<u>**Supplementary Note 1**</u>

The following Supplementary Note provides technical details and rationale behind the metrics and algorithms used in the article "Temporal and spatial topography of cell proliferation in cancer." Specifically, this addendum addresses 1) the scoring of proliferation in fixed human tissues and how the multivariate proliferation index (MPI) compares to single channel gating of proliferation and cell cycle markers, and 2) the derivation of the cell-cycle difference with classical multidimensional scaling (ccD-CMD) representation and how it compares to three recently published time inference algorithms (SCORPIUS[1], Palantir[2], and Cyclum[3]).

1. Classification of proliferation and cell cycle in fixed tissues

1.1 Current tissue-based proliferation scoring metrics

In clinical settings and research studies that use tissues, cellular proliferation is typically measured by determining the "mitotic index" which is a count of the number of mitotic figures per defined area from hematoxylin and eosin (H&E) stained section or by scoring the expression of single protein markers assayed by immunohistochemistry (IHC). Most commonly Ki-67 is used in the clinical setting and additional markers of proliferation used in research studies include phospho-Rb and Cyclin D. Clinical scoring is often performed by a pathologist that visually inspects H&E slides or IHC using light microscopy and assigns an overall proliferation score per tissue specimen.

These approaches have a number of limitations. First, the manual scoring by pathologists is subject to human bias and is not highly quantitative or scalable. Second, single markers are not able to accurately capture proliferation. The markers currently used are biased toward specific parts of the cell cycle; for instance, Ki-67 and phospho-Rb preferentially stain cells that are in the S/G2 phase of the cell cycle, Cyclin D stains cells in the G1 phase, and the mitotic index identifies cells that are in M phase. These markers have a low false positive rate - they are unlikely to stain non-proliferating cells - but they miss proliferative cells from large portions of the cell cycle resulting in numerous false negatives. Other markers of proliferation, such as PCNA and MCM2 (which will be discussed at length in the next section) are more ubiquitously expressed during the cell cycle and are more stable proteins. However, they are often co-expressed with markers of cell cycle arrest (Fig. 1a,b and Extended Data Fig. 1b,c) and hence have a higher rate of false positives. Lastly, single stains do not allow one to discriminate between the various cell types that are present in the tumor areas. Both infiltrating stromal and immune cells proliferate within the tumor microenvironment, and they cannot be visually excluded from the scoring and hence confound proliferation measurements.

To overcome these limitations, we used tissue cyclic immunofluorescence (t-CyCIF) to measure up to 30 protein markers in the same tissue (Fig. 1, Extended Data Fig. 4). The antibody panels include both lineage specific markers (e.g., E-Cadherin, Pan-Cytokeratin, CD45, alpha-Smooth Muscle Actin, Vimentin) as well as proliferation, cell cycle arrest, and cell cycle regulation markers. The lineage markers are necessary to distinguish tumor epithelial cells from infiltrating stromal and immune cells. Once we computationally isolate the cells belonging to the tumor compartment, we use additional markers to calculate a multidimensional metric for proliferation, the Multivariate Proliferation Index, or MPI.

1.2 Multivariate Proliferation Index (MPI)

The Multivariate Proliferation Index (MPI) is a single-cell classification score for proliferation based on a panel of five markers. The MPI is a ternary classification system that categorizes each single epithelial cell into proliferative "+1", non-proliferative "0", or arrested "-1". The aim of the MPI is to provide a sensitive and specific measure of cancer cell proliferation that is not reliant on a single marker, while also identifying cells expressing high levels of arrest markers, even if they concurrently express markers of proliferation. The MPI incorporates information from three markers of proliferation (Ki-67, MCM2, and PCNA) and two markers of cell cycle arrest (p21 and p27).

The values of these 5 markers are compared to assign one of the three MPI labels to each single cell (note that the MPI is calculated using the normalized fluorescence intensity values, see the Methods section for the details regarding how we perform the normalization of cycling immunofluorescence data). A cell is defined as proliferative (MPI +1) if the overall balance of proliferative markers is positive and the cell does not express a cell cycle arrest marker. In the latter case, i.e. if a cell expresses high levels of either p21 or p27, the cell is classified as arrested and is assigned an MPI -1 label. If the cell expresses neither proliferation nor high levels of either cell cycle arrest markers, it is classified as MPI 0 (i.e., non-proliferative; please refer to the Method section for the exact formulaic definitions of the MPI categories).

1.2.1 MPI rationale and components

Using a multidimensional metric such as the MPI has multiple advantages. On a technical level, it is robust to experimental and biologic variability in marker detection. The proliferation and cell cycle arrest markers are partly redundant and overlapping, which reduces the likelihood of erroneous classification because a marker is not detected, or protein expression is absent due to genetic alteration. This is especially relevant in cancer tissues, in which uncontrolled proliferation is often achieved by avoiding cell cycle checkpoints or bypassing the need for an otherwise essential protein in cell cycle progression (for instance both Cyclin D and phospho-Rb are used to define proliferative cells but their expression is often altered among many cancers).

Individual markers have limitations, and hence the MPI markers were chosen to compensate for each other. Ki-67 preferentially marks cells in the G2/M phase of the cell cycle. While it has a clear cell-cycle bias, Ki-67 is also extremely specific to actively proliferating cells, and hence it is hardly ever co-expressed with cell cycle arrest markers p21 and p27 (Fig. 1a,b and Extended Data Fig. 1b,c). Notably, Ki-67 is not essential for proliferation; Ki-67 mutant mice develop normally and cells in culture lacking Ki-67 proliferate efficiently[4]. PCNA is a DNA clamp required for DNA replication, and MCM2 is a licensing factor that accumulates in G1 in pre-replication complexes[5]. Both are required for cell cycle progression and proliferation, and they are broadly expressed throughout the cell cycle. In addition, their protein half-life, estimated to be over 40 and 70 hours respectively[6], is much longer than that of Ki-67 (which is around 8 hours). Their regulation during cell cycle arrest has not been fully characterized, but studies have shown that p21 can trigger cell cycle arrest, in both G1 and G2 cell cycle phases, through a CDK-independent mechanism involving direct binding to PCNA[7,8]. As a result, we observe cells expressing high levels of MCM2 or PCNA concomitantly with high levels of p21 and p27, which are expected to be cell cycle arrested (Extended Data Fig. 1b,c). In conclusion, Ki-67 on its own would miss many proliferative cells, while MCM2 and PCNA would erroneously score arrested cells as proliferative. By combining these markers into a single metric, we overcome the limitations of single markers.

We chose to use p21 and p27 to mark cell cycle arrest. Both of these proteins inhibit the activity of cyclin-dependent kinases (CDKs) and actively trigger cell cycle arrest (rather than simply reporting on it). However, both p21 and p27 also play active roles in cell cycle progression[9,10]. It would be hence incorrect to define cells as being arrested when they express moderate levels of these proteins. The normalization procedure we utilize centers the positive distribution around +0.5 and thus we use this as a threshold for calling cells as arrested. Given the mutual exclusivity of p21 and p27 with Ki-67, the threshold for calling a cell arrested (MPI -1) can be adjusted accordingly, usually within the range of 0.25-0.75 (importantly however, within an imaging experiment involving multiple samples or tissue microarrays, both the normalization and thresholds are changed for all of the samples in the experiment).

1.2.2 Comparison between the MPI and other multidimensional classification strategies

In Fig. 1c and Extended Data Fig. 1d and 1e, we show a direct comparison between MPI calls and UMAP, k-means clustering and t-distributed stochastic neighbor embedding (t-SNE), three commonly

used tools to classify multidimensional datasets from various multiplex techniques (scRNAseq, CyTOF, MIBI, IMC, CODEX). The MPI ternary assignment is completely recapitulated in the UMAP output (Extended Data Fig. 1e). The concordance between the MPI single cell calls and k-means clustering is evident by the fact that the majority of clusters identified by the k-means algorithm consist mainly of a single MPI category (Extended Data Fig. 1d). K-means clustering still involves manual selection of the number of clusters and manual definition of each cluster's identity, which would have to be inspected for each experiment individually. Further, while the MPI definition is deterministic, clustering has a stochastic component that changes the cluster assignment every time the algorithm is run, which makes it not fully reproducible and less scalable.

The t-SNE results are also concordant with the MPI classification, with MPI cells from different categories segregating away from each other (Fig. 1c; the MPI was not used as an input for the t-SNE algorithm). Classifying cells on the t-SNE representation still requires running a clustering algorithm on the t-SNE output (DBSCAN is often used for this purpose), which leads to similar problems as mentioned above for k-means. The major drawback of t-SNE and similar dimensionality reduction algorithms is that they are not computationally scalable and cannot be run on the large datasets generated in our study (in the order of hundreds of thousands to millions of cells). While machine learning techniques exist that permit building classification models from subsets of the data, using these were beyond the scope of this study and did not present clear advantages over the single cell MPI deterministic classification strategies.

1.2.3 Robustness of MPI assignment across adjacent tissue sections

To test whether the MPI assignment was robust across serial sections of the same tissue samples, we performed the MPI marker staining and analysis on 5 sections from a commercial breast tissue microarray (TMA, Pantomics BRC15010). From the epithelial cells in each core, we calculated the MPI fractions and compared the fraction across different sections from the same core. The correlation between sections was extremely high ($R^2 > 0.92$ for all pairwise comparisons of MPI +1 fractions, Extended Data Fig. 1f). Similar results were observed when serial sections from whole slide breast carcinoma sections were analyzed (Extended Data Fig. 1g). Additionally, we calculated the coefficient of variation (CV) across the 5 sections of the breast TMA for the MPI+ 1 fraction; the median value across 74 samples was CV = 0.14 with maximum value of max(CV) = 0.47 (Fig. 1d). In the same dataset, the assessment of proliferation by the Ki-67 positive fraction (median(CV) = 0.3, max(CV) = 0.99) or by phospho-Rb (median(CV) = 0.34, max(CV) = 1.71) showed much lower stability. This indicates that the assessment of proliferation by the MPI is extremely robust to biological replication, whereas single marker measurements provide a much less certain and much more variable metric for assessing proliferation.

1.2.4 Stress-testing MPI marker normalization and MPI assignment

The assignment of the MPI categories depends on the process of normalization of the MPI markers (Ki-67, PCNA, MCM2, p21 and p27). To normalize the data, we employed a two-component Gaussian mixture model (GMM) on single channel distributions. GMM's are commonly used to separate positive and negative populations within fluorescent datasets due to the log-normality of fluorescence measurements. When single markers are used to assign a proliferative index (e.g., Ki-67, phospho-Rb, or phospho-histone H3 which marks mitotic cells) this process is equivalent to using the GMM to select a single threshold because the quantitative amount of positivity is not taken into account, effectively binarizing the data. This leads to a substantial loss of information and it is error prone (see Section 1.2.3 above and Section 2.1 below for more details). Instead, the MPI strategy quantitatively integrates the information from multiple markers, buffering against imprecision in choosing the normalization parameters.

Despite using multiple markers to buffer against errors in single channels normalization, we strived to measure the robustness of the MPI calls to errors in the normalization procedure. In a cohort of 180

breast tumors, we stress-tested the MPI assignment by adding random noise to all five MPI channel distributions and recalculated the MPI scores, repeating this procedure 1000 times. The insertion of noise mimics a random error in the selection of the normalization parameters and hence we modelled the errors as independently normally distributed random variables with zero mean and increasing standard deviation; we started with the errors having 0.1x of the standard deviation (i.e., 10% of the width of the original distribution), and increased the error width up to 1x of standard deviation of the original distribution. This upper limit is an unrealistic extreme as it assumes that the noise distribution is essentially as wide as the original signal. We then assessed the effect of the noise addition in three ways: 1) by linear regression (R^2) of the original MPI score with the noise-added scores (Extended Data Fig. 1h), 2) by Pearson correlation between the two scores (Extended Data Fig. 1i), and 3) by assessing how the ordering between patient samples changed due to the addition of noise (Extended Data Fig. 1j). All three measures show that the MPI measurements are strongly robust to the addition of random noise, with both the R^2 and the Pearson correlation being close to 1 even with the noise being half as wide as the original distribution (in the Extended Data Fig. 1h-j panels this is the 0.5x standard deviation addition). Only when the noise was essentially equal to the signal (1x standard deviation) did the MPI measurements diverge dramatically. From this stress-testing procedure we conclude that the MPI calls are robust to potential errors in the MPI channel normalization.

2. Cell cycle inference from multiplexed imaging data

2.1 Analysis of the cell cycle by combinatorial gating

2.1.1 Single channel gating in t-CyCIF datasets

In single channel immunofluorescence datasets, a conventional strategy for identifying cells expressing a proliferation or cell cycle marker is to visually inspect the single cell distribution and select a threshold or "gate". All cells whose fluorescence signal value is above the threshold are considered "positive" and the others are defined as "negative". The binarization of the data leads to substantial information loss but simplifies the data analysis. The gating strategy assumes that the single cell distribution is bimodal or has a clear positive or negative tail. In the t-CyCIF datasets we produce, lineage markers often have clear bimodal single cell distributions, as shown for example for E-Cadherin, Vimentin and alpha Smooth Muscle Actin (αSMA) in Extended Data Fig. 3c. The bimodality enables us to use the single channel gating strategy to perform basic cell type calling and to isolate epithelial cells away from stromal and immune cells. For example, in the breast tissues shown in Extended Data Fig. 3a-d, epithelial cells were defined as cells positive for E-Cadherin and negative for stromal and immune markers. In our work, the numerical value of the lineage marker expression is still used to resolve "conflicts," i.e., instances for which multiple lineage markers are found to be positive. When these conflicts arise, the marker with the higher normalized value "wins" and is used to define the cell type for the cell in question.

In the case of non-lineage markers, such as the proliferation and cell cycle markers, the single cell distributions from t-CyCIF are generally not bimodal (Extended Data Fig. 3c, lower panels). While the single-cell fluorescent signals are strong and the contrast between nuclear and cytoplasmic signals is evident (Fig. 1b and Extended Data Fig. 3a), the distribution of single cells across tissues is continuous and does not have a strong skew. Inspection of two-channel comparisons with scatter plots shows that the expected relationships between markers, based on the published literature, hold true (Extended Data Fig. 3b). For example, CDT1 and Geminin are generally mutually exclusive, G2 markers like Cyclin A1/2, Geminin, and phospho-Rb are positively correlated with each other and negatively correlated with G1 markers such as Cyclin D1 and p21. To directly confirm that the t-CyCIF protocol is comparable to established immunofluorescence (IF) measurements, we performed standard indirect IF (primary antibody staining followed by secondary fluorescently labelled detection) on five sections of the Pantomics BRC15010 breast tissue microarray and performed the t-CyCIF

protocol on the same slides after the indirect IF. Using this approach, we can compare protein detection using the t-CyCIF multiplexed protocol with the indirect IF within the exact same cells. Extended Data Fig. 4b shows a high degree of correlation between the two measurements. When comparing the overall core detection, we also noted that the t-CyCIF signal while having decreased overall signal provided increase dynamical range (Extended Data Fig. 4c). For instance, when measuring the overall phospho-Rb signal, we observed that the signal was the lowest in non-proliferating cells (MPI 0), increased in proliferative cells early in the cell cycle (MPI +1, Ki-67-), and was highest in proliferating cells later in the cell cycle (MPI +1, Ki-67+) (Extended Data Fig. 4c). While this pattern was observed using both indirect IF and in the t-CyCIF experiment (i.e., direct IF), in the latter the difference between the three bins was more marked. Together with previous validation studies, these correlations serve to increase the confidence that the antibodies used to recognize these cell cycle proteins are indeed binding the intended epitopes.

2.1.2 Lack of accurate quantification of total DNA and DNA replication

A unique limitation in the study of human tissues is the unavailability of two key pieces of information: the quantification of DNA amount and of active DNA replication. DNA staining by intercalating agents such as Hoechst is routinely performed in tissue imaging, and the t-CyCIF method relies on it to align the rounds of imaging to obtain highly multiplexed single-cell datasets. However, the DNA quantification from tissue images does not appear to be an accurate representation of DNA content. Single cell distributions of DNA content from tissue are not bimodal (Extended Data Fig. 4a) which they would be expected to be, and how they appear from quantification of CyCIF images acquired from cells grown in tissue culture (Extended Data Fig. 5a). In addition, genomic instability and aneuploidy are a common feature in human tumors and can vary widely between tumor cells within the same tumor mass, making the interpretation of DNA quantification in solid tumors additionally challenging. Similarly, the detection of DNA replication - usually performed by EdU incorporation in cell culture - is not technically possible in fixed human tissues.

2.1.3 Multichannel gating to infer cell cycle distributions

In order to attempt to understand the cell cycle positions of cells based on a gating approach, we gated single channels and looked at the combinatorial patterns of positivity for multiple channels, or "multichannel gating" (Extended Data Fig. 3d). Binarizing an 8-cell cycle marker panel by gating produces $2^8 = 256$ combinatorial states, and in Extended Data Fig. 3d we show a visual representation of the top 50 most frequently occurring states (from the same breast sample we used for the single distribution and two-channel scatters in panel c. of the same figure). We used the Upset plot package (https://gehlenborglab.shinyapps.io/upsetr/) to display the combination of positive markers in each of the 50 states and the relative abundance of each state. While the abundance of these states is technically quantifiable, it is important to remember that in the absence of bimodal single marker distributions, the choice of gating threshold is somewhat arbitrary. Small changes in any of the cutoffs would drastically modify the relative abundances of all the states.

In the absence of information about the ploidy and DNA replication state of the cells, discriminating the combinatorial states generated by multichannel gating is prohibitive. In cell culture experiments, both DNA amount and EdU incorporation have bimodal distributions, the combination of which clearly defines the G1 (2N, EdU-), S (EdU+) and G2/M (4N, EdU-) phases of the cell cycle. This information is instrumental to define cell cycle states because the dynamics of cyclin expression (as well as CDT1, Geminin and phospho-Rb) are more graded through the cell cycle. In addition, while the dynamics of single proteins through the cell cycle are well-characterized, it is still unclear how they relate to each other. For example, the levels of Cyclin D1 decrease from G1 to S/G2, but how does this quantitatively relate to the increase in Geminin? Similarly, how does the rise of G2-phase cyclins like Cyclin A1/2 relate to the dynamics of decrease in CDT1? Without firm answers to these questions, we are not able classify the multichannel gating data in Extended Data Fig. 3d.

In conclusion, the loss of information that results from binarized gating of the continuous distributions did not simplify the classification problem; it only changed the nature of the problem from continuous to discrete. Thus, because of 1) the lack of interpretability, and 2) the lack of robustness in state abundances, we concluded that a gating strategy would be insufficient for the cell cycle analysis (specifically in our multiplexed tissue imaging datasets). Instead, we pivoted to methods that use the full continuum of markers expression values to create unsupervised metrics for cell cycle analysis.

2.2 Time inference

Time inference is a computational method to model dynamic processes using data generated from a static system (also called trajectory inference). Several approaches have been published to infer temporal dynamics from fixed time data involving imaging from cell lines[11,12] and single-cell RNA sequencing data[1–3,13]. To develop an analogous approach for multiplexed tissue images, we made the following assumptions: 1) measured markers provide coverage of multiple cell cycle transitions, 2) sampling is sufficiently uniform and dense for ergodic theory to be applicable, and 3) there exists a dynamical system governing changes in expression of measured cell cycle proteins. Similar assumptions are needed to apply any time inference framework to static data. If these assumptions are met, the time inference method approximates time series data by inferring an ordering of cells that can be translated into pseudotime.

Time inference algorithms can be implemented in various ways but generally consists of two main steps: 1) reducing high-dimensional data into a low-dimensional representation, and 2) constructing a trajectory model of the cells in low-dimensional space and projecting the cells onto that trajectory, thereby producing a pseudotime ordering of cells. Time inference algorithms usually rely on unsupervised machine learning methods, which have the advantage of being relatively unbiased. By not biasing the ordering with prior knowledge (e.g., protein or gene function), the algorithm can potentially offer novel insights. For canonical presentations of well-understood processes, one might actually achieve a more accurate ordering using a manual approach based on prior knowledge (e.g., ordering cells by DNA content for cell cycle ordering). However, for less understood processes, over-reliance on pre-existing knowledge might lead to a failure to detect biologically significant underlying structures in the data.

Time inference algorithms, however, are not completely unbiased. First, the design of the algorithm usually reflects some assumption about the underlying structure of the data trajectory (e.g., connected vs. unconnected, linear vs. bifurcating). In some algorithms, the user selects a subset of markers to include as input, which generally reflects the assumption that these markers are relevant to the biological process of interest. Finally, some algorithms require a user-defined "start" cell, which again requires prior knowledge and therefore introduces potential bias.

The time inference method we developed for this paper was inspired by previously developed time inference methods but differs from previous methods in that it is designed for multiplexed tissue images and assumes a latent cyclical representation of the data that reflects the cyclical nature of cell cycle progression. Additionally, our method does not rely on DNA or EdU content as inputs; in fact, its aim was not to be highly dependent on any single feature. In human tumor tissue samples, DNA content is difficult to measure robustly, measurement of EdU incorporation cannot be performed, and any individual cell cycle marker could be lost to genetic instability or aberrant epigenetic control. We did, however, use DNA quantification to validate our computational method for time ordering in cell culture experiments (Extended Data Fig. 5).

The following section describes the time inference algorithm we used in our work and compares its output to several existing time inference algorithms.

2.3 Pairwise cell cycle difference (ccD) and classical multidimensional scaling (CMD)

2.3.1 Description of input

For this paper, we constructed trajectories of cell cycle progression from single cell immunofluorescence measurements obtained from fixed tissue images. We utilized the MPI classification to isolate epithelial proliferating cells (i.e., MPI +1 cells), and we used only those cells as input. For the cell culture experiments all the cells measured are used in input, because the cell composition is homogeneous, and the majority of cells were MPI +1. Each cell is represented by a cell cycle marker vector of normalized values for markers of cell cycle proliferation (Ki-67, PCNA, MCM2), cell cycle arrest (p21, p27), and cell cycle progression (phospho-Rb Ser807/811, CDT1, Geminin, and Cyclins A1/2, B1, D1, and E1).

The markers for the ccD-CMD analysis were chosen to represent multiple cell cycle transitions. For instance, p21, CDT1 and Cyclin D1 cover the transition through G1 and into S-phase, and phospho-Rb, Ki-67, Cyclin A1/2 and B1 cover the passage from S-phase into early and then late G2 phase. Although PCNA and MCM2 are used to calculate the MPI, they were not routinely used in the ccD-CMD trajectory inference because their variability within MPI+1 cells is minimal; therefore, they do not provide additional information and potentially add spurious noise. This reasoning also holds true for markers that are highly specific but are only expressed in a small subset of cells (e.g., phospho histone H3 which marks M phase). It is important to note that not all of these markers were usable for every dataset, especially in tissue-based experiments where the staining variability can be high and where individual tumors might genetically or epigenetically lose the expression of single proteins (see Supplementary Table 2 for details of which markers were used for the ccD-CMD calculations in each experiment).

2.3.2 Dimensionality reduction

The ccD-CMD analysis starts by calculating the cell cycle difference (ccD) matrix, which is defined as the absolute value of the pairwise Pearson correlation between the cell cycle marker vector of normalized values of each cell. An example of ccD matrix is presented in Fig. 3c.

In order to be interpreted, the ccD matrix needs to be reduced in dimensions. Classical multidimensional scaling (CMD[14]) is a linear dimensionality reduction method used to approximate the pairwise distances between 'n' points (in our case n = number of single cells) to a representation in lower dimensions. Commonly, reduction to two dimensions is chosen to ease visualization and interpretation. CMD scaling is a linear algorithm, and although assuming linearity is an oversimplification, CMD scaling runs significantly faster than non-linear methods, which is important for scalability as we often need to perform the algorithm on hundreds of samples with upwards of 20,000 cells or tens of samples with over a million cells apiece.

2.3.3 Circular fitting: trajectory model and cell ordering

The reduced two-dimension ccD scatter, referred to as the "ccD-CMD" representation, or landscape, is parametrized by fitting it to a circle by least-squares minimization. This choice was made based on the observation that the two-dimensional representation has an underlying cyclical structure following the dynamics of the cell cycle. This is in contrast with many other time inference algorithms that search for non-cyclical paths through the data.

The circular fitting served two purposes:

1. to perform ordering of the cells around the cell cycle position. Notably the ordering is most accurate only for populations of cells where the ccD-CMD representation forms an evenly distributed torus (referred to as "cell cycle coherent" in the main text), rather than a skewed torus or amorphous point cloud (referred to respectively as "skewed cell cycle" and "non-canonical").

2. to parametrize the ccD-CMD representation and extract quantitative metrics that summarize a sample's overall cell cycle temporal organization.

For each data point (i.e., single cell), two parameters are calculated, 1) the shortest distance between the data point and the fitted circle (d) and 2) the angle relative to the point of origin of the fitted circle (θ) as shown in Fig. 3f. Using the angle, each data point was projected to the nearest point on the circle to order the cells in what is referred to as the "cell cycle ordering". Given the cyclical nature of the ordering, the point of origin (time = 0 or start time) is arbitrary. For each population of cells being analyzed, two overall metrics were extracted from the fit (Fig. 3f):

i) the circle fit distance (**CFD**), which is the mean d of the population,

ii) the inter-octile variation in angle (**IOV**). To calculate the IOV, the angle measurements were binned to 8 equally sized bins (of 45 degrees or $\pi/4$ radiant) and the proportion of cells in each bin were calculated. To ensure lack of positional bias, the proportion calculation was repeated by shifting the bin position by $\pi/8$. The IOV is the coefficient of variation of the bin proportions, which is equal to 0 in a uniformly distributed population.

2.3.4 Measures of cell cycle coherence

The two metrics, CFD and IOV, were used to describe what we call the "cell cycle coherence" of a sampled cell population. A low CFD indicates that the points in CMD space are distributed along the circular path (i.e., the best-fit circle). As more and more cells accumulate in the center of the ccD-CMD representation (forming a cloud of points) the CFD metric increases. In the ccD-CMD representation, the 2D distance between two points is proportional to their similarity in multidimensional space, in our case the cell cycle space. Therefore, a cloud of points means that many cells are just as similar to each other as to many other cells, i.e. there is no coordination between cell cycle markers (e.g., Sample 3, Fig. 4c). If the measured cells are random samples from a deterministically oscillating system, the distance between the cells would be proportional to the time between the positions of the random samplings. In this case, the data points would form a topological toroidal shape and in turn, a low CFD. In Fig. 5h,i, we show how a sharp drop in HER2 expression led to a shift from a toroidal ccD-CMD representation (9 weeks HER2 "on") to a point cloud (2 days HER2 "off"), with a corresponding increase in CFD (from 30 to 50). Our interpretation is that HER2 expression promotes a strong signal for cells to grow; once this is withdrawn, the system drifts in multiple directions and the levels of the cell cycle markers lose coordination.

A low IOV indicates that points are distributed evenly around the circle as opposed to clustered at specific regions. This clustering occurs if cells slow down in a specific part of the cell cycle, hence accumulating at that location. Notably, a similar clustering would occur if cells were to accelerate through a specific part of the cell cycle, but in this case, the accumulation of cells would occur away from the acceleration point. In both instances, the population of cells would be unevenly distributed and have high IOV, a state that we call "skewed cell cycle" because of the uneven distribution of cells around the ccD-CMD torus. The extreme case scenario is cell cycle arrest, for instance triggered in MCF10A and MCF-7 cells by cell cycle inhibitors (Fig. 3g-l). For example, following treatment with palbociclib or nocodazole for either 24 or 48 hours, the recorded IOV is up to three-fold higher than freely cycling MCF10A cells. However, these perturbations are extreme scenarios. In human samples, we focus on the MPI +1 population of cells and eliminate from the analysis the most likely arrested population (MPI -1, with high levels of p21 or p27). Accordingly, most human tumors we measured had IOV between 0.4 and 1.2, while the fully arrested cells have IOV of 1.5-2. The high IOV state potentially represents populations of cells moving towards cell cycle arrest. However, as described above, high IOV could also suggest a more streamlined cell cycle with a shortening in one cell cycle phase and in fact a faster, albeit still less balanced, completion of the cell cycle.

2.3.5 Interpretation of the ccD-CMD representation: closed torus vs point cloud

In the main text, we explain how to make inferences about alterations in cell cycle dynamics from the arrangements of single cell data points in the ccD-CMD representation. Both the cell line experiments, and the tumor sample data catch the two extremes of the ccD-CMD representation, a closed torus and a densely aggregated point cloud.

A general question is whether the current understanding of cell cycle protein dynamics would lead us to expect a torus representation. While the cell cycle is by definition a periodically recurring dynamical system, it is often conceptualized as a linear path starting and ending at cell division. Moreover, the linear path is further abstracted as a sequence of phases (G0, G1, S, G2 and M), separated by discrete molecular events, such as the passing of restriction points. In multidimensional space, these conceptualizations could result in a clustered and disconnected set of point clouds (each of which represent a discrete sub-state in the cell cycle sequence) rather than a continuous topology.

Reasons for a continuous torus: In all the samples we assessed, the ccD-CMD representation did not produce a clustered and disconnected topology. The lack of a clustered and disconnected topology is likely due to both biological and technical reasons. One technical reason is experimental noise, which could blur the edge of distinct clusters, resulting in continuous state transitions. Another technical reason for lack of distinct clusters is that the ccD-CMD population measurement captures a 'mean' cell cycle state, and as a result differences between cell cycle states may be de-emphasized.

However, an important biological reason for a continuum of cell cycle states is that the fluctuation of cell cycle proteins is more gradual than the abstracted sequential and distinct view of the cell cycle would imply. The fluctuations in most proteins we measure (cyclin A1/2, B1, D1, E1, Geminin and CDT1) appear to cross smoothly over multiple adjacent cell cycle phases. In the multiplexed protein space, these gradual changes produce a continuous distribution of cell states that are then experimentally sampled, producing a continuous torus in the ccD-CMD representation.

Reasons for a closed torus: Similar reasoning also applies to explaining why the torus is closed (or nearly closed). During M phase cellular structures are disrupted, the cellular components subdivided in two, and cell division is completed. From an imaging point of view, M phase transitions are the most difficult to capture; cyclic multiplex imaging relies on DNA to identify and outline the nucleus of cells, and when the DNA is compacted the nucleus temporarily "disappears" from the image. However, cell division results in an ~2-fold decrease in protein abundance, which is an order of magnitude lower than the fluctuations that many cell-cycle regulated proteins undergo during the entire cell cycle. In addition, during the cell cycle the protein changes can be uncorrelated, while during cell division the whole proteome drops in synchrony. The ccD-CMD algorithm uses the absolute value of the correlation as the distance metric to calculate the cell cycle difference (the "ccD" part of the algorithm), which focuses on discriminating coordinated changes of protein expression. For these reasons, cells on either side of cell division (i.e., pre-division and post-division) are closer to each other in the ccD-CMD space than cells in other cell cycle phases, leading to the torus shape appearing closed. No extra data manipulation step was undertaken to achieve or bias the representation into forming a closed structure.

2.3.6 Robustness of the ccD-CMD algorithm and coherence metrics

The ccD-CMD algorithm does not have a stochastic component and hence it is an injective function with one-on-one mapping from input to output. However, its application to estimate the coherence metrics IOV and CFD is sensitive to two variables: 1) the number of cells from the tissue used in input, and 2) the markers used to calculate the cell cycle difference (ccD).

Because the ccD is a quantification of the binary distances between pairs of cells, the computing time to calculate the ccD increases non-linearly with the number of cells provided in input. It hence becomes prohibitive to run the ccD-CMD computation on more than a few tens of thousands of cells. On the other hand, when using tissue microarray cores, the number of cells available for analysis is

relatively limited. For both these reasons it is important to estimate how sensitive the performance of the algorithm is to the number of cells used as input. In Extended Data Fig. 5i we calculated the CFD and IOV parameters in 5 tissues using an increasing number of cells ("n" from 50 to 2,000, breast cancer tissue ROIs from Fig. 2 and Extended Data Fig. 7 with at least 20,000 MPI+1 cells used for this analysis). For each 'n', we run the ccD-CMD algorithm 40 times with a different set of 'n' cells from the same tissue and calculate the coefficient of variation (CV) for the IOV and CFD metrics. For both metrics the CV quickly decreased and reached a plateau. This shows that for whole tissue experiments using >2,000 cells would not provide a great increase in precision. In addition, when cell numbers are limited, using at least 500-1,000 cells is required for an acceptable estimate.

Next, we tested the dependency on single markers on the coherence metrics, by eliminating one or two markers from ccD calculation (Extended Data Fig. 5j). For this we used the MCF10A cell line dataset, for which we have experimental controls obtained by perturbing the cell cycle with palbociclib and nocodazole (Fig. 3g-j). Eliminating one marker (out of 10 in the panel) had limited effect, comparable to the biologic variability of the three replicates in Fig. 3j. Eliminating two markers did not produce much larger displacement from the whole-panel estimates (green dot) than the single marker removal. Furthermore, the direction of the changes is spread in all directions of the IOV-CFD phase plane, suggesting that addition of markers converges to a central estimate, the green dot. In conclusion, these results show that the ccD-CMD algorithm is robust to the removal of one or two markers, especially when compared to experimental perturbations.

Finally, the ccD-CMD algorithm is run using only proliferative cells (MPI +1) in order to isolate cells that are actively cycling before assaying the cell cycle dynamics. As a conceptual validation, we characterized the effect on the ccD-CMD metrics of adding cells classified as MPI 0 or MPI -1 to the analysis. To assay this effect, we calculated the ccD-CMD parameters (IOV and CFD) for populations of MPI +1 cells, then added increasing fractions of non-MPI +1 cells from the same tissues, starting from a mix of 90% MPI +1 cells with 10% MPI=0 or MPI -1 cells (i.e., '10% ratio) and ending with 40% MPI +1 cells and 60% MPI 0 or MPI -1 cells (i.e., '60% ratio) (Extended Data Fig. 5k). To understand the effect of different non-MPI +1 populations, we mixed MPI+1 cells with MPI 0, 'MPI -1 p21+' and 'MPI -1 p27+' cells separately. In the case of MPI -1 cells, the hypothesis is that mixing cells from the MPI -1 group would increase the IOV similar to the effect seen in cell lines treated with cell cycle inhibitors. Because the mechanism of arrest that is driven by the two cell cycle arrest markers is part of the MPI calculations (p21 and p27) they each might result in somewhat different molecular signature. Thus, we tested the addition of 'MPI -1 p21+' and 'MPI -1 p27+' cells to the MPI +1 cells separately. In all cases, the increasing addition of arrested cells increased the IOV as expected (Extended Data Fig. 5k, red and magenta lines). Similar addition of MPI 0 cells leads to a decrease in coherence, as expected, but the direction of change in the IOV-CFD phase plane was more variable; in some cases, the CFD only increased while in other cases both the CFD and the IOV increased (Extended Data Fig. 5k, blue lines). We postulate that these different responses occur for two reasons: i) cells in the non-proliferative state MPI 0 either do not express any cell cycle markers, probably having fully exited the proliferative state, or ii) still express markers congruent with an early G1/G0 state. When all cell cycle markers are low their patterned coordination does not resemble any part of the cell cycle dynamics and in turn the CFD increases. Instead, when cells have more recently exited the cell cycle into quiescence, the marker pattern resembles a specific cell cycle phase (early G1/G0) and the IOV increases.

2.3.7 CDT1-Geminin dynamics

In cell culture, the dynamics of the cell cycle are commonly monitored using the FUCCI system[15]. This combines tagged versions of the proteins CDT1 and Geminin to track the progress of live single cells through the cell cycle. To further validate the time dynamics derived from the ccD-CMD algorithm, we measured the reconstructed dynamics of the endogenous CDT1 and Geminin proteins in exemplar tissue areas (Extended Data Fig. 5l). The CDT1 protein is known to function both in M

and G1 phase, and it is downregulated in G2 to stop DNA replication and prevent DNA re-replication. Geminin rises through S and G2 phases and it inhibits CDT1 leading to its degradation. In the cell cycle dynamics reconstructed from tissue, the endogenous CDT1 and Geminin proteins are clearly anti-correlated (Extended Data Fig. 3b and 5l). Geminin levels remained constant and low in the first part of the cell cycle and rose in the middle of the cycle and dropped back at the end of the cycle. Interestingly, the CDT1 dynamics differed slightly in different tissue areas: the CDT1 protein levels either i) peaked at the beginning of the cell cycle, decreasing slowly through G1 and then decreasing sharply in correspondence to the rise of Geminin, or ii) increased through the beginning of the cell cycle and then decreased sharply as Geminin rose. Hence, the reconstructed cell cycle dynamics from the ccD-CMD algorithm are consistent with the known biology of the FUCCI system. However, both the single cell variability and the quantitatively different dynamics observed in the different tissue areas precludes the use of CDT1 and Geminin alone to reconstruct the cell cycle dynamics from static tissue images.

2.4 Comparison of time inference methods across datasets

2.4.1 Description of existing methods

A large number of algorithms have been published to computationally infer ordering of cells based on multidimensional data. The vast majority of the algorithms have been developed to analyze genomics data rather than information from proteomics or antibody-based imaging technologies. We selected a subset of commonly used time inference algorithms, processed our multiplexed tissue imaging data, and compared the results with the ccD-CMD representation and pseudotime ordering from the same datasets. We used the following three algorithms, all originally developed to process single-cell RNA sequencing (scRNA-seq) data:

- **SCORPIUS**[1] calculates a difference matrix based on the entire single cell transcriptome and performs dimensionality reduction. The data in the reduced-dimensional space is clustered into k clusters via k-means clustering. The cluster centroids are connected through a shortest path, and the path is then optimized using the Hastie and Stuetzle principal curve algorithm, which iteratively smooths the trajectory until convergence. The cells are projected onto the trajectory to obtain the cell ordering. The rationale of SCORPIUS is similar to the ccD-CMD, but the difference matrix is different, and the trajectory inference is designed to find a non-cyclical path through the data.
- **Palantir**[2] is used to model cell differentiation into terminal differentiated states. It takes as input scRNA-seq data and projects the data onto diffusion maps to construct a nearest-neighbor graph. The pseudotime for each cell is determined using the shortest path between the cell and a user-defined early cell. As part of the Palantir workflow, the t-SNE algorithm is used to visualize the diffusion data in a two-dimensional plot.
- **Cyclum**[3] is a recently developed neural network algorithm that uses an autoencoder to project cells onto a non-linear, periodic trajectory.

Among the algorithms designed to infer time trajectories from static data a notable mention is **Cycler**[12], an algorithm designed to extract cell cycle information from images of fixed cells. Cycler uses single cell measurements of DNA content, DNA replication and pattern, nuclear area, and local cell crowding. It constructs an ensemble of l-out-of-k-nearest neighbor graphs, calculates the cell trajectory for each graph separately, and makes use of waypoints to construct the final trajectory. In its current version, Cycler requires quantification of DNA amount and DNA replication, which are not available from tissue imaging experiments (see section 2.1 above for a detailed discussion on this subject). Hence, we were not able to process the t-CyCIF data through Cycler.

2.4.2 Comparison of methods across datasets

We ran the ccD-CMD, SCORPIUS, Palantir and Cyclum on three exemplar CyCIF datasets from different experimental sources:

1) on tissue-based CyCIF data from a breast tumor tissue sample from Fig. 2a

2) on plate-based CyCIF data from unperturbed MCF10A cells

3) on synthetic data generated from a mathematical model of the mammalian cell cycle[16]

The results of the comparisons are shown in Extended Data Figure 5. Overall, we found that the results from Palantir and Cyclum did not recapitulate the cell cycle dynamics in any of the experimental settings. Palantir is designed to model cell differentiation into terminal states, which does not describe the trajectory of cell cycle dynamics. Cyclum uses an autoencoder with non-linear periodic transformation functions to infer a latent circular trajectory. Although Cyclum is designed for cell cycle dynamics, it does not appear to be able to detect any discernible structure in multiplexed imaging data. The expected input for Cyclum is scRNA-seq data, which has a) three orders of magnitude more features than our datasets, and b) an extremely different type of noise than imaging data. Additionally, because Palantir and Cyclum use non-linear dimensionality reduction methods, they are significantly slower than CMD scaling, which makes them less suitable for the high-throughput analysis of large datasets.

The time ordering output from SCORPIUS closely resembles the ccD-CMD output. However, the ccD-CMD reduced dimensionality representation is strikingly different from the SCORPIUS output. While the data points in the ccD-CMD representation formed a torus, which allowed us to parametrize the representation and derive the "cell cycle coherence" metrics above, the cloud output from SCORPIUS could not be similarly parametrized.

Comparison of specific datasets:

1. Tissue data - Extended Data Fig. 5f,h
   The ordering of tissue data by Cyclum and Palantir does not capture the basic tenets of cell cycle protein dynamics. In addition, the t-SNE representation of the Palantir output shows a branched structure, which is incompatible with the cell cycle. These observations are true for all of the datasets we describe below. SCORPIUS and the ccD-CMD ordering are comparable, but a direct comparison is not possible as the ground truth cell cycle ordering in tissues is unknown.

2. Culture cell line data - Extended Data Fig. 5g
   The ccD-CMD cell ordering appears comparable to the SCORPIUS ordering when run with single cell data from unperturbed cells (MCF10A). In both orderings, there is a clear inverse relationship between CDT1 and Geminin as well increasing DNA content as the cell cycle progresses. However, many of the markers show the drawbacks of the non-cyclical path derived from SCORPIUS. For example, Geminin, phospho-Rb, and Cyclin B show a rise from low to high through time but do not show the drop from high to low that is expected in the cyclical pattern of cell cycle dynamics. ccD-CMD is able to better capture this cyclical pattern.

3. Synthetic data - Extended Data Fig. 5c-e
   To obtain a dataset that has a ground truth of cell ordering within the cell cycle, we simulated cell cycle protein dynamics with a system of ordinary differential equations (ODEs) based on the model by Csikasz-Nagy and Tyson[16] and generated values of nine cell cycle markers over time. We ran both ccD-CMD and SCORPIUS on this data and compared the reconstructed orderings with the known ordering from the differential equation numerical solution. We found that the ccD-CMD analysis outperformed SCORPIUS. In the ccD-CMD ordering, 93% of cells were within 1% of their correct ordering as opposed to 36% of the cells in the SCORPIUS ordering. It is notable that the two-dimensional representation generated from ccD-CMD has

cells tightly distributed along the circle with the exception of a gap where M phase is expected to be (Extended Data Fig. 5d). The ccD-CMD ordering also shows the most disordering in M phase/early G1 (Extended Data Fig. 5d-e) which reflects the difficulty of detecting and ordering cells in M phase that is also seen in the multiplexed imaging data (see section 2.3.5 above on "*Interpretation of the ccD-CMD representation: closed torus vs point cloud*").

2.4.3 Comparison summary

These comparisons show that the ccD-CMD algorithm orders cells efficiently and that the resulting dynamics are congruent with results on cell cycle biology published in the scientific literature. Both Palantir and Cyclum do not seem able to recapitulate the basic tenets of cell cycle protein dynamics, especially in the tissue exemplar dataset. Time inference obtained using SCORPIUS and the ccD-CMD is comparable. In experimental data (either cell line or tissue data), the ground truth information is not available, and hence quantifying the differences in precision between ccD-CMD analysis and SCORPIUS is not possible. However, on synthetic data ccD-CMD has a higher ordering precision compared to SCORPIUS.

Finally, the main advantage of the ccD-CMD representation over the other time inference algorithms is the ability to provide a reduced-dimensional representation of "coherence" in cell cycle dynamics. In control settings such as cells growing freely in cell culture, the ccD-CMD representation forms the torus-shaped that inspired the circle fit approximation (see section 2.3.3). The reduced representation of cell cycle marker correlation generated by other algorithms does not show an 'actionable' and readily interpretable topology (Extended Data Fig. 5f-h), either because of a lack of reduced dimension representation (Cyclum), a lack of topological shape (SCORPIUS), or a disconnected and branched topology (Palantir). In tissues we detected quantifiable differences in the ccD-CMD representation topology across patients that can be interpreted in the context of changes observed in cell culture using growth factor withdrawal (resulting in high CFD) or cell cycle block using palbociclib or nocodazole (resulting in high IOV) (Fig. 3g-l). We present ccD-CMD representations from three patient samples to portray the range of topologies observed in human tissues (Fig. 4b,c). However, the SCORPIUS and Palantir outputs from the same data show no discernible difference between the three datasets (Extended Data Fig. 5h).

REFERENCES

1. Cannoodt, R. *et al.* SCORPIUS improves trajectory inference and identifies novel modules in

    dendritic cell development. *bioRxiv* 079509 (2016) doi:10.1101/079509.

2. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nature*

    Biotechnology **37**, 451–460 (2019).

3. Liang, S., Wang, F., Han, J. & Chen, K. Latent periodic process inference from single-cell RNA-seq

    data. *Nat Commun* **11**, 1441 (2020).

4. Sobecki, M. *et al.* The cell proliferation antigen Ki-67 organises heterochromatin. *Elife* **5**, e13722

    (2016).

5. Labib, K., Kearsey, S. E. & Diffley, J. F. MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. *Mol Biol Cell* **12**, 3658–3667 (2001).

6. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

7. Cayrol, C., Knibiehler, M. & Ducommun, B. p21 binding to PCNA causes G1 and G2 cell cycle arrest in p53-deficient cells. *Oncogene* **16**, 311–320 (1998).

8. Cazzalini, O. *et al.* p21CDKN1A does not interfere with loading of PCNA at DNA replication sites, but inhibits subsequent binding of DNA polymerase delta at the G1/S phase transition. *Cell Cycle* **2**, 596–603 (2003).

9. Georgakilas, A. G., Martin, O. A. & Bonner, W. M. p21: A Two-Faced Genome Guardian. *Trends Mol Med* **23**, 310–319 (2017).

10.   Sicinski, P., Zacharek, S. & Kim, C. Duality of p27Kip1 function in tumorigenesis. *Genes Dev* **21**, 1703–1706 (2007).

11.   Kafri, R. *et al.* Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* **494**, 480–483 (2013).

12.   Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L. & Liberali, P. Trajectories of cell-cycle progression from fixed cell populations. *Nat Methods* **12**, 951–954 (2015).

13.   Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).

14.   Nguyen, L. H. & Holmes, S. Ten quick tips for effective dimensionality reduction. *PLoS Comput Biol* **15**, e1006907 (2019).

15.   Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).

16. Csikász-Nagy, A., Battogtokh, D., Chen, K. C., Novák, B. & Tyson, J. J. Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys J* **90**, 4361–4379 (2006).