

ODM Basic Training

Maria Borodaenko
Senior Solution Consultant



Program

Welcome to ODM training

- Learning goals
- Brief overview of what ODM does and why
- Training
- Q&A Session

Learning Goals



Learning Goals

- Get an impression how ODM can help me and my colleagues
- Get understanding of ODM capabilities
 - Know how to find relevant data
 - Learn how to import a study
 - Get knowledge about curation
 - Know how to share, export and delete a study

ODM Introduction



Making Data FAIR and Action-Ready for Both Consumers & Curators

DATA CURATORS



Metadata Curation

Harmonise thousands of samples
Enforce minimum metadata model

RESEARCHERS



Integrated Data Catalog

Find data across sources
Explore study-sample-data links

DATA SCIENTISTS



Search Services

Slice-and-dice analysis-ready data
Write R Shiny apps rapidly

A Single Point of Truth for Your Data



Training



Training Plan

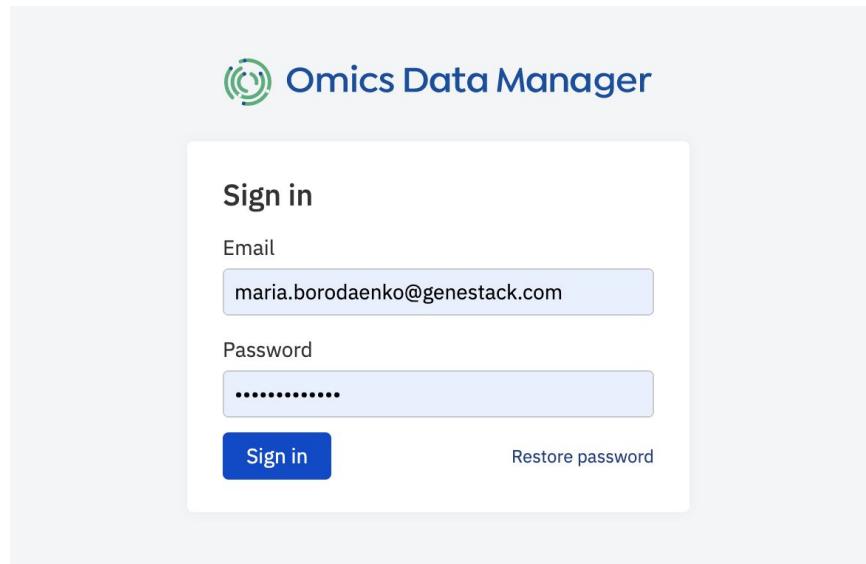
- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- Metadata curation
- Access to studies
- Study export
- Study deletion

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- Metadata curation
- Access to studies
- Study export
- Study deletion

Login

The users are able to login into ODM via SSO



Starting Page

Navigation

-  Dashboard
-  Study Browser
-  Template Editor
-  Users and permissions
-  Groups
-  User Guide

Tasks 2 Maria Borodaenko

 Omics Data Manager

 Invite users to share data with them
Manage groups to regulate data access

 **Browse studies**
82 studies available

 **Set up templates**
Rules for metadata validation

 **Import data using API**
See step-by-step example

[Create new study](#)

Tutorials and documentation

-  Quick start
-  API documentation
-  Generate API token
-  User guide

Omics Data Manager 1.50.0 Release notes Privacy Contact © 2011–2022 Genestack Limited

Starting Page

≡

Tasks 2 Maria Borodaenko

 Omics Data Manager

Invite users to share data with them
Manage groups to regulate data access

 **Browse studies**
82 studies available

 **Set up templates**
Rules for metadata validation

 **Import data using API**
See step-by-step example

Main tasks

Create new study

Tutorials and documentation

- Quick start
- API documentation
- Generate API token
- User guide

Omics Data Manager 1.50.0 Release notes Privacy Contact © 2011–2022 Genestack Limited

Starting Page

≡ Tasks 2 Maria Borodaenko

 Omics Data Manager

[Invite users to share data with them](#)
[Manage groups to regulate data access](#)

[Create new study](#)

Tutorials and documentation

- [Quick start](#)
- [API documentation](#)
- [Generate API token](#)
- [User guide](#)

 **Browse studies**
82 studies available

 **Set up templates**
Rules for metadata validation

 **Import data using API**
See step-by-step example



Omics Data Manager 1.50.0 Release notes Privacy Contact

© 2011–2022 Genestack Limited

Starting Page

Your profile

≡ Tasks 2 Maria Borodaenko

 Omics Data Manager

 Invite users to share data with them
 Manage groups to regulate data access

 **Browse studies**
82 studies available

 **Set up templates**
Rules for metadata validation

 **Import data using API**
See step-by-step example

[Create new study](#)

Tutorials and documentation

-  Quick start
-  API documentation
-  Generate API token
-  User guide

Omics Data Manager 1.50.0 Release notes Privacy Contact © 2011–2022 Genestack Limited

Profile

Profile application shows your permission role, the list of available groups, allows you to edit personal information and to get a API access token.

The screenshot shows the Genestack Profile application interface. At the top, there is a navigation bar with three horizontal lines on the left, the word "Profile" in the center, and "Tasks 1" and "Maria Borodaenko" on the right. Below the navigation bar, the user's profile information is displayed: a circular icon with "MB", the name "Maria Borodaenko", the roles "ADMIN" and "CURATOR", and the email "maria.borodaenko@genestack.com". It also shows the user's groups: "Curator, Maria test group". There are two buttons at the bottom of this section: "Change password" (blue) and "Edit personal information" (light gray). Below this, a section titled "Genestack API tokens" lists a single token: "Summer token odm-demos" (with a copy icon), created on "Jun 16, 2021 14:05". At the bottom of the page, there is a button "Create new token" and a note: "You will be emailed a link to download token as plain text". To the right of the token list, there is a descriptive text about personal access tokens: "Personal access tokens allow you to access your data via the REST API. The tokens are permanent. You can have multiple tokens and revoke them at any time".

Profile

Tasks 1 Maria Borodaenko

Maria Borodaenko ADMIN CURATOR
maria.borodaenko@genestack.com
Groups: Curator, Maria test group

Change password Edit personal information

Genestack API tokens

Summer token odm-demos Jun 16, 2021 14:05

Create new token You will be emailed a link to download token as plain text

Personal access tokens allow you to access your data via the REST API. The tokens are permanent. You can have multiple tokens and revoke them at any time

Training Plan

- Getting started
- **Exploring a Study in Metadata Editor**
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- Metadata curation
- Access to studies
- Study export
- Study deletion

Exploring a Study

Study menu

Multiplexed enrichment and genomic profiling of peripheral blood cell specific immune signature

Study Samples 80 Libraries 80 Preparations 80 EXP Expression

genestack:accession GSF051531

Study Source GEO

Study Source ID GSE120442

Study Description The human immune system consists of many specialized cell subsets that simultaneously carry out a diverse range of functions using overlapping pathways and signals. Subset-specific immune profiling can resolve immune activity in autoimmune disease, cancer immunity, and infectious disease that may not be discoverable or detectable in analyses of crude blood samples. Here, we present a low-input microfluidic system for sorting immune cells into subsets and profiling their cellular states by gene expression analysis using full-length RNA-seq. We validate the device's technical performance by benchmarking its subset enrichment and genomic profiling performance against gold standard benchtop protocols and make the added value of subset-resolved profiling over crude sample analysis clear through in vitro experiments that show subset-specific stimulated responses. Our results show that even the

View only Published by Maria Borodaenko on Nov 21, 22:07

Share Export data Rename Copy accession Save to bookmarks More info Default Template > Template

Maria Borodaenko Export

Exploring a Study

Multiplexed enrichment and genomic profiling of peripheral specific immune signature

Study Samples 80 EXP Expression VAR Variant

genestack:accession GSF316814

Study Source GEO

Study Source ID GSE120442

Study Description The human immune system carries out a diverse range of functions. It identifies and eliminates foreign substances, such as viruses and bacteria, and maintains homeostasis. The immune system also plays a role in the development of certain diseases, such as cancer and autoimmune disorders. In this study, we used multiplexed enrichment and genomic profiling to analyze the immune signature of peripheral blood samples. We found that the immune system responds to various stimuli by activating different subsets of lymphocytes and other immune cells. We also identified specific genetic variants that are associated with the immune response. These findings can help us better understand the complex mechanisms of the immune system and develop new treatments for various diseases.

Study Samples 80 EXP Expression VAR Variant

Filters

Sex	Disease	Age	Age Unit
2 distinct values	2 distinct values	8 distinct values	1 distinct value
<input type="checkbox"/> female 72 <input type="checkbox"/> male 8	systemic lupus erythematosus 40 Healthy 40	59 57	24 8
			year
			80

Edit Published by Maria Borodaenko on Aug 8, 00:40

Multiplexed enrichment and genomic profiling of peripheral specific immune signature

Study Samples 80 EXP Expression VAR Variant

Export dataset

genestack:accession	GSF316814
Experimental Platform	Illumina Genome Analyzer II
Data Processing Method	VarScan
Scale	No value
Genome Version	GRCh38

Exploring Sample Filters

Open filters to analyse data

Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signature Export

Study Samples 80 Libraries 80 Preparations 80 EXP Expression

Filters 41 samples: Disease systemic lupus ery... x

Sex	Disease	Age	Age Unit	Tissue
2 distinct values	2 distinct values	8 distinct values	2 distinct values	2 distinct values
female	37	59	year	Blood
male	4	57	51	No value
	41	51	52	19
	Healthy	52	24	19
	39	24	26	
		26	81	
		81	29	
		29		

Select values of interest

Apply to explore a subset of samples

Apply

Study with Multiple Datasets

List of available datasets

Metadata Editor Multiplexed enrichment and genomic profiling of p... Tasks 5 3 Maria Borodaenko

Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signature

Export

Study	Samples 80	Libraries 80	Preparations 80	EXP Expression
Datasets				Export dataset
training.2.0-tx-expression.gct				genestack:accession GSF051699
training.px-expression.gct				Experimental Platform Illumina NextSeq 550
training.rpkmtx-expression.gct				Data Processing Method Bowtie2, RSEM
				Scale RPKM
				Genome Version GRCh37
				Pipeline ID No value
				Data Files / Raw No value

View only Published by Maria Borodaenko on Nov 21, 22:24

Dataset in focus

Experiment's metadata

Multiple Versions of The Same Dataset

If few versions of dataset are available, they will be displayed. The latest version is active and focused by default, previous version available for exploring and exporting.

Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signature Export

Study	Samples 80	Libraries 80	Preparations 80	EXP Expression
Datasets				v. 2 ACTIVE v. 1 Export dataset
training.2.0-tx-expression.gct				genestack:accession GSF052036
training.px-expression.gct				Experimental Platform Illumina NextSeq 550
training.rpkmtx-expression.gct				Data Processing Method Bowtie2, RSEM
				Scale TPM

Version History

Metadata Editor Multiplexed enrichment and genomic profiling of p... Tasks 7 Maria Borodaenko

Multiple multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signature Export

Study Samples 80 Libraries 80 Preparations 80 EXP Expression

genestack:accession GSF051531

Study Source GEO

Study Source ID GSE120442

Study Description The human immune system consists of many specialized cells that carry out a diverse range of functions using overlapping pathways. Immune profiling can resolve immune activity in autoimmune diseases or infectious disease that may not be discoverable or detectable by other samples. Here, we present a low-input microfluidic system for profiling the cellular states by gene expression.

History of changes

Fixing Sex attribute
Published by Maria Borodaenko on Jun 16, 14:33

uncurated
Published by Fatima Sabitova on Jun 6, 14:33

organism and sex was changed
Published by Fatima Sabitova on May 16, 16:57

Back to latest version Published by Maria Borodaenko on Nov 21, 22:07

View only Published by Maria Borodaenko on Nov 21, 22:07

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- **Searching for available data in Study Browser**
- Data loading
- Metadata templates
- Metadata curation
- Access to studies
- Study export
- Study deletion

Searching for Available Data

Study Browser GSE66728, brain, Parkinson's Disease, GTEx, any metadata ? Tasks 1 Maria Borodaenko

Access
My studies For all users Curator Maria test group

Signal type
Expression 47 Flow Cytometry 2 Variant 8

Organism
 Homo sapiens 16 Human 3 Camellia sinensis 2 Mus musculus 2 koala metagenome 2

more >

Data Type
 RNA-seq 6

Study Source
 Arvados 10 GEO 10 ArrayExpress 2

47 studies found

Filtered by: Signal type Expression ×

Study Title	Sample Count	Type
GTEX version 3	22951	EXP
A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [gene expression levels] - CTA	373	EXP
A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [gene expression levels] - WTA	403	EXP
A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [protein levels]	330	EXP
[UNCURATED] A multi-tissue study of immune gene expression profiling highlights the key role of the nasal epithelium in COVID-19 severity	156	EXP

Created by Fatima Sabitova on Aug 5, 2022

Single-cell, Homo sapiens, Basal pore, Not reported, Alveolar +3 more

Created by Fatima Sabitova on Jul 24, 2022

Single-cell, Homo sapiens, Basal pore, Not reported, Alveolar +3 more

Created by Fatima Sabitova on Jul 24, 2022

Single-cell, Homo sapiens, Basal pore, Seromucinous glands, Alveolar +6 more

Created by Fatima Sabitova on Jul 23, 2022

Homo sapiens, saliva, nasal, Blood, COVID, Control

Created by Fatima Sabitova on Jul 12, 2022

Faceted Search

Access

Study Browser GSE66728, brain, Parkinson's Disease, GTEx, any metadata

47 studies found

Filtered by: Signal type Expression

Access

- My studies
- For all users
- Curator
- Maria test group

Signal type

- Expression 47
- Flow Cytometry 2
- Variant 8

Organism

- Homo sapiens 16
- Human 3
- Camellia sinensis 2
- Mus musculus 2
- koala metagenome 2

Data Type

- RNA-seq 6

Study Source

- Arvados 10
- GEO 10
- ArrayExpress 2

Linked data

Disease

- Search
- Not applicable 3
- COVID-19 2
- Healthy 2
- lung adenocarcinoma 2
- lung squamous cell carcinoma 2
- acinar cell carcinoma 1
- adenocarcinoma 1
- Adenocarcinoma with mixed subtypes 1
- adenosarcoma 1
- Anesthesia 1

Metadata

Therapeutic Area

- Infectious Disease 3
- Immunology 2
- Oncology 2
- Crohn's Disease 1
- Immuno-oncology 1

Disease

- Sex
 - female 8

156 EXP

Studies

- GTEX version 3
- A single-cell and spatial targets of SARS-CoV-2
- A single-cell and spatial targets of SARS-CoV-2
- [UNCURATED] A multi-tissue study of immune gene expression profiling highlights the key role of the nasal epithelium in COVID-19 severity

Full-Text Search

The screenshot shows the Genestack Study Browser interface. At the top, there is a search bar containing the query "GSE66728, brain, Parkinson's Disease, GTEx, any metadata". Below the search bar, it says "47 studies found" and "Filtered by: Signal type Expression X". A black arrow points from the search bar down to the first study entry in the list.

The study list includes:

- GTEX version 3**: GTEX, Created by Fatima Sabitova on Jul 24, 2022, 22951 EXP
- A single-cell targets of SARS-CoV-2 [protein levels]**: Single-cell, Homo sapiens, Basal pore, Not reported, Alveolar +3 more, Created by Fatima Sabitova on Jul 24, 2022
- [UNCURATED] A multi-tissue study of immune gene expression profiling highlights the key role of the nasal epithelium in COVID-19 severity**: Homo sapiens, saliva, nasal, Blood, COVID, Control, Created by Fatima Sabitova on Jul 12, 2022

On the left sidebar, there are filters for Access (My studies, For all users, Curator, Maria test group), Signal type (Expression 47, Flow Cytometry 2, Variant 8), Organism (Homo sapiens 16, Human 3, Camellia sinensis 2, Mus musculus 2, koala metagenome 2), Data Type (RNA-seq 6), and Study Source (Arvados 10, GEO 10, ArrayExpress 2).

A modal window titled "Study Browser" is open over the third study entry. It contains a search bar with "human" and a list of terms from ontologies:

- Homo sapiens (preferred label of humans)
- Human herpesvirus 5
- human granulocytic anaplasmosis
- human intestinal bacterium SNU-Julong732

A red curly arrow points from the text "Terms from ontologies" to the list of terms in the modal window.

Search Results

Study Browser GSE66728, brain, Parkinson's Disease, GTEx, any metadata ? Tasks 1 Maria Borodaenko

Access
My studies
For all users
Curator
Maria test group

Signal type
Expression 47
Flow Cytometry 2 Variant 8

Organism
Homo sapiens 16
Human 3
Camellia sinensis 2
Mus musculus 2
koala metagenome 2

more >

Data Type
RNA-seq 6

Study Source
Arvados 10
GEO 10
ArrayExpress 2

47 studies found

Filtered by: Signal type Expression X

⋮ **GTEX version 3** 22951 EXP
GTEX
Created by Fatima Sabitova on Aug 5, 2022

⋮ **A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [gene expression levels] - CTA** 373 EXP
Single-cell, Homo sapiens, Basal pore, Not reported, Alveolar +3 more
Created by Fatima Sabitova on Jul 24, 2022

⋮ **A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [gene expression levels] - WTA** 403 EXP
Single-cell, Homo sapiens, Basal pore, Not reported, Alveolar +3 more
Created by Fatima Sabitova on Jul 24, 2022

⋮ **A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2 [protein levels]** 330 EXP
Single-cell, Homo sapiens, Basal pore, Seromucinous glands, Alveolar +6 more
Created by Fatima Sabitova on Jul 23, 2022

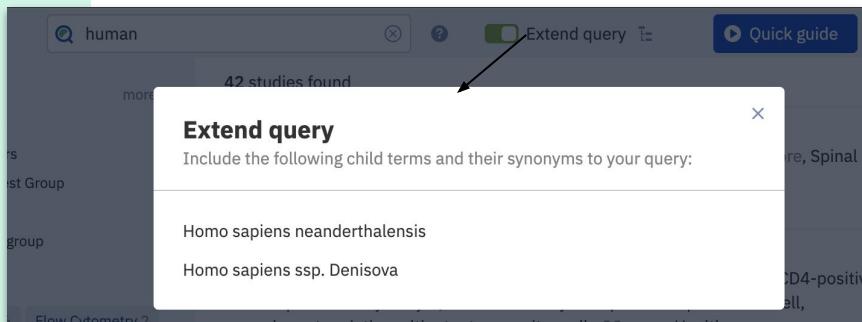
⋮ **[UNCURATED] A multi-tissue study of immune gene expression profiling highlights the key role of the nasal epithelium in COVID-19 severity** 156 EXP
Homo sapiens, saliva, nasal, Blood, COVID, Control
Created by Fatima Sabitova on Jul 12, 2022



Search results

Advanced Search Capabilities

The query can be extended based to include child terms from the ontology



Help button for using logical operators, wildcards and more

A screenshot of the Genestack search interface showing the "Help" section. At the top, there is a "Quick guide" button. Below it, a search bar shows "o sapiens" with a "Clear" button. A large arrow points from the "Help" button in the main interface to this "Help" section. The "Help" section is titled "Help" and contains three main sections: "Searching for phrases", "Using wildcards", and "Using logical operators".

- Searching for phrases**: Use quotes ("") to search for an exact phrase. For example, "*single cell*" will find results that contain the whole phrase only.
- Using wildcards**: Question marks (?) match any single character. Asterisks (*) allow any number of (including zero) wildcard characters. For example, *ale will find "male" and "female".
- Using logical operators**: The AND operator between terms requires both terms to be matched. For example, *melanoma AND braf*

Bookmarks

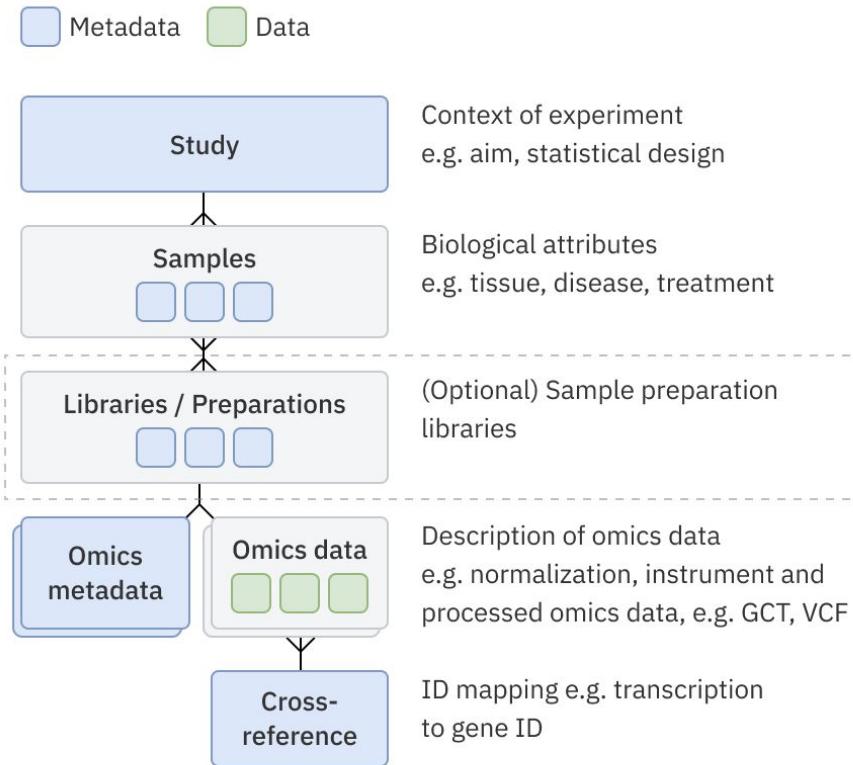
In case you have some studies which you regularly work with, it could be easier and faster to find them among bookmarks.

The screenshot shows the Genestack Study Browser interface. On the left, a sidebar titled "Bookmarks" lists two items: "Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signature" (Created by Maria Borodaenko on Nov 21, 2022) and "GTEx" (Created by Genestack Superuser on Nov 10, 2022). A green callout box with the text "Find it quickly next time" has an arrow pointing to the "Multiplexed enrichment" bookmark. On the right, a study detail view for "GSE66728, brain, Parkinson's Disease, GTEx, any m..." is shown. The study has been "Saved". The "Access" section includes "My studies", "For all users", and "Maria Test Group". The "Signal type" section includes "Expression 7", "Flow Cytometry 1", and "Variant 1". The "Therapeutic Area" section is empty. A green callout box with the text "Save an interesting study" has an arrow pointing to the "Save to bookmarks" button in the study details panel. The "Save to bookmarks" button is highlighted in blue, indicating it is selected. Other options in the panel include "Share", "Export data", "Copy accession", and "More info".

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- **Data loading**
- Metadata templates
- Metadata curation
- Access to studies
- Study export
- Study deletion

Data Model and Supported formats



Import	GUI	Python Script
Study	Manual	TSV
Samples	Manual or TSV import	TSV
Libraries / Preparations (optional)	X	TSV
Omics metadata	X	TSV
Omics data	X	GCT, VCF, FACS

Linking Attributes

Headers in the expression file should correspond to sample metadata so the system recognise what column relates to what sample.

During import user can specify which attribute from the template should be used for linking during import.

If nothing is specified the “Sample Source ID” attribute is used.

Can be any template attribute

Sample file		
Name	Sample Source	Sample Source ID
Sampe_1	Lab_1	Sample_1_ID
Sample_2	Lab_1	Sample_2_ID
Sample_3	Lab_1	Sample_3_ID

Expression file			
#1.2			
24571	60	Sample_1_ID	Sample_2_ID
Name	Description	Sample_1_ID	Sample_2_ID
Gene_1		2.23451	0

Data Loading via Script

During study import a study itself and associated data objects are created sequentially and linked to each other based on the data model.

This is done by the import data script which calls REST APIs. The user should be a curator and should provide an access token, links to the files and the server address. More details about the parameters can be found in the script's help.

```
study GSF052044 was added successfully
samples were added successfully (sample group accession is GSF052044)
Successfully linked: [samples_to_study]
libraries were added successfully (libraries group accession is GSF052210)
Successfully linked: [libraries_to_samples]
Successfully linked: [libraries_to_samples]
expression data were added successfully (group accession is GSF052210)
Successfully linked: [expression_to_libraries]
Execution is finished!
```

Example command

```
python3 import_ODM_data.py \
--token YOUR_TOKEN \
--server YOUR_SERVER \
--study s3://data_source/study.tsv \
--samples s3://data_source/samples.tsv \
--libraries s3://data_source/libraries.tsv \
--expression s3://data_source/expression.gct \
-em s3://data_source/expression.gct.tsv
```

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- **Metadata templates**
- Metadata curation
- Access to studies
- Study export
- Study deletion

Metadata Templates

For effective metadata curation the Templates are used. Template are set by users with “Manage templates” permission in Template Editor.

They describe the attributes and select ontologies for metadata validation. There could be more than 1 template on the instance to have set of attributes specific to a study type. The default template can be set.

The screenshot shows the Genestack Template Editor interface. On the left, there's a sidebar with categories: Study, Sample, Library, Preparation, Expression, and Variant. The 'Sample' category is currently selected, indicated by a blue border. In the main area, there's a table titled 'Sample' with columns for Name, Required, and Meta. The rows contain fields like 'Accession', 'SAMPLE_ID', 'Sample Name', 'Organism', and 'Tissue'. To the right of the table, a modal window titled 'Template Editor' is open. It shows a section for 'Default template' with a button labeled 'Default Template'. Below this, a note says: 'The default template is applied to a study when no other template is selected.' A 'Templates' section lists existing templates: 'Template for BZ' and 'Agri'. At the bottom of the modal, there are dropdown menus for 'NCBI Taxonomy / GSF047496' and 'Uberon Anatomical Entities / GSF047497', both currently set to 'No'. The overall interface is clean with a light gray background and blue highlights for active sections.

Metadata Validation

Curators can choose a template for their studies.
Default template is applied by default.

The study will be validated against the template.
Any inconsistencies will be shown.

Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subclone specific immune signature

Study	Samples 80	Libraries 80	EXP Expression					
	A	D	E	F	G	H	I	J
	genestack:....	Organism	Sex	Disease	Age	Age Unit	Tissue	Cell Type
1	GSF052047		F		24			
2	GSF052048		F		24			

Validation summary
Select value to replace

Organism
No value 80
Not filled

Sex
F 72
Should be a preferred label from dictionary "Sex"
M 8
Should be a preferred label from dictionary "Sex"

Cell Type
No value 80
Not filled

Required Templates Attributes

There are a few attributes required by the system. These attributes should be presented in any template to allow ODM work properly. The following attributes should not be removed from the attribute or altered in spelling.

ATTRIBUTE	ENTITY	COMMENT
“Accession”	All entities	Mandatory for the template.
“Study Title”	Study	Without these attributes the users will get errors upon opening a study in the GUI
“Sample Source ID” (or alternative linking attribute)	Samples	Linking attribute should present in the template for proper linkage

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- **Metadata curation**
- Access to studies
- Study export
- Study deletion

Metadata Curation

For curating metadata one should be a member of the Curator group.

The studies is opened in a view mode, to check validation and to apply changes the curator should switch to the Edit mode.

Changes made in the Edit mode are available for other curators but they are not seen by the researcher.

Start curation

The screenshot shows the Genestack Metadata Editor interface. At the top, there are navigation tabs: 'Study' (selected), 'Samples 80' (highlighted with a blue border), and 'EXP Expression'. Below these are 'Filters' and a search bar with the text '35 samples: Disease systemic lupus erythematosus'. The main area is a table with columns 'genestack:acc...' and 'Sample Source ID'. The table contains several rows of sample data:

	genestack:acc...	Sample Source ID
1	GSF316728	P0772_CD14_1
2	GSF316729	P0772_CD14_2
3	GSF316730	P0772_CD19_1
4	GSF316731	P0772_CD19_2

At the bottom right, there is a note: 'Published by Maria Borodaenko on [date]'. A blue button labeled 'Edit' is located at the bottom center.

Changing a Template

Prior to curating a study it is better to verify that the proper template is used.

The applied template is available in the study menu. It can be explored and changed if needed.

The screenshot shows the Genestack interface with a floating 'Choose template' dialog and a study menu overlay.

Choose template Dialog:

- Template for BZ
- Agri
- Agriculture template
- Antman
- Astex template
- Copy of Copy of Default Template
- Copy of Default Template
- Copy of Default Template
- Copy of Default Template 2
- Copy of Default Template 3
- Copy of Maria Test

Explore button is visible at the bottom right of the dialog.

Apply and **Cancel** buttons are at the bottom of the dialog.

Study Menu Overlay (Maria Test Template selected):

- Share
- Export data
- Rename
- Copy accession
- Save to bookmarks
- More info

Maria Test Template is selected in the list.

Description: Maria Test is used for metadata validation.

Actions: Explore, Apply another

Metadata Curation

Curators can bulk replace values and reassign attributes to make metadata valid.

Species ID	Type
Homo sapiens	T Species ID
Homo sapiens	Not from template
Homo sapiens	Copy values to...
Homo sapiens	Bulk replace
Homo sapiens	Reassign attribute
Homo sapiens	Remove attribute
Homo sapiens	

Reassign attribute

Species ID to

Reassign **Cancel**

Metadata Versioning

- Once the changes are made study should be published.
- The history of published versions is available for each study.
- Any of previous version can be viewed and restored.

Version history

- [Restored] Initial version
Published by Maria Borodaenko on Aug 8, 04:03
- Edited using curation script
Published by Maria Borodaenko on Jun 17, 16:46
- Initial version
Published by Maria Borodaenko on Jun 17, 16:40

Study Design No value

Therapeutic Area Immuno-oncology

Edit Published by Maria Borodaenko on Jun 17, 16:40

17	GSF333659	P1353_CD14_1	10031353	No value
18	GSF333660	P1353_CD14_2	10031353	No value

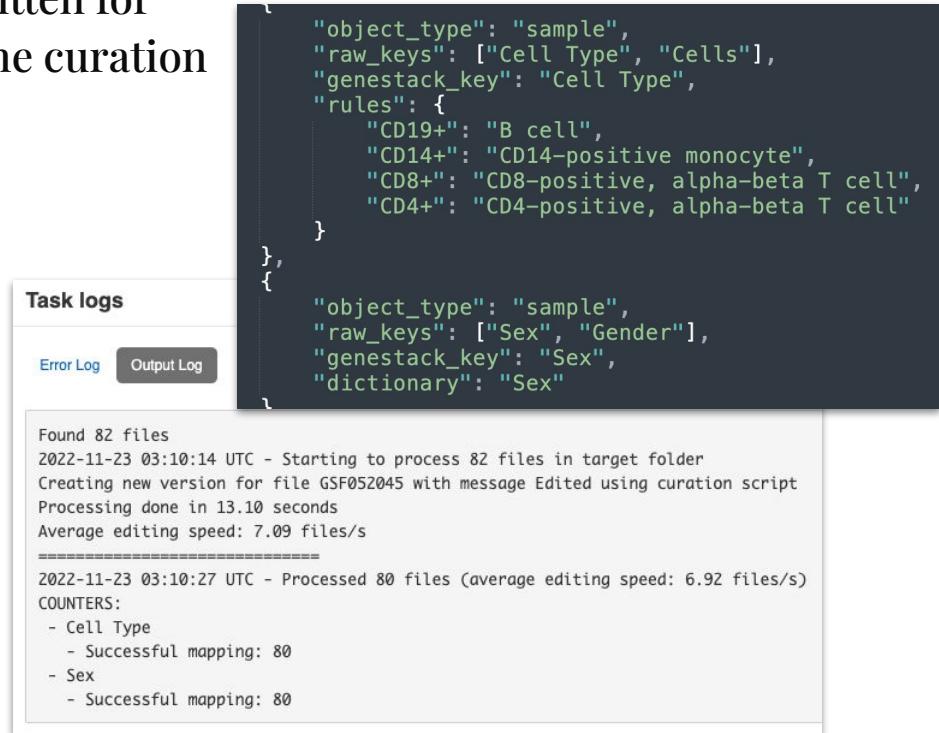
[Back to latest version](#) [Restore](#) Published by Maria Borodaenko on Jun 17, 16:40

Curation Script

Curation script is a python script written for automating frequent operations in the curation process.

The main purpose of the script is to facilitate the curation process, reduce manual operations and reduce errors in metadata curation.

The script could be run on a particular study(ies). Changes done by the script are regulated by JSON file with predefined rules.



The screenshot shows a split-screen interface. On the left, a 'Task logs' panel displays two tabs: 'Error Log' (disabled) and 'Output Log'. The 'Output Log' tab is active, showing the following log entries:

```
Found 82 files
2022-11-23 03:10:14 UTC - Starting to process 82 files in target folder
Creating new version for file GSF052045 with message Edited using curation script
Processing done in 13.10 seconds
Average editing speed: 7.09 files/s
=====
2022-11-23 03:10:27 UTC - Processed 80 files (average editing speed: 6.92 files/s)
COUNTERS:
- Cell Type
  - Successful mapping: 80
- Sex
  - Successful mapping: 80
```

On the right, a code editor displays a JSON configuration file with the following content:

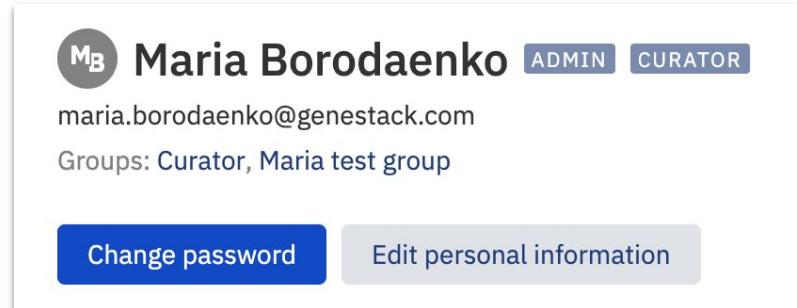
```
{"object_type": "sample",
"raw_keys": ["Cell Type", "Cells"],
"genestack_key": "Cell Type",
"rules": {
    "CD19+": "B cell",
    "CD14+": "CD14-positive monocyte",
    "CD8+": "CD8-positive, alpha-beta T cell",
    "CD4+": "CD4-positive, alpha-beta T cell"
},
{
"object_type": "sample",
"raw_keys": ["Sex", "Gender"],
"genestack_key": "Sex",
"dictionary": "Sex"
```

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- Metadata curation
- **Access to studies**
- Study export
- Study deletion

Access to Studies

- The list of groups available in your profile.
- Curator group is a special group granting edit permissions.
- You can share your study with any group you are a member of.
- You can view with which groups a study is shared.



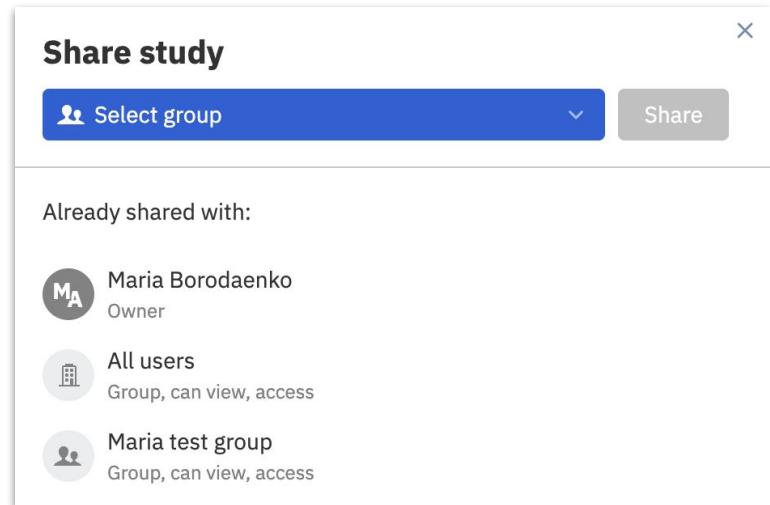
A user profile card for Maria Borodaenko. It includes her name, email (maria.borodaenko@genestack.com), and two groups she belongs to: Curator and Maria test group. There are buttons for changing her password or editing personal information.

Maria Borodaenko ADMIN CURATOR

maria.borodaenko@genestack.com

Groups: Curator, Maria test group

Change password Edit personal information



A dialog box titled "Share study" with a "Select group" dropdown and a "Share" button. Below it, it shows the groups already shared with:

- Maria Borodaenko (Owner)
- All users (Group, can view, access)
- Maria test group (Group, can view, access)

Sharing a Study

To share a study user should be the owner of a study - the one who created it in ODM.

Share option is available in the Study menu.

The screenshot shows the Genestack interface with a study titled "Multiplexed enrichment and genomic profiling of peripheral blood ce specific immune signature". A green box labeled "Study menu" points to the "Share" option in the context menu, which also includes "Export data", "Rename", "Copy accession", "Save to bookmarks", "More info", and "Default Template". Below this, a "Share study" dialog is open, showing a dropdown for "Select group" and a list of users already shared with the study. The list includes "Maria Borodaenko" (Owner), "All users" (Group, can view, access), and "Maria test group" (Group, can view, access). A second "Share study" dialog is shown in a darker overlay, with a search bar containing "Search by name" and a list of users: "All users", "Curator" (highlighted with a blue background), and "Maria test group".

Study menu

Share

Export data

Rename

Copy accession

Save to bookmarks

More info

Default Template

Metadata Editor Multiplexed enrichment and genomic profiling of p...

Multiplexed enrichment and genomic profiling of peripheral blood ce specific immune signature

Share study

Select group

Share

Already shared with:

- Maria Borodaenko Owner
- All users Group, can view, access
- Maria test group Group, can view, access

Share study

Select group

Search by name

All users

Curator

Maria test group

Training Plan

- Getting started
- Exploring a Study in Metadata Editor
- Searching for available data in Study Browser
- Data loading
- Metadata templates
- Metadata curation
- Access to studies
- **Study export**
- Study deletion

Exporting a Study

The screenshot shows the Genestack interface with a study titled "Multiplexed enrichment and genomic profiling of peripheral blood ce... specific immune signature". A "Study menu" button is highlighted in green, and a "Study export" button is also highlighted in green. A tooltip for "Study export" points to the "Export" button in the study menu. Below the study title, there's a "Metadata Editor" tab and a "Share" button. On the left, a sidebar titled "Export Data" lists previous exports, including one for this study with a download link and timestamp. At the bottom, there's a section for "Other exports" with a note about outdated data.

Metadata Editor Share

Multiplexed enrichment and genomic profiling of peripheral blood ce... specific immune signature

Export Data

Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific imm...
Study, size before compression is unknown

Exported on 2022-10-26 11:39:06, 2.06 MB ...

Y2F0aW9uL3ppcC9hL3QvZy9jLzIxNjMwMjE=/Exported_on_2022-10-26_11_39_06.zip Download

Authorization is not required to download with this link

Other exports

These archives are already prepared for export but might contain outdated data.

Exported on 2021-11-18 11:14... Download ...

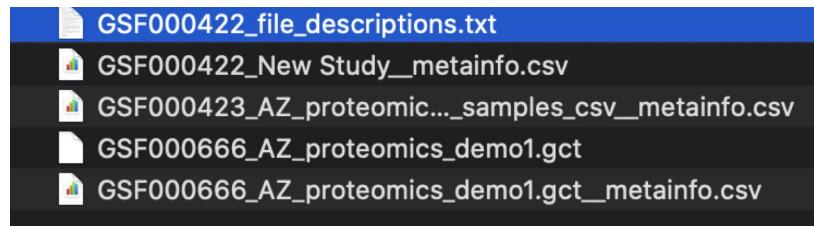
To export a study with all related data click on the Export data either in the study menu or in Metadata Editor.

This will navigate you to the Export Data page where the link for loading an archive with available files will be created. Also, the previous exports are mentioned there.

Exporting a Study

The archive contains all available files and a ReadMe file with the description.

[accession]_file_descriptions.txt contains the accessions of the exported files, their names and the descriptions of the files.



```
--  
Study title: New Study  
Study accession: GSF000422  
Exported on: 2021-08-31 13:26:12 UTC  
  
This archive contains the following:  
  
Omics data files:  
GSF000666_AZ_proteomics_demo1.gct - Expression data file (as imported to ODM) in GCT format.  
  
Metadata files:  
GSF000422_New Study_metainfo.csv - Study metadata (context of study etc.) table in CSV format. The first row contains metadata field names and the second row contains values.  
GSF000423_AZ_proteomics_demo1_samples_csv_metainfo.csv - Samples metadata (sample biological attributes etc.) table in CSV format. The first row contains metadata field names and subsequent rows contain values for each sample.  
GSF000666_AZ_proteomics_demo1.gct_metainfo.csv - Expression metadata table (that accompanies the expression data file of the same name) in CSV format. The first row contains metadata field names and the second row contains values.
```

Exporting a Dataset

It is possible to export just one dataset of interest. For doing it navigate to a tab with analysis of interest and click on “Export dataset” button.

The data file with the corresponding metadata file and the description will be compress into archive.

Multiplexed enrichment and genomic profiling of specific immune signature

Study Samples 80 EXP Expression VAR Variant

Export dataset

genestack:accession GSF316814

Experimental Platform Illumina Genome Analyzer II

Preparations 80 EXP Expression

v. 2 ACTIVE v. 1

Export dataset

genestack:accession GSF051697

Experimental Platform Illumina NextSeq 550

Single file export

Previous version export

Study Deletions

To delete a study “wipeStudy” method can be called either from the Browser console or via the python script. The method finds all entities linked to the study and delete them one by one.

A screenshot of a web browser window. At the top, there is a search bar with the text "brain, Parkinson's Disease, GTEx, any met". Below it, a message says "1 study found". Underneath, a filter section shows "Filtered by: Access Maria test group". A context menu is open over a study entry. The menu items are: "Share", "Export data", "Copy accession" (which is highlighted in blue), "Save to bookmarks", and "More info".

A screenshot of a browser's developer tools console tab. It shows the command `window.wipeStudy('GSF316726')` being typed in.

A screenshot of the Genestack "Users and permissions" page. The title is "All users". A search bar contains the text "mar". There are four users listed: Maria Borodaenko (ADMIN, You, maria.borodaenko@genestack.com) and Maria Lazarenko (DEACTIVATED, maria.lazarenko@genestack.com). To the right, there is a "Permissions" section with checkboxes for "Manage organization" (checked), "Manage groups" (checked), "Create and deactivate users" (unchecked), "Change user passwords" (unchecked), and "Manage user permissions" (unchecked). A red arrow points to the "Manage organization" checkbox.

Prerequisites: The user should have “Manage organisation” permission.

Input: Study accession.

Study Design within ODM

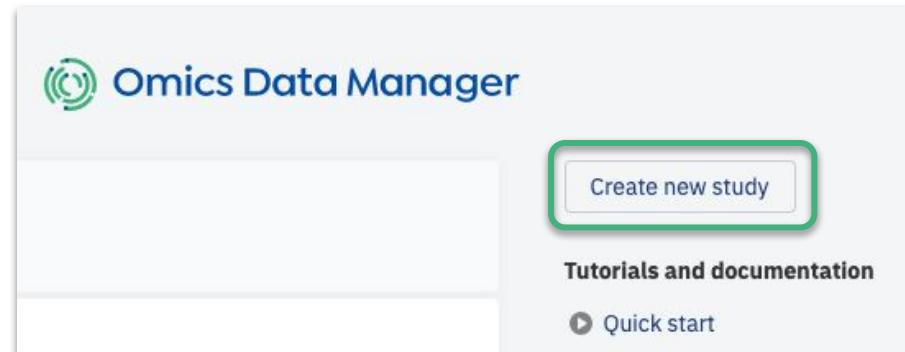
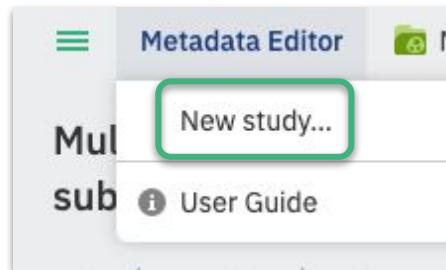


Study Design

A new study can be created directly in ODM without the need of filling in Excel spreadsheets or tsv files. The benefits are:

- Always synced and up-to-date attributes
- Suggestions from the ontologies while filling metadata
- Immediate validation based on the template

New study can be created from Dashboard or from Metadata Editor menu.



Study Design

After clicking on New study, the user will be prompted to enter a name and choose a template.

A new study with the chosen name and blank attributes from the template will be created.

The screenshot shows the Genestack Metadata Editor interface. At the top, there is a navigation bar with 'Metadata Editor', a 'New Study' button, 'Tasks 6', and 'Maria Borodaenko'. Below the navigation bar, the title 'New Study' is displayed. There are two tabs: 'Study' (which is selected) and 'Samples 4 •'. To the right of the tabs is an 'Export' button. In the main area, there are two rows of metadata fields. The first row contains 'genestack:accession' with the value 'GSF052213'. The second row contains 'Study Source' with the value 'No value'.

genestack:accession	GSF052213
Study Source	No value

New study

Name
Untitled Study

Choose template:

- Default Template Default
- Maria Test
- Maria's Template

Create

Adding Samples

A few blank samples will be created with the new study.

New samples can be created by “+” button or imported from .tsv file.

New samples

	F	G
	Disease	Age
1		
2		
3		
4		
5		
6		
7		
8		
...		
N		

Invalid metadata

+ 

25

Maximum 100 samples at a time

Add Cancel

New Study

Study Samples 4

Filters Invalid metadata +

Import samples metadata from TSV file

A	B	C	D	E	F	G
Sample Source ID	Sample Name	Organism	Sex	Diagnosis	Sample Type	Notes
1						
2						
3						
4						
5						
6						
7						
8						
...						
N						

Select tsv file...

Import Cancel

Congratulations



You are now
a trained ODM user!

Q&A



Thank You!

Maria Borodaenko
Senior Solution Consultant
maria.borodaenko@genestack.com
www.genestack.com



Unlocking the Power of Life-Sciences Data