**5-formylcytosine is an activating epigenetic mark for Pol III during zygotic reprogramming**

Eleftheria Parasyraki1,8, Medhavi Mallick1,2,8, Victoria Hatch1,3,8, Viviana Vastolo1, Michael U. Musheev1, Emil Karaulanov1, Alexandr Gopanenko1, Simon Moxon4, Maria Méndez-Lago1, Dandan Han1,5, Lars Schomacher1, Debasish Mukherjee1,6, and Christof Niehrs1,7,9

1 Institute of Molecular Biology (IMB), 55128 Mainz, Germany

2 Present address: Departments of Pediatrics and Cellular & Molecular Medicine, Pediatric Diabetes Research Center, University of California San Diego, La Jolla, California, 92093, USA

3 Present address: EMBL’s European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

4 School of Biological Sciences, University of East Anglia, Norwich, UK

5 Present address: STEMCELL Technologies Germany GmbH, 50933, Cologne, Germany

6 Present address: Cologne Cluster of Excellence: Cellular Stress Responses in Ageing-associated Diseases (CECAD), D-50931 Cologne, Germany

7 Division of Molecular Embryology, DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany

8 These authors contributed equally

9 Lead contact [c.niehrs@imb-mainz.de](mailto:c.niehrs@imb-mainz.de)

Short title: 5fC as epigenetic mark

**Highlights**

* 5fC forms chromocenters in the perinucleolar compartment during *Xenopus* ZGA
* 5fC colocalizes with Pol III and accumulates on active *tRNA* tandem repeats
* 5fC promotes Pol III recruitment and *tRNA* transcription at ZGA
* 5fC promotes transcription of a *tRNA* transgene

**Summary**

5-methylcytosine (5mC) is an established epigenetic mark in vertebrate genomic DNA but whether its oxidation intermediates formed during TET-mediated DNA demethylation possess an instructive role of their own that is also physiologically relevant, remains unresolved. Here we reveal a 5-formylcytosine (5fC) nuclear chromocenter, which transiently forms during zygotic genome activation (ZGA) in *Xenopus* and mouse embryos. We identify this chromocenter as the perinucleolar compartment, a structure associated with Pol III transcription. In *Xenopus* embryos, 5fC is highly enriched on Pol III target genes activated at ZGA, notably at oocyte-type tandem arrayed *tRNA* genes. By manipulating Tet and Tdg enzymes, we show that 5fC is required as regulatory mark to promote TfIIIc and Pol III recruitment, as well as *tRNA* expression. Concordantly, 5fC modification of a *tRNA-iMet*transgene enhances its expression *in vivo*. The results establish 5fC as activating epigenetic mark during zygotic reprogramming of Pol III gene expression.

Keywords: 5-formylcytosine, B-box, perinucleolar compartment, tRNA, Pol III, TDG, TET, tRNA-iMet, Xenopus, ZGA

**Introduction**

The discovery some 10 years ago of three oxidized derivatives of 5-methylcytosine (oxC-marks) occurring in genomic DNA of higher eukaryotes was a striking finding. However, understanding the physiological role of these oxC-marks remains challenging. A crucial unresolved question that we address here is whether they can function as epigenetic mark in their own right, rather than as demethylation intermediates.

The oxC-marks 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) are iterative oxidation products of 5-methylcytosine (5mC), a prominent epigenetic signal associated with gene silencing. It is well established that 5mC oxidation is mediated by TET dioxygenases and that one important role of oxC-marks is to serve as intermediates in the enzymatic removal of 5mC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; Ito et al., 2010; He et al., 2011; Ito et al., 2011; Kohli and Zhang, 2013; Hill et al., 2014; Shen et al., 2014). Thymine DNA glycosylase (TDG) completes demethylation by excising 5fC and 5caC via base excision repair (BER) to restore unmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011; Almeida et al., 2014).

However, oxC-marks might also function as epigenetic marks on their own to regulate gene expression (Cimmino and Aifantis, 2017). In line with this possibility, both 5hmC and 5fC are relatively stable modifications (Bachman et al., 2014; Bachman et al., 2015). Moreover, 5hmC is preferentially enriched at enhancers in mouse and human embryonic stem cells (Pastor et al., 2011; Szulwach et al., 2011; Wu et al., 2011; Xu et al., 2011; Raiber et al., 2012; Cimmino and Aifantis, 2017), whereas 5fC or 5caC accumulate at poised or active enhancers and promoters (Raiber et al., 2012; Song et al., 2013; Lu et al., 2015; Xia et al., 2015). In midstage mouse embryos, 5fC specifically accumulates on active enhancer regions suggesting a role in embryonic development (Iurlaro et al., 2016). In preimplantation stage mouse and human embryos, accumulation of 5fC precedes the induction of gene expression (Yun Gao et al.; Zhu et al., 2017; Gao et al., 2020) Thus, in mammalian embryos, 5fC displays features of an active mark that may promote gene expression. In contrast, in plasmid-based assays, 5fC modestly impairs Pol II mediated gene expression (Kellinger et al., 2012; You et al., 2014; Kitsera et al., 2017) and in the absence of BER, 5fC has almost no effect on reporter gene activity (Müller et al., 2021).

To affect gene expression, 5fC could recruit specific reader proteins, alter the physical properties of DNA, or modulate chromatin accessibility. Proteomic screens have identified proteins associated with 5fC (Iurlaro et al., 2013; Spruijt et al., 2013; Bai et al., 2021), but whether their binding is direct or indirect and physiologically relevant remains unresolved. 5fC can underwind the double helix and it was proposed that formylation of CpG repeats affects DNA supercoiling and packaging in chromatin, potentially affecting gene expression (Raiber et al., 2015). Moreover, while 5mC reduces nucleosome mechanical stability and DNA flexibility, 5fC enhances both features, leading to tighter binding of DNA to nucleosomes *in vitro* (Ngo et al., 2016). 5fC may also organize nucleosomes by forming a covalent Schiff-base with histones (Raiber et al., 2018). Yet, another report concluded that 5fC does not change the global structure of DNA (Hardwick et al., 2017). Furthermore, the proposition of enhanced nucleosome association with 5fC collides with the observation that active regulatory regions, where 5fC specifically accumulates, are generally nucleosome-depleted to ensure accessibility to regulatory proteins (Struhl and Segal, 2013). Concordantly, 5fC in ESCs accumulates in DNase I hypersensitive regions (Song et al., 2013; Xia et al., 2015). Thus, whether 5fC regulates DNA structure and attracts nucleosomes remains controversial.

In summary, while genome-wide mapping in embryos associates 5fC with gene expression activation, *in vitro* and cell-based experiments point to a repressive role. Thus, the key question remains unresolved: Is the presence of 5fC merely a byproduct of active demethylation, or is 5fC an instructive mark that regulates gene expression on its own? If so, does it activate or repress gene activity, and is it physiologically relevant? Conclusively answering these questions has proven formidable: I) 5fC is typically present only at very low levels in the genome and hence relevant genomic sites have to be identified where 5fC stably accumulates. II) When manipulating 5fC levels via TETs, one needs to rule out pleiotropic effects of TETs on other oxC-marks and exclude the non-enzymatic function of TETs as transcriptional repressors. Notably, a function of the 5fC as demethylation intermediate must be ruled out. Thus, merely correlating gene-regulatory effects with TET deficiency, even catalytic-specific, is inconclusive. III) The regulatory function should be physiologically relevant.

Here we report the discovery of a 5fC chromocenter, which forms transiently in nuclei of *Xenopus* embryos during zygotic genome activation (ZGA). ZGA is the unique phase during embryonic development when the transition from the maternal environment of the oocyte to embryonic gene expression occurs. ZGA is accompanied by major transcriptional changes, chromatin remodeling, and nuclear reorganization (Lee et al., 2014; Jukam et al., 2017; Eckersley-Maslin et al., 2018; Blitz and Cho, 2021). Importantly, the 5fC chromocenter forms at ZGA in the virtual absence of 5hmC and 5caC, affording the unique opportunity to study its role in isolation.

We identify the 5fC chromocenter as the perinucleolar compartment (PNC), a unique nuclear structure localized at the periphery of the nucleolus, which is associated with transcription by *RNA* polymerase III (Pol III). Concordantly, genomic profiling reveals massive enrichment of 5fC on Pol III target genes, notably tandem arrayed *tRNA* gene clusters. By manipulating Tet and Tdg enzymes, and by making use of a 5fC chemically modified *tRNA* transgene, we provide direct evidence that 5fC promotes Pol III transcription by acting as an instructive mark rather than as demethylation intermediate. 5fC-Pol III chromocenters also form in mouse zygotes during minor ZGA, supporting evolutionary conservation. These findings establish 5fC as active mark during reprogramming of Pol III gene expression in early vertebrate development.

**Results**

**5fC accumulates in a nuclear chromocenter at ZGA**

We monitored global levels of cytosine modifications during *X. laevis* development using liquid chromatography tandem mass spectrometry (LC-MS/MS) and employing isotopic standards as a highly sensitive method for detecting and quantifying oxC-marks (Schomacher et al., 2016) (Figure S1A). 5mC was rather uniform from ZGA onwards (st9). 5hmC was present at very low levels until organogenesis (st23), while 5caC was undetectable until neurulation (st13), upon which levels of both marks steadily increased, consistent with previous observations in *Xenopus* (Almeida et al., 2014; Bogdanović et al., 2016). Interestingly, unlike the other cytosine marks, 5fC showed a biphasic profile: Following induction at ZGA (st9), the signal levelled off at neurula (st13), to increase steadily from organogenesis onwards, paralleling the profile of 5hmC and 5caC. Zooming-in on ZGA with additional time points, we found that 5fC accumulates abruptly at early ZGA (st8 + 1hr) and is preceded by transient 5hmC accumulation pre-ZGA. Notably, 5fC levels reached ~0.001% of total C, which is up to 5 times higher than in embryonic stem cells (0.0006%) or neurons (0.00016%), thought to harbour top oxC levels (Liu et al., 2013; Schomacher et al., 2016). Moreover at st9, 5fC levels were ~5x higher than 5hmC, which displayed exceptionally low levels (~0.0002%), inverting the abundance-ratio of these oxCs in mammalian cells, where 5hmC is typically 10-100x more prevalent than 5fC (Wagner et al., 2015). This transient 5fC accumulation was intriguing, as it raised the possibility of 5fC functioning as regulatory mark during the critical phase of ZGA.

To confirm 5fC accumulation at ZGA, we employed antibodies specific for 5mC, 5hmC, 5fC and 5caC, to carry out immunofluorescence (IF) microscopy. The specificity of these antibodies was carefully validated. Dot blot control experiments with each antibody recognised only the specific modification it was raised against (Figure S1B). Amplicons modified with each oxC- or methyl mark outcompeted IF signals only of their matching target antibody in histological sections (Figure S1C). Potassium perruthenate (KRuO4) treatment, which oxidizes 5hmC to 5fC (Booth et al., 2012), specifically abolished 5hmC IF signals (Figure S1D). Conversely, sodium borohydride (NaBH4), which reduces 5fC to 5hmC (Booth et al., 2012), abolished 5fC IF signals (Figure S1E). We then carried out IF analysis for each of the oxC-marks, which confirmed the LC-MS/MS results: While 5mC and 5fC clearly stained nuclei at ZGA (st9), 5hmC and 5caC IF signals were close to background (Figure 1B-C), only appearing from organogenesis onwards (Figure S1C). 5fC-positive nuclei were specifically found in the animal hemisphere, unlike 5mC-positive nuclei that were equally distributed between the animal and vegetal hemispheres (Figure 1B). The animal hemisphere gives rise to most of the embryo proper, while the vegetal hemisphere develops into endoderm. 5fC-positive cells became more homogeneously distributed at gastrula and early neurula (st10-13; Figure S1F). No 5fC signals were observed before ZGA (st7, Figure 1D), confirming that 5fC specifically accumulates upon ZGA in *Xenopus*.

Close inspection revealed that 5fC nuclear staining at ZGA was not uniform, unlike 5mC staining, but instead showed marked enrichment in one or two foci, which were often crescent-shaped (Figure 1C-D). 5fC foci gave way to uniform staining during gastrula stages and disappeared by neurula (st13; Figures 1D; S1G). 5fC foci were also observed with an independent 5fC antibody (Figure S1H), while no foci were observed for 5hmC or 5caC.

To be visible as a chromocenter, 5fC must accumulate massively in this nuclear compartment. To distinguish whether the 5fC chromocenter represents a single genomic locus or is composed of loci on different chromosomes coalescing in interphase nuclei, we stained metaphase chromosomes. While 5mC distributed homogenously, 5fC signals distributed discretely across chromosomes during metaphase, with ~30% of chromosomes showing 5fC punctae (Figure 1E). An average of ~ 40 foci per metaphase spread was observed, mostly consisting of very small foci of below 0.1 µm diameter and few larger ones (> 0.15 µm) (Figure S1I). Thus, the 5fC chromocenter is composed of genomic loci residing on different chromosomes, which are coalescing in interphase nuclei to form a single compartment.

One such nuclear compartment of coalescing loci is the nucleolus, hence we carried out co-IF for 5fC and the nucleolar marker xNopp180 (Schmidt-Zachmann et al., 1984) in ZGA (st9) embryos (Figure 1F, top). Interestingly, the 5fC chromocenter located not in- but directly adjacent to the nucleolus, at the perinucleolar compartment (PNC). The PNC is a dynamic structure highly enriched in RNA-binding proteins and newly transcribed RNA, peripherally associated with the nucleolus (Huang et al., 1998). Importantly, it is known that Pol III target genes such as *tRNA*s, *5S rRNA*, *U6* *snRNA* that reside on different chromosomes, are coalescing in the PNC along with their transcripts (Matera et al., 1995; Thompson et al., 2003; Norton et al., 2009; Németh et al., 2010). Co-IF showed that Pol III, like 5fC, localizes adjacent to the nucleolus (Figure 1F, middle; 1G) and that Pol III colocalizes with the 5fC chromocenter (Figure 1F bottom; 1G-H). Collectively, the IF experiments indicate that the 5fC chromocenter corresponds to the Pol lll-rich PNC during ZGA.

**5fC marks Pol III bound *tRNA* genes during ZGA**

To characterize the genomic localization of 5fC, we carried out DNA immunoprecipitation-sequencing (DIP-seq) with 5mC, 5hmC, 5fC, and 5caC antibodies using genomic DNA from *X. tropicalis* embryos harvested at ZGA (st9). We confirmed antibody specificity of the IPs: Firstly, modified amplicons containing each methyl- and oxC mark were spiked into the genomic DNA before pulldown. Subsequent DIP-qPCR showed modification-specific amplicon precipitation (Figure S1J). Secondly, we carried out LC-MS/MS for each of the oxC-marks with input and pulldown samples and found the expected enrichments for modified cytosines (Figure S1K).

DIP-seq reads were aligned to the *X. tropicalis* genome without excluding multi-mapping reads to allow analysis of both unique and repetitive sequences (Dreszer et al., 2011; Papin et al., 2017). Globally, 5hmC correlated with 5caC, while 5fC and 5mC signatures were distinct (Figure 2A). Quantification of DIP-seq peaks (Suppl. Table 1) showed that 5fC peaks overlap mostly with 5mC, rather than 5hmC or 5caC (Figure S2A), consistent with previous observations (Shen et al., 2013). The genomic distribution of all cytosine modifications was similar between 5’UTR, 3’UTR, exon, intron, intergenic regions, promoter and downstream sequences of coding genes (Figure S2B). However, unlike 5hmC and 5caC peaks, 5mC and 5fC peaks were enriched in non-coding RNA genes (Figure S2C), particularly Pol III target genes (henceforth termed 'Pol III genes'), e.g. *5S rRNA, U6 snRNA,* and *tRNAs* (Figure 2B-C; S2D-E). Focusing on regions of high oxC mark density, we found that 5fC peaks selectively clustered over tens of kilobases at Pol III genes (Figure 2D; S2F; Suppl. Table 1). 5fC enrichment at Pol III genes, notably *tRNA* gene repeats, was also observed using only uniquely mapping reads, but at expectedly lower levels (Figure S2G).

To confirm 5fC accumulation at Pol III genes, we carried out DIP-qPCR in *X. tropicalis* embryos across different stages (Figure 2E; S2H). Before ZGA (st7), 5fC signals were undetectable. At ZGA (st9) 5fC robustly occupied Pol III targets (*tRNA-Ile, U6 snRNA, 5S rRNA*) (Figure 2E; S2H). 5fC signals at Pol III genes gradually decreased by neurula (st13) and were absent by organogenesis (st32). Pol II (*chordin*) and Pol I (*5.8S, 18S* and *28S rRNA*) genes were lowly occupied by 5fC (Figure 2E; S2H). Similar to 5fC, 5mC occupancy was induced at ZGA as reported (Bogdanovic et al., 2011). 5hmC and 5caC were at background levels at all tested loci. Thus, 5fC massively accumulated in ZGA and early gastrula by three orthogonal approaches: LC-MS/MS (Figure 1A), PNC chromocenter staining (Figure 1B-D), and 5fC DIP (Figure 2E).

To correlate 5fC-clusters with Pol III occupancy, we generated a Pol III (subunit RPC6) ChIP-seq data set for ZGA (st9) embryos. We found 1,419 Pol III sites, which included the majority of known Pol III genes, including 49 % of *tRNA* genes (Suppl. Table 2). Unlike 5hmC and 5caC, 5fC preferentially occupied long Pol III peaks (Figure S2I), notably at *tRNA* genes (Figure 2F). Pol III distribution across different *tRNA* genes was relatively homogeneous compared to 5fC occupancy, which was very heterogenous, with a predominance on *initiator* *tRNA* (*tRNA-iMet*) gene followed by *tRNA-Gly* and *tRNA-His* genes (Figure 2G). Interestingly, *tRNA-iMet* encodes the rate limiting tRNA for cell growth (Kolitz and Lorsch, 2010). Concordantly, 5fC peaks were also enriched on *tRNA* genes corresponding to ‘optimal’ codons (Figure 2H), which promote transcript stability and favor expression, while non-optimal codons promote RNA clearance (Bazzini et al., 2016; Mishima and Tomari, 2016). A small fraction of 5mC peaks overlapped with *tRNA* genes, while 5hmC and 5caC were almost absent.

We conclude that at ZGA, 5fC is uniquely enriched on Pol III occupied *tRNA* genes and correlates with tRNA features (*tRNA-iMet*, optimal codons) that promote protein biosynthesis.

**5fC marks active *tRNA* tandem repeat genes during ZGA**

A characteristic feature of the 5fC peaks on Pol III genes was their clustering along tens of kilobases (Figure 2D, S2E). These 5fC-marked clusters represent oocyte-type tandem-repeat *tRNA* genes, which are distinct from the low copy-number, somatic type homologs. Oocyte-type *tRNA*s exhibit two waves of expression (Andrews et al., 1991). The first wave occurs during oogenesis, followed by a second prominent ‘ZGA-wave’ which initiates at ZGA and ceases by early neurulation, when low copy number, somatic type *tRNA* genes continue to be transcribed.

Since 5fC accumulated at oocyte-type tandem repeat *tRNA* genes, we examined their location relative to chromatin marks. We centered the metaplots on 5fC and analyzed Pol III occupancy and previously established chromatin marks (Hontelez et al., 2015) (Figure 3A). At ZGA (st9), 5fC-associated *tRNA* genes were broadly occupied by Pol III, as well as by active chromatin marks (H3K4me1, p300). Since Pol III occupancy is a direct proxy of Pol III-dependent transcription and target gene expression (Barski et al., 2010; Moqtaderi et al., 2010; Oler et al., 2010)(Orioli et al., 2016), this Pol III profile suggests ongoing transcription of these tandem array *tRNA*s, as expected at ZGA. Chromatin profiles of 5fC and 5mC were similar (compare Figure 3A vs. Figure S3A) but a distinguishing feature of 5fC peaks was that a fraction of the repressive H3K9me3 marks were offset by ~1.5 kb from the 5fC peak center (Figure 3A). Note that in *Xenopus,* 5mC becomes a repressive mark only after gastrula stage (Bogdanovic et al., 2011).

Metagene profiles zooming in- and centering on *tRNA* genes confirmed that at ZGA, 5fC peaks coincided with high Pol III occupancy and showed a local sink of histone marks over the *tRNA* genes (Figure 3B). The 5fC profile peaked around the B-box (Figure S3B), a CpG-containing promoter element bound by transcription factor IIIC (TfIIIc) (Kirkland et al., 2013). TFIIIC is a multisubunit general transcription factor that recruits Pol III to promote *tRNA* and *5S* *rRNA* transcription (Graczyk et al., 2018).

As embryonic development proceeded, active marks progressively yielded to repressive marks (H3K9me3, H3K9me2, H4K20me3) (Figure 3C-D). For example by st16, p300 signals were almost lost, while H3K9me3 and H4K20me3 strongly increased (Figure 3D). We conclude that at ZGA, 5fC-rich *tRNA* genes in oocyte-type tandem repeats are transiently marked by active chromatin, which ceases by neurula stage, consistent with the reported Pol III gene ‘ZGA expression wave’ (Andrews et al., 1991). *U6 snRNA* gene repeats also featured pronounced 5fC peaks (Figure S3C) while Pol I genes were not enriched for any tested cytosine modifications (Figure S3D).

One prominent oocyte *tRNA* gene cluster in *X. laevis* expressed during the ZGA-wave (Clarkson and Kurer, 1976; Müller et al., 1987), has a 143 kb-sized homolog in *X. tropicalis.* The 143 kb cluster contains arrays of ~3 kb repeats, composed of two *tRNA-iMet* genes followed by single copies for five other *tRNA*s (*Tyr*, *Lys*, *Leu*, *Ala*, *Asn*), tandemly repeated 42 times (Figure 3E-G). At ZGA, this cluster featured mostly active marks (H3K4me1, H3K4me3, p300) (Figure 3E-F). While Pol III occupied each of the seven *tRNA* genes, a prominent 5fC peak located on the two *tRNA-iMet* genes (Figure 3F; line ‘L1’; Figure 3G). Major 5mC peaks were offset by 1.4 kb from the *tRNA-iMet* 5fC peaks, coincident with high H3K9me3 and low Pol III occupancy (Figure 3F; line ‘L2’; Figure 3G). A similar epigenetic profile was observed for a *tRNA-Gly* tandem array, where 5fC correlated with Pol III and p300 but anti-correlated with H3K9me2 and H3K9me3 (Figure S3E). 5hmC and 5caC signals were at background levels for both tandem arrays.

We analyzed base-resolution DNA methylation data for the oocyte *tRNA* gene cluster using a published resource (Bogdanović et al., 2016). There was a good match with our DIP-seq methylation profile (Figure S3F). All *tRNA* genes of this cluster were heavily methylated both at early and advanced embryonic stages. This observation echoes the ~400 mammalian tandemly arrayed *ribosomal RNA* gene repeats in the nucleolus, of which ~50% are constitutively hypermethylated even in metabolically highly active cells (Zillner et al., 2015).

Zooming into the *tRNA-iMet* and *tRNA-Gly* loci, all CpGs were methylated between 85-95% (Figure 3H, S3F). Next, we generated 5fC base-resolution data using methylation-assisted bisulfite amplicon sequencing for selected loci (MAB-seq, Neri et al., 2016). All genic and intergenic CpGs of *tRNA-iMet* and *tRNA-Gly* loci were formylated between 5-17%, including at the B-box, while *gapdh* was devoid of 5fC (Figure 3H, S3G). Note that these 5fC levels are underestimates since both the DNA methylation and formylation data derive from whole st9 blastula embryos, i.e. including chromocenter-negative vegetal halves, effectively diluting the 5fC signal by at least ~twofold. The *tRNA* 5fC levels align with those of mouse embryonic stem cells, known for high 5fC levels that average ~8-20% at statistically significant loci (Booth et al., 2014 , Neri et al., 2015).

We conclude that at ZGA i) 5fC occurs in the absence of other oxC modifications on *tRNA* oocyte-type tandem repeat genes, and ii) they are anti-correlated with the repressive mark H3K9me3 but positively correlated with Pol III occupancy.

**5fC promotes Pol III binding to *tRNA* genesat ZGA**

To investigate the role of 5fC on Pol III binding at ZGA, we manipulated *Tet2* and *Tet3* expression. *Tet2* and *Tet3* are expressed during early *Xenopus* development (Xu et al., 2012), while a *Tet1* orthologue is absent in *Xenopus* (Xu et al., 2012; Bogdanović et al., 2016). We carried out knockdown (KD) experiments with *Tet2* and *Tet3* in *X. tropicalis* antisense Morpholino (Mo) approaches. Of note, adequately controlled Mo are a broadly used research tool in model organisms such as *Xenopus* harboring large stores of maternal RNAs that escape CRISP/Cas9-mediated genome editing (Blum et al., 2015). The here employed *Tet3* Mo was previously characterized (Xu et al., 2012; Schomacher et al., 2016). For *Tet2* Mo we conducted rescue experiments employing Mo-resistant *Tet2* mRNA co-injection to validate specificity (Figures S4A-B). Knockdown of *Tet2* induced lethality during gastrulation in ~70% of embryos (Figure 4A-B). Knockdown of *Tet3* showed milder phenotypes, with surviving embryos displaying microcephaly, as reported previously (Xu et al., 2012; Schomacher et al., 2016). *Tet2/3* double KD led to 100% lethality during gastrulation (Figure 4A-B), as is the case in zebrafish (Bogdanović et al., 2016) and possibly mouse (Arand et al., 2022).

We conducted Pol III ChIP-qPCR at ZGA (st9), which showed that Pol III selectively occupies 5fC-modified *tRNA* genes (*Gly*, *iMet, Asp*) as opposed to non 5fC-modified *tRNA* genes (*eMet, Thr*, *Cys*) (Figure 4C). Double *Tet2/3* Mo KD, which reduced global 5fC levels (Figure S4C), led to reduced Pol III binding to 5fC-modified *tRNA* genes, while Pol II and Pol I target genes were unaffected (Figure 4C). Conversely, Tet2/3 *mRNA* overexpression (OE), induced global 5fC (Figure S4D) and promoted Pol III binding to 5fC-modified *tRNA-iMet* and *U6 snRNA* genes (Figure 4C). Low-level Pol III binding to *tRNA* genes displaying little 5fC (*eMet*, *Thr*, *Cys*) as well as to Pol II or Pol I target genes remained unchanged. We also analyzed the binding of TfIIIc (alpha subunit), the Pol III transcription factor that recognizes the B-box. Similar to Pol III, TfIIIc occupied 5fC-rich Pol III *tRNA* genes at ZGA (Figure 4D). This was specific to ZGA, as both Pol III and TfIIIc occupancy was more homogeneous across *tRNA* genes by late neurula (st18) (Figure S4E-F). Notably, as for Pol III, TfIIIc occupancy at 5fC-rich target genes was significantly reduced upon *Tet2/3* KD (Figure 4D).

The requirement of Tet2/3 for Pol lll and TfIIIc binding at ZGA could be due to different reasons. First, TETs can act as direct transcriptional repressors (Williams et al., 2011) and hence they could be essential for Pol lll recruitment and gene expression at ZGA in a manner unrelated to their catalytic role (e.g. repression of a Pol III repressor). Second, the Tet2/3 requirement for Pol lll recruitment may reflect an essential role of 5fC at Pol lll targets, either as DNA demethylation intermediate or as instructive regulatory mark. To distinguish between these possibilities we manipulated Tdg in *Xenopus*, the enzyme that is essential for removing 5fC during active DNA demethylation (He et al., 2011; Maiti and Drohat, 2011). In *Xenopus*, *Tdg* expression starts only during gastrulation and there are no maternal transcripts ([www.xenbase.org](http://www.xenbase.org)) or protein (Slenn et al., 2014). Accordingly, *Tdg* KD using an established antisense Mo (Schomacher et al., 2016), led to pronounced accumulation of 5fC and 5caC in tailbud stage but not ZGA stage embryos (Figure S4G). Unlike *Tet2/3* morphants, *Tdg* morphants passed through ZGA, gastrulation, and neurulation, and displayed axial defects only at tailbud stages (Figure S4H), similar to mouse *Tdg* null embryos, which die at midgestation (Cortázar et al., 2011).

To specifically eliminate 5fC, we forced 5fC excision by microinjection of human TDG mRNA, which reduced 5fC levels to half (Figure S4I) and induced embryonic lethality and gastrulation defects, observed as a cellular mass extruding from the blastopore in late gastrula (st12) (Figure 4E-F). TDG OE also delayed gastrulation as indicated by retarded blastopore closure (Figure 4E, G). OE of the catalytically inactive TDGN140A (dTDG)(Hardeland et al., 2000) (Figure S4J), did not interfere with gastrulation (Figure 4E-G), despite higher expression (Figure S4K) and merely reduced pigmentation in tadpoles (Figure S4L).

We then tested the effect of TDG OE on Pol III binding by ChIP-qPCR. TDG OE reduced Pol III binding by 50% at 5fC-modified *tRNA* genes, while dTDG had no effect (Figure 4H). TDG OE did not affect low level Pol III binding to Pol II and Pol I target genes (Figure 4H) nor Pol II binding targets (Figure S4M). This key result supports that it is the presence of 5fC at target loci rather than their demethylation, which is essential for Pol III engagement at ZGA.

**5fC promotes *tRNA* transcription during ZGA**

Since Pol III occupancy is a direct proxy of ongoing transcription and was reduced following 5fC loss, we next probed directly whether 5fC promotes *tRNA* transcription at ZGA. We used Northern blot analysis to quantify *X. tropicalis* *tRNA* transcripts since endogenous *tRNA*s are heavily chemically modified, which renders their qPCR and RNA-seq analysis challenging. Further, to discriminate between maternally provided vs. newly transcribed *tRNA*, we employed α-amanitin to acutely inhibit Pol III at ZGA (Chen et al., 2019). Treatment of st9 embryos with α-amanitin completely removed *U6* *snRNA* transcripts (Figure 5A), indicating low levels of maternal *U6* *snRNA*. By contrast, α-amanitin reduced levels of *tRNA-iMet* and *tRNA-Gly* by only ~25-30% (Figure 5B and S5A), indicating large maternal *tRNA* stores and gauging the fraction of newly transcribed *tRNA*. Next, to reduce 5fC, we either injected embryos with *Tet2/3* Mo to inhibit 5fC synthesis or with TDG mRNA to force 5fC excision. *Tet2/3* Mo reduced *tRNA-iMet* and *tRNA-Gly* transcript levels at st9 by 20%, a reduction comparable to that afforded by α-amanitin (Figure 5B and S5A). TDG OE reduced both *tRNA* transcript levels to a similar extent, albeit somewhat later than *Tet2/3* KD, at early gastrula (st10) (Figure 5B and S5A), possibly because Tet KD acts more acutely by preventing 5fC accumulation in the first place. Unlike TDG, dTDG did not reduce *tRNA-iMet* or *tRNA-Gly* transcripts (Figure S5B), supporting that is the catalytic activity of TDG that impairs *tRNA* transcription. Neither *Tet2/3* Mo nor TDG OE reduced levels of other Pol III target transcripts (*U6* *snRNA*, *5S* *rRNA*; Figure S5C-D), supporting that 5fC specifically promotes *tRNA* gene transcription.

To analyze the role of Tetproteins on global gene expression, we analyzed the transcriptomes of ZGA stage (st9) embryos injected with *Tet2/3* Mo. In *Tet2/3* Morphants, expression of 2,876 genes changed significantly (5% FDR), with about equal number up- and downregulated genes (Figure 5C, Suppl. Table 3). Gene Ontology (GO) enrichment analysis of the upregulated genes revealed terms related to protein biosynthesis (e.g. ‘mRNA splicing’, ‘translation’) with high significance. Downregulated genes showed one enriched term related to protein degradation (Figure 5D), and concordantly the top downregulated gene *psmb7* encodes a proteasomal subunit. This gene expression signature is consistent with Tet2/3 deficient embryos upregulating protein biosynthesis by increasing translation while decreasing protein degradation, possibly a compensatory reaction to tRNA deficiency.

Finally, we directly tested the ability of 5fC to promote *tRNA* transcription using a strand-exchange (SE) transgene approach (Lühnsdorf et al., 2012). In brief, we subcloned *tRNA-iMet* into a plasmid and using nicking endonucleases excised the non-coding (template) *tRNA-iMet* strand (92 nt) and replaced it with a synthetic oligonucleotide carrying a desired cytosine modification at two sites (Figure 5E). The two modified cytosines were placed inside and directly downstream of the B-box, as B-box formylation was observed *in vivo*. We then microinjected one-cell stage *X. laevis* embryos with plasmids that either did not undergo strand exchange (‘No SE’) or SE-plasmids carrying unmodified (C), methylated (5mC), or formylated (5fC) cytosine. Of note, chemical synthesis of 5hmC-modified B-box oligonucleotides failed repeatedly and hence 5hmC modification could not be analyzed. Embryos were harvested at different stages for quantification of *tRNA-iMet* levels by qPCR. Both *tRNA-iMet* controls (‘No SE’ and unmodified ‘C’) showed expression at early gastrula (st10) but neither at ZGA onset (st8) nor at mid-gastrula (st11) (Figure 5F), consistent with wave-like transcription of oocyte-type *tRNA* genes specifically around ZGA. 5mC-modified *tRNA-iMet* showed the same temporal profile and comparable expression level to the unmodified transgenes, consistent with the report that DNA methylation is not transcriptionally repressive in early *Xenopus* embryos (Bogdanovic et al., 2011). In contrast, the 5fC-modified transgene showed 6-fold increased *tRNA-iMet* levels compared to controls, and this induction was also restricted to early gastrula. While the average expression increase was robust and statistically highly significant, it was variable between embryo batches. We reasoned that the variability may be due to the site of microinjection. Indeed, we observed an average 20-fold 5fC-induced *tRNA-iMet* expression when plasmid microinjection aimed centrally at the animal pole, while vegetal pole microinjection yielded only negligible expression (Figure 5G). This dramatic animal-vegetal difference echoes our observation that 5fC chromocenter nuclei occur only in cells located in the animal- but not the vegetal hemisphere (Figure 1B) and the reported animal-vegetal gradient of ZGA gene expression in *Xenopus* embryos (Chen et al., 2019).

To further corroborate the specificity of 5fC-induced *tRNA-iMet* expression, we pre-treated SE plasmids with recombinant TDG to remove 5fC. LC-MS/MS verified 90% 5fC removal upon TDG treatment (Figure S5E-F). Consistently, TDG treatment reduced *tRNA-iMet* expression to control levels (Figure 5H). Taken together, the observation that two formylated cytosines in a hemi-modified B-box suffice to induce *tRNA* gene expression *in vivo*, provides direct, unequivocal evidence that 5fC acts as an activating mark for Pol III at ZGA.

**5fC enrichment on *tRNA* genes is conserved in mouse**

To investigate whether the occurrence of transient Pol III 5fC chromocenters is evolutionary conserved, we carried out IF analysis on mouse pre-implantation embryos. We validated 5fC antibody specificity in *TET1,2,3* triple mutant mouse embryonic stem cells (Dawlaty et al., 2014); Figure S6). The paternal mouse pronucleus (PN) is known to specifically stain positive for 5fC, consistent with its undergoing genome-wide DNA demethylation (Inoue et al., 2011; Zeng et al., 2019). We confirmed robust 5fC immunostaining of the paternal PN and weak staining of the maternal PN (Figure 6A). Close inspection revealed that the paternal 5fC staining pattern was not uniform but occurred in puncta in late stage (PN4/5) zygotes, when the minor ZGA occurs (Bouniol et al., 1995; Aoki et al., 1997; Abe et al., 2018). Importantly, the majority of the 5fC chromocenters in PN4/5 zygotes also stained positive for Pol III (Figure 6A, C). 5fC chromocenters were transient and disappeared by the 2-cell stage (major ZGA) giving way to homogenous signal, similar to Pol III staining (Figure 6A, D).

To confirm that 5fC chromocenters and their Pol III colocalization are a feature unique to this DNA modification, we analyzed 5mC and all three oxC-marks in PN4/5 zygotes (Figure 6B). 5mC was only detectable in the maternal PN, consistent with the genome-wide loss of DNA methylation occurring in the paternal PN upon fertilization (Mayer et al., 2000). While 5hmC signals accumulated in both pronuclei, 5caC staining was mainly detectable in the paternal PN. However, except for 5fC, none of the other cytosine marks showed punctae colocalizing with Pol III. We conclude that as in *Xenopus*, 5fC forms transient chromocenters during minor ZGA, which specifically colocalize with Pol III.

**Discussion**

The two main conclusions of this study are that i) transient 5fC chromocenters build up during *Xenopus* ZGA and ii) the 5fC modification functions as an activating mark on *tRNA* genes to promote their expression during *Xenopus* zygotic reprogramming. The observation of 5fC-Pol III chromocenters also in mouse zygotes suggests evolutionary conservation. Thus, our study answers the longstanding question in the DNA demethylation field whether 5fC has a biologically relevant role of its own by demonstrating that it acts as an instructive, gene-activating mark in *Xenopus*.

**5fC is an activating mark in ZGA *tRNA* expression**

*Xenopus* embryos offered a unique opportunity to analyze the regulatory role of 5fC: This mark accumulates at ZGA to exceptional levels in the PNC, at a stage when 5mC-mediated gene silencing is still dormant (Bogdanovic et al., 2011), Tdg is absent (Slenn et al., 2014), and where the analysis of Pol III gene regulation is greatly sensitized by the occurrence of amplified tandem *tRNA* clusters. The PNC occurs in diverse cell types, from yeast to human, where it is prevalent in cancer cells (reviewed in (Pollock and Huang, 2010). It is the site of Pol III gene transcription and is enriched in Pol III-transcribed RNAs (Matera et al., 1995; Thompson et al., 2003; Norton et al., 2009; Németh et al., 2010). In *Xenopus* oocytes, nuclear bodies containing Pol III target loci were described to be attached to lampbrush chromosomes (Nizami and Gall, 2012). The PNC is a dynamic, probably phase-condensed compartment, like the nucleolus with which it associates. For example, the PNC depends on Pol III transcription and Pol III inhibitors cause its rapid disassembly (Wang et al., 2003). The PNC in *Xenopus* embryos stains positive for Pol III and 5fC. Concordantly, genomic profiling revealed 5fC accumulation at oocyte-type Pol III tandem array genes, notably *tRNAs*.

Several lines of evidence indicate that 5fC functions as epigenetic mark promoting Pol III transcription at these loci in *Xenopus*: 1) 5fC becomes detectable at ZGA and ceases by late gastrulation, coinciding with the ZGA-wave of transcription of oocyte-type Pol III tandem array genes (Andrews et al., 1991). 2) 5fC peaks are enriched on *tRNA* genes with ‘optimal’ codons, which promote transcript stability. 3) *tRNA-iMet*, the rate limiting *tRNA* for cell growth (Kolitz and Lorsch, 2010), shows the highest 5fC occupancy. 4) within *tRNA* loci, 5fC peaks are anti-correlated with the repressive mark H3K9me3 and positively correlated with Pol III occupancy. 5) impairing 5fC synthesis by *Tet2/3* KD reduces Pol III- and TfIIIc binding to target genes and *tRNA* expression. 6) lowering 5fC levels by TDG OE impairs Pol III binding to *tRNA* genes and inhibits their expression instead of activating them, supporting that 5fC does not act as demethylation intermediate. Indeed, DNA demethylation may not even be necessary in the first place because early *Xenopus* embryos lack the gene-repressive effect of 5mC (Bogdanovic et al., 2011). 7) As key direct evidence, we show that chemical 5fC modification stimulates expression of *tRNA-iMet* in a manner that is both region- and stage-specific in the embryo.

We note that there is a curious, threefold coincidence of the formyl moiety associated with iMet. First, we find that DNA coding for *tRNA-iMet* shows the highest 5fC occupancy. Second, 5fC is also a unique RNA modification of mitochondrial *tRNA-iMet* that promotes mitochondrial translation (van Haute et al., 2017). Third, N-formyl-methionine is the modified translation-initiating amino acid in bacteria and in eukaryotic organelles, such as mitochondria and chloroplasts (Sherman et al., 1985). Thus, the addition of a formyl moiety to DNA, RNA, and protein is associated with protein biosynthesis, suggesting co-evolution relating to methionine- and one-carbon metabolism that regulates all three biopolymers.

How are 5fC and hence TET enzymes targeted to Pol III genes? High 5fC density at *tRNA* clusters implies that its precursor 5mC must also cluster at these loci, which is what we observe (Figure 3, S3F). DNA methylation starts sharply during ZGA (Bogdanovic et al., 2011) and this work), thus it is conceivable that Tets are attracted to high 5mC density at tandem array *tRNA* genes set by DNA methyltransferases. However, high 5mC density alone cannot explain the animal-vegetal 5fC differences nor preferential formylation of certain *tRNA* genes (*iMet*, *Gly*) and hence there must be additional determinants.

How is the transient deposition of 5fC on Pol III targets regulated? At ZGA, 5fC accumulates while 5hmC remains low and in the near absence of 5caC. The genomic signature of 5fC correlates with that of 5mC. The accumulation of 5fC is highly unusual, since typically 5hmC is more prevalent than 5fC and 5caC and TET enzymes rather tend to stall at 5hmC marks (Ito et al., 2011; Hu et al., 2015). While e.g. SALL4A was shown to bind 5hmC and facilitate its oxidation (Xiong et al., 2016), it would remain unclear in such a scenario what prevents Tet enzymes to process 5fC further to 5caC. Furthermore, for 5fC to accumulate, it must be protected from Tdg action, which is consistent with *Xenopus* Tdg appearing only after ZGA (st9) (Slenn et al., 2014). Similarly in mouse, *Tdg* mRNA is not detected in 2-cell stage embryos (Huang et al., 2017) and *Tdg* deficiency does not affect zygotic DNA demethylation (Santos et al., 2013; Guo et al., 2014), suggesting that absence of TDG protein before ZGA is conserved. Thus, transient 5fC accumulation at Pol III sites at ZGA (st9) and early gastrula (st10) implies that Tet enzymes act in a highly regulated fashion, neither stalling at 5hmC, nor proceeding to 5caC, before Tdg is expressed and removes the mark. Taken together, these results support a model (Figure 6E) where during zygotic genome reprogramming full activation of Pol III target genes, notably tandem array *tRNA* genes, requires 5fC, which serves as an activating epigenetic mark that by an unknown mechanism promotes Pol III and TfIIIc recruitment. At the end of gastrulation, oocyte-type Pol III targets are silenced as DNA methylation marks become repressive, 5fC is lost upon *tdg* expression, and repressive histone marks accumulate.

What may be the mechanism whereby 5fC promotes Pol III binding? 5fC may be recognized by a reader protein, e.g. a transcription factor or it may repel an inhibitor of Pol III transcription. However, the broad 5fC distribution across gene body and intergenic regions as well as the moderate 5fC levels at individual CpGs argue against a very position-specific reader effect. 5fC may regulate nucleosome positioning (Raiber et al., 2018) but 5fC peaks were rather depleted of histones, consistent with the finding that gene bodies of active *tRNA* genes are nucleosome-free (Helbo et al., 2017). On the other hand, even a single copy of 5fC within 90 bp of DNA increases DNA flexibility and bendability (Ngo et al., 2016), which may assist transcription factor recognition. Hence, the DNA structure, rather than the formylated cytosine *per se*, may facilitate TfIIIc/Pol III engagement without a 5fC reader.

**The role of 5fC in mouse zygotes**

As in *Xenopus* ZGA, during mouse minor ZGA, 5fC foci colocalize with Pol III. The observation of 5fC-Pol III chromocenters in mouse zygotes support that an activating function of 5fC on tRNA expression may be evolutionary conserved. These observations are suggestive for a conserved function of 5fC in Pol III gene regulation in mouse ZGA. However, unlike TET-deficient *Xenopus* and zebrafish embryos (Bogdanović et al., 2016), *Tet* DKO and TKO mouse embryos can develop several days past ZGA and die only during gastrulation, albeit with altered expression of a few hundred genes at preimplantation and blastocyst stages (Kang et al., 2015; Dai et al., 2016; Cheng et al., 2022). Intriguingly, acute TET KD revealed that Tet1/3 deficient embryos arrest at the 2-cell stage, suggesting a requirement for completion of ZGA that may be compensated in the reported knockout mice (Arand et al., 2022). Since most studies on *Tet* genes have focussed on Pol II targets, it will be interesting to investigate the role of *Tet* genes more systematically on Pol III targets, which are so important for cell growth in development and disease, including cancer (Marshall and White; Marshall and White; Yeganeh and Hernandez, 2020).

**Limitations of this study**

A 5fC reader on Pol III targets remains unknown. Furthermore, epigenetic marks can carry regulatory memory across cell division, which has not been addressed here. Technical limitations of the study are that *Xenopus* 5fC sequencing has not been conducted at base-resolution beyond select *tRNA* genes, although this does not affect the main conclusions. The functional relevance of the 5fC chromocenter in mouse, albeit highly suggestive, remains to be experimentally confirmed. Our study provides the framework to pursue these and other emerging questions.

**STAR Methods**

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Christof Niehrs (C.Niehrs@imb-mainz.de).

**Materials availability**

Plasmids generated in this study will be shared by the lead contact upon request.

**Data and code availability**

* NGS data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper also analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table.
* This paper does not report original code.
* Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Ethics Statement**

*X. laevis* and *X. tropicalis* experiments (Landesuntersuchungsamt, reference number 23177–07/A17-5-002 HP), as well as mouse husbandry at the Institute for Molecular Biology (IMB; Haltungsgenehmigung, reference number 41a/177-5865-§11 IMB) and the Translational Animal Research Center (TARC; Haltungsgenehmigung, reference number 41a/177-5865-§11 ZVTE) were approved by the state review board of Rheinland Pfalz, Germany and performed according to federal and institutional guidelines.

**Mouse Embryonic Stem Cells (mESCs) culture**

mESCs clones WT 4 and TET TKO 26 (Dawlaty et al., 2014) were cultured on tissue culture plates coated with 0.1% Gelatin (Sigma) in 2i medium (Neurobasal - DMEM/F-12 medium (Gibco), 1x N2- supplement (Gibco), 1x B27 (Gibco), 2 mM L-Glutamine (Gibco), 1000 U/ml Leukemia inhibitory factor (LIF, Millipore), 100 U/ml Pen-Strep (Gibco), 1 μM PD0325901 (Sigma), 3 μM CHIR99021 (Sigma), 50 μg/ml BSA) at 37°C in 5% CO2 and 20% O2.

**METHOD DETAILS**

***Xenopus* embryo fixation and sectioning**

*X. laevis* embryos were fixed in MEMFA (100 mM MOPS pH 7.4, 2 mM EGTA, 1 mM MgSO4, 4% formaldehyde) for 2 hrs at RT and embedded in optimal cutting temperature (OCT) media for cryosectioning. Sections were cut to 12 µm thickness using a Leica CS3050 S cryotome.

***Xenopus* embryo immunofluorescence staining**

Tissue sections were permeabilised with 0.5% Triton X-100 in PBS for 1 hr at RT. For oxC staining, tissue sections were incubated in 4 N HCl for 15 min and neutralised in 100 mM Tris-HCl pH 8.5 for 5 min. Sections were then blocked in PBT (2 mg/ml BSA (Sigma), 0.1% Triton X-100, PBS) for 1 hr at RT, incubated with primary antibody in PBT overnight at 4oC and washed 5x in PBT for 20 min. 5fC and 5caC primary antibodies were diluted in Pierce Immunostain Enhancer (Thermo Scientific) instead of PBT for signal enhancement. For double immunohistochemistry, primary antibodies were added sequentially. Sections were incubated with secondary antibody in PBT overnight at 4oC, then washed like previously. When both antibodies used were raised in the same species the first primary/secondary was blocked using 5% rabbit non-immune serum (31883, Thermo Scientific) for 1 hr and then incubated with goat anti-rabbit H+L Fab fragments (SAB3700970, Sigma) 1:20 for 1 hr then fixed for 10 min in MEMFA before adding the second primary antibody. DNA was stained with Yoyo1 (Life Technologies) overnight at 4oC, then washed like previously. Sections were mounted with Dako mounting medium (Agilent Technologies).

Antibody dilutions used as follow: 5mC (MAb-5MECYT-100, Diagnode) 1:5000, 5hmC (39791, Active Motif) 1:1000, 5fC (61227, Active Motif) 1:100, 5fC (74178, Cell Signaling) 1:25, 5caC (61229, Active Motif) 1:100, POLR3F (ab151495, Abcam) 1:100 and xNopp180 (a kind gift from Marion S. Schmidt-Zachmann) 1:100. Secondary antibodies used were Alexa Fluor® 546 goat anti-mouse (A-11030, Thermo Scientific), Alexa Fluor® 647 goat anti-rabbit (A-21244, Thermo Scientific) and Alexa Flour® 546 goat anti-rabbit (A-11010, Thermo Scientific). Imaging was carried out using a Leica TCS SPE and SP5 confocal microscope.

To validate anti-oxC antibody specificity, control experiments were carried out. 2 µM of 5mC, 5hmC, 5fC or control PCR amplicons or 5caC dNTPs were denatured at 95°C for 10 min, cooled on ice and added to 100 µl antibody mixture for epitope competition experiments. This mixture was then used for immunohistochemistry on st38 *X. laevis* embryos to outcompete antibody binding. Slides containing st38 *X. laevis* sections were treated with 1 mM potassium perruthenate (KRuO4) or 40 µM sodium borohydride (NaBH4, Sigma). This was followed by 5hmC and 5fC immunohistochemistry.

**Metaphase chromosome spreads**

The protocol of (Kieserman and Heald, 2011) was used with modifications: The vitelline membrane and the vegetal pole of st9 *X. laevis* embryos were removed using forceps. The animal pole was incubated in deionized water for 20 min to dissociate the cells and fixed in 200 µl 60% acetic acid for 5 min at RT. Cells were pelleted at 500 g and resuspended in 20 µl 60% acetic acid. The cell suspension was then dropped onto glass slides from a 1 m height to burst the cells. Spreads were fixed using MEMFA (100 mM MOPS pH 7.4, 2 mM EGTA, 1 mM MgSO4, 4% formaldehyde) for 10 min at RT, followed by immunofluorescence staining.

**5mC, 5hmC, 5fC and 5caC DNA Immunoprecipitation (DIP)**

Genomic DNA from st9 *X. tropicalis* embryos was extracted using the DNeasy Blood and Tissue kit (Qiagen). 15 µg DNA was sheared to ~200 bp using a Covaris S2/E210 focused ultrasonicator. 5 pg of 5mC, 5hmC, 5fC or control unmodified amplicon (see Suppl. Table 4) were spiked in and the samples were brought to 450 μl with nuclease-free water. 5% sample was kept aside as input. DNA was denatured for 10 min at 95°C and cooled immediately on ice. Samples were incubated in IP buffer (10 mM Na-Phosphate pH 7, 140 mM NaCl, 0.05% Triton X-100) and 3 µg antibody overnight at 4oC with overhead shaking. Antibodies used: 5mC (MAb-5MECYT-100, Diagnode), 5hmC (39791, Active Motif), 5fC (61227, Active Motif), 5caC (61229, Active Motif) and IgG (ab46540, Abcam). Dynabeads Protein G (10004D, Thermo Scientific) were pre-washed with 0.1% BSA, PBS and resuspended in IP buffer. 40 μl Dynabeads per sample were added and the samples were incubated at 4°C for 2 hrs with overhead shaking. Dynabeads were collected in a magnetic rack, washed 3x in IP buffer, resuspended in Proteinase K digestion buffer (50 mM Tris pH 8, 10 mM EDTA, 0.5% SDS) and Proteinase K (0.3 mg/ml final concentration) was added. Samples were incubated for 3 hrs at 50°C and DNA was purified with phenol/chloroform followed by ethanol precipitation.

**Genome-wide 5mC, 5hmC, 5fC, 5caC DIP-sequencing**

DIP-sequencing libraries were prepared for single-end samples in triplicates. Library prep was performed with Swift Bioschiences's Accel-NGS 1S Plus DNA Library Kit following Accel-NGS 1S Plus DNA Library Kit Reference Guide (2015) (15-0300, 08/15). Libraries were amplified in 7 PCR cycles for high concentration samples and 10 PCR cycles for low concentration samples, profiled in a High Sensitivity DNA chip on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). All samples were pooled in equimolar ratio and sequenced on 4 HiSeq 2500 Rapid Flowcells, SR mode with 59 cycles plus 2x8 cycles for dual index reads. In total 35-68 million reads were generated for each sample. Sequenced reads were first quality checked using FastQC ([*https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)) and then aligned over *X. tropicalis* genome assembly v.7.1 ([*https://ftp.xenbase.org/pub/Genomics/JGI/Xentr7.1/*](https://ftp.xenbase.org/pub/Genomics/JGI/Xentr7.1/)) using weighted multi-mapping strategy employing bowtie2 v.2.2.7 ([*http://bowtie-bio.sourceforge.net/bowtie2*](http://bowtie-bio.sourceforge.net/bowtie2)) with default parameters (Papin et al., 2017).

**Pol III, Pol II and TfIIIc chromatin immunoprecipitation (ChIP)**

The ChIP protocol employed was previously described (Hontelez et al., 2015). In short, 1,000 *X. tropicalis* embryos were cross-linked in 1% methanol-free formaldehyde for 30 min at RT. Cross-linking was quenched with 125 mM glycine for 30 min at RT and the embryos were washed twice with 0.3X Modified Barth’s Saline (88 mM NaCl, 1 mM CaCl2, 1 mP MgSO4, 5 mM HEPES pH 7.8, 2.5 mM NaHCO3). Embryos were homogenized in 200 µl sonication buffer (20 mM Tris–HCl pH 8, 70 mM KCl, 1 mM EDTA, 10% glycerol, 5 mM DTT, 0.125% Nonidet P40 and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)). Chromatin was sheared to ~500bp using a Bioruptor Plus (Diagenode) and centrifuged for 5 min at 16,000 g to remove yolk and pigment. 10% chromatin was kept aside as input control and 100 µl of chromatin was added to 100 µl incubation buffer (50 mM Tris–HCl pH 8, 100 mM NaCl, 2 mM EDTA, 1 mM DTT, 1% Nonidet P40 and cOmplete EDTA-free Protease Inhibitor Cocktail). 5 µg of POLR3F (ab151495, Abcam), Pol II (C15200004, Diagenode), TFIIIC alpha subunit (NBP2-14077, Novus Biologicals) or IgG (ab171870, Abcam) antibody was added and incubated overnight at 4°C with overhead shaking. 50 µl Dynabeads Protein G (10004D, Thermo Scientific) per reaction were blocked with incubation buffer + 0.1% BSA for 10 min at 4oC with overhead shaking. Beads were washed twice in incubation buffer for 5 min at 4oC, chromatin was added and the samples were incubated overnight at 4°C with overhead shaking. Beads were washed sequentially with wash buffer 1 (50 mM Tris–HCl pH 8, 100 mM NaCl, 2 mM EDTA, 1 mM DTT, 1% Nonidet P40, 0.1% deoxycholate, and cOmplete EDTA-free Protease Inhibitor Cocktail), wash buffer 2 (wash buffer 1 +400 mM NaCl), wash buffer 3 (wash buffer 1 + 250 mM LiCl), wash buffer 1 and Tris-EDTA buffer. 200 µl elution buffer (100 mM NaHCO3 pH 8.8, 1% SDS) was added to the beads and incubated for 15 min at RT with overhead shaking. Elution was repeated with additional 200 μl elution buffer. 16 µl 5 M NaCl was added to each elution and input control. Cross-links were reversed by incubating at 65°C overnight with shaking.

**Pol III ChIP-sequencing**

Pol III ChIP-sequencing was performed in triplicates for Pol III and Input samples or in duplicates for IgG samples. ChIP-seq library preparation was performed using NuGEN´s Ovation Ultralow System V2 1-16 (2014). Libraries were prepared with a starting amount of 1 ng of DNA and were amplified in 12 PCR cycles. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). All samples were pooled in equimolar ratio and sequenced on 1 NextSeq 500 Highoutput Flowcell, PE mode for 2x 42 cycles plus 7 cycles for the index read. Sequencing generated a total of 35-58 million reads. For mapping and peak calling, we used a similar strategy as used for the DIP-seq analysis.

***Xenopus* DIP and ChIP-seq data integration and visualization**

We divided the genome into 1000 bp bins and then calculated coverage using deepTools *bamCoverage* ([*https://deeptools.readthedocs.io/*](https://deeptools.readthedocs.io/)). Then 5mC, 5hmC, 5fC, and 5caC replicates were correlated with each other using Pearson correlation and heatmaps were plotted using heatmap.2 function of R-tools package ggplot2 ([*https://cran.r-project.org/web/packages/ggplot2/*](https://cran.r-project.org/web/packages/ggplot2/)). We used BEDTools to count the overlapping DIP-peaks from all four modification marks to plot the Venn diagram and to calculate the overlap of DIP-peaks with various genomic features. We downloaded the RNA pol II gene annotation file as “Xentr7\_2\_Stable.gff3” from Xenbase ([*http://www.xenbase.org*](http://www.xenbase.org)) and subsequently extracted the exon, intron, 5UTR, 3UTR, promoter, and gene downstream (3 kb) regions using R-tools package GenomicFeatures ([*https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html*](https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html)). Those regions that were not covered by the RNA pol II genes annotation, were considered as intergenic. Furthermore, we used a *de novo* ncRNA annotation (see below) to estimate the enrichment of DIP-peaks over ncRNAs. We estimated the percentage of DIP-peaks correlating with Pol I genes (*5.8S, 18S and 28S rRNA*), Pol II genes (annotated in Xenbase), and Pol III genes (*tRNA, U6, 5S rRNA*)*.* For identification of DIP clusters, we used BEDTools *mergeBed* with the parameters of minimum 4 peaks and maximum 2.5 kb distance from each other. BoxplotR ([*http://shiny.chemgrid.org/boxplotr/*](http://shiny.chemgrid.org/boxplotr/)) tool was used to create violin plots. Moreover, we estimated percentage of DIP-peaks over different categories of *tRNAs* representing optimal or non-optimal codons (Bazzini et al., 2016) using χ2 test. For motif analysis, we used 5fC overlapping Pol III genes using HOMER findMotif ([*http://homer.ucsd.edu/homer/motif/*](http://homer.ucsd.edu/homer/motif/)) tool with manual inspection of consensus pattern.

For metaplot analysis, we used published ChIP-seq datasets (<https://bigftp.xenbase.org/pub/sequence_information/RIMLS-SVH/ChIP-seq_Feb2015/xt7_1/>) for st9, 10.5, 16 (Hontelez et al., 2015). We calculated normalized read density using deepTools *computeMatrix* ([*https://deeptools.readthedocs.io/*](https://deeptools.readthedocs.io/)) and heatmaps were plotted using *plotHeatmap*. Visualization of bigWig tracks from DIP-seq, ChIP-seq and published datasets was performed on the UCSC genome browser ([*https://genome.ucsc.edu/*](https://genome.ucsc.edu/)) using customized genome profile of *X. tropicalis* assembly v.7.1. The bigWig browser tracks are generated using deepTools *bamCoverage* for DIP-Seq datasets using parameters “*--normalizeUsing RPKM --extendReads 250”* (as per average fragment size) and for Pol III ChIP-Seq (paired-end datasets) using parameters “*--normalizeUsing RPKM –extendReads”*.

***ncRNA* annotation**

Non-coding RNAs were annotated using Infernal v1.1.2 (Nawrocki and Eddy, 2013) based on Rfam 12.1 models (Nawrocki et al., 2015) using default gathering thresholds (Suppl. Table 5). *miRNA* families were excluded from the Rfam search due to incomplete coverage and results were converted to bed format. *X. tropicalis* *miRNA* hairpins were downloaded from miRBase v21.0 (Kozomara and Griffiths-Jones, 2013) and aligned to the genome using BLAST (Camacho et al., 2009). Coordinates of perfect full-length *miRNA* matches were added to the rest of non-coding RNA annotation.

**Methylation-assisted bisulfite (MAB) amplicon sequencing**

MAB sequencing was performed on *X. tropicalis* st9 gDNA in biological triplicates as previously described (Neri et al., 2016) . In short, gDNA and spike-in DNA underwent six consecutive rounds of M.SssI CpG methylation (NEB). After every second round, DNA was purified with AMPure XP beads (Beckman Coulter). Methylated DNA was bisulfite converted using the EpiTect Bisulfite Kit (Qiagen) as described (Neri et al., 2016). Spike-in DNA had five unmodified CpGs, two 5fCpGs on one strand, and several cytosines in CpN (N=A, T, C) context. Spike-in DNA controled for M.SssI methylation efficiency, 5fC signal recovery, and bisulfite conversion efficiency. The 5fC spike-in was prepared using two complementary synthetic oligonucleotides, one containing a 5’ overhang with two cytosines in CpG context. The oligos were mixed at 1:1 ratio and a fill-in PCR was performed in the presence of 5fdCTP (TriLink) instead of dCTP, to insert two 5fC bases. An analogous but unmodified spike-in was also prepared. 5fC modified and unmodified spike-in oligos were purified with DNA Clean & Concentrator-5 (Zymo research) and mixed in 1:1 ratio to produce 50% hemi-modified 5fC spike-in, which was added to the sequencing samples at 1: 20 ratio (spike-in: gDNA amplicon).

Forward strand specific PCR from MAB-treated DNA was performed with Phusion U Hot Start DNA Polymerase (Thermo Scientific) using primers specific for *tRNA-iMet*, *tRNA-Gly*, *gapdh* and 5fC spike-in DNA. A fraction of the sample was analyzed by 2% agarose gel electrophoresis to confirm product size and purity. PCR products were purified with DNA Clean & Concentrator-5 (Zymo Research) and eluted with 12 µl nuclease-free water. For strand-specific amplicon preparation from M.SssI treated and untreated DNA, we followed a two-step PCR process. To selectively amplify the forward strand, we first performed a 25-cycle linear amplification using Phusion U Hot Start DNA Polymerase and a single primer that adds a 20 nt sequence (F4) to the 5’ end of the forward strand of the target locus. PCR products were purified with DNA Clean & Concentrator-5 and eluted in 12 μl nuclease-free water. 2.5 μl purified PCR product were used as template for a second PCR, performed with Q5 High-Fidelity DNA Polymerase (NEB), F4-specific primer, and the respective second primer for each genomic sequence and spike-in DNA. PCR products were purified with DNA Clean & Concentrator-5.

DNA library preparation was performed using NEBNext Ultra II DNA Library Prep Kit for Illumina Version 7.0\_9/22. Libraries were prepared with 2.6 ng DNA and amplified with 8 PCR cycles. Two post-PCR purification steps were performed to exclude residual primer and adapter dimers. Libraries were profiled in a *High Sensitivity DNA* chip on a 2100 Bioanalyzer (Agilent technologies) and quantified using Qubit 1x dsDNA HS Assay Kit, in a Qubit 4.0 Fluorometer (Invitrogen). All samples were pooled together in equimolar ratio and sequenced on 1 NextSeq2000 P1 flow cell, in paired end mode for 2x 311 cycles plus 2x 8 cycles for the dual index read.

**5mC and 5fC base resolution profiling**

Available bisulfite sequencing (BS-seq) data (Bogdanović et al, 2016) for *X.tropicalis* st9, 12, 30 and 43 were used to obtain base-resolution 5mC profiles. The analysis was performed with bismark (v. 0.22.3) using default parameters with --local mode. Amplicon sequences retrieved from *X. troplicalis* v.7.1 genome were used as reference.

MAB-seq data from amplicon sequencing were quality checked with fastqc (v. 0.11.9) tool, and mapped to *in silico* converted amplicon sequences (C-to-T, except for cytosines in CpG context) using bowtie2 (v. 2.4.5) (--local, -N 0, -L 20). The pileup function from Rsamtools package (v. 2.14.0) was utilized to retrieve all substitution rates (min\_base\_quality = 30, min\_mapq = 10). Bisulfite conversion was estimated to be near 100%. 5fC signal in *tRNA-iMet* and *tRNA-Gly* loci were determined as C-to-T conversion rate (%C-to-TCpG) in MAB-seq samples (M.SssI + BS treated) after subtracting sequencing errors and single nucleotide polymorphism (SNP) at the ‘G’ in CpGs (%G-to-NCpG). 5fC levels were also corrected by a normalization factor (NF) for under-recovery of 5fC in spike-in controls. Thus, 5fC levels were calculated as: %5fC = (%C-to-TCpG – %G-to-NCpG) x NF). 5fC normalization factor is defined as %5fCexpected/%5fCobserved at the 5fC positions of the 5fC spike-in in 2 replicates.

**Genomic DNA preparation and LC-MS/MS analysis**

Genomic DNA was isolated with a DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions after treatment of lysed cells with RNase A (final concentration 1 mg/ml); this was followed by ethanol precipitation with ammonium acetate as salt. Quantification of 5mC and oxC was carried out as previously described (Schomacher et al., 2016).

**Expression constructs**

*X.tropicalis* *Tet2* and *Tet3* and *X. laevis* *Tet2* cDNA were inserted into pCS2FLAG (Addgene plasmid 16331). Human TDG and mutant TDGN140A cDNAs (plasmids pRS210 and pRS211, respectively, encoding both N-terminally HA-tagged proteins) were a kind gift from Primo Schär. For *tRNA-iMet* transgene expression, *X. tropicalis* *tRNA-iMet* gene sequence was inserted into pCS2+ and the CMV promoter of the plasmid was removed to avoid *tRNA-iMet* transcription by Pol II. Nb.Bpu10l restriction sequences were inserted directly upstream and downstream of the *tRNA-iMet* gene sequence that were used at the strand exchange assay (see below). The *tRNA-iMet* sequence was co-transcribed with a downstream 100 bp barcode (5’- AGATCTATTCTCGGCAATCTACGGAGCGACTTGATTATCAACAGCTGTCTAGCAGTTCTAATCTTTTGCCAACATCGTAATAGCCTCCAAGAGATTGATC-3’) which was initially intended to serve as a discriminant between endogenous and transgene *tRNA-iMet* in qPCR. In practice, we ended up using qPCR primers to amplify directly the *tRNA-iMet* sequence (see Suppl.Table 4) rather than the 100 bp barcode, since the endogenous *tRNA-iMet* background was negligible. pCS2+RLuc was prepared by inserting the RLuc sequence under the CMV promoter in a pCS2+ vector.

**Strand exchange assay**

Strand exchange was performed according to (Lühnsdorf et al., 2012) with the following changes: Double nicked *tRNA-iMet* plasmid was prepared by digesting 50 μg plasmid with 65 U Nb.Bpu10l (ER1681, Thermo Scientific) at 37oC for 2 hrs. Using a Biometra Thermal Cycler, the excised 92-mer was melted out at 90oC for 1 sec in presence of 200X molar excess of the synthetic unmodified (C) or modified (5mC or 5fC) strand (see Suppl. Table 6), then cooled to 65oC at a cooling rate 0.02oC/sec and then quickly to 4oC. Plasmids were concentrated with Amicon Ultra 30kD centrifugal filters (Merck), loaded on 0.8% agarose gel and the covalently closed plasmid band was extracted and purified with QIAquick Gel Extraction Kit (Qiagen).

***In vitro* TDG glycosylation assay**

The protocol was adapted from (Schomacher et al., 2016). 75 nM *tRNA-iMet* plasmid was incubated with 750 nM recombinant TDG (1:10 molar ratio) in glycosylase buffer (10 mM HEPES-KOH, pH 7.4, 1 M trimethylamine-N-oxide21, 100 mM KCl, 10 mM EDTA, 0.5 mM DTT, and 100 μg/ml BSA) at 37oC for 1 hr. Mock samples were incubated with buffer alone, without enzyme. Plasmid DNA was subsequently phenol/chloroform purified and ethanol precipitated.

***Xenopus* microinjections**

Embryos were obtained by *in vitro* fertilization as previously described (Gawantka et al., 1995). *X. tropicalis* and *laevis* expression constructs were used as templates to generate *mRNAs* with the MEGAscript® SP6 Transcription kit (Invitrogen) according to the manufacturer's instructions. Total amounts of mRNA injected per embryo was 1 ng *Tet2-* and 1 ng *Tet3 mRNA* for *Tet2/3* overexpression, 0.5 ng TDG/*TDGN140A* mRNA for TDG/dTDG phenotype scoring and 1 ng TDG/*TDGN140A* mRNA for Pol III ChIP-qPCR and northern blots. Control *mRNA* used for injections was bovine preprolactin (*Ppl*). ATG- blocking Morpholino antisense oligonucleotides (Gene Tools) were designed to target the respective genes. Morpholinos (Mo) and *mRNAs* were injected into animal blastomeres at one-cell stage with a total volume of 5 nl per *X. tropicalis* or 10 nl per *X. laevis* embryo. Total amounts of Mo injected per embryo for *Tet2/3* knockdown was 30 ng Tet2 and 30 ng Tet3 Mo or 60 ng of standard control Mo. For *Tdg* knockdown, *X. tropicalis* embryos were injected with 30 ng *Tdg* Mo. For *Tet2* Mo rescue experiment, *X. tropicalis* embryos were co-injected with 30 ng *X. tropicalis* specific Mo and 1 ng *X. laevis* *Tet2* *mRNA* or *PPL* *mRNA*. For Pol III inhibition, *X. tropicalis* embryos were injected with 2 ng α-amanitin (Sigma) per embryo. Total amount of *tRNA-iMet* transgene plasmid injected into *X. laevis* embryos was 200 pg. This plasmid was co-injected with 200 pg pCS2+RLuc per embryo, which was intended to function as a normalization control but finally not used since we normalized instead over the very *tRNA-iMet* transgene plasmid DNA whose expression was analyzed. Embryo images for phenotypes were taken with a Zeiss SteREO Discovery.V20 microscope.

**qPCR**

For DIP and ChIP samples, qPCR was performed using the Roche Universal probe library (UPL). UPL compatible primers were designed using the Roche Assay Design center (<http://lifescience.roche.com/shop/en/mx/overviews/brand/universal‐probe‐library>) (see Suppl. Table 4). For RT-qPCR of endogenous *U6 snRNA*, *5S* oocytic *rRNA* and *5.8S rRNA*, as well as *tRNA-iMet* transgene transcript levels quantification, total RNA was extracted from *X. tropicalis* and *X. laevis* embryos by standard Trizol RNA extraction. RNA was treated with DNase I Amplification Grade (AMPD1, Sigma) and cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) according to the instruction of the supplier. qPCR was performed using Roche LightCycler 480 SYBR Green I Master (see Suppl. Table 4). For *tRNA-iMet* transgene plasmid injected embryos, equal amount of RNA without DNase I and reverse transcriptase treatment was analyzed by qPCR alongside the cDNA, to account for the amount of plasmid DNA injected in each case. All qPCR reactions were performed in technical duplicates using Roche Light Cycler 480 in 384‐well format. Roche LightCycler analysis software was used for quantification.

**mRNA-seq**

50 st9 *X. tropicalis* embryos per sample after *Tet2/3* knockdown and control injection were homogenized in 1 ml of QIAzol. RNA was extracted using the miRNeasy kit (QIAGEN) following manufacturer’s instructions with on-column DNAse digestion. RNA integrity (RIN) values were verified using an RNA 6000 Nano kit on a 2100 Bioanalyzer (Agilent) to be consistently above 9.0. NGS library libraries were prepared with Illumina's TruSeq stranded mRNA LT Sample Prep Kit following Illumina’s standard protocol (Part # 15031047 Rev. E). Libraries were prepared with a starting amount of 1 μg and amplified in 10 PCR cycles. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). All samples were pooled in equimolar ratio and sequenced on 1 NextSeq 500 High Output Flowcell, SR mode for 1x 84 cycles plus 7 cycles for the index read.

Sample demultiplexing and FastQ file generation was performed using Illumina's bcl2fastq Conversion Software v.2.18.0.12. The raw sequence reads were quality assessed with FastQC v.0.11.5, aligned to the *X. troplicalis* v.7.1 genome with gene annotation from Xenbase (<https://ftp.xenbase.org/pub/Genomics/JGI/Xentr7.1> ) using STAR v.2.5.4b (<https://github.com/alexdobin/STAR>) and secondary alignments were removed with Samtools v.1.3.1 (<http://www.htslib.org>). Read counts were summarised on the gene level using Subread featureCounts v.1.5.1 (<https://subread.sourceforge.net>). Differential gene expression analysis was carried out in R v.3.3.2 using the Bioconductor package DESeq2 v.1.14.1 (<https://doi.org/doi:10.18129/B9.bioc.DESeq2>) following the recommended analysis workflow with independent gene filtering and without LFC shrinkage. Gene ontology (GO) enrichment analysis of the differentially expressed genes defined at 5% FDR cutoff (Suppl. Table 3) was performed with DAVID v.2022q3 (<https://david.ncifcrf.gov>) using the better annotated human orthologs, a specific background list of all expressed and statistically tested genes, “BP\_DIRECT” GO annotation database and a 5% FDR significance threshold.

**Dot blot**

Amplicons were generated using 5mC, 5hmC, 5fC dNTPs (TriLink) and control dNTPs using a human pOct4 plasmid as template. 50, 100, 150 ng amplicons and 150 ng unmodified pGL3 plasmid were melted at 95°C for 5 min in 2X SSC buffer. Samples were immediately placed on ice and spotted on a 6X SSC equilibrated Hybond®-N +  membrane (GE Healthcare) using a Bio-Dot microfiltration unit (Bio-Rad) according to manufacturer’s instructions. DNA was crosslinked with Stratalinker UV Crosslinker using 3x auto crosslink mode. The membrane was washed 2x in TBST (0.1% Tween 20, TBS) and blocked in 5% milk-TBST for 1 hr at RT. The blots were incubated with primary antibodies in 5% milk-TBST overnight at 4oC. Antibody concentrations: ssDNA (18731, IBL) 1:2000, 5mC (MAb-5MECYT-100, Diagnode) 1:1000, 5hmC (39791, Active Motif) 1:20,000 and 5fC (61227, Active Motif) 1:20,000. The membrane was washed 3x in TBST at RT, blocked in 5% milk-TBST for 10 min at RT and incubated with goat anti-mouse IgG-HRP conjugate (115-035-146, Dianova) or a goat anti-rabbit IgG-HRP conjugate (111-035-144, Dianova) 1:10,000 for 1 hr at RT. The membrane was washed 6x in TBST and visualized with SuperSignal West Pico (Thermo Scientific) and a Bio-Rad ChemiDoc imaging system using the Image Lab Software v6.1.0.

**Immunoblot**

*X. tropicalis* embryos microinjected with 1 ng *TDG*-, *TDGN140A-* or *Ppl* *mRNA* were collected at st9 and dissolved in NP-40 lysis buffer (2% NP-40, 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 100 mM N-ethylmaleimide, cOmplete EDTA-free Protease Inhibitor Cocktail and PhosSTOPTM phosphatase inhibitor (Roche)), 10 μl buffer per embryo. Homogenates were extracted with Freon and analyzed by SDS-PAGE. Primary antibodies: TDG (61437, Active Motif) 1:1000, α-tubulin (T5168, Sigma) 1:2000. Secondary antibodies: goat anti-rabbit IgG-HRP (111-035-144, Dianova) 1:2000, goat anti-mouse IgG-HRP (115-035-146, Dianova) 1:2000. Blots were visualized with SuperSignal West Pico (Thermo Scientific) and a Bio-Rad ChemiDoc imaging system using the Image Lab Software v. 6.1.0.

**Northern blot**

RNA from *X. tropicalis* st9 and 10 embryos was extracted with Trizol. 10 μg RNA per sample were boiled at 70oC for 10 min and immediately transferred to ice. Samples were loaded on a 15% PAGE (29:1 acrylamide/bisacrylamide) 8 M urea denaturing gel and ran in Tris-Borate-EDTA (TBE) buffer. Transfer was performed onto a Hybond-N+ membrane (GE Healthcare) with a Trans-Blot® TurboTM system (BioRad) at 25 V for 10 min. For chemical crosslinking, the membrane was placed on a Wattman paper soaked in crosslinking buffer (1% 1-methylimidazole, 12.5 mM HCl, 3.1% Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)) RNA facing up and incubated at 55oC for 2 hrs. The blot was blocked in 100 μg/ml heat denatured salmon sperm DNA (Thermo Scientific) in hybridization buffer (5X SSC buffer, 0.1% Ficol, 0.1% polyvinyl pyrolidone, 0.1% BSA) at 37oC for 2 hrs. Probes (see Suppl. Table 4) were 5’ end-labelled with γ32P ATP (Perkin Elmer) and T4 Polynucleotide kinase (NEB) according to manufacturer’s instructions at 37oC for 1 hr and purified with Microspin™ G-25 Columns (GE Healthcare). Blots were probed with 250 ng radioactive probe in hybridization buffer at 37oC overnight and washed 3x in hybridization buffer at 42oC for 30 min before autoradiography. Phosphor imaging was performed on a Typhoon FLA 9500 with Control Software (GE Healthcare).

**Mouse embryo and mESCs immunofluorescence staining**

Mouse zygotes were collected from the ampulla of CD1 superovulated females 8 hrs after the appearance of vaginal plug in M2 medium (Sigma) and cleaned from cumulus cells by 5 min incubation in M2 medium supplemented with 0.3 mg/ml hyaluronidase (Sigma). Embryos were washed 2x in PBS, fixed in 4% PFA, PBS for 20 min, washed 3x times in PBS and incubated for 8 min in Acidic Tyrode′s Solution (Sigma) to remove the zona pellucida. Embryos were washed 3x in 10 mg/ml BSA in PBS (PBS/BSA), permeabilised in 0.5% Triton X-100 in PBS/BSA for 1 hr, washed 3x in PBS/BSA and incubated with 10 mg/ml RNAse A in PBS/BSA for 1 hr at 37ºC. Embryos were blocked in 0.1 % Triton X-100 in PBS/BSA for 1 hr and incubated in the same solution containing rabbit RPC39 (Pol III) antibody (ab151495, Abcam) 1:100 overnight at 4°C. Embryos were subsequently washed 3x in PBS/BSA, fixed in 4% PFA, PBS for 10 min, washed in PBS/BSA, incubated in 4 N HCl for 15 min to denature DNA, and neutralised in 100 mM Tris-HCl (pH 8.5) for 5 min. The embryos were blocked in 0.1 % Triton X-100 in PBS/BSA for 1 hr and incubated with Alexa Flour® 546 goat anti-rabbit (A-11010, Thermo Scientific) overnight at 4°C in the dark. Because the second primary antibody was raised in the same species (rabbit) as the first primary antibody, the embryos were washed 3x in PBS/BSA, blocked in 5% normal rabbit serum (31883, Thermo Scientific) for 1 hr and incubated with goat anti-rabbit H+L Fab fragments (SAB3700970, Sigma) 1:20 for 1 hr. Embryos were then fixed in 4% PFA for 10 min, before incubating with the second primary antibody overnight at 4°C. The following antibodies were used: 5mC (MAb-5MECYT-100, Diagnode) 1:1000, 5hmC (39791, Active Motif) 1:200, 5fC (61227, Active Motif) 1:100, and 5caC (61229, Active Motif) 1:100. Embryos were then washed 3x in PBS/BSA and incubated with the second secondary antibody overnight at 4°C. Secondary antibodies used were Alexa Fluor® 546 goat anti-mouse (A-11030, Thermo Scientific) and Alexa Fluor® 647 goat anti-rabbit (A-21244, Thermo Scientific). Imaging was carried out using a Leica SP5 confocal microscope with a ×60 objective.

For mESCs IF, glass slides were coated overnight with 50 μg/ml poly-D-lysine. Cells were seeded and let attach to the slides for 2 hrs. Subsequently, media were replaced with 2i media containing 1:800 CellMask Plasma Membrane Orange Stain (Thermo Scientific) to stain plasma membranes for 10 min. The cells were washed 3x with PBS, fixed with 4% formaldehyde for 20 min and stained as described above for mouse embryos. Both 5mC and 5fC antibodies were diluted in Pierce Immunostain Enhancer (Thermo Scientific) for signal enhancement. Imaging was carried out with Leica Stellaris 8 confocal microscope with x60 objective.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**ImageJ analysis**

Image analysis was performed with ImageJ v5. Coloc2, a plugin written for ImageJ (NIH) was used to calculate the Mander’s colocalisation coefficient with thresholds using 30 nuclei as regions of interest from 5fC/Pol lll, 5fC/xNopp180 and Pol lll/xNopp180 double immunohistochemistry experiments (Figure 1H). Average number and size of foci found in 5 chromosome spreads was calculated using the Analyze Particles tool in ImageJ (Figure S1I). Distance in μm of 5fC, Pol III and xNopp180 immunohistochemistry signals (Figure 1G) and the blastopore perimeter (Figure 4G) were calculated using the measure tool in ImageJ. Northern blot quantification (Figure 5B, S5A-B) was performed with the Analyze Gels tool in ImageJ.

**Statistical analysis**

Data throughout the manuscript are displayed as arithmetic mean ± SD. In bar graphs, data values of individual biological replicates are represented with individual points. Statistical significance was determined by the statistical tests indicated at the respective figure legends using GraphPad Prism v9 and displayed as \*p ≤ 0.05, \*\* p ≤ 0.01, \*\*\* p ≤ 0.001 and \*\*\*\*p ≤ 0.0001. The exact P-values of statistically significant differences can be found at the respective figure legends.

**Acknowledgments**

We thank Marion S. Schmidt-Zachmann, Primo Schär, Meelad M. Dawlaty for reagents. We gratefully acknowledge technical support and advice by the IMB core facilities Genomics, Bioinformatics, and Microscopy and their members Fridolin Kielisch, and Sandra Ritz. The graphical abstract was created with  [BioRender.com](https://biorender.com/). This work was funded by the DFG (Ni 286/18-1).

**Contributions**

E.P., V.H., V.V. and C.N. designed the experiments. E.P., M.M., V.H., V.V., M.U.M., E.K., A.G., M.M.-L., D.H., L.S., D.M. performed or analyzed experiments. C.N. and E.P. wrote the manuscript. All authors were involved in discussions, planning the experiments and editing the manuscript.

**Declaration of interests**

The authors declare no competing interests.

References

Abe, K.-I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M.G., Schultz, R.M., and Aoki, F. (2018). Minor zygotic gene activation is essential for mouse preimplantation development. Proceedings of the National Academy of Sciences of the United States of America *115*, E6780-E6788.

Almeida, R.D., Loose, M., Sottile, V., Matsa, E., Denning, C., Young, L., Johnson, A.D., Gering, M., and Ruzov, A. (2014). 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. Epigenetics *7*, 383-389.

Andrews, M.T., Loo, S., and Wilson, L.R. (1991). Coordinate inactivation of class III genes during the Gastrula-Neurula Transition in Xenopus. Developmental biology *146*, 250-254.

Aoki, F., Worrad, D.M., and Schultz, R.M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Developmental biology *181*, 296-307.

Arand, J., Chiang, H.R., Martin, D., Snyder, M.P., Sage, J., Reijo Pera, R.A., and Wossidlo, M. (2022). Tet enzymes are essential for early embryogenesis and completion of embryonic genome activation. EMBO reports *23*, e53968.

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. Nat Chem Biol *11*, 555-557.

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nature Chem *6*, 1049-1055.

Bai, L., Yang, G., Qin, Z., Lyu, J., Wang, Y., Feng, J., Liu, M., Gong, T., Li, X., and Li, Z., et al. (2021). Proteome-Wide Profiling of Readers for DNA Modification. Advanced science (Weinheim, Baden-Wurttemberg, Germany) *8*, e2101426.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J., and Zhao, K. (2010). Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. Nat Struct Mol Biol *17*, 629-634.

Bazzini, A.A., del Viso, F., Moreno-Mateos, M.A., Johnstone, T.G., Vejnar, C.E., Qin, Y., Yao, J., Khokha, M.K., and Giraldez, A.J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. The EMBO journal *35*, 2087-2103.

Blitz, I.L., and Cho, K.W.Y. (2021). Control of zygotic genome activation in Xenopus. Current topics in developmental biology *145*, 167-204.

Blum, M., Robertis, E.M. de, Wallingford, J.B., and Niehrs, C. (2015). Morpholinos. Antisense and Sensibility. Developmental Cell *35*, 145-149.

Bogdanovic, O., Long, S.W., van Heeringen, S.J., Brinkman, A.B., Gómez-Skarmeta, J.L., Stunnenberg, H.G., Jones, P.L., and Veenstra, G.J.C. (2011). Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. Genome Res. *21*, 1313-1327.

Bogdanović, O., Smits, A.H., La Calle Mustienes, E. de, Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., and van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet *48*, 417-426.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. Science *336*, 934-937.

Booth, M.J., Marsico, G., Bachman, M., Beraldi, D., and Balasubramanian, S. (2014). Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nature Chem *6*, 435-440.

Bouniol, C., Nguyen, E., and Debey, P. (1995). Endogenous transcription occurs at the 1-cell stage in the mouse embryo. Experimental cell research *218*, 57-62.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+. Architecture and applications. BMC Bioinformatics *10*, 421.

Chen, H., Einstein, L.C., Little, S.C., and Good, M.C. (2019). Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. Developmental Cell *49*, 852-866.e7.

Cheng, Mittnenzweig, Mayshar, Lifshitz, Dunjić, Rais, Ben-Yair, Gehrs, Chomsky, and Mukamel, et al. (2022). The intrinsic and extrinsic effects of TET proteins during gastrulation. Cell *185*, 3169-3185.e20.

Cimmino, L., and Aifantis, I. (2017). Alternative roles for oxidized mCs and TETs. Current opinion in genetics & development *42*, 1-7.

Clarkson, S.G., and Kurer, V. (1976). Isolation and some properties of DNA coding for tRNA1met from Xenopus laevis. Cell *8*, 183-195.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., and Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., Ge, H., Wang, R., Chapman, D.L., and Tang, F., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature *538*, 528-532.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., and Lyko, F., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Developmental Cell *29*, 102-111.

Dreszer, T.R., Karolchik, D., Zweig, A.S., Hinrichs, A.S., Raney, B.J., Kuhn, R.M., Meyer, L.R., Wong, M., Sloan, C.A., and Rosenbloom, K.R., et al. (2011). The UCSC Genome Browser database. Extensions and updates 2011. Nucleic Acids Research *40*, D918-D923.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nature reviews. Molecular cell biology *19*, 436-450.

Gao, Y., Li, L., Yuan, P., Zhai, F., Ren, Y., Yan, L., Li, R., Lian, Y., Zhu, X., and Wu, X., et al. (2020). 5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution. PLoS biology *18*, e3000799.

Gawantka, V., Delius, H., Hirschfeld, K., Blumenstock, C., and Niehrs, C. (1995). Antagonizing the Spemann organizer. Role of the homeobox gene Xvent-1. The EMBO journal *14*, 6268-6279.

Graczyk, D., Cieśla, M., and Boguta, M. (2018). Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. Biochimica et biophysica acta. Gene regulatory mechanisms *1861*, 320-329.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., and Hu, B., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell stem cell *15*, 447-459.

Hardeland, U., Bentele, M., Jiricny, J., and Schär, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. The Journal of biological chemistry *275*, 33449-33456.

Hardwick, J.S., Ptchelkine, D., El-Sagheer, A.H., Tear, I., Singleton, D., Phillips, S.E.V., Lane, A.N., and Brown, T. (2017). 5-Formylcytosine does not change the global structure of DNA. Nat Struct Mol Biol *24*, 544-552.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., and Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303-1307.

Helbo, A.S., Lay, F.D., Jones, P.A., Liang, G., and Grønbæk, K. (2017). Nucleosome Positioning and NDR Structure at RNA Polymerase III Promoters. Scientific reports *7*, 41947.

Hill, P.W.S., Amouroux, R., and Hajkova, P. (2014). DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming. An emerging complex story. Genomics *104*, 324-333.

Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S.J., Bogdanovic, O., Lister, R., and Veenstra, G.J.C. (2015). Embryonic transcription is controlled by maternally defined chromatin state. Nat Commun *6*, 553.

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., and Gong, W., et al. (2015). Structural insight into substrate preference for TET-mediated oxidation. Nature *527*, 118-122.

Huang, S., Deerinck, T.J., Ellisman, M.H., and Spector, D.L. (1998). The perinucleolar compartment and transcription. The Journal of cell biology *143*, 35-47.

Huang, Y., Kim, J.K., Do, D.V., Lee, C., Penfold, C.A., Zylicz, J.J., Marioni, J.C., Hackett, J.A., and Surani, M.A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. eLife *6*.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell research *21*, 1670-1676.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300-1303.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol *14*, R119.

Iurlaro, M., McInroy, G.R., Burgess, H.E., Dean, W., Raiber, E.-A., Bachman, M., Beraldi, D., Balasubramanian, S., and Reik, W. (2016). In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. Genome Biol *17*, 930.

Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic Genome Activation in Vertebrates. Developmental Cell *42*, 316-332.

Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R.S., Thompson, E.C., Surani, M.A., and Koralov, S.B., et al. (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. Proceedings of the National Academy of Sciences of the United States of America *112*, E4236-45.

Kellinger, M.W., Song, C.-X., Chong, J., Lu, X.-Y., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol *19*, 831-833.

Kieserman, E.K., and Heald, R. (2011). Mitotic chromosome size scaling in Xenopus. Cell Cycle *10*, 3863-3870.

Kirkland, J.G., Raab, J.R., and Kamakaka, R.T. (2013). TFIIIC bound DNA elements in nuclear organization and insulation. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 418-424.

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T., and Khobta, A. (2017). Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. Nucleic Acids Research *45*, 11033-11042.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kolitz, S.E., and Lorsch, J.R. (2010). Eukaryotic initiator tRNA. Finely tuned and ready for action. FEBS Letters *584*, 396-404.

Kozomara, A., and Griffiths-Jones, S. (2013). miRBase. Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research *42*, D68-D73.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science (New York, N.Y.) *324*, 929-930.

Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic Genome Activation During the Maternal-to-Zygotic Transition. Annu. Rev. Cell Dev. Biol. *30*, 581-613.

Liu, S., Wang, J., Su, Y., Guerrero, C., Zeng, Y., Mitra, D., Brooks, P.J., Fisher, D.E., Song, H., and Wang, Y. (2013). Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. Nucleic Acids Research *41*, 6421-6429.

Lu, X., Han, D., Zhao, B.S., Song, C.-X., Zhang, L.-S., Doré, L.C., and He, C. (2015). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell research *25*, 386-389.

Lühnsdorf, B., Kitsera, N., Warken, D., Lingg, T., Epe, B., and Khobta, A. (2012). Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. Analytical biochemistry *425*, 47-53.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine. Potential implications for active demethylation of CpG sites. The Journal of biological chemistry *286*, 35334-35338.

Marshall, L., and White, R.J. Non-coding RNA production by RNA polymerase III is implicated in cancer. Nature Reviews Cancer *8*, 911-914.

Matera, A.G., Frey, M.R., Margelot, K., and Wolin, S.L. (1995). A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. The Journal of cell biology *129*, 1181-1193.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Molecular Cell *61*, 874-885.

Moqtaderi, Z., Wang, J., Raha, D., White, R.J., Snyder, M., Weng, Z., and Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nat Struct Mol Biol *17*, 635-640.

Müller, F., Clarkson, S.G., and Galas, D.J. (1987). Sequence of a 3.18 kb tandem repeat of Xenopus laevis DNA containing 8 tRNA genes. Nucleic Acids Research *15*, 7191.

Müller, N., Ponkkonen, E., Carell, T., and Khobta, A. (2021). Direct and Base Excision Repair-Mediated Regulation of a GC-Rich cis-Element in Response to 5-Formylcytosine and 5-Carboxycytosine. International journal of molecular sciences *22*.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., and Tate, J., et al. (2015). Rfam 12.0. Updates to the RNA families database. Nucleic Acids Research *43*, D130-D137.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1. 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010). Initial genomics of the human nucleolus. PLoS genetics *6*, e1000889.

Neri, F., Incarnato, D., Krepelova, A., Parlato, C., and Oliviero, S. (2016). Methylation-assisted bisulfite sequencing to simultaneously map 5fC and 5caC on a genome-wide scale for DNA demethylation analysis. Nature protocols *11*, 1191-1205.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., and Oliviero, S. (2015). Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. Cell Reports *10*, 674-683.

Ngo, T.T.M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., and Ha, T. (2016). Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. Nat Commun *7*, 10813.

Nizami, Z.F., and Gall, J.G. (2012). Pearls are novel Cajal body-like structures in the Xenopus germinal vesicle that are dependent on RNA pol III transcription. Chromosome Research *20*, 953-969.

Norton, J.T., Wang, C., Gjidoda, A., Henry, R.W., and Huang, S. (2009). The perinucleolar compartment is directly associated with DNA. The Journal of biological chemistry *284*, 4090-4101.

Oler, A.J., Alla, R.K., Roberts, D.N., Wong, A., Hollenhorst, P.C., Chandler, K.J., Cassiday, P.A., Nelson, C.A., Hagedorn, C.H., and Graves, B.J., et al. (2010). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nat Struct Mol Biol *17*, 620-628.

Orioli, A., Praz, V., Lhôte, P., and Hernandez, N. (2016). Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. Genome Res. *26*, 624-635.

Papin, C., Ibrahim, A., Le Gras, S., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S., and Hamiche, A. (2017). Combinatorial DNA methylation codes at repetitive elements. Genome Res. *27*, 934-946.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., and Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Pollock, C., and Huang, S. (2010). The perinucleolar compartment. Cold Spring Harbor perspectives in biology *2*, a000679.

Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome Biol *13*, R69.

Raiber, E.-A., Murat, P., Chirgadze, D.Y., Beraldi, D., Luisi, B.F., and Balasubramanian, S. (2015). 5-Formylcytosine alters the structure of the DNA double helix. Nat Struct Mol Biol *22*, 44-49.

Raiber, E.-A., Portella, G., Martínez Cuesta, S., Hardisty, R., Murat, P., Li, Z., Iurlaro, M., Dean, W., Spindel, J., and Beraldi, D., et al. (2018). 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. Nature Chem *10*, 1258-1266.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics & chromatin *6*, 39.

Schmidt-Zachmann, M.S., Hügle, B., Scheer, U., and Franke, W.W. (1984). Identification and localization of a novel nucleolar protein of high molecular weight by a monoclonal antibody. Experimental cell research *153*, 327-346.

Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhöfer, S., Seggern, A. von, and Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. Nat Struct Mol Biol *23*, 116-124.

Shen, L., Song, C.-X., He, C., and Zhang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. Annual review of biochemistry *83*, 585-614.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692-706.

Slenn, T.J., Morris, B., Havens, C.G., Freeman, R.M., JR, Takahashi, T.S., and Walter, J.C. (2014). Thymine DNA glycosylase is a CRL4Cdt2 substrate. The Journal of biological chemistry *289*, 23043-23055.

Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., and Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678-691.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., and Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146-1159.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., and Rudd, M.K., et al. (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet *7*, e1002154.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., and Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, N.Y.) *324*, 930-935.

Thompson, M., Haeusler, R.A., Good, P.D., and Engelke, D.R. (2003). Nucleolar clustering of dispersed tRNA genes. Science (New York, N.Y.) *302*, 1399-1401.

Wagner, M., Steinbacher, J., Kraus, T.F.J., Michalakis, S., Hackner, B., Pfaffeneder, T., Perera, A., Müller, M., Giese, A., and Kretzschmar, H.A., et al. (2015). Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angewandte Chemie *54*, 12511-12514.

Wang, C., Politz, J.C., Pederson, T., and Huang, S. (2003). RNA polymerase III transcripts and the PTB protein are essential for the integrity of the perinucleolar compartment. Molecular biology of the cell *14*, 2425-2435.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & development *25*, 679-684.

Xia, B., Han, D., Lu, X., Sun, Z., Zhou, A., Yin, Q., Zeng, H., Liu, M., Jiang, X., and Xie, W., et al. (2015). Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. Nature methods *12*, 1047-1050.

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.-L., and Wang, H., et al. (2016). Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. Molecular Cell *64*, 913-925.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., and Huang, S., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular Cell *42*, 451-464.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Di Hu, Zhao, B., and Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200-1213.

Yeganeh, and Hernandez (2020). RNA polymerase III transcription as a disease factor. Genes & development *34*, 865-882.

You, C., Ji, D., Dai, X., and Wang, Y. (2014). Effects of Tet-mediated oxidation products of 5-methylcytosine on DNA transcription in vitro and in mammalian cells. Scientific reports *4*, 7052.

Yun Gao, Lin Li, Peng Yuan, Fan Zhai, Yixin Ren, Liying Yan, Rong Li, Ying Lian, Xiaohui Zhu, and Xinglong Wu, et al. 5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution.

Zeng, T.-B., Han, L., Pierce, N., Pfeifer, G.P., and Szabó, P.E. (2019). EHMT2 and SETDB1 protect the maternal pronucleus from 5mC oxidation. Proceedings of the National Academy of Sciences of the United States of America *116*, 10834-10841.

Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., and Yi, C. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. Cell stem cell *20*, 720-731.e5.

Zillner, K., Komatsu, J., Filarsky, K., Kalepu, R., Bensimon, A., and Németh, A. (2015). Active human nucleolar organizer regions are interspersed with inactive rDNA repeats in normal and tumor cells. Epigenomics *7*, 363-378.

**Figure Legends**

**Figure 1. 5fC accumulates in *X. laevis*****embryos****at zygotic genome activation.**

**(A**) LC-MS/MS analysis of genomic 5mC, 5hmC, 5fC and 5caC of stage 6.5-13 *X. laevis* embryos. Embryos were harvested at the beginning of the indicated Nieuwkoop & Faber stages or 0.5 or 1 hr afterwards. Grey shadowing indicates ZGA. n=3 biological replicates +/- SD. Empty circles, signal below detectability.

**(B)** Immunofluorescence microscopy (IF) of 5mC, 5hmC, 5fC and 5caC of sagittally cut st9 (ZGA) *X.* *laevis* embryos. White arrowheads indicate 5fC signal in the animal pole. Yoyo1, DNA stain. a, animal pole; v, vegetal pole; bc, blastocoel cavity. Scale bar 200 µm.

**(C)** IF for 5mC, 5hmC, 5fC and 5caC in nuclei of st9 (ZGA) *X. laevis* embryos. Note the kidney-shaped 5fC chromocenter (white arrow). Yoyo1, DNA stain. Scale bar 5 µm.

**(D)** IF for 5fC in nuclei of *X. laevis* embryos at st7 (blastula), 8 (early ZGA), 9 (ZGA), 10 (early gastrula) and 13 (early neurula). White arrows indicate 5fC chromocenter. Yoyo1, DNA stain. Scale bar 5 µm.

**(E)** IF for 5mC and 5fC on chromosome spreads from st9 (ZGA) *X. laevis* embryos. Note clustered 5fC foci on different chromosomes. Yoyo1, DNA stain. Scale bar 10 µm.

**(F)** Co-IF of 5fC and the nucleolar marker xNopp180 (top), Pol III and xNopp180 (middle) or 5fC and Pol lll (bottom). Note adjoining 5fC-xNopp180 and Pol III-xNopp180 spots, and colocalizing 5fC-Pol III (white arrows). Yoyo1, DNA stain. Scale bar 5 µm.

**(G)** Adjacency of 5fC/Pol III, 5fC/xNopp180 and Pol lll/xNopp180 foci in co-IF measured by foci-distance. Foci number: 5fC/Pol III n=6; 5fC/xNopp180 n=24; Pol III/xNopp180 n=29.

**(H)** Mander’s colocalisation coefficient for 5fC/Pol lll, 5fC/xNopp180, and Pol lll/xNopp180 co-IF. Each point represents an individual nucleus and lines represent median values.

**Figure 2. 5fC clusters at Pol III genes during ZGA.**

**(A)** Heatmap representation of ZGA (st9) *X. tropicalis* embryo 5mC, 5hmC, 5fC and 5caC from DIP sequencing showing Pearson correlation coefficients score. Correlation is ranked from 0 to 1 where 1 is strongly correlated (dark blue) and 0 uncorrelated (pale green). Dashed red box highlights 5fC.

**(B)** Correlation of 5mC, 5hmC, 5fC and 5caC peaks from DIP sequencing with Pol I, Pol II and Pol III target genes.

**(C)** Percent 5fC peaks overlap with different classes of Pol III genes.

**(D)** Percent 5mC, 5hmC, 5fC or 5caC-peak clusters at Pol III genes. Total count of clusters indicated on top of each bar.

**(E)** DIP-qPCR of 5mC, 5hmC, 5fC, 5caC and IgG of *X. tropicalis* embryos for the indicated developmental stages and Pol III, Pol II and Pol I genes. Normalization over *Sox2*. Values <10-5 were set to 0. n= 3 DIP-pulldown samples +/- SD. ND, not detected.

**(F)** Violin plots showing peak length distribution of Pol III peaks overlapping with DIP peaks at *tRNA* genes.

**(G)** Percentage of 5mC, 5hmC, 5fC, 5caC DIP-peaks and Pol III-ChIP peaks overlapping with various *tRNA* genes. Arrows indicate *tRNA-iMet* and *tRNA-eMet*.

**(H)** Correlation of 5fC enriched *tRNA* genes with optimal and non-optimal codons. For 5fC optimal codons P-value =10-5 and is estimated in comparison with background *tRNA* genes using χ2 test.

**Figure 3. 5fC is associated with active chromatin at the *tRNA* genes internal promoter B-box**

**(A)** Frequency plots (top) and heatmaps (bottom) of normalized read density of the indicated features, centered at 5fC peaks associated with *tRNA* gene clusters +/- 5 kb. Comparison is with *X.* *tropicalis* embryo Pol III ChIP sequencing and input samples (data from this study), as well as active (H3K4me1, p300) and inactive (H3K9me3, H4K20me3) histone marks from published studies.

**(B)** Metagene profiles showing average read density of 5fC DIP sequencing signal in comparison to indicated features of ZGA (st9) embryos, centered at *tRNA* genes.

**(C-D)** Frequency plots (top) and heatmaps (bottom) of normalized read density of the indicated features at *X. tropicalis* **(C)** st10.5 and **(D)** st16, centered at 5fC peaks associated with *tRNA* gene clusters +/- 5 kb. Note the successive shutdown of the tandem *tRNA* loci from ZGA (st9) to neurula (st16), as indicated by decreased active and increased inactive mark densities.

**(E-G)** UCSC browser screenshot of a 5fC *tRNA* gene cluster with normalized DIP sequencing and ChIP sequencing coverage tracks of ZGA (st9) *X. tropicalis* embryos. **(E)** Scaffold 58:70,000-230,000 (160 kb): *tRNA* annotation is shown at the bottom, the colors are explained in panel D. **(F)** Scaffold\_58:80,443-91,056: 1st zoom of the highlighted section from B (blue shaded) encompassing 3.5 tandem *tRNA* gene repeats. L1: line 1 centered at 5fC peak. L2: line 2 centered at 5mC peak. **(G)** Scaffold\_58:82,347-86,478: 2nd zoom of the highlighted section from C (blue shaded) indicating the seven *tRNA* genes of one repeat.

**(H)** 5fC (green) and 5mC (blue) methylation assisted bisulfite (MAB) or bisulfite (BS) sequencing and DIP sequencing profiles at *X. tropicalis* ZGA (st9) for *tRNA-iMet* (Scaffold\_58:76,642-77,045) and *tRNA-Gly* (Scaffold\_109:170,885-171,190) forward (non-coding) strand. *gapdh* (Scaffold\_7:7,145,577-7,145,923) serves as a 5fC-negative control.

**Figure 4. 5fC promotes Pol III binding at ZGA**

**(A)** Gastrula stage *X.* *tropicalis* embryos in vegetal view injected with the indicated antisense-Morpholinos (Mo). White arrowheads indicate lysed cells.

**(B)** Phenotype scoring of gastrula stage *X. tropicalis* embryos injected with the indicated Mo. N=90 embryos per condition.

**(C)** Pol lll ChIP-qPCR in ZGA (st9) *X. tropicalis* embryos injected with control Mo and control *mRNA* (Co), *Tet2/3* Morpholino (*Tet2/3* KD) or *Tet2/3* *mRNA* (*Tet2/3* OE). Values are normalized to Pol III occupancy at *tRNA-Gly*. Data are mean of n=3 independent ChIP pulldown samples +/- SD. Adjusted P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Dunnett’s test. iMet: \*\*\*\*<0.0001, \*=0.013; U6: \*=0.04.

**(D)** TfIIIc ChIP-qPCR in ZGA (st9) *X. tropicalis* embryos injected with control (Co) or *Tet2/3* Morpholinos (*Tet2/3* KD). Values are normalized to Pol III occupancy at *tRNA-Gly*. Data are mean of n=3 independent ChIP pulldown samples +/- SD. Adjusted P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Šídák's test. Gly: \*\*\*\*<0.0001; iMet: \*=0.038; Asp: \*\*\*\*<0.0001; eMet: \*\*\*=0.0001.

**(E)** Gastrula stage *X. laevis* embryos in vegetal view injected with TDG- or catalytically inactive dTDG mRNA.

**(F)** Phenotype scoring of gastrula stage *X. laevis* embryos injected with TDG -, dTDG mRNA or untreated control. The number of embryos scored is specified on the top.

**(G)** Blastopore perimeter in mm of gastrula stage *X. laevis* embryos injected TDG-, dTDG *mRNA* or untreated control. Adjusted P-values are calculated with 1-way ANOVA and corrected for multiple comparisons with Tukey’s test. \*\*\*=0.0003, \*\*\*\*<0.0001. N=21 untreated, N= 25 dTDG mRNA, N=24 TDG mRNA embryos.

**(H)** Pol III ChIP – qPCR in ZGA (st9)  *X. tropicalis embryos* injected with Ppl mRNA (Co), TDG mRNA (TDG OE) or dTDG mRNA (dTDG OE). Values are normalized to Pol III occupancy at *tRNA-Gly*. Data are mean of n=4 independent ChIP pulldown samples +/- SD. Adjusted P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Tukey’s test. Gly: \*\*\*=0.0006; iMet: \*\*\*=0.0002; Asp: \*\*\*\*<0.0001.

**Figure 5. 5fC promotes *tRNA* transcription at ZGA**

**(A)** *U6* *snRNA* RT-qPCR at ZGA (st9) *X. tropicalis* embryos untreated or microinjected with 2 ng α-amanitin to inhibit Pol III. *U6* expression is normalized over *5.8S* *rRNA* levels (Pol I target gene) and represented as fold change over the control. n=3 biological replicates +/- SD. P-value calculated with unpaired t-test. \*\*\*=0.0007.

**(B)** Quantification of northern blot of *tRNA-iMet* in *X. tropicalis* embryos microinjected with α-amanitin, *Tet2/3* Mo (*Tet2/3* KD) or TDG mRNA (TDG OE) and their respective controls (Co), at st9 and 10. *tRNA-iMet* levels are normalized over *5.8S* *rRNA* and represented as fold change over control. α-amanitin, TET KD st10: n=3, TET KD st9, TDG OE st9-10: n=4. P-values calculated with paired t-test. α-amanitin: \*\*=0.0057; *Tet2/3* KD: \*\*=0.0085;

**(C)** Differential expression MA-plot of *Tet2/3* KD in ZGA (st9) *X. tropicalis* embryos. The up- and down-regulated genes (5% FDR) are highlighted in red and their numbers are indicated.

**(D)** Gene Ontology terms of significantly enriched up- (top) and downregulated (bottom) genes from (C).

**(E)** *tRNA-iMet* transgene assays shown in (F)-(H). Top: *tRNA-iMet* plasmid with the two modified cytosines in the non-coding (template) strand of the *tRNA-iMet* gene. Middle: Three different variants of the *tRNA-iMet* plasmid, in which the noted cytosines are either C, 5mC, or 5fC. Bottom: *X. laevis* embryos at one-cell stage were microinjected with *tRNA-iMet* plasmid without strand exchange (‘No SE’) or plasmids that had undergone strand exchange with synthetic oligonucleotides that were unmodified (‘C’), or 5mC, or 5fC modified. Embryos were harvested at the indicated stages and *tRNA-iMet* was quantified by RT-qPCR and normalized to the respective *tRNA-iMet* plasmid DNA levels determined by qPCR. Adjusted P-values were calculated with one-way ANOVA (F) or 2-way ANOVA (G)-(H) on ΔCp values (normally distributed) and corrected for multiple comparisons with Tukey’s test, shown +/- SD. Each biological replicate is a pool of 6-12 embryos.

**(F)** *tRNA-iMet* expression analysis as explained in (E). Embryos were harvested at st8, st10, or st11. St8 and 10, n=4; st11, n=5 biological replicates. Adjusted P-values: \*\*\*=0.0004, \*\*\*\*<0.0001.

**(G)** *tRNA-iMet* expression analysis as explained in (E). Embryos were microinjected with C or 5fC SE-plasmids into the animal (AP) or the vegetal pole (VP) and harvested at st10, n=4 biological replicates. Adjusted P-values: \*\*=0.0022, \*\*\*=0.0002.

**(H)** *tRNA-iMet* expression analysis as explained in (E). Injected *tRNA-iMet* plasmids were mock- or TDG treated before injection and microinjected embryos were harvested at st10, n=4 biological replicates. Adjusted P-values: Mock C-5fC: \*=0.018; 5fC mock-TDG: \*=0.016.

**Figure 6. 5fC accumulation at Pol III sites is conserved in mouse ZGA**

**(A, B)** Immunofluorescence microscopy (IF) of Pol lll (RPC39) and 5fC in mouse zygotes and 2-cell stage preimplantation embryos (A) and 5mC, 5hmC, 5fC and 5caC with Pol III in PN4/5 mouse zygotes (B). Yoyo1, DNA staining. White arrows and respective insets show foci of colocalizing 5fC and Pol III. PN, pronuclei; ♂, paternal pronucleus; ♀, maternal pronucleus; Scale bar, 10 µm.

**(C)** Number of 5fC foci, Pol III foci and overlapping 5fC/Pol III foci in PN4/5 stage embryos. n=86 embryos.

**(D)** Quantification of number of 5fC foci per mouse embryo at the indicated stages. PN2/3: n=12, PN4/5: n=141, 2-cell stage: n=41

**(E)** Model for 5-formylcytosine as activating epigenetic mark for Pol III during ZGA. *Xenopus* embryos are special in containing large *tRNA* tandem repeats, which are transcribed from zygotic genome activation stage (ZGA) until gastrula stage. During ZGA, Tet dixoxygenases oxidize these hypermethylated *tRNA* loci, leading to 5fC accumulation. These 5fC marks promote transcription by an unknown mechanism that may involve either recruitment of 5fC readers or changes in DNA structure. At the end of gastrulation, the *tRNA* tandem repeats are silenced as DNA methylation marks become repressive, 5fC is lost upon expression of *thymine DNA glycosylase* and repressive histone marks accumulate.

**Supplementary Figures**

**Supplementary Figure 1: oxC analysis and specificity controls**

**(A**) LC-MS/MS analysis of genomic 5mC, 5hmC, 5fC and 5caC of st7-42 *X. laevis* embryos. Note that 5hmC is not absent in gastrula and neurula stages but present in very low levels. Grey shadowing indicates ZGA. n=3 biological replicates +/- SD. ND, not detected.

**(B)** Dot blot using control (C), 5mC, 5hmC, 5fC amplicons and unmodified plasmid shows specificity of 5mC, 5hmC and 5fC antibodies. α-ssDNA antibody serves as a loading control.

**(C)** Immunofluorescence microscopy (IF) of *X. laevis* tadpole sections for 5mC, 5hmC, 5fC and 5caC. Antibody specificity shown by outcompeting antibody binding with control (Co), 5mC, 5hmC, 5fC amplicons or 5caC dNTPs. Yoyo1, DNA stain.

**(D)** Potassium perruthenate (KRuO4) treatment of *X. laevis* embryos. Left: LC-MS/MS analysis of genomic 5hmC and 5fC following treatment with increasing concentrations of KRuO4. n=3 biological replicates +/- SD. Adjusted P-values were calculated with ordinary 1-way ANOVA and Dunnett's multiple comparisons test. \*\*=0.0047, \*\*\*\*=<0.0001. Right: IF of 5hmC after 1 mM KRuO4 treatment. Yoyo1, DNA stain.

**(E)** Sodium borohydride (NaBH4) treatment of *X. laevis* embryos. Left: LC-MS/MS analysis of genomic 5hmC and 5fC following treatment with increasing concentrations of NaBH4. n=3 biological replicates +/- SD. Adjusted P-values were calculated with ordinary 1-way ANOVA and Dunnett's multiple comparisons test. \*\*=0.001,\*\*\*=0.0001, \*\*\*\*=<0.0001. Right: IF of 5hmC and 5fC after control treatment and 40 µM NaBH4. Yoyo1, DNA stain.

**(F)** IF for 5fC in sagittally cut st10 (early gastrula) and st13 (early neurula) *X.* *laevis* embryos. Yoyo1, DNA stain. a, anterior ; p, posterior. Scale bar 200 µm.

**(G)** Quantification of 5fC IF signal distribution at the indicated stages of *X. laevis* development. Signal distribution (% cells) is categorized as spot, homogeneous/spot or homogenous signal in the nucleus. The number of nuclei counted is noted on top of each bar.

**(H)** IF for 5fC in ZGA (st9) *X.* *laevis* embryo nucleus using an alternative 5fC antibody from Cell Signalling (CS). Yoyo1, DNA stain.

**(I)** Quantification of 5fC foci in ZGA (st9) *X.* *laevis* embryo nuclei metaphase spreads. n= 5 independent experiments +/- SD. S, small; M, medium; L, large. Foci size is in μm.

**(J)** DIP-qPCR using 5mC, 5hmC, 5fC, and IgG antibodies in ZGA (st9) *X.* *tropicalis* embryos. qPCR was carried out for 5mC, 5hmC, 5fC and control amplicons spiked into the DIP pulldown samples. n=3 independent DIP pulldown samples +/- SD.

**(K)** LC-MS/MS analysis of genomic 5mC, 5hmC, 5fC and 5caC levels from ZGA (st9) *X. tropicalis* DIP samples. n=3 independent DIP pulldown samples +/- SD. ND, not detected.

**Supplementary figure 2. oxC marks genomic localization at ZGA**

**(A)** Venn diagram depicting overlap of 5mC, 5hmC, 5fC and 5caC peaks from ZGA (st9) *X. tropicalis* embryo DIP sequencing.

**(B)** Pie chart of distribution of genomic regions overlap with the 5mC, 5hmC, 5fC and 5caC DIP peaks.

**(C)** Enrichment of 5mC, 5hmC, 5fC and 5caC DIP peaks at ncRNA.

**(D)** Percent 5mC peaks overlapping different classes of Pol III genes.

**(E)** Pol III gene classes distribution in *X. tropicalis* genome.

**(F)**Violin plot showing peak length distribution of 5mC, 5hmC, 5fC and 5caC clusters associated with *tRNA* genes.

**(G)** UCSC browser screenshot of the 160 kb *tRNA* gene cluster showing normalized DIP sequencing coverage using only uniquely mapped reads (UM) or all mapped reads including multi-mapped ones (MM).Note that the y axis scale in UM peaks is 100 lower than in MM peaks.

**(H)** DIP-qPCR of 5mC, 5hmC, 5fC, 5caC and IgG of *X. tropicalis* embryos for the indicated developmental stages and Pol III and Pol I genes. Normalization over *Sox2*. Values <10-5 were set to 0. n= 3 DIP-pulldown samples +/- SD. ND, not detected.

**(I)** Violin plots showing peak length distribution of Pol III peaks overlapping with all DIP peaks for each DNA modification.

**Supplementary Figure 3. 5fC is associated with active transcription marks.**

**(A)** Heatmaps and frequency plots of normalized read density of the indicated features, centred at 5mC peaks at *tRNA* gene clusters +/- 5 kb. Comparison is with *X.* *tropicalis* embryo Pol III ChIP-seq and input samples (data from this study, only for st9), as well as active (H3K4me1, p300) and inactive (H3K9me3, H4K20me3) histone marks at the indicated stages.

**(B)** Frequency plot (top) and heat map (bottom) showing normalized read density of 5fC DIP sequencing signal centered at the B-box motif of tRNA genes covered by 5fC peaks at st9 *X. tropicalis* embryos.

**(C-E)** UCSC browser screenshots of 5fC peaks at clusters of **(C)** *U6 snRNA* (Pol III gene), **(D)** *SSU rRNA* (Pol I gene) and **(E)** *tRNA* genes. Vertical line at (E) is aligned to *tRNA-Gly*.

**(F)** *tRNA-iMet* locus methylation profile from *X. tropicalis* 5mC DIP sequencing (data from this study, only for st9) and base resolution bisulfite sequencing (BS) from public data at the indicated stages.

**(G)** MAB sequencing quantification as percentage C-to-T conversion per CpG at the *tRNA-iMet* locus (Scaffold\_58:76,642-77,045) in MAB and control (untreated and M.SssI treated) *X. tropicalis* st9 samples. The two consecutive *tRNA-iMet* genes (red) and their respective B-boxes (black) are indicated at the bottom of the panel. n=3 biological replicates. Values are average +/- SD.

**Supplementary Figure 4: Morpholino, *mRNA* overexpression and ChIP controls**

**(A)** *Tet2* Mo rescue experiment. Phenotypes of st30 *X. tropicalis* embryos after injection with a combination of control- or *Tet2* Mo with a control- (Ppl) or *X.laevis* *Tet2* *mRNA*.

**(B)** Quantification of phenotypes shown in (B). N=30 embryos per condition.

**(C-D)** LC-MS/MS analysis of genomic 5mC and 5fC of ZGA (st9) *X. tropicalis* embryos injected with control Mo or *Tet2/3* Mo **(C)** or control- (Ppl) or *Tet2/3 mRNA* **(D)**. n=5 biological replicates +/- SD. P-value calculated with unpaired t-test.

**(E)** Pol III ChIP-qPCR at st18 *X. tropicalis* embryos. Values are normalized to Pol III occupancy at *tRNA-Gly*. Data are mean of n=3 independent ChIP pulldown samples +/- SD.

**(F)** TfIIIc ChIP-qPCR at st18 *X. tropicalis* embryos. Values are normalized to Pol III occupancy at *tRNA-Gly*. Data are mean of n=3 independent ChIP pulldown samples +/- SD.

**(G)** LC-MS/MS analysis of genomic 5mC, 5fC and 5caC of st9, 28 and 38 *X. tropicalis* embryos microinjected with control and *Tdg* Mo. Data are fold change over the control. st9: n=3, st28 & 38: n=2 biological replicates +/- SD.

**(H)** Phenotypes of st27 *X. tropicalis* embryos microinjected with control or *Tdg* Mo.

**(I)** LC-MS/MS analysis of genomic 5mC and 5fC from ZGA (st9) *X. tropicalis* embryos microinjected with control (Ppl) and TDG mRNA. Data are fold change over the control. n=4 biological replicates +/- SD.

**(J)** LC-MS/MS analysis of genomic 5mC and 5fC from ZGA (st9) *X. tropicalis* embryos microinjected with control (Ppl) and dTDG mRNA. Data are fold change over the control. n=5 biological replicates +/- SD.

**(K)** Immunoblot from ZGA (st9) *X. tropicalis* embryos injected with control (Ppl), TDG or dTDG mRNA. α-tubulin is shown for reference.

**(L)** Phenotypes of *X. tropicalis* tadpoles microinjected with control (Ppl) or dTDG mRNA.

**(M)** Pol II ChIP-qPCR at ZGA (st9) *X. tropicalis* embryos microinjected with control (Ppl) or TDG mRNA. n=3 biological replicates +/-SD. All comparisons between control and TDG mRNA embryos are non-significant, as calculated with 2-way ANOVA.

**Supplementary Figure 5: 5fC promotes expression of *tRNA* genes.**

**(A)** *tRNA-Gly* northern blot of *X. tropicalis* embryos microinjected with α-amanitin, *Tet2/3* morholino (*Tet2/3* KD) or TDG mRNA (TDG OE), and their respective controls (Co), at st9 and 10. *tRNA-Gly* levels are normalized to *5.8S* *rRNA* and represented as fold change over the control. α-amanitin, *Tet2/3* KD st10: n=3, TET KD st9, TDG OE st9-10: n=4 biological replicates +/- SD. P-values calculated with paired t-test. α-amanitin: \*\*=0.0052; *Tet2/3* KD: \*=0.048.

**(B)** *tRNA-iMet* and *tRNA-Gly* northern blot of *X. tropicalis* embryos microinjected with control (Ppl) and catalytically inactive dTDG mRNA (dTDG OE), at st9 and 10. *tRNA-iMet* and *tRNA-Gly* levels are normalized to *5.8S* *rRNA* and represented as fold-change over the control. n=4 biological replicates +/- SD. Paired t-test shows no significance for all differences between Co and dTDG OE.

**(C, D)** *U6* *snRNA* and oocytic type *5S* *rRNA* RT-qPCR of ZGA (st9) *X. tropicalis* embryos microinjected with control and *Tet2/3* Morpholinos (*Tet2/3* KD) **(C)** or control (*Ppl*) and TDG mRNA (TDG OE) **(D)**. *U6* and *5S* expression is normalized to *5.8S* *rRNA* levels and represented as fold change over the control. n=3 biological replicates +/- SD. Unpaired t-test shows no significance for all differences between Co and treatment.

**(E, F)** LC-MS/MS analysis of plasmids 5mC **(E)** and 5fC **(H)** levels after strand exchange with unmodified (C) and formylated (5fC) oligos, after mock- and *in vitro* TDG treatment.

**Supplementary Figure 6: α-5fC antibody control in mESCs.**

Immunofluorescence microscopy (IF) of 5mC, 5fC and cell membrane (CellMask plasma membrane stain) in WT and TET TKO mESCs. Scale bar, 10 µm.