



AMIA Symposium
Washington, DC
November 12, 2003

NLP for Biomedical Applications

Information integration through terminology integration



Olivier Bodenreider

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

Introduction

- ◆ NLP and text mining require
 - Terminology
 - Domain knowledge
- ◆ Biomedical terminologies
 - Usually provide vocabulary
 - May provide some domain knowledge
 - Enable semantic integration
- ◆ Semantic integration may benefit NLP by enabling links to external resources



Terminology integration

The Unified Medical Language System

Unified Medical Language System

- ◆ Started in 1986
- ◆ National Library of Medicine

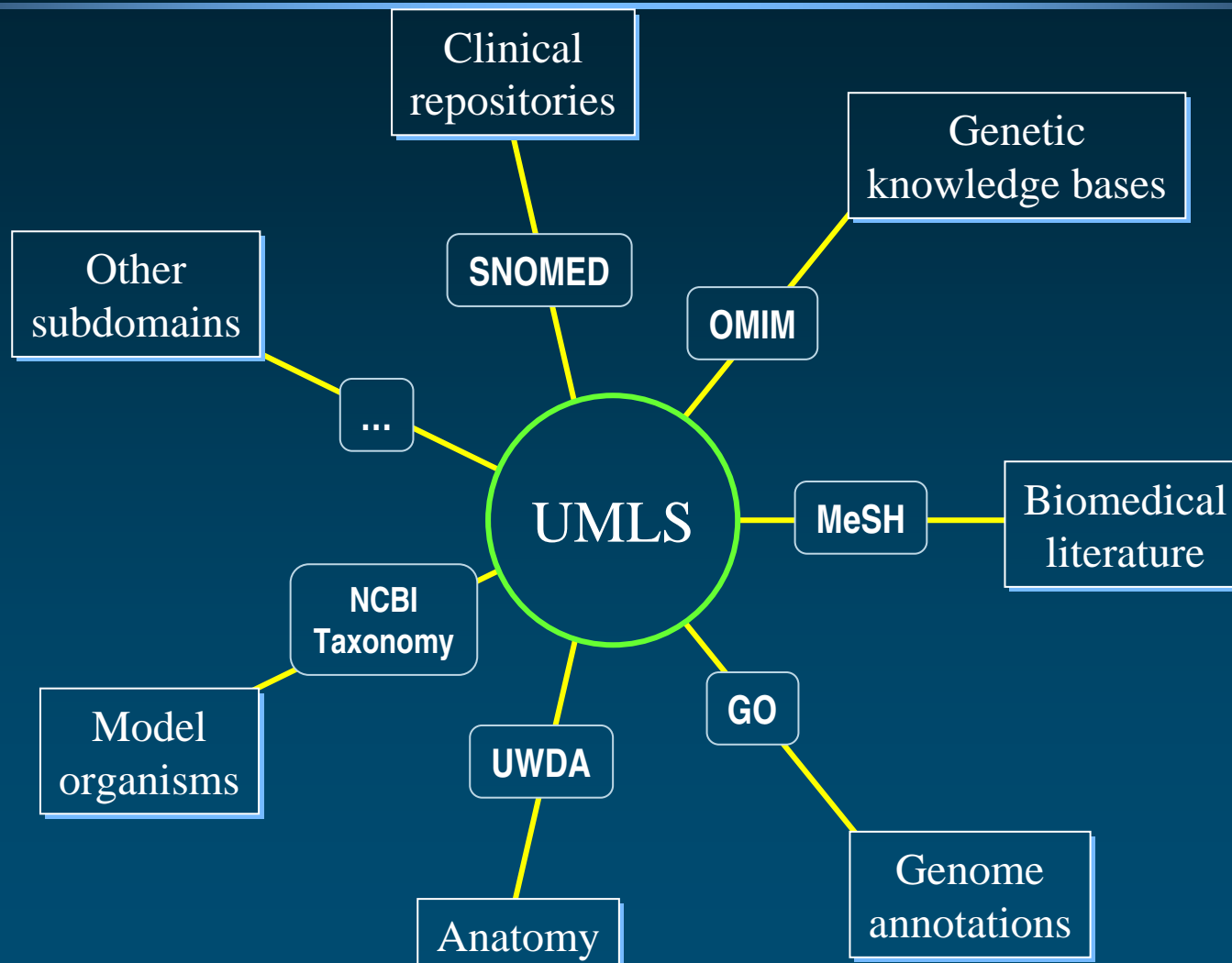
«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»

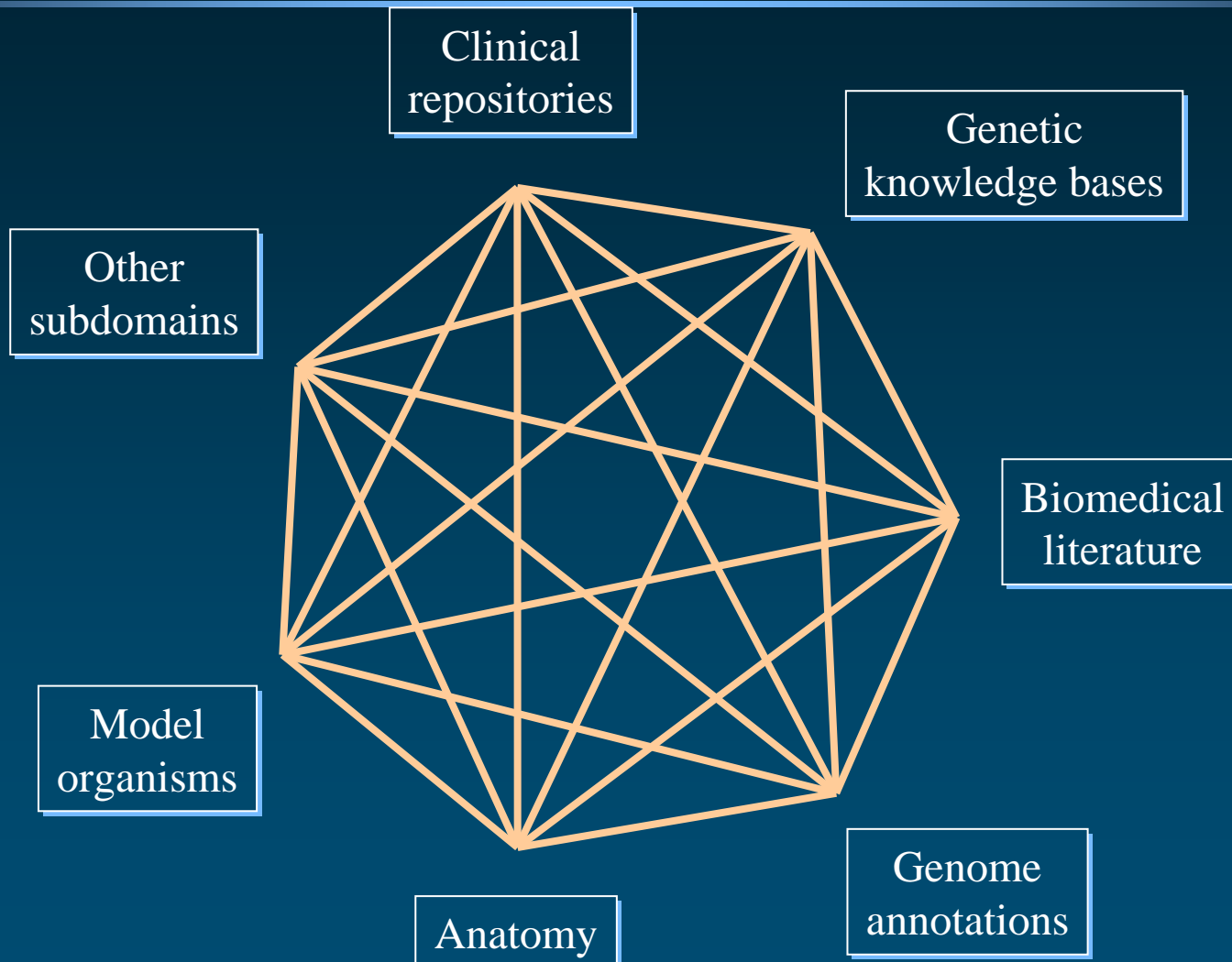
- ◆ Terminology integration
 - 60 families of biomedical vocabularies



Integrating subdomains



Integrating subdomains



Information integration

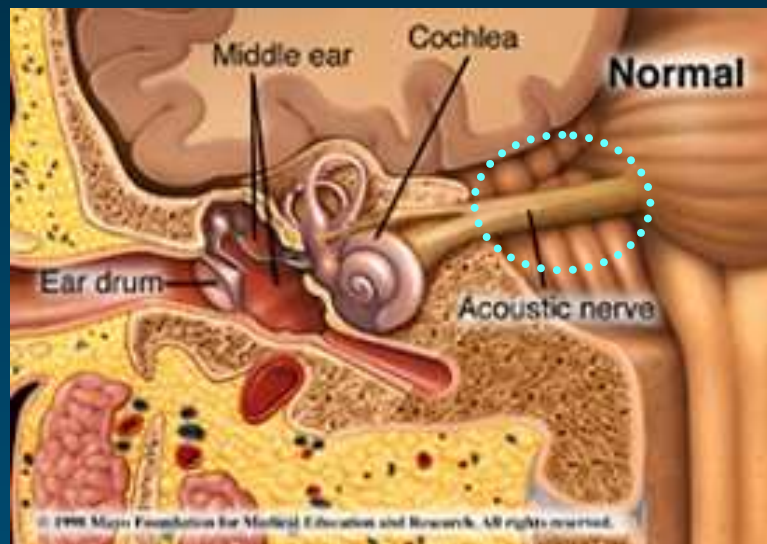
Genetics as an example

NF2 Gene, protein, and disease

Neurofibromatosis 2 is an autosomal dominant disease characterized by tumors called **schwannomas** involving the acoustic nerve, as well as other features. The disorder is caused by mutations of the **NF2** gene resulting in absence or inactivation of the protein product. The protein product of NF2 is commonly called **merlin** (but also **neurofibromin 2** and **schwannomin**) and functions as a tumor suppressor.



Schwannoma (acoustic neuroma)



<http://www.mayoclinic.com>



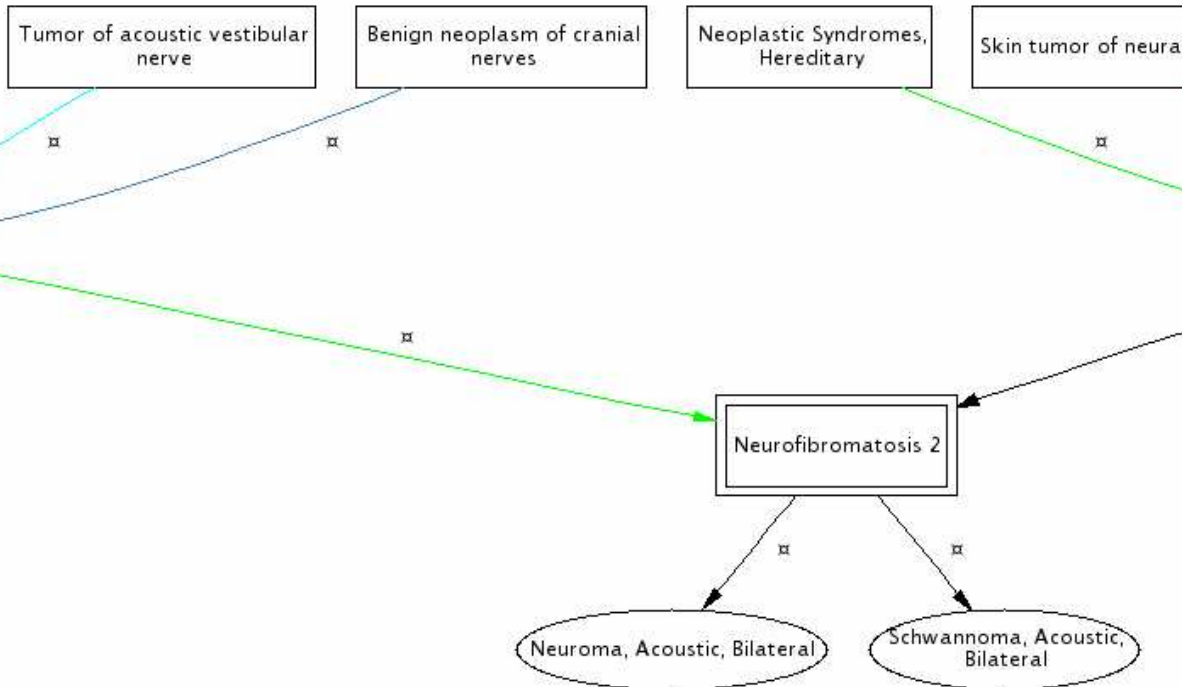
Siblings

Disorders

- Cerebellopontine Angle Acoustic Neuroma
- Diffuse neurofibroma
- Melanocytic Vestibular Schwannoma
- Neurofibromatosis (nonmalignant)
- Neurofibromatosis 1
- neurofibromatosis 1 and 2 (NF1 and NF2)
- Neurofibromatosis 3
- Neurofibromatosis type 3
- NEUROFIBROMATOSIS TYPE IV, OF RICCARDI
- Neuroma, Acoustic, Unilateral
- Segmental neurofibromatosis

(11 siblings)

[direct children and narrower concepts of direct parents and broader concepts]



Other Related Concepts

Anatomy

- Acoustic Nerve

Chemicals & Drugs

- Neurofibromin 2

Disorders

- Familial Acoustic Neuromas
- Neoplasm of uncertain behavior NOS
- Neurofibromatoses
- Neurofibromatosis

- Nerve Sheath Tumors [4]
- Nervous System Neoplasms [6]
- Neurilemmoma [35]
- Neurofibromatosis 1 [38]
- Neuroma, Acoustic [26]
- Peripheral Nervous System Diseases [3]
- Peripheral Nervous System Neoplasms [6]
- Postoperative Complications [9]
- Retinal Diseases [6]
- Skin Neoplasms [9]

BCI **Neurofibromatosis 2** **LEGEND ***

Start again Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS_2003

Type of hierarchical rel: All Parent/Child only Broader/Narrower only

Similar Concepts
(none)

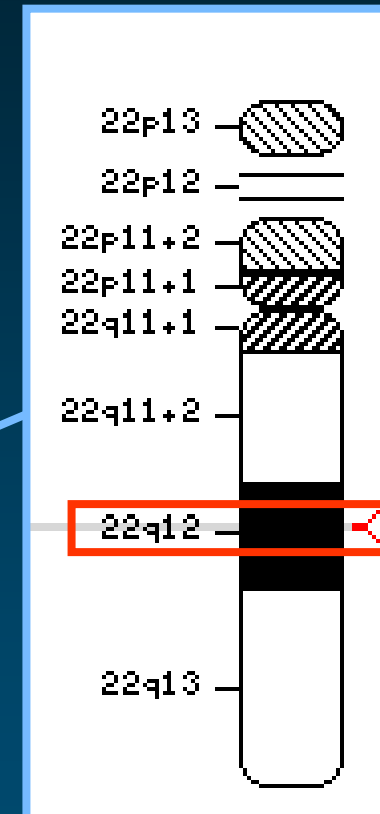
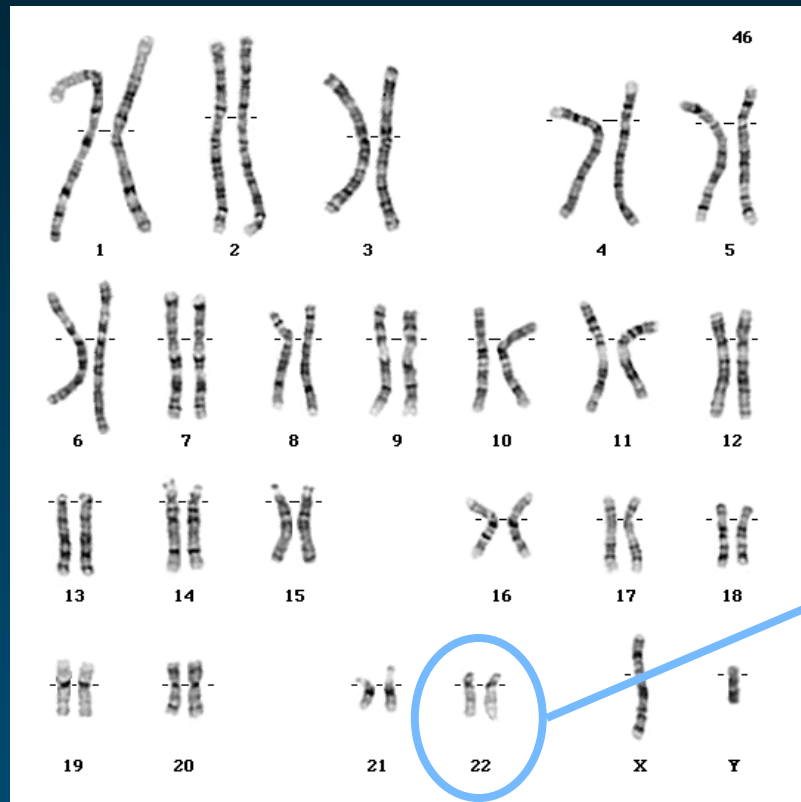
Allegedly Synonyms
• Neurofibromatosis (synonym)

Closest MeSH Terms

Main Headings
• Neurofibromatosis 2

Subheadings

NF2 gene



<http://staff.washington.edu/timk/cyto/human/>

<http://www.ncbi.nlm.nih.gov/mapview/>



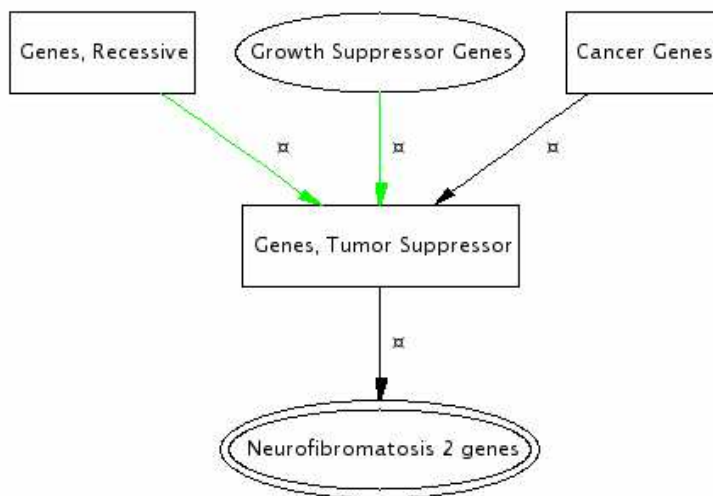
Siblings

Chemicals & Drugs

- ADAM11 protein, human ☒
- DLG5 protein, human ☒
- DPM3 protein, human ☒
- HCCS-1 protein, human ☒
- hssh3bp1 protein, human ☒
- HUGL protein, human ☒
- LAPSER1 protein, human ☒
- mitochondria proteolipid-like protein, human ☒
- MRG protein, human ☒
- p53 gene/protein ☒
- PLAGL1 protein, human ☒
- RARRES3 protein, human ☒
- SEZ6L protein, human ☒
- TES protein, human ☒

Genes & Molecular Sequences

- APC Gene ☒
- BAX Gene ☒
- brca gene ☒
- CDH1 gene ☒
- CHES1 Gene ☒
- cyclin-dependent kinase inhibitor 2A



Other Related Concepts

Chemicals & Drugs

- Neurofibromin 2 ☒

Disorders

- Neurofibromatosis 2 ☒

(2 other related concepts)

- Chromosome Deletion [7] ☒
- Ependymoma [4] ☒
- Glioma [4] ☒
- Loss of Heterozygosity [7] ☒
- Meningeal Neoplasms [25] ☒
- Meningioma [30] ☒
- mesothelioma <1> [4] ☒
- Neoplasms [4] ☒
- Neurilemmoma [20] ☒
- Neurofibromatoses
- Neurofibromatosis 2 [64] ☒
- Neuroma, Acoustic [5] ☒
- Spinal Cord Neoplasms [3] ☒

BCI

Neurofibromatosis 2 genes

LEGEND *

Start again Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS_2003

Type of hierarchical rel: All Parent/Child only

Broader/Narrower only

Similar Concepts

(none)

Allegedly Synonyms

(none)

Closest MeSH Terms

Main Headings

- Genes, Neurofibromatosis 2

Subheadings

Merlin

◆ Synonyms

- Neurofibromin 2
- Schwannomin
- Schwannomerlin
- Neurofibromatosis-2

◆ 10 isoforms

◆ Annotations

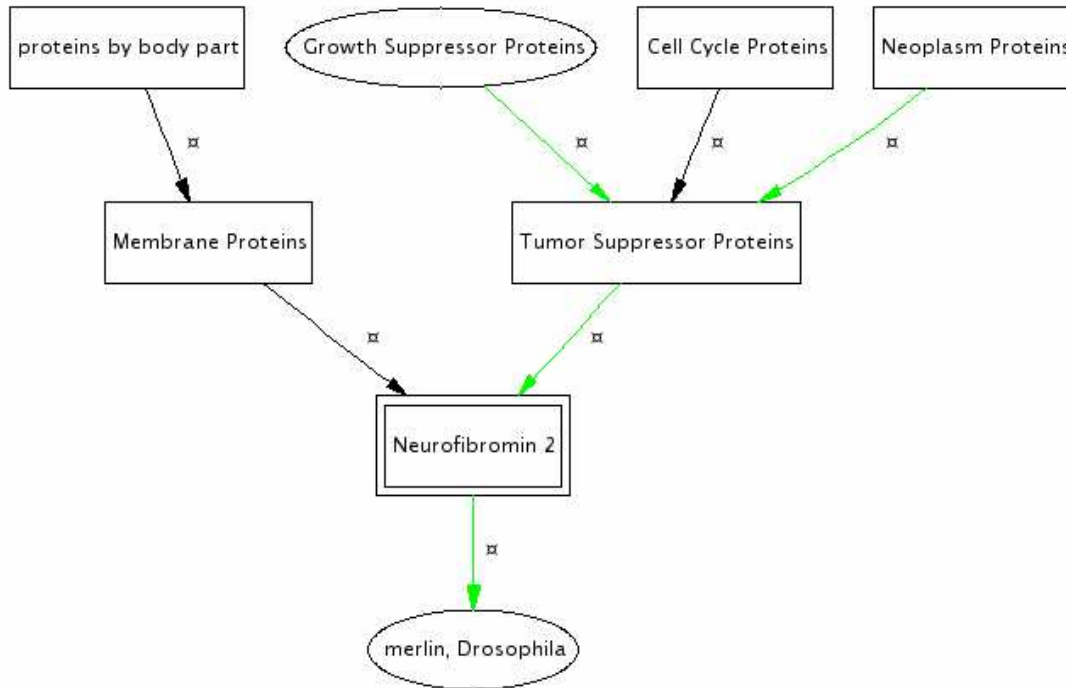
- Negative regulation of cell proliferation
- Cytoskeleton
- Plasma membrane



Siblings

Chemicals & Drugs

- (LA)12 peptide ☒
- (methyl)ammonium uptake carrier, Corynebacterium ☒
- 120-kDa hemocyte-specific membrane protein, flesh fly ☒
- 15a protein, Aedes aegypti ☒
- 22.6-kDa antigen, Schistosoma japonicum ☒
- 36-kDa vesicular integral membrane protein ☒
- 38L protein ☒
- 5-lipoxygenase-activated protein ☒
- 59 kDa dystrophin-associated protein ☒
- A-1 antigen ☒
- A-kinase anchor protein 149 ☒
- A-kinase anchor protein 15 ☒
- A-kinase anchor protein 200 ☒
- A-kinase anchor protein KL ☒
- A14.5L protein ☒
- A15 protein ☒
- ABC-me protein ☒
- ABU-1 protein, C elegans ☒
- AcfB protein ☒
- ACR3 protein ☒



Other Related Concepts

Disorders

- Neurofibromatosis 2 ☒

Genes & Molecular Sequences

- Neurofibromatosis 2 genes ☒

(2 other related concepts)

Co-occurring Concepts

Anatomy

- Arachnoid [1] ☒
- Cell Membrane [1] ☒
- Cerebellum [1] ☒
- Chromosomes, Human, Pair 22 [1] ☒
- Cytoplasm [1] ☒
- Cytoskeleton [2] ☒
- Microfilaments [1] ☒
- Purkinje Cells [1] ☒
- Schwann Cells [1] ☒
- Stem Cells [1] ☒

BCI

Neurofibromin 2

LEGEND *

Start again

Apply new parameters

Restrict to vocabulary:

Show all

Highlight vocabulary:

Nothing

UMLS data:

UMLS_2003

Type of hierarchical rel:

All Parent/Child only

Broader/Narrower only

Similar Concepts

(none)

Allegedly Synonyms

(none)

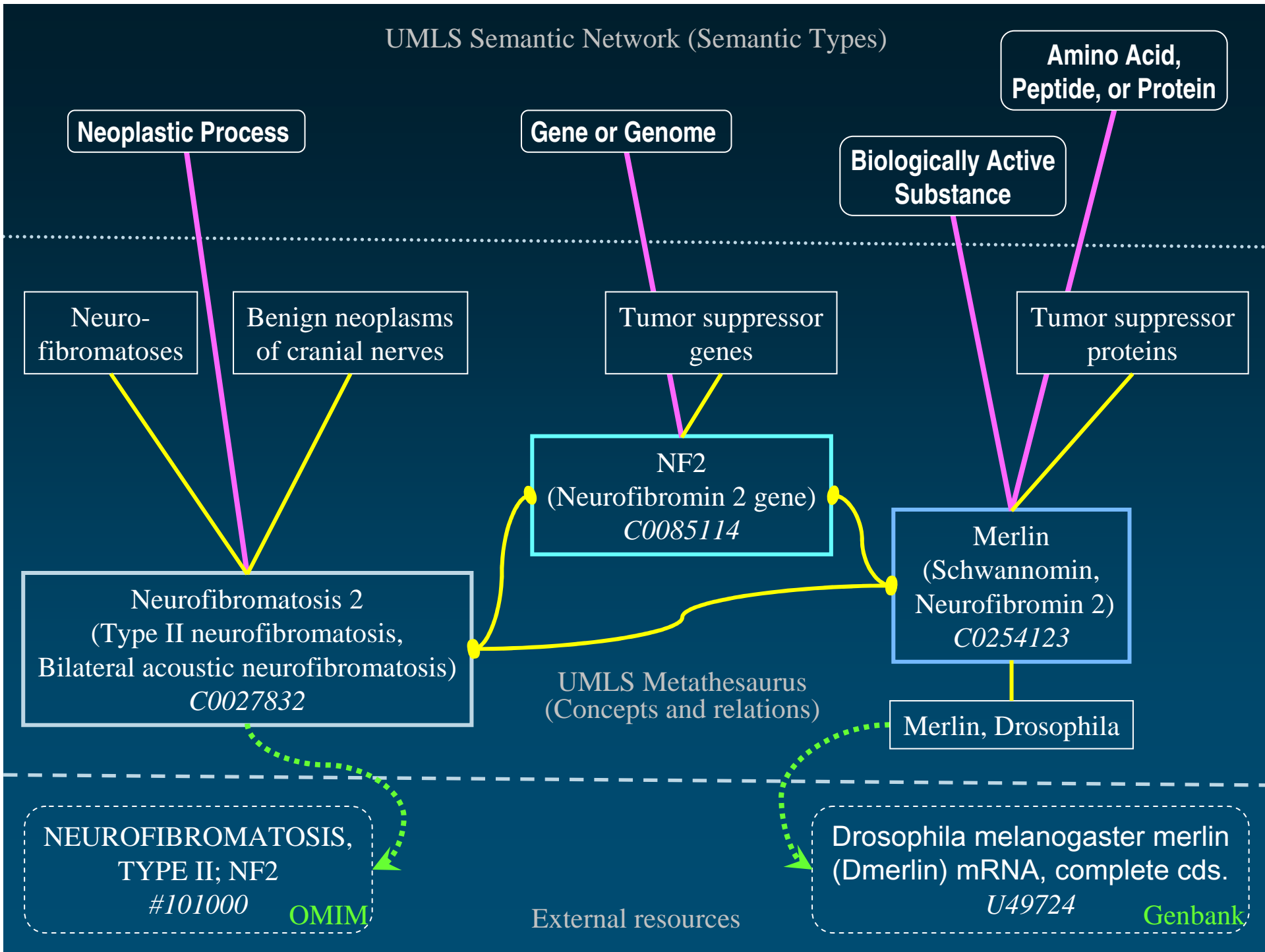
Closest MeSH Terms

Main Headings

- Neurofibromin 2

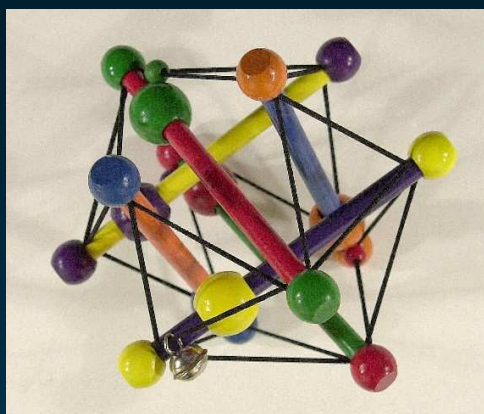
Subheadings

UMLS Semantic Network (Semantic Types)



Limitations

- ◆ Genes not systematically represented
 - Most gene products and diseases are
- ◆ Gene/Gene product-Disease relations
 - Not systematically represented
 - Not explicitly represented (e.g., co-occurrence)
- ◆ Cross-references not systematically represented
- ◆ Naming conventions (genes)



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