**Phase register of Hes1 oscillations with mitoses underlies cell-cycle heterogeneity in ER+ breast cancer cells.**

Nitin Sabherwal[1!]*, Andrew Rowntree[1!], Jochen Kursawe[2] and Nancy Papalopulu[1]*

[1] Division of Developmental Biology and Medicine, School of Medical Sciences, Faculty of Biology, Medicine and Health (FBMH), The University of Manchester, Oxford Road, Manchester, M13 9PT, UK

[2] School of Mathematics and Statistics, University of St Andrews, St Andrews, KY16 9SS, UK

[!] Joint first authors

[*] Joint corresponding authors

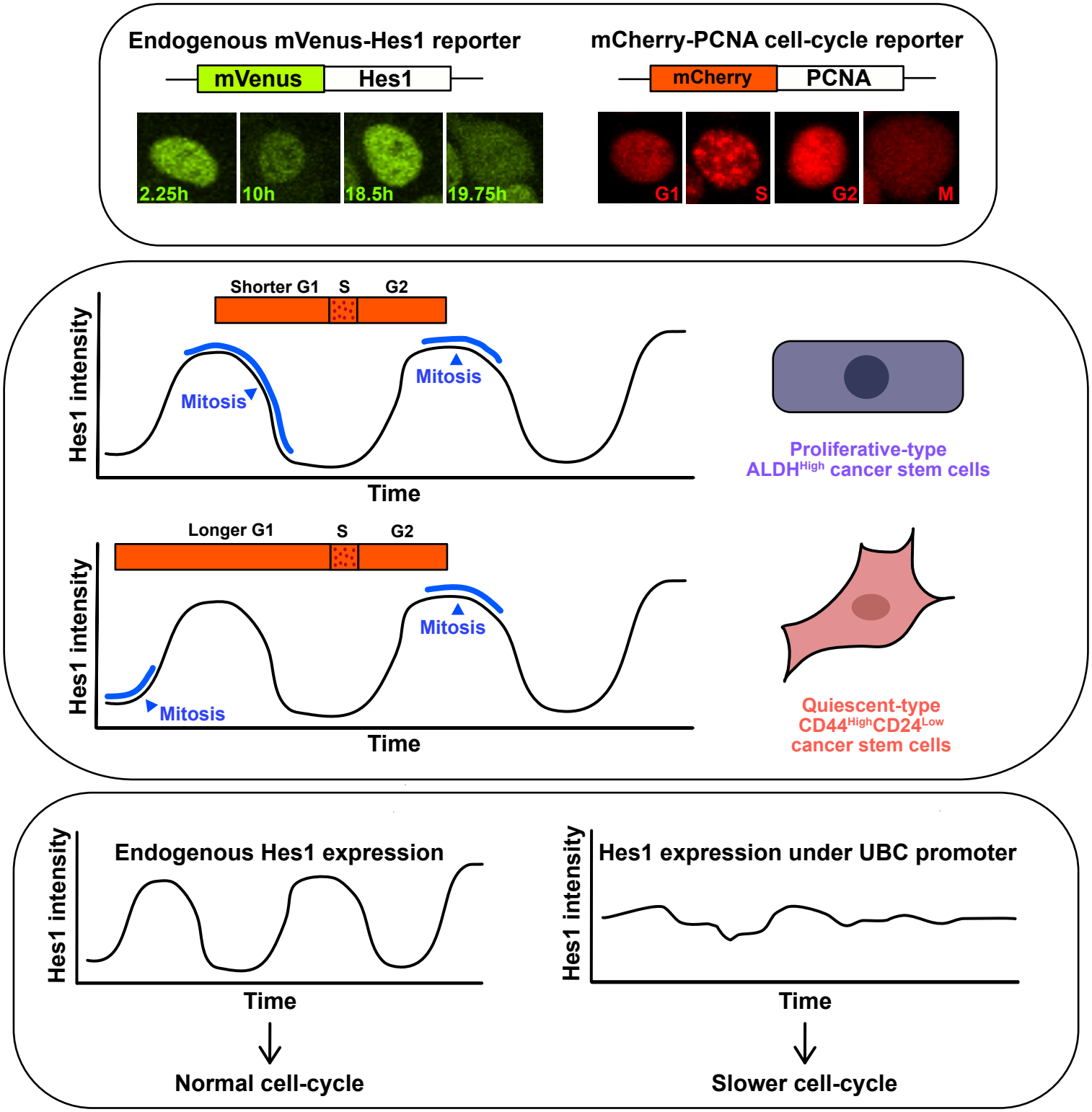
Nitin.Sabherwal@manchester.ac.uk

Nancy.Papalopulu@manchester.ac.uk

**Short title**: Interaction of Hes1 oscillations with cell-cycle

**Key words**: Hes1, oscillations, non-genetic heterogeneity, ER+ breast cancer, cell-cycle, cellular sub-states, quiescence, phase register

# Abstract in picture




**Phase-register of non-ultradian Hes1 oscillations with mitoses underlies cell-cycle heterogeneity in ER+ breast cancer cells.**

**Abstract**

Here, we study the dynamical expression of endogenously labelled Hes1, a transcriptional repressor implicated in controlling cell proliferation, to understand how cell-cycle length heterogeneity is generated in ER+ breast cancer cells. We find that Hes1 shows oscillatory expression and during the cell-cycle has a variable peak in G1, a trough around G1-S transition and a less variable second peak in G2/M. Compared to other subpopulations, the cell-cycle in $CD44^{High}CD24^{Low}$ cancer stem cells is longest and most variable. Most cells divide around the peak of the Hes1 expression wave but mitoses in slow dividing $CD44^{High}CD24^{Low}$ cells appear phase-shifted, resulting in a late-onset Hes1 peak in G1. The position, duration and shape of this peak, rather than the Hes1 expression levels, are good predictors of cell-cycle length. Diminishing Hes1 oscillations impairs proliferation, indicating their functional importance for efficient cell-cycle progression. We propose that the position of mitosis in relation to the Hes1 wave underlies cell-cycle length heterogeneity in cancer cell subpopulations.

**Significance statement**

Tumours exhibit heterogeneities that are not due to mutations, including Cancer Stem Cells with different potencies. We show that the cancer stem cell state predisposed to dormancy *in vivo* has a highly variable and long cell-cycle. Using single-cell live-imaging for the transcriptional repressor Hes1 (a key molecule in cancer), we show a new type of oscillatory expression of Hes1 in all cells in the population. The most potent cancer stem cells tend to divide around the trough of the Hes1 oscillatory wave, a feature predictive of a long cell-cycle. A novel concept proposed here is that the position that a cell is with respect to the Hes1 wave when it divides is predictive of its prospective cell-cycle length and characteristic of its cellular sub-state.

## Introduction

Molecular and phenotypic analyses at the single-cell level have revealed a great deal of heterogeneity in genetically identical populations of cells, which were presumed to be homogeneous based on the population averaging methods. Such non-genetic heterogeneity is thought to have several potential sources, ranging from transcriptional noise (intrinsic) to variability of exposure to environmental signals (extrinsic). Non-genetic heterogeneity has both benefits and pitfalls for the optimal function of biological systems (1). In cancer, it represents a significant challenge because non-genetic heterogeneity is a crucial factor underlying the differential response to chemotherapy, the emergence of treatment resistance in the absence of new mutations, as well as the disease relapse driven by the reactivation of dormant cancer stem cells (2) (3).

At the phenotypic level, non-genetic heterogeneity can manifest itself in the form of the cell-cycle length heterogeneity, which is particularly important in cancer as it may underlie the transition between rapidly dividing, slowly dividing and quiescent cancer cells, with implications for cancer progression, relapse and development of resistance to the treatment (4). Thus, there is a need to understand how cell-cycle heterogeneity may be generated within a population of cells at a mechanistic level.

Most molecular studies of heterogeneity have relied on snapshot measurements of populations of cells where it can be difficult to distinguish the contribution of dynamical gene expression in the time domain versus different levels of expression which could be relatively stable over time. The contribution of dynamical gene expression in creating heterogeneity has been brought to the forefront by studies showing that important regulatory molecules show oscillatory expression (5) (6). Oscillations in gene expression that are asynchronous between cells are not apparent when viewed as a snapshot and on the whole, are an underappreciated source of non-genetic heterogeneity in a population of cells. The ever-growing list of such molecules includes genes associated with the cell-cycle and the circadian clock but also genes such as the DNA damage response protein p53, the pro-neural protein Ascl1 and the neurogenic ligand Dl, to mention just a few (6). Notable examples are the key transcription factors (TF) of the Hes/Her (mammals/zebrafish) family of genes, which have been shown to oscillate synchronously in somitogenesis and asynchronously in neural progenitor cells and whose oscillations enable the cell-state transition to differentiation (7) (8) (9) (10) (11). Among Hes family genes, Hes1 is of particular importance because it has been

implicated in controlling proliferation in a cancer context and shown to control (directly or indirectly) the expression of genes involved in cell-cycle. Indeed, Hes1 is known to affect cell-cycle by repressing both activators (CyD, CyE and E2F) and inhibitors (p21 and p27) of cell-cycle, directly or indirectly (12) (13) (14) (15) (16). However, how these conflicting activities of Hes1 are incorporated in the cell-cycle machinery to modulate its kinetics are not well understood. It is also not clear whether Hes1 oscillates in a cancer model and if so, how does the oscillatory expression of Hes1 interface and possibly control the cell-cycle.

Here, we use an endogenously tagged reporter to ask whether Hes1 oscillates in ER+ breast cancer cells and if so, whether it shows a reproducible correlation with the cell-cycle. We report that Hes1 oscillates with a 26hr periodicity and this nearly matches the average cell-cycle length of ~24hr in MCF-7 cells in most cells. Through clustering analysis of single-cell dynamic traces of Hes1 over the cell-cycle, we find a reproducible relationship between the Hes1 dynamics and the cell-cycle, such that in most cells, division takes place at or near the peak of Hes1 expression. This peak in Hes1 protein expression is then followed by a dip, the onset of which is followed by the G1-S transition in most cells, leading to a second period of increase in Hes1 protein concentration before the next division. A minority of cells divide at the trough rather than the peak of Hes1 expression, causing a delayed Hes1 peak in G1. Such cells tend to have longer cell-cycles, and remarkably, they are enriched for $CD44^{High}CD24^{Low}$ cancer stem cells, which also show the highest cell-cycle heterogeneity. Finally, when we experimentally dampened the Hes1 oscillations, the cell-cycle slowed down, indicating the functional significance of Hes1 oscillations for an efficient cell-cycle progression.

We conclude that Hes1 dynamics are important for sculpting the cell-cycle, such that it occurs efficiently and with little heterogeneity, around a set average cell-cycle length. We propose that differential alignment of Hes1 dynamics with mitoses (i.e. the phase registration) underlies cell-cycle heterogeneity in breast cancer cells and may underlie the propensity of some stem-like cancer cells to become quiescent.

## Materials and methods

### Protein half-life estimation

For estimating protein half-life of exogenous HA-Hes1, MCF7 cells transfected with pCS2-HA-Hes1 plasmid were treated with 100uM Cycloheximide for 30, 60, 90, 120, 150, 180, 210, 240, 270 and 300 mins. Cell lysates from treated cells were used for western blot analysis following standard procedures, using HA-HRP (1:1000, Roche) and tubulin antibodies (as the loading control for the lysates, 1:5000, Sigma).

To estimate protein half-life of endogenous mVenus-Hes1 in Cl19 line, cells treated with Cycloheximide/CHX were live-imaged every 15 mins to capture mVenus-Hes1 mean intensities. Images were processed and analysed using IMARIS software (see below). Decreasing mVenus-Hes1 intensity values over time due to CHX treatment were used to estimate the downward slope and the corresponding mVenus-Hes1 half-life using Microsoft Excel.

### Incucyte growth curve analyses

To compare the growth properties of parental MCF7 *vs* the Cl19 lines, cells were seeded on the 12 well dishes at similar cell densities in triplicates. Cells were imaged every 45 mins using the 4X lens, for around three days period. Following the instructions in the Incucyte software, collected images were processed to lay a confluence mask over the growing cells, which represented cell densities over time. Percentage confluence values over time were used to estimate the growth rates of both parental and Cl19 CRISPR cell lines.

### Fluorescence-activated cell sorting (FACS) to obtain subpopulations of cells from a heterogeneous mixture

To obtain $CD44^{High}CD24^{Low}$ breast cancer stem cells (bCSCs) and bulk (non-stem) subpopulations, cells were trypsinised, resuspended in DMEM/10% FBS and counted using a Neubauer chamber. Ten million cells per FACS sort were centrifuged (500g, 3min), resuspended in blocking solution (1X PBS with 2% FBS, 1000µL) and incubated on ice (30 min). Post blocking, cells were centrifuged again (500g, 3min). Blocking solution was replaced with 100µL blocking solution containing appropriate antibodies (allophycocyanin-conjugated CD44/CD44-APC, Beckton Dickinson and phycoerythrin-conjugated CD24/CD24-PE, Beckman Coulter) at 1:40 and 1:80 dilutions, respectively. Cells were incubated on ice (30-60min). Post-

incubation, cells were washed twice with PBS before resuspending in 1X Hank's Balanced Salt Solution (HBSS) at a concentration of five million cells/ml containing 10µg/ml DAPI. Stained cells were FACS sorted to obtain purified $CD44^{High}CD24^{Low}$ bCSCs and bulk cells from the MCF7 lines using a BD FACSAria Fusion cell sorter (BD Biosciences) as described before (18). To obtain $ALDH^{High}$ bCSCs, cells were stained using AldeRed ALDH Detection Assay kit (Sigma) or ALDHFLOUR kit (STEMCELL Technologies), strictly following the instructions from the manufacturer. Stained cells were FACS sorted using FACSaria cell sorter into $ALDH^{High}$ bCSCs and $ALDH^{Low}$ bulk cells (19).

**Data collection using confocal microscopy and cell tracking using IMARIS**

MCF7 lines expressing fluorescent reporters were live-imaged every 15 mins (unless stated otherwise) for 60-120h using a Nikon A1 confocal microscope. The resulting time-lapse, Z-stacked images were spot-tracked manually using image analysis software IMARIS to obtain 'mean' intensity profiles for Hes1 expression, which represented changes in Hes1 'concentration' during that time, and also to identify precise locations of cell mitoses and various phases of cell-cycle.

The Hes1 mean intensity levels (in arbitrary units) obtained from IMARIS were plotted against time for single cells to obtain single-cell dynamic Hes1 traces over time (time tracks, Fig1D and Supp. Fig5A) for the entire duration of the movie. The entire Hes1 intensity data could be plotted in the form of cell lineage trees (Supp. Fig5B). These lineage trees not only showed the changes in the Hes1 intensity over the entire duration of the movie but also exhibited these changes across cell generations, consisting of mother and daughter cells. We extracted the mean Hes1 intensity changes over time between two consecutive mitoses, representing a full cell-cycle and a full cell generation (generation G1, Supp. Fig5B). These data formed the basis for our analyses of our experiments.

Please **refer to supplementary methods** for the followings -

**Cell culture, plasmids and lentiviral lines**

**Fluorescent tagging of human endogenous Hes1 using CRISPR-Cas9 technology**

**Data presentation, analyses, custom generated codes and statistical methods**

## Results

**1. CRISPR-Cas9 mediated N-terminal tagging of endogenous Hes1 with mVenus in ER+ breast cancer cells (MCF7).** To study Hes1 expression dynamics in ER+ breast cancer MCF7 cells at single-cell level and in real-time, we generated an in-frame fusion of mVenus with the endogenous Hes1. A cDNA cassette expressing mVenus was N-terminally fused to the first exon of Hes1 using a CRISPR-Cas9 genome editing approach (homology-directed repair/HDR), followed by FACS-based clonal selection and expansion of mVenus+ single cells (see materials and methods, m&m, Fig1A). After genotyping, a clonal line (hereafter named as Clone 19/Cl19) was chosen for any further analysis (Supp. Fig1 and Supp. movie1). Cl19 cells were found to be hemizygous for Hes1, with one Hes1 allele having correct insertion of the mVenus cassette and another one knocked-out due to insertion of frame-shift mutation after CRISPR reaction. Cl19 line being hemizygous for Hes1 eliminated any possible interference to the tagged allele by the untagged allele. Cl19 cells recapitulated the endogenous nuclear Hes1 expression, as assessed by immunostaining of parental MCF7 cells using Hes1 antibody (Fig1B and C). During Incucyte based image analysis, both parental MCF7 and Cl19 cells exhibited similar growth rates (Supp. Fig2), suggesting that knocked-out Hes1 allele in Cl19 cells did not adversely affect the cells.

We then compared protein half-lives of mVenus-Hes1 in Cl19 cells by live-imaging against an HA-tagged Hes1 transfected in parental MCF7 cells by western blot analysis, after Cycloheximide/CHX treatment. Both the endogenous mVenus-Hes1 and the exogenous HA-Hes1 exhibited half-lives of around 4hr, which is longer than the mouse protein (24-26min) (27), consistent with the overall increased protein stability in human cells (28). These results suggest that mVenus fusion to endogenous Hes1 does not affect its half-life (Supp. Fig3).

'Unsorted' cells from the Hes1 CRISPR reporter line (Cl19) showed highly heterogenous Hes1 expression among single cells, as shown by snapshots of cultures grown under standard conditions (Fig1B). Immunostaining of parental MCF7 cells using Hes1 antibody (Fig1C) exhibited similar snapshot heterogeneity. Such snapshot heterogeneity can originate (alone or in combination) from two possible scenarios that are, the presence of cells with various Hes1 expression levels, which are stable over time, and/or a temporally dynamic Hes1 expression where cells change Hes1 levels over time, either by random fluctuations or due to oscillatory protein expression around a stable mean. Single-cell live-imaging over time suggested that the expression of Hes1 is temporally dynamic (Fig1D). Thus, population

heterogeneity of Hes1 protein expression in MCF-7 is due, at least in part, to temporal Hes1 dynamics, which may be oscillatory, as we assess in section 4 below.

**2. MCF7 cells exhibit high cell-cycle length heterogeneity, which is higher for the stem cells.** MCF7 cells in culture are mainly representative of non-stem cells, constituting the majority of the population (~99%), referred to as 'bulk' cells. A tiny proportion of cells are classed as breast cancer stem cells (bCSCs) in terms of their functional roles in disease progression; $CD44^{High}CD24^{Low}$ bCSCs tend to be quiescent *in vivo*, while $ALDH^{High}$ type bCSCs are proliferative (29). Therefore, to test whether stem cells might display interesting and useful deviations from the majority of the cells present in culture, we 'FACS-sorted' Cl19 cells into three subpopulations (see m&m) for imaging and analyses, unless stated otherwise. A similar number of cells from each subpopulation ($ALDH^{High}$, $CD44^{High}CD24^{Low}$ and bulk cells) were compared against one another, or, if required, pooled together to generate a virtual 'mixed' population. Both unsorted and FACS-sorted cellular subpopulations from MCF7 cells are highly amenable to single-cell live-imaging experiments. We live-imaged subpopulations of cells from Cl19 line for durations encompassing at least one complete cell-cycle (mitosis to mitosis). To avoid any effects of synchronising agents on either the cell-cycle length or the Hes1 expression, we imaged non-synchronised cells at different stages of their cell-cycle. However, later, for data analyses and comparisons, we pseudo-synchronised Hes1 time-tracks, using the mitoses points as our landmarks.

When characterising the Hes1 levels and dynamics in relation to the cell-cycle, it is crucial to consider the subpopulation heterogeneity that exists in MCF-7 cells. Different cellular sub-types from different breast cancers and their model systems, including MCF7 exhibit cell-cycle differences when grown both *in vivo* and *in vitro*. We estimated the cell-cycle length from single-cell live-imaging data collected for subpopulations of Cl19 MCF7 cells. The data showed a mean cell-cycle length of 29.3h (SD = 11.8h, median = 25.8h, IQR = 9.6h) with high heterogeneity in the cell-cycle lengths with a heavy-tailed distribution towards longer cell-cycles (Fig2A). 'Bulk' cells which represent the majority of cells in the unsorted, heterogeneous cell population had a mean cell-cycle length of 23.4+/-6.6h (mean with SD). By contrast, both stem cell populations had significantly longer cell-cycle lengths, which were on average 29.3+/-10.5h for $ALDH^{High}$ cells and 33.1+/-14.3h for $CD44^{High}CD24^{Low}$ stem cells (Fig2B). In addition, both stem cell populations exhibited higher levels of cell-cycle

heterogeneity than the bulk cells, and this was the highest for the $CD44^{High}CD24^{Low}$ population, evidenced by the higher 'spread' of cell-cycle length values ($CD44^{High}CD24^{Low}$ interquartile rage/IQR = 15.2h, $ALDH^{High}$ IQR = 8.25h and Bulk IQR= 5.3h) (Fig2B). The numbers agreed with the pre-existing notion that while $CD44^{High}CD24^{Low}$ cells represent the quiescent-like stem cells (29), dividing the slowest in normal culture conditions, $ALDH^{High}$ type stem-cells are more proliferative. The non-stem 'bulk' cells representing the majority of the heterogeneous MCF7 cells proliferate the fastest. Here, we extended the previous findings by showing for the first time that the cell-cycle length heterogeneity in $CD44^{High}CD24^{Low}$ cells is higher than in bulk and $ALDH^{High}$ cells, in addition to a prolonged cell-cycle duration on average (Fig2B).

**3. The cell-cycle length shows very weak correlation with the mean Hes1 level.** Given the significant heterogeneities in cell-cycle lengths among MCF7 cells, we also asked if there is a correlation between aspects of Hes1 expression with the length of the cell-cycle, that would imply a functional role. We first examined the correlation of Hes1 protein level with the cell-cycle length by plotting the mean Hes1 levels against cell-cycle lengths obtained from full cell-cycle data. A Spearman correlation coefficient (R) value of only 0.33 suggested only a very weak correlation between the two (Fig 2C); the corresponding $R^2$ value of 0.11 would mean that only 11% of the variance in the cell-cycle length can be attributed to its correlation with the Hes1 mean levels. Furthermore, we found no correlations between the cell-cycle lengths and the initial Hes1 mean intensity (immediately after mitosis) (R = 0.02) (Fig2D). We also noticed that the mean Hes1 expression in the beginning of the cell-cycle is pretty identical to the mean Hes1 expression at the end of the previous cell-cycle, meaning there is no sudden loss or gain of Hes1 protein over mitosis (data not shown). This also suggested that the mean Hes1 levels towards the end of previous cell-cycle would not affect the next cell-cycle. Overall, these data show that while a very weak correlation between mean Hes1 levels and cell-cycle duration may exist, the high cell-cycle heterogeneity observed in MCF7 cells (Fig 2A) cannot be driven by differences in Hes1 levels alone. We hence proceeded by analysing to what extent dynamics of Hes1 expression, rather than its overall expression level, can have an influence on the cell-cycle.

**4. Endogenous mVenus-Hes1 protein expression in MCF7 cells is oscillatory with a circadian-like periodicity.** When we plotted single-cell live-imaging tracking data of Hes1 mean intensity against time, we noticed large scale fluctuations, as shown in Fig1D. These observations prompted us to investigate if Hes1 shows oscillatory expression in MCF7 subpopulations as has been reported in many developmental systems (8) (9) (30) (31) (32) (33) (34). We ran mVenus-Hes1 time traces through pyBOAT/Wavelet, a periodicity analyses platform (Fig3) (22) (23). The power threshold values for the analyses of actual Hes1 traces (Fig. 3A, C) were set based on control background traces, which were analysed similarly but showed no periodicity (Fig3E, G). Fourier analysis showed that unlike the control traces (Fig. 3F), mVenus-Hes1 expression had a dominant mean periodicity of around 26h (Fig. 3B) and no statistical differences for cells from all three subpopulations were observed (Fig3D). In addition to the pyBOAT platform, we also ran our mVenus-Hes1 single-cell time trace data through Lomb-Scargle periodogram (LSP) (Supp. Fig4A-C) (24). As a biological negative control for our periodicity analyses, we used a viral reporter line – Ubc:NuVenus (expressing mVenus tagged with SV40 nuclear localisation signal under the Ubc promoter) – which is not expected to exhibit oscillatory dynamics. LSP analysis gave similar periodicity values for the Hes1 traces, while the control viral reporter line (Ubc:NuVenus) in unsorted MCF7 cells showed that mVenus on its own does not exhibit a periodic expression (Supp. Fig4A). This confirmed that the dynamics of Hes1 protein expression conferred the periodicity observed with the endogenous mVenus-Hes1 reporter. Periodicity analyses using other platforms, including Hilbert transform (25) and Autocorrelation methods (26) showed similar distributions for Hes1 periodicity values (Supp. Fig4D), clearly suggesting that the periodicity values for Hes1 picked by the pyBOAT and the LSP platforms are correct.

**5. Dynamic Hes1 expression between two consecutive mitoses points is biphasic.** Next, we asked whether the Hes1 dynamic expression has a reproducible relationship with the cell-cycle. To facilitate this and for the comparison of large numbers of dynamic Hes1 traces, we visualised all Hes1 traces simultaneously using heatmaps of expression from mitosis to mitosis (see Supp. m&m). To highlight differences in Hes1 expression dynamics rather than their levels we normalised Hes1 traces using Z-scores [(raw intensity - mean intensity)/standard deviation], which enforced that each visualised trace has the same mean and variance. We organised these normalised heat maps in the increasing order of the cell-cycle lengths using

second mitoses as a common end-point (Fig4A). We also aligned individual, normalised mVenus-Hes1 time tracks using second mitoses as a common end-point (Fig4D). Both these alignments showed a biphasic Hes1 expression among all cells, wherein a first Hes1 protein peak was followed by a dip and then by an ascending Hes1 expression towards the end of the cell-cycle (Fig4A and D). To ensure that this dynamic behaviour was specific to Hes1, we expressed Venus and nuclear Venus under the control of the UBC constitutive promoter as controls (Fig4B and C). We analysed their expression traces in the same way as for the mVenus-Hes1. No deducible or reproducible dynamics were observed with either of these controls.

We developed a MATLAB based automatic dip detection pipeline (Supp. m&m), which detected dips in Hes1 time traces with 92% accuracy (4% false positive and 4% false negative, Fig4E and Supp. Fig6). When mVenus-Hes1 single-cell traces were aligned with the second mitosis as the end-point, the duration of the first episode of Hes1 expression was highly variable, ranging from as short as 6.5h to as long as 72.5h (mean = 18.1h, SD = 11.0h, IQR = 8.1h, Fig4A, D and F). By contrast, the position in the time of the dip in Hes1 relative expression in relation to the second mitosis was highly reproducible, which occurred at around 11h after the dip (SD = 3.8h and IQR = 3.2h) (Fig. 4A, D and F).

**6. The dip in the biphasic Hes1 expression dynamics 'gates' the cells for the G1-S phase transition of the cell-cycle.** Next, we wanted to see how Hes1 dynamics might overlap with the cell-cycle phases. Following on from a previous methodology (35), we developed a lentiviral mCherry-PCNA fusion construct as a single-colour, texture-based reporter of the cell-cycle kinetics. In brief, cells expressing PCNA show smooth nuclear staining in G1 and G2 phases of cell-cycle, while this nuclear staining picks up a punctate pattern during S phase, marking DNA replication foci. The PCNA protein spills over the entire cell during mitoses due to the breakdown of the nuclear envelope (Fig5A and Supp. movie2). We performed imaging and analysis on unsorted Cl19 cells virally transduced with mCherry-PCNA so that we could correlate Hes1 dynamics with the cell-cycle kinetics (Fig5A and B). Detailed analyses on these cells co-expressing both Hes1 and PCNA fluorescent reporters showed more than 90% (n = 45) of the cells analysed had an overlap between the Hes1 expression dip and G1 to S phase transition as assessed by the onset of the appearance of PCNA puncti in mCherry expression. Furthermore, our 'Congo flag' analysis (Fig5C) showed that in the majority of the cells (93.3%),

G1 to S transition either happened within the Hes1 dip (50%) or after the onset of the Hes1 dip (43.3%); only a minority of the cells (6.7%) showed G1 to S transition before the appearance of Hes1 dip. The S to G2 transition happened when Hes1 expression started increasing again after the dip in the majority (84%) of the cells (example shown in Fig5B). These data suggested that the dip in Hes1 expression dynamics mostly precedes the G1 to S transition and might permit, or 'gate', cells for this transition during the cell-cycle.

From co-tracking mVenus-Hes1 and mCherry-PCNA in these cells, we conclude that the earlier, variable Hes1 peak marks the G1 phase of cell-cycle, while the time-invariant, second-half Hes1 peak after the dip aligns well with phases between mid-late S and mitosis (Fig5A and B). To interrogate this further, we plotted the duration of Hes1 expression before the expression dip (G1 phase Hes1 expression) against the overall cell-cycle length. A robust correlation was observed (R = 0.95, Supp. Fig6C) indicating that the observed heterogeneity in the variability of the cell-cycle length can be accounted for by the variable Hes1 expression during the G1 phase. This was also evident when the Hes1 expression heat maps were ordered in the order of the cell-cycle length (Fig4A and D).

**7. Gaussian mixture model (GMM) clustering sorts Hes1 dynamics into three distinct classes, with distinct cell-cycle lengths and cellular subpopulation distributions.** To understand how Hes1 dynamics interface with the cell-cycle and whether they fall into distinct classes, we 'stretched' the Hes1 heatmaps to identical lengths using linear interpolation to represent dynamics at pseudo-time scales. This was followed by classifying these pseudo-timed Hes1 heat maps using GMM clustering methods (see Supp. m&m) to ask whether Hes1 heat maps can be classified into discernible, individual groups. Clustering of stretched Hes1 time-series data showed three distinct dynamical Hes1 behaviours during the cell-cycle progression with the following relative expression features (Fig6A and B) - Cluster 1 with high-low-high, cluster 2 with medium-low-high and cluster 3 with low-high-low-high relative Hes1 expression. These differences in the dynamics of Hes1 traces during the cell-cycle were also apparent in unstretched but organised data as per above described clusters, suggesting that they are not an artefact of the pseudo-time analysis (Supp Fig8) and were also evident when the traces were averaged (Fig6B). Further analyses showed that clusters 1 and 2 had similar in cell-cycle lengths (mean length = ~25h) and were both enriched in $ALDH^{High}$ bCSCs (Fig6C and D). Interestingly, cluster 3 was enriched in $CD44^{High}CD24^{Low}$ bCSCs (Fig6C

and D) with highly heterogeneous cell-cycle lengths of up to 90h (mean length 45.2h and IQR 29.5h) (Fig6C). We found that cluster 3 cells share these cell-cycle properties with $CD44^{High}CD24^{Low}$ cells as also shown above in section 2, Fig2B. These data suggest that there are distinct classes of Hes1 dynamics during the cell-cycle and these dynamics are associated with both cellular subtypes and their cell-cycle lengths.

**8. The onset of the first Hes1 peak correlates with the cell-cycle length.** Since durations of the second Hes1 peak (from the dip) and the associated cell-cycle phases afterwards (S-G2-M) are less variable, we next sought to understand which feature of the early Hes1 dynamics correlated best with the cell-cycle length. One could hypothesise that cells tend towards a fixed minimum level of Hes1 to gate G1-S phase transition; therefore, cells with higher peak will take longer to reach this minimum and thus may exhibit a longer cell-cycle. However, we found no evidence for a defined low level of Hes1 that the cells reached as they went through G1 to S transition. Instead, a high starting value of Hes1 correlated with a higher Hes1 value at its lowest level, suggesting towards a constant fold-change. The fold-change in Hes1 expression between the high Hes1 in G1 and low Hes1 at G1-S had a mean value of 1.5-2. When we plotted Hes1 intensity at G1 peak against Hes1 intensity at the G1-S trough, a strong positive correlation (R = 0.76) confirmed the invariability in the Hes1 fold-change (Fig7B); these fold-change values did show no correlation with the cell-cycle length either (R = 0.10, Fig7C). We next looked at the position of the first Hes1 peak by examining the time taken for the cells to reach the first Hes1 peak (Fig7D) and found a good positive correlation between time to peak and the cell-cycle length (R = 0.79) (Fig7E). We also looked at the area under the curve (AUC, Fig7D) for the Hes1 peak in G1 (as a proxy for the shape of the Hes1 peak), and found an excellent positive correlation between the AUC and the cell-cycle length (R = 0.81) (Fig7F), suggesting that the greater amount of time that Hes1 spends above its lower value correlates with longer cell-cycle length. These findings show that the position and the shape of the Hes1 protein peak in G1 are the best predictors of overall cell-cycle length compared to any other dynamic features.

**9. Mitosis is phase-shifted in relation to the Hes1 oscillator in cells with the longer cell-cycle.** The late onset of Hes1 expression in cells with longer cell-cycle prompted to examine the phase relations between the mitoses at the beginning of a cell-cycle and the Hes1 periodic

wave. We constructed phase relation diagrams using Hilbert Transform of the dynamic Hes1 traces using the MATLAB platforms (Fig8A and B, upper panel). Phase diagrams for cells (histograms in Fig8B and circular plots in Fig8C) suggested that for the majority of cells from clusters 1 and 2 with shorter cell-cycles, the mitoses are distributed around the Hes1 peak and the descending slope towards the Hes1 trough; hardly any cells divide at the ascending slope of the Hes1 wave. Since the majority of the cells in the culture divide around the peak ($\pi/2$ and $3\pi/2$, Fig. 8B) we defined this as the dominant phase ("in-phase") of the Hes1 oscillator and mitosis. Interestingly, cluster 3 cells originate from divisions that preferentially take place on a descending Hes1 wave slope (Fig. 8B, last panel and 8C), therefore these cells start their new cell-cycle with a relatively low Hes1, which ascends to an apparently delayed peak of Hes1 expression in G1; we termed this as "out-of-phase" or "phase-shifted" register of the Hes1 oscillator with mitosis (Fig8C). Furthermore, when we presented our data for cellular subpopulations and additionally splitting it into cells with cell-cycle length up to 40h and cells with cell-cycle longer than 40h (Fig8D), it was clear that more than 80% of the cells with longer cell-cycle lengths (>40h) originated from the mitoses which were out of phase with the Hes1 wave and that such cells were preferentially of $CD44^{High}CD24^{Low}$ type (Fig8D). Taken together, these data show that majority of the in MCF7 population exhibit in-phase mitoses around Hes1 peak, resulting in shorter cell-cycles. In contrast, out-of-phase mitoses, i.e. mitoses occurring closer at a Hes1 trough result in cells with longer cell-cycle, which were mainly representing the $CD44^{High}CD24^{Low}$ type cancer stem cells (Fig8C and D).

**10. Sustained Hes1 expression under a ubiquitous promoter slows down the cell-cycle in MCF7 cells.** To ask whether the influence from Hes1 oscillations to cell-cycle length is causative, we directly manipulated the Hes1 expression dynamics. We made use of a viral expression system wherein mVenus-Hes1(cDNA) is expressed under the Ubc promoter (Ubc:mVenus-Hes1) for a sustained rather than a dynamic mVenus-Hes1 expression in the parental MCF7 cells (Fig9). MCF7 cells expressing this reporter showed variability in the levels of the exogenous mVenus-Hes1 expression. To avoid any impact asserted by the over-expression of mVenus-Hes1, we focused on cells with low to medium expression levels of the exogenous mVenus-Hes1. Periodicity analysis of mVenus-Hes1 traces using the LSP platform and the heat maps generated from the normalised expression confirmed that the mVenus-Hes1 expression generated from this viral reporter was not periodic (Fig9A and B). Observed

protein fluctuations were aperiodic, and the characteristic Hes1 dip marking G1 to S transition at a fixed position before mitosis was absent (Fig 9C and D). When these low-medium Ubc: mVenus-Hes1 expressers were compared to the control MCF-7 cells expressing nuclear Venus alone (Ubc:NuVenus), the experimental cells showed significantly longer cell-cycle lengths (Fig9E) and overall, fewer cell divisions (Fig. 9F). These data show a causative influence of the Hes1 dynamics to the cell-cycle length, and suggest that dampening Hes1 dynamics by sustained expression slows down the cell-cycle and impairs efficient cell division.

**Discussion**

The oscillatory protein expression of Hes1, a transcriptional repressor of the helix-loop-helix family, has been shown to be important for cell-state transitions in neural progenitor cells (8) (9) (30). Even though such progenitor cells divide, it was an open question how Hes1 oscillations relate to the process of cell division. In this paper, we have asked whether Hes1 oscillations are observed in breast cancer cells and whether they have a reproducible and functional relation to their cell-cycle.

Using an endogenous CRISPR/Cas9 fluorescent fusion reporter, we show that Hes1 oscillates with an average periodicity of 26h in MCF7 breast cancer cells. This is significantly longer than the previously reported 3-4h ultradian periodicity in mouse cells. There are a number of possibilities that can explain this difference, including species-specific differences, as this is the first report of Hes1 protein dynamics in human cells. We did notice that ultradian oscillations are also "embeded" in these longer duration oscillations (as shown in example in Supp. Fig6A) but our continuous live-imaging of 50-70h duration, which far exceeded any previous live-imaging for Hes1 protein using mouse-based (or any other) experimental systems, uncovered this longer dominant periodicity. During our longer imaging time windows, we noticed bigger fold-change values for Hes1 (Hes1 levels at peak/Hes1 levels at trough) with respect to the bigger oscillations, in comparison to the ultradian ones (analyses not shown), which prompted us to focus on these novel, non-ultradian, circadian-like oscillations of Hes1.

MCF-7 cells are a good model for a non-genetic heterogeneity in breast cancer because they contain "bulk'' cells, i.e. cells with limited potential for tumour (re-)initiation, as well as populations of cells that can act as cancer stem cells when transplanted *in vivo*, namely $ALDH^{High}$ epithelial-like stem cells and $CD44^{High}CD24^{Low}$ mesenchymal-like stem cells (17 and references within). Here, we report that the majority of the MCF7 cells (represented by bulk cells) have a consistent cell-cycle length of approx. 24h while cells sorted for the known cancer stem cell markers show greater variability in their cell-cycle length. Interestingly, stem cells expressing $CD44^{High}CD24^{Low}$ markers, which are more likely to become dormant *in vivo*, show more significant variability than the average of the population and significantly longer cell-cycle, which may predispose them to quiescence, in the right environment.

To understand how these long cell-cycles may be generated, we asked whether the cell-cycle interfaces with the Hes1 oscillations in a manner, which may be causal. Indeed, we report here that, in most cells, Hes1 oscillations have a reproducible relationship to the cell-cycle such that mitosis tends to take place at, or around, the peak of Hes1 expression ('In-phase' register). Following mitosis, a trough of Hes1 expression precedes or coincides with the G1 to S transition, suggesting that the lowering of Hes1 level may gate this transition, perhaps by de-repressing the expression of some key molecules. After this trough, Hes1 protein concentration picks up again as the cells progress to the second mitosis. Thus, because of the tendency of mitosis to take place around the peak, Hes1 expression appears biphasic within every cell-cycle, with two periods of increased protein concentration in G1 and G2. This biphasic expression may reconcile conflicting activities of Hes1, which is known to repress both activators (such as CyD, CyE and E2F) (14) (15) (16) and inhibitors of the cell-cycle (such as p21 and p27) (12) (13), by providing a temporal separation of repressing phases with a period of de-repression at the trough. We suggest that de-repression of molecules such as E2-F1 or CyclinE (a target of Hes1 in MCF-7 cells and known for its function in G1-S transition, https://www.encodeproject.org/experiments/ENCSR109ODF/) due to Hes1 expression dip may be a key for the timing of the G1-S transition. The trough in Hes1 expression at around S-phase is most likely transcriptional as it is not observed when Hes1 is expressed under a ubiquitous promoter and may be due to the loss of Notch signalling in S-phase, reported in developmental systems (36).

The observed variability in the cell-cycle length was specifically associated with Hes1 expression during G1, while the time between the Hes1 trough in G1/S and the subsequent mitosis was less variable. This fits well with reports in the literature wherein neural stem cells, the cell-cycle length is primarily determined by the length of G1 (37). Given this, we further analysed the expression of Hes1 in G1 phase of cell-cycle, to see which Hes1 expression feature shows the best correlation with the cell-cycle length, implying a functional decoding role. One could hypothesise that cells that start with high Hes1 level would take longer to reach a set low Hes1 level, which could be required for G1-S. However, we found that neither the starting levels nor the mean Hes1 levels during the cell-cycle showed a good correlation with cell-cycle length. Furthermore, the trough of Hes1 level had a little varied fold-change relation to the starting level, suggesting that there is no fixed absolute level of low Hes1 that the dynamics tend to. This finding implies that any downstream targets of Hes1 sense a fold-

difference rather than an absolute minimum level of Hes1 at or around G1-S transition during the cell-cycle progression. Thus, Hes1 oscillations may be an underappreciated example of a Fold Change Detection (FCD) system, a possible feature of gene regulatory networks with feedback (38).

Instead, we found an excellent correlation between the shape of the first Hes1 expression period, as recorded in live-imaging traces, and the cell-cycle length. In cells with a delayed onset Hes1 peak during the G1 phase of cell-cycle, that is, a Hes1 peak positioned with some temporal separation from the previous mitosis, tended to have a longer cell-cycle going forward. Thus, the onset and shape of the Hes1 dynamics was a better predictor of cell-cycle length than its level of expression. Since Hes1 shows oscillatory expression, we hypothesised that the delayed Hes1 expression peak is an indication of an 'out-of-phase' registration with the cell-cycle. Indeed, phase reconstruction showed that cells with longer cell-cycles tended to have an unusual phase register with the cell-cycle such that their mitoses tended to take place at the descending arm, rather than at the peak of Hes1 wave. It follows that a delayed Hes1 expression peak in G1 has a knock-on effect in delaying the trough, which in turn delays the G1 to S transition of these cells.

Furthermore, intriguingly, we found that $CD44^{High}CD24^{Low}$ stem-like cells had noticeable dynamic differences from the $ALDH^{High}$ stem-like cells or the bulk cells of the population. Specifically, they were more likely to show a late-onset Hes1 peak in G1 and a higher propensity to have phase shifted registration of Hes1 dynamics with the cell-cycle. These findings are important because these $CD44^{High}CD24^{Low}$ cells show greater heterogeneity in their cell-cycle length (see above) and a higher propensity to become dormant *in vivo* (29). Our findings contrast with previous studies that have focused solely on the level of expression of Hes1, showing that high Hes1 levels are associated with cellular dormancy (31) (39) (40). While this may be true under conditions of high level of overexpression, our findings suggest that under normal conditions, the dynamic phase registration of Hes1 with the cell-cycle rather than the level of expression, underlies an elongation of the cell-cycle, which may predispose some cells to dormancy. Our findings have some analogies to a previous exciting report in development where the phase shift of signalling encodes cellular information (41).

To investigate further if Hes1 oscillations are necessary for the cell-cycle, we misexpressed mVenus-Hes1 under a ubiquitous promoter (Ubc). In cells expressing this

construct, Hes1 did not oscillate showing only some random fluctuations in its expression. In this population, we found a higher proportion of cells that had longer cell-cycle or that failed to divide at all, suggesting that Hes1 oscillations are necessary for an efficient cell-cycle with a consistent cell-cycle length.

In conclusion, Hes1 expression appears heterogeneous in snapshots of ER+ breast cancer cells but this is due to asynchronous temporal oscillations with a circadian-like periodicity. Most likely through regulation of the cell-cycle molecules, these Hes1 dynamics sculpt the cell-cycle to an efficient nearly 24h periodicity. While most cells divide at a specific phase of Hes1 (i.e. mitosis around the Hes1 peak), some cell divide closer to a trough. Such "out of phase" register between Hes1 oscillations and the mitosis generates a cell with a longer cell-cycle, suggesting that a different (perhaps defective) phase coupling of the two processes may underlie the heterogeneity of the cell-cycle. Notably, this is a feature most often observed in $CD44^{High}CD24^{Low}$ stem-like cells, which are prone to quiescence. This suggests that a misalignment of the Hes1 oscillations with the cell-cycle, may eventually contribute to their tendency to drop-out of the cell-cycle *in vivo*, which may have important implications for the progression and relapse of the breast disease.

**Acknowledgements and funding sources**

We thank Profs Robert Clarke from the Manchester Breast Centre and Keith Brennen from the Cell Matrix Centre, FBMH, UoM for providing the parental MCF7 cells and the Human Hes1 cDNA plasmids, respectively. We are grateful to Dr Gareth Howell and Mr Michael Jackson from the FACS facility, FBMH for all their help and expertise in sorting and analysing the cell lines used in this study. We also thank the Genomics facility, FBMH for their sequencing services and Dr Anthony Adamson, from the Genome Editing Unit, FBMH for his guidance on the CRISPR-Cas9 designing and implementation. The work was sponsored by the Wellcome Trust grant no 106185/Z/14/Z to Prof Nancy Papalopulu.

**References**


1. A. Raj, A. van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).
2. A. Brock, H. Chang, S. Huang. Non-genetic heterogeneity-a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* **10**, 336–42 (2009).
3. P. B. Gupta, I. Pastushenko, A. Skibinski, C. Blanpain, C. Kuperwasser, Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* (2019) https:/doi.org/10.1016/j.stem.2018.11.011.
4. W. Chen, J. Dong, J. Haiech, M. C. Kilhoffer, M. Zeniou. Cancer stem cell quiescence and plasticity as major challenges in cancer therapy. *Stem Cells Int.* **2016** (2016).
5. J. H. Levine, Y. Lin, M. B. Elowitz. Functional Roles of Pulsing. *Science (80-. ).* **342**, 1193–1200 (2013).
6. A. Isomura, R. Kageyama. Ultradian oscillations and pulses: Coordinating cellular responses and cell fate decisions. *Dev.* **141**, 3627–3636 (2014).
7. H. Hirata, *et al.* Instability of Hes7 protein is crucial for the somite segmentation clock. *Nat. Genet.* **36**, 750–754 (2004).
8. M. Goodfellow, N. E. Phillips, C. Manning, T. Galla, N. Papalopulu. MicroRNA input into a neural ultradian oscillator controls emergence and timing of alternative cell states. *Nat. Commun.* (2014) https:/doi.org/10.1038/ncomms4399.
9. N. E. Phillips, *et al.* Stochasticity in the miR-9/Hes1 oscillatory network can account for clonal heterogeneity in the timing of differentiation. *Elife* **5**, 1–33 (2016).
10. X. Soto, *et al.* Dynamic properties of noise and Her6 levels are optimized by miR-9, allowing the decoding of the Her6 oscillator. *EMBO J.*, 1–23 (2020).
11. C. S. Manning, *et al.* Quantitative single-cell live imaging links HES5 dynamics with cell-state and fate in murine neurogenesis. *Nat. Commun.* **10** (2019).
12. P. Castella, S. Sawai, K. Nakao, J. A. Wagner, M. Caudy. HES-1 Repression of Differentiation and Proliferation in PC12 Cells: Role for the Helix 3-Helix 4 Domain in Transcription Repression. *Mol. Cell. Biol.* **20**, 6170–6183 (2000).
13. K. Murata, *et al.* Hes1 Directly Controls Cell Proliferation through the Transcriptional Repression of p27Kip1. *Mol. Cell. Biol.* **25**, 4262–4271 (2005).
14. C. Cenciarelli, *et al.* The interference of Notch1 target Hes1 affects cell growth,

differentiation and invasiveness of glioblastoma stem cells through modulation of multiple oncogenic targets. *Oncotarget* **8**, 17873–17886 (2017).

15. N. Noda, S. Honma, Y. Ohmiya. Hes1 is required for contact inhibition of cell proliferation in 3T3-L1 preadipocytes. *Genes to Cells* **16**, 704–713 (2011).
16. J. Hartman, *et al.* HES-1 inhibits 17β-estradiol and heregulin-β1-mediated upregulation of E2F-1. *Oncogene* **23**, 8826–8833 (2004).
17. J. Bagnall, *et al.* Quantitative dynamic imaging of immune cell signalling using lentiviral gene transfer. *Integr. Biol. (United Kingdom)* **7**, 713–725 (2015).
18. H. Harrison, *et al.* Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer Res.* **70**, 709–718 (2010).
19. A. Sarmiento-Castro, *et al.* Increased Expression of Interleukin-1 Receptor Characterizes Anti-estrogen-Resistant ALDH+ Breast Cancer Stem Cells. *Stem Cell Reports* **15**, 307–316 (2020).
20. D. P. Geoffrey J. McLachlan, *Finite Mixture Models* (2000).
21. Leonard Kaufman, Peter J. Rousseeuw. *Finding groups in data : An introduction to cluster analysis* (Wiley-Interscience, 1990).
22. G. A. Moenke G, Sorgenfrei FA, Schmal C. Optimal time frequency analysis for biological data - pyBOAT. *bioRxiv* **179**, 985–986 (2020).
23. S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing* (1999).
24. J. T. VanderPlas. Understanding the Lomb–Scargle Periodogram. *Astrophys. J. Suppl. Ser.* **236**, 16 (2018).
25. Petre Stoica and Randolph Moses. *SPECTRAL ANALYSIS OF SIGNALS* (1997).
26. J. R. B. Oppenheim, Ronald W. Schafer. "Discrete-Time Signal Processing" in (1999).
27. T. Kobayashi, *et al.* Deubiquitinating enzymes regulate Hes1 stability and neuronal differentiation. *FEBS J.* **282**, 2475–2487 (2015).
28. T. Rayon, *et al.* Species-specific pace of development is associated with differences in protein stability. *Science* **369** (2020).
29. S. Liu, *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78–91 (2014).
30. S. Ochi, Y. Imaizumi, H. Shimojo, H. Miyachi, R. Kageyama. Oscillatory expression of Hes1 regulates cell proliferation and neuronal differentiation in the embryonic brain. *Dev.* **147** (2020).

31. R. Sueda, R. Kageyama. Regulation of active and quiescent somatic stem cells by Notch signaling. *Dev. Growth Differ.* **62**, 59–66 (2020).

32. R. Sueda, I. Imayoshi, Y. Harima, R. Kageyama. High Hes1 expression and resultant Ascl1 suppression regulate quiescent vs. active neural stem cells in the adult mouse brain. *Genes Dev.* **33**, 511–523 (2019).

33. R. Kageyama, H. Shimojo, T. Ohtsuka. Dynamic control of neural stem cells by bHLH factors. *Neurosci. Res.* **138**, 12–18 (2019).

34. T. Kobayashi, R. Kageyama. *Expression dynamics and functions of hes factors in development and diseases*, 1st Ed. (Elsevier Inc., 2014).

35. L. Leung, A. V. Klopper, S. W. Grill, W. A. Harris, C. Norden. Apical migration of nuclei during G2 is a prerequisite for all nuclear motion in zebrafish neuroepithelia. *Dev.* **139**, 2635 (2012).

36. F. A. Carrieri, *et al.* CDK 1 and CDK 2 regulate NICD 1 turnover and the periodicity of the segmentation clock . *EMBO Rep.* **20**, 1–22 (2019).

37. N. Sabherwal, R. Thuret, R. Lea, P. Stanley, N. Papalopulu. aPKC Phosphorylates p27Xic1, Providing a Mechanistic Link between Apicobasal Polarity and Cell-Cycle Control. *Dev. Cell* **31**, 559–571 (2014).

38. O. Shoval, *et al.* Fold-change detection and scalar symmetry of sensory input fields. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15995–16000 (2010).

39. R. J. Sang L, Coller HA. Control of the Reversibility of Cellular Quiescence by the Transcriptional Repressor HES1. *Science (80-. ).*, 1095–1100 (2008).

40. L. Sang, J. M. Roberts, H. A. Coller. Hijacking HES1: how tumors co-opt the anti-differentiation strategies of quiescent cells. *Trends Mol. Med.* **16**, 17–26 (2010).

41. K. F. Sonnen, *et al.* Modulation of Phase Shift between Wnt and Notch Signaling Oscillations Controls Mesoderm Segmentation. *Cell* **172**, 1079-1090.e12 (2018).

**Figure legends**

**Fig1. CRISPR-Cas9 mediated N-terminal tagging of the endogenous human Hes1 with mVenus in MCF7 cells shows snapshot heterogeneity due to temporal dynamics.** (A) Schematic of the Hes1 genomic locus, the donor plasmid, guide RNA and how CRISPR-Cas9 mediated homology-directed repair (HDR) reaction achieves tagging of Hes1, and schematic of the Hes1 genomic locus after tagging of the first exon of Hes1 with mVenus cDNA cassette. (B) Imaging of the sequence-verified clonal CRISPR-Cas9 line (Cl19) shows nuclear localisation of Hes1 and snapshot heterogeneity in the level of endogenous mVenus-Hes1 expression, which were recapitulated by the immuno-staining of parental MCF7 cells using an anti-Hes1 antibody (C); the left panels in C show no primary antibody control. (D) Example of a single cell (highlighted with a box in B) tracked over time. Time-lapse images of this example cell show that Hes1 protein levels fluctuate over time (left panel), resulting in a dynamic Hes1 expression as quantified by IMARIS-based cell-tracking (right panel). This dynamic Hes1 expression will generate temporal heterogeneity in mVenus-Hes1 at single-cell levels.

**Fig2. Cell-cycle length heterogeneity in MCF7 cells shows no relationship with the Hes1 mean levels.** (A) Cell-cycle length of 'mixed' MCF7 cells has a mean value of around 29h but shows high heterogeneity (notice the heavy-tailed distribution of the cell-cycle length). (B) Cell-cycle is longer and more heterogeneous for stem cell subpopulations than bulk cells; $CD44^{High}CD24^{Low}$ cells showed the longest and the most heterogeneous cell cycle. Heterogeneity in these cells was even higher than $ALDH^{High}$ stem cells with Bartlett's test showing that $CD44^{High}CD24^{Low}$ and $ALDH^{High}$ cells have significantly different variances (*, $p = 0.0192$). (C) The correlation between the Hes1 mean intensity and the cell-cycle length was overall very modest ($R = 0.33$) as only 11% ($R^2 = 0.11$) of cell-cycle length variance could be attributed to its correlation with the Hes1 levels. (D) The cell-cycle length also did not correlate with the initial Hes1 mean intensity in single cells (immediately after mitoses, in the earlier G1 phases). Overall, these data suggest that Hes1 mean expression levels are not a dominant influence on the cell-cycle length and heterogeneity.

**Fig3. Wavelet analyses of mVenus-Hes1 single-cell time tracks from cellular sub-population showed evidence for one Hes1 wave with a nearly 24hr periodicity**. (A) Using pyBoat platform, mVenus Hes1 time series are detrended prior to periodicity analysis. The black line

shows an example of a raw mVenus recorded time series. Blue is the corresponding detrended time series. The orange line denotes the trend. (B) Fourier Transform for the same input Hes1 trace shows a dominant period peak at around 25h. (C) The detrended mVenus-Hes1 trace is shown again in the above panel, with the lower panel showing the associated period-time-power heat map collapsed in 2D. The periods with the highest power at each time point are used to estimate the mean period and power across the duration of the input trace. (D) All cellular subpopulations from Cl19 line have a mean period of around 26h with no statistically significant differences among them. Each data point in the graph represents the mean wavelet period from one trace. (E-G) Fourier and wavelet analysis of a background time series, with no dominant period for the entire track as confirmed through the Fourier power spectrum (F). (G) The power values at the strongest wavelet periods are lower than in (C), indicating that the trace has no dominant period over time. Notice more than 10-fold higher Fourier power for Hes1 trace in comparison to the background trace (B *vs* F), and also highly coherent period-power values (bright yellow colour in C, shown as a ridge in red) for Hes1, compared to low, non-coherent period-power values for background trace (very low yellow content in G).

**Fig4. Dynamic Hes1 expression between two consecutive mitosis points is biphasic.** (A) mVenus-Hes1 traces from mixed Cl19 cells plotted as heat maps (above) and organised in increasing order of cell-cycle length (below). Notice the presence of the Hes1 expression dip (highlighted by a black line), preceded by a highly variable and followed by a less invariable Hes1 peak, giving rise to the biphasic Hes1 profile during the cell-cycle. When mVenus (B) or Nu-mVenus (C) were expressed under the Ubc promoter as controls, their expression profiles during the cell-cycles showed no such characteristics. (D) When we plotted the same individual Hes1 tracks (as in A) as line plots, collapsed at the second mitoses at the right-hand side of the graph, we noticed a similar biphasic Hes1 expression dynamics with a variable, first Hes1 peak, followed by a dip in its expression (again marked by a black line), which was followed by a Hes1 peak towards the end of the cell-cycle. (E) A MATLAB-based dip-detection method in the mVenus-Hes1 expression profile during the cell-cycle detected dips in the Hes1 expression with 92% accuracy, as compared to manual annotation (see Supp. Fig6). The duration after the dip (until the second mitosis) showed a tight distribution for all the cells with a mean value of around 11h, while the duration to the dip (from the first mitosis) showed

very high variability (6.5-86h) (F). Together, these data showed that Hes1 expression between two consecutive mitoses is biphasic, with a variable peak of expression in G1 phase, a peak in G2 phase and a dip in its expression in between.

**Fig5. Correlating Hes1 expression dynamics with the cell-cycle kinetics using PNCA-based live reporter shows a trough in the biphasic Hes1 expression dynamics, which 'gates' the cells for the G1-S phase transition of the cell-cycle.** (A) Example images of a single cell over time expressing both mVenus-Hes1 and mCherry-PCNA. Cl19 unsorted cells were virally transduced with mCherry-PCNA reporter to show how differentially labelled cell-cycle phases correlate with the Hes1 protein expression dynamics. Schematic of Hes1 and PCNA reporters are shown above the single-cell images. (B) Example of a time trace of Hes1 mean intensity from a single cell expressing PCNA reporter, which was used to mark G1-S and S-G2 transitions, shown as red and green asterisks on top of the Hes1 expression profile. (C) 'Congo flag' plots correlating time to G1-S transition (time from first mitosis to G1-S transition marked by the onset of appearance of PCNA puncti) and time to Hes1 dip (time from first mitosis to the minima in the Hes1 expression, with the dip lasting around 10h as shown in Fig4D) showed that in more than 90% of the cells, the dip in the Hes1 expression precedes or coincides with the G1-S transition, suggesting that the Hes1 dip may gate the cells for this transition.

**Fig6. Gaussian mixture model (GMM) clustering of the mVenus-Hes1 traces shows 3 clusters with distinct Hes1 expression dynamics; these clusters map to cellular sub-states and cell-cycle lengths.** (A) GMM clustering classifies Hes1 time traces (stretched, Z scored heat maps) into three distinct classes (clusters 1, 2 and 3, upper panel). The panel underneath shows cell-cycle lengths of individual cells represented as heat maps, showing that cluster 3 is enriched for cells with longer cell-cycles. (B) These clusters can be defined in terms of distinct mean and normalised Hes1 expression dynamics with cluster 1, 2 and 3 having high-low-high, med-low-high and low-high-low-high Hes1 dynamics, respectively. This is clearly evidenced by the individual (upper panel) and averaged (lower panel) Hes1 tracks. Dark blue line in the lower panels show the mean Hes1 behaviour per cluster, with lighter blue shaded representing the standard deviation. (C) Cell-cycle length distributions of individual clusters confirm that cluster 3 consisted of cells with the longest and the most heterogeneous cell-cycle, while clusters 1 and 2 are enriched for cells with shorter cell-cycles with very little heterogeneity.

(D) While clusters 3 is enriched for $CD44^{High}CD24^{Low}$ bCSCs, clusters 1 and 2 were enriched for $ALDH^{High}$ bCSCs. Overall, Hes1 dynamic behaviour can be classified into three classes with distinct cell-cycle lengths and sub-population distributions.

**Fig7. The shape and the onset of the first Hes1 peak can predict the length of the cell-cycle.** (A) Example of a Hes1 trace with automatically detected peak of the Hes1 in G1 and the dip during G1-S transition. These Hes1 peak (maxima) and dip (minima) values were used to estimate fold-change in Hes1 expression when a cell transitions from G1 to S phase of the cell-cycle. (B) A strong correlation between Hes1 peak and dip values (R = 0.76) suggested that the Hes1 fold-change values for cells remain within the same range (1.5-2 fold), irrespective of whether a cell started with higher or lower Hes1 intensity. (C) When these estimated fold-change values were plotted against the cell-cycle lengths, no correlation was found (R = 0.10), clearly showing that the cell-cycle lengths in the MCF7 subpopulations were not influenced by the Hes1 fold change values. (D) The same dip detection pipeline was used to estimate how long it took for a cell to reach the peak of Hes1 expression at G1 phase (t), and also the area under the curve (AUC) for Hes1 expression during the G1 phase of cell-cycle, which represented the shape of the first phase of the Hes1 expression. (E and F) Both time to peak and AUC showed strong positive correlations with the cell-cycle lengths (R = 0.78 and 0.81, respectively). These data clearly showed that the shape and the onset of the Hes1 expression in its first peak were the best predictors of the cell-cycle lengths in MCF7 subpopulations.

**Fig8. The 'phase register' of mitoses with the Hes1 oscillator and its influence on the length of the cell-cycle - mitosis is phase-shifted in relation to the Hes1 oscillator in cells with the longer cell-cycle.** (A) Upper panel shows an example of a Hes1 time series (blue) over two generations with smoothened trace (orange) and linear trend (yellow) line superimposed. Lower panel shows smoothened but detrended trace (blue) from upper panel and phase reconstruction via Hilbert transform (orange) with phase readouts on right-hand side y-axis. Phases 0 or $2\pi$ represent trough and phase $\pi$ represents a peak in the Hes1 wave. The green line shows the phase readout at the beginning mitosis. (B) Upper panel is a schematic of a basic Hes1 wave showing one full period (i.e. trough to trough or phase 0 to $2\pi$) and phase readouts at intervals of $\pi/2$ along this wave. This panel was used as reference to interpret

where on the Hes1 wave, mitoses that precede the cell-cycle took place for cells from clusters 1 and 2 (middle panel) and cluster 3 (lower panel). The middle panels show that for cells from clusters 1 and 2, the majority of the mitoses took place around the Hes1 peak (between $\pi/2$ and $3\pi/2$), with a small fraction of cells showing mitoses at the descending Hes1 wave (between $3\pi/2$ and $2\pi$). The lower panel shows that for cells from cluster 3, the majority of the mitoses happened at the downward Hes1 wave (near $2\pi$), around its trough. (C) Scatter plot in polar coordinates where the angle (reading anti-clockwise) represents the phase read out at the mitosis point preceding a cell-cycle. The distance from the centre of the circle represents the cell-cycle length of that cell. Cluster 1 cells are in red, cluster 2 in blue and cluster 3 in green. (D) Scatter plots similar plot to C, except here $ALDH^{High}$ bCSCs are in blue, $CD44^{High}CD24^{Low}$ bCSCs are in orange and Bulk cells are in yellow. The left panel shows cells with cell-cycle lengths up to 40h and the right panel shows cells with cell-cycle lengths over 40h. Histograms in B and Scatter plots in C and D show that preceding mitoses for cells with longer cell-cycles, mostly belonging to cluster three and enriched for $CD44^{High}CD24^{Low}$ bCSCs, register with the Hes1 'out of phase', i.e. the mitoses occur around the trough, while cells with shorter cell-cycles mitose around the peak of the Hes1 wave.

**Fig9. Dampening Hes1 oscillations by sustained expression of Hes1 elongates or impairs the cell-cycle.** (A) Schematic of constructs for endogenous expression of mVenus-Hes1 (orange) and mVenus-Hes1 under the Ubc promoter (red). (B) Normalised LSP power shows a strong peak in the power spectrum obtained from the endogeneous mVenus Hes1 expression, in contrast to the the mVenus-Hes1 expression under the Ubc promoter. This suggested that endogenous mVenus-Hes1 shows coherent oscillations while the same protein does not oscillate over time when expressed under the Ubc promoter. (C) Traces of Ubc-mVenus-Hes1 cells, normalised and arranged as a heatmap from mitosis to mitosis with time running vertically from top to bottom. These are aligned in time at the second mitosis of each cell. The time traces have been arranged in increasing order of the cell-cycle length. The dip in expression found by our automated dip-detection (ADD) pipeline is displayed as a black asterisk in, showing that a dip does not occur at an invariant time from the end of the cell-cycle unlike in the Endogenous mVenus-Hes1 time traces (Fig4A and D). (D) Ubc-mVenus-Hes1 time tracks from cells arranged as normalised time series traces from mitosis to mitosis, aligned in time at the ending mitosis, confirming the absence of a common dip in expression

before division as observed in C, compared to the endogenous mVenus-Hes1 as shown in Fig4A and D. When Ubc:mVenus-Hes1 cells were compared against control cells expressing Nu-Venus alone for their cell-cycle properties, experimental cells were found to exhibit longer cell-cycle (E) and to have lower number of cell-divisions (F). Together these data show that Hes1 oscillations are required for the smooth progression of the cell cycle.

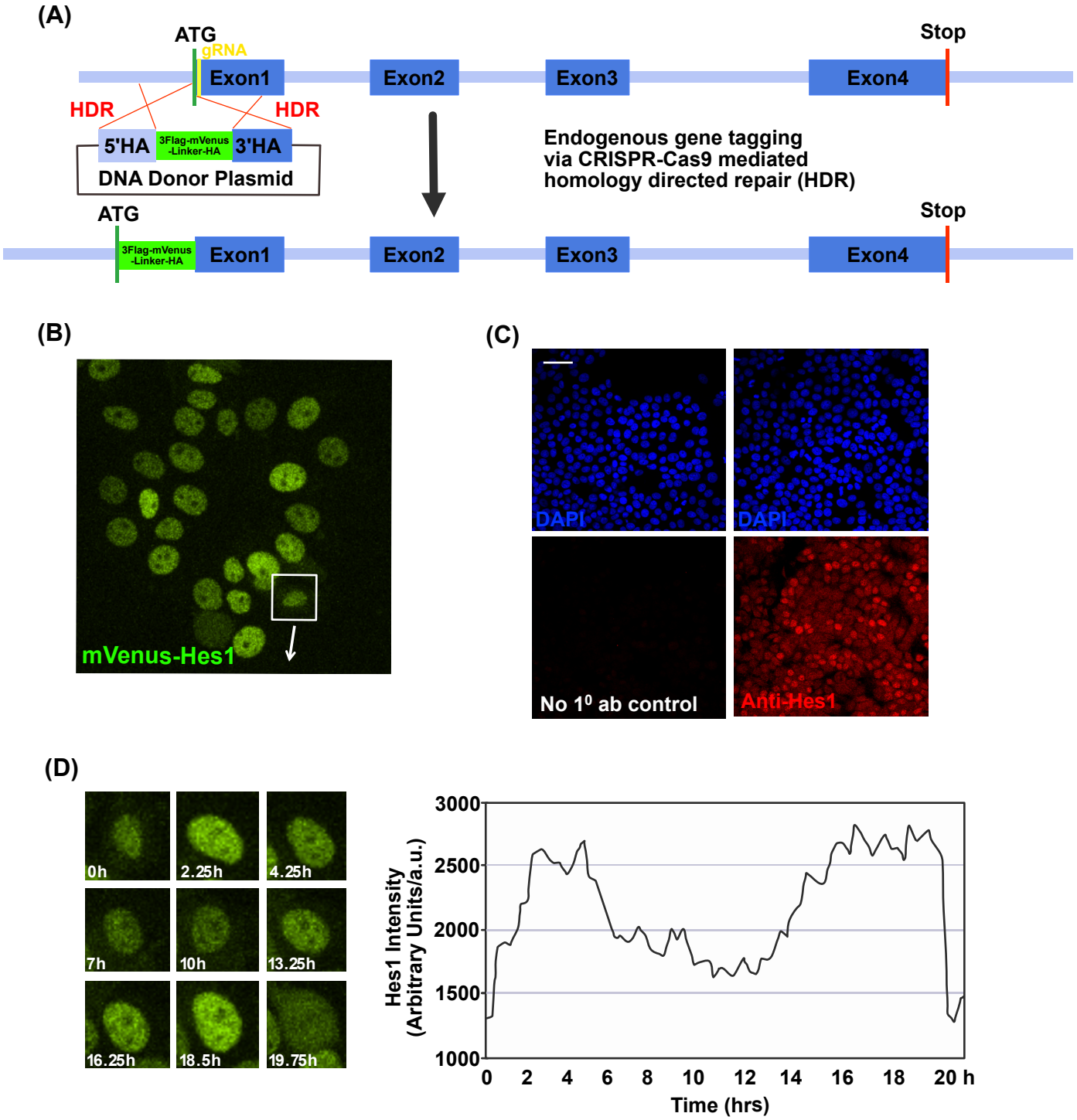



**Figure 1. CRISPR-Cas9 mediated N-terminal tagging of the endogenous human Hes1 with mVenus in MCF7 (ER+ breast cancer) cells shows snapshot heterogeneity due to temporal dynamics.**

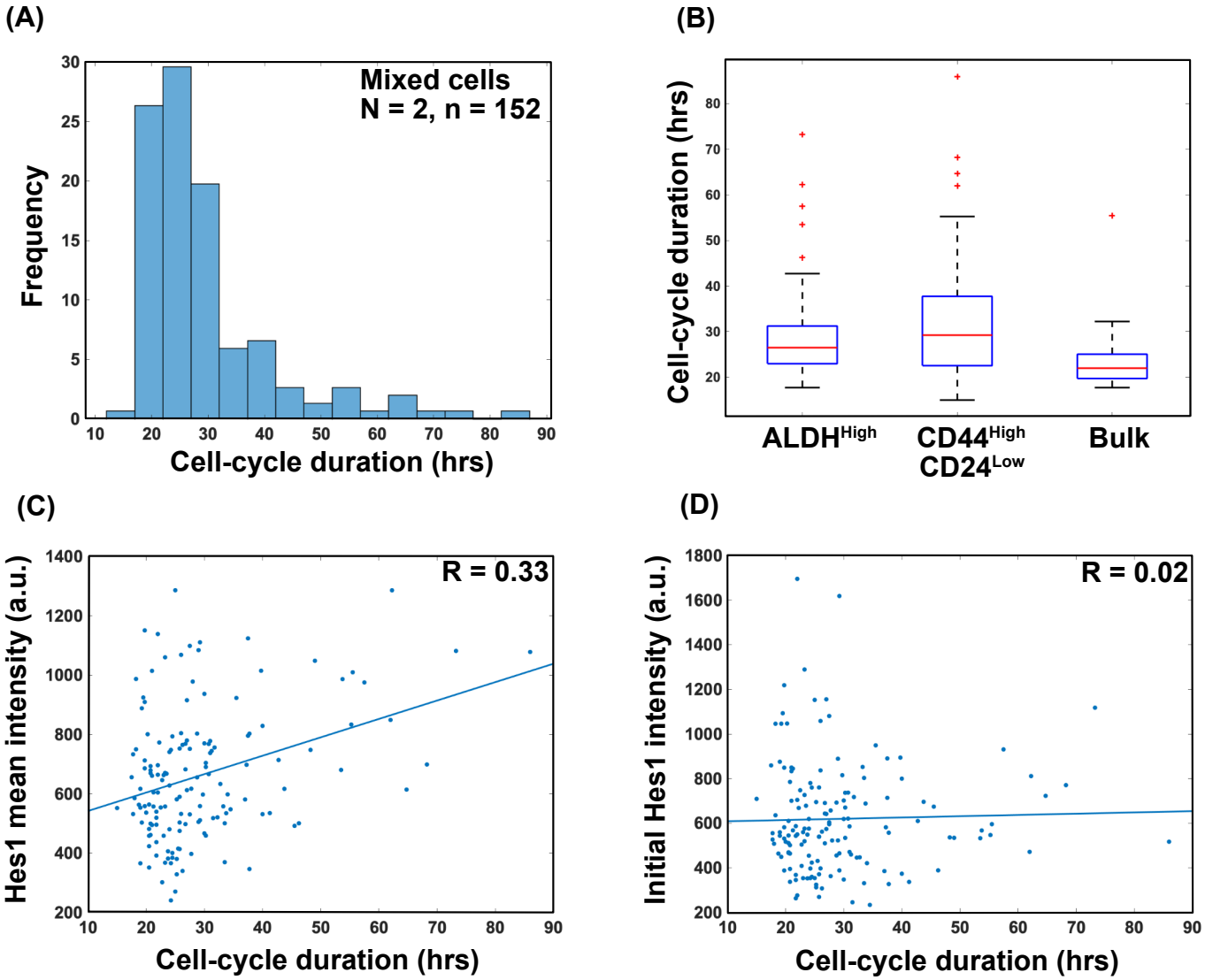



**Figure 2. Cell-cycle length heterogeneity in MCF7 cells shows no relationship with the Hes1 mean levels.**

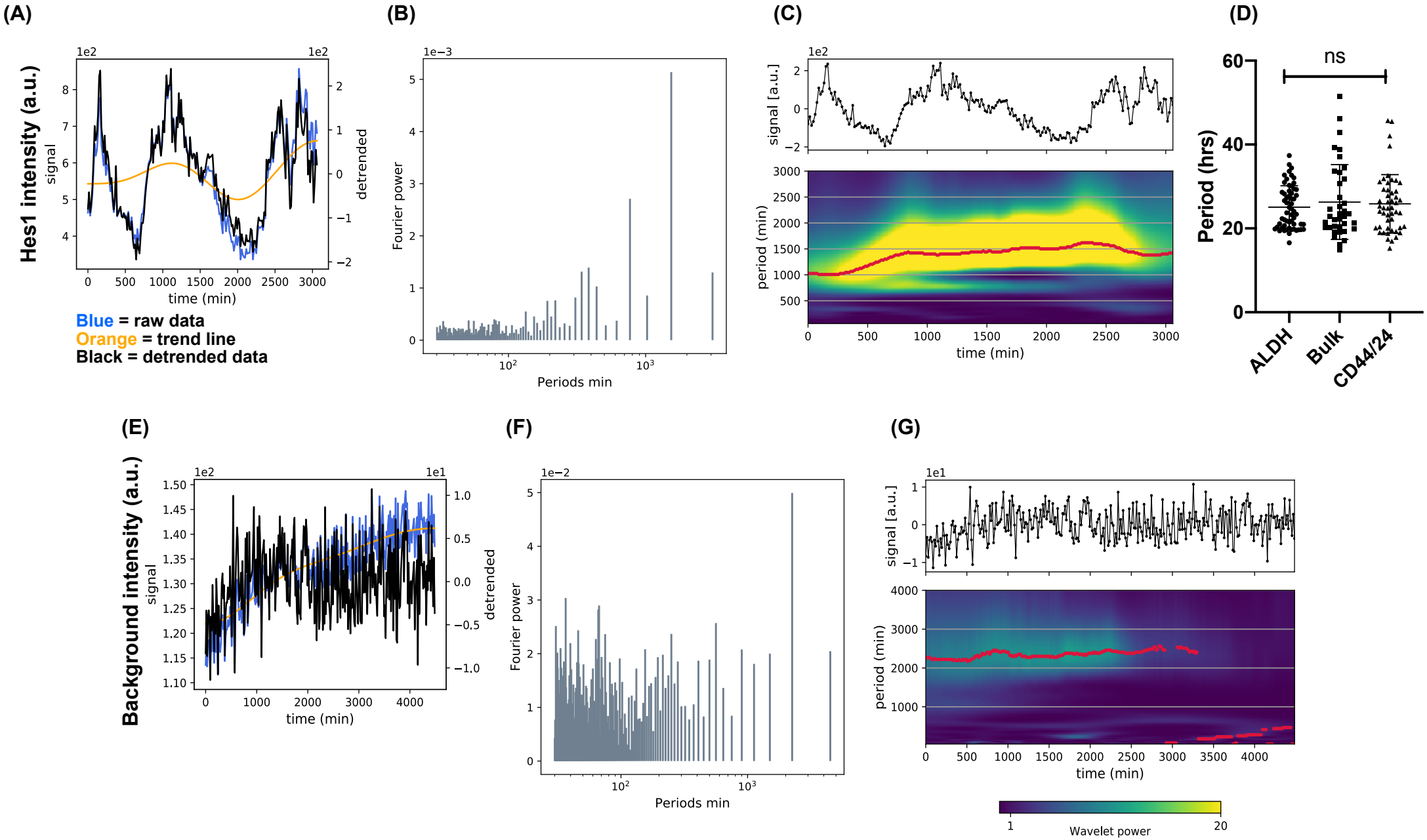



**Figure 3. Wavelet analyses of mVenus-Hes1 single-cell time tracks from cellular sub-population showed evidence for one Hes1 wave with a nearly 24hr periodicity.**

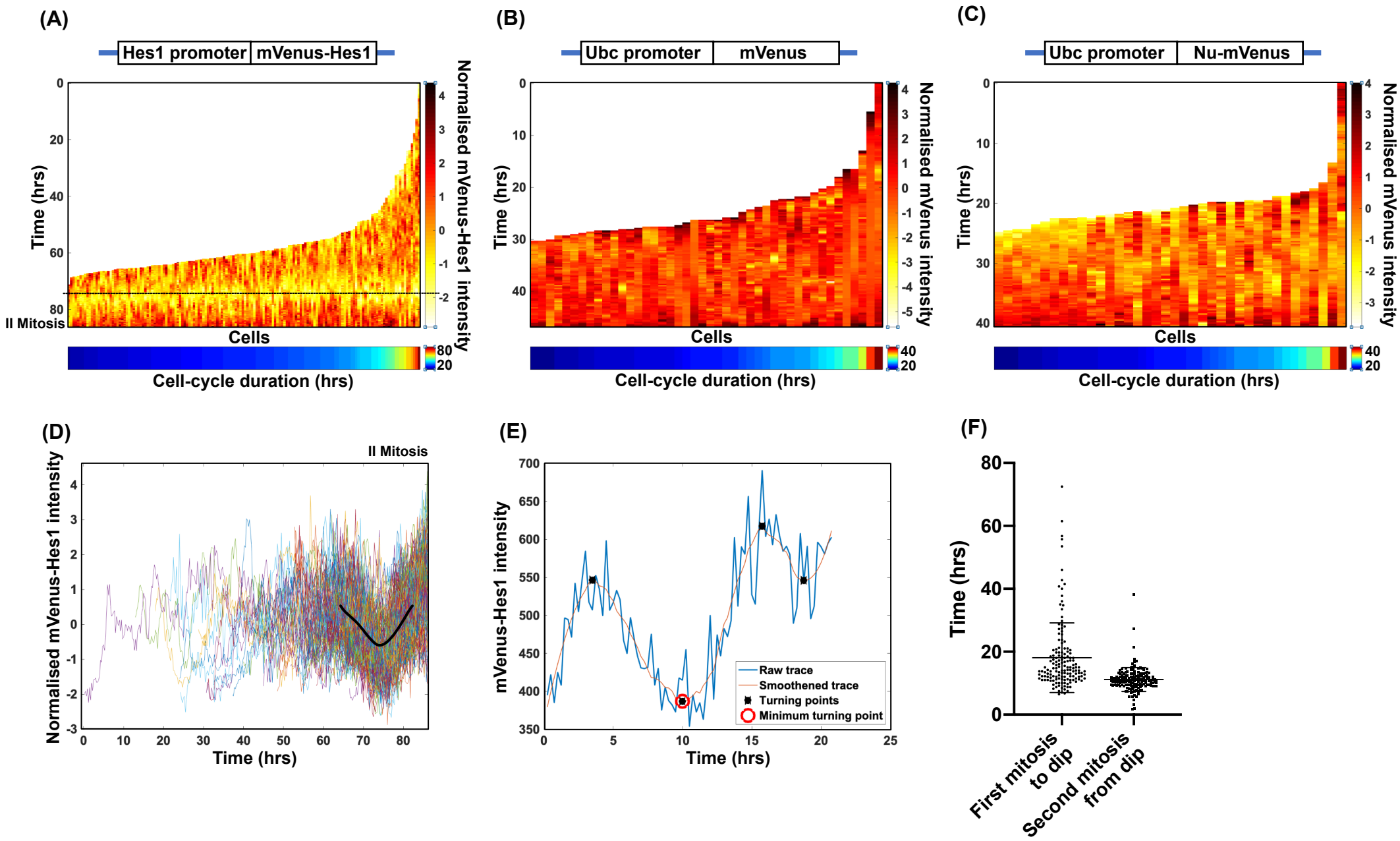



**Figure 4. Dynamic Hes1 expression between two consecutive mitosis points is biphasic.**

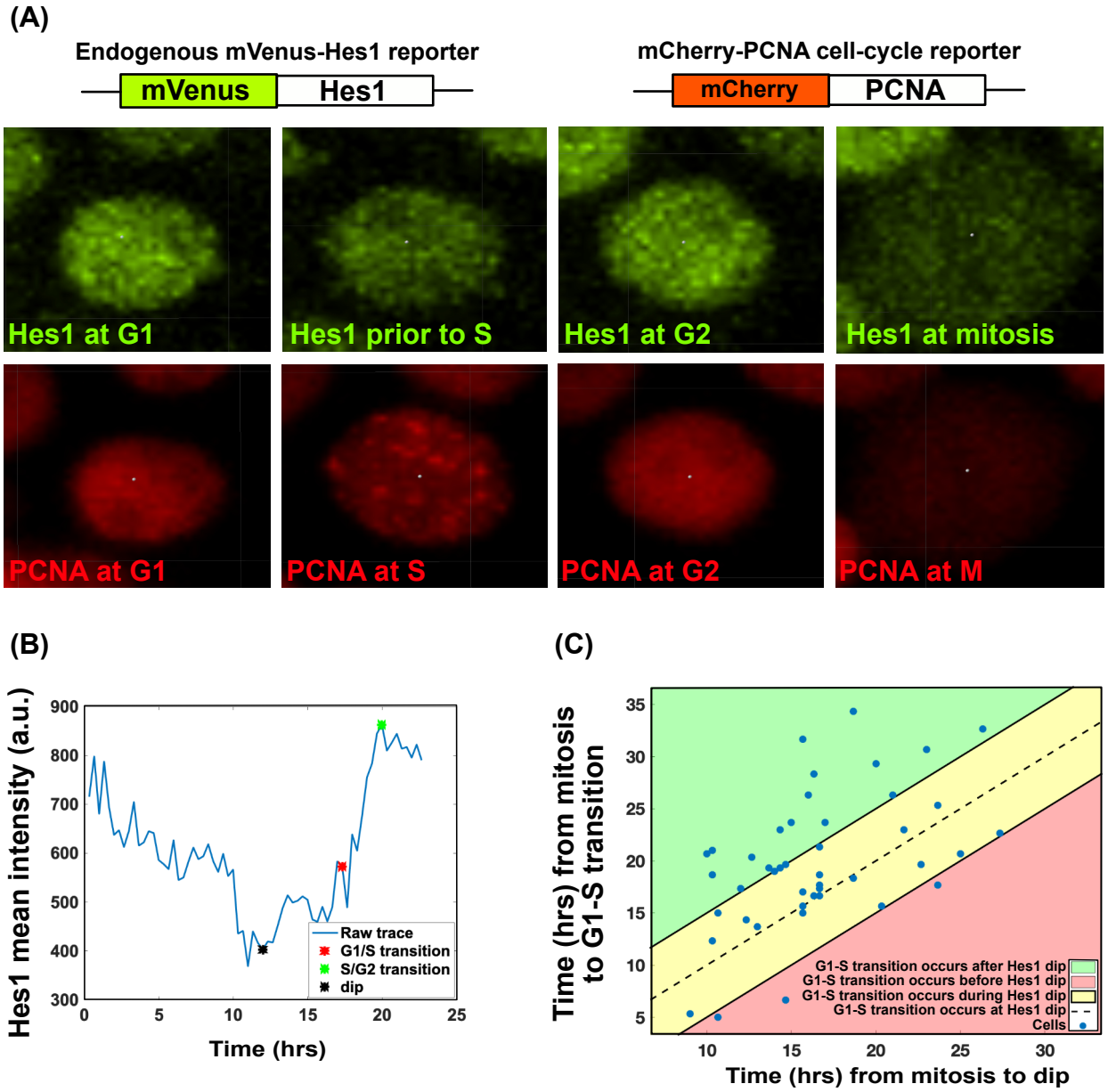



**Figure 5. Live-imaging of endogenous mVenus-Hes1 along with mCherry-PNCA based cell-cycle reporter shows that a trough in the biphasic Hes1 expression dynamics 'gates' the cells for the G1-S phase transition of the cell-cycle.**

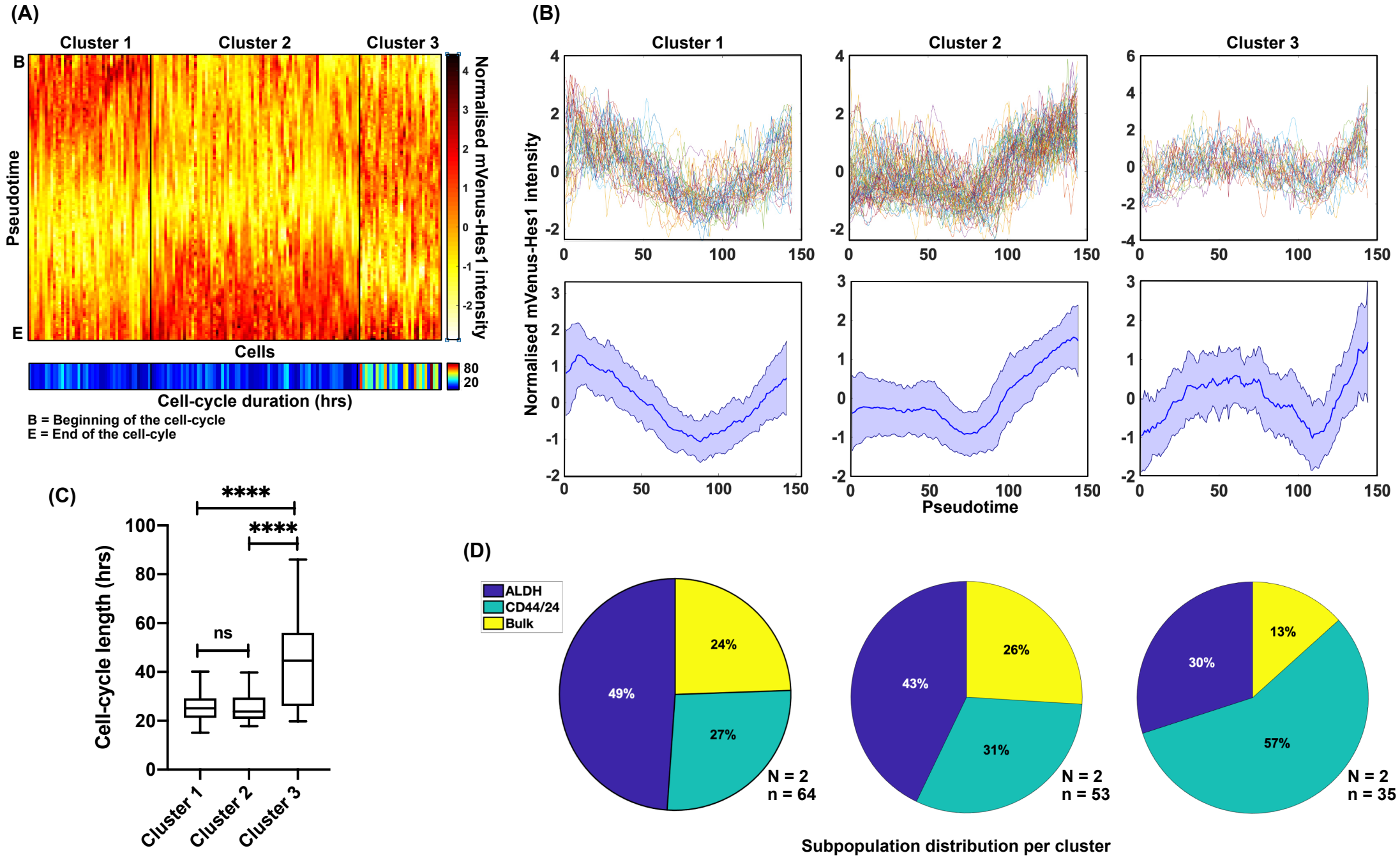



**Figure 6. Gaussian mixture model (GMM) clustering of the mVenus-Hes1 traces shows 3 clusters with distinct Hes1 expression dynamics; these clusters map to different cellular sub-states and cell-cycle lengths.**

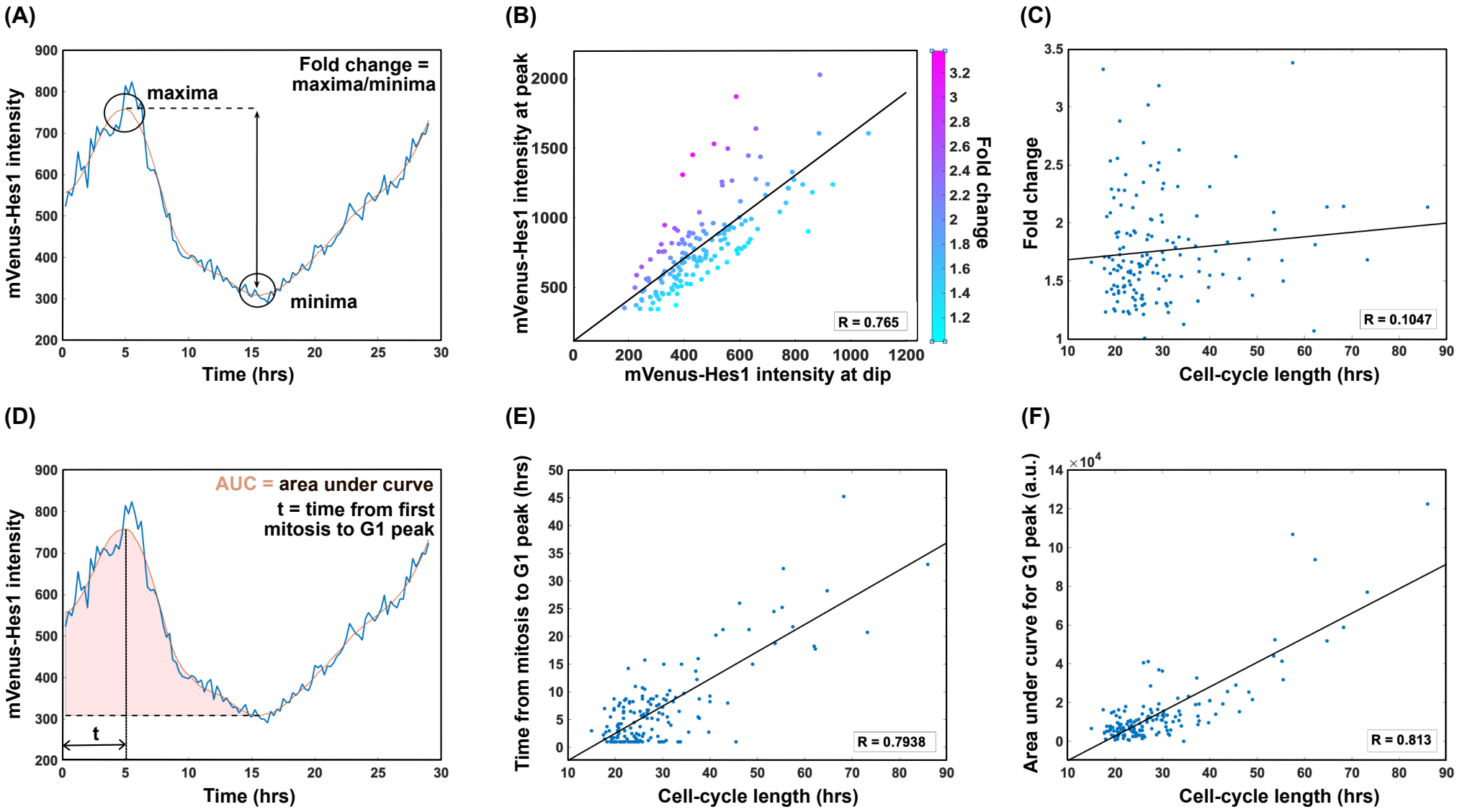



**Figure 7. The shape and the onset of the first Hes1 peak can predict the length of the cell-cycle.**

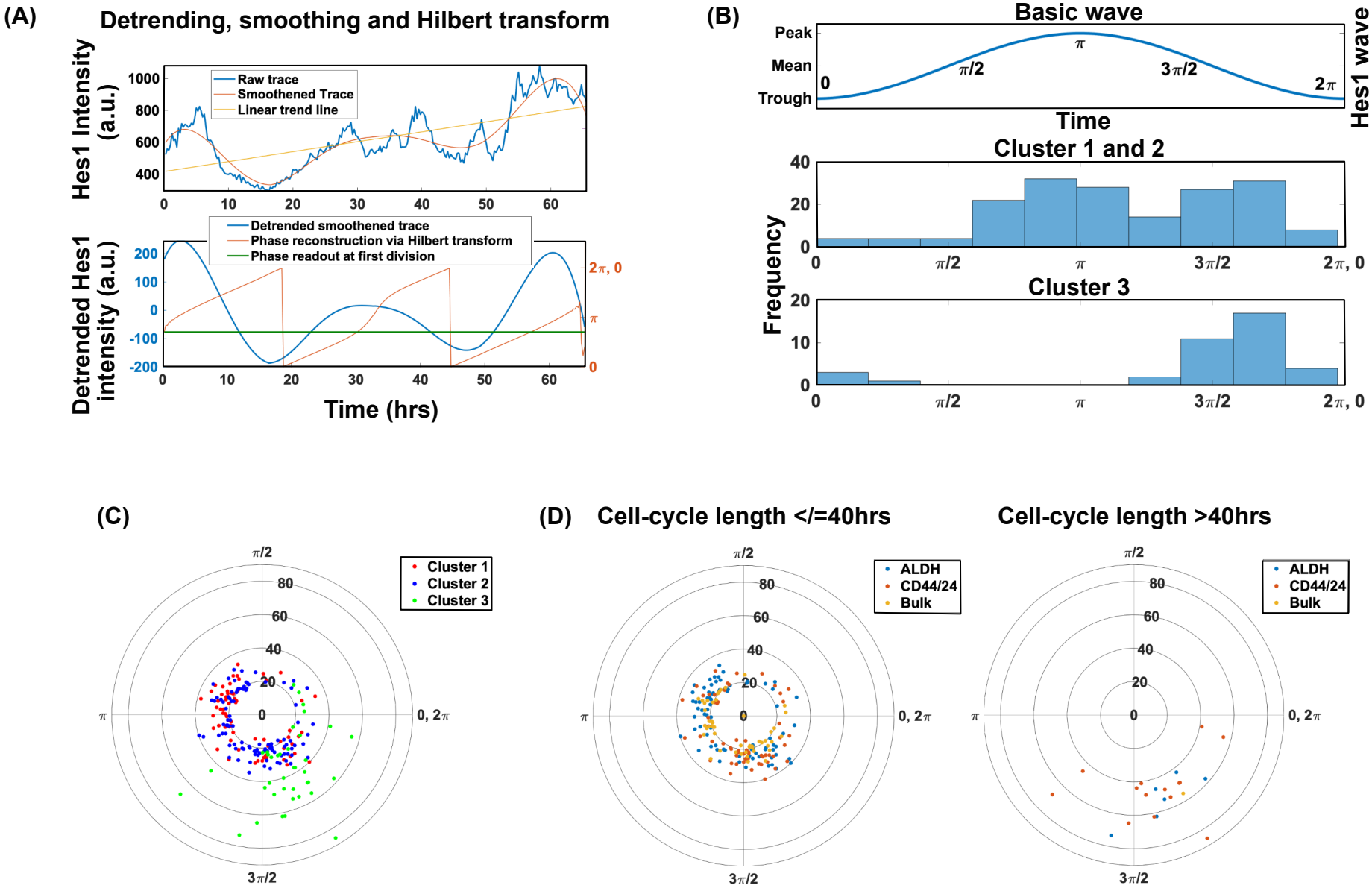



**Figure 8. The 'phase register' of mitoses with the Hes1 oscillator and its influence on the length of the cell-cycle - mitosis is phase-shifted in relation to the Hes1 oscillator in cells with the longer cell-cycle.**

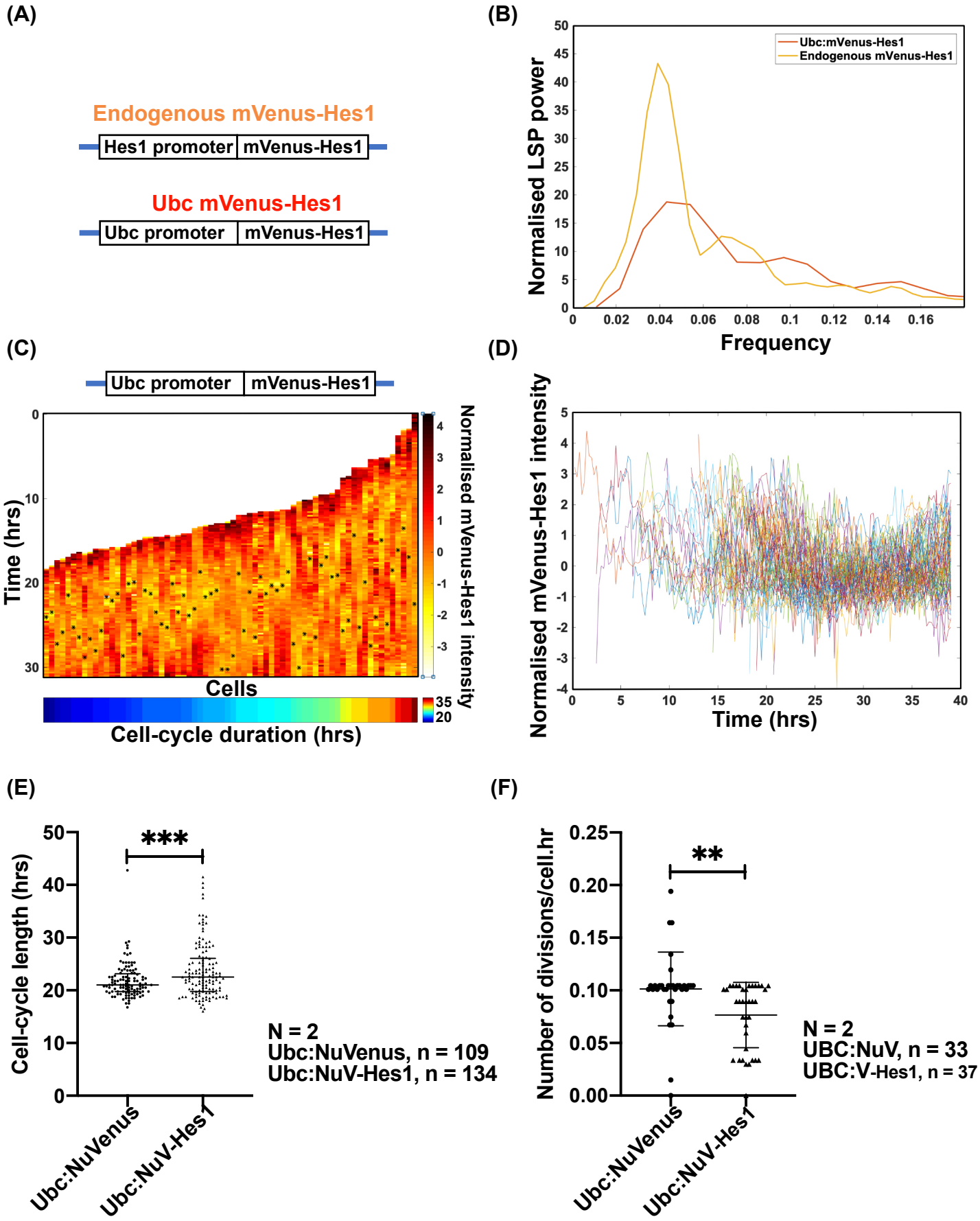



**Figure 9. Dampening Hes1 oscillations by sustained expression of Hes1 elongates or impairs the cell-cycle.**

**Supplementary information**

**A. Supplementary movies**

Supplementary movie 1 shows time-lapse images (15fps) of the Clone 19 reporter line wherein the clonal cells express endogenously tagged Hes1 with mVenus (green). These cells were imaged every 15 mins for a duration covering at least one complete cell-cycle (mitosis to mitosis) of FACS-sorted cellular subpopulations from MCF7 cells, including $CD44^{High}CD24^{Low}$ breast cancer stem cells (bCSCs), $ALDH^{High}$ bCSCs and non-stem bulk cells. This line reported the dynamics of the endogenous mVenus-Hes1.

Clone 19 cells were further transduced with the lentiviral cell-cycle reporter (mCherry-PCNA, red). This new line, shown as supplementary movie 2 (Supp. movie2a showing mVenus-Hes1 signal in green and Supp. movie2b showing mCherry-PCNA signal in Red for the same cells), formed the basis of our analyses connecting dynamic protein expression of mVenus-Hes1 with different phases of the cell-cycle.

## B. Supplementary materials and methods

### Cell culture

Parental MCF7 and any other reporter lines derived from it were propagated in high Glucose DMEM (Sigma-Aldrich, D6429) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich) and 1% Penicillin-Streptomycin (Pen-Strep, Sigma-Aldrich), at $37^0$C under 20% O2 and 5% CO2 culture conditions. Adherent cells were passaged twice weekly using 1X Trypsin-EDTA solution (Sigma-Aldrich) to dissociate them from the culture flask. Cells were routinely checked for any mycoplasma contamination using in-house established PCR based testing method.

### Plasmids and viral reporter lines

To generate pCS2-HA-Hes1 plasmid (expressing human Hes1 fused to HA epitope under CMV promoter), Hes1 was PCR-amplified from the plasmid pBS-SK-HuHes1 using primers with restriction sites introduced. Purified and digested PCR product was sub-cloned in the pCS2-2HA plasmid. Clones were sequence-verified before using them to transfect MCF7 cells using Lipofectamine 3000 using the manufacturer's protocols (Life Technologies).

To generate III generation viral plasmid constructs, we used pLNT-Ubc:mVenus plasmid (Ubc driving mVenus expression) (17) as the backbone for any further subclonings. To generate pLNT-Ubc:NuVenus (nuclear mVenus driven by Ubc promoter) plasmid, mVenus cDNA was amplified using the forward primer containing SV40 NLS and the reverse primer. Digested PCR product was subcloned back in the same vector backbone, and selected clones were sequence verified. Similarly, for the plasmid pLNT-Ubc:mCherry-PCNA (mCherry-PCNA driven by Ubc promoter), primers with introduced restriction sites were designed to amplify mCherry cDNA and human PCNA cDNA separately. Both PCR products were digested and subcloned in the original pLNT vector one after the other, followed by sequence verification of the selected clones. Plasmid for expressing mVenus fused to human cDNA under the Ubc promoter (pLV-Ubc:mVenus-Hes1) was obtained commercially from VectorBuilder. Details for all primers and cloning strategies are available on request. Lentivirus production and transduction of MCF7 cells were performed as described before (17). To maximise the cell population expressing fluorescent reporter of interest, transduced cells were FACS sorted on BD Influx Cell Sorter (BD Biosciences, UK).

**N-terminal in-frame fluorescent tagging of human endogenous Hes1 gene in MCF7 cells using CRISPR-Cas9 genome editing technology**

The donor plasmid for N-terminal tagging of endogenous human Hes1 DNA in MCF7 cells with fluorescent monomeric Venus (mVenus) tag was commercially synthesised from GenScript. It consisted of 800 bps upstream of the Hes1 start codon (5' homology arm/5' HA), 800 bps downstream of the Hes1 start codon (3' HA) and in between these two HAs, a mVenus cassette consisting of ATG-3XFLAG-mVenus(cDNA)-Linker-HA, all cloned in a pUC57 based vector backbone (Fig1A). In the donor plasmid, silent mutations were introduced near the PAM sites where the Sp-Cas9 protein would cut the Hes1 genomic DNA. CRISPR-Cas9 machinery inserts the mVenus cassette downstream of the endogenous Hes1 ATG through homology-directed repair (HDR) reaction using donor plasmid as a template (Fig 1A). gRNAs against the human Hes1 were designed using the Wellcome Trust Sanger Institute portal (https://wge.stemcell.sanger.ac.uk//). The gRNA 5'-gaaaaattcctcgtccccggtgg3-' was commercially obtained from Integrated DNA Technologies (IDT) and used in the CRISPR-Cas9 reaction. tRNA and purified Sp-Cas9 proteins were also obtained from IDT. gRNA (200µM) and tRNA (200µM) stock solutions were mixed in IDT duplex buffer, so their final concentrations were 48µM. The mixture was heated at 95°C for 5 min, followed by slow cooling at room temperature. To make the RNP complex, 5µL of gRNA:tRNA complex was mixed with 2µL of Cas9 (61µM) and 4µL of IDT duplex buffer, making the final concentrations of both gRNA and tRNA as 20.4µM and Cas9 as 11.09µM. The RNP complex was incubated at room temperature for 10 min before it was mixed with 2µg of donor plasmid. The RNP-donor plasmid mix was used to electroporate one million cells using the 4D-Nucleofector X Unit (Lonza) following the standard instructions and the inbuilt program (EN-130) for MCF7 cells. Post electroporation, cells were mixed with growth media and cultured following the standard procedure. Electroporated cells were FACS-sorted for mVenus-positive signal, and single-cell clones were grown in standard culture media. Single-cell clonal lines were genotyped to identify a clone correctly recapitulating the endogenous Hes1 expression. A total of 22 lines were genotyped. A clonal line (Clone 19/Cl19) with a hemizygous mVenus-Hes1 allele, passing the genotyping characterisation was chosen for all further experimentation (Supp. Fig1). Details of the primers used for genotyping are available on request.

**Data analyses methods and custom generated codes**

All data analyses were performed on MATLAB R2018a unless stated otherwise. All the custom codes/algorithms used for data analyses have been deposited in the GitHub (https://github.com/AndyRowntree/Phase-register-of-Hes1-oscillations-with-mitoses-underlies-cell-cycle-heterogeneity).

**a) Generation of temporal Hes1 expression profiles as time tracks and heat maps**

To enable the comparison of large numbers of dynamic time traces in one graph, we decided to display dynamic traces from individual cells as normalised heatmaps. First, due to high heterogeneity in the Hes1 mean intensity levels, intensity values over time from each cell were normalized using z-scores (using the formula $y = (x - \mu)/\sigma$, where y and x represent the normalised and raw intensity levels, respectively, at each time point. $\mu$ and $\sigma$ are the mean and standard deviation, respectively, of the intensity levels across the whole trace) to provide relative traces (Figs 4, 6, 9 and Supp. Figs 5a and 6-8). We then used these normalised Hes1 intensity values to generate heatmaps using the '***imagesc'*** function, where each x-axis point contains one cell's dynamic Hes1 trace from the first mitosis to the second mitosis (representing a complete cell-cycle) with time running vertically from top to bottom.

**b) Interpolation**

In order to classify the heat maps of varying lengths using clustering algorithms, we needed to take each time series and display it as a pseudo-time series of a constant chosen length. We 'stretched' or 'squashed' the data to a pseudo-time scale using the default interpolation option (linear interpolation) using the '***interp1***' function in MATLAB. Due to the requirements of the GMM clustering algorithm (see below), which requires more columns than rows in order to perform clustering, the number of time points in the pseudo-time scale was set to the number of traces minus one. However, when displaying the pseudo-time traces as heatmaps (Fig 6), this number was increased to the length of the longest trace so to encompass more detail into the heatmaps. To ensure that no clusters were produced as artefacts due to this constraint of pseudo-time points, we performed a synthetic control (see below).

**c) Gaussian Mixture Model (GMM) clustering**

Clustering is a standard data analysis technique, classifying data into distinct groups, each containing a specific trait or characteristic. We performed a Gaussian mixture model (GMM) clustering to classify Hes1 heat maps into discernible, individual groups (20). This clustering method fits unlabeled data into a predetermined number (k) of d-dimensional Gaussian components (in our case, d is the number of pseudo-time points). Albeit somewhat similar to k-means clustering, GMM clustering allows for more flexibility as both the mean and variance/covariance of components are considered instead of merely the means. GMMs also allow for distributions, which may be elliptical and not strictly spherical like k-means, allowing for more freedom. We utilised the functions '***fitgmdist***' and '***cluster***' to perform our clustering, it fits Gaussian components to the data using the iterative Expectation-Maximisation (EM) algorithm. Random initial values for component means, covariance matrices, and mixing proportions are chosen, and these parameters are tuned through each iteration until a desired convergence is reached. We use the Silhouette method (21) to determine the optimal number of clusters (k=3). This method provides a value (the silhouette coefficient) which can be used as a measure of how similar each trace is to its assigned cluster and to other clusters. A score (per k) can then be obtained through averaging of these values: the higher the score, the more suitable it is for the data to be classified into k clusters. The method was also used to suggest the most suitable constraints to place upon covariance matrices of the Gaussian components namely 'shared', which means that the shape of distributions for all components in the mixture model are identical and 'diagonal', meaning that Gaussians distributions were only allowed to stretch or squeeze in orthogonal directions during the EM algorithm iterations (Supp. Fig7).

**d) Synthetic Data as a control for GMM clustering**

When interpolating data using a pseudo-timescale, the shape of the Hes1 dynamic expression may be distorted and features in cells with longer cell-cycles might be missed. To ensure that our clustering techniques would perform correctly on our data set and clusters aren't created simply as a result of stretching/squashing, we constructed synthetic data as a control. For this we generated five underlying dynamic shapes: three from polynomial curves fitted to our actual data, and two antiphase sine curves (Supp. Fig7C). These antiphase sine curves would generate two sets of traces that were directly opposed and therefore should never cluster

together. Each shape was generated over short, medium and long length-scales in order to produce 15 base traces (Supp. Fig7D). We then duplicated each base trace ten times and added a vector X of random values, where $X \sim \mathcal{N}\left(0, \frac{1}{2}\right)$ to represent biological noise, which was representative of our actual experimental data, resulting in 150 (n = 152 in our actual experimental data) distinct, synthetic control traces. We then interpolated these traces to provide data of the same length, followed by randomly mixing the (Supp. Fig7E) and applied the same clustering techniques as with our previous analysis. Results showed that all 5 dynamic shapes were clustered together (Supp. Fig7G) using GMM clustering at a cluster number and covariance matrix type as suggested by silhouette analysis (Supp. Fig7F). These results confirmed that the clustering methods used in our experimental analysis were not skewed as a result of data interpolation.

**e) Automated Hes1 expression dip-detection pipeline**

To detect a relative dip in the Hes1 mean expression during the cell-cycle in an automated fashion (Fig 4 and Supp. Fig6), we developed a computational pipeline, into which our data consisting of the Hes1 mean expression time traces were fed. For each Hes1 trace, Savitzgy-Golay (SG) filter was applied using the function '***sgolayfilt***'. Parameters for the SG filter were set as a polynomial order of 3 and a moving frame length of a quarter of the given trace length. The resultant smoothened traces were truncated by three time points at either end to remove artefacts caused by the low polynomial order. All turning points of smoothened, truncated traces were found by finding the points at which the product of the difference between these trace values at three consecutive time points is negative. The dip in expression was then determined as being the point at which the minimum of all such turning points lied. To avoid earlier detection of a dip in any cells starting with low Hes1 levels, dips detected in the first ¼ part of any trace were neglected. The same smoothened traces were utilised when finding the Hes1 peak in G1 phase (Fig 7A): this was accomplished by identifying the maximum value of our traces which lay earlier in the cell-cycle than our previously-given dip.

**f) Periodicity Analyses of Hes1 time traces**

Visual inspection of our mVenus-Hes1 time traces suggested that Hes1 expression was oscillatory with a singly 'full' wave encompassing more than one cell-cycle/generation. Hence,

we found it difficult to reliably extract periodicity data for the Hes1 wave from time traces covering only a single cell-cycle. To confidently extract the periodicity values of Hes1 wave, using multiple analysis pipelines (see below), we concatenated Hes1 time traces for two generations (combining Hes1 mean intensity traces from two generations, mothers and their daughters) (n=145). Our analyses across generations also showed that there was no loss of Hes1 protein during mitoses connecting these two cell-cycles/generations. The mean Hes1 levels at the end of the first cell-cycle and the beginning of the second cell-cycle were nearly identical; the overall Hes1 protein levels during mitosis were also matching to the Hes1 levels in mother and daughter cells (data not shown). These observations confirmed the continuation of a single Hes1 wave across generations, without any loss of Hes1 protein. These periodicity analyses were performed using a variety of methods for corroborative reasons and are described below.

**1 - Continuous-wavelet based time-frequency analysis of Hes1 periodicity using pyBOAT**

To correctly identify time-dependent properties of oscillations within our Hes1 time-series data, we used pyBOAT (22). PyBoat is a computational tool to identify and characterise oscillations in time series data. Its code, detailed installation and user guidelines are available on the GitHub (https://github.com/tensionhead/pyBOAT). It includes built-in functions for Fourier transforms and continuous wavelet transforms. Wavelet transforms can be particularly useful to identify any time-dependent features of non-stationary oscillatory signals. Before Fourier or Wavelet transforms are applied, pyBoat detrends each time series. This detrending relies on a windowed sinc filter (23), which is designed to remove all Fourier components with periods longer than a fixed threshold, $T_c$, from the signal. In all our analysis, we chose $T_c$ equal to the maximum length of the considered time series. Typically, our time series covered a range of 3000-4000 mins, covering two cell-cycles/generation for the MCF7 cell subpopulations imaged. When applying wavelet analysis, we consider periods between 1h and the maximum length of each track.

**2 - Lomb-Scargle periodogram (LSP)**

The Lomb–Scargle periodogram (24) is a well-known algorithm, falling under the Fourier transformation and Least-squares methods for detecting and characterising the periodicity from time-series data. It allows efficient computation of a Fourier-like power spectrum

estimation, resulting in an intuitive means of determining the period of oscillation. We utilised the '***plomb***' function within MATLAB to estimate periodicity via this method. It returns the Lomb-Scargle power spectral density (PSD) estimate of a signal that is sampled at the instants specified. We input linearly detrended traces and received a normalised (for ease of comparison between experiments) periodogram as an output. One particularly high normalised power spike from these spectra imply a strong oscillatory signal. In contrast, readouts consisting of multiple spikes which are relatively lower in power implies the lack of oscillation in our traces.

**3 - Autocorrelation**

We used the '***xcorr***' function to determine a periodicity estimate via autocorrelation. This method finds the cross-correlation (25) (value of a given trace against itself as it is shifted along in time. The autocorrelation sequence of a periodic signal has the same cyclic characteristics as the signal itself. Thus, autocorrelation can help verify the presence of cycles and determine their durations. When more than one period was recorded per trace, we took the mean of the recorded period estimates.

**4 - Peak to peak time detection via Hilbert transform**

A final method we used to estimate periodicity was to apply the Hilbert transform (26) to each smoothened trace using the Savitsky-Golay (SG) filter (order = 4, moving frame window = half of the trace length) which was detrended by a linear fit. Estimates for the period were measured as time between resets of the phase reconstruction obtained by the '***angle***' function of the transformed trace (the idea being that the phase will reset once per period). If more than one period was recorded per trace, the mean of the recorded period estimates was taken.

**g) Phase diagrams**

To extract a phase readout, we used Hilbert transform of Hes1. We use MATLAB functions '***hilbert***' and '***angle***' to output the given phase at all points in each smoothened, linearly detrended trace. Smoothening was performed as above except that we chose an order four polynomial and a moving frame length of half the length of the trace when applying the Savitsky-Golay (SG) filter to account for the longer traces. The linear trend was found and

removed (Fig 8A). Polar scatter plots were plotted using the function '***polarscatter***' (Fig8C and D). They represented the phase of Hes1 wave at the beginning of mitosis, showing phase outputs for each trace at the start of a new cell-cycle.

**Correlation plots**

We used the Spearman correlation coefficient plots to measure the strength and direction of a linear relation between two variables (such as Hes1 mean intensity and cell-cycle length). We used the following textual descriptors associated with R values to describe the relationship between our variables; R = 0 means no linear relationship, +0.30 a weak positive linear relationship, +0.50 a moderate positive relationship. +0.70 a strong positive relationship and 1 a perfect positive relationship. Also note that for a given R value, its square value denotes the proportion of the variance that can be attributed to the correlation; for example, for a correlation between cell-cycle length and mean Hes1 intensity, an R value of 0.33 and its squared ($R^2$) value of 0.11 would mean that only 11% of the variance in the cell-cycle length can be attributed to its correlation with the Hes1 levels, clearly showing why an R value of 0.33 represents a weak positive correlation.

**Statistical analysis**

Data were analysed using multiple hypothesis testing methods, as indicated in the figure legends using Prism 8 (GraphPad, USA). Statistical significance is reported for p-values <0.05 (*), <0.01(**) and <0.001 (***), calculated using unpaired T tests. Errors bars are reported as standard deviation of the mean (mean+/-SD) calculated from pooled data unless otherwise stated. We also used Bartlett's test (through MATLAB) to check if the variances were equal across the groups of data. Significant Bartlett statistic allowed rejection of the null hypothesis of equal variances across groups.

## C) Supplementary figure legends

**Supplementary Fig1. Genotyping of the selected CRISPR clones by PCR, followed by DNA sequencing confirmed the correct insertion of mVenus cassette at the Hes1 N-terminal in Cl19 cells.** (A) Schematic of Hes1 genomic locus, pre- and post-tagging with mVenus. Forward and reverse primers (FP and RP) used for PCR-based genotyping have also been designated. (B) PCR results using FP and RP performed using genomic DNA isolated from different CRISPR clonal lines and the parental MCF7 cells as a control. Take note of a single, smaller band (wild type allele without any insertion) with parental MCF7 cells. Clonal lines 7 and 19 showed two PCR bands, with the upper/larger one suggesting amplification from the Hes1 allele after mVenus insertion; clone 11 also showed a diffused upper band. Both PCR bands from clones 7, 11 and 19 were cloned in TOPO vectors and sequence verified. (C) Sequencing results for the smaller PCR band (allele without insertion) from clone 19. Sequencing data clearly showed substitution and insertion mutations near the PAM sites, resulting in the introduction of a premature stop codon, causing knocking-out of this Hes1 allele. (D) Sequencing results for the larger PCR band (allele with insertion) from clone 19. Sequencing showed the insertion of 3Flag-mVenus-Linker-HA cassette immediately after the start codon of Hes1 coding region. Sequencing data (C and D) confirmed that clone 19 is hemizygous for Hes1, with one allele knocked-out and the other one with the correct insertion of the mVenus cassette.

**Supplementary Fig2. Incucyte based growth comparisons showed that the mVenus-Hes1 CRISPR clonal line (Clone 19/Cl19) and its parental MCF7 line had similar growth properties.** (A) Selected images at different time points from the cells (parental MCF7, upper panel and Cl19, lower panel) growing in the incucyte incubator. Areas with cells growing were masked using the incucyte software to measure the percentage confluence, which, when plotted against time measured the growth rates for both cell lines. The growth rates (dark blue trend lines on the graph) for parental MCF7 and Cl19 cells were identical (B), suggesting that losing a Hes1 allele did not affect the growth properties of Cl19 cells.

**Supplementary Fig3. Protein half-life comparisons between mVenus-Hes1 from the Cl19 cells and HA-Hes1 over-expressed in the parental MCF7 suggested that tagging Hes1 with mVenus did not affect its half-life.** (A) Time course of live-imaging of Cl19 cells while they were treated with either ethanol/EtOH (control, upper panel) or cycloheximide/CHX (lower

panel). (B) Hes1 mean intensities over the time course of EtOH/CHX treatment for single cells, tracked using IMRAIS software. The slops of these intensity plots were used to estimate protein half-lives in Excel. (C) Half-lives of more than 100 cells, which were imaged while being treated with CHX, with a mean half-life of around 4hr. (D) Western blot analysis performed on lysates from MCF7 cells over-expressing HA-Hes1 and treated with cycloheximide/CHX (insight panel). The main panel shows Hes1 protein-life, estimated from the measured band intensities from the western blot, to be 4.1h. These two independent approaches to measure protein half-lives suggested that tagging Hes1 with mVenus did not affect its half-life.

**Supplementary Fig4. Periodicity analyses of Hes1 wave using multiple platforms showed that Hes1 oscillated with a periodicity of nearly 24h.** (A) Example of endogenous mVenus-Hes1 (left panel) and exogenous Ubc:NuVenus (as a non-oscillatory control, middle panel) traces were linearly detrended and run through the Lomb-Scargle periodogram/LSP pipeline, which gives normalised power spectrum for the input traces (right panel). The dominant frequency and the corresponding period values were extracted from these power spectra. Notice a dominant, high peak associated with the mVenus Hes1 trace, compared to multiple, lower peaks associated with the control trace. (B) Normalised power values for the dominant periods for experimental mixed cells were significantly higher in comparison to the Ubc:NuVenus control cells. (C) No significant differences in their periodicities were observed when Hes1 traces from Cl19 subpopulations were analysed using the LSP pipeline and compared against one another; though the low power periodicities for the Ubc:NuVenus controls were significantly different from the ones associated with the endogenous mVenus-Hes1 traces. (D) Periodicity values associated with the endogenous mVenus-Hes1 traces obtained through multiple analyses methods (*viz*. pyBOAT, LSP, Autocorrelation and Hilbert Transform) were similar, validating our findings that mVenus-Hes1 oscillated with a circadian-like periodicity and these oscillations were conferred by the endogenous dynamics of Hes1 protein expression and were not artefacts of individual analysis methods.

**Supplementary Fig5. Representing Hes1 intensities as heat-maps and cell trajectories as cell-lineage trees.** To compare multiple, single-cell Hes1 time traces against one another, we plotted Hes1 mean intensities over time as heat maps. (A) Example showing how an mVenus-Hes1 single-cell time-trace is converted into a heat map after normalising the Hes1 intensity.

Refer to materials and methods (m&m) for details about how intensity values are normalised. (B) MCF7 cells are highly amenable to single-cell live-imaging approaches. Cells can be live-imaged across generations and single cells can be tracked to generate lineage trees. B shows how a lineage tree is generated by imaging a cell across generations, through mitoses/branching points, following its daughters and grand-daughters. Information about Hes1 levels, dynamics and cell-cycle kinetics were obtained for generations with a full cell-cycle, as in this lineage tree for generation 1.

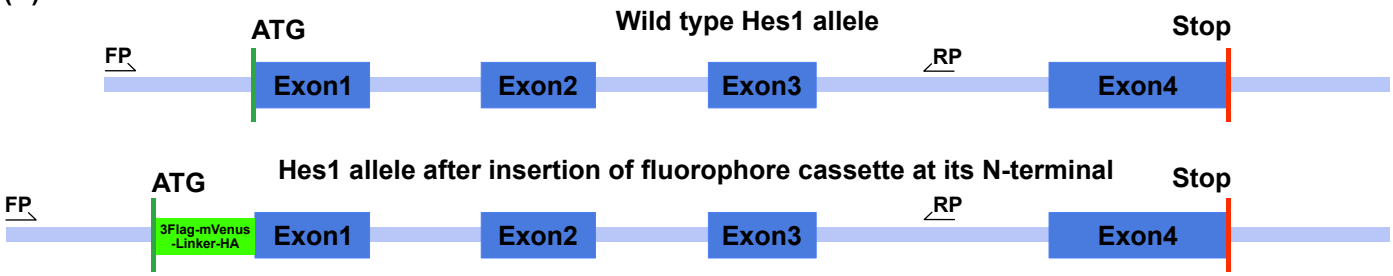
**Supplementary Fig6. Automated dip-detection (ADD) method.** (A) Example of automated dip detection on an mVenus-Hes1 time trace, which detects dips in the Hes1 expression on the basis of the global minima in the normalised Hes1 expression. (B) mVenus-Hes1 heat maps organised in the order of their cell-cycle lengths with the dips detected by the ADD as asterisks. The invariant positions of the asterisks on the heat maps suggested that the duration of the cell-cycle phase post these dips is invariable (quantified in Fig4F), while the duration before the dip is highly variable (Fig4F). (C) The variable duration of the cell-cycle before the appearance of the Hes1 expression dip strongly correlated with the overall cell-cycle length. The correlation was weak between the remaining duration of the cell cycle after the dip, and the full length of the cell cycle. (D). The data showed that the variability in the cell-cycle lengths in the MCF7 subpopulations is mainly contributed by the part of the cell-cycle prior to the dip.

**Supplementary Fig7. Silhouette analysis, followed by the GMM clustering analysis on the synthetic data justified the number of GMM cluster choice.** A) Line graph of means of multiple, repeated Silhouette scores for GMM clustering at four covariance matrix types (shared and full, shared and diagonal, unshared and full, unshared and diagonal) with number of clusters ranging from 2 to 8. The image shows that the optimal GMM clustering for our data (i.e. the highest point on the graph) was a 3-component mixture with shared (identical covariances) and diagonal (only orthogonally flexible Gaussian components) covariance matrices. (B) Line graph of the orange line in (A) showing again the mean but with +/- one standard deviation error shading of multiple repeats of such clustering. This shows that the optimal component number and covariance matrix type were consistent amongst multiple repeats of the Silhouette score analysis. (C) Five underlying dynamic shapes: 1, 2 and 3 based

on dynamics of actual data and 4, 5 sine and negative sine waves respectively. (D) Heatmap showing 10 noisy (representing biological noise) duplicates of each dynamic shape trace at three different time length scales (representing cell-cycle length heterogeneity). (E) Heatmap of stretched (using identical interpolation as with experimental data) and randomly ordered synthetic traces from (D). (F) Line graph of means of multiple, repeated Silhouette scores for GMM clustering at four covariance matrix types (shared and full, shared and diagonal, unshared and full, unshared and diagonal) with number of clusters ranging from 2 to 8. The image shows that the optimal GMM clustering for our synthetic data was a 5-component mixture with shared diagonal covariance matrices. (G) Heatmap showing GMM clustering using optimal cluster number and covariance matrix type as suggested by Silhouette analysis. This image shows all five original dynamic shapes being clustered together, thus eliminating the possibility of stretching data producing artificial clusters. These control data together clearly show that clustering is not an artefact of data stretching and the numbers of clusters suggested by Silhouette analysis are correct.

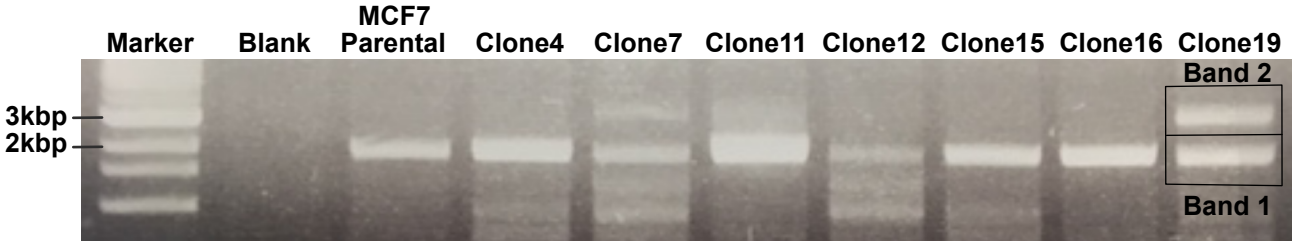
**Supplementary Fig8. GMM clustering performed on stretched, actual data followed by showing the clustered data in unstretched form** confirmed that the normalised Hes1 mean expression patterns seen with three clusters were not artefacts of stretching the Hes1 traces.

(A)




FP = forward primer
RP = reverse primer

(B)

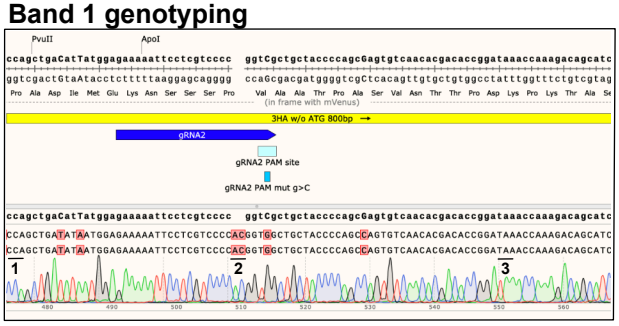



kbp = kilo base pairs

Band 2 = Allele with insert
Band 1 = Allele w/o insert

(C)

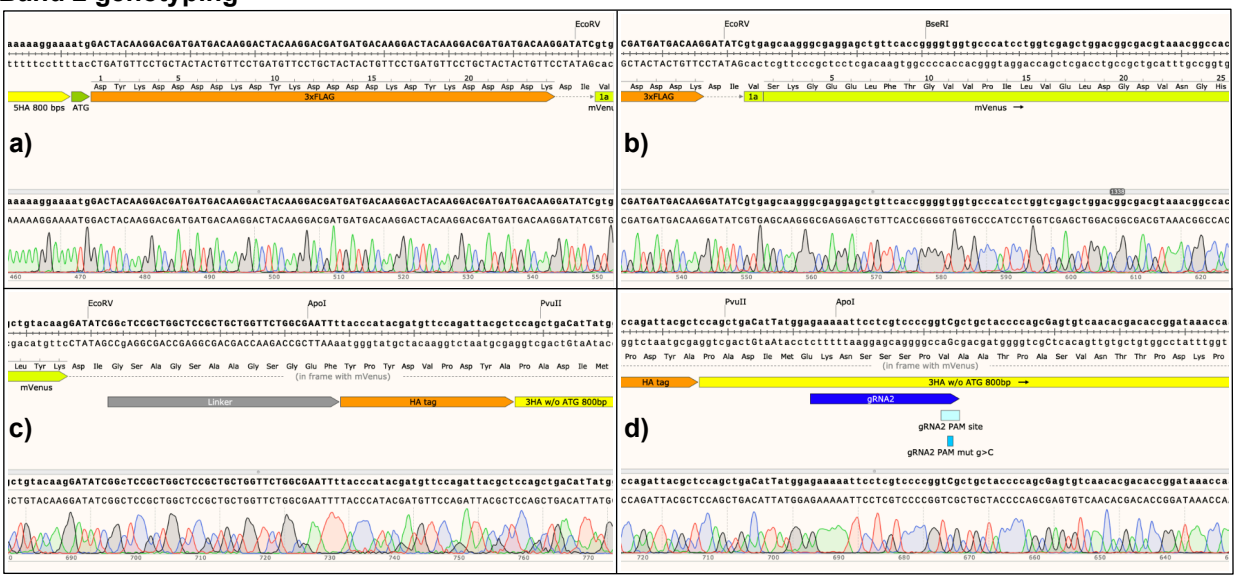
Band 1 genotyping




1 = trinucleotide for the first amino acid post ATG
2 = 2 bp insertion due to Cas9 reaction
3 = premature stop codon
Red boxes = mutations introduced due to Cas9 reaction
HA = Homology arm

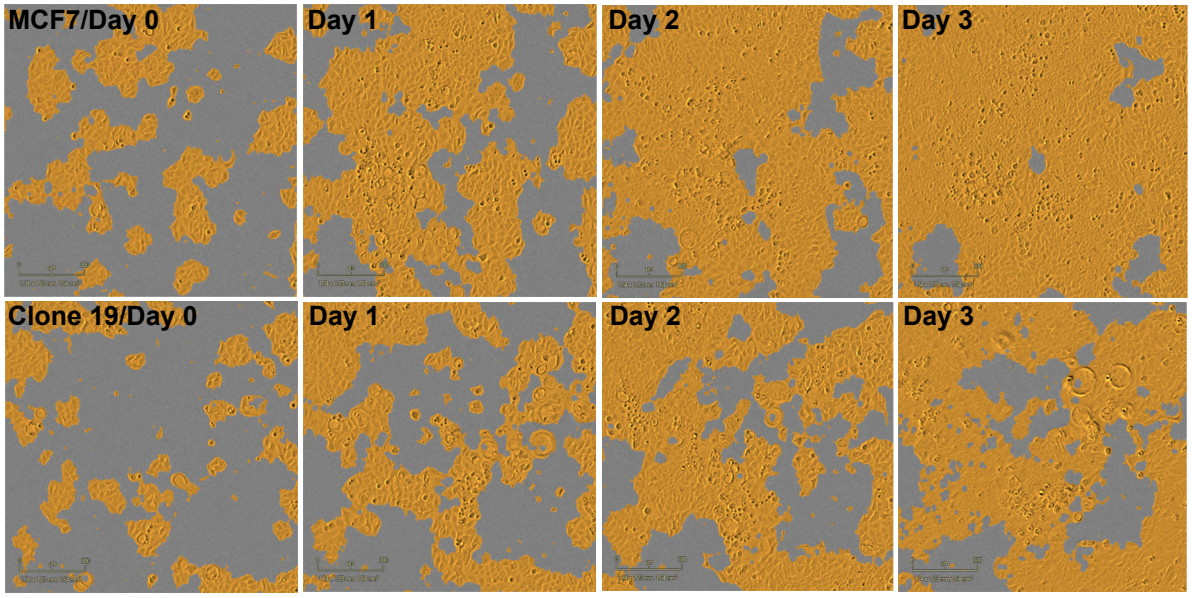
(D)

Band 2 genotyping

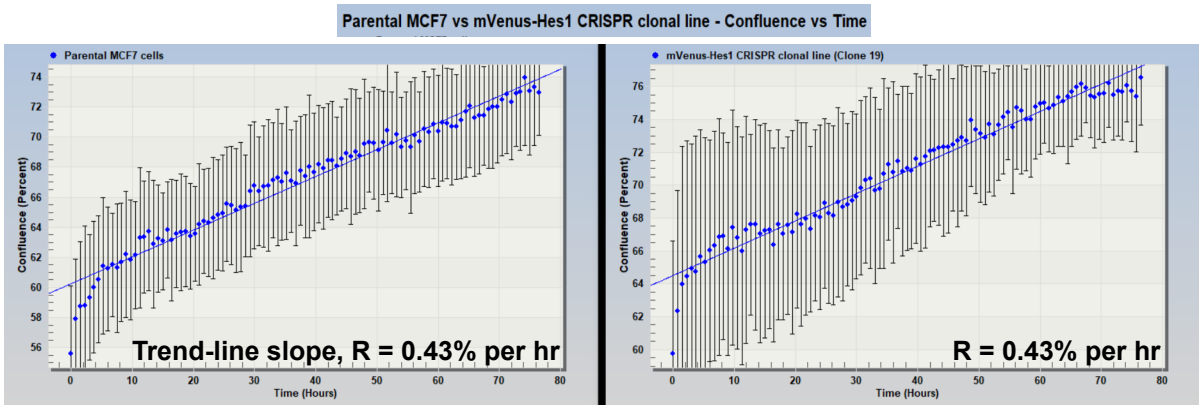



**Supplementary Figure 1. Genotyping of the selected CRISPR clones by PCR, followed by DNA sequencing confirmed the correct insertion of mVenus cassette at the Hes1 N-terminal in Cl19 cells.**

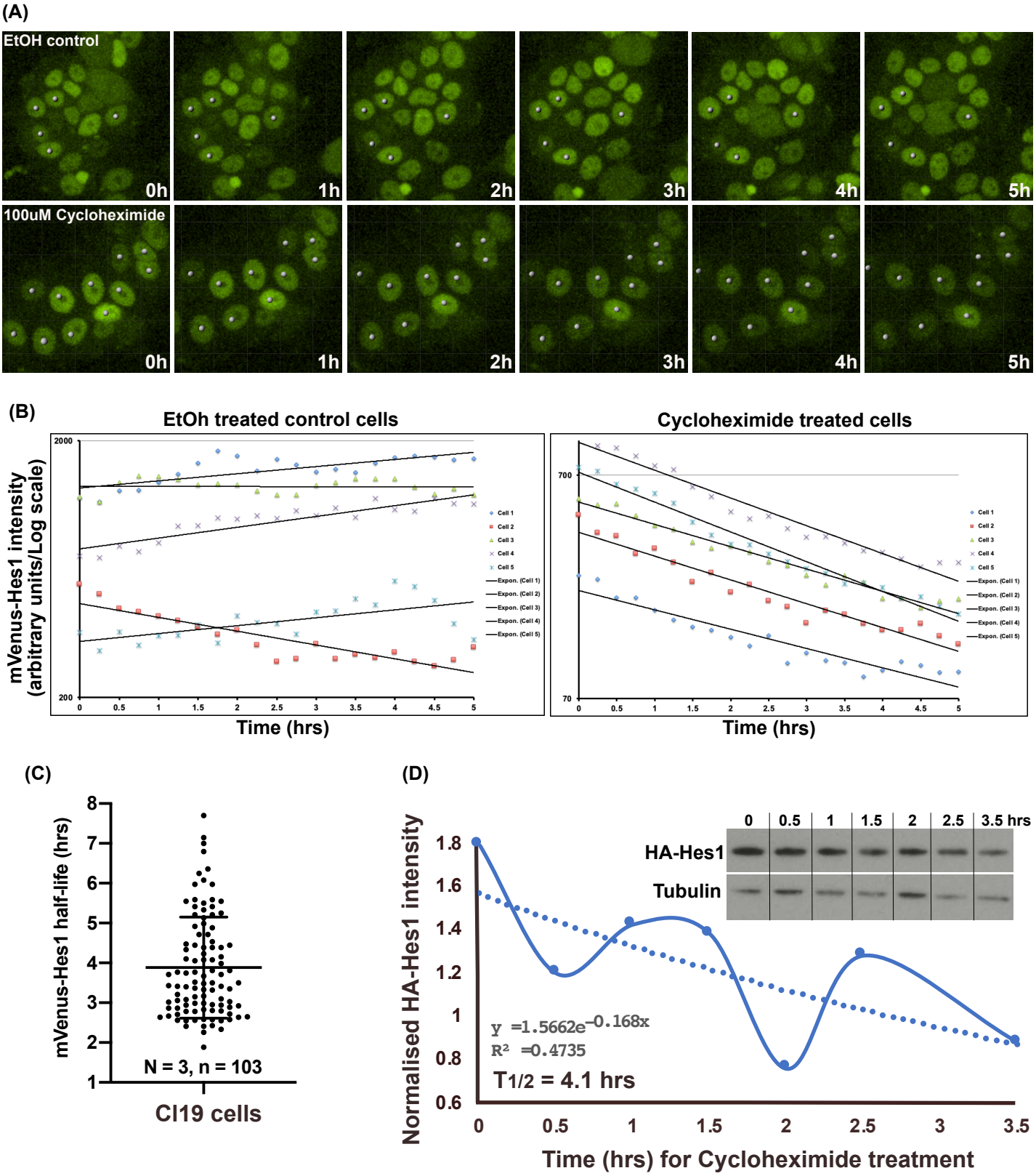
**(A)**



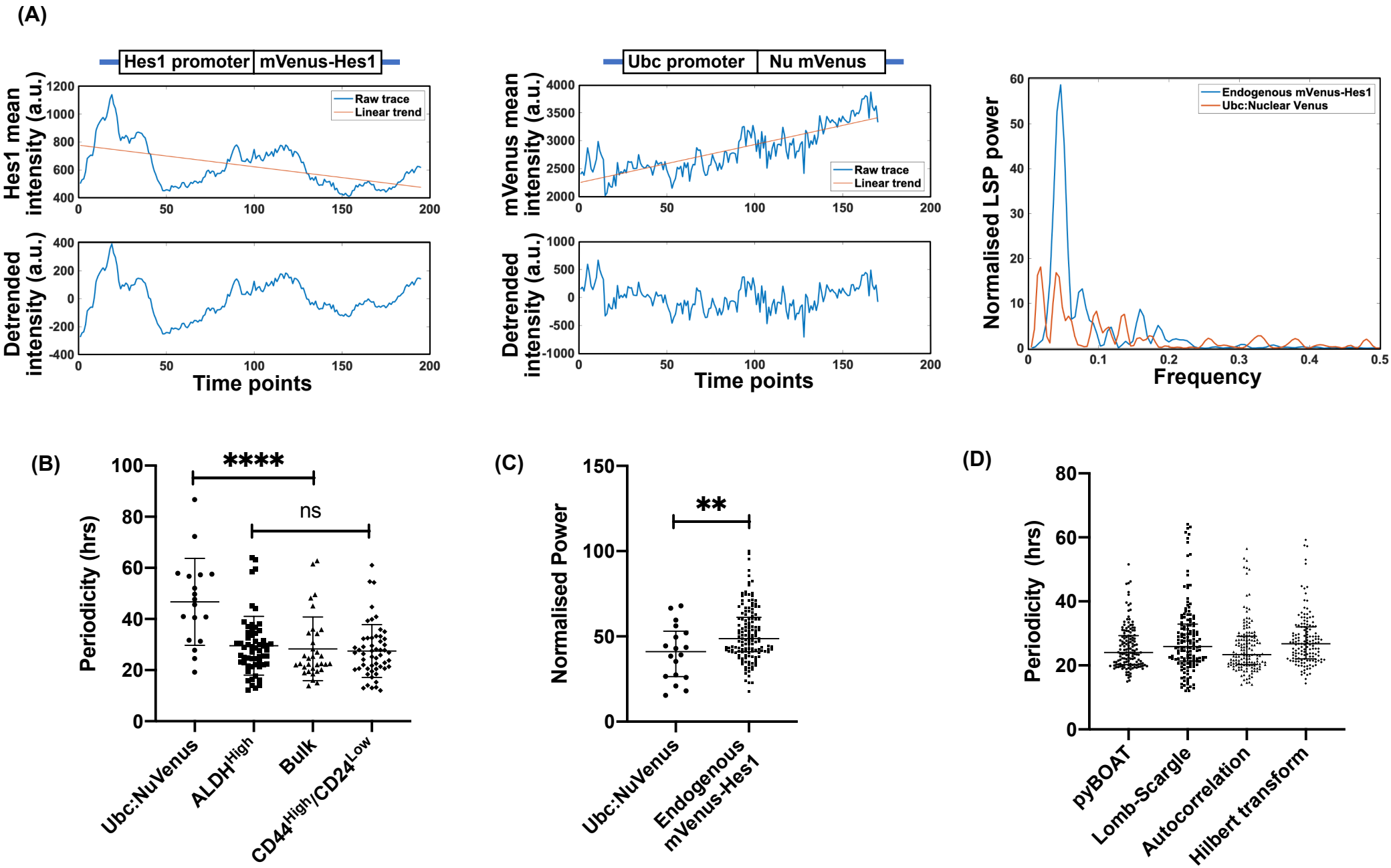

**(B)**



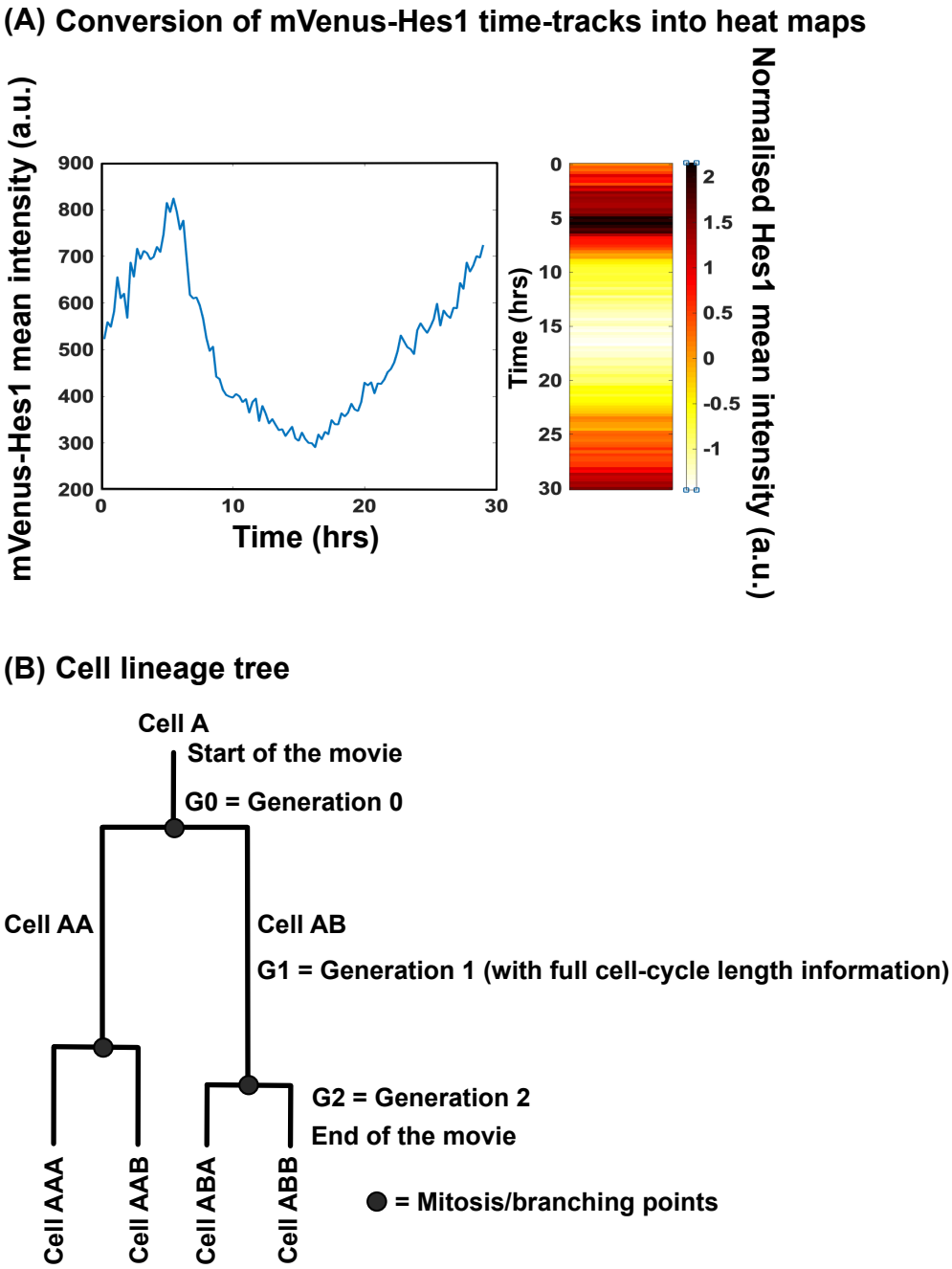

**Supplementary Figure 2. Incucyte based growth comparisons showed that the mVenus-Hes1 CRISPR clonal line (Clone 19/CI19) and its parental MCF7 line had similar growth properties.**



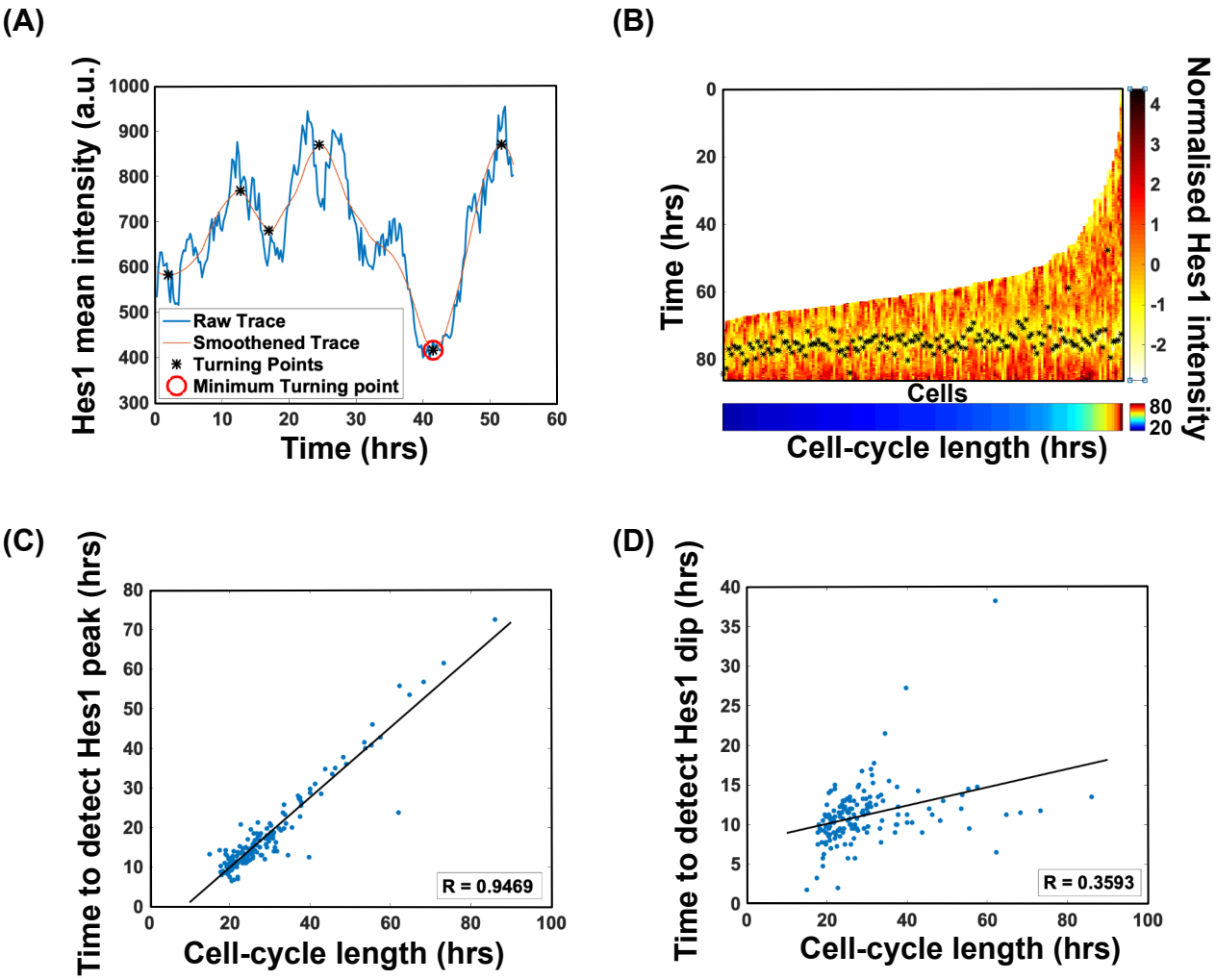

**Supplementary Figure 3. Protein half-life comparisons between mVenus-Hes1 from the CI19 cells and HA-Hes1 over-expressed in the parental MCF7 suggested that tagging Hes1 with mVenus did not affect its half-life.**



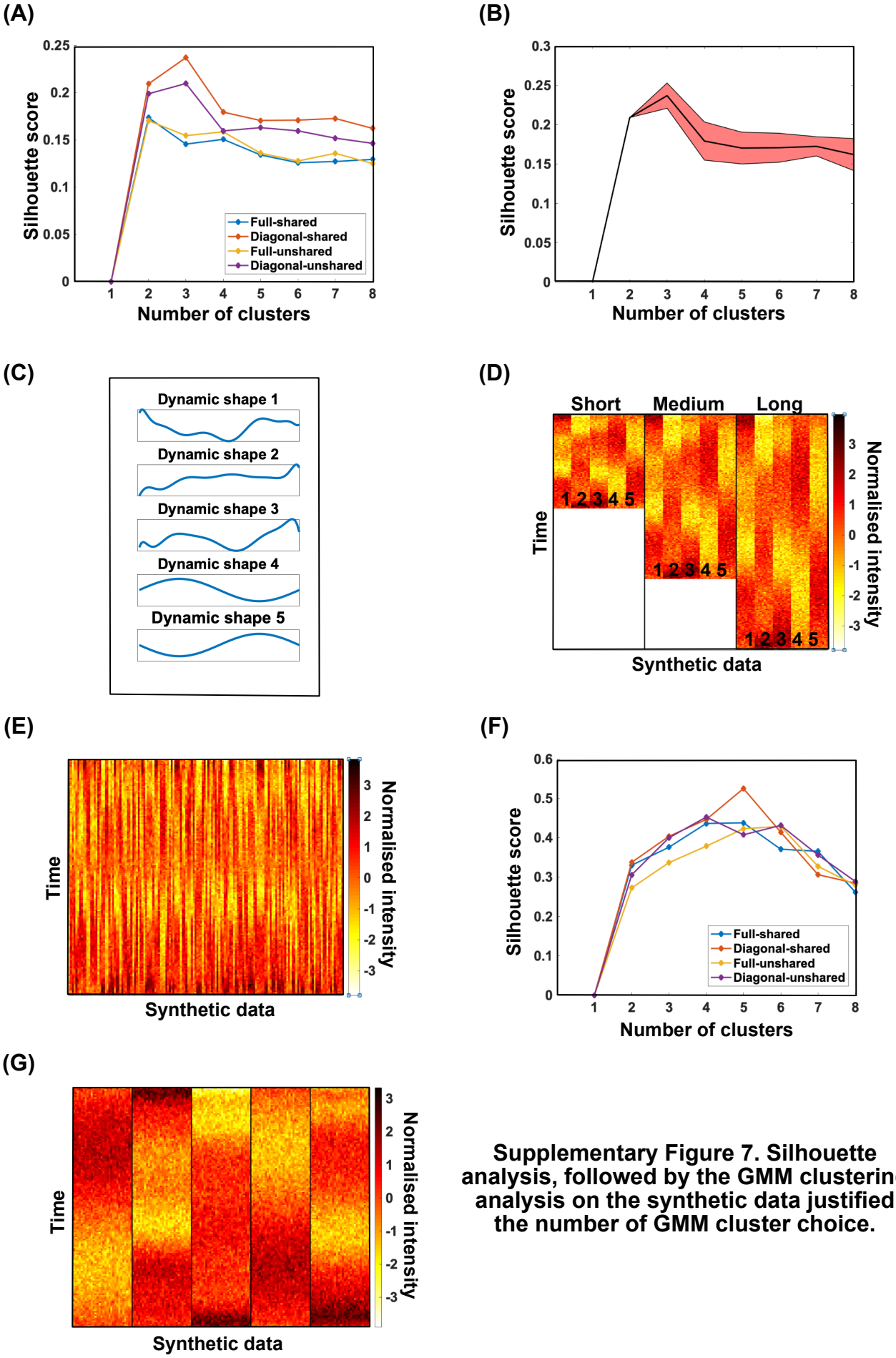

**Supplementary Figure 4. Periodicity analyses of Hes1 wave using multiple platforms showed that Hes1 oscillated with a periodicity of nearly 24h.**



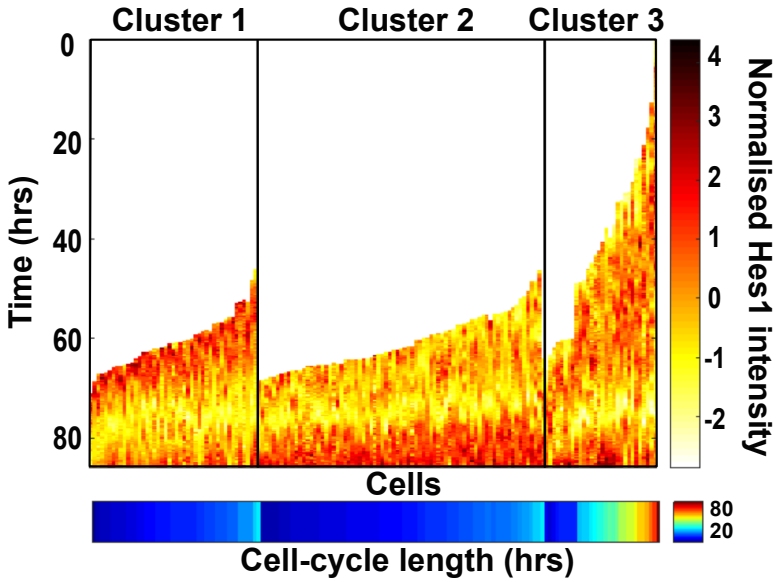

**Supplementary Figure 5. Representing Hes1 intensities as heat-maps and cell trajectories as cell-lineage trees**




**Supplementary Figure 6. Automated dip-detection (ADD) method**




**Supplementary Figure 7. Silhouette analysis, followed by the GMM clustering analysis on the synthetic data justified the number of GMM cluster choice.**




**Supplementary Figure 8. GMM clustering performed on stretched, actual data followed by showing the clustered data in unstretched form confirmed that the normalised Hes1 mean expression patterns seen with three clusters were not artefacts of stretching the Hes1 traces.**