

State of Cell Unjamming Correlates with Distant Metastasis in Cancer Patients

Pablo Gottheil¹, Jürgen Lippoldt¹, Steffen Grosser¹, Frédéric Renner¹, Mohamad Saibah¹, Dimitrij Tschodu¹,
 Anne-Kathrin Poßögel², Anne-Sophie Wegscheider², Bernhard Ulm³, Kay Friedrichs⁴, Christoph Lindner⁵,
 Christoph Engel⁶, Markus Löffler⁶, Benjamin Wolf⁷, Michael Höckel⁷, Bahriye Aktas⁷,
 Hans Kubitschke¹, Axel Niendorf² and Josef A. Käse^{1,*}

¹Peter Debye Institute for Soft Matter Physics, Leipzig University, Linnéstraße 5, 04103, Leipzig, Germany

²MVZ Prof. Dr. med. A. Niendorf Pathologie Hamburg-West GmbH Institute for Histology,
 Cytology and Molecular Diagnostics, Lornsenstraße 4, 22767, Hamburg, Germany

³Independent Statistical Consulting Bernhard Ulm, Kochelseestraße 11, 81371, Bavaria, Germany

⁴Mamma Center Hamburg, Breast Clinic at Jerusalem Hospital,
 Moorkamp 2–6, 20357, Hamburg, Germany

⁵Gynecological Clinic, Agapleosion Diakonieklinikum, Hohe Weide 17, 20259, Hamburg, Germany

⁶Institute for Medical Informatics, Statistics and Epidemiology, University Leipzig,
 Härtelstraße 16-18, 04107 Leipzig, Leipzig, Germany

⁷Department of Gynaecology, University Hospital Leipzig, Liebigstraße 20a, 04103, Leipzig, Germany

 (Received 17 August 2022; revised 14 April 2023; accepted 17 May 2023; published 10 July 2023)

Pathological morphological changes in tumor tissue enable collective cancer cell unjamming, a cellular motility transition. However, fundamental questions remain: Is unjamming essential for tumor progression? Which different unjamming states can be found in patients? Here, vital cell tracking in patient-derived solid tumor explants ($N = 16$) reveals that states of cell unjamming can be recognized by elongated cell and nucleus shape (CeNuS) and low nucleus number density. These static variables serve as a morphodynamic link to map the broad range of morphologies and associated motility states found in histological slides of 1380 breast cancer patients to generate a comprehensive state diagram of cancer cell unjamming. An increase in predicted cell motility in primary tumors through unjamming significantly correlates with distant metastases that may even occur a decade later. Patient risk groups are quantified via a decision boundary in the state space found by machine learning. The resulting clinical prognostic potential is evaluated using a range of quantifiers, including Harrel's concordance index. Using multivariable Cox models, we find that cell unjamming as a prognostic parameter adds a 26% information gain in the concordance index when combined with the established prognostic criteria (tumor diameter, tumor grade, lymph node status) used in the Nottingham index. Unjamming complements the information on affected lymph nodes in patients regarding metastatic risk. The derived state diagram of cancer cell unjamming reconciles conflicting observations regarding shape- or density-induced unjamming and stresses the nuclei's mechanical importance, which is not considered in current theories of cell unjamming. We conclude that cancer cell unjamming is part of the metastatic cascade; thus, an emergent physical phenomenon contributes to tumor progression.

DOI: [10.1103/PhysRevX.13.031003](https://doi.org/10.1103/PhysRevX.13.031003)

Subject Areas: Biological Physics,
 Interdisciplinary Physics,
 Medical Physics

I. INTRODUCTION

Epidemiologically, an estimated 2.1 million women were diagnosed with breast cancer worldwide in 2018. Unfortunately, in the same year, over 600 000 women with

breast cancer died [1], primarily because of the systemic, invasive nature of the disease [2].

Currently, the invasion of nearby lymph nodes (LN) is the key indicator of cancer cell motility in diagnostics, and it constitutes the most important marker for the prognosis of distant metastases [3]. Significant ambiguity remains since around 30% of women without lymph node invasion develop distant metastases. In contrast, around another 30% of women with lymph node invasion remain distant-metastasis-free 10 years after surgery [4]. Current histopathological cell motility markers rely on cancer cells leaving the tumor and invading the lymph or vascular

*jkaes@physik.uni-leipzig.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

systems [5,6]. There are no immediate, established molecular markers for cancer cell motility. Furthermore, no marker—neither molecular nor histopathological—reliably accounts for the onset of cell motility within the cancer cell clusters of a tumor [7], which is an early and essential step in the metastatic cascade [8].

We address this problem by introducing a simple biomechanical cell motility marker based on soft matter physics that describes the squeezed state of a migrating cell in dense tissue and the nucleus number density as discriminators that detect potentially motile cells in static images of primary tumors, thereby acting as a morphodynamic link. We show that these variables describe cell unjamming, which occurs in carcinomas [9], and significantly correlate with distant metastatic risk, thereby suggesting that unjamming in the primary tumor increases cancer aggressiveness by promoting distant metastasis. Furthermore, our patient data-derived unjamming state diagram identifies cell and nucleus shapes in conjunction with nucleus number densities (ρ) as drivers of unjamming. Since jammed cells form a metastable state, we use the term state diagram instead of phase diagram. Cancer cell unjamming as a collective motility transition may be an important early event in the metastatic cascade, different from the epithelial-mesenchymal transition (EMT) [10]. Thus, it may complement the current diagnosis as a prognostic marker for distant metastasis.

The notion of “unjamming” in the context of cellular systems describes a transition from collective arrest of cell motion to mobile tissue behavior. The term emphasizes arrest due to constraints of degrees of freedom, e.g., from neighboring cells. The theoretical framework of glass transitions is also suited to describe collective cell arrest. These transitions are predominantly associated with a diverging viscosity controlled by a critical temperature (for cells, a critical activity). Both concepts are closely related, but here we stick with the notion of “jamming” since, for our interdisciplinary research, this term eases the understanding and intuition, e.g., for physicians or biologists.

In standard histological slides of breast carcinomas used in clinical routines, we typically find regions of dense cancer cell clusters embedded into the extracellular matrix. In our histological data, 90% of cancer cells are in clusters bigger than $600 \mu\text{m}^2$ (compare Fig. 5). These clusters are the focus of this study. By optical microscopy, no space is visible between the cells in the primary breast cancer cell clusters depicted in Fig. 1. With the given optical resolution of roughly $0.25 \mu\text{m}$, Grosser *et al.* [9] estimated the lower bound for the cells’ volume fractions to be around 95%, meaning that the cells are close to volume fraction of 1. This estimate is true for both distinct regions in the clusters with elongated [Fig. 1(b)] and roundish [Fig. 1(a)] cancer cells and nuclei. For these dense cellular islands

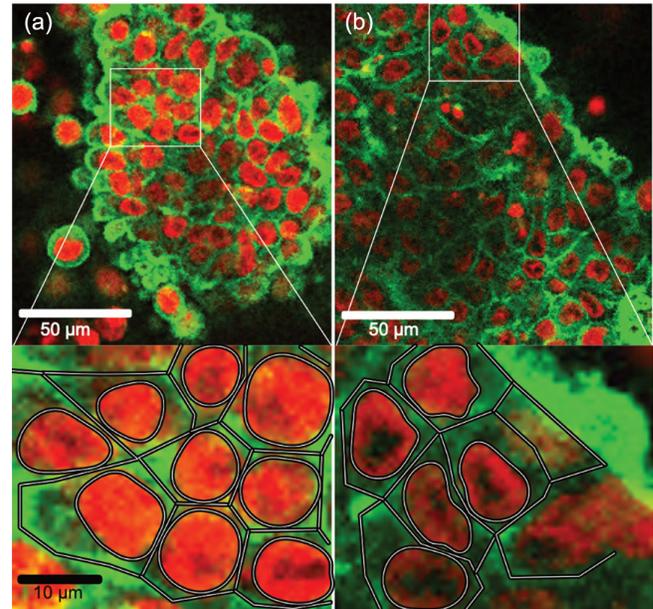


FIG. 1. Cancer cell clusters in patient-derived explants of primary invasive breast cancer stained for actin (green) and DNA (red). White lines emphasize nuclei and cell outlines. The left column (a) represents a region of more roundish cells and nuclei, while the right column (b) contains more elongated cells and nuclei. In both samples, cells are packed close to a volume fraction of 1.

with no space for a cell to move [9], it is not intuitively comprehensible how cells divide or migrate against their obstructing neighbors.

Cell unjamming of a majority of softer deformable cancer cells solves this problem [9,11]. These cells in a cluster move by collectively exchanging positions [9,12–15]. Cell aggregates with elongated cell and nucleus shapes permit cells to pass each other by T1 transitions [15–17]. T1 transitions are topological rearrangement processes that describe neighbor exchanges in confluent tissues [16]. Collective T1 transitions induced by elongated cells result in unjammed, motile regions within the cell clusters, whereas jammed regions are characterized by more roundish cells and nuclei that hinder collective rearrangement [9]. Therefore, one essential prerequisite for cancer cell motility is the active deformation of cells and nuclei to squeeze through and exchange positions with their neighbors, validated below by vital cell tracking and segmentation in patient-derived tumor explants; see Figs. 2 and 4.

This shape-dependent unjamming process is further modulated by the nucleus number density (number of nuclei/cells per area), which—up to now—was not incorporated as a mechanical variable in unjamming models. When cells squeeze by each other, the rigid nuclei touch each other as obstacles by the mechanical stress mediated by the softer surrounding cytoplasm. Since the cytoplasm is much softer than the nuclei, more cytoplasmic space

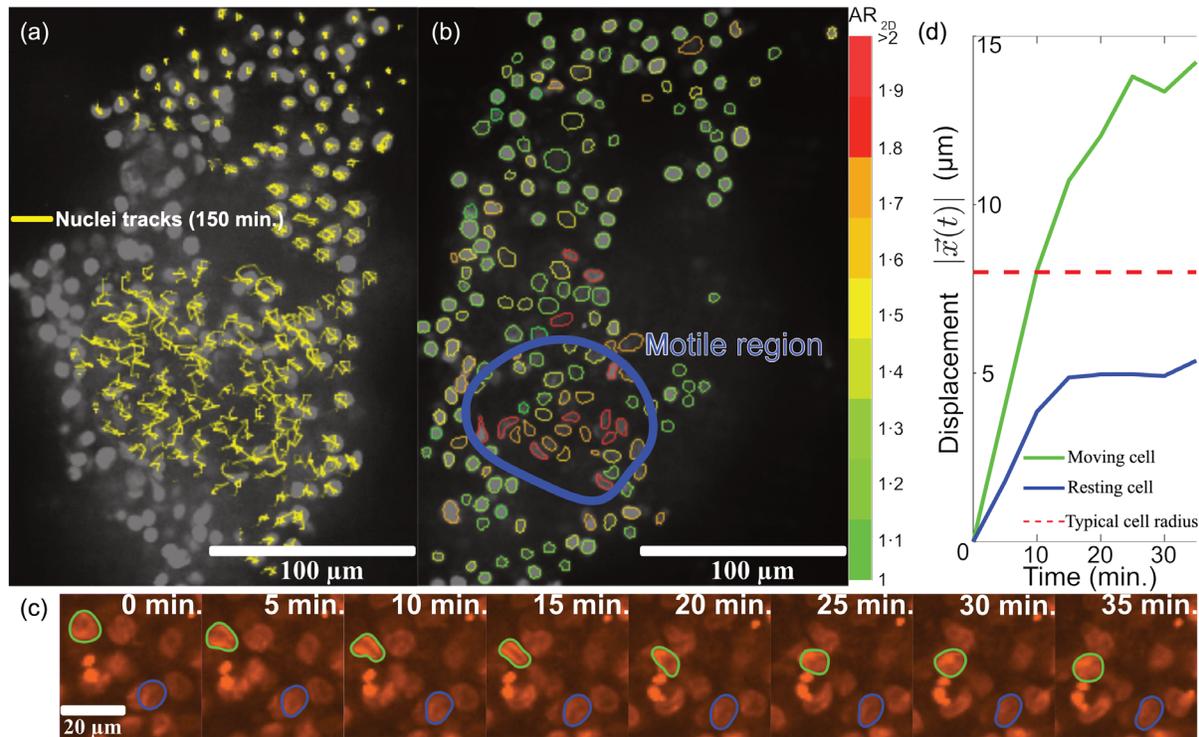


FIG. 2. Vital cell tracking in primary breast cancer explants. (a) Explant of a primary breast carcinoma stained with a vital DNA stain. Yellow lines indicate the tracks of the movement of the nuclei in a 150-minute time interval. An unjammed region of uncorrelated cell motion in the lower middle of the sample becomes visible. This region is surrounded by jammed cells. (b) Cell nuclei were segmented, and their corresponding AR was color-coded according to the color map on the right side (red is for high AR, elongated; green is for small AR, round). The region of motile cells exhibits more elongated nuclei shapes with respect to the surrounding neighbors. This analysis is conducted for the same sample as in panel (a) for an xy plane in the middle of the sample. (c) Zoom of individual tracks of a cell nucleus within the region of the motile cancer cells (green outlines) and of a cell nucleus within the surrounding matrix of jammed cells (blue outlines). The observation time is 35 minutes, displayed in 5-minute time steps. The moving cell is actively deforming and elongating, while the resting one exhibits a constant round shape. (d) Displacement of a motile cell (green line) and a resting cell (blue line). The displacement of the moving cell increases over time and exceeds a typical cell radius (dashed red line, $8 \mu\text{m}$). The displacement of the resting cell only increases for the first 15 minutes due to fluctuations within the cage formed by the surrounding cells.

between the nuclei might foster cell unjamming and less space jamming, respectively. Similar to the stiffening of cytoplasm in atomic force microscopy (AFM) experiments when indenting thin cell regions against the underlying rigid substrate [18], a stiffening between the nuclei of the clusters should occur when the cytoplasmic space that surrounds the nuclei is lowered, which will favor jamming.

In real cancer cell clusters, there are several variables related to packing: One is the cell volume (area) fraction, which is close to or equal to 1 and thus practically invariable, as seen in Figs. 1, 4(c), and 5(a). Another packing density is the nucleus volume (area) fraction discussed later. The critical density under investigation here is the nucleus number density ρ , which is highly variable within the clusters [see Fig. 4(c)] and is equal to the cell number density. Increased cell volumes, i.e., more cytoplasmic space around the nuclei and lower number densities, have been observed for unjammed cancer cell spheroids [9,19,20].

The remainder of this paper is structured as follows: In Sec. II, we present dynamic data of cancer cell unjamming via vital tracking in patient-derived tumor explants. Because of the limited availability of vital patient-derived tumor explants ($N = 16$), they cannot comprehensively probe the entire state space of unjamming. Compensating for this limited size, we conduct our exploratory retrospective histopathological study in Sec. III, where we systematically probe the space of unjamming states with both shape and density as static variables of motility to assign and classify the histological slides of 1380 breast cancer patients with respect to distant metastases. The patient data span an extensive and physiologically relevant state space [Figs. 4(b) and 4(c)] that is highly correlated with metastatic events (Fig. 6). Finally, the conjecture that these metastases were favored by cell motility in the primary tumor as an early event in the metastatic cascade is discussed in Sec. IV, as are the consequences of our state diagram to understand cell unjamming and its potential clinical relevance.

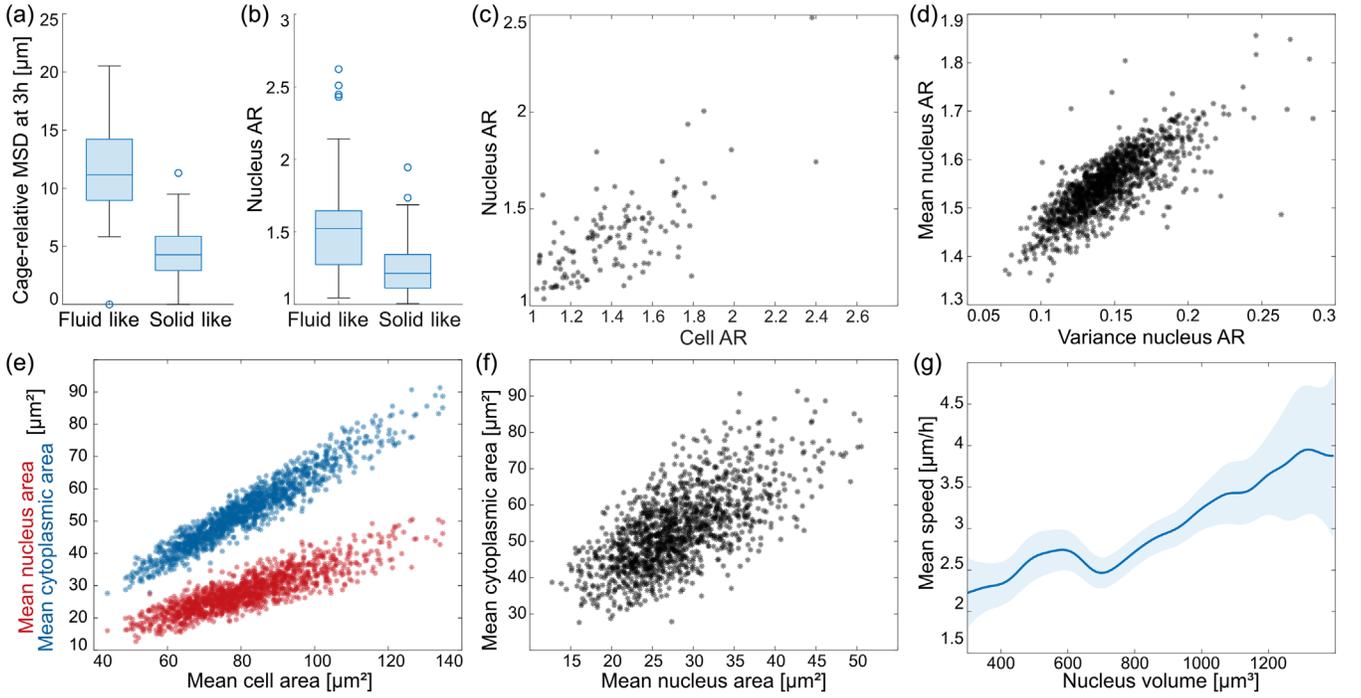


FIG. 3. (a,b) Quantification of the motile and nonmotile regions in vital tumor explants. (a) Measured mean displacements within 3 hours relative to nearest neighbors (cage-relative MSDs). Motile ($N = 22$) regions exhibit significantly (2sKS $p < 0.0001$) higher displacements above an average cell radius compared to jammed regions ($N = 22$) in primary carcinoma (breast, cervix). (b) Nuclei aspect ratios in primary breast carcinoma explants (shown in Fig. 2), which show a more elongated shape for motile ($N = 36$) compared to jammed regions ($N = 145$). 2sKS test: $p < 0.0001$. (c) Relationship between nucleus shape and cell shape. We show the linear correlation between the nucleus and cell aspect ratios, measured in primary breast and cervix cancer, fixated and stained for DNA and actin. The Pearson correlation coefficient is $r = 0.73$. (d) Correlation between mean and variance of the nuclei aspect ratios measured in H&E-stained breast cancer sections. One data point corresponds to one patient from the training or test collective ($N = 1380$). The Pearson correlation coefficient is $r = 0.79$. (e,f) Nucleus packing in histological slides ($N = 1380$) of invasive breast cancer. (e) Blue: correlation between the mean cell area and mean nucleus area in histological images of breast cancer patients. The Pearson correlation coefficient is $r = 0.86$. The red area shows the correlation of the mean cytoplasmic space (cell area minus nucleus area) and the mean cell area in histological images of invasive breast cancer patients. The Pearson correlation coefficient is $r = 0.95$. (f) Correlation of the mean nucleus area and the mean cytoplasmic area in histological images of invasive breast cancer patients. The Pearson correlation coefficient is $r = 0.67$. (g) Correlation of individual cell speed measured by nucleus tracking in vital breast cancer explants with a nucleus volume that scales inversely with nucleus number density [compare Fig. 15(b)]. We measure the mean cage relative speed within 2 hours. Cancer cell motility and, thus, unjamming increase with nucleus volume. We use a moving average with a window width of $100 \mu\text{m}^3$. Details and filtering can be found in Appendix D.

II. DYNAMICS IN VITAL PRIMARY SAMPLES

For dynamic cancer cell tracking data, we initiate pilot experiments with vital patient-derived tumor explants of two different carcinomas (four breast carcinomas and 12 cervix carcinomas) to show that the ability of cells to move in dense tissues can be recognized by cell and nucleus shape. Cancer cells are tracked during a six-hour observation period in 3D within primary tumors by live nucleus staining with a spinning disk microscope [9,21,22]. After the cell tracking experiments, we measure the cell shapes in fixed tumor tissues using an actin cytoskeleton stain. In the vital tumor explants of eight out of 16 patients, we find both areas of motile cancer cells and areas of jammed cancer cells within the cancer cell clusters. In the other explants, the cancer cell clusters are solely jammed. We refer to

Appendix L for detailed information on the vital primary samples and their fluidity status. The motile regions show that cancer cell unjamming occurs in solid tumors.

Motile regions are characterized by cells squeezing through their neighbors, clearly visible by nuclei elongations; see Fig. 2(c), Fig. 4(a), and the video of the vital patient-derived breast cancer explant given in the Supplemental Material [23]. Nuclei tracks cross each other in motile regions, and the displacements exceed cell sizes. In contrast, nonmotile regions exhibit vanishing nucleus movement and only show small fluctuations around their positions [Fig. 3(a)]. Motile, unjammed areas fundamentally differ from the healthy epithelium, which is jammed everywhere [24]. We measure the mean nucleus displacements relative to the next neighbors [cage-relative

mean-squared displacement (MSD)] during a 3-hour period. The nonmotile areas exhibit cage-relative MSDs not exceeding an average cell radius ($8\ \mu\text{m}$) [Fig. 3(a)]. In addition, the median relative distance traveled in motile regions is significantly higher than every movement in the area with nonmotile cells [two-sample Kolmogorov-Smirnov (2sKS) test: $p < 0.0001$; see Fig. 3(a)]. The 150-minute trajectories of the nucleus centers are depicted as yellow lines in Fig. 2(a). Cancer cells can have a fluidlike movement in the unjammed regions, while in the jammed regions, cancer cells are frozen in their positions as in a heterogeneous solid or glass.

Live nucleus tracking data are analyzed in motile and nonmotile regions in exemplary breast cancer explants from our primary samples concerning the 2D aspect ratios (AR) of the nuclei to connect cell motility and the (static) form of cancer cells. Guided by the tracks and a visual evaluation of three-dimensional videos [see the video of the vital sample (shown in Fig. 2) given in the Supplemental Material [23]], regions of jammed and unjammed cells are identified, as depicted in Fig. 2(b). The nuclear ARs are indicated by color-coding in the outlines of the cell nuclei for a spatial and temporal section of the tumor explant, as displayed in Fig. 2(b). Comparing the static ARs of the nuclei within the motile region with respect to the neighboring nonmotile region shows that motile cancer nuclei have significantly higher ARs than the jammed nuclei (2sKS test, $p < 0.0001$), meaning that the motile cells have a more elongated nucleus shape. A detailed look at the dynamics of two cells, one in the motile region and one in the adjacent jammed area, for an observation period of 35 minutes is shown in Fig. 2(c). The nucleus of the motile cell strongly deforms while the cancer cell actively squeezes itself through the dense microenvironment of other cancer cells [green outline, Fig. 2(c)]. The jammed, round nucleus does not change its low AR over time and does not move [blue outline, Fig. 2(c)]. In agreement with our cage-relative MSD measurements, the displacement of the moving cell exceeds the average cell radius during the observation time of 35 minutes, shown in Fig. 2(d). Our tracking data reveal that an elongated nucleus can characterize motile cancer cells squeezing through dense tissue. The nuclei in the displayed image sequence are surrounded by the cells' cytoplasm, which is not fluorescently labeled and thus not directly visible. However, when a nucleus deforms to squeeze through the dense environment of neighboring cells, the surrounding cytoplasm must also deform. Accordingly, the deformation of the whole cell correlates with the nucleus, as shown in Fig. 3(c). A motile cancer cell is not solely squeezed by the surrounding cells in the investigated cell clusters. The effect is reciprocal. The surrounding cells, in turn, are also deformed or elongated due to the moving cell, which results in the collective cell unjamming transition that we observe in vital tumor tissue (see video in the Supplemental Material [23]). Thus, we

conclude that elongated cell and nucleus shapes directly indicate cancer cell motility in static pictures.

However, the degree of deformation depends on the packing of nuclei and the number density since cells cannot move when nuclei get stuck or when they are caged by neighboring cells. Nuclei sizes directly correlate with the cell sizes in tumor clusters [Fig. 3(e)], cytoplasmic sizes around the nuclei [Fig. 3(f)], and the inverse number density [Fig. 15(b)]. Importantly, the effective cytoplasmic distance from the nucleus edge to the cell edge increases with increasing nucleus sizes [Fig. 15(a)]. Consequently, we expect unjamming for large nucleus sizes since this corresponds to low number densities, more cytoplasmic space, and longer distances between nucleus edges and cell membranes, which increases the distance to neighboring nuclei of adjacent cells.

Our vital tracking data in tumor explants, shown in Fig. 2 and the video in the Supplemental Material [23], indeed reveal that nucleus volumes strongly influence the cell motility detected, as shown in Figs. 4(a) and 3(g) (see also Appendix D).

We summarize our tracking data and information on shape and density in a first outline of the state diagram of cancer cell unjamming in Fig. 4(a) (details are given in Appendix D). We plot the binned nucleus volume (scales inversely with number density; compare Appendix N) versus the binned 3D nucleus ellipsoid shape (defined in Appendix D) and color-code the space with the mean cell speed within 2 hours. The black line plots an approximate transition boundary separating caged (jammed) and uncaged (unjammed) cells. The boundary has been determined by a binary support vector machine classification (see Appendix I) of the binned tiles with displacements exceeding a typical cell radius of $8\ \mu\text{m}$ and those not exceeding this radius in a 2-hour interval. Below the transition boundary, average displacements never exceed $8\ \mu\text{m}$ within 2 hours. Consequently, 55 of 55 tiles below the threshold do not exceed a typical cell radius in their displacements. Above the boundary, we find all the tiles that exceed an average displacement of $8\ \mu\text{m}$. Out of a total of 25 tiles above the threshold, 11 tiles exceed a displacement of a typical cell radius. While motility correlations for both axes are strong and depend on the region in the diagram concerning which variable dominates, overall, the nucleus volume as a measure of inverse number density and nucleus spacing (see Appendix N) exhibits a more prominent role in identifying the uncaged sections in the parameter space, as seen from the decision boundary in Fig. 4(a).

The amount of vital tumor tracking data restricts the detail in this state diagram. However, these observations emphasize that nucleus number density and cell and nucleus shapes are essential drivers of cancer cell unjamming. This motivated us to probe the static unjamming state space with an extensive number of retrospective histological images of breast cancer patients. For simplicity and

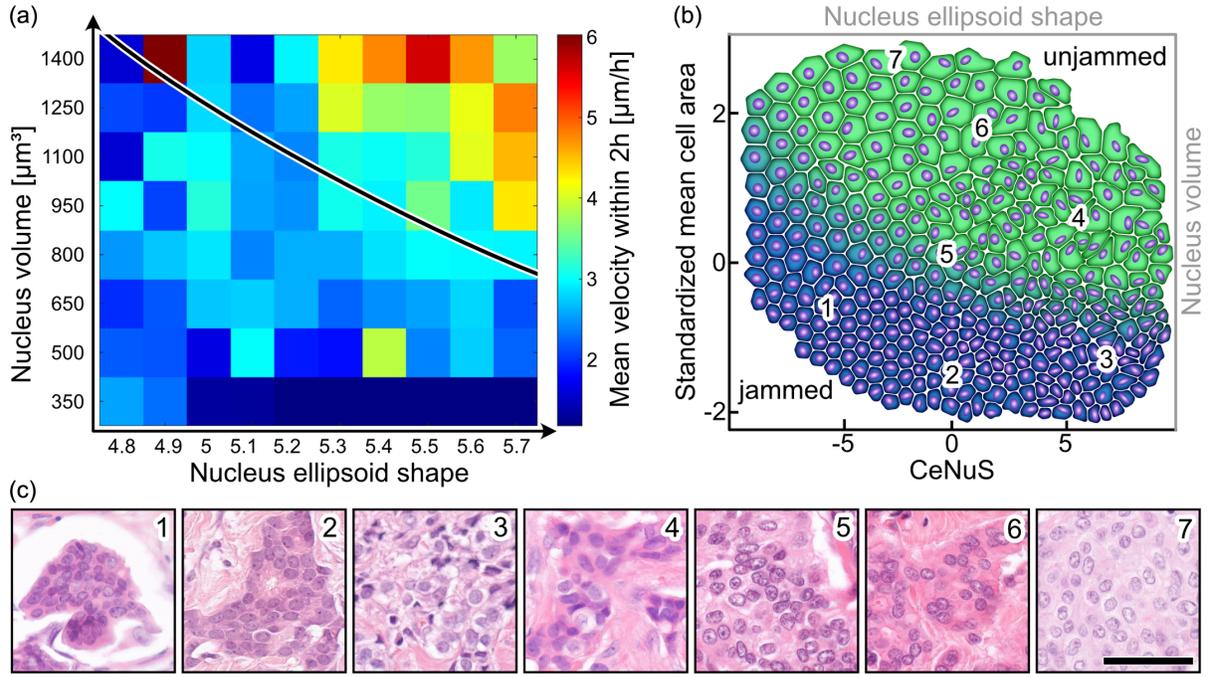


FIG. 4. Derivation of the state diagram of cancer cell unjamming. (a) Tracking data in vital tumor explants. The nucleus volume and nucleus ellipsoid shape s_E are used as axes. These structural observables are binned, and the average speed within 2 hours [$\sqrt{\text{MSD}(t = 2\text{h})}/(2\text{h})$, where MSD is the cage-relative mean-squared displacement] of the tiles is depicted as a heat map. More elongated nuclei and larger nuclei are strongly correlated with higher motility. The black line corresponds to a binary SVM classification between tiles whose displacements exceed a typical cell radius (8 μm) on average and tiles that do not. (b) State diagram of cancer cell unjamming based on histological images from 1380 breast cancer patients connecting tissue morphology with collective cancer cell motility. Nuclei are depicted in purple, and cells are color-coded by a blue-to-green heat map, where blue indicates a jammed state and green a motile state. Note that the diagram does not reflect a linear increase in motility. The region where blue turns to green depicts where we expect the cell unjamming transition to occur. The abscissa varies the cell and nucleus shape (CeNuS) defined in Eq. (1). The ordinate varies the standardized mean cell area \bar{A}_C^{stand} , reflecting the inverse nucleus number density. The variables are measured in tracking experiments [in part (a)] are depicted in gray: The nucleus ellipsoid shape corresponds to CeNuS, and the nucleus volume scales with the inverse nucleus number density represented by \bar{A}_C^{stand} in 2D. (c) Seven images of H&E-stained tissue sections of invasive breast cancer associated with the indicated regions demonstrate that real breast cancer patient data span the whole state space. The scale bar equals 50 μm .

because most medical images are 2D, we use appropriate 2D shape and density variables for our state diagram. As recently shown in [25], 2D observables measured on 2D cuts from 3D cell packings still contain much information with respect to the geometrical cell properties, especially if the number of cells analyzed is high.

III. EXPLORATORY RETROSPECTIVE HISTOPATHOLOGICAL STUDY

This section presents our exploratory retrospective histopathological study with no predetermined trial design. Therein, we analyze the 2D H&E-stained breast cancer tissue sections of 1380 patients and map them to our state diagram with respect to the 2D static observables of unjamming displayed in Fig. 4(b). The large amount of patient data serves as the basis to determine the range of jammed and unjammed states in solid tumors. We

hypothesize that cancer cell unjamming, as the onset of cancer cell motility, favors distant metastasis. If our static state space correlates with distant metastasis risk, it suggests that cancer cell unjamming is an early part of the metastatic cascade via the mechanism of cancer cell motility within the primary tumor and justifies the usefulness of our unjamming-based static cell motility markers.

Representing the 2D-shape axis of the state diagram, we have combined cell and nucleus shapes as static measures of cancer cell motility into one ensemble variable: CeNuS defined in Eq. (1). The nucleus shape is quantified by the AR and the cell shape by a dimensionless measure of cell elongation, the so-called cell shape index used in previous theoretical publications [14,15,26] ($p = P/A^{0.5}$, where P is the perimeter and A the area of a cell); see Fig. 5(e). For a detailed explanation of the choice of quantitative measures, see Sec. V. CeNuS is defined as the sum of the standardized median cell shape index \bar{p}^{stand} , the variance of the cell

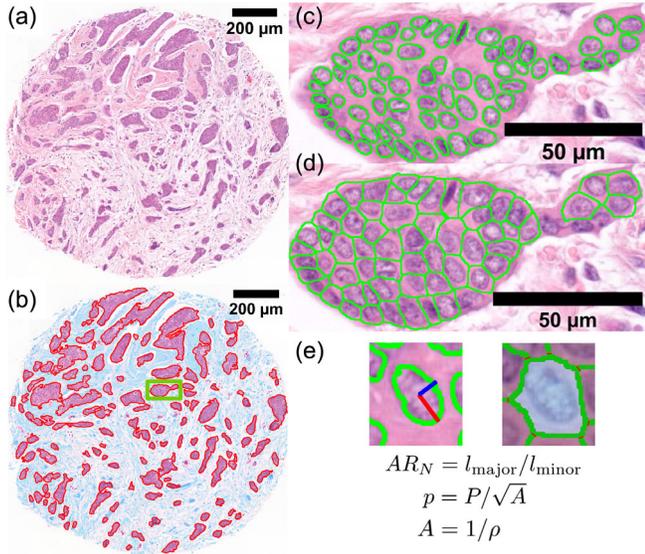


FIG. 5. Image analysis of the histopathological H&E-stained slides to map breast cancer patients within our state diagram. (a) Representative native H&E-stained tumor tissue of cancer cell clusters surrounded by stroma. (c) Nuclei segmentation within the cancer cell clusters to measure the nucleus AR. (d) Cell outline segmentation within the cancer cell clusters to measure the shape index. (e) Quantification of the nucleus shape by the AR (AR_N), the cell shape by the cell shape index (p), and of the inverse nucleus number density by the cell area (A).

shape index $\sigma_p^{2\text{stand}}$, the median nucleus AR $\bar{AR}_N^{\text{stand}}$, and the variance of the nucleus AR $\sigma_{AR_N}^{2\text{stand}}$:

$$\text{CeNuS} = \bar{p}^{\text{stand}} + \sigma_p^{2\text{stand}} + \bar{AR}_N^{\text{stand}} + \sigma_{AR_N}^{2\text{stand}}. \quad (1)$$

Since the cell shape indices and the nucleus AR distributions are right-skewed (Appendix G), the median is used to measure the average in the bulk population. However, highly elongated shapes (cell, nucleus) in the tail of the distribution could also contain vital information about highly motile cells; therefore, the variance from the *mean* is incorporated to reflect the distribution width. We average overall cancer cells in the clusters because this is the most robust measure. A significant linear correlation between ensemble means of the nucleus ARs and their standard deviations are visible for cases from the training and test collective in Fig. 3(d). This agrees with an increase in heterogeneity when regions in the cancer cell clusters collectively unjam, driven by elongated cell and nucleus shapes. The correlation between means and variances also underlies the amorphous structure of the cancer cell aggregates. A deformed nucleus necessitates a deformed cell [Fig. 3(c)]. However, for a low nucleus number density (large cytoplasmic spaces around a nucleus; compare Figs. 3 and 15), the shape of a cell can be strongly deformed while the nucleus is round. Thus, it is better to use CeNuS instead of the sole elongation of the cell or

nucleus. The use of only one of these parameters as state diagram axes is discussed in Sec. III, while the quantitative analysis can be found in Appendix J.

To quantify the nucleus number density in 2D, the average cell area in clusters \bar{A}_C is used since it indicates how much space a single nucleus has around itself with respect to the adjacent nuclei. The mean cell area is proportional to the inverse nucleus number density $\bar{A}_C \sim 1/\rho_{2D}$. The z-standardization of both state variables was performed using the standard deviations and means of a representative sample of a patient cohort of 530 consecutive cases provided by the Institute of Pathology Hamburg-West; see Table VIII in Appendix E for exact values. With these values, other researchers and physicians can standardize their data to apply the presented physical motility marker to evaluate metastatic risk.

The image analysis of H&E-stained breast cancer slides to measure our state variables starts with a tissue segmentation into cancer cell clusters, stroma, and white background, as depicted in Fig. 5(b). Nucleus shapes in the cancer cell cluster are extracted with the StarDist model [27]; see Fig. 5(c). We employ a watershed algorithm to approximate the cell outlines by utilizing the nucleus segments as initialization points and the edges of the cancer clusters as constraints. This results in a good representation of the cell outlines that lie halfway between the nuclei and are constrained to the cancer cluster; see Fig. 5(d). For a detailed description and benchmarking of the reliability of the algorithms, see Methods and Appendix C.

Before dividing our breast cancer patient collective of 1380 patients into a training and a test collective, we map all histopathological slides to our state diagram with respect to CeNuS and cell space. The heterogeneous cell clusters of breast carcinomas span a large state space. Figure 4(b) displays our unjamming state diagram with a schematic representation of the observed tissue morphology as a function of CeNuS and cell area. Figure 4(c) represents examples of histological images from patients in different regions of the state diagram. Based on our cell tracking data [see Fig. 4(a)], we indicate by color-coding where we expect the cells are jammed (blue) or unjammed (green). The 1380 breast tumor cutouts represent the broad range of unjammed or jammed states of cancer cells, which includes regions that are dominated by shape-driven or nucleus-number-density driven behavior. In our retrospective study of breast cancer, a training set consisting of 688 patients and an independent test set with 692 patients are used (see Table XI). By independent, we mean that the collectives are not overlapping with respect to the patients investigated. The training set consists of patients treated only minimally with endocrine (hormone) therapy and/or radiotherapy after surgery. In this group of patients, no patient was treated with cytotoxic chemotherapy. Therefore, we assume that patients in the training set exhibit fewer treatment effects,

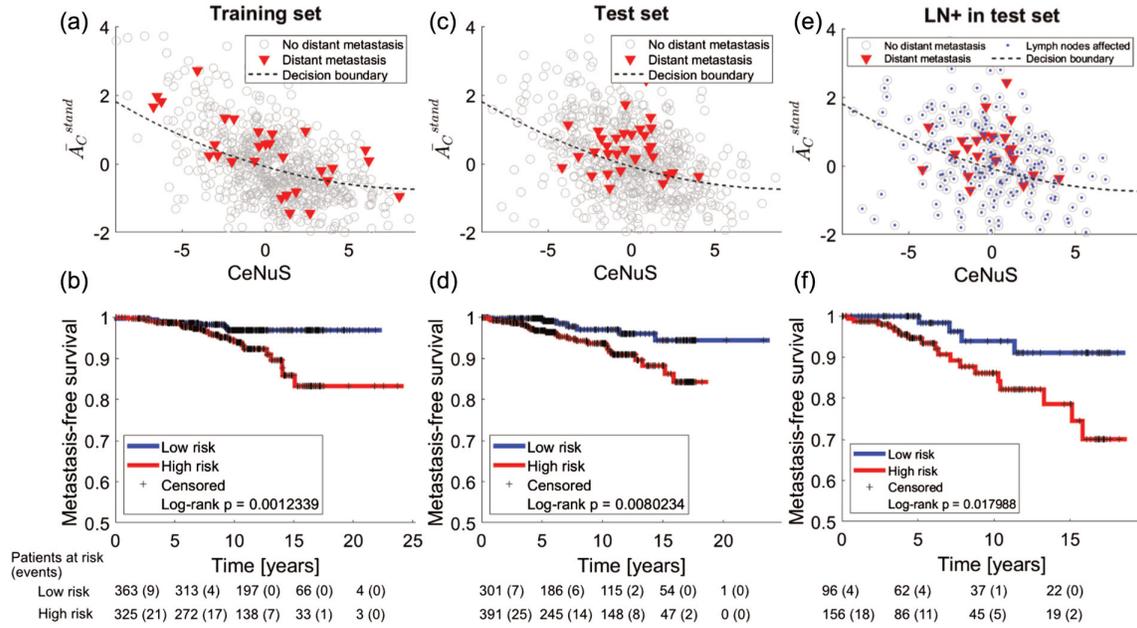


FIG. 6. Distant metastasis and clinical relevance of cancer cell unjamming. The distribution of patient outcomes in terms of distant metastasis is displayed with respect to the variables of the state diagram (cell area, CeNuS) for the training collective ($N = 688$) and for the test collective ($N = 692$). The state diagrams are shown in panels (a) and (c), respectively. Patients who developed distant metastases are indicated by red triangles. Patients who did not develop metastases are indicated by gray circles. The dotted line represents the decision boundary that separates the low-risk patients for distant metastasis (below) from the high-risk group (above) estimated by SVM classification; see Appendix I. Kaplan-Meier estimators are used for assessing the quality of our risk stratification based on cancer cell unjamming. The resulting Kaplan-Meier plots are shown in panel (b) for the training set and panel (d) for the test set. They show that the high-risk group (in red) is well separated from the low-risk group (in blue). The log-rank p -value for the training set is $p = 0.0012$, and for the test set, $p = 0.0080$. In panel (e), we display only the lymph-node-positive (LN+) patients from the test set ($N = 252$) to illustrate the complementarity of lymph node status and cell unjamming. The dotted line represents the decision boundary found in the training set, which yields significant separation between risk groups, as seen in the Kaplan-Meier plot in panel (f) (log-rank $p = 0.018$). “Patients at risk” indicates the total number of patients who have not been censored or have not developed distant metastasis by a specific follow-up time. In parentheses after the “Patients at risk” is the cumulative number of individuals developing distant metastases later.

which may alter the occurrences of distant metastasis. The test collective did not exclude cases where the patients were treated with chemotherapy or the therapy was not documented. After training, this collective of 692 patients (descriptive statistic in Table XI) was used to validate the hypothesis that the static state diagram of unjamming correlates with distant metastasis. Chemotherapy may prevent some of these patients from distant metastasis. Nevertheless, the presented prognostic unjamming criterion is not chemotherapy dependent (see Cox models in Tables V and III) and should thus also be able to stratify the patients with distant metastasis in this collective.

The clinical follow-up data with which we correlate the state diagram are summarized in the descriptive statistics in Tables IX–XI. Distant metastasis refers to the spread of cancer from the primary tumor to distant organs. We have chosen the reported distant metastasis in the patient’s follow-up information as our main event under investigation since metastasis requires cell motility and unjamming supports motility very early on. Our patients’ histories do not discriminate between cancer-related deaths and deaths from other causes. Therefore, distant metastasis as an event

is the more reliable parameter for risk assessment to determine whether our static state diagram of cell unjamming is connected to disease progression and bad prognosis. For an analysis of the prognostic value of unjamming with respect to reported “overall death,” we refer to Sec. IV and Appendix M.

In the next step, we study where distant metastatic events are found within the state diagram among the patients in our training set; see Fig. 6(a). We notice that regions characterized by round cancer cells and nuclei (small CeNuS) and/or high nucleus number densities (small cell areas) are nearly free of patients with distant metastatic events. Conceptually, this fits our hypothesis that cancer cell unjamming is an essential early event within the metastatic cascade since we expect lower cellular motility with rounder nuclei and cells and high nucleus number density. Since current cell unjamming theories do not consider the contribution of the cell nucleus [12] and do not reflect the fact that unjamming depends on shape as well as density, there are no useful quantitative predictions where the cell unjamming transition occurs for our data. We can use our tracking data to estimate where cell unjamming

occurs qualitatively. However, our state diagram cannot fall back on a theoretical, quantitative description of the unjamming boundaries. Thus, we determine an empirical unjamming threshold via the support vector machine (SVM) distant metastasis classification in the state space of our training set [Fig. 6(a)]. A kernel is employed that constrains the boundary to our physical observations from cell tracking [Figs. 4(a) and 4(b)]: Motility through unjamming is correlated with the elongation of cells and nuclei and low nucleus number density. We refer to Appendix I for the model and training details. The resulting boundary (see dotted line in Fig. 6) separates the space into two regions such that a point above the boundary is associated with unjamming and high distant metastatic risk and a point below is associated with a jammed tumor structure and low risk (see Fig. 6).

Among the 688 patients in the training collective, this empirical unjamming threshold identified 21 of 30 patients with distant metastases. The Kaplan-Meier representation of the low- and high-risk groups in the training set in Fig. 6(b) shows significant stratification with a log-rank p -value of $p = 0.0012$. Confirming this in the test set of 692 patients, most distant metastases (25 out of a total of 32 distant metastatic occurrences) are located in the high-risk area [Fig. 6(c)]. The Kaplan-Meier estimators of the low- and high-risk groups in the test set in Fig. 6(d) display significant discrimination with a log-rank p -value of $p = 0.008$. Unjamming as a prognostic marker yields $366/391 = 93.6\%$ false-positive patients in the high-risk group. This is comparable to the false-positive rate in the high-risk group defined by the current clinical standard, the lymph node status, with $230/252 = 91.3\%$ false-positive patients in the high-risk group that do not develop distant metastasis despite their affected lymph nodes. Applying the unjamming threshold to these patients, presented in Figs. 6(e) and 6(f), reduces the false-positive rate in lymph-node-positive patients to $138/156 = 88.5\%$.

Regarding the false negative ratio within the low-risk group in the test set (1-NPV; see Table I) defined by the jamming threshold, we see that $7/301 = 2.3\%$ of patients develop metastasis despite being in the low-risk group. For the lymph node status, we observe $10/440 = 2.3\%$ patients who develop distant metastases despite exhibiting unaffected lymph nodes. Therefore, also in the false-negative ratio (or negative predictive value), the jamming threshold and the lymph node status yield comparable results.

In the test set, unjamming leads to a sensitivity of 0.78 (training 0.70) and a specificity of 0.45 (training 0.54) (Table I; see details in Appendix H). Thus, the sensitivity is higher compared to the lymph node status (0.78 versus 0.69) while the specificity is lower (0.45 versus 0.65).

To put this multitude of measures into perspective, we use Harrel's concordance index (c-index) as a unifying evaluation metric [28] to verify that cancer cell unjamming is highly relevant to distant metastasis prognostication. The

TABLE I. Analyses of the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the c-index for the standardized mean cell area \bar{A}_C^{stand} , CeNuS, the decision boundary presented in Fig. 6 (DB), and the LNS as the standard measure for distant metastatic risk. In brackets (...), we show the order of magnitude of the $\Delta\text{CI}/2$ of the 95% bootstrap confidence interval (CI) for 1000 resamplings. Values are calculated in the test set. We refer to Appendix H for details.

| | \bar{A}_C^{stand} | CeNuS | DB | LNS |
|-------------|----------------------------|---------------------|---------------------|---------------------|
| Sensitivity | 0.719 (10^{-3}) | 0.781 (10^{-3}) | 0.781 (10^{-3}) | 0.688 (10^{-3}) |
| Specificity | 0.473 (10^{-3}) | 0.276 (10^{-3}) | 0.446 (10^{-3}) | 0.652 (10^{-3}) |
| PPV | 0.062 (10^{-4}) | 0.050 (10^{-4}) | 0.063 (10^{-4}) | 0.087 (10^{-3}) |
| NPV | 0.972 (10^{-4}) | 0.963 (10^{-4}) | 0.977 (10^{-4}) | 0.977 (10^{-4}) |
| c-index | 0.606 (10^{-3}) | 0.554 (10^{-3}) | 0.627 (10^{-3}) | 0.659 (10^{-3}) |

c-index measures a prognostic model's or score's ability to differentiate patient outcomes. A c-index of 0.5 corresponds to an entirely random prognosis while a value of 1.0 indicates the perfect ordering of prognosticated outcome probabilities. A c-index of 0 means perfect anticoncordance. One can interpret the c-index as the percentage of patients that can be correctly ordered; e.g., 0.7 indicates correct ordering in 70% of cases.

If cancer cell unjamming is essential to metastasis, which has not yet been considered clinically, it provides additional information with respect to the current diagnosis. This becomes evident if the c-index increases when combining current clinical information with our unjamming criterion. Therefore, we validate via Cox models, survival models based on regression analysis, that our measure of unjamming is a complementary and independent addition to the lymph node status, tumor grade, tumor size, and chemotherapy status with regard to distant metastases.

Considering these established clinical variables, the Nottingham prognostic index [29] and the jamming threshold independently in univariable Cox models (see Table V in Appendix B), all variables (except the grade) provide significant prognosis in the test collective, which demonstrates the diagnostic relevance of these variables.

When combining the established prognostic criteria of tumor diameter, tumor grade, and the lymph node status (all parameters are used in the Nottingham index) with the information of the unjamming threshold in a multivariable Cox model in Table II, only the unjamming threshold and the lymph node status remain significant. The regression coefficient b for the lymph node status is positive [$b = 1.30$, $\exp(b) = 3.69$], indicating a worse prognosis when one or more lymph nodes are affected; that is, the estimated hazard increases 3.69 times. The regression coefficient for the jamming threshold is -1.04 , and the hazard ratio is 0.36. Thus, the jamming threshold shows a better prognosis for subjects whose jamming score is larger since we defined it to be 1 for patients below the

TABLE II. A multivariable Cox model compares our unjamming criteria (decision boundary in Fig. 6) with established prognostic markers of breast cancer progression. The model includes unjamming (1 for jammed states under the decision boundaries, 0 otherwise), the largest tumor diameter, the tumor grade, and the lymph node status (1 if lymph node-positive, 0 otherwise). All variance inflation factors of the used predictor variables are smaller than 1.6, indicating vanishing collinearity. Likelihood Ratio Test $p < 0.0001$. The c-index of the model equals 0.7126. While without unjamming a c-index of only 0.6684 is achieved; see Table VI.

| | Coefficient b | SE(b) | exp(b) | p |
|-------------------|-----------------|-----------|------------|-------|
| Jamming threshold | -1.04 | 0.44 | 0.36 | 0.018 |
| Tumor diameter | 0.19 | 0.14 | 1.20 | 0.199 |
| Grade | -0.19 | 0.28 | 0.83 | 0.513 |
| Lymph node status | 1.30 | 0.43 | 3.69 | 0.002 |

decision boundary presented in Fig. 6 and 0 otherwise, reducing the hazard to 0.36 (36%) for subjects whose tumor morphology indicates jammed cancer cell aggregates. The c-index of this model equals 0.7126. The c-index of the multivariable Cox model using only the established prognostic criteria (tumor size, tumor grade, and lymph node status; see Table VI in Appendix B) without the information of the unjamming threshold equals 0.6684. Thus, using the unjamming threshold in combination with these established prognostic criteria yields a $(0.7126 - 0.6684)/(0.6684 - 0.5) = 0.26 = 26\%$ information gain compared to the use of these established prognostic parameters alone.

We still have to show that chemotherapy does not influence our prognostic results. When adding the chemotherapy status as a predictor variable to the tumor size, tumor grade, lymph node status, and unjamming threshold to the multivariable Cox model shown in Table III, only the jamming threshold and chemotherapy status remain significant. This shows that the low- and high-risk groups found by unjamming are not just caused by a treatment effect. Chemotherapy may prevent distant metastasis in some cases, but it does not alter the observed correlation between cancer cell unjamming and distant metastasis. As expected, the regression coefficient b for the chemotherapy status is positive [$b = 1.55$, $\exp(b) = 4.71$], indicating a worse prognosis when chemotherapy is required: The estimated hazard increases 4.71 times. The regression coefficient for the jamming threshold is -0.97 , and the hazard ratio is 0.41. Thus, a jammed tumor structure consistently indicates a better patient prognosis.

Static observables of cancer cell unjamming strongly correlate with a breast cancer patient's risk for distant metastasis. Our results imply that unjamming as an increase of cancer cell motility in the primary tumor is an early part of the metastatic cascade, previously not considered in diagnosis, since all observables are measured in the primary

TABLE III. A multivariable Cox model is used to evaluate the independent prognostic power of the unjamming threshold (decision boundary in Fig. 6) with respect to established prognostic markers and the chemotherapy status. The model includes the jamming threshold (1 for jammed under the decision boundary, 0 otherwise), the largest tumor diameter, the grade, and the lymph node status (1 if nodes are infiltrated, 0 otherwise) and the chemotherapy status (0 for treatment is unknown or not indicated, 1 for treatment). The jamming threshold and the chemotherapy status are the only significant contributors in the model and thus are not interdependent. All variance inflation factors of the used predictor variables are smaller than 1.7, indicating vanishing collinearity. Likelihood Ratio Test $p < 0.0001$. The concordance index of the model equals 0.7706.

| | Coefficient b | SE(b) | exp(b) | p |
|---------------------|-----------------|-----------|------------|-------|
| Jamming threshold | -0.97 | 0.44 | 0.38 | 0.026 |
| Tumor diameter | 0.15 | 0.16 | 1.16 | 0.334 |
| Grade | -0.35 | 0.29 | 0.70 | 0.221 |
| Lymph node status | 0.82 | 0.43 | 2.27 | 0.055 |
| Chemotherapy status | 1.55 | 0.50 | 4.71 | 0.002 |

tumor and not, for instance, in the blood vessels or lymph nodes.

Unjamming is a new independent variable that provides additional prognostic information to established clinical parameters to better stratify breast cancer patients, leading to more precise medicine as demonstrated via multivariable Cox models. Thus, cancer cell unjamming could help forecast distant metastasis that other diagnostic measures would miss.

The metastatic cascade contains many more complex steps, which are not all known yet, and even parallel pathways through the lymph and vascular systems are possible. Thus, it is impossible to establish a direct causal relation between unjamming and distant metastases without direct measurements of cell dynamics *in vivo* over several (tens of) years. Not only is this unethical, but it is also not technically feasible. The next best thing is to establish a strong correlation between observables of cancer cell unjamming and relapses through distant metastasis. Such correlations are often used to initiate new diagnoses and therapies. If clinically validated, the causal proof is no longer required to be medically valuable. Moreover, the observed correlation confirms the use of our static motility marker and the state diagram.

The additional prognostic information that cancer cell unjamming provides becomes particularly evident in the complementarity of unjamming and lymph node status. In our state diagram for both the training and the test collective, low-risk and high-risk sections are not directly correlated with the lymph node status (Pearson's correlation coefficient 0.08); see Figs. 6(e), 6(f) (test collective), and 7 (training collective). The patients with positive lymph node status who lie below the jamming decision boundary (good prognosis) rarely developed distant metastasis.

TABLE IV. Combining the prognostic lymph node status with the prognostic unjamming criterion: Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and c-Index are calculated for: “Combination 1 (C1)” which is defined such that the low-risk group is lymph node-negative *and* below the jamming threshold (all others are in the high-risk group), “Combination 2 (C2)” which is defined such that the high-risk group consists of patients with affected lymph nodes *and* above the jamming threshold (all other are in the low-risk group), “Unjamming in LN- (ULN-)” which is the jamming threshold applied to all lymph node-negative (LN-) patients and “Unjamming in LN+ (ULN+)” which is the unjamming threshold applied to all lymph node-positive patients (LN+). In brackets (...), we show the order of magnitude of the $\Delta\text{CI}/2$ of the 95% Bootstrap confidence interval (CI) for 1 000 resamplings. Values are calculated in the test set.

| | C1 | C2 | ULN- | ULN+ |
|-------------|---------------------|---------------------|---------------------|---------------------|
| Sensitivity | 0.906 (10^{-3}) | 0.563 (10^{-3}) | 0.700 (10^{-3}) | 0.818 (10^{-3}) |
| Specificity | 0.306 (10^{-3}) | 0.791 (10^{-4}) | 0.470 (10^{-3}) | 0.400 (10^{-3}) |
| PPV | 0.060 (10^{-4}) | 0.115 (10^{-3}) | 0.030 (10^{-4}) | 0.115 (10^{-3}) |
| NPV | 0.986 (10^{-4}) | 0.974 (10^{-4}) | 0.985 (10^{-4}) | 0.958 (10^{-3}) |
| c-Index | 0.605 (10^{-3}) | 0.680 (10^{-3}) | 0.602 (10^{-3}) | 0.635 (10^{-3}) |

Below the jamming threshold in the test collective, only $4/96 = 4.2\%$ (training collective $0/41 = 0\%$) of patients developed distant metastases, while the group above the jamming threshold exhibited $18/156 = 11.5\%$ (training collective $5/60 = 8.3\%$) patients who developed metastases; see Figs. 6(e), 6(f), and 7. Consequently, for all lymph-node-positive patients, we see significant stratification by using our unjamming criterion within the training set (Fig. 7, log-rank $p = 0.009$) but, more importantly, in the test set [Figs. 6(e) and 6(f), log-rank $p = 0.018$]. Thus, risk refinement for nodal-positive patients by the unjamming threshold could lead to a more reliable prognosis and could thereby prevent nodal-positive patients from over-treatment since, from all the nodal-positive patients in our collectives, 36.5% received chemotherapy, while from the nodal-negative patients, only 7.3% received chemotherapy. From a total of 252 lymph-node-positive patients in the test set, 230 did not develop distant metastases and are false positives. From these 252 patients, the jamming threshold can identify a significant (log-rank $p = 0.018$) low-risk group of 96 patients, where the risk of relapse through distant metastases is 0% within the first 5 years and about 10% after 15 years, compared to around 30% after 15 years in the high-risk group.

Within the nodal-positive collective in the test set [Figs. 6(e) and 6(f)], the jamming threshold achieves a sensitivity of 0.82, a specificity of 0.40, and a c-index of 0.64; see Table IV.

In the more detailed analysis in Appendix A, we combine the information on the lymph node status with our prognostic measure based on unjamming. Together, they provide more information, which is suited for finding a

very safe group of patients with a very low risk of distant metastasis [no lymph nodes affected and jammed tumor structure; see Fig. 8(a) and Table I “Combination 1”] and a very vulnerable group of patients for distant metastasis [lymph nodes affected and unjammed tumor structure; see Fig. 8(b) and Table I “Combination 2”].

The prognostic connection between cancer cell unjamming in the primary tumor and distant metastatic events implies that unjamming is essential to tumor progression and provides valuable information that is not covered by the lymph node status.

IV. DISCUSSION

In agreement with Fredberg’s hypothesis that healthy tissues are close to unjamming [30], we have found that the pathological changes in tumors enable cancer cell unjamming as a collective motility transition [9]. Unjammed cells are found in embryonic morphogenesis in zebrafish [31–33]. When development is finished, jamming freezes the tissue in its state. Thus, jammed healthy tissue with its mechanical stability may be close to a cell unjamming transition [30], and small pathological perturbations may induce large mechanical changes [32,33]. Therefore, pathological changes such as neoplasia may induce unjamming in cancer as an early step to disseminate significant amounts of cancer cells into the human body [30]. As already indicated for asthma [26], cell unjamming may play a role in the pathological changes of other diseases. Our findings on the relevance of cancer cell unjamming for tumor progression may be paradigmatic for the importance of unjamming in medicine. In particular, we expect that cancer cell unjamming is relevant to all solid tumors, not only carcinoma, which would concern around 92% of all cancers [34].

This result explains and motivates the recent focus on cell unjamming in soft matter physics [12,35]. Nevertheless, a complete understanding of cell unjamming is lacking. Different experimental observations have led to controversial interpretations, and theories had predictive power for some experiments [26] but do not consider all effects that impact cell unjamming. This illustrates the inherent problem of finding the correct state diagram for cancer cell unjamming since we cannot experimentally control living tissues in a fashion that permits us to probe the whole state space systematically. Previous experiments could only capture small cutouts of the state diagram, leading to a limited view concerning the drivers of cell unjamming.

Here, we address this problem with a unique approach that establishes a morphodynamic link similar to the use of rings in a tree slice in dendrochronology. We use vital cancer cell tracking to obtain fundamental dynamical information. However, the problem remains that we can only study small cutouts of the whole tumor (around $100 \mu\text{m}^3$) and a small ensemble size (breast cancer $N = 4$

and uterine cervical cancer $N = 12$; compare Appendix L). A solid tumor is highly heterogeneous, which may be why we did not find unjamming in all vital tumor pieces (Appendix L). Since we cannot investigate a significantly higher number of vital explants and since we have no sufficiently long patient histories for these patients, we have developed a morphological cancer cell motility marker via CeNuS and \bar{A}_C^{stand} that permits us to evaluate a large number of histological patient slides. With this morphodynamic link, we have sampled and mapped to our state diagram the broad range of jammed or unjammed states present in breast tumors.

Direct immunohistochemical markers for a cancer cell that moves in the primary tumor do not exist. Tumor biology discusses the EMT with its specific molecular changes as the onset of cancer cell motility [36,37]. However, cancer cells do not show a clear, complete transition; instead, they assume a spectrum of molecular transitional states [38,39]. Thus, a molecular state defining cell motility has not yet been found. Therefore, detecting the morphological prerequisites for cell motility by our state diagram for static pictures may become a needed universal histopathological cell motility marker.

We stress the role of the nucleus in cell unjamming since it has not been considered previously. Besides by Grosser *et al.* [9], the nucleus is not mentioned in the cell unjamming literature, to our knowledge, which is surprising considering the central role of nucleus jamming in the stroma [40,41]. Our state diagram paints a completely different picture of the role of the nucleus since the nucleus enters both state variables. Notably, as cell and nucleus shapes are correlated, we find that either one may be used as a state diagram axis to significantly assess distant metastatic risk in histological patient data (see Appendix J for detailed analyses). However, we use CeNuS since this measure better reflects that both cells and nuclei must be elongated during migration in dense tissues and since it yields the most significant separation of risk groups.

The role of nuclei in cell unjamming leads to density effects that are not considered in purely shape-induced cell unjamming. The mean area fraction of the nuclei surrounded by the cells' cytoplasm in the histological images range from about 0.25 up to about 0.45 per patient, with an overall mean of 0.35 and a standard deviation of 0.039. This shows that the nuclei are statistically so far apart—spaced by the cytoplasm—that nuclear jamming in the sense of colloidal jamming cannot occur. Since nuclear jamming is not possible, we assume that the nuclei strongly modulate the effective cell stiffness, which has been demonstrated to modulate jamming behavior [11]. With close spacings between the nuclei [high number densities, see Fig. 3(e); small nuclei, see Fig. 3(f)], the cytoplasm becomes strongly compressed between two nuclei, leading to stiffening and thus fostering jamming. This stiffening

may be attributed to strain stiffening in the semiflexible cytoskeletal network [42], a denser network with less water content [20], a stronger cytoskeleton due to an enhanced expression of cytoskeletal proteins caused by mechanotransduction [43], or epithelial-like contractile behavior [44]. Moreover, the strongly deformed nucleus may lead to pronounced feedback mechanisms on the cells' mechanical properties, such as enhanced cell motility and mechanotransduction by YAP-signaling [45,46]. The dependence on the spacing between nuclei may explain previously observed density effects in cancer cell spheroids [9,19]. Increasing the number density also increases the membrane-to-membrane contacts in the clusters, which could lead to more friction per area and consequently higher viscosity fostering jamming. More compressed cells and nuclei at high number densities may also be connected to more nuclear excess envelope, which results in a lack of actomyosin activity needed to generate the yield stress that a motile cell requires [45]. We conclude that the nucleus is a decisive mechanical element for the motion of cells in dense tissues. In the ECM, this is well known, while in cell clusters, this has previously been ignored.

In contrast to our state diagram, simulations that neglected the nucleus were previously able to describe the dynamics in 2D cell systems [26] but not in 3D cell spheroids [9]. In two dimensions, the nucleus has an additional unconstrained degree of freedom, the free space above the cell monolayer, while at the same time, the nucleus-to-cell area ratio is much smaller compared to 3D tissues [13,14,26], which may be why the nucleus is rendered less important in 2D systems. We conclude that either the cell or nucleus shape is more important for unjamming, which is why we have chosen to describe cell shape and nucleus shape in CeNuS. We have fused density- and shape-dependent jamming into one experimental state diagram, considered the nucleus as a previously neglected mechanical element, and approximated a physiologically relevant transition boundary. An important aspect of our state diagram is that it substantiates the need for refined theories to combine shape and density as well as to include the nucleus as an essential mechanical player.

Furthermore, with vital cell tracking, we can verify the existence of jammed and unjammed regions in tumor tissue in agreement with previous reports [9,11,47]. Thus, we confirm that a motility transition exists in tumors. However, we cannot directly observe the transition's dynamics in our experiments. To our knowledge, this is a common problem for all reported phase diagrams from experimental studies [19,31,47]. Nevertheless, the detected differences estimate where the transition zone must lie, which must be between the variables measured for the jammed and unjammed regions. Since it is not possible to induce controlled unjamming transitions, we cannot determine the transition point more precisely. Thus, we can only display a transition region in the state diagram. No direct experimental

evidence exists for a first-order cell unjamming transition [12,35] or for where precisely this transition occurs.

Besides these inherent limitations concerning the transition, we have solved a key sampling problem that has hampered previous efforts to establish a detailed state diagram of cell unjamming. Our unique approach to sample states using the histological slides of 1380 patients probes the entire physiologically relevant and statically accessible state space. Based on our comprehensive histological data, we present a state diagram in Fig. 4(b) that uses CeNuS (shape) and cell area (density) as state variables and thus unifies the suggested mechanisms of shape-dependent jamming [9,14] and density-dependent jamming [9,19,20,31], as well as adding the essential mechanical effects of the nucleus [40,41]. Our state variables summarize the emergent jamming process on a coarse-grained, collective level. Many parameters introduced by previous diagrams and models that categorize the driving effects of unjamming are captured by our two static state variables. For example, the cell shape as part of CeNuS includes the effects of tissue surface tension, cell adhesion, cortical cell contractility, and cell-cell adhesion. At the same time, nucleus number density reflects mechanical interactions between the cell nuclei, which become relevant at close spacing and higher cell rigidities and even capture the effect of external pressures on the cluster. The complete state diagram has an additional dynamical axis, which is often referred to as “effective temperature” [31] caused by proliferation [48–52], inherent cell motility or activity [14–16], and/or other active internal or external stresses. This axis remains inaccessible using static histological slides. Future yield stress measurements may estimate a “critical activity” to unjam and may thereby help determine the transition points on this axis. The patient data-derived statically accessible state diagram shows the heterogeneity of the cell clusters in breast tumors ranging from a more nucleus-number-density behavior to a more shape-dominated jamming behavior. This may explain why invasive lobular carcinoma (ILC), which often shows rounded nuclei and cell shapes [53], is known for its ability to develop distant metastases.

The most significant advantage of using histopathological patient data for our state diagram is that we can conduct a retrospective study to correlate cancer cell unjamming with distant metastasis and thereby derive its importance for tumor progression. Without precise diagnostic guidance, tumor boards must decide between a broad spectrum of therapy options after surgery. This has resulted in a wave of efforts to establish new prognostic molecular tumor markers, such as novel gene signatures. However, therapy decisions that consult gene expression tests do not necessarily change the survival rate [5,54,55]. More information is needed for a better prognosis. However, the existing markers for motile cancer cells are not sensitive to cell motility in the primary tumor before the cells invade the

lymph and vascular systems. Cancer cells in the primary tumor are at the very beginning of the metastatic cascade since they have not left the primary site. The cells at the rim of cancer cell clusters in breast tumors are in contact with the fibrotic ECM, which can be highly attractive to cancer cells as a tumor promoter through mechanotransduction and durotaxis [56–58]. Cancer cells at the cluster rim are not jammed *per se* with respect to single-cell escape into the ECM. Thus, carcinomas may disseminate cancer cells very early on [59]. Cancer cells in the depth of a cluster cannot escape, which makes collective unjamming necessary for this process. In the cancer cell clusters that we find in our histological images, 64% of the cancer cells are outer cells and 36% are bulk cells. Unjamming can thus additionally transform more than 1/3 of the cancer cells into motile cells. However, more importantly, unjamming changes the quality of cancer cell escape. While the outer cancer cells may escape in a process similar to evaporation as single cells [19,47], depending on the ECM density, unjamming of the inner cells permits collective escape of cancer cell aggregates [47,60–63]. Several reports show that cancer cell escape in clusters increases cancer aggressiveness and worsens prognosis [64–66].

Our histopathological data (Fig. 6) show that cell unjamming is critical for tumor progression. However, we cannot precisely measure the dynamic transition in primary tumors because we use static histological images in our retrospective histopathological study. Thus, we use machine learning to find a decision boundary that divides our state diagram into unjammed regions of high distant metastatic risk and jammed regions of low risk via distant metastasis classification. We have consciously chosen distant metastasis as a clinical event since it requires cell motility as a fundamental prerequisite and connects unjamming with the metastatic cascade. Establishing a direct causality between unjamming and the metastatic cascade is unfeasible since the metastatic cascade contains many more complex mechanisms that are not fully understood. Our unjamming criterion adds a 26% information gain with respect to established prognostic risk parameters used in the standard Nottingham index.

Shape and number density are not separate criteria since unjamming can only be described by both, as shown in our tracking experiments [Fig. 4(a)] and our retrospective clinical study [Figs. 6(a) and 6(e)]. In Appendix H and Table I, we show that the cell area \bar{A}_C^{stand} alone has prognostic power. CeNuS adds further prognostic information to \bar{A}_C^{stand} since we find that it provides an 18% information gain with respect to sensitivity and specificity (details can be found in Appendix H) when combining CeNuS with \bar{A}_C^{stand} instead of only using \bar{A}_C^{stand} . This is confirmed using another measure of prognostic power, the c-index, where the combined use of CeNuS with \bar{A}_C^{stand} yields a 20% information gain over using only \bar{A}_C^{stand} (Appendix H).

Since an unjammed state estimated by our state diagram increases the risk for distant metastasis and even adds new information to current prognosis, we conclude that cancer cell unjamming is an important and previously unknown part of the early metastatic cascade in the primary tumor. We observe no time-to-event correlation of our prognosis, meaning that the correlation between unjamming and distant metastasis does not increase when decreasing the follow-up time. It maintains prognostic power even for events that occur a decade later. We attribute this effect to dormant cancer cells, which are nonproliferating cancer cells that have undergone cell cycle arrest and can hide in distant organs for years or decades before being reactivated and causing disease relapse [67].

Regarding the correlation of our state diagram with the overall death, including deaths that are not related to breast cancer, we find that the trained decision boundary, shown in Fig. 6(a), generalizes to the overall death with a significant log-rank $p = 0.031$. For the detailed analysis, we refer to Appendix M. A recent study including over 750,000 breast cancer patients found that about 50% of the deaths reported in overall death are not caused by breast cancer [68]. Thus, our focus on distant metastases provides a stronger connection to tumor progression.

Our retrospective study has been exploratory. A refined clinical study with a predefined design may improve the prognostic power of unjamming. A technical reason for the restraint of the prognostic power of our unjamming criterion might be that the tissue cutouts used in this study—tissue microarray (TMA) cores of 1.0-mm diameter—do not comprise the whole tumor mount, which is typically much larger, in the range of centimeters (compare the descriptive statistics in Table IX in Appendix K). Pathologists have sampled the sites of the TMA cores from the middle of the respective tumors. Therefore, it is plausible to assume that some TMA cores are not a representative sample of the respective tumor’s jammed or unjammed state. Thus, we assume that analyzing whole tumor mounts or the tumor front in addition to the TMA core would lead to better clinical results.

Unjamming improves the prognosis since it adds new information. Concerning the complementarity of unjamming and the nodal status: not all cancer cells that leave the primary tumor target the nearby lymph nodes. There are other pathways for distant metastasis that bypass the lymph nodes [3]. This agrees with reports that lymph node status has limitations in predicting distant metastasis [4,69], including a large-scale breast cancer study ($N > 24\,000$) [3]. Because of the complementarity to the nodal status, our unjamming criterion can correct for currently made errors in the prognosis and thereby has the potential to reduce overtreatment by identifying a low-risk group among lymph-node-positive patients [see Figs. 6(e) and 6(f)]. Vice versa, the information on the lymph node status also corrects for prognostic mistakes made by unjamming.

With unjamming, we have found a potentially highly clinically relevant histopathological tumor marker, which fills the gap in detecting cancer cell motility in the primary tumor. Unjamming as a prognostic marker provides a new type of clinically relevant data that may add significant information about the distant metastatic risk to existing clinical prognostic indices. We need as many complementary data as possible to obtain an improved picture of the disease and arrive at personalized predictions. We hope our results will gain enough attention so that clinical studies will further determine the prognostic value of unjamming. While vital cancer cell tracking is not clinically feasible, our unjamming state diagram, already parametrized with 2D variables, as a prognostic criterion is well suited for clinical applications because it only requires H&E-stained slides, which are readily available in a clinical routine. For decades, histological sections have been used as a standard procedure after tumor surgery to evaluate the tumor. Thus, our morphodynamic link closes an essential histopathological gap. In the far future, our improved understanding of the metastatic cascade may help identify drug targets for adjuvant migrastatic anticancer therapies, which aim to inhibit cancer cell motility, e.g., by targeting actin polymerization or cell contractility [70]. However, specifically targeting the cytoskeleton of cancer cells is challenging since the cytoskeleton is highly conserved in healthy and malignant cells. Consequently, there exists a severe toxicity problem [70]. Therefore, inhibiting collective unjamming might be one solution to this problem.

Increased cell motility through cancer cell softening and mechanical changes in tumor tissue has been implicated as necessary in tumor progression for two decades [71,72]. While tissue and cell mechanics have become a quintessential part of the physics of cancer and have led to new physics [9,11,26], it is still unclear whether these results are of clinical relevance. Cancer cell unjamming demonstrates that collective emergent effects are needed in the early metastatic cascade, and they provide complementary information for cancer prognosis that cannot be provided by molecular approaches or an established diagnosis.

V. MATERIALS AND METHODS

Vital primary sample preparation.—We obtained primary samples of cervical and breast carcinomas immediately after surgery. Patients were asked to participate through approved ethics votes (cervix ethics vote No. 090-10-19042010, breast ethics vote No. 073-13-11032013). The samples were classified as tumor tissue by pathologists at Leipzig University Hospital.

Samples were transported in a buffer solution based on Ringer’s lactate solution. The basis for the modified Ringer tissue buffer is the Ringer’s lactate solution (B. Braun Medical AG, Cat. No. 3325950, Approval No. 6724011.00.00) supplemented with 5% [w/v] glucose monohydrate and

1% antibiotic–antimycotic solution (PAA, Cat. No. P11-002) to prevent possible contamination. To find cancer cell clusters amidst the stroma and adipose tissue, samples were cut into approximately half-millimeter-size pieces and stored in Dulbecco’s Modified Eagle’s Medium (DMEM). These pieces were then analyzed under a microscope to identify cancerous cell clusters. Cancer cell clusters were separated from the surrounding tissue and subsequently transferred to a 384-well plate with a microscopy bottom (Cat. No. 88416, Ibidi) filled with DMEM and 0.2 μM of SiR-DNA (Cat. No. SC007, Spirochrome) using a 200- μl pipette tip and microscopic vision. Multiple pieces of the same sample were put in one well and independently imaged since they mostly stayed in their positions.

Live imaging of primary samples.—The procedure is similar to the one used for vital cell observations of explanted tumors of mice [22].

We used water immersion objectives to achieve a high NA with Zeiss Immersol W 2010 (Cat. No. 444969-0000-000). The pieces should not be too far from each other in order to keep the immersion for long-term observations. We typically used 2×2 wells of the 384-well plate for imaging.

For vital nuclei staining, we use SiR-DNA, which is minimally toxic (for concentrations less than 0.25 μM [73]; we use 0.2 μM). The excitation wavelength is around 652 nm. In this far-red illumination regime, phototoxic damage to the cells is minimal. The samples were kept in medium containing 0.2 μM SiR-DNA during observation to ensure a stable fluorescence signal.

We used the ZEISS Axio Observer equipped with a Yokogawa CSU-X1A 5000 spinning disk confocal scan unit for vital imaging. Fluorescence signals were recorded every 5 or 7.5 minutes for an observation period of 6 hours. The ZEISS Live Cell Imaging (LCI) Plan-Neofluar 25x/0.8 objective (Cat. No. 420851-9972-000) was used to provide an appropriate field of view.

In order to account for occasionally occurring whole-sample drift and rotation, registration of the 4D stacks (t, x, y, z) was performed with a MATLAB algorithm explained in the Appendix of Grosser *et al.* [9].

To dynamically study the primary tumors’ trajectories of the cancer cells, 3D live tracking of the nuclei was performed using the open-source image analysis software TrackMate [74]. Here, we used a spot size of 10 μm and a threshold of 0.5. The detected spots were linked from frame to frame using the TrackMate LAP Tracker, which penalizes interspot differences in median spot intensity, standard deviation of intensity, and estimated spot diameter. This ensures accurate cell tracking, even when the cells pass each other closely, and prevents tracks from flipping between cells. The maximum allowed linking distance was adjusted for every primary sample such that the maximally allowed step size was slightly larger than the corresponding maximal frame-to-frame velocities. The quality of the resulting tracks was verified by

visual inspection. Tracks shorter than ten frames were excluded from the analysis.

Using nuclei tracking to quantify the cell displacements, we calculate cage-relative (relative to the nearest neighbors) mean-squared displacements (MSD) over time. Cage-relative calculations enable us to measure *single-cell* motility not collective cell motion. From the cage-relative MSDs, we calculate the mean distance traveled per cell at a specific time t as $\sqrt{\text{MSD}(t)}$.

Nucleus shape quantification in primary samples.—To measure the static aspect ratios of the nuclei, nuclei segmentation of static images for $z = \text{const}$ and $t = \text{const}$ was performed with the open-source image analysis library StarDist in PYTHON—which utilizes the predominant star convex shapes of nuclei together with CNN learning [27]—using the pretrained model for fluorescence images. Falsely segmented nuclei were corrected by hand. The aspect ratios were calculated as $\text{AR} = a/b$, where a is the major axis length and b the minor axis length of the ellipse with the same normalized second central moments as the segmented region.

Combined nucleus and cell shape imaging in primary samples.—After live imaging, some tumor explants were fixated with 4% [w/v] paraformaldehyde in PBS for 30 min. Next, fixated tissue was washed with PBS and stored in 1% solution of Triton-X 100 for 24 h or longer, resulting in higher transparency and a better intake of the fluorophore. Next, tumor pieces were stained in Alexa-Fluor 488 Phalloidin actin staining for 24 h and then stored in IMM mounting medium (ibidi) to reduce the fluorophores’ photobleaching to increase visibility by optical clearing. Imaging was carried out using a Leica TCS SP2 laser scanning microscope with a Leica HC PL APO 20X/0.7 CS Corr IMM objective. The refractive index of the mounting medium was measured with a refractometer to $n = 1.445$. To avoid any refractive index mismatch between mounting medium and immersion medium, an 80%/20% glycerol/water mixture ($n = 1.45$) was used as immersion medium.

Patients for retrospective study.—For our retrospective histopathological study, we investigated 1380 breast cancer cases in total. The nucleus and cell shape analyses were based on H&E-stained digitized tissue sections of the corresponding primary tumors. Patients were informed and gave ethical, informed consent, approved by the ethics committee of the Ärztekammer Hamburg (Medical Council Hamburg) with Processing No. PV2946. The considered cases were classified according to World Health Organization histological classification of breast carcinomas, 3rd edition of 2003, which follows the TNM classification according to Union Internationale Contre le Cancer in terms of grading and staging. All data used for this study are from female patients. The clinic evaluated the sex of the patients by specification. All available information can be extracted from the descriptive statistics in Table IX in Appendix K.

H&E-staining and imaging of breast cancer tissue.—H&E staining of patient breast cancer tissue was performed following standard protocols using the Leica ST4040 automat. The formalin-fixated breast cancer tissue sections were then optically inspected and digitized using the Panoramic 1000 Flash IV Scanner of 3DHISTECH with a Zeiss Plan Aplanachromat 40x objective.

Digital image acquisition and analysis.—The raw data of 18 TMAs—each containing about 100 histological images of breast cancer—are available as MIRAX files. We opened each file with QuPath [75] and exported the single TMA cores, which were annotated semiautomatically as TIFF files. Each core, corresponding to one patient, has a diameter of approximately 1200 μm and a μm -to-pixel ratio of 0.249.

We filtered out patients as follows: (1) those with H&E-stained images with less than 200 cancer cells [that is, the total number of nuclei found in the regions that our algorithms labeled as cancer cell clusters (explained below)]; (2) those for which no clear distinction between cancerous and other areas could be made (e.g., no clear separation between breast cancer clusters and intermeshing clusters of cancer-associated fibroblasts was possible); and (3) those whose tissue had slipped out of the focus of the scanner. As orienting information, it is helpful to know that for all 1380 remaining cases, the mean (\pm SD) cancer cell number per patient amounts to 3026 (\pm 1803). To analyze the shapes of cancer nuclei and cells, we use nuclei segmentation of the StarDist algorithm with the pre-trained model of the developers for H&E-stained images [27]. In StarDist, every nucleus segmentation is assigned a probability of its accuracy. We use the suggested 0.69 probability threshold for robust results. We have also tried lower thresholds, as some nuclei are not found at the 0.69 threshold. However, nuclei were segmented incorrectly, and the shapes of the nuclei were often wrong, which is why we decided to use the 0.69 threshold recommended by the developers. Furthermore, distinct nuclei from different z planes can optically overlap in stained tissue sections. The software can account for this to a certain degree by an overlap estimation. We allow an overlap of 30% of the nucleus area.

Tissue segmentation.—To quantify only the shapes of nuclei and cells that are cancerous, we have developed a custom MATLAB algorithm, called tissue segmentation (TS), which distinguishes cancerous tissue from stroma, immune cells, and adipose tissue. TS uses a copy of the mask of segmented nuclei M , that is, 1 for nuclei and 0 otherwise. First, the nuclei of potential immune cells are coarsely filtered by identifying nuclei with a radius smaller than 5 μm and an aspect ratio smaller than 1.2, and they are densely packed compared to surrounding nuclei. Accordingly, only immune cells that exist in joint clusters are filtered. Immune cells that are, for example, isolated in the midst of cancerous tissue are not filtered. However, the

statistics of the nuclear and cellular shapes of the cancerous tissue are hardly touched by these single immune cells since their occurrence is rare. Potential nuclei of fibroblasts are also coarsely filtered out of M by identifying nuclei that exhibit $AR > 2.1$. The remaining nuclei on the mask M are morphologically dilated, resulting in contiguous areas. Contiguous territories with an area that is smaller than $(11 \mu\text{m})^2 \cdot \pi$ are typically groups of fibroblasts and are deleted from M . Subsequently, M is smoothed by thresholding the convolution of the binary matrix M with a $5 \mu\text{m} \times 5 \mu\text{m}$ mask with all entries equal to $1/(5 \mu\text{m})^2$. The resulting mask marks the cancerous regions. For measurements, nuclei filtered in M but lying in the cancer mask are considered since they are only deleted for morphological operations.

Cell and nucleus shape quantification.—The StarDist algorithm provides us with a binary mask labeling nuclei. The TS algorithm results in a mask that identifies the cancerous regions. Combining these two masks, we approximate the cell boundaries in H&E images using the marker-based watershed algorithm [76]. The quality of the approximation is quantified in Appendix C. We calculate the so-called cell shape index using these cell boundaries, where P is the perimeter of the approximated cell boundary and A is the area. Note that P is computed using the MATLAB built-in function *regionprops*, which calculates the distance between each adjoining pair of pixels around the region's border. Thus, small values of P correspond to roundish cells while large values correspond to elongated cells. We quantify nucleus shapes using the aspect ratio of the ellipse's major and minor axis lengths, which have the same normalized second central moments as the nuclei segments. We use aspect ratios for nuclei shape quantification as most nuclei are convex and the aspect ratio is a robust measure for these shapes. We use the cell shape indices for cell shape quantification as cells are often nonconvex and aspect ratios might not capture the true shape of a cell. It is also a common measure for cell shape quantification [14,15].

For the CeNuS and mean cell areas A_C , standardization of variables is performed with respect to the means and standard deviations of a cohort of 530 cases, which the Institute of Pathology Hamburg-West consecutively received: $(X - \bar{X})/\sigma_X$, where X denotes, for example, the ensemble median of nucleus aspect ratios or cell shape indices. The exact values for standardization can be found in Table VIII. Cell and nucleus shapes corresponding to cell segments smaller than $(1 \mu\text{m})^2 \cdot \pi$ or larger than $(28 \mu\text{m})^2 \cdot \pi$ being filtered since they often belong to falsely approximated cell outlines.

Statistical analysis

Two-sample Kolmogorov-Smirnov (2sKS) test.—The 2sKS test is a nonparametric hypothesis test that is commonly used to test if two distributions are statistically different. The test assesses the difference between the

cumulative distribution functions of the distributions of the two samples one wants to compare. The test statistic is defined as $D^* = \max_x \{F_1(x) - F_2(x)\}$, where $F_1(x)$ [$F_2(x)$] is the proportion of x_1 (x_2) values less than or equal to x . We use the MATLAB built-in function *kstest2* to compare distributions.

Kaplan-Meier estimator.—This estimator is a nonparametric statistic used to define survival probabilities from lifetime data. Specifically, it estimates the probability of an individual or group *not* exhibiting a certain event in a time interval. Throughout this study, the event is distant metastasis. The estimator is defined as $\hat{S}(t) = \prod_{i: t_i \leq t} [1 - (d_i/a_i)]$, where t_i are time steps where at least one event happened, d_i is the number of events at t_i , and n_i is the number of individuals known to have no event until time t_i .

Log-rank test.—The log-rank (sometimes referred to as the Mantel-Cox test) test is a nonparametric hypothesis test with a null hypothesis that two groups are sampled from the same population. This test is commonly used to compare risk groups from the Kaplan-Meier estimator and general time-to-event measurements. The test statistic is defined as $\chi^2(\text{logrank}) = \sum_{i=1}^m [(O_i - E_i)^2/E_i]$, where m is the number of groups, O_i the total number of observed events, and E_i the total number of expected events.

Fisher's exact test.—Fisher's exact test is a nonparametric test for testing the null hypothesis that no nonrandom associations exist between two categorical variables. The alternative hypothesis is that there is a nonrandom association between the variables. Especially when few samples are available, it is advised to use this test instead of the also commonly used chi-squared test. We use the test to compare clinical features between patient groups in Appendix K. The MATLAB built-in function *fishertest* is used.

Pearson correlation coefficient (PCC).—PCC is calculated by the covariance of two data arrays divided by the product of their standard deviations. It is commonly used to assess the strength and significance with which data are correlated. A coefficient of 0 equals a nonexistent correlation, and a coefficient of 1 means a perfect correlation. We use the MATLAB built-in function *corrcoef*.

Median absolute deviation (MAD).—We calculate the median absolute deviation by taking the median of the absolute value of the differences of the data from its median, that is, $\text{median}(\|x - \text{median}(x)\|)$, where x is a data vector.

Cox proportional hazard model.—The Cox model defines the hazard for an individual i as $h_i(t) = h_0(t) \cdot e^{X_i(t) \cdot \beta}$, where h_0 is the baseline hazard, X_i is the vector of covariate values for individual i , and β is the coefficient vector. The model is widely used to assess the degree and significance of certain influences X_i on survival or other time-dependent events. When multiple influences X_i are considered, it can be checked if one or multiple influences are independent of other influences regarding hazard ratios. It

should be noted that the model only considers weighted linear sums of the influences on an event. Here, we compare our jamming threshold in the state diagram with conventional clinical prognostic factors such as the tumor grade or tumor size. For Cox analyses, we use the MATLAB built-in function *cophfit*.

Sensitivity and specificity.—In binary classification, the sensitivity is defined as the true positive rate $tp/tp + fn$, where tp is the number of true positives and fn the number of false negatives. In our study, true positives are patients who developed distant metastases, indicating an unjammed tumor structure and false negative patients would be patients who do not develop distant metastases but are associated with the high-risk group.

The specificity is defined as the true negative rate $tn/tn + fp$, where tn is the number of true negatives and fp is the number of false positives. In our study, true negative patients are those who develop no distant metastasis and whose tumor structure indicates a jammed state. False-positive patients are those whose tumor structure indicates a jammed state but who develop distant metastases. Typically for cancer markers, there is a trade-off between sensitivity and specificity.

PPV and NPV.—In binary classification, the PPV is defined as $tp/tp + fp$ and is, therefore, the true positive rate within the group classified as positives. In our study, this would be the distant metastasis rate within the group of patients located above the jamming threshold in Fig. 6. The NPV is defined as $tn/tn + fn$ and corresponds to the number of patients who did not develop distant metastases divided by the number of patients in the low-risk group (below the jamming threshold).

c-index.—The concordance index (c-index, Harrell's c-index) [28] is a measure of prognostic power. It measures the ability of a score to separate patients with favorable and poor outcomes. A c-index of 0.5 corresponds to a completely random score/model, a c-index of 1 to a perfectly correlated score/model, and 0 to a perfectly anticorrelated score/model. One can view the c-index as the percentage of patient pairs that can be correctly ordered with respect to their survival (event) times [28]. In this study, we use a self-written implementation in MATLAB based on the work of Harrell and colleagues [28].

Bootstrap resampling.—Bootstrap resampling is a statistical technique used to estimate the uncertainty associated with a statistic by resampling from the original data set. This technique involves creating a large number of new samples (called bootstrap samples, in this study $N = 1000$) by randomly selecting observations from the original data set with replacement. By generating a large number of bootstrap samples, which have the same size as the original data set, one can estimate the variability of the statistic of interest and construct confidence intervals.

Variance inflation factors (VIFs).—VIFs are a measure of the degree of multicollinearity among predictor variables

in multivariate regression analysis such as multivariable Cox models. VIFs can be used to identify highly correlated predictor variables that may be redundant or provide little additional information in explaining the variation in the dependent variable. VIFs can be defined by the diagonal elements of the inverse of the correlation matrix of the predictor variables [77]. A predictor variable with a VIF of 1 indicates no collinearity. VIFs smaller than 2.5 are conservatively associated with a nonproblematic amount of collinearity regarding multivariate regression analyses.

ACKNOWLEDGMENTS

We thank Denise Helmer and Ole Schamuhn for their technical support. We thank Hanno Krümpelmann, Sebastian Henn, Carla Thomaßen, and Frank Sauer for proofreading. All authors have contributed significantly to this work. P. G. and D. T. performed formal analysis and curated and synthesized the data. P. G. and J. L. contributed the methodology. P. G., J. L., and S. G. developed imaging and analysis of primary samples. P. G., J. L., S. G., H. K., and J. A. K. performed the original draft preparation. F. R. and S. G. treated and imaged vital primary samples. P. G. and H. K. contributed to the visualization of the project. J. L. and M. S. performed vital nucleus tracking. B. W., M. H., and B. A. performed surgery and statistical analysis. C. E. and M. L. contributed to the statistical analyses. A-S. W., A-K. P., B. U., K. F., C. L., and A. N. performed conceptualization, supervision, and project administration of the compilation and formation of the tissue microarrays. A. N. and J. A. K. supervised and administered the entire project and guided the interpretation of the data. All authors provided critical review, commentary, revision, and editing for the final version of the manuscript. The project received funding from the European Research Council (ERC 741350/ HoldCancerBack).

APPENDIX A: COMBINING LYMPH NODE STATUS AND UNJAMMING CRITERIA

This section discusses to what extent our prognostic evaluation can be combined with the lymph node status.

In Fig. 7, the decision boundary presented in the main text (Fig. 6) is applied to only nodal positive cases from the training set, and the resulting stratification is shown in Kaplan-Meier curves. No nodal-positive patient develops metastasis if the patient is below the jamming threshold, while 5/5 metastases are above the threshold in the more unjammed section of the state diagram. The results for the test set can be found in the main text in Figs. 6(e) and 6(f).

When the decision boundary presented in the main text [compare Fig. 6(a)] is applied only to the nodal negative patients within the training and test sets defined in the main text, we find significant stratification within the training set (log-rank $p = 0.018$). Above the unjamming threshold, we

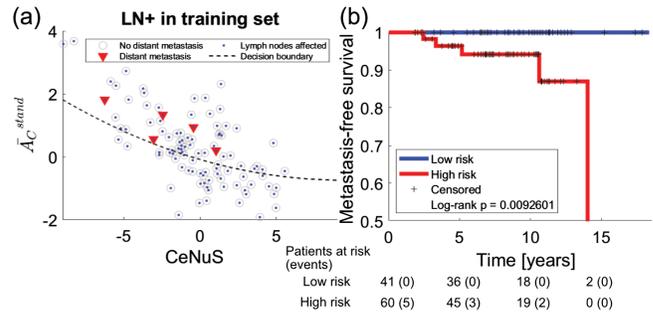


FIG. 7. Stratification of all lymph-node-positive cases in the training set by the decision boundary in the space of CeNuS and \bar{A}_C^{stand} presented in the main text. The y axis shows the probability of the Kaplan-Meier estimator of developing distant metastasis over time. Log-rank $p = 0.01$. “Patients at risk” indicates the total number of patients who have not been censored or have not developed distant metastasis by a specific follow-up time. In parentheses is the cumulative number of individuals who will develop distant metastases at a later time.

find $7/235 = 2.98\%$ (training $16/265 = 6.03\%$) nodal-negative patients with distant metastases (data not shown). Below the jamming threshold, we find that $3/205 = 1.46\%$ (training $9/322 = 2.80\%$) of nodal-negative patients

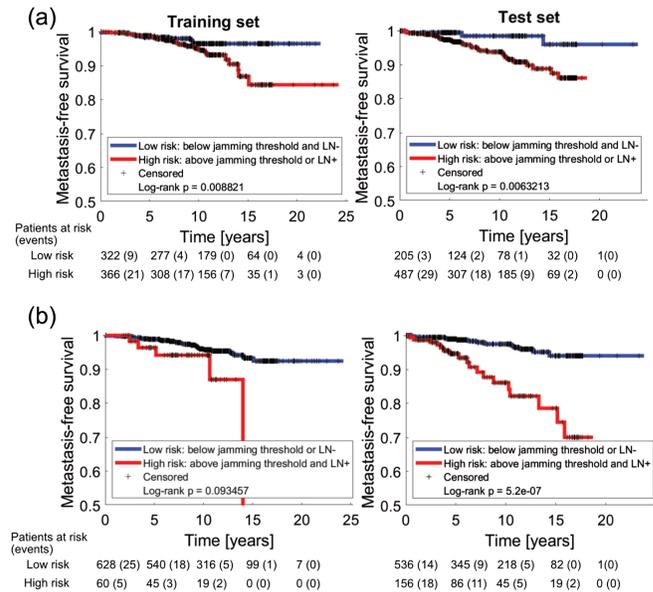


FIG. 8. Combining LN status with the prognostic unjamming criteria. (a) Training and test sets: The low-risk group consists of the patients whose tumor structure indicates a jammed state and who exhibit no affected lymph nodes (LN-). The high-risk group consists of all other patients. (b) Training and test sets: The high-risk group consists of patients whose tumor structure indicates an unjammed state and who exhibit affected lymph nodes (LN+). The low-risk group consists of all other patients. “Patients at risk” indicates the total number of patients who have not been censored or have not developed distant metastasis by a specific follow-up time. In parentheses is the cumulative number of individuals who will develop distant metastases at a later time.

develop distant metastasis. These numbers (summarized in Table I) are promising to establish a high-risk group among nodal-negative patients. However, the risk stratification over time only shows significance in the training set and not in the test set. Stratification in the test set is not significant (probably due to a few events within this collective: $N = 10$), with a log-rank $p = 0.38$ and a distant metastasis rate of 1.46% in the low-risk group and 2.98% in the high-risk group.

In Fig. 8(a), we combine the information on jamming and lymph nodes in order to achieve a very reliable low-risk group, where the tumor structure indicates a more jammed state *and* no lymph nodes were affected; this yields significantly (log-rank $p = 0.006$) separated risk groups in the test set, with a distant metastasis rate in the low-risk group of $3/205 = 1.46\%$ and in the high-risk group $29/487 = 5.96\%$, such that the risk of developing distant metastases is 4 times higher in the high-risk group compared to the low-risk group.

In order to identify a reliable high-risk group, we further combine the unjamming criterion and the lymph node status such that the high-risk group is composed of patients whose tumor structure indicates an unjammed state *and* at least one lymph node is affected; see Fig. 8(b). We see significantly (log-rank $p < 0.001$) separated risk groups in the test set with a distant metastases rate of $14/536 = 2.61\%$ in the low-risk group and $18/156 = 11.54\%$ in the high-risk group.

APPENDIX B: COX PROPORTIONAL HAZARD MODELS

The univariable (Table V) Cox models of classical risk parameters for breast cancer and the presented jamming threshold are presented here while the discussion of the multivariable (Table III) Cox model is discussed in the main text. Patients without distant metastases are censored. In Table V, the univariable Cox models are shown for the variables' jamming threshold (1 if below decision boundaries presented in the main text, 0 otherwise); the largest tumor diameter in cm; the histological grade; the lymph node status (1 if one or

TABLE V. Univariable Cox proportional hazard models for the jamming threshold, the largest tumor diameter d_t , the grade, the lymph node status, the NPI, and chemotherapy.

| | Coefficient b | SE(b) | exp(b) | p |
|-------------------|-----------------|-----------|------------|----------------|
| Jamming threshold | -1.09 | 0.43 | 0.34 | 0.010 |
| Tumor diameter | 0.39 | 0.11 | 1.47 | $\sim 10^{-4}$ |
| Grade | 0.53 | 0.28 | 1.70 | 0.060 |
| Lymph node status | 1.42 | 0.38 | 4.13 | $\sim 10^{-4}$ |
| NPI | 0.52 | 0.17 | 1.69 | 0.002 |
| Chemotherapy | 1.78 | 0.46 | 5.93 | $\sim 10^{-4}$ |

TABLE VI. Multivariable Cox proportional hazard model for standard prognostic parameters: largest tumor diameter d_t , tumor grade, and lymph node status. The likelihood ratio test is $p < 0.0001$. All variance inflation factors are smaller than 1.6, indicating vanishing collinearity. The concordance index equals 0.6684.

| | Coefficient b | SE(b) | exp(b) | p |
|-------------------|-----------------|-----------|------------|-------|
| Tumor diameter | 0.23 | 0.14 | 1.23 | 0.11 |
| Grade | -0.06 | 0.28 | 0.77 | 0.83 |
| Lymph node status | 1.21 | 0.43 | 2.16 | 0.005 |

TABLE VII. Multivariable Cox proportional hazard model for standard clinical parameters: largest tumor diameter d_t , tumor grade, lymph node status, and chemotherapy. The likelihood ratio test is $p < 0.0001$. All variance inflation factors are smaller than 1.7, indicating vanishing collinearity. The concordance index equals 0.7497.

| | Coefficient b | SE(b) | exp(b) | p |
|-------------------|-----------------|-----------|------------|--------|
| Tumor diameter | 0.21 | 0.16 | 1.23 | 0.1750 |
| Grade | -0.26 | 0.29 | 0.77 | 0.3733 |
| Lymph node status | 0.77 | 0.43 | 2.16 | 0.0724 |
| Chemotherapy | 1.57 | 0.50 | 4.82 | 0.0015 |

more lymph nodes are invaded, 0 otherwise); the Nottingham prognostic index (NPI), which was calculated as $0.2 \cdot d_t + N + G$, where N is equal to 1 if no lymph nodes are invaded, 2 if 1–3 nodes are invaded, and 3 otherwise; and the chemotherapy status (1 if the patient was treated with cytotoxic chemotherapy, 0 otherwise). Considered independently, only the tumor grade is insignificant, while chemotherapy status, NPI, lymph node status, tumor diameter, and jamming threshold are significant.

As expected, the hazard for patients whose tumor structure indicates a more jammed state (below threshold in Fig. 6) decreases to 34%. Regarding the other significant parameters, the risk of developing distant metastasis increases with increasing tumor diameter, with affected lymph nodes, with a high NPI and chemotherapy indicated.

In Table VII, the multivariable Cox model is shown for the variables' tumor diameter, tumor grade, lymph node status, and chemotherapy. In Table VI, we present the multivariable Cox model for the variables' tumor diameter, tumor grade, and lymph node status. We present these models without incorporating the jamming information in order to compare these models with the ones including the information on the jamming threshold presented in the main text in Tables III and II. Therefore, comparing the models, one can infer the information gain based on our state diagram, which is discussed in the main text.

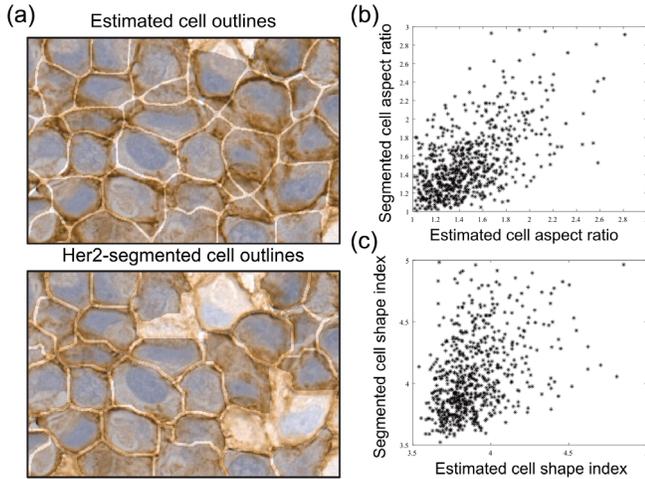


FIG. 9. Comparison of the HER2-segmented cell outlines with the watershed-estimated cell outlines. (a) Exemplary image of the approximated and segmented cell segments (dark masks). (b) Direct correlation of the aspect ratios of the segmented and estimated cell segments ($n = 712$). The Pearson correlation coefficient is 0.59. (d) Direct correlation of the cell shape indices of the segmented and estimated cell segments ($n = 712$). The Pearson correlation coefficient is 0.42.

APPENDIX C: CELL OUTLINE APPROXIMATION IN HISTOLOGICAL SLIDES

We estimate cell outlines by watershedding the segmented nuclei signal and the cancer cluster’s boundaries. To find out how well this approximation matches actual cell shapes, we compare the approximated outlines with segmented outlines of one whole-slide digital image of HER2-stained breast cancer tissue. The HER2 receptor located at cell membranes can be used to find the actual cell outlines. The first step of the segmentation is a color-deconvolution that performs a basis transformation from RGB space to the space of HER2, the blue nucleus signal, and a residuum channel. The HER2-channel and the nucleus channel are further manipulated to enhance proper signals and delete noise. Next, the nucleus channel is used as input to the StarDist algorithm referenced in the main text, which segments the nuclei with the pretrained model for fluorescent nuclei. Finally, the cell outlines are segmented using the watershed algorithm on the gradient image of the HER2 channel with regional minima where segmented nuclei are. We further filter out segments with areas smaller than $(3^2 \cdot \pi) \mu\text{m}^2$ or larger than $(11^2 \cdot \pi) \mu\text{m}^2$, with a weak HER2 signal along the boundary and with too much HER2 signal within the segmentation. Thus, we arrive at a set of segmented cell outlines based on the HER2 segmentation and a set of outlines approximated by the watershed algorithm using the positions of the nuclei and cancer boundaries.

An illustration of the HER2-segmented and watershed-estimated cell outlines can be found in Fig. 9(a), presented

below. The segmented and watershed-estimated cells count as synchronized when their areas overlap by more than 80%. There is a good correlation between segmented and estimated cell aspect ratios, as seen in Fig. 9(b) (Pearson correlation coefficient 0.59). In addition, a moderate correlation is found between the cell shape indices of the estimated and segmented cells (Pearson correlation coefficient 0.42). The estimated cell shape indices statistically underestimate the segmented cell shape indices.

APPENDIX D: DYNAMICS AND STRUCTURAL OBSERVABLES IN VITAL PRIMARY CANCER SAMPLES

We estimate the correlation of the ability of cells to move inside the dense three-dimensional tumor environment on the nucleus ellipsoid shape and nucleus volume to check whether the connection we report in Fig. 6 for the ensemble average is valid on the level of individual cells. This analysis is performed using tissue samples from an exemplary breast tumor with Internal ID No. 106; see Appendix L.

The nucleus ellipsoid shape is defined as $s_E = S \cdot V^{-2/3}$, where S is the surface of the ellipsoid that has the same second moments as the pixel segment and V the respective volume. In 3D, we use this robust measure of cell elongation instead of the aspect ratio because a 3D aspect ratio is not well defined without information loss since there exist three major axes. The nucleus volume is of interest because it is inversely correlated with the number density of the cell packing [9,19]. We track the nuclei as described in Sec. V [see Fig. 10(a)] and estimated the

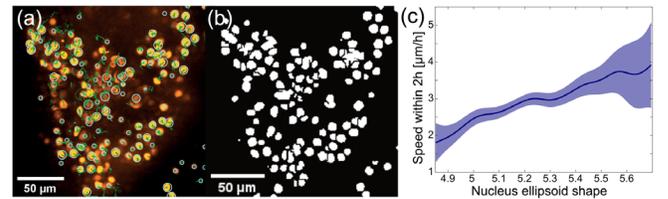


FIG. 10. Correlation of individual cell motility with its nucleus shape. (a) Visualization of 3D tracking in a breast cancer explant. Stained nuclei are depicted in orange, spots found are depicted in light blue, and the tracks are indicated in green. (b) Structural properties of nuclei estimated with adaptive thresholding followed by watershedding of the thresholded signal around the tracked spots, allowing segmentation of nearby nuclei. The black and white image of the thresholded nuclei contains multiple errors that are filtered in a later step. (c) Mean cage-relative speed of cells within 2-hour lag correlated with the nucleus shape (average over each cell). The nucleus shape is estimated as the shape parameter $s_E = A/V^{2/3}$ of the ellipsoid with the same second moments as the nucleus segment. The colored regions indicate the confidence interval of the plots. The data are averaged using a moving average with a window width of 0.1.

nucleus shape with adaptive thresholding followed by morphological opening and a watershedding around the detected tracks, allowing segmentation of very close nuclei. The nuclei shapes contain errors, as shown in Fig. 10(b), filtered out using track statistics. The tracks are filtered in order to minimize the impact of segmentation errors. The first 50 frames are excluded because the nucleus stains are not yet fully drawn in, reducing the tracking quality. Furthermore, tracks with high variances of nuclei volumes, high median absolute deviation of nuclei volumes, and high median differences of nuclei volumes are excluded as these are strong indicators of fluctuating errors in nuclei segmentation. After filtering, we present our results in Figs. 4(a), 3(e), and 10 using about 2600 tracks. Visual inspection of the tumor explants shows that cells that can move stay motile during the experiment and arrested cells rarely move. Therefore, we average the studied quantities for each cell, using the median values. Since motile cells have higher absolute values of cage-relative motion, we use the mean cage-relative speed within a lag of 2 hours as a motility measure. Concretely, we calculate the absolute displacements within 2 hours of the nuclei and their nearest neighbors. We subtract the average displacements of the neighboring nuclei and average the absolute value of the resulting cage-relative displacement within 2 hours. The speed estimation is calculated by dividing by the time frame of 2 hours.

Figures 10, 3(e), and 4(e) show that the shape and volume of cell nuclei in densely packed tumor explants have predictive power on the ability of the respective cells to move inside the tumor. As is visible in Fig. 3(e), low nuclei volumes correlate with low cell motility, and high nuclei volumes correlate with high cell motility. Additionally, round nuclei correlate with low cell motility, and elongated nuclei correlate with high cell motility, as seen in Fig. 10(c). This underlines the importance of both of these static features regarding the estimation of cell motility.

APPENDIX E: STANDARDIZATION VALUES

TABLE VIII. Standardization values for measures presented in the main text. Means and standard deviations (SD) of a cohort of 530 patients for the standardization of cell shape (CS) indices, nucleus AR, and cell areas in μm^2 .

| | Cohort mean | Cohort SD |
|---------------------|-------------|-----------|
| CS indices median | 3.9273 | 0.0270 |
| CS indices variance | 0.0559 | 0.0072 |
| Nucleus AR median | 1.4659 | 0.0645 |
| Nucleus AR variance | 0.1422 | 0.0251 |
| Cell area mean | 79.781 | 14.477 |

APPENDIX F: QUALITY CONTROL OF TISSUE SEGMENTATION

Qualitatively, the results of the TS algorithm were independently verified by pathologists from the Hamburg-West Pathology Department. Quantitatively, we annotated the H&E images of 43 patients. For this purpose, we annotated five areas of representative cancer cell clusters and three representative areas in which there are no cancer cells for each H&E image corresponding to one patient. Annotations were independently validated as being correct by pathologists at Hamburg-West Pathology. In the next step, we compared the regions annotated by hand with the regions of the TS algorithm. Here, we compared the masks that the algorithm found with the annotated ones, which is illustrated for one case in Fig. 11. For cancer (noncancer) regions, we calculate, for each annotated cancer (noncancer) cluster, the fraction of the sum of pixels that the algorithm found correctly in that region divided by the actual pixel number of that region. The median [\pm median absolute deviation (MAD)] ratio of the cancer regions found by the algorithm and the annotated cancer regions is 0.95 ± 0.04 . The corresponding median (\pm MAD) for regions with no cancer is 0.97 ± 0.03 .

APPENDIX G: SHAPE DISTRIBUTIONS

The aspect ratio distributions of the nuclei of the patients are typically highly right skewed, with a skewness (median \pm median absolute deviation) of 1.45 ± 0.47 . On the other hand, the distributions of the cell shape indices of the patients are most often highly right skewed, with a skewness of 1.35 ± 0.19 .

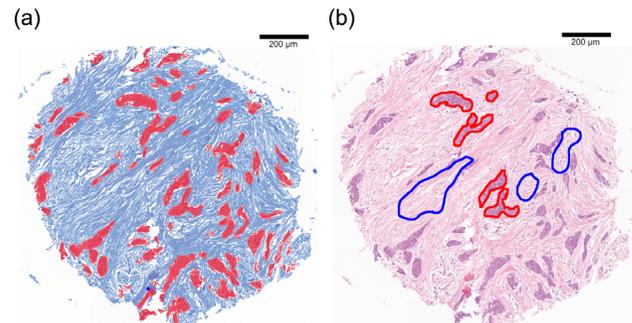


FIG. 11. Illustration of the comparison of the annotated tissue type and the tissue type found by the TS algorithm. (a) Red: cancerous regions found by the TS algorithm; white: background; blue: stroma found by the TS algorithm. (b) Annotations for benchmarking the TS algorithm. Red annotated regions represent cancerous regions while blue annotated regions represent regions with no cancer cells.

APPENDIX H: PROGNOSTIC INFORMATION OF STATE DIAGRAM AXES

For an analysis of the prognostic value, we included the specificity, sensitivity, positive predictive value, negative predictive value, and concordance indices based on CeNuS, cell area (inverse number density), and the decision boundary presented in the main text in Table I. These measures of prognostic relevance were estimated in the test set using the thresholds for the single variables as described in the following procedure, where we also estimated the prognostic information gain using cell area *and* CeNuS together, as in the main text (compare the decision boundary in the state space in Fig. 6) compared to only the cell area.

For the single variables \bar{A}_C^{stand} and CeNuS, we find the threshold that maximizes the separation between the two classes (distant metastasis, no distant metastasis) within the training set. Therefore, we calculate the receiver operating characteristic (ROC) curve for the complete threshold spectrum of the variables and measure the distance for each point on the ROC curve to the nearest point on the line $y = x$ (random model). The threshold with the best separation corresponds to the point on the ROC curve that exhibits the largest distance to a random model. This maximization yields the optimal threshold ($\bar{A}_C^{\text{stand}} = -0.0201$; CeNuS = -1.6288), which we continue to probe in the test set.

Regarding the threshold for \bar{A}_C^{stand} in the test set, we obtain a specificity of 0.4727 and a sensitivity of 0.7188. The distance to a random model $y = x$, D_r , in the ROC space of (1-specificity, sensitivity) in the test set is therefore $D_r^{\text{Ac}} = |(1 - \text{specificity} - \text{sensitivity})|/\sqrt{2} = 0.1354$. With the decision boundary presented in the main text, we calculate the following distance to a random model in the test set $D_r^{\text{SVM}} = |(1 - 0.4455 - 0.7812)|/\sqrt{2} = 0.1603$. Therefore, the use of CeNuS adds $|0.1603 - 0.1354|/0.1354 = 18.40\%$ information gain compared to the single use of the standardized mean cell area regarding sensitivity and specificity.

This is confirmed when using another measure of prognostic power, the so-called concordance index (c-index). The c-index is a measure of how well a score can sort pairs of patients with respect to an event time, where a c-index of 0.5 corresponds to a completely random score and a c-index of 1 or 0 corresponds to a score that is completely correlated or anticorrelated, respectively, with the event times. Using only the threshold for \bar{A}_C^{stand} , the c-index yields $c_{A_C} = 0.6059$ in the test set; using the boundary presented in the main text in the space of mean cell area and CeNuS, the c-index is $c_{\text{SVM}} = 0.6271$. The gain in c-index by combining mean cell area and CeNuS compared to only using the mean cell area is, therefore, $(c_{\text{SVM}} - c_{A_C})/(c_{A_C} - 0.5) = 20.02\%$.

APPENDIX I: SVM MODEL FOR DECISION BOUNDARY IN STATE SPACE(S)

We used the support vector machine algorithm to develop a prognostic classifier with a custom kernel. SVM is a binary classifier introduced by Vapnik [78], which—given labeled training data—will assign unseen examples to one class or another without requiring a probability distribution. Hereby, each data point is represented as an n -dimensional vector. The classification is conducted by constructing an $(n - 1)$ -dimensional separating hyperplane, where a penalty function for points to the maximal margin—defined by so-called support vectors—is minimized. Moreover, nonlinear functions, i.e., kernels, can be used to transform data into a multidimensional space. To build a custom kernel based on the jamming state diagram, we used the hyperbolic tangent function with custom kernel parameters. The PYTHON Scikit-learn package was used to implement the SVM classifier. A grid search was used to find the optimal penalty parameter C and optimal class weights. The code with all hyperparameter values for the different classifications throughout the paper can be found in Ref. [79].

APPENDIX J: PROGNOSIS USING ONLY CELL OR NUCLEUS SHAPE DISTRIBUTIONS INSTEAD OF CeNuS

Since cell and nucleus shapes are correlated, in this section, we discuss how well only cell shape distributions or nuclei aspect ratio distributions can separate risk groups regarding distant metastases in the state space measured in the retrospective histopathological study.

The CeNuS is defined in Eq. (1) as $\text{CeNuS} = \bar{p}^{\text{stand}} + \sigma_p^{2\text{stand}} + \bar{A}_{R_N}^{\text{stand}} + \sigma_{A_{R_N}}^{2\text{stand}}$. To test if only the cell or nucleus shape is sufficient for prognosis, we consequently use only $\bar{p}^{\text{stand}} + \sigma_p^{2\text{stand}}$ or $\bar{A}_{R_N}^{\text{stand}} + \sigma_{A_{R_N}}^{2\text{stand}}$, respectively. In order to make it easier to compare to CeNuS, we multiply the latter measures by 2 such that the interval approximately matches the one of CeNuS.

We observe that both single measures are able to significantly separate risk groups regarding metastatic risk in the corresponding state spaces. For state spaces of training and test sets as well as the corresponding Kaplan-Meier plot, see Fig. 12 (only nucleus aspect ratio distributions) and Fig. 13 (only cell shape distributions). For training, we use the SVM model as described in Appendix I with slightly varying hyperparameters that can be found on the linked GitHub repository.

APPENDIX K: DESCRIPTIVE STATISTICS FOR RETROSPECTIVE HISTOPATHOLOGICAL STUDY

In the following, the descriptive statistics are shown for the comparison of the training and test sets (Table IX), the comparison of the high-risk and low-risk groups in

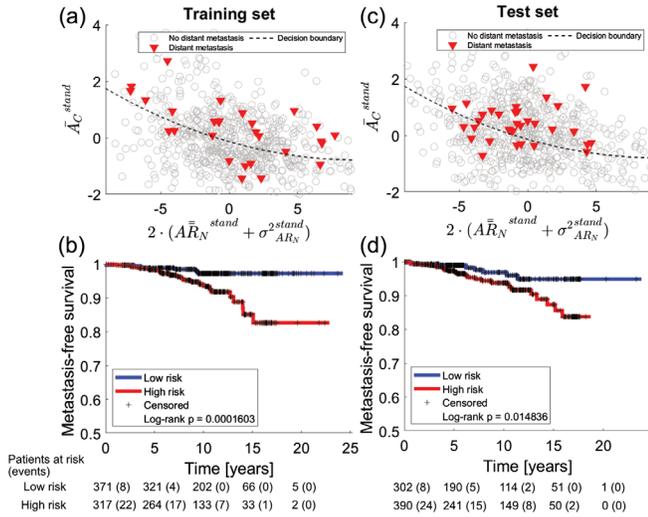


FIG. 12. Clinical relevance of cancer cell unjamming using only nucleus aspect ratio (AR_N) distributions instead of CeNuS as in Fig. 6. The distribution of patient data is displayed with respect to the variables' standardized mean cell area \bar{A}_C^{stand} and $2 \cdot (\bar{AR}_N^{\text{stand}} + \sigma_{AR_N}^{\text{stand}})$ (AR_N denotes the nucleus aspect ratio distribution per patient) for the training collective with minimal treatment ($N = 688$) and in the test collective of the remaining cases ($N = 692$). The state diagrams of the training and test sets are shown in panels (a) and (c), respectively. Patients who developed metastases are indicated by red triangles. Patients who did not develop metastases are indicated by gray circles. The dotted line represents the decision boundary that separates the low-risk patients (below) from the high-risk group (above) estimated by SVM classification; see Appendix I. Kaplan-Meier estimators are used to assess the quality of the risk stratification. The resulting Kaplan-Meier plots are shown in panel (b) for the training set and panel (d) for the test set. These plots show that the high-risk group (in red) is well separated from the low-risk group (in blue). Log-rank p-value training set $p < 0.001$ and test set $p = 0.015$. “Patients at risk” indicates the total number of patients who have not been censored or have not developed distant metastasis by a specific follow-up time. In parentheses is the cumulative number of individuals who will develop distant metastases at a later time.

the training set (Table X), and the comparison of the high-risk and low-risk groups in the test set (Table XI). The p-values are calculated for comparing training and test sets in Table IX and high- and low-risk groups in Tables X and XI. The p-values for continuous variables are calculated with the 2sKS test and for categorical variables with Fisher's exact test.

Below, we give the source and an explanation for all of the entries:

- (i) *Age*.—Source: patient file. Explanation: Cancer is an age-progressive disease; thus, the risk increases with age. On the other hand, the risk of death is higher in younger patients, as these patients are often genetically predisposed and/or the tumor growth is stimulated by hormone levels (e.g., estrogen and progesterone), which may decrease after menopause.

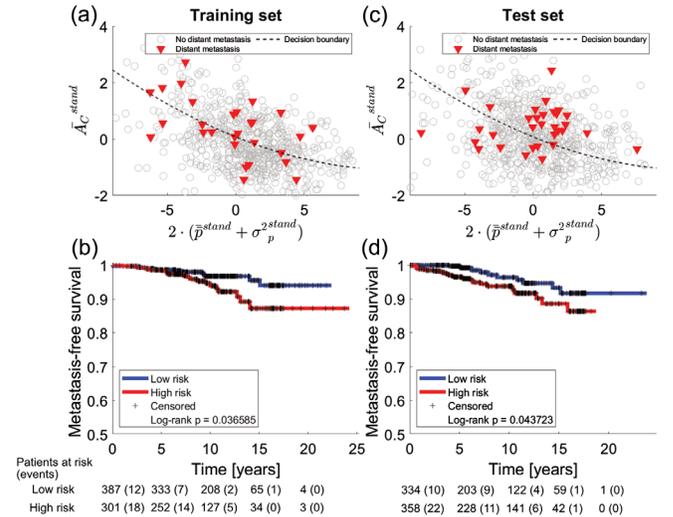


FIG. 13. Clinical relevance of cancer cell unjamming using only cell shape index distributions (p) instead of CeNuS as in Fig. 6. The distribution of patient data is displayed with respect to the variables' standardized mean cell area \bar{A}_C^{stand} and $2 \cdot (\bar{p}^{\text{stand}} + \sigma_p^{\text{stand}})$ (p denotes the cell shape distribution per patient) for the training collective with minimal treatment ($N = 688$) and in the test collective of the remaining cases ($N = 692$). The state diagrams of the training and test sets are shown in panels (a) and (c), respectively. Patients who developed metastases are indicated by red triangles. Patients who did not develop metastases are indicated by gray circles. The dotted line represents the decision boundary that separates the low-risk patients (below) from the high-risk group (above) estimated by SVM classification; see Appendix I. Kaplan-Meier estimators are used to assess the quality of the risk stratification. The resulting Kaplan-Meier plots are shown in panel (b) for the training set and panel (d) for the test set. These plots show that the high-risk group (in red) is well separated from the low-risk group (in blue). Log-rank test in the training set $p = 0.037$ and test set $p = 0.044$. “Patients at risk” indicates the total number of patients who have not been censored or have not developed distant metastasis by a specific follow-up time. In parentheses is the cumulative number of individuals who will develop distant metastases at a later time.

- (ii) *Grade*.—Source: Union Internationale Contre le Cancer (UICC). Explanation: The nature of the cancerous tissue can be an indicator of the aggressiveness or growth rate of the tumor on the basis of the degree of differentiation (extent of degeneration).

Grading following the internationally accepted standard by UICC:

G1: well differentiated (i.e., still similar to the “healthy” cells and less aggressive in growth, thus “low malignant”).

G2: moderately differentiated.

G3: poorly differentiated (i.e., completely degenerated and highly proliferative).

- (iii) *Tumor diameter*.—Source: histological preparation, measured by the examining pathologist. Explanation: The tumor size will affect prognosis no matter how

TABLE IX. Descriptive statistics of all cases (training and test sets combined), the training set, and the test set.

| | All | Training set | Test set | p-value |
|------------------------------|-------------|--------------|------------|----------|
| Number of patients | 1380 | 688 | 692 | |
| Age | | | | 0.0005 |
| Mean | 57.45 | 58.25 | 56.64 | |
| Standard deviation | 8.68 | 8.51 | 8.78 | |
| Range | 26–70 | 29–70 | 26–70 | |
| Grade | | | | |
| 1 | 410 (0.30) | 287 (0.42) | 123 (0.18) | |
| 2 | 771 (0.56) | 391 (0.57) | 380 (0.55) | |
| 3 | 175 (0.13) | 7 (0.01) | 168 (0.24) | |
| Missing information | 24 (0.02) | 3 (0.00) | 21 (0.03) | |
| Tumor diameter | | | | < 0.0001 |
| Mean | 1.66 | 1.32 | 2.01 | |
| Standard deviation | 1.01 | 0.56 | 1.23 | |
| Range | 0–12 | 0.2–4.5 | 0–12 | |
| Lymph node status | | | | < 0.0001 |
| 0 | 1061 (0.77) | 623 (0.91) | 438 (0.63) | |
| 1 | 312 (0.23) | 64 (0.093) | 248 (0.36) | |
| Missing information | 7 (0.01) | 1 (0.00) | 6 (0.01) | |
| Lymph nodes affected | | | | < 0.0001 |
| Mean | 3.71 | 1.36 | 4.51 | |
| Standard deviation | 5.07 | 2.70 | 5.43 | |
| Range | 0–38 | 0–15 | 1–38 | |
| Estrogen receptor status | | | | < 0.0001 |
| 0 | 23 (0.017) | 0 (0) | 23 (0.033) | |
| 1 | 1354 (0.98) | 688 (1) | 666 (0.96) | |
| Missing information | 3 (0.00) | 0 (0.00) | 3 (0.00) | |
| Progesterone receptor status | | | | < 0.0001 |
| 0 | 161 (0.12) | 56 (0.08) | 105 (0.15) | |
| 1 | 1213 (0.88) | 631 (0.92) | 582 (0.84) | |
| Missing information | 6 (0.00) | 1 (0.00) | 5 (0.01) | |
| HER2/Neu receptor status | | | | < 0.0001 |
| 0 | 1018 (0.74) | 589 (0.86) | 429 (0.62) | |
| 1 | 42 (0.03) | 8 (0.01) | 34 (0.05) | |
| Missing information | 320 (0.23) | 91 (0.13) | 229 (0.33) | |
| Endocrine therapy status | | | | < 0.0001 |
| 0 | 47 (0.03) | 0 (0.00) | 47 (0.07) | |
| 1 | 904 (0.66) | 688 (1.00) | 216 (0.31) | |
| Missing information | 429 (0.31) | 0 (0.00) | 429 (0.62) | |
| Chemotherapy status | | | | < 0.0001 |
| 0 | 741 (0.54) | 688 (1.00) | 53 (0.08) | |
| 1 | 204 (0.15) | 0 (0) | 204 (0.30) | |
| Missing information | 435 (0.32) | 0 (0) | 435 (0.63) | |

TABLE X. Descriptive statistics of the training set and the low-risk and high-risk groups resulting from the decision boundary in the space of CeNuS and \bar{A}_C^{stand} presented in the main text.

| | Training set | Low risk | High risk | p-value |
|------------------------------|--------------|------------|------------|----------|
| Number of patients | 688 | 363 | 325 | |
| Age | | | | 0.05 |
| Mean | 58.25 | 57.43 | 59.16 | |
| Standard deviation | 8.51 | 8.83 | 8.05 | |
| Range | 29–70 | 29–70 | 32–70 | |
| Grade | | | | |
| 1 | 287 (0.42) | 215 (0.59) | 72 (0.22) | |
| 2 | 391 (0.57) | 144 (0.40) | 247 (0.76) | |
| 3 | 7 (0.01) | 3 (0.01) | 4 (0.01) | |
| Missing information | 3 (0.004) | 1 (0.00) | 3 (0.01) | |
| Tumor diameter | | | | 0.006 |
| Mean | 1.32 | 1.24 | 1.41 | |
| Standard deviation | 0.56 | 0.49 | 0.62 | |
| Range | 0.2–4.5 | 0.2–3.5 | 0.3–4.5 | |
| Lymph node status | | | | < 0.0001 |
| 0 | 623 (0.91) | 339 (0.93) | 284 (0.87) | |
| 1 | 64 (0.09) | 23 (0.06) | 41 (0.13) | |
| Missing information | 1 (0.00) | 1 (0.00) | 0 (0.00) | |
| Lymph nodes affected | | | | 0.60 |
| Mean | 1.36 | 1.03 | 1.56 | |
| Standard deviation | 2.70 | 2.66 | 2.73 | |
| Range | 0–15 | 0–15 | 0–13 | |
| Estrogen receptor status | | | | |
| 1 | 688 (1.00) | 363 (1.00) | 325 (1.00) | |
| Progesterone receptor status | | | | 0.67 |
| 0 | 56 (0.081) | 28 (0.08) | 28 (0.09) | |
| 1 | 631 (0.92) | 334 (0.92) | 297 (0.91) | |
| Missing information | 1 (0.00) | 1 (0.00) | 0 (0) | |
| HER2/Neu receptor status | | | | 0.48 |
| 0 | 589 (0.86) | 325 (0.90) | 264 (0.81) | |
| 1 | 8 (0.012) | 3 (0.01) | 5 (0.02) | |
| Missing information | 91 (0.13) | 35 (0.10) | 56 (0.17) | |
| Endocrine therapy status | | | | 1.00 |
| 1 | 688 (1.00) | 363 (1.00) | 325 (1.00) | |
| Chemotherapy status | | | | 1.00 |
| 0 | 688 (1.00) | 257 (1.00) | 431 (1.00) | |

many lymph nodes have cancer in them. Breast tumors that are 5 cm or larger are more likely to come back after treatment than smaller tumors. Breast tumors that are smaller than 1 cm and have not spread to the lymph nodes have a very favorable prognosis.

- (iv) *Lymph node status and lymph nodes affected.*—Source: histological preparation, measured by the examining pathologist. Explanation: Nodal involvement

(number of regional nodes positive for cancer) is a strong and independent negative prognostic factor.

When breast cancer begins to spread, the first stop is often one of the neighboring lymph nodes. From there, the tumor is able to spread via the lymphatic system further into more distant lymph nodes and—in the worst case—enters the bloodstream and forms distant metastases (e.g., in the liver, lungs, or bones).

TABLE XI. Descriptive statistics of the test set and the low-risk and high-risk groups resulting from the decision boundary in the space of CeNuS and \bar{A}_C^{stand} presented in the main text.

| | Test set | Low risk | High risk | p-value |
|------------------------------|------------|------------|------------|----------|
| Number of patients | 692 | 301 | 391 | |
| Age | | | | 0.80 |
| Mean | 56.64 | 56.93 | 56.42 | |
| Standard deviation | 8.78 | 8.44 | 9.03 | |
| Range | 26–70 | 35–70 | 26–70 | |
| Grade | | | | |
| 1 | 123 (0.18) | 79 (0.26) | 44 (0.11) | |
| 2 | 380 (0.55) | 166 (0.55) | 214 (0.55) | |
| 3 | 168 (0.24) | 49 (0.16) | 119 (0.30) | |
| Missing information | 21 (0.03) | 7 (0.02) | 14 (0.04) | |
| Tumor diameter | | | | < 0.0001 |
| Mean | 2.01 | 1.79 | 2.18 | |
| Standard deviation | 1.23 | 1.06 | 1.31 | |
| Range | 0–12 | 0–6.5 | 0–12 | |
| Lymph node status | | | | 0.02 |
| 0 | 438 (0.63) | 204 (0.68) | 234 (0.60) | |
| 1 | 248 (0.36) | 93 (0.30) | 155 (0.40) | |
| Missing information | 6 (0.01) | 4 (0.01) | 2 (0.01) | |
| Lymph nodes affected | | | | 0.40 |
| Mean | 4.51 | 4.79 | 4.34 | |
| Standard deviation | 5.43 | 5.79 | 5.21 | |
| Range | 1–38 | 1–38 | 1–38 | |
| Estrogen receptor status | | | | 0.20 |
| 0 | 23 (0.03) | 7 (0.02) | 16 (0.04) | |
| 1 | 666 (0.96) | 292 (0.97) | 374 (0.96) | |
| Missing information | 3 (0.00) | 2 (0.01) | 1 (0.00) | |
| Progesterone receptor status | | | | 0.01 |
| 0 | 105 (0.15) | 33 (0.11) | 72 (0.18) | |
| 1 | 582 (0.84) | 266 (0.88) | 316 (0.81) | |
| Missing information | 5 (0.01) | 2 (0.01) | 3 (0.01) | |
| HER2/Neu receptor status | | | | < 0.0001 |
| 0 | 429 (0.62) | 201 (0.67) | 228 (0.58) | |
| 1 | 34 (0.05) | 7 (0.02) | 27 (0.07) | |
| Missing information | 229 (0.33) | 93 (0.31) | 136 (0.35) | |
| Endocrine therapy status | | | | 0.11 |
| 0 | 47 (0.07) | 23 (0.08) | 24 (0.06) | |
| 1 | 216 (0.31) | 79 (0.26) | 137 (0.35) | |
| Missing information | 429 (0.62) | 199 (0.66) | 230 (0.59) | |
| Chemotherapy status | | | | 0.06 |
| 0 | 53 (0.08) | 23 (0.08) | 30 (0.08) | |
| 1 | 204 (0.30) | 75 (0.25) | 129 (0.33) | |
| Missing information | 435 (0.63) | 203 (0.67) | 232 (0.59) | |

The five-year survival rate decreases in patients with positive lymph nodes versus patients with negative lymph nodes. The higher the number of regional lymph nodes involved, the higher the chance of recurrent disease. Lymph node status contains the information about whether any lymph nodes are affected (lymph node status = 1) or not (lymph node status = 0). Lymph nodes affected is the number of lymphatic sites that were affected during pathological screening.

- (v) *Estrogen or progesterone receptor status.*—Source: immunohistochemical staining, interpreted by the pathologist. Explanation: The hormones estrogen and progesterone can affect the growth of breast cancer cells. They dock to binding sites (hormone receptors, HR) of the cell, which then transmit the growth signal into the cell. In order to determine whether a tumor becomes hormone dependent, we examine how large the proportion of cells and the amount of corresponding HR are. If more than 1% of all tumor cells react to the special marking procedure, it is assumed that the tumor is hormone sensitive. This is expressed by the indication ER+ (estrogen receptor positive) and/or PR+ (progesterone receptor positive). About 75% of breast cancer patients test positive for the estrogen receptor. As soon as one of the two receptor types is positive, it is called hormone receptor positive (HR+). If tumor cells grow hormone dependent, their growth can be slowed down or stopped by (anti)hormone or endocrine therapy, and thus sometimes chemotherapy can be spared.
- (vi) *HER2.*—Source: immunohistochemical staining, interpreted by the pathologist. Explanation: HER2 receptors are binding sites for growth factors on the surface of cancer cells that stimulate the division of these cells. If there is a particularly large number of HER2 receptors on the surface of the cell, this is often accompanied by a more aggressive course of cancer. About 15% of newly diagnosed breast cancer patients have HER2-positive tumors (HER2+). Targeted therapies against HER2 block these receptors and thus inhibit cell growth.
- (vii) *Endocrine therapy status.*—Source: follow-up questionnaire, information provided by the patient and/or treating physician. Explanation: It is indicated for hormone receptor-positive tumors (antihormonal therapy).
- (viii) *Chemotherapy status.*—Source: follow-up questionnaire, information provided by the patient and/or treating physician. Explanation: Through chemotherapy, the cancer cells are attacked and—in the best case—destroyed. This therapy is mainly used for more aggressive growths, such as highly HER2-positive tumors. It is also increasingly recommended for patients under 35 years of age. Before surgery, chemotherapy can be used to shrink the size of the tumor.

TABLE XII. Overview of the investigated clinical samples for dynamical analyses. Grading and staging were given to us by the pathologist. The motility status was detected from live observation as described in the paper. Here, m1mic refers to a micro-metastasis.

| Internal ID | Type | Grade | Stage | Motility status |
|-------------|--------|-------|-------|-----------------|
| 84 | Cervix | 3 | ... | Motile |
| 86 | Cervix | 3 | 4 | Resting |
| 87 | Cervix | ... | ... | Resting |
| 88 | Cervix | ... | ... | Resting |
| 90 | Cervix | 3 | 2b | Motile |
| 93 | Cervix | 3 | 3b | Resting |
| 95 | Cervix | ... | ... | Resting |
| 96 | Mamma | 2 | 2 | Motile |
| 99 | Cervix | 2 | 2b | Motile |
| 100 | Mamma | 2 | 1c | Resting |
| 101 | Mamma | 1 | 1c | Resting |
| 103 | Cervix | ... | ... | Motile |
| 104 | Cervix | 3 | 2b | Resting |
| 105 | Cervix | ... | ... | Motile |
| 106 | Mamma | 2 | m1mic | Motile |
| 108 | Mamma | 3 | 3 | Motile |

APPENDIX L: VITAL PRIMARY SAMPLE INFORMATION

Table XII depicts the vital primary samples we obtained from the University Hospital in Leipzig. The clinical information is shown as well as whether we found motile areas in

TABLE XIII. Known clinical parameters of the vital tumor explants and their respective fluidity status.

| Category | No. moving/no. total samples |
|---------------------------------|------------------------------|
| Total | 8/16 |
| Biopsies (cervix) | 1/7 |
| Mesometrial resections (cervix) | 5/5 |
| Mastectomies | 2/4 |
| Grading known | 6/11 |
| Tumor grade 1 | 0/1 |
| Tumor grade 2 | 3/4 |
| Tumor grade 3 | 3/6 |
| Nodal status known | 5/9 |
| Node negative | 4/6 |
| Node positive | 1/3 |
| Histopathological staging known | 4/9 |
| Stage 1c | 0/2 |
| Stage 2, 2b | 4/5 |
| Stage 3b | 0/1 |
| Stage 4 | 0/1 |

the respective specimens (“Motility status”). Table XIII summarizes the clinical specifications with regard to the information of existing motile areas. We find no significant correlation between stage, grade, or nodal status to the presence of moving areas using Fisher’s exact test.

APPENDIX M: RELATION OF STATE DIAGRAM PROGNOSIS WITH REGARD TO OVERALL SURVIVAL

The decision boundary presented in the main text was derived using distant metastasis classification with support vector machines in the training set, compare Fig. 6(a). The “distant metastasis” event recorded in retrospective patient histories always originates from the primary breast tumor and is therefore unambiguously connected to the disease. Contrary to this, the “overall survival” is the time from the surgery to death, regardless of the cause. Therefore, we choose “distant metastasis” as the event of choice to connect our unjamming criterion to patient outcome. However, in this section, we discuss the prognostic power of the decision boundary trained in the main text (distant metastasis classification) with regard to the overall survival. It is worth noting that a recent study in the US with more than 750,000 breast cancer patients found that only about 50% of the deaths reported by the “overall survival” can be attributed to breast cancer. Of course, the exact numbers for German breast cancer

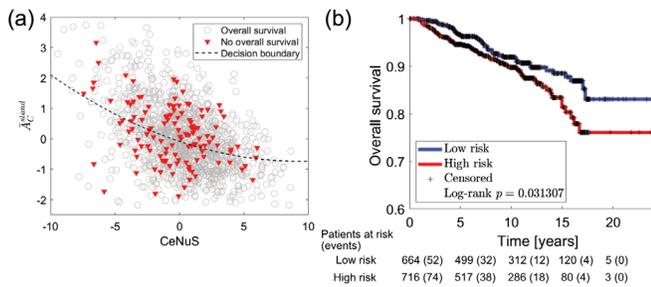


FIG. 14. Clinical relevance of cancer cell unjamming. The distribution of patient data is displayed with respect to the variables of the state diagram (cell area, CeNuS) for all available cases ($N = 1380$). Patients who exhibit no overall survival (death from any cause) are depicted by red triangles. Patients with an overall survival are indicated by gray circles. The dotted line represents the decision boundary that separates the low-risk patients (below) from the high-risk group (above) estimated by SVM classification from the main text in Fig. 6(a). Kaplan-Meier estimators are used for assessing the quality of the risk stratification based on cancer cell unjamming for the overall survival. The resulting Kaplan-Meier plot is shown in panel (b) and exhibits a log-rank p -value training set $p = 0.0313$. “Patients at risk” indicates the total number of patients who have not been censored or have a reported overall death by a specific follow-up time. In parentheses is the cumulative number of individuals who will develop an event (here, overall death) later.

patients and the breast cancer patients in our collective might differ. However, it is clear that the information for the overall survival is burdened by a substantial error when attempting to link it to breast cancer.

Only the event “overall survival” is analyzed in the following analyses, and all other patients (also those developing distant metastasis) are censored.

In Fig. 14(a), we apply the decision boundary presented in the main text to all patients in order to evaluate its significance with regard to overall death. The Kaplan-Meier plot in Fig. 14 shows weakly separated risk groups with a log-rank $p = 0.031$.

APPENDIX N: NUCLEUS PACKING AND SPACING IN CANCER CLUSTERS

In this section, we show that increased nucleus sizes with their associated increased cell [Fig. 3(e)] and cytoplasmic sizes [Fig. 3(f)] correlate with an increase in the effective distance between nucleus edges and cell membranes and therefore with increased cytoplasmic spacing of adjacent nuclei; see Fig. 15(a).

The effective cytoplasmic spacing between the nuclei edge and the cell membrane is the difference in effective radius of the cell segments $r_{\text{eff}}^C = \sqrt{A_C/\pi}$ and nucleus segments $r_{\text{eff}}^N = \sqrt{A_N/\pi}$, respectively. From this, we calculate the (effective) cytoplasmic spacing d between the nucleus border and cell border as $d = r_{\text{eff}}^C - r_{\text{eff}}^N$. In Fig. 15(a), we show that this effective distance between the nucleus border and the cell border increases with increasing nucleus volumes. Furthermore, in Fig. 15(b), we show that the nucleus size scales with the inverse of the number density. A power-law fit reveals that $\bar{A}_N \sim \rho_{2D}^{-\alpha}$, where $\alpha = 0.71 \pm 0.02$.

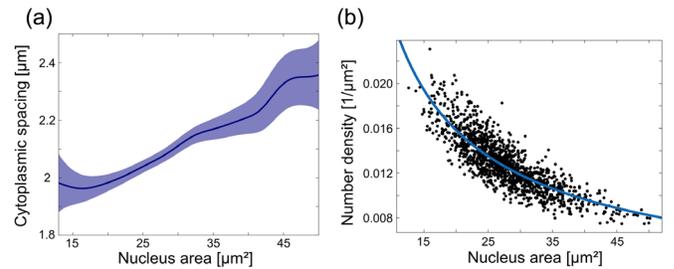


FIG. 15. Nucleus packing and spacing in breast cancer cell clusters measured in $N = 1380$ patients. (a) Cytoplasmic spacing d (effective distance between nucleus edge and cell edge) directly correlating with the nucleus size. The Gaussian moving average window size is $5 \mu\text{m}^2$, and the colored regions correspond to the 95% confidence intervals. (b) Mean nucleus area \bar{A}_N inversely scaling with the number density ρ_{2D} . The blue line corresponds to a power-law fit: $\bar{A}_N \sim \rho_{2D}^{-\alpha}$, where $\alpha = 0.71 \pm 0.02$. The root-mean-squared error is 0.0001. All observables are ensemble averages for one patient.

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, *Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*, *CA: Cancer J. Clinicians* **68**, 394 (2018).
- [2] H. Dillekås, M. S. Rogers, and O. Straume, *Are 90% of Deaths from Cancer Caused by Metastases?*, *Cancer Med.* **8**, 5574 (2019).
- [3] C. L. Carter, C. Allen, and D. E. Henson, *Relation of Tumor Size, Lymph Node Status, and Survival in 24,740 Breast Cancer Cases*, *Cancer* **63**, 181 (1989).
- [4] B. Weigelt, J. L. Peterse, and L. J. Van't Veer, *Breast Cancer Metastasis: Markers and Models*, *Nat. Rev. Cancer* **5**, 591 (2005).
- [5] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, *Breast Cancer*, *Nat. Rev. Disease Primers* **5**, 66 (2019).
- [6] O. Mete and S. L. Asa, *Pathological Definition and Clinical Significance of Vascular Invasion in Thyroid Carcinomas of Follicular Epithelial Derivation*, *Mod. Pathol.* **24**, 1545 (2011).
- [7] T. S. Gerashchenko, N. M. Novikov, N. V. Krakhmal, S. Y. Zolotaryova, M. V. Zavyalova, N. V. Cherdynseva, E. V. Denisov, and V. M. Perelmuter, *Markers of Cancer Cell Invasion: Are They Good Enough?*, *J. Clinical Med.* **8**, 1092 (2019).
- [8] D. Hanahan and R. A. Weinberg, *Hallmarks of Cancer: The Next Generation*, *Cell* **144**, 646 (2011).
- [9] S. Grosser, J. Lippoldt, L. Oswald, M. Merkel, D. M. Sussman, F. Renner, P. Gottheil, E. W. Morawetz, T. Fuhs, X. Xie, S. Pawlizak, A. W. Fritsch, B. Wolf, L.-C. Horn, S. Briest, B. Aktas, M. L. Manning, and J. A. Käs, *Cell and Nucleus Shape as an Indicator of Tissue Fluidity in Carcinoma*, *Phys. Rev. X* **11**, 011033 (2021).
- [10] J. A. Mitchel, A. Das, M. J. O'Sullivan, I. T. Stancil, S. J. DeCamp, S. Koehler, O. H. Ocaña, J. P. Butler, J. J. Fredberg, M. A. Nieto *et al.*, *In Primary Airway Epithelial Cells, the Unjamming Transition Is Distinct from the Epithelial-to-Mesenchymal Transition*, *Nat. Commun.* **11**, 1 (2020).
- [11] T. Fuhs *et al.*, *Rigid Tumours Contain Soft Cancer Cells*, *Nat. Phys.* **18**, 1510 (2022).
- [12] E. Blauth, H. Kubitschke, P. Gottheil, S. Grosser, and J. A. Käs, *Jamming in Embryogenesis and Cancer Progression*, *Front. Phys.* **9**, 328 (2021).
- [13] L. Atia, D. Bi, Y. Sharma, J. A. Mitchel, B. Gweon, S. A. Koehler, S. J. DeCamp, B. Lan, J. H. Kim, R. Hirsch *et al.*, *Geometric Constraints During Epithelial Jamming*, *Nat. Phys.* **14**, 613 (2018).
- [14] D. Bi, J. Lopez, J. M. Schwarz, and M. L. Manning, *A Density-Independent Rigidity Transition in Biological Tissues*, *Nat. Phys.* **11**, 1074 (2015).
- [15] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning, *Motility-Driven Glass and Jamming Transitions in Biological Tissues*, *Phys. Rev. X* **6**, 021011 (2016).
- [16] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning, *Energy Barriers and Cell Migration in Densely Packed Tissues*, *Soft Matter* **10**, 1885 (2014).
- [17] P. Heine, J. Lippoldt, G. A. Reddy, P. Katira, and J. A. Käs, *Anomalous Cell Sorting Behavior in Mixed Monolayers Discloses Hidden System Complexities*, *New J. Phys.* **23**, 043034 (2021).
- [18] R. Mahaffy, C. Shih, F. MacKintosh, and J. Käs, *Scanning Probe-Based Frequency-Dependent Microrheology of Polymer Gels and Biological Cells*, *Phys. Rev. Lett.* **85**, 880 (2000).
- [19] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. A. Mitchel, A. Khan, J. P. Butler, D. Roblyer *et al.*, *A Novel Jamming Phase Diagram Links Tumor Invasion to Non-Equilibrium Phase Separation*, *iScience* **24**, 103252 (2021).
- [20] Y. L. Han, A. F. Pegoraro, H. Li, K. Li, Y. Yuan, G. Xu, Z. Gu, J. Sun, Y. Hao, S. K. Gupta *et al.*, *Cell Swelling, Softening and Invasion in a Three-Dimensional Breast Cancer Model*, *Nat. Phys.* **16**, 101 (2020).
- [21] I. R. Powley, M. Patel, G. Miles, H. Pringle, L. Howells, A. Thomas, C. Kettleborough, J. Bryans, T. Hammonds, M. MacFarlane *et al.*, *Patient-Derived Explants (PDEs) as a Powerful Preclinical Platform for Anti-cancer Drug and Biomarker Discovery*, *Br. J. Cancer* **122**, 735 (2020).
- [22] R. Staneva, F. El Marjou, J. Barbazan, D. Krndija, S. Richon, A. G. Clark, and D. M. Vignjevic, *Cancer Cells in the Tumor Core Exhibit Spatially Coordinated Migration Patterns*, *J. Cell Sci.* **132**, jcs220277 (2019).
- [23] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.13.031003> for video that shows a vital patient-derived breast cancer sample stained with a vital DNA stain where nuclei (cells) squeeze between each other. This is the same specimen as shown in Fig. 2 in the main text. The 2D cut is at 20 μm from the bottom of the sample. This is a native video and not registered. The scale bar is 100 μm in length.
- [24] M. J. O'Sullivan, J. A. Mitchel, A. Das, S. Koehler, H. Levine, D. Bi, Z. D. Nagel, and J.-A. Park, *Irradiation Induces Epithelial Cell Unjamming*, *Front. Cell Dev. Biol.* **8**, 21 (2020).
- [25] T. A. Sharp, M. Merkel, M. L. Manning, and A. J. Liu, *Inferring Statistical Properties of 3D Cell Geometry from 2D Slices*, *PLoS One* **14**, e0209892 (2019).
- [26] J.-A. Park, J. H. Kim, D. Bi, J. A. Mitchel, N. T. Qazvini, K. Tantisira, C. Y. Park, M. McGill, S.-H. Kim, B. Gweon *et al.*, *Unjamming and Cell Shape in the Asthmatic Airway Epithelium*, *Nat. Mater.* **14**, 1040 (2015).
- [27] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, *Cell Detection with Star-Convex Polygons*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, New York, 2018), pp. 265–273.
- [28] F. E. Harrell, Jr., K. L. Lee, and D. B. Mark, *Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors*, *Stat. Med.* **15**, 361 (1996).
- [29] M. H. Galea, R. W. Blamey, C. E. Elston, and I. O. Ellis, *The Nottingham Prognostic Index in Primary Breast Cancer*, *Breast Cancer Res. Treat.* **22**, 207 (1992).
- [30] J. J. Fredberg, *On the Origins of Order*, *Soft Matter* **18**, 2346 (2022).
- [31] A. Mongera, P. Rowghanian, H. J. Gustafson, E. Shelton, D. A. Kealhofer, E. K. Carn, F. Serwane, A. A. Lucio,

- J. Giammona, and O. Campàs, *A Fluid-to-Solid Jamming Transition Underlies Vertebrate Body Axis Elongation*, *Nature (London)* **561**, 401 (2018).
- [32] N. I. Petridou, B. Corominas-Murtra, C.-P. Heisenberg, and E. Hannezo, *Rigidity Percolation Uncovers a Structural Basis for Embryonic Tissue Phase Transitions*, *Cell* **184**, 1914 (2021).
- [33] L. Atia, J. J. Fredberg, N. S. Gov, and A. F. Pegoraro, *Are Cell Jamming and Unjamming Essential in Tissue Development?*, *Cells Dev.* **168**, 203727 (2021).
- [34] T. Dyba, G. Randi, F. Bray, C. Martos, F. Giusti, N. Nicholson, A. Gavin, M. Flego, L. Neamtiu, N. Dimitrova *et al.*, *The European Cancer Burden in 2020: Incidence and Mortality Estimates for 40 Countries and 25 Major Cancers*, *Eur. J. Cancer* **157**, 308 (2021).
- [35] L. Oswald, S. Grosser, D. M. Smith, and J. A. Käs, *Jamming Transitions in Cancer*, *J. Phys. D* **50**, 483001 (2017).
- [36] X. Ye, T. Brabletz, Y. Kang, G. D. Longmore, M. A. Nieto, B. Z. Stanger, J. Yang, and R. A. Weinberg, *Upholding a Role for EMT in Breast Cancer Metastasis*, *Nature (London)* **547**, E1 (2017).
- [37] K. R. Fischer, A. Durrans, S. Lee, J. Sheng, F. Li, S. T. Wong, H. Choi, T. El Rayes, S. Ryu, J. Troeger *et al.*, *Epithelial-to-Mesenchymal Transition Is Not Required for Lung Metastasis but Contributes to Chemoresistance*, *Nature (London)* **527**, 472 (2015).
- [38] M. K. Jolly, S. A. Mani, and H. Levine, *Hybrid Epithelial/Mesenchymal Phenotype(s): The ‘Fittest’ for Metastasis?*, *Biochim. Biophys. Acta* **1870**, 151 (2018).
- [39] P. Chakraborty, J. T. George, S. Tripathi, H. Levine, and M. K. Jolly, *Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum*, *Front. Bioeng. Biotechnol.* **8**, 220 (2020).
- [40] K. Wolf, M. Te Lindert, M. Krause, S. Alexander, J. Te Riet, A. L. Willis, R. M. Hoffman, C. G. Figdor, S. J. Weiss, and P. Friedl, *Physical Limits of Cell Migration: Control by ECM Space and Nuclear Deformation and Tuning by Proteolysis and Traction Force*, *J. Cell Biol.* **201**, 1069 (2013).
- [41] P. M. Davidson, C. Denais, M. C. Bakshi, and J. Lammerding, *Nuclear Deformability Constitutes a Rate-Limiting Step During Cell Migration in 3D Environments*, *Cell. Mol. Bioeng.* **7**, 293 (2014).
- [42] P. A. Janmey, U. Euteneuer, P. Traub, and M. Schliwa, *Viscoelastic Properties of Vimentin Compared with Other Filamentous Biopolymer Networks*, *J. Cell Biol.* **113**, 155 (1991).
- [43] T. Yeung, P. C. Georges, L. A. Flanagan, B. Marg, M. Ortiz, M. Funaki, N. Zahir, W. Ming, V. Weaver, and P. A. Janmey, *Effects of Substrate Stiffness on Cell Morphology, Cytoskeletal Structure, and Adhesion*, *Cell Motil. Cytoskeleton* **60**, 24 (2005).
- [44] R. Vincent, E. Bazellieres, C. Pérez-González, M. Uroz, X. Serra-Picamal, and X. Trepat, *Active Tensile Modulus of an Epithelial Monolayer*, *Phys. Rev. Lett.* **115**, 248103 (2015).
- [45] A. Lomakin, C. Cattin, D. Cuvelier, Z. Alraies, M. Molina, G. Nader, N. Srivastava, P. Saez, J. Garcia-Arcos, I. Zhitnyak *et al.*, *The Nucleus Acts as a Ruler Tailoring Cell Responses to Spatial Constraints*, *Science* **370** (2020). [10.1126/science.aba2894](https://doi.org/10.1126/science.aba2894)
- [46] A. Elosegui-Artola, I. Andreu, A. E. Beedle, A. Lezamiz, M. Uroz, A. J. Kosmalska, R. Oria, J. Z. Kechagia, P. Rico-Lastres, A.-L. Le Roux *et al.*, *Force Triggers YAP Nuclear Entry by Regulating Transport Across Nuclear Pores*, *Cell* **171**, 1397 (2017).
- [47] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G.-J. Bakker, J. Starrau *et al.*, *Cell–Cell Adhesion and 3D Matrix Confinement Determine Jamming Transitions in Breast Cancer Invasion*, *Nat. Cell Biol.* **22**, 1103 (2020).
- [48] J. Devany, D. M. Sussman, T. Yamamoto, M. L. Manning, and M. L. Gardel, *Cell Cycle-Dependent Active Stress Drives Epithelia Remodeling*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1917853118 (2021).
- [49] J. Ranft, M. Basan, J. Elgeti, J.-F. Joanny, J. Prost, and F. Jülicher, *Fluidization of Tissues by Cell Division and Apoptosis*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20863 (2010).
- [50] M. Czajkowski, D. M. Sussman, M. C. Marchetti, and M. L. Manning, *Glassy Dynamics in Models of Confluent Tissue with Mitosis and Apoptosis*, *Soft Matter* **15**, 9133 (2019).
- [51] D. Matoz-Fernandez, K. Martens, R. Sknepnek, J. Barrat, and S. Henkes, *Cell Division and Death Inhibit Glassy Behaviour of Confluent Tissues*, *Soft Matter* **13**, 3205 (2017).
- [52] A. Doostmohammadi, S. P. Thampi, T. B. Saw, C. T. Lim, B. Ladoux, and J. M. Yeomans, *Celebrating Soft Matter’s 10th Anniversary: Cell Division: A Source of Active Stress in Cellular Monolayers*, *Soft Matter* **11**, 7328 (2015).
- [53] A. E. M. Reed, J. R. Kutasovic, S. R. Lakhani, and P. T. Simpson, *Invasive Lobular Carcinoma of the Breast: Morphology, Biomarkers and Omics*, *Breast Cancer Res.* **17**, 1 (2015).
- [54] Z. Varga, P. Sinn, F. Fritzsche, A. von Hochstetter, A. Noske, P. Schraml, C. Tausch, A. Trojan, and H. Moch, *Comparison of Endopredict and Oncotype DX Test Results in Hormone Receptor Positive Invasive Breast Cancer*, *PLoS One* **8**, e58483 (2013).
- [55] I. Sestak, R. Buus, J. Cuzick, P. Dubsy, R. Kronenwett, C. Denkert, S. Ferree, D. Sgroi, C. Schnabel, F. L. Baehner *et al.*, *Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial*, *JAMA Oncol.* **4**, 545 (2018).
- [56] L. Chin, Y. Xia, D. E. Discher, and P. A. Janmey, *Mechanotransduction in Cancer*, *Curr. Opin. Chem. Eng.* **11**, 77 (2016).
- [57] J. A. Espina, C. L. Marchant, and E. H. Barriga, *Durotaxis: The Mechanical Control of Directed Cell Migration*, *FEBS J.* **289**, 2736 (2021).
- [58] C. C. DuFort, M. J. Paszek, and V. M. Weaver, *Balancing Forces: Architectural Control of Mechanotransduction*, *Nat. Rev. Mol. Cell Biol.* **12**, 308 (2011).
- [59] K. Pantel and M. Speicher, *The Biology of Circulating Tumor Cells*, *Oncogene* **35**, 1216 (2016).
- [60] A. Palamidessi, C. Malinverno, E. Frittoli, S. Corallino, E. Barbieri, S. Sigismund, G. V. Beznoussenko, E. Martini, M. Garre, I. Ferrara *et al.*, *Unjamming Overcomes Kinetic and Proliferation Arrest in Terminally Differentiated Cells and Promotes Collective Motility of Carcinoma*, *Nat. Mater.* **18**, 1252 (2019).

- [61] F. Giavazzi, M. Paoluzzi, M. Macchi, D. Bi, G. Scita, M. L. Manning, R. Cerbino, and M. C. Marchetti, *Flocking Transitions in Confluent Tissues*, *Soft Matter* **14**, 3471 (2018).
- [62] C. Blanch-Mercader, V. Yashunsky, S. Garcia, G. Duclos, L. Giomi, and P. Silberzan, *Turbulent Dynamics of Epithelial Cell Cultures*, *Phys. Rev. Lett.* **120**, 208101 (2018).
- [63] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, C. T. Lim, J. M. Yeomans, and B. Ladoux, *Topological Defects in Epithelia Govern Cell Death and Extrusion*, *Nature (London)* **544**, 212 (2017).
- [64] A. A. Khalil, O. Ilina, P. G. Gritsenko, P. Bult, P. N. Span, and P. Friedl, *Collective Invasion in Ductal and Lobular Breast Cancer Associates with Distant Metastasis*, *Clin. Exp. Metastasis* **34**, 421 (2017).
- [65] K. J. Cheung and A. J. Ewald, *A Collective Route to Metastasis: Seeding by Tumor Cell Clusters*, *Science* **352**, 167 (2016).
- [66] E. Wrenn, Y. Huang, and K. Cheung, *Collective Metastasis: Coordinating the Multicellular Voyage*, *Clin. Exp. Metastasis* **38**, 373 (2021).
- [67] T. G. Phan and P. I. Croucher, *The Dormant Cancer Cell Life Cycle*, *Nat. Rev. Cancer* **20**, 398 (2020).
- [68] A. M. Afifi, A. M. Saad, M. J. Al-Husseini, A. O. Elmehraath, D. W. Northfelt, and M. B. Sonbol, *Causes of Death After Breast Cancer Diagnosis: A US Population-Based Analysis*, *Cancer* **126**, 1559 (2020).
- [69] Y. Gu, Y. Liu, L. Fu, L. Zhai, J. Zhu, Y. Han, Y. Jiang, Y. Zhang, P. Zhang, Z. Jiang *et al.*, *Tumor-Educated B Cells Selectively Promote Breast Cancer Lymph Node Metastasis by HSPA4-Targeting IgG*, *Nat. Med.* **25**, 312 (2019).
- [70] A. Gandalovičová, D. Rosel, M. Fernandes, P. Veselý, P. Heneberg, V. Čermák, L. Petruželka, S. Kumar, V. Sanz-Moreno, and J. Brábek, *Migrastatics—Anti-Metastatic and Anti-Invasion Drugs: Promises and Challenges*, *Trends Cancer* **3**, 391 (2017).
- [71] J. Guck, S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H. M. Erickson, R. Ananthakrishnan, D. Mitchell *et al.*, *Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence*, *Biophys. J.* **88**, 3689 (2005).
- [72] M. Lekka, P. Laidler, D. Gil, J. Lekki, Z. Stachura, and A. Hryniewicz, *Elasticity of Normal and Cancerous Human Bladder Cells Studied by Scanning Force Microscopy*, *Eur. Biophys. J.* **28**, 312 (1999).
- [73] O. Sen, A. T. Saurin, and J. M. Higgins, *The Live Cell DNA Stain SiR-Hoechst Induces DNA Damage Responses and Impairs Cell Cycle Progression*, *Sci. Rep.* **8**, 1 (2018).
- [74] J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri, *Trackmate: An Open and Extensible Platform for Single-Particle Tracking*, *Methods* **115**, 80 (2017).
- [75] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman *et al.*, *QuPath: Open Source Software for Digital Pathology Image Analysis*, *Sci. Rep.* **7**, 1 (2017).
- [76] F. Meyer, *Topographic Distance and Watershed Lines*, *Signal Processing* **38**, 113 (1994).
- [77] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (John Wiley & Sons, New York, 2005).
- [78] V. N. Vapnik, *An Overview of Statistical Learning Theory*, *IEEE Trans. Neural Networks* **10**, 988 (1999).
- [79] <https://github.com/pgotth/SVM-classifier>.