

Towards an Ontology for Representing Malignant Neoplasms

William Duncan, Carmelo Gaudioso
and Alexander Diehl



**ROSWELL
PARK**
CANCER INSTITUTE



Objectives

- Review of problems encountered with current classification systems
 - ICD codes, NCI Thesaurus, SNOMED CT
- How to address these issues
 - On carcinomas and other pathological entities
 - Summarize our work
- Conclusions / unaddressed needs

Background & Motivation

- Roswell Park Cancer Institute
 - Founded in 1898 as the nation's first cancer center
 - Active research in immunotherapies, cancer genetics and genomics, cancer cell biology, clinical trials
- Many Databases
 - EMR, Cancer Registry, Lab data, Pathology data (organ, cell, molecule), Treatment data, Outcome data etc.
- Query and analyze data across them
 - Federated queries
 - Term enrichment
 - Ovarian cancers

Review: ICD systems (ICD-9, ICD-10, ICD-O)

- International Statistical Classification of Diseases
 - Maintained by the World Health Organization (WHO)
 - Hierarchically structured codes that represent diseases, disorders, and other health related issues
- ICD-O – C56 8140/3 (ovarian adenocarcinoma)
 - C56 – the site code for an ovary
 - 8140/3
 - 8140 – neoplasm arising from glandular epithelial tissue (morphology code)
 - '/3' – neoplasm is malignant (behavior code)

Review: ICD systems (ICD-9, ICD-10, ICD-O)

- Advantages
 - Allows diseases to be easily grouped and counted for statistical and reporting purposes.
 - ‘8140’ classifies adenocarcinoma patients
- Disadvantages
 - Missing information for treatments, outcomes, molecular disorders, etc.
 - Lacks formal relations

Review: NCI Thesaurus

- Contains over 100,000 concepts with textual definitions and 400,000 cross-links between its concepts
- Malignant neoplasms branch seems adequate
 - Ovarian Adenocarcinoma (C770): An adenocarcinoma that arises from the ovary.
 - Parents: Adenocarcinoma, Ovarian Carcinoma
- Adenocarcinoma (C2852): A common cancer characterized by the presence of malignant glandular cells.
- Ovarian Carcinoma (C4908): A malignant neoplasm originating from the surface ovarian epithelium.

Review: NCI Thesaurus

- Issues once we get away from neoplasms
- Abnormal Cell (C12913): An abnormal human cell type which can occur in either disease states or disease models.
- Neoplastic Cell (C12922): Cells of, or derived from, a tumor.
- Malignant Cell (C12917): Cells of, or derived from, a malignant tumor.

Review: SNOMED CT

- Advantages
 - Large number of clinical terms
 - Treatments (chemotherapy)
 - Cancer staging (Cancer of ovary, stage 4)
- Disadvantages
 - Tumor cell (SCTID 252987004) is defined as subtype of the concept Abnormal cell (SCTID 39266006), but does not specify if the concept Tumor cell represents malignant cells.
 - Malignant tumor cells (SCTID: 88400008) is defined as being a subtype of the concept Malignant neoplasm, primary (SCTID: 86049000).

Review: Disease Ontology

- Advantages
 - OBO Foundry ontology
 - Adequate coverage of cancer types
- Disadvantages
 - Inconsistent use of terms ‘cancer’ and ‘neoplasm’
 - ovary neuroendocrine neoplasm is an ovarian cancer
 - Does not relate cancer to anatomical entities
 - Does not relate ovary epithelial cancer to ovaries
 - Inhibits queries: What are most common anatomical structures in which malignant cells are found?

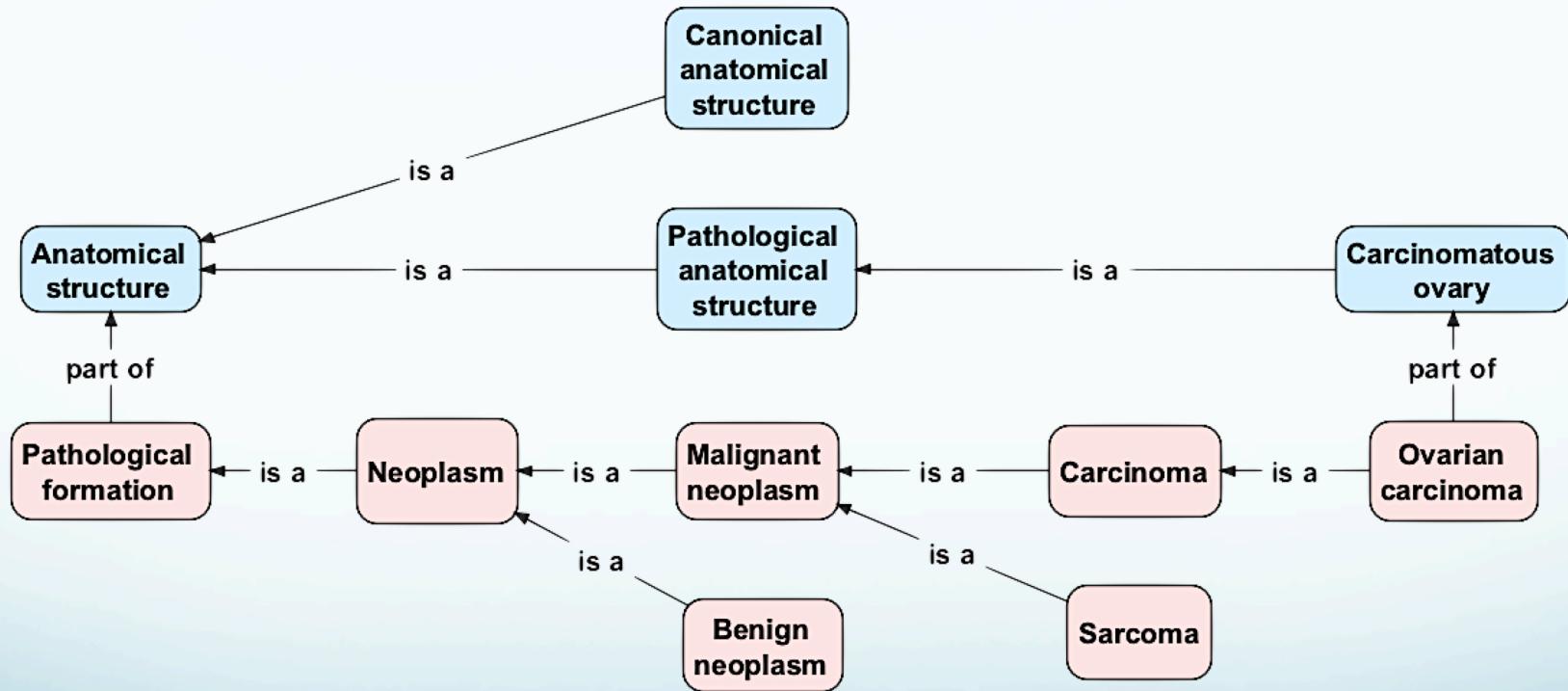
On carcinomas and other pathological entities

- Goal – Expand Foundational Model of Anatomy (FMA) to include pathological material entities:
 - pathological anatomical structures (malignant neoplasm)
 - pathological bodily substances (portion of pus)
- Why?
 - In the FMA, an anatomical structure is canonical
 - ‘Canonical’ means generated by the coordinated expression of the organisms own genes
 - Problem: not all neoplasms are canonical (HPV-related squamous intraepithelial neoplasia)

On carcinomas and other pathological entities

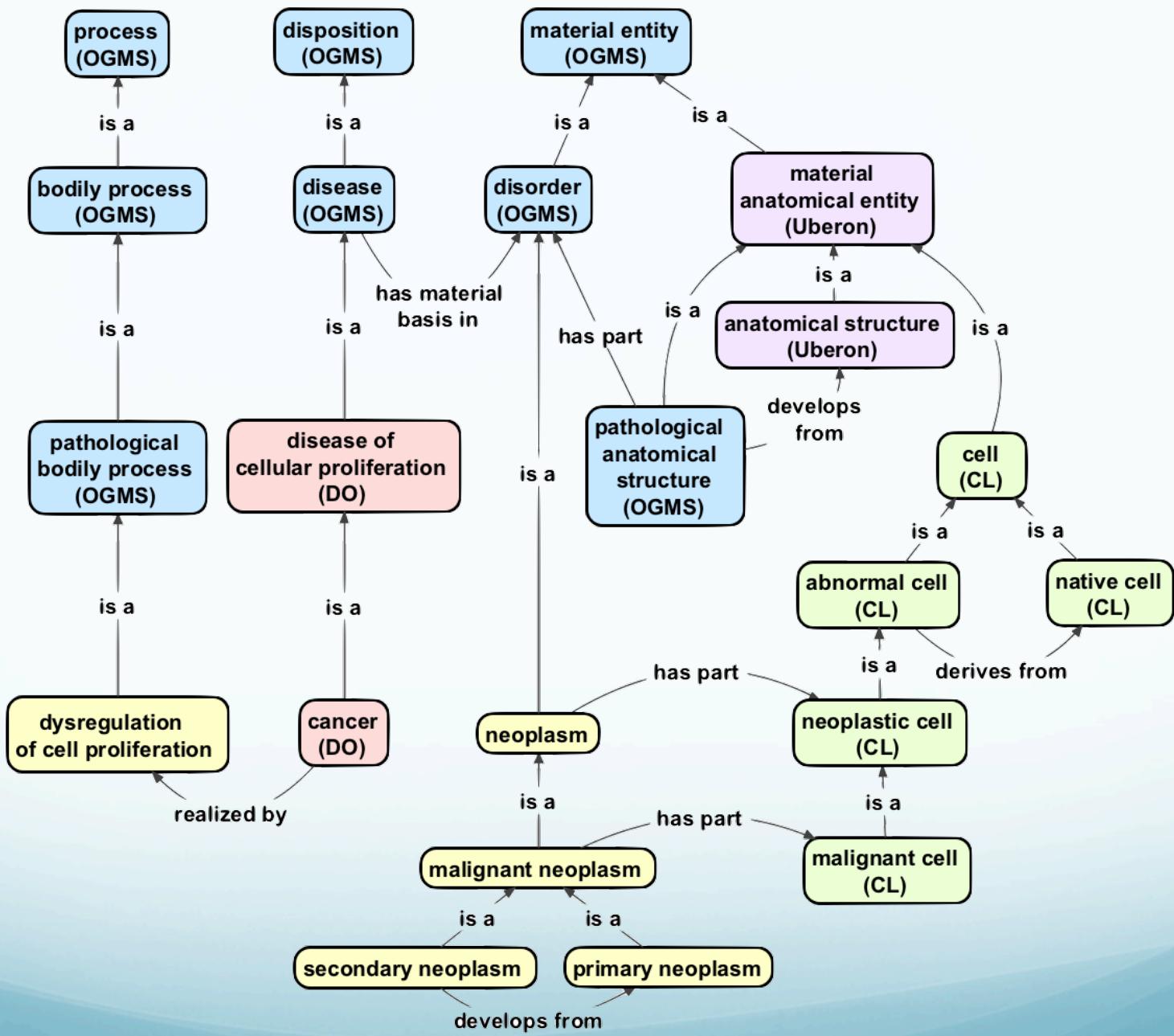
- An entity is **pathological** if:
 - Results from processes other than the expression of the normal complement of genes of an organism of the given type
 - Predisposed to have health-related consequences for the organism in question
- **Pathological formation:** A pathological entity evolving in some larger anatomical structure
- **Pathological anatomical structure:** An anatomical structure that has part some pathological formation

On carcinomas and other pathological entities



Our proposal

- Smith et al. (2005) predates the OBO Foundry
- Extend Smith et al. to use existing OBO Foundry ontologies
 - Cell Ontology (CL)
 - Ontology for General Medical Sciences (OGMS)
 - Disease Ontology (DO)
 - Uberon



Our proposal: CL

Imported classes

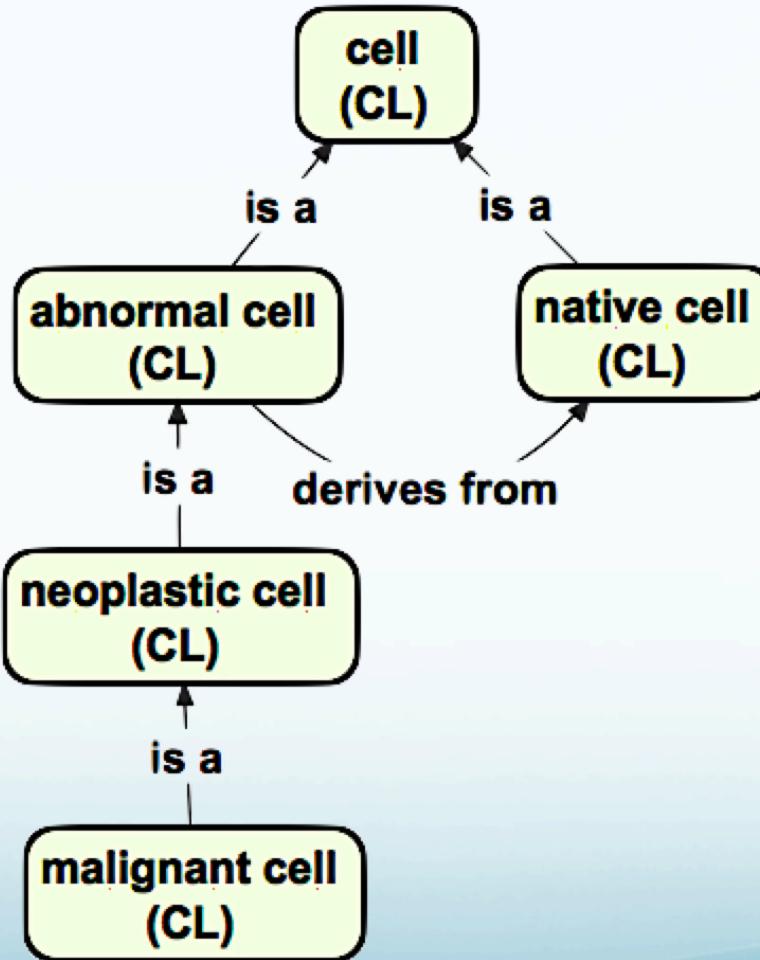
- **abnormal cell**: A cell found in an organism or derived from an organism exhibiting a phenotype that deviates from the expected phenotype of any native cell type of that organism
- **neoplastic cell**: An abnormal cell exhibiting dysregulation of cell proliferation or programmed cell death and capable of forming a neoplasm
- **malignant cell**: A neoplastic cell that is capable of entering a surrounding tissue

Our proposal: CL

Imported relations

- x **derives from** y : holds between an instance x , an instance y , and times t, t' and is such that:
 - changes in y at t result in a new entity x at t'
 - y ceases to exist when x is created
 - x inherits a significant portion of its matter from y
- abnormal cell **derives from** native cell
- x **develops from** y : similar to x **derives from** y , but without the criterion that y ceases to exist
 - secondary neoplasm **develops from** primary neoplasm

Our proposal: CL



Our proposal: OGMS & DO

Dispositional account of disease

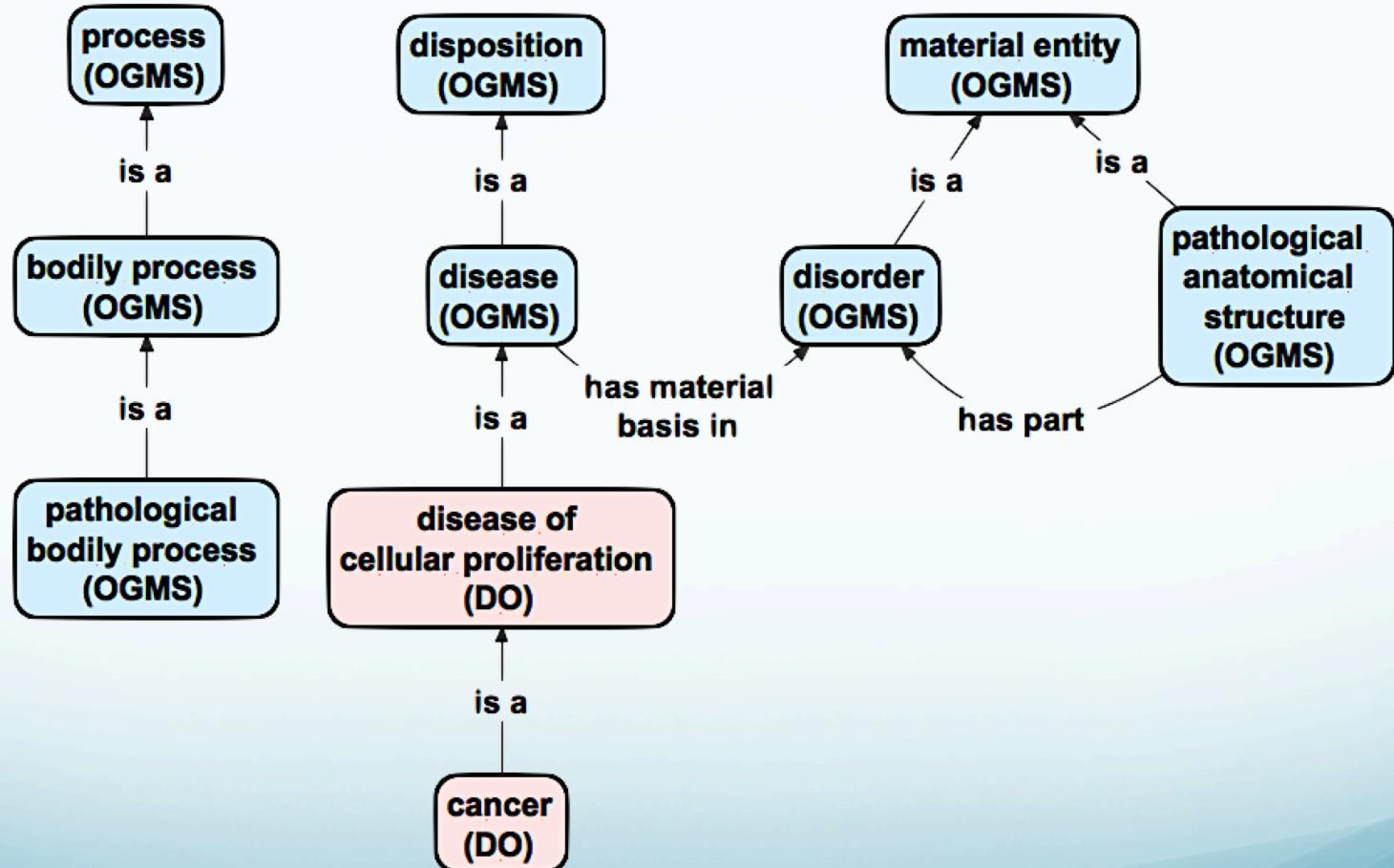
- **disease:** A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism
- **disease of cellular proliferation:** A disease that is characterized by abnormally rapid cell division
- **cancer:** A disease of cellular proliferation that is malignant and primary, characterized by uncontrolled cellular proliferation, local cell invasion and metastasis

Our proposal: OGMS & DO

Relate diseases to material entities

- **disorder:** A material entity which is clinically abnormal and part of an extended organism
- **x has material basis in y:** holds between an instance of a disease x and an instance of a disorder y at particular time t in which x exists because of the physical makeup of some part of y at time t
- disease **has material basis in** disorder
 - cancer **has material basis in** disorder that is **part of** a malignant cell

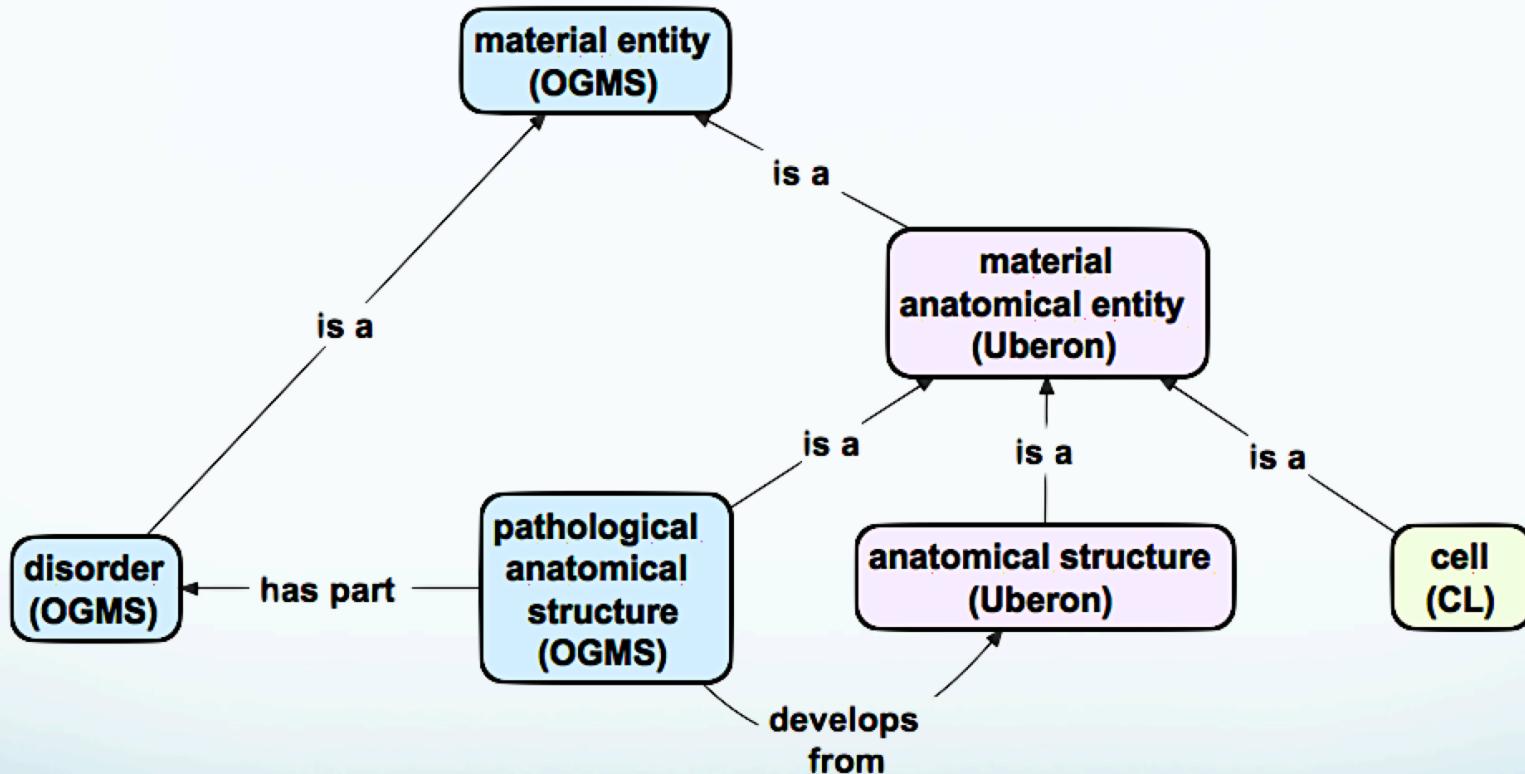
Our proposal: OGMS & DO



Our proposal: Uberon

- **material anatomical entity:** An anatomical entity that has mass
- **anatomical structure:** A material anatomical entity that is a single connected structure with inherent 3D shape generated by coordinated expression of the organism's own genome
- **pathological anatomical structure:** A material entity that comes into being as a result of changes in some pre-existing anatomical structure through processes other than the expression of the normal complement of genes of an organism of the given type, and is predisposed to have health-related consequences for the organism in question

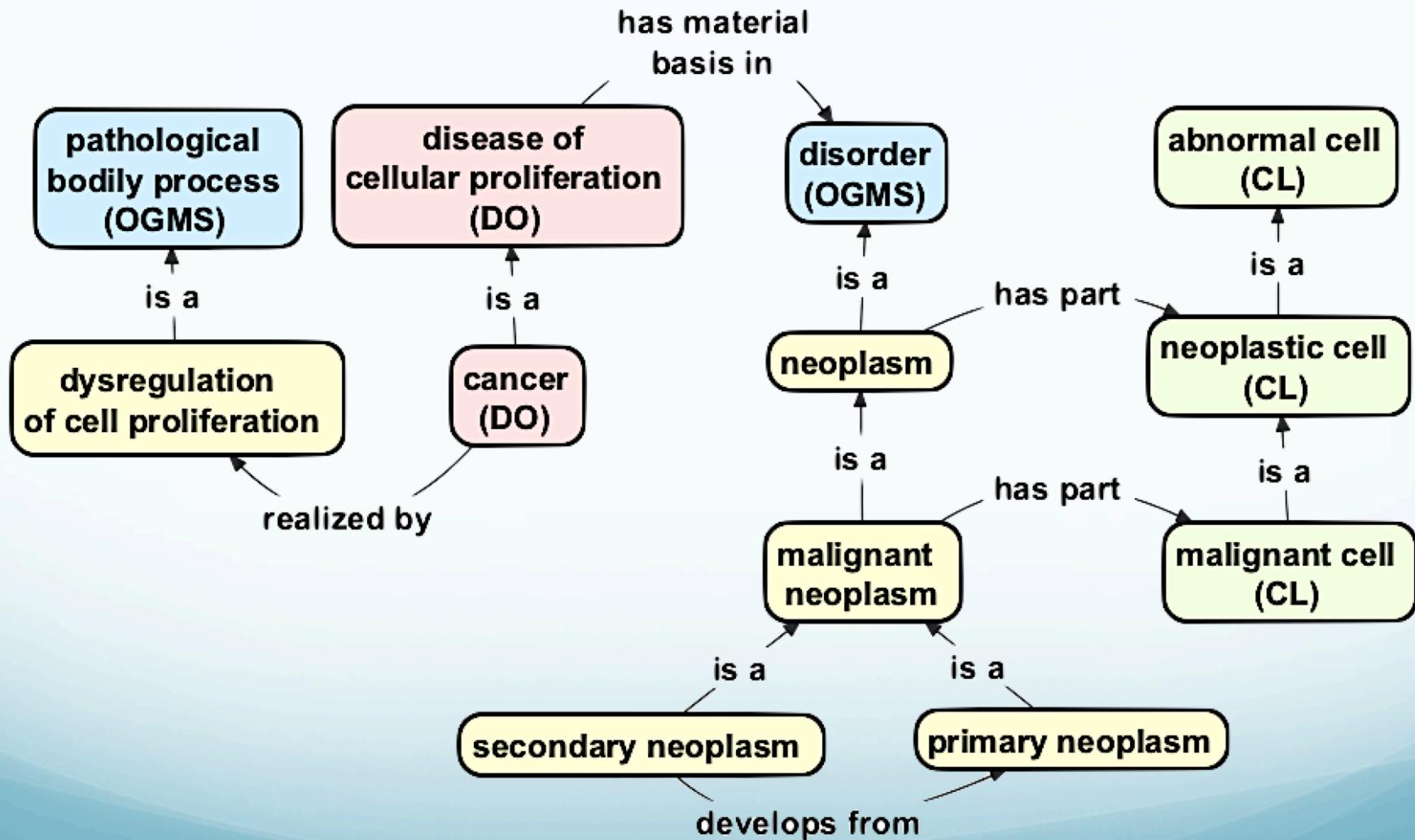
Our proposal: Uberon



Our proposal: New Entities

- **dysregulation of cell proliferation:** A pathological bodily process that realizes uncontrolled cell proliferation
- **neoplasm:** A disorder that results from uncontrolled cell proliferation
- **malignant neoplasm:** A neoplasm that has acquired the disposition to invade surrounding tissues and spread to remote anatomical sites
- **primary neoplasm:** A malignant neoplasm that is found in the site where the malignant cells first began proliferating
- **secondary neoplasm:** A malignant neoplasm that develops from a *primary neoplasm*

Our proposal: New Entities



Conclusions / Needs

- NCI Thesaurus was an excellent resource for finding definitions of malignant neoplasms, but:
 - Encountered issues in other branches
 - abnormal, neoplastic, and malignant cell
 - Upper-level hierarchy needs work
 - Disease, Disorder, or Finding (C7057)
 - Disease or Disorder (C2991)
- Disease Ontology is a good resource for cancer types but:
 - Inconsistently uses the terms ‘cancer’ and ‘neoplasm’
 - Needs to be linked to other ontologies

Conclusions / Needs

- Need well defined hierarchy for pathological entities
- Smith et al. outline the following structure:
 - Anatomical structure
 - Canonical anatomical structure
 - Pathological anatomical structure
- But a definition is not provided for Anatomical structure
- Similar problem for pathological bodily processes

Conclusions / Needs

Much more work to be done!

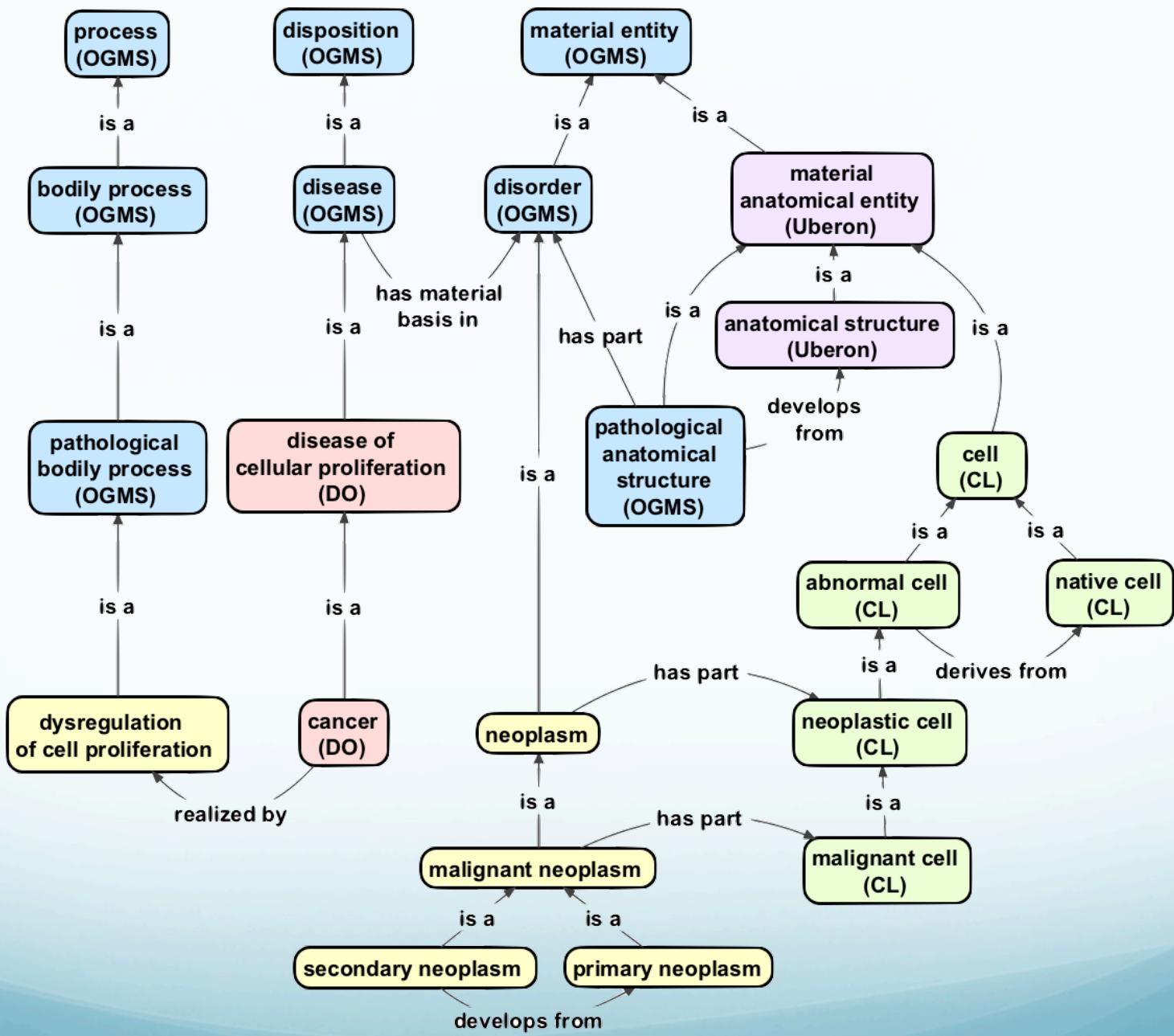
- Cancer staging
- Laboratory methods and results
 - Immunohistochemistry
 - Flow cytometry
- Treatments and outcomes
 - Chemotherapy
 - Radiation therapy
 - Immunotherapy
- Molecular tests
 - DNA sequencing
 - RNA sequencing

Financial Support

- NCI P30CA16056 – Roswell Park Clinical Data Network Shared Resource
 - William Duncan and Carmelo Gaudioso
- NCATS 5UL1TR001412
 - Alexander Diehl
- NCI P50CA159981 – Ovarian Supplemental Grant
 - William Duncan, Carmelo Gaudioso, and Alexander Diehl

Thank you for attending!

Questions?



References

Diehl AD et al. (2016). The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *Journal of Biomedical Semantics*, 7(44). PMCID:PMC4932724.

<https://github.com/obophenotype/cell-ontology>.

Mungall CJ et al. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology*, 13(1), R5. PMCID:PMC3334586.

<http://uberon.github.io>.

Scheuermann RH et al. (2009). Toward an ontological treatment of disease and diagnosis. San Francisco: *Proceedings of the 2009 AMIA Summit on Translational Bioinformatics*, 116–120. PMCID:PMC3041577.

<https://github.com/OGMS/ogms>.

Schrimi LM & Mitraka E (2015). The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mammalian Genome*, 26(9-10): 584–589. PMCID:PMC4602048.

<http://disease-ontology.org>.

Smith B et al. (2005). On Carcinomas and Other Pathological Entities. *Comparative and Functional Genomics*, 6(7-8): 379-387. PMCID:PMC2447494.