and additional links to external databases.**

Mouse Gene	IMPC Mouse	Human Gene	OMIM Disease
Casq2	Casq2	CASQ2	Ventricular tachycardia, catecholaminergic polymorphic, 2
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, 6
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, Kufs type, adult onset
Eogt	Eogt	EOGT	Adams-Oliver syndrome 4
Pcsk1	Pcsk1	PCSK1	Obesity with impaired prohormone processing
Wnt1	Wnt1	WNT1	Osteogenesis imperfecta, type XV
Wnt3	Wnt3	WNT3	Tetra-amelia syndrome 1

Showing 1 to 7 of 7 entries

Previous **1** Next

# **But first, how we got here...**

*(and how you can too)*

# Overview

1. Goals and challenges for a pharmaceutical company
2. A data-driven approach for overcoming these obstacles using R
3. An example case study with this approach (*fabricated*)



# Goals and challenges for a pharmaceutical company

# Goals of a Pharmaceutical Company

(an incomplete list)

**1. Discover** therapeutic treatment approach *hypotheses*.

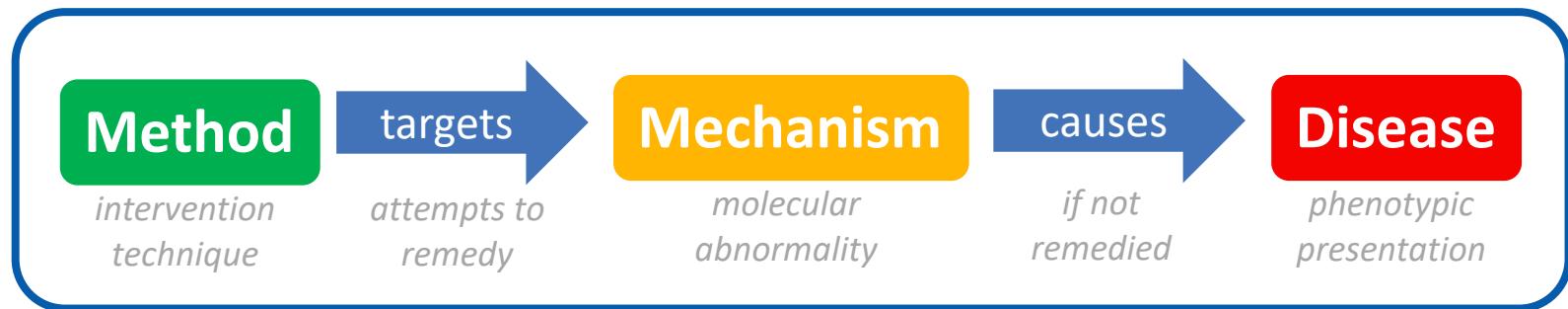
**2. Develop**/hone treatments to maximize potency, safety, and scalability.

**3. Deliver** optimized treatments to patients as efficiently as possible.

[ Heal  
the  
world ]

# A therapeutic treatment hypothesis

(an oversimplification)



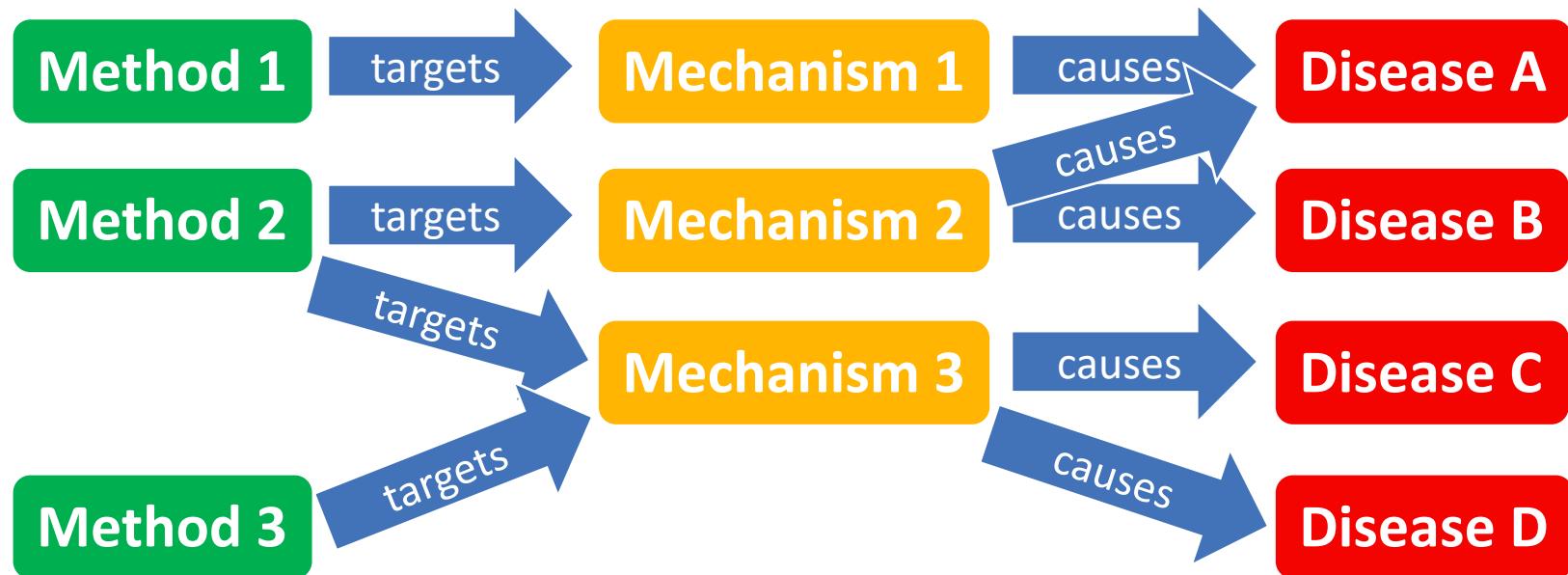
## Hypothesis:

Prevent the mechanism, prevent the disease.

(need the one with the best chance of *actually* helping people)

# Problem: Too many potential hypotheses to choose from!

(even after filtering to only what is currently known)



...and so on

(but some have weak links, enabling elimination)

# Deciding which hypothesis to consider adding to R&D pipeline

(manual, coarse filter requirements)

## Method

- Must be accessible by pharma company.
- Must be understood well enough to yield consistent results.
- Must not generate undesired phenotypes when used to target mechanism.
  - *Early awareness and management of potential side effects could help maximize chances of successful therapeutic targeting downstream.*

## Mechanism

- Must be targetable by intervention method.
  - Some methods may be more effective than others.
- Must be simple enough to recapitulate in a controlled laboratory environment.

## Disease

- Must be feasible to generate a viable model that enables testing of intervention method.
- Must be detectable in model via phenotype presence and, upon rescue, absence.

# Too many options remain

(not an exhaustive list)

## Methods

- Antimetabolites
- Antimitotics
- Antitumor Antibiotics
- Asparagine-Specific Enzymes
- Biosimilars
- Bisphosphonates
- Chemotherapy
- DNA-Damaging Agents (Antineoplastics) and Alkylating Agents
- DNA-Repair Enzyme Inhibitors
- Histone Deacetylase Inhibitors
- Hormones (Corticosteroids)
- Hypomethylating (Demethylating) Agents
- Immunomodulators
- Janus-Associated Kinase (JAK) Inhibitors
- Monoclonal Antibodies
- Phosphoinositide 3-kinase inhibitors (PI3K inhibitors)
- Proteasome Inhibitors
- Selective Inhibitors of Nuclear Export (SINE)
- Tyrosine Kinase Inhibitors

## Mechanisms

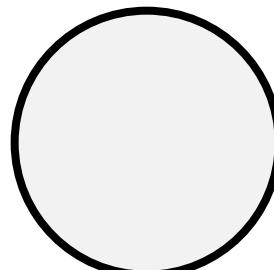
- Genome
- Transcriptome
- Proteome
- Epigenome
- Metabolome/Interactome
- Envirome
- Aging

- Bone
- Immunological
- Cancer
- Metabolic
- Cardiovascular
- Neurological
- Connective Tissue
- Nutritional
- Ophthalmological
- Dermatological
- Developmental
- Ear, Nose, Throat
- Endocrine
- Gastrointestinal
- Psychiatric
- Renal
- Respiratory
- Skeletal
- Unclassified
- Hematological

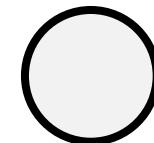
*Source: "The human disease network"  
Goh, Cusick, et al. PNAS 2007*

# Too many options remain

Impossible to investigate *everything*.



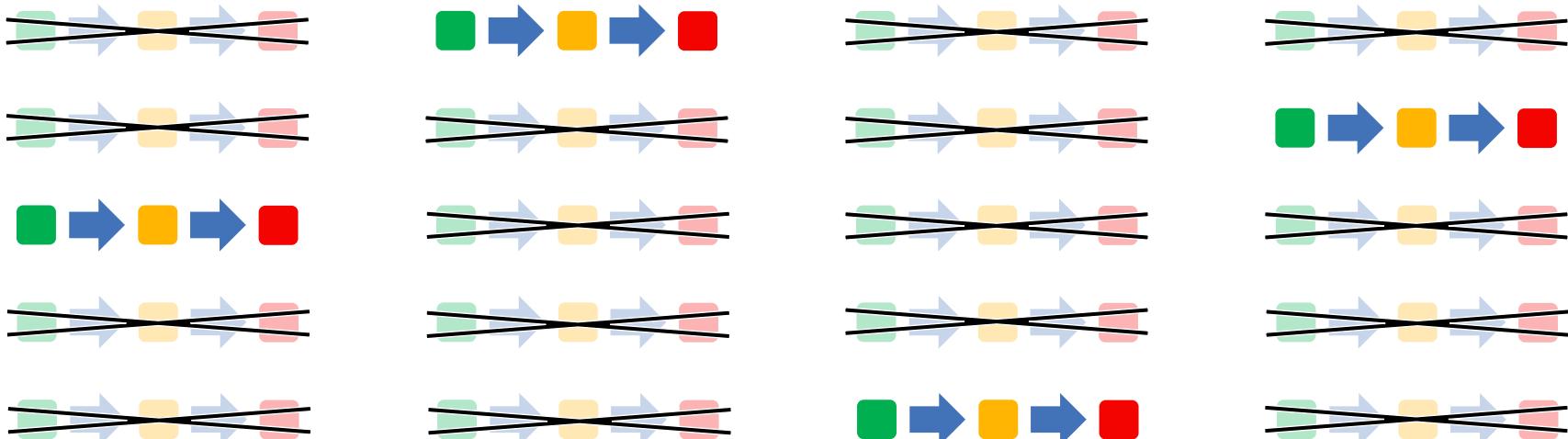
**Testable  
Options**



**Resources**  
*(time, funds, personnel)*

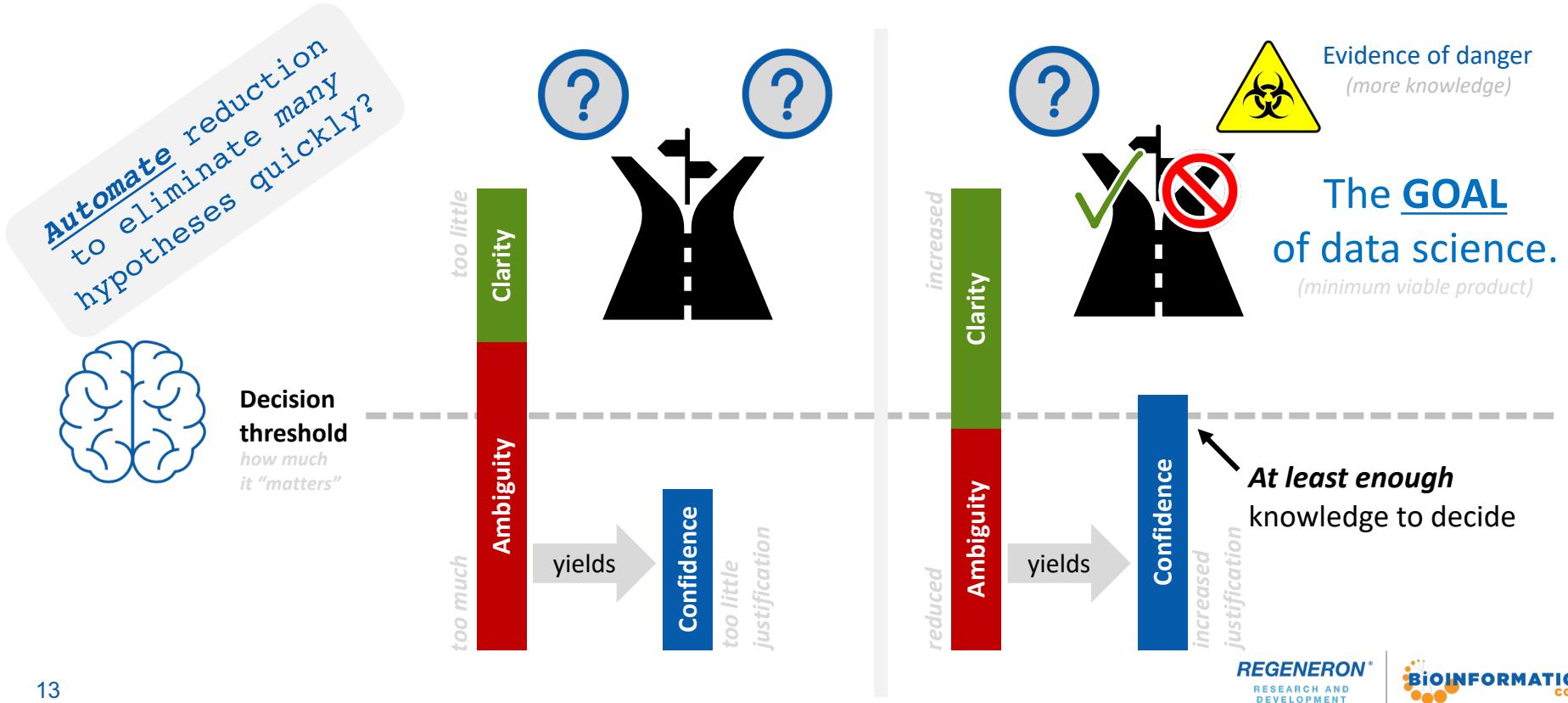
# Need a better way to decide which hypotheses to investigate

Accelerate the identification and elimination of less viable options.



# Decisionmaking is just *reducing ambiguity*.

The brain naturally develops **confidence** when sufficient clarity is achieved.



Need *one more concept* to build a tool for rapidly making decisions *at scale*:

Computational reasoning via ***ontologies***

“

”

The Matrix is everywhere.



Source: *The Matrix*

“

# Ontologies are everywhere.

”



Source: *The Matrix*

“

# Ontologies are dictionaries that can be used programmatically:

Controlled **vocabularies** and their hierarchical defining **relationships**

”



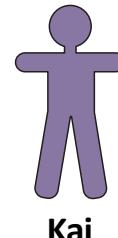
Source: *The Matrix*

# A conceptual example.

(to build an effective decision acceleration tool, must get “computational reasoning” right)

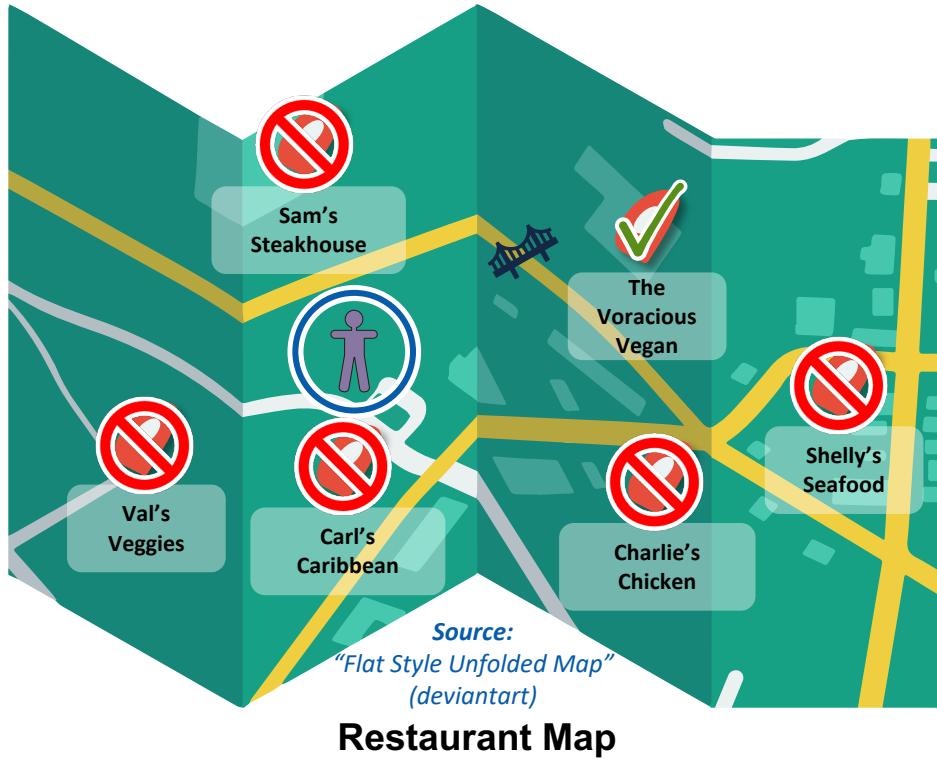
“Today for lunch, Kai wants to have a vegetarian meal without leaving a carbon footprint and be back in an hour.”

**Which restaurant should Kai choose?**



# Choosing an accessible vegetarian meal with a carbon-neutral route

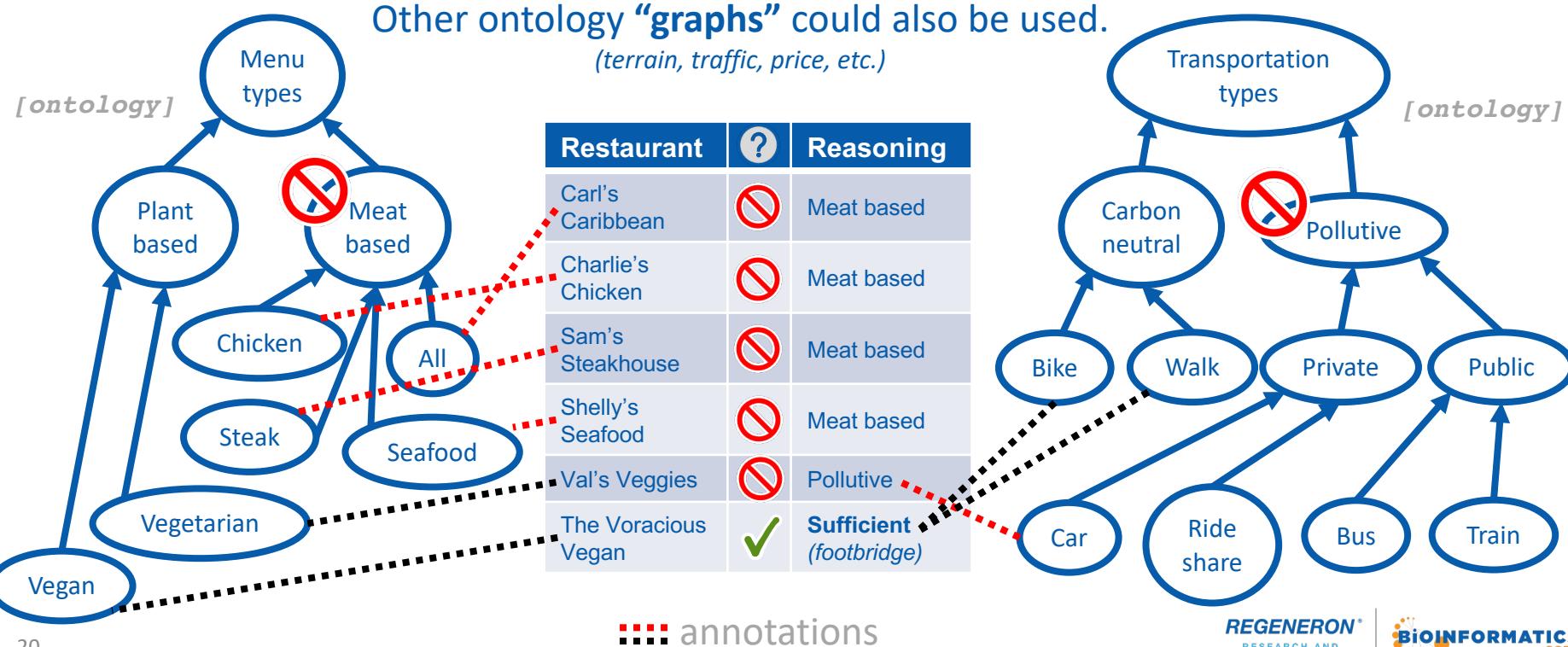
## Human reasoning



Restaurant	?	Reasoning
Carl's Caribbean	🚫	Not vegetarian
Charlie's Chicken	🚫	Not vegetarian
Sam's Steakhouse	🚫	Not vegetarian
Shelly's Seafood	🚫	Not vegetarian
Val's Veggies	🚫	Too far (indirect)
The Voracious Vegan	✓	Sufficient (footbridge)

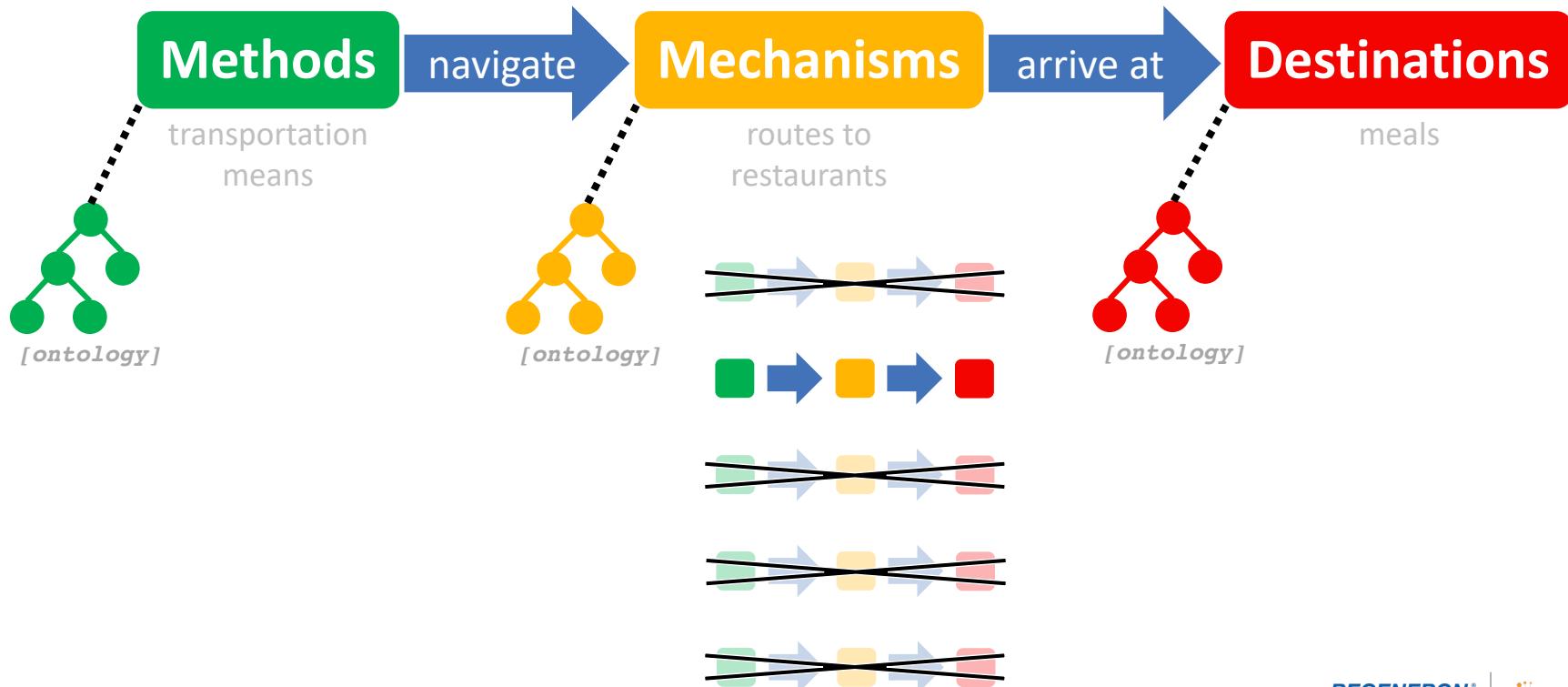
# How could we build software that recommends the same decision?

Applied “*computational* reasoning” (semantics)



# Computationally-reasoned lunch recommendation

Can be scaled to think through extremely large catalogs of options.

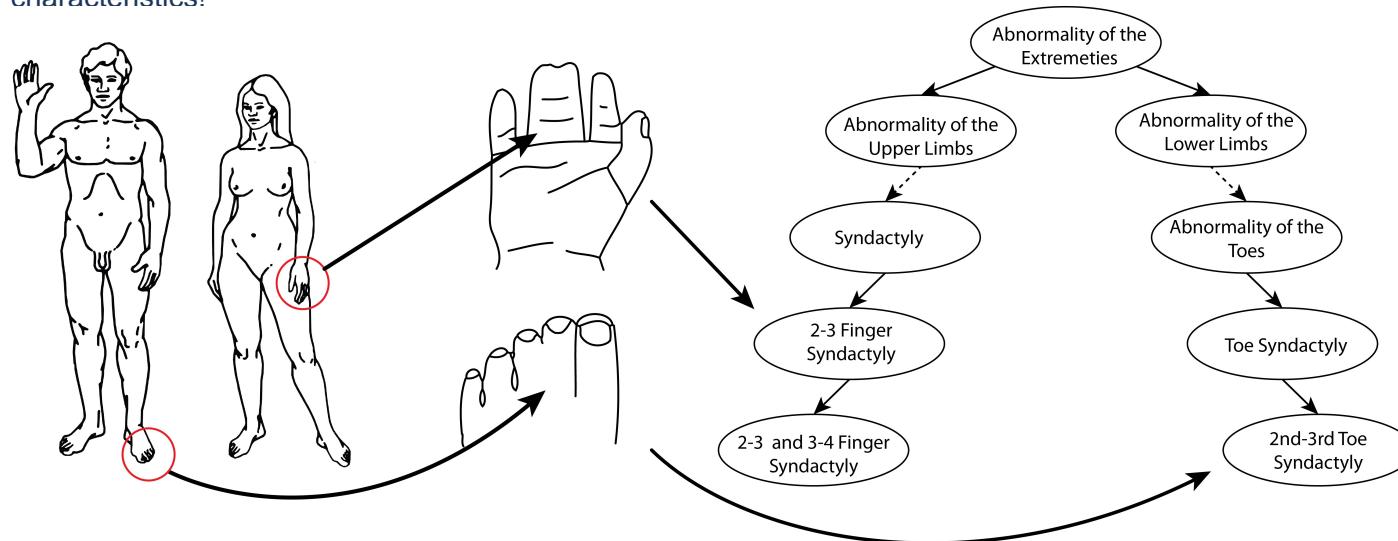


# Applying computational reasoning to biomedicine.

A bounty of public ontologies, biological entity (gene, disease, etc.) catalogs, and their annotations currently exist.

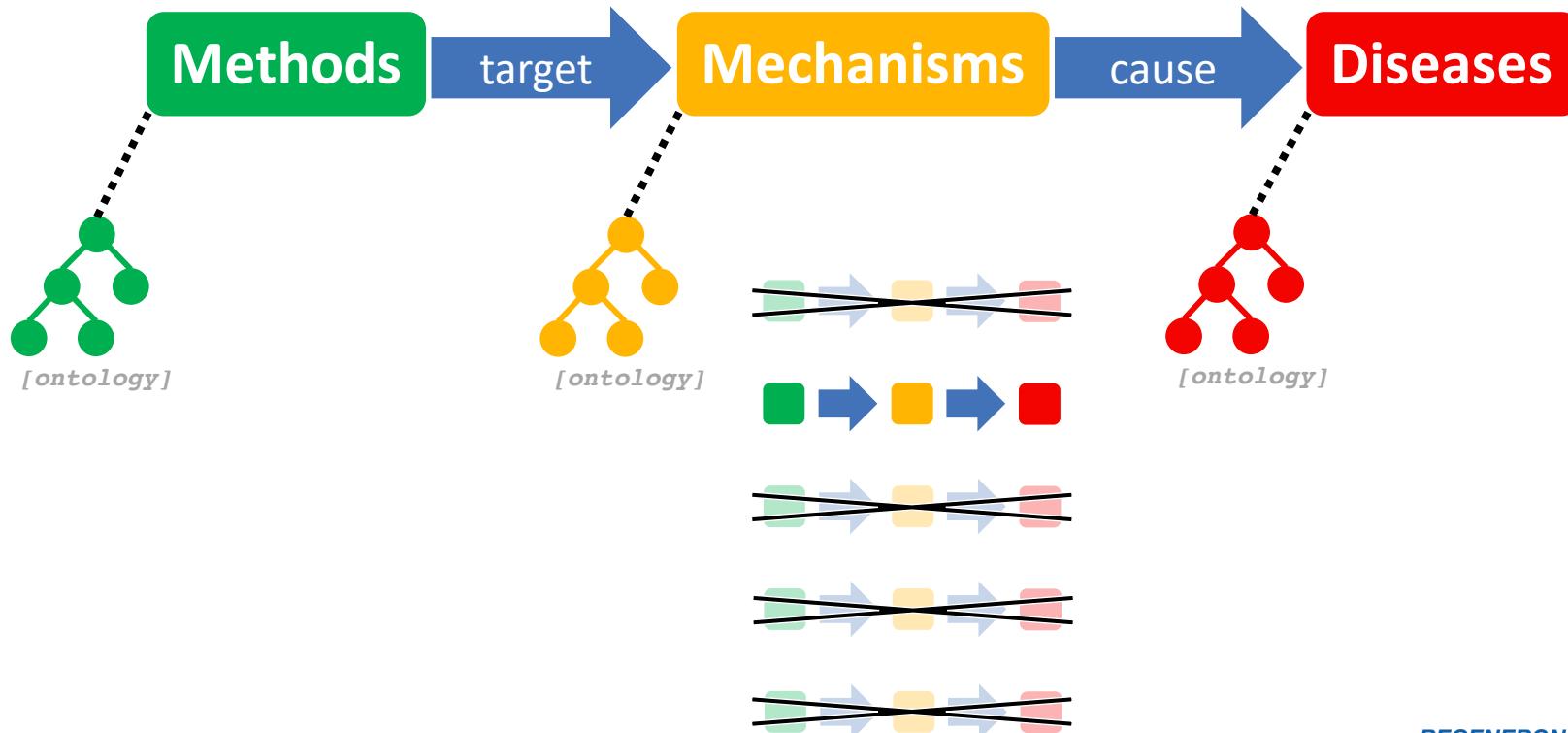
## Example: The Human Phenotype Ontology (HPO)

- Represents the relationships between terms describing human aberrations/diseases (Robinson, et al, 2008)
- 10,000+ characteristics!



# Computationally-reasoned hypothesis recommendation

Can be applied to increasingly complex hypotheses across massive biomedical catalogs.



# Hypothesis recommendations curated w/ computational reasoning



=

The screenshot shows the 'Intermap Target Detector' web application interface. At the top, there's a search bar with the URL [https://bioinfo-connect.regenon.regn.com/bics/intermap\\_target\\_detector/?GOtermIDs=GO:0005788](https://bioinfo-connect.regenon.regn.com/bics/intermap_target_detector/?GOtermIDs=GO:0005788). Below the header, the title 'Intermap Target Detector' is displayed, followed by a sub-instruction: 'Select GO (Gene Ontology) terms to identify genes, mice, and diseases that meet the criteria listed below the menu.' A 'Gene Ontology term(s)' input field contains the term 'endoplasmic reticulum lumen'. To the right, there's a section titled 'Share session (copy & paste URL) or open in new window' with a 'Search:' input field. The main content area features four columns: 'Mouse Gene', 'IMPC Mouse', 'Human Gene', and 'OMIM Disease'. Each column has a dropdown menu set to 'Show 25 entries'. The 'Mouse Gene' column lists seven entries, each with a green arrow pointing to the corresponding entry in the other three columns. The 'IMPC Mouse', 'Human Gene', and 'OMIM Disease' columns also show seven entries each. At the bottom, there are buttons for 'Mouse Gene', 'IMPC Mouse', 'Human Gene', and 'OMIM Disease', and navigation links 'Previous', '1', and 'Next'.

A data-driven  
approach for  
overcoming  
hypothesis selection  
obstacles using R

# The approach



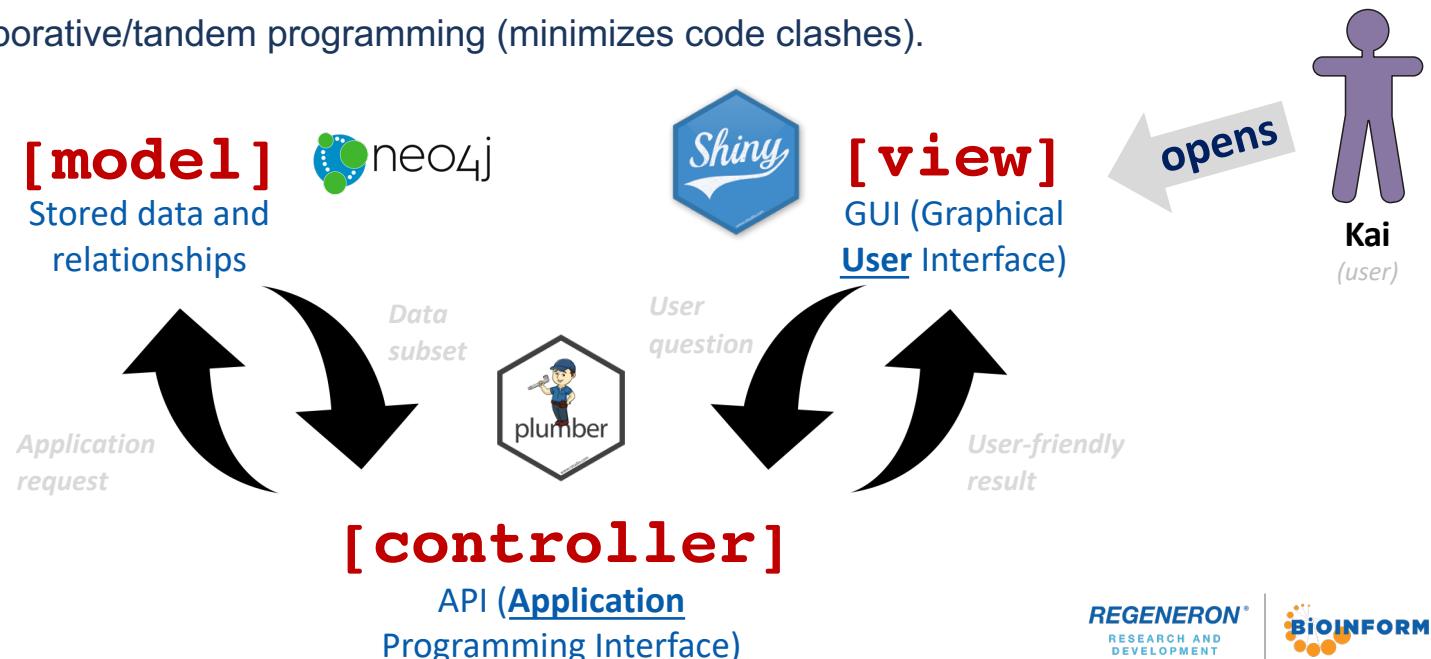
1. Build a living ***knowledgebase*** of biomedical entities, ontologies, and their relationships.
2. Build a programmatic association-mining ***connection*** (API) to that knowledgebase.
3. Build a user-friendly ***interface*** for querying that connection and explaining the reasoning behind recommendations.



# Approach ≈ MVC (model-view-controller) design pattern

Futureproofing made easy.

- Separates/orgанизes application into *distinct* manageable components.
- Easier to add/debug functionality to 1 component without digging through / modifying entire codebase.
- Better for collaborative/tandem programming (minimizes code clashes).



# Resources and techniques necessary for implementation

- Neo4j “graph” database for storing catalogs of published experimental data and their annotations to ontologies
- Construction and optimization of code in Neo4j’s “cypher” language for quickly searching through connected graphs and metadata
- Semantic similarity for generating relevance-ranked gene lists via “transitive prioritization” (guilt-by-association)
- Construction of an interactive API-first (application programming interface) webtool for querying the data and enabling scientific decisionmaking in the shiny framework for R, etc.

## Data



National  
Center for  
Biotechnology  
Information



## Structured Vocabulary & Semantic Reasoning



**GENEOntology**  
Unifying Biology



The Mammalian Phenotype  
Ontology



## Tools & Frameworks



## Database & Platform



**MySQL**

R Studio® Connect



**REGENERON®**  
RESEARCH AND  
DEVELOPMENT

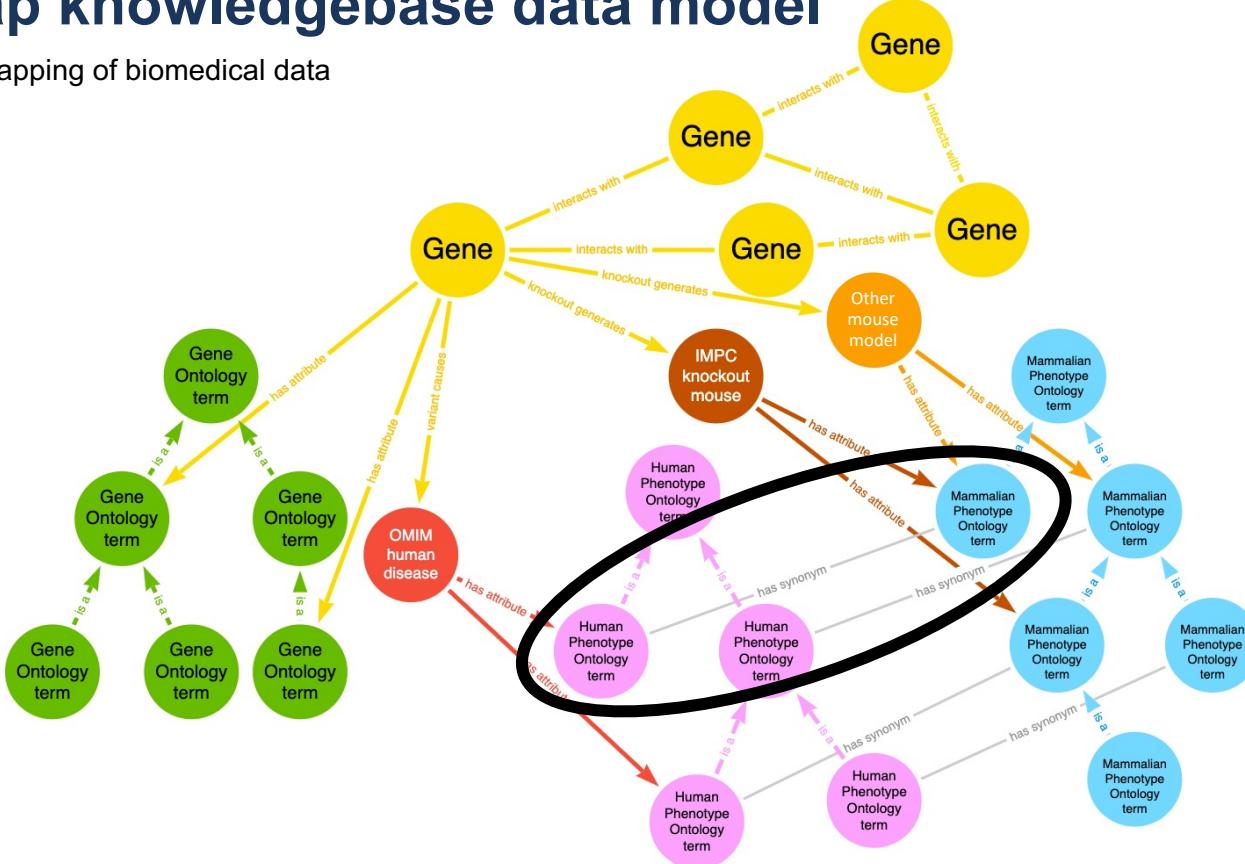
**BIOINFORMATICS**  
CORE

# The knowledgebase

*(and building increasingly complex/complete cypher queries with R)*

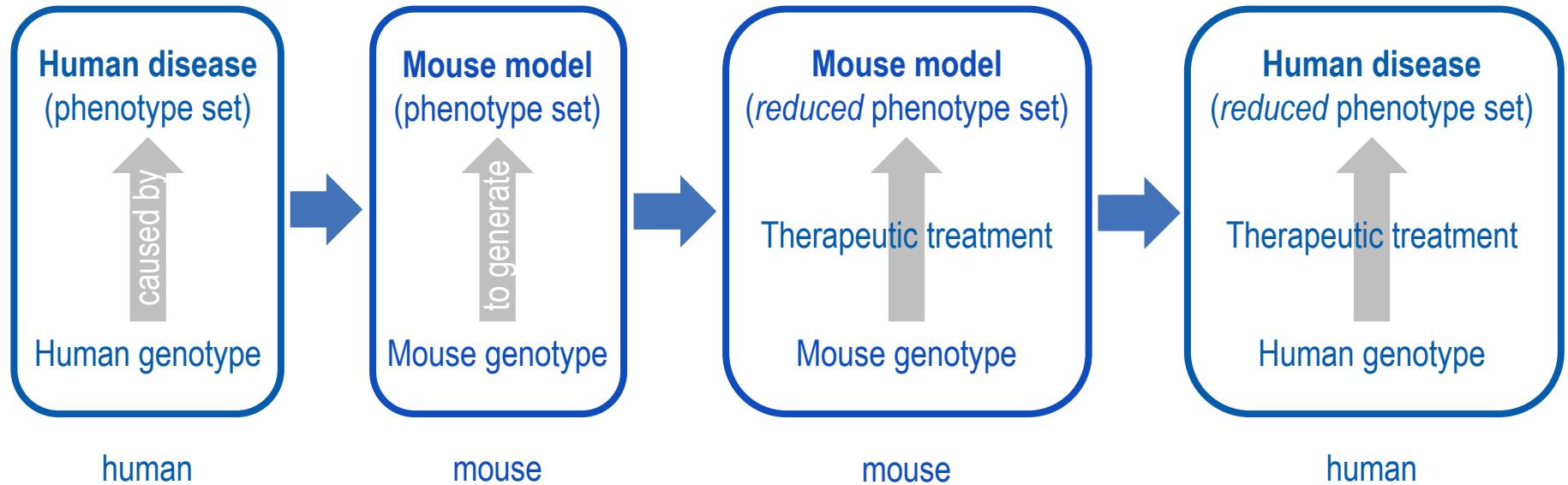
# Intermap knowledgebase data model

## Interspecies mapping of biomedical data



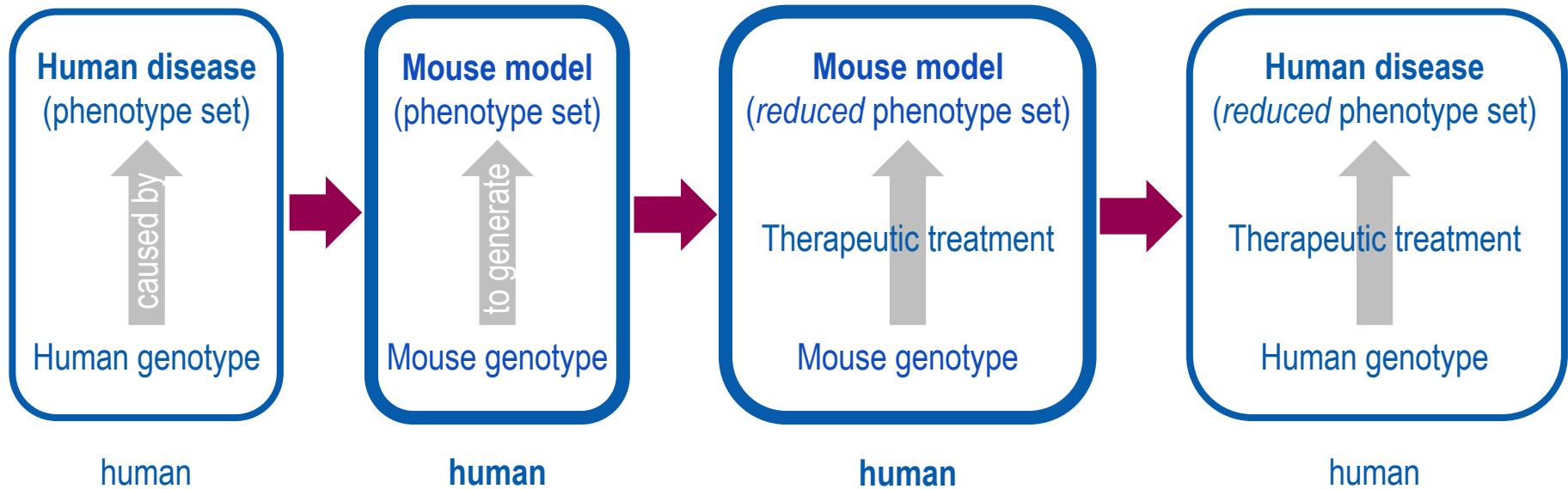
# The necessity of an *interspecies* knowledgebase

Human phenotype might not be directly considered *throughout the entire* therapeutic development process.



# Interspecies mappings for improved computational reasoning

Consider human phenotype more directly *throughout the entire therapeutic development process* by looking at **mouse data in the context of human disease**.

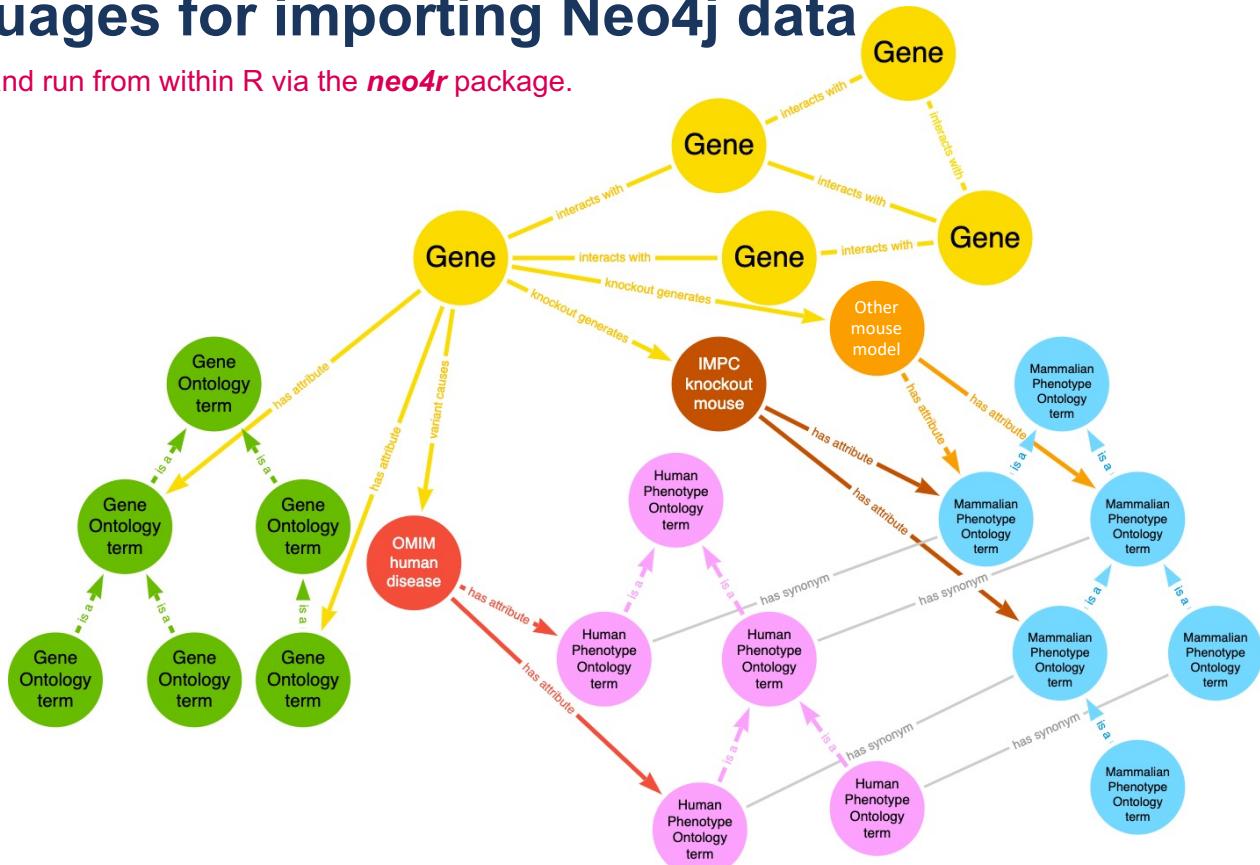


# R and Cypher languages for importing Neo4j data

Used direct cypher + cypher generated and run from within R via the *neo4r* package.

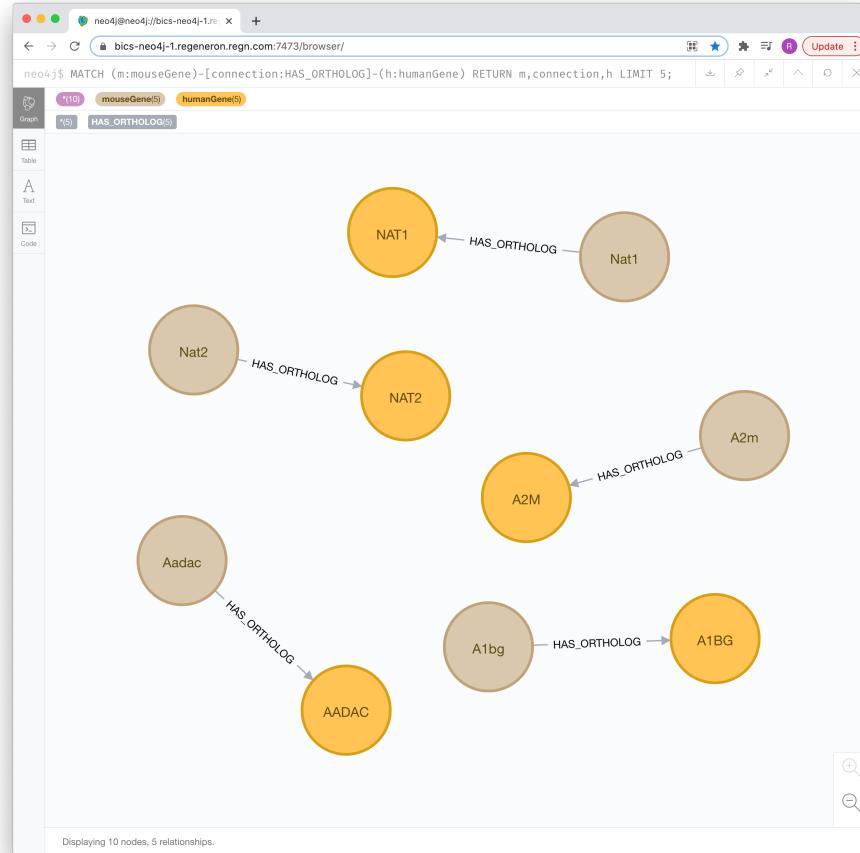
Type	Entity	Quantity
Nodes	GO terms	44085
	HPO terms	16173
	MPO terms	13698
	Mouse genes and derivatives	73139
	Human genes and derivatives	62134
	IMPC mice	6316
	OMIM diseases	7797
Edges	Connections	890564

$$( \text{brown circle} + \text{orange circle} = \text{yellow circle} )$$



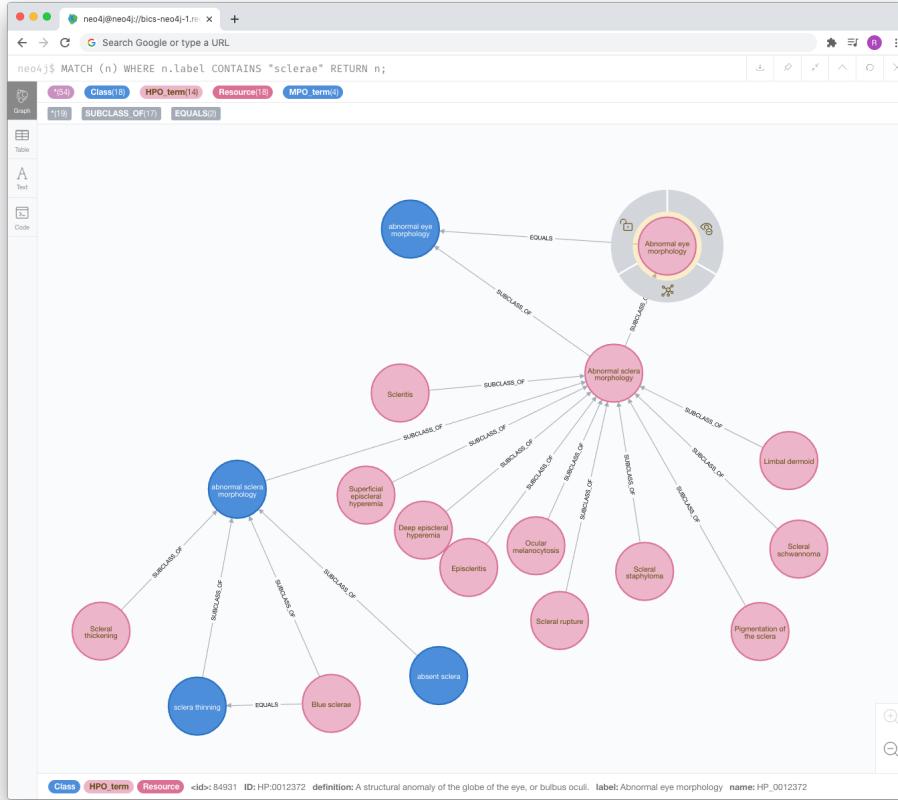
# Neo4j query example: Gene ortholog pairs

Nat2 ≈ NAT2, etc.



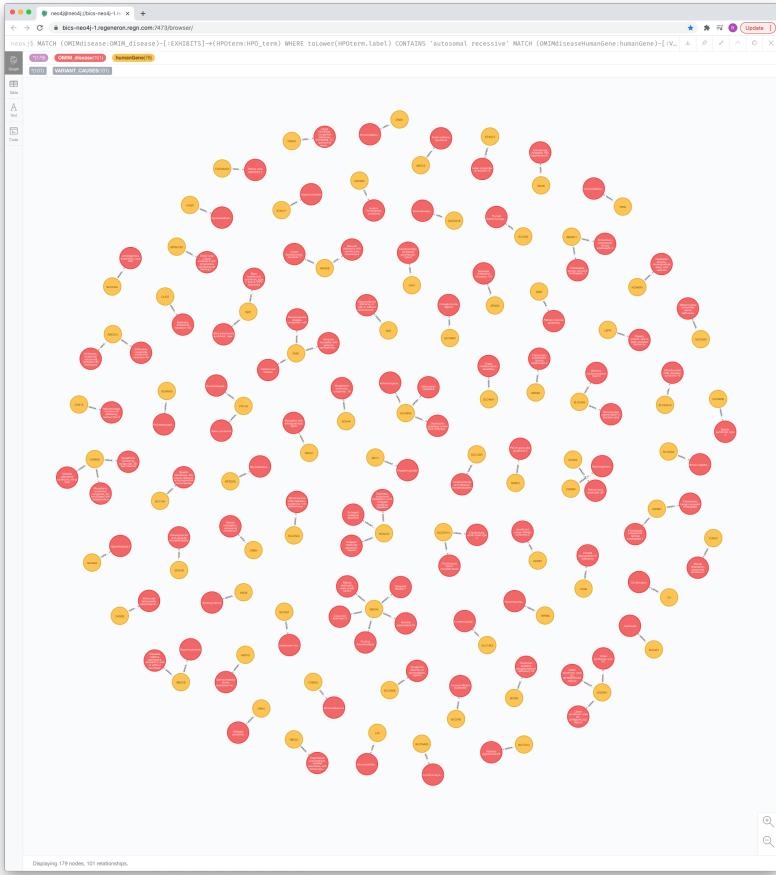
# Neo4j query example: Interspecies translation

“Sclera thinning” in mouse = “Blue sclerae” in human



# Neo4j query example: Monogenic diseases with viable KO

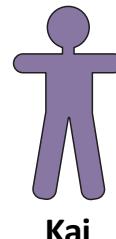
Diseases (red nodes) with only 1 human gene (yellow node) known to be causally connected (monogenic).



# A pharmaceutical example.

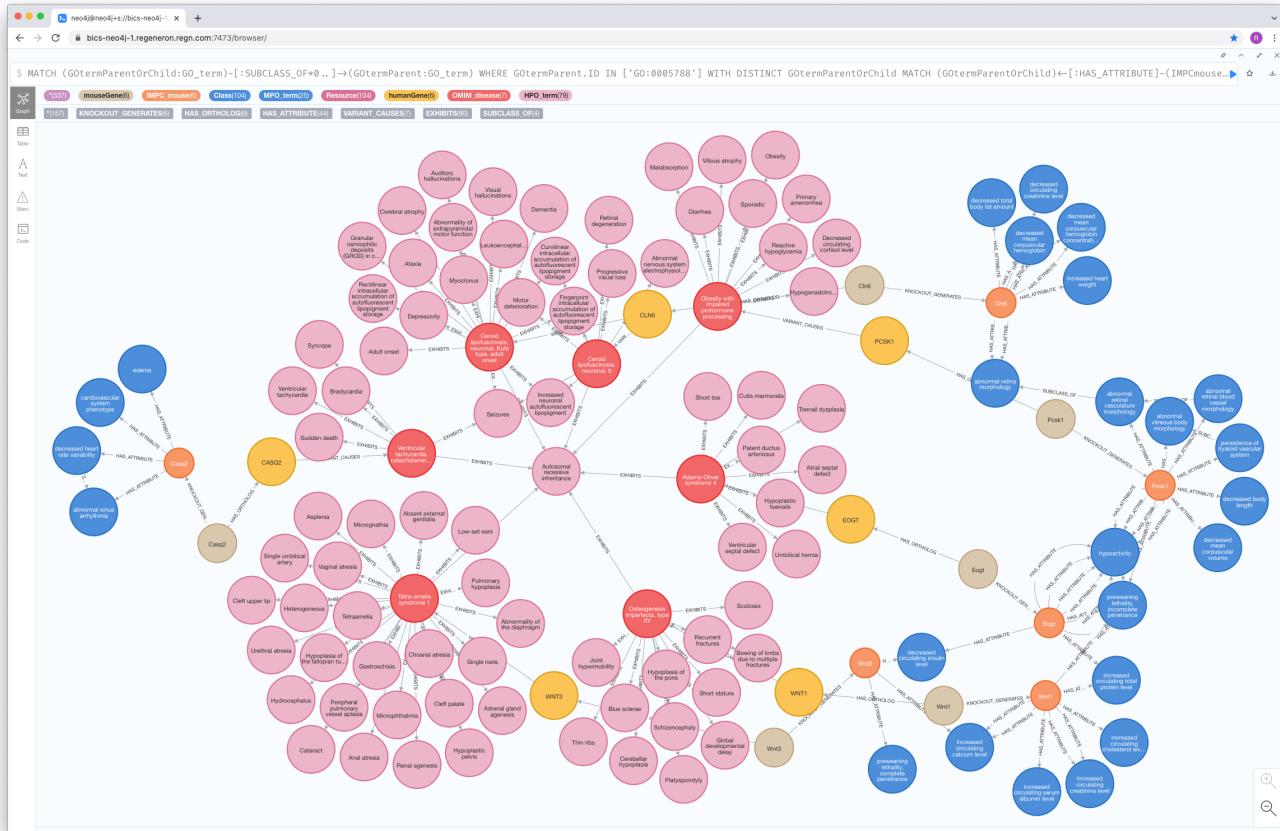
*“Kai aims to remedy disease using a therapeutic intervention technique that works well at targeting **ER (endoplasmic reticulum) lumen** proteins, and thereby genetic diseases caused by loss of function variants to ER lumen genes. The treatment needs to be testable in-house, so it should be relatively straightforward to generate a **disease model**, the **knockout of the causal gene needs to be non-lethal**, and there should be **sufficient time to intervene** before the onset of the disease.”*

***Which disease and mechanism should Kai decide to target?***



## Neo4j query example: Potentially viable diseases (red)

## Intermap subgraph for “endoplasmic reticulum lumen” GO term



# GENEONTOLOGY

Unifying Biology



## The Mammalian Phenotype Ontology





# Neo4j query example: Potentially viable diseases

Intermap subgraph for “endoplasmic reticulum lumen” GO term

```

neo4j@neo4j:s://bics-neo4j-1 ~ + 
← → ⌂ 🔍 https://bics-neo4j-1.regeneron.regn.com:7473/browser/
1 MATCH (G0termParentOrChild:GO_term)-[:SUBCLASS_OF*0 .. ]→(G0termParent:GO_term)
2 WHERE G0termParent.ID IN ['GO:0005788']
3 WITH DISTINCT G0termParentOrChild
4 MATCH (G0termParentOrChild)←[:HAS_ATTRIBUTE]-(IMPCmouseGene:mouseGene)-[:KNOCKOUT_GENERATES]→(IMPCmouse:IMPC_mouse)
5 WITH DISTINCT G0termParentOrChild, IMPCmouseGene, IMPCmouse
6 MATCH (undesiredMPOtermParentOrChild:MPO_term)-[:SUBCLASS_OF*0 .. ]→(undesiredMPOtermParent:MPO_term)
7 WHERE undesiredMPOtermParent.ID IN ['MP:0008762']
8 WITH DISTINCT
9 G0termParentOrChild,
10 IMPCmouseGene,
11 IMPCmouse,
12 SUM(SIZE((IMPCmouse)→(undesiredMPOtermParentOrChild))) AS undesiredMPOtermsPerIMPCmouse
13 WHERE undesiredMPOtermsPerIMPCmouse = 0
14 WITH DISTINCT G0termParentOrChild, IMPCmouseGene, IMPCmouse, undesiredMPOtermsPerIMPCmouse
15 MATCH (IMPCmouseGene)-[:HAS_ORTHOLOG]-(OMIMdiseaseHumanGene:humanGene)
16 WITH DISTINCT G0termParentOrChild, IMPCmouseGene, IMPCmouse, undesiredMPOtermsPerIMPCmouse, OMIMdiseaseHumanGene
17 MATCH (OMIMdiseaseHumanGene:humanGene)-[:VARIANT_CAUSES]→(OMIMdisease:OMIM_disease)-[:EXHIBITS]→(HP0term:HPO_term)
18 WHERE HP0term.ID IN ['HP:0000007']
19 WITH DISTINCT G0termParentOrChild, IMPCmouseGene, IMPCmouse, undesiredMPOtermsPerIMPCmouse, OMIMdiseaseHumanGene, OMIMdisease, HP0term
20 MATCH (OMIMdisease)←[:VARIANT_CAUSES]-(causalGenes:humanGene)
21 WITH
22 IMPCmouseGene,
23 IMPCmouse,
24 undesiredMPOtermsPerIMPCmouse,
25 OMIMdiseaseHumanGene,
26 OMIMdisease,
27 COUNT(DISTINCT((causalGenes)→(OMIMdisease))) AS genesPerOMIMdisease
28 WHERE genesPerOMIMdisease <= 1
29 MATCH (IMPCmouse)-[:HAS_ATTRIBUTE]→(IMPCmouseMPOterm:MPO_term)
30 MATCH (OMIMdisease)-[:EXHIBITS]→(OMIMdiseaseHP0term:HPO_term)
31 RETURN DISTINCT
32 IMPCmouseGene,
33 IMPCmouse,
34 IMPCmouseMPOterm,
35 OMIMdiseaseHumanGene,
36 OMIMdisease,
37 OMIMdiseaseHP0term;

```

# The API connection

# Wrap cypher-generating R code in plumber endpoint

The ease of using the plumber R package for constructing APIs

```

125 #* Find potential gene targets.
126 #* @param maxQtyGeneTargetsPerDisease The maximum quantity of causal genes permitted per OMIM disease.
127 #* @param HumanPhenotypeOntologyTermsToUseForIncludingOMIMDiseases Required array of Human Phenotype Ontology term IDs with which to include OMIM diseases. Default is "HP:0000007", which is "Autosomal recessive inheritance".
128 #* @param MammalianPhenotypeOntologyTermsToUseForExcludingIMPCmice Required array of Mammalian Phenotype Ontology term IDs with which to exclude IMPC mice. Default is "HP:0013292", which is "embryonic lethality prior to organogenesis".
129 #* @param GeneOntologyTermIDs Required array of Gene Ontology term IDs with which to include mouse genes.
130 #* @param maxQtyCausalGenesPermittedPerDisease Maximum quantity of causal genes permitted per OMIM disease.
131 endpoint_entitiesAnnotatedToGeneOntologyTerms <- function(
132   GeneOntologyTermIDs<-c("GO:0031383","GO:0031383","GO:1990701","GO:0030176","GO:0097637","GO:1905103","GO:0030285","GO:1983561","GO:0005765"),
133   MammalianPhenotypeOntologyTermsToUseForExcludingIMPCmice<-"HP:0008762", # embryonic lethality
134   HumanPhenotypeOntologyTermsToUseForIncludingOMIMDiseases<-"HP:0000007", # Autosomal recessive inheritance
135   maxQtyCausalGenesPermittedPerDisease=1
136 )
137 ~ {
138   # Entities
139   # Convert ontology term arrays into a strings of terms to search for.
140   GeneOntologyTermIDs_asString <- GeneOntologyTermIDs %>
141     convertToArrayOfQuotedElements(arrayOfConvert->)
142   MammalianPhenotypeOntologyTermsToUseForExcludingIMPCmice_asString <- MammalianPhenotypeOntologyTermsToUseForExcludingIMPCmice %>
143     convertToArrayOfQuotedElements(arrayOfConvert->)
144   HumanPhenotypeOntologyTermsToUseForIncludingOMIMDiseases_asString <- HumanPhenotypeOntologyTermsToUseForIncludingOMIMDiseases %>
145     convertToArrayOfQuotedElements(arrayOfConvert->)
146   # Retrieve entity results.
147   resultEntities <-
148   {
149     # Build query.
150     postgreQ
151       MATCH (GOtermParentOrChild:GO_term)-[:SUBCLASS_OF*0...]->(GOtermParent:GO_term)
152       WHERE GOtermParent.ID IN `GeneOntologyTermIDs_asString,')
153       WITH DISTINCT GOtermParentOrChild
154       MATCH (GOtermParentOrChild)-[:HAS_ATTRIBUTE]->(IMPCmouse:Gene:mouseGene)-[:KNOCKOUT_GENERATES]->(IMPCmouse:IMPC_mouse)
155       WITH DISTINCT GOtermParentOrChild, IMPCmouse
156       MATCH (undesiredPOTermParentOrChild:PO_term)-[:SUBCLASS_OF*0...]->(UndesiredPOTermParent:PO_term)
157       WHERE UndesiredPOTermParent.ID IN `MammalianPhenotypeOntologyTermsToUseForExcludingIMPCmice_asString,')
158       WITH DISTINCT
159         GOtermParentOrChild,
160         IMPCmouse,
161         IMPCmouse,
162         SIZE([currAnnotation IN COLLECT(DISTINCT(CIMPMouse))-[:HAS_ATTRIBUTE]->(UndesiredPOTermParentOrChild)) WHERE SIZE(currAnnotation) > 0]) AS undesiredPOTermsPerIMPMouse
163       UNWIND undesiredPOTermsPerIMPMouse AS undesiredPOTerm
164       WITH DISTINCT GOtermParentOrChild, IMPCmouse, undesiredPOTermsPerIMPMouse
165       MATCH (CIMPMouseGene)-[:HAS_ORTHOLOGY]->(OMIMdiseaseHumanGene:humanGene)
166       MATCH (OMIMdiseaseHumanGene)-[:VARIANT_CAUSES]->(OMIMdisease:OMIM_disease)-[:EXHIBITS]->(HPOterm:HPO_term)
167       WHERE HPOterm.ID IN `HumanPhenotypeOntologyTermsToUseForIncludingOMIMDiseases_asString,')
168       WITH DISTINCT GOtermParentOrChild, IMPCmouse, OMIMdisease, IMPCmouse, undesiredPOTermsPerIMPMouse, OMIMdiseaseHumanGene, HPOterm
169       MATCH (OMIMdisease)-[:VARIANT_CAUSES]->(causalGenes:humanGene)
170       WITH
171         IMPCmouseGene,
172         IMPCmouse,
173         undesiredPOTermPerIMPMouse,
174         OMIMdiseaseHumanGene,
175         OMIMdisease,
176         OMIMdisease,
177         SIZE([currAnnotation IN COLLECT(DISTINCT((causalGenes)-[:VARIANT_CAUSES]->(OMIMdisease))) WHERE SIZE(currAnnotation) > 0]) AS genesPerOMIMdisease
178       WHERE genesPerOMIMdisease <> `maxQtyCausalGenesPermittedPerDisease,'
179       RETURN DISTINCT
180         IMPCmouseGene.entrez_id AS mouse_gene_entrez_id,
181         IMPCmouseGene.geneSymbol AS mouse_gene_symbol,
182         IMPCmouseGene.marker_accession_id AS IMPC_mouse_id,
183         IMPCmouseGene.marker_symbol AS IMPC_mouse_id,
184         OMIMdiseaseHumanGene.entrez_id AS human_gene_entrez_id,
185         OMIMdiseaseHumanGene.geneSymbol AS human_gene,
186         OMIMdisease.omim_id AS OMIM_id,
187         OMIMdisease.disease_name AS OMIM_disease_name,
188         genesPerOMIMdisease AS genes_per_OMIM_disease,
189         ORDER BY mouse_gene_symbol, OMIM_disease_name,'
190       )
191   ~>
192   # Submit query.
193   call_neo4j[(
194     query=
195       con$graphDatabaseConnection
196   ) ~>
197     # Convert result list into a tibble of lists.
198     convertNodeListTableIntoTibble(Neo4jListTable<-)
199   # Return final table.
200   return(resultEntities)
201 }
202 ~

```



# Build entire API w/ databases. Publish to RStudio Connect.

The ease of using the plumber R package for constructing APIs

The screenshot shows the Swagger UI interface for the Intermap Engine API to Neo4j Graph Database. The URL is [https://bioinfo-connect.regeneron.regn.com/bics/intermap\\_graph\\_database\\_api/\\_docs\\_/](https://bioinfo-connect.regeneron.regn.com/bics/intermap_graph_database_api/_docs_/). The title is "Intermap Engine API to Neo4j Graph Database" version 1.0.0 OAS3. The API Description is "https://bioinfo-connect.regeneron.regn.com/bics/intermap\_graph\_database\_api/openapi.json". The "Servers" dropdown is set to "https://bioinfo-connect.regeneron.regn.com/bics/intermap\_graph\_database\_api/". The "default" section contains several POST requests:

- /list/available\_gene\_ontology\_terms List all available Gene Ontology terms.
- /list/available\_ontology\_terms List all available terms from a selected ontology.
- /search/identify\_target\_genes\_for\_monogenic\_recessive\_OMIM\_diseases\_with\_nonlethal\_mouse\_knockouts Find potential gene targets.
- /search/entities/entities\_annotated\_with\_go\_terms Find potential gene targets.
- /search/annotations\_per\_entity/go\_terms\_per\_gene Find GO annotations to mouse genes.
- /search/annotations\_per\_entity/mpo\_terms\_per\_impc\_mouse Find MPO annotations to IMPC mice.
- /search/annotations\_per\_entity/hpo\_terms\_per\_omim\_disease Find HPO annotations to OMIM diseases.
- /search/complete\_ontological\_subtree\_for\_IDS Retrieve the ontological subtree including and subsumed by a selection of nodes (all self and descendant nodes).
- /search/ontological\_ancestors\_and\_descendants\_for\_IDS Retrieve the ontological ancestors and descendants of a selection of nodes.

A note at the bottom states: "GET / RStudio Connect added this endpoint to redirect to the API docs by default. Once you define a base handler (i.e.: 'GET /'), RStudio Connect will stop adding this redirector."



# The Intermap API makes future advancements FAIR game.

FAIR principles of good collaboration.



## 1. Findable.

- All selectable terms listed in endpoint.



## 2. Accessible.

- Open to all within intranet.



## 3. Interoperable.

- Works with all programming languages, due to inherent language agnosticism.

## 4. Reusable.

- Can be used repeatedly to build a growing range of scientific decision enablement tools.

# The interface

# Build shiny app w/API. Publish to RStudio Connect.

Focus on rapidly providing both multiomic hypotheses and the computational reasoning behind them.

Intermap Target Detector

Select GO (Gene Ontology) terms to identify genes, mice, and diseases that meet the criteria listed below the menu.

Gene Ontology term(s)

endoplasmic reticulum lumen

Understand the 1 current query term

Inclusion criteria:

1. Mouse genes are known to exhibit ANY of the properties selected in the GO term menu.
2. Mouse genes generate non-embryonic-lethal knockout mice (IMPC).
3. Mouse genes have human gene orthologs that are known to host variants that cause genetic (OMIM) disease.
4. OMIM diseases are monogenic (known to be caused by variants in only 1 gene).
5. OMIM diseases are known to have an autosomal recessive mode of inheritance (HPO terms).

Share session (copy & paste URL) or open in new window

Click the entities (genes, mice, and diseases) below for details on their annotations (Gene Ontology, Mammalian Phenotype Ontology, and Human Phenotype Ontology, respectively), and additional links to external databases.

Show 25 entries

Search:

Mouse Gene	IMPC Mouse	Human Gene	OMIM Disease
Casq2	Casq2	CASQ2	Ventricular tachycardia, catecholaminergic polymorphic, 2
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, 6
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, Kufs type, adult onset
Eogt	Eogt	EOGT	Adams-Oliver syndrome 4
Pcsk1	Pcsk1	PCSK1	Obesity with impaired prohormone processing
Wnt1	Wnt1	WNT1	Osteogenesis imperfecta, type XV
Wnt3	Wnt3	WNT3	Tetra-amelia syndrome 1

Mouse Gene    IMPC Mouse    Human Gene    OMIM Disease

Showing 1 to 7 of 7 entries

Previous    1    Next

Gene Ontology query results.



R Studio Connect



Kai  
(user)

notices

**"Adult onset** suggests that there might be enough time to intervene therapeutically..."

REGENERON®  
RESEARCH AND  
DEVELOPMENT

BIOINFORMATICS  
CORE

# Network and tabular explanation of reasoning

*Results must be annotated to any Gene Ontology term(s) directly-selected with user query or any indirectly-selected (more specific) child term(s).*

*Computational reasoning*

Intermap Target Detector

https://bioinfo-connect.regeneron.regn.com/bics/intermap\_target\_detector/?GOtermIDs=GO:0005788

Ontology subtree selected for query

Qty. terms: 8 (1 direct + 7 indirect)

Graphical view Textual view

Close

Selection Label Definition Subtree

Show entries Search:

Indirect Terminal cisterna lumen The region between the inner and outer lipid bilayers of the terminal cisterna envelope. This space is enriched in calsequestrin. Cellular component

Indirect Longitudinal sarcoplasmic reticulum lumen The region between the inner and outer lipid bilayers of the longitudinal sarcoplasmic reticulum envelope. The longitudinal sarcoplasmic reticulum lumen is continuous with the lumen contained within the terminal cisternae. Cellular component

Direct Endoplasmic reticulum lumen The volume enclosed by the membranes of the endoplasmic reticulum. Cellular component

Indirect Sarcoplasmic reticulum lumen The volume enclosed by the membranes of the sarcoplasmic reticulum. Cellular component

Indirectly selected

Directly selected

Endoplasmic reticulum lumen

Sarcoplasmic reticulum lumen

Cortical endoplasmic reticulum lumen

Terminal cisterna lumen

Endoplasmic reticulum lumen

Sarcolemma

Perinuclear endoplasmic reticulum lumen

Longitudinal sarcoplasmic reticulum lumen

Diagram illustrating the ontology subtree selected for query GO:0005788. The diagram shows a network of nodes representing cellular compartments, with arrows indicating relationships. A central red node is labeled "Endoplasmic reticulum lumen". It has three outgoing arrows to other nodes: "Sarcoplasmic reticulum lumen" (labeled "Directly selected"), "Cortical endoplasmic reticulum lumen", and "Terminal cisterna lumen". The "Cortical endoplasmic reticulum lumen" node has an arrow pointing to a gray node labeled "Endoplasmic reticulum lumen". The "Terminal cisterna lumen" node has an arrow pointing to a gray node labeled "Terminal cisterna lumen". The "Endoplasmic reticulum lumen" node has two outgoing arrows: one to a gray node labeled "Sarcolemma" and another to a gray node labeled "Perinuclear endoplasmic reticulum lumen". The "Sarcolemma" node has an arrow pointing to a gray node labeled "Longitudinal sarcoplasmic reticulum lumen". The "Perinuclear endoplasmic reticulum lumen" node has an arrow pointing to a gray node labeled "Sarcoplasmic reticulum lumen". A legend on the left indicates that red nodes represent "Directly selected" terms and gray nodes represent "Indirectly selected" terms.

# Gene Ontology terms annotated to mouse gene

The screenshot shows a web-based application titled "Intermap Target Detector". The main title bar says "Intermap Target Detector" and the URL is "https://bioinfo-connect.regeneron.regn.com/bics/intermap\_target\_detector/?GOtermIDs=GO:0005788". The main content area is a modal window titled "Mouse Gene: Cln6". Inside, there's a table listing Gene Ontology terms. The table has columns: "Present in Query?", "Gene Ontology Category", "Gene Ontology Term", and "QuickGO". There are 10 entries shown. The first entry is "TRUE" for "cellular\_component" with term "endoplasmic reticulum lumen" and QuickGO ID "GO:0005788". The other 9 entries are all "FALSE" for "biological\_process". The categories include "cellular macromolecule catabolic process", "cholesterol metabolic process", "ganglioside metabolic process", "glycosaminoglycan metabolic process", "locomotion involved in locomotory behavior", "lysosomal lumen acidification", "lysosome organization", "positive regulation of proteolysis", and "visual perception". The QuickGO IDs for these are GO:0044265, GO:0008203, GO:0001573, GO:0030203, GO:0031987, GO:0007042, GO:0007040, GO:0045862, and GO:0007601 respectively. At the bottom of the table are search and filter fields: "Present in Query?", "Gene Ontology Category", "Gene Ontology Term", and "QuickGO". Navigation buttons "Previous", "1", and "Next" are also present.

Present in Query?	Gene Ontology Category	Gene Ontology Term	QuickGO
TRUE	cellular_component	endoplasmic reticulum lumen	GO:0005788
FALSE	biological_process	cellular macromolecule catabolic process	GO:0044265
FALSE	biological_process	cholesterol metabolic process	GO:0008203
FALSE	biological_process	ganglioside metabolic process	GO:0001573
FALSE	biological_process	glycosaminoglycan metabolic process	GO:0030203
FALSE	biological_process	locomotion involved in locomotory behavior	GO:0031987
FALSE	biological_process	lysosomal lumen acidification	GO:0007042
FALSE	biological_process	lysosome organization	GO:0007040
FALSE	biological_process	positive regulation of proteolysis	GO:0045862
FALSE	biological_process	visual perception	GO:0007601

# Mammalian Phenotype Ontology terms annotated to mouse

The screenshot shows the "Intermap Target Detector" application window. In the main panel, under "Gene Ontology term(s)", the term "endoplasmic reticulum lumen" is selected. Below it, the "Inclusion criteria:" section lists five conditions for identifying genes, mice, and diseases. A modal dialog box is open, titled "IMPC Mouse: Cln6". This dialog displays phenotype annotations categorized by MPO Term and OMIM Disease. The MPO Terms listed are: abnormal retina morphology, decreased circulating creatinine level, decreased mean corpuscular hemoglobin, decreased mean corpuscular hemoglobin concentration, decreased total body fat amount, and increased heart weight. The OMIM Diseases listed are: Ventricular tachycardia, catecholaminergic polymorphic, 2, Ceroid lipofuscinosis, neuronal, 6, Ceroid lipofuscinosis, neuronal, Kufs type, adult onset, Adams-Oliver syndrome 4, Obesity with impaired prohormone processing, Osteogenesis imperfecta, type XV, and Tetra-amelia syndrome 1.

Select GO (Gene Ontology) terms to identify genes, mice, and diseases that meet the criteria listed below the menu.

Gene Ontology term(s)

endoplasmic reticulum lumen

Understand the 1 current query term

Inclusion criteria:

1. Mouse genes are known to exhibit ANY of the properties selected in the GO term menu.
2. Mouse genes generate non-embryonic-lethal knockout mice (IMPC).
3. Mouse genes have human gene orthologs that are known to host variants that cause genetic (OMIM) disease.
4. OMIM diseases are monogenic (known to be caused by variants in only 1 gene).
5. OMIM diseases are known to have an autosomal recessive mode of inheritance (HPO terms).

IMPC Mouse: Cln6

Show 10 entries Search:

MPO Term

abnormal retina morphology  
decreased circulating creatinine level  
decreased mean corpuscular hemoglobin  
decreased mean corpuscular hemoglobin concentration  
decreased total body fat amount  
increased heart weight

OMIM Disease

Ventricular tachycardia, catecholaminergic polymorphic, 2  
Ceroid lipofuscinosis, neuronal, 6  
Ceroid lipofuscinosis, neuronal, Kufs type, adult onset  
Adams-Oliver syndrome 4  
Obesity with impaired prohormone processing  
Osteogenesis imperfecta, type XV  
Tetra-amelia syndrome 1

Showing 1 to 6 of 6 entries Previous 1 Next

Showing 1 to 7 of 7 entries Previous 1 Next

# Gene Ontology terms annotated to identified human gene

The screenshot shows a web-based application titled "Intermap Target Detector". The main title bar says "Intermap Target Detector" and the URL is "https://bioinfo-connect.regeneron.regn.com/bics/intermap\_target\_detector/?GOtermIDs=GO:0005788". A modal window is open, titled "Human Gene: CLN6". The modal contains a table with the following data:

Present in Query?	Gene Ontology Category	GO Term	QuickGO
TRUE	cellular_component	endoplasmic reticulum lumen	GO:0005788
FALSE	biological_process	cellular macromolecule catabolic process	GO:0044265
FALSE	biological_process	cholesterol metabolic process	GO:0008203
FALSE	biological_process	ganglioside metabolic process	GO:0001573
FALSE	biological_process	glycosaminoglycan metabolic process	GO:0030203
FALSE	biological_process	locomotion involved in locomotory behavior	GO:0031987
FALSE	biological_process	lysosomal lumen acidification	GO:0007042
FALSE	biological_process	lysosome organization	GO:0007040
FALSE	biological_process	positive regulation of proteolysis	GO:0045862
FALSE	biological_process	protein catabolic process	GO:0030163

Below the table are four input fields: "Present in Query?", "Gene Ontology Category", "GO Term", and "QuickGO". There are also "Previous" and "Next" buttons at the bottom right of the modal.

# Human Phenotype Ontology terms annotated to disease

The screenshot shows a web-based application window titled "Intermap Target Detector". The URL in the address bar is [https://bioinfo-connect.regeneron.regn.com/bics/intermap\\_target\\_detector/?GOtermIDs=GO:0005788](https://bioinfo-connect.regeneron.regn.com/bics/intermap_target_detector/?GOtermIDs=GO:0005788). The main content area displays a list of Human Phenotype Ontology (HPO) terms associated with the specified OMIM Disease: Ceroid lipofuscinosis, neuronal, Kufs type, adult onset. The terms listed are:

- Abnormality of extrapyramidal motor function
- Adult onset
- Ataxia
- Auditory hallucinations
- Autosomal recessive inheritance
- Cerebral atrophy
- Curvilinear intracellular accumulation of autofluorescent lipopigment storage material
- Dementia
- Depressivity
- Fingerprint intracellular accumulation of autofluorescent lipopigment storage material

Below the list, there is a search bar labeled "Search:" and a "Close" button.

# External OMIM documentation page for disease

The screenshot shows the OMIM entry for #204300, which corresponds to CEROID LIPOFUSCINOSIS, NEURONAL, 4A (KUFS TYPE), AUTOSOMAL RECESSIVE; CLN4A. The page includes a sidebar with navigation links like About, Statistics, Downloads, Contact Us, MIMmatch, Donate, Help, and a search bar. The main content area displays the title, phenotype-gene relationships, and a table of inheritance and gene information. A sidebar on the right lists external links to various databases such as Protein, Clinical Resources (Clinical Trials, EuroGentest, Genetic Alliance, GTR, GARD, Orphanet), Variation, and Animal Models.

#204300  
Table of Contents

Title  
Phenotype-Gene Relationships  
Clinical Synopsis  
Phenotypic Series

Text  
Description  
Nomenclature  
Clinical Features  
Inheritance  
Molecular Genetics

See Also  
References  
Contributors  
Creation Date  
Edit History

# 204300

CEROID LIPOFUSCINOSIS, NEURONAL, 4A (KUFS TYPE),  
AUTOSOMAL RECESSIVE; CLN4A

ICD+  
# 204300

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
15q23	Ceroid lipofuscinosis, neuronal, 4A (Kufs type), autosomal recessive	204300	AR	3	CLN6	606725

Clinical Synopsis ▾  
Phenotypic Series ▾  
PheneGene Graphics ▾

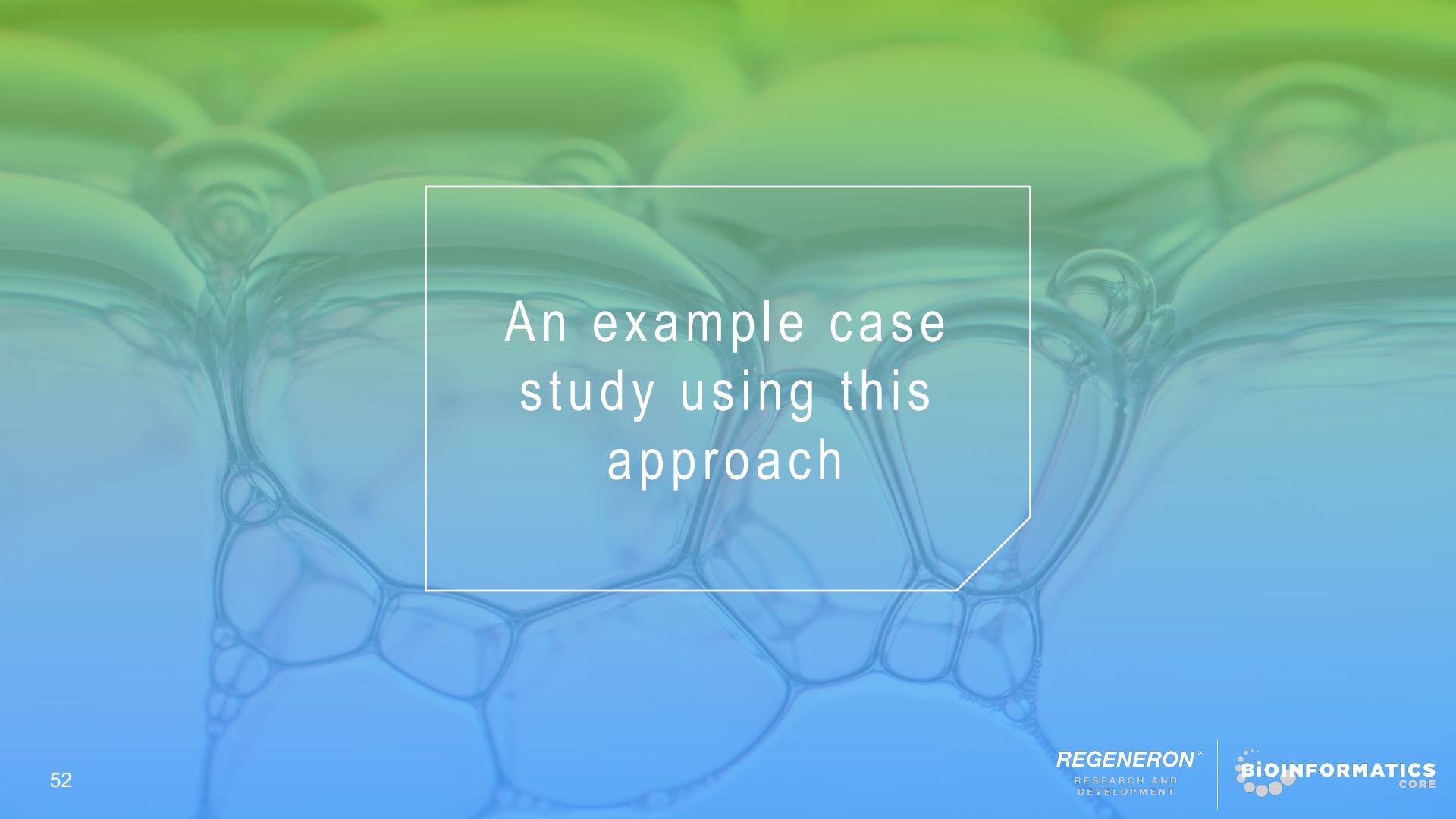
▼ TEXT

A number sign (#) is used with this entry because autosomal recessive neuronal ceroid lipofuscinosis-4A (CLN4A) is caused by homozygous or compound heterozygous mutation in the CLN6 gene (606725) on chromosome 15q23.

Biallelic mutation in the CLN6 gene can also cause a variant late infantile form of CLN (CLN6; 601780).

External Links

- Protein
- Clinical Resources
  - Clinical Trials
    - EuroGentest
    - Genetic Alliance
    - GTR
    - GARD
    - Orphanet
  - Variation
  - Animal Models

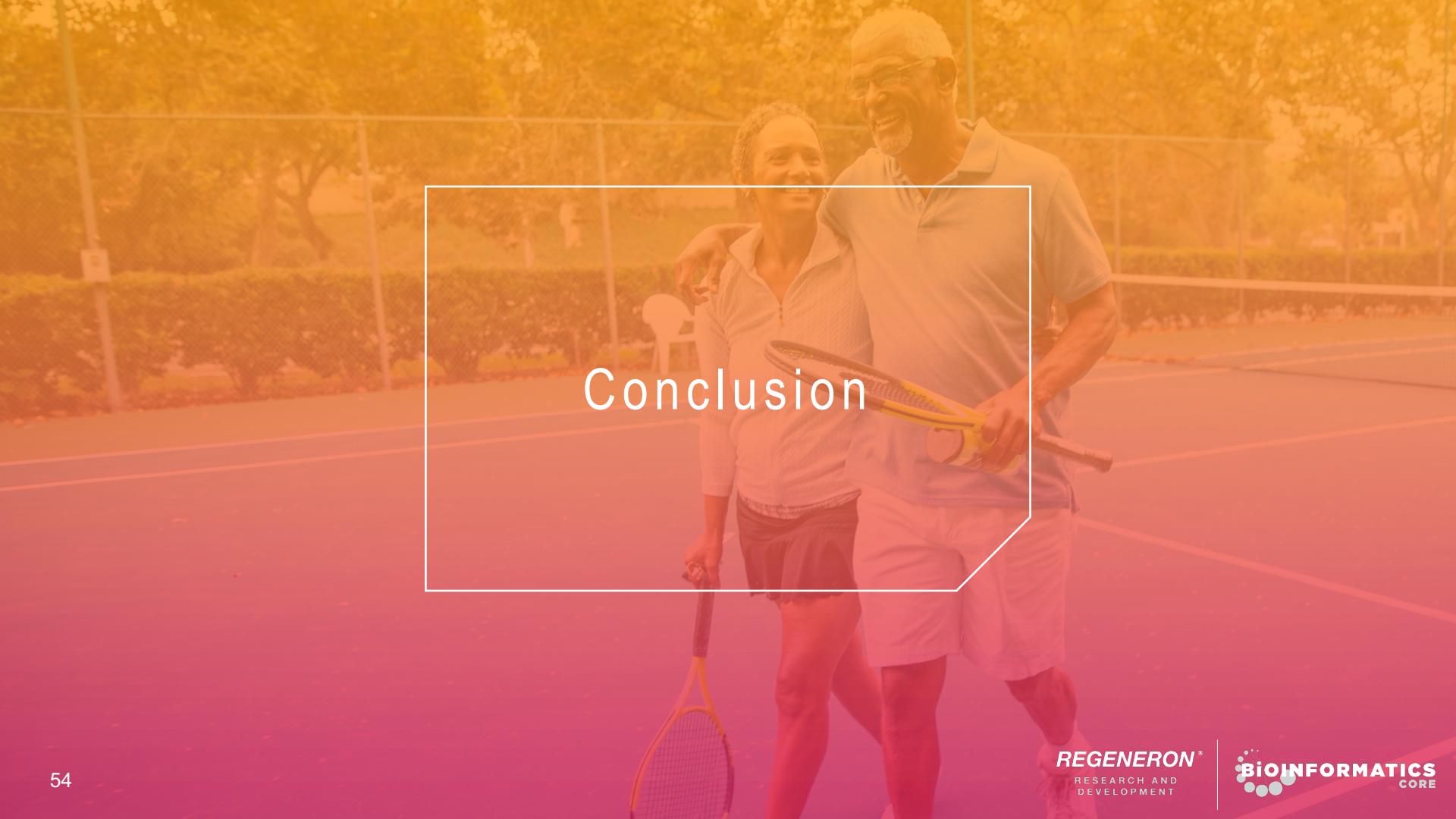


An example case  
study using this  
approach

# An optimized way to generate therapeutic target hypotheses

The screenshot shows the 'Intermap Target Detector' application interface. On the left, a sidebar displays the query term 'endoplasmic reticulum lumen'. Below it, the 'Inclusion criteria:' section lists five conditions for gene selection. The main content area shows a table of 7 entries, each mapping a mouse gene to its human ortholog and associated diseases. The table includes columns for 'Mouse Gene', 'IMPC Mouse', 'Human Gene', and 'OMIM Disease'. The first entry is Casq2, which maps to CASQ2 and is linked to 'Ventricular tachycardia, catecholaminergic polymorphic, 2'. Other entries include Cln6, Eogt, Pcsk1, Wnt1, and Wnt3.

Mouse Gene	IMPC Mouse	Human Gene	OMIM Disease
Casq2	Casq2	CASQ2	Ventricular tachycardia, catecholaminergic polymorphic, 2
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, 6
Cln6	Cln6	CLN6	Ceroid lipofuscinosis, neuronal, Kufs type, adult onset
Eogt	Eogt	EOGT	Adams-Oliver syndrome 4
Pcsk1	Pcsk1	PCSK1	Obesity with impaired prohormone processing
Wnt1	Wnt1	WNT1	Osteogenesis imperfecta, type XV
Wnt3	Wnt3	WNT3	Tetra-amelia syndrome 1

A photograph of a smiling senior couple on a tennis court. The man, wearing a light blue polo shirt and khaki shorts, has his arm around the woman. The woman, wearing a white zip-up top and dark shorts, holds a tennis racket. They are standing on a red clay court with a chain-link fence and trees in the background.

# Conclusion

# **Summary:** A way to use computational reasoning with R to eliminate obstacles to scientific progress

Prevent the mechanism, prevent the disease.

**The Intemap approach is a starting point. Many other uses, including:**

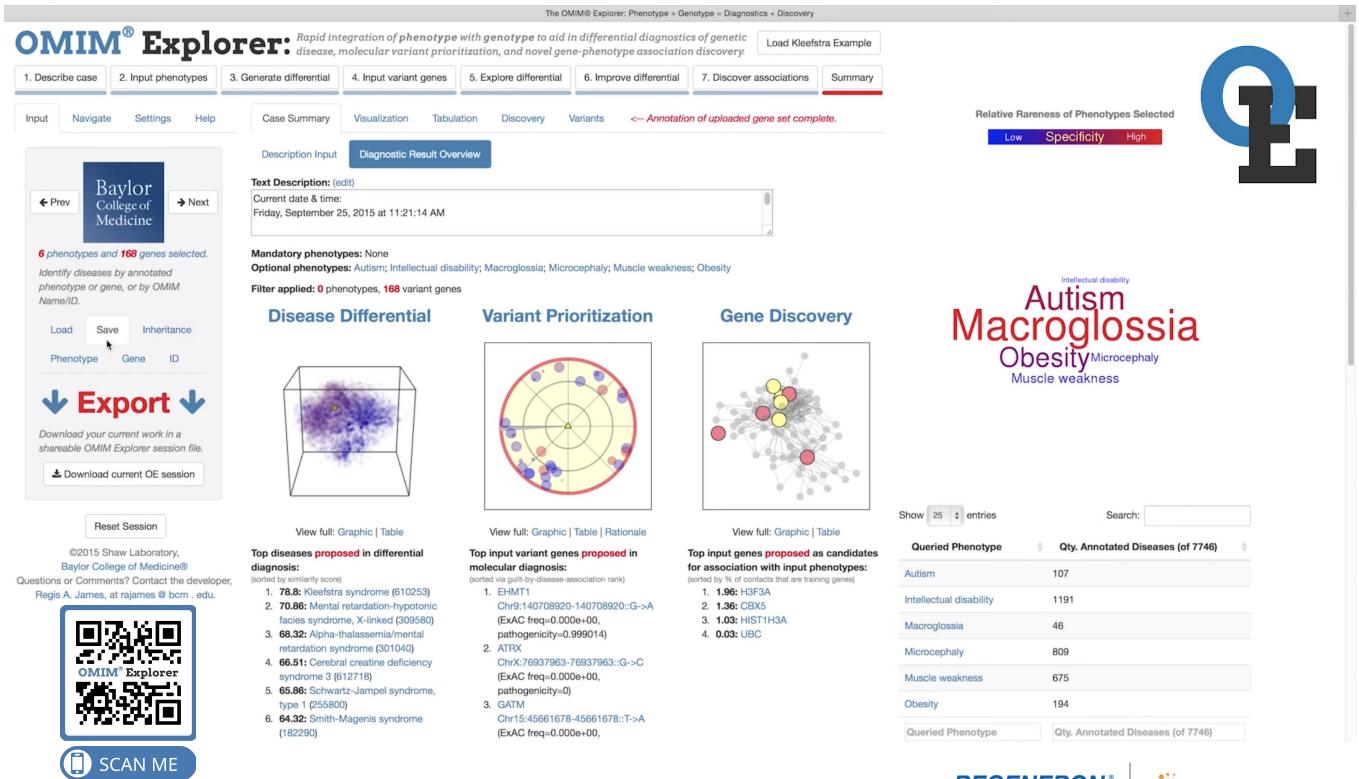
- Drug repurposing
- Clinical diagnostics in genetic disease patients
- Many others



# A previous approach to integrative clinical diagnostics

PhD proof of concept only. You can build a similar tool in your own environment.

1. Applied semantic similarity analysis techniques to a biomedical knowledgebase on-the-fly to help enable clinical geneticists decide on phenotype-based, exome-filtered diagnoses for patients.
2. R-based shiny app that used an older version of the plumber API package and a MySQL backend for caching.



# Acknowledgements

The collaborator community of which I am privileged to be a member.

- Marina Lehmkuhl
- Calvin Chen
- Christian Buhay
- Mickey Atwal
- Maria Praggastis
- Katherine Cygnar
- Eric Engelhard
- Melissa Haendel
- Peter Robinson
- Uniprot
- NCBI
- IMPC
- MGI
- The Monarch Initiative
- OMIM
- The bioinformatic data scientist community at large (building on their public data)
- The R in Pharma Organizing Committee