Observables Ontology for Anatomic and Molecular Pathology

James R. Campbell MD
W. Scott Campbell PhD MBA

Nebraska Medicine
University of Nebraska Medical Center
Omaha, NE

Nebraska Medicine
SERIOUS MEDICINE. EXTRAORDINARY CARE.®

Disclosures

We receive no financial support from any interests vending electronic health records or vocabulary services

Portions of this work have been funded by:

NIH 5U01HG009455-02 Deploying ONC national Standards in Support of Metadata for Big Data

Research Warehouse Management

PCORI CDRN-1501-26643 CDRN 04:The Greater Plains Collaborative

NIH 1U54GM115458-0, Subproject ID 5422 Great Plains Idea-Ctr-Biomedical Informatics and Cyberinfrastructure Enhancement Core

Acknowledgements: Contributors and Collaborators

- NM: Jody Peterson, Susan Griffith, Milissa Gerkin, Anne Naberhaus, Donna Weis
- UNMC: Geoffrey Talmon, Audrey Lazenby, Allison Cushman-Vokoun
- UIowa: Bradley McDowell, Aaron Bossler, Mary Charlton, Amanda Kahl
- Epic: Vikas Biliyar, Brad Strock, Connor Nolan, Adam Thody, Eric Kleinberg
- Royal College of Pathology: Lazlo Iglali
- UK Digital Health: Jeremy Rogers, Deborah Drake
- College of American Pathologists: Raj Dash, Mary Kennedy, Alexis Carter
- Observables Project SNOMED Int: Daniel Karlsson, Ian Green, Farzaneh Ashrafi, Suzanne Santamaria, David Sperzel
- NeHTA, RCPA: Michael Osborne, Meagan Judge
- GO: Manuel Glynias, Nicole Kusold

Interoperation in healthcare

- Widespread implementation of the EHR has posed

University of Nebraska Medical Center | Nebraska Medicine

DEPARTMENT OF PATHOLOGY AND MICROBIOLOGY

Name: ,			Accession No:
Hosp No:	Acct No:	Age:	Date Taken:
DOB:			Date Received:
Sex:			Submitted by:
Loc: **E**	**Room:**		Client:

Surgical Pathology

Final Diagnosis:

LEFT UPPER LOBE OF LUNG, LOBECTOMY:
- INVASIVE MUCINOUS ADENOCARCINOMA.
- GREATEST DIMENSION 4.6 CM.
- THREE OF FIVE LOBAR HILAR AND INTRAPARENCHYMAL LYMPH NODES POSITIVE FOR CARCINOMA (3/5).
  - GREATEST SIZE OF LYMPH NODE METASTASIS 1.0 CM (A4).
- MARGINS (VASCULAR, BRONCHIAL, PARENCHYMAL) NEGATIVE FOR CARCINOMA.

MUTATION DETECTED: EGFR .c2155G>T (G719C) This mutation is associated with increased sensitivity to the EGFR TKIs, erlotinib (Tarceva) and gefitinib (Iressa; Han et al. 2005; Lynch et al. 2004; Rosell et al. 2005; Taron et al. 2005). Of note, in a trial of the irreversible pan-ErbB TKI, neratinib (HKI-272), 3 of 4 patients with EGFR G719X mutation had a partial response (Sequist et al. 2010).
(https://www.mycancergenome.org/content/disease/lung-cancer/egfr/2/)

Terminologies and Datasets for Pathology in Cancer

Research community

- HUGO; Human Gene Nomenclature Committee
- UniProt; Ensemble; Cosmic
- NCBO: OMIM; Orphanet; Protein Ontology
- Global Alliance for Genomics and Health

Clinical community (from ONC S&I framework)

- SNOMED CT 20180130
- LOINC 2.63
- RXNORM
- HIPAA transaction code sets

Public Health community

- NAACCR v18: ICD-O3, ISO, NCHS, SEER codes
- HL7 V2 CDA standard for reporting to cancer registries (CDC)
- CDC National Program of Cancer Rregistries statistics

Terminologies and Datasets for Pathology in Cancer

Research community

- HUGO; Human Gene Nomenclature Committee
- UniProt; Ensemble; Cosmic
- NCBO: OMIM; Orphanet; Protein Ontology
- Global Alliance for Genomics and Health

No meaningful semantic bridge links genetic scientific findings with clinical concept models or public health

Public Health community

- NAACCR v18: ICD-O3, ISO, NCHS, SEER codes
- HL7 V2 CDA standard for reporting to cancer registries (CDC)
- CDC National Program of Cancer Rregistries statistics

AP/MP Cases for Data (Re-)Use

Anatomic pathology (AP) is the discipline expert in making gross anatomical or microscopic diagnoses.

Molecular pathologists employ proteomic or genomic methods to diagnose disease.

- Laboratory risk and safety management
- Retrieving biobank tissue for research protocols

AP/MP Cases for Data (Re-)Use

- Clinical decision support
- Clinical research project planning
- Clinical quality assurance and outcomes assessment
- Laboratory risk and safety management
- Retrieving biobank tissue for research protocols

GPC RCR Quality Assurance: Treatment Audits

- Colon cancer therapy:
  - NRAS or KRAS variant+; EGFR treatments are ineffective and should not be used
  - EGFR+ and *RAS- ➔ cetuximab or panitumumab
  - high degree MicroSatellite Instability ➔ immunotherapy – Pembrolizumab or Nivolumab

Research Use Cases in Precision Medicine

- How many cases in the past two years have we had with advanced stage breast cancer and triple negative status(ER-, PR-, Her2-)?
- What is the survival for early stage colon cancers with/without finding of tumor budding?
- How many cases of stage 1-2 melanoma do I have in the tissue biobank that were BRAF V600E positive?
- How many patients do we currently follow with Lynch syndrome due to MLH1 variant?

Quality Assurance: Molecular Pathologist

- A new research study reports pathogenicity of a CDH1 sequence variant in stomach cancer; how many patients have we diagnosed in the last five years that I need to contact?

- Clinical research reports significance of the BRAF and *RAS genes in colorectal treatment planning, how do I need to change my reports to oncology?

Quality Assurance: Molecular Pathologist

What is needed to meet the needs of these researchers and clinicians is a domain ontology – a structured, fully defined coding hierarchy with knowledge features - capturing detailed AP & MP observations in the EHR which can be employed in data bases organizing the growing body of scientific knowledge and promoting interoperation between the communities of use.

A domain ontology carries the definitional information with the data as it moves from clinical to research to public health Databases.

These data types are called “Observables” within semantics of SNOMED CT and LOINC

UNMC: Project for structured encoding of AP/MP cancer reports

- Objective: Detailed structured and coded reporting of all anatomic and molecular pathology observations for all CAP synoptic cancer worksheets (82 types of malignancies)
- Project plan: Analyze and encode details of CAP worksheets for cancer
- Tooling: Nebraska Lexicon© extension namespace; SNOWOWL authoring platform; DL classifier
- SNOMED International committed to project workplan for 2017-8 with goal of promoting an observables ontology to International release

College of American Pathologists Cancer Checksheets

Histologic Type (select all that apply) (Note B)
___ Adenocarcinoma
___ Mucinous adenocarcinoma
___ Signet-ring cell carcinoma
___ Medullary carcinoma
___ High-grade neuroendocrine carcinoma

+ EGFR by Mutation-Specific Immunohistochemistry
+ EGFR L858R (clone 43B2)
+ ___ Negative#
+ ___ Positive##
+ ___ Equivocal### (explain): _____

+ *EGFR* Exon 19 Deletion (E746_A750del) (clone 6B6)
+ ___ Negative#
+ ___ Positive##
+ ___ Equivocal### (explain): _____

Sample from CAP Colorectal and Lung Cancer Checklists
College of American Pathologists, Northfield, IL USA

College of American Pathologists Cancer Checksheets

Histologic Type (select all that apply) (Note B)
___ Adenocarcinoma
___ Mucinous adenocarcinoma
___ Signet-ring cell carcinoma

+ RESULTS

+ *EGFR* Mutational Analysis (Note B)
+ ___ No mutation detected

+ *ALK* Rearrangement by Molecular Methods (Note C)
+ ___ No rearrangement detected#
+ ___ Rearrangement identified##
+ ___ EML4-ALK (specify variant type, if known): _____
+ ___ KIF5B-ALK
+ ___ TFG-ALK
+ ___ KLC1-ALK
+ ___ Other *ALK* rearrangement (specify, if known): _____
+ ___ Cannot be determined (explain): _____

+ Polysomy:
+ ___ Present###
+ ___ Absent

Sample from CAP Colorectal and Lung Cancer Checklists
College of American Pathologists, Northfield, IL USA

Project Workplan - Terminology

- Analyze terminology requirements in CAP Cancer worksheets
- Consult pathologists regarding query and decision support requirements to refine application of concept model
- Extend SNOMED CT content where necessary
- Vet the application of concept model with Observables Project and create templates to guide concept model application for reproducibility
- Encode in SNOMED CT and map Observables to LOINC; genes to HGNC, GO
- DL Classify and publish domain ontology for AP/MP
- (Implementation to Scott)

Nebraska Lexicon© Observables Ontology: SNOMED CT Concept Model Developments and Content Extensions

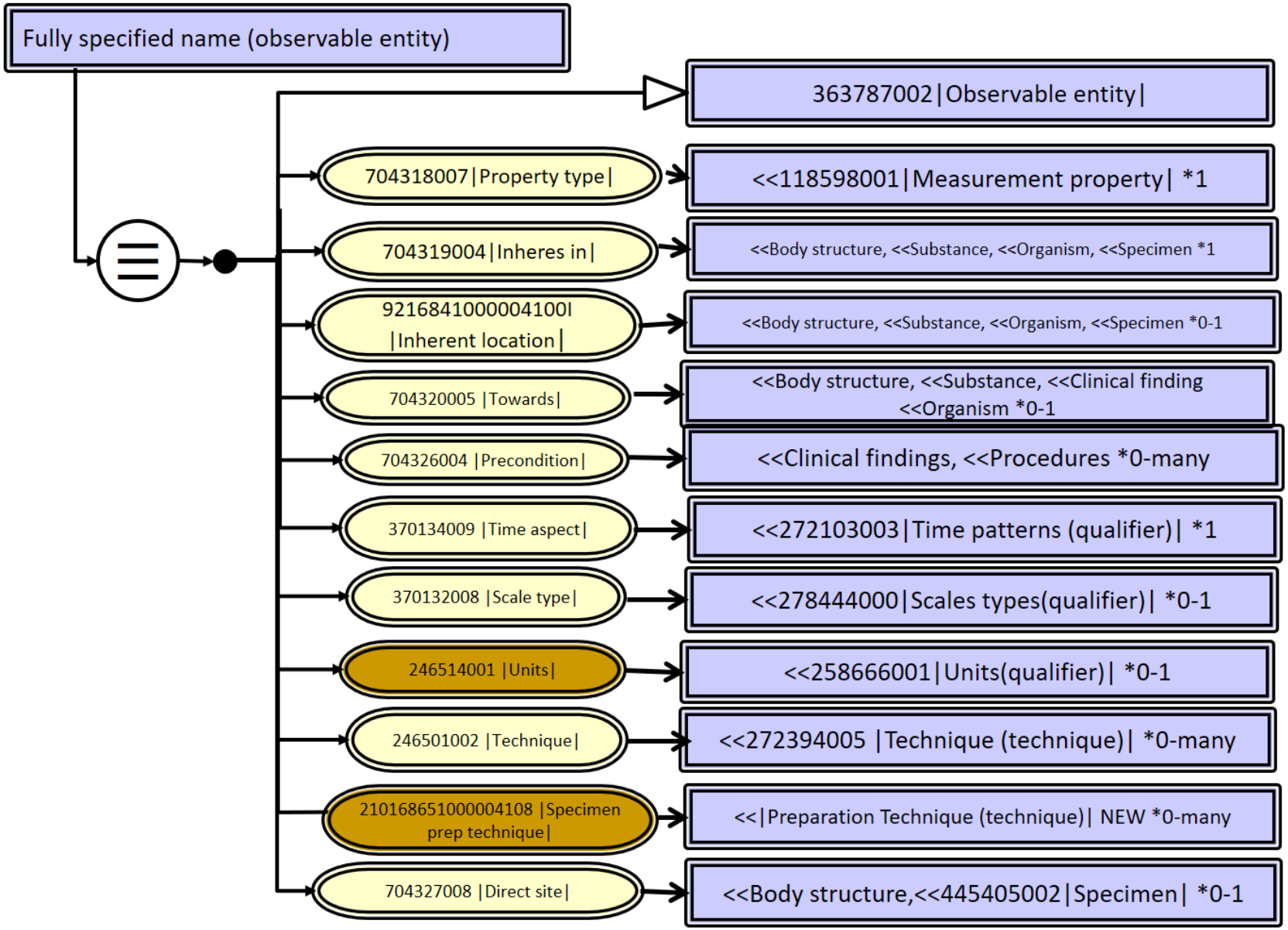
Operational Requirements: AP Observables Ontology

Query and retrieve:

- Gross exam results
- Microscopic:
  - detailed histologic findings by site of origin
  - histologic findings by slide preparation or methods (frozen section vs paraffin)
  - invasion and features of spread
  - special staining techniques

- Information model to link findings by case and episode over time

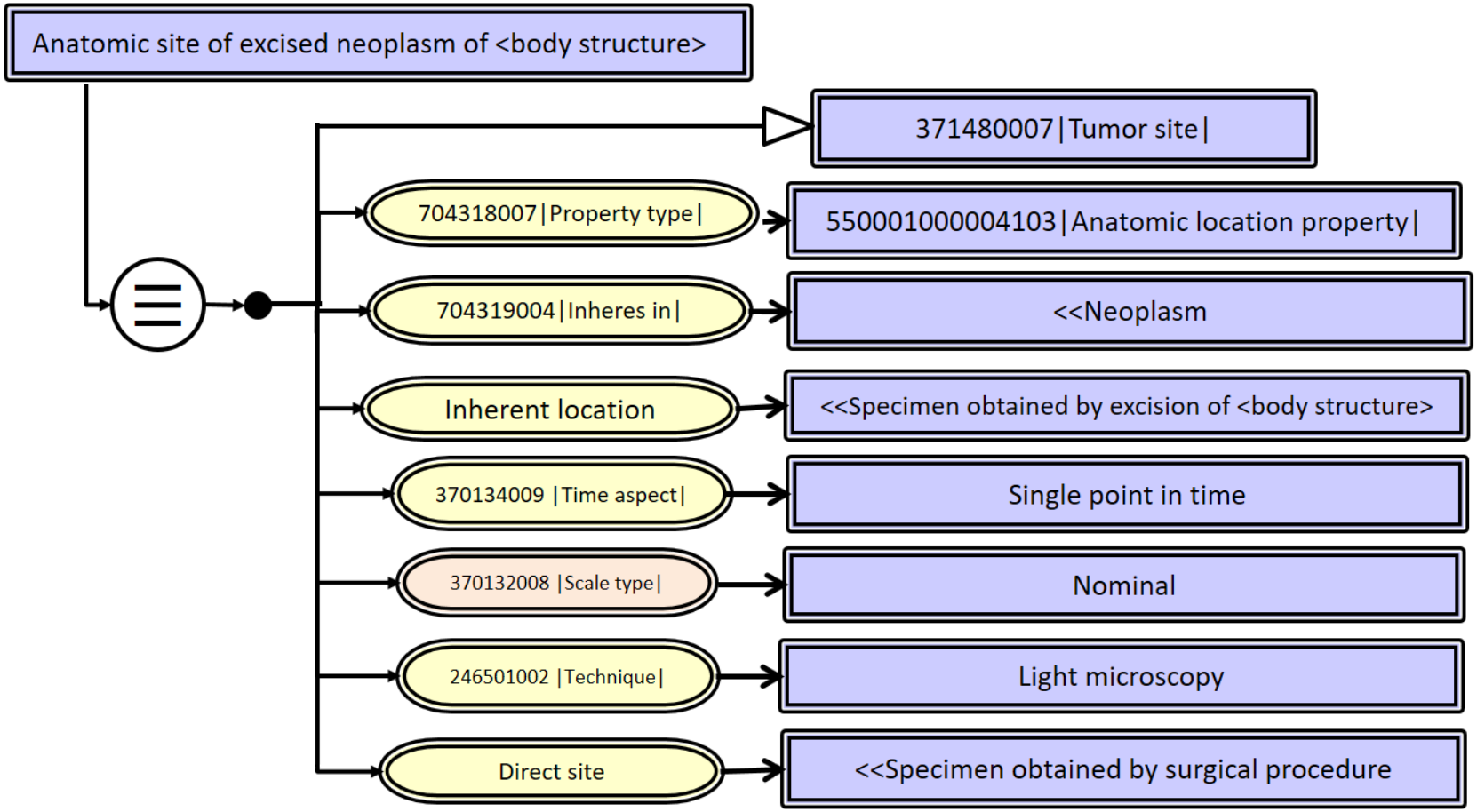
Observables Concept Model



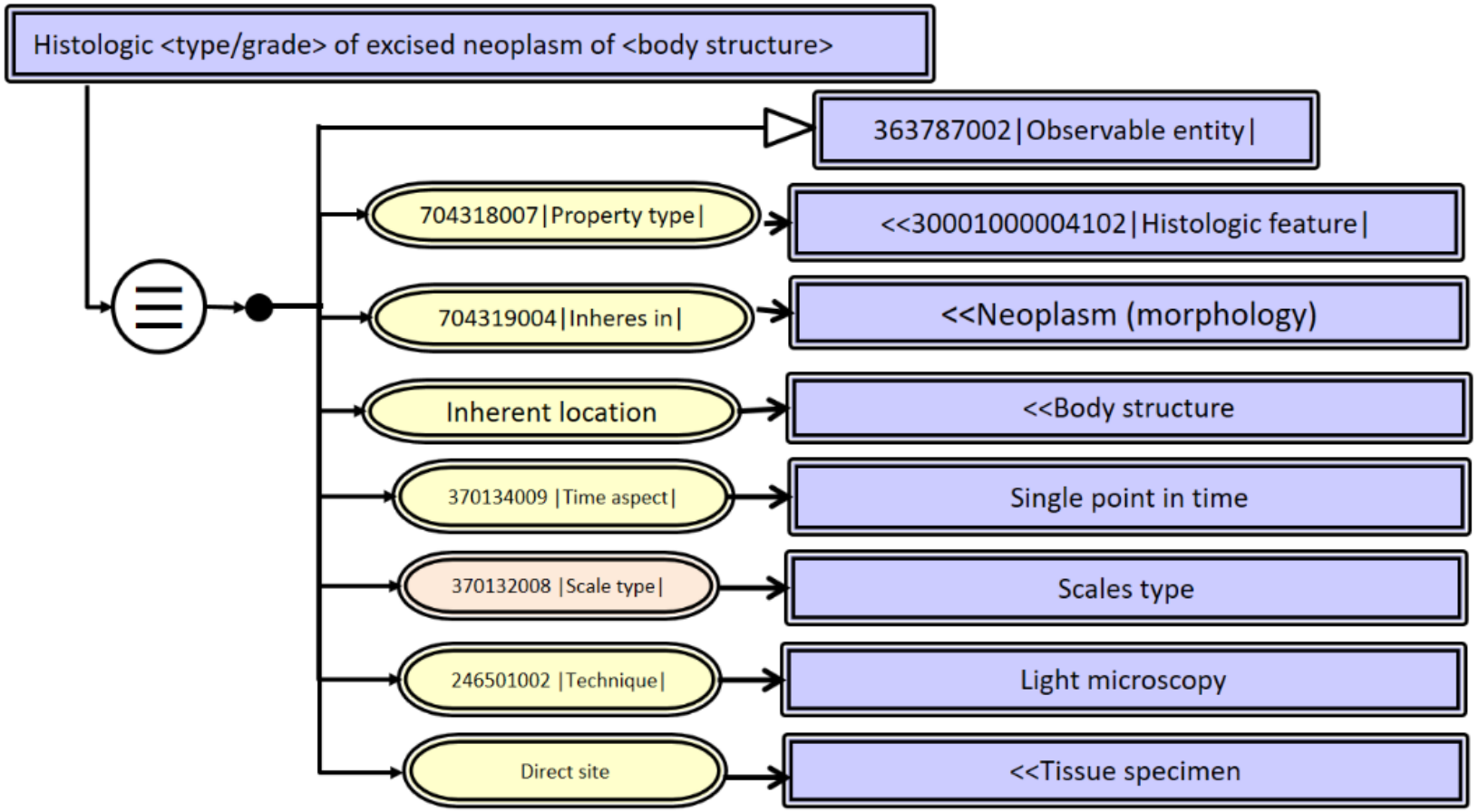
AP SNOMED CT Extensions

- Property types: morphology, histologic feature, cell type, grade, invasiveness
- Techniques: gross, microscopy, light microscopy, immunohistochemistry, immunoperoxidase staining, paraffin embedding, frozen section preparation
- Body structures: surgical margins

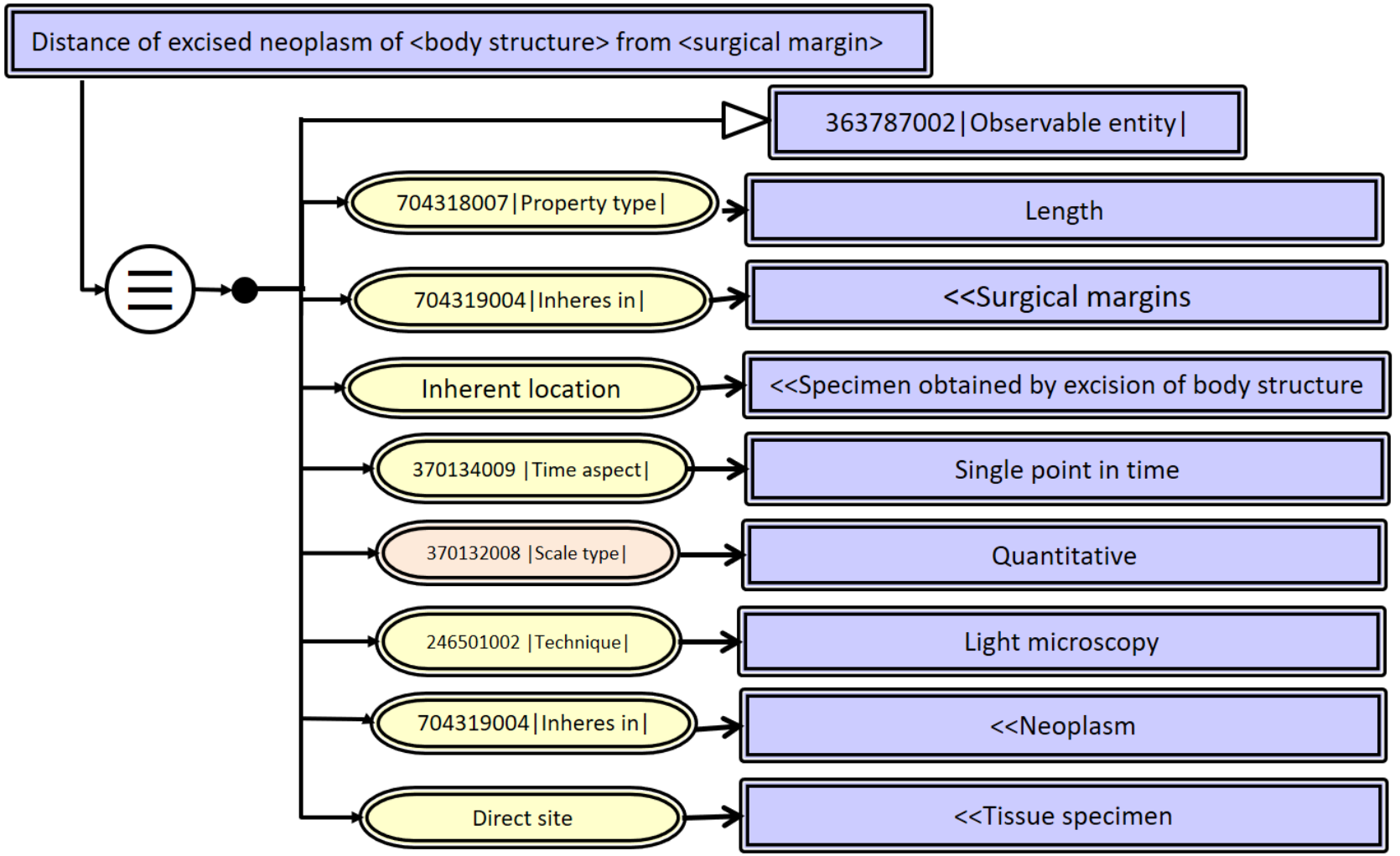
Anatomic site of excised neoplasm of <body structure>



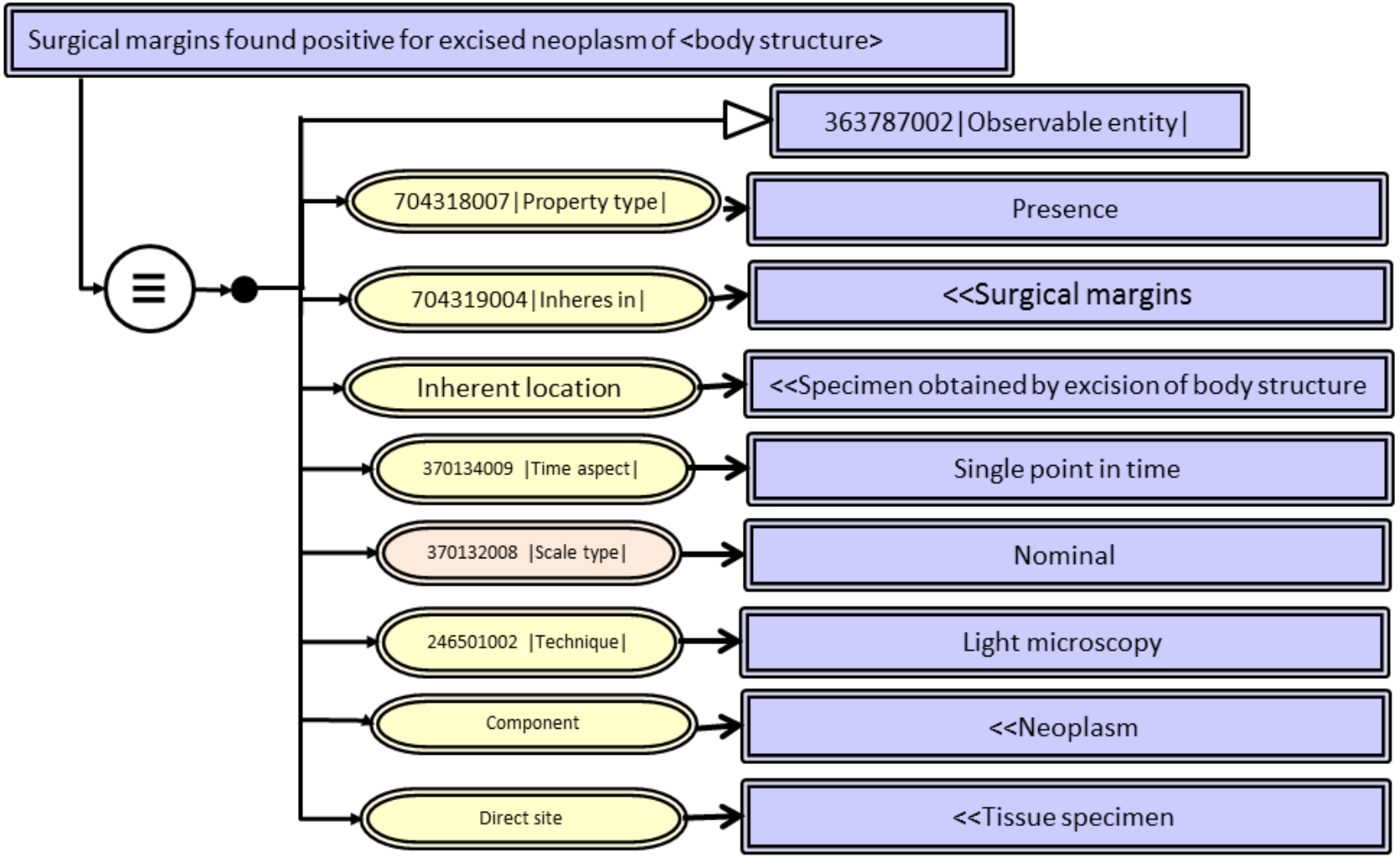
Histologic <feature> of excised <phenotype> of <body structure>



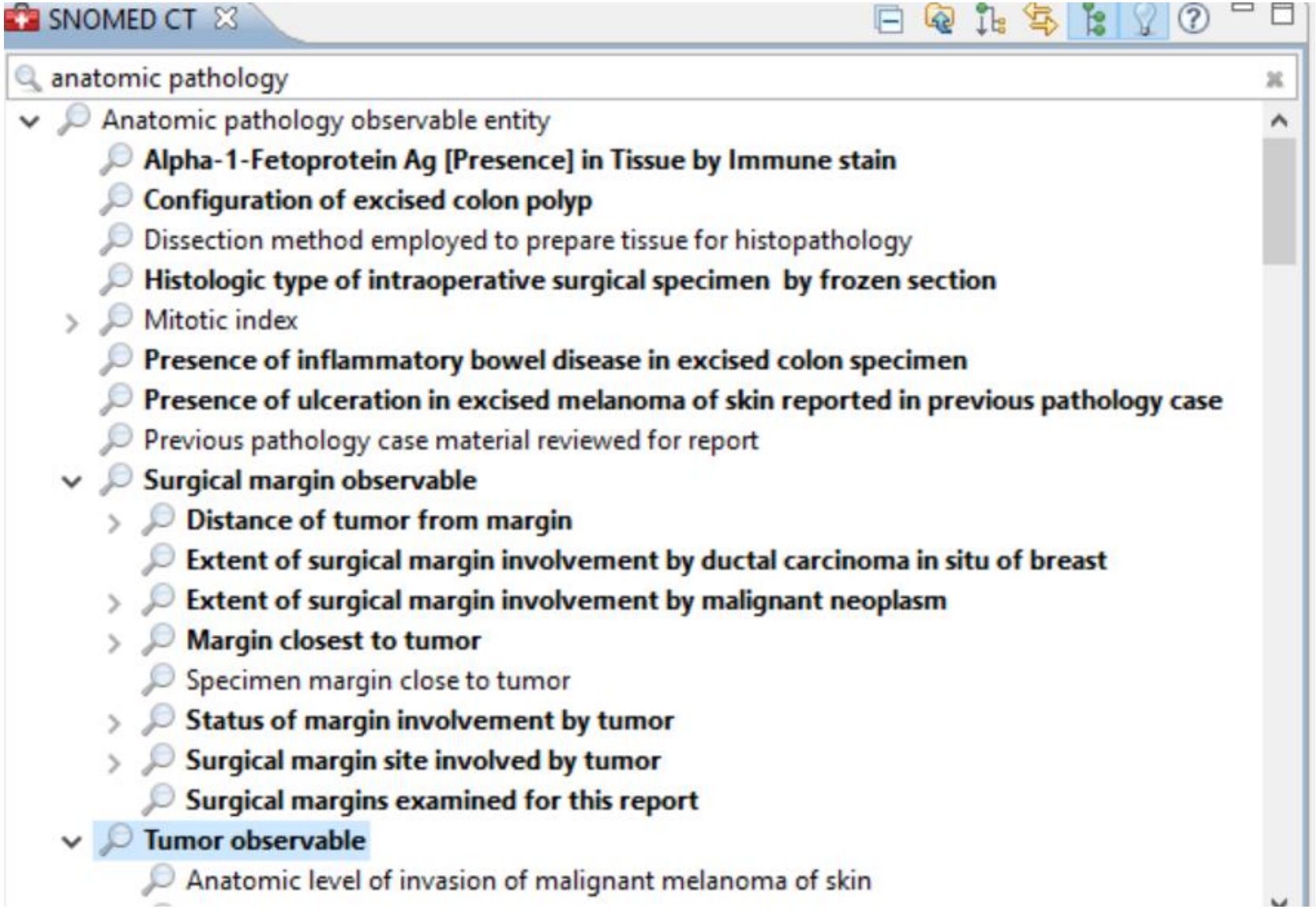
Distance of neoplasm of <structure> from <surgical margin>



Surgical margin found positive for neoplasm of <structure>



AP Observables (classified)



Operational Requirements: MP Observables Ontology

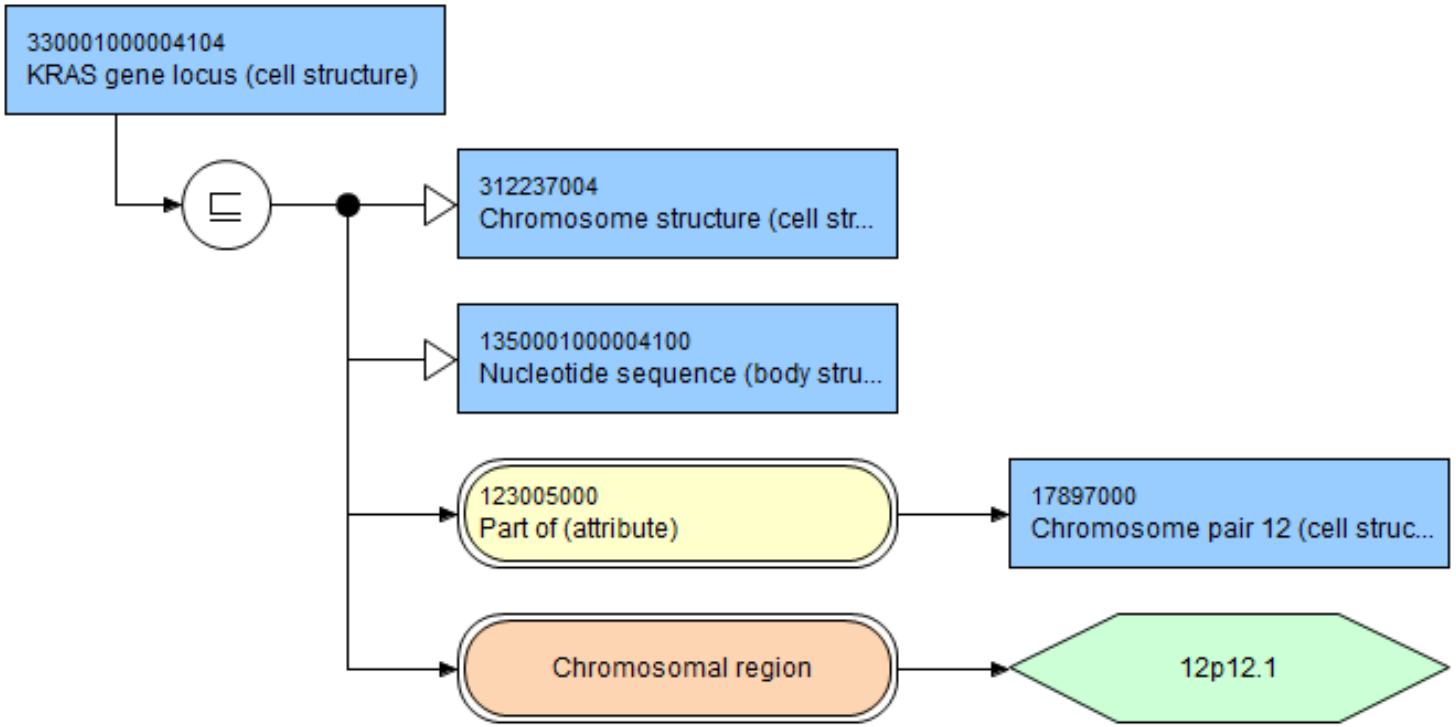
Query and retrieve:

- somatic vs germline results
- findings by gene locus or metabolic pathway
- findings by molecular genomic methods (protein, gene or chromosome)
- specific SNPs by gene and chromosome
- types of chromosomal rearrangement
- specific translocations
- gene amplification and polysomy
- …

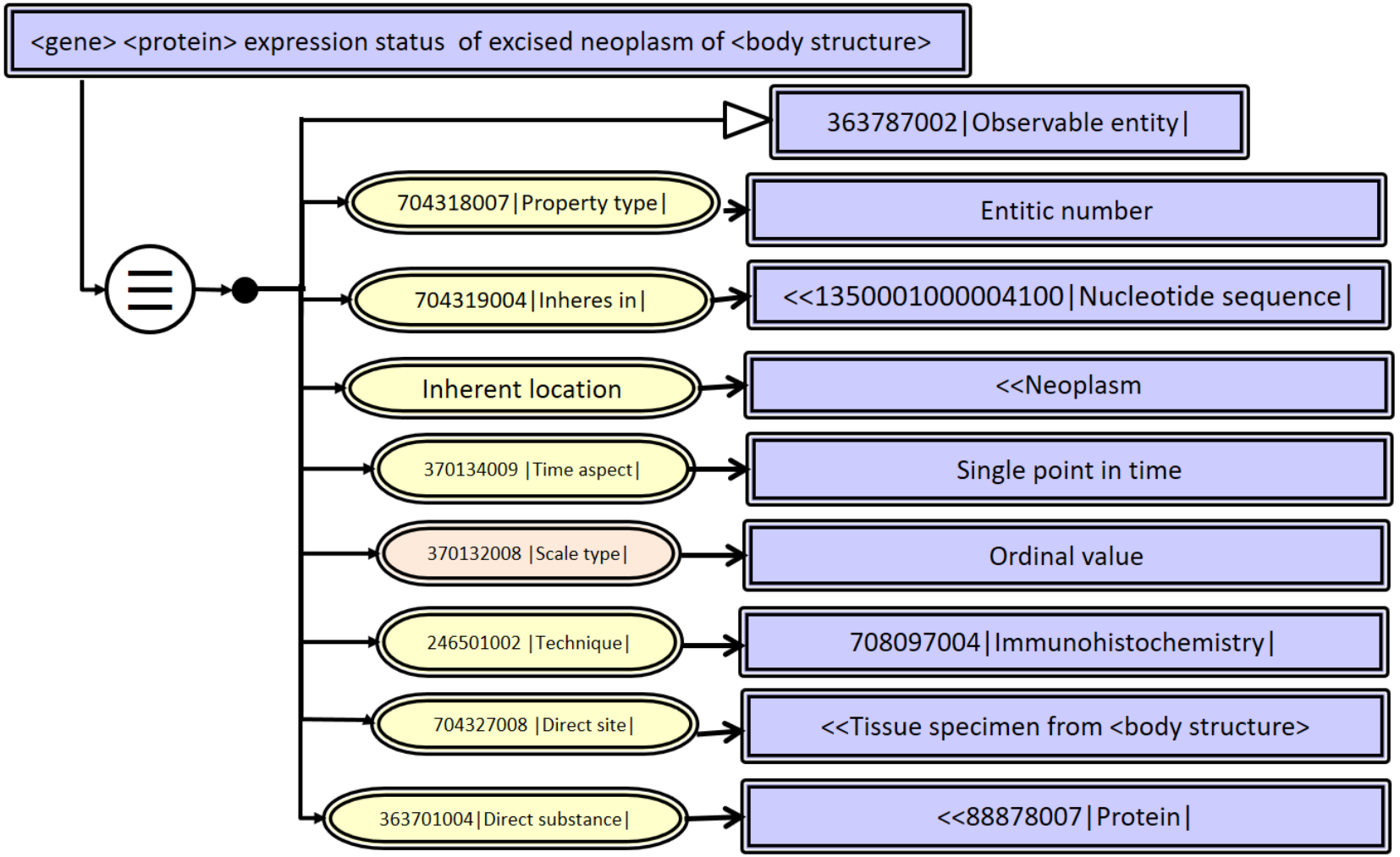
MP SNOMED CT Extensions

- Properties: chromosome rearrangement, gene amplification, sequence variation
- Techniques: molecular genetics, bisulfite treatment, flow cytometry, nucleotide sequencing, FISH, pyrosequencing, PCR
- Body structures: genes, nucleotide sequence alterations
- Substances: proteins and enzymes

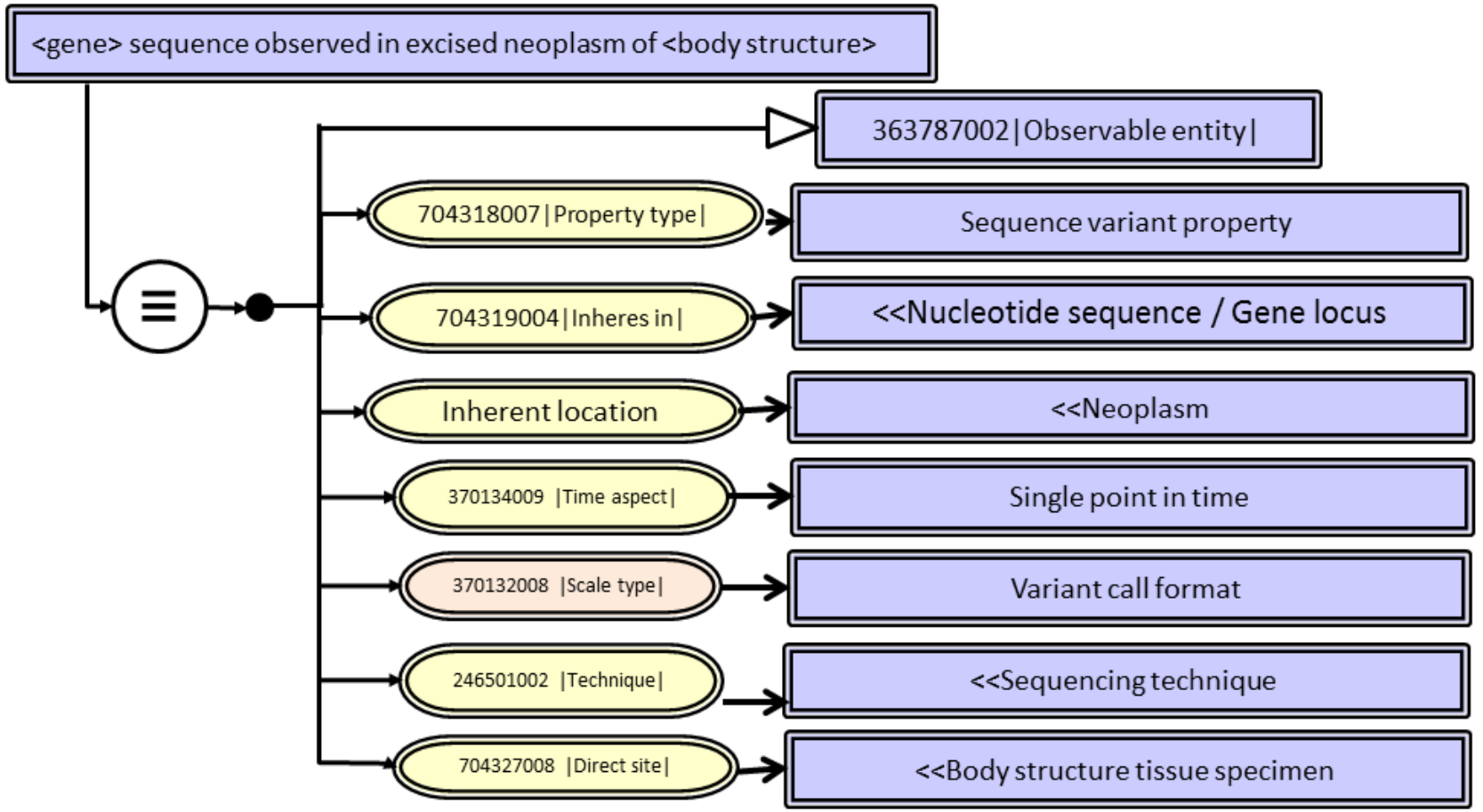
Extending SNOMED CT: 330001000004104|KRAS gene locus (cell structure)|



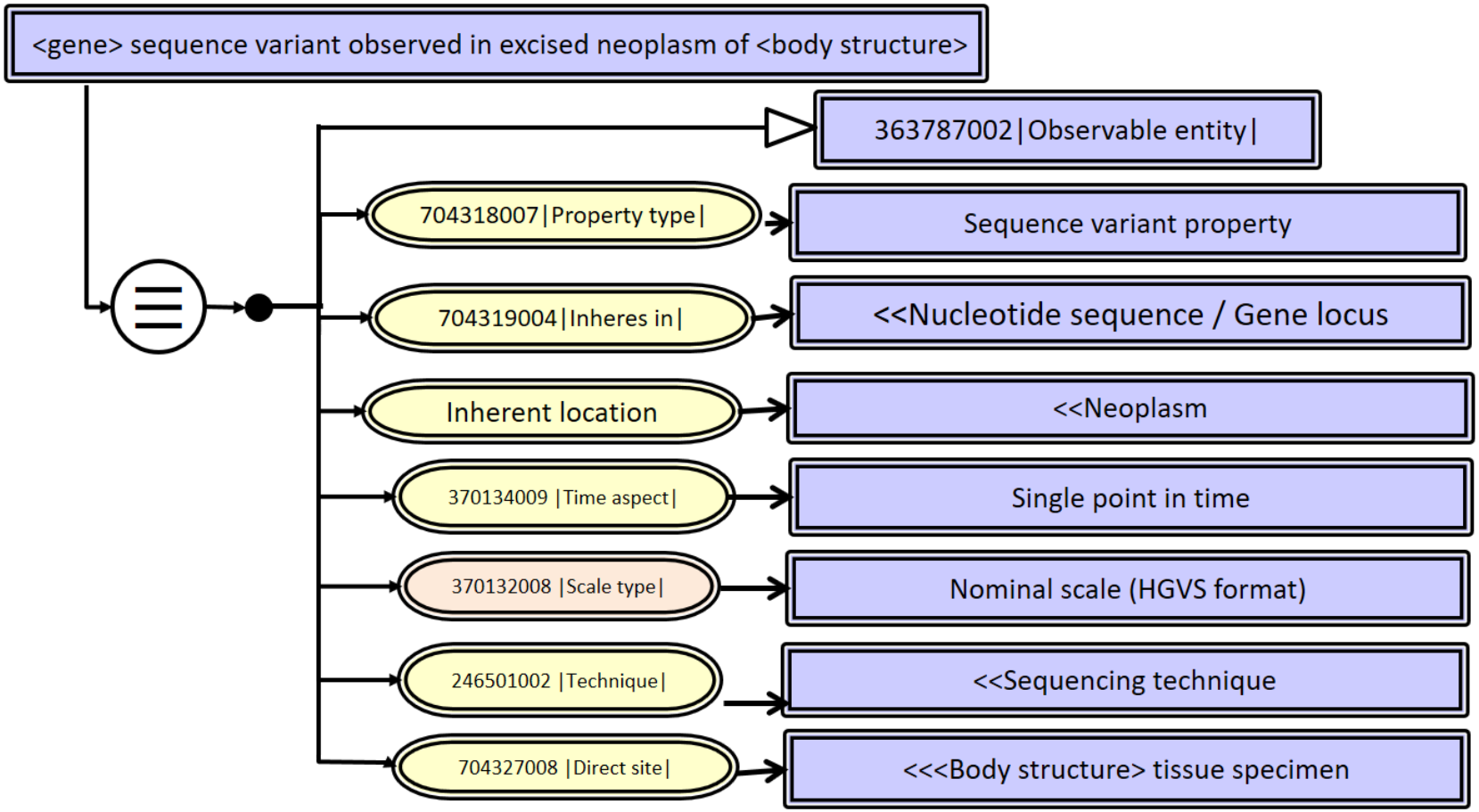
<protein> expression of <gene> assessed by IHC



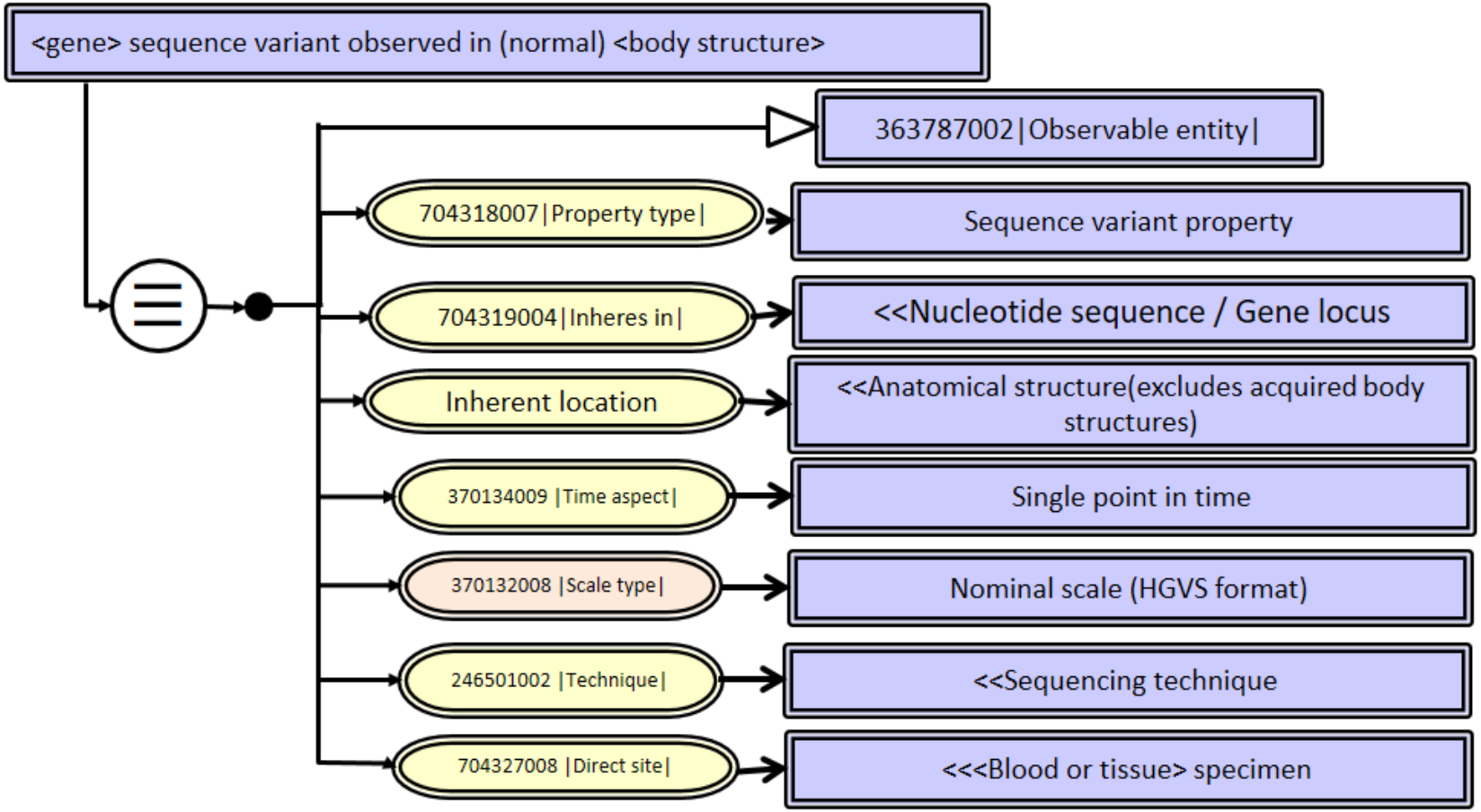
<gene locus> sequence observed in <neoplasm> of <body structure>



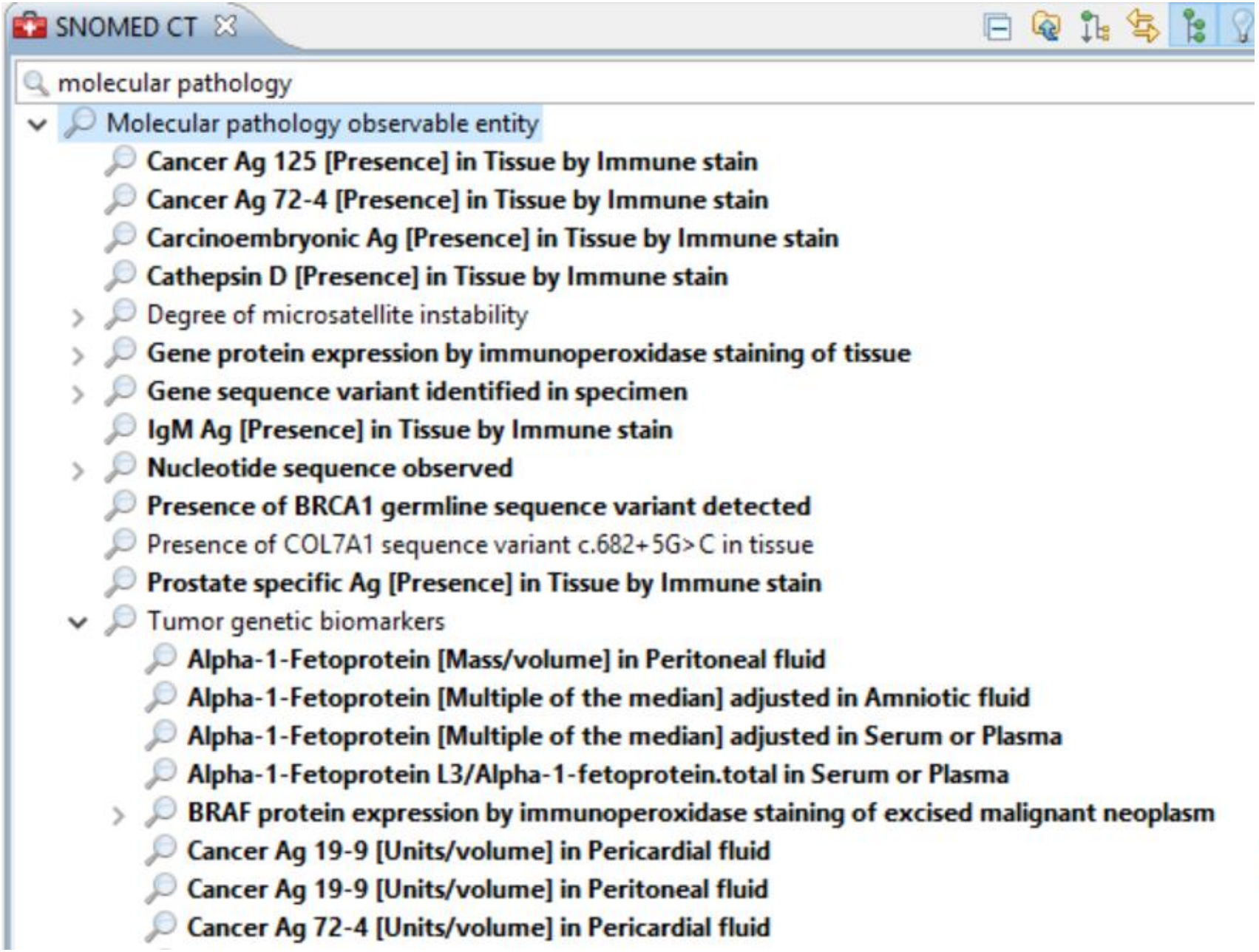
Somatic variants



Germline variants

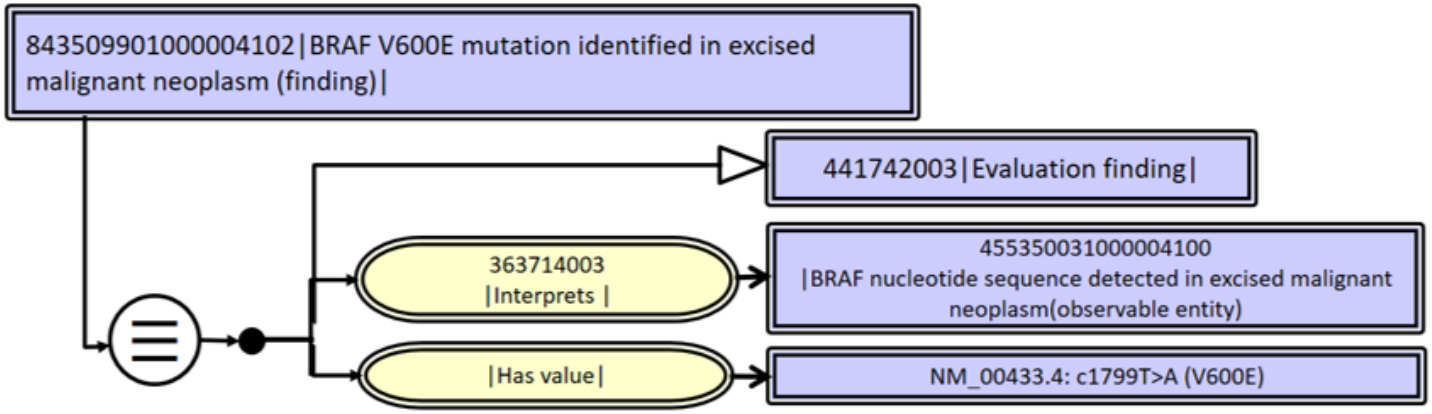


MP Observables

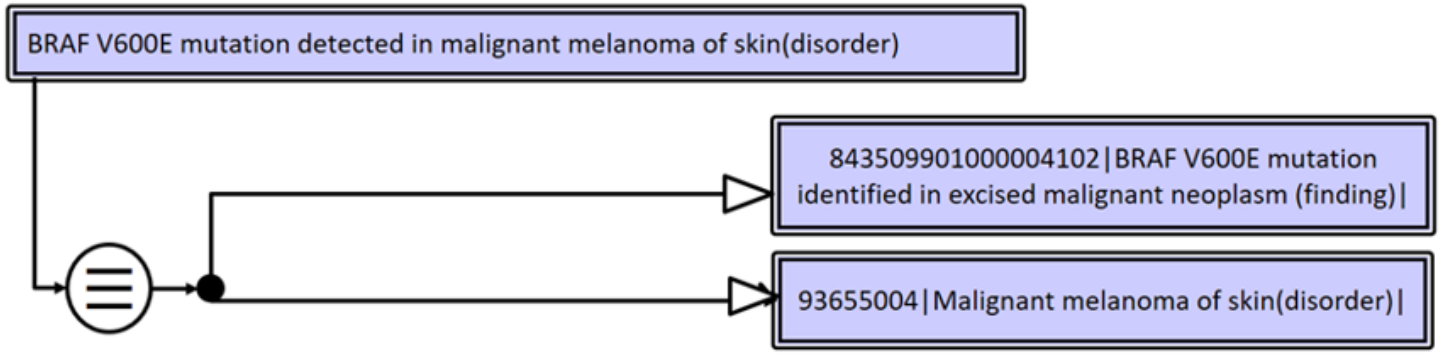


Implications for Clinical Findings

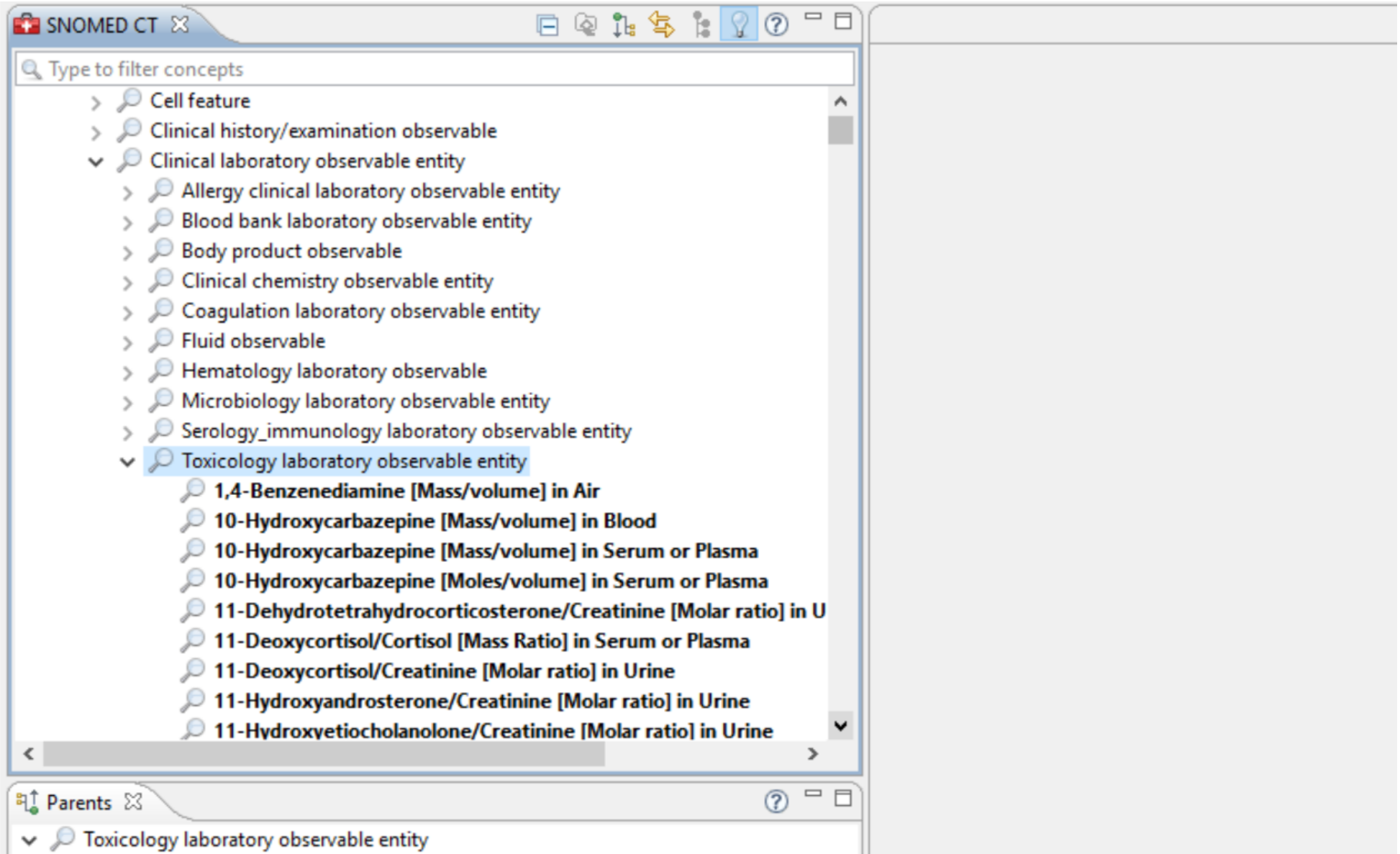
BRAF V600E mutation identified in malignancy



BRAF V600E mutation detected in malignant melanoma of skin



Clinical Laboratory Ontology



Scott to discuss automation of the AP/MP workflows and structured synoptic report

Nebraska Lexicon © extension:
https://www.unmc.edu/pathology/informatics/tdc

James R. Campbell
campbell@unmc.edu

W. Scott Campbell
wcampbel@unmc.edu

Nebraska Medicine

SERIOUS MEDICINE. EXTRAORDINARY CARE.®