Linage priming and cell type proportioning depends on the interplay between stochastic and deterministic factors

William M. Salvidge[1], Chris Brimson[1], Nicole Gruenheit[1], Li-Yao Huang[2], Catherine J. Pears[2], Jason B. Wolf[3] and Christopher R.L. Thompson[1*]

[1] *Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK*

[2] *Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK*

[3] *Milner Centre for Evolution and Department of Life Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK*

[*] *Author for correspondence*

Abstract

Embryonic stem cells (ESCs) can break symmetry and differentiate along different lineages, even when exposed to a seemingly identical environment. It is thought that this priming of cells towards different lineages is due to cell-cell variation, although the underlying mechanisms are poorly understood. To address this, we exploit the tractability of the social amoeba *Dictyostelium discoideum*, where cell fate choice also does not depend on spatial cues. We develop and test a model to explain quantitative experimental single cell observations of probabilistic differentiation. The model suggests that cell cycle position affects lineage choice, as previously shown, but that stochastic cell-cell variation also plays a key role. Single cell sequencing reveals genes strongly associated with fate choice exhibit extensive stochastic cell-cell expression variation. Like lineage priming genes in ESCs, they are associated with specific epigenetic modifications, which when perturbed affect their expression and disrupt fate choice. We suggest this represents an adaptive mechanism that increases developmental robustness against perturbations that affect deterministic signals.

Introduction

The robust spatial and temporal control of symmetry-breaking and cell type differentiation is a fundamental feature of developmental systems. The mechanisms underlying the spatial control of type specialization have been extensively studied (Slack, 2014). However, there are many examples of robust cell fate choice in the absence of spatial cues. Examples include embryonic stem cells (ESCs) in culture (Pauklin and Vallier, 2013), lineage choice in the mouse blastocyst (Saiz et al., 2016; Yamanaka et al., 2010) and 'bet hedging' strategies in microbial populations (Maamar et al., 2007; Süel et al., 2007). In these cases, differentiation is probabilistic and is thought to depend on cell-cell heterogeneity. Intrinsic variation in cell physiology, redox state and gene expression arising from differences in cell cycle position, cell size, inheritance of cell contents upon cell division and transcriptional bursting have all been linked to fate choice (Huh and Paulsson, 2011; Raj et al., 2006). These factors can produce robust cell type proportioning if the probability of cells being in different states is predictable at the cell-population level (if the population size is sufficiently large).

Despite the widespread use of heterogeneity to facilitate developmental decision making, we still have a limited understanding of the underlying mechanisms or how variation is tuned to result in robust proportioning. Work has centred around the identification of factors that 'prime' cells for differentiation before they commit to a particular cell fate and express genes associated with specific cell lineages. For example, genes associated with ESC differentiation exhibit bivalent H3K4me3 and H3K27me3 epigenetic marks, suggesting a role in coordinating their expression (Azuara et al., 2006; Bernstein et al., 2006). H3K4Me3 deposition has also been shown to be sensitive to redox state, thus providing a potential explanation for the link between reactive oxygen species and ESC differentiation (Ulfig and Jakob, 2024). It has also been suggested that the level of stochasticity in the expression of lineage associated genes could influence the number of cells that are in the primed state and thus the number of differentiating cells (Desai et al., 2021). Finally, deterministic differences between cells, such as cell cycle position have been shown to influence the response to differentiation cues and lineage choice (Pauklin and Vallier, 2013).

The social amoeba *Dictyostelium discoideum* provides a uniquely tractable model system to study the mechanisms underlying lineage priming and the control of robust probabilistic differentiation. In response to starvation *D. discoideum* cells undergo a multicellular developmental programme to build a fruiting body consisting

of hardy spore cells and dead stalk cells that aid dispersal from deteriorating environments. There is assumed to be selective pressure to ensure that the 'fittest' or most energy rich cells are chosen as spores, which requires mechanisms that bias differentiation. Furthermore, there is assumed to be strong selective pressure to ensure robust output of 'optimal' cell type proportioning. Excess stalk cell production is costly because it reduces the number of viable spores available for dispersal, while underproduction of stalk can compromise fruiting body architecture, limiting the success of spore dispersal (Buttery et al., 2009; Madgwick et al., 2018; Rodrigues and Gardner, 2022; Wolf et al., 2015).

Symmetry breaking and initial cell type differentiation in *D. discoideum* does not depend on spatial cues. Instead, cell fate choice has been associated with differences in cell cycle position at the time of starvation (Gomer and Firtel, 1987; Gruenheit et al., 2018). Cell cycle position affects the threshold of responsiveness to diffusible inducers of spore or stalk cell differentiation, such as cAMP and DIF-1 (even though all cells experience the same amount of each signal) (Chattwood et al., 2013; Gruenheit et al., 2018). Under this scenario, stochastic variation in cell cycle length generates a steady-state distribution of cell cycle positions. Optimal proportioning of stalk and spore cells is presumably reached because cell fate propensity has been evolutionarily tuned through the cell cycle when a sufficiently large number of cells is randomly sampled. It is likely that this provides an adaptive mechanism to select the least fit cells to differentiate as stalk (Zahavi et al., 2018). Cells that have just divided are primed to become stalk and are more sensitive to stalk inducing signals (and less sensitive to spore inducing signals). Post mitotic cells have just divided their cytoplasm and are likely to be energy poor compared to cells in G2. This idea is supported by the finding that glucose depletion causes cells to arrest around mitosis and differentiate as stalk cells (Gruenheit et al., 2018; Thompson and Kay, 2000).

These observations provide insights into the effects of cell cycle position heterogeneity on cell fate at a population level. However, we still have a poor understanding of how single cells interpret heterogeneity to result in probabilistic decision making or how heterogeneity is tuned to result in robust cell type proportions. To understand how cell-cell variation can be exploited to control fate choice and generate robust proportioning in the absence of cell-cell communication, we have used a combination of mathematical modelling and experimentation. The model is based on observations of probabilistic differentiation in *D. discoideum*. We find quantitative behaviour can only be explained by a model that not only encompasses deterministic

cell cycle effects, but also the effects of stochastic cell-cell variation on the responsiveness of cells to differentiation inducing signals. Experimental observations, using single cell RNA-seq, reveals a set of genes that show extensive stochastic gene expression variation at the time of starvation. These genes show many hallmarks of lineage priming genes. They are upregulated as cells undergo a programme of differentiation and exhibit cell type specific gene expression. A genome wide screen reveals that these genes, together with cell cycle associated genes, are also more likely to be required for cell type differentiation. Stochastically expressed genes also exhibit differential H3K4 methylation. Perturbation of H3K4 methylation preferentially affects the level of their expression (and thus cell-cell variation), as well as cell type proportioning during development. These observations suggest that deterministic variation in cell cycle position acts together with stochastic variation in developmental genes to control probabilistic cell fate choice. Finally, we suggest this system is evolutionarily advantageous because it allows the use of deterministic information on the status or quality of cells (e.g., their energetic state), yet protects against potentially catastrophic effects of large-scale perturbations that cause cells to exhibit inadequate variation in deterministic properties. These studies provide further evidence supporting the key role of stochastic variation in developmental gene expression and cellular decision making.

Results

A combination of stochastic and deterministic variation explains lineage priming and fate choice in *D. discoideum*

To understand the mechanisms underlying lineage priming and fate choice, we first considered an earlier deterministic model that was designed to explain cell cycle dependent fate propensity in *D. discoideum* (Gruenheit et al., 2018). The model assumes all cells experience similar levels of a stalk-inducing factor, such as DIF-1. DIF-1 is easily diffusible and the rapid movement of cells within an aggregation result in a well-mixed population. Furthermore, levels of DIF-1 are regulated by negative feedback that can buffer perturbations in DIF-1 synthesis (Insall et al., 1992). Experimental observations instead suggest cells differ in their intrinsic threshold of responsiveness to this signal, with those above this threshold deterministically differentiating into pre-stalk cells (Chattwood and Thompson, 2011; Stevense et al.,

2010). Consistent with the wealth of experimental evidence, the model assumes that these differences are partly determined by cell cycle position (Gomer and Ammann, 1996; Gomer and Firtel, 1987). It has been proposed that this is because a cell cycle associated factor (CCAF) rapidly accumulates at mitosis and increases stalk propensity (Gruenheit et al., 2018). Following mitosis, the level of CCAF decays, resulting in a decreasing propensity of cells to adopt stalk cell fate (and increasing spore propensity) as the cell cycle progresses. This model provides a relatively good fit to observed population level cell fate proportions. However, a deterministic factor governing cell fate would be expected to drive cells in the same state towards the same fate. Instead, however, cells at the same cell cycle position can adopt different fates (Gruenheit et al., 2018). Furthermore, a truly deterministic model would be expected to result in a discrete change from stalk to spore fate at the point in the cell cycle where CCAF level drops below the threshold that results in stalk fate (see Figure 1 parts A.i and A.ii). Instead, there is a gradual decrease in stalk propensity after mitosis.

To develop a model that generates the sort of probabilistic differentiation observed in *D. discoideum* we first incorporated the deterministic influence of the cell cycle on responsiveness or sensitivity to stalk inducing factors (CCAF) (Figure 1A.i):

$$A_t = A_0 - \beta t \tag{1}$$

where A_0 is the starting level of CCAF, β is its rate of decay, and t is the amount of time after the end of the previous mitosis. CCAF is measured on the scale of its biological effect rather than in terms of its molecular concentration. In theory, the decay in the level of CCAF could show a range of shapes (e.g., exponential/convex, linear, or parabolic/concave), which depends both on how the underlying factor(s) decay at the molecular level and how molecular concentrations translate into biological effects. We assume linear decay both for simplicity, and because fitting alternative models to experimental data indicate that it is the best fit model (see Supplementary text). Cells with a CCAF level above a threshold value (denoted R) adopt stalk cell fate (in response to stalk-inducing factors). As a result, stalk propensity at time t after the end of the previous mitosis (P_t) is determined by the proportion of cells with a value of CCAF (given by A_t) above R. This model would produce the steplike pattern of cell fate through the cell cycle (as described above), where all cells adopt stalk cell fate during the period of the cell cycle where CCAF levels are above the threshold ($A_t \geq R$) and switch to spore cell fate when CCAF levels decay below the threshold ($A_t < R$)

(Figure 1A.ii). This steplike pattern will occur irrespective of how CCAF levels change through the cell cycle (i.e. regardless of whether it is linear or nonlinear) since cells in the same cell-cycle state can only be above or below the CCAF threshold that results in stalk fate.

A purely deterministic model clearly does not recapitulate the quantitative shift in cell fate propensity observed experimentally (Figure 1A.iii). We next reasoned that, for the system to show a deterministic change through the cell cycle, but result in a probabilistic output (i.e., stalk propensity), it must integrate another source of cell-cell variation that is independent of the cell cycle (along with a deterministic cell-cycle dependent factor, CCAF). We therefore hypothesised that there is also a stochastically variable cell-cycle independent factor (CCIF) contributing to sensitivity to stalk-inducing factors. The addition of such intrinsic variability makes cell fate probabilistic by allowing otherwise identical cells to adopt a distribution of possible responsiveness to stalk-inducing factors. We further hypothesized that this noisiness across cells arises from stochastic expression of a set of genes contributing to CCIF. For this, we adapt the logic of the telegraph model of gene expression (Peccoud and Ycart, 1995), where expression of gene i at time t, $e_i = \{0, 1\}$, is a Bernoulli process with probability p_i that $e_i = 1$ (meaning that p_i is scaled to the size of the temporal interval). Although genes might logically vary in the value of p_i, such variability does not impact our results, so we simply consider the average probability of being expressed across all genes: $\bar{p}$. When gene i is expressed, it contributes x_i to the total value of CCIF, making its realised contribution $g_i = e_i x_i$ and expected contribution in any time interval $E[g_i] = p_i x_i = \bar{p} x_i$. We assume that genes vary in their contribution to CCIF according to $x \sim N[\bar{x},\ s^2]$. The total level of CCIF at time t is the sum of the g_i values for the set of G stochastically expressed genes: $\mu = \sum_{i=1}^{G} g_i = G\bar{p}\bar{x}$ and (following the law of total variance (Cornell and Benjamin, 1970)) its variance would be $\sigma^2 = G\bar{p}s^2 + \bar{p}(1-\bar{p})G\bar{x}^2$. This scenario results in a normal distribution of CCIF levels, CCIF$\sim N[\mu, \sigma^2]$, which we use to model the properties of the aggregate effect of all stochastically expressed genes on CCIF.

If we were to only consider the contribution of this stochastic variation, the proportion of cells adopting stalk fate (P_t) would, be determined by the proportion of the distribution that exceeds the threshold value (R) that leads to stalk fate (Figure 1B.i), which is given by the complementary cumulative distribution of CCIF values above the threshold R:

$$P_t = \frac{1}{2}\left(1 + \mathrm{erf}\left[(\mu - R)/\left(\sigma\sqrt{2}\right)\right]\right), \quad (2)$$

where erf is the Gauss error function, defined as $\mathrm{erf}(z) = \left(2/\sqrt{\pi}\right)\int_0^z e^{-x^2}dx$. Under this scenario, the proportion of prestalk cells would remain constant throughout the cell-cycle (Figure 1B.ii), which does not follow the observed progressive shift in stalk fate probability after mitosis (Figure 1B.iii).

We integrate the deterministic influence of cell-cycle state and noisy gene expression by considering that the total level of sensitivity to stalk-inducing factors is the sum of the value of CCAF and CCIF. We achieve this integration by assuming that the mean sensitivity across cells has a constant component (CCIF) given by μ (cf. eqn. 2), and a cell-cycle dependent component (CCAF) given by A_t (cf. eqn. 1). This produces an expression for mean sensitivity at time t (C_t) that is analogous to equation (1): $C_t = \mu + A_0 - \beta t = C_0 - \beta t$, where $C_0 = \mu + A_0$ (i.e., is the sum of the mean CCIF value and starting CCAF value). Note that the change in average sensitivity through the cell cycle depends solely on changes in CCAF levels (βt), but the change in the average stalk propensity does not change linearly through the cell cycle because it depends on the proportion of cells with a value above R (see Figure 1.C.i). Using this expression we can rewrite the overall stalk propensity (P_t) at time t by replacing μ in equation (2) with C_t (see Figure 1C.i):

$$P_t = \frac{1}{2}\left(1 + \mathrm{erf}\left[\frac{C_0 - \beta t - R}{\sigma\sqrt{2}}\right]\right) \quad (3)$$

(which represents the integral over the distribution of sensitivity values above the value R at time t). This 'stochastic-deterministic' model generates a non-linear change in stalk propensity through time (Figure 1C.ii) that matches the probabilistic nature of cell stalk fate after mitosis (Figure 1C.iii). The model is also flexible in that, at its limits, it can produce anywhere from the step function (Figure 1A) expected for a purely deterministic process (i.e., as $\sigma \to 0$) and the constant proportioning (Figure 1B) expected for a purely stochastic process (i.e., as $\beta \to 0$) (see Supplementary text). To facilitate fitting of the model in equation (3) to experimental data, we can reduce the parameter space by combining the three constants (C_0, R, and σ) into a single term, denoted C_0^*, which represents a sort of reference point for the model. This rescaling is achieved by assuming that the parameters are all measured in standard deviation units and that sensitivity is measured as a distance from the threshold that leads to stalk cell

fate (R), such that $C_0^* = C_t - R$. This results in a two parameter version of equation (3), $\frac{1}{2}\left(1 + \mathrm{erf}\left[(C_0^* - \beta t)/\sqrt{2}\right]\right)$, that can be fitted to the stalk propensity data (see below) to estimate the two parameters (C_0^* and β), where the parameter estimates are in standard deviation units (meaning C_0^* represents how far mean sensitivity is from the threshold, measured as a Z-score). To test this model empirically, we started by fitting the model to existing data on stalk propensity through the cell cycle (see Supplementary Data for the original raw data and adjusted values used in the analysis). These data come from an experiment (Gruenheit et al., 2018) where time lapse microscopy was used to group cells according to their cell cycle position at the time of starvation. The fate of each cell was monitored using live cell reporter genes. For our analysis, we used the measures of relative stalk propensity over six hours (after which cells stochastically re-enter mitosis). We find that the stochastic-deterministic model provides an excellent fit to the data (adjusted *R*-squared = 0.92, *AIC* = –56.37), yielding estimates of C_0^* = 0.57 and β = 0.41 (Figure 1C.iii) that correspond to an expected steady-state stalk propensity of 0.35 and an expected starting propensity of 0.72. Moreover, the stochastic deterministic model provides a significantly better fit to the data than the previously proposed model for exponential decay of CCAF (Gruenheit et al., 2018) (it is 98.35 times more likely; see Supplementary text and Supplementary Figure 1). We also evaluated the support for our model assumptions by comparing the fit of other models derived using alternative assumptions. We evaluated the assumption of linear decay in CCAF by comparison to models with exponential, quadratic, and cubic decay functions (which together capture a broad range of possible shapes), and the assumption of Gaussian variation in CCIF by fitting a model based on gamma distributed variation (which can capture a diversity of distributions, including ones approximating normality). These model comparisons all support our model assumptions (see the Supplementary text for the methods and results of these comparisons).

Cell cycle independent stochastic gene expression variation is extensive in growing cells

The stochastic-deterministic model suggests that cell fate choice in *D. discoideum* should depend on stochastic variation in the expression of genes associated with fate choice, as well as deterministic cell cycle dependent cell-cell variation. Single cell RNA-seq (scRNA-seq) data from *D. discoideum* cells prior to starvation and exposure

to differentiation inducing signals has previously been used to identify cell cycle position associated gene expression variation (Gruenheit et al., 2018). Because these data are from a relatively small number of cells, but sequenced to high depth, it allowed us to determine the extent to which there is also cell cycle independent gene expression variation. We first determined the coefficient of variation (CV^2) of expression for all genes. As expected, this tends to decrease as average expression level increases (Supplementary Figure 2). When this trend is accounted for, genes with greater cell-cell variability than expected for their level of expression (FDR < 0.05) can be identified (Figure 2A). Next, we used Area Under the Receiver Operating Characteristic Curve (AUROC) to identify genes with variation that can be explained because their variation in expression is associated with cells previously shown to be in early G2, late G2 or M/S phase (Supplementary Figure 2). Despite using a conservative AUROC threshold (>0.65), when these genes were removed, many genes remain that exhibit variation that cannot be explained by differences in cell cycle position (Figure 2B). Unlike cell cycle genes, principal component analysis does not result in cell groupings (Figure 2C), which suggests that this represents stochastic variation rather than a consequence of hitherto unknown extrinsic cues. Finally, we were able to determine the extent to which cell cycle associated genes exhibit stochastic gene expression. For this, the CV^2 of each gene was recalculated within groups of cells from each of the different cell cycle stages. Most cell cycle dependent genes were found to exhibit greater within group variation than expected (Figure 2D). This is not due to the low level of expression at specific cell cycle stages, as variation is higher at all stages, including when they are maximally expressed (Supplementary Figure 2). These results thus reveal that stochastic gene expression variation is widespread in growing cells.

Stochastically expressed genes are associated with cell fate determination

To test whether genes that exhibit variability in their expression are associated with cell fate choice, we compared the developmental timing of expression of stochastically expressed genes, cell cycle associated genes and non-variable genes. The average expression of each gene during developmental stages was compared to expression during growth to generate a developmental index (where 0 is exclusive to growth and 1 is exclusive to development) (de Oliveira et al., 2019). Stochastic genes were greatly enriched ($p \leq 0.001$, binomial test) for developmental genes (index >= 0.9), whereas cell

cycle and non-variable genes showed no enrichment compared to the genome wide expectation (p>0.5, binomial test) (Figure 3A). Next, we tested whether any of these groups of genes were associated with prestalk or prespore cell fate. First, RNA-seq data from prestalk and prespore cells was analysed to identify genes that exhibit cell type specific gene expression (where 0 is exclusive to prespore cells and 1 is exclusive to prestalk cells). Again, stochastic genes were significantly enriched in cell type specific genes (p≤0.001), with both prestalk (p≤0.001, binomial test) and prespore genes (p= 0.013) contributing to this enrichment (Figure 3B). Next, we determined whether these genes are more likely to be required for fate choice. We performed an unbiased large scale REMI-seq forward genetic screen (Gruenheit et al., 2021) to identify genes required for stalk cell differentiation. REMI-seq technology permits the abundance of thousands of mutants to be simultaneously quantified before or after a selection regime, such as selection imposed by the ability to undergo prestalk cell differentiation. The REMI-seq library was plated at low cell density and treated with cAMP to induce competence to differentiate followed by treatment with DIF-1 to induce prestalk cell differentiation (Figure 3C). Prestalk cells terminally differentiate as dead stalk cells after prolonged DIF-1 incubation, and thus surviving mutants with defects in prestalk cell differentiation can be enriched. After 2 and 6 rounds of growth and selection, gDNA was prepared from each biological replicate for sequencing and quantitative analysis. In order to ensure that enrichment was due to a failure to respond to DIF-1, rather than increased growth rate, we compared this set of mutants to a control selection in which cells were simply taken through an equivalent number of generations of growth (Gruenheit et al., 2021). An additional control screen was performed to identify mutants that are incompetent to differentiate at all (as either stalk or spore) because they fail to respond to cAMP. Cells were incubated in the presence of the cAMP analogue, 8-Br-cAMP, which triggers spore cell differentiation. Cells that cannot differentiate as spores were killed by detergent treatment, thus reducing their frequency in the population. REMI-seq identified 244 mutants with defects in stalk cell differentiation (Figure 3D, Supplementary Data). A subset of these mutants was recreated in the parental strain and tested in stalk cell induction assays and most mutants (8/9) exhibited defects in stalk cell differentiation (Figure 3E), illustrating the quantitative success of the REMI-seq approach. This allowed us to test whether genes required for differentiation are more likely to exhibit cell to cell gene expression variation. REMI mutants were assigned to 199 genes with intragenic insertions, or mutations where the REMI insertion lay within upstream promoter sequences (within

500bp of the transcription start site). These genes were significantly enriched for genes with variable expression ($p \leq 0.001$, binomial test) (Figure 3F). This is due to both cell cycle associated and stochastically expressed genes, as the relative number of genes identified in each class does not vary significantly from expected (chi square, $p=0.77$). These genes also exhibit greater variability than expected when the CV^2 was normalised to their expression (*** $p \leq 0.001$, t-test) (Figure 3F). These results suggest gene expression variation is a feature of genes associated with fate choice and cell type proportioning.

Stochastically expressed developmental genes exhibit differential patterns of H3K4 methylation

Fate choice in *D. discoideum* shares features with lineage priming in embryonic stem cells, including the extensive cell-cell variation of genes associated with differentiation (Chang et al., 2008). The expression of genes associated with the differentiation is thought to be under epigenetic control due to the presence of specific patterns of epigenetic marks, including the co-occurrence of H3K4me3 and H3K27me3. The role of these modifications is not fully understood due to the difficulty with which epigenetic marks can be altered at a genome wide scale in higher organisms. The apparent absence of polycomb like proteins in *D. discoideum* suggests H3K27me3 modification is unlikely to play any role (Kaller et al., 2006). However, H3K4 mono, di, or tri-methylation (H3K4Me1-3), which is dependent on Set1/COMPASS, is present (Chubb et al., 2006b). Similarly, H3K9/K14 acetylation is present, which is consistent with the idea that H3K4me3 targets the Gcn5 H3K9/K14 histone acetyl-transferase to specific loci (Huang et al., 2021). Finally, recent studies have linked histone modifications to the control of transcriptional burst frequency (Weinberger et al., 2012; Wu et al., 2017), which will in turn affect the level of cell-cell variation in transcription. We therefore tested whether genes that exhibit cell-cell variation in expression in *D. discoideum* also show hallmarks of differential regulation by H3K4 methylation. Analysis of ChIP-seq data (Wang et al., 2021) revealed that H3K4 methylation exhibits a characteristic gene expression level dependent pattern around the transcription start site and gene body (Barski et al., 2007; Soares et al., 2017) (Figure 4A). To compare patterns in variable and non-variable genes, it was thus necessary to divide genes into bins with similar gene expression levels. Ten random samples of non-variable genes with the same distribution of expression levels as those seen in variable genes were used to compare

the profile and number of genes with H3K4 methylation. This revealed variably expressed genes exhibit different profiles around the gene promoter and gene body (Figure 4B). In addition, when these were divided into stochastic and cell cycle genes, both groups were enriched for H3K4Me1 (binomial test $p<0.001$, Figure 4C), whilst stochastic genes were depleted for H3K4Me3 (binomial test $p<0.001$, Figure 4C).

H3K4 methylation is required for normal expression of stochastic genes

Variably expressed genes exhibit different patterns of H3K4 methylation compared to non-variable genes. To test whether their expression also depends on H3K4 methylation, we examined a *D. discoideum* Set1 knockout strain in which H3K4 methylation is abolished (Chubb et al., 2006a) (Figure 5A). Bulk RNA-seq was performed on wild type and Set1 mutant cells. Most genes increased in expression, with 459 up-regulated, compared to 132 down-regulated genes in *Set1*$^-$ mutant cells (Figure 5B). The affected genes were strongly enriched for variable genes identified by scRNA-seq (binomial test, $p<0.01$, Figure 5C). The level of enrichment increases as the stringency with which variable genes are defined is increased. This suggests that Set1 dependent genes are among the most variably expressed genes during growth (Figure 5C, Supplementary Figure 5). Furthermore, if variable genes are divided into cell cycle and non-cell cycle associated stochastic genes, Set1 dependent genes are enriched in both groups (binomial test, $p<0.01$) (Figure 5D), suggesting a general role in controlling the expression of variably expressed genes associated with fate choice. These findings were validated with reporter genes in which the promoters of representative Set1 dependent variable genes (*GtaU*, *HspF-2*) were used to drive RFP expression (Supplementary Figure 6). The *actin15* promoter was used as a control because it exhibits little cell-cell expression variation and is unaffected by disruption of Set1 in RNA-seq. Actin15-RFP expression was similar in wild type and Set1$^-$ mutant cells (Supplementary Figure 6), but both *GtaU* and *HspF-2*-RFP expression increased markedly in most cells (Supplementary Figure 6).

Changes in gene expression can be caused by changes in burst frequency or burst size. To better understand the effects of Set1 on gene expression, we adapted the stochastic deterministic model to examine the properties of expression variation of individual genes instead of their aggregate influence on CCIF levels. For this, the probability that a gene is expressed, p_i, can be interpreted as a measure of the burst frequency, since burst frequency will dictate the probability of observing gene i being

expressed in interval t. When gene i is expressed, it produces an amount of transcript given by b_i, which is a measure of burst size. Cells in the 'off' state are assigned a value of $b_i = 0$. There can be variability in the level of expression of gene i among cells in the 'on' state, denoted ε^2, which can reflect both error variation in experimental estimation and biological noise/variability across cells, such that the distribution among these cells is $\sim N[b_i, \varepsilon^2]$. Hence, the average level of expression for gene i across a set of cells (which includes cells in both the on and off states) is $p_i b_i$, while its variance in expression would be $p_i \varepsilon^2 + p_i(1 - p_i){b_i}^2$. Consequently, an increase in burst frequency would be expected to lead to an increase in mean expression. Because both the expected mean and variance in expression levels change as a function of a change in burst frequency, we can measure the impact of a change in burst frequency on expression variability as the squared coefficient of variation (CV^2), which has an expected value of $(1 - p_i)/\ p_i + \varepsilon^2/(p_i {b_i}^2)$. The value of CV^2 would, therefore, be expected to decrease if burst frequency were to increase (i.e., $\Delta p_i > 0$) and vice versa.

To test how Set1 affects gene expression, we examined the effects of Set1 disruption on gene expression in single cells. scRNA-seq data was generated from wild type and *set1*$^-$ mutant cells. The level of gene expression and cell-cell variation (CV^2) was determined for genes shown by bulk RNA-seq to be under the control of Set1 and by high depth scRNA-seq to exhibit significant variation in wild type cells. We find that the average expression level of these genes was affected in *set1*$^-$ mutant cells (Figure 5E). As expected, based on the model predictions, the level of cell-cell variability measured by CV^2 significantly decreased in up-regulated genes (Figure 5F), which is consistent with an increased burst frequency. These findings are supported by quantification of *GtaU* and *HspF-2* reporter gene expression levels and cell-cell variability (Figure 5G and H). The observed changes in gene expression are thus consistent with the idea that disruption of Set1 affects gene expression by increasing burst frequency and thus the number of expressing cells (which also reduces cell-cell gene expression variation) rather than simply further increasing the level of expression in the subpopulation of cells that already express these genes.

Loss of Set1 dependent H3K4 methylation affects cell fate choice

Genes associated with cell fate choice exhibit cell-cell variation in gene expression, they are enriched for Set1 dependent histone modifications and their normal expression is dependent on Set1. We therefore tested whether Set1 disruption

affects fate choice. Disruption of Set1 did not result in defects in developmental timing, but slugs exhibit a slightly elongated and twisted morphology (Supplementary Figure 7). Clear defects were, however, evident in chimeric development between wild type and *set1*$^{-}$ mutant cells (Figure 6A and Supplementary Figure 8). *set1*$^{-}$ mutant cells sorted strongly towards the posterior and collar regions in chimeric slugs (Figure 6A) and to the upper and lower cup of the fruiting body, which are normally occupied by prestalk (pstB and/or pstO) cell subtypes (Supplementary Figure 8). This sorting behaviour suggests H3K4 methylation is required for normal pre-spore cell differentiation, and that mutant cells instead differentiate as pstB and/or pstO cell types. Indeed, when RNA-seq was performed on *set1*$^{-}$ cells separated by FACS after chimeric development, *set1*$^{-}$ cells express pre-stalk genes more strongly, whereas pre-spore genes are poorly expressed (Figure 6B). Chimeric development of wild type and *set1*$^{-}$ strains transformed with prestalkO (ecmO-lacZ), prestalkB (ecmB-lacZ), or prespore (*pspA-RFP*) reporter constructs further confirm these findings. Both *ecmB*-lacZ and *ecmO*-lacZ reporter genes are more strongly expressed in *set1*$^{-}$ cells than wild type (Figure 6C), whilst the pspA-RFP marker gene is only very weakly expressed (Figure 6C). We next used RNA-seq transcriptome profiling as it provides a sensitive method to detect defects in developmental gene expression during clonal development. Genes normally expressed in pre-stalk cells were found to be overexpressed, whereas pre-spore genes are poorly expressed (Figure 6D). Furthermore, representative pre-stalk lacZ marker genes with a long half-life, are strikingly mis-expressed throughout the slug (Figure 6E).

The data from chimeric and clonal development suggest *set1*$^{-}$ mutant cells exhibit an intrinsic defect in pre-spore cell differentiation and tendency to differentiate as pstO and pstB cell types. To determine if this is due to a lack of H3K4 methylation, rather than some hitherto unknown function of Set1, a strain was generated in which another COMPASS complex component, Ash2, was knocked out. Again, this resulted in a severe reduction in the levels of H3K4 methylation (Figure 6F). A strain was also generated in which Gcn5, which is required for H3K9/14 acetylation, was disrupted (Figure 6G). The effects of H3K4 methylation on gene expression are thought to be, in part, due to its effect on targeting H3K9/14 residues for acetylation. Indeed, disruption of *set1* in *D. discoideum* has previously been shown to decrease H3K9/14 acetylation (Hsu et al., 2012). Both *ash2*$^{-}$ and *gcn5*$^{-}$ mutants phenocopied *set1*$^{-}$ knockout cells in chimeric development with wild type cells (Figure 6G). In addition, chimeric development of a hypomorphic Set1-FLAG knock-in allele (which results in reduced

di-methylation and absence of tri-methylation but little effect on mono-methylation (Supplementary Figure 9)) results in a qualitatively weaker sorting phenotype than that seen with *set1*$^{-}$ or *ash2*$^{-}$ mutants (Supplementary Figure 9). Together these results support the idea that H3K4 di and tri methylation is required for normal pre-spore cell differentiation.

Cell cycle position and gene expression variation interact to control cell type proportioning

Set1 disruption predominantly affects the expression of stochastically expressed developmental genes and results in defects in cell fate choice during development. This is consistent with predictions of the stochastic-deterministic model at the single cell level. A perturbation to the regulation of stochastic expression that increases the probability that genes will be expressed (i.e., increases $\bar{p}$) will necessarily lead to an increase in the average level of CCIF (since the mean $\mu = G\bar{p}\bar{k}$, and such a perturbation would increase $\bar{p}$), and thus will increase stalk cell differentiation. This, however, assumes that the regulation of CCAF and CCIF are independent and there are no effects on the cell cycle or the level of CCAF. To test this, we analysed the effects of perturbing cell cycle associated gene expression variation on stochastic variation, and vice versa. Cell cycle progression was blocked through cold shock, which results in widespread changes in cell cycle associated gene expression (Strasser et al., 2012). However, this did not affect the expression of representative Set1 dependent reporter genes that are not associated with the cell cycle (Figure 7A and B). Next, we perturbed the level of expression of variably expressed genes through Set1 disruption. scRNA-seq reveals *set1*$^{-}$ cells still exhibit cell cycle associated differential gene expression with the number of M/S and G2 phase cells unaffected (Figure 7C, Supplementary Figure 10). Furthermore, cell cycle progression and timing of cell division are unaffected by Set1 disruption (Figure 7D). These data thus suggest Set1 dependent variable genes and cell cycle dependent variation are independently controlled.

The stochastic deterministic model predicts an increase in CCIF should increase the likelihood of each single cell responding to stalk inducing signals (such as the stalk inducer DIF-1). To test this idea, DIF responsiveness was quantitatively compared at the single cell level in wild type and *set1*$^{-}$ mutant strains in which a GFP reporter gene had been knocked into the DIF sensitive *ecmA* or *ecmB* prestalk genes

(Supplementary Figure 11). The number of *ecmB* GFP positive cells was significantly higher in *set1*$^{-}$ mutant cells at all DIF-1 concentrations tested, and the number of *ecmA* positive cells was higher at the lowest concentrations (Figure 8A and B). However, the level of GFP expression in each responding cell did not significantly change. This suggests that the number of cells that are in a DIF responsive state is dependent on the level of Set1 dependent stochastic gene expression (CCIF), as well as the position of in the cell cycle (CCAF). We next analysed the relationship between cell cycle phase and cell fate choice in *set1*$^{-}$ mutant type cells. Cells were filmed growing at low density for 12-14 hours, which allowed each cell to be tracked for at least one cell division. The growth medium was then removed, and cell type differentiation induced. Cell tracking was continued, and the final fate of each cell (expression of cell type specific GFP reporter gene) was compared to the cell cycle position (timing of the last division) when differentiation was induced. Consistent with expectation from the stochastic-deterministic model (where expected stalk propensity, $G\bar{p}\bar{x}$,increases will burst frequency), *set1*$^{-}$ mutant cells exhibit a higher probability of adopting the pre-stalk cell fate being higher throughout the cell cycle in *set1*$^{-}$ mutant cells (Figure 8C). Finally, we tested whether the additive effects predicted by the stochastic-deterministic model are also seen during normal development. Developmental timing and fruiting body morphology is normal when cell cycle progression is perturbed in wild type cells (by growth at 11.5°C). Similarly, although changes in the level of expression of variable genes (through mutation of *set1*) affect proportioning, normal fruiting bodies are formed (Figure 8D). In contrast, when both inputs are simultaneously perturbed, the effects on development are dramatic. Few fruiting bodies formed, even 24 hours after development was complete in all other samples (Figure 8D). Together, these results are consistent with a model in which symmetry breaking, fate choice and cell type proportioning in *D. discoideum* depends on the integration of two independent variable inputs. Since variation is widespread in biological systems, there is immense scope for natural selection to harness these properties for robust biological decision making.

Discussion

The ability of isogenic cells to break symmetry and adopt different fates allows unicellular organisms to cope with dynamically changing environments and underpins the division of labour, which is a key feature of the evolution of multicellularity (West and Cooper, 2016). Cell state heterogeneity is an inevitable feature of biological

systems and can provide a reliable substrate for symmetry breaking. For example, stochastic intrinsic variation can be used to predictably 'sample' cells from different states if population sizes are sufficiently large (even if it is impossible to know what state each cell will be in). In other cases, cells may transition reproducibly between states, such as their position in the cell cycle. Moreover, if cell state reflects properties, such as energetic or resource state, it can be used to best determine the utility of cells for different roles.

Set1 dependent methylation controls transcriptional burst frequency

Stochastic cell-cell variation can arise from noisy gene expression, which is an inevitable consequence of transcriptional bursting. Changes in burst frequency will affect the level of transcriptional cell-cell variation, as well as mean expression levels. Bursting parameters and thus levels of transcriptional noise are affected by histone modifications. Roles in differentiation are suggested by studies in embryonic stem cells, as well as other systems suggest that gene networks associated with lineage choice are often associated with specific histone modifications (Hong et al., 2011; Yadav et al., 2018). This includes the presence of marks associated with gene activation (e.g. compass complex dependent H3K4 methylation and polycomb complex dependent H3K27 methylation). Single-cell RNA-seq analysis of mouse embryonic stem cells has also shown that treatment of cells with nucleoside analogues that are removed by the base excision repair pathway result in genome wide increases in cell-to-cell variability in transcript abundance of thousands of genes (Desai et al., 2021). Importantly, this increase in transcriptional noise does not markedly change expression mean. This also affects the likelihood of differentiation, presumably because it changes the probability of cells being in a responsive state to fate inducing signals. It is, however, unknown whether this reflects endogenous control mechanisms that regulate the levels of transcriptional noise.

Our results also support the idea that gene expression noise is important for cell fate. Our studies center on Set1 dependent H3K4 methylation, which is known to recruit various HATs, HDAC's and chromatin remodelers (e.g. ISWI, NURF, CHD) (Bian et al., 2011; Santos-Rosa et al., 2003; Shi et al., 2006; Shi et al., 2007; Sims et al., 2005; Taverna et al., 2006; Wysocka et al., 2006). Although Set1 dependent histone modifications are associated with more highly expressed loci (Barski et al., 2007; Soares et al., 2017), this relationship is not necessarily causal. In fact, genetic

studies have revealed that absence of H3K4 methylation only affects a small subset of genes, and these are typically upregulated (Hsu et al., 2012; Lorenz et al., 2014; Margaritis et al., 2012; Ramakrishnan et al., 2016). In addition, previous studies have also noted a correlation between the breadth of the H3K4 methylation domain and the level of gene expression variability (Benayoun et al., 2014; Rotem et al., 2015; Sze et al., 2020; Wu et al., 2017) . Indeed, disturbance of these patterns affects transcriptional consistency (Sze et al., 2020; Wu et al., 2017). Our studies provide further support for these ideas. Moreover, they support the idea that the modification of chromatin structure provides a means by which the level of gene expression and gene expression variation can be controlled to facilitate cell fate decisions.

Noisy gene expression has the potential to buffer development against environmental perturbations

Cell fate choice and symmetry breaking in *D. discoideum* depend on the interplay between deterministic effects of cell cycle position and stochastically expressed developmental genes on responsiveness to fate inducing signals fate. Why should such a mechanism evolve when a purely stochastic system could achieve correct proportioning simply by random sampling of a sufficiently large population of cells? Similarly, a cell cycle system alone could also achieve correct proportioning if cells are randomly distributed through the cell cycle. *D. discoideum* cells that have recently completed mitosis (i.e., those that have recently divided their resources) are biased towards the stalk fate, Use of the cell cycle position presumably reflects an adaptation that helps ensure aggregations to disproportionately sacrifice 'lower value' cells to the stalk for building the stalk and benefit higher value (higher resource state) cells that form spores. Resource based division of labour is also seen in *Sinorhizobium meliloti*, where cells with highest levels poly-3-hydroxybutyrate (PHB) (which allows survival in long term starvation) do not grow and instead leave resources to be utilised by low PHB containing cells (Ratcliff and Denison, 2010).

While such examples suggest cell state heterogeneity can be utilised adaptively to make cellular decisions, it is also vulnerable to systematic perturbations that affect the proportion of cells in different states. As soil dwellers, *D. discoideum* cell populations are likely to be exposed to varying environmental conditions, such as temperature, pH, or available nutrients that can all potentially affect developmental signalling, cell physiology, and cell type proportions. For example, *D. discoideum* cells

depleted in glucose or that have experienced cold shock will arrest around mitosis (and end up biased towards become stalk), which means that certain environments could lead to extreme stalk fate bias. Therefore, the evolutionary success of this organism means that mechanisms likely evolved to ensure development is buffered against environmental perturbations. Early studies identified a mechanism by which cell type proportioning could be corrected by feedback loops where the excess of one cell type leads to its trans-differentiation into the other (Belcher et al., 2022; Insall et al., 1992; Kay and Thompson, 2001). It has also been shown that disturbances that affect the cell cycle result in compensatory gene expression changes that alter the threshold of responsiveness (Chattwood et al., 2013). Our findings provide another potential solution to this problem, in which the integration of different sources of cell state variation allows systems to be robust against environmental perturbations, while still being able to adaptively exploit differences in cell state. This is because stochastic cell-cell variation, could buffer against catastrophic shifts in cell-type proportioning, by ensuring some cells always adopt different fates (Supplementary Figure 12). This would reduce the range over which feedback loops that refine proportioning need to operate. It is tempting to speculate that this complexity is a consequence of the strength of selection to ensure the correct ratio of stalk and spore cells under fluctuating natural environments that may affect signalling and cell fate choice (e.g. temp, pH, light, salt levels, humidity, food availability). Moreover, it highlights the impacts that cell-cell variation and environmental variation may have on the evolution of developmental signalling pathways. Since variation can promote or hinder developmental robustness, it is likely that many developmental systems will exhibit the goldilocks principle when it comes to adaptive exploitation of heterogeneity, where they have been evolutionarily tuned so that they are reliably and repeatedly exposed to 'just the right amount' of variation.

Figure Legends

Figure 1. Theoretical patterns of stalk proportioning through the cell cycle.
Plots illustrate the model structure, theoretical expectations for stalk proportioning (P_t) through the cell cycle, and corresponding fits to empirical data (from (Gruenheit et al., 2018)) for scenarios in which cell fate is controlled by either cell cycle position (time post mitosis), noisy gene expression, or a combination of the two. The five red points

in the illustrations of the model structure and corresponding theoretical expectations approximately match the five timepoints used in fitting the empirical data.

A. Stalk proportioning through the cell cycle is deterministic. i) In this scenario, sensitivity to stalk-inducing factors depends solely on the level of a cell-cycle associated factor (CCAF), which has a maximal level just after completing mitosis (given by C_0) and decays linearly through the cell cycle (at a rate given by β; see eqn. 1). **ii**) When the CCAF level is above a threshold value (R), cells adopt stalk fate and when the level drops below the threshold, they switch to spore fate. This results in a stepwise change in cell fate. **iii**) The best fit to the data shows a clear lack of fit, with the empirical pattern showing neither the predicted all or none pattern of cell fate nor a sharp shift in stalk propensity.

B. Stalk proportioning depends solely on noisy gene expression. i) In this scenario, sensitivity to stalk-inducing factors depends on a cell-cycle independent factor (CCIF), whose level is normally distributed (with mean μ and a standard deviation σ) because it is determined by the sum across many noisily expressed genes. All cells with a value of CCIF above a threshold value (R) adopt stalk fate, while those below the threshold adopt spore fate (see eqn. 2 for the expected stalk proportioning). **ii**) Because the level of CCIF does not depend on cell cycle, expected stalk proportioning does not change through time. **iii**) The best fit to the data shows a clear lack of fit, with the empirical pattern showing an obvious shift in stalk propensity through the cell cycle.

C. Stalk proportioning depends on the combination of the deterministic cell-cycle dependent change in CCAF and the stochastic input of CCIF from noisy gene expression. i) In this scenario (the stochastic-deterministic model), sensitivity to stalk-inducing factors is determined by the sum of the contribution of CCAF, which decays through the cell cycle, and CCIF, which shows a fixed pattern of variation at all timepoints. The proportion of cells at each timepoint that have a total sensitivity above the threshold value adopt stalk fate, while those below the threshold adopt the spore fate. **ii**) As CCAF decays through the cell cycle, the proportion of cells above the threshold declines, resulting in the non-linear pattern of stalk proportioning seen in the righthand plot. **iii**) The best fit to the data shows a very good match between cell fate behaviour and the pattern predicted by the model.

Figure 2. Identification of gene expression variation.

A. Gene expression variability in single cells. Plot shows $\log_{10}$ transformed relative gene expression variability (observed CV^2/expected CV^2) versus $\log_{10}$ transformed average gene expression level for all genes in 81 single cells. 2073 genes that show significantly greater variation (FDR < 0.05; red dashed line) than expected (black dashed line) were identified (yellow dots).

B. Variable genes can be divided into cell cycle associate and cell cycle independent genes. Variable genes were defined as cell cycle associated (1016 genes) or cell cycle independent (1057 genes). **i**) Hierarchical clustering reveals cell cycle associated genes show a strong cell cycle associated gene expression pattern (blue (S/M phase, 669 genes), orange (G2.1 phase, 247 genes) and red (G2.2 phase, 100 genes). **ii**) Cell cycle independent genes show little or no pattern with respect to the cell cycle groups. Cell order is the same in both plots.

C. Non-cell cycle associated genes do not drive cell groupings. i) Plotting PC1 against PC2 from a PCA of expression variation in cell cycle associated variable genes identifies three cell groups (blue, orange, red circles). **ii**) PCA of non-cell-cycle variable genes does not identify any grouping of cells.

D. Cell cycle dependent genes also exhibit extensive stochastic cell-cell variation. The CV^2 was calculated for all genes within each group of cells at different cell cycle stages and compared to the expected value based on average gene expression level. Genes with more variation than expected (adjusted p value <0.05, above dashed line) are shown in yellow. The 1016 cell cycle associated genes (black) are generally more variable than expected at each cell cycle stage.

Figure 3. Variably expressed genes are important for prestalk cell differentiation.

A. Developmental index of different groups of genes. Average gene expression during developmental stages was compared to average expression during growth + average expression during development. This developmental index ranges from zero to one. Zero is exclusive to growth and one is exclusive to development) (*** $p \leq 0.001$).

B. Cell type index of different groups of genes. Gene expression in prestalk cells was compared to expression in prestalk + prespore cells. This cell type index ranges from zero to one (zero is exclusive to prespore cells and one is exclusive to prestalk cells). (*** $p \leq 0.001$).

C. Schematic of screens carried out to identify mutants defective in prestalk cell differentiation. i) Selection for DIF-1 insensitivity. Cells were incubated with cAMP and DIF-1 to induce stalk cell differentiation. Only cells that fail to respond to these signals remain as amoebae and can re-enter growth upon addition of growth medium. gDNA was extracted for sequencing after two or six rounds of selection. **ii**) Selection for competence to respond to cAMP. Cells were treated with 8-Br-cAMP to induce terminal spore cell differentiation, before treatment with detergent to remove cells that failed to respond and remain as amoeba. A single round of this assay followed by sequencing allowed mutants that dropped out (failed to respond to 8-Br-cAMP) to be identified. **iii**) Selection for mutants with growth advantage. Cells were grown for a similar number of generations to that estimated from re-growth after stalk cell induction.

D. Identification of mutations that affect stalk cell differentiation. Heatmap of all 315 REMI mutant positions that were enriched in round 2 or 6 of the DIF-1 selection. 244 insertions were classified as DIF signalling mutants that affect stalk cell differentiation. Other mutants enriched in the screen were removed because they increased in frequency after only growth (63 growth mutants) or failed to differentiate in response to Br-cAMP (8 cAMP non-responsive mutants).

E. Validation of mutants identified by REMI-seq. Stalk cell inductions were carried out on independently generated mutants from round 2 or 6. 8/9 mutants produce fewer stalk cells than wild type (* $p \leq 0.05$,** $p \leq 0.01$,*** $p \leq 0.001$,**** $p \leq 0.0001$). Results are an average of at least three independent experiments.

F. REMI mutants that affect development are enriched for variable genes. i. Comparison of expected vs observed numbers of REMI mutants with variable, cell cycle or stochastic expression ii. Comparison of normalised CV^2 of genes associated

with REMI mutants that affect development vs the genome wide average (* $p \leq 0.05$,** $p \leq 0.01$,*** $p \leq 0.001$).

Figure 4. Variably expressed genes exhibit differential patterns of histone modifications.

A. The H3K4 methylation profile in *D. discoideum*.

Genes were assigned to ten bins according to their average expression level in vegetative AX4 cells (Supplementary data). Genes were identified that contain H3K4me1 or H3K4me3 peaks around the TSS (-2500bp to +2500bp) and the distribution of peaks for each bin was calculated. Both H3K4me1 (left) and H3K4me3 (right) exhibit characteristic distributions around the TSS, which correlate with expression level.

B. H3K4 methylation patterns differ in variable and non-variable genes.

The peak density of variable genes with at least one H3K4me1 or H3K4me3 peak between 2500bp upstream and 2500bp downstream of the TSS was plotted (blue line). To control for the effect of expression level on peak density, the distribution of expression levels of variable genes was used to select ten random samples of non-variable genes with the same expression distribution (Supplementary Fig 4). The peak density of the ten random samples was averaged (red line).

C. Number of genes with H3K4 methylation

Comparison of the numbers of observed and expected genes associated with H3K4Me1 and H3K4Me3 marks (*** $p \leq 0.001$).

Figure 5. Set1 controls cell-cell gene expression variation.

A. Set1 is required for H3K4 methylation. Western blot of mono, di and tri methylation at position H3K4 in wild type Ax4 and mutant *Set1*$^{-}$ cells. Disruption of *Set1* leads to the complete loss of all three methylation marks. Histone 3 levels are unaffected.

B. Deletion of *Set1* leads to specific changes in gene expression. Volcano plot showing differential gene expression between *Set1*- mutant and wild type cells. Up-regulated (green) and down-regulated (blue) genes are highlighted based on cut offs of two-fold change and an adjusted p-value of < 0.01.

C. Set1 dependent genes are enriched for variable genes. Quantification of enrichment of Set1 dependent genes in variable gene populations. As the threshold for the definition of a variable genes increases, the enrichment for *Set1* repressed genes also increases. *** binomial test, p<0.01.

D. Set1 dependent variable genes are enriched for cell cycle associated and cell cycle independent variable genes. Plot shows the number of genes mis-expressed in Set1 mutant cells compared to a random sampling.

E. Variable genes show increased expression in *Set1*$^-$ single cells. Expression level was calculated from 310 wild type and 310 *Set1*$^-$ cells. Variable genes (FDR < 0.05) were defined as those up- or down-regulated in the bulk *Set1* mutant sequencing. Upregulated genes show a statistically significant increase (p < 0.01) in expression. Downregulated genes show a small non-significant decrease in expression

F. Variable genes show decreased expression variability in *Set1*$^-$ single cells. CV^2 was calculated from 310 wild type and 310 *Set1*$^-$ cells. Variable genes (FDR < 0.05) were defined as those up- or down-regulated in the bulk *Set1* mutant sequencing. Up-regulated genes display a significant decrease in CV^2 in the Set1 mutant compared to WT cells ($p < 0.05$). Downregulated genes show a small non-significant decrease in variability.

G. *GtaU* and *HspF-2* show increased expression in *Set1*$^-$ single cells.
FACS quantification of *GtaU* and *HspF-2* reporter gene expression levels normalised to actin15-GFP.

H. *GtaU* and *HspF-2* show decreased cell-cell variability in *Set1*$^-$ single cells.
Cell-cell variability in normalised expression (CV^2) from FACS analysis of *GtaU* and *HspF-2* reporter genes.

Figure 6. Set1 is required for cell fate choice.

A. *set1*$^{-}$ mutant cells tend to occupy regions of the slug associated with prestalk cell differentiation in chimera with wild type. Schematic showing location of prestalk (pstA, pstO, pstB) and prespore (psp) cell types in the slug. Chimeric development of wild type and *set1*$^{-}$ mutant cells. 10% actin-RFP expressing cells were mixed and developed with 90% unlabelled cells. *Set1*$^{-}$ mutant cells sort to the collar and back of slugs when mixed with wild type cells. Scale bar = 0.5mm.

B. RNA-seq of FACS purified Set1 mutant and wild type cells. Actin15-GFP expressing wild type cells were mixed with actin15-RFP expressing *Set1*$^{-}$ mutant cells in a 50:50 ratio. GFP and RFP cells were separated at the slug stage and subjected to RNA-seq. The cell type index (prestalk expression/(prestalk + prespore expression)) was calculated for genes differentially expressed in wild type (blue bars) or Set1 mutant cells (yellow bars). >0.5 = prestalk, <0.5 = prespore. Interleaved histogram (light grey) shows the average distribution of 1000 random samples of the same numbers of genes. Scale bar = 0.5mm.

C. Chimeric development of wild type and set1 mutant cells expressing cell type specific markers. 10% labelled cells were mixed with 90% unlabelled cells. Set1 mutant cells strongly express pre-stalk markers, whereas a pre-spore marker is expressed weakly. Scale bar = 0.5mm.

D. RNA-seq reveals Set1 mutant slugs exhibit defects in cell type proportioning. Clonal wild type or *Set1* mutant cells were developed to the slug stage. The cell type index (>0.5 = prestalk, <0.5 = prespore) was calculated for genes identified as differentially expressed by RNA-seq in Ax4 (blue bars) and *Set1-* (yellow bars) cells. Interleaved histograms (light grey) show the average distribution of 1000 random samples of the same numbers of genes.

E. Set1 mutant cells overexpress prestalk marker genes. Clonal wild type or *Set1* mutant cells transformed expressing *ecmO*-lacZ- or *ecmB*-lacZ markers were developed to the slug stage. *ecmB* and *ecmO* are expressed in the collar and posterior anterior like cells in wild type. In *Set1* mutant slugs, pre-stalk markers are expressed throughout the length of the slug. Scale bar = 0.5mm.

F. Western blot of H3K4 methylation in Ash2 mutant cells. *ash2* mutant cells have lost H3K4 methylation. Levels of histone H3 are unaffected.

G. Western blot of H3K9/14 acetylation in Gcn5 mutant cells. *Gcn5* mutant cells have lost H3K9 and H3K14 acetylation. Levels of histone H3 are unaffected.

H. Chimeric development of Ash2 mutant cells with wild type cells. Like *set1*$^-$ cells, when labelled *ash2*$^-$ cells are mixed with WT cells they show a strong localisation in the collar and posterior of the slug. In the reciprocal mix, WT cells occupy the anterior pre-spore region. Scale bar = 0.5mm.

I. Chimeric development of Gcn5 mutant cells. Like *Set1*$^-$ and *Ash2*$^-$ when labelled *Gcn5*$^-$ cells are mixed with WT cells they are localised to the collar and back of the slug. Scale bar = 0.5mm.

Figure 7. Cell cycle and stochastic variation are independent.

A. Variation in expression of Set1 dependent stochastic genes is unaffected by cell cycle disruption. Dual reporter lines for actin15, gtaU and hspF-2 were incubated at either 22°C or 11.5°C and expression was quantified by FACS.

B. Level of expression of Set1 dependent stochastic genes is unaffected by cell cycle disruption. Dual reporter lines for actin15, gtaU and hspF-2 were incubated at either 22°C or 11.5°C and average level of expression calculated. Inset shows comparison of proportion of positive fluorescent cells (above background) at 11 and 22°C.

C. Quantification of numbers of cells in different cell cycle phases. Single cells were assigned to different cell cycle stages based on the expression of MS and G2 markers from scRNA-seq (see Supplementary figure 10 and methods).

D. Growth rate is unaffected by Set1 disruption. Cells were plated at low density and tracked by live imaging to determine cell cycle length of individual cells.

Figure 8. Cell type proportioning and responsiveness to differentiation cues depends on both stochastic and cell cycle variation

A and B. Set 1 regulates the threshold of responsiveness to differentiation cues. *ecmA*-GFP (A) and e*cmB*-GFP (B) knock-in lines were generated in wild type and *Set1*- mutant cells. Cells were induced to differentiate in culture in the presence of varying amounts of the prestalk inducer DIF-1. Marker gene expression was quantified by flow cytometry. Expressing cells were divided into bins of 0.05 (for log10 GFP expression).

C. Cell type differentiation at different cell cycle positions. Cells were grouped into hourly blocks according to the time of the last cell division before the addition of differentiation inducing signals. The percentage of cells that differentiate as prestalk cells was calculated. Results show the average of 8 movies.

D. Development is blocked when stochastic and cell cycle gene expression is altered simultaneously. WT and *Set1* mutant cells were incubated at 22°C, or at 11.5°C to inhibit cell cycle progression before being induced to develop.

Supplementary Figure Legends

Supplementary Figure 1

Fit of experimental data to models of fate choice

A. Comparison of the fit of empirical measures of stalk proportioning (P_t) through the cell cycle (from (Gruenheit et al., 2018)) to three different models for the control of proportioning. **i**) Expected stalk proportioning through the cell cycle predicted by a model where there is deterministic exponential decay in CCAF levels (after Gruenheit et al., 2018). The expected pattern of stalk proportioning for the best fit model is shown by the blue line (see Supplementary Information). **ii**) Expected stalk proportioning through the cell cycle predicted by a version of the stochastic-deterministic model for the case where gene expression noise is given by a gamma distribution (instead of being Gaussian) (see Supplementary Information). The gamma distribution has two additional parameters corresponding to the gamma distribution shape and scale

parameters k and θ. The best fit model corresponds to the gamma distribution shown in part B.ii of this figure, which is very similar to the shape of a Gaussian distribution (see part B.i of this figure). **Iii**) Expected stalk proportioning through the cell cycle for a version of the stochastic-deterministic model with an exponential decay in CCAF (see Supplementary Information). The best fit of this model corresponds to the pattern of decay in CCAF shown in part C.i of this figure.

B. The two figures show a comparison of **i**) Gaussian distributed noise and **ii**) the pattern of gamma distributed noise (see eqn. S1) corresponding to the best fit model shown in part A.iI of this figure.

C. The three figures show the pattern of decay in CCAF associated with the best fit models for **i**) a model with exponential decay (see eqn. S2), **ii**) a quadratic model of decay in CCAF, and **iii**) a cubic model of decay in CCAF. In each case, the best fit decay model is approximately linear and in all three cases, the model with linear decay presented in the main text has a better fit to the data.

Supplementary Figure 2

Cell cycle dependent genes defined through AUROC analysis.

A. Previous analyses of scRNAseq of 81 single cells reveals cells can be divided into two or three cell groups based on cell cycle position [31]. To identify markers associated with the cell cycle, relaxed AUROC (Area Under the Receiver Operating Characteristic Curve) thresholds of > 0.65 and $p < 0.05$ were used to compare gene expression in 25 M/S phase, and cells from different G2 phases (31 cells and 25 cells). Additional markers were identified by comparing MS cells to all G2 cells (31 + 25 = 56). A total of 5529 cell cycle markers were identified.

B. Intersection of cell cycle markers identified using two or three cell groups.

C. Intersection of variably expressed genes (2073 genes) with cell cycle markers (5529 genes) shows the cell cycle can account for variation in expression of 1016 of 2073 variable genes.

Supplementary Figure 3

Markers of different cell cycle phases also exhibit more variation within cells at the same cell cycle positions.

Columns represent the level of variability of all genes when calculated based on cells from M/S (25 cells), G2-1 (31 cells) and G2-2 (25 cells) phases. Genes that significantly

deviate from the trend are highlighted in yellow. Rows show the markers associated with each cell group (black dots) mapped onto these plots. Genes are enriched more variable at all cell cycle stages.

Supplementary Figure 4

Normalisation of expression levels of variable and non-variable genes for ChIP analyses

A. Distribution of expression levels of variable and non-variable genes. The average expression of variable genes and non-variable genes (limited to the lowest and highest expressed variable genes).

B. Sampling non-variable genes with the same distribution as variable genes. Ten samples of non-variable genes were taken. To ensure that the samples matched the expression distribution of variable genes, the probabilities of sampling a non-variable gene was weighted based on the percentage of variable genes in 30 different expression bins. Each sample contains 2024 genes, which is the same as the number of variable genes. The expression distribution of the samples matches the variable genes.

C and D. Variable genes and sampled non-variable genes containing H3K4me1 peaks show the same expression distribution. Variable genes and sampled non-variable genes were overlapped with lists of genes that contained H3K4me1 in their TSS adjacent regions. This was done for two ChIP replicates C and D). Sampled non-variable genes with H3K4me1 marks have the same expression distribution as variable genes.

E and F. Variable genes and sampled non-variable genes containing H3K4me3 peaks show the same expression distribution. Variable genes and sampled non-variable genes were overlapped with lists of genes that contained H3K4me3 in their TSS adjacent regions. This was done for two ChIP replicates E and F). Sampled non-variable genes with H3K4me3 marks have the same expression distribution as variable genes.

G and H. Variable genes tend to be enriched for genes with H3K4me1 marks and depleted for H3K4me3. The number of genes that contained at least one H3K4me1

(G) or H3K4me3 (H) peak within the region -2500bp upstream to 2500bp downstream of the TSS was calculated. This number was calculated for variable genes and compared to the average of ten samples of non-variable genes.

Supplementary Figure 5

Set1 dependent genes are more variable than expected.

Plot of gene expression variability (CV^2) against log10 expression in wild type cells. Different thresholds for the definition of variable genes are shown (FDR < 0.05 – red, FDR < 0.01 – green, FDR < 0.001 – pink, FDR < 0.0001 – purple). Set1 dependent genes (black dots) are enriched in variable populations.

Supplementary Figure 6

Dual reporter gene analysis of gene expression variation.

A. Schematic of dual reporter plasmid. Promoters of interest were cloned upstream of RFP in the pDM324 vector which also carries an actin15-promoter-GFP-actin8-terminator cassette.

B. scRNA-seq CV^2 plot illustrating the variability of genes used in dual reporter assay. *GtaU* (DM = 4.8, FDR = 0.001) *and HspF-2* (DM = 6.4, FDR = 5.2×10^{-5}) are highlighted. *Actin-15* is not variable (DM = 1.26, FDR = 0.71).

C. Single cell analysis of representative Set1 dependent reporter genes

FACS analysis of cells expressing RFP under the control of the actin-15 (i), gtaU (ii) and hspf2 (iii) promoters in wild type (blue) and Set1 mutant (yellow) cells.

Supplementary Figure 7

Disruption of set1 does not affect developmental timing but results in slightly aberrant slug morphology

Representative images of wild type and *set1*$^-$ mutant development.

Supplementary Figure 8

Set1$^-$ mutant cells sort to upper and lower cup of fruiting bodies

A. Schematic of cell type differentiation in the *D. discoideum* fruiting body

B. Sorting pattern of wild type and Set1 mutant cells in chimeric fruiting bodies
Set1 mutant cells sort to the upper cup, lower cup and stalk. Wild type cells sort to spore region of the fruiting body.

Supplementary Figure 9
A hypomorphic Set1 mutant exhibits a weaker developmental phenotype than knockout cells.

A. Western blot of H3K4 methylation. Blot of wild type and Set1-FLAG Knock-in cells for different H3K4me marks reveals that tagged cells exhibit a complete loss of H3K4me3 (like null mutant cells), but retain ~30% of H3K4me2 and accumulate increased mono-methylation compared to the WT.

B. Chimeric development of Set1-Flag knock-in cells with wild type cells. Mixing of 90% WT, FLAG-tagged Set1 or mutant *Set1*$^{-}$ cells with 10% WT cells labelled with constitutively expressed actin15-RFP. Wild type cells are largely absent from the prestalk region and are found throughout the prespore zone when mixed with *Set1*$^{-}$ mutant cells.

Supplementary Figure 10
The cell cycle is unaffected in Set1 mutant cells.

A. i. Dimensionality reduction of high-depth scRNA-seq data from 81 wild-type cells was performed. using genes previously defined as markers of M/S or G2 cells [31]. tSNE indicates cells can be assigned to two groups. The cells in each grouping match those defined using the 500 most variable genes in the genome (Gruenheit et al., 2018).Cells are colour-coded by the normalised ratio of the expression of all M/S genes to G2 genes (see methods). **ii** Histogram of cells from split into either M/S (blue) or G2 (pink) groups.

B. scRNA-seq was performed on 310 wild-type and *Set1*$^{-}$ single cells. The threshold ratio of M/S to G2 cells from wild type cells sequenced at high depth (**A, ii,** dotted line) was used as to define cells from wild type (i) or the Set1 mutant (ii) as either M/S or G2.

Supplementary Figure 11

Generation of ecmA and ecmB-GFP knock-in strains

A. Schematic of the construction of ecmA/B-GFP knock-in strains. A plasmid carrying GFP and a blasticidin resistance cassette flanked by loxP sites was integrated into the genome of wild type or Set1 mutant cells through homologous recombination at either the ecmA or ecmB locus. After selection the blasticidin cassette was removed by expression of Cre-recombinase.

B. GFP expression in knock-in strains. Gene replacement does not affect expression timing or positioning. ecmB-GFP expressing cells are found in the stalk tube, basal disc, lower and upper cup. ecmA-GFP expressing cells are found in the stalk tube, apical tip and lower cup.

Supplementary Figure 12

Stochastic variation can buffer against changes in deterministic fate cues

A. Stochastic variation smoothens the pattern of stalk proportioning through the cell cycle. The lines represent the expected stalk proportioning through the cell cycle for the stochastic-deterministic model (given by eqn. 3) with different levels of stochastic variability (σ, see embedded legend). When the level of stochastic variability (arising from noisy gene expression) is close to zero, we see the emergence of the steplike shift from stalk to spore (cf. Figure 1A.ii), and as the level of stochastic variability increases the system shows a more gradual shift in cell fate through the cell cycle. The lines were calculated using equation (3) with the other parameter values (i.e., other than σ) set to: $R = 0.5$, $C_0 = 0.75$, and $\beta = 1$.

B. Stochastic variation makes the system less sensitive to cell cycle biases. The lines represent the relative change in stalk proportioning resulting from a given level of cell cycle bias. The baseline level of stalk proportioning was calculated assuming uniform sampling of cells across the cell cycle. Biased sampling was achieved by changing the sampling of cells such that there was a linear increase or decrease in the sampling probability through the cell cycle (with the degree of sampling bias given by δ; see Supplementary Information). The bias values on the X-axis correspond to the proportion of cells sampled in the first or second half of the cell cycle as a result of

non-uniform sampling (e.g., a value of -¾ indicates that ¾ of cells would be in the first half of the cell cycle instead of expectation of ½ for unfirm sampling, meaning it represents an oversampling of ¼ in the first half of the cell cycle and an under-sampling of ¼ in the second half of the cycle). The relative change in stalk proportion was calculated by taking the difference between the expected stalk proportion (given the level of cell cycle bias) and the baseline proportioning (with uniform sampling across the cell cycle) and dividing by this baseline value. For example, a value 0.75 means that the cell cycle bias results in 75% more stalk cells than that expected without any bias.

Acknowledgements

This work was supported by a Wellcome Trust Investigator Award (WT095643AIA) to C.R.L.T; and grant from NERC (NE/V012002/1) to C.R.L.T and J.B.W.

Methods

Strains, culture and development

Dictyostelium discoideum strains were derived from parental *Ax3* or *Ax4* strains. Cells were cultured in 10cm tissue culture dishes containing HL5 media including glucose (Formedium) or on SM agar plates (Formedium) supplemented with *Klebsiella aerogenes*. For development, cells in tissue culture were diluted to 1 x 10^5 cells/ml and allowed to grow for 2-3 divisions before harvesting. Cells were induced to develop at a density of 5.1 x 10^5 cells/cm^2 on non-nutrient KK2 agar (16.1mM KH_2PO_4, 3.7mM K_2HPO_4) plates containing 1.5% purified agar (Oxoid). For chimeric development, cells were mixed prior to spreading on agar. *Set1* mutant lines were generated using a knock-out construct kindly provide by J. Chubb. *Ash2 and Gcn5* mutants were selected from the GWDI project (https://remi-seq.org) and verified by diagnostic PCR. A strain with a random intergenic insertion (GWDI_42_D_7) was utilised as a control. For cell cycle synchronisation by cold shock, exponentially growing cells were seeded at 1.5 x 10^6 cells/ml and incubated for 20 hours.

Reporter gene analyses

To measure cell-cell variation in gene expression of stochastic genes, promoter sequences 1kb upstream of the start codon were amplified and cloned into pDM324 (Veltman et al., 2009). The vector pDM327 was digested with NgoMIV to add sequences containing the *actin15*-promoter-GFP-*actin8*-terminator. *ecmA*-GFP or *ecmB*-KI GFP knock-in strains were generated by gene replacement (Chattwood et al., 2013). To quantify prestalk cell differentiation by FACS, growing cells were re-suspended at 1 x 10^5 cells/ml in 1.5ml of 1x stalk solution supplemented with 5mM cAMP and incubated in 35mm plastic bottom dishes for 9hrs at 22°C. Inductions were carried out for a further 9hrs at 22°C by replenishing the cAMP and adding different doses of DIF-1. For analysis, cells were washed and collected in 1ml of KK2 + 20mM EDTA. Fluorescence was measured by FACS (Nxt Flow Cytometer (Attune)). Relative proportion of RFP and GFP positive was assessed using FlowJo software. LacZ staining of developed structures was carried out using strains expressing β-galatosidase under promoter control of fate specific markers ecmO and ecmB. Strains were fixed for 10 minutes in 4% formaldehyde and Z-buffer (10mM KCl, 60mM Na_2HPO_4, 40mM NaH_2PO_4, 1mM $MgSO_4$). Structures were washed twice in Z-buffer before being permeabilised for 20 minutes in 0.1% NP40. Another two washes were

carried out before adding staining solution (5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6$, 1mg/ml X-gal, 1mM EGTA).

Western blotting

Growing cells were lysed in nuclei buffer (40mM Tris-pH7.8, 1.5% Sucrose, 0.1mM EDTA, 6mM $MgCl_2$, 40mM KCl, 5mM DTT, 0.4% NP40). Pellets were re-suspended at a density of 1 x 10^7 nuclei/ml in 1x PBS including 1x protease inhibitors (Promega #G6521) and 1X SDS-PAGE buffer. Western blots were probed with antibodies to Histone-3 (ab1791 - polyclonal-rabbit IgG), H3K4-me1 (ab8895 - polyclonal-rabbit IgG), -me2 (Millipore #07-030 - polyclonal-rabbit IgG), -me3 (ab8580 - polyclonal-rabbit IgG), β-actin (Santa-Cruz Biotechnology – monoclonal- mouse IgG). $Gcn5^-$ blots were carried out as previously described (Huang et al., 2021). Blots were imaged using a Chemidoc MP imaging system (Biorad) and Image Lab 5.2.1 software (Biorad).

Single cell analysis of cell cycle and cell fate

To follow differentiating cells, cells were transformed with the *pspA*-GFP cell type specific reporter gene. Log phase cells were resuspended in HL5 glucose media and 7.5 x 10^3 cells were plated in 750μl deposited on a 35mm glass bottom dish (WPi). Cells were grown for a minimum of 16 hours during which time cells were imaged every 5 minutes. The time between cell divisions was determined by manually tracking single cells. After 16 hours, HL5 media was removed and cells were washed five times in KK2 buffer to remove residual growth medium. To induce differentiation, 750μl of conditioned media supplemented with 10nM DIF-1 was added. Conditioned media was collected from cells developed at a high density (1 x 10^6 cells/ml) in 1x stalk solution (10mM MES pH6.2, 1mM $CaCl_2$, 2mM NaCl, 10mM KCl, 200mg/ml streptomycin sulphate) supplemented with 5mM cAMP for 16 hours. Conditioned media was collected and briefly centrifuged before DIF-1 was added to washed cells. Images were taken at 5 minute intervals for another 20 hours. A final fluorescence image was taken to determine the number of prestalk and prespore cells. Cells were tracked by hand to determine the timing of the last cell division prior to the induction of cell differentiation.

REMI-seq screen for DIF-insensitive mutants

A REMI mutant library consisting of >10,000 mutants (Gruenheit et al., 2021) was grown to log phase . Cells were plated in triplicate at 2x10^5 cells/ml in stalk medium in 10cm diameter tissue culture dishes with 5mM cAMP for 24h. Cells were then washed

twice with KK2 before 24h incubation with 10nM DIF. Stalk medium was then removed and replaced with growth medium (HL5) and cells were allowed to grow until reaching confluency. Genomic DNA was prepared from the mutant library following 2 and 6 rounds of this selection and processed for sequencing. To control for mutants that are unable to differentiate in response to cAMP, an 8-Br-cAMP monolayer assay screen was performed. Cells from the mutant library were seeded in triplicate at $2x10^5$ cells/ml in stalk medium supplemented with 10mM 8-Br-cAMP. After 48 hours, detergent was added (0.1% NP40, 10mM EDTA) to remove cells that had not formed spores. Stalk medium was then removed and replaced with growth medium (HL5) and the cells were grown until reaching confluency. Genomic DNA was prepared from the mutant library following 1 round of this selection and processed for sequencing (Gruenheit et al., 2021). Analysis of mutant pools was carried out as previously reported (Gruenheit et al., 2021) using Z-score thresholds of > 1.5 for enriched mutants and < -1 for depleted mutants. Mutants (Supplementary data file 1) were identified as DIF-insensitive if they were enriched in the cAMP removal screen, not down-regulated in 8-Br-cAMP screen and not previously identified as having a growth phenotype (Gruenheit et al., 2021).

Sample preparation and analysis of bulk RNA-seq data

RNA was extracted from log phase growing cells or cells developed for 16.5 hours on 1.5% non-nutrient agar at 22°C. For RNA sequencing of chimeric development, wild type cells expressing *actin15*-GFP were mixed in a 1:1 ratio with *set1*$^-$ cells expressing *actin15*-RFP and developed for 16.5 hours. Chimeric slugs were disaggregated by passing through a 25G needle before sorting (BD FACSaria) into GFP and RFP positive populations. RNA was extracted and RNA integrity number (RIN) of samples was determined by Tapestation (Agilent). Libraries were prepared from samples deemed with a RIN > 8 using an Illumina TruSeq kit and sequencing was undertaken on a Hiseq-4000 (illumina) using 100bp pair-end chemistry. Sequences were trimmed of Truseq adapters and quality controlled (Trimmomatic) by discarding reads shorter than 20bp or those where the average quality score dropped below an average of 15 in a sliding 4 base pair window. Leading and trailing bases of reads below a phred score of 30 were also removed from tags. Reads were aligned (Bowtie2) to the *D. discoideum* genome (an inverted repeat on chromosome 2 was masked), bamfiles were sorted (Samtools) and reads counted using the RPKM_count.py script (RSeQC). DESeq2 v1.26.0 was used for differential expression analyses. Thresholds for differential expression between samples were set at a *p.adj* < 0.01 with a fold change

of > 2 between samples. To calculate differences in cell type specific gene expression, prestalk and prespore RNA-seq data was downloaded from the SRA (PRJNA543665) (de Oliveira et al., 2019). Genes with less than 10 read counts were removed. A cell type index was calculated for the remaining 5319 genes (Cell type index = expression count in prestalk cells / (expression count in prestalk cells + expression count in prespore cells).

ChIP-seq analysis

Bulk RNA-seq data from vegetative cells (this study) was used to rank genes based on their level of expression. Genes with detectable expression (ie > 0 normalised reads) were divided into ten equally sized bins. ChIP-seq data for two H3K4me1 and H3K4me3 replicates was downloaded from the GEO database (accession #GSE137604 Sub-Series GSE137599) as narrowPeak files [39]. Promoter regions around the TSS for each gene in each bin were identified (-2500bp to + 2500bp up/down-stream of the TSS) and annotated. Using functions from the chIPSeeker package and custom scripts (https://github.com/WilliamSalvidge/dictyChipSeq) annotated regions were intersected H3K4-me1 or -me3 peaks as defined by narrowPeak files. Overlaps were averaged for each expression bin and plotted using the *plotAvgProf* function from the chIPSeeker R package. This accounts for differing numbers of peaks in different expression bins and allows patterns of peak density between expression bins to be compared. To compare peak distribution in variable and non-variable genes, an equal number (2024) of genes was sampled using a weighted probability based on the expression of variable genes.

Identification of variable genes using single cell RNA-sequencing

Data for 81 single wild type cells isolated using the Fluidigm C1 platform were downloaded from the SRA (SAMN07833758 - SAMN0783383). Reads were normalised (DESeq2 v1.26.0) and the coefficient of variation squared (CV^2) was calculated and plotted against mean expression. A trend line was fitted to the data using non-linear least squares regression (Scran v1.15.9). Genes were defined as variable (2073 genes) based on a one-sided test assuming a normal distribution around the trend but one where deviation changed depending on the mean expression of a given gene (Scran v1.15.9 - modelGeneCV2) with a FDR of < 0.05. To identify genes with a cell cycle signature, single cells were clustered using M3Drop v1.12.0 as previously reported (Gruenheit et al., 2018). Cells could be organised into three

clusters of 25, 31 and 25 cells. Marker genes for each cluster were identified using logistic regression (M3Drop v1.12.0). A relaxed AUROC threshold (> 0.65) ensured that all genes possessing weak cell cycle signature could be identified (5529 genes). This allowed genes variable genes to be identified that where variation is dependent on cell cycle position (1016 genes) and independent of cell cycle position (1057 genes).

Single cell sequencing of wild type and *set1*$^{-}$ cells

Actin15-GFP expressing wild type or *set1*$^{-}$ cells were grown on tissue cultures dishes and harvested during log-growth. Cells were washed in 1x PBS and incubated with DAPI (2.5μg/ml). Cells were resuspended at a density of 2.8 x 10^4 cells/ml and dispensed into a SMARTer ICELL8 3' DE Chip using the ICELL8 cx Single Cell System. Wells of the 3' DE Chip contain pre-printed oligonucleotides possessing well specific barcodes and UMI connected to a polydT region for hybridisation with polyadenylated transcripts. For each well of the chip, 50nl of stained cell solution was aliquoted to maximise the number of wells containing a single cell. Cells were imaged directly dispensing into nano-wells using DAPI and FITC filters and the 3' DE Chip was frozen at -80°C. Images taken were analysed to identify wells containing individual cells based on DAPI and GFP fluorescence. In total 799 wells were identified that contained single cells (399 wild type and 400 *set1*$^{-}$). The 3' DE Chip was then thawed to lyse cells and loaded onto the ICELL8 cx Single Cell System where components for reverse transcription (RT) and cDNA amplification were dispensed into chosen wells. After RT-PCR, products from separate wells were collected into a single sample, concentrated and purified according to the manufacturer's instructions. Samples were prepared for sequencing using a Nextera XT Library Prep Kit (illumina) and sequenced on a NextSeq 500 (illumina) system utilising one flow cell and a NextSeq High-Output kit (2 x 75bp reads). One read was used to sequence the well barcode and transcript UMI, with the second reading the 3' end of the transcript itself. This yielded >500 million reads. FASTQ files were demultiplexed (mappa), reads were then quality controlled (reads shorter than 15bp discarded, a minimum of 30% N's allowed, phred score of 20) and trimmed of adapters (cutadapt). Reads were aligned (STAR) to the latest version of the *D. discoideum* genome (v2.7), sorted (Samtools) and tags counted (UMI-tools). Cells were quality controlled (Scater v1.14.6) and cells over 2 median associated deviations (MADs) from the median for library size, total number of features

or mitochondrial reads excluded as outliers. This left 310 wild type and 310 *set1*$^-$ cells. Genes with < 5 reads were also removed from further analyses.

Cell cycle analysis

Genes previously defined by scRNA-seq as markers of M/S and G2 cells from 81 Fluidigm-sorted cells (Gruenheit et al., 2018) were used to compare cell cycle patterns in iCell8-sorted Ax4 and *Set1*$^-$ cells. The average expression of all 876 M/S genes and 642 G2 genes with expression in these samples (>0 reads) was determined. The normalised ratio of M/S to G2 marker gene expression was used to define the cell cycle position in each cell.

Model fitting to measurements of stalk cell induction

To evaluate the fit of our stochastic-deterministic model to data, we used data on stalk fate measured at different times in the progression through the cell cycle in population of cells aligned in the cell cycle (Gruenheit et al., 2018). Stalk fate propensity was measured in two genetically different sets of cells (wildtype AX3 and AX3 with a knockout of the gene *gefE*) grown under two conditions ('normal', G+, and low glucose, G–, conditions), which alter the stalk propensity of cells. We used data from the first six timepoints, corresponding to zero to five hours after the last division, and combined the four sets of cells by adjusting the values in set such that their mean propensity matched the overall global mean propensity (and hence there is no difference in average propensity of the four sets; see Supplementary Data for the raw and adjusted values). After combining the four sets, one outlier was identified (*gefE*$^-$ under G– at hour 5), which was consistent with those cells passing a checkpoint where they re-enter mitosis, and was removed, after which data were re-normalised as described above (see Supplementary Information for a comparison of the model fitting with the outlier included). To confirm that the different sets of cells behave similarly, we also fitted the model separately to each class and see no evidence of heterogeneity of model estimates. The stochastic-deterministic model was fitted to these data using the 'NonlinearModelFit' function in Wolfram Mathematica version 14, which uses the Levenberg–Marquardt algorithm for least-squares curve fitting. This model fitting yielded estimates of C_0^* and β, the Akaike Information Criterion (AIC) and of the error and total sums of squares, which were used to calculate the *R*-squared. The details of the alternative models that were fitted and the methods used for comparing the fit of

different models to the stalk propensity data are provided in the Supplementary Information.

STAR Methods

Software

Trimmomatic - http://www.usadellab.org/cms/index.php?page=trimmomatic
Bowtie2 - http://bowtie-bio.sourceforge.net/index.shtml
Samtools - http://samtools.sourceforge.net
RSeQC - http://rseqc.sourceforge.net
Takara mappa - https://takarabiousa.github.io/mappa_userguide.html
Cutadapt - https://github.com/marcelm/cutadapt
STAR aligner manual -
https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf
UMI tools manual - https://github.com/CGATOxford/UMI-tools
***D. discoideum* genome** -
https://protists.ensembl.org/Dictyostelium_discoideum/Info/Index
Mathematica 14.0 - https://www.wolfram.com/mathematica/

R - packages

DESeq2 - v1.26.0 -
https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Scater - v1.14.6 - https://bioconductor.org/packages/release/bioc/html/scater.html
Scran - v1.15.9 - https://bioconductor.org/packages/release/bioc/html/scran.html
M3Drop - v1.12.0 - https://bioconductor.org/packages/release/bioc/html/M3Drop.html
SC3 - v1.14.0 - https://www.bioconductor.org/packages/release/bioc/html/SC3.html
Superheat - v.0.1.0 - https://rlbarter.github.io/superheat/

References

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823-837.
Belcher, L.J., Madgwick, P.G., Kuwana, S., Stewart, B., Thompson, C.R.L., and Wolf, J.B. (2022). Developmental constraints enforce altruism and avert the tragedy of the commons in a social microbe. Proc Natl Acad Sci U S A *119*, e2111233119.
Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., Mancini, E.*, et al.* (2014). H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. Cell *158*, 673-688.
Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.
Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O.*, et al.* (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. The EMBO journal *30*, 2829-2842.
Buttery, N.J., Rozen, D.E., Wolf, J.B., and Thompson, C.R.L. (2009). Quantification of Social Behavior in D. discoideum Reveals Complex Fixed and Facultative Strategies. Current Biology *19*, 1373-1377.
Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544-547.
Chattwood, A., Nagayama, K., Bolourani, P., Harkin, L., Kamjoo, M., Weeks, G., and Thompson, C.R. (2013). Developmental lineage priming in Dictyostelium by heterogeneous Ras activation. eLife *2*, e01067-e01067.
Chattwood, A., and Thompson, C.R.L. (2011). Non-genetic heterogeneity and cell fate choice in Dictyostelium discoideum. Development, Growth & Differentiation *53*, 558-566.
Chubb, J.R., Bloomfield, G., Xu, Q., Kaller, M., Ivens, A., Skelton, J., Turner, B.M., Nellen, W., Shaulsky, G., Kay, R.R.*, et al.* (2006a). Developmental timing in Dictyostelium is regulated by the Set1 histone methyltransferase. Developmental biology *292*, 519-532.
Chubb, J.R., Bloomfield, G., Xu, Q., Kaller, M., Ivens, A., Skelton, J., Turner, B.M., Nellen, W., Shaulsky, G., Kay, R.R.*, et al.* (2006b). Developmental timing in

Dictyostelium is regulated by the Set1 histone methyltransferase. Dev Biol *292*, 519-532.
Cornell, J.R., and Benjamin, C.A. (1970). Probability, Statistics, and Decisions for Civil Engineers. (NY: McGraw-Hill).
de Oliveira, J.L., Morales, A.C., Stewart, B., Gruenheit, N., Engelmoer, J., Brown, S.B., de Brito, R.A., Hurst, L.D., Urrutia, A.O., Thompson, C.R.L.*, et al.* (2019). Conditional expression explains molecular evolution of social genes in a microbe. Nat Commun *10*, 3284.
Desai, R.V., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H.*, et al.* (2021). A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. Science *373*.
Gomer, R.H., and Ammann, R.R. (1996). A cell-cycle phase-associated cell-type choice mechanism monitors the cell cycle rather than using an independent timer. Dev Biol *174*, 82-91.
Gomer, R.H., and Firtel, R.A. (1987). Cell-autonomous determination of cell-type choice in Dictyostelium development by cell-cycle phase. Science *237*, 758-762.
Gruenheit, N., Baldwin, A., Stewart, B., Jaques, S., Keller, T., Parkinson, K., Salvidge, W., Baines, R., Brimson, C., Wolf, J.B.*, et al.* (2021). Mutant resources for functional genomics in Dictyostelium discoideum using REMI-seq technology. BMC Biology *19*, 1-19.
Gruenheit, N., Parkinson, K., Brimson, C.A., Kuwana, S., Johnson, E.J., Nagayama, K., Llewellyn, J., Salvidge, W.M., Stewart, B., Keller, T.*, et al.* (2018). Cell Cycle Heterogeneity Can Generate Robust Cell Type Proportioning. Developmental Cell *47*, 494–508.e494-494–508.e494.
Hong, S.H., Rampalli, S., Lee, J.B., McNicol, J., Collins, T., Draper, J.S., and Bhatia, M. (2011). Cell fate potential of human pluripotent stem cells is encoded by histone modifications. Cell Stem Cell *9*, 24-36.
Hsu, D.-W., Chubb, J.R., Muramoto, T., Pears, C.J., and Mahadevan, L.C. (2012). Dynamic acetylation of lysine-4-trimethylated histone H3 and H3 variant biology in a simple multicellular eukaryote. Nucleic acids research *40*, 7247-7256.
Huang, L.-Y., Hsu, D.-W., and Pears, C.J. (2021). Methylation-directed acetylation of histone H3 regulates developmental sensitivity to histone deacetylase inhibition. Nucleic Acids Research *49*, 3781-3795.
Huh, D., and Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. Nature Genetics *43*, 95-100.

Insall, R., Nayler, O., and Kay, R.R. (1992). DIF-1 induces its own breakdown in Dictyostelium. The EMBO Journal *11*, 2849-2854.
Kaller, M., Nellen, W., and Chubb, J.R. (2006). Epigenetics in Dictyostelium. Methods Mol Biol *346*, 491-505.
Kay, R.R., and Thompson, C.R.L. (2001). Cross-induction of cell types in Dictyostelium: evidence that DIF-1 is made by prespore cells. Development *128*, 4959-4966.
Lorenz, D.R., Meyer, L.F., Grady, P.J.R., Meyer, M.M., and Cam, H.P. (2014). Heterochromatin assembly and transcriptome repression by Set1 in coordination with a class II histone deacetylase. eLife *3*.
Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in Gene Expression Determines Cell Fate in Bacillus subtilis. Science *317*, 526-529.
Madgwick, P.G., Stewart, B., Belcher, L.J., Thompson, C.R.L., and Wolf, J.B. (2018). Strategic investment explains patterns of cooperation and cheating in a microbe. Proceedings of the National Academy of Sciences of the United States of America *115*, E4823–E4832-E4823–E4832.
Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., van Hooff, S., van Leenen, D., Dargemont, C., Géli, V.*, et al.* (2012). Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3′-End Antisense Transcription. PLoS Genetics *8*, e1002952-e1002952.
Pauklin, S., and Vallier, L. (2013). The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. Cell *155*, 135-147.
Peccoud, J., and Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. Theoretical Population Biology *48*, 222-234.
Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biology *4*, e309-e309.
Ramakrishnan, S., Pokhrel, S., Palani, S., Pflueger, C., Parnell, T.J., Cairns, B.R., Bhaskara, S., and Chandrasekharan, M.B. (2016). Counteracting H3K4 methylation modulators Set1 and Jhd2 co-regulate chromatin dynamics and gene transcription. Nature Communications *7*.
Ratcliff, W.C., and Denison, R.F. (2010). Individual-level bet hedging in the bacterium Sinorhizobium meliloti. Curr Biol *20*, 1740-1744.
Rodrigues, A.M.M., and Gardner, A. (2022). Reproductive value and the evolution of altruism. Trends in Ecology & Evolution *37*, 346-358.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nature biotechnology *33*, 1165-1172.
Saiz, N., Williams, K.M., Seshan, V.E., and Hadjantonakis, A.-K. (2016). Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. Nature Communications *7*, 13463-13463.
Santos-Rosa, H., Schneider, R., Bernstein, B.E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2003). Methylation of Histone H3 K4 Mediates Association of the Isw1p ATPase with Chromatin. Molecular Cell *12*, 1325-1332.
Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R.*, et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96-99.
Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G.E., Fingerman, I.M., Briggs, S.D.*, et al.* (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. Journal of Biological Chemistry *282*, 2450-2455.
Sims, R.J., Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. Journal of Biological Chemistry *280*, 41789-41792.
Slack, J. (2014). Establishment of spatial pattern. Wiley Interdiscip Rev Dev Biol *3*, 379-388.
Soares, L.M., He, P.C., Chun, Y., Suh, H., Kim, T.S., and Buratowski, S. (2017). Determinants of Histone H3K4 Methylation Patterns. Molecular Cell *68*, 773–785.e776-773–785.e776.
Stevense, M., Muramoto, T., Muller, I., and Chubb, J.R. (2010). Digital nature of the immediate-early transcriptional response. Development *137*, 579-584.
Strasser, K., Bloomfield, G., MacWilliams, A., Ceccarelli, A., MacWilliams, H., and Tsang, A. (2012). A Retinoblastoma Orthologue Is a Major Regulator of S-Phase, Mitotic, and Developmental Gene Expression in Dictyostelium. PLoS ONE *7*, e39914-e39914.

Süel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B. (2007). Tunability and noise dependence in differentiation dynamics. Science (New York, NY) *315*, 1716-1719.
Sze, C.C., Ozark, P.A., Cao, K., Ugarenko, M., Das, S., Wang, L., Marshall, S.A., Rendleman, E.J., Ryan, C.A., Zha, D.*, et al.* (2020). Coordinated regulation of cellular identity–associated H3K4me3 breadth by the COMPASS family. Science Advances *6*, eaaz4764-eaaz4764.
Taverna, S.D., Ilin, S., Rogers, R.S., Tanny, J.C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L.P., Chait, Brian T.*, et al.* (2006). Yng1 PHD Finger Binding to H3 Trimethylated at K4 Promotes NuA3 HAT Activity at K14 of H3 and Transcription at a Subset of Targeted ORFs. Molecular Cell *24*, 785-796.
Thompson, C.R., and Kay, R.R. (2000). Cell-fate choice in Dictyostelium: intrinsic biases modulate sensitivity to DIF signaling. Dev Biol *227*, 56-64.
Ulfig, A., and Jakob, U. (2024). Redox heterogeneity in mouse embryonic stem cells individualizes cell fate decisions. Dev Cell *59*, 2118-2133 e2118.
Veltman, D.M., Akar, G., Bosgraaf, L., and Van Haastert, P.J.M. (2009). A new set of small, extrachromosomal expression vectors for Dictyostelium discoideum. Plasmid *61*, 110-118.
Wang, S.Y., Pollina, E.A., Wang, I.H., Pino, L.K., Bushnell, H.L., Takashima, K., Fritsche, C., Sabin, G., Garcia, B.A., Greer, P.L.*, et al.* (2021). Role of epigenetics in unicellular to multicellular transition in Dictyostelium. Genome Biology 2021 22:1 *22*, 1-30.
Weinberger, L., Voichek, Y., Tirosh, I., Hornung, G., Amit, I., and Barkai, N. (2012). Expression noise and acetylation profiles distinguish HDAC functions. Molecular cell *47*, 193-202.
West, S.A., and Cooper, G.A. (2016). Division of labour in microorganisms: An evolutionary perspective. Nature Reviews Microbiology *14*, 716-723.
Wolf, J.B., Howie, J.A., Parkinson, K., Gruenheit, N., Melo, D., Rozen, D., and Thompson, C.R.L. (2015). Fitness trade-offs result in the illusion of social success. Current Biology *25*, 1086-1090.
Wu, S., Li, K., Li, Y., Zhao, T., Li, T., Yang, Y.-F., and Qian, W. (2017). Independent regulation of gene expression level and noise by histone modifications. PLOS Computational Biology *13*, e1005585-e1005585.
Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF

couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.
Yadav, T., Quivy, J.P., and Almouzni, G. (2018). Chromatin plasticity: A versatile landscape that underlies cell fate and identity. Science *361*, 1332-1336.
Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development (Cambridge, England) *137*, 715-724.
Zahavi, A., Harris, K.D., and Nanjundiah, V. (2018). An individual-level selection model for the apparent altruism exhibited by cellular slime moulds. J Biosci *43*, 49-58.

Figure 1

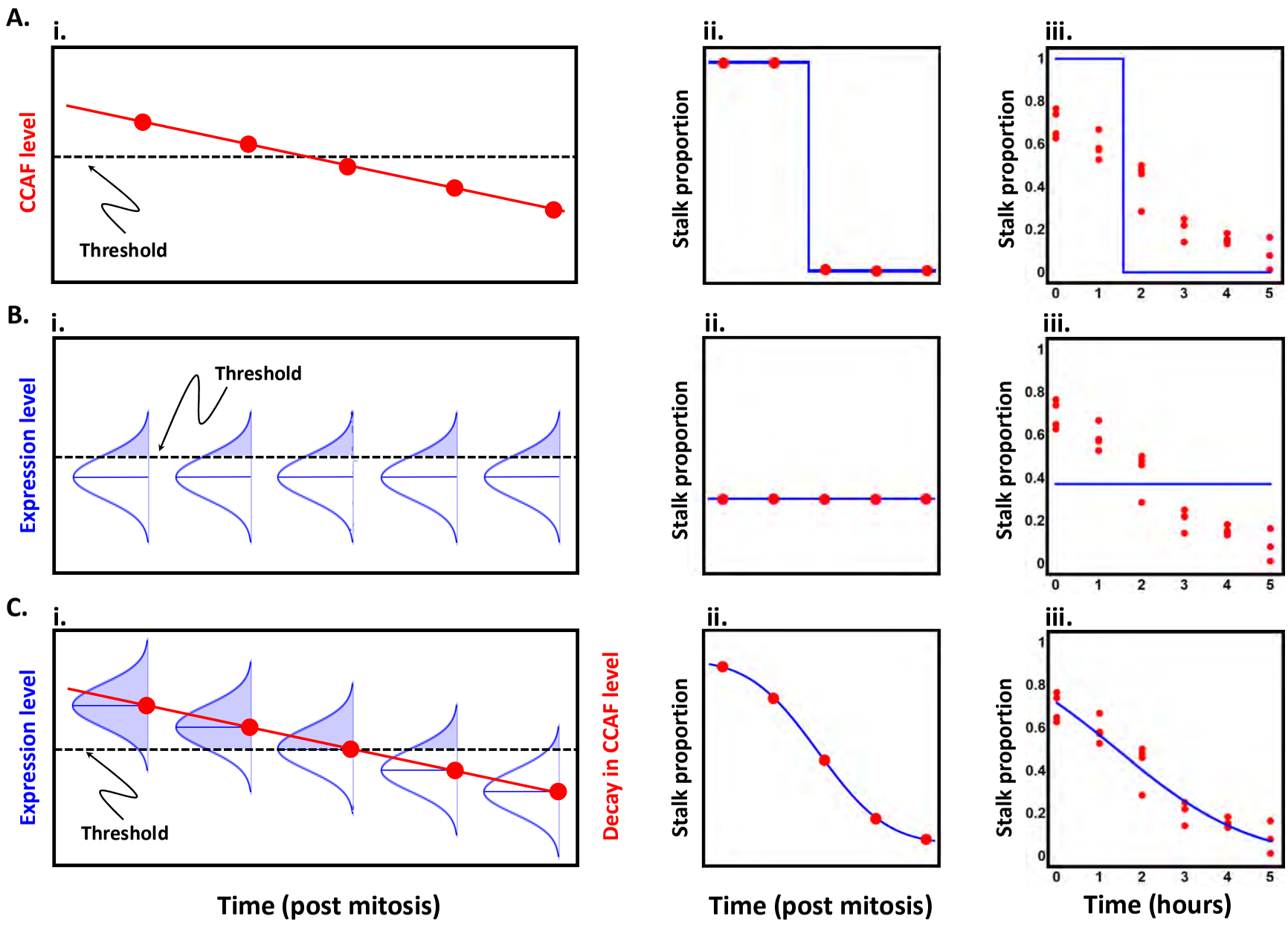

A.
i.
CCAF level
Threshold
ii.
Stalk proportion
iii.
Stalk proportion
B.
i.
Expression level
Threshold
ii.
Stalk proportion
iii.
Stalk proportion
C.
i.
Expression level
Threshold
Decay in CCAF level
ii.
Stalk proportion
iii.
Stalk proportion
Time (post mitosis)
Time (post mitosis)
Time (hours)

Figure 2

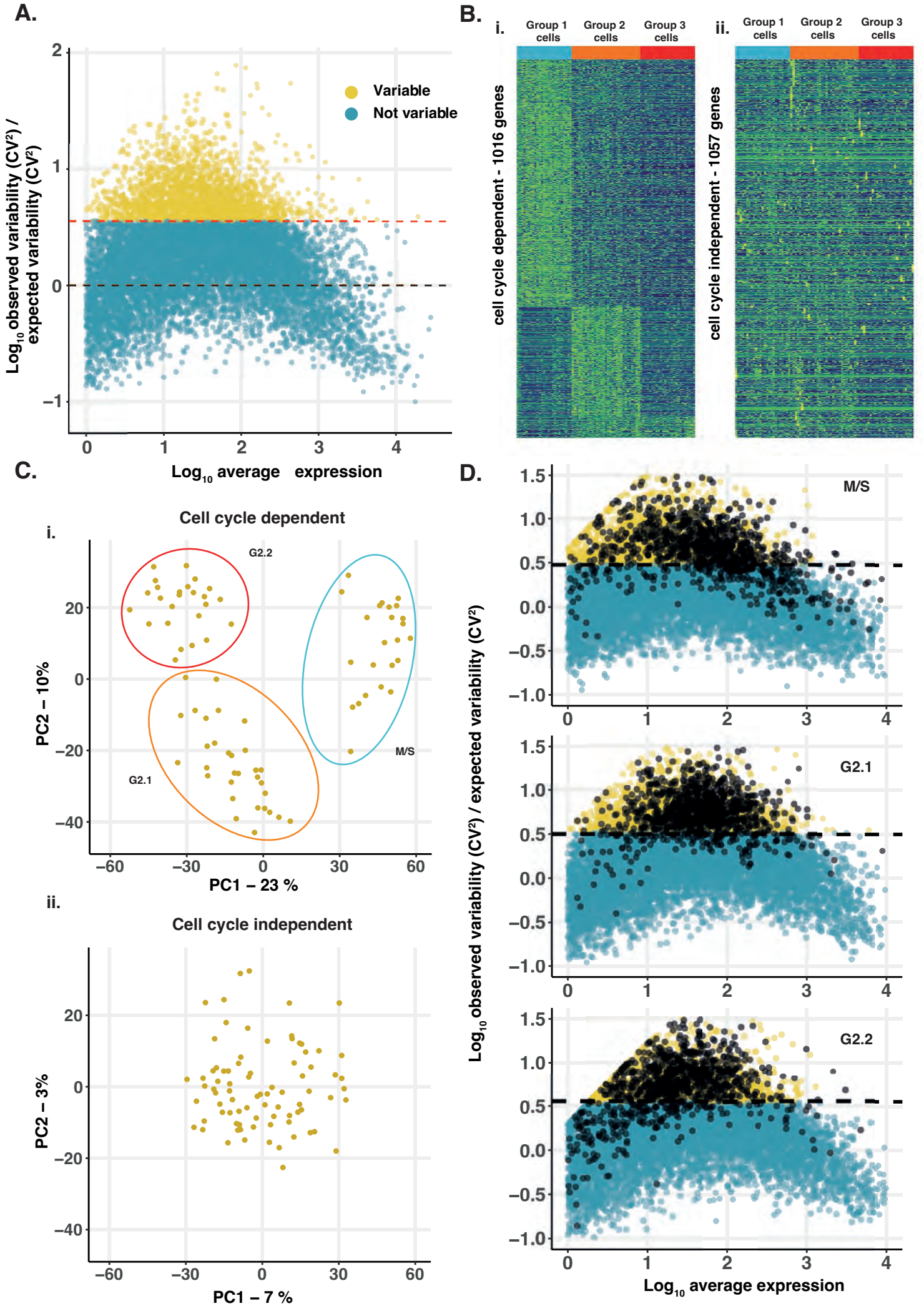

A.
Variable
Not variable
Log10 observed variability (CV²) / expected variability (CV²)
Log10 average expression
B.
i.
ii.
Group 1 cells
Group 2 cells
Group 3 cells
cell cycle dependent - 1016 genes
cell cycle independent - 1057 genes
C.
i.
Cell cycle dependent
G2.2
G2.1
M/S
PC2 – 10%
PC1 – 23 %
ii.
Cell cycle independent
PC2 – 3%
PC1 – 7 %
D.
M/S
G2.1
G2.2
Log10 observed variability (CV²) / expected variability (CV²)
Log10 average expression

Figure 3

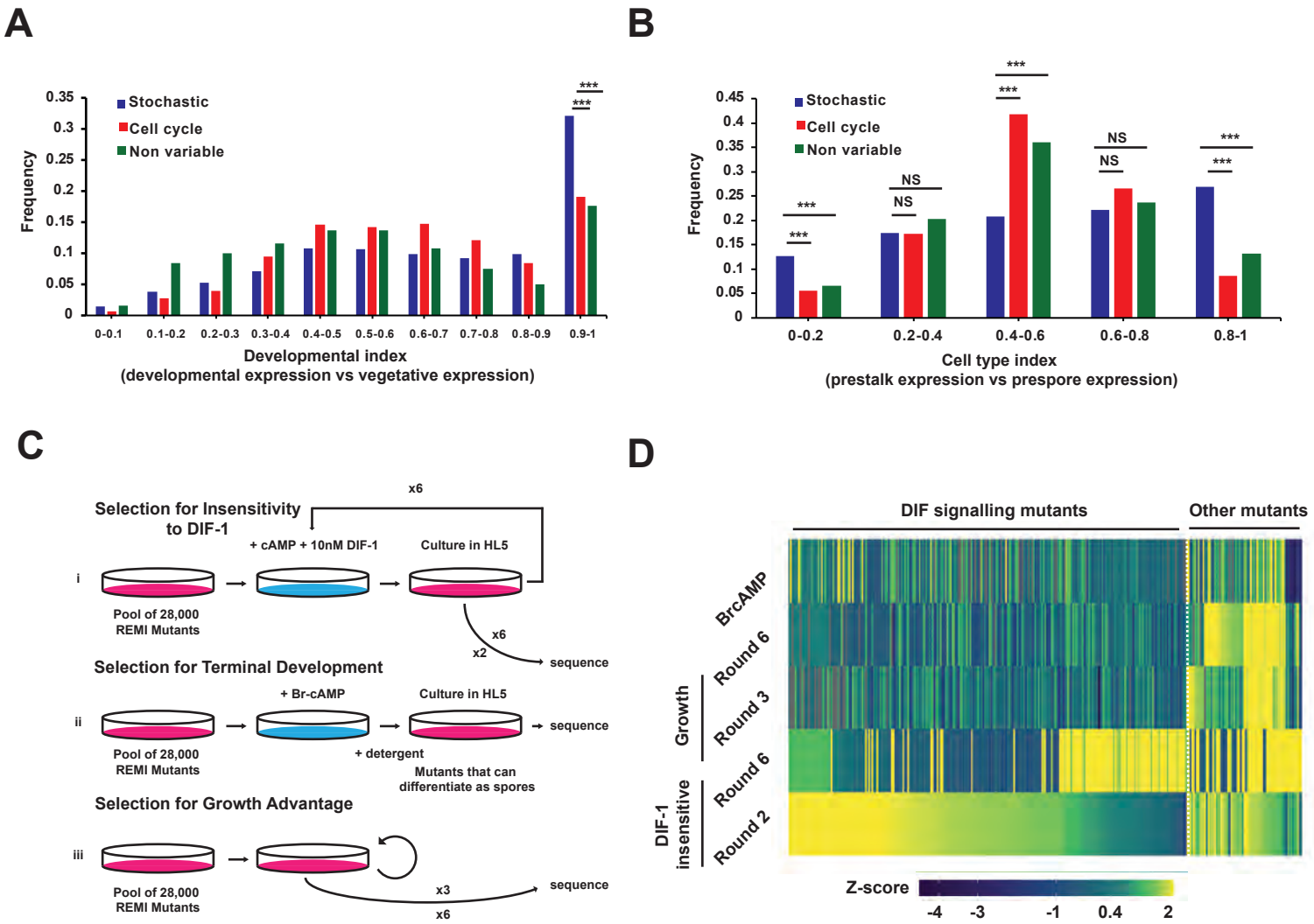

A
Stochastic
Cell cycle
Non variable
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
***
***
0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1
Developmental index
(developmental expression vs vegetative expression)
B
Stochastic
Cell cycle
Non variable
Frequency
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
***
***
NS
NS
***
***
NS
NS
***
***
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1
Cell type index
(prestalk expression vs prespore expression)
C
Selection for Insensitivity to DIF-1
x6
+ cAMP + 10nM DIF-1
Culture in HL5
i
Pool of 28,000 REMI Mutants
x6
x2
sequence
Selection for Terminal Development
+ Br-cAMP
Culture in HL5
ii
Pool of 28,000 REMI Mutants
+ detergent
Mutants that can differentiate as spores
sequence
Selection for Growth Advantage
iii
Pool of 28,000 REMI Mutants
x3
x6
sequence
D
DIF signalling mutants
Other mutants
BrcAMP
Round 6
Round 3
Round 6
Round 2
Growth
DIF-1 insensitive
Z-score
-4
-3
-1
0.4
2
E
% Stalk Cell
100
75
50
25
0
ns
*
**
****
**
****
****
****
****
WT
R2
R6
R2 & R6
Wild type
DDB_G0281087
DDB_G0293590
DDB_G0274923
DDB_G0280389
DDB_G0285055
DDB_G0270586
DDB_G0281971
DDB_G0272474
DDB_G0279425
F
Number of REMI mutants
60
50
40
30
20
10
0
stochastic
cell cycle
expected
*
**
***
stochastic
cell cycle
stochastic + cell cycle
CV² (normalised to expression)
10
9
8
7
6
5
4
3
2
1
0
***
all genes
Cell fate REMI mutants

Figure 4

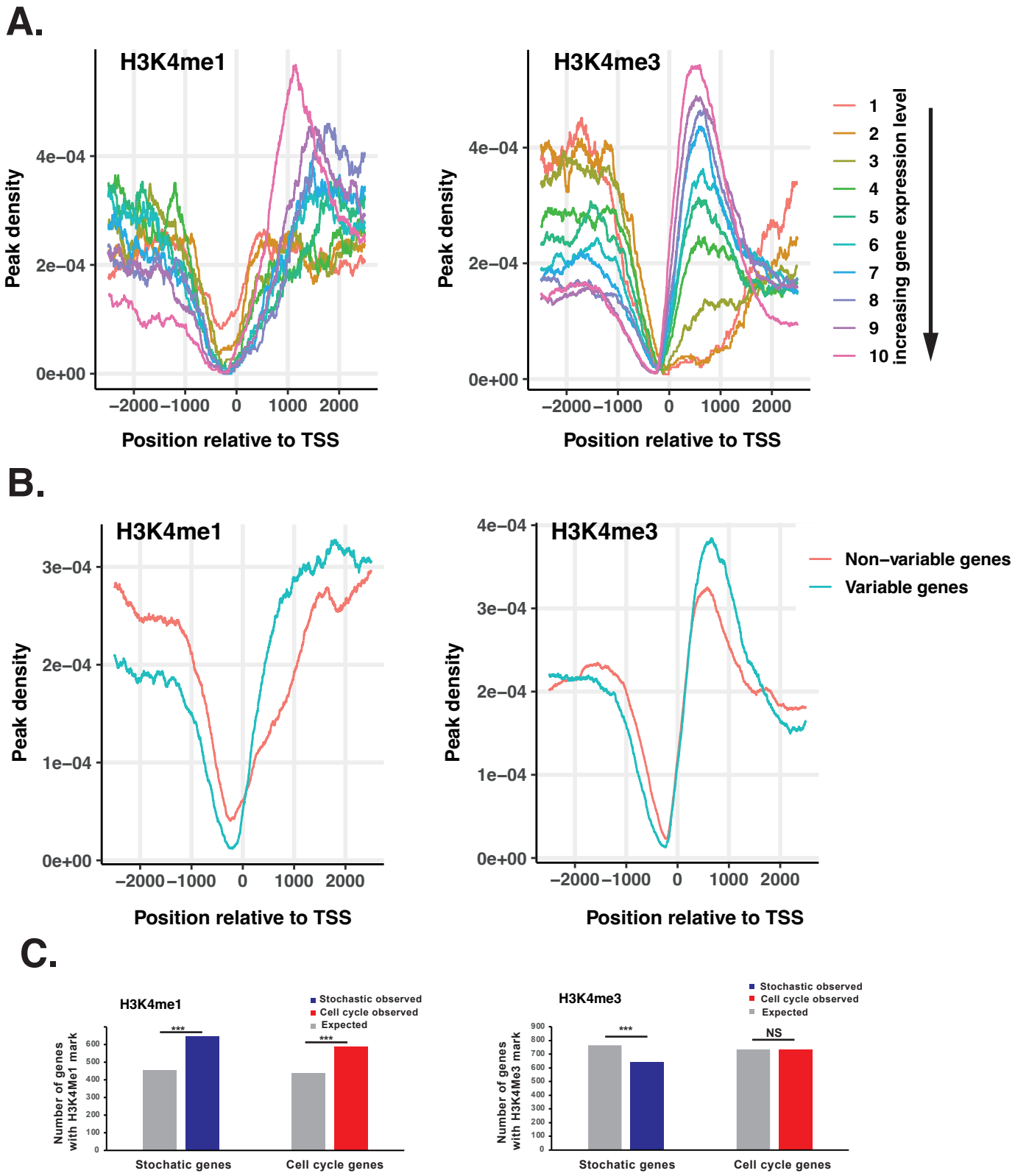

A.
H3K4me1
H3K4me3
Peak density
4e−04
2e−04
0e+00
−2000 −1000 0 1000 2000
Position relative to TSS
1
2
3
4
5
6
7
8
9
10
increasing gene expression level
B.
H3K4me1
H3K4me3
3e−04
1e−04
Non−variable genes
Variable genes
C.
H3K4me1
H3K4me3
Stochastic observed
Cell cycle observed
Expected
Number of genes with H3K4Me1 mark
Number of genes with H3K4Me3 mark
***
NS
Stochatic genes
Cell cycle genes

Figure 5

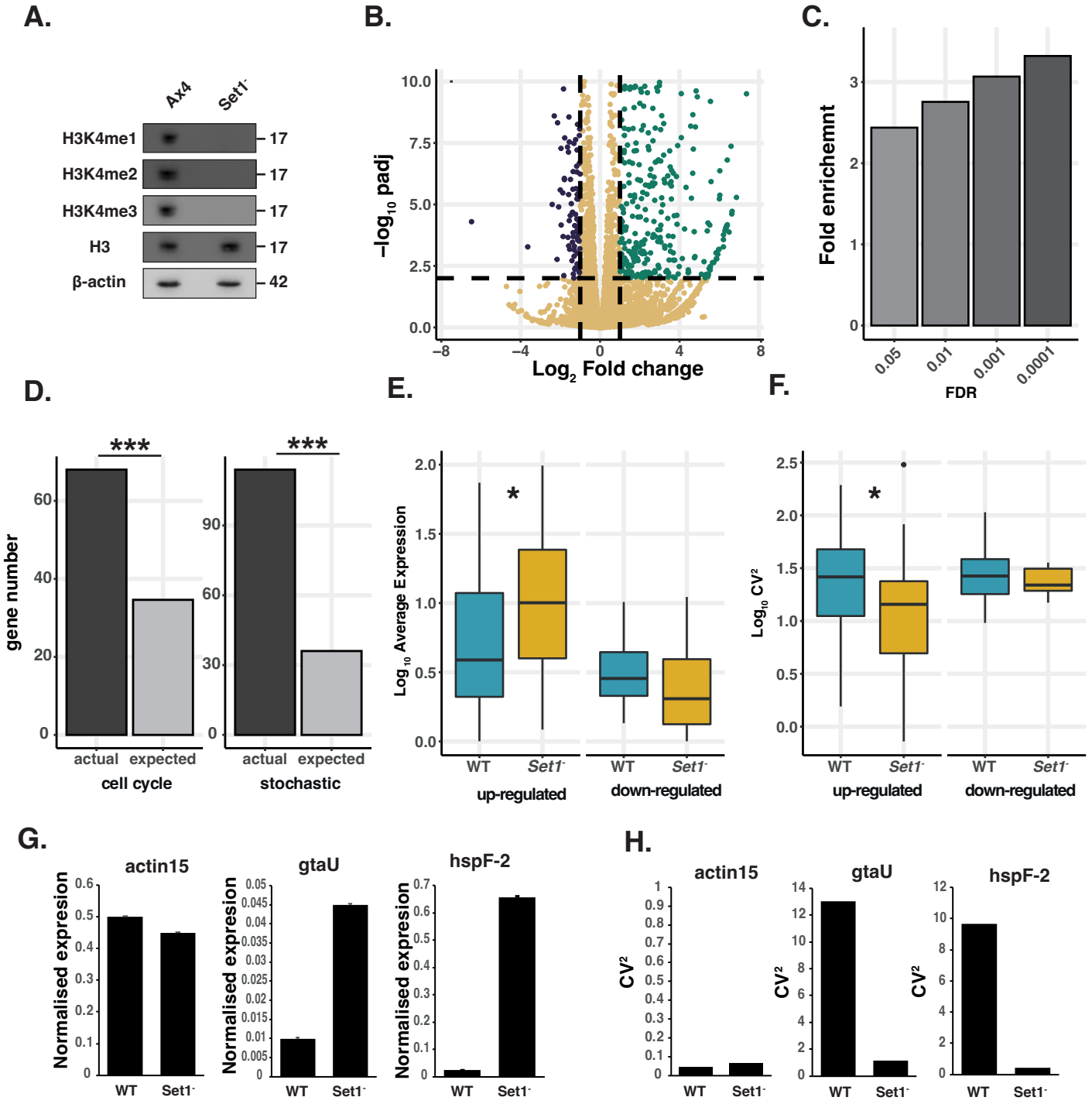

A.
Ax4
Set1-
H3K4me1
H3K4me2
H3K4me3
H3
β-actin
17
17
17
17
42
B.
−log10 padj
Log2 Fold change
C.
Fold enrichemnt
FDR
0.05
0.01
0.001
0.0001
D.
gene number
actual
expected
cell cycle
stochastic
***
E.
Log10 Average Expression
WT
Set1-
up-regulated
down-regulated
F.
Log10 CV²
G.
actin15
gtaU
hspF-2
Normalised expresion
H.
CV²

Figure 6

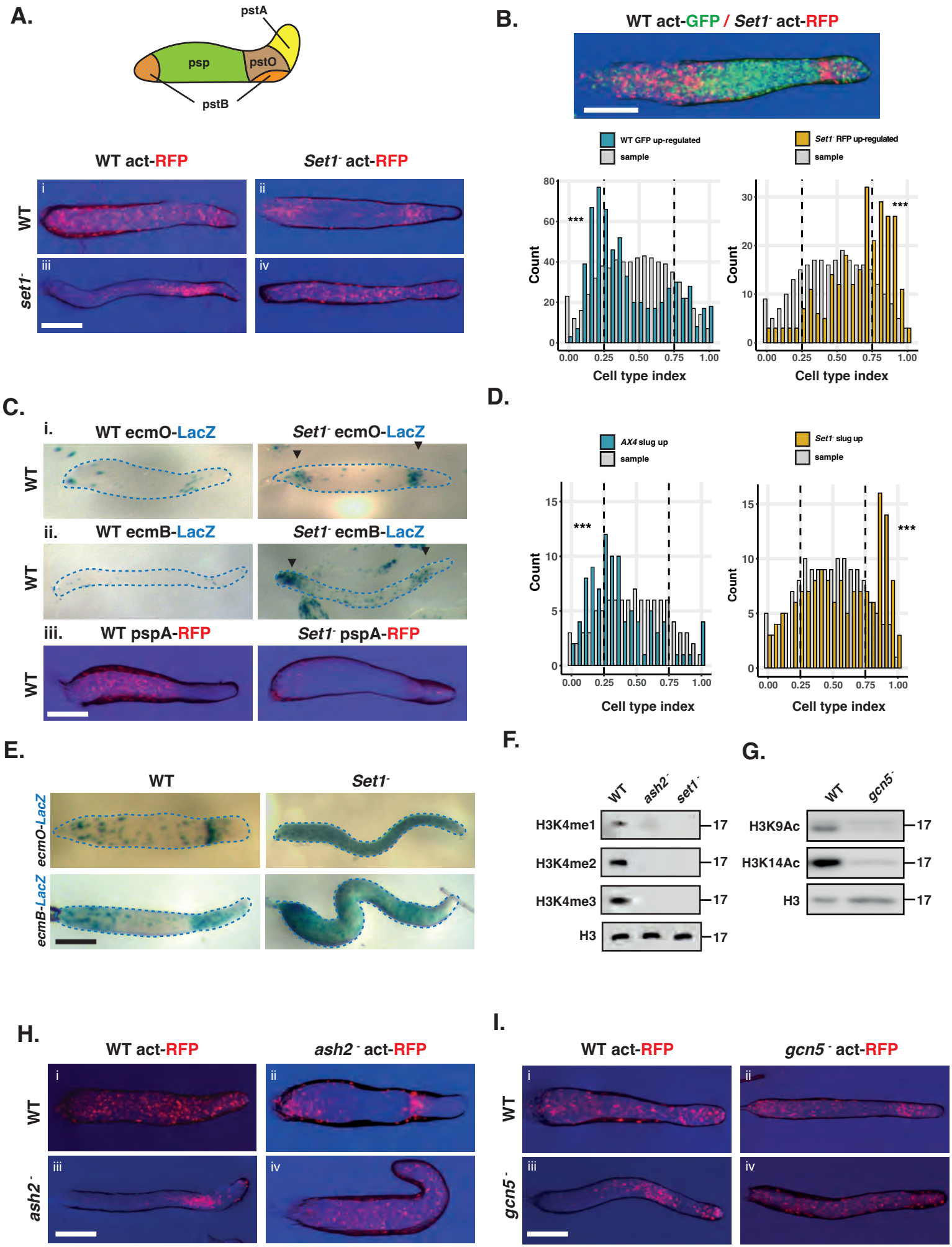

A.
pstA
psp
pstO
pstB
WT act-RFP
Set1- act-RFP
WT
set1-
i
ii
iii
iv
B.
WT act-GFP / Set1- act-RFP
WT GFP up-regulated
sample
Set1- RFP up-regulated
sample
Count
Cell type index
***
C.
i. WT ecmO-LacZ
Set1- ecmO-LacZ
ii. WT ecmB-LacZ
Set1- ecmB-LacZ
iii. WT pspA-RFP
Set1- pspA-RFP
WT
D.
AX4 slug up
sample
Set1- slug up
sample
Count
Cell type index
***
E.
WT
Set1-
ecmO-LacZ
ecmB-LacZ
F.
WT
ash2-
set1-
H3K4me1
H3K4me2
H3K4me3
H3
17
G.
WT
gcn5-
H3K9Ac
H3K14Ac
H3
17
H.
WT act-RFP
ash2- act-RFP
WT
ash2-
i
ii
iii
iv
I.
WT act-RFP
gcn5- act-RFP
WT
gcn5-
i
ii
iii
iv

Figure 7

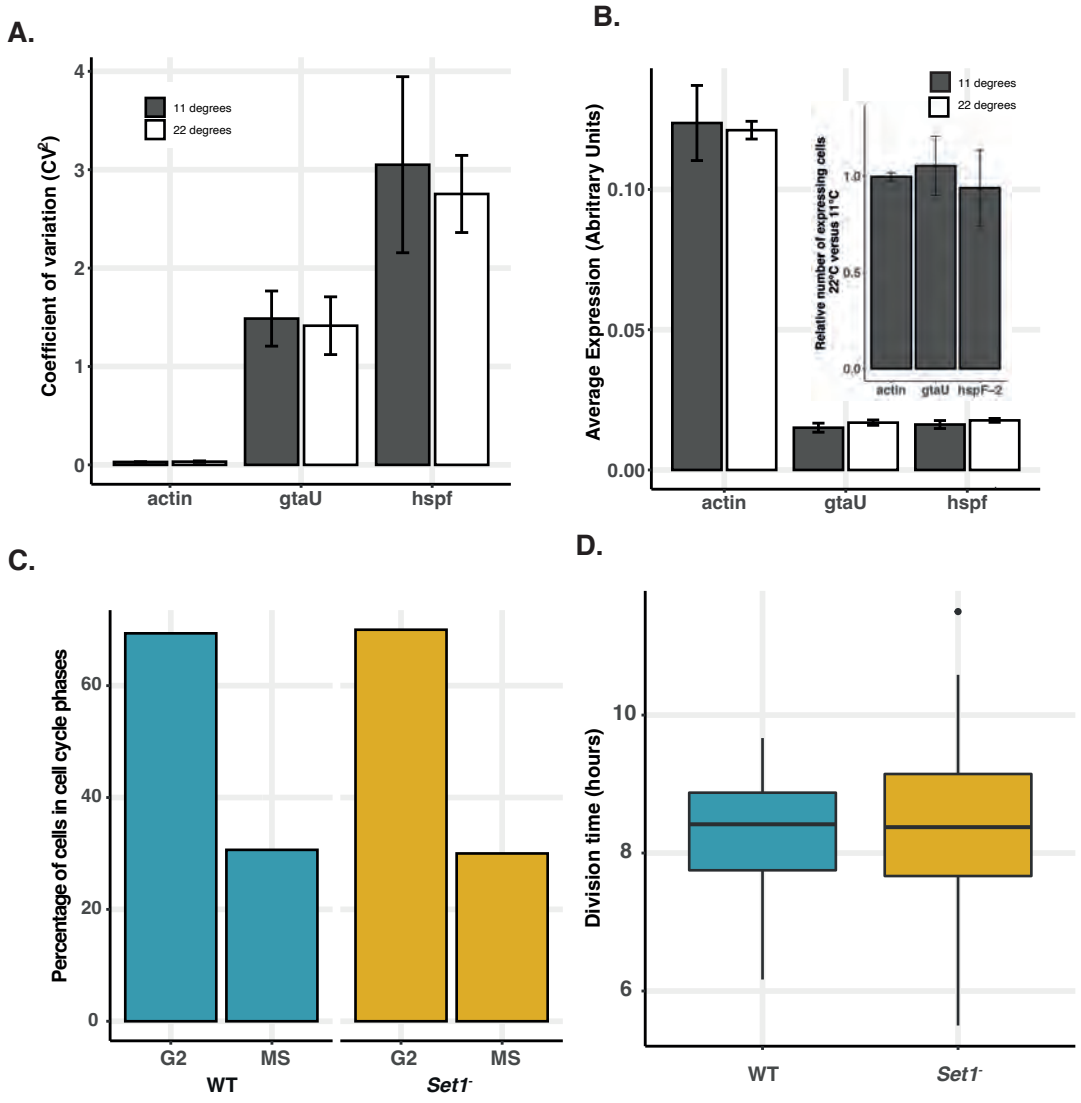
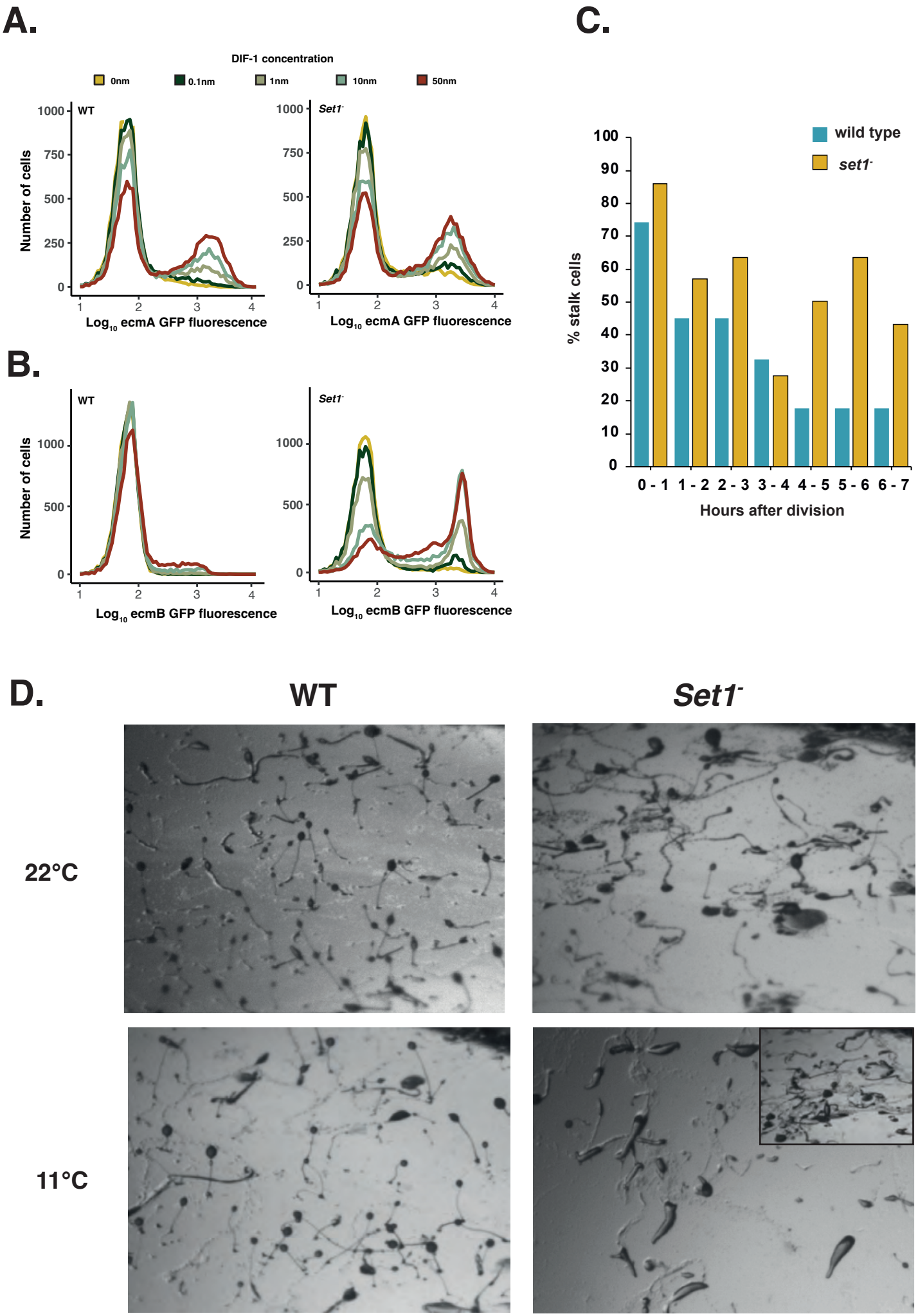
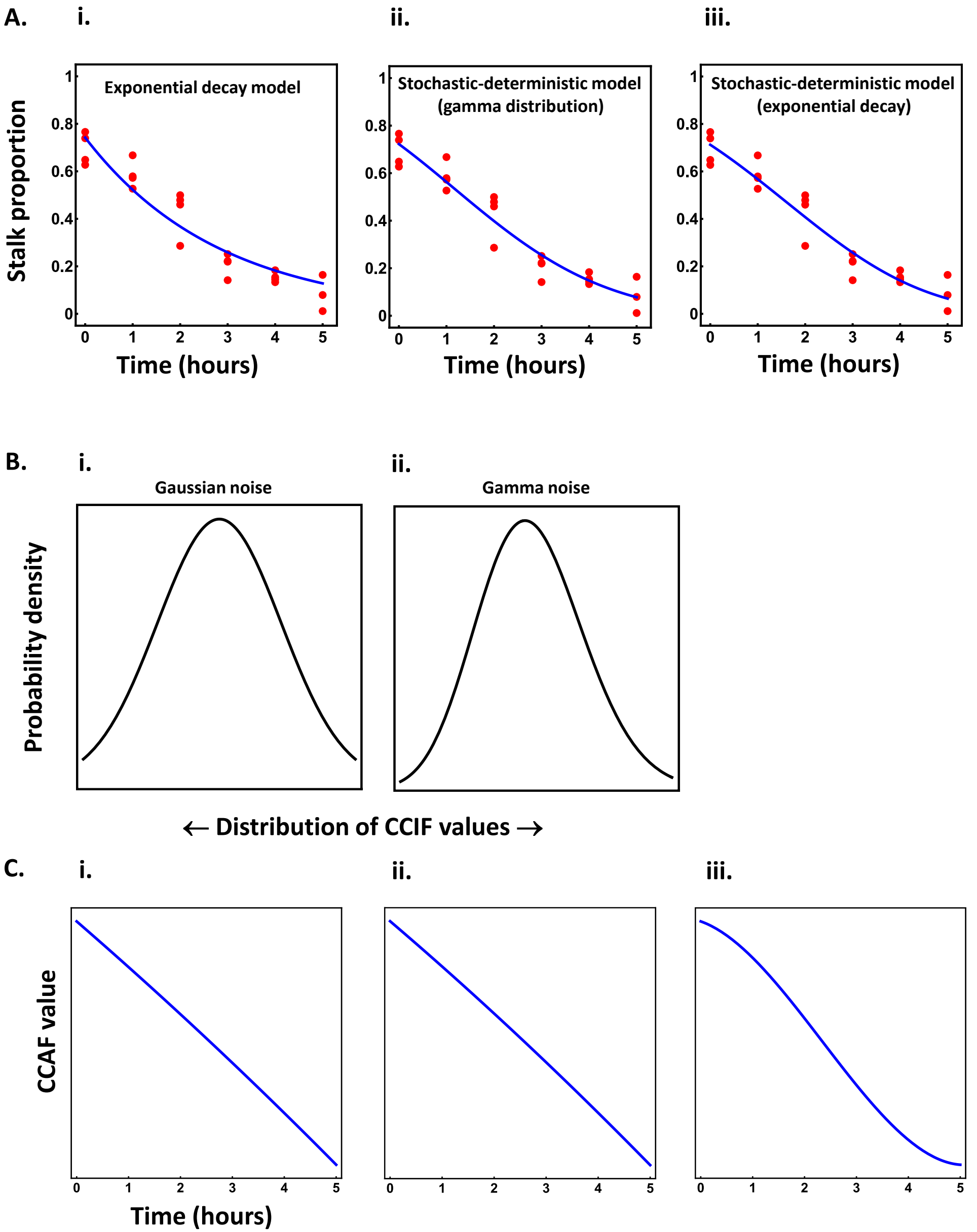

A.
Coefficient of variation (CV²)
11 degrees
22 degrees
actin
gtaU
hspf
B.
Average Expression (Abritrary Units)
11 degrees
22 degrees
Relative number of expressing cells 22°C versus 11°C
actin
gtaU
hspF-2
C.
Percentage of cells in cell cycle phases
G2
MS
WT
Set1⁻
D.
Division time (hours)
WT
Set1⁻

Figure 8

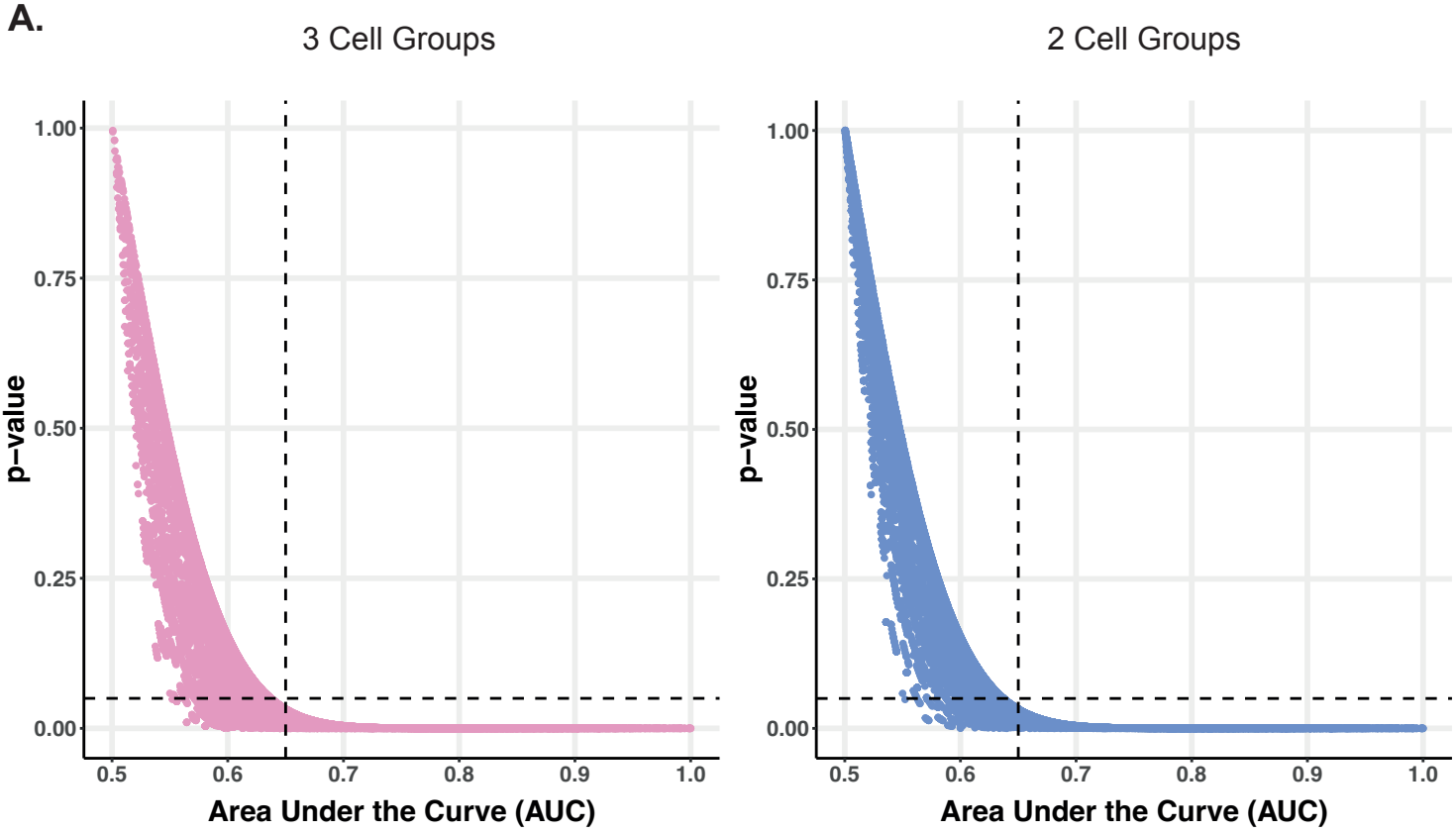

A.
DIF-1 concentration
0nm
0.1nm
1nm
10nm
50nm
WT
Set1-
Number of cells
Log10 ecmA GFP fluorescence
B.
Log10 ecmB GFP fluorescence
C.
wild type
set1-
% stalk cells
Hours after division
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
D.
WT
Set1-
22°C
11°C

Supplementary Figure 1

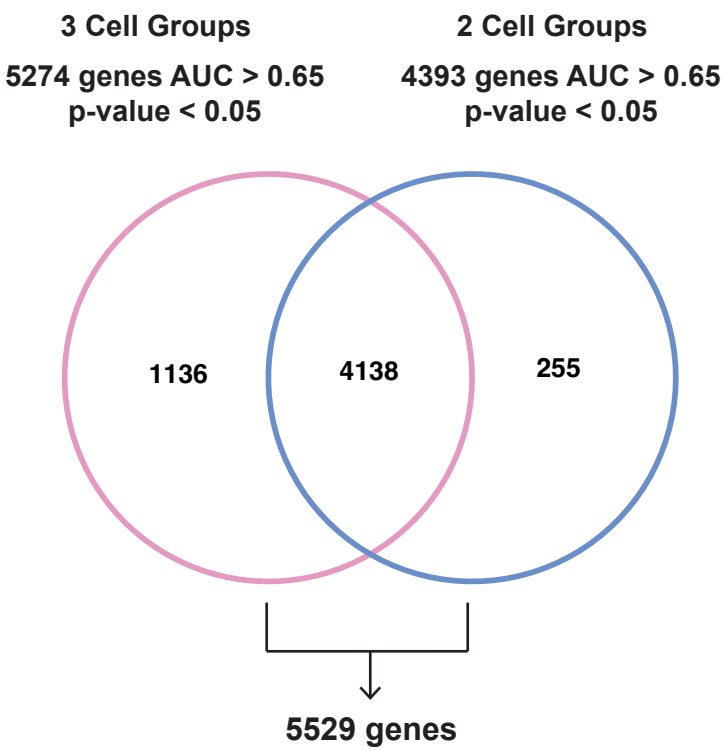


Supplementary figure 2

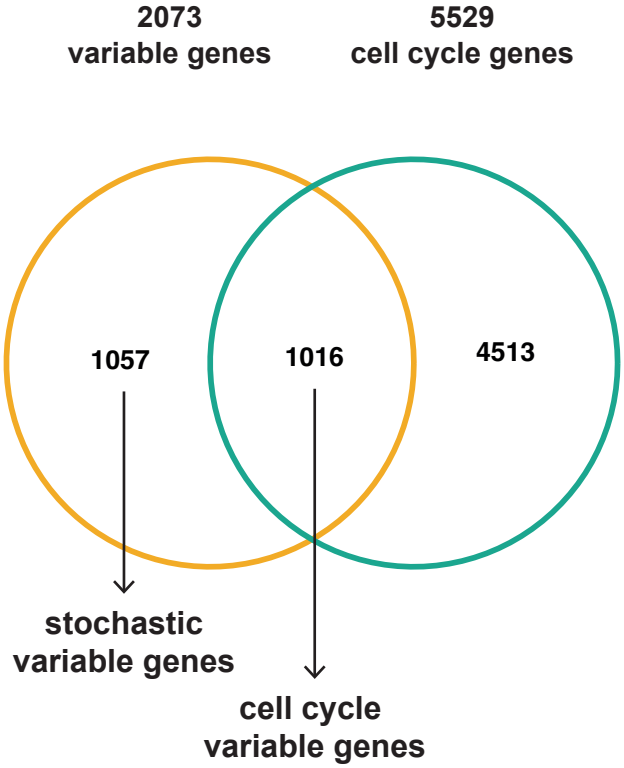
A.



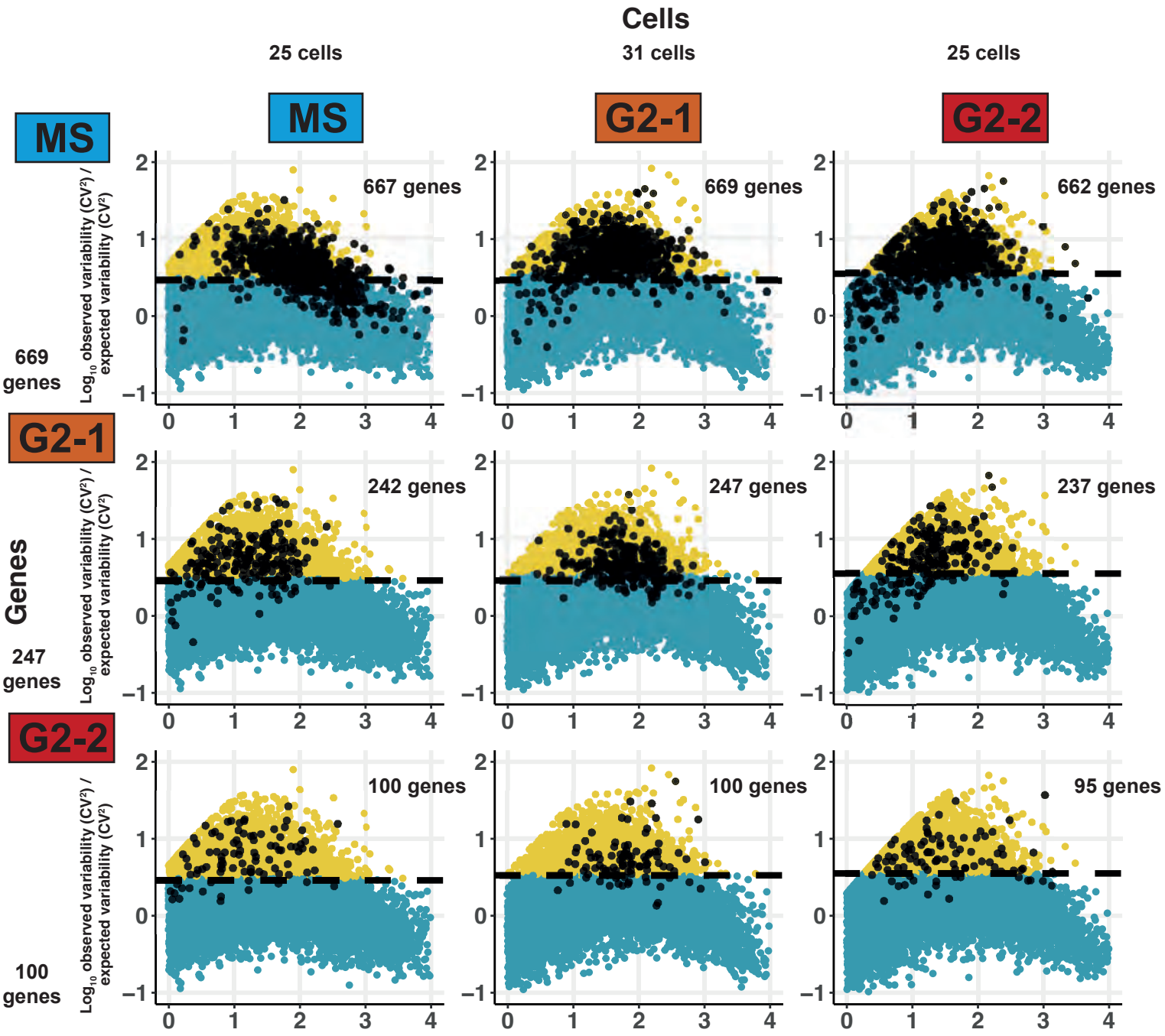
B.



C.

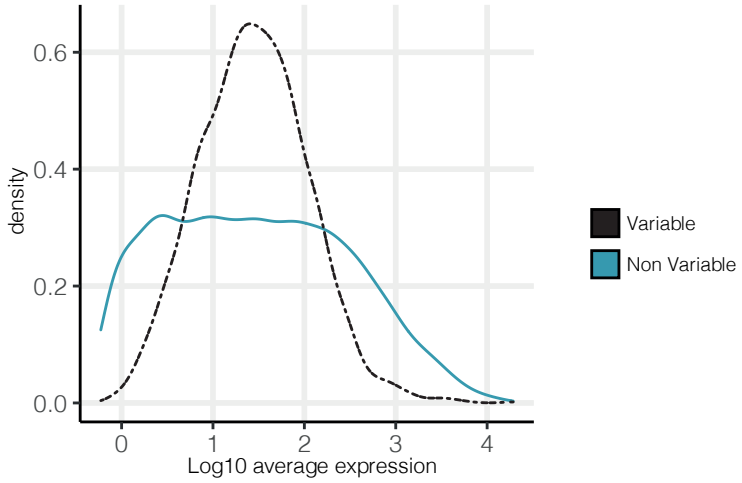


Supplementary figure 3

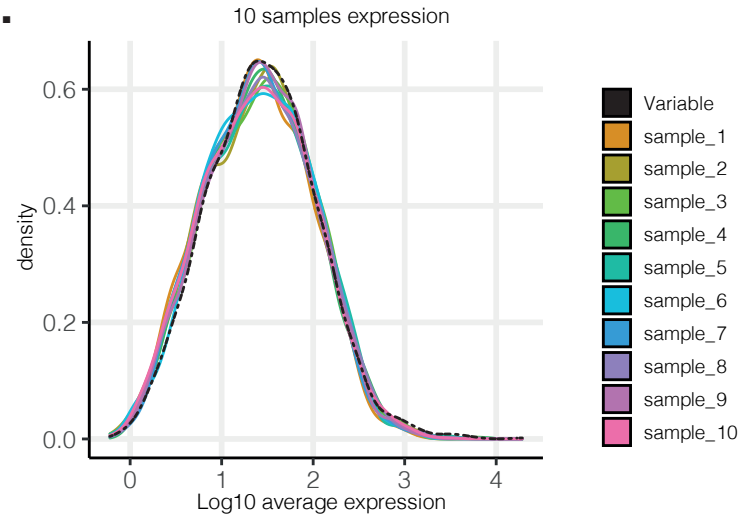


Supplementary figure 4

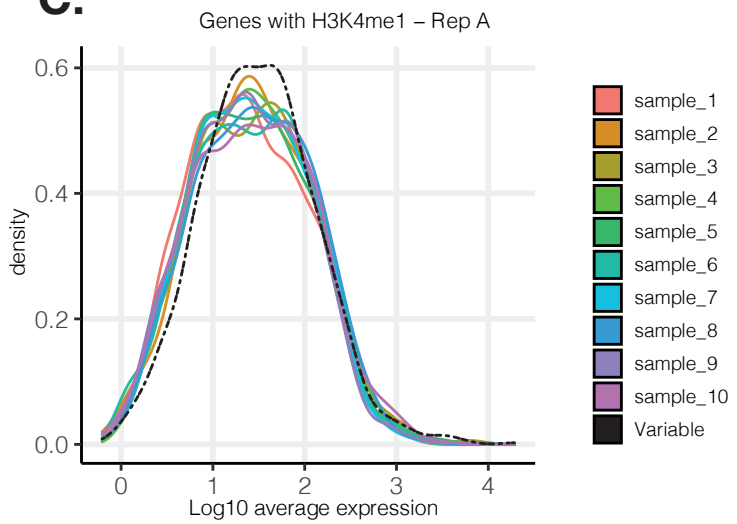
A.



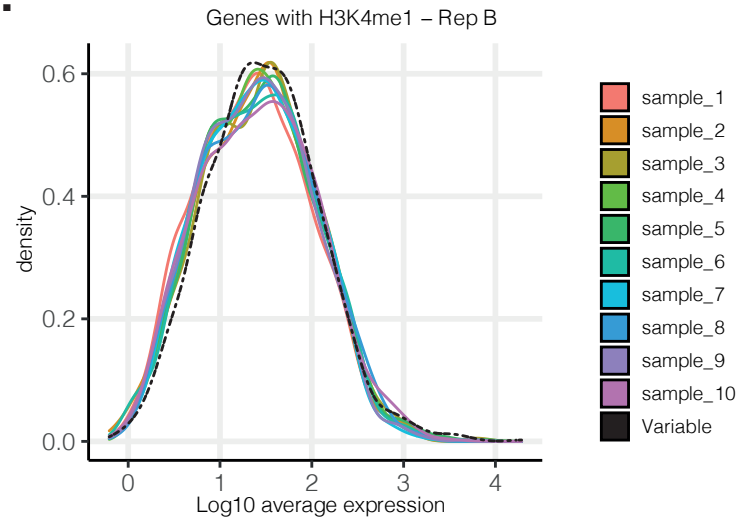
B.



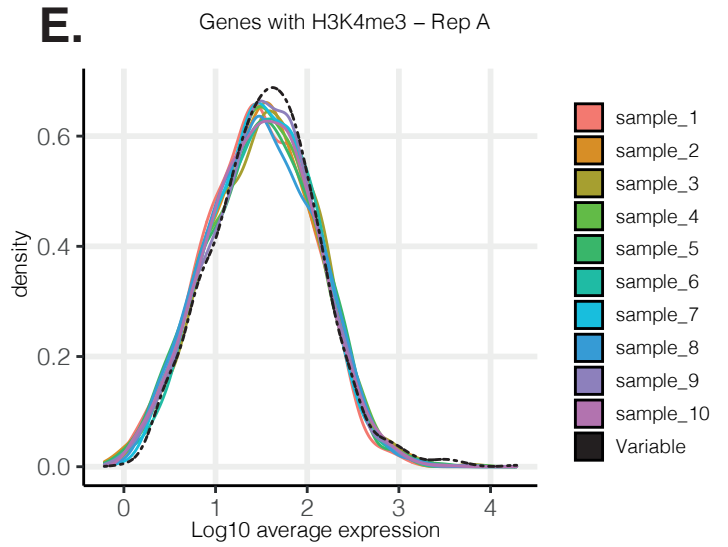
C.



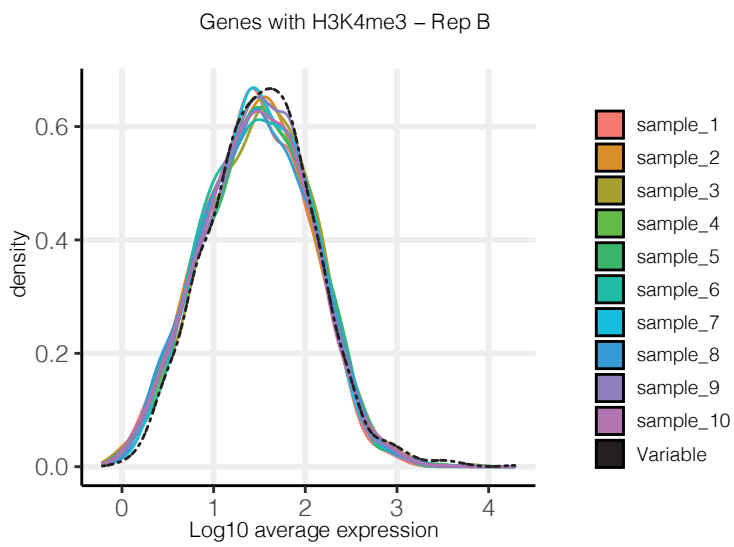
D.



E.



F.



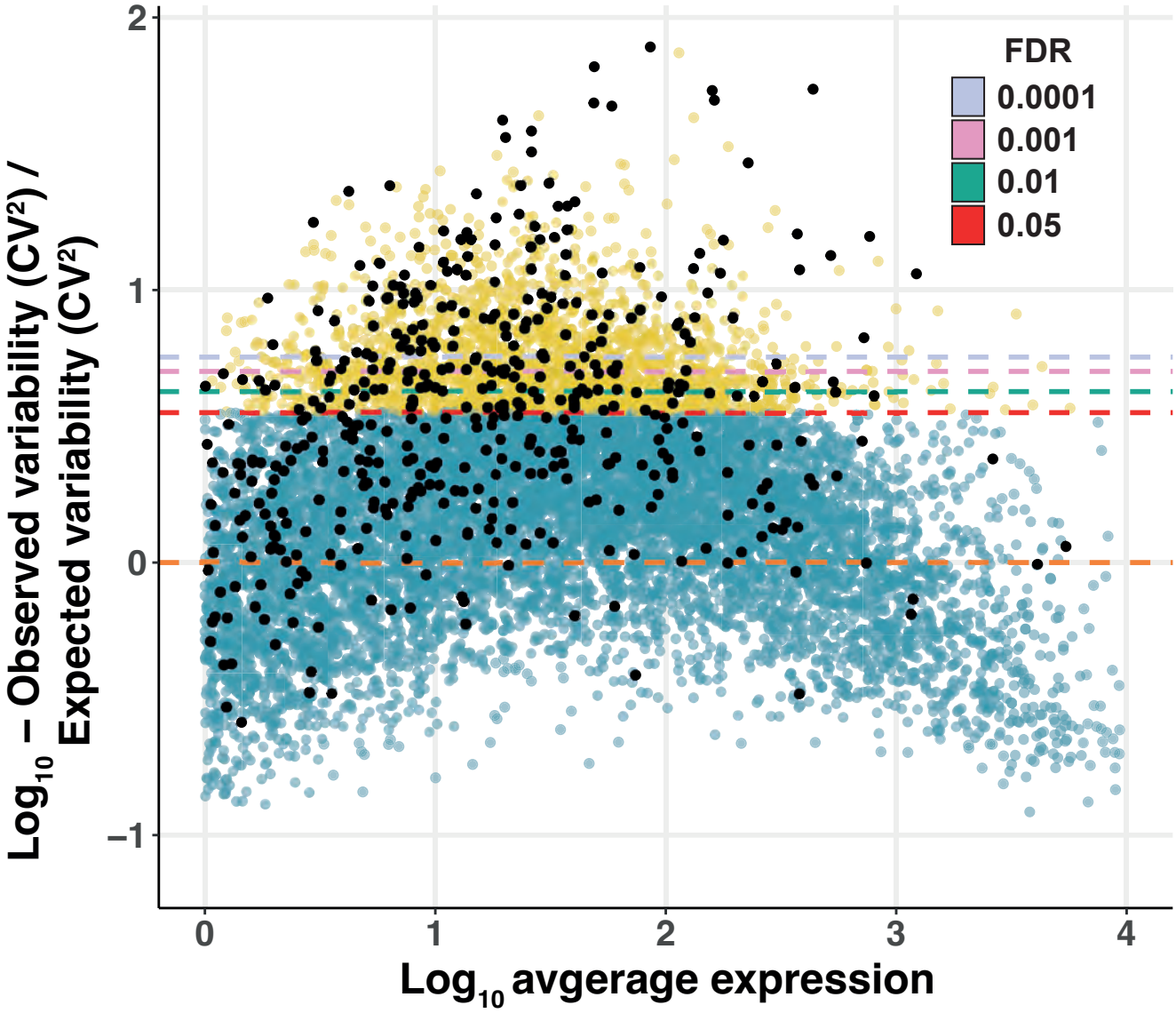
G.

	Replicate_A	Replicate_B
Ten Samples of Non Variable Genes with H3K4me1– Averaged	780.9	1005.4
Variable Genes with H3K4me1	1121	1349

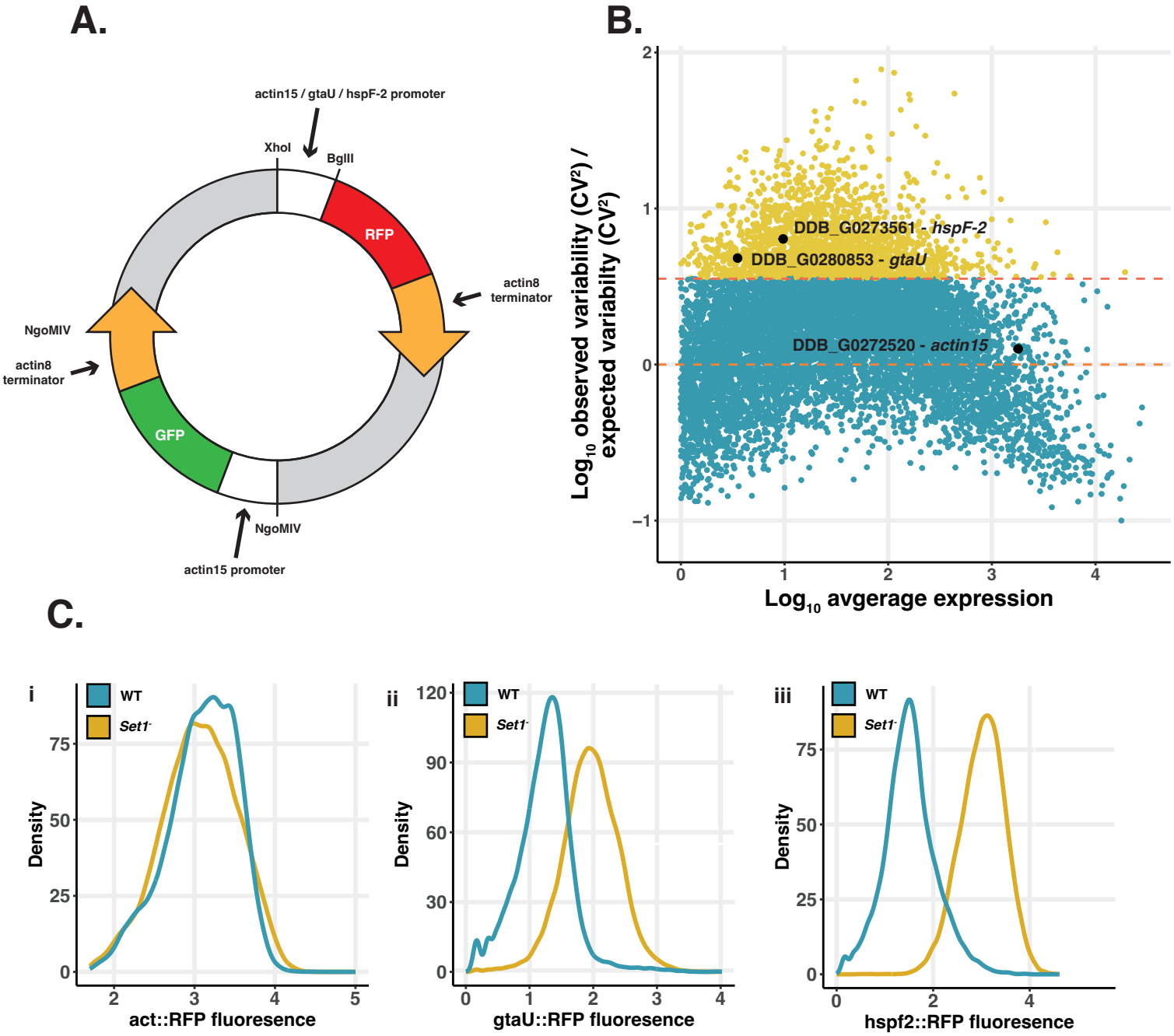
H.

	Replicate_A	Replicate_B
Ten Samples of Non Variable Genes with H3K4me3– Averaged	1370.5	1630.1
Variable Genes with H3K4me3	1233	1505

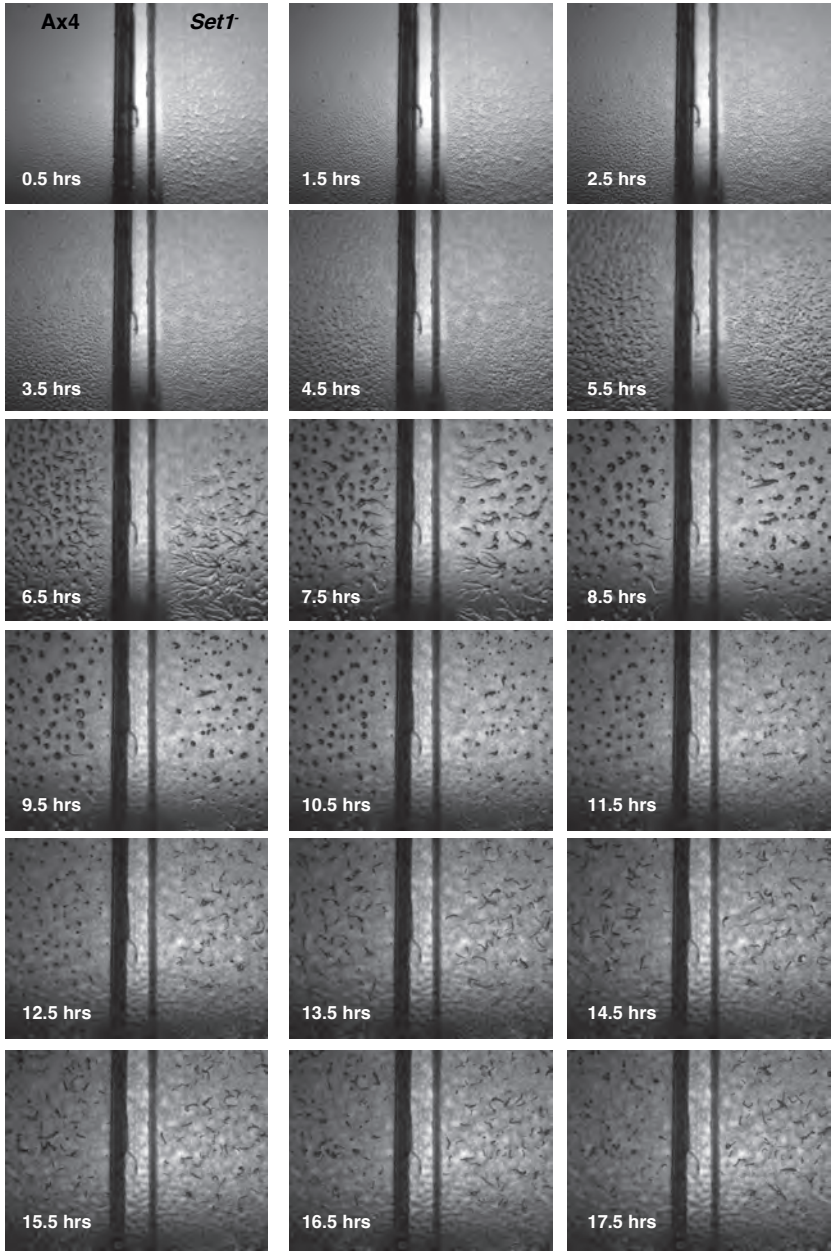
Supplementary figure 5

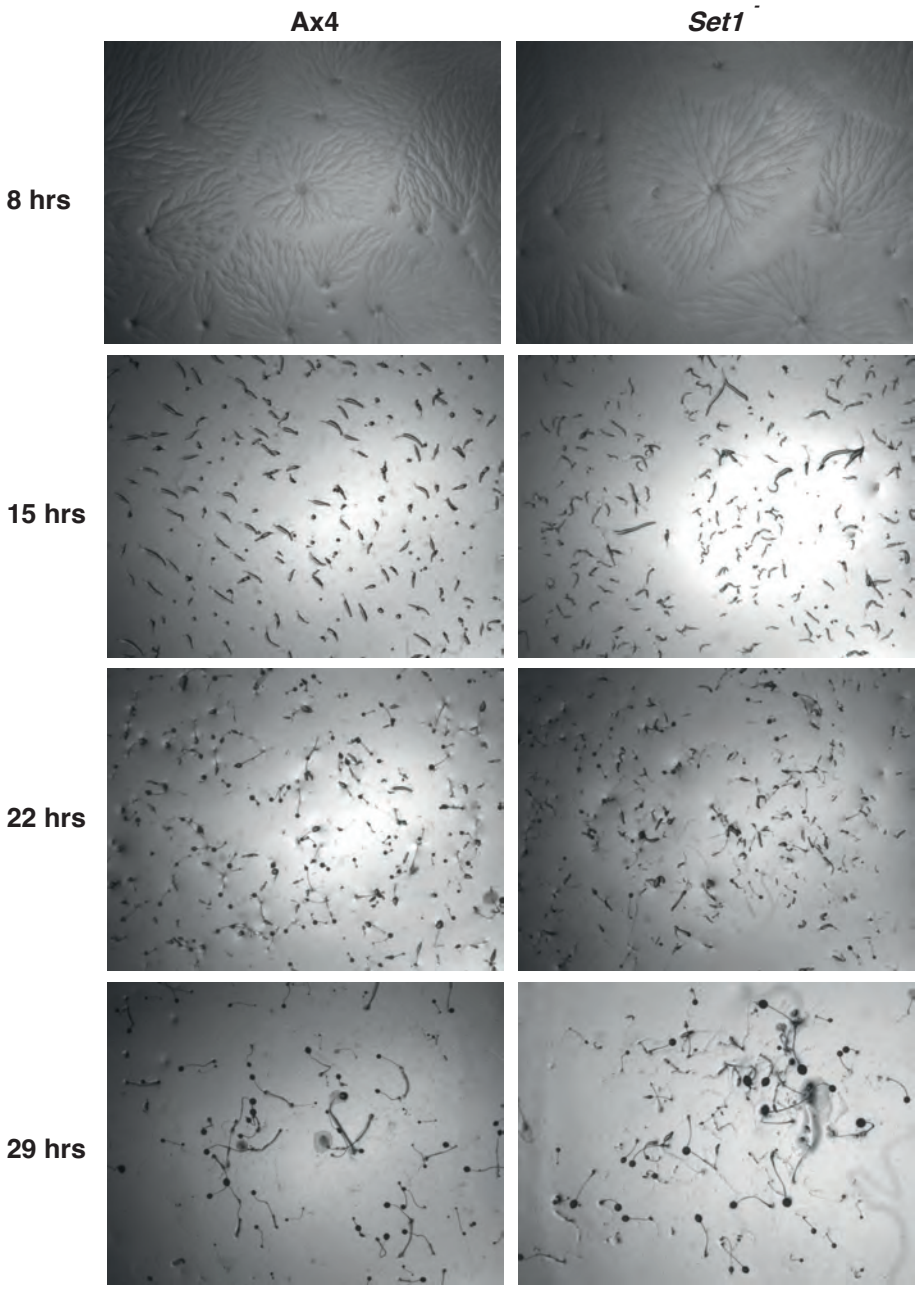


Supplementary figure 6

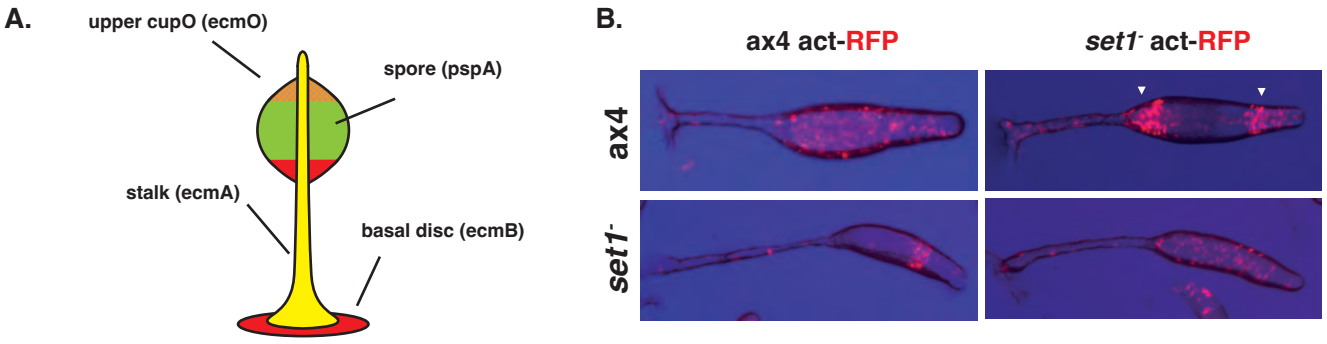


Supplementary figure 7

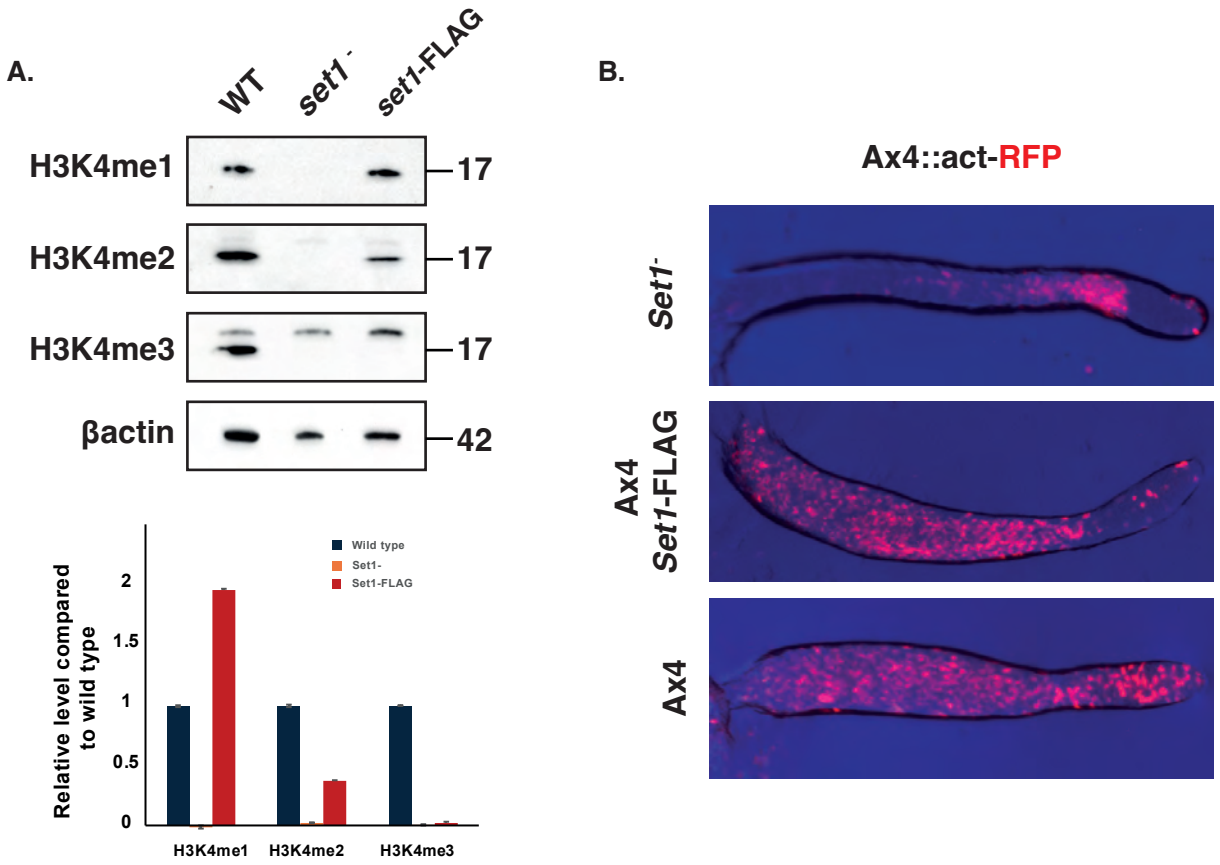




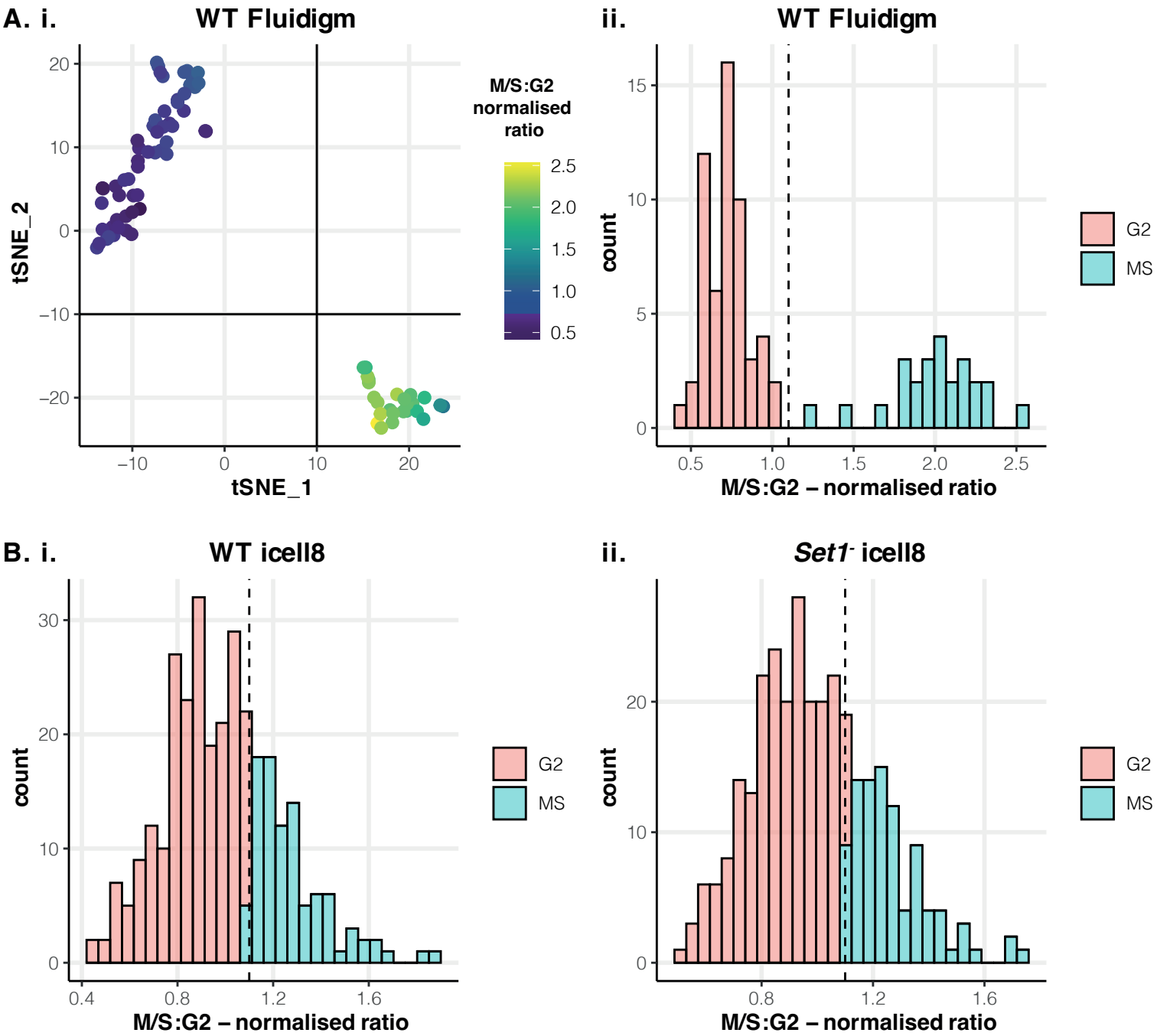
Supplementary figure 8



Supplementary figure 9

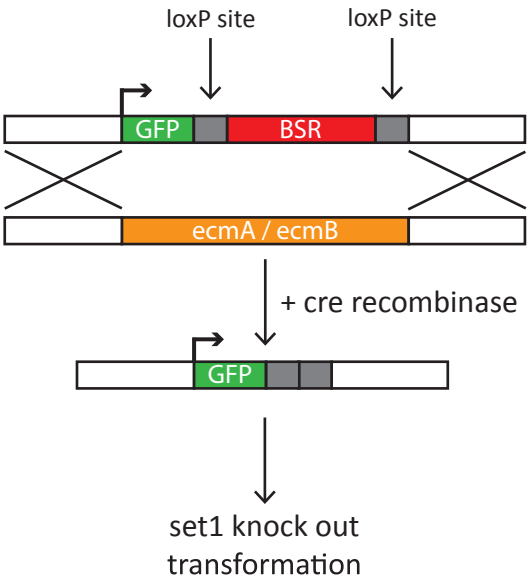


Supplementary figure 10

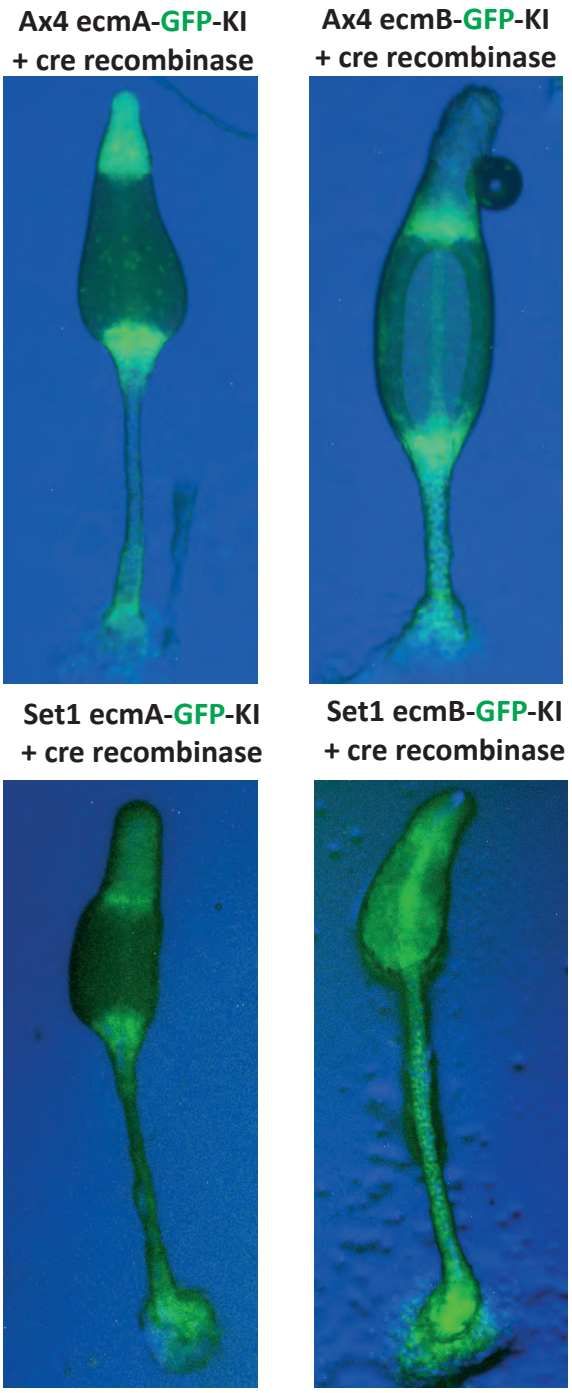
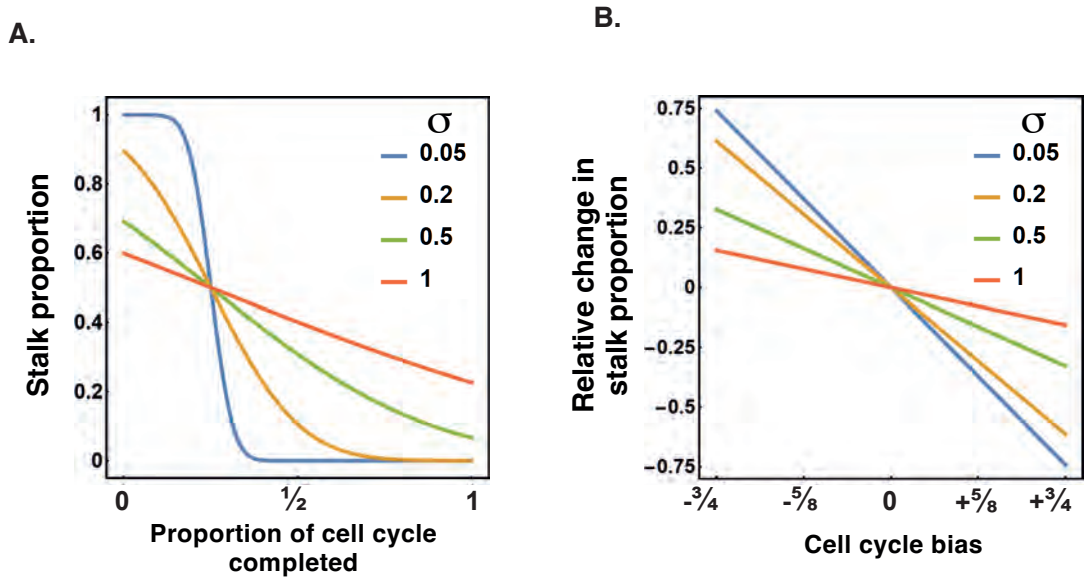


Supplementary figure 11

A



B



Supplementary figure 12