

Perspective

Towards ‘end-to-end’ analysis and understanding of biological timecourse data

 Siddhartha G. Jena^{1,*},  Alexander G. Goglia^{2,*} and  Barbara E. Engelhardt^{3,4}

¹Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, U.S.A.; ²Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, U.S.A.; ³Department of Computer Science, Princeton University, Princeton, New Jersey, U.S.A.; ⁴Gladstone Institutes, San Francisco, U.S.A.

Correspondence: Barbara E. Engelhardt (barbara.engelhardt@gladstone.ucsf.edu)



Petabytes of increasingly complex and multidimensional live cell and tissue imaging data are generated every year. These videos hold large promise for understanding biology at a deep and fundamental level, as they capture single-cell and multicellular events occurring over time and space. However, the current modalities for analysis and mining of these data are scattered and user-specific, preventing more unified analyses from being performed over different datasets and obscuring possible scientific insights. Here, we propose a unified pipeline for storage, segmentation, analysis, and statistical parameterization of live cell imaging datasets.

Introduction

Cellular behavior is, by definition, dynamic. Cells divide, move, and proliferate, carving out physical trajectories in 2- or 3-dimensional space [1]. They also activate and deactivate signaling pathways [2], and produce RNA and proteins, over both short and long time scales [3]. Much of modern biology has focused on developing techniques to measure these processes; however, traditional fixed-cell measurements impose limitations on retrieving potentially important information.

A great deal of relevant biological information is lost when phenomena are observed with techniques that measure static snapshots of cellular processes without actively tracking them over time and space. Many early fixed-cell methods, such as western or northern blotting [4, 5] or immunostaining [6], leverage specific antibody or oligonucleotide probes to measure the abundance of transcribed RNA, translated protein, or post-translational modifications on proteins that underlie many cellular behaviors. However, such bulk-cell measurements erase information about cell-to-cell heterogeneity that can impact signaling, transcription, or translation between cells. Immunostaining and subsequent imaging can reveal differences between individual cells, but erases information about their movement and spatial dynamics within their niche.

Fixed-cell methods also only capture a single snapshot of cellular behavior, which may represent an incomplete picture of a biological process: for instance, at any given point in time, a pathway of interest may be activated in only a subset of seemingly homogeneous, clonal cells [7, 8]. While improved quantification of pathway activity at the single-cell level has helped attribute a component of this disparity to noise in pathway activation [9], we know now that at least some fixed-snapshot heterogeneity in signaling is produced not by long-lived signaling states but rather by asynchronous temporal signaling dynamics (e.g. frequency, duration) [10, 11]. For instance, screening kinase inhibitors for their effects on single cell Erk activity revealed differences in signaling dynamics — and how they impact cell fate — that could not have been inferred solely from fixed-cell measurements [12]. These dynamics are often themselves heterogeneous between cells, particularly in contexts like the emergence of rare cancer cells where dynamics may eventually help predict and define new cell types [13]. Heterogeneity within populations limits the application of common heuristics like estimating the fraction of time spent in the active state based on the active fraction of a fixed population. On the other hand, live-cell imaging allows for precise measurement of dynamic signaling properties in single cells

*These authors contributed equally to this work.

Received: 3 February 2022

Revised: 19 May 2022

Accepted: 20 May 2022

Version of Record published:

17 June 2022

without the need for such indirect estimates. A recent example of this involved the use of live-cell Erk biosensors combined with oncogene expression, revealing that differential Erk dynamics led to the opposing cell fates of cell cycle arrest or proliferation [13].

Live-cell imaging, which uses brightfield imaging or fluorescent reporters to track and measure behavior in living cells over time, offers a possible solution to the challenges of capturing system dynamics at single-cell resolution. Beginning with the discovery and isolation of green fluorescent protein (GFP) [14] and its subsequent chimerization with proteins to visualize their location and activity, thousands of reporters of protein activity, transcription, post-translational modifications, and even intracellular pH or calcium levels [15–17] have been deployed in both cell culture models and *in vivo* (for those interested, an extremely comprehensive list can be found at <https://biosensordb.ucsd.edu/>). Additionally, advancements in imaging resolution, microscope stage-top incubators, and data storage [18–20] have allowed for long-term measurement of cell motility [21, 22], morphology [23, 24], and intracellular activity [25–30] over extended periods of time or at high time resolution, allowing for quantitative statements to be made about dynamic processes, heterogeneity, and information processing in and between cells.

As these imaging modalities have grown more and more common, petabytes of imaging data have been generated; yet, often a single readout or a small subset of cells are used for subsequent analysis and publication, such as a single measurement of GFP across time, leaving many potential findings about cellular phenotypes and dynamics unexplored. In this sense, the live-cell imaging community lags substantially behind the genomics and transcriptomics communities in terms of accessible and easy-to-use deposition of data, and centralized analysis of those data. The current lack of resources for live-cell imaging is largely technical, as there is no standardization between the many live-cell imaging methods and file formats used to collect these data. Even after the data have been collected, researchers looking to conduct similar analyses or share their analysis pipelines are at the whim of journal-to-journal data- and code-sharing practices, which tend to encourage but not enforce data and code sharing and often say nothing about ease of access. This process can be cumbersome and in fact has led most quantitative microscopists to host their analysis pipelines on decentralized sites such as Github with research-quality code not able to generalize beyond the lab-specific data. Rarely are the data made public.

Fortunately, researchers have broadly outlined the phenotypic and dynamics information that they would like to have from live-cell imaging data. While these phenotypes depend heavily on the types of analyses envisioned, the list is both broad and adaptable. The important phenotypes include cellular morphology, population averages, single-cell trajectories, spatial information, and lineage tracing of cell division; more complicated phenotypes related to cellular dynamics include rates and motility, time averages, and pulse detection [31, 32]. Here, we outline and describe a potentially revolutionary vision for the field: a centralized repository of live-cell imaging data coupled with data analysis tools that encompass the range of readouts and downstream analyses imagined by scientists. In particular, we describe the possible uses of these data paired with data processing and analysis tools to develop a broader and deeper understanding of biochemical pathways in cells, as well as local interactions and cellular dynamics in tissues.

The resource

We propose the development of a publicly available and accessible repository for live-cell imaging datasets, and the construction of an accompanying end-to-end workflow that allows scientists to begin with imaging data and an intended set of readouts in mind, and obtain a well-founded and descriptive set of statistics on phenotypes within the data (Figure 1). This system takes as input live-cell imaging data across different experimental conditions (Figure 1a), processes them into a standardized format (Figure 1b), models and controls for biological and experimental confounding, and quantifies both static and dynamic cellular phenotypes (Figure 1c). Importantly, the workflow will include built-in methods for dimension reduction, clustering, and visualization of these experiments, differential analyses across specific conditions, and association tests using metadata (Figure 1d). This allows for both low-dimensional (e.g. 2D principal component analysis, highlighting major sources of statistical differences between two or more samples, shown as an example in the upper panel of Figure 1d) or higher-dimensional data representation (e.g. 3D scatter plots of multiple dimensions quantified from time-course data, such as cell size, movement speed, or signaling activity shown as an example in the lower panel of Figure 1d).

The first step of this process involves the development of a public-facing data repository for live-cell imaging data. To do this, we are developing infrastructure for data and software pipeline servers at Gladstone Institutes.

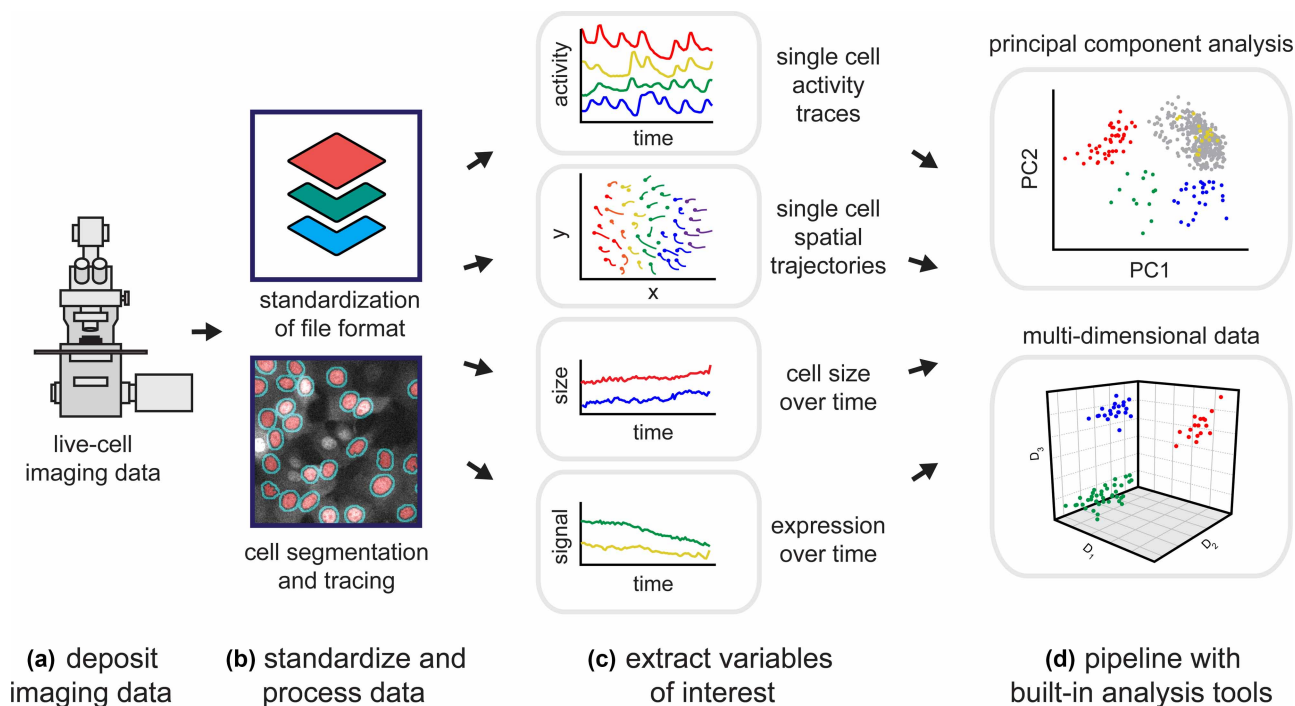


Figure 1. A centralized repository for live-cell imaging data and analysis.

The proposed repository would enable researchers to (a) access deposited imaging data from previous studies in (b) standardized formats, subjected to standardized processing (e.g. segmentation and tracing), with built-in tools to (c) extract and (d) analyze variables of interest.

Approximating the space required, we assume that a multiplexed imaging experiment consisting of 1000 time-points imaged per well in a 384-well plate will produce a file requiring roughly ~5–10 Tb of space (scaling with data resolution and number of imaged channels). Based on the rate of publication of such imaging datasets accompanying papers, we estimate ~30 of such collections being deposited in the database the first year and in subsequent years. However, a far greater number of publications per year (~100–200) display post-processed data without releasing the associated video files. The creation of this database will motivate these libraries to be uploaded as well, leading to an estimated 10K live-cell imaging videos plus metadata per year before accounting for growth.

An important bottleneck in current live-cell imaging data sharing is the diversity of file formats produced using different imaging modalities. Different models of microscope and versions of imaging software output different metadata files, and accurate parsing of these files is required for accurate comparisons between conditions within a dataset, as well as between distinct datasets. Currently, open-source software such as ImageJ [33] and the associated BioFormats [34] package allow users to view a range of proprietary datasets on most personal computers. We envision a database where such software would be used upon data submission to standardize all data and associated metadata into a common, usable format (Figure 1b,c). In this way, data submissions associated with a particular project can be easily re-analyzed and assessed by the community without first having to download massive files to ensure reproducibility, thus enabling researchers to explore and build on their colleagues' findings. Previous efforts have assembled a list of minimum microdata recommendations for biological imaging — notably the Recommended Metadata for Biological Images (REMBI) [35] guidelines — which will guide the data standardization built into our pipeline.

The most novel aspect of this enterprise is the downstream analysis of imaging data within the repository (Figure 1c,d). Currently, some of the most important steps in live-cell imaging analysis — such as the thresholding and watershed functions used in cell segmentation — are largely implemented on a case-by-case basis, and can differ in quality depending on what software is used and how stringent the segmentation parameters are. Unifying segmentation through a trusted workflow using software packages such as TrackMate [36], CellProfiler [37], and DeepCell [38], would make it possible to pass settings seamlessly between users and

reduce differences between downstream results stemming from improper, inconsistent, or non-reproducible segmentation workflows. Once the data have been segmented, additional information like location, fluorescence intensity, and measurable physical cell properties (e.g. morphology, size) will be collected across the timecourse and used for downstream analysis.

For the final step of the analysis pipeline, we envision a framework that takes variables of interest as input (for example, spatial location, cell size, or GFP expression level) and computes the values of these variables over time and space using established statistical metrics. One example is the analysis of spatial ‘clusters’ of cells; rather than the researcher having to manually label groups of cells that demonstrate a similar phenotype, they could define a specific metric calculated over the entire dataset, such as the Geary’s C or Moran’s I spatial correlation metrics [39]. This would allow for qualitative comparisons between cellular behaviors; for instance, when we see that protein X is expressed in clusters of cells, we can then quantify how frequently such clusters are found throughout the dataset, what the average cluster size is, and what other proteins display similar spatial behavior and in what cell types. Another feature of live-cell imaging data of interest is cellular lineage tracking, which is relevant, for example, to the emergence of differentiated cells in a population of stem cells [40–42], or, more generally, for understanding differential behaviors of daughter cells post-mitosis. There are a number of existing algorithms that can be used to track cell development [43–45], which will output generational and lineage information for each cell in the video. The ability to perform sophisticated analyses on computationally identified lineages of cells can allow scientists to understand the events that precede a particular cell-state change, a goal that unifies many fields within biology.

The repository proposed here would therefore constitute a fully ‘end-to-end’ resource for biologists. We envision a future in which a researcher attempting to collect live-cell imaging data can use this resource as a hypothesis generator before embarking on experiments such as the endogenous fluorescent tagging of a gene product or the introduction of a specific signaling pathway reporter; preparation of such cell or organism lines often takes substantial time and resources. Much as existing published sequencing datasets can make the construction of a candidate gene list easier for a researcher seeking to conduct a CRISPR screen, leveraging existing data may allow scientists interested in live-cell imaging to narrow their scope and avoid pouring time and resources into the wrong experiments or cell-line development. Moreover, this will allow scientists both to ask harder questions from existing data and to address more challenging problems with future experiments.

Addressing outstanding questions in cell biology

While the repository we have described would be immensely useful for the centralized storage and easy accessibility of high-quality imaging data, leveraging this huge dataset properly using robust, reproducible statistical tools could also offer answers to wide-reaching questions in cell biology. Since the discovery of canonical meta-zoan signaling pathways over the course of the 20th century [46], much interest has been focused on understanding the behavior of these pathways in individual cells, as well as their impact on neighboring cells and tissues. Over the past decade or so, reporters for signaling pathway activity have enjoyed a renaissance, and newly discovered pathway dynamics have raised a slew of questions that such reporters are uniquely poised to answer. Key measurements to understand the dynamics of a given pathway include the fraction of time spent in an active state, the time that passes between periods of pathway activation in a single cell, and evidence of spatial coupling in pathway activity. With the proper reporters and sufficient cell numbers (usually hundreds of imaged cells), all of these parameters can be recovered using timecourse analysis as well as more sophisticated spatial statistics [12, 47].

A number of signaling pathways, such as the Ras/ERK pathway, can control a range of biological processes [48], and it remains an open question how information encoded in different dynamics of a single pathway can lead to different downstream outcomes. A centralized database of live-cell signaling data would prove indispensable to answering this question. Looking across all examples of signaling dynamics, across cell types, contexts, and organisms, we may ask whether there are limits on the timescales of signaling behaviors of different pathways, or on the ability of pathways to communicate signaling information between cells. A single research group typically cannot study all of the settings in which dynamics of a certain pathway arise, as complex signaling dynamics have been found in a diverse range of contexts, including cancer, development, and wound healing [49, 50 2]. Rather, using data collected by research groups around the world, we may uncover a unified set of quantified dynamics that capture the behavior of a certain pathway across cell types, contexts, and treatment conditions. Since many canonical signaling pathways are also evolutionarily conserved, insight about a

pathway in a particular cell type or system may inform experiments in different contexts, and serve as hypothesis generators for future work.

Small molecule or genetic screening would also benefit from the establishment of a live-cell imaging data repository with robust analysis tools. Imaging is commonly used in assays evaluating cancer cell growth in the presence of a drug, or CAR-T cell efficiency in the presence of genetic or chemical perturbations [51, 52]. However, analysis tools for such assays are largely limited to phenotyping counts, e.g. counting the number of live cells over time based on the presence of a fluorescent dye or other such sensor. Live-cell imaging datasets from these experiments may contain rich and complex cellular phenotypes that capture changes to cell state and behavior across time in response to perturbations, including cell morphology, cell motility, and spatial distributions of cells. ‘Freeing’ these data beyond their current limited summary statistics could provide new mechanistic insights into these cellular processes, thereby informing efforts to develop better drugs and personalized cell therapies. This database would ideally parallel the rapidly growing set of repositories for primary cell genomic, transcriptomic, proteomic, and epigenomic data such as PRIDE [53], Cistrome [54], and TCGA [55], which have been widely used to inform and guide basic and translational research.

Previous initiatives have established multiple high-quality repositories for biological imaging data, including the BioImage Archive [56] and Open Microscopy Environment [57]. We anticipate that our resource will integrate seamlessly with these repositories, allowing users to directly import data from outside sources into our pipeline using dedicated plugins. Thus, while our resource would provide overlapping data-storage functions with these existing repositories, users would still be able to take advantage of the data standardization and downstream analysis tools unique to our pipeline regardless of where the data are stored.

Conclusion

As imaging modalities continue to reveal increasingly higher-resolution details about the cellular world, the push for accessible, easily processed, and statistically interpretable data has increased as well. We feel that a substantial gap in the cell biology community would be filled by the end-to-end repository and analysis resource presented here.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

CRedit Contribution

Barbara E. Engelhardt: Conceptualization, Supervision, Writing — original draft, Writing — review and editing.

Siddhartha Gautama Jena: Conceptualization, Writing — original draft, Writing — review and editing.

Alexander G. Goglia: Conceptualization, Visualization, Writing — original draft, Writing — review and editing.

Abbreviations

GFP, green fluorescent protein; REMBI, Recommended Metadata for Biological Images.

References

- 1 Murase, M. (1992) *The Dynamics of Cellular Motility* (pp. 1–357), John Wiley & Sons, Medford, MA, USA
- 2 Purvis, J.E. and Lahav, G. (2013) Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 <https://doi.org/10.1016/j.cell.2013.02.005>
- 3 Shamir, M., Bar-On, Y., Phillips, R. and Milo, R. (2016) Snapshot: timescales in cell biology. *Cell* **164**, 1302–1302 <https://doi.org/10.1016/j.cell.2016.02.058>
- 4 Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl Acad. Sci. U.S.A.* **74**, 5350–5354 <https://doi.org/10.1073/pnas.74.12.5350>
- 5 Mahmood, T. and Yang, P.C. (2012) Western blot: technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4**, 429 <https://doi.org/10.4103/1947-2714.94940>
- 6 Magaki, S., Hojat, S.A., Wei, B., So, A. and Yong, W.H. (2019) An introduction to the performance of immunohistochemistry. *Methods Mol. Biol.* **1897**, 289–298 https://doi.org/10.1007/978-1-4939-8935-5_25
- 7 Altschuler, S.J. and Wu, L.F. (2010) Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563 <https://doi.org/10.1016/j.cell.2010.04.033>
- 8 Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L. et al. (2016) Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 <https://doi.org/10.1038/nmeth.3734>
- 9 Filippi, S., Barnes, C.P., Kirk, P.D., Kudo, T., Kunida, K., McMahon, S.S. et al. (2016) Robustness of MEK-ERK dynamics and origins of cell-to-cell variability in MAPK signaling. *Cell Rep.* **15**, 2524–2535 <https://doi.org/10.1016/j.celrep.2016.05.024>

- 10 Aoki, K., Kumagai, Y., Sakurai, A., Komatsu, N., Fujita, Y., Shionyu, C. et al. (2013) Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Mol. Cell* **52**, 529–540 <https://doi.org/10.1016/j.molcel.2013.09.015>
- 11 Hiratsuka, T., Bordeu, I., Pruessner, G. and Watt, F.M. (2020) Regulation of ERK basal and pulsatile activity control proliferation and exit from the stem cell compartment in mammalian epidermis. *Proc. Natl Acad. Sci. U.S.A.* **117**, 17796–17807 <https://doi.org/10.1073/pnas.2006965117>
- 12 Goglia, A.G., Wilson, M.Z., Jena, S.G., Silbert, J., Basta, L.P., Devenport, D. et al. (2020) A live-cell screen for altered Erk dynamics reveals principles of proliferative control. *Cell Syst.* **10**, 240–253 <https://doi.org/10.1016/j.cels.2020.02.005>
- 13 Aikin, T.J., Peterson, A.F., Pokrass, M.J., Clark, H.R. and Regot, S. (2020) MAPK activity dynamics regulate non-cell autonomous effects of oncogene expression. *eLife* **9**, e60541 <https://doi.org/10.7554/eLife.60541>
- 14 Tsien, R.Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 <https://doi.org/10.1146/annurev.biochem.67.1.509>
- 15 Corrêa, Jr, I.R. (2014) Live-cell reporters for fluorescence imaging. *Curr. Opin. Chem. Biol.* **20**, 36–45 <https://doi.org/10.1016/j.cbpa.2014.04.007>
- 16 Kudo, T., Jeknić, S., Macklin, D.N., Akhter, S., Hughey, J.J., Regot, S. et al. (2018) Live-cell measurements of kinase activity in single cells using translocation reporters. *Nat. Protoc.* **13**, 155–169 <https://doi.org/10.1038/nprot.2017.128>
- 17 Burgstaller, S., Bischof, H., Gilkes, T., Stryeck, S., Gottschalk, B., Ramadani-Muja, J. et al. (2019) pH-Lemon, a fluorescent protein-based pH reporter for acidic compartments. *ACS Sensors* **4**, 883–891 <https://doi.org/10.1021/acssensors.8b01599>
- 18 Yamagata, K., Suetsugu, R. and Wakayama, T. (2009) Long-term, six-dimensional live-cell imaging for the mouse preimplantation embryo that does not affect full-term development. *J. Reprod. Dev.* **55**, 343–350 <https://doi.org/10.1262/jrd.20166>
- 19 Magidson, V. and Khodjakov, A. (2013) Circumventing photodamage in live-cell microscopy. *Methods Cell Biol.* **114**, 545–560 <https://doi.org/10.1016/B978-0-12-407761-4.00023-3>
- 20 Culley, S., Tosheva, K.L., Pereira, P.M. and Henriques, R. (2018) SRRF: universal live-cell super-resolution microscopy. *Int. J. Biochem. Cell Biol.* **101**, 74–79 <https://doi.org/10.1016/j.biocel.2018.05.014>
- 21 Bunyak, F., Palaniappan, K., Nath, S.K., Baskin, T.L. and Dong, G. (2006) Quantitative cell motility for in vitro wound healing using level set-based active contour tracking. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*, pp. 1040–1043, IEEE, New York City, NY, USA
- 22 Bahnon, A., Athanassiou, C., Koebler, D., Qian, L., Shun, T., Shields, D. et al. (2005) Automated measurement of cell motility and proliferation. *BMC Cell Biol.* **6**, 1–27 <https://doi.org/10.1186/1471-2121-6-19>
- 23 Phillip, J.M., Wu, P.H., Gilkes, D.M., Williams, W., McGovern, S., Daya, J. et al. (2017) Biophysical and biomolecular determination of cellular age in humans. *Nat. Biomed. Eng.* **1**, 1–12 <https://doi.org/10.1038/s41551-017-0093>
- 24 Phillip, J.M., Han, K.S., Chen, W.C., Wirtz, D. and Wu, P.H. (2021) A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nat. Protoc.* **16**, 754–774 <https://doi.org/10.1038/s41596-020-00432-x>
- 25 Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A. et al. (2014) Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373 <https://doi.org/10.1126/science.1254933>
- 26 Hannanta-Anan, P. and Chow, B.Y. (2016) Optogenetic control of calcium oscillation waveform defines NFAT as an integrator of calcium load. *Cell Syst.* **2**, 283–288 <https://doi.org/10.1016/j.cels.2016.03.010>
- 27 Albeck, J.G., Mills, G.B. and Brugge, J.S. (2013) Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Mol. Cell* **49**, 249–261 <https://doi.org/10.1016/j.molcel.2012.11.002>
- 28 Jena, S.G., Yu, C. and Toettcher, J.E. (2021) Dynamics and heterogeneity of Erk-induced immediate-early gene expression. *bioRxiv* <https://doi.org/10.1101/2021.04.30.442166>
- 29 Wilson, M.Z., Ravindran, P.T., Lim, W.A. and Toettcher, J.E. (2017) Tracing information flow from Erk to target gene induction reveals mechanisms of dynamic and combinatorial control. *Mol. Cell* **67**, 757–769 <https://doi.org/10.1016/j.molcel.2017.07.016>
- 30 Rodriguez, J., Ren, G., Day, C.R., Zhao, K., Chow, C.C. and Larson, D.R. (2019) Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* **176**, 213–226 <https://doi.org/10.1016/j.cell.2018.11.026>
- 31 Wang, W., Douglas, D., Zhang, J., Kumari, S., Enameh, M.S., Dai, Y. et al. (2020) Live-cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. *Sci. Adv.* **6**, eaba9319 <https://doi.org/10.1126/sciadv.aba9319>
- 32 Van Valen, D.A., Kudo, T., Lane, K.M., Macklin, D.N., Quach, N.T., DeFelicce, M.M. et al. (2016) Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Comput. Biol.* **12**, e1005177 <https://doi.org/10.1371/journal.pcbi.1005177>
- 33 Abràmoff, M.D., Magalhães, P.J. and Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42
- 34 Moore, J., Linkert, M., Blackburn, C., Carroll, M., Ferguson, R.K. and Flynn, H. (2015) OMERO and Bio-Formats 5: flexible access to large bioimaging datasets at scale. In *Medical Imaging 2015: Image Processing* (Wax, A., ed.), Vol. **9413**, pp. 37–42. SPIE, Bellingham, Washington, USA
- 35 Sarkans, U., Chiu, W., Collinson, L., Darrow, M.C., Ellenberg, J., Grunwald, D. et al. (2021) REMBI: recommended metadata for biological images-enabling reuse of microscopy data in biology. *Nat. Methods* **18**, 1418–1422 <https://doi.org/10.1038/s41592-021-01166-8>
- 36 Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E. et al. (2017) Trackmate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 <https://doi.org/10.1016/j.ymeth.2016.09.016>
- 37 McQuin, C., Goodman, A., Chernyshev, V., Kametsky, L., Cimini, B.A., Karhohs, K.W. et al. (2018) CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 <https://doi.org/10.1371/journal.pbio.2005970>
- 38 Bannon, D., Moen, E., Borba, E., Ho, A., Campilsson, I., Chang, B. et al. (2018) DeepCell 2.0: Automated cloud deployment of deep learning models for large-scale cellular image analysis. (Unpublished)
- 39 Dale, M.R., Dixon, P., Fortin, M.J., Legendre, P., Myers, D.E. and Rosenberg, M.S. (2002) Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* **25**, 558–577 <https://doi.org/10.1034/j.1600-0587.2002.250506.x>
- 40 Weinreb, C., Rodríguez-Fraticelli, A., Camargo, F.D. and Klein, A.M. (2020) Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 <https://doi.org/10.1126/science.aaw3381>
- 41 Kretzschmar, K. and Watt, F.M. (2012) Lineage tracing. *Cell* **148**, 33–45 <https://doi.org/10.1016/j.cell.2012.01.002>
- 42 Maruno, T., Fukuda, A., Goto, N., Tsuda, M., Ikuta, K., Hiramatsu, Y. et al. (2021) Visualization of stem cell activity in pancreatic cancer expansion by direct lineage tracing with live imaging. *eLife* **10**, e55117 <https://doi.org/10.7554/eLife.55117>
- 43 Moen, E., Borba, E., Miller, G., Schwartz, M., Bannon, D., Koe, N. et al. (2019) Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning

- 44 Wolff, C., Tinevez, J.Y., Pietzsch, T., Stamatakis, E., Harich, B., Guignard, L. et al. (2018) Multi-view light-sheet imaging and tracking with the maMUT software reveals the cell lineage of a direct developing arthropod limb. *eLife* **7**, e34410 <https://doi.org/10.7554/eLife.34410>
- 45 Piltti, K.M., Cummings, B.J., Carta, K., Manughian-Peter, A., Worne, C.L., Singh, K. et al. (2018) Live-cell time-lapse imaging and single-cell tracking of *in vitro* cultured neural stem cells—tools for analyzing dynamics of cell cycle, migration, and lineage selection. *Methods* **133**, 81–90 <https://doi.org/10.1016/j.ymeth.2017.10.003>
- 46 Lim, W., Mayer, B. and Pawson, T. (2014) *Cell Signaling*, Taylor & Francis, Milton Park, Abingdon-on-Thames, Oxfordshire, UK
- 47 Verma, A., Jena, S.G., Isakov, D.R., Aoki, K., Toettcher, J.E. and Engelhardt, B.E. (2021) A self-exciting point process to study multicellular spatial signaling patterns. *Proc. Natl Acad. Sci. U.S.A.* **118**, e2026123118 <https://doi.org/10.1073/pnas.2026123118>
- 48 Lavoie, H., Gagnon, J. and Therrien, M. (2020) ERK signalling: a master regulator of cell behaviour, life and fate. *Nat. Rev. Mol. Cell Biol.* **21**, 607–632 <https://doi.org/10.1038/s41580-020-0255-7>
- 49 Ghilardi, S.J., O'Reilly, B.M. and Sgro, A.E. (2020) Intracellular signaling dynamics and their role in coordinating tissue repair. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **12**, e1479 <https://doi.org/10.1002/wsbm.1479>
- 50 Bugaj, L.J., Sabnis, A.J., Mitchell, A., Garbarino, J.E., Toettcher, J.E., Bivona, T.G. et al. (2018) Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. *Science* **361**, eaao3048 <https://doi.org/10.1126/science.aao3048>
- 51 Li, H., Huang, Y., Jiang, D.Q., Cui, L.Z., He, Z., Wang, C. et al. (2018) Antitumor activity of EGFR-specific CAR T cells against non-small-cell lung cancer cells *in vitro* and *in mice*. *Cell Death Dis.* **9**, 1–11 <https://doi.org/10.1038/s41419-017-0012-9>
- 52 Majzner, R.G., Rietberg, S.P., Sotillo, E., Dong, R., Vachharajani, V.T., Labanieh, L. et al. (2020) Tuning the antigen density requirement for CAR T-cell activity. *Cancer Discov.* **10**, 702–723 <https://doi.org/10.1158/2159-8290.CD-19-0945>
- 53 Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D. et al. (2005) PRIDE: the proteomics identifications database. *Proteomics* **5**, 3537–3545 <https://doi.org/10.1002/pmic.200401303>
- 54 Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C. et al. (2016) Cistrome data browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.* **45**, D658–D662 <https://doi.org/10.1093/nar/gkw983>
- 55 Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* **19**, A68
- 56 Hartley, M., Kleywegt, G.J., Patwardhan, A., Sarkans, U., Swedlow, J.R. and Brazma, A. (2022) The bioimage archive - building a home for life-sciences microscopy data. *J Mol Biol* **18**, 167505 <https://doi.org/10.1016/j.jmb.2022.167505>
- 57 Goldberg, I.G., Allan, C., Burel, J.M., Creager, D., Falconi, A., Hochheiser, H. et al. (2005) The open microscopy environment (OME) data model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol.* **6**, R47 <https://doi.org/10.1186/gb-2005-6-5-r47>