bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Oxidative and non-oxidative active turnover of genomic methylcytosine in distinct pluripotent states

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, 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 unveiled an additional mdC oxidation-independent turnover process based on 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.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Methylation at position 5 of deoxycytidine (dC) is the most prominent post-replicative nucleobase modification in higher eukaryotic genomes and its functional relevance has been extensively investigated[1,2]. In mammals, genomic dC is methylated to 5-methyl-2'-deoxycytidine (mdC) by the DNA (cytosine-5)-methyltransferases DNMT1, DNMT3a and DNMT3b[3], mostly, but not exclusively in the context of CpG dinucleotides[4]. These enzymes differentially contribute to establishing and maintaining genomic mdC patterns across cellular and organismal generations. In addition, genomic mdC can be 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[5–7]. As several classes of DNA binding factors display differential affinity for their target sequences depending on the dC modification state[2,1,8], 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 and restriction, imprinting, X chromosome inactivation and genome stability[1,2].

In the last two decades methylation of genomic dC was shown to be rapidly removed in various biological contexts[9]. 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[9] (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[9,10] and Neil1/2[11,12] is best characterized. Additional proposals invoke the involvement of other DNA repair pathways[9] or direct C-C bond cleavage[13–16]. A third line of proposal posits that genomic mdC and hmdC could be deaminated to thymidine (dT) and 5-hydroxymethyl-2'-deoxyuridine (hmdU), respectively, either by a cytosine deaminase of the AID/APOBEC family[9] or by *de novo* DNA methyltransferases DNMT3a/b[17]. The resulting T:G or hmU:G mismatches would be then resolved by BER or non- canonical mismatch repair (ncMMR)[18–21]. 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[20–22]. Currently, all these pathways are highly controversial, as there is no conclusive evidence for their functional relevance *in vivo*[9]. 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[23]. This transition can be modelled by cultivating mouse pluripotent stem cells (mPSCs) under specific conditions and involves dynamic changes in expression and

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

epigenetic landscapes, including a major global gain of mdC[23]. 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[24]. 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

Global genomic mdC oxidation kinetics decrease upon progression to primed pluripotency.

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[25]. 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[26]. Live cell microscopy over the course of 6 days without passaging showed that these cells rapidly formed monolayer colonies that retained a remarkably homogeneous nuclear YFP signal across the culture (Supplementary Fig. S1a), indicating that under these conditions mPSC cultures may not loose 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; Supplementary Fig. S1a). 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 (Supplementary Fig. S1b), 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

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

chromatography coupled with triple quadrupole mass spectrometry (UHPLC-MS^2)[27] to analyse global levels of genomic mdC and hmdC upon priming of mPSCs under the same conditions over a five day time course. As previously reported[28], the transition from the naïve (2i/LIF) to a homogeneously primed pluripotent state was marked by a large increase in global levels of genomic mdC, while hmdC abundance dropped sharply, both modifications reaching plateau levels by day 4 (Supplementary Fig. S2a). 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 was shown to greatly increase the levels of oxidized mdC derivatives[29]. The decrease in hmdC abundance 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[30].

Notably, under 2i/LIF conditions, mPSCs exhibit low levels of genomic cytosine modification[23], while replacement of the Mek inhibitor with the Src inhibitor CGP77675 (conditions referred to as alternative 2i or a2i) results in relatively high genomic mdC levels[31,32]. Nevertheless, gene expression patterns remain similar and developmental potential unaltered relative to those of naïve mPSCs[31,32]. In particular, our expression analysis showed that the pattern of pluripotency and early lineage specification factors in mPSCs under a2i/LIF conditions was intermediate to that of naïve and primed states, with still high levels of Tfcp2l1 and Klf4 transcripts relative to 2i/LIF (naïve) conditions, but intermediate expression of Rex1, Stella and Prdm14 as well as low levels of Fgf5, Oct6, Otx2 and Dnmt3b relative to primed conditions (Fig. S1b). This pattern is highly similar to that previously shown for formative pluripotency, an intermediate state along the trajectory from naïve to primed pluripotency, during which competence for multi-lineage commitment is acquired[33]. In addition, high depth bisulfite amplicon sequencing revealed increased *de novo* methylation at secondary imprints in the transition from a2i/LIF to serum/CR conditions (Supplementary Fig. S2b), which is similar to the progression previously shown from pre- to early post-implantation stages in mouse embryos[34]. Altogether, our expression and genomic methylation analyses support that cultures transited from a2i/LIF to serum/CR conditions recapitulate fundamental molecular processes of the progression to a genuine primed pluripotent state. Importantly, prolonged culture under 2i/LIF conditions leads to loss of imprinted methylation and genetic instability, both are persevered in a2i/LIF long term cultures[31,32]. In addition, a2i/LIF and serum/CR conditions allow comparison of genomic cytosine modification dynamics between functionally distinct pluripotent states with

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 set out to monitor global cytosine modification kinetics during acquisition of primed pluripotency by metabolic labelling of the C5 substituent. To this aim we supplemented L-Methionine (Met)-free medium with isotopically labelled Met (m+4Met), whereby the methyl group provides a mass shift of four units ([*methyl*-^{13}C,d_3], hereafter referred to as m+4). As for naturally occurring Met, m+4Met is incorporated into S-adenosylmethionine (SAM) and m+4 is transferred thereof onto 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)[27].
In order to attain accurate absolute quantification of both naturally occurring (mdC, hmdC, fdC) and heavy (m+4dC, hm+3dC, f+2dC) isotopologues, we synthesised an even heavier series of corresponding isotopologues ([d_4,$^{15}N_2$]-mdC, [$^{13}C_5$,$^{15}N_2$]-hmdC and [$^{13}C_5$,$^{15}N_2$]-fdC; Fig. 1b) and used them as internal standards for quantification by UHPLC-MS^2. In this work cadC and ca+1dC were not considered as isotopologues differing by only one mass unit cannot be reliably discriminated using UHPLC-MS^2. 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 day 5 after transfer to serum/CR medium (Fig.s 1c,d, Supplementary Fig. S3). Relative to a2i/LIF conditions, in priming cultures the rate of m+4dC oxidation to hm+3dC was clearly slower. Accumulation of the further oxidized state f+2dC became undetectable under these conditions. 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 (Supplementary Fig. S2a) and with a recent report of lower levels of chromatin bound Tet proteins in primed cultures[30]

An active and largely Tet enzyme-dependent component of genomic mdC turnover in primed pluripotent cells

To monitor the turnover of modified genomic cytosines we reversed the labelling schema described above. We first pulsed labelled cultures with m+4Met, then chased the label for two days by transferring the cultures to medium with naturally occurring Met and collected samples every 12 h during the chase (Fig. 2a). We again compared cultures maintained constantly under serum/a2i/LIF conditions to cultures undergoing priming in serum/CR medium. In both cases the cultures were pulsed for five days with m+4Met before chasing.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In the case of the 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-pulsed 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; Fig. 2a). Due to its highly efficient incorporation, dT+12 was co-pulsed only during the last day of the m+4Met pulse. Thereafter, dT+12 was chased for two days in parallel to 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 became quite steep, pointing to high turnover of fdC. Although under priming conditions oxidation of mdC to hmdC and further to fdC was hardly detectable in labelling time course experiments (Fig. 1), the fact that in the pulse chase experiments the decline of labelled hmdC and fdC was slower than that of dT+12 clearly shows that some mdC and hmdC oxidation still takes place (Fig 2b, central panel). Importantly, upon priming, the turnover of m+4dC is faster than the dilution rate of dT+12 (Fig. 2b, inset), revealing the presence of a component of the m+4dC turnover that cannot be explained by passive dilution through DNA synthesis.

To assess the contribution of Tet-dependent oxidation to this component we performed the same pulse/chase experiment under priming condition (Fig. 2a, upper part) with E14tg2a mPSCs lacking all three Tet proteins (Tet triple knockout, alias Tet TKO)[35] as well as parental, Tet 1-3 proficient E14tg2a cells (Fig. 3; note that E14tg2a cells are Hprt null[36]). In the parental cell line (Tet1-3 proficient), 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. 2c). 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 mdC turnover through a DNA repair process

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Next to oxidative mdC turnover, we started to investigate a potential deamination based pathway. Deamination of mdC and hmdC to dT (mdU) and hmdU, respectively, has been proposed as a potential alternative first step for active erasure of genomic C methylation. 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) 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[37]. After priming in the presence of m+4Met, m+4dU was clearly detected in the genome of parental wt J1 mPSCs, but not in Dnmt TKO cultures. Thus, generation of m+4dU strictly requires the presence of m+4dC, showing that m+4dU originates exclusively through deamination of m+4dC (Supplementary Fig. S4).

Direct deamination of genomic mdC would imply the formation of T:G mismatches, which, if not efficiently repaired, would be highly mutagenic. It would then be reasonable to expect that a mechanism for highly efficient coupling of mdC deamination with BER of the ensuing T:G mismatches would have co-evolved[19] so as to minimize the steady state levels of such mismatches. In contrast to this idea, we surprisingly detected substantial accumulation of m+4dU in genomic DNA, whenever cultures were provided with m+4Met (gDNA; Fig. 1-5, Supplementary Fig. S3-4, S7). To explain this observation we considered an alternative scenario whereby genomic strands containing m+4dC may first be subject to a DNA repair process leading to the release of soluble m+4dC/m+4dCMP, either by direct exonuclease action, as upon MMR and end resection of double strand breaks, or through digestion of excised strands (e.g., in nucleotide excision repair). Deamination of this material by Cytidine and Deoxycytidylate deaminases (Cda and Dctd)[38], would generate soluble m+4dU/m+4dUMP, which could be phosphorylated and re-incorporated into the genome through DNA synthesis (Fig. 4a). To test if the high levels of m+4dU detected in the genome are generated along this pathway, we supplemented the medium with m+4Met and increasing concentrations of unlabelled dT. If DNA repair generates soluble m+4dU/m+4dUMP, incorporation of the latter into gDNA will be outcompeted by unlabelled dT. In addition, a dT excess will lead to high concentrations of dTMP and dTTP, which are both known to feedback negatively on Dctd through product inhibition and allosteric regulation, respectively[39]. The accumulation of genomic m+4dU in priming mPSC cultures was indeed progressively reduced with supplementation of increasing concentrations of

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

unlabelled dT (Fig. 4b). This result supports the idea that genomic m+4dU originates in large part from a pool of soluble precursor.

To further probe whether deamination of m+4dC takes place within gDNA or at the level of soluble nucleoside/nucleotide pools, we next used CRISPR/Cas9-mediated gene editing to generate mPSC with double deficiency for Cda and Dctd (CD-DKO; Supplementary Fig. 5a,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 Fig S5c-e). All three CD-DKO clones grew at slower rates than parental wt cells, a phenotype which was at least partially relieved by supplementation of the medium with dT (Supplementary Fig. S5f). This indicates that deamination of (d)C/(d)CMP by Cda and/or Dctd provides substantial amounts of (d)U/(d)UMP for biosynthesis of (d)T/(d)TMP by Thymidine synthase to support DNA replication and is fully consistent with previously reported dT auxotrophy of Dctd deficient somatic cell lines[40]. 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 J1 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 the genomic m+4dU measurable in wt cells with our UHPLC-MS^2 method originates from a process involving direct repair of gDNA containing m+4dC, with consequent release of soluble m+4dC/m+4dCMP, followed by deamination to m+4dU/m+4dUMP and (re)-incorporation of the latter into the genome through phosphorylation to the triphosphate state and DNA synthesis. Although our data point to exclusive deamination in the soluble pool, we wanted to further investigate potential deamination events of mdC to T directly in the genome, particularly in light of previous reports in this direction[17,19]. In this respect it is important to consider the limitations of our quantitative UHPLC-MS^2 method. In UHPLC the overwhelming amount of natural dT present in the genome co-elutes with m+4dU and potentially outcompetes it during ionization for the capture of charge. The effect is a strong ion suppression, resulting in a reduction of the m+4dU signal and, consequently, a high limit of detection (LOD). To put this effect into perspective we calculated that our LOD for m+4dU is about 100 times higher than that for fdC (around $1x10^{-5}$ versus $1x10^{-7}$ per nucleotide), which is a relatively stable dC modification considered to have an epigenetically relevant role.

However, deamination of very few m+4dC residues directly within the genome could trigger long patch BER and/or MMR[20,21], leading to the release of much larger amounts of m+4dC in the soluble pool and, subsequently, deamination to m+4dU and re-incorporation into the

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genome, explaining the large amounts of genomic m+4dU observed. In CD-DKO cells direct deamination of genomic m+4dC would be expected to be the only potential source of m+4dU and, as the latter would then be mismatched to G, its steady state levels would be expected to be very low due to efficient repair. In order to probe genomic deamination with higher sensitivity the ion suppression effect had to be substantially reduced. To this end we performed a double labelling experiment with m+4Met and a dC nucleoside where 9 hydrogen atoms are replaced with deuterium (dC[D_9]; Supplementary Fig. S5a). After incorporation into the genome, dC[D_9] will be methylated to m+4dC[D_8], which deaminates to m+4dU[D_8] (a dT with eleven D atoms and a +12 mass shift; Supplementary Fig S6a). This massive H-to-D exchange was chosen as it is known to determine a shift to shorter retention time in UHPLC, 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 (Supplementary Fig. S6b). Notably, m+4dU[D_8] is expected to have an even shorter retention time that would allow it to at least partly escape ion suppression from natural dT. After double labelling with m+4Met and dC[D_9] we were able to detect m+4dU[D_8] confidently only in wt mPSCs (Supplementary Fig. S6c). However, 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, defined as three times the average level of blank samples. Although we cannot exclude that these signals are caused by background processes, we nevertheless performed a calculation taking into account the limits of detection and quantification that were gained by deuterium labelling. This calculation allows us to estimate the maximum number of potential genomic deamination events between 500 and 5000 m+4dU[D_8] molecules per cell in the course of our labelling experiment. According to an estimated rate constant for spontaneous hydrolytic deamination of mdC in solution[41] only around 20 such events would be expected to take place. Therefore, although a reliable and reproducible signal for m+4dU[D_8] could not be detected in CD-DKO cells, we cannot exclude the occurrence of genomic deamination above the rates of spontaneous hydrolysis. However, in the best case this process cannot exceed a level of $1x10^{-6}$-$1x10^{-7}$ events per nucleotide.

Tet enzymes do not trigger the DNA repair process responsible for turnover of genomic mdC into dT

Conversion of m+4dC into m+4dU obviously 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[42]. Also, hmdC may mark DNA damage sites and promote their repair[43] or even

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

target DNA repair by selectively recruiting endonucleases like Srap1[44] and Endonuclease G[45]. 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[27], 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 (Fig. S7). Similarly, analysis of the pulse chase experiment with the same pair of cell lines also revealed a tendency for higher accumulation of m+4dU in Tet TKO mPSCs at later chase time points (Fig. 3, lower panels). Therefore, in pluripotent cells Tet proteins do not seem to trigger DNA repair events that contribute to turnover of mdC into dT.

Oxidation-independent repair of mdC-containing gDNA 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 naïve and priming conditions described above (Fig. 1-3, Supplementary Fig. S3). In time course experiments, progressive accumulation m+4dU was much faster under priming conditions, reaching approximately five times higher levels than under naïve conditions (Fig. 1, Supplementary Fig. S3). 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. 2b,3). Thus, in contrast with the observed decline of mdC oxidation (Fig. 1,2, Supplementary Fig. S3), these results reveal increased rates of oxidation-independent repair of mdC-containing gDNA upon acquisition of primed pluripotency. This anti-correlation suggests that oxidation-independent mdC turnover through DNA repair may serve at least in part as a supplementary pathway for the erasure of dC methylation. However, as in primed Tet TKO mPSCs there was no apparent difference between global mdC turnover and passive dilution through DNA synthesis (Fig. 3: decay of m+4dC and dT+12, respectively), this process cannot involve more than 1% of the methylome.

At last, we compared oxidation-independent mdC turnover through DNA repair in pluripotent states (intermediate serum/a2i/LF as well as primed serum/CR and EpiSCs under CDM/CHIR/IWR1/FGF-2/Activin A conditions) with various lineage committed cell types, including primary and immortalized embryonic fibroblasts, neural stem cells (NSCs)[46] 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 cultured for substantially longer periods, ranging from 6 to 12 days (Fig. 5a). This

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

clearly shows that in lineage committed cells oxidation-independent mdC turnover through DNA repair is substantially lower than in primed pluripotent cells. In addition, the rate of oxidation-independent mdC 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 oxidation-independent mdC 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 oxidation-independent mdC turnover is under developmental and cell lineage-dependent control.

Discussion

Recent methylome analysis of matched clonal and polyclonal cell populations[47] and studies on the contribution of Dnmt3 proteins to DNA methylation maintenance[48,49] 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 pulse-chase experiments with m+4Met to estimate the global turnover of genomic dC modifications in mPSCs cultured under naïve conditions (2i/LIF), using a very short pulse time[50,51]. However, the naïve pluripotent state is characterised by very low global levels of genomic dC modifications, which increase only upon exit from the naïve state and acquisition of primed pluripotency[1,23] , a crucial transition to lay the ground for commitment to somatic lineages.

Here, we report the first 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 and resemble those reported for the formative state[33], the rates of mdC oxidation to hmdC and further to fdC are quite sustained and 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[52]. In contrast to the high levels of mdC oxidation in a2i/LIF conditions, upon transition to a more advanced, primed pluripotency state 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) and 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

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

embryonic development and the reversion from heterogeneous partially primed/formative states present in serum/LIF cultures to the naïve pluripotent state[1,24]. We suggest that against the backdrop of global gain and maintenance of dC methylation occurring 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, especially in the case of loci with a highly accessible chromatin conformation like actively transcribed genes.

Furthermore, by tracing the origin of genomic m+4dU to deamination of soluble m+4dC, we show that a small part of genomic mdC is turned over by a DNA repair process that involves its direct excision from the genome (Fig. 5b). 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 the DNA repair process underlying this turnover pathway. Although the nature of this DNA repair process and the events triggering it remain to be determined, its rate is not proportional to global genomic mdC levels, either between pluripotent states or between lineage committed cells, indicating that the underlying DNA repair process is developmentally regulated. In this regard, it should be noted that in PSCs, including primed human PSCs, the expression of DNA repair factors and the proficiency of various DNA repair pathways were shown to be higher than in somatic cells[53–55]. Interestingly, the extent of oxidation-independent turnover of mdC through DNA repair is not proportional to global genomic mdC levels, either between pluripotent states or lineage committed cells, indicating that the underlying DNA repair process does not take place randomly throughout the methylome. In addition, 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. Whether this is indeed the case requires further investigation.

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 mPSCs and Dnmt TKO J1, respectively; Yi Zhang (Boston Children's Hospital, Boston, MA) for parental and Tet TKO E14tg2a. Anti-DCTD antibody and purified recombinant DCTD protein were generous gifts from Frank Maley (New York State University, NY), Sebastian Bultmann and Christopher

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mulholland for guidance on high depth bisulfite amplicon sequencing (both at Ludwig Maximilian University, Munich).

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, OK and JS 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/MS. FS, TC and SS wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data availability and code availability statements

The raw data that support the findings of this study are available from the corresponding authors upon reasonable request. No own code was used.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science* **361**, 1336–1340 (2018).
3. Jeltsch, A. & Jurkowska, R. Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. *Nucl. Acids Res.* **44**, 8556–8575 (2016).
4. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annual Review of Genomics and Human Genetics* **16**, 55–77 (2015).
5. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324**, 930–935 (2009).
6. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**, 1300–1303 (2011).
7. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, 1303–1307 (2011).
8. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152**, 1146–1159 (2013).
9. Schuermann, D., Weber, A. R. & Schär, P. Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair* **44**, 92–102 (2016).
10. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18**, 517–534 (2017).
11. 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).
12. Schomacher, L. *et al.* Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature Structural & Molecular Biology* **23**, 116–124 (2016).
13. 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).
14. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D. & Klimašauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology* **5**, 400–402 (2009).
15. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C–C bond cleavage *in vivo*. *Nature Chemical Biology* **14**, 72–78 (2017).
16. Schön, A. *et al.* Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells. *Angewandte Chemie International Edition* AOP (2020) doi:10.1002/anie.202000414.
17. Metivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
18. Cortellino, S. *et al.* Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146**, 67–79 (2011).
19. Rai, K. *et al.* DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. *Cell* **135**, 1201–1212 (2008).
20. 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).
21. Franchini, D.-M. *et al.* Processive DNA Demethylation via DNA Deaminase-Induced Lesion Resolution. *PLoS ONE* **9**, e97754 (2014).
22. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics & Chromatin* **6**, 39 (2013).
23. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews Genetics* **18**, 643–658 (2017).
24. 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).
25. 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).
26. 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).
27. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**, 574–581 (2014).
28. 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).

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

29. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
30. 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).
31. Yagi, M. *et al.* Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
32. Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
33. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**, 1221–1234 (2017).
34. Wang, L. *et al.* Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell* **157**, 979–991 (2014).
35. 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).
36. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
37. 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).
38. 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).
39. 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).
40. 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).
41. 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).
42. Ladstätter, S. & Tachibana-Konwalski, K. A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming. *Cell* **167**, 1774-1787.e13 (2016).
43. 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).
44. Kweon, S.-M. *et al.* Erasure of Tet-Oxidized 5-Methylcytosine by a SRAP Nuclease. *Cell Reports* **21**, 482–494 (2017).
45. 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).
46. Liu, N. *et al.* Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. *PLoS ONE* **8**, e62755 (2013).
47. Shipony, Z. *et al.* Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014).
48. Ooi, S. K. *et al.* Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics & Chromatin* **3**, 17 (2010).
49. 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).
50. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* **6**, 1049–1055 (2014).
51. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol* **11**, 555–557 (2015).
52. Rulands, S. *et al.* Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell Systems* **7**, 63-76.e12 (2018).
53. Tichy, E. D. *et al.* Mouse Embryonic Stem Cells, but Not Somatic Cells, Predominantly Use Homologous Recombination to Repair Double-Strand DNA Breaks. *Stem Cells and Development* **19**, 1699–1711 (2010).
54. Tichy, E. D. *et al.* Mismatch and base excision repair proficiency in murine embryonic stem cells. *DNA Repair* **10**, 445–451 (2011).
55. Maynard, S. *et al.* Human Embryonic Stem Cells Have Enhanced Repair of Multiple Forms of DNA Damage. *STEM CELLS* **26**, 2266–2274 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

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 balls in the ribose represent ^{13}C atoms. **c** and **d**) Time course analysis of genomic dC derivatives upon metabolic labelling with m+4Met under naïve (c; a2i/LIF) and priming (d; serum/CR) conditions. The labelling/time course schedules are shown at the top. Global levels of unlabelled (pale yellow) mdC and m+4dC (dark yellow), hm+3C (blue) and f+2C (green) are shown as mean and standard deviation of three technical replicates. See also Fig. S3 for an independent biological replicate. LOD = limt of detection.

Figure 2. Turnover of genomic mdC and its derivatives under naïve and priming conditions

a) Workflow of m+4Met and dT+12 pulse-chase experiments with wild type mPSCs under naïve and priming conditions. **b**) Global profiles of the indicated labelled cytosine derivatives and dT+12 in the genome of naïve (a2i/LIF; left) and primed mPSCs (serum/CR; right and inset) upon chase of m+4Met. Upper panels show values relative to the highest level for each modification. Dashed lines represent first order decay fitted curves. Lower panels show absolute values of genomic m+4Met (note the different scales for m+4dU in the two conditions). Mean values and standard deviation of triplicate measurements are shown. **c**) Percent of difference between fold changes of dT+12 and m+4dC in wt mPSCs under priming conditions. Shown are mean and standard deviation of three independent biological replicates.

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). Profiles of global genomic levels of the indicated labelled cytosine derivatives and dT+12 are shown. Upper and mid panels show m+4dC and dT+12 values relative to the highest data point for each modification. Upper panels display magnifications of mid panels fields in dotted frames. Dashed lines represent first order decay fitted curves. Lower panels display global genomic levels of m+4dU.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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 mean and standard deviation of three technical replicates. LOD = limit of detection.

Figure 5. mdC turnover through oxidation-independent DNA repair is developmentally regulated.

a) Cultures of mPSCs in naïve (a2iL = a2i/LIF) and primed states (CR = CHIR/IWR1; FA = FGF-2/Activin A; CDM = Chemically Defined Medium), 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 measure in the same culture labelled with m+4Met. Mean and standard deviation of three technical replicates are shown. **b**) Schematic summary of the dC modification turnover pathways identified in primed pluripotent cells. Deamination of genomic mdC followed by BER could not be detected and may occur only at levels below the sensitivity of our assay.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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[26,56]. 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, except for PR8 cultures where coating with 2.2 µg/cm^2 Laminin 521 (BioLamina) was necessary. For metabolic labelling experiments with m+4Met under all conditions mPSC basal medium was generated with 0.2 mM L-Methionine- and L-Cysteine-free DMEM (Sigma or Life Technologies) supplemented with m+4Met ([*methyl*-^{13}C,d$_3$]) and L-Cysteine. EpiSCs in chemically defined serum-free medium (CDM) were derived from the Oct4-YFP reporter mPSC line OLY2-1[25]. 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[46]. 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[57] and 20 ng/ml each of FGF-2 and EGF (PerpoTech). Primary mEFs (CF-1, Applied StemCell) were cultured in mPSC basal medium. Somatic cell lines were cultured in DMEM containing 10% FBS (Life Technologies), 2 mM L-glutamine and 0.1 mM β-mercaptoethanol.

bioRxiv preprint doi: https://doi.org/10.1101/846584; this version posted March 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Generation of CD-DKO mPSCs by CRISPR/Cas9 editing

Single guide RNAs were designed over exon1 and 4 of Cda and Dctd, respectively (Fig. S5), 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)[58]. 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[59]. 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[60] were probed with a rabbit anti-DCTD antibody[39].

Isolation of genomic DNA and UHPLC-MS^2 analysis

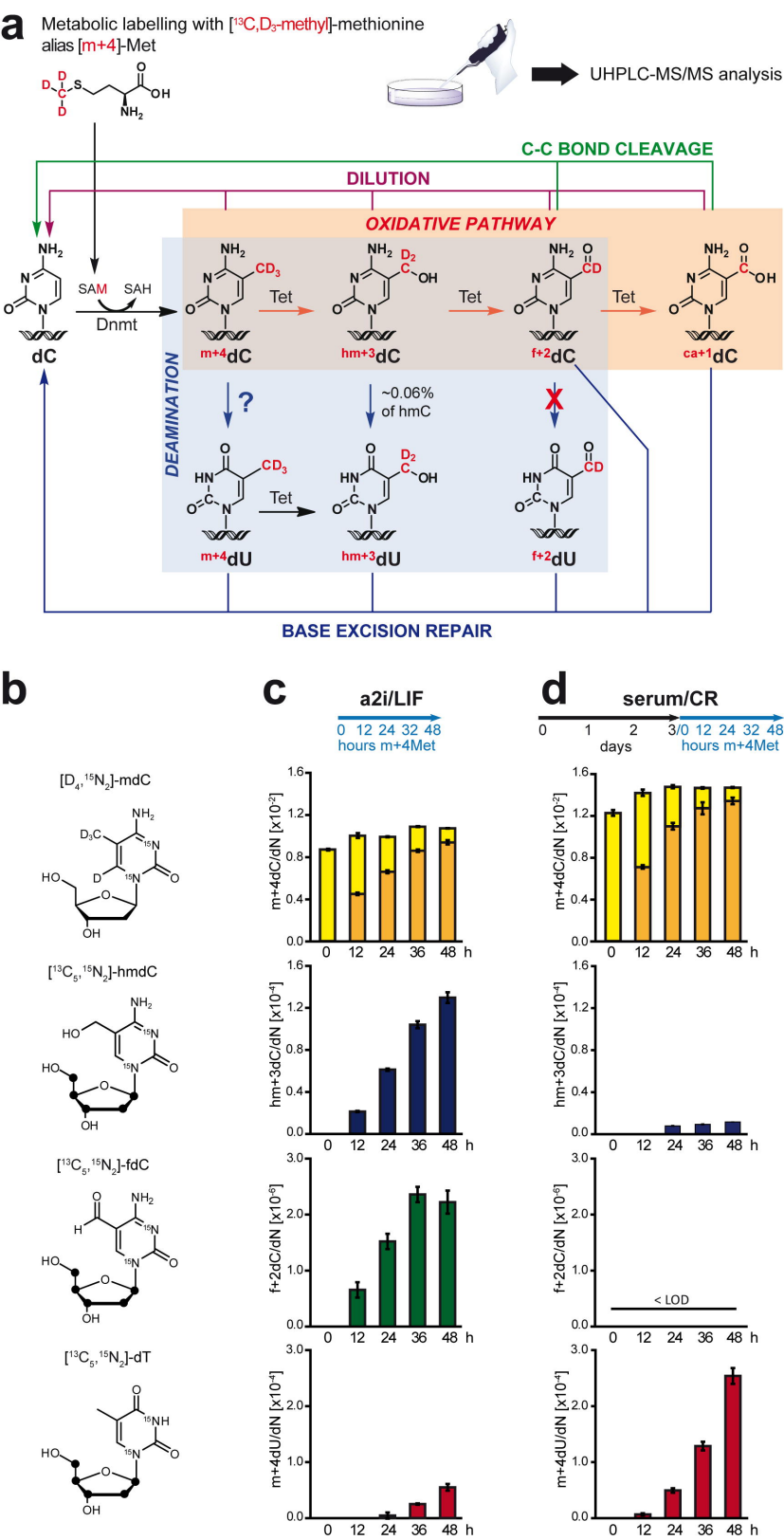
Isolation, digestion and UHPLC-MS^2 analysis of gDNA samples were performed as described in Traube et al.[61] with the following modifications: for every technical replicate 4 µg of gDNA were digested and analysed; the enzyme mixture including S1 Nuclease, Antarctic Phosphatase and snake venom Phosphodiesterase was used for digestion; the digestion mixture was supplemented with 0.8 µM Tetrahydrouridine (Abcam) to inhibit potential cytidine deaminase activity in enzyme preparations.

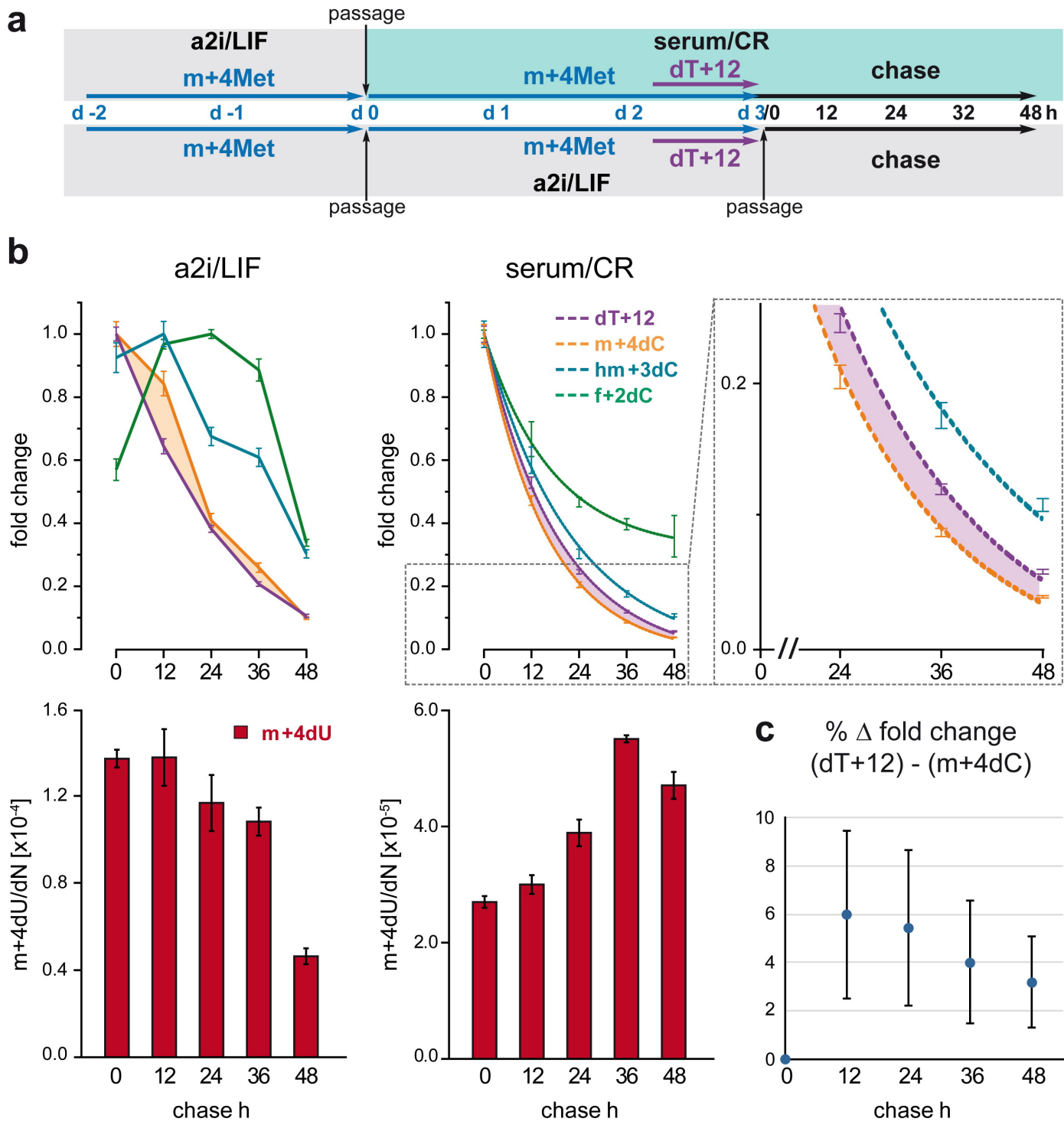
Chemical synthesis

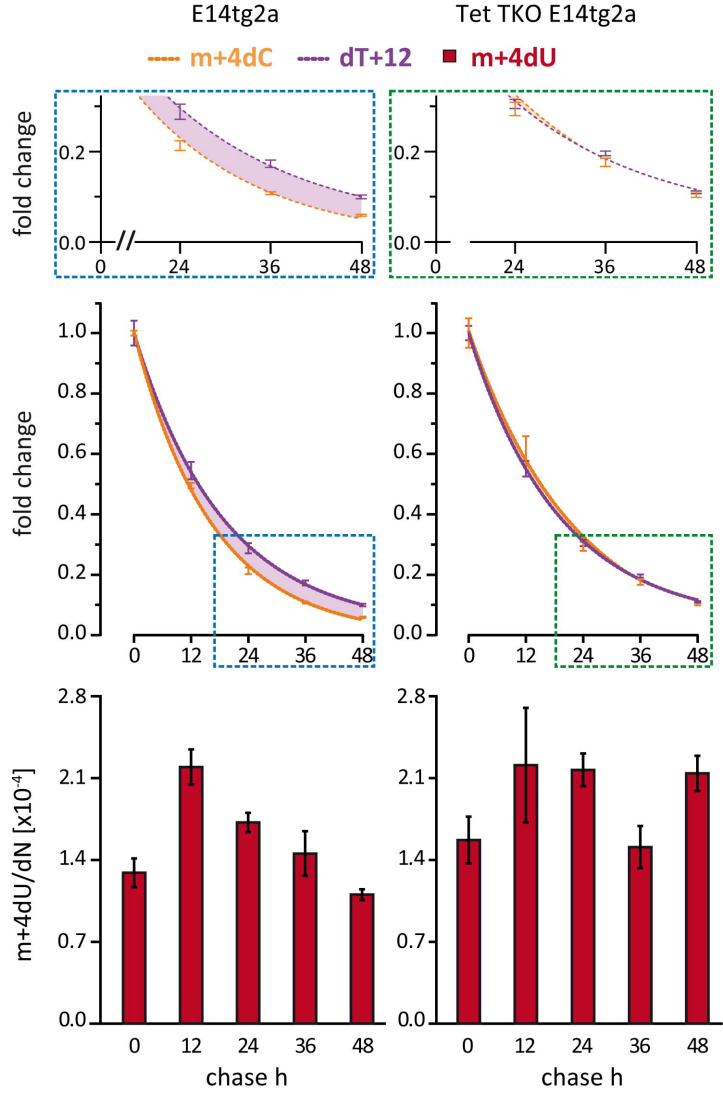
For a description of the synthesis of labelled nucleosides see Supplementary Methods in Supplementary Information file.

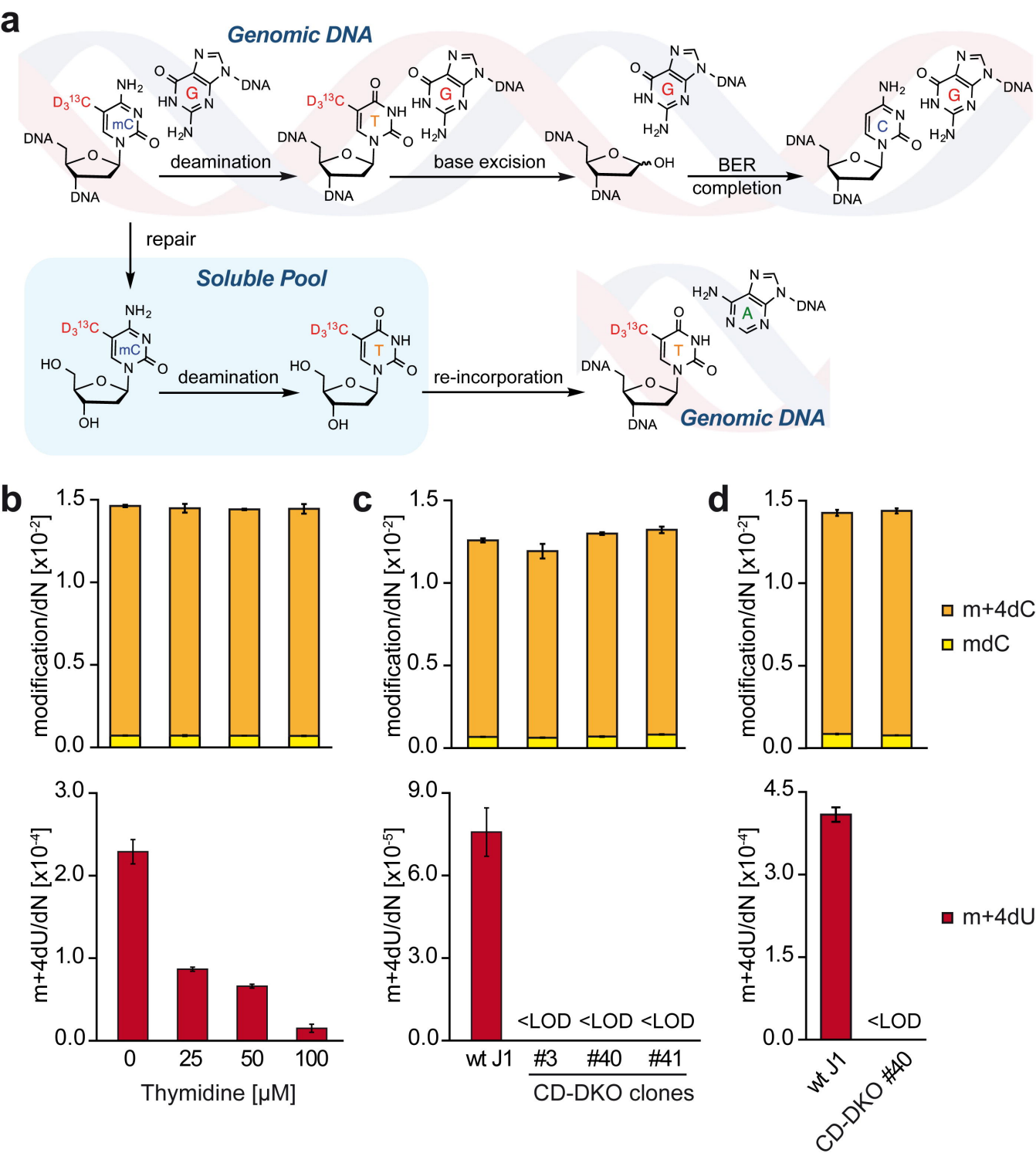
Methods-specific references

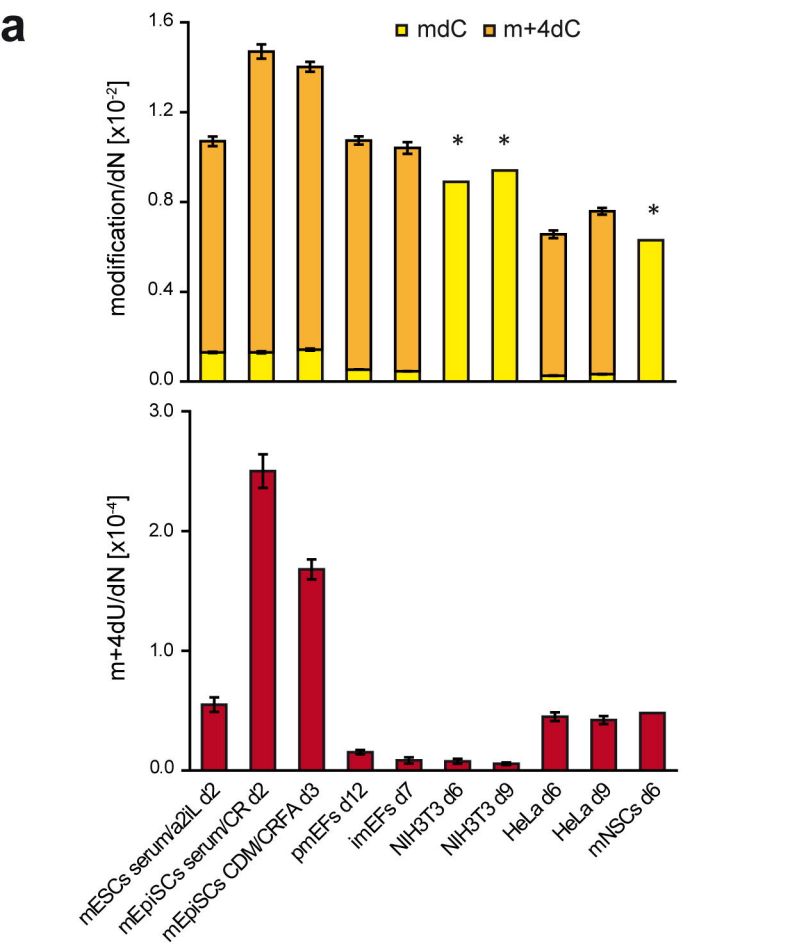
56. Zhou, X., Chadarevian, J. P., Ruiz, B. & Ying, Q.-L. Cytoplasmic and Nuclear TAZ Exert Distinct Functions in Regulating Primed Pluripotency. *Stem Cell Reports* **9**, 732–741 (2017).
57. Nichols, J. & Ying, Q.-L. Derivation and propagation of embryonic stem cells in serum- and feeder-free culture. in *Embryonic Stem Cell Protocols* (ed. Turksen, K.) vol. 329 91–98 (Humana Press, 2006).
58. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
59. 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).
60. 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).
61. Traube, F. R. *et al.* Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nature Protocols* **14**, 283 (2019).

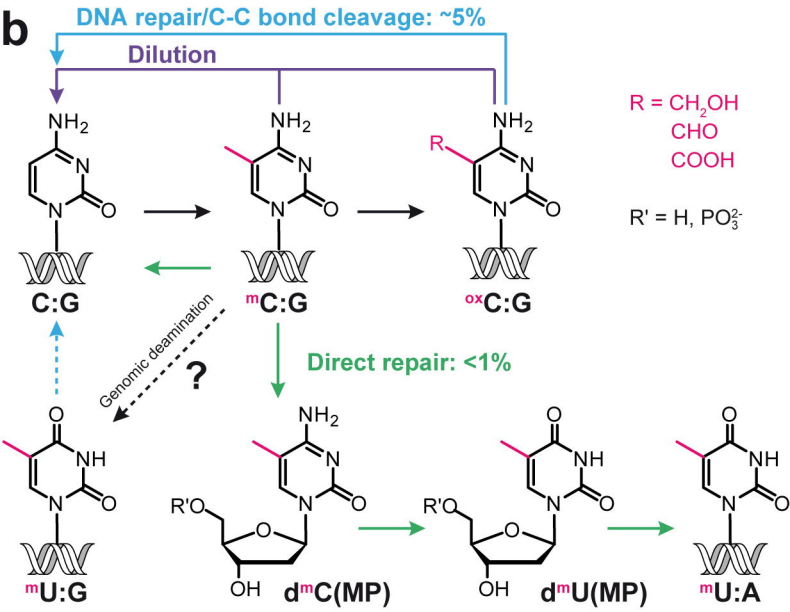

a
Metabolic labelling with [13C,D3-methyl]-methionine
alias [m+4]-Met
UHPLC-MS/MS analysis
C-C BOND CLEAVAGE
DILUTION
OXIDATIVE PATHWAY
SAM
SAH
Dnmt
dC
m+4dC
hm+3dC
f+2dC
ca+1dC
Tet
DEAMINATION
?
~0.06%
of hmC
m+4dU
hm+3dU
f+2dU
BASE EXCISION REPAIR
b
[D4,15N2]-mdC
[13C5,15N2]-hmdC
[13C5,15N2]-fdC
[13C5,15N2]-dT
c
a2i/LIF
0 12 24 32 48
hours m+4Met
m+4dC/dN [x10-2]
hm+3dC/dN [x10-4]
f+2dC/dN [x10-6]
m+4dU/dN [x10-4]
d
serum/CR
0 1 2 3
days
0 12 24 32 48
hours m+4Met
< LOD


a
passage
a2i/LIF
serum/CR
m+4Met
m+4Met
dT+12
chase
d -2
d -1
d 0
d 1
d 2
d 3
0
12
24
32
48 h
m+4Met
m+4Met
dT+12
chase
a2i/LIF
passage
passage
b
a2i/LIF
serum/CR
fold change
dT+12
m+4dC
hm+3dC
f+2dC
m+4dU
m+4dU/dN [x10-4]
m+4dU/dN [x10-5]
chase h
c
% Δ fold change
(dT+12) - (m+4dC)
chase h


E14tg2a
Tet TKO E14tg2a
m+4dC
dT+12
m+4dU
fold change
0.0
0.2
0
24
36
48
1.0
0.8
0.6
0.4
12
2.8
2.1
1.4
0.7
m+4dU/dN [x10-4]
chase h


a
Genomic DNA
deamination
base excision
BER
completion
repair
Soluble Pool
deamination
re-incorporation
Genomic DNA
b
c
d
modification/dN [x10-2]
m+4dC
mdC
m+4dU/dN [x10-4]
m+4dU/dN [x10-5]
m+4dU
<LOD
0
25
50
100
Thymidine [µM]
wt J1
#3
#40
#41
CD-DKO clones
CD-DKO #40


a
mdC
m+4dC
modification/dN [x10-2]
1.6
1.2
0.8
0.4
0.0
*
*
*
m+4dU/dN [x10-4]
3.0
2.0
1.0
0.0
mESCs serum/a2iL d2
mEpiSCs serum/CR d2
mEpiSCs CDM/CRFA d3
pmEFs d12
imEFs d7
NIH3T3 d6
NIH3T3 d9
HeLa d6
HeLa d9
mNSCs d6


b
DNA repair/C-C bond cleavage: ~5%
Dilution
R = CH2OH
CHO
COOH
R' = H, PO3 2-
C:G
mC:G
oxC:G
Genomic deamination
?
Direct repair: <1%
mU:G
dmC(MP)
dmU(MP)
mU:A