The Image Data Resource: update and next steps

OME Community Meeting
May 26th 2020

Sébastien Besson

Frances Wong

University of Dundee

OME

https://idr.openmicroscopy.org

@IDRNews

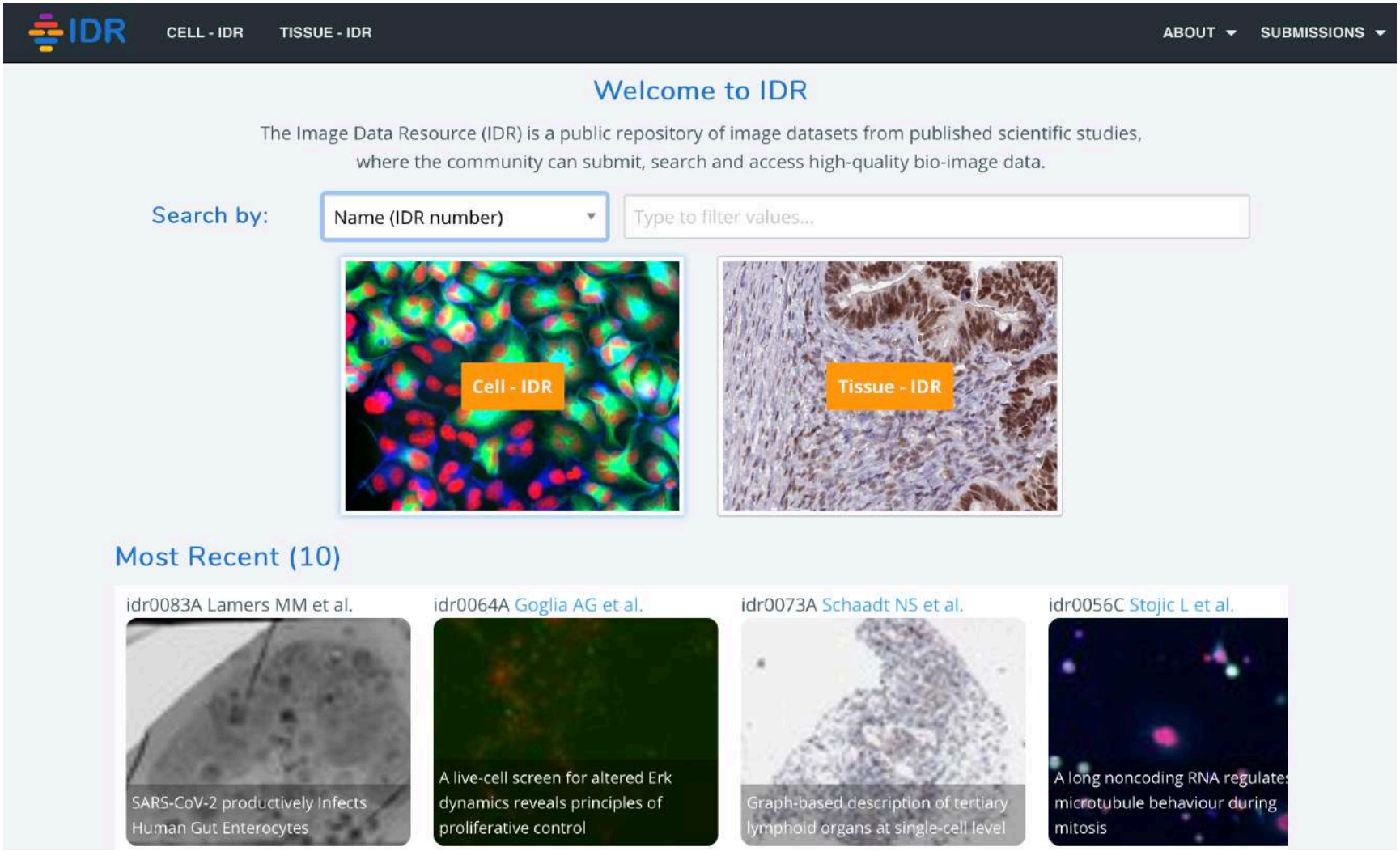
University of BRISTOL

University of Dundee

EMBL-EBI

The IDR at EMBL-EBI Embassy

IDR



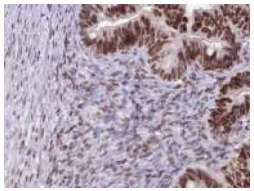
https://idr.openmicroscopy.org/

Williams et al (2017) Nature Methods

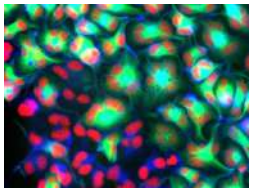
Reference image datasets

IDR

Biological Domains

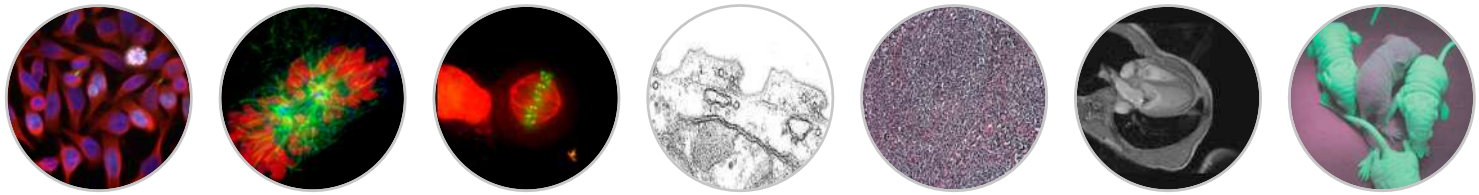


Tissue

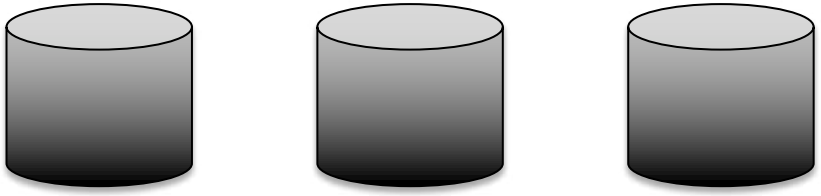


Cell

Imaging Modalities



Biomolecular Resources



Genes Compounds Cell Lines

Controlled Vocabularies

Phenotypes Ground Truth

https://idr.openmicroscopy.org/about/submission.html

IDR studies & publications

IDR

>60 cross-published studies

cross-referenced via accessions and DOIs

eLIFE

SCIENTIFIC REPORTS

nature publishing group npg

SCIENTIFIC DATA

nature

molecular systems biology

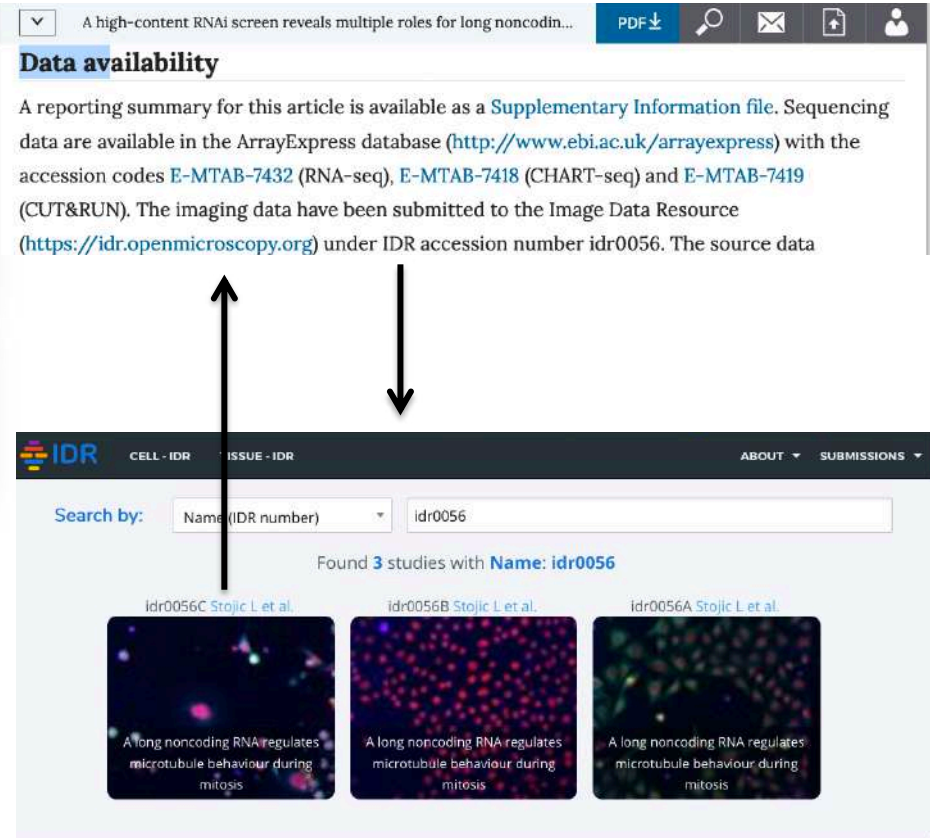
PLOS | ONE

Cell Cell Systems

Development
For advances in developmental biology and stem cells

Science AAAS

BMC Cell Biology



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

IDR metrics

May 2020	
Raw data size	169 TB
Image files	23 M
5D Images	9 M
2D Planes	68 M
Organisms	13
Genes	83 K
Antibodies	7 K
Compounds	34 K

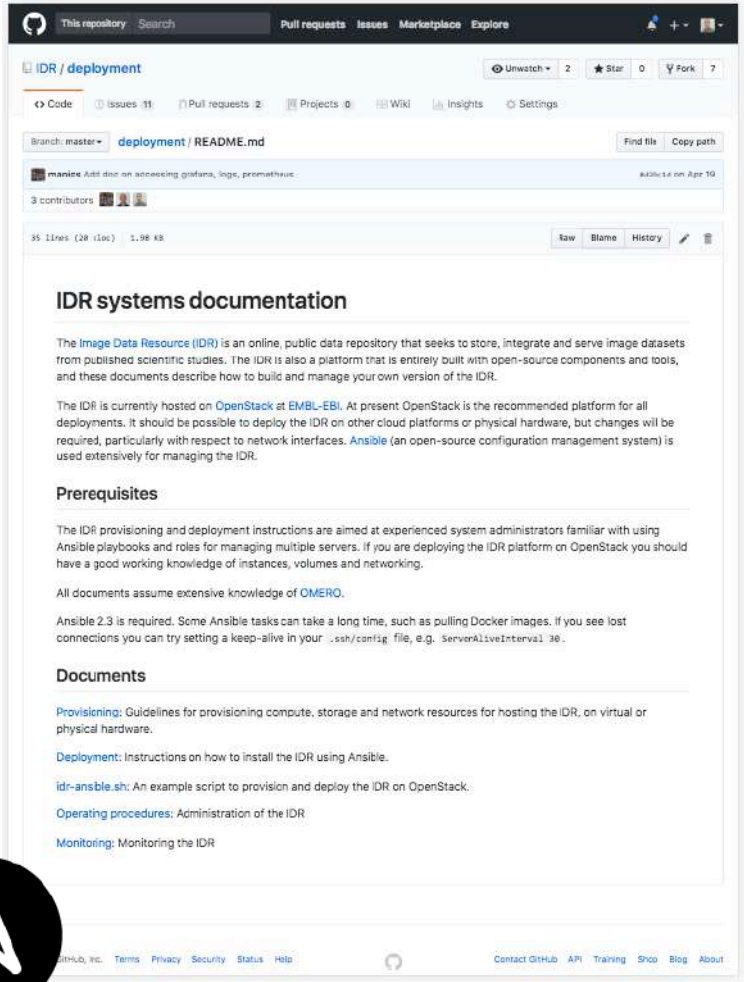
https://idr.openmicroscopy.org/

Scalability and deployment

IDR

```
$ source openstack-credentials.env

$ ansible-galaxy install -r requirements.yml

$ ansible-playbook $details \
  openstack-create-infrastructure.yml

$ ansible-playbook $details \
  idr-00-preinstall.yml idr-01-install.yml \
  idr-03-postinstall.yml
```

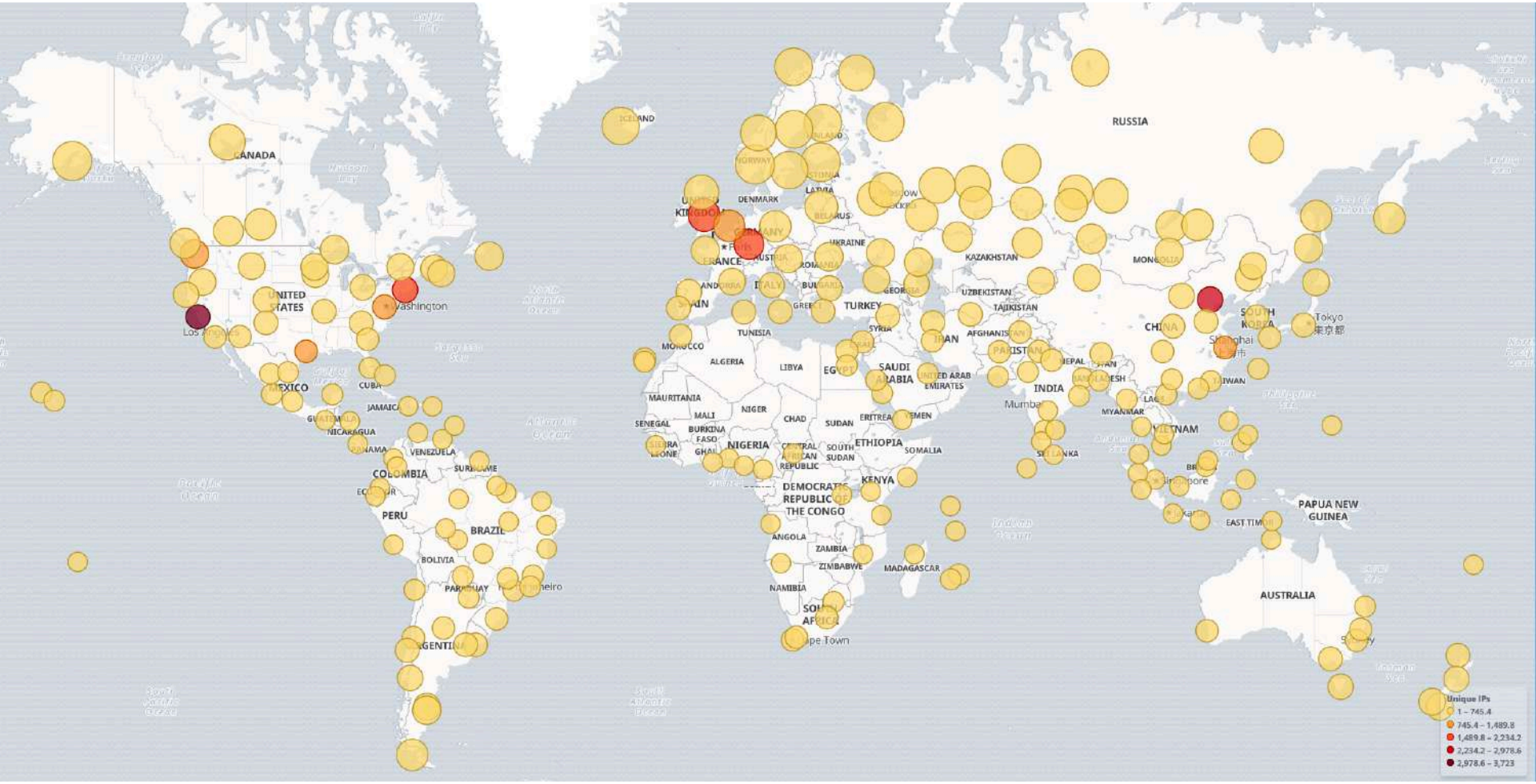
```
TASK [nginx-proxy : nginx | copy proxy ssl certificate] ************************
changed: [test-gateway]

TASK [nginx-proxy : nginx | copy proxy ssl certificate key] ********************
changed: [test-gateway]

TASK [nginx-proxy : nginx | start service] *************************************
changed: [test-gateway]

RUNNING HANDLER [nginx : restart nginx] ****************************************
changed: [test-gateway]

PLAY RECAP *********************************************************************
localhost                  : ok=25   changed=9    unreachable=0    failed=0
test-database              : ok=24   changed=10   unreachable=0    failed=0
test-gateway               : ok=33   changed=19   unreachable=0    failed=0
test-omero                 : ok=81   changed=49   unreachable=0    failed=0
```



ANSIBLE

https://github.com/IDR/deployment

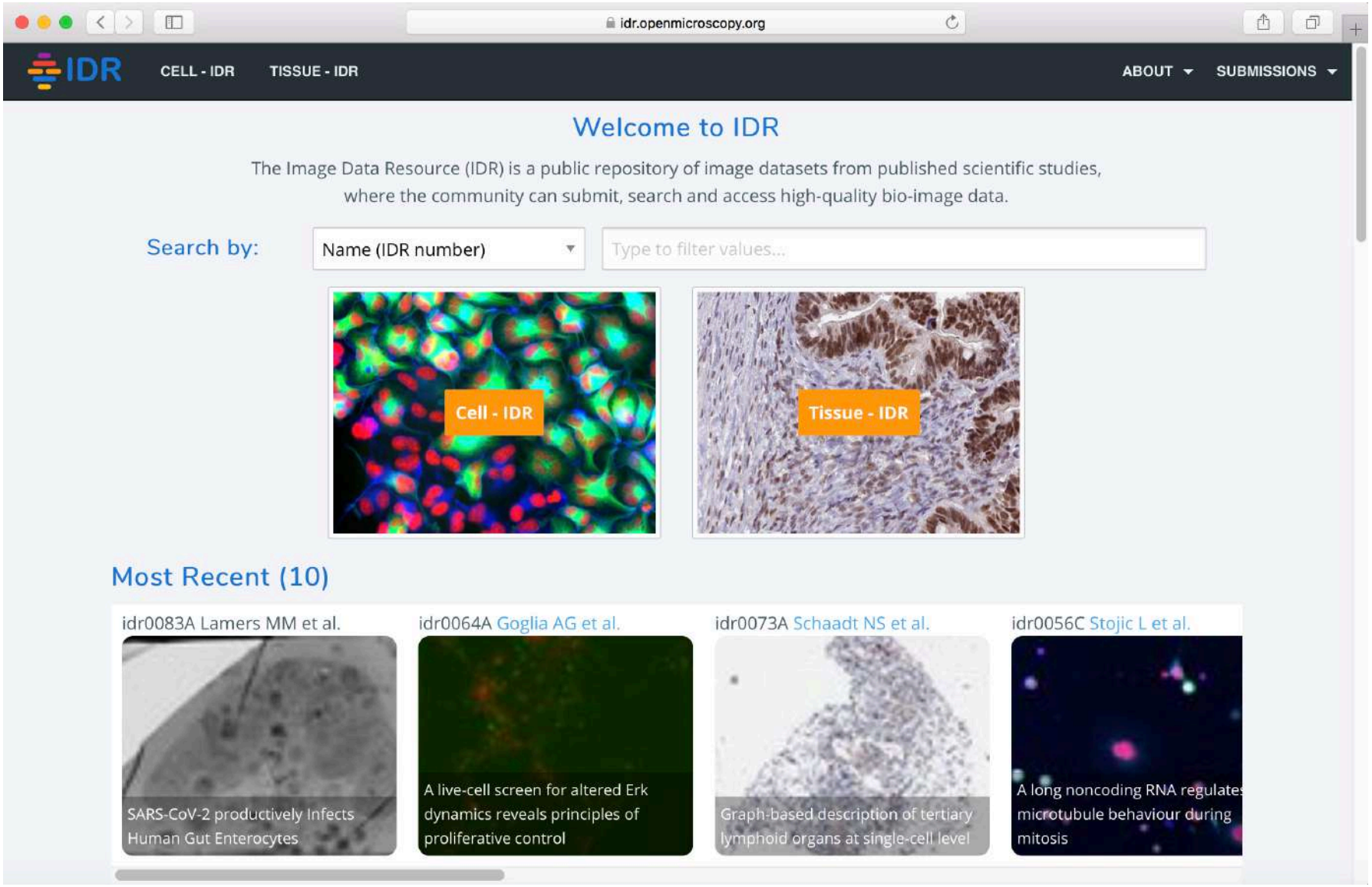
IDR usage



May 2019 - May 2020: 20M hits

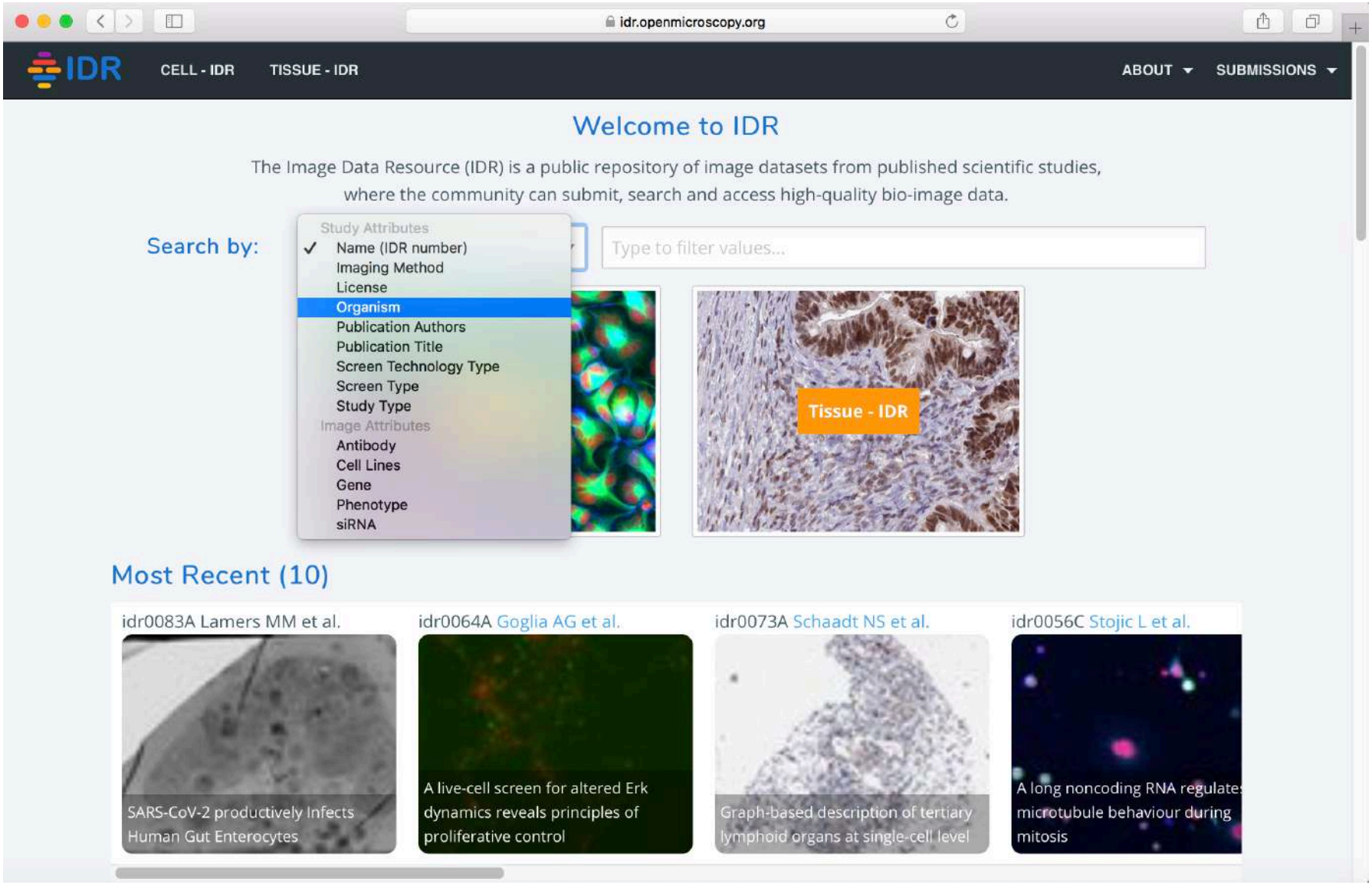
https://idr.openmicroscopy.org

IDR Homepage



https://idr.openmicroscopy.org

IDR Homepage



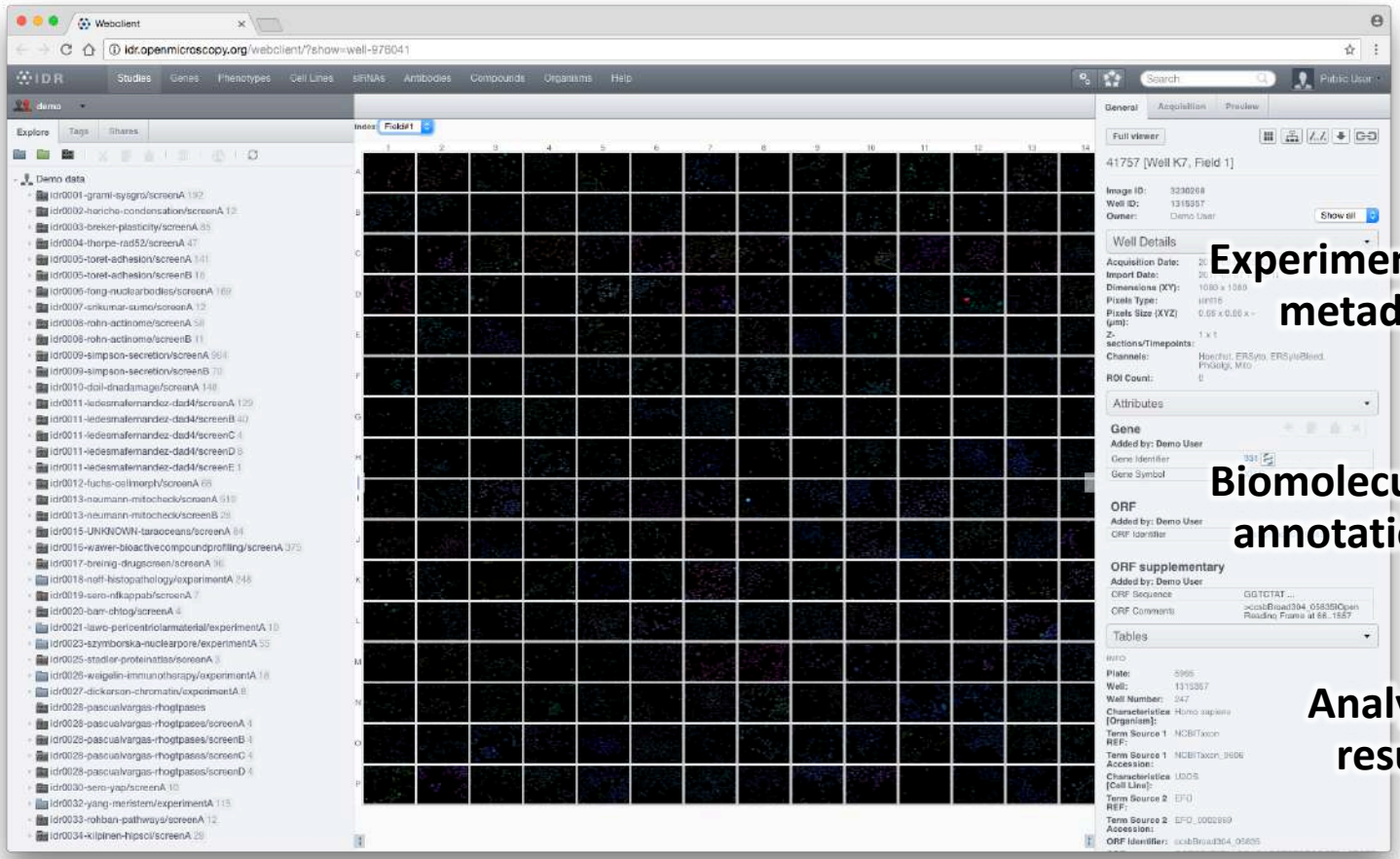
https://idr.openmicroscopy.org

Organism

- Homo sapiens
- Mus musculus
- Drosophila melanogaster
- Saccharomyces cerevisiae
- Schizosaccharomyces pombe
- Arabidopsis thaliana

- Gallus gallus
- Danio rerio
- Escherichia coli
- Apis mellifera
- Human adenovirus 2
- Severe acute respiratory syndrome coronavirus 2

Metadata



Experimental metadata

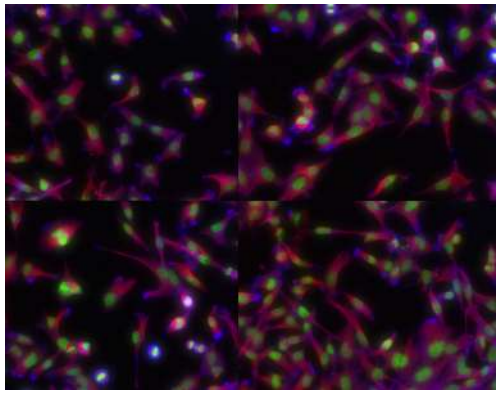
Biomolecular annotations

Analysis results

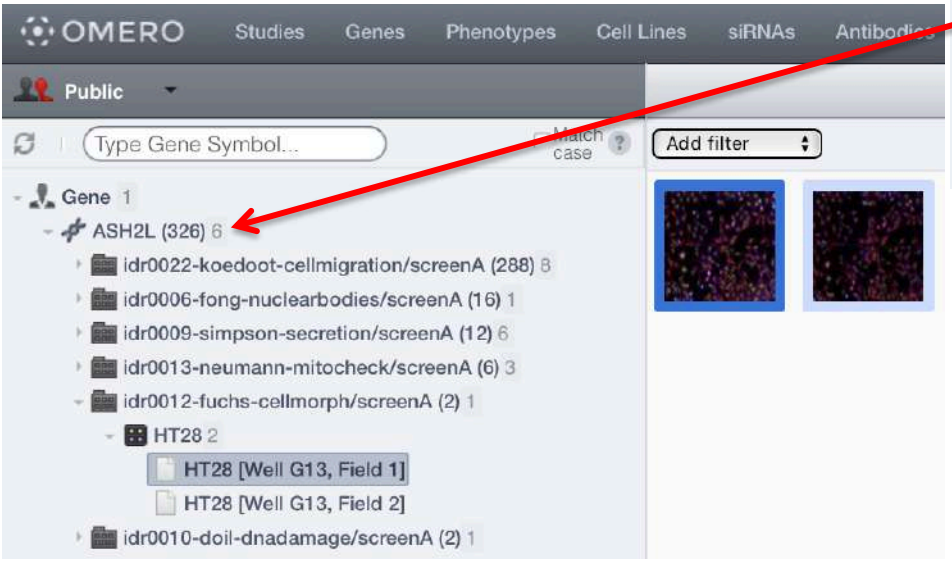
https://idr.openmicroscopy.org

Linked Metadata

IDR



idr0012



Attributes 8

Cell Lines

Added by: Public data

Cell Line	HeLa

Gene

Added by: Public data

Gene Identifier	9070
Gene Symbol	ASH2L

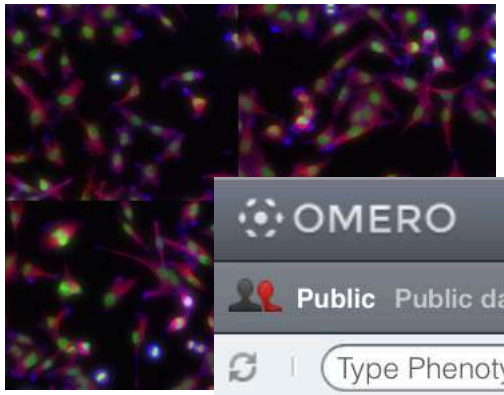
Phenotype

Added by: Public data

Phenotype	elongated cells
Phenotype Term Name	elongated cell phenotype
Phenotype Term Accession	CMPO_0000077

Linked Metadata

IDR



id

OMERO Studies Genes Phenotypes Cell Lines

Public Public data

Type Phenotype... Match case

Phenotype 1

CMPO_0000077 (20872) 8

- idr0028-pascualvargas-rhogtpases/screenD (3920) 4
- idr0028-pascualvargas-rhogtpases/screenA (3808) 4
- idr0028-pascualvargas-rhogtpases/screenC (3220) 4
- idr0033-rohban-pathways/screenA (2916) 12
- idr0028-pascualvargas-rhogtpases/screenB (2856) 4
- idr0001-graml-sysgro/screenA (2788) 122
- idr0008-rohn-actinome/screenB (1098) 10
- idr0012-fuchs-cellmorph/screenA (266) 55

Attributes 8

Cell Lines

Added by: Public data

Cell Line	HeLa
	9070
	ASH2L
	elongated cells
	elongated cell phenotype
	CMPO_0000077

OMERO Studies Genes Phen

Public

Type Gene Symbol...

Gene 1

ASH2L (326) 6

- idr0022-koedoot-cellmigration/screenA
- idr0006-fong-nuclearbodies/screenA (1
- idr0009-simpson-secretion/screenA (12
- idr0013-neumann-mitocheck/screenA (6) 3
- idr0012-fuchs-cellmorph/screenA (2) 1
  - HT28 2
    - HT28 [Well G13, Field 1]
    - HT28 [Well G13, Field 2]
- idr0010-doil-dnadamage/screenA (2) 1

Ontologies

IDR

Organism	NCBITaxon https://www.ncbi.nlm.nih.gov/taxonomy
Study Type	Experimental Factor Ontology (EFO) https://www.ebi.ac.uk/efo/
Screen Type (HCS)	Experimental Factor Ontology (EFO) https://www.ebi.ac.uk/efo/
Screen Technology Type (HCS)	Experimental Factor Ontology (EFO) https://www.ebi.ac.uk/efo/
Library Type (HCS)	Experimental Factor Ontology (EFO) https://www.ebi.ac.uk/efo/
Protocol	Experimental Factor Ontology (EFO) https://www.ebi.ac.uk/efo/
Imaging Method	Biological Imaging Methods Ontology (FBbi) https://www.ebi.ac.uk/ols/ontologies/fbbi
Phenotype	Cellular Microscopy Phenotype Ontology (CMPO) https://www.ebi.ac.uk/cmpo/
Clinical/Pathology	SNOMED CT https://bioportal.bioontology.org/ontologies/SNOMEDCT
Compounds	PubChem https://pubchem.ncbi.nlm.nih.gov/

Compounds

IDR

Compound supplementary

Added by: Public data

Compound Name	Azithromycin
Compound Identifier	CLOUD081
Compound DrugBank Identifier	DB00207
Compound PubChem CID	447043
Compound ChEMBL ID	CHEMBL529
ChEMBL Identifier URL	https://www.ebi.ac.uk/chembldb/target/inspect/CHEMBL529
Compound Concentration (microMolar)	6.7
Compound General Target Class	ribosome
Compound General activity	bacterial ribosome inhibitor
Compound Target Entrez IDs	23569;2737

Compound

Added by: Public data

Compound Name	propoxycaine

Compound supplementary

Added by: Public data

Compound Broad Identifier	BRD-K18250272-003-03-7
Compound Broad Identifier URL	https://pubchem.ncbi.nlm.nih.gov/compound/BRD-K18250272-003-03-7
Compound Broad Identifier Short	BRD-K18250272
Compound Group	BIO
Compound Source	Prestwick Chemical Inc.
Vendor Catalog Identifier	Prestw-1059
Compound SMILES	CCCOc1cc(N)ccc1C(=O)OCCN(CC)CC
PubChem InChIKey	CAJIGINSTLKQMM-UHFFFAOYSA-N
Dose	3.02 micromolar

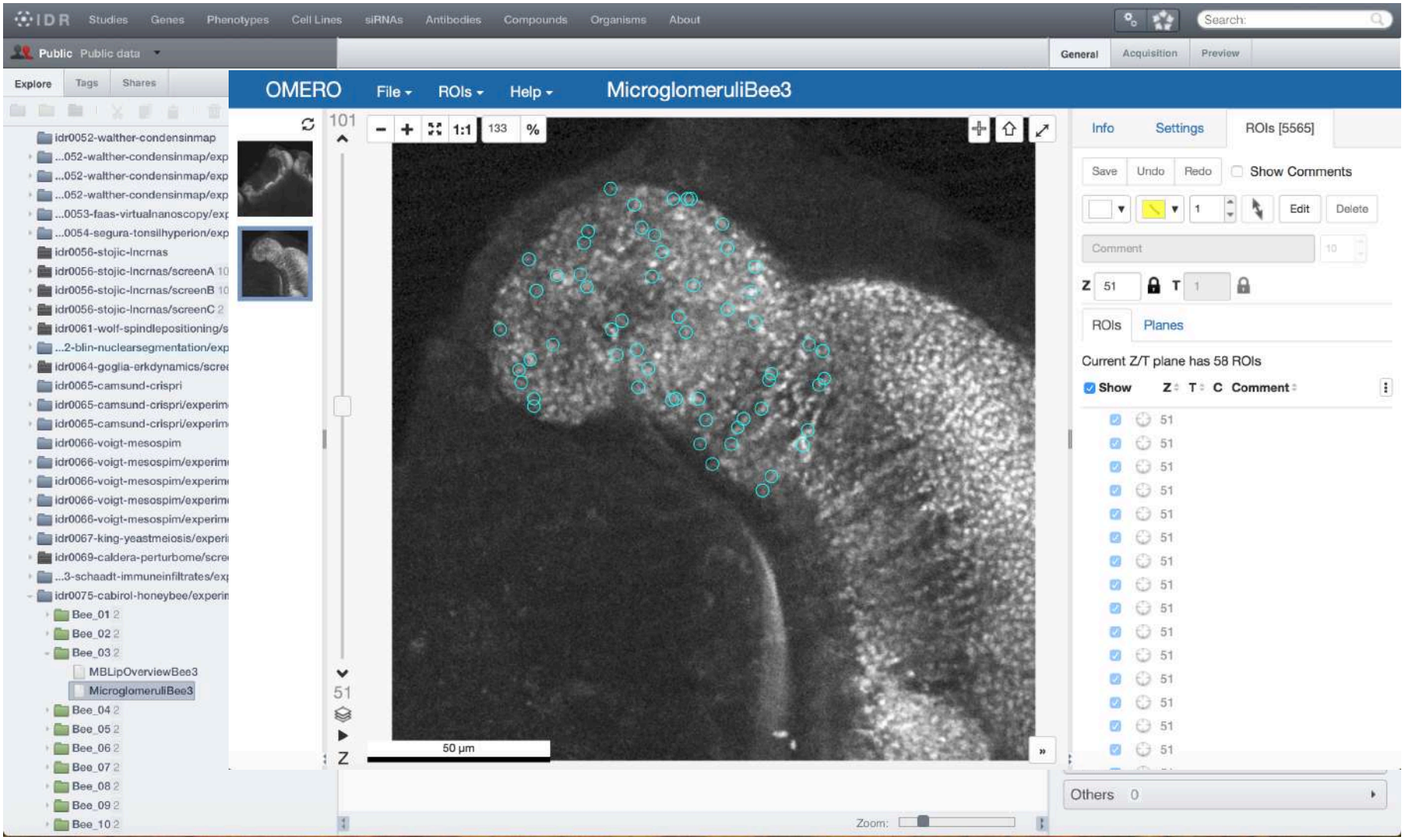
Tissue Data: Honey Bee

IDR

idr0075, Cabirol et al Scientific Reports DOI:10.1038/s41598-019-55974-2

Tissue Data: Honey Bee

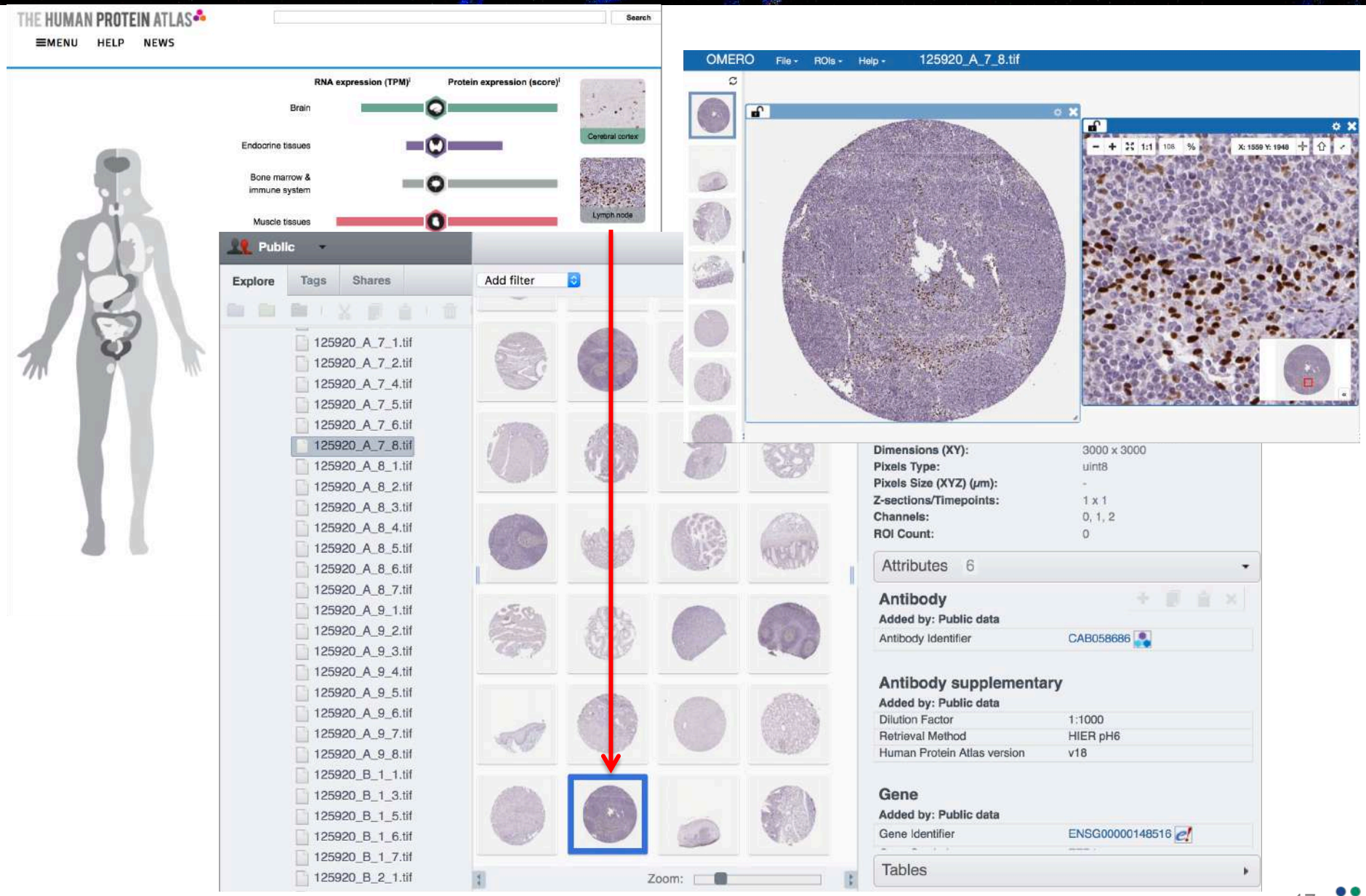
IDR



idr0075, Cabirol et al Scientific Reports DOI:10.1038/s41598-019-55974-2

Tissue Data: Human Protein Atlas

IDR



idr0043, https://www.proteinatlas.org

Tissue Data: Mouse Cell Fate

IDR

idr0044, McDole et al Cell DOI:10.1016/j.cell.2018.09.031

Cell Data: Transcription Profiling in Yeast

IDR

idr0047, Li et al Scientific Data DOI:10.1038/s41597-019-0106-6

Cell Data: Transcription Profiling in Yeast

IDR

idr.openmicroscopy.org

IDR Studies Genes Phenotypes Cell Lines siRNAs Antibodies Compounds Organisms About Search:

Public Public data

Explore Tags Shares

idr0047-neuert-yeastmrna/experimen

Exp1_rep1 63

Exp1_rep1_0min_im1.tif
Exp1_rep1_0min_im2.tif
Exp1_rep1_0min_im3.tif
Exp1_rep1_0min_im4.tif
Exp1_rep1_10min_im1.tif
Exp1_rep1_10min_im2.tif
Exp1_rep1_10min_im3.tif
Exp1_rep1_10min_im4.tif
Exp1_rep1_15min_im1.tif
Exp1_rep1_15min_im2.tif
Exp1_rep1_15min_im3.tif
Exp1_rep1_15min_im4.tif
Exp1_rep1_1min_im1.tif
Exp1_rep1_1min_im2.tif
Exp1_rep1_1min_im3.tif
Exp1_rep1_1min_im4.tif
Exp1_rep1_20min_im1.tif
Exp1_rep1_20min_im2.tif
Exp1_rep1_20min_im3.tif
Exp1_rep1_20min_im4.tif
Exp1_rep1_25min_im1.tif
Exp1_rep1_25min_im2.tif
Exp1_rep1_25min_im3.tif
Exp1_rep1_25min_im4.tif
Exp1_rep1_2min_im1.tif
Exp1_rep1_2min_im2.tif
Exp1_rep1_2min_im3.tif
Exp1_rep1_2min_im4.tif
Exp1_rep1_30min_im1.tif
Exp1_rep1_30min_im2.tif

idr.openmicroscopy.org

OMERO File ROIs Help Exp1_rep1_20min_im2.tif

25 1:1 34 % X: 497 Y: 41

13 Z

Zoom:

Info Settings ROIs [372]

Save Undo Redo Show Comments

1 Edit Delete

Comment

Z 13 T 1

Show Z T C Comment

Comments 0

idr0047, Li et al Scientific Data DOI:10.1038/s41597-019-0106-6

IDR and BioStudies

IDR

ebi.ac.uk

EMBL-EBI | Services | Research | Training | About us | EMBL-EBI

BioStudies.

Search BioStudies

Examples: hyperplasia, PMC516016

Home | Browse | Submit | Help | About BioStudies | Feedback | Login

BIOSTUDIES / STUDIES / S-BIAD1

Release Date: 22 November 2018

{JSON} <XML> .-PageTab.] FTP

A microscopy data set of discrete spatial and temporal single molecule RNA expression in single cells

Gregor Neuert [1]

[1] Vanderbilt University

Accession

S-BIAD1

Description

We report a comprehensive single cell dataset of spatial distributions of nuclear and cytoplasmic mRNA as a function of time. We measure mRNA expression with single molecule RNA fluorescent in-situ hybridization microscopy in the yeast model organism Saccharomyces cerevisiae in tens of thousands of cells. The resulting dataset contains the discrete number of mRNA molecules in the nucleus and the cytoplasm for sixteen time points, two genes and two stress conditions each in biological duplicates or triplicates. We present these datasets as population means, fraction of cells above basal mRNA expression (ON-cells), the variance normalized by the expression mean (Fano factor), marginal probability of nuclear and cytoplasmic mRNA, and the joint probability of nuclear and cytoplasmic RNA expression. The reuse potential are in three areas: (1) development of discrete single cell modelling approaches, (2) building predictive models to study fundamental processes in transcription regulation, (3) development of single cell image processing approaches not possible with continuous, non-spatial datasets of low temporal resolution.

Study type

FISH

Study Organism

Saccharomyces cerevisiae

Data files

Show 5 entries Search:

Name
20181016-ftp/Exp1_rep1/#1_Raw_Images/Exp1_
20181016-ftp/Exp1_rep1/#1_Raw_Images/Exp1_
20181016-ftp/Exp1_rep1/#1_Raw_Images/Exp1_
20181016-ftp/Exp1_rep1/#1_Raw_Images/Exp1_
20181016-ftp/Exp1_rep1/#1_Raw_Images/Exp1_

Showing 1 to 5 of 6,784 entries

Previous 1 2 3 4 5 ... 1357 Next

Attributes 1

Release Date

License

Copyright

Data Publisher

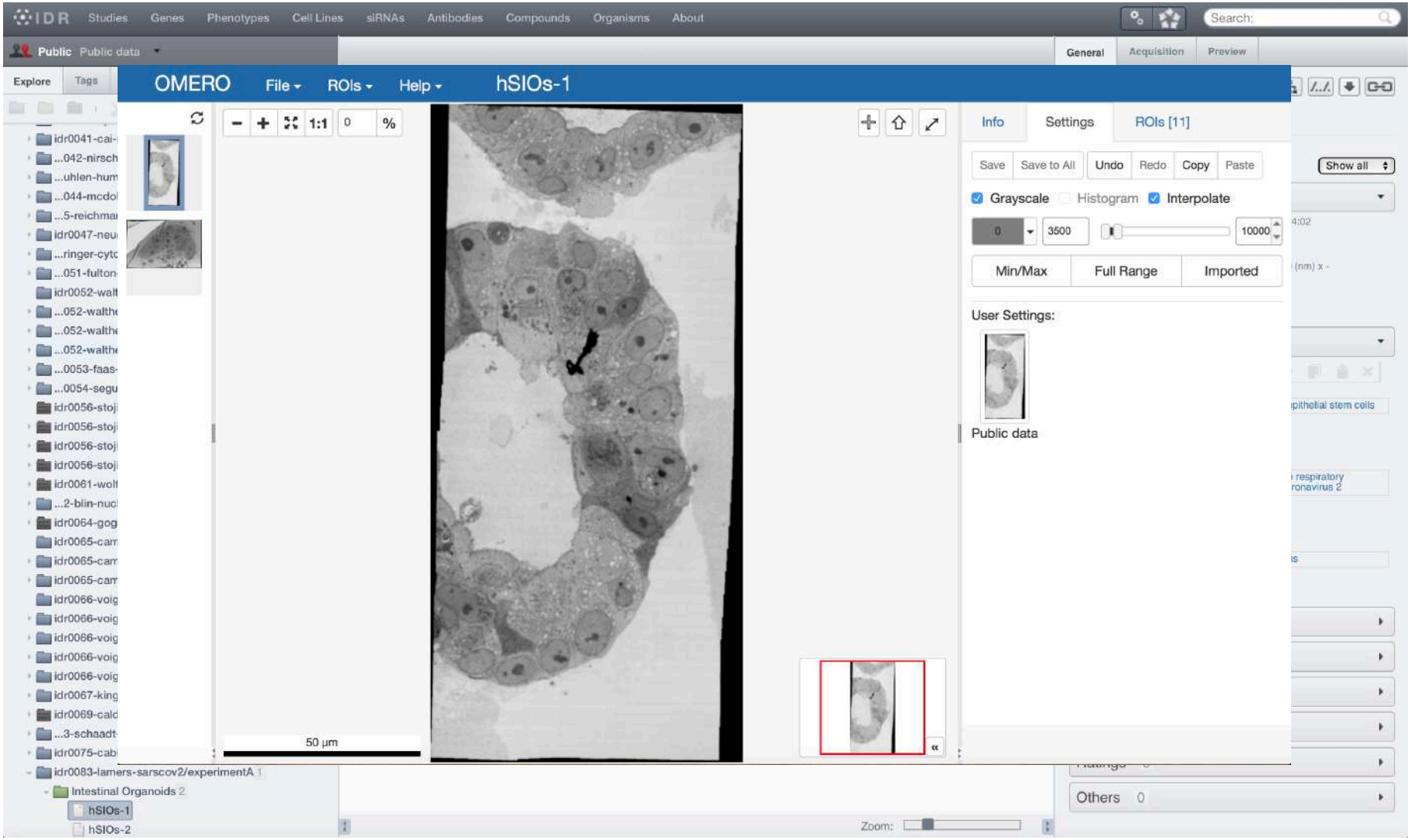
Data DOI

BioStudies Accession

Annotation File

Tissue Data: SARS-CoV-2

IDR



idr0083, Lamers et al Science DOI:10.1126/science.abc1669

Tissue Data: SARS-CoV-2

IDR

idr0083, Lamers et al Science DOI:10.1126/science.abc1669

IDR and EMPIAR

IDR

Attributes	1
Publication DOI	
Release Date	
License	
Copyright	
Data Publisher	
Data DOI	
EMPIAR Accession	
Annotation File	

EMPIAR

EMPIAR home | Deposition | REST API | FAQ | About EMPIAR | Policies | Feedback | Share

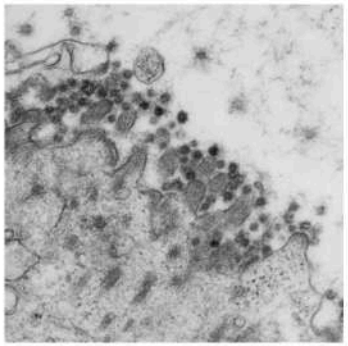
EMPIAR-10404

SARS-CoV-2 productively infects human gut enterocytes

Publication: SARS-CoV-2 productively infects human gut enterocytes
Lamers MM, Beumer J, van der Vaart J, Knoops K, Puschhof J, Breugem T, Ravelli RBG, van Schayck JP, Mykytyn AZ, Duimel HQ, van Donselaar E, Riesebosch S, Kuijpers HJH, Schipper D, van de Wetering WJ, de Graaf M, Koopmans M, Cuppen E, Peters PJ, Haagmans B, Clevers H
Science
DOI: 10.1126/science.abc1669

Related IDR entry: idr0083
Deposited: 2020-04-30
Released: 2020-05-01
Last modified: 2020-05-01
Dataset size: 66.5 GB
Dataset DOI: 10.6019/EMPIAR-10404
Experimental metadata: Download xml

Contains:
stitched maps



Quick links
- EMDB
- PDBe
- BioImage Archive
- EMPIAR Quick tour
- Statistics
- Re-use case study
- EMPIAR@PDBj

EMPIAR citations

Cryo-EM structure of the potassium-chloride cotransporter KCC4 in lipid nanodiscs.
Reid MS, Kern DM, Brohawn SG. (2020)

Development of basic building blocks

IDR and the press!

IDR

The Sun | SPORT | MONEY | TECH | TRAVEL | MOTORS | DEAR DEIDRE | PUZZLES

Mirror | MORE

UN-M
coron
multi

Lucy Jones,
4 May 2020,

ow deadly bug

SARS-CoV-2 virus, showing the deadly

SCIENCE

The team, led by Dundee's Professor of Quantitative Cell Biology, Jason Swedlow, built the Image Data Resource (IDR), a publication system for very large, complex images recorded using the world's most powerful microscopes and imaging systems.

Onl
home

To share the unique images with scientists around the world, the researchers built a new database, called the Image Data Resource (IDR).

Professor Jason Swedlow, who led the project, said: "We're excited to publish these important new datasets in IDR, where they can be seen by researchers around the world, who can also scan the images and view the SARS-CoV-2 virus up close on their computer."

more tha

Image of coronavirus Sars-CoV-2 (dark circles) at the edge of an intestinal cell

Credit: medicalxpres

the human gut

Open cloud & open analytics

IDR

jupyter ARP2-3Complex Last Checkpoint: 09/08/2017 (autosaved)

Control Panel | Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | OMERO Python 2

Markdown

Query IDR for Phenotypes

In [4]:
```
session = create_http_session(idr_url.value)
organism = org_sel.value
idr_base_url = idr_url.value

[query_genes_dataframe, screen_to_phenotype_dictionary] = get_phenotypes_for_genelist(idr_base_url, session, go_gene_li
display(HTML(query_genes_dataframe.to_html( escape=False)))
```

```
[==============================================================] 100.0% ...Iterating through gene list
```

	Entrez	Ensembl	Key	Value	PhenotypeName	PhenotypeAccession	ScreenIds
ARPC2	[10109]	[ENSG00000163466]	GeneName	ARPC2	[decreased cell numbers, geometric cell phenot...	[CMPO_0000274, CMPO_0000299, CMPO_0000276, CMP..	[206, 206, 206, 206, 206, 206, 206, 206, 206, ...
ARPC1A	[10552]	[ENSG00000241685]	GeneName	ARPC1A	[strong decrease in rate of protein secretion]	[CMPO_0000319]	[251]
ARPC5L	[81873]	[ENSG00000136950]	GeneName	ARPC5L	[increased actin localised to the nucleus, elo...	[CMPO_0000261, CMPO_0000287, CMPO_0000288, CMP..	[206, 206, 206, 206, 206, 206, 206, 206, 206, ...
ACTR3C	[653857]	[ENSG00000106526]	EnsemblID	ENSG00000106526	[cell death phenotype, mitosis delayed phenoty...	[CMPO_0000305, CMPO_0000348, CMPO_0000307, CMP..	[1101, 1101, 1101, 1101, 1101, 1101, 1101]
ACTR3B	[57180]	[ENSG00000133627]	GeneName	ACTR3B	[binuclear cell phenotype]	[CMPO_0000213]	[1101]
HEL-68	[10552]	[ENSG00000241685]	EnsemblID	ENSG00000241685	[strong decrease in rate of protein secretion]	[CMPO_0000319]	[251]
ARPC4	[10093]	[ENSG00000241553]	GeneName	ARPC4	[mild decrease in rate of protein secretion]	[CMPO_0000318]	[251]
ACTR3	[10096]	[ENSG00000115091]	GeneName	ACTR3	[strong decrease in rate of protein secretion,...	[CMPO_0000319, CMPO_0000319]	[251, 803]

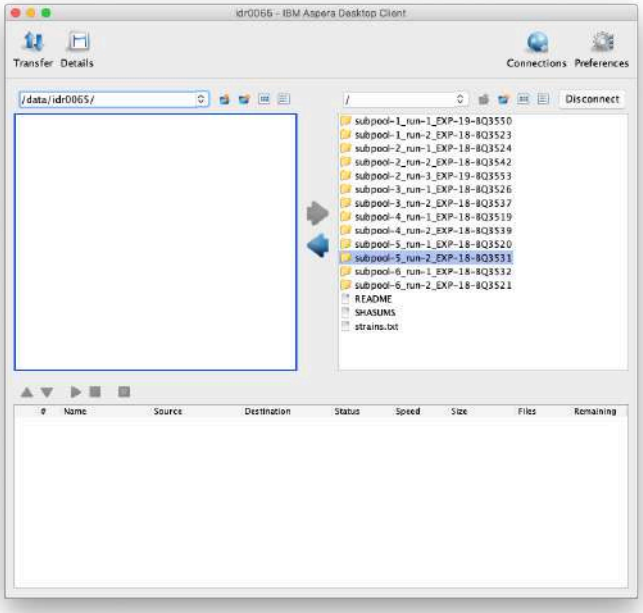
https://idr-analysis.openmicroscopy.org/

Redesigning data access for re-analysis

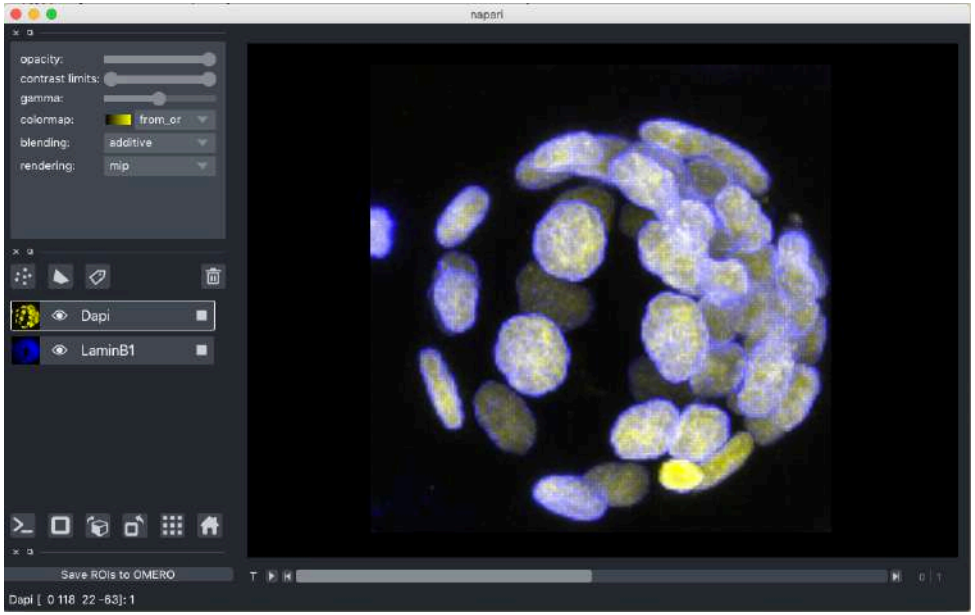
IDR

From file download

... to scalable remote data access



Aspera



Zarr + napari

See Next Generation File Formats talk

Day 2 of the 2020 OME Community meeting

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

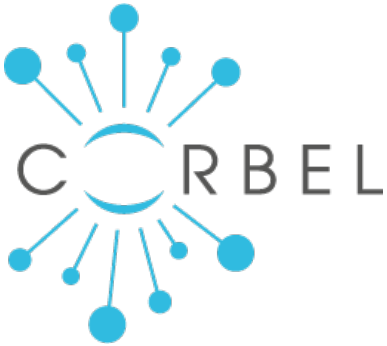
Thanks to the funders

IDR

Biotechnology and Biological Sciences Research Council

Chan Zuckerberg Initiative

EOSC-*Life*



GLOBAL BIOIMAGING
growing collaboration

MULTI MOT

Thanks to the IDR Team

IDR

OME

Jason Swedlow

Sebastien Besson

Simon Li

Dominik Lindner

Frances Wong

Josh Moore

Jean-Marie Burel

Petr Walczysko

Will Moore

David Gault

Mark Carroll

Past members

Eleanor Willams
Simone Leo
Gabry Rustici
Ola Tarkowska
Balaji Ramalingam
Riad Gozim
Rafael Carazo-Salas
Balint Antal
Anatole Chessel
Simon Jupp
Tony Burdett

EMBL-EBI

Alvis Brazma

Ugis Sarkans