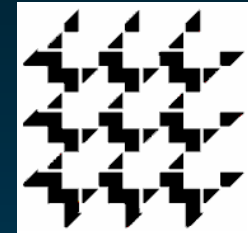


Institute for Discrete Sciences  
Workshop on Associating Semantics with Graphs



Rutgers University  
April 16, 2007



# Unified Medical Language System

*The graph behind the forest*



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA

# Biomedical trees



<http://www.tolweb.org/tree/>



**Lineage** (full): [root](#); [cellular organisms](#); [Eukaryota](#); [Fungi/Metazoa group](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Euarchontoglires](#); [Glires](#); [Rodentia](#); [Sciurognathi](#); [Muroidea](#); [Muridae](#); [Murinae](#); [Mus](#)

- **[Mus musculus](#)** (house mouse) *Click on organism name to get more information.*
  - **[Mus musculus bactrianus](#)** (southwestern Asian house mouse)
  - **[Mus musculus castaneus](#)** (southeastern Asian house mouse)
  - **[Mus musculus domesticus](#)** (western European house mouse)
  - **[Mus musculus gentilulus](#)**
  - **[Mus musculus homourus](#)**
  - **[Mus musculus molossinus](#)** (Japanese wild mouse)
  - **[Mus musculus musculus](#)** (eastern European house mouse)
  - **[Mus musculus praetextus](#)**
  - **[Mus musculus wagneri](#)**

# Medical Subject Headings



[Amino Acids, Peptides, and Proteins \[D12\]](#)

[Proteins \[D12.776\]](#)

[Cytoskeletal Proteins \[D12.776.220\]](#)

[Amino Acids, Peptides, and Proteins \[D12\]](#)

[Proteins \[D12.776\]](#)

[Contractile Proteins \[D12.776.210\]](#)

[Muscle Proteins \[D12.776.210.500\]](#)

[Amino Acids, Peptides, and Proteins \[D12\]](#)

[Proteins \[D12.776\]](#)

[Membrane Proteins \[D12.776.543\]](#)

[Ankyrins \[D12.776.543.080\]](#)

[Arrestins \[D12.776.543.090\] +](#)

[Bacterial Outer Membrane Proteins \[D12.776.543.100\] +](#)

[Calnexin \[D12.776.543.162\]](#)

[Connexins \[D12.776.543.225\] +](#)

▶ [Dystrophin \[D12.776.543.250\]](#)

[Dystrophin-Associated Proteins \[D12.776.543.268\] +](#)

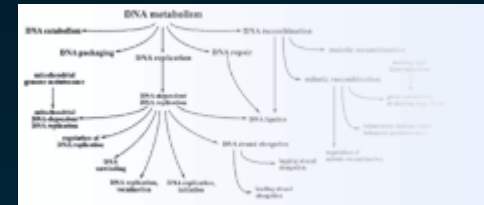
[Ephrins \[D12.776.543.287\] +](#)

[Heterotrimeric GTP-Binding Proteins \[D12.776.543.325\] +](#)

<http://www.nlm.nih.gov/mesh/2007/MBrowser.html>



# Gene Ontology



+ all : all [189459]

+ ⓘ GO:0008150 : biological\_process [137743]

+ ⓘ GO:0065007 : biological regulation [19935]

+ ⓘ GO:0050789 : regulation of biological process [18154]

+ ⓘ GO:0019222 : regulation of metabolic process [9816]

+ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**

+ ⓘ GO:0008152 : metabolic process [54684]

+ ⓘ GO:0044238 : primary metabolic process [44818]

+ ⓘ GO:0006629 : lipid metabolic process [3940]

+ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**

+ ⓘ **GO:0019222 : regulation of metabolic process [9816]**

+ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**

<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>

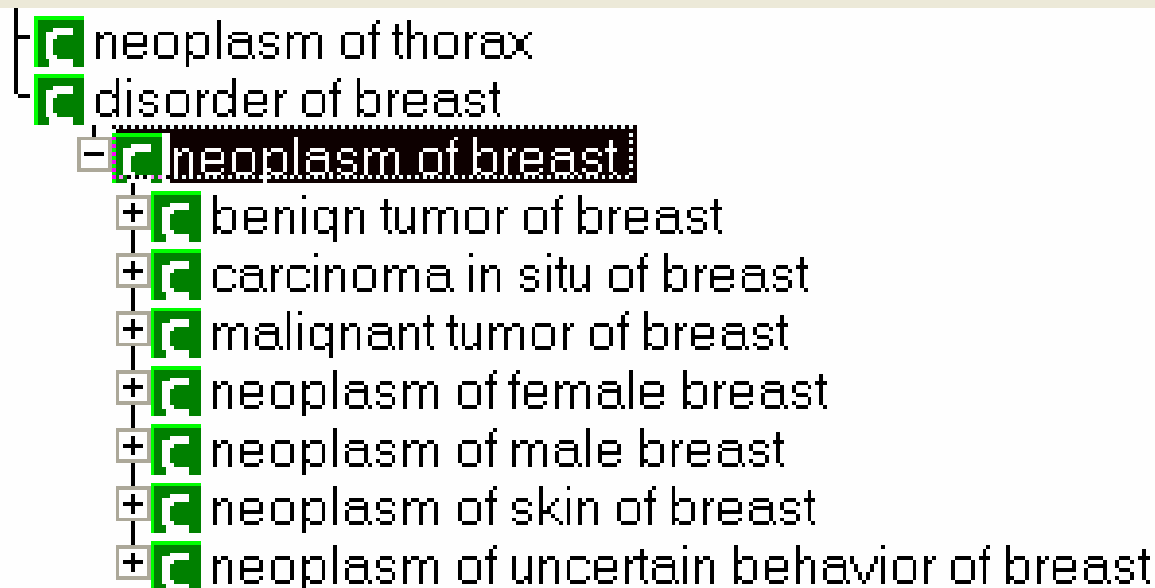


# SNOMED Clinical Terms



Hierarchy for 'neoplasm of breast'

Subtype hierarchy



<http://www.clininfo.co.uk/clue5/clue.htm>



# Biomedical trees revisited



# Medical Subject Headings



[Amino Acids, Peptides, and Proteins \[D12\]](#)

[Proteins \[D12.776\]](#)

[Cytoskeletal Proteins \[D12.776.543.325\]](#)

[Amino Acids, Peptides](#)

[Proteins \[D12.776\]](#)

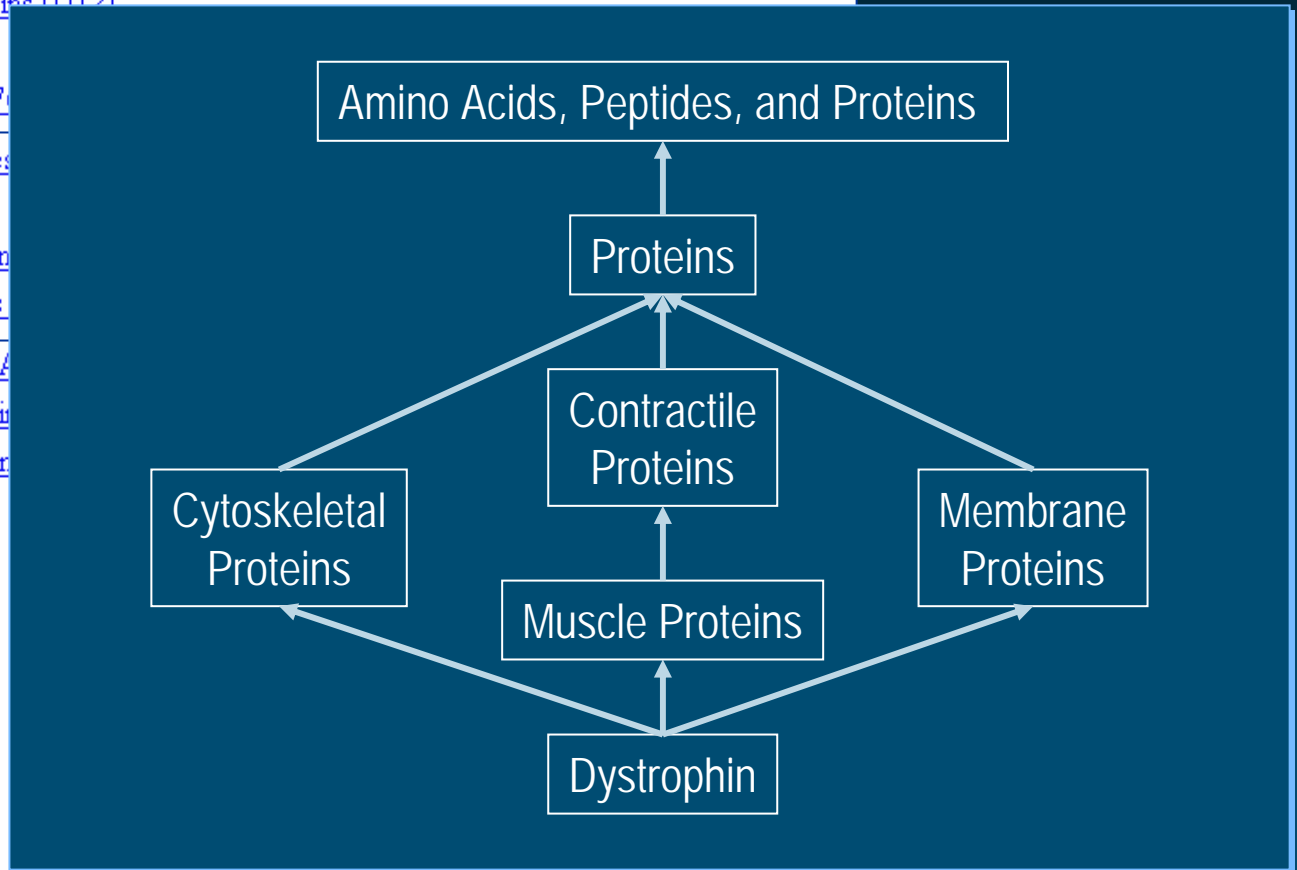
[Contractile Protein](#)

[Muscle Proteins](#)

[Amino A](#)

[Protein](#)

[Mer](#)



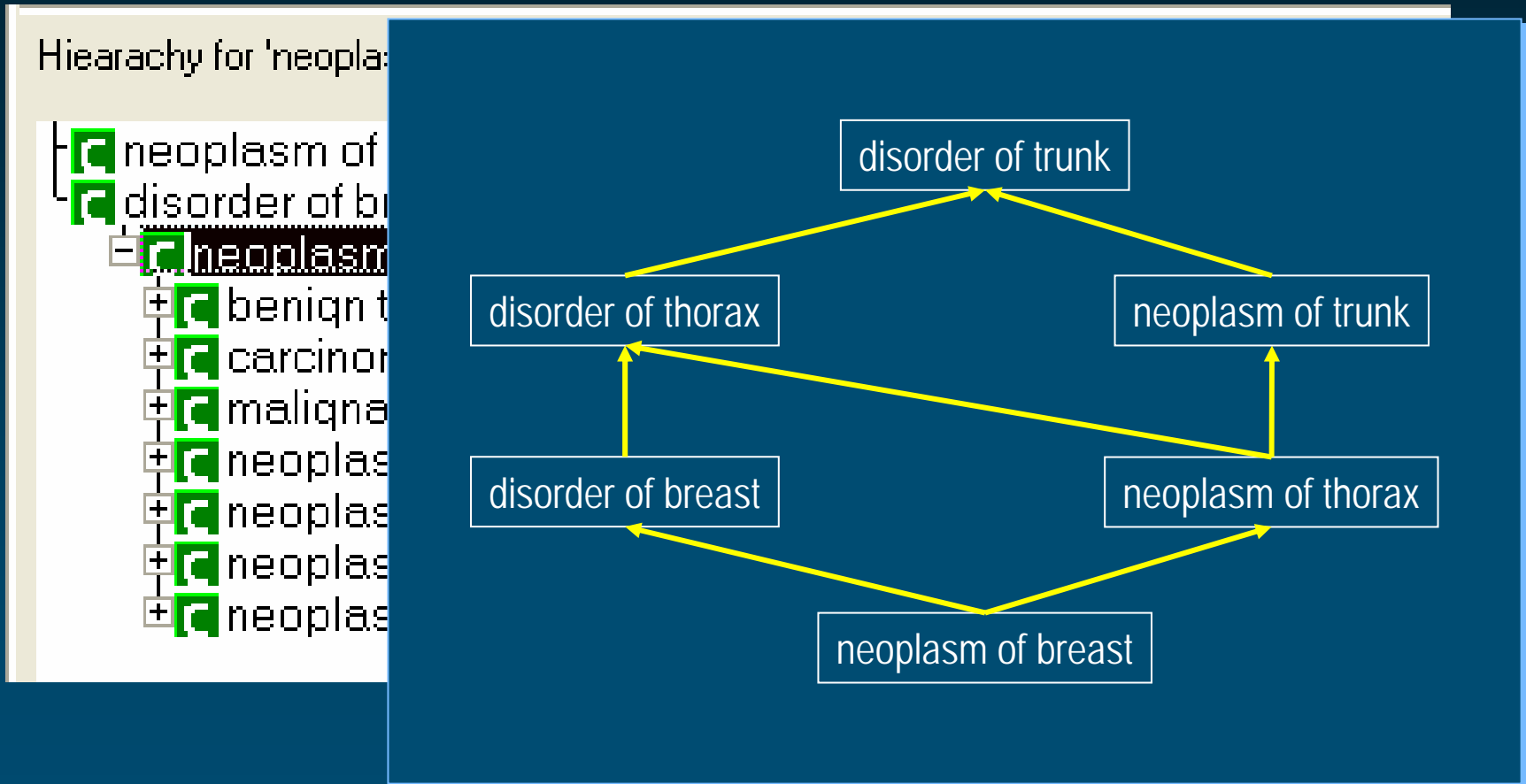
[Heterotrimeric GTP-Binding Proteins \[D12.776.543.325\] +](#)

<http://www.nlm.nih.gov/mesh/2007/MBrowser.html>





# SNOMED Clinical Terms



<http://www.clininfo.co.uk/clue5/clue.htm>



# Terminology integration

## *Unified Medical Language System*



# Addison's disease in medical vocabularies

## ◆ Synonyms

- Addisonian syndrome
  - Bronzed disease
  - Addison melanoderma
  - Asthenia pigmentosa
  - Primary adrenal deficiency
  - Primary adrenal insufficiency
  - Primary adrenocortical insufficiency
  - Chronic adrenocortical insufficiency
- } eponym
- } symptoms
- } clinical variants

# Organize terms

- ◆ Synonymous terms clustered into a concept
- ◆ Preferred term
- ◆ Unique identifier (CUI)

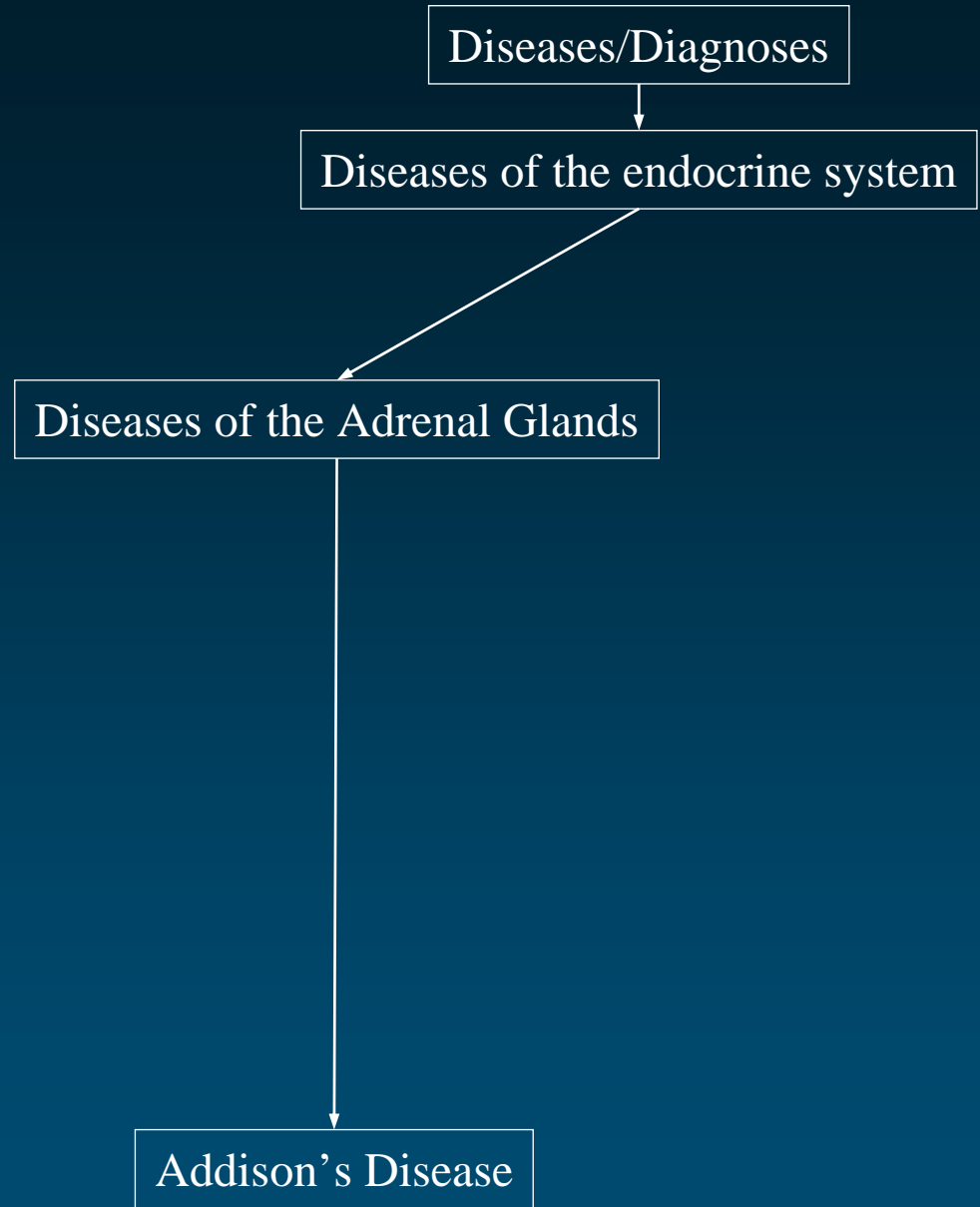
Addison Disease	MeSH	D000224
Primary hypoadrenalism	MedDRA	10036696
Primary adrenocortical insufficiency	ICD-10	E27.1
Addison's disease (disorder)	SNOMED CT	363732003

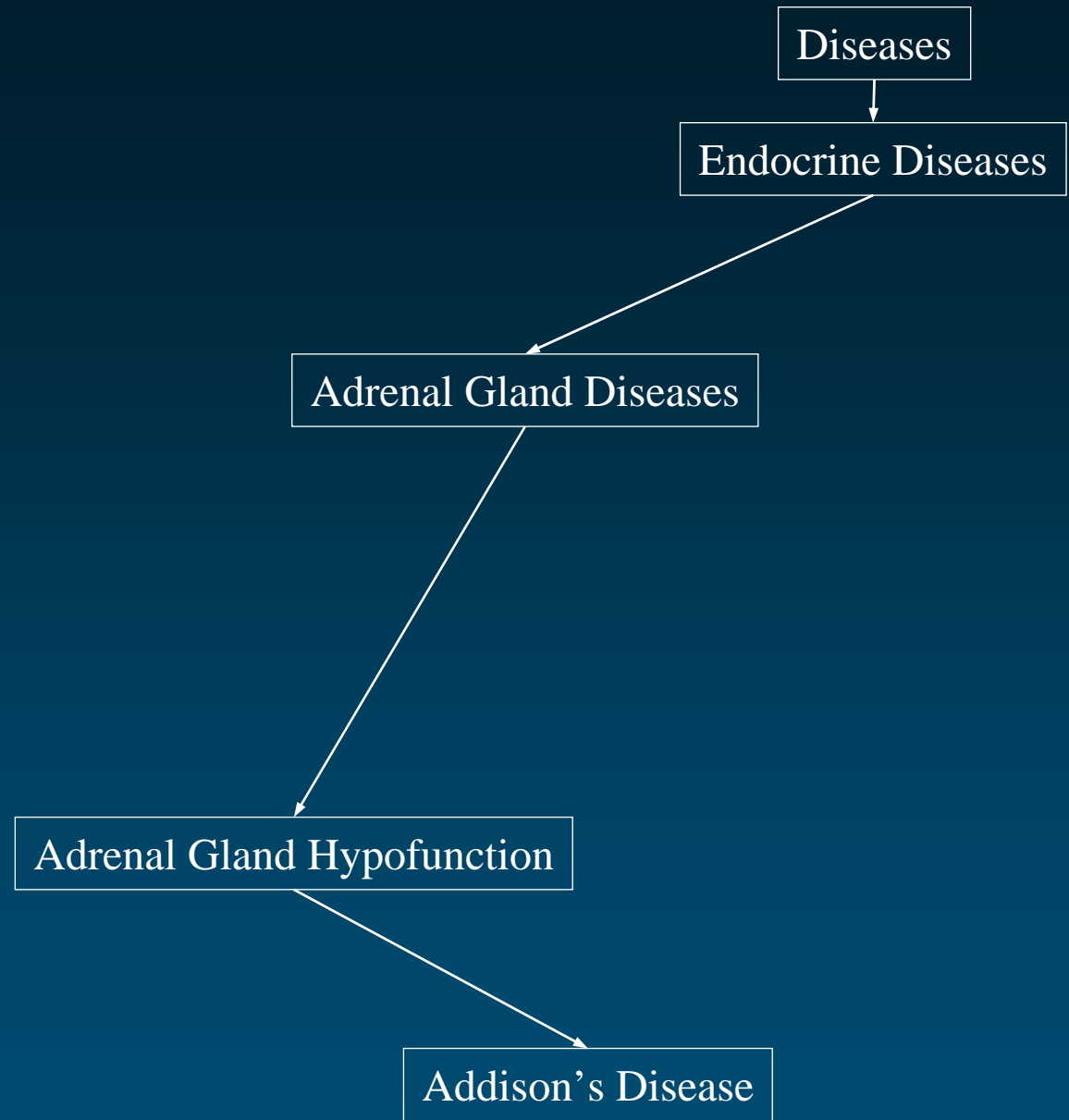
C0001403

Addison's disease



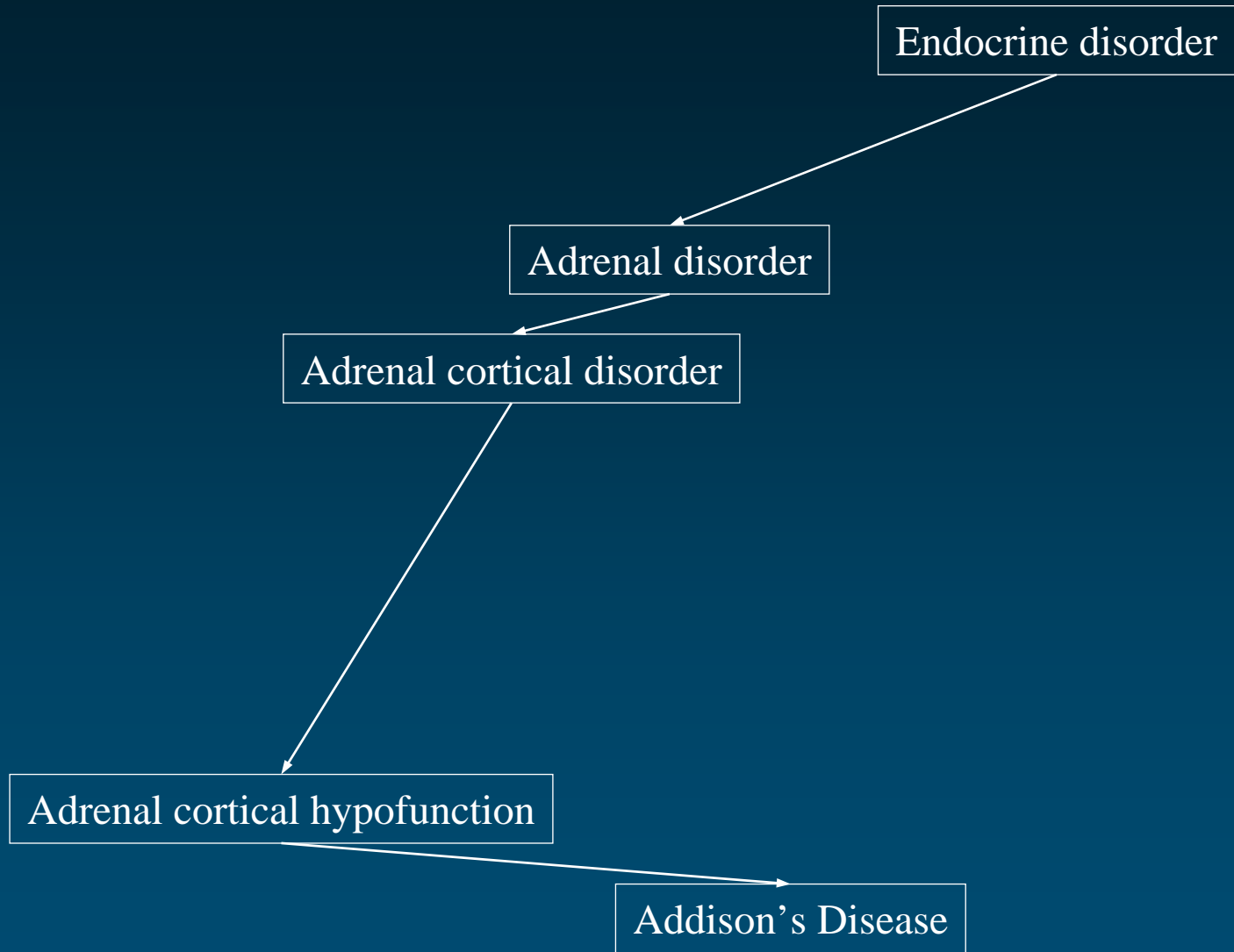
# SNOMED International



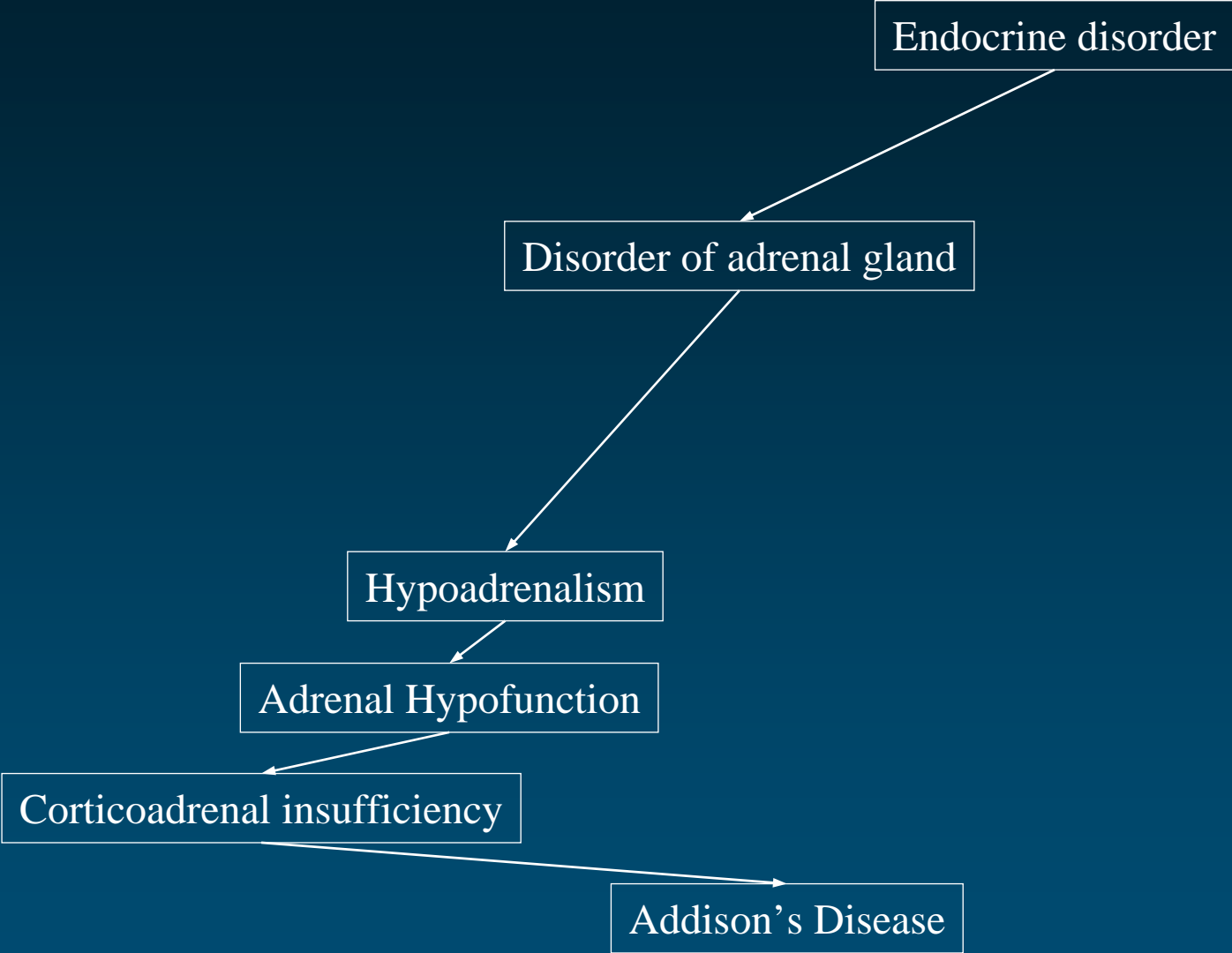




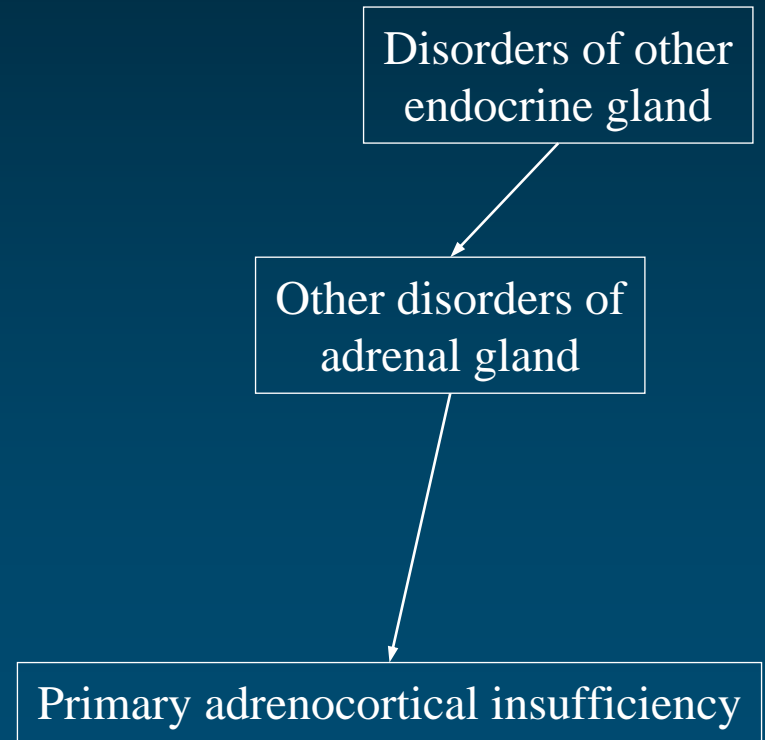
# AOD



# Read Codes

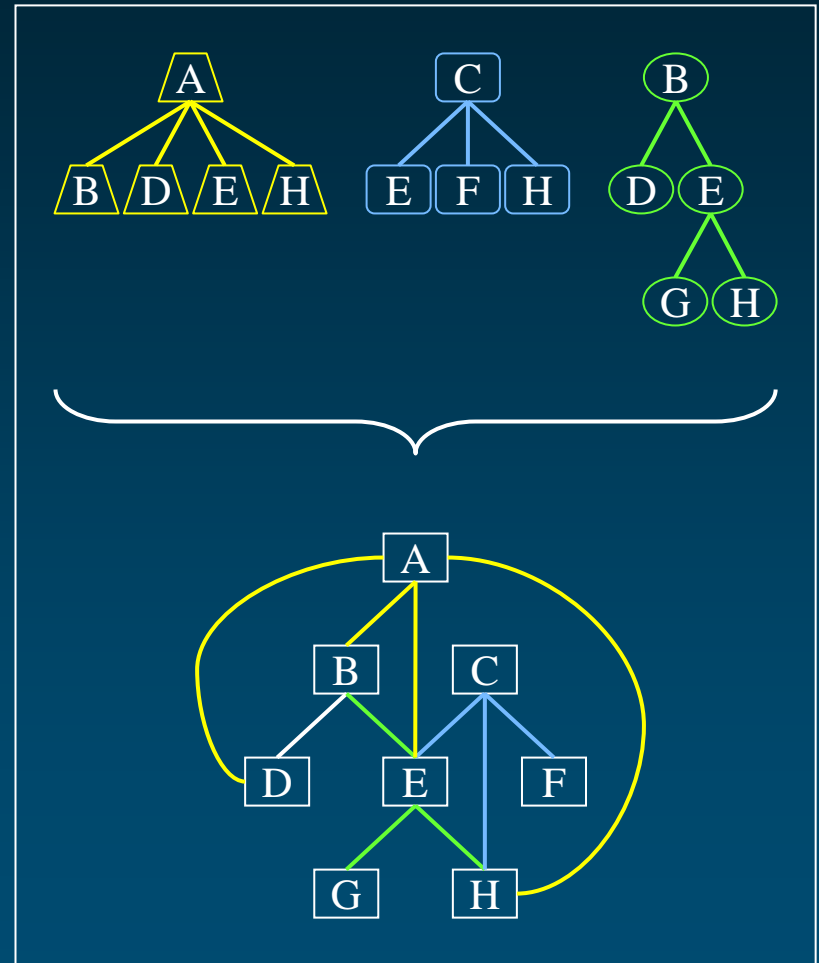


# ICD-10

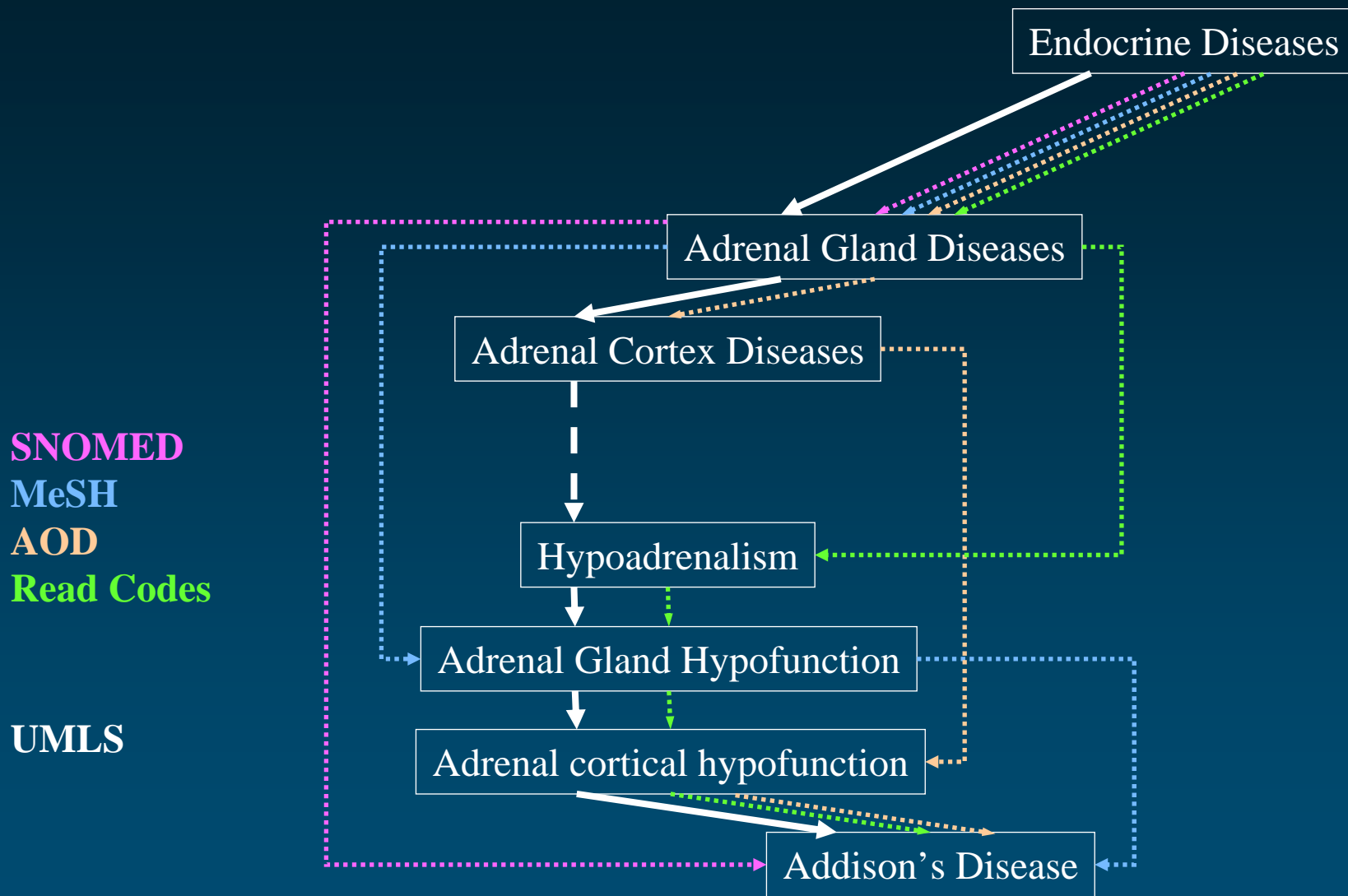


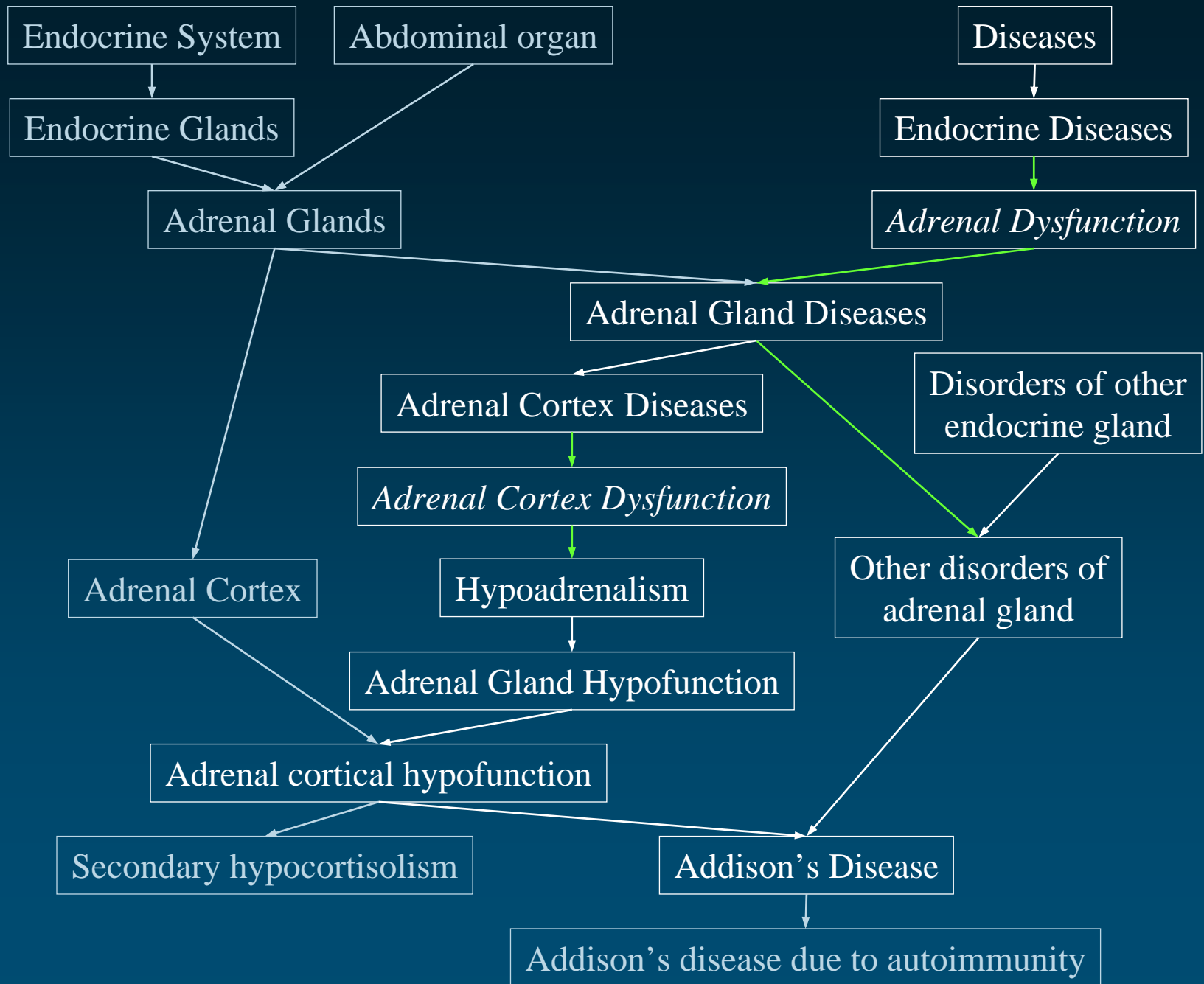
# Organize concepts

- ◆ Inter-concept relationships: hierarchies from the source vocabularies
- ◆ Redundancy: multiple paths
- ◆ One graph instead of multiple trees (multiple inheritance)



*organize concepts*





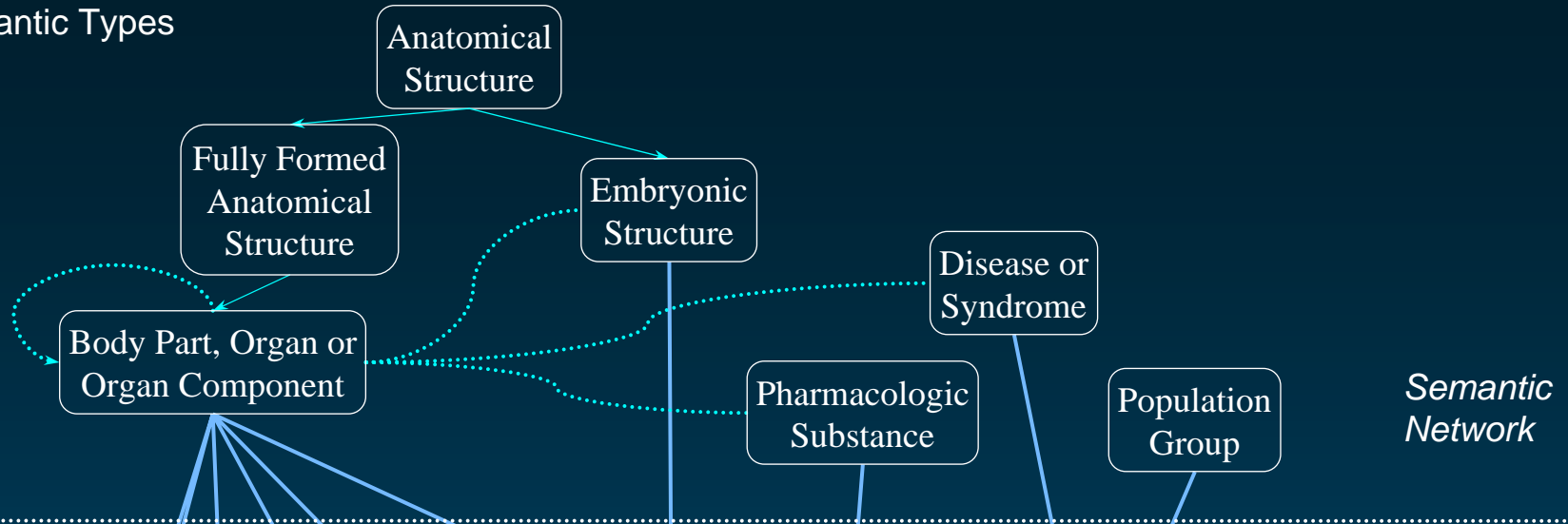


# Source Vocabularies

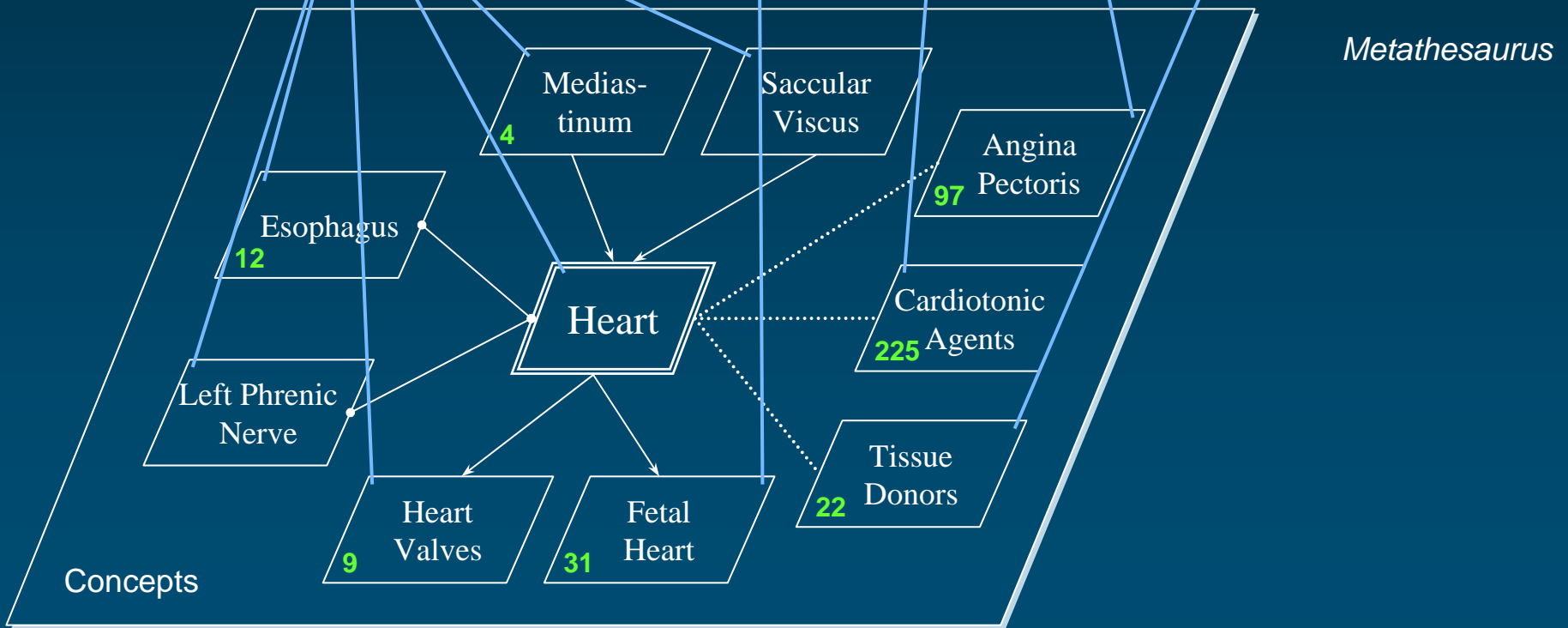
(2007AA)

- ◆ 139 source vocabularies
  - 17 languages
- ◆ Broad coverage of biomedicine
  - 5.5M names
  - 1.4M concepts
  - 16M relations
- ◆ Common presentation

Semantic Types



*Semantic Network*



*Metathesaurus*

Concepts



# Biomedical forest vs. graph

# UMLS Knowledge Source Server



## UMLS Knowledge Source Server (UMLSKS)

UMLSKS Version 5.0      UMLS Releases: 2002 2002AB 2002AC 2002AD 2003AA 2003AB 2003AC 2004AA 2004AB 2004AC 2005AA 2005AB 2005AC 2006AA 2006AB 2006AC 2006AD 2007AA

[Metathesaurus](#)

[Semantic Network](#)

[SPECIALIST Lexicon](#)

[Logout](#)

### About the UMLSKS

- [Home](#)
- [Overview](#)
- [Frequently Asked Questions](#)
- [Edit Views/Profile](#)

### Downloads

- [UMLS Knowledge Sources](#)
- [RxNorm Files](#)
- [Mappings](#)
- [VA/KP Problem List](#)
- [Developer's API](#)

### Documentation



Metathesaurus



Semantic N/W



SPECIALIST  
Lexicon

### Quick Search

Select UMLS  
Release:

2007AA

Enter search  
value:

Addison's disease

Metathesaurus  
Concept  
Search

[Search Tips...](#)

Semantic  
Network  
Search

[Search  
Tips...](#)

SPECIALIST  
Lexicon  
Search

[Search  
Tips...](#)

### Advanced Searches

#### [Metathesaurus Advanced Search](#)

Facilitates advanced searching of the UMLS Metathesaurus, including restricting vocabularies, performing batch searches, performing XML queries, and using a command-line type interface.

<http://umlsks.nlm.nih.gov/>



# Addison's disease in UMLS/SKS (1)

## Ancestors:

### MeSH

[MeSH Descriptors \[\]](#)  
[Index Medicus Descriptor \[\]](#)  
[Diseases \(MeSH Category\) \[C\]](#)  
[Endocrine System Diseases \[C19\]](#)  
[Adrenal Gland Diseases \[C19.053\]](#)  
[Adrenal Insufficiency \[C19.053.500\]](#)  
[Addison Disease \[C19.053.500.263\]](#)

### MeSH

[MeSH Descriptors \[\]](#)  
[Index Medicus Descriptor \[\]](#)  
[Diseases \(MeSH Category\) \[C\]](#)  
[Immune System Diseases \[C20\]](#)  
[Autoimmune Diseases \[C20.111\]](#)  
[Addison Disease \[C20.111.163\]](#)

### MeSH German

[Endokrine Krankheiten \[\]](#)  
[Nebennierenkrankheiten \[\]](#)  
[Nebennierenunterfunktion \[\]](#)  
[Addison-Krankheit \[\]](#)



# Addison's disease in UMLSKS (2)

Ancestors:	
MeS	<b>Alcohol and Other Drug Thesaurus</b>
MeS	<u>health and disease [G]</u>
InC	<u>disorder by body system or organ function [GK]</u>
D	<u>endocrine disorder [GV]</u>
	<u>adrenal disorder [GV14]</u>
	<u>adrenal cortical disorder [GV14.02]</u>
	<u>adrenal cortical hypofunction [GV14.02.06]</u>
	<u>Addison's disease [GV14.02.06.02]</u>
MeS	<b>CRISP Thesaurus</b>
MeS	<u>disease/disorder []</u>
InC	<u>endocrine disorder []</u>
D	<u>adrenal disorder []</u>
	<u>hypoadrenalism []</u>
	<u>Addison's disease []</u>
MeS	<b>CRISP Thesaurus</b>
End	<u>disease/disorder []</u>
Ne	<u>immunopathology []</u>
N	<u>autoimmune disorder []</u>
	<u>Addison's disease []</u>
	<u>ADDISON-DISEASE []</u>

# Addison's disease in UMLS SKS (3)

Ancestors:	
MeS <u>Me:</u> <u>Inc</u> <u>D</u>	<p><b>Alcohol and Other Drug Thesaurus</b></p> <p><b>ICD-10</b></p> <ul style="list-style-type: none"> <li><u>Endocrine, nutritional and metabolic diseases []</u></li> <li><u>Disorders of other endocrine glands []</u></li> <li><u>Other disorders of adrenal gland []</u></li> <li><u>Primary adrenocortical insufficiency []</u></li> </ul> <hr/> <p><b>ICD-10 Austral Mod</b></p> <ul style="list-style-type: none"> <li><u>Endocrine, nutritional and metabolic diseases []</u></li> <li><u>Disorders of other endocrine glands []</u></li> <li><u>Other disorders of adrenal gland []</u></li> <li><u>Primary adrenocortical insufficiency []</u></li> </ul>
MeS <u>Me:</u> <u>Inc</u> <u>D</u>	<p><b>CRI:</b></p> <ul style="list-style-type: none"> <li><u>Disorders of other endocrine glands []</u></li> <li><u>Other disorders of adrenal gland []</u></li> <li><u>Primary adrenocortical insufficiency []</u></li> </ul> <hr/> <p><b>MedDRA</b></p> <ul style="list-style-type: none"> <li><u>Endocrine disorders []</u></li> <li><u>Adrenal gland disorders []</u></li> </ul>
MeS <u>End</u> <u>Ne</u> <u>N</u>	<p><b>CRI:</b></p> <ul style="list-style-type: none"> <li><u>Adrenal cortical hypofunctions []</u></li> <li><u>Addison's disease []</u></li> </ul> <hr/> <p><b>MedDRA</b></p> <ul style="list-style-type: none"> <li><u>Metabolism and nutrition disorders []</u></li> <li><u>Metabolism disorders NEC []</u></li> <li><u>Metabolic disorders NEC []</u></li> <li><u>Addison's disease []</u></li> </ul>

# Addison's disease in UMLS SKS (4)

Ancestors:	
MeS	Alcohol and Other Drug Thesaurus
MeS	ICD-10
MeS	MedDRA Japanese
MeS	内分泌障害 []
MeS	副腎障害 []
MeS	副腎皮質機能低下 []
MeS	原発性副腎機能不全 []
MeS	ICD-10
MeS	MedDRA Japanese
MeS	代謝および栄養障害 []
MeS	代謝障害 NEC []
MeS	代謝障害 NEC []
MeS	原発性副腎機能不全 []
MeS	MedDRA Portuguese
MeS	Doenças do metabolismo e da nutrição []
MeS	Afecções metabólicas NC []
MeS	Alterações metabólicas NC []
MeS	Doença de Addison []
MeS	MedDRA Portuguese
MeS	Doenças endócrinas []
MeS	Doenças das glândulas supra-renais []
MeS	Hipofuncionamento cortical suprarenal []
MeS	Doença de Addison []

# Addison's disease in UMLS SKS (5)

Ancestors:

MeS	Alcohol and Other Drug Thesaurus	MeS	heal	ICD-10	MedDRA Japanese	SNOMED Clinical Terms
<a href="#">MeS</a>	<a href="#">heal</a>	<a href="#">MeS</a>	<a href="#">dis</a>	<a href="#">End</a>	<a href="#">Di</a>	<a href="#">内</a>
<a href="#">Inc</a>	<a href="#">dis</a>	<a href="#">End</a>	<a href="#">e</a>	<a href="#">Di</a>	<a href="#">副</a>	<a href="#">Clinical finding</a> []
<a href="#">D</a>	<a href="#">e</a>	<a href="#">C</a>		<a href="#">C</a>	<a href="#">置</a>	<a href="#">Finding by site</a> []
						<a href="#">Finding of body region</a> []
						<a href="#">Finding of trunk structure</a> []
						<a href="#">General finding of abdomen</a> []
						<a href="#">Abdominal organ finding</a> []
						<a href="#">Disorder of adrenal gland</a> []
						<a href="#">Disorder of adrenal cortex</a> []
						<a href="#">Adrenal cortical hypofunction</a> []
						<a href="#">Primary adrenocortical insufficiency</a> []
MeS	CRI	MeS	dise	ICD-10	Med	SNOMED Clinical Terms
<a href="#">MeS</a>	<a href="#">dise</a>	<a href="#">MeS</a>	<a href="#">en</a>	<a href="#">End</a>	<a href="#">代</a>	<a href="#">Clinical finding</a> []
<a href="#">Inc</a>	<a href="#">en</a>	<a href="#">End</a>	<a href="#">a</a>	<a href="#">Di</a>	<a href="#">代</a>	<a href="#">Finding by site</a> []
<a href="#">D</a>	<a href="#">a</a>	<a href="#">C</a>		<a href="#">C</a>	<a href="#">付</a>	<a href="#">Finding of body region</a> []
						<a href="#">Finding of trunk structure</a> []
						<a href="#">General finding of abdomen</a> []
						<a href="#">Disorder of abdomen</a> []
						<a href="#">Disorder of adrenal gland</a> []
						<a href="#">Hypoadrenalism</a> []
						<a href="#">Adrenal hypofunction</a> []
						<a href="#">Adrenal cortical hypofunction</a> []
						<a href="#">Primary adrenocortical insufficiency</a> []
MeS	CRI	MeS	dise	ICD-10	Med	SNOMED Clinical Terms
<a href="#">End</a>	<a href="#">dise</a>	<a href="#">MeS</a>	<a href="#">imm</a>	<a href="#">End</a>	<a href="#">代</a>	<a href="#">Clinical finding</a> []
<a href="#">Ne</a>	<a href="#">imm</a>	<a href="#">End</a>	<a href="#">a</a>	<a href="#">Di</a>	<a href="#">代</a>	<a href="#">Finding by site</a> []
<a href="#">N</a>	<a href="#">a</a>	<a href="#">C</a>		<a href="#">C</a>	<a href="#">付</a>	<a href="#">Finding of body region</a> []
						<a href="#">Finding of trunk structure</a> []
						<a href="#">General finding of abdomen</a> []
						<a href="#">Disorder of adrenal gland</a> []
						<a href="#">Adrenal hypofunction</a> []
						<a href="#">Adrenal cortical hypofunction</a> []
						<a href="#">Primary adrenocortical insufficiency</a> []

# UMLS Semantic Navigator

The screenshot displays the UMLS Semantic Navigator interface. The central area shows a hierarchical diagram of concepts related to Addison's disease. The central node is "Addison's disease", which is connected to "Adrenal gland hypofunction" and "Adrenal cortical hypofunction". "Adrenal gland hypofunction" is connected to "Non-Neoplastic Adrenal Gland Disorder" and "Metabolic disorders NEC". "Adrenal cortical hypofunction" is connected to "Adrenal cortical dysfunction" and "Dysfu...". The diagram uses arrows and § symbols to indicate relationships between concepts.

**Siblings**

**Concepts & Ideas**

- Clinical Syndromes □

**Disorders**

- Aarskog syndrome □
- Abstinence Syndrome □
- Achard Syndrome □
- Acroparesthesia Syndrome □
- ACTH deficiency □
- Addisonian crisis □
- Adie Syndrome □
- Adrenal cortex atrophy □
- Adrenal insufficiency due to adrenal

**Chemicals & Drugs**

- Acetaminophen 25 MG/ML / Dextromethorphan 1 MG/ML / Guaifenesin 10 MG/ML / Pseudoephedrine 3 MG/ML Oral Solution □

**Co-occurring Concepts**

**Anatomy**

- Adrenal Cortex [14] □
- Adrenal Glands [18] □
- Liver [2] □

**Chemicals & Drugs**

**BCI Addison's disease LEGEND**

Start again Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

**Similar Concepts**

- Adrenal cortical hypofunction □

**Closest MeSH Terms**

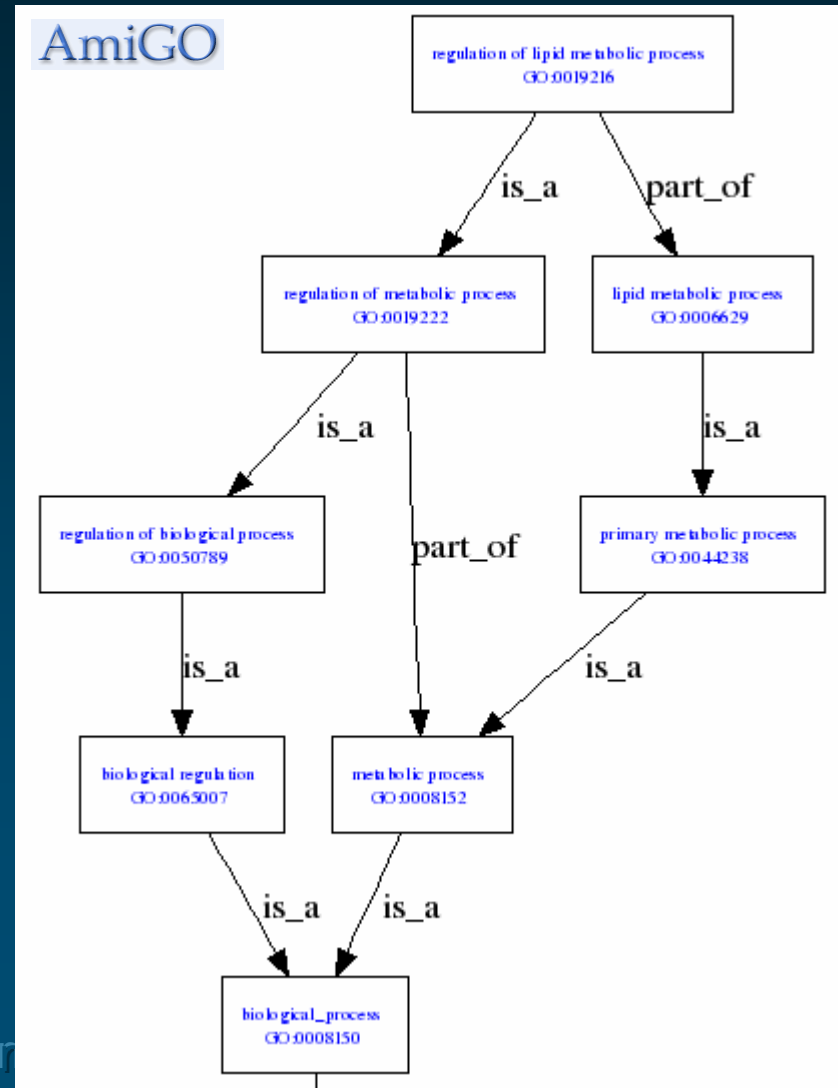
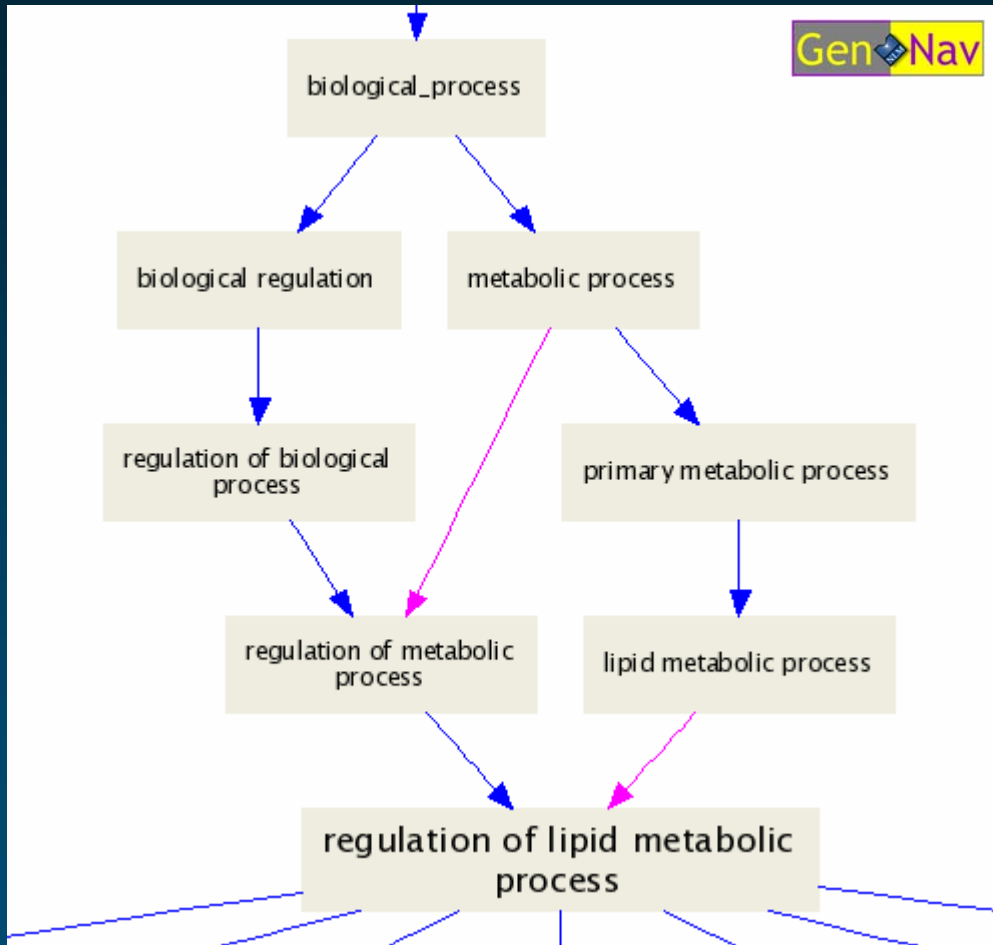
Main Headings



- ⊕ all : all [189459]
  - ⊕ ⓘ GO:0008150 : biological\_process [137743]
    - ⊕ ⓘ GO:0065007 : biological regulation [19935]
      - ⊕ ⓘ GO:0050789 : regulation of biological process [18154]
        - ⊕ ⓘ GO:0019222 : regulation of metabolic process [9816]
          - ⊕ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**
  - ⊕ ⓘ GO:0008152 : metabolic process [54684]
    - ⊕ ⓘ GO:0044238 : primary metabolic process [44818]
      - ⊕ ⓘ GO:0006629 : lipid metabolic process [3940]
        - ⊕ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**
  - ⊕ ⓘ GO:0019222 : regulation of metabolic process [9816]
    - ⊕ ⓘ **GO:0019216 : regulation of lipid metabolic process [108]**

<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>

# GenNav



<http://mor.nlm.nih.gov/perl/gennav.pl>

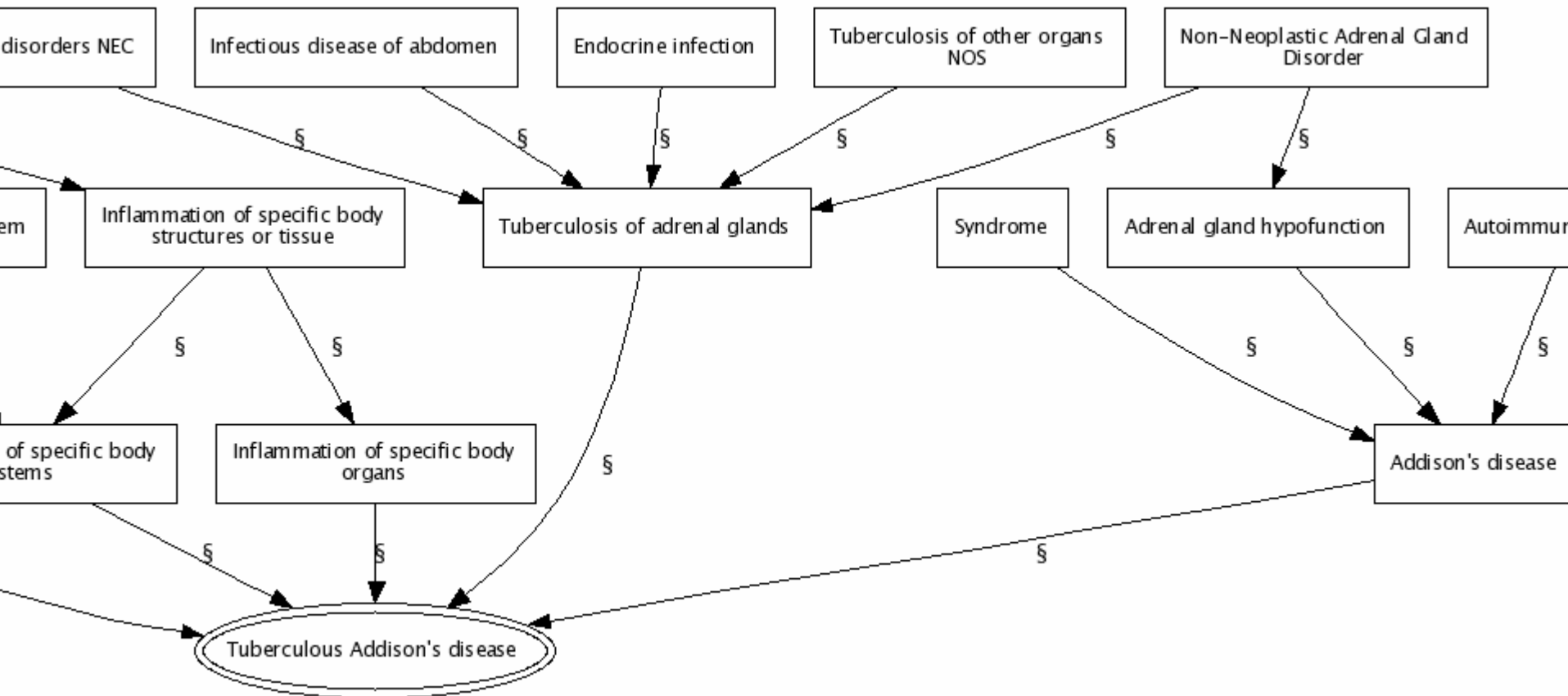


Lister Hill National Center for Biomedical Research and Innovation

# Semantics of the UMLS graph

## *Issues and challenges*

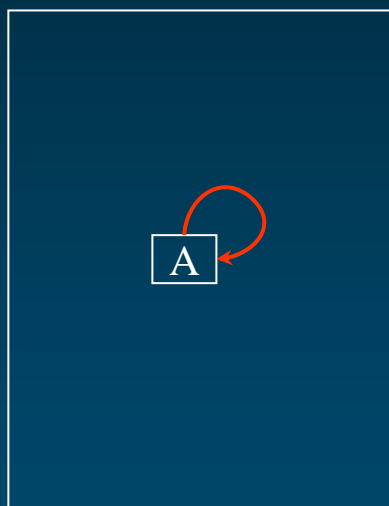
# Visualization of large graphs



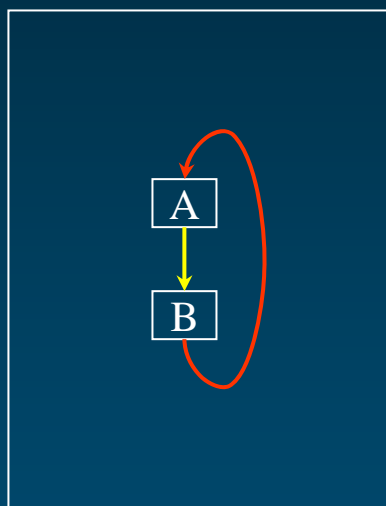


# Acyclicity

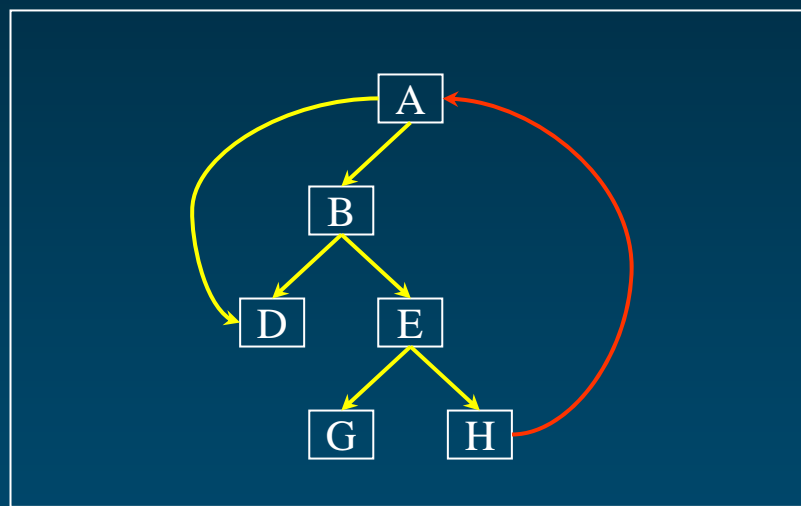
“back edge” from a child concept to a parent concept



Reflexive  
13,000



Direct  
1800



Indirect  
120

# Underspecification of relationships

- ◆ Relationship “attribute” not always present
- ◆ Relations used to create hierarchies vs. hierarchical relations

[Environment and Public Health \[G03\]](#)

[Public Health \[G03.850\]](#)

▶ [Accidents \[G03.850.110\]](#)

[Accident Prevention \[G03.850.110.060\]](#) +

[Accidental Falls \[G03.850.110.085\]](#)

[Accidents, Aviation \[G03.850.110.185\]](#)

[Accidents, Home \[G03.850.110.205\]](#)

[Accidents, Occupational \[G03.850.110.250\]](#) +

[Accidents, Radiation \[G03.850.110.285\]](#)

[Accidents, Traffic \[G03.850.110.320\]](#)

[Drowning \[G03.850.110.500\]](#) +

# Which tasks?

- ◆ Information integration
- ◆ Mapping
  
- ◆ Depending on the degree of human involvement
  - Hypothesis generation / validation
  - Knowledge discovery
  - Automated reasoning
- ◆ Knowledge standardization
  - Common format
  - Common semantics



# Which formalisms?

- ◆ SKOS – Thesaurus
  - Simple Knowledge Organization Schema
- ◆ RDF – Concept-Relationship-Concept triples
  - Resource Description Framework
- ◆ Description Logics / Frames
  - OWL Web Ontology Language
  - Protégé (frames / OWL)
  - OBO Open Biomedical Ontology
- ◆ Rule languages
- ◆ Formal logic

# Which identifiers?

## ◆ For concepts

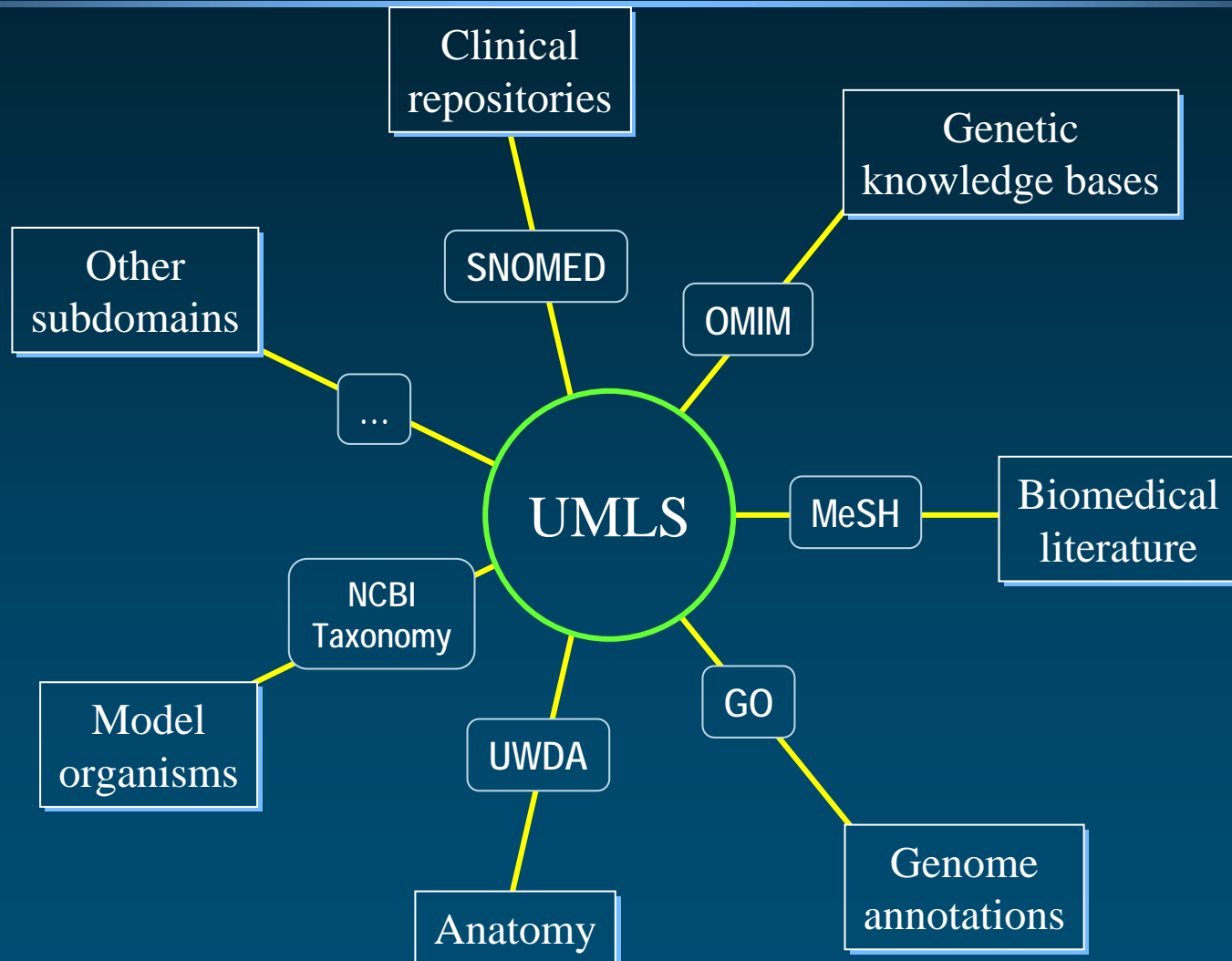
- Namespaces, ontologies, knowledge bases
  - OBO – Open Biomedical Ontologies
  - UMLS – Unified Medical Language System
  - NCBI Entrez (Entrez Gene, GenBank, UniGene, ...)
- Mappings across information sources

## ◆ For relationships

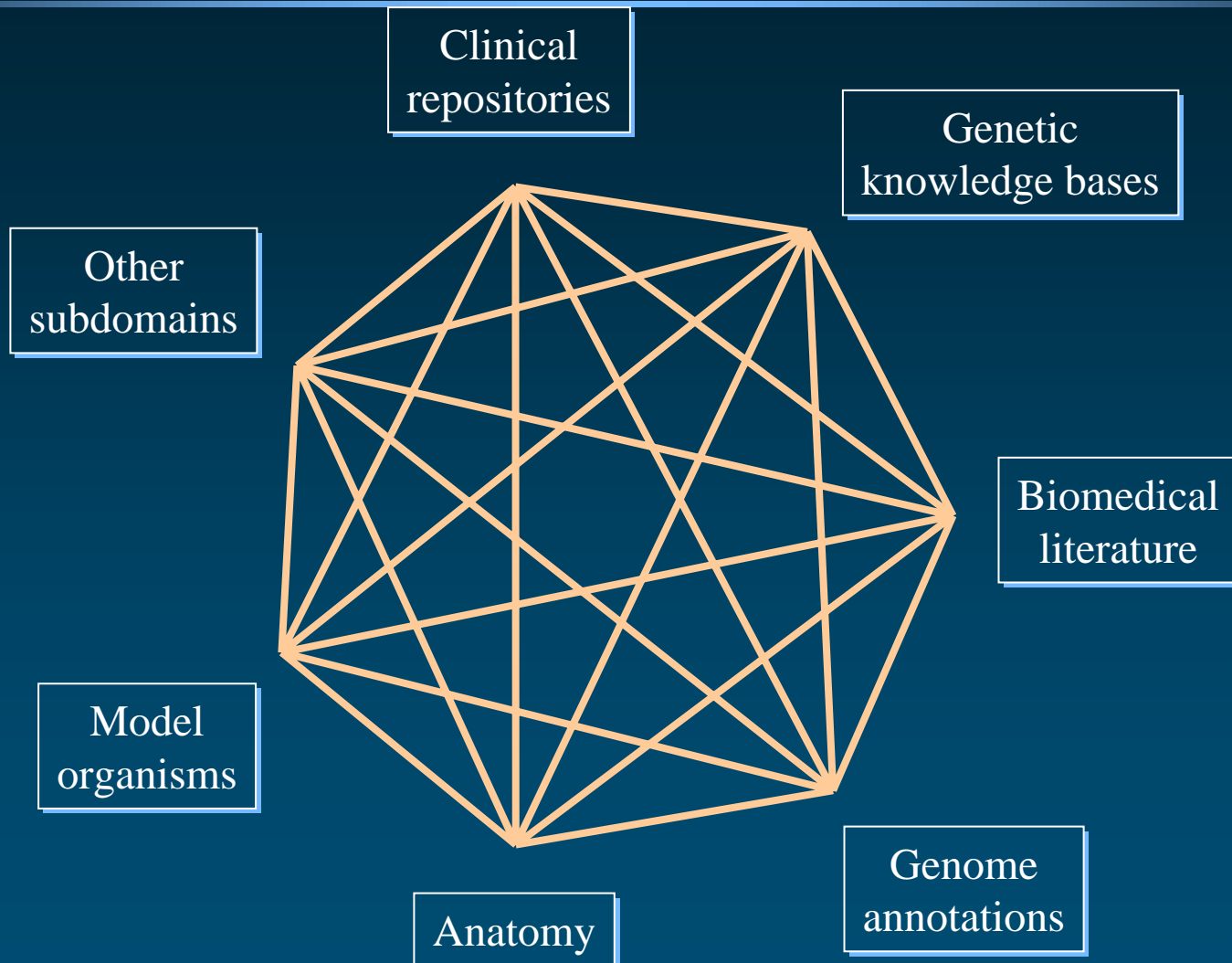


# Conclusions

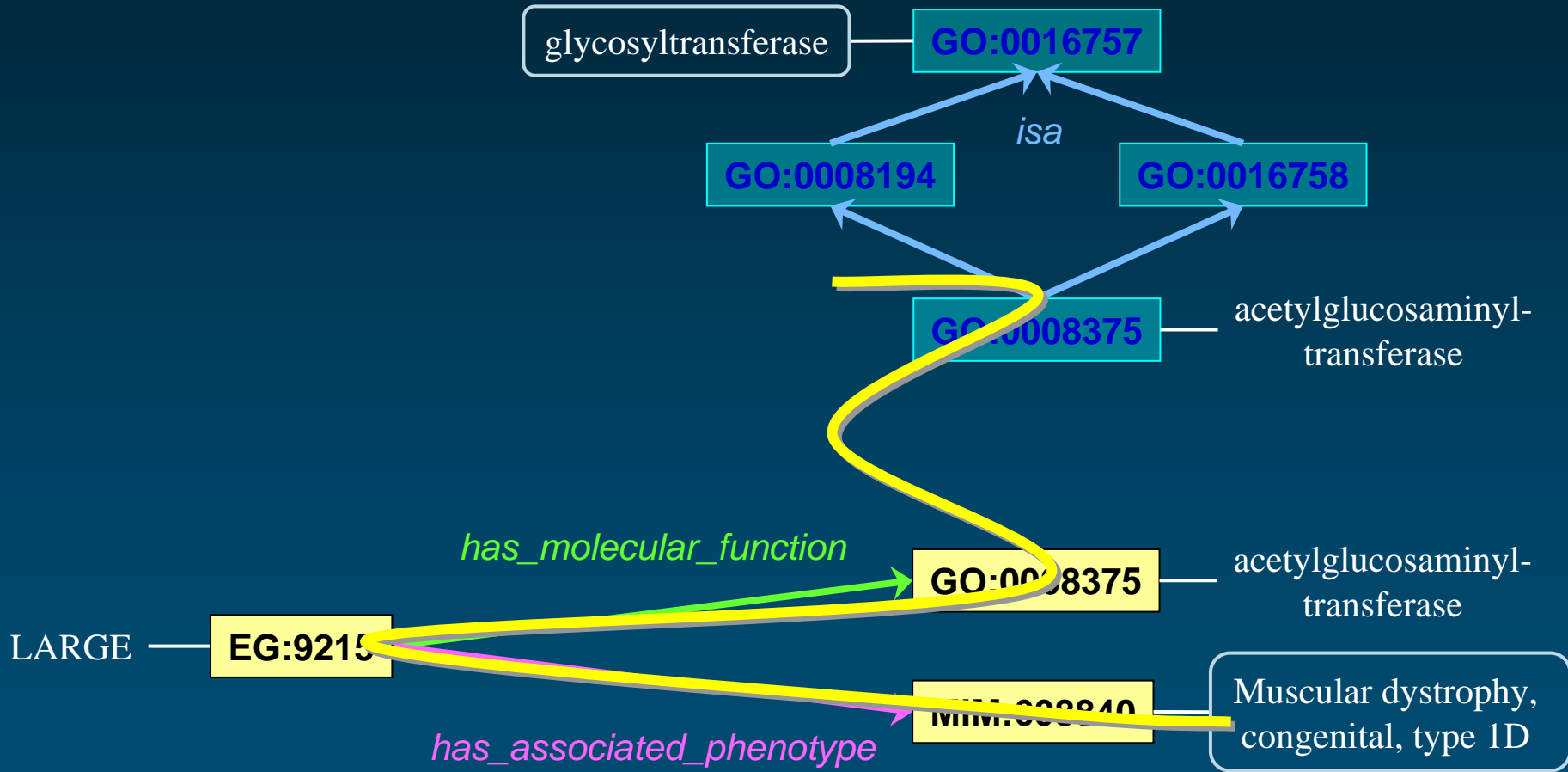
# Integrating subdomains

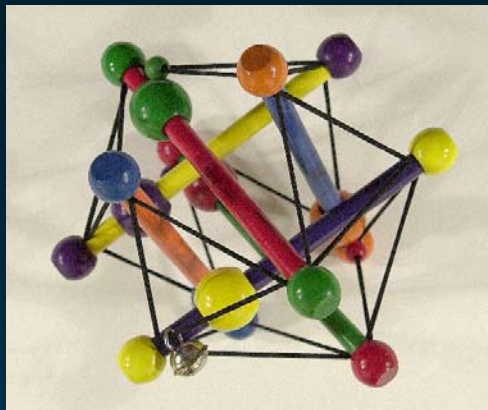


# Integrating subdomains



# From *glycosyltransferase* to *congenital muscular dystrophy*





# Medical Ontology Research

Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)

Web: [mor.nlm.nih.gov](http://mor.nlm.nih.gov)



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA

# UMLS References

## ◆ UMLS

[umlsinfo.nlm.nih.gov](http://umlsinfo.nlm.nih.gov)

## ◆ UMLS browsers

(free, but UMLS license required)

- Knowledge Source Server: [umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov)
- Semantic Navigator:  
<http://mor.nlm.nih.gov/perl/semnav.pl>
- RRF browser  
(standalone application distributed with the UMLS)





# UMLS References

## ◆ Gentle introduction

- Bodenreider O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*; D267-D270.  
<http://mor.nlm.nih.gov/pubs/pdf/2004-nar-ob.pdf>

## ◆ Seminal paper

- Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods Inf Med*, 32(4), 281-91.



# Biomedical information integration through RDF

## ◆ Biomedical perspective

- Sahoo S, Zeng K, Bodenreider O, Sheth AP. (2007). From “glycosyltransferase” to “congenital muscular dystrophy”: Integrating knowledge from NCBI Entrez Gene and the Gene Ontology. *Proceedings of Medinfo (in press)*.  
<http://mor.nlm.nih.gov/pubs/pdf/2007-medinfo-ss.pdf>

## ◆ Semantic Web perspective

- Sahoo S, Zeng K, Bodenreider O, Sheth AP. (2007). An experiment in integrating large biomedical knowledge resources with RDF: Application to associating genotype and phenotype information. *Proceedings of the workshop on Health Care and Life Sciences Data Integration for the Semantic Web at the 16th International World Wide Web Conference (WWW2007) (in press)*.  
[http://mor.nlm.nih.gov/pubs/pdf/2007-www\\_hcls-ss.pdf](http://mor.nlm.nih.gov/pubs/pdf/2007-www_hcls-ss.pdf)

