



# Tour of the Image Data Resource

**Eleanor Williams, Josh Moore, Gabriella  
Rustici, Aleksandra (Ola) Tarkowska**



University  
of Dundee



# IDR Tour Workshop

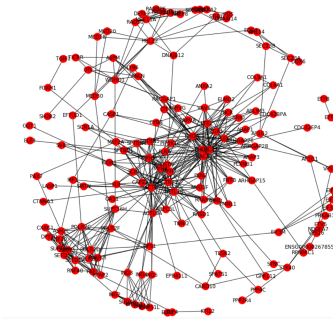
- Querying IDR annotations using the web user interface



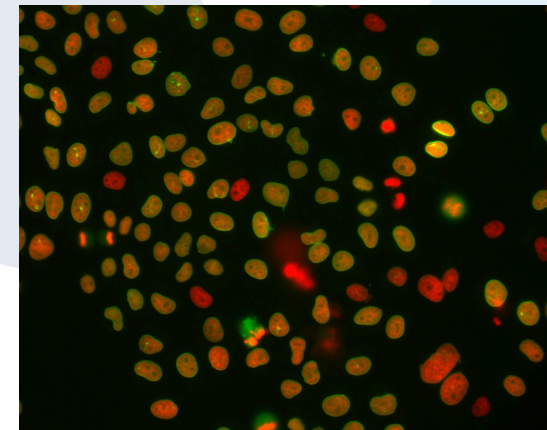
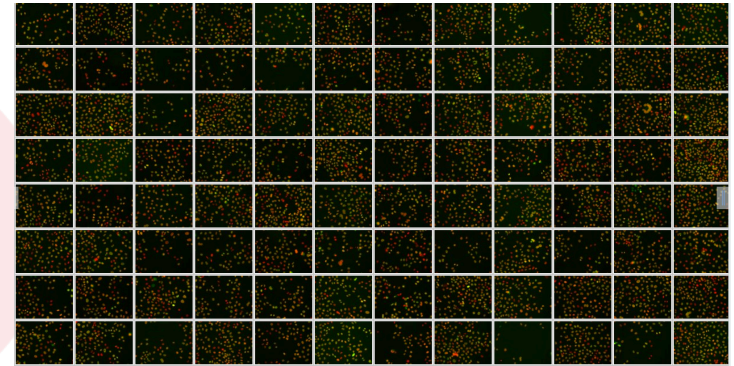
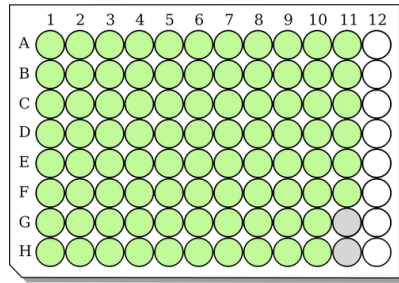
- Using the API to query annotations



- Performing more complex queries and analyses



# High Content Screens



Cells in wells are treated in parallel:

- **genes knocked down or out**, or
- **proteins tagged** for visualization, or
- treated with chemical **compounds**

Using the web user interface



[idr.openmicroscopy.org](https://idr.openmicroscopy.org)





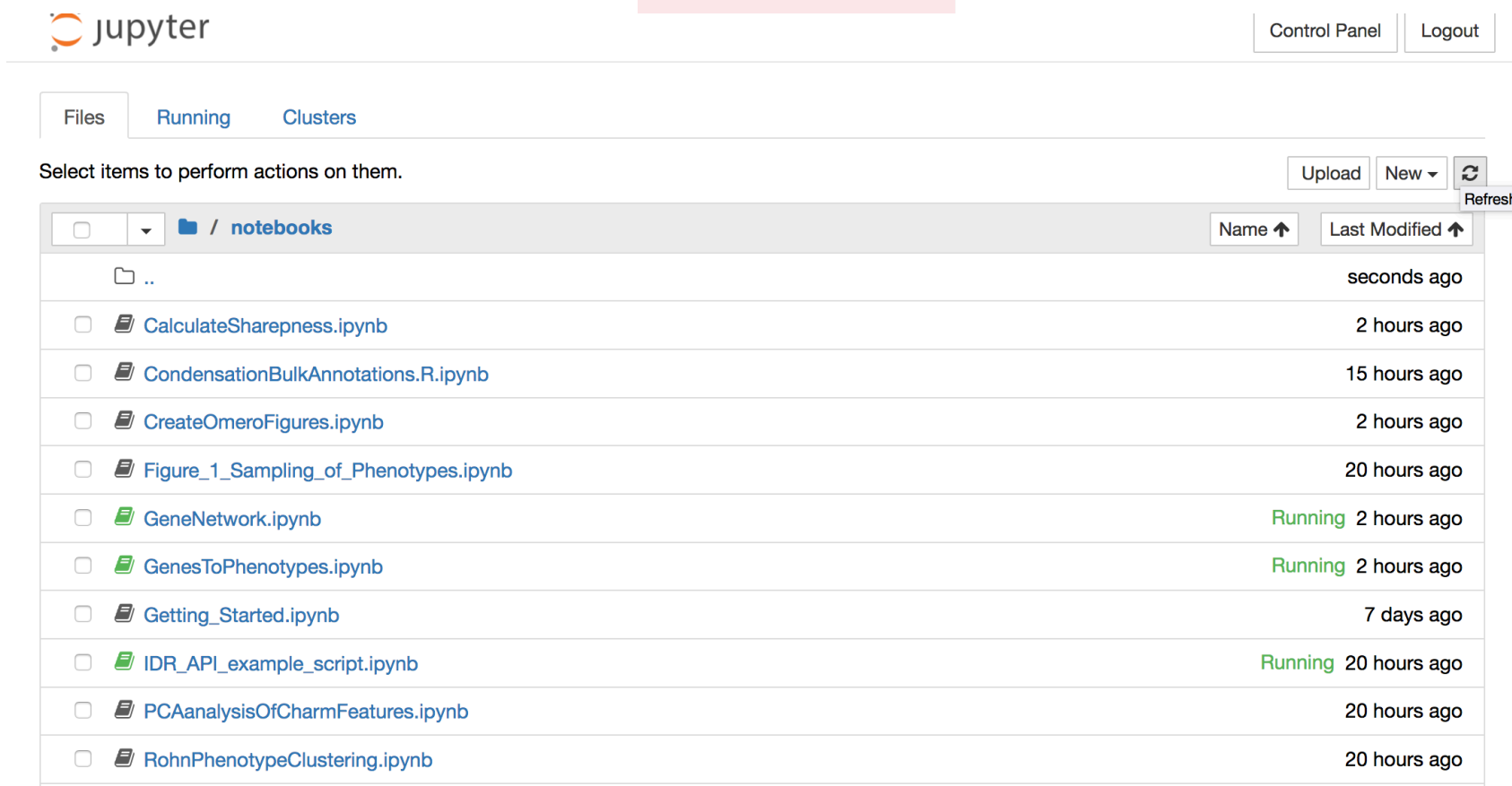
# Using the API to query data annotations

- IDR datasets, thumbnails and metadata are accessible through the web-based API in JSON format
- The full API specification for MAPR is available from <https://github.com/ome/omero-mapr/tree/master/docs>
- Further API documentation for the IDR including an example Jupyter notebook to load data from IDR is available from <http://idr.openmicroscopy.org/about/api.html>.



# Jupyter – follow along with workshop

<https://jupyter.openmicroscopy.org/jupyter/>



The screenshot shows the Jupyter web interface. At the top, there is a header with the Jupyter logo, a "Control Panel" button, and a "Logout" button. Below the header, there are tabs for "Files", "Running", and "Clusters". The "Files" tab is active, and it shows a message "Select items to perform actions on them." with buttons for "Upload", "New", and "Refresh". The main content area displays a list of notebooks in the "notebooks" directory. The list has columns for "Name" and "Last Modified".

Name	Last Modified
..	seconds ago
CalculateSharepness.ipynb	2 hours ago
CondensationBulkAnnotations.R.ipynb	15 hours ago
CreateOmeroFigures.ipynb	2 hours ago
Figure_1_Sampling_of_Phenotypes.ipynb	20 hours ago
GeneNetwork.ipynb	Running 2 hours ago
GenesToPhenotypes.ipynb	Running 2 hours ago
Getting_Started.ipynb	7 days ago
IDR_API_example_script.ipynb	Running 20 hours ago
PCAanalysisOfCharmFeatures.ipynb	20 hours ago
RohnPhenotypeClustering.ipynb	20 hours ago



## More complex queries and analyses

- Have a list of genes you are interested in, what phenotypes are they associated with in IDR?
- Interested in the elongated cell phenotype. What genes are associated with this phenotype in IDR, across species?

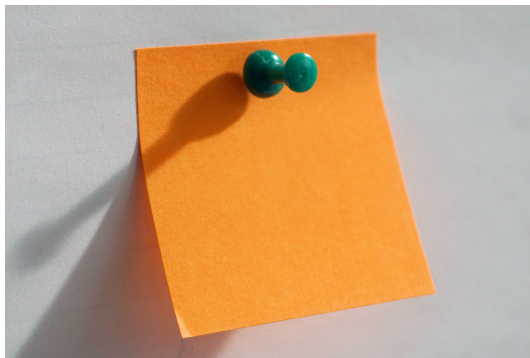
<https://github.com/IDR/idr-notebooks>



# IDR – your opinions



**GREEN** – one thing you **like** about the IDR



**ORANGE** – one thing you would like **added or changed** in the IDR

# Thanks to the Funders



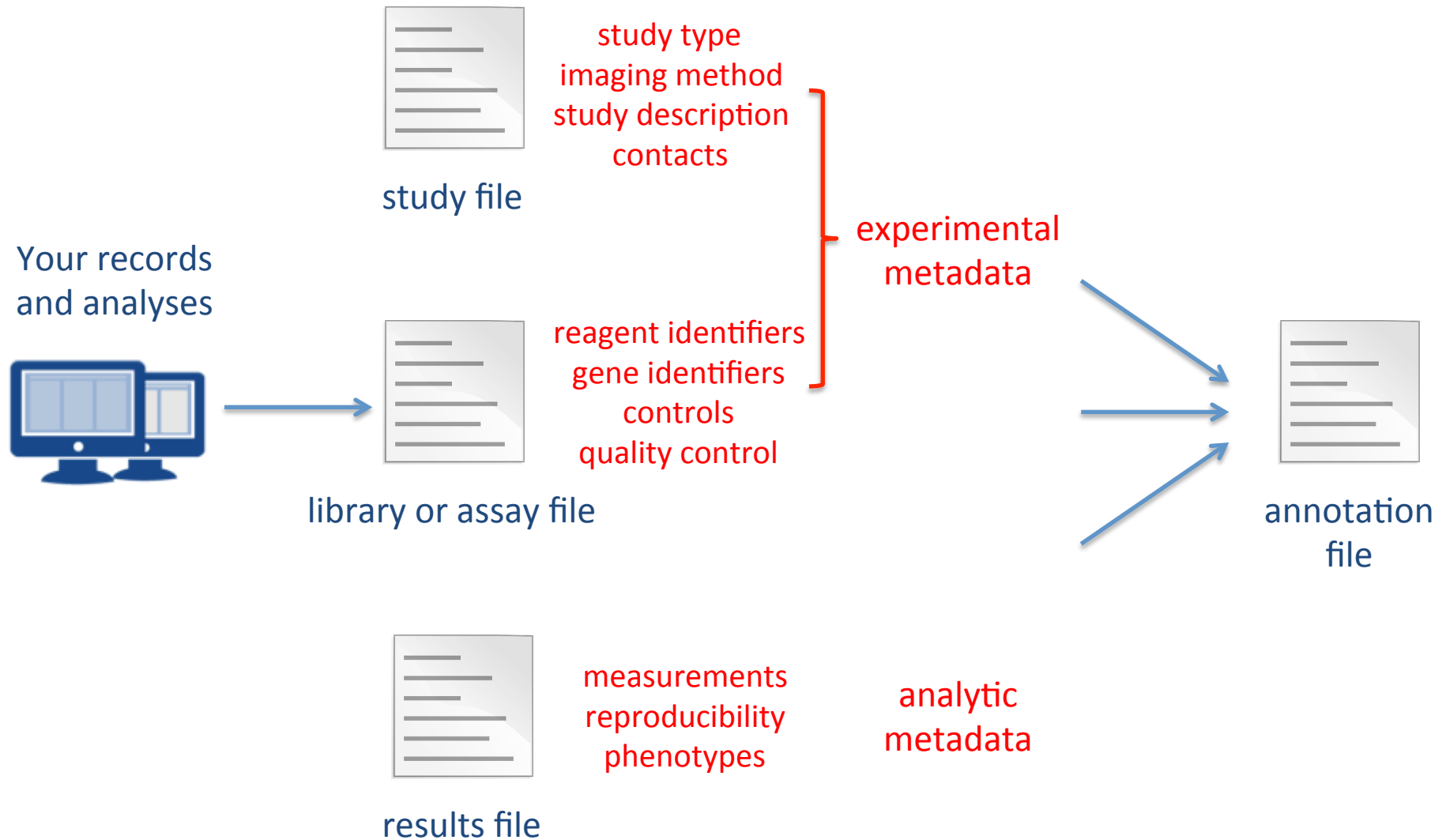


# Annotating your own images in OMERO and preparing data for submission

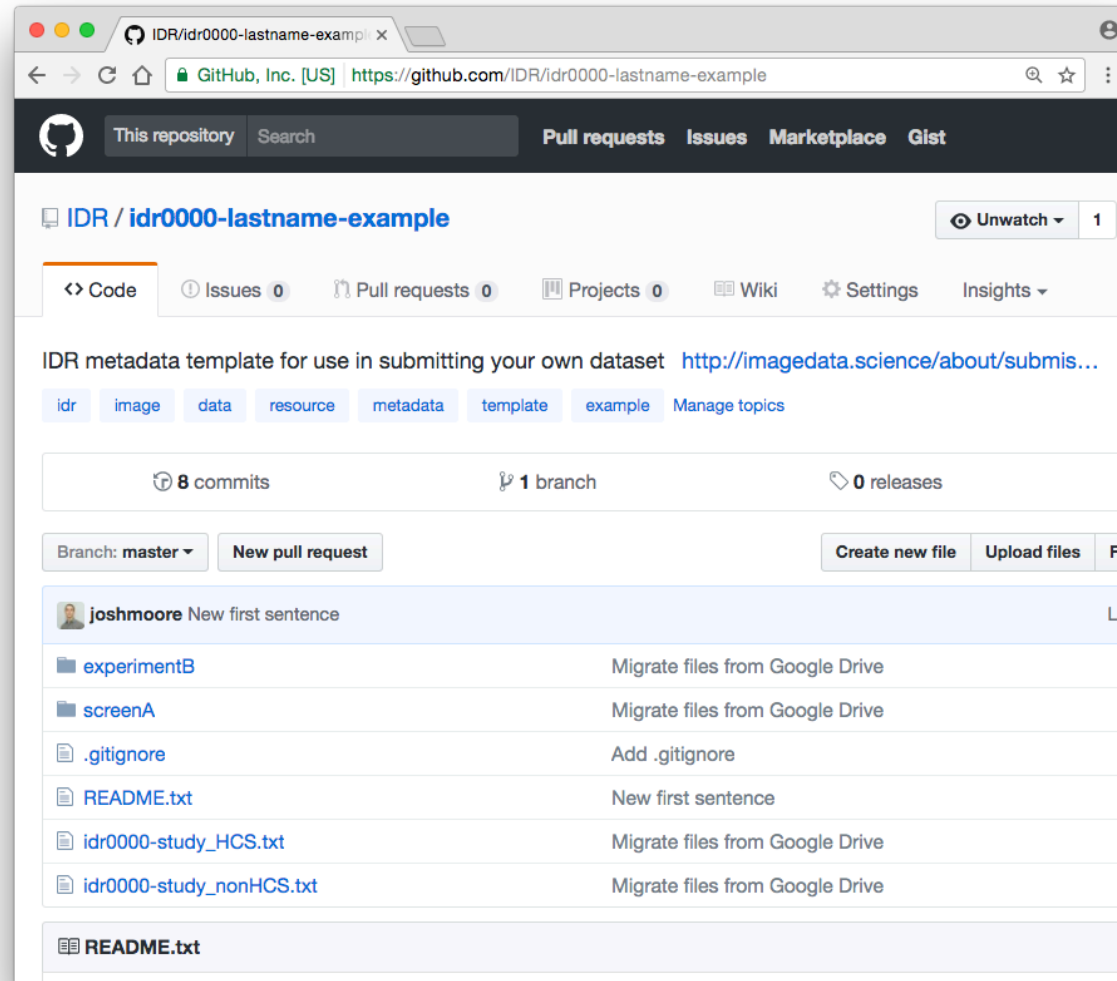
- What is the format for recording IDR annotations?
- How are the annotations added to the images in IDR?



# IDR metadata templates



# How are the IDR annotations formatted?



<https://github.com/IDR/idr-metadata>

# How do we add the annotations to IDR?

## 1. Add the annotations.csv file to a screen or experiment

```
/bin/omero metadata populate --file idr0002-screenA-  
annotation.csv Screen:102
```

```
/bin/omero metadata populate --file idr0021-experimentA-  
annotation.csv Project:51
```

# How do we add the annotations to IDR?

The screenshot shows the IDR web interface. On the left, a tree view under 'Demo data' lists various studies, including 'idr0002-heriche-condensation/screenA 12'. The main panel displays the 'General' tab for this study, showing attributes like 'Study Type' (high content screen), 'Screen Type' (primary screen), 'Screen Technology Type' (RNAi screen), and 'Imaging Method' (fluorescence microscopy). Below the attributes, the 'Attachments' section is visible, showing a list of files. A red arrow points from the 'Attachments' section in the main panel to a detailed view of the attachments at the bottom of the slide.

Attributes	
Added by: Demo User	
Study Type	high content screen
Screen Type	primary screen
Screen Technology Type	RNAi screen
Imaging Method	fluorescence microscopy
Publication Title	Integration of biological data by kernels on graph nodes allows prediction of new genes involved in mitotic chromosome condensation.
Publication Authors	Heriche JK, Lees JG, Morilla I, Walter T, Petrova B, Roberti MJ, Hossain MJ, Adler P, Fernandez JM, Krallinger M, Haering CH, Vilo J, Valencia A, Ranea JA, Orengo C, Eilenberg J.
PubMed ID	24943848 <a href="http://www.ncbi.nlm.nih.gov/pubmed/24943848">http://www.ncbi.nlm.nih.gov/pubmed/24943848</a>
PMC ID	PMC4142622 <a href="http://europepmc.org">http://europepmc.org</a>

Attachments	
bulk_annotations (777.17 KB)	
idr0002-heriche-condensation-screenA-plate-all.h5 (992.65 GB)	

## Attachments

bulk\_annotations (777.17 KB)

Annotations.csv file once attached is named bulk\_annotations

# How do we add the annotations to IDR?

The screenshot displays the IDR web interface. On the left, a sidebar shows a tree view of 'Demo data' with various screen and plate entries. The main area features a grid of microscopy images, with a red arrow pointing to a specific well (row A, column 4). The right panel, titled 'General', contains a 'Tables' dropdown menu and a detailed metadata section for the selected image. The metadata includes information about the plate, well, organism, cell line, siRNA, and gene annotations.

Table	Value
Has Phenotype	yes
Phenotype Annotation Level	gene
envelope breakdown	

Table	Value
INFO	
Plate:	422
Well Number:	4
Well:	67114
Characteristics [Organism]:	Homo sapiens
Term Source 1 REF:	NCBITaxon
Term Source 1 Accession:	NCBITaxon_9606
Characteristics [Cell Line]:	HeLa
Term Source 2 REF:	EFO
Term Source 2 Accession:	EFO_0001185
siRNA Identifier:	s15534
Sense Sequence:	GCAGGUUUCUCAUCGAUUUtt
Antisense Sequence:	AAAUCGAUGAGAAACCUGCca
Reagent Design Gene Annotation Build:	GRCh37, Ensembl release 61, Feb 2011
Gene Identifier:	ENSG00000083168
Gene Symbol:	MYST3
Gene Symbol Synonyms:	KATE6A
Analysis Gene Annotation Build:	GRCh37, Ensembl release 61, Feb 2011
Quality Control:	pass
Channels:	H2B- mCherry/Cy3:chromatin;eGFP:nuclear lamina and report on nuclear envelope breakdown
Median Deviation Fraction - Shorter	-0.09

# How do we add the annotations to IDR?

2. Convert selected annotations from the bulk table to key value pairs (map annotations)

```
/bin/omero metadata populate --context bulkmap --cfg  
idr0002-screenA-bulkmap-config.yml Screen:102
```

# How do we add the annotations to IDR?

```
#####  
# mapr groups  
#####  
  
- group:  
  namespace: openmicroscopy.org/mapr/organism  
  columns:  
    - name: Characteristics [Organism]  
      clientname: Organism  
      include: yes  
  
- group:  
  namespace: openmicroscopy.org/mapr/sirna  
  columns:  
    - name: siRNA Identifier  
      include: yes  
  
- group:  
  namespace: openmicroscopy.org/mapr/sirna/supplementary  
  columns:  
    - name: Sense Sequence  
      include: yes  
    - name: Antisense Sequence  
      include: yes  
    - name: Reagent Design Gene Annotation Build  
      include: yes  
  
- group:  
  namespace: openmicroscopy.org/mapr/gene  
  columns:  
    - name: Gene Identifier  
      include: yes  
    - name: Gene Identifier  
      clientname: Gene Identifier URL  
      clientvalue: http://www.ensembl.org/id/{{ value|urlencode }}  
      include: yes  
    - name: Gene Symbol  
      include: yes  
  
- group:  
  namespace: openmicroscopy.org/mapr/gene/supplementary  
  columns:  
    - name: Gene Symbol Synonyms  
      include: yes  
    - name: Gene Annotation Comments
```

Idr0002-screenA-bulkmap-config.yml



# How do we add the annotations to IDR?

Tables	
INFO	
Plate:	422
Well Number:	4
Well:	67114
Characteristics [Organism]:	Homo sapiens
Term Source 1 REF:	NCBITaxon
Term Source 1 Accession:	NCBITaxon_9606
Characteristics [Cell Line]:	HeLa
Term Source 2 REF:	EFO
Term Source 2 Accession:	EFO_0001185
siRNA Identifier:	s15534
Sense Sequence:	GCAGGUUUCUCAUCGAUUUtt
Antisense Sequence:	AAAUCGAUGAGAAACCUGCca
Reagent Design Gene Annotation Build:	GRCh37, Ensembl release 61, Feb 2011
Gene Identifier:	ENSG00000083168
Gene Symbol:	MYST3
Gene Symbol Synonyms:	KATE6A
Analysis Gene Annotation Build:	GRCh37, Ensembl release 61, Feb 2011
Quality Control:	pass
Channels:	H2B- mCherry/Cy3:chromatin;eGFP:nuclear lamina and report on nuclear envelope breakdown
Median Deviation Fraction - Shorter	-0.09

Bulk table

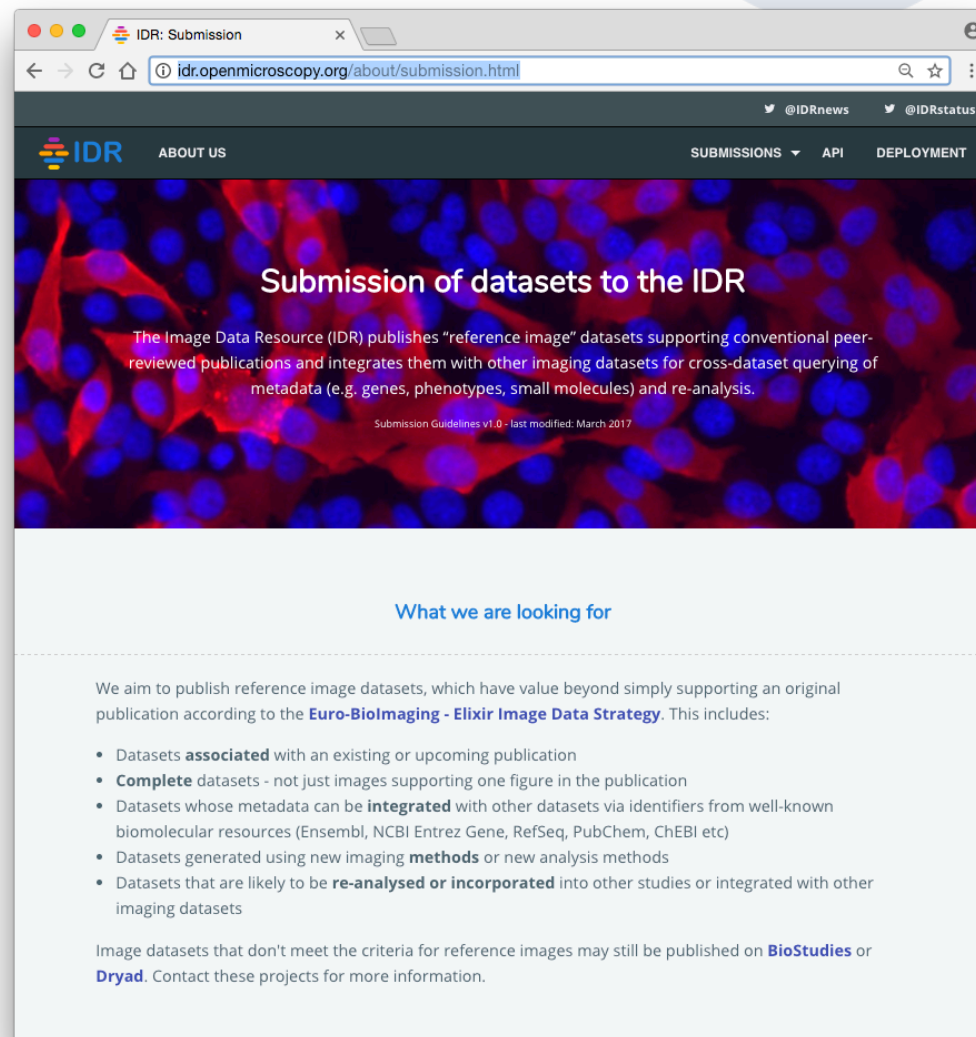
bulkmap-config

Attributes	
<b>Gene</b>	
Added by: Demo User	
Gene Identifier	ENSG00000083168
Gene Symbol	MYST3
<b>Gene supplementary</b>	
Added by: Demo User	
Gene Symbol Synonyms	KATE6A
Analysis Gene Annotation Build	GRCh37, Ensembl release 61, Feb 2011
<b>Organism</b>	
Added by: Demo User	
Organism	Homo sapiens

Key value pairs  
(aka map annotations)



# Submitting data

A screenshot of a web browser showing the IDR Submission page. The browser's address bar displays 'idr.openmicroscopy.org/about/submission.html'. The page features a dark header with the IDR logo, 'ABOUT US', and navigation links for 'SUBMISSIONS', 'API', and 'DEPLOYMENT'. Social media handles '@IDRnews' and '@IDRstatus' are also present. The main content area has a background image of red and blue fluorescent cells. It includes the title 'Submission of datasets to the IDR', a paragraph explaining the IDR's mission, and a link to 'Submission Guidelines v1.0'. Below this, a section titled 'What we are looking for' lists criteria for dataset acceptance, such as being associated with a publication, complete, integrated with other datasets, generated with new methods, or likely to be re-analysed. It also mentions alternative publication venues like BioStudies or Dryad for datasets not meeting the criteria.

Submission of datasets to the IDR

The Image Data Resource (IDR) publishes “reference image” datasets supporting conventional peer-reviewed publications and integrates them with other imaging datasets for cross-dataset querying of metadata (e.g. genes, phenotypes, small molecules) and re-analysis.

Submission Guidelines v1.0 - last modified: March 2017

### What we are looking for

We aim to publish reference image datasets, which have value beyond simply supporting an original publication according to the [Euro-BioImaging - Elixir Image Data Strategy](#). This includes:

- Datasets **associated** with an existing or upcoming publication
- **Complete** datasets - not just images supporting one figure in the publication
- Datasets whose metadata can be **integrated** with other datasets via identifiers from well-known biomolecular resources (Ensembl, NCBI Entrez Gene, RefSeq, PubChem, ChEBI etc)
- Datasets generated using new imaging **methods** or new analysis methods
- Datasets that are likely to be **re-analysed or incorporated** into other studies or integrated with other imaging datasets

Image datasets that don't meet the criteria for reference images may still be published on [BioStudies](#) or [Dryad](#). Contact these projects for more information.

# Thanks to the Funders

