

ASK for confident decisions: Harnessing the power of semantics



Erich Gombocz

Vice President, Chief Scientific Officer
IO Informatics, Inc., Berkeley, CA, USA.

Outline

Challenges
in Predictive
Biology

Semantic
Integration

Network
Graph to
Model

Semantic
Queries
(SPARQL)

ASK:
Collection
s of
Biomarker
Signatures

Confident
Patient
Screening

Implications
for PM

Challenges

Data Coherence

- Source heterogeneity
- Taxonomies, ontologies
- Synonyms, non-standardized vocabularies
- Normalization challenges

Integration Woes

- Hesitance in Data Sharing
- Security, compliance
- Complexity in meaningful integration of experimental and clinical data

Lack of Tools

- Intuitive, science-driven tools for hypothesis and decision support hard to find
- Requirement for complex ad-hoc queries demanding
- Hypothesis generation requires more than domain expertise

Qualification

- Biomarker classifier validation with mechanistic and functional insights is demanding
- Unifying public resources and internal datasets and proper weighing of markers is non-trivial

Applicability

- Ease of access and use for clinicians at multiple levels required
- Confidence criteria for decision support need to be accepted for patient screening



Semantic Integration

Dynamic, extensible
data model –

*modeling data as a
graph is more flexible
and intuitive than a
using a table*

Relational databases
scale well to size -

*but what about
scaling to complexity
of biological
systems?*

Methodology

Create Semantic Framework

- Merge and map results from multiple modalities into a semantic framework
- Visualize, investigate and analyze data relationships

Qualify Biological Validity

- Qualify viability of experimental correlation networks through incorporation of mechanistic public knowledge to ensure their functional biology

Build Model(s)

- Save the resulting sub-network as semantic query (SPARQL)
- Each semantic query (SPARQL profile) represents a signature of a biological process
- Set criteria for ranges and weights on biomarkers

Applied Semantic Knowledgebase (ASK)

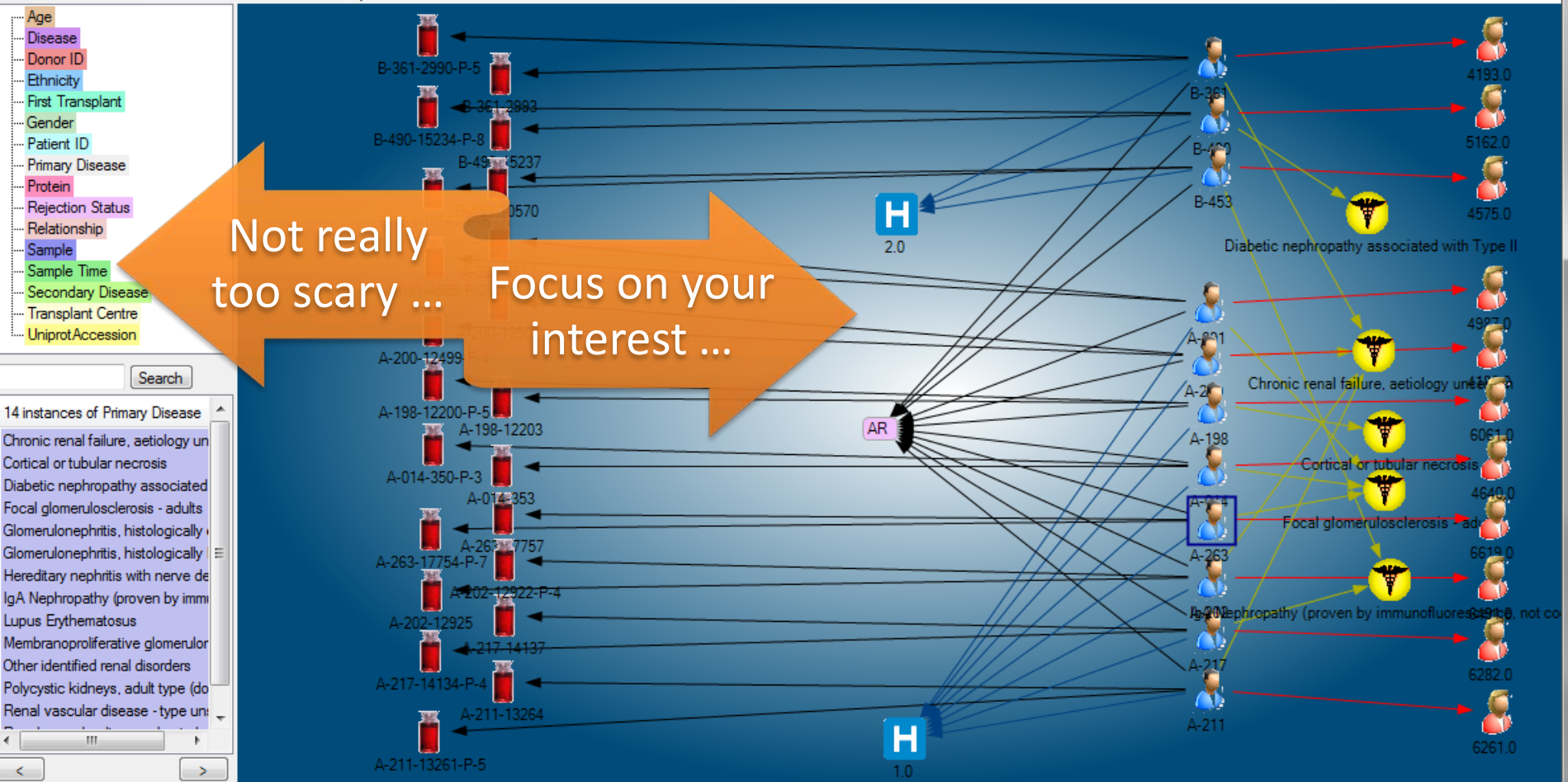
- Represent your model as array of semantic queries in an Applied Semantic Knowledgebase (ASK)
- Validate each model through iterative refinement with additional test cases

Decision supported Screening

- Apply ASK arrays to unknowns for screening
- Use scoring of the match ("hits-to-fit") for informed decision-making with high confidence.
- Provide a web-based real-time actionable patient care

-

Not really too scary ... Focus on your interest ...



Entity	Relationship	Entity
A-263	hasDonor	6619
A-263	hasEthnicity	Caucasian
A-263	hasPrimaryDisease	Focal glomerulosclerosis - adults
A-263	hasSecondaryDisease	NULL
A-263	isFirst Transplant	Yes

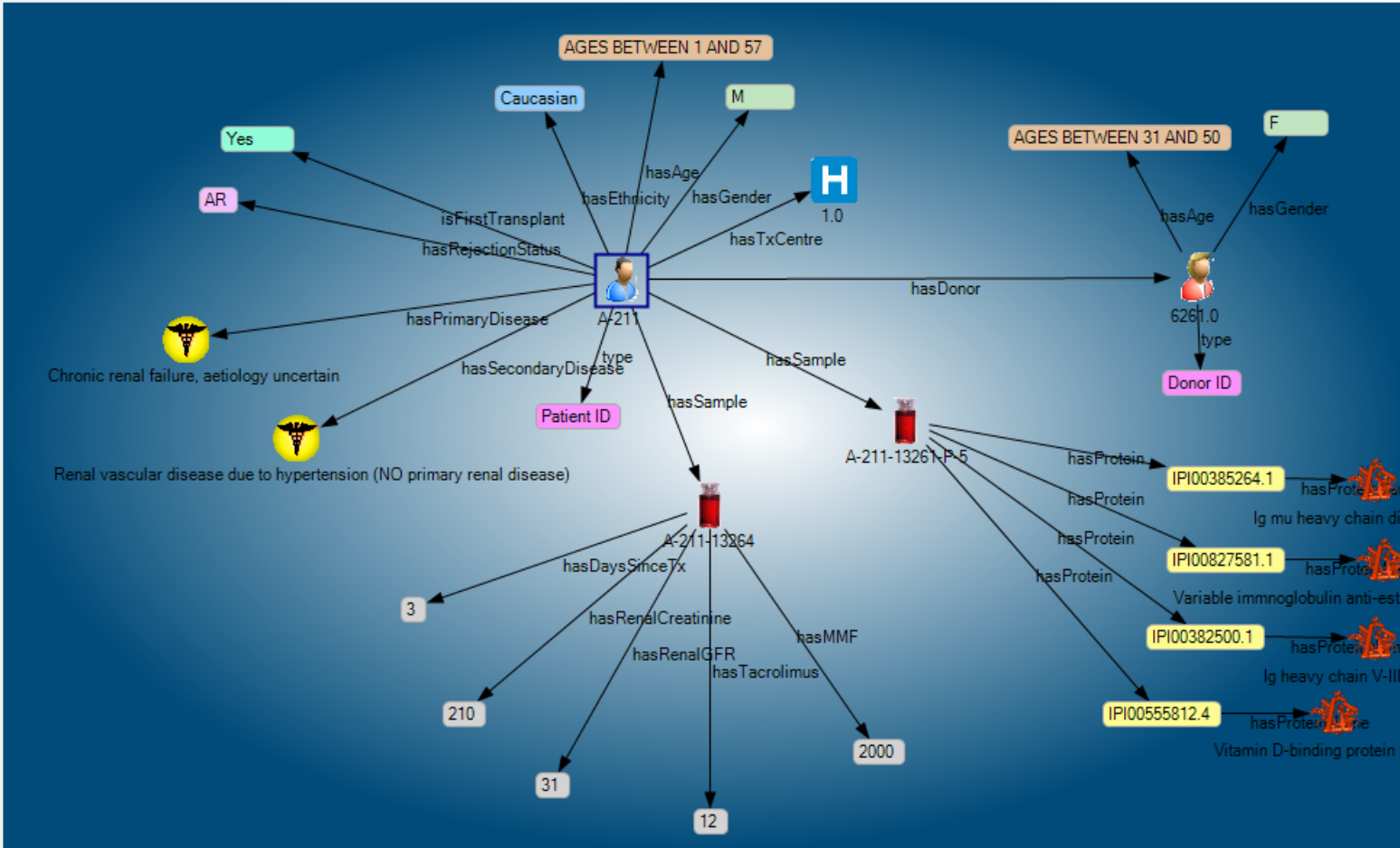
- Age
- Disease
- Donor ID
- Ethnicity
- First Transplant
- Gender
- Patient ID
- Primary Disease
- Protein
- Rejection Status
- Relationship
- Sample
- Sample Time
- Secondary Disease
- Transplant Centre
- UniprotAccession

Search

500 instances of Protein(0...500)

- Amyloid lambda 6 light chain variable re
- Amyloid lambda 6 light chain variable re
- Anchor protein
- Angiopoietin-related protein 3 precursor
- Angiotensinogen precursor
- AngRem104
- ANKRD26 protein
- ANKS3 protein
- Ankyrin-3
- Ankyrin G119
- Ankyrin repeat and SAM domain-contai
- Ankyrin repeat and zinc finger domain-c
- Ankyrin repeat domain-containing prote

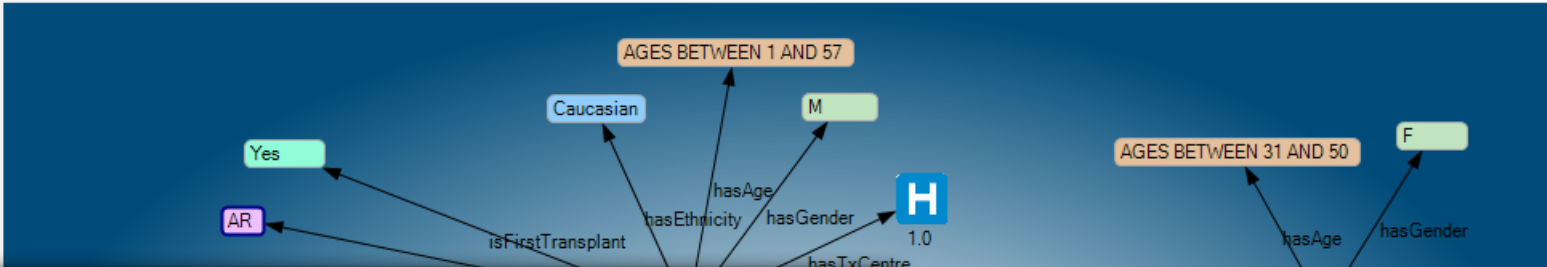
< >



Entity	Relationship	Entity
A-211	hasDonor	6261.0
A-211	hasEthnicity	Caucasian
A-211	hasPrimaryDisease	Chronic renal failure, aetiology uncertain
A-211	hasSecondaryDisease	Renal vascular disease due to hypertension (NO primary renal disease)
A-211	isFirstTransplant	Yes

- Relations
- hasSample
 - hasRejectionStatus
 - hasTxCentre
 - hasGender
 - hasAge
 - hasEthnicity
 - hasPrimaryDisease
 - hasSecondaryDisease
 - isFirstTransplant

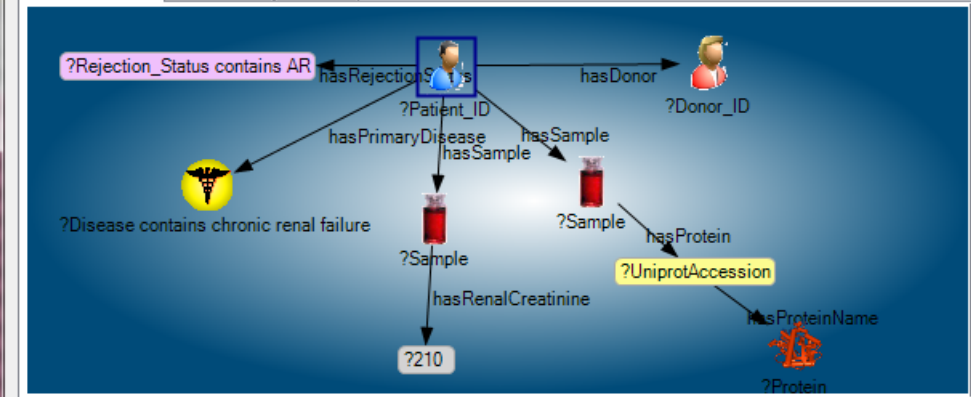
- Age
- Disease
- Donor ID
- Ethnicity
- First Transplant
- Gender
- Patient ID
- Primary Disease
- Protein
- Rejection Status



Query Tool

SPARQL Endpoint: Currently Loaded Dataset

Graphical Query Text Search SPARQL



Patient ID	Donor ID	Rejecti...	Disease	Protein
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Protein
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Isoform A of Coagulation factor VII p
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Keratin-type I cytoskeletal 10
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Peroxisomal membrane protein PEX
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Protein kinase C gamma type
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Isoform 1 of Neurofibromin
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	hypothetical protein LOC124565 iso
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Isoform 1 of Solute carrier family 12
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Apolipoprotein
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	Isoform 1 of Netrin receptor UNC5D
A-211	6261.0	AR	Chronic renal failure, aetiology uncertain	614 kDa protein

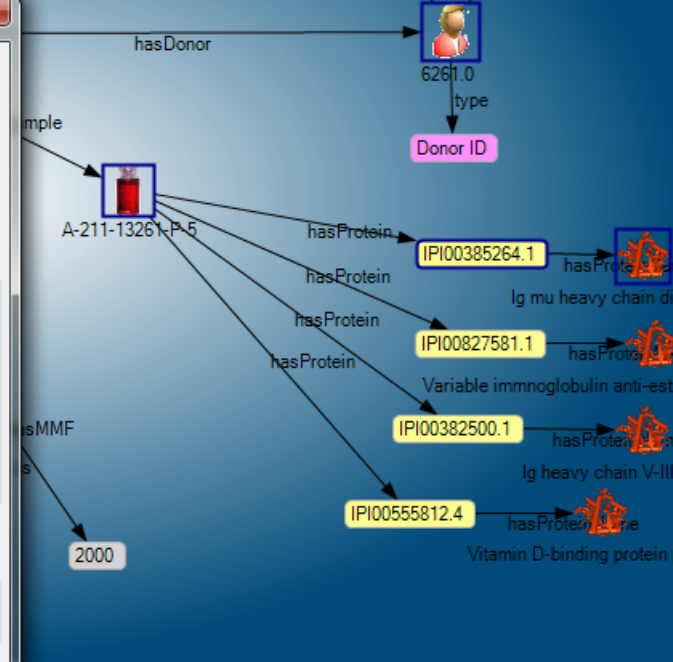
2993 results returned.

Run Query
Stop Query

Query Graph
Clear
Load
Save
Graph Query

Main Graph
Clear Main Graph

Results
Graph Results
Import to Memory
Export



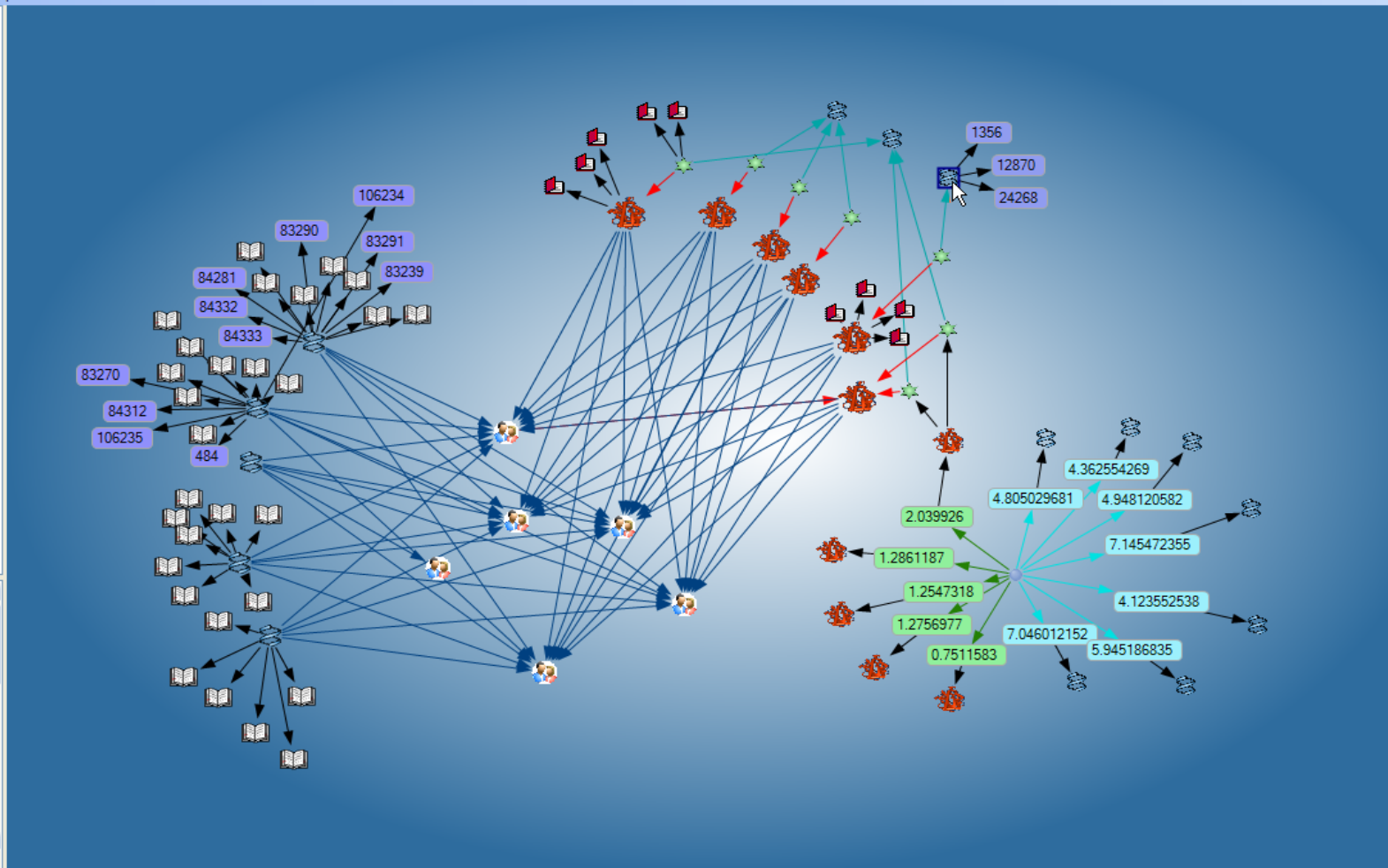
Update ☒ Automatically

- Relations
- hasSample
- hasRejectionStatus
- hasTxCentre
- hasGender
- hasAge
- hasEthnicity
- hasPrimaryDisease
- hasSecondaryDisease
- isFirstTransplant

- Entrez
 - BioSystems
 - Gene
 - OMIM
 - Protein
- Gene
 - Chromosomal Location
 - Gene ID
 - Gene Name
 - Gene Symbol
 - Gene Title
- Patient
 - Controls
- Platform
 - Gene Expression
- Protein
 - EC Enzyme Classification
 - IPI Accession Number
 - PGC
 - Protein ID
 - Protein Name
 - Protein Symbol
 - Protein Value
- Scaler
- Visualizer

38 instances of OMIM

ACERULOPLASMINEMIA
ADIPOCYTE-, C1q-, AND COLLAGEN DOM
ADIPONECTIN, SERUM LEVEL OF, QUAN
ADIPONECTIN, SERUM LEVEL OF, QUAN
ADIPONECTIN, SERUM LEVEL OF, QUAN
ANTIGEN DEFINED BY MONOCLONAL AN
BARE LYMPHOCYTE SYNDROME, TYPE
BETA-1,3-GLUCURONYLTRANSFERASE
BETA-2-MICROGLOBULIN; B2M
BODY MASS INDEX; BMI
CAPPING PROTEIN, ALPHA-3; CAPZA3
CEREBRAL PALSY, ATAXIC, AUTOSOMA
CEREBRAL PALSY, SPASTIC QUADRIPLE
CERULOPLASMIN; CP
CERULOPLASMIN; CP



Entity List Entity Details Relationship

Entity	Relationship	Entity
CP	hasEntrezItem	12870
CP	hasEntrezItem	1356
CP	hasEntrezItem	24268
CP	hasEntrezItem	3562578
CP	hasEntrezItem	3882101
CP	geneTotalRecords	5731
CP	hasEntrezItem	6390855
CP	hasEntrezItem	7012354
CP	hasEntrezItem	7040095
CP	hasEntrezItem	8050338
CP	hasEntrezItem	8206865
IPI00017601.1	hasGene Symbol	CP

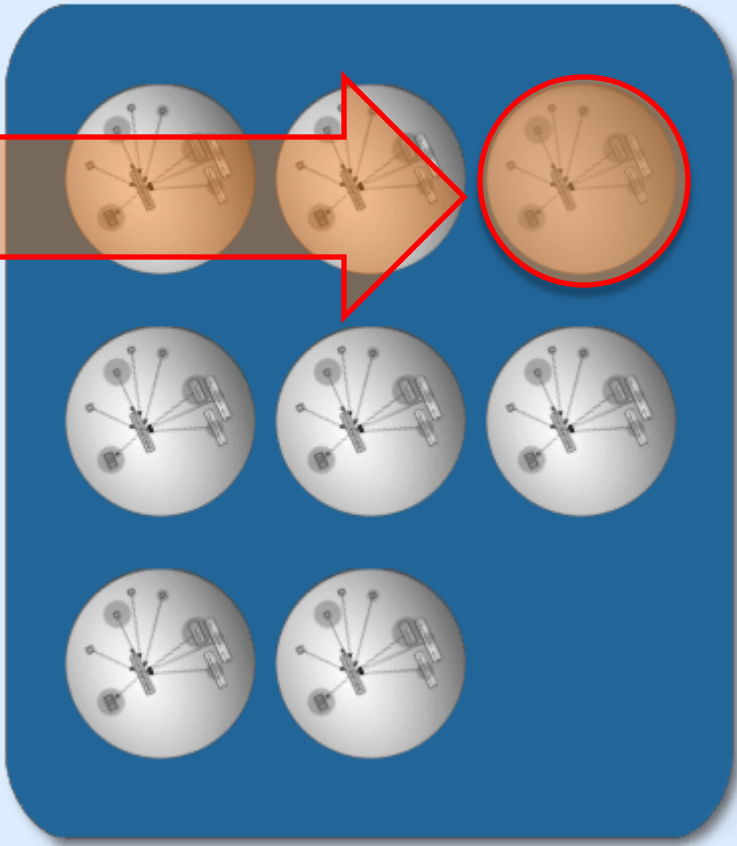
Relations

geneTotalRecords
hasEC Enzyme Classification
hasEntrezItem
hasGene Symbol



ASK Arrays

Choose Array: Organ failures Search



A Acute Rejection, m

B Acute Rejection, m

C Acute Rejection, h

D Acute Rejection, w

E Non Rejection, me

F Non Rejection, me

G Acute Rejection, lo

H Acute Rejection, u


IO Informatics Sentient Suite
Welcome Administrator: [Profile](#) [Logout](#)

[Browse](#) [Search](#) [Site Administration](#) [About](#)

[My Queries](#) [ASK Arrays](#) [Query Builder](#) [Custom SQL](#) [SPARQL](#) [Administration](#) [Migration](#)

ASK Arrays

Choose Array:



- A Acute Rejection, medium confidence [Gene BMs]
- B Acute Rejection, medium confidence [Protein BMs]
- C Acute Rejection, high confidence [Combinatorial BMs]
- D Acute Rejection, very tight range (ALL Combinatorial BMs)
- E Non Rejection, medium confidence [Protein BMs]
- F Non Rejection, medium confidence [5 Protein BMs, 1 Gene BM]
- G Acute Rejection, low confidence [Gene BMs]
- H Acute Rejection, uncertain [1 weak Gene BM]

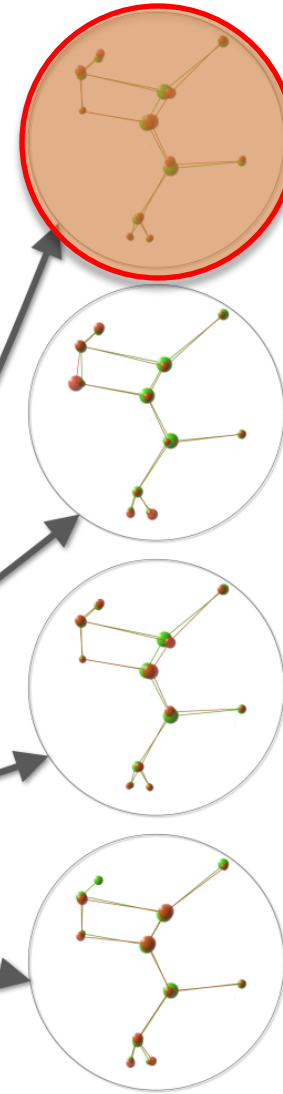
[Acute Rejection, medium confidence \[Gene BMs\]](#) [Open in Knowledge Explorer\(RQ\)](#)

Your search returned 4 record(s)
Export: [Excel](#) [Text \(TSV\)](#) [HTML](#) [XML](#) [Open in Knowledge Explorer \(FQML\)](#) [Chart](#)

Controls	Value	Value1	Value2	Value3	Value4	Score
AR4	4.792655711	4.517395827	4.287688659	4.516525668	5.629179083	0.7457
AR3	5.781567879	4.63077352	3.847728603	4.771930762	6.209432992	0.7882
AR2	4.922748721	4.743109483	3.96468382	4.02275853	5.829296888	0.8852
AR5	5.150558709	4.716695571	3.373487898	5.368614783	5.983364312	0.9975

[Show/Hide SQL](#)

Copyright 2006 - 2010, IO Informatics, Inc. (rev. 2.1.0.65)

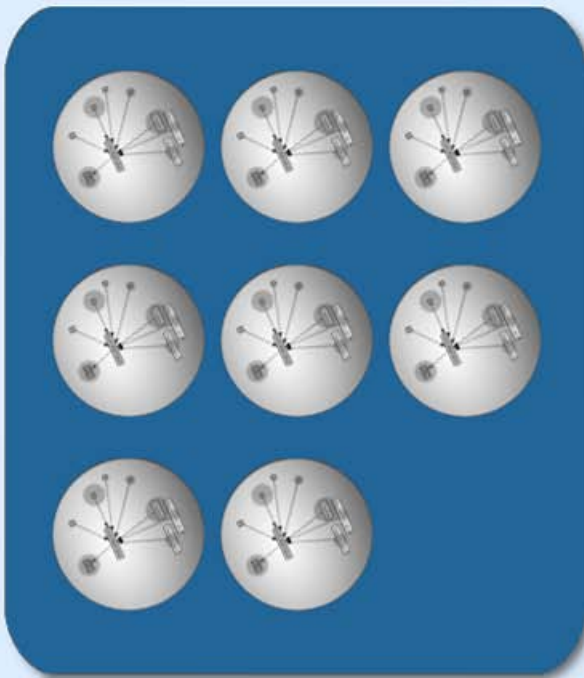


Select an
Array,
then
click
"Search"

ASK Arrays

Choose Array: Organ failures

Search

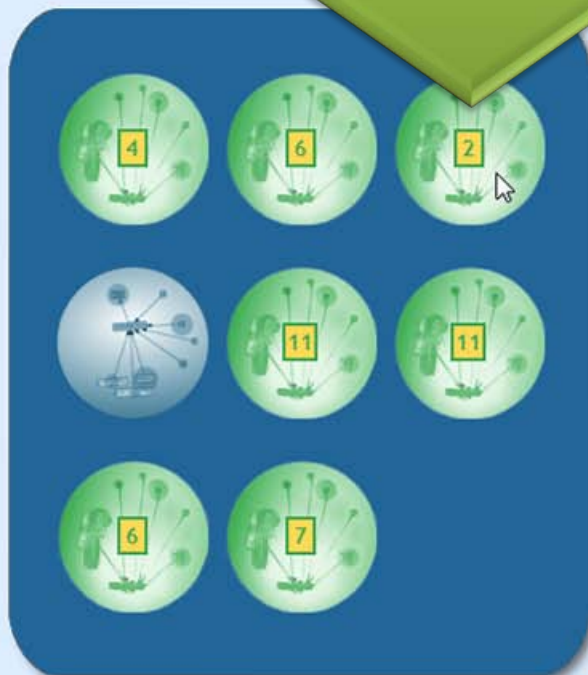


- A Acute Rejection, medium confidence [Gene BMs]
- B Acute Rejection, medium confidence [Protein BMs]
- C Acute Rejection, high confidence [Combinatorial BMs]
- D Acute Rejection, very tight range (ALL Combinatorial BMs)
- E Non Rejection, medium confidence [Protein BMs]
- F Non Rejection, medium confidence [5 Protein BMs, 1 Gene BM]
- G Acute Rejection, low confidence [Gene BMs]
- H Acute Rejection, uncertain [1 weak Gene BM]

ASK Arrays

Choose Array: Organ failure

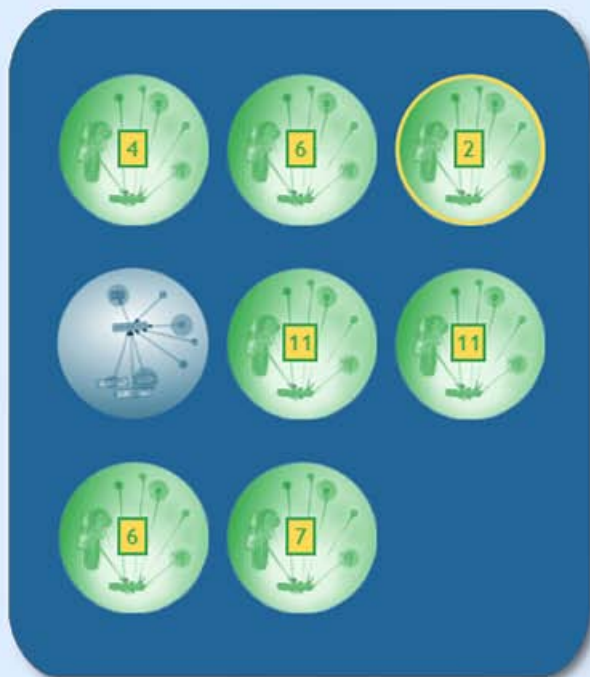
Click on a
profile
for
matching
patients



- A Acute Rejection, medium confidence [Gene BMs]
- B Acute Rejection, medium confidence [Protein BMs]
- C Acute Rejection, high confidence [Combinatorial BMs]
- D Acute Rejection, very tight range (ALL Combinatorial BMs)
- E Non Rejection, medium confidence [Protein BMs]
- F Non Rejection, medium confidence [5 Protein BMs, 1 Gene BM]
- G Acute Rejection, low confidence [Gene BMs]
- H Acute Rejection, uncertain [1 weak Gene BM]

ASK Arrays

Choose Array: Organ failures Search



- A Acute Rejection, medium confidence [Gene BMs]
- B Acute Rejection, medium confidence [Protein BMs]
- C Acute Rejection, high confidence [Combinatorial BMs]

- D Acute Rejection, very tight range [ALL Combinatorial BMs]
- E Non Rejection, medium confidence [Protein BMs]
- F Non Rejection, medium confidence [5 Protein BMs,

- G Acute Rejection, low confidence [Gene BMs]
- H Acute Rejection, uncertain [1 weak Gene BM]

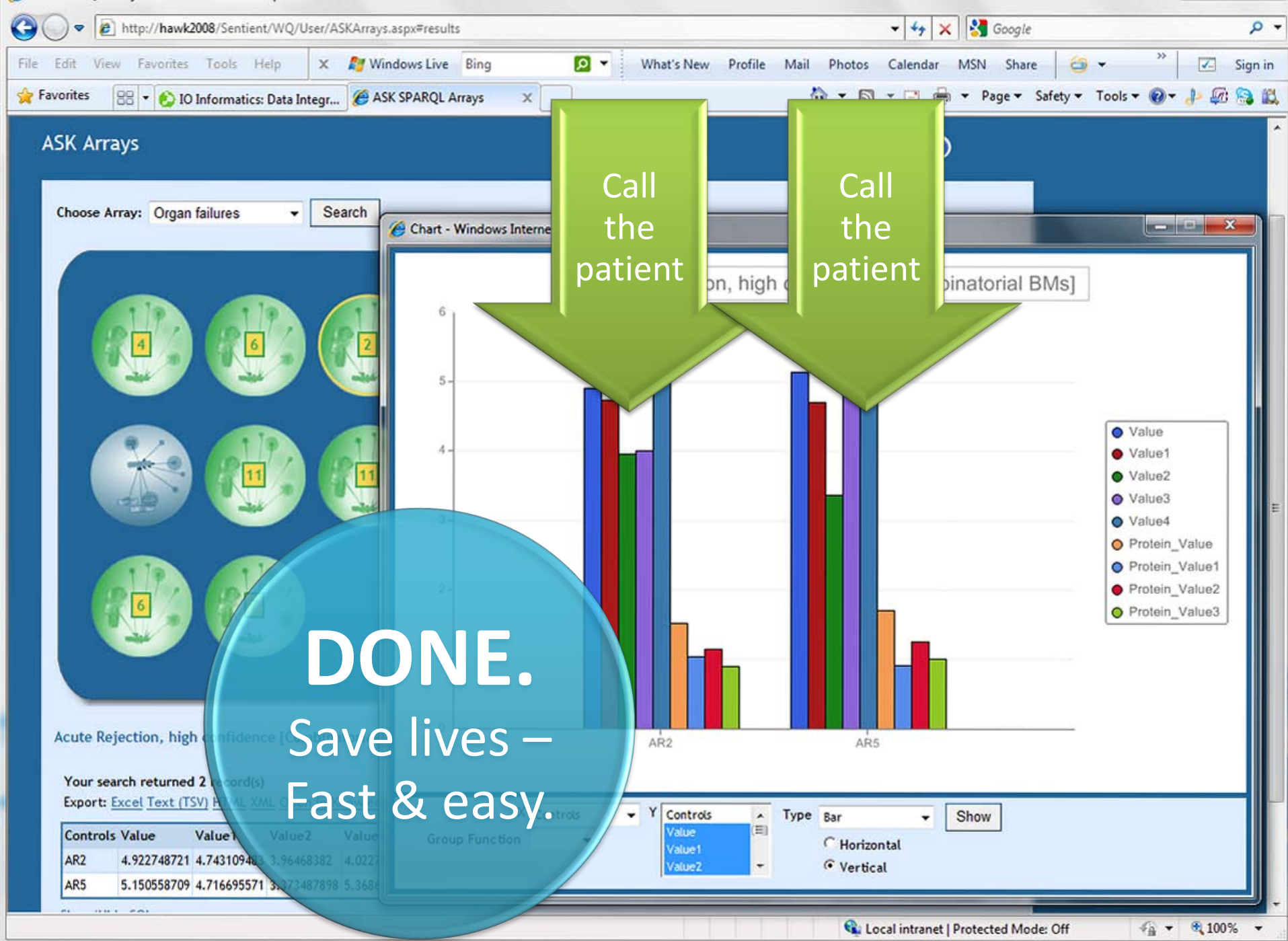
Review the
patient's
biomarkers
and scoring

Acute Rejection, high confidence [Combinatorial BMs] [Open in Knowledge Explorer\(RQ\)](#)

Your search returned 2 record(s)

Export: [Excel Text \(TSV\)](#) [HTML XML](#) [Open in Knowledge Explorer \(FQML\)](#) [Chart](#)

Controls	Value	Value1	Value2	Value3	Value4	Protein_Value	Protein_Value1	Protein_Value2	Protein_Value3	Score	Score2	Score3
AR2	4.922748721	4.743109483	3.96468382	4.02275853	5.829296888	1.534376	1.0509057	1.164798	0.9107218	0.0579	0.0127	0.027
AR5	5.150558709	4.716695571	3.373487898	5.368614783	5.983364312	1.725311	0.9254474	1.2732773	1.0197685	0.1586	0.1144	0.1091



Implications for PM

Actionable inferences about disease states and treatments using clinical, multi-OMIC, molecular phenotypic data and mechanistic insights from public knowledge networks combined, is a remarkable step towards patient-centric personalized medicine.

Applied Semantic Knowledgebases (ASK) provide researchers faced with complex biological questions relying on solid decision-support with a novel, directly applicable approach.

To use, share and apply knowledge based on sophisticated network models via an intuitive web tool - hiding complexity, yet providing concise information which data (disease states, patients) and how good it fits the model - *is changing the way how knowledge is built, refined and applied in life sciences and personalized medicine.*

Thank you!

Questions?

