



Expanding TheCellVision.org: a central repository for visualizing and mining high-content cell imaging projects

Myra Paz David Masinas,^{1,†} Athanasios Litsios,^{1,†} Anastasia Razdaibiedina,^{1,2,3} Matej Usaj,¹ Charles Boone ,^{1,2,*} Brenda J. Andrews ^{1,2,*}

¹The Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada

²Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada

³Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1N1, Canada

*Corresponding authors: The Donnelly Centre, Department of Molecular Genetics, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. Email: brenda.andrews@utoronto.ca; Email: charlie.boone@utoronto.ca

[†]These authors contributed equally to this work.

We previously constructed [TheCellVision.org](https://thecellvision.org), a central repository for visualizing and mining data from yeast high-content imaging projects. At its inception, [TheCellVision.org](https://thecellvision.org) housed two high-content screening (HCS) projects providing genome-scale protein abundance and localization information for the budding yeast *Saccharomyces cerevisiae*, as well as a comprehensive analysis of the morphology of its endocytic compartments upon systematic genetic perturbation of each yeast gene. Here, we report on the expansion of [TheCellVision.org](https://thecellvision.org) by the addition of two new HCS projects and the incorporation of new global functionalities. Specifically, [TheCellVision.org](https://thecellvision.org) now hosts images from the Cell Cycle Omics project, which describes genome-scale cell cycle-resolved dynamics in protein localization, protein concentration, gene expression, and translational efficiency in budding yeast. Moreover, it hosts PIFiA, a computational tool for image-based predictions of protein functional annotations. Across all its projects, [TheCellVision.org](https://thecellvision.org) now houses >800,000 microscopy images along with computational tools for exploring both the images and their associated datasets. Together with the newly added global functionalities, which include the ability to query genes in any of the hosted projects using either yeast or human gene names, [TheCellVision.org](https://thecellvision.org) provides an expanding resource for single-cell eukaryotic biology.

Keywords: TheCellVision.org; cell vision; high-content screening; quantitative image analysis; subcellular morphology; subcellular localization; protein abundance; Cell Cycle Omics; PIFiA; image analysis tools

Introduction

Single-cell phenomics, which entails the highthroughput (HTP) acquisition of multiparametric phenotypic information from individual cells, has emerged as a powerful approach for dissecting the genotype-to-phenotype relationship and decoding the system-level functioning of organisms. Single-cell phenotypic information in phenomics is typically acquired via high-content screening (HCS), which combines HTP microscopy with computational data analysis. This procedure results in the generation of large numbers of cell images (often in the range of millions) accompanied with vast datasets representing the extracted phenotypic features. Together with the logistical challenges associated with handling and analyzing this information, comes the challenge of making the data accessible in biologically meaningful ways to end-users and intuitively reusable for subsequent analyses by other researchers. To address these issues, we previously developed [TheCellVision.org](https://thecellvision.org) (Masinas et al. 2020), a centralized database and website for hosting datasets associated with yeast HCS projects in an easily interpretable and accessible way.

At its inception, [TheCellVision.org](https://thecellvision.org) housed two large-scale datasets whose generation was facilitated by the yeast GFP-collection (Huh et al. 2003); an array of strains expressing GFP-fusion proteins which enables proteome-scale identification of protein abundance and localization. Specifically, [TheCellVision.org](https://thecellvision.org) houses CYCLOPs (Collection of Yeast Cell and Localization Patterns (Koh et al. 2015)), which consists of ~390,000 images from two HCS studies from our lab (Chong et al. 2015; Kraus et al. 2017), and their analyses describing proteome-scale localization patterns with resolution of ~15 localization classes in reference conditions as well as during different genetic or chemical perturbations. Endocytic Compartment Morphology, the second large-scale study housed in [TheCellVision.org](https://thecellvision.org), consists of ~190,000 images depicting the morphology of yeast endocytic compartments upon genome-scale genetic perturbations (Mattiuzzi Usaj et al. 2020). [TheCellVision.org](https://thecellvision.org) has more than 11,000 unique users worldwide since its launching date in 2020 (Fig. 1), highlighting the utility of such HCS resources and their accessible dissemination.

We updated [TheCellVision.org](https://thecellvision.org) to include data from two additional large-scale HCS screens that we recently generated

Received on 16 January 2024; accepted on 13 March 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Sessions

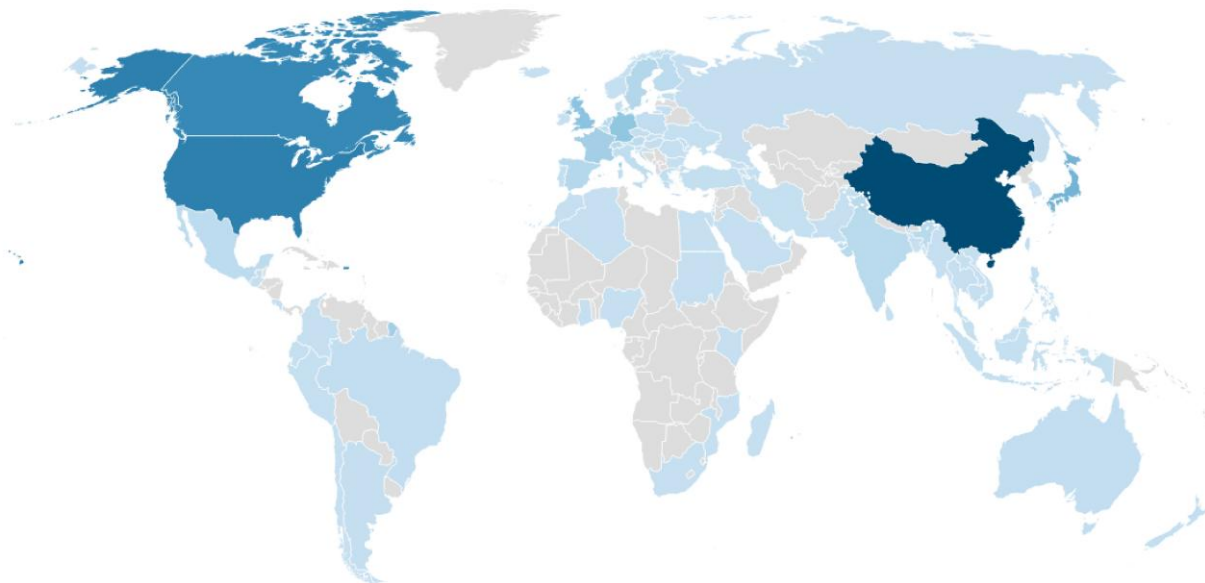


Fig. 1. Geographical distribution of [TheCellVision.org](https://www.thecellvision.org) sessions. Number of [TheCellVision.org](https://www.thecellvision.org) active sessions per country from 2020 November 1, until 2023 December 5. Heatmap was generated using Datawrapper (<https://www.datawrapper.de>) and data were obtained from Google Analytics.

(Litsios et al. 2024; Razdaibiedina et al. 2024), which complement and extend the biological scope of the currently housed datasets. Specifically, [TheCellVision.org](https://www.thecellvision.org) now houses Cell Cycle Omics, a resource that describes the first genome-scale annotation of protein localization changes during the cell cycle in any organism (Litsios et al. 2024). This resource presents in an integrative way cell cycle-resolved yeast multiomics, including protein localization, protein concentration, gene transcription, and translational efficiency measurements. Moreover, [TheCellVision.org](https://www.thecellvision.org) now also houses PIFiA (Protein Image-based Functional Annotation), a self-supervised machine-learning pipeline that can be used for protein functional annotation prediction based on features extracted from single-cell imaging data (Razdaibiedina et al. 2024). Finally, to enhance the cross-species utility of both the existing and new resources, [TheCellVision.org](https://www.thecellvision.org) now has a series of new global functionalities that facilitate its use by researchers with no familiarity with yeast biology.

Materials and methods

Highthroughput imaging screens

At present, [TheCellVision.org](https://www.thecellvision.org) houses four large-scale projects, two existing since the inception of [TheCellVision.org](https://www.thecellvision.org), and two new ones (Fig. 2):

- i) CYCLOPs (Collection of Yeast Cell and Localization Patterns) (Chong et al. 2015; Kraus et al. 2017).
- ii) Endocytic Compartment Morphology (Mattiuzzi Usaj et al. 2020).
- iii) Cell Cycle Omics. This is a newly added project that describes genome-scale protein concentration and localization dynamics during the yeast cell cycle based on protein-GFP fusions (Litsios et al. 2024). Cell cycle-resolution has been achieved via in silico synchronization of single cells expressing fluorescent cell cycle markers from static

microscopy images. The proteome data are complemented with population-level, cell cycle-resolved gene expression and translational efficiency data.

- iv) PIFiA. This is another newly added project which provides a tool for predictions of protein function, based on genome-scale comparative analysis of extracted features of images of single cells endogenously expressing each protein fused to GFP (Razdaibiedina et al. 2024). Along with a fully interactive t-SNE map, PIFiA allows users to explore micrographs associated with each GFP-tagged protein and perform enrichment analysis based on filtered nearest neighbors.

Databased development and schema

We use PostgreSQL, a relational database management system, to store and organize the data in [TheCellVision.org](https://www.thecellvision.org). The database schema has been expanded to include two new main clusters—one for each of the new research projects: Cell Cycle Omics and PIFiA (Supplementary Fig. 1). We have also updated the core cluster and created a table for Human Orthologs wherein we have imported 6,005 unique entries associated with our yeast gene collection.

Results and discussion

New projects available on [TheCellVision.org](https://www.thecellvision.org) Cell Cycle Omics

Cell Cycle Omics contains genome-scale data for the cell cycle-resolved protein concentration, protein localization, gene expression, and translational efficiency dynamics in budding yeast. The data include 129,525 microscopy images of >20 million live cells (Litsios et al. 2024) expressing proteins of interest fused with endogenous C-terminal GFP tags (Huh et al. 2003). Cells were imaged using HTP microscopy, and proteome localization and concentration were quantified using DeepLoc, a deep

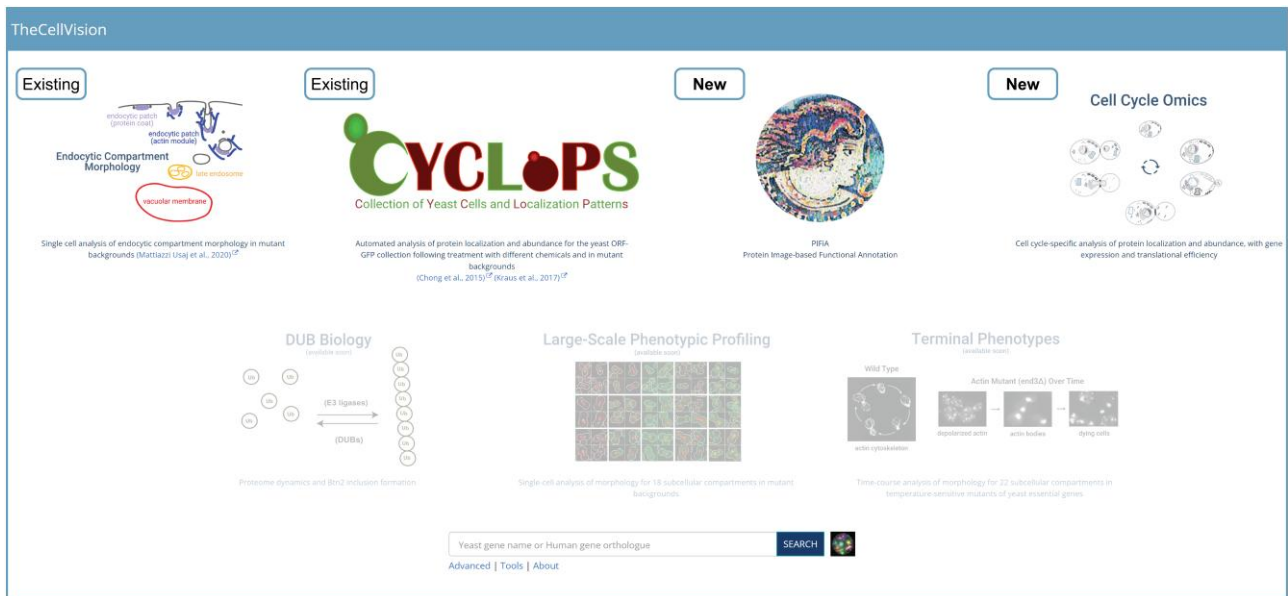


Fig. 2. HCS projects housed in [TheCellVision.org](https://thecellvision.org). Snapshot of [TheCellVision.org](https://thecellvision.org)'s landing page containing an overview of hosted projects. Existing and newly added projects are annotated as such. Three projects presented with reduced opacity represent additional projects to be hosted on [TheCellVision.org](https://thecellvision.org) soon.

convolutional neural network (CNN) for automated classification of proteins to different subcellular compartments (Kraus *et al.* 2017). Cell cycle-resolution was achieved using CycleNet, a newly developed CNN for in silico synchronization of single cells from microscopy images of unperturbed cell populations (Litsios *et al.* 2024). Cell cycle-resolved gene expression and translational efficiency measurements were obtained at the population level using RNA sequencing (Couvillion and Churchman 2017) and Ribosome Profiling (Ingolia 2010), after synchronization of cells with the α -factor mating hormone (Breedon 1997). All omics datasets were scored using the same statistical metrics to identify proteins/transcripts with cell cycle-periodic regulation.

Upon selecting the Cell Cycle Omics project at the landing page of [TheCellVision.org](https://thecellvision.org), the user is prompted to enter a gene/protein name of interest into the query box (Fig. 3a). Either yeast or human gene names can be used, which is a new global functionality of [TheCellVision.org](https://thecellvision.org) described further below. The user is then navigated to the relevant results page, where information about the gene name is displayed, along with a brief description of its function (Fig. 3b). At the upper right side of this page, the user can select to download all data associated with the query gene in the Cell Cycle Omics project ("DOWNLOAD" option). Underneath the description, the user can select to open the cell viewer, where microscopy images of single cells are displayed across all available biological replicates (Fig. 3c). Here, the protein encoded by the gene of interest is displayed in green (GFP channel), and a nuclear and cytoplasmic marker (same across all strains) in red (RFP channel) and blue (FarRed channel), respectively. The overlay of all channels is displayed by default, but the user can select to view specific channels individually. Moreover, the image's brightness is adjustable and there is an option to magnify regions of interest within the image. Next to the cell viewer, the multiomics cell cycle profile of each gene is shown, with the option to select and enlarge it (Fig. 3d). Underneath the cell viewer and the multiomics profile, the actual cell cycle data are presented. These include protein localization dynamics with resolution of 22 localization classes and protein concentration dynamics, all across six cell cycle phases

(G1 Pre-START, G1 Post-START, S/G2, Metaphase, Anaphase, Telophase) and for three biological replicates (Fig. 3e). This is followed by cell cycle-resolved gene expression and translational efficiency data for five cell cycle phases (G1 Post-START, S/G2, Metaphase, Anaphase, Telophase) and two biological replicates (Fig. 3f). Finally, information about whether the gene/protein was scored as cell cycle-periodic upon our statistical analysis is presented for each of the omics datasets (Fig. 3e).

Protein Image-based Functional Annotation

The PIFiA computational pipeline offers a tool for predictions of protein function based on genome-scale comparative analysis of features extracted from images of single cells endogenously expressing GFP-fusion proteins (Razdaibiedina *et al.* 2024). It relies on a self-supervised learning approach which was applied to 3,058,961 single-cell images acquired via confocal microscopy, from strains of the yeast GFP-collection (Huh *et al.* 2003) into which an additional nuclear and a cytosolic marker were introduced using SGA technology (Tong *et al.* 2001). The PIFiA workflow consists of a feature extraction step performed by a deep neural network and subsequent analysis steps on the extracted feature profiles. The downstream analysis enables prediction of protein localization, identification of functional modules or subsets of proteins with related cellular roles (e.g. members of the same protein complex), and prediction of protein function.

The main page for PIFiA on the CellVision website contains an interactive visualization of the extracted feature profiles for each protein in a two-dimensional space using t-SNE (van der Maaten and Hinton 2008), where each protein (node) is colored according to its subcellular localization prediction (Fig. 4a). On the right of the t-SNE projection is a list of clickable protein complexes which can be toggled on/off to highlight their members on the t-SNE map. The user can select their gene of interest by either typing the yeast or human gene name in the search bar or by choosing a protein node on the map itself. Upon selecting their gene of interest, the respective protein is highlighted in the t-SNE map

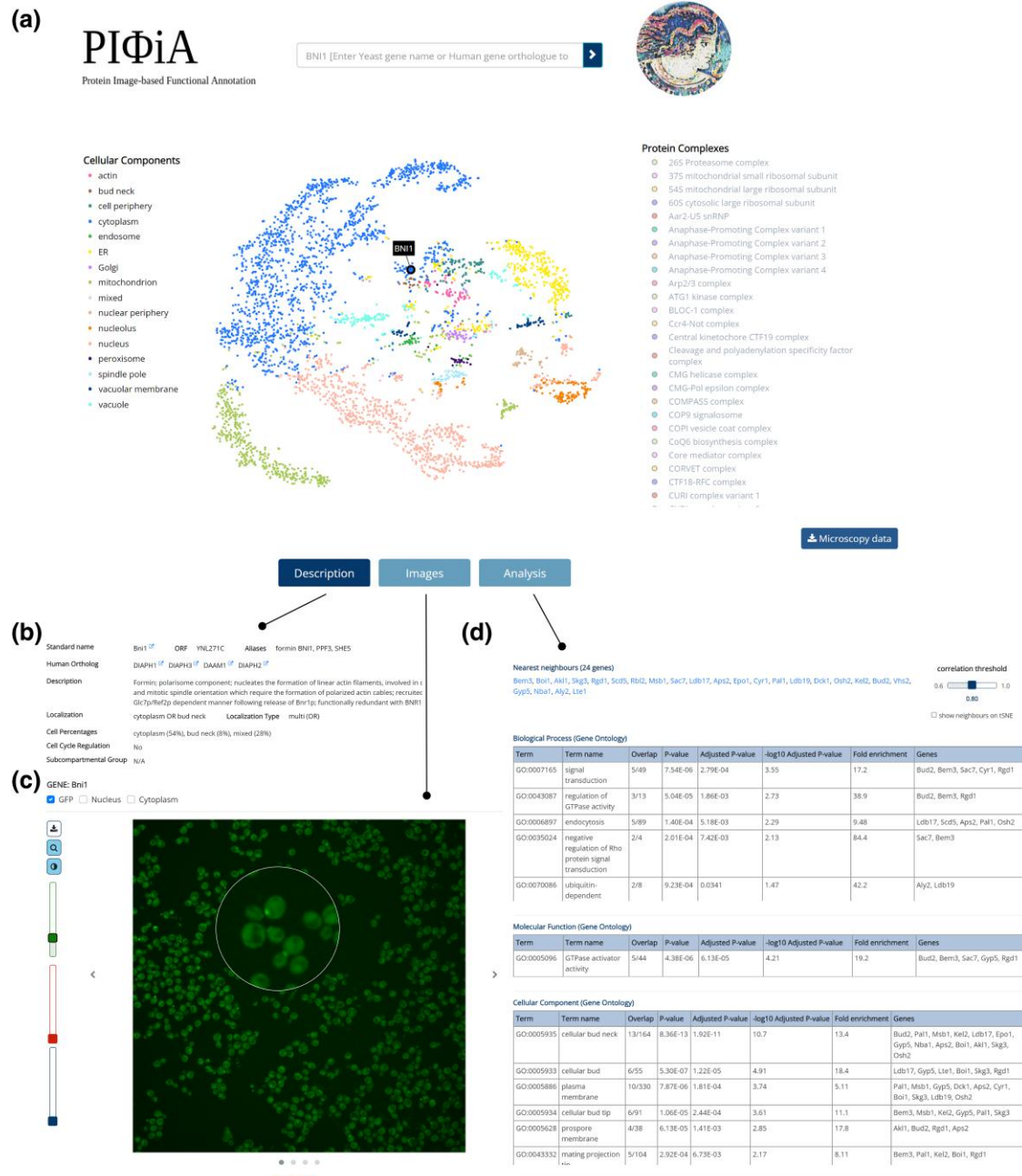


Fig. 4. Layout overview of PIFiA results page. a–d) Snapshots of various parts of the results section after the search for an example gene (BNI1 in this case). a) Interactive T-SNE map generated from PIFiA’s feature profiles, with indication of BNI1’s position. Color-coding denotes subcellular localization. Protein complexes that can be selected to display their members in the t-SNE map are shown on the right. At the bottom, there is the option to view b) various types of descriptive information related to the queried protein, c) associated microscopy images, and d) other neighboring proteins in the t-SNE map (proteins with similar feature profiles with the queried protein) along with their functional enrichments.

browse through the rest of the associated image set. Moreover, other functionalities included are the option to select which channel(s) to view, download the image set, magnify regions of interest within the image and adjust image brightness. Lastly, the “Analysis” tab (Fig. 4d) is used to obtain more precise predictions for the function of the queried protein, based on PIFiA’s image-based self-supervised learning approach. We use a “guilt-by-association” approach in which the function of the queried protein is determined by the function of its neighboring proteins (in the feature space). This tab shows a list of proteins with similar feature profiles to the protein of interest, along with their Gene

Ontology annotations and enrichment scores. The nearest neighbors list can be adjusted by changing the value of the similarity threshold (correlation is used as a similarity metric). A checkbox to mark the neighbors on the t-SNE map is also available.

New global functionalities of TheCellVision.org

Given the importance of yeast as a model for eukaryotic cell biology, we added a series of new global functionalities (available for all hosted projects) in TheCellVision.org, to facilitate its utilization for cross-species research. Specifically, we have added (i) a new results subsection where the human orthologs are displayed next to

the respective yeast gene, (ii) options to select the displayed human orthologs and be automatically navigated to <https://www.alliancegenome.org/> for retrieval of relevant gene information, and (iii) searchability of the website directly via the use of human gene names, so that the resource can be easily used by scientists that have no familiarity with yeast genes.

Conclusions and future directions

The continuous advances in wet-lab and computational approaches for the generation of image-based omics datasets and the HTP phenotypic profiling of single cells, have created a need to accessibly store and disseminate the resultant datasets. [TheCellVision.org](https://www.thecellvision.org) serves as a central repository for fluorescence microscopy images and associated quantitative data generated using high-content screening of budding yeast images. Moreover, it provides the tools for meaningfully exploring the data and directly accessing them for independent analyses. The four large-scale projects currently hosted at [TheCellVision.org](https://www.thecellvision.org) include genome-scale information about genetically, chemically, and cell cycle-induced protein localization and concentration dynamics, as well as a tool for guiding protein functional annotation, and a comprehensive characterization of the endocytic compartment morphology caused by single gene perturbations. All these resources are now searchable directly via human gene names, facilitating their use for cross-species research.

[TheCellVision.org](https://www.thecellvision.org) will be soon supplemented with three additional large-scale phenotypic screens, which include a comprehensive morphological analysis of 18 subcellular yeast compartments upon genome-wide perturbations, a time-course analysis of the morphology of 22 subcellular compartments in temperature-sensitive mutants of yeast essential genes, and a screen for protein concentration and localization changes in response to perturbation of yeast deubiquitinating enzymes. In the future, [TheCellVision.org](https://www.thecellvision.org) is intended to support a global Advanced Search function pertaining to all hosted projects, which will harness the diversity and complementarity of these projects to provide multidimensional biological information for each search query. The comprehensiveness, diversity, and accessibility of the datasets hosted in [TheCellVision.org](https://www.thecellvision.org), provide a resource for eukaryotic cell biology that can be readily used for easy exploration of new hypotheses, independent data reanalyses, and dataset benchmarking.

Acknowledgments

We thank Helena Friesen for helpful comments and discussions.

Data availability

Microscopy images, datasets, and computational tools associated with the previous projects are available as described in [Masinas et al. \(2020\)](#). Raw images and segmented single-cell crops associated with PIFiA are available for download at https://thecellvision.org/pifia_files. Bulk download of images from Cell Cycle Omics can be provided upon request. Source code and usage examples for PIFiA are available at <https://github.com/arazd/PIFiA>. The code for CycleNET, the supervised neural network used in Cell Cycle Omics for protein localization and cell cycle classification is available at <https://github.com/BooneAndrewsLab/CycleNET>.

Supplemental material

Supplemental material available at GENETICS online.

Funding

A.L. holds a Canadian Institutes of Health Research Fellowship (MFE - 187913). This work was supported by grants from the National Institutes of Health (R01HG005853 to B.A., C.B.), and the Canadian Institutes of Health Research (PJT-180259 to B.A.).

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

- Breeden LL. 1997. Alpha-factor synchronization of budding yeast. *Methods Enzymol.* 283:332–341. doi:10.1016/s0076-6879(97)83027-3.
- Chong YT, Koh JLY, Friesen H, Duffy K, Cox MJ, Moses A, Moffat J, Boone C, Andrews BJ. 2015. Yeast proteome dynamics from single cell imaging and automated analysis. *Cell.* 161(6):1413–1424. doi:10.1016/j.cell.2015.04.051.
- Couvillion MT, Churchman LS. 2017. Mitochondrial ribosome (mitoribosome) profiling for monitoring mitochondrial translation in vivo. *Curr Protoc Mol Biol.* 119(1):4.28.1–4.28.25. doi:10.1002/cpmb.41.
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. *Nature.* 425(6959):686–691. doi:10.1038/nature02026.
- Ingolia NT. 2010. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol.* 470:119–142. doi:10.1016/S0076-6879(10)70006-9.
- Koh JLY, Chong YT, Friesen H, Moses A, Boone C, Andrews BJ, Moffat J. 2015. CYCLOps: a comprehensive database constructed from automated analysis of protein abundance and subcellular localization patterns in *Saccharomyces cerevisiae*. *G3 (Bethesda).* 5(6):1223–1232. doi:10.1534/g3.115.017830.
- Kraus OZ, Grys BT, Ba J, Chong Y, Frey BJ, Boone C, Andrews BJ. 2017. Automated analysis of high-content microscopy data with deep learning. *Mol Syst Biol.* 13(4):924. doi:10.15252/msb.20177551.
- Litsios A, Grys B, Kraus O, Friesen H, Ross C, Masinas MPD, Forster DT, Couvillion MT, Timmermann S, Billmann M, et al. 2024. Proteome-scale movements and compartment connectivity during the eukaryotic cell cycle. *Cell.* 187(6):1490–1507. doi:10.1016/j.cell.2024.02.014.2024.
- Masinas MPD, Usaj MM, Usaj M, Boone C, Andrews BJ. 2020. [TheCellVision.org](https://www.thecellvision.org): a database for visualizing and mining high-content cell imaging projects. *G3 (Bethesda).* 10(11):39693976. doi:10.1534/g3.120.401570.
- Mattiazzi Usaj M, Sahin N, Friesen H, Pons C, Usaj M, Masinas MPD, Shuteriqi E, Shkurin A, Aloy P, Morris Q, et al. 2020. Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. *Mol Syst Biol.* 16(2):e9243. doi:10.15252/msb.20199243.
- Razdaibiedina A, Brechalov A, Friesen H, Usaj MM, Masinas MPD, Garadi SH, Wang K, Boone C, Ba J, Andrews B. 2024. PIFiA: self-supervised approach for protein functional annotation from single-cell imaging data. *Mol Syst Biol.* 1–28. doi:10.1038/s44320-024-00029-6.
- Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghizadeh S, Hogue CWV, Bussey H, et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 294:2364–2368. doi:10.1126/science.1065810.
- Van Der Maaten L, Hinton G. 2008 Visualizing data using t-SNE. *J Mach Learn Res* 9: 2579–2605. <https://jmlr.org/papers/v9/vandermaaten08a.html>