

## ARTICLE OPEN



# Epigenetic OCT4 regulatory network: stochastic analysis of cellular reprogramming

Simone Bruno <sup>1</sup>, Thorsten M. Schlaeger <sup>2</sup> and Domitilla Del Vecchio <sup>1</sup> ✉

Experimental studies have shown that chromatin modifiers have a critical effect on cellular reprogramming, i.e., the conversion of differentiated cells to pluripotent stem cells. Here, we develop a model of the OCT4 gene regulatory network that includes genes expressing chromatin modifiers TET1 and JMJD2, and the chromatin modification circuit on which these modifiers act. We employ this model to compare three reprogramming approaches that have been considered in the literature with respect to reprogramming efficiency and latency variability. These approaches are overexpression of OCT4 alone, overexpression of OCT4 with TET1, and overexpression of OCT4 with JMJD2. Our results show more efficient and less variable reprogramming when also JMJD2 and TET1 are overexpressed, consistent with previous experimental data. Nevertheless, TET1 overexpression can lead to more efficient reprogramming compared to JMJD2 overexpression. This is the case when the recruitment of DNA methylation by H3K9me3 is weak and the methyl-CpG-binding domain (MBD) proteins are sufficiently scarce such that they do not hamper TET1 binding to methylated DNA. The model that we developed provides a mechanistic understanding of existing experimental results and is also a tool for designing optimized reprogramming approaches that combine overexpression of cell-fate specific transcription factors (TFs) with targeted recruitment of epigenetic modifiers.

*npj Systems Biology and Applications* (2024)10:3; <https://doi.org/10.1038/s41540-023-00326-0>

## INTRODUCTION

Through the process of cellular differentiation, embryonic stem cells evolve into a variety of specialized cell types. By contrast, cellular reprogramming converts differentiated cells to induced pluripotent stem cells (iPSCs)<sup>1</sup>. Since human iPSCs have functions almost identical to those of embryonic stem cells (ESCs), they can be used to replace damaged cells, representing a promising alternative to ESCs for regenerative medicine<sup>2,3</sup>. The first iPSC reprogramming approach, introduced by Yamanaka et al.<sup>4,5</sup>, is based on overexpression of four key transcription factors (TFs), OCT4, Sox2, Klf4, and c-Myc (OSKM). Because the efficiency of the initial reprogramming process was very low<sup>6–9</sup>, a plethora of follow-up studies have aimed at improving efficiency<sup>10–14</sup>.

We can largely group these studies into those that keep the TF cocktail to OSKM and investigate the extent to which the levels of these factors influence efficiency<sup>11,15–20</sup> and those that add factors to the original OSKM cocktail, such as epigenetic modifiers<sup>21–24</sup>. Epigenetic modifiers are enzymes that either add or remove covalent modifications to histones and DNA, which ultimately affect chromatin compaction and hence gene expression. In particular, the extent to which chromatin is compactified determines how easily a gene can be transcribed<sup>25</sup>. The extent of chromatin compaction, in turn, is dictated by enzymatic modifications to the histones around which DNA is wrapped, such as H3K9 methylation (H3K9me3) or H3K4 methylation/acetylation (H3K4me3/ac), and by specific enzymatic modifications to DNA itself<sup>26</sup>. Therefore, chromatin state provides an additional layer of transcriptional regulation, which can be modulated by epigenetic modifier enzymes. Accordingly, the studies using epigenetic modifiers are grounded on the fact that in terminally differentiated cells, such as in fibroblasts used in cellular reprogramming, genes that are highly expressed in

pluripotent stem cells are “shut off”, often due to highly compactified chromatin<sup>26–30</sup>.

In the last decade, a plethora of experimental studies investigated how chromatin modifications affect cellular reprogramming<sup>21–24,31</sup>. These investigations aimed to determine which chromatin modifications impact the reprogramming process in terms of improving efficiency and reducing latency variability. Here, efficiency refers to the percentage of reprogrammed cells within a given time frame and latency variability refers to variability of the time that an individual cell takes until it gives rise to a daughter iPSC cell<sup>32</sup>. For instance, Chen et al.<sup>22</sup> demonstrated that H3K9 methylation acts as a barrier during the reprogramming of mouse embryonic fibroblasts (MEFs) into iPSCs. They observed a significant increase in the number of reprogrammed cells by adding JMJD2, an enzyme that removes H3K9me3, in the OSKM cocktail. Similarly, Gao et al.<sup>24</sup> showed that the inclusion of TET1, an enzyme that removes DNA methylation, to the original OSKM cocktail resulted in a twofold enhancement of iPSC reprogramming efficiency. Furthermore, it was shown that the removal of Methyl-CpG-binding domain (MBD) proteins can further increase this improvement. MBD proteins bind to methylated CpG dinucleotides<sup>33</sup>, protecting them from being bound by TET1<sup>34</sup>. In refs. <sup>21,31</sup>, knock downs of MBDs were performed, resulting in more than a tenfold increase in iPSC reprogramming efficiency.

In this paper, we focus on the dynamics of reactivation of TF OCT4 during reprogramming, since it is well known that overexpression of OCT4 alone is sufficient for iPSC reprogramming<sup>15,19,35–37</sup> and that OCT4 is a key regulator of TET and JMJD2<sup>26,38–42</sup>. To this end, we created a model for the three-gene network composed of OCT4, TET1, and JMJD2 and defined this network the *epigenetic OCT4 gene regulatory network* (Epi OCT4 GRN). We then simulated the model using Gillespie’s Stochastic

<sup>1</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. <sup>2</sup>Boston Children’s Hospital Stem Cell Program, Boston Children’s Hospital, 300 Longwood Avenue, Boston, MA 02115, USA. ✉email: [ddv@mit.edu](mailto:ddv@mit.edu)

Simulation Algorithm (SSA)<sup>43</sup> in order to determine the efficacy of different reprogramming approaches with respect to efficiency and latency variability. We also analyzed how biochemical parameters, such as proliferation rate and the concentration of MBDs, affect the reprogramming process, with the aim of providing a mechanistic understanding of the outcomes that have been experimentally observed.

It is important to point out that a physics-based biochemical reaction model that incorporates histone modifications, DNA methylation, and transcription factor-mediated regulation has been recently developed<sup>44</sup>. However, this model has not yet been exploited to construct a OCT4 gene regulatory network aimed at investigating the impact of epigenetic modifiers on iPSC reprogramming. On the other side, some models that include histone modifications or DNA methylation into gene expression regulation to investigate iPSC reprogramming have appeared in the past years<sup>45–48</sup>. However, none of these models include both histone modifications and DNA methylation. Furthermore, none of these models include the effect of MBD proteins on TET1 activity, thus limiting the ability to study their influence on the reprogramming process.

This paper is organized as follows. In the “Results” section, we introduce the Epi OCT4 GRN and present the results of our computational analysis. Finally, in the “Discussion” section we present a discussion and conclusive remarks.

## RESULTS

In this section, we first introduce the model of the epigenetic OCT4 gene regulatory network developed in this paper. We then provide detailed explanations of the results obtained from our computational analysis. Specifically, we first studied how DNA methylation affects the dynamics of cellular differentiation, which we captured in our model by the progressive inactivation of the OCT4 gene starting from a high OCT4 level, corresponding to the pluripotent state<sup>49,50</sup>. We then compared different reprogramming approaches by studying the impact of overexpressing OCT4 alone and in combination with TET1 and JMJD2 on the dynamics of OCT4 gene reactivation. Specifically, we examined process efficiency and latency variability for each approach, as these metrics have been commonly employed in published experimental studies to assess the success of cellular reprogramming<sup>21–24,31</sup>. To this end, we performed a computational study of the temporal trajectories of the system by simulating the reactions associated with the Epi OCT4 GRN (Fig. 2) with Gillespie’s Stochastic Simulation Algorithm (SSA)<sup>43</sup>.

### Model of the epigenetic OCT4 gene regulatory network

Our model includes one TF gene, OCT4, and two chromatin modifier genes, TET1 and JMJD2. We consider OCT4 as the only TF in our model because it has been shown that overexpression of OCT4 alone leads to reprogramming<sup>15,19,35–37</sup>. The chromatin modifiers JMJD2 and TET1 are enzymes that catalyze the erasure of histone modification H3K9me3 and DNA methylation, respectively<sup>26,38–42</sup>, two chromatin modifications associated with compacted chromatin state and gene silencing<sup>51</sup>. Specifically, while JMJD2 directly erases H3K9me3<sup>26,38–40</sup>, TET1 recognizes CpGme dinucleotides and converts methylated CpG to carboxylcytosine through multiple intermediate forms<sup>41,42</sup>, none of which is recognized by DNMT1, the enzyme responsible for copying the CpGme pattern on the nascent DNA strand during DNA replication<sup>51</sup>. Transcription factor OCT4 recruits writers of H3K4me3 to its own gene<sup>52</sup> and to the JMJD2<sup>53</sup> and TET1<sup>54</sup> genes. This leads to a model in which OCT4 self-activates and also activates TET1 and JMJD2 by recruiting writers of activating chromatin modifications, while TET1 and JMJD2 self-activate and mutually activate each other and OCT4 by erasing repressive

chromatin modifications. We call this system the epigenetic OCT4 gene regulatory network (Epi OCT4 GRN).

The chromatin modifiers TET1 and JMJD2 act on the chromatin modification circuit within each gene, which has been developed in ref. <sup>44</sup>. This circuit includes H3K9 methylation (H3K9me3), DNA methylation (CpGme), H3K4 methylation/acetylation (H3K4me3/ac), and their known interactions. The first two modifications are associated with a repressed gene state<sup>51</sup>, while H3K4me3/ac is associated with an active gene state<sup>26,55</sup>.

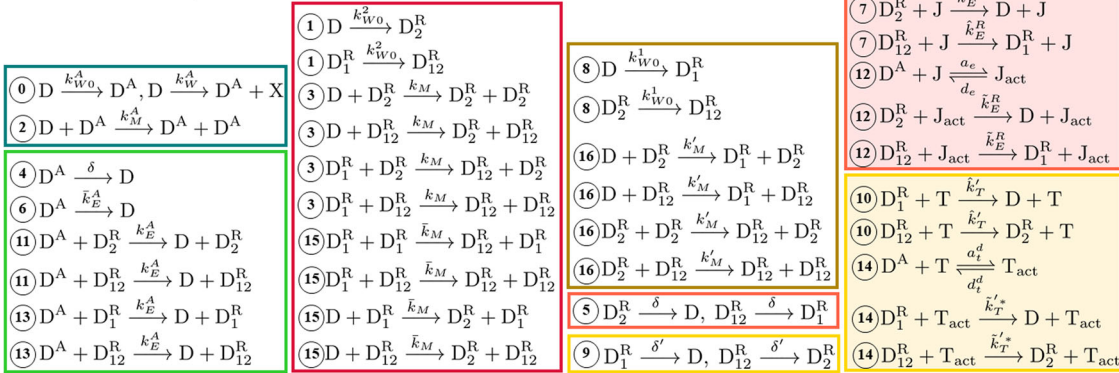
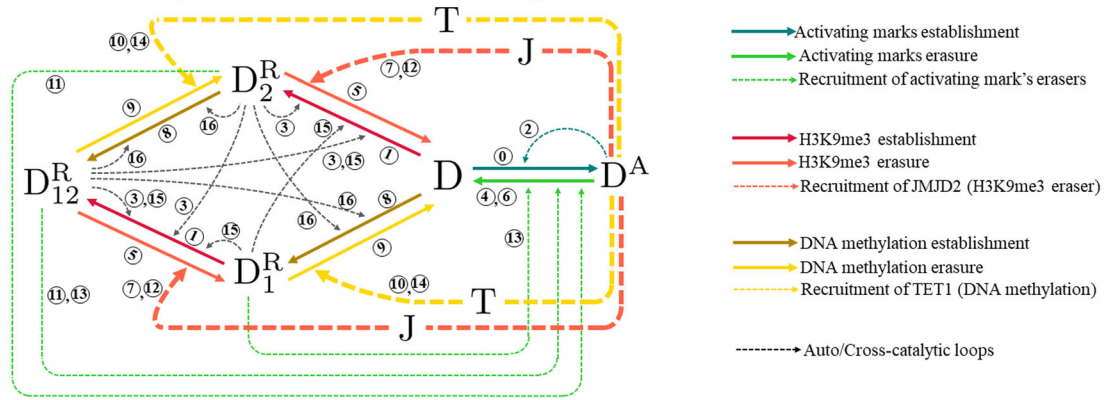
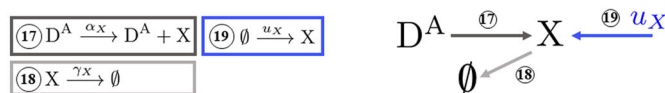
We next describe the chromatin modification circuit and the model of gene expression (see ref. <sup>44</sup> for details). In terms of species, in this model we have D (nucleosome with DNA wrapped around it), D<sup>A</sup> (nucleosome with H3K4me3/ac), D<sub>2</sub><sup>R</sup> (nucleosome with H3K9me3), D<sub>1</sub><sup>R</sup> (nucleosome with CpGme), and D<sub>12</sub><sup>R</sup> (nucleosome with both H3K9me3 and CpGme) (Fig. 1a). Then, the expression rate of each gene will be determined by the number of nucleosomes with activating (D<sup>A</sup>) or repressive (D<sub>1</sub><sup>R</sup>, D<sub>2</sub><sup>R</sup>, D<sub>12</sub><sup>R</sup>) chromatin modifications. In terms of molecular interactions, both histone modifications and DNA methylation can be de novo established (process encapsulated in reactions ①, ②, and ③). Then, histone modifications can enhance the establishment of marks of the same kind to nearby nucleosomes via a read-write mechanism, generating auto-catalytic reactions (encapsulated in ④, ⑤). Analogously, repressive histone modifications enhance the establishment of DNA methylation, and vice versa, generating cross-catalytic reactions (encapsulated in ⑥, ⑦). Finally, each modification can be passively removed through dilution, due to DNA replication (reactions ⑧, ⑨, and ⑩), or through the action of eraser enzymes (basal erasure) (reactions ⑪, ⑫, and ⑬). These erasers can be also recruited by the opposite modifications (recruited erasure), that is, repressive modifications recruit activating modification’s erasers and vice versa (reactions ⑭, ⑮, ⑯, and ⑰). In this model, the rate of the processes described above for H3K9me3 (DNA methylation) is assumed not to change if the other repressive mark is present on the same nucleosome. All the reactions described above are collected in the list of Fig. 1b, in which the reactions involving TET1 and JMJD2 are shaded in yellow and pink respectively. A diagram of the chromatin modification circuit corresponding to the reactions in Fig. 1b is provided in Fig. 1c. In this system, the transcriptional self-activation is modeled as a Hill function with cooperativity 1. Specifically,  $k_W^A$  (Fig. 1b) is a monotonically increasing function of the abundance of X (X), which can be written as  $k_W^A = \tilde{k}_W^A (X/K_A) / (1 + (X/K_A))$ , in which  $\tilde{k}_W^A$  is a coefficient that does not depend on X, and  $K_A$  is the dissociation constant of the binding reaction between X and DNA<sup>44</sup>.

In our model, transcription is allowed only by D<sup>A</sup>, and transcription and translation are lumped together (reaction ⑰). The transcription by D is considered negligible because it is assumed that transcription of D by RNA polymerase II occurs concurrently with H3K4me3 deposition (i.e., conversion of D to D<sup>A</sup>), as observed in ref. <sup>56</sup>. Furthermore, the gene product X is subject to dilution due to cell division and degradation (reaction ⑩). Finally, the production of X can be artificially increased through overexpression (reaction ⑱). A diagram representing these reactions is shown in Fig. 1d. Then, based on the interactions among OCT4, TET1, and JMJD2, we can wire the three chromatin modification circuits to obtain the Epi OCT4 GRN circuit that we analyze in this paper (Fig. 2). In the single gene’s chromatin modification circuit proposed in ref. <sup>44</sup> the concentrations of TET1 and JMJD2 impact the rate constants as they do here. However, these concentrations are constant parameters themselves. By contrast, in the Epi OCT4 GRN introduced here the concentrations of TET1 and JMJD2 are state variables and, as such, they vary dynamically under the effect of each other concentrations and the concentration of OCT4.

Let  $D_{tot} = D_{tot}/\Omega$ , in which  $D_{tot}$  represents the total number of modifiable nucleosomes within the gene of interest and  $\Omega$

**a Chromatin modification circuit model: nucleosome modifications**

D	D <sup>A</sup>	D <sub>1</sub> <sup>R</sup>	D <sub>2</sub> <sup>R</sup>	D <sub>12</sub> <sup>R</sup>
Unmodified nucleosome	nucleosome with H3K4me3/ac	nucleosome with CpGme	nucleosome with H3K9me3	nucleosome with both H3K9me3 and CpGme

**b Chromatin modification circuit model: reaction list**

**c Chromatin modification circuit model: reaction diagram**

**d Gene expression model**

**e**

Param.	Definition	Interpretation
$r$	$k_M^R/k_M^A$	$r$ is the relative speed between the establishment of DNA methylation through repressive histone modifications and the self-establishment of repressive histone modifications
$\varepsilon_d$	$\delta/(k_M^A D_{tot})$	$\varepsilon_d$ scales the ratio between the rate of the dilution process and the rate of the auto-catalysis for each mark
$\varepsilon_e$	$k_E^A/(k_M^A D_{tot})$	$\varepsilon_e$ scales the ratio between the rate of the basal erasure and the rate of the auto-catalysis for each mark
$\varepsilon'$	$k_E^R/k_M^A$	$\varepsilon'$ scales the ratio between the rate of the recruited erasure and the rate of the auto-catalysis for each mark
$\tilde{\mu}$	$(k_E^R D_{tot})/k_E^A$	$\tilde{\mu}$ is the relative speed between the erasure rates of repressive and activating histone modifications
$\tilde{\mu}'$	$(k_E^R D_{tot})/k_E^R$	$\tilde{\mu}'$ is the relative speed between the erasure rates of DNA methylation and activating histone modifications

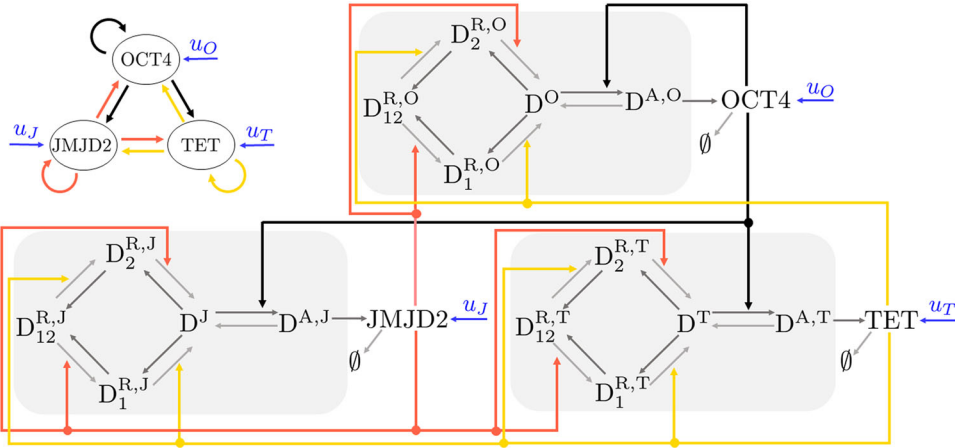
**Fig. 1 The role of TET1 (T) and JMJD2 (J) in the gene's inner chromatin modification circuit.** **a** Nucleosome modifications considered in the model. **b** Reactions associated with the chromatin modification circuit of each gene X, with X = O (OCT4), T (TET1), J (JMJD2). Here, reactions associated with activating marks, H3K9me3 and DNA methylation are enclosed in green boxes, pink boxes, and yellow boxes, respectively. Dark shades are associated with reactions describing the establishment of the modifications and light shades are associated with reactions describing the erasure of the modifications. Furthermore, shaded boxes enclose reactions involving J (pink) and T (yellow). Finally, each reaction rate constant is subscripted with  $W$ ,  $M$ , or  $E$  to indicate the association of its corresponding reaction with the writing (establishment), maintenance (auto/cross-catalytic reactions), or erasure of a chromatin modification. **c** Diagram representing the chromatin modification circuit for each gene X. **d** Reactions associated with the gene expression model. Specifically, reactions associated with production (dark gray), dilution/degradation (light gray), and artificial overexpression  $u_X$  (blue) of the gene product X. The reactions are described in the "Model of the epigenetic OCT4 gene regulatory network" subsection. **e** Definitions and interpretations of  $r$ ,  $\varepsilon_d$ ,  $\varepsilon_e$ ,  $\varepsilon'$ ,  $\tilde{\mu}$ , and  $\tilde{\mu}'$ .

represents the reaction volume, and the normalized time  $\tau = tk_M^A D_{tot}$ . Let us define the dimensionless parameters  $a = k_M^R/k_M^A$ ,  $\bar{a} = \bar{k}_M^R/\bar{k}_M^A$ , and  $a' = k_M^R/k_M^A$ , in which  $a$  represents the normalized rate constant of the auto-catalytic reaction for repressive histone modifications, and  $\bar{a}$  and  $a'$  represent the normalized rate constants of the cross-catalytic reactions between

repressive histone modifications and DNA methylation. Let us also introduce:

$$r = \frac{a'}{a} = \frac{k_M^R}{k_M^A}, \quad (1)$$





**Fig. 2 The epigenetic OCT4 gene regulatory network (Epi OCT4 GRN).** Diagram of the Epi OCT4 GRN, in which OCT4 self-activates and also activates TET1 and JMJD2 by recruiting writers of activating chromatin modifications on all genes (black arrows), while TET1 and JMJD2 self-activate and mutually activate each other and OCT4 by recruiting erasers for repressive chromatin modifications on all genes (pink and yellow arrows, respectively). For simplicity of illustration, in each gene's chromatin modification circuit we used gray for the solid arrows indicating the establishment and erasure and we did not represent the dashed arrows indicating recruitment and catalysis.

that is, the ratio between the rate at which repressive histone modifications enhance the establishment of DNA methylation through cross-catalytic reactions ( $k'_M$ ) and the rate at which repressive histone modifications enhance their own establishment through auto-catalytic reactions ( $k_M$ ). Furthermore, we use  $\eta$  to represent the efficiency of the maintenance process of DNA methylation by DNMT1<sup>26</sup>. The expression of  $\eta$  was derived in ref. <sup>44</sup> by accounting for the dilution of DNA methylation due to DNA replication and the maintenance process in which DNMT1 replicates CpG methylation on the newly synthesized DNA strand based on the pattern of the mother strand<sup>26,57</sup>. The expression of  $\eta$  is given by  $\eta = \delta'/\delta$ , in which  $\delta$  represents the rate constant of the passive erasure through dilution and  $\delta'$  represents the effective passive erasure rate constant obtained from the balance between the dilution and the maintenance process. In particular,  $\eta = 1$  if DNMT1 is completely absent (no maintenance) and  $\eta = 0$  if the maintenance process is 100% efficient. Now, we define:

$$\tilde{\mu} = \frac{\tilde{k}_E^R D_{tot}}{k_E^A}, \quad \tilde{\mu}' = \frac{\tilde{k}_T^* D_{tot}}{k_E^A}, \quad \varepsilon_d = \frac{\delta}{k_M^A D_{tot}}, \quad \varepsilon_e = \frac{\tilde{k}_E^A}{k_M^A D_{tot}}, \quad \varepsilon' = \frac{k_E^A}{k_M^A}, \quad (2)$$

with  $\tilde{\beta} = O(1)$  and  $\tilde{b} = O(1)$  such that  $(\tilde{k}_T^* D_{tot})/k_E^A = \tilde{\beta}\tilde{\mu}'$  and  $(\tilde{k}_E^R D_{tot})/k_E^A = \tilde{b}\tilde{\mu}$ , respectively. Specifically,  $\tilde{\mu}$  is a dimensionless parameter quantifying the asymmetry between the erasure rates of repressive and activating histone modifications, while  $\tilde{\mu}'$  is a dimensionless parameter quantifying the asymmetry between the erasure rates of DNA methylation and activating histone modifications. Furthermore,  $\varepsilon_d = \delta/k_M^A D_{tot}$  is the normalized rate constant associated with dilution due to DNA replication, and, since  $k_E^R/k_M^A = \tilde{b}\tilde{\mu}\varepsilon_e$ ,  $\tilde{k}_T^*/k_M^A = \tilde{\beta}\tilde{\mu}'\varepsilon_e$ ,  $\tilde{k}_E^R D_{tot}/k_M^A = \tilde{\mu}\varepsilon'$  and  $\tilde{k}_T^* D_{tot}/k_M^A = \tilde{\mu}'\varepsilon'$ , the dimensionless parameter  $\varepsilon_e$  ( $\varepsilon'$ ) scales the ratio between the rate of the basal erasure (recruited erasure) and the one of the auto/cross-catalysis of chromatin modifications. We collect the definitions and interpretations of these parameters in Fig. 1e.

Finally, for modeling the gene expression process, we introduce  $\bar{p}_O = a_O/\gamma_O$ ,  $\bar{p}_T = a_T/\gamma_T$ ,  $\bar{p}_J = a_J/\gamma_J$ ,  $\bar{u}_O = (u_O\Omega)/\gamma_O$ ,  $\bar{u}_T = (u_T\Omega)/\gamma_T$ ,  $\bar{u}_J = (u_J\Omega)/\gamma_J$  and  $n_X^A$ , with  $X = O, T, J$ , that is, the total amount of nucleosomes modified with activating chromatin modifications for each gene  $X$ .

We next describe the relationship between the abundance of MBD proteins ( $B$ ) and the rate coefficients associated with the erasure of DNA methylation by the action of TET1. Specifically, these coefficients are  $\tilde{k}_T^*$  for the reactions in which TET1 is

recruited by  $D^A$  (reactions ⑭) and  $\tilde{k}_T^*$  for the reactions in which TET1 is not recruited by  $D^A$  (reactions ⑩). As derived in<sup>44</sup>,  $\tilde{k}_T^*$  and  $\tilde{k}_T$  can be written as:

$$\tilde{k}_T = \frac{F_1}{F_2 \frac{B}{K_B} + 1}, \quad \tilde{k}_T^* = \frac{F_3}{F_4 \frac{B}{K_B} + 1}, \quad (3)$$

in which  $K_B$  is the dissociation constant of the binding reaction between  $B$  and methylated DNA and  $F_1, F_2, F_3$  and  $F_4$  are parameters independent of  $B$  or  $K_B$ . Then,  $\tilde{k}_T$  and  $\tilde{k}_T^*$  increase when  $B/K_B$  decreases. From (2),  $\tilde{\mu}'$  is therefore an increasing function of  $\tilde{k}_T^*$  that can be written as:

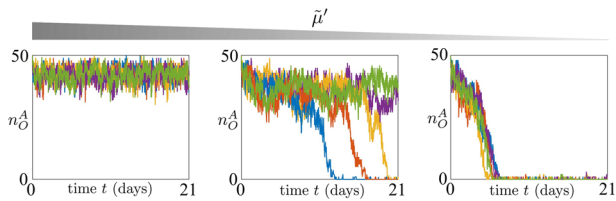
$$\tilde{\mu}' = \frac{\tilde{k}_T^* D_{tot}}{k_E^A} = \frac{D_{tot}}{k_E^A} \frac{F_3}{F_4 \frac{B}{K_B} + 1} = \frac{F_5}{F_2 \frac{B}{K_B} + 1}, \quad (4)$$

in which  $F_5 = F_3 D_{tot}/k_E^A$  is independent of  $B$  or  $K_B$ . From (4), we can conclude that knocking down MBD proteins ( $B = 0$ ) or locally preventing their binding to methylated DNA ( $K_B \rightarrow \infty$ ) allows to increase  $\tilde{\mu}'$ .

### Effect of DNA methylation on the dynamics of OCT4 during differentiation

It has been experimentally observed that, during cellular differentiation, the OCT4 gene undergoes progressive silencing, with concurrent establishment of DNA methylation<sup>50</sup>. Here, we investigate how the rate constant of the DNA methylation erasure process by TET1  $\tilde{\mu}'$  affects the first time that, without any external stimulus, the OCT4 gene reaches the repressed state  $n_O^A \approx 0$  corresponding to a differentiated state<sup>36</sup>, starting from the active state  $n_O^A/D_{tot} \approx 1$  (Fig. 3). For values of  $\tilde{\mu}'$  sufficiently large, none of the trajectories reach the OCT4 repressed state in the time window of the simulation, indicating stability of the high OCT4 expression state. By reducing  $\tilde{\mu}'$ , most of the trajectories reach, although in a stochastic manner, the OCT4 repressed state. Finally, if we keep reducing  $\tilde{\mu}'$ , then the trajectories reach  $n_O^A \approx 0$  quickly and in a more synchronous fashion. This result suggests that, even if the initial state, where OCT4 gene is active, is devoid of DNA methylation, the stability of this active state is contingent on a sufficiently fast erasure of DNA methylation.

Furthermore, given that TET1 is the enzyme catalyzing the erasure of DNA methylation (reactions ⑩ and ⑭ in Fig. 1b), this erasure is faster in the presence of high levels of TET1. We should then expect that stem cells have higher levels of TET1 and that, during differentiation, TET1 levels decline. This is consistent with



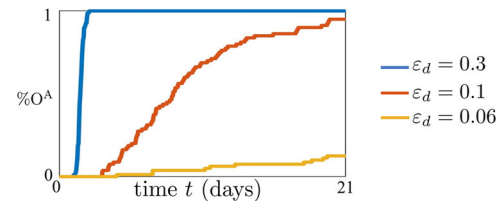
**Fig. 3 Small DNA methylation erasure rate leads to a faster and more synchronous differentiation process.** Time trajectories of  $n_O^A$  (total amount of nucleosomes modified with activating chromatin modifications for the OCT4 gene) starting from the OCT4 fully active state for different values of  $\tilde{\mu}$ . In all plots, on the x axis we have the time (days). The parameter values used for these simulations can be found in Supplementary Table 1. In particular, we set  $\tilde{\mu}' = 1, 0.15, 0.07$ ,  $\varepsilon_d = 0.3$ ,  $\bar{p}_O = \bar{p}_T = \bar{p}_J = 3.2$ ,  $\eta = 0.1$ ,  $\tilde{\mu} = 1$ ,  $\varepsilon_e = 0.3$  and  $\varepsilon' = 1$ . In our model, parameter  $\tilde{\mu}'$  quantifies the asymmetry between the erasure rates of DNA methylation and activating histone modifications. Mathematical definition of  $\tilde{\mu}'$  can be found in Eq. (2). For all simulations, we implemented the reactions listed in Fig. 1 with the SSA<sup>43</sup> and we considered a time span of 21 days ( $\tau = 201.6$ ) and  $D_{\text{tot}} = 50$  (see Supplementary Note 1).

in vivo studies showing that the level of TET1, highly expressed in embryonic stem cells (ESCs), decreases during cellular differentiation, reaching a low expression level after gastrulation<sup>58</sup>. This is also consistent with in vitro studies showing that in ESCs that start to differentiate, the level of OCT4 declines together with the level of TET1<sup>59</sup>.

Finally, it is worth noting that H3K9me3 decay rate is estimated to be much higher than the decay rate of DNA methylation<sup>44,60</sup>. Therefore, decreasing the erasure rate constant of repressive histone modifications  $\tilde{\mu}$  has a lesser impact on both the speed of the differentiation process and the stability of the high OCT4 state compared to decreasing the erasure rate constant of DNA methylation (see Supplementary Fig. 1).

### Effect of proliferation rate on reprogramming through OCT4 overexpression

Here, we investigate how the proliferation rate  $\varepsilon_d$  affects the efficiency of reprogramming when overexpressing OCT4 only. To this end, we analyzed the trajectory of the active chromatin state of the OCT4 gene,  $n_O^A$ , starting from a fully repressed state ( $n_O^A = n_T^A = n_J^A = 0$ ), when we artificially overexpress OCT4, by letting  $\bar{u}_O > 0$  in the reactions listed in Fig. 1d. Now, we define the process efficiency  $\%O^A$  as the percentage of trajectories of  $n_O^A$  that reach the active state by a prefixed time. Then, for a fixed input  $\bar{u}_O$ , we evaluated  $\%O^A$  for different values of  $\varepsilon_d$ . Simulations show that increasing  $\varepsilon_d$  speeds up OCT4 reactivation dynamics, making the process more efficient (Fig. 4). These findings align with the experimental results reported by Hanna et al.<sup>32</sup>. In their study<sup>32</sup>, the authors investigated the impact of various factors, including proliferation rate and addition of Nanog to the original OSKM TF cocktail, on the kinetics of iPSCs formation from somatic cells. Notably, their findings demonstrate that a higher proliferation rate accelerates the reprogramming process. The reason why increasing  $\varepsilon_d$  makes the OCT4 reactivation process more efficient is that a higher  $\varepsilon_d$  leads to a higher decay rate of all modifications (Eq. (2), Figs. 1 and 2). Given that the initial state is the OCT4 repressed state, mainly characterized by repressive chromatin marks, then higher  $\varepsilon_d$  leads to a faster erasure of DNA methylation and repressive histone modifications. This, in turn, allows faster establishment of activating histone modifications, leading to a faster, and also more efficient reactivation of the OCT4 gene.

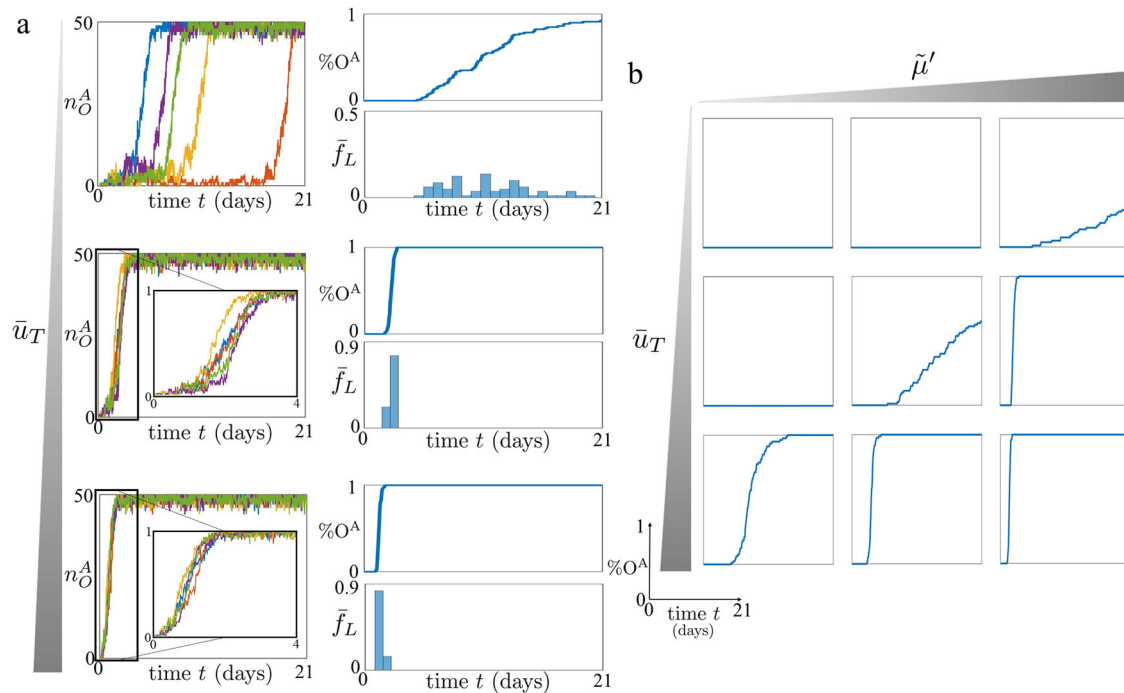


**Fig. 4 Higher proliferation rate speeds up reprogramming through overexpression of OCT4.**  $\%O^A$  for different values of  $\varepsilon_d$ , when we set  $\bar{u}_O > 0$ , capturing OCT4 overexpression. The parameter values used for these simulations can be found in Supplementary Table 1. In particular, we consider  $\varepsilon_d = 0.3, 0.1, 0.06$  and we set  $\bar{u}_O = 320$ ,  $\bar{u}_T = 0$ ,  $\bar{u}_J = 0$ ,  $\tilde{\mu}' = 1$ ,  $\bar{p} = 5$ , in which  $\bar{p} = \bar{p}_O = \bar{p}_T = \bar{p}_J$ ,  $\eta = 0.1$ ,  $\varepsilon_e = 0.3$ ,  $\tilde{\mu} = 1$  and  $\varepsilon' = 1$  (see Supplementary Fig. 3 for how the parameters  $\bar{u}_O$ ,  $\varepsilon_e$ ,  $\tilde{\mu}'$ , and  $\bar{p}$  influence the impact of  $\varepsilon_d$  on the reprogramming process). In our model, parameter  $\varepsilon_d$  represents the normalized rate constant associated with dilution due to DNA replication. For all simulations, we implemented the reactions listed in Fig. 1 with the SSA<sup>43</sup> and we considered a time span of 21 days ( $\tau = 201.6$ ) and  $D_{\text{tot}} = 50$  (see Supplementary Note 1).

### Reprogramming through concurrent OCT4 and TET1 overexpression

Here, we consider a reprogramming approach in which the enzyme TET1 is also overexpressed. We first determined how efficiency  $\%O^A$  and the latency variability are affected by different levels of TET1 overexpression  $\bar{u}_T > 0$ . To this end, we conducted several simulations for a fixed  $\bar{u}_O$ , capturing a fixed level of OCT4 overexpression and different values of  $\bar{u}_T$  (Fig. 5). Simulations show that adding overexpression of TET1 makes the OCT4 reactivation process more efficient and reduces the latency variability (Fig. 5a). Specifically, the reduction of latency variability is captured by the narrower distribution in the values of  $\bar{f}_L$ , that is, the normalized frequency of a specific latency value across all the simulations. This reduction can be also captured by observing the shape of the curves  $\%O^A$ -time in Fig. 5a. As the level of TET1 overexpression increases, the synchronization of OCT4 reactivation events also increases, resulting in steeper curves. Overall, these results can be justified by the fact that activating histone modifications can be established only on unmodified nucleosomes, as shown in the chromatin modification circuit diagram (Fig. 1c), and therefore the reprogramming process cannot start until repressive modifications are erased. Accordingly, while OCT4 recruits writers for activating chromatin modifications, TET1 erases DNA methylation. OCT4 overexpression can have an effect only when unmodified nucleosomes are present, which is allowed with higher probability by the presence of TET1. These results are in agreement with experimental data showing that the addition of TET1 to the original OSKM TF cocktail increases iPSC reprogramming efficiency, promoting the formation of OCT4<sup>+</sup> colonies<sup>24</sup>.

However, in these experiments, the efficiency of the reprogramming process was not increased pronouncedly. Specifically, the authors obtained only a twofold increase in the number of OCT4<sup>+</sup> colonies observed at day 13 after doxycycline (Dox) induction<sup>24</sup>. This is surprising if one does not consider the role of MBD proteins, which is captured in our model by Eq. (4). Specifically, according to this model, MBD proteins bind to methylated CpG dinucleotides<sup>33</sup> and then protect them from being bound by TET1<sup>34</sup>. This, in turn, results in the erasure rate of DNA methylation not increasing substantially when TET1 is added if MBD proteins (B) level is high. Therefore, TET1 overexpression only scarcely enhances the erasure rate of DNA methylation unless MBD proteins are either knocked down or prevented from binding DNA (see expression of  $\tilde{\mu}'$  in Eq. (4)). In order to verify how different levels of MBD affect the OCT4-TET1 overexpression reprogramming approach, we thus evaluated  $\%O^A$  for several values of  $\bar{u}_T$  and  $\tilde{\mu}'$  (Fig. 5b). The higher  $\tilde{\mu}'$  is, obtained by lowering



**Fig. 5** Concurrent OCT4 and TET1 overexpression leads to a more efficient and less stochastic reprogramming process, under specific parameter regimes. **a** Left hand side plots: time trajectories of  $n_O^A$  (total amount of nucleosomes modified with activating chromatin modifications for the OCT4 gene) starting from the OCT4 fully repressed state ( $n_O^A = 0$ ) for different values of  $\bar{u}_T$ . Right hand side plots:  $\%O^A$ , that is, the normalized amount of  $N = 100$  time trajectories which reach  $n_O^A \geq 40$ , starting from  $n_O^A = 0$ , and  $\bar{f}_L$ , that is, the normalized frequency of a latency value across all the  $N = 100$  simulations. In all plots, on the x axis we have the time (days). The parameter values used for these simulations can be found in Supplementary Table 1. In particular, we consider three values of  $\bar{u}_T$  (i.e.,  $\bar{u}_T = 0, 50, 160$ ), and we set  $\bar{u}_O = 320$ ,  $\bar{u}_J = 0$ ,  $\tilde{\mu}' = 0.5$ ,  $\epsilon_d = 0.2$ ,  $\epsilon_e = 0.2$ ,  $\bar{p}_O = \bar{p}_T = \bar{p}_J = 5$ ,  $\eta = 0.1$ ,  $\tilde{\mu} = 1$  and  $\epsilon' = 1$  (see Supplementary Fig. 4 for how the parameters  $\bar{u}_O$ ,  $\epsilon_d$ ,  $\epsilon_e$ ,  $\tilde{\mu}'$ , and  $\bar{p}$  influence the impact of  $\bar{u}_T$  on the reprogramming process). **b**  $\%O^A$  for different values of  $\tilde{\mu}'$  and  $\bar{u}_T$ . The parameter values used for these simulations can be found in Supplementary Table 1. In particular, we consider  $\bar{u}_T = 0, 50, 160$ ,  $\tilde{\mu}' = 3, 0.5, 0.2$ ,  $\epsilon_d = 0.1$ ,  $\epsilon_e = 0.1$ , and all the other parameter values equal to the ones considered for the simulations in (a) (see Supplementary Fig. 4 for how the parameters  $\bar{u}_O$ ,  $\epsilon_d$ ,  $\epsilon_e$ , and  $\bar{p}$  influence the impact of  $\bar{u}_T$  and  $\tilde{\mu}'$  on the reprogramming process). In our model,  $\tilde{\mu}'$  quantifies the asymmetry between the erasure rates of DNA methylation and activating histone modifications. For all simulations, we implemented the reactions listed in Fig. 1 with the SSA<sup>43</sup> and we considered a time span of 21 days ( $\tau = 201.6$ ) and  $D_{\text{tot}} = 50$  (see Supplementary Note 1).

the MBD proteins level, the more efficient and less variable the reactivation process is. These results are consistent with previous experimental data showing that knock down of MBDs increases significantly the efficiency of iPSC reprogramming and reduces its latency variability<sup>21,31</sup>.

Our results also show that, independent of the MBD protein level, by sufficiently increasing TET1 expression, OCT4 reactivation achieves an almost constant latency and high efficiency (Fig. 5b). Although this is in principle possible, it is also plausible that in practice the high levels of TET1 required to achieve this may be toxic to the cell or may not even be practically reachable.

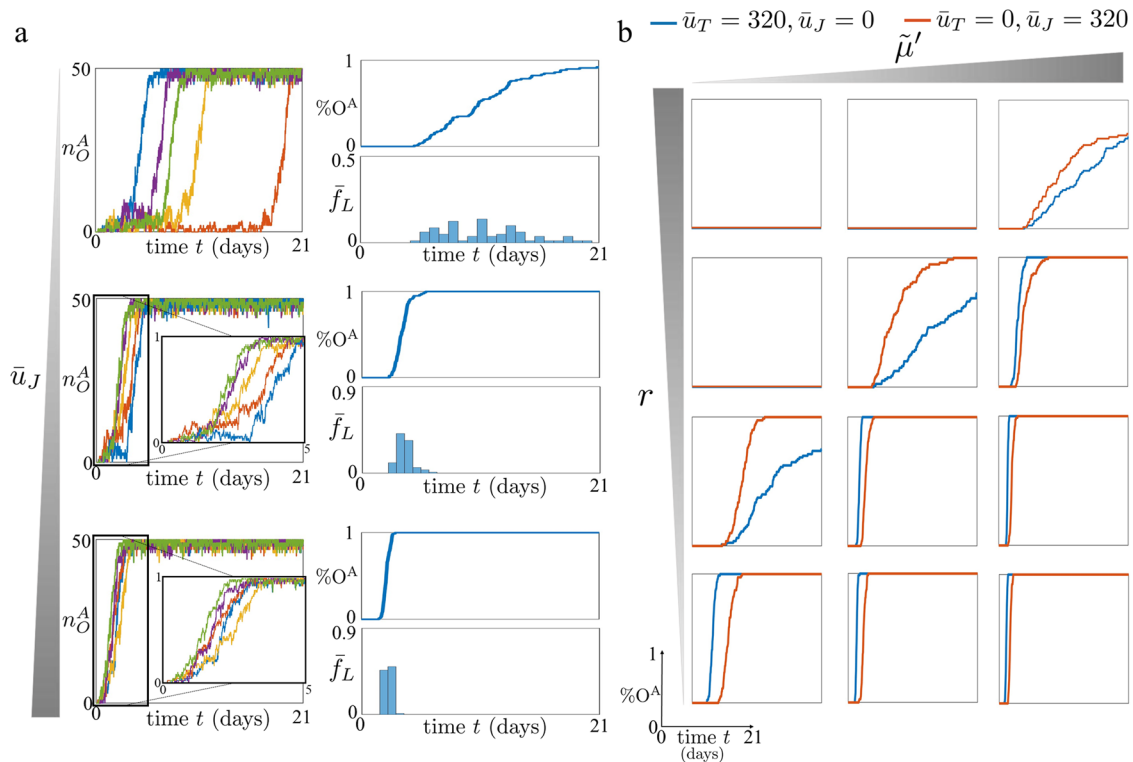
### Reprogramming through concurrent OCT4 and JMJD2 overexpression

Based on experimental studies conducted in the last decade, the repressive histone modification H3K9me3 seems to be a similarly crucial barrier for the reprogramming process<sup>22,61</sup>. Indeed, it was shown that the addition of JMJD2 in the OSKM cocktail enhances the reprogramming efficiency, resulting in a  $\approx 1.75$ -fold increase in the number of iPSC colonies obtained within the experimental time frame (8 days)<sup>22</sup>.

Thus, we evaluated efficiency and latency variability for three different levels of JMJD2 overexpression,  $\bar{u}_J$  (Fig. 6a). Furthermore, in order to properly compare this reprogramming approach to the one based on concurrent OCT4 and TET1 overexpression, we considered the same value as used in the previous analysis for the OCT4 overexpression level ( $\bar{u}_O$ ). Simulations show that JMJD2

overexpression reduces the latency variability of the reprogramming process and increases the number of trajectories of  $n_O^A$  that reach the active state by a fixed time period (Fig. 6a), in agreement with the experimental data in ref. <sup>22</sup>.

We next compared efficiency and latency variability of the OCT4–TET1 overexpression approach to those of the OCT4–JMJD2 overexpression approach (Fig. 6b). The results show how efficiency and latency variability of both approaches are affected by  $\tilde{\mu}'$  and  $r = \alpha'/\alpha$ . Parameter  $r$  is the ratio between the rate of the cross-catalytic reaction with which repressive histone modifications enhance the establishment of DNA methylation and the rate of the auto-catalytic reaction with which repressive histone modifications enhance their own establishment (Eq. (1)). For high values of DNA methylation erasure rate,  $\tilde{\mu}'$ , the OCT4–JMJD2 overexpression approach is more efficient than the OCT4–TET1 overexpression approach only when  $r \gg 1$ , that is, repressive histone modifications enhance the establishment of DNA methylation more than their own establishment (Fig. 6b). In fact, in this parameter regime, DNA methylation is rapidly erased ( $\tilde{\mu}'$  large), but it is also quickly re-established by repressive histone modifications ( $r$  large). Thus, overexpression of JMJD2 leads to a fast erasure of repressive histone modifications. As a result, the absence of repressive histone modifications prevents the quick re-establishment of DNA methylation, leading to a rapid erasure of DNA methylation and reactivation of the OCT4 gene. In this case, using OCT4–JMJD2 overexpression for reprogramming could be more efficient and exhibit less variability in latency compared to OCT4–TET1 overexpression. For lower  $\tilde{\mu}'$ , overexpression of



**Fig. 6 JMJD2 and TET1 overexpression can have different effects on the reprogramming efficiency and latency variability.** **a** Left hand side plots: time trajectories of  $n_O^A$  (total amount of nucleosomes modified with activating chromatin modifications for the OCT4 gene) starting from the OCT4 fully repressed state ( $n_O^A = 0$ ) for different values of  $\bar{u}_J$ . Right hand side plots:  $\%O^A$ , that is, the normalized amount of  $N = 100$  time trajectories which reach  $n_O^A \geq 40$ , starting from  $n_O^A = 0$ , and  $\bar{f}_L$ , that is, the normalized frequency of a specific latency value across all the  $N = 100$  simulations. In all plots, on the x axis we have the time (days). The parameter values used for these simulations can be found in Supplementary Table 2. In particular, we consider three values of  $\bar{u}_J$  (i.e.,  $\bar{u}_J = 0.5, 50, 160$ ), and we set  $\bar{u}_T = 0, \bar{u}_O = 320, \tilde{\mu}' = 0.5, \epsilon_d = 0.2, \epsilon_e = 0.2, \bar{p}_O = \bar{p}_T = \bar{p}_J = 5, \eta = 0.1, \tilde{\mu} = 1$  and  $\epsilon' = 1$  (see Supplementary Fig. 5 for how the parameters  $\bar{u}_O, \epsilon_d, \epsilon_e, \tilde{\mu}'$ , and  $\bar{p}$  influence the impact of  $\bar{u}_J$  on the reprogramming process). **b**  $\%O^A$  for different values of  $\tilde{\mu}'$  and  $r$ . The parameter values used for these simulations can be found in Supplementary Table 2. In particular, we consider  $\bar{u}_T = 320, \bar{u}_J = 0$  (blue lines),  $\bar{u}_T = 0, \bar{u}_J = 320$  (red lines) and, for both cases  $\bar{u}_O = 160, \tilde{\mu}' = 1, 0.5, 0.2, r = 10, 5, 1, 0.2, \tilde{\mu} = 1, \epsilon_d = 0.16, \epsilon_e = 0.16, \bar{p}_O = \bar{p}_T = \bar{p}_J = 5, \eta = 0.1$  and  $\epsilon' = 1$  (see Supplementary Fig. 6 for how  $\epsilon_d, \epsilon_e$ , and the initial overexpression level influence the impact of  $\bar{u}_T, \bar{u}_J, \tilde{\mu}'$ , and  $r$  on the reprogramming process). In our model,  $\tilde{\mu}'$  quantifies the asymmetry between the erasure rates of DNA methylation and activating histone modifications and  $r$  the ratio between the rate at which repressive histone modifications enhance the establishment of DNA methylation through cross-catalytic reactions and the rate at which repressive histone modifications enhance their own establishment through auto-catalytic reactions. Mathematical definitions of  $r$  and  $\tilde{\mu}'$  can be found in (1) and (2), respectively. For all simulations, we implemented the reactions listed in Fig. 1 with the SSA<sup>43</sup> and we considered a time span of 21 days ( $\tau = 201.6$ ) and  $D_{\text{tot}} = 50$  (see Supplementary Note 1).

OCT4–JMJD2 becomes more effective even with lower values of  $r$  (Fig. 6b). In fact, if  $\tilde{\mu}'$  is low, even a high level of TET1 overexpression may be insufficient to erase DNA methylation. However, a similar level of JMJD2 overexpression could be sufficient to erase repressive histone modifications. Then, similarly to the previous case, if the enhancement of DNA methylation by H3K9me3 is significant ( $r$  is sufficiently high), erasing repressive histone modifications prevents the strong re-establishment of DNA methylation by H3K9me. This leads to an acceleration in DNA methylation erasure, resulting in a faster and less stochastic reactivation of the OCT4 gene.

Overall, these results suggest that, due to the enhancement of DNA methylation by the positive reinforcement with H3K9me3, the addition of JMJD2 to OCT4 overexpression may be more effective than the addition of TET1. However, for  $\tilde{\mu}'$  sufficiently high, OCT4–TET1 overexpression approach is always more efficient than the OCT4–JMJD2 overexpression approach.

In previously conducted experimental studies, it was observed that the addition of JMJD2 to the OSKM cocktail resulted in an approximate 1.75-fold increase in the number of iPSC colonies<sup>22</sup>. This efficiency improvement is smaller compared to the effect achieved by adding TET1 to the OKSM cocktail, resulting in a

twofold increase<sup>24</sup>, or even a minimum of tenfold increase when MBDs were knocked down<sup>21,31</sup>. These results indicate that in practical scenarios, we are likely to encounter a parameter regime where the enhancement of DNA methylation establishment through repressive histone modifications is not highly effective (low  $r$ ). According to our theoretical findings, this leads to a situation where the OCT4–TET1 overexpression reprogramming approach is more efficient compared to the OCT4–JMJD2 reprogramming approach. Furthermore, this efficiency improves even further when MBD is reduced or knocked down (higher  $\tilde{\mu}'$ ).

## DISCUSSION

In this work, we introduced the epigenetic OCT4 gene regulatory network (Epi OCT4 GRN), a network comprising a unique TF gene, OCT4, and two genes expressing chromatin modifiers, TET1 and JMJD2. The TF OCT4 and the two modifiers TET1 and JMJD2 are positively autoregulated and mutually activate each other, although through different mechanisms. Specifically, OCT4 self-activates and activates TET1 and JMJD2 by recruiting writers of activating chromatin modifications, while TET1 and JMJD2 self-activate and activate each other and OCT4 by erasing repressive



chromatin modifications. Each gene also has an inner chromatin modification circuit on which the chromatin modifiers act<sup>44</sup> ("Model of the epigenetic OCT4 gene regulatory network" subsection and Fig. 1c, d). Employing this model, we conducted a computational analysis to study the dynamics of reactivation of OCT4 during three reprogramming approaches based on overexpression of OCT4 alone, overexpression of OCT4 and TET1, and overexpression of OCT4 and JMJD2 by using Gillespie's Stochastic Simulation Algorithm (SSA) ("Results" section).

Our analysis indicates that, for the same level of OCT4 overexpression, the simultaneous overexpression of either TET1 or JMJD2 increases efficiency and reduces reactivation latency variability (Figs. 5a and 6a). The correlation between higher efficiency and reduced latency variability arises from the fact that increased levels of TET1 or JMJD2 lead to a greater probability of erasing repressive chromatin marks. As a result, within a given time span, more of these erasure events take place, reducing the stochastic nature of the process. Additionally, this results in the occurrence of the same number of events within a shorter timeframe. Consequently, the overall OCT4 reactivation process becomes more efficient and exhibits decreased latency variability. Comparing these two reprogramming approaches, our results suggest that the former is more efficient if the enhancement of DNA methylation establishment by H3K9me3 is sufficiently weak ( $r$  sufficiently low) and DNA methylation erasure is sufficiently fast ( $\bar{\mu}'$  sufficiently large) (Fig. 6b). By comparing previously obtained experimental data, the addition of TET1 seems to be more effective than the addition of JMJD2. This observation suggests that the parameter regime characterized by a sufficiently low value of  $r$  and a sufficiently large value of  $\bar{\mu}'$  is the more plausible scenario we may encounter. Physically, this parameter regime corresponds to a situation where the recruitment of DNA methylation by H3K9me3 is weak and the MBD proteins do not significantly hamper TET1 binding to methylated DNA.

We also compared the effectiveness of adding transient overexpression of TET1 or JMJD2, instead of constant overexpression (Supplementary Note 3). Our simulations show that transient overexpression of TET1 or JMJD2, although less effective than constant overexpression, still enhances the efficiency of the OCT4 reactivation process (Supplementary Figs. 7–10). Additionally, we studied the effectiveness of sequential, transient overexpression of TET1 and JMJD2. Our results show that transient overexpression of TET1 first and JMJD2 later can be as or more effective than constant overexpression of either one of the two erasure enzymes (Supplementary Figs. 11 and 12). This is because, in the parameter regime considered, the addition of JMJD2 to erase repressive histone modifications is not effective unless DNA methylation is quickly erased. To achieve a rapid erasure of DNA methylation, the addition of TET1 is necessary. More precisely, our computational analysis reveals that when the initial level of transient overexpression for both TET1 and JMJD2 is sufficiently high and matches the level of constant overexpression for either TET1 or JMJD2, the sequential transient overexpression of TET1 followed by JMJD2 can be nearly as effective as constant TET1 overexpression and more effective than JMJD2 constant overexpression. Moreover, the TET1–JMJD2 sequential transient overexpression with an initial overexpression level higher than that of constant TET1 or JMJD2 overexpression can show higher efficiency. This is practically relevant since cells may be able to tolerate higher overexpression levels if applied for a shorter duration. Overall, these results suggest that simpler reprogramming approaches, which involve the transient transfection of TET1 (and/or JMJD2) instead of genetically modifying the cells to obtain constant overexpression of TET1 (and/or JMJD2), could be a valuable option for enhancing reprogramming efficiency.

Another computational study was conducted with the aim of determining how the dosage of OCT4 overexpression affects the reprogramming process (Supplementary Note 2). The results

suggest that higher levels of OCT4 improve the process efficiency and reduce latency variability. However, this conclusion is based on a single TF model of the pluripotency GRN, while it is well known that the pluripotency GRN includes and requires the reactivation of two additional TFs: Nanog and Sox2<sup>32,62</sup>. These TFs are linked to OCT4 through mutual activation interactions. Models of this multi-TF GRN have shown that multiple stable steady states can arise and that the pluripotent state does not necessarily correspond to the highest OCT4 level<sup>63,64</sup> in accordance with experimental results in mouse<sup>65</sup> and human<sup>15,66</sup>. It is therefore plausible that an intermediate overexpression level of OCT4 may be preferable to increase reprogramming efficiency in a model that can capture multiple stable steady states, each corresponding to different lineages. Thus, in order to use this model not only to study the OCT4 reactivation dynamics, but also the dynamics of the PL GRN, future studies will need to combine the OCT4–Nanog–Sox2 GRN model<sup>63,67,68</sup>, with the chromatin modification circuit model to enable concurrent investigation of the effect of TF overexpression dosage and chromatin state on the PL GRN dynamics.

Finally, in our model, dilution due to cell growth and division is captured by an effective decay reaction with first-order kinetics, which is one of the standard models used<sup>69</sup>. However, more elaborated models for dilution may be implemented to account for the cell cycle and the binomial partitioning of molecules at cell division<sup>70,71</sup>. We then conducted an additional computational study, described in Supplementary Note 4, in which we removed the first-order decay reactions and introduced the binomial partitioning of molecules at the end of the cell cycle. This study reveals that, while the time trajectories generated by an Epi OCT4 GRN model with binomial partitioning are less smooth compared to those from our original model, the trend with which TET1 and JMJD2 affect OCT4 reprogramming efficiency and latency variability does not significantly change (Supplementary Figs. 14 and 15). Furthermore, it has been shown that OCT4 gene transcription is affected by a dosage compensation effect<sup>70</sup>. This can be qualitatively captured by our model by varying the effective expression rate of the gene (i.e.,  $\bar{p}_O$ ). If  $\bar{p}_O$  increases, due to imperfect compensation, then, a lower level of overexpression is required to achieve the same improvement in reprogramming efficiency (Supplementary Figs. 4, 5, 7, 9 and 13).

In conclusion, the model developed in this paper not only allows to mechanistically compare the effect of different reprogramming approaches on the OCT4 reactivation dynamics, but can also aid the rational design of new gene reactivation approaches and their application to cell fate reprogramming.

## METHODS

All simulations in this paper were conducted using Gillespie's Stochastic Simulation Algorithm (SSA)<sup>43</sup>.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## CODE AVAILABILITY

MATLAB codes used to generate the graphs in the paper are available on a GitHub repository at <https://github.com/simonbruno100/EpiOct4GRNpaper2023>.

Received: 12 July 2023; Accepted: 8 December 2023;

Published online: 06 January 2024

## REFERENCES

1. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).



2. Jopling, C., Boue, S. & Belmonte, J. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* **12**, 79–89 (2011).
3. Hirschi, K., Li, S. & Roy, K. Induced pluripotent stem cells for regenerative medicine. *Annu. Rev. Biomed. Eng.* **16**, 277–94 (2014).
4. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
5. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblast by defined factors. *Cell* **131**, 861–872 (2007).
6. Malik, N. & Rao, M. A review of the methods for human iPSC derivation. *Methods Mol. Biol.* **997**, 23–33 (2013).
7. Goh, P. et al. A systematic evaluation of integration free reprogramming methods for deriving clinically relevant patient specific induced pluripotent stem (iPS) cells. *PLoS ONE* **8**, e81622 (2013).
8. Schlaeger, T. et al. A comparison of non-integrating reprogramming methods. *Nat. Biotech.* **33**, 58–63 (2015).
9. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183–193 (2016).
10. Theunissen, T. & Jaenisch, R. Molecular control of induced pluripotency. *Cell Stem Cell* **14**, 729–734 (2014).
11. Carey, B. et al. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* **9**, 588–598 (2011).
12. Shu, J. et al. Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963–975 (2013).
13. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
14. Heng, B. & Fussenegger, M. Integration-free reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) without viral vectors, recombinant DNA, and genetic modification. *Methods Mol. Biol.* **1151**, 75–94 (2010).
15. Papapetrou, E. et al. Stoichiometric and temporal requirements of oct4, sox2, klf4, and c-myc expression for efficient human iPSC induction and differentiation. *PNAS* **106**, 12759–12764 (2009).
16. Tiemann, U. et al. Optimal reprogramming factor stoichiometry increases colony numbers and affects molecular characteristics of murine induced pluripotent stem cells. *Cytometry A* **79**, 426–35 (2011).
17. Meng, X. et al. Efficient reprogramming of human cord blood cd34+ cells into induced pluripotent stem cells with oct4 and sox2 alone. *Mol Ther.* **20**, 408–416 (2012).
18. Nagamatsu, G. et al. Optimal ratio of transcription factors for somatic cell reprogramming. *J. Biol. Chem.* **287**, 36273–36282 (2012).
19. Radziszewska, A. et al. A defined oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* **15**, 579–590 (2013).
20. Sui, D. et al. Fine-tuning of iPSC derivation by an inducible reprogramming system at the protein level. *Stem Cell Rep.* **2**, 721–733 (2014).
21. Rais, Y. et al. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
22. Chen, J. et al. H3k9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
23. Yamaguchi, S., Shen, L., Liu, Y., Sender, D. & Zhang, Y. Role of tet1 in erasure of genomic imprinting. *Nature* **504**, 460–464 (2013).
24. Gao, Y. et al. Replacement of oct4 by tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
25. Huisinga, K., Brower-Toland, B. & Elgin, S. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**, 110–122 (2006).
26. Allis, C., Caparros, M., Jenuwein, T. & Reinberg, D. *Epigenetics* 2nd edn (Cold Spring Harbor Laboratory Press, 2015).
27. Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
28. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
29. Wang, J., Jia, S. & Jia, S. New insights into the regulation of heterochromatin. *Trends Genet.* **32**, 284–294 (2016).
30. Penagos-Puig, A. & Furlan-Magaril, M. Heterochromatin as an important driver of genome organization. *Front. Cell Dev. Biol.* **8**, 579137 (2020).
31. Luo, M. et al. Nurd blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* **31**, 1278–1286 (2013).
32. Hanna, J. et al. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
33. Boland, M., Nazor, K. & Loring, J. Epigenetic regulation of pluripotency and differentiation. *Circ. Res.* **115**, 311–324 (2014).
34. Ludwig, A. et al. Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acid Res.* **45**, 2438–2457 (2017).
35. Radziszewska, A. & Silva, J. Do all roads lead to oct4? The emerging concepts of induced pluripotency. *Trends Cell Biol.* **24**, 275–284 (2014).
36. Shi, G. & Jin, Y. Role of oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res. Ther.* **1**, 39 (2010).
37. Hammachi, F. et al. Transcriptional activation by oct4 is sufficient for the maintenance and induction of pluripotency. *Cell Rep.* **1**, 99–199 (2012).
38. Whetstine, J. et al. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467–481 (2006).
39. Cloos, P. et al. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone h3. *Nature* **442**, 307–311 (2006).
40. Fodor, B. D. et al. Jmjd2b antagonizes h3k9 trimethylation at pericentric heterochromatin in mammalian cells. *Genes Dev.* **20**, 1557–1562 (2006).
41. Rasmussen, K. D. & Helin, K. Role of tet enzymes in DNA methylation, development, and cancer. *Genes Dev.* **30**, 733–750 (2016).
42. Jin, C. et al. Tet1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* **42**, 6956–6971 (2014).
43. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55 (2007).
44. Bruno, S., Williams, R. J. & Del Vecchio, D. Epigenetic cell memory: the gene's inner chromatin modification circuit. *PLoS Comput. Biol.* **18**, 1–27 (2022).
45. Artyomov, M. N., Meissner, A. & Chakraborty, A. K. A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.* **6**, e1000785 (2010).
46. Flöttmann, M., Scharp, T. & Klipp, E. A stochastic model of epigenetic dynamics in somatic cell reprogramming. *Front. Physiol.* **3**, 216 (2012).
47. Ashwin, S. S. & Sasai, M. Effects of collective histone state dynamics on epigenetic landscape and kinetics of cell reprogramming. *Nat. Rep.* **5**, 16746 (2015).
48. Chen, T., Al-Radhawi, M. & Sontag, E. A mathematical model exhibiting the effect of DNA methylation on the stability boundary in cell-fate networks. *Epigenetics* **16**, 436–457 (2021).
49. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–20 (2013).
50. Athanasiadou, R. et al. Targeting of de novo DNA methylation throughout the oct-4 gene regulatory region in differentiating embryonic stem cells. *PLoS ONE* **5**, e9937 (2010).
51. Huang, S., Litt, M. & Blakey, C. A. *Epigenetic Gene Expression and Regulation* (Academic Press, 2015).
52. Ang, Y.-S. et al. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
53. Loh, Y.-H., Zhang, W., Chen, X., George, J. & Ng, H.-H. Jmjd1a and jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21**, 2545–2557 (2007).
54. Wu, Y. et al. Oct4 and the small molecule inhibitor, sc1, regulates tet2 expression in mouse embryonic stem cells. *Mol. Biol. Rep.* **40**, 2897–2906 (2013).
55. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications—writers that read. *EMBO Rep.* **16**, 1467–1481 (2015).
56. Kim, T. & Buratowski, S. Dimethylation of h3k4 by set1 recruits the set3 histone deacetylase complex to 50 transcribed regions. *Cell* **137**, 259–272 (2009).
57. Brenner, C. et al. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **24**, 336–346 (2005).
58. Khoueiry, R. et al. Lineage-specific functions of tet1 in the postimplantation mouse embryo. *Nat. Genet.* **49**, 1061–1072 (2017).
59. KP, K. et al. Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
60. von Meyenn, F. et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).
61. Chen, T. & Dent, S. Y. R. Chromatin modifiers: regulators of cellular differentiation. *Nat. Rev. Genet.* **15**, 93–106 (2014).
62. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–24 (2007).
63. Del Vecchio, D., Abdallah, H., Qian, Y. & Collins, J. J. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Syst.* **4**, 109–120.e11 (2017).
64. Bruno, S., Al-Radhawi, M., Sontag, E. D. & Del Vecchio, D. Stochastic analysis of genetic feedback controllers to reprogram a pluripotency gene regulatory network. in *Proc of American Control Conference* 5089–5096 (2019).
65. Niwa, H., Miyazaki, J. I. & Smith, A. G. Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nat. Genet.* **24**, 372–376 (2000).
66. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for nanog, oct4, and sox2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440–454 (2012).

67. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
68. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
69. Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
70. Skinner, S. O. et al. Single-cell analysis of transcription kinetics across the cell cycle. *Elife* **5**, e12175 (2016).
71. Beentjes, C. H. L., Perez-Carrasco, R. & Grima, R. Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. *Phys. Rev. E* **101**, 032403 (2020).

## ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (NIH/NIBIB Grant Number R01EB024591) and by the National Science Foundation (NSF Collaborative Research grant MCB-2027949).

## AUTHOR CONTRIBUTIONS

D.D.V. designed the research. S.B. developed the mathematical models and performed the computational study. S.B. and D.D.V. analyzed the results. T.M.S. assisted with the interpretation and understanding of the results. S.B. and D.D.V. wrote the paper. S.B., T.M.S., and D.D.V. reviewed and contributed to the completion of this manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41540-023-00326-0>.

**Correspondence** and requests for materials should be addressed to Domitilla Del Vecchio.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024