Spatial self-organization of cancer stem cell niches revealed by live single-cell imaging

Mathilde Brulé, Anaïs Horochowska, Emeline Fontaine, Raoul Torero-Ibad, Flavie Woesteland, Marie Denoulte, Jean Pesez, Eric Adriaenssens, Robert-Alain Toillon, Xuefen Le Bourhis, et al.

Spatial self-organization of cancer stem cell niches revealed by live single-cell imaging

Mathilde Brulé[1†], **Anaïs Horochowska**[1†], **Emeline Fontaine**[1], **Raoul Torero-Ibad**[2], **Flavie Woesteland**[1], **Marie Denoulet**[1§], **Jean Pesez**[2], **Eric Adriaenssens**[1], **Robert-Alain Toillon**[1], **Xuefen Le Bourhis**[1], **Benjamin Pfeuty**[2], **Chann Lagadec**[1,3,*‡], **François Anquez**[2*‡]

***For correspondence:**
chann.lagadec@inserm.fr;
francois.anquez@univ-lille.fr

[†]These authors contributed equally to this work
[‡]These authors also contributed equally to this work

Present address: [§]Nantes Université, INSERM, CNRS, Université d'Angers, CRCI2NA, Nantes, France

[1]Univ. Lille, CNRS, Inserm, CHU Lille, Centre Oscar Lambret, UMR9020 – UMR-S 1277 - Canther – Cancer Heterogeneity, plasticity and Resistance to Therapies, F-59000 Lille, France ; [2]Univ. Lille, CNRS, UMR 8523 - PhLAM - Physique des Lasers Atomes et Molécules, F-59000 Lille, France; [3]Institut pour la Recherche sur le Cancer de Lille (IRCL), 59000 Lille, France

Abstract

Phenotypic plasticity is a major factor of tumor heterogeneity and treatment resistance. In particular, cancer stem cells (CSCs) represent a small subpopulation within tumors with self-renewal and tumor-forming capabilities. Understanding reprogramming, maintenance, and lineage properties of CSCs requires dedicated tools to disentangle the respective influences of phenotypic inheritance and cell-cell interactions. Here we set up ultra-wide field microscopy of breast cancer cell lines expressing a stemness fluorescent reporter for several days. The fluorescent reporter distinguishes three phenotypes : cancer stem cells (CSCs), cancer differentiated cells (CDCs) and intermediate/transiting cancer cells (iCCs). Spatial statistics indicate significant zonation, aka phenotypic niches, with CSC clustering near each other but away from CDCs. Surprisingly, single cell time series reveal spontaneous reprogramming events from CDC to CSC even in unperturbed populations. We identify that such transitions are prone to arise during the cell cycle. Moreover, lineage analysis shows that the phenotype is partially inherited from ancestor cells. However, such heredity is not sufficient to explain the spatial properties of the cell population, which also depend on cell-cell interactions. Indeed, we identified that phenotypic transitions of cancer cells are influenced by the phenotypic state of neighboring cells. Reprogramming into CSCs is respectively promoted and inhibited by the presence of CSCs and CDCs in the neighborhood. Altogether, our results disentangle how phenotypic inheritance and intercellular interactions orchestrate the spatio-temporal self-organization of cancer cell heterogeneity, maintaining a subpopulation of CSCs within niches.

Abbreviations

CSCs: Cancer stem cells; iCCs: intermediate cancer cells; CDCs: Cancer differentiated cells; EMT: Epitheial-Mesenchyma Transition; pALDH1A1: promoter of the *ALDH1A1* (Aldehyde Deshydrogenase 1A1) gene; PDF: probability density function ; PCF: Point Correlation Function, CSR: complete spatial randomness; SCTS: single-cell time series; GAP statistics ; RFU : Relative Fluorescence Units.

Introduction

Tumors constitute a diverse array of cells, encompassing transformed cancer cells, supportive cells, and cells infiltrating the tumor. Tumor heterogeneity extends beyond malignant cells alone, as tumors constitute intricate ecosystems housing various cell types such as endothelial cells, macrophage and lymphocyte cells, cancer-associated fibroblasts, and an intricate extracellular matrix network, contributing to spatial and temporal disparities in the tumor microenvironment (***Junttila and De Sauvage, 2013***; ***Lu et al., 2012***). Moreover, these diversities result in variations among patients, but also within individual tumors, influencing the spatial and temporal characteristics of tumors. This variability significantly impacts how tumors respond to drugs and ultimately affects the outcome of the disease (***Kashyap et al., 2022***).

Throughout the cancer cell population, heterogeneity arises from mutation and epigenetic origins, but also from the regulation of different phenotypes within the cancer cell population. The clonal evolution model postulates that stochastic mutations in individual tumor cells provide a basis for adaptation and selection of the fittest tumor clones, driving intra-tumor heterogeneity through selection. Clones with advantageous traits proliferate, while those less fit are outcompeted and may eventually disappear. Importantly, these advantages may vary over time and with surrounded environment, different areas of the tumor favoring distinct clone types based on environmental conditions (***Anderson et al., 2006***; ***Sottoriva et al., 2010***; ***Waclaw et al., 2015***). While initial observations indicated the presence of sub-clones within tumors exhibiting differences in genetic makeup and response to chemotherapy (***Shapiro et al., 1981***; ***Yung et al., 1982***), recent profiling efforts utilizing comprehensive sequencing and methylation analysis across various tumor regions have unveiled multiple distinct clones harboring unique genetic mutations and epigenetic patterns within a single tumor (***Anderson et al., 2011***; ***Gerlinger et al., 2012***). Meanwhile, the cancer stem cell (CSC) model suggests that only a subset of cancer cells possess the capacity for indefinite self-renewal, giving rise to progenitors and differentiated cells which sustain tumor growth in a hierarchical manner akin to normal tissue hierarchy maintained by healthy normal stem cells. Consequently, CSCs generate cellular diversity by establishing a differentiation hierarchy within the tumor (***Colacino et al., 2018***). However, this hierarchy is not strictly unidirectional, as terminally differentiated cells can revert to a stem cell-like state under specific conditions, a phenomenon known as cell plasticity (***Friedmann-Morvinski and Verma, 2014***; ***Wahl and Spike, 2017***; ***Brown et al., 2022***). The concept of cell plasticity reconciles elements of both stochastic and CSC models, whereby mutations in differentiated cells can confer self-renewal ability, establishing new hierarchical CSC clones and increasing functional diversity within the tumor (***Wahl and Spike, 2017***). The capacity of cells to transition between states through various programs, such as Epithelial-Mesenchymal Transition (EMT), suggests that CSCs may not always be predetermined; instead, stemness can be considered as a cellular state that can be gained or lost reversibly. In essence, cellular plasticity enables dynamic transitions between CSCs and non-CSCs (***Chaffer et al., 2011***; ***Gupta et al., 2019***). Furthermore, distinct subsets of CSCs may occupy different positions along the epithelial-mesenchymal axis and have the potential to undergo inter-conversion (***Liu et al., 2014***; ***Bocci et al., 2018***, ***2019a***).

Upon homeostatic conditions, cancer cells exhibit stable populations of both stem-like and differentiated cells. The activity of CSCs is orchestrated by a multitude of pluripotent transcription factors including OCT4, SOX2, NANOG, KLF4, and MYC (***Zhang and Wang, 2008***). Furthermore, numerous intracellular signaling pathways play pivotal roles in regulating CSC behavior. These pathways include Wnt (***Kanwar et al., 2010***), Notch (***Wang et al., 2010***), Sonic Hedgehog (***Li et al., 2007***), NF-κB, JAK-STAT, PI3K/AKT/mTOR, TGF/SMAD, and PPAR (***Yang et al., 2020***). In addition to these intracellular pathways, various extracellular influences are also crucial. These influences include vascular niches, hypoxia, tumor-associated macrophages, cancer-associated fibroblasts, cancer-associated mesenchymal stem cells, extracellular matrix, and exosomes (***Plaks et al., 2015***; ***Yoshida and Saya, 2016***; ***Oshimori et al., 2021***). Both paracrine and juxtacrine mediated signalling spatially shape the

tumor leading to the formation of CSC niches. Interestingly a complex equilibrium maintains a balance between CSCs and differentiated cancer cells. As an example, differentiated cancer cells secrete factors like Interleukin-6 (IL6) (***Liu et al., 2015***) and BDNF-NTRK2-VGF (***Wang et al., 2018***), NGF/p75NTR axis (***Tomellini et al., 2015***), fostering the survival and self-renewal of CSCs in breast cancer and glioblastoma. While CSCs secrete factors like DKK1 pushing forward differentiation (***Wu et al., 2022***). Perturbing this equilibrium results in unbalancing the system and induces phenotypic switch. Indeed, phenotypic plasticity of CSC, also called reprogramming from non-CSC into CSC, was first observed upon perturbations such as anticancer treatments ((***Lagadec et al., 2012***)) or specific environmental condition as hypoxia (***Li et al., 2009***). After treatment, enriched Breast Cancer Stem Cells (BCSCs) swiftly restore the parental cell composition, indicating the inclination of BCSCs to differentiate under such circumstances (***Bidan et al., 2019***; ***Gupta et al., 2011***; ***Iliopoulos et al., 2011***).

In view of these numerous pathways involved in phenotypic balance, in order to propose efficient therapeutic strategies, there is a need for *ex vivo* assays emulating the spatial arrangement of cell types in tumors, as can be done to recapitulate essential steps of development such as germ layer and axial patterning (***Warmflash et al., 2014***). However, the diversity of phenotypes and interaction motifs and signalling pathways involved make it difficult to understand how CSCs niche auto-organizes in space and time. Some general insights about phenotypic heterogeneity and spatial organization of stem cell niche could be gained through system-level approaches. Indeed, regardless the signaling pathways at play, the main sources of spatial self-organization of cell phenotypes are fundamentally influenced by a few cell-fate processes, namely differentiation, division, motility and death (***Kicheva et al., 2012***; ***Grace and Hütt, 2015***; ***Landge et al., 2020***). While the complexity of tumor heterogeneity arises from the diversity of intracellular and extracellular signaling mechanisms that couple all these processes, some key determinants can be unveiled by focusing on the initial and spontaneous trends to self-organize in unperturbed environments. We therefore hypothesize that cultured cancer cell lines show some rudimentary forms of self-organization featured with a spatial segregation of cancer stem cell niche.

Live cell imaging tools provide most spatio-temporal information sufficient to track the main processes involved in spatial self-organization. This wealth of data facilitates a system-level approach to investigate the intricate interplay between the niche effect and the homeostatic capacity of tumor models. It also aids in identifying key factors that could disrupt the cancer stem cell niche and cellular plasticity. In this work, we investigate the interplay between the temporal transitions and the spatial distribution of stemness phenotypes in population of breast cancer cell upon unperturbed and steady culture conditions. To this aim, we developed an ultra-wide field imaging system capable of tracking thousands of single-cells in space and time for days. Using mNeptune fluorescent protein expression under the control of the promoter of ALDH1A1 (pALDH1A1:mNeptune) as a stemness reporter for breast cancer cells (***Bidan et al., 2019***), we conducted a series of analyses to characterize the spatial and temporal features of phenotypic heterogeneity. From the broad distribution of fluorescent marker over the population, statistical analysis identifies two main phenotypes associated to stemness and differentiation traits, as well as an intermediate state. Point pattern analysis uncovered spatial segregation between stem-like and differentiated phenotypes. Notably, we observed phenotypic transitions, including reprogramming events from CDCs to CSCs phenotypes, even within unperturbed cancer cell populations. Analysis of cell fate further underscored the significant role of intercellular signaling in phenotypic transitions and CSC reprogramming. Altogether, our findings demonstrate that cancer stem cell niche spontaneously emerge in unconstrained population of cancer cells through the interplay of phenotypic inheritance across generations and intercellular communications that stabilize the stemness phenotype.

Results

Ultra-wide field imaging characterizes cancer cell phenotypic diversity and its spatial distribution

To distinguish CSCs from non-CSCs and to track the spatio-temporal dynamics of cells, we previously developed a reporter based on mNeptune fluorescent protein expression under the control of the promoter of ALDH1A1 (pALDH1A1:mNeptune) (***Bidan et al., 2019***). In 2D cell culture conditions, CSC frequency is rather low (from <1% up to 5%) (***Fillmore and Kuperwasser, 2008***). Thus, we set up ultra-wide field microscopy to investigate the phenotypic diversity of breast cancer cells with sufficient statistics.

SUM159-PT and MDA-MB-231 cells were imaged with a high Numerical Aperture objective to allow detection of the faint pALDH1a1:mNeptune signal. A 6.8mm $\times$ 6.8 mm wide field of view is reconstructed by assembling 60X high resolution images (***Figure 1***). Images reveal a large diversity in single cell mNeptune fluorescence intensity, with majority of cells exhibiting no or low fluorescence and a small fraction of cells expressing high fluorescence (***Figure 1***C). This observation qualitatively aligns with a low CSC fraction reported in breast cancer cells (***Fillmore and Kuperwasser, 2008***). At larger scale (***Figure 1***A), reconstructed images indicate high spatial heterogeneity with small regions populated by cells exhibiting high fluorescence.

To further quantify distribution of SUM159-PT sub-population, in addition to the CSC reporter, we used cell segmentation and image processing (see material and methods). At cellular level, the CSC reporter signal is evenly spread over the cell volume (***Figure 1***C). Thus, we used average nuclear mNeptune fluorescence as a proxy for single cell phenotype. We extracted single cell fluorescence for up to 10^4 individual cells per reconstructed field of view. We observed that single cell fluorescence show a tailed distribution (***Figure 1***D and ***Figure 1—figure Supplement 1*** A) with $\sim 95\%$ of cells with fluorescence below 130 Relative Fluorescence Units (RFU). More precisely the empirical fluorescence distribution is well described by the sum of three probability density functions (lower panel of ***Figure 1***D). A normal distribution centered at low fluorescence level (2.6 RFU) likely corresponds to cancer differentiated cells (CDCs) while, on the other side, a gamma distribution with a mean of $\sim$ 150 RFU corresponds to CSCs. Additionally, a third distribution, also fitted by a gamma distribution (mean $\sim$ 42 RFU) indicates the presence of intermediate cancer cells (iCCs) with characteristics that presumably lie between CDCs and CSCs. The cell fluorescence distribution is reproducible from one experiment to another, although weights of each sub-population may slightly vary (***Figure 1—figure Supplement 1***A). Similar results are obtained with another cell line, MDA-MB-231 (***figure Supplement 2***A), for which the fluorescence distribution is also captured by the sum of three distinct probability density functions centered on low, intermediate and high fluorescence intensity. The intersection between the fitted probability density functions (PDF) defines thresholds that are further used to approximate the fraction of cells in the respective phenotypic states, CDC, iCC, or CSC.

Finally, our setup allows time-lapse imaging of live cells (***Figure 1***E and***Figure 1—video 1***). Unperturbed SUM159-PT cells are tracked for up to 5 days at an acquisition rate of one frame every 45 min. Analyzing cell growth over time, we find that the growth rate is similar under our imaging system compared to cells grown in a standard incubator (***Figure 1***F). We also do not observed qualitative morphological differences between cells growing under imaging and in standard conditions. This indicates that imaging with our top stage incubator does note introduce major side effects one th cancer cell population. By means of cell segmentation and image processing, we recovered the time evolution of the three identified populations (***Figure 1***G). After a small drop during the first 30h, the fraction of CSCs remains constant ($\sim 5\%$) for the entire duration of the experiment. This observation is consistent with previous reports in breast cancer cells with our CSC reporter (***Bidan et al., 2019***) or using other CSC markers (***Lagadec et al., 2012***). On the other hand, CDC and iCC sub-populations are much more dynamic during a transient phase that last approximately 24 to 36h after which their proportions stabilize to $\sim 65\%$ for CDCs and $\sim 30\%$ for iCCs. This behaviour is

reproducible from one experiment to another (***figure Supplement 1***B). And we observe a similar trend with the other cell line, MDA-MB-231 (***figure Supplement 2***B).

Interestingly, we do not find a significant difference between the mean division rates (0.02 to $0.04h^{-}1$) of all three populations with SUM159-PT (***figure Supplement 1***C) and MDA-MB-231 (***figure Supplement 2***C) cell lines. Specifically, the division rate of CDCs does not significantly exceed that of iCCs during the transient phase. This suggests that the progressive establishment of steady proportion of cancer cell phenotypes relies on dynamics balance between cell-fate transition events, initially dominated by transitions from intermediate to differentiated phenotypes.

Point pattern analysis unveils spatial segregation of phenotypes with clusters of CSCs

To quantitatively characterize spatial heterogeneity, we employed functions from point pattern analysis (***Cressie, 2015***), a methodology recently applied to biological systems (***Parra, 2021***). Each cell is characterized by both its spatial coordinates (X,Y) and a specific phenotype (CDC, iCC or CSC (***Figure 2***A left panel). We estimated the empirical Point Correlation Function (PCF), denoted as $g(r)$, which is a function of a distance, r. The value of $g(r)$ is computed by counting events within a thin annulus of radius r with a thickness of $5\ \mu m$ (see Material and Methods). PCF measures the increase or decrease of the likelihood of finding an event at a distance r compared to what would be expected under complete spatial randomness (CSR). $g(r)$ greater than 1 indicates a more clustered pattern than CSR while a value smaller than 1 indicates a more dispersed pattern than CSR (***Figure 2***A, right panels).

We first computed the univariate PCF, for which cell phenotypes are not distinguished (***figure Supplement 1***). At very short distances ($r \leq 10\ \mu m$), the univariate PCF remains below 1, indicating that cells do not overlap in our experimental conditions. Subsequently, the function peaks at approximately 1.5 to 2 for radii of around 40 to $60\ \mu m$ (depending on the experiment), reflecting the close proximity of sister cells and the formation of micro-colonies. At larger scales, the PCF gradually approaches 1, indicating a complete spatial randomness.

To investigate the spatial correlation between phenotypes, we computed the bivariate PCF. For each phenotype, we screened the presence of the same or other phenotypes within its surroundings. Because the univariate PCF has a well defined shape reflecting tissue structure, we needed to disentangle properties due to putative relationship between phenotypes from the univariate spatial organization. To achieve this, any statistical test must account for the sample size to correctly estimate the confidence interval (***Cressie, 2015***). We thus compared the experimental bivariate PCF with the one obtained from phenotype shuffling in which cell fluorescence are randomly permuted (see materials and methods). Overall, no significant differences between shuffling and bivariate PCF are found for iCC versus iCC or CSC (***Figure 2***C and F). In contrast, CSCs (respectively CDCs) are more clustered with CSCs (respectively CDCs) compared with shuffling or complete spatial randomness (CSR) (***Figure 2***D and G). This effect is more pronounced with CSCs for which $g_{++}(r \approx 15\ \mu m)$ is about 20 times greater than CSR and 10 times greater than shuffling (***Figure 2***D). Fitting $g_{++}(r)$ to an exponential curve indicated a characteristic length of $\sim 50\ \mu m$ at the beginning of the experiment slowly increasing to $\sim 100\ \mu m$ after 4 days (***Figure 2—figure Supplement 2***). Finally, at short range ($r \sim 15\ \mu m$) both CSC and iCC are excluded from regions with a high density of CDC (***Figure 2***B and E). Again, the effect is more pronounced with CSC versus CDC for which $g_{+-}(r \approx 15\ \mu m)$ is found to $\sim$ 10 times smaller than CSR and 15 times smaller than shuffling (***Figure 2***B). This clustered pattern is observed at all time points (***figure Supplement 2***) and is reproducible from one experiment to another (***figure Supplement 3***). We also applied point pattern analysis to data collected for the other breast cancer cell line, MDA-MB-231, and similar results are found, though with weaker correlation (***figure Supplement 4***).

In summary, we observe a spatial patterning of cancer cell phenotypes, where cells given phenotype (low or high fluorescence intensities) tend to cluster with cells with the same phenotype. Notably, CSCs tend to form clusters of averaged diameter of $100\ \mu m$ (about 10 cell width). Inter-

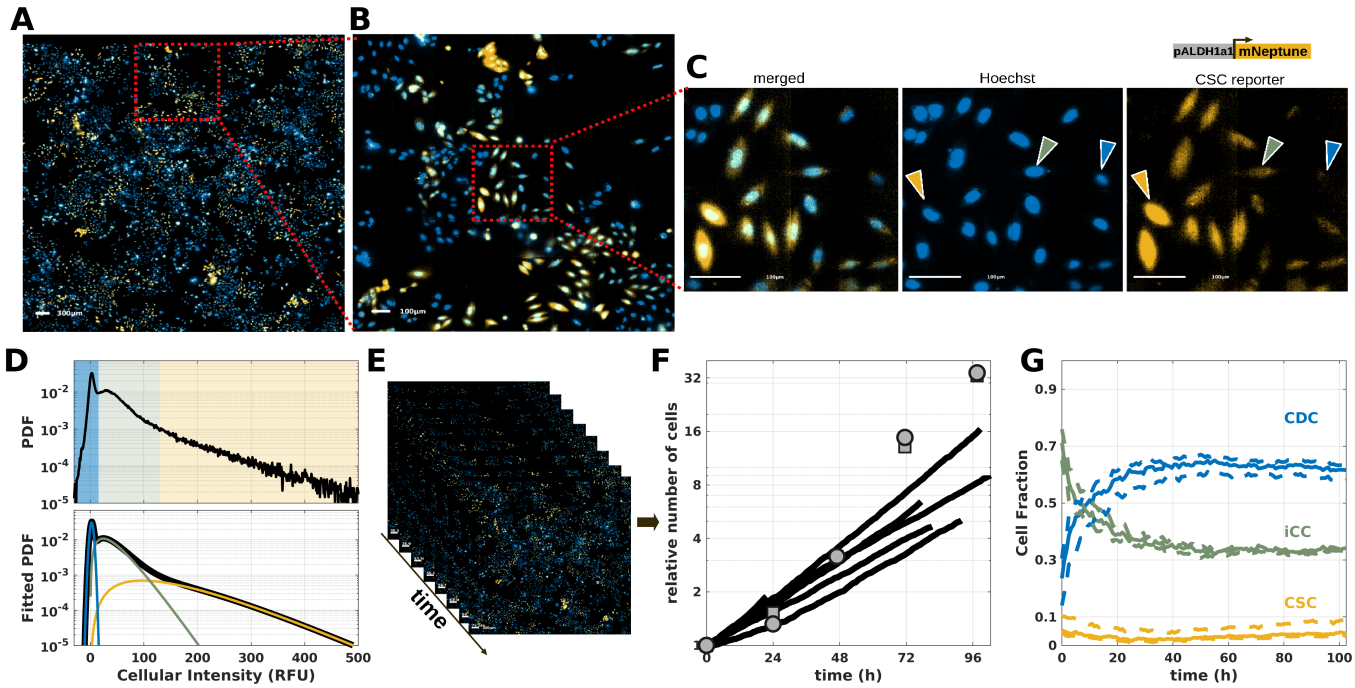


Figure 1. Ultra-wide field time-lapse imaging of CSC fluorescent reporter with single cell resolution.
(**A**-**C**) Representative epi-fluorescence images of SUM159-PT breast cancer cells stably transfected with the CSC reporter, pALDH1a1:mNeptune (***Bidan et al., 2019***). Hoechst (blue). CSC reporter (yellow). (**A**) Reconstructed ~6.8mm × ~6.8 mm field of view 72h after the beginning for a representative experiment. The full field of view is reconstructed by assembling 32 × 32 higher resolution images. Scale bar $300\mu m$. (**B**) A 4X zoom on a subset of the same image. Scale bar $100\mu m$. (**C**) A 16X zoom on a subset of the full width image. CSC reporter signal only (right), Hoechst signal only (middle) and merged Hoechst and CSC reporter (left). Scale bar $100\mu m$. The yellow, green and blue arrows respectively indicate high, medium and low fluorescent cells. (**D**) Upper panel : probability density function of single cell CSC reporter signal (extracted by means of image processing, see material and methods) for one representative experiment. Lower panel : fit of the PDF with a the sum of three probability distributions (black line, see material and methods). A normal distribution for cells with low intensities (blue, mean $\mu \approx 2.6$ and standard deviation $\sigma \approx 3.3$) and a gamma distribution for cells with intermediate intensities (green, shape parameter $k \approx 2.3$ and scale parameter $\theta \approx 18.2$) and another gamma distribution for cells with higher fluorescence level (yellow, shape parameter $k_+ \approx 2.7$ and scale parameter $\theta_+ \approx 55.7$). For all experiments, we define two fluorescence thresholds ($I_- = 15.5$ and $I_+ = 129.5$) to separate the three cell populations (see material and methods) : cells with signal lower than I_- are considered as Differentiated Cancer Cells (CDCs) ; cells with signal higher than I_+ are considered as Cancer Stem Cells (CSCs) ; and cells with signal between than I_- and I_+ are labelled as intermediate fluorescence Cancer Cells (iCCs). (**E**) Time-resolved analysis *via* time-lapse imaging where a full field of view (~6.8mm × ~6.8 mm) is acquired every ~ 45 min. (**F**) Time evolution of the relative cell number for 5 different experiments under the microscope (black lines) or grown in standard conditions with (gray squares) or without (gray circles) Hoechst staining. The growth curve is fitted (not shown) with an exponential function and the mean doubling time is between 27 and 36h. (**G**) Time evolution of the cell fraction of the three phenotypes (CDC blue line, iCC green line and CSC yellow line) defined by the intensity thresholds I_- and I_+. Dashed lines shows a sensitivity to thresholds (I_- and I_+) analysis. Time evolution of proportions of each phenotype is shown in dashed lines for $12.5 \leq I_- \leq 17.5$ and $84.5 \leq I_+ \leq 148.5$.

Figure 1—video 1. Representative time-lapse of SUM159-PT breast cancer cells stably transfected with the CSC reporter

Figure 1—figure supplement 1. Repeatability of fluorescence distribution and time evolution in SUM159-PT cells

Figure 1—figure supplement 2. Repeatability of fluorescence distribution and time-evolution in MDA-MB-231 cells

mediate phenotypes does not show significant spatial correlation with other CSCs nor iCCs, but display negative correlation with CDCs, thereby suggesting that iCCs are likely to transit between CDCs and CSCs at the interface of CSC-dense and CDC-dense areas.

Machine learning applied to single-cell time series identifies phenotypic transitions such as CSC reprogramming events

Because proliferation rate is similar for all phenotypes, the robust establishment of a specific proportion of different phenotypes throughout the population is expected to solely rely on dynamic transitions between stem, intermediate and differentiated states. It is questionable whether these transitions mainly occur during the initial transient phase to establish a steady proportion or throughout the entire course of steady population growth. To address these issues, we used unsupervised partitioning analysis of single-cell time series (SCTS) to, without prior knowledge of time scales or specific shapes, identify distinctive temporal patterns of differentiation or reprogramming events. Automated tracking of cell trajectories, detection of division events and single-cell fluorescence extracted by means of image processing (see materials and methods) allow us to reconstruct single-cell time series (SCTS). Then, as depicted on ***Figure 3***A, cells for which the full cell cycle could be monitored were selected for further analysis. For SUM159-PT cells, we collected 19,620 SCTS from 5 different experiments. We used cell-cycle progression (time relative to the cell cycle duration) as a metric for time evolution and each time series is re-sampled so that all SCTS share the same number of points along cell cycle progression. Euclidean distance is used as a di-similarity measurement and k-medoids (***Kaufman, 1990***) is used to partition SCTS and the GAP statistics (***Tibshirani et al., 2001***) is used to determine the relevant number of clusters. This procedure results in 16 clusters. However the last cluster (#16) contains SCTS from only 2 experiments, while all other clusters cover all experiments (***figure Supplement 1***A and B). This last cluster was excluded from further analysis and we finally considered 15 clusters (***Figure 3***) with 99.9% of the SCTS clustered. The resulting clustering did not show significant bias for any experiment (***figure Supplement 1***A and B).

Partitioning analysis effectively captures the phenotypic diversity as the procedure results in SCTS categorized based on their average fluorescence signal (***Figure 3***B). More generally, clustered SCTS could be classified into two main categories : approximately 70% of cells exhibiting a stable signal (***figure Supplement 2***B), while approximately 30% of the cells within a transiting state undergoing an increase toward higher fluorescence (***figure Supplement 2***A). It is important to note that these transitions occur gradually throughout the cell cycle, without any abrupt changes. We obtained similar results for the other cell line MDA-MB-231 (***figure Supplement 3***) for which we found approximately 50% transitioning SCTS. However we note that transitioning SCTS clusters of MDA-MB-231 also include SCTS with decreasing fluorescence ((***figure Supplement 3*** clusters #14, 7 and 8).

Notably, the partitioning analysis unveils clusters (#10 and 11) with ascending signal and for which cells transition from a fluorescence level close to the CDC threshold to a fluorescence level close or above CSC threshold (***Figure 3***B and ***figure Supplement 2***A). Clusters #10 and 11 comprise 4.8% of the analyzed time series (***Figure 3***B) and are not limited the transient phase at the beginning of the experiment (***figure Supplement 1*** C). This last observation support the idea that steady proportion of phenotypes arises from dynamic conversions throughout experiment time course. More specifically, among cells within ascending clusters, we found representative instances of single cell transitioning from CDC to CSC (***Figure 3***C). Interestingly, progenitors from the same mother cell can have very different fates. Indeed, the sister of the reprogramming cell depicted in ***Figure 3***C does not exhibit transition nor significant fluorescence variation (***Figure 3***C left panel).

To qualitatively visualize whether phenotype is transmitted to the cell progeny, we plot the clustered SCTS together with the mother SCTS and daughters SCTS when available (***figure Supplement 4***). While the fluorescence level of a mother seems to be preserved for its daughter, cell division nevertheless introduces some variability, which will be studied more systematically in next

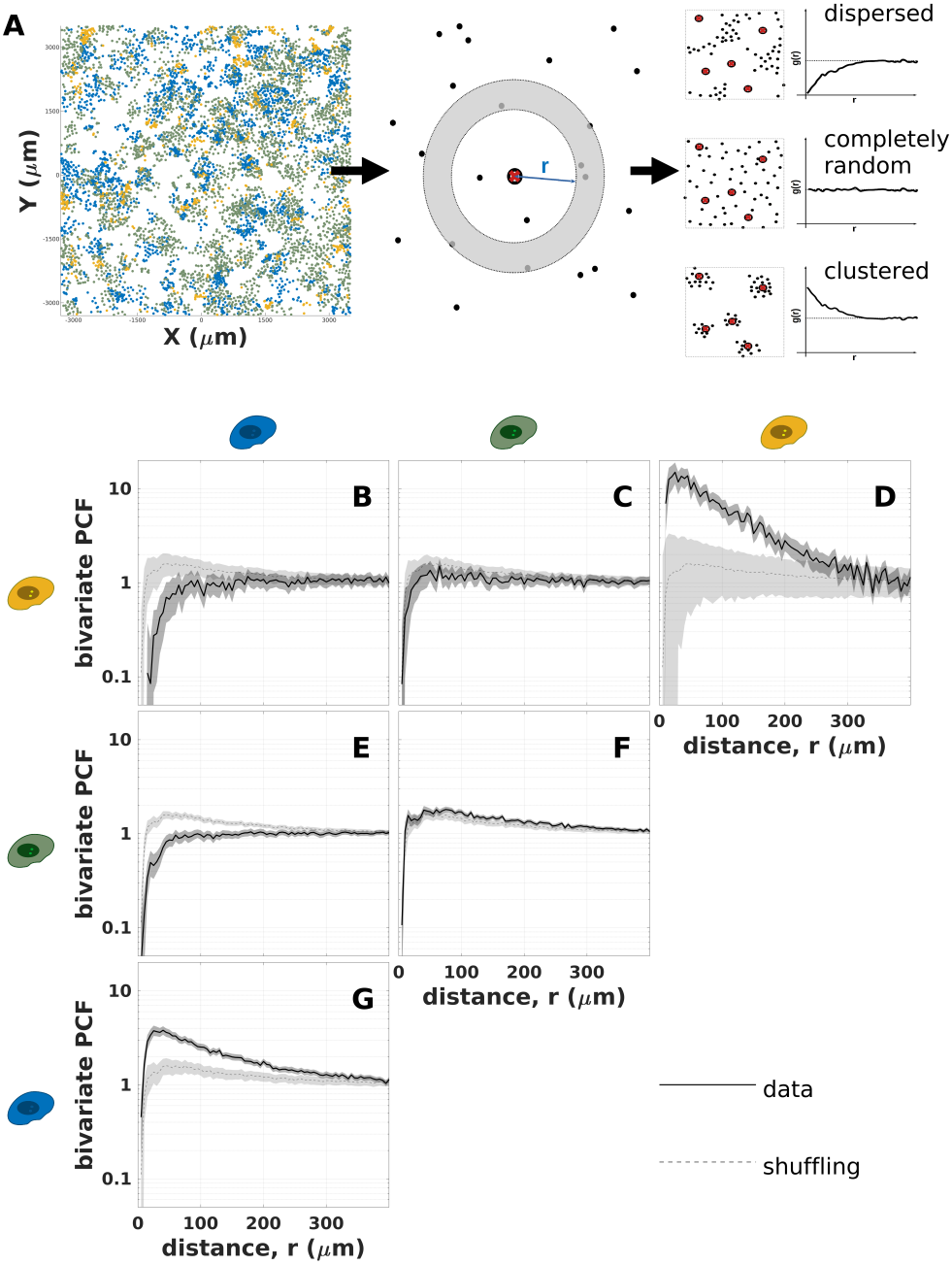


Figure 2. Statistical analysis of the spatial distribution of SUM159-PT phenotypes.
(**A**) Workflow and interpretation of Point Pattern Analysis (see Material and Methods for details). For a given experiment and time point (here 90h), detected cells are characterized by their spatial coordinates (X,Y) and phenotypes (CDC in blue, iCC in green or CSC in yellow). The corresponding Point Correlation Function (PCF, $g(r)$) is computed. PCF measures the increase or the decrease of the likelihood of finding an event at a distance r compared to what would be expected under complete spatial randomness (CSR). $g(r)$ greater than 1 indicates a more clustered pattern than CSR while a value smaller than 1 indicates a more dispersed pattern than CSR. (**B**-**G**) Bivariate PCF, $g_{xy}(r) = g_{yx}(r)$, for a representative experiment 90h after beginning of imaging where x and y represent cell phenotypes (CDC, iCC or CSC). The dark lines are the measured PCFs (data). The gray shaded area is the 99% confidence interval obtained from bootstrap resampling (sensitivity analysis, see Material and Methods). The dashed gray lines are the PCF obtained from phenotype shuffling (control). The gray shaded area is the 99% confidence interval obtained from 1500 repeats of shuffling. (**B**) Bivariate PCF measuring spatial correlation between CSC and CDC. (**C**) Bivariate PCF measuring spatial correlation between CSC and iCC. (**D**) Bivariate PCF measuring spatial auto-correlation of CSC. (**E**) Bivariate PCF measuring spatial correlation between iCC and CDC. (**F**) Bivariate PCF measuring spatial auto-correlation of iCC. (**G**) Bivariate PCF measuring spatial auto-correlation of CDC.

Figure 2—figure supplement 1. Univariate (regardless of phenotype) Point Correlation Function of SUM159PT cells

Figure 2—figure supplement 2. Time evolution of bivariate Point Correlation Functions of SUM159-PT cells

Figure 2—figure supplement 3. Repeatability of bivariate Point Correlation Function of SUM159-PT cells

Figure 2—figure supplement 4. Repeatability of bivariate Point Correlation Function of MDA-MB-231 cells

section. Remarkably, transiting SCTS are not limited to the transient phase (0 to 24/36h) for which we observed exchange between iCC and CDC (***Figure 1***G). These reprogramming SCTS are also detected during the second phase of the experiment when steady state is already reached at the population level (***figure Supplement 1***C). This strongly suggests that a steady proportion of CSC fraction is maintained through a dynamic equilibrium between reprogramming and differentiating trends.

Lineage analysis reveals phenotypic inheritance through symmetric division

To investigate how cell cycle influences phenotypic changes, we examined the fluorescence variations across generations. ***Figure 4***A shows an example of lineage reconstruction with a depth of three generations. From a mother cell, we kept track of its 2 daughters which gave rise to 4 grand-daughters for which we could detect 6 out of 8 great grand-daughters.

We first consider fluorescence correlation between mothers and daughters. To do so, we compared the fluorescence intensity measured at the end of the mother's cycle with the one measured at the beginning of the daughter's cycle (see materials and methods). From the 5 experiments with SUM159-PT cells, we could retrieve 37,582 of such relationships. As expected daughter's fluorescence strongly correlates with the one of its mother (***Figure 4***B center panel). The determination coefficient, which measures the fraction of variance explained by linear correlation, is $R^2 = 0.70$. Indicating that 70% of the daughter's signal variance is explained by a linear relationship with mother's signal. However, fluorescence variation during mitosis is systematically biased toward negative values, indicating that cell mitosis may coincide with chromatin changes that could impact either ALDH1A1 promoter transcription activity or protein degradation (***Alber et al., 2018***). Relative fluorescence variation indeed shows a bell shape distribution centered around -0.25 (right panel of ***Figure 4***B). This behaviour is reproducible considering each experiment separately (***figure Supplement 1***A). Repeating the same analysis for MDA-MB-231 cells (***figure Supplement 3***A) shows a weaker but significant correlation between mother and daughter with $R^2 = 0.39$. Importantly, divisions are mostly symmetric (see ***figure Supplement 2***). Indeed, fluorescence of two sister cells are strongly correlated ($R^2 = 0.66$ for SUM159PT cells).

We then revisited the measurements of fluorescence variations during cell cycle by comparing signal measured at the beginning of cell cycle with the one measured at the end (center panel of ***Figure 4***C). We used the 19620 SUM159PT cells for which the full cell cycle is monitored. The determination coefficient is $R^2 = 0.59$, indicating a weaker correlation between these two signals compared to the correlation between mother and daughter. As opposed to signal variation during mitosis, fluorescence variation during cell cycle is distributed around zero but with a tail biased toward higher values (right panel of ***Figure 4***C). Indeed, $\sim 30\%$ of the cells at least double their fluorescence level. These data are in agreement with the time series clustering from which we found $\sim 30\%$ of cells in transiting clusters. This behaviour is reproducible considering each experiment separately (***figure Supplement 1***B). Analysis of MDA-MB-231 reveals higher variability of phenotype during cell cycle with $R^2 = 0.16$ (***figure Supplement 3***B).

Finally, to better understand how the fluorescent phenotype is transmitted across generations, we computed the determination coefficient between the fluorescence intensities of the mother cell and its progeny of first, second and third generation (***Figure 4***D). The fluorescence level is measured at the beginning of cell cycle for each generation. As expected, correlation decreases as generation increases with a determination coefficient of $R^2 = 0.51$ for generation 1, $R^2 = 0.32$ for generation 2 and $R^2 = 0.25$ for generation 3. Repeating the same analysis for MDA-MB-231 cells we found similar behavior but with weaker correlation between a cell and its daughters ($R^2 = 0.16$). Interestingly, the correlation between a cell and its progeny is found to be stronger than in case of a memory-less chain of processes (***Figure 4*** D). In such a memory-less model, phenotypic inheritance upon division only depends on the state (fluorescence, I_m) of the mother cell and, similarly, fluorescence variation during cell cycle ($I_2 - I_1$) only depends on fluorescence at the beginning of cell cycle (I_1) (see material and methods). The observed correlation between a cell and its progeny prompted

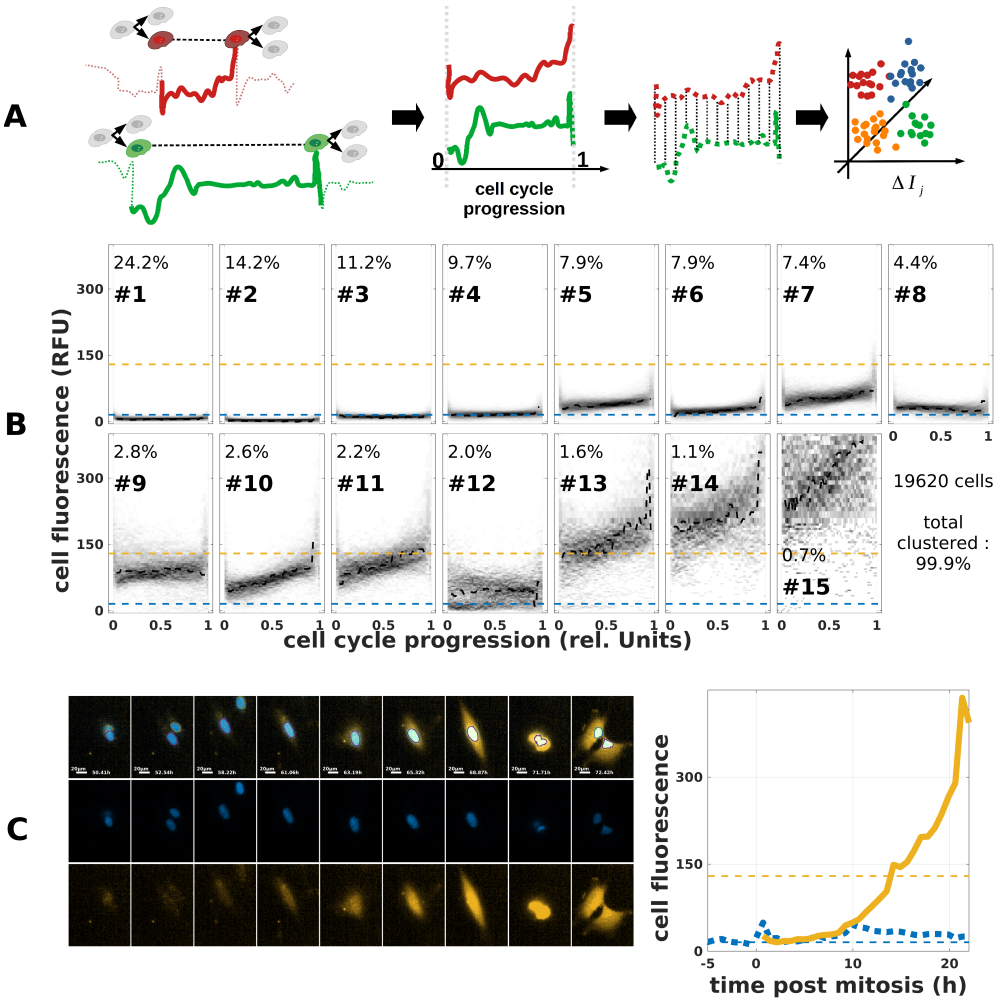


Figure 3. Fluorescent signal dynamics at single cell level.
(**A**) **Schematic representation of single cell time-series analysis:** Selection of cells for which signals is recorded throughout a whole cell cycle. Time series are (*i*) resampled to 60 points using cell cycle progression as a time metric for time, (*ii*) compared using a Euclidean distance, and (*iii*) partitioned around 16 medoids. (**B**) **Results of single cell time-series clustering for SUM159PT cells :** Clusters are numbered by increasing size. The percentage of cells within each cluster is indicated. Are shown scatter plot of each time-series within each cluster. (**C**) Example of a cell with a significant increase of fluorescent signal associated to de-differentiation into CSC. Left panel : time-lapse images (scale bar $20\mu m$). bottom : m-Neptune fluorescence ; middle Hoechst and top : merged. Right panel : Time evolution of fluorescence signal of the cell (yellow trace) and the one of its mother cell (blue dotted) and sister cell (blue dashed).

Figure 3—figure supplement 1. Meta-analysis of identified SCTS clusters

Figure 3—figure supplement 2. Transitionning and non-transitionning SCTS clusters in SUM159PT cells

Figure 3—figure supplement 3. Transitionning and non-transitionning SCTS clusters in MDA-MB-231 cells

Figure 3—figure supplement 4. SCTS across generations in SUM159PT cells

us to hypothesize a coupling between fluorescence variation during cell cycle and signal decrease after mitosis. We indeed find a weak correlation between ΔI_{21} and ΔI_{dm} ($R^2 = 0.35$). However, such relationship is fully explained by simulation of a memory-less chain ($R^2 = 0.36$), ruling out the coupling hypothesis.

Altogether, these data highlight a significant influence of the cell division events on the change of fluorescent signal. Though phenotypic inheritance is observed, the broad distribution of the relative signal increase during the cell cycle reflects the probabilistic nature of differentiating and reprogramming events. In contrast, the relative decrease of signals occurring during mitosis pertains all cells. The strong correlation between a cell and its progeny (that is found to be stronger than in case of a memory-less chain model) suggests the existence of a hidden mechanism that maintain phenotype across cell cycle and upon division.

Contribution of phenotypic inheritance to the spatial clustering of CSCs

A mechanistic hypothesis to explain the spatial clustering of CSC (***Figure 2***B-G) relies on the phenotypic inheritance described above (***Figure 4***D). Divisions where sister cells display similar fluorescence intensities would lead to create spatial correlation due to the proximity of daughter and mother cells. The patterning role of this mechanism depends on the degree of signal correlation between sister cells as well as on the characteristics of cell motility. We need a statistical test to estimate to which extent cell displacement and lineage can solely explain the clustered pattern associated with bivariate PCFs (***Figure 2*** B-G). To guarantee that the test accurately account for observed cell motility, we used cell trajectories reconstructed from experimental data, while fluorescence variation is obtained from Monte Carlo simulations. At initial time, cells are attributed the fluorescence signal measured from experiment. Time evolution of fluorescence is assumed to be driven by two independent processes (see materials and methods) : transitions occur (*i*) upon division (***Figure 4***B) and (*ii*) during cell cycle (***Figure 4***C).

On the one hand, simulations reproduce correctly the time evolution of each phenotype at the population level (***Figure 5***B). After a transient period of $\sim$ 24 to 36 hours, a steady state is reached with proportions of each phenotype comparable to experimental data. On the other hand, we find a determination coefficient between mother and offsprings of first, second and third generation weaker than the one measured on experimental data (***Figure 5***C). Similar results are obtained for MDA-MB-231 (***figure Supplement 2***).

Importantly, simulations of SUM159PT cells fail to reproduce the spatial clustering of CSC (***Figure 5***E). For instance, the maximum bivariate PCF, g_{++}, (found at $r \approx 15\ \mu m$), is only 3 to 5 times greater than shuffling for all Monte-Carlo simulations, whereas the experimental value is 10 to 25 times higher than shuffling (***Figure 5***E and ***figure Supplement 1***C). This behaviour is reproducible from one experiment to another (***figure Supplement 1***C) with p-values below 0.01 for all experiments. Moreover, simulations of SUM159PT cells fail to explain mutual exclusion of CSC and CDC (***Figure 5***D and ***figure Supplement 1***A) or mutual exclusion of iCC and CDC (***figure Supplement 1***D). Indeed, bivariate PCF, $g_{+-}(r \approx 15\ \mu m)$, is only 2 fold lower than shuffling for all Monte-Carlo simulations, while it is more than 10 fold lower for all experiments. Simulations results are more mitigated for MDA-MB-231 cell line for which we found more variability from one experiment to another (***figure Supplement 3***A-I). We, however, note that simulations fail to reproduce mutual exclusion of CSC and CDC (***figure Supplement 3***A and I) with p-values below 0.05 for all experiments.

Altogether, these data suggest that phenotypic inheritance of progeny that tends to be colocalized is not enough to explain the observed spatial pattern, probably because of cell motility and the progressive loss of phenotypic memory. Another mechanism is therefore required to promote the formation of CSC clusters, which presumably involves phenotypic transitions driven by environment-specific cues.

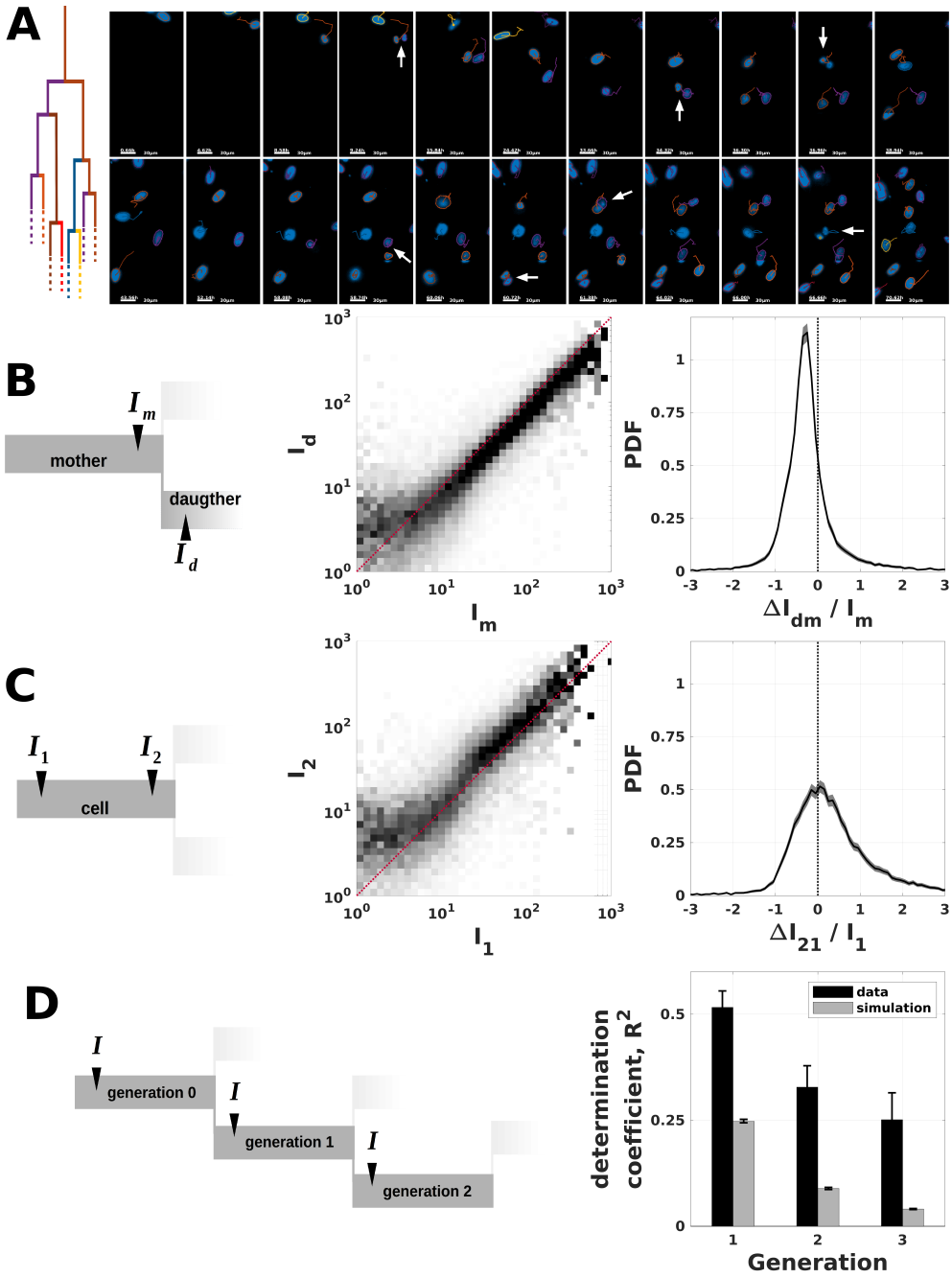


Figure 4. Statistical analysis of phenotypic inheritance of SUM159-PT cells
(**A**) **Representative example of cell lineage:** Hoechst images of a mother cell for which we kept track of its 2 daughters ; 4 grand-daughters and 6 out of 8 great grand-daughters. Segmentation masks and cell trajectories are overlaid. Arrows indicate cell division. Scale bar $20\mu m$.
(**B**) **Correlation of CSC reporter signal between mother and daughters.** Left panel : Mother fluorescence intensity, I_m, is measured at the end of cell cycle and daughter, I_d, fluorescence intensity is measured at the beginning of cell cycle. See materials and methods for details. Middle panel : scatter plot of daugther's fluorescence as a function of mother's fluorescence. Right panel : Probability density function of fluorescence variation upon mitosis, $\Delta I_{dm} = I_d - I_m$, relative to mother's fluorescence, I_m. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods). (**C**) **Correlation of CSC reporter signal between beginning and end of cell cycle for the same cell.** Left panel : Fluorescence intensity is measured both at the beginning of cell cycle (I_1) and at the end (I_2). See materials and methods for details. Middle panel : scatter plot of I_2 as a function of I_1. Right panel : Probability density function of fluorescence variation during cell cycles, $\Delta I_2 = I_2 - I_1$, relative to fluorescence at beginning of cell cycle, I_1. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods). (**D**) **Correlation of CSC reporter signal across generations.** Left panel : Fluorescence intensity, I, is measured both at the beginning of cell cycle for each cell. See materials and methods for details. Right panel : Determination coefficient between signal of a mother cell and signal of its progeny as a function of generation (1 : daugthers ; 2 : grand-daughters and 3 : great grand-daughters). Black bars correspond to data and gray bars correspond to numerical simulation of a memory-less chain model (see materials and methods). Error bars represent the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods).

Figure 4—figure supplement 1. Repeatability of fluorescence variations during cell cycle and upon mitosis in SUM159PT cells

Figure 4—figure supplement 2. symmetric and asymmetric division rates of SUM159PT cells

Figure 4—figure supplement 3. Statistical analysis of phenotypic inheritance of MDA-MB-231 cells

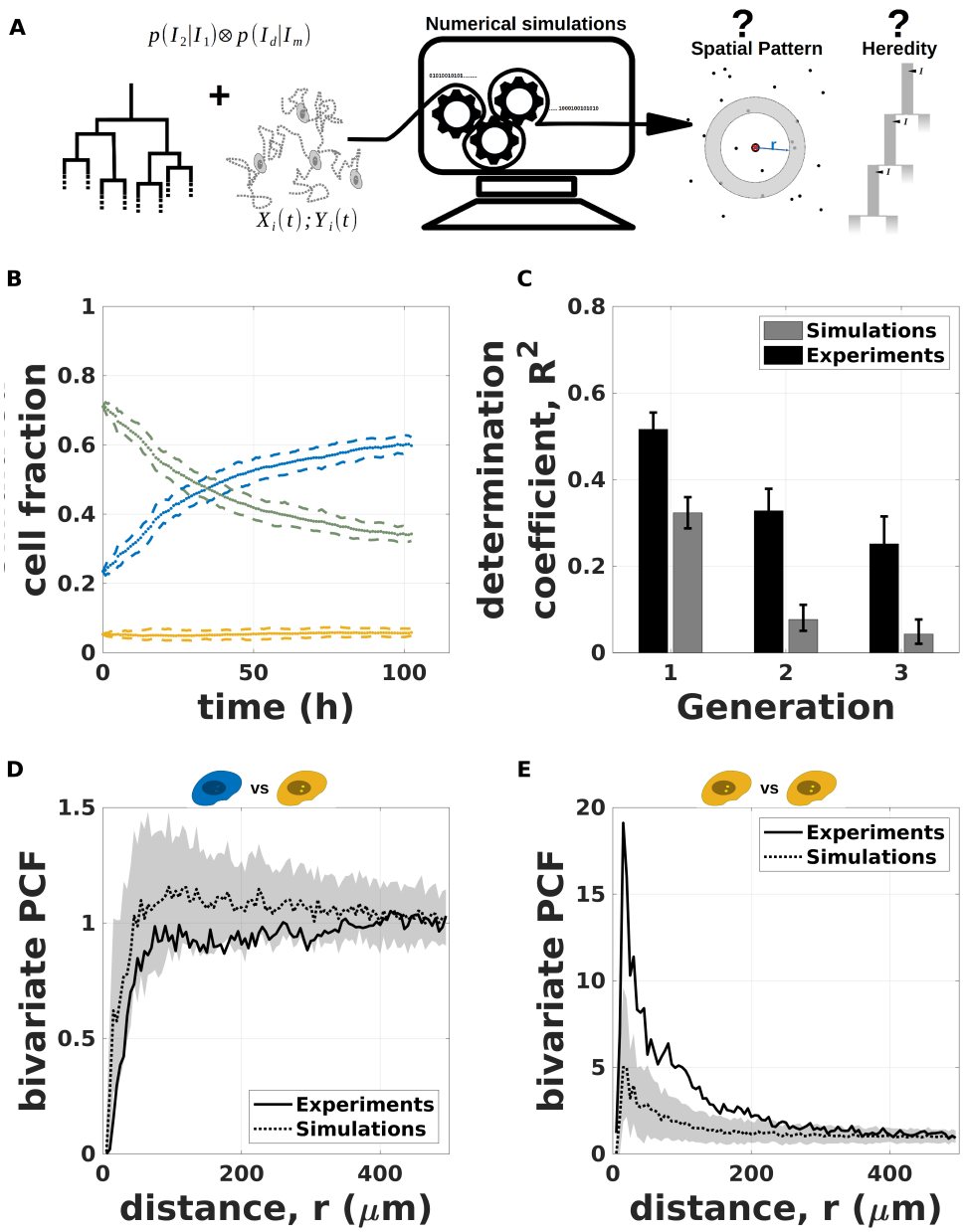


Figure 5. Spatio-temporal simulations of phenotypic inheritance of SUM159PT cells: (**A**) **Simulation workflow :** Cell lineage and cells trajectories are known from experimental data. Probability density function for fluorescence variation are extracted from experimental data. Monte Carlo method is used to simulate fluorescence variation, while cell trajectories are kept identical to experiments. Relevant quantities such as PCF or phenotypic inheritance are computed as for experiments. (**B**) **Simulated time evolution of the cell fraction of the three phenotypes** (CDC blue , iCC green and CSC yellow) defined by the intensity thresholds I_- and I_+. Mean (points) and 99% confidence interval (dashed lines) obtained from 500 independent simulations based on trajectories of a representative experiment (same as figure ***Figure 1***). (**C**) **Simulated correlation of fluorescence signal across generations.** Is shown determination coefficient between signal of a mother cell and signal of its daughter. Black bars correspond to data and gray bars correspond to numerical simulation (see Material and Methods). Error bars represent the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see Materials and Methods). (**D**) **Simulated bivariate PCF for CSC versus CDC.** Continuous black line represents experimental PCF (same as figure ***Figure 2***). Dashed line is the PCF of simulated data for the same experiment. The gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see Material and Methods). (**E**) **Simulated bivariate PCF for CSC versus CSC.** Continuous black line represents experimental PCF (same as figure ***Figure 2***). Dashed line is the PCF of simulated data for the same experiment. The gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see Material and Methods).

Figure 5—figure supplement 1. Repeatability of spatio-temporal simulations with SUM159PT cells

Figure 5—figure supplement 2. Spatio-temporal simulations of phenotypic inheritance of MDA-MB-231 cells

Figure 5—figure supplement 3. Repeatability of spatio-temporal simulations with MDA-MB-231 cells

Contribution of intercellular coupling to phenotypic transitions and CSC reprogramming

Motility and division processes are not sufficient to fully explain the spatial patterning of cancer cell phenotypes. Moreover, phenotypic inheritance across several generations is stronger than expected for memory-less chain (***Figure 4***D). We hypothesize that a spatial control of phenotypic transitions must therefore occur and promote a clustered cell distribution. CSCs display the most pronounced spatial correlation. This pattern, reminiscent of a niche-like effect, is likely to involve cell-to-cell communication.

Figure 6—video 1 A and B qualitatively shows how CSC niche is progressively established from slow fluorescence increase and phenotypic inheritance. We note that fluorescence is stabilized in cells surrounded by CSC (in the middle ***Figure 6—video 1*** B) of while cells in the vicinity of CDCs seem to exhibit lower signal (upper right of ***Figure 6—video 1*** B). More specifically, focusing on two different lineage descending from the same mother cell (***Figure 6—video 1*** C) indicates that cell fate may strongly depends on environment.

To investigate the contribution of inter-cellular signaling, we examined fluorescence variation with respect to a quantity reflecting the cell environment. To do so, we computed fluorescence variation conditional to both phenotype and the local spatial density of a given phenotype (see Material and Methods). We considered both the relative density of CSC, f^+ and the relative density of CDC, f^-.

We first focus on signal variation during cell cycle (ΔI_{21}) conditional to spatial environment. Regions of high fraction of CSC phenotypes tend to promote signal increase in iCC and CSC , thus promoting CSC maintenance and iCC reprogramming into CSC (***Figure 6***A). For instance, for cells classified as CSC at beginning of cell cycle, $< \Delta I_{21} | f^+ >$ is negative (differentiation) when $f^+ < 0.05$ while it increases by 50% (reprogramming) when $f^+ > 0.9$. A consistent trend is obtained in regions with low CDC fraction, except that the signal increase associated to reprogramming extends to CDC phenotypes (***Figure 6***B, lower panel). For CDC, $< \Delta I_{21} | f^- >$ value increased by 50% $f^- < 0.1$ while the value drops close to zero for $f^- > 0.9$. The MDA-MB-231 cell line also exhibits fluorescence variations that depend on the local cellular environment (***figure Supplement 2***). In sharp contrast, fluorescence variations upon mitosis (ΔI_{dm}) show no significant dependency on the local environment (***figure Supplement 1***), consistently with a more peaked distribution observed in ***Figure 4***. Similar results are obtained for the MDA-MB-231 cell line (***figure Supplement 3***). Altogether these results highlight a central role cell-cell communication in spatial patterning and phenotypic inheritance across several generations.

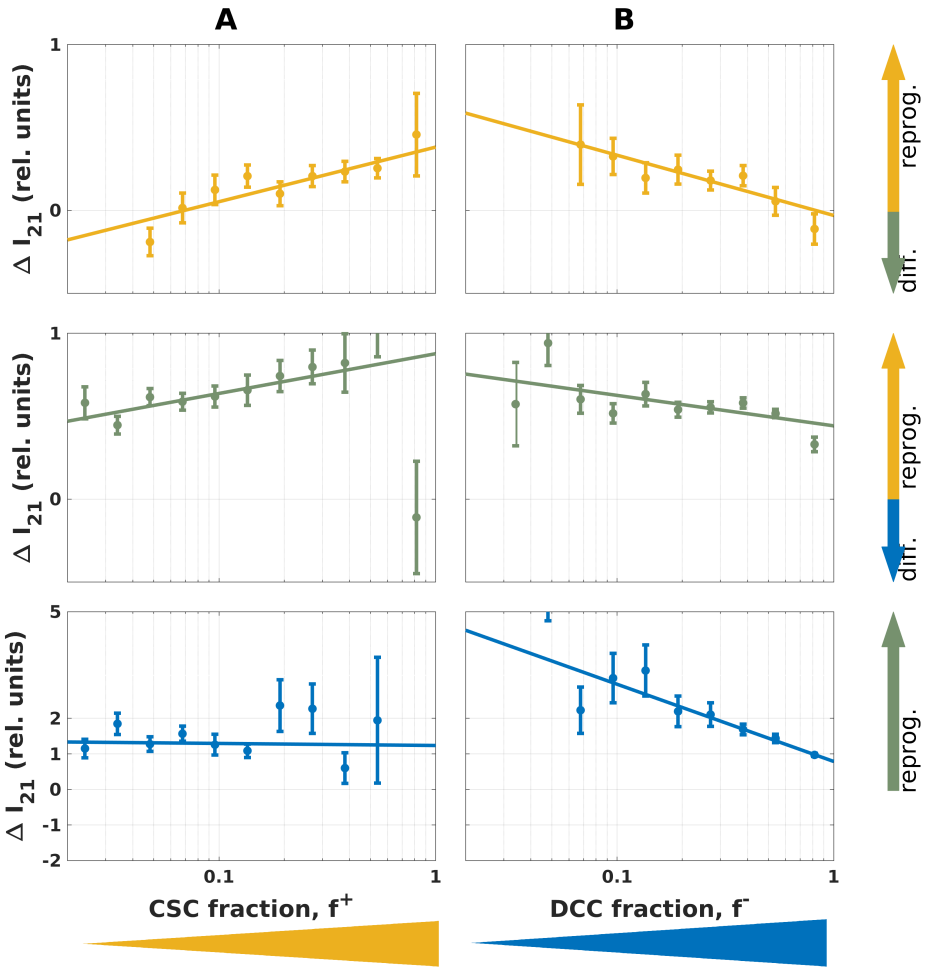


Figure 6. Influence of local environment on fluorescence variation during cell cycle in SUM59PT cells.
Average fluorescence variation during cell cycle, ΔI_{21} as a function of the local CSC fraction, f^+, (**A**) or the local CDC fraction, f^-, (**B**). From top to bottom, data are shown as function of the initial cell phenotypes : CSC (yellow), iCC (green) and CDC (blue). ΔI_{21} is normalized to population average fluorescence intensity of CSC (~ 150 RFU), iCC (~ 42 RFU) and CDC (~ 2.5 RFU). Positive values for ΔI_{21} indicate differentiation and negative values indicate reprogramming. Each point represent conditional mean and the height of error bars two standard deviations.

Figure 6—video 1. Time lapse showing progressive establishment of CSC niche

Figure 6—figure supplement 1. Influence of local environment on fluorescence variation upon mitosis in SUM59PT cells

Figure 6—figure supplement 2. Influence of local environment on fluorescence variation during cell cycle in MDA-MB-231 cells

Figure 6—figure supplement 3. Influence of local environment on fluorescence variation upon mitosis in MDA-MB-231 cells

Discussion

Intra-tumor heterogeneity and cancer stem cells (CSCs) have been recognized to play a role in tumor resistance and invasiveness. Effectively targeting CSC for treatment requires a comprehensive understanding of their multifaceted and intricate plasticity. Despite efforts to characterize the transcriptomic and metabolic profiles of CSCs (REFs), their self-organizing capacity to spontaneously reform stem-like cell niches and spatially-heterogeneous tumors remains poorly understood. In this study, live cell tracking of cultured breast cancer cells marked with a stemness fluorescent reporter enables us to link phenotypic transitions at single-cell level with the emergence of spatiotemporal organization of phenotypic heterogeneity at the population level. ALDH1A expression/activity is commonly used as a CSC marker where high expression is associated to an hybrid Epithelial/mesenchymal CSC phenotype (***Grosse-Wilde et al., 2015***; ***Colacino et al., 2018***). In our experiments, careful analysis of fluorescent intensity distribution reveals two well-distinct populations characterized with high and low intensities, and therefore classified as CSC-like and CDC-like phenotypes. Interestingly, a population with intermediate level of reporters is also observed and possibly associated with a metastable transiting phenotype or manifesting the existence of diverse non-stem cell types (***Colacino et al., 2018***).

The phenotype organization suggests a fluid hierarchical arrangement of differentiated states, displaying nevertheless asymmetry between differentiation and dedifferentiation events. Indeed, live cell tracking of fluorescent signal provides insights into the temporal characteristics of both differentiation and dedifferentiation transitions. Temporal resolution of phenotypic transitions is constrained by the time interval of image recording (1 hour) and the degradation timescale of the fused fluorescent proteins (~ 10h). Given these resolution limitations, dedifferentiation events can be clearly identified during the cell cycle over the course of about ten hours. The timescale of differentiation events are more difficult to extract from signal noise through clustering analysis restricted to one cell cycle. Instead, signs of signal decrease are found systematically just after mitosis which can be due to chromatin condensation and a global transcription arrest during mitosis (***Hsiung et al., 2016***) or to a symmetric division with accumulation of differentiation factors (***Morrison and Kimble, 2006***). This suggests that differentiation occurs more progressively at time scales much larger than the cell cycle, or through intermediate or primed states (***Sha et al., 2019***; ***Pfeuty et al., 2018***). This is overall in good agreement with cell sorting data. Indeed, after selecting only CSC, steady proportions are restored after a long timescale (more than 10 days), while selecting only CDC, new steady proportions are established after only a few divisions (REF thèse justine). Interestingly also, while cancer cell reprogramming into stem-like cells has been previously shown to actively occur in the contexts of radiotherapy or chemotherapy (***Lagadec et al., 2012***; ***Auffinger et al., 2014***), the present work highlights a significant number of reprogramming events found to occur in an unperturbed cell population. A balance between reprogramming and differentiation events maintains a dynamic equilibrium between heterogeneous phenotypes at the population level.

The advantage of being able to track the phenotypes of a cancer cell population in space and time is to infer the system-level self-organizing mechanism resulting in spatial pattern of cell types. General principles for spatial self-organization relies on diverse cellular processes such as signaling feedbacks, motility, division, where each process involves deterministic and stochastic contributions (***Kicheva et al., 2012***; ***Grace and Hütt, 2015***; ***Landge et al., 2020***). Spatio-temporal point pattern analysis of cell phenotypes clearly identifies clusters of stem-like cells with a typical size of 150 to 300 μm. The formation of CSC clusters and the spatial exclusion with CDCs are found to originate from the interplay of phenotypic inheritance and signaling cues provided by neighboring cells. All these observations were found in two distinct breast cell lines, SUM159-PT and MDA-MB-231. In both cell lines, symmetric division constitutes a local positive feedback mechanism that is prone to generate clusters of cells with similar phenotypes, while cell-cell interactions constitute population-level feedback that shape and stabilizes long-range spatial structures. In contrast,

the contribution of cell motility in cancer cell cultures is not a morphogenetic cell-sorting process (***Strandkvist et al., 2014***) but rather a source of spatial noise mixing cell phenotypes. Intracellular noise and cell motility thus provide multiple sources of stochasticity influencing the level of spatial correlation and the nature of the patterns. Variation of fluorescent reporter expression and associated noise together with different motile behaviour between cell lines might contribute to the less pronounced spatial segregation and influence of cell to cell interactions found in MDA-MB-231 compared to SUM159-PT.

Can mechanistic signaling insights be inferred from the spatial correlation of phenotypes and phenotypic changes ? The clustering of CSC phenotypes is much more marked than for other phenotypes and the reprogramming events are promoted by neighboring CSC cells, which altogether supports a lateral induction mechanism. Lateral induction is a common cell-cell interaction mechanism that can typically generate spatial cell-fate patterns with wavelength of a dozen of cells (***Owen et al., 2000***; ***Sjöqvist and Andersson, 2019***). In the present case, CSC clustering is observed up to $300\mu m$ range. Given a typical diffusivity of $D \approx 100\ \mu m^2 s^{-1}$ for a soluble signaling factor. Such a scale is rather compatible with paracrine signaling with a ligand lifetime of 5 to $15 min$ (***Handly et al., 2015***). In contrast, the mutual exclusion between CSC and CDC is observed at a much shorter length scale of 25 to $50\ \mu m$ which rather supports juxtacrine signaling or paracrine signaling with a ligand of short lifetime or diffusibility. Such a short lifetime is nevertheless unlikely given the low cell density and the geometry of the culture dish in which cells are plated on a surface in contact with a reservoir. The spatial pattern characterized with two very different length scales of the spatial correlations between phenotypes is thus proposed to reflect the involvement of both paracrine and juxtacrine mechanisms in intercellular interactions.

Notch-mediated juxtacrine signaling plays a crucial role in maintaining stemness in cancer stem cells (***Meurette and Mehlen, 2018***) where Jagged-Notch interactions have been proposed to be involved in the spatial segregation of an hybrid E/M phenotype at the interior of the tumor (***Bocci et al., 2019b***). Alternatively, self-renewal or differentiation of CSC has been proposed to be sensitive to a wealth of diffusible signaling factors, such as SHH ligands modulating hedgehog pathway (***Kim et al., 2013***), cytokines modulating JAK/STAT pathway (***Ruiz and Altaba, 2011***) or DKK1 ligand modulating WNT pathway (***Wu et al., 2022***).

A mathematical modeling approach would be valuable to test those diverse signaling hypothesis. In particular, naive mathematical models of spatiotemporal phenotypic dynamics in cancer cell population (***Olmeda and Ben Amar, 2019***) would allow to identify the contribution of diverse intercellular mechanisms to, respectively, the observed spatial patterns and the homeostatic establishment of cell-type fractions. Refining our understanding of the feedback mechanisms that empower cancer stem cells to rapidly reestablish intratumor heterogeneity holds promise for cancer therapy.

Materials and Methods

Cell culture, plasmid transfections, and generation of stable cell lines

Cell culture

The experiments are carried out on two breast cancer cell lines, SUM159PT and MDA-MB-231. SUM159PT cell line is obtained from Asterand and cultured in F12 Nut Mix media (Gibco) supplemented with fetal bovine serum (5%, FBS, PAN-Biotech), insulin (5 $\mu g/ml$), HEPES (10 nM, 15630080, Gibco), hydrocortisone (1 $\mu g/ml$) and Zell Shield. MDA-MB-231 cell line is obtained from ATCC and cultured in MEM media (Gibco) supplemented with 10% of fetal bovine serum (FBS, PAN-Biotech), 1X of non-essential amino acids (11140035, Gibco) and Zell Shield. Cells were maintained in a 5% CO2/air environment.

Plasmid Transfections

The pALDH1A1-mNeptune vector is construct with mNeptune fluorescent protein under control of ALDH1A1 promoter, a breast CSC marker, and is obtained as previously studied (***Bidan et al., 2019***). Briefly, the high-fidelity DNA polymerase (Q5 DNA polymerase, New England Biolabs) was employed to PCR-amplify the human ALDH1A1 promoter region (-1248 to +52) from genomic DNA of SUM159PT cell line. The mNeptune-TK fused protein coding sequence, replication origin (ori), and neomycin resistance gene were PCR-amplified using various templates vectors. Twenty cycles of PCR were performed, employing specific primers with flanking BsaI sites. The flanking overhangs were selected to complementarily ligate with overhangs from other PCR fragments. Therefore, the PCR fragments can orderly assemble and form a circular vector. The resulting PCR fragments were purified, and close circular plasmids were assembled via a single restriction-ligation reaction with BsaI enzyme (R3535L, New England Biolabs) and T4 DNA ligase (M0202L, New England Biolabs). The assembled plasmids were transformed into competent cells (C404003, Invitrogen). The plasmids were extracted with QIAGEN kits. Sequencing primers were synthesized by Eurogentech and the vector was sequenced by GATC (Sanger sequencing).

Thanks to this construction, we can sort the mNeptune high cells that exhibit stemness characteristics (self-renewal, differentiation and tumorigenicity) (***Bidan et al., 2019***).

The cells were transfected with the vectors using nucleofection. Five hundred thousand cells are resuspended in 100 μL of buffer (kit V, VCA-1003, Lonza) with 1 μg of DNA and electroporated using the X-013 protocol of the Nucleofector II Device (Amaxa).

Generation of stable cell lines

Stable cell lines pALDH1A1-mNeptune are obtained as previously explained (***Bidan et al., 2019***). The mNeptune fluorescence was examined 24 hours post-transfection and the positive cells were selected with 1 mg/ml of G418 (Invitrogen). Subsequent to selection, cells positive for fluorescence were sorted using fluorescence-activated cell sorting (BD FACS Aria III). Several days after cell sorting, a heterogeneous cell population with mNeptune positive and mNeptune negative cells is regenerated as expected (***Bidan et al., 2019***). Regularly, cell cultures are checked for functional stemness reporter. To do so, negative cells are sorted using fluorescence-activated cell sorting (BD FACS Aria III) and we look for establishement of heterogeneous cell population with mNeptune positive and mNeptune negative cells.

After establishment of pALDH1A1-mNeptune stable cell line, cells were transfected with a pCMV-Grx1roGFP2-Hygromycin vector then the stable cell lines were selected with hygromycin B (the vector was modified from pEIGW-Grx1-roGFP2, 64990, Addgene). The information obtained from Grx1-roGFP2 fluorescence is not utilized in this work.

Time-lapse Microscopy

Ten thousands cells were seeded into 35 mm glass bottom dish in 2 mL of cell media (D35-20-1.5-N, Cellvis). After 24 hours, cells are stained with 1 $\mu g/ml$ of Hoechst 33342 (3570, Invitrogen) in PBS 1X

during 20 min. Cells were washed with PBS 1X and 3 mL of cell media is added for the time-lapse experiment.

Samples were placed on a Nikon Ti-E microscope equipped with a motorized filters wheel (Nikon) and a XY-motorized stage (Applied Scientific Instrumentation). We used a custom-built top-stage incubator to regulate temperature, humidity, and atmosphere. The incubator was described in (***Guilbert et al., 2020***). Cells were maintained at 37°C and atmosphere is regulated at 5% CO_2. The microscope and top-stage incubator are placed in an enclosure with temperature maintained at 35°C. This limits temperature gradient between immersion objective and sample.

Cells were imaged on a sCMOS camera (Orca-Flash LT, Hamamatsu) through a 60X microscope objective (NA = 1.4, Nikon). We set the camera binning to 4 resulting in an effective pixel size of $0.43\mu m$. Illumination for fluorescence and brightfield images was achieved through custom-built optical system (components from Thorlabs) which allows synchronization of illumination with other apparatus. Exposure time was set to 50 ms for all experiments and for all channels. Hoechst was excited with a 365nm LED (M365LP1, Thorlabs) passing through a band-pass filter (FF01-390/40, semrock) resulting an average light intensity of $\sim 1.0mW.mm^{-2}$ on the sample. Fluorescence was collected and image on the camera *via* a dichroic mirror (FF416-Di01, semrock) and a band-pass filter (FF01-445/20, semrock). m-Neptune was excited with a 590nm LED (M590L3, Thorlabs) passing through a band-pass filter (FF01-593/40, semrock) resulting an average light intensity of $\sim 106mW.mm^{-2}$ on the sample. Fluorescence was collected and image on the camera *via* a dichroic mirror (Di02-R635, semrock) and a band-pass filter (FF01-680/42, semrock).

We use a custom-built acquisition software written in labview to control the setup. The sample is scanned to record 1024 overlapping images ($\sim 222\mu m \times \sim 222\mu m$). Channels were recorded sequentially for each given positions. Images stitching lead to a ∼6.8mm × ∼6.8 mm wide field of view. Lateral overlap between high-resolution images was $10\mu m$ in both directions. Scanning duration was ∼ 50 min. And scanning was repeated to perform time-lapse imaging with time resolution of ∼ 50 min.

Image processing, cell segmentation and tracking

Image processing was performed offline using custom written code in Matlab R2019a. Cell segmentation was performed on images of cell nuclei (Hoechst channel) assuming one nucleus per cell. After image restoration, cell segmentation and tracking were performed simultaneously. Tracking data were indeed used to improve segmentation as proposed in (***Chalfoun et al., 2016***). The SUM159PT dataset was fully curated by human intervention. Cell lineage and single cell fluorescence signal were subsequently extracted.

Image restoration

Image restoration aims at correcting shading in-homogeneity. We took advantage of the very high-throughput of our experiments to extract both background and foreground profiles. Such a method have been discussed previously and our methodology is inspired from both (***Kask et al., 2016***) and (***Peng et al., 2017***). Prior to shading correction, camera offset was estimated by recording dark images and was then subtracted. Then We use the following expression for shading correction(***Kask et al., 2016***) :

$$\widetilde{i}(x,y) = \frac{m(x,y) - bf(x,y)}{g(x,y)} \tag{1}$$

where $\widetilde{i}$ and m are respectively an estimation of the true image and the distorted (measured) images, b is background intensity, g and f are respectively the multiplicative and additive modulating functions.

Thanks to the very high throughput of our experiments, both modulating function, f and g, were estimated retrospectively from the data. Indeed, each time-lapse experiment involve recording $\sim 10^5$ vignettes which all have the same modulating functions. In each image, sub-regions

could be classified (*via* image segmentation) into either foreground or background. Thus f and g could be estimated one after the other by averaging. Modulating function were estimated by combining data from at least 3 different time-lapse experiments. For each experiment, because cells randomly cover the field of view, we separately accumulated data by blocks from either background or foreground. Blocks were averaged to reconstruct both spatial profiles, f and g, up to a multiplicative factor (***Kask et al., 2016***). In order to limit the influence of outliers, we only retain values between 5th and 95th percentiles for both background and foreground. Background was estimated first and non-uniform background was subtracted prior to foreground estimation. For background estimations, images were binned 2 times resulting a 256×256 pixels estimation of f. Similarly, for foreground estimation, images were binned 16 times resulting a 32×32 pixels estimation of g. Averaged modulating functions were then resized to original size (512×512) and smoothed with a Gaussian filter (radius of 2 pixels for background and 16 pixels for foreground). After normalization with respect to their maximum, equation 1 was applied pixel wised. The b factor was estimated by averaging $m(x, y)/f(x, y)$ over all background pixels of each image separately.

Masks obtained by segmentation from Hoechst channel were used to reconstruct modulating functions of all channels. Foreground mask was defined as region with identified nuclei while background masks were obtained by excluding disks of radius $22\mu m$ around each detected cells. While, given cell segmentation, reconstruction of f and g is straightforward, restoration of Hoechst images suffers from a chicken-and-egg problem. To work around this issue, image restoration of Hoechst images was performed in two steps. First, a single modulating function was estimated by assuming the same non-homogeneous profile for background and foreground ($f = g$). A first segmentation was performed which was used only to extract Hoechst channel modulating functions, f and g, as described above. Then, the final segmentation of nuclei was performed after restoration using equation 1.

Cell segmentation and tracking

We used tracking information to correct segmentation errors as suggested in (***Chalfoun et al., 2016***). Here, we first describe segmentation then cell tracking and finally explain how tracking is used as a feedback to further refine cell segmentation.

Shading correction was performed as described above. Then, images of nuclei were blurred with a gaussian filter of width $6.5\mu m$ to remove tiny details useless for nuclear shape segmentation. This procedure resulted in bell shaped intensity distributions centered on nuclei. We then detected local maxima to assign putative cell centers. Individual masks were initialized to disks of radius $22\mu m$ around putative centers. Nuclear masks were refined by iteratively removing border pixels using Otsu thresholding (***Otsu et al., 1975***). This procedure is followed by several morphological operations. First, holes are filled, then we performed erosion followed by dilatation with kernel of 2 pixels. Each mask was finally automatically screened to detect putative neighbouring cells for which masks were merged. To split joined masks of neighbouring cells, we compared mask boundary to its convex hull. To do so, for all subset of boundaries found inside convex hull, we selected the point closest to mask center and then computed the ratio between (*i*) distance of this point to center and (*ii*) distance of closest section of convex hull to mask center. If at least two points were found with this ratio smaller than 0.75 the mask was split in two parts.

Cell segmentation as described above was performed for each vignette separately. Then the reconstructed field of view was automatically screened for duplicate cell masks caused by image overlap. Overlapping masks found in two different but neighbouring vignettes were identified as duplicate and only the largest mask was kept for further analysis. Information on nuclei area and Hoechst fluorescence intensity were then collected for the whole field of view and at each time point. These data were then used to filter false positives and, in particular, small masks with low fluorescence signal. Detected cells from all vignettes were assembled to assign coordinates (X, Y) for tracking. Tracking was performed by connecting all cells detected at a given time point to its nearest neighbour at the next time point. We used the algorithm described in (***Sbalzarini***

and Koumoutsakos, 2005) minimizing the following cost function : $c = \sum_{ij}(X_i - X_j)^2 + (Y_i - Y_j)^2 + \alpha((W_i - W_j)^2 + (w_i - w_j)^2)$ where summation runs over all paired cells (i, j), X, Y are cells coordinates and W, w are the lengths of major and minor axis of the mask. The parameter α was set to 0.12. Association of cost function higher than $30\mu m$ were not considered as described in (***Sbalzarini and Koumoutsakos, 2005***).

As proposed in (***Chalfoun et al., 2016***), we used tracking as a feedback to enhance cell segmentation. A common error in nuclei segmentation is that two neighbouring cells come in close contact and lead to detection of a single mask for both cells. We call this event "cells collision". Collision events can easily be identified from tracking data because the trajectory of one cell prematurely ends. Conversely, the mask from a single cell can be correctly detected at one frame but split in the following frame. We named this event "cell over-split". Over-split events can be confounded with natural cell division. However, during mitosis the cell transiently become brighter in the bright-field channel. To distinguish between mitosis and over-split we thus used a contrast parameter estimated from brightfield images. We computed histogram of all background pixels. The contrast parameter was defined as the fraction of mask pixels out of the 99% confidence interval. Cells with contrast parameter higher than a user defined threshold (typically 0.2) were considered as mitotic ruling out over-split. Screening the tracking data for over-split or collision events allowed us to correct the segmentation. Tracking was then run again. This procedure was repeated 10 times and the number of collision/over-split events was decaying progressively close to zero.

Single cell fluorescence distribution and fluorescence thresholds

Shading correction for the CSC reporter channel was performed as described above. After shading correction we applied a median filter in a window of 3×3 pixels to remove outlier pixels caused by noise amplification in region of low foreground modulating function. Because the CSC reporter is homogeneously distributed across the cell we used fluorescence averaged over the nuclear region as a proxy for single cell fluorescence, I.

For both cell lines, single cell fluorescence distribution was fitted to a compound distribution : A normal distribution for cells with low intensities (mean μ and standard deviation σ) and a gamma distribution for cells with intermediate intensities (shape parameter k and scale parameter θ) and another gamma distribution for cells with higher fluorescence level (shape parameter k_+ and scale parameter θ_+). The probability density function for single cell fluorescence, I, reads :

$$p(I) = f_- \frac{e^{-\frac{1}{2}(\frac{I-\mu}{\sigma})^2}}{\sigma\sqrt{2\pi}} + f\frac{I^{k-1}e^{-I/\theta}}{\Gamma(k)\theta^k} + f_+\frac{I^{k_+-1}e^{-I/\theta_+}}{\Gamma(k_+)\theta_+^{k_+}} \quad (2)$$

where f_-, f and f_+ are respectively the weights of the CDC, iCC and CSC phenotypes. Equation 2 was fitted to data from all experiments all pooled together. We found an optimal set of parameters (all experiments pooled) to be : $\mu \approx 2.6$, $\sigma \approx 3.3$, $k \approx 2.3$, $\theta \approx 18.2$, $k_+ \approx 2.7$ and $\theta_+ \approx 55.7$ for SUM159-PT. For MDAMB-231 the optimal parameter set was : $\mu \approx 47.3$, $\sigma \approx 34.8$, $k \approx 3.0$, $\theta \approx 32.4$, $k_+ \approx 1.8$ and $\theta_+ \approx 115.5$. We note that weights of each sub-populations, f_-, f and f_+, could vary slightly from one experiment to another.

To attribute phenotype to single cells we used intensity thresholds (I_- and I_+) to separate the three cell populations : cells with signal lower than I_- are considered as Differentiated Cancer Cells (CDC) ; cells with signal higher than I_+ are considered as Cancer Stem Cells (CSC) ; and cells with signal between than I_- and I_+ are labelled as intermediate fluorescence Cancer Cells (iCC). Unique pair of thresholds were determined based on the fitted distribution for given cell line. I_- was chosen so that complementary cumulative distribution function of the CDC cells was bellow 1%. These values were computed independently of the weights, f_-, and thus depends only on μ and σ. Similarly, I_+ was chosen so that complementary cumulative distribution function of the CDC cells was bellow 1%. We found $I_- = 15.5$ and $I_+ = 129.5$ for SUM159-PT cells and $I_- = 115$ and $I_+ = 245$ for MDA-MB-231 cells.

Data curation

The above segmentation and tracking procedure lead to a false positive rate of $FP \sim 5\%$, a false negative rate of $FN \sim 3\%$, a sensitivity of $s = TP/(TP + FN) \sim 97\%$ and a positive predictive value of $ppp = TP/(TP + FP) \sim 95\%$. While these results are quite good (***Caicedo et al., 2019***), such an error may be a problem for cell tracking. For instance, assuming an average track length of 20 frames, one expects the probability to detect the cell at all time points to be $s^{20} = 0.54$. In other words, the cell is missed at least once in half of the trajectories. Such a situation will be a problem in particular for lineage reconstruction.

The segmented and tracking data were thus manually corrected using custom-built software in matlab. In brief, segmentation masks were overlaid with images of Hoechst staining and bright-field. Data were screened by a human to detect segmentation errors which were classified into three categories : false positives, missed cells or masks to fuse. Manual correction was saved and corresponding masks were corrected automatically. Tracking was performed once again. Manual correction of a full time-lapse took 40 hours for an untrained user which reduced to 25h after one round. The SUM159PT dataset presented here was fully corrected.

Lineage reconstruction

Lineage was reconstructed from tracking data by detecting mitotic events. We note that tracking assigns the same identifier to a mother cell and its closest daughter. Putative divisions were detected by screening creation of new trajectory in the vicinity of existing one (distance lower than $60\mu m$). Again we used the contrast parameter estimated from brightfield images to validate mitotic events. If the contrast parameter of either the mother cell or daughter cells was higher than a user defined threshold (typically 0.2) the event was considered as a mitosis.

After detection of all mitotic events, we screened for cells for which we could detect beginning and end of cell cycle, *ie.* cells for which birth and subsequent division were captured by the time-lapse and also detected by the algorithm. Only cells with cell cycle duration greater than $15h$ were retained for time-resolved analysis (lineage analysis and time-series clustering). On the other hand, all cells were used for spatial analysis.

Point pattern analysis

Ripley function and edge correction

The empirical Ripley, $K(r)$ function was estimated by counting events within a disk of radius r (***Cressie, 2015***) and averaging over all points of interest. However, observation of the sample in a finite area may lead to biased estimation because information on neighbours for points close to the edge is missing. This effect was corrected by introducing a weight, $w_i(r)$, that rescales counting for points, i, close to the edge (***Cressie, 2015***). We used uniform correction for which, $w_i(r)$ is the ratio between the area of a disk of radius r and the actual observed area within radius r (***Ripley, 1976***, ***1977***). $w_i(r)$ equals one for points far from the edge and is greater than one for points with truncated observation area.

The empirical Ripley function then reads :

$$K_{12}(r) = \frac{1}{N_1} \sum_{i \in C_1} \frac{1}{N_2} \sum_{\substack{j \in C_2 \\ i \neq j}} w_i(r) \delta_{ij}(r) \tag{3}$$

where $\delta_{ij}(r)$ equals 1 if distance between i and j is smaller than r and zero otherwise. C_1 and C_2 are two class of points. In the univariate case, $C_1 = C_2$ and N_1 is the number of points in C_1 and $N_2 = N_1 - 1$. In the bivariate case, $C_1 \neq C_2$ and N_1 and N_2 are respectively the number of points in C_1 or C_2. We note that the Ripley function is symmetrical by construction, *ie.* $K_{12} = K_{21}$.

The Ripley K function was estimated for all r up to $r = 500\mu m$ by steps of $\Delta r = 5\mu m$. It was computed in the univariate case (without distinguishing phenotypes) or the bivariate cases where cells were separated in two of the three sub-classes : CSC, iCC, CDC.

Point correlation function

The Ripley K function is useful to distinguish clustered or dispersed pattern compared to complete spatial randomness (CSR). However, this function scales with r^2 rendering its visualization and interpretation difficult for all spatial scales together. We instead used the Point Correlation Function (PCF), $g(r)$ which was estimated as follow (***Cressie, 2015***) :

$$g_{12}(r) = \frac{K_{12}(r + \Delta r) - K_{12}(r)}{2\pi r \Delta r + \pi \Delta r^2} \tag{4}$$

where, K_{12}, r and Δr were defined above. The PCF, $g_{12}(r)$ can be interpreted as the increase or the decrease of the likelihood of finding an type-2 event at a distance r of an type-1 event compared to what would be expected under CSR. $g_{12}(r)$ greater than 1 indicates a more clustered pattern than CSR while a value smaller than 1 indicates a more dispersed pattern than CSR.

Single cell time-series clustering

The aim of time-series clustering was to identify, without *a priori* knowledge on time scales nor shape, families of single cell temporal patterns of the CSC reporter signal.

To do so, we selected single cell traces for which we could detect beginning and end of cell cycle. Doing so, we obtained 19620 single cell time-series. We used time relative to cell cycle division, τ, to measure cell cycle progression : $\tau = 0$ refers to beginning of cell cycle and $\tau = 1$ to its end. The signals were resampled (zero-order hold resampling (***Pohlmann, 2000***)) so that all time-series shared the same number of points, $N = 60$. Euclidean distance, d_{kl}, was used to compare single cell time-series : $d_{kl}^2 = \sum_\tau (I_k(\tau) - I_l(\tau))^2$ where I_k and I_l are signals from two different time series.

All cells drastically change shape during mitosis. They transiently round up and their apparent area was thus smaller than during the rest of cell cycle. This transient morphological change caused a bias in the fluorescence signal estimation during mitosis compared to the rest of the cell cycle. The estimated signal indeed abruptly increased during mitosis and was restored during cycle when the cell was plated back. To avoid clustering on these parts of the time-series, we excluded the six first time points and the six last time points of resampled time-series to search for temporal pattern.

To favor detection of transitioning temporal patterns, clustering was first performed on 6 subsets independently. The 3 first subsets were cells found to be either CDC, iCC or CSC at the beginning of cell cycle and which have changed phenotype at the end ; the 3 other subsets were cells that do not change phenotype. We partitioned time series around medoids (***Kaufman, 1990***; ***Fränti, 2018***). In the initialization step, time series were clustered using a hierarchical procedure where the number of clusters of subset s, k_s, was chosen based on a elbow plot. For all subsets, k_s was between 10 and 25% of the total number of time-series in the subset. A second step aims at optimizing the selection of k_s-centroids. We randomly swapped an existing cluster center (medoid) with a non-medoid time-serie (***Fränti, 2018***). The permutation was retained if it lead to a decrease of global explained variance, W, (***Fränti, 2018***). This procedure was iterated $1000k_s$ times. In a final step, all clusters we merged using a hierarchical procedure.

The final number of clusters, k, was chosen using the Gap-statistics (***Tibshirani et al., 2001***). This method uses a synthetic datasets to monitor how $W(k)$ decreases with k if there where no significant temporal pattern to find. We generated reference datasets of the same size as the experimental one. All time-series of the reference datasets were assumed constant but with additive noise. To do so we randomly initialized synthetic time-series with signal according to the fitted fluorescence distribution. The noise level was chosen by examining mean squared deviation of constant experimental traces. The synthetic datasets was then clustered using the same procedure as for the experimental dataset. To select the relevant number of clusters, k^*, we compared global explained variance of clustering of experimental data to mean and standard deviation of 20 simulated datasets. As expected, for low values of k, $W(k)$ decreases faster for the simulated datasets compared to the experimental data. The optimal number of clusters was chosen as described in the original article (***Tibshirani et al., 2001***).

Statistical analysis

Bootstrap resampling

Bootstrap resampling (***Efron, 1992***) was used to estimate sensitivity of several quantities without knowledge of the underlying error distribution. To do so, bootstrap randomly resamples the data with replacement. The procedure is repeated N_{bs} times to estimate mean and confidence interval for the quantity of interest. For all uses of bootstrap resampling, we chose N_{bs} to ensure that value obtained with bootstrap coincide with the empirical mean estimated without bootstrap(***Efron, 1992***).

Phenotype shuffling

Phenotype shuffling was used as a statistical test for bivariate point pattern analysis. Indeed, the confidence interval of point correlation function (PCF) strongly depends on the number of sample. Moreover, the univariate PCF exhibits a structure indicating that cells display spatial clustering independently of their phenotype.

With phenotype shuffling we aim at deciding whether correlations and anti-correlations between phenotypes revealed by bivariate PCF could be solely explained by the univariate spatial pattern or whether such correlations characterize specific interactions between phenotypes. To do so phenotype shuffling compute phenotype distribution as if phenotypes were randomly distributed across univariate distribution. At a given time point, we used measured cells positions but fluorescence intensities are attributed randomly by permuting all cells intensities. The bivariate PCF are then computed as described above with the simulated intensities. The procedure is repeated 1500 times to compute mean and confidence interval.

p-values

p-values are used in this work to assess significance of spatial correlations (anti-correlations) between phenotypes compared to (*i*) phenotype shuffling or (*ii*) numerical simulations. p-values estimate probabilities of the null hypothesis.

For figures 3 and 4, the null hypothesis is "the estimated bivariate PCF does not differ from bivariate PCF upon phenotype shuffling". To estimate p-values, we first used bootstrap resampling to estimate mean, μ and standard deviation σ of the bivariate PCF at a given radius. Then, we applied phenotype shuffling and computed the corresponding bivariate PCF 1500 times. p-value is defined as the fraction of shuffling for which the corresponding bivariate PCF falls within the range $[\mu - \sigma; \mu + \sigma]$ at the desired radius.

For figures 1 and 3, the null hypothesis is "the experimental bivariate PCF does not differ from the simulated bivariate PCF". To estimate p-values, we first used bootstrap re-sampling to estimate mean, μ and standard deviation σ of the experimental bivariate PCF at a given radius. We ran spatial simulations and computed the corresponding bivariate PCF 500 times. p-value is defined as the fraction of simulated data for which the corresponding bivariate PCF falls within the range $[\mu - \sigma; \mu + \sigma]$ at the desired radius.

Determination coefficient

Determination coefficient, R^2 is used to examine correlation between fluorescence signal at different stages of the cell cycle or between a mother cell and its progeny. The determination coefficient between random variable Y and X examine the variance explained by a linear relationship, $Y^{th} = aX + b$. Coefficients a and b are obtained by fitting the data using matlab built-in fucntion. For a given sample, R^2 reads :

$$R^2 = 1 - \frac{\sum_i Y_i - Y_i^{th})}{\sum_i (Y_i - < Y >)} \tag{5}$$

where summation run over all samples and $< Y >$ is the sample mean.

Numerical simulations

Numerical simulations are used for statistical hypothesis testing. Cells are assumed to have an internal variable, I, which represents the fluorescence intensity of the CSC reporter. Cells are assumed to have two states 1 and 2. The first state correspond the beginning of cell cycle and the second to end of cell cycle. Cells can make transition from state 1 to 2 (which correspond to cell cycle evolution) and then from 2 to 1 (which correspond to mitosis).

Memory-less chain model

In this model, the internal variable, I, is assumed to change only during state transition. The new value of I depends on its value at the previous state and follows the empirical distribution shown at figure 4 B and C for SUM159PT or figure 3 for MDA-MB-231. To simulate the new value of the intensity, I_2, at transition $1 \leftarrow 2$, we first compute the empirical cumulative of the conditional probability density function, $p(I_2|I_1)$, of I_2 given I_1. We used logarithmic sampling for intensity with 62 bins and intensity comprised between -100 and 5000 RFU. A uniform pseudo-random number is generated *via* matlab built-in funciton and I_2 is obtained by inverse transform of the conditional CDF. The new value of the intensity, I_1, at transition $2 \leftarrow 1$, is computed the same but we used the conditional probability density function, $p(I_d|I_m)$, of I_d given I_m where I_m is replaced by I_1. Simulations are run 500 times for 10000 cells. Then quantities described in the text are computed the same way as for experimental data.

Memory-less spatial model

Motility and divisions are not simulated. Instead we used trajectories and lineage extracted from experiments. Thus for each trajectories, state transition $1 \leftarrow 2$ and $2 \leftarrow 1$ are defined by experimentally determined lineage. Again, the internal variable, I, is assumed to change only during state transition. Because time evolution of fluorescence of transiting cells was found to be monotonous, we assumed linear time evolution during cell cycle. Evolution of the internal variable, I, is calculated the same way as for Memory-less chain model. Doing so we could simulate intensities at each time point for each cell of the experiment. The simulations were repeated 500 times for each experiments and quantities described in the text are computed the same way as for experimental data.

Phenotype density estimation

To estimate density of a given phenotype, for each cell, we counted the number of cell within a circle of radius $R = 300\mu m$ centered at the cell position. For cells at the edge, we applied correction as for empirical estimation of the Ripley, K, function. For SUM159PT cells, f_+ is defined as the number of CSC divided by the total number of cells. Similarly, f_- is defined as the number of CDC divided by the total number of cells. For MDA-MB-231 cells, f_+ is defined as the number of CSC and iCC divided by the total number of cells. This smooth variations given the very low number of CSC in MDA. f_- is defined the same way as for SUM159PT.

References

Alber AB, Paquet ER, Biserni M, Naef F, Suter DM. Single live cell monitoring of protein turnover reveals intercellular variability and cell-cycle dependence of degradation rates. Molecular cell. 2018; 71(6):1079–1091.

Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell. 2006; 127(5):905–915.

Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, Kempski H, Moorman AV, Titley I, Swansbury J, Kearney L, Enver T, Greaves M. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011; 469(7330):356–361. https://doi.org/10.1038/nature09650, doi: 10.1038/nature09650.

Auffinger B, Tobias A, Han Y, Lee G, Guo D, Dey M, Lesniak M, Ahmed A. Conversion of differentiated cancer cells into cancer stem-like cells in a glioblastoma model after primary chemotherapy. Cell Death & Differentiation. 2014; 21(7):1119–1131.

876 **Bidan N**, Bailleul-Dubois J, Duval J, Winter M, Denoulet M, Hannebicque K, El-Sayed IY, Ginestier C, Forissier V,
877 Charafe-Jauffret E, et al. Transcriptomic analysis of breast cancer stem cells and development of a pALDH1A1:
878 mNeptune reporter system for live tracking. Proteomics. 2019; 19(21-22):1800454.

879 **Bocci F**, Gearhart-Serna L, Boareto M, Ribeiro M, Ben-Jacob E, Devi GR, Levine H, Onuchic JN, Jolly MK. Toward
880 understanding cancer stem cell heterogeneity in the tumor microenvironment. Proceedings of the National
881 Academy of Sciences. 2019; 116(1):148–157.

882 **Bocci F**, Gearhart-Serna L, Boareto M, Ribeiro M, Ben-Jacob E, Devi GR, Levine H, Onuchic JN, Jolly MK. Toward
883 understanding cancer stem cell heterogeneity in the tumor microenvironment. Proceedings of the National
884 Academy of Sciences. 2019; 116(1):148–157.

885 **Bocci F**, Jolly MK, George JT, Levine H, Onuchic JN. A mechanism-based computational model to capture the
886 interconnections among epithelial-mesenchymal transition, cancer stem cells and Notch-Jagged signaling.
887 Oncotarget. 2018; 9(52):29906.

888 **Brown MS**, Abdollahi B, Wilkins OM, Lu H, Chakraborty P, Ognjenovic NB, Muller KE, Jolly MK, Christensen BC,
889 Hassanpour S, et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs
890 disease progression and metastasis in breast cancer. Science advances. 2022; 8(31):eabj8002.

891 **Caicedo JC**, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, Heng C, Becker T, Doan M, McQuin
892 C, et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nature methods.
893 2019; 16(12):1247–1253.

894 **Chaffer CL**, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak
895 K, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proceedings
896 of the National Academy of Sciences. 2011; 108(19):7950–7955.

897 **Chalfoun J**, Majurski M, Dima A, Halter M, Bhadriraju K, Brady M. Lineage mapper: A versatile cell and particle
898 tracker. Scientific reports. 2016; 6(1):36984.

899 **Colacino JA**, Azizi E, Brooks MD, Harouaka R, Fouladdel S, McDermott SP, Lee M, Hill D, Madden J, Boerner J,
900 et al. Heterogeneity of human breast stem and progenitor cells as revealed by transcriptional profiling. Stem
901 Cell Reports. 2018; 10(5):1596–1609.

902 **Cressie N**. Statistics for spatial data. John Wiley & Sons; 2015.

903 **Efron B**. Bootstrap methods: another look at the jackknife. In: *Breakthroughs in statistics: Methodology and*
904 *distribution* Springer; 1992.p. 569–593.

905 **Fillmore CM**, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise
906 to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Research. 2008; 10(2):R25.

907 **Fränti P**. Efficiency of random swap clustering. Journal of big data. 2018; 5(1):1–29.

908 **Friedmann-Morvinski D**, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. EMBO
909 reports. 2014; 15(3):244–253.

910 **Gerlinger M**, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A,
911 Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR,
912 Nohadani M, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing.
913 New England Journal of Medicine. 2012; 366(10):883–892. https://doi.org/10.1056/NEJMoa1113205, doi:
914 10.1056/NEJMoa1113205, pMID: 22397650.

915 **Grace M**, Hütt MT. Regulation of spatiotemporal patterns by biological variability: General principles and ap-
916 plications to Dictyostelium discoideum. PLoS computational biology. 2015; 11(11):e1004367.

917 **Grosse-Wilde A**, Fouquier d'Hérouël A, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, Del Sol A, Walters KA,
918 Huang S. Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor
919 survival. PloS one. 2015; 10(5):e0126522.

920 **Guilbert M**, Anquez F, Pruvost A, Thommen Q, Courtade E. Protein level variability determines phenotypic
921 heterogeneity in proteotoxic stress response. The FEBS Journal. 2020; 287(24):5345–5361.

922 **Gupta PB**, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. Stochastic state transitions give
923 rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011; 146(4):633–644.

Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell stem cell. 2019; 24(1):65–78.

Handly LN, Pilko A, Wollman R. Paracrine communication maximizes cellular response fidelity in wound signaling. Elife. 2015; 4:e09652.

Hsiung CCS, Bartman CR, Huang P, Ginart P, Stonestrom AJ, Keller CA, Face C, Jahn KS, Evans P, Sankaranarayanan L, Giardine B, Hardison RC, Raj A, Blobel GA. A hyperactive transcriptional state marks genome reactivation at the mitosis–G1 transition. Genes & Development. 2016; 30(12):1423–1439. http://genesdev.cshlp.org/content/30/12/1423.abstract, doi: 10.1101/gad.280859.116.

Iliopoulos D, Hirsch HA, Wang G, Struhl K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proceedings of the National Academy of Sciences. 2011; 108(4):1397–1402.

Junttila MR, De Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature. 2013; 501(7467):346–354.

Kanwar SS, Yu Y, Nautiyal J, Patel BB, Majumdar AP. The Wnt/β-catenin pathway regulates growth and maintenance of colonospheres. Molecular cancer. 2010; 9:1–13.

Kashyap A, Rapsomaniki MA, Barros V, Fomitcheva-Khartchenko A, Martinelli AL, Rodriguez AF, Gabrani M, Rosen-Zvi M, Kaigala G. Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends in Biotechnology. 2022; .

Kask P, Palo K, Hinnah C, Pommerencke T. Flat field correction for high-throughput imaging of fluorescent samples. Journal of microscopy. 2016; 263(3):328–340.

Kaufman L. Partitioning around medoids (program pam). Finding groups in data. 1990; 344:68–125.

Kicheva A, Cohen M, Briscoe J. Developmental pattern formation: insights from physics and biology. Science. 2012; 338(6104):210–212.

Kim SY, Kang JW, Song X, Kim BK, Yoo YD, Kwon YT, Lee YJ. Role of the IL-6-JAK1-STAT3-Oct-4 pathway in the conversion of non-stem cancer cells into cancer stem-like cells. Cellular signalling. 2013; 25(4):961–969.

Lagadec C, Vlashi E, Della Donna L, Dekmezian C, Pajonk F. Radiation-induced reprogramming of breast cancer cells. Stem cells. 2012; 30(5):833–844.

Landge AN, Jordan BM, Diego X, Müller P. Pattern formation mechanisms of self-organizing reaction-diffusion systems. Developmental biology. 2020; 460(1):2–11.

Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, Wicha M, Clarke MF, Simeone DM. Identification of pancreatic cancer stem cells. Cancer research. 2007; 67(3):1030–1037.

Li Z, Bao S, Wu Q, Wang H, Eyler C, Sathornsumetee S, Shi Q, Cao Y, Lathia J, McLendon RE, Hjelmeland AB, Rich JN. Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. Cancer Cell. 2009 Jun; 15(6):501–513.

Liu CC, Lin JH, Hsu TW, Su K, Li AFY, Hsu HS, Hung SC. IL-6 enriched lung cancer stem-like cell population by inhibition of cell cycle regulators via DNMT1 upregulation. Int J Cancer. 2015 Feb; 136(3):547–559.

Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, Martin-Trevino R, Shang L, McDermott SP, Landis MD, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem cell reports. 2014; 2(1):78–91.

Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. Journal of cell biology. 2012; 196(4):395–406.

Meurette O, Mehlen P. Notch signaling in the tumor microenvironment. Cancer cell. 2018; 34(4):536–548.

Morrison SJ, Kimble J. Asymmetric and symmetric stem-cell divisions in development and cancer. nature. 2006; 441(7097):1068–1074.

Olmeda F, Ben Amar M. Clonal pattern dynamics in tumor: The concept of cancer stem cells. Scientific Reports. 2019; 9(1):15607.

970 **Oshimori N**, Guo Y, Taniguchi S. An emerging role for cellular crosstalk in the cancer stem cell niche. The
971 Journal of pathology. 2021; 254(4):384–394.

972 **Otsu N**, et al. A threshold selection method from gray-level histograms. Automatica. 1975; 11(285-296):23–27.

973 **Owen MR**, Sherratt JA, Wearing HJ. Lateral induction by juxtacrine signaling is a new mechanism for pattern
974 formation. Developmental biology. 2000; 217(1):54–61.

975 **Parra ER**. Methods to Determine and Analyze the Cellular Spatial Distribution Extracted From Multiplex Im-
976 munofluorescence Data to Understand the Tumor Microenvironment. Frontiers in Molecular Biosciences.
977 2021; 8.

978 **Peng T**, Thorn K, Schroeder T, Wang L, Theis FJ, Marr C, Navab N. A BaSiC tool for background and shading
979 correction of optical microscopy images. Nature communications. 2017; 8(1):14836.

980 **Pfeuty B**, Kress C, Pain B. Network features and dynamical landscape of naive and primed pluripotency. Bio-
981 physical Journal. 2018; 114(1):237–248.

982 **Plaks V**, Kong N, Werb Z. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor
983 cells? Cell Stem Cell. 2015 Mar; 16(3):225–238.

984 **Pohlmann KC**. Principles of digital audio. McGraw-Hill Professional; 2000.

985 **Ripley BD**. The second-order analysis of stationary point processes. Journal of applied probability. 1976;
986 13(2):255–266.

987 **Ripley BD**. Modelling spatial patterns. Journal of the Royal Statistical Society: Series B (Methodological). 1977;
988 39(2):172–192.

989 **Ruiz I**, Altaba A. Hedgehog signaling and the Gli code in stem cells, cancer, and metastases. Sci Signal. 2011;
990 4(200):19.

991 **Sbalzarini IF**, Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in cell biology.
992 Journal of structural biology. 2005; 151(2):182–195.

993 **Sha Y**, Haensel D, Gutierrez G, Du H, Dai X, Nie Q. Intermediate cell states in epithelial-to-mesenchymal transi-
994 tion. Physical biology. 2019; 16(2):021001.

995 **Shapiro JR**, Yung WKA, Shapiro WR. Isolation, karyotype, and clonal growth of heterogeneous subpopulations
996 of human malignant gliomas. Cancer research. 1981; 41(6):2349–2359.

997 **Sjöqvist M**, Andersson ER. Do as I say, Not (ch) as I do: Lateral control of cell fate. Developmental biology.
998 2019; 447(1):58–70.

999 **Sottoriva A**, Verhoeff JJ, Borovski T, McWeeney SK, Naumov L, Medema JP, Sloot PM, Vermeulen L. Cancer stem
1000 cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer research.
1001 2010; 70(1):46–56.

1002 **Strandkvist C**, Juul J, Baum B, Kabla AJ, Duke T. A kinetic mechanism for cell sorting based on local variations
1003 in cell motility. Interface focus. 2014; 4(6):20140013.

1004 **Tibshirani R**, Walther G, Hastie T. Estimating the Number of Clusters in a Data Set via the Gap Statistic. Journal
1005 of the Royal Statistical Society Series B (Statistical Methodology). 2001; 63(2):411–423. http://www.jstor.org/
1006 stable/2680607.

1007 **Tomellini E**, Touil Y, Lagadec C, Julien S, Ostyn P, Ziental-Gelus N, Meignan S, Lengrand J, Adriaenssens E, Po-
1008 lakowska R, et al. Nerve growth factor and proNGF simultaneously promote symmetric self-renewal, quies-
1009 cence, and epithelial to mesenchymal transition to enlarge the breast cancer stem cell compartment. Stem
1010 Cells. 2015; 33(2):342–353.

1011 **Waclaw B**, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial model predicts that dispersal
1012 and cell turnover limit intratumour heterogeneity. Nature. 2015; 525(7568):261–264.

1013 **Wahl GM**, Spike BT. Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity.
1014 NPJ breast cancer. 2017; 3(1):14.

1015 **Wang J**, Wakeman TP, Lathia JD, Hjelmeland AB, Wang XF, White RR, Rich JN, Sullenger BA. Notch promotes
1016 radioresistance of glioma stem cells. Stem cells. 2010; 28(1):17–28.

Wang X, Prager BC, Wu Q, Kim LJ, Gimple RC, Shi Y, Yang K, Morton AR, Zhou W, Zhu Z, et al. Reciprocal signaling between glioblastoma stem cells and differentiated tumor cells promotes malignant progression. Cell stem cell. 2018; 22(4):514–528.

Warmflash A, Sorre B, Etoc F, Siggia ED, Brivanlou AH. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat Methods. 2014 Aug; 11(8):847–854.

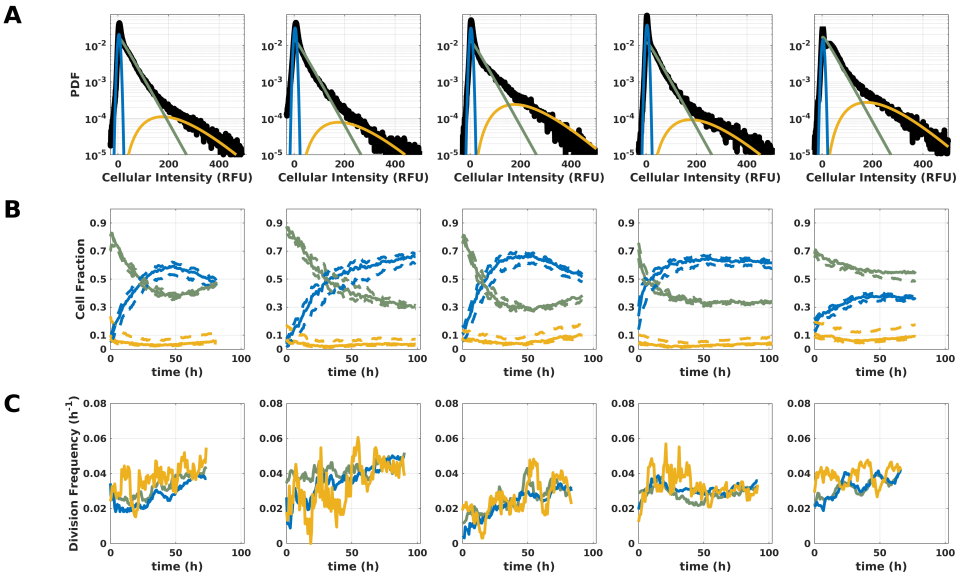
Wu M, Zhang X, Zhang W, Chiou YS, Qian W, Liu X, Zhang M, Yan H, Li S, Li T, et al. Cancer stem cell regulated phenotypic plasticity protects metastasized cancer cells from ferroptosis. Nature Communications. 2022; 13(1):1371.

Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, Zhang G, Wang X, Dong Z, Chen F, Cui H. Targeting cancer stem cell pathways for cancer therapy. Signal Transduction and Targeted Therapy. 2020; 5(1):8. https://doi.org/10.1038/s41392-020-0110-5, doi: 10.1038/s41392-020-0110-5.

Yoshida GJ, Saya H. Therapeutic strategies targeting cancer stem cells. Cancer science. 2016; 107(1):5–11.

Yung WKA, Shapiro JR, Shapiro WR. Heterogeneous chemosensitivities of subpopulations of human glioma cells in culture. Cancer research. 1982; 42(3):992–998.

Zhang H, Wang ZZ. Mechanisms that mediate stem cell self-renewal and differentiation. Journal of cellular biochemistry. 2008; 103(3):709–718.



1033

Figure 1—figure supplement 1. Repeatability of fluorescence distribution and time-evolution in SUM159-PT cells. SUM159-PT breast cancer cells were stably transfected with the CSC reporter, pALDH1a1:mNeptune (***Bidan et al., 2019***). Cells were imaged as described in the main text. Each column corresponds to an independent experiment. Individual cell nuclei were segmented and the nuclear average signal is used as a proxy for cell phenotype. (**A**) **Fitting of single cell fluorescence distribution.** The probability density function is well fitted by the sum of three distribution (thick black line) : A normal distribution for cells with low intensities (blue, mean $\mu \approx 2.6$ and standard deviation $\sigma \approx 3.3$) and a gamma distribution for cells with intermediate intensities (green, shape parameter $k \approx 2.3$ and scale parameter $\theta \approx 18.2$) and another gamma distribution for cells with higher fluorescence level (yellow, shape parameter $k_+ \approx 2.7$ and scale parameter $\theta_+ \approx 55.7$). Curve fitting was done by Maximum Likelihood Estimation. All experiments were first pooled to determine shape parameters (μ,σ,k,θ,k_+,θ_+) of each sub-distribution. Then the fitting procedure was repeated for each experiment fixing shape parameters to determine weights of each distribution.(**B**) **Time evolution of the three populations** (CDCs blue line, iCCs green line and CSCs yellow line) defined by the intensity thresholds ($I_- = 15.5$ and $I_+ = 129.5$). Dashed lines show the same but with varying thresholds ($12.5 \leq I_- \leq 17.5$ and $84.5 \leq I_+ \leq 148.5$) (**C**) **Instantaneous division rate of all three populations.** We first counted the number of division events detected within a time-windows of 8h for each sub-population. Then, division frequency was estimated by normalizing count by time-window duration and total number of cells of the given sub-population.

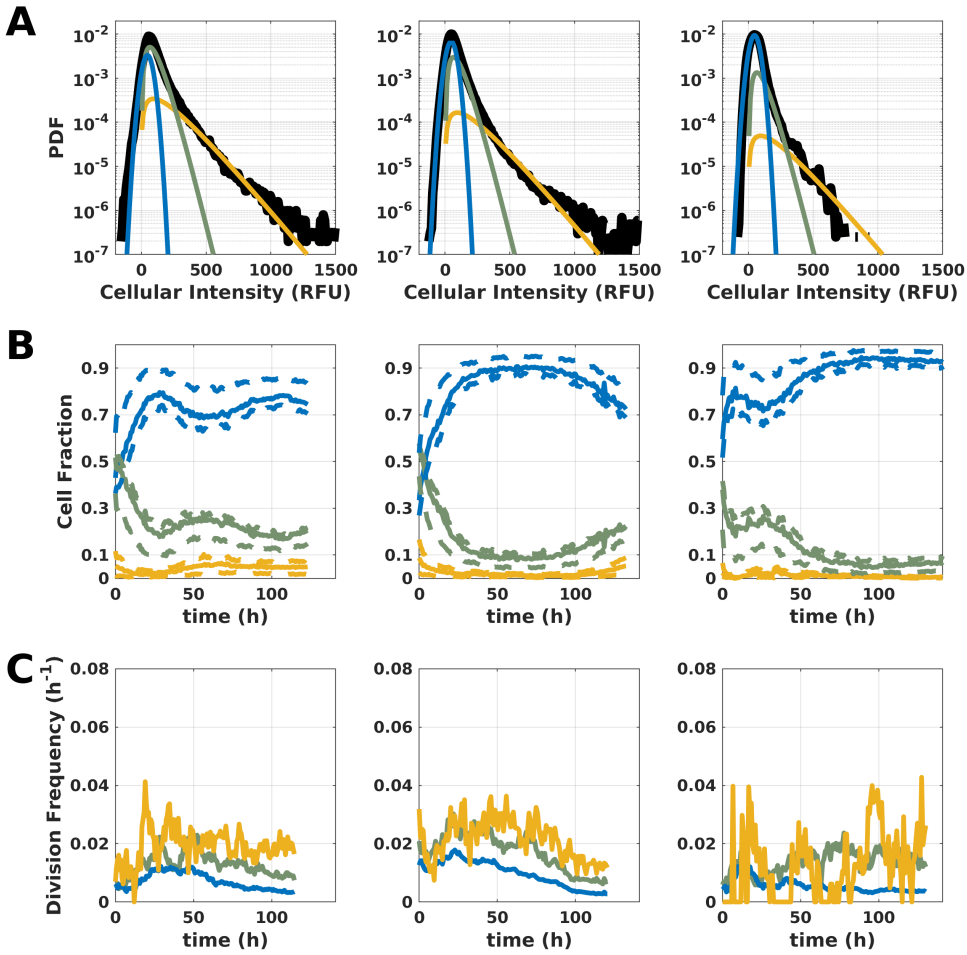


Figure 1—figure supplement 2. Repeatability of fluorescence distribution and time-evolution in MDA-MB-231 cells. MDA-MB-231 breast cancer cells were stably transfected with the CSC reporter, pALDH1a1:mNeptune (***Bidan et al., 2019***). Cells were imaged as described in the main text. Each correspond to an independent experiment. Individual cell nuclei were segmented and the nuclear average signal is used as a proxy for cell phenotype. (**A**) **Fitting of single cell fluorescence distribution.** The probability density function is well fitted by the sum of three distribution (thick black line) : A normal distribution for cells with low intensities (blue, mean $\mu \approx 47.3$ and standard deviation $\sigma \approx 34.8$) and a gamma distribution for cells with intermediate intensities (green, shape parameter $k \approx 3.0$ and scale parameter $\theta \approx 32.4$) and another gamma distribution for cells with higher fluorescence level (yellow, shape parameter $k_+ \approx 1.8$ and scale parameter $\theta_+ \approx 115.5$). Curve fitting was done by Maximum Likelihood Maximization. All experiments were first pooled to determine shape parameters (μ,σ,k,θ,k_+,θ_+) of each sub-distribution. Then the fitting procedure was repeated for each experiment fixing shape parameters to determine weights of each distribution.(**B**) **Time evolution of the three populations** (CDC blue line, iCC green line and CSC yellow line) defined by the intensity thresholds ($I_- = 115$ and $I_+ = 245$). Dashed lines show the same but with varying thresholds ($105 \leq I_- \leq 145$ and $205 \leq I_+ \leq 345$) (**C**) **Instantaneous division rate of all three populations.** We first counted the number of division events detected within a time-windows of 8h for each sub-population. Then division frequency was estimated by normalizing count by time-window duration and total number of cells of the given sub-population.

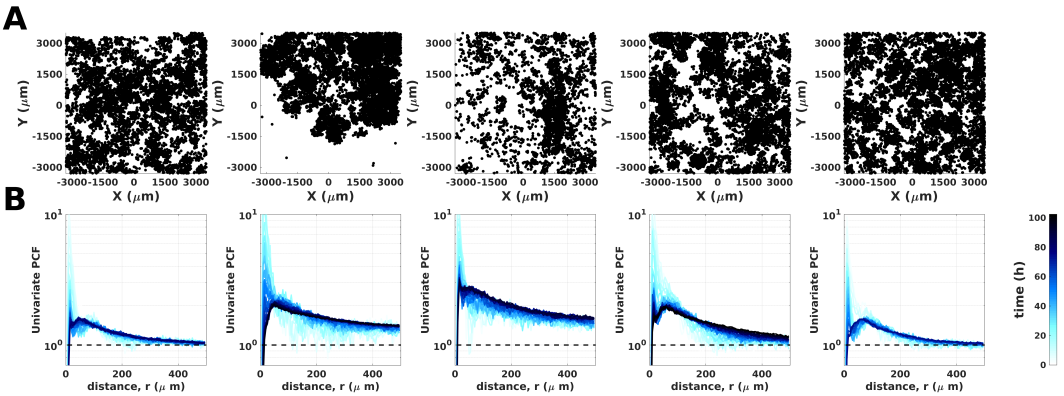


Figure 2—figure supplement 1. Univariate Point Correlation Function of SUM159-PT cells for all time lapse experiments.

(**A**) **Spatial distribution of cells** regardless of their phenotype at the end of the experiment. Each cell is represented by a black dot at its spatial coordinates. (**B**) **Univariate PCF** for all time points of the time lapse. The PCF, $g(r)$ is computed at each time point (see Material and methods) and averaged over 5 frames. Line color from light blue to dark blue codes for time (0 to 100h).

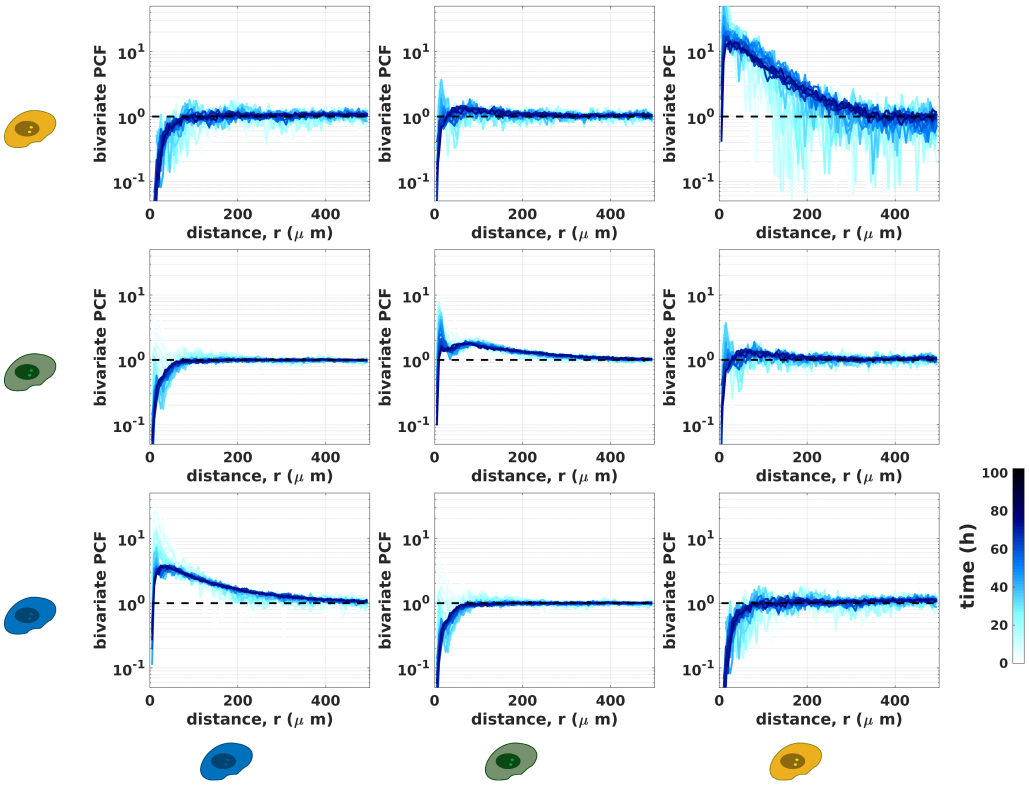


Figure 2—figure supplement 2. Time evolution of bivariate Point Correlation Functions of SUM159-PT cells for a representative experiment. The experiment is the same as the one shown in main figure. Line color from light blue to dark blue codes for time (0 to 100h). First line : bivariate PCF measuring spatial correlation between CSC and either CDC (first column) , iCC (second column) or CSC (third column). Second line : bivariate PCF measuring spatial correlation between iCC and either CDC (first column) , iCC (second column) or CSC (third column). Third line : bivariate PCF measuring spatial correlation between CDC and either CDC (first column) , iCC (second column) or CSC (third column).

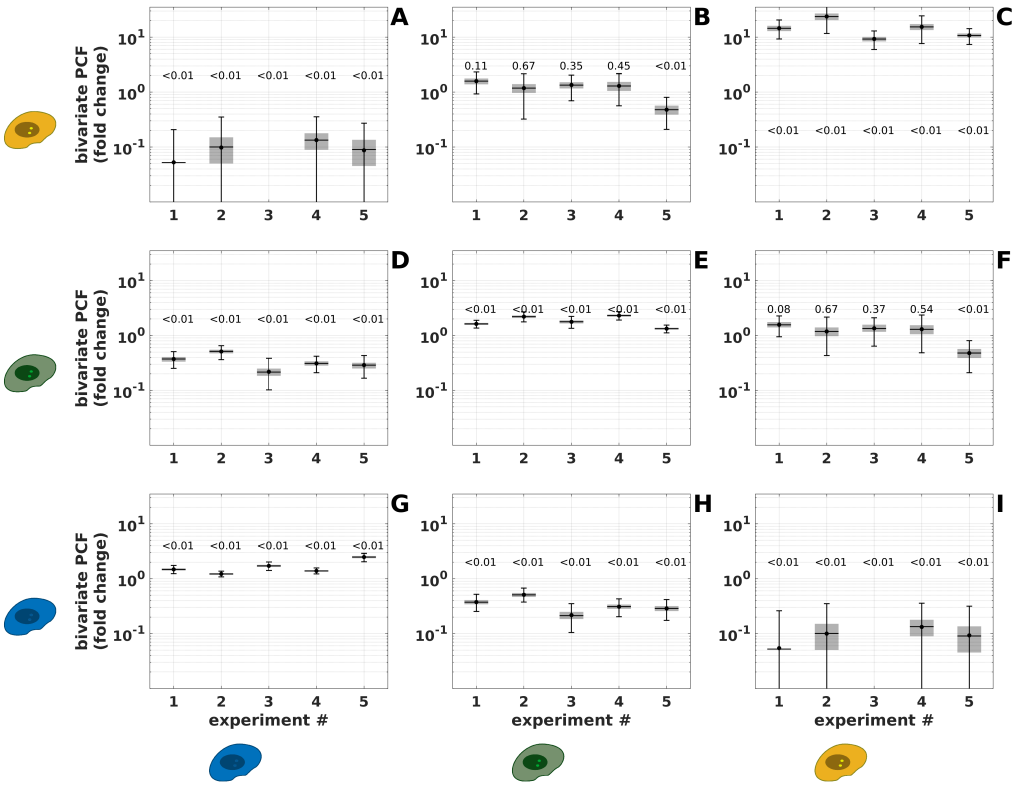


Figure 2—figure supplement 3. Repeatability of bivariate Point Correlation Function of SUM159-PT cells. We report the ratio between the measured bivariate PCF (data) and the one obtained for phenotype shuffling (control). This ratio is estimated at $r = 15\mu m$ for which the shuffling PCF is maximum. Black dot : actual ratio. Black horizontal line : median obtained from bootstrap resampling. Gray shaded box : 50% confidence interval. Error bars : 99% confidence interval. Note that data point is absent when the measured bivariate PCF is null because of the logarithmic scale. p-values are shown for each data point. p-values are estimated by 1,500 repeats of shuffling to test the null hypothesis (shuffling identical to data within the 68% confidence interval obtained from bootstrap resampling). (**A**) Bivariate PCF measuring spatial correlation between CSC and CDC. Data point is absent for experiment 3 because the measured bivariate PCF is null (cannot be shown on logarithmic scale). (**B**) Bivariate PCF measuring spatial correlation between CSC and iCC. (**C**) Bivariate PCF measuring spatial auto-correlation of CSC. (**D**) Bivariate PCF measuring spatial correlation between iCC and CDC. (**E**) Bivariate PCF measuring spatial auto-correlation between iCC. (**F**) Bivariate PCF measuring spatial correlation between iCC and CSC. (**G**) Bivariate PCF measuring spatial auto-correlation of CDC. (**H**) Bivariate PCF measuring spatial correlation between CDC and iCC. (**I**) Bivariate PCF measuring spatial correlation between CDC and CSC. Data point is absent for experiment 3 because the measured bivariate PCF is null (cannot be shown on logarithmic scale).

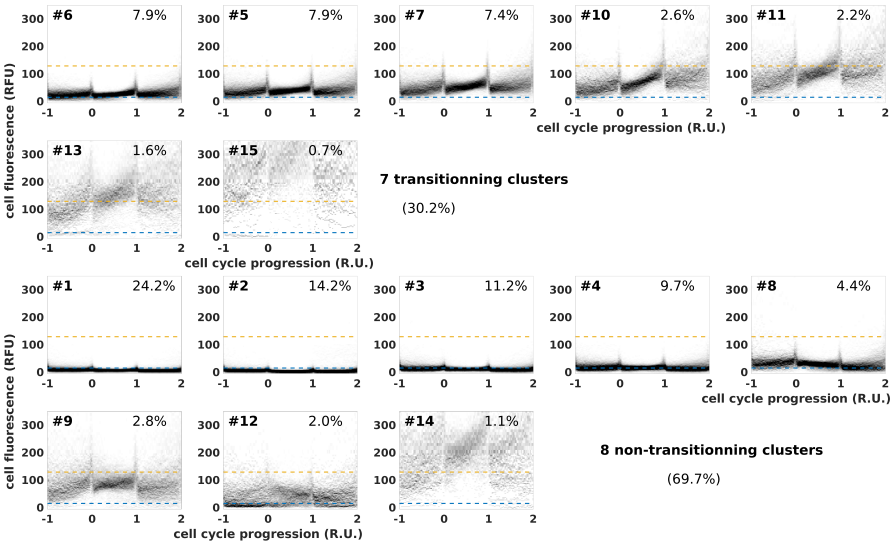


Figure 2—figure supplement 4. Repeatability of bivariate Point Correlation Function of MDA-MB-231 cells. We report the ratio between the measured bivariate PCF (data) and the one obtained for phenotype shuffling (control). This ratio is estimated at $r = 15\mu m$ for which the shuffling PCF is maximum. Black dot : actual ratio. Black horizontal line : median obtained from bootstrap resampling. Gray shaded box : 50% confidence interval. Error bars : 99% confidence interval. p-values are shown for each data point. p-values are estimated by 1500 repeats of shuffling to test the null hypothesis (shuffling identical to data within the 68% confidence interval obtained from bootstrap resampling). (**A**) Bivariate PCF measuring spatial correlation between CSC and CDC. (**B**) Bivariate PCF measuring spatial correlation between CSC and iCC. (**C**) Bivariate PCF measuring spatial auto-correlation of CSC. (**D**) Bivariate PCF measuring spatial correlation between iCC and CDC. (**E**) Bivariate PCF measuring spatial auto-correlation between iCC. (**F**) Bivariate PCF measuring spatial correlation between iCC and CSC. (**G**) Bivariate PCF measuring spatial auto-correlation of CDC. (**H**) Bivariate PCF measuring spatial correlation between CDC and iCC. (**I**) Bivariate PCF measuring spatial correlation between CDC and CSC.

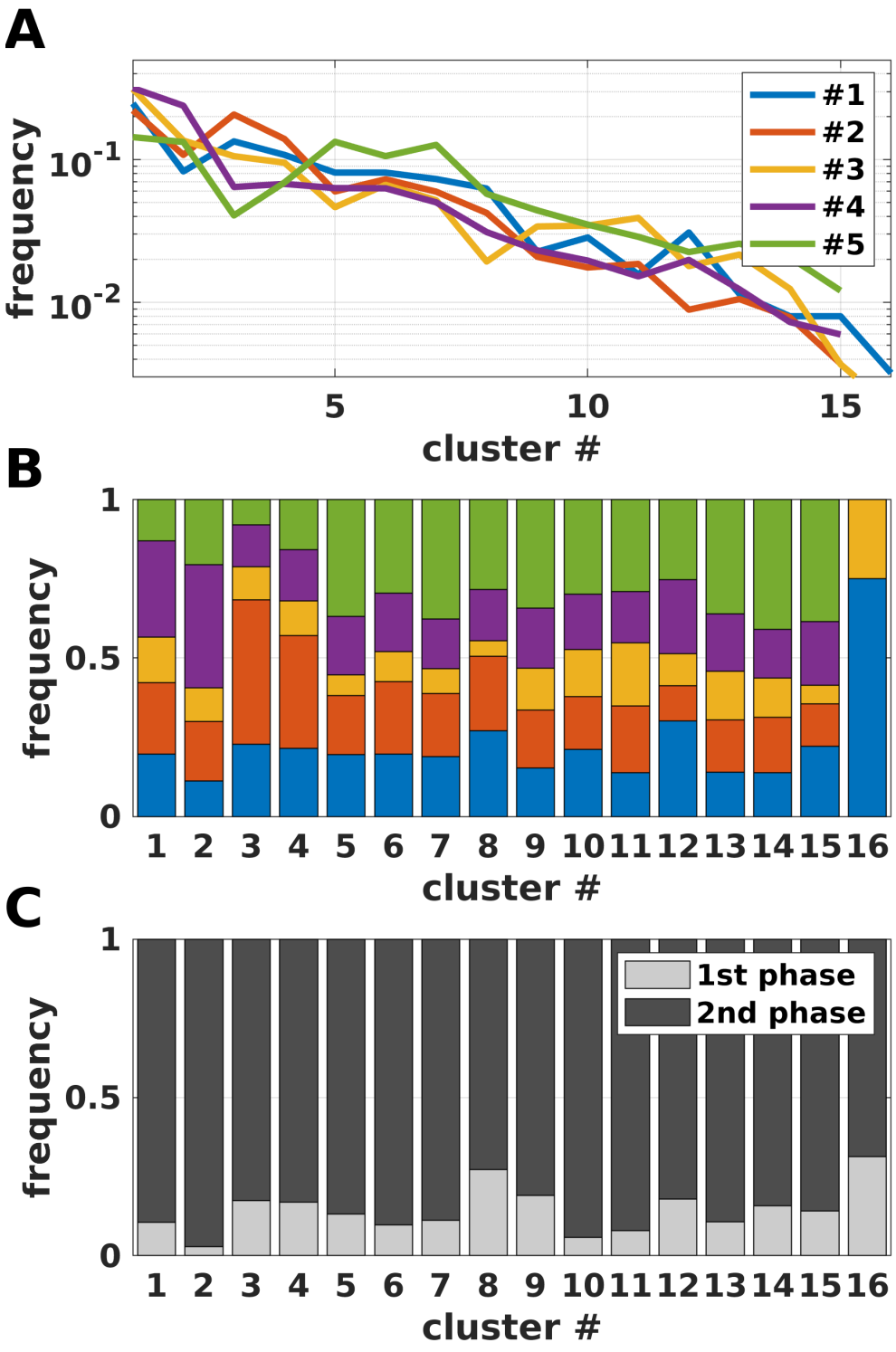


Figure 3—figure supplement 1. Meta-analysis of identified SCTS clusters : (**A**) Fraction (ordinate) of cells from a given experiment partitioned into a given cluster (abscissa). Each experiment is plotted in a different color. (**B**) Distribution of clusters among each experiment. Color code for experiment number is the same as for A. (**C**) Distribution of clusters among phases of the experiment. First phase is the transient phase and second phase is the stationary one for which steady proportion are reached for all phenotypes.

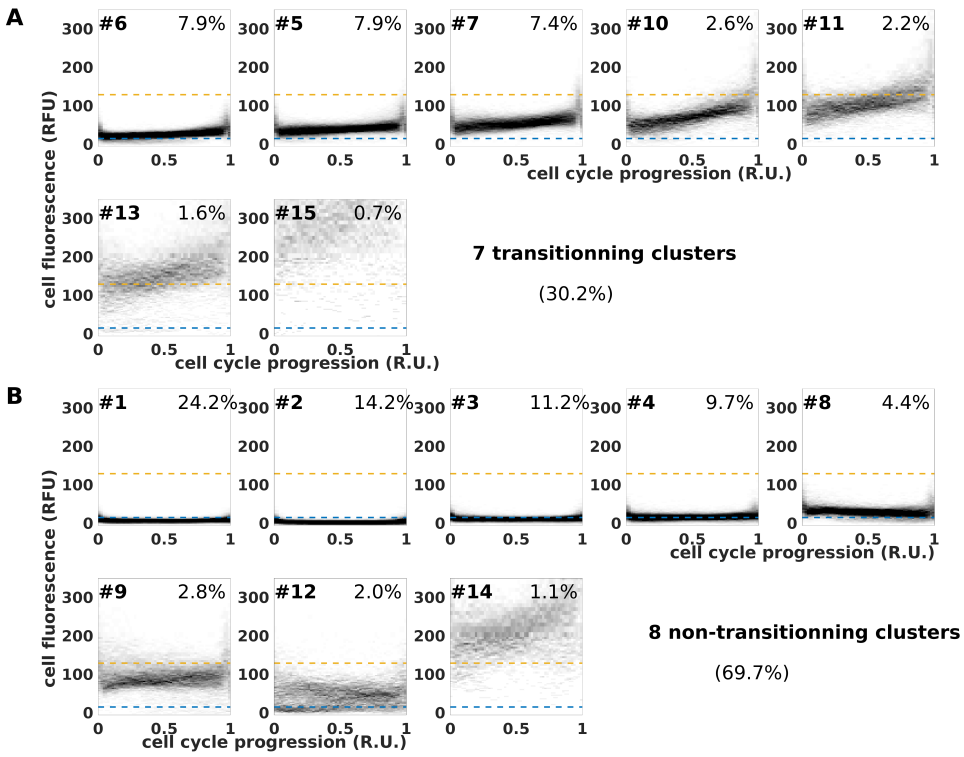


Figure 3—figure supplement 2. Transitionning (**A**) and non-transitionning SCTS clusters (**B**) in SUM159PT cells. Data are the same as for the main figure. Clusters are numbered by increasing size. The percentage of cells within each cluster is indicated.

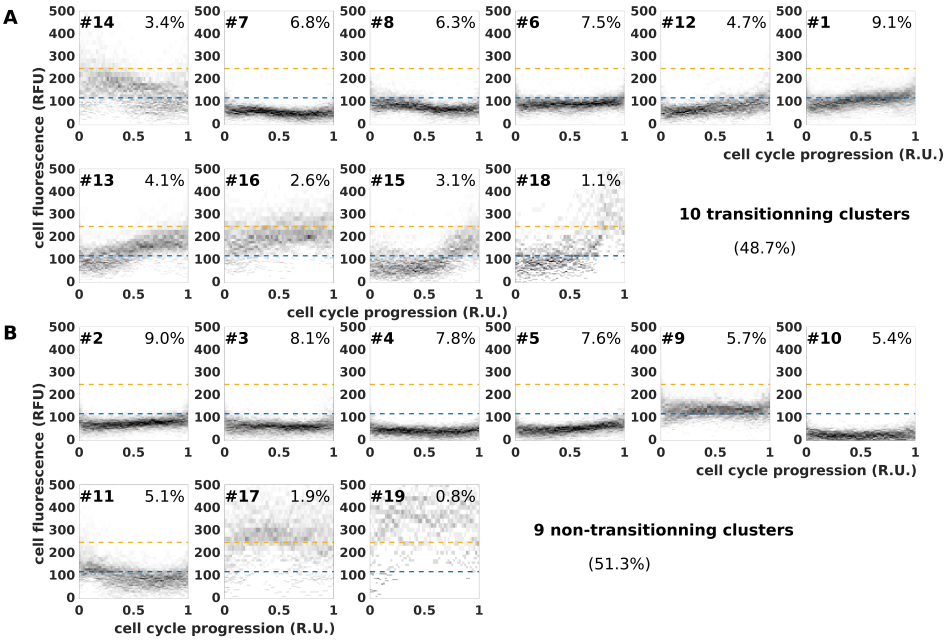


Figure 3—figure supplement 3. Transitionning (**A**) and non-transitionning (**B**) SCTS clusters in MDA-MB-231 cells. Clusters are numbered by increasing size. The percentage of cells within each cluster is indicated.

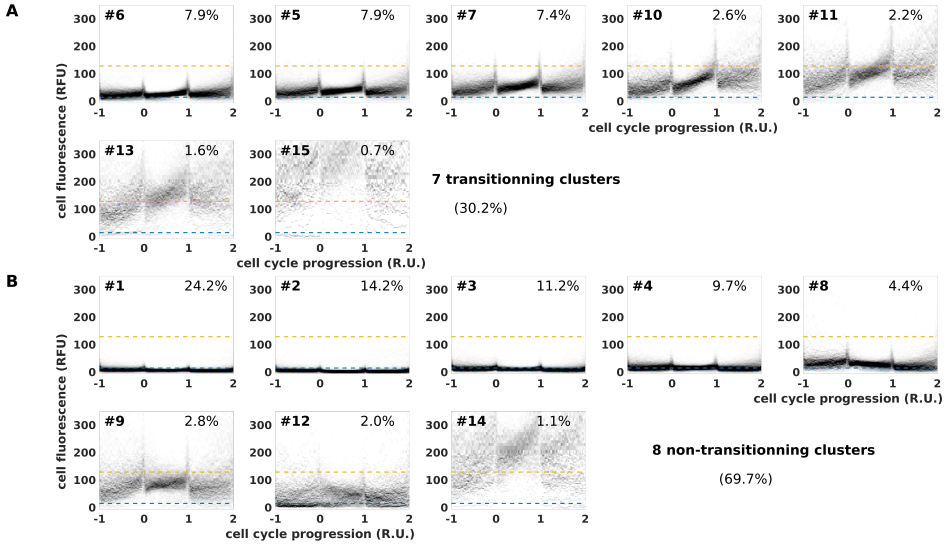


Figure 3—figure supplement 4. Data are the same as for the main figure but mother signal and daugthers signals are shown when available. Transitionning (**A**) and non-transitionning SCTS clusters (**B**) in SUM159PT cells.

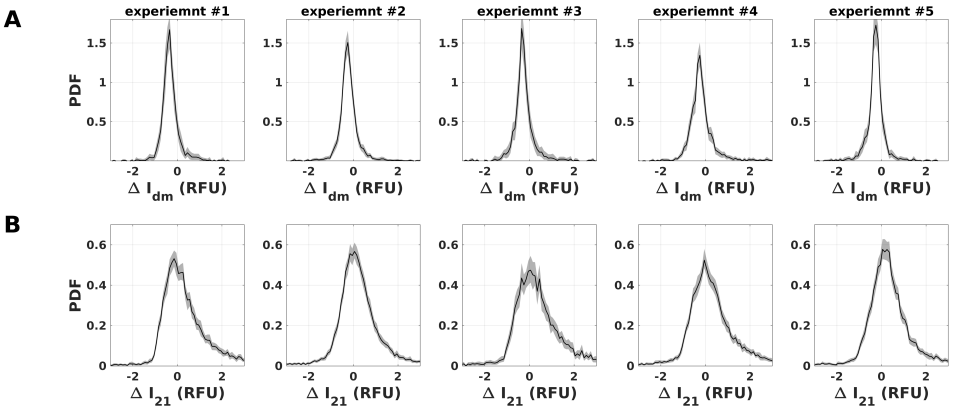


Figure 4—figure supplement 1. Repeatability of fluorescence variations during cell cycle and upon mitosis in SUM159PT cells : Each column correspond to an experiment.(**A**) Probability density function of fluorescence variation upon mitosis, $\Delta I_{dm} = I_d - I_m$, relative to mother's fluorescence, I_m. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods). (**B**) Probability density function of fluorescence variation during cell cycles, $\Delta I_2 = I_2 - I_1$, relative to fluorescence at beginning of cell cycle, I_1. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods).

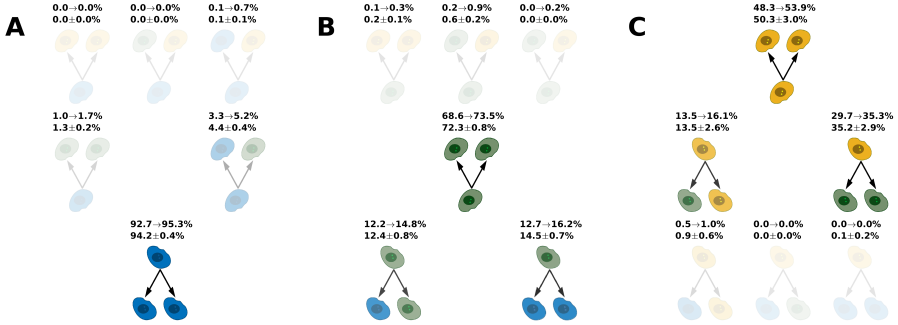


Figure 4—figure supplement 2. Symmetric and asymmetric division rates of SUM159PT cells: Phenotype are defined according to fluorescence thresholds, I_+ and I_-. Blue cell pictogram represents CDC, Green cell pictogram represents iCC and Yellow cell pictogram represents CSC. Percentages indicate rate of a given division type for (**A**) CDC, (**B**) iCC and (**C**) CSC. The lower values represent mean and standard deviation obtained by bootstrap resampling (sensitivity analysis, see materials and methods). The upper interval indicates sensitivity to thresholds : lower and higher values by varying thresholds ($12.5 \leq I_- \leq 17.5$ and $84.5 \leq I_+ \leq 148.5$).

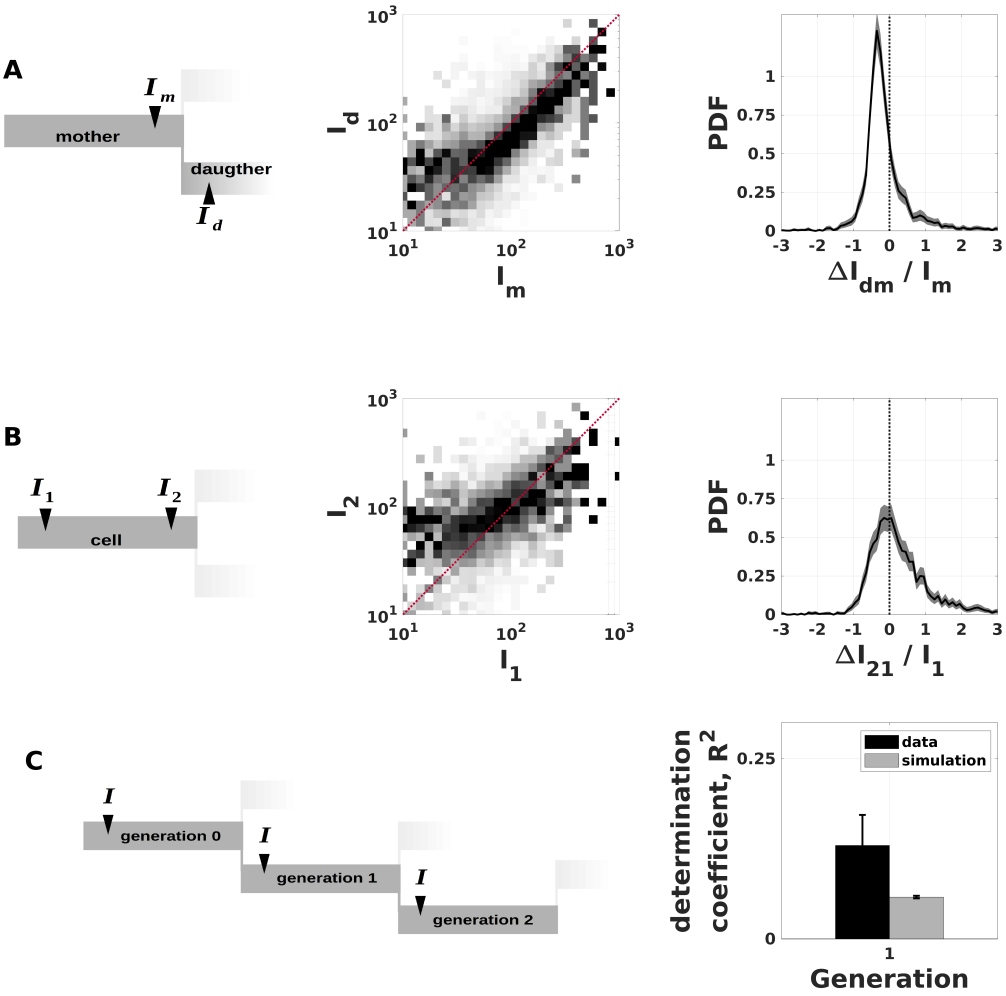
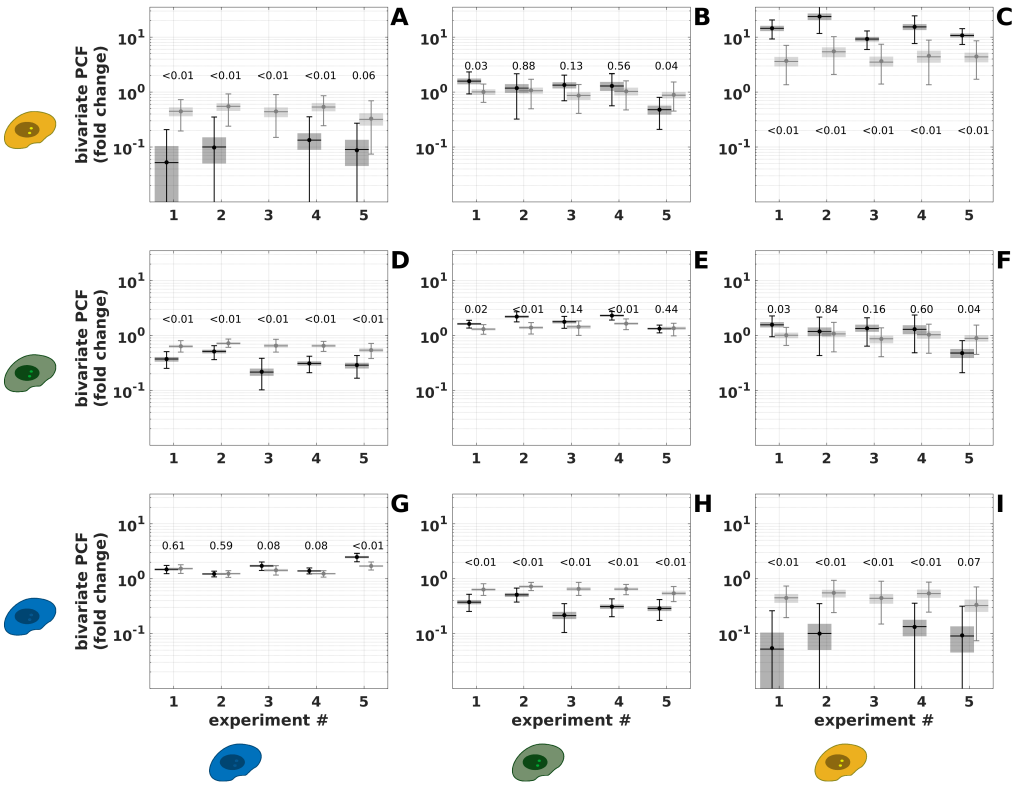


Figure 4—figure supplement 3. Analysis of fluorescence correlations in the lineage tree of MDA-MB-231 cells. (**A**) **Correlation of CSC reporter signal between mother and daughters.** Left panel : Mother fluorescence intensity, I_m, is measured at the end of cell cycle and daughter, I_d, fluorescence intensity is measured at the beginning of cell cycle. See materials and methods for details. Middle panel : scatter plot of daughter's fluorescence as a function of mother's fluorescence. Right panel : Probability density function of fluorescence variation upon mitosis, $\Delta I_{dm} = I_d - I_m$, relative to mother's fluorescence, I_m. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods). (**B**) **Correlation of CSC reporter signal between beginning and end of cell cycle for the same cell.** Left panel : Fluorescence intensity is measured both at the beginning of cell cycle (I_1) and at the end (I_2). See materials and methods for details. Middle panel : scatter plot of I_2 as a function of I_1. Left panel : Probability density function of fluorescence variation during cell cycles, $\Delta I_2 = I_2 - I_1$, relative to fluorescence at beginning of cell cycle, I_1. Black line represents the estimated PDF. Gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods). (**C**) **Correlation of CSC reporter signal across generations.** Left panel : Fluorescence intensity, I, is measured both at the beginning of cell cycle for each cell. See materials and methods for details. Right panel : Determination coefficient between signal of a mother cell and signal of its daughter. Black bars correspond to data and gray bars correspond to numerical simulation of a memory-less chain model (see materials and methods). Error bars represent the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see materials and methods).



1046

Figure 5—figure supplement 1. Repeatability of simulated bivariate Point Correlation Function of SUM159-PT cells. We report the ratio between the measured bivariate PCF and the one obtained for phenotype shuffling (control) for data (black) and simulations (gray). This ratio is estimated at $r = 15\mu m$ for which the shuffling PCF is maximum. Dot : actual ratio. Horizontal line : median obtained from bootstrap resampling. Gray shaded box : 50% confidence interval. Error bars : 99% confidence interval. p-values are shown for each data point. p-values are estimated by 500 repeats of simulations to test the null hypothesis (simulations identical to data within the 68% confidence interval obtained from bootstrap resampling). (**A**) Bivariate PCF measuring spatial correlation between CSC and CDC. Data point is absent for experiment 3 because the measured bivariate PCF is null (cannot be shown on logarithmic scale). (**B**) Bivariate PCF measuring spatial correlation between CSC and iCC. (**C**) Bivariate PCF measuring spatial auto-correlation of CSC. (**D**) Bivariate PCF measuring spatial correlation between iCC and CDC. (**E**) Bivariate PCF measuring spatial auto-correlation between iCC. (**F**) Bivariate PCF measuring spatial correlation between iCC and CSC. (**G**) Bivariate PCF measuring spatial auto-correlation of CDC. (**H**) Bivariate PCF measuring spatial correlation between CDC and iCC. (**I**) Bivariate PCF measuring spatial correlation between CDC and CSC. Data point is absent for experiment 3 because the measured bivariate PCF is null (cannot be shown on logarithmic scale).

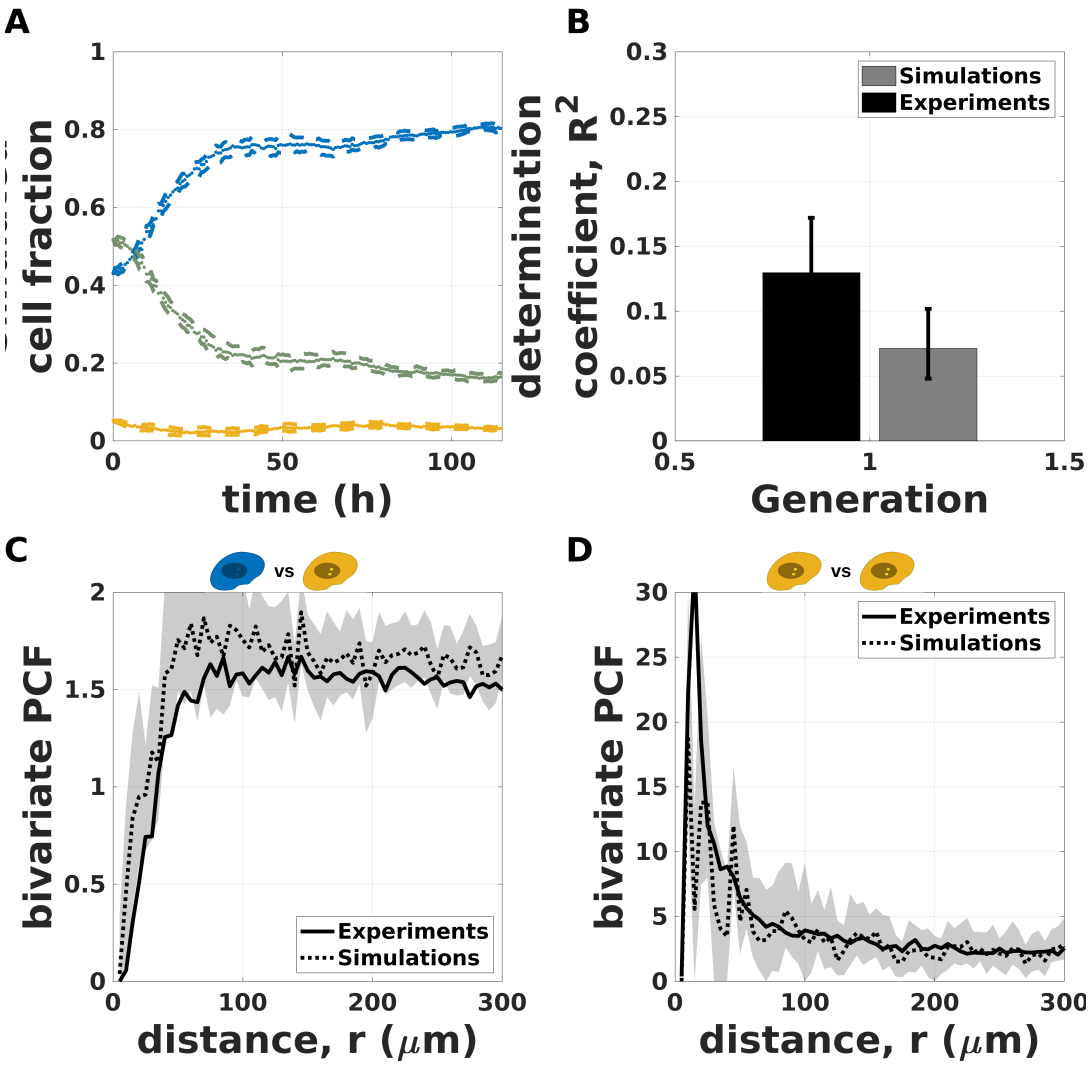


Figure 5—figure supplement 2. Spatio-temporal simulations of phenotypic inheritance of MDA-MB-231 cells (**A**) **Simulated time evolution of the cell fraction of the three phenotypes** (CDC blue , iCC green and CSC yellow) defined by the intensity thresholds I_- and I_+. Mean (points) and 99% confidence interval (dashed lines) obtained from 500 independent simulations based on trajectories of a representative experiment (same as figure ***Figure 1—figure Supplement 2***). (**B**) **Simulated correlation of fluorescence signal across generations.** Is shown determination coefficient between signal of a mother cell and signal of its daughter. Black bars correspond to data and gray bars correspond to numerical simulation (see material and methods). Error bars represent the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see material and methods). (**C**) **Simulated bivariate PCF for CSC versus CDC.** Continuous black line represents experimental PCF (same as figure ***Figure 2—figure Supplement 4***). Dashed line is the PCF of simulated data for the same experiment. The gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see material and methods). (**D**) **Simulated bivariate PCF for CSC versus CSC.** Continuous black line represents experimental PCF (same as figure ***Figure 2—figure Supplement 4***). Dashed line is the PCF of simulated data for the same experiment. The gray shaded area is the 99% confidence interval obtained by bootstrap resampling (sensitivity analysis, see material and methods).

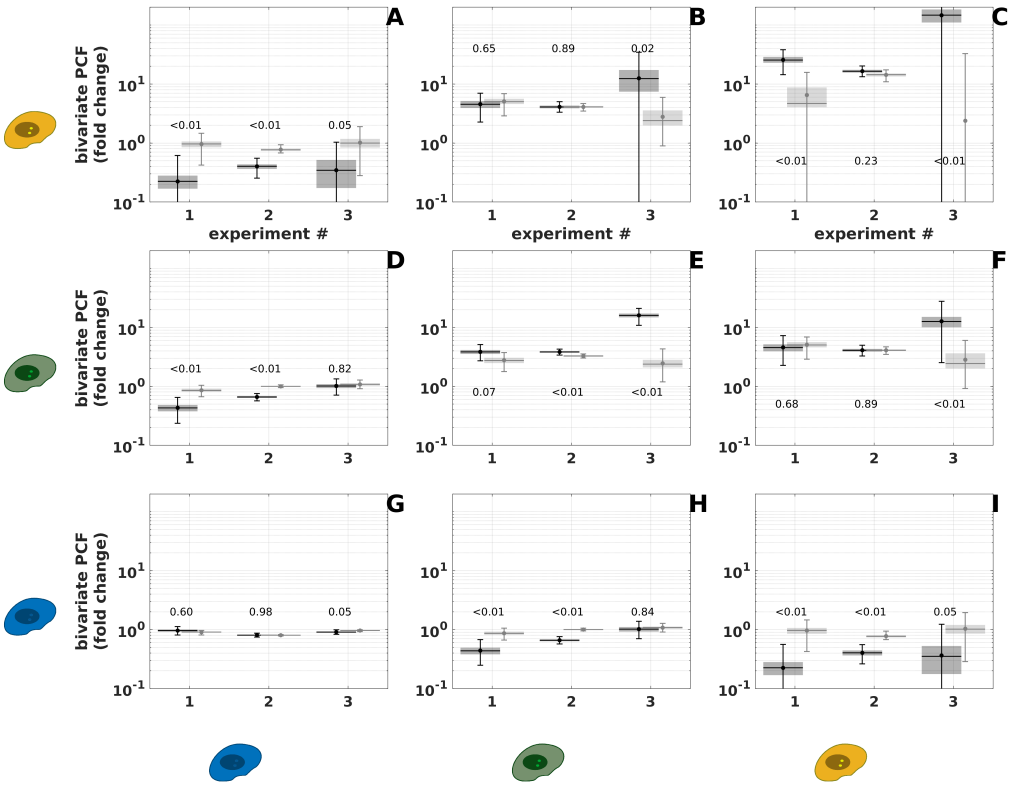


Figure 5—figure supplement 3. Repeatability of spatio-temporal simulations with MDA-MB-231 cells We report the ratio between the measured bivariate PCF and the one obtained for phenotype shuffling (control) for data (black) and simulations (gray). This ratio is estimated at $r = 15\mu m$ for which the shuffling PCF is maximum. Dot : actual ratio. Horizontal line : median obtained from bootstrap resampling. Gray shaded box : 50% confidence interval. Error bars : 99% confidence interval. p-values are shown for each data point. p-values are estimated by 500 repeats of simulations to test the null hypothesis (simulations identical to data within the 68% confidence interval obtained from bootstrap resampling). (**A**) Bivariate PCF measuring spatial correlation between CSC and CDC. (**B**) Bivariate PCF measuring spatial correlation between CSC and iCC. (**C**) Bivariate PCF measuring spatial auto-correlation of CSC. (**D**) Bivariate PCF measuring spatial correlation between iCC and CDC. (**E**) Bivariate PCF measuring spatial auto-correlation between iCC. (**F**) Bivariate PCF measuring spatial correlation between iCC and CSC. (**G**) Bivariate PCF measuring spatial auto-correlation of CDC. (**H**) Bivariate PCF measuring spatial correlation between CDC and iCC. (**I**) Bivariate PCF measuring spatial correlation between CDC and CSC.

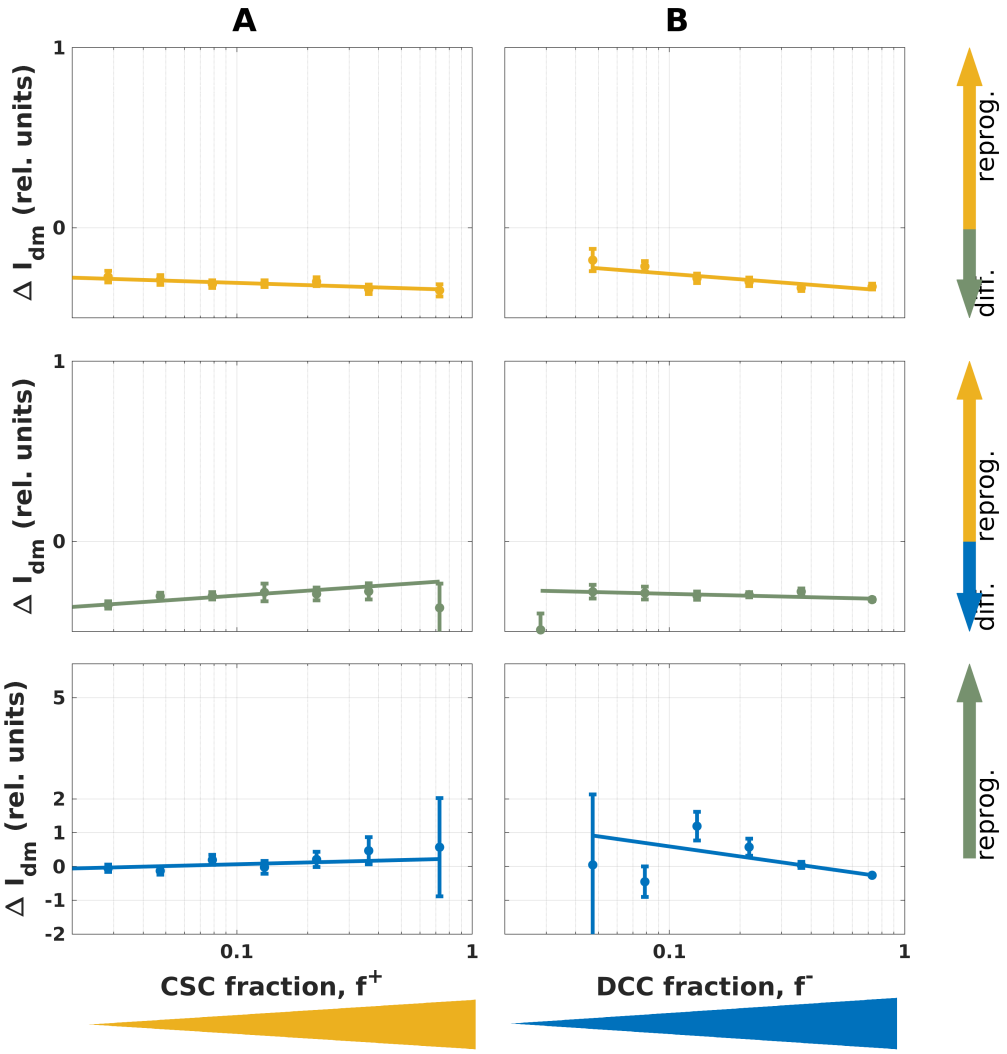


Figure 6—figure supplement 1. Influence of local environment on fluorescence variation upon mitosis in SUM59PT cells.

Average fluorescence variation upon mitosis, ΔI_{dm} conditional to either CSC fraction, f^{+}, (**A**) or CDC fraction, f^{-}, (**B**). From top to bottom data are shown for CSC (yellow), iCC (green) and CDC (blue). ΔI_{dm} is normalized to population average fluorescence intensity of CSC (~ 150 RFU), iCC (~ 42 RFU) and CDC (~ 2.5 RFU). Positive values for ΔI_{12} indicate differentiation and negative values indicate reprogramming. Each point represent conditional mean and the height of error bars two standard deviations.

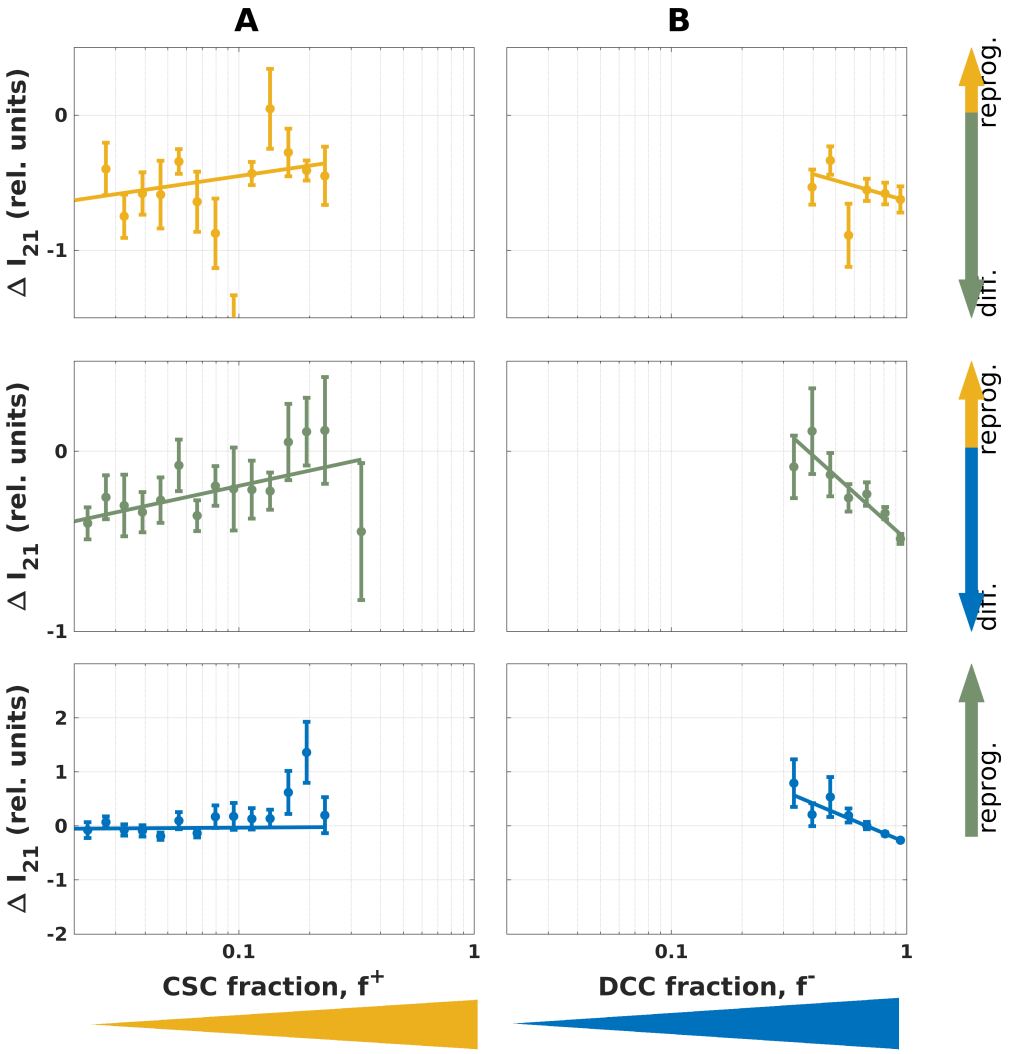


Figure 6—figure supplement 2. Influence of local environment on fluorescence variation during cell cycle in MDA-MB-231 cells

Average fluorescence variation during cell cycle, ΔI_{21} as function of the local CSC fraction, f^{+}, (**A**) or the localCDC fraction, f^{-}, (**B**). From top to bottom data are shown for CSC (yellow), iCC (green) and CDC (blue). ΔI_{21} is normalized to population average fluorescence intensity of CSC (~ 150 RFU), iCC (~ 42 RFU) and CDC (~ 2.5 RFU). Positive values for ΔI_{21} indicate differentiation and negative values indicate reprogramming. Each point represent conditional mean and the height of error bars two standard deviations.

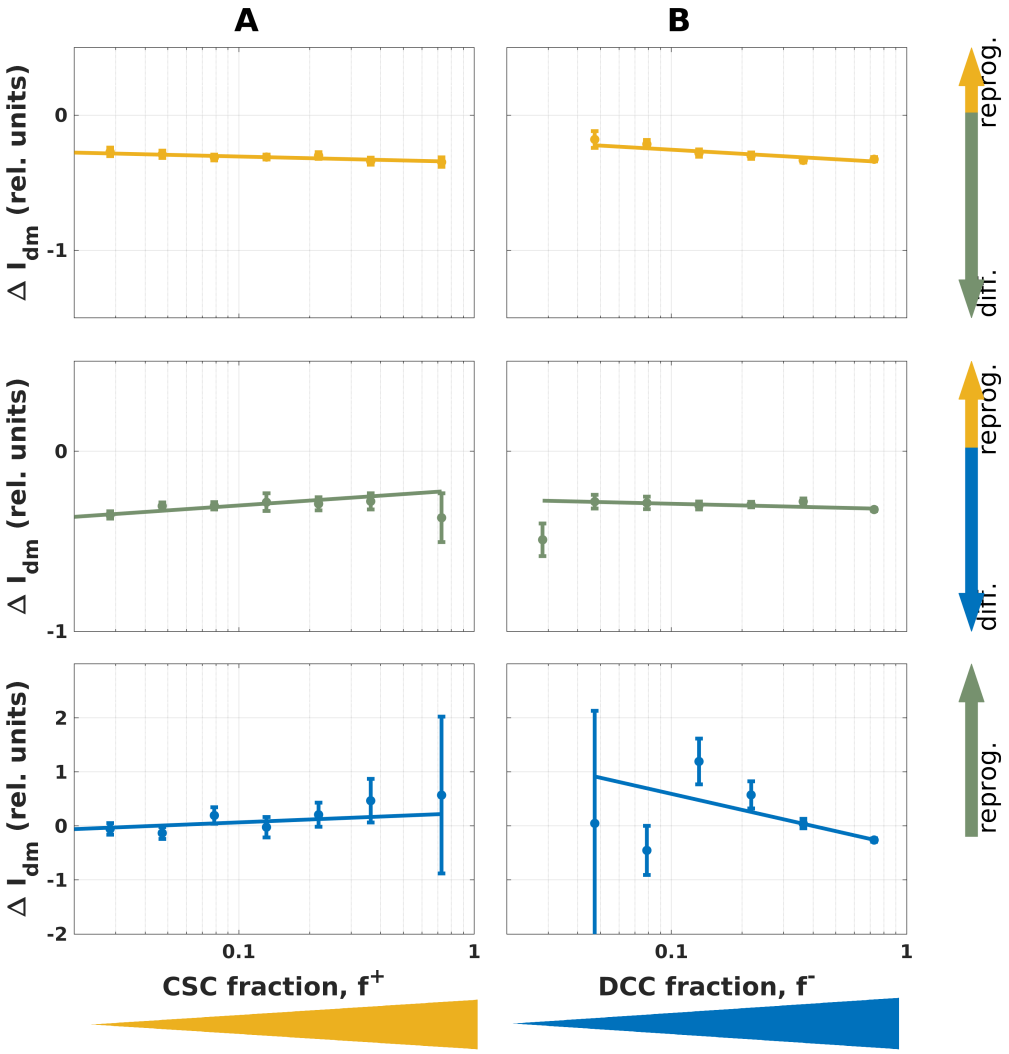


Figure 6—figure supplement 3. Influence of local environment on fluorescence variation upon mitosis in MDA-MB-231 cells

We report average fluorescence variation upon mitosis, ΔI_{dm} conditional to either CSC fraction, f^{+}, (**A**) or CDC fraction, f^{-}, (**B**). From top to bottom data are shown for CSC (yellow), iCC (green) and CDC (blue). ΔI_{dm} is normalized to population average fluorescence intensity of CSC (~ 150 RFU), iCC (~ 42 RFU) and CDC (~ 2.5 RFU). Positive values for ΔI_{12} indicate differentiation and negative values indicate reprogramming. Each point represent conditional mean and the height of error bars two standard deviations.