



<https://idr.openmicroscopy.org>

Image Data Integration and Publication at Scale: Update on The IDR

OME Community Meeting
May 29th 2024

Frances Wong & Will Moore

University of Dundee

OME

@openmicroscopy, @IDRNews

Overview of IDR

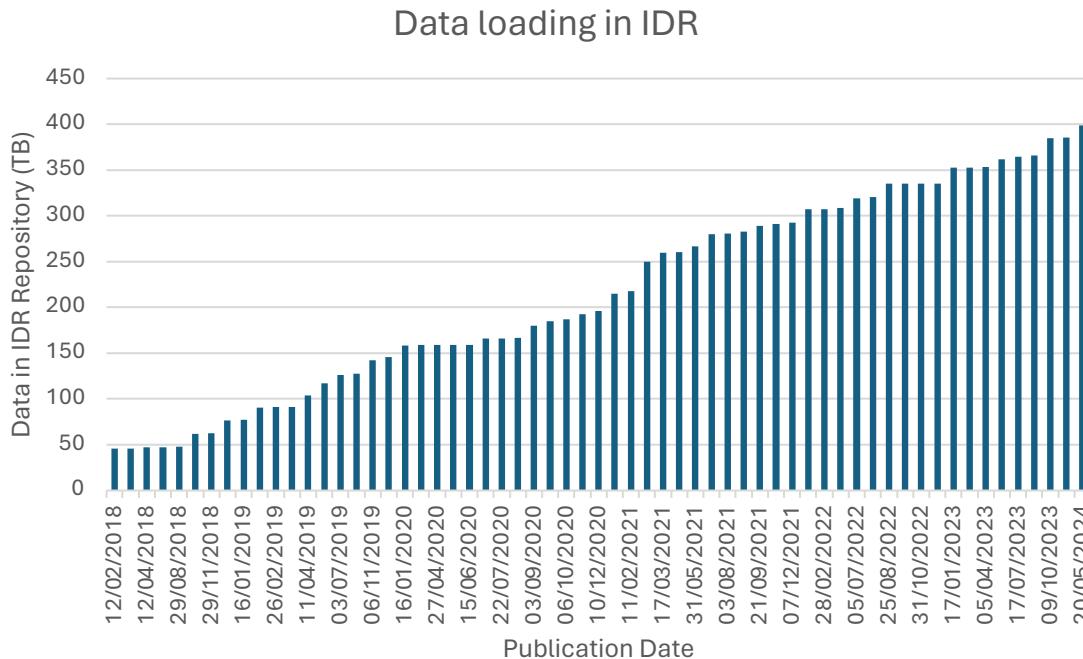
- **Public access** (<https://idr.openmicroscopy.org>)
- **Reference datasets** - complete datasets containing molecular and functional annotations, associated with an existing or upcoming publication.
- **Study integration** - integrating studies or datasets with other datasets via **genes, compounds or phenotypes**.
- **Curated metadata**
- **Cloud re-analysis**

IDR metrics

May 2024

Raw data size	399 TB
Image files	38 M
5D Images	14 M
2D Planes	118 M
Organisms	84
Genes	89 K
Antibodies	12 K
Compounds	40 K

IDR metrics



IDR studies and publications

**>125 cross-published
studies**



***cross-referenced via
accessions and DOIs***

A high-content RNAi screen reveals multiple roles for long noncoding... PDF | | | |

Data availability

A reporting summary for this article is available as a [Supplementary Information file](#). Sequencing data are available in the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress>) with the accession codes [E-MTAB-7432](#) (RNA-seq), [E-MTAB-7418](#) (CHART-seq) and [E-MTAB-7419](#) (CUT&RUN). The imaging data have been submitted to the Image Data Resource (<https://idr.openmicroscopy.org>) under IDR accession number [idr0056](#). The source data

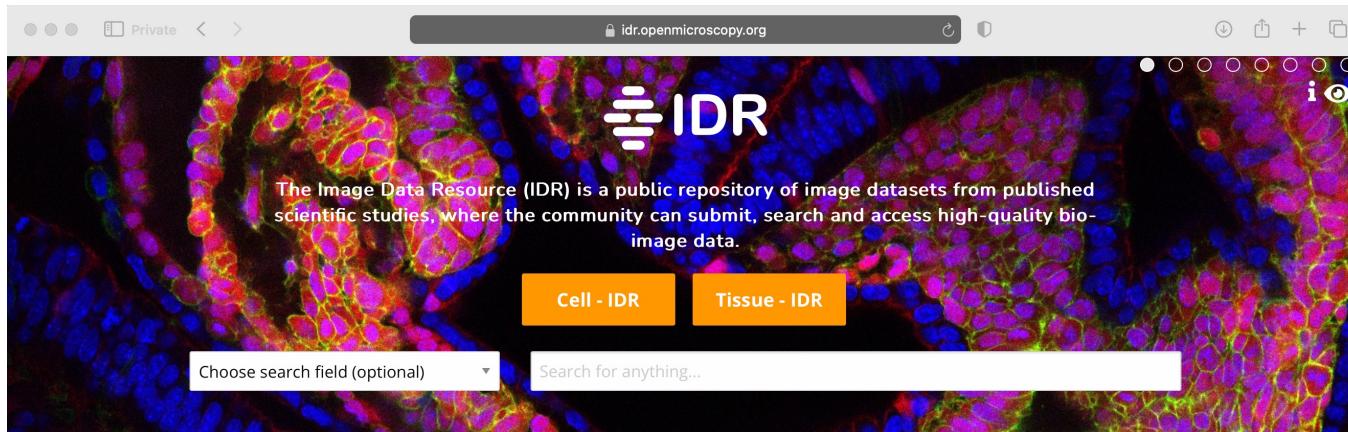
An upward-pointing arrow originates from the search bar on the IDR website, and a downward-pointing arrow points to the search results page below it.

The IDR website interface shows a search bar with "Search by: Name (IDR number)" and "idr0056". Below the search bar, it says "Found 3 studies with Name: idr0056". Three study entries are listed, each with a thumbnail image and a brief description:

- idr0056C Stojic L et al. A long noncoding RNA regulates microtubule behaviour during mitosis
- idr0056B Stojic L et al. A long noncoding RNA regulates microtubule behaviour during mitosis
- idr0056A Stojic L et al. A long noncoding RNA regulates microtubule behaviour during mitosis

<https://idr.openmicroscopy.org/about/studies.html>

IDR Homepage

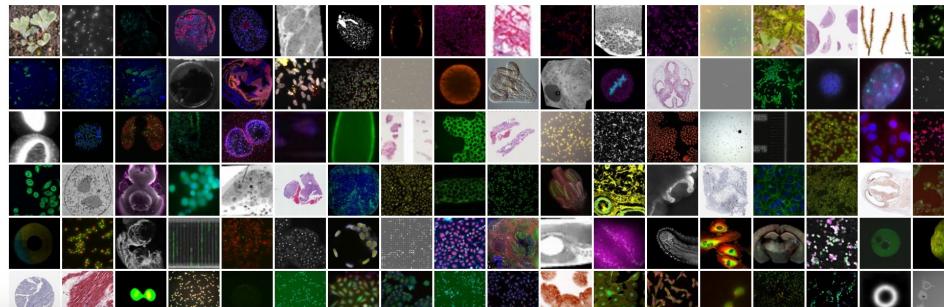


128 Studies

14,018,967 Images

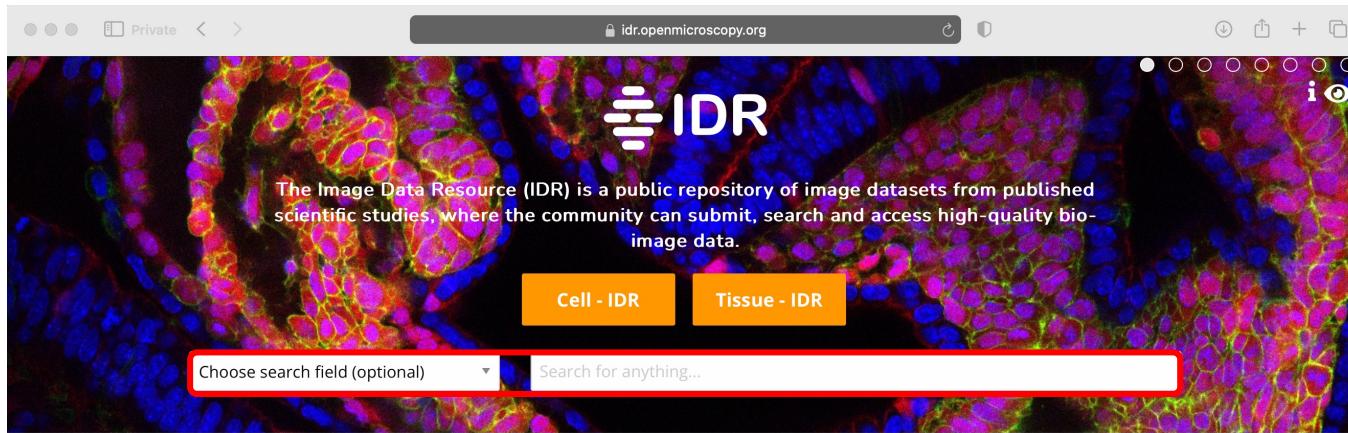
399 TB

Group Studies by type



<https://idr.openmicroscopy.org>

IDR Search

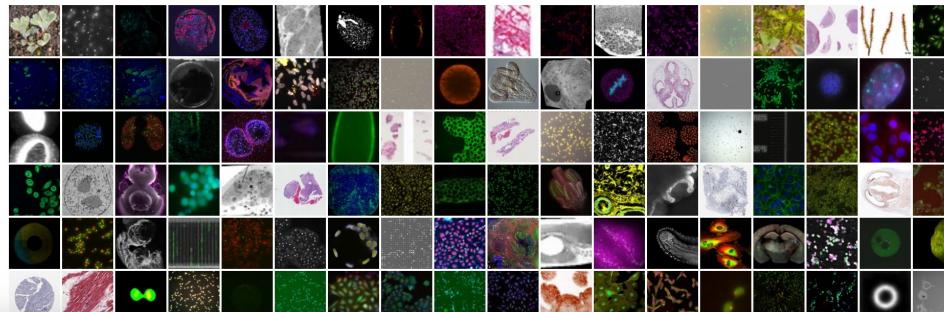


128 Studies

14,018,967 Images

399 TB

Group Studies by type



<https://idr.openmicroscopy.org>

IDR Search

The Image Data Resource (IDR) is a public repository of image datasets from published scientific studies, where the community can submit, search and access high-quality bio-image data.

Cell - IDR Tissue - IDR

Choose search field (optional) X

2342 Images matched **Gene Symbol:** pax6

991 Images matched **Antibody:** pax6

9 Images matched **HGNC Gene Symbol:** pax6

6 Images matched **Nearest Human Homologue:** pax6

4 Images matched **Original Gene Target:** pax6

3 Images matched **Gene Symbol Synonyms:** eyeless tumor-head-63 l(4)102cdh ey eyel pax-6 ok107 ey/pax6 eye l(4)33 pax6 ey dpax-6 l(4)102cdr pax6 cg1464

<https://idr.openmicroscopy.org>

IDR Search

The screenshot shows the IDR Search interface on a web browser. The URL in the address bar is idr.openmicroscopy.org. The main search query is "Gene Symbol equals pax6". The results show 2342 images across 9 experiments/screens. One result, "idr0070A", is highlighted with a red box.

ID	Images	Publication Title
idr0043A	1027	Proteomics. Tissue-based map of the human proteome.
idr0070A	855	The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene ...
idr0022A	288	Uncovering the signaling landscape controlling breast cancer cell migration identi...
idr0114A	136	Enabling research with human embryonic and fetal tissue resources
idr0009A	14	Genome-wide RNAi screening identifies human proteins with a regulatory functio...
idr0013A	9	Phenotypic profiling of the human genome by time-lapse microscopy reveals cell ...
idr0093A	9	High content genome-wide siRNA screen to investigate the coordination of cell siz...
idr0010A	2	RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to a...
idr0012A	2	Clustering phenotype populations by genome-wide RNAi and multiparametric im...

<https://idr.openmicroscopy.org>

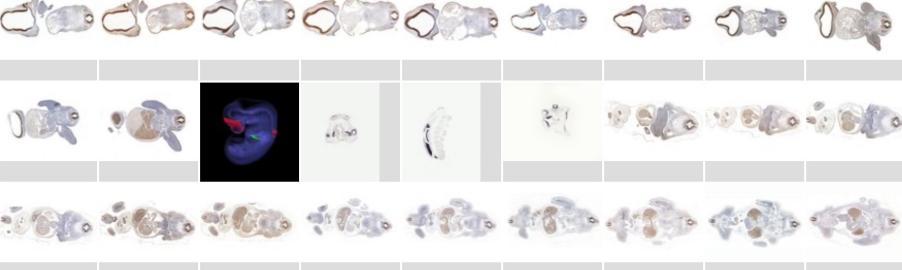


IDR Search

Screenshot of the IDR Search interface showing search results for "Gene Symbol equals pax6".

Search Query: Gene Symbol equals pax6

Results: 2342 images in 9 experiments/screens

ID	Images	Publication Title
idr0043A	1027	Proteomics. Tissue-based map of the human proteome.
idr0070A	855	The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene ...
		
idr0022A	288	Uncovering the signaling landscape controlling breast cancer cell migration identi...
idr0114A	136	Enabling research with human embryonic and fetal tissue resources

IDR Search

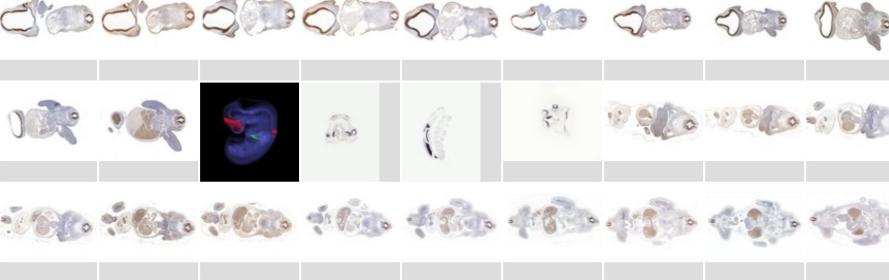
Screenshot of the IDR Search interface showing a search for gene symbol "pax6".

The search query is highlighted with a red box:

Attribute	Operator	Value
Gene Symbol	equals	pax6

Search results:

Search **Gene Symbol** equals **pax6** found **2342** images in **9** experiments/screens

ID	Images	Publication Title
idr0043A	1027	Proteomics. Tissue-based map of the human proteome.
idr0070A	855	The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene ...
		
idr0022A	288	Uncovering the signaling landscape controlling breast cancer cell migration ident...
idr0114A	136	Enabling research with human embryonic and fetal tissue resources

IDR Search

Attribute Operator Value

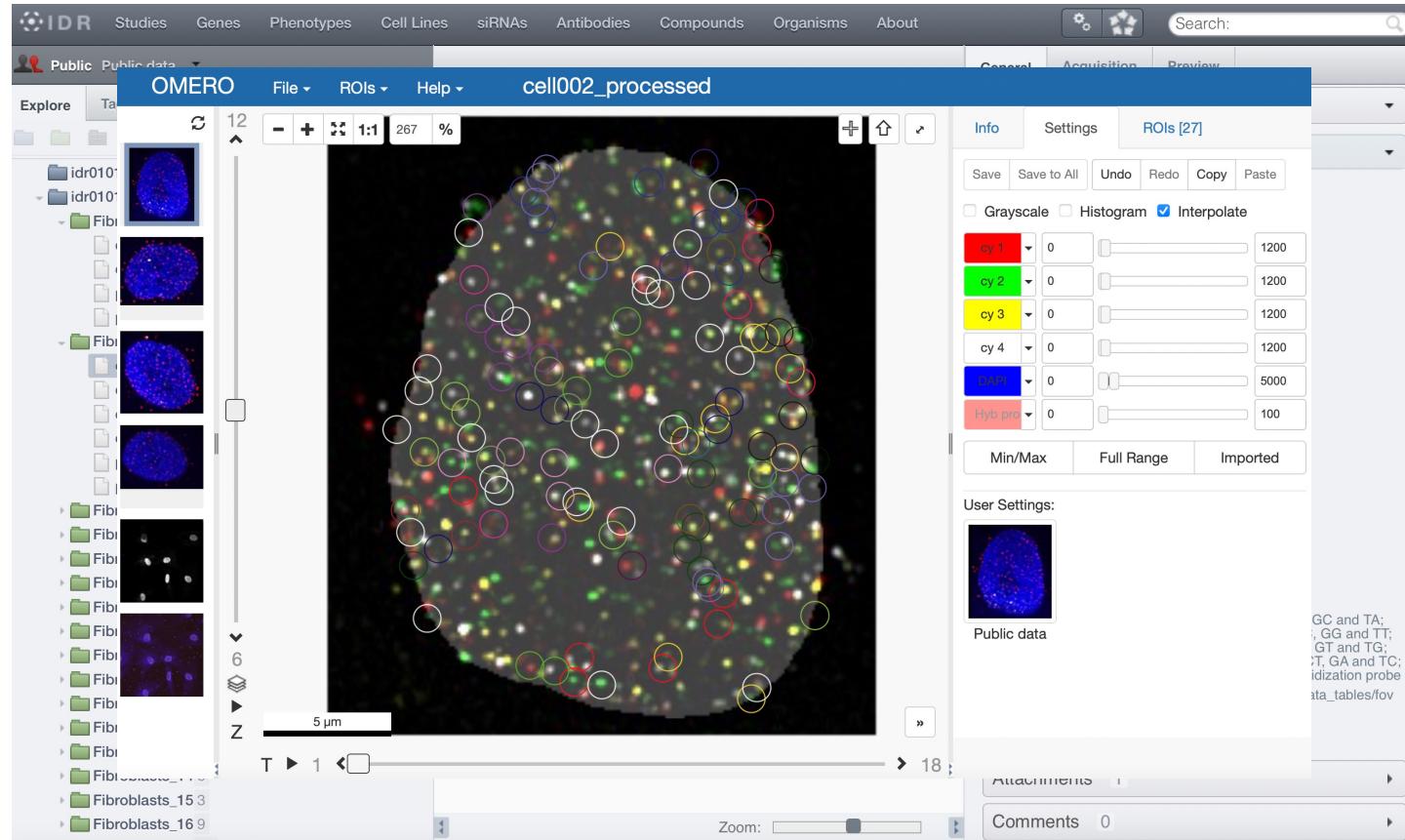
Gene Symbol equals pax6

Any
Study
Imaging Method
License
Name (IDR number)
Organism
Publication Authors
Publication Title
Screen Technology Type
Screen Type
Study Type
Image
Antibody
Antibody Identifier
Cell Line
Compound Name
Gene Identifier
✓ Gene Symbol
InChIKey
Pathology
Pathology Identifier
Phenotype
Phenotype Term Accession
Protein
PubChem CID
siRNA Identifier
siRNA Pool Identifier

contains
✓ equals

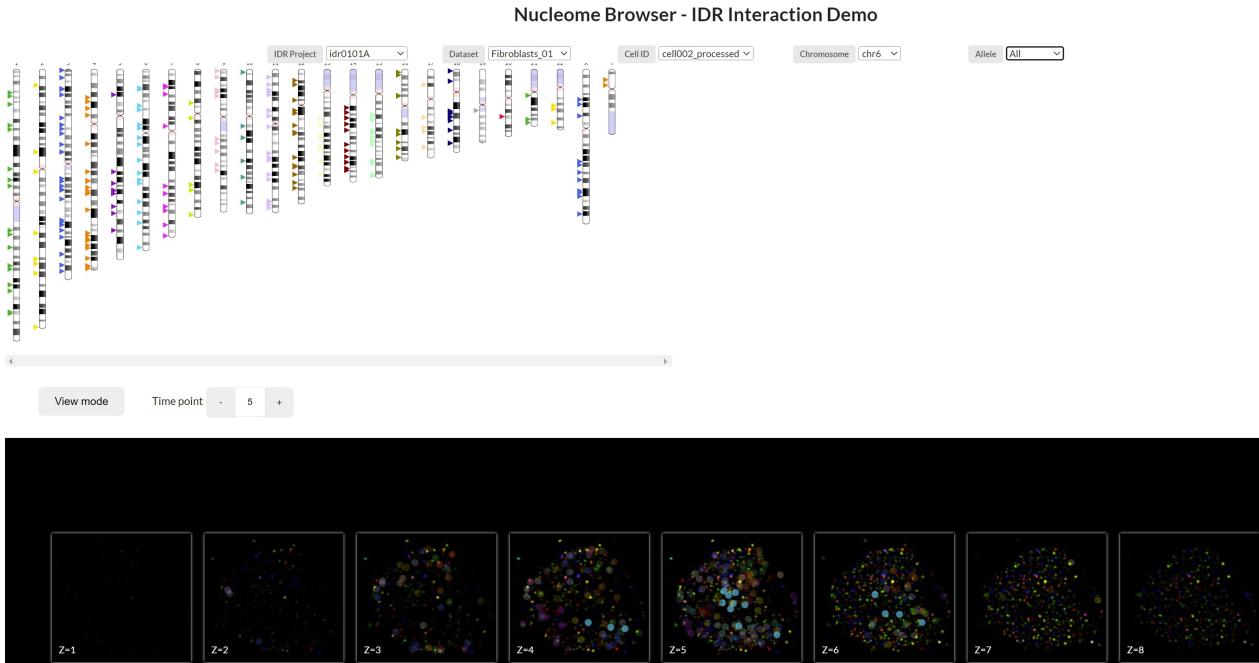
The screenshot shows the IDR Search interface with a search query. The 'Attribute' field contains 'Gene Symbol', the 'Operator' field contains 'equals', and the 'Value' field contains 'pax6'. A red box highlights the 'Gene Symbol' field and the '>equals' operator. Red arrows point from these highlighted fields to a dropdown menu below, which contains the options 'contains' and '✓ equals'. The '✓ equals' option is selected.

In Situ Sequencing - Linking Multimodal Metadata



idr0101, Payne et al Science DOI:10.17867/10000169

In Situ Sequencing - Linking Multimodal Metadata



Jhu et al, 2022, *Nature Methods*, 10.1038/s41592-022-01559-3
https://vis.nucleome.org/static/apps/IDR_demo/

Data Download and Reuse



Imaging data for all studies published in IDR is available for download using FTP or Globus -

<https://idr.openmicroscopy.org/about/download.html>



Datasets are published under the Creative Commons Attribution 4.0 International license (CC BY 4.0) or a more permissive license.



- **Share** — copy and redistribute the material in any medium or format.
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.
- **Attribution** — credit must be given to the creator.

Accessing IDR Data

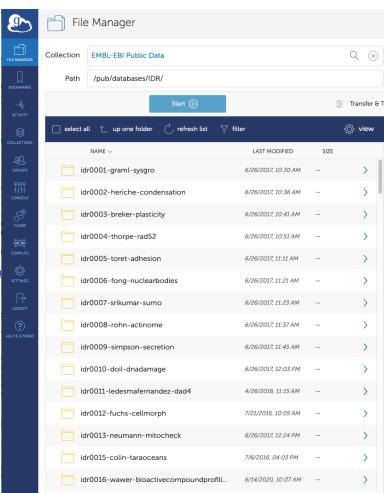
Download

<https://ftp.ebi.ac.uk/pub/databases>IDR/>

Index of /pub/databases>IDR

Name	Last modified	Size	Description
------	---------------	------	-------------

Parent Directory
idr0001-grant-sysgro/
idr0002-heriche-condensation/
idr0003-breker-plasticity/
idr0004-thorpe-rad52/
idr0005-toret-adhesion/
idr0006-fong-nuclearbodies/
idr0007-srikumar-sumo/
idr0008-rohn-actinome/
idr0009-simpson-secretion/
idr0010-doll-dnamage/
idr0011-ledesmafernandez-dad4/
idr0012-fuchs-cellmorph/
idr0013-neumann-mitochondria/
idr0015-colin-taraocceans/
idr0016-wawer-bioactivecompound/
idr0017-brening-drugscreen/
idr0018-neff-histopathology/
idr0019-sero-infkappp/
idr0020-barr-chtop/



FTP

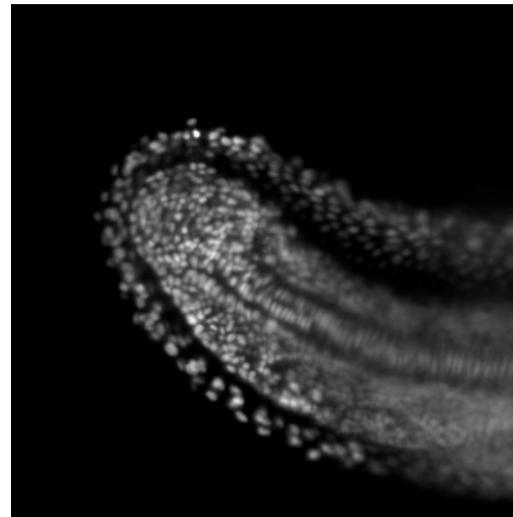
Globus

OMERO

API

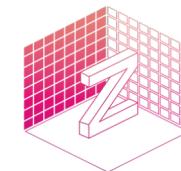
```
conn = BlitzGateway("username", "password")
conn.connect()
image = conn.getObject("Image", image_id)
plane = image.getPrimaryPixels().getPlane(theT=10)
```

Web Client



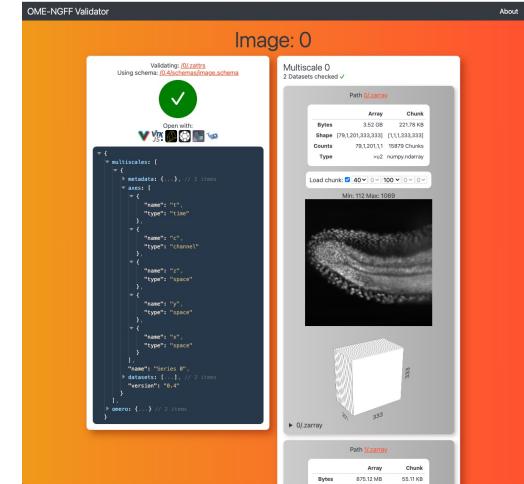
idr0051, Attardi et al, Development DOI:10.1242/dev.166728

OME-NGFF



Coming soon!

OME-Zarr



Validator



Support for OME-NGFF Data in IDR

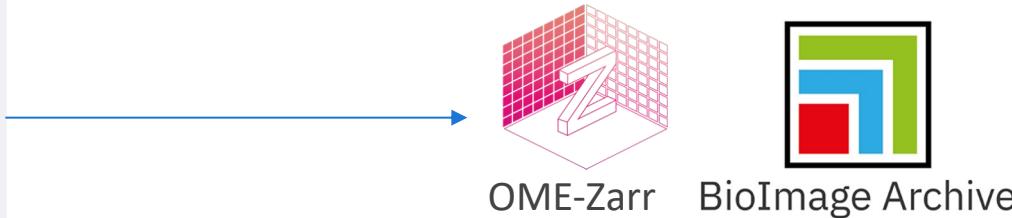
What have we done?

NB: Not released yet

The screenshot shows the IDR (Image Data Resource) web interface. At the top, there's a navigation bar with 'Public' and 'Public data' dropdowns, and tabs for 'Explore', 'Tags', and 'Shares'. Below the navigation is a toolbar with various icons for file operations like upload, download, and delete. The main area is titled 'Public data' and contains a tree view under 'OME-NGFF'. The tree view lists 17 studies, each with a folder icon and a study ID followed by a screen name and count. The studies are:

- idr0004-thorpe-rad52/screenA 46
- idr0010-doil-dnadmamage/screenA 148
- idr0011-ledesmafernandez-dad4/screenA 129
- idr0011-ledesmafernandez-dad4/screenB 40
- idr0011-ledesmafernandez-dad4/screenC 4
- idr0011-ledesmafernandez-dad4/screenD 8
- idr0011-ledesmafernandez-dad4/screenE 1
- idr0012-fuchs-cellmorph/screenA 68
- idr0013-neumann-mitochcheck/screenA 510
- idr0013-neumann-mitochcheck/screenB 28
- idr0015-colin-taraoceans/screenA 83
- idr0016-wawer-bioactivecompoundprofiling/screenA 413
- idr0025-stadler-proteinatlas/screenA 3
- idr0026-weigelin-immunotherapy/experimentA 18
- idr0033-rohbar-pathways/screenA 12
- idr0035-caie-drugresponse/screenA 55
- idr0036-gustafsdottir-cellpainting/screenA 20
- idr0051-fulton-tailbulightsheet/experimentA 3
- idr0054-segura-tonsilhyperion/experimentA 1
- idr0064-goglia-erkdynamics/screenA 9
- idr0090-ashdown-malaria/screenA 22
- idr0091-julou-lacinduction/experimentA 22

Converted 17 studies into OME-Zarr



Data hosted on BioImage Archive S3 and served via OMERO (IDR)

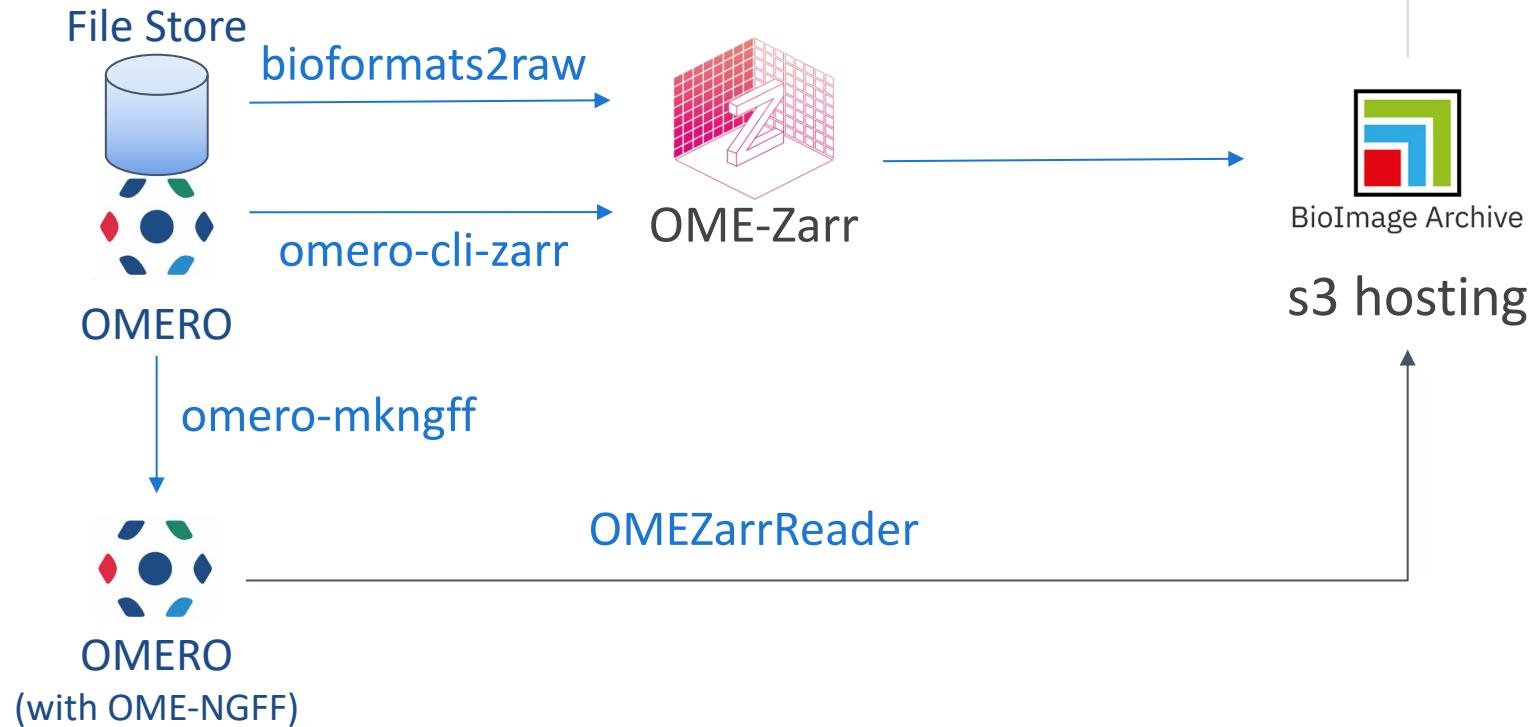
~47 TB, ~1600 Plates, 44 Datasets

Support for OME-NGFF Data in IDR

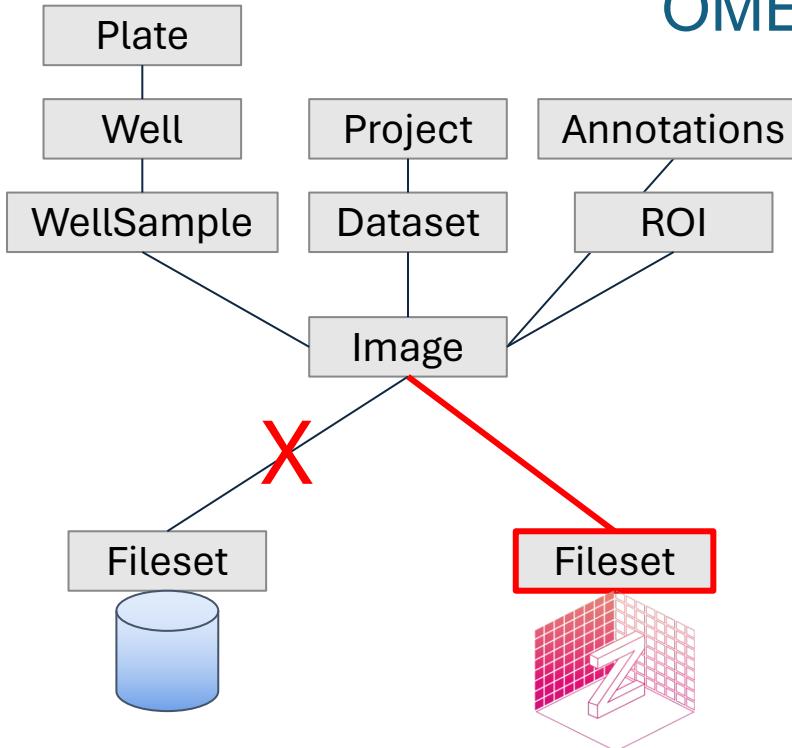
Why have we done this?

- IDR fork of Bio-Formats was required to read IDR-specific data formats
 - Want to bring IDR into line with mainline Bio-Formats
 - Solution: Convert legacy formats into OME-NGFF
- Host the data on the EBI Embassy Object Store
 - Improved scalability compared with NFS file system
 - New submissions of OME-NGFF data hosted by submitters
- OME-NGFF data available directly via S3

https://uk1s3.embassy.ebi.ac.uk/bia-integrator-data/pages/idr_ngff_data.html



OMERO-mkngff

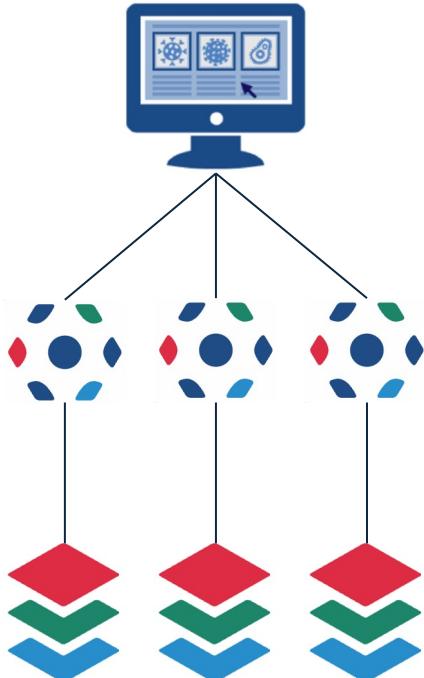


Creates a new NGFF-based Fileset in the OMERO database, replacing original Fileset

<https://github.com/IDR/omero-mkngff>

```
$ omero mkngff sql [fileset ID] /path/to/fileset.zarr > myNgff.sql  
$ psql -U omero -d idr -h $DBHOST -f myNgff.sql
```

OME-NGFF performance improvements



OMERO.iviewer: reduce requests rate

OMERO micro-services: distributed rendering

Bio-Formats OMEZarrReader: read directly via S3

OME-NGFF in IDR

IDR Studies Genes Phenotypes Cell Lines siRNAs Antibodies Compounds Organisms About

Public

Explore Tags Shares Index Field#1

Public data

- OME-NGFF
 - idr0004-thorpe-rad52/screenA 46
 - idr0010-doll-dnamage/screenA 148
 - idr0011-ledesmafernandez-dad4/screenA 129
 - idr0011-ledesmafernandez-dad4/screenB 40
 - idr0011-ledesmafernandez-dad4/screenC 4
 - idr0011-ledesmafernandez-dad4/screenD 8
 - idr0011-ledesmafernandez-dad4/screenE 1
 - idr0012-fuchs-cellmorph/screenA 68
 - HT01
 - HT02
 - HT03
 - HT04
 - HT05
 - HT06
 - HT07
 - HT08
 - HT09
 - HT10
 - HT11
 - HT12
 - HT13
 - HT14
 - HT15
 - HT16
 - HT17
 - HT18

General Acquisition Preview

Full viewer

Fileset Info
Fileset ID: 6312771

General Acquisition Preview

Full viewer

Fileset Info
Fileset ID: 6312771

File count: 4725

Imported from:

<https://uk1s3.embassy.ebi.ac.uk/bia-integrator-data/S-BIAD845/409227a9-2f3f-421c-8086-b14253283e42/409227a9-2f3f-421c-8086-b14253283e42.zarr.zattrs>

<https://uk1s3.embassy.ebi.ac.uk/bia-integrator-data/S-BIAD845/409227a9-2f3f-421c-8086-b14253283e42/409227a9-2f3f-421c-8086-b14253283e42.zarr.zgroup>

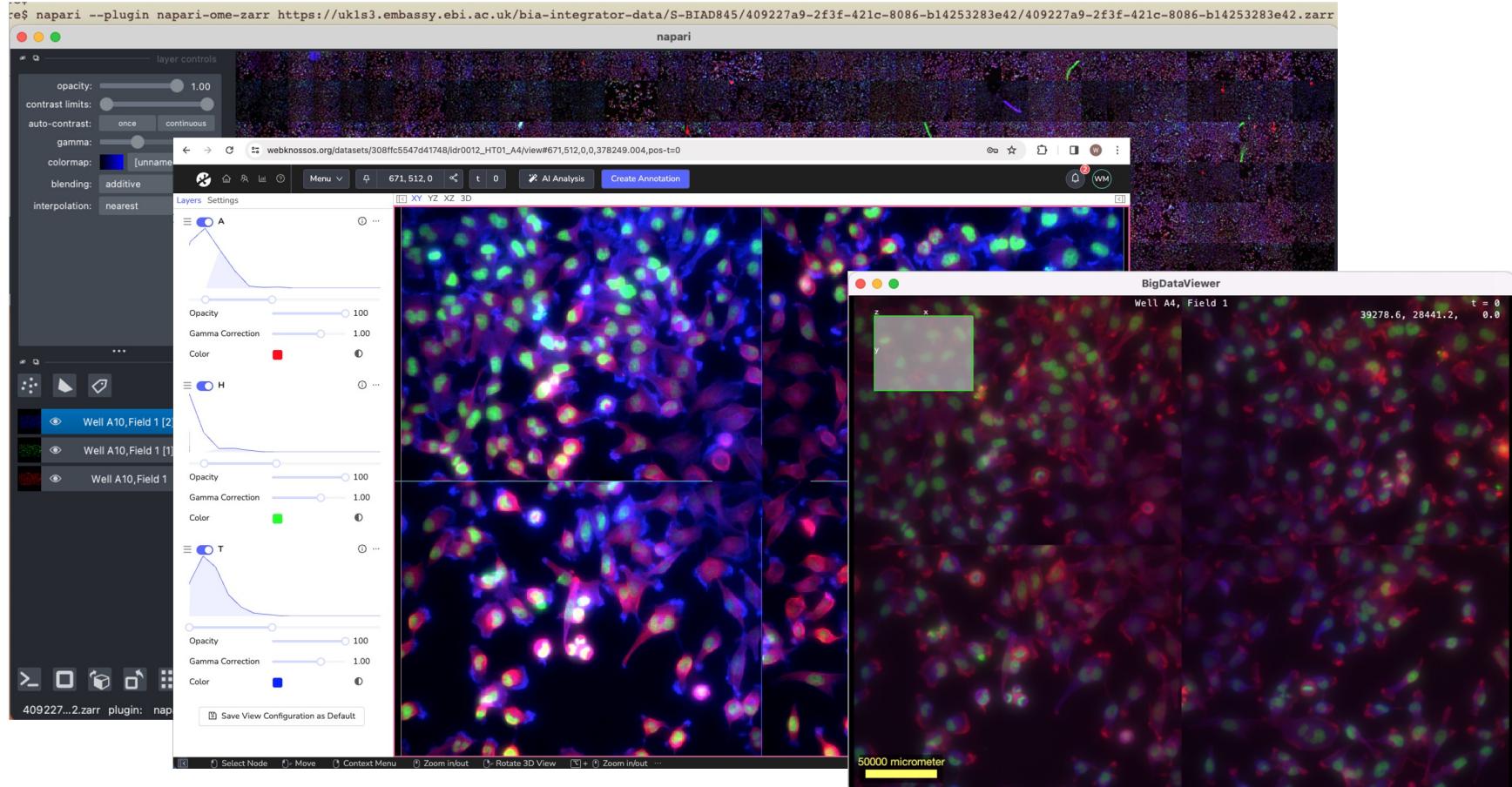
Show more...

Paths on server:

demo_2/2016-05/16/14-53-28.229_mkngff/409227a9-2f3f-421c-8086-b14253283e42.zarr.zattrs

demo_2/2016-05/16/14-53-28.229_mkngff/409227a9-2f3f-421c-8086-b14253283e42.zarr.zgroup

OME-NGFF in other tools



IDR supports FAIR principles

Findable

The screenshot shows the IDR web application interface. At the top, there is a search bar with the following fields: Attribute (Gene Symbol), Operator (equals), and Value (pax6). Below the search bar, a message states "Search Gene Symbol equals pax6 found 2342 images in 9 experiments/screens". The main area displays two experiment cards: "idr0043" and "idr007C". Each card includes a thumbnail image, a list of screen names, and a grid of small images representing individual wells. To the right of the cards, there is a file viewer window showing details for File ID: 6312771, including URLs for the dataset and a Zarr archive.

Accessible

This block contains two side-by-side images. On the left, a Jupyter notebook cell titled "jupyter idr0022 (read only)" is shown, displaying a series of histograms and a 3D scatter plot of microscopy data. On the right, a "BigDataViewer" interface shows a fluorescence microscopy image of a cell with multiple channels (red, green, blue) overlaid. A yellow scale bar at the bottom left indicates "50000 micrometer". The viewer also displays coordinates (39278.6, 28441.2) and time (t = 0.0).

Interoperable

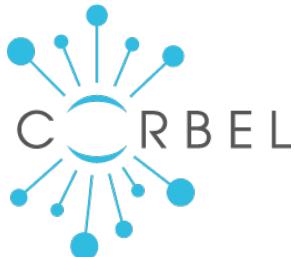
Reusable

Thanks to the funders



Biotechnology and
Biological Sciences
Research Council

Chan
Zuckerberg
Initiative 



GLOBAL
BIOIMAGING
growing collaboration



Members of the IDR Team



Jason
Swedlow



Frances
Wong



Dominik
Lindner



Josh
Moore



Jean-Marie
Burel



Petr
Walczysko



Will
Moore



David
Gault



Khaled
Mohamed

Past members

Simon Li

Mark Carroll

Riad Gozim

Eleanor Williams

Simone Leo

Gabry Rustici

Ola Tarkowska

Balaji Ramalingam

Rafael Carazo-Salas

Balint Antal

Anatole Chessel

Simon Jupp

Tony Burdett



Sebastien
Besson



Alvis
Brazma



Ugis
Sarkans



Matthew
Hartley



Kolawole
Babalola

OME 2024

Presentations available @

[https://downloads.openmicroscopy.org/presentations/2024/
Dundee](https://downloads.openmicroscopy.org/presentations/2024/Dundee)