

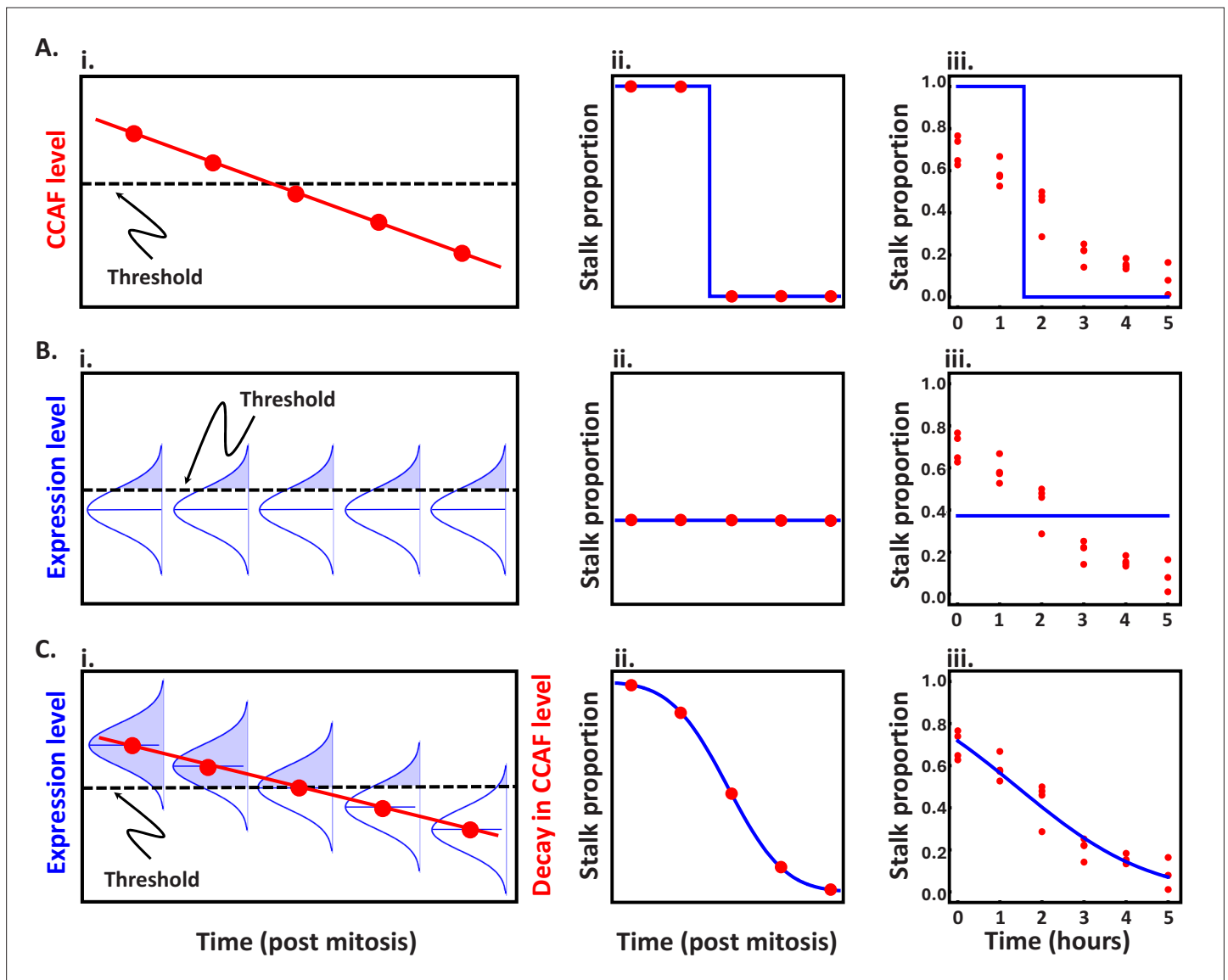


---

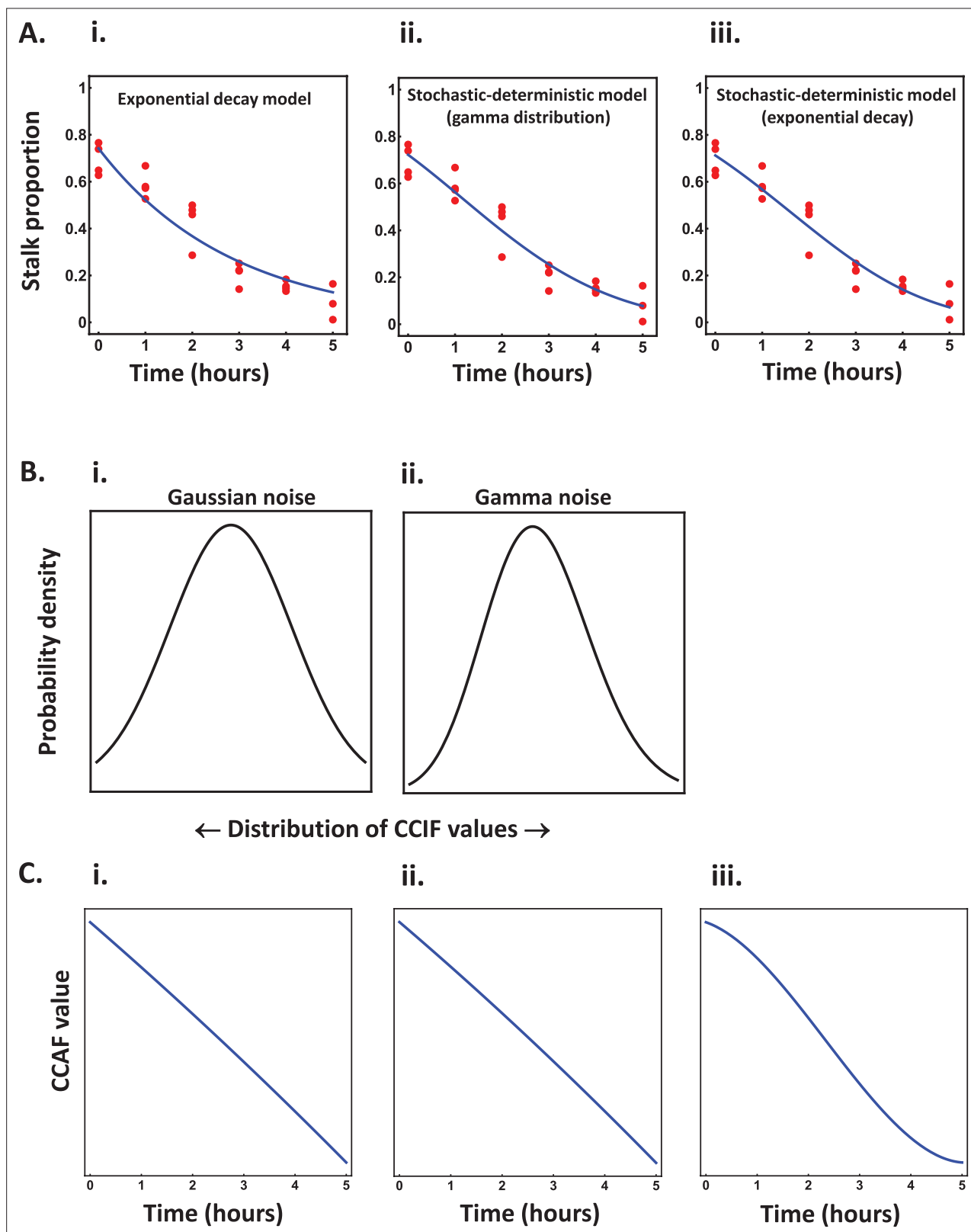
## Figures and figure supplements

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

**William Salvidge *et al.***



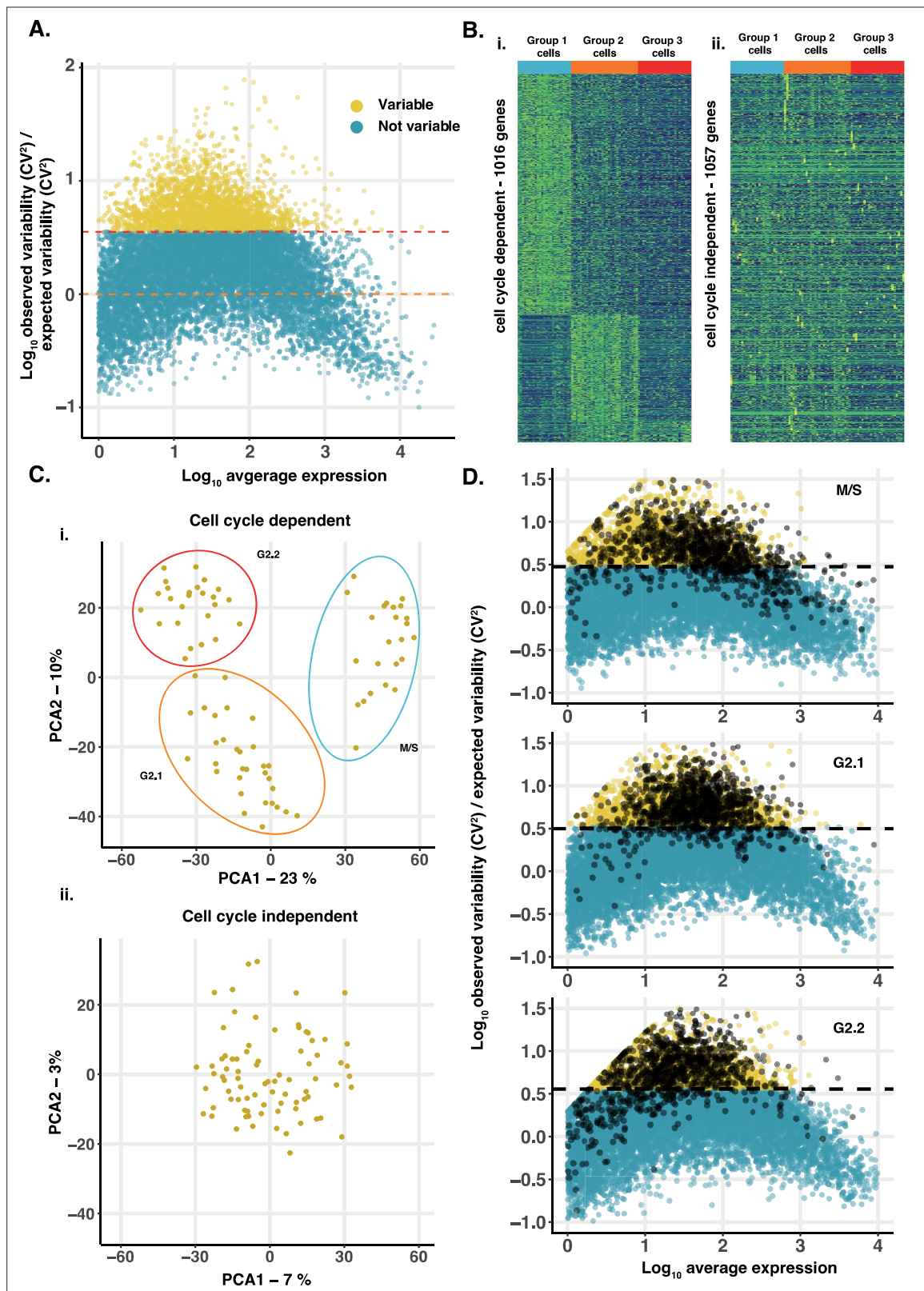
**Figure 1.** Theoretical patterns of stalk proportioning through the cell cycle. Plots illustrate the model structure, theoretical expectations for stalk proportioning ( $P_t$ ) through the cell cycle, and corresponding fits to empirical data (from [Gruenheit et al., 2018](#)) for scenarios in which 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  $\beta$ ; see [Equation 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  $\mu$  and a standard deviation  $\sigma$ ) 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 [Equation 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 right-hand 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 1—figure supplement 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 cell-cycle associated factor (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 Appendix Figure 1—figure supplement 1 continued on next page

*Figure 1—figure supplement 1 continued*

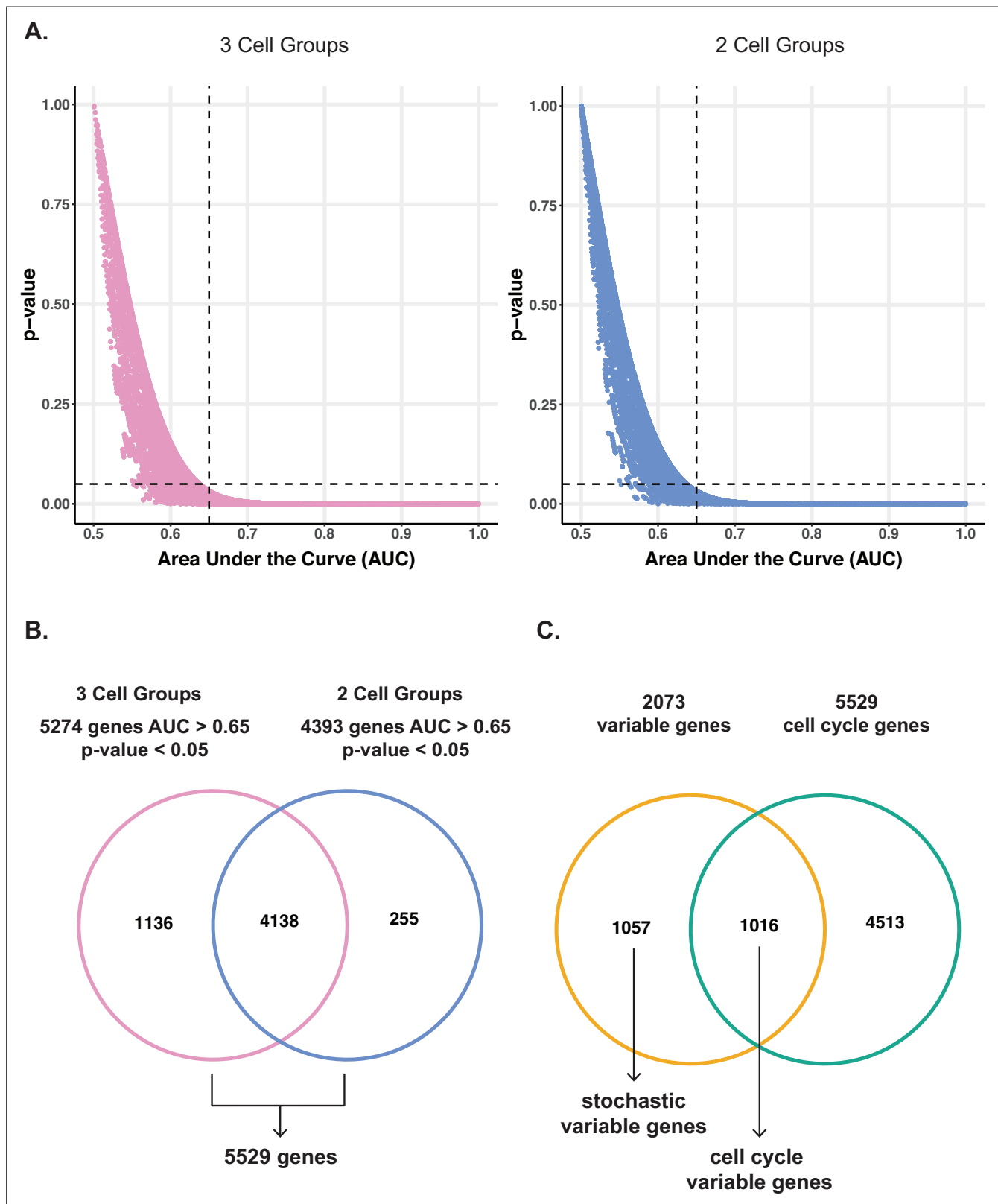
1). **(ii)** Expected stalk proportioning through the cell cycle predicted by a version of the stochastic-deterministic model for the case where gene expression noise is given by a gamma distribution (instead of being Gaussian) (see Appendix 1). The gamma distribution has two additional parameters corresponding to the gamma distribution shape and scale parameters  $k$  and  $\theta$ . 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 Appendix 1). 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 **Equation 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 **Equation 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.



**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 associated and cell cycle-independent genes. Variable genes were defined as cell cycle-associated (1016 genes) or cell cycle-independent (1057 genes). (C) Cell cycle dependent and independent genes were identified using PCA. (D) Cell cycle dependent genes were identified using PCA. Figure 2 continued on next page

*Figure 2 continued*

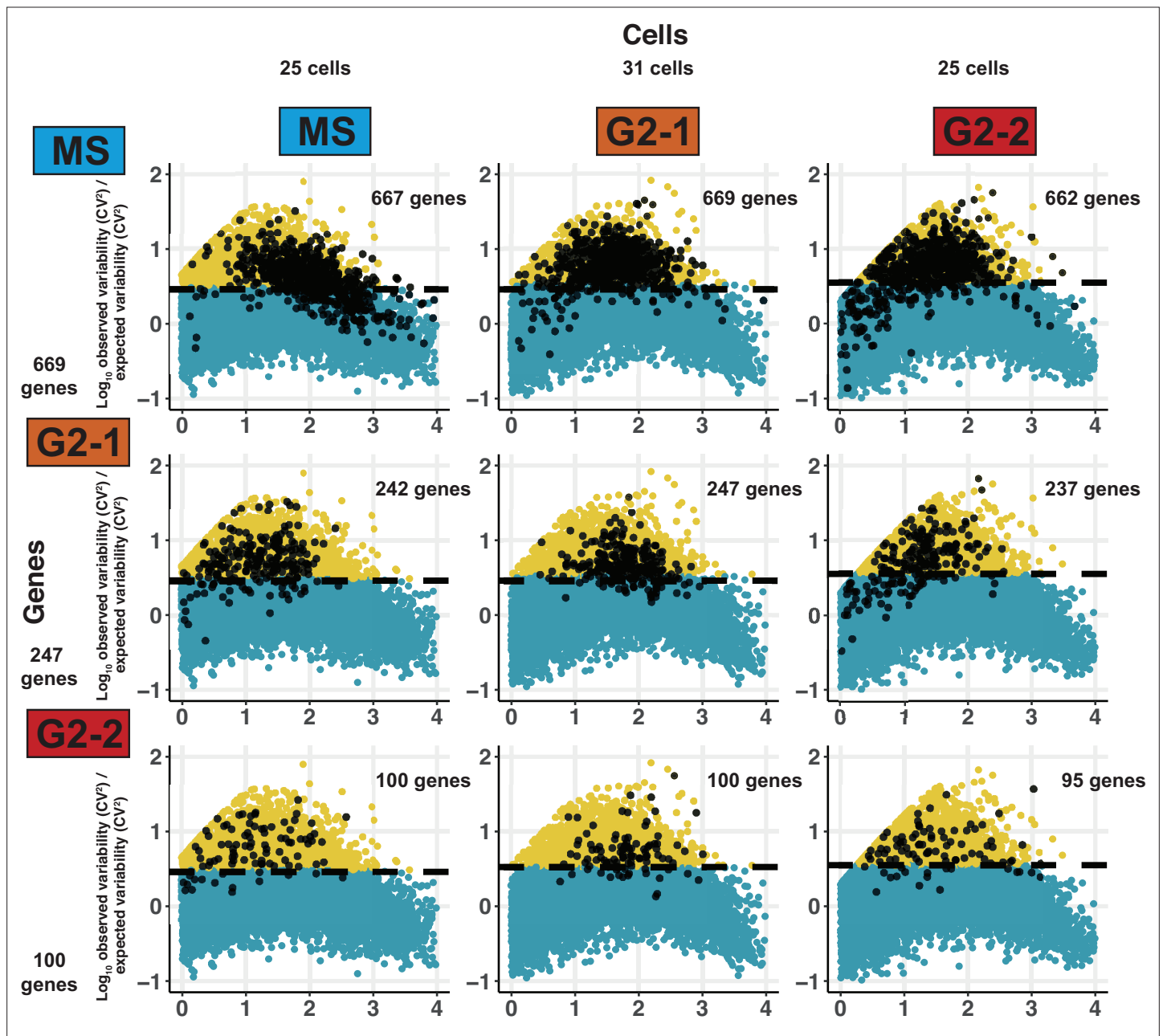
cycle-independent (1057 genes). **(i)** Hierarchical clustering reveals cell cycle-associated genes show a strong cell cycle-associated gene expression pattern blue (S/M phase, 669 genes), orange (G2.1 phase, 247 genes) and red (G2.2 phase, 100 genes). **(ii)** Cell cycle independent genes show little 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.



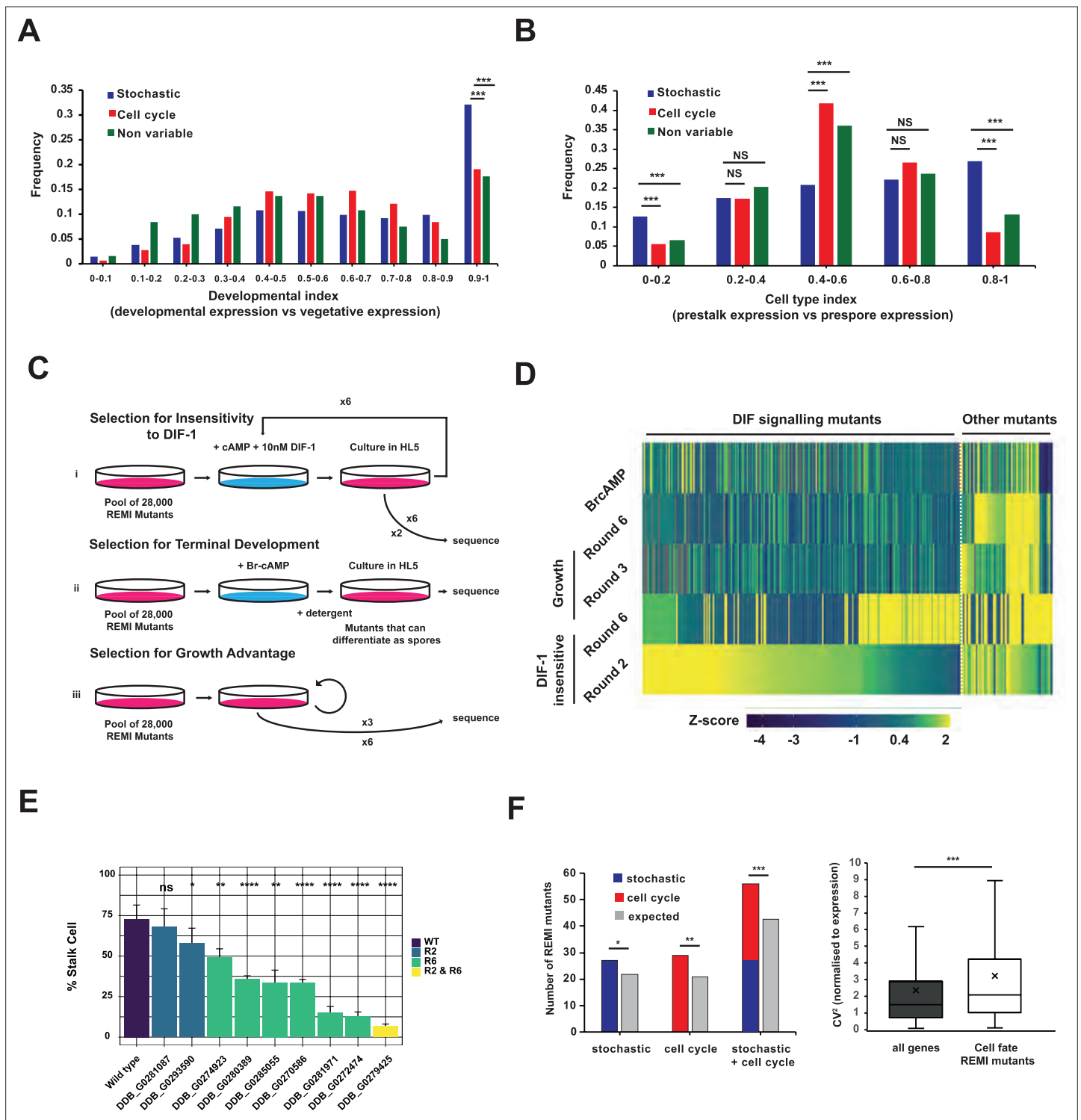
**Figure 2—figure supplement 1.** Cell cycle dependent genes defined through AUROC analysis. (A) Previous analyses of scRNAseq of 81 single cells reveals cells can be divided into two or three cell groups based on cell cycle position (Gruenheit et al., 2018). To identify markers associated with the cell cycle, relaxed AUROC (Area Under the Receiver Operating Characteristic Curve) thresholds of >0.65 and  $p < 0.05$  (dashed lines) 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 Figure 2—figure supplement 1 continued on next page

Figure 2—figure supplement 1 continued

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.



**Figure 2—figure supplement 2.** 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.

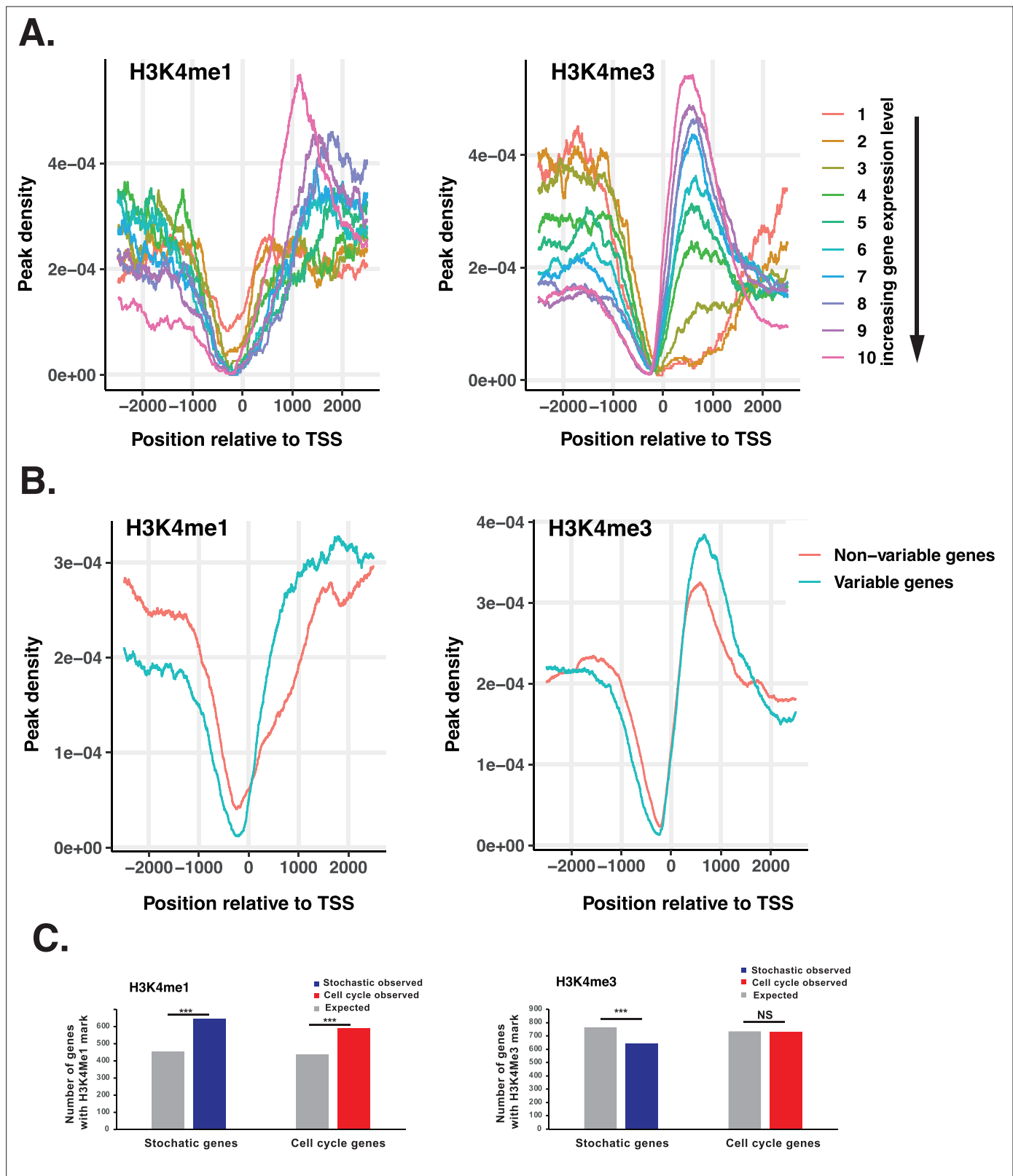


**Figure 3.** Variably expressed genes are important for prestalk cell differentiation. **(A)** Developmental index of different groups of genes. Average gene expression during developmental stages was compared to average expression during growth + average expression during development. This developmental index ranges from zero to one. Zero is exclusive to growth and one is exclusive to development ( $***p < 0.001$ ). **(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). ( $***p < 0.001$ ). **(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

Figure 3 continued on next page

## Figure 3 continued

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. 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). 244 insertions were classified as DIF signalling mutants that affect stalk cell differentiation. Level of enrichment is shown as a Z-score. (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 ( $*p \leq 0.05$ ,  $** P \leq 0.01$ ,  $*** P \leq 0.001$ ).

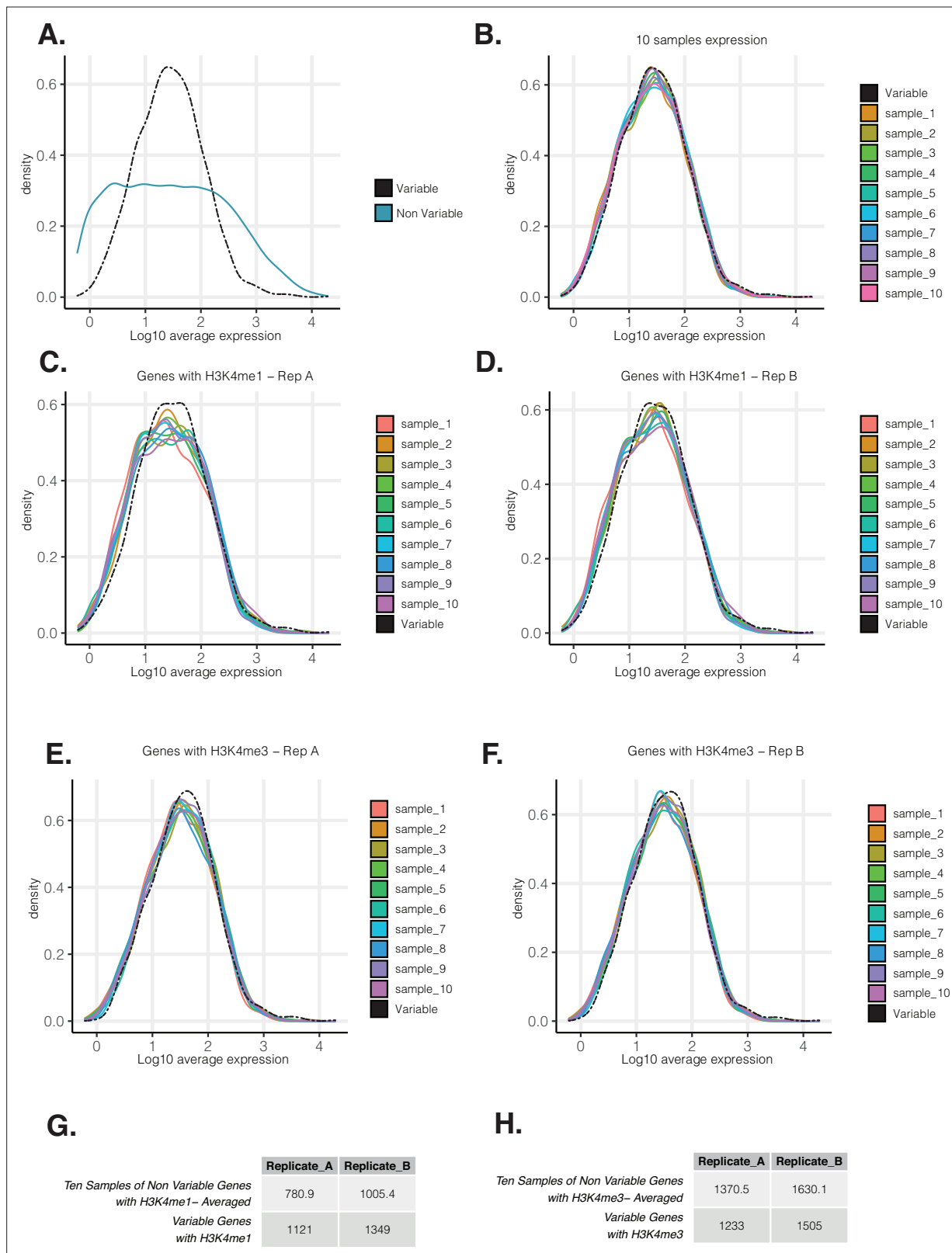


**Figure 4.** Variably expressed genes exhibit differential patterns of histone modifications. **(A).** The H3K4 methylation profile in *D. discoideum*. Genes were assigned to ten bins according to their average expression level in vegetative AX4 cells (Supplementary data). Genes were identified that contain H3K4me1 or H3K4me3 peaks around the TSS (–2500 bp to + 2500 bp) 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

Figure 4 continued on next page

*Figure 4 continued*

variable and non-variable genes. The peak density of variable genes with at least one H3K4me1 or H3K4me3 peak between 2500 bp upstream and 2500 bp 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 (**Figure 4—figure supplement 1**). 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 (\*\* $p \leq 0.001$ ).

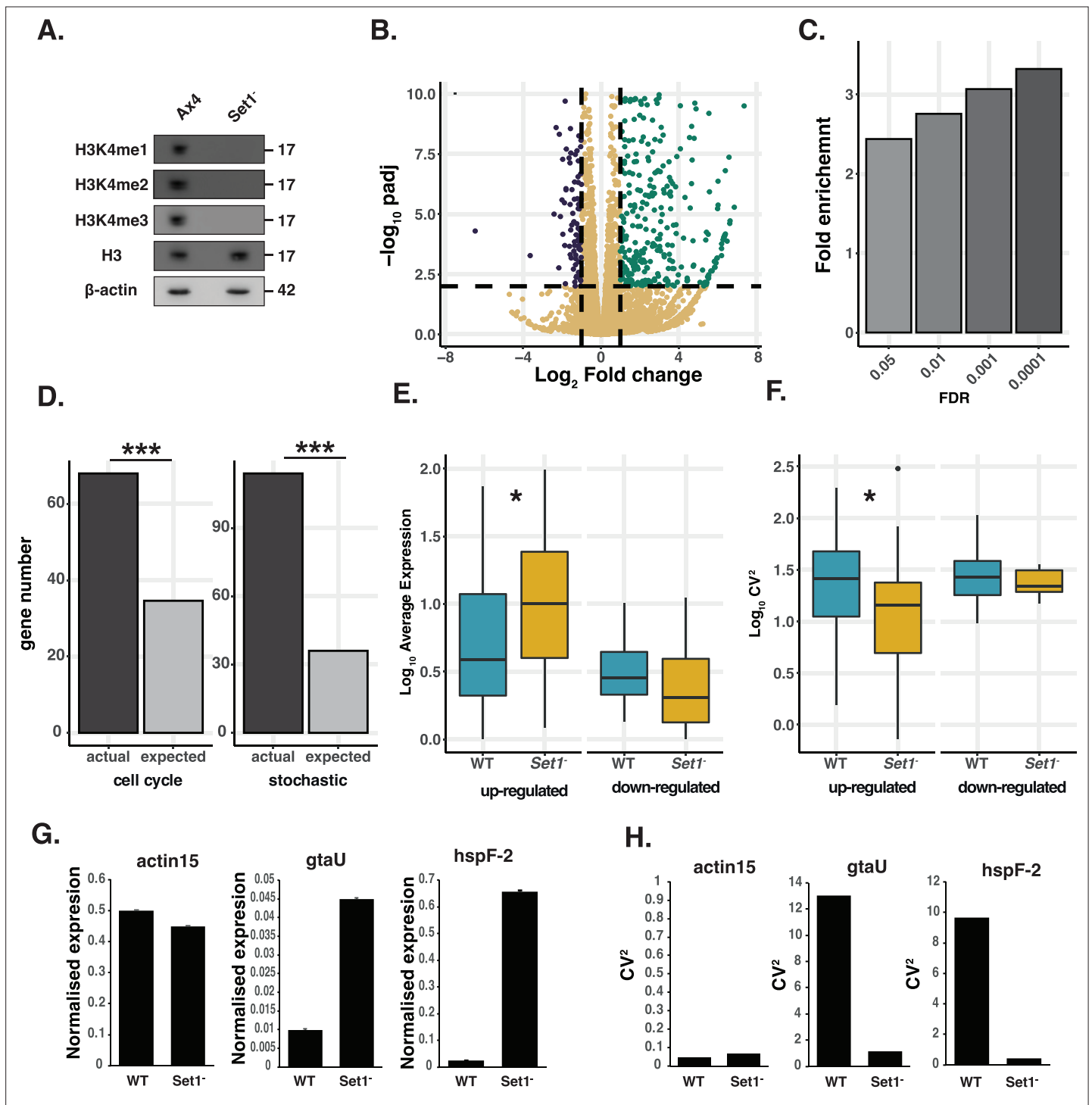


**Figure 4—figure supplement 1.** Normalisation of expression levels of variable and non-variable genes for ChIP analyses. **(A)** Distribution of expression levels of variable and non-variable genes. The average expression of variable genes and non-variable genes (limited to the lowest and highest expressed 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

Figure 4—figure supplement 1 continued on next page

*Figure 4—figure supplement 1 continued*

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 -2500 bp upstream to 2500 bp downstream of the TSS was calculated. This number was calculated for variable genes and compared to the average of 10 samples of non-variable genes.

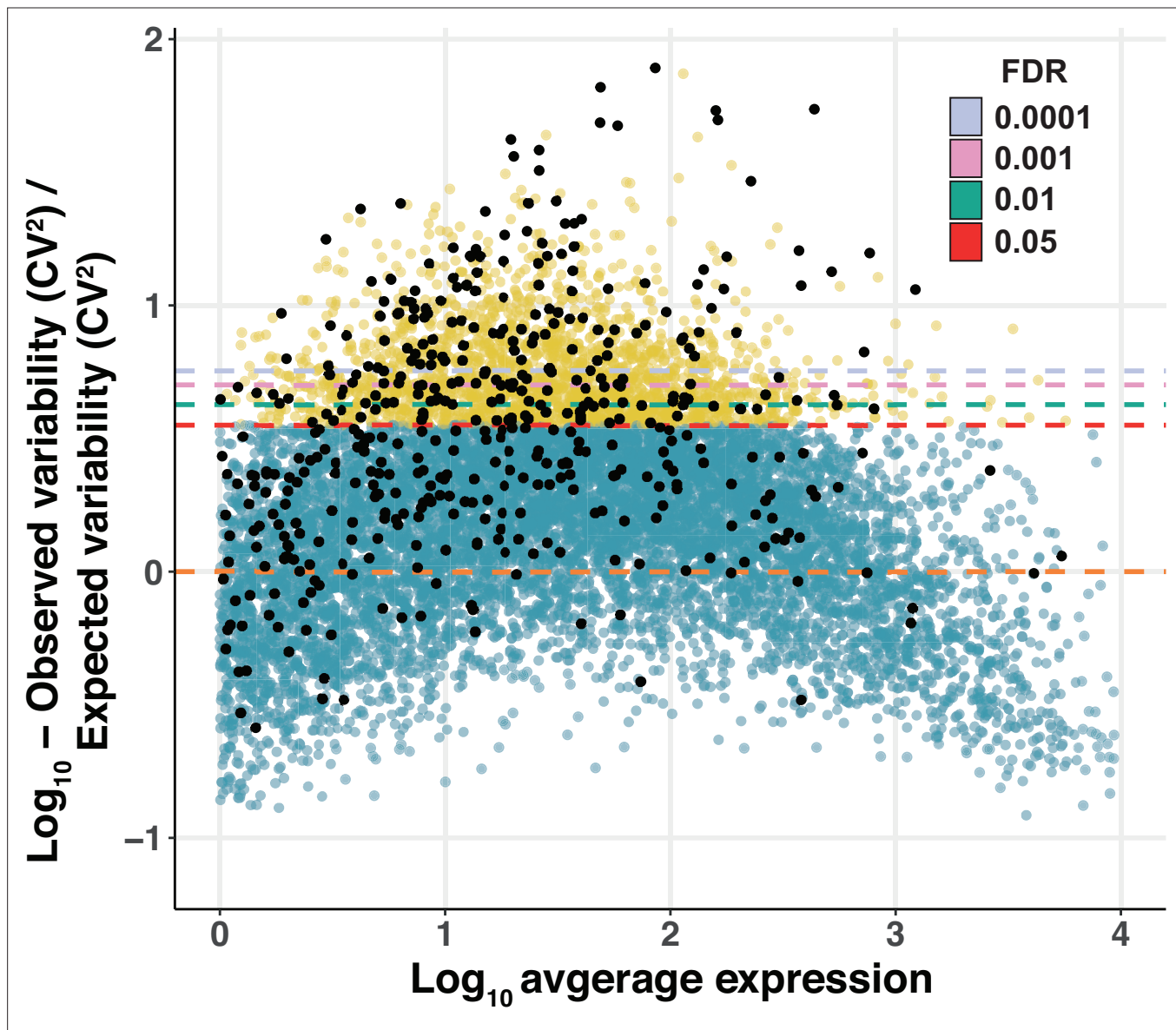


**Figure 5.** Set1 controls cell-cell gene expression variation. (A). Set1 is required for H3K4 methylation. Western blot of mono, di, and tri methylation at position H3K4 in wild-type Ax4 and mutant Set1<sup>-</sup> cells. Disruption of Set1 leads to the complete loss of all three methylation marks. Histone 3 levels are unaffected. (B). Deletion of Set1 leads to specific changes in gene expression. Volcano plot showing differential gene expression between Set1<sup>-</sup> 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 repressed genes in variable gene populations. As the threshold for the definition of a variable genes increases, the enrichment for Set1 repressed genes also increases. (D). Set1-dependent variable genes are enriched for cell cycle-associated and cell cycle independent stochastically variable genes. Plot shows the number of cell cycle-associated and cell cycle independent stochastically variable genes that are affected by Set1 disruption. This is compared to a random sample of the same number of genes from the genome. \*\*\* binomial test, p<0.01. (E). Variable genes show increased expression in Set1<sup>-</sup> single cells. Expression

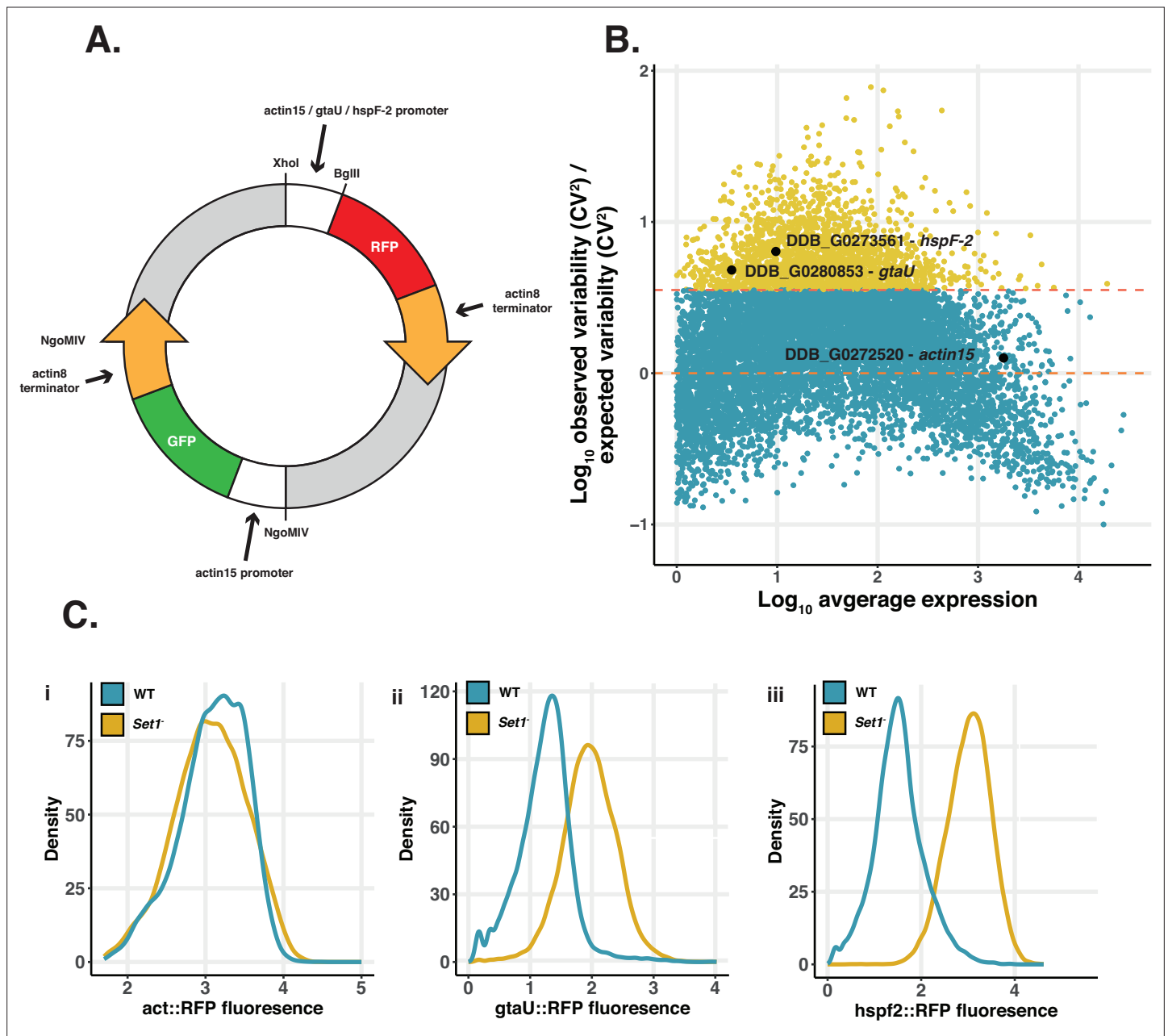
Figure 5 continued on next page

## Figure 5 continued

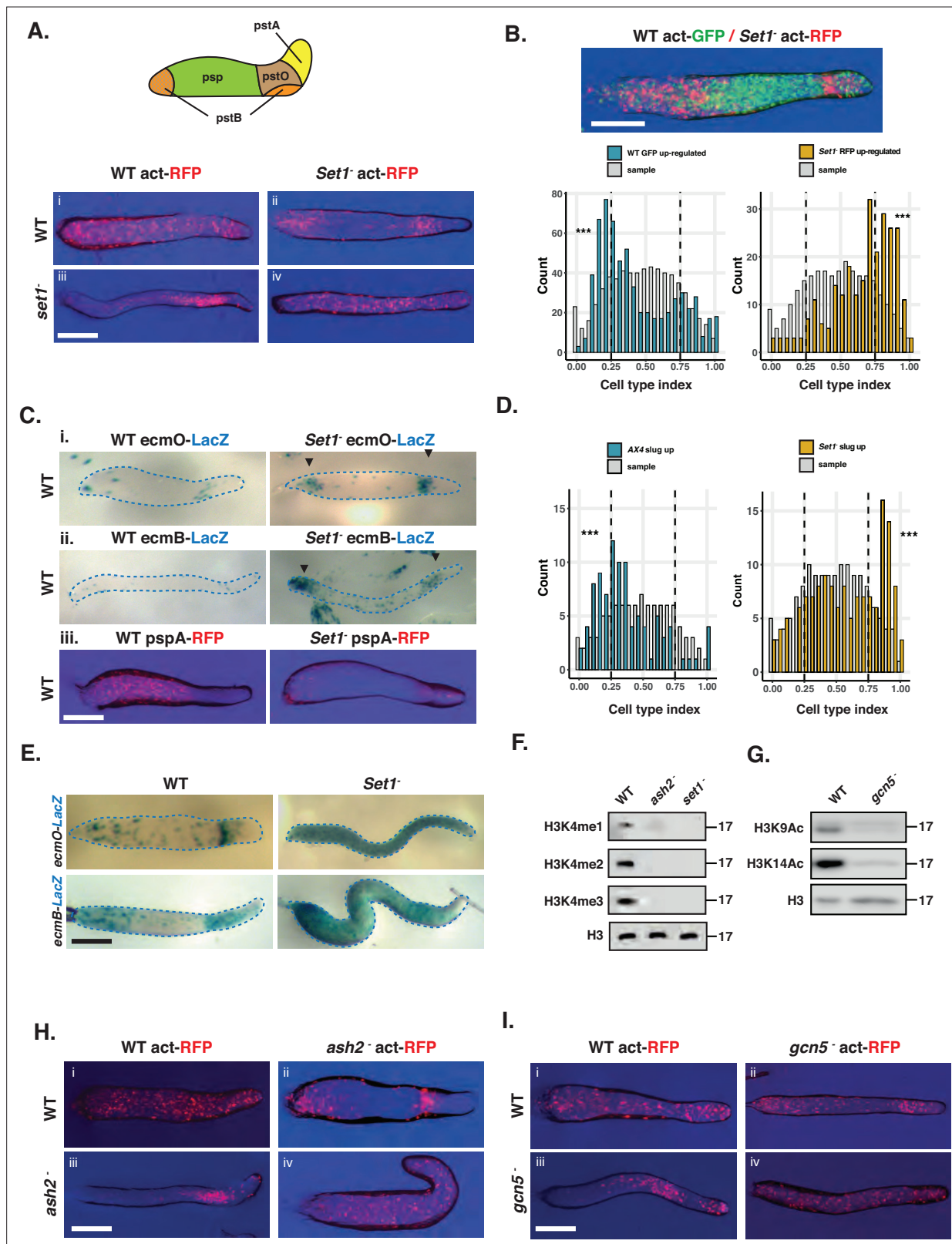
level was calculated from 310 wild-type and 310 *Set1*<sup>-</sup> cells. Variable genes (FDR <0.05) were defined as those up- or down-regulated in the bulk *Set1* mutant sequencing. Upregulated genes show a statistically significant increase ( $p < 0.01$ ) in expression. Downregulated genes show a small non-significant decrease in expression (**F**). Variable genes show decreased expression variability in *Set1*<sup>-</sup> single cells.  $CV^2$  was calculated from 310 wild-type and 310 *Set1*<sup>-</sup> 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*<sup>-</sup> 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*<sup>-</sup> single cells. Cell-cell variability in normalised expression ( $CV^2$ ) from FACS analysis of *GtaU* and *HspF-2* reporter genes.



**Figure 5—figure supplement 1.** Set1-dependent genes are more variable than expected. Plot of gene expression variability ( $\text{CV}^2$ ) against  $\text{log}_{10}$  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.



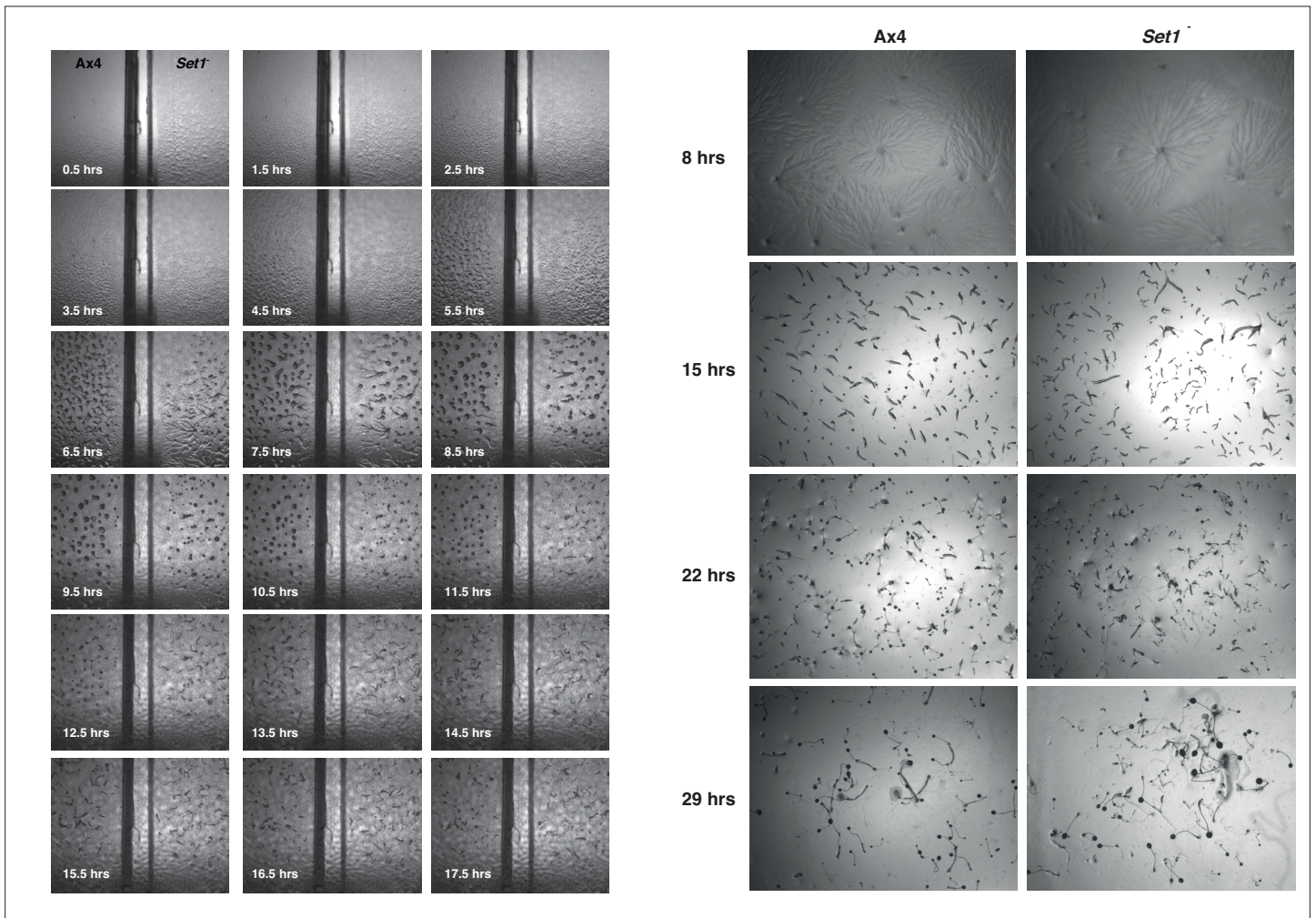
**Figure 5—figure supplement 2.** Dual reporter gene analysis of gene expression variation. (A) Schematic of dual reporter plasmid. Promoters of interest were cloned upstream of RFP in the pDM324 vector which also carries an *actin15*-promoter-GFP-*actin8*-terminator cassette. (B) scRNA-seq  $CV^2$  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 =  $5.2 \times 10^{-5}$ ) 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 *hspF-2* (iii) promoters in wild-type (blue) and *Set1* mutant (yellow) cells.



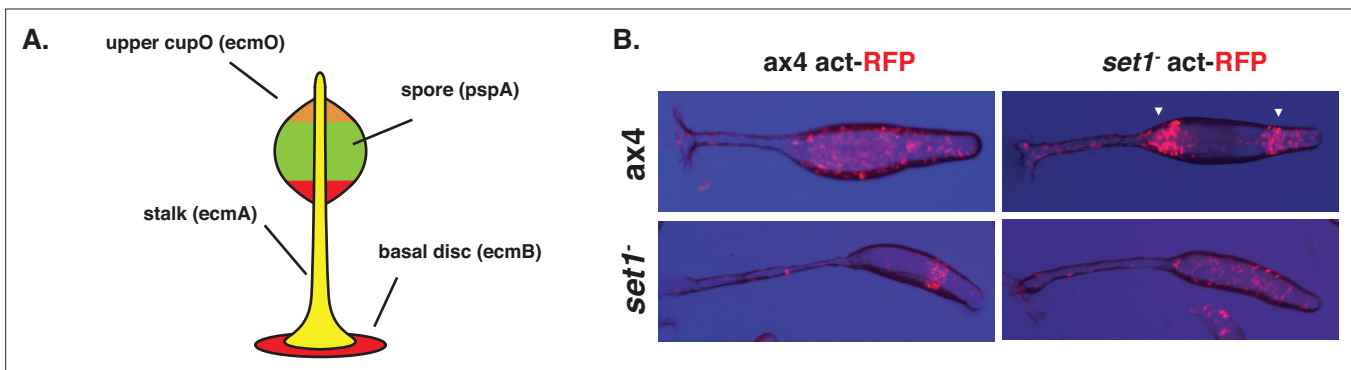
**Figure 6.** *Set1* is required for cell fate choice. (A) *set1*<sup>-</sup> mutant cells tend to occupy regions of the slug associated with prestalk cell differentiation in chimera with wild-type. Schematic showing location of prestalk (pstA, pstO, pstB) and prespore (psp) cell types in the slug. Chimeric development of wild-type and *set1*<sup>-</sup> mutant cells. Genotypes of 90% unlabelled cells are shown in rows and genotypes of 10% actin-RFP expressing cells shown in columns. *Set1*<sup>-</sup> mutant cells sort to the collar and back of slugs when mixed with wild-type cells. Scale bar = 0.5 mm. (B) RNA-seq of FACS purified *Set1* cells. (C) Chimeric development of WT and *Set1*<sup>-</sup> mutant cells. (D) RNA-seq of FACS purified *Set1* cells. (E) Chimeric development of WT and *Set1*<sup>-</sup> mutant cells. (F) Immunoblot showing H3K4me1, H3K4me2, H3K4me3, and H3 levels in WT, *ash2*<sup>-</sup>, and *set1*<sup>-</sup> mutant cells. (G) Immunoblot showing H3K9Ac, H3K14Ac, and H3 levels in WT and *gcn5*<sup>-</sup> mutant cells. (H) Chimeric development of WT and *ash2*<sup>-</sup> mutant cells. (I) Chimeric development of WT and *gcn5*<sup>-</sup> mutant cells. Figure 6 continued on next page

## Figure 6 continued

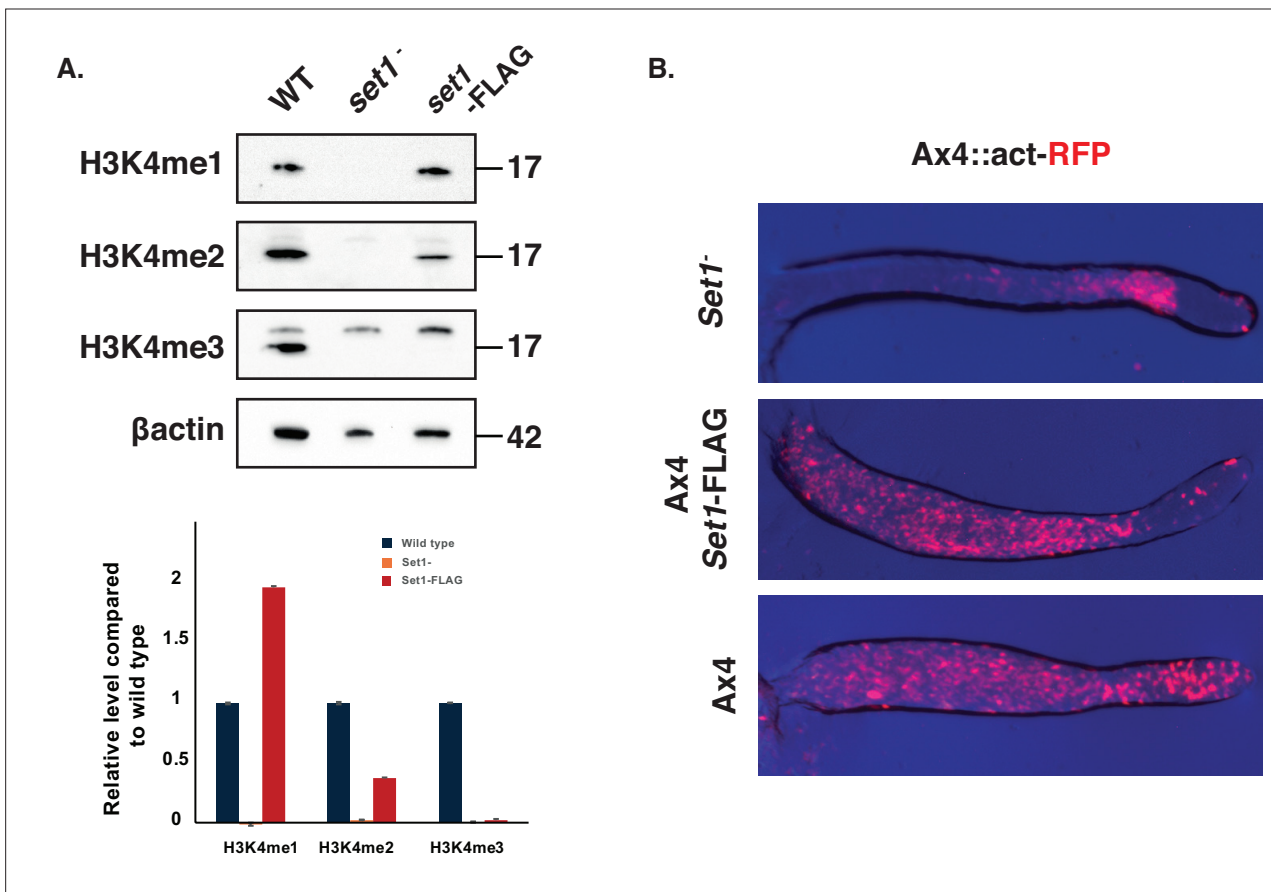
mutant and wild-type cells. Actin15-GFP expressing wild-type cells were mixed with actin15-RFP expressing *Set1*<sup>-</sup> mutant cells in a 50:50 ratio. GFP and RFP cells were separated at the slug stage and subjected to RNA-seq. The cell type index (prestalk expression/(prestalk + prespore expression)) 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 gray) shows the average distribution of 1000 random samples of the same numbers of genes. Scale bar = 0.5 mm. **(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 = 0.5 mm. **(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*<sup>-</sup> (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 = 0.5 mm. **(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*<sup>-</sup> cells, when labelled *ash2*<sup>-</sup> 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 = 0.5 mm. **(I)** Chimeric development of *Gcn5* mutant cells. Like *Set1*<sup>-</sup> and *Ash2* when labelled *Gcn5*<sup>-</sup> cells are mixed with WT cells they are localised to the collar and back of the slug. Scale bar = 0.5 mm.



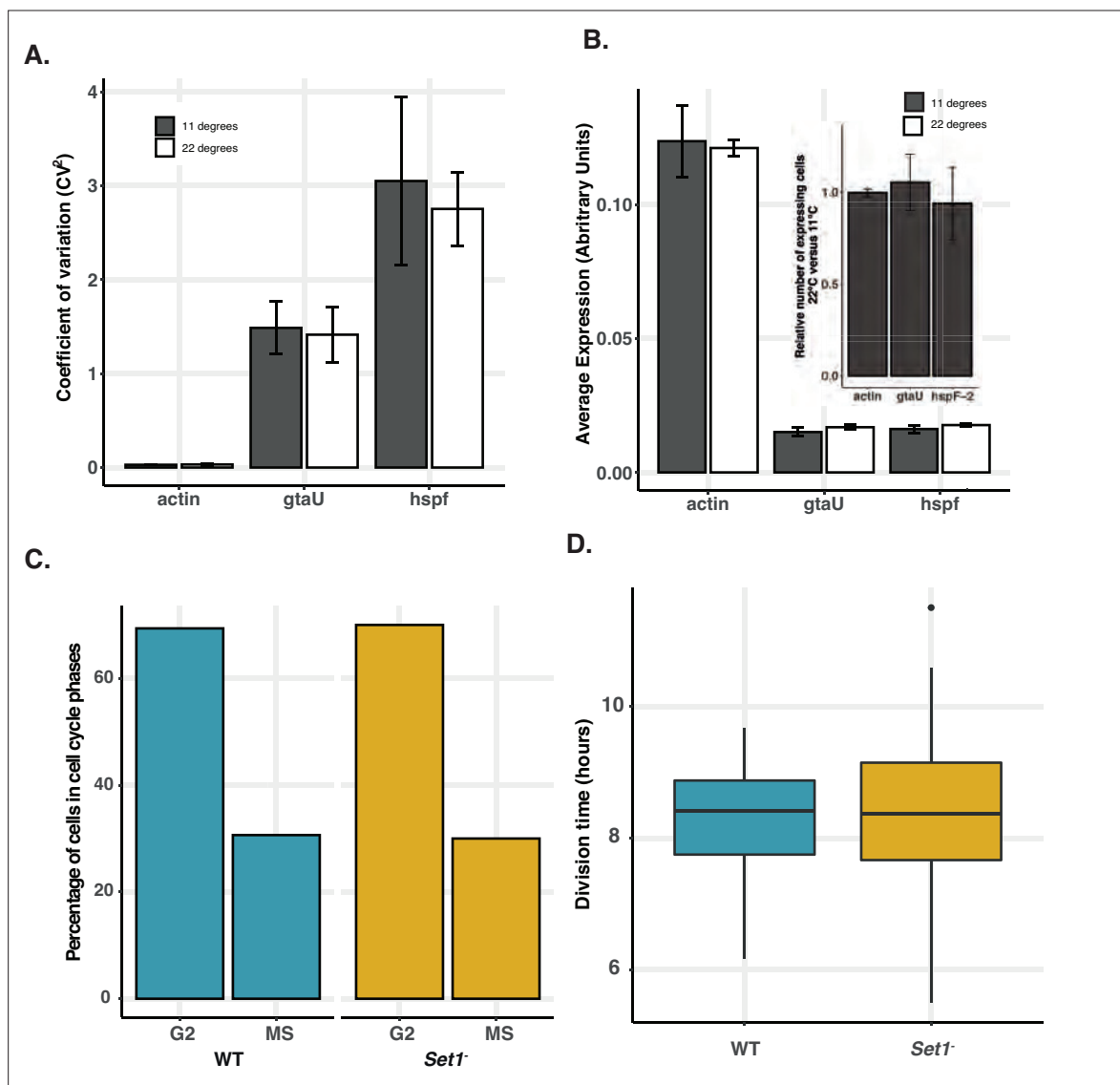
**Figure 6—figure supplement 1.** Disruption of *set1* does not affect developmental timing but results in slightly aberrant slug morphology. Representative images of wild-type and *set1*<sup>-</sup> mutant development.



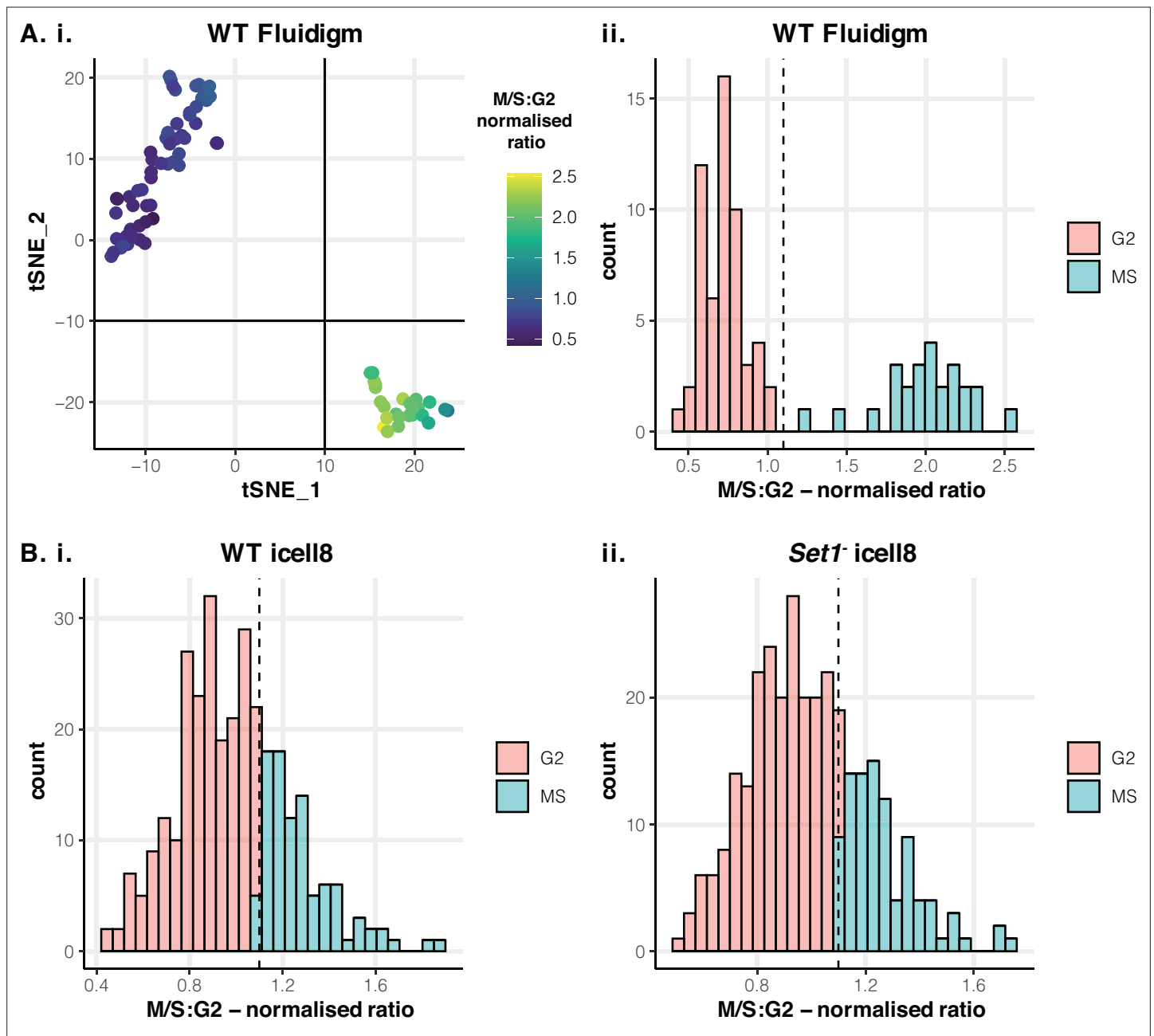
**Figure 6—figure supplement 2.** *Set1*<sup>-</sup> mutant cells sort to upper and lower cup of fruiting bodies. (A) Schematic of cell type differentiation in the *D. discoideum* fruiting body. (B) Sorting pattern of wild-type and *Set1* mutant cells in chimeric fruiting bodies. *Set1* mutant cells sort to the upper cup, lower cup, and stalk. Wild-type cells sort to spore region of the fruiting body.



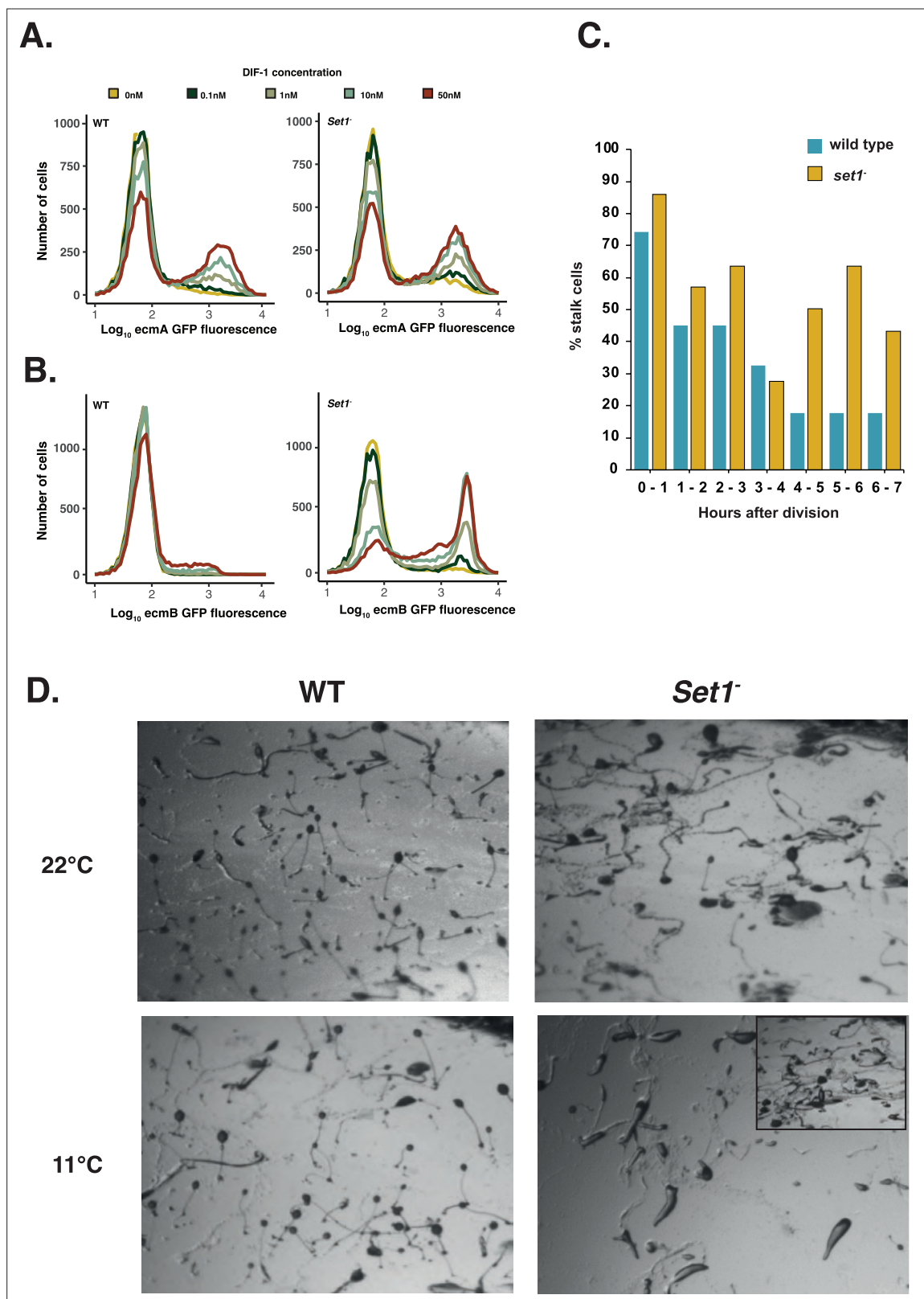
**Figure 6—figure supplement 3.** A hypomorphic *Set1* mutant exhibits a weaker developmental phenotype than knockout cells. **(A)** Western blot of H3K4 methylation. Blot of wild-type and *Set1*-FLAG Knock-in cells for different H3K4me marks reveals that tagged cells exhibit a complete loss of H3K4me3 (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*<sup>-</sup> 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*<sup>-</sup> mutant cells.



**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 **Figure 7—figure supplement 1** 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 7—figure supplement 1.** 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<sup>31</sup>. 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*<sup>-</sup> 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 to define cells from wild-type **(i)** or the *Set1* mutant **(ii)** as either M/S or G2.

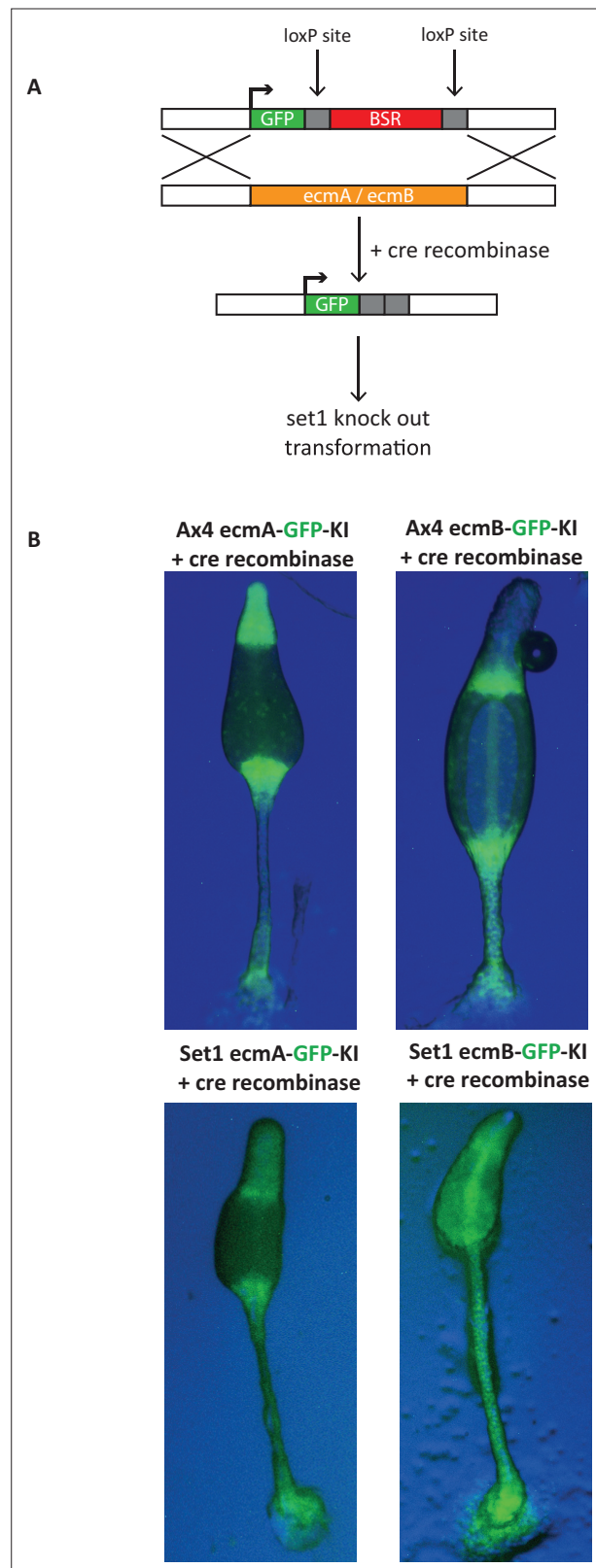


**Figure 8.** Cell type proportioning and responsiveness to differentiation cues depends on both stochastic and cell cycle variation. **(A and B)** *Set 1* regulates the threshold of responsiveness to differentiation cues. *ecmA*-GFP **(A)** and *ecmB*-GFP **(B)** knock-in lines were generated in wild-type and *Set1<sup>-</sup>* 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 log<sub>10</sub> GFP expression). **(C)** Cell type differentiation at different cell cycle

Figure 8 continued on next page

*Figure 8 continued*

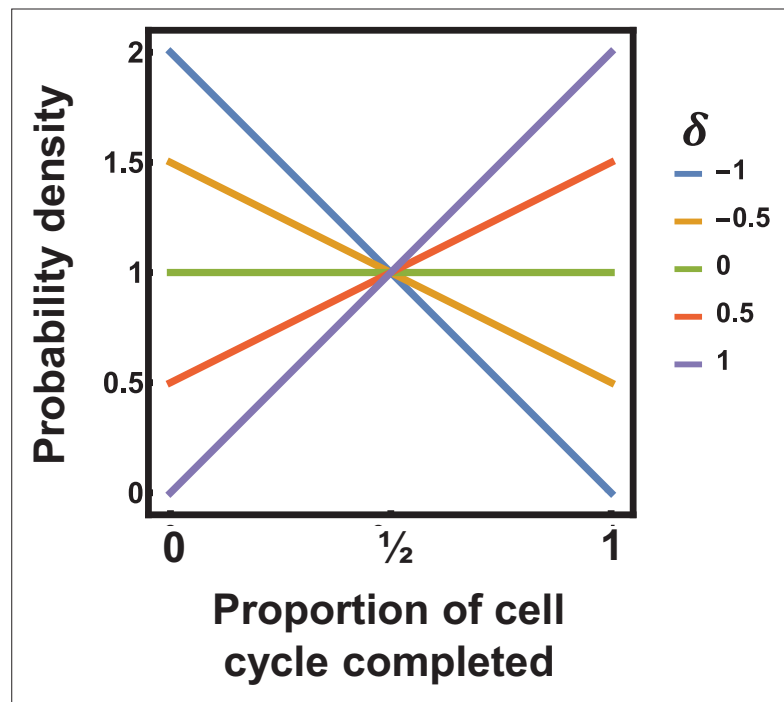
positions. Cells were grouped into hourly blocks according to the time of the last cell division before the addition of differentiation inducing signals. The percentage of cells that differentiate as prestalk cells was calculated. Results show the average of eight movies. **(D)** Development is blocked when stochastic and cell cycle gene expression is altered simultaneously. Wild-type (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. Scale bar = 1 mm.



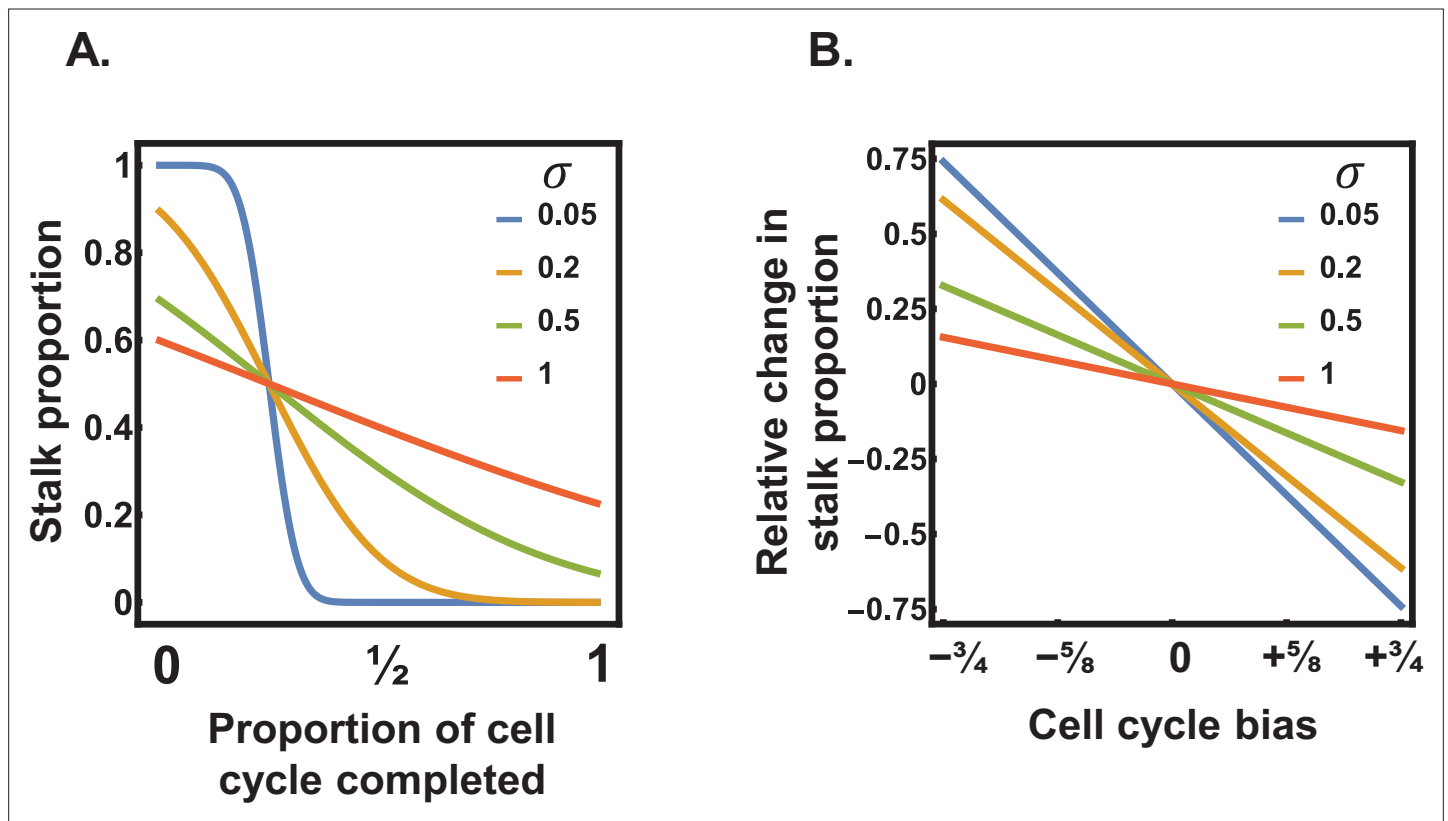
**Figure 8—figure supplement 1.** Generation of ecmA and ecmB-GFP knock-in strains. **(A)** Schematic of the construction of ecmA/B-GFP knock-in strains. A plasmid carrying GFP and a blasticidin resistance cassette flanked by loxP sites was integrated into the genome of wild-type or Set1 mutant cells through homologous recombination at either the ecmA or ecmB locus. After selection, the blasticidin cassette was removed by expression of Cre-  
 Figure 8—figure supplement 1 continued on next page

Figure 8—figure supplement 1 continued

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.



**Figure 8—figure supplement 2.** Probability density functions for biased sampling from the cell cycle. The plot illustrates the probability density functions (PDFs) associated with biased sampling from the cell cycle. The proportion of time through the cell cycle is given by  $t$  (e.g.  $t = 0.5$  means that cells are halfway through the cell cycle). The degree of biased sampling is given by  $\delta$ , which ranges from  $-1$  (bias towards the first half of the cell cycle) and  $+1$  (bias towards the second half of the cell cycle), where  $\delta = 0$  returns the uniform PDF. The influence of such bias on the PDF for sampling from the cell cycle is given by the equation  $Q_t = 2\delta t - \delta$  (where  $Q_t$  gives the relative change in probability density caused by sampling bias), which results in a PDF for sampling from the cell cycle given by  $N(t) = f(t) + Q_t f(t)$  (where  $f(t)$  corresponds to the probability density from the uniform distribution).



**Figure 8—figure supplement 3.** 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 **Equation 3**) with different levels of stochastic variability ( $\sigma$ , 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 (**Figure 1Aii**), 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  $\sigma$ ) 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  $\delta$ ; see Appendix 1). 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  $-3/4$  indicates that  $3/4$  of cells would be in the first half of the cell cycle instead of expectation of  $1/2$  for uniform sampling, meaning it represents an oversampling of  $1/4$  in the first half of the cell cycle and an under-sampling of  $1/4$  in the second half of the cycle). The relative change in stalk proportioning was calculated by taking the difference between the expected stalk proportioning (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.