Published in final edited form as:

ACS Nano. 2018 May 22; 12(5): 4687–4694. doi:10.1021/acsnano.8b01272.

Single-Cell Digital Lysates Generated by Phase-Switch Microfluidic Device Reveal Transcriptome Perturbation of Cell Cycle

Yan Chen^{†,∇}, Joshua Millstein[‡], Yao Liu^{†,○}, Gina Y. Chen[†], Xuelian Chen[†], Andres Stucky[†], Cunye Qu[†], Jian-Bing Fan[§], Xiao Chang^{||}, Ava Soleimany^{||}, Kai Wang^{||}, Jiangjian Zhong[⊥], Jie Liu[‡], Frank D. Gilliland[‡], Zhongjun Li^{*,†,‡,§}, Xi Zhang^{*,†,○}, and Jiang F. Zhong^{*,†,⊥,iD}

[†]Division of Periodontology, Diagnostic Sciences and Dental Hygiene, and Division of Biomedical Sciences, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, California 90089, United States

[‡]Division of Biostatistics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, United States

[§]Illumina Inc., San Diego, California 92122, United States

^{||}Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, United States

[⊥]Z-genetic Medicine Inc., Temple City, California 91780, United States

Abstract

With conventional gene expression profiling, information concerning cellular heterogeneity is often lost in the physical mixing and averaging of millions of cells. Single-cell transcriptome analysis has the potential to address these issues. However, there is a need to determine how many cells are needed to draw meaningful conclusions in each single-cell study. Here, we introduce the concept of “digital lysate” for assessing cellular heterogeneity with a phase-switch microfluidic platform and apply it to construct a molecular map of transcriptome perturbation during the cell cycle. Using a phase-switch droplet microfluidic platform and next-generation sequencing, we obtained transcriptomes of single cells by random sampling. Digital lysates were generated by permutating and averaging multiple single-cell transcriptomes. In our studied cell populations, digital lysates converged to physical lysates ($r = 0.93$), and the sample-to-sample repeatability was comparable to that of conventional analysis of a physical lysate ($r = 0.98$). After determining the number of cells needed, single-cell transcriptomes were used to organize cells into a map by molecular similarity, and the map was validated by cell cycle-specific markers ($p = 0.003$). Cell

*Corresponding Authors. jzhong@usc.edu, zhangxxi@sina.com, johnneyusc@gmail.com.

ORCID

Jiang F. Zhong: 0000-0003-0551-5161

[∇]Present Addresses

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[○]Department of Hematology, The Second Affiliated Hospital, Third Military Medical University, Chongqing, China

[‡]Department of Blood Transfusion, Lab of Radiation Biology, The Second Affiliated Hospital, Third Military Medical University, Chongqing, China

The authors declare no competing financial interest.

cycle regulatory genes were inferred using this molecular map and verified with siRNA assays. The study described here provides an effective approach, the generation and analysis of digital lysates, to investigate cellular heterogeneity.

Graphical Abstract

Keywords

microfluidic; single cell; digital lysate; transcriptome; human embryonic stem cell

Single-cell transcriptome analysis can provide insights into molecular biology and has become a common approach in many laboratories. Many single-cell capture devices have been reported, including microfluidic devices we have developed.¹⁻³ Commercial single-cell technologies, such as BD FACSAria, Fluidigm C1, Silicon Biosystems DEPArray, and AVISO CellSelector, are associated with high equipment and reagent costs.⁴⁻⁷ Each single cell costs as much as a bulk lysate in next-generation sequencing (NGS). In order to lower the cost per sample, barcoding has been applied to enable pooling cells together for RNA-seq. By mixing a cell and a barcoding bead in a microwell or droplet and carefully controlling the mixing concentrations, hundreds to thousands of cells can be pooled for one RNA-seq reaction to significantly lower the cost of RNA-seq per cell.⁸⁻¹⁰ However, the coverage of each cell (reads per cell) still depends on the sequencing depth, which significantly contributes to the cost and is not reduced by pooling. Therefore, knowing how many single cells can represent a specific cell population so that the number can be minimized is perhaps the best way to effectively reduce the cost of profiling a heterogeneous cell population. Here, we report a phase-switch microfluidic processor to perform nanoliter reverse transcription (RT) for high-quality cDNA. Because RT is a reaction that depends on

mRNA concentration, the device switches each single cell from aquatic medium into a hydrophobic oil droplet to minimize reaction volume (therefore increasing RNA concentration) for the RT reaction. The resulting high-quality cDNAs enable digital lysate generation for assessing the minimal cell numbers needed to represent a heterogeneous cell population. This approach has the potential to profile a heterogeneous cell population with minimal cost.

Previously, we reported a highly efficient approach to generating whole-genome transcriptomes for a large number of single cells using a microfluidic platform.^{2,3,11} We have demonstrated that our microfluidic platform can produce high-quality data relative to a conventional population cell lysate.^{1,3,11} In this study, we generated thousands of digital lysates from 35 single-cell transcriptomes with a phase-switch microfluidic platform to assess cellular heterogeneity of a H9 cell population. The microfluidic device integrates multiple single-cell isolation chambers and on-site droplet generators. The isolation chambers contain hydrodynamic cell traps to ensure only one cell occupies each trap, and the special design of droplet generator enables fast oil loading and replacement to switch the cells from an aquatic phase (cell media) into a hydrophobic phase (oil droplets). The phase-switch mechanism minimizes the carry-over volume of each cell and enables nanoliter RT with a very high mRNA concentration, which is essential for obtaining high-quality cDNA.

Using cDNA obtained in this manner, we demonstrate the digital lysate approach here by showing how we were able to construct a transcriptome perturbation map of the cell cycle with only 29 single-cell transcriptomes. With this sequential perturbation map of the cell cycle, cell cycle regulatory genes were identified. One of these genes, *DMNT3B*, was verified with a functional siRNA assay. The molecular mapping technique (sequential perturbation of all genes during the cell cycle) and other methods described here could be applied to various cell populations in order to reconstruct other sequences of biological events similar to the cell cycle. The phase-switch microfluidic device and the digital lysate approach provide an effective method to investigate cellular heterogeneity with limited numbers of single-cell transcriptomes. It can be applied to a broad range of molecular biology studies of heterogeneous cell populations, including rare circulating tumor cells.

RESULTS

Phase-Switch Microfluidic Device for Nanoliter RT

In order to encapsulate a live mammalian cell and minimize the carry-over volume for RT, we have developed a phase-switch single-cell capture method. A key factor for nanoliter RT is controlling the volume of individual single-cell reactions, preferably achieving 10 nL to increase mRNA concentration. To precisely control the volume of each cell suspension, we utilized oil to strip off the carrier liquid by switching the cell from an aquatic phase into a hydrophobic oil phase. This phase-switch removed most carry-over liquid (cell medium) and allowed the cells (with almost no carry-over medium) to be added directly into 10 nL of RT master mix for achieving high mRNA concentration. In addition, the oil phase prevented the evaporation of the RT master mix. The entirety of this single-cell extraction experiment was monitored in real time under a computer-controlled microscope. Single cells were captured individually in the sieve structure in each capture unit (Figure 1). The cell loading process

was rapid, taking less than 2 s, with a single-cell occupation rate of 97% on average. To encapsulate a human embryonic stem cell (hESC) in a volume-defined droplet, we used 3 M fluid (Novec 7500) containing 1% pegylated surfactant as the oil phase to switch each cell from the aquatic phase (media) into the hydrophobic oil, which stripped off all carry-over liquid. First, oil filled all the flow channels except the center channel next to the sieve structure. After cell capturing, each cell loading channel was then partitioned by valves, and oil was pushed from the left side to strip all liquid off the captured cells (Figure 1, top inset). The captured cells could then flow in the oil into a chamber with 10 nL of RT master mix (Figure 1, middle inset). Once the cells were in the chamber, centrifugation of the chip was used to bring the cells from the oil layer (top of the chamber) into the aquatic RT master mix (bottom of the chamber) for high mRNA concentration RT (Figure 1, bottom inset). The high-quality single-cell cDNAs were then subjected to RNA-seq and digital lysate generation after being diluted to 10 μ L and recovered from the cyclic olefin copolymer (COC) microwells.

Digital Lysate Converts to Physical Lysates

We use a digital lysate to determine how many single cells could represent a specific cell population with next-generation sequencing (RNA-seq). Due to cellular heterogeneity, transcriptomes of single cells are much less similar to each other than transcriptomes of multicell lysates, which average many thousands of single cells. To fully represent a heterogeneous cell population would require single-cell profiling of the same number of cells as processed in a bulk population cell lysate (thousands to millions of cells), a task that is not financially feasible. To circumvent this hurdle, we created “digital lysates” which average multiple single-cell transcriptomes digitally and compared these digital lysates to conventional multicell lysates, which we call ‘physical lysates’. The digital lysates were computed using differing numbers of single-cell transcriptomes without replacement. Therefore, thousands of digital lysates could be generated from 35 single-cell transcriptomes. Comparing these digital lysates to physical lysates can assess cellular heterogeneity because, in a less heterogeneous population, digital lysates with fewer cells will converge on the physical lysate.

A total of 35 human embryonic stem cells (H9) were subjected to single-cell RNA-seq to obtain transcriptome profiles as described previously.^{1–3,11} When we created a digital lysate consisting of just a single cell, a substantial amount of heterogeneity was indicated by differences in the correlations between the single-cell transcriptomes and the physical lysates with 1000 cells (Figure 2). However, digital lysates containing more cells quickly converged on the physical lysate with an apparent asymptote at approximately $r = 0.93$ when 22 or more cells were included. The trajectory of the correlations as the number of single cells in the digital lysates increases is a measure of both heterogeneity and the validity of the platform. That is, if a true asymptote is reached, then adding more single cells to the average would not increase the correlation, and the remaining difference is due to technical variations. The qPCR analysis of specific well-known genes also confirmed the validity of the system. As expected, correlations between H9 digital lysate and physical lysate of various cancer cells are substantially lower than those calculated with H9 physical lysate (Figure 2, black dots). The correlations between H9 and kidney cancer calculated by digital

lysates were $r = 0.64$ when 22 or more cells were used in the digital lysate (Figure 2, green dots). The correlation was similar to the correlation calculated by physical lysates. Similarly, for breast cancer lysate, the correlations calculated with digital lysates were stable at $r = 0.48$ with 22 or more cells and were similar to the correlations calculated with physical lysate (Figure 2, red dots). These results indicated that 22 or more H9 cells are sufficient to represent this H9 cell population.

Sequential Perturbation of the Transcriptome during a Cell Cycle

A time series of transcriptome perturbations is the most informative way to infer gene regulation, but requires a highly homogeneous cell population to obtain reliable data at each time point. The single-cell approach can circumvent the need for homogeneous cell populations, which are very difficult if not impossible to obtain. Differentiation/maturation of a cell is orchestrated by sequential expression of a series of genes. Therefore, mRNA expression profiles (transcriptomes) from consecutive developmental stages are more similar than those from disparate stages. With random sampling, gene expression profiles from single cells at various developmental stages can be obtained and organized by similarity into a sequential order.

We isolated 29 individual cells that carried fluorescent cell cycle indicators (fucci) and obtained single-cell transcriptomes with our microfluidic platform. Based on digital lysates created from the single-cell profiles, we identified that 15 cells should be sufficient to represent this fucci cell population (Figure 3A, inset), which has specific fluorescent colors at different cell cycle stages. A similarity matrix was calculated based on known cell cycle genes (GO: 0022402). The cells were then organized based on transcriptome similarity without using the fluorescent cell cycle color for reference (Figure 3A). In agreement with our estimation that 15 cells would be sufficient to represent the cell population, random sampling revealed two pairs of cells with very similar profiles (indicated by arrows in Figure 3A). The molecular map with the sequential ordering of the single-cell transcriptomes was confirmed by the fluorescent color of each cell, which was recorded before obtaining the transcriptomes but was not used for the map construction ($P = 0.0031$, Figure 3B). When a random gene list was used to organize the transcriptomes, the map was much less correlated with the cell cycle stages indicated by the fluorescent indicators. These results indicate that single-cell transcriptomes can be organized into a molecular map with limited prior knowledge. The molecular map can then provide a dynamic perturbation profile of the temporal and quantitative changes in gene expression across the entire cell cycle and can be further used to infer gene functions.

Functional Validation of Regulatory Genes with siRNA

To demonstrate the utility of the molecular map, the expression levels of known cell cycle genes were plotted in the corresponding sequential order (Figure 4). As expected, known cell cycle genes had expression perturbation profiles that agreed with previously reported studies of physical population cell lysates (Figure 4). In addition to known cell cycle genes, genes indicated by the self-organizing map (SOM) analysis¹² were plotted onto the cell cycle map to identify candidate cell cycle genes including DNA methyltransferase (*DMNT3B*), which was selected for confirmation by siRNA knockdown. *DMNT3B* was

found to have gradually increasing expression during G1/S/G2/M progression. Furthermore, we compared the molecular map expression profile of *DNMT3B* to a panel of cell cycle genes (GO: 0022402) and 20 additional well-characterized cell cycle genes in single cells at different cell cycle stages and found the expression of *DNMT3B* was consistently correlated throughout the cell cycle to known cell cycle genes including *CDK9* (cc = 0.82), *PINX1* (cc = 0.76), and *PMS2L4* (cc = 0.83).

DNMT3B is a member of the DNA methyltransferase family, which is important in *de novo* DNA methylation and plays key roles in transcriptional repression. It was found to be down-regulated in arrested MCF7 breast cancer cells in G1, yet its significance in cell cycle progression has remained elusive.¹³ Using siRNA targeting *DNMT3B* resulted in an accumulation of cells in the S phase (yellow cells, Figure 5). Not all cells were blocked in the G1/S phase due to the efficiency of transfection, which can be explained by the fact that siRNA suppression is transitory and not all cells are exposed to the same amount of siRNA. The siRNA effect was most obvious 24 h post-transfection, as shown in the fluorescent images (Figure 5). The functional assay thus further confirmed the cell cycle regulatory role of *DNMT3B* as inferred from the single-cell molecular map.

DISCUSSION

Although a multicell lysate provides a large amount of RNA for analysis, averaging all cells in a population cell lysate erases information on cell-specific regulatory states. Heterogeneity in natural cell populations can be due to a variety of factors including cell cycle asynchrony, the presence of multiple cell types, and a spectrum of microenvironments. The best way to profile a heterogeneous cell population is obtaining a single-cell transcriptome from each cell, but this is financially infeasible. With the microfluidic digital lysate approach reported here, it is possible to describe a heterogeneous cellular population with a limited number of single-cell transcriptomes using random sampling.

This system was externally validated by performing qPCR of known cell cycle genes, next-generation sequencing, and microarray. The fact that agreement increased as we averaged data across an increasing number of single cells implies that a large component of the data is measuring the same underlying values as would be measured in a physical population cell lysate and that the remaining component is largely random. In addition to validating the technology, this analytic approach essentially quantifies heterogeneity in a cell population. The internal validity of the system was demonstrated to be similar to that of the conventional approach, considering that an equally high correlation was observed among independent digital lysates constructed from 22 and 15 cells as these from physical lysates for H9 and fucci cells, respectively.

Using a digital lysate, we determined that 15 single-cell transcriptomes could represent the fucci cell population and constructed a perturbation map of the cell cycle of fucci cells. The sequential perturbation of individual genes during the cell cycle was estimated by organizing single-cell transcriptomes according to the similarity of the expression of known cell cycle genes. Fluorescent biomarkers for cell cycle stages supported these results. The sequential perturbation map reported here should be approximately equivalent to what one would

obtain from longitudinal gene expression in a perfectly synchronized cell population, which is very difficult to obtain. Sequential perturbation data from single cells can be used to dissect molecular mechanisms of cellular development including the cell cycle. For example, in this study we leveraged the map to identify unknown cell cycle genes, one of which, *DNMT3B*, we experimentally validated using siRNA. Suppression of *DNMT3B* was found to cause an accumulation of cells in S interphase with fewer cells in mitosis. The exact role that *DNMT3B* plays in the cell cycle is still unclear.

The reported microfluidic platform and analytic approach address an important question in single-cell molecular analysis, namely, how many cells are required to represent a cell population. The strategy demonstrated here was to compare digital lysates (averages of various numbers of single-cell transcriptomes) to physical population cell lysates. In our two studies (H9 cells and fucci cells), 22 and 15 or more single-cell transcriptomes were sufficient to characterize the two populations, respectively. It is expected that more than 30 cells will be needed to describe primary cells from tissues. This approach is particularly useful in the study of complex tissues. With digital lysates, the number of cells needed to represent the study population can be estimated to ensure that meaningful conclusions can be drawn, and enrichment can be applied to obtain a cell population that is amenable to single-cell transcriptome profiling.

METHODS

Microfluidic Device Fabrication

The microfluidic device consisted of a three-layer elastomeric structure and a thermoplastic substrate. The elastomeric part of the device was made from polydimethylsiloxane (PDMS, RTV615, General Electric) reversibly bonded to a cyclic olefin copolymer (COC) substrate. The microwell structures on the COC substrate were made using a computer-controlled micromilling system (CNC). The COC plate had a dimension of 5 cm × 5 cm × 2 mm, and arrays of 32 holes with 2 mm diameter were milled as microwell RT chambers.

The molds for casing PDMS elastomeric devices were made by standard photolithography with an EVG mask aligner (EVG 610, EV Group). The flow mold was a two-layer mold, with 7 μm-tall cell trap features defined in SU8-2010 photoresist (MicroChem Corp.) and 15 μm tall flow channels made in AZ-50XT positive photoresist (AZ Electronic Materials). The control mold was fabricated using SU8-2035 photoresist (Microchem Corp.) to deposit valve features of 20 μm in height.

The three-layer PDMS device was fabricated using multilayer soft lithography. The flow layer was made by pouring a mixture of PDMS (RTV615, ratio 5:1) onto the flow mold, after which it was degassed and cured for 50 min at 80 °C. The control layer was made by spin-coating PDMS (RTV615, ratio 20:1) and baking for 40 min at 80 °C. After baking, the PDMS flow layer was peeled off and aligned to the control layer and then baked at 80 °C for 60 min. Meanwhile, a blank layer with through-hole structures was prepared by spinning PDMS (RTV615, ratio 20:1) on a Si wafer with pillar structures and curing it. The two-layer structure was separated from the control mold, mounted onto the blank layer with through-hole structures, and baked for 4 h at 80 °C. Finally, the three-layer bonded structure was

treated with oxygen plasma using a reactive ion etcher (RIE-10NR, Samco, Japan), reversibly bonded to the COC plate with microwell structures, and baked overnight at 80 °C.

Device Operation

A homemade pneumatic control system with a touch screen, a single-chip microprocessor, and solenoid valves (The Lee Company, USA) was used to implement semi-automated device operation, as shown in Figure 1. The device operation requires control of only five pneumatic valves, and a control program was written in the single-chip microprocessor for chip automation. Tygon tubing was used to connect the solenoids to the microfluidic control line ports. FC-40 electronic liquid (Fluorinert, 3M, USA) was used to fill the control lines, and the valves were actuated with 25 psi of pressure.

Cell Culture and Single-Cell cDNA Synthesis

H9 human embryonic stem cells (WA09, WiCell Research Institute, Inc.) were maintained with a feeder-free protocol as previously described.^{14,15} HeLa F. (RIKEN Cell Bank RCB2812) cells (fucci cells)¹⁶ were grown under standard conditions and cultured in advanced DMEM (Invitrogen, Carlsbad, California), 1% fetal bovine serum (HyClone, South Logan, Utah), and 0.5% penicillin-streptomycin (Invitrogen, Carlsbad, California). Cells were trypsinized using TrypLE (GIBCO) and resuspended in 1× phosphate-based buffer (PBS). Single-cell encapsulation was performed with the custom-fabricated microfluidic device described above. Single-cell encapsulation, cDNA synthesis, and amplifications on the nanoliter scale were carried out as previously reported.^{1,3,11} In brief, single cells were encapsulated, and the carrier liquid was removed with the microfluidic device as described above. These cells in oil then were manipulated into 10 nl RT master mix (Qiagen, USA) for RT reaction and then diluted to 10 µL. The resulting products were then stored at −20 °C until they were used for quantitative PCR, microarray gene expression profiling and RNA-seq (Illumina, USA). Samples were prepared for PCR amplification and subsequent analysis by adding preamplification master mix (Illumina, San Diego, CA) as described previously.^{3,11} For comparison, a Single-Cell Preamplification Kit (Life-Technologies, CA) was also used for cDNA synthesis and subsequently for qPCR verification. After the cDNA quality was evaluated, the libraries were constructed with a library construction kit (Illumina, USA) for RNA-seq and further analyses.

RNA-seq Analysis

Single-cells along with bulk population cell lysates were subjected to RNA-seq. For bulk population cell lysate, total RNA was extracted from 1×10^6 cells using RNeasy mini kit, in combination with RNAase-free DNAase to remove the potential genomic DNA contamination (Qiagen). RNA concentration was quantified by Nanodrop 2000C spectrophotometer, and 100 ng of total RNA was used for the mRNA-seq library preparation by the TruSeq Stranded mRNA Library Prep kit (Illumina). For single-cell RNA-seq, single cells were isolated with our microfluidic device, and cDNAs were obtained with REPLI-g WTA Single Cell Kit (Qiagen) according to manufacturer's protocol. The library length was detected by BioAnalyzer 2100 (Agilent), and the RNA concentration was confirmed by StepOnePlus Real-Time PCR System (Thermo Fisher, 4376600). Fragmentation of amplified cDNA was performed with NEBNext dsDNA Fragmentase (NEB). For each

sample, 100 ng of small fragments (50–500bp) were used for RNA-seq library preparation by the NEBNext Library Prep kit (NEB). All libraries were quantified by BioAnalyzer 2100 (Agilent) and KAPA Library Quantification Standards Kits for Illumina platforms (KAPA) and were sequenced on the Illumina HiSeq 2000 platform (Illumina, USA). The raw reads generated were filtered according to sequencing quality and with regard to adaptor contamination and duplicated reads. Thus, only high-quality reads were used in the genome assembly. The RNA-seq data were analyzed with Partek Flow version 4 (Partek Inc., USA). Bases with Phred score <20 were trimmed from both ends of the raw sequencing reads, and trimmed reads shorter than 25 nt were excluded from downstream analyses. Both pre- and post-alignment QA/QC was carried out with default settings as part of the workflow.

Trimmed reads were mapped onto human genome hg38 using Tophat 2.0.8 as implemented in Flow with default settings and using Gencode 20 annotation as guidance. Gencode 26 annotation (www.gencodegenes.org) was used to quantify aligned reads to genes/transcripts using Partek E/M method.¹⁷ Read counts per gene in all samples were normalized by adding 0.001 followed by reads per kilobase of transcript per million mapped reads (RPKM) normalization and analyzed for differential expression using Partek's gene specific analysis method (genes with <10 reads in any sample were excluded). To generate significantly differentially expressed genes among different samples, a cutoff of FDR adjusted $p < 0.05$ (Poisson regression) and fold change $> |2|$ was applied.

Gene Expression Normalization and QC

Illumina gene expression data from each sample were processed and normalized independently of the other samples. Analyses were restricted to genes significantly expressed in all single cells and population cell lysates at a nominal p -value of 0.01, yielding 2181 significant expression features out of 29,377 annotated features. For assessing agreement between single-cell data and population cell lysate data, both were processed using a log 2 transformation, followed by quantile normalization. To maximize the signal-to-noise ratio for fine mapping of single cells to the cell cycle, a more comprehensive approach was taken. The Lumi R package¹⁸ was used to employ a variance-stabilizing transformation¹⁹ that utilizes technical replicates followed by robust spline normalization. Extreme outliers across single cells were identified on a gene-by-gene basis, and the values were set to “missing” if they were more extreme than three interquartile ranges away from the first or third quartiles. However, no more than one outlier exceeded this criterion. Missing data were then inputted using a nearest-neighbor averaging method.

siRNA Transfection

Cells (2×10^5) were seeded in 6-well plate and cultured until around 60% confluence. The siRNAs to *DNMT3B* (Santa Cruz) were transfected into the cells according to the manufacturer's instructions with slight modification. Briefly, 6 μ L siRNA and 6 μ L siRNA transfection reagent (Invitrogen) were resuspended in 100 μ L Opti-MEM media (Invitrogen), and the mixture was incubated at room temperature for 20 min, then added to cells cultured in serum-free, antibiotic-free medium. After 6 h, the medium was replaced with regular cell culture medium. At different time points, cells were fixed for FACS

analysis to determine the distribution of cells in different cell cycle stages based on stage-specific fluorescent biomarkers.

Statistical Analysis

In cells that are approximately homogeneous with respect to factors other than their cell cycle stages, expression of cell cycle-related genes can be used to determine the position in the cell cycle of each cell in relation to others. That is, similarity between cells in expression data should reflect similarity in cell cycle stage. We restricted this analysis not only to expressed genes (as described above) but also to genes for which existing literature provides some evidence of their involvement in the cell cycle. Specifically, the inclusion criteria required that genes be classified in the GO category *cell cycle process* (GO: 0022402). Among the many approaches that have been applied to clustering genes based on expression patterns, Ray *et al.*²⁰ applied a traveling salesman problem (TSP) solver called Concorde²¹ to obtain a one-dimensional ordering of genes within clusters. We applied Concorde to gene expression data, but here we used it to estimate the shortest Hamiltonian path (a variation of the TSP where the path does not end at the starting position) based on Euclidean distance through the single cells rather than genes. This approach required the construction of an adjacency matrix for single cells, which was generated using a network-based approach usually applied to the identification of coexpressed modules of genes.²²

The cell cycle status of individual cells forms a perturbation map, that is, a diagram of how expression of a gene is perturbed as a cell passes through the cell cycle. This perturbation map was used with the self-organizing maps (SOM)¹² approach to identify clusters of genes with the same perturbation signatures. Genes with perturbation patterns which are similar to known cell cycle genes were inferred as cell cycle genes, yielding a candidate cell cycle gene, *DMNT3B*, which was selected for confirmation by functional assay (siRNA knockdown, as described above).

Acknowledgments

This work was supported in part by the National Institutes of Health (NCI, R01CA164509 and R01CA197903, NIEHS, R01ES021801-04S1), National Science Foundation (CHE1213161), and a Seed Grant from the University of Southern California.

References

1. Zhong JF, Chen Y, Marcus JS, Scherer A, Quake SR, Taylor CR, Weiner LP. A Microfluidic Processor for Gene Expression Profiling of Single Human Embryonic Stem Cells. *Lab Chip*. 2008; 8:68–74. [PubMed: 18094763]
2. Li Z, Zhang C, Weiner LP, Zhang Y, Zhong JF. Molecular Characterization of Heterogeneous Mesenchymal Stem Cells with Single-Cell Transcriptomes. *Biotechnol. Adv.* 2013; 31:312–317. [PubMed: 23266308]
3. Fan JB, Chen J, April CS, Fisher JS, Klotzle B, Bibikova M, Kaper F, Ronaghi M, Linnarsson S, Ota T, Chien J, Laurent LC, Loring JF, Nisperos SV, Chen GY, Zhong JF. Highly Parallel Genome-Wide Expression Analysis of Single Mammalian Cells. *PLoS One*. 2012; 7:e30794. [PubMed: 22347404]
4. Shaw JA, Guttery DS, Hills A, Fernandez-Garcia D, Page K, Rosales BM, Goddard KS, Hastings RK, Luo J, Ogle O, Woodley L, Ali S, Stebbing J, Coombes RC. Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts. *Clin. Cancer Res.* 2017; 23:88–96. [PubMed: 27334837]

5. Pieper IL, Radley G, Chan CH, Friedmann Y, Foster G, Thornton CA. Quantification Methods for Human and Large Animal Leukocytes Using DNA Dyes by Flow Cytometry. *Cytometry, Part A*. 2016; 89:565–574.
6. Neumann MH, Schneck H, Decker Y, Schomer S, Franken A, Endris V, Pfarr N, Weichert W, Niederacher D, Fehm T, Neubauer H. Isolation and Characterization of Circulating Tumor Cells Using a Novel Workflow Combining the Cellsearch® System and the Cellcelector. *Biotechnol. Prog.* 2017; 33:125–132. [PubMed: 27126369]
7. Genshaft AS, Li S, Gallant CJ, Darmanis S, Prakadan SM, Ziegler CG, Lundberg M, Fredriksson S, Hong J, Regev A, Livak KJ, Landegren U, Shalek AK. Multiplexed, Targeted Profiling of Single-Cell Proteomes and Transcriptomes in a Single Reaction. *Genome Biol.* 2016; 17:188. [PubMed: 27640647]
8. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015; 161:1202–1214. [PubMed: 26000488]
9. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015; 161:1187–1201. [PubMed: 26000487]
10. Fan HC, Fu GK, Fodor SP. Combinatorial Labeling of Single Cells for Gene Expression Cytometry. *Science*. 2015; 347:1258367. [PubMed: 25657253]
11. Chen Y, Zhang B, Feng H, Shu W, Chen GY, Zhong JF. An Automated Microfluidic Device for Assessment of Mammalian Cell Genetic Stability. *Lab Chip*. 2012; 12:3930–3935. [PubMed: 22814625]
12. Kohonen T. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*. 1982; 43:59–69.
13. Robertson KD, Keyomarsi K, Gonzales FA, Velicescu M, Jones PA. Differential Mrna Expression of the Human DNA Methyltransferases (Dnmts) 1, 3a and 3b During the G(0)/G(1) to S Phase Transition in Normal and Tumor Cells. *Nucleic Acids Res.* 2000; 28:2108–2113. [PubMed: 10773079]
14. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*. 1998; 282:1145–1147. [PubMed: 9804556]
15. Xu C, Inokuma MS, Denham J, Golds K, Kundu P, Gold JD, Carpenter MK. Feeder-Free Growth of Undifferentiated Human Embryonic Stem Cells. *Nat. Biotechnol.* 2001; 19:971–974. [PubMed: 11581665]
16. Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, Imamura T, Ogawa M, Masai H, Miyawaki A. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*. 2008; 132:487–498. [PubMed: 18267078]
17. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, et al. GENCODE: The Reference Human Genome Annotation for the ENCODE Project. *Genome Res.* 2012; 22:1760–1774. [PubMed: 22955987]
18. Du P, Kibbe WA, Lin SM. Lumi: A Pipeline for Processing Illumina Microarray. *Bioinformatics*. 2008; 24:1547–1548. [PubMed: 18467348]
19. Lin SM, Du P, Huber W, Kibbe WA. Model-Based Variance-Stabilizing Transformation for Illumina Microarray Data. *Nucleic Acids Res.* 2008; 36:e11. [PubMed: 18178591]
20. Ray SS, Bandyopadhyay S, Pal SK. Gene Ordering in Partitive Clustering Using Microarray Expressions. *J. Biosci.* 2007; 32:1019–1025. [PubMed: 17914244]
21. Applegate D, Cook W, Rohe A. Chained Lin-Kernighan for Large Traveling Salesman Problems. *INFORMS J. on Computing*. 2003; 15:82–92.
22. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* 2005; 4:17.

Figure 1.

Phase-switch microfluidic platform. The microfluidic device has multiple layers with microwells (RT chambers) in a COC layer. The valves and fluid loading are controlled by a programmable control unit (blue arrow). A cell (red arrow) is captured in the microstructure, and the carrying liquid (medium) is stripped off by oil (top inset). The cell in oil flows off the microstructure and into the RT chamber (middle inset). The bottom inset shows the multiple layers of the microfluidic chip. The RT chambers are in the COC layer.

Figure 2.

Digital lysates calculated with increasing single-cell transcriptomes converge with a physical lysate. Digital lysates are the mathematical averages of different combinations of single-cell transcriptomes. The correlations between H9 digital lysates and physical lysate increase when digital lysates are generated from 2 to 35 single-cell transcriptomes (black dots). The correlation stabilizes at $r = 0.93$ when the digital lysate includes 22 or more cells. The correlations between H9 digital lysates and unrelated physical lysate also increase with cell numbers in digital lysate calculation, but, as expected, are much lower and stabilize at 0.64 for kidney cancer (green dots) and at 0.48 for breast cancer (red dots) when 22 or more cells are included. Blue and black dotted lines show the correlations of physical lysates for H9 vs kidney cancer and H9 vs breast cancer, respectively. The correlations between physical lysates agree with those calculated by digital lysates.

Figure 3.

Clustering single-cell transcriptomes by similarity reconstructs stepwise cell cycle events. Single-cell transcriptomes were obtained from randomly selected single fucci cells using a microfluidic device. (A) Weighted gene co-expression network analysis (WGCNA) and a shortest Hamiltonian path algorithm were used to organize the transcriptomes into a circular map based on similarity. The distance between two cells represents their similarity. Digital lysate indicates that 15 or more cells represent the cell population (inset). Arrows indicate two pairs of cells with very similar expression profiles. (B) As shown in the center inset, the fucci cells are genetically modified to express cell cycle-specific colors (red: G1 phase with Cdt1-RFP; green: G2 phase with Geminin-GFP). The color of each cell was recorded by fluorescent imaging before processing with the microfluidic device. When the color of each cell was matched to the corresponding ID and plotted, most cells in the same cell cycle phase clustered together, as expected. The p -value ($P = 0.0031$) denotes the empirically computed probability of the observed ordering of cells due to chance and indicates that the similarity clustering agrees with the fluorescent imaging data.

Figure 4.

Sequential perturbations of cell-cycle-specific genes. After organizing single-cell transcriptomes by similarity into a sequencing order, expression levels of various cell-cycle-specific genes were plotted to visualize the sequential perturbation of individual genes during the cell cycle. Expression levels were scaled from 0 (undetectable) to 1 (maximum expression). Cell cycle phases were defined and colored based on the cell cycle molecular map. As expected, G0/G1-specific genes had higher expression levels in the G0/G1 phase (left panel), and an S-phase specific gene was mainly expressed within the S phase (middle panel). G2/M-specific genes had high expression levels in the G2/M phase and the early G0/G1 phase (right panel). The sequential expression order suggests that mRNAs of many G2/M-specific genes are not degraded until the early G0/G1 phase, after cell division.

Figure 5.

Perturbation of the cell cycle by *DNMT3B* knockdown. *DNMT3B* was suppressed with siRNA and the numbers of cells in different cell cycle stages were compared to cells treated with control siRNA. The distribution of cells in each cell cycle phase was altered by siRNA silencing of *DNMT3B* at 6, 24, and 48 h post-siRNA suppression. There were significantly more cells in the transition stages of S phase (yellow) at all time points, and there were fewer cells in G1 phase (red) at 6 and 24 h. There were fewer cells in M phase (colorless) at 6 and 48 h post-knockdown. Solid columns: control siRNA; Striped columns: *DNMT3B*

knock-down. Two representative fluorescent images are presented (bottom panel) to show alternation of cells at different cell cycle stages at 24 h.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript