BioSamples: Quick tour

Luca Cherubin [1]

- DNA & RNA
- Gene Expression
- Ontologies
- Proteins

- Beginner
- 0.5 hour

This quick tour provides a brief introduction to the EBI's BioSamples data resource. Undergraduate-level understanding of biology would be an advantage.

Learning objectives:

- A basic understanding of the BioSamples resource, how to use it and how to find out more

What is BioSamples?

The BioSamples Database aggregates information for reference samples (such as cell lines or stock strains), as well as samples for which data exist in one of the EBI's other data resources (e.g. ArrayExpress [2], the European Nucleotide Archive [3], PRIDE [4]). Samples are described with a set of annotations. These annotations can be linked to ontology [5] (or controlled vocabulary [6]) terms, which provide synonyms and more specific terms. Samples can also have defined relationships to other samples e.g. cell line derived from a donor or parent/child relationships.

What is a BioSample?

For the BioSamples Database, the term 'BioSample' has a broad definition. Generally, each BioSample relates to a discrete physical object that is composed of biological material and has, or will have, assay [7] data associated with it. These include blood samples, cell cultures or individual organisms. However, a BioSample can also be an environmental sample (for instance, for meta-genomic analysis), a hybrid between two species, or a parasite culture.

What is a BioSample group?
The BioSamples Database also has a concept of a group. A group is a collection of samples with similar annotations, typically used together in a study, project or experiment. In cases where individual sample information cannot be made public (for example, as a result of ethical constraints) a group containing aggregated annotation may be used instead.

**What can I do with BioSamples?**

Using BioSamples, you can:

- Find sources or experimental results for samples of interest e.g. samples with acute myloid leukemia.
- Discover links to public assay data on relevant samples across multiple technologies.
- Link to experimental data on the same cell line/strain/etc from around the world.
- Pre-register your samples before submitting experimental data. You will then receive a stable and unique accession number that is suitable for use in publications, and for use in multiomic data submissions.

BioSamples also supports direct submissions of sample information, typically in advance of submitting assay data to other EBI resources. For more information on the submission process and accepted file formats, see the BioSamples submission page.

**Searching and visualising data in BioSamples**

BioSamples home page
BioSamples - database of sample descriptions

The BioSamples database aggregates sample information for reference samples (e.g. Coriell Cell lines) and samples for which data exist in one of the EBI's assay databases such as ArrayExpress, the European Nucleotide Archive or PRoteomics IdentifIcates DatabaseE. It provides links to assays specific samples, and accepts direct submissions of sample information.

- Help pages about how to search BioSamples, how to submit data, and FAQ.
- Programmatic access to query and download data using web services.
- Latest news about BioSamples database.
- The BioSamples database now offers access to RDF, and a supporting SPARQL endpoint as part of the EBI's RDF platform.
- Contact us by emailing biosamples@ebi.ac.uk

**External links**
- Experimental Factor Ontology
- Human Induced Pluripotent Stem Cells Initiative (HiPSC)
- 1000 Genomes
- Encyclopedia of DNA Elements (ENCODE)
- Catalogue Of Somatic Mutations In Cancer (COSMIC)

**Data Content**
- 2,040,998 Samples
- 92,589 Groups

Figure 1 BioSamples website: home page

All the web pages in the BioSamples database feature a query box in top right-hand corner, with the option to search samples or sample groups.

There are also links to browse by sample or sample group, and links to documentation which includes instructions on how to submit.

**Search results**
Figure 2 Search results list for the term 'leukemia'

- **Query reminder:** The exact search query is reproduced at the top of the results table. It is also kept in the search box, where it can easily be edited to refine the query.
- **Highlighting:** Matches to the search query are highlighted in the results table, including exact matches (green), synonyms (yellow) or more specific terms (red).
- **Links to sample or group pages:** Each result has a link on the left-hand side to a specific BioSamples Database web page.
- **Links to sources:** Where applicable, the right-hand side of each result has link(s) to other web pages.

**Sample group**
Figure 3.1 Search results page (upper section) using 'search by Groups' for the terms "chronic myelogenous leukemia [12]", and selecting the group accession [10] 'SAMEG14728'

On the upper section of the sample group page, notice that:

- Search term highlighting (colour coded) persists from the initial search (including synonyms and more specific terms).
- General information about the group is displayed first, such as publications, release dates and dates of updates.
- Attributes that are common to all the samples in this group are also displayed here.
Figure 3.2 Search results page (lower) using ‘search by Groups’ for the terms "chronic myelogenous leukemia [12]", and selecting the group accession [10] ‘SAMEG14728’

In the lower section of the sample group page, notice that:

- For large sample groups, only some samples are shown at once. The other pages of samples can be accessed by links in the top left.
- If there are a larger number of attributes, a horizontal scroll bar allows you to see all of them.
- There is a search box to find particular samples of interest; if no samples are displayed, check that this has been cleared.

Single sample
Figure 4 Single sample page, resulting from a search for the sample accession "SAMEA249878"

Each sample in the BioSamples Database has a specific web page with a unique and permanent URL, e.g.
http://www.ebi.ac.uk/biosamples/samples/SAMEA249878 [13]. This can be used to refer to a specific sample in publications and other documents.

On this page you can notice the following features:

- Search query highlighting;
- Where attributes are linked to a **controlled vocabulary** [6] or **ontology** [5], those links are displayed here;
- The sample group(s) that this sample is part of are listed.

### Getting data from BioSamples

Public samples are available from BioSamples. Samples can be displayed on the website's user interface, accessed programmatically via the **BioSamples API** [14], or downloaded in our **SampleTab file format** [15] from the [FTP] [16] site.

To help you find relevant samples of interest, the search function in BioSamples includes synonyms and more specific terms - for example, a search for “cancer” will also include "leukemia" [12]”. These are described in more detail in the **BioSamples search documentation** [17].

Sample information from BioSamples is also included in the **EBI's RDF platform** [18].

### Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

### Get help and support on BioSamples

**Support**

- If you have any questions or want help submitting data please contact the biosamples [at] ebi.ac.uk (BioSamples support desk)

  and we will endeavour to reply within 48 hours.

- If you would like to be notified of changes and improvements to BioSamples, you can subscribe to our low-traffic [announcement mailing list] [19].

### References

- Gostev M., Faulconbridge A. et al. (2012). **The BioSample Database (BioSD) at the European Bioinformatics Institute** [21]. Nucleic Acids Res. 40 [Database issue], D64-70

### Funding
BioSamples and the team's related activities are supported by member states of the European Molecular Biology Laboratory; European Commission: BioMedBridges [284209]; Wellcome Trust and Medical Research Council (MRC): Human Induced Pluripotent Stem Cells Initiative (HipSci); the EBiSC Innovative Medicines Initiative Joint Undertaking under grant agreement 115582.

Contributors

Adam Faulconbridge [22]

EMBL-EBI
Bioinformatician - Parkinson team: Samples, Phenotypes and Ontologies

Adam has been a bioinformatician on the BioSamples Database since he joined EMBL-EBI in 2011, and is involved in data management, curation, and software development activities. Prior to EMBL-EBI, Adam completed his PhD and MRes (Bioinformatics) at the University of York, and his BSc in Genetics at the University of Nottingham.

Source URL: https://www.ebi.ac.uk/training/online/course/biosamples-quick-tour

Links
[1] https://www.ebi.ac.uk/training/online/trainers/cherubin_9980
[2] https://www.ebi.ac.uk/training/online/glossary/arrayexpress
[3] https://www.ebi.ac.uk/training/online/glossary/european-nucleotide-archive
[4] https://www.ebi.ac.uk/training/online/glossary/pride
[5] https://www.ebi.ac.uk/training/online/glossary/ontology
[6] https://www.ebi.ac.uk/training/online/glossary/controlled-vocabulary
[7] https://www.ebi.ac.uk/training/online/glossary/assay
[8] https://www.ebi.ac.uk/training/online/glossary/annotation
[9] https://www.ebi.ac.uk/biosamples/samples?text=acute+myeloid+leukemia
[10] https://www.ebi.ac.uk/training/online/glossary/accession
[11] https://www.ebi.ac.uk/biosamples/sampletab/submission
[12] https://www.ebi.ac.uk/training/online/glossary/leukemia
[14] https://www.ebi.ac.uk/biosamples/docs/references/api
[15] https://www.ebi.ac.uk/biosamples/docs/references/sampletab
[16] https://www.ebi.ac.uk/training/online/glossary/ftp
[17] https://www.ebi.ac.uk/biosamples/docs/references/filters
[18] https://www.ebi.ac.uk/rdf/
[22] https://www.ebi.ac.uk/training/online/trainers/faulcon