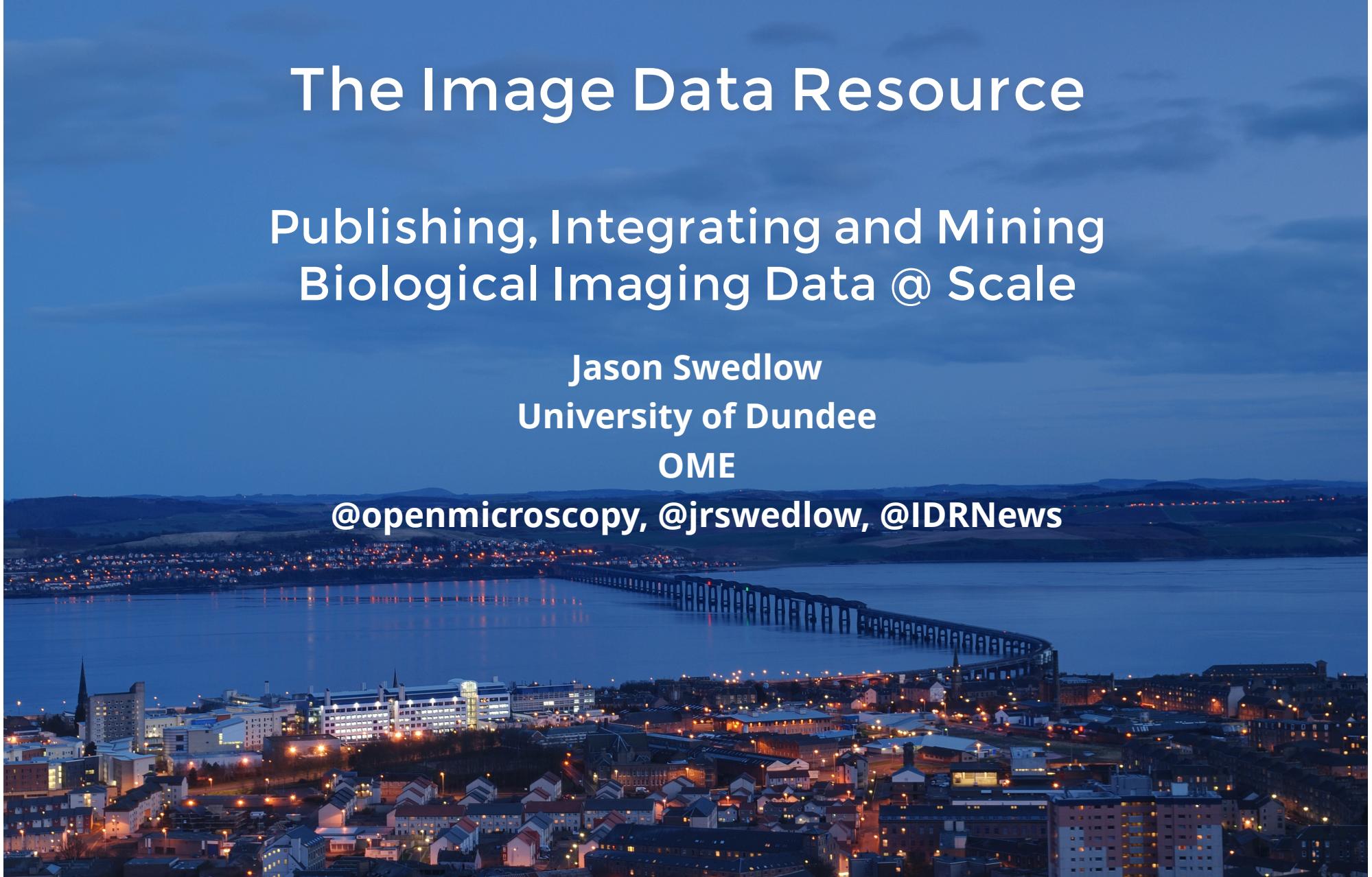


The Image Data Resource

Publishing, Integrating and Mining
Biological Imaging Data @ Scale

Jason Swedlow
University of Dundee
OME

@openmicroscopy, @jrswedlow, @IDRNews

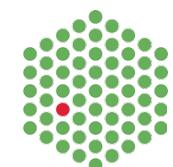


University of
BRISTOL

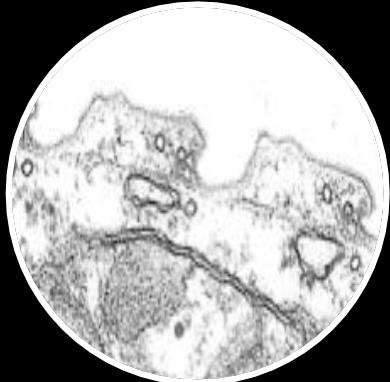


University
of Dundee

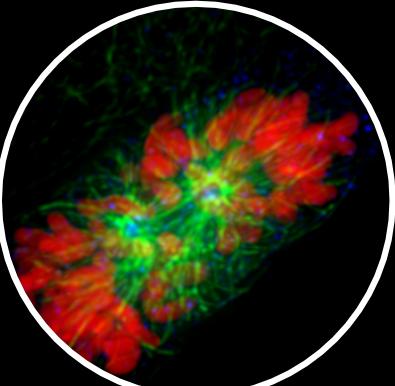
EMBL-EBI



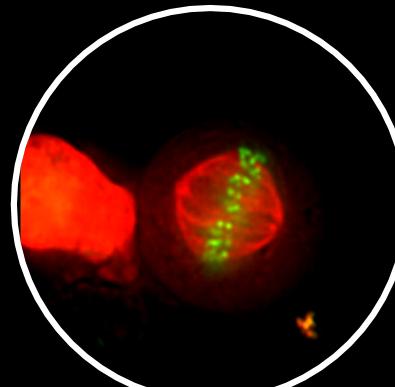
The Image Problem is Ubiquitous



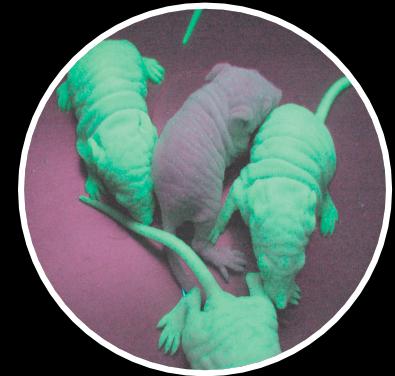
Organelles



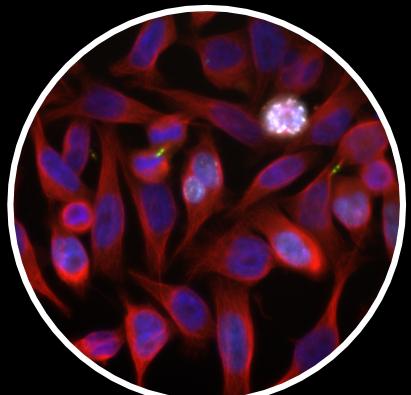
Cells



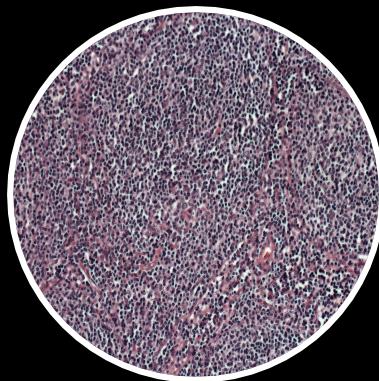
Dynamics



Physiology



Lead Discovery
Target Validation



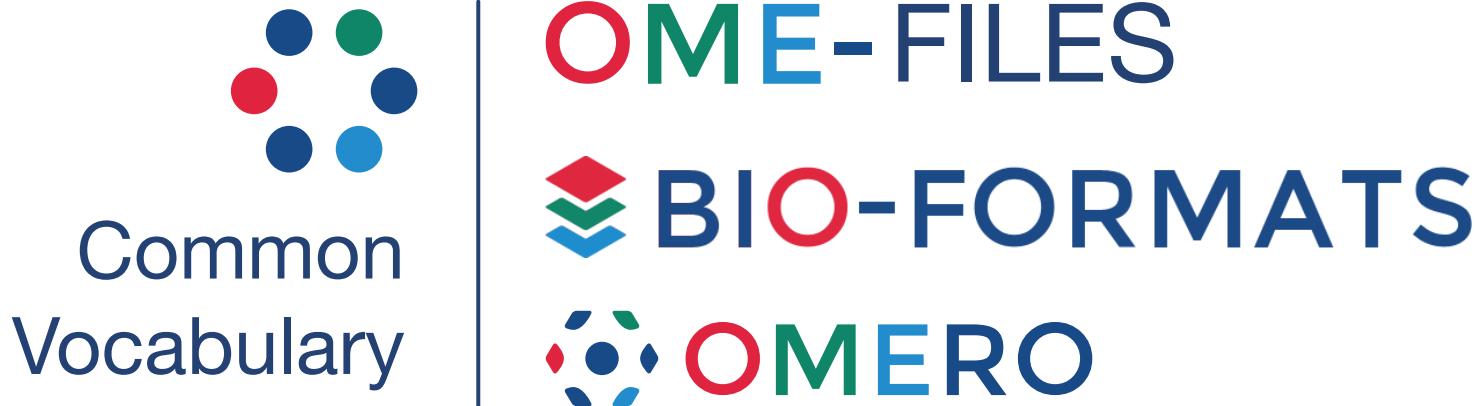
Pathology



In Vivo

A pretty picture?
A measurement?
A resource?

Open Tools for Data Access

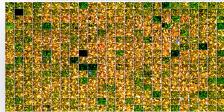


<http://openmicroscopy.org>

Image Data Resource Vision



Gene Product
Targeting HCS



Genetic HCS



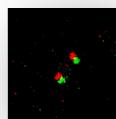
Geographic HCS



Chemical HCS



Histopathology



3D-SIM



Super-resolution

Integrated studies

Thumbnail (of 5D Images)

Experimental metadata

Biomolecular annotations

Analysis results

The screenshot shows a 5D image grid with 14 columns and 14 rows. Each cell in the grid contains a small thumbnail of a 5D image. To the right of the grid, there is a detailed panel for "Well Details" and "Attributes". The "Well Details" section includes fields for Image ID, Well ID, Owner, Acquisition Date, Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ), Z, sections/timepoints, Channels, ROI Count, and Attributes. The "Attributes" section lists Gene, ORF, and ORF supplementary information. The "Biomolecular annotations" section shows tables for ORF supplementary data. The "Analysis results" section displays tables for INFO, Plate, Well, Well Number, Characteristics, Organism, Term Source 1, Term Source 2, and Term Source 3.



Cross-data
browsing

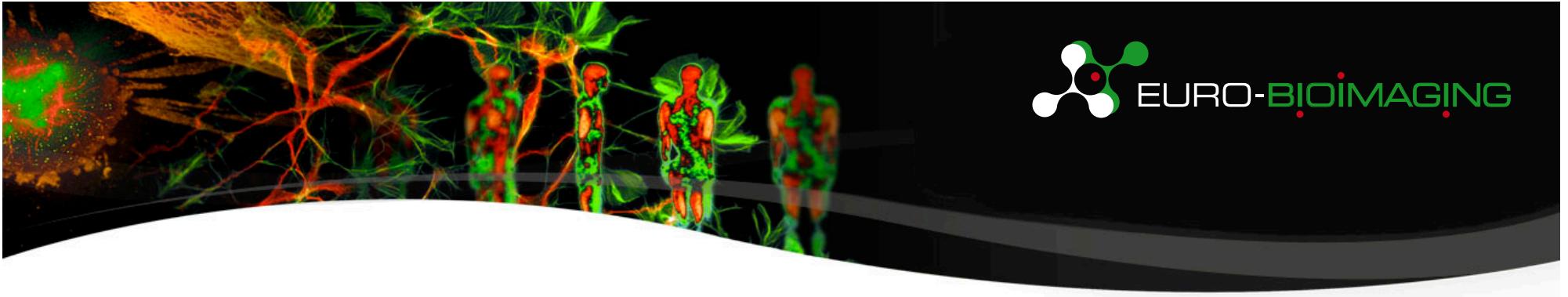


Cloud
analysis



Download
(local analysis)





Euro-BioImaging

an open access resource, removing
barriers and enabling world-class research

- **Access** to biological and medical imaging technologies in Europe
- **Support** from expert technical staff
- **Image data** repositories and analysis tools
- **Training** opportunities in imaging for everyone

How to Source Data



Simon Li
@crucifixkiss



Following

Another 20TB of #BigData for the
@openmicroscopy @BBSRC @emblebi image
repository



RETWEET

1

LIKES

3



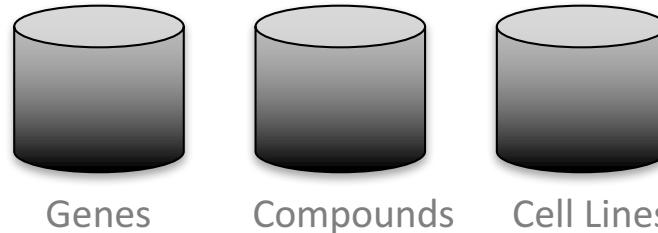
1:43 PM - 15 Sep 2015



Reference images



Biomolecular Resources

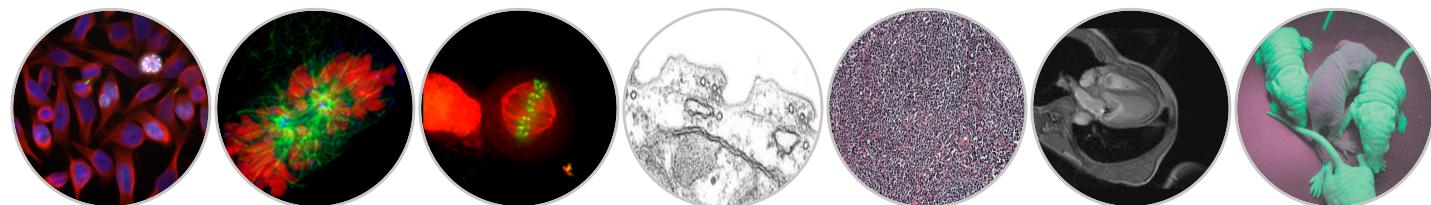


Genes

Compounds

Cell Lines

Imaging Domains



Controlled Vocabularies



Phenotypes

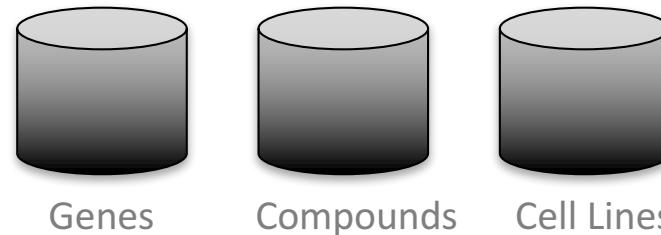


Ground Truth

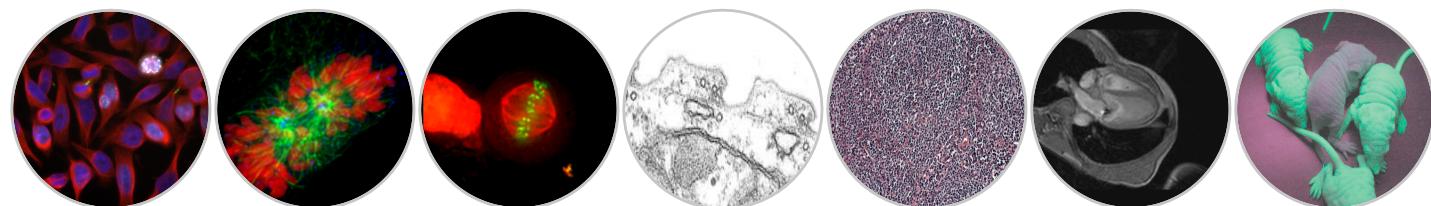
Reference images



Biomolecular Resources



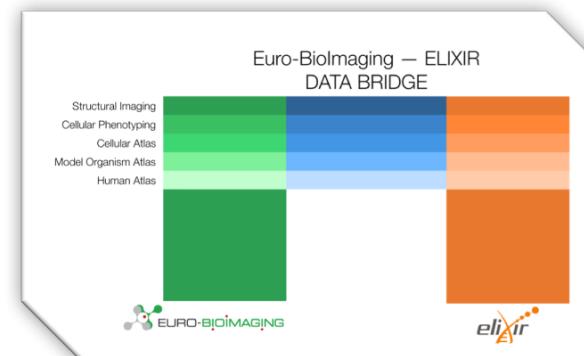
Imaging Domains



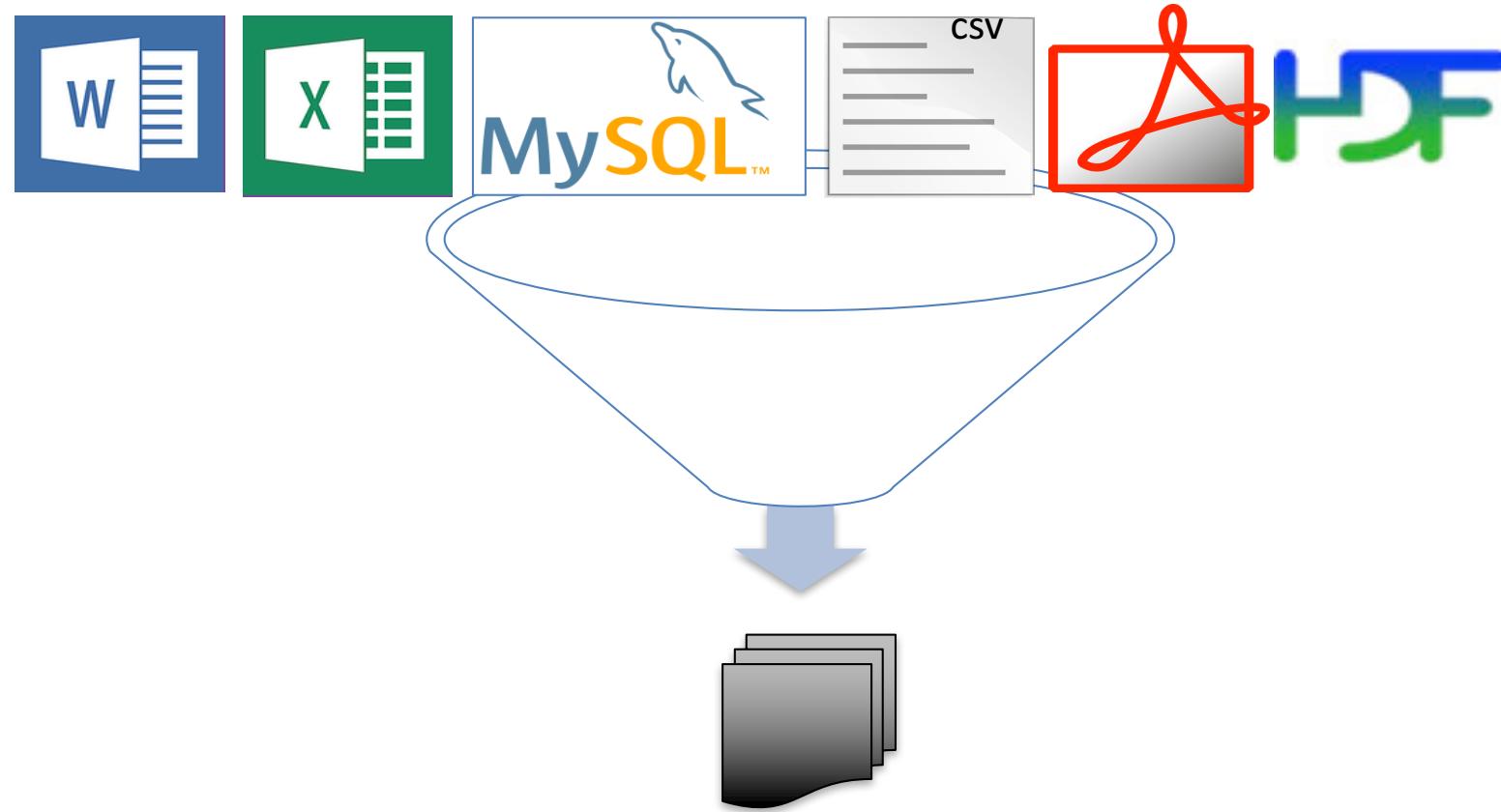
Controlled Vocabularies



Google: "Euro-BioImaging/Elixir Data Strategy"



Submitted Metadata

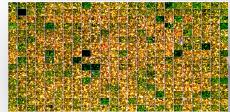


In the style of: MAGE-TAB **isatab** or, Cellular Phenotype Database

The IDR @ EBI Embassy



Gene Product
Targeting HCS



Genetic HCS



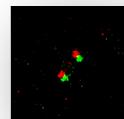
Geographic HCS



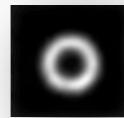
Chemical HCS



Histopathology



3D-Sim



Super-resolution

<https://idr.openmicroscopy.org>

The screenshot shows the IDR Webclient interface. On the left, a sidebar lists various studies under 'Demo data'. In the center, a grid of thumbnails represents 5D images from 'Field 1' across 14 columns and 14 rows. To the right, detailed experimental metadata is displayed, including acquisition parameters like date, time, and resolution, as well as biomolecular annotations and analysis results.

Integrated studies

Experimental metadata

Biomolecular annotations

Analysis results

Thumbnails (of 5D Images)

Well Details

General

Acquisition

Preview

Full viewer

41757 [Well K7, Field 1]

Image ID: 3230268
Well ID: 1315357
Owner: Demo User

Well Details

Acquisition Date: 2018-07-10
Import Date: 2018-07-10
Dimensions (XY): 1080 x 1080
Pixels Type: uint16
Pixels Size (XYZ): 0.66 x 0.66 x -
Z: sections/timepoints: 1 x 1
Channels: Hoechst, ERSyto, ERSytoBleed, PhGolg; Mito
ROI Count: 0

Attributes

Gene

Added by: Demo User
Gene Identifier: 331
Gene Symbol:

ORF

Added by: Demo User
ORF Identifier:

ORF supplementary

Added by: Demo User
ORF Sequence: GGTCTAT ...
ORF Comments: >ccsbBroad304_05835/Open
Reading Frame at 66..1557

Tables

INFO

Plate: 5985
Well: 1315357
Well Number: 247
Characteristics Homo sapiens
[Organism]:
REF:
Term Source 1 NCBI Taxon
Accession:
Characteristics U2OS
(Cell Line):
Term Source 2 EFO
REF:
Term Source 2 EFO_0002869
Accession:
ORF Identifier: ccsbBroad304_05835



Cross-data
browsing



Cloud
analysis



Download
(local analysis)

Vital Stats: IDR @ EBI Embassy



<http://idr.openmicroscopy.org>

Metric	Sept 2017
Image Data Size	42.8 TB
Image files	14.6 M
Datasets	3805
Total Images	2.68 M
Planes	37.3 M
Experiments	1.04M
Genes	19,605
Annotated Images	386 k
Phenotypic Classes	161
External Links	172 k

→~\$25,000/month on AWS←

Williams et al (2017) Nature Methods

IDR reality



42 Studies **Cross-published:** **Models:** **Projects:**

4 Incoming

Datasets & papers linked via DOIs

The screenshot shows a grid of microscopy images from a field labeled 'Field #1'. Overlaid on the images are logos for several journals and databases: BMC Cell Biology, Cell Systems, eLIFE, molecular systems biology, and SCIENTIFIC DATA. To the left of the grid, there is a sidebar with a tree view of datasets and papers linked via DOIs.

Models:

- S. cerevisiae*
- S. pombe*
- Drosophila*
- Arabidopsis*
- Mus*
- Human*
- Plankton*

Projects:

- Mitocheck**
- Sysgro**
- HPA**
- MULTIMOT**
- BBBC**
- Tara**
- HipSci**

SCIENTIFIC DATA

nature

<http://idr.openmicroscopy.org>

IDR datasets



Lawo et al, 2013

Neff et al, 2013

Dickerson et al, 2015

Super-resolution

Histopathology

Cell Dynamics

Brightfield

Geographic screen

Yang et al, 2016

Tara Oceans



IDR datasets- ASH2/ASH2L



Well Details

Import Date: 2016-05-17 01:37:04
Dimensions (XY): 1344 x 1024
Pixels Type: uint16
Pixels Size (XYZ) (μ m): 338.67 x 338.67
Z-sections/Timepoints: 1 x 1
Channels: 0, 1, 2
ROI Count: 0

Tags: 9070

Gene Identifier	9070
Gene Identifier URL	http://www.ncbi.nlm.nih.gov/gene/?term=907
Gene Symbol	ASH2L
RefSeq Target	NM_004674+NM_001105214
Phenotype 5	elongated cells
Phenotype 5 Term Name	elongated cell phenotype
Phenotype 5 Term Accession	CMPO_0000077
Phenotype 5 Term Accession URL	http://www.ebi.ac.uk/cmpo/CMPO_0000077

cmpo - Cellular Microscopy Phenotype Ontology

[Home](#) [Term Request](#) [Developers](#) [About](#)

elongated cell phenotype

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

A phenotype observation at the level of the cell shape where the cell is elongated, with a length notably greater than its width

parents

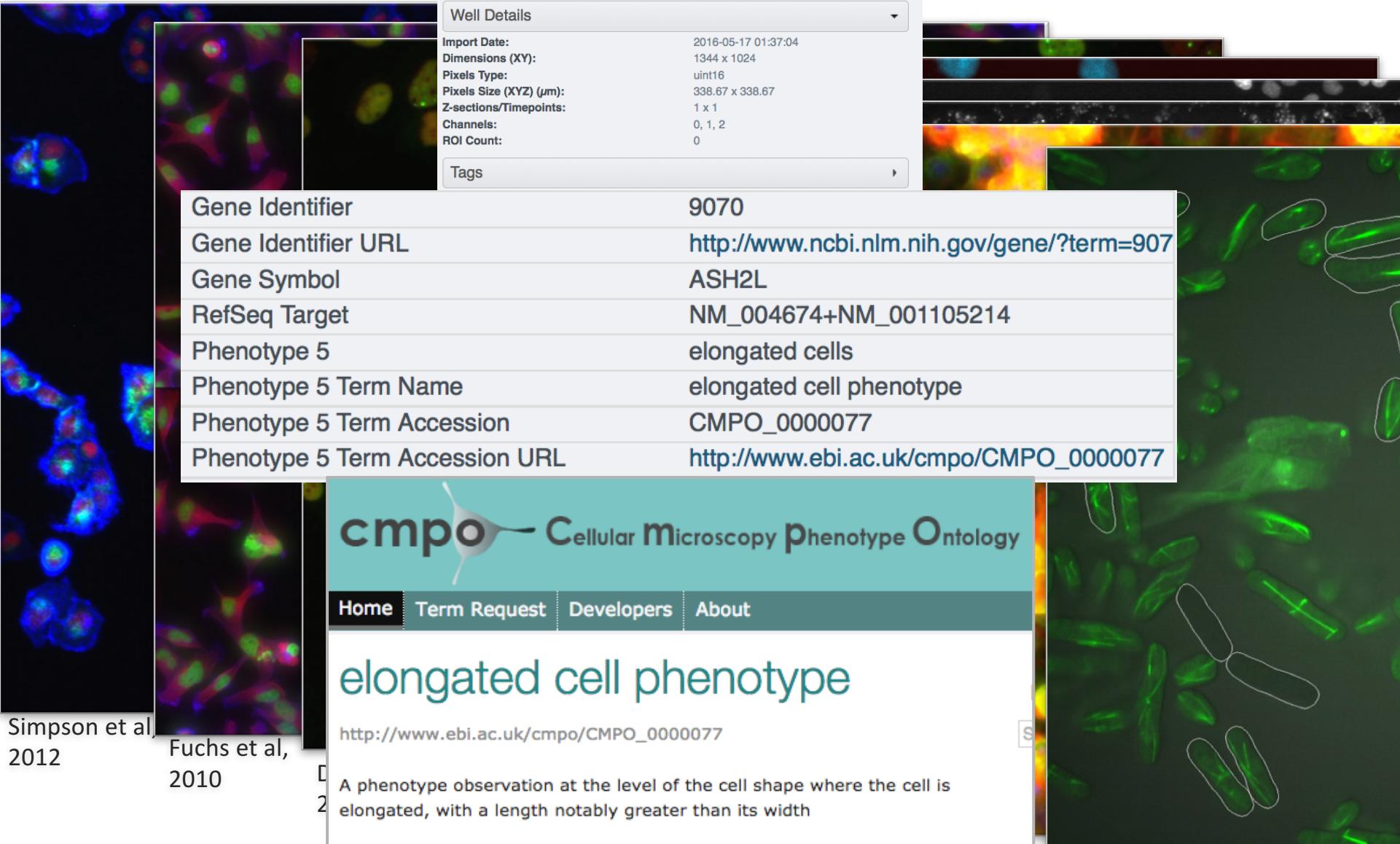
[cell morphology phenotype](#)

LA: 0
P: 0.64
Cluster: 5

Simpson et al., 2012

Fuchs et al., 2010

Graml et al., 2015





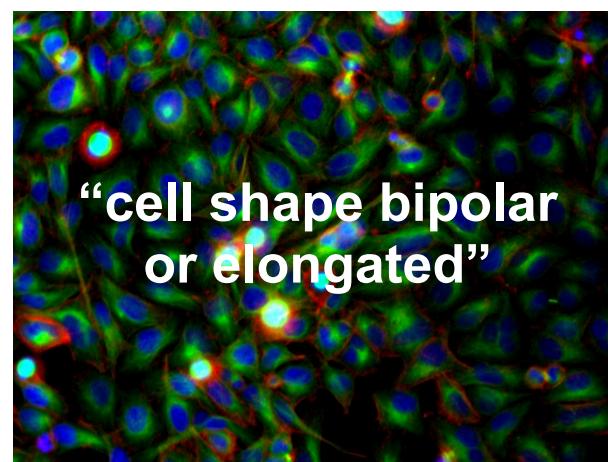
Integration by Phenotype

elongated cell phenotype

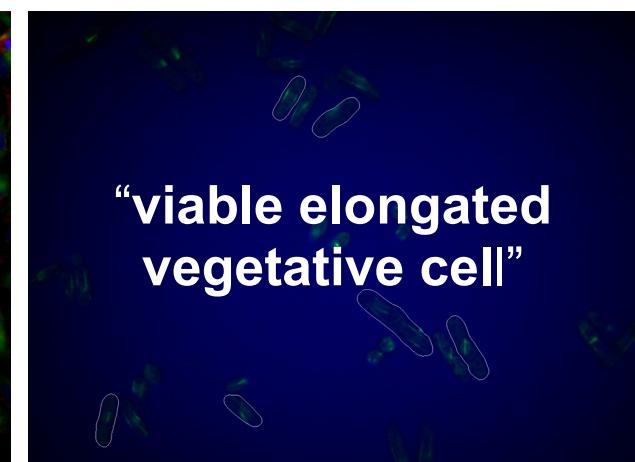
http://www.ebi.ac.uk/cmpo/CMPO_0000077



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

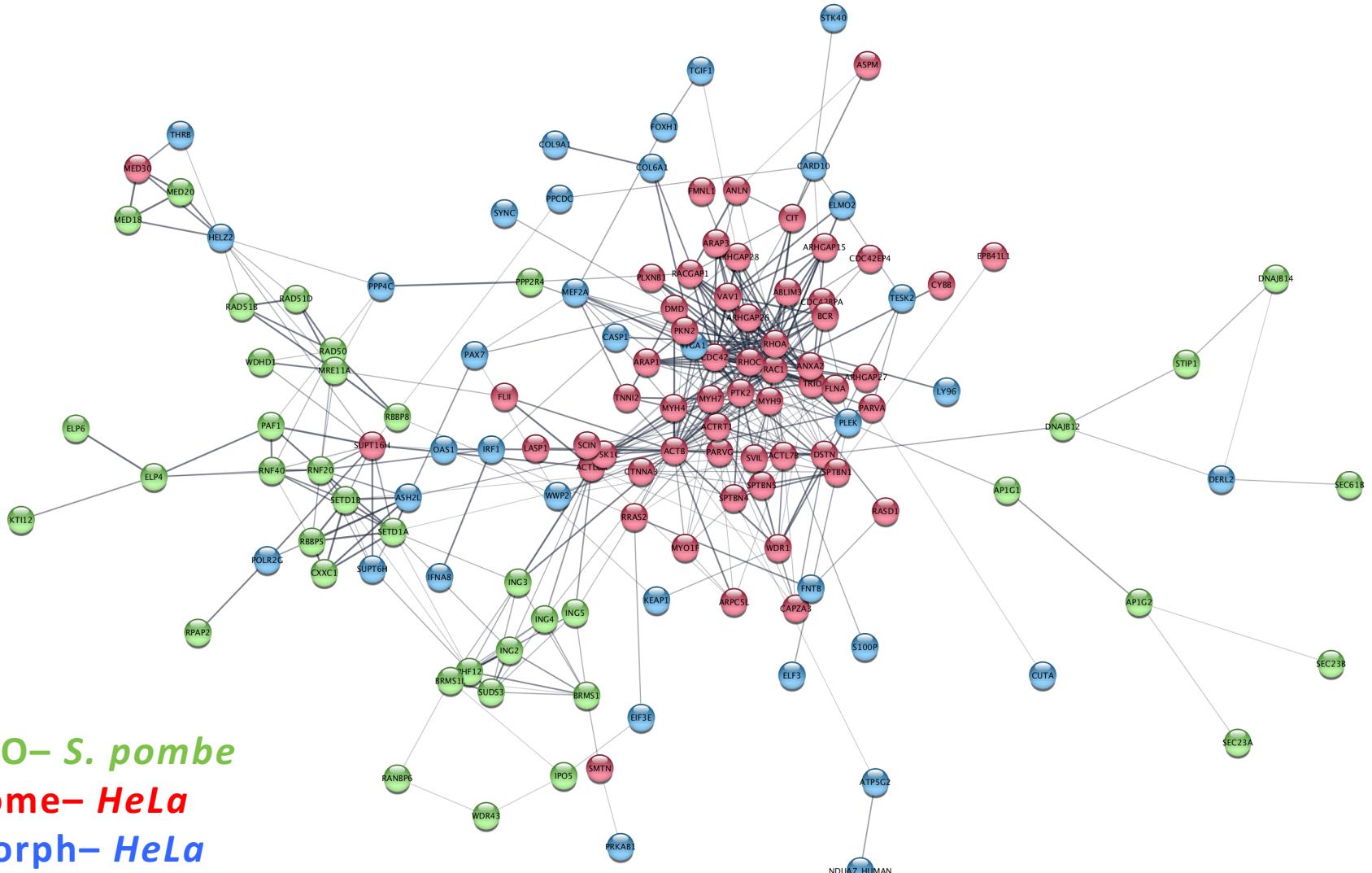


idr0008-B Actinome, Rohn et al 2011, human HeLa targeted actinome RNAi screen (image-109296)



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

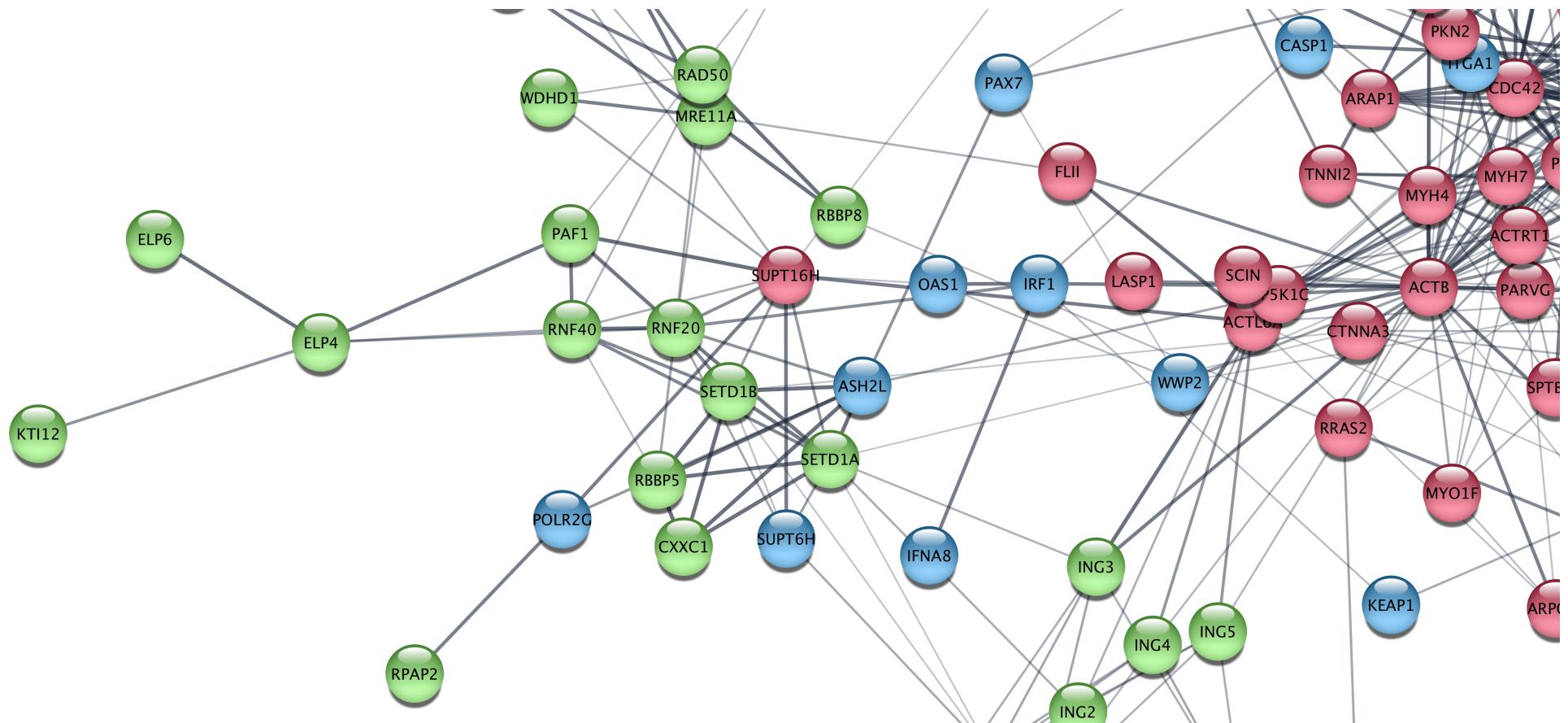
CMPO phenotype: elongated cells



SYSGRO– *S. pombe*
 Actinome– *HeLa*
 CellMorph– *HeLa*



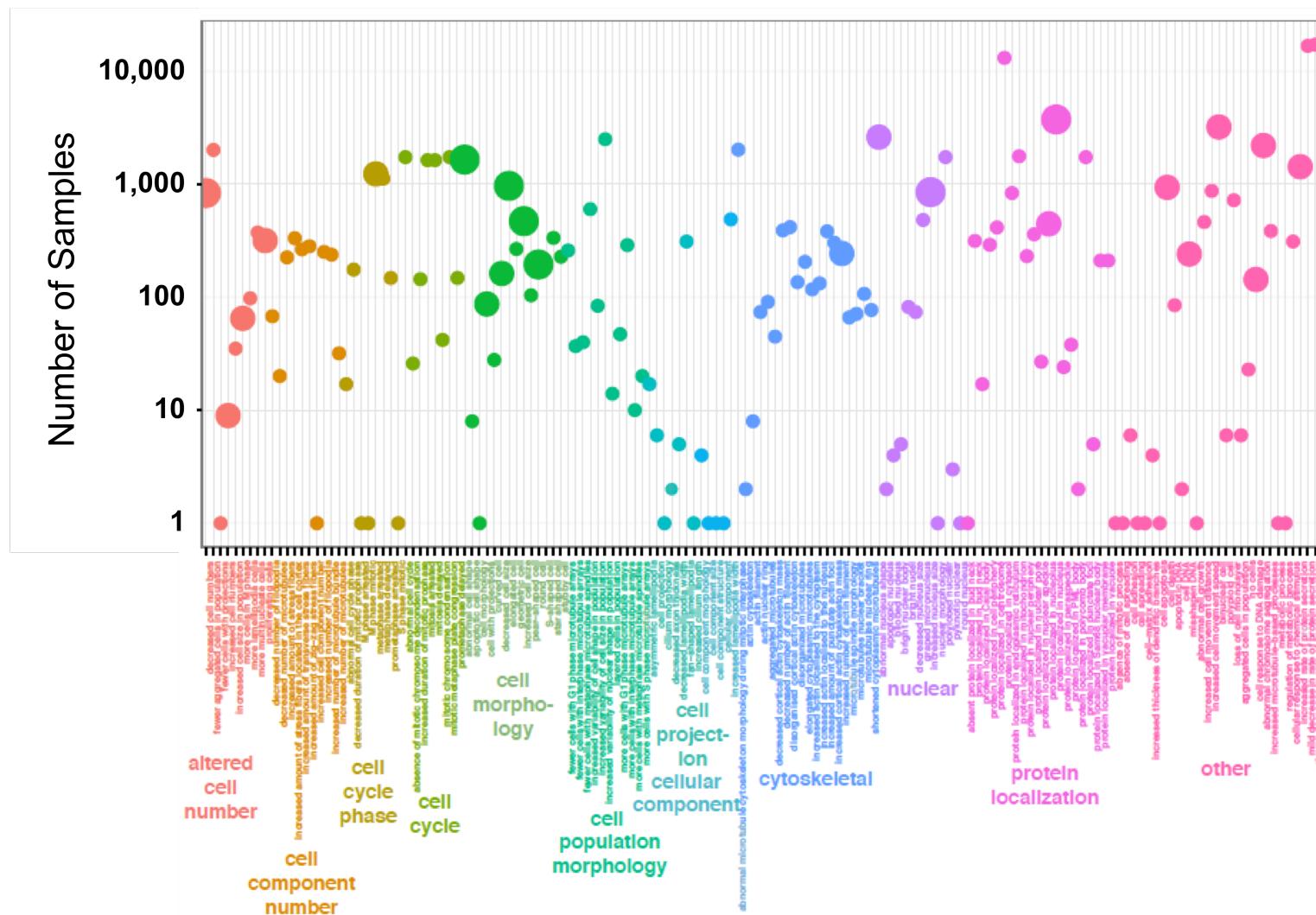
CMPO phenotype: elongated cells



SYSGRO- *S. pombe*
Actinome- *HeLa*
CellMorph- *HeLa*

- Eleanor Williams, Dundee (with help from Cytoscape & String DB)

IDR Phenotypes



Phenotype Class

Williams et al (2017) Nat Methods

Interactive Analysis with IPython in 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



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

	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]



Deploying your own IDR



```
$ source openstack-credentials.env  
  
$ ansible-playbook -i  
inventory/openstack.py -e  
@vars/ome2016-overrides.yml os-idr-
```

```
4. spli@test-gateway:~ (bash)  
  
ssl-certs.yml for test-gateway  
  
TASK [nginx-proxy : nginx | create proxy ssl certificate directory] *****  
changed: [test-gateway] => (item=/etc/nginx/ssl/idr-fullchain.pem)  
ok: [test-gateway] => (item=/etc/nginx/ssl/idr-privkey.pem)  
  
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  
  
(idrstaging)spli@ansible (openstack-playbooks-multivm)$
```

<http://github.com/idr/deployment>

Submissions

A screenshot of a web browser displaying the IDR Submission website at idr.openmicroscopy.org/about/submission.html. The page has a dark background featuring a fluorescence microscopy image of cells. The main title is "Submission of datasets to the IDR". Below it, a paragraph explains the purpose of 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." A link to "Submission Guidelines v1.0 - last modified: March 2017" is provided. The page also includes sections for "What we are looking for" and information about where datasets can be published if they don't meet the criteria.

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

Submissions



idr0000-lastname-example / idr0000-study_HCS.txt

joshmoore Migrate files from Google Drive aaad6a7 on Mar 14

1 contributor

111 lines (95 sloc) | 11.6 KB

Raw Blame History

```
1 # FILL IN AS MUCH INFORMATION AS YOU CAN. HINTS HAVE BEEN PUT IN SOME FIELDS AFTER THE HASH # SYMBOL. REPLACE THE HINT WITH TEXT WHERE APPROPRIATE

# STUDY DESCRIPTION SECTION
"# Section with generic information about the study including title, description, publication details (if applicable) and contact details"

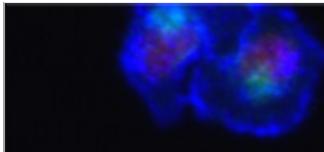
Comment[IDR Study Accession]    # leave blank
Study Title      # fill in
Study Type       high content screen
Study Type REF    EFO
Study Type Term Source REF   NCBITaxon
Study Type Term Accession Number  EFO_0007550
Study Description     # a brief description of the overall aim of the study. The publication abstract can be entered here.
Study Organism "# if more than one organism was studied, enter them in separate columns"
Study Organism Term Source REF  NCBITaxon
Study Organism Term Accession Number  # leave blank
Study Screens Number   # enter how many screens are in the study
Study External URL     # if there is an existing web page related to the study enter it here
Study Public Release Date   # enter a date of when the data can be made public

# Study Publication
Study PubMed ID # fill in if known
Study Publication Title # fill in if known
Study Author List      # fill in if known
Study PMC ID      # fill in if known
Study DOI        # fill in if known

# Study Contacts
Study Person Last Name "# fill in, you can add more people, each in their own column if you wish"
Study Person First Name # fill in
Study Person Email      # fill in
```

<https://github.com/IDR/idr0000-lastname-example>

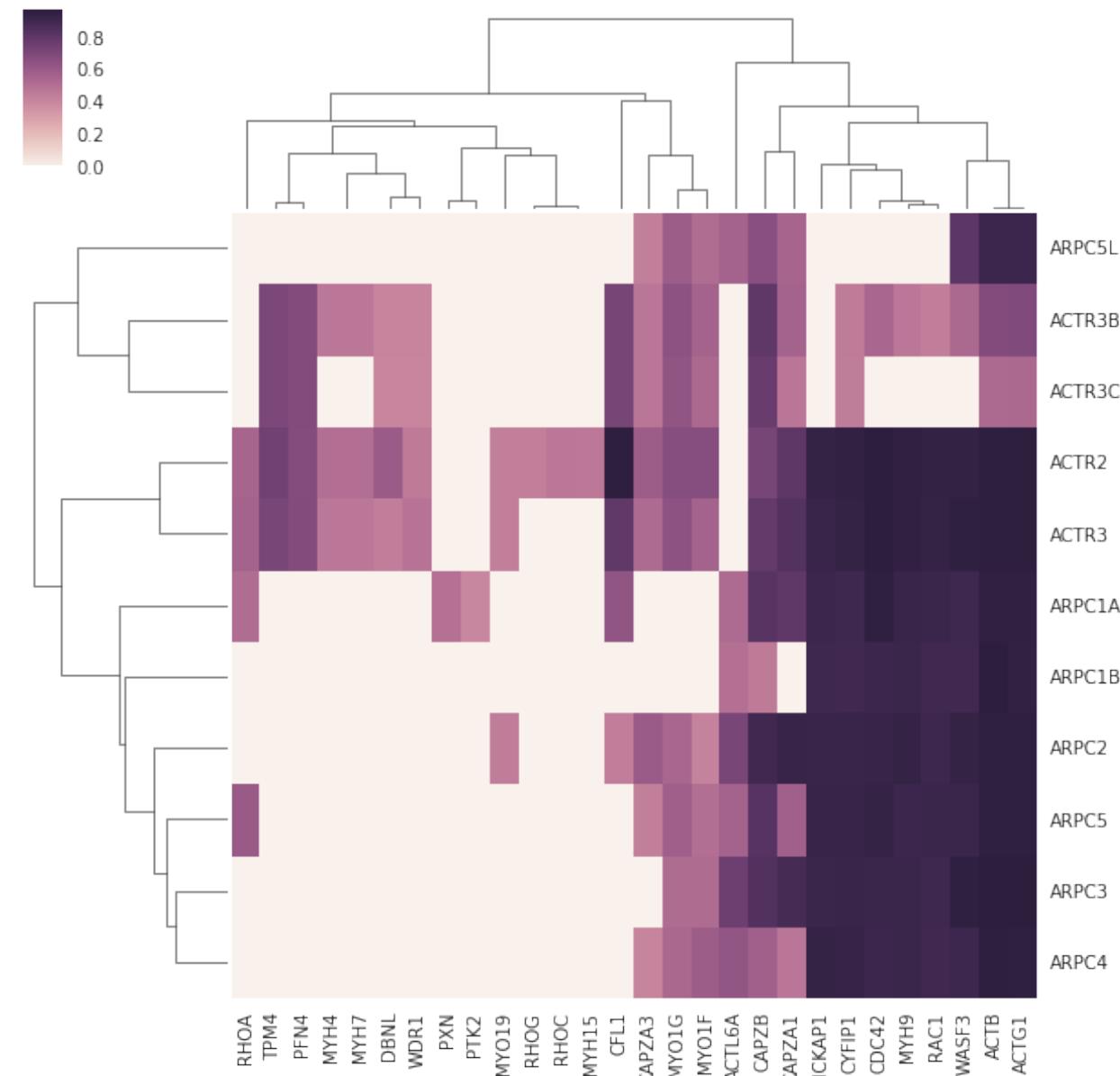
Next Steps



Genes ↔ Phenotypes



StringDB
Significance



Arp2/3
“Query”

Balaji Ramalingam, Dundee

Primary Interactors

29



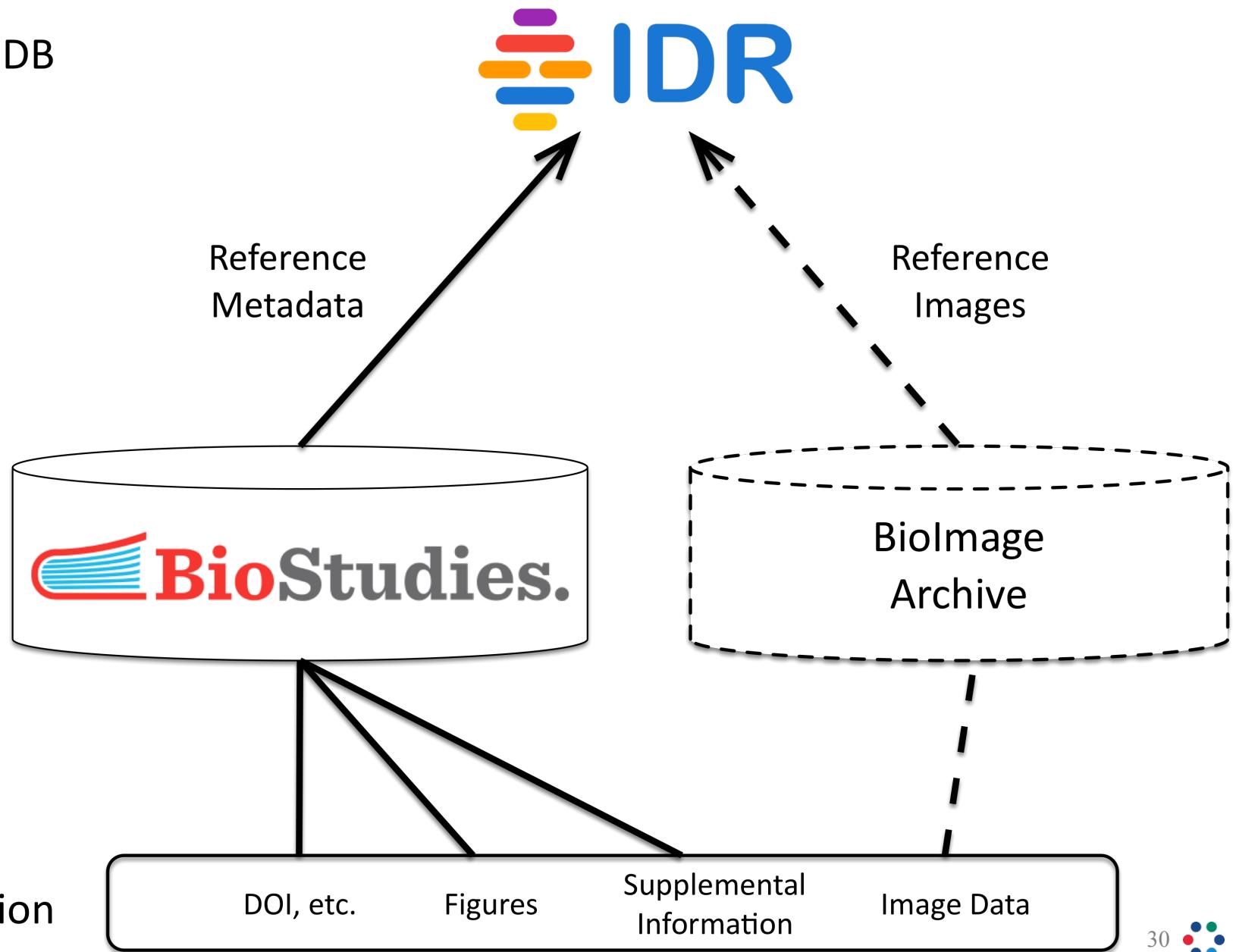
Data Resources Vision



Added Value DB

Data Archive

Publication



Thanks to the Funders



GLOBAL
BIOIMAGING
growing collaboration



Thanks to the IDR Team



Jason
Swedlow



Josh
Moore



Simon
Li



Eleanor
Williams



Simone
Leo



Ola
Tarkowska



Gabry
Rustici



Alvis
Brazma



Ugis
Sarkans



Simon
Jupp



Tony
Burdett

EMBL-EBI



Rafael
Carazo-Salas



Balint
Antal



Anatole
Chessel



/



University of
BRISTOL





Questions?



@IDRnews



@IDRstatus