



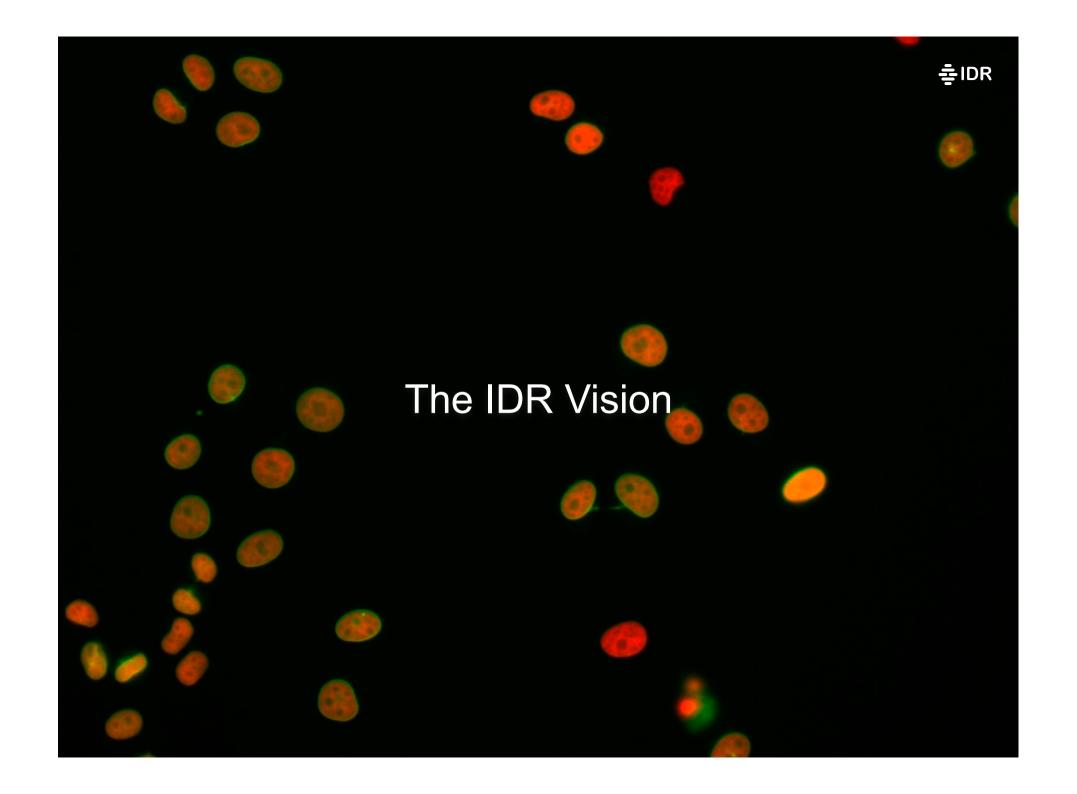




Talk Outline

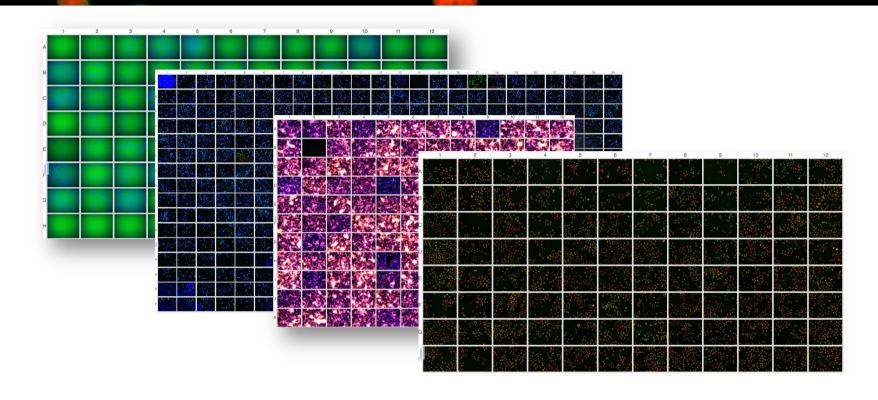


- o The IDR Vision
- What's New
 - Pre-publication submissions and data growth
 - Improved annotation search with OMERO.mapr
- Working with Linked Data
 - Annotations: "elongated cell phenotype", etc.
 - Image and feature data: Jupyter, workflows, etc.
- Next steps



Data **Sharing** in OMERO





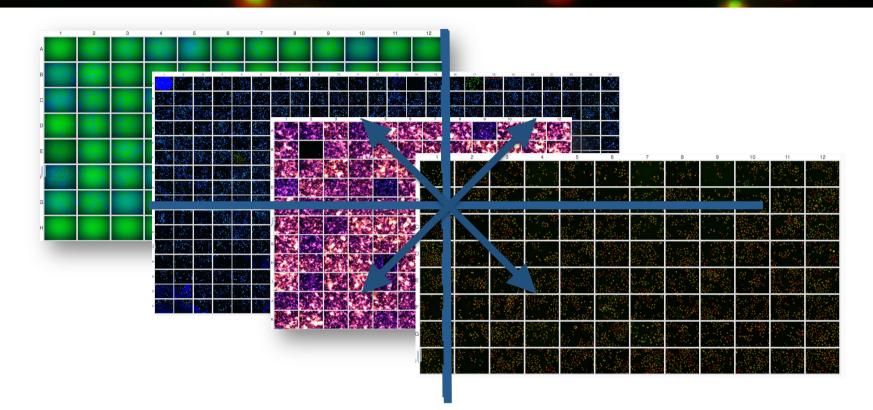






Data Integration in OMERO

















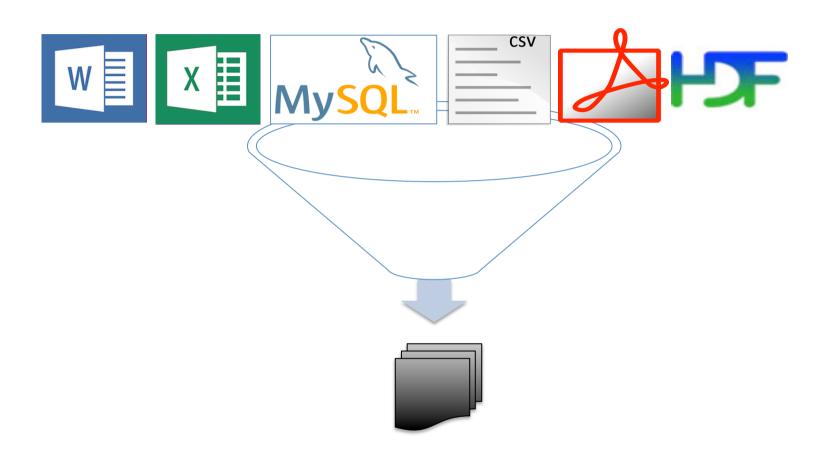






Submitted metadata





In the style of: MAGE-TAB isatab or, Cellular Phenotype Database

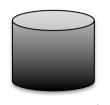


Reference images



Biomolecular Resources





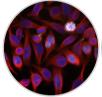


C

Compounds

Cell Lines

Imaging Domains















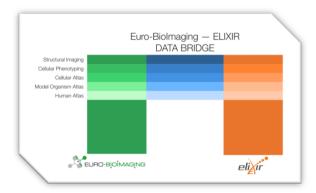
Controlled Vocabularies



Phenotypes



Ground Truth





The IDR @ EBI







Genetic HCS



Geographic HCS



Chemical HCS



Histopathology

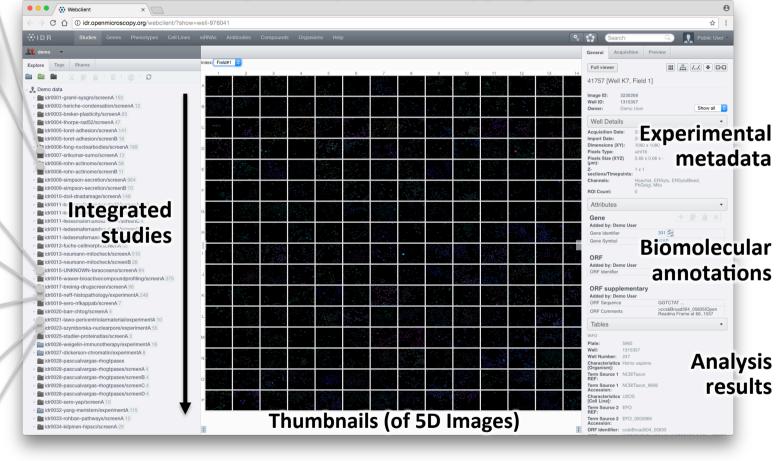


3D-Sim



Super-resolution

https://idr.openmicroscopy.org





Cross-data browsing



Cloud analysis



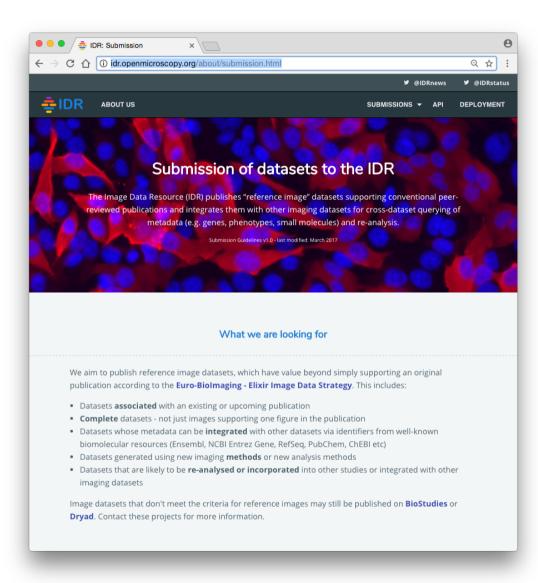
Download (local analysis)





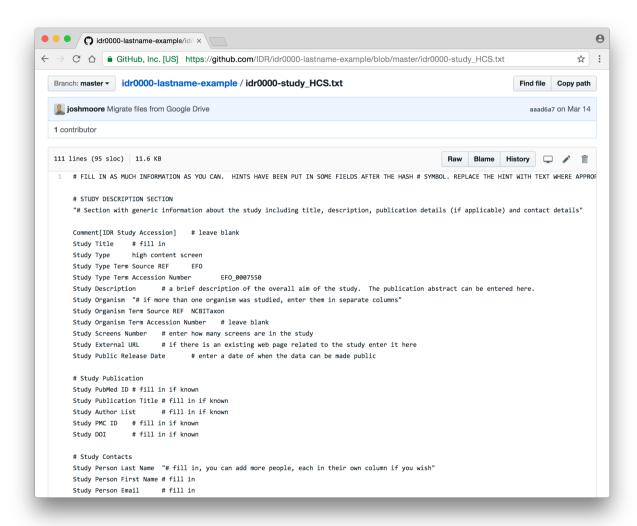
Submissions





Submissions





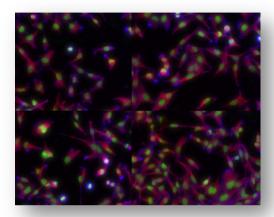
IDR Statistics



	May 2016	May 2017
Screens or Projects	21	40
Raw Image Data	38 Tb	42 Tb
Images	2.3 M	2.7 M
Experiments	800K	1 million
Genes	19K	19K
Species	Human, Plankton, Drosophila, M. musculus S. cerevisiae, S. pombe	+ Arabidopsis
Annotation Search Categories	Genes, Phenotypes, Compounds, siRNA	+ Antibodies, ORFs, Cell Lines
Author Phenotypes	212 (89% mapped)	286 (88% mapped)
Consortia	MitoCheck, Tara Oceans, Sysgro	BBBC, HipSci, HPA
Journals	Dev. Cell, JCB, MBoC, Nature, PNAS	BMC Cell Biology, Cell Systems, eLife, Scientific Data

Improved searching: OMERO.mapr





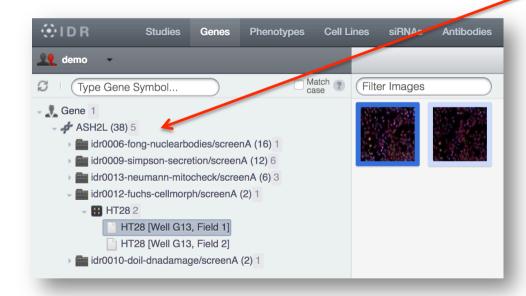
idr0012

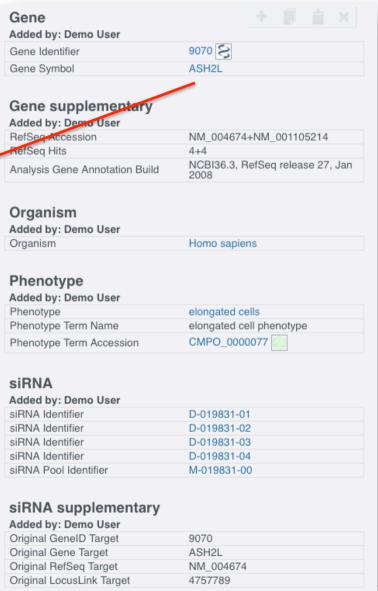


9070 S ASH2L		
ASH2L		
NM_004674+NM_001105214		
4+4		
NCBI36.3, RefSeq release 27, Jan 2008		
II.		
Homo sapiens		
elongated cells		
elongated cell phenotype		
CMPO_0000077		
D-019831-01		
D-019831-02		
D-019831-03		
D-019831-04		
M-019831-00		
9070		
9070 ASH2L NM 004674		

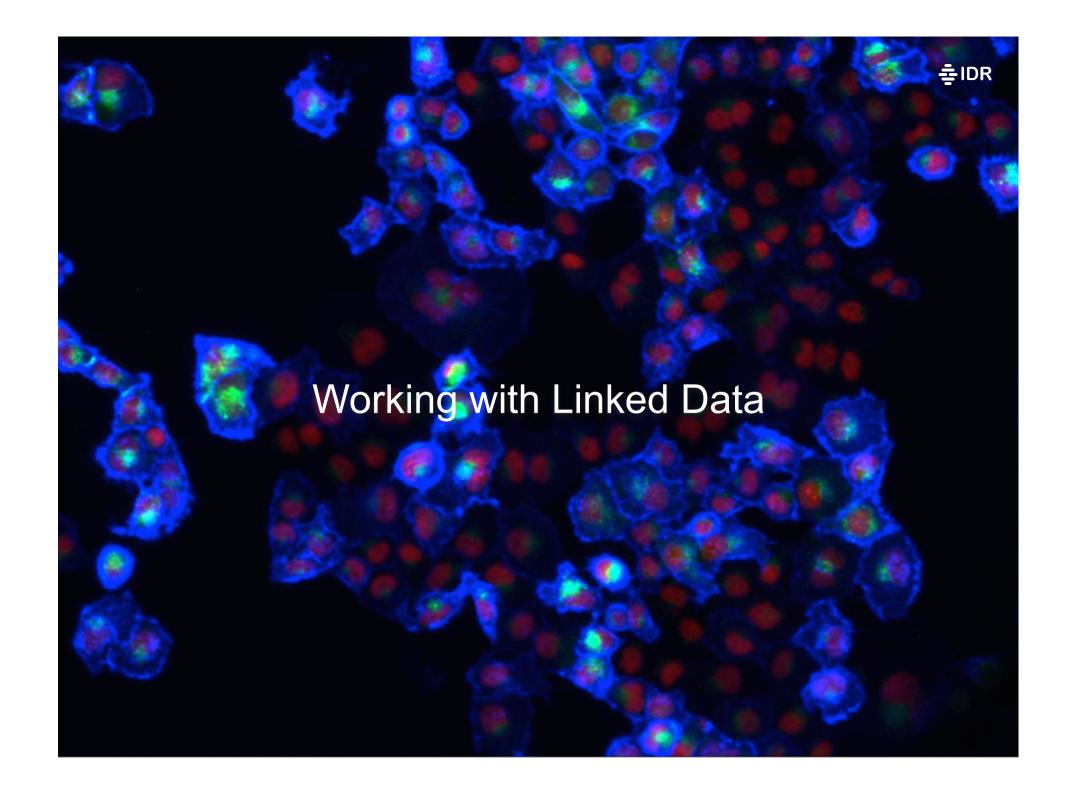
Improved searching: OMERO.mapr



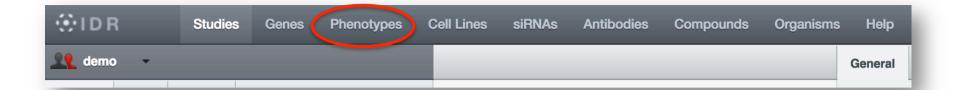






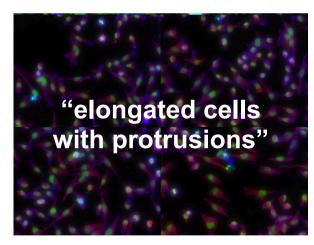


Many links between datasets

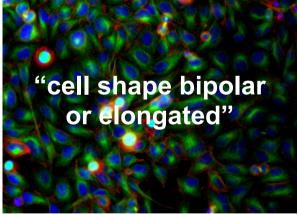


elongated cell phenotype

http://www.ebi.ac.uk/cmpo/CMPO_0000077



idr0012-A CellMorph, Fuchs et al 2010, human HeLa genome wide RNAi screen (image-1812396)

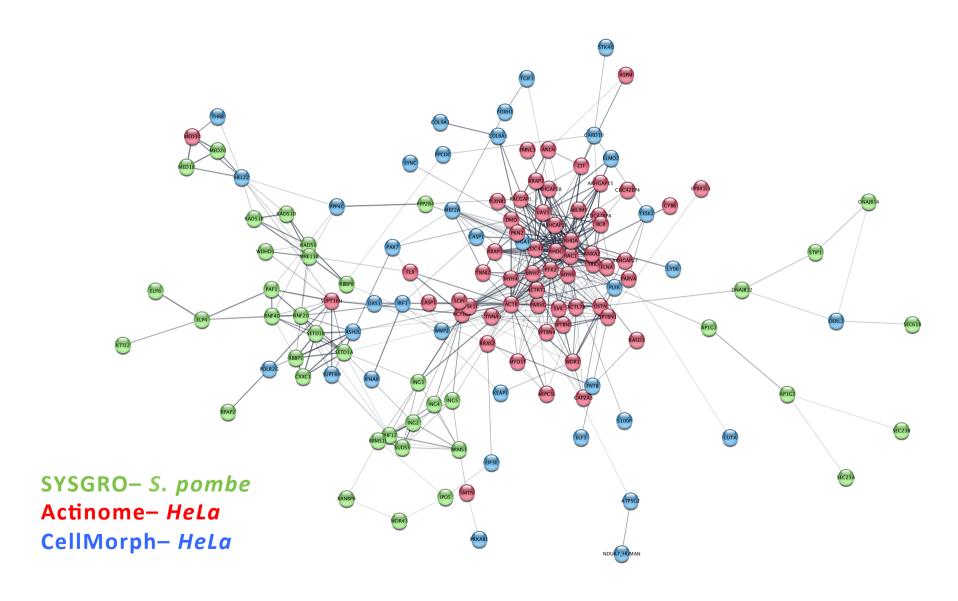


idr0008-B Actinome, Rohn et al 2011, human HeLa targeted actinome RNAi screen (image-109296)



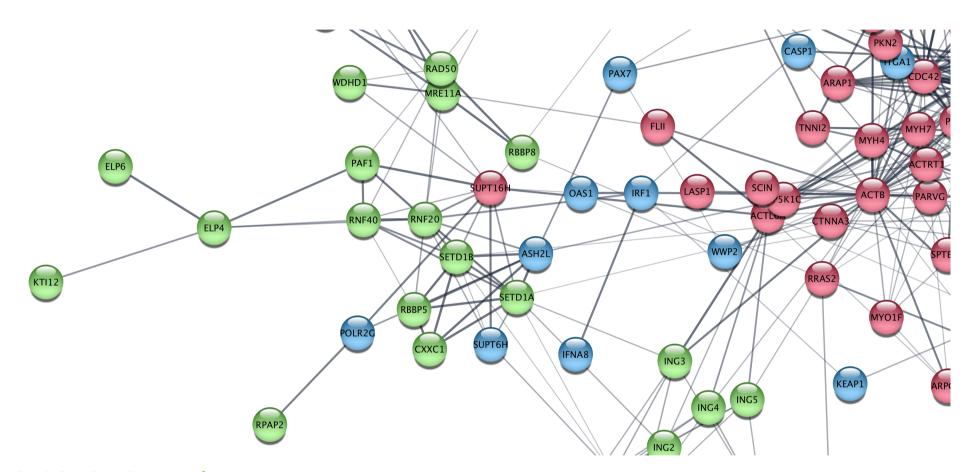
idr0001-A Sysgro, Graml et al 2014, S. pombe genome wide gene knockout screen (image-1233538)

CMPO phenotype: elongated cells





CMPO phenotype: elongated cells

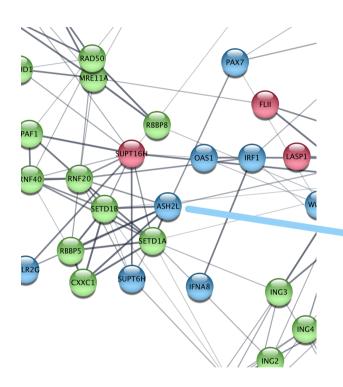


SYSGRO- S. pombe Actinome- HeLa CellMorph- HeLa



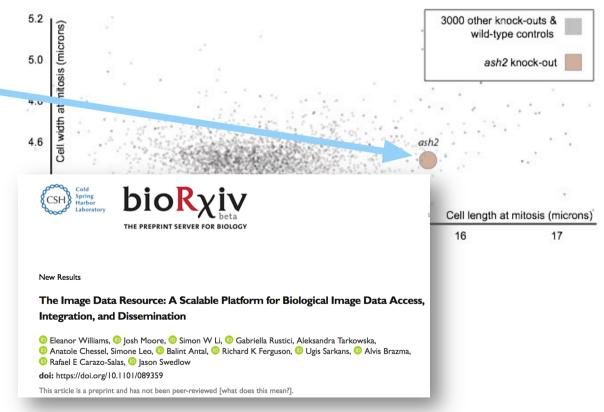
Retrospective analysis





SYSGRO- S. pombe Actinome- HeLa CellMorph- HeLa

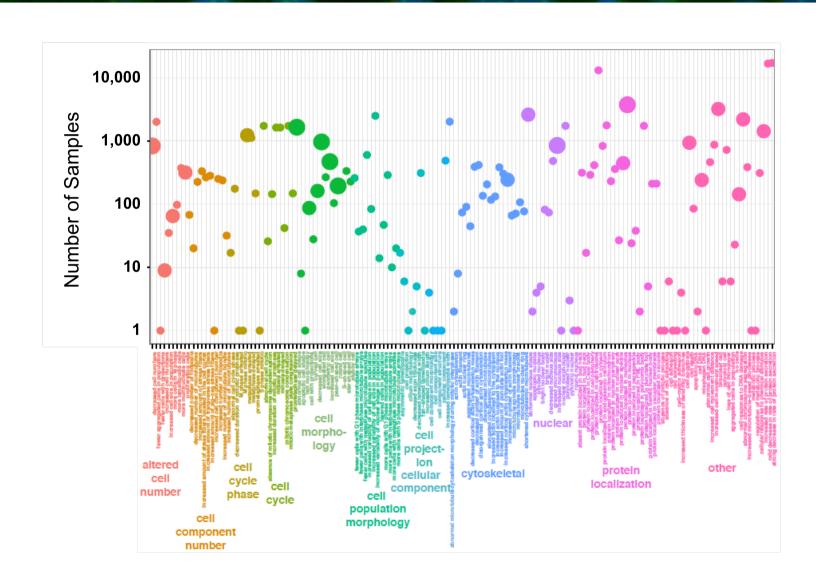
ash2 knock-out leads to abnormal cell elongation in S. pombe





Data exploration opportunities





Data exploration opportunities



Reproducing Fig. 1

This notebook exemplifies how to reproduce Figure 1 of the article. The annotations from all screens will be downloaded and parsed to build statistics on phenotypes, which will be displayed using bokeh.

Dependencies

```
In [9]: TOOLS = "pan, wheel_zoom, reset"

    Matplotlib

    NumPy

                                             phenotypes = figure(title = "Fig 1",

    Pandas

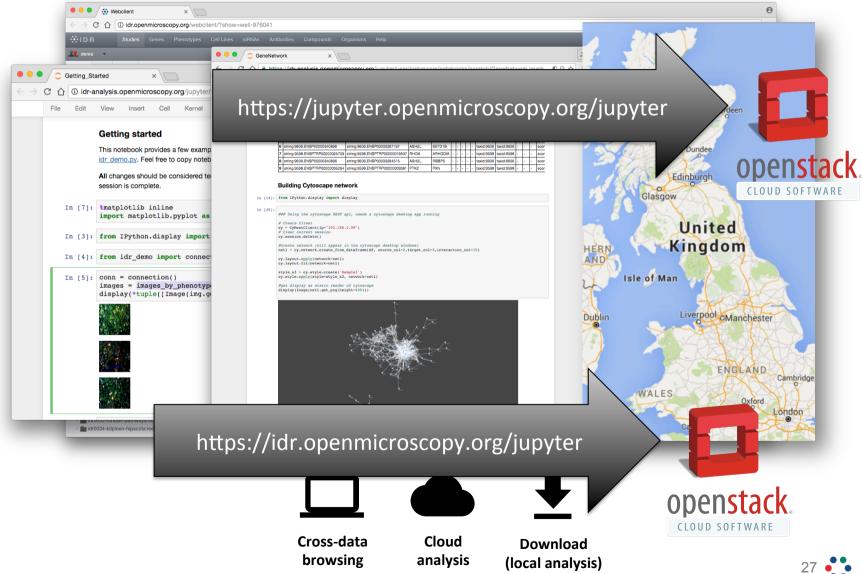
                                                                  tools=TOOLS,

    Bokeh

                                                                 y_axis_type="log",
                                                                  width=1200,
                                                                 toolbar location="above")
In [1]: import omero
         from idr import conn
                                             source = ColumnDataSource(
                                                 data=dict(
                                                     ph=[ph['FigureCmpoName'] for ph in phalls],
         import numpy as np
                                                     n=[ph['n'] for ph in phalls],
         import matplotlib.py
                                                     names=[ph['screens'] for ph in phalls],
         from pandas import D
         from pandas import c
         from pandas import r
         from bokeh.models im
         from bokeh.plotting
         from bokeh.plotting
         from bokeh.plotting
                                             label dat
         from bokeh.plotting
                                             cir = phe
         from bokeh.charts im
         from bokeh.models.fc
         from bokeh.models.fc
                                             hover =
         from bokeh.models.fc
         from bokeh.models.fc
         from bokeh.models.fc
         from bokeh.models im
         from bokeh.models im
         import bokeh.palette
         output_notebook()
         *matplotlib inline
                                             phenotype
          BokehJS 0.12.5 succe
                                             phenotype
                                             phenotype
                                             show(pher
                                                                     10^1
                                                                     10/0
```

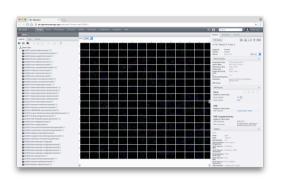
A platform, not a silo



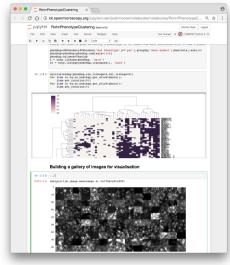


Data services vision





Browse & Search



Compute

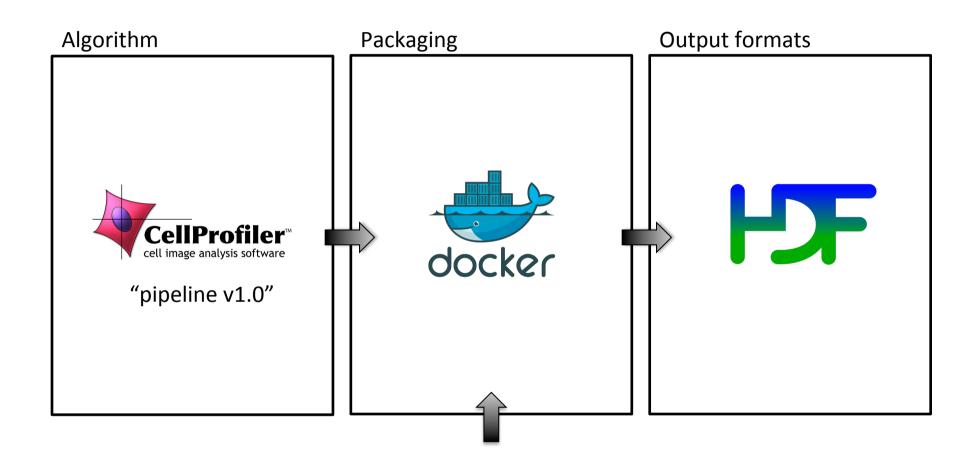
Image Tools Resource (ITR)

Workflows













Current features



	Language	Datasets	Size	Format
QC: sharpness	Python	All	120MB	Annotation
WND-CHARM (v1)	u	5	2TB	HDF5
WND-CHARM (v2)	u	2	50GB	HDF5
Image Similarity	R	N/A	N/A	RDS
Nuclei	CellProfiler	2	50GB	ROIs
			•••	•••

Training Denominating Neproducionity Ne-analys	Training	Benchmarking	Reproducibility	Re-analysis
--	----------	--------------	-----------------	-------------

In collaboration with











Reference images



Biomolecular Resources



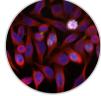


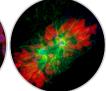


Compounds

Cell Lines

Imaging Domains















Controlled Vocabularies



Phenotypes



32

Reference data



Biomolecular Resources



Genes



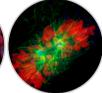


Compounds

Cell Lines

Imaging Domains















Controlled Vocabularies



Phenotypes

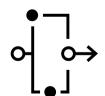


Ground Truth

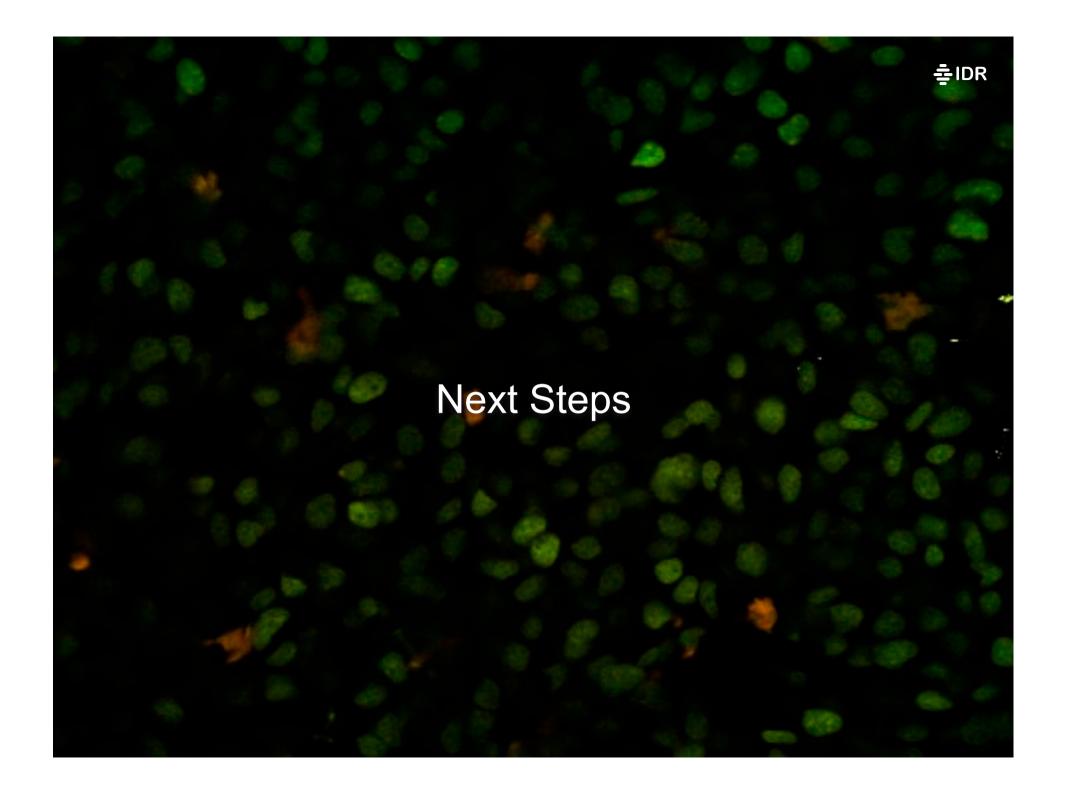




Features



Workflows

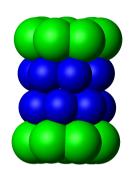


Next Steps – datasets and annotations



https://zh.wikipedia.org/ wiki/File:

1G0U_subunits_sideview.png License: GFDL



Further analyses e.g. proteasome genes



Develop phenotype descriptions for collective vs single phenotypes

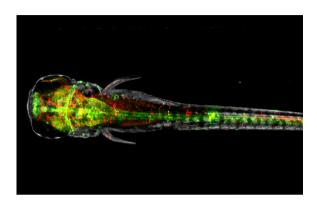
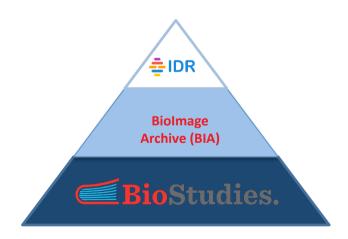


Image: Zeiss Microscopy License: https:// creativecommons.org/ licenses/by-nc-nd/2.0/

Cross-scale imaging, e.g. LSFM



Scale up submission process



Next Steps – driving involvement



- Submit your reference datasets
- Submit your reference workflows
- o Let us know what your definition of "reference" is
- o Finally, get in touch early!

Thanks to the Funders















Thanks to the IDR Team



















Jason **Swedlow**

Josh Moore

Simon Li

Eleanor Williams

Simone Leo

Ola **Tarkowska**

Gabry Rustici







Ugis Sarkans



Simon Jupp



Tony **Burdett**









Balint Antal



Anatole Chessel







