OME’s Bio-Formats, OMERO & IDR: Open Tools for Comprehensive Cell Phenotyping

Petr Walczysko, Jean-Marie Burel
School of Life Sciences
University of Dundee

The OME Consortium
openmicroscopy.org
@openmicroscopy

Euro-BioImaging
eurobioimaging.eu
@eurobioimaging
The Image Problem is Ubiquitous

Organelles
Cells
Dynamics
Physiology

Lead Discovery
Target Validation
Pathology
In Vivo

A pretty picture?
A measurement?
A resource?
DATA SHARING & PUBLICATION
# Security Model

<table>
<thead>
<tr>
<th>PERMISSIONS</th>
<th>Read</th>
<th>Annotate</th>
<th>Write</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Group-Read</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>You</td>
</tr>
<tr>
<td>Group-Annotate</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>Your Group</td>
</tr>
<tr>
<td>Group-Write</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Public-Read</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>Anyone</td>
</tr>
<tr>
<td>Public-Annotate</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Public-Write</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>
Data Sharing with OMERO
Integrated studies

Thumbnails (of 5D Images)

Gene Product
Targeting HCS

Genetic HCS

Chemical HCS

Histopathology

3D-SIM

Super-resolution

Experimental metadata

Biomolecular annotations

Analysis results

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

Visit: www.eurobioimaging.eu
Email: helpdesk@eurobioimaging-interim.eu

Funded by the Horizon 2020 Framework Program of the European Union
DATA SERVICES VISION

European life scientists as users

HUB

Web-access portal provides virtual access

Image Data Storage and Analysis

IMAGING TOOL RESOURCE
Repository & access point for software tools (user friendly and interoperable)

CLOUDs... Deployable Portable

IMAGE DATA RESOURCE
Resource for reference image data sets (for sharing and re-use)

EuBI PP II / GBI

EuBI PP II
Integrated studies

Analysis results

Experimental metadata

Biomolecular annotations

Thumbnails (of 5D Images)

Gene Product Targeting HCS

Genetic HCS

Geographic HCS

Chemical HCS

Histopathology

3D-Sim

Super-resolution

Cross-data browsing

Cloud analysis

Download (local analysis)
Reference images

Biomolecular Resources
- Genes
- Compounds
- Cell Lines

Imaging Domains

Controlled Vocabularies
- Phenotypes
- Ground Truth

Google: “Euro-BioImaging/Elixir Data Strategy”
Submission of datasets to 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.

Submission Guidelines v1.0 - Last Modified March 2017

What we are looking for

We aim to publish reference image datasets, which have value beyond simply supporting an original publication according to the Euro-BioImaging - Elixir Image Data Strategy. This includes:

• Datasets associated with an existing or upcoming publication
• Complete datasets - not just images supporting one figure in the publication
• Datasets whose metadata can be integrated with other datasets via identifiers from well-known biomolecular resources (Ensembl, NCBI Entrez Gene, RefSeq, PubChem, ChEBI etc)
• Datasets generated using new imaging methods or new analysis methods
• Datasets that are likely to be re-analysed or incorporated into other studies or integrated with other imaging datasets

Image datasets that don’t meet the criteria for reference images may still be published on BioStudies or Dryad. Contact these projects for more information.

https://idr.openmicroscopy.org/about/submission.html
46 Studies
Cross-published:
Models:
Projects:

4 Incoming
Datasets & papers linked via DOIs

https://idr.openmicroscopy.org
## Vital Stats: IDR @ EBI Embassy

https://idr.openmicroscopy.org

<table>
<thead>
<tr>
<th>Metric</th>
<th>August 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Data Size</td>
<td>48.2 TB</td>
</tr>
<tr>
<td>Image files</td>
<td>15.4 M</td>
</tr>
<tr>
<td>Datasets</td>
<td>4288</td>
</tr>
<tr>
<td>Total Images</td>
<td>2.85 M</td>
</tr>
<tr>
<td>Targets</td>
<td>166,114</td>
</tr>
<tr>
<td>Experiments</td>
<td>1.04M</td>
</tr>
<tr>
<td>Genes</td>
<td>19,605</td>
</tr>
<tr>
<td>Annotated Images</td>
<td>386 k</td>
</tr>
<tr>
<td>Phenotypic Classes</td>
<td>161</td>
</tr>
<tr>
<td>External Links</td>
<td>172 k</td>
</tr>
</tbody>
</table>

→$25,000/month on AWS←

IDR datasets

Lawo et al, 2013
Super-resolution
Yang et al, 2016
Brightfield

Neff et al, 2013
Histopathology
Kilpinen et al, 2017
Phenotypic screen

Dickerson et al, 2015
Cell Dynamics
Tara Oceans
Geographic screen

https://idr.openmicroscopy.org
Linked Metadata

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
type elongated cell phenotype
Type Term Accession CMPO_0000077

Type Gene Symbol...

ASH2L (38) 5

Filter Images

HT28 [Well G13, Field 1]
HT28 [Well G13, Field 2]
idr0010-dool-dnadamage/screenA (2) 1
Linked Metadata

idr0012
Virtual Analysis Environment (VAE)

The IDR Virtual Analysis Environment (VAE) supports the open and reproducible analysis of data in the IDR. It is built on JupyterHub and is available to anyone interested in exploring and mining the diverse and vast range of image data and metadata in the IDR.

All notebook sessions are subject to available resources, are limited in the network resources they can access, and will automatically terminate after a period of inactivity. Do not store anything you wish to keep.

If you have any queries about the VAE, including questions about usage, API access and available data, please open an issue on the idr-contrib/community GitHub repository.

Public access
We have a set of Jupyter Notebooks that can be run by anyone without a login.

Click here to obtain a temporary session

GitHub login

https://idr-analysis.openmicroscopy.org/
**Query IDR for Phenotypes**

In [4]:
```python
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

<table>
<thead>
<tr>
<th>Entrez</th>
<th>Ensembl</th>
<th>Key</th>
<th>Value</th>
<th>PhenotypeName</th>
<th>PhenotypeAccession</th>
<th>ScreenIds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARPC2</td>
<td>[10109]</td>
<td>GeneName</td>
<td>ARPC2</td>
<td>[decreased cell numbers, geometric cell phenot...]</td>
<td>[CMPO_0000274, CMPO_0000299, CMPO_0000276, CMPO...]</td>
<td>[206, 206, 206, 206, 206, 206, 206, 206, 206, 206, ...]</td>
</tr>
<tr>
<td>ARPC1A</td>
<td>[10552]</td>
<td>GeneName</td>
<td>ARPC1A</td>
<td>[strong decrease in rate of protein secretion]</td>
<td>[CMPO_0000319]</td>
<td>[251]</td>
</tr>
<tr>
<td>ARPC5L</td>
<td>[81873]</td>
<td>GeneName</td>
<td>ARPC5L</td>
<td>[increased actin localised to the nucleus, elo...]</td>
<td>[CMPO_0000261, CMPO_0000287, CMPO_0000288, CMPO...]</td>
<td>[206, 206, 206, 206, 206, 206, 206, 206, 206, 206, ...]</td>
</tr>
<tr>
<td>ACTR3C</td>
<td>[653857]</td>
<td>EnsemblID</td>
<td>ENSG000000106526</td>
<td>[cell death phenotype, mitosis delayed phenot...]</td>
<td>[CMPO_0000305, CMPO_0000348, CMPO_0000307, CMPO...]</td>
<td>[1101, 1101, 1101, 1101, 1101, 1101, 1101, 1101, 1101, 1101]</td>
</tr>
<tr>
<td>ACTR3B</td>
<td>[57180]</td>
<td>GeneName</td>
<td>ACTR3B</td>
<td>[binuclear cell phenotype]</td>
<td>[CMPO_0000213]</td>
<td>[1101]</td>
</tr>
<tr>
<td>HEL-68</td>
<td>[10552]</td>
<td>EnsemblID</td>
<td>ENSG000000241685</td>
<td>[strong decrease in rate of protein secretion]</td>
<td>[CMPO_0000319]</td>
<td>[251]</td>
</tr>
<tr>
<td>ARPC4</td>
<td>[10093]</td>
<td>GeneName</td>
<td>ARPC4</td>
<td>[mild decrease in rate of protein secretion]</td>
<td>[CMPO_0000318]</td>
<td>[251]</td>
</tr>
<tr>
<td>ACTR3</td>
<td>[10096]</td>
<td>GeneName</td>
<td>ACTR3</td>
<td>[strong decrease in rate of protein secretion,...]</td>
<td>[CMPO_0000319, CMPO_0000319]</td>
<td>[251, 803]</td>
</tr>
</tbody>
</table>

https://idr-analysis.openmicroscopy.org/
(3) Enabling independent IDR

```bash
$ 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
```

https://github.com/IDR/deployment
(3) Enabling independent IDRs

- EBI / EuBI
- Human Cell Atlas
- Digital Pathology
- EMBRC
- ...
- ...
- ...
(1) Integration with data archives

Added Value DB

Data Archive

Publications

[Diagram showing integration between IDR, BioStudies, and BioImage Archive]

Ellenberg et al (2018) arxiv
Thanks to the Funders
Thanks to the IDR Team

Jason Swedlow
Josh Moore
Simon Li
Eleanor Williams
Seb Besson
Dominik Lindner
Gabry Rustici
Alvis Brazma
Ugis Sarkans
Simon Jupp
Tony Burdett
Rafael Carazo-Salas
Balint Antal
Anatole Chessel