Supplementary information

Supplementary Methods and Materials

Microfluidic device fabrication. The PDMS-based pillar-lattice-arrays were designed by using software (AutoCAD) and cast from SU8 mold.([1](#)) The distance of two adjacent cylindrical pillars is spaced identically to 35 μm, which would allow for the unrestricted motility of T cells. Cuboid pillars, which are about 220 μm in diameter, make it easier to peel off the solidified PDMS compared to all the same cylindrical pillars. Standard photolithography techniques were used to fabricate the SU8 mold of the pillar lattices. Briefly, the photoresist (SU8-2100) was evenly coated onto a silicon wafer at a 150-μm thickness (10s at 100 rpm, and 30s at 1800 rpm), followed by processes of the soft bake, UV-light exposure, post-exposure bake, develop and hard bake. The mold was exposed to chlorotrimethylsilane (Sigma, USA) for 5min to facilitate the peeling of the PDMS pillars. Subsequently, the PDMS mixture was poured onto the wafer (Sylgard 184, USA, weight ratio A/B = 10:1) and coated upon SU8 mold by using a spinning coating machine (10s at 100 rpm, 30s at 500 rpm). A thin PDMS layer was obtained and degassed in a vacuum desiccator for 15 minutes. The wafer with the PDMS layer was baked in a conventional oven at 80 °C for 1 hour. After curing, the thin PDMS layer was carefully peeled off and tailored to fit the standard 96-well plate. Finally, PDMS-based pillar-lattice-arrays were autoclaved and placed in an imaging 96-well plate (MatTek). The pillars were then soaked in 1% (w/v) Pluronic 108 solution overnight, sterilized with 70% ethanol for 15 minutes, and washed twice with DMEM culture medium.

Cell isolation and culture.

Cancer cell culture: The mouse melanoma cell line B16F10 was purchased from ATCC. The mouse pancreatic tumor cell line UN-KC6141 was a kind gift from Dr. Surinder K. Batra (University of Nebraska).([2](#)) Both tumor cell lines were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) (Gibco, USA) and penicillin-streptomycin (100 U/mL) (Gibco, USA). Cells were maintained in a 5% CO_2 supplemented, 37°C humidified incubator, and passaged at 70% confluency by trypsin-EDTA (Gibco, USA).

Primary CAF culture. Cancer-associated fibroblasts (CAFs) were isolated from orthotopic tumors. Briefly, 200,000 UN-KC6141 or B16F10 tumor cells in 30% (v/v) Matrigel (BD, USA) were injected into the pancreas and subcutaneously (s.c.), respectively, into 4-6 weeks old C57BL6 mice. All procedures were approved by the Indiana University Institutional Animal Care and Use Committee (IACUC). Tumors were allowed to grow for 3 weeks *in vivo*. After 3 weeks, mice were euthanized to collect xenograft tumor tissue. Tumor tissue was digested by a mouse tumor dissociation kit (Miltenyi, Germany) using a gentleMACS dissociator. Dissociated tumor single-cell suspensions were filtered using a 70 μm cell strainer, resuspended in fibroblast growth medium-2 (FGM-2) (Lonza, USA), and the immediately cultured adherently in a 5% CO_2 supplemented, 37°C humidified incubator for 3 weeks to allow for fibroblasts outgrowth. The CAF cells were validated using FAP/alpha-SMA immunofluorescence staining to confirm their identity.

CD8+ T cell isolation and culture. OT-I mice [C57BL/6-Tg(TcraTcrb)1100Mjb/J] were purchased from the Jackson Laboratory (JAX) and bred in house. All procedures were approved by the Indiana University Institutional Animal Care and Use Committee (IACUC). The OT-I mice contain Tcra-V2 and Tcrb-V5 transgenes and produce MHC class I-restricted, ovalbumin-specific, CD8+ T cells (OT-I cells). Newborn offspring were genotyped for OT-I T cell receptor (TCR) expression according to protocols provided by JAX. Genotyped newborns of 4-10 weeks old were euthanized to collect spleen. Spleen was grinded using a 30 μm cell strainer. Cells passed through the strainer were then treated with ACK lysis buffer (Gibco) to lyse red blood cells. Then the CD8+ T cells were isolated using a naive CD8a T Cell isolation kit (Miltenyi, Germany). The isolated CD8+ T cells were cultured in DMEM medium supplemented with 10% FBS, 30U/mL rIL-2 (Peprotech, USA) and penicillin-streptomycin (100 U/mL). Dynabeads CD3/CD28 were added to T cell culture at a bead-to-cell ratio of 1:1 to stimulate T cell expansion *in vitro*.

On-chip investigation of T cell-tumor infiltration.

Hybrid tumor-stroma spheroid formation. To form hybrid tumor-CAF spheroids, 0.4 million tumor cells (UN-KC6141-OVA or B16F10-OVA) were labeled with membrane DiO dye (Green) (Invitrogen, USA) and seeded on the pillar-lattice-arrays in each well of 96 well imaging plate (MatTek, USA). After 24 h, tumor cells spontaneously aggregated into tumor spheroids. Next, 0.6 million CAF cells were labeled with DiI dye (red) (Invitrogen, USA) and seeded on the pillar-lattice-arrays. The newly seeded CAF cells spontaneously assembled on top of the tumor spheroids to form a tumor core surrounded by a shell of CAF cells, enclosing the tumor core in 12 h (**Fig.S1**).

T cell track analysis. To analyze T cell infiltration tracks, raw images captured on the confocal microscopes were exported as TIFF image stacks in ImageJ. The image stacks were imported into Imaris 9.0 software (Bitplane, Switzerland) for T cell track detection and analysis. T cells were detected by analyzing the blue CMAC tracker channel with a spot radius of 10 μm with background subtraction. Tracking was performed using the Brownian motion model with a maximum displacement of 50 μm and a maximum gap of 1 frame. Tumor spheroids were detected by a surface with a radius of 300 μm. The detected tumor spheroid was analyzed by distance transformation (Matlab plugin) to calculate the distance inside and outside the spheroid surface. The final infiltration depth (Id) was calculated by deducting the "outside distance" from the "inside distance". We filtered out T cells that interacted with the spheroid by thresholding the maximum T cell track distance of the Id at 0.1 μm. T cell cytotoxicity was analyzed as described previously,([3], [4]) briefly dead tumor cells were detected by the staining of SYTOX deep red dye at 0.2 μM, detected in the deep red channel with a spot radius of 10 μm. T cell and dead cell colocalization were detected using a distance threshold of 10 μm.

Deep-learning-based TIL score analyzer.

Data and TIL maps. Whole-slide images and clinical follow-up data were obtained from TCGA (https://gdc.cancer.gov/) Skin Cutaneous Melanoma (SKCM) projects. The TIL localization inside tumors were extracted from the images of diagnostic H&E-stained sections from the Skin Cutaneous Melanoma as previously report.([5]) The TIL map were colored in red and blue representing TILs (or T cells) and tumor. A total of 411 whole-slide images from 397 unique patients were processed to generate

TIL maps for analysis. The TIL maps were sampled with 128 x 128 pixels and resized to 256 x 256 pixels.

Network architecture and training procedures. The TIL scoring deep learning system was developed to score TIL infiltration based on clinical pathology data and patient survival data. The TIL scoring algorithm combines elements of the 18-layer residual neural network (Resnet18)([6]) with a discrete-time hazards model to predict time-to-event data from images.([7], [8]) Image feature extraction is achieved by five groups of convolutional layers arranged in residual blocks. The extracted features are output to a sequence of two fully connected layers containing 1,000, 20 nodes, respectively. The terminal fully connected layer outputs a prediction of the hazard function $h(\tau_j|x) = \frac{1}{1+\exp[-\phi_j(x)]}$ at the discrete time-point τ_j associated with the image, where $\phi(x) \in R^m$ represents the output of the network to m (20 in our case) discrete-time point, x is the input of network corresponding to each image sample. The loss function for backpropagation is the modified cross-entropy function to account for the censored data. The full loss function can be written as follows:([9], [10])

$$loss = -\frac{1}{n}\sum_{1=1}^{n}\left(d_i\log[\sigma_{k(t_i)}(t_i)(\phi(x_i))] + (1-d_i)\log\left[\sum_{k=\kappa(t_i)+1}^{m+1}\sigma_k(\phi(x))\right]\right)$$

where $k(t_i)$ denotes the duration index of individual i's event time. As the hazard function was trained, a 3-year survival classification was used to give TIL score.

Training sampling. To establish the scoring algorithm that scores TIL pattern based on patient survival time, our deep learning algorithm was fine-tuned on pre-trained Resnet18 for 30 epochs. The stochastic gradient descent (SGD) optimizer was used to minimize the negative log-likelihood via backpropagation to optimize model weights of fully connected layers.([11]) Model weights of modified fully connected layers were initialized using the variance scaling method, and a weight decay was applied to the fully connected layers during training (decay rate = 0.1). Models were trained for 30 epochs (1 epoch is one complete cycle through all training samples) using mini-batches consisting of 16 image samples each. Minibatch samples were random assigned at the beginning of each epoch for robustness. During training, a single area was sampled from each slide, and these sampled areas were treated as semi-independent training samples. Each sample was labeled with the corresponding patient survival time for training, duplicating survival time for patients that preserved multiple slides. The areas were sampled randomly at the beginning of each training epoch and regarded as an entirely new set of samples. These sample areas were randomly transformed to improve the robustness to account for tissue orientation, color variations, and anisotropy. The contrast[0.5], brightness[0.5], saturation[0.5] and hue[0.5] of the samples were also randomly transformed using the "colorJitter" PyTorch operations. These sampling and transformation procedures have the effect of augmenting the effective size of the limited clinical labeled training data. Similar augmentation approaches have been widely adapted and shown considerable improvements in neural network performance in a variety of imaging applications.([12])

Prediction sampling and averaging. Sampling was also performed to increase the robustness and stability of predictions. (*i*) Nine areas are first sampled from slides corresponding to each patient. (*ii*) The survival time expectation (Se) of each sample for each patient is then predicted using the trained

deep learning system. (*iii*) The median Se of nine predictions from samples is calculated as the final survival time prediction for the patient.

Validation procedures. To train and evaluate model performance, patients were randomly assigned to non-overlapping training (80%) and test (20%) sets. If a patient was assigned to one set, then all H&E slides corresponding to that patient were assigned to the corresponding set. This ensured that no slices from a single patient were assigned to both training and testing sets to avoid overfitting and ensure validation accuracy. The randomized assignment of patient datasets was performed 15 times and each of these training/testing sets was employed to train and evaluate a model. Prediction accuracy was measured using a time-dependent concordance index (c index) to measure the concordance between predicted survival and actual survival for testing samples.([12]) A c index of 1 indicates perfect concordance between predicted risk and overall survival, and a c index of 0.5 corresponds to random concordance. The following 3-year classification outcomes were also validated by cross validation, and roc curves from the cross validation were drawn to evaluate the classification performance.

Screening of epigenetic drugs.

Epigenetic drug library screening. To screen effective epigenetic drugs that promote T cell infiltration, tumor-stroma spheroids at day 3 (24 hours post CAF cell addition) were treated with 5 μM of chemicals of an epigenetic drug library (Cayman Chemicals, USA, Epigenetics Screening Library, Catalog# 11076) for 24 hours. Post epigenetic drug treatment, culture medium was refreshed, and OT-I T cells, labeled with blue CMAC tracker dye (Invitrogen, USA), were added at 50,000 cells per well. Additionally, anti-mouse-PD-1 antibody (clone RMP1-14, Bioxcell, USA) was added at a concentration of 3.5 μg/mL together with OT-I T cells. SYTOX deep red nucleic acid stain was loaded at a final concentration of 0.2 μM to visualize cell death as it stains newly killed cells' nuclei once the cell membrane integrity is compromised. The plate was then set up for time-lapse imaging on a Leica SP8 confocal microscope or Olympus OSR spinning disk confocal microscope for 12 h in a 5% CO_2 supplemented, 37°C humidified incubation chamber. Confocal images were taken every 15 min with 4 channels.

Drug screening result TIL scoring. The final images of T cell distributions in tumor-stroma spheroids after 24 hours of screening were analyzed for TIL scores by our deep learning algorithm. For each treatment, ten spheroids' images were collected and used for TIL scoring. Each screening result images was given a TIL Score$_i$ of "0" when the TIL infiltration pattern was classified as "hazard"; or a TIL Score$_i$ of "1" when it was classified as "non-hazard". The final infiltration score for each drug, calculated by averaging the Score$_i$ from 10 screening results from the same drug treatment condition.

Validation of selected drug candidates.

In vivo validation of GSK-LSD1. To validate the effect of the anti-tumor activity of GSK-LSD1 *in vivo*, 200,000 B16F10 cells were injected subcutaneously (s.c.) into 40 C57BL6 mice of 5 weeks age (Envigo). On day 7, mice were randomized into 4 groups to have similar starting average tumor size (p>0.05), with 10 mice per group. Each animal was injected intraperitoneal (i.p.) with 200 μg anti-PD-1 (clone RMP1-14, Bioxcell, USA) and/or GSK-LSD1 (20mg/kg) every other day from day 8 to day 17. Tumor size was measured every other day and calculated as tumor volume = (length*width^2)/2.

Spheroid and tumor immunofluorescence staining. To visualize infiltrated T cells, the spheroid and tumor immunofluorescence staining were conducted using our previously developed protocols.([13-16](#)) Tumor spheroids or tumor tissues were fixed with 4% PBS buffered formalin (Sigma, USA) overnight then transferred to 70% ethanol for dehydration. The dehydrated spheroid/tissues were then embedded in paraffin and sectioned at 30 μm thickness and mounted onto slides. The mounted tissue sections were then treated for antigen retrieval by heat-mediated antigen retrieval and stained with Alexa 594 labelled anti-CD8 antibody (Cat# 100758, Biolegend, USA) to identify infiltrating T cells. Stained slides were visualized using an Olympus epi-fluorescence microscope.

Statistical analyses. The c indices generated by Monte Carlo cross-validation were performed using the Wilcoxon signed-rank test. This paired test was chosen because each method was evaluated using identical training/testing sets. Statistical analysis of Kaplan–Meier plots was performed using the Mantel-Cox log-rank test. Other statistical analysis comparing 2 groups was performed using the student's t-tests, comparing 3 or more groups was performed using one-way ANOVA. P-value was denoted as following: *$p<0.05$, **$p<0.01$, ***$p<0.005$, ****$p<0.001$.

Supplementary Discussion

Discussion. S1 Impact of CAF layer and tumor antigen on T cell infiltration and killing. Tumor stroma was reported to prohibit T cell infiltration via physical ECM barrier and secretory factors (e.g., chemokines, cytokines, and microRNA), and tumor with thicker stroma such as pancreatic cancer tends to permit less immune infiltration.([17](#)) However, it has been not systematically studied how tumor stromal components affect T cell penetration quantitatively. Thus, using our microfluidic platform, we fabricated tumor spheroids with stromal "shell" with tunable thickness by tuning tumor cell to CAF ratio. We found that CAF can effectively prohibit T cell infiltration and killing (**Fig. 2b**), Moreover, higher CAF composition resulted in less T cell infiltration, as well as reduced T cell killing. Interestingly, although fewer T cells can enter the spheroid in the higher CAF ratio conditions, the ones that did successfully enter the spheroid did not halt their infiltration until they reached the tumor core/CAF shell boundary, which still could perform killing (**Fig. S2**), suggesting the physical barrier created by CAF has the limited effect to prohibit T cell infiltration, which is consistent to that observed in patient tumor sections([18](#)). Another key factor to impact T cell infiltration is the tumor antigen presentation, as acquired ICI resistance can be partially attributed to antigen loss/downregulation of antigen presentation molecules as seen from patients' pathology slides.([19](#)) To study this phenomenon, we also tuned antigen-positive tumor cell ratio in our model to observe its impact on T cell infiltration. We mixed OVA+ UN-KC6141 cells and OVA- wild type UN-KC6141 cells at a ratio of 0, 25, 50, and 100%. We observed that T cell has minimum infiltration inside the 0 and 25% low OVA+ cell ratio spheroids, whereas T cell infiltration number and depth significantly increased in the 50% and 100% OVA+ group, indicating tumor antigen presentation is another key factor in T cell infiltration (**Fig. S3**). Surprisingly, the 50% OVA+ spheroids permit a slightly higher number and deeper T cell infiltration as compared with 100% OVA+ spheroids, despite more killing are seen with 100% OVA+ spheroids. This is likely due to the frequent pauses of T cells at the surface of 100% OVA+ spheroids to perform killing, limiting

its penetration (**Fig. S3**). This observation also highlights the importance of studying T cell infiltration as an independent event, as it does not always correlate positively with antigen presentation or T cells' capacity to kill.
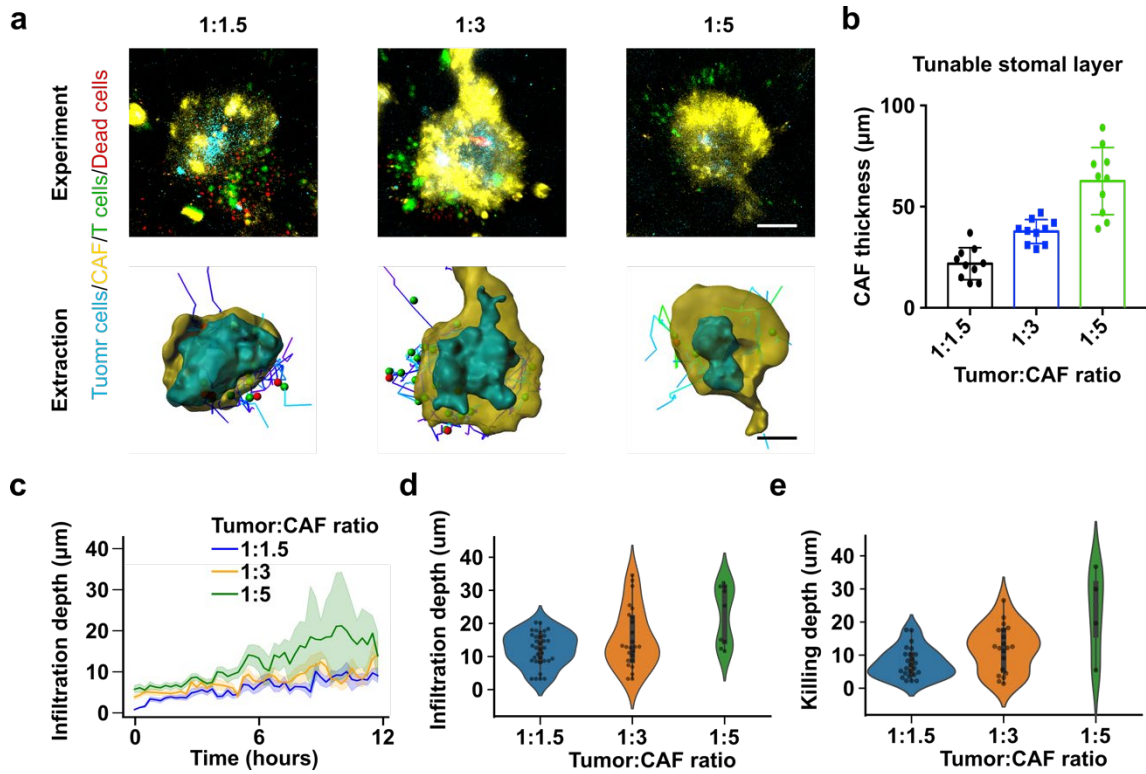
Discussion. S2 Development of the deep-learning-based TIL score analyzer. We trained the deep-learning-based TIL score analyzer using clinical data of pathology H&E images and survival data from 411 slides, 397 cases from The Cancer Genome Atlas (TCGA) Skin Cutaneous Melanoma (SKCM) projects.([20](#)) The TIL score is to best assign higher TIL scores to patients with better survival. Although cancer patient survival can be attributed to many factors such as treatment history and molecular status of the tumors in addition to TIL infiltration, abundant evidence have shown that patients' survival correlate well with TIL scores([21-23](#)). Thus, we attempted to develop the analyzer that projects TIL maps (e.g., TIL number, the relative depth, distribution, and clusters of TILs within tumors) to the patient survival. The core element of the deep learning algorithm consisted of convolutional neural network (CNN) modules, image sampling, and risk filtering to improve prediction accuracy and stability (**Fig.S4**). The TILs and tumor were first extracted from H&E-stained tissue sections to digitized images colored in red (TILs) and blue (tumor).([5](#)) These digitized images of TIL/tumor distribution were then used to train a deep convolutional network integrated with a discrete-time hazards model for predicting outcomes. The network consisted of interconnected residual blocks and nonlinear functions to transfer the images to highly predictive prognostic features. Fully connected layers perform additional transformations on these image-derived features, and a discrete-time hazard layer generates a prediction of the survival likelihood distribution,([8](#)) and then a SoftMax layer outputs the classification of 3-year survival potential. To improve the deep learning performance, a sampling and risk filtering technique was adapted to address intra-tumoral heterogeneity and limited clinical samples. In the training, new samples were randomly sampled from each slide image at the start of each training iteration, providing the CNN with a fresh look at each patient's TIL distribution and capturing heterogeneity within the slides. Each sample is preprocessed using data augmentation techniques that randomly transform the images to reinforce network robustness to tumor orientation and variations.([24](#)) For prospective prediction, we first took multiple samples within each slide to generate a representative batch of fields for each patient, and then, risk output from these samples were sorted and filtered to predict a more robust patient-level risk that reflects the aggressiveness of their disease. These sampling and filtering procedures were described in detail in Methods. The prognostic accuracy of our deep learning algorithm was assessed using Monte Carlo cross-validation. We randomly split our cohort into paired training (80%) and testing (20%) sets to generate 15 training/testing set pairs. We trained the model using each training set and then, evaluated the prognostic accuracy of these models on the paired testing sets, generating a total of 15 accuracy measurements. Using the sampling and filtering techniques, our deep learning system reached a median c index([25](#)) of 0.674 to predict the survival probability distribution. To access the performance of the classifier, we employed 10 times cross-validation to test the sensitivity and specificity. And the 3-year survival classification obtained from the survival probability to score the TIL patterns reached a median AUC of 0.8051 (**Fig.4Sc**), where AUC evaluates the performance of the classifier globally, an AUC of 0.8-0.9 means a good classification. Compared with current methods, there are two main improvements we've made: (1) Whereas the previous reference (27 in main text) calculated "Banfield Raftery" index ("count of TIL clusters") or

"Ball Hall" index ("cluster extend") from the extracted TIL map and analyzed the correlation of these 2 indexes with patient survival, we further developed this by directly inputting these TIL maps into a deep convolutional neural network and trained it to associate the TIL patterns with discrete patient survival time to generate a TIL score. This TIL score correlated better with patient survival as compared to any single parameter in the previous study including the "Banfield Raftery" index or "Ball Hall" index alone (**Fig.S5**). (2) This scoring algorithm in combination with our microfluidic platform allowed us to apply this to evaluate our in vitro screening TIL pattern images to find epigenetic drugs that resulted in better TIL infiltration patterns. Thus, with minimal pre-processing, we could directly input our TIL pattern images into the trained convolutional neural network and obtain the corresponding TIL scores.

Discussion. S3 Throughput of our system. Our platform can fabricate 80 spheroids per well in a 96-well plate format. In a typical drug screening experiment using 20 plates, we could achieve a throughput of 153,600 spheroids per screen in combination with a high-content imager. For T cell infiltration analysis, the throughput is limited by an additional factor: the limitation of high-speed imaging. In this study, our strategy was to perform an initial T-cell infiltration screening using a fast-scanning Olympus OSR spinning disk microscope allowing for 2-color (T cells: Green and tumor-CAF spheroids: Yellow) imaging of 100 positions with a 15-minutes imaging interval for 12 hours. This enabled high-dimensional time-series data recording of single T cell infiltration dynamics. Top candidates were then selected and validated with detailed interrogation of T cell infiltration and killing dynamics in detail using the 4-color (T-cell, CAF shell, tumor core, cell death indicator dye) Leica SP8 confocal microscope. Due to the limited scanning speed of our Leica SP8 confocal microscope set-up, we were only able to perform a detailed validation of 30 spheroids at a time with a 15-minutes imaging interval. Moreover, we only used the 12-hour end-point T cell infiltration position map for drug scoring in our experiments. In the future, the deep-learning-guided system can be developed to eliminate the prolonged time-lapse recording, and scan all 7,680 spheroids at the end, allowing for screening of 768 agents with 10 spheroids per drug or 1,536 agents with 5 spheroids per drug, greatly expand the screening throughput.

Fig. S1 Microfluidic fabrication of "core/tumor-shell/stroma" spheroid array. a. Formation of "core/tumor-shell/stroma" spheroids by sequential addition of tumor and CAF cells. b. Tumor spheroids' viability over the prolonged culture period. **b.** Tunable CAF layer by changing tumor cell UN-KC6141 and CAF ratio. **c.** Tunable tumor/CAF ratio resulted in variable stromal thickness in "core/tumor-shell/stroma" spheroids. **d.** Controlled initial seeding of ~9,000 cells per spheroid with pure 9,000 UN-KC6141 cells or 3,000 UN-KC6141 cells + 6,000 CAF cells resulted in differential spheroid growth. **e.** Distinct speed and track straightness was observed in killing and non-killing T cells. Speed and track straightness was calculated in Imaris software based on the Euclidean distance between T cell position displacement in consecutive imaging frames. Scale bar: 200 μm.

Fig. S2 Impact of the tunable stromal layer of tumor spheroids on T cell tumor infiltration and killing behavior. a. The "core/tumor-shell/stroma" spheroids with tunable CAF layer thickness were subjected to T cell infiltration (tumor cells: green, CAF cells: yellow). **b.** Tunable stromal layer thickness in the various tumor: CAF ratio conditions. **c.** T cell infiltration depth over time in heterotypic spheroids with 1:1.5, 1:3, and 1:5 tumor to CAF ratio. **d.** Total infiltrated T cell number and infiltration depth distribution in all 3 types of spheroids with various CAF ratios. **e.** Total killing events and killing depth distribution in all 3 types of tumor spheroids with various CAF ratios. Swarm plot width is scaled by event numbers. Scale bar: 50 μm.
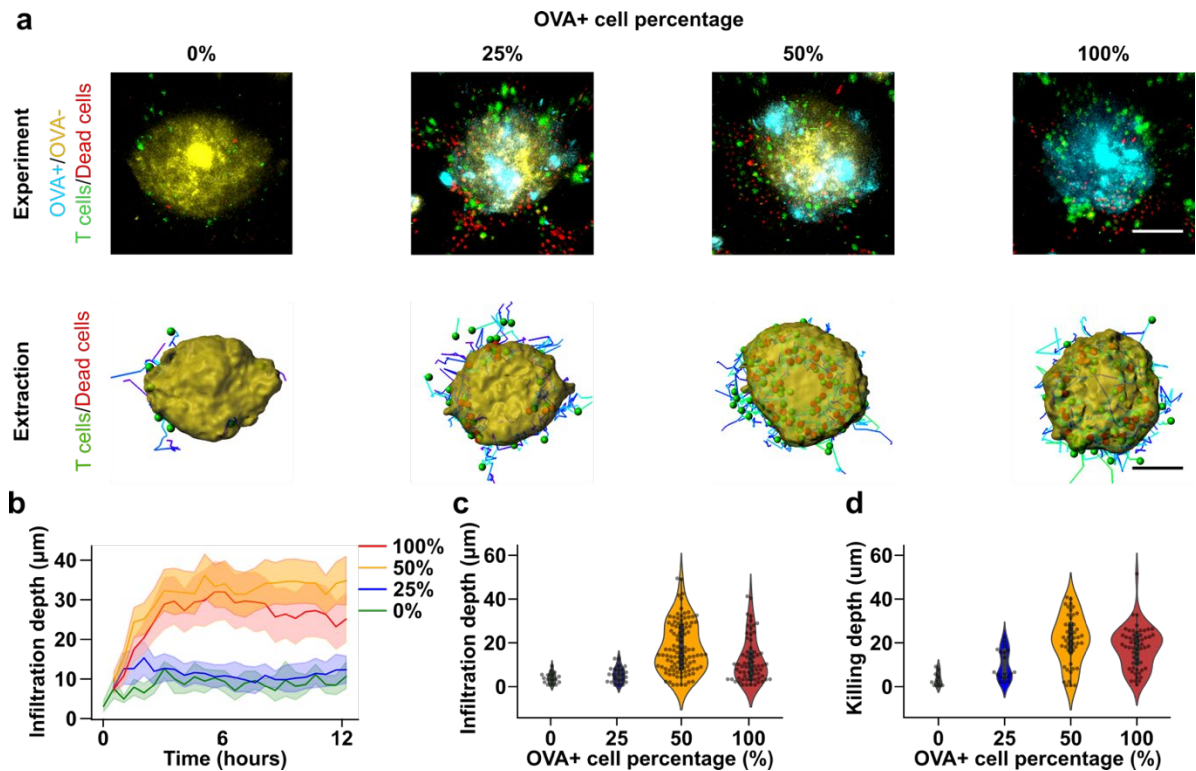
Fig. S3 Impact of tumor antigen presentation on T cell tumor infiltration and killing behavior. a. Tuning OVA+ versus OVA- cell ratio in "core/tumor-shell/stroma" spheroids to makeup OVA+ cell ratio as 0%, 25%, 50%, and 100%. **b.** T cell infiltration depth over time inside all four different types of spheroids with various OVA+ cell ratio. **c.** Total infiltrated T cell number and infiltration depth distribution in all 4 types of spheroids. 100% OVA+ spheroids have less infiltrating T cells likely due to the frequent pausing and killing of T cells during the 12-hour imaging period. **d.** Total killing events and killing depth distribution in all 4 types of spheroids. 100% OVA+ spheroids have less infiltrated T and infiltration depth yet more killing events as compared with 50% OVA+ group. Swarm plot width is scaled by event numbers. Scale bar: 30 μm.
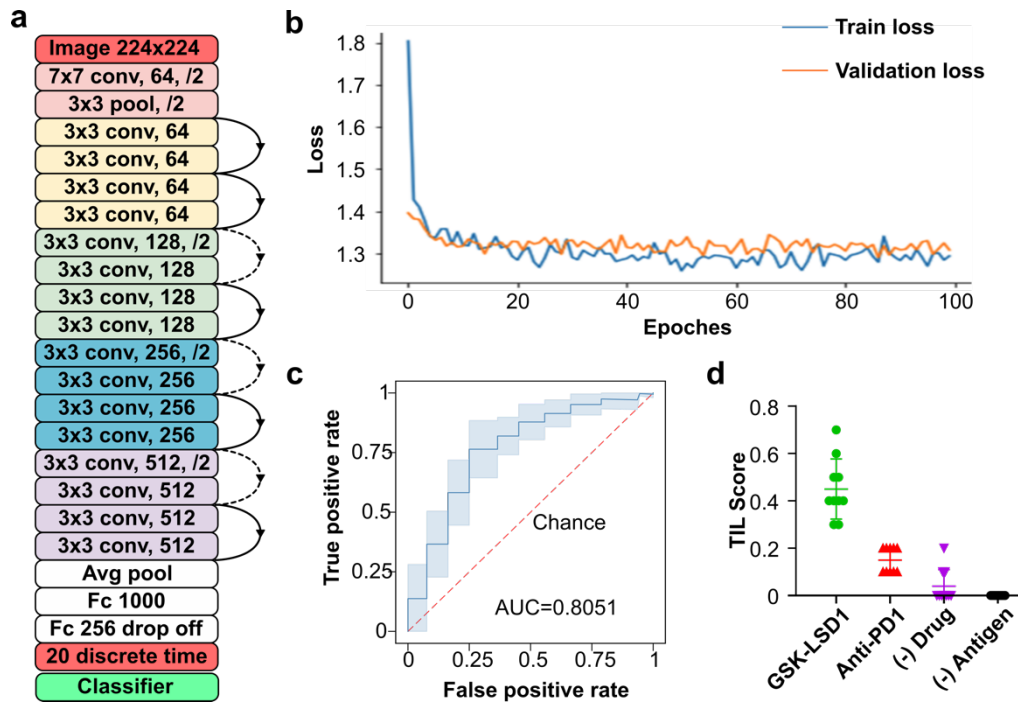
Fig. S4 The deep-learning-based TIL scoring system. a. Detailed diagram of the architecture. The architecture is a variation of the Resnet18 network with modified output layers. **b.** Representative training curves of training loss and validation loss. **c.** The cross-validated time-dependent ROC curve was generated by the classifier for 3-year survival predictions with an AUC of 0.8051. **d.** Predicted TIL score of different drugs (n=10).
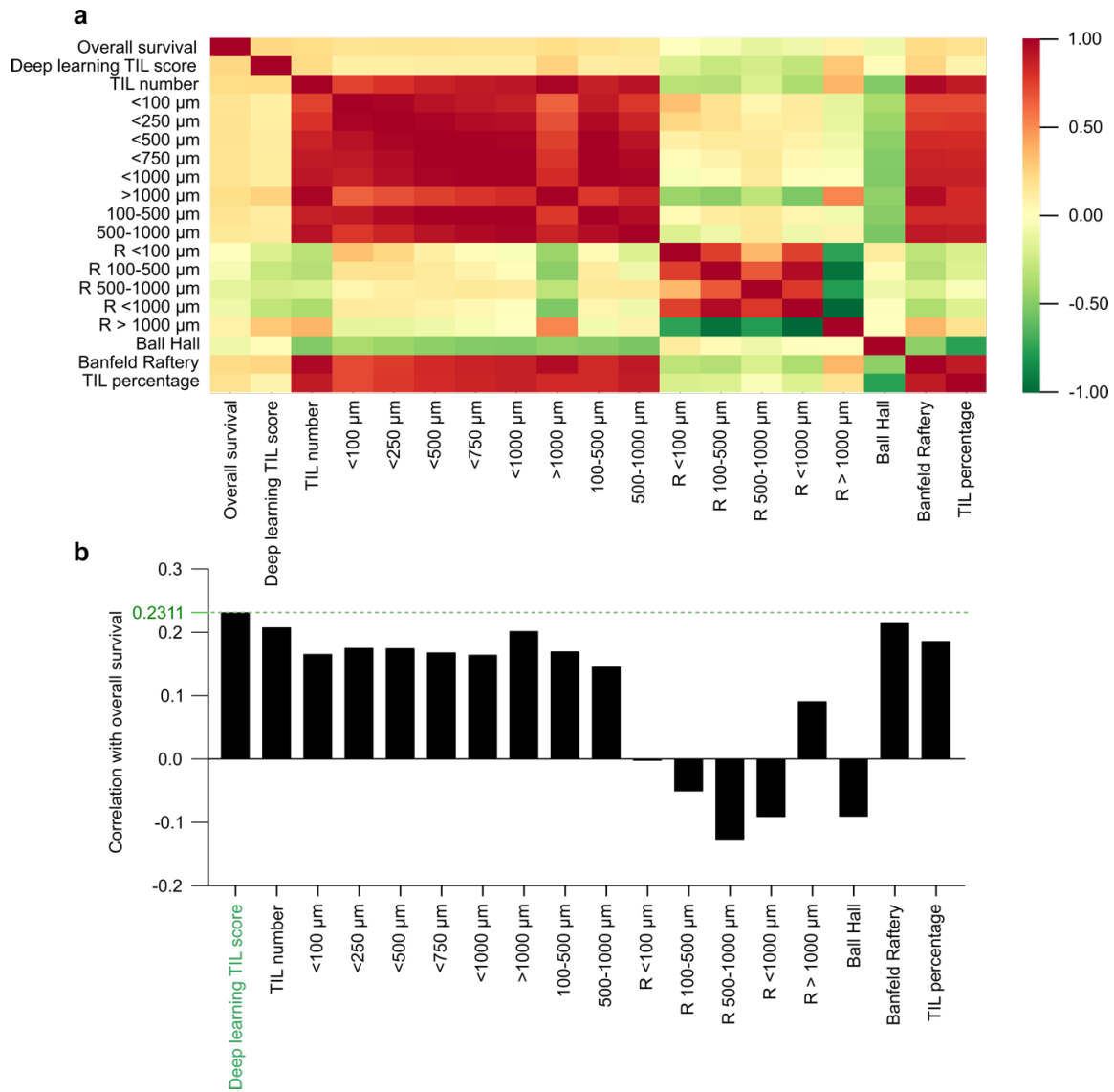
Fig. S5 Deep learning TIL scoring method is better than other methods based on single TIL parameter. a. Spearman's rank correlation coefficient of deep-learning-based TIL scores and other single T cell infiltration parameters with patients' overall survival (OS). **b.** Comparison of deep learning TIL scores spearman's rank correlation coefficient with other T cell single-parameter based readouts from TCGA skin cancer pathology slides to predict patients' overall survival.

Epigenetic drug toxicity screening
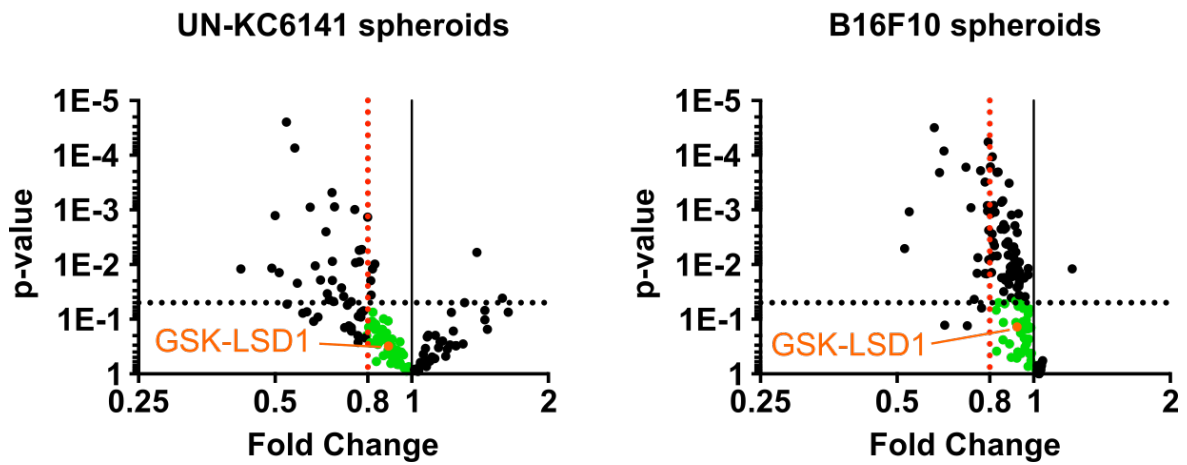
Fig. S6 Cytotoxicity screening of epigenetic drugs. UN-KC6141 spheroids and B16F10 spheroids were treated with epigenetic drug at 5 µM for 24 hours and cytotoxicity was measured by CCK-8 assay. Drugs without cytotoxicity (cell viability>80% and p-value>0.05) highlighted by green were chosen for T cell infiltration screening.
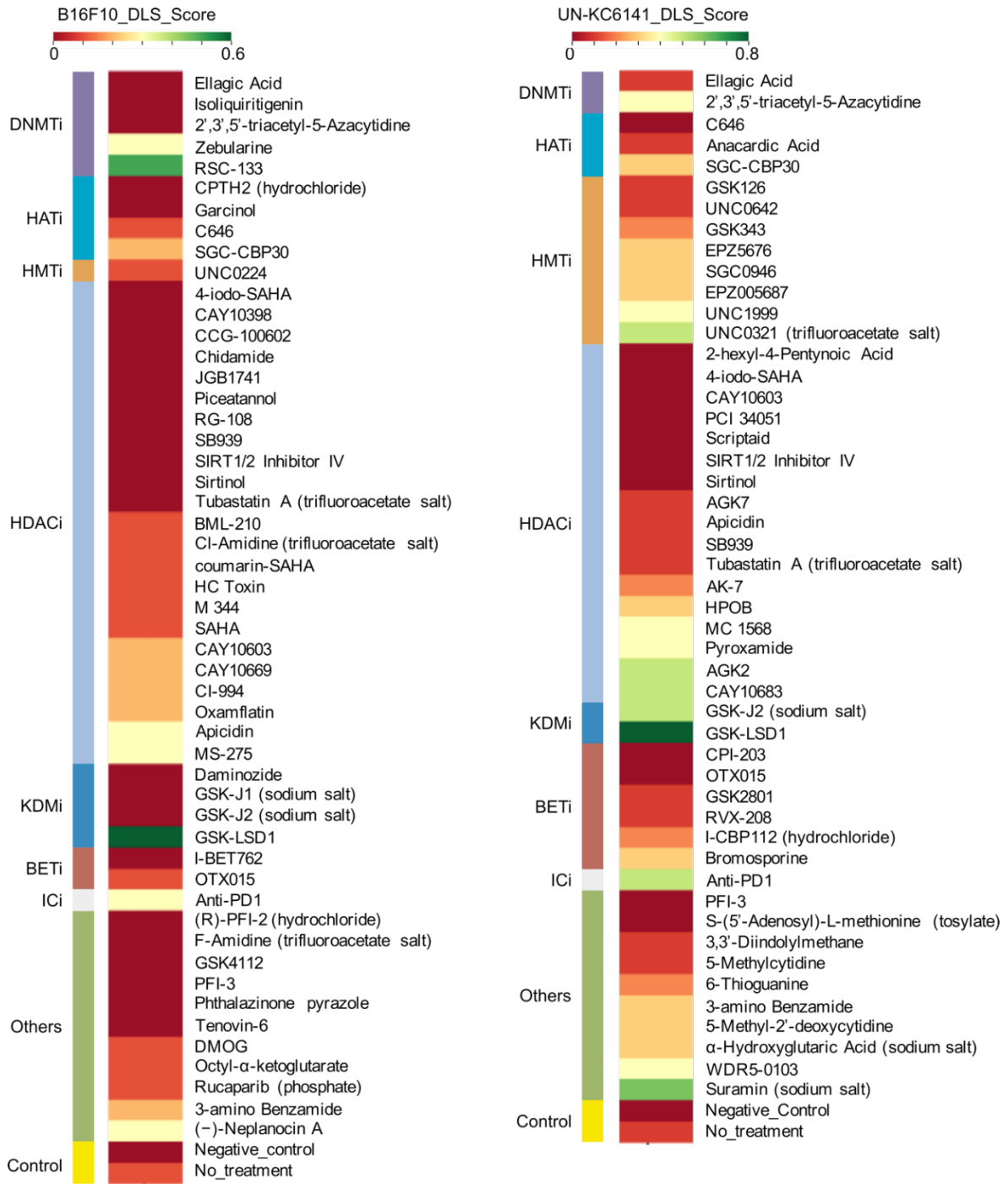
Fig. S7 Drug cores of epigenetic drug library in "core/tumor-shell/stroma" spheroids (B16F10 and UN-KC6141). Negative control refers to a score of OVA negative cell spheroids against T cells (OVA-specific CD8+ T cells). No treatment control refers to OVA positive cell spheroids against T cells (OVA-specific CD8+ T cells).
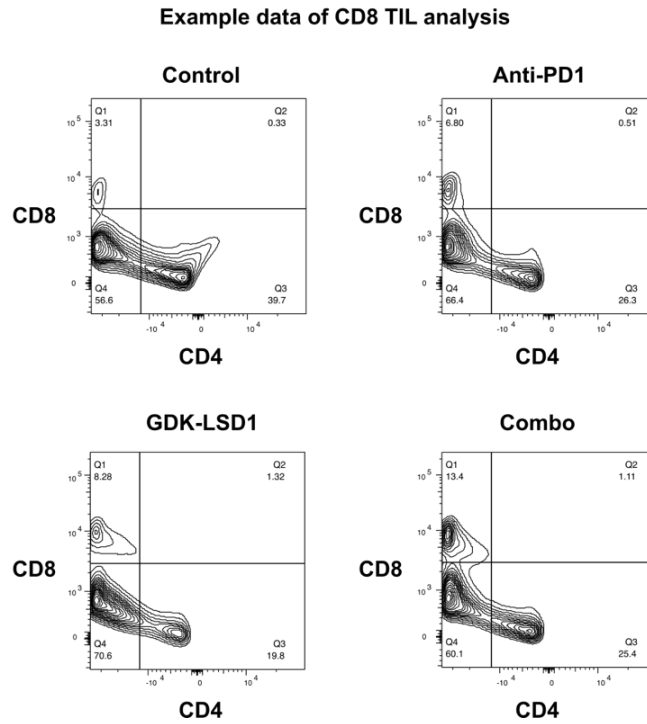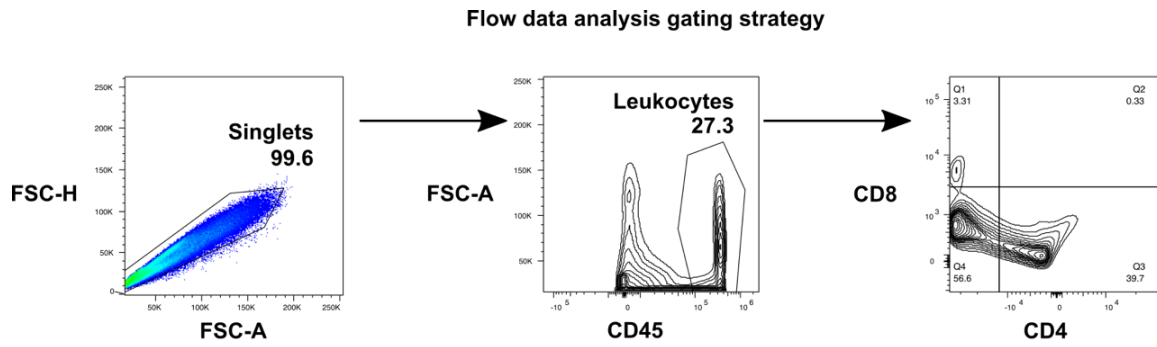
Fig. S8 Flow cytometry gating strategy for quantification of tumor infiltrating CD8+ T cells. The total tumor cell population was first gated for FSC-H and FSC-A for single cells, then CD45+ cells were selected and further gated into CD4+ and CD8+ T cells.

Supplementary Movies
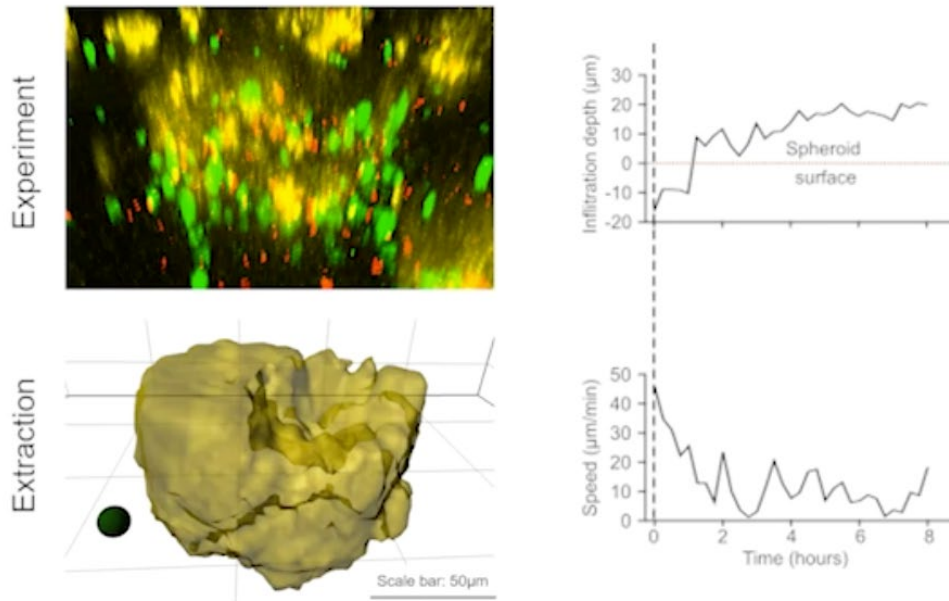
Movie. S1 Tracking of simultaneous T cell infiltration and killing. An individual T cell's killing of two tumor cells is recorded as the T cell (green) infiltrates into "core/tumor-shell/stroma" spheroids (yellow surface) and kills (red). Track color scale: timestamped.
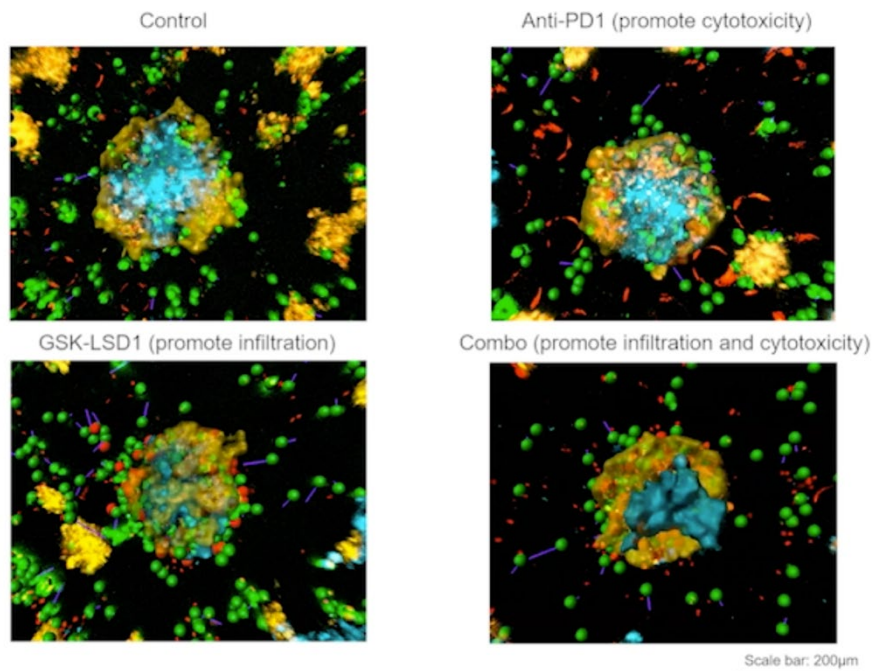
Movie. S2 Effect of treatments on T cell infiltration. T cell infiltration and killing of tumor cells inside the "core/tumor(cyan)-shell/stroma(yellow)" spheroids. Videos show different T cell behaviors under control (no treatment), anti-PD1 treatment, GSK-LSD1 treatment, and combo (anti-PD1+ GSK-LSD1) treatments. Track color scale: timestamped.

References

1. Y. Wang *et al.*, In situ differentiation and generation of functional liver organoids from human iPSCs in a 3D perfusable chip system. *Lab on a Chip* **18**, 3606-3616 (2018).
2. B. Chen *et al.*, High-throughput acoustofluidic fabrication of tumor spheroids. *Lab on a Chip* **19**, 1755-1763 (2019).
3. Z. Ao *et al.*, Rapid Profiling of Tumor‐Immune Interaction Using Acoustically Assembled Patient‐Derived Cell Clusters. *Advanced Science* **9**, 2201478 (2022).
4. Z. Ao *et al.*, Evaluation of cancer immunotherapy using mini-tumor chips. *Theranostics* **12**, 3628 (2022).
5. J. Saltz *et al.*, Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Reports* **23**, 181-193.e187 (2018).
6. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Ieee, "Deep Residual Learning for Image Recognition" in 2016 Ieee Conference on Computer Vision and Pattern Recognition. (Ieee, New York, 2016), 10.1109/cvpr.2016.90, pp. 770-778.
7. C. C. Brown, On the Use of Indicator Variables for Studying the Time-Dependence of Parameters in a Response-Time Model. *Biometrics* **31**, 863-872 (1975).
8. E. Biganzoli, P. Boracchi, L. Mariani, E. Marubini, Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Statistics in Medicine* **17**, 1169-1186 (1998).
9. M. F. Gensheimer, B. Narasimhan, A scalable discrete-time survival model for neural networks. *PeerJ* **7**, e6257-e6257 (2019).
10. H. Kvamme, Ø. Borgan, Continuous and discrete-time survival prediction with neural networks. *arXiv preprint arXiv:1910.06724* (2019).
11. L. Bottou (2010) Large-Scale Machine Learning with Stochastic Gradient Descent. in *Proceedings of COMPSTAT'2010*, eds Y. Lechevallier, G. Saporta (Physica-Verlag HD, Heidelberg), pp 177-186.
12. A. Krizhevsky, I. Sutskever, G. E. Hinton (2012) Imagenet classification with deep convolutional neural networks. in *Advances in neural information processing systems*, pp 1097-1105.
13. Z. Ao *et al.*, Controllable fusion of human brain organoids using acoustofluidics. *Lab on a Chip* **21**, 688-699 (2021).
14. H. Cai *et al.*, Intelligent acoustofluidics enabled mini-bioreactors for human brain organoids. *Lab on a Chip* **21**, 2194-2205 (2021).
15. Z. Ao *et al.*, One-stop microfluidic assembly of human brain organoids to model prenatal cannabis exposure. *Analytical chemistry* **92**, 4630-4638 (2020).
16. Z. Ao *et al.*, Understanding Immune-Driven Brain Aging by Human Brain Organoid Microphysiological Analysis Platform. *Advanced Science* **9**, 2200475 (2022).
17. J. Watt, H. M. Kocher, The desmoplastic stroma of pancreatic cancer is a barrier to immune cell infiltration. *Oncoimmunology* **2**, e26788 (2013).
18. X. Li *et al.*, Infiltration of CD8(+) T cells into tumor cell clusters in triple-negative breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 3678-3687 (2019).

19. V. Anagnostou *et al.*, Evolution of neoantigen landscape during immune checkpoint blockade in non–small cell lung cancer. *Cancer discovery* **7**, 264-276 (2017).

20. N. Cancer Genome Atlas Research *et al.*, The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**, 1113-1120 (2013).

21. A. Ladányi, Prognostic and predictive significance of immune cells infiltrating cutaneous melanoma. *Pigment cell & melanoma research* **28**, 490-500 (2015).

22. G. Erdag *et al.*, Immunotype and immunohistologic characteristics of tumor-infiltrating immune cells are associated with clinical outcome in metastatic melanoma. *Cancer research* **72**, 1070-1080 (2012).

23. J. Haanen *et al.*, Melanoma-specific tumor-infiltrating lymphocytes but not circulating melanoma-specific T cells may predict survival in resected advanced-stage melanoma patients. *Cancer immunology, immunotherapy* **55**, 451-458 (2006).

24. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **60**, 84-90 (2017).

25. L. Antolini, P. Boracchi, E. Biganzoli, A time-dependent discrimination index for survival data. *Statistics in Medicine* **24**, 3927-3944 (2005).