Cellular Microscopy Phenotype Ontology (CMPO): Quick tour

Gabry Rustici [1]

- Ontologies
- Cross-domain
- Beginner
- 0.5 hour

This quick tour provides a brief introduction to the Cellular Microscopy Phenotype Ontology (CMPO). It covers the scientific principles that led to the creation of this ontology and how it can be used.

Learning objectives:

- Gain a basic understanding of CMPO and the benefits it brings in annotating cellular phenotype data;
- How to browse the ontology and use it to annotate data.

What is CMPO?

The Cellular Microscopy Phenotype Ontology [2] (CMPO) provides a species-neutral controlled vocabulary [3] for describing phenotypic qualities relating to whole cells, cellular components, cellular processes and cell populations. CMPO terms can be used to annotate phenotype descriptions from high-content screening databases and cellular image repositories.

Why do we need CMPO?

Phenotype data associated with images from high-content screening (HCS) are often annotated using free-text descriptors. As a result, the same phenotype can be described using different words by various researchers. This makes it difficult to search for and to analyse images that have phenotypes in common. By annotating a dataset with an ontology we add value to our data, and allow for it to be exploited computationally to support smarter querying and facilitate data integration. Researchers can explore multiple datasets for results of interest and discern new scientific insights.

For example, a researcher could examine data on mouse knockouts to compare mouse phenotypes with data on human disease phenotypes and generate new hypotheses as to the underlying genetic causes of human disease.
What can you do with CMPO?

You can use CMPO to:

- Annotate cell phenotype data with broad category terms;
- Analyse annotated data to search for biologically relevant insights;
- Improve data discoverability by incorporating CMPO into existing web applications.

Figure 1: CMPO can be used to annotate phenotype data at different biological levels. By linking these different types of data by annotating them with the same controlled vocabulary [3], we can enable researchers to gain insight into disease mechanisms, and predict novel drug targets and biomarkers.

Accessing CMPO

You can access CMPO in a number of ways:

- Via the CMPO webpage [2] on the EMBL-EBI website - where you can find additional information, contact details, download the ontology [4] (in OWL [5] and OBO formats), search CMPO and request new terms to be added.
- Via the CMPO entry [6] on the BioPortal [7] repository or the EMBL-EBI's Ontology Lookup Service [8] (OLS) - both of which allow you to explore the ontology and find terms of interest.
Exploring CMPO

Searching CMPO in OLS

The Ontology Lookup Service (OLS) allows you to search for terms within and across a number of biomedical ontologies. We can use OLS to explore CMPO for concepts or terms that we are interested in.

First, select CMPO from the drop-down "Search Ontology" menu. Then begin typing in the Term Name box - you can input term names you already know, or use free text. The OLS search includes a partial search term function, so as soon as you start typing you will begin to see suggested terms. The image below shows a search in progress for the term "elongated".
The approach described above is useful if you want to find the CMPO terms to describe an observed phenotype. If you leave the "Search Ontology" drop-down menu showing the default "search in all ontologies" option, you can run a query across all the ontologies supported by OLS. This can help you to discover other ontologies with shared or similar terms; this is useful if you are planning to combine a number of ontologies or if you are integrating different annotated datasets (mouse and human disease data, for example).

If you have a previously annotated dataset that includes CMPO identifiers - CMPO:000xxxx - then you can use the "Simple Term ID Search" option shown above. Running a query on "CMPO:0000316" will return the CMPO entry for "abnormal cell growth" [9].

**Browsing CMPO in OLS**

We can also use OLS to browse the entire ontology, instead of individual terms, allowing us to navigate the ontology structure and see more granular phenotypes. To do this, select CMPO from the drop-down "Search Ontology" menu; then click the "Browse" button on the right. This will open a new page with a visualisation window where the entire ontology will appear in branch form. You can expand the ontology by clicking on the "+" icon to the left of a term; child terms will then be visible. You can click the "-" icon to collapse branches back.
Figure 4 Ontology branches down to "abnormal nucleus shape phenotype".

**CMPO term metadata**

Once you have located a CMPO term - using the "Search Ontology" [4]", "Simple Term ID Search" or the "CMPO Ontology Browser" options - you can learn more about the term by viewing the associated metadata [10]. When using the "Search Ontology" or "Simple Term ID Search" functions a box containing the metadata will appear below the search box. This includes information on: the definition of the term, the preferred name of the term, its broader synonyms, and links to PubMed [11] where the phenotype was reported (if available).
CMPO term hierarchy

CMPO also allows you to explore a term within the context of the entire ontology. By clicking on the "Browse" button next to the "Search Ontology" drop-down menu, you will be taken to the OLS Browser. On the right-hand side you will find a box that contains graphical representation of where the selected term sits in the ontology hierarchy and its relationships to neighbouring terms. In the example below "mitosis delayed phenotype" is a term at the base of the ontology tree with no child terms. If we were investigating "delayed process phenotype", then we would be able to browse terms both above and below our term of interest.
You can zoom the ontology browser by clicking on a term in the graph.
Annotating data with CMPO

Annotating datasets

As we have already seen there are many benefits to annotating data with ontology terms. Whilst a daunting prospect for those new to ontologies, annotating data is being made easier with simple tools designed to be used by biologists.

EMBL-EBI provides the Webulous tool suite for collaborative ontology development, allowing users to populate ontology design patterns via simple user interfaces. Webulous has been designed to support a scenario where knowledge about a domain is collected in spreadsheets and transformed into statements within an ontology according to pre-defined ontology design patterns. CMPO is available via a Webulous Google Add-on for Google Spreadsheets allowing annotation of your own data using drop-down validation boxes. Populated templates are then sent back to a server for processing. This approach can be used by domain experts to annotate their own data and generate a basic ontology themselves.

Automating annotation

Of course, you may want to annotate much larger datasets that would take a long time to complete by hand. To speed up the annotation of high-throughput -omics data we need to explore ways of automating the process. ZOOMA from EMBL-EBI is one such method; it aims to provide an automated service to take free-text data descriptors and then map them computationally to existing ontology terms. This removes the need for the user to conduct the mapping process.
Getting help and support on CMPO

Additional information on CMPO can be found at:

- CMPO About [15] webpage
- CMPO developers [16]' webpage

If you cannot find the information you require on these pages, you can also email the CMPO team [17] for help.

CMPO Development

CMPO was developed as part of the BioMedBridges project. More information about the need for the ontology [4] can be found in the BioMedBridges report on Mapping of standards and ontologies between the different image reference data sets [18].

Funding

CMPO and this e-learning module were developed as part of the BioMedBridges [19] project, funded by FP7 Capacities Specific Programme, grant agreement no. 284209.

![BioMedBridges](image)

Contributors

Gabry Rustici [1]

[1]
EMBL-EBI
Training coordinator, Functional Genomics

Gabriella Rustici was previously the Research and Training Coordinator in the Functional Genomics Group at the European Bioinformatics Institute in Cambridge, UK. She earned her PhD in Genetics from Cambridge University in 2004, working on transcription profiling of the fission yeast cell cycle, and has since been involved in the analysis of functional genomics data.

Tom Hancocks
EMBL-EBI
Scientific Training Officer - BioMedBridges

Tom Hancocks works as a Scientific Training Officer for the Training Team at EMBL-EBI.

He studied Human Genetics at the University of Leeds and McMaster University in Hamilton, Ontario; before completing an MSc in Analytical Genomics at the University of Birmingham.

Tom has worked for the NHS in diagnostic genetics and as a bioinformatics trainer for healthcare scientists and clinicians.

Source URL: http://www.ebi.ac.uk/training/online/course/cellular-microscopy-phenotype-ontology-cmpo-quick

Links
[1] http://www.ebi.ac.uk/training/online/trainers/gabry
[5] http://www.ebi.ac.uk/training/online/glossary/owl
[10] http://www.ebi.ac.uk/training/online/glossary/metadata
[16] http://www.ebi.ac.uk/cmpo/developers
[17] mailto:jupp@ebi.ac.uk,gabry@ebi.ac.uk?subject=CMPO%20Question%20(pan%20form)
[18] https://zenodo.org/record/14129#.VdSM3yxVhHw
[20] http://www.ebi.ac.uk/training/online/trainers/Tom