Active turnover of genomic methylcytosine in pluripotent cells

Fabio Spada[1]*, Sarah Schiffers[1§], Angie Kirchner[1§], Yingqian Zhang[1,2], Gautier Arista[1], Olesea Kosmatchev[1], Eva Korytiakova[1], René Rahimoff[1§], Charlotte Ebert[1] and Thomas Carell[1]*

[1]Department of Chemistry, Ludwig Maximilians University Munich and Center for Integrated Protein Science Munich (CIPS[M])
[2]College of Chemisty, State Key Laboratory of Elemento-Organic Chemistry and Department of Chemical Biology, Nankai University, Tianjin, China

[§]Current addresses: S.S., National Cancer Institute, Center for Cancer Research, Bethesda, MD, USA; A.K., Cancer Research UK Cambridge Institute, UK; R.R., Department of Chemistry, University of California, Berkeley, CA, USA.

*Corresponding authors: Fabio Spada fabio.spada@cup.lmu.de; Thomas Carell: Thomas.Carell@cup.lmu.de

Abstract

Epigenetic plasticity underpins cell potency, but the extent to which active turnover of DNA methylation contributes to such plasticity is not known and the underlying pathways are poorly understood. Here we use metabolic labelling with stable isotopes and mass spectrometry to quantitatively address the global turnover of genomic methylcytidine (mdC), hydroxymethylcytidine (hmdC) and formylcytidine (fdC) across mouse pluripotent cell states. High rates of mdC/hmdC oxidation and fdC turnover characterize a formative-like pluripotent state. In primed pluripotent cells the global mdC turnover rate is about 3-6% faster than can be explained by passive dilution through DNA synthesis. While this active component is largely dependent on Tet-mediated mdC oxidation, we unveil additional oxidation-independent mdC turnover, possibly through DNA repair. This process accelerates upon acquisition of primed pluripotency and returns to low levels in lineage committed cells. Thus, in pluripotent cells active mdC turnover involves both mdC oxidation-dependent and -independent processes.

Introduction

Methylation at position 5 of deoxycytidine (dC) is the most prominent nucleobase modification in higher eukaryotic genomes[1]. In mammals, genomic dC is methylated to 5-methyl-2'-deoxycytidine (mdC) by the DNA (cytosine-5)-methyltransferases DNMT1, DNMT3a and DNMT3b, mostly, but not exclusively in the context of CpG dinucleotides[1,2]. These enzymes differentially contribute to establishing and maintaining genomic mdC patterns across cellular and organismal generations. Genomic mdC is further modified by iterative oxidation to 5-hydroxymethyl-, 5-formyl- and 5-carboxyl-dC (hmdC, fdC and cadC, respectively) through the action of Ten-Eleven Translocation (TET) di-oxygenases[3–5]. Several classes of DNA binding factors display differential affinity for their target sequences depending on the dC modification state[1]. This is why genomic patterns of dC modification contribute to consolidation and propagation of gene expression states and chromatin structure, thus affecting fundamental processes such as cell fate determination, imprinting, X chromosome inactivation and genome stability[1].

In the last two decades methylation of genomic dC was shown to be rapidly removed in various biological contexts[6]. Several erasure mechanisms were proposed, including prevention of methylation maintenance, which results in passive dilution of the modified base through DNA replication, and active demethylation by enzymatic processes[6] (Fig. 1a). Among the latter, oxidation of mdC to fdC or cadC by TET enzymes, followed by base excision repair (BER), mediated by the DNA glycosylases TDG[5,7] and Neil1/2[8,9], is best characterized. Additional proposals invoke the involvement of other DNA repair pathways[6] or direct C-C bond cleavage[10–14]. A third line of proposal posits deamination of genomic mdC and hmdC to thymidine (dT) and 5-hydroxymethyl-2'-deoxyuridine (hmdU), respectively, either by a cytosine deaminase of the AID/APOBEC family[15–17] or by *de novo* DNA methyltransferases DNMT3a/b[18]. The resulting T:G or hmU:G mismatches would be then resolved by BER or non-canonical mismatch repair (ncMMR)[15,16,19,20]. Finally, it was suggested that unmodified dC could be deaminated by AID and that long patch BER or ncMMR could lead to indirect co-removal of adjacent ^{m}C residues[17,19,20]. Currently, all these pathways are highly controversial. Conclusive evidence for their functional relevance *in vivo* is lacking[6]. It is however conceivable that different pathways operate at different stages during development.

During early embryonic development, the epiblast transits through a spectrum of pluripotent states[21]. This transition can be modelled by cultivating mouse pluripotent stem cells (mPSCs) under specific conditions, which recapitulates dynamic expression and epigenetic changes, including a major global gain of mdC[21]. Conversely, reversion from partially and heterogeneously primed mPSC cultures to the naïve state involves downregulation of *de novo*

methyltransferases and collapse of DNA methylation maintenance with consequent global and passive (Tet enzyme-independent) loss of genomic methylation[22]. In contrast, little is known about the extent and nature of the turnover of modified genomic cytosines during the forward, developmentally relevant transition from naïve to primed pluripotent states. Here, we used metabolic labelling with stable isotope standards for highly accurate mass spectrometry based quantification of the global genomic mdC, hmdC and fdC turnover through a progression of pluripotent states. We show that upon acquisition of primed pluripotency roughly 3-6% of the methylome is actively turned over. Most of this active turnover is due to Tet-mediated mdC oxidation, while a minor part involves an unknown mdC oxidation-independent DNA repair process.

Results

Genomic mdC oxidation kinetics decrease in primed mPSCs

We cultured mPSCs under specific conditions to model the progression of the embryonic epiblast from naïve to primed pluripotency and investigate the kinetics and turnover of dC modification during this transition. In order to monitor priming conditions, we first made use of a mPSC line expressing an Oct4-YFP reporter[23]. After adaptation to naïve conditions in the presence of Gsk3 inhibitor CHIR99021 (CHIR), Mek1/2 inhibitor PD0325901 (also known as 2i) and LIF, the cells were plated at low density in serum-containing medium supplemented with CHIR and the Tankyrase inhibitor IWR1-endo (serum/CR) as previously described[24]. Live cell microscopy over the course of 6 days without passaging showed that these cells rapidly formed monolayer colonies, which retained a remarkably homogeneous nuclear YFP signal across the culture (Extended Data Fig. 1a), indicating that under these conditions mPSC cultures may not lose pluripotency. Similarly intense and homogeneous YFP signals were observed when the same reporter line was permanently primed to mouse Epiblast Stem Cells (mEpiSCs) in serum-free, ~~chemically defined~~ medium supplemented with CHIR, IWR1, FGF-2 and Activin A (CDM/CRFA; Extended Data Fig. 1a). Expression analysis of pluripotency factors in serum/CR cultures and CDM/CRFA EpiSCs revealed similarly reduced and increased transcript levels of naïve and primed pluripotency factors, respectively (Extended Data Fig. 1b), showing that these conditions support similarly primed states. Thus, even without prolonged passaging or in chemically defined conditions, CHIR/IWR1 treatment supports primed pluripotency.

We then used ultra-high pressure liquid chromatography coupled to triple quadrupole mass spectrometry (UHPLC-MS^2)[25] to analyse global levels of genomic mdC and hmdC upon priming of mPSCs under the same conditions over five days. As previously reported[26], the

transition from the naïve (2i/LIF) to a homogeneously primed pluripotent state involved a large increase of the global mdC levels, while the hmdC abundance dropped sharply. Both modifications reaching plateau levels by day 4 (Extended Data Fig. 2a). Global levels of genomic mdC and hmdC, similar to those in serum/CR cultures, were detected in Oct4-YFP mEpiSCs under serum-free conditions, excluding potential confounding effects of undefined serum composition. Note that under both serum-containing and serum-free conditions, we avoided supplementation with ascorbic acid, which greatly increases the levels of oxidized mdC derivatives[27]. The decrease of hmdC upon priming is consistent with a recent proteomic analysis showing that the levels of chromatin bound Tet1 and 2 proteins are markedly reduced in mEpiSCs compared to naïve mPSCs, while Tet3 protein is undetectable in both states[28].

Notably, replacing the Mek inhibitor with the Src inhibitor CGP77675 (conditions referred to as alternative 2i or a2i) results in relatively high genomic mdC levels, while gene expression patterns remain similar to those of naïve mPSCs [29,30]. Our analysis showed that the expression pattern of pluripotency and early lineage specification factors under a2i/LIF conditions is intermediate between that of naïve and primed states. We found high levels of Tfcp2l1 and Klf4 transcripts relative to 2i/LIF (naïve) conditions, intermediate expression of Rex1, Stella and Prdm14, and low levels of Fgf5, Oct6, Otx2 and Dnmt3b relative to primed conditions (Extended Data Fig. 1b). This pattern is similar to that present in formative pluripotency, an intermediate state along the trajectory from naïve to primed pluripotency, during which competence for multi-lineage commitment is acquired[31]. High depth bisulfite amplicon sequencing revealed a gradual increase of CpG methylation at secondary imprints during the transition from a2i/LIF to serum/CR conditions (Extended Data Fig. 2b), as observed in early embryos[32]. Altogether, these data support that cultures transferred from a2i/LIF to serum/CR conditions recapitulate fundamental molecular processes of the progression to a genuine primed pluripotent state. Importantly, a2i/LIF and serum/CR conditions allow comparison of genomic cytosine modification dynamics between functionally distinct pluripotent states with substantial and comparable steady state levels of genomic mdC. Thus, we adopted a2i/LIF conditions both for routine mPSC maintenance and as starting state for further time course analyses of mPSC priming.

We then monitored global cytosine modification kinetics during acquisition of primed pluripotency by metabolic labelling of the C5 substituent. To this end, we supplemented L-Methionine (Met)-free medium with isotopically labelled Met (m+4Met), whereby the methyl group provides a m+4 mass shift [*methyl*-^{13}C, d_3]. As for naturally occurring Met, m+4Met is incorporated into S-adenosylmethionine (SAM) and the $^{13}CD_3$ (m+4) is then transferred to genomic dC by Dnmts, generating m+4dC. The latter can be sequentially oxidised by Tet proteins to hm+3dC, f+2dC and ca+1dC (Fig. 1a)[25].

For accurate quantification of natural mdC, hmdC, fdC and heavy m+4dC, hm+3dC, f+2dC, we synthesised a corresponding series of even heavier isotopologues (Fig. 1b-[1-3]) and used these as internal standards for quantification by UHPLC-MS^2. In this work cadC and ca+1dC were not quantified because a difference of one mass unit cannot be discriminated. We first compared the modification kinetics in cultures maintained constantly under a2i/LIF conditions to cultures undergoing priming in CHIR/IWR1 medium. In both cases we performed time course analyses over 2 days of labelling in the presence of m+4Met, which in the case of priming cultures was from day 3 to 5 after transfer to serum/CR medium (Fig. 1c,d; Extended Data Fig. 3). In priming cultures the rate of m+4dC oxidation to hm+3dC was clearly slower than under a2i/LIF conditions and accumulation of the further oxidized state f+2dC became undetectable upon priming. Thus, the transition to primed pluripotency determines a major decrease in global oxidation of genomic mdC, which is consistent with the drastically reduced steady state levels of genomic hmdC (Extended Data Fig. 2a).

Oxidative turnover of genomic mdC in mPSCs

To monitor the turnover of modified genomic cytosines we reversed the labelling schema described above. We first labelled cultures with m+4Met for five days and, after transferring the cultures to medium with natural Met, we measured the disappearance of the label over two days (Fig. 2a). Again, we compared cultures maintained constantly under serum/a2i/LIF conditions to cultures undergoing priming in serum/CR medium. For primed cultures the first two pulse days were in a2i/LIF medium, while the last three were in serum/CR medium, thus corresponding to the first three days of priming. To generate a reference for passive dilution of a genomic nucleoside through DNA synthesis, we co-labelled with an isotopically labelled 2'-deoxythymidine having a mass shift of 12 units ([$^{15}N_2$,$^{13}C_{10}$]-dT; hereafter referred to as dT+12). Due to its highly efficient incorporation, dT+12 was co-pulsed only during the last day of the m+4Met pulse (Fig. 2a). Thereafter, dT+12 was chased together with m+4Met. Remarkably, while in most cases the levels of modified cytosines readily declined with the onset of the chase, under a2i/LIF conditions hm+3dC and f+2dC accumulated for the first 12 and 24 hours, respectively (Fig. 2b). This shows again that the rate of Tet-mediated oxidation is quite high in mPSCs under a2i/LIF conditions and declines sharply upon priming, as already seen with the time course analysis during m+4Met labelling (Fig. 1b,c). Notably, under a2i/LIF conditions the decline of f+2dC after 36 h is steep, pointing to high turnover of fdC. Although under priming conditions oxidation of mdC to hmdC and fdC was hardly detectable in labelling time course experiments (Fig. 1), in pulse chase experiments the decline of labelled hmdC and fdC was slower than that of dT+12, clearly showing that some mdC and hmdC oxidation still takes place (Fig 2b). Importantly, upon priming m+4dC turnover was faster than the dilution

rate of dT+12 (Fig. 2c; Extended Data Fig. 4a), revealing the presence a component that cannot be explained by passive dilution through DNA synthesis.

To assess the contribution of Tet-dependent oxidation to this active component, we performed the same pulse/chase experiment under priming conditions (Fig. 2a, upper part) with E14tg2a mPSCs lacking all three Tet proteins (Tet triple knockout, alias Tet TKO)[33] as well as parental, Tet 1-3 proficient E14tg2a cells (Fig. 3). In the parental cell line, turnover of genomic m+4dC was again clearly faster than passive dilution of dT+12. After averaging the data from three such pulse chase experiments with wt mPSC lines, a paired two-tailed T-test showed high significance (P = 0.0053) for the deviation between m+4dC and dT+12 decay rates. This accounts for an active component of global mdC turnover involving 3-6% of the methylome, depending on the time point (Fig. 2f). In contrast, in the genome of Tet TKO cells the decline rate of m+4dC was essentially indistinguishable from that of dT+12 (Fig. 3). This shows that the active component of the mdC turnover is indeed largely dependent on the presence of Tet enzymes, and thus mostly due to mdC oxidation.

Additional non-oxidative mdC turnover in mPSCs

Next to oxidative mdC turnover, we investigated potential deamination of mdC to dT (mdU)[15]. Upon metabolic labelling with m+4Met, deamination of m+4dC would generate m+4dU, that is dT bearing the m+4 methyl group (Fig. 1a). In order to investigate this process quantitatively we synthesised the isotopologue dT-[$^{13}C_5$,$^{15}N_2$] (Fig. 1b-[4]) and used it as a standard for UHPLC-MS^2 analysis. As the source of methyl group for *de novo* dT biosynthesis is serine through 5,10-methylenetetrahydrofolate and not methionine through SAM, it is expected that, upon supplementation of the medium with m+4Met, m+4dC is the only source of m+4dU. To verify this expectation we made use of mPSCs lacking all catalytically active Dnmts (Dnmt TKO), and therefore devoid of genomic mdC[34]. After priming in the presence of m+4Met, m+4dU was clearly detected in the genome of wt mPSCs, but not in Dnmt TKO cultures (Extended Data Fig. 3). Thus, generation of m+4dU strictly requires the formation of genomic m+4dC, showing that m+4dU originates exclusively through deamination of m+4dC.

Direct deamination of genomic mdC would generate T:G mismatches, which, if not efficiently repaired, would be highly mutagenic. Despite this, we detected substantial accumulation of m+4dU in genomic DNA (gDNA), whenever cultures were provided with m+4Met (Fig. 1-5, Extended Data Fig. 3-5,7). To explain this observation we considered an alternative scenario in which m+4dC is released from the genome as soluble m+4dC/m+4dCMP, possibly through a DNA repair process. Deamination of soluble m+4dC/m+4dCMP by Cytidine and Deoxycytidylate deaminases (Cda and Dctd)[35], would generate soluble m+4dU/m+4dUMP,

which could be phosphorylated and re-incorporated into the genome through DNA synthesis (Fig. 4a). To test whether m+4dU is generated along this pathway, we supplemented the medium with m+4Met and increasing concentrations of unlabelled dT. The latter was added to outcompete the incorporation of soluble m+4dU/m+4dUMP into gDNA and to allosterically inhibit Dctd[36]. Indeed, we detected a dT dose-dependent reduction of genomic m+4dU in priming mPSC cultures (Fig. 4b). This result supports the idea that genomic m+4dU originates in large part from a pool of soluble precursor.

To probe whether deamination of m+4dC takes place within gDNA or at the level of soluble nucleoside/nucleotide pools, we generated mPSC with double deficiency for Cytidine deaminase (Cda) and Deoxycytidylate deaminase (Dctd) through CRISPR/Cas9-mediated gene editing (hereafter referred to as CD-DKO; Supplementary Fig. 1a,b). Genomic sequencing showed that three of the isolated clonal lines bear compound biallelic *Cda* and *Dctd* mutations incompatible with the expression of functional enzymes (Supplementary Figures 1c-e and 2a,b). All three CD-DKO clones grew at slower rates than parental wt cells, a phenotype which was at least partially relieved by supplementation of dT (Supplementary Fig. 1f). This indicates that deamination of (d)C/dCMP by Cda and/or Dctd provides substantial amounts of (d)U/dUMP for biosynthesis of (d)T/dTMP by Thymidine synthase to support DNA replication. This observation is fully consistent with previously reported dT auxotrophy of Dctd deficient somatic cell lines[37]. As two of the CD-DKO clones grew particularly poorly under serum/CR priming conditions, we first used serum/LIF medium for metabolic labelling with m+4Met under priming conditions. Importantly, while the genomic content of mdC/m+4dC was the same as in parental wt cells, only background levels of m+4dU could be measured in all three CD-DKO clones (Fig 4c). The same result was obtained with CD-DKO clone 40 upon priming in serum/CHIR/IWR1 medium (Fig. 4d). These results clearly show that all measurable genomic m+4dU originates from a process involving direct release of m+4dC/m+4dCMP from the genome, its deamination to m+4dU/m+4dUMP in the soluble pool and (re)-incorporation into the genome through phosphorylation and DNA synthesis.

Although our data do not support deamination of mdC directly within the genome, it is important to consider the limitations of our UHPLC-MS^2 method. In UHPLC the overwhelming amount of natural dT present in the genome co-elutes with m+4dU and outcompetes it during ionization for the capture of charge. This strong ion suppression effect reduces the m+4dU signal and results in a high limit of detection (LOD). To put this into perspective, our LOD for m+4dU ($1x10^{-5}$ per nucleotide) is about 10-100 times higher than typical global levels of fdC, which is a relatively stable dC modification with an epigenetically relevant role[38,39]. In regard to this and previous reports of direct mdC deamination within gDNA[15,18], we cannot fully

exclude that deamination of a few m+4dC residues within the genome trigger long patch BER and/or ncMMR[19,20], leading to the release of larger amounts of m+4dC.

In order to minimize ion suppression and gain sensitivity, we double labelled wt and CD-DKO mPSCs with m+4Met and a dC nucleoside where 9 hydrogen atoms are replaced with deuterium. After incorporation into the genome, dC[D_9] is methylated to m+4dC[D_8], which deaminates to m+4dU[D_8] (a dT with eleven D atoms and a +12 mass shift; Extended Data Fig. 6a). This massive H-to-D exchange was expected to the shift UHPLC retention time slightly, allowing m+4dU[D_8] to escape co-elution with dT. Indeed, co-injection tests showed a slightly shorter retention time for dT[D_8] relative to dT (Extended Data Fig. 6b). For m+4dU[D_8] we saw an even shorter retention time. After double labelling with m+4Met and dC[D_9] we detected m+4dU[D_8] confidently in wt mPSCs (Extended Data Fig. 6c). In addition, we sporadically detected small m+4dU[D_8] peaks in technical replicates of CD-DKO samples. After integration of these sporadic signals, the corresponding values were invariably below LOD. Although we cannot exclude that these signals are caused by background processes, taking them into consideration allows us to estimate between 500 and 5000 genomic deamination events per cell in the course of our labelling experiment. According to an empirical rate constant for spontaneous hydrolytic deamination of mdC in solution[40], only around 20 such events would be expected to take place. Therefore, we cannot exclude the occurrence of a small amount of genomic deamination above the rate of spontaneous hydrolysis at levels of $1x10^{-6}$-$1x10^{-7}$ per nucleotide.

Tet enzymes do not trigger mdC-to-dT turnover

Conversion of m+4dC into m+4dU does not require oxidized intermediates. However, it was recently suggested that in the zygote Tet3 generates undefined DNA lesions, triggering repair-mediated erasure of genomic methylation from the paternal genome[41]. Also, hmdC may mark DNA damage sites and promote their repair[4242] or even elicit DNA repair by selectively recruiting endonucleases like Srap1[43] and Endonuclease G[44]. We therefore considered whether Tet proteins contribute to the turnover of genomic mdC into dT by generating hmdC or low levels of DNA lesions such as hmdU[25], which triggers BER initiated by the DNA glycosylase Smug1. To this aim we primed Tet TKO and parental E14tg2a mESCs in the presence of m+4Met. Surprisingly, higher levels of genomic m+4dU were detected in the absence of Tet enzymes (Extended Data Fig. 7). Similarly, analysis of the pulse chase experiment also revealed a tendency to higher accumulation of m+4dU in Tet TKO mPSCs at later chase time points (Fig. 3b). Therefore, in pluripotent cells Tet proteins do not seem to trigger DNA repair events that may contribute to turnover of mdC into dT.

mdC-to-dT turnover is developmentally regulated

To investigate whether oxidation-independent mdC turnover is regulated during the transition to primed pluripotency, we first compared the genomic abundance of m+4dU in the time course and pulse-chase experiments under a2i/LIF and priming conditions described above (Fig. 1-3, Extended Data Fig. 3). In time course experiments, progressive accumulation m+4dU was much faster and about five times higher under priming conditions (Fig. 1, Extended Data Fig. 3). Also, m+4dU was the only m+4dC derivative that accumulated during label chase under priming conditions, though to different extents in the two cell lines tested (Fig. 2e,3b and Extended Data Fig. 4b). Thus, in contrast to the observed decline of mdC oxidation (Fig. 1,2, Extended Data Fig. S3), these results reveal increased rates of mdC-to-dT turnover upon acquisition of primed pluripotency. This anti-correlation suggests that oxidation-independent mdC turnover may serve at least in part as a supplementary pathway for the erasure of dC methylation. However, this process cannot involve more than 1% of the methylome, as in primed Tet TKO mPSCs there was no apparent difference between global mdC turnover and DNA replication-dependent passive dilution (Fig. 3a: decay of m+4dC and dT+12, respectively).

At last, we compared oxidation-independent mdC-to-dT turnover in pluripotent states with various lineage committed cell types, including primary and immortalized embryonic fibroblasts, neural stem cells (NSCs)[45] and HeLa cells. Although primed pluripotent cells were labelled with m+4Met for only two to three days, m+4dU reached higher levels in their genome than in those of any of the somatic cell types labelled for substantially longer periods, ranging from 6 to 12 days (Fig. 5a). This clearly shows that in lineage committed cells oxidation-independent mdC turnover is substantially lower than in primed pluripotent cells. In addition, the rate of mdC-to-dT turnover is clearly not proportional to the global genomic abundance of mdC either between pluripotent states or among pluripotent and lineage committed cells. In particular, among the latter, mdC-to-dT turnover is clearly more active in NSCs than in fibroblasts, despite the mdC genomic content being at least 20% lower in NSCs. These observations indicate that mdC-to-dT turnover is under developmental and cell lineage-dependent control.

Discussion

Recent methylome analysis of matched clonal and polyclonal cell populations[46] and studies on the contribution of Dnmt3 proteins to DNA methylation maintenance[47,48] suggested that in heterogeneously primed mPSCs cultures (serum/LIF) genomic mdC undergoes higher

dynamic turnover as compared to differentiated somatic cells. However, the extent and especially the nature of this turnover have not been addressed. Previously, Bachmann *et al.* reported a pulse-chase experiment with m+4Met and a short pulse time to estimate the global turnover of genomic mdC and hmdC in mPSCs under naïve conditions (2i/LIF)[49], where the levels of dC modifications are far lower.

Here, we report the analysis of on- and off-rate kinetics for genomic dC modifications upon transition to primed pluripotency. We show that in mPSCs under a2i/LIF conditions, which display expression and genomic methylation patterns that are intermediate between naïve and primed pluripotent states (formative state)[31], the rates of mdC oxidation to hmdC and further to fdC are quite sustained. The removal/conversion of fdC is very rapid. This points to a very dynamic equilibrium for fdC under these conditions. It is tempting to speculate that these high levels of oxidative turnover may underlie the frequent periodical oscillations in global mdC levels recently identified through single cell analysis in partially (and heterogeneously) primed serum/LIF cultures[50]. In contrast to the high levels of mdC oxidation in a2i/LIF conditions, upon transition to primed pluripotency, the global oxidation rates of mdC and hmdC as well as the removal of hmdC and fdC decline drastically. Despite this, the global turnover rate of genomic mdC is 3-6% faster than can be accounted for by passive dilution through DNA synthesis (DNA replication and repair; Fig. 5b). This difference is largely due to Tet-mediated mC oxidation. Evidence for passive erasure of dC methylation based on impairment of its maintenance has been provided for primordial germ cells, preimplantation embryonic development and the reversion from heterogeneous partially primed/formative states present in serum/LIF cultures to the naïve pluripotent state[1,22]. We suggest that against the backdrop of global gain and maintenance of dC methylation in the forward developmental transition from naïve to primed pluripotency an active demethylation mechanism may be more attainable than preventing maintenance of dC methylation at selected sites. This could be especially the case at loci with a highly accessible chromatin conformation like actively transcribed genes.

By tracing the origin of genomic m+4dU to the deamination of soluble m+4dC, we show that a small part of genomic mdC is turned over by a process that involves the direct excision of mdC from the genome (Fig. 5b). Although the nature of this process and the events triggering it remain to be determined, we speculate that DNA repair is responsible for the release mdC. While we show that TET proteins are clearly not involved in triggering or targeting this non-oxidative mdC turnover, our data neither support nor rule out that low levels of direct enzymatic deamination of genomic mdC may trigger a DNA repair process underpinning the turnover pathway. Importantly, this process is not proportional to global genomic mdC levels, either between pluripotent states, among lineage committed cell types or between pluripotent and committed cells, indicating that it could be developmentally regulated. Interestingly, the levels of oxidation-independent mdC turnover show an inverse correlation with mdC oxidation rates

across pluripotent states and seem to be slightly more sustained in Tet deficient cells, pointing to a possible function as alternative or supplementary mechanism dedicated to active erasure of cytosine methylation.

In summary, we show that upon priming of pluripotent cells a significant fraction of mdC (3-6%) vanishes after oxidation to hmdC, fdC and cadC. A smaller fraction of mdC is removed from the genome by cleavage of phosphodiester bonds, likely as part of an unknown DNA repair process. Once released in the soluble pool, mdC is deaminated to dT, which is reintegrated into the genome. Finally our data are inconclusive regarding a very low level of genomic mdC deamination, which would generate dT:dG mismatches and trigger DNA repair.

Acknowledgements

We are very thankful to the following colleagues: Markus Möser (Max Planck Institute for Biochemistry, Martinsried, Germany) for K3 mPSCs (Kindlin3+/+); Hitoshi Niwa and Masaki Okano (both at Kumamoto University, Japan) for Oct4-YFP-Puro E14tg2a and Dnmt TKO J1 mPSCs, respectively; Yi Zhang (Boston Children's Hospital, Boston, MA) for parental and Tet TKO E14tg2a, Sebastian Bultmann and Christopher Mulholland (both at Ludwig Maximilian University, Munich) for guidance on high depth bisulfite amplicon sequencing. Anti-DCTD antibody and purified recombinant DCTD protein were generous gifts from Frank Maley (New York State University, NY).

Funding was provided by the Deutsche Forschungsgemeinschaft via the programs SFB1309 (TP: A4), SFB1361 (TP: 2), GRK 2338 and SPP-1784. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912). YZ is supported by the China Scholarship Council (CSC Nr. 201806200069).

Author Contributions

FS and TC conceived the study and designed experiments. FS, SS, AK, YZ, GA and OK performed experiments and analyzed data. FS, TC and SS interpreted data. RR, CE and EK synthesized isotopically labelled nucleosides used as standards for LC-MS^2. FS, TC and SS wrote the manuscript.

Competing Financial Interests Statement

The authors declare no competing interests.

References

1. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 1–18 (2019) doi:10.1038/s41580-019-0159-6.
2. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annual Review of Genomics and Human Genetics* **16**, 55–77 (2015).
3. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930–935 (2009).
4. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, 1303–1307 (2011).
6. Schuermann, D., Weber, A. R. & Schär, P. Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair* **44**, 92–102 (2016).
7. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry* **286**, 35334 (2011).
8. Slyvka, A., Mierzejewska, K. & Bochtler, M. Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Scientific Reports* **7**, 9001 (2017).
9. Schomacher, L. *et al.* Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature Structural & Molecular Biology* **23**, 116–124 (2016).
10. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. The Mammalian de novo DNA Methyltransferases Dnmt3a and Dnmt3b Are Also DNA 5-Hydroxymethyl Cytosine Dehydroxymethylases. *J. Biol. Chem.* **287**, 33116–33121 (2012).
11. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D. & Klimašauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology* **5**, 400–402 (2009).
12. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C–C bond cleavage *in vivo*. *Nature Chemical Biology* **14**, 72–78 (2017).
13. Liutkevičiūtė, Z. *et al.* Direct Decarboxylation of 5-Carboxylcytosine by DNA C5- Methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
14. Schiesser, S. *et al.* Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing. *Angewandte Chemie International Edition* **51**, 6516–6520 (2012).
15. Rai, K. *et al.* DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. *Cell* **135**, 1201–1212 (2008).
16. Cortellino, S. *et al.* Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146**, 67–79 (2011).
17. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics & Chromatin* **6**, 39 (2013).
18. Metivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
19. Grin, I. & Ishchenko, A. A. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucl. Acids Res.* **44**, 3713–3727 (2016).
20. Franchini, D.-M. *et al.* Processive DNA Demethylation via DNA Deaminase-Induced Lesion Resolution. *PLoS ONE* **9**, e97754 (2014).
21. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews Genetics* **18**, 643–658 (2017).
22. von Meyenn, F. *et al.* Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell* **62**, 848–861 (2016).
23. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
24. Kim, H. *et al.* Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat Commun* **4**, 2403 (2013).
25. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**, 574–581 (2014).
26. Shirane, K. *et al.* Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Developmental Cell* **39**, 87–103 (2016).
27. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

28. van Mierlo, G., Wester, R. A. & Marks, H. A mass spectrometry survey of chromatin-associated proteins in pluripotency and early lineage commitment. *Proteomics* **19**, e1900047 (2019).
29. Yagi, M. *et al.* Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
30. Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
31. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**, 1221–1234 (2017).
32. Wang, L. *et al.* Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell* **157**, 979–991 (2014).
33. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28**, 2103–2119 (2014).
34. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–14 (2006).
35. Jekunen, A., Puukka, M. & Vilpo, J. Exclusion of exogenous 5-methyl-2'-deoxycytidine from DNA in human leukemic cells: A study with [2-14C]- and [methyl-14C]5-methyl-2'-deoxycytidine. *Biochemical Pharmacology* **32**, 1165–1168 (1983).
36. Maley, G. F., Lobo, A. P. & Maley, F. Properties of an affinity-column-purified human deoxycytidylate deaminase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1162**, 161–170 (1993).
37. de Saint Vincent, B. R., Déchamps, M. & Buttin, G. The modulation of the thymidine triphosphate pool of Chinese hamster cells by dCMP deaminase and UDP reductase. Thymidine auxotrophy induced by CTP in dCMP deaminase-deficient lines. *J. Biol. Chem.* **255**, 162–167 (1980).
38. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol* **11**, 555–557 (2015).
39. Bilyard, M. K., Becker, S. & Balasubramanian, S. Natural, modified DNA bases. *Current Opinion in Chemical Biology* **57**, 1–7 (2020).
40. Shen, J.-C., Rideout, W. M. & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucl. Acids Res.* **22**, 972–976 (1994).
41. Ladstätter, S. & Tachibana-Konwalski, K. A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming. *Cell* **167**, 1774-1787.e13 (2016).
42. Jiang, D., Wei, S., Chen, F., Zhang, Y. & Li, J. TET3-mediated DNA oxidation promotes ATR-dependent DNA damage response. *EMBO reports* **18**, 781–796 (2017).
43. Kweon, S.-M. *et al.* Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell Reports* **21**, 482–494 (2017).
44. Robertson, A. B., Robertson, J., Fusser, M. & Klungland, A. Endonuclease G preferentially cleaves 5-hydroxymethylcytosine-modified DNA creating a substrate for recombination. *Nucl. Acids Res.* **42**, 13280–13293 (2014).
45. Liu, N. *et al.* Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. *PLoS ONE* **8**, e62755 (2013).
46. Shipony, Z. *et al.* Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014).
47. Ooi, S. K. *et al.* Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics & Chromatin* **3**, 17 (2010).
48. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**, 5594–605 (2003).
49. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* **6**, 1049–1055 (2014).
50. Rulands, S. *et al.* Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell Systems* **7**, 63-76.e12 (2018).

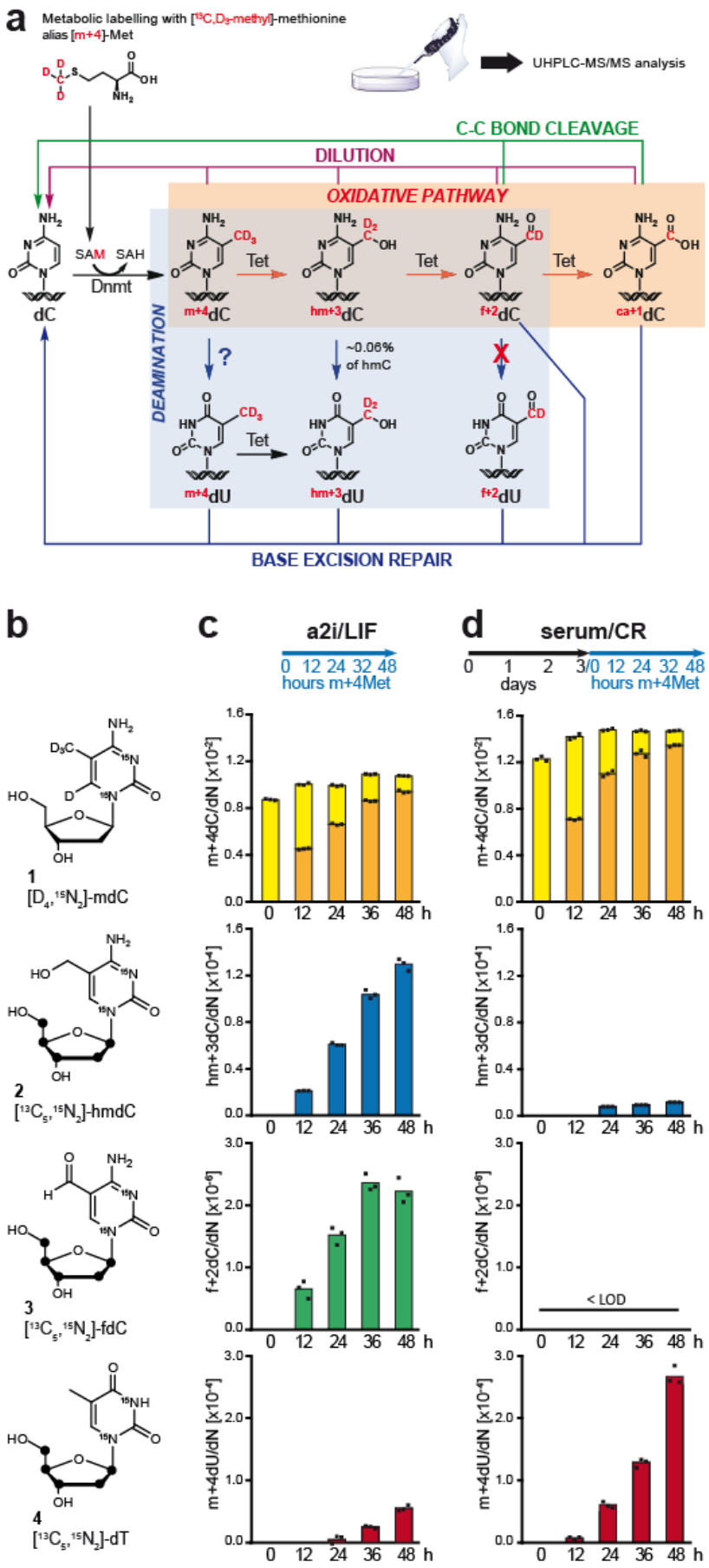


Figure 1. Drastic reduction of global genomic mdC oxidation and increase of mdC to mdU conversion upon transition to primed pluripotency.

a) Proposed de-modification pathways for genomic C5-cytosine modifications and expected labelled mdC derivatives upon metabolic labelling with m+4Met. **b**) Nucleoside isotopologues used as internal standards for UHPLC-MS^2 analysis. Black dots in the ribose represent ^{13}C atoms. **c** and **d**) Time course analysis of genomic dC derivatives upon metabolic labelling with m+4Met under a2i/LIF (c) and priming conditions (serum/CR; d). The labelling/time course schedules are shown at the top. Global levels of unlabelled mdC (pale yellow) and m+4dC (dark yellow), hm+3C (blue) and f+2C (green) are shown as individual values of replicate measurements (n=3: black squares) and their means (bars) from single samples. An independent biological replicate for both c and d is shown in Extended Data Fig. 3. LOD = limit of detection.

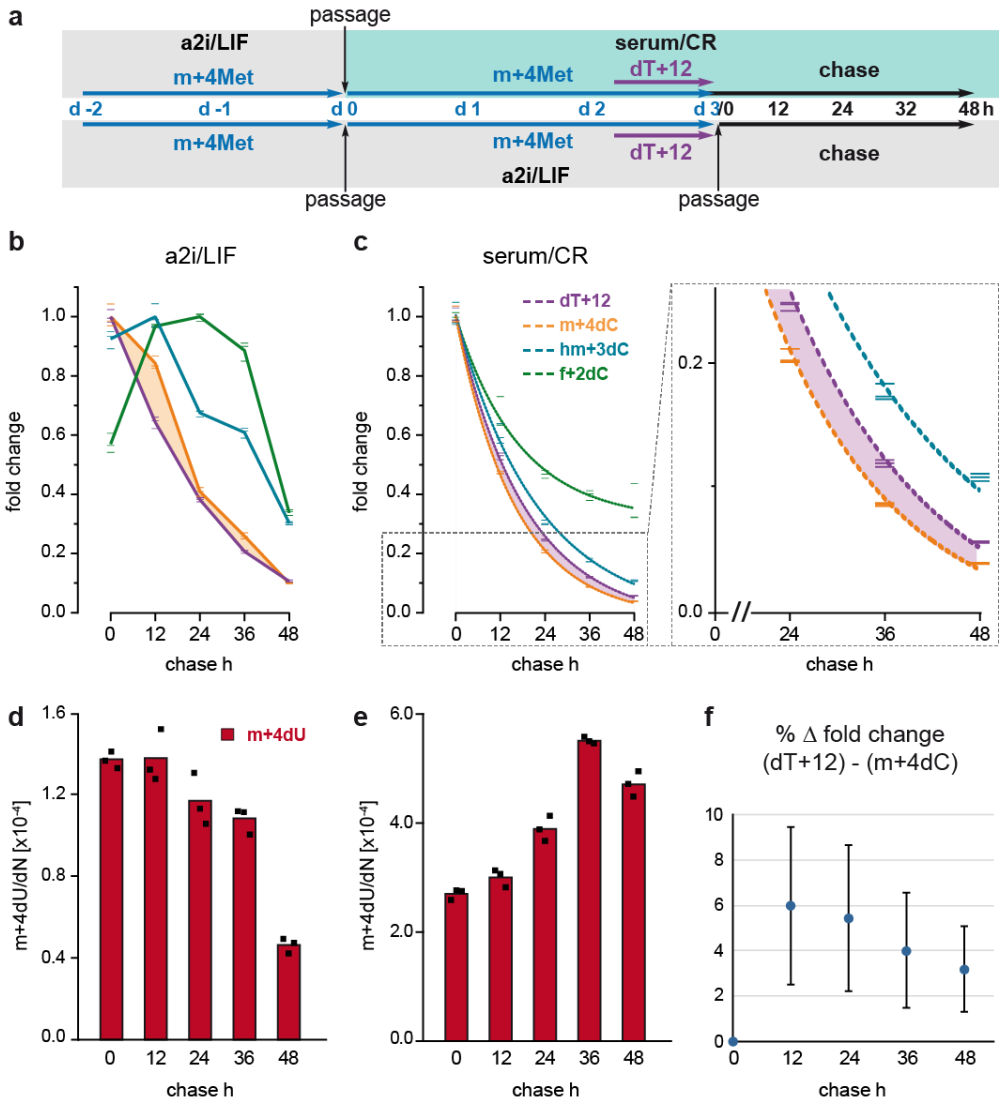


Figure 2. Turnover of genomic mdC and its derivatives under a2i/LIF and priming conditions.

a) Workflow of m+4Met and dT+12 pulse-chase experiments with wild type mPSCs under a2i/LIF and priming (serum/CR) conditions. **b** and **c**) Global profiles of the indicated labelled cytosine derivatives and dT+12 in the genome of mPSCs under a2i/LIF (b) and priming conditions (serum/CR; c) upon chasing of m+4Met. Each panel shows fold change values relative to the highest level for each modification from an individual experiment. Dashes represent single technical replicate measurements (n=3) from single samples. Full lines in b join mean values, while in c dashed lines represent first order decay curves fitted to mean values. In b and c the area between the m+4dC and dT+12 lines is shadowed in yellow and purple, respectively. **d** and **e**) Absolute levels of genomic m+4Met from the same samples as in b and c, respectively. Black squares represent single technical replicate measurements (n=3) from single samples and bars show their mean values. The experiment under priming conditions (c,e) was independently repeated once in the same mPSC line (Extended Data Figure 4) and once in E14tg2a (Fig. 3). **f**) Percent of difference between fold changes of dT+12 and m+4dC in wt mPSCs under serum/CR conditions from independent biological experiments (n=3; Fig. 2c,3a and Extended Data Fig. 4a). Large blue dots and error bars represent mean values and SD, respectively.

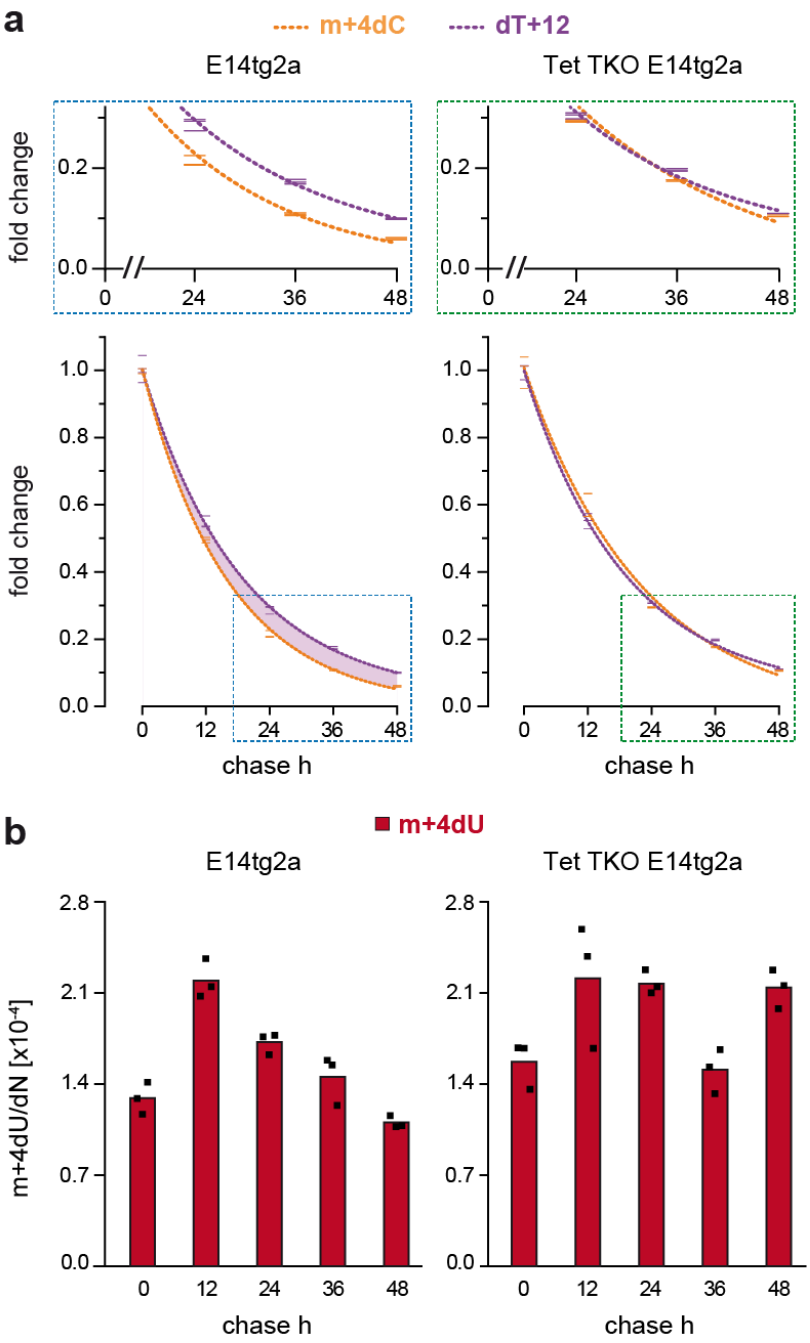


Figure 3. The active component of mdC turnover in primed pluripotent cells largely depends on Tet proteins.

m+4Met and dT+12 pulse-chase experiment with parental and Tet TKO E14tg2a mPSCs under priming conditions (serum/CR; equivalent to Fig. 2c,e). Profiles of global genomic levels of the indicated labelled cytosine derivatives and dT+12 are shown. **a**) m+4dC and dT+12 fold change values relative to the highest data point for each modification. Upper panels display magnifications of dashed frames in lower panels. Dashes represent single technical replicate measurements (n=3) from single samples and dashed lines are first order decay curves fitted to their means. In middle panels the area between the m+4dC and dT+12 lines is shadowed in purple. **b**) Absolute levels of genomic m+4Met from the same samples as in a. Black squares represent individual technical replicate measurements (n=3) from single samples and bars show their mean values. This experiment was not independently replicated.

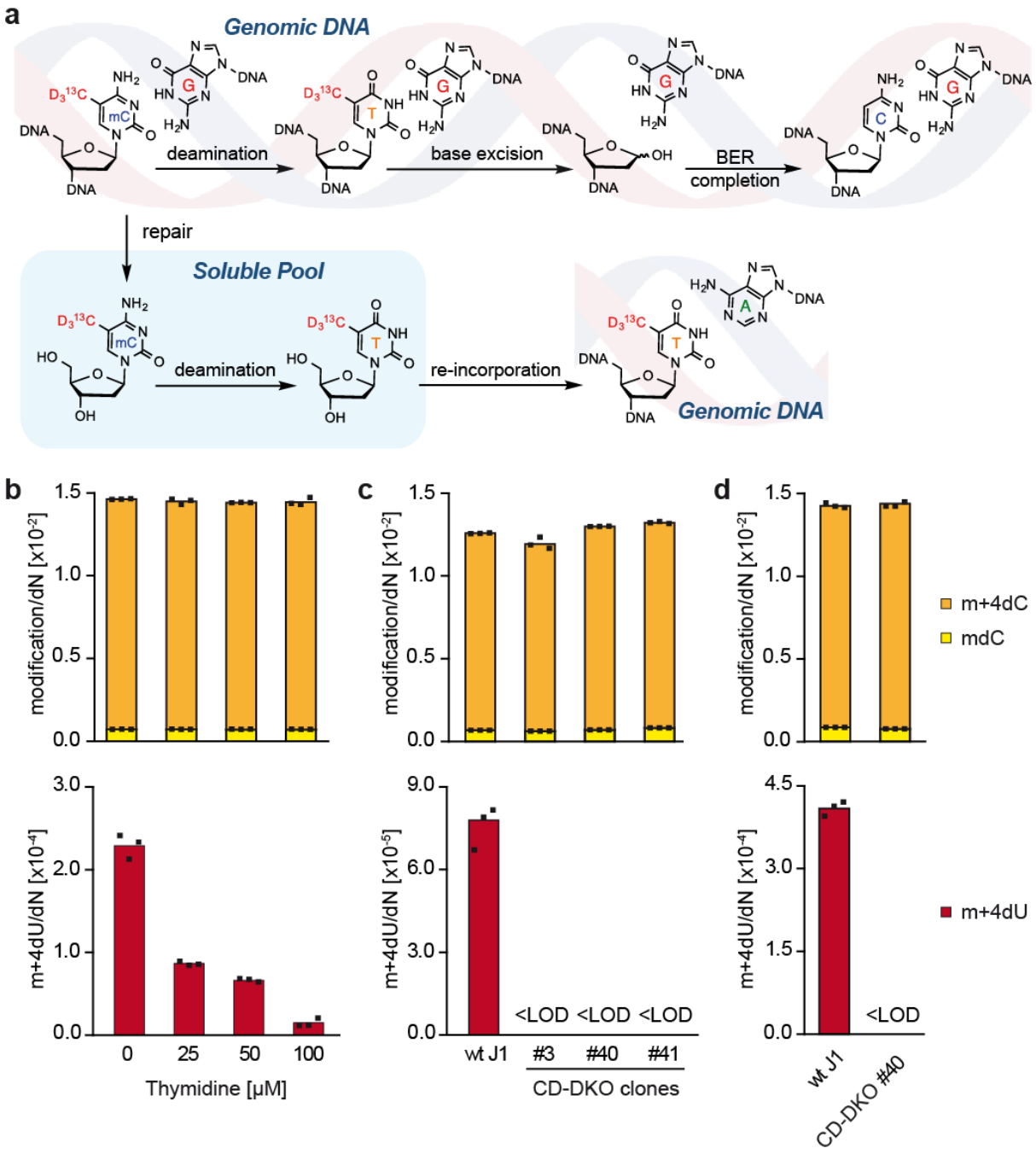


Figure 4. All detectable mdC deamination takes place in the soluble pool.

a) Potential pathways for the conversion of genomic m+4dC into genomic m+4dU. **b**) wt mPSCs were primed under CHIR/IWR1 (serum/CR) conditions in the presence of m+4Met and growing concentrations of unlabelled dT in the culture medium as indicated. **c**) wt and CD-DKO clones mPSCs were grown under serum/LIF conditions in the presence of m+4Met. **d**) wt and CD-DKO clone #40 mPSCs were metabolically labelled with m+4Met under serum/CR priming conditions. In b-d priming and labelling was for five days. Global genomic contents of mdC, m+4dC (upper panels) and m+4dU (lower panels) are shown as individual values of replicate measurements (n=3: black squares) from single samples and their means (bars). LOD = limit of detection. Experiments in b and c were repeated three times with very similar results.

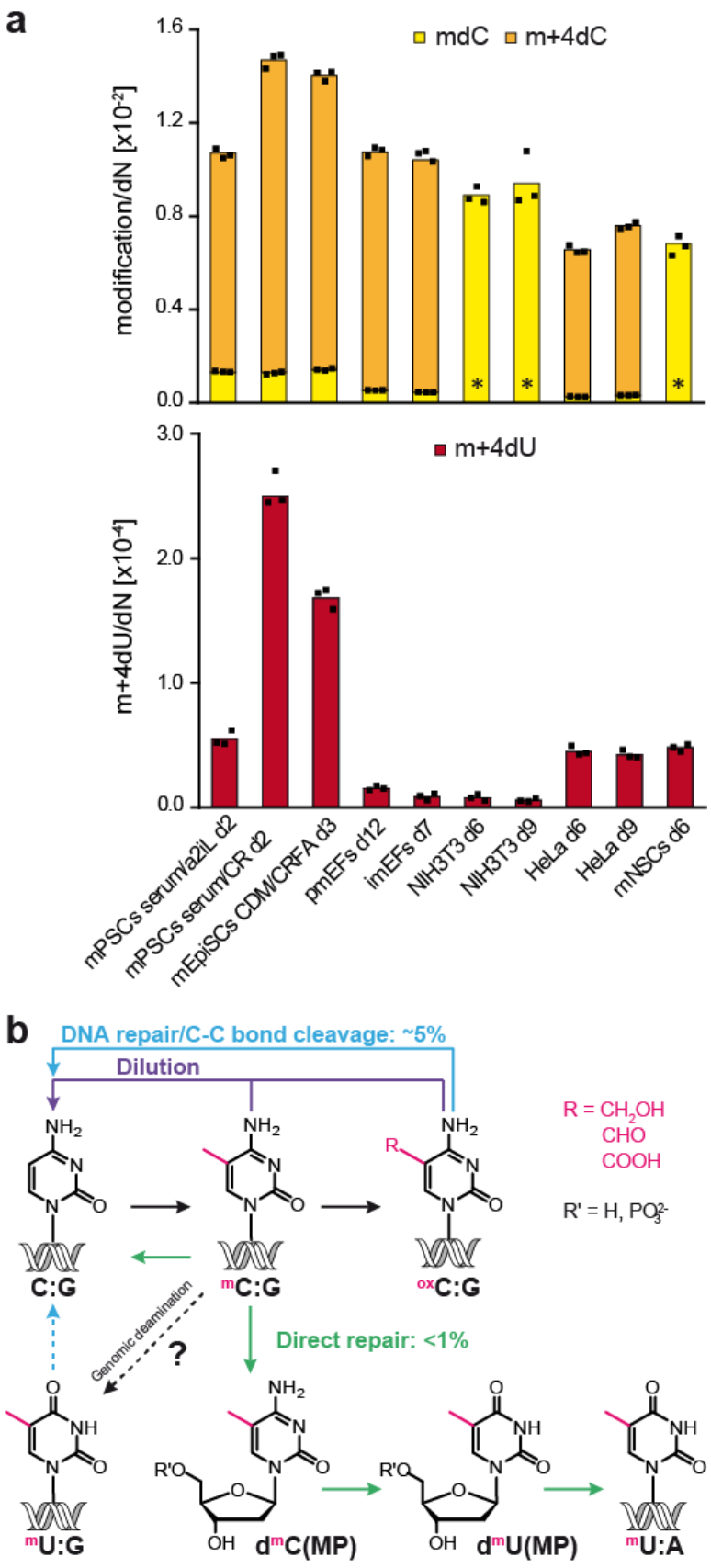


Figure 5. Oxidation-independent mdC-to-dT turnover is developmentally regulated.

a) Cultures of mPSCs in a2i/LIF and primed states, primary (pmEFs) and immortalized mouse embryonic fibroblasts (imEFs and NIH3T3), HeLa and mPSC-derived neural stem cells (mNSCs) were labelled for the indicated number of days (d) with m+4Met. Asterisks indicate cases where distinct parallel cultures in the absence and presence of m+4Met were used to measure global levels of genomic mdC and m+4dU, respectively. In all other cases mdC, m+4dC and m+4dU were measured from the same culture labelled with m+4Met. Individual values of replicate measurements (n=3: black squares) from single samples and their means (bars) are shown. Only the first two samples were independently replicated a second time. a2iL = a2i/LIF; CR = CHIR/IWR1; FA = FGF-2/Activin A; CDM = Chemically Defined Medium. **b**) Schematic summary of the dC modification turnover pathways identified in primed pluripotent cells. Deamination of genomic mdC could not be detected and may occur only at levels below the sensitivity of our assay.

Online Methods

Cell culture

Medium components were from Sigma unless specified otherwise. Basal medium for mPSC culture was DMEM high glucose containing 10% FBS (Pan-ES, Pan-Biotech), 2 mM L-Alanyl-L-Glutamine, 1x MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol. mPSC lines were routinely maintained in basal medium supplemented with 1000 U/mL LIF (ORF Genetics), 3 µM CHIR99021 and 7.5 µM CGP77675 (a2i). For adaptation to 2i/LIF conditions CGP was replaced with 1 µM PD0325901. For priming, mPSC basal medium was supplemented with 1.5 µM CHIR99021 and IWR1-endo at 2.5 µM as previously reported[24,51]. For experiments under serum/LIF conditions basal medium was supplemented exclusively with 1000 U/ml LIF. Small molecule inhibitors were purchased from Selleckchem (CHIR, PD, IWR1), Axon Medchem (CHIR, CGP, PD, IWR1) or MedChemExpress (CHIR). Gelatin coating was used for all mPSCs cultures in serum containing medum. Wherever not otherwise stated experiments with wild type mPSCs were performed with an embryonic stem cell line derived from F1 129/C57Bl6 hybrid embryos[52]. For metabolic labelling experiments with m+4Met under all conditions mPSC basal medium was generated with L-Methionine- and L-Cysteine-free DMEM (Sigma or Life Technologies) supplemented with 0.2 mM m+4Met ([*methyl*-$^{13}C,d_3$]) and L-Cysteine (both from Sigma). In pulse-chase experiments, dT+12 ([$^{15}N_2$,$^{13}C_{10}$]-dT; BACH UG, Hamburg) was supplemented at 5 µM. dC[D_9] (Silantes GmbH, Munich) was supplemented to culture medium at 100 µM. EpiSCs in chemically defined serum-free medium (CDM) were derived from the Oct4-YFP reporter mPSC line OLY2-1[23]. These were first adapted to CDM supplemented with CHIR, PD and LIF as reported above, then transferred to CDM supplemented with CHIR, CGP, 12 ng/ml FGF-2 (Miltenyi Biotec) and 20 ng/ml Activin A (PeproTec) and maintained in the latter medium by passaging as small cell clusters using 5 mM EDTA in Hank's balanced salt solution without Mg^{2+}, Ca^{2+} and sodium carbonate supplemented with 10 mM HEPES buffer. CDM was DMEM high glucose supplemented with 1x NEAA, 2 mM L-Alanyl-L-Glutamine, 15 µg/ml human recombinant Insulin, 10 µg/ml human holo-Transferrin (Merck), 12.5 mg/ml AlbuMAX I (Life Technologies) 0.7 µg/ml vitamin B12, 1.8 ng/ml biotin, 0.75 µM $ZnSO_4$ and 2.6 nM $CuSO_4$. Laminin 521 coating was used for all mPSC cultures in CDM. The neural stem cell line ENC1 was generated and cultured as previously described[45]. ENC1 cells were labelled with m+4Met ([*methyl*-$^{13}C,d_3$]) in DMEM/F-12 generated with L-Methionine-free DMEM supplemented with N2[53] and 20 ng/ml each of FGF-2 and EGF (PerpoTech). Primary mEFs (CF-1, Applied StemCell) were cultured in mPSC basal medium. Fibroblast cell lines and HeLa cells were cultured in DMEM containing 10% FBS (Life Technologies), 2 mM L-glutamine and 0.1 mM β-mercaptoethanol.

Generation of CD-DKO mPSCs by CRISPR/Cas9 editing

Single guide RNAs were designed over exon1 and 4 of *Cda* and *Dctd*, respectively (Supplementary Fig. 1), using CRISPR Design Tool (http://crispr.mit.edu/). Oligonucleotides used for construction of Cas9/gRNA expression vectors are listed in Supplementary Table 1. The oligonucleotides were annealed and cloned in the BbsI site of pSpCas9-2A-Puro (PX 459; Addgene Plasmid 48139)[54]. Cas9/gRNA expression vectors targeting both *Cda* and *Dctd* were co-transfected in wt J1 mPSCs using Lipofectamine 2000 (Life Technology) according to the manufacturer's instructions. Two days after transfection cultures were selected with 1 µg/ml puromycin for two days and subcloned by limiting dilution. Clones were screened using the Surveyor Mutation Detection Kit (Transgenomics) and potential compound biallelically targeted clones were subject to sequencing. For reverse transcription-PCR, RNA was isolated with the ZR-Duet DNA/RNA MiniPrep Kit (Zymo Research) and first strand cDNA synthesis was as previously described[55]. Primers were used to amplify the Cda cDNA are reported in Supplementary Table 1. For western blot analysis of Dctd whole cell extracts and purified human recombinant DCTD protein[56] were probed with a rabbit anti-DCTD antibody[36].

Isolation of genomic DNA and UHPLC-MS^2 analysis

Isolation, digestion and UHPLC-MS^2 analysis of gDNA samples were performed as described previously described[25,57] with the following modifications: for every technical replicate 4 µg of gDNA were digested and analysed; the enzyme mixture including S1 Nuclease, Antarctic Phosphatase and snake venom Phosphodiesterase was used for digestion; the digestion mixture was supplemented with 0.8 µM Tetrahydrouridine (Abcam) to inhibit potential cytidine deaminase activity in enzyme preparations.

Reverse transcription (RT)-qPCR analysis

Total RNA was isolated from the lysate flow-throughs obtained upon genomic DNA isolation with silica spin columns. RNA isolation from the flow-throughs was performed as described in the Quick-DNA/RNA Miniprep kit (Zymo Research, D7001). In-column DNase I digestion was performed to minimize carryover of residual genomic DNA. One microgram of total RNA was reverse transcribed with the iScript cDNA Synthesis Kit and the cDNA product equivalent to 50 ng of total RNA were amplified using iTaq Univesal SYBR Green Supermix (both from BioRad) on a qTOWER^3/G cycler (Jena Biosciences). qPCR primers are reported in Supplementary Table 1. Each primer pair spanned across one intron except fom Oct6 cDNA amplifications as Pou3f1 (the gene coding Oct6) is intronless. qPCR reactions were also performed in parallel for every target/sample by omitting the reverse transcriptase to control for signals generated from residual genomic DNA (intronless Oct6 transcript, relatively short introns and spliced pseudogenes). An RT-qPCR assay for Gapdh transcripts was used as

housekeeping transcript reference to calculate ΔCt values. Fold change values were calculated with the ΔΔCt method.

High depth targeted bisulfite amplicon sequencing and analysis

Genomic DNA (1-2 µg) was bisulfite treated using the EpiTect Bisulfite Kit (Qiagen 59104). 5' overhangs compatible with Illumina TruSeq and Nextera adapters were appended to the locus specific primers (Supplementary Table 1). Approximately 10 ng of bisulfite converted DNA were amplified with HotStar-Taq plus DNA polymerase (Qiagen). PCR products were analyzed by agarose gel electrophoresis and quantified by densitometry using ImageJ (version 1,52d). Similar amounts of these primary amplicons were used as templates for a second round of amplification with i5 and i7 Indexing Primers (Supplementary Table 1). PCR products were again analyzed by agarose gel electrophoresis and, after purification with CleanNGS magnetic beads (CleanNA), they were quantified by fluorometry with Quant-iT PicoGreen Reagent (Thermo Fisher). Equimolar amounts of amplicons were then pooled and the size distribution of the final library was assessed on a Bioanalyzer (Agilent). The 2x300 output mode for dual indexed sequencing was used on a MiSeq instrument (Illumina). Size, numbers of CpG sites and genomic coordinates of targeted regions are reported in Supplementary Table 2. Sequenced reads were trimmed for adaptor sequences and masked for low-quality sequence using TrimGalore version 0.5.0, Cutadapt version 1.17. Trimmed reads were mapped to imprint-containing, bisulfite converted chromosomes 2, 7, 11,12 and 17 from mm10 using bsmap v2.90. Methylation ratios were extracted from bam files using a custom Python script and tabulated as a matrix table from which the box plot in Extended Data Fig. 2b was generated by means of an R script. Raw sequencing data and matrix table are available at GEO with accession number GSE152174.

Synthetic Procedures

The synthesis of labelled nucleosides standards is reported in the Supplementary Information file as Supplementary Note.

Data availability statement

The targeted bisulfite amplicon sequencing data on which Extended Data Fig. 2b is based are available at GEO with accession code GSE152174. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files.

Methods-only references

51. Zhou, X., Chadarevian, J. P., Ruiz, B. & Ying, Q.-L. Cytoplasmic and Nuclear TAZ Exert Distinct Functions in Regulating Primed Pluripotency. *Stem Cell Reports* **9**, 732–741 (2017).
52. Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M. & Fässler, R. Kindlin-3 is essential for integrin activation and platelet aggregation. *Nature Medicine* **14**, 325–330 (2008).
53. Nichols, J. & Ying, Q.-L. Derivation and propagation of embryonic stem cells in serum- and feeder-free culture. in *Embryonic Stem Cell Protocols* (ed. Turksen, K.) vol. 329 91–98 (Humana Press, 2006).
54. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
55. Rahimoff, R. *et al.* 5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017).
56. Weiner, K. X., Weiner, R. S., Maley, F. & Maley, G. F. Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in Escherichia coli. *J. Biol. Chem.* **268**, 12983–12989 (1993).
57. Traube, F. R. *et al.* Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nature Protocols* **14**, 283 (2019).

SUPPLEMENTARY INFORMATION

Active turnover of genomic methylcytosine in pluripotent cells

Fabio Spada[1]*, Sarah Schiffers[1], Angie Kirchner[1], Yingqian Zhang[1,2], Olesea Kosmatchev[1], Eva Korytiakova[1], René Rahimoff[1§], Charlotte Ebert, and Thomas Carell[1]*

[1]Department of Chemistry, Ludwig Maximilians University Munich and Center for Integrated Protein Science Munich (CIPS[M])

[2]College of Chemisty, State Key Laboratory of Elemento-Organic Chemistry and Department of Chemical Biology, Nankai University, Tianjin, China

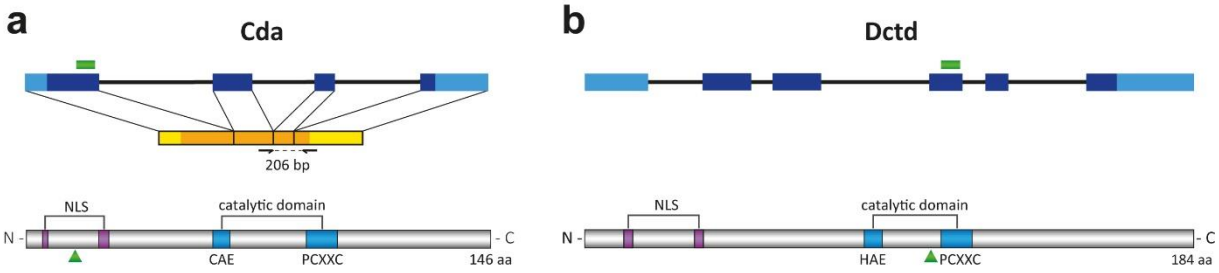
[§]Current addresses: R.R., Department of Chemistry, University of California, Berkeley, USA

*Corresponding authors: Fabio Spada fabio.spada@cup.lmu.de; Thomas Carell: Thomas.Carell@cup.lmu.de

Supplementary Figure 1 and 2

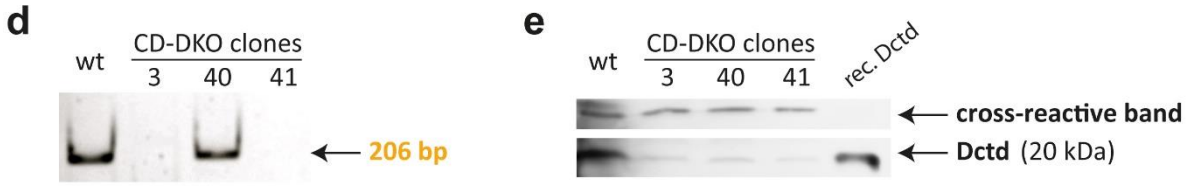
Supplementary Table 1 and 2

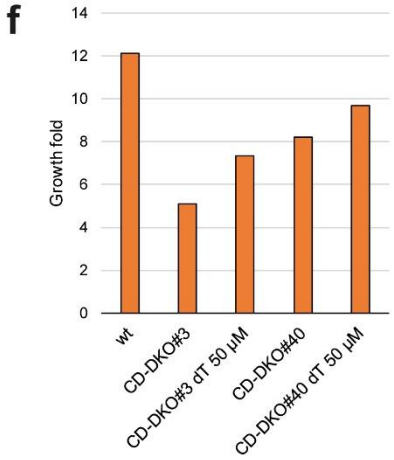
Supplementary Note: Synthetic Procedures



c

Genomic sequence:		Protein sequence:
Wild type Cda	...GGACAGCAGCAGCCGCTGGACGTGCTC**GGG**CTCCA...	MAQERPSCAVEPEHVQRLLLSSREAKKSAYC...
CD-DKO clone 03_Cda	...GGACAGCAGCAGCCGCCCGGTTC-CTCGGGCTCCA...	MAQERPSCAVEPEEPGGCCCPLVRPRSQPT...
CD-DKO clone 40_Cda	...GGACAGCAGCAGCCGCTGGACG--CTCGGGCTCCA...	MAQERPSCAVEPERPAAAAVLS*GQEVSLL...
CD-DKO clone 41_Cda	...GGACAGCAGCAGCCGCT----------GGGCTCCA...	MAQERPSCAVEPSGCCCPLVRPRSQPTAPT...
Wild type_Dctd	...AGCACATTCGTTACACGGGAA-CA**AGG**CGACGTAC...	...MNKNSADVKGCSMYVALFPCNECAKLIIQ...
CD-DKO clone 03_Dctd	...AGCACATTCGTTACACGGG--------CGACGTAC...	...MNKNSADVKGCSMYVAFPV*RMC*THH...
CD-DKO clone 03_Dctd	...AGCACATTCGTTACACGGGAA---AGGCGACGTAC...	...MNKNSADVKGCSMYVARVTNVLNSSSRQV*...
CD-DKO clone 40_Dctd	...AGCACA----------------AAGGCGACGTAC...	...MNKNSADVKGCSMYVAFVLNSSSRQV*K...
CD-DKO clone 40_Dctd	...AGCACATTCGTTACACGGGAA---AGGCGACGTAC...	...MNKNSADVKGCSMYVAFPV*RMC*THH...
CD-DKO clone 41_Dctd	...AGCACATTCGTTACACGGGAAACAAGGCGACGTAC...	...MNKNSADVKGCSMYVAFPV*RMC*THH...
CD-DKO clone 41_Dctd	...AGCACATTCGTTACACGGGAA---AGGCGACGTAC...	...MNKNSADVKGCSMYVALFPV*RMC*TH...

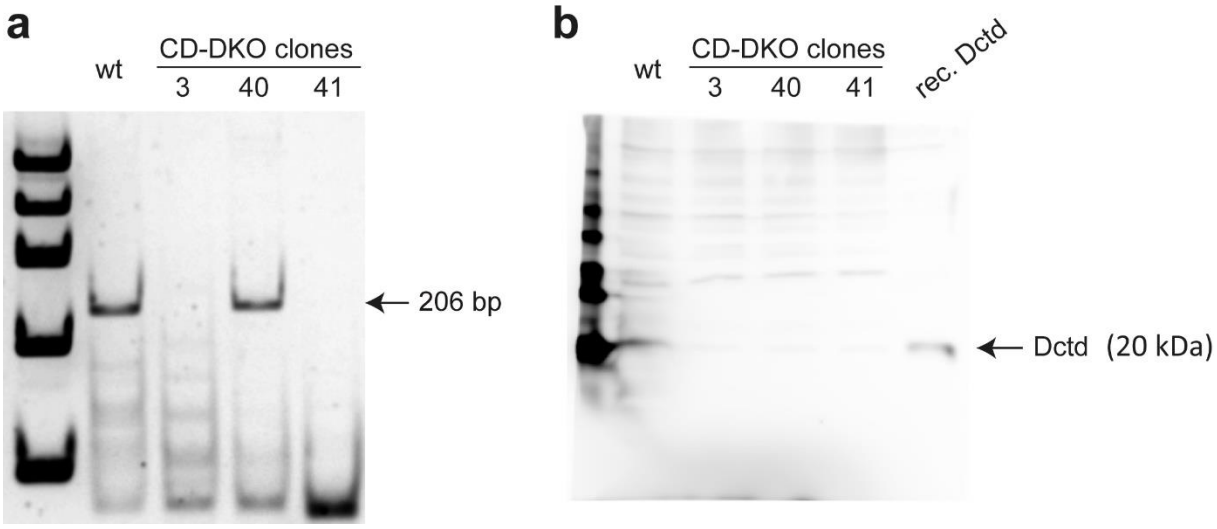




Supplementary Figure 1. Generation and characterization of CD-DKO mPSCs (related to Fig. 4).

a and **b**) Schematic representations of gene body (top), transcript (middle, A only) and protein (bottom) for mouse Cda (a) and Dctd (b). Exons are represented by blue boxes, with UTR

regions in lighter colour. The positions of gRNAs are indicated by green rectangles over the gene, while green triangles indicate the approximate positions of the expected truncation of the protein. The catalytic domains are shown along with the two Zn^{2+} co-ordinating motifs, which are conserved across kingdoms of life among all cytosine nucleoside and monophosphate nucleotide deaminases, including the vertebrate-specific Aid/Apobec family. **c**) Genomic sequences of *Cda* and *Dctd* alleles as well as respective protein sequences for Cda and Dctd in CD-DKO clones. Antisense strands are shown so as to highlight the PAM sequences (bold type in wt sequence). The expected Cas9 cut sites are indicated by green triangles. The three clones are homozygous and heterozygous with respect to Cda and Dctd mutations, respectively. Mutations in *Dctd* were targeted to exon 4 and lead to termination of the frame before the second conserved catalytic motif. Note that, though exons 3 and 5 are in frame, alternative splicing events skipping the frameshifts in exon 4 would generate a protein lacking both conserved motifs involved in Zn^{2+} co-ordination, which would therefore lack catalytic activity. **d**) Reverse transcription-PCR products from CDA transcripts in wt cells and CD-DKO clones resolved on a 10% non-denaturing polyacrylamide gel. The experiment was repeated twice with the same result. **e**) Western blot with an anti-Dctd antibody on whole cell extracts from wt and CD-DKO cells and purified recombinant Dctd protein. Cross-reactive bands are shown in the upper panel as loading control. As all the alleles detected in CD-DKO clones are expected to result in substantial protein truncations, the week signals visible in CD-DKO lanes at approximately the same position as in wt and recombinant Dctd lanes are likely due to unspecific cross-reaction. **f**) Growth delay of CD-DKO mPSCs and partial rescue with dT. The same number of cells was plated for two days under a2i/LIF conditions with or without addition of 50 µM dT to the medium (only CD-DKO clones). Cell counts after two days of culture are shown.



Supplementary Figure 2. Source data for Supplementary Fig. 1d and e
a and **b**) Uncropped and unprocessed source images for Supplementary Fig. 1d and e, respectively. Unlabelled lanes on the left show molecular weight standards.

Supplementary Table 1

Oligonucleotide/primer sequences used in this study. RT = Reverse Transcription.

Name	Sequence	Application	Reference
mCda gRNA F	CACCGAGCAGCCGCTGGACGTGCTC	gRNA construct	this work
mCda gRNA R	AAACGAGCACGTCCAGCGGCTGCTC	gRNA construct	this work
mDctd gRNA F	CACCGACATTCGTTACACGGGAACA	gRNA construct	this work
mDctd gRNA R	AAACTGTTCCCGTGTAACGAATGTC	gRNA construct	this work
mCda F	AAGGCCATCTCCGAAGGGTA	RT-PCR	this work
mCda R	TCTTCAGGTCCAAACGAGGC	RT-PCR	this work
Dnmt3b Fw	CTGGCACCCTCTTCTTCATT	RT-qPCR	1
Dnmt3b Rv	ATCCATAGTGCCTTGGGACC	RT-qPCR	1
Fgf5 Fw	AGGGGATTGTAGGAATACGAG	RT-qPCR	3
Fgf5 Rv	TCTTGGAATCTCTCCCTGAAC	RT-qPCR	3
Klf4 Fw	CGGGAAGGGAGAAGACACT	RT-qPCR	3
Klf4 Rv	GAGTTCCTCACGCCAACG	RT-qPCR	3
Nanog Fw	AAGTACCTCAGCCTCCAGCA	RT-qPCR	4
Nanog Rv	GCTTGCACTTCATCCTTTGG	RT-qPCR	4
Otx2 Fw	CATGATGTCTTATCTAAAGCAACCG	RT-qPCR	5
Otx2 Rv	GTCGAGCTGTGCCCTAGTA	RT-qPCR	5
Oct6 Fw	TTTCTCAAGTGTCCCAAGCC	RT-qPCR	3
Oct6 Rv	ACCACCTCCTTCTCCAGTTG	RT-qPCR	3
Oct4 Fw	AGGAAGCCGACAACAATGAG	RT-qPCR	4
Oct4 Rv	ACTCCACCTCACACGGTTCT	RT-qPCR	4
Prdm14 Fw	CTCTGGAGACAGGCCATACC	RT-qPCR	3
Prdm14 Rv	GTCTGATGTGTGTTCGGAGTATGC	RT-qPCR	3
Tfcp2l1 Fw	GGGGACTACTCGGAGCATCT	RT-qPCR	3
Tfcp2l1 Rv	TTCCGATCAGCTCCCTTG	RT-qPCR	3
Rex1 Fw	GCTCCTGCACACAGAAGAAA	RT-qPCR	5
Rex1 Rv	GTCTTAGCTGCTTCCTTCTTGA	RT-qPCR	5
Stella Fw	AAAGTCGACCCAATGAAGGA	RT-qPCR	4
Stella Rv	ACACCGGGGTTTAGGGTTAG	RT-qPCR	4
Gapdh Fw	TGGAGAAACCTGCCAAGTATGA	RT-qPCR	1
Gapdh Rv	GTTGCTGTTGAAGTCGCAGG	RT-qPCR	1
Dlk1 S1 F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GTGTTTATTAGTTTGGTGGTG	Bsulfite PCR	this work
Dlk1 S1 R	CGACGCTCTTCCGATCTACTTAAATCTCCTCATCACCA	Bsulfite PCR	this work
H19 S2 F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAGAGAGAAGAA GGAGATTTATTTGT	Bsulfite PCR	this work
H19 S2 R	CGACGCTCTTCCGATCTCATCTAAAATCTAACAAAAATATTAAAAA	Bsulfite PCR	this work
H19 S3 F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTATAATGGGA ATTTGAGGGTA	Bsulfite PCR	this work
H19 S3 R	CGACGCTCTTCCGATCTTAACCAAAACCAAACTATAAAATAACTAAT	Bsulfite PCR	this work
Nesp S1a F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGTTTTTTGAA TTTGAGTTTGAT	Bsulfite PCR	this work
Nesp S1a R	CGACGCTCTTCCGATCTTAAATAACTAATTAAAAATAACACCC	Bsulfite PCR	this work
Nesp S2 F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGAGGTTAAAGG TTTGTTGG	Bsulfite PCR	this work
Nesp S2 R	CGACGCTCTTCCGATCTACATAAAAAACAAATAACTTACCTTT	Bsulfite PCR	this work
Magel2-Mrkn3 S1b F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGGGGATGGA GAAGGGT	Bsulfite PCR	this work
Magel2-Mrkn3 S1b R	CGACGCTCTTCCGATCTATTCTCCAACTTCTACTCTCA	Bsulfite PCR	this work
Cdkn1c 2 F	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTGTGAAATTG AAAATATTATATTATGTT	Bsulfite PCR	this work
Cdkn1c 2 R	CGACGCTCTTCCGATCTTAAATAAAACCCCTTACACAACC	Bsulfite PCR	this work
Nextera i7	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTC GG	Indexing	this work
TruSeq i5	AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTT CCCTACACGACGCTCTTCCGATCT	Indexing	this work

Supplementary Table 2 (related to Extended Data Figure 2b)

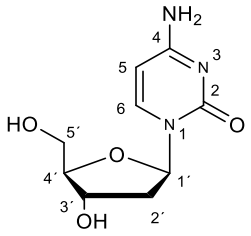
Size, numbers of CpG sites and genomic coordinates of regions targeted within secondary imprints (sDMRs).

sDMR	genomic coordinates (mm10)	Size (bp)	CpGs
Dlk1	chr12:109,460,142-109,460,363	221	9
H19 S2	chr7:142,577,743-142,578,043	300	9
H19 S3	chr7:142,578,512-142,578,757	245	7
Nesp S1a	chr2:174,284,913-174,285,161	248	23
Nesp S2	chr2:174,285,645-174,285,827	182	10
Magel2-Mrkn3	chr7:62,406,768-62,407,043	275	15
Cdkn1c	chr7:143,460,795-143,461,100	305	28

SYNTHETIC PROCEDURES

Unless noted otherwise, all reactions were performed using oven dried glassware under an atmosphere of argon. Molsieve-dried solvents were used from *Sigma Aldrich* and chemicals were bought from *Sigma Aldrich*, *TCI*, *Carbolution* and *Carbosynth*. For extraction and chromatography purposes, technical grade solvents were distilled prior to their usage. Reaction controls were performed using TLC-Plates from *Merck* (Merck 60 F_{254}), flash column chromatography purifications were performed on *Merck* Geduran Si 60 (40-63 μM). Visualization of the TLC plates was achieved through UV-absorption. NMR spectra were recorded in deuterated solvents on *Varian VXR400S*, *Varian Inova 400*, *Bruker AMX 600*, *Bruker Ascend 400* and *Bruker Avance III HD*. HR-ESI-MS spectra were obtained from a *Thermo Finnigan* LTQ FT-ICR. IR-measurements were performed on a *Perkin Elmer Spectrum BX*. HPLC purifications were performed on a *Waters Breeze* system (2487 dual array detector, 1525 binary HPLC pump) using a Nucleosil VP 250/10 C18 column from *Macherey Nagel*. HPLC-grade MeCN was purchased from *VWR*.

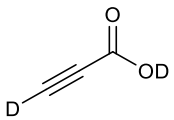
General numbering of the pyrimidine nucleosides



Synthesis of 5-[D_3]-Methyl-2'-deoxy-[6-D, 1,3-$^{15}N_2$]-cytidine

[D_2]-Propionic acid

5

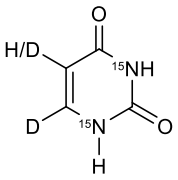


Na_2CO_3 (6.57 g, 62.0 mmol, 1.5 Equiv.) was dissolved in D_2O (25 mL), propionic acid (2.60 mL, 42.0 mmol, 1.0 Equiv.) was added dropwise over 3 min. The reaction was stirred for 18 h at rt, the solvent was removed under reduced pressure, another portion of D_2O (21 mL) was added and the mixture was stirred for 14 h at rt. Through addition of D_2SO_4 the pH value of the mixture was adjusted to 1. The mixture was extracted with DCM, the organic layers were dried over $MgSO_4$, all the volatiles were removed under reduced pressure (600 mbar, 40 °C). Purification

by fractional distillation (27 mbar,39 °C) yielded [D_2]-propionic acid (452 mg, 6.27 mmol, 15 %) as a colorless liquid and was directly used for further reactions.

[D_2,$^{15}N_2$]-Uracil

6



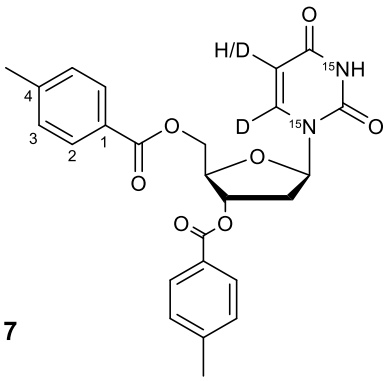
[$^{15}N_2$]-Urea (150 mg, 2.41 mmol, 1.0 Equiv.) was dissolved in D_2SO_4 (3 mL) and [D_2]-propiolic acid (0.161 mL, 2.65 mmol, 1.1 Equiv.) was added dropwise. The mixture was stirred at 105 °C for 18 h. The brown suspension was diluted with water (50 mL) and the pH value was adjusted to 10 with Na_2CO_3. Volatiles were removed under reduced pressure and the crude was purified by flash column chromatography (DCM/MeOH 4:1). A mixture of [D_2, $^{15}N_2$]-uracil (101 mg, 0.864 mmol, 36 %) und [D, $^{15}N_2$]-uracil (50 mg, 0.43 mmol, 18 %) was obtained as a colorless solid.

^{1}H-NMR (400 MHz, DMSO-d6) δ 11.03 (dd, ^1J = 77.2 Hz, ^4J = 1.8 Hz, 1H, N3-H), 10.80 (dd,^1J = 82.6 Hz, ^4J = 1.9 Hz, 1H, N1-H), 5,46 (dd, ^3J = 4.3 Hz, ^4J = 2.5 Hz, C5-H) ppm.

^{13}C-NMR (101 MHz, DMSO-d6) δ 164.9 (t, ^{1}J = 9.7 Hz, C4), 151.9 (m, C2), 142.3 (m, C6), 101.4-99.5 (m, C5) ppm.

[D_2,$^{15}N_2$]-3',5'-Bis-O-(p-Toluoyl)-2'-deoxyuridine

7



The mixture of [D_2, $^{15}N_2$]-uracil und [D, $^{15}N_2$]-uracil (37 mg, 0.32 mmol, 1.0 Equiv.) was dissolved in MeCN (2 mL). Bis-(trimethylsilyl)acetamide (0.3 mL, 0.96 mmol, 3.0 Equiv.) was added dropwise and the reaction mixture was stirred for 3 h at 80 °C. All volatiles were removed under reduced pressure and the residue was coevaporated with chloroform (3 x 6 mL). The residue was dissolved in chloroform(2 mL) and Hoffer's chlorosugar (149 mg,0.383 mmol, 1.2 Equiv.) was added and stirred at rt for 17 h. After removing of the volatiles under reduced

pressure, purification by flash column chromatography (*i*Hex/EtOAc 1:1) yielded a mixture of α-[D_2, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-2'-deoxyuridine β-[D, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-2'-deoxyuridine (41 mg, 0.088 mmol, 28 %, α/β: 1.28:1) as a yellow foam.

α-[D_2, $^{15}N_2$]-3',5'-Bis-O-(p-toluoyl)-2'-deoxyuridine:

^{1}H-NMR (400 MHz, CDCl3): δ = 9.06 (d, *^{1}J* = 91.0 Hz, 1H, NH), 7.89 (d, *^{3}J* = 8.0 Hz, 2H, Tol2), 7.78 (d, *^{3}J* = 8.0 Hz, 2H, Tol2), 7.31-7.20 (m, 4H, Tol3), 6.31 (d, *^{3}J* = 6.8 Hz, 1H, C1'- H), 5.72 (dt, *^{3}J* = 4.7 Hz, *^{3}J* = 2.4 Hz, 1H, C5-H), 5.61 (m, 1H, C3'-H), 4.86 (t, *^{3}J* = 4.3 Hz, 1H, C4'-H), 4.54 (m, 2H, C5'-H), 3.01-2.89 (m, 1H, C2'-H), 2.55 (d, *^{3}J* = 15.5 Hz, 1H, C2'-H), 2.43 (s, 3H, CH_3), 2.41 (s, 3H, CH_3) ppm.

^{13}C-NMR (101 MHz, CDCl3): δ = 165.7 (COO), 166.1 (COO), 150.1 (C2), 144.9 (Tol4), 144.4 (Tol4), 129.7 (Tol2), 129.6 (Tol2), 129.5 (Tol3), 129.4 (Tol3), 126.5 (Tol1), 125.9 (Tol1), 101.4 (C5), 88.0 (d, 1J = 10.0 Hz, C1'), 85.7 (C4'), 74.7 (C3'), 64.0 (C5'), 39.0 (C2'), 21.7 (CH_3) ppm.

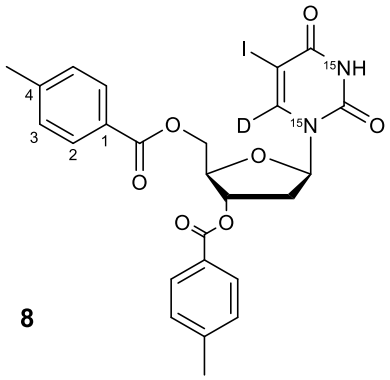
β-[D_2, $^{15}N_2$]-3',5'-Bis-O-(p-toluoyl)-2'-deoxyuridine:

^{1}H-NMR (400 MHz, CDCl3): δ = 8.93 (d, *^{1}J* = 91.0 Hz, 1H, NH), 7.94 (d, *^{3}J* = 8.0 Hz, 2H, Tol2), 7.93 (d, *^{3}J* = 8.0 Hz, 2H, Tol2) 7.31-7.20 (m, 4H, Tol3), 6.41 (dd, *^{3}J* = 8.3 Hz, *^{3}J* = 5.7 Hz, 1H, C1'-H), 5.60 (m, 2H, C3'-H und C5-H), 4.73 (dd, *^{3}J* = 12.3 Hz, *^{4}J* = 3.1 Hz, 1H, C5'-H), 4.67 (dd, *^{3}J* = 12.3 Hz, *^{4}J* = 3.5 Hz, 1H, C5'-H), 4.60-4.49 (m, 1H, C4'-H), 2.75 (dd, *^{3}J* = 14.3 Hz, *^{3}J* = 5.7 Hz, 1H, C2'-H), 2.43 (s, 6H, CH3), 2.30 (m, 1H, C2'-H) ppm.

^{13}C-NMR (101 MHz, CDCl3): δ = 165.7 (COO), 166.1 (COO), 150.1 (C2), 144.7 (Tol4), 129.9 (Tol2), 129.5 (Tol2), 129.4 (Tol3), 129.3 (Tol3), 126.5 (Tol1), 125.9 (Tol1), 101.4 (C5), 85.4 (d, 1J = 12.8 Hz, C1') 83.0 (C4'), 74.7 (C3'), 64.0 (C5'), 38.4 (C2'), 21.7 (CH_3) ppm.

LRMS (ESI$^+$): m/z calculated $[C_{25}H_{23}D_2O_7{}^{15}N_2]^+$ ([M+H]$^+$): 469.17, found: 469.15.

β-[D, $^{15}N_2$]-3',5'-Bis-O-(p-toluoyl)-5-iodo-2'-deoxyuridine



8

A mixture of α-[D_2, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-2'-deoxyuridine and β-[D_2, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-2'-deoxyuridine (267 mg, 0.570 mmol, 1.0 Equiv.) was dissolved in MeCN (6 mL), LiI (92 mg, 0.68 mmol, 1.2 Equiv.) and ceric(IV)-ammoniumnitrate (625 mg, 1.14 mmol,

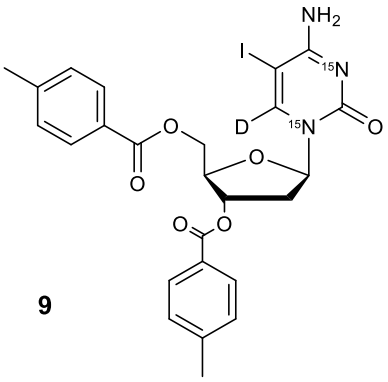
2.0 Equiv.) were added. The suspension was stirred at 80 °C for 3 h. After full conversion of the starting material, saturated aq. NaCl-solution (30 mL) was added, mixture was extracted with DCM, the combined organic layers were dried oved $MgSO_4$ and all volatiles were removed under reduced pressure. Purification by flash column chromatography (*i*Hex/EtOAc 1:1) yielded β-[D, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-5-iodo-2'-deoxyuridine 33 (177 mg,0.298 mmol, 52 %) as a yellow solid.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.99 (d, ^{1}J = 93.6 Hz, 1H, NH), 7.94 (d, ^{3}J = 3.5 Hz, 2H, Tol2), 7.92 (d, ^{3}J = 3.6 Hz, 2H, Tol2), 7.27 (d, ^{3}J = 7.7 Hz, 4H, Tol3), 6.37 (dd, ^{3}J = 8.7 Hz, ^{3}J = 5.4 Hz, 1H, C1'-H), 5.61 (d, ^{3}J = 6.6 Hz, 1H, C3'-H), 4.73 (t, ^{3}J = 2.7 Hz, 2H, C5'-H), 4.57 (q, ^{3}J = 2.7 Hz, 1H, C4'-H), 2.78 (dd, ^{3}J = 14.3, 3J = 5.5 Hz, 1H, C2'-H), 2.43 (s, 3H, CH_3), 2.42 (s, 3H, CH3), 2.28 (m, 1H, C2'-H) ppm.

^{13}C-NMR (101 MHz, $CDCl_3$): δ = 166.1 (COO), 166.0 (COO), 159.6 (d, ^{1}J = 10.4 Hz, C4), 149.7 (t, ^{1}J = 18.9 Hz, C2), 144.7 (Tol4), 144.6 (Tol4), 129.9 (Tol2), 129.7 (Tol2), 129.6 (Tol3), 129.3 (Tol3), 126.4 (Tol1), 126.2 (Tol1), 85.8 (d, ^{1}J = 12.3 Hz, C1'), 83.5 (C4'), 74.9 (C3'), 68.8 (d, ^{2}J = 9.0 Hz, C5), 64.2 (C5'), 38.8 (C2'), 21.8 (CH_3) ppm.

HRMS (ESI$^+$): m/z calculated $[C_{25}H_{23}DO_7{}^{15}N_2I]^+$ ([M+H]$^+$): 592.0480, found:592.0476.

β-[D, $^{15}N_2$]-3',5'-Bis-O-(p-Toluoyl)-5-iodo-2'-deoxycytidine



9

1,2,4-Triazole (42 mg, 0.60 mmol, 9.0 Equiv.) was dissolved in MeCN (1 mL), $POCl_3$ (0.028 mL, 0.29 mmol, 4.4 Equiv.) was added dropwise at 0 °C, the mixture was stirred for 10 min at 0 °C, NEt_3 (0.084 mL,0.60 mmol, 9.0 Equiv.) was added and the reaction solution was stirred at rt for 20 min. A solution of β-[D, ^{15}N2,]-3',5'-bis-O-(p-Toluoyl)-5-iodo-2'-deoxyuridine (40 mg, 0.07 mmol, 1.0 Equiv.) in MeCN (1.5 mL) was added to the reaction and stirred at 30 °C for 18 h. After full conversion of the starting material, NEt_3 (0.4 mL) und H_2O (0.1 mL) were added and stirred for 10 min at 30 °C. Saturated aq. NaCl-solution (5 mL) and saturated aq. $NaHCO_3$-solution (5 mL) were cooled down to 0 °C and added to the reaction mixture. The solution was extracted with DCM, the combined organic layers were dried over

$MgSO_4$ and all volatiles were removed under reduced pressure. The residue was dissolved in 1,4-dioxane (10 mL) and a concentrated NH_3-solution (1.5 mL) was added. The reaction mixture was stirred for 15 min at 30 °C, afterwards an aqueous NH_4Cl-solution (10 mL) was added and extracted with DCM. The combined organic layers were dried over $MgSO_4$ and the volatiles were removed under reduced pressure. Purification of the crude by flash column chromatography (DCM/MeOH 50:1) yielded [D, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-5-iodo-2'-deoxycytidine (17 mg, 0.029 mmol, 43 %) as a yellow solid.

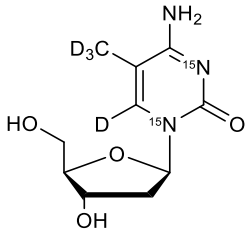
^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.94 (d, ^3J = 8.0 Hz, 2H, Tol2), 7.90 (d, ^3J = 8.1 Hz, 2H, Tol2), 7.29-7.22 (m, 4H, Tol3), 6.36 (dd, ^3J = 8.4 Hz, ^3J = 5.4 Hz, 1H, C1'-H), 5.58 (d, ^3J = 6.4 Hz, 1H, C3'-H), 5.49 (s, 2H, NH2), 4.76 (dd, ^3J = 12.3 Hz, ^3J = 3.0 Hz, 1H, C5'-H), 4.68 (dd, ^3J = 12.3 Hz, ^3J = 3.4 Hz, 1H, C5'-H), 4.59 (q, ^3J = 2.7 Hz, 1H, C4'-H), 3.02-2.92 (m, 1H, C2'-H), 2.42 (s, 3H, CH_3), 2.41 (s, 3H, CH_3), 2.17 (m, 1H, C2'-H) ppm.

^{13}C-NMR (101 MHz, CDCl3): δ = 166.1 (COO), 163.7 (C4), 154.9 (C2), 144.5 (Tol4) 144.5 (Tol4), 129.8 (Tol2), 129.7 (Tol2), 129.5 (Tol3), 129.3 (Tol3), 126.4 (Tol1), 126.3 (Tol1), 87.1 (d, ^1J = 11.2 Hz, C1'), 83.6 (C4'), 75.3 (C3'), 64.2 (C5'), 56.1 (C5), 39.5 (C2'), 21.8 (CH_3) ppm.

HRMS (ESI$^+$): m/z calculated $[C_{25}H_{24}DO_6N^{15}N_2I]^+$ ([M+H]$^+$): 593.0792, found: 593.0791.

5-[D_3]-Methyl-2'-deoxy-[6-D, 1,3-$^{15}N_2$]-cytidine

1



Ni(dppp)Cl_2 (8 mg, 0.01 mmol, 0.5 Equiv.) was dried in a schlenk flask under high vacuum for 30 min. A solution of β-[D, $^{15}N_2$]-3',5'-bis-O-(p-toluoyl)-5-iodo-2'-deoxycytidine (17 mg, 0.029 mmol, 1.0 Equiv.) in THF (1 mL) was added at 0 °C. Afterwards D_3CMgI (0.072 mL, 0.072 mmol, 2.5 Equiv.) was added dropwise and the mixture was stirred at rt for 1.5 h. Another portion of D_3CMgI (0.072 mL, 0.072 mmol, 2.5 Equiv.) was added dropwise and stirred for 1 h at rt. A third portion of D_3CMgI (0.20 mL, 0.20 mmol, 6.9 Equiv.) and Ni(dppp)Cl_2 (8 mg, 0.01 mmol, 0.5 Equiv.) were added to the mixture and stirred for17 h at rt. A saturated aqueous NH_4Cl-solution (8 mL) was added to the reaction solution and extracted with EtOAc. The combined organic layers were dried $MgSO_4$ and all volatiles were removed under reduced pressure The residue was purified by flash column chromatography (DCM/MeOH 20:1) to yield a mixture of β-5-[D_3]-methyl-3',5'-bis-O-(p-toluoyl)-2'-deoxy-[D,$^{15}N_2$,]-cytidine and β-3',5'-bis-O-(p-toluoyl)-2'-deoxy-[D, $^{15}N_2$,]-cytidine as a yellow foam.

The resulting mixture was dissolved in MeOH (1.2 mL), K_2CO_3 (13 mg, 0.094 mmol, 4.7 Equiv.) was added and the solution was stirred for 18 h at rt. Another portion of K_2CO_3 (13 mg,

0.094 mmol, 4.7 Equiv.) and MeOH (1.2 mL) was added and stirred for 17 h at rt. The volatiles were removed under reduced pressure and the residue was purified via HPLC (Macherey-Nagel, Nucleosil 100-7 C18, 10 × 250 mm, linear gradient, 0 % - 10 % MeCN in water in 45 min) to yield [D_4, $15N_2$]-m5dC (0,9 mg,0,004 mmol, 12 %) as a colorless solid.

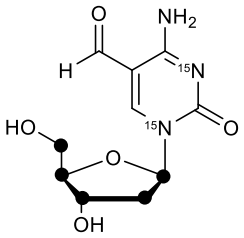
^{1}H-NMR (800 MHz, D_2O): δ = 6.30 (t, ^3J = 6.7 Hz, 1H, C1'-H), 4.46 (dt, ^3J = 6.6 Hz, ^3J = 4.1 Hz, 1H, C3'-H), 4.05 (q, ^3J = 4.2 Hz, 1H, C4'-H), 3.86 (dd, ^3J = 12.5 Hz, ^3J = 3.6 Hz, 1H, C5'-H), 3.78 (dd, ^3J = 12.5 Hz, ^3J = 5.1 Hz, 1H, C5'-H), 2.47-2.36 (m, 1H, C2'-H), 2.38-2.25 (m, 1H, C2'-H) ppm.

^{13}C-NMR (201 MHz, D_2O): δ = 89.1 (C4'), 88.4 (d, ^1J = 13.0 Hz, C1'), 73.1 (C3'), 63.8 (C5'), 41.8 (C2') ppm.

HRMS (ESI$^+$): m/z calculated [$C_{10}H_{12}D_4O_4N^{15}N_2$]+ ([M+H]+): 248.1327, found: 248.1326.

Synthesis of 5-formyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, *N^1,N^3*-$^{15}N_2$)-cytidine

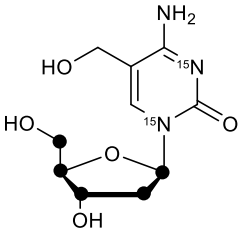
3



5-formyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N1,N3-$^{15}N_2$)-cytidine was prepared according to Iwan *et al.*[6]

1.3 Synthesis of 5-hydroxymethyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, *N^1,N^3*-$^{15}N_2$)-cytidine

2



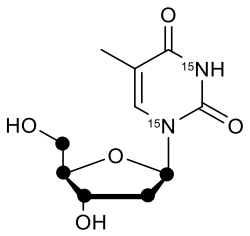
5-formyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N1,N3-$^{15}N_2$)-cytidine (7.4 mg, 0.028 mmol, 1.0 Equiv.), $CeCl_3$*$7H_2O$ (32 mg, 0.085 mmol, 3.0 Equiv.) and $NaBH_4$ (1.0 mg, 0.028 mmol, 1.0 Equiv.) were dissolved in MeOH and stirred for 5 min at rt. Saturated aqeous NH_4Cl solution was added and the mixture was concentrated under reduced pressure. Purification via HPLC (Macherey-Nagel, Nucleosil 100-7 C18, 10 × 250 mm, linear gradient, 0 % - 10 % MeCN in

water in 60 min) yielded 5-hydoxymethyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N1,N3-$^{15}N_2$)-cytidine (quant.).

HRMS (ESI$^+$): m/z calculated $[C_5{}^{13}C_5H_{16}O_5N^{15}N_2]^+$ ([M+H]$^+$): 265.1193, found: 265.1190.

1.4 Synthesis of 5-methyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, *N*^{1},*N*^{3}-$^{15}N_2$)-uridine

4



The detailed synthesis of this compound will be published in a forthcoming manuscript.

References

1. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**, 574–581 (2014).
2. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**, 1221–1234 (2017).
3. Tsukiyama, T. & Ohinata, Y. A Modified EpiSC Culture Condition Containing a GSK3 Inhibitor Can Support Germline-Competent Pluripotency in Mice. *PLoS ONE* **9**, e95329 (2014).
4. Respuela, P. *et al.* Foxd3 Promotes Exit from Naive Pluripotency through Enhancer Decommissioning and Inhibits Germline Specification. *Cell Stem Cell* **18**, 118–133 (2016).
5. Kurek, D. *et al.* Endogenous WNT Signals Mediate BMP-Induced and Spontaneous Differentiation of Epiblast Stem Cells and Human Embryonic Stem Cells. *Stem Cell Reports* **4**, 114–128 (2015).
6. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C–C bond cleavage *in vivo*. *Nature Chemical Biology* **14**, 72–78 (2017).