Standardization of the Histopathology Cancer Report: An Ontological Approach

Anna Maria Masci, Shannon McCall, Alessandro Racioppi,
Helena Judge Ellis, Jihad S. Obeid, Barry Smith,
Christian Stoeckert and Jie Zheng

August 1, 2018

Standardization of the Histopathology Cancer Report: An Ontological Approach

Anna Maria Masci[1], Shannon McCall[2], Alessandro Racioppi[3], Helena Judge Ellis[4] , Jihad S. Obeid[5], Barry Smith[6] , Christian Stoeckert[7], Jie Zheng[7]
1 Department of Biostatistics and Bioinformatics, Duke University. Durham, NC annamaria.masci@duke.edu
2 Department of Pathology, Duke University Health System, Durham, NC
3 University of North Carolina, Chapel Hill, NC
4 Biobanking Without Borders, LLC, Durham, NC
5 Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC
6 Department of Philosophy, University of Buffalo NY
7 Department of Genetics, University of Pennsylvania, Philadelphia, PA

Extended Abstract

Keywords (ontology, histopathology report, histopathology, cancer, tumor, cancer pathology, pathology standards)

Background

In recent years**,** the complexity of cancer pathology reporting has increased significantly. The pathology report covers not only general information such as the presence or absence of cancer, but includes a collection of specific parameters such as tumor size, grade, margin, lymphatic or vascular involvement as well as molecular testing e.g. proteomics and genomics (Figure 1). Soon, biomarkers and immune profiling will play an increasingly more important role in determining the eligibility for particular therapies, along with genetic predisposition and social risk factors. The increased use of digital pathology, which allows streamlined sharing of images, has highlighted the importance of clear communication of the information displayed in the pathology report. In the past years**,** significant effort has been devoted to redefining the way that histopathology report information is recorded. The College of American Pathologists (CAP) (http://www.cap.org/), a leading organization of board-certified pathologists, introduc***ed*** synoptic cancer reports**,** a structured checklist to standardize clinical documentation. Despite continuous improvement and generation of electronic reports, formal representation [1] is still lacking. This lack of standardization limits the ability to integrate pathology information with other genomic and proteomic data and often results in loss of information.

Ontology

Standardized, computable representations, in the form of ontologies and structured data, are foundational methods for sharing and integrating data. Ontologies constitute heterogeneous resources combinable for reasoning and hypothesis testing. Ontology as we practice it draws on the example of Gene Ontology (http://www.geneontology.org) [2], which provides a controlled structured vocabulary for the description of biological processes, molecular functions and cellular components. The methodology is standardized through the Open Biomedical Ontologies (OBO) Foundry initiative (http://www.obofoundry.org) [3]. Ontologies that follow this methodology represent complex bodies of knowledge drawing from various sources by using a consistent framework, a feature that simpler network representations lack. Ontologies allow us to use standard W3C (https://www.w3.org) inference tools to reason over the data that are annotated with ontology terms. They allow us to enhance the quality of our knowledge by expanding the power of data retrieval and analysis.

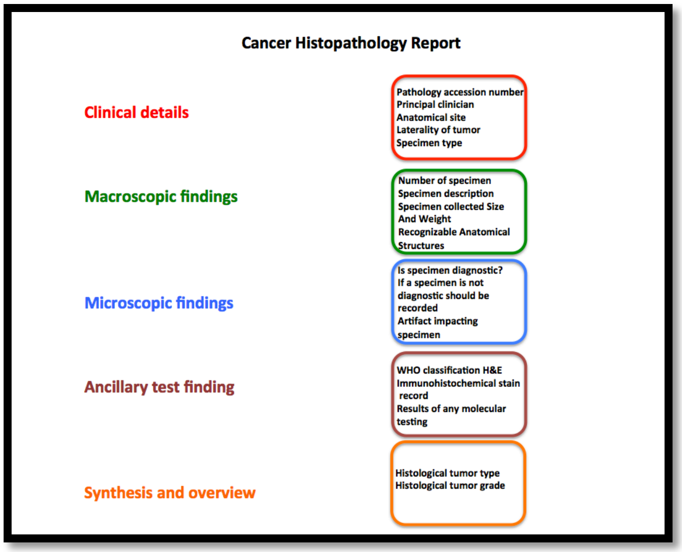

Figure 1: Categories represented in Histopathology Report Ontology

Methods

To address the lack of formal terminology and computational resources critical to improving the use of the histopathology reports, we are developing an ontological representation of the required elements for cancer histopathology. As proof of concept, we are creating a representation of the required terms that are shared by most cancer histopathology reports. We are developing our representation by applying the formalism used in the OBO Foundry ontologies, including the use of upper level BFO as common architecture (http://ifomis.uni-saarland.de/bfo/).

Duke University Department of Biostatistics and Bioinformatics Pilot grant 391-8538

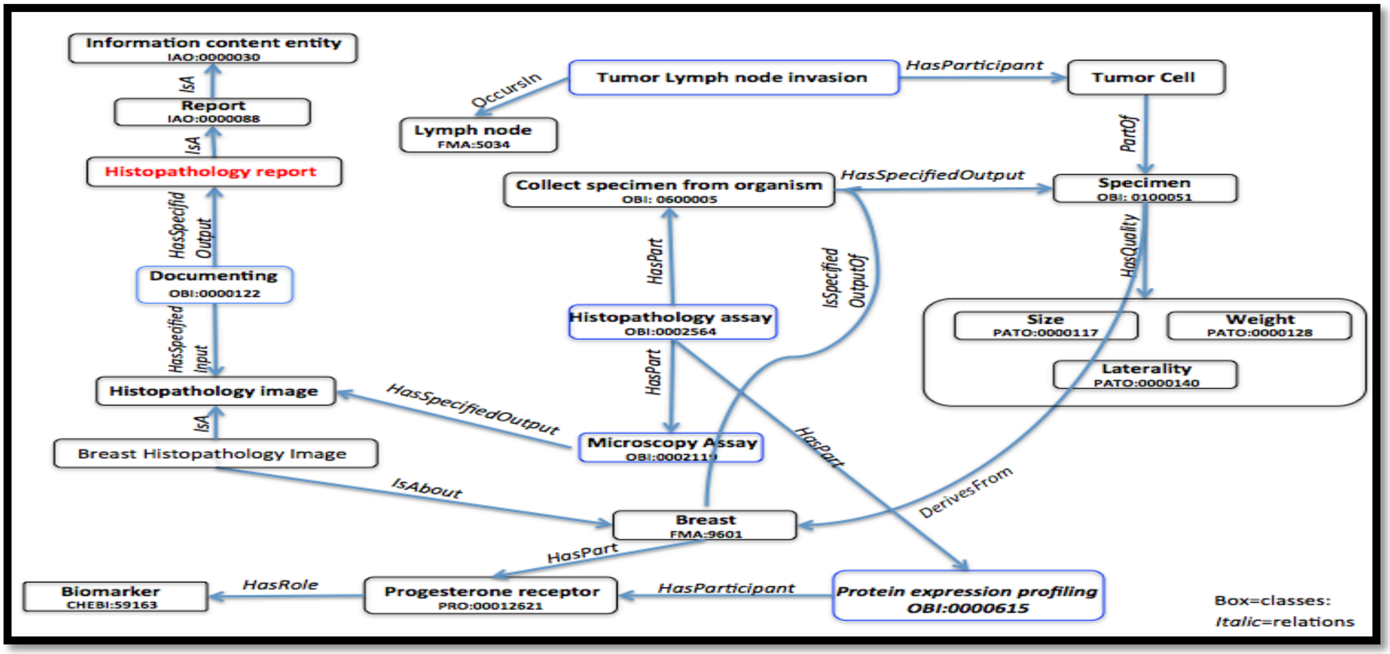

Figure 2; Example of the classes and relations captured in Histopathology Report Ontology

Our development activities are taking place in close collaboration with the developers of several OBO ontologies, notably, Ontology of BioBanking (OBIB) (https://github.com/biobanking/biobanking) [4], Ontology of Biomedical Investigations (OBI) (http://obi-ontology.org) [5], Quantitative Histopathology Image Ontology (QHIO) [6], and Information Artifact Ontology (IAO) https://github.com/information-artifact-ontology/IAO/.

As a first step we identified the ontological classes corresponding to the data elements within the report. We next searched for the terms of interest in other OBO Foundry ontologies to avoid duplication. For the ones that were not already in another OBO Foundry ontology, we created new terms with clear definitions. These new terms will be submitted to the appropriate OBO Foundry ontology through their tracker system for approval. In order to capture the complexity represented in the pathology report we have begun to connect the different classes. The result is a graph representation, where the classes are the nodes and the relations are the edges. To link the different classes we rely on the relations defined in Relation Ontology (RO) [7] an example of the use of relations is shown in Figure 2.

Advantages

We are constructing an ontological network that will allow queries and inferences concerning diverse information in the report. Examples are the following:

- *If the tissue under observation is a breast tissue, a progesterone receptor will automatically be a biomarker.*
- *Any specimen must have size, weight, and laterality.*
- *If lymph node invasion has been reported, one of the specimens has to be a lymph node.*

This prototype for ontological representation of histopathology report is an innovative approach that will facilitate maximal integration of information from various types of cancer reports, leading to a comprehensive picture of all the parameters captured in the report. Implementation of this approach can take different forms from tagging data collected in existing systems to instantiation of the ontological representation in semantic graph databases. Applying this ontological approach will not only improve the annotation capability of histopathology reporting, but additionally would enhance the ability to share information and exponentially increase the power of data retrieval.

References

1. Campbell, W.S., et al., A computable pathology report for precision medicine: extending an observables ontology unifying SNOMED CT and LOINC. J Am Med Inform Assoc, 2017.
2. Ashburner, M., et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000. **25**(1): p. 25-9.
3. Smith, B., et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol, 2007. **25**(11): p. 1251-5.
4. Brochhausen, M., et al., *OBIB-a novel ontology for biobanking.* J Biomed Semantics, 2016. **7**: p. 23.
5. Bandrowski, A., et al., *The Ontology for Biomedical Investigations.* PLoS One, 2016. **11**(4): p. e0154556.
6. Gurcan, M.N., et al., Developing the Quantitative Histopathology Image Ontology (QHIO): A case study using the hot spot detection problem. Journal of biomedical informatics, 2017. **66**: p. 129-135.
7. Smith, B., et al., *Relations in biomedical ontologies.* Genome Biol, 2005. **6**(5): p. R46.