Release Notes Table of Contents

[0. Introduction 2](#_Toc528862914)

[1. Navigating the Ontology 2](#_Toc528862915)

[Ontology structure 2](#_Toc528862916)

[Top-level indices 3](#_Toc528862917)

[Author index 3](#_Toc528862918)

[Disease index 3](#_Toc528862919)

[Intervention index 3](#_Toc528862920)

[Reference Index 3](#_Toc528862921)

[Regimen Index 3](#_Toc528862922)

[Medication Annotations 5](#_Toc528862923)

[Regimen Properties and Annotations 6](#_Toc528862924)

[Regimen Properties 6](#_Toc528862925)

[Regimen Annotations 8](#_Toc528862926)

[Reference Properties and Annotations 10](#_Toc528862927)

[Reference Properties 10](#_Toc528862928)

[Reference Annotations 10](#_Toc528862929)

[Author Annotations 11](#_Toc528862930)

[Example: find all rituximab-containing regimens and R-CHOP in particular 12](#_Toc528862931)

[Step 1: Open the ontology 12](#_Toc528862932)

[Step 2: Find rituximab 13](#_Toc528862933)

[Step 3: Open up the rituximab entity 14](#_Toc528862934)

[Step 4: Find all regimens in which rituximab is used 15](#_Toc528862935)

[Step 5: Find R-CHOP 16](#_Toc528862936)

[Step 6: Open up the R-CHOP regimen 17](#_Toc528862937)

[Step 7: Explore the supportive medications. 18](#_Toc528862938)

[Step 8: Explore a reference 19](#_Toc528862939)

[Step 9: Explore an author 20](#_Toc528862940)

[2. New features 21](#_Toc528862941)

[3. Bugs fixed 21](#_Toc528862942)

[4. To be completed 21](#_Toc528862943)

[5. License 21](#_Toc528862944)

[6. Citing this work 22](#_Toc528862945)

[7. Prior release stats 22](#_Toc528862946)

[Version 2018-08-25 22](#_Toc528862947)

[Version 2018-05-08 22](#_Toc528862948)

[Version 2018-02-02 23](#_Toc528862949)

[Version 2017-12-05 23](#_Toc528862950)

[Version 2017-11-12 23](#_Toc528862951)

[Version 2017-10-22 23](#_Toc528862952)

[Version 2017-10-15 23](#_Toc528862953)

[8. Prior release new features 23](#_Toc528862954)

[Version 2018-08-25 23](#_Toc528862955)

[Version 2018-05-08 23](#_Toc528862956)

[Version 2018-02-02 24](#_Toc528862957)

[Version 2017-12-05 24](#_Toc528862958)

[Version 2017-11-12 24](#_Toc528862959)

[Version 2017-10-22 24](#_Toc528862960)

[Version 2017-10-15 25](#_Toc528862961)

[9. Prior bugs fixed 25](#_Toc528862962)

[Version 2018-08-25 25](#_Toc528862963)

[Version 2018-05-08 25](#_Toc528862964)

[Version 2018-02-02 26](#_Toc528862965)

[Version 2017-12-05 26](#_Toc528862966)

[Version 2017-11-12 26](#_Toc528862967)

[Version 2017-10-22 26](#_Toc528862968)

[Version 2017-10-15 26](#_Toc528862969)

# 0. Introduction

HemOnc.org is the largest freely available wiki of chemotherapy drugs and regimens. This derived OWL-based ontology is an ongoing effort to formally represent most of the content of HemOnc.org in a computable format, mapping to established standards wherever possible. As of this version, the ontology contains:

* **222,542** axioms
* **30,513** individuals (drugs, regimens, authors, etc.)
  + 1,365 distinct treatment regimens
  + 3,665 references
  + 24,935 authors
* **1,471** classes and subclasses (categories of drugs, regimens, etc.)

# 1. Navigating the Ontology

The HemOnc.org ontology is meant to be self-explanatory, and this section gives an overview with some examples of how to browse the ontology. All examples are using Protégé, and this section will be expanded over time. Users are expected to have a passing knowledge of Protégé, but if not, [excellent tutorials are available online](https://protegewiki.stanford.edu/wiki/Protege4UserDocs).

## Ontology structure

Currently, the ontology uses four native OWL datatypes:

* **Individuals:** anything that is a leaf (terminating) node of a hierarchy is instantiated as an individual entity (purple diamond in the default view of Protégé); this includes all regimens, medications, authors, and references.
* **Classes:** anything that is a non-terminating node in a hierarchy is instantiated as a class (yellow circle in the default view of Protégé); this includes categories of medications, etc.
* **Object properties:** certain individuals are assigned properties that explicitly link them to other entities; e.g., individual medications are properties of regimens
* **Annotation properties:** certain individuals also have annotations, such as the number of authors of a reference, synonyms for drug names, etc.

## Top-level indices

The ontology is currently organized into five top-level classes/indices. These are meant to be mutually exclusive, i.e., an entity (individual or subclass) that is a member of one of the top-level classes can’t belong to another top-level class.

Author index: this index contains the authors of any extracted references (see below). The authors are organized by the disease(s) in which the regimen is studied.

Disease index: a polyhierarchy of the diseases represented on HemOnc.org, grouped by the clinical context in which they are usually treated. For example, gastrointestinal malignances are grouped, and gynecologic malignancies are grouped. With some exceptions, each disease subtype belongs to just one group.

Intervention index: a polyhierarchy of the interventions present on HemOnc.org, organized into four major subclasses: 1) biologic products (e.g., coagulation factor replacements); 2) devices; 3) drugs; and 4) radiation. See below for a list of properties and annotations.

Reference Index: this index contains the references that are cited by the treatment regimens (below). Nearly every treatment regimen on HemOnc.org has at least one reference, with many having multiple references. Generally, these references are restricted to prospective clinical trials with at least some demonstrated efficacy for one or more of the arm studied; see [HemOnc.org’s Eligibility Criteria](https://hemonc.org/wiki/Eligibility_criteria) for more details. For the purposes of the ontology, references that are published in abstract form only are excluded.

Regimen Index: a polyhierarchy of the treatment regimens present on HemOnc.org, organized by a variety of definitions such as the class of regimen (**Regimens\_by\_Class**; see **Table 1**) and the episode in which the treatment is given (**Regimens\_by\_Context**; see **Examples** below). Note that exact numbers are subject to change as content on the website is updated.

|  |  |
| --- | --- |
| **Table 1.** Regimens by class | |
| **Class of regimen** | **Number of distinct regimens** |
| Chemotherapy regimens | 1,170 |
| Chemoimmunotherapy regimens | 29 |
| Chemoradiotherapy regimens | 86 |
| Endocrine therapy regimens\* | 26 |
| Growth factor therapy regimens | 12 |
| Immunosuppressive therapy regimens | 33 |
| Immunotherapy regimens | 31 |
| Radiotherapy regimens | 41 |
| Other regimens | 134 |

\* we’re in the process of renaming endocrine therapy to hormonotherapy so this number is low because “hormonotherapy” is currently mapping to Other regimens.

## Medication Annotations

Medications (drugs) are the base unit for most antineoplastic regimens on HemOnc.org. Each medication is primarily named by its most common generic name, and can have one or more of the following annotations:

RxNorm**:** this is the RxNorm unique identifier (RXCUI) for a medication, if it’s available. For example, the RXCUI for **paclitaxel** is RXCUI:56946. Note that some medications, especially those newly approved, approved only in non-US countries, or under investigation in clinical trials, do not have an associated RXCUI. **Currently, 453 of 527 drugs (86%) have an associated RxCUI.**

Drug\_Code\_Name: this is the code name(s) used in development of the drug. For example, **imatinib** was known as **STI-571** prior to assignment of a generic name. Some drugs may have had more than one code name; e.g., **vemurafenib** was known as **RO5185426**, **RG7204**, and **PLX4032** during different periods of its development.

Drug\_Generic\_Name: these are alternate generic names for a drug that is better known by another generic name. Example: **aclarubicin** is also known as **aclacinomycin**.

Drug\_Brand\_Name: these are brand names for a drug. Example: **anastrozole** has 13 brand names listed on HemOnc.org.

Drug\_Brand\_Name\_Preferred: this is the brand name that we have determined is the preferred brand name, which is for the most part US-centric. Example: **anastrozole** has the preferred brand name of **Arimidex** in the US.

Drug\_Alias: certain medications on HemOnc.org have unorganized tables of drug aliases; in this case, the synonyms are not categorized as code, generic, or brand names but generally annotated as aliases; over time, we plan to eliminate this generic category.

Sig: this is the prescription for the drug, of the following general format: *1.* *Dose (Cap) 2. Route 3. Infusion time 4. Daily timing 5. Schedule timing 6. Sequence 7. (Notes)*. Here are several examples:

1. 100 mg PO BID
2. 1.4 mg/m<sup>2</sup> (maximum dose of 2 mg) IV once on day 1
3. 50 mg/m<sup>2</sup> IV over 30 minutes once on day 1
4. 2400 mg/m<sup>2</sup> IV continuous infusion over 46 hours, started on day 1, given third

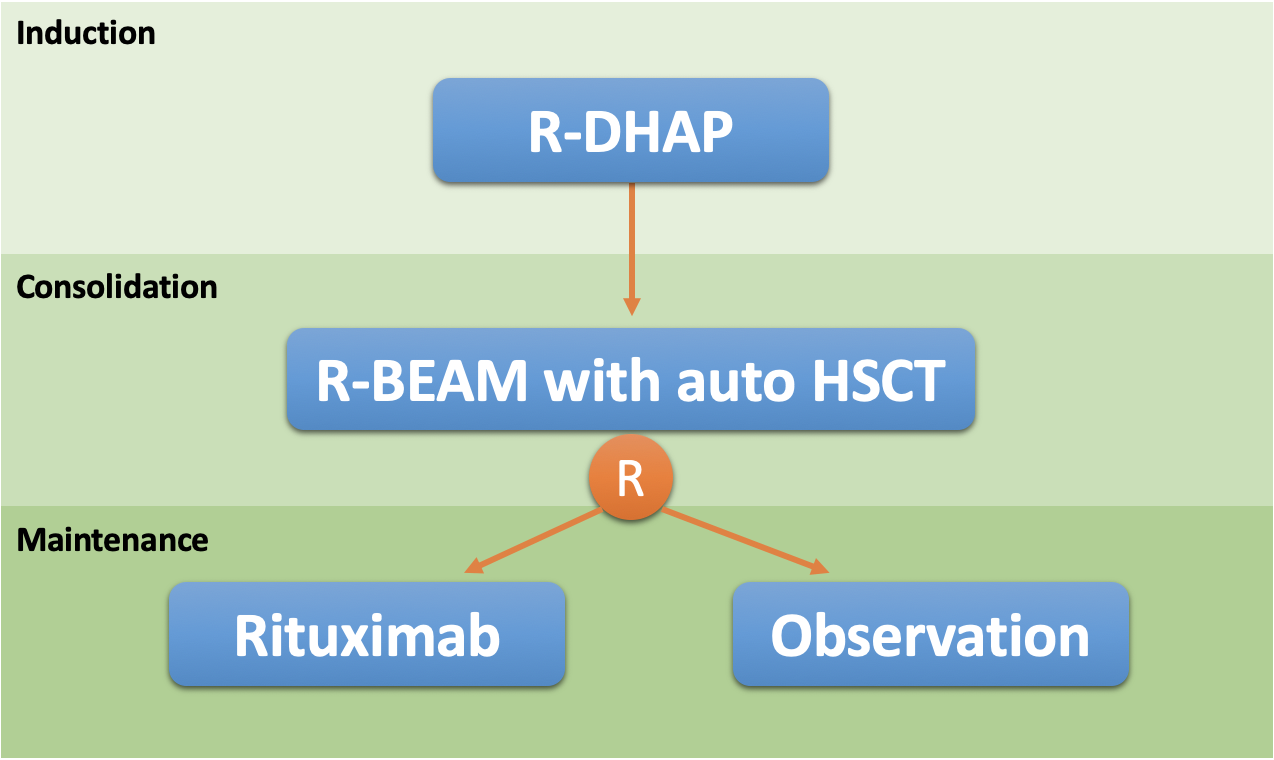
**Note that this data element is newly added to this release, and represents a major addition to the ontology; we have begun with the antineoplastics.** For now, the prescriptions are completely post-coordinated, but future work will focus on decomposing them into their respective components. Note also that some drugs have a large number of associated prescriptions, which reflects the underlying complexity of the chemotherapy regimens. Eventually, specific prescriptions will be tied back to specific regimens, but for now they are represented as unordered annotations of the drug entity.

|  |  |
| --- | --- |
| Table 2. Medication-level annotations and number of instances | |
| **Annotation** | **Number of instances** |
| RxNorm | 453 |
| Drug\_Code\_Name | 287 |
| Drug\_Generic\_Name | 343 |
| Drug\_Brand\_Name | 2,102 |
| Drug\_Brand\_Name\_Preferred | 263 |
| Drug\_Alias | 2,269 |
| Sig (Drug prescription instruction) | 3,988 |

## Regimen Properties and Annotations

Regimen Properties: In addition to their location in the regimen index polyhierarchy, there are seven properties that can be assigned to regimens:

1. hasAntineoplasticComponentOf**:** this is the property that associates antineoplastic drugs, i.e., drugs that exert a direct or indirect cytotoxic effect, with their respective regimens. For example, the regimen **AC** has antineoplastic components of **Doxorubicin** and **Cyclophosphamide**.
2. hasSupportiveComponentOf**:** this is the property that associates supportive drugs (or categories of drugs, as many published regimen protocols are nonspecific in what they recommend for supportive care) with their respective regimens. For example, many paclitaxel-containing regimens will have an associated supportive steroid, H1-blocker, H2-blocker, and antiemetic.
3. hasCNSTherapyComponentOf**:** this is the property that associates antineoplastic and/or supportive drugs that are used for CNS prophylaxis or treatment, with their respective regimens. For example, the regimen **R-HyperCVAD/R-MA** uses intrathecal cytarabine and methotrexate.
4. hasImmunosuppressiveComponentOf: this is the property that associates immunosuppressive drugs with their respective regimens. Take note that some drugs might be antineoplastic or immunosuppressive (or both), depending on the context in which they are used. For example, rituximab used to treat follicular lymphoma is antineoplastic, whereas rituximab used to treat immune thrombocytopenic purpura is immunosuppressive.
5. hasPrecedingTreatmentOf: this is one of the two properties that links multipart regimens. For example, Le Gouill et al. 2017 (LyMa) specifies a first-line regimen for mantle cell lymphoma with an induction component (**R-DHAP**), a consolidation component (**R-BEAM with autologous hematopoietic stem cell transplant [auto HSCT]**), and a maintenance component (**rituximab monotherapy** versus **observation**). In this example, **R-BEAM with auto HSCT** *hasPrecedingTreatmentOf* **R-DHAP**, and both **rituximab monotherapy** and **observation** *hasPrecedingTreatmentOf* **R-BEAM with auto HSCT**. This progression through components of the regimen is depicted here:



1. hasSubsequentTreatmentOf: this is the paired property to the above, to link multipart regimens. Many regimens have complex conditional requirements before proceeding from one part to the next (e.g., consolidation is contingent on having a certain response to induction treatment, etc.). In fact, the “simple” regimen described above (LyMa) has a condition requirement of achieving a complete response to progress from the induction to the consolidation phase. Conditional transitions are not yet captured in the ontology. Note also that surgery is not yet captured as an entity in the ontology, such that the fact that adjuvant regimens are preceded by surgery is not yet correctly represented.
2. hasReferenceOf: the vast majority of regimens on HemOnc.org have at least one supporting reference, and some have many more than that. For the purposes of the ontology, this property is reserved for *published prospective clinical trials*; this is distinct from the website which also contains some references to meta-analyses, reviews, abstracts, and retrospective case series.

|  |  |
| --- | --- |
| Table 3. Regimen properties and number of instances. | |
| **Regimen Property** | **Number of Instances** |
| hasAntineoplasticComponentOf | 3,675 |
| hasSupportiveComponentOf | 1,205 |
| hasCNSTherapyComponentOf | 146 |
| hasImmunosuppressiveComponentOf | 136 |
| hasPrecedingTreatmentOf | 724 |
| hasSubsequentTreatmentOf | 747 |
| hasReferenceOf | 5,882 |

### Regimen Annotations

In addition to properties, there are several annotations for regimen entities:

regimen: this is a logical annotation that is **TRUE** when an entity is a regimen.

biomarker**:** this annotation is used to tag regimens that are used within the context of a particular protein variation or over-expression. For example, the regimen **Dabrafenib & Trametinib** is annotated with BRAF-mutated in NSCLC, BRAF-mutated in melanoma, and BRAF-mutated in thyroid cancer since it is used in all of those contexts. This is a work in progress and does not include self-evident facts such as that all regimens used to treat acute promyelocytic leukemia are used in the context of *PML-RARA* mutations. Current biomarker tags include:

1. ALK-positive in NSCLC
2. BRAF-mutated in: melanoma, NSCLC, thyroid cancer, and disease-agnostic
3. BRCA-mutated in breast cancer
4. EGFR-mutated in NSCLC & disease-agnostic
5. ERBB2-mutated (disease-agnostic)
6. ERBB3-mutated (disease-agnostic)
7. ER/PR-positive in breast cancer
8. FLT3 positive in AML
9. Hedgehog pathway mutated (disease-agnostic)
10. HER2 positive in breast cancer
11. IDH-mutated in AML
12. KRAS wild-type in colon cancer
13. MSI-H or dMMR (disease-agnostic)
14. NRAS-mutated in melanoma
15. Philadelphia chromosome (Ph) positive in B-ALL
16. ROS1-positive in NSCLC
17. Triple-negative in breast cancer (TNBC; i.e., no expression of ER, PR, or HER2)

empty**:** this means that the regimen does not contain any recognized entities (e.g., medications). This may be legitimate (e.g., **Observation** is a “treatment” without any actual treatment) or may represent content that has not yet been completed.

historical**:** this means that the regimen is historical or obsolete *for a specific treatment context*. It is possible that the regimen may still be considered current in other contexts. For example, **CHOP** is no longer used in the treatment of B-cell lymphomas but is still current for T-cell lymphomas.

Regimen\_Alias: many regimens are known by aliases or short-hand. While some regimens are named by their conventional alias, e.g., **R-CHOP**, there may be other aliases by which the regimen is known (e.g., **CHOP-R** in this example). This annotation captures aliases which are included on the HemOnc.org website and the number of aliases will increase over time as we flesh out this particular content.

|  |  |
| --- | --- |
| Table 4. Regimen-level annotations and number of instances | |
| **Annotation** | **Number of instances** |
| regimen | 1,365 |
| biomarker | 148 |
| empty | 61 |
| historical | 184 |
| Regimen\_Alias | 550 |

## Reference Properties and Annotations

As explained above, references are named individuals in the Reference Index hierarchy. By convention, we use the trial name whenever possible, e.g., **NSABP\_B-28**. If we do not have a trial name available, the individual reference is named using the primary reference shorthand as described below, pasted to the primary reference URL (so as to exclude the possibility of merging two different references for different regimens, such as Smith et al. 2009 and Smith et al. 2009).

Reference Properties: at this time, the only properties that a reference has are related to its authors; this will likely expand over time, for example we plan to convert journal from an annotation to a property (at which point we will also create a journal index):

1. hasFirstAuthor: this is the first listed author.
2. hasLastAuthor: for references with two or more named authors, this is the last listed author. Note that in many cooperative group references, the “last author” is the cooperative group itself; we have removed these types of “authors” from the ontology.
3. hasMiddleAuthor: for references with three or more named authors (often many more than three), each middle author is referred to through this property.

|  |  |
| --- | --- |
| Table 5. Reference properties and number of instances. | |
| **Reference Property** | **Number of Instances** |
| hasFirstAuthor | 3,587 |
| hasLastAuthor | 3,577 |
| hasMiddleAuthor | 46,950 |

Reference Annotations: further information about references is available through annotations:

reference: this is a logical annotation that is **TRUE** when an entity is a reference.

NumberOfAuthors: this is a count of the total number of authors for a given reference.

PrimaryReference: this is the short descriptive name for a reference describing a regimen. For any given trial, we consider this to be *the first published manuscript with results*. By convention, the name is the last name of the first author plus “et al.” followed by the year of publication (or e-publication – whichever came first). For example, the shorthand for the first manuscript describing results for NSABP B-28 is **Mamounas et al. 2005**.

PrimaryReferenceJournal: this is the MEDLINE-approved abbreviation for the journal in which the reference is published.

PrimaryReferenceTitle: this is the title of the manuscript. As with journal names, there are some parsing errors here, which we are actively working on.

PrimaryReferenceURL: this is the direct hyperlink to the publication described by *PrimaryReference*, above. You can click the link within the Protégé browser and it will take you directly there. When a PubMed Central manuscript is available, we link to that; otherwise the links will generally take you to the original journal article on the parent journal’s website.

PubMedURL: most of the references included on HemOnc.org are also indexed in MEDLINE. When this is the case, we also include a direct hyperlink to the PubMed abstract. This link is also clickable.

YearOfPublication: this information is pulled directly from the PubMed API.

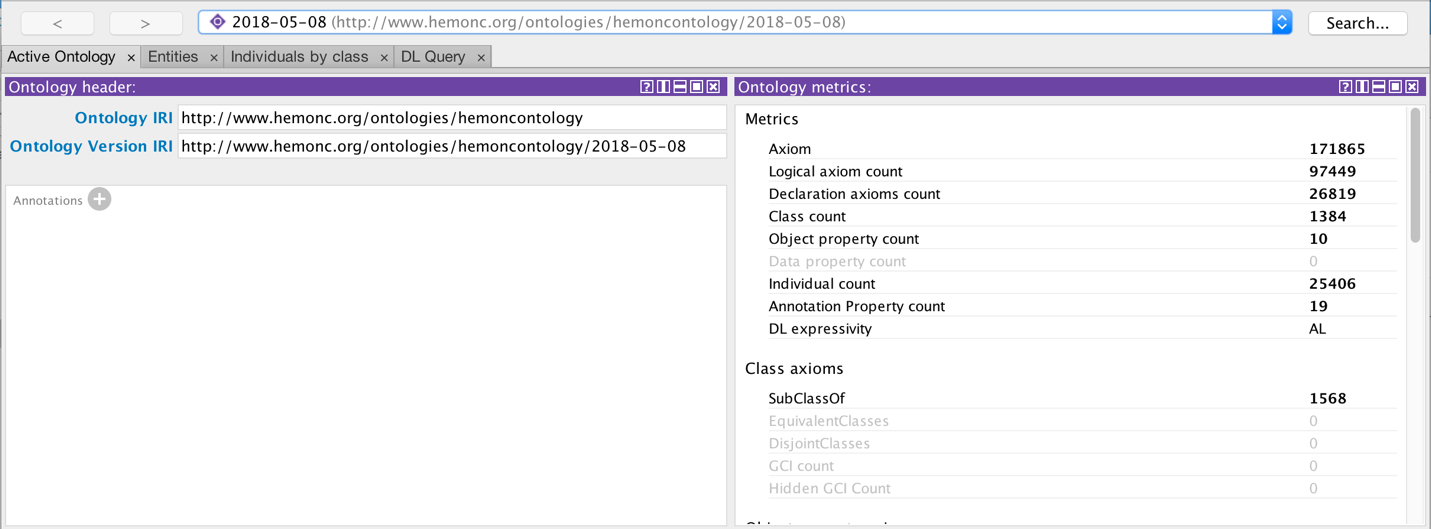
## Author Annotations

Right now, the only annotation for authors is a logical indicator, *author*, which is set to **TRUE**. In the future, we intend to leverage the PubMed and ORCID APIs to enrich this content, for example by separating first and last name, and adding institutional affiliations.

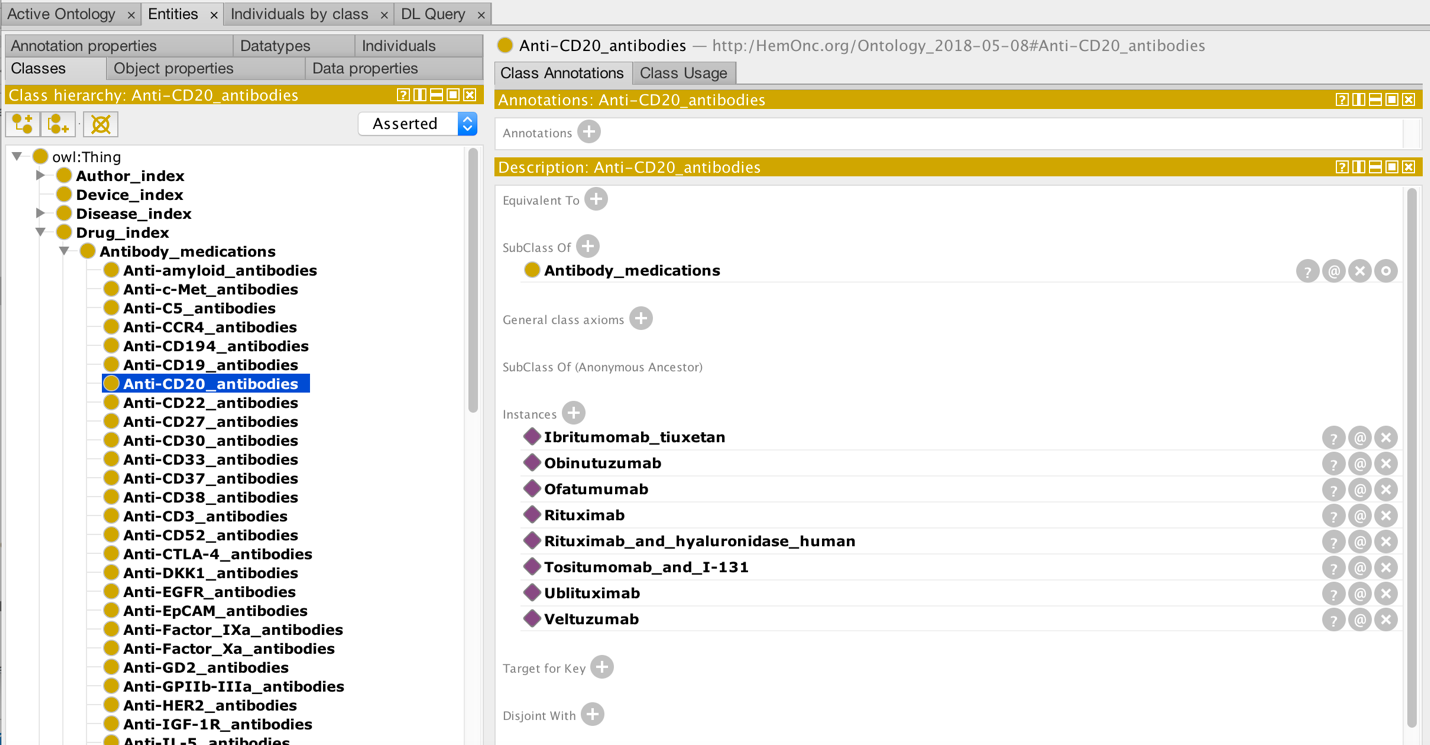
## Example: find all rituximab-containing regimens and R-CHOP in particular

As an example of how you can navigate the ontology (using Protégé), the following screenshots will show you how to quickly identify all regimens that contain the drug rituximab, find the rituximab-containing regimen R-CHOP, and learn more about references that support this medication. While this example is shown by “walking” the ontology, you can also use the “Search…” function to find concepts even more quickly! (*Note: these screenshots are from the 2018-05-08 version of the ontology but the appearance should be very similar to the current version*)

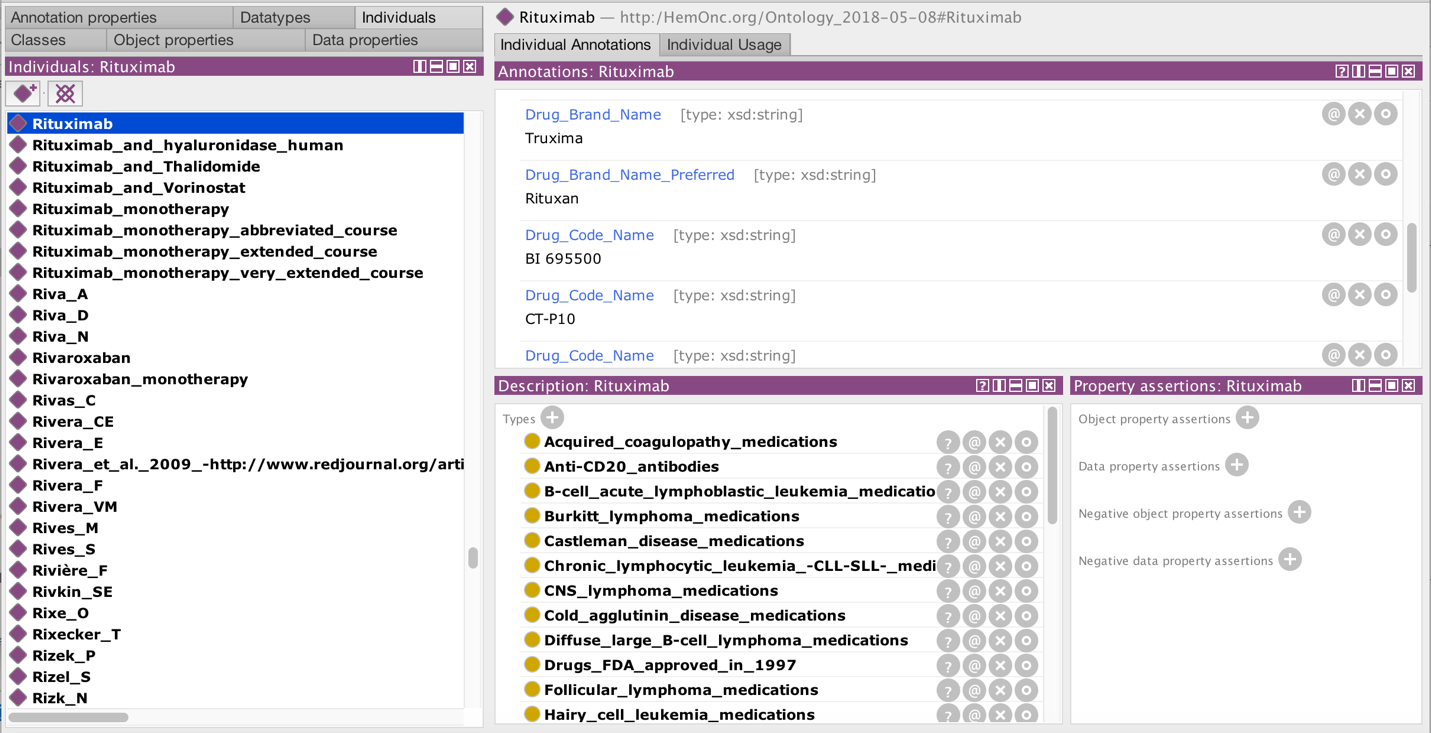
Step 1: Open the ontology. When first opening the OWL file, you should see the main landing page in Protégé as shown below. From here, click on the **Entities** tab.



Step 2: Find rituximab. This is done by clicking on the **Entities** tab, which defaults to the **Classes** tab (see arrow), then opening up **Drug\_index** and then opening **Antibody\_medications**. Finally, clicking on **Anti-CD20\_antibodies** brings up the panels on the right. Rituximab is the fourth medication in the list of eight anti-CD20 antibodies.

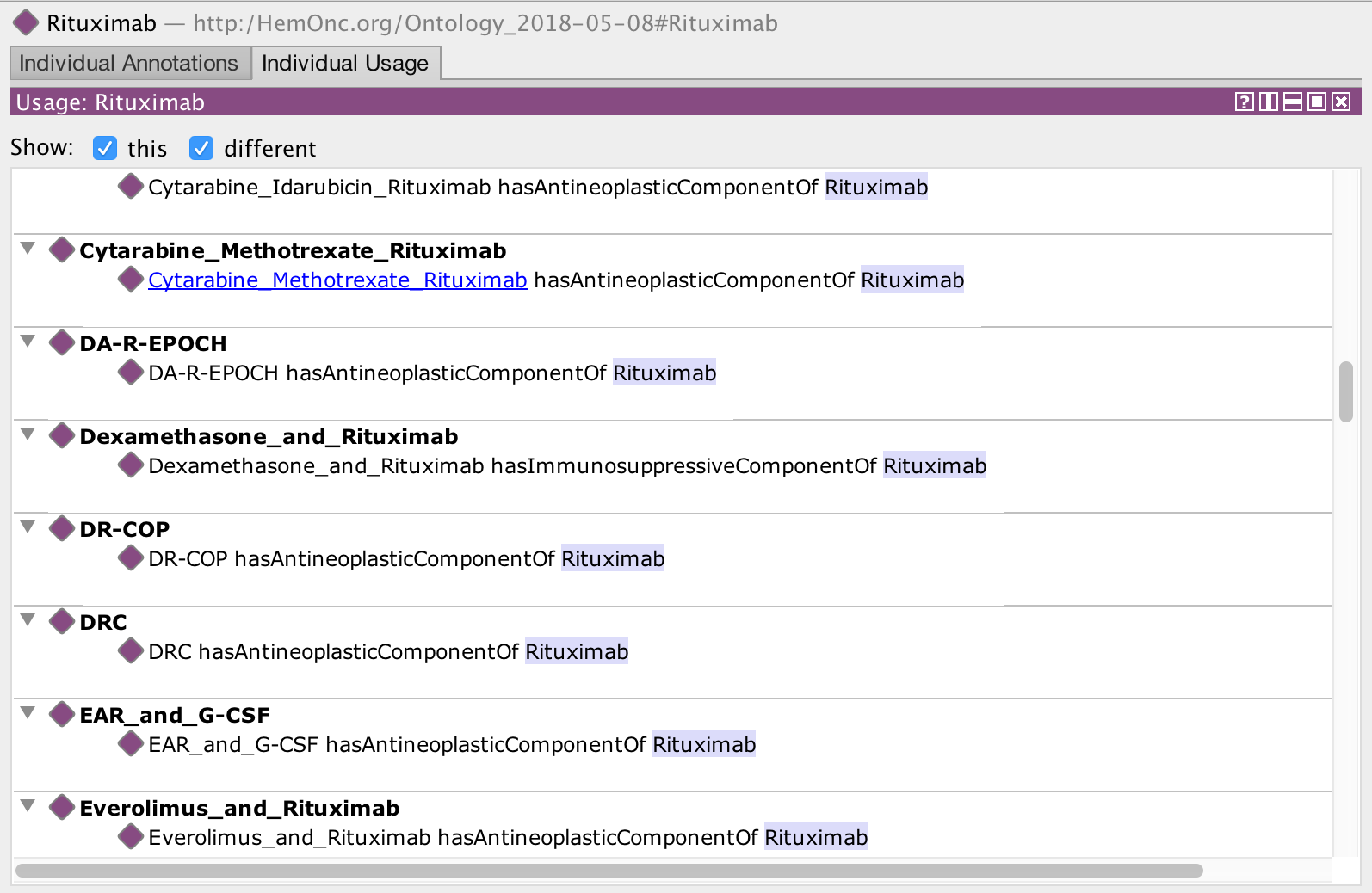


Step 3: Open up the rituximab entity. By clicking on the **Rituximab** link, you are taken to the view shown below. Note that you are now in the **Individuals** tab (see arrow), and all individuals in the ontology are listed on the left, in alphabetical order. You can see that there are a mix of drugs, regimens, authors, and references here. In the bottom right panel are shown just some of the many classes that rituximab belongs to. In the top right panel, some of the annotations for rituximab are shown.

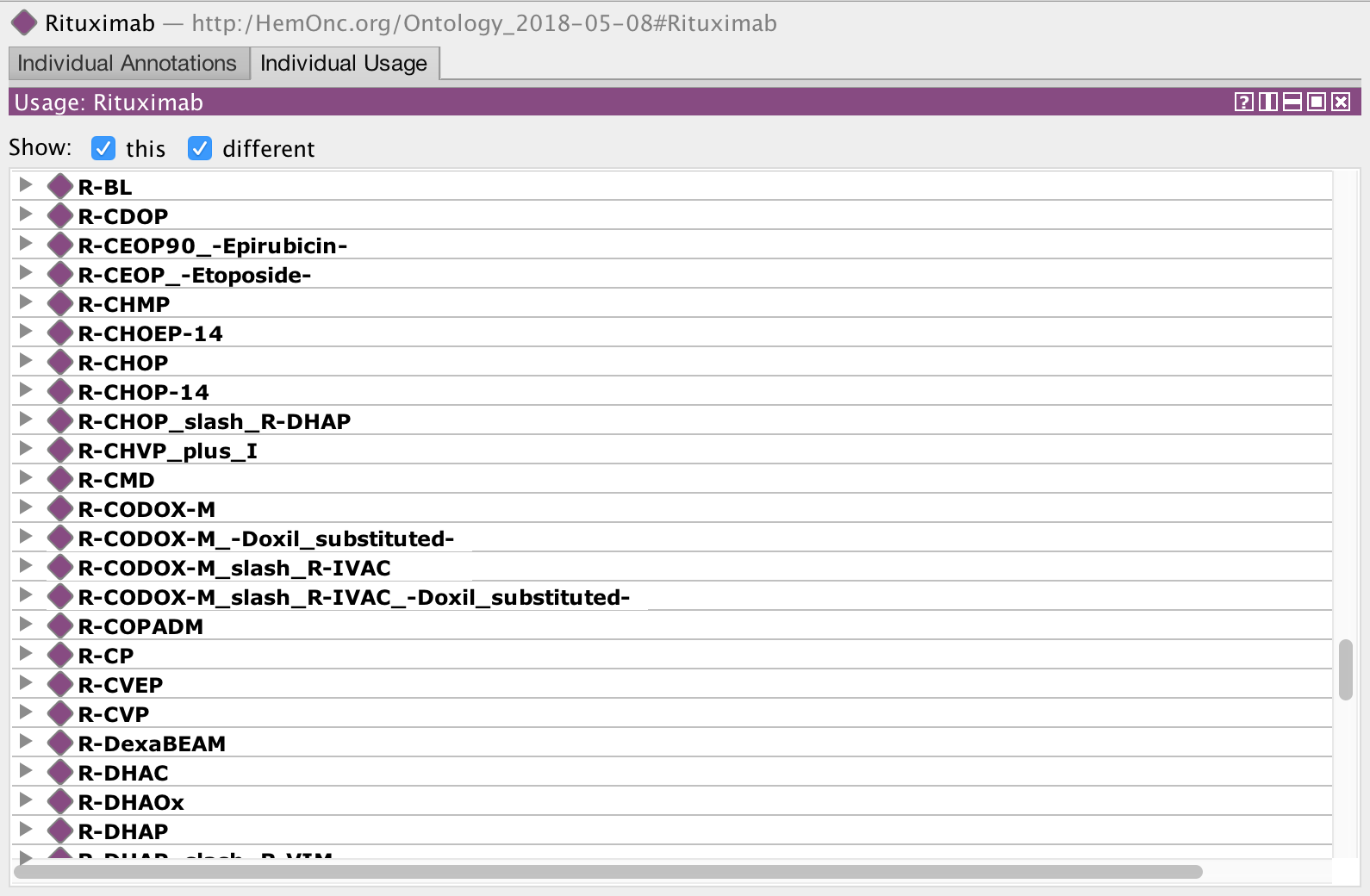


Step 4: Find all regimens in which rituximab is used. Just above the top panel, you can see a tab marked **Individual Annotations**, which is the default view. If you switch to the **Individual Usage** tab (see arrow), you will see the following screenshot. Note that in this screenshot, you can see regimens where rituximab is an antineoplastic component (e.g., DA-R-EPOCH), as well as a regimen where rituximab is an immunosuppressive component (dexamethasone & rituximab).

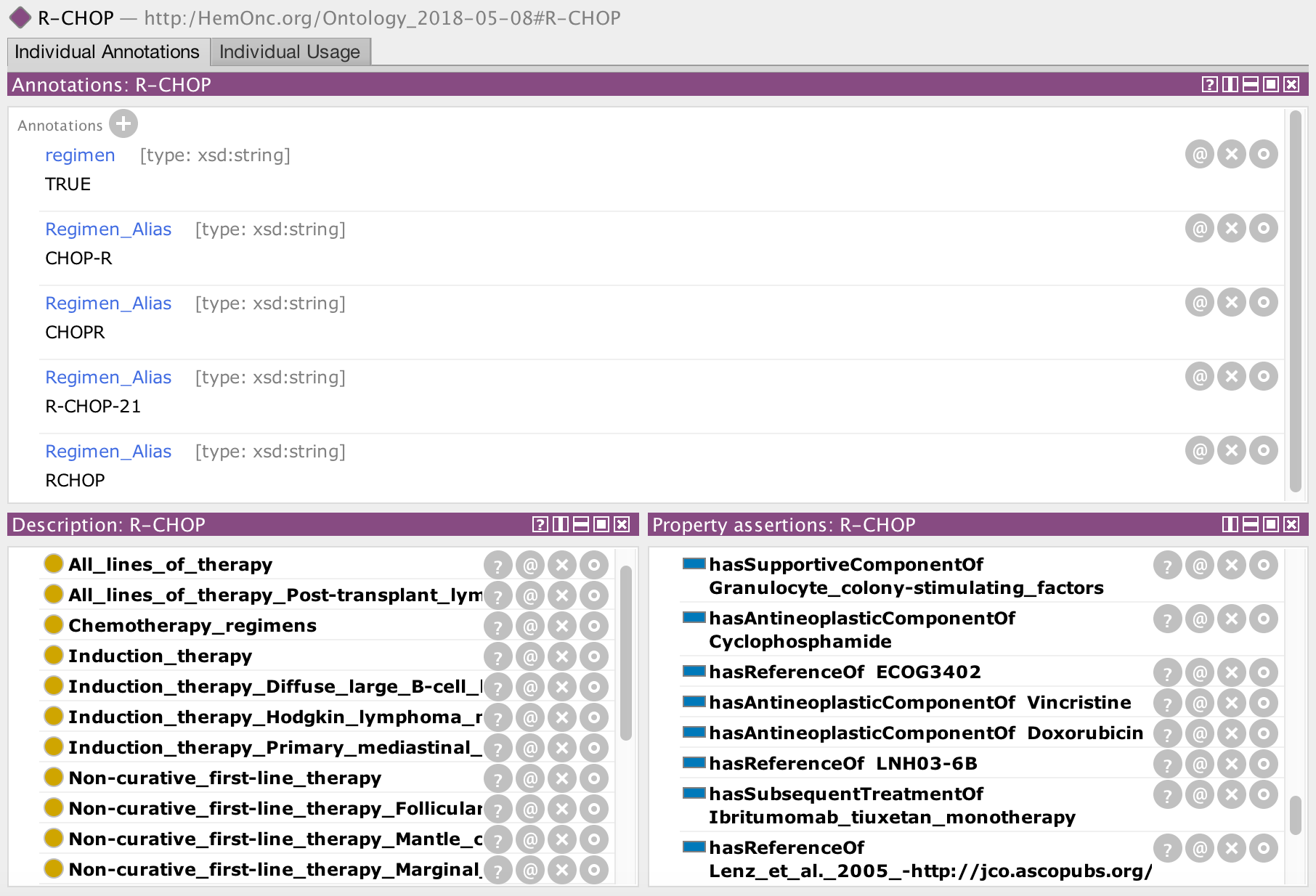
If you search for “hasAntineoplasticComponentOf Rituximab”, you will see that rituximab is used as an antineoplastic in 134 different treatment regimens within the ontology; if you search for “hasImmunosuppressiveComponentOf Rituximab” you will find three regimens.



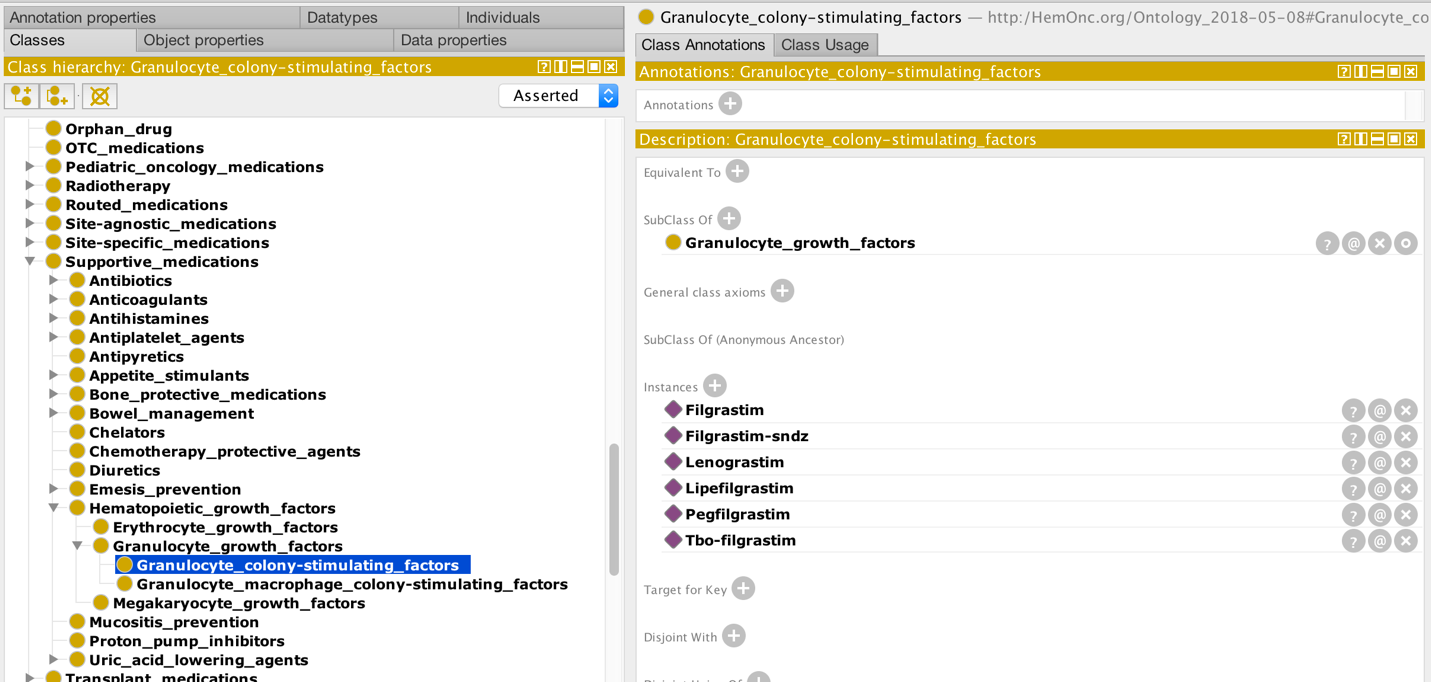
Step 5: Find R-CHOP. By scrolling through the individual usages, you eventually get to **R-CHOP**, as shown below. Note that you could have found R-CHOP directly by using the “Search…” function in the upper right of Protégé. Note that there are many regimens with names similar to R-CHOP but these are all distinct regimens.



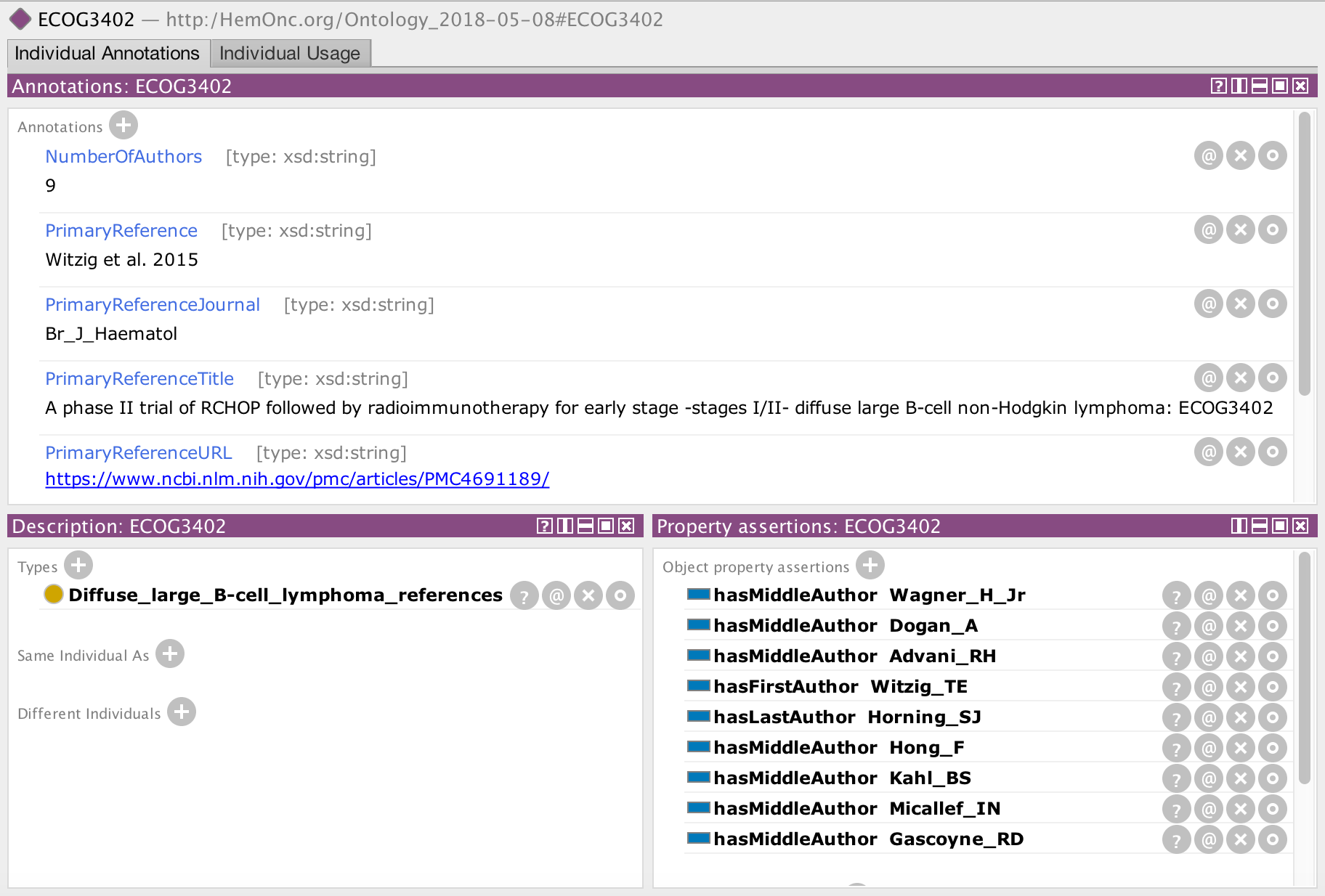
Step 6: Open up the R-CHOP regimen. By clicking on **R-CHOP** and then the **Individual Annotations** tab, you will see the screenshot below. Note that while annotations and parent classes are displayed in alphabetical order, the properties are not and are seemingly random; we’re working on fixing this so that properties are also in alphabetical order. Shown in the top panel are four regimen aliases (in addition to R-CHOP). In the bottom left panel are some of the many parent classes that R-CHOP belongs to. Note in particular that R-CHOP belongs to the generic class of **Induction therapy** and also to three disease-specific induction contexts: DLBCL, NLPHL, and PMBCL – these are considered to be potentially curative with chemotherapy alone. Conversely, other parent clasess are non-curative (e.g., first-line therapy for mantle cell lymphoma). In the bottom right panel you can see several of the antineoplastic and supportive medications that comprise various R-CHOP regimens; you can also see that a subsequent treatment is ibritumomab tiuxetan (in certain circumstances). Finally, you see several properties that are references (see step 8). All of these properties are clickable.



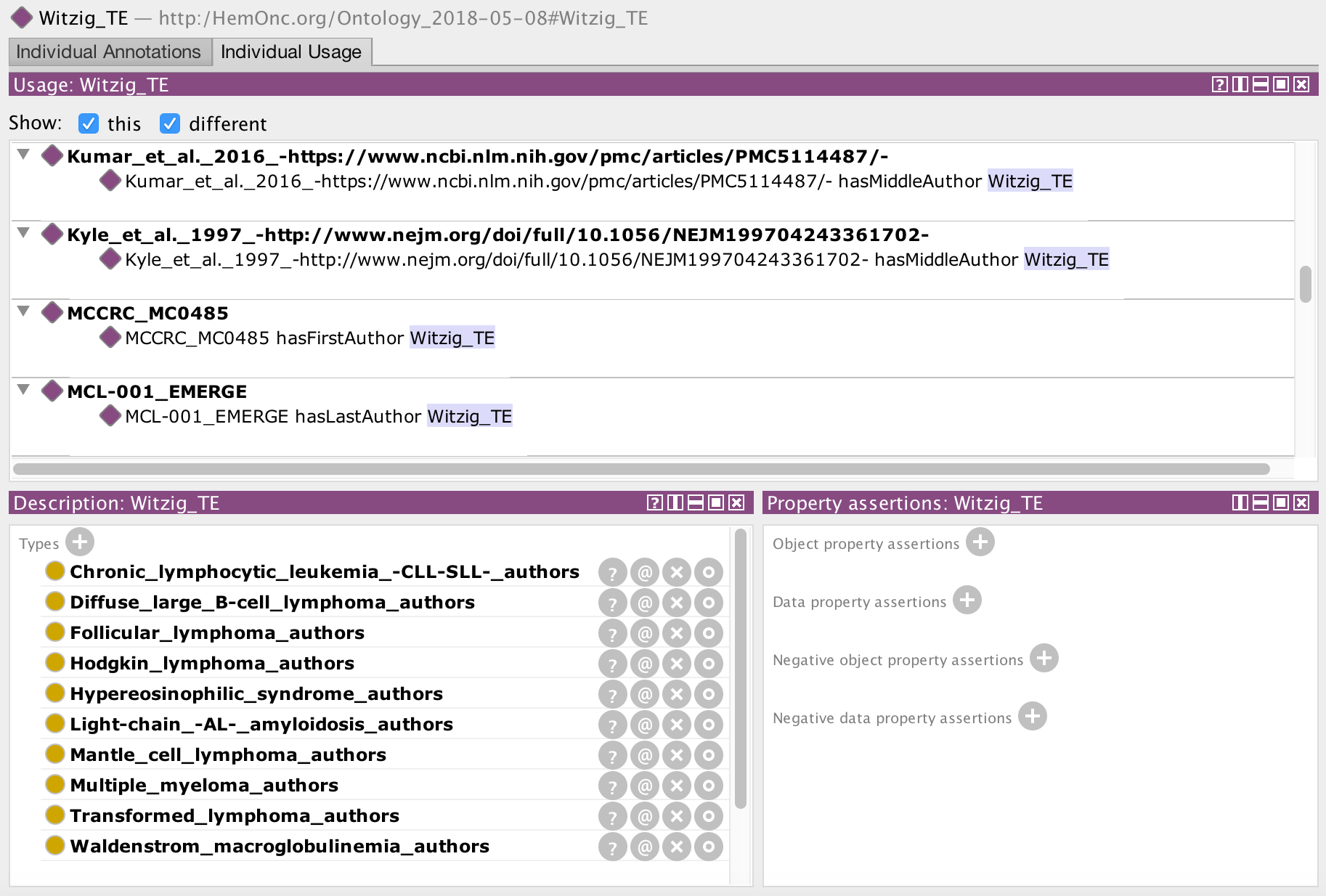
Step 7: Explore the supportive medications. You will notice that most, but not all, of the **hasComponentOf** properties of R-CHOP point towards specific drugs that comprise the regimen. However, there are also two categories of medications here: **Granulocyte\_colony-stimulating\_factors** and **PCP\_prophylaxis**. This reflects the fact that many regimens will specify classes or categories of supportive medications, rather than specific medications. Let’s explore this further, by clicking on **Granulocyte\_colony-stimulating\_factors**. This will bring you back to the class hierarchy, which is where you started in Step 2. Here, you will see that there are six granulocyte colony-stimulating factors.



Step 8: Explore a reference. Going back to the main R-CHOP regimen view, click on **ECOG3402** in the lower right panel. This will take you to a reference concept, which has a completely different set of annotations and properties, as shown below. At the top, you see several important pieces of metadata such as the number of authors, the journal in which the reference was published, and the title of the manuscript. The link under *PrimaryReferenceURL* is clickable; in this case the link will take you to the PubMed Central version of the article. On the lower left, you see that there is only one parent class, in this case DLBCL references. On the lower right, you see the list of authors assigned to one of three roles: first, middle, or last.



Step 9: Explore an author. Finally, click on the first author, Dr. Witzig. As you did in step 4, switch from **Individual Annotations** to the **Individual Usage** tab. Looking at the parent classes (lower left), you can see that Dr. Witzig has co-authored prospective clinical trials in numerous lymphoid and a few other malignancies. In the top panel, you see four of these trials; two in which Dr. Witzig was middle author, one where he was first author, and one where he was last author. This is the conclusion of the example tutorial.



# 2. New features

* Added prescription signature information tied to medications

# 3. Bugs fixed

* Minor bug fixes

# 4. To be completed

This is a partial list of items that we plan to address in future releases, in no certain order:

* Expand prescription signature information to include all classes of medications, not just antineoplastics
* Classify “other regimens” to the extent possible
* Capture surgeries as entities
* Transform journal names from annotations into properties
* Finish addressing how to handle regimens that specify categories of things (e.g., 5-HT3 antagonists)
* Formalize the connection between regimens and diseases
* Add a preferred name term to all individual entities
* Order words in classes to be consistently adjective\_subject
* Replace hasCNSTherapyComponentOf property with hasProphylacticComponentOf for CNS prophylaxis and hasLocalTherapyComponentOf property for CNS treatment?
* Replace hasAntineoplasticComponentOf with hasSystemicAntineoplasticComponentOf and hasLocalTherapyComponentOf
* Add expected duration of treatment property
* Add an annotation for regimens that are not stand-alone (i.e., they’re a component of a multipart treatment protocol)
* Store cooperative group information separately, when it occurs (e.g., EGOG, CALGB, Intergroup, etc.)
* Re-order properties to make them easier to read through
* Make sure that PMC and original articles are correctly labeled
* Revise temporal links to include conditional requirements (e.g., attainment of PR/CR)
* Add Pubchem information using the Pubchem API

# 5. License

The HemOnc.org ontology is licensed by version under the **CreativeCommons NonCommercial-ShareAlike 4.0 International License**. Under this license, users are free to:

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material

Under the following terms:

**Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

**NonCommercial** — You may not use the material for commercial purposes.

**ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

**No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

# 6. Citing this work

If you plan to publish a manuscript, book chapter, etc. using information derived from the HemOnc.org ontology, we would appreciate citation. Please cite:

Malty AM, Jain SK, Yang PC, Harvey K, Warner JL. Computerized approach to creating a systematic ontology of hematology/oncology regimens. ***JCO Clinical Cancer Informatics***. 2018 May 11. DOI: 10.1200/CCI.17.00142

This work is currently supported in part by National Institutes of Health (NIH) Grants No. U24 CA184407, U24 CA194215, and U01 CA231840. Previously, this work was supported by NIH Grant No. L30 CA171123. This work was conducted using the Protégé resource, which is supported by Grant No. GM10331601 from the NIH National Institute of General Medical Sciences.

# 7. Prior release stats

## Version 2018-08-25

* **193,215 axioms**
* 28,931 individuals (drugs, regimens, authors, etc.)
  + 1,280 distinct treatment regimens
  + 3,199 references
  + 22,440 authors
* 1,436 classes and subclasses (categories of drugs, regimens, etc.)

## Version 2018-05-08

* **171,865 axioms**
* 25,406 individuals (drugs, regimens, authors, etc.)
  + 1,275 distinct treatment regimens
  + 2,827 references
  + 20,778 authors
* 1,384 classes and subclasses (categories of drugs, regimens, etc.)

## Version 2018-02-02

* **31,257 axioms**
* 1,747 individuals (drugs, regimens, etc.)
* 4 top-level classes: drug index, device index, disease index, and regimen index
* 1,094 subclasses (categories of drugs, regimens, etc.)

## Version 2017-12-05

* **30,526 axioms**
* 1,728 individuals (drugs, regimens, etc.)
* 4 top-level classes: drug index, device index, disease index, and regimen index
* 1,192 subclasses (categories of drugs, regimens, etc.)

## Version 2017-11-12

* **24,742 axioms**
* 1,749 individuals (drugs, regimens, etc.)
* 4 top-level classes: drug index, device index, disease index, and regimen index
* 1,010 subclasses (categories of drugs, regimens, etc.)

## Version 2017-10-22

* **23,996 axioms**
* 1,728 individuals (drugs, regimens, etc.)
* 4 top-level classes: drug index, device index, disease index, and regimen index
* 1,097 subclasses (categories of drugs, regimens, etc.)

## Version 2017-10-15

* **19,358 axioms**

• 1794 individuals (drugs, regimens, etc.)

• 4 top-level classes: drug index, device index, disease index, and regimen index

• 1124 subclasses (categories of drugs, regimens, etc.)

# 8. Prior release new features

## Version 2018-08-25

* Changed the RxNav API call to include non-approved drugs
* Began to pull publication information from the PubMed API
* Added year of publication as an annotation to references

## Version 2018-05-08

* Added regimen aliases (e.g., CHOP-R is an alias for R-CHOP)
* Made references into entities (instead of annotations) with annotations and properties
* Created a new top-level reference index to contain reference entities
* Created a new top-level author index to contain author entities
* Created a subsequent treatment property to pair with the preceding treatment property
* Created uniform placebo, best supportive care, observation concepts

## Version 2018-02-02

* Preserved temporality in multipart regimens that are within the same treatment context (e.g., AC comes before Taxol) – for most disease types
* Changed “obsolete” concept to “historical” concept to better reflect the nature of these regimens

## Version 2017-12-05

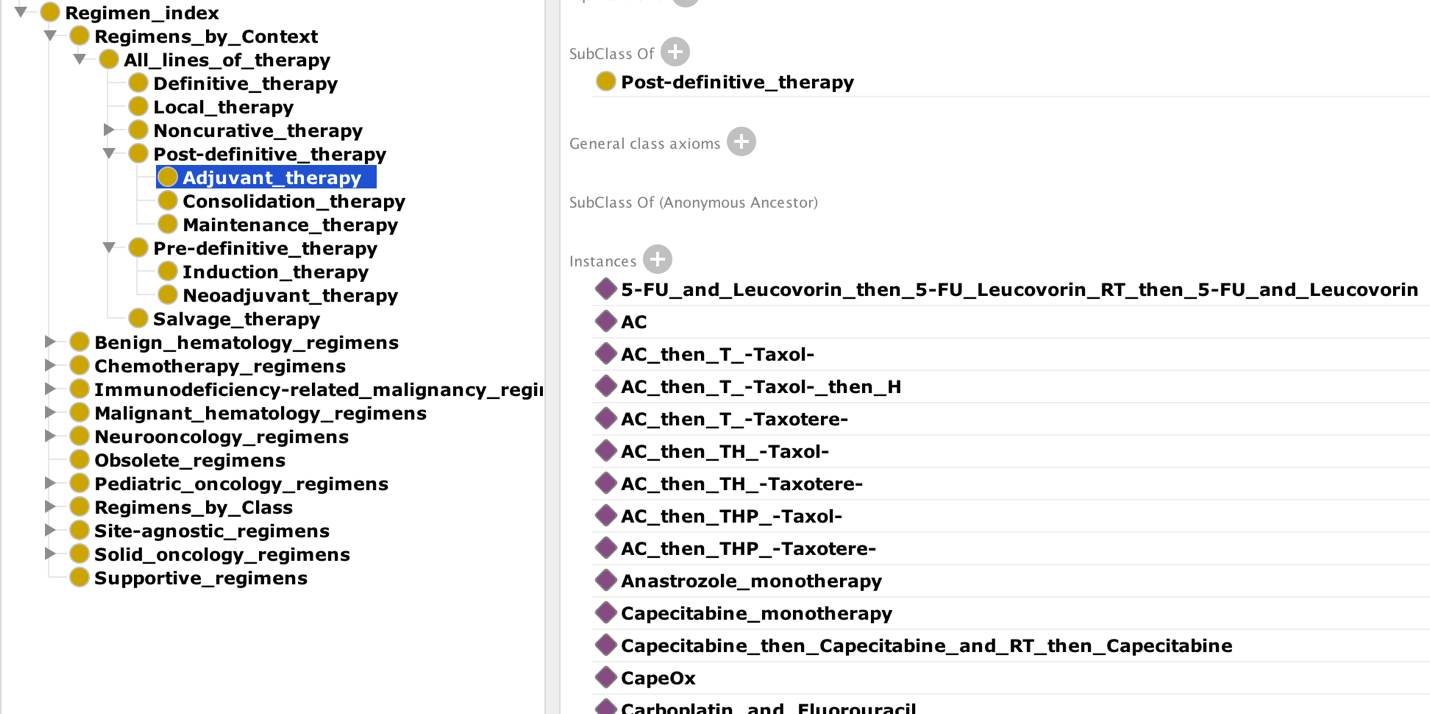
* Added over 5000 drug synonyms
* Began to address how to handle regimens that specify categories of things (e.g., 5-HT3 antagonists)
* Eliminated most duplicative regimen names
* Labeled medication entities simply by generic name, with brand name now as an annotation

## Version 2017-11-12

* Added property to distinguish CNS treatments from other antineoplastic treatments
* Added first-line, second-line, and third-line non-curative categories
* Split biomarker-specific pages and added a biomarker annotation to their regimens
* Added a property linking to preceding regimens for multipart regimens
* Added “empty” and “to be completed” annotations

## Version 2017-10-22

* Added a regimen-by-episode class to capture the therapeutic context in which treatment is given. Therapeutic contexts are nested within a “**Regimens by Context**” concept; see screenshot for an example of some of the >150 regimens belonging to the “**Adjuvant therapy**” class.



## Version 2017-10-15

* Added growth factor and immunosuppressive therapy to regimen categories
* Changed mappings to be more consistent with OWL conventions
* Mapped many headings to match episode categories (neoadjuvant, adjuvant, etc.)
* Added an obsolete regimen annotation at the regimen level
* Adopted a convention to replace most non-alphanumeric characters in the ontology:
  + “&” replaced with **and**
  + “+” replaced with **plus**
  + “/” replaced with **slash**
  + **commas** removed
  + “ö” replaced with **o**

# 9. Prior bugs fixed

## Version 2018-08-25

* Completed the elimination of duplicative regimen names
* Fixed bug where some preceding treatments weren't parsed correctly
* Fixed parser to capture components that don't come right after the bullet point

## Version 2018-05-08

* Fixed a bug that was pulling subsequent treatments into the main antineoplastic category
* Removed several “NA” references that resulted from parsing errors

## Version 2018-02-02

* Various minor bug fixes

## Version 2017-12-05

* Fixed the RxNorm API calling routine

## Version 2017-11-12

* Fixed a bug related to classification of first-line, second-line, and third-line

## Version 2017-10-22

* Fixed several issues where references weren’t being parsed correctly or represented correctly in the OWL format, leading to a notable increase in reference annotations.
* Commas removed from all classes and entities; slashes replaced with “-“ for all classes (the word “slash” is still used for entities); some white space issues fixed.
* Removed references to the steroid conversions and antiemetics reference pages from regimens; instead they now point to the categories of steroids and antiemetics.
* Fixed the "other" regimens-by-class of therapy category.

## Version 2017-10-15

• Fixed an issue of nesting regimen categories appropriately

• Fixed an issue where the OWL file was three times larger than it should have been