Supplementary Materials for

Oncostreams: Self-organization and dynamics determine spatial heterogeneity and malignant behavior in gliomas

Andrea Comba[1,2,4], Sebastien Motsch[3], Patrick J. Dunn[1,2,4], Todd C. Hollon[1], Daniel B. Zamler[1,2,4], Anna E. Argento[1,2], Celina G. Kleer[4,5], Maria G Castro[1,2,5], and Pedro R. Lowenstein[1,2,5,*]

*Correspondence to: pedrol@umich.edu

This file includes:

Materials and Methods

Figures S1-S17

Tables S1, S3 and S4

Other Supplementary Materials:

Movies S1 to S8

Table S2

Materials and Methods

Glioma cell lines and culture conditions:

Mouse glioma cells (NPA, GL26) and human glioma cells (MSP-12, SJGBM2) were maintained at 37 °C with 5% CO2 and their respective media as described before (*14, 15*) Mouse NPA (N-Ras, shp53-GFP and shATRX-GFP) neurospheres were derived from genetically engineered tumor using the Sleeping Beauty (SB) transposase system as previously described (*14, 18*). Mouse GL26 glioma cell were generated by Sugiura K and obtained from the frozen stock maintained by the National Cancer Institute (Bethesda, MD). MSP-12 human glioma cell lines were provided by Christine Brown, City of Hope, and SJGBM2 human glioma cells were provided by Children's Oncology Group (COG) Repository, Health Science Center, Texas Tech University.

Intracranial implantable syngeneic mouse gliomas:

Glioma tumors were generated by stereotactic intracranial implantation into the mouse striatum of 3.0 x 104 mouse glioma cells (either, NPA, or, GL26) in C57BL/6 mice and human glioma cells in immune-deficient NSG mice (SJGBM2) as described before (14*, 15, 27*) . To test whether oncostreams tumor cells move other cells throughout the tumor we generate a co-implantation glioma model by intracranial implantation of high malignant GL26-citrine cells with low aggressive human MSP12 glioma cells at a ratio of 1:30 (1,000 GL26-citrine cells and 30,000 MSP12 cells) in immune-deficient NSG mice. In addition, NSG mice were implanted with 30,000 MSP12 cells alone or 1,000 GL26-citrine cells alone as controls.

Experiments were conducted according to the guidelines approved by the Institutional Animal Care (IACUC) and Use Committee at the University of Michigan. Stereotactic implantation was performed as previously described (14).

Organotypic brain slice culture glioma model and time-lapse confocal imaging:

For the analysis of the dynamic of gliomas, we generate a 3D organotypic slice culture glioma model, by intracranial implantation of 3 x 10^4 NPA neurospheres. C57BL6 animals were used for glioma cell movement analysis in the tumor core and B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J- transgenic mice from Jackson laboratory (STOCK 007676) were used to study the behavior of tumor cells at the tumor: brain interface. Mice have euthanized at 19 days post-implantation or an IVIS detection signal of 10^6 (photon/s). The brain was removed, dissected,

and embedded in a 4% solution of low melting temperature agarose and kept on ice for 5 minutes. Solidified embedded brains were submerged in ice-cold and oxygenated media (DMEM High-Glucose without phenol red, 21063045, GibcoTM, USA) and sectioned in a Leica VT100S vibratome (Leica, Buffalo Grove, IL) set to 300 μm in the z-direction. 300 μm thick brain tumor sections were transferred to laminin-coated Millicel Cell Culture Insert (PICM0RG50, Millipore Sigma, USA) placed into a Nunc glass base dish, 27mm diameter (150682, Thermo Scientific) containing culture medium (D-MEM F-12 media supplemented with 25% FBS, Penicillin-Streptomycin 10.000 U/ML). All steps were performed under sterile conditions in a BSL2 laminar flow hood. Tumor slices were then maintained at 37 °C with a 5% CO2 atmosphere for 24 hours. Then, the medium was changed to normal NPA medium (DMEM-F12 media supplemented with B27 2%, N2 1%, Normocin 0.2 %, Penicillin-Streptomycin 10.000 U/ML and growth factors EGF and FGF 20 ng/ml). For time-lapse imaging of tumor cell migration slices were placed in the incubator chamber of a single photon laser scanning confocal microscope model LSM 880 (Carl Zeiss, Jena, Germany). at 37 °C with a 5% CO2. Imaging was obtained every ten minutes for 100-300 cycles. Following movie acquisition, sections were fixed in 4% paraformaldehyde (PFA) for 2 days. Fixed sections were embedded in 2% agarose. For H&E and immunohistochemistry analysis sections were then processed and embedded in paraffin at the University of Michigan Microscopy & Image Analysis Core Facility using a Leica ASP 300 paraffin tissue processor/Tissue-Tek paraffin tissue embedding station (Leica, Buffalo Grove IL).

Mathematical analysis of tumor cell movement

To determine the movement of cells in different areas of the tumor we performed localized statistics analysis in different zones and on a given time interval. We selected localized areas based on the organization of cells in clusters, group of cells that are nearby and seem to move together with similar distribution. Raw data of 4 movies from the tumor core and 4 movies from the tumor border were analyzed for 293 cycles (core) and 186 cycles (border) for a frame rate of $\Delta t = 10$ min per image. To track the evolution of cells motion, we used the software Fiji with the plugin Track-Mate. We used as parameters for the cell size (called 'blob') 20μm and a threshold of 1 together with the DoG method (Difference of Gaussian detectors). We then obtain for each experiment several paths for many different cells. To reduce the erratic behavior, we filtered the trajectories. From the data obtained with TrackMate, we then smoothed out the paths which

allowed to estimate the velocity of each cell at any time (**Fig. S1A**). We simply use a Gaussian Kernel as a filter (with standard deviation σ = 2 and a stencil of 9 points). Thus, we had an estimation of the positions xi(t) and velocities vi(t) of the cells i. From the velocity vi(t), we also deduced the velocity direction θi (t).

After obtaining these parameters we applied several statistics to investigate cell behavior (**Fig. S1B**). Each cell is characterized by a position xi Є R2 and a velocity vi Є R2. The velocity vector v can be decomposed into speed (scalar) c = |v| and angle direction θ such that v = c (cos θ, sin θ). We utilized three statistics to visualize the distribution of velocity: (i) distribution of velocity angle θ (provides indication of the overall direction of the cells); (ii) distribution of speed c and (iii) distribution of the velocity vector v = (vx, vy). Since v is a vector, we need to plot a 'heatmap' to visualize its distribution (**Fig. S1B**).

To understand the organization of oncostream dynamics we analyzed the cell-cell correlation using two further statistics using pair wise information: (i) Neighbor density: taking a reference cell xi, we estimate the probability to have another cell xj nearby, and, (ii) velocity correlation: depending on the position of a nearby xj, estimate the correlation: $\omega i \cdot \omega j$ where ωi (resp. ωj) is the (normalized) velocity direction of cell i (resp. j) (**Fig. 1K and Fig. S1, C to E**). A correlation of +1/-1 indicates that cells are moving in the same/opposite direction, whereas 0 indicates that they move in orthogonal direction. Orthogonal direction is when $\omega i \cdot \omega j = 0$. Details on how to estimate numerically the correlation functions are described in Supplementary Materials and summarized in **Fig. S1 C to E**.

Classification of glioma migration patterns

To classify the collective cellular motion behavior of the three type of patterns called flock, stream and swarm illustrated in **Fig. 1A** we used as criteria the orientation of each cell described by its unique velocity angle denoted θi. More precisely, we transformed the Angle Velocity Distribution graph (**Fig. 1G**) into a histogram (**Fig. 1I**) where we examined the distribution of all the values θi. A schematic representation of these distributions is depicted in **Fig. 1H and 1I**. Considering a data-set θn n=1…N of orientations where N is the total number of cells, $\theta n \in [0, 2\pi]$ is the direction of the cell n. We tested three types of distributions ρ to describe the dataset and give a likelihood in each case.

Flock: wrapped normal distribution: In a flock, the distribution of orientation should be *uni-modal* and thus it can represented as a *wrapped* normal distribution ρ_{flock}:

$$\rho_{flock}(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-\frac{|\theta-\mu+2\pi k|^2}{2\sigma^2}},$$

where μ and σ are respectively the mean and standard deviation of the distribution. The 'peak' of the distribution is at μ whereas σ measures the spreading of the distribution. The summing in n is necessary to ensure that the distribution is $2\pi-$periodic.
The estimation of the two parameters μ and σ are obtained by maximizing the *likelihood*, i.e. the probability to observe the data-set $\{\theta_n\}_n$ given the distribution ρ_{flock}:

$$\max_{\mu,\rho} \mathcal{L}(\mu,\rho) \qquad \text{with} \quad \mathcal{L}(\mu,\rho) = \prod_{n=1}^{N} \rho_{flock}(\theta_n). \tag{2.2}$$

After some standard computations (taking the log), we find an explicit expression of the parameters (2.3):

$$\mu = \mathrm{Arg}(\bar{\mathbf{z}}) \quad , \quad \sigma^2 = -\ln R_e^2$$

with

$$\bar{\mathbf{z}} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{i\theta_n} \text{ (using complex number)}, \; R_e^2 = \frac{N}{N-1}\left(|\bar{\mathbf{z}}|^2 - \frac{1}{N}\right).$$

Stream: symmetric wrapped normal distribution: In a stream, the distribution of orientation is supposed to be $\pi-$periodic (same distribution of cells moving in both direction). Thus, we parametrize the distribution with a symmetric version of the expression (2.4):

$$\rho_{stream}(\theta) = \frac{1}{2}\Big(\rho_{\mu,\sigma}(\theta) + \rho_{\mu,\sigma}(\theta+\pi)\Big) \tag{2.4}$$

where $\rho_{\mu,\sigma}$ is defined in (2.4). The estimation of the two parameters μ and σ can be estimated as previously since

$$\begin{aligned}\rho_{stream}(\theta) &= \frac{1}{2\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \left(\mathrm{e}^{-\frac{|2\theta-2\mu+4\pi k|^2}{2(2\sigma)^2}} + \mathrm{e}^{-\frac{|2\theta-2\mu+4\pi k+2\pi|^2}{2(2\sigma)^2}}\right) \\ &= \frac{1}{(2\sigma)\sqrt{2\pi}} \sum_{\tilde{k}=-\infty}^{\infty} \mathrm{e}^{-\frac{|2\theta-2\mu+2\pi\tilde{k}|^2}{2(2\sigma)^2}} = \rho_{2\mu,2\sigma}(2\theta).\end{aligned}$$

Thus, the parameters μ and σ are estimated using (2.3) with $\bar{\mathbf{z}}_{\mathbf{2}} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{i2\theta_n}$ in lieu of $\bar{\mathbf{z}}$.

Swarm: uniform distribution: In a swarm, the orientation is supposed to be 'random' meaning that all directions are equally probable. Thus, the distribution is constant.

Model selection: Given a data-set $\{\theta_n\}_{n=1..N}$, we need to find a criteria to determine whether the underlying distribution is *more like* a flock, a stream or a swarm. A first method consists in comparing the likelihood (2.2) for each model after we have maximized the parameters μ and σ. But this method will never select a swarm since there is no parameter involved and a flock or stream distribution can be made uniform with σ large. One can compensate using information criterion

(e.g. AIC, BIC) which penalized adding parameters. However, in practice, swarm distributions will still not be selected, the model selection will pick a flock or a stream with a large σ.
For this reason, we *fix* a priori the standard deviation of the flock and stream distribution. In other words, a distribution will be considered a flock if the distribution is not "too" flat. We choose:

$$\sigma^2_{flock} = 2 \quad , \quad \sigma^2_{stream} = 1. \tag{2.5}$$

We give an illustration with the experiment "scene 5" zone D in figure 3. The associated log-likelihood gives:

$$\ln(\mathcal{L}_{flock}) = -46798.93, \;\; \ln(\mathcal{L}_{stream}) = -43583.16, \;\; \ln(\mathcal{L}_{swarm}) = -43732.28.$$

Thus, in this example, we select a stream formation since the log-likelihood is the highest.

Genetically engineered mouse glioma models (GEMM):

We used genetic engineered mouse glioma models for survival analysis and histopathological analysis. Murine glioma tumors harboring different genetic drivers were generated using the Sleeping Beauty (SB) transposon system as described before (*14, 15, 28*). Genetic modifications were induced in postnatal day 1 (P01) male and female wild-type C57BL/6 mice (Jackson Laboratory), according to IACUC regulations. Genetic models generated include the following gene expression/inhibitions: (i) shp53, NRAS and shATRX **(NPA)**, (ii) shp53, NRAS, shATRX and IDH1-R132H **(NPAI),** (iii) shp53, NRAS and PDGFβ **(NPD).** Plasmid sequences were verified by Sanger sequencing. Plasmid sequences described below were used to generate tumors: (i) pT2C-LucPGK-SB100X, transposon & luciferase expression; (ii) pT2-NRASSV12, NRAS expression; (iii) pT2-shp53-GFP4, p53 knock-down; (iv) pT2-shATRx-GFP4, ATRX knock-down; (v) pKT-IDH1(R132H)-IRES-Katushka, mIDH1 expression; (vi) pT2-shp53-PDGFβ-GFP4, p53 knock-down in combination with PDGFβ ligand overexpression. The pT2CAG-NRASV12 and pT2-shp53-GFP4 plasmids were the generous gift of Dr. John Ohlfest (University of Minnesota). Mice were injected with the necessary plasmids according to the protocol described (*1, 2*). Plasmid uptake in pups and tumor growth was monitored by IVIS® Spectrum imaging. Adult mice displaying symptoms of morbidity were transcardially perfused (*14, 15*).

Analysis of Oncostreams in human glioma tissue

Oncostream presence was analyzed in H&E sections of paraformaldehyde-fixed paraffin-embedded (PFPE) human glioma samples obtained from primary surgery from the (The University

of Michigan Medical School Hospital and TCGA database). To determine the presence of Oncostream in a large cohort of human glioma tissues we used the biospecimens of 'The Cancer Genome Atlas Research Network" (TCGA) data from the Genomic Data Commons Data Portal, National Cancer Institute, NIH (https://portal.gdc.cancer.gov). We analyzed primary Glioblastoma multiforme (TCGA-GBM) and Low-Grade Glioma (TCGA-LGG) database. We selected cases that have available the Slide Image and diagnostic Slide. The diagnostic slides are available for TCGA-GBM: 389 patients and TCGA-LGG: 491 patients. The presence of Oncostreams was scored on 100 TCGA-GBM Grade IV tissue samples and 120 TCGA-LGG samples. Oncostream presence was analyzed on tumors classified by histology grade (grade II, III and IV) as it is shown in **Table S1.** H&E histology samples were analyzed at high magnification using the Slide Image Viewer from the data portal. Histological material containing brain tumors and oncostreams of both rodent and human gliomas were evaluated by CGK (board certified pathologist), AC and PRL. Concordance on the diagnosis regarding the presence or absence of oncostreams, between the three evaluators, was >90%. Due to the need for good quality morphological preservation, only paraformaldehyde-fixed paraffin-embedded (PFPE) sections were used. Clinical data including age, sex, pathology, survival, treatment information, and data including information of MGMT DNA methylation status, IDH1 mutation status, G-CIMP DNA methylation status, were obtained from http://firebrowse.org, http://gliovis.bioinfo.cnio.es and https://portal.gdc.cancer.gov data portals (**Table S2**).

Cell aspect ratio and alignment analysis in H&E tumor sections

Images were obtained using bright-field microscopy of H&E stained paraffin sections (Olympus BX53 Upright Microscope from Olympus). Tumors were imaged using 40X and 20X objectives. Images were processed using the program ImageJ. Briefly, image processing included color deconvolution to isolate nuclei, transforming images to 8-bit, and adjusting the threshold to remove background. Elliptical overlay masks were then imposed over the nuclei and matching their shape. The shapes of these masks were then analyzed for the shape descriptors aspect ratio and circularity. For cell alignment analysis images were processed using ImageJ. Briefly, Image processing included creating a ROI around the stream area and the non-stream area. These ROIs were then saved as separate images. Both images underwent color deconvolution to isolate nuclei stain, 8-bit conversion, and threshold adjustment to remove background. Images were then analyzed for the

Feret's angle of the nucleus. The Feret's angle is the angle (0º-180º) that is taken from the x-axis to a line parallel with the longest distance across the particle being considered. Histograms were generated with these data in MATLAB using the Matplotlib plugin.

Immunohistochemistry on paraffin embedded brain tumors (IHC-DAB):

This protocol was performed as described before (*14*). Briefly, following perfusion, mouse brains were fixed in 4% paraformaldehyde (PFA) for an additional 48 hours at 4 °C. Brains were then processed and embedded in paraffin at the University of Michigan Microscopy & Image Analysis Core Facility using a Leica ASP 300 paraffin tissue processor/Tissue-Tek paraffin tissue embedding station (Leica, Buffalo Grove IL). Tissue was sectioned using a rotary microtome (Leica) set to 5 μm in the z-direction. Endogenous Peroxidase Quenching was completed through a 0.3% H2O2 incubation for 5 minutes at room temperature. Heat-induced antigen retrieval was performed using 10mM Citric Acid, 0.05% Tween 20, pH 6.0. Tissue permeabilization and blocking were completed using PBS 0.2% Tween-20 with 5% goat serum for one hour at room temperature. Tissue was incubated with primary antibodies at 4º C overnight at concertation detailed in **Table S4.** Tissue sections were then incubated with secondary biotinylated antibodies at a 1:1000 dilution in PBS with 0.2% Tween-20 overnight at 4º C. ABC Avidin-Biotin-Complex Binding reagent (Vectastain Elite ABC kit) and Betazoid DAB Chromogen detection kit (BioCare BDB2004) were used according to the manufacturer's instructions. Images were obtained using a bright-field from five independent biological replicates (Olympus BX53 Upright Microscope from Olympus). Ten different fields of each section were selected at random for study to include heterogeneous tumor areas.

Immunofluorescence on paraffin embedded sections from brain tumors:

This protocol was performed as described before (*14*). Paraformaldehyde fixed and paraffin-embedded tissues were sectioned and then de-paraffinized and re-hydrated. Heat-induced antigen retrieval was performed using 10mM Citric Acid, 0.05% Tween 20, pH 6.0. Permeabilization was performed using 0.5% TritonX-100 in PBS for 30 minutes at room temperature while shaking. Tissue sections were blocked in 10% horse serum and 3% BSA in PBS. Primary antibodies were incubated at 4˚C overnight in a humid chamber in 3% BSA in PBS at concertation detailed in **Table S4.** Tissues were then incubated for 90 minutes at room temperature with Alexa Fluor™

488 or Alexa Fluor™ 594 conjugated secondary antibodies (Invitrogen, Thermofisher Scientific). Nuclei were stained with DAPI (1:1000) in PBS for 5 minutes. Images were acquired with a laser scanning confocal microscope (LSM 880, Axio Observer, Zeiss, Germany).

Immunohistochemistry on vibratome brain tumor sections

Brains were left in 4% paraformaldehyde fixation for 48 hours and then moved to PBS 0.1% sodium azide for an additional 24 hours at 4°C. A Leica VT100S vibratome was used to obtain 50 µm coronal brain sections. The immunohistochemistry protocol was performed as previously described (28). Briefly, vibratome sections were permeabilized in TBS-Triton-X 0.1% for 60 minutes. Antigen retrieval was performed in 10mM sodium citrate. Non-specific antibody binding was blocked with 10% goat serum in TBS-Triton-X 0.1% for 1 hour at room temperature. Brain sections were then incubated with primary antibody diluted in TBS-Triton-X 0.1%, 1% goat serum, and 0.1% sodium azide for 24hours at RT, in the dark. Sections were then washed 6 times in TBS-Triton-X 0.1% and then incubated with the secondary antibody diluted in 1% goat serum in TBS-Triton-X 0.1% for 24hrs at RT, in the dark. Finally, sections were washed 6 times and incubated with 5µg/ml of 4',6-diamidino-2-phenylindole (DAPI) (Life technologies, D21490) in PBS for 5 minutes. Sections were washed again 3 times and mounted on microscope slides with prolong gold anti-fade reagent (Invitrogen, P36930).

Deep learning analysis for oncostreams detection on H&E staining of glioma tissue

A fully convolutional neural network (fCNN) was trained in order to identify and segment oncostreams in histologic images (*30*). We implemented a U-Net architecture to provide semantic segmentation of glioma specimens using deep learning (31). Our *oncostream* dataset consisted of images from sacrificed mice and open-source images from The Cancer Genome Atlas (TCGA). A total of 109 hematoxylin and eosin (H&E) stained histologic mice images and 64 from TCGA were reviewed and oncostreams were manually segmented by the study authors (AC, A.E.A and P.R.L.). Images from both datasets were then augmented by randomly sampling regions within each image to generate unique patches (~ 300 patches/image). The location and scale of each patch was randomly chosen to allow for oncostream segmentation to be scale invariant. First, using the mice dataset only, six iterations of training/validation set splits (80%/20%) were generated by randomly sampling unique, non-overlapping regions from each of labelled histologic images and

used for model selection/hyper-parameter tuning and model testing, respectively. Our fCNN was trained for binary classification of foreground (oncostream) and background (non-oncostream) using a binary cross-entropy loss function. Both pixel-level classification accuracy and intersection over union (IOU) were used as metrics to evaluate model performance. The model was trained using the Adam optimizer with an initial learn rate $\alpha = 0.0001$, $\beta1 = 0.9$, $\beta2 = 0.999$, and $\varepsilon = 10E-7$ (*33*). We used a scheduled learning rate decrease such that the rate was halved every 10 epochs and the model was trained for a total of 75 epochs for each iteration. The best performing model on the mouse dataset was then used as a pre-trained fCNN to initialize training on the TCGA data. We used both mouse and TCGA images for fCNN training with 20% of the TCGA data held out for model validation. Our fCNN was implemented using the model-level Python-based API, Keras (version 2.2.0), with a TensorFlow (*32*) (version 1.8.0) backend running on two NVIDIA GeForce 1080 Ti graphical processing units.

Statistical Analysis:

All *in vivo* experiments were performed in independent biological replicates, as indicated in the text and figures for each experiment. Data are shown as the mean ± SEM. Any difference was considered statistically significant when $p < 0.05$ using the ANOVA test. In experiments that included one variable, the one-way ANOVA test was used. In experiments with two independent variables, the two-way ANOVA test was employed. A posterior Tukey's multiple comparisons test was used for mean comparisons. Student t-test was used to compare unpaired data from two samples. Survival data were entered into Kaplan-Meier survival curves plots, and statistical analysis was performed using the Mantel log-rank test. The effect size is expressed in median survival (MS). Significance was determined if $p<0.05$. All analyses were conducted using GraphPad Prism (version 6.01) or SAS (version 9.4, SAS Institute, Cary, NC). Each statistical test used is indicated within the figure legends.

Fig. S1

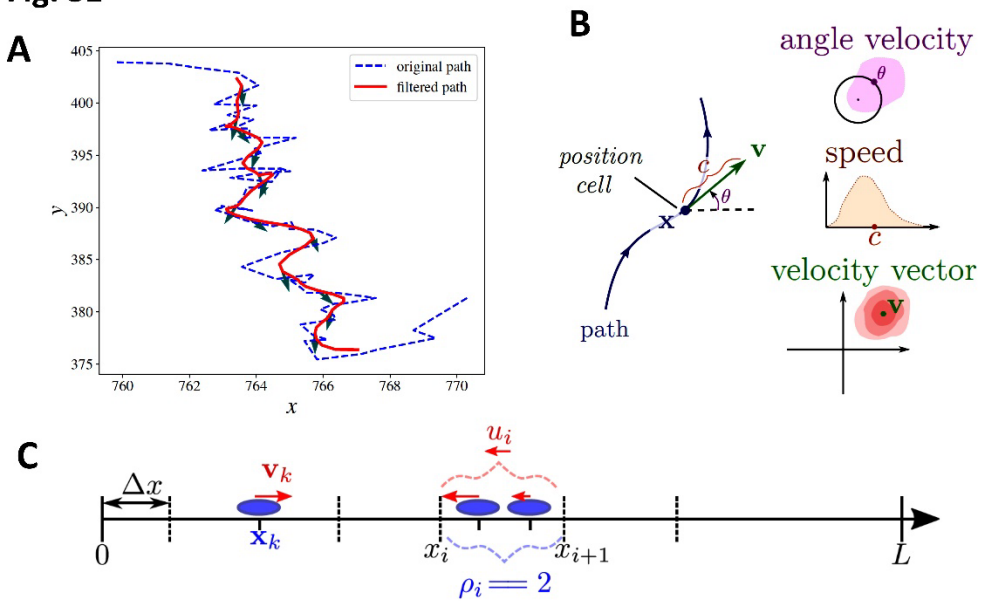
A
original path
filtered path
B
angle velocity
position cell
speed
velocity vector
path
C
D
Each cell k
All the pairs (p, q)
histogram
averaging
density
average velocity
radial distribution
correlation function
E
front
left
right
back
correlation velocity
F
Vectors showing four individual orientations
Vector towards 0°-90°
Vector towards 90°-180°
Vector towards 180°-270°
Vector towards 270°-0°
Stream
Swarm
Flock
G
Movie Core #1: Velocity Vector
Zone A
Swarm
Zone B
Stream
Zone C
Flock

Fig. S1: Statistical analysis performed to study glioma dynamics on the tumor core. A) Example of a path obtained with Image-J before (dashed line), and after filtering (red line). The smooth path allows to estimate the velocity of the cell at each time step. **B)** Three Individual statistics were explored to visualize the distribution of velocity: distribution of velocity angle (θ), speed distribution (c), and the velocity vector (v) distribution. **C)** Correlation functions: estimation of the density ρ_i and average velocity u_i. **D)** Radial distribution and correlation function are estimated in the same way as density and average velocity, but we use pairwise information (e.g. distances between neighbors). **E)** Heat map plot of the pair wise correlation of the velocity. We estimated the correlation depending on the position of a nearby x_j in different positions (front-back and left-right). Positive correlation (+) indicates cells are moving in same directions (red) and negative correlation (-) cells are moving in opposite directions (blue). **F)** Vectors indicating average direction per cell for each area (from Fig. 1E). **G)** Heat map plot of the distribution of the velocity vector in different zones of the tumor core. Velocity is shown in the x-axis and y-axis in μm/h.

Fig. S2

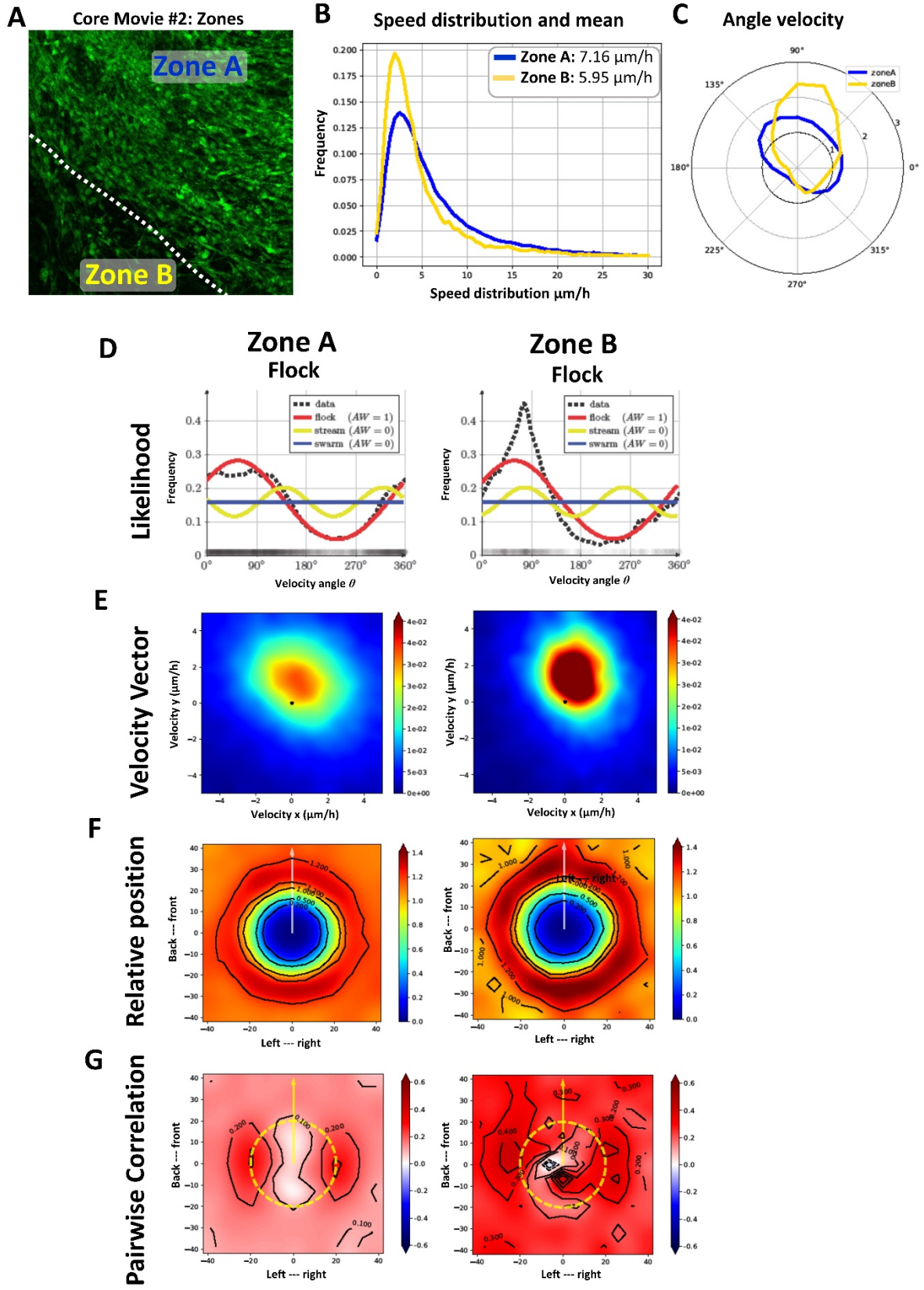
A
Core Movie #2: Zones
Zone A
Zone B
B
Speed distribution and mean
Zone A: 7.16 µm/h
Zone B: 5.95 µm/h
Frequency
Speed distribution µm/h
C
Angle velocity
zoneA
zoneB
D
Zone A
Flock
Zone B
Flock
Likelihood
data
flock (AW = 1)
stream (AW = 0)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S2: Collective dynamic patterns observed by confocal imaging within the glioma tumor core: analysis of core Movie #2. A) Statistical analysis of different regions of the movie, Zone A and Zone B, illustrated on a representative time lapse confocal still image. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic pattern formation. Zone A: flock, Zone B: flock. AW: Akaiki Weight. **E)** Heat map of the distribution of velocity vectors in each zone. **F)** Histogram plot showing relative position with nearby neighbors within each zone. X and Y axes are in μm. Scale bar shows frequency represented in colors. **G)** Histograms of pair-wise correlation with nearby neighbors for each zone. X and Y axes are in μm. Scale bar shows correlation represented as colors.

Fig. S3

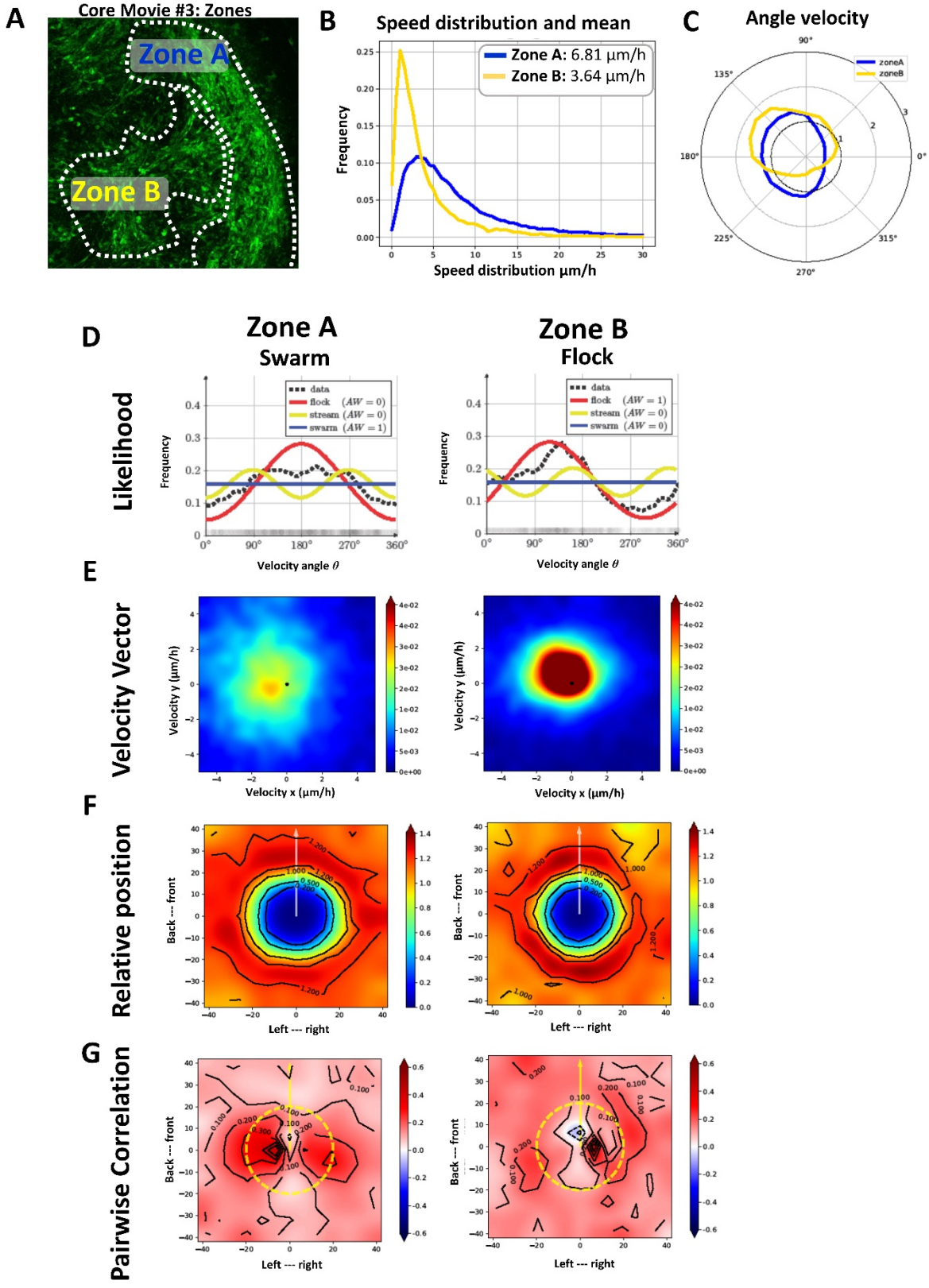
A
Core Movie #3: Zones
Zone A
Zone B
B
Speed distribution and mean
Zone A: 6.81 µm/h
Zone B: 3.64 µm/h
Frequency
Speed distribution µm/h
C
Angle velocity
zoneA
zoneB
D
Zone A
Swarm
Zone B
Flock
Likelihood
data
flock (AW = 0)
stream (AW = 0)
swarm (AW = 1)
flock (AW = 1)
stream (AW = 0)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S3: Collective dynamic patterns observed by confocal imaging within the glioma tumor core: analysis of core movie #3. A) Statistical analysis of different regions of the movie, Zone A and Zone B, illustrated on a representative time lapse confocal still image. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean of the speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic pattern formation. Zone A: flock, Zone B: flock. AW: 0 or AW:1. **E)** Heat map plot of the distribution of the velocity vector in each zone. **F)** Histogram plot showing interposition with nearby neighbor for each zone. X and Y axes are in μm. Scale bar shows frequency represented in colors. **G)** Histograms of pair-wise correlation with nearby neighbors for each zone. X and Y axes are in μm. Scale bar shows correlation represented in colors.

Fig. S4

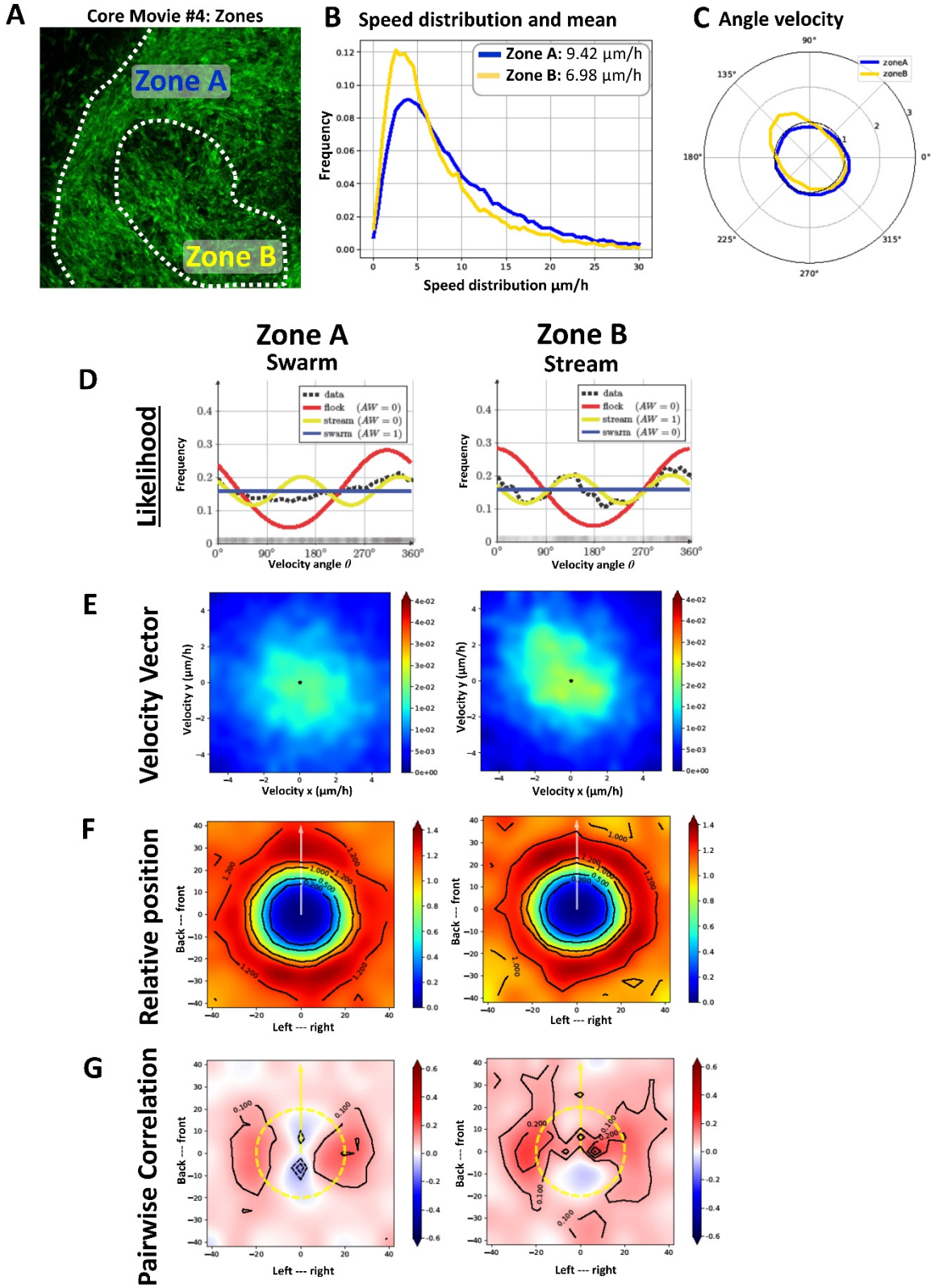
A
Core Movie #4: Zones
Zone A
Zone B
B Speed distribution and mean
Zone A: 9.42 µm/h
Zone B: 6.98 µm/h
Frequency
Speed distribution µm/h
C Angle velocity
zoneA
zoneB
Zone A
Swarm
Zone B
Stream
D
Likelihood
data
flock (AW = 0)
stream (AW = 0)
swarm (AW = 1)
stream (AW = 1)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S4: Collective dynamic patterns observed by confocal imaging within the glioma tumor core: analysis of core movie #4. A) Statistical analysis of different regions of the movie, Zone A and Zone B, illustrated on a representative time lapse confocal still image. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean of the speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic pattern formation. Zone A: swarm, Zone B: Stream. AW: 0 or AW:1. **E)** Heat map plot of the distribution of the velocity vector in each zone. **F)** Histogram plot showing interposition with nearby neighbor for each zone. X and Y axes are in μm. Scale bar shows frequency represented in colors. **G)** Histograms of pair-wise correlation with nearby neighbors for each zone. X and Y axes are in μm. Scale bar shows correlation represented in colors.

Fig. S5

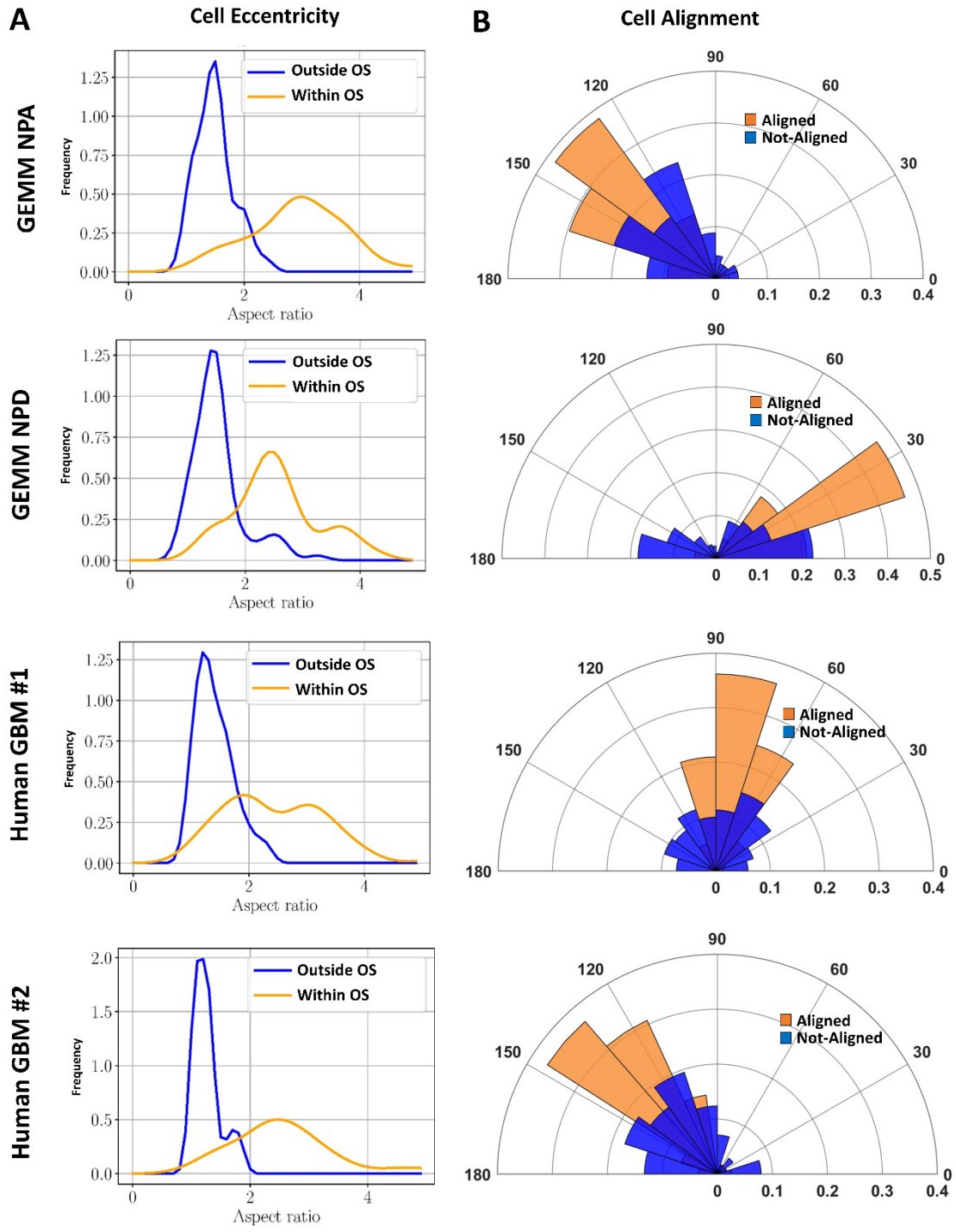

Fig. S5: Oncostream characterization in mouse and human glioma tissue. A-B) Cellular eccentricity and alignment analysis on mouse genetically engineered glioma models (NPA and NPD) and human glioma. Histograms of cellular aspect ratio **(A)** and cell alignment **(B)** shows that cells within areas of oncostreams (Within-OS) are elongated and aligned, whereas outside of oncostreams (Outside-OS) they tend to be rounded and are not-aligned.

Fig. S6

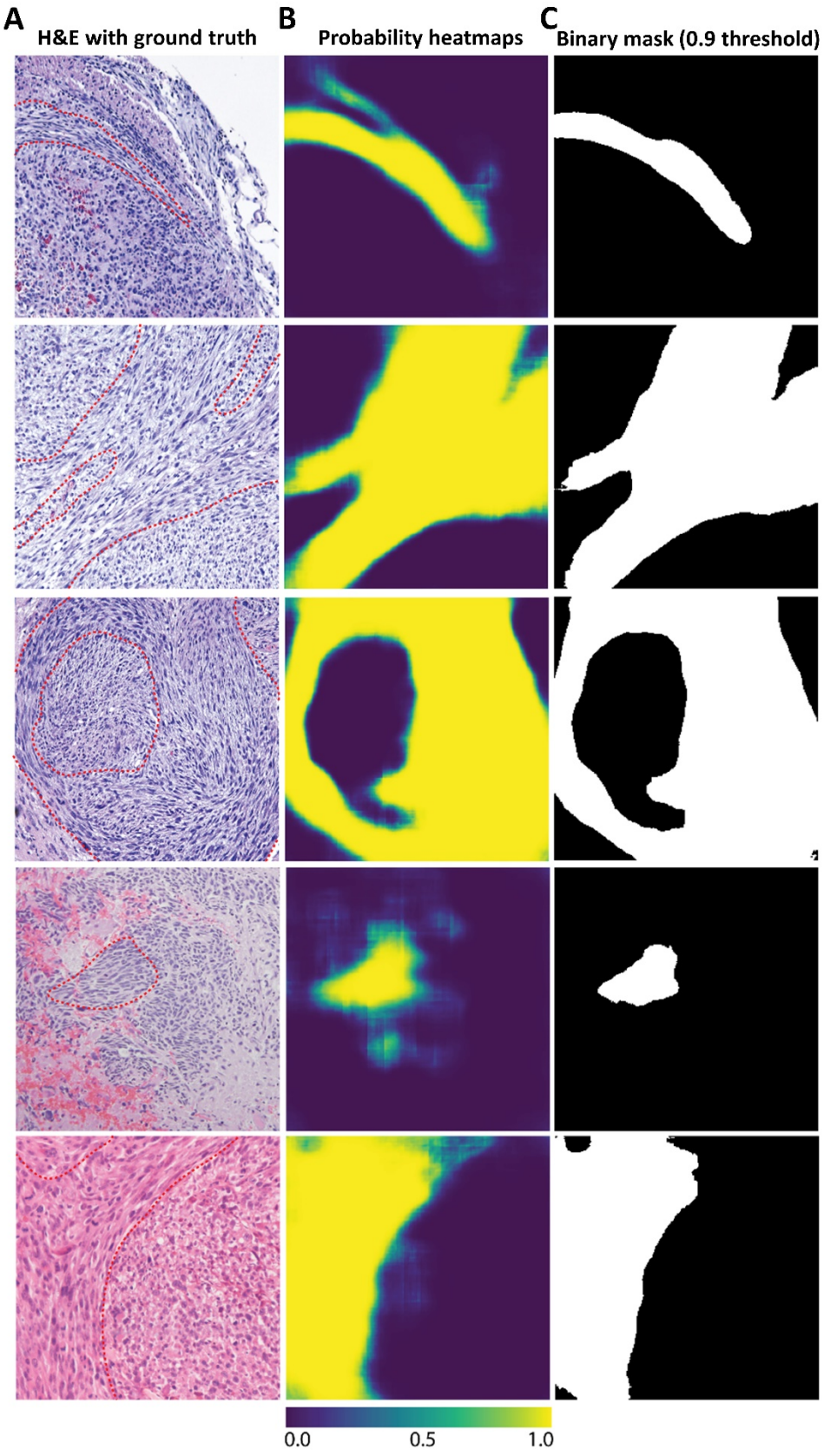
A
H&E with ground truth
B
Probability heatmaps
C
Binary mask (0.9 threshold)
0.0
0.5
1.0

Fig. S6: A deep learning model identifies oncostreams in mouse glioma H&E histological images. We implemented a U-Net architecture to provide semantic segmentation of oncostream fascicles in glioma histological sections. **A)** Mouse H&E images are shown with oncostreams manually segmented (stippled red lines). **B)** Model output is a semantic segmentation probability heatmap with each pixel being assigned a probability of being within an oncostream (foreground, yellow), or not (background, deep purple). Probability heatmaps for each corresponding H&E images are shown. **C)** Probability heatmaps can be converted to binary masks (foreground versus background) using probability thresholding. Oncostream binary masks with probability threshold of >0.9 are shown.

Fig. S7

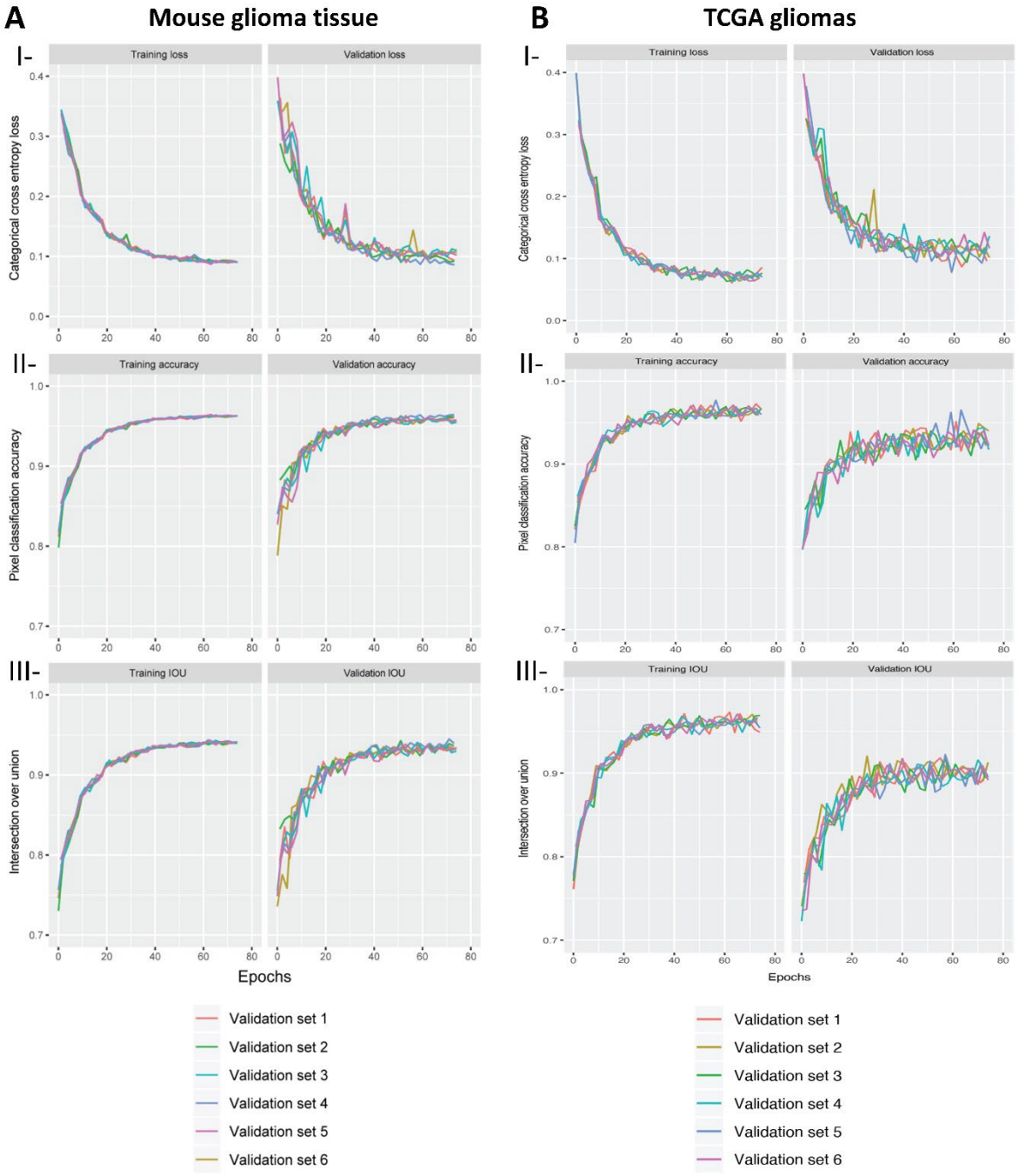

Fig. S7. Training and validation curves. A-B) Training and validation curves for a fully convolutional neural network trained using the mouse dataset (**A**) and mouse and TCGA dataset (**B**) is shown for the **(I)** cross entropy loss, **(II)** pixel classification accuracy, and **(III)** intersection over union (IOU) metric. We performed 6 iterations of random training-validation dataset splits (80%/20%), and holding out only the TCGA data for validation. Training and validation metrics stabilized after 75 epochs.

Fig. S8

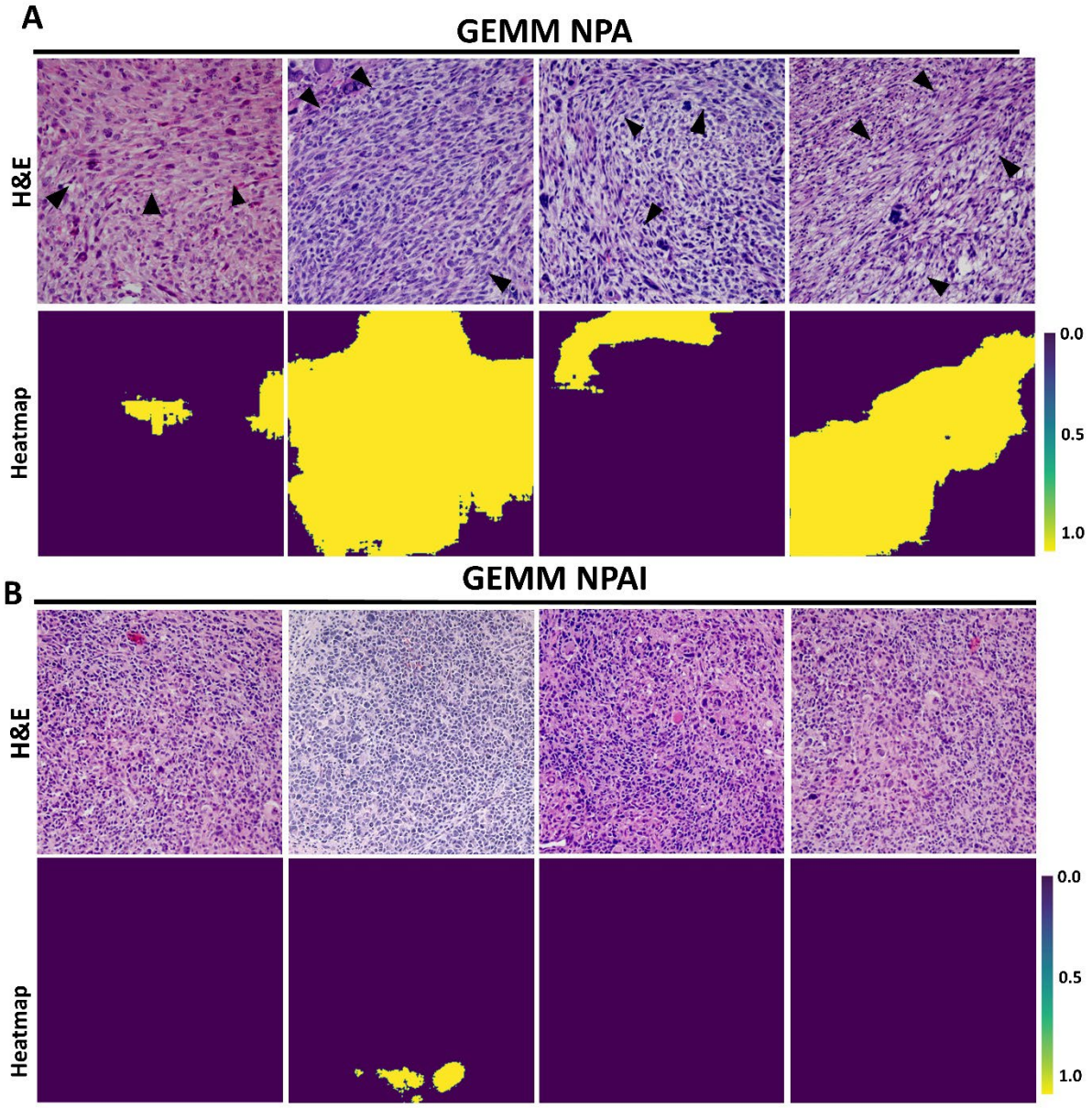

Fig. S8: Semantic segmentation of oncostreams areas on IDH-WT (NPA) and IDH-Mutant (NPAI) mouse genetically engineered glioma (GEMM). A-B) Representative images of oncostreams manually segmented on H&E stained sections of NPA (IDH-WT) gliomas (**A**) and NPAI (IDH-Mut) gliomas (**B**); (oncostreams are indicated with black arrowheads). 'Heatmap' row illustrates the semantic segmentation probability heatmaps for each corresponding H&E image.

Fig. S9

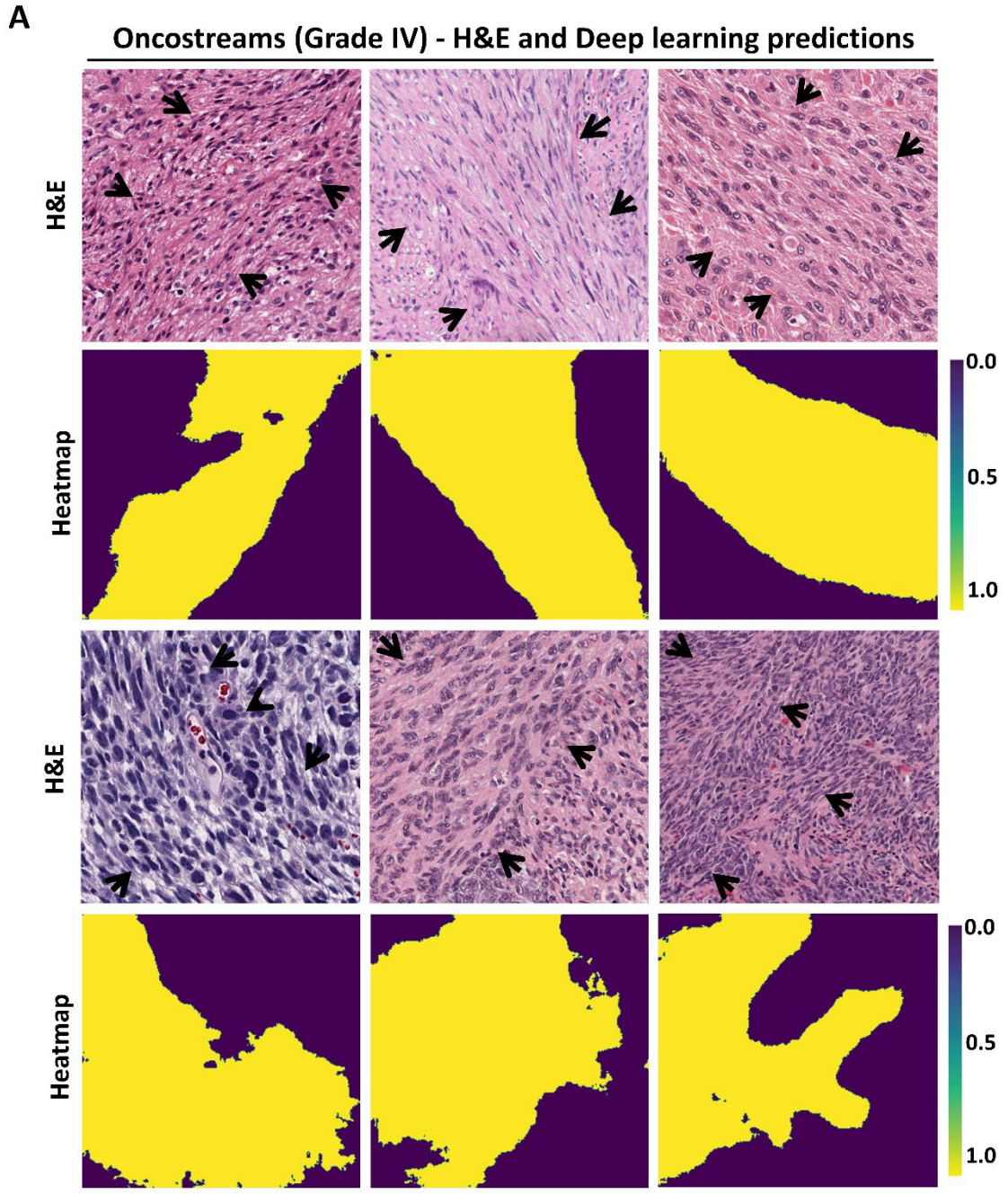

Fig. S9: Oncostreams are identified by deep learning in human high grade gliomas (Glioblastoma - Grade IV). A) Representative H&E images of TCGA-glioblastoma multiforme (WHO Grade IV) diagnostic slides show the presence of manually segmented oncostreams (indicated by black arrows). Second and fourth rows show semantic segmentation probability heatmaps for each corresponding H&E image.

Fig. S10

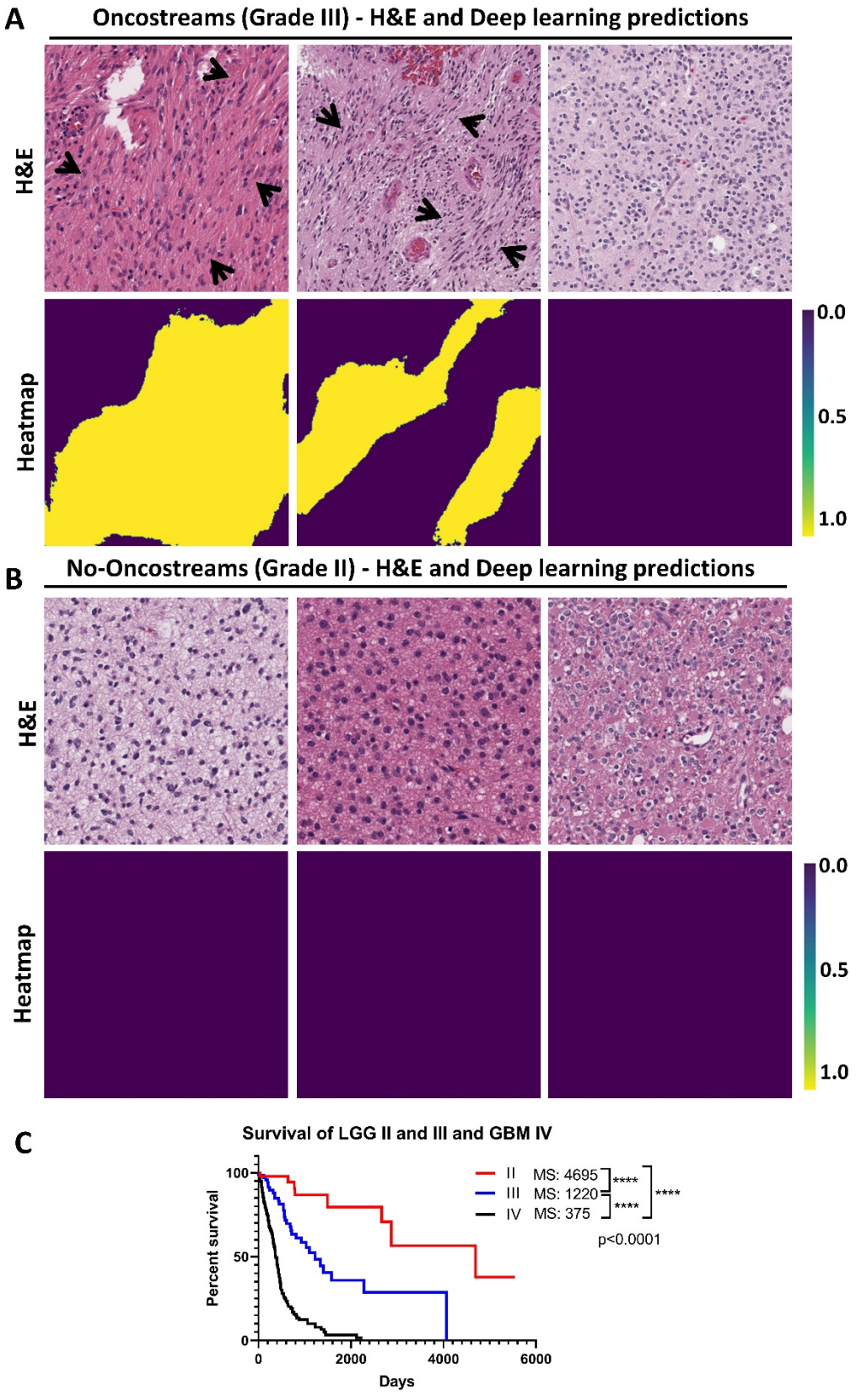
A
Oncostreams (Grade III) - H&E and Deep learning predictions
H&E
Heatmap
0.0
0.5
1.0
B
No-Oncostreams (Grade II) - H&E and Deep learning predictions
H&E
Heatmap
0.0
0.5
1.0
C
Survival of LGG II and III and GBM IV
Percent survival
100
50
0
0
2000
4000
6000
Days
II MS: 4695
III MS: 1220
IV MS: 375
****
****
****
p<0.0001

Fig. S10: Oncostreams are observed in a low percentage of human gliomas WHO grade III but not in WHO grade II. A-B) Representative H&E images of TCGA gliomas WHO Grade III **(A)** and Grade II **(B)** samples show the presence of manually segmented oncostreams (indicated by black arrows). Semantic segmentation probability heatmaps for each corresponding H&E image are shown in the row labeled 'heatmap'. **C)** Kaplan-Meier survival curve by histology grade determined from the 100 TCGA-GBM and 120 TCGA-LGG analyzed cases.

Fig. S11

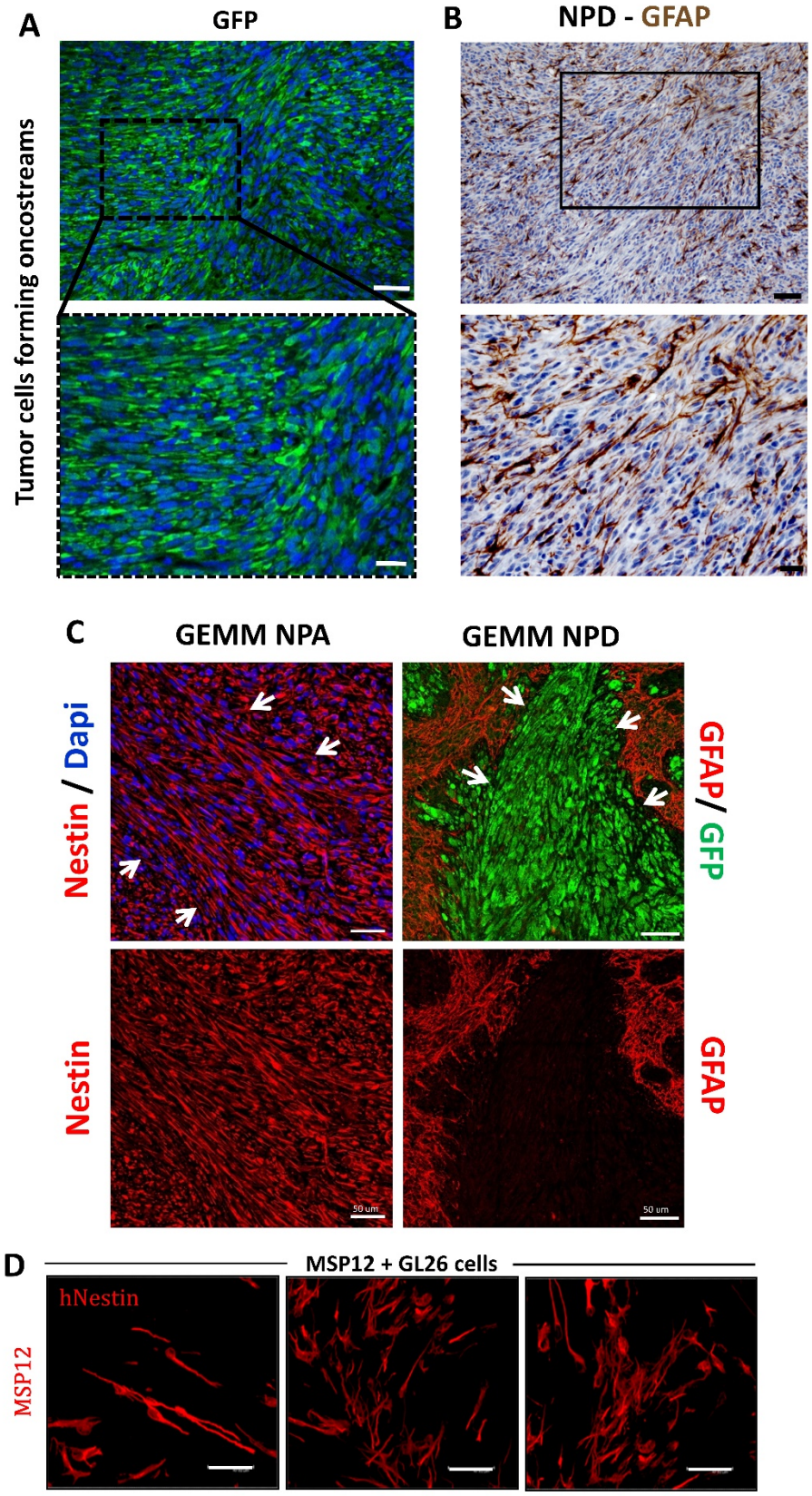
A
GFP
Tumor cells forming oncostreams
B
NPD - GFAP
C
GEMM NPA
GEMM NPD
Nestin / Dapi
GFAP/ GFP
Nestin
GFAP
50 um
50 um
D
MSP12 + GL26 cells
MSP12
hNestin

Fig. S11: Oncostreams are formed by aligned tumor cells and includes cells from the tumor microenvironment. Oncostreams' cellular heterogeneity was analyzed by immunofluorescence staining. **A)** GFP expression in NPA tumors show oncostreams shaped by tumor cells (green = tumor cells; and blue = DAPI stained nuclei). Scale bar: 100 μm (top) and 50 μm (bottom). **B)** Immunohistochemistry analysis on GEMM of glioma (NPD) shows alignment of GFAP+ cells within oncostreams. Scale bars: 50 μm (top) and 20 μm (bottom). **C)** Representative confocal images of GEMM **NPA** glioma illustrates the presence of nestin (red) within oncostreams. Nuclei were stained with DAPI (left). Confocal images of GEMM **NPD** glioma show the expression of tumor (GFP+, green) within streams, and GFAP+ cells (red) surrounding the stream (right panel). Scale bar: 50 μm. **D**) Immuno-fluorescence images of human-nestin (red) highlighting MSP12 cells. MSP-12 cells were implanted with GL26-citrine cells. MSP12 cells adopt a bipolar structure only when aligned to GL26-citrine cells. Scale bar: 47.62 μm.

Fig. S12

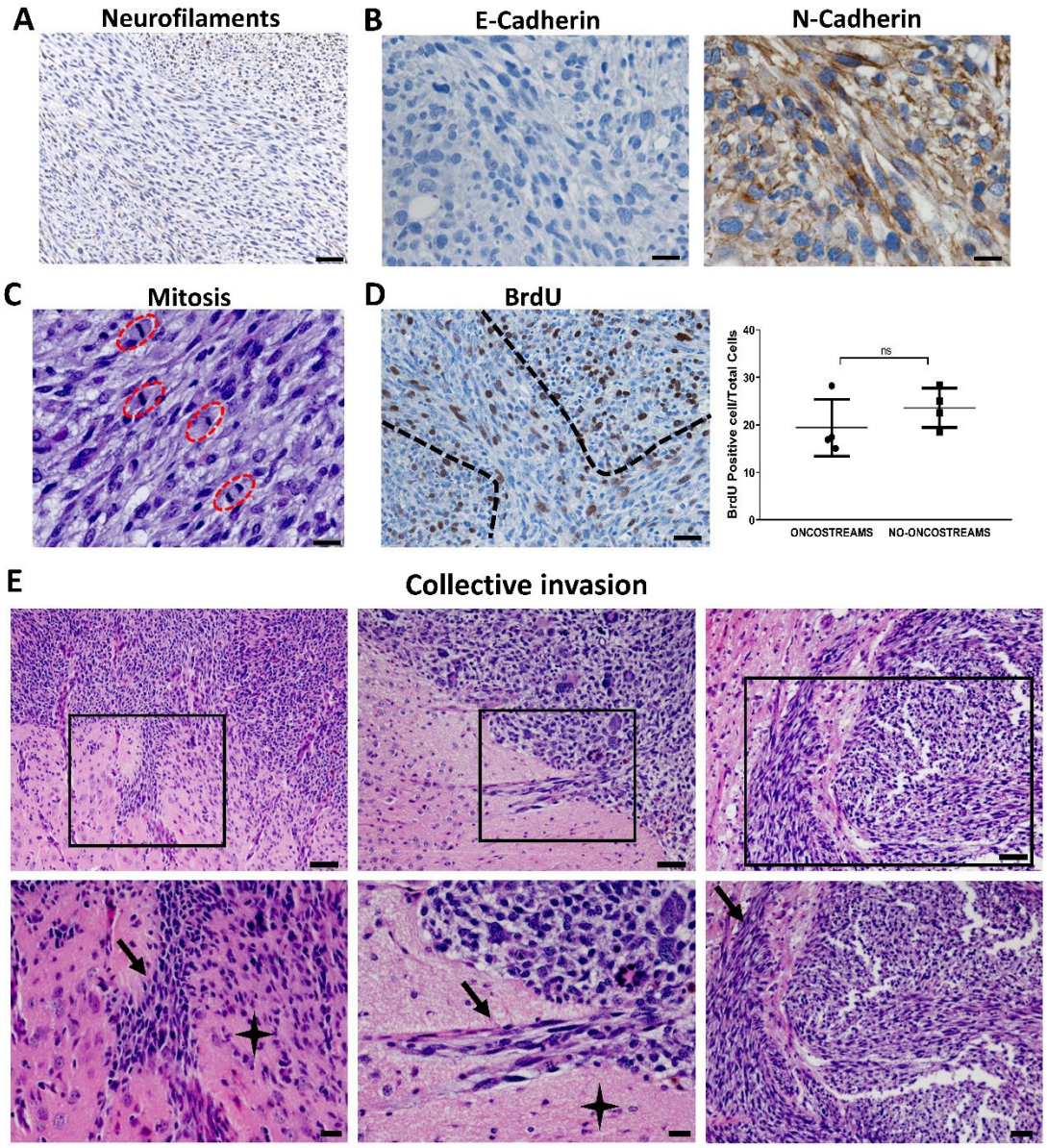

Fig. S12: Oncostreams cellular functions characterization. A) Immunohistochemistry staining for neurofilaments in NPA tumors was negative. Scale bars: 50 µm. **B)** Immunohistochemistry analysis of E-cadherin and N-cadherin expression within NPA gliomas. Gliomas are negative for E-cadherin and positive for N-cadherin both within and outside oncostream structures. **C)** Analysis of glioma cells' mitosis within oncostreams in H&E stained sections. Mitosis are orientated in the same orientation of oncostreams. **D)** Proliferation analysis comparing oncostream (dotted lines) with no-oncostream areas. Positive BrDu cells were counted by Image-J software. Scale bars: 50 µm. BrDu positive cells per total cells in the visual field were counted; n=4. Ten fields of each section were selected at random. Error bars represent ± SEM; paired t-test. **E)** Representative images of oncostream invasion (black boxes) from H&E stained sections of genetic engineered NPA gliomas. Arrows indicate oncostream collective invasion. Stars indicate areas of single cell invasion. Scale bars: 50 µm (top) and 20 µm (bottom).

Fig. S13

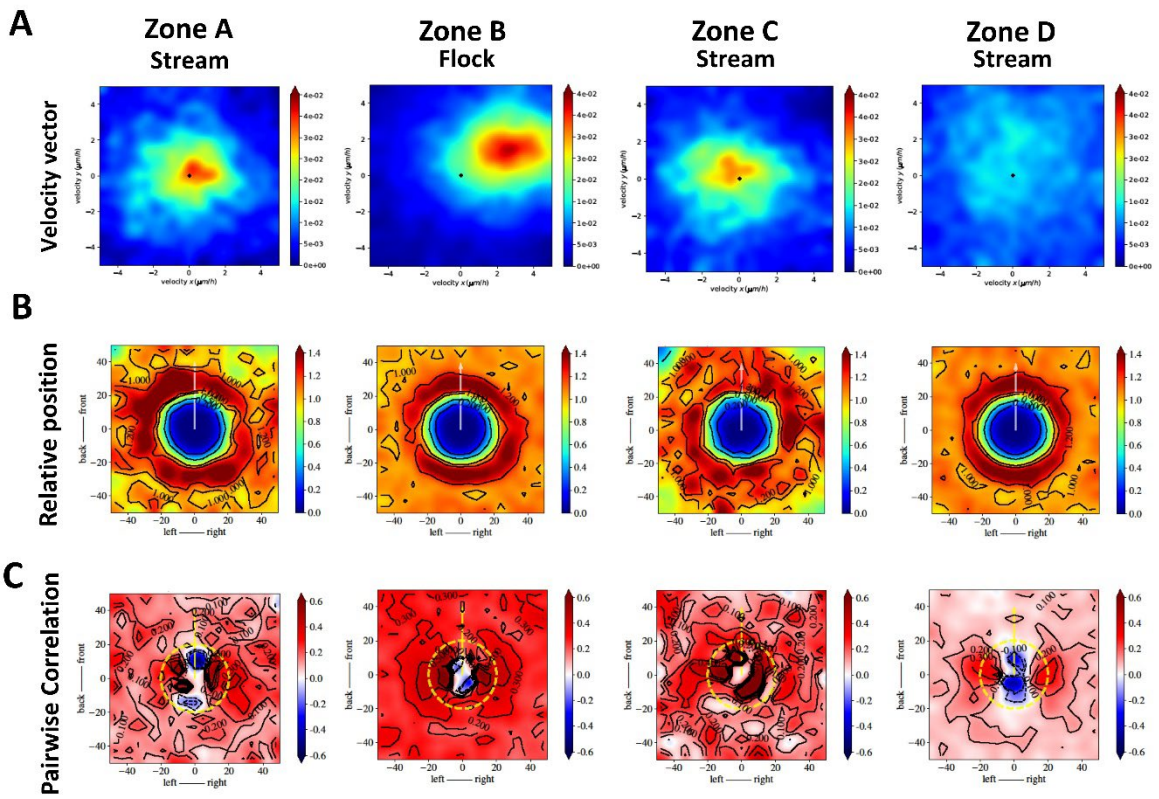

Fig. S13: Invasion analysis using confocal time-lapse imaging on the tumor border movie #1. **A)** Heatmap plot of the distribution of velocity vectors in zones A, B, C and D of border movie #1 (shown in **Fig. 6)**. **B)** Histogram of relative cell positions in relationship with nearby neighbors. For each cell, x_i, we estimate the probability to find another cell xj nearby. Probability scale bars are represented by colors. Axes shows cells position left-right and front-back in µm. **C)** Pairwise correlation of motion with nearby neighbors. Scale bars of probability are represented by colors. Axes indicate cell position and distance in µm from neighboring cells. Dotted yellow line indicates the location of x_i.

Fig. S14

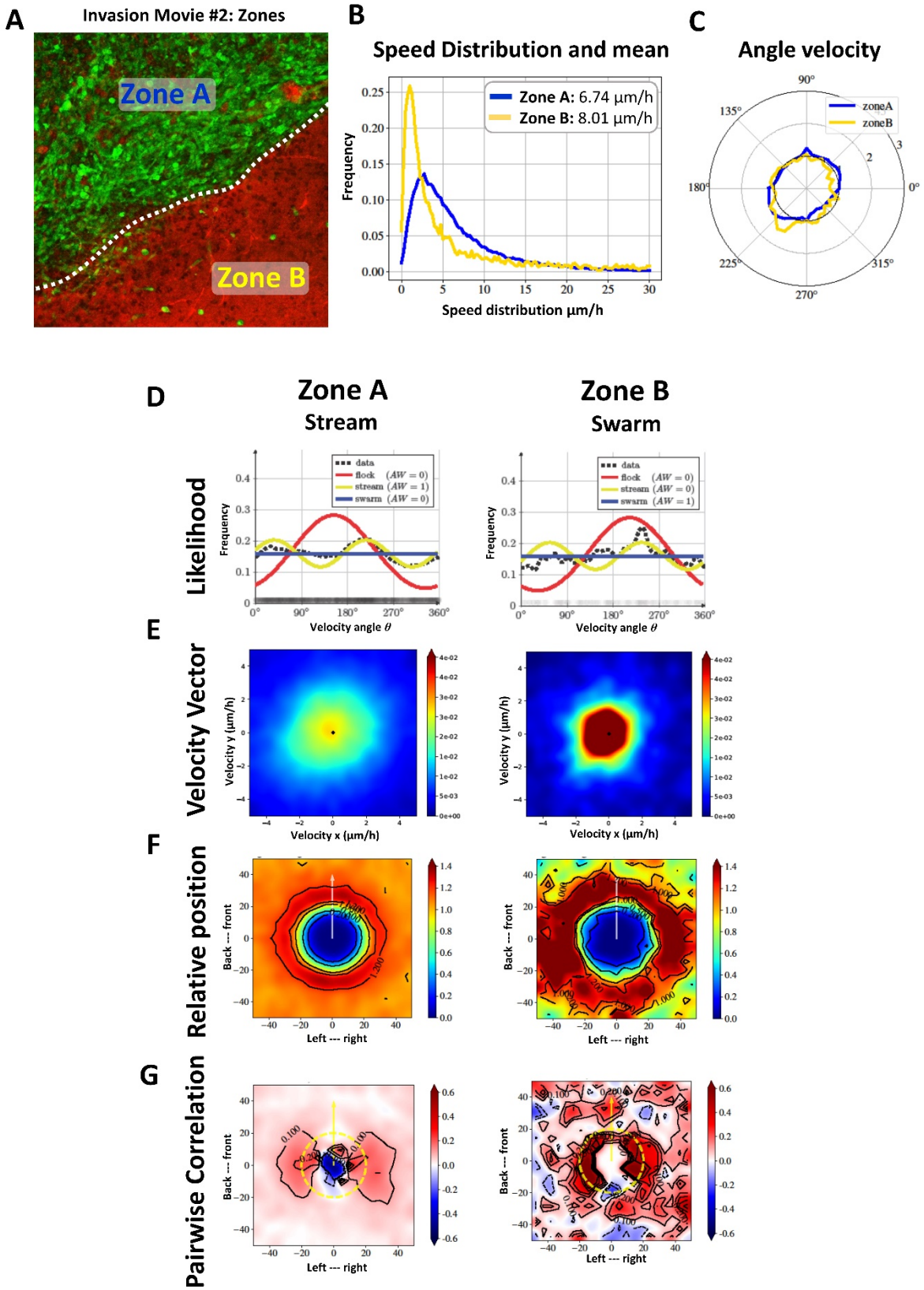
A
Invasion Movie #2: Zones
Zone A
Zone B
B
Speed Distribution and mean
Zone A: 6.74 µm/h
Zone B: 8.01 µm/h
Frequency
Speed distribution µm/h
C
Angle velocity
zoneA
zoneB
D
Zone A
Stream
Zone B
Swarm
Likelihood
data
flock (AW = 0)
stream (AW = 1)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S14: Invasion analysis using confocal time-lapse imaging on the tumor border movie #2). **A)** Labeling of different regions of the tumor border for movie #2. Representative time lapse confocal images subdivided into dynamic Zones A and B. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic motion patterns. Zone A: stream, Zone B: swarm. AW: 0 or AW:1. **E)** Heat map plot of the distribution of the velocity vectors in each zone. **F)** Histogram plot showing the relative position with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate probability values in color. **G)** Histograms of pairwise correlations with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate correlation values in colors.

Fig. S15

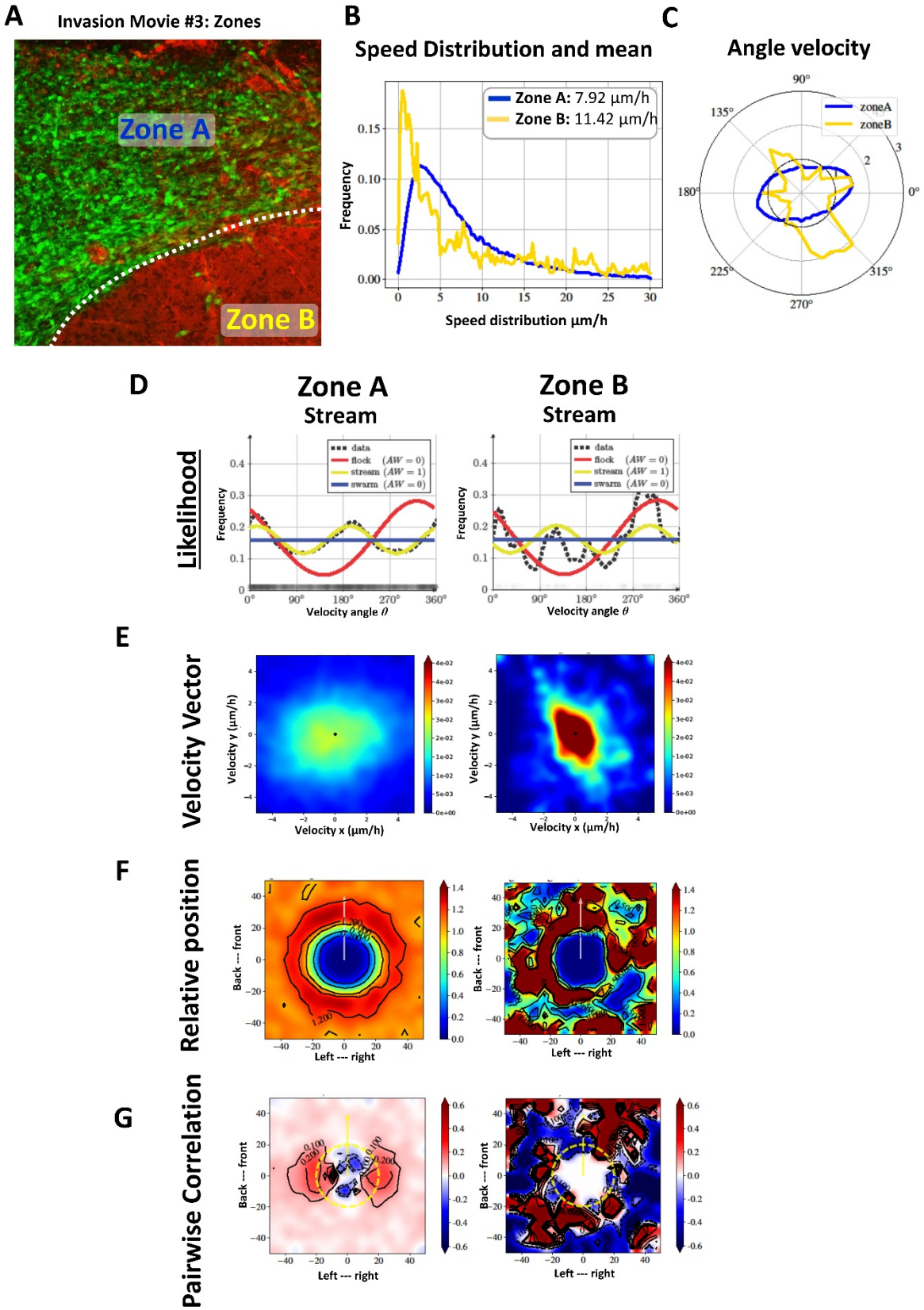
A
Invasion Movie #3: Zones
Zone A
Zone B
B
Speed Distribution and mean
Zone A: 7.92 µm/h
Zone B: 11.42 µm/h
Frequency
Speed distribution µm/h
C
Angle velocity
zoneA
zoneB
D
Zone A
Stream
Zone B
Stream
Likelihood
data
flock (AW = 0)
stream (AW = 1)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S15: Invasion analysis using confocal time-lapse imaging of the tumor border movie #3. **A)** Labeling of different regions of the tumor border for movie #2. Representative time lapse confocal images subdivided into dynamic Zones A and B. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic motion patterns. Zone A: stream, Zone B: stream. AW: 0 or AW:1. **E)** Heat map plot of the distribution of the velocity vectors in each zone. **F)** Histogram plot showing the relative position with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate probability values in color. **G)** Histograms of pairwise correlations with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate correlation values in colors.

Fig. S16

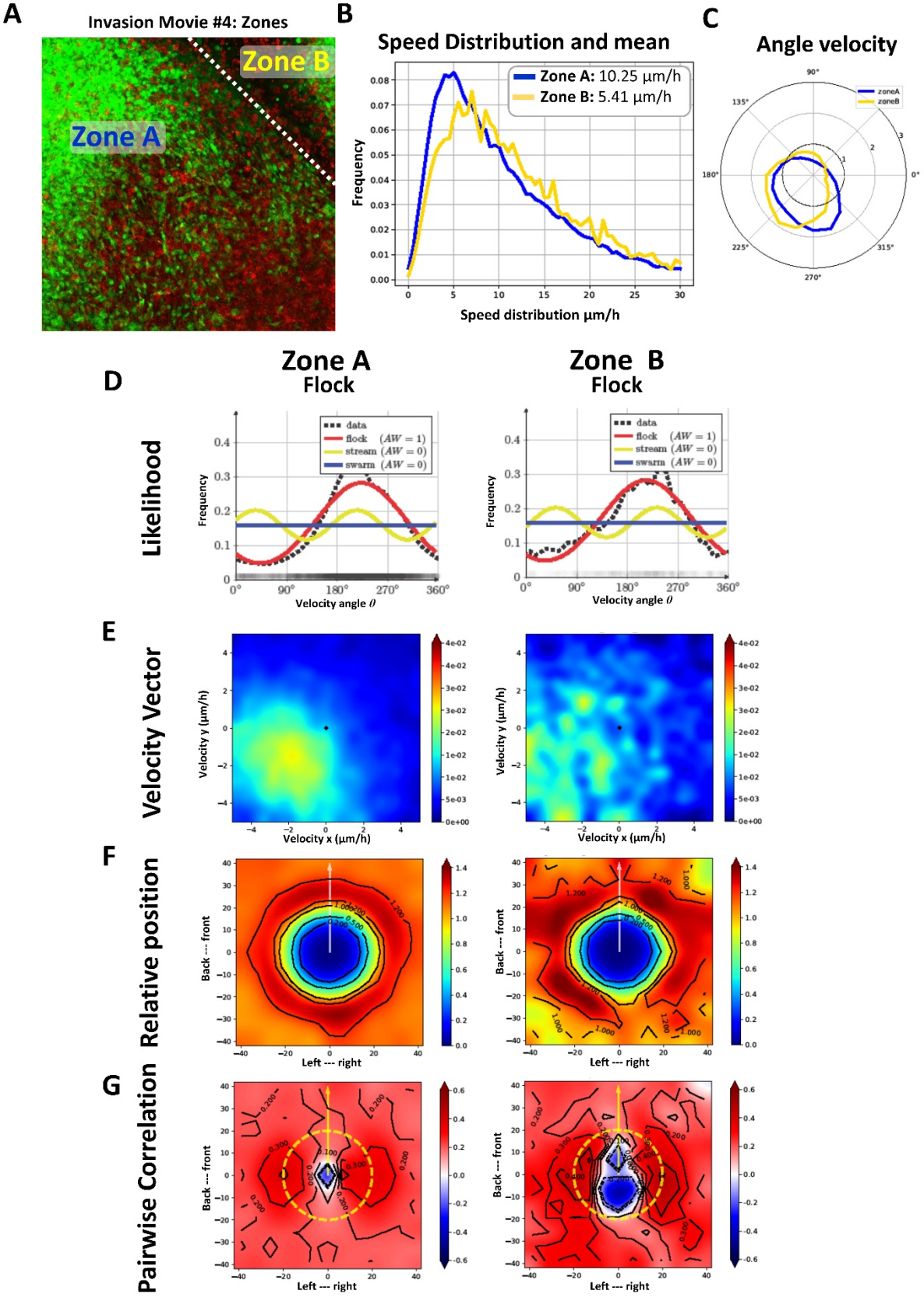
A
Invasion Movie #4: Zones
Zone B
Zone A
B
Speed Distribution and mean
Zone A: 10.25 µm/h
Zone B: 5.41 µm/h
Frequency
Speed distribution µm/h
C
Angle velocity
zoneA
zoneB
D
Zone A
Flock
Zone B
Flock
Likelihood
data
flock (AW = 1)
stream (AW = 0)
swarm (AW = 0)
Frequency
Velocity angle θ
E
Velocity Vector
Velocity y (µm/h)
Velocity x (µm/h)
F
Relative position
Back --- front
Left --- right
G
Pairwise Correlation
Back --- front
Left --- right

Fig. S16: Invasion analysis using confocal time-lapse imaging on the tumor border movie #4. **A)** Labeling of different regions of the tumor border for movie #2. Representative time lapse confocal images subdivided into dynamic Zones A and B. **B)** Speed distribution ($\mu m/h$) in Zone A (blue) and B (yellow). Inner panel shows mean speed for each zone. **C)** Angle Velocity distribution analysis (θ) performed by zones. **D)** Likelihood analysis histograms to classify dynamic motion patterns. Zone A: flock, Zone B: flock. AW: 0 or AW:1. **E)** Heat map plot of the distribution of the velocity vectors in each zone. **F)** Histogram plot showing the relative position with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate probability values in color. **G)** Histograms of pairwise correlations with nearby neighbors for each zone. *x* and *y* axes are in μm. Bars to the right of each figure indicate correlation values in colors.

Fig. S17

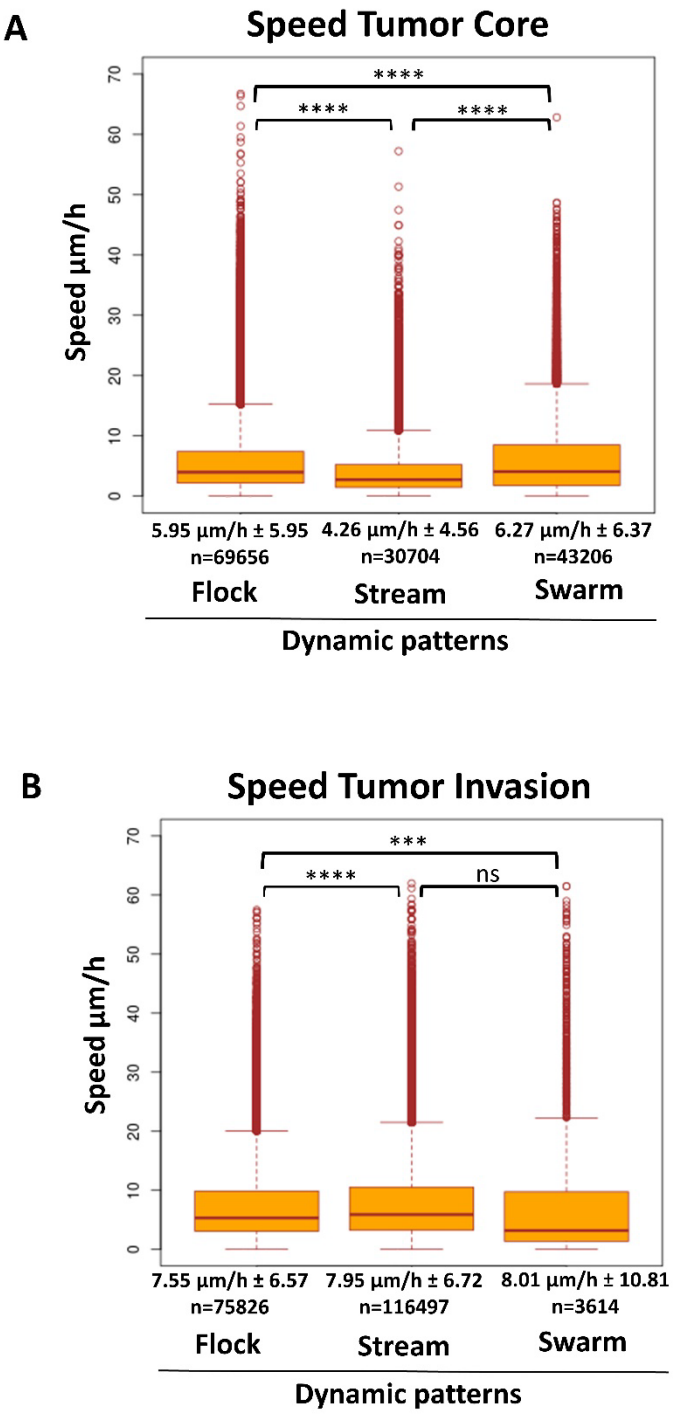

Fig. S17: Speed analysis of cells in either flocks, streams, or swarms with the tumor core or the tumor border. **A)** Box plots of cell speed (µm/h) within flocks, streams, or swarms in the tumor core. Stream n=30704, Flock **n= 69656**, Swarm n=43206. Mean ±SEM are shown; One-way ANOVA, ***p<0.001, ***p<0.0001. **B)** Box plot of speed of cells (µm/h) within flocks, streams, or swarms within the tumor border. Stream n=116497, Flock **n=75826**, Swarm n=3614. Mean ±SEM are shown; One-way ANOVA, ***p<0.001, ***p<0.0001.

Table S1. Analysis of oncostreams on TCGA glioma diagnostic slides from the Genomic Data Commons Portal

TCGA glioma analysis				
Grade	Recurrence	Total Tumors	OS Positive	OS Negative
GBM-Grade IV	Primary	100	**47**	53
LGG-Grade III	Primary	70	**6**	64
LGG-Grade III	Primary	50	**0**	50

Table S3. Summary results of oncostreams detection using manually histopathological analysis and deep learning segmentation

Accuracy of methodologies for Oncostreams analysis					
Tumor	Method	Oncostreams		Concordance	
		Positive	Negative	Images #	Percentage (%)
GBM - Grade IV	Manually	78	31	92/109	84.4
	Deep Learning	78	31		
LGG - Grade III	Manually	29	97	113/126	89.7
	Deep Learning	24	102		
LGG - Grade II	Manually	0	61	60/61	98.4
	Deep Learning	1	60		

Table S4 –List of antibodies used for immuno-histochemistry analysis

Antibody name	Company	Catalog #	Host	Dilution
Anti-Nuclei, Clone 235-1 (HuNu)	Millipore Sigma	MAB1281	Mouse	1:100
Anti-Green Fluorescent Protein (GFP)	Rockland	600-101-215	Goat	1:1000
Anti-Alpha Smooth Muscle Actin (α-SMA)	Abcam	ab5694	Rabbit	1:500
Anti-Glial Fibrillary Acidic Protein (GFAP)	Millipore Sigma	AB5804	Rabbit	1:1000
Anti-Neurofilament-L (C28e10)	Cell Signaling	2837	Rabbit	1:100
Anti-E-Cadherin (24E10)	Cell Signaling	3195	Rabbit	1:400
Anti-N-Cadherin	Abcam	AB18203	Rabbit	1:1000
Anti-Nestin	Novus	NB100-1604	Chicken	1:800
Anti Sox2	Invitrogen	MA1-014	Mouse	1:200
Anti-Iba1 [EPR16588]	Abcam	ab178846	Rabbit	1:500
Anti-Brdu (Bu20a)	Cell Signaling	5292	Mouse	1:200

Movies S1 to S4. Glioma dynamics at the tumor core.
Time lapse confocal imaging of organotypic brain slice cultures of NPA glioma cores. Movement of tumor cells GFP positive (green) were analyzed within the tumor core. Imaging was obtained every 10 minutes for the duration of 293 cycles.

Movies S5 to S8. Glioma dynamics at the tumor border.
Glioma dynamic at the tumor border were analyzed using organotypic slice cultures glioma model by intracranial implantation of GFP positive NPA cells into the striatum of B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-**tdTomato**-EGFP)Luo/J- transgenic mice. Normal brain parenchyma is visualized in red. Imaging was obtained every 10 minutes for the duration of 186 cycles.