

TheCellVision.org: A Database for Visualizing and Mining High-Content Cell Imaging Projects

Myra Paz David Masinas,^{*1} Mojca Mattiazzi Usaj,^{*1,2} Matej Usaj,^{*} Charles Boone,^{*1,2} and

Brenda J. Andrews^{*1,2}

^{*}The Donnelly Centre, University of Toronto, Ontario, M5S 3E1, Canada and [†]Department of Molecular Genetics, University of Toronto, Ontario, M5S 3E1, Canada

ORCID IDs: 0000-0002-7389-8478 (M.M.U.); 0000-0003-0675-2847 (M.U.); 0000-0002-3542-6760 (C.B.); 0000-0001-6427-6493 (B.J.A.)

ABSTRACT Advances in genome engineering and high throughput imaging technologies have enabled genome-scale screens of single cells for a variety of phenotypes, including subcellular morphology and protein localization. We constructed TheCellVision.org, a freely available and web-accessible image visualization and data browsing tool that serves as a central repository for fluorescence microscopy images and associated quantitative data produced by high-content screening experiments. Currently, TheCellVision.org hosts ~575,590 images and associated analysis results from two published high-content screening (HCS) projects focused on the budding yeast *Saccharomyces cerevisiae*. TheCellVision.org allows users to access, visualize and explore fluorescence microscopy images, and to search, compare, and extract data related to subcellular compartment morphology, protein abundance, and localization. Each dataset can be queried independently or as part of a search across multiple datasets using the advanced search option. The website also hosts computational tools associated with the available datasets, which can be applied to other projects and cell systems, a feature we demonstrate using published images of mammalian cells. Providing access to HCS data through websites such as TheCellVision.org enables new discovery and independent re-analyses of imaging data.

KEYWORDS

high-content
screening
quantitative
image analysis
subcellular
morphology
subcellular
localization
protein
abundance
penetration
image analysis
tools

High-content screening (HCS), which combines HTP microscopy with multi-parametric quantitative image analysis, has been used in recent years to explore different aspects of cell biology in diverse cell systems. HCS experiments produce massive image datasets, catalyzing an increasing demand for computational image analysis methods that enable extraction of quantitative information about a variety of cell features, including subcellular morphology, protein localization

and abundance, and the dynamics of proteins or cellular compartments in response to genetic or environmental perturbations (Chong *et al.* 2015; de Groot *et al.* 2018; Heigwer *et al.* 2018; Mattiazzi Usaj *et al.* 2020; Mattiazzi Usaj *et al.* 2016; Styles *et al.* 2016; Yin *et al.* 2013). Several tools have been established in budding yeast that allow the exploration of gene and protein function using fluorescence microscopy. One key collection is the array of strains expressing GFP-fusion proteins (Huh *et al.* 2003) which enables proteome-scale exploration of protein abundance and localization (Breker *et al.* 2013; Chong *et al.* 2015; Denervaud *et al.* 2013; Kraus *et al.* 2017; Tkach *et al.* 2012). Collections of strains carrying deletions of non-essential yeast genes (Giaever *et al.* 2002) or temperature-sensitive alleles of essential genes (Costanzo *et al.* 2016; Li *et al.* 2011) can be modified using automated yeast genetics to express fluorescent reporters of compartments or biological processes of interest, or stained with fluorescent dyes, and then used in imaging screens to quantify the phenotypic changes associated with genetic perturbations (Eapen *et al.* 2017; Mattiazzi Usaj *et al.* 2015; Mattiazzi Usaj *et al.* 2020; Neumuller *et al.* 2013; Ohya *et al.* 2005; Sauerwald *et al.* 2015; Styles *et al.* 2016; Wolinski *et al.* 2009).

Copyright © 2020 Masinas *et al.*

doi: <https://doi.org/10.1534/g3.120.401570>

Manuscript received July 16, 2020; accepted for publication August 28, 2020; published Early Online September 15, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.12950441>.

¹These authors contributed equally to this work.

²Corresponding authors: The Donnelly Centre, University of Toronto, 160 College Street, Room 1308, Toronto, ON M5S 3E1, Canada.

E-mail: m.mattiaziusaj@utoronto.ca; charlie.boone@utoronto.ca; brenda.andrews@utoronto.ca

As noted above, high-content microscopy screens involve the acquisition of hundreds of thousands of images, from which single cell data are extracted and analyzed, producing several GBs to TBs of image data and associated numerical values. Recently, there have been important community efforts to develop standards and repositories for biological images and associated raw data (Orloff *et al.* 2012; Williams *et al.* 2017), which will accelerate re-analysis and higher-level mining of image datasets. However, the huge variety of biological images – including clinical, whole organism and single cell images – means that project- or organism-specific websites continue to serve a valuable complementary role. For example, several stand-alone websites allow users to search images of the yeast GFP-fusion collection (Breker *et al.* 2014; Dubreuil *et al.* 2019; Huh *et al.* 2003; Koh *et al.* 2015), with most accessible through the *Saccharomyces Genome Database* (SGD) (Cherry *et al.* 2012).

To examine subcellular phenotypes, we have performed several high-content screens using arrays of budding yeast mutants (Mattiuzzi Usaj *et al.* 2020; Styles *et al.* 2016). Images and associated results from these and other studies that have systematically assessed (sub)cellular morphologies of single-gene mutant strains (Eapen *et al.* 2017; Mattiuzzi Usaj *et al.* 2015; Neumuller *et al.* 2013; Ohya *et al.* 2005; Sauerwald *et al.* 2015; Wolinski *et al.* 2009; Zhao *et al.* 2016) are often not incorporated into community databases, such as SGD, and thus they are less accessible for mining. Access to images and data from these types of HCS projects is needed to facilitate further analysis of the data, to allow for multi-project data mining to answer additional questions not addressed in primary publications, and to serve as benchmarking datasets for the development of novel image analysis methods. To address this need, we developed a centralized database and website, TheCellVision.org, that hosts both HCS projects employing the collection of GFP-fusion proteins and collections of single-gene mutant strains. In addition to the images and processed data obtained from HCS projects, we have made the analysis tools available on TheCellVision.org.

MATERIALS AND METHODS

High-throughput imaging screens and analyses

At present, TheCellVision.org houses two published datasets:

- i. CYCLOPs (Collection of Yeast Cell and Localization Patterns). CYCLOPs consists of high-throughput imaging screens of yeast GFP-tagged strains, as well as the analysis of protein abundance and localization, which was obtained using support vector machines and convolutional neural networks, as described in Chong *et al.* and Kraus *et al.* (Chong *et al.* 2015; Kraus *et al.* 2017). The GFP collection was imaged in standard conditions, in the presence of 3 chemicals (HU – hydroxyurea; RAP – rapamycin, AF – alpha factor), and one mutant background (deletion of *RPD3* gene). The previous version of the CYCLOPs web-accessible database (Koh *et al.* 2015) has now been migrated to TheCellVision.org.
- ii. Endocytic Compartment Morphology dataset. This dataset includes images from high-throughput screens of yeast single-gene mutants combined with single cell analysis of subcellular, endocytic compartment morphology, obtained using neural networks, as described in Mattiuzzi Usaj *et al.* (Mattiuzzi Usaj *et al.* 2020).

Database development and schema

TheCellVision.org is a Django-based web application written in Python that allows access to multiple datasets obtained from various research projects. To store and manage large-scale data, TheCellVision.org uses a PostgreSQL object-relational database system on its

backend. The web interface is developed using HyperText Markup Language (HTML5), Cascading Style Sheets (CSS3) and JavaScript. Open-source libraries such as Bootstrap and jQuery are utilized for easier implementation and to ensure an efficient web design across modern browsers. An nginx web server is used to process incoming client requests over Hypertext Transfer Protocol (HTTP) and serves responses through a Web Server Gateway Interface (uWSGI).

Currently, TheCellVision.org houses two research projects (see above), which together comprise ~575,590 fluorescence microscopy images from ~9,767 yeast strains and the associated quantitative data. The database schema is organized into three main clusters – one for each project and a core group which consists of tables that are shared by the projects (Figure S1). The database schema will continue to evolve as new experimental data are added to the database; the design is general and flexible, enabling the addition of new image datasets without database reconstruction.

Analysis of the BBBC021 dataset

To test the image analysis tools developed for the Endocytic Compartment Morphology dataset (Mattiuzzi Usaj *et al.* 2020) on an independent image dataset, we used the BBBC021v1 benchmarking set available from the Broad Bioimage Benchmark Collection (Caie *et al.* 2010; Ljosa *et al.* 2012). The dataset consists of fluorescence micrographs of MCF-7 breast cancer cells acquired on a high-content screening microscope (Caie *et al.* 2010). From the dataset, we collected subsets of images from a unique compound-concentration combination for each of 13 phenotypic classes (Caie *et al.* 2010).

CellProfiler (Carpenter *et al.* 2006) was used for object segmentation and extraction of cellular features such as area, shape, intensity, texture, granularity and radial distribution. The segmentation pipeline is available at <https://data.broadinstitute.org/bbbc/BBBC021/> (section ‘CellProfiler pipelines’). To segment the cells, primary objects (nuclei) were first identified from the DNA channel, then the F-actin channel was used to identify the secondary objects (cells). Lastly, to determine cytoplasmic regions, each identified nucleus was subtracted from its corresponding cell object. Cells touching the border of the image were discarded. In total, we extracted quantitative information for 22,529 cells from 160 image sets (~1.2% of the dataset). The single cell classification tool used for analysis of the Endocytic Compartment Morphology dataset (2-hidden-layer fully connected neural network) was then applied to the extracted data (Mattiuzzi Usaj *et al.* 2020).

Data availability

Bulk download of images from the CYCLOPs dataset will be provided upon request. Raw and processed images associated with the Endocytic Compartment Morphology dataset were deposited to the Image Data Resource (<https://idr.openmicroscopy.org>) under accession number idr0078. Datasets can be downloaded from each project’s ‘Supplemental Files’ tab. This includes protein localization and abundance, penetrance and phenotype data. Details on the CYCLOPs datasets and the Endocytic Compartment Morphology dataset are described elsewhere (Chong *et al.* 2015; Kraus *et al.* 2017; Mattiuzzi Usaj *et al.* 2020).

Code for the 2 hidden-layer fully connected neural network for single cell classification, and other developed tools is available at <https://thecellvision.org/tools>.

All supplemental material is available at figshare. Figure S1 illustrates the database schema for each of the clusters: A) core, B) Endocytic Compartment Morphology project, C) CYCLOPs project.

Table S1 provides the number of cells for each class of the BBBC021 dataset included in the analysis, classification accuracy for each class, and the effect of training set size on classification accuracy. Supplemental material available at figshare: <https://doi.org/10.25387/g3.12950441>.

RESULTS AND DISCUSSION

Image collections available on TheCellVision.org

The first release of TheCellVision.org houses two large-scale datasets: the Collection of Yeast Cell and Localization Patterns dataset (CYCLOPs) (Chong *et al.* 2015; Kraus *et al.* 2017), and the Endocytic Compartment Morphology dataset (Mattiuzzi Usaj *et al.* 2020). The CYCLOPs dataset comprises 385,984 images: 330,688 images from 18 screens described in Chong *et al.* (Chong *et al.* 2015), and 55,296 images from an alpha-factor time-course screen described in Kraus *et al.* (Kraus *et al.* 2017). These are 2-channel images: the green (GFP) channel represents a GFP-fused protein from the ORF-GFP collection (Huh *et al.* 2003), while the red (RFP) channel identifies the cytosol and was used for cell segmentation purposes. In total the CYCLOPs dataset covers 4,139 proteins from *S. cerevisiae*. These proteins have been classified into 16 localization classes (Figure 1A) (Chong *et al.* 2015).

The Endocytic Compartment Morphology dataset comprises 189,606 images, which are single-channel maximum z-projections depicting the subcellular morphology of 4 endocytic markers (Figure 1B) in 5,628 yeast mutant strains (corresponding to 5,267 *S. cerevisiae* ORFs). The screened mutant strains include non-essential gene deletion strains, and strains with temperature-sensitive alleles of essential genes. All temperature-sensitive strains were screened at two temperatures (26° and 37°). The identification of significant morphology mutants only included temperature-sensitive strains screened at the non-permissive temperature (37°). However, all screens done at the permissive temperature (26°) are included in TheCellVision.org together with micrographs, and results from penetrance and phenotype analyses.

TheCellVision.org functionalities

Our aim was to design a database that provides the user with a simple yet informative interface for accessing data from our image-based yeast screens. The landing page contains icons for each of the main research projects that lead to their respective home page (Figure 1C). From the landing page, the user can also access the image analysis tools ('Tools'), and an advanced search option ('Advanced'). In addition, a search box allows the user to search for genes of interest across all open projects. Clicking on TheCellMap.org icon on the right side of the search bar will open a separate window displaying the genetic interaction results for the queried genes on our sister site, TheCellMap.org, which hosts data associated with the global yeast synthetic genetic interaction mapping project (Costanzo *et al.* 2016; Usaj *et al.* 2017).

Search box and general search results: Typing into the search box on the landing page will prompt the user with a list of available matching gene names. Alternatively, the user also has the option to directly paste in a list of genes. Clicking on the search icon triggers the request to retrieve available data associated with the input genes across all projects. The search results page displays the input list of genes on the left pane. On top of the results page is a brief summary of the selected gene with a link to the gene's summary page on SGD. For easier navigation between projects, the results page has a

tab-separated interface. Searching for a gene that has not been screened will return a "No results found" message in the main results window. (Figure 2A).

The Endocytic Compartment Morphology tab provides basic information on the imaged strain with the number of cells and replicates included in the analysis, penetrance data for the 4 screened markers, fraction of cells with each phenotype for the 17 mutant phenotypes (collapsible table), and all available microscopy images, including examples of wild-type strain micrographs. A brief legend can be accessed by clicking on the collapsible 'Information' icon. On the CYCLOPs tab, the displayed data include information on protein subcellular localization and abundance, as well as localization changes in the tested conditions (described in Chong *et al.* (Chong *et al.* 2015)), and micrographs from all CYCLOPs screens for the searched strain. Links to external protein localization databases are included at the bottom of the page. CYCLOPs data can also be accessed through SGD's Localization Resources. Finally, a download button is available in the top right corner which allows the user to download a spreadsheet containing the data provided on the results page for all queried genes. An information icon beside the download button acts as a link to a separate page listing the column information for the downloaded file.

Image modals: The image thumbnails displayed on the results page are in PNG format, and are generated instantaneously for each searched gene from original TIF images stored in the database. Clicking on a thumbnail opens up a modal with the enlarged image (Figure 2B). The modal's header provides general information on the strain and screen being displayed (*e.g.*, *SEC12*, vacuole in the example shown in Figure 2A). A magnifying glass icon is available on the left side of the modal. Once the icon is clicked and the user hovers the pointer over the image, the pointer transforms into a magnifying glass which allows the user to zoom into cells of interest (Figure 2B). By default, the images have their intensity values scaled to the maximum pixel value in the image. Clicking on the 'Adjust brightness' icon opens two sliders for adjusting the image brightness: the brightness settings on each slider will be propagated to all images of the selected replicate of the searched mutant stain and all displayed control wild-type strain images, respectively. Navigation buttons at the bottom of the displayed image can be used to browse through individual images in all available replicates. For the Endocytic Compartment Morphology project, the mutant (left) and wild-type images from the same screen (right) are displayed side by side for phenotype comparison (Figure 2B). The displayed wild-type images are randomly selected from the same imaging plate as the mutant. To download the set of images for the currently selected strain and marker as a single .zip file, the user can click the 'Download images' icon. The downloaded images are all 16-bit uncompressed grayscale TIFs.

For CYCLOPs, toggle buttons are available for viewing individual fluorescence channels. By default, an overlay of both channels is shown. In addition, the user can enable the dual view (eye icon) that allows side-by-side comparison of images from different CYCLOPs screens.

Image viewer: For a more comparative visualization of images, an 'Image Viewer' tab is included on the general results page. For a selected gene, the image viewer displays a gallery of all available micrographs across all projects, including mutants of the gene with various endocytic compartments tagged with fluorescent proteins and GFP-tagged versions of the encoded protein. Alternatively, a gallery of all available micrographs across all searched genes for a selected

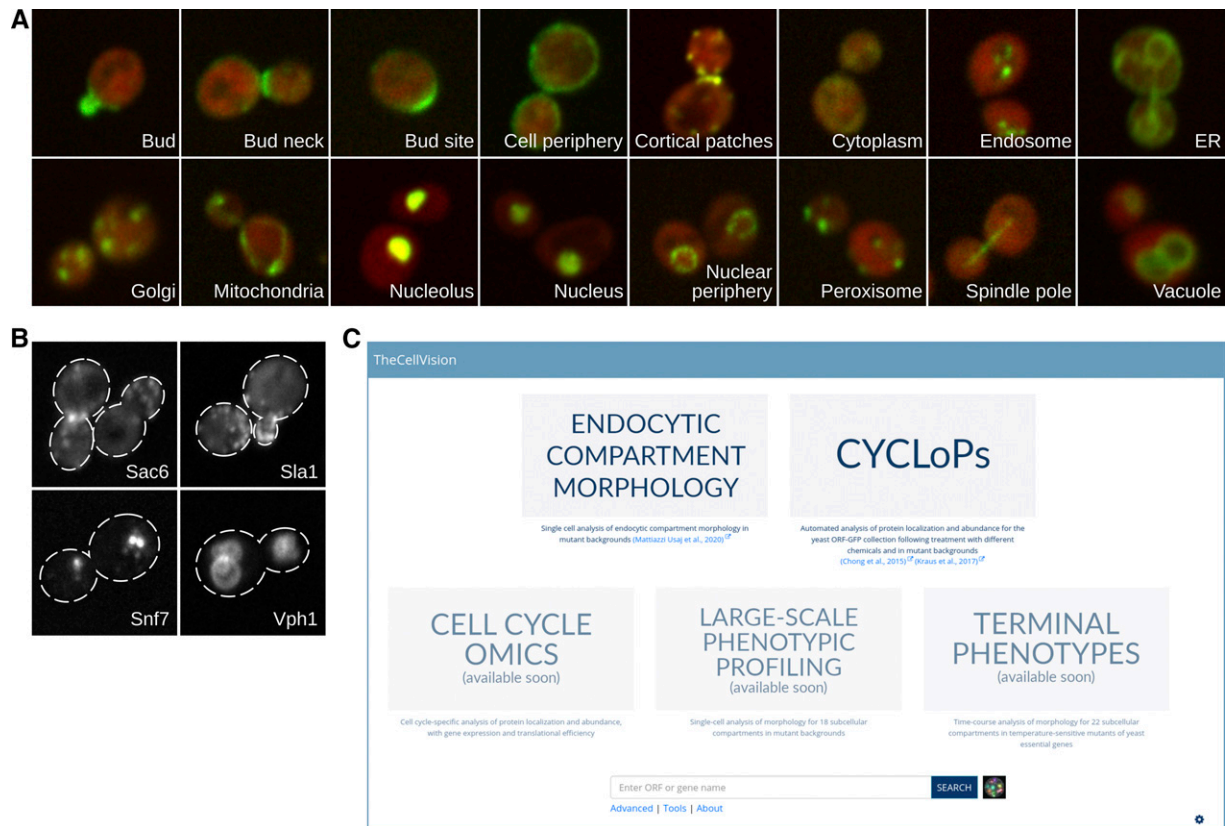


Figure 1 A) Example micrographs of the 16 localization classes of the Collection of Yeast Cell and Localization Patterns dataset (CYCLoPs). B) Example micrographs of the 4 endocytic markers of the Endocytic Compartment Morphology dataset. C) TheCellVision.org landing page.

project may be viewed ('All genes' option). The magnifying glass and brightness adjustment icons can be found on the top right side.

Advanced search: From the main landing page, clicking the 'Advanced' link takes the user to the advanced search page. This page allows the user to search for more complex combinations of terms across different projects and/or screens using logical AND, OR, NOT operations (Figure 2C). The user can select from a list of projects, and predefined project-specific data types, markers, screens, phenotypes, or localizations. For example, a user might want to ask whether any endoplasmic reticulum proteins are important for proper actin patch morphology. In this case, a search for genes that when mutated cause a defect in actin patch morphology, and whose products localize to the endoplasmic reticulum returns 82 genes (Figure 2D) indicating that many ER proteins play an important role in actin patch morphology. For search results involving more than 20 genes, a pagination bar appears on top of the gene list that allows the user to navigate between pages.

The advanced search can also combine different data types from the same project. For example, using the CYCLoPs dataset the user can look for proteins that go to the vacuole for degradation after rapamycin treatment (example input parameters for this search are shown in Figure 2E). The search returns a list of 7 proteins that are involved in translation or nutrient transport (Figure 2E) – inhibition of these processes is consistent with known effects of TOR inhibition with rapamycin (Loewith and Hall 2011).

Image analysis tools: The field of bioimage analysis has seen rapid growth in recent years. There are several available software tools that

perform different image processing and analysis steps (Smith *et al.* 2018). All the tools and code developed as part of the projects hosted by TheCellVision.org are available in the 'Tools' tab. The link next to each tool will take the user to its respective GitHub page. These tools can be adapted for the analysis of other fluorescent markers and subcellular compartments, and can also be used, as we show below, to analyze images in a species-independent manner.

One of the tools, the single cell labeling tool, offers a simple way to manually annotate single cells when building a classification training set (Figure 3A). This graphical user interface (GUI) application allows the user to view and label single cell images on a grid layout. The required input is a simple spreadsheet containing single cell data, such as image path location and cell coordinates, and a list of possible labels. The user can annotate single cells with one of the labels and export the classification training set. This classification training set can be used in combination with the 2-hidden-layer fully connected neural network (ODNN in 'Tools' in TheCellVision.org) developed in Mattiazzi Usaj *et al.* (Mattiazzi Usaj *et al.* 2020) to classify phenotypes.

Because the inputs are quantitative single cell descriptors, we predicted that the 2-hidden-layer fully connected neural network classifier could be used for the analysis of both yeast and other cells. To test the usability and performance of the classifier on human cell images, we took advantage of the BBBC021 benchmarking set available from the Broad Bioimage Benchmark Collection (Caie *et al.* 2010; Ljosa *et al.* 2012). The dataset consists of fluorescence micrographs of MCF-7 human breast cancer cells acquired on a high-content screening microscope. We first tested the performance of the classifier on a set of 22,529 cells extracted from only ~1.2% of images

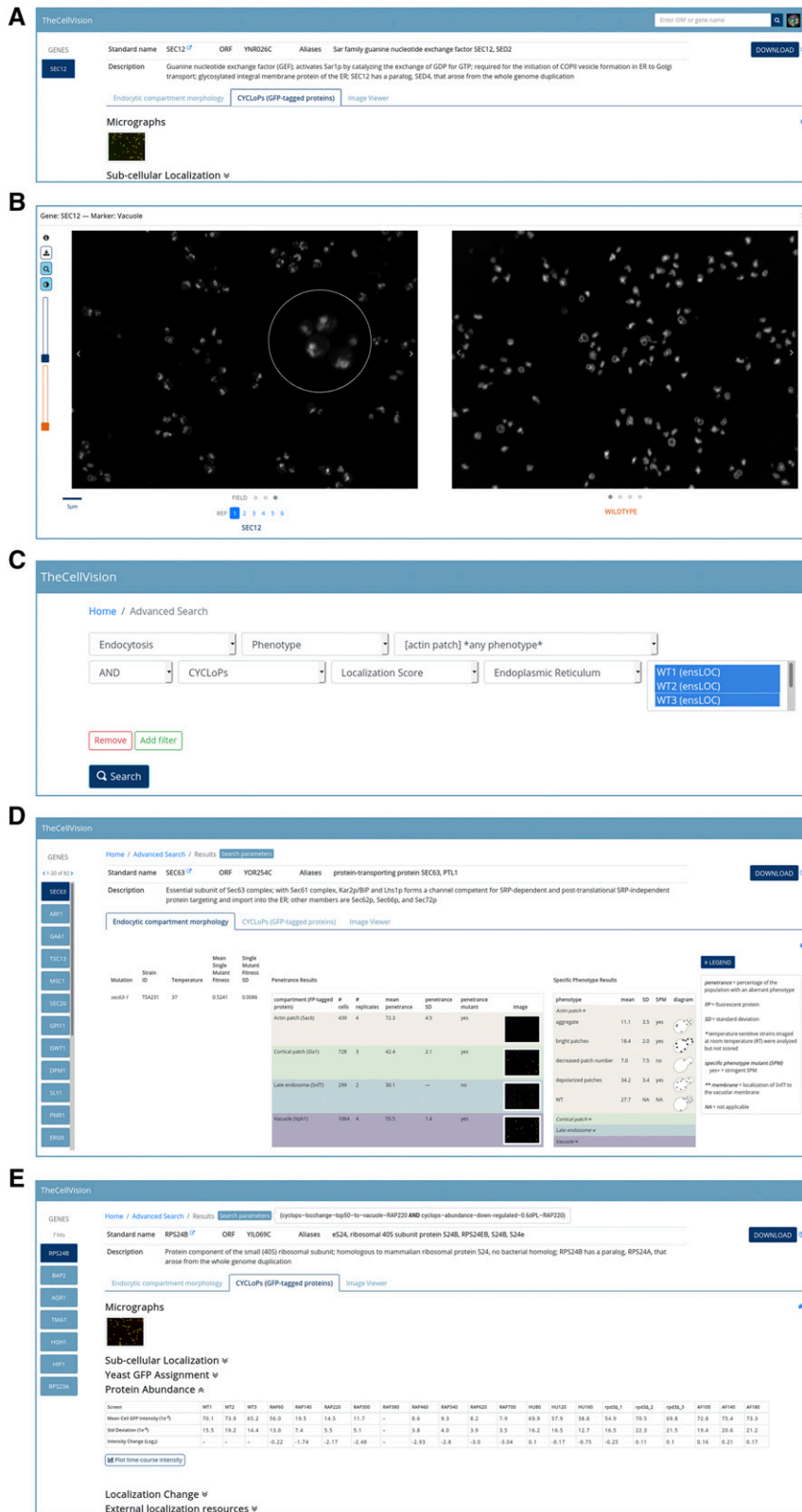


Figure 2 TheCellVision.org functionalities. A) Example results page of a simple search. B) Clicking on a thumbnail opens up an image modal with the enlarged image. Among the features available in the modal are the magnifying glass and brightness adjustment. C) The advanced search allows the user to search for more complex combinations of terms across different projects and/or screens using logical AND, OR, NOT operations. D) Results of the example advanced search from C). E) Example of advanced search results combining Localization change and Protein abundance data from the CYCLOPs dataset.

of the BBBC021 dataset and achieved ~85% classification accuracy without any hyper-parameter optimization (Table S1). We next asked if we could get better classification accuracy by removing mislabeled cells from the training set. In the dataset, all cells from a given

compound-concentration combination are assigned the same mode of action (MOA) label, but all cells do not actually display the same phenotype. We excluded from the dataset 1474 objects (~6.5%) that were classified to a different MOA class in more than 50% of a total of

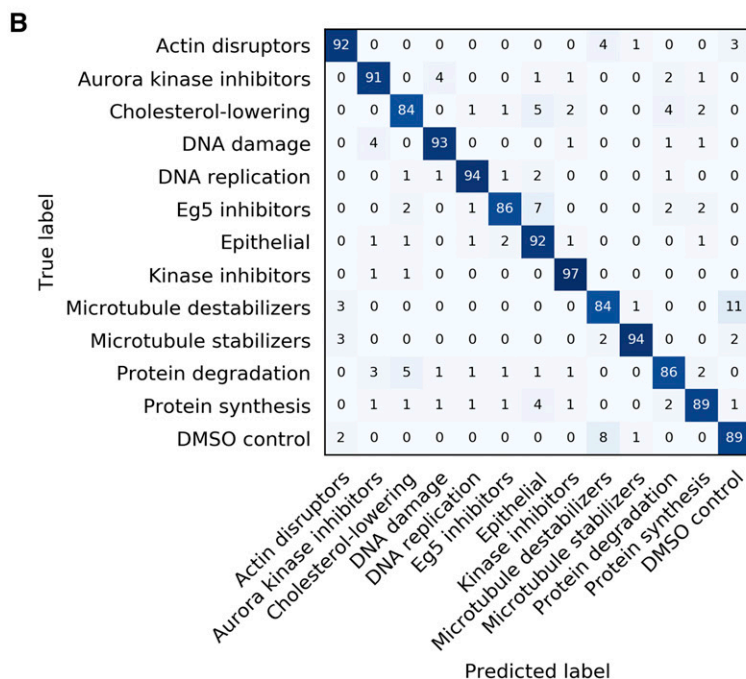
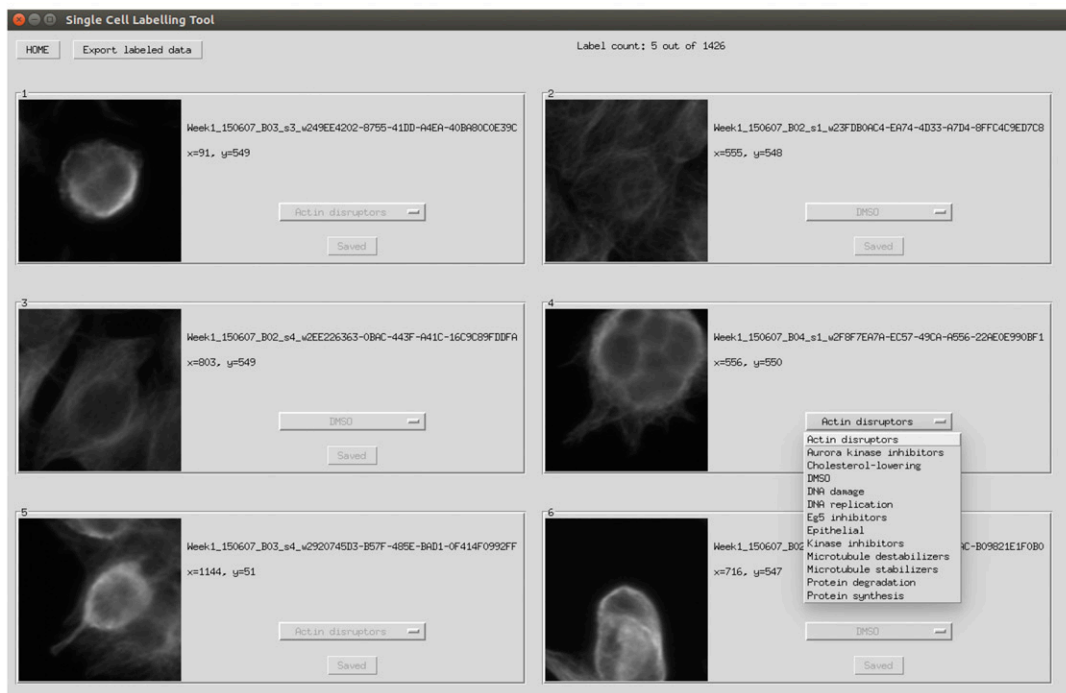
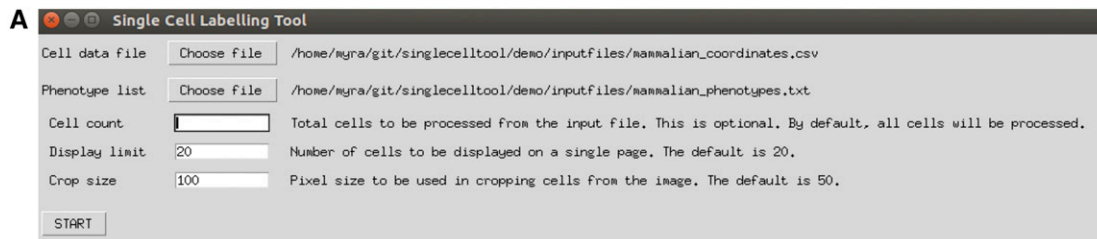


Figure 3 Image analysis tools available through TheCellVision.org. A) The single cell labeling tool offers a simple way to manually annotate single cells when building a classification training set. Top: Input GUI. Bottom: Labeling tool GUI. B) Confusion matrix illustrating the classification accuracy of the 2-hidden-layer fully connected neural network for the 13 classes of the BBBC021 dataset.

150 predictions from 3 independent runs. The classification accuracy for the cleaned training set was ~90% (Figure 3B). The final number of cells for each class in the cleaned dataset is available in Table S1. We also tested the effect of training set size on classification accuracy. Three independent runs were performed for training sets comprised of 50 to 1000 cells per class (Table S1). Similarly to the Endocytic Compartment Morphology dataset (Mattiuzzi Usaj *et al.* 2020), we observed the largest differences in classification accuracy between smaller training sets; training sets with sizes of 500 cells or more showed modest increases in classification accuracy (4% higher accuracy with training sets of 1000 cells compared to that of 500 cells).

In summary, we show that our 2-hidden-layer fully connected neural network classifier accurately classifies at least two disparate cell types, yeast and human cells (Figure 3B and Mattiuzzi Usaj *et al.* (Mattiuzzi Usaj *et al.* 2020)). We therefore predict that our single cell labeling tool and classifier will be useful for analyzing image data from different cell types.

Individual project pages: Each project has a different customized interface for browsing the analysis results and images for the searched genes. The results of individual projects can be accessed either through the general search (as described above) or through the project's home page. On each project's home page, the user can also access project-specific information and help pages ('About', 'Help'), the single-project advanced search option ('Advanced search'), and any supplemental files published with the primary paper as well as additional unpublished data ('Supplemental files'). For example, the additional files available for the Endocytic Compartment Morphology project include the classification training sets (quantitative features for all single cells used for training), phenotype profiles of the identified morphology mutants and pairwise phenotype profile similarities (Mattiuzzi Usaj *et al.* 2020).

CONCLUSIONS AND FUTURE DIRECTIONS

Each high-throughput imaging screen produces thousands of images, with gigabytes of data. TheCellVision.org provides single-platform interactive browsing of images and data generated with different high-throughput imaging screens and projects and is the first database that allows a multi-parametric search across results from different HCS projects. The current implementation houses two large-scale projects performed in budding yeast, comprising ~575,590 images, as well as the associated image analysis tools. The developed tools can be used in the analysis of different cell types (see section *Image analysis tools*), and we anticipate TheCellVision.org will provide a valuable resource to members of the community.

We continue to generate new datasets that answer diverse biological questions, and new data, including non-yeast data, will be deposited in TheCellVision.org as it is ready for release. Upcoming datasets include 'Cell Cycle Omics' - cell cycle-specific analysis of protein localization and abundance, with gene expression and translational efficiency data, 'Large-scale Phenotypic Profiling' - single-cell analysis of morphology for 18 subcellular compartments in mutant backgrounds, and 'Terminal Phenotypes' - time-course analysis of morphology for 22 subcellular compartments in temperature-sensitive mutants of yeast essential genes. In addition, we are further developing the advanced search function that allows the user to simultaneously search across different projects, and we are expanding the set of image analysis tools. In the future, we plan on adding a 'multi-omics' search option that would allow the user to search genes of interest across TheCellVision.org and its sister-site, TheCellMap.org, that houses synthetic genetic interaction data (Usaj *et al.* 2017).

The database will be a valuable tool for biologists interested in exploring different aspects of eukaryotic cell biology, and those looking for simple ready-to-use tools to analyze their image dataset, as well as for researchers developing image analysis tools in need of new benchmarking datasets.

ACKNOWLEDGMENTS

We thank Helena Friesen, Nil Sahin, Michael Costanzo and other members of the Andrews and Boone labs for helpful discussions and critical comments. This work was supported by the Canadian Institutes of Health Research (FDN-143264 and FDN143265 to CB and BA, respectively) and the National Institutes of Health (R01HG005853). High throughput imaging infrastructure was purchased with funds from the Canadian Foundation for Innovation and the Ontario Research Fund (36441).

LITERATURE CITED

- Breker, M., M. Gymrek, O. Moldavski, and M. Schuldiner, 2014 LoQATe-Localization and Quantitation ATlas of the yeast proteome. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast. *Nucleic Acids Res.* 42: D726–D730. <https://doi.org/10.1093/nar/gkt933>
- Breker, M., M. Gymrek, and M. Schuldiner, 2013 A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J. Cell Biol.* 200: 839–850. <https://doi.org/10.1083/jcb.201301120>
- Caie, P. D., R. E. Walls, A. Ingleston-Orme, S. Daya, T. Houslay *et al.*, 2010 High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol. Cancer Ther.* 9: 1913–1926. <https://doi.org/10.1158/1535-7163.MCT-09-1148>
- Carpenter, A. E., T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang *et al.*, 2006 CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 7: R100. <https://doi.org/10.1186/gb-2006-7-10-r100>
- Cherry, J. M., E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley *et al.*, 2012 Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40: D700–D705. <https://doi.org/10.1093/nar/gkr1029>
- Chong, Y. T., J. L. Koh, H. Friesen, S. K. Duffy, M. J. Cox *et al.*, 2015 Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* 161: 1413–1424. <https://doi.org/10.1016/j.cell.2015.04.051>
- Costanzo, M., B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons *et al.*, 2016 A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353: aaf1420. <https://doi.org/10.1126/science.aaf1420>
- de Groot, R., J. Luthi, H. Lindsay, R. Holtackers, and L. Pelkmans, 2018 Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol. Syst. Biol.* 14: e8064. <https://doi.org/10.15252/msb.20178064>
- Denervaud, N., J. Becker, R. Delgado-Gonzalo, P. Damay, A. S. Rajkumar *et al.*, 2013 A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proc. Natl. Acad. Sci. USA* 110: 15842–15847. <https://doi.org/10.1073/pnas.1308265110>
- Dubreuil, B., E. Sass, Y. Nadav, M. Heidenreich, J. M. Georgeson *et al.*, 2019 YeastRGB: comparing the abundance and localization of yeast proteins across cells and libraries. *Nucleic Acids Res.* 47: D1245–D1249. <https://doi.org/10.1093/nar/gky941>
- Eapen, V. V., D. P. Waterman, A. Bernard, N. Schiffmann, E. Sayas *et al.*, 2017 A pathway of targeted autophagy is induced by DNA damage in budding yeast. *Proc. Natl. Acad. Sci. USA* 114: E1158–E1167. <https://doi.org/10.1073/pnas.1614364114>
- Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418: 387–391. <https://doi.org/10.1038/nature00935>
- Heigwer, F., C. Scheeder, T. Miersch, B. Schmitt, C. Blass *et al.*, 2018 Time-resolved mapping of genetic interactions to model rewiring of signaling pathways. *eLife* 7: e41074. <https://doi.org/10.7554/eLife.40174>

- Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson *et al.*, 2003 Global analysis of protein localization in budding yeast. *Nature* 425: 686–691. <https://doi.org/10.1038/nature02026>
- Koh, J. L., Y. T. Chong, H. Friesen, A. Moses, C. Boone *et al.*, 2015 CYCLOPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 5: 1223–1232 (Bethesda). <https://doi.org/10.1534/g3.115.017830>
- Kraus, O. Z., B. T. Grys, J. Ba, Y. Chong, B. J. Frey *et al.*, 2017 Automated analysis of high-content microscopy data with deep learning. *Mol. Syst. Biol.* 13: 924. <https://doi.org/10.15252/msb.20177551>
- Li, Z., F. J. Vizeacoumar, S. Bahr, J. Li, J. Warringer *et al.*, 2011 Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat. Biotechnol.* 29: 361–367. <https://doi.org/10.1038/nbt.1832>
- Ljosa, V., K. L. Sokolnicki, and A. E. Carpenter, 2012 Annotated high-throughput microscopy image sets for validation. *Nat. Methods* 9: 637. <https://doi.org/10.1038/nmeth.2083>
- Loewith, R., and M. N. Hall, 2011 Target of rapamycin (TOR) in nutrient signaling and growth control. *Genetics* 189: 1177–1201. <https://doi.org/10.1534/genetics.111.133363>
- Mattiazzi Usaj, M., M. Brloznic, P. Kaferle, M. Zitnik, H. Wolinski *et al.*, 2015 Genome-Wide Localization Study of Yeast Pex11 Identifies Peroxisome-Mitochondria Interactions through the ERMES Complex. *J. Mol. Biol.* 427: 2072–2087. <https://doi.org/10.1016/j.jmb.2015.03.004>
- Mattiazzi Usaj, M., N. Sahin, H. Friesen, C. Pons, M. Usaj *et al.*, 2020 Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. *Mol. Syst. Biol.* 16: e9243. <https://doi.org/10.15252/msb.20199243>
- Mattiazzi Usaj, M., E. B. Styles, A. J. Verster, H. Friesen, C. Boone *et al.*, 2016 High-Content Screening for Quantitative Cell Biology. *Trends Cell Biol.* 26: 598–611. <https://doi.org/10.1016/j.tcb.2016.03.008>
- Neumuller, R. A., T. Gross, A. A. Samsonova, A. Vinayagam, M. Buckner *et al.*, 2013 Conserved regulators of nucleolar size revealed by global phenotypic analyses. *Sci. Signal.* 6: ra70. <https://doi.org/10.1126/scisignal.2004145>
- Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani *et al.*, 2005 High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. USA* 102: 19015–19020. <https://doi.org/10.1073/pnas.0509436102>
- Orloff, D. N., J. H. Iwasa, M. E. Martone, M. H. Ellisman, and C. M. Kane, 2012 The cell: an image library-CCDB: a curated repository of microscopy data. *Nucleic Acids Res.* 41: D1241–D1250. <https://doi.org/10.1093/nar/gks1257>
- Sauerwald, J., T. Jores, M. Eisenberg-Bord, S. G. Chuartzman, M. Schuldiner *et al.*, 2015 Genome-Wide Screens in *Saccharomyces cerevisiae* Highlight a Role for Cardiolipin in Biogenesis of Mitochondrial Outer Membrane Multispan Proteins. *Mol. Cell. Biol.* 35: 3200–3211.
- Smith, K., F. Piccinini, T. Balassa, K. Koos, T. Danka *et al.*, 2018 Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays. *Cell Syst.* 6: 636–653. <https://doi.org/10.1016/j.cels.2018.06.001>
- Styles, E.B., K.J. Founk, L.A. Zamparo, T.L. Sing, D. Altintas *et al.*, 2016 Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci. *Cell Syst* 3: 264–277.e10. <https://doi.org/10.1016/j.cels.2016.08.008>
- Tkach, J. M., A. Yimit, A. Y. Lee, M. Riffle, M. Costanzo *et al.*, 2012 Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat. Cell Biol.* 14: 966–976. <https://doi.org/10.1038/ncb2549>
- Usaj, M., Y. Tan, W. Wang, B. VanderSluis, A. Zou *et al.*, 2017 TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 (Bethesda)* 7: 1539–1549 (Bethesda). <https://doi.org/10.1534/g3.117.040220>
- Williams, E., J. Moore, S. W. Li, G. Rustici, A. Tarkowska *et al.*, 2017 The Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* 14: 775–781. <https://doi.org/10.1038/nmeth.4326>
- Wolinski, H., U. Petrovic, M. Mattiazzi, J. Petschnigg, B. Heise *et al.*, 2009 Imaging-based live cell yeast screen identifies novel factors involved in peroxisome assembly. *J. Proteome Res.* 8: 20–27. <https://doi.org/10.1021/pr800782n>
- Yin, Z., A. Sadok, H. Sailem, A. McCarthy, X. Xia *et al.*, 2013 A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. *Nat. Cell Biol.* 15: 860–871. <https://doi.org/10.1038/ncb2764>
- Zhao, L., Q. Yang, J. Zheng, X. Zhu, X. Hao *et al.*, 2016 A genome-wide imaging-based screening to identify genes involved in synphilin-1 inclusion formation in *Saccharomyces cerevisiae*. *Sci. Rep.* 6: 30134. <https://doi.org/10.1038/srep30134>

Communicating editor: M. Tyers