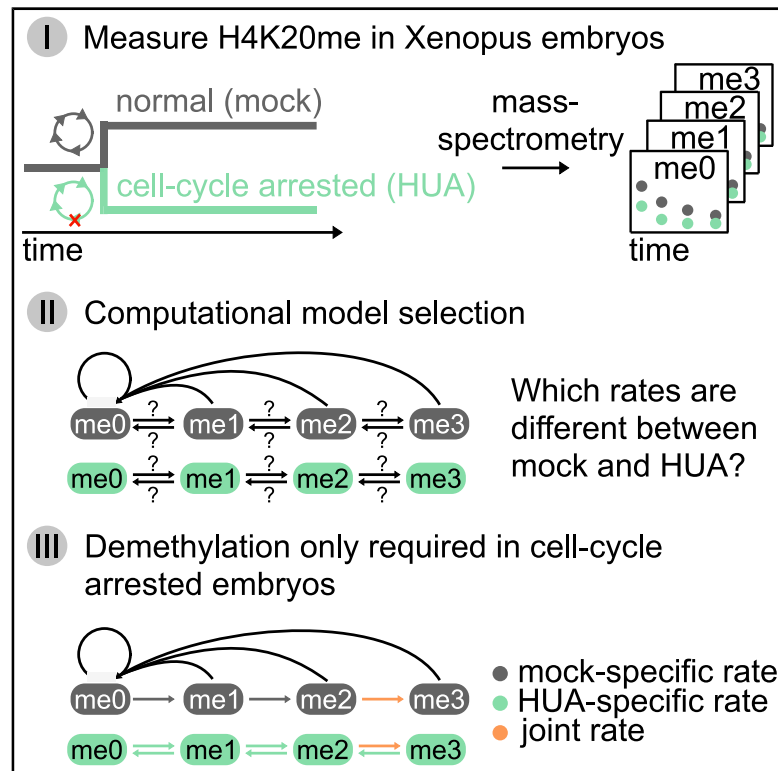


H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested *Xenopus* Embryos

Graphical Abstract



Authors

Lea Schuh, Carolin Loos, Daniil Pokrovsky, Axel Imhof, Ralph A.W. Rupp, Carsten Marr

Correspondence

carsten.marr@helmholtz-muenchen.de

In Brief

Schuh et al. introduce a computational model to describe H4K20me kinetics in normal and cell-cycle-arrested *Xenopus* embryos. This quantitative model invokes specific methylation and unspecific demethylation and correctly predicts cell-cycle durations and cell-cycle dependencies. Active demethylation is not required to explain H4K20me kinetics of cycling cells, suggesting that overall H4K20me dilution through DNA replication is dominant. So only once cells stop cycling during embryogenesis, active H4K20 demethylation may contribute to shape histone methylation.

Highlights

- Computational model explains H4K20me kinetics in cycling and arrested embryos
- Specific methylation rates are necessary—unspecific demethylation seems sufficient
- Active demethylation is required in cell-cycle-arrested embryos only
- Cell cycle may actively shape the H4K20me landscape during embryogenesis *in vivo*



Report

H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested *Xenopus* Embryos

Lea Schuh,^{1,2} Carolin Loos,^{1,2,3,4} Daniil Pokrovsky,⁵ Axel Imhof,⁵ Ralph A.W. Rupp,⁵ and Carsten Marr^{1,6,*}

¹Institute of Computational Biology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg 85764, Germany

²Department of Mathematics, Technical University of Munich, Garching 85748, Germany

³Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA

⁴Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁵Department of Molecular Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, 82152, Germany

⁶Lead Contact

*Correspondence: carsten.marr@helmholtz-muenchen.de

<https://doi.org/10.1016/j.cels.2020.11.003>

SUMMARY

DNA replication during cell division leads to dilution of histone modifications and can thus affect chromatin-mediated gene regulation, raising the question of how the cell-cycle shapes the histone modification landscape, particularly during embryogenesis. We tackled this problem by manipulating the cell cycle during early *Xenopus laevis* embryogenesis and analyzing *in vivo* histone H4K20 methylation kinetics. The global distribution of un-, mono-, di-, and tri-methylated histone H4K20 was measured by mass spectrometry in normal and cell-cycle-arrested embryos over time. Using multi-start maximum likelihood optimization and quantitative model selection, we found that three specific biological methylation rate constants were required to explain the measured H4K20 methylation state kinetics. While demethylation is essential for regulating H4K20 methylation kinetics in non-cycling cells, demethylation is very likely dispensable in rapidly dividing cells of early embryos, suggesting that cell-cycle-mediated dilution of H4K20 methylation is an essential regulatory component for shaping its epigenetic landscape during early development. A record of this paper's transparent peer review process is included in the Supplemental Information.

INTRODUCTION

All cells in our body contain the same genetic information encoded in the DNA. However, we are constituted out of many different cell types all performing their own specialized functions. Chromatin, mainly composed of DNA and histone octamers (two copies of histone H2A, H2B, H3, and H4 each), is an instructive DNA scaffold that aids extracting cell-specific information for gene expression. Histone tails are subject to various post-translational modifications, such as methylation, acetylation, phosphorylation, and ubiquitination (Bannister and Kouzarides, 2011), which play a fundamental role in altering chromatin accessibility. Dynamic regulation of gene expression is central for executing cell internal programs (proliferation, differentiation, etc.) and reacting to cell external signals with an appropriate response. Particularly during development, where cells continuously divide and differentiate, a fast and economical control of gene expression is required. Histone modifications are believed to regulate the progression throughout development (Jambhekar et al., 2020). In *Xenopus laevis*, a model organism for developmental biology, stage-specific histone modifications have been observed during the transit

from pluripotent to differentiated states, a process called epigenome maturation (Schneider et al., 2011). However, cells divide rapidly during early development. With each cell cycle newly formed, largely unmodified histones are incorporated into the DNA leading to an overall dilution of most histone modifications (Jasencakova et al., 2010). How is the histone modification landscape shaped by the cell cycle *in vivo*?

Histone methylation is known to play important roles in many biological processes (Greer and Shi, 2012), and its deregulation is linked to cancer and aging in humans (Fraga et al., 2005; Klutstein et al., 2016). The methylation of lysine 20 on histone H4 (H4K20) is one of the most frequent lysine methylation sites observed in HeLa cells, mouse embryonic fibroblasts and several other cell types (Everitts et al., 2013; Leroy et al., 2013; Pesavento et al., 2008; Schotta et al., 2008). It is evolutionarily conserved from *Schizosaccharomyces pombe* to humans (Lachner et al., 2004) and is known to have a strong cell-cycle dependence. H4K20 occurs in four different states, un-, mono-, di-, and tri-methylation. Each methylation state plays a different functional role ranging from DNA-damage repair and chromatin condensation observed in fission yeast *Schizosaccharomyces pombe* (Sanders et al.,



2004), over transcriptional regulation shown in human T cells (Bar-ski et al., 2007) and *Xenopus* embryos (Nicetto et al., 2013), mitotic progression found in *Drosophila melanogaster* (Sakaguchi and Steward, 2007), to cell-cycle control (Schotta et al., 2008), and silencing of repetitive DNA and transposons observed in mouse models (Schotta et al., 2004) and *Xenopus* embryos (van Kruijsbergen et al., 2017). H4K20me is regulated by three methyltransferases: KMT5A (also known as PR-Set7) for mono-methylation, first identified in *Drosophila* (Fang et al., 2002; Nishioka et al., 2002; Xiao et al., 2005), and SUV4-20H1 and SUV4-20H2 for both di- and tri-methylation, first identified in mammalian cells (Schotta et al., 2004). Whether there is a specificity of SUV4-20H1/2 for di- or tri-methylation is still debated (Schotta et al., 2008). The level of mono-methyltransferase KMT5A is cell-cycle dependent, and its degradation in G1 phase leads to a decline of H4K20me1 in late G1 as observed in human cell lines and *Xenopus* egg extracts (Abbas et al., 2010; Centore et al., 2010; Zee et al., 2012). H4K20me1 reaches its lowest level in S phase while increasing in G2 phase and peaking during mitosis. Both H4K20me2 and H4K20me3 levels have also been found to be cell-cycle dependent in HeLa cells though in a less dramatic fashion (Pesavento et al., 2008). The cell-cycle-dependent presence of H4K20 methyltransferases allows H4K20me2 and H4K20me3 to be reestablished only after mitosis in the next cell cycle (Jørgensen et al., 2013). For demethylation, unspecific enzymes such as PHF8 have been observed in human cell lines (Feng et al., 2010), but their functional importance has recently been questioned (Alabert et al., 2020; Jørgensen et al., 2013; Re-verón-Gómez et al., 2018). It has even been suggested that the loss of histone mark H3K27me3 in mammalian cells may occur only by dilution during chromatin replication rather than by active removal (Jadhav et al., 2020). Finally, homologs of all H4K20-modifying enzymes are present in the *Xenopus* genome (Bowes et al., 2010).

To address the role of the cell cycle for epigenome maturation in *Xenopus* development, we have measured histone modification proportions in sibling embryo populations, which either proliferate or are arrested at the G1/S transition. Using quantitative mass spectrometry data for H4K20 we compared over 200 model hypotheses describing H4K20me kinetics in the cycling and cell-cycle-arrested population. With only a few assumptions, our computational model is able to explain H4K20me kinetics, retrieves correct cell-cycle durations and known cell-cycle dependencies of H4K20me. Furthermore, our approach allows us to estimate cell numbers over time and reveals the importance of three specific biological methylation rate constants and a shared biological demethylation rate constant, which is essential to establish the observed histone modification profile in the cell-cycle arrested but not required in the cycling population of *Xenopus* embryos.

RESULTS

Cell-Cycle Arrest Changes H4K20me Patterns during *Xenopus* Embryogenesis

After *in vitro* fertilization of a *Xenopus* oocyte, cells rapidly divide in a state of transcriptional quiescence up to 5.5 h post fertilization (hpf) (Heasman, 2006). Only then a regular zygotic cell cycle containing G1 and G2 phases is initiated (Newport and Kirschner,

1982). To identify how H4K20 methylation (H4K20me) is shaped by cell-cycle, we compared a population of normal *Xenopus* embryos (from now on called “mock”) with a cell-cycle-arrested population. For this, half of the embryos were continuously incubated with hydroxyurea/aphidicolin (from now on called “HUA”) from gastrulation onward (11 hpf). This treatment arrests cells at the G1/S boundary and is compatible with embryonic development (Harris and Hartenstein, 1991). HUA treatment applied before 11 hpf is lethal (Harris and Hartenstein, 1991; Pokrovsky et al., 2020). Correct and robust establishment of the cell-cycle arrest by HUA in the *Xenopus* embryos has been shown by Pokrovsky et al. (2020). Mass spectrometry measurements of H4K20me states, averaging over all cells in the embryos and all histones in the cells, were conducted at 14.75, 19.75, 27.5, and 40 hpf corresponding to late gastrula (NF13), neurula (NF18), tailbud (NF25), and tadpole (NF32) stages, respectively (Figure 1A). H4K20me proportions of mock and HUA showed significant differences across three biological replicates in all four H4K20me states (Figure 1B). In HUA-treated embryos, methylation accumulates in the di- and tri-methylation states in comparison to mock. Upon DNA replication, newly synthesized and unmodified histones are incorporated in mock, while in HUA, only little DNA replication takes place and hence only little incorporation of newly synthesized and unmodified histones occurs. All three biological replicates result in highly reproducible H4K20me proportions across all four developmental stages suggesting a high accuracy and quality of the mass spectrometry measurements.

Specific Methylation Rate Constants Are Necessary to Explain H4K20me in Mock Embryogenesis While Demethylation Is Not Essential

To identify how H4K20me kinetics are shaped by cell cycle, we defined models for untreated embryos (mock) and fitted them to the data (see STAR Methods). Mock models are composed of four H4K20me states corresponding to un- (me0), mono- (me1), di- (me2), and tri-methylated (me3) H4K20, allowing for successive methylation and demethylation with biological rate constants m_i and d_i , $i \in \{1,2,3\}$, respectively (see Figure 2A and STAR Methods for a detailed model description). For mock, where the cells undergo cell division, newly synthesized and unmethylated histones are incorporated into replicating DNA leading to a continuous dilution of methylated H4K20. Considering methylation proportions (defined as the frequency of a particular methylation state divided by the sum of all methylation states as measured by mass spectrometry), cell-cycle results in an overall increase of unmethylated H4K20 mediated by an outflow of H4K20me states with population growth rate $g(t) = \ln(2)/c(t)$, where $c(t)$ is the average cell-cycle duration c across all cells as a function of experiment time t (see Figure 2A and STAR Methods). Having measured average H4K20me proportions across whole *Xenopus* embryos, our cell-cycle function accordingly models average cell-cycle durations across all cells constituting the *Xenopus* embryos at the respective developmental stages. By considering average H4K20me proportions across asynchronous cell populations (Boterenbrood et al., 1983), we assume H4K20me dilution to occur continuously. The most general model is parameterized with six biological rate constants, where a biological rate constant is defined as the proportion of H4K20 in a particular methylation state being methylated/demethylated per hour (h^{-1}). Although no actual enzymatic rate

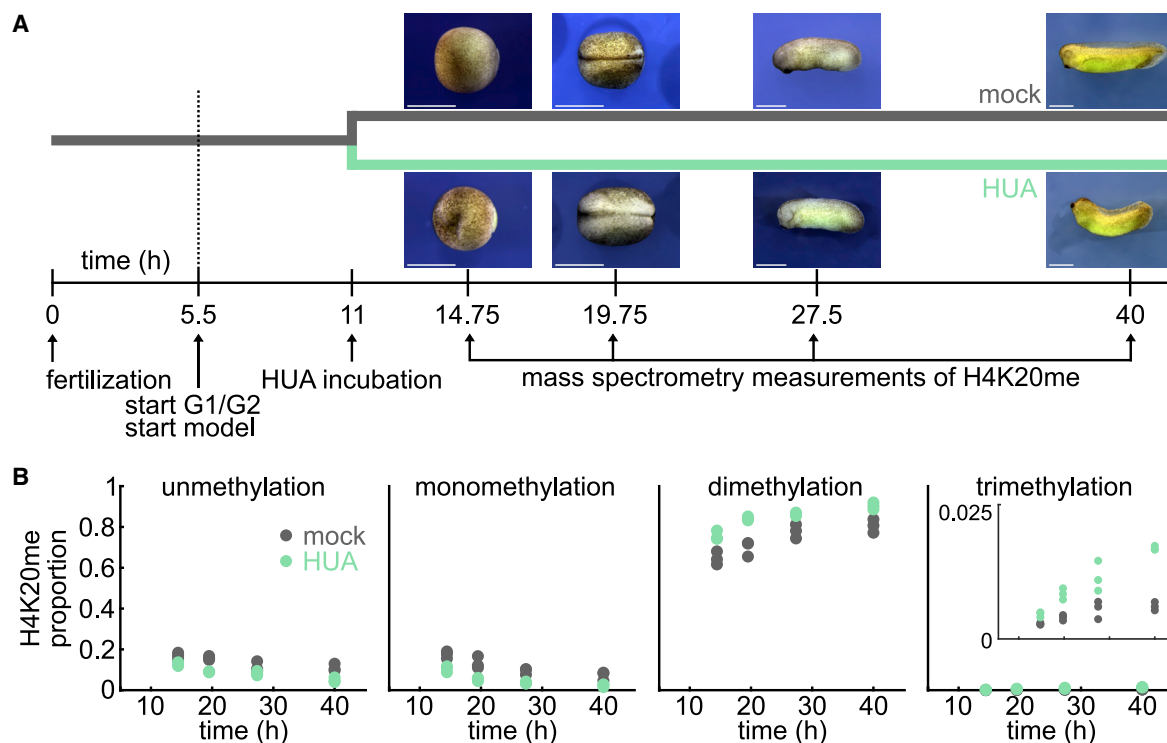


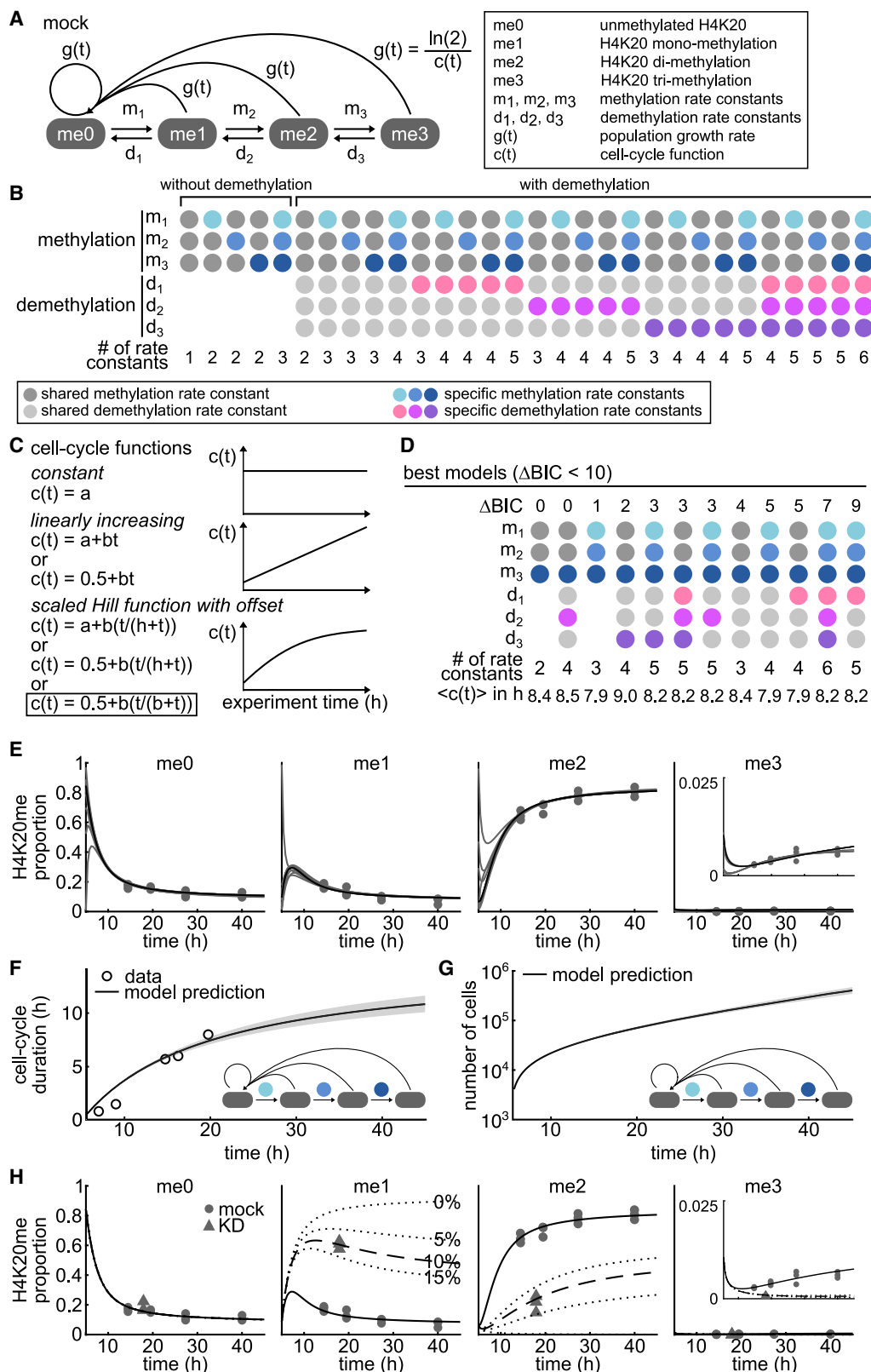
Figure 1. H4K20 Methylation Kinetics during *Xenopus* Embryogenesis Are Altered upon HUA-Induced Cell-Cycle Arrest

(A) *Xenopus* eggs are fertilized *in vitro* at time point 0. For the next 5 hpf, the embryonic cell-cycle consists of S and M phases only. At 5.5 hpf, G1 and G2 phases appear. At 11 hpf, half of the embryos are incubated with hydroxyurea/aphidicolin (HUA), arresting cells at the G1/S boundary. Mass spectrometry measurements of H4K20 methylation (H4K20me) are performed at 14.75, 19.75, 27.5, and 40 hpf in embryos with dividing (mock) or non-dividing cells (HUA). HUA incubated embryos are viable and visually remarkably similar to mock embryos (scale bar 1 mm).

(B) H4K20me kinetics differ significantly between mock (gray) and HUA treated (green) embryo populations (two-sample t test for all three biological replicates of mock and HUA for each time point resulted in p values < 0.05 for 15 out of 16 time points). In HUA H4K20 un- and mono-methylation is decreased while H4K20 di- and tri-methylation (see inset) is increased.

constants are derived we will refer to the biological rate constants as rate constants from here on. The most general model contains three rate constants for methylation m_1 , m_2 , and m_3 and three rate constants for demethylation d_1 , d_2 , and d_3 (Figure 2B, rightmost model). However, we also considered models with less parameters: rate constants shared between two or more reactions are termed “shared methylation/demethylation rate constants” (Figure 2B, gray) and rate constants specific to one reaction are termed “specific methylation/demethylation rate constants” (Figure 2B, colored). Intrigued by the question whether demethylation is important for methylation kinetics at all (as its existence was recently challenged at least for histone H3 lysine 27 tri-methylation; Reverón-Gómez et al., 2018), we also considered 5 models without demethylation. In total, the 30 models we consider comprise between 1 and 6 rate constants (Figure 2B; STAR Methods). In addition to the rate constants, we inferred another 4 model parameters: 3 initial H4K20me proportions at 5.5 hpf (denoted as me_{0_1} , me_{0_2} , me_{0_3} with $me_{0_1} = 1 - me_{0_2} - me_{0_3}$), and one noise parameter σ , determining the width of the Laplacian noise distribution (STAR Methods). As we were interested in H4K20me kinetics under the influence of the cell cycle, we started our mock model at 5.5 hpf (Figure 1A), when a regular zygotic cell-cycle with G1/G2 phases is initiated (Newport and Kirschner, 1982). Since cell cycle has been shown to vary substan-

tially with embryonic age, we considered 6 different cell-cycle functions $c(t)$ to model cell cycle over the experiment time t : constant, linearly increasing, or gradually plateauing (using a scaled Hill function with Hill coefficient 1 and offset) (Figure 2C). The number of model parameters for the cell-cycle functions varied from 1 (for the constant cell-cycle function) to 3 parameters (for the gradually plateauing cell-cycle function) (STAR Methods). We performed multi-start maximum likelihood optimization and model selection on 180 models (30 models times 6 different cell-cycle functions). Including prior biological knowledge about the short cell-cycle at 5.5 hpf of ~30 min (Anderson et al., 2017; Gelens et al., 2015), we found that only one of the six tested cell-cycle functions was able to predict a biologically meaningful average cell-cycle duration of around 8 h: a constrained scaled Hill function with Hill coefficient 1 and offset 0.5, $c(t) = 0.5 + b/(b + t)$ (Table S1). All models with other cell-cycle functions estimated average cell-cycle durations of at least 70 h. Using a constrained scaled Hill function, we found 12 models that outperformed other models with a BIC (Bayesian information criterion) difference of $\Delta BIC > 10$, which is considered to be an appropriate threshold for model rejection (Kass and Raftery, 1995) (Figure 2D). The two best models (with $\Delta BIC = 0$) show specificity in tri-methylation and shared rate constants for mono- and di-methylation. Overall, the best models with and without demethylation showed specificity



(legend on next page)

in either all three methylation rate constants or only in the tri-methylation rate constant. Varying numbers of demethylation rate constants were possible, ranging from 0 to 3. Fits to these 12 top models were able to capture the kinetics underlying H4K20me during mock embryogenesis (Figure 2E). Together, we found that either three specific methylation rate constants or one specific tri-methylation rate constant were necessary to explain the data from untreated embryos and that active demethylation was not required.

Validation of Mock Model by Comparing Cell-Cycle Durations to Experimental Data

We validated one of the best-performing models by comparing it to the average cell-cycle durations experimentally measured in *Xenopus* neural progenitors at various developmental stages (Graham and Morgan, 1966; Sabherwal et al., 2014; Thuret et al., 2015). We are aware that this comparison is drawn between the average cell-cycle durations of heterogeneous cell populations of *Xenopus* embryos and potentially more homogeneous cell populations of neural progenitors. However, to the best of our knowledge, this is the only available data on average cell-cycle durations during early *Xenopus* embryogenesis, which we could use for comparison. The cell-cycle durations from the mock model with three specific methylation rate constants but no demethylation (Figure 2F, inset) showed good agreement with measured cell-cycle durations (Figure 2F). Using this model, we can also predict the absolute number of cells within a normally developing embryo, which is experimentally challenging. For the same model (Figure 2G, inset), the number of cells was predicted to rise exponentially from roughly 20,000 cells after 10 h to 300,000 cells after 40 h (Figure 2G and STAR Methods). Similar results are obtained for the other best-performing mock models. Additionally, we predicted the effect of morpholino knockdowns of the di- and tri-methyltransferases SUV4-20H1/2 (KD) on H4K20me kinetics with the same model (Figure 2H). We found that a complete reduc-

tion of the di- and tri-methylation rate constants did not match the data perfectly. However, under the assumption that either the knockdown efficacy is not 100% or that there exist other enzymes performing di- and/or tri-methylation leading to a leaky reduction of the original di- and tri-methylation rate constants to 10%, the model is able to capture the perturbation.

Specific Methylation Rate Constants and Demethylation Are Necessary to Model H4K20me in HUA Embryogenesis

In contrast to mock, methylated H4K20 is not diluted in the cell-cycle-arrested HUA embryo population. We thus modeled HUA with the same set of reactions, however, without a cell-cycle function $g(t) = 0$ (Figure 3A). Similarly, to the mock model we performed multi-start maximum likelihood optimization and model selection on 30 HUA models with and without demethylation. We found that the five best-performing models (with $\Delta BIC < 10$) all required three specific methylation rate constants and demethylation (Figure 3B). The number of demethylation rate constants varied between 0 and 3 (Figure 3B). The single best-performing HUA model without demethylation (rightmost model in Figure 3B) was substantially outperformed by the HUA models with demethylation ($\Delta BIC = 13$), suggesting that demethylation was essential to explain the HUA data. The model fits of the five best HUA models were able to capture the kinetics underlying H4K20me during HUA embryogenesis (Figure 3C). Together, we found that three specific methylation rate constants were necessary to explain the HUA data and that demethylation was essential.

Joint Model Is Able to Retrieve Cell-Cycle Dependence of H4K20me and Finds Demethylation to Be Essential in HUA but Not Necessary in Mock

The models performing best in mock and HUA required three specific methylation rate constants and were indecisive about demethylation ranging from no demethylation over one shared to three

Figure 2. Demethylation Is Not Necessary to Explain Data of Cycling Mock Cells

- (A) Model of cycling mock population composed of four H4K20 states: un- (me0), mono- (me1), di- (me2), and tri-methylation (me3). m_1 , m_2 , and m_3 represent the mono-, di-, and tri-methylation rate constants and d_1 , d_2 , and d_3 represent the demethylation rate constants. An overall dilution of methylation happens due to cell division, parametrized with population growth rate $g(t)$, which is dependent on the cell-cycle function $c(t)$.
- (B) All possible parameter combinations result in 5 models without demethylation and 25 models with demethylation. Rate constants specific to a particular methylation or demethylation step are indicated in color, rate constants shared between methylation or demethylation steps are shown in gray. The number of rate constants ranges between 1 for the simplest model with no demethylation and shared methylation rate constant and 6 for the most complex model, where each methylation and demethylation rate constant is specific.
- (C) Only a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 gives an average cell-cycle duration in the expected range of 8 h (marked by the black box). All other cell-cycle functions $c(t)$ predicted average cell-cycle durations of at least 70 h, which is biologically not meaningful and reflects a population of non-cycling cells.
- (D) The 12 best-performing models are ordered by increasing BIC. All models with $\Delta BIC < 10$ require either three specific methylation rate constants (m_1 , m_2 , and m_3) or a specific tri-methylation rate constant. However, if present, demethylation may take on any of the 5 possible rate constant combinations. The best-performing models without demethylation perform similarly well as the best-performing models with demethylation ($\Delta BIC = 0$ and 1). The estimated average cell-cycle duration $\langle c(t) \rangle$ is in a biologically realistic range of around 8 h.
- (E) All 12 best-performing models fit the data. The model with three specific methylation rate constants but with no demethylation is shown in black.
- (F) Model prediction of the cell-cycle duration (median, 25th and 75th percentiles of MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but with no demethylation (inset)) agrees with experimental measurements of different papers.
- (G) The model with three specific methylation rate constants but with no demethylation (inset) predicts an increase of cell numbers from roughly 20,000 cells after 10 h to 300,000 cells after 40 h (using the median, 25th and 75th percentiles of the MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but with no demethylation (inset)) in a developing *Xenopus* embryo.
- (H) The model with three specific methylation rate constants but with no demethylation is able to predict the effects on H4K20me upon morpholino knockdowns of the di- and tri-methyltransferases SUV4-20H1/2 (KD) assuming a reduction to 10% of the original di- and tri-methylation rate constants. The dotted lines are the H4K20me kinetics predictions corresponding to 0%, 5%, and 15% of the original di- and tri-methylation rate constants. The solid line shows the previous fit with 100% of the original di- and tri-methylation rate constants.

See also Table S1.

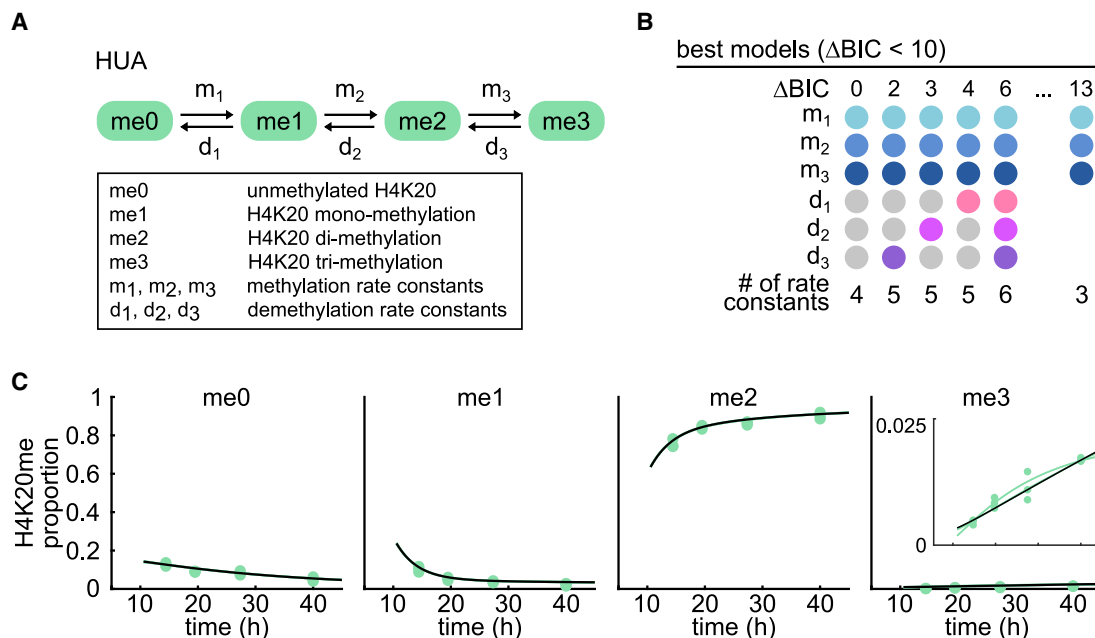


Figure 3. Demethylation Is Essential to Explain Data of Cell-Cycle-Arrested HUA Cells

(A) Model of cell-cycle-arrested HUA population. In contrast to the mock model (Figure 2A), the HUA cells do not divide ($g(t) = 0$), and no dilution of methylated H4K20 is required.

(B) The 5 best-performing HUA models with $\Delta\text{BIC} < 10$ all require 3 specific methylation rate constants (m_1 , m_2 , and m_3) and demethylation. However, demethylation may take on any of the 5 possible rate constant combinations. The single best-performing HUA model without demethylation (right) is outperformed by the HUA models with demethylation ($\Delta\text{BIC} = 13$).

(C) Model fits of top 5 HUA models with demethylation overlap strongly and show the ability to explain the HUA data. The best-performing model is highlighted in black.

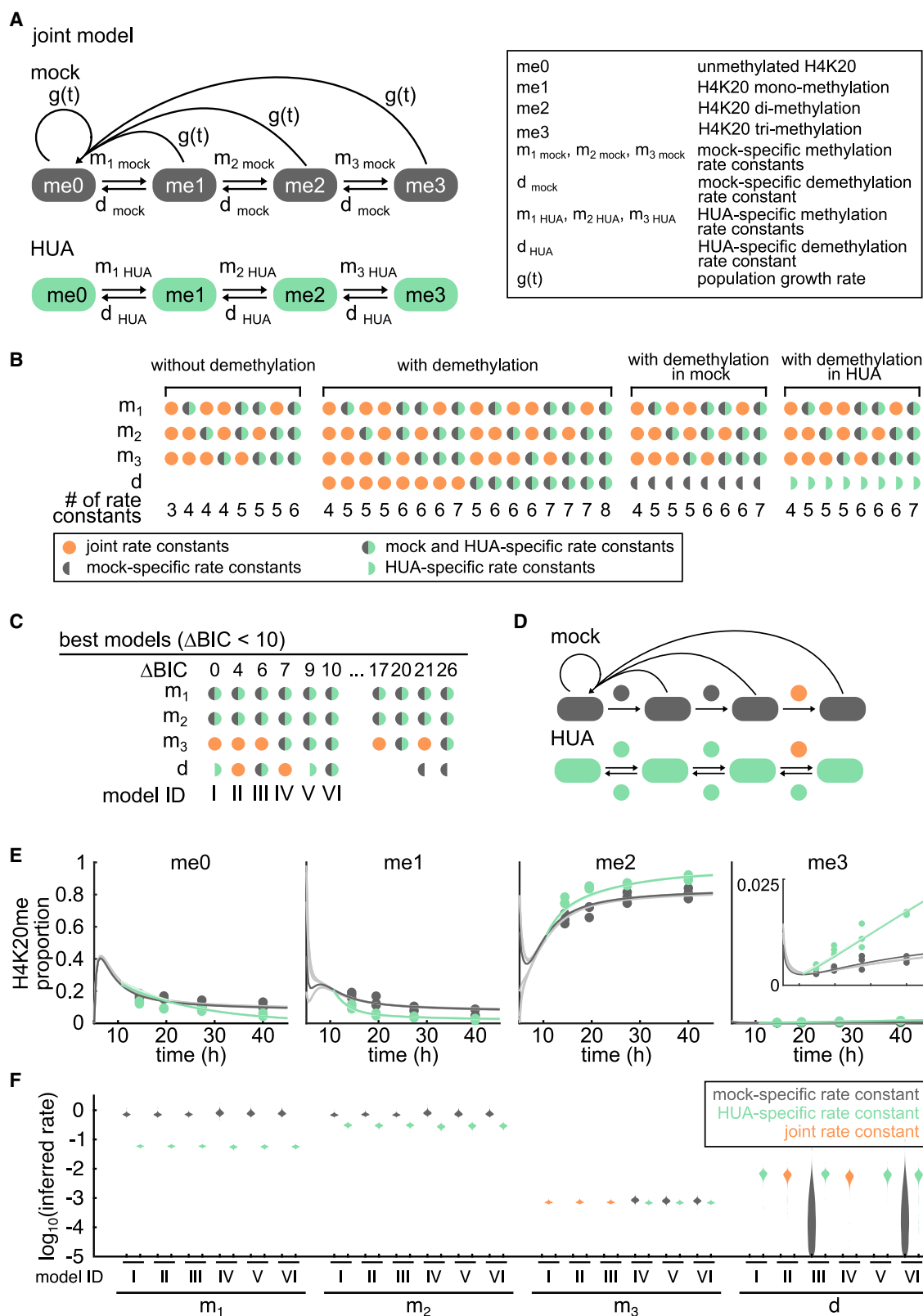
See also Table S1.

specific demethylation rate constants (Figures 2D and 3B). To determine which rates are substantially different between the two *Xenopus* populations we considered these findings and devised a joint model considering mock and HUA data simultaneously. For the most general hypothesis (Figure 4A), we allowed for three mock- and three HUA-specific methylation rate constants (visualized by a half gray and half green dot for m_1 , m_2 , or m_3 in Figure 4B). We also allowed for joint methylation rate constants shared between specific mock and HUA methylation steps (visualized as an orange dot for m_1 , m_2 , or m_3 in Figure 4B) reducing the number of parameters. As demethylation was not necessary to explain the mock data and one demethylation rate constant shared between methylation steps was sufficient for HUA, we here restricted demethylation to the simplest case of at most one shared demethylation rate constant d_{mock} and d_{HUA} (Figure 4A). We allowed for mock- and HUA-specific demethylation rate constants (visualized again by a half gray and half green dot for d in Figure 4B) or a joint demethylation rate constant for mock and HUA (visualized as an orange dot for d in Figure 4B). Furthermore, as a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 was the only function that led to biologically meaningful cell-cycle durations (see above and Figures 2C and 2F), we did not consider different cell-cycle functions thereby reducing the set of possible models to 8 joint models without demethylation (Figure 4B left), 16 joint models with demethylation and 2×8 models with demethylation in either mock or HUA (Figure 4B right). To identify joint models that are able to explain our measured data, we again fitted

the models using multi-start maximum likelihood optimization and model selection.

All 6 best-performing joint models ($\Delta\text{BIC} < 10$) required mock- and HUA-specific mono- and di-methylation rate constants (Figure 4C). However, they were not conclusive about tri-methylation and, if present, demethylation rate constants (Figure 4C). Specificity in one or more rate constants highlights that the differences in H4K20me proportions of mock and HUA are not explicable by the missing cell cycle alone but that the overall H4K20me kinetics are cell-cycle dependent. The model structure of the best-performing joint model (model I) is shown in Figure 4D. Joint models with demethylation in HUA only (models I and V in Figure 4C) performed just as well as joint models with demethylation in both HUA and mock (models II, III, IV, and VI in Figure 4C) while joint models without demethylation ($\Delta\text{BIC} = 17$ and 20) and joint models with demethylation in mock only ($\Delta\text{BIC} = 21$ and 26) were substantially outperformed. This suggested that demethylation was essential for HUA only, in accordance with the results from the separate models (Figures 2D and 3B).

The top 6 joint models (models I–VI in Figure 4C) showed good overall agreement with mock and HUA data (Figure 4E) and strongly consistent rate constants (Figure 4F). We determined the marginal distributions for all rate constants by Markov chain Monte Carlo (MCMC) sampling, where the credibility ranges are the 25th and 75th percentiles of the marginal distributions (see STAR Methods). We found strong discrepancies between mono- and di-methylation rate constants for mock and HUA, decreasing



(legend on next page)

10-fold and 2-fold, respectively (Figure 4F). The mock-specific mono- and di-methylation rate constants of the top 6 joint models had overlapping credibility ranges suggesting that for mock, a shared rate constant for mono- and di-methylation would suffice (Figure 4F). Similarly, mock- and HUA-specific tri-methylation rate constants show overlapping credibility ranges suggesting that a joint tri-methylation rate constant would suffice. In joint models with demethylation (models III and VI) we found the mock-specific demethylation rate constants to take on very small values while the HUA-specific demethylation rate constants were small but substantially larger: for model VI the median mock-specific demethylation rate constant was estimated to be 2.0×10^{-4} (with $0.5\text{--}8.3 \times 10^{-4}$ credibility range), while the HUA-specific demethylation rate constant was estimated to be 5.9×10^{-3} ($4.9\text{--}7.0 \times 10^{-3}$) (Figure 4F).

DISCUSSION

The joint demethylation rate constants (in models II and IV) were estimated to similar values as HUA-specific demethylation rate constants (the joint demethylation rate constant for model II and the HUA-specific demethylation rate constant for model VI were both estimated to be 5.9×10^{-3} [$4.9\text{--}7.0 \times 10^{-3}$]). This suggests that joint demethylation rate constants are overshadowed by the HUA model, strengthening the hypothesis that demethylation is not necessary in mock but essential in HUA. We would like to note that we here consider bulk H4K20me mass spectrometry data. However, demethylation might act highly localized and specific on only a few promoter nucleosomes with very important functions. Biologically, this could mean that while demethylases are present during embryogenesis, their effect in cycling cells is minute due to an overall dilution by unmodified histones. Only when cells stop to cycle (as modeled with the HUA treatment in our approach) demethylation kicks in and stabilizes post-translational modifications specifically, thereby potentially driving differentiation. To verify this experimentally, a knockdown of the H4K20 demethylases in HUA-treated embryos would be required where we hypothesize severe phenotypes due to the cell's incapability of reversing methylation. In contrast, such a knockdown should show little to no effect in untreated embryos as we hypothesize little to no active demethylation here. However, known H4K20 de-

methyltransferases e.g., PHF8, ROSBIN, and PHF2 are not specific to H4K20, and no global inhibitors of H4K20 demethylation are yet known. Hence, an experimental validation of our model predictions is currently not feasible due to technical limitations and therefore beyond the scope of this work.

Our findings can be interpreted in light of the current knowledge on methyltransferases. The mono-methyltransferase KMT5A (also known as PR-Set7) was found to be cell-cycle dependent, getting degraded by the proteasome in G1 phase (Abbas et al., 2010; Centore et al., 2010). In the absence of KMT5A, mono-methylation might be compensated by SUV4-20H1/2 but with lower activity (Southall et al., 2014; Yang et al., 2008). HUA treatment blocks the cell-cycle at the G1/S boundary, suggesting that none to little KMT5A is present in HUA to mono-methylate H4K20. This is reflected by a 10-fold decrease in the HUA-specific mono-methylation rate constant in all best-performing joint models (Figure 4F). As HUA-specific mono-methylation rate constants were necessary to explain the data (see Figure 4C), the joint model is able to retrieve this known cell-cycle dependence of H4K20me. H4K20me2 is also regulated in a cell-cycle dependent manner, however, peaking in G1 phase (Pesavento et al., 2008). In contrast, all best-performing joint models estimate the HUA-specific di-methylation rate constants to be decreased 2-fold in comparison to the mock-specific di-methylation (Figure 4F). We hypothesize this unexpected decrease of HUA-specific di-methylation to be due to either compensatory effects of SUV4-20H1/2, when the enzymes additionally mono-methylate H4K20, or so far unknown effects.

The separate model for mock identified either only tri-methylation or all three methylation steps to be specific (Figure 2D). The joint model reflects the same specificities regarding methylation in mock. Even though the joint model allows for specificity in all three methylation steps, the credibility ranges of mock-specific mono- and di-methylation rate constants in the joint models overlap (Figure 4F left). This suggests that a shared rate constant for mock mono- and di-methylation would suffice resulting in a mock model with specificity in tri-methylation only. However, in the joint models the HUA-specific mono- and di-methylation rate constants have non-overlapping credibility ranges with respect to the mock-specific rate constants (10-fold and 2-fold decrease) nor to each other. Under the assumption that the mock and HUA models are based on the same model structure, allowing for three

Figure 4. Joint Computational Modeling Allows Direct Comparisons between Mock and HUA Rate Constants and Reveals that Demethylation Is Overshadowed by HUA

- (A) Joint model allows for three methylation and one demethylation rate constants for both mock and HUA as suggested by the best models for mock and HUA. (B) We fit 16 models with demethylation and 8 models each for without demethylation in mock and/or HUA to the joint data to infer mock- and HUA-specific rate constants. The joint rate constants of mock and HUA are shown in orange, the rate constants present in both the mock and HUA models but taking on mock- and HUA-specific values are indicated in gray/green, the rate constants only present in the mock or HUA model are shown in gray and green half-circles, respectively. The model structure of the most complex of models is shown in (A). The number of rate constants ranges between 3 and 8. (C) The best-performing models on the combined dataset are ordered according to their BIC value. All models require HUA-specific mono- and di-methylation rate constants but are indecisive about tri-methylation and demethylation. Joint models where demethylation is present in either only HUA or in both mock and HUA perform equally well. Joint models where demethylation is not present in either only HUA or in both mock and HUA perform considerably worse. Model IDs of all considerably best-performing models are given (I–VI). (D) Model structure of the simplest best-performing joint model with demethylation in only HUA (model I). (E) All best-performing joint models are able to explain both the mock and HUA data. The estimated initial conditions vary between the models. Joint model I is highlighted. (F) The violin plots of the marginal distributions of all best-performing joint models show high consistency between the estimated methylation and demethylation rate constants. HUA-specific mono- and di-methylation rate constants are considerably decreased. Tri-methylation rate constants between mock and HUA have strongly overlapping marginal distributions. Demethylation seems to be dominated by the HUA population and is negligible in the mock population if a mock-specific demethylation rate is allowed. See also Table S1.

specific methylation rate constants in the joint model was thus necessary (Figure 4F) to resolve these differences.

All three joint models with specificity in tri-methylation (models IV, V, and VI) result in slightly lower BIC values (Figure 4C), which is likely due to an increased penalization term for an additional estimated parameter and not due to a decreased likelihood. The estimated tri-methylation rate constants are small (on the order of 10^{-3}) and the credibility ranges for mock- and HUA-specific tri-methylation overlap in all three joint models suggesting that a joint tri-methylation rate constant would suffice. When we interpret differences in HUA and mock rates as indications for cell-cycle dependence rates, we find no evidence for cell-cycle dependence for H4K20 tri-methylation. To clarify if the corresponding enzymes are indeed homogeneously expressed is up to further research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Embryos Handling and HUA Treatment
 - Nuclear Histone Extraction
 - Mass Spectrometry Sample Preparation
 - Mass Spectrometry Analysis
 - Histone Modifications Quantification
- METHOD DETAILS
 - Models
 - Noise Models
- OPTIMIZATION AND PARAMETER ESTIMATION
 - Model Selection
 - Parameter Uncertainty
 - Validation - Cell-Cycle Durations
 - Prediction of Number of Cells
 - Implementation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cels.2020.11.003>.

ACKNOWLEDGMENTS

L.S. was funded by the BMBF project TIDY (031L0170B). L.S. is especially grateful to the Technical University of Munich's Department of Mathematics, whose generous Entrepreneurial Award (within the Program "Global Challenges for Women in Math Science") contributed to the completion of this project. D.P., A.I., and R.R. were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID213249687 - SFB 1064. C.M. acknowledges funding from the European Research Council (ERC), grant agreement no. 866411.

AUTHOR CONTRIBUTIONS

Conceptualization, A.I., R.A.W.R., and C.M.; Methodology, L.S. and C.L.; Software, L.S. and C.L. Validation, C.L.; Formal Analysis, L.S.; Investigation, L.S.

and D.P.; Resources, A.I., R.A.W.R., and C.M.; Data Curation, D.P.; Writing - Original Draft, L.S.; Visualization, L.S.; Supervision, A.I., R.A.W.R., C.L., and C.M.; Project Administration, A.I., R.A.W.R., and C.M.; Funding Acquisition, A.I., R.A.W.R., and C.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 14, 2020

Revised: August 5, 2020

Accepted: November 11, 2020

Published: December 8, 2020

REFERENCES

- Abbas, T., Shibata, E., Park, J., Jha, S., Karnani, N., and Dutta, A. (2010). CRL4(Cdt2) regulates cell proliferation and histone gene expression by targeting PR-Set7/Set8 for degradation. *Mol. Cell* 40, 9–21.
- Alabert, C., Loos, C., Voelker-Albert, M., Graziano, S., Forné, I., Reveron-Gomez, N., Schuh, L., Hasenauer, J., Marr, C., Imhof, A., and Groth, A. (2020). Domain model explains propagation dynamics and stability of histone H3K27 and H3K36 methylation landscapes. *Cell Rep.* 30, 1223–1234.e8.
- Anderson, G.A., Gelens, L., Baker, J.C., and Ferrell, J.E., Jr. (2017). Desynchronizing embryonic cell division waves reveals the robustness of *Xenopus laevis* development. *Cell Rep.* 21, 37–46.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395.
- 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.
- Boterenbrood, E.C., Narraway, J.M., and Hara, K. (1983). Duration of cleavage cycles and asymmetry in the direction of cleavage waves prior to gastrulation in *Xenopus laevis*. *Wilhelm Roux Arch. Dev. Biol.* 192, 216–221.
- Bowes, J.B., Snyder, K.A., Segerdell, E., Jarabek, C.J., Azam, K., Zorn, A.M., and Vize, P.D. (2010). Xenbase: gene expression and improved integration. *Nucleic Acids Res.* 38, D607–D612.
- Centore, R.C., Havens, C.G., Manning, A.L., Li, J.M., Flynn, R.L., Tse, A., Jin, J., Dyson, N.J., Walter, J.C., and Zou, L. (2010). CRL4(Cdt2)-mediated destruction of the histone methyltransferase Set8 prevents premature chromatin compaction in S phase. *Mol. Cell* 40, 22–33.
- Everitts, A.G., Manning, A.L., Wang, X., Dyson, N.J., Garcia, B.A., and Collier, H.A. (2013). H4K20 methylation regulates quiescence and chromatin compaction. *Mol. Biol. Cell* 24, 3025–3037.
- Faber, J., and Nieuwkoop, P.D. (1994). Normal Table of *Xenopus laevis* (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorphosis (Garland Science).
- Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr. Biol.* 12, 1086–1099.
- Feng, W., Yonezawa, M., Ye, J., Jenuwein, T., and Grummt, I. (2010). PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. *Nat. Struct. Mol. Biol.* 17, 445–450.
- Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* 37, 391–400.
- Fröhlich, F., Kaltenbacher, B., Theis, F.J., and Hasenauer, J. (2017). Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Comput. Biol.* 13, e1005331.
- Gelens, L., Huang, K.C., and Ferrell, J.E., Jr. (2015). How does the *Xenopus laevis* embryonic cell cycle avoid spatial chaos? *Cell Rep.* 12, 892–900.
- Graham, C.F., and Morgan, R.W. (1966). Changes in the cell cycle during early amphibian development. *Dev. Biol.* 14, 439–460.

- Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357.
- Harris, W.A., and Hartenstein, V. (1991). Neuronal determination without cell division in *Xenopus* embryos. *Neuron* **6**, 499–515.
- Hass, H., Loos, C., Raimúndez-Álvarez, E., Timmer, J., Hasenauer, J., and Kreutz, C. (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics* **35**, 3073–3082.
- Heasman, J. (2006). Patterning the early *Xenopus* embryo. *Development* **133**, 1205–1217.
- Jadhav, U., Manieri, E., Nalapareddy, K., Madha, S., Chakrabarti, S., Wucherpfennig, K., Barefoot, M., and Shivdasani, R.A. (2020). Replicational dilution of H3K27me3 in mammalian cells and the role of poised promoters. *Mol. Cell* **78**, 141–151.e5.
- Jambhekar, A., Dhall, A., and Shi, Y. (2020). Author Correction: roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **21**, 59.
- Jasencakova, Z., Scharf, A.N.D., Ask, K., Corpet, A., Imhof, A., Almouzni, G., and Groth, A. (2010). Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol. Cell* **37**, 736–743.
- Jørgensen, S., Schotta, G., and Sørensen, C.S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res.* **41**, 2797–2806.
- Kass, R.E., and Raftery, A.E. (1995). Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795.
- Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA methylation in cancer and aging. *Cancer Res.* **76**, 3446–3450.
- Lachner, M., Sengupta, R., Schotta, G., and Jenuwein, T. (2004). Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. *Cold Spring Harbor Symp. Quant. Biol.* **69**, 209–218.
- Leroy, G., Dimaggio, P.A., Chan, E.Y., Zee, B.M., Blanco, M.A., Bryant, B., Flaniken, I.Z., Liu, S., Kang, Y., Trojer, P., and Garcia, B.A. (2013). A quantitative atlas of histone modification signatures from human cancer cells. *Epigenet. Chromatin* **6**, 20.
- Liebler, D.C., and Zimmerman, L.J. (2013). Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry* **52**, 3797–3806.
- MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968.
- Maier, C., Loos, C., and Hasenauer, J. (2017). Robust parameter estimation for dynamical systems from outlier-corrupted data. *Bioinformatics* **33**, 718–725.
- Newport, J., and Kirschner, M. (1982). A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686.
- Nicetto, D., Hahn, M., Jung, J., Schneider, T.D., Straub, T., David, R., Schotta, G., and Rupp, R.A.W. (2013). Suv4-20h histone methyltransferases promote neuroectodermal differentiation by silencing the pluripotency-associated Oct-25 gene. *PLoS Genet.* **9**, e1003188.
- Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuiikov, S., Valenzuela, P., Tempst, P., Steward, R., et al. (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201–1213.
- Pesavento, J.J., Yang, H., Kelleher, N.L., and Mizzen, C.A. (2008). Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol. Cell. Biol.* **28**, 468–486.
- Pokrovsky, D., Forne, I., Straub, T., Imhof, A., and Rupp, R. (2020). Mitotic activity shapes stage-specific histone modification profiles during *Xenopus* embryogenesis. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2020.08.04.200550v1?rss=1>.
- Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906.
- Reverón-Gómez, N., González-Aguilera, C., Stewart-Morgan, K.R., Petryk, N., Flury, V., Graziano, S., Johansen, J.V., Jakobsen, J.S., Alabert, C., and Groth, A. (2018). Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. *Mol. Cell* **72**, 239–249.e5.
- Sabherwal, N., Thuret, R., Lea, R., Stanley, P., and Papalopulu, N. (2014). aPKC phosphorylates p27Xic1, providing a mechanistic link between apico-basal polarity and cell-cycle control. *Dev. Cell* **31**, 559–571.
- Sakaguchi, A., and Steward, R. (2007). Aberrant monomethylation of histone H4 lysine 20 activates the DNA damage checkpoint in *Drosophila melanogaster*. *J. Cell Biol.* **176**, 155–162.
- Sanders, S.L., Portoso, M., Mata, J., Bähler, J., Allshire, R.C., and Kouzarides, T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603–614.
- Schneider, T.D., Arteaga-Salas, J.M., Mentale, E., David, R., Nicetto, D., Imhof, A., and Rupp, R.A.W. (2011). Stage-specific histone modification profiles reveal global transitions in the *Xenopus* embryonic epigenome. *PLoS One* **6**, e22548.
- Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262.
- Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callén, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., et al. (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* **22**, 2048–2061.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.
- Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early development of *Xenopus laevis*: a laboratory manual (Vancouver: CSHL Press).
- Southall, S.M., Cronin, N.B., and Wilson, J.R. (2014). A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. *Nucleic Acids Res.* **42**, 661–671.
- Stapor, P., Weindl, D., Ballnus, B., Hug, S., Loos, C., Fiedler, A., Krause, S., Hroß, S., Fröhlich, F., and Hasenauer, J. (2018). PESTO: parameter Estimation TOolbox. *Bioinformatics* **34**, 705–707.
- Thuret, R., Auger, H., and Papalopulu, N. (2015). Analysis of neural progenitors from embryogenesis to juvenile adult in *Xenopus laevis* reveals biphasic neurogenesis and continuous lengthening of the cell cycle. *Biol. Open* **4**, 1772–1781.
- van Kruisbergen, I., Hontelez, S., Elurbe, D.M., van Heeringen, S.J., Huynen, M.A., and Veenstra, G.J.C. (2017). Heterochromatic histone modifications at transposons in *Xenopus tropicalis* embryos. *Dev. Biol.* **426**, 460–471.
- van Nuland, R., and Gozani, O. (2016). Histone H4 lysine 20 (H4K20) methylation, expanding the signaling potential of the proteome one methyl moiety at a time. *Mol. Cell. Proteomics* **15**, 755–764.
- Villar-Garea, A., Forne, I., Vetter, I., Kremmer, E., Thomae, A., and Imhof, A. (2012). Developmental regulation of N-terminal H2B methylation in *Drosophila melanogaster*. *Nucleic Acids Res.* **40**, 1536–1549.
- Xiao, B., Jing, C., Kelly, G., Walker, P.A., Muskett, F.W., Frenkiel, T.A., Martin, S.R., Sarma, K., Reinberg, D., Gamblin, S.J., and Wilson, J.R. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev.* **19**, 1444–1454.
- Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. *J. Biol. Chem.* **283**, 12085–12092.
- Zee, B.M., Britton, L.M., Wolle, D., Haberman, D.M., and Garcia, B.A. (2012). Origins and formation of histone methylation across the human cell cycle. *Mol. Cell. Biol.* **32**, 2503–2514.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
H4K20me proportions	This work	https://doi.org/10.5281/zenodo.4046502
Software and Algorithms		
MATLAB2017a (including the Statistics and Optimization Toolbox)	Mathworks	https://www.mathworks.com
AMICI	Fröhlich et al., 2017	http://icb-dcm.github.io/AMICI/
PESTO	Stapor et al., 2018	https://github.com/ICB-DCM/PESTO
Code – parameter estimation and model selection	This work	https://doi.org/10.5281/zenodo.4046502

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Carsten Marr (carsten.marr@helmholtz-muenchen.de).

Materials Availability

This study did not generate new materials.

Data and Code Availability

- H4K20 methylation proportions have been deposited at <https://github.com/marrlab/HistonesXenopus> and are publicly available at <https://doi.org/10.5281/zenodo.4046502>.
- Original MATLAB code is publicly available at <https://github.com/marrlab/HistonesXenopus> and <https://doi.org/10.5281/zenodo.4046502>.
- The scripts used to generate the figures reported in this paper are available at <https://github.com/marrlab/HistonesXenopus> and <https://doi.org/10.5281/zenodo.4046502>.
- Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal work has been conducted in accordance with Deutsches Tierschutzgesetz; Xenopus experiments were approved by the Government of Oberbayern.

Embryos Handling and HUA Treatment

Xenopus laevis eggs were collected, in vitro fertilized and handled by standard methods (Sive et al., 2000). The staging was done according to Nieukoop and Faber (Faber and Nieuwkoop, 1994). When embryos reached the desired stage (NF10.5), they were separated into two groups and incubated continuously into either HUA or mock solutions in parallel. HUA solution: 20mM Hydroxyurea (USBiological, H9120) and 150μM Aphidicolin (BioVotica, BVT-0307) in 0.1x MBS solution (Harris and Hartenstein, 1991). Mock solution: 2% DMSO (dissolvent for Aphidicolin) in 0.1x MBS solution. The embryos were collected at the four developmental stages (NF13, NF18, NF25 and NF32) for the mass spectrometry analysis.

Nuclear Histone Extraction

Around 50 to 200 embryos developed to desired stages (NF13, NF18, NF25 NF32). They were harvested and histone proteins were purified by acid extraction from nuclei (Pokrovsky et al., 2020; Schneider et al., 2011). Each developmental stage is represented by three biological replicates. Each biological replicate derived from a different mating pair.

Mass Spectrometry Sample Preparation

The pellet from the nuclear histone extraction was dissolved in an appropriate amount of Lämmli Buffer to reach $1.37 \cdot 10^6$ nuclei/μL in each sample. 15μL were loaded on an 8-16% gradient SDS-PAGE gel (SERVA Lot V140115-1) and stained with Coomassie Blue to

visualize the histone bands. Histone bands were excised and propionylated (as described in (Villar-Garea et al., 2012)). As an internal and inter-sample control, a library consisting of heavy-labelled peptides mimicking H4K20 methylation states which contain a heavy Arginine (R10 peptides) was used (product of JPT company). R10 peptides were mixed in the library with the equimolar concentration and the mix was added to each analyzed sample before in-gel trypsin digestion. Digested peptides were sequentially desalted using C18 Stagetips (3M Empore) and porous carbon material (TipTop Carbon, Glygen) as described in (Rappsilber et al., 2007) and resuspended in 15 μ l of 0.1% FA.

Mass Spectrometry Analysis

To identify and measure the proportion of the histone modifications a parallel reaction monitoring method (PRM) was used (Liebler and Zimmerman, 2013). The mass spectrometer was operated in the scheduled PRM mode to identify and quantify specific fragment ions of N-terminal peptides histone proteins. In this mode, the mass spectrometer automatically switched between one survey scan and 9 MS/MS acquisitions of the m/z values described in the inclusion list containing the precursor ions, modifications and fragmentation conditions. Survey full scan MS spectra (from m/z 270–730) were acquired with resolution 60,000 at m/z 400 (AGC target of 3×10^6). PRM spectra were acquired with resolution 30,000 to a target value of 2×10^5 , maximum IT 60 ms, isolation window 0.7 m/z and fragmented at 27% or 30% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C.

Histone Modifications Quantification

Data analysis was performed with Skyline (version 3.7) (MacLean et al., 2010) by using doubly and triply charged peptide masses for extracted ion chromatograms (XICs). Selection of respective peaks was identified based on the retention time and fragmentation spectra of the spiked in heavy-labelled peptides. Integrated peak values (Total Area MS1) were exported as csv file for further calculations. Total area MS1 from endogenous peptides was normalized to the respective area of heavy-labelled peptides. The sum of all normalized total area MS1 values of the same isotopically modified peptide in one sample resembled the amount of total peptide. The proportions of the different K20 methylation states were calculated and displayed as percentages of the overall K20 peptide amount.

METHOD DETAILS

Models

HUA and Mock Models

We consider the proportions of un- (me0), mono- (me1), di- (me2) and tri-methylated (me3) H4K20 within a *Xenopus* embryo population, defined as

$$meX = \frac{meX_{MS}}{\sum_{i=0}^3 me_{iMS}},$$

where meX_{MS} is the H4K20 methylation as measured by mass spectrometry and $X \in \{0,1,2,3\}$. We assume successive methylation and demethylation of H4K20 (van Nuland and Gozani, 2016) resulting in three possible methylation rate constants for mono-, di-, and tri-methylation with rate constants m_1 , m_2 , m_3 , respectively, and three possible demethylation rates with rate constants d_1 , d_2 , d_3 (Figures 2A and 3A). However, reactions might share rate constants. The simplest model (Figure 2B left) comprises one shared methylation rate constant for mono-, di- and tri-methylation. We successively added model-specific rate constants to this simplest model (Figure 2B). Models allowing for two specific methylation rate constants are identical to a model allowing for three specific methylation rate constants. Hence, we do not consider models with two specific methylation rate constants separately. This results in $2^3 - 3 = 5$ models for methylation - three methylation rate constants with either a shared or specific rate constant minus the three cases where we assume only two of the three rate constants to be specific. We have the same for demethylation resulting overall in $(2^3 - 3) \times (2^3 - 3) = 25$ possible HUA models.

Joint Models

The joint model considers both mock and HUA data sets. We based the joint model on our previous findings assuming three specific methylation rate constants and at most one demethylation rate constant for both mock (Figure 2D) and HUA (Figure 3B) as well as a scaled Hill function with Hill coefficient 1 and offset 0.5 as cell-cycle function. In general, the joint model would allow for $(2^3 - 3)^4 = 625$ distinct models. By constraining both the HUA and mock model to allow for three methylation and one demethylation rate constants, we are able to reduce the number of possible models to 16. The simplest joint model is comprised of 3 rate constants which are shared for mock and the HUA reactions (Figure 4B left). We successively added model complexity by allowing for HUA-specific rate constants, totaling to 16 models for the joint model with demethylation in mock and HUA and 8 models for the joint model with demethylation present in either one or none (Figure 4B).

For all models we describe the temporal changes in these proportions by systems of ordinary differential equations (ODEs) using mass action kinetics (see below).

HUA Model

We first derive the system of ODEs for the absolute numbers of H4K20me states, given by $m\tilde{e}0$, $m\tilde{e}1$, $m\tilde{e}2$, and $m\tilde{e}3$, for the model with the largest number of rate constants (Figure 3A):

$$\dot{m\tilde{e}0} = -m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1$$

$$\dot{m\tilde{e}1} = m_1 \times m\tilde{e}0 - (m_2 + d_1) \times m\tilde{e}1 + d_2 \times m\tilde{e}2$$

$$\dot{m\tilde{e}2} = m_2 \times m\tilde{e}1 - (m_3 + d_2) \times m\tilde{e}2 + d_3 \times m\tilde{e}3$$

$$\dot{m\tilde{e}3} = m_3 \times m\tilde{e}2 - d_3 \times m\tilde{e}3$$

$$\dot{N} = 0,$$

where N is the total number of histone tails. As the HUA model assumes no cell-cycle, the number of histones over time is constant and its derivative is zero. The proportions $me0$, $me1$, $me2$ and $me3$ are given by $meX = \frac{m\tilde{e}X}{N}$, for $X \in \{0, 1, 2, 3\}$ (Alabert et al., 2020) and the corresponding ODEs are given by

$$\dot{meX} = \frac{\dot{m\tilde{e}X}}{N} - \frac{m\tilde{e}X \times \dot{N}}{N^2}$$

simplifying to

$$\dot{meX} = \frac{\dot{m\tilde{e}X}}{N}$$

in the HUA model. The full ODE system for the proportions is given by

$$\dot{me0} = -m_1 \times me0 + d_1 \times me1$$

$$\dot{me1} = m_1 \times me0 - (m_2 + d_1) \times me1 + d_2 \times me2$$

$$\dot{me2} = m_2 \times me1 - (m_3 + d_2) \times me2 + d_3 \times me3$$

$$\dot{me3} = m_3 \times me2 - d_3 \times me3$$

$$\dot{N} = 0.$$

Mock Model - Constant Cell-Cycle Duration

According to the HUA model, we first formulate the ODE system of the absolute numbers of methylation states, $m\tilde{e}0$, $m\tilde{e}1$, $m\tilde{e}2$, and $m\tilde{e}3$. During DNA replication newly synthesized and unmodified histones are incorporated, leading to a constant increase in unmethylated H4K20 with a population growth rate $g(t) = \ln(2)/c(t)$, where $c(t)$ is the cell-cycle function that allows cell-cycle durations to change with time. In the simplest case, we assume the cell-cycle duration to be constant over time, denoted by a :

$$c(t) = a.$$

Then the full ODE system of the absolute numbers of methylation states is given by

$$\dot{m\tilde{e}0} = -m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a} \times (m\tilde{e}0 + m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3)$$

$$\dot{m\tilde{e}1} = m_1 \times m\tilde{e}0 - (m_2 + d_1) \times m\tilde{e}1 + d_2 \times m\tilde{e}2$$

$$\dot{m\tilde{e}2} = m_2 \times m\tilde{e}1 - (m_3 + d_2) \times m\tilde{e}2 + d_3 \times m\tilde{e}3$$

$$\dot{m\tilde{e}3} = m_3 \times m\tilde{e}2 - d_3 \times m\tilde{e}3$$

$$\dot{N} = \frac{\ln(2)}{a} \times N,$$

where N is the total number of histone tails, $N(t) = N_0 \times e^{\frac{\ln(2)}{a} \times t}$ and $N(t_0) = N_0$ the number of histone tails at the beginning of the model. Then the ODE system of the methylation proportions is given by

$$\begin{aligned}
 \dot{m\tilde{e}0} &= \frac{-m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a} \times (m\tilde{e}0 + m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3)}{N} - \frac{m\tilde{e}0 \times \frac{\ln(2)}{a} \times N}{N^2} \\
 &= -m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a} \times (m\tilde{e}0 + m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3) - m\tilde{e}0 \times \frac{\ln(2)}{a} \\
 &= -m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a} \times (m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3) \\
 \dot{m\tilde{e}1} &= m_1 \times m\tilde{e}0 - \left(m_2 + d_1 + \frac{\ln(2)}{a}\right) \times m\tilde{e}1 + d_2 \times m\tilde{e}2 \\
 \dot{m\tilde{e}2} &= m_2 \times m\tilde{e}1 - \left(m_3 + d_2 + \frac{\ln(2)}{a}\right) \times m\tilde{e}2 + d_3 \times m\tilde{e}3 \\
 \dot{m\tilde{e}3} &= m_3 \times m\tilde{e}2 - \left(d_3 + \frac{\ln(2)}{a}\right) \times m\tilde{e}3 \\
 \dot{N} &= \frac{\ln(2)}{a} \times N,
 \end{aligned}$$

leading to a constant increase of the unmethylated H4K20 proportion and a constant decrease of the methylated H4K20me proportions.

Mock Model - Linearly Increasing Cell-Cycle

In the case of a linear cell-cycle function

$$c(t) = a + b \times t$$

we first derive the ODE system for the absolute numbers of H4K20 methylation $m\tilde{e}0$, $m\tilde{e}1$, $m\tilde{e}2$, and $m\tilde{e}3$:

$$\begin{aligned}
 \dot{m\tilde{e}0} &= -m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a + b \times t} \times (m\tilde{e}0 + m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3) \\
 \dot{m\tilde{e}1} &= m_1 \times m\tilde{e}0 - (m_2 + d_1) \times m\tilde{e}1 + d_2 \times m\tilde{e}2 \\
 \dot{m\tilde{e}2} &= m_2 \times m\tilde{e}1 - (m_3 + d_2) \times m\tilde{e}2 + d_3 \times m\tilde{e}3 \\
 \dot{m\tilde{e}3} &= m_3 \times m\tilde{e}2 - d_3 \times m\tilde{e}3 \\
 \dot{N} &= \frac{\ln(2)}{a + b \times t} \times N,
 \end{aligned}$$

with N the total number of histone tails. Accordingly, the relative H4K20me proportions $me0$, $me1$, $me2$ and $me3$ are given by $meX = \frac{m\tilde{e}X}{N}$, for $X \in \{0, 1, 2, 3\}$ (Alabert et al., 2020) and the corresponding ODEs are given by the chain rule:

$$\dot{meX} = \frac{\dot{m\tilde{e}X}}{N} - \frac{m\tilde{e}X \times \dot{N}}{N^2}.$$

When plugging in the equations for the absolute H4K20 methylation states into the above equation, we receive the following ODE system for the proportional H4K20 methylation states:

$$\begin{aligned}
 \dot{me0} &= \frac{-m_1 \times m\tilde{e}0 + d_1 \times m\tilde{e}1 + \frac{\ln(2)}{a + b \times t} \times (m\tilde{e}0 + m\tilde{e}1 + m\tilde{e}2 + m\tilde{e}3)}{N} - \frac{m\tilde{e}0 \times \frac{\ln(2)}{a + b \times t} \times N}{N^2} \\
 &= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + b \times t} \times (me0 + me1 + me2 + me3) - me0 \times \frac{\ln(2)}{a + b \times t} \\
 &= -m_1 \times me0 + d_1 \times me1 + \frac{\ln(2)}{a + b \times t} \times (me1 + me2 + me3) \\
 \dot{me1} &= m_1 \times me0 - \left(m_2 + d_1 + \frac{\ln(2)}{a + b \times t}\right) \times me1 + d_2 \times me2
 \end{aligned}$$

$$\dot{m}e_2 = m_2 \times m_1 - \left(m_3 + d_2 + \frac{\ln(2)}{a + b \times t} \right) \times m_2 + d_3 \times m_3$$

$$\dot{m}e_3 = m_3 \times m_2 - \left(d_3 + \frac{\ln(2)}{a + b \times t} \right) \times m_3$$

$$\dot{N} = \frac{\ln(2)}{a + b \times t} \times N.$$

To constrain the system to biologically meaningful cell-cycle durations we included prior knowledge from literature: at 5.5 hpf the cell-cycle in *Xenopus* has been found to be ~0.5 h (Heasman, 2006). Hence, we assumed a second linearly increasing cell-cycle function

$$c(t) = 0.5 + b \times t.$$

Mock Model - Scaled Hill function

Similarly, we derive the ODE system of the methylation proportions m_0 , m_1 , m_2 and m_3 for the cell-cycle function $c(t) = a + b \times \frac{t}{h+t}$, a scaled Hill function with Hill coefficient 1 and offset:

$$\begin{aligned} \dot{m}e_0 &= \frac{-m_1 \times m\tilde{e}_0 + d_1 \times m\tilde{e}_1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (m\tilde{e}_0 + m\tilde{e}_1 + m\tilde{e}_2 + m\tilde{e}_3)}{N} - \frac{m\tilde{e}_0 \times \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times N}{N^2} \\ &= -m_1 \times m_0 + d_1 \times m_1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (m_0 + m_1 + m_2 + m_3) - m_0 \times \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \\ &= -m_1 \times m_0 + d_1 \times m_1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times (m_1 + m_2 + m_3) \\ m\tilde{e}_1 &= m_1 \times m_0 - \left(m_2 + d_1 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \right) \times m_1 + d_2 \times m_2 \\ m\tilde{e}_2 &= m_2 \times m_1 - \left(m_3 + d_2 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \right) \times m_2 + d_3 \times m_3 \\ m\tilde{e}_3 &= m_3 \times m_2 - \left(d_3 + \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \right) \times m_3 \\ \dot{N} &= \frac{\ln(2)}{a + \frac{b \times t}{h+t}} \times N. \end{aligned}$$

Similar to the mock model with linearly increasing cell-cycle function we tested three different scaled Hill functions with Hill coefficient 1 and offset as cell-cycle functions:

$$c(t) = a + \frac{b \times t}{h+t}$$

$$c(t) = 0.5 + \frac{b \times t}{h + t}$$

$$c(t) = 0.5 + \frac{b \times t}{b + t}$$

We again reduced the number of model parameters in the second equation by inserting prior knowledge about the cell-cycle duration at the start of the model (see [Mock Model - Linearly Increasing Cell-Cycle Duration](#)). Additionally, we reduced the number of model parameters further by assuming the scaling b and the dissociation constant h to be identical in the third equation. In comparison to the former two cell-cycle functions, the third equation constrains the parameter space more strictly. E.g. for upper and lower boundaries of 100 and 0.0001 for parameters a , b and h the first equation will allow for cell-cycle durations up to $100 + 100 \times (42 - 5.5) / (0.0001 + 42 - 5.5) \approx 200$ hours while the third equation only allows for cell-cycle durations for up to $0.5 + 100 \times (42 - 5.5) / (100 + 42 - 5.5) \approx 27$ hours. Additionally, we tested whether inferring the Hill coefficient with the other parameters rather than fixing it to 1 would lead to similar or improved results. Hill functions with Hill coefficient > 1 first increase slowly before a rapid increase and a gradual plateau follows. This does not reflect the biologically observed cell-cycle dynamics. Hill coefficients ≤ 1 lead to a fast initial increase which could reflect known cell-cycle dynamics. We ran the optimization for the best performing mock model without demethylation this time inferring the Hill coefficient with the other parameters (lower and upper boundaries of 0.001 and 1, respectively). We found the inferred Hill coefficient to be 0.5. While the BIC values for both models are comparable (-23 and -21 for the models with Hill coefficient = 1 and inferred Hill coefficient, respectively) the average cell-cycle durations differ (8 h and 15 h, respectively) ([Table S1](#)). By choosing a Hill coefficient = 1, we receive biologically meaningful average cell-cycle durations while reducing the number of inferred parameters by 1 and maintaining the same goodness of fit. Hence, all analyses were performed using the scaled Hill function with Hill coefficient 1 and offset 0.5.

Noise Models

As experimental data is generally noise corrupted, we evaluated all models with an underlying Laplacian noise model. Maier et. al. ([Maier et al., 2017](#)) have shown that Laplacian noise models may outperform Gaussian ones due to their increased robustness against outliers ([Maier et al., 2017](#)). All model parameters are comprised in the parameter vector θ and the experimental measurement i at time point k is denoted by \bar{y}_i^k . The log-likelihood for the Laplacian noise model is given by

$$\log L(\theta) = - \sum_{i,k} \left(\log(2\sigma) + \frac{\log(\bar{y}_i^k) - \log(y_i(t_k, \theta))}{\sigma} \right).$$

By performing maximum likelihood estimation we obtain the optimal model parameters.

OPTIMIZATION AND PARAMETER ESTIMATION

The model parameters include the initial proportions, $me1_0$, $me2_0$ and $me3_0$. Without loss of generality, we fix relative initial proportion $me0_{0rel}=0.1$ to obtain structural identifiability, where the relative initial proportions are given by

$$meX_0 = \frac{meX_{0rel}}{\sum_{i=0}^3 me_{i0}},$$

with $X \in \{0, 1, 2, 3\}$ and meX_{0rel} the relative initial proportions. We additionally infer one noise parameter, the model-specific rate constants of (de-)methylation and potentially up to three constants (mock models) describing the cell-cycle function. We tested whether the fixation of the relative initial proportion $me0_{0rel}$ influences the robustness of the optimization by fixing $me0_{0rel}=0.01$ and $me0_{0rel}=1$ for the best performing mock model without demethylation. We found that the optimized rate constant parameter sets are robust to the initializations of the relative initial proportion of unmethylated H4K20 (see [Table S1](#)). For numerical reasons we optimized the parameters in a \log_{10} scale ([Hass et al., 2019](#)). The lower and upper bounds for the rate constants, initial states, noise parameter and cell-cycle parameters were initiated in \log_{10} scale at -10 to 2, -4 to 2, -2 to 0 and -10 to 10, respectively. We performed multi-start local optimization of the negative log-likelihood using the parameter estimation toolbox PESTO ([Stapor et al., 2018](#)) and simulated the models with AMICI ([Fröhlich et al., 2017](#)). We performed at least 100 local optimization runs per model, initialized by latin-hyper cube-sampled starts. For the models not converging upon these initializations (where by 'not converging' we mean that the likelihood value of the second best run differs more than 0.1) we decreased the width between upper and lower bounds to increase the probability of convergence. For this, we assured that the optimization bounds were wide enough such that the optimal values are not in the bounds for the rate constants and the cell-cycle parameters. As the initial states are unidentifiable we ignored optimal values which ran into these boundaries as long as other optimal values were found within. For models where this was not the case we expanded the boundaries of the rate constants and initial states up to -20 to 10 and -10 to 10,

respectively, as we assumed any smaller or larger values to be biologically non-informative. For the initially non-converged joint models we also increased the number of starts to 800. Models not having converged upon manually adjusting the boundaries and running for 800 starts were determined to not converge. All mock and HUA models converged. We determined 5 out of the 40 joint models to not converge (Table S1). The given likelihood values of these joint models are lower bounds of the true optimal likelihood values obtainable upon convergence. As the likelihood values of all 5 non-converged joint models still resulted in considerably lower BIC values in comparison to the other tested models we can safely report them as best performing models for the respective demethylation hypothesis. As the BIC values between the demethylation hypotheses allowing and not allowing for demethylation in HUA differ considerably we assume the comparison between different demethylation hypotheses to be valid and the resulting conclusions to be justified.

Model Selection

We use the Bayesian Information Criterion (BIC) (Schwarz, 1978) for model comparisons:

$$\text{BIC} = \ln(n) \times k - 2 \times \log L,$$

where n is the number of data points, k is the number of estimated parameters or the overall model complexity and $\log L$ is the log-likelihood value for the maximum likelihood estimate of the model parameters. The BIC rewards high likelihood values and penalizes model complexity. Hence, low BIC values are preferable. In comparison to other model selection methods such as the Akaike Information Criterion (AIC) the BIC penalizes additional model complexity more strongly. We consider a $\Delta\text{BIC} > 10$ between two models to be enough evidence to reject the model with the higher BIC (Kass and Raftery, 1995).

Parameter Uncertainty

To receive the uncertainties for the estimated model parameters we performed Markov Chain Monte Carlo (MCMC) sampling of the posterior distribution

$$p(\theta|D) \propto L(\theta)\pi(\theta),$$

with uniform prior $\pi(\theta)$ defined over the optimization boundaries, likelihood function $L(\theta)$ and data D . We sampled the posterior for all six best performing joint models and the mock model with three specific methylation rate constants and no demethylation (PESTO-internal function `getParameterSamples`). We employed parallel tempering with five parallel chains initiated at the five most optimal parameter estimates per model obtained during optimization and performed 10^6 iterations. Upon performing a Geweke test (first 10% versus last 50% of the final MCMC chains), we discarded the first 10% of the samples as burn-in phase and thinned the chains keeping only every 100th sample. The marginal posterior distributions are plotted via violin plots (plotting function `violin`, Hoffman, H. (2015). `violin.m` - Simple violin plot using matlab default kernel density estimation. (<https://de.mathworks.com/matlabcentral/fileexchange/45134-violin-plot>), MATLAB Central File Exchange. Retrieved November 13, 2019.)).

Validation - Cell-Cycle Durations

We used the median and the 25th and 75th percentiles of the MCMC chain determined during the parameter uncertainty analysis for the cell-cycle parameter b , and evaluated the median and the 25th and 75th percentiles of the cell-cycle function according to

$$c(t) = 0.5 + \frac{b \times t}{b + t},$$

for $t \in [0, 40]$, where the cell-cycle duration of 0.5 hours at 5.5 hpf (start of model) is taken from (Anderson et al., 2017; Gelens et al., 2015).

Prediction of Number of Cells

Using the median and the 25th and 75th percentiles of the cell-cycle parameter b (as determined in the validation analysis), we determined the theoretical number of cells a *Xenopus* embryo is on average composed of between 5.5 hpf and 45.5 hpf according to $dN/dt = \ln(2)/(0.5 + b \times t/(b + t)) \times N$, where

$$\frac{\partial N(t)}{\partial t} = \frac{\ln(2)}{0.5 + \frac{b \times t}{b + t}} \times N$$

$$N(t) = N_0 \times e^{(2 \times \ln(2)) / (2 \times b + 1)^2 \times (2 \times b^2 \times \ln(2 \times b \times t + b + t) + 2 \times b \times t + t - 2 \times b^2 \times \ln(b))},$$

where $N(t)$ is the number of cells at time t , N_0 the initial number of cells and $0.5 + b \times t/(b + t)$ the cell-cycle function (constrained scaled Hill function with Hill coefficient 1 and offset 0.5). At the start of the model (at 5.5 hpf) we take the initial number of cells N_0 to be 4096 (Heasman, 2006).

Implementation

The toolboxes used for the analysis of the manuscript for ODE simulation (AMICI (Fröhlich et al., 2017)) and parameter estimation (PESTO (Stapor et al., 2018)) are available under <https://github.com/ICB-DCM>. The MATLAB code corresponding to this manuscript is available via <https://github.com/marrlab/HistonesXenopus>. The analysis was performed with MATLAB 2017a.

QUANTIFICATION AND STATISTICAL ANALYSIS

For comparing H4K20me data for mock and HUA (Figure 1B), a two-sample t-test at the 0.05 significance level was used for all three biological replicates of mock and HUA for each time point. In Figures 2F and 2G, the model predictions are given for the median, 25th and 75th percentiles of the MCMC samples of the cell-cycle parameter of the model with three specific methylation rate constants but no demethylation. Stated values of specific (de-)methylation rate constants are given by the median and the credibility ranges from the 25th to the 75th percentiles. All statistics and analyses were performed with MATLAB 2017a.

Cell Systems, Volume 11

Supplemental Information

**H4K20 Methylation Is Differently Regulated
by Dilution and Demethylation in Proliferating
and Cell-Cycle-Arrested *Xenopus* Embryos**

Lea Schuh, Carolin Loos, Daniil Pokrovsky, Axel Imhof, Ralph A.W. Rupp, and Carsten Marr

H4K20 methylation kinetics are differently regulated by dilution and demethylation in proliferating and cell-cycle arrested *Xenopus* embryos

Lea Schuh, Carolin Loos, Daniil Pokrovsky, Axel Imhof, Ralph A. W. Rupp, Carsten Marr

Summary

Initial Submission: Received May 14, 2020
Preprint: <https://doi.org/10.1101/2020.05.28.110684>
Deposited on bioRxiv, May 30, 2020
Scientific editor: Bernadett Gaál, DPhil.

First round of review: Number of reviewers: 4
4 confidential, 0 signed
Revision invited July 3, 2020
Major changes anticipated
Revision received Aug 5, 2020

Second round of review: Number of reviewers: 4
4 original, 0 new
4 confidential, 0 signed
Accepted Nov. 11, 2020

Data freely available: Yes
Code freely available: Yes

Editor's View

This paper is addressing the salient biological problem of whether the cell cycle plays an "active" role in shaping the histone modification landscape in proliferating cells, given that DNA replication "passively" (by default) dilutes histone modifications. The answer to this question will be of interest to biologists interested

in how histone methylation is maintained and regulated, particularly in rapidly dividing cells during embryonic development.

The problem addressed in this paper is a very nicely constrained one and proliferating vs. cell-cycle arrested *Xenopus* embryos offer an exceptionally well-suited *in vivo* model in which to address it. We also noted that the computational modelling here "an example of a well-placed model that's necessary for insight", that is, an exceptionally good example of good use of modelling. First, the succinct description of the model hypotheses that could describe H4K20 methylation kinetics in cycling and cell-cycle arrested cells gives a clear view of the problem at hand and of the space of possibilities. Using multi-start maximum likelihood optimization and quantitative model selection and the experimental data, the authors arrive at biological insights that would be difficult to attain relying solely on currently available experimental tools and without modelling.

We are always on the lookout for studies that address salient biological questions through a combination of well-designed experiments and good use of modeling, so this was not a difficult initial decision to make. During peer review we were therefore primarily looking to make sure that our assessment of the suitability of the experimental system and computational modelling was correct and that the experiments and analysis are robust and technically correct.

Accordingly, in my first decision letter I emphasized the need to address the questions the reviewers had about the robustness of the model and the key conclusions, and their concerns about whether the quantification of histone methylation was sufficiently accurate for the purposes of this work, and whether the HUA treatment arrested cells as expected and had no major off-target effects that could affect the results. The authors were able to address the concerns raised by the reviewers in one round of revisions.

This Transparent Peer Review Record is not systematically proofread, type-set, or edited. Special characters, formatting, and equations may fail to render properly. Standard procedural text within the editor's letters has been deleted for the sake of brevity, but all official correspondence specific to the manuscript has been preserved.

Editorial decision letter with reviewers' comments, first round of review

Dear Carsten,

I'm enclosing the comments that reviewers made on your paper, which I hope you will find useful and constructive. As you'll see, they express interest in the study, but they also have a number of criticisms and suggestions. Based on these comments, it seems premature to proceed with the paper in its current form; however, if it's possible to address the concerns raised with additional experiments and/or analysis, we'd be interested in considering a revised version of the manuscript.

As a matter of principle, I usually only invite a revision when I'm reasonably certain that the authors' work will align with the reviewers' concerns and produce a publishable manuscript. In the case of this manuscript, the first priority should be to thoroughly address the concerns the reviewers raise about some of the assumptions of the model, its robustness to these and those with respect to data quality and accuracy. I've highlighted some of the most important of these below.

1. Can we know whether the cell cycle duration function capture the heterogeneous cell cycle durations in the developing embryo or at least can we be confident that the results and conclusions are robust to this heterogeneity? (raised by Reviewer 1)
2. Are the results and conclusions robust to the way that the model handles H4K20 methylation dilution? (Reviewer 3 point 1)
3. Reviewer 4 is concerned about the potential for overfitting.
4. Given that high accuracy quantification of histone methylation is exceptionally important in this context, the technical concerns raised in this regard by Reviewer 1 are important to address.
5. Please provide demonstration that HUA treatment is acting as anticipated, as requested by Reviewer 2.

If it's technically possible to experimentally distinguish between the predictions of the models with and without demethylation, along the lines of what Reviewer 3 suggests (point 2), adding this would certainly strengthen the conclusions.

I've also highlighted portions of the reviews that strike me as particularly critical. I'd also like to be explicitly clear about an almost philosophical stance that we take at Cell Systems.

We believe that understanding how approaches fail is fundamentally interesting: it provides critical insight into understanding how they work. We also believe that all approaches do fail and that it's unreasonable, even misleading, to expect otherwise. Accordingly, when papers are transparent and forthright about the limitations and crucial contingencies of their approaches, we consider that to be a great strength, not a weakness. Please keep this in mind when discussing what conclusions can and what cannot be drawn based on your analysis and approach.

As you address these concerns, it's important that you and I stay on the same page. I'm always happy to talk, either over email or by phone, if you'd like feedback about whether your efforts are moving the manuscript in a productive direction. We also appreciate that the COVID-19 pandemic challenges and limits what you and your lab can do, so if you would like to talk through your revision plan then let's schedule a Zoom call.

Do note that we generally consider papers through only one major round of revision, so the revised manuscript would be either accepted or rejected based on the next round of comments we receive from the reviewers. If you have any questions or concerns, please let me know. More technical information and advice about resubmission can be found below my signature. Please read it carefully, as it can save substantial time and effort later.

I look forward to seeing your revised manuscript.

All the best,

Bernadett

Bernadett Gaal, DPhil
Scientific Editor, Cell Systems

Reviewers' comments:

Reviewer #1: In this interesting manuscript, Schuh et al developed a computation model to determine how histone H4K20 methylation (me0/me1/me2/me3) states are integrated and regulated by the cell-cycle and *Xenopus laevis* embryogenesis. To build their model, the authors first use H4K20 methylation mass spectrometry of histone H4 from nuclei of differentiating *Xenopus* embryos and then use a variety of approaches to model the cell cycle and "methylation" rate constants. They then use these models to back-fit the cell cycle terms to derive a set of best-fit joint models. Using their model, the authors tested the question of whether or not demethylation is necessary to account for the observed dynamics of K20 methylation. Their model showed that demethylation is likely essential for explanation of the HUA-dependent H4K20 dynamics, but "redundant" in mock-treated embryos. This is an important observation that addresses some long-standing questions in the field. As this is just a model, the authors don't experimentally test this question. They propose, in the discussion, a fascinating biological consequence of this distinction that demethylases are only important as a trigger for differentiation. As this is likely the most important consequence of this manuscript, this result is begging for an experimental test. This may be beyond the scope of the current manuscript, but at least proposing how the authors would perform the test would be helpful (e.g. a PHF8 knockdown, a *Drosophila* screen of KDMs for cell-cycle elongation coupled differentiation defects, or something).

Overall, this manuscript is likely of interest to many in the developmental biology, chromatin, and molecular modeling fields. It will be of use to many in developing similar models for other histone and non-histone PTMs that may regulate the cell cycle. However, before I recommend publication, I suggest that the authors consider the critique above and the specific points below:

I'm confused by the statement at the bottom of page 3: "In HUA, methylation accumulates in the di- and tri-methylation states in comparison to mock. There, newly synthesized and unmodified histones are incorporated upon DNA replication, leading to an overall dilution of H4K20me and hence a higher proportion of lowly methylated un- and mono-methylated states." If HUA leads to an increase in di- and tri-methylation, and there is no DNA replication, that means that newly synthesized unmodified histones are NOT incorporated in HUA? This sentence is not written clearly.

Figure 1B - each time point of the H4K20me proportion should be subject to a t-test.

The difference in H4K20 methylation between mock and HUA is rather modest and less than one would predict from the simple hypothesis of cell cycle dependence. To truly demonstrate that this difference is

due to altered methyltransferase activities, a PR-Set7 or SUV420H morpholino injection would help strengthen this experiment. However, this could be challenging. To demonstrate that H4K20 methylation kinetics are truly altered upon HUA cell cycle arrest, perhaps an alternative experiment would be to perform this experiment in cell free extracts. The cell cycles are much faster than in post-MBT embryos and the consequences of altered H4K20 dynamics would be more robust.

The definition and derivation of the cell-cycle duration function $c(t)$ is complicated. The cell-cycle duration of pre- and post-MBT and gastrula cells are highly variable, going from exceedingly fast pre-MBT to very slow during gastrulation and neurulation. But do 3 functions (constant, linearly increasing, or gradually plateauing) account for the biological cell cycles observed in the very heterogeneous developing embryo? While the authors compared one of their top models with an "average cell-cycle duration" in neural progenitors, this still a conflation of a model derived from total embryo histone mass spectrometry to a heterogeneous population of cells. Some additional analysis of this question is warranted.

For the histone mass spectrometry, it is not clear that in-gel trypsin digestion and recovery is adequately quantitative and representative of the entire histone pool. Similarly, it does not appear that missed cleavages are accounted for; while propionylation should eliminate this issue, the authors should at least include searches for missed cleavage. While the authors have published this approach before, some additional justification that this approach matches the recovery from RP-HPLC followed by propionylation/digestion is warranted in this study. This is particularly important as the model development needs high accuracy quantification to be valid.

"Methylation rate constants" should probably be renamed. Maybe "biological methylation rate constants". Rate constants have a specific biochemical meaning for each enzyme (e.g. k_{cat} for the turnover rate and K_m for the Michaelis-Menten constant representing substrate affinity). As the authors are not deriving actual enzymatic rates constants, the authors should ensure that readers are not misled.

What is the authors' proposed biological/biochemical reason for the Hill coefficient being not 1?

In Figure 4C and F, the annotation for the model ID (e.g. a-f) is confusing. The figure subpanels are also a-f.

Reviewer #2: This article by Schuh and colleagues addresses a fundamental puzzle in developmental cell biology: as cells proliferate, how do they maintain their epigenetic state and combat dilution of histone marks through DNA replication? Schuh and colleagues tackle this question using the multiple methylation states of H4K20 in developing *Xenopus laevis* embryos as a model. In this study, they examine the levels of un- mono- di- and tri-methylated H4K20 across a developmental time course, beginning at gastrulation stages (stage 11), when *X. laevis* embryos have re-acquired a full cell cycle with G1, S, G2 and M phases. The principle perturbation used in hydroxyurea/aphidicolin (HUA) which induces cell cycle arrest at the G1/S transition, and the authors use mathematical modeling to infer the kinetics of methylation and demethylation rates in normal and HUA-treated embryos. They find that three independent methylation rate constants are needed to generate each of the methylation states with the dynamics observed, and interestingly, predict that demethylation is dispensable in normally-cycling cells. The underlying fundamental question of how epigenetic states are maintained across a proliferating and heterogeneous

cell population is exciting and should be interesting to a broad audience. The application of this modeling approach to developing embryos has the potential to be broadly useful across many biological settings. Therefore I think the work is potentially appropriate for Cell Systems.

I have one major concern:

Given the very long incubation times with HUA, I think it's **essential to demonstrate that the HUA treatment is acting as anticipated**. This is especially since, by the authors' calculations, the number of cells should be increasing from 20k to 300k over this period.

In particular, the authors **should confirm that HUA is robustly introducing the expected G1/S cell cycle arrest throughout the embryo**.

I can imagine several experimental strategies for this, but the most obvious would be to use FACS together with DNA staining (Hoechst e.g.) to illustrate and quantify the degree of cell cycle arrest in HUA versus control embryos at each of the stages used.

I also think it would be prudent **to confirm that drastic off target effects are not being caused that might conflate the conclusions**; the most obvious of these to me would be cell death, which could be checked by TUNEL staining at each stage.

Minor comment:

In the Introduction, when discussing the roles of H4K20 modifications and the associated enzymatic activities that deposit them, the authors move rapidly between discussing results observed in cell culture and those obtained across a wide range of model organisms. I request that the authors specify the experimental system in which each result was obtained, because it's potentially misleading to assume that conclusions reached in human cell lines will also hold true in early *Xenopus* embryos.

Reviewer #3: In this manuscript the authors analyze the cell cycle dependent kinetics of H4K20 methylation. Their main result is that active demethylation is probably dispensable during active cell cycles but not in cases where the cell cycle has been arrested. In the former case, dilution at DNA replication is potentially sufficient for regulating H4K20me levels. In my view, this work is a useful contribution to the field, as quantitative characterization of histone PTM kinetics is rarely performed, despite the importance of such analyses. However, I have some difficulties with the analysis, both from a theoretical and experimental angle, as I detail below. Furthermore, one of the principle results, namely that active demethylation is less important for active cell cycles is already a known result for some other histone modifications, particularly H3K27me. This detracts somewhat from the manuscript's novelty.

* **The model handles dilution through an outflow of H4K20 states dependent on the cell cycle duration c.** However, in reality in a single cell, dilution is a discrete event. At a population level, this may not matter **but only if the cells are unsynchronized**. In the experiments presented here, the cells will start out synchronised by virtue of fertilisation at $t=0$. Given this synchrony, **it is not clear whether the desynchronization assumption is true at the later timepoints, between 14.75 and 40 hpf, when the average cell cycle is about 8 h long**. This issue needs to be discussed.

* The author's favorite model is the one without any active demethylation. However, a model with active demethylation still fits equally well (Fig 2D), although is structurally more complex. The authors say in the abstract that active demethylation is redundant. I think that is going too far: to rule out active demethylation, what seems to be required is measurement of H4K20me at earlier time points (<10 hpf), where the model predictions with/without active demethylation diverge (Fig 2E). In my view, it is essential

to perform these experiments to validate the model prediction. It would also be useful to plot the prediction from the best active demethylation model to the data of Fig 2F,G to see how it compares.

* In the methods, it is mentioned that $me0_0$ is set to 0.1 for structural identifiability. Do other initial conditions lead to different answers? If so, I would be concerned about the robustness of the methodology.

* I think there may also be an error in the equations for $d/dt(me0)$ for the cases where the cell cycle time is itself a function of time. For the case where $c(t)=a+bt$, then $d/dt(me0)$ should have a term which is $(\ln 2 a)/(a+bt)^2 \times (me0+me1+me2+me3)$. A similar issue occurs in the Hill function cases. If so, the analysis would have to be redone with potentially different conclusions.

Reviewer #4: Schuh et al. present a study on computational modeling of H4K20 methylation. On the whole this is an interesting angle: modeling methylation rates on the basis of quantitative mass-spec measurements. It is a manuscript with a relatively narrow, but well-specified focus. It is well-written and the results are clearly presented. The code to reproduce the analyses is available: good.

I have mostly one major concern which has to do with the number of measurements in relation to the number of models tested. For instance, the authors fit 180 models to the normally developing embryos (mock-treated). The data consists of 16 measurements (four methylation states at four stages). To me this sounds like this has a huge potential for overfitting. By fitting this large number of models I'm not sure this actually is certain to result in something biologically meaningful in the end. Could not any combination of 16 measurements be fitted to some degree? In the end, the models and fitted parameters can be biologically explained, but how trustworthy are these models in the end? This concern is not alleviated by the results in the manuscript, as a model fitted on the mock-treated embryos cannot explain the data of the HUA embryos. In this case, a demethylation rate is necessary. How useful is a model if it cannot predict the effect of perturbations? In my view, a perturbation experiment would really be necessary to validate a model. Here, the assumption would be that a good model could predict the effect of a perturbation. For instance, inhibition or knock-out of a methyltransferase or demethylase. However, due to functional redundancy and different specificities, this is likely not a trivial experiment. Added to that, I am hesitant to require further experiments in these uncertain times.

In conclusion, I think the manuscript is interesting and original. However, due to the combined experimental and computational setup, I'm not sure how valid the fitted models are and to what extent these can be interpreted biologically.

Other remarks:

* Abstract: "suggesting that cell-cycle mediated dilution of chromatin marks is an essential regulatory component for shaping the epigenetic landscape during early embryonic development." This too broad of a statement. The manuscript described H4K20 methylation, which has been linked to the cell cycle. This may or may not hold for other chromatin marks, but this is outside the scope of this manuscript.

* It is stated that the model retrieves correct cell-cycle durations. This is not surprising, as a specific model is chosen to agree with the cell-cycle durations: "only a constrained scaled Hill function with Hill coefficient 1 and offset 0.5 gives a cell-cycle duration in the expected range." If you choose a model that fits the data, it is not unexpected that this model yields correct predictions.

* Why were these four stages chosen? The earliest measured stage is late gastrula (NF 13), however, establishment of H4K20 methylation would be interesting to measure. In *Xenopus*, a large fraction of the

epigenome is established at or around MBT (blastula stage), therefore it would be interesting to include several stages that also cover blastula and early and mid-gastrula.

Authors' response to the reviewers' first round comments

Attached.

Editorial decision letter with reviewers' comments, second round of review

Dear Carsten,

I'm very pleased to let you know that the reviews of your revised manuscript are back, the peer-review process is complete, and only a few minor, editorially-guided changes are needed to move forward towards publication.

I've made some suggestions about your manuscript within the "Editorial Notes" section, below. Please consider my editorial suggestions carefully, ask any questions of me that you need, make all warranted changes, and then upload your final files into Editorial Manager. ***We hope to receive your files within 5 business days, but we recognize that the COVID-19 pandemic may challenge and limit what you can do. Please email me directly if this timing is a problem or you're facing extenuating circumstances.***

I'm looking forward to going through these last steps with you. More technical information can be found below my signature, and please let me know if you have any questions.

All the best,
Bernadett

Bernadett Gaal, DPhil
Scientific Editor, Cell Systems

Editorial Notes

Transparent Peer Review: Thank you for electing to make your manuscript's peer review process transparent. As part of our approach to Transparent Peer Review, we ask that you add the following sentence to the end of your abstract: "A record of this paper's Transparent Peer Review process is included in the Supplemental Information." Note that this ***doesn't*** count towards your 150 word total!

Also, if you've deposited your work on a preprint server, that's great! Please drop me a quick email with your preprint's DOI and I'll make sure it's properly credited within your Transparent Peer Review record.

Abstract: I've gone over your abstract with the goals of making it more succinct and concrete, largely by reworking and recombining parts of your original version. See what you think. Please feel free to revert anything that you don't like or that you feel distorts your meaning! I apologize if there are instances of the latter.

"DNA replication during cell division leads to dilution of histone modifications and can thus affect chromatin-mediated gene regulation, raising the question of how the cell-cycle shapes the histone modification landscape, particularly during embryogenesis. We tackled this problem by manipulating the cell-cycle during early *Xenopus laevis* embryogenesis and analysing in vivo histone H4K20 methylation kinetics. The global distribution of un-, mono- di- and tri-methylated histone H4K20 was measured by mass spectrometry in normal and cell-cycle arrested embryos over time. Using multi-start maximum likelihood optimization and quantitative model selection, we found that three specific biological methylation rate constants were required to explain the measured H4K20 methylation state kinetics. While demethylation is essential for regulating H4K20 methylation kinetics in non-cycling cells, demethylation is very likely dispensable in rapidly dividing cells of early embryos, suggesting that cell-cycle mediated dilution of H4K20 methylation is an essential regulatory component for shaping its epigenetic landscape during early development."

Manuscript Text: The text is compelling and clear, but the Discussion seems rather too long. We favour slim Discussions that do not reiterate what's found in the Results beyond a brief transitional summary and are limited to around four medium sized paragraphs. There are important points in the Discussion, including those that you included in response to Reviewers 1 and 3. Please focus on these and limit reiteration of the results to what's necessary to convey these points. Please also remove subheadings from the Discussion section. Please feel free to ignore the strict character limits on the main text listed in the checklist that I had sent you some time ago, but please do work on making the Discussion shorter and more succinct.

Also:

- House style disallows editorializing within the text (e.g. strikingly, surprisingly, importantly, etc.), especially the Results section and in the Summary. These terms are a distraction and they aren't needed—your excellent observations are certainly impactful enough to stand on their own. Please remove these words and others like them. “Notably” is suitably neutral to use once or twice if absolutely necessary.
- We don't allow “priority claims” (e.g. new, novel, first, etc.). For a discussion of why, read: <http://crosstalk.cell.com/blog/getting-priorities-right-with-novelty-claims>, <http://crosstalk.cell.com/blog/novel-insights-into-priority-claims>.

Figures and Legends:

Please ensure that all figures included in your point-by-point response to the reviewers' comments are present within the final version of the paper, either within the main text or within the Supplemental Information.

STAR Methods:

Please replace your Materials and Methods section with STAR Methods and please follow the STAR Methods format for reporting experimental procedures, methods, and analysis, including the Key Resources Table. Cell Press introduced the STAR Methods format to help improve the rigour in reporting methods and resources for reproducibility. This section replaces the Experimental Procedures and Supplemental Experimental Procedures sections. For detailed instructions on STAR Methods and a template for the Key Resources Table, see our [STAR Methods](#) webpage. Please contact me if you have any questions about restructuring your manuscript using the STAR Methods format.

Cell Press has recently changed the way it approaches "availability" statements for the sake of ease and clarity. Please revise the first section of your STAR Methods as follows, noting that the particular examples used might not pertain to your study.

RESOURCE AVAILABILITY

Lead Contact: Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jane Doe (janedoe@qwerty.com).

Materials Availability: This study did not generate new materials. *-OR-* Plasmids generated in this study have been deposited at [Addgene, name and catalog number]. *-OR-* etc.

Data and Code Availability:

- **Source data statement** (described below)
- **Code statement** (described below)
- **Scripts statement** (described below)
- Any additional information required to reproduce this work is available from the Lead Contact.

Starting in August of 2020, Cell Systems papers will need to contain a comprehensive and structured "Data and Code Availability" statement. These statements will exceed standard STAR Methods requirements, so please note that **the instructions below supersede the instructions found [here](#).**

Data and Code Availability statements pertain to the source data and original code reported in the study. In this context, **source data** is defined as the collection of individual, unprocessed observations used to generate the figures reported in the paper. Examples include scRNA-seq and proteomic datasets, but also CSV spreadsheets used to generate graphs, and original micrographs in TIFF format. **Code** is defined as any computationally implemented program, algorithm, or pipeline necessary to reproduce the

analysis or conclusions reported in a paper. Smaller **scripts** that have been used to visualize data and generate figures should also be included in the statement, as described below.

Data and Code Availability statements are reported in the first section of the STAR Methods. **They have four parts and each part must be present. Each part should be listed as a bullet point, as indicated above. For convenience, a .docx template for Data and Code availability statements can be downloaded [here](#).**

Part 1 pertains to source data. Examples can be used in any number or combination, making sensible modifications as necessary:

- [Data-type] source data have been deposited at [data-type-specific repository] and are publicly available under the accession numbers: [Insert].
- [Data-type] source data have been deposited at [general repository] and are publicly available at [insert DOI].
- [Data-type] source data are available in the paper's Supplemental Information.
- The [data-type] source data reported in this study have not been deposited in a publicly available repository because [reason why data are not public] . They have been archived locally [insert archiving plan]. To request access [insert instructions].
- This paper analyzes existing, publicly available data. These datasets' accession numbers are provided in the Key Resource Table.
- Source data are not provided in this paper but are available from the Lead Contact on request. *(Note: Cell Systems discourages this practice. If you need to make this statement, please discuss it with your editor first.)*

Part 2 pertains to original code. Examples can be used in any number or combination, making sensible modifications as necessary:

- [Adjective] original code is publicly available at [repository name and DOI].
- [Adjective] original code is available in this paper's Supplemental Information.
- The original code reported in this study is not publicly available repository because [reason why data are not public]. Original code has been archived locally [insert archiving plan]. To request access [insert instructions].
- This paper does not report original code.

Part 3 pertains to scripts used to generate figures. Examples to be used in any number or combination:

- The scripts used to generate the figures reported in this paper are available at [repository name and DOI].
- The scripts used to generate the figures reported in this paper are available in this paper's Supplemental Information.

- The scripts used to generate the figures reported in this paper are available in the [name software package, with version, and provide reference or URL] and their use is described in the STAR Methods.
- Scripts were not used to generate the figures reported in this paper.
- Scripts used to generate the figures presented in this paper are not provided in this paper but are available from the Lead Contact on request. *(Note: Cell Systems discourages this practice. If you need to make this statement, please discuss it with your editor first.)*

Part 4 is a statement: “Any additional information required to reproduce this work is available from the Lead Contact.”

Please ensure that the large datasets generated in this paper has been archived in at least one publicly accessible repository (e.g. GEO, PRIDE, etc.). If there is no community-approved repository for your large-scale data, we recommend that you deposit them on Mendeley Data. Please provide your datasets' DOIs within both the Data and Software Availability section of the STAR Methods and the Key Resources Table. Thank you!

Please ensure that custom code has been archived in at least one publicly accessible repository (e.g. GitHub, Zenodo, etc) and that a DOI is provided within both the Data and Software Availability section of the STAR Methods and the Key Resources Table. Thank you!

You do not currently have a **Key Resources Table** (KRT). The KRT highlights key reagents and resources used in the paper. The table need not list every item used or generated in the study and does **not** replace the detailed explanation of the methods and materials used in the study in the STAR Methods text. For details, please refer to the [Table Template](#) or feel free to ask me for help.

All datasets and code generated or used in this paper must be listed, with their unique identifiers, within appropriate sections of the Key Resources Table. For the work generated in this paper, please indicate "This work" as the source.

Thank you!

Reviewer comments:

Reviewer #1: The authors have satisfactorily addressed my critique.

Reviewer #2: The authors have satisfactorily addressed my comments.

Reviewer #3: The authors have satisfactorily addressed my questions.

Reviewer #4: The authors have sufficiently addressed my concerns.