



The Arborist and *tmtk*: A researcher friendly data loading toolkit for transSMART

Jochem Bijlard¹, Ward Weistra¹, Stefan Payralbe¹, Natalia Boukharov², Mariska Bierkens³, Annemieke Hiemstra³, Remond Fijneman³, Jan-Willem Boiten⁴, Gerrit Meijer³

[1] The Hyve, Utrecht, The Netherlands, [2] The Hyve, Cambridge, USA, [3] Netherlands Cancer Institute, Amsterdam, The Netherlands, [4] Lygature, Utrecht, The Netherlands

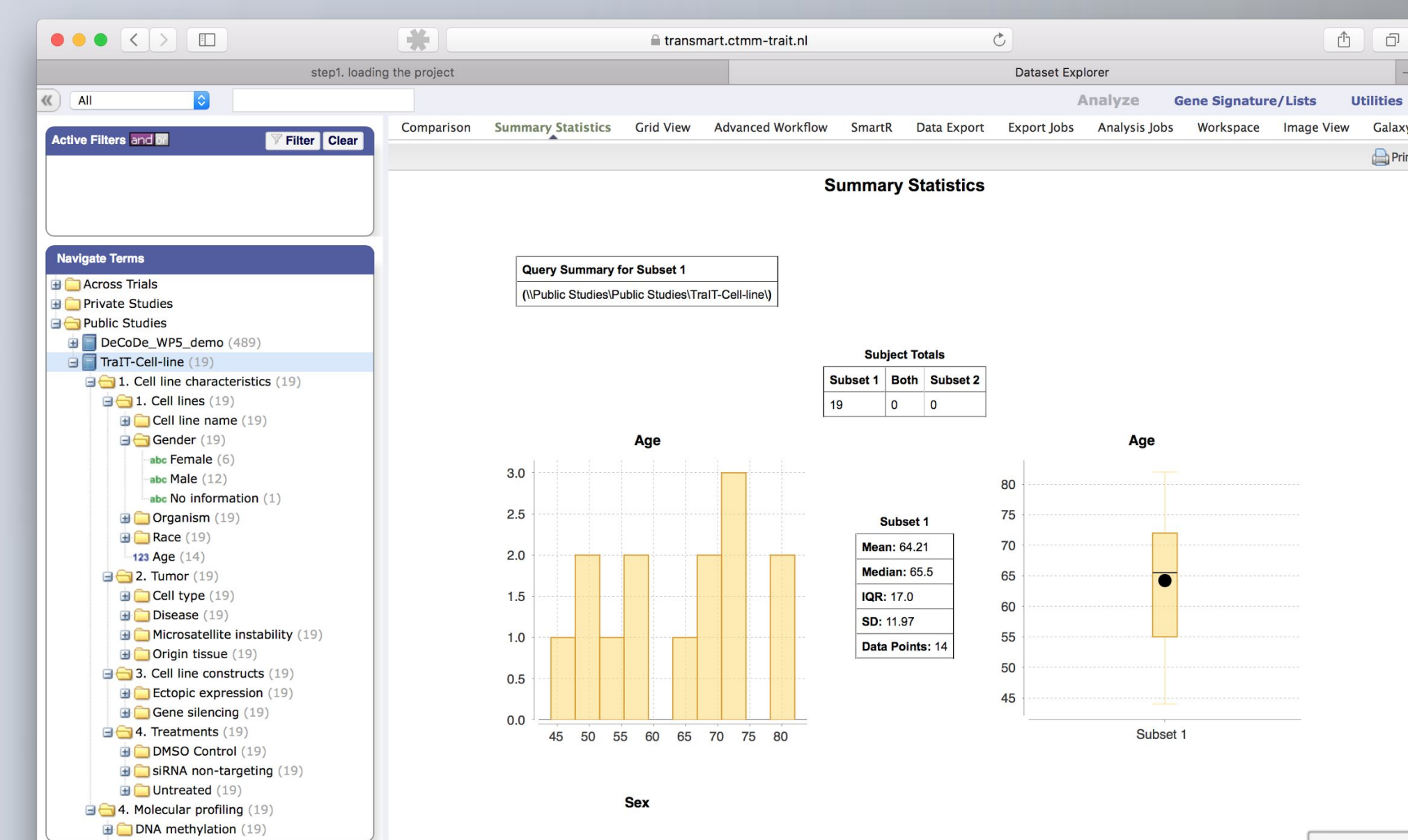
Correspondence: jochem@thehyve.nl, ward@thehyve.nl

Your study data in Excel

Subject ID	Cell line name	Organism	Origin tissue	Cell type	Disease	Ethnicity	Age	Gender	Ectopic MU
1	CACO2	Homo sapiens	Colon	Epithelial adherent cells	Colorectal adenocarcinoma	Caucasian	72	Male	No
2	CACO2_AURKA	Homo sapiens	Colon	Epithelial adherent cells	Colorectal adenocarcinoma	Caucasian	72	Male	No
3	COLO205	Homo sapiens	Colon	Round & refractile	Colorectal adenocarcinoma	Caucasian	70	Male	No
4	COLO320	Homo sapiens	Colon	Round & refractile	Colorectal adenocarcinoma	Caucasian	50	Female	No
5	DLD1	Homo sapiens	Colon	Epithelial cells	Colorectal adenocarcinoma	Caucasian	Male	No	No
6	HCT116	Homo sapiens	Colon	Epithelial-like cells	Colorectal carcinoma	Caucasian	Male	No	No
7	HCT116_MLH1	Homo sapiens	Colon	Epithelial-like cells	Colorectal carcinoma	Caucasian	Male	Yes	No
8	HCT15	Homo sapiens	Colon	Epithelial-like cells	Colorectal adenocarcinoma	Caucasian	Male	No	No
9	HT29	Homo sapiens	Colon	Epithelial-like cells	Colorectal adenocarcinoma	Caucasian	44	Female	No
10	HT29	Homo sapiens	Colon	Epithelial-like cells	Colorectal adenocarcinoma	Caucasian	44	Female	No
11	UM1863	Homo sapiens	Colon	Organoids, columnar cells	Colorectal carcinoma	Caucasian	74	Female	No
12	LS1747	Homo sapiens	Colon	Epithelial adherent cells	Colorectal carcinoma	Caucasian	38	Female	No
13	LS1747	Homo sapiens	Colon	Epithelial adherent cells	Colorectal carcinoma	Caucasian	63	Male	No
14	PC346c	Homo sapiens	Prostate	Epithelial cells	Prostatic carcinoma: derived f	Caucasian	68	Male	No
15	RND	Homo sapiens	Colon	Epithelial cells	Colorectal carcinoma	Caucasian	82	Female	No
16	SW639	Homo sapiens	Colon	Epithelial cells	Colorectal carcinoma	Caucasian	82	Female	No
17	SW48	Homo sapiens	Colon	Epithelial cells	Colorectal adenocarcinoma	Caucasian	82	Female	No
18	SW480	Homo sapiens	Colon	Epithelial cells	Colorectal adenocarcinoma	Caucasian	50	Male	No
19	SW480_AURKA	Homo sapiens	Colon	Epithelial cells	Colorectal adenocarcinoma	Caucasian	50	Male	No
20	VCAp	Homo sapiens	Prostate	Epithelial cells	Prostatic carcinoma: derived f	Caucasian	59	Male	No

From Excel to transSMART in five simple steps

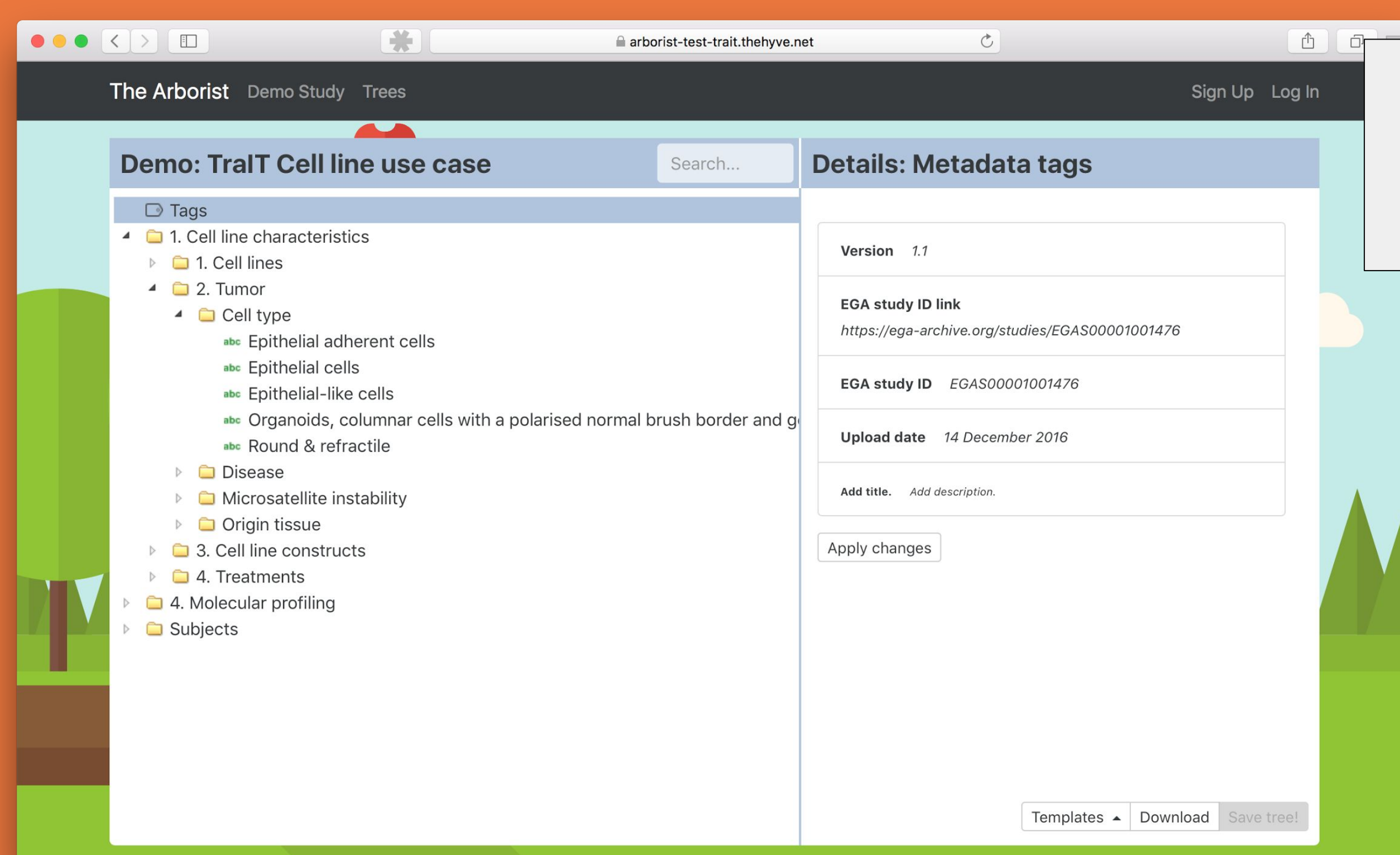
Your study loaded in transSMART



The Arborist Visual editor

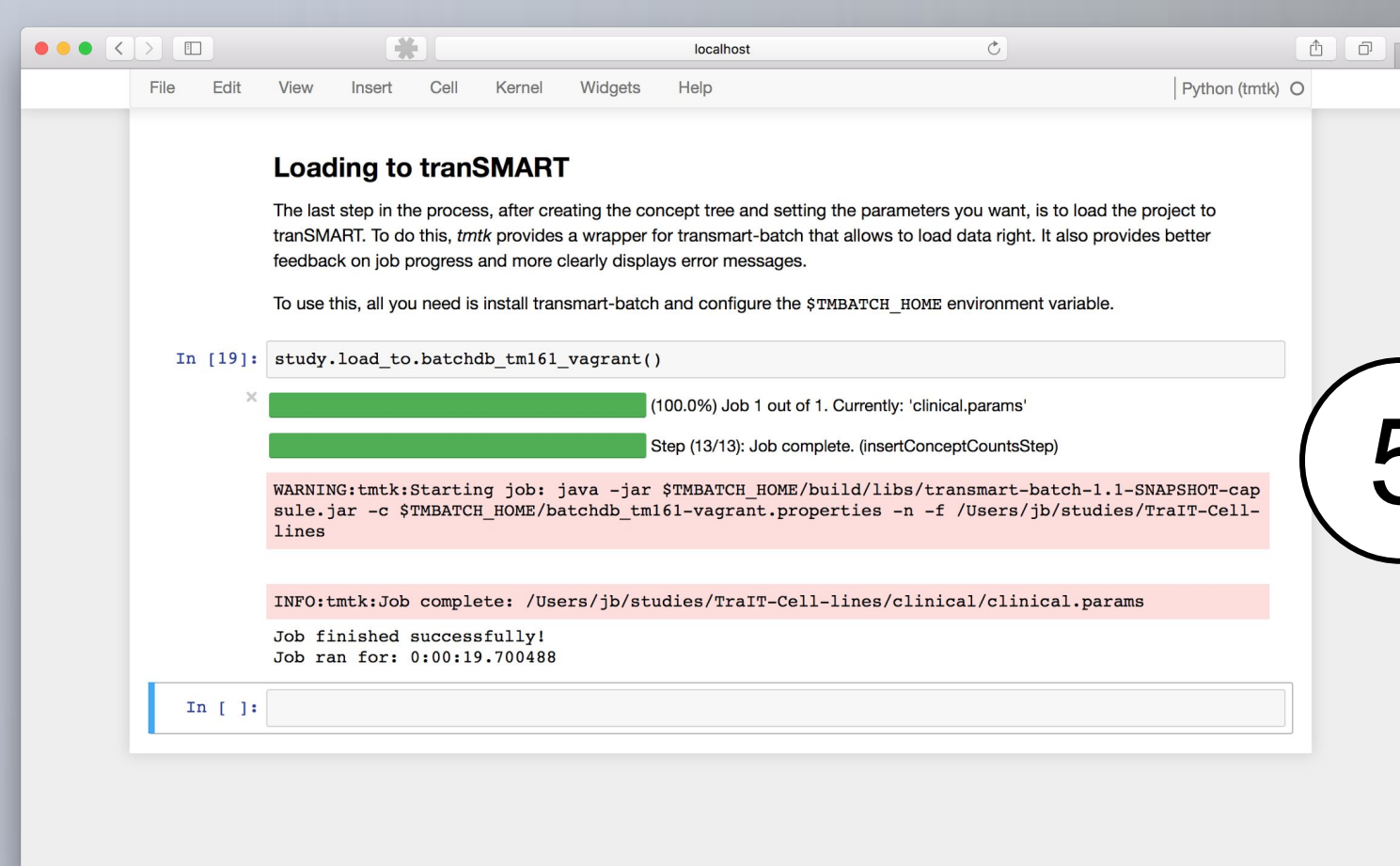
Collaborate on data modelling with non-technical data experts in the secure Arborist web application.

- Restructure the transSMART tree with **drag and drop**
- **Rename** variables and values
- Add and edit **metadata** for any tree node
- Work with both **low and high dimensional data**

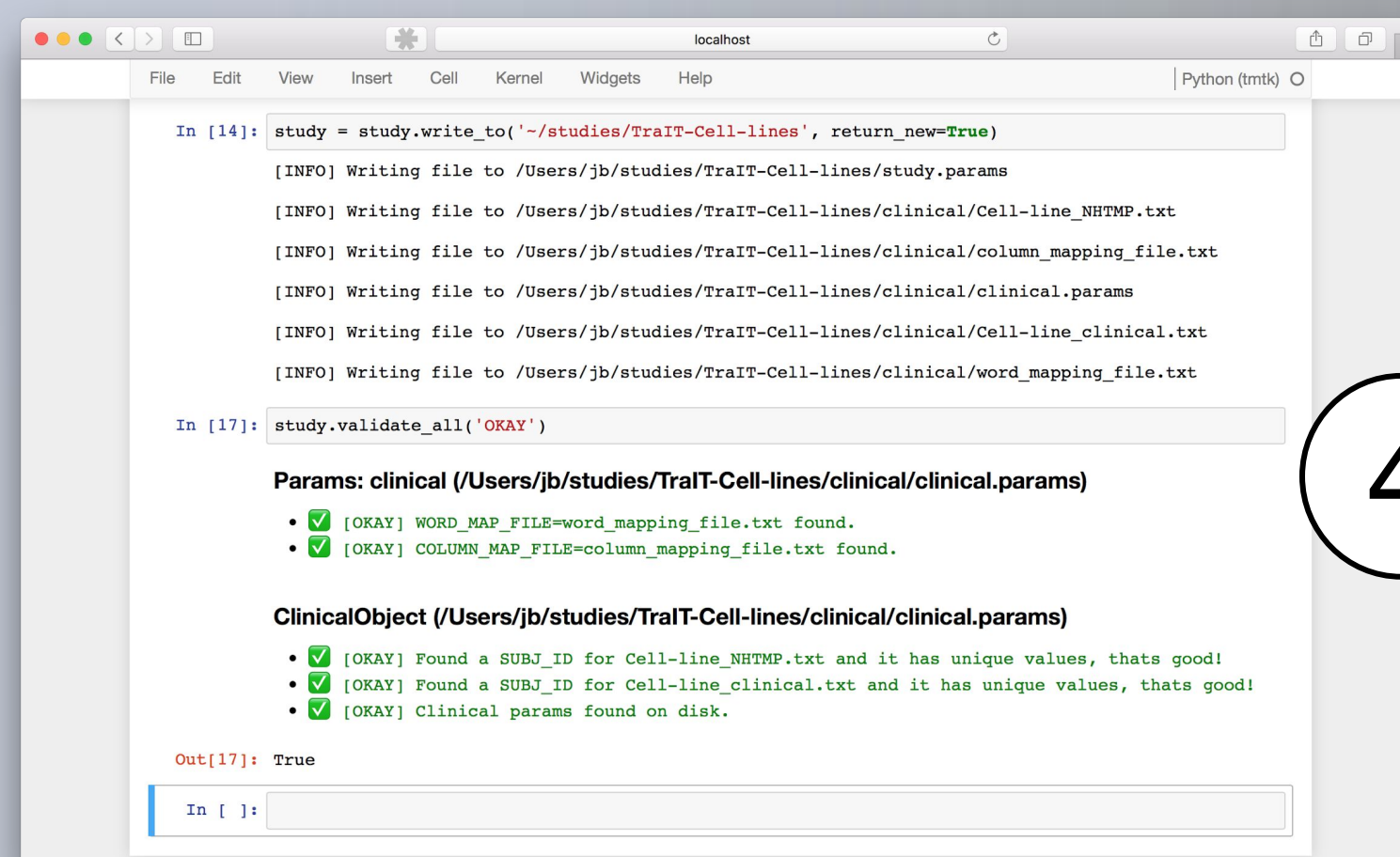


Send to the Arborist web application for easy collaboration!

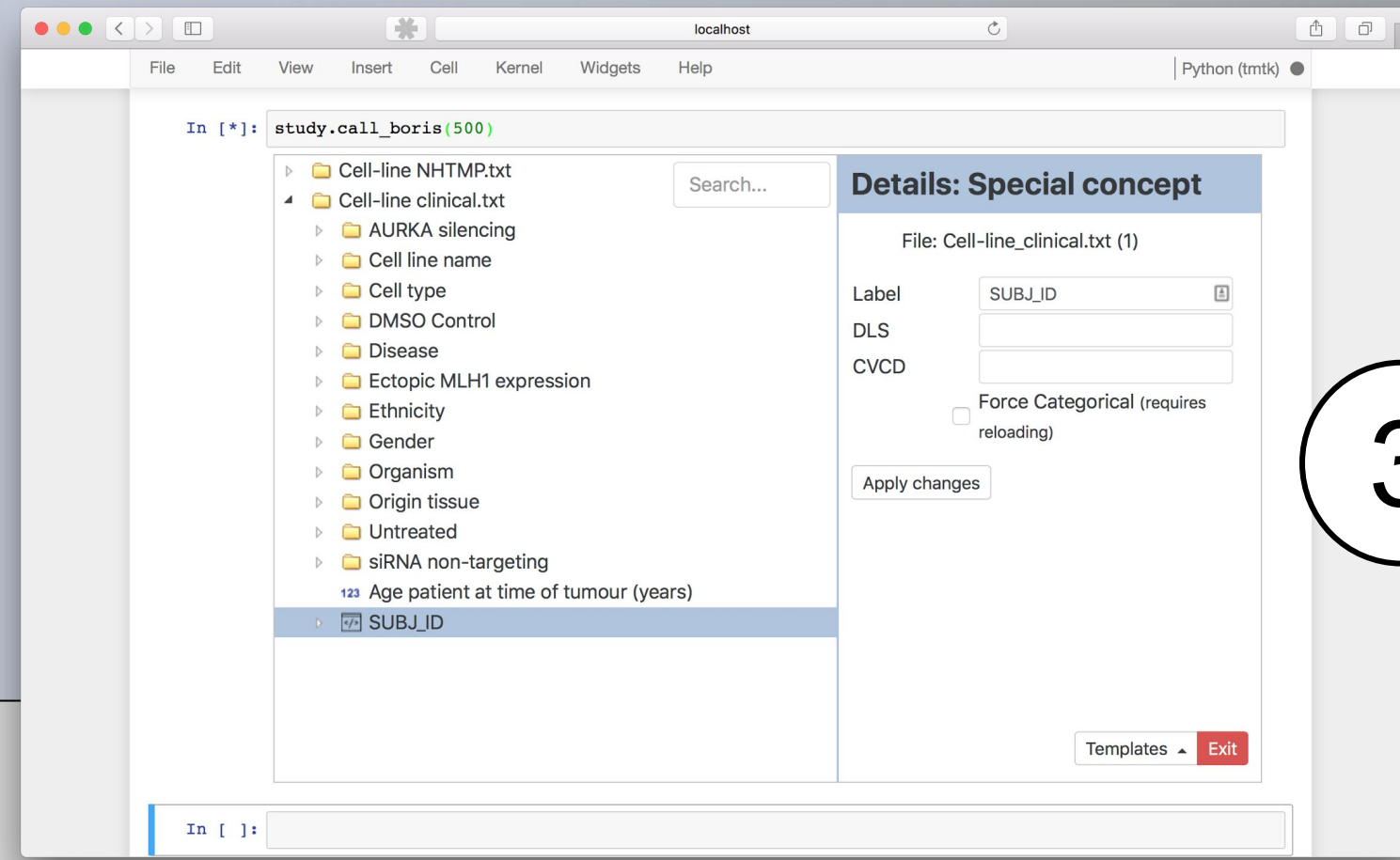
Load: use transmart-batch to load your data to transSMART.



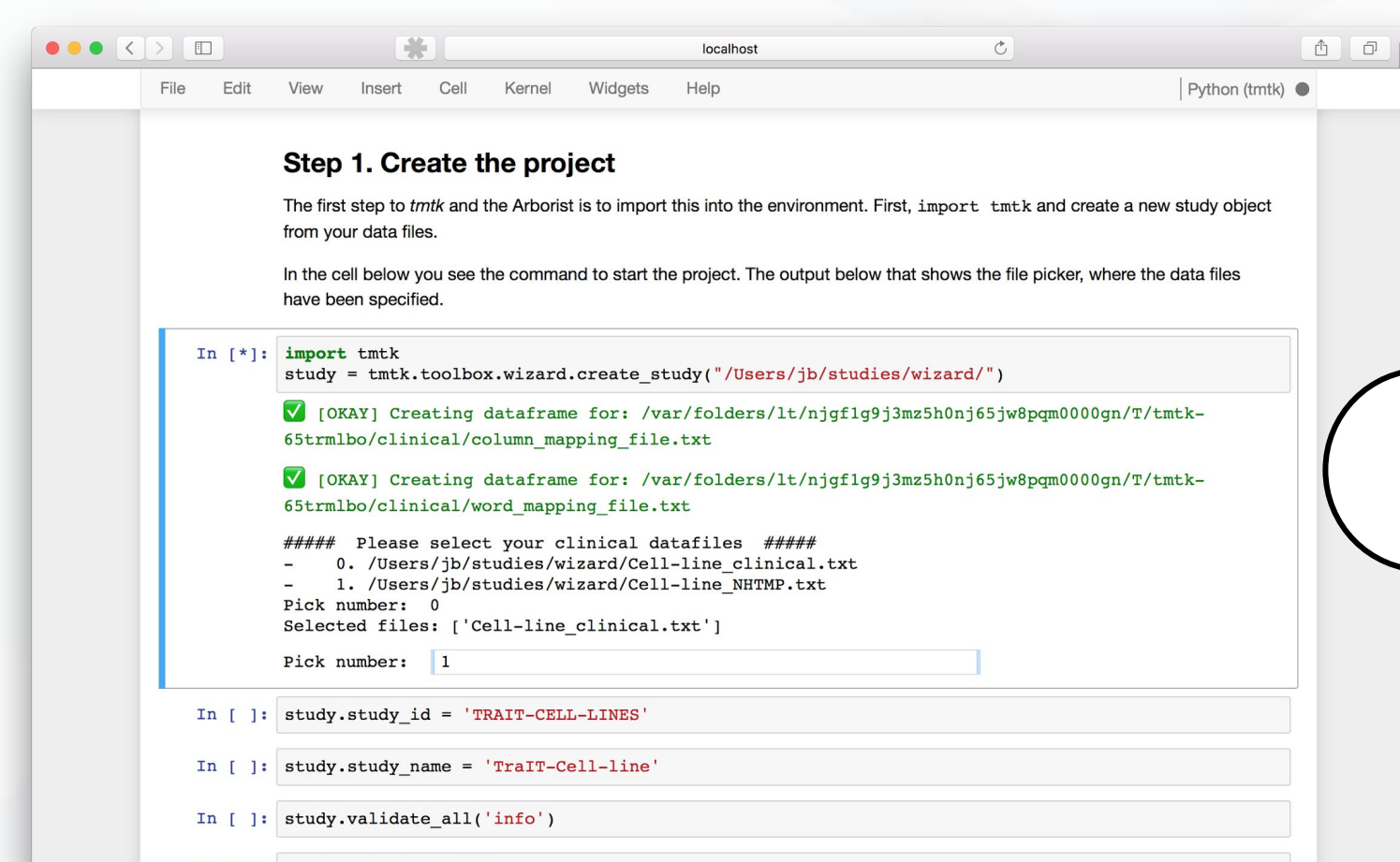
Save: store the study on disk as transSMART-ready staging files.



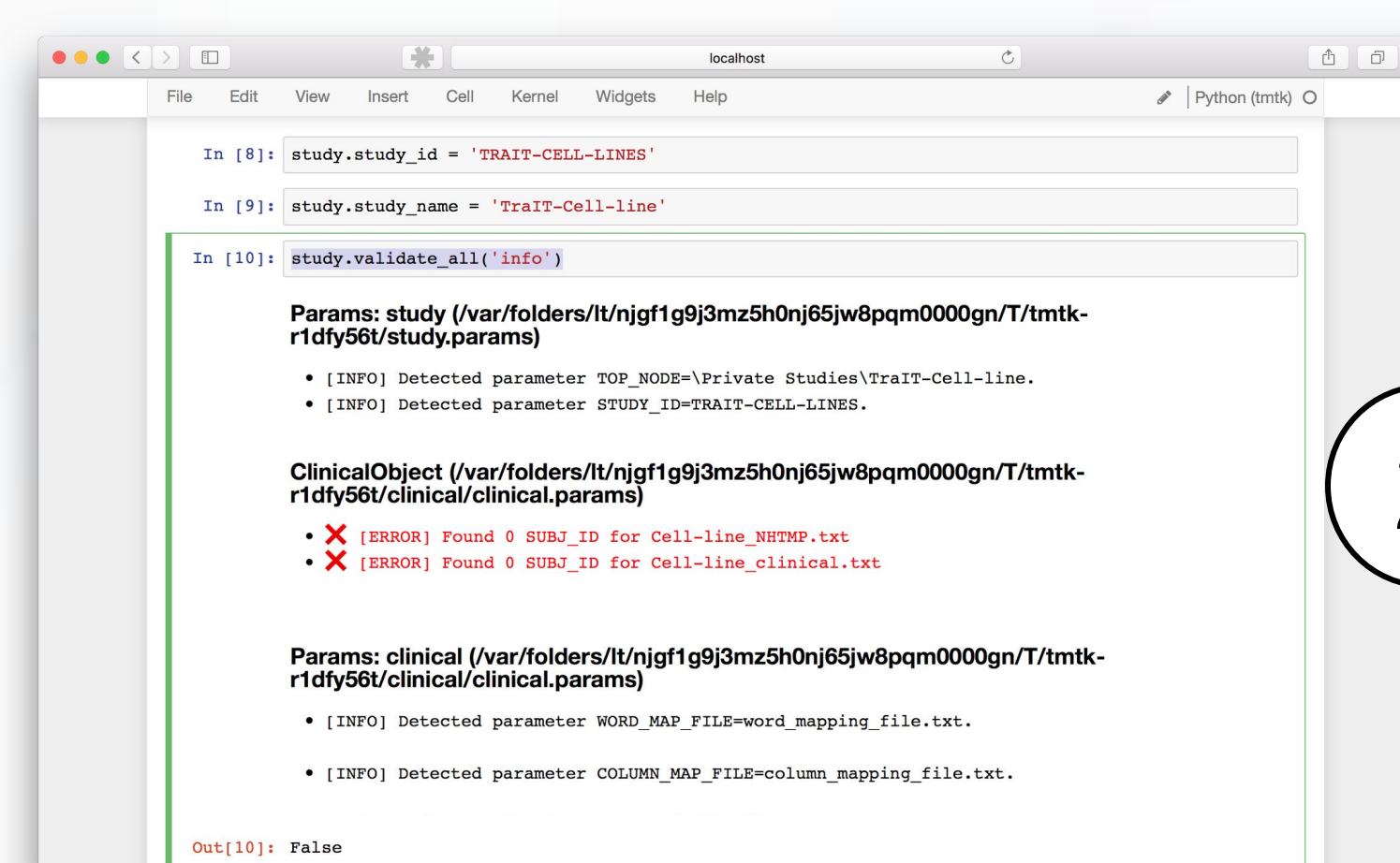
Edit: make changes to your tree with the visual Arborist editor.



Import: start the import wizard to create a study based on your study data.



Validate: let the toolkit check the transSMART-specific requirements.



tmtk Python library

Library that allows users to create and load studies without the need for transSMART specific knowledge:

- **Quickstart** studies from tabular files (e.g. XLS, TSV, CSV)
- Extensive dataset **validation**
- Use **The Arborist** directly embedded into Jupyter Notebook
- Load studies to **The Arborist** web application for collaboration
- Many functions to work with **low and high dimensional data**
- **Minimal** technical and transSMART specific knowledge required

Install for Python3: \$ pip install tmtk

Documentation: <https://tmtk.readthedocs.io>

Code at <https://github.com/thehyve/tmtk> under GPL v3 license.

tmtk notable python commands

The main object in the *tmtk* workflow is the *Study*. It provides an API for modifying and validating your data. Below are the key methods and features provided by *tmtk*.

Starting a study

create_study_from_templates()
Create Study from **TraIT** templates. A way to create an entire Study from filled in templates.

wizard.create_study()
Create Study from tabular files. Quickstart your transmart study.

RandomStudy()
Generate fully randomized Study object. Great for testing stuff!

Validation of the data

.validate_all()
Many of the objects in *tmtk* have validating methods. These methods can easily be extended by adding more.

Loading the data
tmtk provides a wrapper for transmart-batch for easy use and better progress bars!

.load_to()
Load your study to transmart from Jupyter or the console.

Transmart arborist

Visual drag and drop editor for the transSMART concept tree. Use it to shape the concept tree, change word mappings, add metadata, and map concepts to ontologies.

.call_boris()
Launch the Arborist embedded into Jupyter.

.publish_to_baas()
Send data tree to Arborist web application for easy collaboration.

