Optofluidic real-time cell sorter for longitudinal CTC studies in mouse models of cancer

Bashar Hamza[a,b,1], Sheng Rong Ng[b,c,1], Sanjay M. Prakadan[b,d,e,f,g,1], Francisco Feijó Delgado[b,h], Christopher R. Chin[b], Emily M. King[b], Lucy F. Yang[b,h], Shawn M. Davidson[b,c], Kelsey L. DeGouveia[b,i], Nathan Cermak[b,j], Andrew W. Navia[b,d,e,f,g], Peter S. Winter[b,d,e,f,g], Riley S. Drake[b,d,e,f,g], Tuomas Tammela[b], Carman Man-Chung Li[b,c], Thales Papagiannakopoulos[b], Alejandro J. Gupta[b,d,e,f,g], Josephine Shaw Bagnall[b,h], Scott M. Knudsen[b], Matthew G. Vander Heiden[b,c,f], Steven C. Wasserman[h], Tyler Jacks[b,c,f,k,2], Alex K. Shalek[b,d,e,f,g,l,m,2], and Scott R. Manalis[b,h,n,2]

[a]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142; [c]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; [d]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [f]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [g]Ragon Institute of MGH, MIT and Harvard University, Cambridge, MA 02139; [h]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [i]Department of Biomedical Engineering, Wentworth Institute of Technology, Boston, MA 02115; [j]Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; [k]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; [l]Harvard–MIT Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology, Cambridge, MA 02139; [m]Department of Immunology, Massachusetts General Hospital, Boston, MA 02114; and [n]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Tyler Jacks, December 2, 2018 (sent for review August 15, 2018; reviewed by Daniel Chiu, Caroline Dive, and Klaus Pantel)

Circulating tumor cells (CTCs) play a fundamental role in cancer progression. However, in mice, limited blood volume and the rarity of CTCs in the bloodstream preclude longitudinal, in-depth studies of these cells using existing liquid biopsy techniques. Here, we present an optofluidic system that continuously collects fluorescently labeled CTCs from a genetically engineered mouse model (GEMM) for several hours per day over multiple days or weeks. The system is based on a microfluidic cell sorting chip connected serially to an unanesthetized mouse via an implanted arteriovenous shunt. Pneumatically controlled microfluidic valves capture CTCs as they flow through the device, and CTC-depleted blood is returned back to the mouse via the shunt. To demonstrate the utility of our system, we profile CTCs isolated longitudinally from animals over 4 days of treatment with the BET inhibitor JQ1 using single-cell RNA sequencing (scRNA-Seq) and show that our approach eliminates potential biases driven by intermouse heterogeneity that can occur when CTCs are collected across different mice. The CTC isolation and sorting technology presented here provides a research tool to help reveal details of how CTCs evolve over time, allowing studies to credential changes in CTCs as biomarkers of drug response and facilitating future studies to understand the role of CTCs in metastasis.

microfluidic | GEMM | circulating tumor cells | metastasis | single-cell RNA-Seq

Circulating tumor cells (CTCs) are an intermediate in the hematogenous spread of tumors during metastasis (1). Given their accessibility and potential prognostic and diagnostic value, CTCs have been the focus of significant clinical research efforts monitoring response to therapy and predicting risk of relapse (2–4). Over the past decade, novel microfluidic liquid biopsy-based techniques, as well as in vivo, vein catheter-based methods, have been developed to detect and collect CTCs directly from the blood of human patients (2–10). Combined with recently developed single-cell profiling methods, such as single-cell RNA sequencing (scRNA-Seq) (11–14), in-depth examination of CTCs is now possible. Such studies can provide new insights into the genomic properties of CTCs, as well as their relationship to matched primary and metastatic tumors (3, 4, 15–18).

Genetically engineered mouse models (GEMMs) of cancer, which mimic the natural multistage evolution of their human counterparts, facilitate characterization of both acute perturbations (e.g., drug treatment) and long-term phenotypic changes (e.g., tumor evolution) not possible in human subjects. However, despite the usefulness of GEMMs in cancer research, the combination of the small total murine blood volume (~1.5 mL) and the rarity of CTCs in circulating blood (fewer than 100 cells

Significance

Despite the usefulness of genetically engineered mouse models in cancer research, their small total blood volume and the rarity of circulating tumor cells (CTCs) preclude the use of existing liquid biopsy techniques for longitudinal CTC studies in mice. We have devised a method for collecting CTCs from an unanesthetized mouse longitudinally, spanning multiple days or weeks, to study acute perturbations (e.g., drug treatment) or potentially long-term phenotypes (e.g., tumor progression) within the same mouse. Here, we show that our optofluidic-based approach eliminates confounding biases driven by intermouse heterogeneity that can occur when CTCs are collected across different mice.

Author contributions: B.H., S.R.N., S.M.P., F.F.D., C.R.C., S.M.D., N.C., T.T., C.M.-C.L., T.P., J.S.B., S.M.K., M.G.V.H., S.C.W., T.J., A.K.S., and S.R.M. designed research; B.H., S.R.N., S.M.P., F.F.D., C.R.C., E.M.K., L.F.Y., S.M.D., K.L.D., A.W.N., R.S.D., and A.J.G. performed research; B.H., S.R.N., S.M.P., and F.F.D. contributed new reagents/analytic tools; B.H. and F.F.D. built the system and performed computer simulations; S.R.N. generated mice and performed mouse viral infection and in vivo bioluminescence imaging; S.M.P., A.W.N. and A.J.G. performed RNA sequencing; S.M.P. performed statistical analysis on the RNA-sequencing data; C.R.C., E.M.K., and S.M.D. performed mouse arteriovenous surgery; K.L.D. processed terminal mouse blood and tumor tissue; N.C. provided crucial input for the design of the system; P.S.W. assisted S.M.P. with the RNA-sequencing analysis; R.S.D. performed immunofluorescence staining; and B.H., S.R.N., S.M.P., P.S.W., T.J., A.K.S., and S.R.M. analyzed data; B.H., S.R.N., S.M.P., T.J., A.K.S., and S.R.M. wrote the paper.

Reviewers: D.C., University of Washington; C.D., University of Manchester; and K.P., UKE Hamburg.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE122233).

[1]B.H., S.R.N., and S.M.P contributed equally to this work.

[2]To whom correspondence may be addressed. Email: tjacks@mit.edu, shalek@mit.edu, or srm@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1814102116/-/DCSupplemental.

Published online January 23, 2019.

per milliliter) (3, 19) precludes the use of existing liquid biopsy techniques for longitudinal CTC studies in mice. When repeated blood samples are required at short intervals, a maximum of 1.0% of an animal's total blood volume can be removed every 24 h (~16.5 μL for a 25-g mouse) (20), a miniscule volume that does not yield a sufficient sample of CTCs for analysis.

GEMMs have been developed that combine genetic perturbations (manipulation of oncogenes and tumor suppressor genes) with genetically encoded fluorescent markers, enabling the unbiased detection and isolation of CTCs from the bloodstream. In vivo flow cytometry techniques have been used to enumerate CTCs in ear capillaries or tail veins longitudinally without euthanizing animals (21–23), but these techniques do not permit isolation and downstream molecular characterization of CTCs. On the other hand, in vivo, vein catheter-based techniques allow for the direct capture and isolation of CTCs from much larger blood volumes in real time (7, 8). However, these techniques are currently limited to detecting epithelial cell adhesion molecule (EpCAM)-expressing CTCs, which may result in only a subpopulation of CTCs being detected and isolated (24).

Results

Optofluidic Platform Design and Characterization. To enable longitudinal, in-depth studies of CTC biology in GEMMs and other murine cancer models, we have developed an optofluidic system capable of detecting and capturing fluorescent CTCs in living mice over several hours, days, or weeks. Key components of the system include a polydimethylsiloxane-based microfluidic CTC sorter chip, a fluorescence detector, and computer-controlled pneumatic valves (Fig. 1 *A* and *B*). A cannulated mouse with two permanent catheters easily accessible on its back allows for continuous blood withdrawal from the left carotid artery and return through the right jugular vein. Blood flows at a rate of 30 μL·min^{-1} into the CTC sorter chip. Two closely spaced laser beam lines illuminate the main flow channel of the chip. As such, each fluorescent cell that passes through the device emits two pulses of light, which are detected by a photomultiplier tube (Fig. 1*C* and *SI Appendix*, Fig. S1). The second laser line allows the controller to compute the velocity of the cells, which is essential to ensure reliable CTC capture. Similar to the ensemble-decision aliquot ranking technique for sorting CTC-containing aliquots of blood (9, 10), upon fluorescent cell detection, the controller

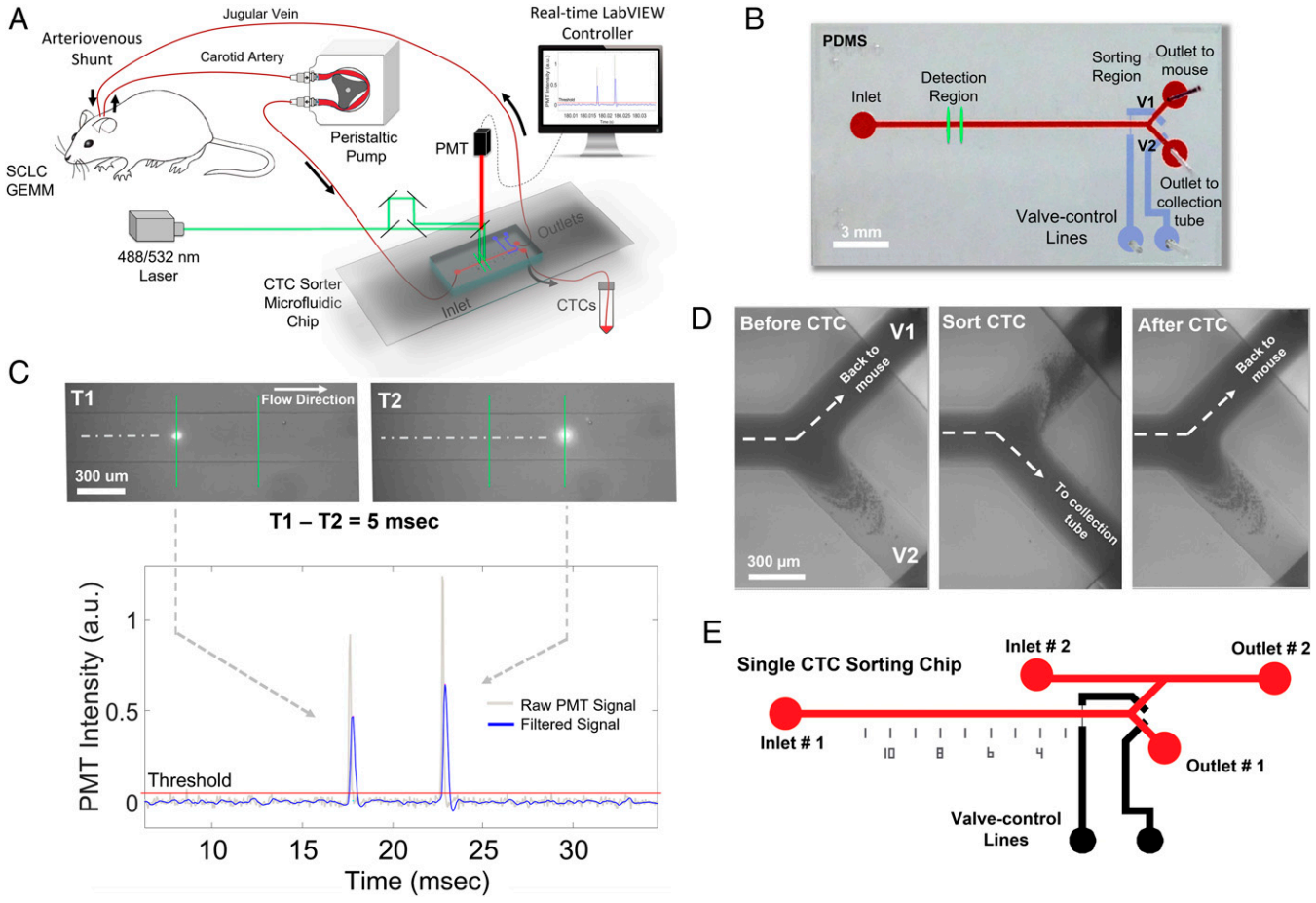

Fig. 1. Microfluidic sorter for longitudinal CTC studies in GEMMs. (*A*) Peristaltic pump withdraws blood from a surgically implanted cannula in the carotid artery of a mouse at a flow rate of 30 μL·min^{-1}. The blood is directed into the main flow channel of the CTC sorter chip. For tdTomato-positive cells, a green (532-nm) laser illuminates two points along the main flow channel of the CTC chip separated by a known distance. Thus, fluorescent CTCs emit two red-shifted pulses of light, which are detected by a photomultiplier tube (PMT). Based on the timing of the pulses, a LabVIEW program computes the velocity of the cells and operates computer-controlled pneumatic valves to redirect fluorescent CTCs toward a collection tube. After exiting the chip, CTC-depleted blood returns to the jugular vein of the mouse via a second surgically implanted cannula. (*B*) Top-view image of the CTC sorter microfluidic chip showing the inlet, outlets, and valve actuation lines (V1 and V2). (*C*) Illustration of the CTC detection mechanism using the two excitation laser lines. A low-pass filter is applied to the raw data for determining true peaks. (*D*) Outlet by which blood is returned to the mouse is briefly sealed while the opposite outlet is opened to allow for CTC isolation in real time. (*E*) After collection, CTCs are further enriched by a secondary CTC sorting chip designed with a parallel channel to flush CTCs into wells containing cell lysis buffer for downstream scRNA-Seq.

instantly operates pneumatic valves (25) to redirect a small blood volume that includes the CTC toward a collection tube (mean ± SD = 127 ± 47 nL per sort event; Fig. 1*D* and *SI Appendix*, *Supplementary Information Text* and Fig. S2). Blood from the collection tube can then be further enriched for CTCs and run through a secondary single-CTC sorting chip for downstream characterization using techniques such as scRNA-Seq (Fig. 1*E*).

Validation of the Platform's Detection Limits with Fluorescent Microbeads and Cells. To ascertain the detection limit of our CTC sorter, we passed a sample of healthy mouse blood spiked with flow cytometry calibration beads through the system. The reference beads comprised five fluorescence intensity groups, including one with zero fluorescence. The system consistently detected the two brightest fluorescence levels (peaks 4 and 5) and approximately the brightest 30% of level 3 (peak 3) (*SI Appendix*, Fig. S3 *A–C*). This sensitivity was sufficient to detect nearly the entire population of tdTomato-expressing murine small cell lung cancer (SCLC) cells spiked in healthy mouse blood. We also tested blood isolated from autochthonous SCLC tumor-bearing mice that exhibit metastasis to distant organs in a pattern similar to metastatic spread in human patients (26). Tumors in these mice were initiated by Cre-mediated deletion of tumor suppressor genes *Trp53*, *Rb1*, and *Pten* in the murine lung epithelium (26). This GEMM also includes a Cre-activated tdTomato allele (27) that engenders fluorescence in all tumor cells after tumor initiation, including CTCs. The majority of CTCs from the blood of SCLC tumor-bearing mice were above the detection threshold (*SI Appendix*, *Supplementary Information Text* and Fig. S3*D*).

Validation of Platform's Sorting Functionality. After establishing that the sensitivity of the optical detector was sufficient, we characterized and optimized the sorting efficacy using blood samples from healthy mice spiked with low concentrations of tdTomato-expressing murine SCLC cells. In samples containing 100 cells per milliliter or more, over 80% of detected tdTomato-positive cells were successfully captured. For samples with only 10 cells spiked into 500 μL of healthy mouse blood, the sorted sample contained 6.0 ± 0.7 cells (mean ± SD, $n = 3$ repeats). Applying a slight delay in actuating the pneumatic valves until the cell has moved closer to the sorting region decreased the collected blood volume per CTC to 76 ± 28 nL (mean ± SD) without compromising the capture efficiency. At this volume, on the order of 700 neighboring white blood cells (WBCs) and over 70,000 red blood cells (RBCs) and platelets in the bloodstream are collected in addition to the target CTC on each valve actuation. These experiments demonstrate that the CTC sorter is capable of isolating fluorescent CTCs from blood even at very low concentrations (*SI Appendix*, Fig. S4).

Longitudinal CTC Collection from SCLC Tumor-Bearing Mice. Next, we conducted a longitudinal study of CTCs collected from autochthonous SCLC tumor-bearing mice treated with the BET bromodomain inhibitor JQ1, which has been demonstrated to have antiproliferative effects in SCLC (28–30). CTCs were isolated from mice over a 2-h period before treatment (0 h) and at 24-h intervals following treatment initiation, continuing over 96 h (*SI Appendix*, *Supplementary Information Text* and Figs. S5 and S6*A*). CTCs were enriched from the samples by RBC lysis, followed by WBC depletion, using magnetic-activated cell sorting (*SI Appendix*, Fig. S6*B*) and, finally, by passing through a secondary, single-cell CTC sorting chip. Enriched CTCs were then processed using Smart-Seq2 (31) (Fig. 1*E* and *SI Appendix*, *Supplementary Information Text* and Fig. S6*C*). Cells with insufficient gene complexity for downstream analysis after scRNA-Seq were eliminated computationally, in addition to cells with high expression of immune and platelet signature genes (3, 4, 16) (*SI Appendix*, *Supplementary Information Text*). The overall yields of the process (from blood to a successful scRNA-Seq library) were 11.5% and 5.3% for samples from treated and untreated mice, respectively (median values with a range of 7.4–31% for treated samples and a range of 3.3–6.7% for untreated samples; *SI Appendix*, Fig. S7).

Analysis of Single-CTC Transcriptomes Across Different Mice and Within Each Mouse. We then examined our data to determine how the information collected longitudinally from the same mouse with our system compared with the common approach of capturing CTCs across different mice using asynchronous terminal bleeds (16, 19). To analyze our longitudinal CTC data, we pooled our collected CTC transcriptomes across all mice, performed a principal component analysis (PCA) visualized by t-distributed stochastic neighbor embedding (tSNE) (12, 32, 33), and identified clusters (using k-nearest neighbors clustering) over the significant principal components (PCs) (14, 34) (Fig. 2 *A–C* and *SI Appendix*, *Supplementary Information Text*). This unsupervised analysis revealed that mouse of origin contributed significantly to the variation observed in our dataset, with cluster representation driven primarily by individual mice (3, 4, 6) (Fig. 2*C*). We next performed PCA on CTCs collected from each mouse individually. Here, we found that PC1 significantly correlated (Spearman correlation) with time since treatment ($P < 0.05$, Student's t test following a Lilliefors test for normality) when independently calculated for each of the treated mice but not for either control (Fig. 2 *D* and *E* and *SI Appendix*, Fig. S8). This suggests that by isolating CTCs from the same animal longitudinally, we are able to eliminate potentially confounding differences between animals that could otherwise mask biologically relevant gene expression changes that occur over time.

In comparison, the conventional approach for performing a longitudinal CTC analysis would be to begin the experiment with a cohort of mice and obtain terminal bleeds from a subset at each time point. We simulated this approach from our measurements by selecting a different treated mouse to represent each of the 0-h, 48-h, and 96-h time points (*SI Appendix*, Fig. S9); here, regardless of which mouse was chosen to represent which time point, we found that the mean PC1 coordinate of treated mouse 1 existed outside the interquartile range of the other mice, suggesting a consistent mouse-specific effect that dominates the first PC (Fig. 2*F* and *SI Appendix*, Fig. S10). As such, conclusions drawn from analysis of CTCs from terminal bleeds at different time points across mice would be confounded by organism-specific features from the different mice.

Supervised Analysis of Single-CTC Transcriptomes. To more formally examine treatment-induced shifts in gene expression, we calculated differential expression across all pairs of time points within each longitudinally profiled mouse (14, 35) and, once again, simulated terminal bleed data (*SI Appendix*, *Supplementary Information Text* and Fig. S9). Our analyses showed that the majority of the differentially expressed genes within each mouse (per mouse) were unique (*SI Appendix*, Fig. S11). Furthermore, each per-mouse differentially expressed gene set shared little overlap ($P < 0.05$, hypergeometric test) with those calculated from mock terminal bleed datasets (*SI Appendix*, Fig. S12), regardless of the chosen mouse for the different time points. At each time point, differentially expressed genes in the mock terminal bleed data were enriched for several functional processes, such as mitochondrial function, cellular organization, and metabolism (36, 37); however, upon further inspection of the different mock terminal bleed permutations, we found that these enrichments were linked primarily to mouse rather than time point (Fig. 2*F* and *SI Appendix*, Figs. S10 and S12; e.g., housekeeping genes or ribosome), suggesting confounding mouse-to-mouse heterogeneity. This is evocative of the marked inter- and

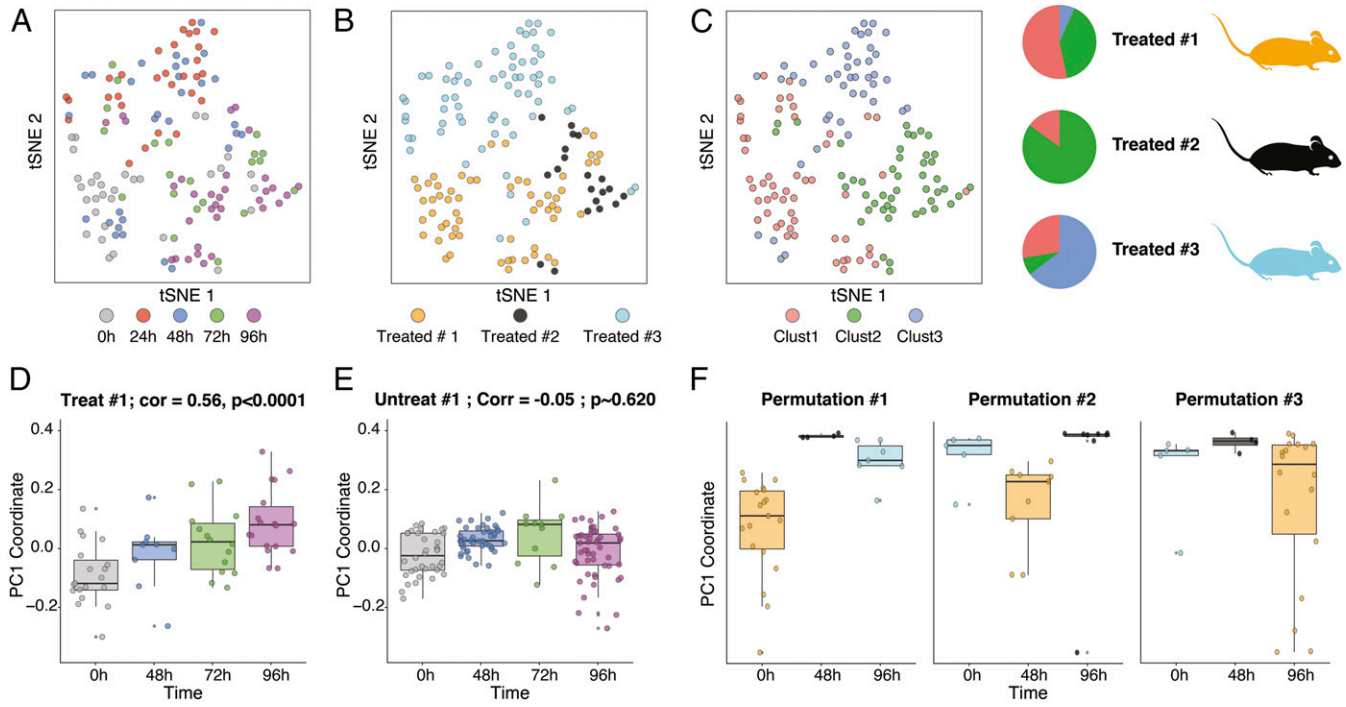

Fig. 2. scRNA-Seq of captured CTCs demonstrates the utility of intramouse CTC profiling. The tSNE of all CTCs collected across three JQ1-treated mice is colored by time point posttreatment (*A*), mouse (*B*), and cluster of assignment based on *k*-nearest neighbors clustering (*C*). (*Top Right*) Pie charts show the fractional representation of each cluster in each treated mouse. Boxplots of the first PC of CTC transcriptomes from PCAs were obtained from longitudinally following the same treated mouse [*D*, correlation (Corr) = 0.56] or untreated mouse (*E*, Corr = −0.05). Each point represents a CTC. (*F*) Boxplots of the first PC from three different "mock terminal bleed" permutations across three treated mice (*SI Appendix, Supplementary Information Text* and Fig. S9).

intrapatient heterogeneity observed in CTCs longitudinally collected from human patients (5–7), and suggests the importance of examining the same mouse over time.

Discussion

The platform outlined here represents an important advancement in the detection and continuous capture of single CTCs from the same mouse over time. Our method enables CTCs to be isolated in low blood volumes and prepares them for downstream characterization. Here, we used scRNA-Seq to show that continuous CTC capture from the same mouse eliminates biases driven by intermouse heterogeneity that can occur when CTCs are collected across different mice. Although future work will be needed to elucidate the underlying drivers of this variability, given the baseline genetic homogeneity of the animals used to generate our GEMM, one potential explanation could be underlying differences in the cellular composition of the primary tumors across different animals. scRNA-Seq results from the primary tumor samples harvested from each animal after terminal CTC collection (96 h) are consistent with this hypothesis (Fig. 3 *A* and *B* and *SI Appendix, Supplementary Information Text*). These data suggest that primary tumors from each mouse

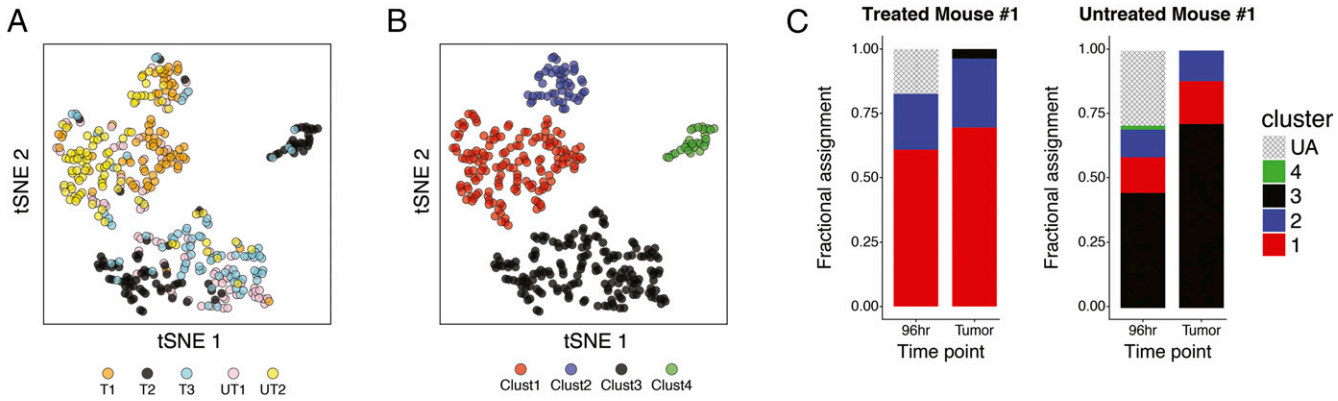

Fig. 3. scRNA-Seq of end-point primary tumors demonstrates heterogeneity in phenotype. (*A*) The tSNE of primary tumor cells across treated and untreated mice, colored by mouse called from *k*-nearest neighbors clustering. (*B*) tSNE of primary tumor cells across treated and untreated mice, colored by clusters. (*C*) Computational cluster assignments (*SI Appendix, Supplementary Information Text*) for 96-h CTCs next to their matched primary for a representative treated mouse and untreated mouse plotted as bar plots ($n = 18$ and $n = 82$ cells for treated mouse 1 96-h CTCs and tumor cells, respectively; $n = 52$ and $n = 84$ cells for untreated mouse 1 96-h CTCs and tumor cells, respectively). UA, unassigned. Neither pairing is significantly different ($P = 0.99$ and $P = 0.66$ for treated mouse 1 and untreated mouse 1, respectively, by Fisher's exact test).

may contain multiple malignant gene expression states (32, 33), which appear to be shared across mice. Although some of these differences could be attributed to the presence of multiple independently evolving primary tumors within each mouse, the fact that each gene expression state is composed of cells from multiple mice suggests that these states may be a shared feature of our SCLC GEMM, although future experiments will be needed to robustly validate this finding. Intriguingly, we were able to computationally match ∼67% (62 cells) of the terminally collected CTCs (96 h, $n = 92$ cells from five mice) to one of these shared states (*SI Appendix*, *Supplementary Information Text* and Fig. S13). Furthermore, we observed comparable state frequencies between a mouse's CTCs and its corresponding primary tumor sample at the terminal time point ($P > 0.5$, Fisher's exact test; Fig. 3*C*). We note that mice with fewer than 10 CTCs, the statistical power in our comparison was limited in some cases (*SI Appendix*, Fig. S14). While further experimentation will be needed to corroborate this preliminary finding, our data demonstrate the potential value of having matched primary tumor samples as a reference in mouse models of cancer, and that analysis of CTCs in our murine model of SCLC may reveal similar biology to primary tumors from the same mouse, suggesting their utility as a surrogate for matched tumors under specific circumstances.

Future work of this kind has the potential to shed new light on the relationship between CTCs, primary tumors, and metastases, allowing for the exploration of their utility as biomarkers and facilitating examination of how individual CTCs contribute to metastasis. Moreover, they may help elucidate the features that inform shifts observed upon perturbation, such as drug treatment. Ultimately, these data show that our platform opens the door for novel CTC experimentation, such as examining longitudinal drug responses and comparing CTCs with primary tumors (shown here), characterizing their relationship to metastases, and measuring the rate of CTC production in an acute window. With additional development, our device could enable longitudinal studies in mice to find associations between individual CTCs and clusters of CTCs, profile rare immune cells (e.g., using genetic reporters or based on tetramer staining), monitor mesenchymal cells in a variety of contexts (including wound healing and tumor formation), and measure induction rates of drugs or nanoparticles in circulating mononuclear cells.

Materials and Methods

All RNA-sequencing data generated in this study (raw data and processed data matrices) have been deposited into the Gene Expression Omnibus database hosted at the National Center for Biotechnology Information under the accession code GSE122233. Information on mouse models, cell culture, shunt surgery, optofluidic platform design and fabrication, real-time data processing and analysis, CTC and tumor cell processing and enrichment, single-cell RNA-sequencing sample preparation, and data analysis is available in *SI Appendix*. All animal-based procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care, Division of Comparative Medicine.

ACKNOWLEDGMENTS. We thank R. Kimmerling, S. Olcum, M. Stevens, B. A. Goods, A. S. Genshaft, C. G. K. Ziegler, N. Calistri, L. Atta, and M. B. Cole for helpful discussions. We also thank Samira Daswani for technical assistance. This work was supported, in part, by the Thomas and Sarah Kailath Fellowship (to B.H.); the A*STAR (Agency for Science, Technology and Research, Singapore) National Science Scholarship (to S.R.N.); the Lustgarten Foundation (M.G.V.H.); the Ludwig Center at MIT (S.R.M., T.J., and M.G.V.H.); Stand Up To Cancer (M.G.V.H.); the Howard Hughes Medical Institute (HHMI) Faculty Scholars Award (to M.G.V.H.); the HHMI Investigator Program (T.J.); the Searle Scholars Program (A.K.S.); the Beckman Young Investigator Program (A.K.S.); NIH Grant 1R01 CA184956 (to S.R.M. and T.J.); NIH New Innovator Award 1DP2GM119419 (to A.K.S.); the Pew-Stewart Scholars (A.K.S.), a Sloan Fellowship in Chemistry (to A.K.S.); NIH Grant 5U24AI118672 (to A.K.S.); NIH Grant 1U54CA217377 (to A.K.S. and S.R.M.); NIH Grant 1R33CA202820 (to A.K.S.); NIH Grant 2U19AI089992 (to A.K.S.); NIH Grant 1R01HL134539 (to A.K.S.); NIH Grant 2RM1HG006193 (to A.K.S.); NIH Grant 2P01AI039671 (to A.K.S.); and Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute.

1. Pantel K, Speicher MR (2016) The biology of circulating tumor cells. *Oncogene* 35:1216–1224.
2. Yu M, Stott S, Toner M, Maheswaran S, Haber DA (2011) Circulating tumor cells: Approaches to isolation and characterization. *J Cell Biol* 192:373–382.
3. Miyamoto DT, et al. (2015) RNA-seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* 349:1351–1356.
4. Ting DT, et al. (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep* 8:1905–1918.
5. Ozkumur E, et al. (2013) Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 5:179ra47.
6. Sarioglu AF, et al. (2015) A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat Methods* 12:685–691.
7. Gorges TM, et al. (2016) Enumeration and molecular characterization of tumor cells in lung cancer patients using a novel in vivo device for capturing circulating tumor cells. *Clin Cancer Res* 22:2197–2206.
8. Vermesh O, et al. (2018) An intravascular magnetic wire for the high-throughput retrieval of circulating tumour cells in vivo. *Nat Biomed Eng* 2:696–705.
9. Schiro PG, et al. (2012) Sensitive and high-throughput isolation of rare cells from peripheral blood with ensemble-decision aliquot ranking. *Angew Chem Int Ed Engl* 51:4618–4622.
10. Johnson ES, Anand RK, Chiu DT (2015) Improved detection by ensemble-decision aliquot ranking of circulating tumor cells with low numbers of a targeted surface antigen. *Anal Chem* 87:9389–9395.
11. Zeisel A, et al. (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347:1138–1142.
12. Shalek AK, et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498:236–240.
13. Macosko EZ, et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
14. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33:495–502.
15. Lohr JG, et al. (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* 32:479–484.
16. Aceto N, et al. (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158:1110–1122.
17. Vishnoi M, et al. (2015) The isolation and characterization of CTC subsets related to breast cancer dormancy. *Sci Rep* 5:17533.
18. Alix-Panabières C, Pantel K (2016) Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. *Cancer Discov* 6:479–491.
19. Rhim AD, et al. (2012) EMT and dissemination precede pancreatic tumor formation. *Cell* 148:349–361.
20. Parasuraman S, Raveendran R, Kesavan R (2010) Blood sample collection in small laboratory animals. *J Pharmacol Pharmacother* 1:87–93.
21. Georgakoudi I, et al. (2004) In vivo flow cytometry: A new method for enumerating circulating cancer cells. *Cancer Res* 64:5044–5047.
22. Zettergren E, et al. (2012) Instrument for fluorescence sensing of circulating cells with diffuse light in mice in vivo. *J Biomed Opt* 17:037001.
23. Nedosekin DA, Verkhusha VV, Melerzanov AV, Zharov VP, Galanzha EI (2014) In vivo photoswitchable flow cytometry for direct tracking of single circulating tumor cells. *Chem Biol* 21:792–801.
24. Chudziak J, et al. (2016) Clinical evaluation of a novel microfluidic device for epitope-independent enrichment of circulating tumour cells in patients with small cell lung cancer. *Analyst* 141:669–678.
25. Unger MA, Chou HP, Thorsen T, Scherer A, Quake SR (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* 288:113–116.
26. McFadden DG, et al. (2014) Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 156:1298–1311.
27. Madisen L, et al. (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* 13:133–140.
28. Lenhart R, et al. (2015) Sensitivity of small cell lung cancer to BET inhibition is mediated by regulation of ASCL1 gene expression. *Mol Cancer Ther* 14:2167–2174.
29. Kato F, et al. (2016) MYCL is a target of a BET bromodomain inhibitor, JQ1, on growth suppression efficacy in small cell lung cancer cells. *Oncotarget* 7:77378–77388.
30. Jahchan NS, et al. (2016) Identification and targeting of long-term tumor-propagating cells in small cell lung cancer. *Cell Rep* 16:644–656.
31. Picelli S, et al. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9:171–181.
32. Patel AP, et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344:1396–1401.
33. Tirosh I, et al. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352:189–196.
34. Chung NC, Storey JD (2015) Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* 31:545–554.
35. McDavid A, et al. (2013) Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* 29:461–467.
36. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
37. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550.

PNAS
www.pnas.org

Supplementary Information for

Optofluidic real-time cell sorter for longitudinal CTC studies in mouse models of cancer

Bashar Hamza[1], Sheng Rong Ng[1], Sanjay M. Prakadan[1], Francisco Feijó Delgado, Christopher R. Chin, Emily M. King, Lucy F. Yang, Shawn M. Davidson, Kelsey L. DeGouveia, Nathan Cermak, Andrew W. Navia, Peter S. Winter, Riley S. Drake, Tuomas Tammela, Carman Man-Chung Li, Thales Papagiannakopoulos, Alejandro J. Gupta, Josephine Shaw Bagnall, Scott M. Knudsen, Matthew G. Vander Heiden, Steven C. Wasserman, Tyler Jacks[2], Alex K. Shalek[2], and Scott R. Manalis[2]

[1]B.H., SR.N., and S.M.P contributed equally to this work.
[2]To whom correspondence may be addressed: T.J. (tjacks@mit.edu), A.K.S. (shalek@mit.edu), or S.R.M. (srm@mit.edu).

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1814102116

Supplementary Information Text

Mouse model and cell lines. The *Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Pten$^{fl/fl}$; Rosa26$^{LSL\text{-}Luciferase/LSL\text{-}Luciferase}$* (PRP-L/L) mouse model of SCLC has been described previously (1). *Rosa26$^{LSL\text{-}tdTomato/LSL\text{-}tdTomato}$* mice were obtained from Jackson Laboratories (*Gt(ROSA)26Sor$^{tm14(CAG\text{-}tdTomato)Hze}$*) and crossed into the PRP-L/L model to obtain *Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Pten$^{fl/fl}$; Rosa26$^{LSL\text{-}tdTomato/LSL\text{-}Luciferase}$* mice. Tumors were initiated by intratracheal delivery of 2×10^8 plaque-forming units (p.f.u.) of adenovirus expressing Cre recombinase under the control of a CGRP promoter (Ad5-CGRP-Cre (2)), as previously described (3). Adenoviral stocks were purchased from the Viral Vector Core Facility at the University of Iowa Carver College of Medicine. Candidates for arteriovenous shunt surgery were identified by *in vivo* bioluminescence imaging using the IVIS Spectrum In Vivo Imaging System (PerkinElmer).

Murine SCLC cell lines (AF1281-M1, AF3291LN) were generated from mSCLC tumors isolated from *Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Rosa26$^{LSL\text{-}tdTomato/+}$* or *Trp53$^{fl/fl}$; Rb1$^{fl/fl}$; Pten$^{fl/fl}$; Rosa26$^{LSL\text{-}tdTomato/LSL\text{-}Luciferase}$* mice as previously described (4).

Shunt Surgery. All animal-based procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care (CAC), Division of Comparative Medicine (DCM). The arteriovenous shunt method was validated as an approach for real-time manipulation of nutrient levels in the serum of anesthetized mice (5–7). We adopted the same technique for continuous sampling of mouse blood for CTCs. Briefly, catheters are inserted into the right jugular vein and the left carotid artery and are externalized using standard cannulation surgical techniques. A peristaltic pump (Instech Laboratories Inc., Plymoth Meeting, PA, USA) is then connected to the catheters for blood sampling and

return through the carotid artery and jugular vein, respectively, in the conscious mouse (**Fig. 1A**). During the four-day longitudinal studies, the total collected blood volume is monitored. If depleted blood volume exceeds 1% of the animal's body weight (for example 260 µL for a 25g mouse), per MIT DCM guidelines, healthy-mouse donor blood (of same strain and sex) is infused directly into the mouse using the jugular vein catheter, equivalent to the amount removed.

Optical Detection Platform. The optical system is comprised of two optical trains, making up two compact vertical microscopes. The top optical train divides the laser beam (OBIS 532 LS, Coherent Inc) into two separate beams that are focused along one axis to produce two illumination lines projected at the sample focal plane (perpendicular to the blood flow channel) for precise velocity measurements of the flowing CTCs (**Fig. S1A-B**). The laser passes through a line filter and polarizing beam splitters to generate the two beams with minimal losses. Next, the two laser lines pass through a cylindrical lens to focus the two beams into lines. The focused lines are then projected onto the microfluidic channel with a 4F optical system. The dichroic mirror and longpass filters, placed directly above the detection region, pass a filtered fluorescence signal to the PMT (Hamamatsu H10722-20) by blocking the 532 nm laser line signal with a notch filter (532 nm StopLine single-notch filter, Semrock). A 90:10 beam splitter is added before the PMT to allow for imaging of the illumination region for device alignment purposes.

The bottom optical train is similar in configuration to the top train and uses a second green laser to illuminate a circular region covering the valves, enabling fluorescent and bright field imaging for confirming successful CTC sorting and valve functionality throughout the experiment.

Microfluidic Device Design and Fabrication. The microfluidic device for sorting minute amounts of blood containing single CTCs was designed to have one inlet to a 1000μm-long microfluidic channel that bifurcates into two channel outlets (90º apart, **Fig. 1B-C**); one for returning the blood to the mouse and the other for collecting the sorted CTC-containing blood sample. The fabrication was performed using standard soft lithographic techniques on two four-inch wafers. A single layer of photoresist (SU8 2050, Microchem, Newton, MA) was patterned to create the pneumatic channels on the valve control wafer. For the blood flow channel, AZ9260 positive resist (MicroChemicals) was exposed, developed, and then reflowed at 120˚C for 10 minutes to create the half ellipsoid channel profile necessary for a complete valve seal (**Fig. S2**) (8).

A mixture of PDMS (Polydimethylsiloxane) and its curing agent (SYLGARD 184 A/B, Dowcorning, Midland, MI, USA) at a 10:1 ratio was spun on top of the actuation wafer to a thickness of 50 μm and baked in an oven set to 65 ˚C for at least 3 hours. For the flow channel layer, the mixture was poured to a thickness of ~5 mm and cured at 65 ˚C for 3 hours. Afterwards, the flow channel layer was peeled off and punched with a 0.75 mm puncher (Harris Uni-Core, Ted Pella Inc., Reading, CA) to define the inlet and outlets to and from the channel, respectively, and diced to prepare for bonding. The flow channel devices and the actuation layers were then treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec). Next, the flow layer was aligned to the actuation layer and transferred to a hot plate set to 85 ˚C. After 15 minutes, the assembled PDMS layers were peeled off and punched with a 0.75 mm puncher to define inlets to the actuation channels. The assembled PDMS layers were treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec) for irreversible bonding to a glass slide (Fisherbrand 1x3”, Fisher Scientific, Pittsburgh, PA).

Prior to flow experiments, the device was aligned to project the two laser lines across the flow channel approximately 8 mm away from the valve actuation region. The device was then primed with Heparinized-Saline (diluted to 100 USP units per mL, NDC 25021-400-30) to prevent any clotting within the microfluidic channel.

Real-Time Data Processing and System Control. The PMT module generates an output voltage that is sampled by a NI USB-6009 (National Instruments) 14 bit analog-to-digital converter. This analog-to-digital converter is also used to output the control voltage signals for the peristaltic pump and the PMT gain. The PMT voltage is acquired and displayed in real time in LabVIEW (National Instruments) at 20,000 samples per second. The raw data is filtered with a low pass filter to remove spurious noise and the output of a running median filter is subtracted to remove low frequency drift or any DC offset. Upon transit of a fluorescent cell or particle, the PMT output signal passes a specific threshold, triggering the NI USB-9472 device to actuate the SMC Solenoid Valves (S070A-6DC, 8 ports) that are connected to an external 6-volt power supply.

During non-sorting operation of the system (i.e. cell counting experiments), the first valve (V1) region is kept sealed to force the blood to flow directly through to the opposite outlet and back into the jugular vein catheter of the mouse (Top-view image, **Fig. 1b**). When the sorting functionality is activated, and upon detection of a cell, its velocity is calculated (**Fig. S4A**) and used to estimate the time of its arrival at the valve region of the channel, all using a virtual instrument (VI) program in LabVIEW (**Fig. S4B**). V2 then closes immediately and V1 opens for an amount of time equal to twice the estimated time of arrival of the cell at the valve to deflect a 127±47 nL (or 76±28 nL with a valve actuation delay) blood bolus containing the cell (**Fig. S4C**). This amount of blood per single sort

provides an upper limit for the number of events we can collect from a mouse over the four-day study. For a 25g mouse, a maximum number of ~2,600 total events can be collected without the need for fluid replacement, per MIT DCM guidelines.

After CTC enrichment with RBC lysis and MACS (described below), the cells were sorted using our secondary single-CTC sorting chip. Upon detection of a single fluorescent CTC in the secondary single-cell sorting chip (**Fig. 1E**), micro-valves actuate to push the cell into the parallel channel. The micro-valves then seal the primary channel and sample flow is stopped to introduce fresh buffer into the parallel channel, releasing the cell out of the chip into the tubing and then into a collection well containing TCL buffer (QIAGEN) with 1% 2-mercaptoethanol for downstream scRNA-seq (**Fig. S6C**).

CTC enrichment by RBC lysis and MACS. Following CTC isolation, the collected blood sample is first treated with ACK Lysing Buffer (Gibco A10492-01) to remove RBCs, then rinsed, filtered using a 30μm Pre-Separation Filter (Miltenyi Biotec, #130-041-407), and processed through magnetic activated cell sorting (MACS) using mouse CD45 MicroBeads (Miltenyi #130-052-301) and MS columns (Miltenyi #130-042-201) to remove CD45-positive cells, according to the manufacturers' protocols. The final product is then diluted to a total volume of 2 mL of MACS buffer and flowed through the secondary single-CTC sorting chip.

Characterization with Cell Line and Beads. To validate the sensitivity of the system to detect fluorescent cells in blood, 5-peak FACS calibration beads (Sphero PE 5-peak, Spherotech Inc) were spiked into mouse blood and flowed through the device at 30 μL/min (**Fig. S3**). To validate the sorting functionality of the system, a tumor cell line (AF1281-M1) was established from a mouse with autochthonous small cell lung cancer expressing

a bright and stable cytoplasmic tdTomato fluorescent protein. Cells were then counted using Coulter Counter (Multisizer 4, Beckman-Coulter) and re-suspended in 500 µL of mouse blood at different densities (10, 100, 500, and 1000 cells, **Fig. S4**).

JQ1 Treatment of Tumor-Bearing Mice. Tumor-bearing SCLC mice were treated with 10 mg/kg JQ1 (Cayman Chemical) by intravenous injection daily for the duration of the study. JQ1 was dissolved in DMSO to make a 20 mg/mL stock, then diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline to obtain a working concentration of 2 mg/mL. Vehicle-treated control mice received an equivalent dose of DMSO diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline (**Fig. S5**).

Dissociation of Tumor Samples for Single-Cell RNA-Sequencing Analysis. Primary tumors from tumor-bearing animals were dissected, dissociated into single cells using a lung dissociation kit according to the manufacturer's protocol (Miltenyi Biotec #130-095-927), then stained with APC-conjugated antibodies against CD11b (eBioscience #17-0112-82), CD31 (BioLegend #102510), CD45 (eBioscience #17-0451-83) and TER-119 (BD Biosciences #557909). tdTomato-positive, APC-negative cells were single-cell sorted by FACS into TCL buffer (QIAGEN) containing 1% 2-mercaptoethanol, then frozen at -80ºC for downstream processing for RNA-seq.

Single-Cell RNA-Sequencing Sample Preparation. Both CTC samples and primary tumor samples collected in TCL buffer were processed through Smart-Seq2 as follows. Total nucleotide material from lysed single cells was extracted with RNA-clean AMPure nucleotide extraction beads (Beckman-Coulter) and washed with 80% ethanol before undergoing reverse transcription with Maxima enzyme (Thermoscientific), followed by PCR with a KAPA Hotstart Readymix 2x kit (KAPA biosystems). Following

quantification and quality control analysis by Qubit DNA quantification (Thermoscientific) and tape station (Agilent), whole transcriptome amplifications (WTAs) of each single cell were transformed into sequencing libraries with a Nextera XT kit (Illumina) and barcoded with unique 8-bp DNA barcodes. cDNA libraries were pooled, quantified, and sequenced on an Illumina NextSeq 500 to an average depth of 1.2M reads/CTC.

Analysis of Raw Sequencing Data. Following sequencing, BCL files were converted to merged, demultiplexed FASTQs. Paired-end reads were mapped to mm10 mouse transcriptome (UCSC) with Bowtie. Expression levels of genes were log-transformed transcript-per-million (TPM[i,j]) for gene i in sample j, estimated by RSEM in paired-end mode. For each cell, we enumerated genes for which at least one read was mapped, and the average expression level of a curated list of housekeeping genes. We excluded from analysis profiles with fewer than 500 detected genes or an average housekeeping expression below 0.5 $\log_2$(TPM).

Identification of leukocytes and correction of platelet effect. A matrix of TPM estimates for all genes across all cells resulted from preprocessing of RNA-Seq data. A raw principal component analysis was run for each sample, and the first PC separated cells of immune lineage from cells expressing epithelial genes. We selected from overlaps between this first PC and a curated set of immune-related genes (**Table S1**) to score CTCs according to their immune score. Cells with an average immune score below 0.5 $\log_2$(TPM) were taken forward for analysis of their transcriptomes as CTCs.

Likewise, a curated set of platelet-related genes (**Table S1**) was used to develop a platelet signature for each CTC transcriptome. The signal of this platelet signature was then regressed out using the RegressOut function in Seurat.

Principal Component Analysis of CTCs. We performed a gene selection based on a binning strategy across expression using Seurat to define a unique set of variable genes for each treated and untreated mouse, as well as combined sets of treated and untreated mice. Principal component analysis was performed over these variable gene sets for each case (per-mouse or combined). PCs were correlated to time using Spearman correlation. To test for significance of PC1 correlation with time, we permuted assignments of time (N = 1000) and calculated statistics for each set of CTCs per mouse. The sets of correlations for each mouse were tested with Lilliefors test for normality and used to determine PCs with significant correlations with time.

Generation of Mock Data. To simulate current methods of terminal bleed assays for CTC collection across an acute time scale, we generated a series of mock datasets using our true, continuously collected data. At each time point, the CTCs from a single treated mouse were selected without replacement. This process was exhaustively repeated five times to generate all unique treated mock data over which further differential expression analysis was performed.

Differential Expression Analysis. For each of the treated mice, as well as the mock data, we performed differential expression using Seurat's built-in single-cell differential expression tool, with a bimodal distribution model. Differential expression was performed between all pairs of time points available for each dataset. Genes with avg_diff > 1 and p-value < 0.01 were selected and visualized using the DoHeatmap function in Seurat.

Furthermore, these genes were enriched for upstream regulators using gene set enrichment analysis (GSEA) through the Broad Institute.

Analysis of Primary Tumor Data. Variable genes across primary tumor cells of all mice were calculated and principal component analysis was performed as described above. Relevant PCs were determined by visualizing percent variance explained in an elbow plot, significance by Jackstraw (9), and manual inspection of loadings and coordinates. Following inspection, PCs 1-5 were selected for downstream visualization of the primary tumor cells by tSNE in Seurat, with perplexity set to 15, and 2500 iterations run.

Assignment of CTCs to Primary Tumor Cluster. Differential expression, as reported above, was used to identify marker genes that describe the resultant clusters in the primary tumor. These marker genes were used to develop a signature score for each cluster for each of the 96-hr CTCs, using weighted averaging of the genes for each signature (**Table S2**). Next, we permuted random sets of genes with similarly binned expression distribution and size to create cluster-independent background scores for each CTC. CTCs with cluster-specific signatures above their cluster-independent background were "assigned" to that cluster. CTCs were assigned to all clusters which were above background – if a CTC scored above background for more than one cluster, it was assigned to all those clusters; if a CTC scored above background for no clusters, it was classified as "unassigned" (**Fig. S13**). The total number of CTCs per mouse for each cluster was visualized by a stacked bar plot, generated through ggplot2 in R. Distributions of assignment between 96-hr CTCs and primary tumors from mice were tested by Fisher's Exact Test, with a cutoff for statistically significant difference set at $p<0.05$.

Supplementary Figures

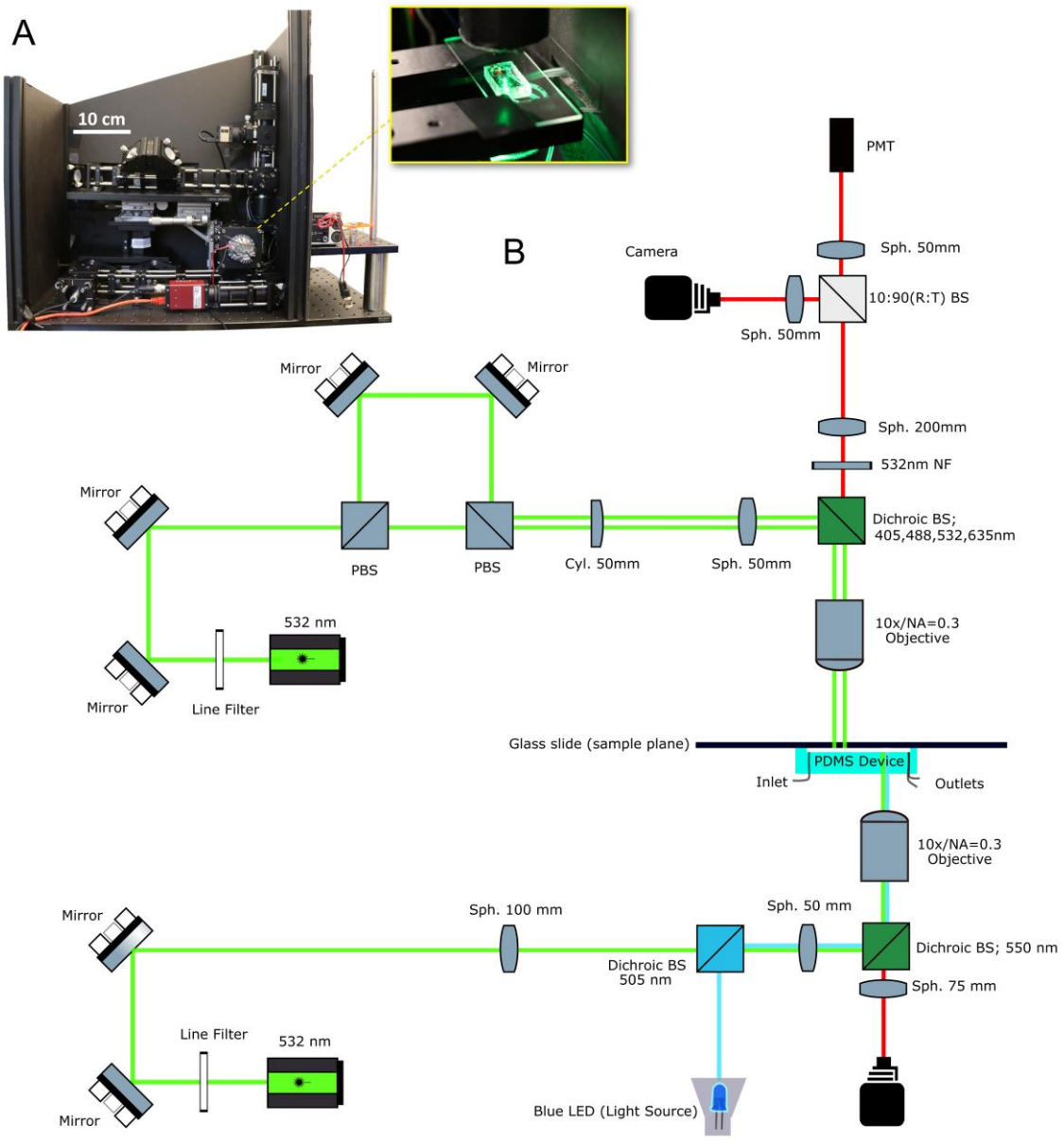

Fig. S1. The optical configuration of the CTC sorter. (**A**) Side-view image showing the optical components of the enclosed CTC sorter system. Inset indicates the location of the microfluidic device that is sandwiched between two vertical optical trains. (**B**) Top optical train splits the excitation source and focuses it into two lines across the microfluidic channel near the inlet. Emitted light is then filtered and passed to a PMT for real-time CTC detection. Bottom optical train combines a diffused excitation source at the bifurcation region for continuous imaging of the valve region (Key – Sph: Spherical lens, Cyl: Cylindrical lens, PBS: Polarizing Beamsplitter, R: Reflection, T: Transmission, NF: Notch filter, and NA: Numerical Aperture) (10).

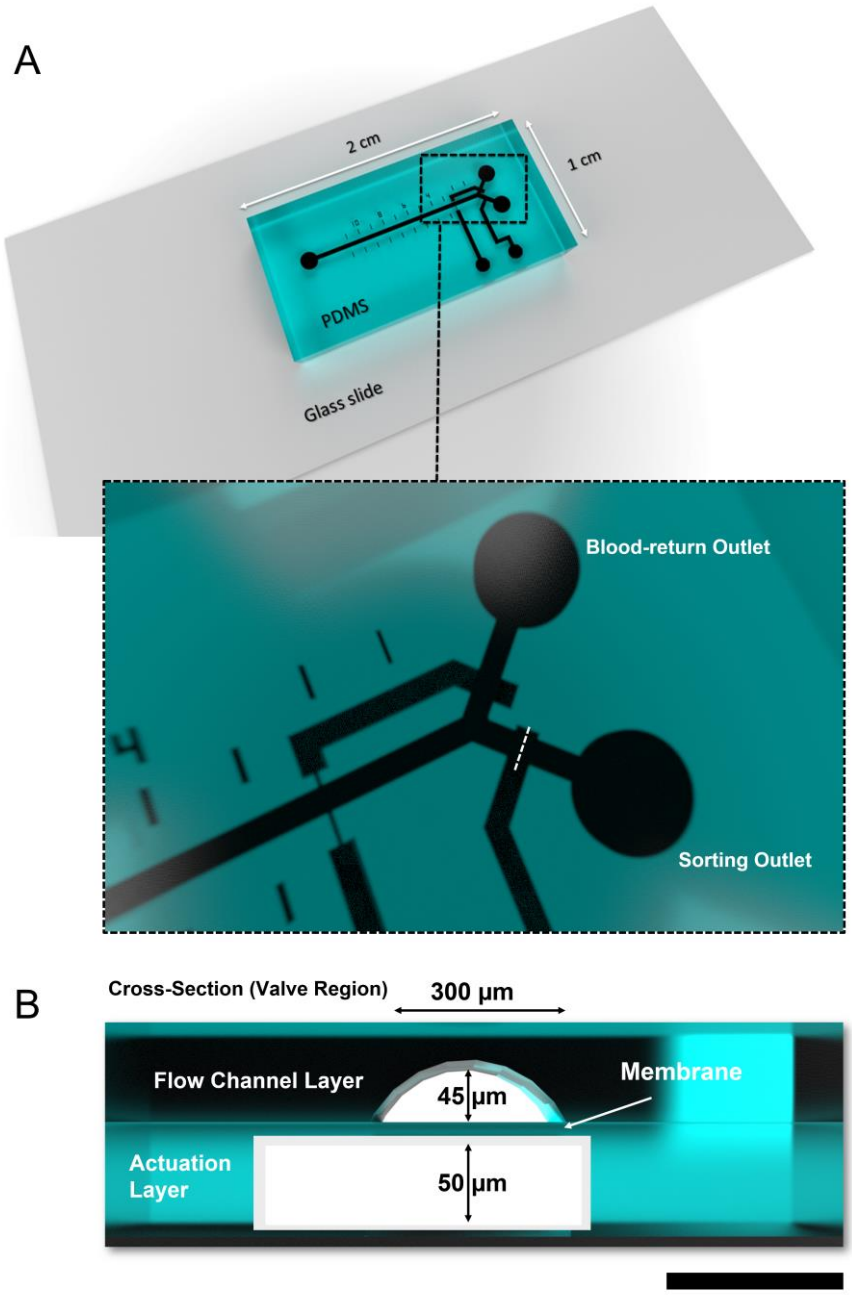

Fig. S2. Three-dimensional rendering of the microfluidic CTC sorting chip. (**A**) The device is fabricated using standard photo- and soft-lithographic techniques to create a two-layer PDMS device bonded to a 1×3-inch glass slide. (**B**) The microfluidic design contains a single inlet and two outlets. Actuation air channels (50 µm in height) lying underneath the outlet regions of the blood flow channels (half ellipsoid cross section, height = 45 µm, width = 300 µm) deflect a thin PDMS membrane upward or downward to stop or resume blood flow above, respectively.

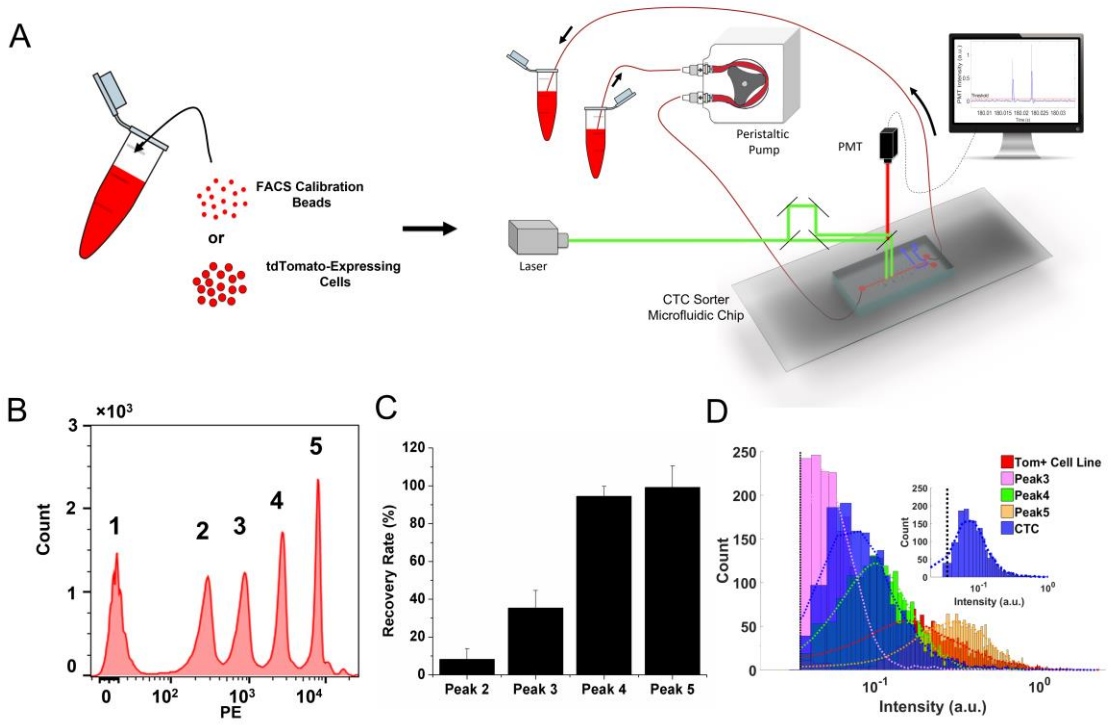

Fig. S3. CTC sorter system characterization. (**A**) Healthy mouse blood spiked with either fluorescent beads or tdTomato-positive cells is used to determine the detection limits of the system. Each sample is flowed through the sorter at a flow rate of 30 µL/min. (**B**) Five separate peaks representing the five different intensity levels of the SpheroTech PE Calibration beads as detected by FACS. (**C**) Beads representing fluorescent peaks 2 through 5 were spiked separately into four separate 500 µL samples of healthy mouse blood. Each sample was run through the CTC sorter to measure the recovery rate (i.e. Recovery Rate for Peak 2 sample = total detected/1,000×100). N = 3, results demonstrated as mean±s.d. (**D**) Comprehensive histogram plot with Kernel smoothing function fit (dotted envelope) demonstrating the result of all spiked-blood experiments using beads, a tdTomato-positive fluorescent cell line, or terminal bleed samples from SCLC tumor-bearing mice containing tdTomato-positive CTCs (N=1,900 events per population. For CTC population, six terminal bleed samples accounted for the 2,000 events). Inset represents the CTC population that lies above the detection threshold (vertical dotted line).

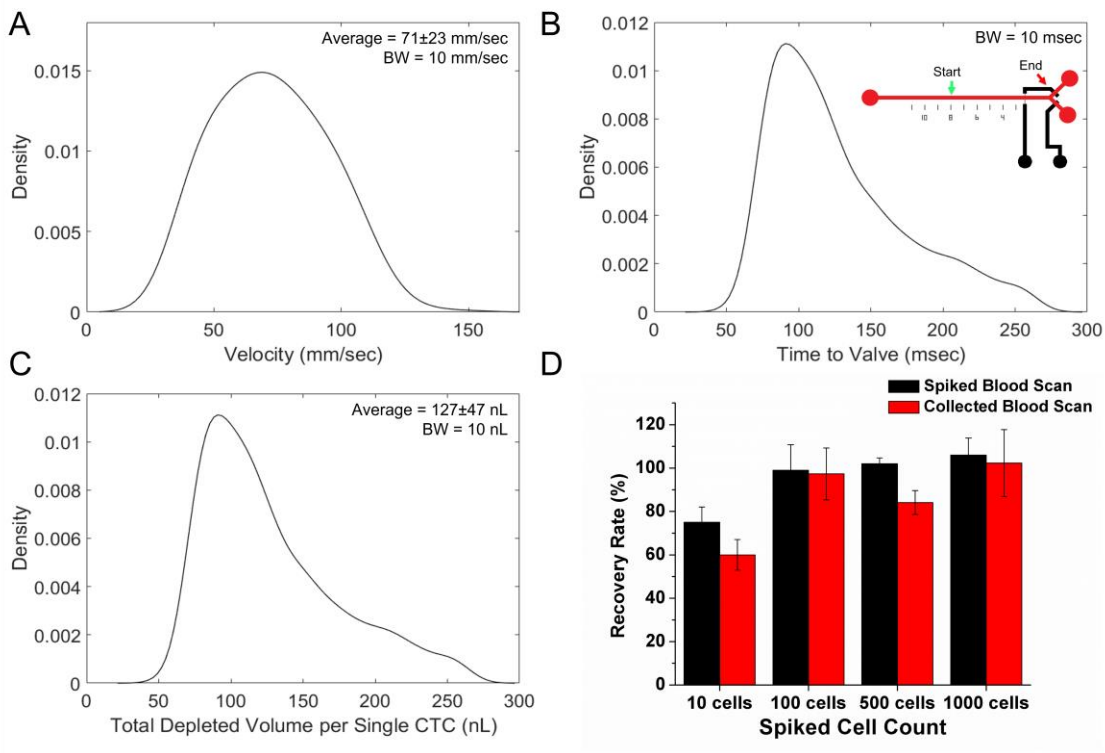

Fig. S4. Characterization of sorting reliability. (**A**) Velocity distribution of cells passing through the dual-excitation detection region of the channel. (**B**) Estimated time of arrival at the outlet distribution. Inset demonstrates the distance between the detection region (Start) and the valve region (End) which equals 8 mm. (**C**) Distribution of the total depleted blood volume per single sort (N=2,531 cells, Kernel smoothing function with bandwidth (BW) = 10 mm/sec, msec, or nL for parts a through c, respectively). (**D**) Healthy mouse blood samples spiked with tdTomato-expressing cells at physiologically-relevant densities (10, 100, 500, and 1,000 counted using a Coulter Counter) were run through the sorter to first identify the actual number of cells in each sample (black columns) with the sorting functionality activated. Sorted samples were then diluted and flowed through the sorter again (red columns) to measure the recovery yield of the sorted samples (N=3 repeats, mean±s.d.).

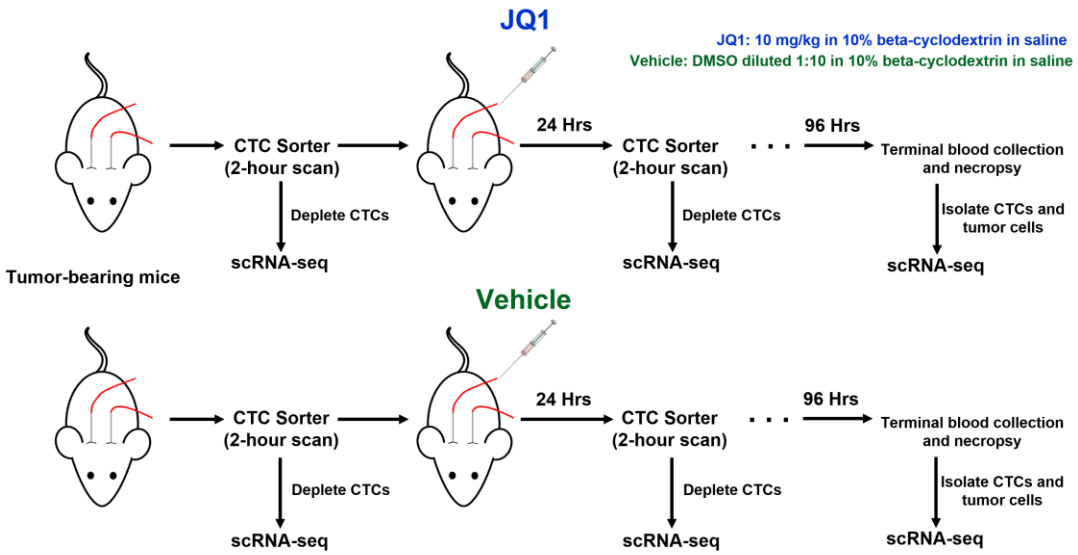

Fig. S5. Longitudinal JQ1 study timeline. Autochthonous SCLC tumor-bearing mice undergo arteriovenous surgery, followed by treatment with either the BET bromodomain inhibitor JQ1 (top row), or vehicle (bottom row). Each mouse is scanned to collect an initial CTC population before treatment. Intravenous injection of either JQ1 or vehicle then follows and continues daily following a two-hour scan for collection of CTCs. Collected blood samples from each mouse then undergo enrichment processes to sort single CTCs for downstream scRNA-Seq.

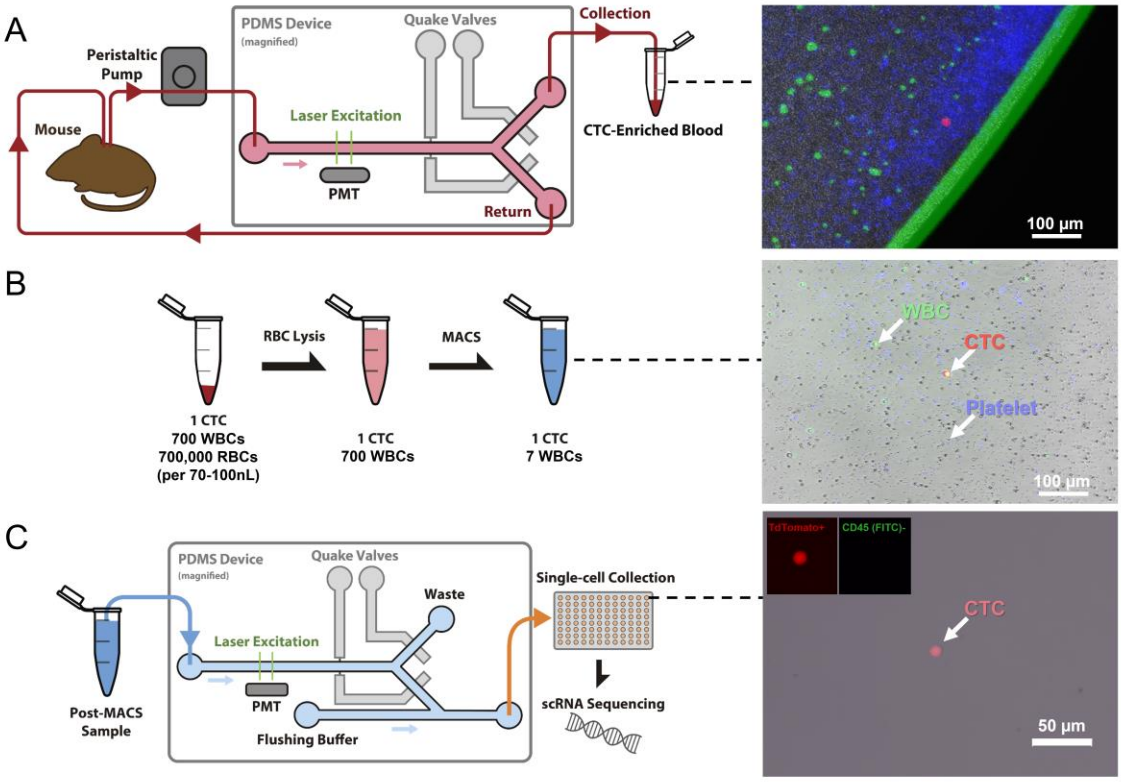

Fig. S6. CTC enrichment protocol. (**A**) A cannulated mouse is first connected to the CTC sorter system to collect a population of CTCs during a two-hour scan. Top microscopy image shows a tdTomato-expressing CTC within a well of a glass-bottom, 96-well plate containing whole mouse blood with stained platelets (blue) and WBCs (green). (**B**) Depleted blood volume containing the CTCs then undergoes RBC lysis and MACS to deplete both RBCs and WBCs and enrich for CTCs. Middle microscopy image shows the product of this step with several WBCs, platelets, and debris remaining in the product. (**C**) A secondary single-cell sorting chip is designed with a secondary parallel channel to interrogate the diluted MACS product (of step b) and sort single CTCs into wells containing cell-lysis buffer for downstream scRNA-Seq. Bottom microscopy image shows the end result of this enrichment protocol with a tdTomato-expressing CTC within a well of a 96-well plate (insets represent the tdTomato-positive (TRITC) and CD45-positive (FITC) channels).

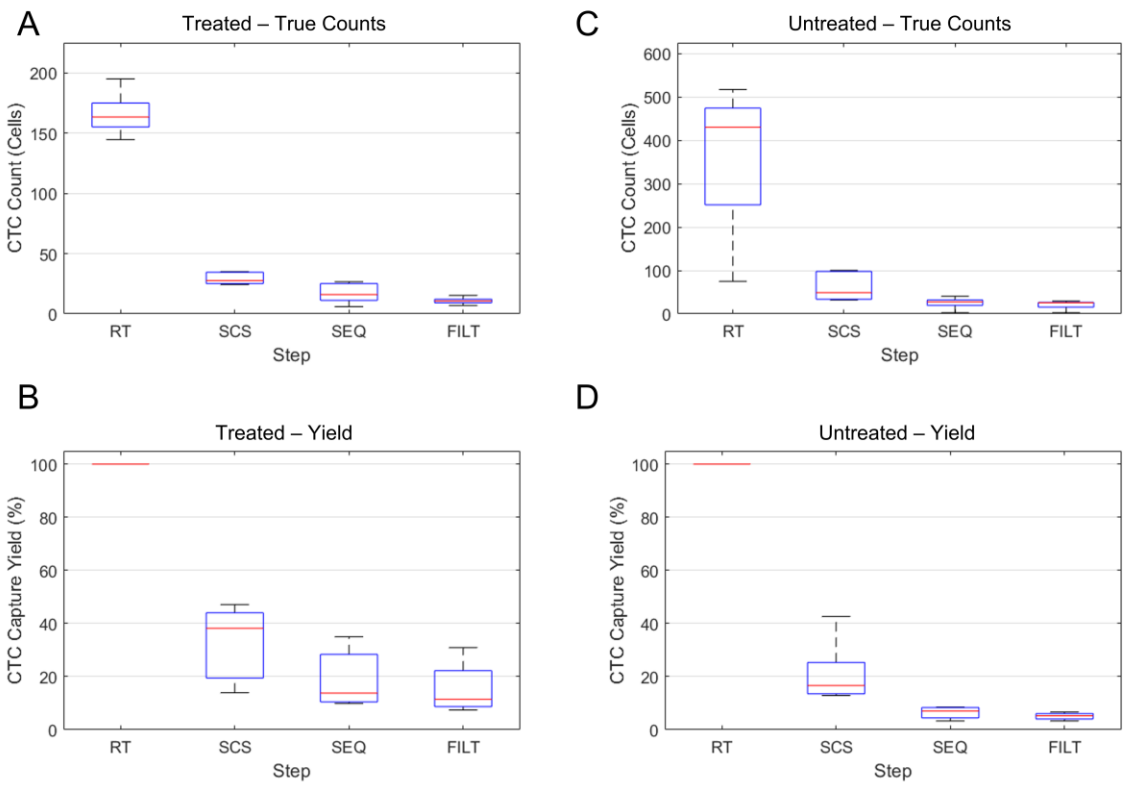

Fig. S7. Boxplots of CTC counts (**A, C**) and overall yields (**B, D,** normalized to RT count) of CTCs undergoing the enrichment process of: 1. real-time CTC collection during the two-hour mouse-blood scan (RT); 2. single-cell sorting (SCS) post-leukocyte depletion using the secondary single-cell sorting chip; 3. technical filtering of cells with sufficient gene complexity post-RNA sequencing (SEQ); and 4. final biological selection (filtering = FILT) of single CTCs that lack immune and platelet signature genes, for both the JQ1-treated samples (**A, B**) and untreated samples (**C, D**), respectively (the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively).

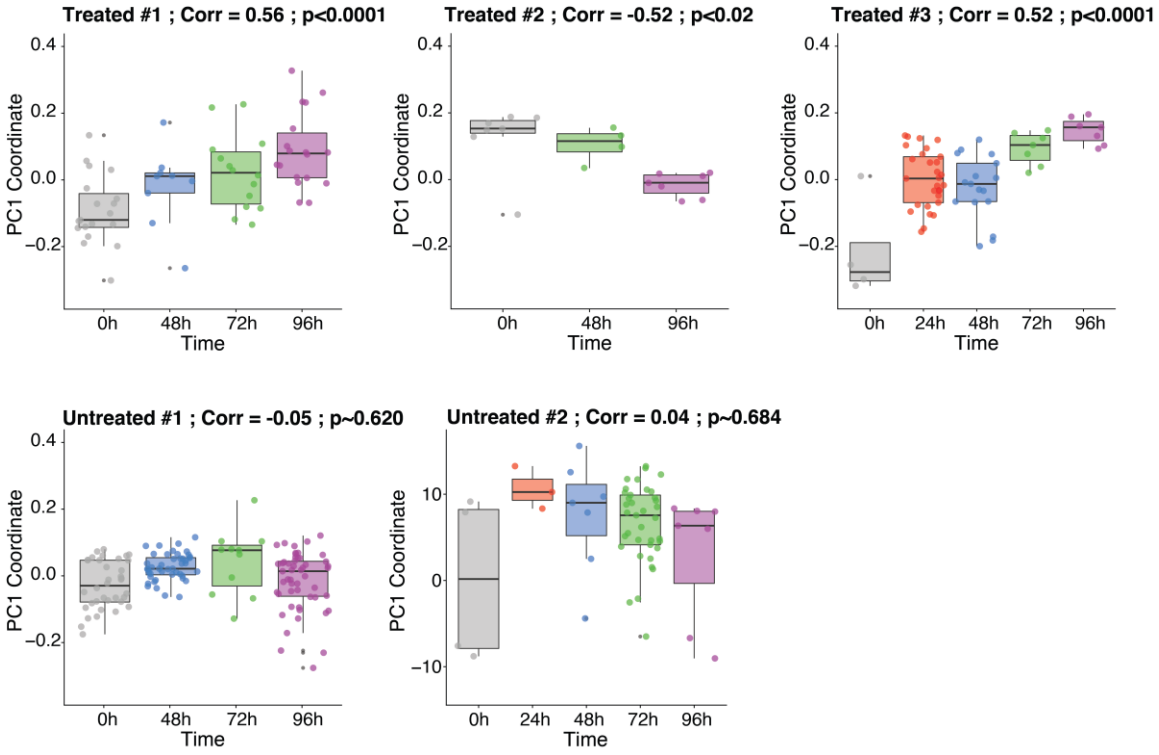

Fig. S8. Boxplots of the PC1 coordinate for all CTCs from treated mice (top row) and untreated mice (bottom row) over variable genes with correlations and p-values of correlations as determined by Student's t-test of significance, with all correlations of principal components following a Normality test (Lillefors). Data are delineated by time point of collection for each mouse presented. Treated #1 and Untreated #1 are also presented in **Fig. 2**; Treated #'s 2, 3 and Untreated #2 are replicate experiments.

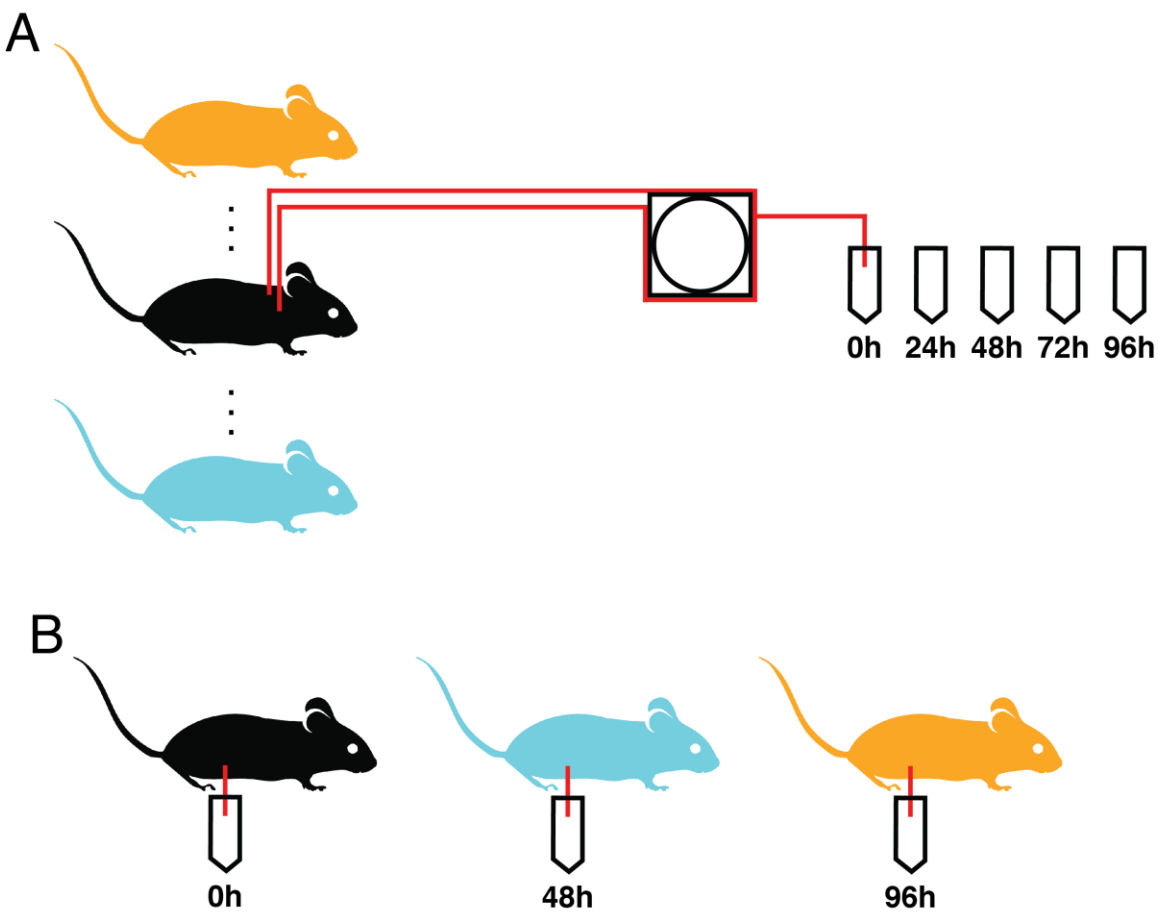

Fig. S9. Schematic of true (**A**) and mock (**B**) data for simulations of experiments. (**A**) shows the process by which single CTCs are collected using our CTC sorter. (**B**) shows one of the combinations of mice that could lead to a time course collection of CTCs, which simulates a terminal bleed experiment of CTCs. NB: we analyzed all six possible permutation combinations of (**B**).

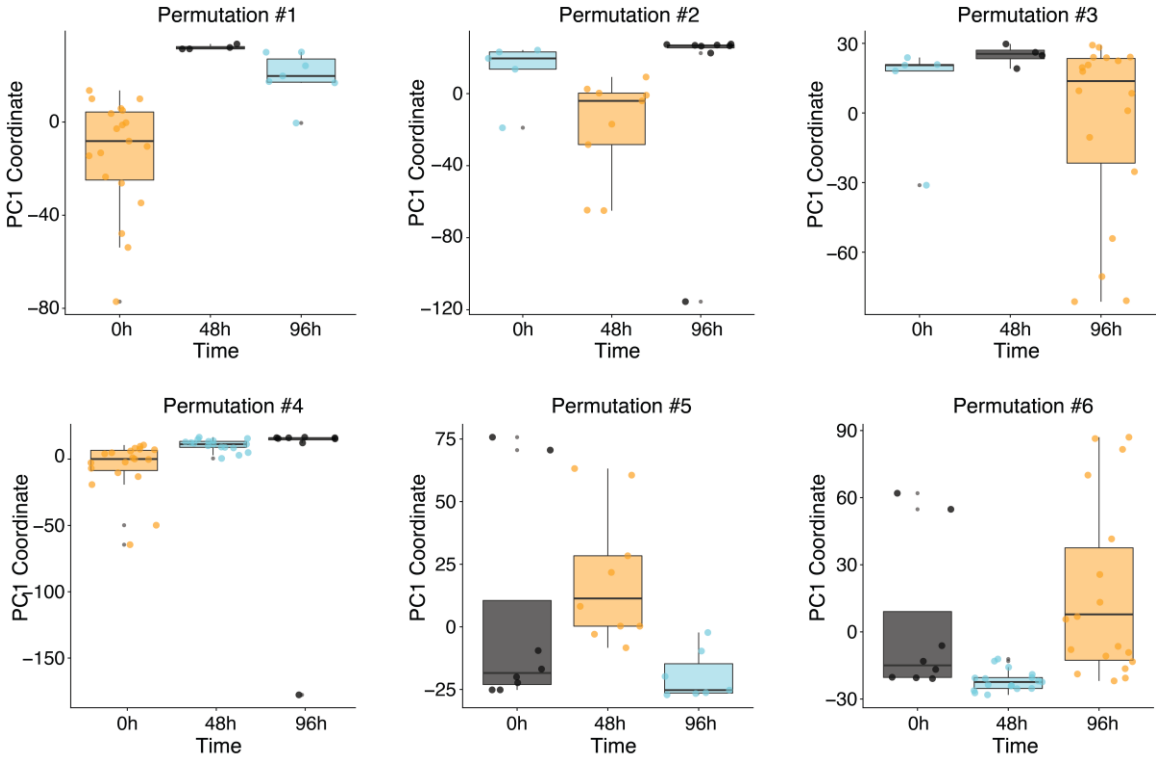

Fig. S10. PC1 boxplots of all combinations of treated mouse simulated data (see **Fig. S9B**). The mean PC1 coordinate of the time point from which the CTCs of Treated #1 were selected (yellow box) was outside the interquartile range of the other two treated mice in all possible combinations. NB Permutations 1-3 are also shown in **Fig. 2**.

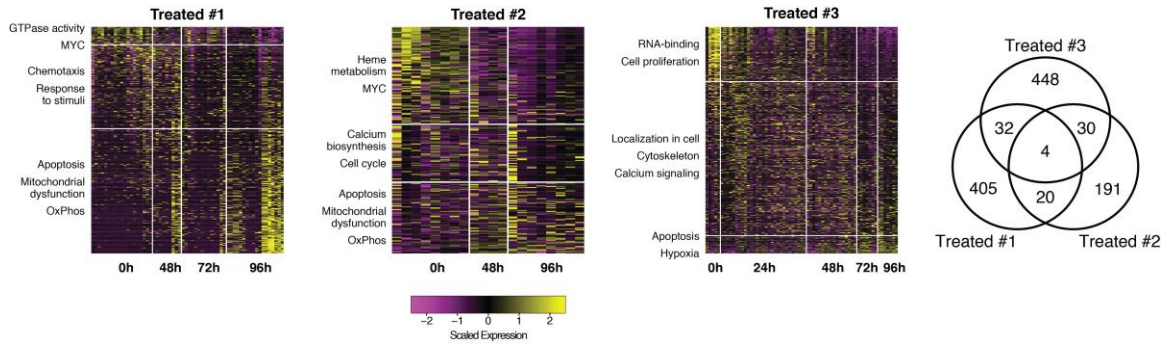

Fig. S11. Left: Heatmaps of differentially expressed genes between all pairs of time points within each treated mouse. Rows represent genes, columns represent cells. Time points sampled are provided on the x-axis and selected gene set enrichments for each differentially expressed gene module (identified by hierarchical clustering; see **Materials and Methods**) are displayed to the left of the relevant cluster (for a full list of enrichments, see **Table S1**)**.** The majority of these gene set enrichments are unique to each pair of mouse and time point. **Right:** Venn diagram shows overlap of differentially expressed genes across all treated mice.

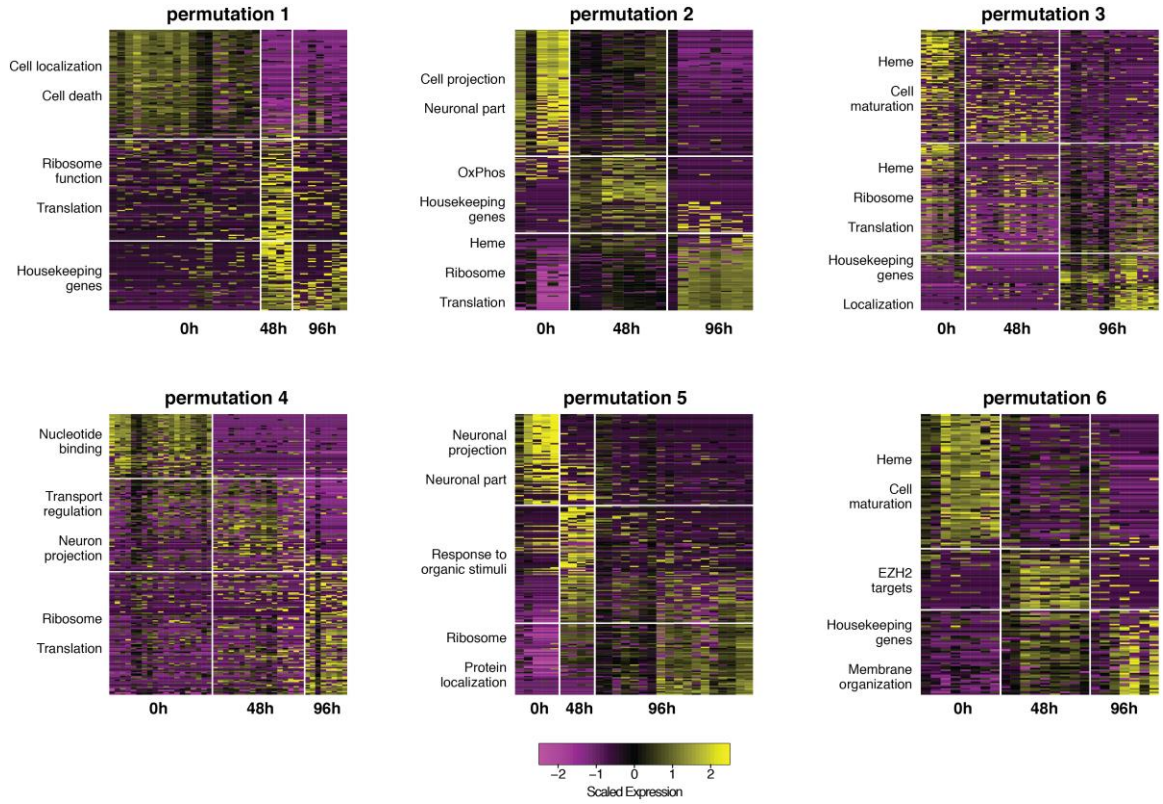

Fig. S12. Heatmaps of differentially expressed genes between all pairs of time points within each possible permutation for a mock treatment time course using the common 0, 48 and 96h time points (See **Fig. S9** and **S10**). Rows represent genes, columns represent cells. Time points sampled are provided on the x-axis and selected gene set enrichments for each differentially expressed gene module (identified by hierarchical clustering; see **Materials and Methods**) are displayed to the left of the relevant cluster (for a full list of enrichments, see **Table S2**). The majority of these gene set enrichments are unique to each pair of mouse and time point.

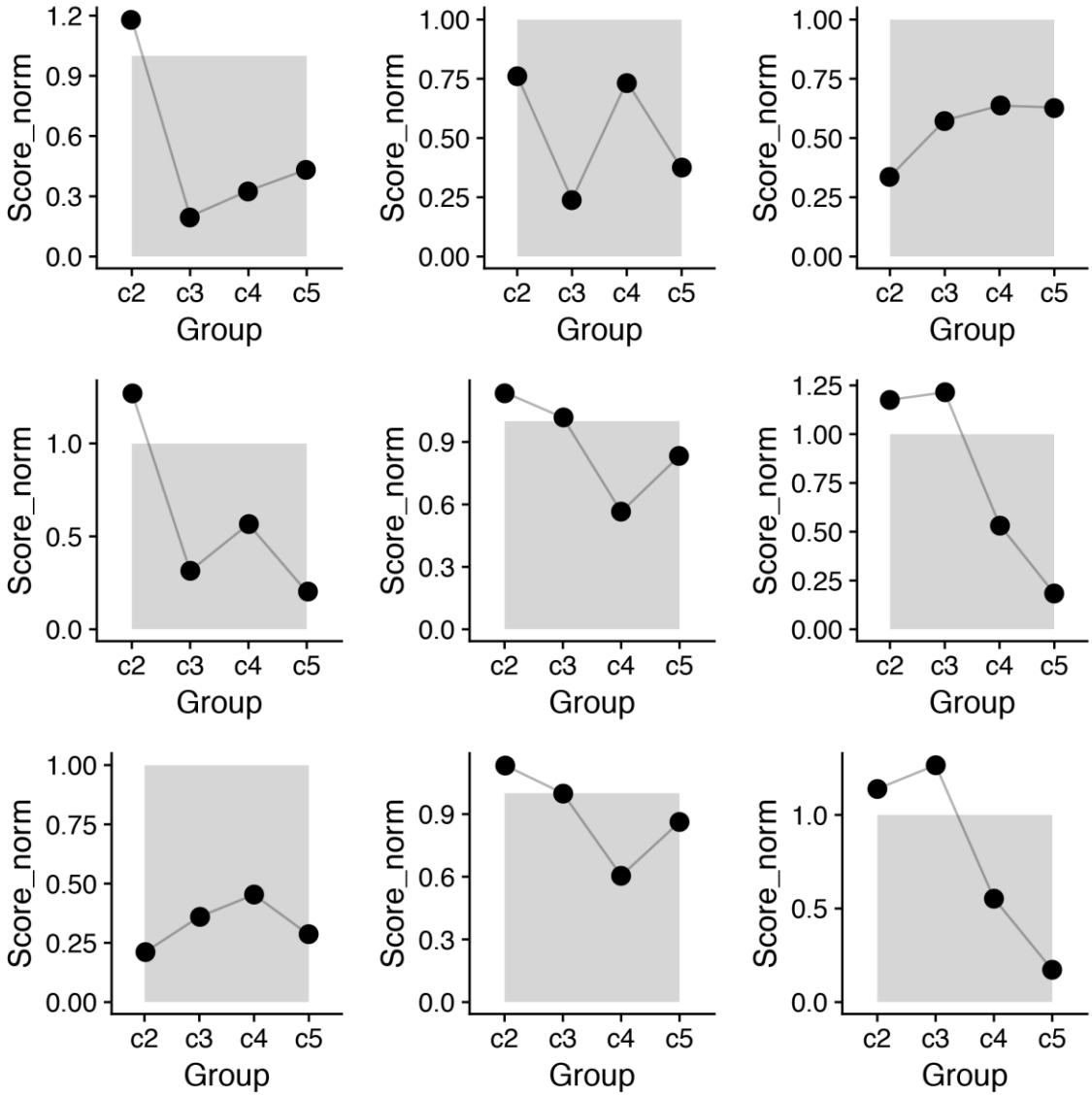

Fig. S13. Schematic representation of scoring algorithm used to assign CTCs above background to specific clusters of phenotypes from primary tumor data. CTC signatures were calculated for each cluster of primary tumor cells in **Fig. S13A** (black dots) and compared to a background generated by permuting random sets of genes with similar mean bins of expression (gray background). CTCs were scored to clusters above background (e.g. top left) or not scored if no clusters were above background (e.g. top middle). (See **Materials and Methods** for full description of assignment framework).

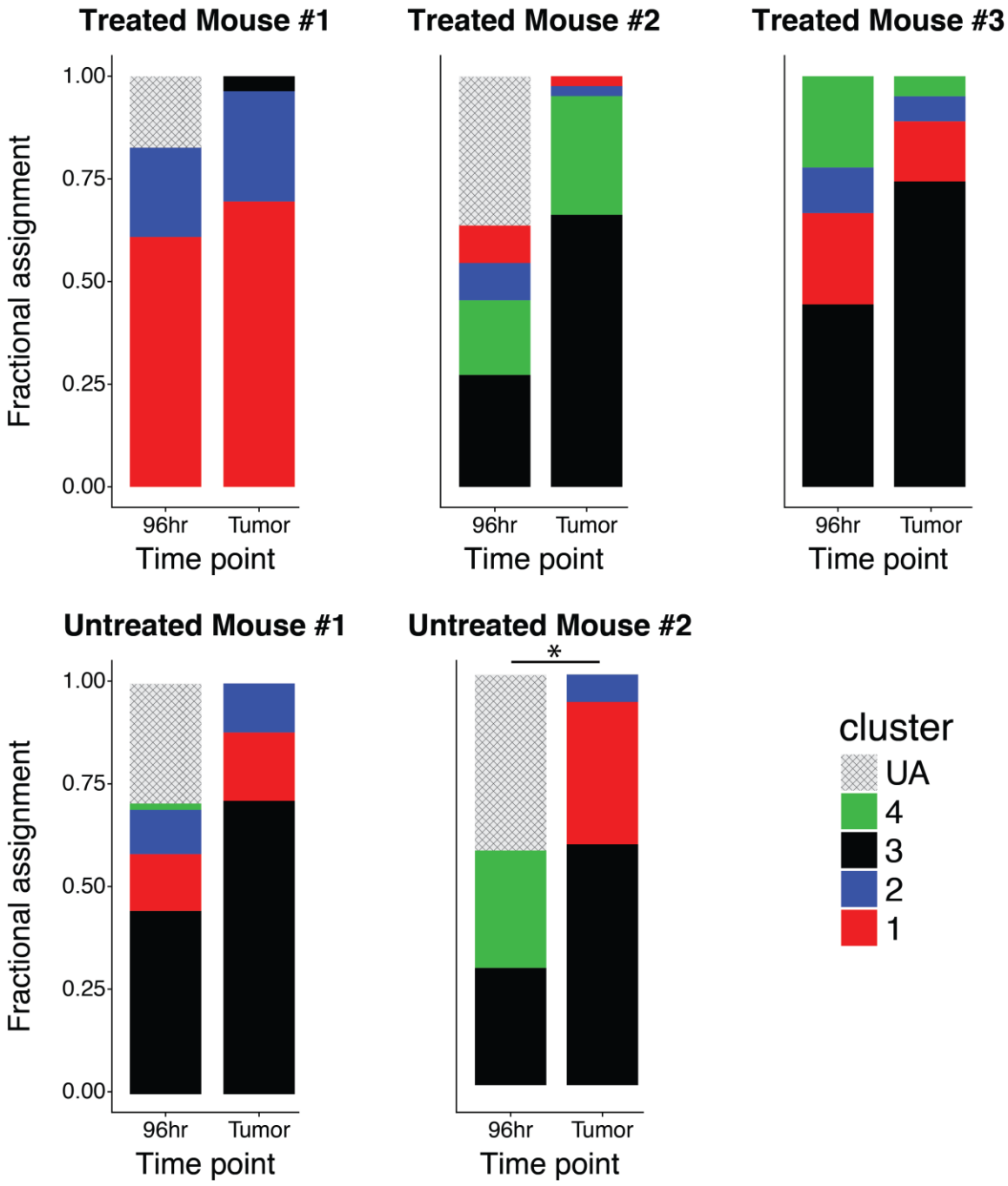

Fig. S14. Assignments for 96-hr CTCs and primary tumors based on clusters identified by scRNA-Seq of primary treated and untreated mice. CTCs were assigned (or not assigned) to clusters based on the highest signature score above permuted background (**Materials and Methods**; **Fig. S13**), and the results were visualized using stacked bar plots. UA: unassigned. Similarity tested by Fisher's Exact test ($p = 0.99$, 0.08, 0.16, and 0.66 for treated mice 1 through 3 and untreated mouse #1, respectively. ✳ = $p < 0.05$ for untreated mouse # 2). N=18, 8, 7, 52, and 7 CTCs for treated mice 1 through 3 and untreated mice 1 and 2, respectively. N=82, 82, 79, 84, and 75 primary tumor cells for treated mice 1 through 3 and untreated mice 1 and 2, respectively.

Table S1: List of genes used to derive platelet signature for each CTC

Pf4	Tmem140	Vcl
Kif2a	Snca	Ypel5
Tsc22d1	Pdlim1	Ppm1a
Wipf1	Pgrmc1	Ptgs1
Asah1	Ctsa	Itga2b
Gnas	Snn	Mpp1
Rnf11	Itgb3	Dapp1
Nap1l1	Fhl1	Eif2ak1
Thbs1	Mylk	Prkar2b
Rsu1	Prdx6	Prkcb
App	Rabgap1l	Cxcl5
Pip4k2a	Mfap3l	Tpm1
Pdgfra	F13a1	Pecam1
Rab31	Endod1	Mmd
Max	Tax1bp3	Leprot
Odc1	Nrgn	Tuba4a
Gng11	Rgs10	Sparc
Ppbp	Itgb5	Tubb2a
Mlh3	Cttn	
Ptpn12	Gla	

Table S2: List of marker genes for each primary tumor cell cluster

C2	C3	C4	C5
Tuba1a	Cd209c	Prom2	Selenbp1
Nsg1	Hist1h2ao	Cnpy1	F3
Nfib	Gm13399	Glb1l2	Liph
Uchl1	Hist1h2ae	Car8	C920009B18Rik
Nfix	Hist1h2br	Rdh9	Mettl7a1
Adcy1	Hist1h2bq	Gpx3	Ier3
Stmn1	Hist1h4d	Kcnk16	Emp2
Ptp4a2	E030030I06Rik	Dnajb1	Anxa2
Dbn1	Nanog	Ovol1	Fgfr3
Miat	Hist1h2ah	Rab27b	Sh2d4a
Rufy3	4930519F09Rik	Egfr	Cdkn1c
Nme1	Aurka	Nov	Gm10190
Sox11	Cep55	Gm166	Net1
Acat1	Kif15	Cib3	D630045M09Rik
Cacna1c	Hist1h4i	Ccdc24	Gm4841
Gria2	Hist1h1e	Zfp36	Kctd14
Nova1	Nr1d1	Csrnp1	Sdc1
Mapt	Ttk	Cdc14a	Epas1
Erh	Hist2h3c2	Tspan1	Pls3
Pdlim7	Ska1	Ccdc154	Xkrx
Tubb3	Troap	Ces1d	Prdx6
Vma21	Ccnb1	Btg2	Ptprf
Tsn	Pbld1	Slc44a3	Msn
Cisd1	Esco2	Krt18	Lama5
Lrrn1	Ankle1	Slc16a5	Neat1
Cdv3	Cit	Cdc14b	Npc2
Dpysl4	Cdca2	Kcnip2	Cryab
Ncam1	Sgol1	Kcnq1	Rin2
Elavl4	Mis18bp1	Adamts17	Ugdh
Shank1	Kif11	Me3	1700128F08Rik
Hmgb3	Kif20b	Dsp	Epn3
Fli1	Bub1	Igf2as	Ttc7
Trim44	Spag5	Egln3	Ly6e
Faim	Sapcd2	Itm2c	Myh14
Gnb1	Espl1	Cldn7	Tspan2
Dcx	Arhgef39	Pqlc1	Lmo7
Nme2	Nuf2	Tmem184a	Tppp3
Pfas	Shcbp1	Sqrdl	4921515E04Rik
Map1b	Omd	Cd59a	Notch2
Ttc9b	Aurkb	Gem	Cyp2j5
Grik3	Cdca5	Sytl1	Dapk1

Fabp5	Hmmr	Ap1m2	Nfkbia
Vat1l	Kif22	Tmem30b	Myo1c
Slc29a1	Ckap2l	B4galt2	H2-Q9
Got1	Aspm	Ptprn	Sat1
Brd3	Arrdc3	Ncoa7	Me1
Rbx1	Plk1	Pirt	Myadm
Celf5	Ncapg	Vwf	Pcx
Cacna2d1	Fam64a	Lgals9	Pmp22
Pcna	Prc1	Igf2	Aim2
Marcksl1	Kifc5b	Grasp	Mlph
Lrdd	Nusap1	Krt8	Plin3
Med1	Kifc1	Aldh3b1	Taok3
Dek	Ckap2	Arc	F830016B08Rik
Syt11	Ndc80	Fhl2	Prss35
Fam92a	Kif18b	Bcr	Syt12
Hnrnpa0	Ccna2	Piezo2	Lpin2
9030624G23Rik	Birc5	Cdh13	Vegfa
Atp1a3	Ect2	Inpp5j	Sfn
Gabrb3	Gm11974	Insrr	Bcam
Sez6	Fbxo5	Ets2	Psg28
Tm2d2	Casc5	Plcb4	Wfdc2
Cdip1	Tpx2	Aldh1a7	Trip10
Btbd17	Prr11	Atf3	2410004P03Rik
Basp1	Ube2c	Tmc4	Clip4
Nrxn2	C330027C09Rik	Fosb	Mboat1
Nr2f2	Kif18a	Hhatl	Lipi
C1galt1c1	Hist1h2bc	Mest	Eaf2
Foxp1	Cdc20	St14	Sh3rf3
Mtx2	Cdc25c	Calca	Gclc
Mcfd2	Ska3	Lyplal1	Prkci
Polr2c	Dbf4	Espn	P4ha2
Tmeff2	Rad54l	Egr2	Ppl
Trp53inp2	Kif2c	Actn1	Pkp2
0610037L13Rik	Sgol2	Ret	Akap5
Slc6a8	Pbk	Tgfb3	A430107P09Rik
Arf2	Cdca8	Hspa4l	Cat
Gm20337	2700099C18Rik	Atp2a3	Anxa5
Trappc1	Gas2l3	Glt28d2	Gsn
Ptprs	Kif23	Kcnn3	Vamp8
	Chek2		Gabre
			Mprip

References

1. McFadden DG, et al. (2014) Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 156(6):1298–1311.
2. Sutherland KD, et al. (2011) Cell of origin of small cell lung cancer: Inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* 19(6):754–764.
3. DuPage M, Dooley AL, Jacks T (2009) Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* 4(8):1064–1072.
4. Dooley AL, et al. (2011) Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev* 25(14):1470–5.
5. Ayala JE, Bracy DP, Mcguinness OP, Wasserman DH (2006) Considerations in the Design of Hyperinsulinemic- Euglycemic Clamps in the Conscious Mouse. *Diabetes* 55(October 2005):390–397.
6. Ayala JE, et al. (2010) Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Dis Model Mech* 3(9–10):525–534.
7. Ayala JE, et al. (2011) Hyperinsulinemic-euglycemic Clamps in Conscious, Unrestrained Mice. *J Vis Exp* (57):1–8.
8. Unger MA, Chou HP, Thorsen T, Scherer A, Quake SR (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* 288(5463):113–6.
9. Zheng GXY, et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:1–12.
10. Component Library by Alexander Franzen is licensed under a Creative Commons Attribution-Non Commercial 3.0 Unported License.