cBiT manual

1. What is cBiT?

The Compendium for Biomaterial Transcriptomics, or cBiT, is a repository that incorporates data archives containing material science and transcriptomics-based cell biology (i.e. "materiomics"). One of the main goals of cBiT is to describe all meta-information necessary for reproducibility and downstream analysis of experiments with respect to the culturing conditions, transcriptomics procedure, and biomaterial fabrication and properties.

This is a step towards:

- A better comparability between biomaterial-based experiments, allowing for new and more efficient data analysis strategies
- A sustainable infrastructure
- Maximum and effective use of public data

Data archives in cBiT are prepared using ISAcreator (http://isa-tools.org/) which is essentially dedicated spreadsheet software that enables the inclusion of transcriptomics data (or other Big Data types) and allows for an unambiguous characterization of meta-information.

cBiT was designed to be intuitive in use and with this manual users can get all the information they need to start looking for data, downloading interesting data sets, and analyzing them!

2. Main lay-out

cBiT is available at https://cbit.maastrichtuniversity.nl/ and should work fine in recent versions of Chrome, Firefox, and Safari. cBiT is not compatible with Internet Explorer and for those users we suggest switching to Chrome. cBiT consists of three sections divided into tabs, the Welcome tab, the About tab, and the Browse tab. The Browse tab is the most important for cBiT users but we will first briefly go into the Welcome and About tabs.

2.1 Welcome tab

The Welcome tab is cBiT's homepage and contains a short explanatory video of what cBiT is all about. Please have a look and don't hesitate to distribute the video amongst colleagues. The more people know about cBiT, the better!

2.2 About tab

The About tab describes the idea behind cBiT in a bit more detail, offers the cBiT manual for download, provides instructions on how you can contribute your data, and shows our contact details.

2.3 People

Here you can read a short bio on the people behind cBiT.

2.4 FAQ tab

This tab contains a list of frequently asked questions (FAQ) that you may find useful.

2.5 Enter cBiT

The Enter cBiT tab is what cBiT is all about. Here, you can browse through the available data sets, look for specific data using either the search box or the quick selection menu, and download data of interest for further analysis. In the next paragraph the Enter cBiT tab will be explained in more detail and examples will be given on how to use cBiT.

3. Using cBiT

When clicking Enter cBiT you will first be taken to the *Dashboard* which gives you a quick overview of what kind of biomaterial studies cBiT has. You can search and download biomaterial studies in the *Biomaterial Studies* page, underneath Enter cBiT. This page is divided into a results section on the right side of the screen showing the cBiT studies and a search and selection menu section on the left to perform queries (**Figure 1**).

> *Note*: *There is also a second page present underneath Enter cBiT, called Tendon Studies. This page will contain a simple overview of transcriptomics studies focused on tendon research. While this is not the main focus of cBiT, we consider these studies also useful within the biomaterial field of research. However, it does not have the search and download functionalities of the Biomaterial Studies page.*

Before showing how to search the cBiT repository for biomaterial studies, let's first look into the "study cards" in the results section.
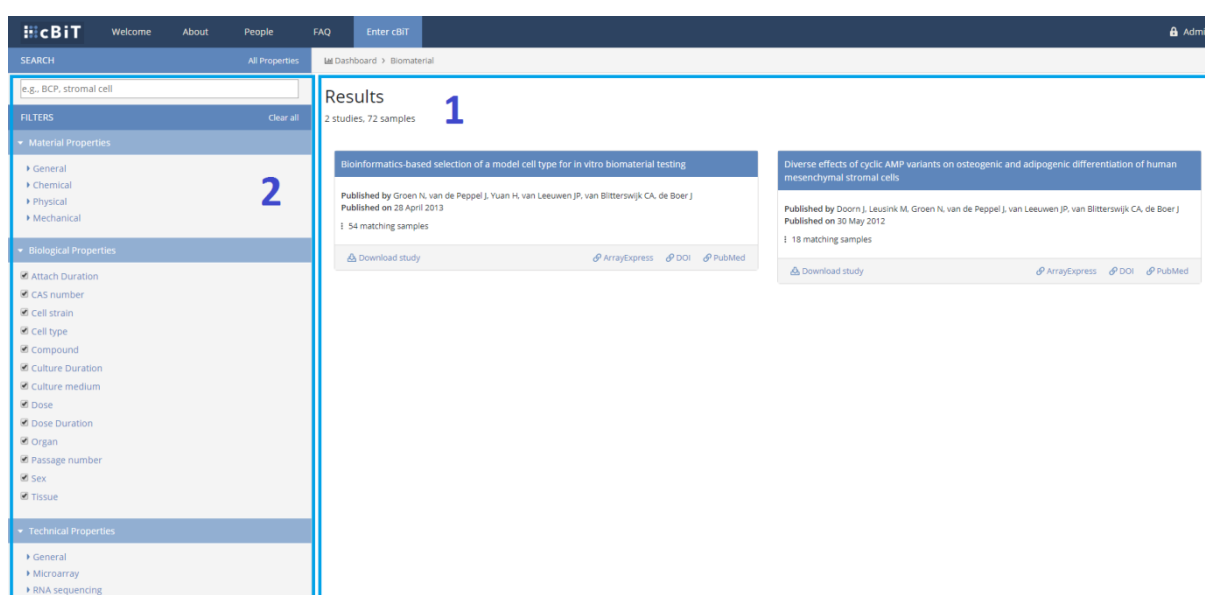

Figure 1: The two sections of the Biomaterial Studies page, underneath the Enter cBiT tab, i.e. the results section (1) and a search and selection menu section (2).

3.1 The study cards

The results section shows all studies as "study cards" with the study title, the authors/researchers involved and the study publishing date (**Figure 2**). In the lower part of the study card you can find several links:

- **Download study**: this will generate a zip file which contains the study details, metadata, protocols and any relevant supplementary files.

- **ArrayExpress**: links directly to the location of the raw and processed data in the ArrayExpress database developed by EBI.

- **DOI**: this DOI links directly to the article on the journal website where the study was published (when applicable).

- **PubMed**: links to the PubMed location of the published study (when applicable).

By default, all studies present in cBiT are shown. Clicking on a study title will take you to that study's page with the abstract and additional details about the study and samples in dropdown menus. It is a good way to get a grasp of what the study was about and whether it is of interest to you. Under the dropdown menus "Common properties" and "Distinguishing properties" you can find a list of properties that the samples all share and properties that they show variations in respectively. In the study page it is also possible to download the study protocols by clicking on the Protocol button (it is automatically included in the Study zip file as well).
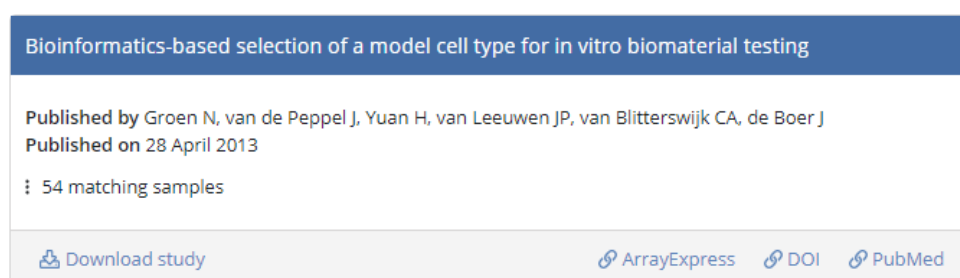

Figure 2: Example of a study card.

3.2 Performing a query

In order to query the cBiT database, the underlying search engine will find matches in any of the recorded study properties of the studies and samples in cBiT. For a list of study properties available in cBiT (Material, Biological, Technical) click on "Full list of properties" just underneath the search box. A specific cBiT query can be performed using either the search box or the selection menu (**Figure 1**, section 2) in the following way:

1. Use the search field in the upper left corner of the screen to look for specific study terms. Most study properties are indexed when new data sets are integrated into cBiT and all these properties are being searched for query matches. For example, searching for "MG-63" will return all studies using the MG-63 cell line (included in the "Cell strain" biological property).

The results section on the right immediately updates in real-time the studies that contain samples matching the query.

2. Use the quick selection menu on the left of the screen to narrow down studies. The menu items correspond with study properties and selecting/de-selecting items will immediately update the studies in the results section that contain samples matching the selection. Properties are divided into Material, Biological, and Technical Properties and further sub-categories per item.

Most properties have values that can be checked or unchecked (e.g. "Material") but there are also properties containing numerical values where a range of values can be selected using a slider (e.g. "Culture duration") (**Figure 3**). For properties with the slider option the unit of measurement can also be selected (e.g. hour, day, week, etc.). The number of samples matching each property value is shown on the right of each property name.

An overview of all available biomaterial properties included in cBiT can be viewed by clicking "Full list of fields" just underneath the search box.

The search box and selection menu can also be used in parallel, e.g. first search for a specific term and then narrow done further by using the selection menu. To start a query from scratch, click on "Clear all".

Figure 3: An example of a check box study property (left) and a slider property (right).

3.3 Downloading studies

Based on a query (and after exploring the study cards), you may decide to download certain studies. As already mentioned above, the data associated with a study can be found in 2 different locations shown in the study card:

- **Download study**: this contains all the study metadata, which includes all of the material, biological and technical properties. Furthermore it contains the protocols and any supplementary data associated with the study, e.g. a separate metadata Excel file containing more complex study property data, such as graphs or images.

- **ArrayExpress**: this contains all the raw and processed transcriptomics data corresponding with the study samples. The link directly takes you the accession page for the data in ArrayExpress where at the bottom you can find all the raw and processed data files and any other info you may need (such as an annotation file to identify genes). The reason we do not offer the transcriptomics data in cBiT directly is a very practical one: it would take up too much server capacity. By using ArrayExpress we can avoid this while still offering the search functionalities and supplementary data within cBiT which is not possible in ArrayExpress, thereby combining the best of both worlds (**Figure 4**).

Figure 4: A cBiT study card and its associated ArrayExpress accession page containing the transcriptomics data.

3.4 The cBiT zip archive and ArrayExpress data

Once you have downloaded the cBiT zip archive and the transcriptomics data from ArrayExpress, you have all the files you need for data analysis purposes, such as differential expression analysis, correlation analysis, pathway analysis, gene network construction, etc. An overview of the types of files you may encounter is shown in **Table 1**.

Table 1: List of file types typically present in an extracted cBiT zip archive and the associated ArrayExpress (AE) data set.

File type	Source	Description
Study detail and property overview.xlsx	cBiT	A list of all study properties contained in cBiT with an explanation of each
s_study_sample.txt	cBiT	Values of all measured or documented study properties for all samples in the study, i.e. the metadata (empty columns can be ignored)
i_Investigation.txt	cBiT	Contains some additional information on the study such as the study title, abstract and authors/researchers
Protocols StudyID_xxx.pdf	cBiT	The protocols associated with the study
Biomaterial graphs file*	cBiT	A separate metadata Excel file, containing more complex study property data, such as graphs or images
Supplementary file(s)*	cBiT	One or more supplementary files containing additional measured or relevant data, not included in the s_study_sample.txt file or Biomaterial graphs file.
Raw data file(s)	AE	The raw data file(s) of the transcriptomics assay corresponding with the study samples
Processed data file	AE	The processed data file(s) of the transcriptomics assay corresponding with the study samples and having undergone processing steps such as normalization and vst or log2 transformation
Array design file (annotations)	AE	Additional annotations for the genes in the raw and processed data files (e.g. EntrezGene ID, Gene Symbol, etc.)

* Not always present in a study

3.6 Invitation to participate

cBiT's usefulness as a dedicated database increases as more studies come in, not only in terms of the number of datasets, but also their level of completeness and variation. We therefore invite research groups who are interested in depositing their own biomaterial-associated data in a sustainable repository and who are eager to contribute to the cBiT project, to contact us. It is possible to either deposit your own data, or provide us with biomaterials that have clinical potential which we will proceed to investigate, analyze, and incorporate into cBiT. This can also include data analysis of data already present in cBiT. In short, the possibilities offered by cBiT are:

1. Browse, download and analyze data yourself

2. Browse, download and find an external party to analyze the data

3. Contact us with an interesting question and ask us to analyze cBiT data

4. Supply us with an interesting biomaterial and ask us to generate data

5. Upload your own dataset

If you would like to participate, please send an email to Dennie Hebels at:
d.hebels@maastrichtuniversity.nl

4. Questions and feedback

If you have any questions on how to use cBiT which are not covered in this tutorial or if you run into any problems during your queries, please contact Dennie Hebels (d.hebels@maastrichtuniversity.nl). Recommendations for improvements to the cBiT infrastructure, data archives or the tutorial are also much appreciated!