Recommendations to model data from Pathway Interaction Database in caBIO

Version 1

Release 4.2

Last Updated:

Owner: caBIO Team

Email: klemmj@mail.nih.gov

National Cancer Institute Center for BioInformatics

6116 Executive Blvd.

Rockville, MD 20852



1 Overview of PID Data and Proposed Object Model Changes 4

1.1 Description 4

1.2 Use Cases 4

1.2.1 Gene-Pathway Association 4

1.2.2 Protein-PathwayInteraction Association 5

1.2.3 Gene-PostTranslationalModification Association 6

1.2.4 Evidence-Gene Association 8

1.3 Proposed Object Model Changes 9

1.3.1 Existing Domain Model 9

1.3.2 Extended Domain Model 10

1.3.3 Existing Data and Associations 12

1.3.3.1.1 Observations 12

1.4 PID Dataset 12

1.4.1 Data Statistics 13

1.4.2 Doubts / Imponderables / Not Modeled 16

1.4.3 From Pathway Data to Proposed Model 16

1.4.4 From Interactions Data to Proposed Model 17

1.4.5 From Molecule Data to Proposed Model 19

2 Data Exemplars 21

2.1.1 Pathway Data 21

2.1.2 Interaction Data 21

2.1.3 Molecules Data 22

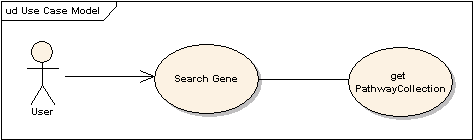
# Overview of PID Data and Proposed Object Model Changes

## Description

The Pathway Interaction Database is a highly-structured, curated collection of information about known biomolecular interactions and key cellular processes assembled into signaling pathways. It is a collaborative project between the US National Cancer Institute (NCI) and Nature Publishing Group (NPG), and is an open access online resource. It exposes data regarding pathways, complexcomponents and molecular interactions from NCI-Nature-curated dataset, Reactome and BioCarta.

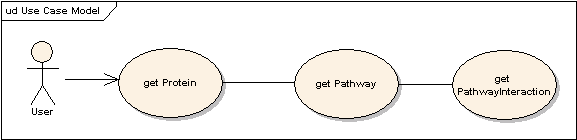
## Use Cases

### Gene-Pathway Association



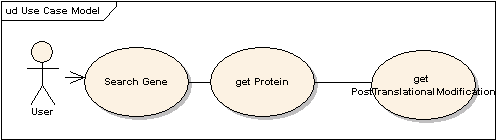
|  |  |
| --- | --- |
| **Use Case Name** | **Gene-Pathway Association** |
| **Use Case ID** | 2.1 |
| **Primary Actor** | Public |
| **Brief Description** | The *Gene-Pathway Association* use case allows users to search for pathways associated with a Gene. The results of high throughput data analysis, such microarray, are often lists of genes that exhibit certain behavior of interests to the biologist, who wants to know whether any of these genes are associated with known pthways in the literature. |
| **Workflow** | 1. Query for caBIO Gene object by commonly used IDs, such as HUGO symbol, or Genbank accession. 2. Get pathwayAssociations to retrieve associated pathways which in turn can be used to retrieve interactions, etc. |
| **Trigger** | The user searches for a Gene |
| **Pre-conditions** | 1. The user navigates to the caBIO Home Page and clicks “Continue” |
| **Flow of Events** | 1. The user selects the Gene object 2. The user provides a valid Gene Id such as a HUGO Symbol, etc 3. The user clicks the submit button to initiate the search 4. A list of Genes appear 5. The user can navigate to pathwayAssociations by clicking on the link, *getPathwayAssociation*. |
| **Post Conditions** | *Success Condition:* The user successfully searches the database and obtains a list of all pathways associated with a Gene  *Alternate Condition:*  AC2: No result is found  *Error Conditions:*  EC1: System Error during Request Processing  EC2: Specified Gene does not exist  EC3: Specified Association does not identify a valid pathway |
| **Alternative Flow** | Alternate Condition:  AC2: No result is found   1. “No results found” message is displayed |
| **Related Use Cases** | Protein-PathwayInteraction Associations, Gene-PostTranslationalModification Association, Evidence-Gene Association |
| **Issues/Notes** | 1. The user can also execute the same search using the Java API or Web Service 2. This use case is already supported in caBIO 4.1, but with the introduction of new datasets, additional data would be retrievable through this use-case |

### Protein-PathwayInteraction Association



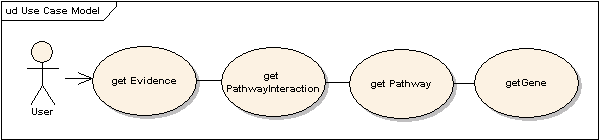
|  |  |
| --- | --- |
| **Use Case Name** | **Protein-PathwayInteraction Association** |
| **Use Case ID** | 2.3 |
| **Primary Actor** | Public |
| **Brief Description** | The use case allows a user to retrieve Pathway Interactions associated with a given protein |
| **Workflow** | 1. Search for a protein by its Uniprot Id. 2. Get PathwayCollection to retrieve associated pathways. 3. Click on interactionCollection to retrieve associated interactions. 4. Filter by interaction type. |
| **Trigger** | The user desires to search for Pathway Interactions |
| **Pre-conditions** | The user navigates to the caBIO Home Page and clicks “Continue” |
| **Flow of Events** | 1. Click on Protein 2. Search for a Protein by its Uniprot Id 3. Get Pathways by clicking on *getPathwayCollection*. 4. Click on *getInteractionCollection* to get associated interactions |
| **Post Conditions** | *Success Condition:* The user successfully searches the database and obtains a list of all interactions associated with the pathways in which a given protein participates  *Alternate Condition:*  AC2: No result is found  *Error Conditions:*  EC1: System Error during Request Processing  EC2: Specified Protein does not exist  EC3: Specified Protein is not associated with any Pathway |
| **Alternative Flow** | Alternate Condition:  AC2: No result is found  1. “No results found” message is displayed |
| **Related Use Cases** | Gene-Pathway Associations, Gene-PostTranslationalModification Association, Evidence-Gene Association |
| **Issues/Notes** | The user can also execute the same search using the Java API or Web Service |

### Gene-PostTranslationalModification Association



|  |  |
| --- | --- |
| **Use Case Name** | **Browse Gene-PostTranslationalModification Association** |
| **Use Case ID** | 2.4 |
| **Primary Actor** | Public |
| **Brief Description** | Given a gene, find all post translational modifications associated with the expressed proteins on account of participation in a pathway. |
| **Workflow** | 1. Start from a gene. 2. Get corresponding proteins. 3. Get the post translational modification information for the protein, if available. |
| **Trigger** | The user desires to search for publicly available genes |
| **Pre-conditions** | 1. The user navigates to the caBIO Home Page and clicks “Continue” |
| **Flow of Events** | 1. Click on Gene 2. Search for a Gene by its HUGO Symbol, Unigene Cluster Id, etc. 3. Get expressed proteins by clicking on *getProtein*. 4. Click on *getPostTranslationalModification* to get associated post translational modification information |
| **Post Conditions** | *Success Condition:* The user successfully searches the database and obtains a list of all available post translational modification information for the expressed gene.  *Alternate Condition:*  AC2: No result is found  *Error Conditions:*  EC1: System Error during Request Processing  EC2: Specified Protein does not exist  EC3: Specified Protein has not undergone any post translational modification or has not participated in any pathways stored in caBIO |
| **Alternative Flow** | Alternate Condition:  AC2: No result is found  1. “No results found” message is displayed |
| **Related Use Cases** | Gene-Pathway Association, Protein-PathwayInteraction Association, Evidence-Gene Association |
| **Issues/Notes** | The user can also execute the same search using the Java API or Web Service |

### Evidence-Gene Association



|  |  |
| --- | --- |
| **Use Case Name** | **Browse Evidence-Pathway Association** |
| **Use Case ID** | 2.4 |
| **Primary Actor** | Public |
| **Brief Description** | Given a publication which documents a particular pathway interaction, find all genes associated with the same. |
| **Workflow** | 1. Start from a publication/Evidence retrievable using its PubMed Id. 2. Get corresponding Pathway Interactions. 3. Get associated Pathways 4. Get associated Genes |
| **Trigger** | The user desires to search for publication information |
| **Pre-conditions** | 1. The user navigates to the caBIO Home Page and clicks “Continue” |
| **Flow of Events** | 1. Click on Evidence 2. Search for a Evidence using its PubMed Id 3. Get associated PathwayInteractions. 4. Click on *getPathwayCollection* to get associated pathways 5. Click on *getGene* to get associated Genes |
| **Post Conditions** | *Success Condition:* The user successfully searches the database and obtains a list of all genes associated with a given publication on account of the pathways and the interactions it is involved in  *Alternate Condition:*  AC2: No result is found  *Error Conditions:*  EC1: System Error during Request Processing  EC2: Specified Publication does not exist  EC3: Specified Publication is not associated with any PathwayInteraction |
| **Alternative Flow** | Alternate Condition:  AC2: No result is found  1. “No results found” message is displayed |
| **Related Use Cases** | Gene-Pathway Association, Protein-PathwayInteraction Association, Gene-PostTranslationalModification |
| **Issues/Notes** | The user can also execute the same search using the Java API or Web Service |

## Proposed Object Model Changes

### Existing Domain Model

The follow objects from the current caBIO domain model are relevant to the modeling this data in caBIO.

Figure 1 describes the current caBIO model that associates these objects.

|  |  |
| --- | --- |
| *caBIO Domain Object* |  |
| Pathway | An object representation of a molecular/cellular pathway in human and mouse compiled by BioCarta. It provides access to further details about the associated Genes and any available histopathological information for genes involved in clinical trials through getGeneCollection and getHistopathologyCollection methods.  It gets populated with data from CGAP and is composed of BioCarta Pathways. It provides links to genes associated with every pathway as well as histopathological information pertaining to these Genes through the associated Histopathology object. |
| NucleicAcidSequence | An object representation of a gene sequence; provides access to the clones from which it was derived, the ASCII representation of the residues it contains, and the sequence ID and a link to the Location Objects that in turn provide further details of any SNP, Cytoband, EST or MRNAs associated with that sequence.  It is populated with data pertaining to human and mouse from Unigene. |
| Gene | Gene objects are the effective portal to most of the genomic information provided by the caBIO data services; It provides access to the associated proteins, gene-targets, diseases, histopathologies, ontologies, pathways, aliases, clones and sequence data through their respective objects such as Histopathology, DiseaseOntology, OrganOntology, Pathway, GeneAlias, Clone and NucleicAcidSequence respectively.  The location of the gene on the chromosome is available through the Location object.  This object gets populated with data pertaining to human and mouse genes from NCBI’s Unigene’s repository. Gene Ids from NCBI’s OMIM Databank, Ensembl, and Enzyme Commission are available through DatabaseCrossReference. |
| Protein | An object representing the various names (scientific, common, abbreviated, etc.) for human and mouse genomes. It can be used to retrieve the associated Genes, Chromosomes, Pathways if any, Proteins, or Tissues. |
| Evidence | An object representing the publication from PubMed that serves as an Evidence supporting the association between a caBIO Gene and a DiseaseOntology or an Agent. Based on the publication, an Evidence can be associated with multiple EvidenceCodes. |
| EvidenceCode | A class that models the different types of EvidenceCodes associated with an Evidence, based on the data in the publication. |

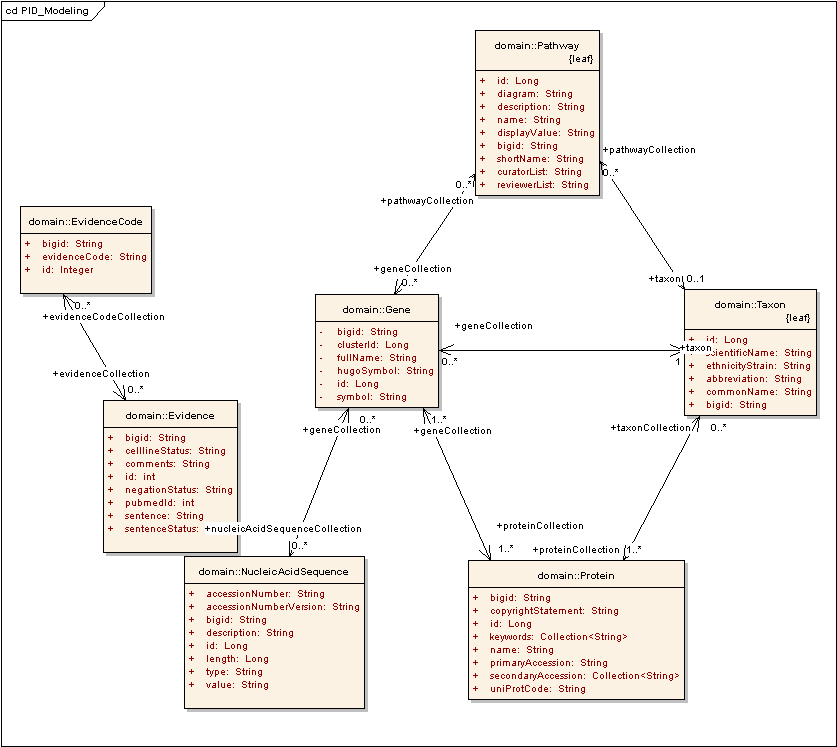


Figure Current caBIO Model

### Extended Domain Model

The extended model adds supports for new objects to model the pathwayInteractions, the different components in an interaction and its association to an Evidence or EvidenceCode.

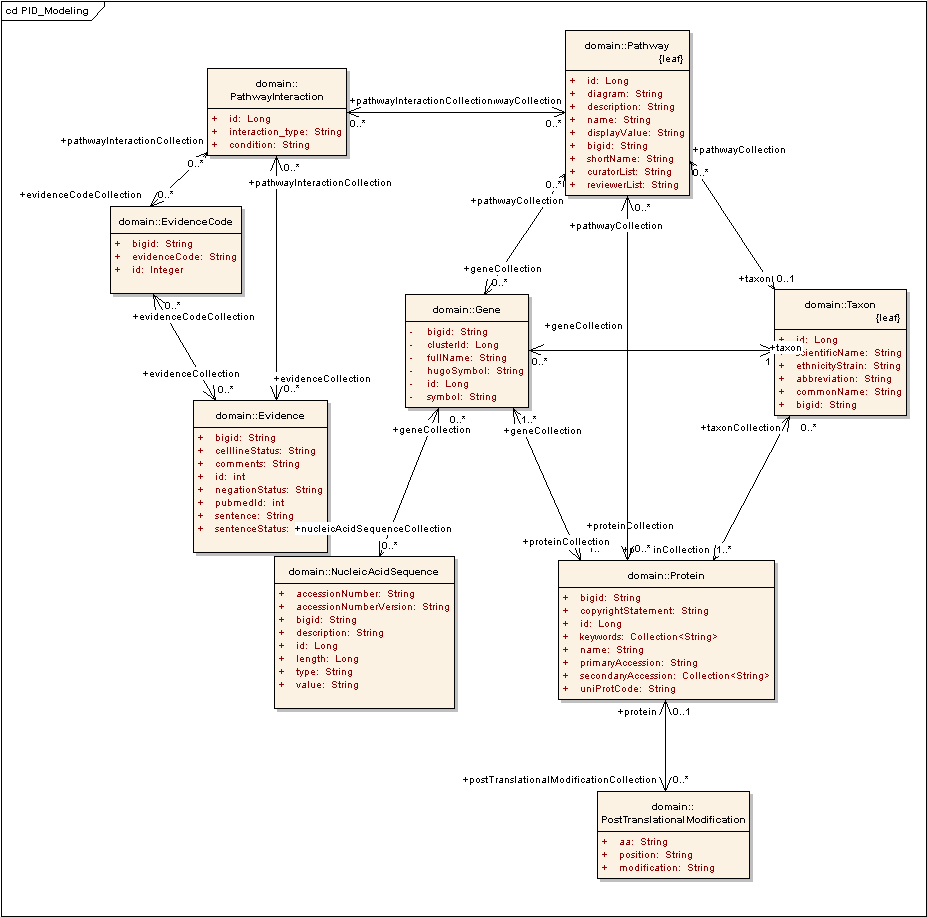


Figure Proposed Model for Pathway Data Modeling

Note: New classes or attributes are highlighted in red

**Pathway**

New attributes would be added to model the curator, reviewer and source information

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Diagram | Pathway Diagram in SVG Format |
| Description | Sentence from PubMed publication |
| Name | Short name for the pathway |
| LongName | Long name for the pathway |
| DisplayValue | Display value for the diagram |
| CuratorList | List of curators |
| ReviewerList | List of reviewers |
| Source | Pathway Source (Reactome, BioCarta, NCI-Nature) |

**PathwayInteraction**

New class to model the interactions that constitute a Pathway

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Role | Role of an interaction in a Pathway |

**PostTranslationalModification**

New class to model the Post Translational Modification information for a protein where available

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Aa | Amino acid |
| Position | Position at which modification has occurred |
| Modification | Type of modification |

### Existing Data and Associations

Currently caBIO procures BioCarta data from CMAP which gives a Locus Link Id that allows us to associate a caBIO Gene with a Pathway.

This would need to be retained for backward compatibility since caMOD and Rembrandt amongst others use this data.

##### Observations

1. Retain existing data and associations (316 pathways currently in caBIO)
2. Set PathwaySource as ‘BioCarta’ for existing data (254 pathways from PID datasets)
   1. 244 of these 254 pathways are already present in caBIO
3. Add data to model associations to Proteins and Interactions for the human pathways from these 244 records
4. Add remaining 10 pathways
5. Outcome:
   * 1. 244 Pathways would have an association to Gene AND Protein and Interaction
     2. Mouse Pathways in caBIO's Pathway will have no association to PathwayInteractions; Only the Human Pathways will.
     3. 10 newly added Pathways will have no association to Gene, but will have an association to PathwayInteractions.
     4. Resulting total number of BioCarta pathways : 316 + 10 = 317

## PID Dataset

Based on the above model, the tables below summarize the data available from the PID dataset and their proposed association to caBIO.

According to the DTD, a *Pathway* is composed of one or more *PathwayInteractions*, and has associated *Evidences*.

It is composed of multiple *PathwayInteractions* which play one of 227 roles in the *Pathway*.

Every *PathwayInteraction* also has associated *Evidences* or *EvidenceCodes*.

A *PathwayInteractionCondition* influences a *PathwayInteraction*

Multiple *PathwayInteractionComponents* (Molecules) participate in a *PathwayInteraction* with each Component being defined by its role in the Interaction. Every *Molecule* can function as either as an input, output, agent or inhibitor in a *PathwayInteraction*. There are some records with “process-type” as the role which would be pruned as junk data

However there may be many different aspects to each such role which is modeled by *PathwayInteractionComponentLabel.*

These molecules are identified by aliases or Uniprot accession numbers or descriptions in some cases.

Some Proteins have subunit information available and it may be these subunits that may participate in a Pathway.

It may also be “collections” of Complex/Compounds/Proteins that participate in a PathwayInteraction.

### Data Statistics

|  |  |  |
| --- | --- | --- |
| **Pathways** | **Description** | **Notes** |
| ZSTG\_PATHWAYS | Pathway Data (consists of the source, long/short name, organism)  PID uses “7” to refer to Reactome and “5” to NCI-Nature  Only Human Data available. Hence organism = Hs for all records | Reactome: 765  NCI-Nature: 111  Source Id No. 3: 154 *(presumably “3” also refers to BioCarta)???*  BioCarta: 101 |
| ZSTG\_PATHWAYCURATORS / ZSTG\_PATHWAYREVIEWERS | A Pathway can be curated/ reviewed by multiple people | **Pathway Curators**  Reactome: 656  NCI-Nature: 106  BioCarta: 0  **Pathway Reviewers**  Reactome: 7  NCI-Nature: 224  BioCarta: 0 |
| ZSTG\_PATHWAYCOMPONENTS | Gives the Interaction Ids associated with a Pathway | Reactome: 3219  NCI-Nature: 4095  BioCarta: 1557  Source Id No.3: 2227 |
|  |  | *DTD implies a 1-to-many association between Pathway and Reference/Evidence, but there is no data to support this.*  *All References are associated only with an Interaction and not with a Pathway.* |
|  |  |  |
| **Interactions** |  |  |
| ZSTG\_INTERACTIONS | Lists the interactions in each Pathway | Reactome: 3020  NCI-Nature: 3941  BioCarta: 1759  2: 1244 |
| ZSTG\_INTERACTIONEVIDENCE | Lists the multiple evidence-Codes associated with each Interaction | Reactome: 3020  NCI-Nature: 5468  “3”: 1759  BioCarta: 1244 |
| ZSTG\_INTERACTIONREFERENCE | Lists the multiple publications associated with each Interaction | Reactome: 3922  NCI-Nature: 4956  BioCarta: 0  3: 0 |
| ZSTG\_INTERACTIONCONDITION | Lists the multiple conditions associated with each Pathway Interaction | Reactome: No data  NCI Nature: 250  3: 50  BioCarta: 70  *DTD implies a 1-to-many association, but PID data shows a 1-to-1 association between PathwayInteraction and PathwayInteractionCondition.*  *On the basis of this “condition” could be modeled simply as an attribute of Interaction* |
|  |  |  |
| ZSTG\_INTERACTIONCOMPONENTLABEL | Lists the different molecules participating in each Interaction along with their roles (agent, inhibitor, input, output) and the 5 different lable-types(i.e. activity-state, location, function, ptm) | What is the difference between Role-Type and Label-Type? All InteractionComponents have only one role, but many have multiple Component Labels.  NCI-Nature:agent:1593  Reactome:agent:3048  3:agent:967  BioCarta:agent: 662  NCI-Nature:inhibitor:2  Reactome:inhibitor:346  3:inhibitor:128  BioCarta:inhibitor:14  NCI-Nature:input:6253  Reactome:input:3361  3:input:1116  BioCarta:input:484  NCI-Nature:output:5540  Reactome:output:3795  2:output:1366  BioCarta:output: 934  **Label-types for the different InteractionComponents** (activity-state, location, function, ptm)  BioCarta:activity-state:1266  BioCarta:location:828  3:activity-state:1308  3:location:2269  NCINature:activity-state:4972  NCINature:function:43  NCINature:location:5535  Reactome:activity-state:145  Reactome:location:13237  Reactome:process-type:5  Reactome:ptm:1 |
| ZSTG\_INTERACTIONCOMP-PTM-TERMS | An Interaction involves multiple components/molecules, each identified by a moleculeId (which is used to provide further information on the molecule) and each playing a different role in the Interaction. Some of these have associated PostTranslationalModification (PTM) data  Lists the PTM data for these molecules, if available | Reactome:1183  NCINature:2862  BioCarta: 0  Source No “3”: 0 |
|  |  |  |
| **Molecules** |  |  |
| ZSTG\_MOLECULENAMES | Lists the type of molecule and the value (an identifier/description). There are only 5 types of molecules in the data: complex, compound, protein, and rna.  Records of type “molecule-type” would be purged  NCINature:complex:1870  NCINature:compound:201  NCINature:protein:4121  NCINature:rna:7  Reactome:complex:3332  Reactome:compound:2613  Reactome:molecule-type:754  Reactome:protein: 7624  Reactome:rna:67  (null):complex:884  (null):compound:644  (null):protein:7268  (null):rna:14 | There is insufficient data to create a separate Complex or Compound object. The “value” column of this table is composed of comments/aliases/descriptions/symbols and is the only new data/attribute that could added to this new object.  Almost all the entries for “rna” molecules are comments and some “protein” molecules have accession numbers.  *Approximately only 3000-4000 of the 33,000+ molecules is associated with caBIO Identifiers.* |
| ZSTG\_MOLECULE-COMP-PTMTERMS | Only some “complex” Molecules have PTM data associated with them. (6300) | Reactome:1691  NCINature:4609 |
| ZSTG\_MOLECULEPARTS | Subunits of “protein” molecules | NCINature: 194  Reactome: 549  Null Source Id: 68 (from BioCarta file..hence possibly BioCarta??) |
| ZSTG\_MOLECULEFAMILIES | Molecules of type “protein”, “complex” and “compound” have family-data associated with them | NCI Nature: 738  Null: 1408 |

### Doubts / Imponderables / Not Modeled

1. Complex
2. Compound
3. Would be classes with one attribute
4. Little to no associations to existing caBIO Protein(?) objects based on (10)
5. PathwayInteractionComponent’s role (Role of a molecule participating in a PathwayInteraction)
6. PathwayInteractionComponentLabel (Additional information on the role of the molecule in a pathway interaction)
7. ProteinSubUnit Information
8. Family information where available (Complexes being collections of other complexes, etc)
9. No data to warrant association between Pathway and Evidence.
10. *Insufficient information to map caBIO Ids on most of the participating molecules in the Pathway Data*

### From Pathway Data to Proposed Model

|  |  |  |
| --- | --- | --- |
| **Data** | **Model** | **Notes** |
| <Pathway id="500317" subnet="true"> | **Pathway** |  |
| <Organism>Hs</Organism> | Used to make association to Taxon (Humans) |  |
| <LongName>Signalling to RAS</LongName> | **Pathway**.name |  |
| <ShortName>Signalling\_to\_RAS</ShortName> | **Pathway**.shortName |  |
| <Source id="7">Reactome</Source> | **Pathway**.source |  |
| <CuratorList> |  |  |
| <Curator>Nasi, S, Annibali, D</Curator> | **Pathway**.curatorList | Concatenate all curators into a single record |
| <Curator>Jassal, B</Curator> |  |  |
| </CuratorList> |  |  |
| <ReviewerList> |  |  |
| <Reviewer>Greene, LA</Reviewer> | **Pathway**.reviewerList | Concatenate all reviewers into a single record |
| <Reviewer>Dr. Dan Dumont</Reviewer> |  |  |
| <Reviewer>Dr. Andreas Papapetropoulos</Reviewer> |  |  |
| </ReviewerList> |  | Neither Reactome nor NCI Nature datasets have any data to support association between **Pathway and Evidence**; Data exists to support only the association between Interaction and Evidence. |
| <PathwayComponentList> |  |  |
| <PathwayComponent interaction\_idref="500704"/> | To make appropriate association to **PathwayInteractions** |  |
| <PathwayComponent interaction\_idref="501216"/> |  |  |
| <PathwayComponent interaction\_idref="503170"/> |  |  |
| <PathwayComponent interaction\_idref="500703"/> |  |  |
| <PathwayComponent interaction\_idref="501218"/> |  |  |
| <PathwayComponent interaction\_idref="501217"/> |  |  |
| <PathwayComponent interaction\_idref="503168"/> |  |  |
| <PathwayComponent interaction\_idref="503522"/> |  |  |
| </PathwayComponentList> |  |  |
| </Pathway> |  |  |

### From Interactions Data to Proposed Model

|  |  |  |
| --- | --- | --- |
| **Data** | **Model** | **Notes** |
| <Interaction interaction\_type="modification" id="500460"> | PathwayInteraction.interaction\_type  Id: used to make an association between **Pathway** and **PathwayInteraction** | 227 different interaction types.  Unclear about  **Pathway** is associated with multiple **PathwayInteractions** |
| <Condition condition\_type="retrograde axon cargo transport">retrograde a  xon cargo transport</Condition> | **PathwayInteraction.**condition\_type | DTD shows 1-to-many association between Interaction and Condition (hypothetical example shown alongside).  However data shows only one condition for every interaction.  Hence condition\_type could be attribute of **PathwayInteraction** |
| <Condition condition\_type="B-Lymphocyte">B-Lymphocyte</Condition> |  |  |
| <Condition condition\_type="activation of NF-kappaB transcription factor">activation of NF-kappaB transcription factor</Condition> |  |  |
| <Abstraction pathway\_idref="500682" pathway\_name="Orc1 removal from chromatin" external\_pathway\_id="Orc1\_removal\_from\_chromatin"/> |  | This information is already captured in **Pathway** |
| <Source id="7">Reactome</Source> | Used to make associations to **Pathways** from appropriate source |  |
| <EvidenceList> |  |  |
| <Evidence value="NIL">NIL</Evidence> | **EvidenceCode**.evidenceCode |  |
| </EvidenceList> |  |  |
| <ReferenceList> |  |  |
| <Reference pmid="16612002">16612002</Reference> | **Evidence**.pubmedId |  |
| <Reference pmid="9687533">9687533</Reference> |  |  |
| <Reference pmid="12660731">12660731</Reference> |  |  |
| </ReferenceList> |  |  |
| <InteractionComponentList> | Not modeled | **PathwayInteraction** is composed of many PathwayInteractionComponents |
| <InteractionComponent role\_type="input" molecule\_idref="501651"> | Not modeled | 4 different roles for a **PathwayInteractionComponent**: 4 (input, output, inhibitor, agent) |
| <Label label\_type="location" value="cytoplasm"/> | Not modeled | Data from NCI Nature shows 1-to-many association between **PathwayInteractionComponent** and **PathwayInteractionComponentLabel**  5 different label-types: Activity-state, function, ptm, location, process-type  96 different Label-values |
| <Label label\_type="activity-state" value="active"/> |  |  |
| </InteractionComponent> |  |  |
| <InteractionComponent role\_type="input" molecule\_idref="505232"> |  |  |
| <Label label\_type="location" value="plasma membrane"/> | Not Modeled |  |
| <PTMExpression> | **PostTranslationalModification** | Would have many-to-one association with Protein |
| <PTMTerm protein="Q6PIZ9" position="63" aa="Y" modification="phosphorylation"/> | Protein accession number would be used to make association with caBIO Protein  **PostTranslationalModification**.position  **PostTranslationalModification**.aa  **PostTranslationalModification**.modification |  |
| <PTMTerm protein="Q6PIZ9" position="79" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="Q6PIZ9" position="110" aa="Y" modification="phosphorylation"/> |  |  |
| </PTMExpression> |  |  |
| </InteractionComponent> |  |  |
| <InteractionComponent role\_type="output" molecule\_idref="501654"> |  |  |
| <Label label\_type="location" value="plasma membrane"/> |  |  |
| </InteractionComponent> |  |  |
| </InteractionComponentList> |  |  |
| </Interaction> |  |  |

### From Molecule Data to Proposed Model

|  |  |  |
| --- | --- | --- |
| **Data** | **Model** | **Notes** |
| <Molecule molecule\_type="protein" id="200158"> | Not Modeled | **PathwayInteractionComponent** is one of multiple Molecule-types. A Molecule can be a protein, complex, compound, rna or “molecule-type”, the last possibly being (junk data) |
| <Name name\_type="PF" long\_name\_type="preferred symbol" value="FRS2 family-active"/> | Not Modeled |  |
| <FamilyMemberList> |  |  |
| <Member member\_molecule\_idref="201289"> |  |  |
| <Label label\_type="activity-state" value="active"/> |  |  |
| <PTMExpression> | **PostTranslationalModification** |  |
| <PTMTerm protein="Q8WU20" position="2" aa="G" modification="myristoylation"/> |  |  |
| <PTMTerm protein="Q8WU20" position="196" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="Q8WU20" position="306" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="Q8WU20" position="349" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="Q8WU20" position="392" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="Q8WU20" position="436" aa="Y" modification="phosphorylation"/> |  |  |
| </PTMExpression> |  |  |
| </Member> |  |  |
| <Member member\_molecule\_idref="201290"> |  |  |
| <Label label\_type="activity-state" value="active"/> |  |  |
| <PTMExpression> |  |  |
| <PTMTerm protein="O43559" position="2" aa="G" modification="myristoylation"/> |  |  |
| <PTMTerm protein="O43559" position="417" aa="Y" modification="phosphorylation"/> |  |  |
| <PTMTerm protein="O43559" position="455" aa="Y" modification="phosphorylation"/> |  |  |
| </PTMExpression> |  |  |
| </Member> |  |  |
| </FamilyMemberList> |  |  |
| <ComplexComponentList> |  |  |
| <ComplexComponent molecule\_idref="200523"> | Not Modeled |  |
| </ComplexComponent> |  |  |
| <ComplexComponent molecule\_idref="201460"> |  |  |
| </ComplexComponent> |  |  |
| </ComplexComponentList> |  |  |
| <Part whole\_molecule\_idref="200268" part\_molecule\_idref="201327" start="217" end="1013"/> | Not Modeled |  |
| </Molecule> |  |  |

# Data Exemplars

### Pathway Data

<Pathway id="500317" subnet="true">

<Organism>Hs</Organism>

<LongName>Signalling to RAS</LongName>

<ShortName>Signalling\_to\_RAS</ShortName>

<Source id="7">Reactome</Source>

<CuratorList>

<Curator>Nasi, S, Annibali, D</Curator>

<Curator>Jassal, B</Curator>

</CuratorList>

<ReviewerList>

<Reviewer>Greene, LA</Reviewer>

</ReviewerList>

<PathwayComponentList>

<PathwayComponent interaction\_idref="500704"/>

<PathwayComponent interaction\_idref="501216"/>

<PathwayComponent interaction\_idref="503170"/>

<PathwayComponent interaction\_idref="500703"/>

<PathwayComponent interaction\_idref="501218"/>

<PathwayComponent interaction\_idref="501217"/>

<PathwayComponent interaction\_idref="503168"/>

<PathwayComponent interaction\_idref="503522"/>

</PathwayComponentList>

</Pathway>

### Interaction Data

<Interaction interaction\_type="modification" id="501031">

<Source id="7">Reactome</Source>

<EvidenceList>

<Evidence value="NIL">NIL</Evidence>

</EvidenceList>

<ReferenceList>

<Reference pmid="14657031">14657031</Reference>

<Reference pmid="11859075">11859075</Reference>

</ReferenceList>

<InteractionComponentList>

<InteractionComponent role\_type="agent" molecule\_idref="502748">

<Label label\_type="location" value="nucleus"/>

</InteractionComponent>

<InteractionComponent role\_type="input" molecule\_idref="501810">

<Label label\_type="location" value="nucleus"/>

</InteractionComponent>

<InteractionComponent role\_type="output" molecule\_idref="501810">

<Label label\_type="location" value="nucleus"/>

</InteractionComponent>

</InteractionComponentList>

</Interaction>

### Molecules Data

### 

<Molecule molecule\_type="compound" id="503535">

<Name name\_type="CH" long\_name\_type="Chemical Entities of Biological Int

erest" value="15554"/>

<Name name\_type="AS" long\_name\_type="alias" value="Prostaglandin H2"/>

<Name name\_type="AS" long\_name\_type="alias" value="(5Z,13E)-(15S)-9alpha

,11alpha-Epidioxy-15-hydroxyprosta-5,13-dienoate"/>

<Name name\_type="PF" long\_name\_type="preferred symbol" value="Prostaglan

din H2"/>

</Molecule>

#### 

|  |
| --- |
| END OF DOCUMENT |