

The Image Data Resource (IDR): a scalable resource for FAIR biological imaging data

OME Team University of Dundee @openmicroscopy, @IDRNews

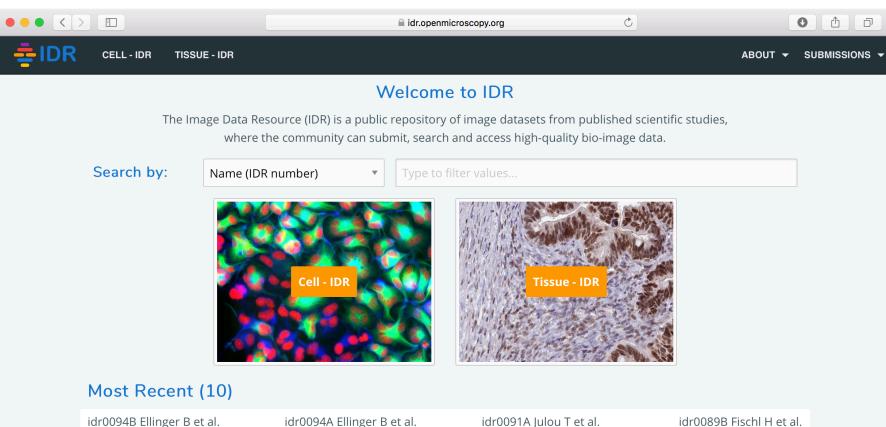




University of Dundee E



IDR Homepage

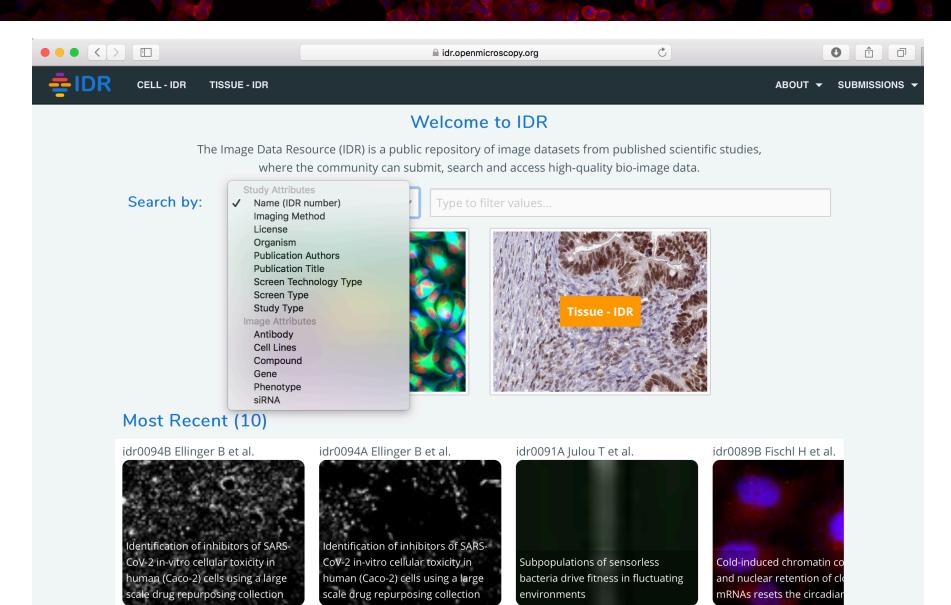


출IDR



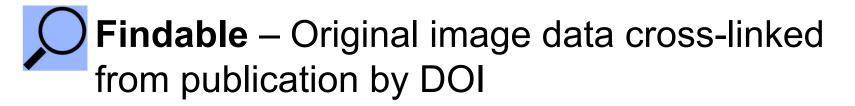
https://idr.openmicroscopy.org

IDR Homepage



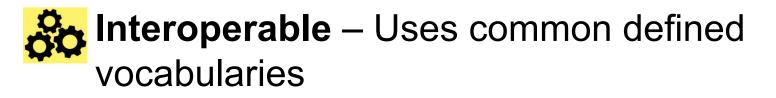
출IDR

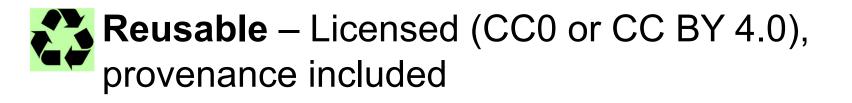
https://idr.openmicroscopy.org



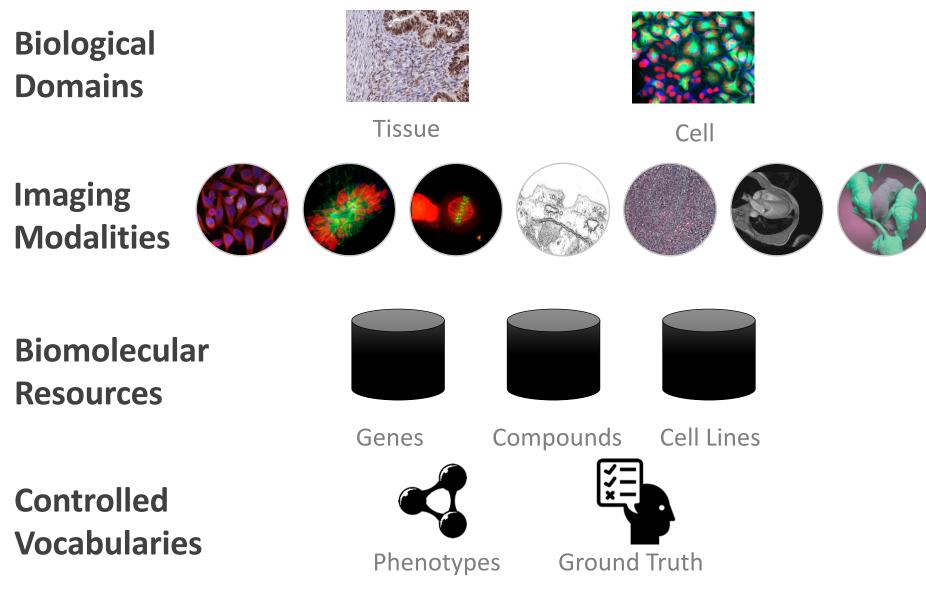


Accessible – Metadata retrievable, open API





Reference datasets



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

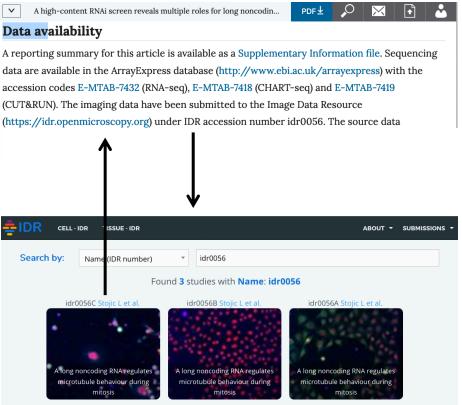
October 2020			
Raw data size	190 TB		
Image files	24 M		
5D Images	10 M		
2D Planes	71 M		
Organisms	13		
Genes	83 K		
Antibodies	8 K		
Compounds	38 K		

IDR studies and publications

>85 cross-published studies

SCIENTIFIC REPORTS eLIFE nature publishing group SCIENTIFIC DATA molecular systems biology nature PLOS ONE Cell Development Science For advances in developmental biology and stem cells BMC Cell Biology AAAS

cross-referenced via accessions and DOIs



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

The IDR @ EMBL-EBI Embassy

Gene Product Targeting HCS



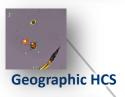
🙎 de

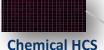
Explore

🗀 🛍 📾

🐧 Demo data

Genetic HCS







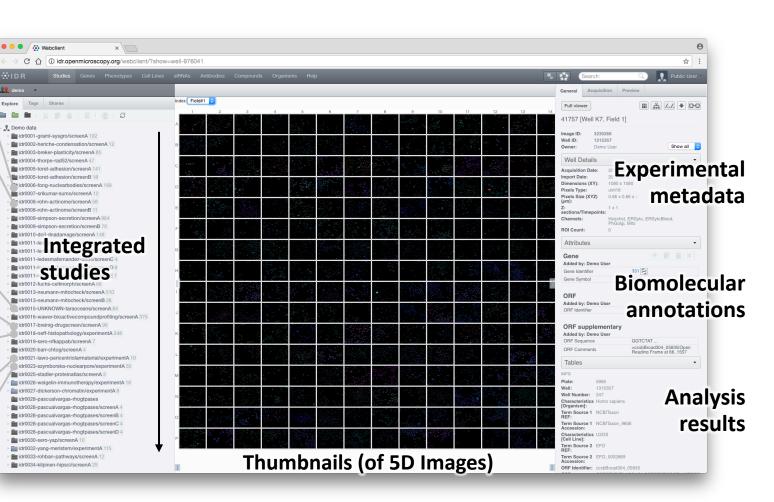
Histopathology





	. /
_	
Ο	











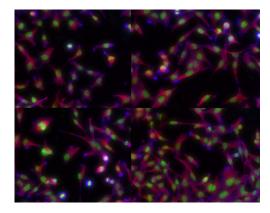


Cross-data browsing

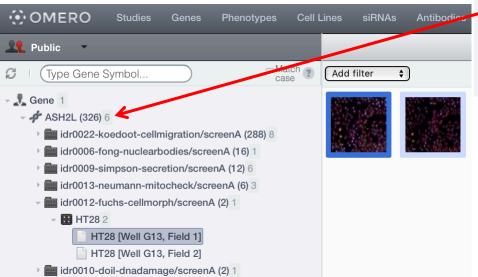
Cloud analysis

Download (local analysis)

Linked Metadata



idr0012



Attributes 8	•
Cell Lines	+ 🗊 💼 🗙
Added by: Public data	
Cell Line	HeLa
Gene	
Added by: Public data	
Gene Identifier	9070 8
Gene Symbol	ASH2L
Phenotype	ſ
Added by: Public data	
Phenotype	elongated cells
Phenotype Term Name	elongated cell phenotype
Phenotype Term Accession	CMPO_0000077

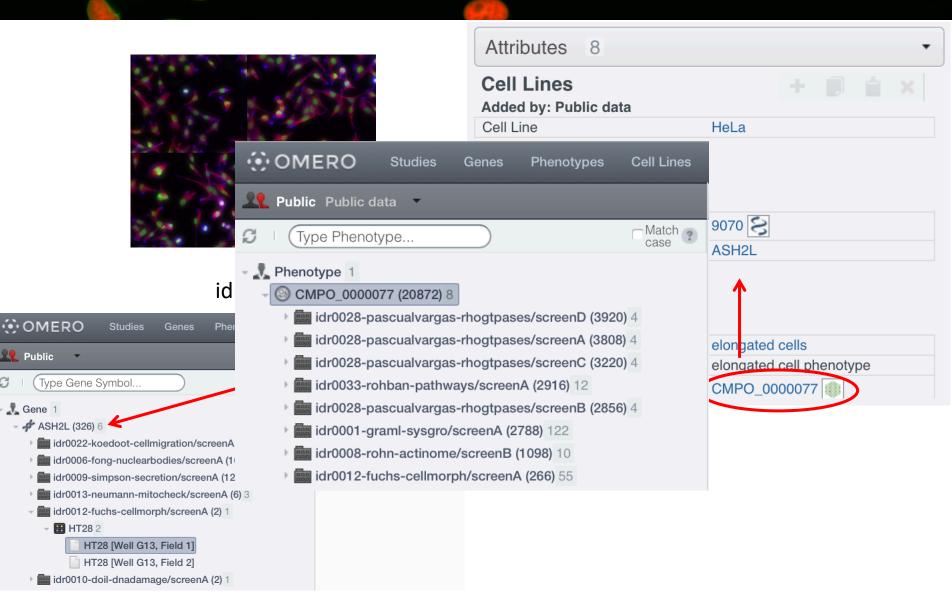
출IDR

Linked Metadata

Rublic

👢 Gene 1

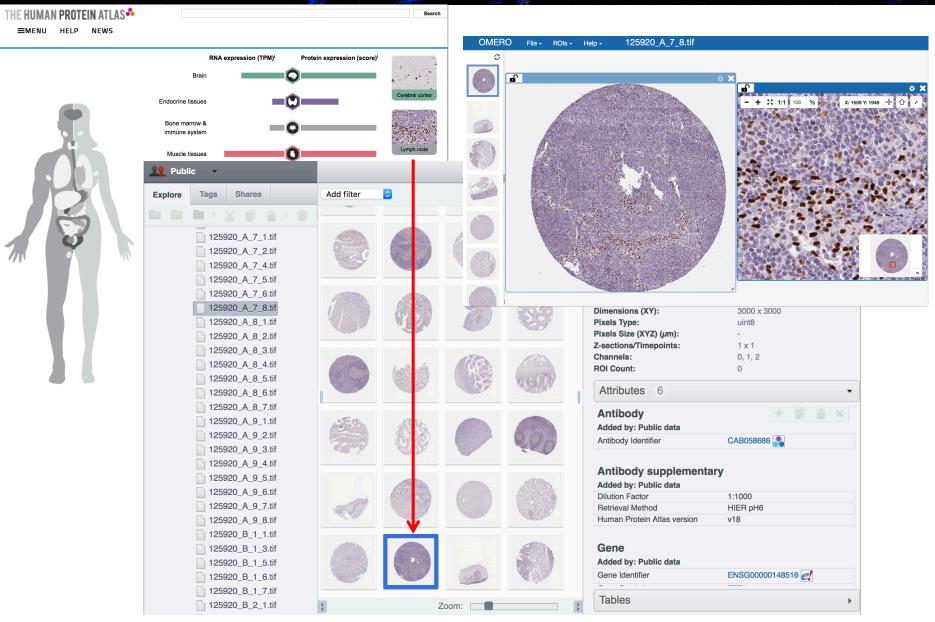
G



Ontologies

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 <u>https://bioportal.bioontology.org/ontologies/SNOMED</u> <u>CT</u>
Compounds	PubChem https://pubchem.ncbi.nlm.nih.gov/

Tissue Data: Human Protein Atlas



idr0043, Uhlen et al Science DOI:10.1126/science.1260419, https://www.proteinatlas.org

출IDR

IDR and **EMPIAR**

	EMF			
Attributes 1	EMPIAR home Depo	sition REST API FAQ About EMPIAR Polic	ies 🗣 Feedback < Share	
Publication DOI	EMPIAR-1	10404		Quick links
Release Date				𝔗 EMDB
License	SARS-CoV-2	productively infects human gu	it enterocytes	& PDBe
A	Publication:	SARS-CoV-2 productively infects human gut	Contains:	& Biolmage Archive
Copyright Data Publisher		enterocytes Lamers MM (b, Beumer J (b, van der Vaart J (b,	stitched maps	A EMPIAR Quick tour
Data DOI		Knoops K (16), Puschhof J, Breugem T, Ravelli RBG (16) , van Schayck JP (16), Mykytyn AZ, Duimel HQ, van Donselaar E, Riesebosch S, Kuijpers HJH,		<u></u> <u> </u>
EMPIAR Accession		Schipper D, van de Wetering WJ, de Graaf M, Koopmans M (), Cuppen E (), Peters PJ, Haagmans B (), Clevers H ()		 Re-use case study EMPIAR@PDBj
Annotation File	Related IDR entry: Deposited: Released:	Badgmans B B Clevers H B Science DOH: 10, 1126/science.abc1669 1000000000000000000000000000000000000		EMPIAR citations Cryo-EM structure of the potassium-
	Last modified: Dataset size: Dataset DOI: Experimental metadata:	2020-05-01 66.5 GB 10.6019/EMPIAR-10404 Download xml		chloride cotransporter KCC4 in lipid nanodiscs. Reid MS, Kern DM, Brohawn SG. (2020) Development of basic building blocks

Development of basic building blocks

Open cloud and open analytics

Jupyter ARP2-3Complex Last Checkpoint: 09/08/2017 (autosaved)	Control Panel	Logout
File Edit View Insert Cell Kernel Widgets Help Not Trusted	OMERO Py	rthon 2 O
H K		

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_list) display(HTML(query genes dataframe.to html(escape=False)))

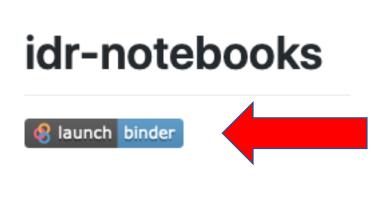
[======] 100.0% ...Iterating through gene list

	Er	ntrez	Ensembl	Key	Value	PhenotypeName	PhenotypeAccession	ScreenIds
AF	RPC2 [10	0109]	[ENSG00000163466]	GeneName	ARPC2	[decreased cell numbers, geometric cell phenot	[CMPO_0000274, CMPO_0000299, CMPO_0000276, CMP	[206, 206, 206, 206, 206, 206, 206, 206,
ARP	PC1A [10)552]	[ENSG00000241685]	GeneName	ARPC1A	[strong decrease in rate of protein secretion]	[CMPO_0000319]	[251]
ARF	PC5L [81	1873]	[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,
АСТ	R3C [653	3857]	[ENSG00000106526]	EnsembIID	ENSG00000106526	[cell death phenotype, mitosis delayed phenoty	[CMPO_0000305, CMPO_0000348, CMPO_0000307, CMP	[1101, 1101, 1101, 1101, 1101, 1101, 1101]
ACT	R3B [57	7180]	[ENSG00000133627]	GeneName	ACTR3B	[binuclear cell phenotype]	[CMPO_0000213]	[1101]
HE	L-68 [10)552]	[ENSG00000241685]	EnsemblID	ENSG00000241685	[strong decrease in rate of protein secretion]	[CMPO_0000319]	[251]
AR	RPC4 [10	0093]	[ENSG00000241553]	GeneName	ARPC4	[mild decrease in rate of protein secretion]	[CMPO_0000318]	[251]
AC	CTR3 [10	0096]	[ENSG00000115091]	GeneName	ACTR3	[strong decrease in rate of protein secretion,	[CMPO_0000319, CMPO_0000319]	[251, 803]

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

Open cloud and open analytics

https://github.com/IDR/idr-notebooks can be launched in MyBinder, just click the badge



Ç Jupyter
Files Running Clusters
Select items to perform actions on them.
D D binder
□ □ docker
□ □ includes
CalculateSharpness.ipynb
Figure_1_Sampling_of_Phenotypes.ipynb
GeneNetwork.ipynb
GenesToPhenotypes.ipynb
Getting_Started.ipynb

https://github.com/ome/omero-guide-fiji https://github.com/ome/omero-guide-cellprofiler https://github.com/ome/omero-guide-python

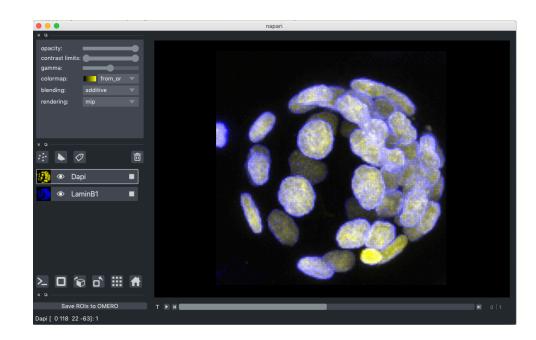
contain IDR-specific notebooks too !

Redesigning data access for re-analysis

From file download

idr0065 - IBM Aspera Desktop Client Â. M Transfer Details Preference Connections /data/idr0065/ 🗊 🤞 💕 📖 🚍 🗘 🍵 💕 📖 😑 🛛 Disconnect subpool-1 run-1 EXP-19-BO3550 subpool-1 run-2 EXP-18-BQ3523 subpool-2 run-1 EXP-18-BQ3524 subpool-2_run-2_EXP-18-BQ3542 subpool-2_run-3_EXP-19-BQ3553 ubpool-3_run-1_EXP-18-BQ3526 subpool-3_run-2_EXP-18-BQ3537 subpool-4 run-1 EXP-18-BO3519 subpool-4 run-2 EXP-18-BQ3539 subpool-5 run-1 EXP-18-BQ3520 subpool-5_run-2_EXP-18-BQ3531 subpool-6_run-1_EXP-18-BQ3532 subpool-6_run-2_EXP-18-BQ3521 README SHASUMS strains txt **b** III × A W Aspera

... to scalable remote data access



Zarr + napari

As seen in sharing big image data in the

cloud workshop tomorrow

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



Publicly available, curated studies...

submitted by the community...

in a searchable, scalable platform...

that links metadata ...

and enables reanalysis ...

that can be deployed by others.

Thanks to the funders





Thanks to the IDR Team





Sebastien Besson



Simon Li



Dominik Lindner

Frances Wong

OME









Past members

Josh Moore

Jean-Marie **Burel**

Petr Walczysko

Will Moore

David Gault



Alvis Brazma



Sarkans

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