Visualization Tools for Biomedical Knowledge

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

◆ Issues and Challenges

◆ **SemNav** (UMLS Semantic Navigator)  
*Visualizing terminological knowledge*

◆ **GenNav**  
*Visualizing gene annotations*

◆ **RxNav**  
*Visualizing drug information*
Issues and Challenges
Issues

◆ Size
  ● Large number of concepts (>1 million)

◆ Complexity
  ● Polyhierarchical structures
  ● Multiple information sources
  ● Multiple properties

◆ Lack of formality
  ● Redundant relations
  ● Hierarchies vs. hierarchical relations
Challenges

◆ Restrict information space
  ● To selected information sources (SemNav)
  ● To selected organisms (GenNav)
◆ Reduce complexity (SemNav)
  ● Group concepts by semantic groups
  ● Transitive reduction on hierarchical relations
  ● Select co-occurring concepts
◆ Reduce the cognitive burden on the user
  ● Use graph-based rather than tree-based representations
UMLS Semantic Navigator

SemNav

http://umlsks.nlm.nih.gov*

► SN Resources ► Semantic Navigator

(* free UMLS registration required)
Unified Medical Language System®

- Developed at NLM since 1990
- 139 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.1M names
  - 1.3M concepts
  - 16M relations
- Integration
  - Synonymous terms are clustered in a concept
  - Hierarchies (trees) are combined in a graph structure
Terminology integration

Terms

- Duchenne muscular dystrophy
- Duchenne's muscular dystrophy
- Duchenne de Boulogne muscular dystrophy
- Duchenne type progressive muscular dystrophy
- Pseudohypertrophic muscular dystrophy
- X-liked recessive muscular dystrophy
- Severe generalized familial muscular dystrophy

MeSH, SNOMED
CTV3, Jablonski,
CRISP, DxPlain,
MedDRA, LOINC

COSTAR
Jablonski
SNOMED
MeSH, CTV3
SNOMED
Jablonski
SNOMED
Terminology integration

- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)
UMLS  A two-level structure

- **Two-level structure**
  - Semantic Network
    - 135 Semantic Types (STs)
    - 54 types of relationships among STs
  - Metathesaurus
    - >1M concepts
    - ~12 M inter-concept relationships
  - Link = categorization
The diagram illustrates the semantic network of the heart, categorizing terms under different semantic types. The heart is central, with links to various concepts such as esophagus, left phrenic nerve, heart valves, fetal heart, and mediastinum. The network also includes categories like anatomical and embryonic structure, pharmacologic substance, disease or syndrome, and tissue donors. The numbers next to each concept suggest the frequency or prominence of these terms within the network.
Concept: Dystrophin

CUI: C0079259

Semantic Type: Amino Acid, Peptide, or Protein
Biologically Active Substance

Definition:
A muscle protein localized in surface membranes which is the product of the Duchenne/Becker muscular dystrophy gene. Individuals with Duchenne muscular dystrophy usually lack dystrophin completely while those with Becker muscular dystrophy have dystrophin of an altered size. It shares features with other cytoskeletal proteins such as SPECTRIN and alpha-actinin but the precise function of dystrophin is not clear. One possible role might be to preserve the integrity and alignment of the plasma membrane to the myofibrils during muscle contraction and relaxation. MW 400 kDa. (MeSH)

large, structural, spectrin-like protein expressed in skeletal muscle, genetic defect is linked to Duchenne and Becker muscular dystrophy. (CRISP Thesaurus)

Synonyms:
Dystrophin

Ancestors:
MeSH
- MeSH Descriptors [I]
- Index Medicus Descriptor [I]
- Chemicals and Drugs (MeSH Category) [D]
- Amino Acids, Peptides, and Proteins [I]
Amino Acids, Peptides, and Proteins [D12]
Proteins [D12,776]

MeSH Browser

Amino Acids, Peptides, and Proteins [D12]
Proteins [D12,776]

Actinin [D12,776,210,500,095]
Actins [D12,776,210,500,100]
Actomyosin [D12,776,210,500,154]
Calsequestrin [D12,776,210,500,220]
CapZ Actin Caping Protein [D12,776,210,500,227]
Cavolin 3 [D12,776,210,500,235]
Cofilin 2 [D12,776,210,500,242]

Dystrophin [D12,776,210,500,250]
Dystrophin-Associated Proteins [D12,776,210,500,410] +
Myogenic Regulatory Factors [D12,776,210,500,570] +
Myoglobin [D12,776,210,500,580]
Myosins [D12,776,210,500,600] +
Parvalbumin [D12,776,210,500,750]
Profilins [D12,776,210,500,775]
Ryanodine Receptor Calcium Release Channel [D12,776,210,500,800]
Tropomodulin [D12,776,210,500,847]
Tropomyosin [D12,776,210,500,895]
Tropinin [D12,776,210,500,910] +

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] +
Erbins [D12,776,543,287] +
siblings:
- 120-kDa hemocytometer-specific membrane protein
- 15a protein, desmin delta
- 180-kDa gant
- Substosoma japonicum C
- 20kDa polypeptide, Micrasterias tabacum C
- 3D3 type protein, human C
- 3D3-type protein mouse C
- 5-lipoxygenase ad protein C
- 7 kDa protein, carassius C
- A14-5L protein C

other related concepts:
- Disorders
  - Muscular Dystrophies C
  - Muscular Dystrophy Duchene C
- Living Beings
  - Mice, inbred C57BL6
- Co occurring
  - Concepts
  - Anatomy
    - Astrocytes C
    - Brain C
    - Cell membrane C
    - Cytoskeleton C
    - Displacment (Anatomy) C
    - Heart C
    - Heart (Accumulation) C

similar concepts:
- Closest MeSH Terms
- (none)

allegedly:
- (none)
Lister Hill National Center for Biomedical Communications

Siblings
- 120-kDa hemocyte-specific membrane protein, flesh fly
- 15a protein, Aedes aegypti
- 22.6-kDa antigen, Schistosoma japonicum
- 22kDa polypeptide, Nicotiana tabacum
- 3D3-lyric protein, human
- 3D3-lyric protein, mouse
- 4.1B protein, mouse
- 5-lipoxygenase-acting protein
- 7 kDa protein, carlavirin
- A14.5L protein, cytoskeletal proteins

Other Related Concepts
- Disorders
  - Muscular Dystrophies
  - Muscular Dystrophy
  - Duchenne
- Living Beings
  - Mice, inbred mdx

Co-occurring Concepts
- Anatomy
  - Astrocyte [4]
  - Brain [24]
  - Cell membrane [7]
  - Cytoskeleton [5]
  - Diaphragm (Anatomy) [6]
  - Heart [9]
  - Hepatocellular (Biology) [10]
Amino Acid, Peptide, or Protein

- affects
- causes
- disease or syndrome
- complications
- produced by

Biologically Active Substance

- affects
- causes
- disease or syndrome
- complications
- produced by

Dystrophin

- Biologically Active Substance
- Amino Acid, Peptide or Protein
- Muscle Dystrophy, Duchenne

Semantic Types

Disease or Syndrome

Amino Acid, Peptide or Protein

Muscular Dystrophy, Duchenne

Frequency = 190

Metathesaurus Relationships

C1 otherwise related to C2

- MeSH

C1 co-occurs with C2

- MEDLINE

Semantic Network Relationships

Amino Acid, Peptide, or Protein

- affects
- causes
- disease or syndrome

Biologically Active Substance

- affects
- causes
- disease or syndrome
- complications
- produced by

Dystrophin

- Biologically Active Substance
- Amino Acid, Peptide or Protein
- Muscle Dystrophy, Duchenne

Semantic Types

Disease or Syndrome

Amino Acid, Peptide or Protein

Muscular Dystrophy, Duchenne

Frequency = 190

Metathesaurus Relationships

C1 otherwise related to C2

- MeSH

C1 co-occurs with C2

- MEDLINE

Semantic Network Relationships

Amino Acid, Peptide, or Protein

- affects
- causes
- disease or syndrome

Biologically Active Substance

- affects
- causes
- disease or syndrome
- complications
- produced by

Dystrophin

- Biologically Active Substance
- Amino Acid, Peptide or Protein
- Muscle Dystrophy, Duchenne

Semantic Types

Disease or Syndrome

Amino Acid, Peptide or Protein

Muscular Dystrophy, Duchenne

Frequency = 190
Technical details

◆ Simple web/cgi technology (apache, Perl)
◆ dot (GraphViz)
  ● PNG file (-Tpng)
  ● Client-side map (-Tcmap)
◆ Precompute the transitive closure on hierarchical relations to perform the transitive closure fast
◆ Remove cycles (UMLS)
Gene Ontology browser

Gene Ontology™

- Developed by the GO Consortium
- Several components (GO database)
  - Ontology (~17,000 concepts)
    - Molecular functions
    - Cellular components
    - Biological processes
  - Gene products (~1.6M)
  - Associations between Gene products and GO concepts (~6.8M)
all : all ( 179339 )
- GO:0008150 : biological_process ( 127417 )
- GO:0005575 : cellular_component ( 116421 )
  - GO:0005623 : cell ( 85785 )
    - GO:0005622 : intracellular ( 69449 )
      - GO:0005856 : cytoskeleton ( 2376 )
    - GO:0043229 : intracellular organelle ( 62778 )
      - GO:0043232 : intracellular non-membrane-bound organelle ( 7453 )
    - GO:0005856 : cytoskeleton ( 2376 )
  - GO:0043226 : organelle ( 62799 )
    - GO:0043229 : intracellular organelle ( 62778 )
      - GO:0043232 : intracellular non-membrane-bound organelle ( 7453 )
    - GO:0005856 : cytoskeleton ( 2376 )
  - GO:0043228 : non-membrane-bound organelle ( 7453 )
    - GO:0043232 : intracellular non-membrane-bound organelle ( 7453 )
    - GO:0005856 : cytoskeleton ( 2376 )
- GO:0003674 : molecular_function ( 121803 )
- GO:0004486 : cellular_component : obsolete biological_process ( 0 )
- GO:0004486 : cellular_component : obsolete cellular_component ( 0 )
- GO:0004486 : molecular_function : obsolete molecular_function ( 0 )
<table>
<thead>
<tr>
<th>Molecular functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• actin binding [TAS]</td>
</tr>
<tr>
<td>• structural constituent of cytoskeleton [TAS]</td>
</tr>
<tr>
<td>• protein binding [IPI]</td>
</tr>
<tr>
<td>• structural constituent of muscle [IDA, TAS]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biological processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• muscle contraction [NR]</td>
</tr>
<tr>
<td>• muscle development [NAS]</td>
</tr>
<tr>
<td>• peptide biosynthesis [IDA]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cellular components</th>
</tr>
</thead>
<tbody>
<tr>
<td>• cytoskeleton [TAS]</td>
</tr>
<tr>
<td>• dystrophin-associated glycoprotein complex [TAS]</td>
</tr>
<tr>
<td>• costamere [IDA]</td>
</tr>
<tr>
<td>Full name</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Nucleoside diphosphate kinase, NDP kinase</td>
</tr>
</tbody>
</table>

**Organisms**

- Arabidopsis thaliana (thale cress)
- Bos taurus (cow)
- Caenorhabditis elegans
- Danio rerio (zebrafish)
- Dictyostelium discoideum
- Drosophila melanogaster (fruit fly)
- Gallus gallus (chicken)
- Homo sapiens (human)
- Human immunodeficiency virus 1
- Human immunodeficiency virus type 1 (BRU ISOLATE)
- Human immunodeficiency virus type 1 (CLONE 12)
- Human immunodeficiency virus type 1 (PV22 ISOLATE)
- Macaca fascicularis (crab-eating macaque)
- Mus musculus (house mouse)
- Plasmodium falciparum (malaria parasite P. falciparum)
- Rattus norvegicus (Norway rat)
- Saccharomyces cerevisiae (baker’s yeast)
- Schizosaccharomyces pombe (fission yeast)
- Trypanosoma brucei
- Trypanosoma brucei TREN927
Term:

What: GO concept

Field: Automatic

String matching method: Automatic

Species restriction:

Arabidopsis thaliana (thale cress)
Bos taurus (cow)
Caenorhabditis elegans
Chlamydomonas reinhardtii
Danio rerio (zebrafish)
Dictyostelium discoideum
Drosophila melanogaster (fruit fly)
Escherichia coli
Hepatitis C virus
Homo sapiens (human)
Magnaporthe grisea (rice blast fungus)
Mus musculus (house mouse)
Mycoplasma pneumoniae
Neurospora crassa
Oryza sativa (rice)
Plasmodium falciparum (malaria parasite P. falciparum)
Pneumocystis carinii
Rattus norvegicus (Norway rat)
Saccharomyces cerevisiae (baker's yeast)

Comments? Feedback? Quit

Contact: Olivier Bodenreider

Copyright and Privacy Notice

Interface version: 1.07

Done
RxNorm browser

Normalized form

Strength
4mg/ml

Ingredient
Fluoxetine

Dose form
Oral Solution

Semantic clinical drug component

Strength

Ingredient

Dose form

Semantic clinical drug form
Generic vs. Brand

**Generic**
- Ingredient (IN)
- Clinical drug form (SCDF)
- Clinical drug component (SCDC)
- Clinical drug (SCD)

**Brand**
- Brand name (BN)
- Branded drug form (SBDF)
- Branded drug component (SBDC)
- Branded drug (SBD)

*tradename_of*
Relations among drug entities
RxNorm database

◆ Data sources
- Master Drug Data Base
- Multum MediSource Lex.
- Micromedex DRUGDEX
- FDA National Drug Code Directory
- National Drug Data File Plus Source Vocabulary
- VA National Drug File
- SNOMED Clinical Terms

◆ Content
- 5,570 ingredients
- 10,788 brand names
- 22,724 clinical drug comp.
- 29,734 clinical drugs
- 17,149 branded drugs
- 16,447 branded drug comp.
- 13,516 clinical drug forms
- 13,035 branded drug forms
- 140 dose forms

(as of February 28, 2006)
Relations among drug entities
Retrieved "Zoloft" for String "zoloft"
Retrieved "Sertraline 20 MG/ML Oral Solution".
Branded Drug:
- Acetate 0.141 MEQ/ML / Amino Acids 85 MG/ML / Chloride Ion 0.07 MEQ/ML / inorganic phosphate 0.03 MMOL/ML / Magnesium 0.01 MEQ/ML / Potassium 0.06 MEQ/ML / Sodium 0.95 MEQ/ML / Water 1.00 UNT/ML
- Acetate 0.142 MEQ/ML / Amino Acids 85 MG/ML / Chloride Ion 0.038 MEQ/ML / inorganic phosphate 0.03 MMOL/ML / Magnesium 0.01 MEQ/ML / Potassium 0.066 MEQ/ML / Sodium 0.95 MEQ/ML / Water 1.00 UNT/ML
- Acetate 0.148 MEQ/ML / Amino Acids 100 MG/ML / Potassium 0.0054 MEQ/ML Injectable Solution [Aminosyn]
- Gentamicins 0.003 MG/MG / prednisolone 0.006 MG/MG Ophthalmic Ointment [Pred-G S.O.P.]
- Gentamicins 3 MG/ML / prednisolone 10 MG/ML Ophthalmic Suspension [Pred-G]
- Neomycin 3.5 MG/ML / Polymyxin E 10000 UNT/ML / prednisolone 5 MG/ML Ophthalmic Suspension [Poly Pred]
- prednisolone 0.002 MG/MG / Sulfacetamide 0.1 MG/MG Ophthalmic Ointment [Blephamide S.O.P.]
- prednisolone 0.0025 MG/MG / Sulfacetamide 0.1 MG/MG Ophthalmic Ointment [Cetapred]
- prednisolone 0.005 MG/MG / Sulfacetamide 0.1 MG/MG Ophthalmic Ointment [Predsulfair]
- prednisolone 0.005 MG/MG / Sulfacetamide 0.1 MG/MG Ophthalmic Ointment [Vasocidin]
- prednisolone 0.005 MG/MG / Sulfacetamide Sodium 0.1 MG/MG Ophthalmic Ointment [Medasulf]
- prednisolone 0.005 MG/MG / Sulfacetamide Sodium 0.1 MG/MG Ophthalmic Ointment [Metimyd]
- prednisolone 1 MG/ML Oral Solution [Bubbli-Pred]

Clinical Drug Form:
- Acetate / Amino Acids / Chloride Ion / Magnesium / Potassium
- Acetate / Amino Acids / Chloride Ion / Sodium
- Acetate / Magnesium / Potassium / Sodium

Dose Form:
- Disintegrating Tablet
- Enema
- Enteric Coated Tablet

Branded Drug Form:
- Acetate / Amino Acids / Chloride Ion / Magnesium / Potassium
- Acetate / Amino Acids / Chloride Ion / Sodium
- Acetate / Magnesium / Potassium / Sodium

Retrieval Status or Detailed View of an RxNorm Entry (RXCUI | UMLSCUI | STR)
208993|C0710662|prednisolone 0.005 MG/MG / Sulfacetamide Sodium 0.1 MG/MG Ophthalmic Ointment [Metimyd]
Medical
Ontology
Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

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