Briefing Note for Editorial Advisory Group, Content Managers Advisory Group, Clinical Leads, Member Forum, and Community of Practice

Representation of Human Genes in SNOMED CT

Version 1.5

Status: Open

Request for input and Consultation on impact

Version history

Version	Date	Comment
0.1	2025-05-08	Initial version for review by the Member Forum
0.5	2025-08-18	Revision for general review by Content team
0.8	2025-09-03	Revised based on comments from the Member Forum and content team
1.0	2025-11-03	Version for community review
1.5	2026-01-06	Revision following community review. Refinement of scope. Naming of new attribute with definition. Clarification on proposed modeling and addition of chromosomal bands.

Purpose

The purpose of this briefing note is to introduce a proposed structure for the representation of human gene names and their associated chromosomal locations in the context of body structures, specifically as subcellular structures. This shift from the current location of genes within SNOMED CT reflects the biological reality of genes as physical, organized, and functional units within chromosomes, aligns SNOMED CT with contemporary clinical practice, and future-proofs the terminology for the era of precision medicine.

This proposal defines a structural representation for clinically relevant human gene names within SNOMED CT, **solely** to support the authoring, definition, and classification of SNOMED CT content such as Clinical findings, Disorders, and Observable entities. This proposal does not aim to provide a comprehensive catalog of human genes, nor does it intend to replicate all of the genomic relationships that might be represented in a gene ontology such as the HUGO Gene Nomenclature Committee (HGNC), although it seeks to align with external genomic standards. The focus is to associate clinically relevant genes with their representative chromosomes. This proposal provides non-normative examples of how gene names may be used in modeling other content (Appendix 1) such as Observable entities, Clinical findings, or Diseases.

This proposal also does not address the notion of a gene as an informational artifact (an alternative representation supported by organizations such as NCBI), but focuses on the structural relationship of a gene to its associated chromosome.

Background

Stimulated by inquiries from the UK and New Zealand regarding the planned SNOMED CT representation of gene names, during the April 2025 Pathology and Laboratory Medicine Working Group meeting in Oslo, Norway, a discussion concerning the proper placement of gene names in the SNOMED taxonomy was held. Currently, a small number of gene names are located in the Substance hierarchy under 67271001 |Gene (substance)|. Of the 322 subtypes of this grouper, 274 are located under 260071009 |Human leukocyte antigen allele (substance)|. Only very few other named genes are contained in this subhierarchy. Furthermore, there are no gene names under 82256003 |Human gene (substance)| and 49046007 |Human structural gene (substance)|. Of the existing gene substances, only 14 are used to model 14 evaluation procedure concepts.

Based on the provided use cases, discussion by members of the working group on the proper positioning of genes within the SNOMED taxonomy led to the consensus that the proper location was as a subtype of the body structure hierarchy. This is consistent with other ontologies such as the Foundational model of Anatomy, which treats genes as physical parts of chromosomes and the Sequence ontology (http://www.sequenceontology.org/) that defines a gene as a biochemical feature located at a specific position on a sequence, i.e. on a chromosome.

The UK currently has identified a list of approximately 4900 gene names that it proposes would be useful to add to SNOMED to support their genomics efforts. New Zealand has added 382391000210108 |Gene structure (cell structure)|, multiple genetic alteration morphologies, and 274 gene concepts under the 312237004 |Chromosome structure (cell structure)| in their national extension. The proper structure to support the location of genes within the SNOMED hierarchy is critical to support their and other member countries efforts in the genomics domain. Other member countries (BE, AU, SE) have expressed support for addition of gene names to the International release of SNOMED CT.

Justification for this proposal

While some references refer to genes as primarily informational units as opposed to structural entities, the addition of genes as structural components of a chromosome is consistent with the HGNC relationship of “chromosomal location” as a defining characteristic of a gene.

The lack of name genes and inability to reference specific genes in Clinical finding and Disease models has been an issue for many years in SNOMED CT with content tracker tickets dating back to 2012. The need to address this issue was made more apparent during the creation of the LOINC extension to SNOMED where specific gene observables are needed.

Use cases

- Provide a structured list of clinically relevant genes names associated with their respective chromosomes to support the authoring, definition, and classification of SNOMED CT content such as Clinical findings, and Disorders.
- Provide the ability to sufficiently define Observable entity concepts related to molecular pathology
- Provide an indirect link between SNOMED CT and more detailed gene standards such as the HGNC database
- Provides the ability to link a structural gene with its protein transcript, useful in sufficiently defining specific deficiencies.

Issues/problem statement

In addition to the inclusion of gene names in the body structure hierarchy, the proper relationship to their chromosomes must also be considered. Currently, SNOMED CT represents chromosomes as both individual structures (e.g. 1366639000 |Chromosome 1 structure (cell structure)|) and chromosome pairs (e.g. 46507000 |Chromosome pair 1 (cell structure)|) under 91272006 |Chromosome (cell structure)|. The following issues with the current structure in SNOMED have been identified. As additional use cases are developed, more issues may arise.

1. Gene names as subtype of 115668003 |Biological substance (substance)| do not allow for their association with a particular chromosome (no existing attribute).
2. While genes (usually) reside on both members of a chromosome pair, the precise variant of a gene may differ between the members of a chromosome pair.
3. 312237004 |Chromosome structure (cell structure)| is intended to represent the grouper of all or parts of a chromosome, although the content of this area of the terminology is relatively unpopulated.
    a. E.g. the concept 312242007 |Long arm of chromosome (cell structure)|, a subtype of 312237004 |Chromosome structure (cell structure)| has no subtypes. If this were a logical grouper, then the long arm of each individual chromosome would be represented as subtypes. These, along with other chromosomal structures would need to be created to support the location of genes on the chromosome.
4. There is a gap in adherence to the established SEP model for cell and subcellular structures. Specifically, the requirement to include an 'entire' subconcept under each 'structure' concept in the cell subhierarchy is not consistently met. The terms 'structure' and 'entire' are often absent, with different semantic tags like `(cell structure)` and `(cell)` used in their place.

Proposed solution

To address the need for the appropriate representation of gene names as body structures (i.e. cell structure) in SNOMED, the following changes are proposed:

1. For individual chromosome structure concept subtypes of 312237004 |Chromosome structure (cell structure)|, individual “entire” chromosome concepts have been added (i.e. “Entire chromosome X”. As these are extant in the Body structure hierarchy, this implies that all subtypes will represent human chromosomes. ***This has been completed.***
2. Existing Chromosome pair concepts have been moved as subtypes of their appropriate Chromosome structure concept. ***This has been completed.***
3. The concept 312237004 |Chromosome structure (cell structure)| will be retained as the grouper for “all-or-part” of a chromosome and will be assigned the PT = Chromosome. The current subtype 91272006 |Chromosome (cell structure)| has been inactivated and the existing 264328003 |Whole chromosome (cell structure)| has been retained and retermed “Entire chromosome” . This is consistent with the SNOMED SEP model. ***This has been completed.***
4. 312242007 |Long arm of chromosome (cell structure)|, 278145009 |Short arm of chromosome (cell structure)|, and 70187007 |Centromere (cell structure)|, would remain as direct subtypes of 312237004 |Chromosome structure (cell structure)| and have “structure” added to the FSN. ***This has been completed.***
5. Each part of identified human chromosomes (e.g. Chromosome 1 structure) will be added as subtypes of the above part groupers. For the long and short arms, the terming will include the “p” and “q” representations, e.g. “Chromosome 1p” for the short arm. ***This has been completed.***

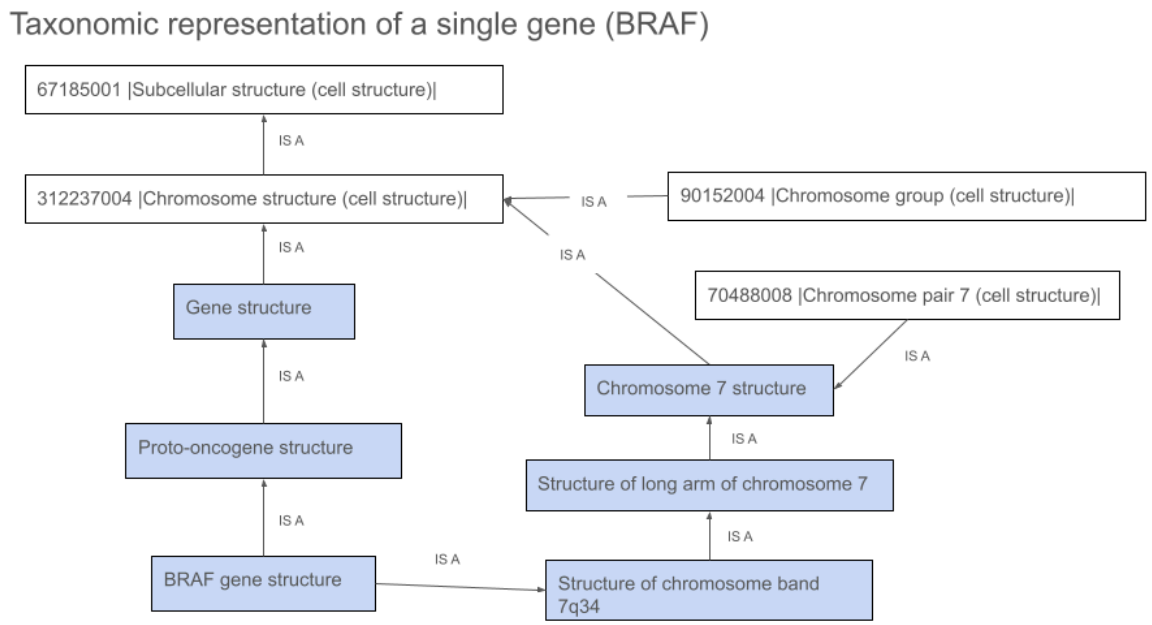
6. Chromosome bands representing the specific locations of genes will be added as subtypes of the corresponding chromosome arm.
7. Gene names will be added as subtypes of a new concept, *Gene structure* (see 6a), which is a subtype of 312237004 |Chromosome structure (cell structure)|, with an additional parent representing the specific chromosome band in which the gene resides.
    a. New Zealand has added a concept of "Gene structure (cellular structure) which will be requested for promotion and represent the "all-or-part-of" grouper for genes.
    b. E.g. "BRAF gene structure" IS A Gene structure (actually a proto-oncogene) and IS A Chromosome band 7q34 structure (see diagram below).
8. The current concepts 82256003 |Human gene (substance)| and 49046007 |Human structural gene (substance)| will be inactivated.
9. Human gene names will be added under the specific band and arm of their associated chromosome using the current name(s) as represented by the HGNC guidelines (https://www.genenames.org/about/guidelines/#!#genenames) (e.g. FSN= B-Raf proto-oncogene serine/threonine kinase gene structure, PT = BRAF gene)
10. Each gene will be modeled with one primary chromosomal location, with exceptional cases (e.g. complex loci, alternative assemblies) being out of scope.
11. Non-human genes will remain as substances until a general approach for non-human body structures has been developed.
12. To support modeling of genes with an etiological association to a disorder, a new attribute *HAS ASSOCIATED GENE* is proposed. A working definition is "Represents a recognized association between a disorder (or clinical finding) and a gene, where the gene is implicated in the condition but is not itself the anatomical site of the abnormality. This relationship does not assert a specific causal mechanism and may represent causative, susceptibility, modifying, or otherwise biologically relevant involvement."

As a result of these changes, the addition of individual chromosomes and their associated parts involving the 22 autosomes plus chromosomes X and Y, each with the long arm, short arm, and centromere has been completed. This resulted in the addition of 72 new concepts (24 × 3) to support the top-level chromosomal structures needed to properly add gene names.

While the use cases for adding specific chromosomal regions (bands), such as 1q21.1, to support karyotype-defined disorders (e.g. 1q21.1 microdeletion syndrome) and molecular pathology concepts associated with the LOINC extension are well recognized, the proposal does not require the immediate addition of a comprehensive set of all cytogenetic bands. Instead, chromosome band concepts will be added incrementally, prioritizing those needed to support defined clinical content and gene localization, with further expansion undertaken as resources allow. An automated process for adding these bands is being investigated. Similarly, a mechanism for appropriately representing chromosomal band ranges is under investigation and will be addressed separately from this proposal.

Diagrammatic representation of proposal

Proposal diagram using the BRAF gene as an example. Cells in blue represent new content.



Benefits of adding Gene names as Body Structures

- Allows for the use of the existing FINDING SITE attribute for Clinical findings to represent the location and type of changes to the gene associated with a disorder. The type of change can be described with new morphologic abnormality concepts (e.g. frameshift, translocation, deletion, etc.)
- Allows for the use of gene names to be used as the value for the INHERES IN attribute in the Observable entity concept model
- Will provide a mechanism to model genes associated with a disorder (requires the new HAS ASSOCIATED GENE attribute, See examples in Appendix 1).
- As conformant with the SNOMED SEP model for anatomy, the IS A relationship between structure concepts supports the classification hierarchy. The "part of" relationships are represented between entire concepts. This approach will utilise description logics for automatic classification and construction of IS A relationships between structure concepts.

Clarification on the use of “FINDING SITE” vs. HAS ASSOCIATED GENE

- The use of the FINDING SITE relationship is constrained to only those concepts where the specific chromosomal structure (e.g. arm or band) or gene itself is the locus of the abnormality (e.g. a mutation or deletion affecting the chromosome or gene structure).
- In other cases where a gene is referenced to indicate an involvement rather than the anatomical site of change, the gene would be modeled using the proposed “HAS ASSOCIATED GENE” attribute.

Suggested alternative: linking directly to HGNC

During review of the proposal, a suggested alternative was to avoid adding gene names as SNOMED CT concepts and instead link disorders directly to HGNC using external identifiers or gene symbols. While this approach aims to minimize duplication of externally governed content, it would significantly constrain how genes can be represented and used within SNOMED CT. Treating genes only as external references would prevent them from participating fully in the SNOMED CT concept model, including hierarchical organization, explicit relationships to chromosomal structures, and consistent use in defining Disorders and Observable entities. It would also increase maintenance complexity, as changes in gene nomenclature would need to be reflected wherever those references are used. For these reasons, and to support established genomics and molecular pathology use cases, the proposal adopts a controlled, needs-driven introduction of gene concepts within SNOMED CT that remain aligned with HGNC nomenclature while being managed through existing SNOMED CT editorial and release governance.

Examples of potential use for gene names in modeling other concepts

Other hierarchies within SNOMED CT that benefit from the addition of gene names include the Observable entity hierarchy and the Clinical finding/Disease hierarchy.

Observable entity

In the case of Observable entities, the primary use case is the detection of some form of alteration to the underlying structure of the gene. This may include mutations, deletion, insertions, duplications, translocation, etc. In these cases, the type of genetic alteration is what is being measured and the alteration INHERES_IN the gene. Examples of an Observable entity of this type are shown in Appendix 1. (**NB:** The examples are for illustration purposes only. They do NOT necessarily represent the final international model for Observable entities using gene names.)

Clinical findings/Disease

The primary use case for gene names in the Clinical findings/Disease hierarchy is to associate the specific gene known to contribute to the clinical condition, such as specific enzyme deficiency diseases, and other genetic diseases.

Limitations of the proposal

- The proposal ignores the notion of genes as information units often represented in the literature and by organizations such as the US National Center for Biotechnology Information (NCBI)
- Only addresses human gene names. Genes from other organisms are not addressed due to a lack of a terminological structure to represent anatomy (gross, cellular, and subcellular) of non-human organisms
- While named variants of genes can be represented in this structure, specific ***sequences*** that may be of clinical value are out of scope
- This proposal does not address RNA transcripts or protein translations. Proteins associated with structural genes will be addressed in a future proposal
- The proposal does not definitively describe the relationship of Chromosomes with other subcellular structures

Impact to content users

The addition of gene names and the use of these concepts to more fully define existing content in SNOMED CT has the potential to affect a large number of existing concepts (>1000); however, it is anticipated that the changes resulting from the use of these new concepts will result in a majority of the impacted concepts having a more detailed logical model allowing them to become sufficiently defined, but will have little effect on the taxonomic structure of SNOMED CT.

Next Steps

- Seek input on this proposal from the community and specifically those Members that are actively involved in representing genes in SNOMED CT.
- Implement the consensus proposal in the International release of SNOMED CT.
- Add genes (or promote from extensions) currently needed to support modeling of existing international content.

Release date

Following community engagement and approval of the proposed approach, the necessary chromosomal structures and gene names will be added incrementally over a series of releases beginning Q1 2026. Input on prioritization of the creation and application of gene names will be solicited from Members in support of the Content Development Roadmap.

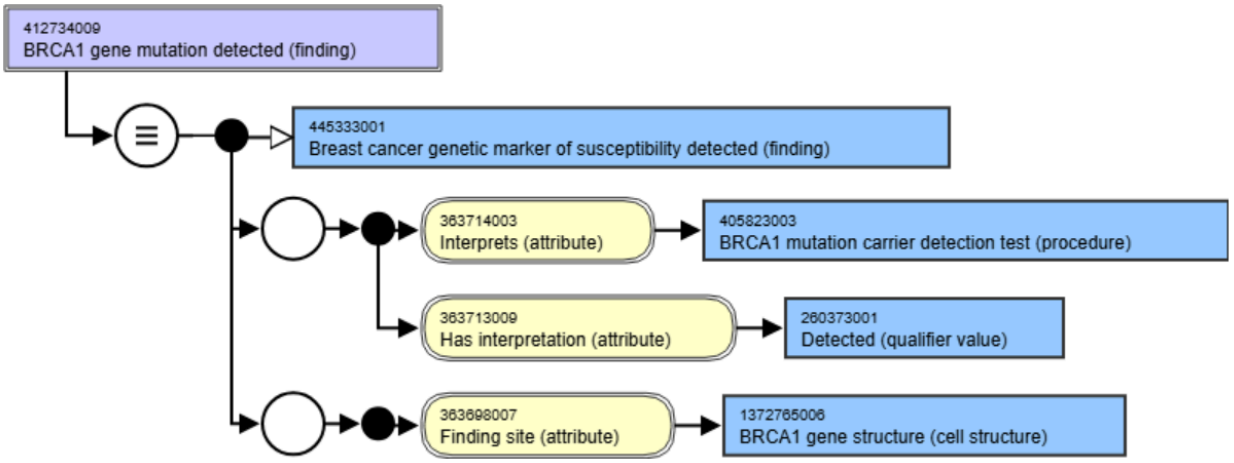
Request: Please provide comments, questions, and suggestions on the Briefing Notes page by February 28, 2026

Approvals	Date	Name
Chief Terminologist	2025-10-15	James T. Case
Director of Content and Mapping	2025-11-14	Monica Harry
CSRM representative	2025-11-13	Kelly Kuru

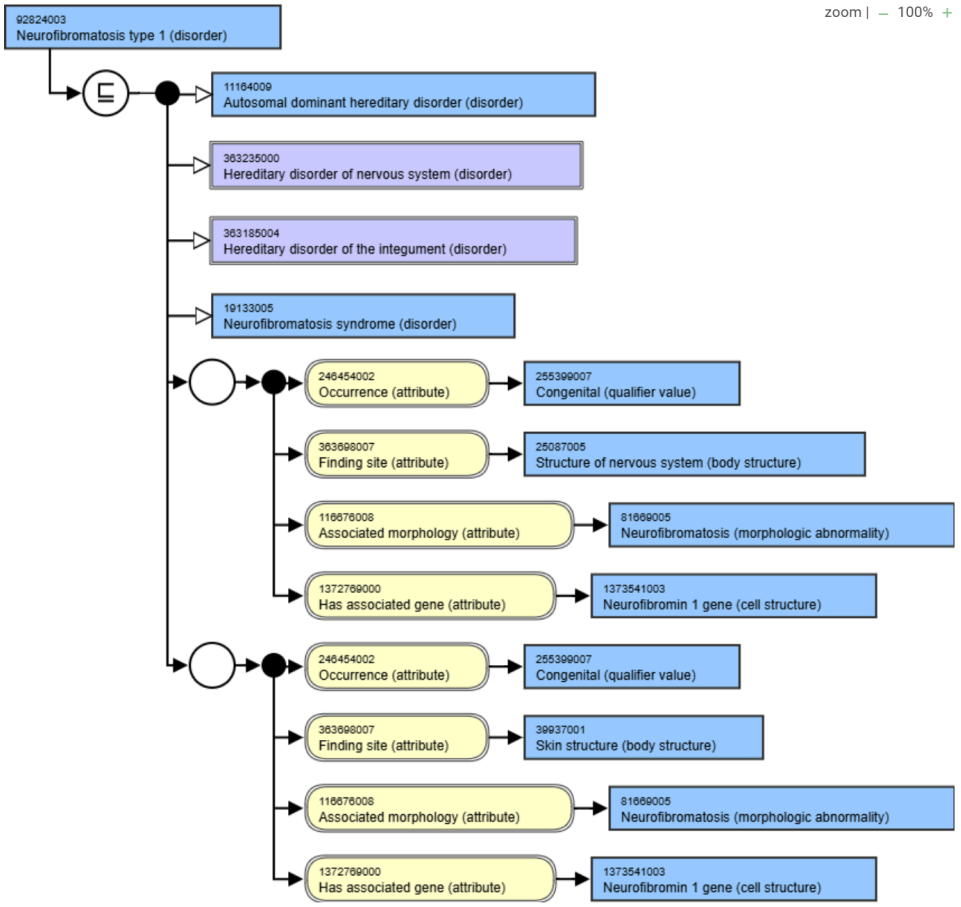
James T. Case 2025-08-12

Appendix 1 - Modeling examples using Chromosome structures and gene names

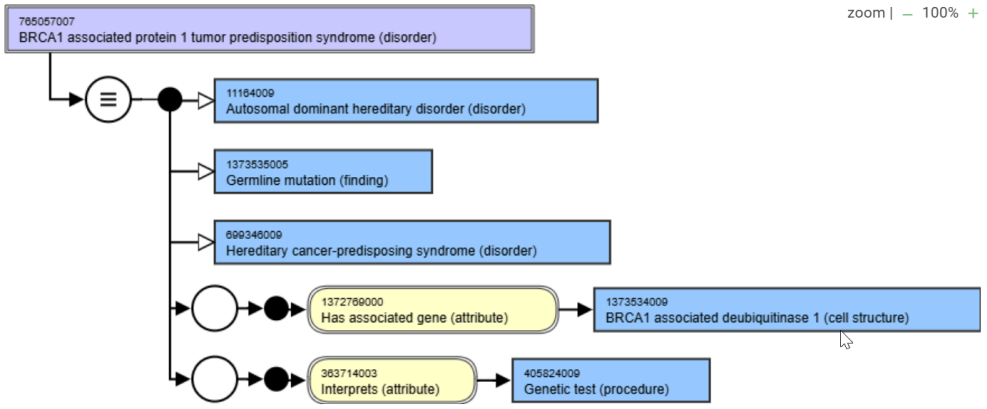
Example 1: 412734009 |BRCA1 gene mutation detected (finding)|



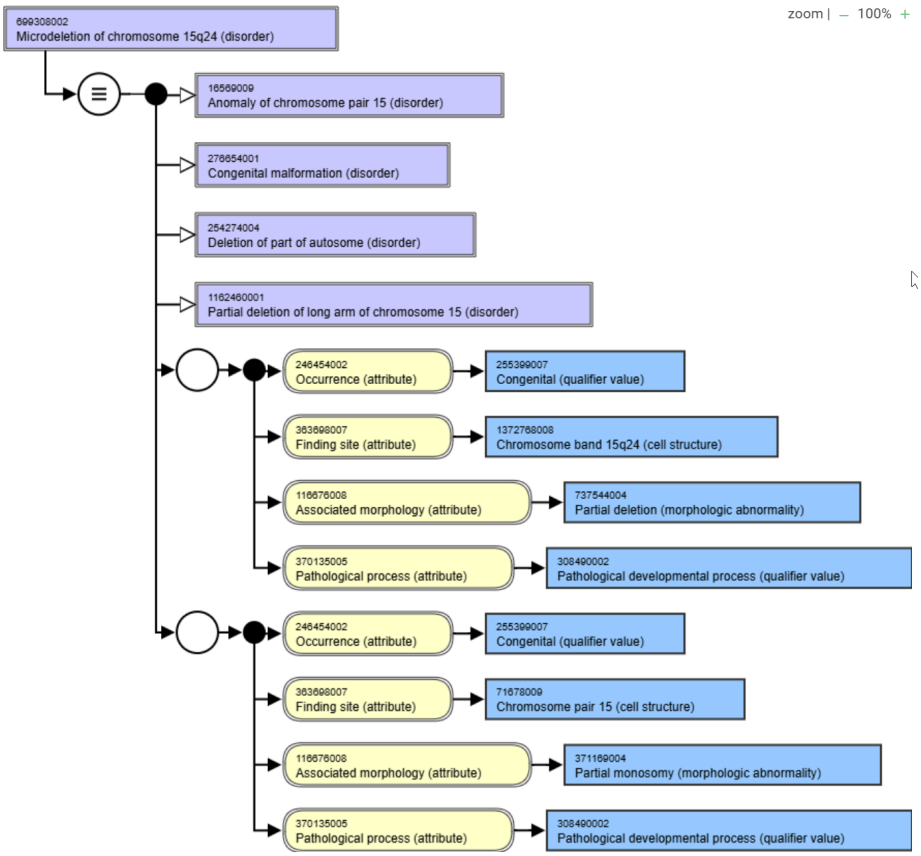
Example 2: 92824003 |Neurofibromatosis type 1 (disorder)|



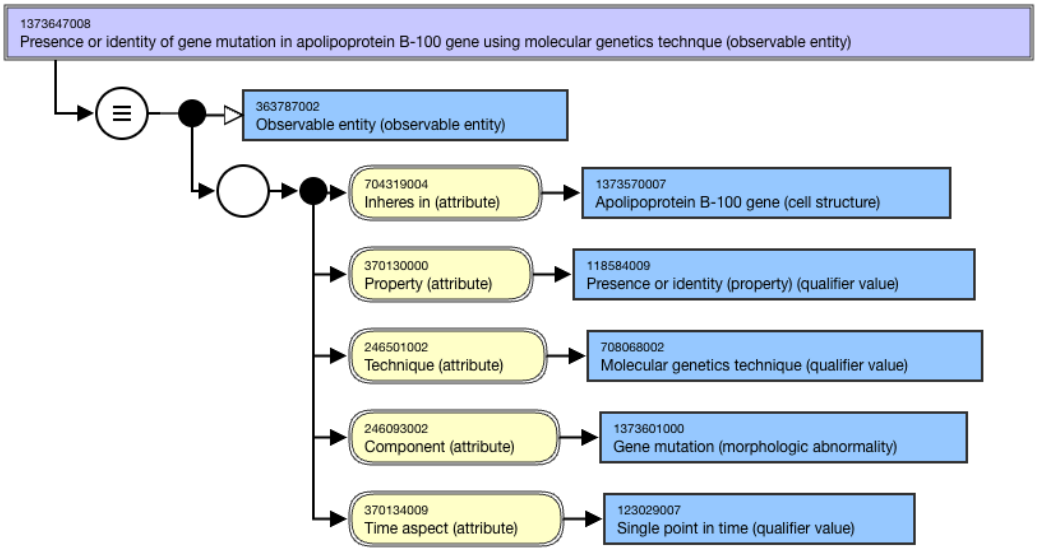
Example 3: 765057007 BRCA1 associated protein tumor 1 predisposition syndrome



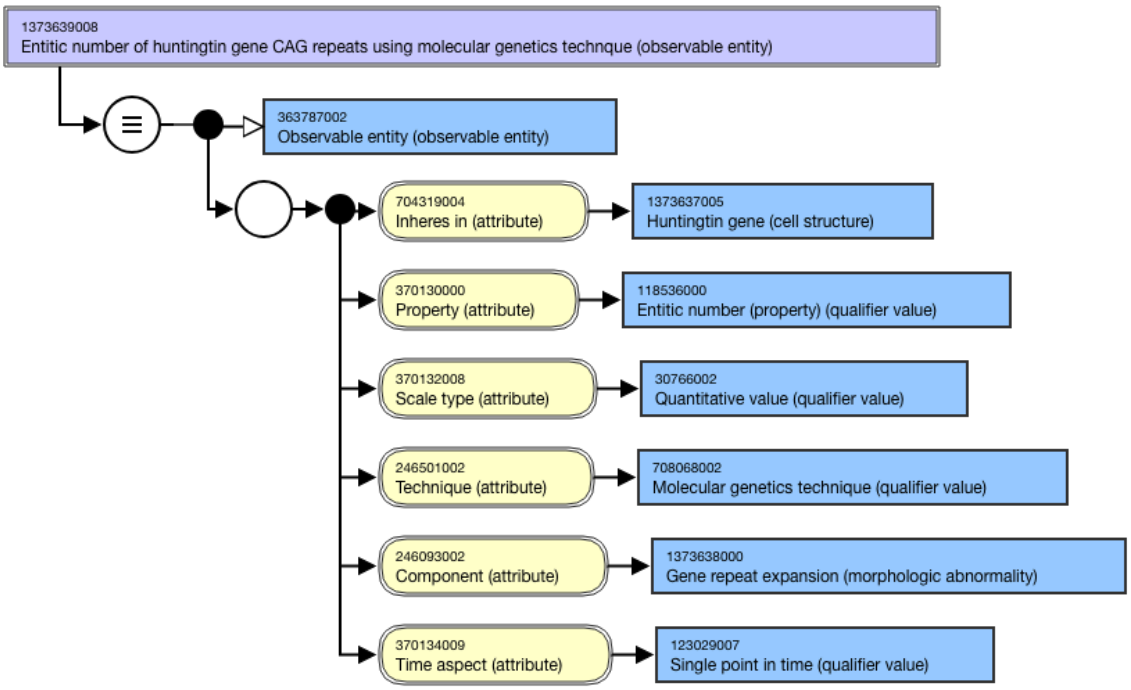
Example 4: 699308002 |Microdeletion of chromosome 15q24 (disorder)|



Example 5: 444032005| Detection of mutation in apolipoprotein B-100 gene (procedure)
*Modeled as an observable in accordance with SNOMED CT Editorial Policy



Example 6: 437742006|Huntington disease gene mutation carrier detection test (procedure)|
*Modeled as an observable in accordance with SNOMED CT Editorial Policy (Similar to LOINC code 53782-9)



Example 7: 738541005 |Thiopurine S-methyltransferase normal metabolizer (finding)|