Title

Microfluidics-free single-cell genomics with templated emulsification

Journal

Nature Biotechnology, 41(11)

ISSN

1087-0156

Authors

Clark, Iain C

Fontanez, Kristina M

Meltzer, Robert H

et al.

Publication Date

2023-11-01

DOI

10.1038/s41587-023-01685-z

nature biotechnology

Article

https://doi.org/10.1038/s41587-023-01685-z

Microfluidics-free single-cell genomics with templated emulsification

Received: 18 May 2022

Accepted: 20 January 2023

Published online: 06 March 2023

Check for updates

Iain C. Clark[1], Kristina M. Fontanez[2], Robert H. Meltzer[2], Yi Xue[2], Corey Hayford[2], Aaron May-Zhang[2], Chris D'Amato[2], Ahmad Osman[2], Jesse Q. Zhang[2], Pabodha Hettige[2], Jacob S. A. Ishibashi[2], Cyrille L. Delley[3], Daniel W. Weisgerber[3], Joseph M. Replogle[4], Marco Jost[4,12], Kiet T. Phong[5], Vanessa E. Kennedy[6], Cheryl A. C. Peretz[7], Esther A. Kim[8], Siyou Song[8], William Karlon[9], Jonathan S. Weissman[4,10], Catherine C. Smith[6], Zev J. Gartner[5,11] & Adam R. Abate[3]

Current single-cell RNA-sequencing approaches have limitations that stem from the microfluidic devices or fluid handling steps required for sample processing. We develop a method that does not require specialized microfluidic devices, expertise or hardware. Our approach is based on particle-templated emulsification, which allows single-cell encapsulation and barcoding of cDNA in uniform droplet emulsions with only a vortexer. Particle-templated instant partition sequencing (PIP-seq) accommodates a wide range of emulsification formats, including microwell plates and large-volume conical tubes, enabling thousands of samples or millions of cells to be processed in minutes. We demonstrate that PIP-seq produces high-purity transcriptomes in mouse–human mixing studies, is compatible with multiomics measurements and can accurately characterize cell types in human breast tissue compared to a commercial microfluidic platform. Single-cell transcriptional profiling of mixed phenotype acute leukemia using PIP-seq reveals the emergence of heterogeneity within chemotherapy-resistant cell subsets that were hidden by standard immunophenotyping. PIP-seq is a simple, flexible and scalable next-generation workflow that extends single-cell sequencing to new applications.

Single-cell RNA sequencing (scRNA-seq) is an essential technology in the biological sciences because it reveals how the properties of tissues arise from the transcriptional states of numerous interacting cells. Defining the gene expression signatures of individual cells allows cell-type classification, the discovery of unique cell states during development and disease and the prediction of regulatory mechanisms that control these states. As a result, bulk sequencing is being rapidly replaced by single-cell methods. The first single-cell approaches isolated cells and

[1]Department of Bioengineering, University of California, Berkeley, California Institute for Quantitative Biosciences, Berkeley, CA, USA. [2]Fluent Biosciences, Watertown, MA, USA. [3]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [4]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. [6]Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [7]Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA. [8]Division of Plastic and Reconstructive Surgery, University of California San Francisco, San Francisco, CA, USA. [9]Departments of Pathology and Laboratory Medicine, University of California San Francisco, San Francisco, CA, USA. [10]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Chan Zuckerberg Biohub, San Francisco, CA, USA. [12]Present address: Department of Microbiology, Harvard Medical School, Boston, MA, USA. e-mail: adam@abatelab.org

prepared them individually for sequencing[1–4]. While improvements in molecular biology increased data quality[5,6], the requisite isolation and processing of separate cells ultimately limited throughput. Implementation of valve-based microfluidics reduced hands-on time[7] but failed to substantially increase cell number and thus could not capture the heterogeneity intrinsic to most tissues. Advances in high-throughput droplet microfluidic barcoding have expanded single-cell sequencing to tens of thousands of cells[8,9] and fueled biological discovery but require expensive instruments located in core facilities and therefore remain inaccessible to many labs. Methods for direct combinatorial indexing of cells[10,11], the use of nanowell arrays[12] or sample multiplexing[13,14] have overcome some limitations of microfluidics, but no current method simultaneously accommodates both low (10) and high (>10^6) cell numbers, can be applied to hundreds of independent samples and can be rapidly implemented without custom equipment.

The scalability of single-cell methods is important for many applications, including tissue atlas projects[15–18], million-cell perturbation experiments[19], drug development pipelines[20] and developmental studies[21]. Droplet microfluidics has an intrinsic disadvantage at high cell numbers due to the upper limit on drop generation speed. At high fluid velocities, droplet generation becomes uncontrolled, resulting in polydispersed emulsions and poor bead loading that reduces single-cell data quality[22,23]. Therefore, to sequence millions of cells requires long run times, parallel droplet generators with complex designs that are prone to clogging or implementation of additional barcoding steps before encapsulation[24]. More generally, droplet microfluidics relies on an expensive instrument usually located in a core facility, which necessitates sample transport or fixation that can alter RNA profiles. Centralized processing also reduces access to many labs and does not fit experiments that need rapid or point-of-collection sample handling, such as remote fieldwork or studies using infectious samples requiring biosafety precautions[12,25].

Much effort has thus gone into developing microfluidics-free single-cell methods. Split-pool ligation[10,11] and tagmentation[26,27] perform direct combinatorial barcoding of bulk suspensions and substantially increase cell number; however, these laborious workflows require enormous numbers of pipetting operations and are poorly suited for low cell inputs. Moreover, while scalable, these methods require substantial expertise[28], and broad adoption of split-pool barcoding will likely require robotic automation in a centralized facility. Alternatively, methods based on nanowells prioritize simplicity and cost-effectiveness[12,25]. No microfluidics are required, and wells are loaded by sedimentation, providing an instrument-free and point-of-use solution. However, nanowell array chips do not efficiently scale in cell or sample number; the planar arrays capture cells on a two-dimensional surface and, thus, cannot compete with emulsions or combinatorial indexing using a three-dimensional volume that easily scales to millions of cells. Moreover, unless combined with multiplexing[13,14], nanowell chips are poorly suited for processing many separate samples because they require one array per sample and thus hundreds of arrays for hundreds of samples. To advance the field of single-cell genomics, next-generation technologies must simultaneously innovate on speed, scale and ease of use. An ideal system would be compatible with the barcoding of separate samples in well plates, accommodate orders-of-magnitude differences in cell number, be completed in minutes and be easy to run at the bench or in the field without specialized instrumentation.

Here, we describe a flexible, scalable and instrument-free scRNA-seq method based on rapid templated emulsification of cells and barcoded hydrogel templates without microfluidics[29]. In contrast to microfluidic emulsification, in which droplets are created sequentially and thus their number scales with instrument run time, templated emulsification generates monodispersed droplets in parallel by bulk self-assembly, and, thus, the number of droplets (and cells that can be barcoded) scales only with container volume. The result is a scalable, user-friendly scRNA-seq method that we call particle-templated instant partition sequencing (PIP-seq). Templated emulsification produces drops that are equivalent to those generated with microfluidics and compatible with the latest innovations in multiomic measurements. Here, we show that PIP-seq generates accurate single-cell gene expression profiles from human tissues and is compatible with multimodal measurements of RNA and single guide RNA (sgRNA; CRISPR droplet sequencing (CROP-seq)) or RNA and protein (cellular indexing of transcriptomes and epitopes sequencing (CITE-seq)). Finally, we demonstrate the use of PIP-seq to monitor the response of individuals with mixed phenotype acute leukemia (MPAL) to chemotherapy, revealing heterogeneity within cells with similar immunophenotypes. In summary, PIP-seq fills an unmet technical need by improving the speed, scalability and ease of use of single-cell sequencing.

Results

Overview of the technology

PIP-seq uses particle templating to compartmentalize cells, barcoded hydrogel templates and lysis reagents in monodispersed water-in-oil droplets (Fig. 1a). Rapid emulsification with a standard vortexer allows cells to be encapsulated at the bench or point of collection in minutes. The cells are lysed by increasing the temperature to 65 °C, which activates proteinase K (PK), releasing cellular mRNA that is captured on polyacrylamide beads decorated with barcoded poly(T) sequences (Fig. 1b). PIP-seq emulsions can be stored for days at 0 °C without change in data quality (Extended Data Fig. 1), allowing samples to be banked for future processing. After resuming, oil is removed, beads are transferred into a reverse transcription buffer, and full-length cDNA is synthesized, amplified and prepared for sequencing (Fig. 1c,d).

A unique and valuable feature of PIP-seq is that cell encapsulation in droplets is performed in parallel using bead size to control droplet volume. In contrast to microfluidics, the number of droplets scales with total container volume, not emulsification time. For example, at a 6% collision rate that includes cell doublets and barcode reuse, we estimate that 3,500 cells can be barcoded with 35 µl of barcoded hydrogel templates in a 500-µl tube, 225,000 cells can be barcoded with 2 ml of barcoded hydrogel templates in a 15-ml conical tube, and 1 million cells can be barcoded with 10 ml of barcoded hydrogel templates in a 50-ml conical tube (Fig. 1e). Regardless of the tube size, only 2 min of vortexing is required for cell capture. PIP-seq is equally scalable to large sample numbers. Encapsulation can be performed directly in 96-, 384- or 1,536-well plates (Fig. 1f and Extended Data Fig. 2), greatly simplifying experiments testing hundreds of different conditions and streamlining integration with robotic handling systems. Thus, compared to current scRNA-seq technologies, PIP-seq has the greatest flexibility to cover combinations of cell and sample numbers (Fig. 1g).

scRNA-seq with particle-templated emulsification

High-throughput single-cell sequencing requires efficient cell lysis and reverse transcription of mRNA using barcoded primers. In the absence of microfluidics, barcoded hydrogel templates, cells and lysis reagents must be combined before emulsification. To prevent cell lysis before compartmentalization, we use PK, a protease that has minimal activity at 4 °C but can be activated at higher temperatures. After emulsification, the sample is heated to efficiently lyse cells. To illustrate this process, we stained cells with calcein, performed templated emulsification at 4 °C with PK and imaged the droplets before and after thermal activation. Intact cells appeared as compact puncta before lysis but rapidly released calcein into the bulk of the droplets after the temperature was increased (Fig. 2a and Extended Data Fig. 2a,b). Thus, cells can be mixed with PK in bulk before emulsification, and thermal activation triggers the release of mRNA for barcoding after emulsification.

To ensure that temperature-activated lysis and bulk agitation do not prelyse cells and result in mRNA cross-contamination, we

https://doi.org/10.1038/s41587-023-01685-z

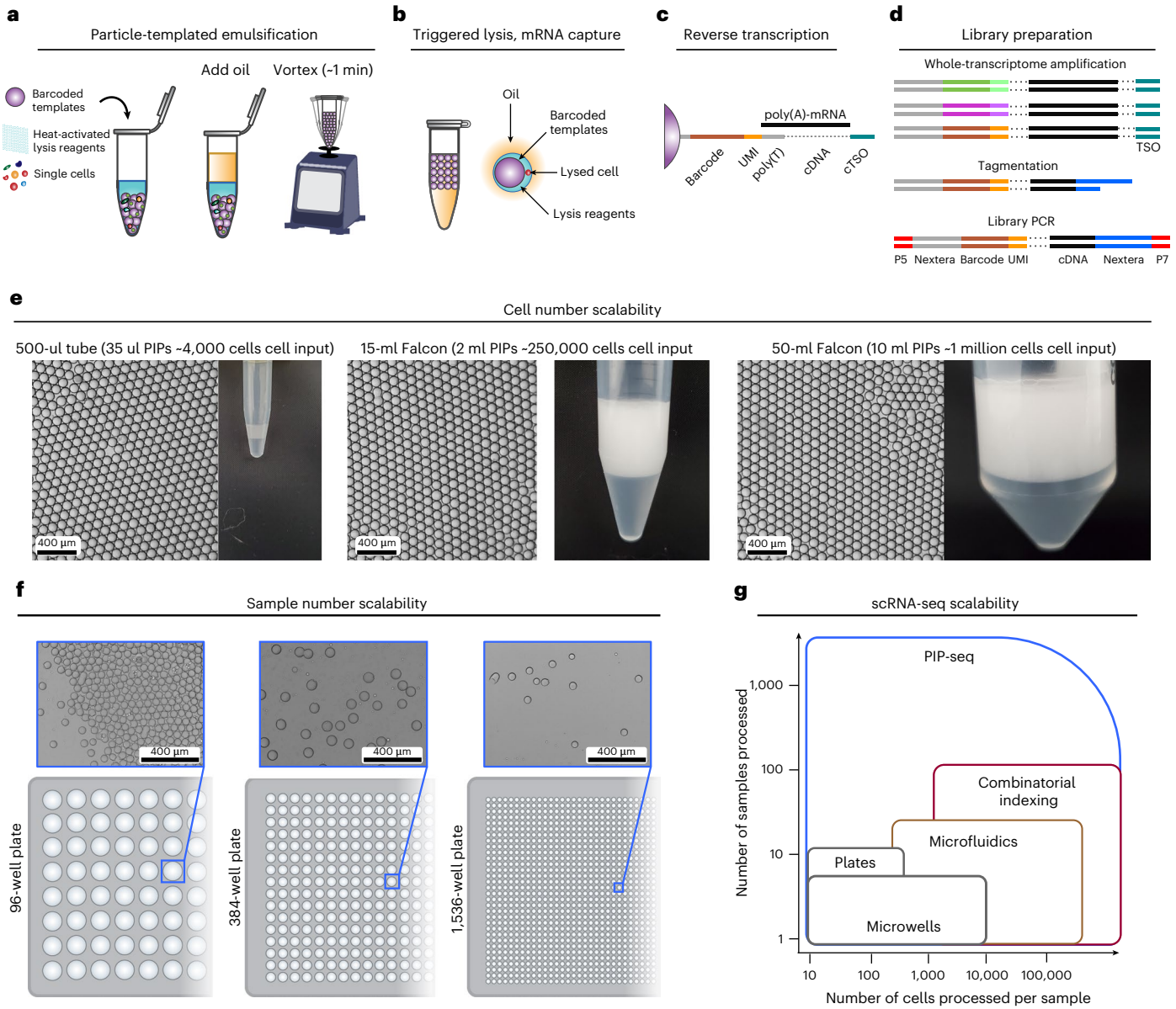


Fig. 1 | Rapid and scalable templated emulsification for single-cell genomics. **a**–**d**, PIP-seq enables the encapsulation, lysis and barcoding of single cells. **a**, Schematic of the emulsification process. Barcoded particle templates, cells and lysis reagents are combined with oil and vortexed to generate monodispersed droplets. **b**, Heat activation of PK results in lysis and release of mRNA that is captured on bead-bound barcoded poly(T) oligonucleotides. **c**, Oil removal is followed by bulk reverse transcription of mRNA into cDNA. cTSO is the complement of the template switch oligonucleotide. **d**, Barcoded whole-transcriptome-amplified cDNA is prepared for Illumina sequencing. **e**–**g**, Efficient single-bead, single-drop encapsulation at scale. **e**, Particle-templated emulsification in different-sized tubes (1.5 ml, 15 ml and 50 ml) produces monodispersed emulsions capable of barcoding orders of magnitude different cell numbers. **f**, PIP-seq is compatible with plate-based emulsification, including 96-, 384- and 1,536-well plate formats. Representative images are shown from experiments completed three times. **g**, The estimated ability of different technologies to easily scale with respect to cell and sample number.

performed mouse–human cell line mixing studies. We synthesized barcoded polyacrylamide beads with poly(T) sequences by using split-pool ligation of four 6-base pair (bp) randomers[30]. Beads contained ~10^8 (96^4) unique barcodes, providing ample sequence space to label 1 million cells. PIP-seq barcode rank plots for mixed mouse–human cell suspensions allowed cell identification by unique molecular identifier (UMI) abundance (Fig. 2b). The fraction of mouse reads in human transcriptomes was below 3%, and transcriptomes containing multiple cells were rare and consistent with Poisson encapsulation of two cells (Fig. 2c,d). These results illustrate that PIP-seq yields high-purity scRNA-seq data with minimal transcriptome mixing and low doublet formation.

Accurate and scalable reconstruction of single-cell phenotypes in complex tissue

An important application of single-cell sequencing is atlasing cell types in heterogeneous tissue. To investigate the feasibility of atlasing studies, we applied PIP-seq to samples derived from healthy breast tissue. In tandem, we performed scRNA-seq on tissues from the same individuals using a commercially available scRNA-seq technology (10x Genomics, Chromium v3). We integrated PIP-seq data across participants and recovered expected cell types by dimensionality reduction, including the two lineages of luminal epithelial cells (LEP1 and LEP2), myoepithelial cells, fibroblasts, vascular cells and immune cells (Fig. 3a and Extended Data Fig. 3a,b)[31]. To compare transcriptome capture between

https://doi.org/10.1038/s41587-023-01685-z

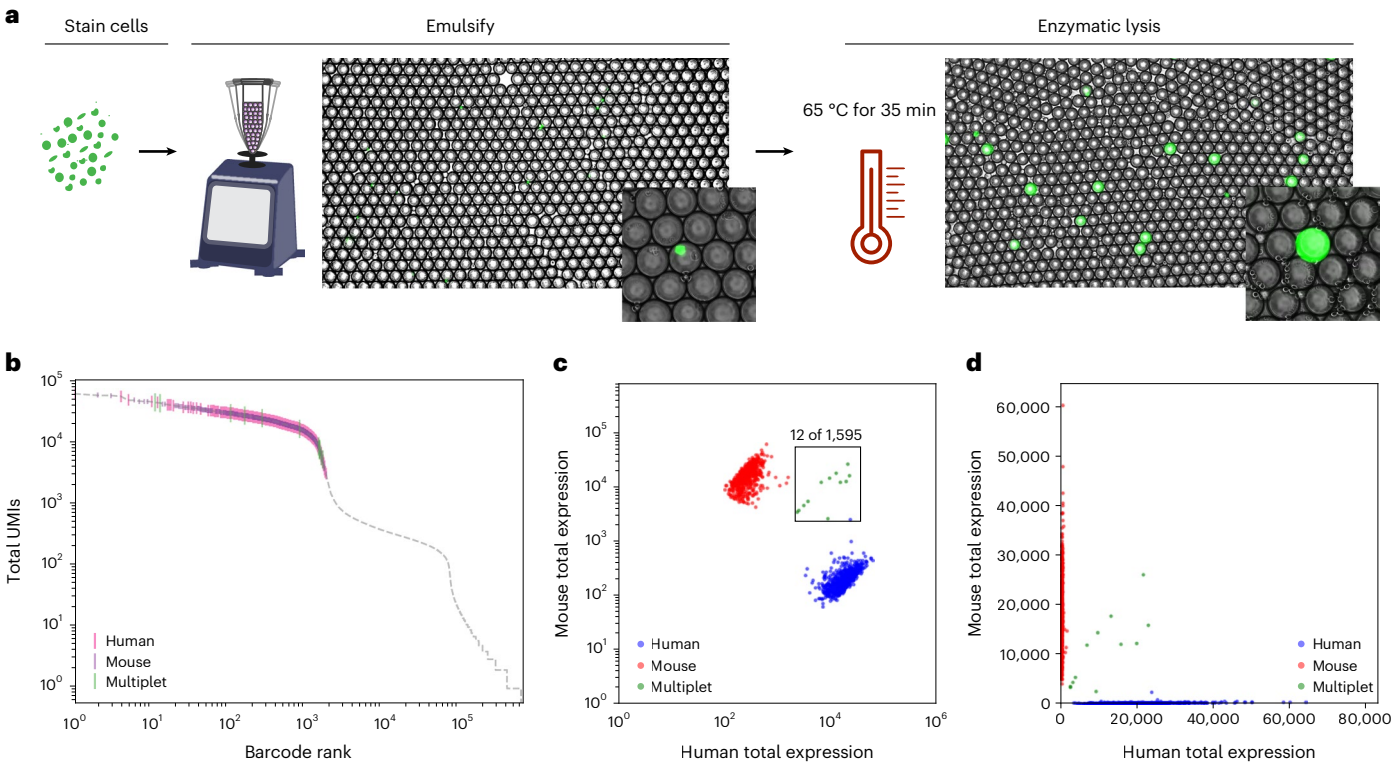


Fig. 2 | Heat-activated enzymatic lysis yields high-purity single-cell transcriptomes. a, Fluorescence microscopy (brightfield and green fluorescent protein) of calcein-stained cells emulsified with barcoded bead templates before and after heat-activated lysis. Inset images show cell puncta (left) and release of calcein (right) after lysis. Representative images are shown from experiments completed at least three times. **b**–**d**, Cell purity assessed with mouse–human mixing studies. **b**, Distribution of total UMIs as a function of cell barcode rank. The gray line represents all barcode groups, with called cells colored by species. **c**,**d**, Purity analysis of cell transcriptomes assessed using barnyard plots. Cells are colored by cell type (red, mouse reads; blue, human reads; green, mixed reads). Representative data are shown from species-mixing experiments completed over ten times.

platforms, we downsampled the 10x Chromium and PIP-seq datasets to an equivalent number of cells and reads (2,400 cells and 36,500 reads per cell). Chromium detected more unique genes (2,298 versus 1,757, median) and transcripts (7,491 versus 3,394) per cell, with similar percentages of reads assigned to mitochondrial transcripts (2.34% versus 1.32%; Extended Data Fig. 3c). To compare the transcriptome accuracy of PIP-seq, we downsampled each dataset to an equivalent number of UMIs per cell (2,400 cells and 1,500 UMIs), integrated the data, performed dimensionality reduction and identified clusters (Fig. 3b,c). We compared marker genes and the correlation between gene expression profiles by cluster. Predicted marker genes were concordant between methods (Fig. 3d), gene expression was highly correlated (Fig. 3e and Extended Data Fig. 4a), and breast tissue markers from previous reports were segregated identically within integrated clusters (Extended Data Fig. 4b). Comparison of PIP-seq to publicly available data from 10x (v3 and v2) and previously published scRNA-seq workflows demonstrated that PIP-seq produced high-quality transcriptomes across a range of sequencing depths (Extended Data Fig. 5). Next, we validated the scalability of PIP-seq, capturing and performing scRNA-seq on 138,146 breast tissue cells in a single-tube reaction and on 65,000 peripheral blood mononuclear cells (PBMCs; Extended Data Fig. 6a–c). At high cell numbers, we identified a population of $CD34^+$ hematopoietic stem/progenitor cells in the PBMC sample, highlighting the importance of scalability in detecting rare cell types (Extended Data Fig. 6b,c). Last, we validated that PIP-seq is compatible with antibody-based cell hashing (Extended Data Fig. 6d,e). Hashing can be used to further increase the number of cells and conditions processed. Thus, PIP-seq is an easy-to-use, accurate and scalable method to profile complex tissues.

PIP-seq for single-cell pooled CRISPR screens

CRISPR perturbations combined with single-cell sequencing allow unbiased discovery of genotype–phenotype relationships[32–34]. Expanding this approach to genome-wide sgRNA libraries can elucidate gene function on an unprecedented scale. However, such studies require sequencing millions of cells to characterize all perturbations in libraries with tens or hundreds of thousands of individual sgRNAs[19]. To demonstrate how the throughput of PIP-seq enables perturbation studies at scale, we profiled the transcriptional changes associated with a CRISPR interference (CRISPRi) allelic series CROP-seq library[35]. This library expressed sgRNA and a polyadenylated copy of the guide sequence from separate promoters. gRNAs were captured and barcoded with the cell's polyadenylated mRNA, making this approach immediately compatible with PIP-seq. The library is designed to quantitatively titrate gene expression using sgRNAs with target site mismatches[35], allowing us to compare measured gene expression to expected knockdown efficiency across each gene's allelic series (Fig. 4a). We transduced K562 cells containing a stable dCas9-KRAB with the CRISPRi lentiviral library and performed PIP-seq to capture the transcriptional profiles and sgRNA identity of individual cells (Fig. 4b,c). For cells with single gRNA assignments, previously reported knockdown efficiencies[35] correlated with the normalized counts of targeted genes (Fig. 4d) and were most significant for highly expressed genes (Extended Data Fig. 7a,b). In addition, the knockdown of genes produced known transcriptional changes. For example, gRNA targeting *HSPA5* resulted in endoplasmic reticulum stress and increased the unfolded protein response (Fig. 4e). These results validate the use of PIP-seq for CROP-seq experiments, paving the way for routine million-cell experiments that map genotype–phenotype relationships at the genome scale.

https://doi.org/10.1038/s41587-023-01685-z

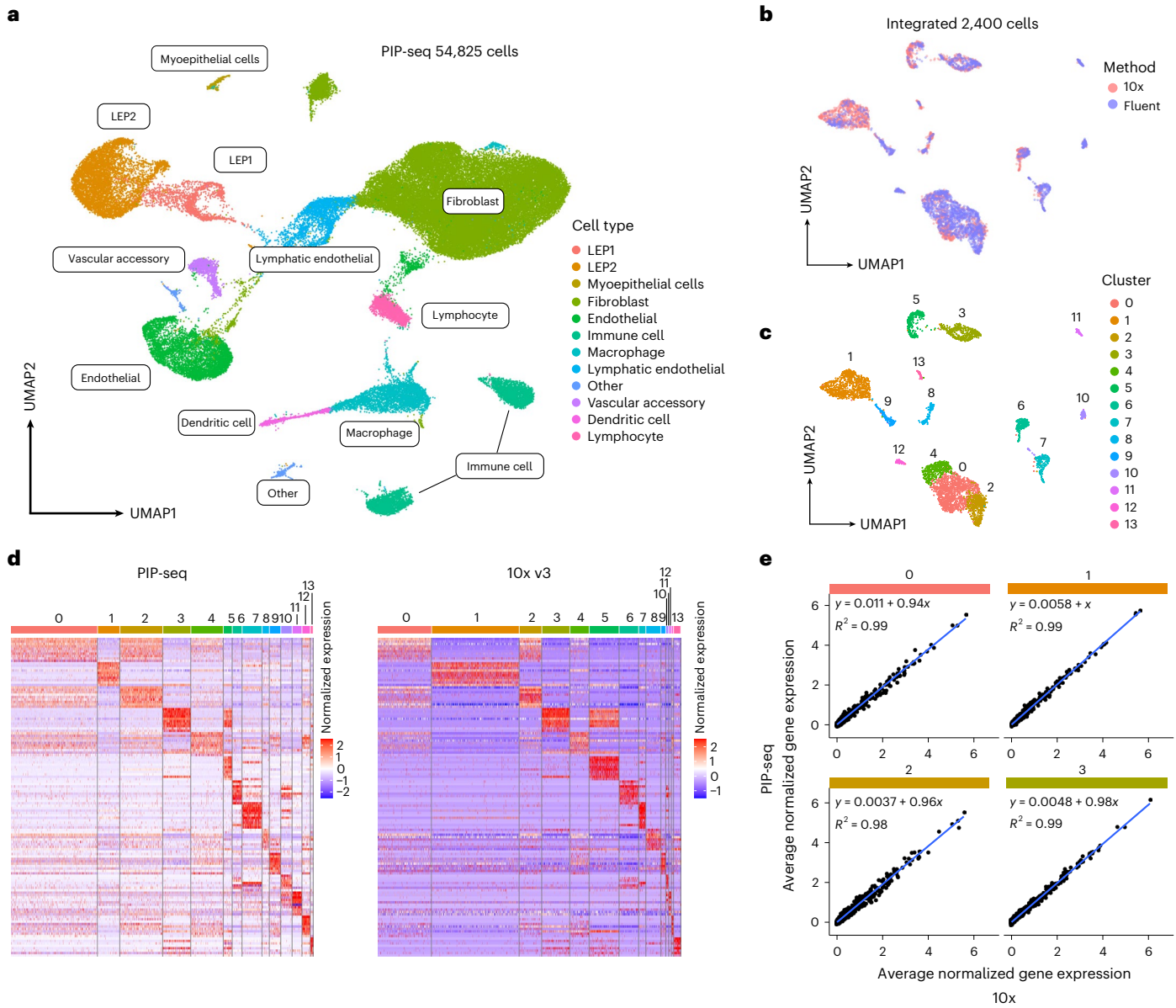


Fig. 3 | Accurate single-cell transcriptional profiling of healthy breast tissue using PIP-seq. a, Clustering and identification of cell types from PIP-seq data (54,825 cells from two individuals). **b**–**e**, Comparison of PIP-seq data to 10x Genomics data collected from the same tissue. **b**, Integration of PIP-seq and 10x data. **c**,**d**, Cell clustering and comparison of marker genes between platforms. **d**, Heat maps of marker gene expression show similar patterns in PIP-seq and 10x data. **e**, Correlations in normalized gene expression, by cluster, between platforms (see also Extended Data Fig. 4a).

Transcriptomic signatures of MPAL relapse

Monitoring of cancer in response to therapy is an emerging application of single-cell sequencing that benefits from rapid sample processing at the point of collection and the ability to delay cDNA synthesis and library preparation until multiple samples have been collected. We investigated the utility of PIP-seq for understanding cancer dynamics by first validating the single-cell transcriptional responses of two cancer cell lines (H1975 and PC9) to gefitinib, an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor. We treated H1975 and PC9 cells with DMSO (vehicle control) or 1 µM gefitinib overnight and performed PIP-seq (Fig. 5a). A transcriptional response in H1975, which is resistant to gefitinib due to EGFR mutations L858R and T790M, was not observed, while gefitinib-sensitive PC9 cells showed a substantial shift in gene expression (Fig. 5b). Differential gene expression analysis revealed increased levels of tumor-associated calcium signal transducer 2 (*TACSTD2*) in PC9 cells, consistent with its known modulation during lung adenocarcinoma tumor growth[36] (Fig. 5c), and decreased expression of cyclin dependent kinase 4 (*CDK4*), which is known to enhance sensitivity to EGFR inhibitors[37] (Extended Data Fig. 8). In addition, drug-resistant H1975 cells spiked into a background of sensitive cells (1:9 H1975:PC9) could be detected solely by their single-cell phenotypes and at roughly the expected frequency (4.7%; Fig. 5d). Thus, PIP-seq recovered genes with reported roles in lung cancer drug resistance and could identify resistance phenotypes within a background of drug-sensitive cells.

Next, we applied PIP-seq to study MPAL, a high-risk disease characterized by multiple hematopoietic lineages[38,39]. Recurrence and changes in immunophenotype with chemotherapy are typically monitored using flow cytometry of surface markers during diagnosis, treatment and relapse, but this provides limited insight into the drivers of relapse after drug treatment. Like other scRNA-seq methods, PIP-seq can be multiplexed to simultaneously characterize single-cell

https://doi.org/10.1038/s41587-023-01685-z

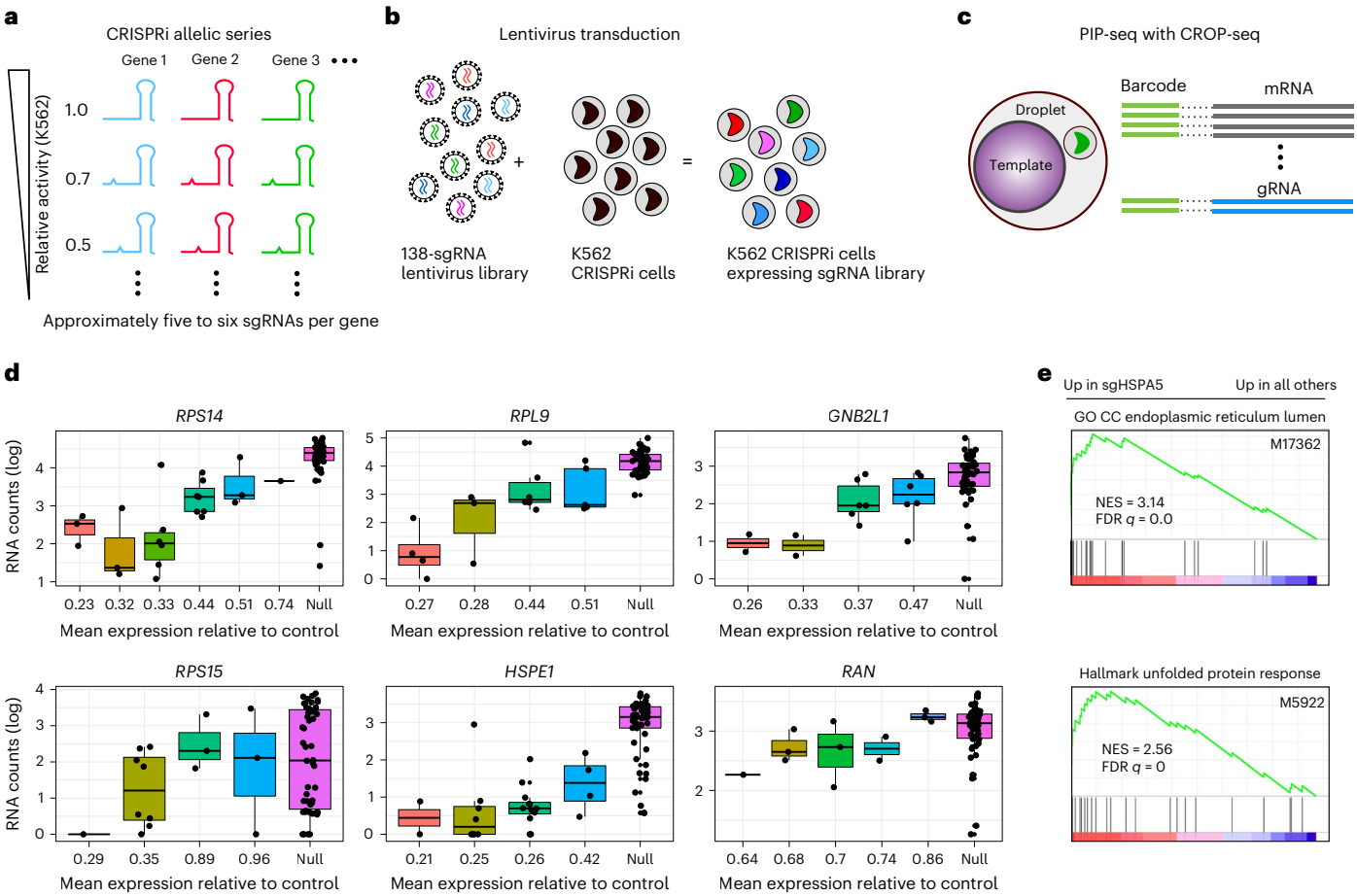


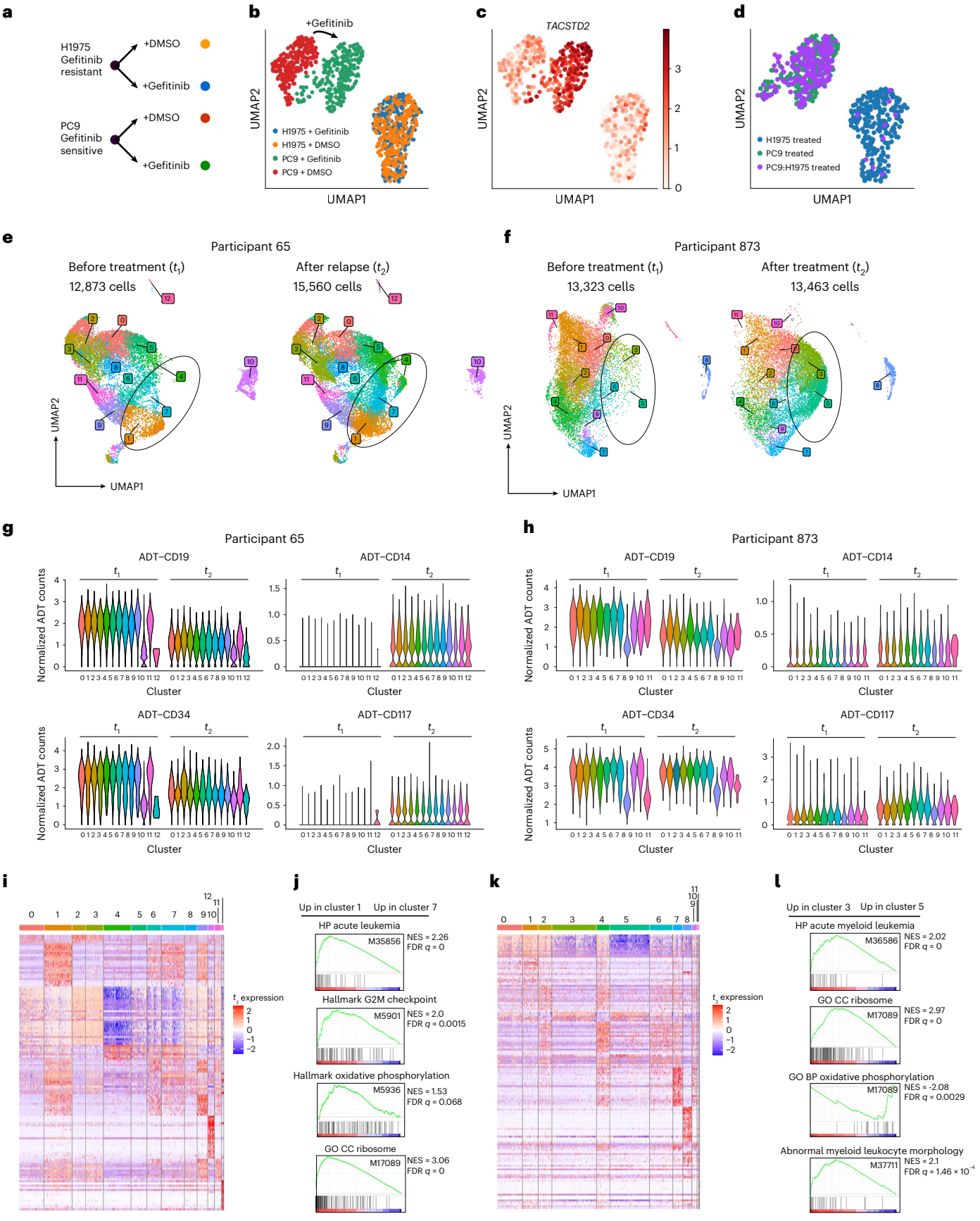
Fig. 4 | Transcriptome and gRNA sequencing using PIP-seq. a, Schematic of the CROP-seq sgRNA library designed with target mismatches to modulate the activity of essential genes. **b**, Lentiviral transduction of the CRISPRi library in K562 cells. **c**, Schematic of the capture and barcoding of polyadenylated mRNA and sgRNA using PIP-seq. RNA and sgRNA libraries are prepared separately and pooled for sequencing. **d**, Quantification of gene expression of sgRNAs within an allelic series. sgRNAs are ordered from high to low predicted knockdown efficiency[35]. Non-targeting sgRNAs are denoted as "Null". Box plots indicate the median, with the lower and upper hinges corresponding to the 25th and 75th percentiles, respectively, and raw data points are displayed (with slight jitter). **e**, Preranked gene set enrichment analysis (GSEA) of scRNA-seq data comparing sgHSPA5-transduced cells to non-sgHSPA5-transduced cells shows enrichment in genes related to endoplasmic reticulum stress and unfolded protein response; GO CC, Gene Ontology cellular component; NES, normalized enrichment score; FDR, false discovery rate.

gene expression and surface immunophenotype[40]. Using PIP-seq, we performed antibody-derived tag (ADT) sequencing (CITE-seq) on longitudinal samples collected from individuals with MPAL treated with chemotherapy. PIP-seq confirmed the diagnosis of these samples as B/myeloid MPAL and identified aberrant expression of immune and stem cell markers that matched with clinical immunophenotypes determined by flow cytometry (Supplementary Table 2 and Extended Data Figs. 9a and 10a). However, PIP-seq revealed an additional layer of complexity undetectable by traditional immunophenotyping. Dimensionality reduction identified cell clusters that emerged after drug treatment (Fig. 5e,f and Extended Data Figs. 9 and 10). These clusters had similar immunophenotypes (Fig. 5g,h) but contained notable transcriptional heterogeneity (Fig. 5i,k and Supplementary Tables 4 and 5). Cell populations upregulating genes and pathways (oxidative phosphorylation, G2M checkpoint modulation and ribosome biogenesis) implicated in a variety of cancers, including acute lymphoblastic leukemia[41–50], but not previously linked to MPAL were observed (Fig. 5j,l). Taken together, our results highlight the value of single-cell methodologies for studying the heterogeneous response of cancer subpopulations to chemotherapy and the potential for the integration of simple and reliable scRNA-seq into clinical workflows.

Discussion

Genomics has progressed rapidly to high-throughput, multimodal single-cell analysis[40,51–54]. Further improvements in data quality, the ability to measure additional cellular properties and new computational approaches for understanding and integrating single-cell information[55–57] will continue to refine our understanding of cell states. At the same time, there remains an unmet need for simplified workflows that scale in cell number and sample size and that allow for breaks in processing after initial sample collection. PIP-seq is a microfluidics-free scRNA-seq method that produces high-quality data using a simplified emulsification technique. Like other high-throughput single-cell approaches, PIP-seq is fundamentally a strategy to barcode mRNA from cells so that material can be pooled and sequenced. The core advantage of PIP-seq is the speed and simplicity of sample processing. Particle-templated emulsification forms monodispersed bead-containing emulsions in minutes with a standard laboratory vortexer, removing the need for instrumentation located in core facilities or hours of multichannel pipetting to perform split-pool indexing in plates. This expands access to single-cell technologies in several ways. First, PIP-seq reduces the need for sample transport, enabling immediate processing by technicians without prior training and collection and banking of samples from remote locations, including field sites.

https://doi.org/10.1038/s41587-023-01685-z



Second, PIP-seq allows infectious samples that require special precautions to be processed at the point of collection or in the biosafety facilities where they are stored. More generally, rapid sample processing eliminates the need for fixatives and minimizes transcriptional perturbations and batch artifacts associated with processing many samples in series.

Fig. 5 | Molecular signatures of drug-resistant cancer phenotypes in cell lines and human samples. a, A two-by-two experimental study design using lung adenocarcinoma cell lines (H1975 and PC9) treated with gefitinib or DMSO. **b**, Clustering of scRNA-seq data after drug treatment shows transcriptional perturbations in gefitinib-sensitive PC9, but not gefitinib-resistant H1975, cells. **c**, Increased expression of *TACSTD2* in PC9 cells challenged with gefitinib. **d**, Identification of drug-resistant H1975 cells spiked into drug-sensitive PC9 cells based on gefitinib-induced transcriptional perturbation. **e**–**l**, PIP-seq RNA and barcoded antibody (CITE-seq) analysis of MPAL. **e**, Clustering of single cells for participant 65 before (left) and after (right) chemotherapy. **f**, Clustering of single cells for participant 873 before (left) and after (right) chemotherapy. **g**,**h**, ADT abundance, by cluster, before (t_1) and after (t_2) chemotherapy. ADTs change as a function of chemotherapy but are consistent among clusters for both participant 65 (**g**) and participant 873 (**h**), with the exception of T cell subsets. **i**–**l**, Analysis of transcriptional heterogeneity in MPAL samples. **i**, Heat map of top differentially expressed marker genes by cluster after relapse in participant 63. **j**, GSEA preranked analysis comparing transcriptomic differences between clusters 1 and 7 in participant 65 using the following gene sets: Human Phenotype acute leukemia (M35856), hallmark G2M checkpoint (M5901), hallmark oxidative phosphorylation (M5936) and Gene Ontology cellular component (GO CC) ribosome (M17089). **k**, Heat map of top differentially expressed marker genes by cluster after relapse in participant 873. **l**, GSEA preranked analysis comparing transcriptomic differences between clusters 3 and 5 in participant 873 using gene sets Human Phenotype acute myeloid leukemia (M36586), Gene Ontology cellular component ribosome (M17089), Gene Ontology biological process (GO BP) oxidative phosphorylation (M17089) and abnormal myeloid leukocyte morphology (M37711).

In addition to workflow simplicity, PIP-seq is intrinsically scalable, handling cell inputs over five orders of magnitude (10 to 10^6), making it well suited for screening genome-wide Perturb-seq experiments and large cell atlas studies. While methods based on combinatorial indexing scale efficiently to large cell numbers, PIP-seq has a simpler workflow and is also compatible with high-throughput processing of samples in plates, allowing many conditions and replicates to be run simultaneously. This has implications for data quality and biological discovery in single-cell experiments because the detection of true positives and reduction in false positives in differential expression analysis is improved by incorporating replicates and statistical methods that account for biological variability[58]. Increased flexibility in the number of samples that can be processed also enables difficult experimental designs, such as dose–response curves, time-course studies, combinatorial perturbations, single-cell sequencing of organoids and large drug screens. In addition, because PIP-seq can directly emulsify in plates, it integrates with robotic fluid handling and therefore comprises a drop-in solution for single-cell readouts in high-throughput experiments in academia or industry.

We confirmed the accuracy of PIP-seq as a single-cell genomics tool by profiling heterogeneous tissue and directly comparing our results to a commercial scRNA-seq platform (10x Genomics). PIP-seq cell-type classification, marker identification and gene expression levels were tightly matched with 10x data but detected fewer genes per cell. We attribute these differences to the extensive optimization that the commercial platform has undergone and suspect that, like other single-cell techniques[3–6,12,25,59], further improvements to PIP-seq molecular biology will increase sensitivity. In addition, because PIP-seq emulsions are functionally equivalent to those made with microfluidics, our approach is immediately compatible with emerging advances, including improvements to the molecular biology of myriad multiomic profiling methods developed for other droplet microfluidic barcoding systems[40,57,60].

Finally, we demonstrated the utility of PIP-seq in processing clinical samples. In combination with barcoded antibodies, we profiled the relapse of MPAL after chemotherapy. MPAL is a subtype of leukemia characterized by poor prognosis[61], lineage ambiguity, lack of consensus regarding therapy and considerable intratumoral genetic and immunophenotypic heterogeneity[62,63]. The molecular mechanisms underlying treatment resistance in this complex disease remain undefined. Changes in gene expression have been linked to prognosis and treatment resistance in multiple cancers. However, tumor heterogeneity makes it unlikely that bulk sequencing methods would identify strong gene signatures associated with resistance in clinical samples. Using PIP-seq of longitudinal samples from two individuals with MPAL with disease progression after initial therapy, we identified transcriptional heterogeneity beyond that observed by immunophenotype and speculate that this heterogeneity may play a role in MPAL treatment resistance. We observed upregulation of genes and pathways previously associated with acute lymphoblastic leukemia in several cell subsets that emerged after chemotherapy and modulation of ribosomal genes in both individuals. Control of translation has been previously implicated in many cancers[41–46,64], including leukemia, but has not yet been linked to MPAL progression and drug resistance, suggesting that therapeutics targeting ribosomal biogenesis and/or protein translation may also have therapeutic potential in MPAL[65]. Our results motivate the use of single-cell technologies for understanding MPAL tumor heterogeneity and response to chemotherapy and suggest that the broad adoption of such technologies for monitoring cancer progression (and tailoring treatment) is within reach. In summary, scRNA-seq provides unparalleled insight into cell heterogeneity but remains underutilized in many settings. PIP-seq addresses this with a simple, rapid and scalable workflow that can be used by any lab containing standard molecular biology equipment.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-023-01685-z.

References

1. Eberwine, J. et al. Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
2. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
3. Hagemann-Jensen, M. et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
4. Hahaut, V. & Picelli, S. Full-length single-cell RNA-sequencing with FLASH-seq. *Methods Mol. Biol.* **2584**, 123–164 (2023).
5. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
6. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
7. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643 (2017).
8. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
9. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Methods* **12**, 44–73 (2016).
10. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
11. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
12. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).

13. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
14. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
15. Tabula Sapiens Consortium* et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
16. Haniffa, M. et al. A roadmap for the Human Developmental Cell Atlas. *Nature* **597**, 196–205 (2021).
17. Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
18. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
19. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).
20. Heath, J. R., Ribas, A. & Mischel, P. S. Single-cell analysis tools for drug discovery and development. *Nat. Rev. Drug Discov.* **15**, 204–216 (2016).
21. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
22. Utada, A. S., Fernandez-Nieves, A., Stone, H. A. & Weitz, D. A. Dripping to jetting transitions in coflowing liquid streams. *Phys. Rev. Lett.* **99**, 094502 (2007).
23. Clark, I. C. & Abate, A. R. Microfluidic bead encapsulation above 20 kHz with triggered drop formation. *Lab Chip* **18**, 3598–3605 (2018).
24. Datlinger, P. et al. Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat. Methods* **18**, 635–642 (2021).
25. Aicher, T. P. et al. Seq-Well: a sample-efficient, portable picowell platform for massively parallel single-cell RNA sequencing. *Methods Mol. Biol.* **1979**, 111–132 (2019).
26. Amini, S. et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014).
27. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
28. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
29. Hatori, M. N., Kim, S. C. & Abate, A. R. Particle-templated emulsification for microfluidics-free digital biology. *Anal. Chem.* **90**, 9813–9820 (2018).
30. Delley, C. L. & Abate, A. R. Modular barcode beads for microfluidic single cell genomics. *Sci. Rep.* **11**, 10857 (2021).
31. Murrow, L. M. et al. Mapping hormone-regulated cell-cell interaction networks in the human breast at single-cell resolution. *Cell Syst.* **13**, 644–664.e8 (2022).
32. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
33. Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
34. Replogle, J. M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
35. Jost, M. et al. Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* **38**, 355–364 (2020).
36. Lin, J. C. et al. TROP2 is epigenetically inactivated and modulates IGF-1R signalling in lung adenocarcinoma. *EMBO Mol. Med.* **4**, 472–485 (2012).
37. Malumbres, M. CDK4/6 inhibitors restore therapeutic sensitivity in HER2$^+$ breast cancer. *Cancer Cell* **29**, 243–244 (2016).
38. Alexander, T. B. et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373–379 (2018).
39. Kotrova, M. et al. Distinct bilineal leukemia immunophenotypes are not genetically determined. *Blood* **128**, 2263–2266 (2016).
40. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
41. Barna, M. et al. Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971–975 (2008).
42. Kang, J. et al. Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. *Signal Transduct. Target. Ther.* **6**, 323 (2021).
43. Kampen, K. R., Sulima, S. O., Vereecke, S. & De Keersmaecker, K. Hallmarks of ribosomopathies. *Nucleic Acids Res.* **48**, 1013–1028 (2020).
44. Fancello, L., Kampen, K. R., Hofman, I. J., Verbeeck, J. & De Keersmaecker, K. The ribosomal protein gene *RPL5* is a haploinsufficient tumor suppressor in multiple cancer types. *Oncotarget* **8**, 14462–14478 (2017).
45. Rao, S. et al. Inactivation of ribosomal protein L22 promotes transformation by induction of the stemness factor, Lin28B. *Blood* **120**, 3764–3773 (2012).
46. De Keersmaecker, K. et al. Exome sequencing identifies mutation in *CNOT3* and ribosomal genes *RPL5* and *RPL10* in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 186–190 (2013).
47. Chen, C. et al. Oxidative phosphorylation enhances the leukemogenic capacity and resistance to chemotherapy of B cell acute lymphoblastic leukemia. *Sci. Adv.* **7**, eabd6280 (2021).
48. Nelson, M. A. et al. Intrinsic OXPHOS limitations underlie cellular bioenergetics in leukemia. *eLife* **10**, e63104 (2021).
49. Lobrich, M. & Jeggo, P. A. The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat. Rev. Cancer* **7**, 861–869 (2007).
50. Didier, C. et al. G2/M checkpoint stringency is a key parameter in the sensitivity of AML cells to genotoxic stress. *Oncogene* **27**, 3811–3820 (2008).
51. Demaree, B. et al. Joint profiling of DNA and proteins in single cells to dissect genotype–phenotype associations in leukemia. *Nat. Commun.* **12**, 1583 (2021).
52. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 44447 (2017).
53. Gaiti, F. et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**, 576–580 (2019).
54. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
55. Hao, Y. et al. Dictionary learning for integrative, multimodal, and scalable single-cell analysis. Preprint at *bioRxiv* https://doi.org/10.1101/2022.02.24.481684 (2022).
56. Ghazanfar, S., Guibentif, C. & Marioni, J. C. StabMap: mosaic single cell data integration using non-overlapping features. Preprint at *bioRxiv* https://doi.org/10.1101/2022.02.24.481823 (2022).
57. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
58. Squair, J. W. et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
59. Hughes, T. K. et al. Second-strand synthesis-based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894 (2020).
60. De Rop, F. V. et al. Hydrop enables droplet-based single-cell ATAC-seq and single-cell RNA-seq using dissolvable hydrogel beads. *eLife* **11**, e73971 (2022).

61. Mejstrikova, E. et al. Prognosis of children with mixed phenotype acute leukemia treated on the basis of consistent immunophenotypic criteria. *Haematologica* **95**, 928–935 (2010).
62. Takahashi, K. et al. Integrative genomic analysis of adult mixed phenotype acute leukemia delineates lineage associated molecular subtypes. *Nat. Commun.* **9**, 2670 (2018).
63. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
64. Ebright, R. Y. et al. Deregulation of ribosomal protein expression and translation promotes breast cancer metastasis. *Science* **367**, 1468–1473 (2020).
65. Khot, A. et al. First-in-human RNA polymerase I transcription inhibitor CX-5461 in patients with advanced hematologic cancers: results of a phase I dose-escalation study. *Cancer Discov.* **9**, 1036–1049 (2019).

Methods

PK-triggered cellular lysis and mRNA capture

Mammalian cells were stained with Calcein AM (Thermo Fisher, C3099) in 1 ml of PBS with 0.04% bovine serum albumin (BSA) according to manufacturer's instructions. After 30 min of incubation at room temperature on a rotisserie incubator (Isotemp, Fisher Scientific), cell suspensions were quantified with a Luna-FL automated cell counter and diluted in 1× PBS with 0.04% BSA. Calcein-stained cells (1,500) in 5 µl of 1× PBS with 0.04% BSA were added to 35 µl of barcoded hydrogel templates with 29 U ml^{-1} PK (NEB, P8107S) and 70 mM DTT (Sigma, D9779) and mixed for 10 pipette strokes. Care was taken to avoid generating bubbles when mixing cells with barcoded hydrogel templates. Two hundred and eighty microliters of 0.5% ionic Krytox in HFE 7500 oil[66] was added to the cell–bead mixture and vortexed at 3,000 r.p.m. for 15 s horizontally and then 2 min vertically with a custom vortexer (Fluent BioSciences, FB0002776). Oil was removed from below the emulsion such that less than 100 µl remained. The PIP emulsion was subsampled on a C-Chip disposable hemacytometer (Fisher Scientific, DHCN015) before lysis, with each subsample consisting of 3.5 µl of PIP emulsion per field of view. The C-chip was imaged in brightfield at ×2 magnification. The remaining PIP emulsion was subjected to enzymatic lysis at 65 °C for 35 min on a PCR thermocycler (Eppendorf Mastercycler Pro) with the lid temperature set to 105 °C. After lysis was complete, fluorescence images were captured using a Nikon 2000 microscope with 470-nm excitation (Thorlab, M470L5).

Synthesis of barcoded bead templates

Prototype barcode bead fabrication proceeded according to previous reports[30]. Briefly, a simple coflow microfluidic device was used to combine acrylamide premix (6% (wt/vol) acrylamide, 0.1% bis-acrylimide, 0.3% (wt/vol) ammonium persulfate, 0.1× Tris-buffered saline–EDTA (TBSET: 10 mM Tris-HCl (pH 8.0), 137 mM NaCl, 20 mM EDTA, 1.4 mM KCl and 0.1% (vol/vol) Triton X-100), 50 µM acrydited primer (/5Acryd/TTTTTTTAAGCAGTGGTATCAACGCAGAGTACGACTCCTCTTTCCCTACACGACGCTCTTCC) with oil (HFE 7500, 3M Novec) containing 2% (wt/vol) surfactant (008-Fluoro-surfactant, Ran Technologies) and 0.4% (vol/vol) tetramethylethylenediamine). The emulsion was solidified at room temperature for 12 h, and beads were removed using 1*H*,1*H*,2*H*,2*H*-perfluoro-1-octanol (Sigma-Aldrich) and washed three times with Tris-EDTA-Tween buffer (TET: 10 mM Tris-HCl (pH 8.0), 10 mM EDTA and 0.1% (vol/vol) Tween 20), followed by two washes with 30 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM $MgCl_2$ and 0.1% Tween 20. The final bead size was 80 µm. Split-pool barcode assembly used the ligation assembly approach as described previously[30]. Beads were resuspended in T4 ligation buffer (NEB, B0202S), heated with a complementary oligonucleotide to 75 °C for 2 min and cooled to room temperature to anneal. One hundred microliters of beads was distributed into each well of a 96-well plate containing a unique barcode with 1× T4 ligation buffer and 1.9 U $µl^{-1}$ T4 DNA ligase (NEB, M0202M). Ligations were incubated at 25 °C for 1 h and heat inactivated at 65 °C for 10 min. Well contents were combined and washed five times in 15 ml of TET. The process was repeated to add four barcodes and a UMI with poly(T) (NNNNNNNNNNNNTTTTTTTTTTTTTTTTTTV). Quality control steps were identical to previous reports[30]. Bead manufacturing methods were transferred to Fluent BioSciences for scaled production, validation and distribution. Commercially produced beads were used for several experiments, as noted.

Varied format emulsification

PIP emulsification in varied formats was performed in 0.5-ml microcentrifuge tubes, 15-ml conical tubes and 50-ml conical tubes. Briefly, PIP particles were suspended in buffer with 29 U ml^{-1} PK (NEB, P8107S) and 70 mM DTT (Sigma, D9779) and pelleted through centrifugation. Barcoded hydrogel templates were then distributed at 35-µl, 0.5-ml and 8-ml volumes in 0.5-ml, 15-ml and 50-ml tubes, respectively. Fluorinated oil with surfactant (Fluent Biosciences, FB0001804) was added to each tube at 200-µl, 8-ml and 32-ml volumes, respectively. Emulsification was conducted on a Vortex Genie 2 with a custom adapter (Fluent, FBS-SCR-8VX) at maximum r.p.m. for 1 min. After emulsification, the samples were allowed to settle for 30 s, and excess oil was removed via syringes using 22-gauge blunt needles. The emulsion was subsampled, loaded on a C-Chip disposable hemacytometer (Fisher Scientific, DHCN015) and imaged under brightfield microscopy (DIAPHOT300, Nikon) at ×2 and ×4 magnification.

Emulsification in well plates was tested using two bead buffer conditions. First, to test emulsification in 96-, 384- and 1,536-well plates, PIP particles were suspended in 2% (vol/vol) Triton X-100 (Sigma, X100-5ML) in 10 mM Tris-HCl (Teknova, T1075) and centrifuged at 6,000*g*; the supernatant was then removed (Fig. 1 and Extended Data Fig. 2c). Depending on the well plate working volume, 38 µl, 8 µl or 3 µl of the centrifuged barcoded hydrogel templates was added to 96-, 384- or 1,536-well plates, respectively. For 96- and 384-well plates, 2 µl of sample was added to each well, and for 1,536-well plates, 1 µl was added to each well. PIP and sample volumes totaled 25% of the volume of each well. Each plate type was then sealed (Applied Biosystems, 4306311) and shaken for 5 min (IKA, 253614 and 3426400) to ensure complete mixing. Each plate type was centrifuged at 200*g* for 1 min before removing the seal. Then, 80 µl, 20 µl or 8 µl of 2% (wt/wt) fluorosurfactant (Ran BioTechnologies, 008 Fluorosurfactant) in HFE oil (3M, Novec 7500) was added to each well in 96-well (Applied Biosystems, N8010560), 384-well (Applied Biosystems, A36931) or 1,536-well (Nunc, 253614) plates, respectively. The addition of oil represented 50% of the volume of each well for a total volume of 75% consisting of PIP, sample and oil. After resealing, PIP emulsification was performed by vortexing for 30 s at 3,200 r.p.m. (Benchmark Scientific, BV1003). The emulsified plate was centrifuged at 200*g* for 1 min before removing the seal and imaging droplets from individual wells on a fluorescence microscope (EVOS FL Auto).

Second, to test well plate emulsification with cells in 96- and 384-well plates, PIP particles were suspended in buffer with 29 U ml^{-1} PK (NEB, P8107S) and 70 mM DTT (Sigma, D9779) and pelleted through centrifugation. For 96-well plates (Eppendorf, 0030129300), 25 µl of barcoded hydrogel templates was then distributed into each well with 4,000 cells per well (2,000 cells per µl × 2 µl). Fluorinated oil with surfactant (150 µl; Fluent Biosciences, FB0001804) was added to each well. Emulsification was conducted on a Vortex Genie 2 with a flat-head adapter at 3,000 r.p.m. for 2 min. For 384-well plates (Corning, 3347), 15 µl of barcoded hydrogel templates was then distributed into each well with 3,000 cells per well (2,000 cells per µl × 1.5 µl). Fluorinated oil with surfactant (105 µl; Fluent Biosciences, FB0001804) was added to each well. Emulsification was conducted on a Vortex Genie 2 with a flat-head adapter at 3,000 r.p.m. for 2 min (Fig. 1 and Extended Data Fig. 2a,b).

PIP-seq protocol

Unless otherwise noted, cells were centrifuged at 300*g* for 5 min, washed twice in 1× PBS without calcium or magnesium (Thermo Fisher, 70011044) with 0.04% BSA, filtered with a 70-µm cell strainer and resuspended in 1× PBS with 1% Pluronic F127 (Sigma, P2443). Prealiquoted barcoded hydrogel templates were thawed on ice. Volumes of barcoded hydrogel templates, cells and oil varied based on the number of cells as noted in each experimental subsection below. The following protocol was used for a standard small-format run: 5 µl of 500 cells per µl was added to 35 µl of barcoded hydrogel templates with 29 U ml^{-1} PK and 70 mM DTT (Fluent BioSciences, FB0001876) and mixed for 10 strokes. Care was taken to avoid generating bubbles when mixing cells with barcoded hydrogel templates. Oil (280 µl; Fluent Biosciences, FB0001804) was added to the cell–bead mixture and vortexed (Vortex Genie 2, Scientific Industries) using a custom adapter (Fluent BioSciences, FB0002100) at the maximum r.p.m. for 15 s horizontally and

2 min vertically. Excess oil (230 µl) was removed, and the emulsion and enzymatic lysis was completed at 65 °C for 35 min with a 4 °C hold on a PCR thermocycler with the lid temperature set to 105 °C. The remaining oil was removed. The emulsion was broken using the following protocol. Using a multichannel pipette, 180 µl of room temperature high-salt buffer (250 mM Tris-HCl (pH 8), 375 mM KCl, 15 mM $MgCl_2$ and 50 mM DTT) was added to the top of the emulsion followed by 40 µl of 100% *1H,1H,2H,2H*-perfluoro-1-octanol (Sigma-Aldrich, 370533). The samples were vortexed for 3 s and briefly centrifuged, and the bottom oil phase was removed. Barcoded hydrogel templates were transferred into a 1.5-ml Eppendorf tube and washed three times with 2× RT buffer (100 mM Tris-HCl (pH 8.3), 150 mM KCl, 6 mM $MgCl_2$ and 20 mM DTT) with 1% Pluronic F68 (Gibco, 24040032). After washing, the beads were pelleted, the aqueous layer was removed, and the remaining bead and buffer volume was 25 µl. To this bead buffer mixture, 25 µl of reverse transcription master mix comprising 4.8% PEG8000, 4% PM400, 2.5 µM template switch oligonucleotide (PIPS_TSO), 1 mM dNTPs (NEB), 1 U μl^{-1} RNase inhibitor (NxGen, Lucigen) and 1 U μl^{-1} reverse transcriptase (Thermo Fisher, Maxima H-minus EP0751) was added. The reaction was thoroughly mixed, and cDNA synthesis was completed for 30 min at 25 °C and 90 min at 42 °C, followed by 10 min at 85 °C and a 4 °C hold. Whole-transcriptome amplification (WTA) was performed directly on reverse transcription product without purification by adding 50 µl of 2× KAPA HiFi master mix and 0.25 µM primer (PIPS_WTA_primer) and thermocycling (95 °C for 3 min, 16 cycles of 98 °C for 15 s, 67 °C for 20 s and 68 °C for 4 min, followed by 72 °C for 5 min and a hold at 4 °C). After WTA, barcoded hydrogel templates were removed using Corning Spin-X filter columns (1 min at 13,000*g*), and amplified cDNA was purified using 0.6× Ampure XP. Libraries were generated from WTA amplified material using the Nextera XT DNA library preparation kit with a custom primer (PIPS_P5library) and standard Nextera P7 indexing primers (N70x). Libraries were pooled and sequenced using an Illumina NextSeq 2000 instrument with 15% PhiX. Oligonucleotides used in this study are supplied in Supplementary Table 1.

Human–mouse mixing studies

Human HEK 293T cells (ATCC, CRL-3216) were grown in DMEM (Thermo Fisher, 11995073) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher, A3840001) and 1% penicillin–streptomycin–glutamine (Thermo Fisher, 10378016). Mouse NIH 3T3 cells (ATCC, CRL-1658) were grown in DMEM (Thermo Fisher, 11995073) supplemented with 10% bovine calf serum (ATCC, 30-2030) and 1% penicillin–streptomycin–glutamine. Cells were grown to a confluence of ~70% and treated with TrypLE Express with Phenol red (Thermo Fisher, 12605010) for 3 min, quenched with an equal volume of growth medium and centrifuged for 5 min at 200*g*. The supernatant was removed, and the cells were resuspended in 1× DPBS without calcium or magnesium. Cells were diluted to their final concentration in 1× DPBS with 0.04% BSA and mixed evenly to create a 50:50 human:mouse mixture. Cell viability was evaluated using acridine orange/propidium iodide stain (Logos Bio, F23001) and quantified with a Luna-FL automated cell counter. Cells were processed using the PIP-seq protocol as described above.

Seventy-two-hour hold experiments

Five microliters of a 50:50 mixture of human HEK 293T cells and mouse NIH 3T3 cells (800 cells per µl) was added to 35 µl of barcoded hydrogel templates (Fluent BioSciences, FB0003067) with 29 U ml^{-1} PK and 70 mM DTT and mixed for 10 strokes. Oil (280 µl; Fluent Biosciences, FB0001804) was added to the cell–bead mixture, which was vortexed on a digital vortexer using a custom adapter (Fluent BioSciences, FB0002084) at 3,000 r.p.m. for 15 s horizontally and 2 min vertically. Excess oil (230 µl) was removed, and the emulsion was placed in a preheated digital dry bath at 66 °C for 38 min and 4 °C for 11 min. Control samples proceeded to emulsion breaking, while 0 °C hold samples were placed in an ice bucket in the refrigerator (4 °C) for 72 h before breaking emulsions. Breaking, mRNA extraction, reverse transcription, WTA and cDNA isolation, adapter ligation-based library preparation and Illumina sequencing were performed as previously described.

Healthy breast tissue comparison to 10x data

Fresh reduction mammoplasty tissue was processed as previously described[31,67]. Use of breast tissue specimens to conduct the studies described was approved by the University of California San Francisco Committee on Human Research under Institutional Review Board protocols 16-18865 and 10-01532. Tissues were obtained as deidentified samples, and all participants provided written informed consent. Bulk mammary tissues were mechanically processed into a slurry and digested overnight with collagenase type 3 (200 U ml^{-1}, Worthington Biochem CLS-3) and hyaluronidase (100 U ml^{-1}; Sigma-Aldrich, H3506) in medium containing charcoal:dextran-stripped FBS (GeminiBio, 100-119). The digested fragments were size filtered into a below-40-µm fraction and an above-100-µm fraction and cryopreserved. For PIP-seq, cells were thawed and resuspended in PBS + 0.04% BSA and passed through a 70-µm FlowMi cell strainer (Sigma, BAH136800070). For 10x Genomics data, the 100-µm fraction was thawed and further digested with trypsin, followed by dispase (Stemcell Technologies, 07913) and DNaseI (Stemcell Technologies, 07469) digestion to achieve single-cell suspensions. For PIP-seq, 20 µl of cells (1,500 cells per µl in PBS + 0.04% BSA) was added to 200 µl of barcoded hydrogel templates (Fluent BioSciences, FB0002617) and mixed for 10 strokes. Oil (1,000 µl; Fluent Biosciences, FB0001804) was added to the cell–bead mixture and vortexed on a digital vortexer using a custom adapter (Fluent BioSciences, FB0002100) at 3,000 r.p.m. for 15 s horizontally and 2 min vertically. Excess oil (800 µl) was removed, and the emulsion was placed on a preheated digital dry bath at 66 °C for 38 min and 4 °C for 11 min. Breaking, mRNA extraction, reverse transcription, WTA and cDNA isolation were performed under standard conditions. Adapter ligation-based library preparation was performed according to manufacturer's instructions (Watchmaker Genomics, 7K0019-024). Samples were sequenced on an Illumina NextSeq 2000, with four participant samples pooled per P3 cartridge, and sequenced at a read depth of approximately 36,500 reads per cell. For 10x Genomics, cells from each participant were labeled with MULTIseq barcodes[13] and were pooled and stained with DAPI to be sorted for DAPI-live cells. Single-cell libraries were prepared according to the 10x Genomics Single Cell V3 protocol (v3.1 Rev D) with the standard MULTIseq sample multiplexing protocol. The libraries were sequenced on a NovaSeq S4 lane at a read depth of about 70,000 reads per cell. To compare platforms, we downsampled PIP-seq and 10x data, which had different numbers of cells and sequencing depth per cell. The PIP-seq data had 54,825 cells, sequenced at approximately 36,500 reads per cell, while the 10x data had 2,420 cells sequenced at approximately 70,000 reads per cell. Data were downsampled to 2,400 cells and 36,500 reads in R (downsampleReads, DropletUtils). For correlation and marker gene comparisons, data were downsampled to 2,400 cells and 1,500 UMIs in R (SampleUMI, Seurat v4.1.0). Markers used for breast tissue cluster cell-type calling are available in Supplementary Table 2.

Single-tube large-format breast tissue study

PIP-seq was performed as previously described, except that cells were counted and diluted with PBS + 0.04% BSA to a concentration of 10,000 cells per µl. Cell suspension (40 µl) was added to 800 µl of barcoded hydrogel templates (Fluent BioSciences, FB0003067). Oil (4,000 µl; Fluent Biosciences, FB0001804) was added to the cell–bead mixture and vortexed on a digital vortexer using a custom adapter (Fluent BioSciences, FB0002659) at 3,000 r.p.m. for 15 s horizontally and 2 min vertically. Excess oil was removed using a 3-ml syringe with a 22-gauge blunt-bottom syringe needle. Lysis proceeded using 3,300 µl of a lysis emulsion (Fluent BioSciences, FB0003039) added to the cell–bead emulsion. The mixture was placed in a preheated digital dry bath at

37 °C for 45 min and 4 °C for 10 min. Breaking, mRNA extraction, reverse transcription, WTA and cDNA isolation were performed under the same conditions as described previously. Adapter ligation-based library preparation was performed according to manufacturer's instructions (Watchmaker Genomics, 7K0019-024). cDNA (80 ng) was used to prepare four replicate library preparations, which were pooled and sequenced on two Illumina NextSeq 2000 P3 cartridges at a read depth of 13,025 reads per cell after concatenation.

CROP-seq

K562 CRISPRi cells were cultured in RPMI-1640 (Gibco, 11875093) with 10% FBS (Thermo Fisher Scientific, 10438026) and 1% penicillin–streptomycin (Thermo Fisher Scientific, 15140148) in an incubator at 37 °C with 5% CO_2. K562 CRISPRi cells were transduced with a lentivirus library containing 138 sgRNAs[35] at a multiplicity of infection of 0.1. Lentivirus-infected cells (BFP^+) were sorted to high purity using a BD FACS Aria III (100-µm nozzle) and processed according to the PIP-seq scRNA-seq workflow. Cells (3 µl; 333 cells per µl) were added to 28 µl of barcoded hydrogel templates with 29 U ml^{-1} PK and 70 mM DTT and mixed for 10 strokes. One hundred and fifty microliters of 0.5% ionic Krytox in HFE 7500 oil was added to the cell–bead mixture and vortexed at 3,000 r.p.m. for 1 min on a Vortex Genie 2 with a custom tube adapter. cDNA was processed according to the standard PIP-seq protocol to obtain sequence-ready libraries containing transcriptome information. To recover sgRNA sequences, we implemented an additional amplification step. We amplified 1 ng of cDNA in a 50-µl reaction using primers P5-PE1 (0.5 µM) and Weissman_U6 (0.25 µM; Supplementary Table 1) with 1× Kappa HiFi. Reactions were thermocycled at 95 °C for 3 min followed by 10 cycles of 95 °C for 20 s, 70 °C for 30 s (−0.2 °C per cycle) and 72 °C for 20 s, followed by 8 cycles of 95 °C for 20 s, 68 °C for 30 s and 72 °C for 20 s, followed by 72 °C for 4 min and hold at 4 °C. Library PCR product enriched in sgRNA sequences was purified with a double-sided 0.5×/0.8× Ampure XP bead cleanup, and the size was determined (Agilent Tapestation).

Transcriptome and sgRNA libraries were pooled at 20:1 before sequencing. Reads were first processed to extract sgRNA sequences. The bioinformatics pipeline was run using a custom index built from the full human transcriptome (GENCODE v32) and gRNA sequences (Salmon v1.2.0.). This approach led to the recovery of >14,000 unique gRNA counts across all cell-associated barcodes. Cells were assigned to gRNA groups using a previously reported approach[32]. Briefly, cells were classified as uniquely expressing a single gRNA species if the guide's expression was at least tenfold higher than the sum of all other gRNAs. Similarly, cells were classified as containing multiple gRNAs in cases where the difference was smaller than 1. For the 581 single cells sequenced, 2 did not have any gRNA, 441 contained a single gRNA, and 138 contained multiple gRNAs. Cell barcodes were processed using Seurat v4.1.0. All gRNAs in the list of features were excluded from the identification of variable transcripts (feature selection) and in subsequent stages of dimensionality reduction and clustering. To understand the relationship between gRNAs and mRNA expression, gRNAs were ranked according to their expected level of knockdown, as reported previously[35], and a generalized additive model was used to assess groupwise trends for each set of gRNAs.

Lung adenocarcinoma cell line experiments

PC9 cells were obtained from the RIKEN Bio Resource Center (RCB4455). H1975 cells were obtained from ATCC (CRL-5908). Cells were cultured in RPMI-1640 (Gibco, 11875093) with 10% FBS, penicillin and streptomycin in an incubator at 37 °C with 5% CO_2. Gefitinib (1 µM; Frontier Scientific, 501411677) or DMSO was added to culture flasks 24 h before cells were collected for processing. PC9 and H1975 cells were both treated with gefitinib and DMSO. To perform the cell mixing study, gefitinib-treated H1975 cells and gefitinib-treated PC9 cells were mixed at a ratio of 1:9 H1975:PC9. Five microliters of cells (400 cells per µl) was added to 28 µl of barcoded hydrogel templates with 22.8 U ml^{-1} PK and 28 mM DTT and mixed for 10 pipette strokes. One hundred and fifty microliters of 0.5% ionic Krytox in HFE 7500 oil[66] was added to the cell–bead mixture and vortexed at 3,000 r.p.m. for 1 min on a Vortex Genie 2 with a custom tube adapter. Triplicate tubes of 400 cells were processed per treatment condition. Data were analyzed using Seurat v4.1.0.

Healthy PBMCs

Cryopreserved PBMCs were obtained from a commercial provider (AllCells). Cells were thawed and prepared for PIP-seq as previously described in the MPAL study, except that the final cell dilution was made in 1× PBS + 0.04% BSA. For the high-cell-count PBMC study, PIP-seq was performed as previously described in the high-cell-number breast tissue study except that cells were counted and diluted with PBS + 0.04% BSA to a concentration of 4,300 cells per µl, and 44 µl of cell suspension was added to 800 µl of barcoded hydrogel templates (Fluent BioSciences, FB0003067). Cryopreserved PBMCs used for cell hashing were obtained from a commercial provider (AllCells) and prepared for PIP-seq as described previously. For the cell hashing study, cell staining and PIP-seq were performed according to the PIP-seq Single Cell Epitope Sequencing user guide (FB0002079). Briefly, 1 million PBMCs were resuspended in 47.5 µl of cell staining buffer (BioLegend, 420201), and 2.5 µl of TruStain FcX block (BioLegend, 422301) was added before mixing and incubating for 10 min on ice. Next, 1 µg of TotalSeqA antibody was diluted in cell staining buffer, and 50 µl of this antibody dilution was added to the blocked cells before incubation on ice for 30 min. Stained cells were washed in cell staining buffer three times and resuspended in 1× PBS + 0.04% BSA at 2,000 cells per µl. For PIP-seq, 20 µl of this cell resuspension was added to 200 µl of barcoded hydrogel templates (Fluent BioSciences, FB0002617) and processed through PIP-seq.

MPAL

Participants whose samples were used in this study were treated at the University of California San Francisco. Samples were collected in accordance with the Declaration of Helsinki under Institutional Review Board-approved tissue banking protocols, and written informed consent was obtained from all participants. Sample clinical characteristics are available in Supplementary Table 3. Cryopreserved PBMCs were thawed by hand until approximately 85% of ice remained. Using a 5-ml serological pipette, 1 ml of 4 °C defrosting medium (DMEM with 20% FBS and 2 mM EDTA) was added dropwise to each sample, and, without disturbing the remaining ice pellet, the sample was carefully transferred dropwise to a preprepared 40-ml aliquot of 4 °C defrosting medium. This was repeated until the contents of the entire cryovial were transferred into the 50-ml conical of defrosting medium. The sample was inverted four to five times and centrifuged at 114*g* for 15 min at 4 °C with no brake. The supernatant was aspirated, and 10 ml of room temperature RPMI-1640 with 1% penicillin–streptomycin–glutamine was used to gently resuspend the cells. Cell clumps were manually removed, and, if necessary, cells were filtered through a 70-µm cell strainer into a fresh 50-ml conical. The sample was inverted two to three times and centrifuged at 114*g* for 10 min with low brake at room temperature. The supernatant was aspirated, and cells were resuspended in an appropriate volume of 1× PBS + 5% FBS. Cells were quantified with Acridine Orange (AO)/Propidium Iodide (PI), and viability was evaluated on the Luna-FL. One to 2 million cells were aliquoted into a new 15-ml conical tube and centrifuged at 350*g* for 4 min at 4 °C, the supernatant was aspirated, and the tube was placed on ice. Forty-five microliters of cold cell staining buffer (BioLegend, 420201) was added per 1 million cells and resuspended gently. Five microliters of Trustain FcX block (BioLegend, 422301) was added per 1 million cells and gently mixed 10 times with a wide-bore pipette tip. Cells were blocked on ice for 15 min. A custom pool of 19 TotalSeqA antibodies was obtained from BioLegend and diluted according to the manufacturer's instructions.

Immediately before use, antibodies were mixed and centrifuged at 10,000*g* for 4 min at 4 °C; 4.6 µl of 0.5 µg µl^{-1} antibody pool was added per 1 million blocked cells and gently mixed 10 times with a wide-bore pipette tip. The samples were incubated on ice for 60 min. Next, 3.5 ml of cold cell staining buffer was added, gently mixed with a wide-bore pipette tip and slowly inverted twice to mix. Cells were centrifuged at 350*g* for 4 min at 4 °C, and the supernatant was removed. The addition of cold cell staining buffer was repeated twice for a total of three washes. After the final supernatant aspiration, stained cells were resuspended in 1× PBS with 0.04% BSA and mixed five to ten times until cells were completely suspended without visible clumps. Cell concentration was determined with AO/PI, and viability was evaluated on a Luna-FL. Final dilutions were made in 1× PBS with 0.04% BSA. Twenty microliters of cells was added to 200 µl of barcoded hydrogel templates (1,000 cells per µl) and processed according to the PIP-seq Single Cell Epitope Sequencing user guide (FB0002079). Marker genes identified for participants 65 and 873 are available in Supplementary Tables 4 and 5, respectively. Clinical FACS data from participants 65 and 873 were analyzed with FlowJo.

PIP-seq bioinformatic analysis

Analysis of sequencing data was performed using custom scripts to generate gene expression matrices starting from processed FASTQ sequences. The pipeline is composed of four basic steps: (1) barcode identification and error correction, (2) mapping to reference sequences, (3) cell calling and (4) gene expression matrix generation. Briefly, after demultiplexing the sequencing data, each read in the FASTQ is matched against a 'whitelist' of known barcodes. Reads were matched with a hamming distance tolerance of 1, meaning that the barcode portion of a read can differ from a whitelist entry by one base and can still be matched to that barcode. Reads that did not match any barcode in the whitelist were discarded from further analysis. Matching reads were output to a new intermediate FASTQ file that was then used for mapping against an appropriate transcriptome reference. Reference transcriptomes matching the species of each sample were prepared using the Salmon 'index' function with the default *k*-mer size of 31 (ref. [68]). GENCODE references were used to build the transcriptome indexes, including GRCh38.p13 for human, GRCm38.p6 for mouse and the combination thereof for HEK 293T/NIH 3T3 cell mixture studies. Following barcoding, Salmon 'alevin' v1.2.0 (ref. [69]) was used to map reads to the full transcriptome. The intermediate FASTQ files generated during barcoding were provided as input into alevin along with a list of all whitelisted barcodes contained in raw reads. After mapping, data were output as UMI count matrices (sparse matrix, gene list and barcode list) with dimensions of 'all barcodes x all genes in index'. An in-house Python implementation of emptyDrops[70], a standard scRNA-seq method to separate putative cells from background, was then applied. A custom threshold for each experiment was set, beneath which no true cell barcodes were expected to fall. As with emptyDrops, an estimated ambient profile across all barcodes beneath that threshold was created. A *P* value was computed by comparing the gene expression profile for each barcode above the threshold against the ambient profile. Barcodes with a statistically significant difference (Benjamini–Hochberg-adjusted *P* value of <0.001) from the ambient background profile were categorized as cell-containing barcodes. The alevin output matrices were then subset to only include called cell barcodes. Gene expression matrices were normalized before performing unsupervised clustering and uniform manifold approximation and projection (UMAP) dimensionality reduction. Gene expression counts for each cell were first divided by the total counts for that cell and multiplied by a scaling factor of 10,000. The data were then transformed to natural log scale using log1p(). The Seurat package (v4.1.0) was used to perform downstream clustering, marker gene determination and visualization in R. Seurat's FindClusters() and RunUMAP() commands were used with default settings.

For saturation curve comparisons, PIP-seq and 10x samples were downsampled to matching depths of 5,000–80,000 reads per called cell. Downsampling was performed using seqtk for PIP-seq samples and using the DropletUtils read10xMolInfo() function with a molecule_info.h5 file directly downloaded from the 10x website. Inflection point-based cell calling was used to standardize cell calls across platforms. Median transcripts per cell and genes per cell values were calculated from the cell fraction of the resulting count matrices. For violin plot comparisons, samples were prepared to match the same processing configuration used by Ding et al.[28]. Samples were first downsampled to 53,000 reads per called cell and trimmed to 50 bp for read 2 before processing, sampling in the same manner described above. Each violin plot represents the cell fraction from a single replicate of an HEK 293T/NIH 3T3 cell mixture, with human and mouse split out into separate plots.

Analysis of PBMC data for the high-cell-count study was performed using custom scripts, as described above, until the completion of mapping. Cell calling, clustering and differential expression were performed using PIPseeker v1.0.0 (Fluent Biosciences) in 'reanalyze' mode using –force-cells 65000. The top differentially expressed genes from the PIPseeker graph-based clustering result were used to determine cell types by comparing to a reference gene list (Supplementary Table 7). The log-normalized expression values for key genes (for example, *CD34*) were overlaid on the UMAP projection to highlight markers associated with specific cell types (color bars are in $\log_{10}$ scale). Analysis of PBMC data for the cell hashing study was performed using PIPseeker v1.0.0 in 'count' mode using STAR (v2.7.10a) and the PIPseeker human reference (https://www.fluentbio.com/products/pipseeker-for-data-analysis/). ADT analysis was conducted by performing barcode error correction with PIPseeker v1.0.0 (count mode) and custom scripts to trim read two to the first 16 bp. Error-corrected and trimmed FASTQ files were input to CITE-seq Count (v1.4.3) using the following settings: -t (hashtag whitelist) -cbf 1 -cbl 16 -umif 17 -umil 28 –cells (number of called cells from RNA cell calling). The hashtag whitelist contained two TotalSeqA anti-human antibody hashes (A0253, TTCCGCCTCTCTTTG; A0255, AAGTATCGTTTCGCA). The filtered matrix output by PIPseeker for the RNA data was merged with the UMI count matrix from CITE-seq Count on cell barcode to create a merged matrix. The hashing data were demultiplexed in Seurat using HTODemux (positive.quantile=0.99). Downstream analysis was performed in Seurat using SCTransform() along with RunPCA(), FindNeighbors(dims=1:15) and RunUMAP(dims=1:15). Cell-type annotation was performed with singleR (v1.4.1) and used an annotated 10x Genomics v1 chemistry dataset as a reference. Cells were classified by their max hash identity and projected in the RNA-based UMAP space. The hash tag oligonucleotide data were subjected to clustering in Seurat using the HTOHeatmap() function to visualize singlets, doublets and unclassified cells.

For 72-h hold experiments, analysis was performed using custom scripts, as previously described above. Samples were normalized to the same depth (45,000 reads per cell). Cell types were then annotated as human (HEK 293T) or mouse (NIH 3T3) using a purity threshold of >85% single-species content per barcode. Barcodes from each species were subset, and transcript counts were summed for each gene to generate two pseudobulk count tables per sample. Samples were aggregated separately for each species and analyzed with DESeq2. A contrast of 0 versus 72 h was performed for each species while controlling for batch effects associated with different users. For the correlation analysis, pseudobulk counts derived above were normalized to transcripts per million and transformed using log(1 + x). Pearson correlations (*R*) and slopes (*m*) were calculated by fitting a linear model to the data. Data were then plotted in R with ggplot2 v3.3.5 and were aggregated into a grid using GGally v2.1.2. Additionally, the distribution of cells in UMAP space at 0 and 72 h after lysis was examined. After processing data in Seurat, as described, harmony batch correction was used to integrate datasets.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Sequencing data were deposited into Gene Expression Omnibus SuperSeries accession number GSE202919.

Code availability
Code for processing raw FASTQ reads into count tables and UMAPs is available at https://www.fluentbio.com/products/pipseeker-software-for-data-analysis/. All other code will be made available from the corresponding author upon request.

References
66. DeJournette, C. J. et al. Creating biocompatible oil-water interfaces without synthesis: direct interactions between primary amines and carboxylated perfluorocarbon surfactants. *Anal. Chem.* **85**, 10556–10564 (2013).
67. Labarge, M. A., Garbe, J. C. & Stampfer, M. R. Processing of human reduction mammoplasty and mastectomy tissues for cell culture. *J. Vis. Exp.* https://doi.org/10.3791/50011 (2013).
68. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
69. Srivastava, A., Malik, L., Smith, T., Sudbery, I. & Patro, R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol.* **20**, 65 (2019).
70. Lun, A. T. L. et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).

Acknowledgements
We thank members of the Abate laboratory for helpful advice and discussions. I.C.C. was supported by K22AI152644 and DP2AI154435 from the NIH. This work was supported by grants 1R44GM145185, 1R43CA239978-01 and 1R43GM137648-01A1 to Fluent BioSciences. M.J. was supported by K99GM130964 from the NIH. J.M.R. was supported by F31NS115380 from the NIH. J.S.W. was supported by NIH 1RM1 HG009490-01 and Howard Hughes Medical Institute. C.C.S. was supported by the Damon Runyon Cancer Research Foundation (CI-99-18), the American Cancer Society (132032-RSG-18-063-01-TBG) and the Leukemia & Lymphoma Society.

Author contributions
Conceptualization, I.C.C. and A.R.A.; methodology, I.C.C., K.M.F., R.H.M., C.L.D. and A.R.A.; investigation, I.C.C., K.M.F., Y.X., D.W.W., K.T.P., C.D.A., A.O., J.Q.Z., P.H., J.S.A.I., V.E.K. and C.A.C.P.; formal analysis, I.C.C., K.T.P., C.H. and A.M.-Z.; resources K.M.F., R.H.M., J.M.R., M.J., E.A.K., S.S., W.K., J.S.W., C.C.S., Z.J.G. and A.R.A.; paper preparation, I.C.C., C.C.S., C.A.C.P. and A.R.A.; supervision, I.C.C., K.M.F., R.H.M., C.C.S., Z.J.G. and A.R.A.

Competing interests
A.R.A. filed a patent related to templated emulsification and is a founder of Fluent Biosciences. I.C.C. consults for Fluent Biosciences and is on its Scientific Advisory Board. K.M.F., R.H.M., Y.X., C.H., A.O., P.H., J.S.A.I., J.Q.Z., A.M.-Z. and C.D.A. are employees at Fluent Biosciences and are working to commercialize the PIP-seq technology. M.J. consults for Maze Therapeutics and Gate Biosciences. J.M.R. consults for Maze Therapeutics and Waypoint Bio. J.S.W. declares outside interest in 5 AM Venture, Amgen, Chroma Medicine, DEM Biosciences, KSQ Therapeutics, Maze Therapeutics, Tenaya Therapeutics, Tessera Therapeutics and Velia Therapeutics. All other authors have no competing interests to declare.

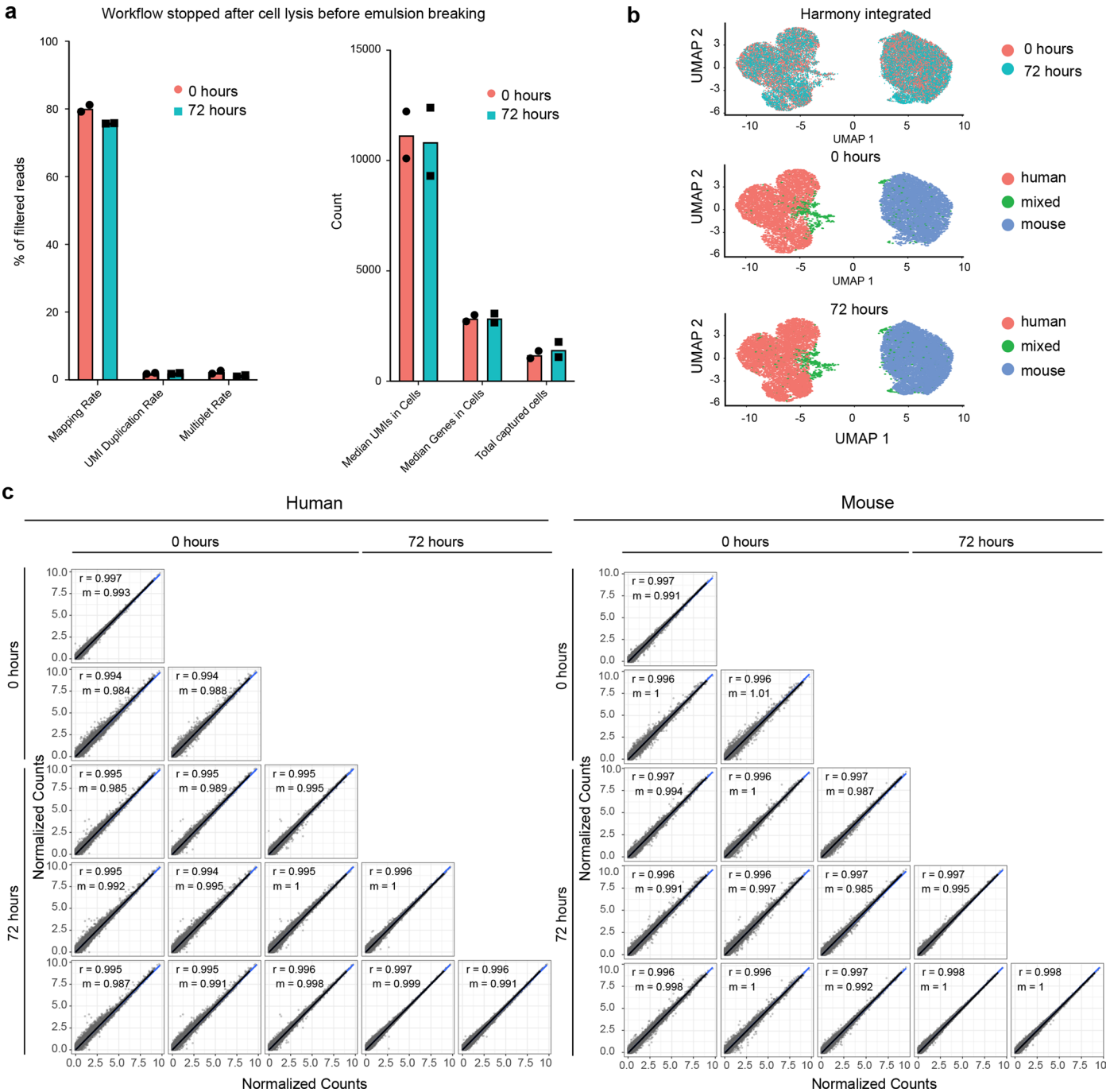
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41587-023-01685-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-023-01685-z.

Correspondence and requests for materials should be addressed to Adam R. Abate.

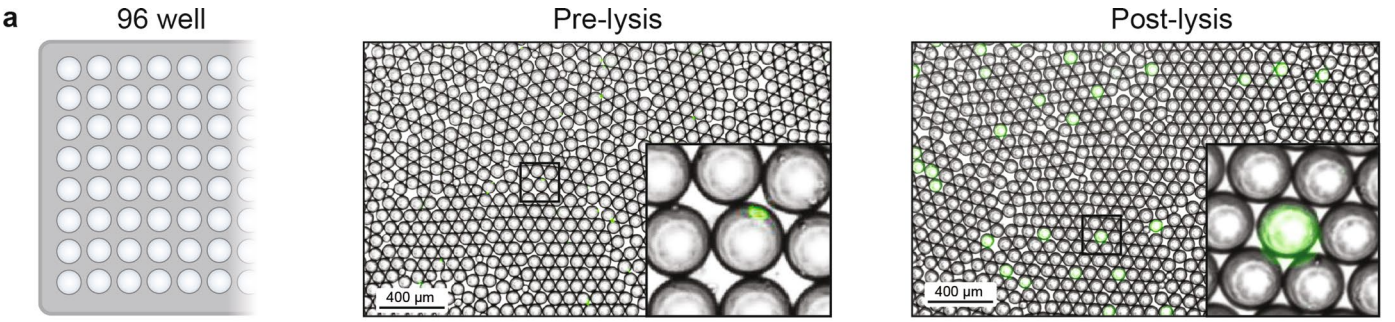
Peer review information *Nature Biotechnology* thanks Linas Mazutis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

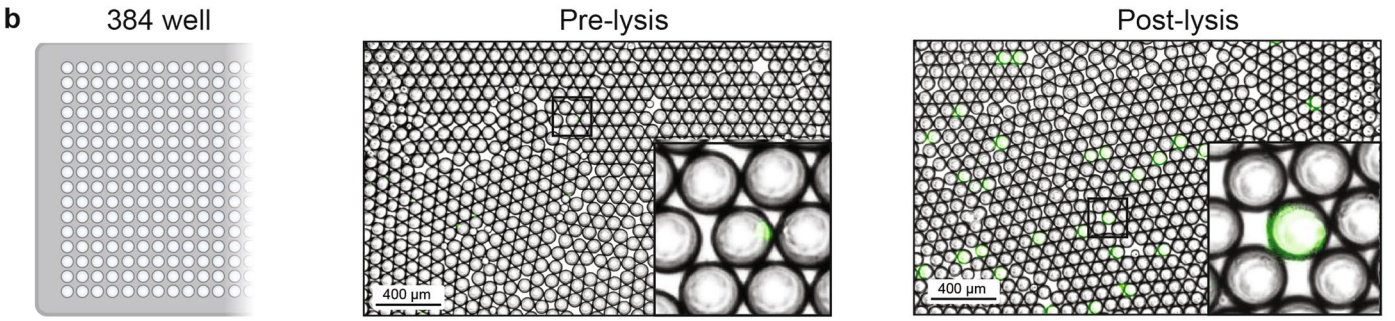
https://doi.org/10.1038/s41587-023-01685-z

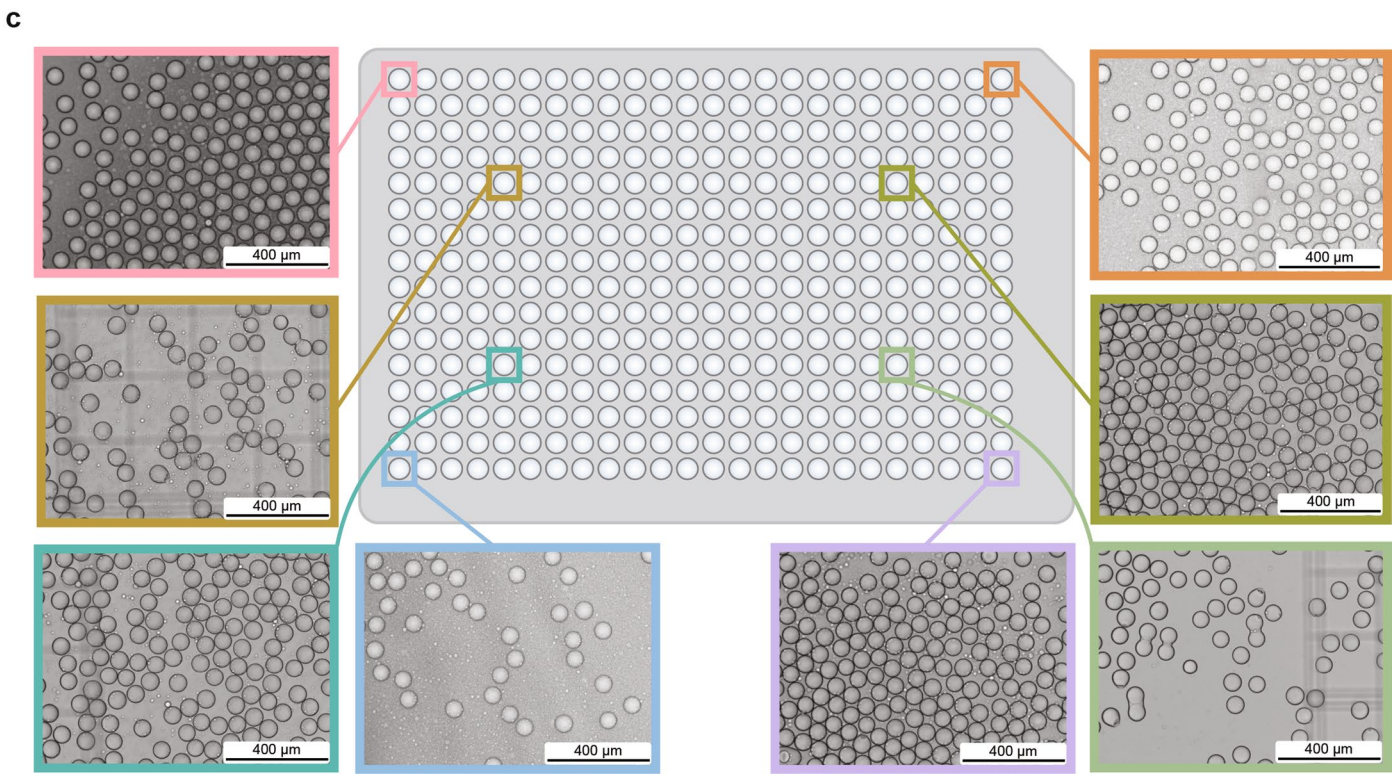


Extended Data Fig. 1 | Flexible sample processing with PIP-seq. (a) Storage of droplets after emulsification for 72 hours at 0 °C did not change quality metrics. Each bar represents the average of two biologically independent experiments with the individual data points shown. **(b)** Data integration between time points. **(c)** Correlations in normalized gene expression, by sample, between time points for mouse and human cells.

https://doi.org/10.1038/s41587-023-01685-z

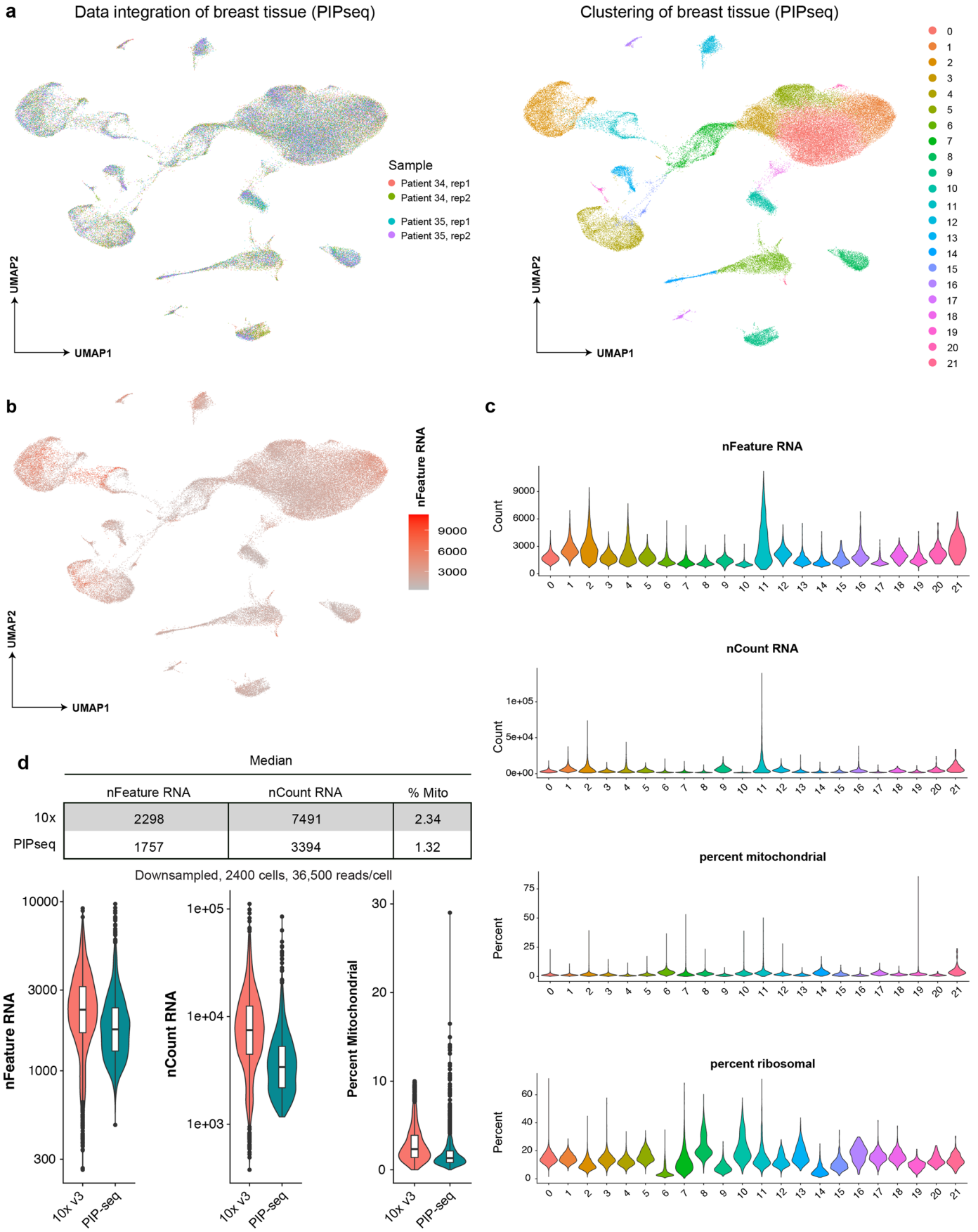






Extended Data Fig. 2 | High cell and sample number experiments with PIP-seq. (a-c) Microscope images of droplets and cells in plate emulsification experiments. Representative images are from experiments completed at least three times. **(a,b)** Barcode bead templates, stained cells (puncta), and lysis reagents are combined with oil and vortexed in **(a)** 96-well and **(b)** 384-well plates to generate monodispersed droplets. Heat activation of Proteinase K results in lysis and release of calcein dye, and full-drop fluorescence. **(c)** Microscope images of droplets from random wells of a 384-well plate emulsification experiment.

https://doi.org/10.1038/s41587-023-01685-z

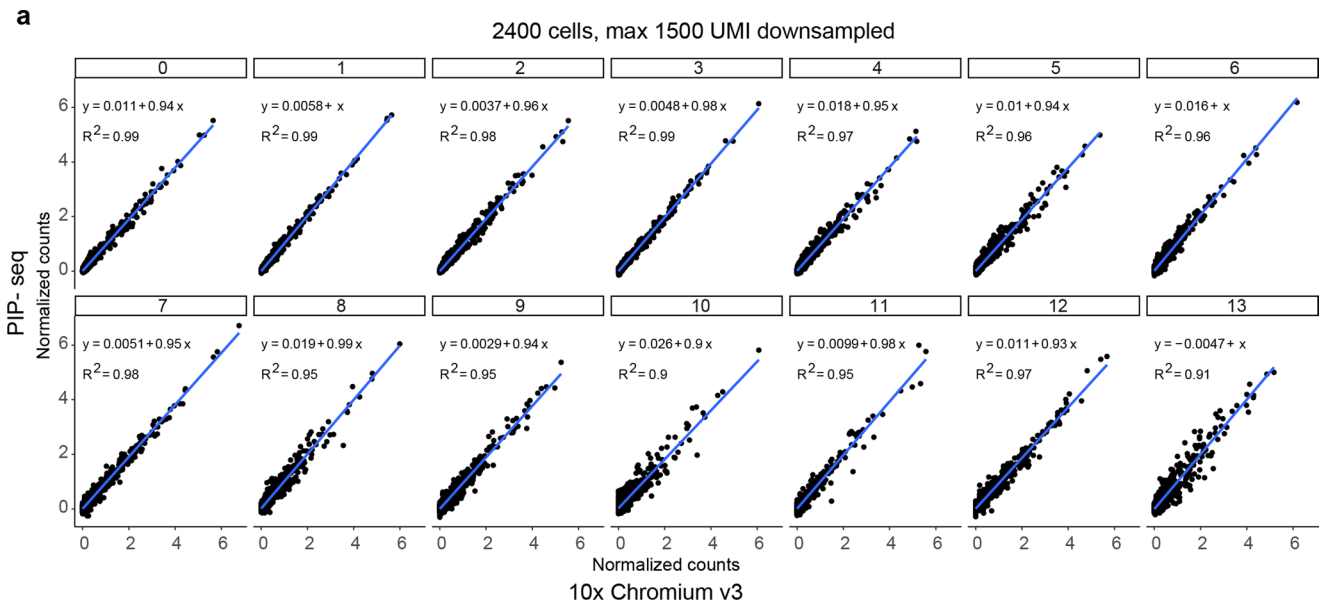


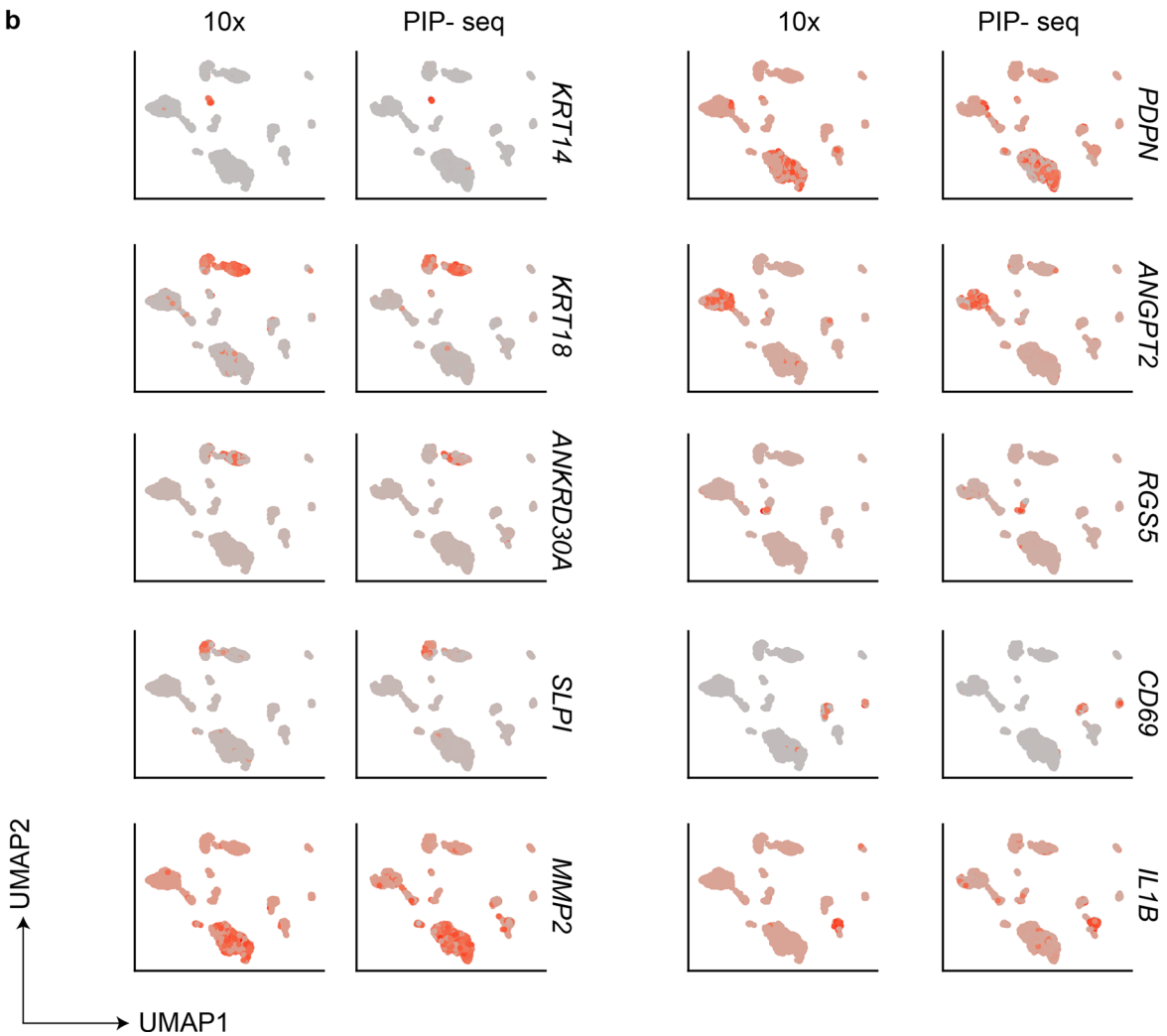
Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41587-023-01685-z

Extended Data Fig. 3 | Quality control analysis of PIP-seq using healthy breast tissue. (a) Integration and clustering of 54,825 cells from 2 patients with 2 replicates per patient. **(b)** Coloring of UMAP by the number genes (nFeature RNA) for each cell. **(c)** The number of unique genes (nFeature RNA), transcripts (nCount RNA), percent mitochondrial reads, and percent ribosomal reads as a function of cluster. **(d)** Comparison between 10X Genomics' and PIP-seq data after downsampling 2400 cells to 36,500 reads per cell. Box plots indicate the median with the lower and upper hinges corresponding to the 25th and 75th percentiles. **(c,d)** Each violin represents a combination of 4 individual samples (2 replicates from 2 patients).

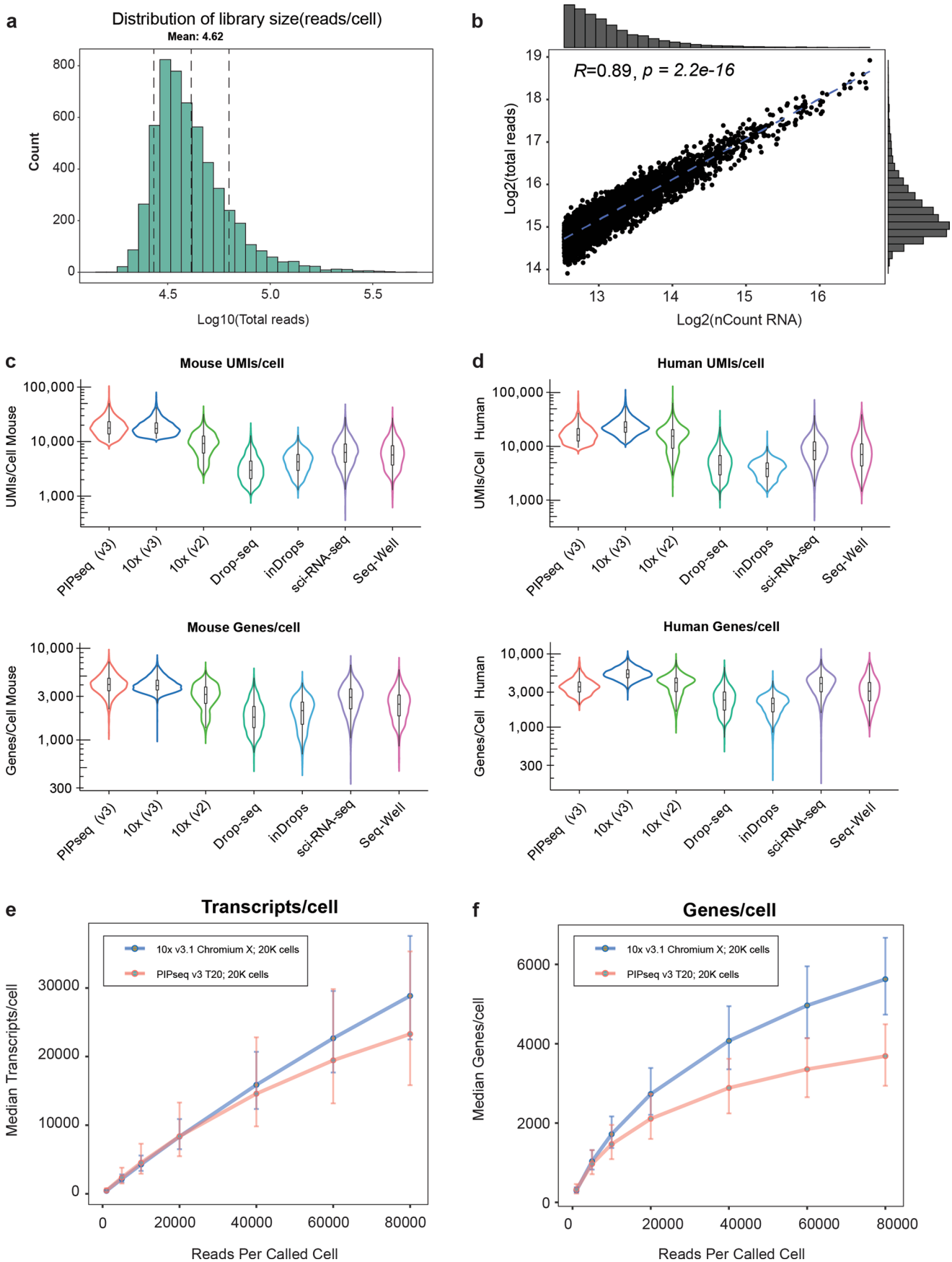
https://doi.org/10.1038/s41587-023-01685-z





Extended Data Fig. 4 | Comparison, after down-sampling (2400 cells 1500 UMIs) of the larger breast tissue PIP-seq dataset to 10x Genomics data generated from identical samples. (a) Correlations in normalized gene expression, by cluster, between platforms. **(b)** Expression of marker genes overlayed on clusters is consistent between platforms.

https://doi.org/10.1038/s41587-023-01685-z

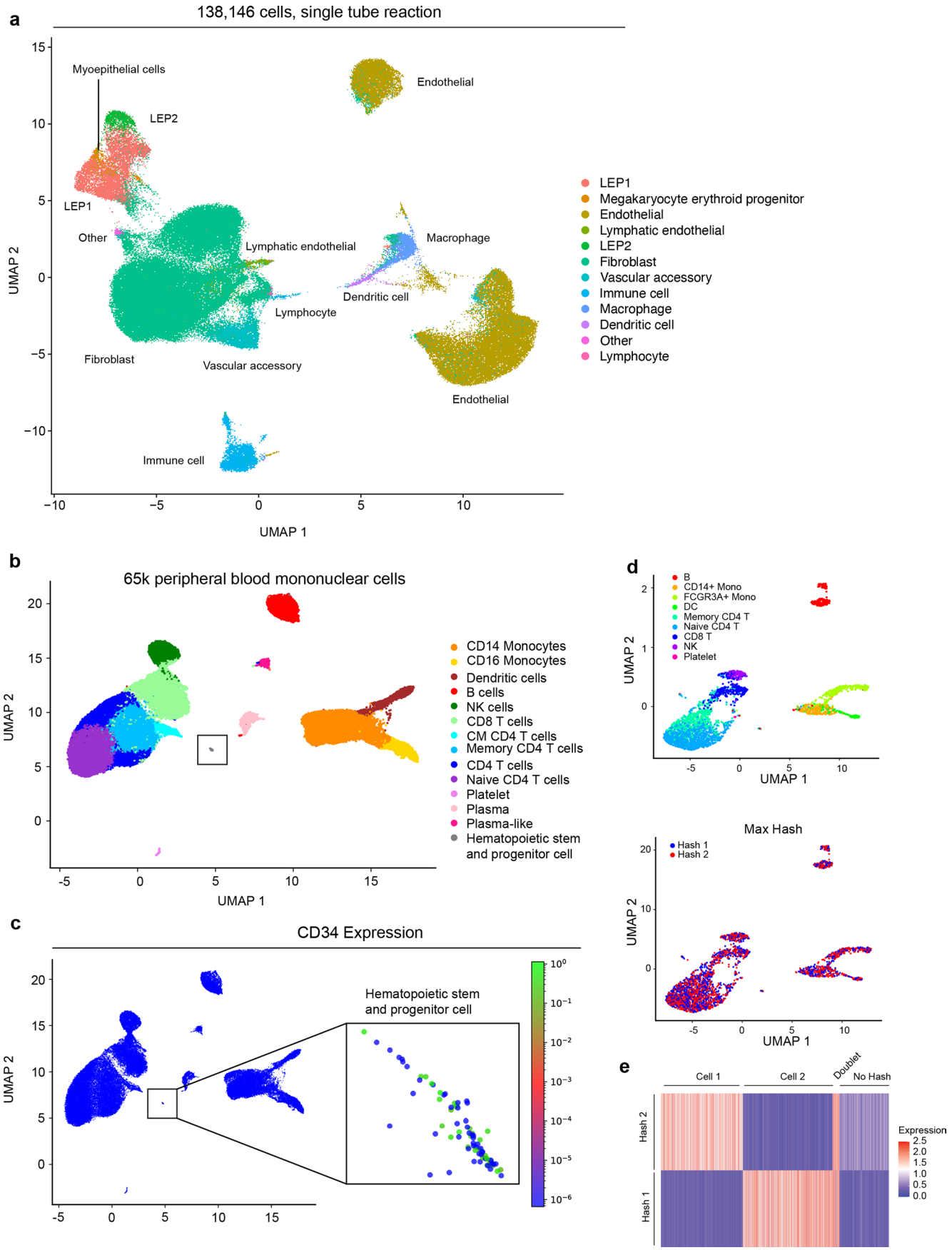


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41587-023-01685-z

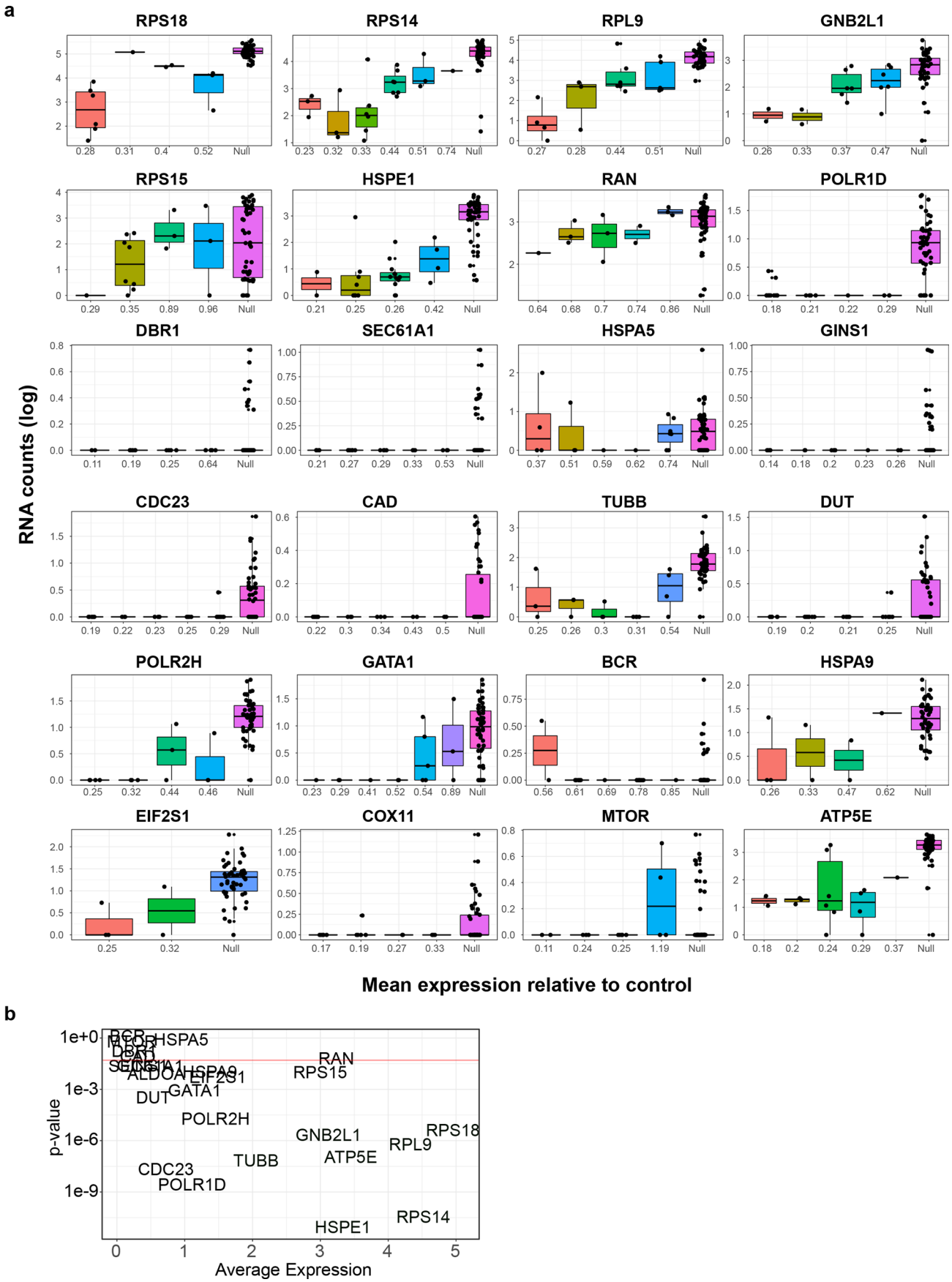
Extended Data Fig. 5 | Data quality assessment of PIP-seq. (a) Representative distribution of reads per cell. **(b)** Correlation between reads and genes per cell. Spearman's R and p values were calculated in R v4.1.0. **(c,d)** Comparison of UMIs/cell and genes/cell in current single-cell methods. Plots display a violin for a single representative sample for each platform. Transcripts per cell and genes per cell are separated by cell species (mouse (**c**) and human (**d**)), identified using the 85% species thresholding technique, as described in the methods. Box plots show the median, 25th and 75th percentiles. **(e,f)** Comparison of PIP-seq to 10X Genomics across a range of sequencing depths (0-80,000 reads/cell) (e) UMIs/cell and (f) genes/cell 80k cells down sampled from one biological replicate. Points represent the median with the lower and upper error bars corresponding to the 25th and 75th percentiles, respectively.

https://doi.org/10.1038/s41587-023-01685-z



Extended Data Fig. 6 | High cell number PIP-seq. (a) scRNA-seq of 138,146 single cells from breast tissue using a single-tube emulsification in 2-minutes. **(b)** scRNA-seq of 65k peripheral blood mononuclear cells (PBMCs) recovers a small population of CD34 cells. **(c)** Coloring of UMAP by the normalized expression of *CD34* for each cell. **(d,e)** Hashing of PBMCs demonstrates compatibility of PIP-seq with barcoded antibodies.

https://doi.org/10.1038/s41587-023-01685-z

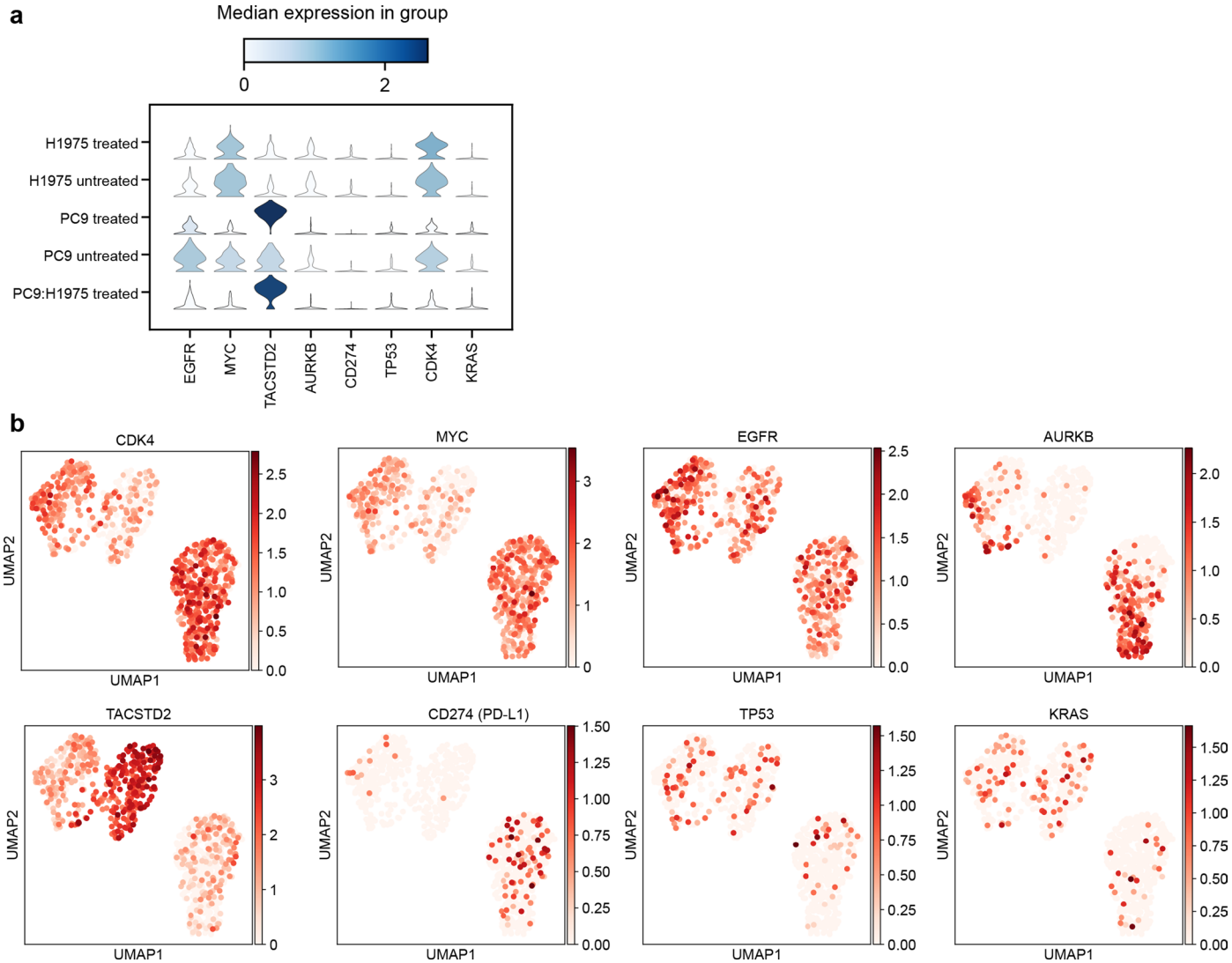


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41587-023-01685-z

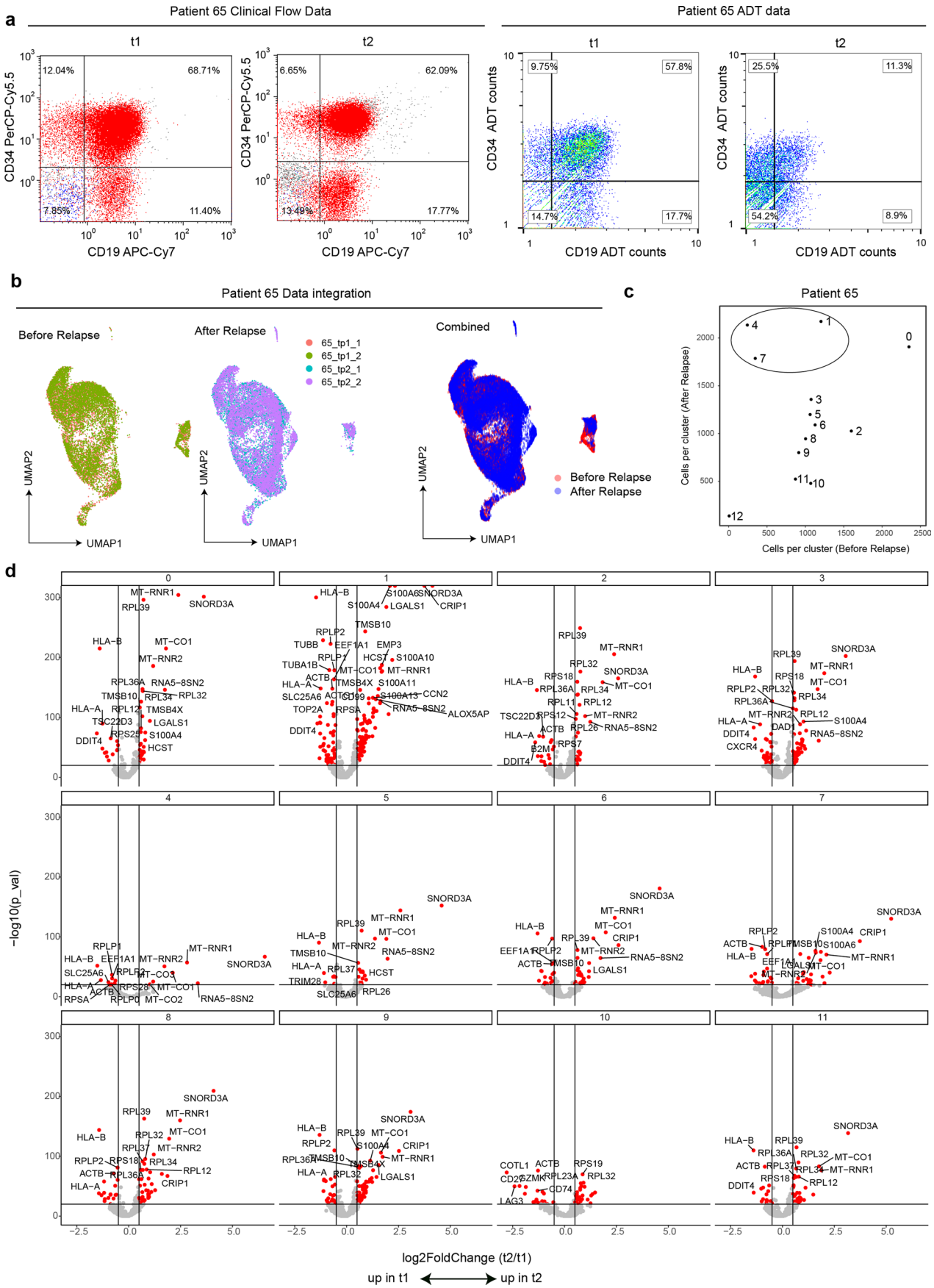
Extended Data Fig. 7 | CROP-seq with PIP-seq. (a) Gene expression for each sgRNA within an allelic series for all genes in the CRISPRi library. Each sgRNA is ordered from predicted high to low knockdown efficiency. Non-targeting sgRNA are denoted as "Null." Data is from one CROP-seq experiment. Box plots indicate the median with the lower and upper hinges corresponding to the 25th and 75th percentiles. Raw data points are displayed with a slight jitter. **(b)** The relationship between gene expression and predicted knockdown of each gene. Expected changes in transcription across the allelic series were prominent in highly expressed genes. p-value represents the significance of the generalized additive model relating gRNA identity to knockdown efficiency for each gene. P-values for each model were directly plotted along with the average expression for each gene (using log1p of the normalized counts). The horizontal red line shows the significance level of p = 0.05.

https://doi.org/10.1038/s41587-023-01685-z



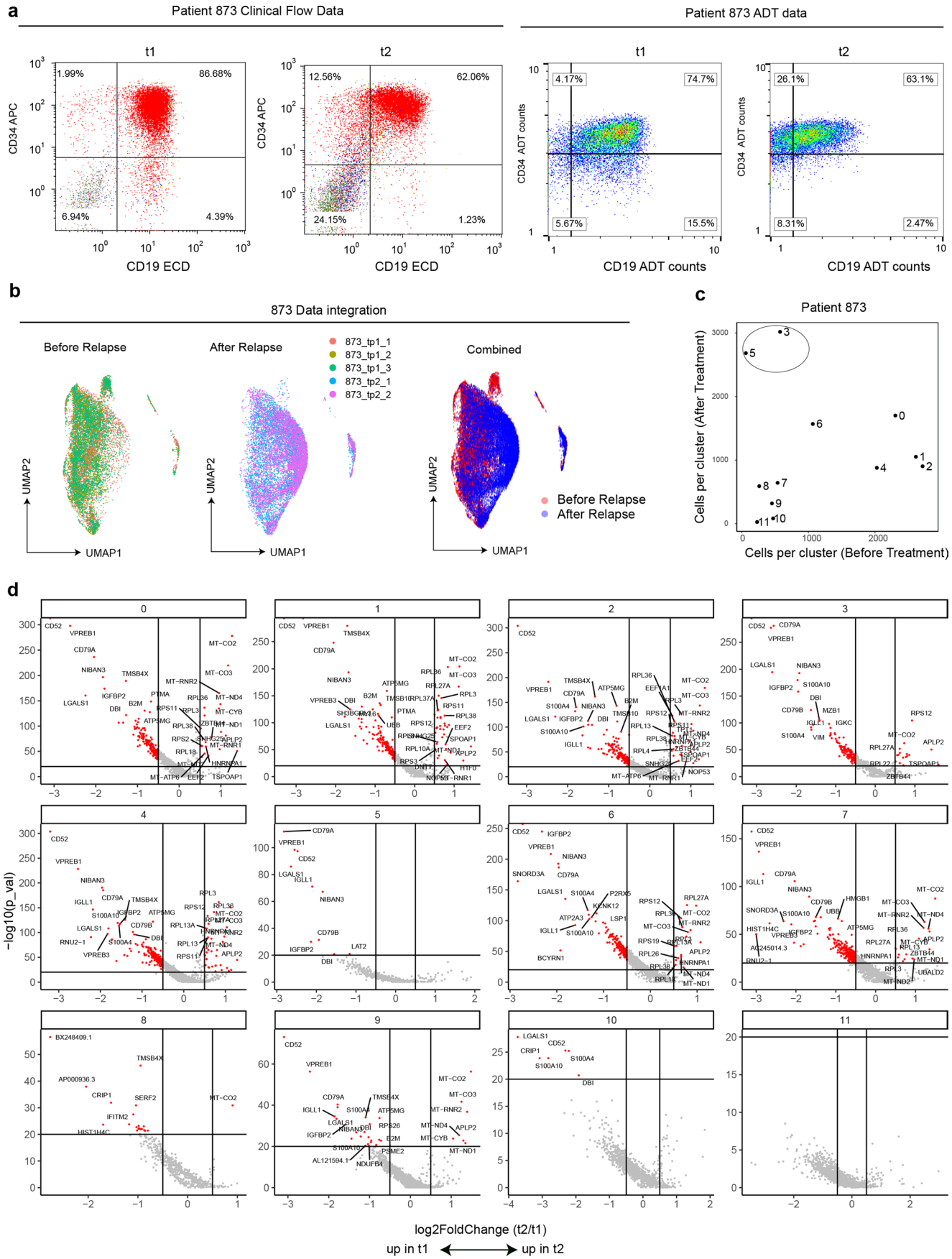
Extended Data Fig. 8 | Identification of Gefitinib-specific transcriptional responses in cancer cell lines. (a) Violin plots of median expression values for selected differentially expressed genes. **(b)** The expression of selected differentially expressed genes superimposed on H1975 and PC9 cell clusters.

https://doi.org/10.1038/s41587-023-01685-z



Extended Data Fig. 9 | Analysis of PIP-seq data from MPAL Patient 65. (a) Clinical flow cytometry and corresponding antibody derived tag (ADT) data for patient 65 with mixed phenotypical acute leukemia (MPAL). **(b)** Integration of replicates and time points. **(c)** Correlation between the number of cells in each cluster before and after relapse identifies expansion of clusters 1,4, and7. **(d)** Volcano plots showing differential gene expression, by cluster, between t1 (before treatment) and t2 (after relapse).

https://doi.org/10.1038/s41587-023-01685-z



Extended Data Fig. 10 | Analysis of PIP-seq data from MPAL Patient 873. (a) Clinical flow cytometry and corresponding antibody derived tag (ADT) data for patient 873 with mixed phenotypical acute leukemia (MPAL). **(b)** Integration of replicates and time points. **(c)** Correlation between the number of cells in each cluster before and after treatment identifies the expansion of clusters 3 and 5. **(d)** Volcano plots showing differential gene expression, by cluster, between t1 (before treatment) and t2 (after relapse).

nature research

Corresponding author(s): Adam Abate

Last updated by author(s): Jan 13, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BD FACS Aria III
Data analysis	R (4.1.2); Seurat (v. 4.1.0); dplyr (1.0.8); FlowJo (10.8.0); ggplot2 (v3.3.5); GSEA (v 4.2.1); Adobe Illustrator (26.1); scanpy (1.4.4); anndata (0.7.1); umap (0.3.10); numpy (1.18.1); scipy (1.4.1); pandas (1.0.3); scikit-learn (0.22.1); CITE-seq-Count (v1.4.3); STAR (v2.7.10a); PIPseeker (v1.0.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data were deposited into GEO SuperSeries accession number GSE202919. The data has been made public. All other materials will be made available upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Most experiments related to single cell sequencing. Single cell RNA-seq was performed on 500 to 50,000 cells, consistent with samples sizes used by others in the field (PMID: 26000488, PMID: 26000487)
Data exclusions	No data were excluded from the study.
Replication	The replication number of each experiment is provided in the manuscript. Each replication was successful.
Randomization	For in vitro drug experiments, samples were randomly allocated into treatment groups. No other samples were allocated into experimental groups because samples (tissue, blood) were processed directly without treatment.
Blinding	Blinding was not performed. In most experiments, experimenters were required to know the conditions of each well.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CITE-seq Antibodies: ADT-CD14 (TotalSeq™-A0081 anti-human CD14 Antibody) #M5E2; ADT-CD64 (TotalSeq™-A0162 anti-human CD64 Antibody) #10.1; ADT-CD22 (TotalSeq™-A0393 anti-human CD22 Antibody) #S-HCL-1; ADT-CD10 (TotalSeq™-A0062 anti-human CD10 Antibody) #HI10a; ADT-CD13 (TotalSeq™-A0364 anti-human CD13 Antibody) #WM15; ADT-CD117 (TotalSeq™-A0061 anti-human CD117 (c-kit) Antibody) #104D2; ADT-CD5 (TotalSeq™-A0138 anti-human CD5 Antibody) #UCHT2; ADT-CD7 (TotalSeq™-A0066 anti-human CD7 Antibody) #CD7-6B7; ADT-CD56 (TotalSeq™-A0084 anti-human CD56 (NCAM) Recombinant Antibody) #QA17A16 ; ADT-HLA-DR (TotalSeq™-A0159 anti-human HLA-DR Antibody) #L243 ; ADT-CD11b (TotalSeq™-A0161 anti-human CD11b Antibody) #ICRF44 ; ADT-CD4 (TotalSeq™-A0922 anti-human CD4 Antibody) #OKT4 ; ADT-IgG1 (TotalSeq™-A0090 Mouse IgG1, κ isotype Ctrl Antibody) #MOPC-21; ADT-CD3 (TotalSeq™-A0049 anti-human CD3 Antibody) #SK7; ADT-CD19 (TotalSeq™-A0050 anti-human CD19 Antibody) #HIB19; ADT-CD30 (TotalSeq™-A0028 anti-human CD30 Antibody) #BY88; ADT-CD33 (TotalSeq™-A0052 anti-human CD33 Antibody) #P67.6; ADT-CD45 (TotalSeq™-A0391 anti-human CD45 Antibody) #HI30; ADT-CD34 (TotalSeq™-A0054 anti-human CD34 Antibody) #581. Sample HHMTB0065: CD34 PerCP-Cy5.5 – clone 8G12, BD Biosciences Cat No 347213; CD19 APC-Cy7 – clone SJ25C1, BD Biosciences Cat No 348804. Sample HMTB00873; CD34 APC – clone 581, Beckman Coulter Cat No IM2472U; CD19 ECD – clone J3-119, Beckman Coulter Cat No IM2708U. Hash Antibodies: TotalSeq™-A0253 anti-human Hashtag 3, BioLegend Cat. No. 394605, Lot No. B364258; TotalSeq™-A0255 anti-human Hashtag 5 Cat. No. 394609, Lot No. B366838.
Validation	All antibodies were commercial in origin and validated by the company. BioLegend (https://www.biolegend.com/nl-nl/reproducibility) purifies antibodies from cell line supernatant using affinity chromatography followed by validation using proper negative and positive controls in functional assays. This is relevant for all antibodies used for flow cytometry and Ab sequencing. BD (https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/quality-and-reproducibility) confirms specificity using multiple methodologies that may include a combination of flow cytometry, immunofluorescence, immunohistochemistry or western blot. Beckman Coulter (https://www.beckman.com/reagents/coulter-flow-cytometry/antibodies-and-kits/single-color-antibodies/quality-standards).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human HEK 293T cells (ATCC # CRL-3216), Murine NIH/3T3 cells (ATCC # CRL-1658) , H1975 (ATCC#CRL-5908); K562i were a gift from the Weissman lab; PC9 was obtained from RIKEN Bio Resource Center (RCB4455).
Authentication	Cell lines are authenticated by the manufacturer with STR profiling. K562i were generated and authenticated by the Weissman lab (PMID: 31932729)
Mycoplasma contamination	Cell lines are verified to be free of Mycoplasma contamination by the manufacturer.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Human research participants

Policy information about studies involving human research participants

Population characteristics	Breast tissue study: Samples were derived from a human breast tissue biobank at Univeristy of California in San Francisco (UCSF). This biobank was established from patients who underwent elective breast reduction or mastectomy surgeries at UCSF for cosmetic reasons. These patients are either cis-women or trans-men who are older than 18. Cancer study: Two human participants were included in this study.Patient 65 is a 65 year old male diagnosed with B/myeloid mixed-phenotypic acute leukemia. This patient had an MLL rearrangement. Genotypic information (germline or somatic) was not obtained. The patient was treated with cytotoxic chemotherapy (cytarabine and duanorubicin).Patient 873 is a 76 year old female diagnosed with B/myeloid mixed-phenotypic acute leukemia. Genotypic information includes somatic mutations in TP53 (c.743G>A) and SF3B1 (c.1739C>T). The patient was treated with cytotoxic chemotherapy (cyclophosphamide and dexamethasone) and the antibody-drug conjugate inotuzumab.
Recruitment	Breast tissue study: Patients were selected from those who have already been scheduled for surgery, and those who are not excluded by being <18yrs of age or currently pregnant. Patients were notified by their physician during a clinical visit and the research team reached out to them before their surgeries to obtain consent. Cancer study: Patients consented to have samples prospectively banked in the UCSF tumor bank at time of routine bone marrow biopsy. The samples were then retrospectively selected from the UCSF tumor bank for this project. Self selection bias, should it exist, is not expected to impact the results of single cell RNA-seq studies that characterize cell heterogeneity.
Ethics oversight	Breast tissue study: This study uses human tissues previously collected in a biobank approved by UCSF IRB. Cancer study: The prospective banking of patient samples is approved by the UCSF IRB (IRB# 11-06477). Samples were collected in accordance with the Declaration of Helsinki under institutional review board-approved tissue banking protocols, and written informed consent was obtained from all patients.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Bone Marrow Aspirates (BMA) and Peripheral Blood Specimens are prepared using the Bulk Lysis Technique. Briefly, using a 3mL syringe with a 22 gauge blunt needle, bone marrow aspirate was drawn up and expelled back into the tube a minimum of three times to break up the spicules and release cells. Any clots in the sample were removed from the tube. Red blood cell lysis was performed using ammonium chloride. After tube inversion for 10 minutes, cells were centrifuges at 1400rpm for 5 minutes. Supernatant was removed, and the pellet was resuspended by adding, dropwise, 1mL of PBS + 0.1%NaN3 + 0.5% BSA to the tube while vortexing gently. An additional 13mL of PBS + 0.1%NaN3 + 0.5% BSA was added, clumps were removed manually with a pipette, and cells were centrifuges at 1400rpm for 5 minutes. Pellets were resuspended in 600µL of PBS-P solution (10% heat-inactivated (56°C for 30 min) newborn calf serum), with 10% Acid Citrate Dextrose (Sigma C3821-50ML).

Instrument	Sample HMTB0065: Instrument: BD FacsCanto (6-color) Sample HMTB00873: Instrument: Beckman-Coulter Navios (10-color)
Software	Kaluza (Beckman Coulter)
Cell population abundance	Abundance is reported in figures where relevant.
Gating strategy	Sample HMTB0065: Gating strategy: none (all events displayed) Sample HMTB00873: Gating strategy: none (all events displayed)

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.