Published online 10 December 2021

Nucleic Acids Research, 2022, Vol. 50, No. 5 ***e28***
https://doi.org/10.1093/nar/gkab1201

Grouped-seq for integrated phenotypic and transcriptomic screening of patient-derived tumor organoids

Yushuai Wu[1], Kaiyi Li[1], Yaqian Li[2], Tao Sun[3], Chang Liu[1], Chunhui Dong[2], Tian Zhao[4], Decong Tang[5], Xiaojie Chen[5], Xiaofang Chen[2,4,*] and Peng Liu[1,*]

[1]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China, [2]School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China, [3]Department of General Surgery, Peking University Third Hospital, Beijing 100191, China, [4]Beijing Organobio Corporation, Beijing 102206, China and [5]Beijing NeoAntigen Biotechnology Co. Ltd, Beijing 102206, China

Received September 07, 2021; Revised November 17, 2021; Editorial Decision November 17, 2021; Accepted November 23, 2021

ABSTRACT

Patient-derived tumor organoids (PDOs) have emerged as a reliable *in vitro* model for drug discovery. However, RNA sequencing-based analysis of PDOs treated with drugs has not been realized in a high-throughput format due to the limited quantity of organoids. Here, we translated a newly developed pooled RNA-seq methodology onto a superhydrophobic microwell array chip to realize an assay of genome-wide RNA output unified with phenotypic data (Grouped-seq). Over 10-fold reduction of sample and reagent consumption together with a new ligation-based barcode synthesis method lowers the cost to ∼$2 per RNA-seq sample. Patient-derived colorectal cancer (CRC) organoids with a number of 10 organoids per microwell were treated with four anti-CRC drugs across eight doses and analyzed by the Grouped-seq. Using a phenotype-assisted pathway enrichment analysis (PAPEA) method, the mechanism of actions of the drugs were correctly derived, illustrating the great potential of Grouped-seq for pharmacological screening with tumor organoids.

INTRODUCTION

Compelling evidence has proved patient-derived tumor organoids (PDOs) recapitulate the phenotypic and genotypic characteristics of the original tumor tissues in many types of cancers (1–5), including colorectal cancer (CRC) (6–8), breast cancer (9), lung cancer (LC) (10,11), etc. In addition, PDOs are relatively easy to handle and can proliferate *in vitro* indefinitely, making the PDO a perfect tumor model for drug discovery and precision medicine (1–3). Indeed, many pioneering studies have demonstrated that drug screening based on patient-derived tumor organoids can be employed to investigate gene-drug associations, to discover new and novel therapeutics, and to evaluate drug actions (12,13). However, the generation of a large number of organoids is time-consuming (3,4) and the use of an extracellular matrix that supports the growth of organoids makes the handling of organoids cumbersome. As a result, the application of tumor organoids in a high-throughput screening (HTS) is still rare. Moreover, the responses of PDOs to drugs are usually evaluated with some simple, easy-to-measure readouts, such as cell viability (3,8,11), which do not require many cells. Apparently, the true values of the PDOs as a more biologically relevant model for drug discovery have not been fully explored.

Microfluidic technology has been adopted to reduce the sample consumption, cost, and turnaround time of organoid culture and analysis by transferring assays onto miniaturized and parallelized microfabricated devices (14,15). For instance, Brandenberg *et al.* developed microcavity array devices together with scalable and automated protocols for suspension culture and real-time analysis of gastrointestinal organoids (16). Schuster *et al.* demonstrated a high-throughput microfluidic system for 3D organoid culture and analysis, enabling combinatorial and dynamic drug treatments to hundreds of organoid cultures (17). Recently, our group also demonstrated that a superhydrophobic microwell array device can expedite the drug sensitivity test of patient-derived lung cancer organoids to a week (10). The efficient utilization of organoids was realized on these microdevice due largely to the nanoliter-scale microfabricated structures, in which the reduced sample consumption matched the limited quantities of

*To whom correspondence should be addressed. Tel: +86 10 62798732; Email: pliu@tsinghua.edu.cn
Correspondence may also be addressed to Xiaofang Chen. Tel: +86 10 82339861; Email: xfchen@buaa.edu.cn

organoids, though the assays still centered on phenotypic analyses.

Actually, high-throughput phenotypic screening performed using simple measurements, such as cell viability, cell growth, protein-protein interactions, etc., often failed to elucidate the mechanisms of action (MoA) of the compounds unambiguously (18–20). A proposed strategy that may overturn this situation is to combine the phenotype-based screening with molecular-level approaches, in which a set of genes or genome-wide expression profiles are employed to depict the actions of drugs (21–23). Towards this goal, targeted transcriptional profiling platforms, such as RASL-seq and Luminex L1000, were firstly developed due to the consideration of costs and assay complexity (24,25). However, the limited gene panels can only test pre-defined hypotheses and any unexpected gene perturbations off the panels would be missed out. Recently, many sample-barcoding methodologies were developed to enable the pooling of cDNA from many samples prior to the sequencing library construction, leading to the significant reduction of sequencing costs (26–28). For example, PLATE-seq employs automated liquid handling to deliver well-specific, barcoded oligo (dT) primers to every sample in a 96-well plate, allowing 96 cell samples to be sequenced at ∼$15 per sample (27). DRUG-seq, a cost-effective digital transcriptional profiling method used for high-throughput screening, further increased the throughput to 384 or 1536 and reduced the costs to $2–4 per sample (28). While these sequencing-based methods have demonstrated great potential in broad pharmacological applications, all of them require a large number of cells and fall short in processing patient-derived organoids, due to the use of microliter-scale multi-well plates. Even though a recent study successfully developed a targeted RNA sequencing (TORNADO-seq) method for organoid-based drug screens, the requirement of ∼10 000 cells per sample and the limited genes that can be measured illustrate the urgent need of inventing a more suitable screening platform for tumor organoids (29).

Here, we developed a brand-new high-throughput screening technology called genome-wide RNA output unified with phenotypic data (Grouped-seq) for the integrated phenotypic and transcriptomic screening of patient-derived tumor organoids on a superhydrophobic microwell array chip (SMARchip). Compared with the previous reports, our platform has three key improvements: (a) a nanoliter-scale, high-throughput microwell array with flexible operations facilitating the culture, imaging, and sequencing of cell and organoid samples with a minimum number of starting materials, (b) a newly-designed library construction method combined with a low-cost microwell barcoding strategy enabling the pooling of cDNA prior to library construction and the backtracking of RNA-seq data to the phenotypic measurements and (c) a phenotype-assisted pathway enrichment analysis method for identifying compound MoAs from combined phenotypic and transcriptomic data. Collectively, our Grouped-seq platform provides over 10-fold reductions in reagent and sample consumptions without any loss in assay accuracy and reproducibility at an estimated cost of ∼$2 per sample.

MATERIALS AND METHODS

Preparation of CRC organoids

The recipe of the CRC organoid culture medium (CRCM) is listed in Supplementary Table S1. A surgically resected colorectal cancer tissue was obtained from a patient with an informed consent at the Peking University Third Hospital. The tissue was first minced into 2-mm cubes using a razor blade, followed by the filtration through a 100-μm cell strainer (Falcon) using a syringe plunger. After that, the filtrate was re-filtered through another 40-μm cell strainer (Falcon) and the tumor tissue retained in the strainer was collected and suspended into the CRCM for culture overnight. To culture CRC organoids in a 24-well cell culture plate, organoids in suspension were first centrifuged for 5 min at 500 × g in 4°C and resuspended in the cold growth factor-reduced Matrigel (BD Biosciences) with CRCM (1:1). Then, 50-μl drops of the Matrigel cell cluster suspension were added into the wells and allowed to solidify at 37°C for 20 min. After Matrigel stabilization, 500 μl of CRCM was added into the wells and the plates were transferred to a humidified 37°C/5% CO_2 incubator. The CRCM was replenished every 3–4 days and the organoids were passaged every 1–4 weeks. When passaging, organoids in gel were first transferred to a six-well low attachment plate and incubated with 2 ml organoids harvesting solution (R&D Systems) on an orbital shaker at 4°C for 1.5–2 h to dissolve the Matrigel. After that, organoids were mechanically sheared through a pipette and centrifuged at 1000 rpm for 3 min. Sheared organoids were washed with cold basal medium and centrifuged at 1000 rpm again. Finally, organoids were resuspended in the basal medium and Matrigel mixture (1:1) and reseeded at a suitable ratio (1:1 to 1:4).

Histology and sequencing of CRC organoids

The reagents used for CRC organoids handling are listed in Supplementary Table S2. The harvested organoids were washed with cold PBS, suspended in 40 μl of 10 mg/ml fibrinogen solution (Sigma-Aldrich), and then immediately mixed with 20 μl of thrombin reagent (Solarbio) for fibrin polymerization. After that, the fibrin hydrogel containing organoids together with the matched tissues were fixed in 1 ml of 4% paraformaldehyde (Sigma-Aldrich), followed by dehydration, paraffin embedding, sectioning, and a standard H&E staining protocol.

After dissolving the Matrigel, CRC organoids containing about 10^6 cells were collected for whole-exome sequencing. First, genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. DNA concentration was measured using the Qubit® DNA Assay Kit in Qubit® 3.0 Flurometer (Invitrogen). About 0.4 μg of genomic DNA was used to generate a sequencing library using the Agilent SureSelect Human All Exon V6 kit (Agilent). Products were purified using the AMPure XP system (Beckman Coulter) and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. Finally, the DNA library was sequenced on the Illumina Hiseq platform and

150 bp paired-end reads were generated. The sequencing reads were aligned to the human reference genome hg19 by Burrows-Wheeler Aligner (BWA) software v0.7.16a (30) to get the original mapping results in BAM format. SAMtools v1.9 (31) and Picard (http://broadinstitute.github.io/picard/) were used to sort BAM files. Then, duplicate marking, local realignment, and base quality recalibration were performed to generate final BAM file for computation of the sequence coverage and depth. The somatic SNV was detected by MuTect (32) and the somatic InDel was by Strelka (33).

Capture oligo synthesis and verification

The 114-base-long capture oligo that attaches to the streptavidin-coated Dynabeads™ is constructed by linking two short oligos (P1 and P2) together using a linker via enzymatic ligation. The sequences are listed in Supplementary Table S3, 4. A 100-μl mixture for the ligation is composed of 3 μl of P1 (10 μM), 1 μl of P2 (10 μM), 3 μl of linker (10 μM), 78 μl of nuclease-free water, 5 μl of T4 DNA ligase, and 10 μl of T4 DNA ligase buffer (all from NEB). The reaction is performed at 65 °C for 5 min, followed by rapid cooling on ice.

The ligation-based synthetic method was verified using flow cytometry (FCM) and sanger sequencing. In the experiment of flow cytometry, P1 was labeled with biotin for linking to beads and P2 was labeled with FAM for detection. Four samples were analyzed in the FCM assay, including two synthesis conditions (the P1/P2 ratio of 3/1 and 1/3), a positive control (PC, directly synthesized capture oligo labeled with biotin and FAM on two ends, respectively), and a negative control (NC, P1 and P2 without linkers). Each sample (10 μl) was mixed thoroughly with 10 μl of Dynabeads™ beads (Thermo Fisher) in a centrifugal tube and incubated for 15 min. Then, the tubes were put on a magnetic stand and washed with 20 μl of binding and washing buffer for three times. The beads were resuspended with nuclease-free water and used for FCM. Next, sanger sequencing was performed to further verify the sequence accuracy of the synthesized capture oligo. Since UMI and poly-T sequence were not suitable for PCR, we replaced them with known sequences as the primer binding site (Supplementary Table S5). 2 μl of the synthesized capture oligo was mixed with 23 μl of PCR mix which consists of 12.5 μl of 2× HotStart Readymix (Kapa Biosystems), 1 μl of 10-μM forward primer, 1 μl of 10-μM reverse primer, and 8.5 μl of nuclease-free water. The PCR program was as follows: 98°C 5 min, 30 cycles of 98°C 20 s, 58°C 15 s, and 72°C 20 s and 72°C 5 min. The amplicons were purified using the Hipure Gel Pure DNA Mini Kit (Magen) prior to sequencing.

Encoding magnetic beads on the bead array

The reagents used in the beads encoding are listed in is listed in Supplementary Table S6. Magnetic beads (Dynabeads™, Thermo Fisher) were washed three times with the binding and washing buffer on a magnet, followed by the loading to the microwells using the droplet rolling method. The microwells on the SMARchip were modified with 2% BSA to minimize non-specific adsorption of magnetic beads. An oligo array containing barcoded capture oligos, was aligned to the bead array upside and down and incubated for 15 min, in which the capture oligos were linked to beads through biotin-streptavidin binding. After the incubation, the oligo array was removed, the bead array was submerged into the PBS buffer to wash away unbound oligos for three times. Finally, a piece of salinized glass slide, on which a droplet array of 2× lysis buffer (200 mM Tris–HCl, 20 mM EDTA, 1% Sarkosyl, 50 mM DTT) was spotted, was aligned and covered upside down to the bead array to deliver the lysis buffer to the microwells. Since the droplet volume of bead array and lysis array are both ∼200 nl, the bead array has a final concentration of 1× lysis buffer for the following cell lysis.

Cell lysis and mRNA capture

The reagents used in the cell lysis are listed in is listed in Supplementary Table S6. By aligning the bead array upside and down to the cell array, magnetic beads settled down to the cells on the bottom of the microwells of the cell array. At the same time, lysis buffer diffused and lysed the cells *in situ*. The released mRNAs were immediately captured by the magnetic beads, which rested on the cells. After a 15-min incubation, the beads were gently washed off from the SMARchip into a 1.5 ml centrifuge tube with 1 ml of 6× SSC solution using a pipette. The beads were then washed twice with 600 μl of washing buffer A (10 mM Tris–HCl, 15 M NaCl, 1 mM EDTA and 0.1% Sarkosyl), once with 300 μl of washing buffer B (10 mM Tris–HCl, 15 M NaCl and 1 mM EDTA), and once with 50 μl of 5× Maxima H RT buffer (Thermo Fisher). The washed beads were finally suspended in 4 μl of 5× Maxima H RT buffer and stored on ice.

Reverse transcription and template switching

The reagents used in the reverse transcription are listed in is listed in Supplementary Table S6. The beads were mixed with 16 μl RT mix, which contained 10.5 μl of Nuclease-Free water, 2 μl of 10× dNTP, 2 μl of TSO (25 μM), 1 μl of Reverse Transcriptase (200 U/μl) and 0.5 μl of Rnase inhibitor (40 U/μl) (Thermo Fisher). Then, the beads were incubated at room temperature for 30 min, followed by 42°C for 90 min to obtain cDNA. Ten cycles of extended reaction (50°C for 3 min, 42°C for 3 min) could be chosen to increase the cDNA output when needed. The beads were then washed once with 20 μl of TE/SDS (1 × TE and 0.5% sodium dodecyl sulfate), twice with 20 μl of TE/TW (1 × TE and 0.01% Tween 20), once with 20 μl of ddH2O, and finally, re-suspended in 12 μl of ddH2O.

PCR pre-amplification

The reagents used in the PCR pre-amplification are listed in is listed in Supplementary Table S6. PCR mix containing 20 μl of 2× HotStart Readymix (Kapa Biosystems), 4 μl of 5′ end biotin-modified P7 primer (10 μM, Supplementary Table S7) and 4 μl of TSO primer (10 μM, Supplementary Table S7) was added to the beads. The PCR program was

as follows: 95 °C for 3 min; then 4 cycles of: 98 °C for 20 s, 65 °C for 45 s, and 72 °C for 3 min; then 8 cycles of 98 °C for 20 s, 67 °C for 20 s and 72 °C for 3 min; then a final extension step of 5 min at 72 °C. The PCR products were purified using 0.6× Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's instruction, and eluted into 50 μl of ddH_2O.

Fragmentation and adapter ligation

To prepare 3′-end cDNA fragments for sequencing, PCR products were fragmented by Covaris M220 (Covaris) following the manufacturer's instruction. After DNA shearing, end repair, 5′ phosphorylation, dA-tailing, adaptor ligation, size selection, and PCR enrichment were sequentially conducted using the NEBNext Ultra™ II DNA Library Prep Kit for Illumina (NEB). At last, PCR products were purified (AMPure XP system) and the library quality was assessed on the Agilent Bioanalyzer 2100.

Sequencing and data analysis

The libraries were paired-end sequenced (150 nt each) on the HiSeq-PE150 instrument (Illumina). The Grouped-seq data comprise of two paired-end reads: Read 1 contains a sequence that typically mapped to 3′ end of a mRNA transcript and Read 2 contains a wellcode (16 bases) that identifies a specific microwell on the SMARchip, and a UMI (10 bases) that corrects the PCR and sequencing bias. A single dataset (with two or four types of sample index) was generated from a single SMARchip for the downstream analysis.

The pipeline for bioinformatics analysis can be found on the Github (available at https://github.com/wuys13/Grouped-seq-analysis-pipeline). Read1 Fastq files were aligned to an appropriate reference genome using STAR v2.4.0a with the default settings (34). mm10 was used for mouse cells and hg38 was used for human cells. For samples with mouse and human cell mixtures, the union of hg38 and mm10 were used. Read2 Fastq files were used to extracted wellcodes and UMIs based on the wellcode-UMI-Poly T pattern (16-nt wellcode, 10-nt UMI, 16-nt poly T). RNNS algorithm was developed for searching wellcodes, which either totally match or 1-nt mismatch with the references. Since all of the designed wellcodes are more than 2-Hamming distance away from each other, 1-Hamming distance difference between the designed and the measured wellcode is allowed to correct the error caused by PCR or sequencing. PCR duplicates were identified when the same wellcode, UMI and gene ID (ENSEMBL human GRCh38, mouse GRCm38) were found. The gene abundances of the samples were estimated using the feature Counts (Version 1.5.0-p1) (35). The normalized counts of genes as well as differentially expressed genes were calculated using the DESeq2 package (36).

ERCC experiment

To assess the quantitative capability of Grouped-seq, we spotted a 192-droplet array of ERCC mRNA spike-ins (Ambion, Life Technologies) on a piece of glass slide and each droplet has a volume of 200 nl. After Grouped-seq, the sequencing data was aligned to ERCC sequence list, followed by UMI merging to obtain the measured expression. Finally, we calculated the ERCC capture efficiency with the measured expression and input ERCC molecules.

Mixed-species experiment

In the mixed-species experiment, human (A549) and mouse (3T3) cells were manually pipetted into the microwells of a single SMARchip in an alternate line format with a density of 300 cells per microwell. Cells were cultured for three days and imaged, followed by Grouped-seq. The capture of mRNA was performed on the chip, and then all the beads were pooled together for sequencing.

qPCR assays

To optimize the quantities of the magnetic beads and the capture oligos for mRNA capture, about 300 Jurkat cells were loaded into each microwell, followed by cell lysis and mRNA capture using the magnetic beads. After collection of the beads and reverse transcription, 2 μl of beads solution was used for each qPCR assay. The beads solution was mixed with 12.5 μl of Power Up SYBR Green Master Mix (Applied Biosystems), 1 μl of forward primer (10 μM), 1 μl of reverse primer (10 μM), and 8.5 μl of nuclease-free water. Quantitative real-time PCR was performed in three replicates on the ABI-7900HT Real-Time PCR System (Applied Biosystems). The qPCR primers for the target genes (P7/TSO, MTND4L, ACTB, PPP1R26, GPR17, Supplementary Table S8) were designed with Primer3 software (RRID:SCR_003139).

Comparison of Grouped-seq to conventional RNA-Seq

In Grouped-seq, 300 A549 cells were allocated into each microwell and a total of 192 libraries were constructed using the pooled procedure described above. Meantime, 10^5 of A549 cells from the same batch were collected to extract total RNA using a RNeasy column (Qiagen) and a sequencing library was prepared from the total RNA using the Illumina TruSeq poly(A) + RNA-Seq library construction according to the manufacturer's instruction.

Comparison of gene expression profiles between cells cultured in 2D and 3D

To demonstrate the feasibility of Grouped-seq for analyzing 3D cultured cells, we performed Grouped-seq from 2D- and 3D-cultured A549 cells in parallel. DE (differential expression) analysis was performed using DESeq2 package after library normalization (36). DEGs were detected with a cutoff of $|Log_2FC| \geq 2$ and FDR ≤ 0.01. Finally, the pathway enrichment analysis was performed using clusterProfiler package (37). The enriched pathways were ordered by their adjusted p-values.

Drug responses of CRC organoids analyzed on the SMARchip by Grouped-seq

After the establishment of CRC organoids, Grouped-seq was performed to decipher MoAs of different drugs. Four common anti-CRC drugs across eight concentrations with five repeats were applied to the CRC organoids. The concentration gradients were 0.256, 1.28, 6.4, 32, 160, 800,

4000 and 20000 nM in irinotecan and in oxaliplatin combined with fluorouracil; 2.048, 10.24, 51.2, 256, 1280, 6400, 32000 and 160 000 nM in fluorouracil, and 0.16384, 0.4096, 1.024, 2.56, 6.4, 16, 40 and 100 ng/ml in cetuximab.

Eight concentrations in the gradient of each drug were divided into the low concentration and the high concentration groups. The gene expression data were log-normalized for principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE). In the tSNE method, cell clustering was performed based on the shared nearest neighbor (SNN) method and the top 10 highly variable genes (HVGs) were sorted out for each cluster.

Phenotype-assisted pathway enrichment analysis (PAPEA)

The phenotype-assisted pathway enrichment analysis consists of two rounds of analyses of the DEGs with the aid of the cell viability analysis. In the first round, only the genes that changed synchronously with the phenotypes of the tumor organoids (named as phenotypic change-related genes, PCR genes). To find out the PCR genes, the benchmark dose (BMD), at which the organoid viability starts to change dramatically, is first determined for the drug using the DRC. Then, the concentration gradient is divided into the low and the high concentration groups (LCG and HCG) at the BMD (BMD is in the LCG). By employing criteria of log fold change (LFC) >0.5 and P.adjust < 0.05, PCR genes are determined by comparing the gene expression levels between the LCG and the HCG using the unpaired t-test. Then, the MoA-related pathways (MoA pathways) can be identified based on the PCR gene set using the GO analysis (37). The enriched pathways with criteria of P.adjust <0.05 and the involved PCR gene number ≥ 3 are sorted out and top 6 pathways with the most minimal p.adjust are recognized as MoA pathways for the following analysis.

In the second round of analysis, the genes involved in the top MoA pathways are re-examined with loose criteria (LFC > 0.35 and P.adjust < 0.05) and the top genes are selected as the MoA pathway-related genes (MPR genes). Since these genes may not change synchronously with the phenotype, we repeatedly divide the concentration gradient into two groups at each concentration to calculate the corresponding LFCs. The BMD of each gene is determined by selecting the concentration, at which the LFC of this gene is maximum. The MPR genes are then sorted into three phases (low, middle, and high) according to their BMDs. After that, based on the percentages of the MPR genes in each phase, the order of the MoA pathways was determined. A possible scenario of the drug action could be deduced based on the functions of the MPR genes, the MoA pathways, and the additional sequential information.

RESULTS

Grouped-seq performed on superhydrophobic microwell array chip

We developed a high-throughput screening platform termed as Grouped-seq (Genome-wide RNA output unified with phenotypic data) for combined imaging- and sequencing-based analysis of tumor organoids with limited quantities on a superhydrophobic microwell array chip (SMARchip) (38). Tumor organoids can be derived from surgically resected tumor tissues using a mechanical processing method developed by our group (10). After a short period of expansion in culture plates, tumor organoids can be loaded into a SMARchip with a density of 10 organoids per microwell for drug treatments followed by integrated phenotypic and transcriptomic analysis (Figure 1A). The nanoliter-scale SMARchip consists of a layer of synthesized superhydrophobic polymers (poly(butyl methacrylate-co-ethylene dimethacrylate), BMA-EDMA) on the top of a PDMS microwell array. When the excess solution is aspirated out, a droplet array in the microwells can spontaneously form due to this superhydrophobic layer (Figure 1B). To further enhance the resistance of the SMARchip to liquids containing surfactants, such as cell lysis buffer, hydrophobic oil, FC-40, can be infused into the superhydrophobic polymer networks (Supplementary Figure S1A) (39). This oil-enhanced SMARchip can withstand the lysis buffer for a longer period of time, facilitating the cell lysis in the microwells (Supplementary Figure S1B). The operation of the SMARchip is flexible: cells can be loaded into the microwells by either random seeding or droplet rolling (Supplementary Figure S2A), and cultured in either the immersion mode or the droplet mode (Supplementary Figure S2B). In addition, reagents can be loaded in the microwells either as a whole by the immerse-aspirate method or individually by the spot-cover method (Supplementary Figure S2C). Any complicated liquid and cell handling, such as cell culture, imaging, and RNA-seq, can be realized in a high-throughput manner by combing the above operations on the SMARchip.

The principle of the Grouped-seq is outlined as follows (Figure 1C): cells or organoids are first seeded into the microwells of a SMARchip (cell array) by either random seeding or droplet rolling. After cell culture, a drug array, on which multiple drug gradients are spotted using a robotic spotter, is covered onto the cell array upside down to deliver drugs to the cells, which has been validated no cross contamination in our previous studies (38,40). Following the drug treatment, the cells in the microwells can be imaged to obtain the phenotypic data. Next, an oil-enhanced SMARchip (bead array), on which each microwell contains magnetic beads conjugated with barcoded poly-(dT) capture oligos in the lysis buffer, is aligned to the cell array. The beads fall down to the cells and the lysis buffer diffuses to lyse the cells, leading to the capture of the exposed mRNA onto the beads. After the collection of all the beads to a tube, cDNA synthesis, amplification, and library construction are performed in the tube for sequencing. The use of the magnetic beads facilitates the easy recovery of mRNA from the microwells as well as the washes performed in the tube.

Design and characterization of a low-cost, high-capacity barcoding system

Since we need to know the barcode in each microwell to enable the combined analyses of the phenotypic and the transcriptomic data, we developed a ligation-based barcode

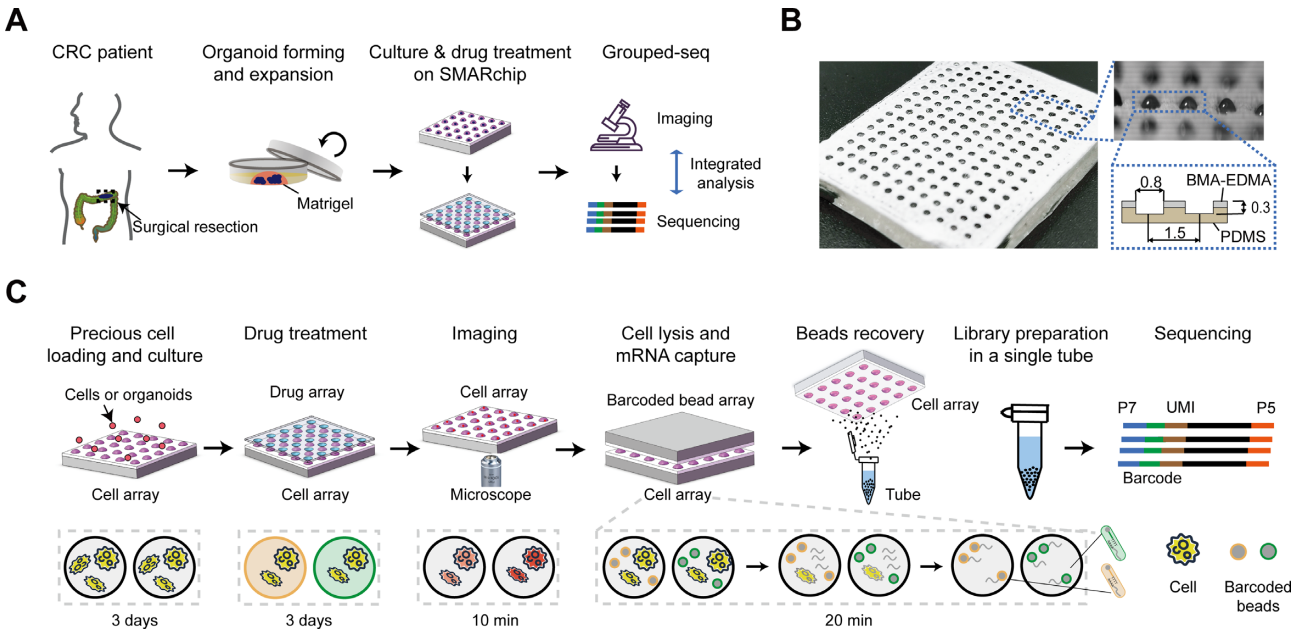

Figure 1. Design and operation of SMARchip for Grouped-seq. (**A**) Workflow of Grouped-seq for in-depth analysis of tumor organoids treated with drugs. After the forming of from surgically resected tumor tissues and the expansion in a culture plate, the CRC organoids are loaded and cultured on a superhydrophobic microwell array chip (SMARchip), followed by drug treatments and Grouped-seq. (**B**) Photograph of a SMARchip with an array of 192 microwells and the cross-section view of the chip. A droplet array can be formed spontaneously due to the superhydrophobic layer. The microwells of the 192-well SMARchip have dimensions of 800 μm in diameter, 200 μm in depth, and 700 μm in pitch, and the BMA-EDMA layer is about 100 μm thick. (**C**) Diagram illustrating the operation procedure of the Grouped-seq performed on the SMARchip to obtain the phenotypic and the transcriptomic data in a single run. Top panel: macro-views of the on-SMARchip operations and the in-tube library preparation and sequencing. Bottom panel: micro-views of on-SMARchip operations including cell culture for 3 days, drug treatment for 3 days, cell imaging for 10 min., and loading of barcoded beads, cell lysis, and mRNA capture for 20 min. Grouped-seq performs integrated phenotypic and transcriptomic screening within a week.

synthesis method to expand the barcoding capacity and to lower the costs of the oligo synthesis for Grouped-seq. The capture oligo that links to the streptavidin-coated Dynabeads™ consists of two parts: part1 (P1, 55 bases), which contains the 5′ biotin-modified P7 adapter, a 6-base sample index (Supplementary Supplementary Table S3), and the first half of the Read2 sequences, and part2 (P2, 59 bases), which is made up of the rest half of the Read2, a 16-base wellcode, an 8-base Unique Molecular Index (UMI), and a Poly-(dT) sequence (Figure 2A, and Supplementary Table S4). We employed T4 DNA ligase together with a 19-base linker to link these two parts together, forming the final capture probe with a total length of 114 bases. Since P1 and P2 contain a 6-base sample index and a 16-base wellcode, respectively, M × N types of full-length capture oligos with a 22-base sample code can be formed by using M types of P1 and N types of P2 with the cost of synthesizing M + N short oligos instead of M × N 114-base-long oligos (Figure 2C). With this capture oligo, mRNA can be captured onto the beads and reverse transcribed to the first-strand cDNA followed by template switching (41). Then, cDNA is amplified with a biotinylated P7 primer and a template switch oligo (TSO). After the ultrasonic fragmentation and the modified adaptor ligation, 3′-ends of mRNAs are sequenced to identify barcodes, UMIs, and transcript information (Figure 2B).

The assembly of the magnetic beads with barcoded capture probes is performed on an oil-enhanced SMARchip (bead array), on which beads are first loaded (Supplementary Figure S3A, and Supplementary Movie S1). The ligated capture oligos are then loaded into the microwells of another SMARchip (oligo array) using a nanoliter centrifugal liquid dispenser (NanoCLD) developed by our group (Supplementary Figure S3B) (40). The bead and the oligo arrays are sandwiched together, so that the capture oligos and the beads are linked together via the biotin-streptavidin bonding. After washes with phosphate-buffered saline (PBS), the excess unlinked oligos are removed while the beads are retained in the microwells owing to the sedimentation. Finally, lysis buffer is spotted onto a glass slide and then transferred to the bead array.

We first verified that the full-length capture oligo can be synthesized by this ligation method using Sanger sequencing and flow cytometry (Supplementary Figure S4, and Supplementary Table S5). Next, to achieve the best mRNA capture efficiency, we evaluated two mixing ratios of P1/P2, 1/3 and 3/1, which represent either P1 or P2 is used up during the ligation, respectively. We also tested whether it is necessary to remove excess oligos less than 100 bases after the ligation using the QIAquick PCR Purification Kit (Qiagen). The quantitation of captured mRNA demonstrated that the P1/P2 ratio of 3/1 provided the best capture efficiencies even without the purification step (Figure 2D, and Supplementary Data S1). RNA-seq using 192 synthesized capture oligos also proved that the 3/1 ratio without purification generated much more uniform gene numbers across

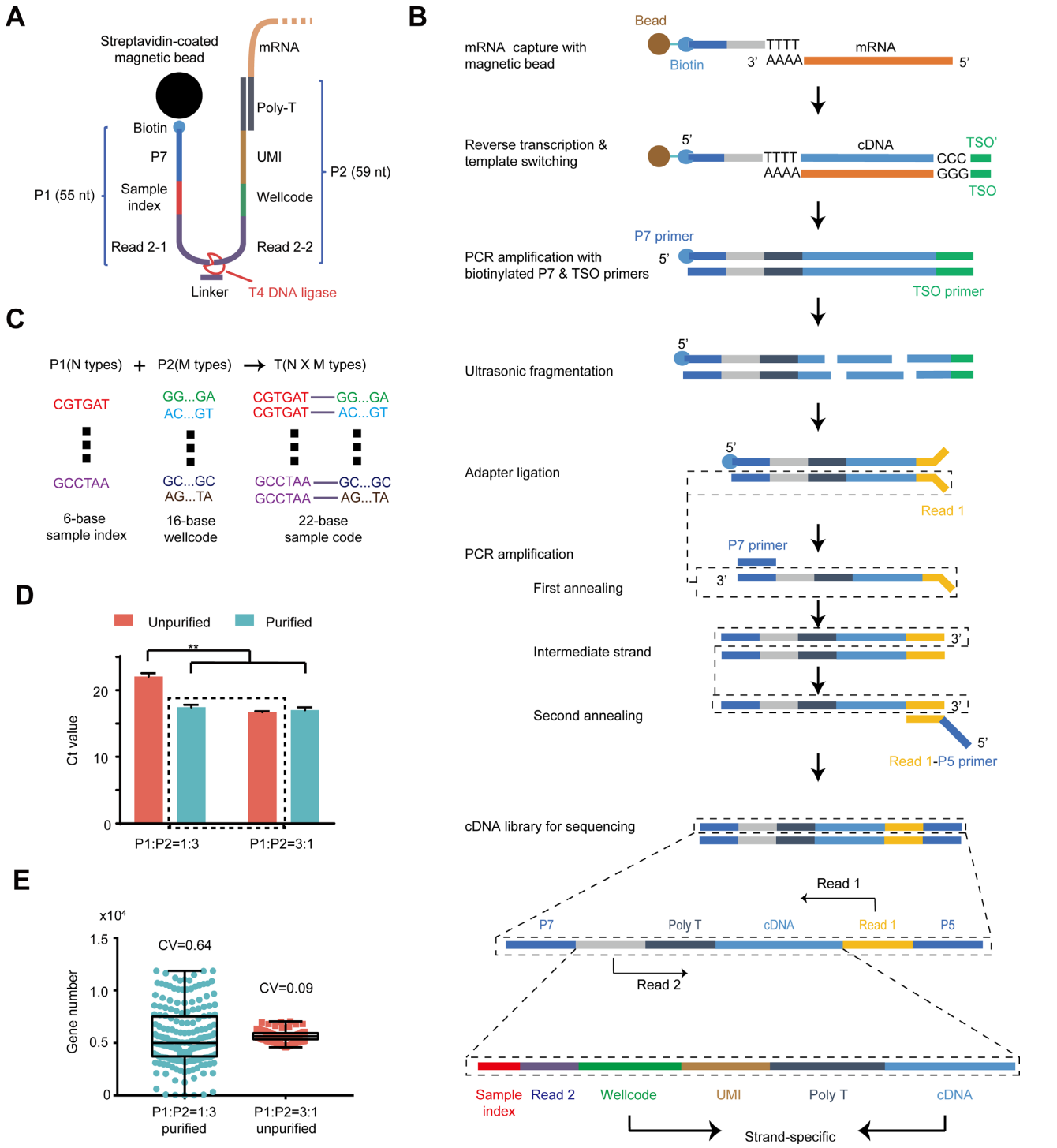

Figure 2. Design and characterization of the ligation-based barcoding system of the Grouped-seq. (**A**) Schematic of the design of the barcoded capture oligo with two parts that are ligated together with a linker. The assembled 114-base-long capture oligo contains P7 adapter, a 6-base sample index, the Read2 sequence, a 16-base wellcode, an 8-base UMI, and a Poly-(dT). (**B**) Diagram illustrating the Grouped-seq protocol. mRNA was first captured by the capture oligos linked to magnetic beads via the streptavidin-biotin binding. Then, the mRNA was reverse transcribed to cDNA using the template switch method, followed by PCR amplification with P7 and TSO primers. The amplicons were purified, and then fragmented to ~300 bp using ultrasound, followed by adapter ligation. Due to the biotinylated modification of the P7 primer, no adapter was added to the P7 ends. Finally, another round of amplification was performed to obtain the cDNA library for sequencing. The Read1 and Read2 primers were used to sequence the cDNA fragment and the wellcode/UMI, respectively. (**C**) N × M types of full-length capture oligos generated by the ligation-based barcode synthetic method. N types of P1 and M types of P2 are linked using T4 ligases to generate N × M types of full-length oligos. P1 contains 6-base sample index with a barcoding capacity of $4^6 = 4096$ (N < 4096) and P2 contains 16-base wellcode with a barcoding capacity of $4^{16} = 4294967296$ (M < 4294967296). Such a combined barcoding capacity allows us to choose the best oligo sequences and benefits the decoding efficiency. (**D**) Bar graph comparing the mRNA capture efficiencies (Ct value of the quantitative PCR) using the capture oligos synthesized from two mixing ratios of P1/P2 combined with or without purification ($n = 3$ independent repeats, $P < 0.05$, Student's *t*-test). (**E**) Comparison of gene number distributions across 192 synthesized capture oligos using different methods, proving the P1/P2 ratio of 3/1 without purification generated more uniform gene numbers with a CV of 9%.

all the microwells with a coefficient of variation (CV) of 9% (Figure 2E). We optimized the amounts of the beads and the capture oligos in the microwells for mRNA capture on the SMARchip. Quantitative PCR analyses of four genes (MTND4L, ACTB, PPP1R26, and GPR17) representing the abundancies from high to low levels demonstrated that the capture condition of 5 mg/ml beads (Figure 3A, and Supplementary Data S1) and 0.1 μM oligos (Figure 3B, and Supplementary Data S1) provides the best efficiency. We also optimized the PCR cycle numbers of the library amplification, showing 5 cycles should be sufficient to generate a library with the highest gene number and mapping percentage (Supplementary Figure S5A, and Supplementary Data S1). Since only 192 microwells were barcoded with 22 bases, the relative nearest neighborhood searching (RNNS) method can be used to extract the wellcode information, providing a wellcode extraction efficiency of 93.8%, which was 8% higher than that of the totally matched searching (TMS) (Supplementary Figure S5B, and Supplementary Data S2).

Performance characterization of the Grouped-seq technology

To characterize the performance of the Grouped-seq, we first conducted a 192-well RNA-seq using ∼300 A549 cells per microwell on the SMARchip. On average, we detected ∼11 570 genes per microwell with a CV of 4.0% from ∼ 903 012 mapped reads per microwell with a CV of 22.7% at a sequencing depth of 1 M reads per microwell (Figure 4A, and Supplementary Data S3). The correlation coefficients of the gene expression profiles between any two microwells are all above 0.95 (Figure 4B) and the top 15 genes from all the microwells are also highly consistent (Figure 4C), illustrating the excellent sequencing uniformity of the Grouped-seq on the SMARchip. We did find the uniformity of the gene numbers across the entire array is going down with a CV from 4.0% to 11% while the sequencing depth is decreased from 1 M to 0.025 M reads per microwell (Supplementary Figure S6). Considering the 0.025-M sequencing depth just corresponds to ∼1.5 G of data for 192 samples, such a variation is acceptable.

To assess the quantification capability, we loaded the External RNA Controls Consortium (ERCC) mRNA spike-ins, a set of 92 synthetic mRNA molecules covering a broad range of concentrations, to the microwells and performed the Grouped-seq at a sequencing depth of 0.2 M reads/microwell. We first combined the data from all the 192 microwells together and compared the measured mean abundance with the number of input molecules, yielding a strong correlation (Pearson coefficient $\underline{r} = 0.97$) between the measured and the predicted molecules, though low-quantity ERCC transcripts showed increased noise levels (Figure 4D, and Supplementary Data S4). We then calculated the correlation coefficient in each individual microwell, showing most of the coefficients are around 0.94. Only two of the microwells are in the range of 0.80–0.90 (Supplementary Figure S7A).

To verify no cross-contamination among microwells during the sample processing, we conducted a mixed-species experiment with human (A549) and mouse (3T3) cells, which were loaded into the microwell array in an alternate line format. After the capture of mRNA on the chip, all the beads were pooled together for the following reverse transcription (RT) reaction as well as the sequencing. We found samples were perfectly separated by species, suggesting no cross-contamination during sample processing (Figure 4E, and Supplementary Data S5). We compared the RNA-seq data obtained from the SMARchip and the off-chip TruSeq (Illumina), showing a good agreement with a correlation of 0.93 (Figure 4F). To test the reproducibility of our platform, three repeats of the Grouped-seq were performed independently using about 300 A549 cells per microwell at a sequencing depth of 0.05 M/microwell, generating average gene numbers of 5639 ± 555, 5646 ± 555 and 5901 ± 554, respectively (Supplementary Figure S7B, and Supplementary Data S6). The correlation coefficients between two repeats are 0.98, illustrating the high reproducibility of the Grouped-seq platform (Supplementary Figure S7C).

To explore the effect of the sequencing depth to Grouped-seq, we sequenced ∼300 A549 cells per microwell at the sequencing depth from 0.025, 0.05, 0.1, 0.5, to 1 M reads/microwell. While the mapped read numbers increased proportionally with the sequencing depth, the gene numbers became gradually saturated and over 10 000 genes could be detected from all the microwells when the read number exceeded 500 000 (Figure 4G, and Supplementary Data S3), which corresponds to about 28.8 G of data for 192 microwells in the paired-end sequencing. We compared the gene expressions obtained at the lowest and the highest sequencing depths (0.025 and 1 M/microwell), yielding correlation coefficients of 0.78, 0.94, 0.99 and 1 for the genes within the expression ranges of $0 <$ Counts per million (CPM) < 1, $1 <$ CPM < 10, $10 <$ CPM < 100 and $100 <$ CPM, respectively (Supplementary Figure S7D). Unique Molecular Index (UMI) is employed in our system for monitoring the potential PCR amplification redundancy. Our Grouped-seq platform demonstrated that the UMI correction rate, which is defined as read number/UMI number, was increased from 1.22 ± 1.3% (mean ± CV), 1.24 ± 1.7%, 1.28 ± 2.2%, 1.51 ± 5.1%, to 1.88 ± 8.5% with the increase of the sequencing depth (Figure 4H, and Supplementary Data S3). Such a low rate is due to the small PCR cycle number during the library preparation. Finally, we evaluated the limit of detection (LOD) of the gene expression level that can be detected at the different sequencing depths. Since the exact amounts of mRNA in the cells were unknown, we used the CPM of the genes obtained at the 1-M sequencing depth as the standards and the LOD of a sequencing depth is defined as the upper limit of the CPMs (obtained at 1-M sequencing depth) of the genes that can be detected at least once with this sequencing depth. The LODs was determined to be 0.01, 0.32, 0.40, 1.00 and 1.58 CPM, for the sequencing depths of 1, 0.5, 0.1, 0.05 and 0.025 M/microwell, respectively (Figure 4I). This calculation is important because it tells what the minimum sequencing depth is in order to measure a gene with a certain expression level.

Comparison of gene expression profiles between cells cultured in 2D and 3D

Since three-dimensional (3D) culture is essential for tumor organoids, we developed a protocol for perform-

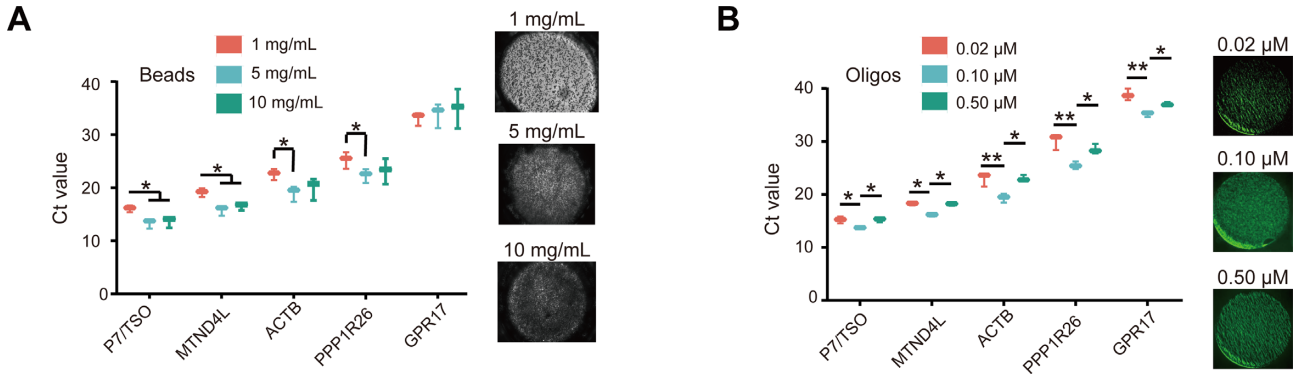

Figure 3. Optimization of concentration of beads and capture oligos used for Grouped-seq. (**A**, **B**) Optimization of amounts of beads and capture oligos in the microwells for mRNA capture. Quantitative PCR analyses demonstrated that the capture with 5 mg/ml beads and 0.1 μM oligos has the best efficiency ($n = 3$ independent repeats, ** $P < 0.01$, * $P < 0.05$, Student's t-test).

ing Grouped-seq from cells cultured in Matrigel on the SMARchip. First, A549 cells were cultured in Matrigel on the SMARchip for two days. Prior to sequencing, the harvest solution was delivered to the microwells for Matrigel digestion. After that, the microwell array was centrifugated to sediment the cells to the bottom of the microwells, followed by the Grouped-seq process (Figure 5A). As a comparison, A549 cells were also cultured in the microwells in the conventional 2D monolayer. Both the 2D and 3D samples were processed in parallel and sequenced at a depth of 0.1 M reads per microwells. We compared the statistical results of the RNA-seq data between the 2D and the 3D cultures and found almost no differences in the levels of the read mapping rate (∼92%) and the gene number (∼7400 genes) (Figure 5B, and Supplementary Data S7). Although the exon mapping percentage and the mapped mitochondrial RNA (mtRNA) demonstrated certain differences between these culture methods, the changes are still less than 10%. We also demonstrated that the correlation coefficients between any two microwells are higher than 0.95 in the 3D cell culture. Meanwhile, the coefficients between the 2D and the 3D microwells are all in the narrow range of 0.82–0.95 (Figure 5, C and D), due probably to the short culture time on the SMARchip. Nevertheless, this high similarity proved that the quality of the Grouped-seq was not affected by Matrigel during the sample processing.

Next, we used DESeq2 to detect differentially expressed genes (DEGs) between the 2D and the 3D cultures using the criteria of q-value < 0.01and |log$_2$Fold change| > 2. Only 0.5% (61 of 11395, including 30 up-regulated and 31 down-regulated) of genes changed significantly while the rest (11 334 of 11 395) did not. These DEGs include the cell proliferation genes, such as MMP7, PPP1R15, ADM, JUN, ALDH3A1 and OGT, and the tumor development genes, such as BCL2L1, CCNB1, CCNB2, CDK1, FAS, SESN2, SHISA5 and TP53 (Figure 5E, and Supplementary Data S8). Furthermore, the GO analysis indicates these DEGs are associated with cell proliferation, cell cycles, and tumor development, as other previous studies have proven (Figure 5F, and Supplementary Data S9) (42–44).

Drug responses of CRC organoids analyzed on the SMARchip by Grouped-seq

Previously, we have demonstrated that lung cancer organoids can be employed to predict patients' drug responses within one week using our SMARchip system, on which tumor organoids were cultured and characterized using the cell viability testing (10). To analyze the responses of PDOs to anti-cancer drugs in a more comprehensive way, we apply the Grouped-seq technology to the PDO-based drug tests in the current study. A CRC organoid line was first established following the protocol developed by our group (Supplementary Figure S8) (10). Then, the organoids mixed in the Matrigel were cultured on the SMARchip in a density of ∼10 organoids per microwell for three days. Four common anti-CRC drugs including irinotecan (I), fluorouracil (F), oxaliplatin combined with fluorouracil (OF), and cetuximab (C), each of which has eight different concentrations with five repeats (Figure 6A, and Supplementary Data S10), were delivered to the organoids for two-day treatment. To eliminate the variation caused by the uneven numbers and the sizes of organoids, the viability of the CRC organoids was measured both before and after the drug treatment using the alamarBlue™ kit and the relative cell viability is represented by the ratio of these two measurements. Finally, the Grouped-seq was performed using these organoids on the SMARchip (Figure 6B).

We first plotted the dose response curves (DRCs) of the CRC organoids to evaluate the drug effects (Figure 6C, and Supplementary Data S11). Based on the phenotypic analysis, the transcriptomic data from the Grouped-seq were separated into nine groups: four drugs combined with the low and the high concentration ranges, and the negative control, for the subsequent PCA (Principal Components and Clustering Analysis) and the t-SNE (t-Distributed Stochastic Neighbor Embedding) (Figure 6D and E). Interestingly, we found the responses of the CRC organoids to these four drugs are totally distinct. First, cetuximab, an epidermal growth factor receptor (EGFR) inhibitor, had no effect to the CRC organoids which harbor no EGFR mutations (Figure 6C, and Supplementary Figure S8D). Since cetuximab is a monoclonal antibody, the gene expres-

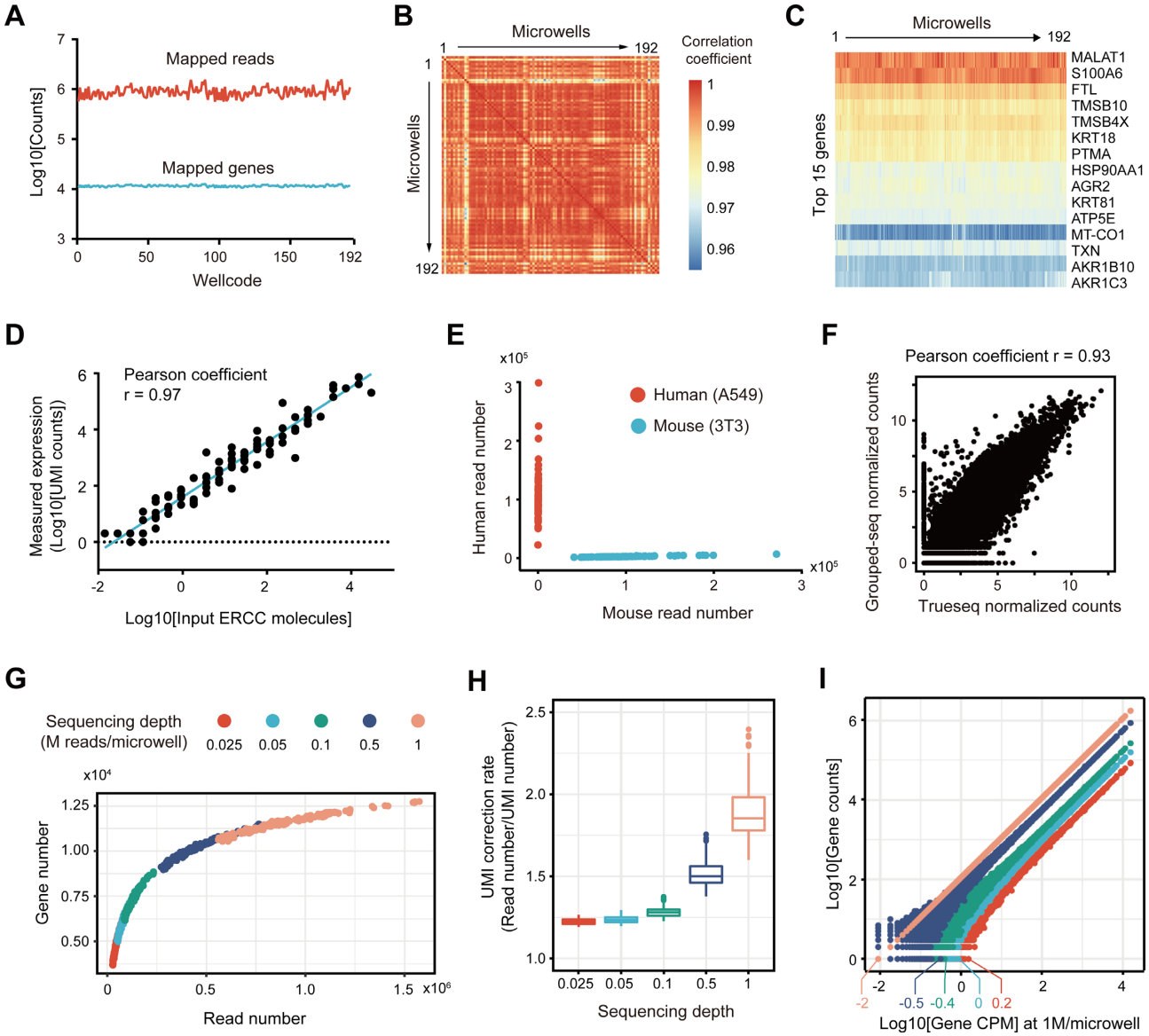

Figure 4. Performance characterization of the Grouped-seq platform. (**A**) Mapped read and gene numbers across 192 microwells of the SMARchip. (**B**) Heatmap of correlation coefficients between any two of microwells. (**C**) Top 15 genes with the highest expression levels in all the microwells. (**D**) Quantitative capability of the Grouped-seq evaluated by ERCC sample. (**E**) Mixed-species experiment with human (A549) and mouse (3T3) cells performed on a single SMARchip. (**F**) High correlation between the Grouped-seq and the off-chip TruSeq. (**G**) Read and gene numbers of 192 microwells obtained at different sequencing depth from 0.025, 0.05, 0.1, 0.5, to 1 M reads per microwell. (**H**) UMI correction rates at different sequencing depth. UMI correction rate is defined as the ratio of read number to UMI number. (**I**) Limit of detection (LOD) of the gene expression level that can be detected at the different sequencing depths. The counts per million (CPMs) of the genes obtained at the 1-M sequencing depth are used as the standards and the LOD of a sequencing depth is defined as the upper limit of the CPMs (obtained at 1-M sequencing depth) of the genes that can be detected at least once with this sequencing depth.

sions of the organoids had no changes and grouped with the NC in the entire dose range (Figure 6D and E). Second, fluorouracil is a widely used chemotherapy medication to treat cancer by blocking the production of DNA. Our CRC organoids demonstrated a strong resistance to fluorouracil in the entire dose range (Figure 6C). However, the gene expression profiles of both the low and the high concentration ranges kept a distance from the NC group, showing the organoids adjusted the gene expressions to eliminate the effect of the drug (Figure 6, D and E). The DEGs mainly focused on DNA/RNA/protein binding and regulation, including LMO7, SLPI, YBX1, TPM3, FAM111B, SLIRP and RANBP1 (Figure 6F). Third, the combined use of oxaliplatin and fluorouracil did induce the death of tumor organoids with an IC_{50} of 231.6 nM (Figure 6C). Like the fluorouracil group, the gene expressions of the organoids treated with OF were clustered together and the DEGs are related to cell death and autophagy (AKR1B10, AKR1C2 and AKR1C3) and signal transduction (wnt, MAPK, and PIK-Akt) (Figure 6D–F). Finally, irinotecan works by blocking type I topoisomerase which results in DNA damage and death of cancer cells. In the low dose range, the CRC organoids resisted the treatment of irinotecan and the gene expression profiles closely as-

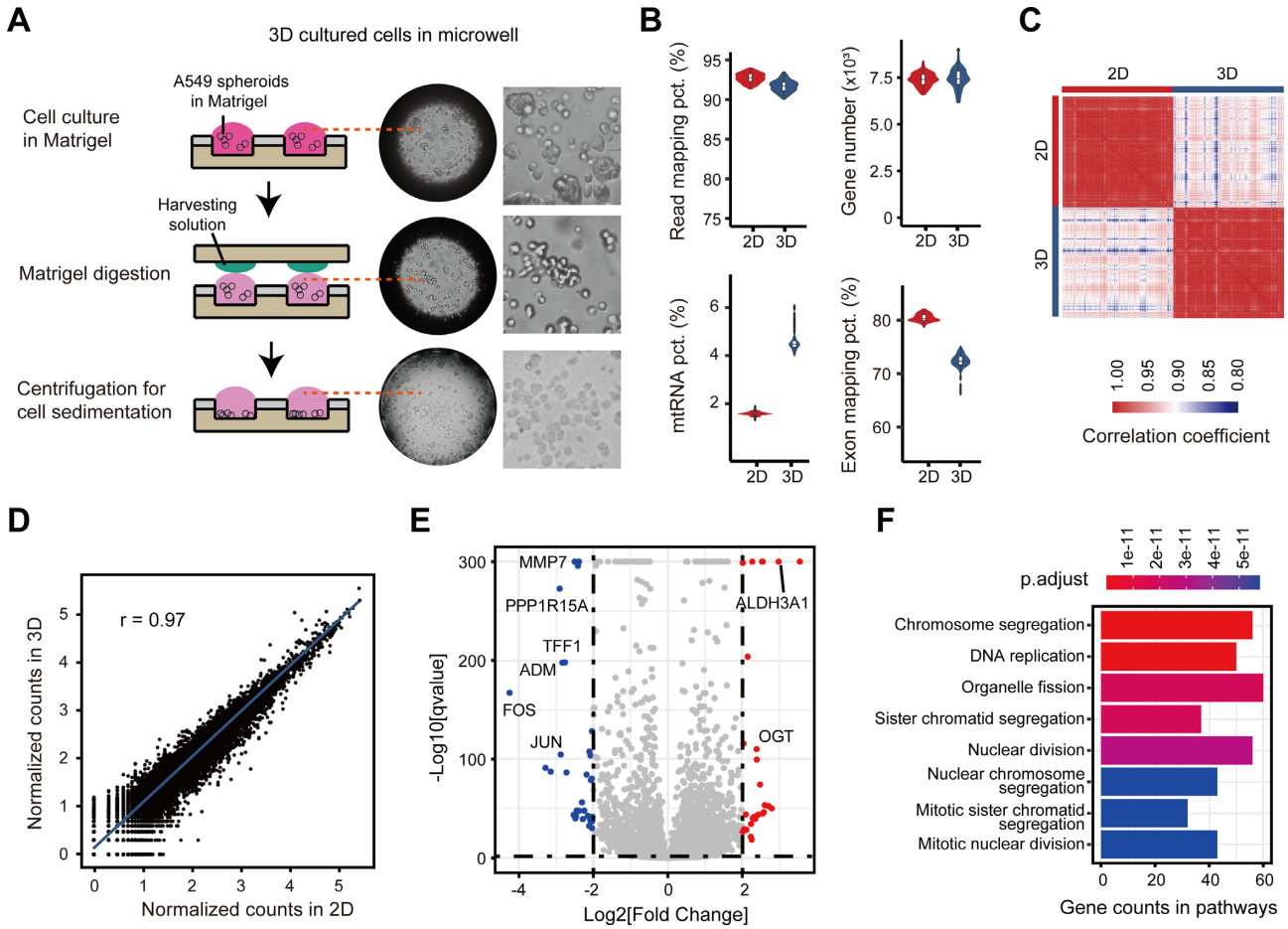

Figure 5. Three-dimensional cell culture followed by Grouped-seq on the SMARchip. (**A**) Diagram illustrating the procedure of performing Grouped-seq from cells cultured in Matrigel on the SMARchip. Photographs show A549 cells cultured in Matrigel (top), in digested Matrigel (middle), and on the bottom of the microwells (bottom). (**B**) Comparisons of sequencing data obtained from 2D and 3D cultured cells on read mapping percentages, gene numbers, mitochondrial RNA percentages, and exon mapping percentages. (**C**) Heatmap of the correlation coefficients. The correlations are high among microwells in the same culture modes, but relatively low between the 2D and 3D modes. (**D**) Collective correlation coefficients between the 2D and the 3D cultured cells. (**E**) Differential gene expression analysis between 2D and 3D cultures. Only 30 up-regulated and 31 down-regulated genes show q-value <0.01 and $|\log_2 \text{Fold change}| > 2$. (**F**) GO pathway enrichment analysis of differentially expressed genes of the 3D culture.

sociated with those of the NC group, showing irinotecan has no effect to the cells (Figure 6C–E). However, when irinotecan increased to the high dose range, the viability of tumor organoids decreased significantly with an IC_{50} of 68.9 nM (Figure 6C). Likewise, the gene expressions of the organoids moved away from the NC with several DNA conformation-related genes that were changed dramatically, including TOP2A, HMGN2 and HMGN1 (Figure 6F). Overall, we can clearly differentiate the drug responses of the CRC organoids to different drugs and the phenotypic and the transcriptomic data are highly correlated with each other.

Phenotype-assisted pathway enrichment analysis (PAPEA) for Grouped-seq

To gain mechanistic insights into the drug responses of the CRC organoids, we developed a phenotype-assisted pathway enrichment analysis (PAPEA) method coupled with the Grouped-seq platform. Unlike the conventional pathway enrichment analysis (PEA), in which the pathways are directly drawn from all the differentially expressed genes, this new strategy consists of two rounds of analyses of the DEGs with the aid of the cell viability analysis (Figure 7A, and Supplementary Data S12): in the first round, only the genes that changed synchronously with the phenotypes of the tumor organoids (named as phenotypic change-related genes, PCR genes) are considered in order to eliminate interferences from any accidentally changed genes, the expression changes of which are not strongly correlated with the phenotypic data. The PCR genes and the DRC have the same benchmark dose (BMD) (45), at which the gene expression level and the viability of the organoids start to significantly change, respectively. Then, the MoA-related pathways (MoA pathways) can be identified based on the PCR genes using the GO analysis. In the second round of analysis, the genes involved in the top MoA pathways are examined again and the genes that have significant expression changes at different BMDs (MoA pathway-related genes, MPR genes) are selected and sorted into three phases (low, middle, and high) according to their BMDs. A scenario of the drug action could be deduced based on the functions of the MPR genes and their changing sequence

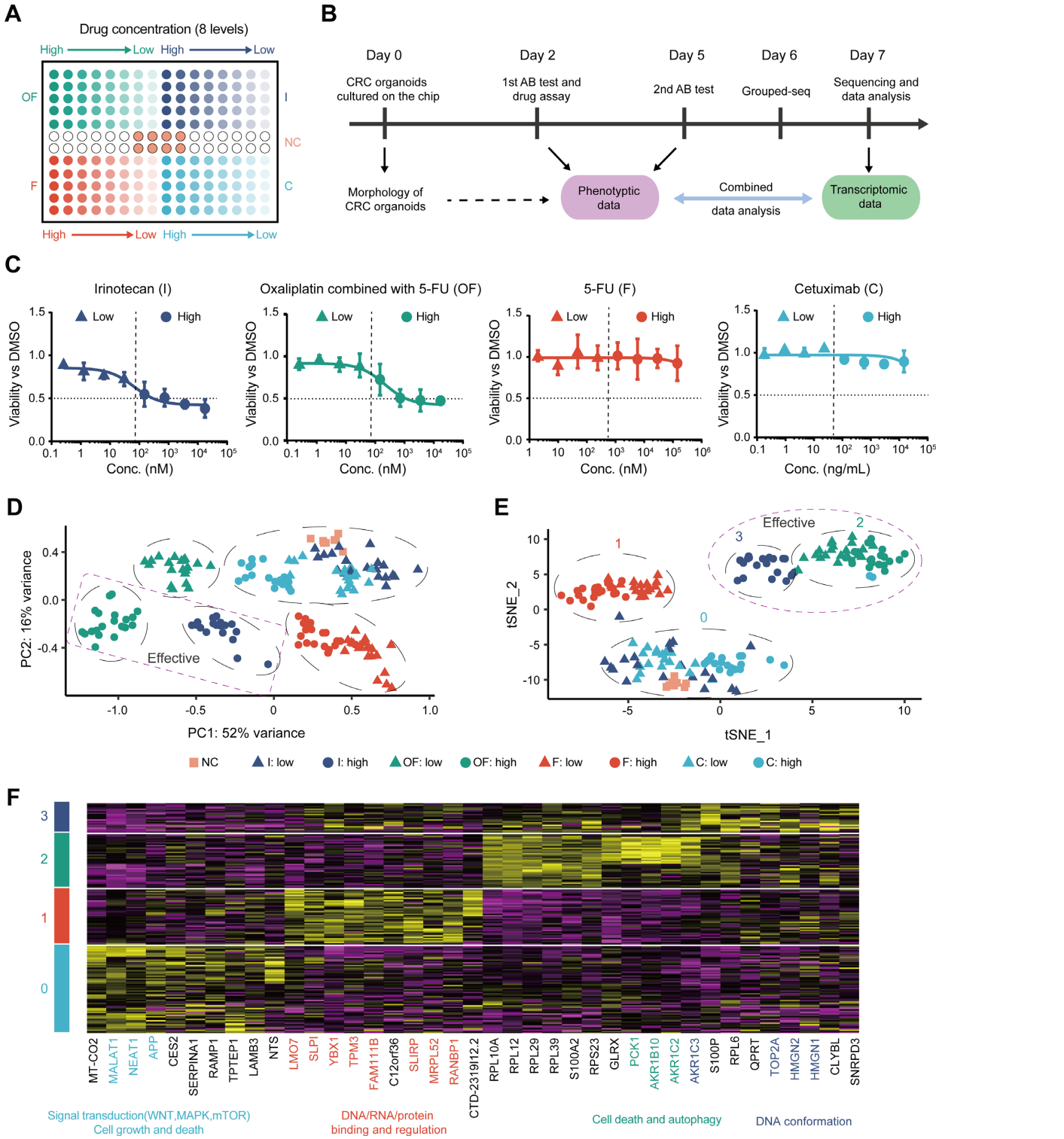

Figure 6. Drug responses of CRC organoids analyzed on the SMARchip by Grouped-seq. (**A**) Schematic of the layout of the 192-microwell SMARchip, on which four different drugs with eight doses and five repeats as well as negative controls are tested. (**B**) Diagram illustrating the procedure of the culture of CRC organoids, drug treatment, and Grouped-seq performed on the SMARchip. (**C**) Dose–response curves of the organoid viabilities under the treatments of different drugs. (**D**, **E**) PCA and t-SNE plots of CRC organoids treated with different drugs. Different colors represent different drugs: pink (negative control, NC), purple (irinotecan, I), green (oxaliplatin combined with fluorouracil, OF), red (fluorouracil, F) and blue (cetuximab, C). Different shapes represent different drug concentrations: square (NC), triangle (low concentration range) and circle (high concentration range). (**F**) Identification of cell subpopulations in the t-SNE map. Top 10 marker genes for each subpopulation are listed. The marker genes in Group 0 (NC, I: low, C: low, and C: high) are associated with signal transduction (WNT, MAPK and mTOR) and cell growth and death. The marker genes in Group 1 (F: low and F: high) are with DNA/RNA/protein binding and regulation. The marker genes in Group 2 (OF: low and OF: high) are cell death and autophagy. The marker genes in Group 3 (I: high) are DNA conformation.

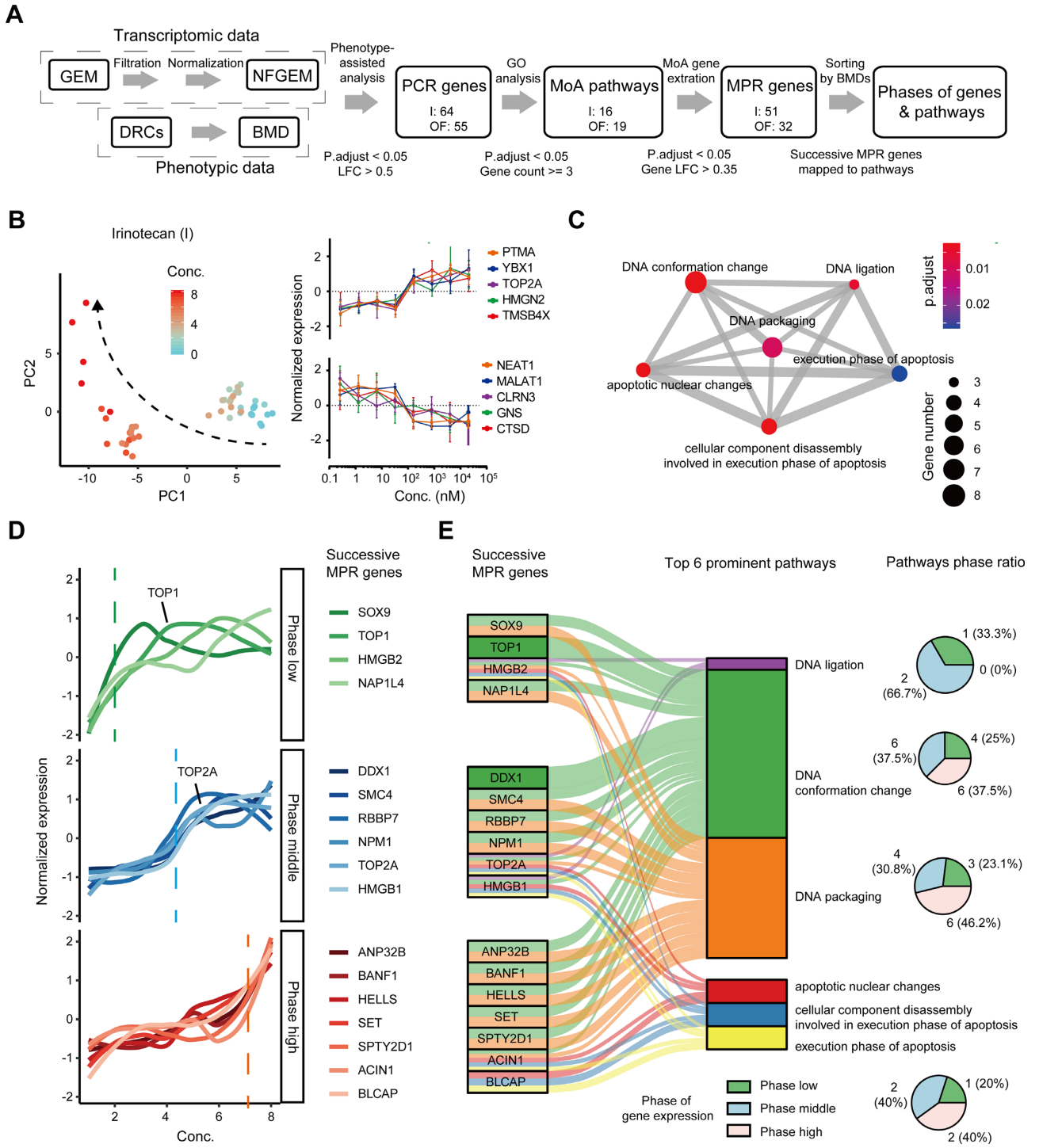

Figure 7. CRC organoids treated with irinotecan and analyzed using the PAPEA protocol. (**A**) Workflow of the phenotype-assisted pathway enrichment analysis (PAPEA). The detailed description can be seen in the Materials and Methods. (**B**) PCA plot of organoids treated with irinotecan at different concentrations (left), demonstrating a clear dependence of the gene expression on the drug concentration. The typical up- and down-regulated PCR genes show a strong correlation with the DRCs (right). (**C**) Network plot of top 6 MoA pathways, demonstrating these pathways are strongly correlated with each other. The colors represent the p.adjust while the circle sizes represent the enriched gene numbers. The widths of lines linked to two pathways represent the shared gene numbers. (**D**) Total 17 MPR genes extracted for irinotecan and grouped into the phase low, middle, and high according to their BMDs. TOP1 increased at the phase low while TOP2A was at the phase middle, showing a great agreement with the reported MoA of irinotecan. (**E**) Top MoA pathways sorted based on the phase percentages of the MPR genes. Irinotecan demonstrated that the pathways related to the DNA conformation changes are involved in all the phases, while the cell apoptosis-related pathways are only in the phase high.

Following the PAPEA protocol, we first analyzed the Grouped-seq data of the CRC organoids treated with irinotecan. The PCA analysis of irinotecan with NC clearly shows a gradually changing pattern of the gene expression with the drug concentration (Figure 7B). In the first round of analysis, the BMD was determined to be 160 nM for irinotecan from the drug response curve (Figure 6C). By employing criteria of log fold change (LFC) >0.5 and adjusted *P*-value (*P*.adjust) <0.05, 64 PCR genes were identified and enriched into 16 pathways. By contrast, when the conventional PEA was employed, 369 DEGs were identified with the same criteria as those of the PAPEA and 75 pathways, such as cellular protein complex disassembly, nuclear transport and RNA splicing, causing more difficulties in the deduction of the drug MoA. This stark contrast illustrates the necessity of the phenotype assistance. In the second round of analysis, we focused on the top 6 pathways, which are highly related to each other (Figure 7C, and Supplementary Data S13). A total of 51 genes are involved in these top pathways and 17 of them (MPR genes) have LFC >0.35 and *P*.adjust <0.05 at different BMDs (four in the low, six in the middle and seven in the high phase) (Figure 7D, and Supplementary Data S14). The results clearly show that the major pathways including DNA ligation, DNA conformation change, and DNA packaging are all related to DNA structures in the low dose range, while the pathways related to cell apoptosis play an important role in the high dose range (Figure 7E). When examining the specific genes, we found irinotecan first causes the expression increase of TOP1, which encodes type I topoisomerase, at the lower dose. With the increase of irinotecan concentration, the expression of TOP2A encoding type IIA topoisomerase increases as well. This finding is in line with the reported MoA of irinotecan, which mainly targets the enzyme topoisomerase I (TOP1) (46–48). Type IIA topoisomerase was also affected by irinotecan, due to the compensation of Type IIA topoisomerase to Type I topoisomerase (48,49).

Similarly, we analyzed the Grouped-seq data of the CRC organoids treated with oxaliplatin combined with fluorouracil (OF), which also show a high dependence on the drug concentrations (Supplementary Figure S9A). A total of 55 PCR genes were identified, leading to 19 enriched MoA pathways (Supplementary Figure S9, A and B, and Supplementary Data S15). In the top 6 pathways, 13 MPR genes, which have significant LFCs, were found and separated into three phases according to their BMDs (Supplementary Figure S9C, and Supplementary Data S16). These pathways are all related to cellular metabolic processes, by which cells counteract the effects of cytotoxic drugs (Supplementary Figure S9D). Although the exact MoA of OF is still unclear (50–52), these combined chemotherapeutic drugs usually produce several kinds of cytotoxicity to inhibit the proliferation of colorectal cancer cells (53), which are proved by our results. When we analyzed the organoids treated with fluorouracil and cetuximab, it is difficult to use the phenotype-assisted method as the viability of the organoids show no responses to both drugs. The PCA analyses of each drug with NC also demonstrated the gene expressions lacked the dependence on the drug concentrations (Supplementary Figure S10, A and C). We then tried to use the medial concentration as the BMDs for these two drugs and found 120 and 37 PCR genes for fluorouracil and cetuximab, respectively. Unfortunately, no MoA pathways were enriched to pass the threshold (Supplementary Figure S10B and D, and Supplementary Data S17, and Supplementary Data S18). As a result, the MoAs of fluorouracil and cetuximab can only be analyzed using the way shown in Figure 6F.

DISCUSSION

The Grouped-seq platform successfully translated a high-throughput RNA-seq sample preparation process onto a superhydrophobic microwell array platform for combined phenotypic and transcriptomic analysis of drug responses of tumor organoids. Several key innovations implemented in the Grouped-seq are: (i) a newly-designed library construction method combined with a low-cost, efficient microwell barcoding strategy for pooled RNA-seq, (ii) a nanoliter-scale, high-throughput microwell array suitable for analyzing samples with limited quantities, such as tumor organoids and (iii) a more efficient pathway enrichment analysis method assisted by the phenotypic changes.

First of all, a reliable sample barcoding strategy is the key to achieve a pooled RNA-seq for high-throughput analysis. Unlike the random barcode synthesis methods adopted in most of the single-cell RNA-seq (54–56), we need to know the barcode information of each individual microwell to link the phenotypic results with the sequencing data. Thereby, we developed a ligation-based barcode synthesis method to lower the cost of the capture oligos and to achieve a high-capacity coding system with an excellent performance. We found that it only takes $1862 to obtain 192 capture oligos with a final length of 114 bases and a 5′-biotin modification using our method, in stark contrast with the cost of ∼$32 000 using the direct synthesis method (Supplementary Table S9). More importantly, our capture oligo is barcoded with a 6-base sample index and a 16-base well-code. Such a long coding sequence allows us to design 192 barcodes with appropriate GC contents, minimum poly-N fragments, and >2 Hamming-distance to one another. As a result, we are able to employ the relative nearest neighborhood searching to extract the barcode information with an efficiency of 93.8%, resulting in the excellent mapping rate of the Grouped-seq.

The nanoliter-scale superhydrophobic microwell array chip coupled with the pooled RNA-seq protocol significantly reduces the sample and the reagent consumptions for the RNA-seq. We demonstrated that 10 tumor organoids can be loaded into each microwell for combined drug response analysis followed by RNA-seq. By contrast, conventional 96- or 384-well plates may need 10 times or even 100 times more organoids to accomplish the same assays (29). The reduced sample consumption can significantly lower the costs of organoid expansion *in vitro* and shorten the sample preparation time prior to screening. In addition, the low reagent consumption together with the pooled library construction reduces the cost to ∼$2 per RNA-seq sample (Supplementary Table S10), which is similar to the reported DRUG-seq but dramatically lower than the conventional RNA-seq (Supplementary Table S11) (28). This cost reduc-

tion was achieved without sacrificing the qualities of the sequencing. With a sequencing depth of 1 M reads per microwell, ∼11 570 genes can be obtained from each microwell with a CV of 4.0% and the correlation coefficient between any two microwells is always >0.96, surpassing previously reported platforms (Supplementary Table S12) (26–28).

The pathway enrichment analysis is the critical step to decipher the drug responses of tumor organoids in our study. We employed the phenotype-assisted pathway enrichment analysis method to locate the most phenotype-related pathways and genes, leading to the precise identification of the MoAs of irinotecan and oxaliplatin combined with fluorouracil. Apparently, for the drugs, such as fluorouracil and cetuximab, to which the tumor organoids are insensitive, this method is unable to identify the central pathways and the conventional pathway enrichment method based on the comparison with the negative controls should be used to explore the changes in gene expressions.

In conclusion, the Grouped-seq platform provides an excellent high-throughput screening platform for various pharmacological applications, such as elucidation of new compound MoAs, repositioning of existing drugs, and finding of potential side effects of drugs, using not only cell lines but also precious patient-derived primary cells or organoids, which are more closely related to the *in vivo* situation. Additionally, since we have demonstrated high-throughput cell chemical transfection and electroporation on the SMARchip (57,58), the Grouped-seq platform should be able to conduct RNAi- or CRISPR-Cas9-based screening assays for gene function annotation or new drug target discovery in a more scalable and economical manner in the future.

DATA AVAILABILITY

All data are available in the main text or the supplementary materials. The sequencing data presented in this paper have been deposited in the Sequence Read Archive (SRA) under BioProject accession number PRJNA771467 and the processed sequencing data are available at NAR Online. The pipeline for bioinformatics analysis can be found on Github (https://github.com/wuys13/Grouped-seq-analysis-pipeline).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Y. Wang for discussion on the design of capture oligos.

Author contributions: P.L., Y.W. and X.C. conceived the idea. Y.W. fabricated the SMARchip, developed library preparation pipeline, and completed the development of the method. P.L. and X.C. supervised the work. K.L. and Y.L. developed the CRC organoid model and performed the drug tests. T.S. provided CRC patient samples. C.D. conducted WES analysis. C.L. and T.Z. helped to conduct TruSeq library preparation. D.T. and X.C. helped to develop RNA-seq data analysis pipeline. P.L. and Y.W. wrote the manuscript.

FUNDING

National Natural Science Foundation of China [81771931]; Beijing Municipal Science and Technology Commission [Z190017, Z191100007619018]. Funding for open access charge: National Natural Science Foundation of China [81771931]; Beijing Municipal Science and Technology Commission [Z190017, Z191100007619018].

Conflict of interest statement. None declared.

REFERENCES

1. Weeber,F., Ooft,S.N., Dijkstra,K.K. and Voest,E.E. (2017) Tumor organoids as a pre-clinical cancer model for drug discovery. *Cell Chem. Biol.*, **24**, 1092–1100.
2. Aboulkheyr Es,H., Montazeri,L., Aref,A.R., Vosough,M. and Baharvand,H. (2018) Personalized cancer medicine: an organoid approach. *Trends Biotechnol.*, **36**, 358–371.
3. Drost,J. and Clevers,H. (2018) Organoids in cancer research. *Nat. Rev. Cancer*, **18**, 407–418.
4. Bleijs,M., van de Wetering,M., Clevers,H. and Drost,J. (2019) Xenograft and organoid model systems in cancer research. *EMBO J.*, **38**, e101654.
5. Vivarelli,S., Candido,S., Caruso,G., Falzone,L. and Libra,M. (2020) Patient-derived tumor organoids for drug repositioning in cancer care: a promising approach in the era of tailored treatment. *Cancers (Basel)*, **12**, 3636.
6. van de Wetering,M., Francies,H.E., Francis,J.M., Bounova,G., Iorio,F., Pronk,A., van Houdt,W., van Gorp,J., Taylor-Weiner,A., Kester,L. *et al.* (2015) Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*, **161**, 933–945.
7. Vlachogiannis,G., Hedayat,S., Vatsiou,A., Jamin,Y., Fernandez-Mateos,J., Khan,K., Lampis,A., Eason,K., Huntingford,I., Burke,R. *et al.* (2018) Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science*, **359**, 920–926.
8. Ooft,S.N., Weeber,F., Dijkstra,K.K., McLean,C.M., Kaing,S., van Werkhoven,E., Schipper,L., Hoes,L., Vis,D.J., van de Haar,J. *et al.* (2019) Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Transl. Med.*, **11**, e2574.
9. Sachs,N., de Ligt,J., Kopper,O., Gogola,E., Bounova,G., Weeber,F., Balgobind,A.V., Wind,K., Gracanin,A., Begthel,H. *et al.* (2018) A living biobank of breast cancer organoids captures disease heterogeneity. *Cell*, **172**, 373–386.
10. Hu,Y., Sui,X., Song,F., Li,Y., Li,K., Chen,Z., Yang,F., Chen,X., Zhang,Y., Wang,X. *et al.* (2021) Lung cancer organoids analyzed on microwell arrays predict drug responses of patients within a week. *Nat. Commun.*, **12**, 2581.
11. Kim,M., Mun,H., Sung,C.O., Cho,E.J., Jeon,H.J., Chun,S.M., Jung,D.J., Shin,T.H., Jeong,G.S., Kim,D.K. *et al.* (2019) Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. *Nat. Commun.*, **10**, 3991.
12. Miyamoto,S., Narita,T., Komiya,M., Fujii,G., Hamoya,T., Nakanishi,R., Tamura,S., Kurokawa,Y., Takahashi,M. and Mutoh,M. (2019) Novel screening system revealed that intracellular cholesterol trafficking can be a good target for colon cancer prevention. *Sci. Rep.*, **9**, 6192.
13. Zhan,T., Ambrosi,G., Wandmacher,A.M., Rauscher,B., Betge,J., Rindtorff,N., Haussler,R.S., Hinsenkamp,I., Bamberg,L., Hessling,B. *et al.* (2019) MEK inhibitors activate Wnt signalling and induce stem cell plasticity in colorectal cancer. *Nat. Commun.*, **10**, 2197.
14. Sackmann,E.K., Fulton,A.L. and Beebe,D.J. (2014) The present and future role of microfluidics in biomedical research. *Nature*, **507**, 181–189.
15. Yeo,L.Y., Chang,H.C., Chan,P.P. and Friend,J.R. (2011) Microfluidic devices for bioapplications. *Small*, **7**, 12–48.
16. Brandenberg,N., Hoehnel,S., Kuttler,F., Homicsko,K., Ceroni,C., Ringel,T., Gjorevski,N., Schwank,G., Coukos,G., Turcatti,G. *et al.* (2020) High-throughput automated organoid culture via stem-cell aggregation in microcavity arrays. *Nat. Biomed. Eng*, **4**, 863–874.

17. Schuster,B., Junkin,M., Kashaf,S.S., Romero-Calvo,I., Kirby,K., Matthews,J., Weber,C.R., Rzhetsky,A., White,K.P. and Tay,S. (2020) Automated microfluidic platform for dynamic and combinatorial drug screening of tumor organoids. *Nat. Commun.*, **11**, 5271.
18. Bansal,M., Yang,J., Karan,C., Menden,M.P., Costello,J.C., Tang,H., Xiao,G., Li,Y., Allen,J., Zhong,R. *et al.* (2014) A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.*, **32**, 1213–1222.
19. Bisikirska,B., Bansal,M., Shen,Y., Teruya-Feldstein,J., Chaganti,R. and Califano,A. (2016) Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression. *Cancer Res.*, **76**, 664–674.
20. Woo,J.H., Shimoni,Y., Yang,W.S., Subramaniam,P., Iyer,A., Nicoletti,P., Rodriguez Martinez,M., Lopez,G., Mattioli,M., Realubit,R. *et al.* (2015) Elucidating compound mechanism of action by network perturbation analysis. *Cell*, **162**, 441–451.
21. Wagner,B.K. and Schreiber,S.L. (2016) The power of sophisticated phenotypic screening and modern mechanism-of-action methods. *Cell Chem. Biol.*, **23**, 3–9.
22. Olivier,M., Asmis,R., Hawkins,G.A., Howard,T.D. and Cox,L.A. (2019) The need for multi-omics biomarker signatures in precision medicine. *Int. J. Mol. Sci.*, **20**, 4781.
23. Sun,Y.V. and Hu,Y.J. (2016) Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv. Genet.*, **93**, 147–190.
24. Li,H., Qiu,J. and Fu,X.D. (2012) RASL-seq for massively parallel and quantitative analysis of gene expression. *Curr. Protoc. Mol. Biol.*, https://doi.org/10.1002/0471142727.mb0413s98.
25. Subramanian,A., Narayan,R., Corsello,S.M., Peck,D.D., Natoli,T.E., Lu,X., Gould,J., Davis,J.F., Tubelli,A.A., Asiedu,J.K. *et al.* (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, **171**, 1437–1452.
26. Alpern,D., Gardeux,V., Russeil,J., Mangeat,B., Meireles-Filho,A.C.A., Breysse,R., Hacker,D. and Deplancke,B. (2019) BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.*, **20**, 71.
27. Bush,E.C., Ray,F., Alvarez,M.J., Realubit,R., Li,H., Karan,C., Califano,A. and Sims,P.A. (2017) PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat. Commun.*, **8**, 105.
28. Ye,C., Ho,D.J., Neri,M., Yang,C., Kulkarni,T., Randhawa,R., Henault,M., Mostacci,N., Farmer,P., Renner,S. *et al.* (2018) DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.*, **9**, 4307.
29. Norkin,M., Ordonez-Moran,P. and Huelsken,J. (2021) High-content, targeted RNA-seq screening in organoids for drug discovery in colorectal cancer. *Cell Rep.*, **35**, 109026.
30. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
31. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
32. Cibulskis,K., Lawrence,M.S., Carter,S.L., Sivachenko,A., Jaffe,D., Sougnez,C., Gabriel,S., Meyerson,M., Lander,E.S. and Getz,G. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.*, **31**, 213–219.
33. Saunders,C.T., Wong,W.S., Swamy,S., Becq,J., Murray,L.J. and Cheetham,R.K. (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*, **28**, 1811–1817.
34. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
35. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
36. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
37. Yu,G., Wang,L.G., Han,Y. and He,Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
38. Zhang,P., Zhang,J., Bian,S., Chen,Z., Hu,Y., Hu,R., Li,J., Cheng,Y., Zhang,X., Zhou,Y. *et al.* (2016) High-throughput superhydrophobic microwell arrays for investigating multifactorial stem cell niches. *Lab Chip*, **16**, 2996–3006.
39. Wong,T.S., Kang,S.H., Tang,S.K., Smythe,E.J., Hatton,B.D., Grinthal,A. and Aizenberg,J. (2011) Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature*, **477**, 443–447.
40. Wang,Y., Wu,Y., Chen,Y., Zhang,J., Chen,X. and Liu,P. (2018) Nanoliter centrifugal liquid dispenser coupled with superhydrophobic microwell array chips for high-throughput cell assays. *Micromachines*, **9**, 286.
41. Picelli,S., Bjorklund,A.K., Faridani,O.R., Sagasser,S., Winberg,G. and Sandberg,R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
42. Luca,A.C., Mersch,S., Deenen,R., Schmidt,S., Messner,I., Schafer,K.L., Baldus,S.E., Huckenbeck,W., Piekorz,R.P., Knoefel,W.T. *et al.* (2013) Impact of the 3D microenvironment on phenotype, gene expression, and EGFR inhibition of colorectal cancer cell lines. *PLoS One*, **8**, e59689.
43. Xu,X., Farach-Carson,M.C. and Jia,X.Q. (2014) Three-dimensional in vitro tumor models for cancer research and drug evaluation. *Biotechnol. Adv.*, **32**, 1256–1268.
44. Breslin,S. and O'Driscoll,L. (2013) Three-dimensional cell culture: the missing link in drug discovery. *Drug Discov. Today*, **18**, 240–249.
45. Phillips,J.R., Svoboda,D.L., Tandon,A., Patel,S., Sedykh,A., Mav,D., Kuo,B., Yauk,C.L., Yang,L.L., Thomas,R.S. *et al.* (2019) BMDExpress 2: enhanced transcriptomic dose-response analysis workflow. *Bioinformatics*, **35**, 1780–1782.
46. Li,F., Ling,X., Harris,D.L., Liao,J., Wang,Y., Westover,D., Jiang,G., Xu,B., Boland,P.M. and Jin,C. (2017) Topoisomerase I (Top1): a major target of FL118 for its antitumor efficacy or mainly involved in its side effects of hematopoietic toxicity? *Am. J. Cancer Res.*, **7**, 370–382.
47. Meisenberg,C., Gilbert,D.C., Chalmers,A., Haley,V., Gollins,S., Ward,S.E. and El-Khamisy,S.F. (2015) Clinical and cellular roles for TDP1 and TOP1 in modulating colorectal cancer response to irinotecan. *Mol. Cancer Ther.*, **14**, 575–585.
48. Pommier,Y. (2006) Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer*, **6**, 789–802.
49. Chavez,J.D., Schweppe,D.K., Eng,J.K., Zheng,C., Taipale,A., Zhang,Y., Takara,K. and Bruce,J.E. (2015) Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.*, **6**, 7928.
50. Alcindor,T. and Beauger,N. (2011) Oxaliplatin: a review in the era of molecularly targeted therapy. *Curr. Oncol.*, **18**, 18–25.
51. Alian,O.M., Azmi,A.S. and Mohammad,R.M. (2012) Network insights on oxaliplatin anti-cancer mechanisms. *Clin. Transl. Med.*, **1**, 26.
52. Kelland,L. (2007) The resurgence of platinum-based cancer chemotherapy. *Nat. Rev. Cancer*, **7**, 573–584.
53. Longley,D.B., Harkin,D.P. and Johnston,P.G. (2003) 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer*, **3**, 330–338.
54. Macosko,E.Z., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
55. Zhang,M., Zou,Y., Xu,X., Zhang,X., Gao,M., Song,J., Huang,P., Chen,Q., Zhu,Z., Lin,W. *et al.* (2020) Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.*, **11**, 2118.
56. Zheng,G.X., Terry,J.M., Belgrader,P., Ryvkin,P., Bent,Z.W., Wilson,R., Ziraldo,S.B., Wheeler,T.D., McDermott,G.P., Zhu,J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
57. Bian,S.T., Zhou,Y.C., Hu,Y.W., Cheng,J., Chen,X.F., Xu,Y.C. and Liu,P. (2017) High-throughput in situ cell electroporation microsystem for parallel delivery of single guide RNAs into mammalian cells. *Sci. Rep.*, **7**, 42512.
58. Zhang,J.X., Hu,Y.W., Wang,X.Q., Liu,P. and Chen,X.F. (2019) High-throughput platform for efficient chemical transfection, virus packaging, and transduction. *Micromachines*, **10**, 387.