Direct, high-throughput linking of single cell imaging and gene expression

Catherine K Xu[1,2]*, Georg Meisl[3], Nikita Moshkov[4], Karolis Goda[5], Alexey Shkarin[1], Maximilian F Schlögel[1], Tuomas PJ Knowles[3,6], Linas Mazutis[5], Jochen Guck[1,2]*

**To whom correspondence should be addressed: catherine.xu@mpzpm.mpg.de, jochen.guck@mpl.mpg.de*

[1]Max Planck Institute for the Science of Light, Erlangen, Germany

[2]Max Planck Zentrum für Physik und Medizin, Erlangen, Germany

[3]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK

[4]Institute of Computational Biology, Helmholtz Munich, Munich, Germany

[5]Institute of Biotechnology, Vilnius University, Vilnius, Lithuania

[6]Cavendish Laboratory, University of Cambridge, Cambridge, UK

Abstract

Single-cell technologies have transformed our ability to dissect cellular heterogeneity by enabling measurements of individual molecular modalities, from genome and transcriptome to proteome and metabolome. However, at the single-cell level, the physical properties of cells, such as size, morphology, and mechanical state, remain largely disconnected from molecular profiling, limiting our understanding of the relationship between these aspects of cellular phenotype and gene expression. We introduce im-seq, a high-throughput, flow-based platform that integrates live-cell imaging with droplet-based mRNA sequencing at the single-cell level. By optically barcoding individual cells, im-seq enables the joint capture of physical and transcriptional profiles from single cells. We demonstrate that this multimodal approach can resolve physical and molecular features across cell lines, to reveal genes associated with phenotypic properties at unprecedented resolution. Our results establish im-seq as a versatile high-throughput framework for linking

genetic information to physical properties, providing the large scale, information-dense datasets needed to power the next generation of data-driven discoveries in cell biology.

Main Text

Single-cell technologies have become a crucial tool for investigating biological systems by enabling detailed characterisation of cellular heterogeneity. Measurements of individual modalities at the single cell level, such as the transcriptome or proteome, have revealed dynamic cellular states and trajectories that are obscured in bulk analyses [1, 2]. More recently, the emergence of multi-omic approaches, which capture multiple molecular identities from the same cell, has provided further insights into how cells function [3]. These integrative technologies demonstrate the importance of high-dimensional, information-rich data for dissecting complex biological processes across areas such as development, immunity, and disease.

However, in contrast to these advances in molecular profiling, the physical dimension is almost entirely lacking in the space of multimodal single-cell technologies [4]. Physical phenotypes including cell size, shape, and mechanical properties, are fundamental to cellular function and identity, influencing processes such as migration, differentiation, and immune responses [5]. Yet, their relationship to cellular biochemistry such as gene expression remains poorly understood, due to the lack of technologies that can connect physical and molecular readouts at single-cell resolution.

Previous efforts to link optical phenotype data with sequencing readouts at the single-cell level have relied on imaging cells under static conditions, combined with positional barcoding on arrays [6], optical decoding of barcode bead identities [7, 8, 9, 10, 11], or precise cell handling for deterministic barcoding [12] or library preparation [13]. While these approaches have enabled important demonstrations of image–transcriptome coupling, static imaging methods are inherently limited in throughput and do not scale easily, especially when compared to flow-based systems such as imaging flow cytometry or droplet-based single-cell RNA sequencing [14, 15, 16].

To overcome these limitations, we developed a high-throughput, scalable, flow-based platform that integrates live-cell imaging, droplet-based scRNA-seq, and

optical barcode decoding. This approach enables simultaneous capture of physical phenotypes and transcriptomic profiles from individual cells in flow, combining the versatility of imaging-based characterisation with the scalability of droplet sequencing workflows.

In standard droplet-based scRNA-seq, single cells are co-encapsulated with uniquely barcoded beads, allowing transcriptomes to be indexed during reverse transcription [15, 16, 17]. However, because these DNA barcodes are randomly distributed across optically identical beads, there is no way to associate optical phenotypes with sequencing output. To overcome this, im-seq introduces an additional optical barcode into each droplet: a combinatorial tag that can be identified by both imaging and sequencing (Figure 1). These multimodal barcodes comprise multiple hydrogel beads, each carrying a unique sequence and known, unique optical signature [10, 18, 19]. By randomly combining multiple such beads together, we can construct immense numbers of multi-component barcodes: as the complete barcode is encoded by the presence or absence of each bead type, 2^n unique barcodes can be generated from just n bead species. For example, 30 beads yield over 1 billion unique combinations. By imaging droplets to identify their optical barcode and sequencing to read the same barcode, we can directly link physical phenotypes to gene expression at the single-cell level.

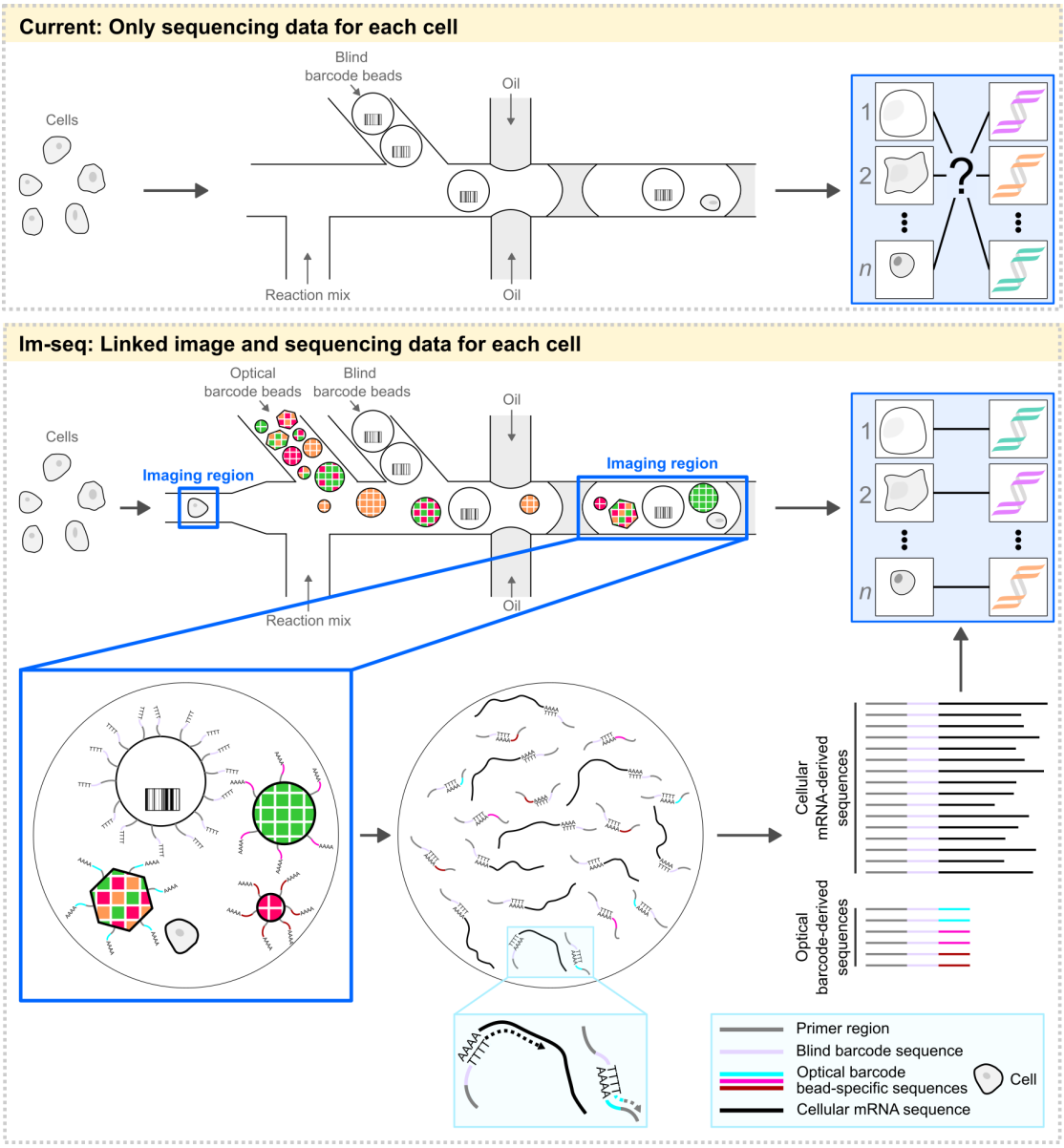

Figure 1: **Overview of im-seq.** Current single cell mRNA expression profiling yields data unconnected to the physical phenotypes of cells. In contrast, im-seq is able to link together both imaging and sequencing data for single cells in high throughput. In standard scRNA-seq, cells are encapsulated into droplets with barcode beads of unknown sequences ("blind barcode beads"). In im-seq, cells are imaged immediately prior to encapsulation with both blind barcode beads and optical barcode beads. The optical barcode beads, which are functionalised with known DNA sequences, are then used to enable direct image-to-sequence linking of single cells.

In order to synthesise uniquely identifiable optical barcode components, we employed droplet microfluidics to generate highly homogeneous populations of hydrogel beads. Unique bead types were produced by varying bead size, shape, and colour, defined by the ratiometric combinations of three fluorescent dyes [19, 20, 21] (Figure 2A, B). Each bead type was reversibly conjugated to a known DNA sequence to enable independent identification by sequencing. Following optical barcode assembly in droplets, barcodes could be read both by the optical identification of component beads according to their size, shape, and colour (Figure 2C), and by their associated DNA sequences. In each droplet, the blind barcode bead tags all encapsulated nucleic acids: optical barcode DNA and cell mRNA, with the same unique sequence [22, 23]. The combinatorial optical barcode thereby links together optical and sequencing data at the single-cell level.

We fully exploited our flow-based system by incorporating mechanical profiling into our cell phenotyping. By imaging cells while they are subjected to shear forces under flow, their relative mechanical stiffness can be determined from their degree of deformation [24]. The resulting images contain information on the cell size, morphology, and mechanical properties, a feature inaccessible to existing image-sequence linking methods such as imaging FACS and spatial transcriptomics [25, 26]. Furthermore, chemical crosslinking may alter cell mechanical properties, necessitating measurements on live, rather than fixed, cells [27]. Following the phenotypic characterisation, cells are immediately encapsulated into droplets alongside the randomly generated optical barcode and scRNA-seq reagents, enabling almost simultaneous capture of its gene expression profile.

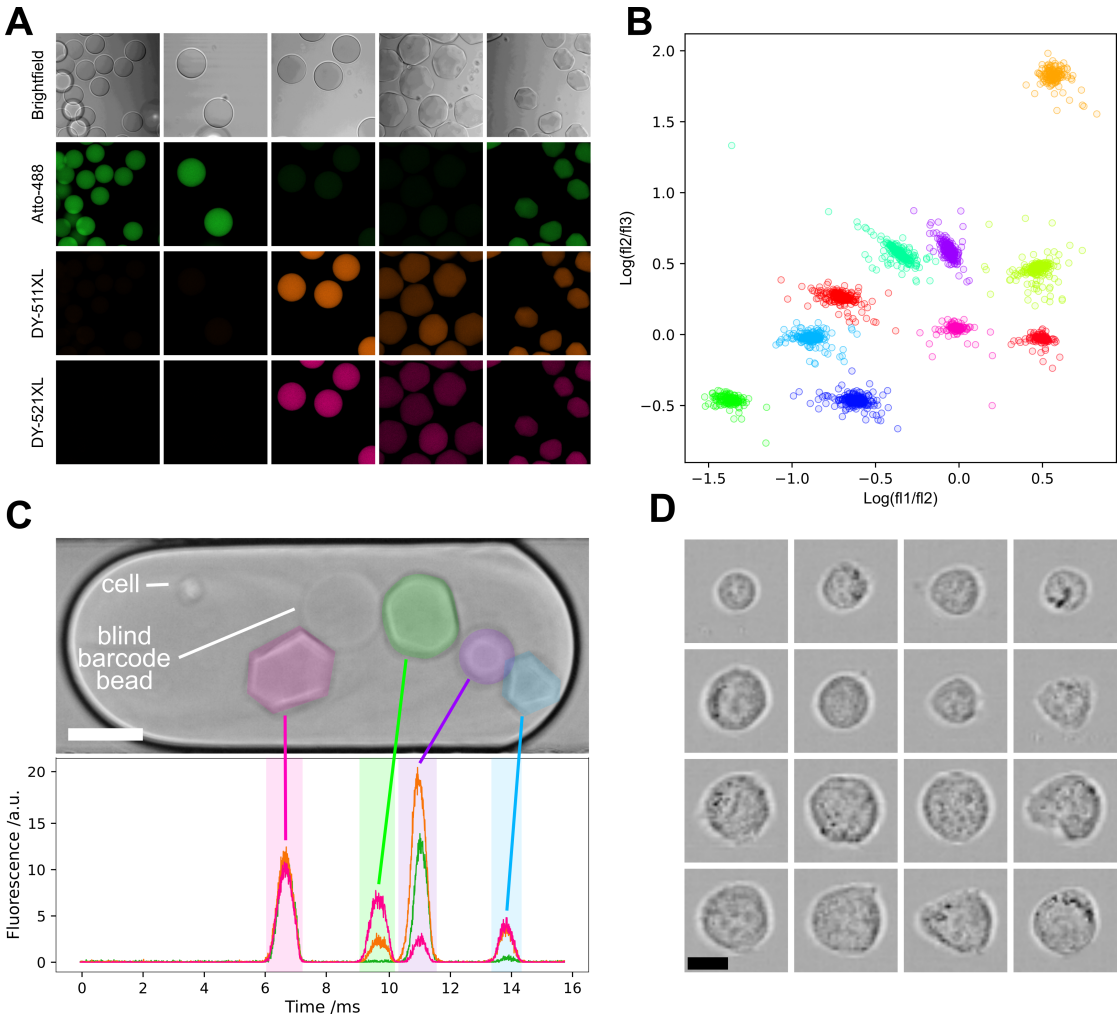

Figure 2: **Optical barcodes can be identified on-chip to label cell images.** (**A**) Example brightfield and fluorescence images of types (columns) of optical barcode beads, which are each uniquely identifiable based on their combination of size, shape, and colour. (**B**) Ten ratiometric colours defined by the combination of three fluorescent dyes. The three fluorescence signals (fl1, fl2, fl3) were reduced to two axes by taking the pairwise ratios, removing noise arising from variations in absolute fluorescence intensity. (**C**) Droplet containing a cell, blind barcode bead, and four optical barcode beads which comprise the optical barcode. Fluorescence traces are acquired simultaneously to determine the colour of each bead (coloured as in (**B**)) (Scale bar: 50 µm.). (**D**) Example images of cells subjected to shear forces, immediately prior to encapsulation in droplets. In addition to size and morphology, the deformation of the cell reports on its mechanical stiffness. (Scale bar: 10 µm.)

We sought to validate that our multimodal barcodes can be correctly uniquely linked between imaging and sequencing data. We mixed data from two experiments together, to yield one set of barcoded cell-containing droplets detected optically and another by sequencing, both originating from a mixture of experiment classes 1 and 2. We then analysed both optical and sequencing datasets to extract droplet barcode identities and identified droplets with the same unique barcodes detected in both modalities. Comparison of the linked optical-sequencing data with the known experiment classes confirmed that over 98% of matched droplets were correctly classified (Figure 3A).

Having validated that our multimodal barcodes can correctly link optical and sequence data at the single droplet level, we next applied im-seq to a mixture of two cell lines, namely OCI-LY1 and THP-1. These cell lines differ in physical phenotype at the population level (Figure 3E), particularly in cell size, but individual cells can overlap in properties. We therefore employed these cell lines to showcase the importance of linking phenotype and gene expression at the single cell level, as population- and sorting-based approaches require binning of cells [28]. By pairing cells with their encapsulating droplet, the barcode was read both optically and by sequencing, to link cell image to mRNA expression (Figure 3B-D). Droplet generation for ten minutes yielded linked image and gene expression data for approximately one thousand single cells. Cell type classification based on mRNA expression reproduced the phenotype characterisation, confirming the correct linking of image and sequence data at the single cell level (Figure 3F), thus allowing us to dissect phenotype-genotype relationships at unprecedented resolution. Our determined error rate of <2% for optical–sequence linking (Figure 3A) indicates that the overlapping cells arise from true shared properties between the two classes, rather than technical artifacts. While just two image-based parameters are displayed here, the dataset for each cell contains the whole image (Figure 2D and 3C), enabling the interrogation of the gene expression changes that underlie physical phenotypes.

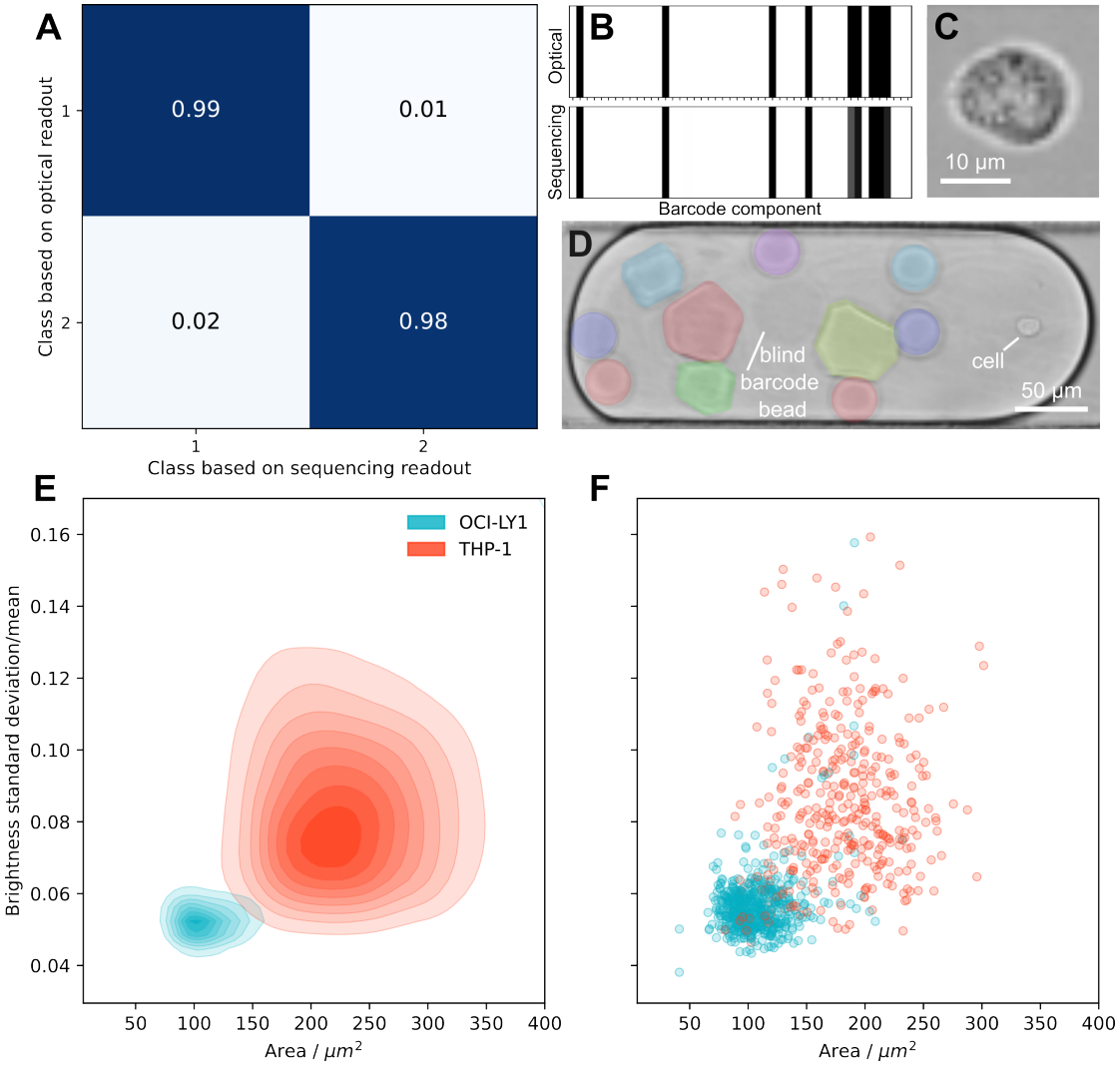

Figure 3: **Im-seq enables linking of single cell image and mRNA expression.** (**A**) Ground truth validation of multimodal barcode linking across optical and sequencing data, by combining data from two separate experiments. In >98% of linked droplets, sequencing data were paired with optical droplet data from the same experiment class, confirming that droplets are both uniquely barcoded and correctly identified. (**B-D**) Example single cell data. Cells are imaged (**C**) and encapsulated in droplets alongside blind barcode beads and optical barcode beads (**D**: beads coloured as in Figure 2). The barcode can be constructed from both optical and sequencing data (**B**). Here, each line represents the probability (from light to dark) of the presence or absence of each optical barcode bead type, based on the respective data types. (**E**) Area and ratio of standard deviation to mean brightness extracted from images of OCI-LY1 (turquoise) and THP-1 (orange) cells in separate measurements, showing distinct physical properties. (**F**) Plot of image features of 943 single cells, coloured according to the gene expression-based cell type classification.

In conclusion, we have developed im-seq, a general and scalable strategy for linking image-based physical phenotypes with single-cell transcriptomic profiles at high throughput. By enabling the joint capture of optical and sequencing data in flow, im-seq provides a powerful framework for investigating the relationship between cell phenotype and molecular composition. The core concept of im-seq is modular and compatible with a broad range of cell phenotype characterisation modalities, including fluorescence imaging flow cytometry [25]. Moreover, since im-seq operates on live cells, it can also be integrated with functional assays such as secretion profiling in droplets [29] and perturbation screens [30]. The relative simplicity of im-seq is ideal for such integrative workflows, as it does not require complex liquid handling such as valves and controlled cell/droplet dispensing [12, 13, 31]. We anticipate that im-seq will open new avenues for studying how physical properties emerge from gene expression programs, and how multimodal signatures can be leveraged to understand cell behaviour in health and disease.

Acknowledgements

We are grateful to Cornelia Liebers, Manuela Hauke, and Christine Schweitzer for culturing cells, to Dr Irina Harder, Parth Patel, and Dr Salvatore Girardo for their support in the clean room, and to Dr Arif Ekici and Dr Rinho Kim for their support with sequencing. We also thank Professor Fabian Theis, Dr Niklas Schmacke, Dr David Morse, Dr Tomasz Kaminski, and Professor Justus Kebschull for their helpful discussions. This research received funding from: the Peter und Traudl Engelhorn Stiftung fellowship (C.K.X), the HFSP Cross-Disciplinary Fellowship (C.K.X.), the Marie Skłodowska-Curie fellowship READ-seq (agreement no. 101068803) (C.K.X.).

Competing interests

A patent has been filed for the technology (inventors C.K.X., G.M., T.P.J.K., and J.G) by the institutions Cambridge Enterprise and Max Planck Innovation (WO 2025/146377).

References

[1] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell rna sequencing technologies and bioinformaticspipelines. *Experimental & Molecular Medicine*, 50(8):1–14, Aug 2018.

[2] Hayley M. Bennett, William Stephenson, Christopher M. Rose, and Spyros Darmanis. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nature Methods*, 20(3):363–374, Mar 2023.

[3] Alev Baysoy, Zhiliang Bai, Rahul Satija, and Rong Fan. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, 24(10):695–713, Oct 2023.

[4] Camille L. G. Lambert, Guido van Mierlo, Johannes J. Bues, Orane J. Guillaume-Gentil, and Bart Deplancke. Engineering next-generation microfluidic technologies for single-cell phenomics. *Nature Genetics*, 57(6):1344–1356, Jun 2025.

[5] Marta Urbanska and Jochen Guck. Single-cell mechanics: Structural determinants and functional relevance. *Annual Review of Biophysics*, 53, 2024.

[6] Jesse Q. Zhang, Christian A. Siltanen, Leqian Liu, Kai-Chun Chang, Zev J. Gartner, and Adam R. Abate. Linked optical and gene expression profiling of single cells at high-throughput. *Genome Biology*, 21(1):49, Feb 2020.

[7] Jinzhou Yuan, Jenny Sheng, and Peter A. Sims. Scope-seq: a scalable technology for linking live cell imaging and single-cell rna sequencing. *Genome Biology*, 19(1):227, Dec 2018.

[8] Zhouzerui Liu, Jinzhou Yuan, Anna Lasorella, Antonio Iavarone, Jeffrey N. Bruce, Peter Canoll, and Peter A. Sims. Integrating single-cell rna-seq and imaging with scope-seq2. *Scientific Reports*, 10(1):19482, Nov 2020.

[9] Fumiko Kawasaki, Yuka Mori, Takahiro Mimori, Issei Sato., and Sadao Ota. Identification of in-droplet multicellular communities by light-induced combinatorial dna barcoding. *Chemistry – A European Journal*, 29(53):e202301133, Sep 2023.

[10] Fumiko Kawasaki, Takahiro Mimori, Yuka Mori, Hiroyuki Aburatani, Nozomu Yachie, Issei Sato, and Sadao Ota. Computational design of synthetic optical

barcodes in microdroplets. *Advanced Optical Materials*, 12(12):2302564, Apr 2024.

[11] Arata Tsuchida, Taikopaul Kaneko, Kaori Nishikawa, Mayu Kawasaki, Ryuji Yokokawa, and Hirofumi Shintaku. Opto-combinatorial indexing enables high-content transcriptomics by linking cell images and transcriptomes. *Lab on a Chip*, 24(8):2287–2297, 2024.

[12] Tyler N. Chen, Anushka Gupta, Mansi D. Zalavadia, and Aaron Streets. μcb-seq: microfluidic cell barcoding and sequencing for high-resolution imaging and sequencing of single cells. *Lab on a Chip*, 20(21):3899–3913, 2020.

[13] Keara Lane, David Van Valen, Mialy M. DeFelice, Derek N. Macklin, Takamasa Kudo, Ariel Jaimovich, Ambrose Carr, Tobias Meyer, Dana Pe'er, Stéphane C. Boutet, and Markus W. Covert. Measuring signaling and rna-seq in the same cell links gene expression to dynamic patterns of nf-κb activation. *Cell Systems*, 4(4):458–469.e5, Apr 2017.

[14] Paul Rees, Huw D. Summers, Andrew Filby, Anne E. Carpenter, and Minh Doan. Imaging flow cytometry. *Nature Reviews Methods Primers*, 2(1):86, Nov 2022.

[15] Allon M. Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A. Weitz, and Marc W. Kirschner. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201, May 2015.

[16] Evan Z. Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R. Bialas, Nolan Kamitaki, Emily M. Martersteck, John J. Trombetta, David A. Weitz, Joshua R. Sanes, Alex K. Shalek, Aviv Regev, and Steven A. McCarroll. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, May 2015.

[17] Rapolas Zilionis, Juozas Nainys, Adrian Veres, Virginia Savova, David Zemmour, Allon M. Klein, and Linas Mazutis. Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols*, 12(1):44–73, Jan 2017.

[18] Paul H. Dannenberg, Jie Wang, Yue Zhuo, Sangyeon Cho, Kwon-Hyeon Kim, and Seok-Hyun Yun. Droplet microfluidic generation of a million optical microparticle barcodes. *Optics Express*, 29(23):38109–38118, Nov 2021.

[19] Yinnian Feng, Adam K. White, Jamin B. Hein, Eric A. Appel, and Polly M. Fordyce. Mrbles 2.0: High-throughput generation of chemically functionalized spectrally and magnetically encoded hydrogel beads using a simple single-layer microfluidic device. *Microsystems & Nanoengineering*, 6(1):109, Nov 2020.

[20] Daniel W. Weisgerber, Makiko Hatori, Xiangpeng Li, and Adam R. Abate. Polyhedral particles with controlled concavity by indentation templating. *Analytical Chemistry*, 94(21):7475–7482, May 2022.

[21] Sebastian Seiffert and Wilhelm Oppermann. Amine-functionalized polyacrylamide for labeling and crosslinking purposes. *Macromolecular Chemistry and Physics*, 208(16):1744–1752, Aug 2007.

[22] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, Sep 2017.

[23] Vanessa M. Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C. Wilson, Renee Moore, Terrill K. McClanahan, Svetlana Sadekova, and Joel A. Klappenbach. Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology*, 35(10):936–939, Oct 2017.

[24] Oliver Otto, Philipp Rosendahl, Alexander Mietke, Stefan Golfier, Christoph Herold, Daniel Klaue, Salvatore Girardo, Stefano Pagliara, Andrew Ekpenyong, Angela Jacobi, Manja Wobus, Nicole Töpfner, Ulrich F. Keyser, Jörg Mansfeld, Elisabeth Fischer-Friedrich, and Jochen Guck. Real-time deformability cytometry: on-the-fly cell mechanical phenotyping. *Nature Methods*, 12(3):199–202, Mar 2015.

[25] Daniel Schraivogel, Terra M. Kuhn, Benedikt Rauscher, Marta Rodríguez-Martínez, Malte Paulsen, Keegan Owsley, Aaron Middlebrook, Christian Tischer, Beáta Ramasz, Diana Ordoñez-Rueda, Martina Dees, Sara Cuylen-Haering, Eric Diebold, and Lars M. Steinmetz. High-speed fluorescence image–enabled cell sorting. *Science*, 375(6578):315–320, Jan 2022.

[26] Cameron G. Williams, Hyun Jae Lee, Takahiro Asatsuma, Roser Vento-Tormo, and Ashraful Haque. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, 14(1):68, Jun 2022.

[27] Seong-Oh Kim, Joonhui Kim, Takaharu Okajima, and Nam-Joon Cho. Mechanical properties of paraformaldehyde-treated individual cells investigated by atomic force microscopy and scanning ion conductance microscopy. *Nano Convergence*, 4(1):5, Mar 2017.

[28] Tianben Ding, Kelvin C. M. Lee, Kevin K. Tsia, T. Nicolai Siegel, Dino Di Carlo, and Keisuke Goda. Image-activated cell sorting. *Nature Reviews Bioengineering*, Jul 2025.

[29] Klaus Eyer, Raphaël C. L. Doineau, Carlos E. Castrillon, Luis Briseño-Roa, Vera Menrath, Guillaume Mottet, Patrick England, Alexei Godina, Elodie Brient-Litzler, Clément Nizak, Allan Jensen, Andrew D. Griffiths, Jérôme Bibette, Pierre Bruhns, and Jean Baudry. Single-cell deep phenotyping of igg-secreting cells for high-resolution immune monitoring. *Nature Biotechnology*, 35(10):977–982, Oct 2017.

[30] Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P. Fulco, Livnat Jerby-Arnon, Nemanja D. Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M. Norman, Eric S. Lander, Jonathan S. Weissman, Nir Friedman, and Aviv Regev. Perturb-seq: Dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *Cell*, 167(7):1853–1866.e17, Dec 2016.

[31] Russell H. Cole, Shi-Yang Tang, Christian A. Siltanen, Payam Shahi, Jesse Q. Zhang, Sean Poust, Zev J. Gartner, and Adam R. Abate. Printed droplet microfluidics for on demand dispensing of picoliter droplets and cells. *Proceedings of the National Academy of Sciences*, 114(33):8728–8733, Aug 2017.