PAPER

Check for updates

Opto-combinatorial indexing enables high-content transcriptomics by linking cell images and transcriptomes†

Arata Tsuchida, [iD] [ab] Taikopaul Kaneko, [iD] [a] Kaori Nishikawa,[a] Mayu Kawasaki,[a] Ryuji Yokokawa [iD] [b] and Hirofumi Shintaku [iD] *[abc]

We introduce a simple integrated analysis method that links cellular phenotypic behaviour with single-cell RNA sequencing (scRNA-seq) by utilizing a combination of optical indices from cells and hydrogel beads. With our method, the combinations, referred to as joint colour codes, enable the link *via* matching the optical combinations measured by conventional epi-fluorescence microscopy with the concatenated DNA molecular barcodes created by cell-hydrogel bead pairs and sequenced by next-generation sequencing. We validated our approach by demonstrating an accurate link between the cell image and scRNA-seq with mixed species experiments, longitudinal cell tagging by electroporation and lipofection, and gene expression analysis. Furthermore, we extended our approach to multiplexed chemical transcriptomics, which enabled us to identify distinct phenotypic behaviours in HeLa cells treated with various concentrations of paclitaxel, and determine the corresponding gene regulation associated with the formation of a multipolar spindle.

Introduction

In the latest single-cell RNA sequencing (scRNA-seq) procedures, thousands of cells are assayed per experiment by combining compartmentalisation of cells with microfluidics and tagging cDNA with cell barcodes to profile the gene expression of single cells.[1–4] The tagging approach has been extended for profiling other omics layers including surface proteins,[5,6] nuclear proteins,[7] and chromatin accessibility.[8] However, most of the omics approaches are incapable of linking the measured molecular profile to cellular phenotypes, such as morphology and molecular localisation.[9,10]

Single-cell optical phenotyping and expression (SCOPE-seq and SCOPE-seq2)[11,12] is a method for linking scRNA-seq with live cell imaging. In SCOPE-seq, a single cell and a bead bearing barcoded DNA are co-isolated in a microwell, images are obtained of the cellular morphology, and mRNA from a single cell on the bead is captured for pooled scRNA-seq. SCOPE-seq links the image of the single cell to the scRNA-seq by optically decoding the barcode of each bead using cyclic hybridisation of fluorescently labelled oligo (oligonucleotide) probes, followed by fluorescence microscopy.

Herein, we propose a novel and simple approach for optical indexing that leverages the combination of cells and hydrogel beads dual-labelled with optical indices and DNA molecular barcodes (we refer to this dual label as a "colour code") for linking cellular images with scRNA-seq. To link the combinations of optical indices decoded from the imaging to the cell barcodes in scRNA-seq, our approach creates concatenated fragments of barcoded DNA oligos (DNA tags) derived from the cells and barcoded dT primers derived from the hydrogel beads. The concatenated fragments are then sequenced with the scRNA-seq library, and provide a look-up table to link the combinations of optical indices and cell barcodes.

Our approach is free of automated microfluidic controls and offers fewer on-chip steps, which is advantageous because this approach can be easily implemented in a standard laboratory setup. We demonstrate our approach with multiplexing up to 256 combinations of colour codes (joint colour codes), using 16 pools of colour-coded cells and 16 pools of colour-coded hydrogel beads, and decoding them with four colours of fluorescence *via* standard epi-fluorescence microscopy.

[a] *Cluster for Pioneering Research, RIKEN, Main Research Building 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*

[b] *Department of Micro Engineering, Graduate School of Engineering, Kyoto University, Kyotodaigaku-katsura, Nishikyo-ku, Kyoto 615-8540, Japan. E-mail: shintaku@infront.kyoto-u.ac.jp*

[c] *Institute for Life and Medical Science, Kyoto University, 53 Kawara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3lc00866e

Results

Strategies to link cell images and whole transcriptomes by utilizing colour coding

Our strategy to link a single-cell image and gene expression leverages a joint colour code created by a co-isolated single cell and a hydrogel bead (Fig. 1A). The cells and hydrogel beads are respectively labelled by fluorescence dye and the corresponding DNA molecular barcodes (Fig. 1B and S1 and S2†) that are read by epi-fluorescence microscopy (Fig. 1C) and next-generation sequencing. The joint colour codes increase the possible unique codes by the combination and enable the linking of single-cell images and gene expression profiles in the two data pools (Fig. 1D).
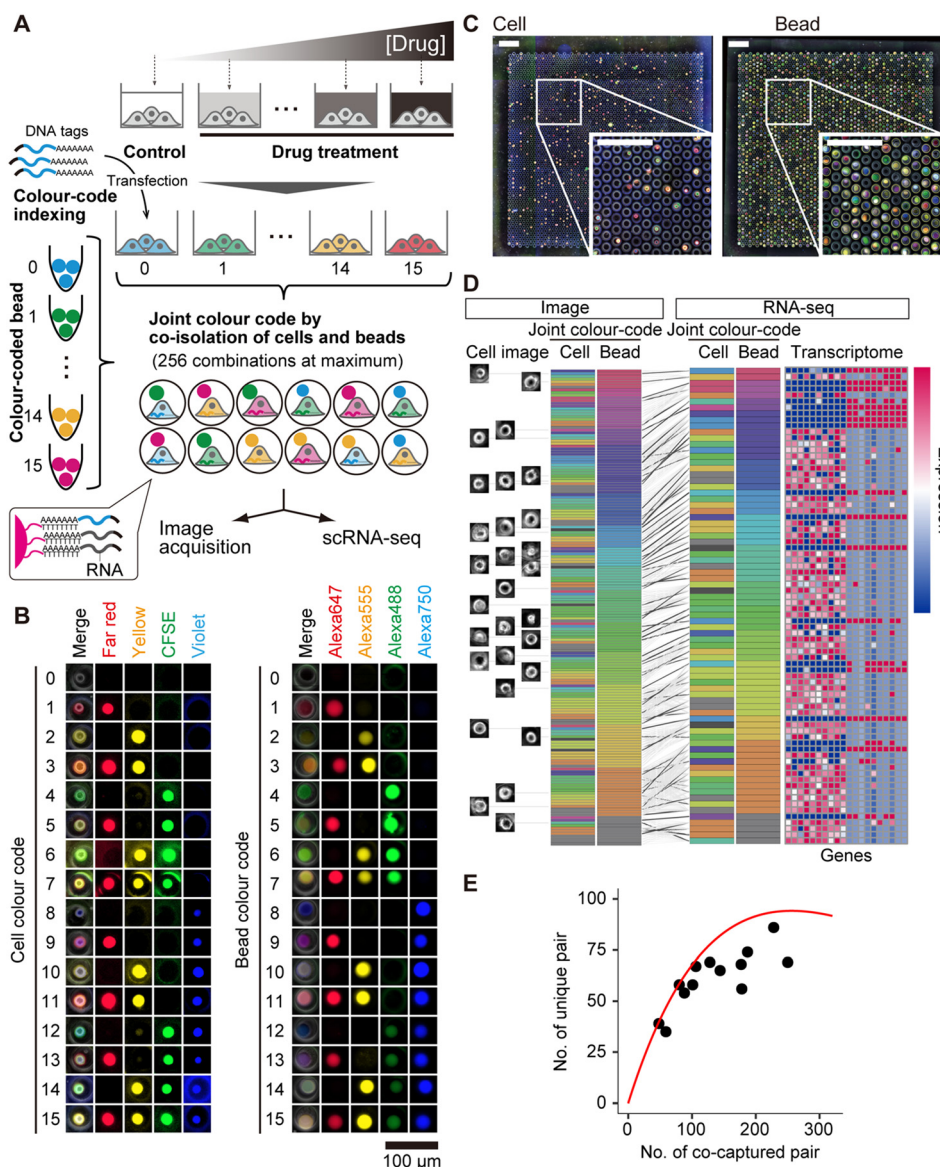

Fig. 1 Linking a cellular image to single-cell RNA-seq with a combination of colour codes. (A) Cells were colour-coded with matching DNA tags and a set of fluorophores as per the 16 different conditions. Hydrogel beads were also colour-coded with matching barcode sequences and fluorophores. Colour-coded cells and hydrogel beads were co-isolated in microwells. We imaged fluorescence combinations of a cell and a hydrogel bead, followed by the generation of concatenated DNA fragments from the DNA tag of the cell and the barcode sequence of the hydrogel bead, along with reverse transcription of cDNA from mRNA. Cell images and transcriptome data were linked by matching the fluorescence combinations from imaging with the library of the joint colour code generated from the concatenated DNA fragments. (B) Representative fluorescence images of colour-coded cells (left) and beads (right) in microwells. Scale bar = 100 μm. (C) Single cells were co-isolated with single hydrogel beads out of a pool of those bearing 16 different colour codes in microwells. Cells are outlined with red borders, while bead-captured wells are outlined with yellow borders. The isolated single cells were processed to yield a scRNA-seq library and a joint colour code library that produced the gene expression and the colour code combination, respectively. Scale bar = 300 μm. (D) Single cells with unique joint colour code combinations resulted in linked datasets of cellular morphology and gene expression. (E) The expected number of unique joint colour codes per experimental run follows the Poisson distribution. The red line and black dots indicate the theoretical distribution and the experimental values from colour-decoded images, respectively.

In our demonstration, we designed 16 colour codes, which were bright or dim combinations of four fluorescence dyes ($2^4 = 16$) and which also corresponded to 16 different sequences of DNA barcodes, respectively, for cells and hydrogel beads (see the Methods section). Thus, a maximum of 256 joint colour codes (16 × 16) can be registered. The expected number of cell–bead pairs with unique joint colour codes per experimental run was predicted to attain a maximum of approximately 94 when assaying 256 single cells on the basis of Poisson distribution, excluding the cells with duplicated joint colour codes (Fig. 1E).

To demonstrate our protocol, we performed mixed-species experiments using HeLa cells (human) and NIH/3T3 cells (mouse) (Fig. S3†). We prepared a pool of 16 differently colour-coded cells (eight sub-pools each of HeLa and NIH/3T3 cells) that were each labelled with a combination of four different dyes (CellTrace Violet, CFSE, Yellow, and Far Red from Thermo Fisher Scientific) (Fig. 1B and S1†) and corresponding DNA tags, which contained an 8 nt barcode, poly A sequence, and a PCR handle (Table S2†), *via* electroporation. We then isolated single cells from the pool of 16 colour codes in microwells and imaged them by epi-fluorescence and in a bright field. The hydrogel beads bearing barcoded primers with colour codes (Fig. 1B and S2†) were subsequently isolated in the microwells to capture mRNA and DNA tags.

To retain the molecules released from the cells within the microwells, we sealed the microwells with a track-etched membrane with nanopores of 10 nm in diameter, chemically lysed the cells in the microwells, and captured the mRNA and DNA tags by the hydrogel beads *via* hybridisation. After removing the track-etched membrane, we imaged the fluorescence of the hydrogel beads in the microwells and read the colour codes. We registered the joint colour codes with the images of the single cells by integrating the microscopic images of cells and hydrogel beads. We finally transferred the hydrogel beads to a standard PCR tube to synthesise libraries of the scRNA-seq and DNA tag by off-chip reactions (see the Methods section). The latter library yielded a look-up table that linked the cell barcodes in the cDNA fragments and joint colour codes by creating concatenated fragments of colour codes of hydrogel beads and cells (Fig. S4†).

Our microwell chip, which contained 2511 wells per chip, captured 149 ± 70 cells and 1326 ± 462 hydrogel beads per run ($n = 13$), and created 137 ± 64 pairs of single cells and hydrogel beads on average. The image-based decoding of the joint colour codes showed that the number of unique joint colour codes matched the theoretical prediction (Fig. 1E).

Linking single-cell images to scRNA-seq

The sequence reads with the same cell barcode were dominantly mapped to the homosapiens genome (GRCh38.p12) or the *Mus musculus* genome (GRCm38.p6), supporting successful RNA-seq at single-cell resolution (Fig. 2A). Using scRNA-seq, 1296 ± 586 genes were detected per cell (*i.e.*, 2839

± 1829 unique molecular identifiers (UMIs) per cell) and 1203 ± 541 genes per cell (*i.e.*, 2633 ± 1716 UMIs per cell), for HeLa and NIH/3T3 cells, respectively (the sequence read per cell was 30 865 on average). Of the 360 unique joint colour codes identified by fluorescence microscopy in six experimental runs, 137 were also identified in the DNA tag library and successfully linked to the scRNA-seq data. Of those, 122 cells, *i.e.*, 89.1%, were consistent for the species (Fig. 2A–C).

To link the cell barcode in scRNA-seq to single-cell images *via* joint colour codes, we devised a framework that optimises pairs of cell barcodes and single-cell images by maximizing the sum of the similarity between the decoded colour code from the images and counts of DNA tags (Fig. 2D and E). We benchmarked the framework in terms of the accuracy and number of linked datasets using the mixed-species data, computed with various metrics of similarity and normalisation approaches for the DNA tag counts. The results showed that the most optimal performance among those tested was obtained for the cosine similarity in combination with the centred log ratio (CLR per feature) for normalisation of DNA tags among those tested (Fig. 2F). The framework with the cosine similarity and CLR yielded a consistency of 91% for species at a threshold of 0.5 (Fig. 2A). We employed the same framework throughout this study.

Labelling cells with DNA tags

Next, we benchmarked two different approaches, electroporation and lipofection, for labelling cells with fluorescently labelled DNA tags, with subsequent flow cytometry analysis (Fig. 3A and B). The data revealed that lipofection outperformed electroporation in delivering a greater number of DNA tags to cells, while the amount of DNA tags resulted in a relatively large cell-to-cell variation. Furthermore, lipofection was less efficient for NIH/3T3 cells as compared to HeLa cells. To gain similar sensitivity in detecting the DNA tags on HeLa and NIH/3T3 cells, we employed electroporation in the experiments with mixed species. Alternatively, tuning the sensitivity was accomplished by modulating the concentration of DNA tags for lipofection (Fig. 3A and B). We employed lipofection in the other experiments with a cell line.

We also assessed the durability of the DNA tags within the cells by quantifying their presence over time (Fig. 3C). After 48 h of labelling, the DNA tags remained detectable, indicating the potential for combining our approach with longitudinal live-cell imaging. Notably, the linking rate, a metric representing the fraction of cell images linked to scRNA-seq among those identified from the cell images, exhibited no degradation over time (Fig. 3D). Interestingly, the linking rate associated with electroporation labelling increased over time. We hypothesise that this trend may be attributed to selection bias in favour of healthy cells within the electroporated cell population. Furthermore, we conducted additional analyses to confirm the integration of labelled cells with unlabelled cells in the transcriptomic
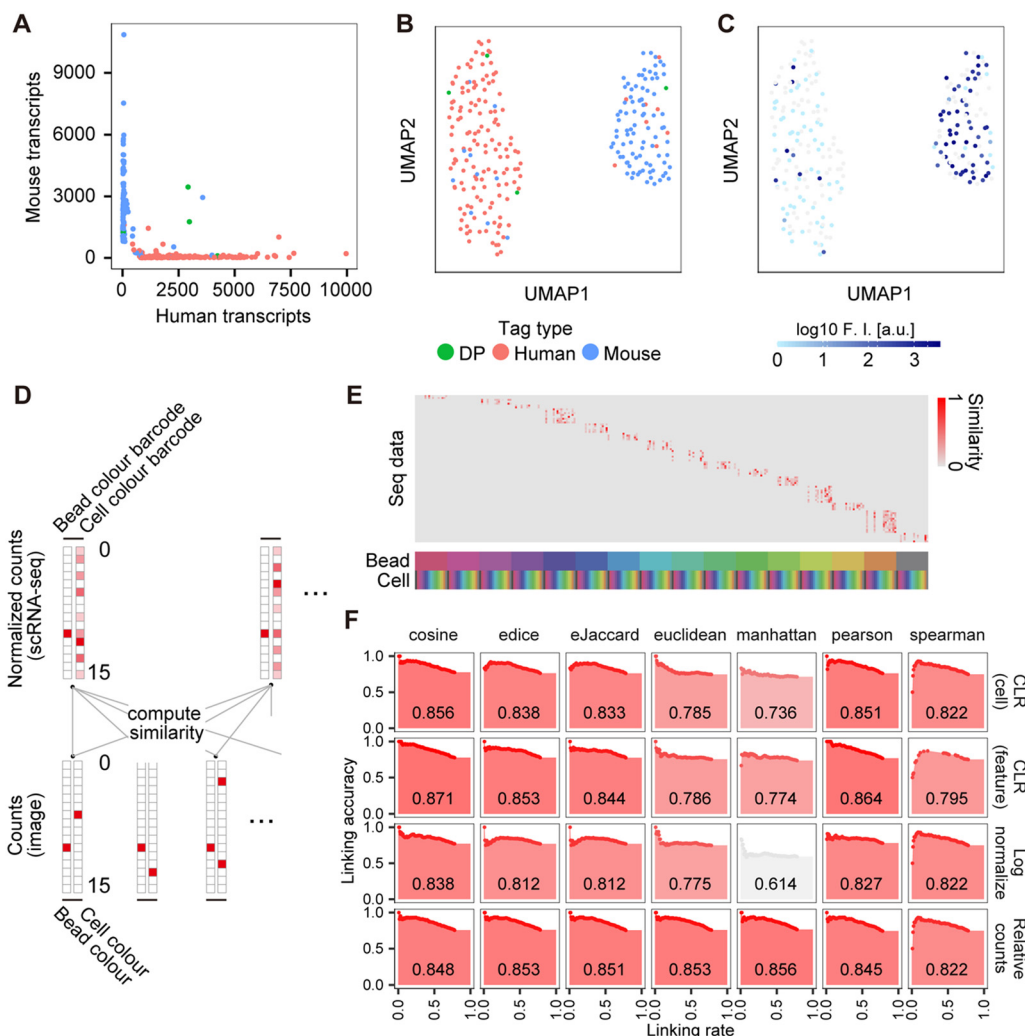
Fig. 2 Mixed species experiment (mouse and human) validating the linkage between the transcriptomic data and imaging data. (A) The number of detected transcripts associated with individual cell barcodes. The colour represents the cell type identified from linked cell images. Uniform manifold approximation and projection (UMAP) of the cells. The colour represents (B) the cell type identified from linked cell images (DP is linked to cells co-captured with cells from different species) and (C) the logarithmic fluorescence intensities of cells unique to mice. (D) Schematic image of the computation for the cosine similarity. (E) A look-up table that shows the matching of sequencing data (columns) and colour codes decoded from the image data (rows). (F) The area under the curve of linking rates (the number of linked joint colour codes over the number of joint colour codes detected in the image) *versus* linking accuracies (the number of linked data with consistent species over the number of linked data) using different similarities and normalizations of DNA tag counts. We identified the species of cells by the abundance of the transcripts when one species exceeds 70% of the total unique molecular identifier (UMI) counts.

space (Fig. 3E and F). These data serve as a clear benchmark for cell tagging achieved through DNA delivery *via* electroporation and lipofection.

Exploring chemical perturbation with high-content transcriptomics

Next, we sought to determine if our approach could enhance the insights gained from chemical screening. We investigated the cell-to-cell heterogeneity in the response of HeLa cells to the chemical impact of paclitaxel, which is a chemotherapy drug used in the clinical treatment of lung, ovarian, and breast cancer that inhibits the growth of cancer cells by blocking cell division. Traditionally, it was believed to induce cell death through mitotic arrest. However, recent studies

have suggested that tumour regression is not solely dependent on mitotic arrest, but is influenced by multipolar spindle formation,[13] leading to cell death.[14] In our study, we aimed to dissect the nuclear phenotype associated with multipolar spindles induced by paclitaxel and its underlying transcriptomic basis.

To understand the intricate relationship between the phenotypic and transcriptomic responses at the single-cell level, we subjected the DNA-tagged and colour-coded HeLa cells to paclitaxel treatment at eight distinct concentrations, ranging from 0.5 to 500 nM, over a 24 h period. Subsequently, we analysed the combined samples using our established approach. We utilised a single fluorescence channel to monitor the emergence of multipolar spindles as a phenotypic response to paclitaxel by staining the DNA with Hoechst 33342 dye. The
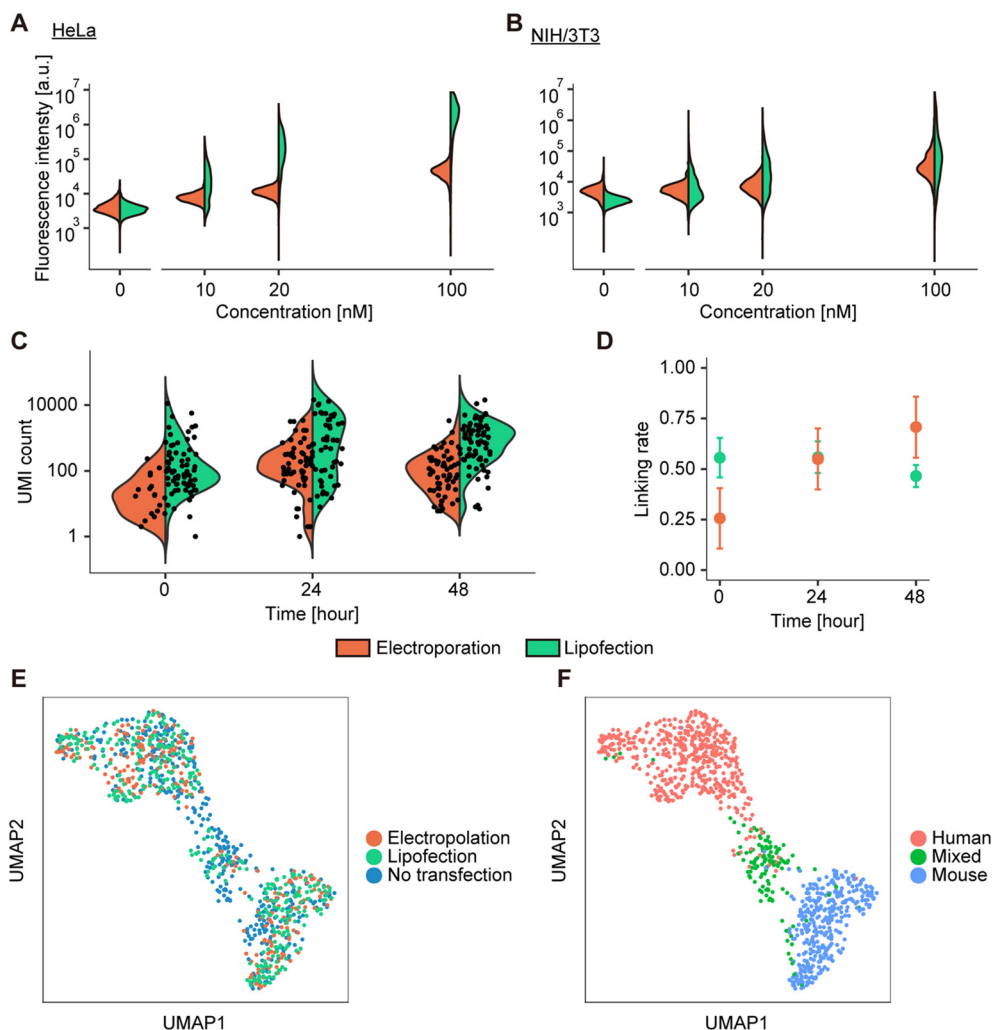
Fig. 3 Comparison of electroporation and lipofection for DNA-tag labelling. Quantities of FAM-labelled DNA tags at various concentrations for (A) HeLa cells and (B) NIH/3T3 cells. (C) Unique molecular identifier (UMI) counts of DNA tags against the different durations of incubation at 20 nM for lipofection, and 100 nM for electroporation. (D) Linking rates against the different durations of incubation. Uniform manifold approximation and projection (UMAP) of (E) single cells labelled by electroporation and lipofection and (F) species determined by the ratio of UMI counts.

colour-decoded images of individual cells revealed that multipolar spindles were more prevalent at higher concentrations of paclitaxel. Notably, even at identical concentrations of paclitaxel, the number of spindles exhibited considerable heterogeneity across the cells (Fig. 4A and B). These observations were in accordance with the findings from non-pooled assays conducted in separate dishes (Fig. S5E and F†).

To determine the mechanism underlying the heterogeneous cellular response, we leveraged the transcriptomic data linked to the phenotypic responses. There were 1364 ± 350 genes per cell (*i.e.*, 4357 ± 2366 UMIs per cell, with $106\,119$ sequence reads per cell on average), according to the transcriptomic data. The integrated multimodal data by weighted nearest neighbour (WNN) analysis enabled two distinct trajectories to be inferred, which is related to the formation of multipolar spindles, or lack thereof, in response to the paclitaxel burden within the transcriptomic data (Fig. 4C–E). Gene set enrichment analysis (GSEA) revealed that the genes associated with mitosis (mitotic

cell cycle, mitotic cell cycle process, cell cycle process, cell cycle G2/M phase transition, cell cycle phase transition, G2/M transition of mitotic cell cycle, and regulation of cell cycle) were downregulated with increasing paclitaxel concentration, irrespective of the presence of multipolar spindle formation (Fig. 4F). The genes included in the Gene Ontology (GO) terms consistently exhibited the downregulation (Fig. 4G).

Subsequently, in trajectory 1 (no multipolar spindle formation), GSEA highlighted the induction of endoplasmic reticulum (ER) stress, which, if sustained or is severe, can potentially trigger apoptosis (Fig. 4H).[15] Conversely, in trajectory 2, characterised by multipolar spindle formation, the *RFC4* gene, known for its role in DNA replication and repair,[16] consistently exhibited upregulation in response to increasing paclitaxel exposure (Fig. 4I). These findings underscore the remarkable power of integrated multimodal data analysis, and its ability to effectively distinguish the gene regulation between two trajectories associated with distinct phenotypic outcomes.
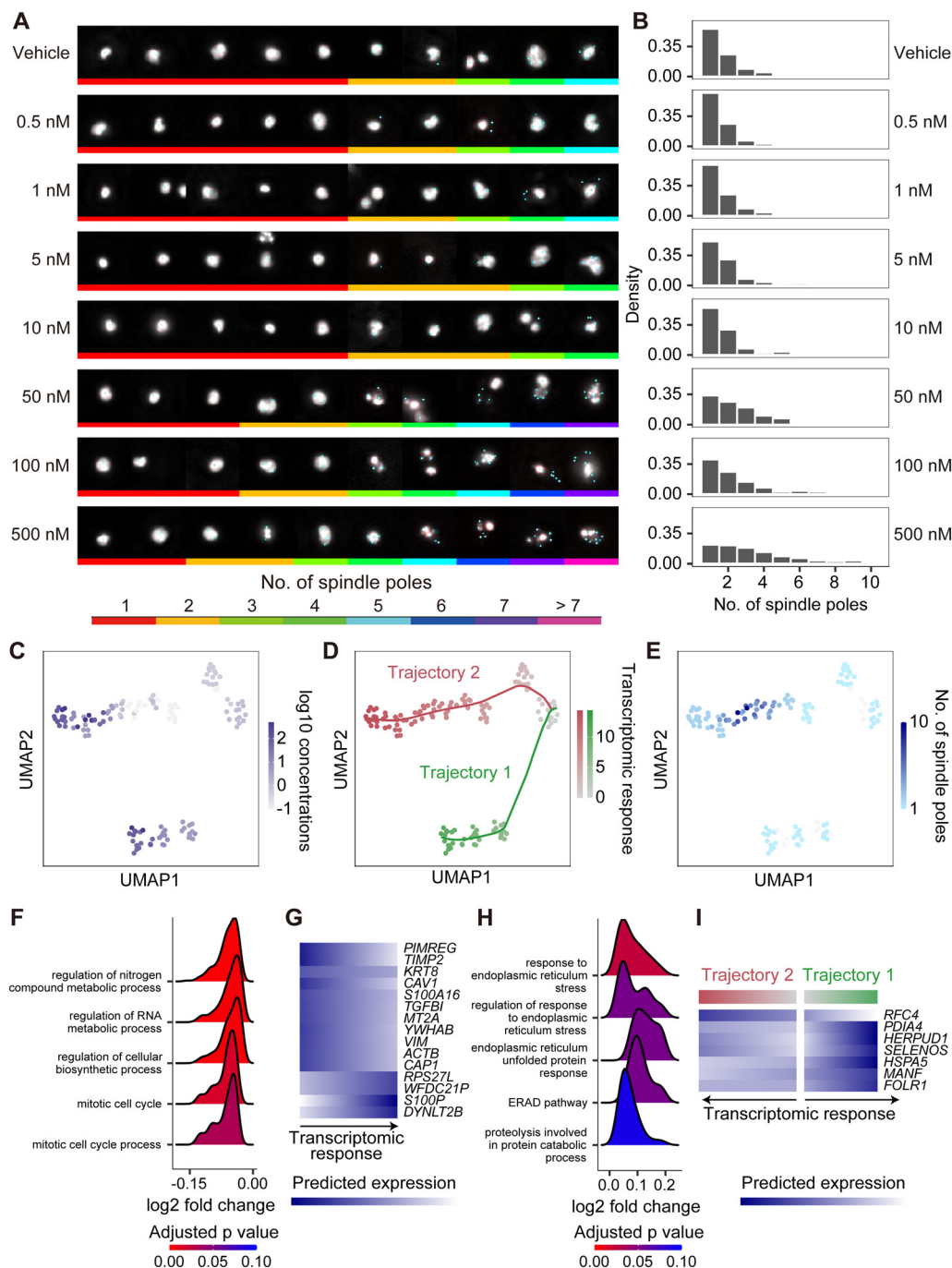
Fig. 4 Integrated analysis of paclitaxel-induced transcriptomic and phenotypic response. (A) On-chip images of nuclei in HeLa cells treated with different concentrations of paclitaxel. The cyan points denote the positions of spindle poles. The colour labels on the bottom of the images indicate the number of spindle poles. The images are stratified according to the number of spindle poles. (B) Distributions of the number of spindle poles in HeLa cells. Uniform manifold approximation and projection (UMAP) of the integrated dataset by weighted nearest neighbour (WNN) analysis. (C) The colour intensity represents the concentrations of paclitaxel. (D) The magnitudes of transcriptomic responses in the two distinct trajectories, and (E) the numbers of spindle poles. (F) Gene set enrichment analysis (GSEA) of differentially expressed genes associated with the increase in transcriptomic response shared among the two trajectories, and (G) its predicted expression of individual genes. (H) GSEA comparing the two distinct trajectories with and without multipolar spindle formation, and (I) predicted expression of individual genes.

Discussion

Multiplex chemical transcriptomics provides mechanistic insights into the cellular responses to chemical perturbations at the molecular level and offers a comprehensive understanding across pooled conditions, suppressing the batch effect.[17,18] Cellular tagging is a key to demultiplex genetically identical cells, and its applicability can be

extended to study chemical-dependent or dose-dependent responses. However, difficulty remains in transcriptomics-based screening in linking molecular responses to key phenotypic expression, such as cell proliferation and morphological change, which are typically profiled by quantitative optical microscopy. The integration of microscopical phenotyping and molecular profiling provides a unique opportunity to dissect the molecular cascades that cause the specific phenotypic expression.[19]

There are two major strategies for integrated phenotypic and transcriptomics analyses. The first is the optical decoding of the barcode sequence by sequential fluorescence *in situ* hybridisation, and the second is physical isolation of the cell of interest and then indexing by known barcode tags. SCOPE-seq2 employs the former strategy, and decodes the cell barcode of the hydrogel beads by performing cyclic hybridisation and readout with automated microfluidic control and microscopic imaging.[12] As an example of the latter strategy, an automated cell-picking system was employed to isolate single cells into 96-well plates and then process them for scRNA-seq.[20]

In contrast to these methods, our approach to link the cellular phenotype with transcriptomics uses a combination of colour codes of cells and hydrogel beads to optically index pairs of single cells and hydrogel beads. Automated microfluidic controls and robotic systems are not required for this method, which incorporates fewer on-chip steps, and is compatible with standard epi-fluorescence microscopy, which are advantageous features because they can be implemented in a standard laboratory.

As demonstrated in our experiments to measure the paclitaxel burden on HeLa cells, the cell colour code also works as cell hashing for multiplex chemical screening. Our analysis revealed two distinct trajectories in transcriptomic response by paclitaxel treatment, which correspond to distinct phenotypic reactions. The trajectory that involved no multipolar spindle formation showed upregulation of the ER stress response. Prolonged and severe ER stress may induce apoptosis, or lead to the acquisition of drug resistance[21] through the activation of unfolded protein response (UPR), a signalling pathway involved in adaptive and apoptotic response.[22] The second trajectory exhibited generations of multipolar spindles, leading to chromosome missegregation and cell death.[14]

Our approach has the potential to be extended to the integrated analysis of dynamic phenotyping and transcriptomics using longitudinal live-cell imaging. The DNA tags were retained within the cells 48 h after labelling. For instance, the integration of our approach with longitudinal imaging of leukocytes at the sites of active inflammation can be used to potentially classify leukocytes by spatiotemporal behaviours[23] and determine the molecular background. Furthermore, our approach can be readily integrated with the profiling of surface proteins *via* CITE-seq,[6] thereby enabling the analysis to be coupled with another omics layer. The proposed opto-combinatorial indexing is also compatible with cell-

hashing using DNA-tagged antibodies[24] or lipids.[25] The transfection-based approaches (electroporation or lipofection) used in our demonstration are robust and cost-effective, for instance, when assaying cells from non-model organisms.

In analysing the drug response of HeLa cells to paclitaxel, we labelled nuclei with a fluorescent dye (Hoechst 33342) to observe the nuclear morphology, resulting in a reduction in the number of cell colour codes. We envision that increasing the number of fluorescence channels by quantitatively demultiplexing the fluorophores with spectral overlap is the key to improving scalability and increasing observable phenotypic parameters. In the future, we hope to demonstrate high-content and improved multiplexing by increasing the fluorescence channels and using unmixing approaches.[26]

In conclusion, opto-combinatorial indexing provides a simplified strategy to simultaneously analyse the image and gene expression from single cells, and effectively dissect the molecular background of distinct phenotypic behaviours by integrating cellular phenotype and transcriptomics data.

Methods

Synthesis of colour-coded hydrogel beads

To allow optical readout of the bead colour codes, we designed the branched DNA that hybridises with the bead colour code and converts the nucleotide sequence to a bright or dim combination of four fluorophores by hybridising four readout oligos with or without fluorophores, creating $2^4 = 16$ different colour combinations (Table S1†). This approach minimizes the number of readout oligos labelled with fluorophores and significantly reduces oligo synthesis costs. We synthesized the polyacrylamide hydrogel beads with poly(dT) sequences through two rounds of split-pool ligation.[27]

We generated droplets of acrylamide premix with 50 μM acrydited primer (5′-Acryd/AAGCAGTGGTATCAACGCAGAGTA CGACGCTCTT-3′) using a simple coflow microfluidic device. The final bead size was approximately 40 μm. We then ligated the first barcode fragments containing the bead colour code to the first part of the cell barcode (Table S1,† Stem_CC00_ID01–Stem_CC15_ID25). In the second round of ligation, we added fragments with the second part of the cell barcode, UMI and poly(T) (NNNNNNNNNTTTTTTTTTTTTTTTT TTTTTTTTTTTTVN) (Table S1,† ID00_dT–ID32_dT). The combination of the first and second parts of the cell barcode created 6400 unique barcodes.

To stain beads with colour codes, we pooled the beads with cell barcodes in a single tube and combined 3×10^4 beads, a mixture of 6 μM branched oligos (Table S1,† Branch_00_NNNN–Branch_15_BGPR), a mixture of 12 μM readout oligos (with Alexa 488, Alexa 555, Alexa 647, and Alexa 750, Table S1,† Readout_Alexa647–Readout_Alexa750), 12 μM oligo without fluorophore (Table S1,† Readout_R0–Readout_B0 to fill the sequence in branched oligos for dim beads), and 0.1 mg mL^{-1} salmon sperm DNA in hybridisation

buffer (5 mM Tris-HCl (pH 8.0), 1 M KCl, 5 mM EDTA, 0.05% (vol/vol) Tween-20). We incubated the mixture at 94 °C for 5 min and cooled it by 5 °C every 5 min to 25 °C, and then maintained it at 4 °C. Excess probes were washed three times with an ice-cold hybridisation buffer.

We reasoned that hybridisation-based staining of hydrogel beads insignificantly affects the synthesis of cDNA and the PCR amplification because the branched oligo hybridises downstream of the cDNA extension during reverse transcription, and the concentration of the branched oligo in PCR is estimated at 3.52 nM per colour code, while that of the PCR primer is at 240 nM. Furthermore, the melting temperature of the branched oligo is lower at 66.4 °C than that of the PCR primer at 77.5 °C (under conditions of 50 mM Na^+ and 3 mM Mg^{2+} as an example), while the annealing temperature for PCR is 65 °C.

Colour-coded cell preparation

We cultured HeLa (RCB0007, RIKEN BRC) and NIH/3T3 cells (RCB2767, RIKEN BRC) in Dulbecco's modified Eagle's medium (DMEM, 08456-65, Nacalai Tesque) containing 10% fetal bovine serum (FBS, 26140-079, Gibco) and 1% penicillin–streptomycin (P/S, P4333-100ML, Sigma-Aldrich).

For the species-mixing experiment, we separately seeded HeLa cells and NIH/3T3 cells in a 100 mm dish at a concentration of 2.0×10^5 cells per mL for each, and cultured them for 1 day. After trypsinisation, we aliquoted cells equally into eight sub-pools per cell type in 16 tubes and individually stained them with 16 different combinations of four types of CellTrace dyes (5 μM Violet, 5 μM CFSE, 5 μM Yellow, and 1 μM Far Red, Invitrogen™) at the concentration of 1.0×10^6 cells per mL. We then individually suspended the stained cells in Gene Pulser® Electroporation Buffer (Bio-Rad) with 100 nM of DNA tag (Table S2†) corresponding to the respective fluorescence colour.

Immediately after electroporation by the Gene Pulser Mxcell™ Electroporation System (Bio-Rad, voltage: 250 V, capacitance: 2000 μF, resistance: ∞ Ohm, duration: 20 ms for HeLa cells, and 400 V, capacitance: 950 μF, resistance: ∞ Ohm, duration: 20 ms for NIH/3T3 cells), we added a five-fold volume of culture medium and incubated the cells for 1 h at 37 °C in 5% CO_2. We washed the cells three times with 1× phosphate-buffered saline (PBS, 14249-24, Nacalai Tesque), and combined the 16 sub-pools in loading buffer (1% polyvinylpyrrolidone in 1× PBS).

For the paclitaxel treatment experiments, we seeded the HeLa cells at the concentration of 5.0×10^4 cells per mL in eight wells of a 24-well plate and cultured them for one day. We individually transfected cells with different types of DNA tags using Lipofectamine® 3000 reagent (Invitrogen™) according to the manufacturer's protocol (incubation for 4 h at 20 nM DNA tag concentration). After lipofection, we washed the cells three times with culture medium. We individually stained the cells in each well with eight combinations of three CellTrace dyes (5 μM CFSE, 5 μM Yellow, and 1 μM Far Red). We cultured cells for one day in culture medium containing paclitaxel (163-28163, FUJIFILM) at different concentrations. We washed the cells three times with 1× PBS. After trypsinisation, the eight sub-pools of cells at equal concentration were pooled equally into one tube. Subsequently, the cells were stained with 10 μg mL^{-1} Hoechst 33342 and resuspended in loading buffer.

For the tag retention assay, we seeded HeLa cells at a concentration of 5.0×10^4 cells per mL and cultured them for one day, followed by lipofection with 20 nM DNA tags and culturing for 4 h. In addition, we performed electroporation on HeLa cells at 1.0×10^6 cells per mL in Gene Pulser® Electroporation Buffer with 100 nM DNA tag under the same conditions described above. We then cultured the cells in a 24-well plate for varying durations up to 48 h and stained them with six different combinations of three CellTrace dyes (5 μM CFSE, 5 μM Yellow, and 1 μM Far Red). After three PBS washes, we pooled all of the cells at the same concentration and resuspended them in loading buffer.

Chip fabrication

The workflow is based on the protocol reported in previous works.[1,2] We fabricated microwell arrays from polydimethylsiloxane (PDMS, SILPOT 184, Dow Corning) by soft lithography using an SU-8 mould. To hydrophilise the microwell array and to perform efficient sealing in the cell lysis and hybridisation step, we functionalized the array according to a previously reported protocol.[2]

On-chip experimental workflow

We placed a PDMS slab with the microwell array superstructure onto a glass-based dish (3961-035, IWAKI), dispensed PBS over the microwells, and maintained it under vacuum for 15 min to remove any bubbles in the microwells. We then dropped pooled cells suspended in loading buffer onto the microarray and incubated them for 5 min at room temperature to allow the cells to settle. After washing the microwell array with PBS, we added 2 mL of DMEM without phenol red (08490-05, Nacalai Tesque) and subsequently acquired scanned images of the microwell array containing the cells. In every experiment, we adjusted the exposure times to effectively use the full dynamic range of the camera, and we used the same setting for the entire chip.

We dropped 20 μL of the colour-coded bead suspension at the concentration of 1.0×10^6 beads per mL onto the microwell array and incubated them for 10 min to capture the beads in the microwells, followed by tapping and resting for 1 min at 37 °C, which was repeated five times. To seal the microwells with a track-etched membrane with nanopores of 10 nm in diameter (Sterlitech), the membrane was pre-treated with atmospheric plasma (BD-20, Electro-Technic Products) for 60 s and then hydrated in PBS. Then, we pressed the membrane and the PDMS slab by placing a glass slide (8 mm square per side, S7214, MATSUNAMI) and a 100 g weight at 37 °C for 30 min. We removed the glass slide by adding 3 mL of PBS. We lysed

the cells with 3 mL of cell lysis buffer (5 M guanidine thiocyanate, 1 mM EDTA, 0.50% sarkosyl (*N*-laurylsarcosine), 1.0% 2-mercaptoethanol), and then agitated them in a microplate shaker at 5–60 rpm for 20 min.

Next, we washed the microwell array with 3 mL of hybridisation buffer (2 M NaCl, 0.64% PEG-8000, 0.52× PBS) to hybridise the released mRNA and DNA tag with the primers on the hydrogel beads by agitation at 5–60 rpm for 40 min. After removing the membrane over the microwell array, we acquired images of the microwell array containing the hydrogel beads with adjusted exposure times. To collect beads, we placed the microwell array directly into a 200 μL tube and flushed it with 200 μL of wash solution 1 (2 M NaCl, 3 mM MgCl$_2$, 20 mM Tris-HCl (pH 8.0), 0.64% PEG-8000, 0.1% Tween 20). We exchanged wash solution 1 for wash solution 2 (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 6 mM MgCl$_2$, 0.4 U μL^{-1} Recombinant RNase Inhibitor (2313A, Takara), 0.1% Tween 20) by repeating centrifugation (3000 × *g*, 4 °C, 3 min) and the buffer exchange twice.

Library preparation

To construct cDNA and DNA-tag libraries, we performed reverse transcription (RT) and polymerase chain reaction (PCR) according to the CITE-seq protocol[6] with modifications. We added 10 μL of RT mix (1× first-strand buffer, 4.8 μM biotinylated template-switching oligonucleotide (TSO, Qiagen), 2 mM dNTP mix, 4 mM dithiothreitol (DTT), 2 U μL^{-1} Recombinant RNase Inhibitor, and 20 U μL^{-1} SMARTScribe Reverse Transcriptase (Takara)) to 10 μL of suspended beads and incubated them in a thermal cycler at 42 °C for 90 min to obtain first-strand cDNA by reverse transcription. Then, the beads were heated at 70 °C for 10 min to stop the reaction. To remove the excess RT primers, we added 2 μL of 2.5 U μL^{-1} Exonuclease I (2650A, Takara) and incubated the beads at 37 °C for 50 min, followed by inactivation by heating at 80 °C for 20 min.

The first-strand cDNA was amplified by PCR in 50 μL of reaction solution containing 0.24 μM primer 2, 9 nM additive primer, 1× SeqAmp PCR Buffer and 0.025 U μL^{-1} SeqAmp DNA Polymerase (Takara) (Table S2†) using the following program: 95 °C for 1 min; 16–18 cycles of 98 °C for 10 s and 65 °C for 30 s; 68 °C for 4 min; and 72 °C for 10 min. We purified the mRNA-derived cDNAs (long cDNA) and the DNA tag-derived cDNAs (<200 bp short cDNA) by size fractionation using SPRIselect beads (B23318, Beckman Coulter).

We mixed the cDNA product with 0.6× SPRIselect beads and placed it on the DynaMag™-Spin Magnet (Invitrogen™) to capture the beads. We then transferred the first supernatant containing the DNA-tagged product to a new tube and further purified it with 1.4× SPRI beads. The magnetic beads were washed three times (long) and twice (short) on a magnetic stand with 0.2 mL of fresh 80% (vol/vol) ethanol and were then air-dried for 2.5 min at room temperature. The cDNAs derived from mRNA and DNA tags

were eluted with 13 and 11 μL of elution buffer (10 mM Tris-HCl, pH 8.5), respectively.

For the mRNA-derived cDNA, we examined the yield, quality, and size distribution using a Qubit 4 Fluorometer (Thermo Fisher Scientific) and quantitative real-time PCR (qPCR) targeting *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase, Hs02758991_g1, Thermo Fisher Scientific) with an Agilent High Sensitivity DNA Kit and the Bioanalyzer 2100 (Agilent). We then performed the tagmentation of 600 pg of cDNA using a Nextera XT DNA Library Prep Kit (Illumina) and subsequent PCR with custom indexing primers. The PCR products were then purified with 0.6× SPRIselect beads and eluted with 6.5 μL of Resuspension Buffer (RSB, Illumina).

To construct the DNA tag library, the short cDNA was amplified in 20 μL of 1× KAPA Hifi Hotstart Ready Mix (Roche) containing 1.6 μL of 10-fold diluted templates and 0.25 or 0.5 μM of indexing primers (Table S2†) according to the following program: 98 °C for 2 min; 2 cycles of 98 °C for 20 s and 74 °C for 30 s; 12–18 cycles of 98 °C for 20 s and 72 °C for 30 s; and 72 °C for 5 min. The library was then purified with 1.5× SPRI beads and eluted with 8 μL of elution buffer. We assessed the yield and length of the library using a Qubit™ dsDNA HS Assay Kit and an Agilent High Sensitivity DNA Kit (Agilent), respectively.

The typical size of the mRNA-derived cDNA library was approximately 500 bp, while the size of DNA tag libraries was 224 bp. Finally, we quantified the library using KAPA Library Quantification Kits (Roche) according to the manufacturer's protocol. The library was sequenced on a HiSeq X (Illumina) instrument with 2 × 150 bp paired-end reads.

Cell and bead image processing

We scanned the entire PDMS chip with the Micro-Magellan[28] for both the cell imaging and the hydrogel bead imaging. We performed flat-field correction on all fluorescence images using a built-in MATLAB function before stitching them together. We visually detected the cells and beads captured in the wells and registered their respective colour codes. To identify cell and bead pairs co-captured in the same wells, we applied an affine transformation to cell images to align the positions of the microwells of cell images to the microwells of bead images.

When the centres of the cells were within the radii of the bead-captured wells, we assigned them as co-captured pairs. To compensate for the different focal lengths of the nuclei of paclitaxel-treated cells in the microwells, z-stack images were obtained with a 10× lens (Olympus UPlanFL N 10×) at 3 μm intervals from the bottom to the top of the wells. We then executed an extended depth-of-field algorithm[29] provided by Fiji software. We counted the spindle poles using CellProfiler[30] by enhancing the speckles with the 'EnhabceOrSuppressFeature' module and segmenting them into individual poles with the 'IdenfifyPrimaryObjects' module.

Single-cell RNA data processing

We used UMI-tools[31] to demultiplex the sequence reads derived from cDNA and DNA tags into each cell barcode and each UMI. We demultiplexed DNA-tag UMI counts into each tag type using the CITE-seq-Count program. We mapped cDNA reads to the reference genomes and transcriptomes of GRCh38 (human, p12 for the species-mixing experiments and p13 for the paclitaxel treatment experiments) and GRCm38. p6 (mouse) using the STAR (version 2.7.10b) mapping program.[32] We removed all cells with less than 800 UMI counts. To link the sequencing data and the image data, we assessed the cosine similarities between the centred log ratio (CLR) of the UMI counts of DNA tags from the sequencing data and dummy variables of the joint colour code from the image data.

We assumed that cosine similarity is a function of the signal-to-noise ratio of tag counts expressed by the following equation:

$$(\text{Cosine similarity}) = \boldsymbol{c}\cdot\boldsymbol{e}/(|\boldsymbol{c}||\boldsymbol{e}|) = s/\left(1\cdot\sqrt{(s^2+(k-1)n^{-2})}\right)$$
$$= 1/\sqrt{(1+(k-1)n^{-2}/s^2)}$$

where $\boldsymbol{c}$ denotes a vector of the one-hot encoded joint colour code, $\boldsymbol{e}$ denotes a vector of the tag count, which consists of an element of signal s and others of noise n ($\bar{n}$ is the mean value), and k denotes the pooling number of tags. We optimized the combinations that maximized the sum of the cosine similarities within each chip and further filtered out the linked data whose signal-to-noise ratios were less than $\sqrt{5}$. To remove batch effects, we integrated the data from different batches with the functions of '*SelectIntegrationFeatures*', '*FindIntegrationAnchors*', and '*IntegrationData*' from the Seurat (version 4.3.0.1)[33] package.

Weighted nearest-neighbour analysis for paclitaxel-treated cells

For paclitaxel-treated cells, we filtered out the cells with less than 2000 UMI count, excluded data linked to doublet cells in a well, and genes with low detection rates below 0.2, and mitigated the batch effect as described above. Subsequently, we isolated cells whose joint colour code was unique on the chip. We then transformed gene expression into the principal components using the '*runPCA*' function from the Seurat package.

Furthermore, we combined paclitaxel concentrations (log 10-transformed with a 0.1 offset) and the number of spindle poles to create principal components using the '*prcomp*' function from the stats package. To incorporate both sets of principal components in subsequent analyses, we performed a WNN analysis[33] utilizing the '*FindMultiModalNeighbors*' function (with a k-nearest neighbors' parameter of 25) from the Seurat package.

Extraction of transcriptomic response

We clustered cells and projected them on the uniform manifold approximation and projection (UMAP) based on the neighbouring information from WNN using the '*FindCluster*' and '*RunUMAP*' functions from the Seurat package. Using the clusters and UMAP, we performed a trajectory analysis with slingshot[34] and extracted the transcriptomic response, ψ, with the '*slingAvgPseudotime*' function. To determine the differentially expressed genes that share the increase in transcriptomic response within trajectories, we fitted the gene expression with the following model using the edgeR[35] package:

$$\log(\mu_{gi}) = \beta_0 + \beta_\psi \psi + \log(N_i)$$

where μ_{gi} denotes an expected expression of gene g in a cell i calculated by edgeR, ψ denotes the transcriptomic response, and N_i denotes a size factor of cell i. We assessed the significance of to derive fold changes and p values with the quasi-likelihood F-test. To determine differentially expressed genes between trajectories, we fit the gene expression with the following model:

$$\log(\mu_{gi}) = \beta_0 + \beta_{\psi,1}t_1\psi + \beta_{\psi,2}t_2\psi + \log(N_i)$$

where t_k indicates a factor denoting the assignment to trajectory k. We assessed the significance of $\beta_{\psi,2} - \beta_{\psi,1}$ to derive fold changes and p values.

Data availability

The sequencing data generated in this study have been deposited in the NCBI BioProject under accession code PRJNA1027139.

Author contributions

A. T., K. N., and M. K. performed experiments; A. T., T. K., and H. S. analysed the data. R. Y. assisted in analysing the data. A. T., T. K., and H. S. designed experiments. H. S. supervised the project. A. T., K. N., T. K., and H. S. wrote the original manuscript. All authors approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

References

1 T. M. Gierahn, M. H. Wadsworth, 2nd, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S. Fortune, J. C. Love and A. K. Shalek, *Nat. Methods*, 2017, **14**, 395–398.

2 T. K. Hughes, M. H. Wadsworth, 2nd, T. M. Gierahn, T. Do, D. Weiss, P. R. Andrade, F. Ma, B. J. de Andrade Silva, S. Shao, L. C. Tsoi, J. Ordovas-Montanes, J. E. Gudjonsson, R. L. Modlin, J. C. Love and A. K. Shalek, *Immunity*, 2020, **53**, 878–894 e877.

3 A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner, *Cell*, 2015, **161**, 1187–1201.

4 E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev and S. A. McCarroll, *Cell*, 2015, **161**, 1202–1214.

5 V. M. Peterson, K. X. Zhang, N. Kumar, J. Wong, L. Li, D. C. Wilson, R. Moore, T. K. McClanahan, S. Sadekova and J. A. Klappenbach, *Nat. Biotechnol.*, 2017, **35**, 936–939.

6 M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija and P. Smibert, *Nat. Methods*, 2017, **14**, 865–868.

7 H. Chung, C. N. Parkhurst, E. M. Magee, D. Phillips, E. Habibi, F. Chen, B. Z. Yeung, J. Waldman, D. Artis and A. Regev, *Nat. Methods*, 2021, **18**, 1204–1212.

8 S. Chen, B. B. Lake and K. Zhang, *Nat. Biotechnol.*, 2019, **37**, 1452–1457.

9 M. N. Abdelmoez, K. Iida, Y. Oguchi, H. Nishikii, R. Yokokawa, H. Kotera, S. Uemura, J. G. Santiago and H. Shintaku, *Genome Biol.*, 2018, **19**, 66.

10 Y. Oguchi, Y. Ozaki, M. N. Abdelmoez and H. Shintaku, *Sci. Adv.*, 2021, **7**, eabe0317.

11 J. Yuan, J. Sheng and P. A. Sims, *Genome Biol.*, 2018, **19**, 227.

12 Z. Liu, J. Yuan, A. Lasorella, A. Iavarone, J. N. Bruce, P. Canoll and P. A. Sims, *Sci. Rep.*, 2020, **10**, 19482.

13 L. M. Zasadil, K. A. Andersen, D. Yeum, G. B. Rocque, L. G. Wilke, A. J. Tevaarwerk, R. T. Raines, M. E. Burkard and B. A. Weaver, *Sci. Transl. Med.*, 2014, **6**, 229ra243.

14 C. M. Scribano, J. Wan, K. Esbona, J. B. Tucker, A. Lasek, A. S. Zhou, L. M. Zasadil, R. Molini, J. Fitzgerald, A. M. Lager, J. J. Laffin, K. Correia-Staudt, K. B. Wisinski, A. J. Tevaarwerk, R. O'Regan, S. M. McGregor, A. M. Fowler, R. J. Chappell, T. S. Bugni, M. E. Burkard and B. A. Weaver, *Sci. Transl. Med.*, 2021, **13**, eabd4811.

15 P. C. Liao, S. K. Tan, C. H. Lieu and H. K. Jung, *J. Cell. Biochem.*, 2008, **104**, 1509–1523.

16 Y. Li, S. Gan, L. Ren, L. Yuan, J. Liu, W. Wang, X. Wang, Y. Zhang, J. Jiang, F. Zhang and X. Qi, *Am. J. Cancer Res.*, 2018, **8**, 1343–1355.

17 S. R. Srivatsan, J. L. McFaline-Figueroa, V. Ramani, L. Saunders, J. Cao, J. Packer, H. A. Pliner, D. L. Jackson, R. M. Daza, L. Christiansen, F. Zhang, F. Steemers, J. Shendure and C. Trapnell, *Science*, 2020, **367**, 45–51.

18 J. M. McFarland, B. R. Paolella, A. Warren, K. Geiger-Schuller, T. Shibue, M. Rothberg, O. Kuksenko, W. N. Colgan, A. Jones, E. Chambers, D. Dionne, S. Bender, B. M. Wolpin, M. Ghandi, I. Tirosh, O. Rozenblatt-Rosen, J. A. Roth, T. R. Golub, A. Regev, A. J. Aguirre, F. Vazquez and A. Tsherniak, *Nat. Commun.*, 2020, **11**, 4296.

19 T. N. Chen, A. Gupta, M. D. Zalavadia and A. Streets, *Lab Chip*, 2020, **20**, 3899–3913.

20 J. Jin, T. Ogawa, N. Hojo, K. Kryukov, K. Shimizu, T. Ikawa, T. Imanishi, T. Okazaki and K. Shiroguchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2210283120.

21 J. R. Cubillos-Ruiz, S. E. Bettigole and L. H. Glimcher, *Cell*, 2017, **168**, 692–706.

22 G. C. Shore, F. R. Papa and S. A. Oakes, *Curr. Opin. Cell Biol.*, 2011, **23**, 143–149.

23 G. Crainiciuc, M. Palomino-Segura, M. Molina-Moreno, J. Sicilia, D. G. Aragones, J. L. Y. Li, R. Madurga, J. M. Adrover, A. Aroca-Crevillen, S. Martin-Salamanca, A. S. Del Valle, S. D. Castillo, H. C. E. Welch, O. Soehnlein, M. Graupera, F. Sanchez-Cabo, A. Zarbock, T. E. Smithgall, M. Di Pilato, T. R. Mempel, P. L. Tharaux, S. F. Gonzalez, A. Ayuso-Sacido, L. G. Ng, G. F. Calvo, I. Gonzalez-Diaz, F. Diaz-de-Maria and A. Hidalgo, *Nature*, 2022, **601**, 415–421.

24 M. Stoeckius, S. Zheng, B. Houck-Loomis, S. Hao, B. Z. Yeung, W. M. Mauck, 3rd, P. Smibert and R. Satija, *Genome Biol.*, 2018, **19**, 224.

25 C. S. McGinnis, D. M. Patterson, J. Winkler, D. N. Conrad, M. Y. Hein, V. Srivastava, J. L. Hu, L. M. Murrow, J. S. Weissman, Z. Werb, E. D. Chow and Z. J. Gartner, *Nat. Methods*, 2019, **16**, 619–626.

26 J. Seo, Y. Sim, J. Kim, H. Kim, I. Cho, H. Nam, Y. G. Yoon and J. B. Chang, *Nat. Commun.*, 2022, **13**, 2475.

27 R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein and L. Mazutis, *Nat. Protoc.*, 2017, **12**, 44–73.

28 H. Pinkard, N. Stuurman, K. Corbin, R. Vale and M. F. Krummel, *Nat. Methods*, 2016, **13**, 807–809.

29 B. Forster, D. V. D. Ville, J. Berent, D. Sage and M. Unser, Extended Depth-of-Focus for Multi-Channel Microscopy Images: A Complex Wavelet Approach, *IEEE International Symposium on Biomedical Imaging*, 2004, vol. 1, pp. 660–663.

30 D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini and A. Goodman, *BMC Bioinf.*, 2021, **22**, 433.

31 T. Smith, A. Heger and I. Sudbery, *Genome Res.*, 2017, **27**, 491–499.

32 A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras, *Bioinformatics*, 2013, **29**, 15–21.

33 Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert and R. Satija, *Cell*, 2021, **184**, 3573–3587 e3529.

34 K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom and S. Dudoit, *BMC Genomics*, 2018, **19**, 477.

35 M. D. Robinson, D. J. McCarthy and G. K. Smyth, *Bioinformatics*, 2010, **26**, 139–140.